

Measurement of the Direct CP Asymmetry in  
 $b \rightarrow s\gamma$  via  
Sum of Exclusive  $B$  Meson Decays using  
the *BABAR* Detector

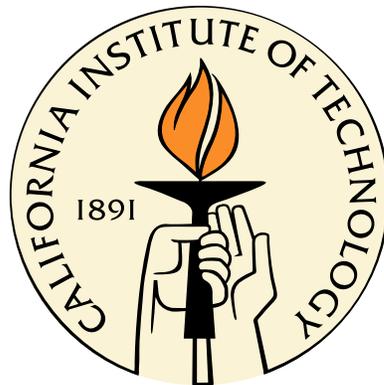
Thesis by

Piti Ongmongkolkul

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2013

(Defended May 1, 2013)



# Acknowledgements

First of all thanks Papa and Mama, without whom my existence would be very questionable assuming causality, and all my sisters and brothers growing up with me and making my childhood awesome.

I would like to express my deepest gratitude to David Hitlin, my advisor, and Frank Porter for their excellent guidance and the extraordinary amount of freedom they gave me. The atmosphere they created allows me to try new things, to take risk and try to be creative without the pressure. Of course, a lot of stuff I tried turned out to be just plain stupid. But, once in a while, it ended in quite an impressive feat.

Thanks to David Doll for his work in this thesis and his wisdom; our office feels empty since he graduated. Thanks to Chih-hsiang Cheng, to whom I usually defer questions, for his help on this thesis; his vast amount of knowledge impresses me every time. Also, I would like to thank Bertrand Echenard for his advice and insight on the analysis.

My friends also contribute significantly in the form of keeping me entertained and being partners in my culinary adventure. This includes but is not limited to Noon, Au, Manow, Kob, Tara, Ped, Best, Ploy, Kook, Nun, Nok, Peera, Kay, Nuknik, Pinto, Tenth, Mee, Daniel and Heywood. My time at Caltech would not be as enjoyable without them.

# Abstract

We perform a measurement of direct  $CP$  violation in  $b \rightarrow s\gamma$ ,  $A_{CP}$ , and the measurement of a difference between  $A_{CP}$  for neutral  $B$  and charged  $B$  mesons,  $\Delta A_{X_s\gamma}$ , using 429  $fb^{-1}$  of data recorded at the  $\Upsilon(4S)$  resonance with the *BABAR* detector.  $B$  mesons are reconstructed from 16 exclusive final states. Particle identification is done using an algorithm based on Error Correcting Output Code with an exhaustive matrix. Background rejection and best candidate selection are done using two decision tree-based classifiers. We found  $A_{CP} = 1.73\% \pm 1.93\% \pm 1.02\%$  and  $\Delta A_{X_s\gamma} = 4.97\% \pm 3.90\% \pm 1.45\%$  where the uncertainties are statistical and systematic respectively. Based on the measured value of  $\Delta A_{X_s\gamma}$ , we determine a 90% confidence interval for  $\text{Im} \frac{C_{8g}}{C_{7\gamma}}$ , where  $C_{7\gamma}$  and  $C_{8g}$  are Wilson coefficients for New Physics amplitudes, at  $-1.64 < \text{Im} \frac{C_{8g}}{C_{7\gamma}} < 6.52$ .

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background</b>	<b>4</b>
2.1 <i>CP</i> Violation Primer . . . . .	4
2.2 Effective Field Theory Primer . . . . .	7
2.3 Direct <i>CP</i> Violation in $B \rightarrow X_s \gamma$ . . . . .	13
<b>3 PEP-II and the <i>BABAR</i> Detector</b>	<b>15</b>
3.1 Introduction . . . . .	15
3.2 PEP-II . . . . .	17
3.3 Silicon Vertex Tracker(SVT) . . . . .	19
3.4 Drift Chamber(DCH) . . . . .	20
3.5 Detector of Internally Reflected Cherenkov Light . . . . .	22
3.6 Electromagnetic Calorimeter(EMC) . . . . .	24
3.7 Instrumented Flux Return(IFR) . . . . .	25
3.8 Trigger . . . . .	27
<b>4 Particle Identification(PID)</b>	<b>28</b>
4.1 Introduction . . . . .	28
4.2 Error-Correcting Output Code(ECOC) . . . . .	29
4.3 Exhaustive Matrix . . . . .	31

4.4	<i>BABAR</i> Implementation . . . . .	34
4.5	Training Sample . . . . .	36
4.6	Performance . . . . .	37
<b>5</b>	<b>Measurement of <math>A_{CP}</math></b>	<b>39</b>
5.1	$B$ Reconstruction and Event Selection . . . . .	39
5.1.1	Overview . . . . .	39
5.1.2	$B$ Reconstruction . . . . .	40
5.1.3	Event Selection . . . . .	41
5.1.3.1	Best Candidate Selection . . . . .	42
5.1.3.2	Background Rejecting Classifier, BRC . . . . .	44
5.1.4	Cut Optimization . . . . .	47
5.2	$A_{CP}$ Extraction . . . . .	50
5.2.1	Fitting Procedure . . . . .	50
5.2.2	Peaking Background Dilution . . . . .	52
5.2.3	The effect on Statistical Uncertainty of Peaking Background Contamination . . . . .	54
5.2.4	Toy Study . . . . .	55
5.2.4.1	Fitting Strategy . . . . .	56
5.2.4.2	Sensitivity, Bias and Robustness at Non-zero $A_{CP}$ . . . . .	58
5.3	Detector $A_{CP}$ . . . . .	64
5.3.1	Source of Detector $A_{CP}$ . . . . .	64
5.3.2	Detector $A_{CP}$ Subtraction . . . . .	67
5.3.3	Measurement of Detector $A_{CP}$ . . . . .	69
5.3.3.1	Sideband $A_{CP}$ . . . . .	69
5.3.3.2	Systematic Uncertainty Associated with Detector $A_{CP}$ . . . . .	71
5.4	Results . . . . .	76
	<b>Bibliography</b>	<b>83</b>

<b>A Particle Identification Appendix</b>	<b>87</b>
A.1 PID Variables . . . . .	87
A.2 PID Variable Importance . . . . .	88
A.3 PID Performance and Comparison to Previous Generation . . . . .	93
A.3.1 Kaon . . . . .	94
A.3.1.1 Super Loose Kaon Selector . . . . .	94
A.3.1.2 Very Loose Kaon Selector . . . . .	95
A.3.1.3 Loose Kaon Selector . . . . .	96
A.3.1.4 Tight Kaon Selector . . . . .	97
A.3.1.5 Very Tight Kaon Selector . . . . .	98
A.3.1.6 Super Tight Kaon Selector . . . . .	99
A.3.2 Pion . . . . .	100
A.3.2.1 Super Loose Pion Selector . . . . .	100
A.3.2.2 Very Loose Pion Selector . . . . .	101
A.3.2.3 Loose Pion Selector . . . . .	102
A.3.2.4 Tight Pion Selector . . . . .	103
A.3.2.5 Very Tight Pion Selector . . . . .	104
A.3.2.6 Super Tight Pion Selector . . . . .	105
A.3.3 Proton . . . . .	106
A.3.3.1 Super Loose Proton Selector . . . . .	106
A.3.3.2 Very Loose Proton Selector . . . . .	107
A.3.3.3 Loose Proton Selector . . . . .	108
A.3.3.4 Tight Proton Selector . . . . .	109
A.3.3.5 Very Tight Proton Selector . . . . .	110
A.3.3.6 Super Tight Proton Selector . . . . .	111
A.3.4 Electron . . . . .	112
A.3.4.1 Super Loose Electron Selector . . . . .	112
A.3.4.2 Very Loose Electron Selector . . . . .	113
A.3.4.3 Loose Electron Selector . . . . .	114
A.3.4.4 Tight Electron Selector . . . . .	115

A.3.4.5	Very Tight Electron Selector . . . . .	116
A.3.4.6	Super Tight Electron Selector . . . . .	117
<b>B</b>	<b>Data Set</b>	<b>118</b>
<b>C</b>	<b><math> \Delta E </math>-minimization as a Baseline for SSC</b>	<b>121</b>
<b>D</b>	<b><math>\pi^0</math> Veto</b>	<b>122</b>
<b>E</b>	<b>SSC and BRC Variables</b>	<b>129</b>
<b>F</b>	<b>Final Cut Optimization Miscellaneous Tables and Figures</b>	<b>132</b>
<b>G</b>	<b>Peaking <math>B\bar{B}</math> Background</b>	<b>138</b>
<b>H</b>	<b>Toy Generation Procedure</b>	<b>140</b>
H.1	$K^*$ Region . . . . .	140
H.2	Full Mass Range . . . . .	146
<b>I</b>	<b>Detector <math>A_{CP}</math> Appendix</b>	<b>151</b>
<b>J</b>	<b>Unblind Results Appendix</b>	<b>155</b>
J.1	$K^*$ region . . . . .	155
J.1.1	All $B$ . . . . .	155
J.1.2	Charged $B$ . . . . .	157
J.1.3	Neutral $B$ . . . . .	159
J.2	Full Mass Range . . . . .	161
J.2.1	All $B$ . . . . .	161
J.2.2	Charged $B$ . . . . .	163
J.2.3	Neutral $B$ . . . . .	165
<b>K</b>	<b>Detector <math>A_{CP}</math> Appendix</b>	<b>167</b>
<b>L</b>	<b>Distributions for PDF</b>	<b>170</b>

M Effect of Charged/Neutral $B$ Efficiency/Lifetime Difference and Isospin Asymmetry on Total Flavor Asymmetry	172
N $CPT$ and $\Delta\Gamma_d$ Assumptions	177

# List of Figures

1.1	A Feynman diagram of $b \rightarrow s\gamma$ process . . . . .	2
2.1	A diagram showing $B$ meson decays to $X_s$ and a photon. The amplitude can be separated into two parts: long distance (blue) and short distance (red). . . . .	7
2.2	An example diagram of $b \rightarrow s\gamma$ decay. . . . .	8
2.3	Diagram $O_{7\gamma}$ . . . . .	10
2.4	Diagram for effective interaction involved in $B \rightarrow X_s\gamma$ , along with the leading order diagram and its contribution to $B \rightarrow X_s\gamma$ . . . . .	11
2.5	An example of SUSY contribution to $C_7$ . . . . .	14
3.1	Longitudinal cross-section of <i>BABAR</i> detector. . . . .	16
3.2	SLAC site, PEP-II an <i>BABAR</i> detector . . . . .	17
3.3	Integrated luminosity delivered by PEP-II and recorded by the <i>BABAR</i> detector . . . . .	19
3.4	Layout of the SVT microstrip layers. . . . .	20
3.5	Longitudinal cross-section of drift chamber. . . . .	21
3.6	$dE/dx$ for various types of charge particle. . . . .	22
3.7	$\theta_c$ for various types of particle. . . . .	23
3.8	DIRC layout and principle of operation. . . . .	23
3.9	Schematics of EMC layout. . . . .	24
3.10	Layout of IFR and RPC. . . . .	26
3.11	Schematic for each RPC module. . . . .	26

4.1	Comparison of PID efficiency (circles with error bars) and $\pi$ as $K$ fake rate (triangles with error bars) from ECOC-based PID selector using the exhaustive matrix (red), 1 vs 1 matrix (green) and a likelihood-based PID selector (blue). The $\pi$ as $K$ fake rate is multiplied by 4 for visibility.	38
5.1	Normalized distribution of maximum SSC response for events in which the best candidate was correctly selected (black) and incorrectly selected (red) . . . . .	44
5.2	Comparison between the signal rate and fake rate for SSC and $\Delta E$ minimization method . . . . .	45
5.3	The response of the BRC for background candidates with $m_{ES} > 5.265$ (dashed), background candidates with $m_{ES} < 5.265$ (dotted), and signal candidates (solid line) for run 3 MC. We also compare offpeak data (red) to continuum MC (green) for $m_{ES} < 5.265$ (solid) and $m_{ES} > 5.265$ (dashed). . . . .	47
5.4	$X_s$ mass distribution before(a) and after (b) the optimal cut. . . . .	49
5.5	Example fit to a toy sample using the fix alpha strategy. The fit to the $b$ $m_{ES}$ distribution is shown on the left and the fit to the $\bar{b}$ $m_{ES}$ distribution is shown on the right. The continuum distribution is shown with light blue, the peaking distribution is shown in red and the sum of the two is shown in green. Monte Carlo components are shown in dashed lines. The toys used in this figure are from all $B$ charges and full $X_s$ mass range.	52
5.6	Bias and sensitivity of each fitting strategy. The description of each fitting strategy can be found in Table 5.5. The blue error bars represent the 50% percentile of the MINOS error in the MINUIT package and the green error bars represent the 75% percentile of the MINOS error. . . .	57
5.7	Sensitivity and bias of signal $A_{CP}$ at various values of signal $A_{CP}$ for all $B$ (top), charged $B$ (mid) and neutral $B$ (bottom) for the full mass range. Blue points with error bars indicate the central value of the fitted $A_{CP}$ and its sensitivity. The true signal $A_{CP}$ are shown as red crosses.	60

5.8	Sensitivity and bias of signal $A_{CP}$ at various values of signal $A_{CP}$ for all $B$ (top), charged $B$ (mid) and neutral $B$ (bottom) for the $K^*$ range. Blue points with error bars indicate the central value of fitted $A_{CP}$ and its sensitivity. True signal $A_{CP}$ are shown as red crosses. . . . .	61
5.9	Fitted $A_{CP}$ of toy with 0 $A_{CP}$ and varying cross feed $A_{CP}$ for full $X_s$ mass range. From top to bottom: all $B$ , charged $B$ and neutral $B$ . The error bars shown are statistical sensitivities. . . . .	62
5.10	Fitted $A_{CP}$ of toy with 0 $A_{CP}$ and varying crossfeed $A_{CP}$ for $K^*$ mass region. From top to bottom: all $B$ , charged $B$ and neutral $B$ . The error bars shown are statistical sensitivities. . . . .	63
5.11	PID efficiencies for $K^+$ and $K^-$ . The efficiency, uncertainties on efficiencies, and efficiency ratios shown are calculated after summing over $\theta$ bins. These uncertainties on efficiencies are vastly underestimated due to the fact that momentum bin is too wide compared to variation of efficiencies, especially at low momentum. . . . .	65
5.12	PID efficiencies for $\pi^+$ and $\pi^-$ . The efficiency, uncertainties on efficiencies, and efficiency ratios shown are calculated after summing over $\theta$ bins. This ratio of efficiencies is much flatter than that for kaons. . . .	66
5.13	Normalized charged $K$ momentum distribution for the full mass range for $b$ (left) and $\bar{b}$ (right) using the all $B$ sample. The distribution of the sideband region is shown in blue while the distribution of the peaking region is shown in green. The distribution shown is, of course, after the efficiency penalty. But, since the distributions after selection are different, we can infer that the distributions before selection is different. The same plot for other $X_s$ mass range and other $B$ sample can be found in Appendix I. . . . .	70
5.14	Fitted $m_{ES}$ distribution for full mass range using all $B$ sample. . . . .	76

5.15	68% and 90% confidence interval for $\text{Im} \frac{C_8}{C_7}$ for possible values of $\tilde{\Lambda}_{78}$ . The irregular shape comes from the fact that it is a product of a range and a number. Care must be taken when performing a global fit with other observables. . . . .	79
5.16	68% and 90% confidence interval for $\text{Im} \frac{C_8}{C_7}$ for the given value of $\tilde{\Lambda}_{78}$ . . . . .	80
5.17	68% and 90% confidence interval for $\frac{r_8}{r_7}$ and $\theta_8 - \theta_7$ from all possible value of $\tilde{\Lambda}_{78}$ . . . . .	81
5.18	68% and 90% confidence interval for $\frac{r_8}{r_7}$ and $\theta_8 - \theta_7$ for a given value of $\tilde{\Lambda}_{78}$ . . . . .	82
A.1	Variable importance of $K$ vs $\pi, p, e$ classifier . . . . .	89
A.2	Variable importance of $K, \pi$ vs $p, e$ classifier . . . . .	89
A.3	Variable importance of $K, \pi, p$ vs $e$ classifier . . . . .	90
A.4	Variable importance of $K, \pi, e$ vs $p$ classifier . . . . .	90
A.5	Variable importance of $K, p$ vs $\pi, e$ classifier . . . . .	91
A.6	Variable importance of $K, e$ vs $\pi, p$ classifier . . . . .	91
A.7	Variable importance of $K, p, e$ vs $\pi$ classifier . . . . .	92
A.8	Kaon selection efficiencies and misidentification rates for Super Loose kaon selector as $K$ of $\pi$ (a), $p$ (b), and $e$ (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . .	94
A.9	Kaon selection efficiencies and misidentification rates for Very Loose kaon selector as $K$ of $\pi$ (a), $p$ (b), and $e$ (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . .	95

- A.10 Kaon selection efficiencies and misidentification rates for Loose kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 96
- A.11 Kaon selection efficiencies and misidentification rates for Tight kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 97
- A.12 Kaon selection efficiencies and misidentification rates for Very Tight kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 98
- A.13 Kaon selection efficiencies and misidentification rates for Super Tight kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 99
- A.14 Pion selection efficiencies and misidentification rates for Super Loose pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 100

- A.15 Pion selection efficiencies and misidentification rates for Very Loose pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. 101
- A.16 Pion selection efficiencies and misidentification rates for Loose pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . 102
- A.17 Pion selection efficiencies and misidentification rates for Tight pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . 103
- A.18 Pion selection efficiencies and misidentification rates for Very Tight pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. 104
- A.19 Pion selection efficiencies and misidentification rates for Super Tight pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 105

- A.20 Proton selection efficiencies and misidentification rates for Super Loose proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 106
- A.21 Proton selection efficiencies and misidentification rates for Very Loose proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 107
- A.22 Proton selection efficiencies and misidentification rates for Loose proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 108
- A.23 Proton selection efficiencies and misidentification rates for Tight proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 109
- A.24 Proton selection efficiencies and misidentification rates for Very Tight proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 110

- A.25 Proton selection efficiencies and misidentification rates for Super Tight proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 111
- A.26 Electron selection efficiencies and misidentification rates for Super Loose electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 112
- A.27 Electron selection efficiencies and misidentification rates for Very Loose electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 113
- A.28 Electron selection efficiencies and misidentification rates for Loose electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 114
- A.29 Electron selection efficiencies and misidentification rates for Tight electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . . 115

A.30	Electron selection efficiencies and misidentification rates for Very Tight electron selector as $e$ of $K$ (a), $\pi$ (b), and $p$ (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . .	116
A.31	Electron selection efficiencies and misidentification rates for Super Tight electron selector as $e$ of $K$ (a), $\pi$ (b), and $p$ (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively. . . . .	117
D.1	Sources of background high energy photons. The three stack plots for each type of source represent the following: the left stack is without a $\pi^0$ veto; the middle is after a mass window cut $\in [0.11, 0.15]$ GeV; the right is with a cut on the $\pi^0$ classifier described in the text (the cut location is 0.2 for demonstrative purposes only). . . . .	123
D.2	Normalized distribution of variables used in $\pi^0$ veto classifier. . . . .	124
D.3	Truth match probability of a $\pi^0$ candidate to be a true $\pi^0$ . . . . .	126
D.4	The classifier response to a $\pi^0$ candidate, depending on the candidate's minimum photon energy vs. invariant mass. . . . .	127
D.5	Maximum classifier response for each high energy photon as described in the text. . . . .	128
E.1	Correlation between variables used in training the SSC. In each of the 2D histograms, the red indicates a higher concentration of correctly reconstructed candidates and blue a lower concentration of correctly reconstructed candidates. The bottom row plots are projections of each variable for mis-reconstructed candidates (blue) and correctly reconstructed candidates (red). . . . .	130

E.2	Correlation plot of variables used in BRC training. B Momentum Flow 2-17 is omitted from the plot. In the 2D histogram red represents more concentration of signal events and blue represents less concentration of signal events. . . . .	131
F.1	Comparison between 3 methods of making optimized cut on SSC and BRC. Classifier method, Global Mass Range and Individual Mass Range cut are shown in blue, green red, respectively. . . . .	134
F.2	Mode distribution before and after the optimal cut . . . . .	137
G.1	$B^+ \rightarrow D^+\pi^0$ is doubly Cabbibo suppressed. $D^+ \rightarrow K^+\pi^0$ is also suppressed analogously. . . . .	139
H.1	Fit to MC peaking $B\bar{B}$ subsample for $K^*$ mass region. . . . .	141
H.2	Fit to MC continuum subsample for $K^*$ mass region. . . . .	142
H.3	Fit to MC Xfeed subsample for $K^*$ mass region. . . . .	143
H.4	Fit to MC signal subsample for $K^*$ mass region . . . . .	144
H.5	Fit to MC peaking $B\bar{B}$ and XFeed subsample for $K^*$ mass region. . . .	145
H.6	Fit to MC for peaking $B\bar{B}$ subsample. . . . .	146
H.7	Fit to MC for continuum subsample. . . . .	147
H.8	Fit to MC for xfeed subsample. . . . .	148
H.9	Fit to MC for peaking signal subsample. . . . .	149
H.10	Fit to MC for combined peaking $B\bar{B}$ and XFeed subsample. . . . .	150
I.1	Normalized charged $K$ momentum distribution for full mass range for $b$ (left) and $\bar{b}$ (right) using all $B$ sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green. . . . .	151
I.2	Normalized charged $K$ momentum distribution for full mass range for $b$ (left) and $\bar{b}$ (right) using neutral $B$ sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green. . . . .	152

I.3	Normalized charged $K$ momentum distribution for full mass range for $b$ (left) and $\bar{b}$ (right) using charged $B$ sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green. . . . .	152
I.4	Normalized charged $K$ momentum distribution for $K^*$ mass range for $b$ (left) and $\bar{b}$ (right) using all $B$ sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green. . . . .	153
I.5	Normalized charged $K$ momentum distribution for $K^*$ mass range for $b$ (left) and $\bar{b}$ (right) using neutral $B$ sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green. . . . .	153
I.6	Normalized charged $K$ momentum distribution for $K^*$ mass range for $b$ (left) and $\bar{b}$ (right) using charged $B$ sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green. . . . .	154
J.1	Fitted $m_{ES}$ distribution for $K^*$ mass range using all $B$ sample. . . . .	155
J.2	Fitted $m_{ES}$ distribution for $K^*$ mass range using charged $B$ sample. . . . .	157
J.3	Fitted $m_{ES}$ distribution for $K^*$ mass range using neutral $B$ sample. . . . .	159
J.4	Fitted $m_{ES}$ distribution for full mass range using all $B$ sample. . . . .	161
J.5	Fitted $m_{ES}$ distribution for full mass range using charged $B$ sample. . . . .	163
J.6	Fitted $m_{ES}$ distribution for full mass range using neutral $B$ sample. . . . .	165
K.1	$K^-$ nucleon cross section[1] . . . . .	168
K.2	$K^+$ nucleon cross section[1] . . . . .	169
L.1	Argus Distribution. $\chi$ is Argus slope. $p$ is Argus Power. $m_0$ is resonance mass. And, $A$ is the normalization. . . . .	170
L.2	CrystalBall Distribution. . . . .	170

- L.3 Novosibirsk Distribution.  $m_0$  is the peak location.  $\sigma$  is the width of the distribution.  $t$  is tail of the distribution . . . . . 170
- L.4 Cruijff Distribution. A bifurcated gaussian function with second order polynomial correction to width of each side.  $m_0$  is the peak location.  $\sigma_L$  and  $\sigma_R$  are the width for each side of the distribution.  $\alpha_L$  and  $\alpha_R$  are the width correction factors for each side of the distribution.  $A$  is the normalization factor. . . . . 171

# List of Tables

2.1	Contribution of New Physics to $C_7$ and $C_8$ with new physics parameterized by $\xi = 3 \frac{C_7^{\text{new}}}{C_8^{\text{new}}}$ and categorized by the particle in penguin diagrams[2]	14
3.1	Cross section of relevant processes within <i>BABAR</i> detector coverage. . .	18
4.1	Indicator matrix of a multiclass classifier from combining two binary classifiers. $t_\alpha$ and $t_\beta$ represents how the given class is treated in the training sample for classifiers $\alpha$ and $\beta$ respectively. 1 means signal and 0 means background. . . . .	30
4.2	Hamming distances from comparing the output of $\alpha$ and $\beta$ that is 1 and 0, respectively, with indicator matrix given in Table 4.1 . . . . .	31
4.3	When using only two classifiers, we only need one classifier to make a mistake to obtain a wrong class. . . . .	31
4.4	All possible classifiers for the four classes problem. There are only 7 distinct usable classifiers. . . . .	32
4.5	Exhaustive matrix for 4 classes. . . . .	32
4.6	Hamming distances of the exhaustive matrix when 1 classifier makes mistake. . . . .	33
4.7	Hamming distances of the exhaustive matrix when 2 classifiers make mistake. . . . .	33

4.8	Hamming distances of the exhaustive matrix when 3 classifiers make mistakes. Note that there is a tie between class $A$ , $B$ and $C$ . In our implementation of ECOC, the output from each classifier is a real number and we use the sum of squared difference to extend the Hamming distance into a real number. There are no ties in our implementation. . . . .	33
4.9	The exhaustive matrix is the indicator matrix used in the ECOC-based selector. Each entry indicates whether the training sample of the given type should be treated as signal(1) or background(-1). . . . .	35
4.10	The selection criteria for each type of particle in ECOC-based selector.	36
5.1	The 38 modes we reconstruct in this analysis; BiType identifies the numeric value we assign to each mode for bookkeeping; charge conjugation is implied. . . . .	42
5.2	The 16 modes used for $A_{CP}$ analysis; BiType identifies the numeric value we assign to each mode for bookkeeping; charge conjugation is implied.	43
5.3	Optimal cut value. . . . .	48
5.4	Value of $S$ , $\Delta$ and $\delta A$ for the $K^*$ region and the whole mass range. . .	54
5.5	Description of each fitting strategy. The parameters that are not listed in any of the two columns are floated separately. . . . .	58
5.6	Expected Sensitivity for $A_{CP}$ at 0% signal $A_{CP}$ and 0% background $A_{CP}$ .	59
5.7	Sideband $A_{CP}$ ( $\delta_{\text{side}}$ ) and associated statistical uncertainty for each $X_s$ mass region and $B$ type. . . . .	70
5.8	Summary of $A_{CP}$ results. . . . .	76
B.1	The run-by-run integrated luminosity of the data. . . . .	118
B.2	The run-by-run number of MC events, cross section, equivalent luminosity, and weighting factor for each background mode. . . . .	119
B.3	The run by run number of signal MC and luminosity weighting factors.	120
C.1	The optimal requirements in each mass bin for the $\frac{ \Delta E }{\sigma_E}$ selected candidate analysis. . . . .	121

D.1	Summary of $\pi^0$ veto classifier. It should be underscored that this classifier is identifying the high energy photons that originate from $\pi^0$ s as “signal” . . . . .	125
F.1	Comparison of number of signal, number of background, and precision for each mode reduction strategy(mode-before/mode-after), pion and kaon tightness level. All the numbers presented here are from the optimal cuts. We also present the comparison of different ways of combining SSC and BRC output to make the final selection. Precision shown here is the total precision for for events with $m_{ES} > 5.27$ GeV (including precision for Mass Range Individual column). . . . .	133
F.2	Expected number of events with $m_{ES} > 5.24$ GeV. The uncertainty given here is calculated from $\sqrt{\sum weight}$ . It may appear that in some cases mass range individual gives lower precision for certain mass range. This is because the PID choice that gives best global FOM for each one of the methods is different. The strategy we selected is Mass Range Global. . . . .	135
F.3	Expected number of events with $m_{ES} > 5.27$ GeV. For each strategy. The uncertainty given here is calculated from $\sqrt{\sum weight}$ . Note that in some of the mass ranges the Mass Range Individual method may give a lower precision than the Mass Range Global method. This is due to the difference in PID selectors. . . . .	136
G.1	The 16 modes we used for $A_{CP}$ analysis. And some examples of peaking $B\bar{B}$ background source for each mode. Charged conjugation is implied.	138
H.1	PDF for each type of sample used for generating toy. . . . .	140
J.1	Fitted shape parameters for $K^*$ mass range with all $B$ sample. . . . .	156
J.2	Correlation among shape parameters for $K^*$ mass range with all $B$ sample.	156
J.3	Fitted shape parameters for $K^*$ mass range with charged $B$ sample. . .	158

J.4	Correlation among shape parameters for $K^*$ mass range with charged $B$ sample. . . . .	158
J.5	Fitted shape parameters for $K^*$ mass range with neutral $B$ sample. . .	160
J.6	Correlation among shape parameters for $K^*$ mass range with neutral $B$ sample. . . . .	160
J.7	Fitted shape parameters for full mass range with all $B$ sample. . . . .	162
J.8	Correlation among shape parameters for full mass range with all $B$ sample.	162
J.9	Fitted shape parameters for full mass range with charged $B$ sample. . .	164
J.10	Correlation among shape parameters for full mass range with charged $B$ sample. . . . .	164
J.11	Fitted shape parameters for full mass range with neutral $B$ sample. . .	166
J.12	Correlation among shape parameters for full mass range with neutral $B$ sample. . . . .	166

# Chapter 1

## Introduction

One of the great mysteries of modern physics is the matter-antimatter asymmetry. For most interactions, matter and antimatter seems to be produced in equal amounts. For example, pair production always produces an electron and a positron; one particle and one antiparticle. But, in the vicinity around us there are many electrons and no positrons. This is called the problem of Baryogenesis.

Sakharov pointed out that one of the necessary ingredients for Baryogenesis is charge conjugation and parity ( $CP$ ) violation[3]. This means some observables, e.g. decay rates, must be different for particle and the corresponding antiparticle. The standard model (SM) does provide  $CP$  violation in the weak interaction via the phase of Cabibo-Kobayashi-Maskawa (CKM) matrix[4][5]. However,  $CP$  violation in CKM matrix is not enough to explain the amount of asymmetry we see[6][7][8]; this hints that there must be some new physics generating extra  $CP$  violation.

$CP$  violation has been studied extensively for the  $B$  meson system for both direct (non mixing-induced)  $CP$  violation and indirect (mixing-induced)  $CP$  violation, giving us a set of very stringent test of the Standard Model (SM). So far, the SM has passed the test. This thesis focuses on a measurement of direct  $CP$  violation in a rare  $b$  decay,  $b \rightarrow s\gamma$ <sup>1</sup>. This decay is forbidden at tree level in the SM. Figure 1.1 shows a Feynman diagram of this process.

Theoretically, calculating the properties of the  $b$  quark can be done more precisely than the properties of  $B$  mesons. Yet, experimentally, we can only directly measure

---

<sup>1</sup>Charge conjugation is implied throughout the document unless specified.

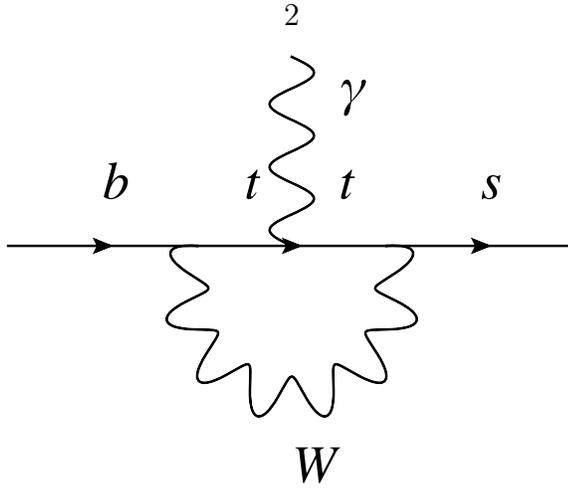


Figure 1.1: A Feynman diagram of  $b \rightarrow s\gamma$  process

the properties of  $B$  mesons. The technique we used to infer the properties of  $b \rightarrow s\gamma$  decay from directly measured property of exclusive  $B$  decays to  $s$  quark-containing final states ( $X_s$ ) and a photon is summing up as many exclusive decay modes of  $B \rightarrow X_s\gamma$  as possible.

The goal of our analysis is to measure the  $A_{CP}$  of the process  $b \rightarrow s\gamma$ , defined as

$$A_{CP} = \frac{\Gamma_{b \rightarrow s\gamma} - \Gamma_{\bar{b} \rightarrow \bar{s}\gamma}}{\Gamma_{b \rightarrow s\gamma} + \Gamma_{\bar{b} \rightarrow \bar{s}\gamma}} \quad (1.1)$$

and the difference of the  $A_{CP}$  of the process  $B \rightarrow X_s\gamma$  for charged  $B$  and for neutral  $B$ ,

$$\Delta A_{X_s\gamma} = A_{X_s^-\gamma} - A_{X_s^0\gamma}, \quad (1.2)$$

both of which are expected to be small in SM[9][10][11].

The first quantity has been measured by many experiments such as *BABAR*[12], *Belle*[13] and *CLEO*[14]. The current world average[1] is

$$A_{CP} = -0.8\% \pm 2.9\%.$$

The Standard Model prediction for  $A_{CP}$ , which is dominated by long distance effects, is  $-0.6\% < A_{CP}^{SM} < 2.8\%$  [11]. The second quantity,  $\Delta A_{X_s\gamma}$ , is suggested by [11] as a sensitive new probe of physics beyond the SM. This observable is especially important

as it can be used to place constraint on the Wilson coefficient  $C_8$  which do not yet have a strong constraint. We will be presenting the first measurement of  $\Delta A_{X_s\gamma}$ .

In the following sections, we will describe the theoretical background for  $CP$  asymmetry. Then, we will describe the *BABAR* detector we use for this measurement. We will then digress a little bit to the subject of how we identify particle type, an important component of the measurement. We will then present our measurement of direct asymmetry observables using  $471 \times 10^6$   $B\bar{B}$  pairs starting with  $B$  meson reconstruction, signal event selection, the  $A_{CP}$  extraction procedure, the results, and their implication for wilson coefficients of new physics amplitudes.

# Chapter 2

## Background

### 2.1 $CP$ Violation Primer

$CP$  violation in general needs two ingredients: two complex phases with different transformation under  $CP$  and interference. Let us show why this is true. Naively, we would not expect to see any difference in any observable from particle and antiparticle no matter what the phase change under a transformation is since observables is the square of some amplitudes. The phase of a complex number and its complex conjugate thus always cancels. For example, let us consider the decay  $x \rightarrow z$  with amplitude

$$A_{x \rightarrow z} = M e^{i\phi}.$$

To find the corresponding amplitude for antiparticle process  $A_{\bar{x} \rightarrow \bar{z}}$ , we apply a  $CP$  transformation to the amplitude. The tricky part is that there are two type of phases: one that change sign under a  $CP$  transformation ( $CP$  odd) such as those from the weak interactions,

$$CP A_{x \rightarrow z}^{\text{weak}} = A_{\bar{x} \rightarrow \bar{z}}^{\text{weak}} = M e^{-i\phi},$$

and the the other that is invariant under  $CP$  such as those from the strong interaction ( $CP$  even)

$$CP A_{x \rightarrow z}^{\text{strong}} = A_{\bar{x} \rightarrow \bar{z}}^{\text{strong}} = M e^{i\phi}.$$

By itself nothing interesting happens, since the observable is the amplitude squared;

whether the sign flips under  $CP$  or not, the square still stays unchanged:

$$|A_{x \rightarrow z}^{\text{weak}}|^2 = M^2 = |A_{\bar{x} \rightarrow \bar{z}}^{\text{weak}}|^2, \quad (2.1)$$

$$|A_{x \rightarrow z}^{\text{strong}}|^2 = M^2 = |A_{\bar{x} \rightarrow \bar{z}}^{\text{strong}}|^2, \quad (2.2)$$

which means there is no  $CP$  violation.

The calculation gets more interesting if there is more than one way that  $x$  can turn into  $z$  and each one involves a different interaction. For example, if  $x$  can go to  $z$  via  $x \xrightarrow{\text{strong}} z$  and  $x \xrightarrow{\text{weak}} z$ . The amplitude for  $x$  turning into  $z$  is the sum of amplitudes of all possible ways of  $x$  turning into  $z$ . This means

$$A_{x \rightarrow z} = A_{x \xrightarrow{\text{strong}} z} + A_{x \xrightarrow{\text{weak}} z} = M_1 e^{i\phi_{\text{strong}}} + M_2 e^{i\phi_{\text{weak}}}.$$

We apply a  $CP$  transformation to the above expression to obtain the amplitude of  $\bar{x} \rightarrow \bar{z}$ . Remember that the strong phase stays the same while the weak phase flips.

Thus,

$$A_{\bar{x} \rightarrow \bar{z}} = A_{\bar{x} \xrightarrow{\text{strong}} \bar{z}} + A_{\bar{x} \xrightarrow{\text{weak}} \bar{z}} = M_1 e^{i\phi_{\text{strong}}} + M_2 e^{-i\phi_{\text{weak}}}.$$

Now let us look at the square of the two amplitudes. For  $x \rightarrow z$  we have

$$|A_{x \rightarrow z}|^2 = M_1^2 + M_2^2 + M_1 M_2 e^{i(\phi_{\text{strong}} - \phi_{\text{weak}})} + M_1 M_2 e^{i(\phi_{\text{weak}} - \phi_{\text{strong}})} \quad (2.3)$$

$$(2.4)$$

and for  $\bar{x} \rightarrow \bar{z}$  we have

$$|A_{\bar{x} \rightarrow \bar{z}}|^2 = M_1^2 + M_2^2 + M_1 M_2 e^{-i(\phi_{\text{strong}} + \phi_{\text{weak}})} + M_1 M_2 e^{i(\phi_{\text{weak}} + \phi_{\text{strong}})}. \quad (2.5)$$

Subtracting the two yields

$$|A_{x \rightarrow z}|^2 - |A_{\bar{x} \rightarrow \bar{z}}|^2 = M_1 M_2 e^{i(\phi_{\text{strong}} - \phi_{\text{weak}})} + M_1 M_2 e^{i(\phi_{\text{weak}} - \phi_{\text{strong}})} \quad (2.6)$$

$$- M_1 M_2 e^{-i(\phi_{\text{strong}} + \phi_{\text{weak}})} - M_1 M_2 e^{i(\phi_{\text{weak}} + \phi_{\text{strong}})} \quad (2.7)$$

$$= -4 \operatorname{Im} M_1 e^{i\phi_{\text{strong}}} \operatorname{Im} M_2 e^{i\phi_{\text{weak}}}. \quad (2.8)$$

The last expression in general is non zero; thus, we have  $CP$  violation. This is actually a general principle;  $CP$  violation needs 2 ingredients:

1. Two phases that transform differently under  $CP$ .
  - Weak phase,  $CP$ -odd.
  - Strong phase,  $CP$ -even.
2. The interference between those two terms.

## 2.2 Effective Field Theory Primer

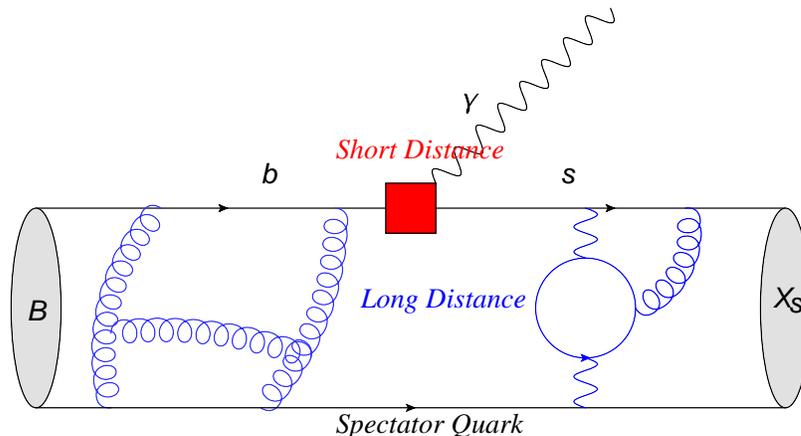


Figure 2.1: A diagram showing  $B$  meson decays to  $X_s$  and a photon. The amplitude can be separated into two parts: long distance (blue) and short distance (red).

In this section, we will explain how one would calculate observables in of  $B$  meson decay in principle, leaving out all the complicated details. The real calculation of the amplitude of decays of  $B$  meson is extremely involved, but the concept is easy enough to understand. The description we use will be very informal to highlight important concepts; for a more rigorous treatment see [15] and [16].

We want to calculate the amplitude of  $B$  decays into a particular final state. Let us consider here the decay of a  $B$  meson to an  $s$  quark-containing resonance  $X_s$  and a photon. As usual in quantum mechanics, we evolve the initial state  $|B\rangle$  with Hamiltonian and see how much it ends up in  $\langle X_s \gamma |$ . Thus the amplitude is given by:

$$A(B \rightarrow X_s \gamma) = \langle X_s \gamma | \exp -iHt | B \rangle, \quad (2.9)$$

where  $H$  is the Hamiltonian describing the amplitude of how one type of matter turns into others. If we do a Taylor expansion on  $\exp -iHt$ , we will obtain the usual notion of summing up amplitudes of all possible ways of going from one state to another.

We could use the Standard Model Hamiltonian. However, as Figure 2.1 shows the decay of a  $B$  meson is extremely complicated and we do not really know what  $|B\rangle$  is and how the quark operators really act on the  $|B\rangle$  state. We will see later on that we

can separate the part we know how to calculate perturbatively from the part that is non-perturbative.

One thing we know about Eq. 2.9 is that if the product of the Hamiltonians does not end up turning a  $b$  quark into  $s$  quark and a photon, the amplitude vanishes. So, let us consider an example of how  $b$  quark can turn into  $s$  quark in the standard model shown in Figure 2.2. The Hamiltonian term of such diagram looks like<sup>1</sup>

$$H_{\tilde{O}_{\tau\gamma}} = (\text{create } s) \times (\text{create } \gamma) \times (\text{propagators} \times \text{couplings}) \times (\text{annihilate } b) \quad (2.10)$$

$$= s\gamma \leftarrow (\text{propagators} \times \text{couplings}) \leftarrow b \quad (2.11)$$

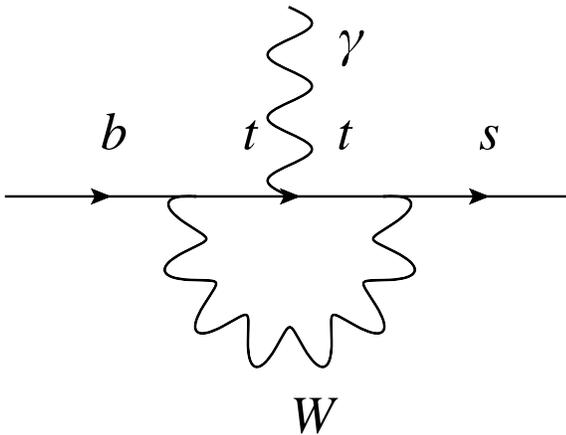


Figure 2.2: An example diagram of  $b \rightarrow s\gamma$  decay.

We can actually do something with the  $W$  propagator. It will have a the term that goes like

$$\frac{1}{p_W^2 - m_W^2}$$

for the  $W$  propagator. Since, in the end, we will be integrating over all the external momenta  $p_b$  and  $p_\gamma$  to find the matrix element, the momentum of the propagator  $p_t$  and  $p_W$  is limited by the energy of the  $b$  quark,  $m_b$ , which is on the order a few GeV, much smaller than  $m_W$  and  $m_t$ . This means we can expand in terms of  $p_W^2/m_W^2$  for  $W$  propagators,

$$\frac{1}{p_W^2 - m_W^2} \rightarrow 1 + \frac{p_W^2}{m_W^2} + \dots,$$

<sup>1</sup>This is a very non-standard and informal notation. We use this to highlight the important idea.

and can do the same for  $t$  propagators. The important part here is that the first order is a constant and the second term is small. We can pull it right off the bra-ket along with the couplings that have to do with heavy particles from the expression. This yields<sup>2</sup>

$$\langle X_s \gamma \left| H_{\tilde{O}_{\tau\gamma}} \right| B \rangle = \text{constant} \times \langle X_s \gamma \left| s\gamma \leftarrow (\text{couplings no } W, t) \leftarrow b \right| B \rangle \quad (2.12)$$

$$= \text{constant} \times \langle X_s \gamma \left| s\gamma \leftarrow \bullet \leftarrow b \right| B \rangle. \quad (2.13)$$

It may look like we did not do anything drastic here. The important point is that inside the bra-ket there are no heavy particle operators. All the effect of heavy particles are factored outside the bra-ket. This means we can separate out the contribution of short distance interaction from the long distance one. The second term does not involve heavy particles such as  $W$  or  $t$ . The details of heavy particles propagators/interactions are in the constant in front<sup>3</sup>. This is a general principle: we can always integrate out propagators heavier than the energy scale involved and separate this contribution from heavy particles which we can calculate perturbatively<sup>4</sup>, from the contribution of the details inside the  $B$  meson which we can not really do much with<sup>5</sup>. We call the theory without heavy particles an Effective Theory.

Dropping the bra and the ket we have

$$H_{\tilde{O}_{\tau\gamma}} = C \times (s\gamma \leftarrow \bullet \leftarrow b) \quad (2.14)$$

$$= C \times O_{\tau\gamma} \quad (2.15)$$

where  $C$  is called the Wilson coefficient and  $O$  corresponds to the operator for the theory without heavy particles. This particular operator is called  $O_{\tau\gamma}$  which represents the interaction of  $b$  turning into  $s$  and  $\gamma$  shown in Figure 2.3.

---

<sup>2</sup>Again, the red circle is a non standard notation. It just signify that the interaction here contains no  $W$  and  $t$ . This matches with red circles in figures.

<sup>3</sup>One actually has to run down the constant to appropriate scale but those are details we will skip. For details see [17].

<sup>4</sup>At least to the first order. The second order is much more complicated[18][19].

<sup>5</sup>Calculating it typically requires one to use optical theorem and to model what  $|B\rangle$  is. For details see [17].

The Wilson coefficients have two important features. First, Wilson coefficients are generic to many decays. For example,  $B \rightarrow X_s l^+ l^-$  also receives a contribution from the diagram shown in Figure 2.3. Other details are hidden inside the second term. This means we can experimentally constrain the Wilson coefficients from many observables. This brings us the second feature, if we have a new physics model, we can calculate its contribution to the Wilson coefficients perturbatively since it involves just short distance interactions. We can then test the new physics contribution against the constraints.

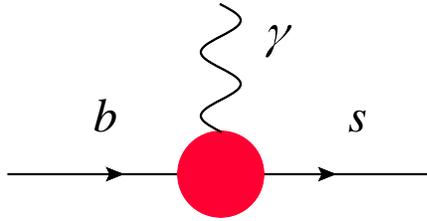


Figure 2.3: Diagram  $O_{7\gamma}$ .

Since the amplitude is the sum of all possible ways  $B$  can turn in to  $X_s \gamma$ , one can enumerate the possible types of interaction in a theory with no heavy particles i.e. Effective Theory. Thus, in general, the Hamiltonian can be written as

$$H_{eff} = \frac{G_F}{\sqrt{2}} \sum_i C_i O_i,$$

where  $G_F$  is Fermi coupling constant and the overall factor is purely conventional. Diagrams of interactions that contribute to  $B \rightarrow X_s \gamma$  at leading order are shown in Figure 2.4. There are only three. Therefore we can write our effective Hamiltonian as

$$H_{eff} = \frac{G_F}{\sqrt{2}} C_1 O_1 + \frac{G_F}{\sqrt{2}} C_{7\gamma} O_{7\gamma} + \frac{G_F}{\sqrt{2}} C_{8g} O_{8g} + \text{higher order}.$$

Using this we can guess the form of  $A_{CP}(B \rightarrow X_s \gamma)$  from what we learned in

Section 2.1; that  $A_{CP}$  arises from interference terms<sup>6</sup>

$$A_{CP} = \frac{\Gamma(b \rightarrow s\gamma) - \Gamma(\bar{b} \rightarrow \bar{s}\gamma)}{\Gamma(b \rightarrow s\gamma) + \Gamma(\bar{b} \rightarrow \bar{s}\gamma)} \quad (2.16)$$

$$= M_{78} \times F_1(C_{8g}, C_{7\gamma}) + M_{17} \times F_2(C_1, C_{7\gamma}) + M_{18} \times F_3(C_{8g}, C_1). \quad (2.17)$$

	Effective Interaction	Leading Order Diagram	Contribution to $B \rightarrow X_s\gamma$
$O_{7\gamma}$			
$O_{8g}$			
$O_1$			

Figure 2.4: Diagram for effective interaction involved in  $B \rightarrow X_s\gamma$ , along with the leading order diagram and its contribution to  $B \rightarrow X_s\gamma$ .

In conclusion, the decay of  $B$  meson involves both short distance and long distance contributions. For the short distance, the momentum of the propagators is limited by the mass of the  $b$  quark, which is much smaller than the mass of the  $W$  and the  $t$  propagator, allowing us to Taylor expand the propagators, integrate them out and separate the interaction that involve heavy particles from those that involve the details of the  $B$  meson. We then generalize the technique to other types of interaction

<sup>6</sup>For the exact expression see [20].

and introduce Wilson coefficient, which can be used to test new physics. Lastly, we found that effective Hamiltonian for  $B \rightarrow X_s \gamma$  consists of three terms shown in Figure 2.4.

## 2.3 Direct CP Violation in $B \rightarrow X_s \gamma$

Direct  $CP$  violation in  $B \rightarrow X_s \gamma$  decay arises from the interference of Wilson Coefficients<sup>7</sup>  $C_1$ ,  $C_{7\gamma}$  and  $C_{8g}$  as shown in the previous section. The theoretical prediction, including recently found long distance effects, for  $A_{CP}$  [11] is<sup>8</sup>

$$-0.6\% < A_{CP}^{SM} < 2.8\%. \quad (2.18)$$

The isospin difference of  $A_{CP}$ ,  $\Delta A_{X_s \gamma}$ , is more interesting. It arises from the interference term that involves the value of the charge of the spectator quark  $e_{\text{spec}}$ . The spectator quark for neutral  $B$  is  $d/\bar{d}$  with electric charge of  $\mp 1/3$  while the spectator quark for a charged  $B$  is  $u/\bar{u}$  with electric charge of  $\pm 2/3$ . Thus, if we subtract  $A_{CP}$  for the charged  $B$  and neutral  $B$ , we have the following expression[11]:

$$\Delta A_{X_s \gamma} = A_{X_s^- \gamma} - A_{X_s^0 \gamma} \approx 4\pi^2 \alpha_s \frac{\tilde{\Lambda}_{78}}{m_b} \text{Im} \frac{C_{8g}}{C_{7\gamma}} \quad (2.19)$$

$$\approx 12\% \times \frac{\tilde{\Lambda}_{78}}{100 \text{ MeV}} \text{Im} \frac{C_{8g}}{C_{7\gamma}}, \quad (2.20)$$

where  $\tilde{\Lambda}_{78}$  is approximately

$$17 \text{ MeV} < \tilde{\Lambda}_{78} < 190 \text{ MeV}.$$

Both  $C_{8g}$  and  $C_{7\gamma}$  are real in the SM, thus the prediction for  $\Delta A_{X_s \gamma}$  is zero. Currently, even though  $C_{7\gamma}$  is constrained by  $b \rightarrow sl^+l^-$  observables and  $Br(B \rightarrow X_s \gamma)$ [21][22], there is no strong constraint on  $C_{8g}$ . Our measurement of  $\Delta A_{X_s \gamma}$  will provide one of the first to constrain  $C_{8g}$ .

The New Physics contribution to these Wilson coefficients can be calculated. For example, the contribution from the two Higgs doublet model[23], or Supersymmetric

<sup>7</sup>See also Paz's presentation at Moriond QCD 2011 <http://moriond.in2p3.fr/QCD/2011/TuesdayMorning/Paz.pdf>

<sup>8</sup>The theoretical prediction, excluding long distance effect[10], is  $(0.44_{-0.10}^{+0.15} \pm 0.30_{0.09}^{0.19})\%$  where the uncertainties refer to the quark mass ratio  $m_c/m_b$ , CKM parameters, and higher order perturbative correction.

model[24] could make  $A_{CP}$  as large as 15%[24]. An example SUSY contribution to  $C_{7\gamma}$  is shown in Figure 2.5.  $C_{8g}$  has a similar contribution from New Physics. The contribution from new physics to  $C_7$  and  $C_{8g}$  is summarized in Table 2.1.

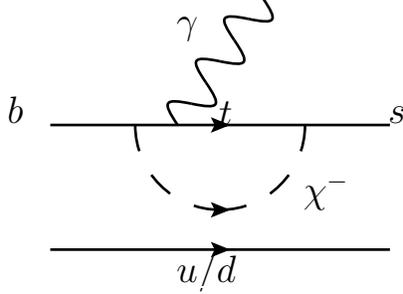


Figure 2.5: An example of SUSY contribution to  $C_7$

Table 2.1: Contribution of New Physics to  $C_7$  and  $C_8$  with new physics parameterized by  $\xi = 3 \frac{C_{7\gamma}^{\text{new}}}{C_{8g}^{\text{new}}}$  and categorized by the particle in penguin diagrams[2]

Model	$\xi$
neutral scalar-vector like quark	1
gluino-squark ( $m_{\tilde{g}} < 1.37m_{\tilde{q}}$ )	-(0.13-1)
techniscalar	$\approx -0.5$
scalar diquark-top	4.8-8.3
gluino-squark ( $m_{\tilde{g}} > 1.37m_{\tilde{q}}$ )	-(1-2.9)
charged Higgs-top	-(2.4-3.8)
left-right W-top	$\approx -6.7$
Higgsino-stop	-(2.6-24)

## Chapter 3

# PEP-II and the *BABAR* Detector

### 3.1 Introduction

In this section we will briefly describe the PEP-II collider and each component of the *BABAR* detector[25]. The main goal of the *BABAR* experiment is to study *CP* violation in the *B* meson system. The detector and beam profile are chosen primarily for this purpose.  $e^-$  and  $e^+$  beams collide asymmetrically primarily at the invariant mass of  $\Upsilon(4S)$ . The information on the collision products is obtained through various detector components.

The schematic of the *BABAR* detector is shown in Figure 3.1. From the innermost layer of the *BABAR* detector, we have a silicon vertex tracker (SVT) which tracks charged particles at high precision close to the interaction point (IP). The next layer is the drift chamber (DCH). It provides tracking information and momentum measurement along with *K* and  $\pi$  particle identification at low momentum ( $<500$  GeV). The *K*/ $\pi$  separation at higher momentum is done through detection of internally reflected Cherenkov light (DIRC). Photon and electron energy is measured at high precision with electromagnetic calorimeter (EMC). SVT, DCH and EMC are operated inside a 1.5T superconducting solenoid. At the outermost layer of the detector we have an instrumented flux return, which identifies muon and detects long lived kaons. The triggering is done in two stages to prevent overloading of the data storage system.

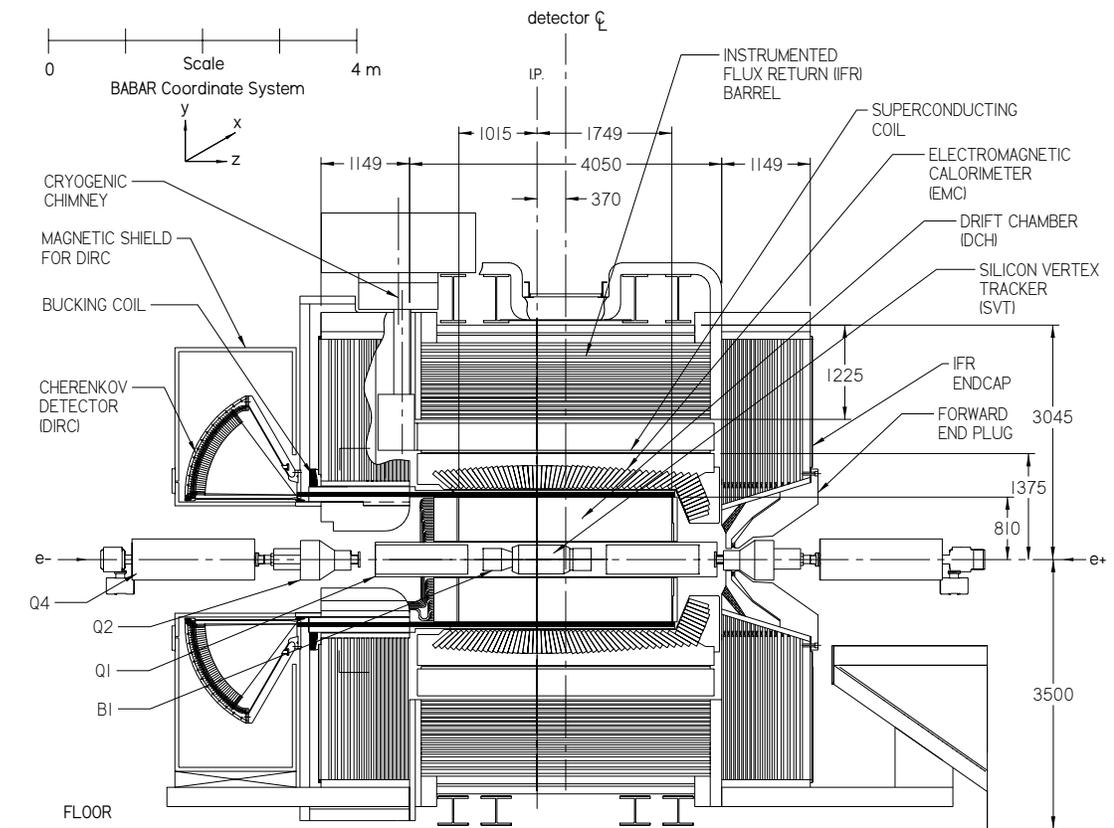


Figure 3.1: Longitudinal cross-section of *BABAR* detector.

## 3.2 PEP-II

The PEP-II accelerator is an asymmetric  $e^+e^-$  collider located at the Stanford Linear Accelerator Center (SLAC). The schematic of PEP-II is given in Figure 3.2. It consists of two storage rings. High energy storage ring (HER) delivers electrons at energy of 9.0 GeV. Low energy storage ring (LER) delivers 3.1 GeV positrons. Together, they collide at the center of mass energy,  $\sqrt{s}$ , of 10.58 GeV – the resonance mass of the  $\Upsilon(4S)$  which is just above the mass of two  $B$  mesons; it decays  $B\bar{B}$  more than 96% of the time.

