Fermilab FERMILAB-THESIS-2002-11

AN UPDATED MEASUREMENT OF $\sin 2\beta$ WITH MULTIPLE FLAVOR TAGS USING A TAG-COMBINING ALGORITHM

Tianjie Gao

A DISSERTATION

in

PHYSICS AND ASTRONOMY

Presented to the Graduate Faculty of the University of Pennsylvania

in Partial Fulfillment of the Requirement for the Degree of Doctor of Philosophy

2001

Supervisor of Dissertation _____

Graduate Group Chairman _____

To my family, who has always been there for me.

ACKNOWLEDGEMENT

I would like to thank my advisor Professor Nigel Lockyer for introducing me to the field of experimental particle physics then guiding and supporting me through a path I eventually become proud of. His keen interests in "fundamentals" have influenced my research and will benefit me in years to come. I cannot thank enough my colleague, Dr. Joel Heinrich, for the wonderful research experience I have had. Learning something new from him everyday is literally true. His silky skills in mathematics and computing take me by surprise often and till this day. If "to iterate is human; recurs divine", he recurses. This work cannot be done without him.

Thanks to Professor Joe Kroll and Professor Gino Segrè for sitting on my committee and watching my progress. I thank Professor Kevin Pitts of University of Illinois at Urbana-Champaign for his help on flavor tagging. I enjoyed the great lectures given by Professor Michael Cohn and Professor Gino Segrè. I feel fortunate to have Walter Kononenko as colleague and am grateful for the help he offered.

I would also like to thank Dr. Fumihiko Ukegawa for his friendship and Dr. Owen Long for his help during my stay at Fermilab. Thank you, Chunhui Chen and Shin-Shan Yu, my fellow graduate students, for the many conversations we had. Last but not least, I thank the PENN CDF group for being part of my unforgettable graduate school years. I certainly will miss the High Energy Lunch and the pick-up basketball games with George Loomis, Tom McKenna and company.

ABSTRACT

AN UPDATED MEASUREMENT OF $\sin 2\beta$ WITH MULTIPLE FLAVOR TAGS USING A TAG-COMBINING ALGORITHM

This thesis reports an updated measurement of Standard Model CP violation parameter $\sin 2\beta$ using the CDF detector at Fermilab. The signal samples of $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ are extracted from the 110 pb^{-1} of p-$\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. The flavor of the neutral B mesons is identified at the time of production by three tagging algorithms: a same-side tag, a soft-lepton tag, and an opposite-side-jet tag. The opposite-side-jet tag is a combination of the previously used jet-charge tag and a few new tags. A generic algorithm is developed to combine the tags to improve the tagging performance. A maximum likelihood fitting method is used to determine $\sin 2\beta = 0.91^{+0.37}_{-0.36}$. The uncertainties are improved over the previous CDF measurement. The value is consistent with the standard model prediction of a large positive CP asymmetry in this decay mode. It is also consistent with the world average value of $\sin 2\beta$.