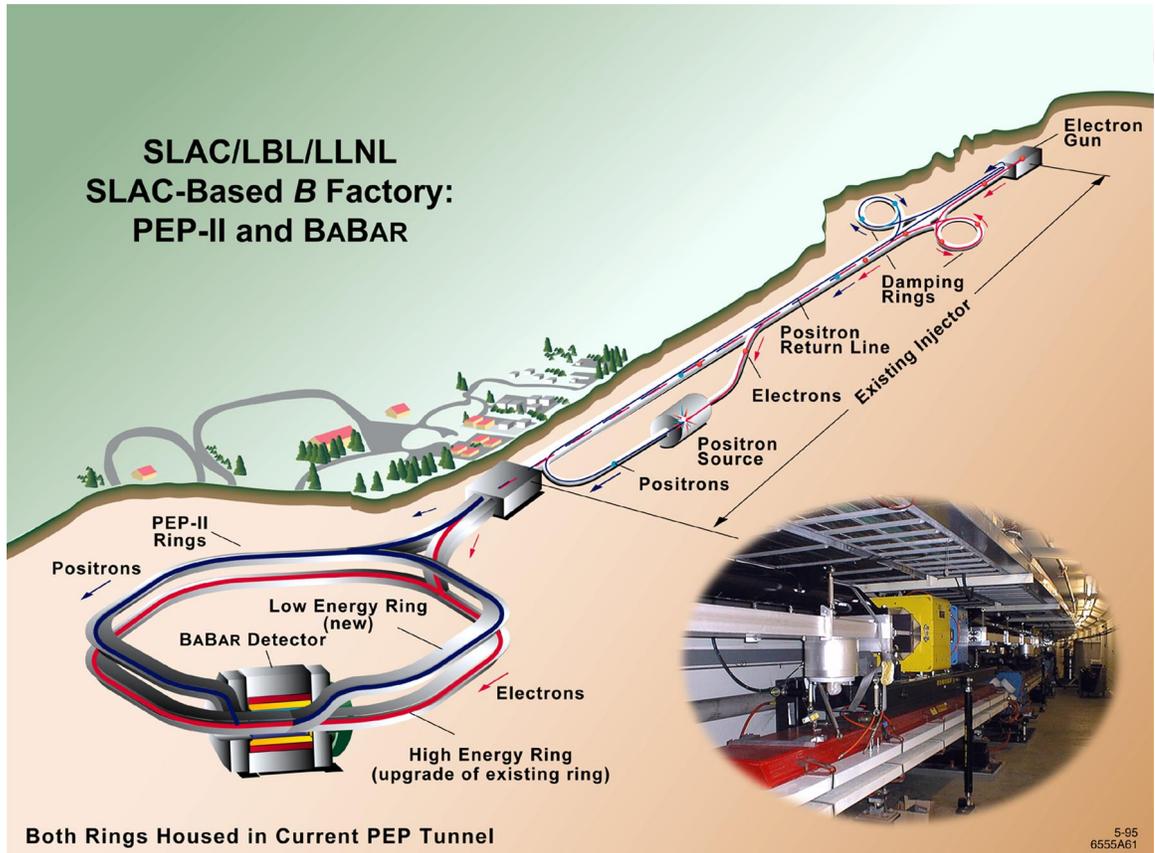


Figure 3.2: SLAC site, PEP-II and *BABAR* detector

At this collision energy, there are many interactions besides  $e^+e^- \rightarrow \Upsilon(4S)$ . In particular, we have Bhabha scattering, muon pair production, tau pair production and light quark pair production. The cross-section for these processes are listed in Table 3.1. These processes serve as background to most  $B\bar{B}$  analysis. To study these

processes with real data, we took some data at 40 MeV below  $\Upsilon(4S)$  resonance so that we have a set of data without  $B\bar{B}$ .

Table 3.1: Cross section of relevant processes within *BABAR* detector coverage.

Process	Cross-section
$e^+e^- \rightarrow \Upsilon(4S)$	1.05nb
$e^+e^- \rightarrow e^+e^-$	$\sim 40$ nb
$e^+e^- \rightarrow \mu^+\mu^-$	1.16nb
$e^+e^- \rightarrow u\bar{u}$	1.39nb
$e^+e^- \rightarrow d\bar{d}$	0.35nb
$e^+e^- \rightarrow s\bar{s}$	0.35nb
$e^+e^- \rightarrow c\bar{c}$	1.30nb

The asymmetric energy of HER and LER beam yields  $\beta\gamma(\Upsilon(4S)) = 0.56$ . This separates the  $B$  and  $\bar{B}$  from  $\Upsilon(4S)$  decays by  $250 \mu m$ , well within the tracking resolution of the *BABAR* detector. This separation is very important for the study of mixing-induced  $B$  decay which yields the very important Standard Model parameter,  $\sin(2\beta)$ .

The designed luminosity of the collider was  $3 \text{ nb}^{-1}s^{-1}$ . By the end of the experiment, the recorded peak luminosity was  $12.1 \text{ nb}^{-1}s^{-1}$ ; 4 times the design luminosity. Throughout 7 runs of the *BABAR* experiment(1999-2008), PEP-II delivered an integrated luminosity of  $553.48 \text{ fb}^{-1}$ , of which  $432.89 \text{ fb}^{-1}$  were recorded by *BABAR* detector at the  $\Upsilon(4S)$  and  $53.85 \text{ fb}^{-1}$  at the off-peak energy. This is shown in Figure 3.3 Our analysis uses  $429 \text{ fb}^{-1}$  of data recorded at  $\Upsilon(4S)$  resonance, which corresponds to  $471 \times 10^6 B\bar{B}$  pairs.

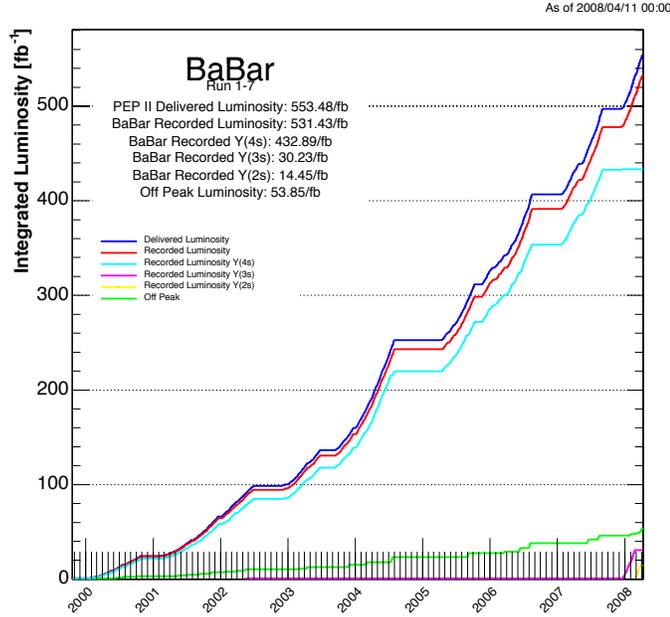


Figure 3.3: Integrated luminosity delivered by PEP-II and recorded by the *BABAR* detector

### 3.3 Silicon Vertex Tracker(SVT)

The SVT is one of the two *BABAR* tracking subsystems. It consists of 5-layer double sided silicon microstrips arranged as shown in Figure 3.4. The microstrips on the two sides of each layer run orthogonal to each other. The strips parallel to the beam pipe measure the azimuthal angle  $\phi$  and the strips which run orthogonal to the beam pipe measures the position along the beam pipe ( $z$ ). The ring closest to the beam pipe is 33 mm away from the IP and the furthest ring is 146 mm away.

The main purpose of the SVT is to measure the precise position and decay angle of particles close to the interaction point. This serves several purposes. First, the resolution of the SVT is typically 20-40 $\mu m$ , depending on the angle. This is less than typical separation of the two  $B$ 's from  $\Upsilon(4S)$  ( $\sim 256\mu m$ ). This high resolution allows reconstruction of  $B$  and  $D$  mesons with high purity, since the daughter of each particle should reconstruct to the same point. The SVT is the only tracking information available for particles with transverse momentum ( $p_t$ )  $< 120$  MeV. Beside tracking,

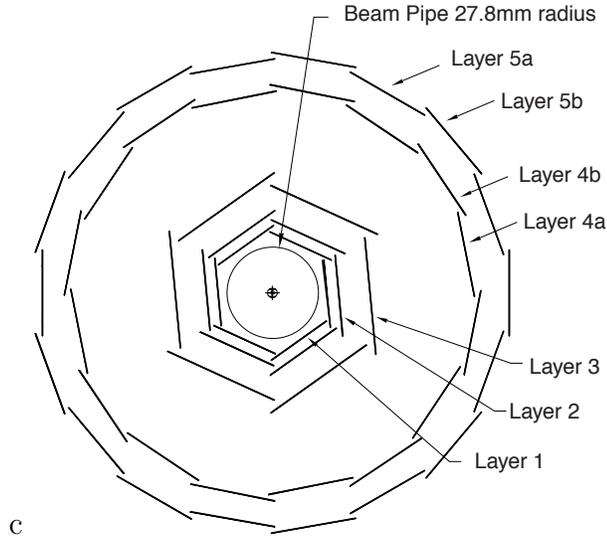


Figure 3.4: Layout of the SVT microstrip layers.

energy loss,  $dE/dx$ , information from the SVT can also be use to do  $K/\pi$  separation. For charged particles with transverse momentum( $p_t$ )  $< 120$  MeV, the  $dE/dx$  from the SVT is the only information available for particle identification (PID). At higher  $p_t$ , the precise knowledge of the angle can be combined the Cherenkov cone angle information from the DIRC effectively. Lastly, since the SVT operates in a magnetic field, measurement of the curvature of the track provides us with the momentum of the particle.

### 3.4 Drift Chamber(DCH)

The DCH consists of 40 layers of small hexagonal cells. The total number of drift cell is 7,104. Each cell consists of 1 tungsten-rhenium sense wire and 6 aluminum field wires. The longitudinal cross-section view DCH is shown in Figure 3.5. The chamber is about 3m long and filled with a mixture of 80% helium and 8-9.5%isobutane to minimize multiple scattering<sup>1</sup>.

Once a charged particle enters the chamber, it ionizes the gas and the electrons will then move toward the sense wire which is at 1930V. On its way to the sense wire,

<sup>1</sup>The rest is argon and carbondioxide

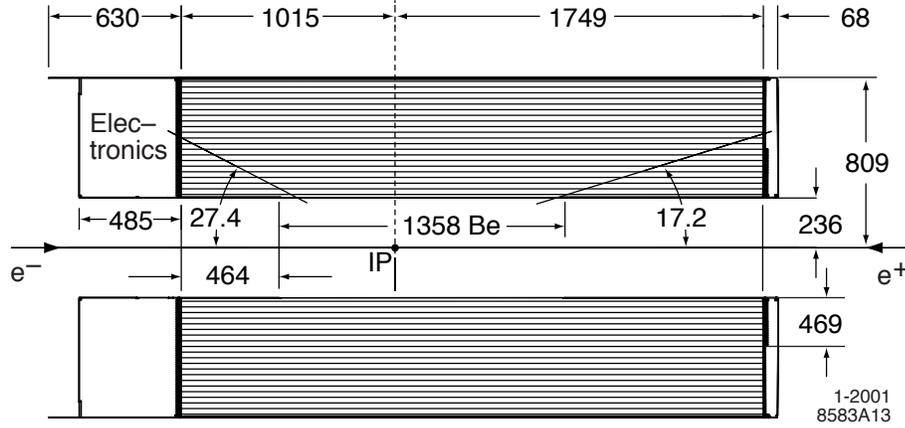


Figure 3.5: Longitudinal cross-section of drift chamber.

the electrons ionizes more gas and produce more electrons, creating an avalanche of electrons moving toward sense wire. The time of arrival measures the distance of closest approach and the integrated charge measures the energy deposited which we use to measure  $dE/dx$ .

There are two sets of sense wires inside the DCH: one parallel to the beam axis and the other are placed at small angle relative to beam axis. Together, they provide both azimuthal and longitude information.

The DCH provides tracking information with a resolution of  $100 - 400\mu m$  depending on the angle. The momentum of the charged track, which is obtained from curvature, has a resolution of

$$\sigma(p_t)/p_t = 0.45\% + 0.15\%p_t(\text{GeV}/c).$$

The DCH also provides energy loss information,  $dE/dx$ , at a resolution of approximately 7%. This information is an excellent  $K/\pi$  PID discriminant at momenta lower than  $700\text{ MeV}/c$ .  $dE/dx$  for given particle momentum for various type of charged particles is shown in Figure 3.6.

Most  $K_S$  particles, which are crucial to many analysis including ours, live long enough to travel through the SVT, leaving no energy, and then decays to  $\pi^+\pi^-$ . The DCH is the sole tracking device for  $K_S$  reconstruction.

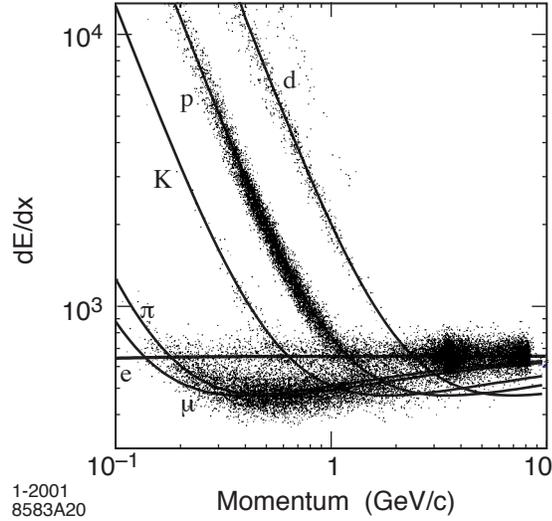


Figure 3.6:  $dE/dx$  for various types of charge particle.

### 3.5 Detector of Internally Reflected Cherenkov Light

The Detector of Internally Reflected Cherenkov light (DIRC) operates on the principle that when a charge particle travels through a fused silica bar, it emits Cherenkov light. The opening cone angle ( $\theta_c$ ) is related to the velocity in which the particle moves through the material. This information, together with the momentum obtained from the SVT and DCH, can be used to distinguish between types of charged particles. Figure 3.7 shows  $\theta_c$  for various types of particles. The DIRC system complements the  $dE/dx$  from the SVT and DCH at  $p_t > 700 \text{ MeV}/c$ , where  $dE/dx$  becomes a less effective discriminant.

Our DIRC system consists of thin synthetic fused silica bars with one end equipped with a mirror and the other end connected to a photo multiplier tube (PMT) array viewing a water container. Figure 3.8 shows the setup and principle of operation of the DIRC. To measure  $\theta_c$ , we have the light internally reflected within the quartz bar. At each reflection, the  $\theta_c$  is preserved. The light will eventually emerge out from the quartz bar at the end with the PMT array. The PMT array then catch the light and we use pattern recognition to determine  $\theta_c$ .

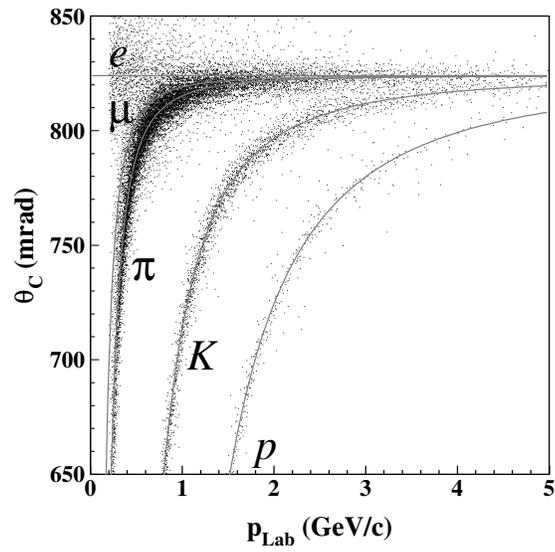
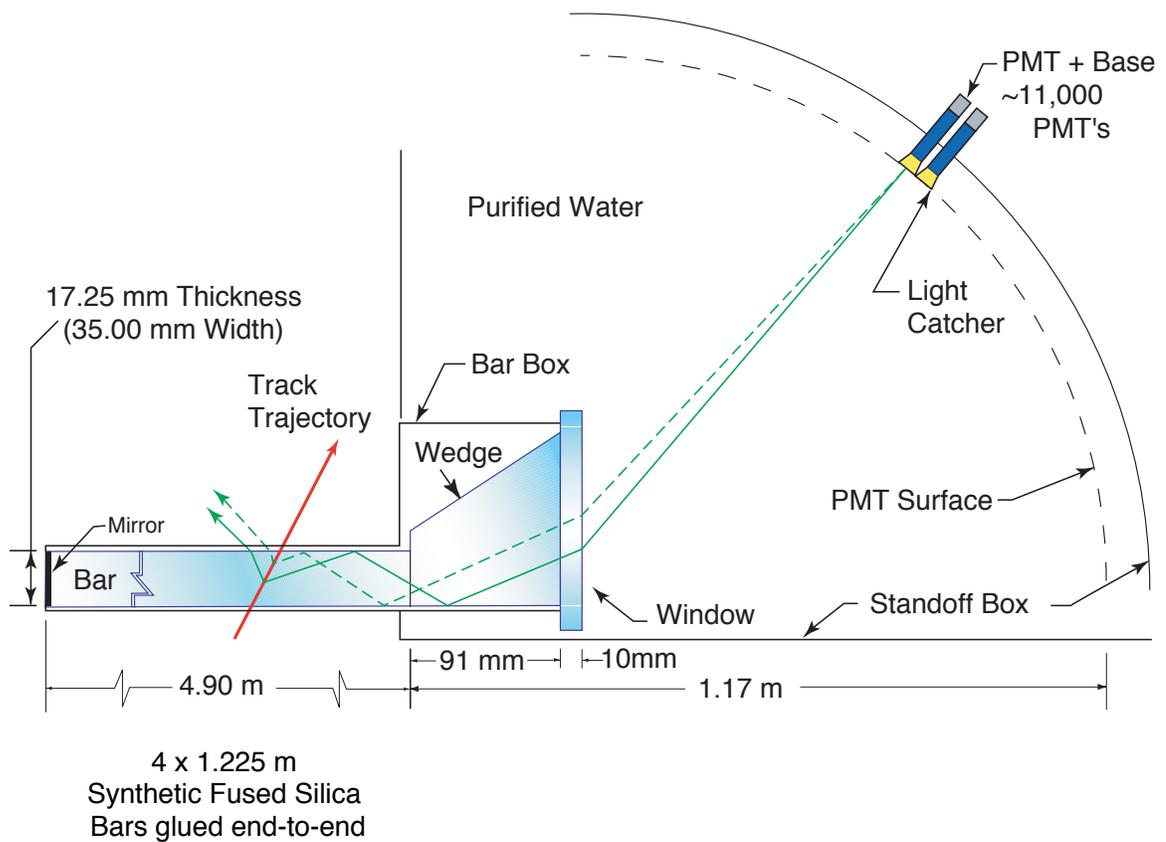
Figure 3.7:  $\theta_c$  for various types of particle.

Figure 3.8: DIRC layout and principle of operation.

### 3.6 Electromagnetic Calorimeter(EMC)

Our EMC is made up of 6,580 CsI(Tl) crystals arranged projectively to the interaction point (IP), as shown in Figure 3.9 covering both barrel and end cap. CsI(Tl) has a radiation length of 1.85cm and Molière radius of 3.8 cm. The crystals are 4.7 cm x 4.7 cm in cross-section at the entrance face; the length of the crystal varies from 29.6 cm in the backward direction to 32.4 cm in the forward direction. Each crystal has 2 PIN diodes for collecting the scintillation light. Since the light output degrades as it receives more radiation, the EMC is constantly re-calibrated through the lifetime of the experiment with activated fluorinert fluid and Bhabha scattering events.

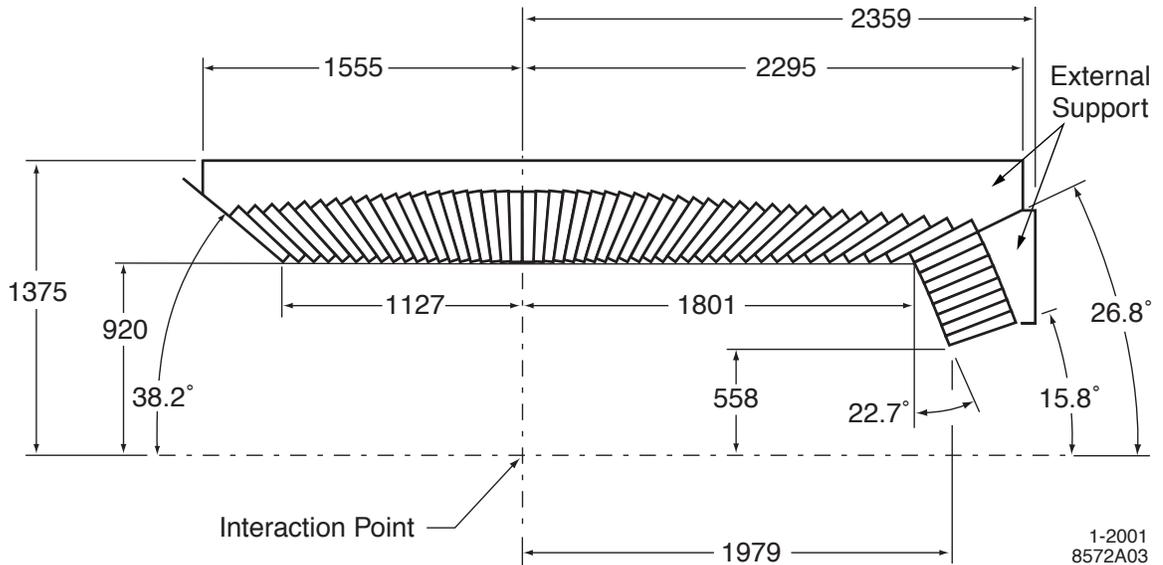


Figure 3.9: Schematics of EMC layout.

The primary purpose of EMC is to measure the energy of photon and electron. When a photon or an electron travels through the crystal it loses almost all of its energy due to bremsstrahlung by emitting lower energy photons through electromagnetic cascade. For CsI(Tl), typically 5,000 photons are created for each MeV of energy deposited. Light are then reflected inside the crystal and collected at the end with PIN diodes. For heavier particle such as  $\pi$ ,  $K$ ,  $p$  or  $\mu$ , it loses its energy through ionizing crystals and leaves only fraction of its energy close to minimum ionizing energy. For this reason,  $E/p$  is an excellent discriminating variable of  $e$  from all the other

heavy particles.

The energy resolution of the EMC comes from two effects. The first is photon statistics. PIN diodes collect only a fraction of photons. Thus, the number of photons reaching the PIN diodes fluctuates by Poisson statistics. The second effect is the longitudinal uniformity of the crystal response. The light output depends on where the shower develops along the crystal. The total resolution is parameterized as follows:

$$\frac{\sigma_E}{E} = \frac{(2.32 \pm 0.30)\%}{\sqrt[4]{E(\text{GeV})}} \oplus (1.85 \pm 0.12)\%.$$

The first term represents the photon statistic and the second term represents effects from longitudinal non-uniformity and energy leakage.

### 3.7 Instrumented Flux Return(IFR)

The main purpose of the IFR system is to identify muons and long lived neutral hadrons. The IFR uses the steel flux return as muon a filter. Muons, unlike other particles, pass through many steel plates; thus, counting how many steel plates the particles pass through can tell us how likely that the track is a muon.

In the early phase of the experiment, resistive plate chambers (RPC) are installed between the steel plates. There are 19 RPC layers total in the barrel and 18 in the endcaps as shown in Figure 3.10. Each RPC consists of 2 layer Aluminum, foam, capacitive sensor, graphite Bakelite sandwiching freon-argon-isobutane gas mixture as shown in Figure 3.11. The RPC works on the principle of detecting streamers from ionization left by charged particles.

However, we found a significant degradation of RPC performance during the experiment. We replaced RPC with limited streamer tubes(LST). LSTs are long rectangular cells of graphite coated PVC and silver-plated wire. Each cell is filled with CO<sub>2</sub>-isobutane-argon gas mixture. The principle of operation is as same for the RPC.

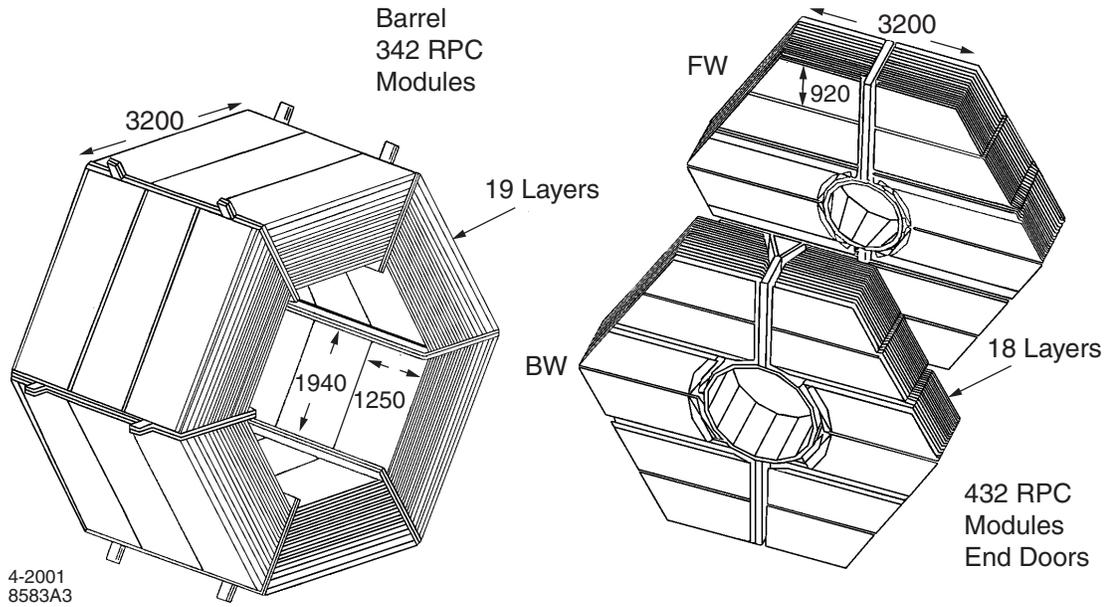


Figure 3.10: Layout of IFR and RPC.

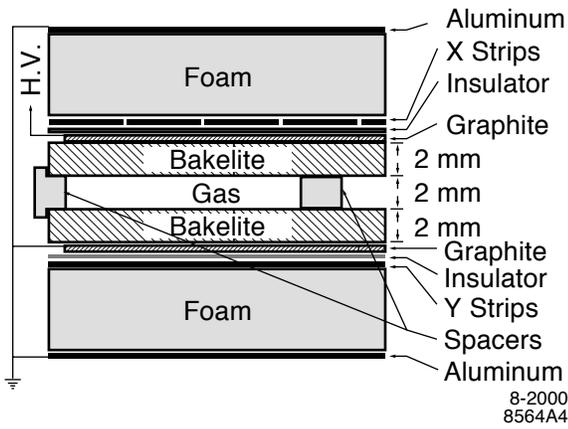


Figure 3.11: Schematic for each RPC module.

## 3.8 Trigger

The main purpose of the trigger is to reduce the data flow rate to the data storage system preventing overloading while retaining most of the interesting events. The *BABAR* trigger system consists of two stages: a hardware-based L1 trigger, and a software-based L3 filter. The trigger is designed to handle 10 times the PEP-II background rate at the design luminosity.

The L1 trigger is based on thresholding the  $p_t$  in the DCH, showers in the EMC and the number of tracks detected in the IFR. The whole process takes around 12.8  $\mu\text{s}$ . At the luminosity of  $3 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ , L1 Trigger rate is approximately 170 Hz, while retaining more than 99.9% of  $B\bar{B}$  events.

The L3 filter decision comes primarily from two orthogonal inputs. One is based exclusively on DCH data and the other is based only on EMC data. The DCH data based one input ensures that tracks are above a certain  $p_t$  with some vertexing requirement. The EMC data-based one ensures that there are two clusters with energy at least 350 MeV and event mass (assuming all tracks are massless) greater than 1.5 GeV. The L3 filter retains more than 99% of  $B\bar{B}$  events. Certain classes of events, such as Bhabha scattering output from L3 are also used for luminosity measurement.

# Chapter 4

## Particle Identification(PID)

### 4.1 Introduction

Particle Identification (PID) is a very important part of the the *BABAR* experiment, used in almost every *BABAR* analysis. The most common use of PID is in determining the flavor of  $B$  or  $D$  and reducing combinatoric background.

re-weigh

In every study of  $CP$  asymmetry, one is required to determine the flavor of the  $B$  of interest, either directly through direct reconstruction, or by inferring it from the flavor of the other  $B$ <sup>1</sup>. The flavor of the a  $B$  meson can be deduced from the charge of the track identified as a kaon or lepton.

To be specific, a  $B$  meson typically cascade down through  $W$  and  $c$  quark emission;  $b \rightarrow W^- c$ . The  $c$  quark then cascades to a strange quark;  $c \rightarrow W^+ s$ . The  $s$  quark then form  $K^-$ . This means that  $\bar{B}^0$  and  $B^-$ , which contain a  $b$  quark, can be identified by  $K^-$  in the final states<sup>2</sup> and  $B^0$  and  $B^+$ , for the same reason, can be identified by a  $K^+$  in the final state. This technique relies heavily on our ability to identify  $K$  and  $\pi$  meson.

Another popular flavor tagging method is via semi-leptonic decay of  $\bar{B}^0 \rightarrow X l^- \bar{\nu}_l$  where  $l \in \{e, \mu\}$ . The flavor of the  $B$  meson can be deduced from the charge of the lepton in the final state. Thus, PID of electrons and muons will be crucial for this

---

<sup>1</sup>since  $\Upsilon(4S)$  decays to two  $B$ s of the opposite flavor

<sup>2</sup>modulo some rare Cabibbo suppressed modes.

technique.

Another use of PID is to reduce combinatoric background. For example, if a study requires reconstruction of a particle from a known final state, for example,  $J/\psi \rightarrow e^+e^-$  or  $K^* \rightarrow K^+\pi^-$ . With PID, the analyst can combine only the combination with the right PID instead of combining all charged tracks.

The *BABAR* detector is designed with this capability in mind. Each detector system contributes information that can be used to perform PID. The SVT and DCH provide  $dE/dx$  and momentum along with tracking information. The DIRC provides the Cherenkov cone angle. The EMC provides various shower shape parameters and energy deposited. The IFR can be used to identify muons. This creates another question: how should we combine all this information to classify a charged track? *BABAR* PID algorithm evolves over the lifetime of the experiment. Starting from simple set of rectangular cuts on the variables, it evolved to a likelihood-based method. In the final iteration of *BABAR* PID, we used a multiclass multivariate algorithm called Error-Correcting Output Code[26] with an exhaustive matrix which will be the focus of this chapter. We will start by introducing the original and pedagogical version of ECOC first. The exhaustive matrix will then be introduced. Adapting the algorithm to suit the *BABAR* requirements will be explained next. Training samples used will also be discussed briefly. Then we will compare the performance of the ECOC-based PID with previous generations of PID.

## 4.2 Error-Correcting Output Code(ECOC)

ECOC[26] is an algorithm for making a multiclass classifier from binary classifiers. It was invented around 1995 by T. Dieterich and G. Bakari. What will be described in this section is a little different from the implementation at *BABAR*, but the basic idea is the same. This section provides a more pedagogical example. The actual *BABAR* implementation will be described in a later section.

Let us consider the problem of classifying data into 4 classes  $\{A, B, C, D\}$  using binary classifiers. For our problem, this translates to determining if a charged track

is a  $p$ ,  $\pi$ ,  $K$ , or  $e$  using classifiers which have only two outputs: 1 or 0.

One simple approach would be to have 2 classifiers:  $\alpha$  and  $\beta$ . The classifier  $\alpha$  would be trained to separate classes  $\{A, B\}$  from class  $\{C, D\}$  and classifier  $\beta$  would be trained to separate  $\{A, C\}$  from class  $\{B, D\}$ , as shown in Table 4.1. This table is called indicator matrix.

Class	$t_\alpha$	$t_\beta$
$A$	1	1
$B$	1	0
$C$	0	1
$D$	0	0

Table 4.1: Indicator matrix of a multiclass classifier from combining two binary classifiers.  $t_\alpha$  and  $t_\beta$  represents how the given class is treated in the training sample for classifiers  $\alpha$  and  $\beta$  respectively. 1 means signal and 0 means background.

The rows and columns of the indicator matrix can be interpreted as the following:

- Each row represents the ideal output from each classifier for a given class. This is the template for us to compare the output from the classifier.
- Each column represents ideal output for each class for a given classifier. This is the way we should train each classifier.

Suppose we wish to classify a data point  $d$  into one of the four classes. To accomplish this, we ask classifier  $\alpha$  and  $\beta$  to classify  $d$ . The output from  $\alpha$  ( $O_\alpha$ ) and  $\beta$  ( $O_\beta$ ) will be a string of 0s and 1s. For example, if  $\alpha$  says  $d$  is a signal(1) and  $\beta$  says  $d$  is a background(0), then we say the output is 10. We then compare the output string to the indicator matrix to see which class is the “closest” one.

To determine this, we can use the Hamming distance which is the number of bits that are different between the output and the corresponding row in the indicator matrix. Intuitively, the Hamming distance measures how close the output string is to each template answer. The lower the Hamming distance, the closer they are. Let us suppose that the true class of  $d$  is  $A$  and we get correct answers from both  $\alpha$  and  $\beta$  that are 1 and 1, respectively. We can then use this answer and compute the hamming distance between the answer and the template for each class shown in Table 4.2.

$O_\alpha$	$O_\beta$	class	$t_\alpha$	$t_\beta$	Hamming distance
1	1	$A$	1	1	0
		$B$	1	0	1
		$C$	0	1	1
		$D$	0	0	2

Table 4.2: Hamming distances from comparing the output of  $\alpha$  and  $\beta$  that is 1 and 0, respectively, with indicator matrix given in Table 4.1

Number of Mistakes	$O_\alpha$	$O_\beta$	Most probable class
0	1	1	$A$
1	1	0	$B$
	0	1	$C$
2	0	0	$D$

Table 4.3: When using only two classifiers, we only need one classifier to make a mistake to obtain a wrong class.

Using Table 4.2, we would say that the most probable class for the data point  $d$  is class  $A$ , because it has the lowest Hamming distance. However, in practice,  $\alpha$  and  $\beta$  are not perfect; they might give a wrong answer for  $d$ . With the indicator matrix we just considered, we only need to have one classifier make a mistake for us to misclassify the class for  $d$ . This is shown in Table 4.3. We will discuss how to make a better indicator matrix in the next section.

### 4.3 Exhaustive Matrix

As we see in the the previous section, our multiclass classifier constructed from 2 binary classifiers will misclassify the class for data  $d$  if one of the two classifiers makes a mistake. We can improve this by adding more classifiers and trying to make sure that when one of them provides an incorrect answer, we can still recover the correct answer.

Let us first count how many different binary classifiers we can build for a fixed number of classes. For the problem of  $n$  classes, we can represent a classifier by a binary string of length  $n$ , similar to the columns in Table 4.1 (eg. 1010). Therefore

<i>A</i>	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
<i>B</i>	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0
<i>C</i>	1	1	0	0	1	1	0	1	1	0	0	1	1	0	0
<i>D</i>	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0
d	all 1	good ones						complement							

Table 4.4: All possible classifiers for the four classes problem. There are only 7 distinct usable classifiers.

Class	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$
<i>A</i>	1	1	1	1	1	1	1
<i>B</i>	0	1	0	1	0	1	0
<i>C</i>	1	0	0	1	1	0	0
<i>D</i>	1	1	1	0	0	0	0

Table 4.5: Exhaustive matrix for 4 classes.

we can train a binary classifier in  $2^n$  ways.

However, we are double-counting some classifiers. The classifiers which are the complement of each other are actually the same classifier. For example, the classifier represented by 1100 and 0011 are both trained to separate  $\{A, B\}$  from  $\{C, D\}$ ; we do nothing but swap the definition of signal and background. Further, the classifier which is represented by a column of all 0's or all 1's are not valid classifiers; they do not really separate anything from anything. Therefore, in total, we can have a maximum of  $2^{n-1} - 1$  binary classifiers. This calculation is illustrated in Table 4.4.

The indicator matrix that has all possible classifiers is called an exhaustive Matrix. An example of an exhaustive matrix for the 4 classes problem is shown in Table 4.5. This matrix is the one used in the ECOC-based PID. The exhaustive matrix is quite special in terms of its recovery power. To illustrate this property, let us consider our problem of classifying data  $d$  for which the true class is  $A$ . The ideal answer from the classifiers would be the first row of Table 4.5: 1111111. However, classifiers can make mistakes. The Hamming distance for this answer with various numbers of classifier mistakes is shown in Tables 4.6-4.8.

As we can see from Tables 4.6-4.8, we can tolerate up to 2 misclassifications and

Output	Class	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	Hamming distance
0111111	<i>A</i>	1	1	1	1	1	1	1	1
	<i>B</i>	0	1	0	1	0	1	0	3
	<i>C</i>	1	0	0	1	1	0	0	5
	<i>D</i>	1	1	1	0	0	0	0	4

Table 4.6: Hamming distances of the exhaustive matrix when 1 classifier makes mistake.

Output	Class	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	Hamming distance
0011111	<i>A</i>	1	1	1	1	1	1	1	2
	<i>B</i>	0	1	0	1	0	1	0	4
	<i>C</i>	1	0	0	1	1	0	0	3
	<i>D</i>	1	1	1	0	0	0	0	6

Table 4.7: Hamming distances of the exhaustive matrix when 2 classifiers make mistake.

Output	Class	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	Hamming distance
0001111	<i>A</i>	1	1	1	1	1	1	1	3
	<i>B</i>	0	1	0	1	0	1	0	3
	<i>C</i>	1	0	0	1	1	0	0	3
	<i>D</i>	1	1	1	0	0	0	0	7

Table 4.8: Hamming distances of the exhaustive matrix when 3 classifiers make mistakes. Note that there is a tie between class *A*, *B* and *C*. In our implementation of ECOC, the output from each classifier is a real number and we use the sum of squared difference to extend the Hamming distance into a real number. There are no ties in our implementation.

still correctly classify  $d$ . The reason for this property is the row Hamming distance. For our 4 class exhaustive matrix, the hamming distance between any two rows is 4. That means classifiers can make two mistakes, in the worst case, at the bits that are not the same between two rows and we can still recover the correct answer.

As shown in Table 4.8, we can run into tie situations. In the actual *BABAR* implementation, we avoid this problem by using a real number for the classifier output and generalize the Hamming distance to a sum or a squared difference between each bit.

There is one problem with the exhaustive matrix. Since exhaustive matrix contains all possible classifiers, the number of classifiers needed for this indicator matrix grows exponentially with the number of classes. This indicator matrix is, therefore, not suitable for a large number of classes. The indicator matrix for a large number of classes, is discussed in Section 3 of [26].

## 4.4 *BABAR* Implementation

At *BABAR*, each analysis has a different requirement for efficiency and fake rate for PID. Efficiency measures how many of real particles of type  $A$  are identified as type  $A$ , while fake rate measures how much of real particle of type other than  $A$  is identified as type  $A$ . These two numbers couple tightly together. To satisfy the different needs of each analysis, we need to create several tightness levels. Each tightness has a different efficiency and fake rate. The original ECOC, as described in previous sections gives only one answer, the best guess; there is no tightness level. This section explains how we modify the ECOC to suit *BABAR* needs.

Let us start with specific information. Our goal is to classify 4 classes of particles:  $e, \pi, K$  and  $p$ . Muons have dedicated PID selector which will not be discussed here. We choose the exhaustive matrix shown in Table 4.9 as our indicator matrix. It consists of 7 binary classifiers trained differently, as shown by the columns of Table 4.9. The algorithm we choose for each binary classifier is a Bootstrap Aggregate Decision Tree and each classifier is trained on 31 variables, listed in Appendix A.1.

Class	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$
$K$	1	1	1	1	1	1	1
$\pi$	-1	1	-1	1	-1	1	-1
$p$	1	-1	-1	1	1	-1	-1
$e$	1	1	1	-1	-1	-1	-1

Table 4.9: The exhaustive matrix is the indicator matrix used in the ECOC-based selector. Each entry indicates whether the training sample of the given type should be treated as signal(1) or background(-1).

To classify a given track, we ask each classifier to give an output between -1 and 1 according to each classifier’s definition of signal and background. The output from the binary classifiers can be represented by a string of real number between -1 and 1 with length 7. We then compute the sum of squared differences (generalized Hamming distances) of the output string with each row of Table 4.9. From this, we have 4 number, corresponding to each particle hypothesis.

We can select the hypothesis that gives the minimum Hamming distance and call it our best guess. In *BABAR* experiment, however, we need multiple tightness levels for different demands from analyses.

To supply multiple tightness levels, we noticed that selecting the minimum can be written as the comparison between the ratios and one:

$$A < B \text{ and } A < C \iff \frac{A}{B} < 1 \text{ and } \frac{A}{C} < 1 \quad (4.1)$$

We can generalize this relation by changing 1’s to constants. This allows us to adjust those constants for different levels of tightness. So, for a given particle hypothesis we have 4 values to select on: the Hamming distance and 3 ratios. To make this clear, let  $H_a$  be the hamming distance between the output and the template for particle of type  $a$ . Variables for selecting each type of particle are shown in Table 4.10.

It should be noted here that each of the variables on which we place the selection on has intuitive meanings. The Hamming distance of the particle itself represents how likely it is that this track is that type of particle. The three ratios represent how sure we are that it is not another type of particle. This interpretation is also very

Particle type	Selection Criteria
$K$	$H_K, \frac{H_\pi}{H_K}, \frac{H_p}{H_K}, \frac{H_e}{H_K}$
$\pi$	$H_\pi, \frac{H_K}{H_\pi}, \frac{H_p}{H_\pi}, \frac{H_e}{H_\pi}$
$p$	$H_p, \frac{H_\pi}{H_p}, \frac{H_K}{H_p}, \frac{H_e}{H_p}$
$e$	$H_e, \frac{H_K}{H_e}, \frac{H_\pi}{H_e}, \frac{H_p}{H_e}$

Table 4.10: The selection criteria for each type of particle in ECOC-based selector.

useful in tuning. For instance, if we are looking at kaon selector and we found that the pion-as-kaon misidentification rate is too high; it is quite obvious that the value we should tighten the selection on is either  $\frac{H_\pi}{H_K}$  or  $H_K$ .

## 4.5 Training Sample

Our training sample comes from purely from data. There are decay modes where we can determine the particle type of a charged track without using PID on the track of interest. PID may be applied on the tracks that we are not interested in.

To obtain a sample of kaon with high purity, we look at  $D^{*+} \rightarrow D^0\pi^+$  and  $D^0 \rightarrow K^-\pi^+$ . We reconstruct  $D$ s from two oppositely-charged tracks. Then form a  $D^*$  from a  $D$  and another charged track. The track of the same charge is then assumed to be a  $\pi$ . Invariant mass and vertexing attributes must be consistent with  $D$  and  $D^*$ . We also place minimum momentum requirement on  $D$  and  $D^*$ . We apply  $\pi$  likelihood PID to identify pions. It should be emphasized that only track which we assumed to be a  $\pi$  have a PID requirement; we do not apply any PID on the track we will pick as  $K$ . Using this method, we obtain very clean sample of  $K$  without applying any PID requirement on the  $K$  track.

For the pion sample, we use the same  $D^{*+} \rightarrow D^0\pi^+$ ;  $D^0 \rightarrow K^-\pi^+$  sample. We use a similar technique to obtain the pion sample: apply kaon PID to identify kaon and use other tracks as pions.

For electrons, we select Bhabha scattering events. Two tracks with total momen-

tum of 0 are selected. We then apply a cut on the ratio of energy deposited in the EMC and momentum ( $E/p$ ) to one track to make sure that it is an electron/positron then use the other track as our electron/positron sample.

For protons, we use the decay  $\Lambda \rightarrow p\pi$ . We select two charged tracks and combine them to make a  $\Lambda$  with requirements on the invariant mass and flight distance.  $K_S$  veto is applied, assuming the mass of the two tracks to be  $\pi$ 's, to remove those with invariant mass consistent with  $K_S$ . Again, we apply PID to identify pions and use the other track as the  $p$  sample.

## 4.6 Performance

The performance of a PID method has two aspects: the efficiency and the fake rate. These are coupled together tightly. The looser we cut the greater the efficiency but the greater the fake rate and vice versa. Across all types PID selections:  $e, \pi, p$  and  $K$ , the new algorithm allows us to tune the classifiers to provide higher efficiency and lower fake rate compared to all previous generations of *BABAR* PID algorithm. All PID selection performance plots is summarized Appendix A.3.

Let us highlight sample here. Figure 4.1 show a comparison of a kaon PID selector, known internally in *BABAR* as *KKMLoose*. Across all momentum ranges we are able to gain higher efficiency especially at momentum  $\sim 1$  GeV, while lowering the fake rate from  $\pi$  and all other types of particles. This momentum range is very important for hadronic  $B$  flavor tagging since a lot of  $K$ s that are used for  $B$  tagging fall in this momentum range. Higher efficiency allows for higher  $B$  tagging efficiency and lower fake rate allows lower flavor dilution. This selector is also used in the  $B \rightarrow X_s \gamma$   $A_{CP}$  analysis which will be described in the next chapter.

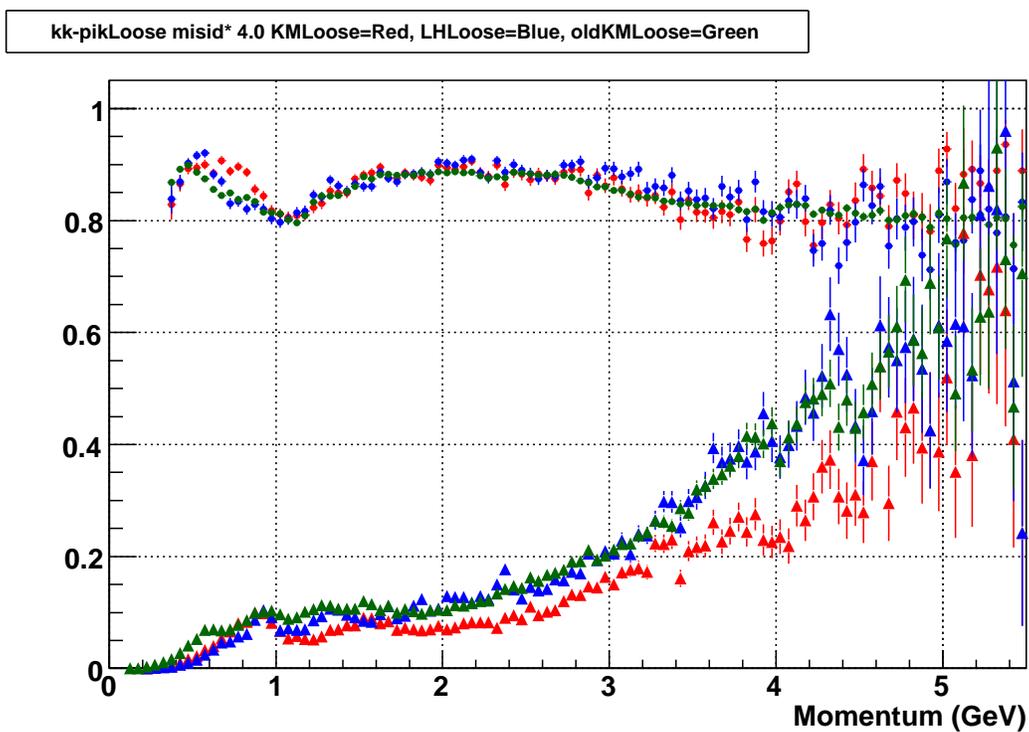


Figure 4.1: Comparison of PID efficiency (circles with error bars) and  $\pi$  as  $K$  fake rate (triangles with error bars) from ECOC-based PID selector using the exhaustive matrix (red), 1 vs 1 matrix (green) and a likelihood-based PID selector (blue). The  $\pi$  as  $K$  fake rate is multiplied by 4 for visibility.

# Chapter 5

## Measurement of $A_{CP}$

### 5.1 $B$ Reconstruction and Event Selection

#### 5.1.1 Overview

Our ultimate goal is to select  $B \rightarrow X_s \gamma$  events with correctly reconstructed  $X_s$  candidates and a correctly chosen primary photon for the 16 final states listed in Table(5.2), and to use these events to extract  $A_{CP}$  using the formula

$$A_{CP} = \frac{N(b \rightarrow s\gamma) - N(\bar{b} \rightarrow \bar{s}\gamma)}{N(b \rightarrow s\gamma) + N(\bar{b} \rightarrow \bar{s}\gamma)}, \quad (5.1)$$

where  $N(b \rightarrow s\gamma)$  and  $N(\bar{b} \rightarrow \bar{s}\gamma)$  are the number of events in which a  $b$  of each flavor decaying to a corresponding  $s$  quark and a  $\gamma$ . Eq. 5.1 and Eq. 1.1 are the same with a few assumptions typically found in  $B$  asymmetry analyses. We list our assumptions in Appendix N.

Our selection is performed using a Monte Carlo (MC) sample<sup>1</sup>. We reconstruct our  $B$  candidates from 38 decay modes listed in Table 5.1. We describe this procedure in Section 5.1.2. After this procedure, we will have multiple  $B$  candidates for each event, since there might be more than one way to pick a set of tracks from an event that combines to look very similar to a  $B$  meson. The entire event may not even be from  $B \rightarrow X_s \gamma$ ; they could arise from  $uds$  or  $c\bar{c}$  or generic  $B\bar{B}$ . So, we have to do a

---

<sup>1</sup>Our dataset is listed in Appendix B

couple of things.

First, we need to select the best candidate for each event. We build a random forest classifier, discriminating between correctly reconstructed candidate and misreconstructed candidates in Section 5.1.3.1. We named it Signal Selecting Classifier (SSC). We then pick, for each event, the candidate that has the best SSC score. We also use SSC output to reject generic  $B\bar{B}$  background. Second, we need to remove continuum events using various event shape variables. We, again build a random forest classifier for this. We called this classifier Background Rejecting Classifier(BRC).

After having selected the best candidate for each event, we then make cut on SSC and BRC based on the  $X_s$  mass. This is because the amount of each type of background varies with  $X_s$  mass: at low  $X_s$  mass we have clean  $B \rightarrow K^*\gamma$  resonance and at high  $X_s$  mass we are overwhelmed by continuum and generic  $B\bar{B}$  and continuum background. After the final SSC and BRC cut we reduce our samples from 38 modes to the final 16 modes in which we can determine the flavor of the  $B$  from its final states. This procedure will be described in Section 5.1.4.

### 5.1.2 $B$ Reconstruction

We reconstructed  $B$  meson candidates from the 38 final states listed on Table 5.1. Note that some of the final states we reconstruct are not  $CP$  eigenstates: only modes where we can determine the  $B$  flavor from the final states, shown in Table 5.2, are actually used in the  $A_{CP}$  extraction process. Extra final states are reconstructed for rejection purposes. In total, we use 10 charged  $B$  final states and 6 neutral  $B$  final states in the  $A_{CP}$  measurement.

Charged kaons and charged pions are selected from charged tracks classified with the error-correcting output code (ECOC) algorithm [26] described in Chapter 4. The classification uses SVT, DIRC, DCH and EMC information. A kaon particle identification (PID) algorithm gives us roughly 90%  $K$  efficiency with a pion-as-kaon fake rate of about 1%. Pion identification is roughly 99% efficient with a 15%  $K$ -as- $\pi$  fake rate.

Neutral Kaons are reconstructed from the decay  $K_s^0 \rightarrow \pi^+\pi^-$ . The invariant mass of the two oppositely charged tracks is required to be between 489 and 507 MeV with a flight distance greater than 0.2 cm from the interaction point. The flight significance (defined as the flight distance divided by the uncertainty of the flight distance) of the  $K_s^0$  must be greater than 3.  $K_L^0$  and  $K_s^0 \rightarrow \pi^0\pi^0$  decays are not used in our analysis.

$\pi^0$  and  $\eta$  mesons are reconstructed from two photons. We require each photon to have an energy of at least 30 MeV for reconstructing  $\pi^0$ s and at least 50 MeV for reconstructing  $\eta$ s. The invariant mass of the two photons must be in the range of [115,150] MeV for  $\pi^0$ s and in the range of [470,620] MeV for  $\eta$ s. Only  $\pi^0$ s and  $\eta$ s with momentum greater than 200 MeV are used. Although we do not reconstruct  $\eta \rightarrow \pi^+\pi^-\pi^0$  explicitly, some are included in final states that contain  $\pi^+\pi^-\pi^0$ .

Each event is required to have at least one photon with energy  $1.6 < E_\gamma^* < 3.0$  GeV, where the \* denotes variables measured in CM frame. These photons are used as the primary photon in reconstructing  $B$  mesons. This photon must have lateral moment less than 0.8 and the nearest EMC cluster must be at least 15 cm away at the face of the calorimeter. The angle of the photon momentum with respect to the beam axis must satisfy  $-0.74 < \cos \theta < 0.93$ .

The invariant mass of  $X_s$  (all daughters of the  $B$ , excluding the primary photon) must satisfy  $0.6 < m_{X_s} < 3.2$  GeV. The  $X_s$  candidate is then combined with the a primary photon to form a  $B$  candidate, which is required to have an energy-substituted mass  $m_{ES} = \sqrt{s/4 - p_B^{*2}}$ , where  $p_B^*$  is the momentum of the  $B$  in the center of mass (CM) frame, greater than 5.24 GeV. We also require that the difference between half of the beam total energy and the energy of the reconstructed  $B$  in the CM frame,  $|\Delta E| = |E_{\text{Beam}}^*/2 - E_B^*|$  be less than 0.15 GeV. The angle between the thrust axis of the rest of the event and the primary photon must satisfy  $|\cos \theta_{T\gamma}^*| < 0.85$ .

### 5.1.3 Event Selection

After reconstruction, most events have multiple  $B$  candidates. In addition, generic  $B\bar{B}$  and continuum events will have some candidates that looks very much like a  $B$

Table 5.1: The 38 modes we reconstruct in this analysis; BiType identifies the numeric value we assign to each mode for bookkeeping; charge conjugation is implied.

BiType	Final State	BiType	Final State
1	$B^+ \rightarrow K_S \pi^+ \gamma$	20	$B^0 \rightarrow K_S \pi^+ \pi^- \pi^+ \pi^- \gamma$
2	$B^+ \rightarrow K^+ \pi^0 \gamma$	21	$B^0 \rightarrow K^+ \pi^+ \pi^- \pi^- \pi^0 \gamma$
3	$B^0 \rightarrow K^+ \pi^- \gamma$	22	$B^0 \rightarrow K_S \pi^+ \pi^- \pi^0 \pi^0 \gamma$
4	$B^0 \rightarrow K_S \pi^0 \gamma$	23	$B^+ \rightarrow K^+ \eta (\rightarrow \gamma \gamma) \gamma$
5	$B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$	24	$B^0 \rightarrow K_S \eta (\rightarrow \gamma \gamma) \gamma$
6	$B^+ \rightarrow K_S \pi^+ \pi^0 \gamma$	25	$B^+ \rightarrow K_S \eta (\rightarrow \gamma \gamma) \pi^+ \gamma$
7	$B^+ \rightarrow K^+ \pi^0 \pi^0 \gamma$	26	$B^+ \rightarrow K^+ \eta (\rightarrow \gamma \gamma) \pi^0 \gamma$
8	$B^0 \rightarrow K_S \pi^+ \pi^- \gamma$	27	$B^0 \rightarrow K^+ \eta (\rightarrow \gamma \gamma) \pi^- \gamma$
9	$B^0 \rightarrow K^+ \pi^- \pi^0 \gamma$	28	$B^0 \rightarrow K_S \eta (\rightarrow \gamma \gamma) \pi^0 \gamma$
10	$B^0 \rightarrow K_S \pi^0 \pi^0 \gamma$	29	$B^+ \rightarrow K^+ \eta (\rightarrow \gamma \gamma) \pi^+ \pi^- \gamma$
11	$B^+ \rightarrow K_S \pi^+ \pi^- \pi^+ \gamma$	30	$B^+ \rightarrow K_S \eta (\rightarrow \gamma \gamma) \pi^+ \pi^0 \gamma$
12	$B^+ \rightarrow K^+ \pi^+ \pi^- \pi^0 \gamma$	31	$B^0 \rightarrow K_S \eta (\rightarrow \gamma \gamma) \pi^+ \pi^- \gamma$
13	$B^+ \rightarrow K_S \pi^+ \pi^0 \pi^0 \gamma$	32	$B^0 \rightarrow K^+ \eta (\rightarrow \gamma \gamma) \pi^- \pi^0 \gamma$
14	$B^0 \rightarrow K^+ \pi^+ \pi^- \pi^- \gamma$	33	$B^+ \rightarrow K^+ K^- K^+ \gamma$
15	$B^0 \rightarrow K_S \pi^0 \pi^+ \pi^- \gamma$	34	$B^0 \rightarrow K^+ K^- K_S \gamma$
16	$B^0 \rightarrow K^+ \pi^- \pi^0 \pi^0 \gamma$	35	$B^+ \rightarrow K^+ K^- K_S \pi^+ \gamma$
17	$B^+ \rightarrow K^+ \pi^+ \pi^- \pi^+ \pi^- \gamma$	36	$B^+ \rightarrow K^+ K^- K^+ \pi^0 \gamma$
18	$B^+ \rightarrow K_S \pi^+ \pi^- \pi^+ \pi^0 \gamma$	37	$B^0 \rightarrow K^+ K^- K^+ \pi^- \gamma$
19	$B^+ \rightarrow K^+ \pi^+ \pi^- \pi^0 \pi^0 \gamma$	38	$B^0 \rightarrow K^+ K^- K_S \pi^0 \gamma$

meson. In this section, we detail two different Random Forest (RF)[27][28] decision tree[29] methods used to select the best candidate for each event and reject many of these false  $B$  candidates.

### 5.1.3.1 Best Candidate Selection

Since each event has multiple reconstructed  $B$  candidates, our goal in this section is to select the  $B$  candidate with the correct reconstruction. Our method is based on a Bagged Decision Tree built using six variables:  $\Delta E/\sigma_E$ , where  $\sigma_E$  is the uncertainty of the  $B$  candidate energy, the thrust of the  $B$ , the  $\pi^0$  momentum, the invariant mass of the  $X_s$ , and the zeroth and fifth Fox-Wolfram moments. The distribution of these variables along with their correlations can be found in Figure E.1.

The main idea is to distinguish between correctly reconstructed  $B$  candidates and misreconstructed ones. The appropriate training sample for this task is signal MC

Table 5.2: The 16 modes used for  $A_{CP}$  analysis; BiType identifies the numeric value we assign to each mode for bookkeeping; charge conjugation is implied.

BiType	Final State
1	$B^+ \rightarrow K_S \pi^+ \gamma$
2	$B^+ \rightarrow K^+ \pi^0 \gamma$
3	$B^0 \rightarrow K^+ \pi^- \gamma$
5	$B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$
6	$B^+ \rightarrow K_S \pi^+ \pi^0 \gamma$
7	$B^+ \rightarrow K^+ \pi^0 \pi^0 \gamma$
9	$B^0 \rightarrow K^+ \pi^- \pi^0 \gamma$
11	$B^+ \rightarrow K_S \pi^+ \pi^- \pi^+ \gamma$
12	$B^+ \rightarrow K^+ \pi^+ \pi^- \pi^0 \gamma$
13	$B^+ \rightarrow K_S \pi^+ \pi^0 \pi^0 \gamma$
14	$B^0 \rightarrow K^+ \pi^+ \pi^- \pi^- \gamma$
16	$B^0 \rightarrow K^+ \pi^- \pi^0 \pi^0 \gamma$
23	$B^+ \rightarrow K^+ \eta(\rightarrow \gamma\gamma) \gamma$
27	$B^0 \rightarrow K^+ \eta(\rightarrow \gamma\gamma) \pi^- \gamma$
33	$B^+ \rightarrow K^+ K^- K^+ \gamma$
37	$B^0 \rightarrow K^+ K^- K^+ \pi^- \gamma$

that provides candidates in which all tracks correctly map to the corresponding MC truth information. These are used as signal in the training sample; the rest of the candidates from signal MC are used as background. We call this classifier Signal Selecting Classifier (SSC).

To select the best  $X_s$  candidate for each event, the response for every candidate that has  $m_{ES} > 5.24 \text{ GeV}/c^2$  is calculated. We then select the candidate that has the maximum SSC response. The distribution of the maximum response for events in which the best candidate is selected correctly, and events in which the best candidate selected is not a true signal candidate is shown in Figure 5.1. Figure 5.2 shows the comparison of signal rate and fake rate of our method and the nominal  $|\Delta E|$  minimization method, which was used in the previous analysis. We found that SSC has a lower the amount of cross feed background for every level of signal efficiency, which gives us a big gain in statistical sensitivity compared to the previous analysis.

The response of the SSC is used not only in selecting the best candidate, but also in rejecting crossfeed and  $B\bar{B}$  background (by cutting on the value of SSC re-

sponse). The optimization of the final cut location on this classifier will be discussed in Section 5.1.4.

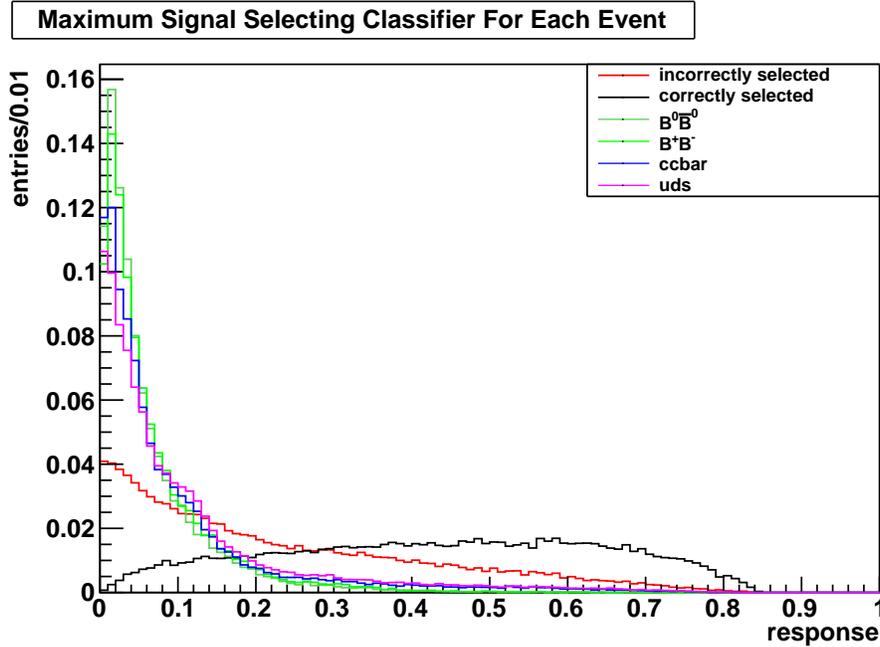


Figure 5.1: Normalized distribution of maximum SSC response for events in which the best candidate was correctly selected (black) and incorrectly selected (red)

### 5.1.3.2 Background Rejecting Classifier, BRC

Much of our background arises from continuum events ( $e^+e^- \rightarrow q\bar{q}$  where  $q \in \{u, d, s, c\}$ ) which have very different kinematics from  $B\bar{B}$  events. For continuum events, since the total energy of the beam is much more than the mass of the light quark pairs, these light quarks fly apart back-to-back with high kinetic energy. The  $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$  process is different. Since the mass of  $\Upsilon(4S)$  is very close to the mass of 2  $B$  mesons, the  $B$  mesons do not have much kinetic energy. They decay almost at rest into many tracks, making the event look spherical. In this section, we will describe how we train a random forest classifier (RF) build specifically to reject continuum background.<sup>2</sup>

<sup>2</sup>This study was done by David A. Doll

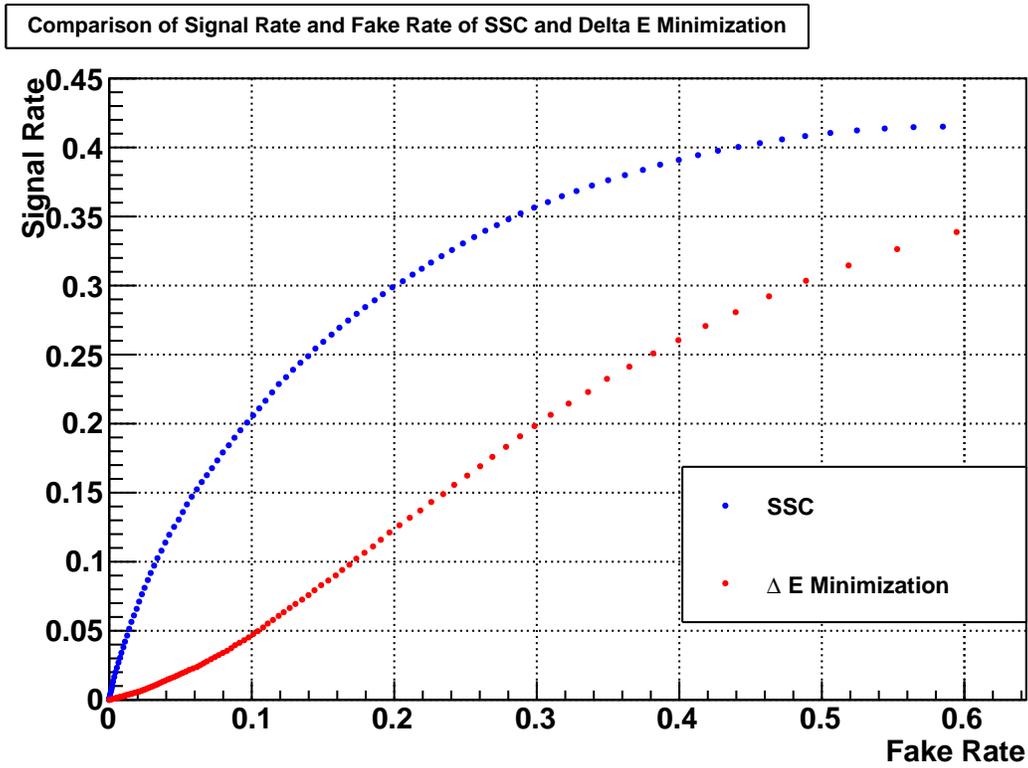


Figure 5.2: Comparison between the signal rate and fake rate for SSC and  $\Delta E$  minimization method

To reject continuum background, we use information from both the reconstructed  $B$  and the rest of the event, which would be  $B$  if the event is  $\Upsilon(4S) \rightarrow B\bar{B}$ . We list below the variables we included in this study as well as a brief description of each. Plots of these variables, along with correlations between each pair of variables can be found in Figure E.2.

- $\pi^0$  score based on the output of a classifier that was trained to separate a photon from  $\pi^0$  from the signal photon. The main purpose is to ensure that the primary photon is not from  $\pi^0 \rightarrow \gamma\gamma$  decay. This is described in detail in Appendix D.
- BmtmFlow1-17: Momentum flow cones in the CM frame about the reconstructed  $B$  direction in  $10^\circ$  increments. The idea being that the distribution of energy can separate more jet-like continuum events from the isotropic/spherical signal. These variables don't show too much rejection power but are included for historical reasons.
- ROE Legendre Monomials along the Photon Axis: The zeroth, first and second order Legendre monomials computed in the CM frame along the primary photon axis. The ROE (rest of the event) corresponds to all particles not involved in the signal  $B$  reconstruction. Since the continuum event is more jet-like than signal, the jet component and the primary photon momentum will be mostly along the same direction.
- $L_{12}/L_{10}$ : The ratio of the second order Legendre monomial to the zeroth when both are computed using the ROE particles (particles not involved in signal  $B$  reconstruction) along the primary photon axis (for the  $X_s$ -mass binned study) or along the thrust axis of the  $B$ -candidate (for the photon energy-binned study).
- $|\cos \theta_B^*|$ : The absolute value of the cosine of the angle of the  $B$  flight direction with respect to the  $z$ -axis, computed in the CM frame.
- $|\cos \theta_T^*|$ : The absolute value of the cosine of the angle between the thrust axis

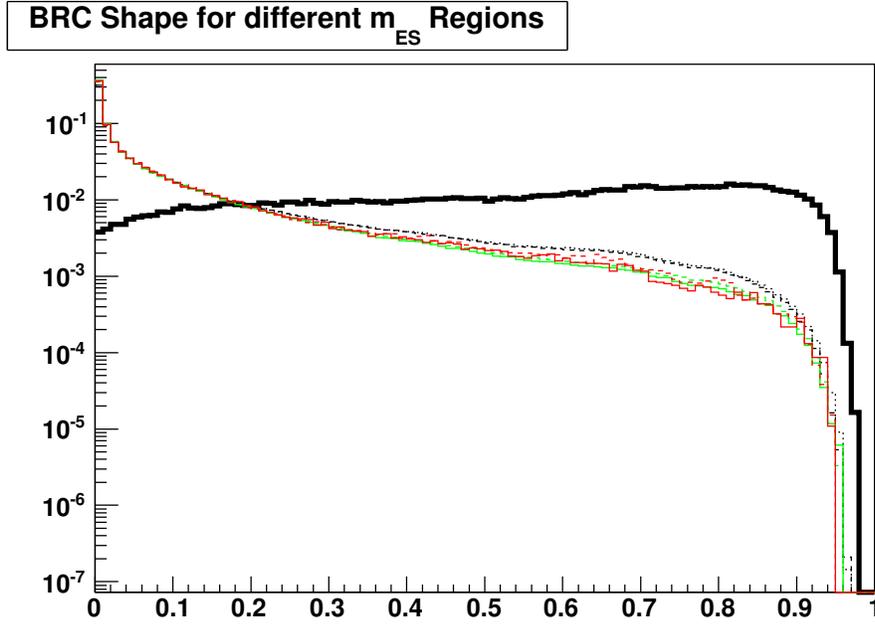


Figure 5.3: The response of the BRC for background candidates with  $m_{ES} > 5.265$  (dashed), background candidates with  $m_{ES} < 5.265$  (dotted), and signal candidates (solid line) for run 3 MC. We also compare offpeak data (red) to continuum MC (green) for  $m_{ES} < 5.265$  (solid) and  $m_{ES} > 5.265$  (dashed).

of the  $B$  candidate and the thrust axis of the ROE (particles not involved in the  $B$  reconstruction), calculated in the CM frame.

- $|\cos \theta_{\gamma T}^*|$ : The absolute value of the cosine of the angle between the primary photon and the ROE particles' thrust, calculated in the CM frame.

The classifier output for various  $X_s$  mass region is shown in Figure 5.3.

#### 5.1.4 Cut Optimization

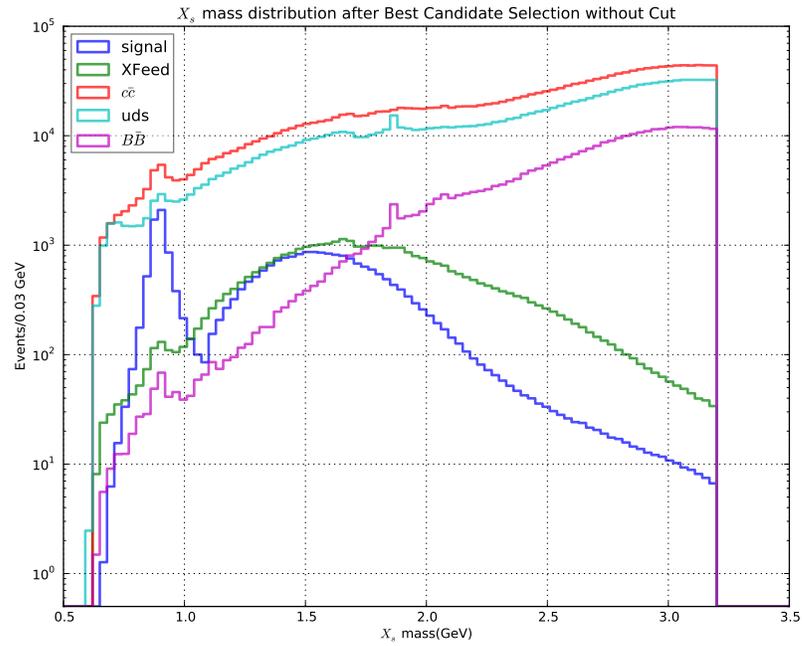
To obtain the best sensitivity, we simultaneously optimize, using MC samples, the SSC and BRC cuts of 4  $X_s$  mass ranges ( $[0.6-1.1]$ ,  $[1.1-2.0]$ ,  $[2.0-2.4]$  and  $[2.4-2.8]$  GeV), maximizing  $S/\sqrt{S+B}$ , where  $S$  is the number of expected signal events and  $B$  is the number of expected background events with  $m_{ES} > 5.27$  GeV. The optimized cut values shown in Table 5.3 are the same for both  $b$  and  $\bar{b}$  flavors.

Table 5.3: Optimal cut value.

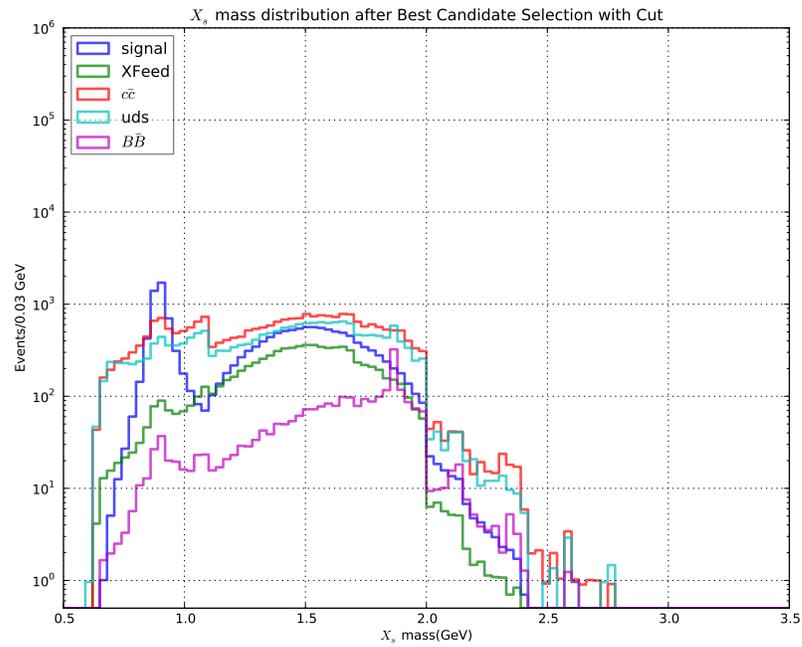
$X_s$ mass( GeV)	SSC	BRC
0.6-1.1	> 0.14	> 0.24
1.1-2.0	> 0.22	> 0.38
2.0-2.4	> 0.39	> 0.52
2.4-2.8	> 0.48	> 0.46

The  $X_s$  mass distribution before and after selection is show in Figure 5.4. Since the efficiencies varies based on  $X_s$  mass, one could be concerned that this could introduce a photon model dependence on the analysis because the cuts was made from model we used in the MC sample. However, since we are measuring  $A_{CP}$  which is effectively the ratio of branching fractions, the dependence of the efficiency on the choice of photon model which used in weighting the signal MC (that is used for training the classifier) would not affect the result much; it will just change the absolute efficiency which cancels out in the ratio. Moreover,  $A_{CP}$  dependence on  $X_s$  mass is expected to be small [2].

Appendix F contains precision for all optimization strategies we tried and also the expected number of each type of event for various  $X_s$  mass region can be found in Table F.2.



(a) Before optimal cut



(b) After optimal cut

Figure 5.4:  $X_s$  mass distribution before(a) and after (b) the optimal cut.

## 5.2 $A_{CP}$ Extraction

In this section we will describe a simple  $A_{CP}$  extraction method based on simultaneous fit of  $B m_{ES}$ <sup>3</sup>. We will then discuss quantitatively two major concerns about the fitting method. Then, we go on to fill out the details of the fitting method and show that the statistical sensitivity we expect is better than  $1/\sqrt{N}$  scaling compared to the previous analysis[12].

### 5.2.1 Fitting Procedure

Our goal here is to extract  $A_{CP}$  from the yield of  $b$  and  $\bar{b}$  flavor events. To achieve this, we fit the  $m_{ES}$  distribution of  $b$  and  $\bar{b}$  flavor simultaneously with the signal yield of both flavors linked together with  $A_{CP}$ . An example of an  $m_{ES}$  distribution for both flavors with a breakdown for each component can be found in Figure 5.5.

Specifically, our PDF for  $m_{ES}$  for each flavor is composed to two parts: continuum and peaking. The continuum part is described by an Argus distribution (Equation L.1) and the peaking part is described by a Cruijff distribution (Equation L.7). To be exact, we use the following extended PDF:

$$\text{PDF}(m_{ES}; c, \chi, p, m_0, \sigma_L, \sigma_R, \alpha_L, \alpha_R, N_{cont}, N_{peak}) = \\ N_{cont} \times \text{Argus}(m_{ES}, c, \chi, p) + N_{peak} \times \text{Cruijff}(m_{ES}, m_0, \sigma_L, \sigma_R, \alpha_L, \alpha_R).$$

To extract  $A_{CP}$ , we find the shape parameters that minimize the sum of the negative log likelihood (NLL) of  $b$  and  $\bar{b}$  flavor.

$$\text{total NLL} = \text{NLL}(\text{PDF}(m_{ES}; \dots, N_{cont}^b, N_{peak}^b), \text{data}^b) + \\ \text{NLL}(\text{PDF}(m_{ES}; \dots, N_{cont}^{\bar{b}}, N_{peak}^{\bar{b}}), \text{data}^{\bar{b}}),$$

where we rewrite  $N_{cont}^b$ ,  $N_{cont}^{\bar{b}}$ ,  $N_{peak}^b$ ,  $N_{peak}^{\bar{b}}$  in term of  $A_{cont}$ ,  $A_{peak}$ ,  $T_{cont}$  and  $T_{peak}$

---

<sup>3</sup>Invariant mass of  $B$  but use half the energy of the beam instead of reconstructed energy.  $m_{ES} = \sqrt{s/4 + p_b^{*2}}$

defined as

$$A_{cont} \equiv \frac{N_{cont}^b - N_{cont}^{\bar{b}}}{N_{cont}^b + N_{cont}^{\bar{b}}}, \quad (5.2)$$

$$T_{cont} \equiv N_{cont}^b + N_{cont}^{\bar{b}}, \quad (5.3)$$

$$A_{peak} \equiv \frac{N_{peak}^b - N_{peak}^{\bar{b}}}{N_{peak}^b + N_{peak}^{\bar{b}}}, \quad (5.4)$$

$$T_{peak} \equiv N_{peak}^b + N_{peak}^{\bar{b}}, \quad (5.5)$$

with the following expressions:

$$N_{cont}^b = \frac{T_{cont}}{2} \times (1 + A_{cont}), \quad (5.6)$$

$$N_{cont}^{\bar{b}} = \frac{T_{cont}}{2} \times (1 - A_{cont}), \quad (5.7)$$

$$N_{peak}^b = \frac{T_{peak}}{2} \times (1 + A_{peak}), \quad (5.8)$$

$$N_{peak}^{\bar{b}} = \frac{T_{peak}}{2} \times (1 - A_{peak}). \quad (5.9)$$

We compute our log likelihood using a binned Poisson likelihood with 200 bins between  $m_{ES} \in (5.24, 5.29)$  GeV. We then use MINUIT [30] to perform multidimensional minimization based on the Davidon-Fletcher-Powell formula and find  $A_{peak}$  along with other shape parameters, that give the minimum likelihood. An example of a fit to toy sample is shown in Figure 5.5.

We intentionally left out several details in this description. In simultaneous fitting of two very similar distributions, we can share some shape parameters of the two distributions to gain more statistical power via reduction of correlation. We can also fix the values of some shape parameters that are unstable or would introduce unnecessary correlation. We will explore these choices in Section 5.2.4.1.

In equation 5.4, where we express  $N_b$  and  $N_{\bar{b}}$  in terms of  $A$  and  $T$ , we made an implicit assumption that  $N_b$  and  $N_{\bar{b}}$  consists purely of signal yield. This is not entirely correct. One may notice that we have a couple types of background: continuum, peaking  $B\bar{B}$  and cross feed. The latter two peak at the same  $m_{ES}$  as our signal. But,

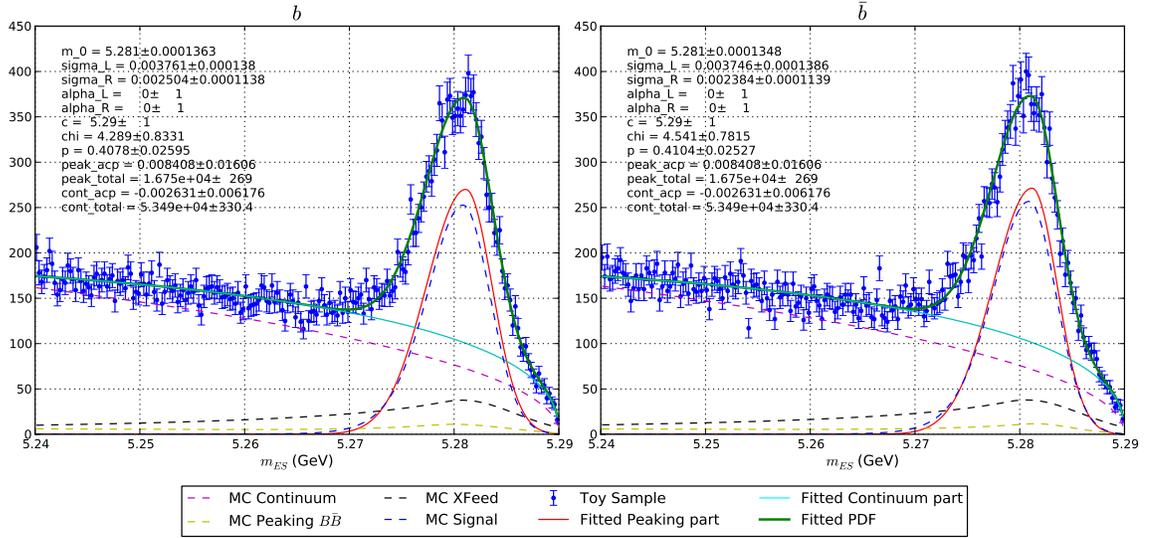


Figure 5.5: Example fit to a toy sample using the fix alpha strategy. The fit to the  $b$   $m_{ES}$  distribution is shown on the left and the fit to the  $\bar{b}$   $m_{ES}$  distribution is shown on the right. The continuum distribution is shown with light blue, the peaking distribution is shown in red and the sum of the two is shown in green. Monte Carlo components are shown in dashed lines. The toys used in this figure are from all  $B$  charges and full  $X_s$  mass range.

the fitting method described made no effort to separate peaking background from signal. This would mean that peaking component yield will be a mixed of signal, cross feed and  $B\bar{B}$  background; thus, invalidating our assumption. The effect can be seen clearly in Figure 5.5 that the red fitted signal do not line up exactly with blue dashed line of MC signal shape. In Section 5.2.2 and 5.2.3, we will show quantitatively that because of our ability of reject peaking  $B\bar{B}$  and cross feed background, this dilution effect on the total  $A_{CP}$  is small compared to expected sensitivity.

## 5.2.2 Peaking Background Dilution

As described in the previous section, our peaking yield will be a mix of signal and peaking background. This will dilute our measured  $A_{CP}$ . We want to quantify the effect of signal contamination.

To begin with, we want to extract the following quantity

$$\text{Pure } A_{CP} = \frac{n_b - n_{\bar{b}}}{n_b + n_{\bar{b}}}, \quad (5.10)$$

where  $n_b$  and  $n_{\bar{b}}$  are the number of signal events for each flavor.

However,  $n_b$  and  $n_{\bar{b}}$  may have peaking background mixed in. Thus, what we actually extract from the fit is

$$\text{Contaminated } A_{CP} = \frac{(n_b + p_b) - (n_{\bar{b}} + p_{\bar{b}})}{(n_b + p_b) + (n_{\bar{b}} + p_{\bar{b}})}, \quad (5.11)$$

where  $p_b$  and  $p_{\bar{b}}$  are the amount of contaminating peaking background for each flavor.

We are interested in the how much contamination will affect our measured  $A_{CP}$ . The quantity of interest is

$$\delta A = \text{Contaminated } A_{CP} - \text{Pure } A_{CP} \quad (5.12)$$

$$= \frac{(n_b + p_b) - (n_{\bar{b}} + p_{\bar{b}})}{(n_b + p_b) + (n_{\bar{b}} + p_{\bar{b}})} - \frac{n_b - n_{\bar{b}}}{n_b + n_{\bar{b}}} \quad (5.13)$$

$$= \left( \frac{p_b + p_{\bar{b}}}{n_b + n_{\bar{b}} + p_b + p_{\bar{b}}} \right) \times \left( \frac{n_b - n_{\bar{b}}}{n_b + n_{\bar{b}}} - \frac{p_b - p_{\bar{b}}}{p_b + p_{\bar{b}}} \right). \quad (5.14)$$

We want to emphasize that Equation 5.14 is an exact expression, and this choice of factorization have an intuitive meaning. The first factor,

$$S = \frac{p_b + p_{\bar{b}}}{n_b + n_{\bar{b}} + p_b + p_{\bar{b}}},$$

is the ratio of total peaking background over the total number of peaking component. The better our selection in eliminating peaking background, the smaller the  $S$ . In the limit where there is no peaking background contamination ( $p_b + p_{\bar{b}} = 0$ ),  $\delta A$  vanishes as expected.

The second factor,

$$\Delta = \frac{n_b - n_{\bar{b}}}{n_b + n_{\bar{b}}} - \frac{p_b - p_{\bar{b}}}{p_b + p_{\bar{b}}},$$

is the difference between the  $A_{CP}$  of pure signal and the  $A_{CP}$  of the peaking back-

ground. In the limit where the  $A_{CP}$  of the peaking background and the signal are the same, we do not really care about the amount peaking background contamination.

We can estimate  $\delta A$  by estimating  $S$  and  $\Delta$ . For  $S$ , we use the ratio of peaking background and total peaking component with  $m_{ES}$  over 5.27 GeV found in Table F.3. We found  $S$  to be  $\sim 0.11$  for the  $K^*$  region ( $m_{X_s} \in [0.6, 1.1]$  GeV) and  $\sim 0.26$  for the full  $X_s$  mass range. This calculation is actually an overestimate of  $S$  since part of the  $B\bar{B}$  component and cross feed background will be picked up by the Argus distribution as well. This can be seen in Figure 5.5 where most of peaking component is actually picked up by Argus distribution.

To estimate the value of  $\Delta$ , we use the values at the two ends of the Standard Model prediction with long distance effects included[11]. We use  $\Delta \sim 3.4\%$ . Together with our estimated value of  $S$  we found the upper bound of  $\delta A$  to be 0.37% for the  $K^*$  region and 0.8% for the whole mass range. This is very small compared to the expected sensitivity of our analysis, 1.6%. The results are summarized in Table 5.4.

Table 5.4: Value of  $S$ ,  $\Delta$  and  $\delta A$  for the  $K^*$  region and the whole mass range.

Mass Range	$B$ Sample	$S$	$\Delta$	$\delta A$
$K^*$	All $B$	0.11	3.4%	0.38%
$K^*$	Charged $B$	0.15	3.4%	0.52%
$K^*$	Neutral $B$	0.09	3.4%	0.30%
All	All $B$	0.26	3.4%	0.88%
All	Charged $B$	0.28	3.4%	0.80%
All	Neutral $B$	0.24	3.4%	0.97%

### 5.2.3 The effect on Statistical Uncertainty of Peaking Background Contamination

Another effect from peaking background contamination comes from the fact that our signal yield will appear to be larger than what it actually is; thereby making the statistical uncertainty on  $A_{CP}$  smaller than what it is supposed to be. In the limit where  $n_b \approx n_{\bar{b}}$ , the statistical uncertainty of  $A_{CP}$  (neglecting the correlation from

shape parameters) is given by.

$$\sigma_{stat} \approx \frac{\sqrt{\sigma_{n_b}^2 + \sigma_{n_{\bar{b}}}^2}}{n_b + n_{\bar{b}}}.$$

Assuming a Poisson uncertainty on  $n_b$  and  $n_{\bar{b}}$ , we have

$$\sigma_{stat} \approx \frac{1}{\sqrt{n_b + n_{\bar{b}}}} \approx \frac{1}{\sqrt{n}}.$$

Let the contamination factor be  $k$  such that total number signal events and the contamination is  $k \times n$ . We have

$$\sigma_{contaminated} \approx \frac{1}{\sqrt{kn}} = \frac{1}{\sqrt{k}} \sigma_{stat}.$$

Thus

$$\sigma_{stat} = \sqrt{k} \times \sigma_{contaminated}.$$

We estimate  $\sqrt{k}$  using the ratio of peaking background over the sum of all peaking components with  $m_{ES}$  over 5.27 GeV found in Table F.3. We found that  $\sqrt{k} = 1.06$  for  $K^*$  region and  $\sqrt{k} = 1.16$  for full mass range. This multiplicative factor adds very little to overall statistical uncertainty.

## 5.2.4 Toy Study

In this section, we will show how we use the toy samples to

1. determine fitting strategy (which parameter to share between the two flavors and which parameters to fix).
2. determine the sensitivity and bias in various situations.

We generate 6 sets of toy samples categorized by the charge of the  $B \in \{\text{neutral, charged, all}\}$  and  $b$  flavor  $\in \{b, \bar{b}\}$  using shape parameters obtained from a fit to the MC. It should be emphasized that toy samples for  $b$  and  $\bar{b}$  flavor are generated using

a different set of shape parameters. The detailed procedure for generating the toy sample can be found in Appendix H.

#### 5.2.4.1 Fitting Strategy

In section 5.2.1 we left out some details of our fitting strategy. Specifically, we left out the information on which parameters to be fixed and which parameters to be shared between the two flavors. We have several issues to consider here. If we share parameters, we will be able to reduce the correlation between the continuum  $A_{CP}$  and the peaking  $A_{CP}$ . We do not expect the shape of the two to differ drastically, but if we were to share shape parameters, we would need to evaluate the systematics for such a decision. This can be difficult. Another issue to consider is which parameters to be fixed fix. Notably,  $\alpha_L$  of the Cruijff distribution, which defines the tail of peaking distribution, contributes significantly to the correlation between the peaking  $A_{CP}$  ( $A_{\text{peak}}$ ) and the continuum  $A_{CP}$  ( $A_{\text{cont}}$ ). The reason is that the shape produced by  $\alpha_L$  at low  $m_{\text{ES}}$  ( $<5.27$  GeV) is very similar to the Argus distribution.

To determine which parameters to share and which parameters to fix, we find the statistical sensitivity on  $A_{\text{peak}}$  of all  $B$  toy samples generated with 0% signal  $A_{CP}$  using the various fitting strategies listed on Table 5.5. We fit 1000 toy samples for each strategy to find statistical sensitivity and bias. The results are shown in Figure 5.6.

We observe several patterns from Figure 5.6. First, sharing the Argus shape parameters introduces a small ( $\sim 0.5\%$ ) bias. This is because the toy we use for each  $b$  flavor is generated with different sets of shape parameters. When we share the shape parameter for the Argus distribution, the Argus distribution for each flavor systematically absorb different amount of peaking background; thus, giving us the bias.

Second, separately floating  $\alpha_L$ , which defines the left side tail of Cruijff distribution, introduces large and unnecessary correlations between the continuum  $A_{CP}$  and the peaking  $A_{CP}$ , thus increasing the uncertainty. Our signal distribution has a very short tail. The long tail of the peaking distribution will be composed purely

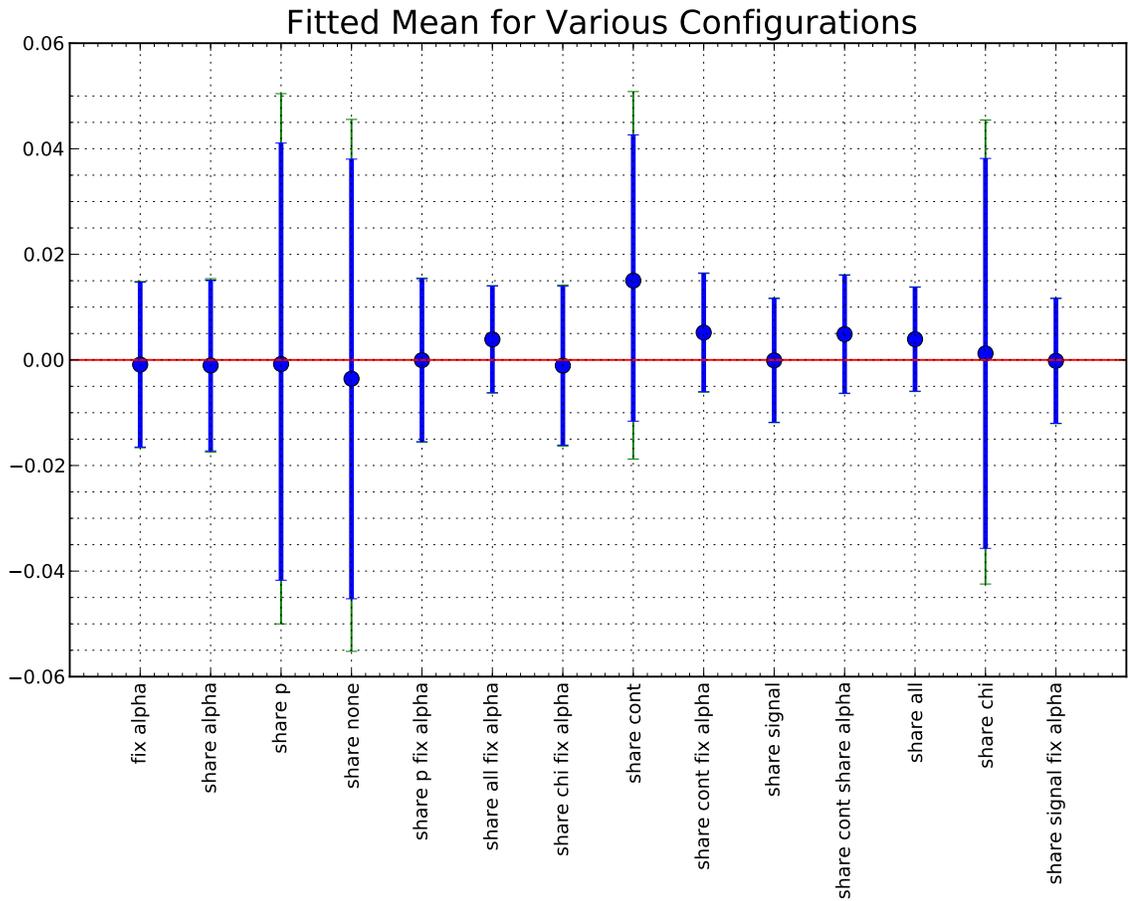


Figure 5.6: Bias and sensitivity of each fitting strategy. The description of each fitting strategy can be found in Table 5.5. The blue error bars represent the 50% percentile of the MINOS error in the MINUIT package and the green error bars represent the 75% percentile of the MINOS error.

of crossfeed and peaking  $B\bar{B}$  and is very similar in shape to the tail of the Argus distribution. We want the yield of the signal part to be of as pure as possible; thus fixing  $\alpha_L$  is desirable. Further, we found that can describe our signal distribution quite well with just a simple bifurcated gaussian (see Figure 5.5).

Table 5.5: Description of each fitting strategy. The parameters that are not listed in any of the two columns are floated separately.

Strategy	Fixed Parameters	Shared Parameters
Fix $\alpha$	$\alpha_L=0$	
Share $\alpha$		$\alpha_L$
Share $p$		$p$
Share $\chi$		$\chi$
Share none		
Share $p$ fix $\alpha$	$\alpha_L=0$	$p$
Share chi fix $\alpha$	$\alpha_L=0$	$\chi$
Share cont		$\chi, p$
Share cont fix $\alpha$	$\alpha_L=0$	$\chi, p$
Share cont share $\alpha$		$\chi, p, \alpha_L$
Share signal		$\sigma_L, \sigma_R, \alpha_L$
Share signal fix $\alpha$	$\alpha_L=0$	$\sigma_L, \sigma_R$
Share all		$\chi, p, \sigma_L, \sigma_R, \alpha_L$
Share all fix $\alpha$	$\alpha_L=0$	$\chi, p, \sigma_L, \sigma_R$

From this study, we choose fix  $\alpha$  as our fitting strategy. It can be described simply as using a bifurcated gaussian as our signal PDF and an Argus distribution to model our background. In addition, all shape parameters are floated separately for each  $b$  flavor. This means that this fitting strategy does not rely on shape parameter information from the MC at all. Furthermore, this strategy gives us a reasonable statistical sensitivity with negligible bias. Also, with this strategy we do not need to evaluate the systematics of the shared parameter since we let them all float separately. Evaluating systematics of such effects is non-trivial.

#### 5.2.4.2 Sensitivity, Bias and Robustness at Non-zero $A_{CP}$

After selecting our fitting strategy, we want to test the robustness of our  $A_{CP}$  extraction method. Specifically, we want to answer the following two questions:

1. What is the sensitivity and bias at non-zero signal  $A_{CP}$ ?
2. What is the sensitivity and bias when peaking background has non-zero  $A_{CP}$ ?

For the first question, the sensitivity theoretically depends on the value of  $A_{CP}$ . Fortunately, however, we expect the deviation to be very small for the range of expected  $A_{CP}$  of a few percents. We also expect a small bias whenever there is a difference in  $A_{CP}$  of the peaking background and the signal  $A_{CP}$  as described in Section 5.2.2.

To answer the question about sensitivity and bias at non-zero  $A_{CP}$ , we fit 1000 toy samples each at different signal  $A_{CP}$  with crossfeed  $A_{CP}$  fixed at zero. The results are shown in Figure 5.7 for the full  $X_s$  mass range, and in Figure 5.8 for the  $K^*$  region. Since there is a difference in signal  $A_{CP}$  and cross feed  $A_{CP}$ , in some of the toy set we use, we expect the dilution effect as described in Section 5.2.2 making the slope of the fitted  $A_{CP}$  and the true  $A_{CP}$  a bit less than 1, and this is confirmed in both figures.

We found the statistical sensitivity of the all  $B$  full  $X_s$  mass range  $A_{CP}$  to be 1.6% and we found that the sensitivity does not depend much on the value of  $A_{CP}$  itself. Sensitivities for other  $B$  categories are summarized in Table 5.6. Our analysis contains roughly 20% more data than the previous analysis. We found that our expected sensitivity improve by more than  $1/\sqrt{N}$  law from the previous *BABAR* analysis of 3%[12]. This is achieved mostly because of better best candidate selection and improved peaking background rejection. However, our  $K^*$  sensitivity is worse than the previous dedicated  $B \rightarrow K^*\gamma$  analysis of  $\sim 1.7\%$ [31]. This is mainly because we do not use the helicity angle of the  $K$  and  $K^*$ .

Table 5.6: Expected Sensitivity for  $A_{CP}$  at 0% signal  $A_{CP}$  and 0% background  $A_{CP}$ .

$B$ Category	$K^*$ Mass Range	Full Mass Range
All $B$	2.5%	1.6%
Charged $B$	5%	2.3%
Neutral $B$	2.8%	2.0%

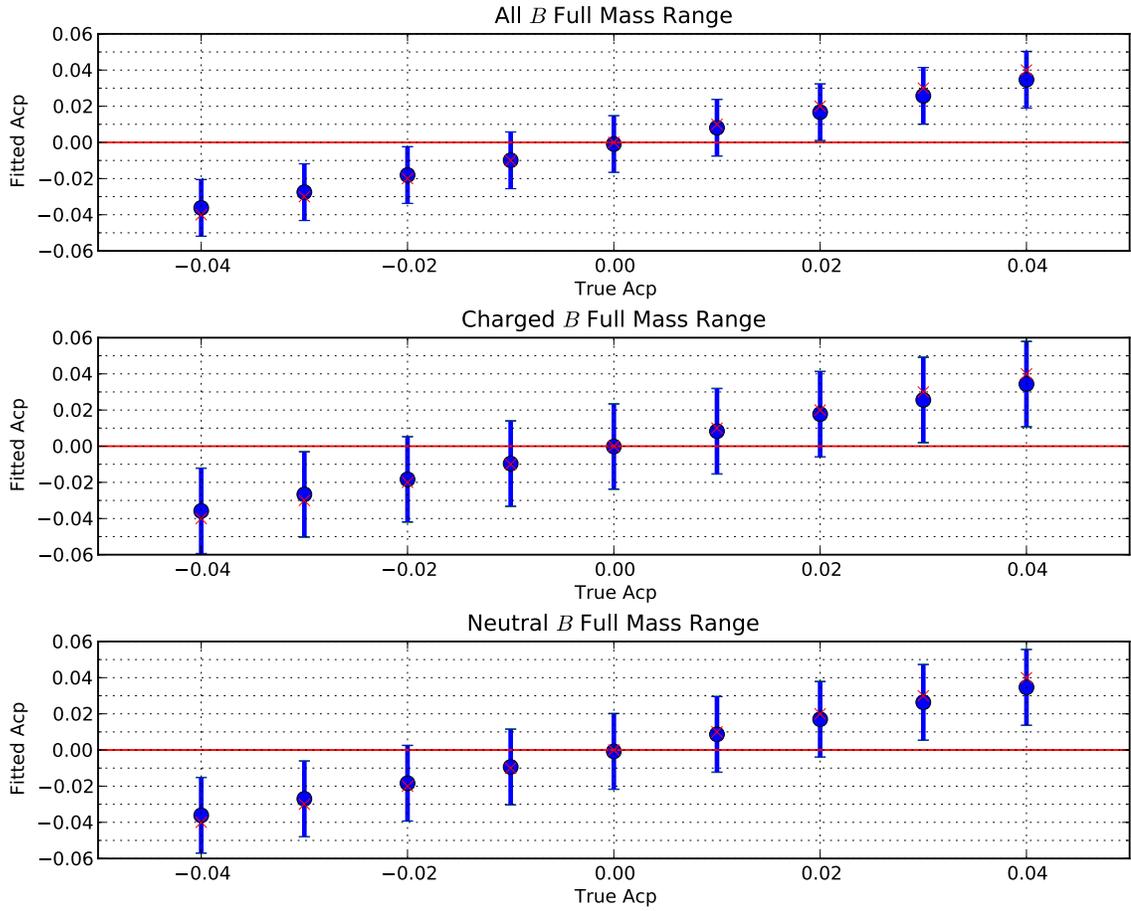


Figure 5.7: Sensitivity and bias of signal  $A_{CP}$  at various values of signal  $A_{CP}$  for all  $B$  (top), charged  $B$  (mid) and neutral  $B$  (bottom) for the full mass range. Blue points with error bars indicate the central value of the fitted  $A_{CP}$  and its sensitivity. The true signal  $A_{CP}$  are shown as red crosses.

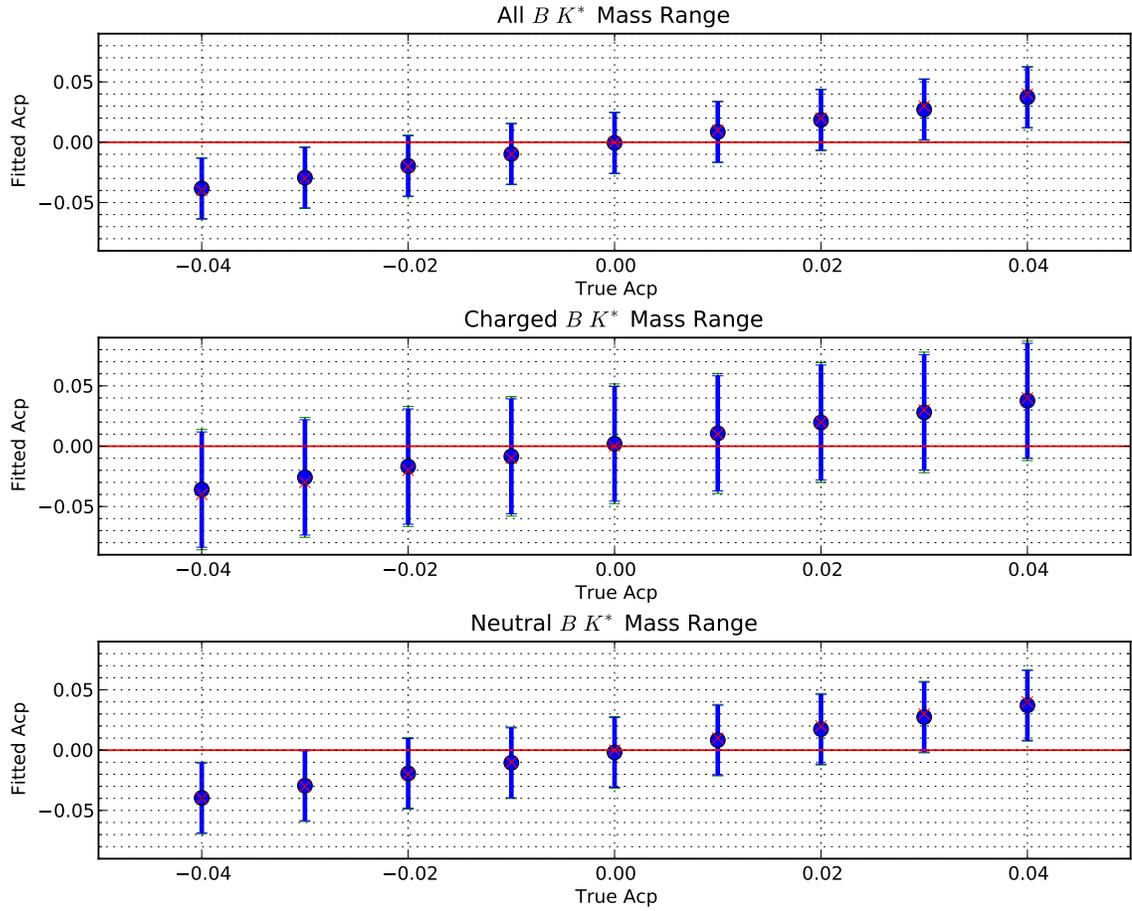


Figure 5.8: Sensitivity and bias of signal  $A_{CP}$  at various values of signal  $A_{CP}$  for all  $B$  (top), charged  $B$  (mid) and neutral  $B$  (bottom) for the  $K^*$  range. Blue points with error bars indicate the central value of fitted  $A_{CP}$  and its sensitivity. True signal  $A_{CP}$  are shown as red crosses.

We did a similar study to answer the second question. We fit 1000 toy samples each at different crossfeed  $A_{CP}$  while fixing the signal  $A_{CP}$  at 0. The results are shown in Figure 5.10 for the  $K^*$  region and Figure 5.9 for the full  $X_s$  mass range. Again, since there is a different in  $A_{CP}$  between signal and peaking background, we expect the dilution effect to make the slope of fitted signal  $A_{CP}$  and crossfeed  $A_{CP}$  slightly positive. We confirm the estimate done in Section 5.2.2 that the deviation should be less than 0.8% for the whole mass range and negligibly small for the  $K^*$  mass region.

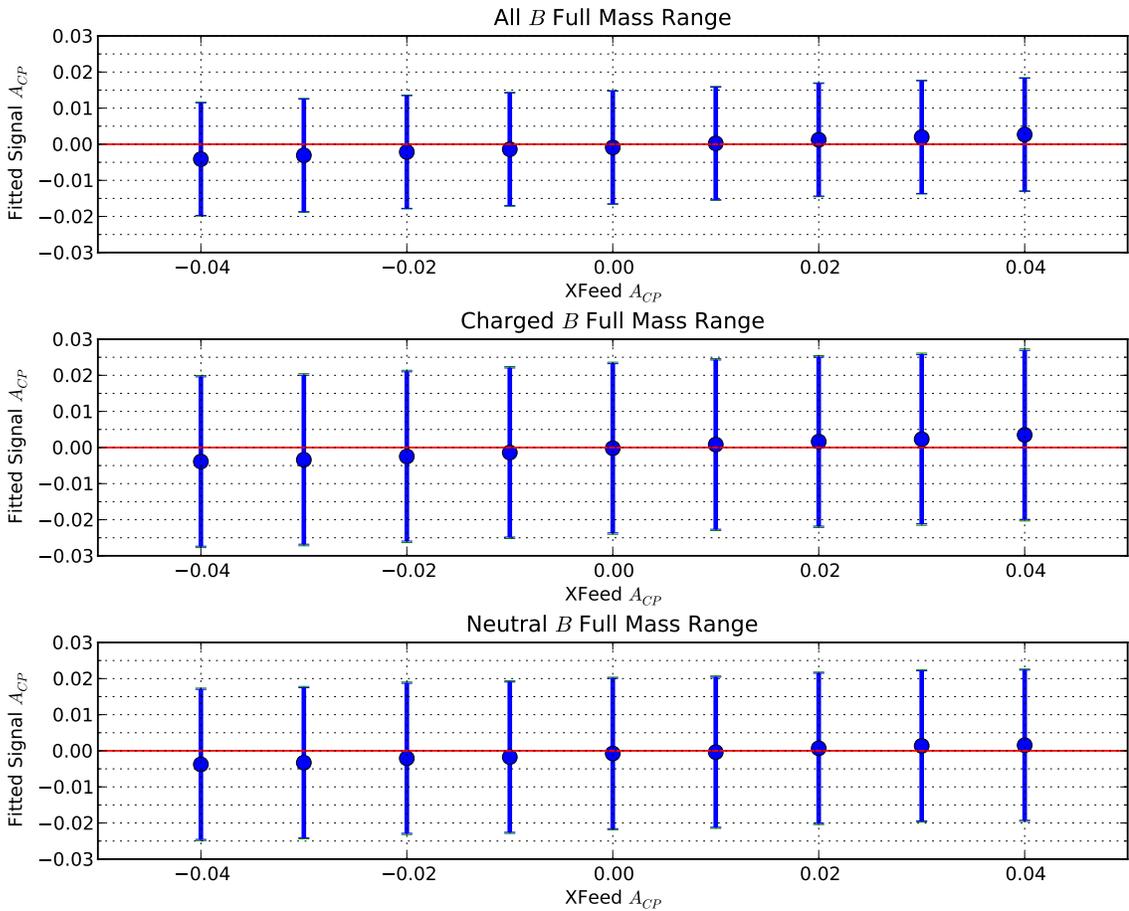


Figure 5.9: Fitted  $A_{CP}$  of toy with 0  $A_{CP}$  and varying cross feed  $A_{CP}$  for full  $X_s$  mass range. From top to bottom: all  $B$ , charged  $B$  and neutral  $B$ . The error bars shown are statistical sensitivities.

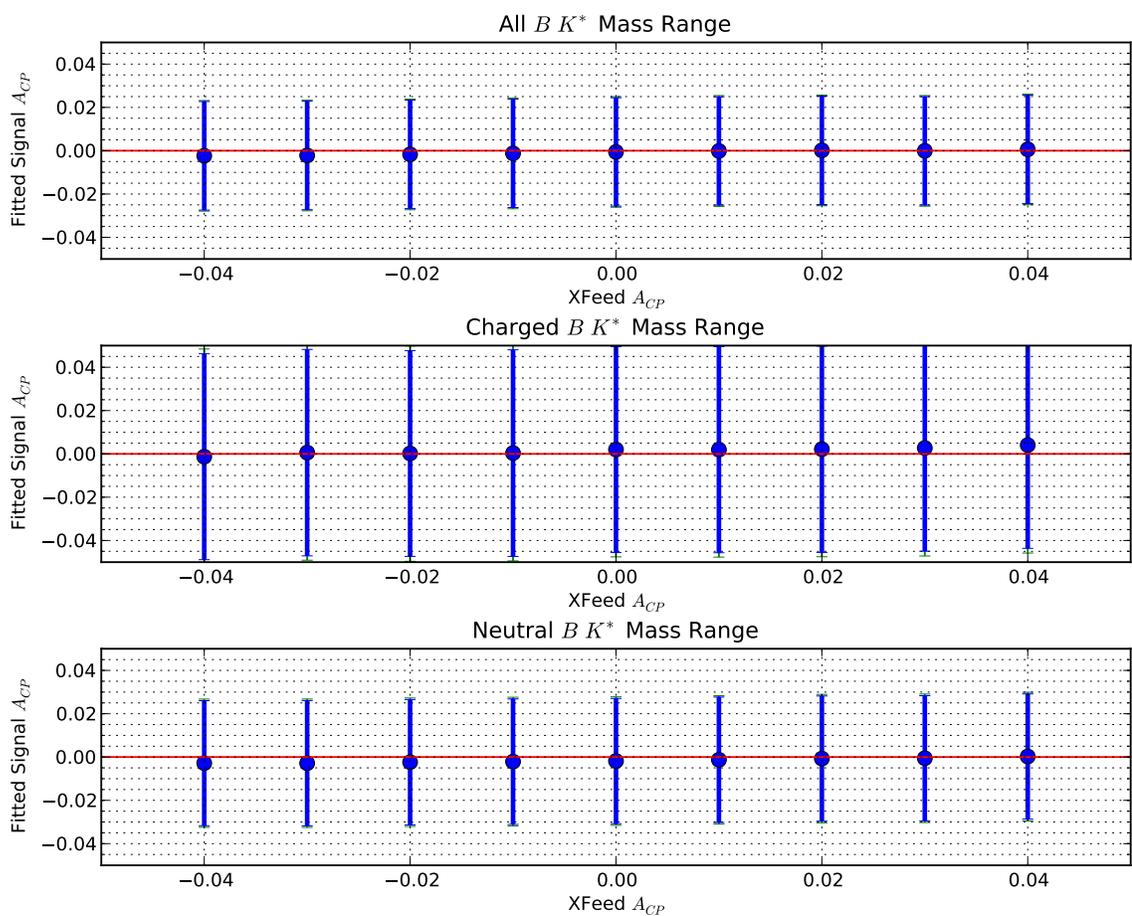


Figure 5.10: Fitted  $A_{CP}$  of toy with 0  $A_{CP}$  and varying crossfeed  $A_{CP}$  for  $K^*$  mass region. From top to bottom: all  $B$ , charged  $B$  and neutral  $B$ . The error bars shown are statistical sensitivities.

## 5.3 Detector $A_{CP}$

Before we go on to the unblind results, there is one issue that we swept under the rug when we describe the  $A_{CP}$  extraction procedure: our detector has an inherent  $CP$  asymmetry, since it is made of matter. This asymmetry will affect the  $A_{CP}$  obtained from the described procedure. In this section, we will show the source of asymmetry and how to correct for it.

### 5.3.1 Source of Detector $A_{CP}$

Our detector is made from matter and not anti-matter. There is a difference in the kaon-nucleon cross section, especially at low ( $< 1 \text{ GeV}/c$ ) momentum (see Appendix K) due to the additional process of  $K^-$  capture by nuclei. This translates to the slightly higher chance for a  $K^-$  to shower before it reaches the DIRC<sup>4</sup> or EMC,<sup>5</sup> thus lowering our ability to use the DIRC information to separate  $K^-$  and  $\pi^-$ . the  $K^-$  has a slightly lower efficiency than a  $K^+$  at low momentum.

To illustrate this effect, let us take a look at the difference in Particle Identification (PID) efficiency of  $K$  and  $\pi$ . We look at the decay sample  $D^{*\pm} \rightarrow D^0 \pi_{slow}^\pm; D^0 \rightarrow K^+ \pi^-$ . These events can be selected very cleanly by cutting on the mass of  $D^*$  and  $D$ . We can identify  $K$  and  $\pi$  without relying on PID for the particular track of interest. It is very important that we do not rely on PID for the particular track, since we want to measure the PID efficiency. A  $K$  can be identified by applying  $\pi$  PID on the other track. Similarly, we can identify  $\pi$  and  $\pi_{slow}$  by applying  $K$  PID on the other track. To measure the efficiency, we applied  $K$  PID to  $K$  sample we obtained without using  $K$  PID. We take the ratio of the total  $K$  sample and the  $K$  sample that passes  $K$  PID. The efficiency for  $K^+$  and  $K^-$  along with the ratio is plotted against track momentum in Figure 5.11, the efficiency for  $\pi^+$  and  $\pi^-$  along with the ratio is shown on Figure 5.12.

As Figure 5.11 illustrates, the difference between the efficiencies varie from 0-2.5%

---

<sup>4</sup>Detection of Internally Reflected Cherenkov light

<sup>5</sup>EletroMagnetic Calorimeter

for  $K$  depending on the momentum of the track. It should be noted that the error bars shown are underestimated. The error bars do not properly take into account the variation from one bin to the next. This is for illustration purpose only. For  $\pi$ , shown in Figure 5.12, the difference is very small and does not depend on track momentum.

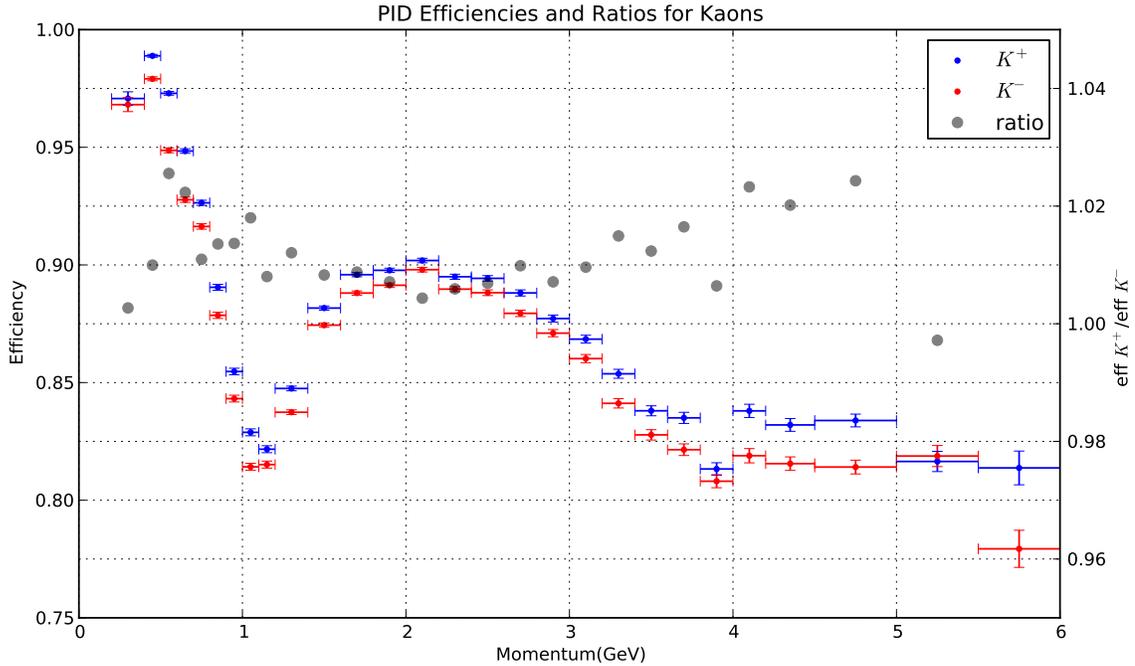


Figure 5.11: PID efficiencies for  $K^+$  and  $K^-$ . The efficiency, uncertainties on efficiencies, and efficiency ratios shown are calculated after summing over  $\theta$  bins. These uncertainties on efficiencies are vastly underestimated due to the fact that momentum bin is too wide compared to variation of efficiencies, especially at low momentum.