Contents

1 Introduction **1**

1.1 Standard Model and CKM Matrix 1

1.2 ***CP*** Violation in Neutral ***B*** Decay Due to Mixing 2

1.3 The Golden Mode: $B^0/\bar{B}^0 \to J/\psi K_S$ 5

2 The CDF Detector **7**

2.1 Tracking 7

2.2 Calorimeters 8

2.3 Muon Detection 9

2.4 Event and Tracks 9

3 dE/dx and Particle Identification **11**

3.1 CDF Run 1 dE/dx 11

3.1.1 CDF Run I CTC dE/dx 12

3.1.2 The Bethe-Bloch Equation 13

3.2 Sample Selections 14

3.2.1 Muons 15

3.2.2 Pions 16

3.2.3 Protons 17

3.2.4 Electrons and Positrons 19

3.2.5 Low Momentum Tracks 21

3.3 MINUIT Fit 21

3.3.1 Low $\beta\gamma$ Region 22

3.3.2 Resolution as a Function of NCTC 23

3.3.3 Overall Fit 23

3.4 $\boldsymbol{dE/dx}$ For Particle ID 26

4 Flavor Tagging **28**

4.1 Same-Side Tag 29

4.2 Soft-Lepton Tag 29

4.3 Jet-Charge Tag 29

4.4 Kaon Tag 30

4.5 Neutral Tag 30

5 An Algorithm of Combining Tags **31**

5.1 Introduction 31

5.2 The Algorithm 32

5.3 The Tags 35

5.3.1 Jet-Charge Tag 35

5.3.2 Jet-related Neutral Tags 36

5.3.3 dE/dx Tag 38

5.4 Optimization Using the Monte Carlo Sample 38

5.4.1 Monte Carlo Sample 38

5.4.2 Binning 40

5.4.3 Dilution Matrices and Dilution Functions 40

5.5 Tag Association Order 46

5.6 Gory Details 49

5.7 Tag Correlation 50

5.8 Summary 51

6 Measuring $\sin 2\beta$ 55

6.1 Introduction 55

6.2 Tagging With Asymmetry 56

6.2.1 Efficiency and Dilution Functions 56

6.2.2 A Relation Between Efficiency and Dilution Asymmetry 57

6.2.3 Efficiency Asymmetry Using Inclusive J/ψ 59

6.2.4 Source Of Tagging Asymmetry 61

6.3 Calibration Using $B^{\pm} \to J/\psi K^{\pm}$ 62

6.3.1 Event Selection 62

6.3.2 $Q_{\mathbf{final}}$ Tag 63

6.3.3 Dilution vs. $Q_{\mathbf{final}}$ 63

6.3.4 The Calibration Fit 64

6.3.5 The Improved Tagging Power 70

6.3.6 Validity of Dilution vs. $\boldsymbol{Q_{\mathbf{final}}}$ Relation 70

6.3.7 No-jet Events 74

6.4 Tags Revisited 75

6.4.1 Neutral Tags in $\boldsymbol{b}$-jets vs. non-$\boldsymbol{b}$-jets 75

6.4.2 Tags in Monte Carlo vs. data 76

6.4.3 $\boldsymbol{Q_{\mathbf{final}}}$ sensitivity to samples with different $\Delta\phi$ characteristics 76

6.5 Measuring $\mathbf{sin\,2\beta}$ 77

6.5.1 $\boldsymbol{B^0/\bar{B}^0 \to J/\psi K_S}$ Event Selection 78

6.5.2 $\boldsymbol{B^0/\bar{B}^0 \to \psi(2S) K_S}$ Event Selection 80

6.5.3 Calculate $\boldsymbol{Q_{\mathbf{final}}}$ 80

6.5.4 Fit for $\mathbf{sin\,2\beta}$ 80

6.6 The Goodness-of-fit 101

6.6.1 Event binning in 5 dimensions 101

6.6.2 Goodness-of-fit $\pounds$ 104

6.7 Fits to Toy Monte Carlo Data 104

6.7.1 Generation of tags 105

6.7.2 Fitter Bias of $\mathbf{sin\,2\beta}$ 108

6.7.3 Deviation from linear calibration relation 111

7 Conclusion **114**

A Combining Tags Does Not Reduce ϵD^2 **116**

A.1 Background 116

A.2 Preparation 117

A.3 Flavor-neutral Information 118

B A Model for Bias in $\sin 2\beta$ Toy MC **120**

B.1 Introduction 120

B.2 Model $\sin 2\beta$ Fit 122

B.3 Model $\sin 2\beta$ Fit Results 126

B.4 Conclusions 128

List of Figures

1.1 (a): The triangle relation between V_{CKM} elements from the unitary constraint on the first and third columns. (b): The same relation expressed in Wolfenstein parameters with the baseline normalized to 1. 2

1.2 Box diagrams of the neutral B mixing. 4

2.1 A quarter of the CDF detector. 8

2.2 A typical CDF B event. The trigger is the decay $B^0 \to J/\psi K^*$ and subsequently $J/\psi \to \mu^+\mu^-$ and $K^* \to K^+\pi^-$. The other b-quark decay can also be seen at the lower portion of the plot as a jet. Notice that, because of the B meson's lifetime, both secondary vertices are significantly separated from the primary vertex. 10

3.1 The cross section of the CTC. It has 9 superlayers: 5 stereo superlayers that are capable of measuring the z coordinate of tracks, and 4 axial superlayers that can only measure the x-y coordinate of tracks. All superlayer cells are tilted by the Lorentz angle (determined by the magnetic field) from the radial direction. 12

3.2 Reconstructed J/ψ mass. Events with $M_{\mu\mu}$ within $\sigma_{M_{\mu\mu}}$ of $M_{J/\psi}$ are selected. . . . 15

3.3 Reconstructed K_S mass. The parameters in the plot are the number of signal events (p_1), the center of the Gaussian (p_2), the width of the Gaussian (p_3), the mean height of the background (p_4) and the slope of the background (p_5). 16

3.4 Reconstructed Λ mass. Events with $M_{p\pi}$ within $2\,\sigma_{M_{p\pi}}$ of M_{Λ^0} are selected. The same notation for parameters as in Fig. 3.3 is used. 18

3.5 Schematic of electron conversions. The electron and positron tracks are represented as arcs that are part of two circles. $\Delta\phi$ is the azimuthal angle between the tracks at the intersection point. s is the distance between tracks at their tangential points. d is the photon impact parameter. 19

3.6 Reconstructed conversion electron-positron mass. 20

3.7 QCTC resolution and its dependence on NCTC. The first plot shows the histogram of NCTC, the peaks at a few NCTC values are caused by the truncated mean operation; the other plots show different fits to σ as a function of NCTC, where p_1 and p_2 are constants in relations shown in the plots. From these fits, $\sigma = \frac{p_1}{\mathrm{NCTC}}$ is a simple and accurate relation to use. 24

3.8 QCTC dE/dx universal curves. For $\beta\gamma < 1$, tracks marked as K, π and D have their $\beta\gamma$ calculated using the proton mass. Upper left: 1A; upper right: 1B, run number < 67390; lower left: 1B, run number > 67390; lower right: all three curves. The two Run 1B curves can be well related by a constant ratio of 1.2. 25

3.9 QCTC separation power for tracks of NCTC=31. 27

3.10 QCTC separation power for tracks of NCTC=11. 27

5.1 Top: The dilution matrix. Bottom: The fit. 41

5.2 Top: The dilution matrix. Bottom: The fit. 43

5.3 Top: The dilution matrix. Bottom: The fit. 44

5.4 Top: The dilution matrix. Bottom: The fit. 45

5.5 Top: The dilution matrix. Bottom: The fit. 47

6.1 The histogram is the Q_{final} distribution for inclusive J/ψ events that have jets. The solid line is a fit to the efficiency function of the form described in the text. The dashed line is a fit to the efficiency function without the two odd Chebyshev polynomial terms. 61

6.2 Q_{final} distribution for signal region $B^{\pm} \to J/\psi K^{\pm}$ SVX events that have jets. All events satisfy $|m_{\mu\mu} - 5.279| < 5\sigma$; $c\tau_B > 0.005$cm. 65

6.3 Q_{final} distribution for signal region $B^{\pm} \to J/\psi K^{\pm}$ CTC events that have jets. All events satisfy $|m_{\mu\mu} - 5.279| < 5\sigma$; $c\tau_B > 0.005$cm. 65

6.4 Dilution vs. Q_{final} in $B^{\pm} \to J/\psi K^{\pm}$ sample. 83

6.5 Percentages of tracks from opposite-side b and same-side b. The jet clustering algorithm has removed much of the same-side tracks, resulting in their low percentage in the jet. 84

6.6 Solid: b-jet; Dash: non-b jet. Plots show that the neutral variables differentiate b jets from non-b jets. 85

6.7 The dash line shows the tag distribution in the Monte Carlo. The points with error bars show the tag distribution in the $B^{\pm} \to J/\psi K^{\pm}$ data. In the n_{used} tag plot, a small disagreement can be seen. The algorithm is verified to be not sensitive to this difference. 86

6.8 Q_{final}'s calculated using two sets of dilution matrices are plotted as 2D histogram. The diagonal line is $y = x$. The Q_{final} calculated using the proper dilution matrices have slightly higher amplitudes. A linear relation is retained between the two Q_{final}'s, i.e., we can still use the linear $D(Q)$ calibration relation. 87

6.9 The normalized mass of $B^0/\bar{B}^0 \to J/\psi K_S$ 88

6.10 $B^0/\bar{B}^0 \to \psi(2S)K_S$ mass. 89

6.11 The negative log-likelihood function near the $\sin 2\beta$ minimum. 90

6.12 The negative log-likelihood function near the $\sin 2\beta$ minimum. Only the Q_{final} tag is used in the fit. 90

6.13 The true asymmetry ($\sin 2\beta \sin \Delta m_d t$) as a function of the lifetime for $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ events. The data points are sideband-subtracted. The non-SVX events are shown on the right indicating a time-integrated measurement of $\sin 2\beta$. Courtesy of Kevin Pitts, UIUC. 91

6.14 Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 0.25$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 0.25)/\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is -0.002 ± 0.035, which is consistent with zero. . . . 107

6.15 Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 0.8$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 0.8)/\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is 0.041 ± 0.034. 108

6.16 Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 1.0$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 1.0)/\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is 0.102 ± 0.027. 109

6.17 The average pull shift at different generated $\sin 2\beta$ values. A linear fit is returns a slope of 0.111 ± 0.016. If the shift were zero, the points would be aligned around $y = 0$. 110

6.18 Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 1.0$ and the non-linear $D(Q)$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 1.0)/\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is -0.025 ± 0.034. 112

6.19 A non-linear calibration relation of D vs. Q_{final}. 113

7.1 The world average is calculated using $\sin 2\beta$ results from five experiments: OPAL, ALEPH, BELLE, BABAR and CDF. 115

List of Tables

3.1 Parameters of Run 1 universal dE/dx fit. 26

5.1 Tag Association Order 48

5.2 Adding a Second Tag to Jet Charge Tag (part 1 of 3) 52

5.2 Adding a Second Tag to Jet Charge Tag (part 2 of 3) 53

5.2 Adding a Second Tag to Jet Charge Tag (part 3 of 3) 54

6.1 $B^{\pm} \to J/\psi K^{\pm}$ calibration fit parameters and their definitions (part 1 of 3) 66

6.1 $B^{\pm} \to J/\psi K^{\pm}$ calibration fit parameters and their definitions (part 2 of 3) 67

6.1 $B^{\pm} \to J/\psi K^{\pm}$ calibration fit parameters and their definitions (part 3 of 3) 68