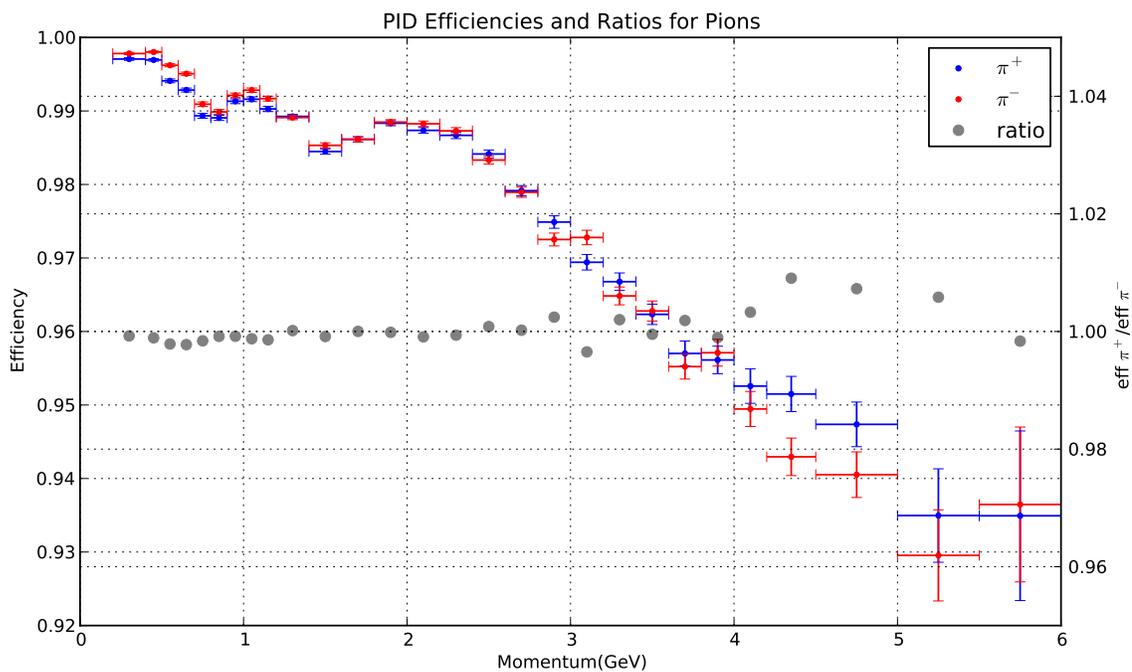


Figure 5.12: PID efficiencies for  $\pi^+$  and  $\pi^-$ . The efficiency, uncertainties on efficiencies, and efficiency ratios shown are calculated after summing over  $\theta$  bins. This ratio of efficiencies is much flatter than that for kaons.

### 5.3.2 Detector $A_{CP}$ Subtraction

One obvious way to correct for detector  $A_{CP}$  is to subtract the detector  $A_{CP}$  from the measured  $A_{CP}$ . It is actually not obvious that this kind of correction will do what we want. Are the  $A_{CP}$  really additive? What exactly do we mean by detector  $A_{CP}$  after all? In this section, we will show that the obvious method works to a very good approximation; subtracting the detector  $A_{CP}$  from the measured  $A_{CP}$  really gives us the real  $A_{CP}$ .

Let us start with our notation. We need to distinguish between real  $A_{CP}$  and measured  $A_{CP}$ .

$$\text{Real } A_{CP} = \frac{N_b - N_{\bar{b}}}{N_b + N_{\bar{b}}}. \quad (5.15)$$

Here we use uppercase  $N$  for the number of real  $b$  events without efficiency penalty. This is ideally what we want to measure.

Similarly, we define

$$\text{Measured } A_{CP} = \frac{n_b - n_{\bar{b}}}{n_b + n_{\bar{b}}}. \quad (5.16)$$

Here we use lowercase  $n$  to represent the number of detected event for each flavor. This represents what our  $A_{CP}$  extraction procedure gives us.

$n$  and  $N$  are related by a simple efficiency relation. We want to emphasize that  $b$  and  $\bar{b}$  may have slightly different efficiency. Our choice of efficiency relation is the following:

$$n_b = \epsilon \times (1 + \delta)N_b, \quad (5.17)$$

$$n_{\bar{b}} = \epsilon \times (1 - \delta)N_{\bar{b}}, \quad (5.18)$$

where  $\epsilon$  represents a central value for the efficiency and  $\delta$  models the difference of the two efficiencies. The reason for this choice will become clear later on.

Substituting Equation 5.17 and 5.18 into 5.16 and canceling  $\epsilon$ , we have

$$\text{Measured } A_{CP} = \frac{(1 + \delta)N_b - (1 - \delta)N_{\bar{b}}}{(1 + \delta)N_b + (1 - \delta)N_{\bar{b}}}. \quad (5.19)$$

Rearranging this expression, we have

$$\text{Measured } A_{CP} = \frac{N_b - N_{\bar{b}} + \delta(N_b + N_{\bar{b}})}{N_b + N_{\bar{b}} + \delta(N_b - N_{\bar{b}})}. \quad (5.20)$$

One thing to notice is that Equation 5.20 has a very nice limit. In the limit where  $N_b = N_{\bar{b}}$ , the expression becomes

$$\lim_{N_b=N_{\bar{b}}} \text{Measured } A_{CP} = \delta. \quad (5.21)$$

The limit in Equation 5.21 defines  $\delta$  as what we would obtain as the measured  $A_{CP}$  if there is no  $A_{CP}$  in the underlying physics; hence the name detector  $A_{CP}$ .

Let us continue with our analysis and substitute

$$X = N_b - N_{\bar{b}} \quad (5.22)$$

$$Y = N_b + N_{\bar{b}} \quad (5.23)$$

into Equation 5.20. We have

$$\text{Measured } A_{CP} = \frac{X + \delta Y}{Y + \delta X}. \quad (5.24)$$

A Taylor expansion of the above expression around  $\delta = 0$  to the first order in  $\delta$  is

6

$$\text{Measured } A_{CP} = \frac{X}{Y} + \delta \left( 1 - \frac{X^2}{Y^2} \right) + O(\delta^2) \quad (5.25)$$

$$\approx \text{Real } A_{CP} + \delta. \quad (5.26)$$

We obtain the second expression by using the fact that that  $X^2/Y^2 \ll 1$ . Subtracting  $A_{CP}$  from both sides gives us the obvious answer on how to correct for detector  $A_{CP}$

---

<sup>6</sup>In  $\delta^2$  term there is actually a  $\delta^2 X/Y$  term that is comparable to  $\delta X^2/Y^2$  term. However, in our case, both terms are much smaller than  $\delta$ .

to a very good approximation: just subtract it off from measured  $A_{CP}$ ,

$$\text{Real } A_{CP} = \text{Measured } A_{CP} - \delta. \quad (5.27)$$

### 5.3.3 Measurement of Detector $A_{CP}$

It was shown in the previous section that detector  $A_{CP}$  ( $\delta$ ) is the measured  $A_{CP}$  when the underlying physics has no flavor asymmetry. Our sideband ( $m_{\text{ES}} < 5.27 \text{ GeV}$ ), which consists mostly of  $e^+e^- \rightarrow$  light quark pair, is a very good candidate for such a system. However, it is not the perfect candidate. As shown in Figure 5.11, the detector asymmetry clearly depends on the  $K$  momentum distribution but the sideband  $K$  momentum and peaking region one are slightly different (see Figure 5.13). Correcting  $\delta$  for the  $K$  momentum difference is difficult, since it requires precise knowledge of the efficiency in each  $K$  momentum bin. Fortunately, a bound on such a correction can be calculated. So, our plan here is to use sideband  $A_{CP}$  as the central value for detector  $A_{CP}$  along with its statistical uncertainty. The bound on the correction will be used as the systematic uncertainty associated with detector  $A_{CP}$ .

#### 5.3.3.1 Sideband $A_{CP}$

Let us start with the simpler part: the central value. We define our sideband data as  $m_{\text{ES}} < 5.27 \text{ GeV}$ . We count the number of events in the sideband region for each flavor and calculate the sideband detector  $A_{CP}$  using

$$\delta_{\text{side}} = \frac{n_b^{\text{side}} - n_{\bar{b}}^{\text{side}}}{n_b^{\text{side}} + n_{\bar{b}}^{\text{side}}}, \quad (5.28)$$

where the statistical error on  $A_{CP}$  is calculated using the following formula:

$$\sigma_{\delta_{\text{side}}} = \frac{1}{\sqrt{n_b^{\text{side}} + n_{\bar{b}}^{\text{side}}}}. \quad (5.29)$$

The results are summarized in Table 5.7.

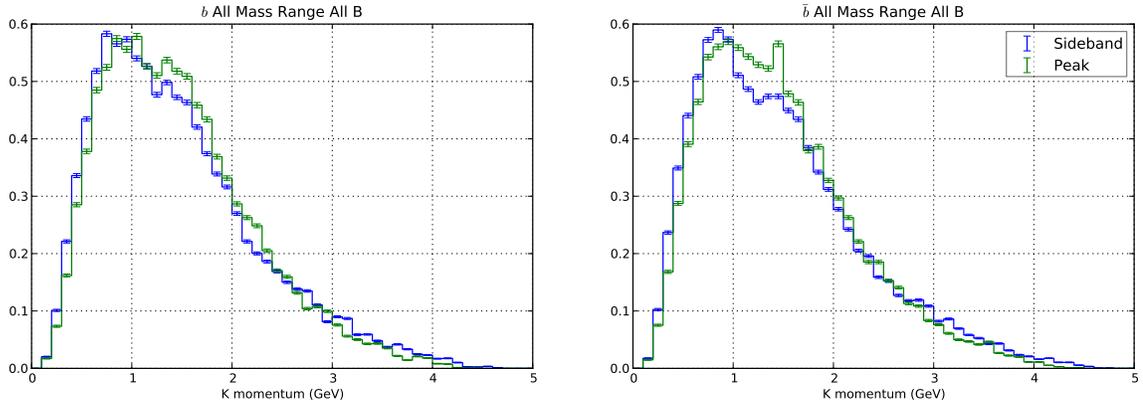


Figure 5.13: Normalized charged  $K$  momentum distribution for the full mass range for  $b$  (left) and  $\bar{b}$  (right) using the all  $B$  sample. The distribution of the sideband region is shown in blue while the distribution of the peaking region is shown in green. The distribution shown is, of course, after the efficiency penalty. But, since the distributions after selection are different, we can infer that the distributions before selection is different. The same plot for other  $X_s$  mass range and other  $B$  sample can be found in Appendix I.

Table 5.7: Sideband  $A_{CP}$  ( $\delta_{\text{side}}$ ) and associated statistical uncertainty for each  $X_s$  mass region and  $B$  type.

$X_s$ Mass Range	$B$ type	$n_b^{\text{side}}$	$n_{\bar{b}}^{\text{side}}$	$\delta_{\text{side}} \pm \text{stat.}$	Systematic
$K^*$	all $B$	4959	5130	$-1.69\% \pm 1.00\%$	0.51%
$K^*$	charged $B$	2376	2460	$-1.74\% \pm 1.43\%$	0.51%
$K^*$	neutral $B$	2538	2670	$-1.66\% \pm 1.37\%$	0.51%
Full	all $B$	20358	20935	$-1.40\% \pm 0.49\%$	0.51%
Full	charged $B$	10861	11101	$-1.09\% \pm 0.67\%$	0.51%
Full	neutral $B$	2538	2670	$-1.74\% \pm 0.72\%$	0.51%

### 5.3.3.2 Systematic Uncertainty Associated with Detector $A_{CP}$

The sideband detector  $A_{CP}(\delta_{\text{side}})$  is a good representation of the detector  $A_{CP}$  for the peaking region ( $\delta$ ). But, as we mentioned earlier, our sideband  $K$  momentum distribution is slightly different from the peak region one (see Figure 5.13). Calculating the exact difference between the two detector  $A_{CP}$ s due to this difference can be difficult since it requires knowledge of the efficiency for each  $K$  momentum bin. Fortunately, the bound on the difference of the detector  $A_{CP}$  of the two regions can be calculated.

The calculation we will show in this section may look mathematically intensive but the idea is very simple. Let us go through the idea first. Of course, we do not know the extent for which the true (before efficiency correction)  $K$  momentum for the sideband region differs from that for the peaking region<sup>7</sup>. Fortunately, the total effect on the detector  $A_{CP}$  is bounded by the variation of the  $K^+$  and  $K^-$  efficiency ratio in each  $K$  momentum bin, which we can find from the  $D^* \rightarrow D\pi$  sample shown in Figure 5.11. For example, if  $K^+$  and  $K^-$  efficiency difference were to have no  $K$  momentum dependence, the sideband detector  $A_{CP}$  would be the same as the peaking region detector  $A_{CP}$ . We will find in the end that the variation in  $\delta$  due to the difference in  $K$  momentum is half the variation of the ratio of  $K^-$  efficiency and  $K^+$  efficiency. From this we can calculate the corresponding systematic associated with the difference between the sideband  $A_{CP}$  and the peaking region  $A_{CP}$ .

Before we go on with our analysis, it should be mentioned that the most obvious way to do the correction is wrong. It is very tempting to calculate the correction by reweighing the events in sideband region according to the  $K$  momentum such that the  $K$  momentum distribution for the peaking region and the sideband region matches, then use the new reweighted sample to count  $n_b$  and  $n_{\bar{b}}$  and calculate the sideband  $A_{CP}$ . This, in fact, does not make any change to the calculated sideband  $A_{CP}$ . Reweighting preserves the yield in the sideband region. This means that the

---

<sup>7</sup>In principle, one could find efficiency for each  $K$  momentum bin and correct it back etc. But, this will introduce a large uncertainty from efficiencies since we only reconstruct from subset of final states the  $K$  distribution of the missing fraction is, by definition, unknown.

$n_b$  and  $n_{\bar{b}}$  stay the same after reweighting resulting in the same  $A_{CP}$  before and after the reweighting. In short, the reweighting does not do anything.

First, we consider a hypothetical situation where there is no  $A_{CP}$  in the underlying physics of  $b \rightarrow s\gamma$ . This translates to  $N_b = N_{\bar{b}}$ , where  $N$  denotes the number of signal event of each flavor *before* the efficiency penalty. In such a situation, we have

$$\delta = \frac{\nu_b - \nu_{\bar{b}}}{\nu_b + \nu_{\bar{b}}}. \quad (5.30)$$

Here we use  $\nu_b$  and  $\nu_{\bar{b}}$  instead of  $n_b$  and  $n_{\bar{b}}$  to emphasize that this equation is true only in the hypothetical no- $A_{CP}$  situation.

Let us understand  $\nu_b$  and  $\nu_{\bar{b}}$  a little better. We can write  $\nu_b$  and  $\nu_{\bar{b}}$  in terms of a sum of products of efficiencies and the PDF for signal events in all of its variables. Assuming the effect of the correlation of  $K$  momentum and other variables is small, we have

$$\nu_f = \sum_{p_K} \sum_{x_1} \dots \sum_{x_n} \epsilon_f^{\text{others}}(x_1, \dots, x_n) \epsilon_f^{p_K}(p_K) p_f^{\text{others}}(x_1, \dots, x_n) p_f^{p_K}(p_K) N_f, \quad (5.31)$$

where  $f \in \{b, \bar{b}\}$ ,  $\epsilon$  represents the efficiency function, which depends on the input variables,  $p$  represents the true (before efficiency correction) probability distribution function of signal event and  $p_K$  is the kaon momentum. We separate out  $p_K$  on purpose, since we know that detector asymmetry originates from asymmetry in  $K^+/K^-$  efficiency.

Rearranging the previous equation, we have

$$\nu_f = \underbrace{\left( \sum_{x_1} \dots \sum_{x_n} \epsilon_f^{\text{others}}(x_1, \dots, x_n) p_f^{\text{others}}(x_1, \dots, x_n) \right)}_{G_f^{\text{others}}} \left( \sum_{p_K} \epsilon_f^{p_K}(p_K) p_f^{p_K}(p_K) \right) N_f \quad (5.32)$$

$$= G_f^{\text{others}} N_f \sum_{p_K} \epsilon_f^{p_K}(p_K) p_f^{p_K}(p_K). \quad (5.33)$$

Substituting the above expression into 5.30 we have

$$\delta = \frac{G_b^{\text{others}} N_b \sum_{p_K} \epsilon_b^{p_K}(p_K) p_b^{p_K}(p_K) - G_{\bar{b}}^{\text{others}} N_{\bar{b}} \sum_{p_K} \epsilon_{\bar{b}}^{p_K}(p_K) p_{\bar{b}}^{p_K}(p_K)}{G_b^{\text{others}} N_b \sum_{p_K} \epsilon_b^{p_K}(p_K) p_b^{p_K}(p_K) + G_{\bar{b}}^{\text{others}} N_{\bar{b}} \sum_{p_K} \epsilon_{\bar{b}}^{p_K}(p_K) p_{\bar{b}}^{p_K}(p_K)} \quad (5.34)$$

Let us impose two assumptions

1. Efficiency functions for all variables except  $p_K$  are the same for both flavors.

This means

$$\epsilon_b^{\text{others}} = \epsilon_{\bar{b}}^{\text{others}}. \quad (5.35)$$

2. The underlying probability distribution for all variables including  $p_K$  are the same for both flavors. This means

$$p_b^{\text{others}} = p_{\bar{b}}^{\text{others}} \quad (5.36)$$

$$p_b^{p_K} = p_{\bar{b}}^{p_K} = p^{p_K}. \quad (5.37)$$

These two assumptions imply

$$G_b^{\text{others}} = G_{\bar{b}}^{\text{others}}. \quad (5.38)$$

Using these two assumptions, Equation 5.34 becomes

$$\delta = \frac{\sum_{p_K} \epsilon_b^{p_K}(p_K) p_b^{p_K}(p_K) - \sum_{p_K} \epsilon_{\bar{b}}^{p_K}(p_K) p_{\bar{b}}^{p_K}(p_K)}{\sum_{p_K} \epsilon_b^{p_K}(p_K) p_b^{p_K}(p_K) + \sum_{p_K} \epsilon_{\bar{b}}^{p_K}(p_K) p_{\bar{b}}^{p_K}(p_K)}. \quad (5.39)$$

Our  $b$  sample consists mostly of  $K^-$  and our  $\bar{b}$  sample consist mostly of  $K^+$ . Thus,

$$\epsilon_{\bar{b}}^{p_K}(p_K) \propto \epsilon_{K^+}(p_K) = F(p_K) \epsilon_{K^+}(p_K) \quad (5.40)$$

$$\epsilon_b^{p_K}(p_K) \propto \epsilon_{K^-}(p_K) = F(p_K) \epsilon_{K^-}(p_K). \quad (5.41)$$

This proportionality assumption does not make us lose the generality due to the fact that we also have modes that do not contain any charged  $K$ s. We can think of this case as a special value of  $p_K$ . The proportional constant  $F(p_K)$  is assumed to be the

same for both flavors, since we factor out the part that depends on flavor.

Also, let us write

$$\epsilon_{K^+}(p_K) = (1 + \lambda(p_K))\epsilon_{K^-}(p_K). \quad (5.42)$$

We know from Figure 5.11 that  $\lambda(p_K) \in [0\%, 2.5\%] \forall p_K$ .

Substituting Eq. 5.40 and 5.41 into Eq. 5.39, we have

$$\delta = \frac{\sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)p^{p_K}(p_K) - \sum_{p_K} F(p_K)\epsilon_{K^+}(p_K)p^{p_K}(p_K)}{\sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)p^{p_K}(p_K) + \sum_{p_K} F(p_K)\epsilon_{K^+}(p_K)p^{p_K}(p_K)} \quad (5.43)$$

$$= \frac{\sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)p^{p_K}(p_K) - \sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)(1 + \lambda(p_K))p^{p_K}(p_K)}{\sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)p^{p_K}(p_K) + \sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)(1 + \lambda(p_K))p^{p_K}(p_K)} \quad (5.44)$$

$$= \frac{-\sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)\lambda(p_K)p^{p_K}(p_K)}{2\sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)p^{p_K}(p_K) + \sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)\lambda(p_K)p^{p_K}(p_K)}. \quad (5.45)$$

Here is the key part. Let us look at

$$S = \sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)\lambda(p_K)p^{p_K}(p_K) \quad (5.46)$$

$F(p_K)$ ,  $\epsilon_{K^-}(p_K)$ , and  $p^{p_K}(p_K)$  inside the sum are positive for all  $p_K$ . Thus, the sum is bounded by the maximum and minimum value of  $\lambda(p_K)$ . Thus, we have

$$\lambda_{\min} \times T < S < \lambda_{\max} \times T, \quad (5.47)$$

where

$$T = \sum_{p_K} F(p_K)\epsilon_{K^-}(p_K)p^{p_K}(p_K). \quad (5.48)$$

Therefore we can write

$$S = \tilde{\lambda}T \quad (5.49)$$

where  $\tilde{\lambda} \in [0\%, 2.5\%]$ . One trivial but very important fact about this expression is that this is true for all momentum distributions including any mix of neutral  $K$  and charged  $K$ . In particular, this relation is true in both the peaking region and the side band region and for all types of  $B$  sample.

Plugging this into Equation 5.45 and using the fact that  $\tilde{\lambda} \ll 1$  we have

$$\delta = -\frac{\tilde{\lambda}}{2}. \quad (5.50)$$

This simple expression relates the detector  $A_{CP}$  to the  $K^+/K^-$  efficiency difference. In addition, it implies that the detector  $A_{CP}$  for both peaking region and the sideband region must lie within the range  $[-1.25\%, 0\%]$ . This agrees with our measured values from the sideband within statistical uncertainty, shown in Table 5.7. It also gives us a bound on the difference in the sideband region detector  $A_{CP}$  ( $\delta_{\text{side}}$ ) and the peaking region detector  $A_{CP}$  ( $\delta$ ) due to the difference in  $p_K$  momentum. The difference is at most  $\pm 1.25\%$ .

We have to be a bit careful about this  $\pm 1.25\%$  uncertainty on the difference, since it is not the usual gaussian-distributed uncertainty. This uncertainty originates from the difference of two flat distributions<sup>8</sup>. So, the shape is a triangle centered at 0 with a width of 1.25% on each side. In particular, the shape is described by

$$P(s) = \int Q(x) \times Q(y = s + x) dx = \begin{cases} 0 & \text{if } |s| > 0.0125 \\ \frac{1}{0.0125^2}(-|s| + 0.0125) & \text{if } |s| \leq 0.0125, \end{cases} \quad (5.51)$$

where  $Q$  is a flat distribution between  $[-1.25\%, 0\%]$ . The standard deviation for such a distribution is given by<sup>9</sup>  $1.25\%/\sqrt{6} = 0.51\%$ . Therefore, we will use 0.51% as our systematic uncertainty related to the detector  $A_{CP}$ . The results are summarized in Table 5.7.

---

<sup>8</sup>We assume a flat distribution for the detector  $A_{CP}$  that lies somewhere in  $[-1.25\%, 0\%]$ .

<sup>9</sup> $\langle x^2 \rangle = \int_{-d}^d \frac{x^2}{d^2}(-|x| + d) dx = 2 \int_0^d \frac{x^2}{d^2}(-x + d) dx = \frac{1}{6}d^2$

## 5.4 Results

We fit our data with the procedure described in Section 5.2 to find  $A_{CP}$  for each  $B$  sample. We subtract the detector  $A_{CP}$  described in Section 5.3. Statistical uncertainties from both fitted  $A_{CP}$  and detector  $A_{CP}$  are added in quadrature. Systematics uncertainties from peaking background dilution and detector  $A_{CP}$  are also added in quadrature. The results are summarized in Table 5.8. The  $m_{ES}$  fit for the full  $X_s$  mass range and the all  $B$  sample is shown in Figure 5.14. Fitted  $m_{ES}$  Distribution for others along with their shape parameters and correlations can be found in Appendix J.

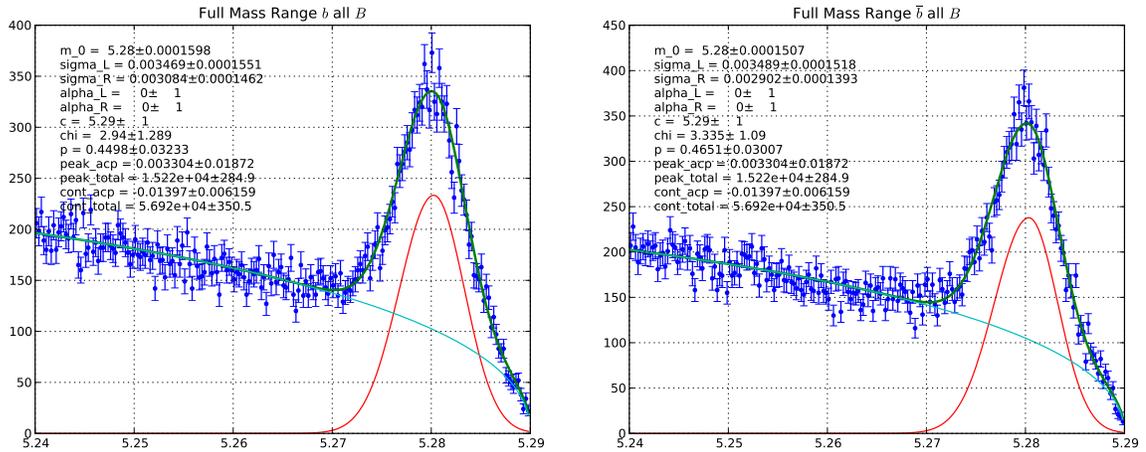


Figure 5.14: Fitted  $m_{ES}$  distribution for full mass range using all  $B$  sample.

Table 5.8: Summary of  $A_{CP}$  results.

Mass Range	$B$ Sample	Fitted $A_{CP}$	Dilution Sys.	Detector $A_{CP} \pm \text{stat} \pm \text{sys}$	$A_{CP} \pm \text{stat} \pm \text{sys}$
$K^*$	All $B$	$-1.44\% \pm 2.39\%$	0.38%	$-1.69\% \pm 1.00\% \pm 0.51\%$	$0.25\% \pm 2.59\% \pm 0.63\%$
$K^*$	Charged $B$	$2.97\% \pm 4.44\%$	0.52%	$-1.74\% \pm 1.43\% \pm 0.51\%$	$4.71\% \pm 4.66\% \pm 0.73\%$
$K^*$	Neutral $B$	$-3.75\% \pm 2.84\%$	0.30%	$-1.66\% \pm 1.37\% \pm 0.51\%$	$-2.09\% \pm 3.15\% \pm 0.59\%$
Full	All $B$	$0.33\% \pm 1.87\%$	0.88%	$-1.40\% \pm 0.49\% \pm 0.51\%$	$1.73\% \pm 1.93\% \pm 1.02\%$
Full	Charged $B$	$3.14\% \pm 2.86\%$	0.80%	$-1.09\% \pm 0.67\% \pm 0.51\%$	$4.23\% \pm 2.93\% \pm 0.95\%$
Full	Neutral $B$	$-2.48\% \pm 2.47\%$	0.97%	$-1.74\% \pm 0.72\% \pm 0.51\%$	$-0.74\% \pm 2.57\% \pm 1.10\%$

We found  $A_{CP}$  for  $K^*\gamma$  to be  $+0.25\% \pm 2.59\% \pm 0.63\%$ . This value is consistent with the previous  $BABAR$  dedicated  $K^*\gamma$   $A_{CP}$  analysis result [31] of  $-0.3\% \pm 1.7\% \pm 0.7\%$  and the world average value [1] of  $-0.3\% \pm 1.7\%$ . Our statistical uncertainty is larger

than the dedicated  $K^*\gamma$  analysis, due to the fact that we did not use the helicity angle of  $K^*$ 's daughters<sup>10</sup>. Figure J.1, J.2 and J.3 show fitted the  $m_{ES}$  distribution for the  $K^*$  mass range using all  $B$ , charged  $B$  and neutral  $B$  accordingly.

The fitted result from the full mass range is

$$A_{CP} = 1.73\% \pm 1.93\% \pm 1.02\%. \quad (5.52)$$

This value agrees with the value from previous *BABAR* analysis [12] of  $-1.1\% \pm 3.0\% \pm 1.4\%$ <sup>11</sup> and the world average [1] of  $-0.8\% \pm 2.9\%$  and the theoretical prediction of  $-0.6\% < A_{CP} < 3.2\%$ . The improvement on the statistical uncertainty is better than Poisson scaling compared to the previous analysis and the uncertainty is smaller than the current world average. The gain on the statistical uncertainty comes mostly from the new best candidate selection method and the improvement on the systematic uncertainty also comes as an added bonus from better peaking background rejection which allows us to use a simpler fitting method. The  $m_{ES}$  distribution fit for the full mass range is shown in Figure J.4, J.5 and J.6 for all  $B$ , neutral  $B$  and charged  $B$ , respectively.

Using the measured values of  $A_{CP}$  for charged  $B$  and neutral  $B$ , we found

$$\Delta A_{X_s\gamma} = 4.97\% \pm 3.90\% \pm 1.45\%, \quad (5.53)$$

where statistical and systematics uncertainty comes from the square root of the sum of the uncertainties from charged and neutral  $B$ .

We can turn this value into a confidence interval on  $\text{Im} \frac{C_{8g}}{C_{7\gamma}}$  by using Eq 14 from [11],

$$\Delta A_{X_s\gamma} = A_{X_s^-\gamma} - A_{X_s^0\gamma} \approx 12\% \times \frac{\tilde{A}_{78}}{100 \text{ MeV}} \text{Im} \frac{C_{8g}}{C_{7\gamma}}, \quad (5.54)$$

---

<sup>10</sup>Our original plan for the analysis is to measure the full mass range for all spin of  $X_s$ . Plus, helicity angle will not help us much for the final states with more than two particles.

<sup>11</sup>The uncertainty of the difference assuming this result is uncorrelated with our result is 3.9% and 1.2% assuming the result is fully correlated. The agreement uses the fact that our central lies within the uncertainty of the previous measurement.

where  $\tilde{\Lambda}_{78}$  is approximated to be

$$17 \text{ MeV} < \tilde{\Lambda}_{78} < 190 \text{ MeV},$$

to calculate 68% and 90% confidence interval for  $\text{Im} \frac{C_{8g}}{C_{7\gamma}}$  from  $\Delta A_{X_s\gamma}$ . We find

$$0.04 < \text{Im} \frac{C_{8g}}{C_{7\gamma}} < 4.48 \quad 68\% \text{ C.I.} \quad (5.55)$$

$$-1.64 < \text{Im} \frac{C_{8g}}{C_{7\gamma}} < 6.52 \quad 90\% \text{ C.I.} \quad (5.56)$$

To obtain the intervals above we assume the value of  $\tilde{\Lambda}_{78}$  from [11] to have a flat prior. For a given value of  $\text{Im} \frac{C_{8g}}{C_{7\gamma}}$ , we calculate the minimum  $\chi^2$  from all possible  $\tilde{\Lambda}_{78}$  values. Our 68% and 90% confidence intervals are obtained from the interval in which  $\min \chi^2 < 1$  and  $\min \chi^2 < 4$ , respectively. The shape of the minimum  $\chi^2$  is non-parabolic, stemming from the fact that we use the minimum  $\chi^2$  from all possible values of  $\tilde{\Lambda}_{78}$ . The plot of the minimum  $\chi^2$  is shown in Figure 5.15, and the plot of the confidence interval for a given value of  $\tilde{\Lambda}_{78}$  is shown in Figure 5.16. The full confidence interval is dominated by the range of possible  $\tilde{\Lambda}_{78}$  especially at low values of  $\tilde{\Lambda}_{78}$ . A better constraint at the lower value of  $\tilde{\Lambda}_{78}$  would greatly improve the confidence interval for  $\text{Im} \frac{C_{8g}}{C_{7\gamma}}$ .

We also use  $\Delta A_{X_s\gamma}$  to constrain the ratio of the magnitude of the Wilson coefficients  $r_8/r_7$  and the difference of the phase of the Wilson coefficients  $\theta_8 - \theta_7$  defined in [11]. We assume a flat prior on  $\tilde{\Lambda}_{78}$  with the range given by [11] and use  $\min \chi^2$  to calculate the confidence intervals. The result is shown in Figure 5.17. Since there is a strong dependence of the confidence interval on the value of  $\tilde{\Lambda}_{78}$ , we also provide the confidence interval for a given value of  $\tilde{\Lambda}_{78}$  in Figure 5.18.

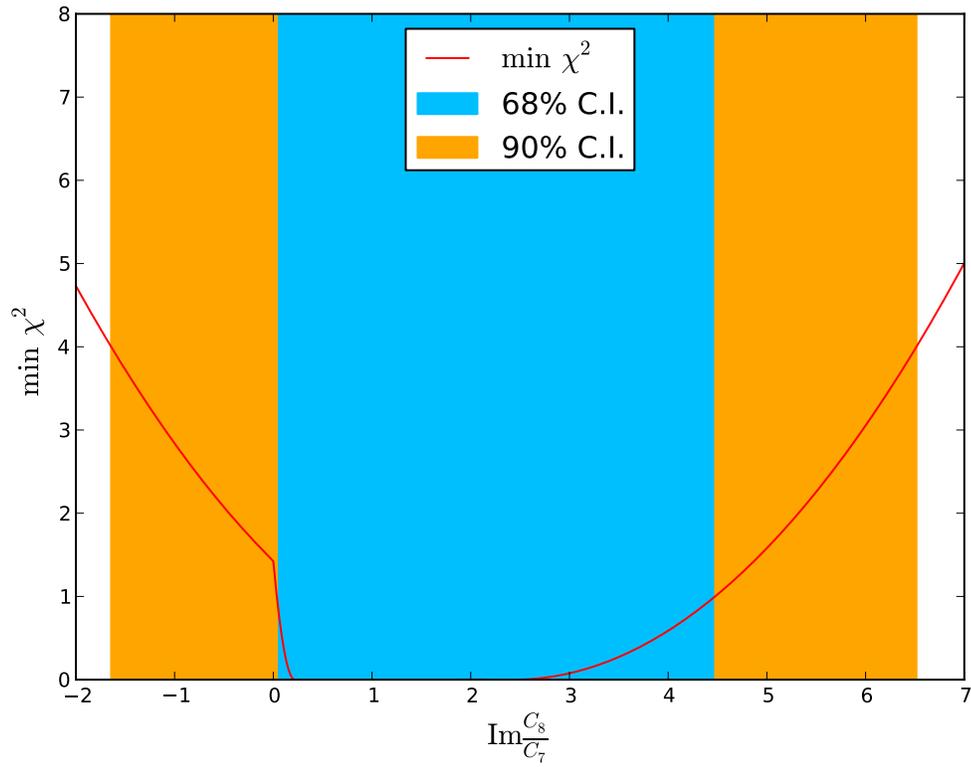


Figure 5.15: 68% and 90% confidence interval for  $\text{Im} \frac{C_8}{C_7}$  for possible values of  $\tilde{\Lambda}_{78}$ . The irregular shape comes from the fact that it is a product of a range and a number. Care must be taken when performing a global fit with other observables.

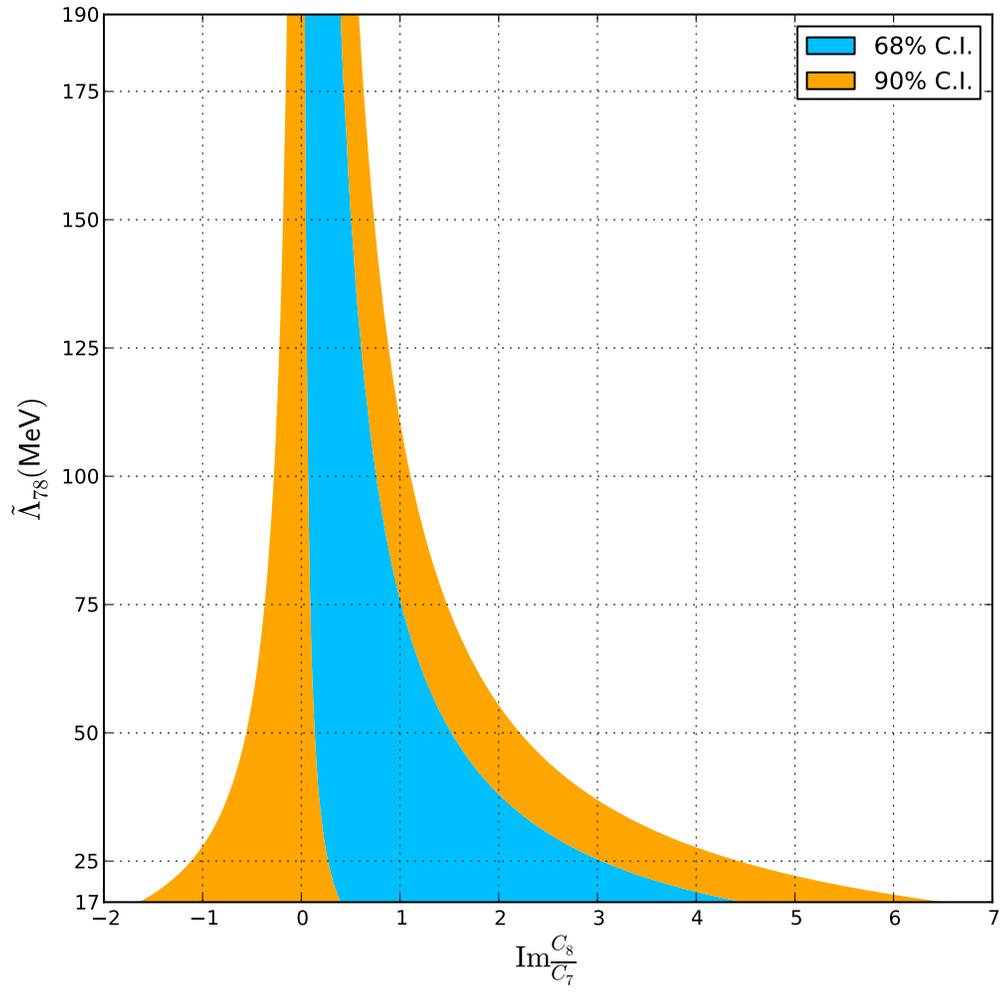


Figure 5.16: 68% and 90% confidence interval for  $\text{Im} \frac{C_8}{C_7}$  for the given value of  $\tilde{\Lambda}_{78}$ .

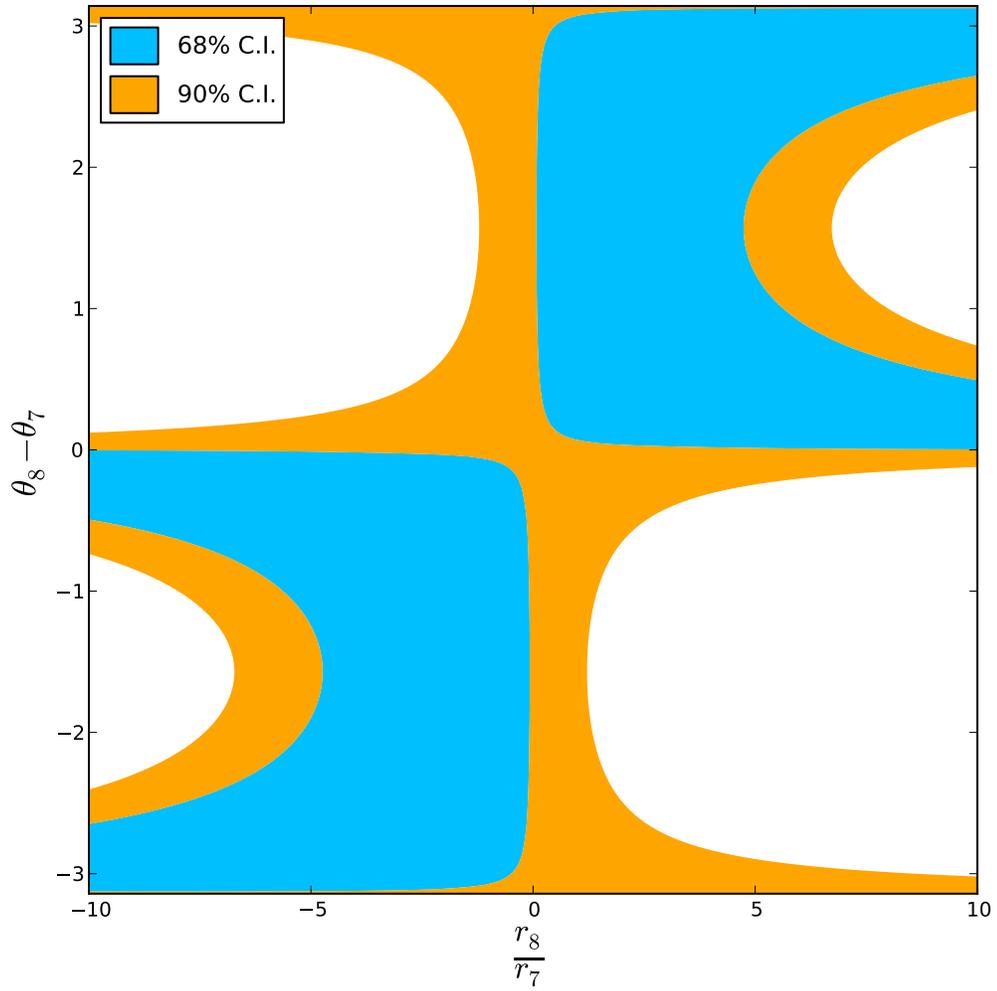


Figure 5.17: 68% and 90% confidence interval for  $\frac{r_8}{r_7}$  and  $\theta_8 - \theta_7$  from all possible value of  $\tilde{\Lambda}_{78}$ .

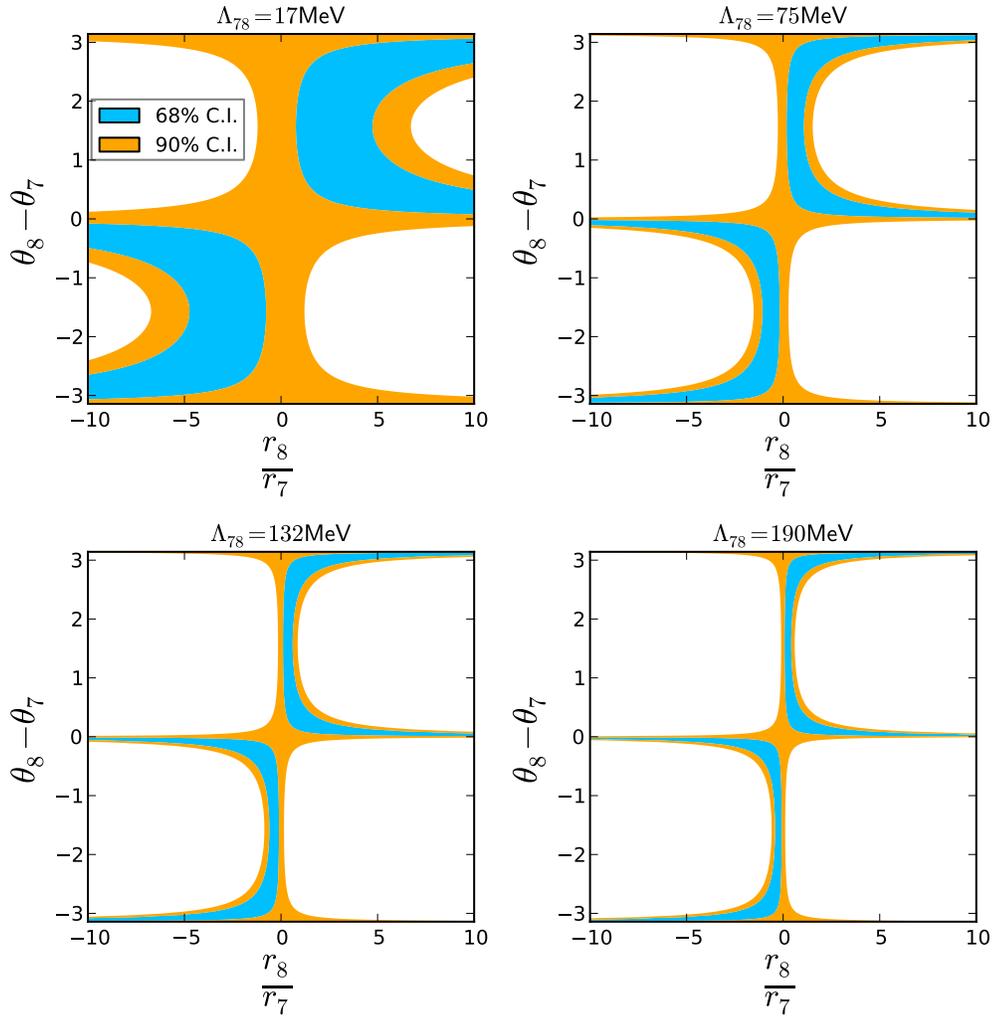


Figure 5.18: 68% and 90% confidence interval for  $\frac{r_8}{r_7}$  and  $\theta_8 - \theta_7$  for a given value of  $\tilde{\Lambda}_{78}$ .

# Bibliography

- [1] Particle Data Group, K. Nakamura *et al.*, “Review of particle physics,” *J.Phys.G* **G37** (2010) 075021.
- [2] A. L. Kagan and M. Neubert, “Direct CP violation in  $\rightarrow X_s$  gamma decays as a signature of new physics,” *Phys.Rev.* **D58** (1998) 094012, [arXiv:hep-ph/9803368](https://arxiv.org/abs/hep-ph/9803368) [hep-ph].
- [3] A. Sakharov, “Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe,” *Pisma Zh.Eksp.Teor.Fiz.* **5** (1967) 32–35.
- [4] N. Cabibbo, “Unitary Symmetry and Leptonic Decays,” *Phys.Rev.Lett.* **10** (1963) 531–533.
- [5] M. Kobayashi and T. Maskawa, “*cp*-violation in the renormalizable theory of weak interaction,” *Progress of Theoretical Physics* **49** no. 2, (1973) 652–657. <http://ptp.ipap.jp/link?PTP/49/652/>.
- [6] G. R. Farrar and M. Shaposhnikov, “Baryon asymmetry of the universe in the minimal Standard Model,” *Phys.Rev.Lett.* **70** (1993) 2833–2836, [arXiv:hep-ph/9305274](https://arxiv.org/abs/hep-ph/9305274) [hep-ph].
- [7] P. Huet and E. Sather, “Electroweak baryogenesis and standard model CP violation,” *Phys.Rev.* **D51** (1995) 379–394, [arXiv:hep-ph/9404302](https://arxiv.org/abs/hep-ph/9404302) [hep-ph].
- [8] P. Huet, “Electroweak baryogenesis and the standard model,” [arXiv:hep-ph/9406301](https://arxiv.org/abs/hep-ph/9406301) [hep-ph].

- [9] A. L. Kagan and M. Neubert, “Direct CP violation in  $B \rightarrow X(s)$  gamma decays as a signature of new physics,” *Phys.Rev.* **D58** (1998) 094012, [arXiv:hep-ph/9803368](#) [hep-ph].
- [10] T. Hurth, E. Lunghi, and W. Porod, “Untagged anti- $B \rightarrow X(s+d)$  gamma CP asymmetry as a probe for new physics,” *Nucl.Phys.* **B704** (2005) 56–74, [arXiv:hep-ph/0312260](#) [hep-ph].
- [11] M. Benzke, S. J. Lee, M. Neubert, and G. Paz, “Long-Distance Dominance of the CP Asymmetry in  $B \rightarrow X_{s,d} + \text{gamma}$  Decays,” *Phys.Rev.Lett.* **106** (2011) 141801, [arXiv:1012.3167](#) [hep-ph].
- [12] BABAR Collaboration, B. Aubert *et al.*, “A Measurement of CP Asymmetry in  $b \rightarrow s\gamma$  using a Sum of Exclusive Final States,” *Phys.Rev.Lett.* **101** (2008) 171804, [arXiv:0805.4796](#) [hep-ex].
- [13] BELLE Collaboration, S. Nishida *et al.*, “Measurement of the CP asymmetry in  $B \rightarrow X(s)$  gamma,” *Phys.Rev.Lett.* **93** (2004) 031803, [arXiv:hep-ex/0308038](#) [hep-ex].
- [14] CLEO Collaboration, T. Coan *et al.*, “Bounds on the CP asymmetry in  $b \rightarrow s$  gamma decays,” *Phys.Rev.Lett.* **86** (2001) 5661–5665, [arXiv:hep-ex/0010075](#) [hep-ex].
- [15] M. B. W. Aneesh V. Manohar, *Heavy Quark Physics*. Cambridge University Press, 2000.
- [16] A. J. Buras, “Flavor physics and CP violation,” [arXiv:hep-ph/0505175](#) [hep-ph].
- [17] A. L. Kagan and M. Neubert, “QCD anatomy of  $B \rightarrow X(s)$  gamma decays,” *Eur.Phys.J.* **C7** (1999) 5–27, [arXiv:hep-ph/9805303](#) [hep-ph].

- [18] Z. Ligeti, L. Randall, and M. B. Wise, “Comment on nonperturbative effects in anti- $B \rightarrow X(s) \gamma$ ,” *Phys.Lett.* **B402** (1997) 178–182, [arXiv:hep-ph/9702322](#) [hep-ph].
- [19] T. Hurth and M. Nakao, “Radiative and Electroweak Penguin Decays of B Mesons,” *Ann.Rev.Nucl.Part.Sci.* **60** (2010) 645–677, [arXiv:1005.1224](#) [hep-ph].
- [20] M. Benzke, S. J. Lee, M. Neubert, and G. Paz, “Long-Distance Dominance of the CP Asymmetry in  $B \rightarrow X_{s,d} + \gamma$  Decays,” *Phys.Rev.Lett.* **106** (2011) 141801, [arXiv:1012.3167](#) [hep-ph].
- [21] W. Altmannshofer, P. Paradisi, and D. M. Straub, “Model-Independent Constraints on New Physics in  $b \rightarrow s$  Transitions,” *JHEP* **1204** (2012) 008, [arXiv:1111.1257](#) [hep-ph].
- [22] W. Altmannshofer and D. M. Straub, “Cornering New Physics in  $b \rightarrow s$  Transitions,” *JHEP* **1208** (2012) 121, [arXiv:1206.0273](#) [hep-ph].
- [23] L. Wolfenstein and Y. L. Wu, “Cp violation in the decay  $b \rightarrow s \gamma$  in the two-higgs-doublet model,” *Phys. Rev. Lett.* **73** (Nov, 1994) 2809–2812. <http://link.aps.org/doi/10.1103/PhysRevLett.73.2809>.
- [24] C.-K. Chua, X.-G. He, and W.-S. Hou, “CP violating  $b \rightarrow s \gamma$  decay in supersymmetric models,” *Phys.Rev.* **D60** (1999) 014003, [arXiv:hep-ph/9808431](#) [hep-ph].
- [25] BABAR Collaboration, B. Aubert *et al.*, “The BaBar detector,” *Nucl.Instrum.Meth.* **A479** (2002) 1–116, [arXiv:hep-ex/0105044](#) [hep-ex].
- [26] T. G. Dietterich and G. Bakiri, “Solving Multiclass Learning Problems via Error-Correcting Output Codes,” *Journal of Artificial Intelligence Research* **2** (1995) 263–286. also available at <http://www.jair.org/media/105/live-105-1426-jair.pdf>.

- [27] L. Breiman, “Random forests,” *Mach. Learn.* **45** no. 1, (Oct., 2001) 5–32. <http://dx.doi.org/10.1023/A:1010933404324>.
- [28] L. Breiman, “Bagging predictors,” *Mach. Learn.* **24** no. 2, (Aug., 1996) 123–140. <http://dx.doi.org/10.1023/A:1018054314350>.
- [29] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Chapman & Hall, New York, NY, 1984.
- [30] F. James and M. Roos, “Minuit - a system for function minimization and analysis of the parameter errors and correlations,” *Computer Physics Communications* **10** no. 6, (1975) 343 – 367. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.158.9157&rep=rep1&type=pdf>.
- [31] BABAR Collaboration, B. Aubert *et al.*, “Measurement of Branching Fractions and CP and Isospin Asymmetries in  $B \rightarrow K^*(892)\gamma$  Decays,” *Phys.Rev.Lett.* **103** (2009) 211802, [arXiv:0906.2177](https://arxiv.org/abs/0906.2177) [hep-ex].
- [32] A. V. Telnov, “A new method of track-level dE/dx calibration for the BaBar Drift Chamber and Silicon Vertex Tracker.”. <http://www.slac.stanford.edu/babar-internal/BAD/doc/detail.html?docNum=1500>.
- [33] I. Narsky, “StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data,” *ArXiv Physics e-prints* (July, 2005) , [arXiv:physics/0507143](https://arxiv.org/abs/physics/0507143).
- [34] BABAR Collaboration, B. Aubert *et al.*, “Measurements of the  $B \rightarrow X_s\gamma$  branching fraction and photon spectrum from a sum of exclusive final states,” *Phys.Rev.* **D72** (2005) 052004, [arXiv:hep-ex/0508004](https://arxiv.org/abs/hep-ex/0508004) [hep-ex].
- [35] J. Luisier, “Measurements of B Meson Lifetime Ratios with the LHCb Detector,”. <http://lphe.epfl.ch/publications/theses/these.jl.pdf>.

# Appendix A

## Particle Identification Appendix

### A.1 PID Variables

We use the following 31 variables for our binary classifiers:

- Momentum;
- Charge;
- Polar angle  $\theta$ ;
- Azimuthal angle  $\phi$ ;
- SVT and DCH  $\frac{dE}{dx}$  pulls for all 4 particle types [32];
- DIRC likelihood for  $\pi$ ,  $K$ ,  $p$ ;
- 6 SVT\*DCH\*DRlikelihood ratios (KvsPi, KvsPro, KvsEle, ProvsPi, ProvsEle, PivsEle);
- Number of signal and background photons in the DIRC;
- Last layer with a hit in the DCH (out of 40);
- Number of layers hit in the SVT;
- EMC energy;

- EMC calorimeter energy divided by momentum (dominates in  $e$  and  $\pi$  separation)
- EMC lateral momentum Ratio of 1) to 2)
  - 1) sum of energies of all but the 2 most energetic crystals, weighted by the square of distance to the cluster center;
  - 2) sum of 1) and the energies of the 2 most energetic crystals, which are weighted by  $r^2$ .  $r$  is the length scale of a crystal, 5 cm;
- Number of crystals in the EMC cluster;
- EMC Zernike moments (2,0) and (4,2) ;
- EMC s1s9 (the ratio of the sums of the energies of the central crystal to the central 9 crystals surrounding the centroid);
- EMC s9s25 (the ratio of the sums of the energies of the central 9 crystals to the central 25 crystals surrounding the centroid.);
- EMC second moment in  $\theta$  and  $\phi$ ;
- EMC longitudinal shower depth.

## A.2 PID Variable Importance

This section shows variable importance for each classifier used in *BABAR* PID. They are calculated by summing weighted differences in figures of merit with and without creating binary decision split using that variable. The larger pie chart section indicates more importance.

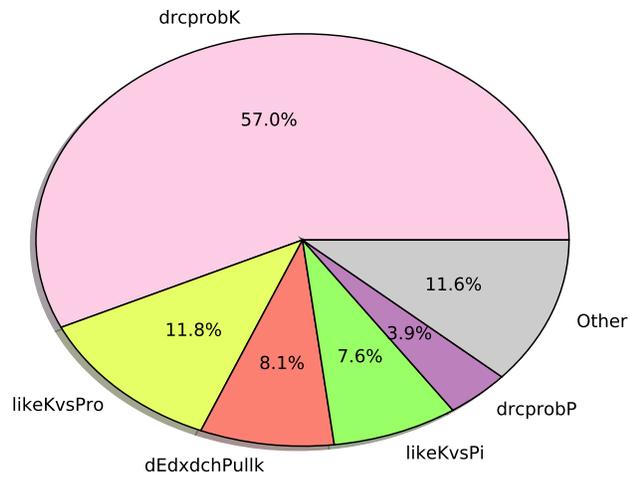


Figure A.1: Variable importance of  $K$  vs  $\pi$ ,  $p$ ,  $e$  classifier

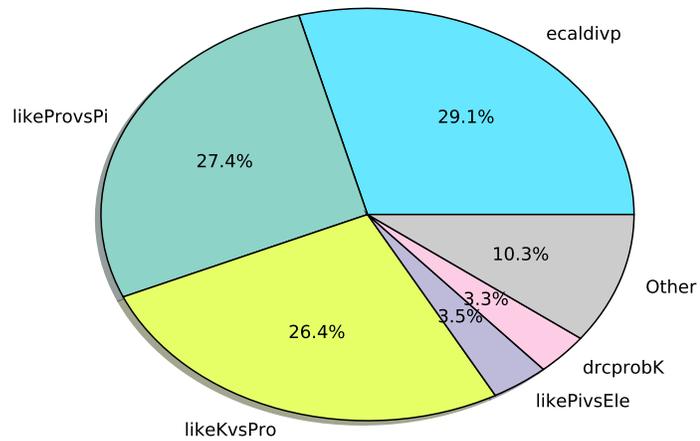


Figure A.2: Variable importance of  $K$ ,  $\pi$  vs  $p$ ,  $e$  classifier

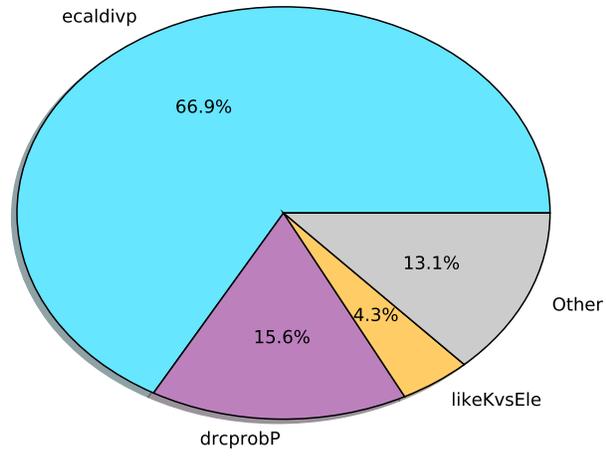


Figure A.3: Variable importance of  $K$ ,  $\pi$ ,  $p$  vs  $e$  classifier

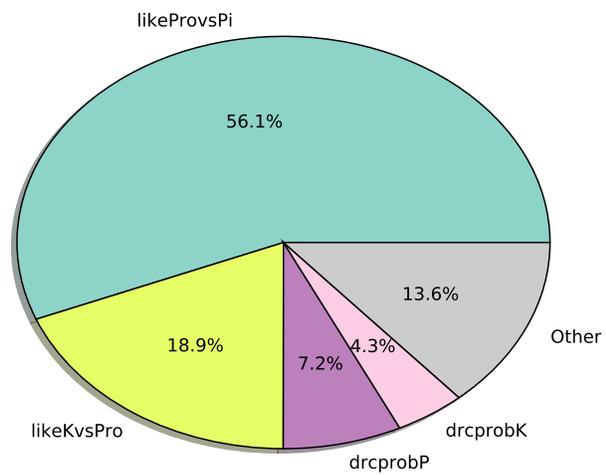


Figure A.4: Variable importance of  $K$ ,  $\pi$ ,  $e$  vs  $p$  classifier

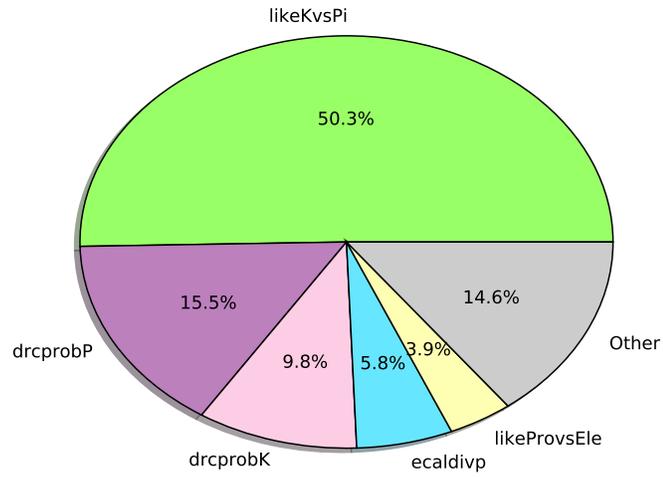


Figure A.5: Variable importance of  $K, p$  vs  $\pi, e$  classifier

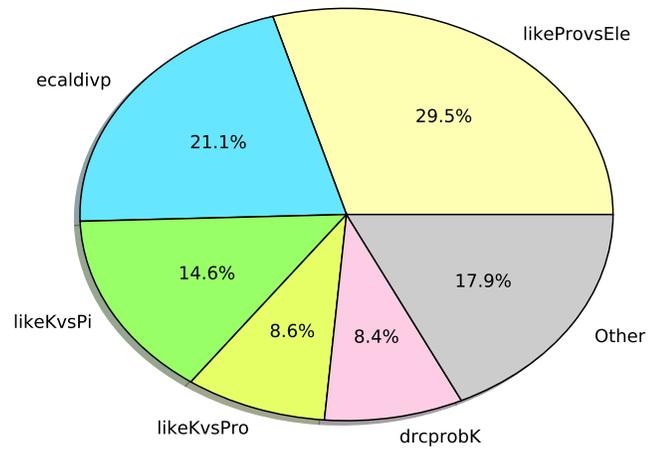


Figure A.6: Variable importance of  $K, e$  vs  $\pi, p$  classifier

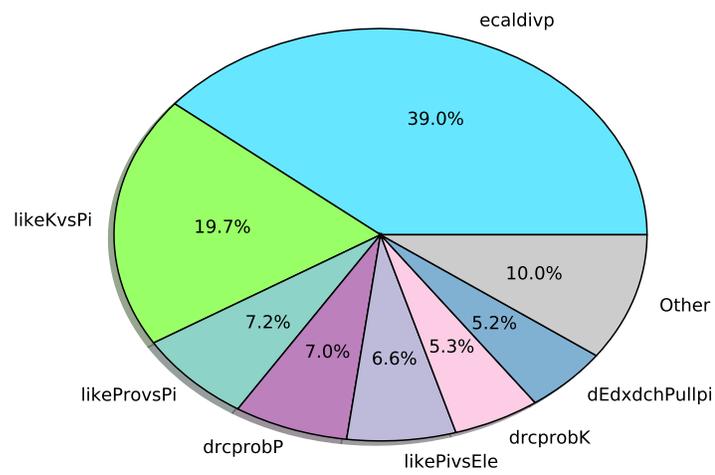


Figure A.7: Variable importance of  $K$ ,  $p$ ,  $e$  vs  $\pi$  classifier

## A.3 PID Performance and Comparison to Previous Generation

The comparisons of efficiencies and misidentification rates of likelihood-based selector and ECOC-based selector using 1vs1 indicator matrix and exhaustive indicator matrix of all tightness level are shown in sections A.3.1-A.3.4. The efficiencies are shown with circles and misidentification rates are represented by triangles. The magnification factors for misidentification rates for each set of plots are different. The color red, green and blue represents the efficiency and misidentification rates from exhaustive matrix, 1vs1 matrix and likelihood selector respectively.

## A.3.1 Kaon

### A.3.1.1 Super Loose Kaon Selector

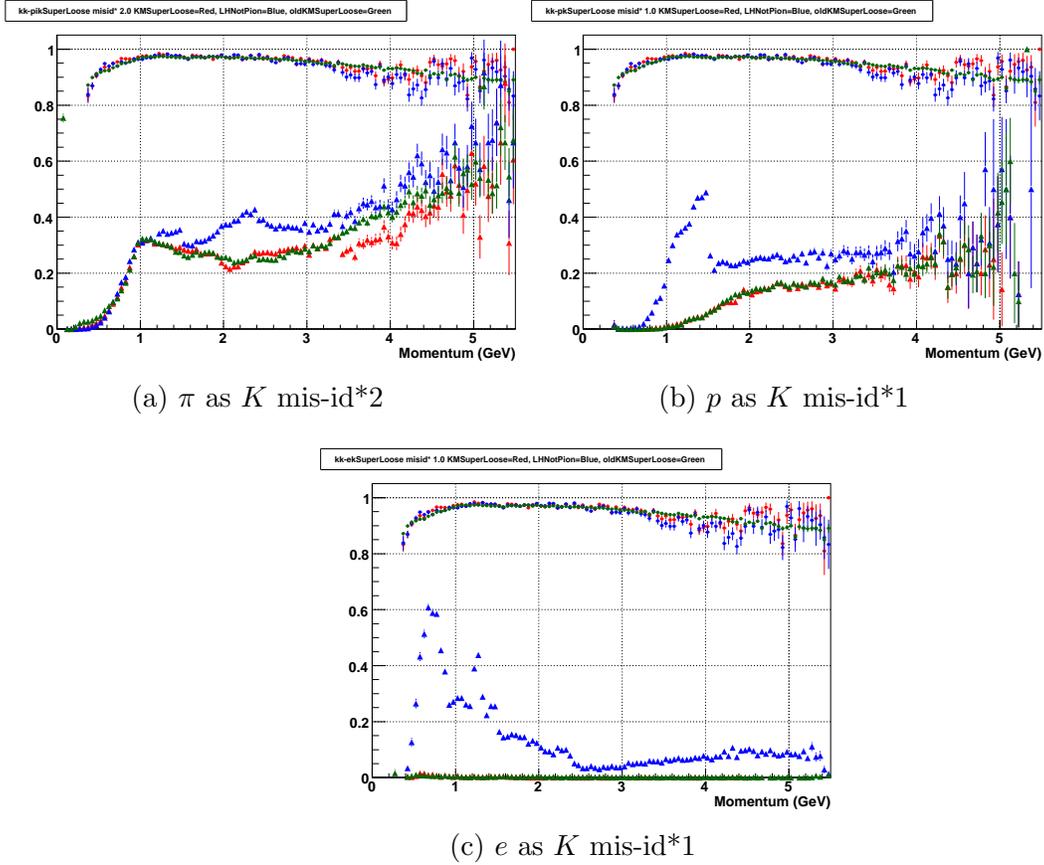


Figure A.8: Kaon selection efficiencies and misidentification rates for Super Loose kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.1.2 Very Loose Kaon Selector

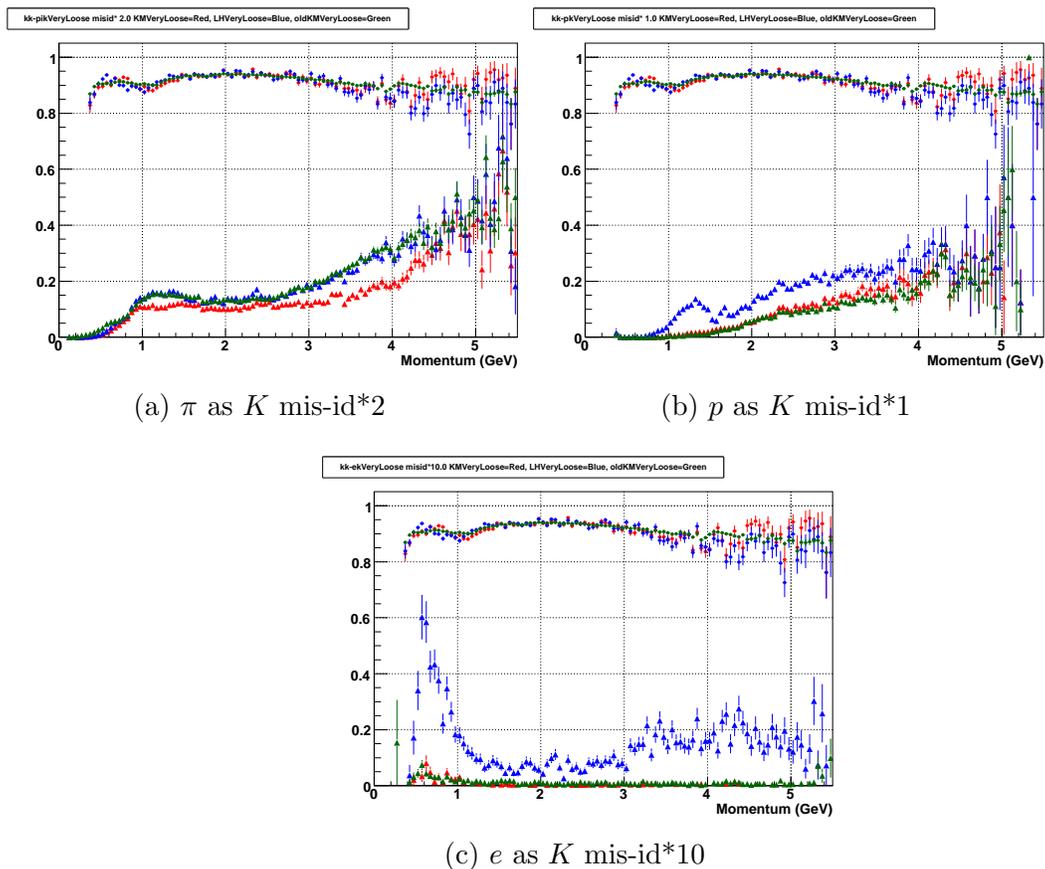


Figure A.9: Kaon selection efficiencies and misidentification rates for Very Loose kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.1.3 Loose Kaon Selector

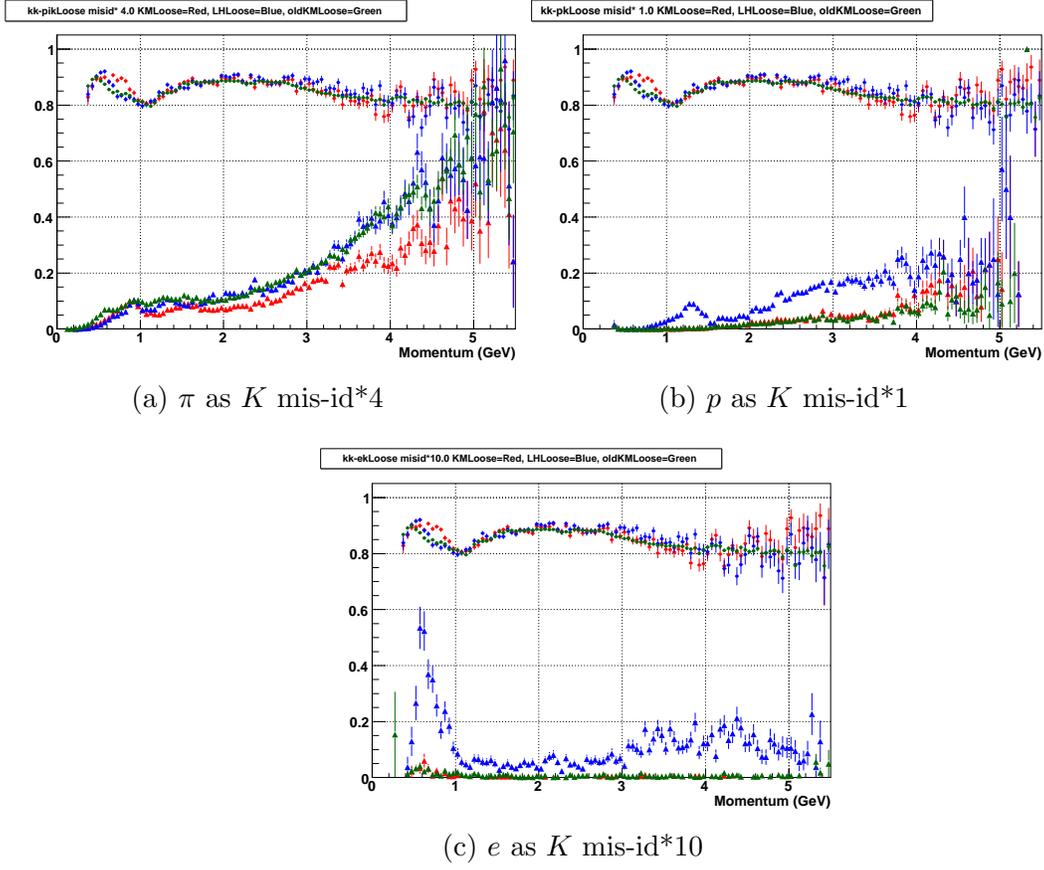


Figure A.10: Kaon selection efficiencies and misidentification rates for Loose kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.1.4 Tight Kaon Selector

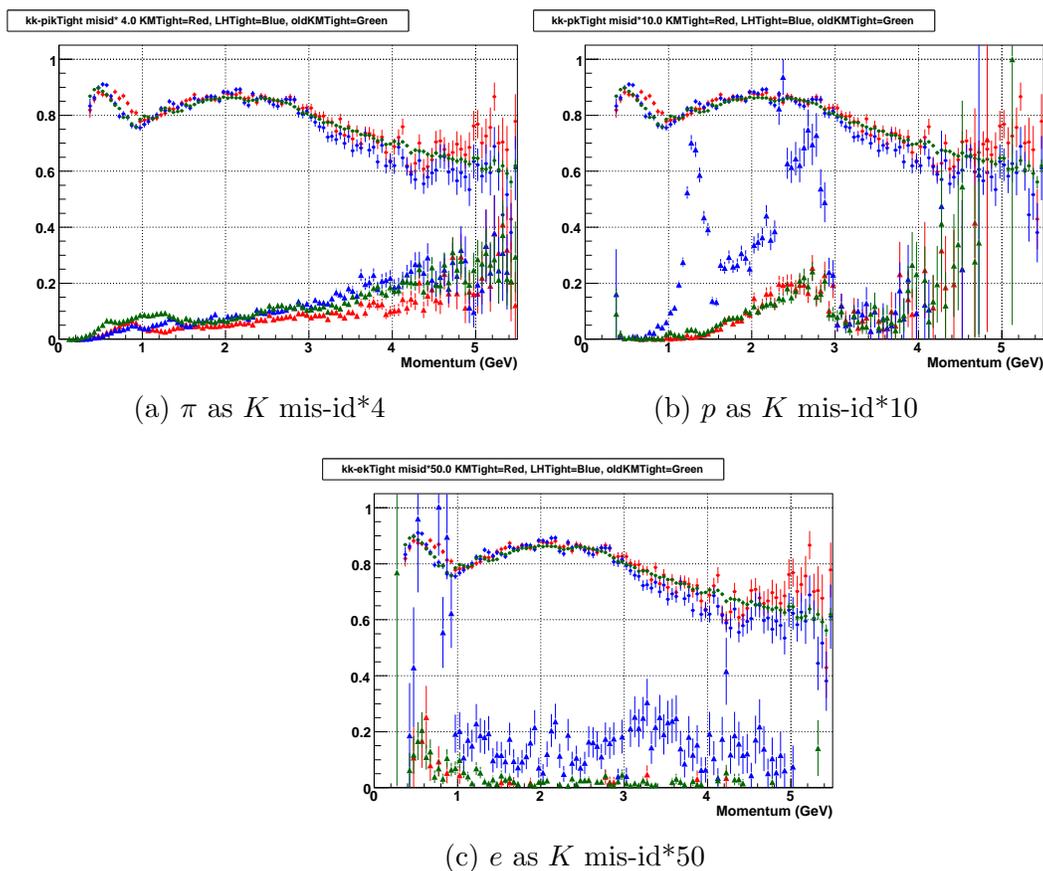


Figure A.11: Kaon selection efficiencies and misidentification rates for Tight kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.1.5 Very Tight Kaon Selector

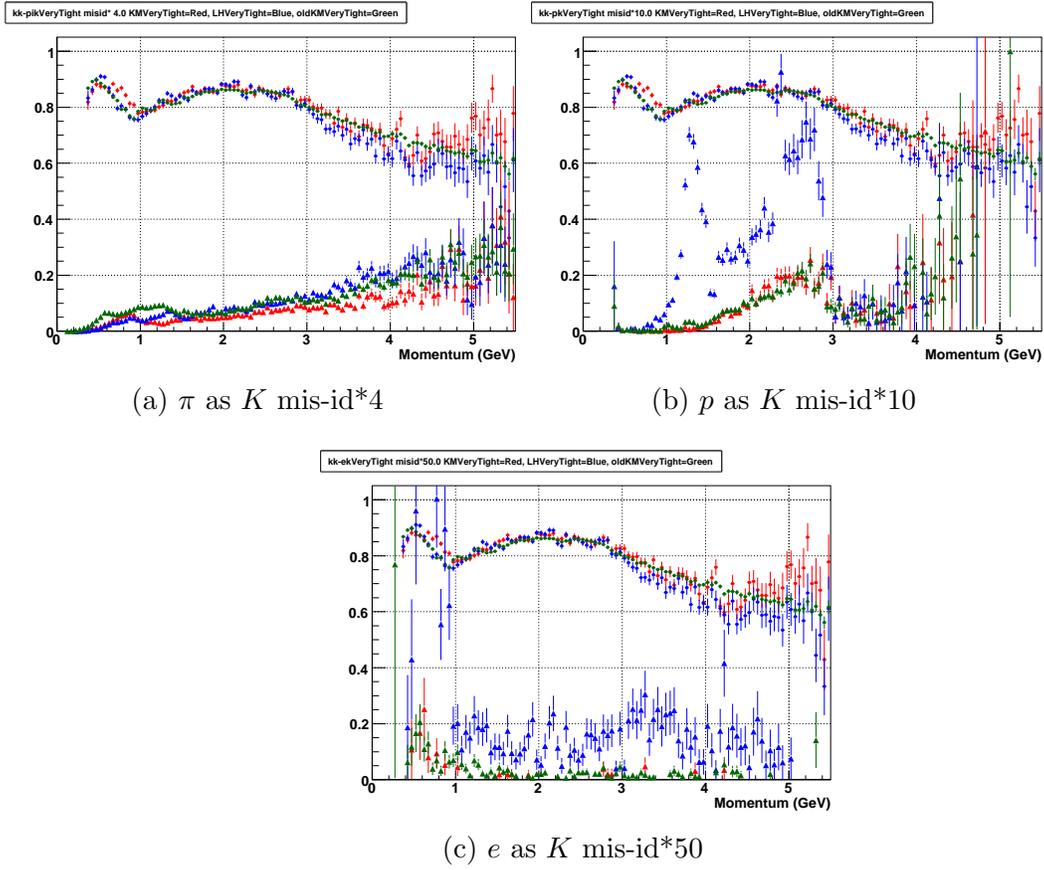


Figure A.12: Kaon selection efficiencies and misidentification rates for Very Tight kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.1.6 Super Tight Kaon Selector

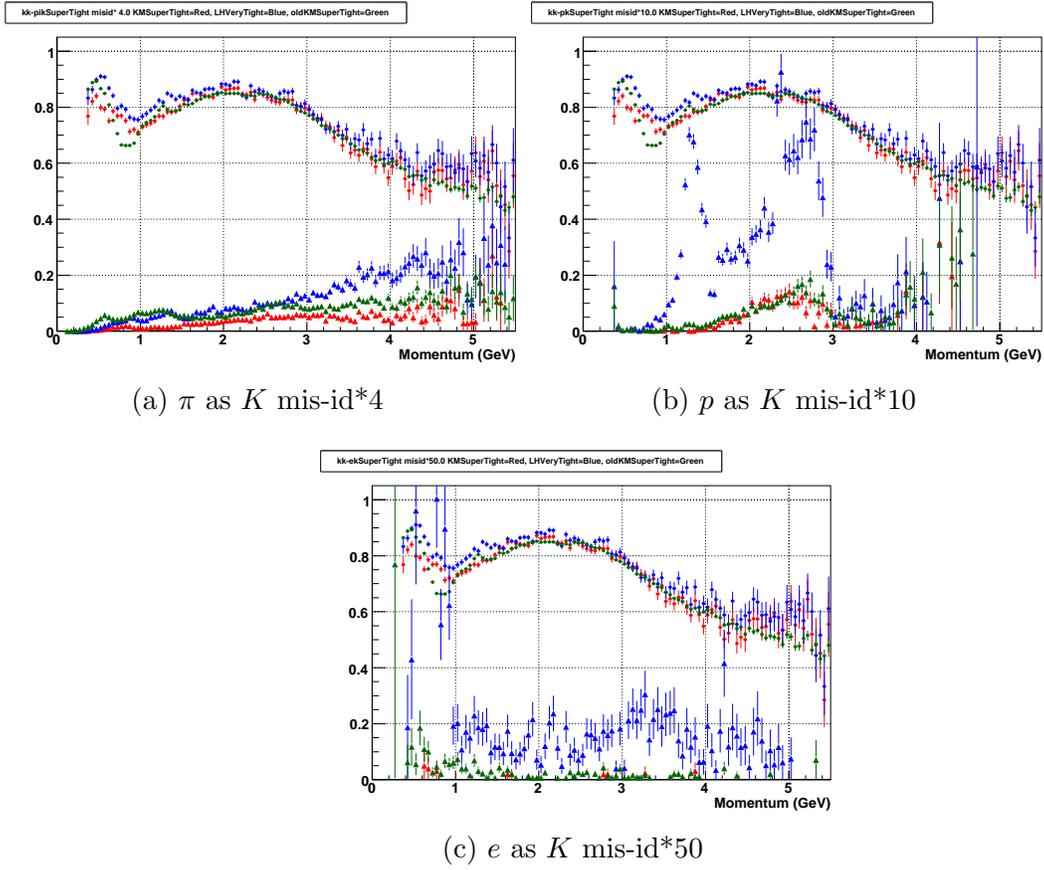
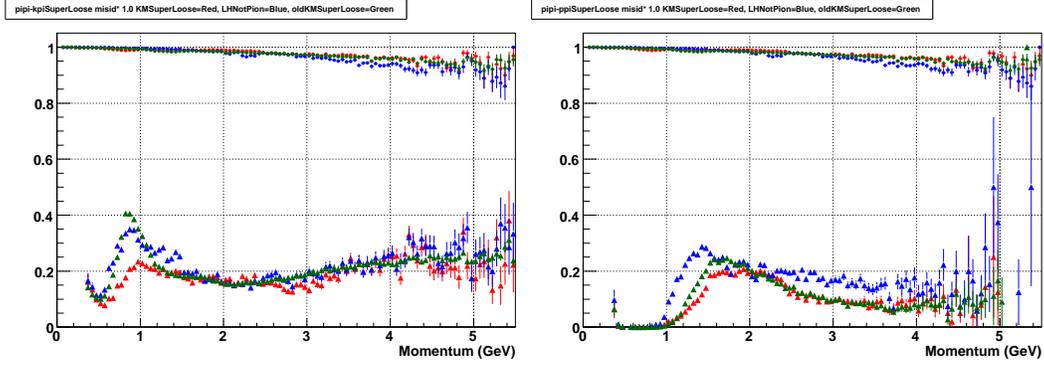


Figure A.13: Kaon selection efficiencies and misidentification rates for Super Tight kaon selector as  $K$  of  $\pi$  (a),  $p$  (b), and  $e$  (c). Kaon selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

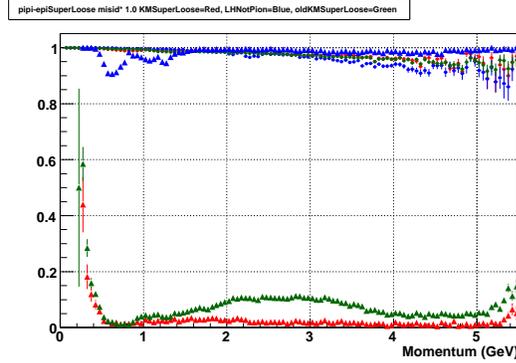
## A.3.2 Pion

### A.3.2.1 Super Loose Pion Selector



(a)  $K$  as  $\pi$  mis-id\*1

(b)  $p$  as  $\pi$  mis-id\*1



(c)  $e$  as  $\pi$  mis-id\*1

Figure A.14: Pion selection efficiencies and misidentification rates for Super Loose pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.2.2 Very Loose Pion Selector

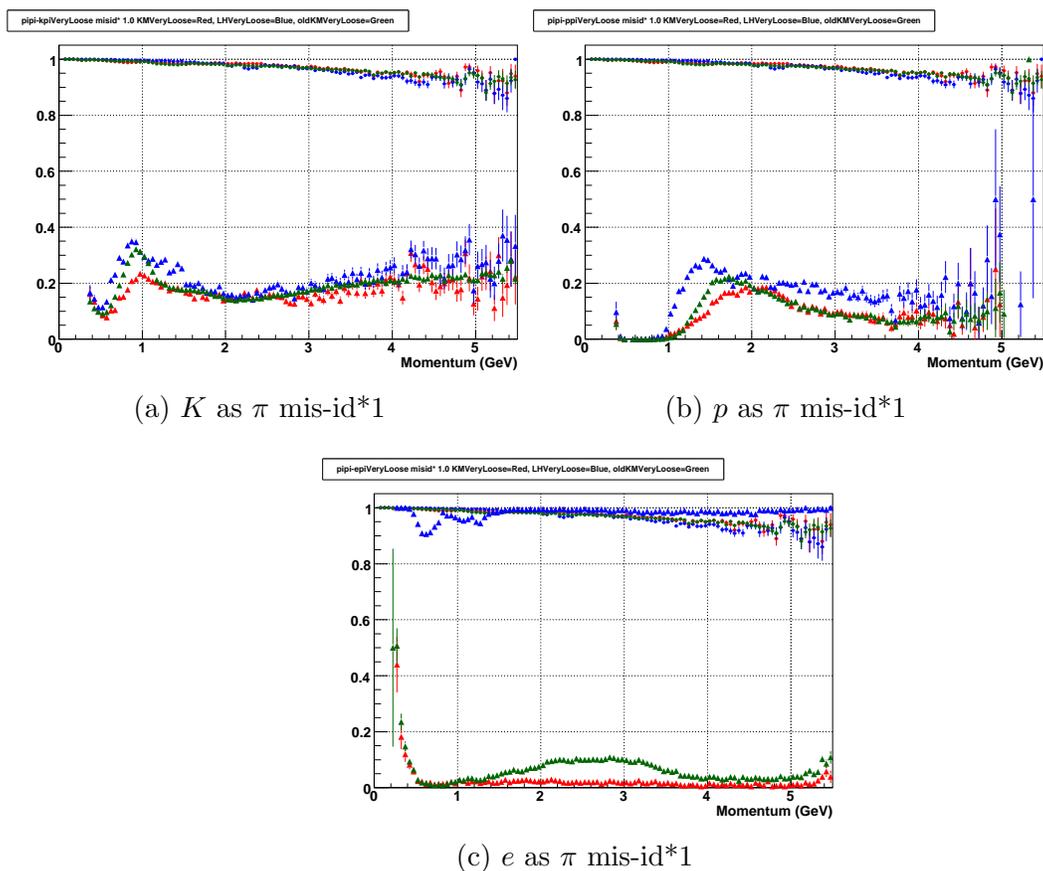


Figure A.15: Pion selection efficiencies and misidentification rates for Very Loose pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.2.3 Loose Pion Selector

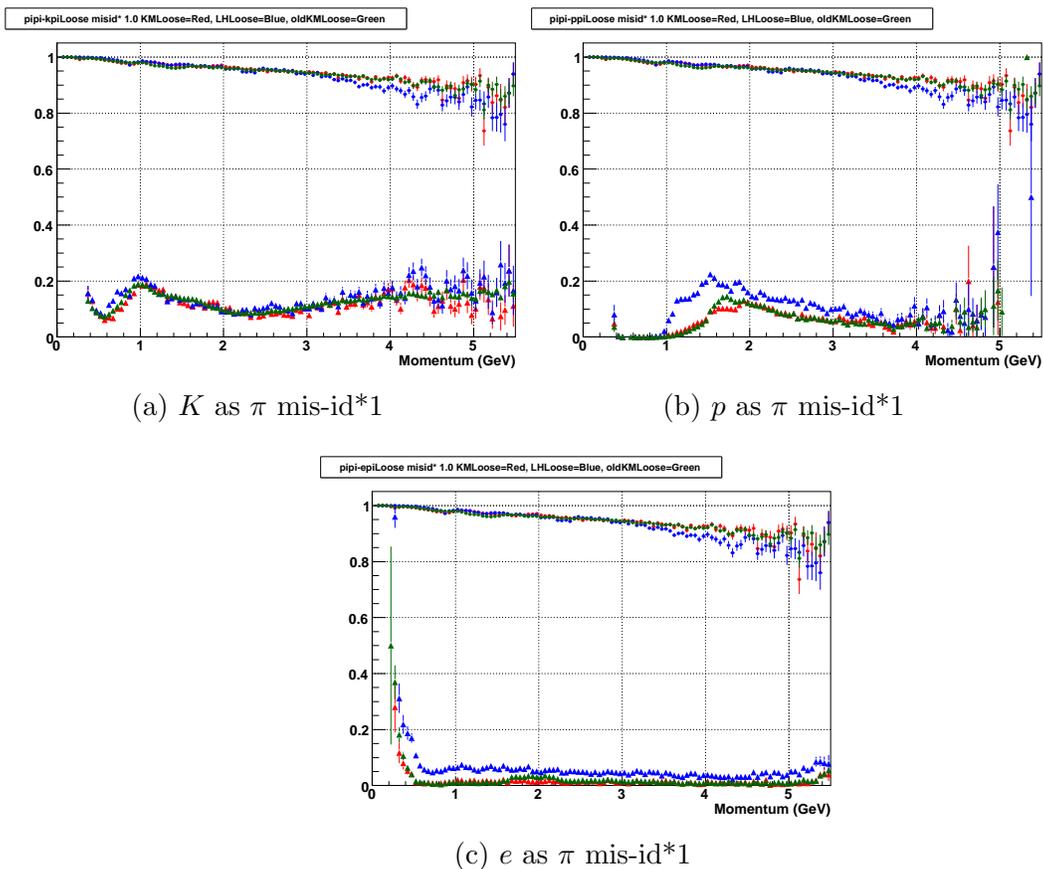


Figure A.16: Pion selection efficiencies and misidentification rates for Loose pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.2.4 Tight Pion Selector

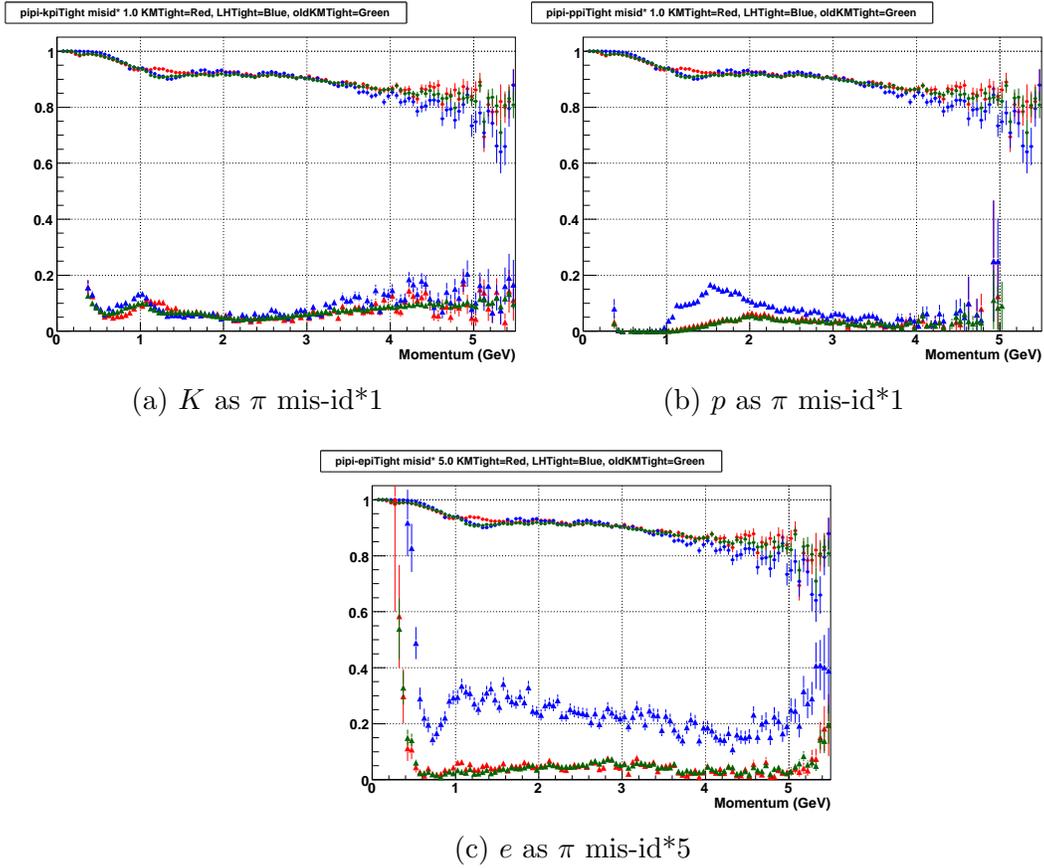


Figure A.17: Pion selection efficiencies and misidentification rates for Tight pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.2.5 Very Tight Pion Selector

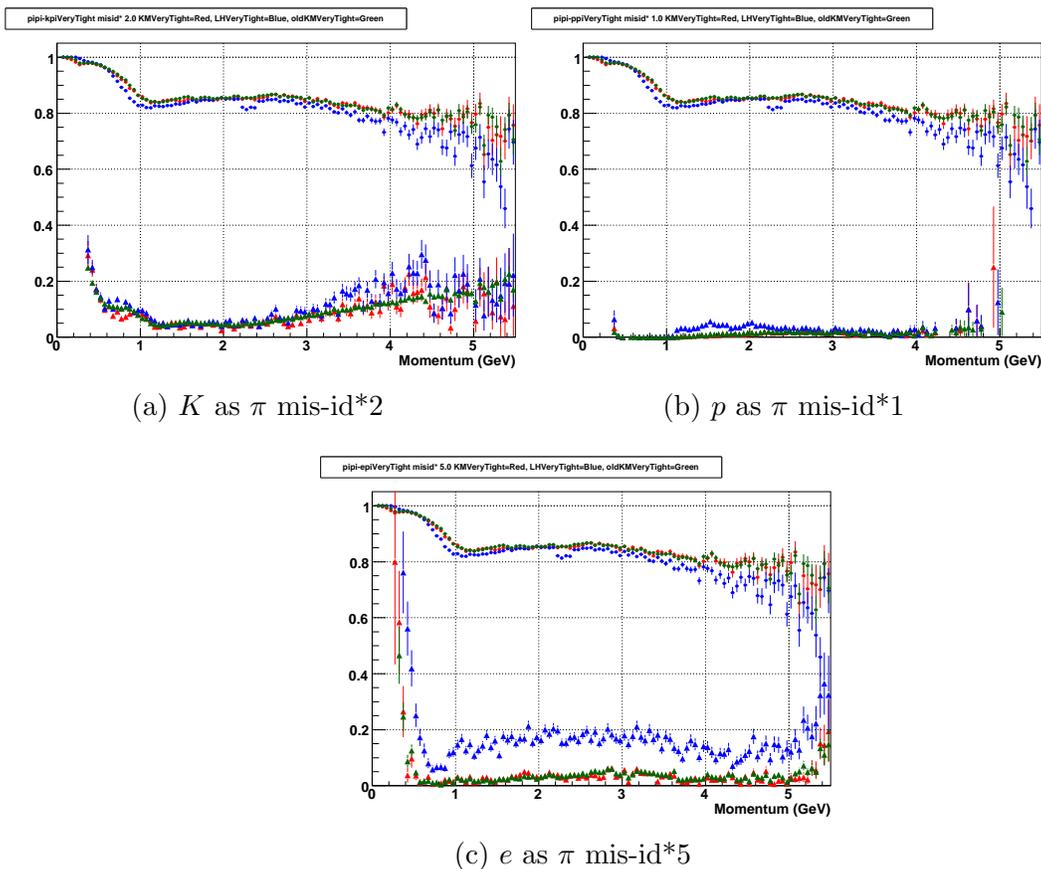


Figure A.18: Pion selection efficiencies and misidentification rates for Very Tight pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.2.6 Super Tight Pion Selector

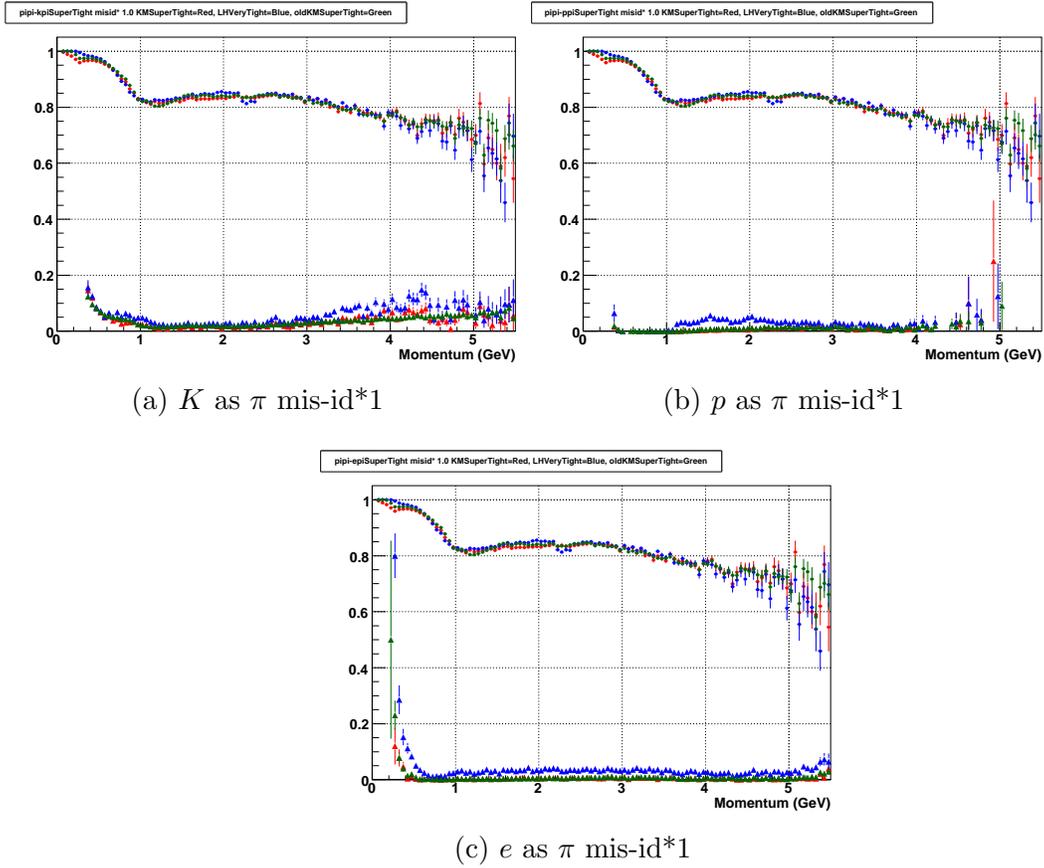
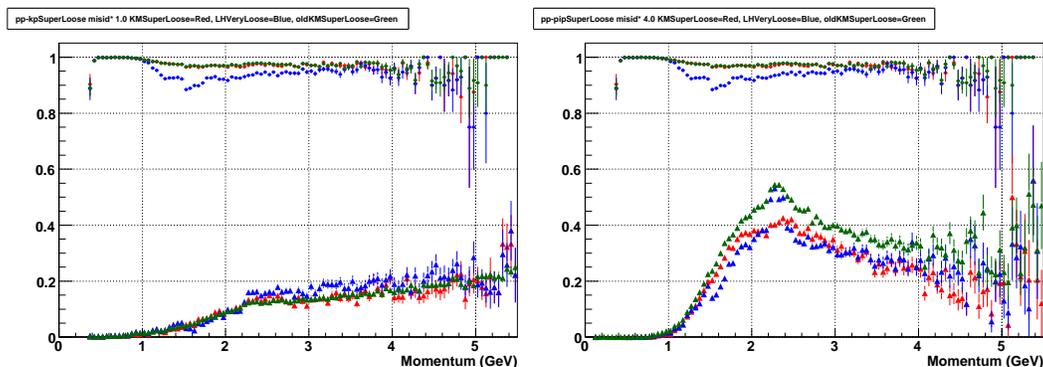


Figure A.19: Pion selection efficiencies and misidentification rates for Super Tight pion selector as  $\pi$  of  $K$  (a),  $p$  (b), and  $e$  (c). Pion selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiency and misidentification rate of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

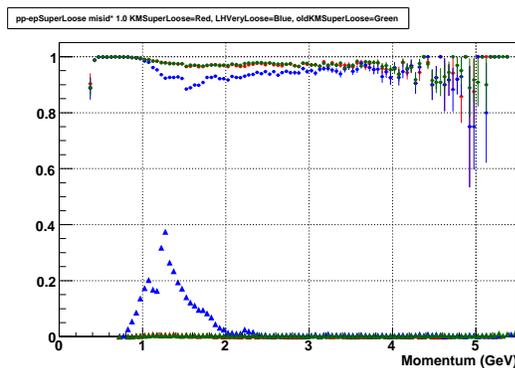
### A.3.3 Proton

#### A.3.3.1 Super Loose Proton Selector



(a)  $K$  as  $p$  mis-id\*1

(b)  $\pi$  as  $p$  mis-id\*4



(c)  $e$  as  $p$  mis-id\*1

Figure A.20: Proton selection efficiencies and misidentification rates for Super Loose proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.3.2 Very Loose Proton Selector

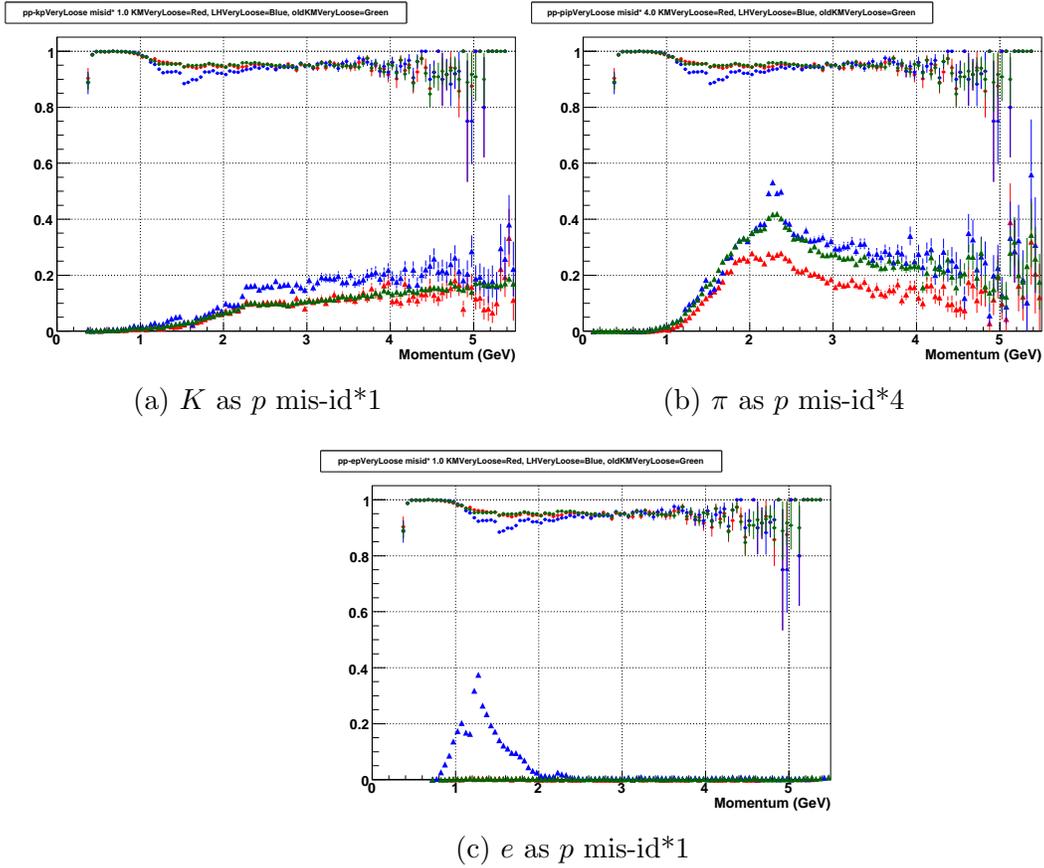


Figure A.21: Proton selection efficiencies and misidentification rates for Very Loose proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.3.3 Loose Proton Selector

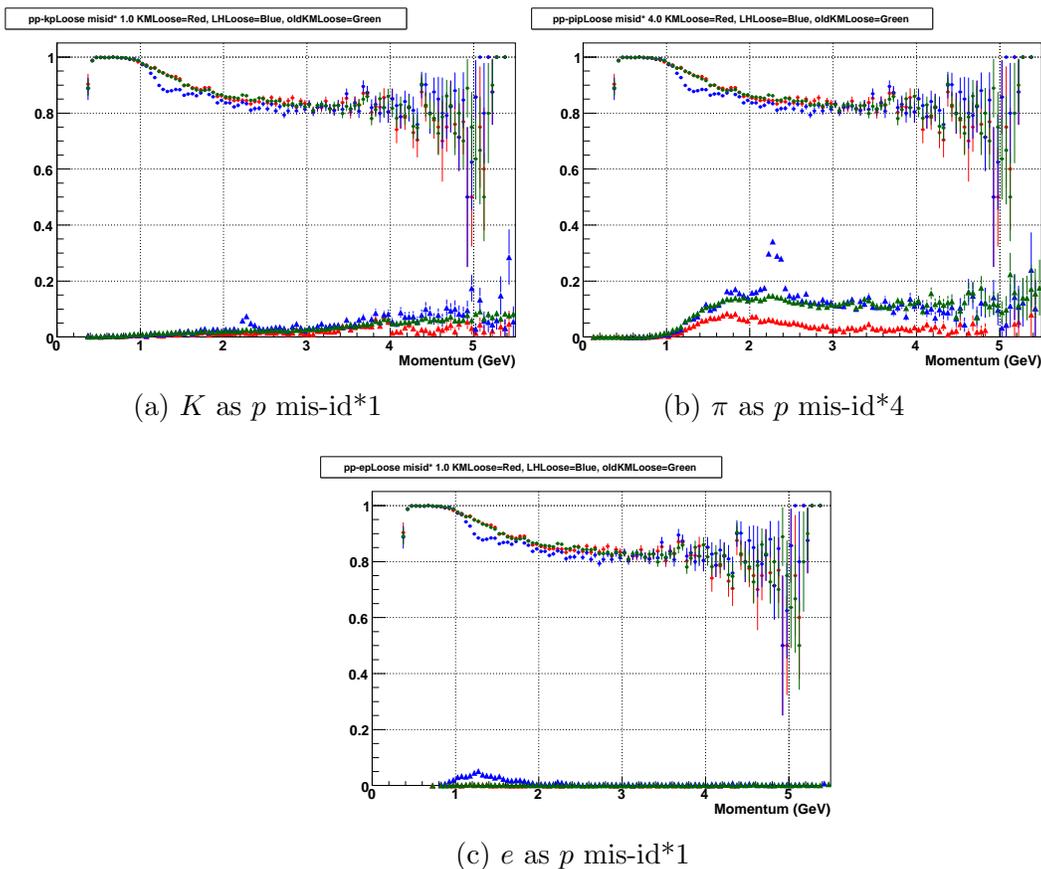


Figure A.22: Proton selection efficiencies and misidentification rates for Loose proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.3.4 Tight Proton Selector

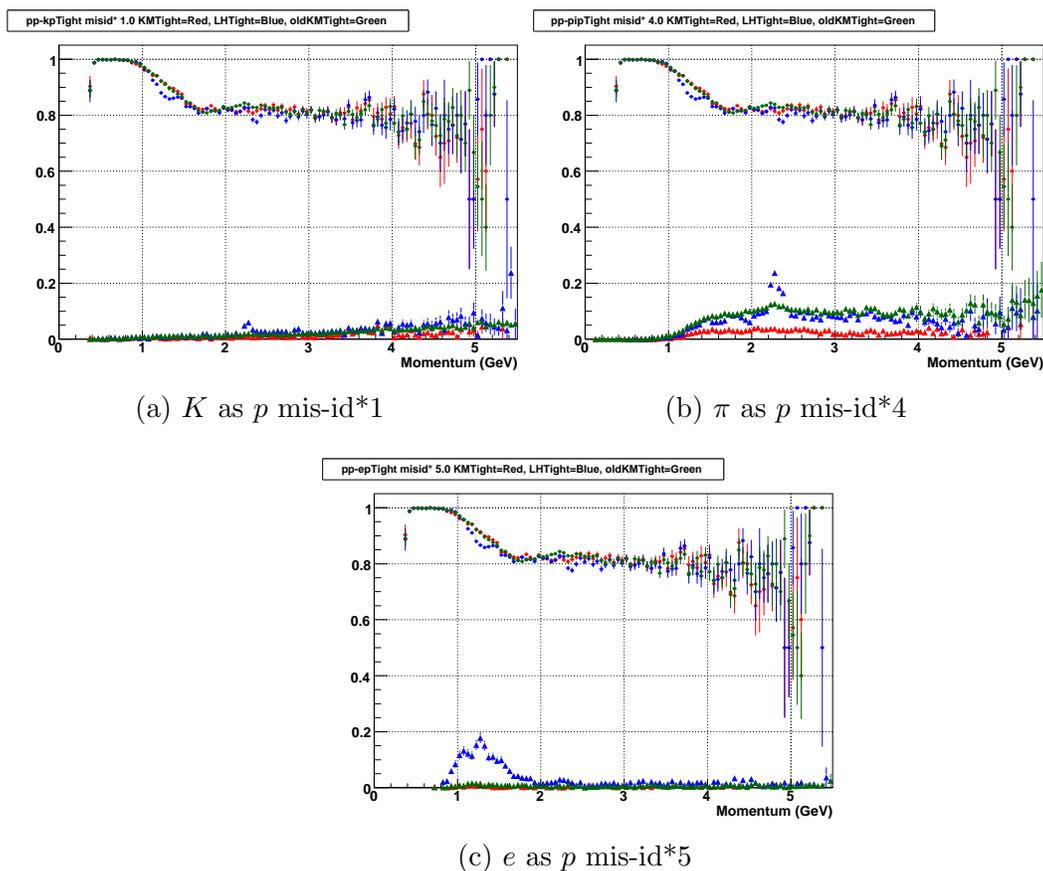


Figure A.23: Proton selection efficiencies and misidentification rates for Tight proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.3.5 Very Tight Proton Selector

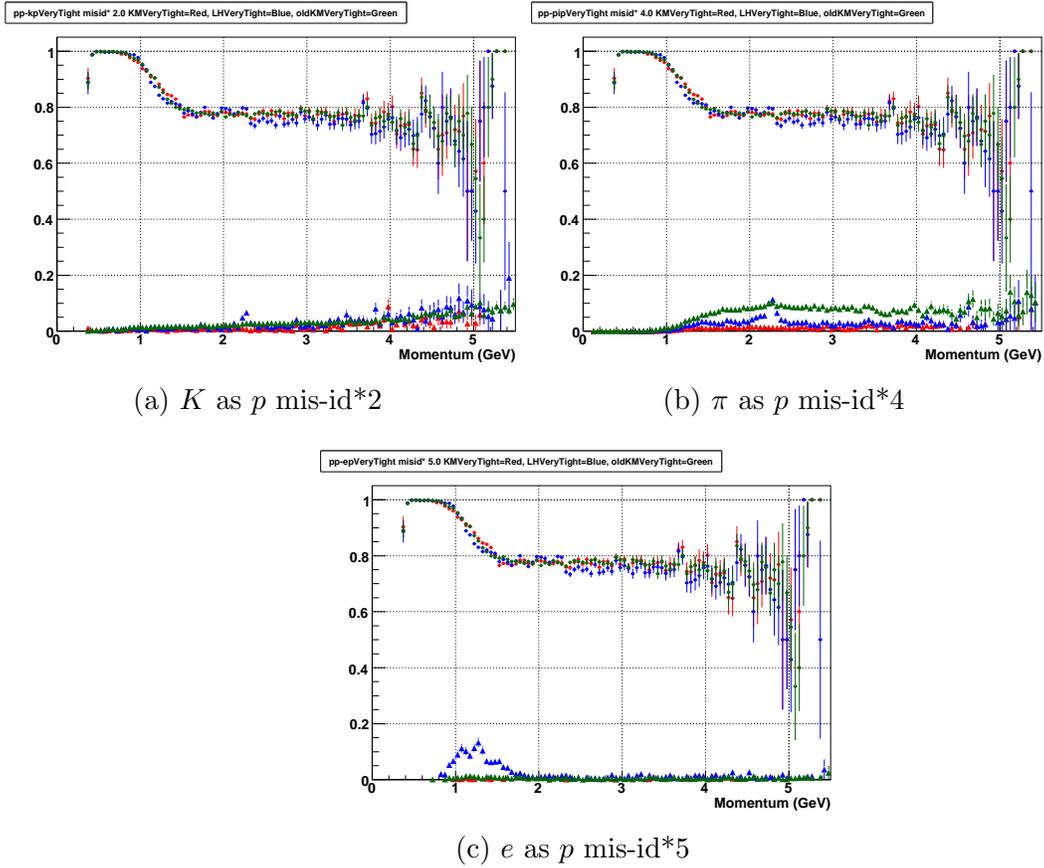


Figure A.24: Proton selection efficiencies and misidentification rates for Very Tight proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.3.6 Super Tight Proton Selector

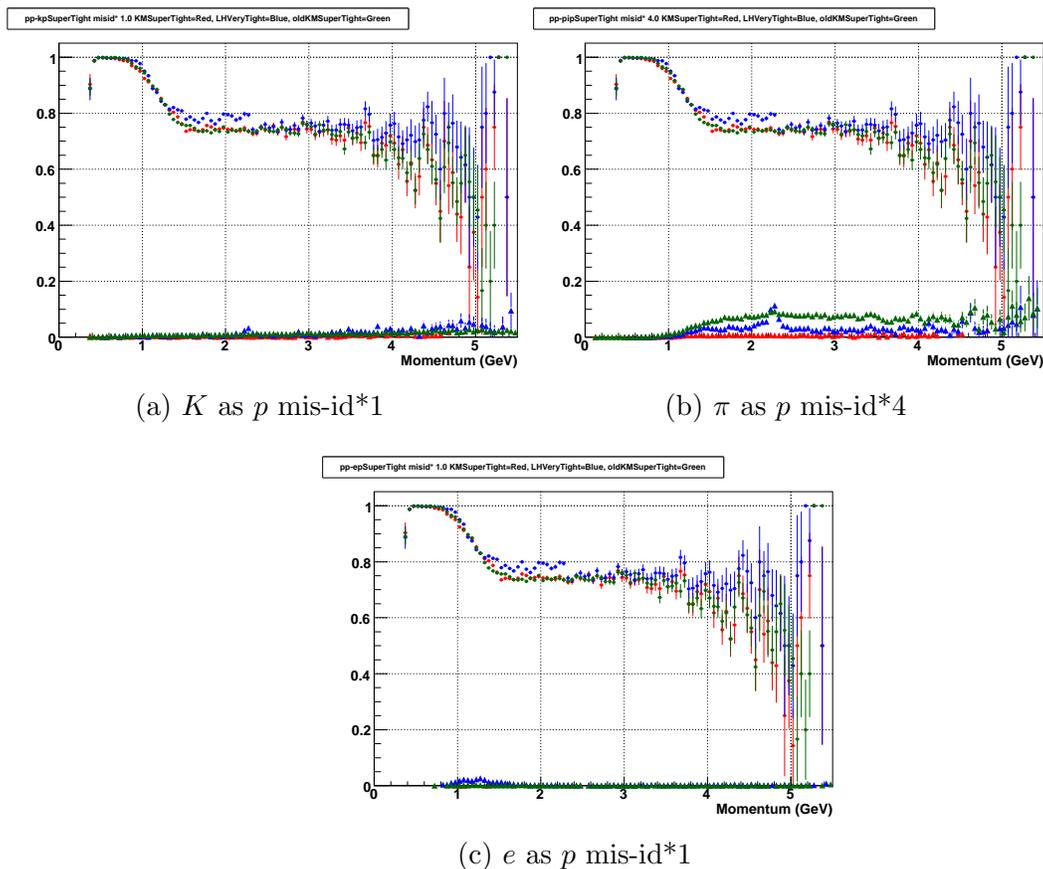


Figure A.25: Proton selection efficiencies and misidentification rates for Super Tight proton selector as  $p$  of  $K$  (a),  $\pi$  (b), and  $e$  (c). Proton selection efficiencies are shown with circles and misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.4 Electron

#### A.3.4.1 Super Loose Electron Selector

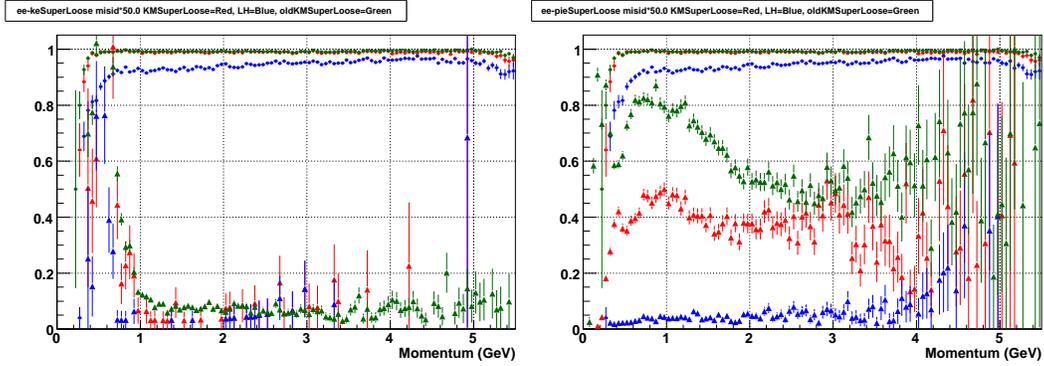
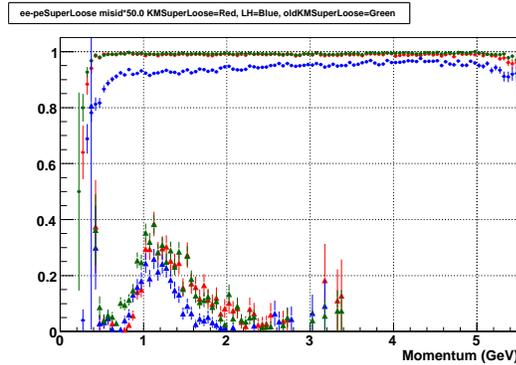
(a)  $K$  as  $e$  mis-id\*50(b)  $\pi$  as  $e$  mis-id\*50(c)  $p$  as  $e$  mis-id\*50