6.2 Calibration fit result with $B^{\pm} \to J/\psi K^{\pm}$ (part 1 of 3) 71

6.2 Calibration fit result with $B^{\pm} \to J/\psi K^{\pm}$ (part 2 of 3) 72

6.2 Calibration fit result with $B^{\pm} \to J/\psi K^{\pm}$ (part 3 of 3) 73

6.3 Free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 1 of 4) 92

6.3 Additional free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 2 of 4) 93

6.3 Additional free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 3 of 4) 94

6.3 Additional free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 4 of 4) 95

6.4 $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 1 of 4) 96

6.4 $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 2 of 4) 97

6.4 $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 3 of 4) 98

6.4 $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 4 of 4) 99

6.5 Fit $\sin 2\beta$ results 100

6.6 Event Distribution Comparison (part 1 of 2) 102

6.6 Event Distribution Comparison (part 2 of 2) 103

Chapter 1

Introduction

1.1 Standard Model and CKM Matrix

In the Standard Model fermions are arranged in three generations. Since the quark weak eigenstates differ from their mass eigenstates, mixing matrices are needed to connect weak and mass eigenstates. By convention, in the weak interaction Lagrangian the charge $-\frac{1}{3}$ quarks are mixed by the Cabibbo-Kobayashi-Maskawa (CKM) [1] matrix V_{CKM}

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.1}$$

The Wolfenstein [2] notation is

$$V_{\mathrm{CKM}} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix}, \tag{1.2}$$

where parameters λ, A, ρ and η are all real numbers. The parameter $\lambda \sim 0.22$ corresponds to the Cabibbo mixing through $\lambda = \sin\theta_c$, where θ_c is the Cabibbo angle [3]. The parameter $A \sim 0.8$ and ρ and η are close to unity. $V_{\rm CKM}$ has four real parameters and is unitary, i.e. $V_{\rm CKM}^\dagger V_{\rm CKM} = 1$. One particular relation from the unitary constraint on $V_{\rm CKM}$ is

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \tag{1.3}$$

which, in the ρ-η plane, is shown in Fig. 1.1.

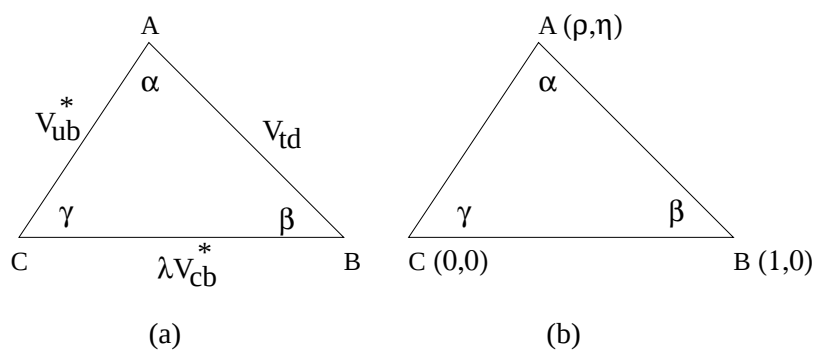


Figure 1.1: (a): The triangle relation between $V_{\rm CKM}$ elements from the unitary constraint on the first and third columns. (b): The same relation expressed in Wolfenstein parameters with the baseline normalized to 1.

1.2 CP Violation in Neutral B Decay Due to Mixing

The symmetries C (charge conjugation) and P (space inversion) both hold for strong and electromagnetic interactions. In the weak interaction, they are violated. In 1964 CP violation was observed in K^0 decays at a level represented by the parameter $\epsilon = 2.3 \times 10^{-3}$ [4]. If the CKM matrix is the cause for that, larger CP violation is expected in B^0 decays, where we can compare

B^0 and $\bar{B}^0$ decays. The CP asymmetry is defined as

$$A_{CP} = \frac{\Gamma(B^0_{\rm phys}(t) \to f) - \Gamma(\bar{B}^0_{\rm phys}(t) \to \bar{f})}{\Gamma(B^0_{\rm phys}(t) \to f) + \Gamma(\bar{B}^0_{\rm phys}(t) \to \bar{f})}, \tag{1.4}$$

where Γ is the decay rate and f stands for the final state of the decay and $\bar{f}$ is the CP conjugate state of f. We will see how it is related to neutral B mixing. The mass eigenstates of neutral B mesons (light and heavy) are

$$|B_{\rm L}\rangle = p|B^0\rangle + q|\bar{B}^0\rangle \tag{1.5}$$

$$|B_{\rm H}\rangle = p|B^0\rangle - q|\bar{B}^0\rangle \tag{1.6}$$

where $|\frac{q}{p}| = 1$ but $\frac{q}{p}$ may have a non-zero phase. Therefore, an initial pure flavor eigenstate of the neutral B evolves over time as

$$|B^0_{\rm phys}(t)\rangle = g_+(t)|B^0\rangle + (q/p)g_-(t)|\bar{B}^0\rangle \tag{1.7}$$

$$|\bar{B}^0_{\rm phys}(t)\rangle = (p/q)g_-(t)|B^0\rangle + g_+(t)|\bar{B}^0\rangle \tag{1.8}$$

where

$$g_+(t) = e^{-i(M-\frac{i}{2}\Gamma)t}\cos(\Delta M t/2) \tag{1.9}$$

$$g_-(t) = e^{-i(M-\frac{i}{2}\Gamma)t} i\sin(\Delta M t/2) \tag{1.10}$$

and $M = \frac{1}{2}(M_{\rm L} + M_{\rm H})$ and $\Delta M = M_{\rm H} - M_{\rm L}$; since $\Gamma_{\rm H} \sim \Gamma_{\rm L}$ we use Γ for the decay rates. That is, a pure flavor state B^0 or $\bar{B}^0$ evolves over time into a mixture of both.