Figure A.26: Electron selection efficiencies and misidentification rates for Super Loose electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

## A.3.4.2 Very Loose Electron Selector

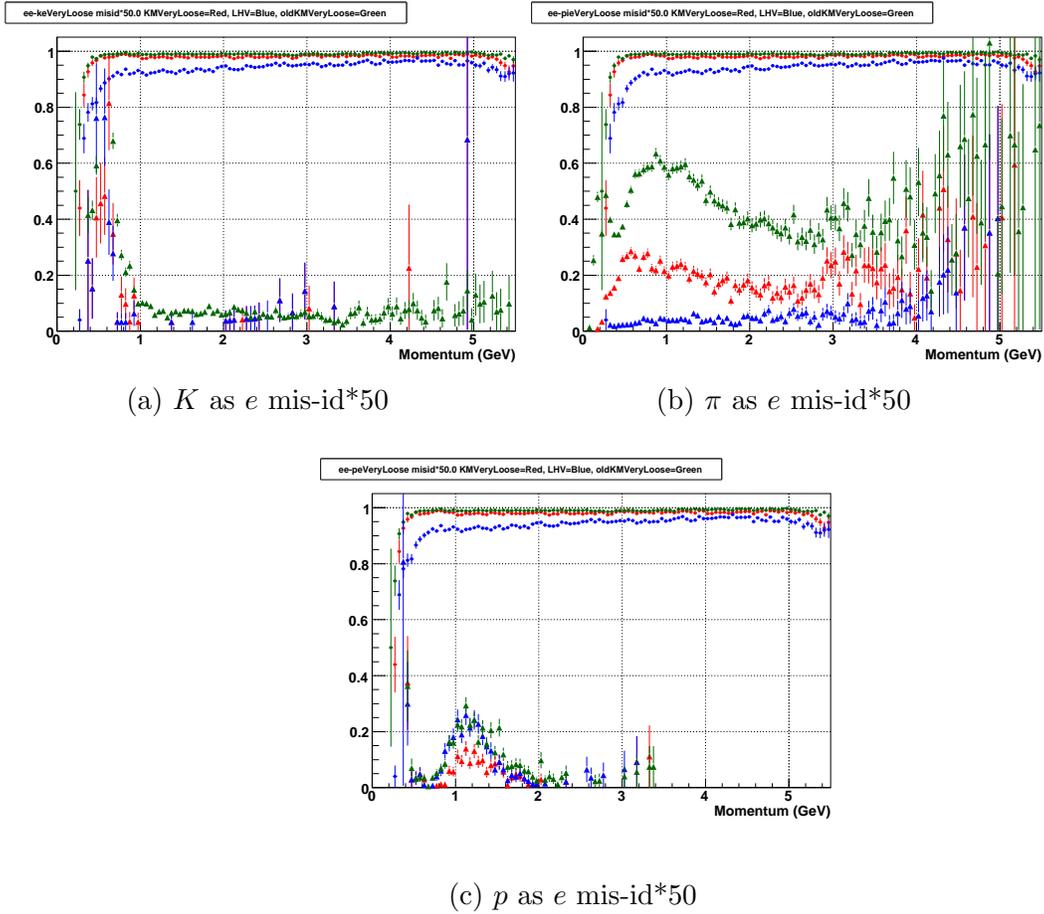


Figure A.27: Electron selection efficiencies and misidentification rates for Very Loose electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.4.3 Loose Electron Selector

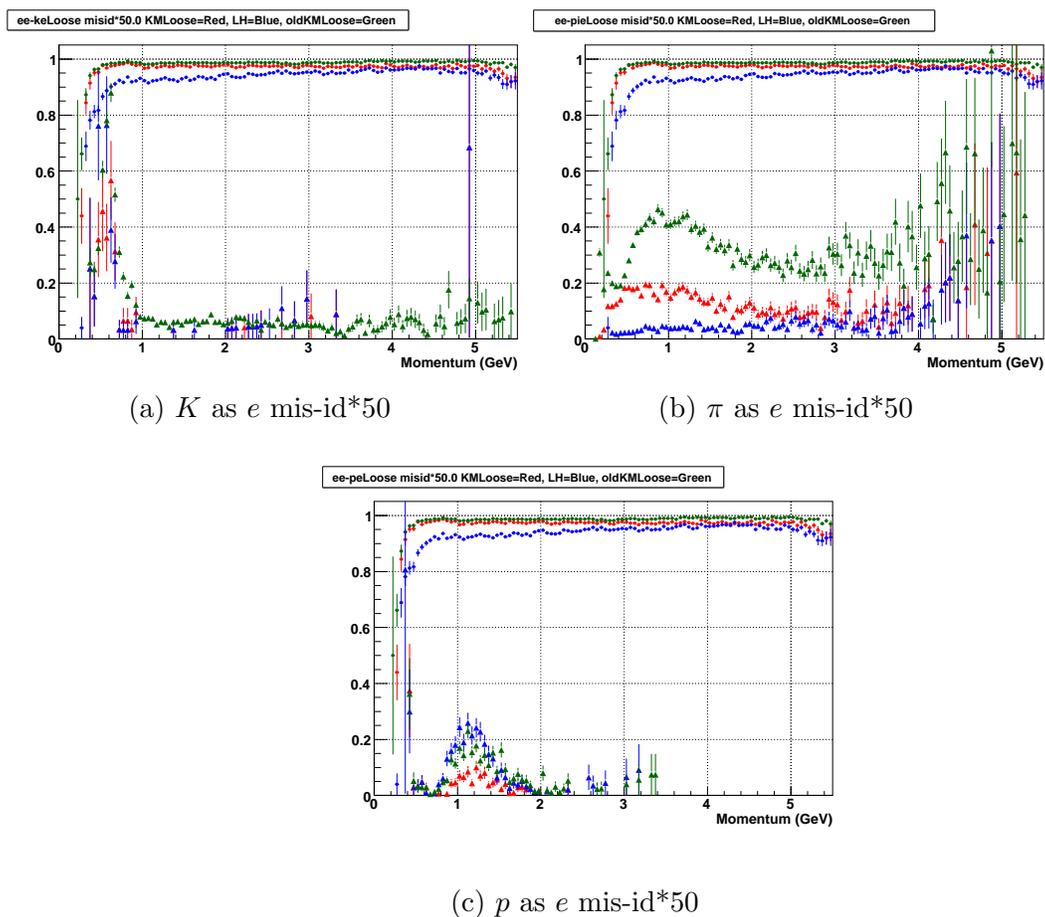


Figure A.28: Electron selection efficiencies and misidentification rates for Loose electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

### A.3.4.4 Tight Electron Selector

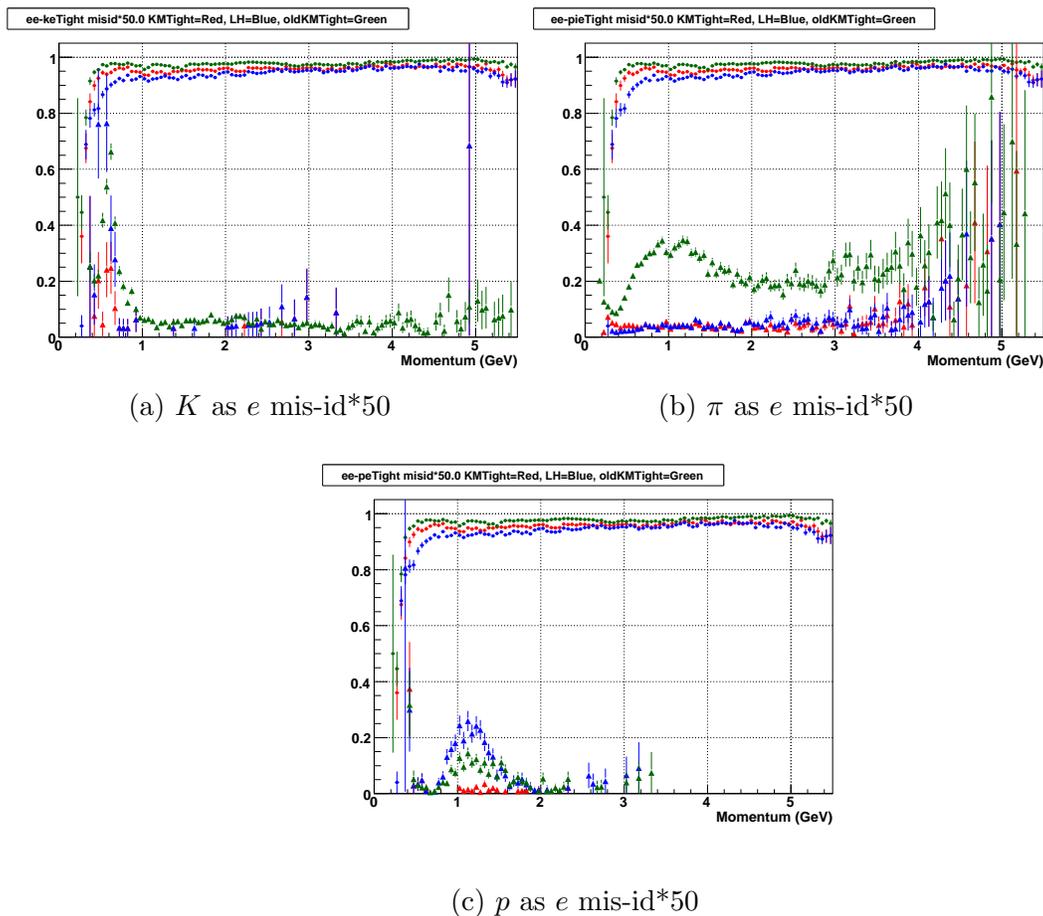


Figure A.29: Electron selection efficiencies and misidentification rates for Tight electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

## A.3.4.5 Very Tight Electron Selector

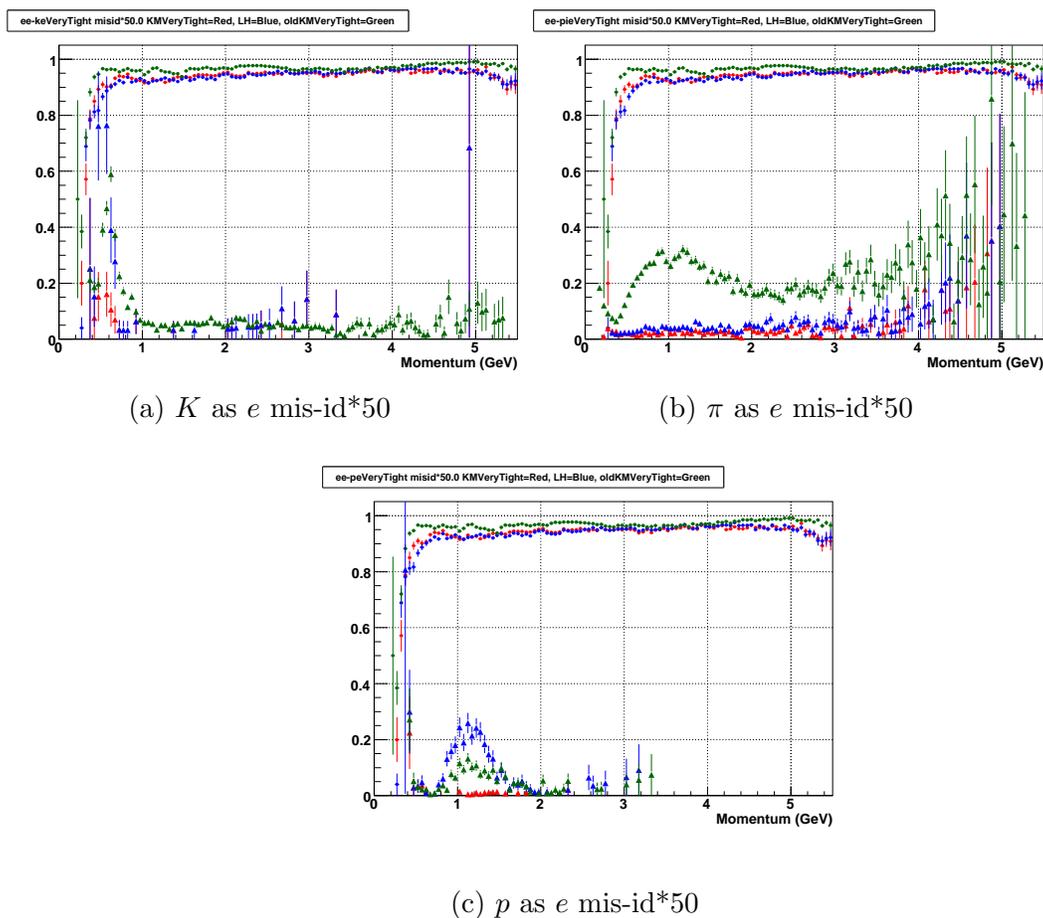


Figure A.30: Electron selection efficiencies and misidentification rates for Very Tight electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

## A.3.4.6 Super Tight Electron Selector

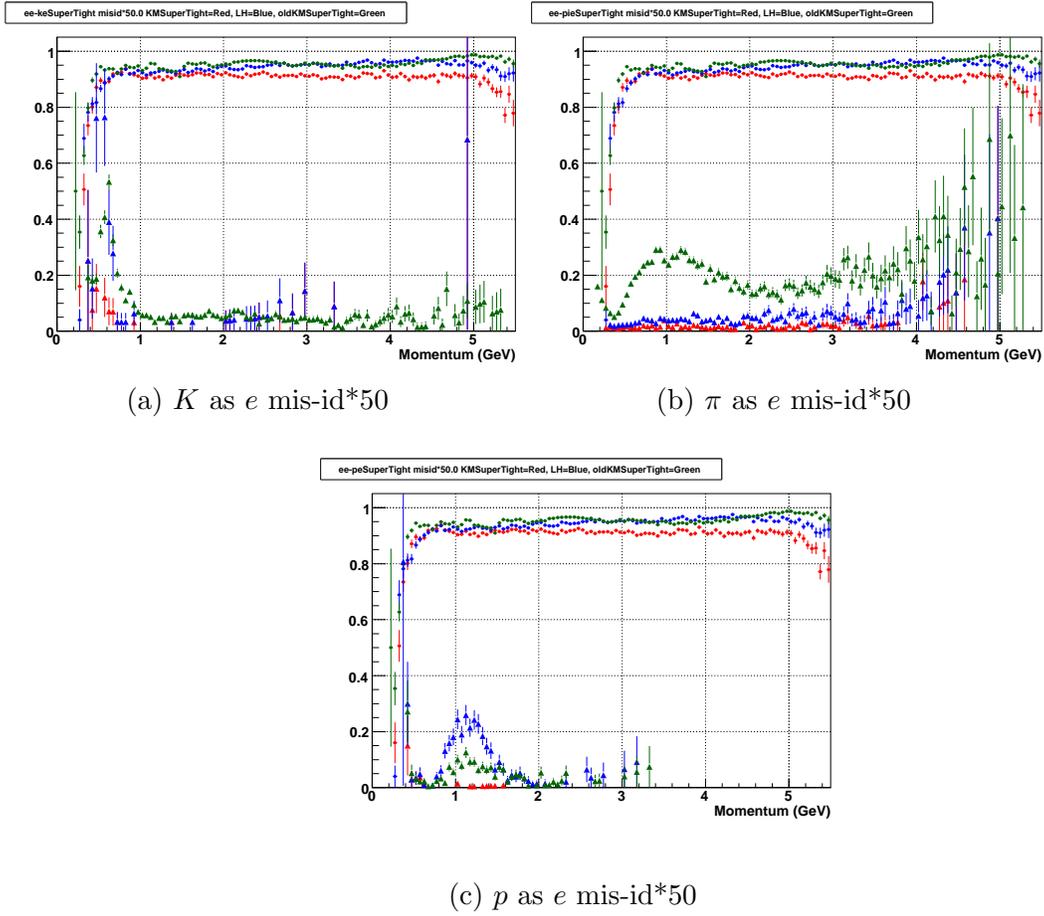


Figure A.31: Electron selection efficiencies and misidentification rates for Super Tight electron selector as  $e$  of  $K$  (a),  $\pi$  (b), and  $p$  (b). Electron selection efficiencies are shown with circles and the misidentification rates are shown with triangles. The efficiencies and misidentification rates of exhaustive matrix, 1vs1 matrix, and likelihood selector are shown in red, green and blue, respectively.

# Appendix B

## Data Set

The data and MC used in this analysis come from the *BABAR* internal framework Release 24, AllEventsSkim. The total integrated luminosity of both the on peak and offpeak data sets is given in Table B.1. The total MC for each potential background mode is given in Table B.2 and our signal modes in Table B.3.

Table B.1: The run-by-run integrated luminosity of the data.

	On Peak ( $fb^{-1}$ )	Off Peak ( $fb^{-1}$ )
Run 1	20.60	2.62
Run 2	62.07	7.03
Run 3	32.67	2.50
Run 4	100.78	10.23
Run 5	133.85	14.55
Run 6	79.03	7.89
total	429.0	44.81

Table B.2: The run-by-run number of MC events, cross section, equivalent luminosity, and weighting factor for each background mode.

Process	Generated Events ( $10^6$ )	Cross Section (nb)	Equivalent Luminosity ( $fb^{-1}$ )	Weighting Factor (data/MC)
$e^+e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}$ (mode SP-998)	44.59	2.09	21.33	0.965
	185.90		88.949	0.698
	137.54		65.809	0.496
	409.30		195.839	0.515
	526.64		251.979	0.531
	327.03		156.475	0.505
Total $u\bar{u}, d\bar{d}, s\bar{s}$	1631.00		780.39	0.550
$e^+e^- \rightarrow c\bar{c}$ (mode SP-1005)	55.25	1.30	42.503	0.485
	164.72		126.709	0.490
	88.32		67.939	0.481
	267.31		205.622	0.490
	344.28		264.827	0.505
	208.66		160.511	0.492
Total $c\bar{c}$	1128.54		868.11	0.494
$e^+e^- \rightarrow \tau^+\tau^-$ (mode SP-3429)	19.69	0.94	20.944	0.983
	57.19		60.845	1.020
	49.00		52.130	0.627
	180.08		191.571	0.526
	237.09		252.228	0.531
	137.66		146.446	0.540
Total $\tau^+\tau^-$	680.71		724.16	0.592
$e^+e^- \rightarrow B^+B^-$ (mode SP-1235)	34.88	0.55	64.415	0.320
	105.56		191.929	0.323
	56.04		101.882	0.321
	166.78		303.244	0.332
	215.17		391.215	0.342
	130.34		236.975	0.333
Total $B^+B^-$	708.76		1289.66	0.333
$e^+e^- \rightarrow B^0\bar{B}^0$ (mode SP-1237)	34.94	0.55	63.527	0.324
	104.19		189.433	0.328
	57.89		105.251	0.310
	169.80		308.729	0.326
	215.95		392.642	0.341
	135.22		245.862	0.321
Total $B^0\bar{B}^0$	718.00		1305.44	0.329

Table B.3: The run by run number of signal MC and luminosity weighting factors.

Process	Generated Events ( $10^6$ )	Cross Section (nb)	Luminosity ( $fb^{-1}$ )	Weighting Factor (data/MC)
$B^+ \rightarrow X_s \gamma$ (mode SP-6432)	0.632	$0.55 \times 2 \times 3.56 \times 10^{-4}$ $= 0.0003916$	1613.9	0.0128
	1.89		4816.1	0.0129
	0.99		2535.8	0.0129
	3.04		7755.4	0.0130
	4.00		10206.8	0.0131
	2.35		6006.1	0.0132
Total $B^+ \rightarrow X_s \gamma$	12.90		32934.1	0.0130
$B^0 \rightarrow X_s \gamma$ (mode SP-6431)	0.632	$0.55 \times 2 \times 3.56 \times 10^{-4}$ $= 0.0003916$	1613.9	0.0128
	1.89		4816.1	0.0129
	0.99		2535.8	0.0129
	3.04		7755.4	0.0130
	4.00		10206.8	0.0131
	2.35		6006.1	0.0132
Total $B^0 \rightarrow X_s \gamma$	12.90		32934.1	0.0130
$B^+ \rightarrow K^{*+} \gamma$ (mode SP-3135)	0.317	$0.55 \times 2 \times 4.03 \times 10^{-5}$ $= 4.433 \times 10^{-5}$	7150.9	0.0029
	0.94		21249.7	0.0029
	0.50		11188.8	0.0029
	1.52		34220.6	0.0029
	2.00		45071.1	0.0030
	1.18		26596	0.0030
Total $B^+ \rightarrow K^{*+} \gamma$	6.45		145477	0.0029
$B^0 \rightarrow K^{*0} \gamma$ (mode SP-3134)	0.317	$0.55 \times 2 \times 4.03 \times 10^{-5}$ $= 4.433 \times 10^{-5}$	7150.9	0.0029
	0.94		21249.7	0.0029
	0.50		11188.8	0.0029
	1.52		34220.6	0.0029
	2.00		45071.1	0.0030
	1.18		26596	0.0030
Total $B^0 \rightarrow K^{*0} \gamma$	6.45		145477	0.0029

## Appendix C

# $|\Delta E|$ -minimization as a Baseline for SSC

Since SSC is a new method for selecting the best candidate, we want to compare it the previous method of minimizing  $|\Delta E|$ . To do that, we selected the best candidate based on minimization of  $|\Delta E|$ . and then we scanned the cut values for  $|\Delta E|$ , BRC to find the values that gives the best precision ( $S/\sqrt{S+B}$ ) for each mass regions. The optimal cut for each mass  $X_s$  mass range is shown in table C.1.

Table C.1: The optimal requirements in each mass bin for the  $\frac{|\Delta E|}{\sigma_E}$  selected candidate analysis.

$m_{X_s}$ (GeV/ $c^2$ )	$\Delta E$ (GeV)	min. bump distance (cm)	BRC
0.6-1.1	$-0.13 < \Delta E < 0.07$	25	0.29
1.1-2.0	$-0.15 < \Delta E < 0.07$	25	0.38
2.0-2.4	$-0.13 < \Delta E < 0.07$	25	0.53
2.4-2.8	$-0.13 < \Delta E < 0.06$	25	0.43

# Appendix D

## $\pi^0$ Veto

Majority of high energy photon backgrounds contamination from the decays of  $\pi^0$ s. The distribution of amount of contamination from each of the different sources is shown in Figure D.1. These high energy photons from sources other than a  $b \rightarrow s\gamma$  transition increase the number of fake  $B$  candidates in an event.

Our strategy to get rid of these fake primary photons is to find photons that come from  $\pi^0$ s and prevent them from combining with our  $X_s$  candidates to make  $B$  candidates. To achieve this, we look at all combinations of the high energy photon ( $E_\gamma^* \in [1.6, 3.0]$ ) with all other photons in the event in the same event to form a  $\pi^0$  candidate. It can be seen in Figure D.1 that we could reject a reasonable amount of high energy photon backgrounds, originating from  $\pi^0$ s, simply by imposing a mass-window cut. However, this would also reject some of the primary photons from the  $b \rightarrow s\gamma$  transitions; a typical mass window cut of  $m_{\pi^0} \in [0.115, 0.150]$  GeV/ $c^2$  removes roughly 10% of our primary signal photons. Using a classifier, we can recover some of these primary signal photons that would otherwise be removed by imposing a  $\pi^0$  mass-window cut. Rather than place a hard cut on the classifier response (described below), we choose to use the output of the classifier as input to a more general background rejecting classifier, described in Section 5.1.3.2. When trying to determine where to place the cut on the classifier response, we would have to decide on some quantity to optimize, generally referred to as a figure of merit (FOM). By using the  $\pi^0$ -veto classifier as input to a more general background rejecting classifier, we can optimize our FOM at a later stage in the analysis procedure, thereby ensuring that a choice of

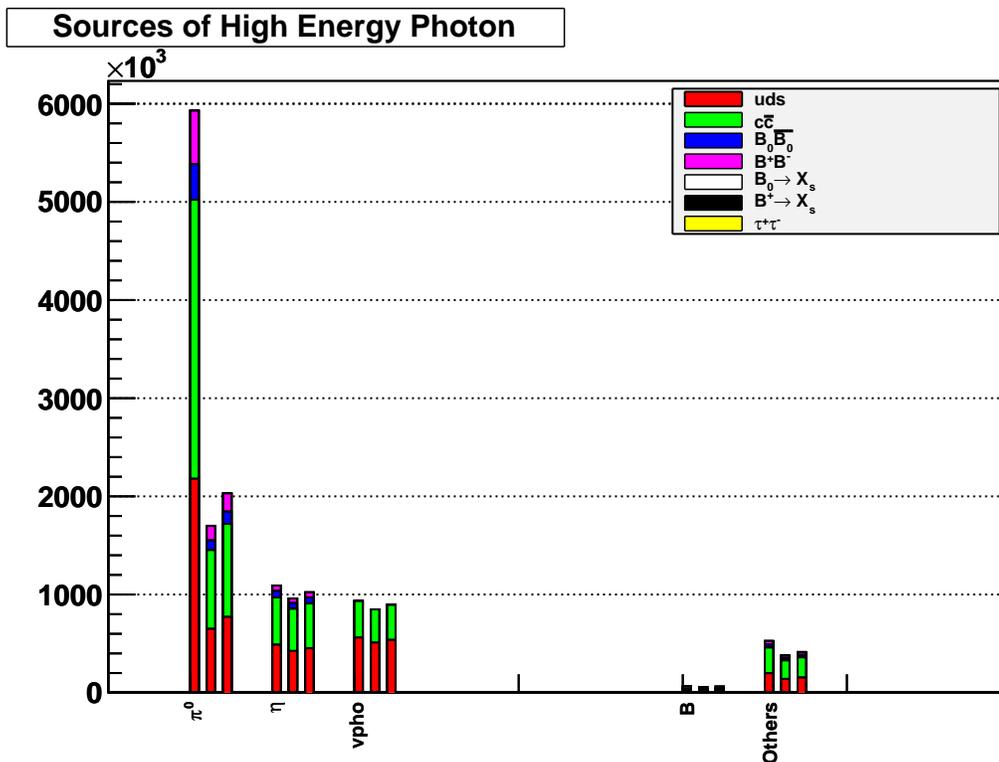


Figure D.1: Sources of background high energy photons. The three stack plots for each type of source represent the following: the left stack is without a  $\pi^0$  veto; the middle is after a mass window cut  $\in [0.11, 0.15]$  GeV; the right is with a cut on the  $\pi^0$  classifier described in the text (the cut location is 0.2 for demonstrative purposes only).

maximum  $\pi^0$ -veto classifier response at this stage does not have unnecessarily poor impact on signal efficiency.

We train a Bagged Decision Tree [29] [28] to separate between true  $\pi^0$  candidates (signal) and fake  $\pi^0$  candidates (background) using SPR[33]. The detailed summary of the classifier is given in Table D.1. The training sample we used in this study consists of  $\pi^0$  candidates from 0.25% of the total  $uds$ ,  $c\bar{c}$ ,  $B\bar{B}$ , and signal MC events. We reconstruct  $\pi^0$  candidates from 2 photons. We require that at least one photon in the  $\pi^0$  candidate passes the primary photon cuts. We define a  $\pi^0$  candidate to be signal if both photons truthmatch back to photons from the same  $\pi^0$  and background otherwise.

We trained the classifier based on 2 variables: 1) the energy of the less energetic photon measured in center of mass frame (minEGamma), and 2) the invariant mass of the  $\pi^0$  candidate. The normalized distributions of each variable for true  $\pi^0$  candidates and fake  $\pi^0$  candidates are shown in Figure D.2 and the probability for a  $\pi^0$  candidate to be a real  $\pi^0$  is shown in Figure D.3. The response from the classifier ranges from 0 (the photon is not likely to be originating from a  $\pi^0$ ) to 1 (more likely to be originating from a true  $\pi^0$ ). The classifier response for given invariant masses and minEGammas is shown in Figure D.4.

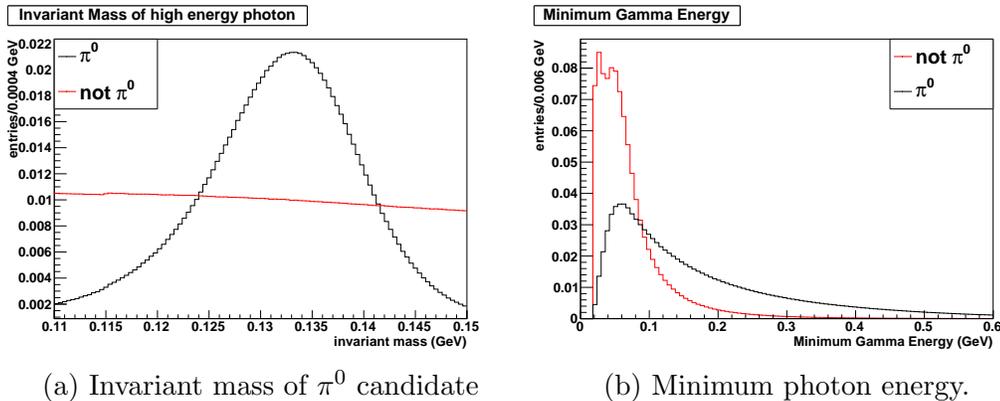


Figure D.2: Normalized distribution of variables used in  $\pi^0$  veto classifier.

A single high energy photon may have multiple  $\pi^0$  candidates. To assign a score to a photon, we compute the maximum classifier response from all  $\pi^0$  candidates that

Table D.1: Summary of  $\pi^0$  veto classifier. It should be underscored that this classifier is identifying the high energy photons that originate from  $\pi^0$ s as “signal”.

Sample Used:	$\pi^0$ candidates from 0.25% of the MC events that pass the preliminary cuts from the following MC samples <ul style="list-style-type: none"> <li>• <math>uds</math></li> <li>• <math>c\bar{c}</math></li> <li>• generic <math>B\bar{B}</math></li> <li>• signal MC</li> </ul>
Training Sample Requirements:	<ul style="list-style-type: none"> <li>• Event passes preliminary cuts given in section 5.1.</li> <li>• <math>\pi^0</math> candidates are formed with photons from Good-PhotonLoose list.</li> <li>• At least one of the photons passes the primary photon requirements.</li> <li>• The invariant mass of the <math>\pi^0</math> is <math>\in [0.11,0.15] \text{ GeV}/c^2</math>.</li> </ul>
Signal Definition:	<ul style="list-style-type: none"> <li>• Both photons are truth matched as photons</li> <li>• The parent of both truth matched photons is the same <math>\pi^0</math></li> <li>• otherwise defined as background</li> </ul>
Variables Used:	<ul style="list-style-type: none"> <li>• The energy of the lower energy photon, measured in the center of mass frame.</li> <li>• Invariant Mass of the two photons.</li> </ul>

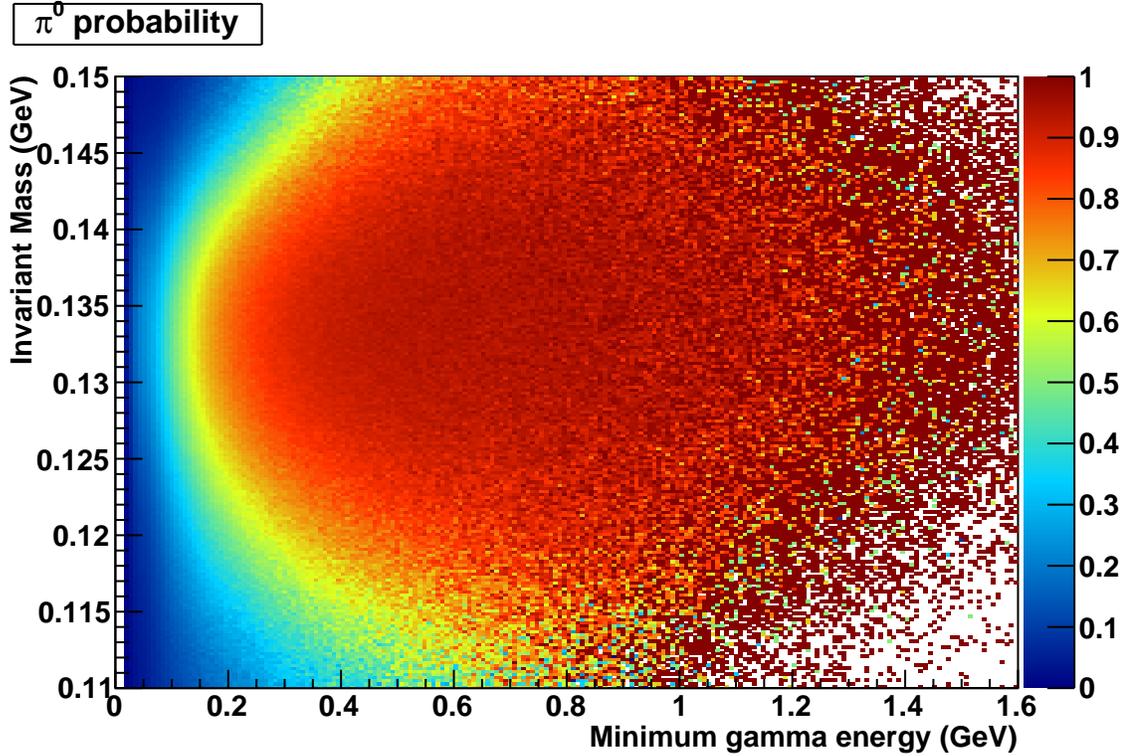


Figure D.3: Truth match probability of a  $\pi^0$  candidate to be a true  $\pi^0$ .

can be made with a given high energy photon. In the case that the photon could not be combined with any other photons to make a  $\pi^0$  candidate within the mass range of  $[0.11, 0.15]$   $\text{GeV}/c^2$ , we give it a score of -1. The negative value is chosen so that the lower the score, the less likely that a given high energy photon originates from a  $\pi^0$ , and vice versa. The normalized distribution of the score given to the high energy photons is shown in Figure D.5.

As mentioned above, we do not place a hard cut on the  $\pi^0$  classifier response. We instead use it as an input to the background rejecting classifier described in Section 5.1.3.2. It should be noted that if we did place a reasonable cut on the response of the classifier, we would still find that high energy photon contamination from  $\pi^0$  decays is greater than the contamination from the next highest source:  $\eta$  decays as illustrated Figure D.1. We still investigated if an explicit  $\eta$ -veto would be beneficial, or at least an  $\eta$  mass-window cut. We found that any mass-window vetos applied would decrease our precision ( $S/\sqrt{S+B}$ ), and a classifier-based veto would

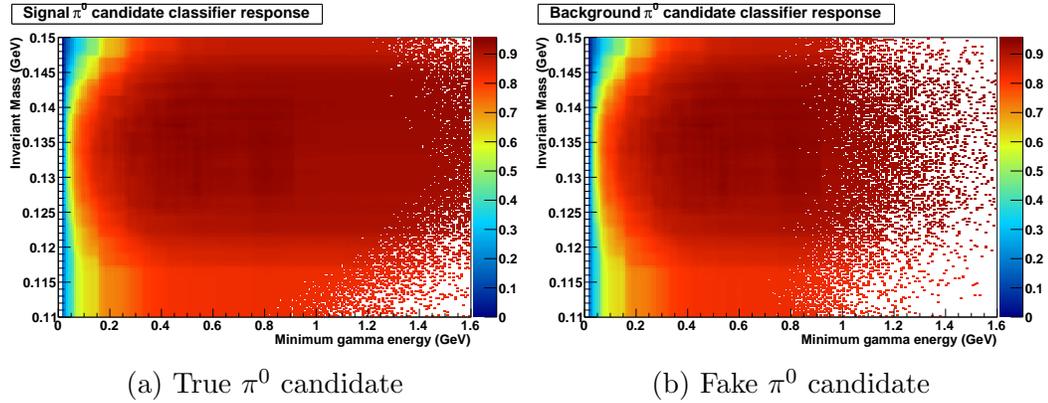


Figure D.4: The classifier response to a  $\pi^0$  candidate, depending on the candidate's minimum photon energy vs. invariant mass.

therefore probably only have marginal improvement at best (and negative impact on the FOM at worst). Because of this, we do not incorporate a dedicated  $\eta$ -photon veto.

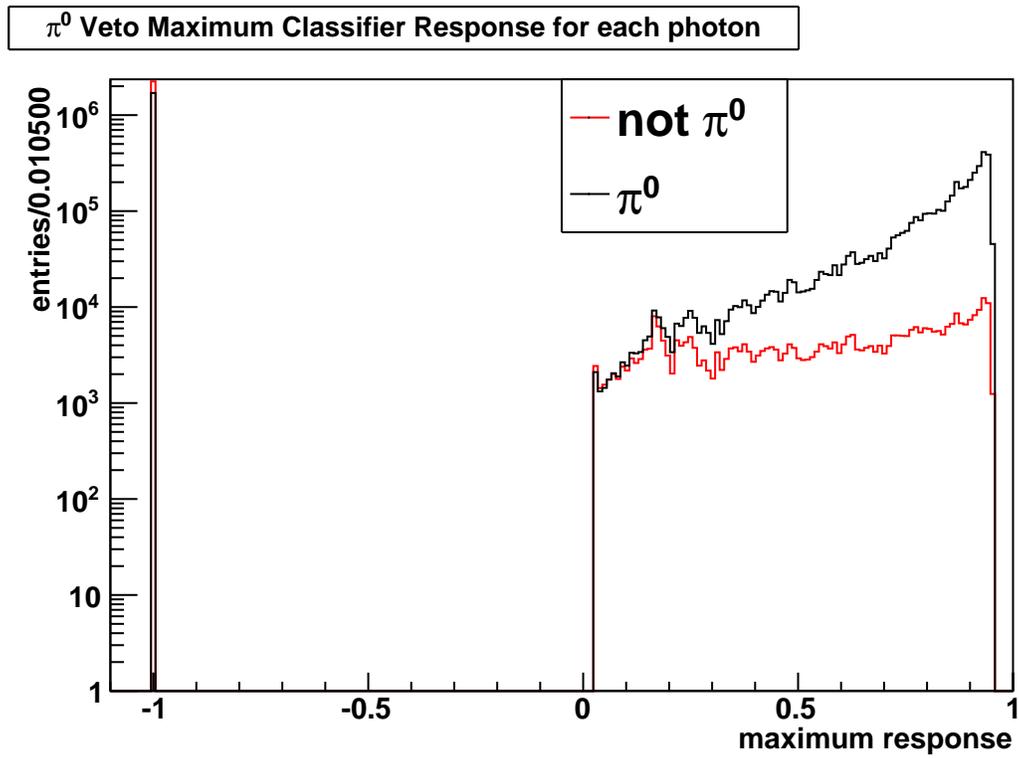


Figure D.5: Maximum classifier response for each high energy photon as described in the text.

## Appendix E

### SSC and BRC Variables

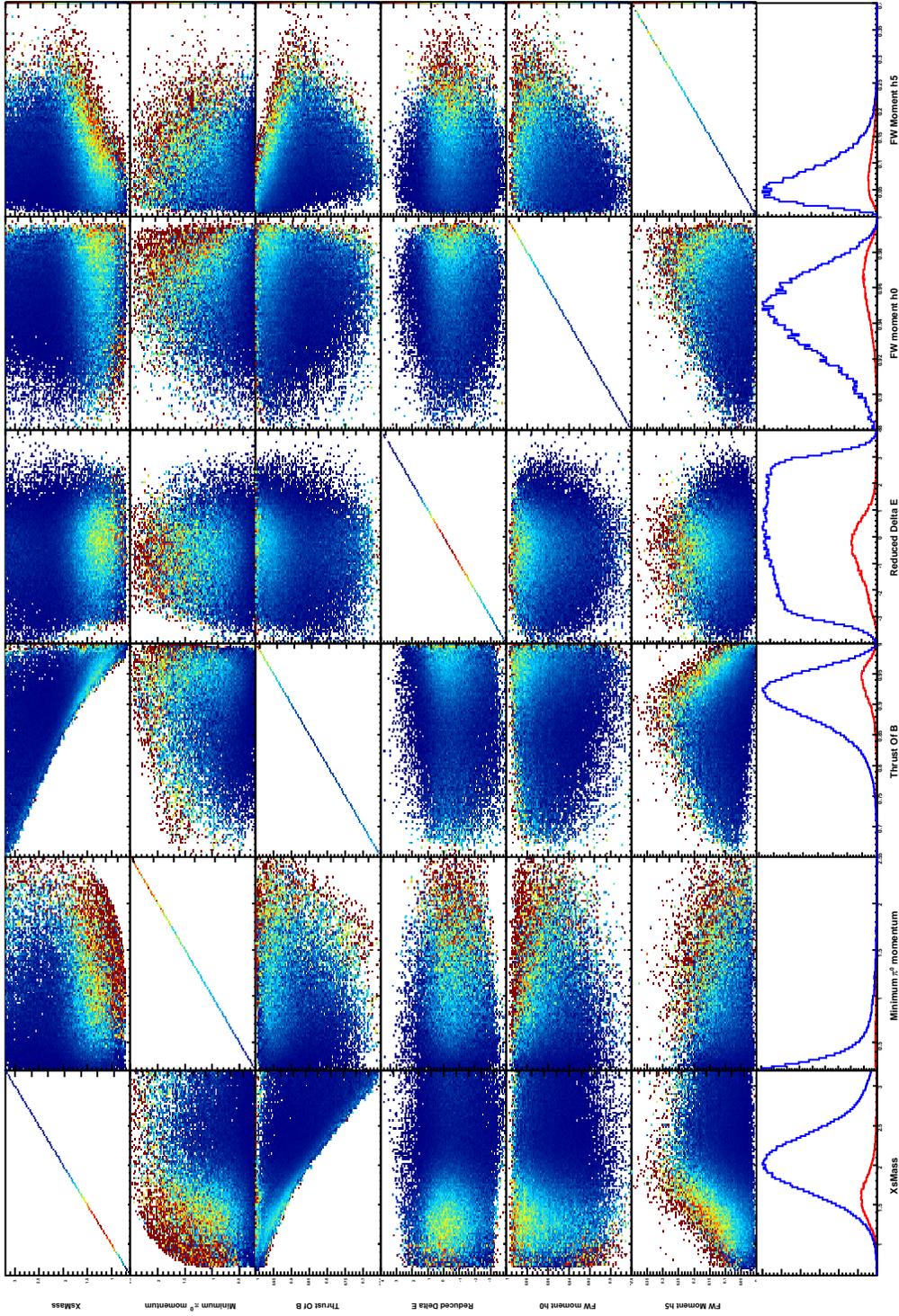


Figure E.1: Correlation between variables used in training the SSC. In each of the 2D histograms, the red indicates a higher concentration of correctly reconstructed candidates and blue a lower concentration of correctly reconstructed candidates. The bottom row plots are projections of each variable for mis-reconstructed candidates (blue) and correctly reconstructed candidates (red).

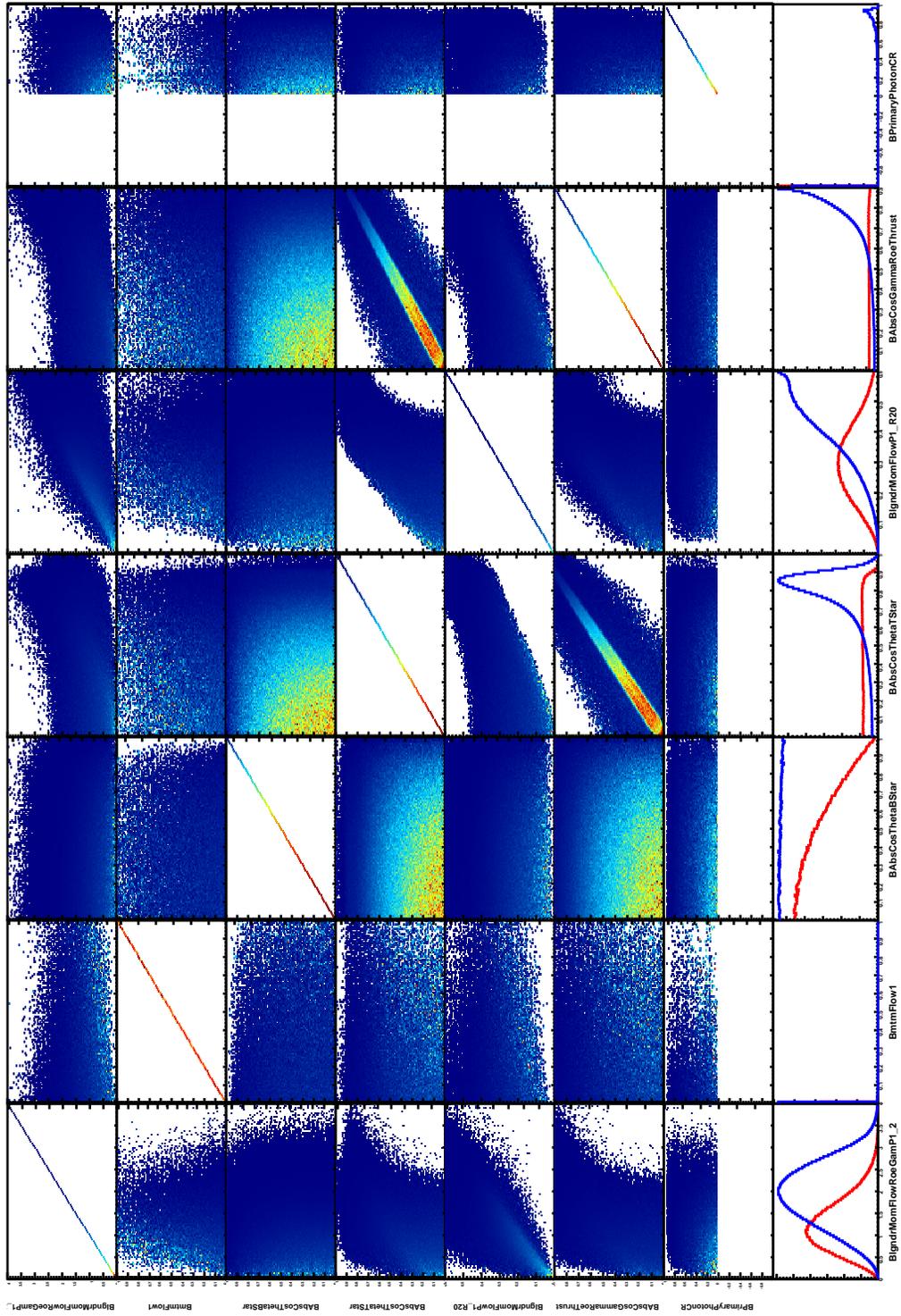


Figure E.2: Correlation plot of variables used in BRC training. B Momentum Flow 2-17 is omitted from the plot. In the 2D histogram red represents more concentration of signal events and blue represents less concentration of signal events.

## Appendix F

### Final Cut Optimization Miscellaneous Tables and Figures

Table F.1: Comparison of number of signal, number of background, and precision for each mode reduction strategy (mode-before/mode-after), pion and kaon tightness level. All the numbers presented here are from the optimal cuts. We also present the comparison of different ways of combining SSC and BRC output to make the final selection. Precision shown here is the total precision for for events with  $m_{\text{ES}} > 5.27$  GeV (including precision for Mass Range Individual column).

Strategy	Kaon Tightness	Pion Tightness	Classifier		Mass Range Global		Mass Range Individual				
			Signal	Background	Precision	Signal	Background	Precision	Signal	Background	Precision
modeafter	VeryLoose	VeryLoose	16387.5	16417.9	90.5	16015.5	18239.0	86.5	17437.0	33070.2	77.6
modeafter	VeryLoose	Loose	15994.0	15391.2	90.3	15820.3	17831.1	86.2	17191.2	31683.1	77.8
modeafter	VeryLoose	Tight	15080.9	13627.1	89.0	15272.2	16968.6	85.1	16554.7	29361.6	77.3
modeafter	VeryLoose	VeryTight	14404.7	13070.2	86.9	14206.9	15400.1	82.6	15499.0	27294.2	74.9
modeafter	Loose	VeryLoose	15822.8	14404.8	91.0	15560.6	16736.6	86.6	16501.8	27401.5	78.8
modeafter	Loose	Loose	15669.3	13984.0	91.0	15365.3	16356.3	86.3	16324.4	27240.2	78.2
modeafter	Loose	Tight	15020.7	12988.3	89.8	14814.0	15497.3	85.1	15976.4	26501.6	77.5
modeafter	Loose	VeryTight	14186.6	12336.5	87.1	13835.1	14275.0	82.5	14907.3	23983.2	75.6
modeafter	Tight	VeryLoose	15348.5	13162.2	90.9	15055.1	15487.0	86.1	15996.1	26207.6	77.9
modeafter	Tight	Loose	15177.8	12950.5	90.5	14852.9	15099.2	85.8	15780.6	25449.6	77.7
modeafter	Tight	Tight	14503.5	11934.2	89.2	14565.1	15068.5	84.6	15166.0	23134.7	77.5
modeafter	Tight	VeryTight	13745.5	11449.6	86.6	13592.3	13861.2	82.0	14562.1	23216.5	74.9
modeafter	VeryTight	VeryLoose	15367.0	13269.9	90.8	15050.5	15478.5	86.1	15991.4	26195.1	77.9
modeafter	VeryTight	Loose	15192.9	12962.8	90.5	14848.3	15090.7	85.8	15785.0	25574.6	77.6
modeafter	VeryTight	Tight	14477.9	11857.7	89.2	14560.5	15060.3	84.6	15161.3	23122.3	77.5
modeafter	VeryTight	VeryTight	13933.1	11987.0	86.5	13588.0	13853.1	82.0	14557.7	23204.3	74.9
modebefore	VeryLoose	VeryLoose	16266.1	16114.9	90.4	15965.2	18172.7	86.4	17504.4	33493.0	77.5
modebefore	VeryLoose	Loose	15946.8	15411.4	90.1	15769.2	17759.5	86.1	17287.1	32623.5	77.4
modebefore	VeryLoose	Tight	15403.6	14620.6	88.9	15349.6	17320.8	84.9	16668.0	30808.9	76.5
modebefore	VeryLoose	VeryTight	14572.6	13889.6	86.4	14152.4	15317.2	82.4	15583.6	28353.5	74.3
modebefore	Loose	VeryLoose	15843.1	15022.7	90.2	15455.4	16520.9	86.4	16613.2	28708.0	78.0
modebefore	Loose	Loose	15627.5	14608.3	89.9	15213.4	16007.1	86.1	16394.2	27908.2	77.9
modebefore	Loose	Tight	15135.6	13994.9	88.7	14893.2	15866.1	84.9	15799.2	26234.0	77.1
modebefore	Loose	VeryTight	14037.5	12019.3	87.0	13854.5	14444.7	82.4	14781.2	24410.9	74.7
modebefore	Tight	VeryLoose	14983.7	12872.7	89.8	15263.5	16247.2	86.0	16077.7	27051.0	77.4
modebefore	Tight	Loose	15136.9	13522.3	89.4	15064.9	15863.5	85.7	15859.5	26258.4	77.3
modebefore	Tight	Tight	14904.3	13469.9	88.5	14401.5	14694.3	84.4	15265.8	24235.1	76.8
modebefore	Tight	VeryTight	13844.6	11780.2	86.5	13549.7	13841.5	81.9	14404.4	23088.1	74.4
modebefore	VeryTight	VeryLoose	15214.7	13485.1	89.8	15258.9	16238.4	86.0	16072.8	27037.3	77.4
modebefore	VeryTight	Loose	14945.7	12999.9	89.4	15060.1	15854.5	85.7	15854.7	26244.4	77.3
modebefore	VeryTight	Tight	14666.4	12770.8	88.5	14396.9	14686.2	84.4	15260.9	24221.5	76.8
modebefore	VeryTight	VeryTight	13795.3	11662.2	86.5	13545.3	13833.0	81.9	14399.9	23075.0	74.4



Table F.2: Expected number of events with  $m_{ES} > 5.24$  GeV. The uncertainty given here is calculated from  $\sqrt{\sum weight}$ . It may appear that in some cases mass range individual gives lower precision for certain mass range. This is because the PID choice that gives best global FOM for each one of the methods is different. The strategy we selected is Mass Range Global.

Mass range (GeV)	MC Sample	Mass Range Global	Mass Range Individual	Classifier
0.6-1.1	uds	$6814.6 \pm 61.6$	$5397.3 \pm 54.9$	$7620.0 \pm 65.1$
	c $\bar{c}$ bar	$4924.6 \pm 48.8$	$3920.3 \pm 43.6$	$5159.1 \pm 50.0$
	$B\bar{B}$	$218.9 \pm 8.5$	$192.5 \pm 7.9$	$204.2 \pm 8.2$
	XFeed	$850.4 \pm 4.2$	$768.8 \pm 3.9$	$739.2 \pm 3.8$
	Signal	$5236.7 \pm 3.9$	$4955.9 \pm 3.8$	$5705.0 \pm 4.1$
	Precision	39.0	40.2	40.9
1.1-2.0	uds	$17201.4 \pm 97.6$	$19966.1 \pm 105.2$	$21190.4 \pm 108.3$
	c $\bar{c}$ bar	$14040.0 \pm 82.4$	$16275.0 \pm 88.7$	$17802.8 \pm 92.7$
	$B\bar{B}$	$2184.9 \pm 26.6$	$2551.2 \pm 28.7$	$2766.5 \pm 29.9$
	XFeed	$6962.2 \pm 15.2$	$7720.3 \pm 16.0$	$8375.6 \pm 16.8$
	Signal	$10372.4 \pm 19.7$	$10914.2 \pm 20.2$	$11654.6 \pm 21.0$
	Precision	46.0	45.5	46.9
2.0-2.4	uds	$364.5 \pm 14.3$	$6611.4 \pm 60.6$	$901.8 \pm 22.4$
	c $\bar{c}$ bar	$294.8 \pm 11.9$	$5524.9 \pm 51.6$	$845.0 \pm 20.2$
	$B\bar{B}$	$98.5 \pm 5.6$	$3138.8 \pm 31.9$	$428.7 \pm 11.8$
	XFeed	$39.5 \pm 1.0$	$898.3 \pm 5.0$	$197.6 \pm 2.4$
	Signal	$111.3 \pm 1.3$	$665.4 \pm 3.1$	$276.7 \pm 2.1$
	Precision	3.7	5.1	5.4
2.4-2.8	uds	$17.3 \pm 3.1$	$7553.6 \pm 64.6$	$54.2 \pm 5.5$
	c $\bar{c}$ bar	$12.0 \pm 2.4$	$6037.6 \pm 54.0$	$41.9 \pm 4.5$
	$B\bar{B}$	$3.8 \pm 1.1$	$4258.1 \pm 37.1$	$21.1 \pm 2.6$
	XFeed	$0.5 \pm 0.1$	$261.9 \pm 2.1$	$2.1 \pm 0.2$
	Signal	$2.5 \pm 0.1$	$138.2 \pm 0.8$	$6.9 \pm 0.2$
	Precision	0.4	1.0	0.6
0.6-2.8	uds	$24398.3 \pm 116.4$	$39528.7 \pm 148.0$	$29766.9 \pm 128.4$
	c $\bar{c}$ bar	$19271.9 \pm 96.5$	$31758.2 \pm 123.8$	$23849.2 \pm 107.3$
	$B\bar{B}$	$2506.1 \pm 28.5$	$10140.6 \pm 57.3$	$3420.5 \pm 33.3$
	XFeed	$7852.6 \pm 15.8$	$9649.4 \pm 17.4$	$9314.5 \pm 17.4$
	Signal	$15723.1 \pm 20.1$	$16673.8 \pm 20.8$	$17643.3 \pm 21.5$
	Precision	59.5	50.8	60.9

Table F.3: Expected number of events with  $m_{ES} > 5.27$  GeV. For each strategy. The uncertainty given here is calculated from  $\sqrt{\sum weight}$ . Note that in some of the mass ranges the Mass Range Individual method may give a lower precision than the Mass Range Global method. This is due to the difference in PID selectors.

Mass range (GeV)	MC Sample	Mass Range Global	Mass Range Individual	Classifier
0.6-1.1	uds	$1898.6 \pm 32.5$	$1480.8 \pm 28.7$	$2071.4 \pm 33.9$
	ccbar	$1260.6 \pm 24.7$	$986.9 \pm 21.9$	$1299.8 \pm 25.1$
	$B\bar{B}$	$142.9 \pm 6.8$	$127.3 \pm 6.4$	$139.8 \pm 6.7$
	XFeed	$508.9 \pm 2.9$	$463.7 \pm 2.7$	$458.6 \pm 2.6$
	Signal	$5184.7 \pm 3.9$	$4909.5 \pm 3.8$	$5648.3 \pm 4.1$
	Precision	54.7	55.0	57.6
1.1-2.0	uds	$4531.7 \pm 50.1$	$5265.0 \pm 54.0$	$5480.3 \pm 55.0$
	ccbar	$3393.3 \pm 40.5$	$3941.4 \pm 43.6$	$4283.1 \pm 45.5$
	$B\bar{B}$	$985.8 \pm 17.8$	$1125.9 \pm 19.1$	$1165.6 \pm 19.4$
	XFeed	$3790.1 \pm 11.3$	$4169.3 \pm 11.8$	$4502.6 \pm 12.4$
	Signal	$10268.5 \pm 19.6$	$10801.6 \pm 20.1$	$11536.2 \pm 20.8$
	Precision	67.8	67.9	70.2
2.0-2.4	uds	$79.9 \pm 6.6$	$1589.2 \pm 29.7$	$212.7 \pm 10.8$
	ccbar	$64.3 \pm 5.6$	$1219.2 \pm 24.3$	$182.6 \pm 9.4$
	$B\bar{B}$	$52.5 \pm 4.1$	$1448.5 \pm 21.6$	$213.9 \pm 8.3$
	XFeed	$20.1 \pm 0.7$	$365.5 \pm 3.1$	$95.2 \pm 1.6$
	Signal	$110.5 \pm 1.3$	$659.1 \pm 3.1$	$274.3 \pm 2.1$
	Precision	6.1	9.1	8.8
2.4-2.8	uds	$2.0 \pm 1.0$	$1842.3 \pm 32.0$	$15.6 \pm 2.9$
	ccbar	$1.9 \pm 1.0$	$1446.9 \pm 26.4$	$13.4 \pm 2.5$
	$B\bar{B}$	$1.6 \pm 0.7$	$1838.0 \pm 24.4$	$13.4 \pm 2.1$
	XFeed	$0.3 \pm 0.0$	$89.3 \pm 1.1$	$1.0 \pm 0.1$
	Signal	$2.5 \pm 0.1$	$137.2 \pm 0.8$	$6.8 \pm 0.2$
	Precision	0.9	1.9	1.0
0.6-2.8	uds	$6512.1 \pm 60.1$	$10177.3 \pm 75.1$	$7780.0 \pm 65.6$
	ccbar	$4720.1 \pm 47.8$	$7594.4 \pm 60.6$	$5779.0 \pm 52.8$
	$B\bar{B}$	$1182.8 \pm 19.5$	$4539.8 \pm 38.3$	$1532.7 \pm 22.3$
	XFeed	$4319.4 \pm 11.6$	$5087.7 \pm 12.6$	$5057.4 \pm 12.7$
	Signal	$15566.3 \pm 20.0$	$16507.5 \pm 20.7$	$17465.6 \pm 21.3$
	Precision	86.6	78.8	90.1

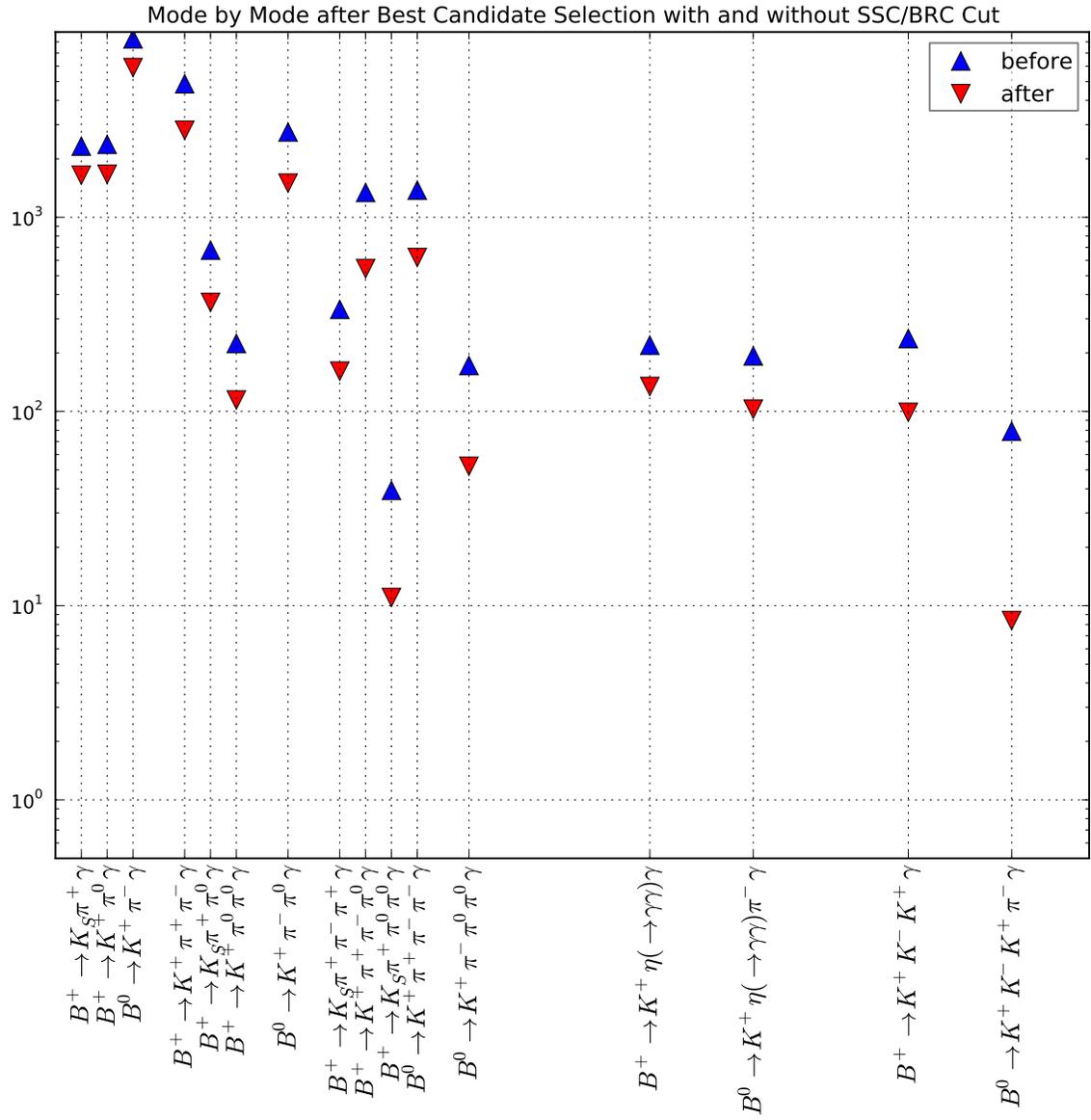


Figure F.2: Mode distribution before and after the optimal cut

# Appendix G

## Peaking $B\bar{B}$ Background

Most of the  $B\bar{B}$  peaking background are coming from  $B \rightarrow D\pi^0$  or  $B \rightarrow D\rho$  and we reconstructed one of  $\pi^0$ 's daughter as the primary photon and  $D$  decays to final states we are trying to reconstruct  $X_s$ . Since we capture most of the final states of  $B$ , the  $m_{ES}$  distribution peaks at  $B$  mass. Figure H.6 shows the  $m_{ES}$  distribution of peaking background for both types of  $B$ . Table G.1 lists some example of the source of background for modes with high contributions. This is by no mean a comprehensive list but to give some idea what the major source of background is for each mode.

Table G.1: The 16 modes we used for  $A_{CP}$  analysis. And some examples of peaking  $B\bar{B}$  background source for each mode. Charged conjugation is implied.

BiType	Final State	Peaking $B\bar{B}$
1	$B^+ \rightarrow K_S \pi^+ \gamma$	Doubly Cabibbo Suppressed $B^+ \rightarrow D^+ \pi^0$ ; $D^+ \rightarrow K_S^0 \pi^+$
2	$B^+ \rightarrow K^+ \pi^0 \gamma$	Quadruply Cabibbo Suppressed $B^+ \rightarrow D^+ \pi^0$ ; $D^+ \rightarrow K^+ \pi^0$
3	$B^0 \rightarrow K^+ \pi^- \gamma$	$B^0 \rightarrow \bar{D}^0 \pi^0$ ; $\bar{D}^0 \rightarrow K^+ \pi^-$
5	$B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$	$B^+ \rightarrow \bar{D}^0 \rho^+$ ; $\bar{D}^0 \rightarrow K^+ \pi^-$ , $\rho^+ \rightarrow \pi^+ \pi^0$
6	$B^+ \rightarrow K_S \pi^+ \pi^0 \gamma$	rare
7	$B^+ \rightarrow K^+ \pi^0 \pi^0 \gamma$	rare
9	$B^0 \rightarrow K^+ \pi^- \pi^0 \gamma$	$B^0 \rightarrow \bar{D}^0 \pi^0$ ; $\bar{D}^0 \rightarrow K^+ \pi^- \pi^0$
11	$B^+ \rightarrow K_S \pi^+ \pi^- \pi^+ \gamma$	rare
12	$B^+ \rightarrow K^+ \pi^+ \pi^- \pi^0 \gamma$	$B^+ \rightarrow \bar{D}^0 \rho^+$ ; $\bar{D}^0 \rightarrow K^+ \pi^-$ , $\rho^+ \rightarrow \pi^- \pi^0$
13	$B^+ \rightarrow K_S \pi^+ \pi^0 \pi^0 \gamma$	rare
14	$B^0 \rightarrow K^+ \pi^+ \pi^- \pi^- \gamma$	$B^0 \rightarrow \bar{D}^0 \pi^0$ ; $\bar{D}^0 \rightarrow K^+ \pi^- \pi^+ \pi^-$
16	$B^0 \rightarrow K^+ \pi^- \pi^0 \pi^0 \gamma$	rare
23	$B^+ \rightarrow K^+ \eta \gamma$	rare
27	$B^0 \rightarrow K^+ \eta \pi^- \gamma$	$B^0 \rightarrow \bar{D}^0 \eta$ ; $\bar{D}^0 \rightarrow K^+ \pi^-$
33	$B^+ \rightarrow K^+ K^- K^+ \gamma$	rare
37	$B^0 \rightarrow K^+ K^- K^+ \pi^- \gamma$	rare

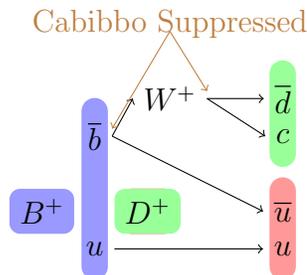


Figure G.1:  $B^+ \rightarrow D^+ \pi^0$  is doubly Cabibbo suppressed.  $D^+ \rightarrow K^+ \pi^0$  is also suppressed analogously.

There is a noticeable difference in that shape of  $m_{\text{ES}}$  distribution of peaking background for neutral  $B$  and charged  $B$  (see top two row of figure H.6). The neutral  $B$  one is more peaking than the charged  $B$  one. The difference in  $m_{\text{ES}}$  distribution of peaking background for charged and neutral  $B$  is from the fact that  $B^+ \rightarrow \bar{D}^0 \pi^0$  is doubly Cabibbo suppressed as figure G.1 illustrated that we need  $\bar{b} \rightarrow W^+ \bar{d}$  and  $W^+ \rightarrow \bar{s}u$  to make a  $B^+ \rightarrow \bar{D}^0 \pi^0$ . On top of that, we also ignore  $K_L^0$ . Thus, the peaking background for mode 1 ( $B^+ \rightarrow K_S \pi^+ \gamma$ ) and 2 ( $B^+ \rightarrow K^+ \pi^0 \gamma$ ) is much less than the peaking background of mode 3 ( $B^0 \rightarrow K^+ \pi^- \gamma$ ).

# Appendix H

## Toy Generation Procedure

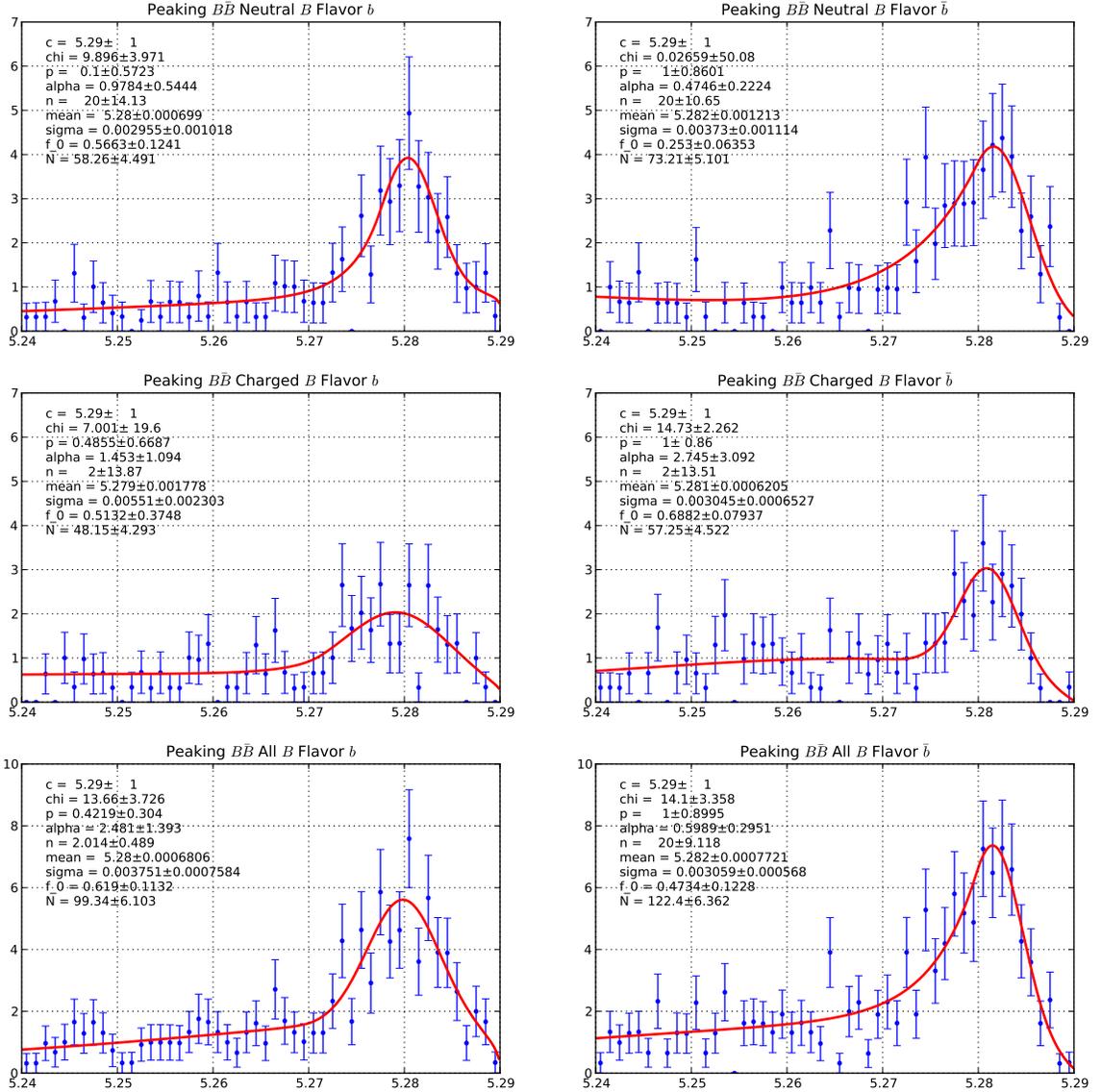
To obtain the shape parameters so that we can generate the toy sample and perform toy study. We fit, the each component of Monte Carlo separately. The fit method we are using is binned likelihood fit with 50 bins. The PDF used are listed on Table H.1

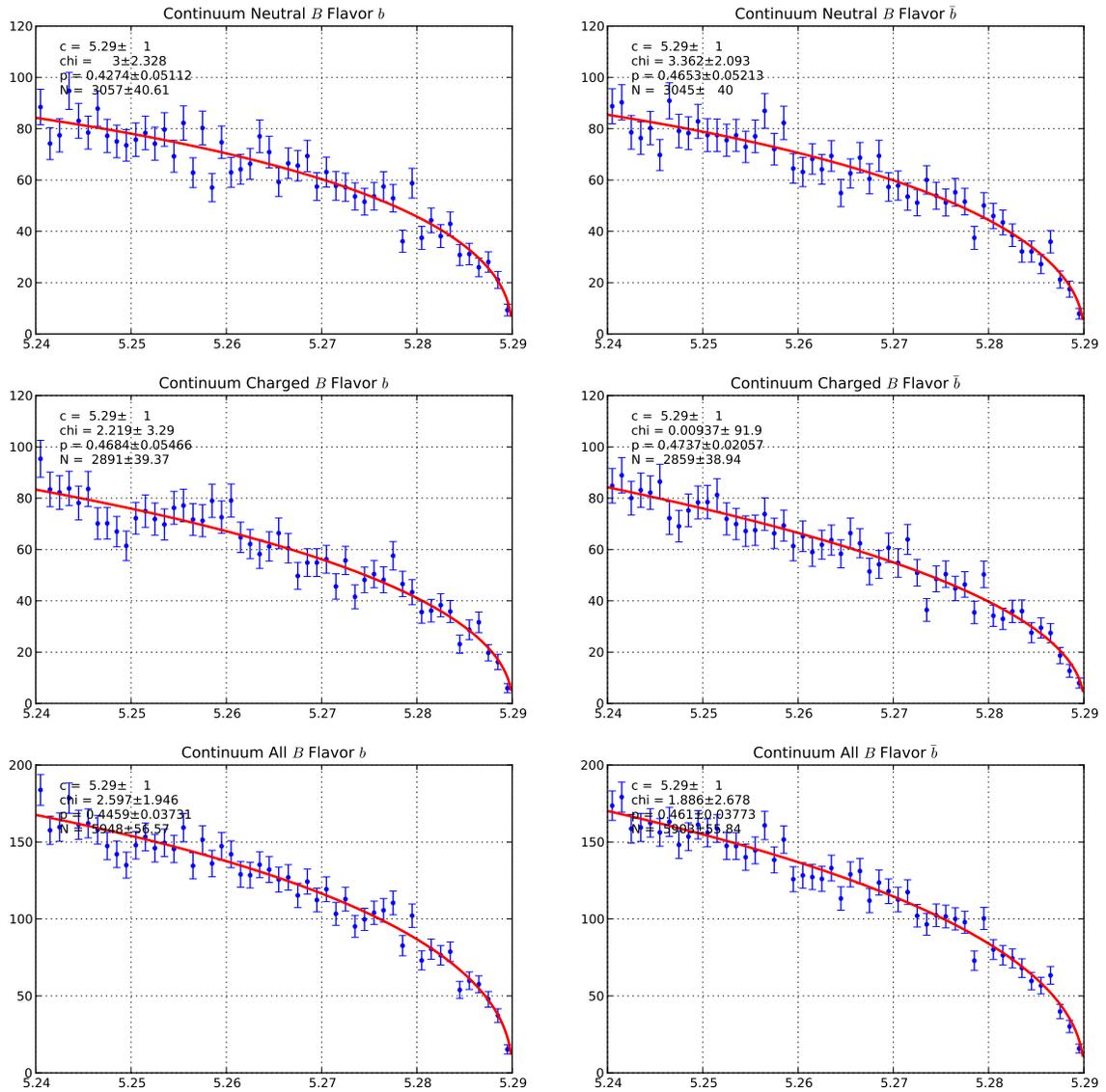
Table H.1: PDF for each type of sample used for generating toy.

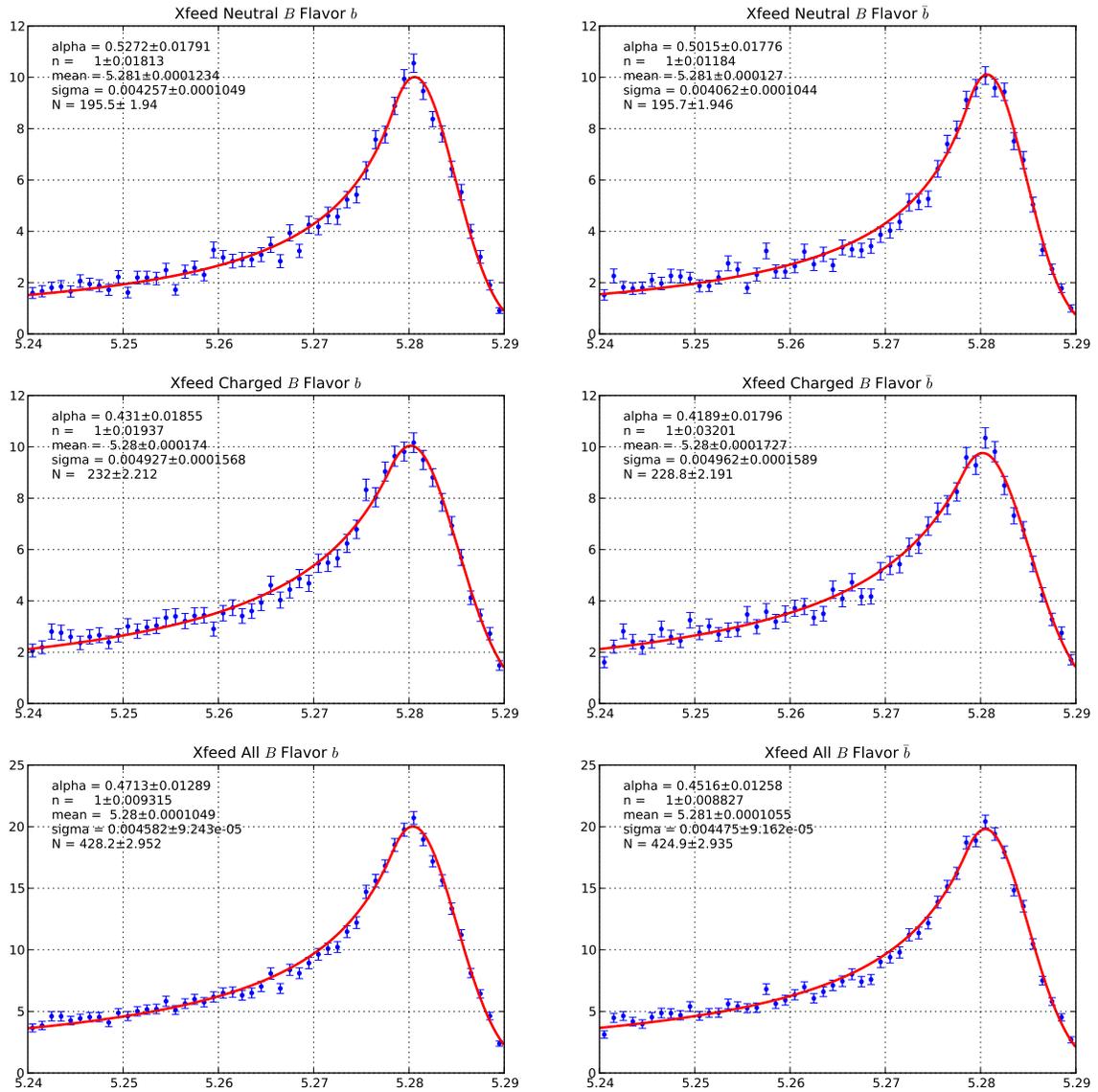
Sample	PDF	Fixed Parameters
Continuum	$\text{Argus}(x; \chi, p, m_0)$	$m_0 = 5.29 \text{ GeV}$
Peaking $B\bar{B}$	$k \cdot \text{CrystalBall}(x; \alpha, n, \bar{x}, \sigma) + (1 - k) \cdot \text{Argus}(x; \chi, p, m_0)$	$m_0 = 5.29 \text{ GeV}$ $p = 0.5$
XFeed	$\text{CrystalBall}(x; \alpha, n, \bar{x}, \sigma)$	None
Signal	$\text{Cruijff}(x; m_0, \sigma_L, \sigma_R, \alpha_L, \alpha_R)$	None

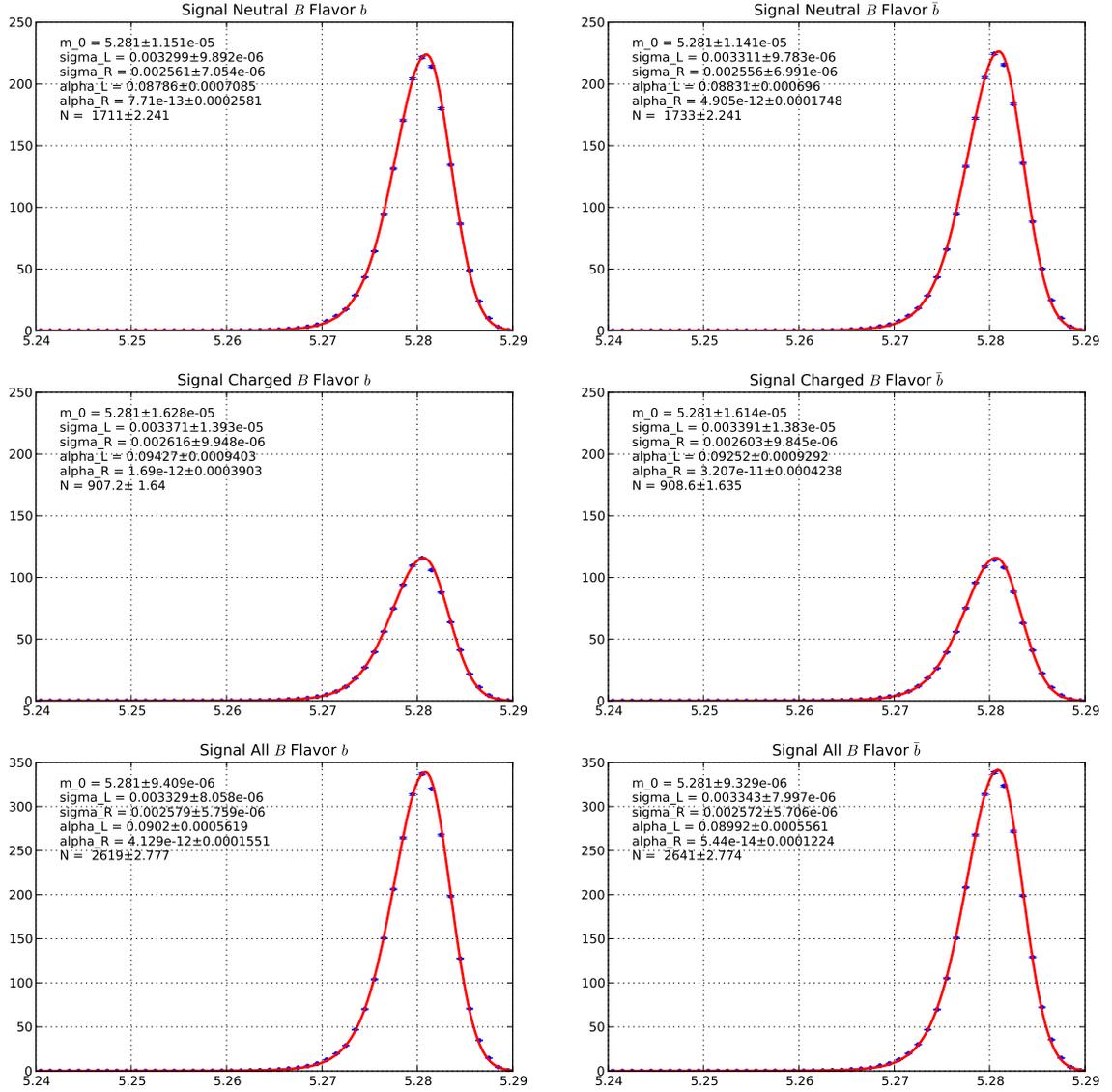
### H.1 $K^*$ Region

Plots are on the next page.

Figure H.1: Fit to MC peaking  $B\bar{B}$  subsample for  $K^*$  mass region.

Figure H.2: Fit to MC continuum subsample for  $K^*$  mass region.

Figure H.3: Fit to MC Xfeed subsample for  $K^*$  mass region.

Figure H.4: Fit to MC signal subsample for  $K^*$  mass region

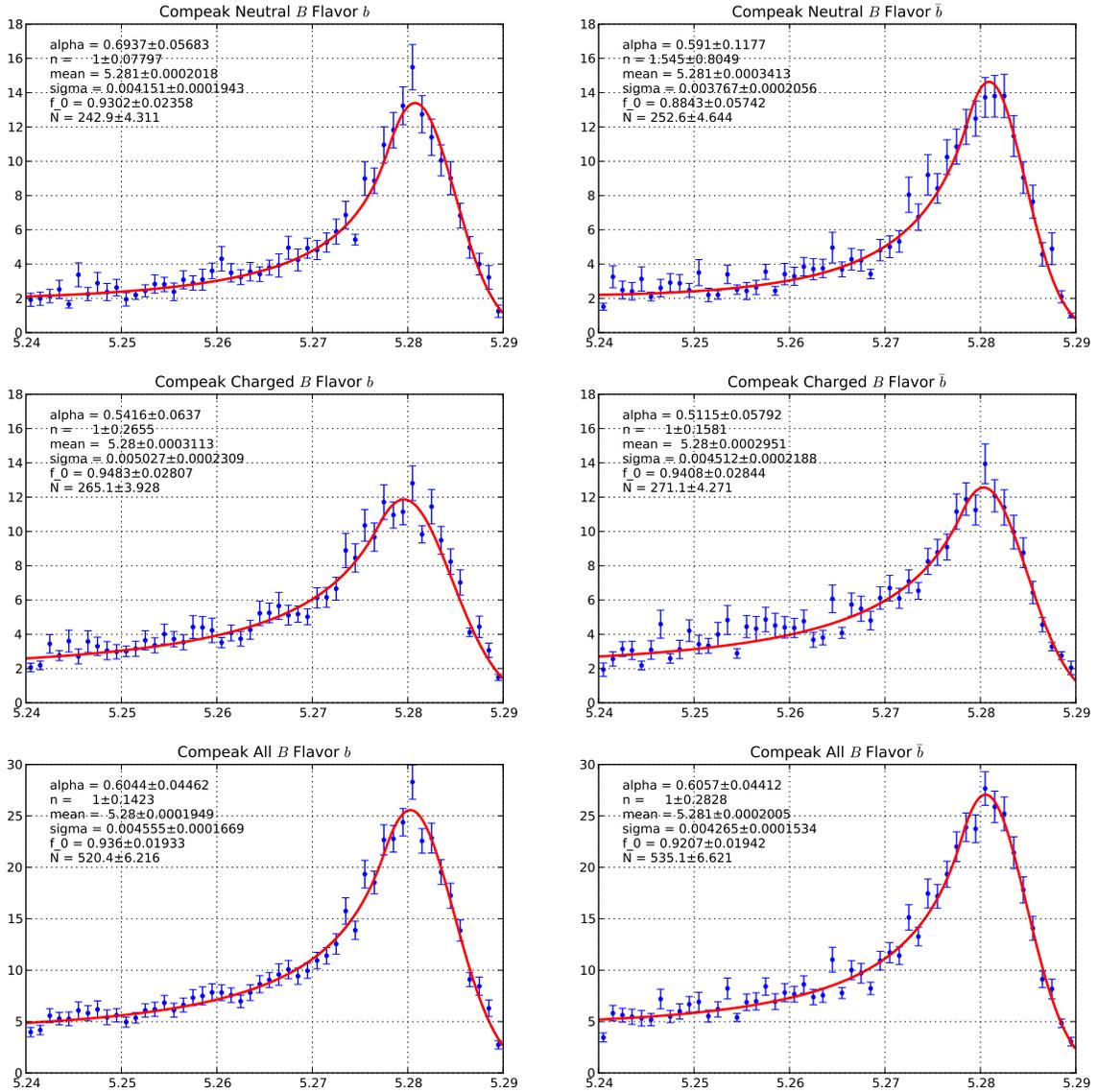


Figure H.5: Fit to MC peaking  $B\bar{B}$  and XFeed subsample for  $K^*$  mass region.

## H.2 Full Mass Range

Plots are on the next page

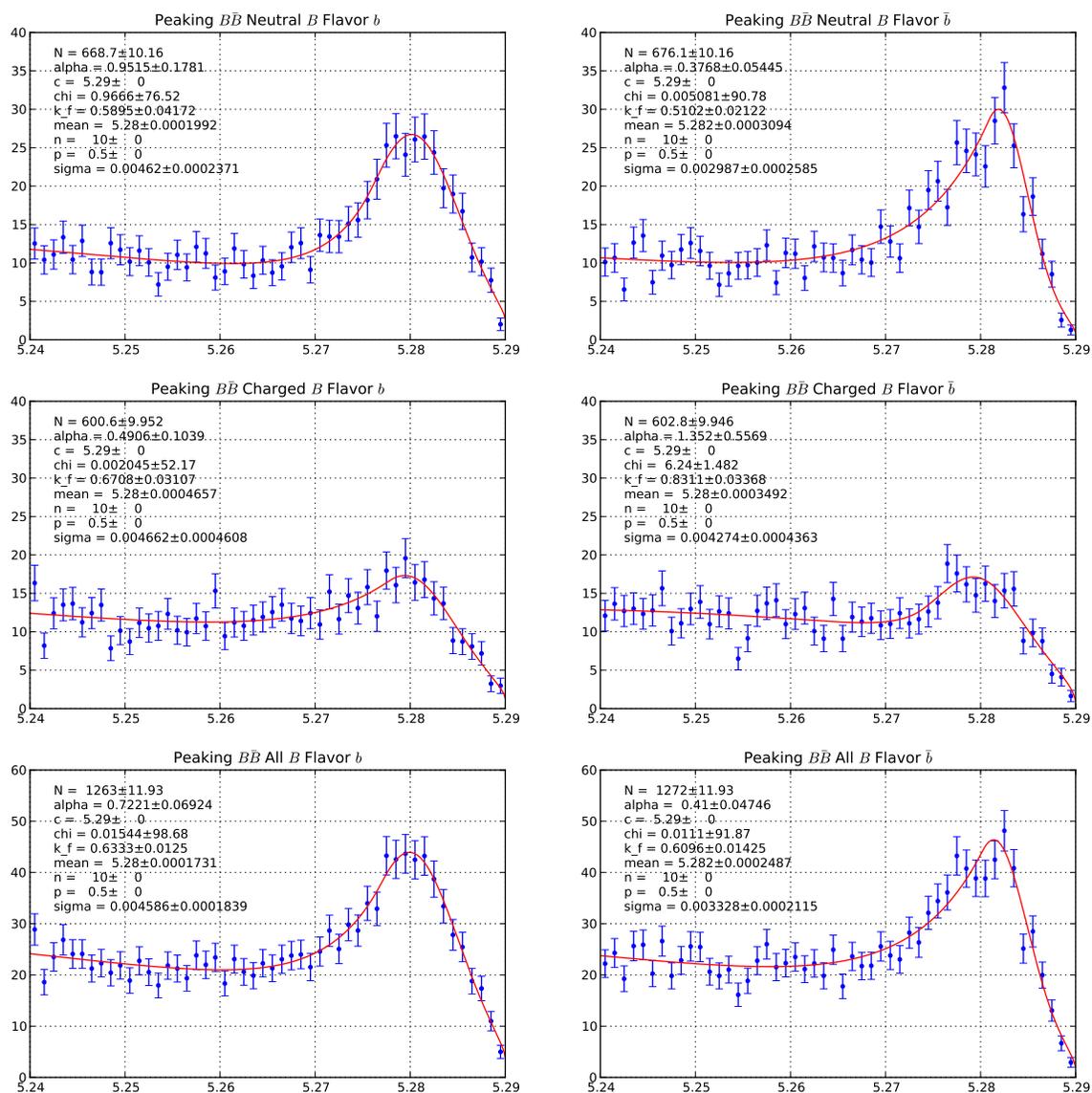


Figure H.6: Fit to MC for peaking  $B\bar{B}$  subsample.

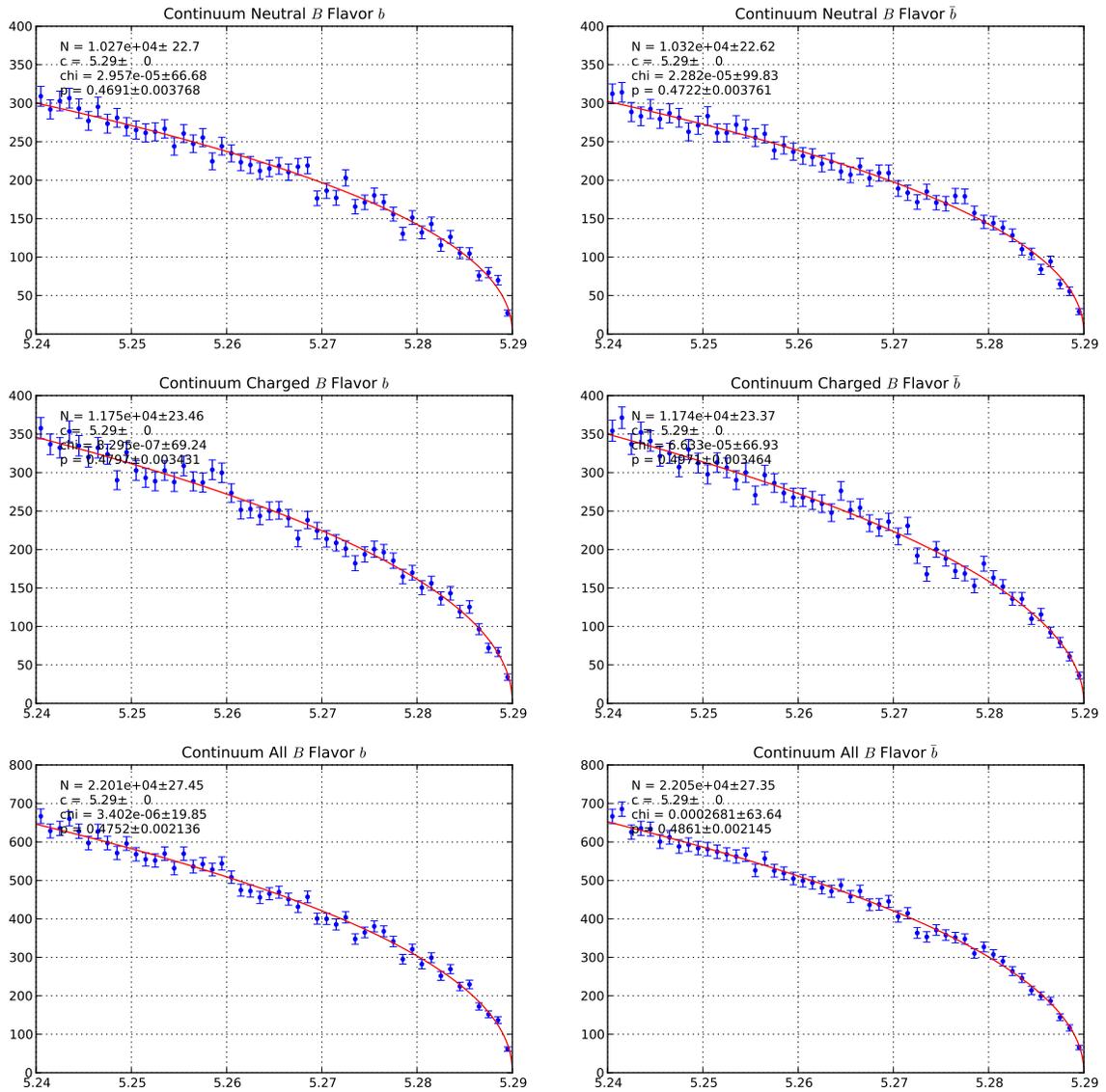


Figure H.7: Fit to MC for continuum subsample.

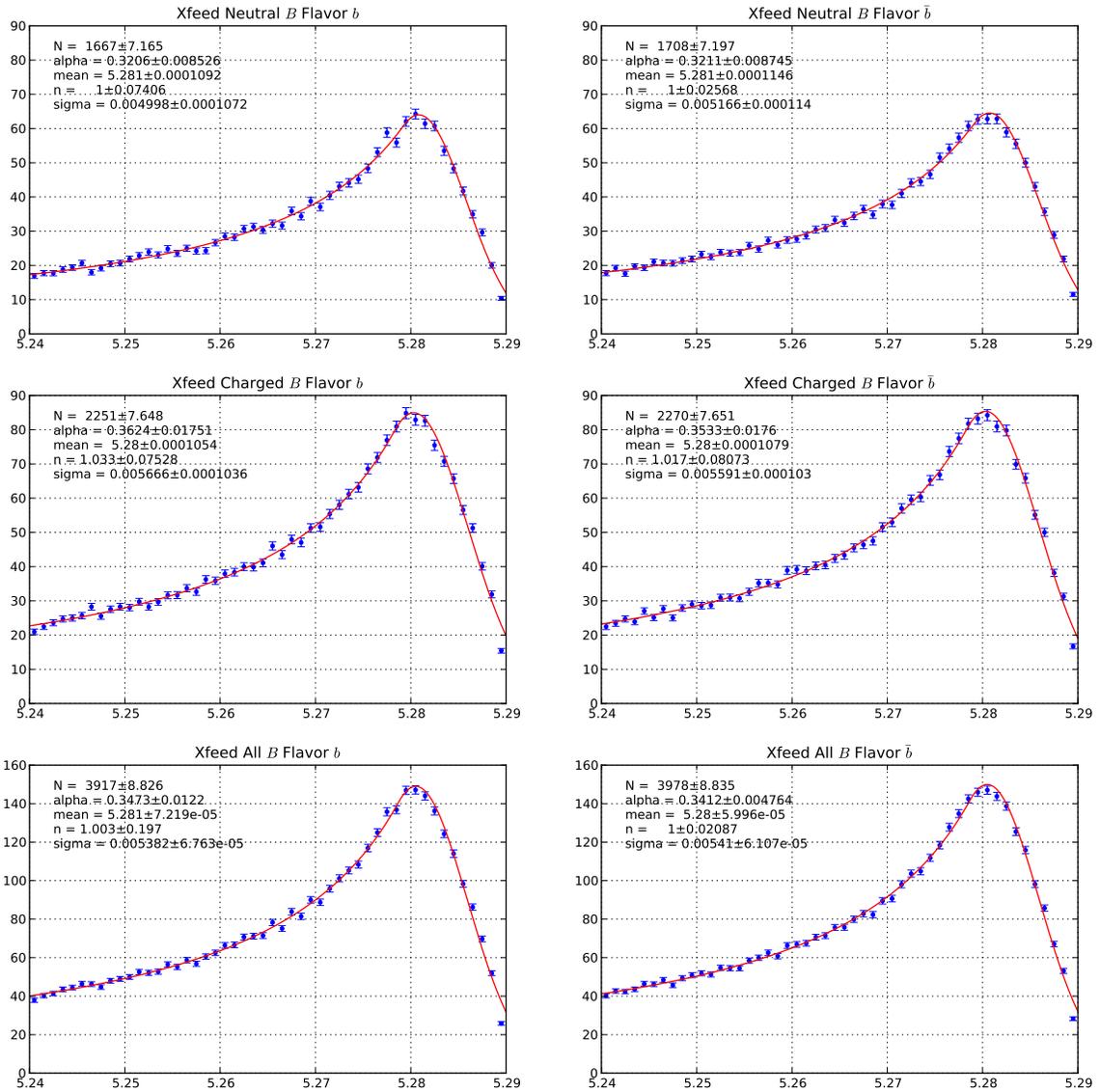


Figure H.8: Fit to MC for xfeed subsample.

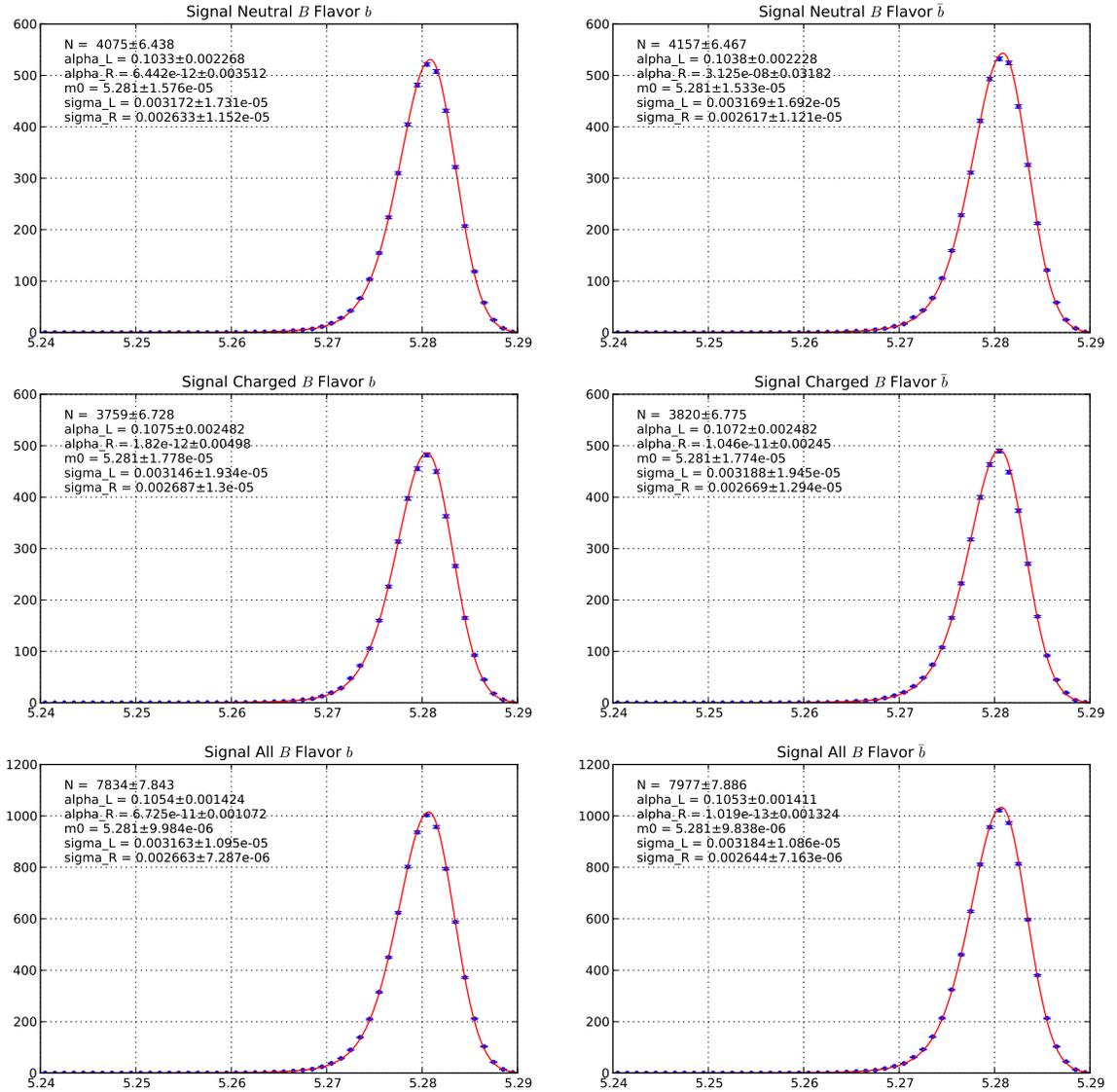
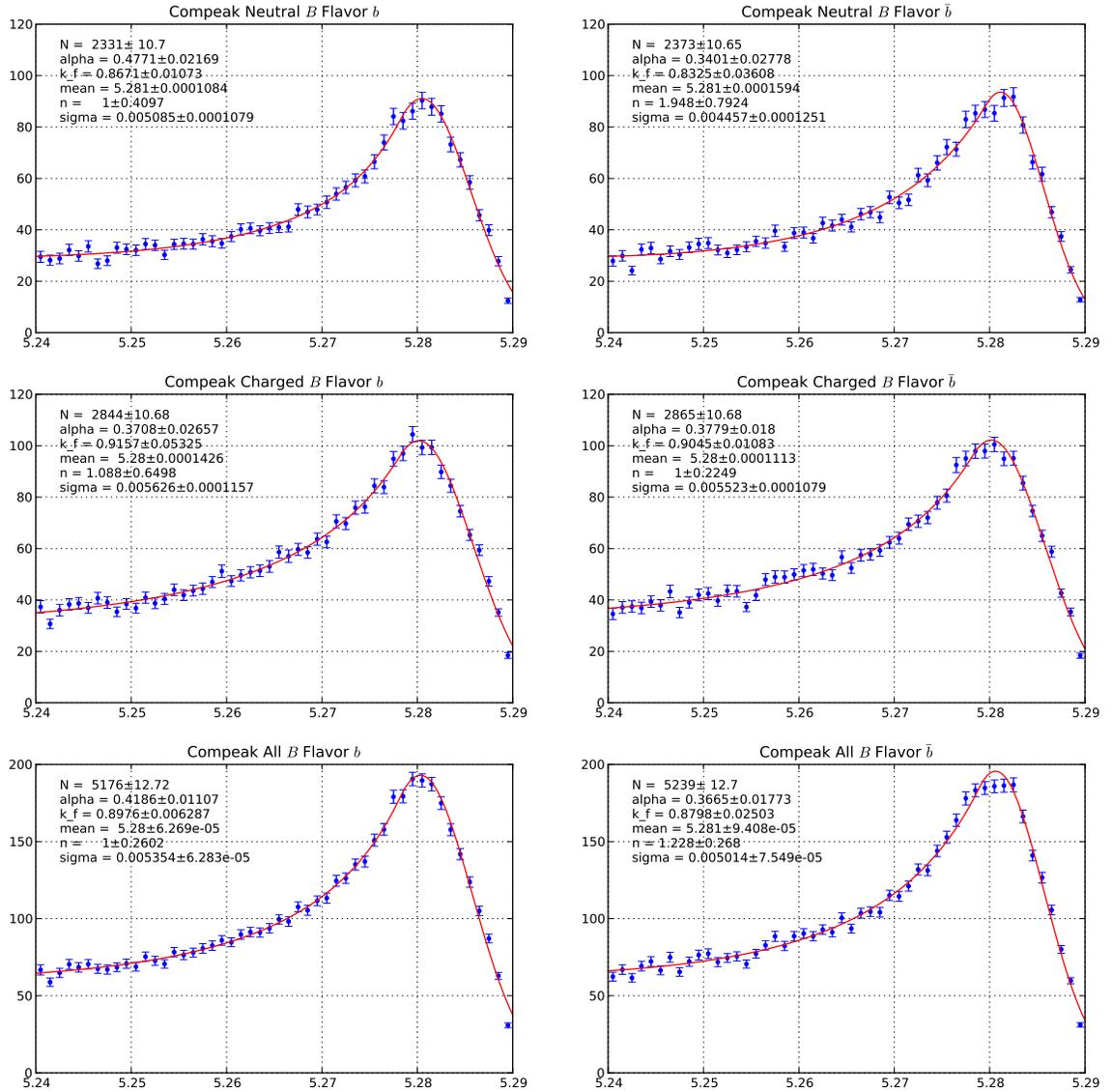


Figure H.9: Fit to MC for peaking signal subsample.

Figure H.10: Fit to MC for combined peaking  $B\bar{B}$  and XFeed subsample.

# Appendix I

## Detector $A_{CP}$ Appendix

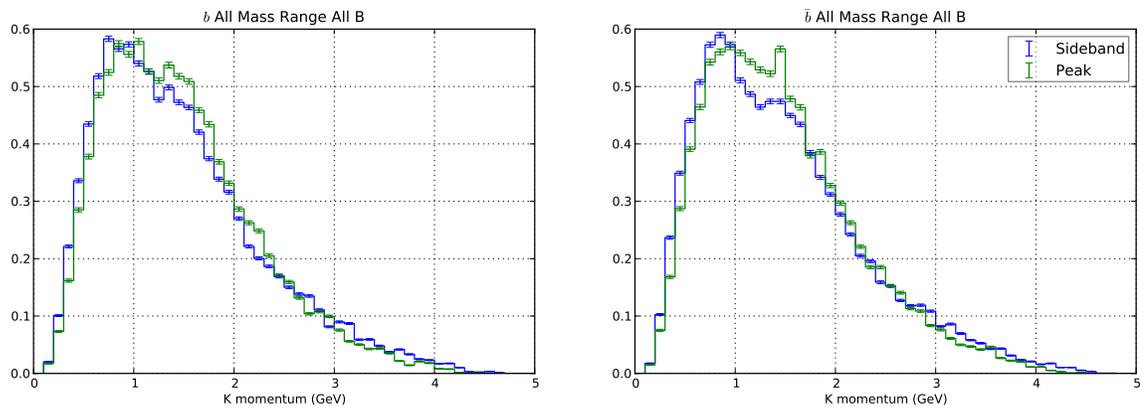


Figure I.1: Normalized charged  $K$  momentum distribution for full mass range for  $b$ (left) and  $\bar{b}$ (right) using all  $B$  sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green.

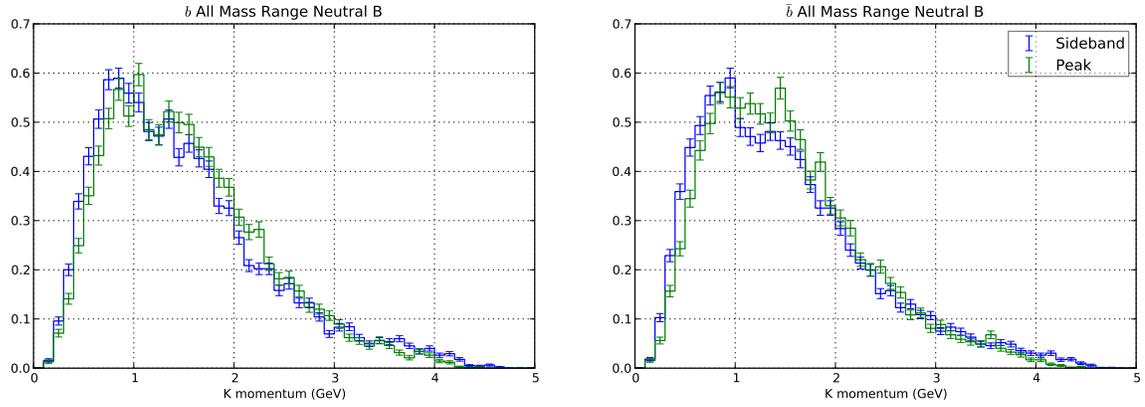


Figure I.2: Normalized charged  $K$  momentum distribution for full mass range for  $b$ (left) and  $\bar{b}$ (right) using neutral  $B$  sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green.

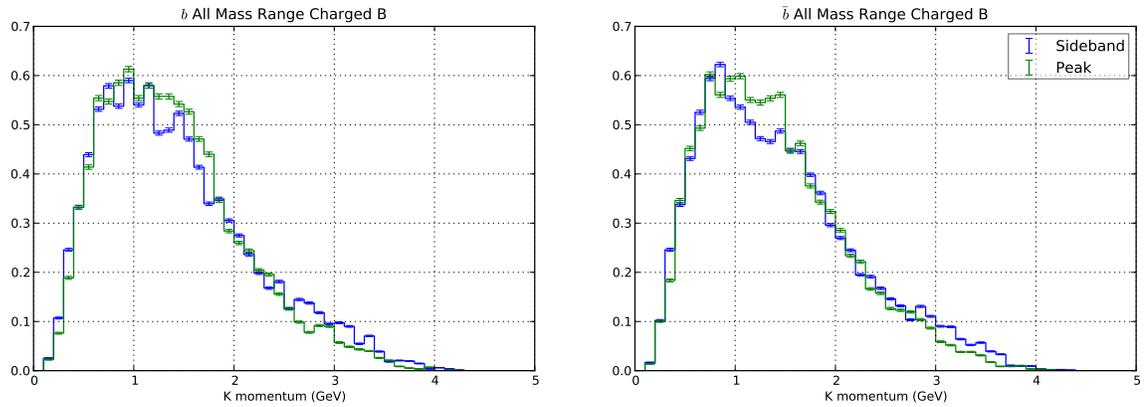


Figure I.3: Normalized charged  $K$  momentum distribution for full mass range for  $b$ (left) and  $\bar{b}$ (right) using charged  $B$  sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green.

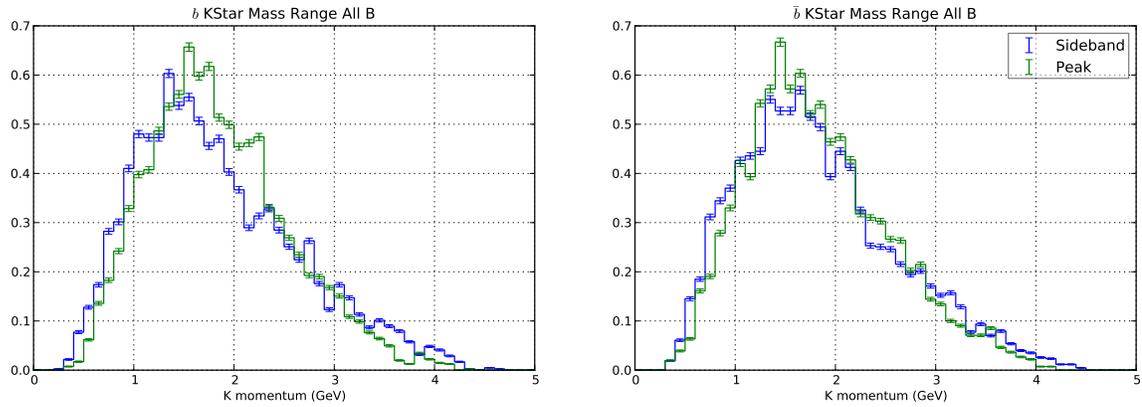


Figure I.4: Normalized charged  $K$  momentum distribution for  $K^*$  mass range for  $b$ (left) and  $\bar{b}$ (right) using all  $B$  sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green.

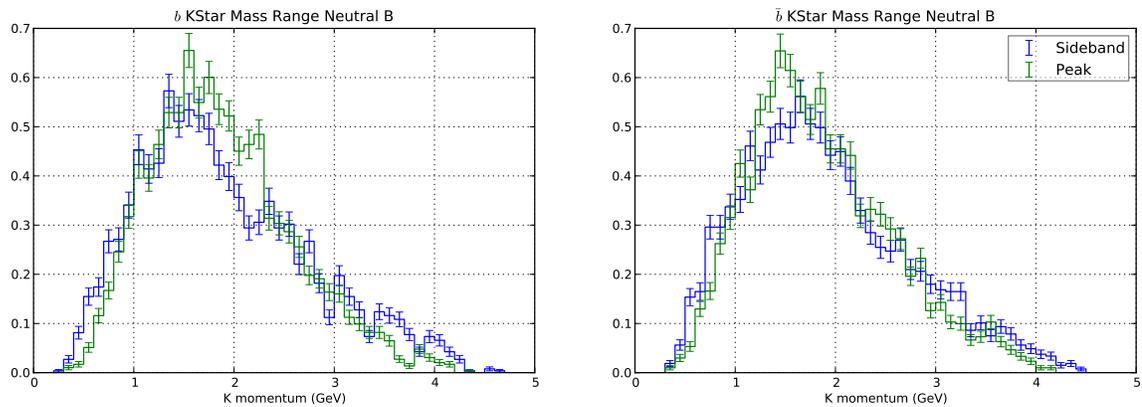


Figure I.5: Normalized charged  $K$  momentum distribution for  $K^*$  mass range for  $b$ (left) and  $\bar{b}$ (right) using neutral  $B$  sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green.

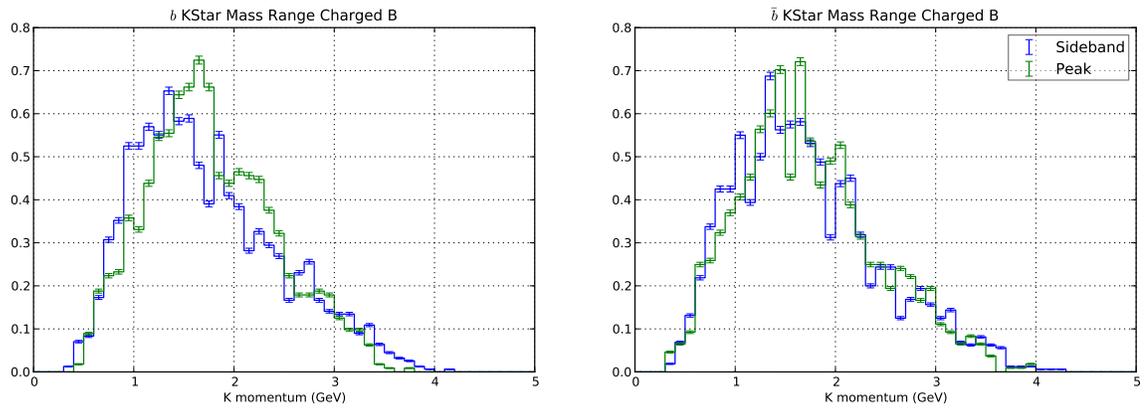


Figure I.6: Normalized charged  $K$  momentum distribution for  $K^*$  mass range for  $b$ (left) and  $\bar{b}$ (right) using charged  $B$  sample. Distribution from the sideband region is shown in blue while distribution from peaking region is shown in green.

# Appendix J

## Unblind Results Appendix

This section contains fitted shape parameters and correlation among them. For each table, the column Para Error refers to uncertainty assuming parabolic shape at the minimum and Error+ and Error- refers to asymmetric error obtained by scanning the likelihood around the minimum.