Let A be the decay amplitude of $B^0 \to f$, $\bar{A}$ the amplitude of $\bar{B}^0 \to \bar{f}$ and define $\lambda_0 = \frac{q\bar{A}}{pA}$. We have the decay rates

$$\Gamma(B^0_{\rm phys}(t) \to f) = |A|^2 e^{-\Gamma t}[\frac{1+|\lambda_0|^2}{2} + \frac{1-|\lambda_0|^2}{2}\cos(\Delta M t) - {\rm Im}\lambda_0 \sin(\Delta M t)] \tag{1.11}$$

$$\Gamma(\bar{B}^0_{\rm phys}(t) \to \bar{f}) = |A|^2 e^{-\Gamma t}[\frac{1+|\lambda_0|^2}{2} + \frac{1-|\lambda_0|^2}{2}\cos(\Delta M t) + {\rm Im}\lambda_0 \sin(\Delta M t)] \tag{1.12}$$

Therefore the CP asymmetry due to mixing is

$$A_{CP}(t) = \frac{(1 - |\lambda_0|^2)\cos(\Delta M t) - 2\mathrm{Im}\lambda_0 \sin(\Delta M t)}{1 + |\lambda_0|^2}. \tag{1.13}$$

To measure A_{CP} requires knowledge of λ_0. There are two components in λ_0, the mixing and the decay amplitude. The neutral B mixing mechanism is illustrated in Fig. 1.2.

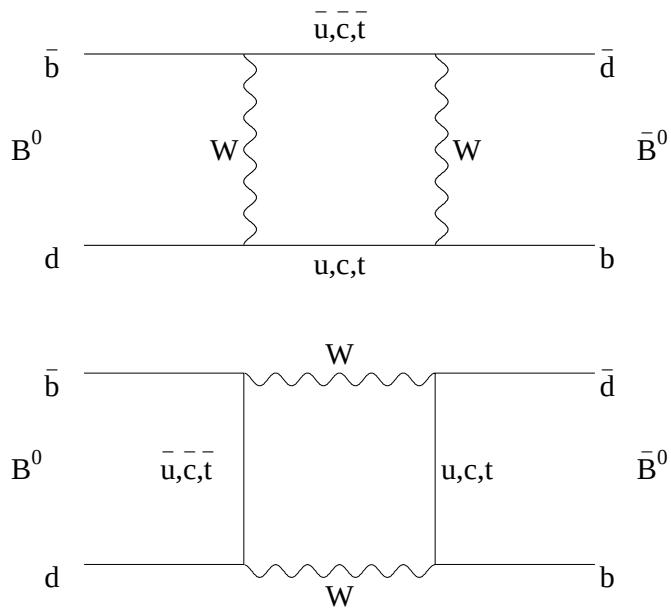


Figure 1.2: Box diagrams of the neutral B mixing.

The box diagram is dominated by the top quark due to its large mass. Therefore, we have

$$\frac{q}{p} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}. \tag{1.14}$$

1.3 The Golden Mode: $B^0/\bar{B}^0 \to J/\psi K_S$

Now we are ready for the decay-amplitude component of λ_0. In a hadron collider the decay mode of choice is $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ because of their relatively clean signals and the absence of competing diagrams with different weak phases. Through the decay chain $b \to c\bar{c}s$, we have

$$\frac{\bar{A}}{A} = \frac{V_{cb} V_{cs}^*}{V_{cb}^* V_{cs}}. \tag{1.15}$$

However, the decay product K^0 can also mix due to the same mechanism of neutral B mixing albeit with the charm quark dominating the box diagram. This contributes another component to λ_0

$$\frac{q'}{p'} = \frac{V_{cs} V_{cd}^*}{V_{cs}^* V_{cd}}. \tag{1.16}$$

With all three components included we have

$$\lambda_0 = \frac{V_{td} V_{cb} V_{cd}^*}{V_{td}^* V_{cb}^* V_{cd}}. \tag{1.17}$$

From the unitary triangle we have

$$\mathrm{Im}\lambda_0 = -\sin 2\beta. \tag{1.18}$$

Since A_{CP} is induced only by mixing, that is, $|\frac{A}{\bar{A}}| = 1$, we have

$$A_{CP} = \sin 2\beta \sin \Delta M t. \tag{1.19}$$

Another decay mode, $B^0/\bar{B}^0 \to \psi(2S)K_S$ with $\psi(2S) \to \mu^+\mu^-$ or $\psi(2S) \to \mu^+\mu^-\pi^+\pi^-$, having the same CP property as $B^0/\bar{B}^0 \to J/\psi K_S$, can be added to increase the sample size. We will use both decay modes to measure $\sin 2\beta$.

Chapter 2

The CDF Detector

The Collider Detector at Fermilab (CDF) is a general purpose magnetic detector for studying 1.8 TeV center-of-mass energy p-$\bar{p}$ collisions at the Fermilab Tevatron [5]. The detector covers all of the solid angle down to 2° from the beam direction and the complete azimuthal angle. Event analyses utilize charged particle tracking, muon identification and calorimetry information, etc.

For this analysis, we will concentrate on the central part of the detector, where our samples of B decays are selected and the tracking system works to its utmost capability. Fig. 2.1 shows a quarter of the detector.

2.1 Tracking

Tracking of charged particles is achieved by a silicon detector (SVX) immediately surrounding the beam pipe, a vertex detector (VTX) outside the SVX, and a central tracking drift chamber (CTC) outside the VTX. A 1.5 Tesla magnetic field in the beam direction is used for measurement of the

Figure 2.1: A quarter of the CDF detector.

track momentum. A typical CDF event is shown in Fig. 2.2. The CTC also has the capability of measuring the dE/dx of the charged tracks, which is encoded as the pulse width from the charge deposited by each hit. The average width of 80% wire hits is recorded for each track. The top 20% of the width distribution is not used due to its long attenuating shape (Landau tail). The dE/dx information is used for particle identification [25].

2.2 Calorimeters

Calorimeters are used to measure the energy flow of particles from the collisions. They are situated outside the tracking system and segmented into towers. Each tower occupies 15° in azimuth and

0.1 in pseudorapidity, defined as $\eta = -\ln\tan\frac{\theta}{2}$ where θ is the polar angle from the proton beam direction (z). The towers all project back to the nominal interaction point on the beamline.

2.3 Muon Detection

The CDF muon detectors consist of CMU (Central MUon detector), CMP (Central Muon uPgrade) and CMX (Central Muon eXtension). CMU and CMP are located outside the central electromagnetic and hadronic calorimeters. They cover the detector region of $|\eta| < 0.7$, where muons penetrating the calorimeters are detected.

2.4 Event and Tracks

Each track in an event is uniquely described by a set of five helix parameters $(d_0, z, c, \cot\theta, \phi_0)$, where d_0 is the impact parameter, z is z-coordinate of the track production point, c is the curvature of the charged track in the magnetic field, $\cot\theta$ is the helix angle of the track, and ϕ_0 is the azimuthal angle of the track production point. From these parameters, with the knowledge of an event's primary vertex and the precision of the SVX detector, the track probability T_p [24]—the probability that a track is from the primary vertex—can be derived. Tracks of less stable particles can be reconstructed from their stable daughters. For example, the CDF subroutine CTVMFT uses 4-momenta conservation and a vertex constraint to derive the parent particle's momentum and mass. The transverse displacement of the reconstructed secondary vertex from the primary vertex is defined as $L_{xy} = \frac{\vec{L}\cdot\vec{p_T}}{|\vec{p}|}$. The tracking system applies a cutoff at $0.4\,\mathrm{GeV}/c$ on the track p_T.

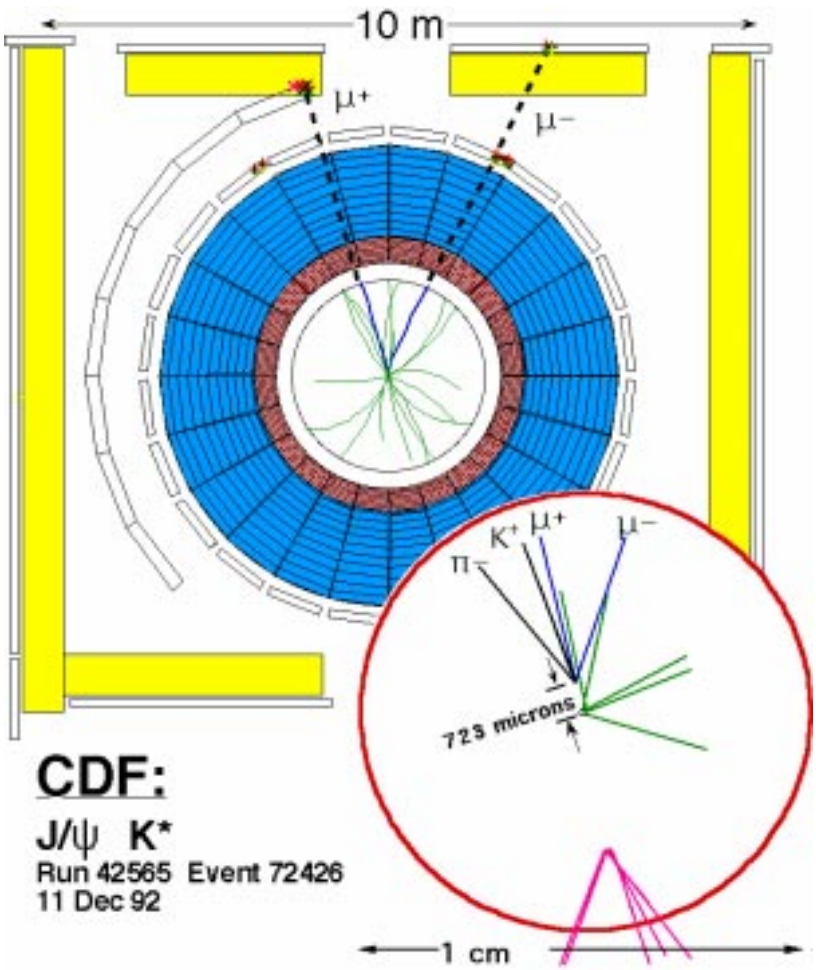


Figure 2.2: A typical CDF B event. The trigger is the decay $B^0 \to J/\psi K^*$ and subsequently $J/\psi \to \mu^+\mu^-$ and $K^* \to K^+\pi^-$. The other b-quark decay can also be seen at the lower portion of the plot as a jet. Notice that, because of the B meson's lifetime, both secondary vertices are significantly separated from the primary vertex.