### J.1 $K^*$ region

#### J.1.1 All $B$

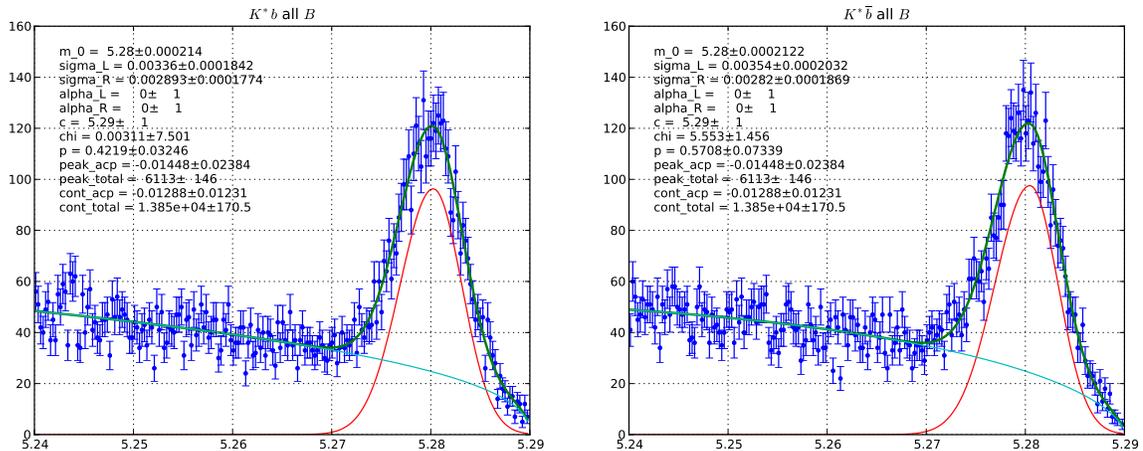


Figure J.1: Fitted  $m_{ES}$  distribution for  $K^*$  mass range using all  $B$  sample.

Table J.1: Fitted shape parameters for  $K^*$  mass range with all  $B$  sample.

	Name	Value	Para Error	Error+	Error-	Limit+	Limit-	FIXED
1	b $m_0$	5.280e+00	2.140e-04					
2	b $\sigma_L$	3.360e-03	1.842e-04					
3	b $\sigma_R$	2.893e-03	1.774e-04					
4	b $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
5	b $\alpha_R$	0.000e+00	1.000e+00					FIXED
6	b $c$	5.290e+00	1.000e+00					FIXED
7	b $\chi$	3.110e-03	7.501e+00					
8	b $p$	4.219e-01	3.246e-02					
9	$A_{\text{peak}}$	-1.448e-02	2.384e-02	-2.389e-02	2.384e-02	-1.000e+00	1.000e+00	
10	$T_{\text{peak}}$	6.113e+03	1.460e+02					
11	$A_{\text{cont}}$	-1.288e-02	1.231e-02			-1.000e+00	1.000e+00	
12	$T_{\text{cont}}$	1.385e+04	1.705e+02					
13	antib $m_0$	5.280e+00	2.122e-04					
14	antib $\sigma_L$	3.540e-03	2.032e-04					
15	antib $\sigma_R$	2.820e-03	1.869e-04					
16	antib $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
17	antib $\alpha_R$	0.000e+00	1.000e+00					FIXED
18	antib $c$	5.290e+00	1.000e+00					FIXED
19	antib $\chi$	5.553e+00	1.456e+00					
20	antib $p$	5.708e-01	7.339e-02					

Table J.2: Correlation among shape parameters for  $K^*$  mass range with all  $B$  sample.

	b $m_0$	b $\sigma_L$	b $\sigma_R$	b $\chi$	b $p$	$A_{\text{peak}}$	$T_{\text{peak}}$	$A_{\text{cont}}$	$T_{\text{cont}}$	antib $m_0$	antib $\sigma_L$	antib $\sigma_R$	antib $\chi$	antib $p$
b $m_0$	1.00	0.78	-0.80	-0.01	-0.16	-0.05	-0.05	0.04	0.04	-0.00	0.00	0.00	0.00	0.00
b $\sigma_L$	0.78	1.00	-0.47	-0.01	0.11	0.17	0.17	-0.14	-0.14	0.00	-0.00	-0.00	-0.00	-0.00
b $\sigma_R$	-0.80	-0.47	1.00	0.01	0.50	0.26	0.25	-0.22	-0.22	0.00	-0.00	-0.00	-0.00	-0.00
b $\chi$	-0.01	-0.01	0.01	1.00	0.06	-0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00
b $p$	-0.16	0.11	0.50	0.06	1.00	0.45	0.45	-0.39	-0.38	0.00	-0.00	-0.00	-0.00	-0.00
$A_{\text{peak}}$	-0.05	0.17	0.26	-0.00	0.45	1.00	-0.09	-0.61	0.07	0.03	-0.24	-0.27	-0.01	-0.28
$T_{\text{peak}}$	-0.05	0.17	0.25	-0.00	0.45	-0.09	1.00	0.07	-0.61	-0.03	0.24	0.28	0.01	0.28
$A_{\text{cont}}$	0.04	-0.14	-0.22	0.00	-0.39	-0.61	0.07	1.00	-0.06	-0.03	0.20	0.23	0.01	0.24
$T_{\text{cont}}$	0.04	-0.14	-0.22	0.00	-0.38	0.07	-0.61	-0.06	1.00	0.03	-0.21	-0.24	-0.01	-0.24
antib $m_0$	-0.00	0.00	0.00	-0.00	0.00	0.03	-0.03	-0.03	0.03	1.00	0.76	-0.80	-0.29	-0.31
antib $\sigma_L$	0.00	-0.00	-0.00	0.00	-0.00	-0.24	0.24	0.20	-0.21	0.76	1.00	-0.45	-0.34	-0.19
antib $\sigma_R$	0.00	-0.00	-0.00	0.00	-0.00	-0.27	0.28	0.23	-0.24	-0.80	-0.45	1.00	0.43	0.61
antib $\chi$	0.00	-0.00	-0.00	0.00	-0.00	-0.01	0.01	0.01	-0.01	-0.29	-0.34	0.43	1.00	0.86
antib $p$	0.00	-0.00	-0.00	0.00	-0.00	-0.28	0.28	0.24	-0.24	-0.31	-0.19	0.61	0.86	1.00

## J.1.2 Charged $B$

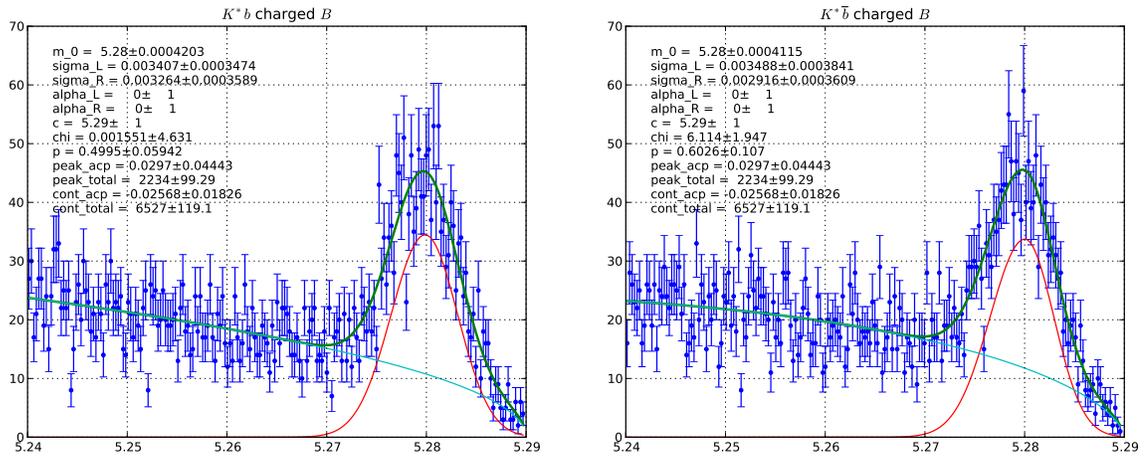


Figure J.2: Fitted  $m_{ES}$  distribution for  $K^*$  mass range using charged  $B$  sample.

Table J.3: Fitted shape parameters for  $K^*$  mass range with charged  $B$  sample.

	Name	Value	Para Error	Error+	Error-	Limit+	Limit-	FIXED
1	b $m_0$	5.280e+00	4.203e-04					
2	b $\sigma_L$	3.407e-03	3.474e-04					
3	b $\sigma_R$	3.264e-03	3.589e-04					
4	b $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
5	b $\alpha_R$	0.000e+00	1.000e+00					FIXED
6	b $c$	5.290e+00	1.000e+00					FIXED
7	b $\chi$	1.551e-03	4.631e+00					
8	b $p$	4.995e-01	5.942e-02					
9	$A_{\text{peak}}$	2.970e-02	4.443e-02	-4.455e-02	4.447e-02	-1.000e+00	1.000e+00	
10	$T_{\text{peak}}$	2.234e+03	9.929e+01					
11	$A_{\text{cont}}$	-2.568e-02	1.826e-02			-1.000e+00	1.000e+00	
12	$T_{\text{cont}}$	6.527e+03	1.191e+02					
13	antib $m_0$	5.280e+00	4.115e-04					
14	antib $\sigma_L$	3.488e-03	3.841e-04					
15	antib $\sigma_R$	2.916e-03	3.609e-04					
16	antib $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
17	antib $\alpha_R$	0.000e+00	1.000e+00					FIXED
18	antib $c$	5.290e+00	1.000e+00					FIXED
19	antib $\chi$	6.114e+00	1.947e+00					
20	antib $p$	6.026e-01	1.070e-01					

Table J.4: Correlation among shape parameters for  $K^*$  mass range with charged  $B$  sample.

	b $m_0$	b $\sigma_L$	b $\sigma_R$	b $\chi$	b $p$	$A_{\text{peak}}$	$T_{\text{peak}}$	$A_{\text{cont}}$	$T_{\text{cont}}$	antib $m_0$	antib $\sigma_L$	antib $\sigma_R$	antib $\chi$	antib $p$
b $m_0$	1.00	0.80	-0.82	-0.00	-0.21	-0.08	-0.08	0.07	0.07	-0.00	0.00	0.00	0.00	0.00
b $\sigma_L$	0.80	1.00	-0.50	-0.00	0.09	0.15	0.16	-0.14	-0.13	0.00	-0.00	-0.00	-0.00	-0.00
b $\sigma_R$	-0.82	-0.50	1.00	0.01	0.57	0.31	0.32	-0.28	-0.27	0.00	-0.00	-0.00	-0.00	-0.00
b $\chi$	-0.00	-0.00	0.01	1.00	0.02	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00
b $p$	-0.21	0.09	0.57	0.02	1.00	0.53	0.54	-0.46	-0.45	0.00	-0.00	-0.00	-0.00	-0.00
$A_{\text{peak}}$	-0.08	0.15	0.31	0.00	0.53	1.00	-0.01	-0.65	0.01	0.04	-0.21	-0.26	0.00	-0.26
$T_{\text{peak}}$	-0.08	0.16	0.32	0.00	0.54	-0.01	1.00	-0.03	-0.65	-0.04	0.21	0.26	-0.00	0.25
$A_{\text{cont}}$	0.07	-0.14	-0.28	-0.00	-0.46	-0.65	-0.03	1.00	0.02	-0.03	0.17	0.21	0.00	0.21
$T_{\text{cont}}$	0.07	-0.13	-0.27	-0.00	-0.45	0.01	-0.65	0.02	1.00	0.03	-0.17	-0.22	-0.00	-0.21
antib $m_0$	-0.00	0.00	0.00	0.00	0.00	0.04	-0.04	-0.03	0.03	1.00	0.79	-0.82	-0.30	-0.33
antib $\sigma_L$	0.00	-0.00	-0.00	-0.00	-0.00	-0.21	0.21	0.17	-0.17	0.79	1.00	-0.52	-0.36	-0.23
antib $\sigma_R$	0.00	-0.00	-0.00	-0.00	-0.00	-0.26	0.26	0.21	-0.22	-0.82	-0.52	1.00	0.42	0.60
antib $\chi$	0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.30	-0.36	0.42	1.00	0.87
antib $p$	0.00	-0.00	-0.00	-0.00	-0.00	-0.26	0.25	0.21	-0.21	-0.33	-0.23	0.60	0.87	1.00

### J.1.3 Neutral $B$

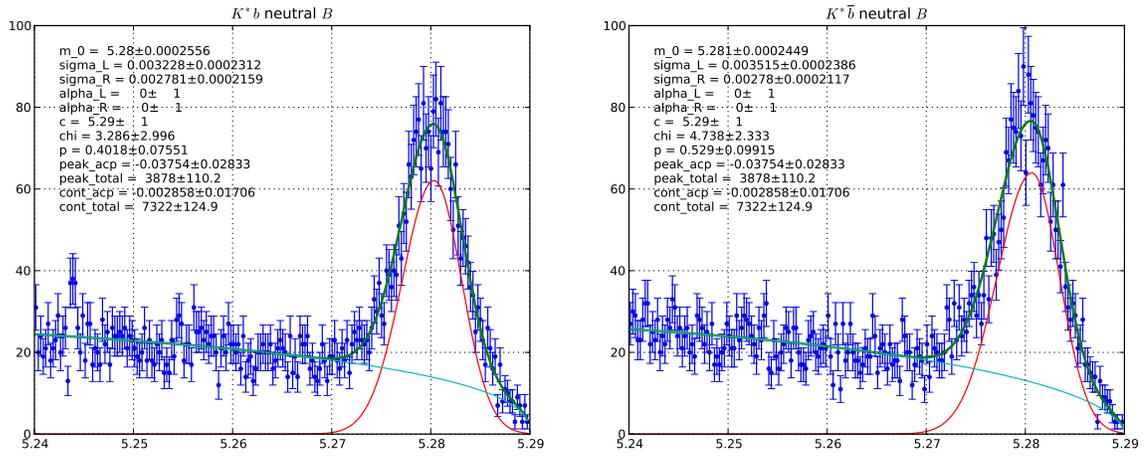


Figure J.3: Fitted  $m_{ES}$  distribution for  $K^*$  mass range using neutral  $B$  sample.

Table J.5: Fitted shape parameters for  $K^*$  mass range with neutral  $B$  sample.

	Name	Value	Para Error	Error+	Error-	Limit+	Limit-	FIXED
1	b $m_0$	5.280e+00	2.556e-04					
2	b $\sigma_L$	3.228e-03	2.312e-04					
3	b $\sigma_R$	2.781e-03	2.159e-04					
4	b $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
5	b $\alpha_R$	0.000e+00	1.000e+00					FIXED
6	b $c$	5.290e+00	1.000e+00					FIXED
7	b $\chi$	3.286e+00	2.996e+00					
8	b $p$	4.018e-01	7.551e-02					
9	$A_{\text{peak}}$	-3.754e-02	2.833e-02	-2.840e-02	2.831e-02	-1.000e+00	1.000e+00	
10	$T_{\text{peak}}$	3.878e+03	1.102e+02					
11	$A_{\text{cont}}$	-2.858e-03	1.706e-02			-1.000e+00	1.000e+00	
12	$T_{\text{cont}}$	7.322e+03	1.249e+02					
13	antib $m_0$	5.281e+00	2.449e-04					
14	antib $\sigma_L$	3.515e-03	2.386e-04					
15	antib $\sigma_R$	2.780e-03	2.117e-04					
16	antib $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
17	antib $\alpha_R$	0.000e+00	1.000e+00					FIXED
18	antib $c$	5.290e+00	1.000e+00					FIXED
19	antib $\chi$	4.738e+00	2.333e+00					
20	antib $p$	5.290e-01	9.915e-02					

Table J.6: Correlation among shape parameters for  $K^*$  mass range with neutral  $B$  sample.

	b $m_0$	b $\sigma_L$	b $\sigma_R$	b $\chi$	b $p$	$A_{\text{peak}}$	$T_{\text{peak}}$	$A_{\text{cont}}$	$T_{\text{cont}}$	antib $m_0$	antib $\sigma_L$	antib $\sigma_R$	antib $\chi$	antib $p$
b $m_0$	1.00	0.79	-0.81	-0.24	-0.28	-0.01	-0.01	0.01	0.01	-0.00	0.00	0.00	0.00	0.00
b $\sigma_L$	0.79	1.00	-0.51	-0.35	-0.23	0.21	0.20	-0.17	-0.17	0.00	-0.00	-0.00	-0.00	-0.00
b $\sigma_R$	-0.81	-0.51	1.00	0.29	0.50	0.20	0.20	-0.17	-0.17	0.00	-0.00	-0.00	-0.00	-0.00
b $\chi$	-0.24	-0.35	0.29	1.00	0.84	-0.12	-0.11	0.10	0.10	-0.00	0.00	0.00	0.00	0.00
b $p$	-0.28	-0.23	0.50	0.84	1.00	0.14	0.14	-0.12	-0.12	0.00	-0.00	-0.00	-0.00	-0.00
$A_{\text{peak}}$	-0.01	0.21	0.20	-0.12	0.14	1.00	-0.07	-0.60	0.06	0.01	-0.25	-0.25	0.00	-0.26
$T_{\text{peak}}$	-0.01	0.20	0.20	-0.11	0.14	-0.07	1.00	0.08	-0.60	-0.01	0.26	0.26	-0.00	0.27
$A_{\text{cont}}$	0.01	-0.17	-0.17	0.10	-0.12	-0.60	0.08	1.00	-0.07	-0.01	0.23	0.23	-0.00	0.24
$T_{\text{cont}}$	0.01	-0.17	-0.17	0.10	-0.12	0.06	-0.60	-0.07	1.00	0.01	-0.23	-0.23	0.00	-0.24
antib $m_0$	-0.00	0.00	0.00	-0.00	0.00	0.01	-0.01	-0.01	0.01	1.00	0.75	-0.79	-0.28	-0.29
antib $\sigma_L$	0.00	-0.00	-0.00	0.00	-0.00	-0.25	0.26	0.23	-0.23	0.75	1.00	-0.43	-0.34	-0.17
antib $\sigma_R$	0.00	-0.00	-0.00	0.00	-0.00	-0.25	0.26	0.23	-0.23	-0.79	-0.43	1.00	0.41	0.60
antib $\chi$	0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.28	-0.34	0.41	1.00	0.85
antib $p$	0.00	-0.00	-0.00	0.00	-0.00	-0.26	0.27	0.24	-0.24	-0.29	-0.17	0.60	0.85	1.00

## J.2 Full Mass Range

### J.2.1 All $B$

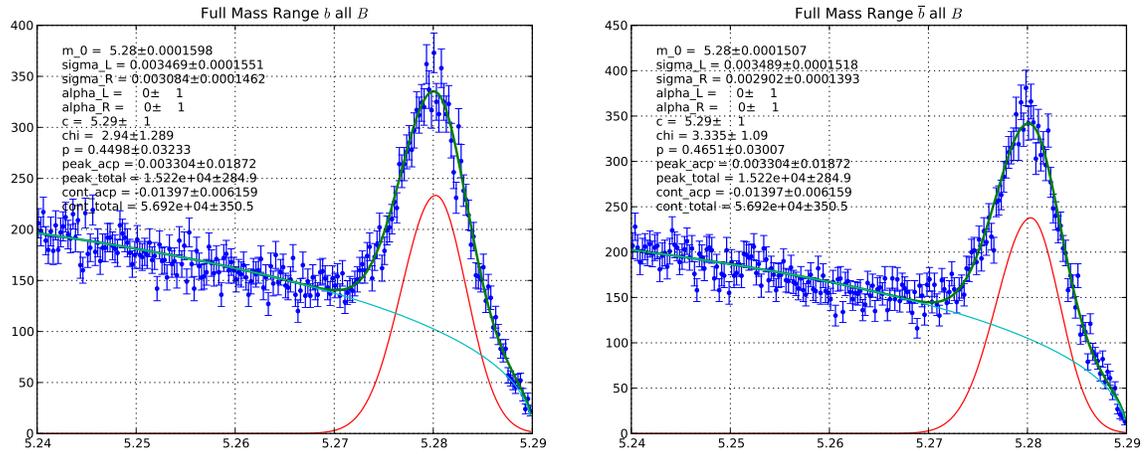


Figure J.4: Fitted  $m_{ES}$  distribution for full mass range using all  $B$  sample.

Table J.7: Fitted shape parameters for full mass range with all  $B$  sample.

	Name	Value	Para Error	Error+	Error-	Limit+	Limit-	FIXED
1	b $m_0$	5.280e+00	1.598e-04					
2	b $\sigma_L$	3.469e-03	1.551e-04					
3	b $\sigma_R$	3.084e-03	1.462e-04					
4	b $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
5	b $\alpha_R$	0.000e+00	1.000e+00					FIXED
6	b $c$	5.290e+00	1.000e+00					FIXED
7	b $\chi$	2.940e+00	1.289e+00					
8	b $p$	4.498e-01	3.233e-02					
9	$A_{\text{peak}}$	3.304e-03	1.872e-02	-1.872e-02	1.874e-02	-1.000e+00	1.000e+00	
10	$T_{\text{peak}}$	1.522e+04	2.849e+02					
11	$A_{\text{cont}}$	-1.397e-02	6.159e-03			-1.000e+00	1.000e+00	
12	$T_{\text{cont}}$	5.692e+04	3.505e+02					
13	antib $m_0$	5.280e+00	1.507e-04					
14	antib $\sigma_L$	3.489e-03	1.518e-04					
15	antib $\sigma_R$	2.902e-03	1.393e-04					
16	antib $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
17	antib $\alpha_R$	0.000e+00	1.000e+00					FIXED
18	antib $c$	5.290e+00	1.000e+00					FIXED
19	antib $\chi$	3.335e+00	1.090e+00					
20	antib $p$	4.651e-01	3.007e-02					

Table J.8: Correlation among shape parameters for full mass range with all  $B$  sample.

	b $m_0$	b $\sigma_L$	b $\sigma_R$	b $\chi$	b $p$	$A_{\text{peak}}$	$T_{\text{peak}}$	$A_{\text{cont}}$	$T_{\text{cont}}$	antib $m_0$	antib $\sigma_L$	antib $\sigma_R$	antib $\chi$	antib $p$
b $m_0$	1.00	0.78	-0.79	-0.28	-0.31	-0.02	-0.02	0.01	0.01	-0.00	0.00	0.00	-0.00	0.00
b $\sigma_L$	0.78	1.00	-0.47	-0.37	-0.22	0.24	0.24	-0.20	-0.20	0.00	-0.00	-0.00	0.00	-0.00
b $\sigma_R$	-0.79	-0.47	1.00	0.40	0.61	0.27	0.27	-0.22	-0.22	0.00	-0.00	-0.00	0.00	-0.00
b $\chi$	-0.28	-0.37	0.40	1.00	0.84	-0.04	-0.04	0.04	0.04	-0.00	0.00	0.00	-0.00	0.00
b $p$	-0.31	-0.22	0.61	0.84	1.00	0.25	0.25	-0.21	-0.20	0.00	-0.00	-0.00	0.00	-0.00
$A_{\text{peak}}$	-0.02	0.24	0.27	-0.04	0.25	1.00	0.00	-0.66	-0.00	0.02	-0.26	-0.27	0.09	-0.22
$T_{\text{peak}}$	-0.02	0.24	0.27	-0.04	0.25	0.00	1.00	-0.01	-0.66	-0.02	0.26	0.27	-0.09	0.22
$A_{\text{cont}}$	0.01	-0.20	-0.22	0.04	-0.21	-0.66	-0.01	1.00	0.01	-0.02	0.21	0.22	-0.07	0.17
$T_{\text{cont}}$	0.01	-0.20	-0.22	0.04	-0.20	-0.00	-0.66	0.01	1.00	0.02	-0.21	-0.22	0.07	-0.18
antib $m_0$	-0.00	0.00	0.00	-0.00	0.00	0.02	-0.02	-0.02	0.02	1.00	0.75	-0.78	-0.26	-0.30
antib $\sigma_L$	0.00	-0.00	-0.00	0.00	-0.00	-0.26	0.26	0.21	-0.21	0.75	1.00	-0.43	-0.37	-0.21
antib $\sigma_R$	0.00	-0.00	-0.00	0.00	-0.00	-0.27	0.27	0.22	-0.22	-0.78	-0.43	1.00	0.35	0.58
antib $\chi$	-0.00	0.00	0.00	-0.00	0.00	0.09	-0.09	-0.07	0.07	-0.26	-0.37	0.35	1.00	0.83
antib $p$	0.00	-0.00	-0.00	0.00	-0.00	-0.22	0.22	0.17	-0.18	-0.30	-0.21	0.58	0.83	1.00

## J.2.2 Charged $B$

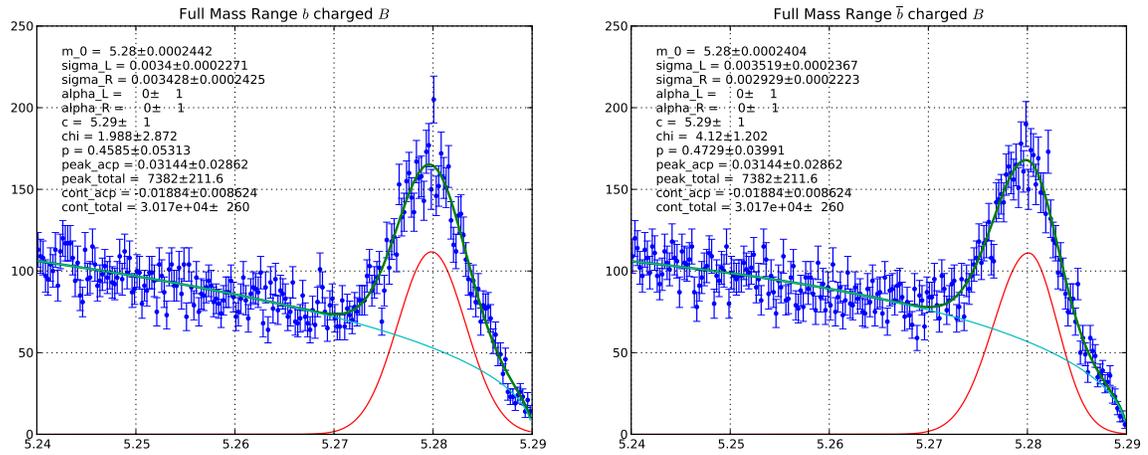


Figure J.5: Fitted  $m_{ES}$  distribution for full mass range using charged  $B$  sample.

Table J.9: Fitted shape parameters for full mass range with charged  $B$  sample.

	Name	Value	Para Error	Error+	Error-	Limit+	Limit-	FIXED
1	b $m_0$	5.280e+00	4.203e-04					
2	b $\sigma_L$	3.407e-03	3.474e-04					
3	b $\sigma_R$	3.264e-03	3.589e-04					
4	b $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
5	b $\alpha_R$	0.000e+00	1.000e+00					FIXED
6	b $c$	5.290e+00	1.000e+00					FIXED
7	b $\chi$	1.551e-03	4.631e+00					
8	b $p$	4.995e-01	5.942e-02					
9	$A_{\text{peak}}$	2.970e-02	4.443e-02	-4.455e-02	4.447e-02	-1.000e+00	1.000e+00	
10	$T_{\text{peak}}$	2.234e+03	9.929e+01					
11	$A_{\text{cont}}$	-2.568e-02	1.826e-02			-1.000e+00	1.000e+00	
12	$T_{\text{cont}}$	6.527e+03	1.191e+02					
13	antib $m_0$	5.280e+00	4.115e-04					
14	antib $\sigma_L$	3.488e-03	3.841e-04					
15	antib $\sigma_R$	2.916e-03	3.609e-04					
16	antib $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
17	antib $\alpha_R$	0.000e+00	1.000e+00					FIXED
18	antib $c$	5.290e+00	1.000e+00					FIXED
19	antib $\chi$	6.114e+00	1.947e+00					
20	antib $p$	6.026e-01	1.070e-01					

Table J.10: Correlation among shape parameters for full mass range with charged  $B$  sample.

	b $m_0$	b $\sigma_L$	b $\sigma_R$	b $\chi$	b $p$	$A_{\text{peak}}$	$T_{\text{peak}}$	$A_{\text{cont}}$	$T_{\text{cont}}$	antib $m_0$	antib $\sigma_L$	antib $\sigma_R$	antib $\chi$	antib $p$
b $m_0$	1.00	0.79	-0.77	-0.32	-0.33	-0.05	-0.05	0.04	0.04	-0.00	0.00	0.00	-0.00	0.00
b $\sigma_L$	0.79	1.00	-0.47	-0.35	-0.21	0.20	0.21	-0.17	-0.17	0.00	-0.00	-0.00	0.00	-0.00
b $\sigma_R$	-0.77	-0.47	1.00	0.51	0.69	0.33	0.34	-0.28	-0.28	0.00	-0.00	-0.00	0.00	-0.00
b $\chi$	-0.32	-0.35	0.51	1.00	0.87	0.07	0.07	-0.06	-0.06	0.00	-0.00	-0.00	0.00	-0.00
b $p$	-0.33	-0.21	0.69	0.87	1.00	0.34	0.35	-0.29	-0.28	0.00	-0.00	-0.00	0.00	-0.00
$A_{\text{peak}}$	-0.05	0.20	0.33	0.07	0.34	1.00	0.04	-0.68	-0.03	0.03	-0.24	-0.27	0.11	-0.19
$T_{\text{peak}}$	-0.05	0.21	0.34	0.07	0.35	0.04	1.00	-0.07	-0.68	-0.03	0.24	0.26	-0.11	0.18
$A_{\text{cont}}$	0.04	-0.17	-0.28	-0.06	-0.29	-0.68	-0.07	1.00	0.05	-0.02	0.19	0.21	-0.09	0.15
$T_{\text{cont}}$	0.04	-0.17	-0.28	-0.06	-0.28	-0.03	-0.68	0.05	1.00	0.03	-0.19	-0.21	0.09	-0.15
antib $m_0$	-0.00	0.00	0.00	0.00	0.00	0.03	-0.03	-0.02	0.03	1.00	0.77	-0.80	-0.27	-0.32
antib $\sigma_L$	0.00	-0.00	-0.00	-0.00	-0.00	-0.24	0.24	0.19	-0.19	0.77	1.00	-0.47	-0.38	-0.25
antib $\sigma_R$	0.00	-0.00	-0.00	-0.00	-0.00	-0.27	0.26	0.21	-0.21	-0.80	-0.47	1.00	0.33	0.57
antib $\chi$	-0.00	0.00	0.00	0.00	0.00	0.11	-0.11	-0.09	0.09	-0.27	-0.38	0.33	1.00	0.83
antib $p$	0.00	-0.00	-0.00	-0.00	-0.00	-0.19	0.18	0.15	-0.15	-0.32	-0.25	0.57	0.83	1.00

### J.2.3 Neutral $B$

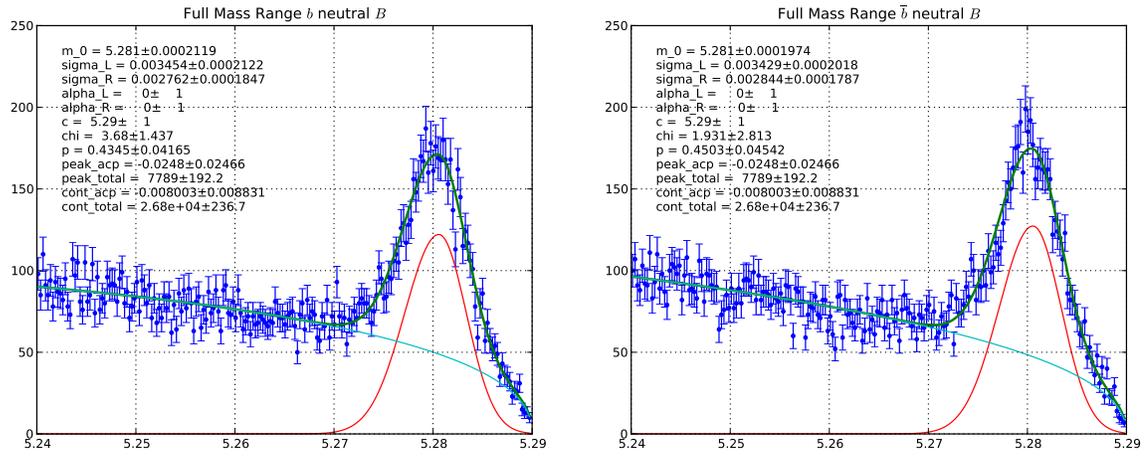


Figure J.6: Fitted  $m_{ES}$  distribution for full mass range using neutral  $B$  sample.

Table J.11: Fitted shape parameters for full mass range with neutral  $B$  sample.

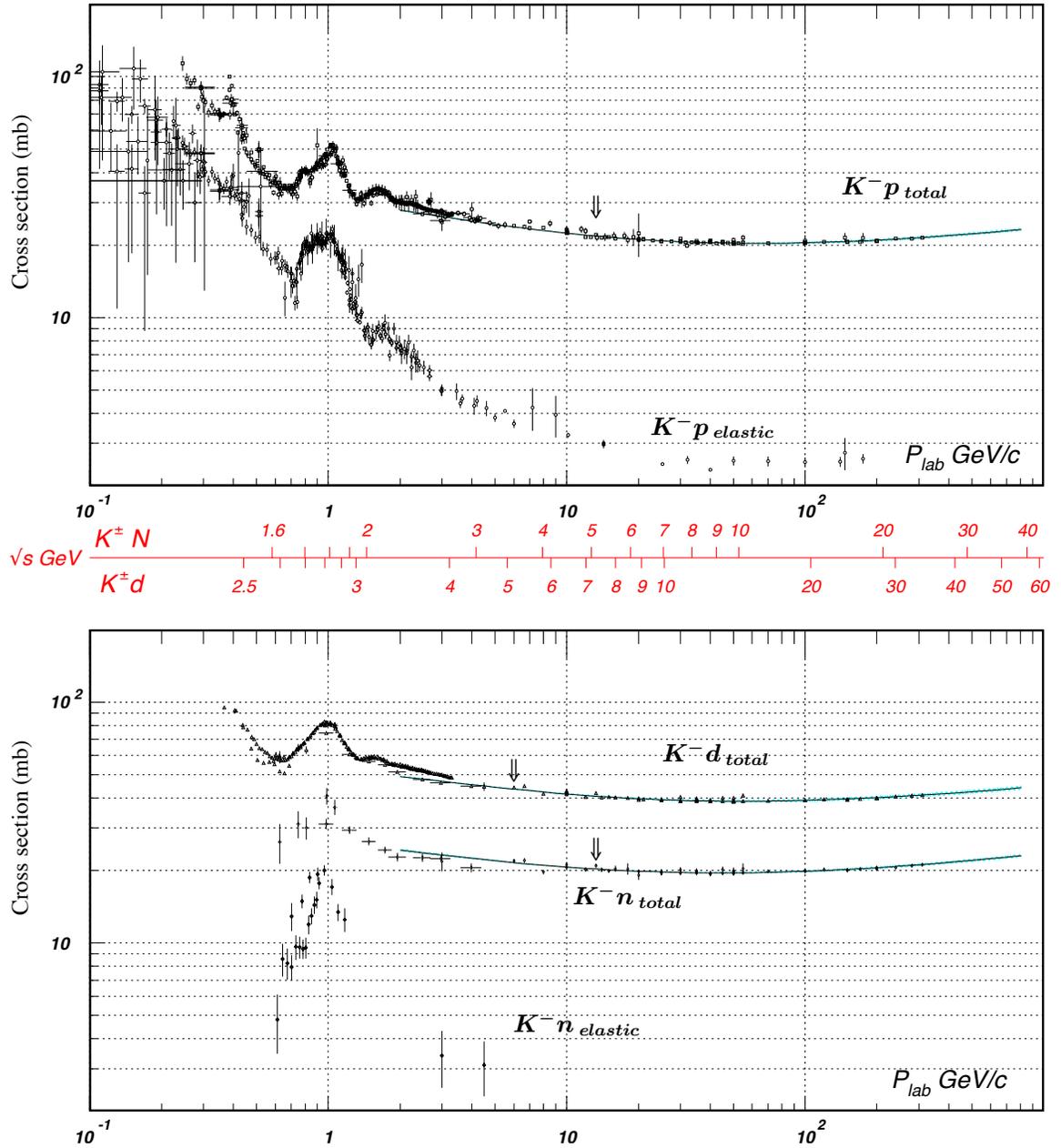
	Name	Value	Para Error	Error+	Error-	Limit+	Limit-	FIXED
1	b $m_0$	5.281e+00	2.119e-04					
2	b $\sigma_L$	3.454e-03	2.122e-04					
3	b $\sigma_R$	2.762e-03	1.847e-04					
4	b $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
5	b $\alpha_R$	0.000e+00	1.000e+00					FIXED
6	b $c$	5.290e+00	1.000e+00					FIXED
7	b $\chi$	3.680e+00	1.437e+00					
8	b $p$	4.345e-01	4.165e-02					
9	$A_{\text{peak}}$	-2.480e-02	2.466e-02	-2.467e-02	2.469e-02	-1.000e+00	1.000e+00	
10	$T_{\text{peak}}$	7.789e+03	1.922e+02					
11	$A_{\text{cont}}$	-8.003e-03	8.831e-03			-1.000e+00	1.000e+00	
12	$T_{\text{cont}}$	2.680e+04	2.367e+02					
13	antib $m_0$	5.281e+00	1.974e-04					
14	antib $\sigma_L$	3.429e-03	2.018e-04					
15	antib $\sigma_R$	2.844e-03	1.787e-04					
16	antib $\alpha_L$	0.000e+00	1.000e+00			1.000e-06	2.000e+00	FIXED
17	antib $\alpha_R$	0.000e+00	1.000e+00					FIXED
18	antib $c$	5.290e+00	1.000e+00					FIXED
19	antib $\chi$	1.931e+00	2.813e+00					
20	antib $p$	4.503e-01	4.542e-02					

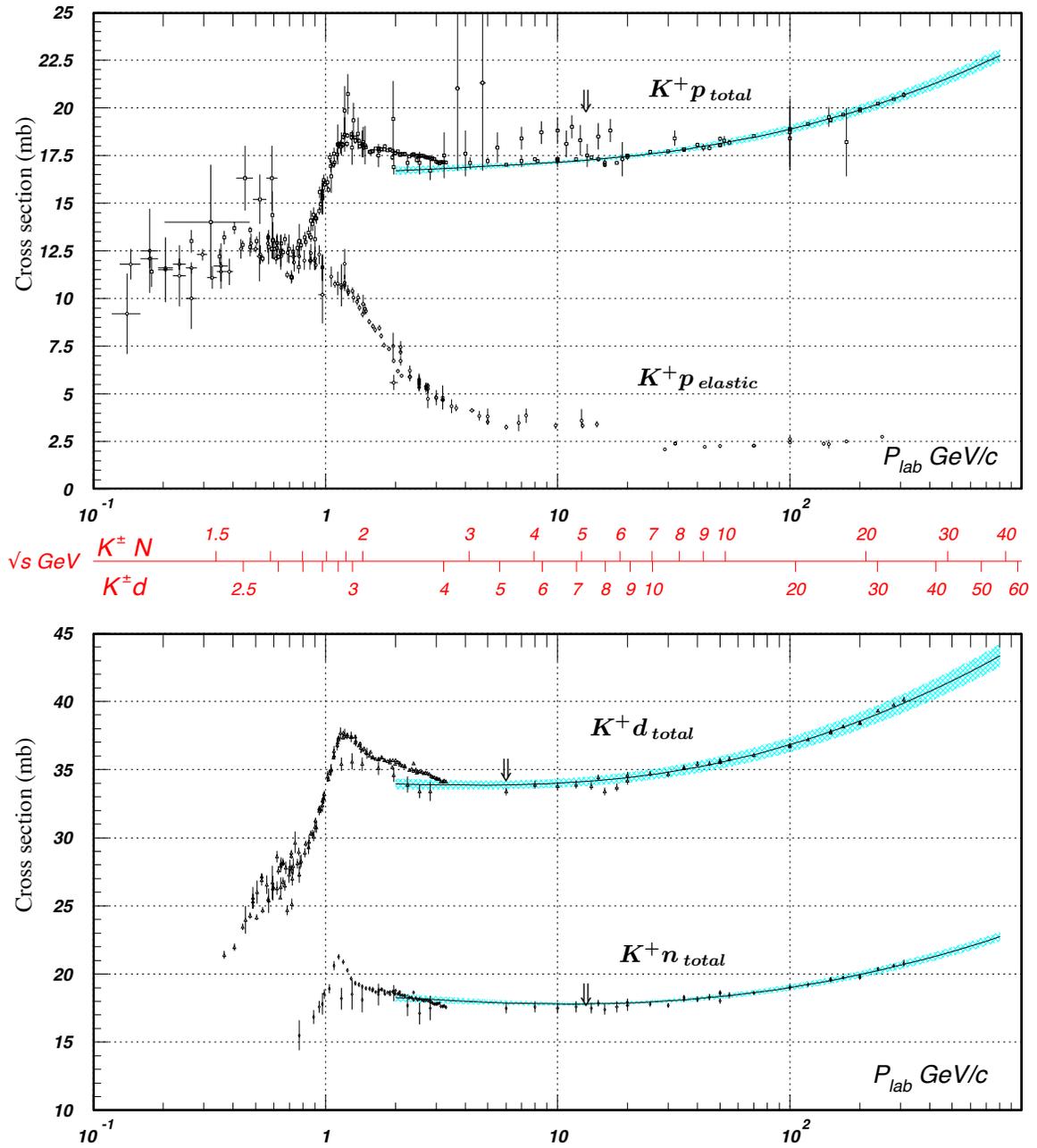
Table J.12: Correlation among shape parameters for full mass range with neutral  $B$  sample.

	b $m_0$	b $\sigma_L$	b $\sigma_R$	b $\chi$	b $p$	$A_{\text{peak}}$	$T_{\text{peak}}$	$A_{\text{cont}}$	$T_{\text{cont}}$	antib $m_0$	antib $\sigma_L$	antib $\sigma_R$	antib $\chi$	antib $p$
b $m_0$	1.00	0.77	-0.81	-0.27	-0.30	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00
b $\sigma_L$	0.77	1.00	-0.49	-0.38	-0.24	0.26	0.26	-0.21	-0.21	0.00	-0.00	-0.00	0.00	-0.00
b $\sigma_R$	-0.81	-0.49	1.00	0.32	0.54	0.23	0.22	-0.18	-0.18	0.00	-0.00	-0.00	0.00	-0.00
b $\chi$	-0.27	-0.38	0.32	1.00	0.84	-0.12	-0.12	0.09	0.09	-0.00	0.00	0.00	-0.00	0.00
b $p$	-0.30	-0.24	0.54	0.84	1.00	0.17	0.17	-0.14	-0.14	0.00	-0.00	-0.00	0.00	-0.00
$A_{\text{peak}}$	0.00	0.26	0.23	-0.12	0.17	1.00	-0.01	-0.64	0.01	0.00	-0.27	-0.26	0.07	-0.22
$T_{\text{peak}}$	0.00	0.26	0.22	-0.12	0.17	-0.01	1.00	0.02	-0.64	-0.00	0.28	0.26	-0.07	0.23
$A_{\text{cont}}$	-0.00	-0.21	-0.18	0.09	-0.14	-0.64	0.02	1.00	-0.02	-0.00	0.22	0.21	-0.06	0.18
$T_{\text{cont}}$	-0.00	-0.21	-0.18	0.09	-0.14	0.01	-0.64	-0.02	1.00	0.00	-0.23	-0.21	0.06	-0.19
antib $m_0$	0.00	0.00	0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	1.00	0.75	-0.78	-0.27	-0.30
antib $\sigma_L$	-0.00	-0.00	-0.00	0.00	-0.00	-0.27	0.28	0.22	-0.23	0.75	1.00	-0.43	-0.37	-0.20
antib $\sigma_R$	-0.00	-0.00	-0.00	0.00	-0.00	-0.26	0.26	0.21	-0.21	-0.78	-0.43	1.00	0.37	0.58
antib $\chi$	0.00	0.00	0.00	-0.00	0.00	0.07	-0.07	-0.06	0.06	-0.27	-0.37	0.37	1.00	0.84
antib $p$	-0.00	-0.00	-0.00	0.00	-0.00	-0.22	0.23	0.18	-0.19	-0.30	-0.20	0.58	0.84	1.00

# Appendix K

## Detector $A_{CP}$ Appendix

Figure K.1:  $K^-$  nucleon cross section[1]

Figure K.2:  $K^+$  nucleon cross section[1]

# Appendix L

## Distributions for PDF

$$\text{Argus}(x; \chi, p, m_0) = Ax \left(1 - \frac{x^2}{m_0^2}\right)^p \exp \left\{ -\frac{1}{2} \chi^2 \left(1 - \frac{x^2}{m_0^2}\right) \right\} \quad 0 \leq x \leq m_0 \quad (\text{L.1})$$

Figure L.1: Argus Distribution.  $\chi$  is Argus slope.  $p$  is Argus Power.  $m_0$  is resonance mass. And,  $A$  is the normalization.

$$\text{CrystalBall}(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot \left(B - \frac{x-\bar{x}}{\sigma}\right)^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases} \quad (\text{L.2})$$

where

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \quad (\text{L.3})$$

and

$$B = \frac{n}{|\alpha|} - |\alpha| \quad (\text{L.4})$$

Figure L.2: CrystalBall Distribution.

$$\text{NVB}(x; m_0, \sigma, t) = \exp \left\{ \frac{1}{2} \left( \frac{\ln q_y}{t} \right)^2 + t^2 \right\} \quad (\text{L.5})$$

where

$$q_y = 1 + t \frac{(x - m_0) \sinh \left\{ t \sqrt{\ln 4} \right\}}{\sigma t \sqrt{\ln 4}} \quad (\text{L.6})$$

Figure L.3: Novosibirsk Distribution.  $m_0$  is the peak location.  $\sigma$  is the width of the distribution.  $t$  is tail of the distribution

$$\text{Cruijff}(x; m_0, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = A \begin{cases} \exp \left[ -\frac{(x-m_0)^2}{2\sigma_L^2 + \alpha_L(x-m_0)^2} \right] & \text{if } x < m_0 \\ \exp \left[ -\frac{(x-m_0)^2}{2\sigma_R^2 + \alpha_R(x-m_0)^2} \right] & \text{if } x > m_0 \end{cases} \quad (\text{L.7})$$

Figure L.4: Cruijff Distribution. A bifurcated gaussian function with second order polynomial correction to width of each side.  $m_0$  is the peak location.  $\sigma_L$  and  $\sigma_R$  are the width for each side of the distribution.  $\alpha_L$  and  $\alpha_R$  are the width correction factors for each side of the distribution.  $A$  is the normalization factor.

## Appendix M

# Effect of Charged/Neutral $B$ Efficiency/Lifetime Difference and Isospin Asymmetry on Total Flavor Asymmetry

One may be concerned about the fact that our total asymmetry was calculated using combined charged and neutral  $B$  which have different reconstruction efficiencies and that the flavor asymmetry of the two may be different. In this section we will show that the total effect is suppressed. The analysis in this section is similar to what we show in Section 5.3.

Let us start off with some numbers to get a feeling of the scale of what we are dealing with

- First the isospin asymmetry which quantify the difference of the width of charged  $B$  and neutral  $B$  [34]

$$\Delta_{0-} = \frac{\Gamma(\bar{B} \rightarrow X_s \gamma) - \Gamma(B^- \rightarrow X_s \gamma)}{\Gamma(\bar{B} \rightarrow X_s \gamma) + \Gamma(B^- \rightarrow X_s \gamma)} = -0.006 \pm 0.058 \pm 0.009 \pm 0.024,$$

where the first error is statistical, the second is systematics and the third is due production ratio of  $\bar{B}^0$  and  $B^-$ . The value of  $\Delta_{0-}$  is at the order of a few percents.

- Isospin efficiency asymmetry of neutral  $B$  ( $\epsilon^0$ ) and charged  $B$  ( $\epsilon^+$ ). The efficiency of charged and neutral  $B$  just happen to be similar. We estimate this

isospin efficiency asymmetry from total number of events we found from charged  $B$  and neutral  $B$  fit:

$$\theta = \frac{\epsilon^0 - \epsilon^+}{\epsilon^0 + \epsilon^+} \approx \frac{7789 - 7382}{7789 + 7382} = 2.7\%.$$

We found the number to be at the order of a few percent. This estimation is actually a combination of  $\Delta_{0-}$  and  $\theta$ . But, the important point here is that its value is just a few percent.

Further, using this definition of  $\theta$ , we have

$$\epsilon^0 = \epsilon(1 + \theta) \tag{M.1}$$

$$\epsilon^+ = \epsilon(1 - \theta). \tag{M.2}$$

We will use this relation later on.

- $\Delta A_{X_s^- \gamma}$  as we measure;

$$\Delta A_{X_s^- \gamma} = 5.62\% \pm 3.78\% \pm 1.45\%,$$

is also at the order of a few percent.

To see the effect of these numbers on the total  $A_{CP}$ , let us start with we actually measure:

$$\text{Measured } A_{CP} = \frac{(\epsilon^+ N_b^+ + \epsilon^0 N_b^0) - (\epsilon^+ N_{\bar{b}}^+ + \epsilon^0 N_{\bar{b}}^0)}{(\epsilon^+ N_b^+ + \epsilon^0 N_b^0) + (\epsilon^+ N_{\bar{b}}^+ + \epsilon^0 N_{\bar{b}}^0)} \tag{M.3}$$

where

- $N_b^+$  and  $N_{\bar{b}}^+$  are the true number of event of charged  $B$  for each flavor before selection.
- $N_b^0$  and  $N_{\bar{b}}^0$  are the true number of event of neutral  $B$  for each flavor before selection.
- $\epsilon^+$  is the efficiency for charged  $B$  and  $\epsilon^0$  is the efficiency for neutral  $B$ .

What we want to measure is actually

$$\text{True } A_{CP} = \frac{(N_b^+/\tau^+ + N_b^0/\tau^0) - (N_{\bar{b}}^+/\tau^+ + N_{\bar{b}}^0/\tau^0)}{(N_b^+/\tau^+ + N_b^0/\tau^0) + (N_{\bar{b}}^+/\tau^+ + N_{\bar{b}}^0/\tau^0)},$$

where  $\tau^+$  and  $\tau^0$  is the lifetime of  $B^+$  and  $B^0$  accordingly. We divide number of event by the lifetime to convert branching fraction into width where we can combine. Here we assume CPT that  $\tau^+ = \tau^-$  and  $\tau^{B^0} = \tau^{\bar{B}^0}$ . To make it easier to read, let us define  $\omega^{+/0} = \frac{1}{\tau^{+/0}}$ .

$$\text{True } A_{CP} = \frac{(N_b^+\omega^+ + N_b^0\omega^0) - (N_{\bar{b}}^+\omega^+ + N_{\bar{b}}^0\omega^0)}{(N_b^+\omega^+ + N_b^0\omega^0) + (N_{\bar{b}}^+\omega^+ + N_{\bar{b}}^0\omega^0)},$$

The ratio of  $\tau^+$  and  $\tau^0$  is very well measured [1]

$$\frac{\tau^+}{\tau^0} = 1.079 \pm 0.007. \quad (\text{M.4})$$

This makes<sup>1</sup>

$$\omega^0 = \omega(1 + \phi) \quad (\text{M.5})$$

$$\omega^+ = \omega(1 - \phi) \quad (\text{M.6})$$

where

$$\phi = 0.038.$$

Again, the important thing is that  $\phi$  is at the order of a few percent.

Let us make M.3 looks a bit more like M by multiplying  $\omega^+\omega^0$  to the numerator and denominator.

$$\text{Measured } A_{CP} = \frac{(\epsilon^+\omega^0[\omega^+N_b^+] + \epsilon^0\omega^+[\omega^0N_b^0]) - (\epsilon^+\omega^0[\omega^+N_{\bar{b}}^+] + \epsilon^0\omega^+[\omega^0N_{\bar{b}}^0])}{(\epsilon^+\omega^0[\omega^+N_b^+] + \epsilon^0\omega^+[\omega^0N_b^0]) + (\epsilon^+\omega^0[\omega^+N_{\bar{b}}^+] + \epsilon^0\omega^+[\omega^0N_{\bar{b}}^0])} \quad (\text{M.7})$$

---

<sup>1</sup>One may also include the effect of difference in  $Br(\Upsilon(4S) \rightarrow B^0\bar{B}^0)$  and  $Br(\Upsilon(4S) \rightarrow B^+B^-)$ , which is also at the order of 1%[1], in a similar manner. This does not change our conclusion.

To the first order in  $\phi$  and  $\theta$

$$\epsilon^+ \omega^0 \approx \epsilon \omega (1 + \phi - \theta) = \epsilon \omega (1 + k) \quad (\text{M.8})$$

$$\epsilon^0 \omega^+ \approx \epsilon \omega (1 - \phi + \theta) = \epsilon \omega (1 - k) \quad (\text{M.9})$$

Further, to make the expression more concise let us write

$$\omega^0 N_{b/\bar{b}}^0 = M_{b/\bar{b}}^0, \quad \omega^+ N_{b/\bar{b}}^+ = M_{b/\bar{b}}^+, \quad (\text{M.10})$$

Substitute these into Eq. M.11 and cancel  $\epsilon \omega$ , we have

$$\text{Measured } A_{CP} = \frac{\left[ (M_b^+ + M_b^0) - (M_{\bar{b}}^+ + M_{\bar{b}}^0) \right] + k \left[ (M_b^+ - M_{\bar{b}}^+) - (M_b^0 - M_{\bar{b}}^0) \right]}{\left[ (M_b^+ + M_b^0) + (M_{\bar{b}}^+ + M_{\bar{b}}^0) \right] + k \left[ (M_b^+ + M_{\bar{b}}^+) - (M_b^0 + M_{\bar{b}}^0) \right]} \quad (\text{M.11})$$

Let us divide denominator and numerator by

$$M_b^+ + M_b^0 + M_{\bar{b}}^+ + M_{\bar{b}}^0.$$

We have

$$\text{Measured } A_{CP} = \frac{\overbrace{\left[ (M_b^+ + M_b^0) - (M_{\bar{b}}^+ + M_{\bar{b}}^0) \right]}^{\text{True } A_{CP}} + k \overbrace{\left[ (M_b^+ - M_{\bar{b}}^+) - (M_b^0 - M_{\bar{b}}^0) \right]}^{\Delta A_{X_s^- \gamma} / 2 + \text{higher order}}}{1 + k \underbrace{\left[ (M_b^+ + M_{\bar{b}}^+) - (M_b^0 + M_{\bar{b}}^0) \right]}_{\Delta_{0-}}} \quad (\text{M.12})$$

$$\approx \frac{\text{True } A_{CP} + \frac{k}{2} \Delta A_{X_s^- \gamma}}{1 + k \Delta_{0-}} \quad (\text{M.13})$$

$$\approx \text{True } A_{CP} + \frac{k}{2} \Delta A_{X_s^- \gamma} - k \Delta_{0-} \quad (\text{M.14})$$

We can see from Eq. M.14 that the effect from isospin asymmetry and  $\Delta A_{X_s^- \gamma}$  on the total flavor asymmetry is suppressed by the efficiency asymmetry. The total effect is at the order of  $10^{-4}$  which is much smaller compared to other systematics.

## Appendix N

### *CPT* and $\Delta\Gamma_d$ Assumptions

Theoretical  $A_{CP}$  is defined as a ratio of widths as shown in Eq.1.1. The experimental measurement of  $A_{CP}$  is done using number of events as shown in Eq. 5.4. In order to use ratio of number of events to represent the ratio of width we need to use a couple assumptions. These assumption are common but they are usually not spelled out. Let us go through step by step the assumptions we used.

Let us start with

$$A_{CP} = \frac{\Gamma_{b \rightarrow s\gamma} - \Gamma_{\bar{b} \rightarrow \bar{s}\gamma}}{\Gamma_{b \rightarrow s\gamma} + \Gamma_{\bar{b} \rightarrow \bar{s}\gamma}}$$

First, using  $\Gamma = Br/\tau$  yields

$$A_{CP} = \frac{\frac{\tau_{\bar{b}}}{\tau_b} Br(b \rightarrow s\gamma) - Br(\bar{b} \rightarrow \bar{s}\gamma)}{\frac{\tau_{\bar{b}}}{\tau_b} Br(b \rightarrow s\gamma) + Br(\bar{b} \rightarrow \bar{s}\gamma)},$$

where  $\tau_b$  and  $\tau_{\bar{b}}$  are the life time of the  $B$  meson containing  $b$  quark and  $\bar{b}$  quark accordingly. Assuming *CPT* invariance  $\tau_b$  and  $\tau_{\bar{b}}$  are the same. This is typically assumed in  $A_{CP}$  analysis. The only measurement of a related quantity to  $\frac{\tau_{\bar{b}}}{\tau_b}$  is from LHCb[35]:

$$\frac{\tau_{B^-}}{\tau_{B^+}} = 1.024 \pm 0.024 \pm 0.007,$$

which is the ratio of the lifetime of two charged  $B$ . The direct measurement on the ratio of lifetime of two  $B$  was not done due to difficulty arises from neutral  $B$  mixing.

So let's assume  $\tau_b = \tau_{\bar{b}}$ , we have

$$A_{CP} = \frac{Br(b \rightarrow s\gamma) - Br(\bar{b} \rightarrow \bar{s}\gamma)}{Br(b \rightarrow s\gamma) + Br(\bar{b} \rightarrow \bar{s}\gamma)},$$

The branching fraction is just

$$Br(b \rightarrow s\gamma) = \frac{N(b \rightarrow s\gamma)}{N(b \text{ produced})},$$

$$Br(\bar{b} \rightarrow \bar{s}\gamma) = \frac{N(\bar{b} \rightarrow \bar{s}\gamma)}{N(\bar{b} \text{ produced})}.$$

If

$$N(b \text{ produced}) = N(\bar{b} \text{ produced}),$$

then we get the expression we are looking for:

$$A_{CP} = \frac{N(b \rightarrow s\gamma) - N(\bar{b} \rightarrow \bar{s}\gamma)}{N(b \rightarrow s\gamma) + N(\bar{b} \rightarrow \bar{s}\gamma)}.$$

This assumption is true for charged  $B$  since we produce  $B$ s from  $\Upsilon(4S)$  which always decays into two  $B$ s of the opposite flavor. Neutral  $B$ s, however, oscillate after being produced. Since we are measuring time integrated quantity, the frequency of oscillation does not affect our observables. Yet, we need to assume that the probability of  $B^0 \rightarrow \bar{B}^0$  is the same as  $\bar{B}^0 \rightarrow B^0$  that is that is  $CP$  violation  $B$  mixing,  $\Delta\Gamma_d$ , is zero. This is typically assumed in  $B_d$  asymmetry measurement and the Standard Model predicts this to be  $10^{-3}$ , much smaller than our sensitivity. The current world average of this quantity is  $\text{sgn}(\lambda_{CP})\Delta\Gamma_d/\Gamma_d = 0.015 \pm 0.018$ [1]. However, one must be careful about this assumption when measuring  $B_s$  asymmetry since it has been established that  $CP$  violation is  $B_s$  mixing is large and non-zero ( $\Delta\Gamma_s/\Gamma_s = +0.150 \pm 0.020$ )[1].