Chapter 3

dE/dx and Particle Identification

CDF Run 1 is capable of measuring a (80% truncated) mean value of track dE/dx using the CTC. The cross section of the CTC is shown in Fig. 3.1, where superlayers and cells can be seen. The dE/dx information is used to improve the jet-charge tag. A brief evaluation of how this information is obtained and used is presented here.

3.1 CDF Run 1 dE/dx

The rate a charged particle loses energy in media through ionization is defined as dE/dx. To a good approximation, dE/dx is a function of the particle's velocity [6]. If the momentum of the particle is measured, dE/dx is able to shed some light on the particle's identity. Although a significant amount of work has been done on dE/dx by A. Wicklund [7], H. Keutelian [8] and M. Peters [9] *et al.*, we took another look at its usefulness for B flavor tagging in Run I as a means of preparing for Run II. In particular we are interested in low p_T kaon tagging.

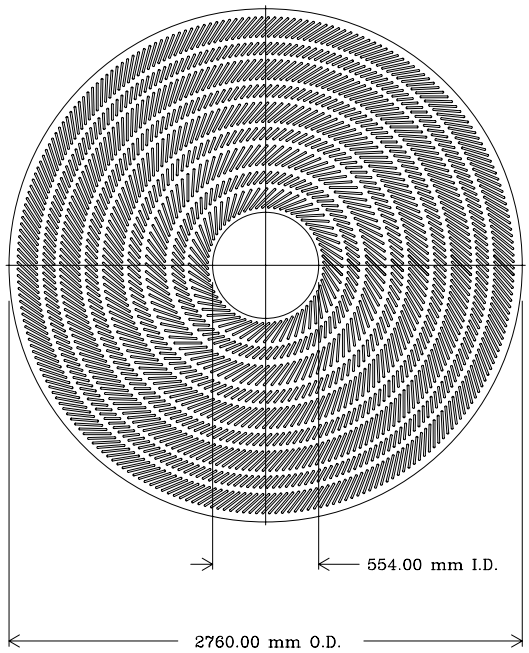


Figure 3.1: The cross section of the CTC. It has 9 superlayers: 5 stereo superlayers that are capable of measuring the z coordinate of tracks, and 4 axial superlayers that can only measure the x-y coordinate of tracks. All superlayer cells are tilted by the Lorentz angle (determined by the magnetic field) from the radial direction.

3.1.1 CDF Run I CTC dE/dx

CTC dE/dx uses an 80% truncated mean value (QCTC) of all wire hits that make up a track. The mean and number of hits (NCTC) of a track are available. There are several corrections to QCTC in the standard Run I offline code, which are applied in the subroutine `DEDXUN`. Corrections have been applied to QCTC for path length, saturation, NCTC dependence, pulse height and pulse width. We will not focus on correcting and refining QCTC here. Our goal is to obtain a universal dE/dx curve that is useful for particle identification, in particular π-K separation. Due to an abrupt run

number dependence of QCTC in Run IB at run 67390 (see Fig. 3.8), we fit three universal curves to represent events in Run 1A, Run 1B (< 67390) and Run 1B (> 67390).

3.1.2 The Bethe-Bloch Equation

The Bethe-Bloch equation [10] gives the mean energy loss rate of a charged particle in material due to ionization of the medium,

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi N e^4}{mc^2\beta^2} z^2 \left(\ln \frac{2mc^2\beta^2\gamma^2}{I} - \beta^2 - \frac{\delta(\beta)}{2} \right). \tag{3.1}$$

In this equation, N is the electron density of the medium, m is electron mass, z is the incident particle charge, I is the mean excitation energy of the medium atoms, $\beta = v/c$, $\gamma = 1/\sqrt{1-\beta^2}$, and $\delta(\beta)$ reflects a correction to the density effect at high β. The energy transferred to an electron is an integral of the force (f) on the electron over the displacement (d) of the electron. When the particle travels with low β, a classical model is suitable to describe the dE/dx. The force $f \sim \frac{1}{r^2}$ is only significant within a range r_0. It is in this range that the electron gets accelerated. The displacement is then $d = \frac{1}{2}at^2$, where t is the duration that the incident particle spends within r_0 and is approximately $\frac{r_0}{\beta c}$. Thus we have $d \propto \frac{1}{\beta^2}$ and the transferred energy $\Delta E \sim \langle f \rangle d \propto \frac{1}{\beta^2}$, noting that $\langle f \rangle$ is the average force in the range d and a constant. This describes the fall of dE/dx at low $\beta\gamma$ before reaching the minimum. For larger β, relativistic effects take place and the classical model is no longer suitable. Dynamically, the field of the traveling particle becomes stronger and thus increases the ionization cross-section; kinematically, with larger β the maximum transfer energy can go higher. The two factors combine to give the logarithmic term. The relativistic rise is dampened by the shielding of the field of the incident particle — medium atoms that surround the particle

are polarized. The medium polarization does not cancel the kinematic part of the logarithmic rise. The kinematic part is responsible for the residual slow rise of dE/dx in very large β.

In practice, instead of the energy lost by the incident particle, it is the energy deposited in the electrons that is measured. CDF measures dE/dx via the pulse width (units of nanoseconds) returned from CTC sense wires. This quantity is not translated into a charge (coulombs). The dE/dx of a track has a wide distribution. So it is necessary to measure dE/dx many times along the track to extract a fair representation of it. The distribution has a long tail in large dE/dx, which makes the average dE/dx diverge logarithmically [10]. Thus a truncated mean of 80% of hits on a track is the estimated dE/dx at CDF. Considering these practical reasons, we do not fit to the exact form of the Bethe-Bloch equation. Instead, we have adopted an empirical variation from H. Keutelian [8] given by

$$Q_0 = \frac{1}{\beta^2}\left[c_0 + c_1 \ln(\frac{\beta\gamma}{\beta\gamma + b})\right] + a_1(\beta - 1) + a_2(\beta - 1)^2. \tag{3.2}$$

This form (pseudo Bethe-Bloch equation) has all the features that are present in the Bethe-Bloch equation. Parameters c_0 and c_1 gives the intensities of the $1/\beta^2$ fall and the logarithmic rise respectively. Parameter b is associated with the CTC gas properties, e.g. mean excitation energy of the gas atoms. Parameters a_1 and a_2 provide further adjustment, especially in low $\beta\gamma$ region.

3.2 Sample Selections

The ultimate goal is to tag kaons from B decays for CP violation studies. Using the J/ψ dataset for the dE/dx is appropriate because the track p_T distributions are well matched to our needs.

3.2.1 Muons

We use the $J/\psi \rightarrow \mu^+\mu^-$ date set to obtain a clean sample of muons. Fig. 3.2 shows the reconstructed J/ψ mass. The signal-to-background ratio (S/B) is roughly 10 : 1. The selection requirements are:

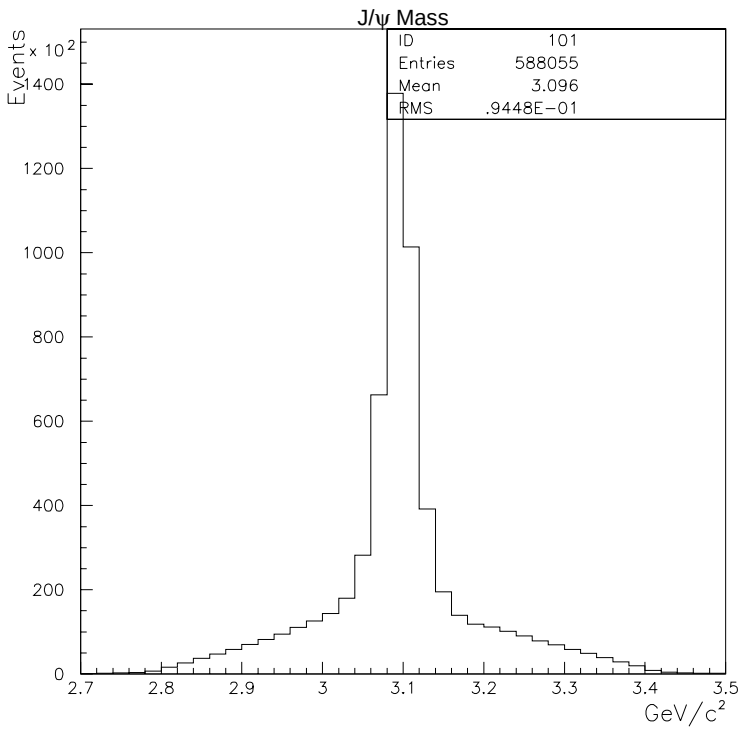


Figure 3.2: Reconstructed J/ψ mass. Events with $M_{\mu\mu}$ within $\sigma_{M_{\mu\mu}}$ of $M_{J/\psi}$ are selected.

- muon matching cuts (between CTC track and muon system track):
  - CMU: matching in x-direction $\chi^2_x < 9$, matching in z-direction $\chi^2_z < 12$
  - CMP: matching in x-direction $\chi^2_x < 9$
  - CMX: matching in x-direction $\chi^2_x < 9$

- two-track ($\mu^+\mu^-$) vertex fit:
  - CTVMFT (CDF vertex fitting subroutine) $\chi^2 < 7$
  - $|m_{\mu\mu} - m_{J/\psi}| < \sigma_{\mu\mu}$

3.2.2 Pions

We use pions from $K_S \to \pi^+\pi^-$. The reconstructed events are shown in Fig. 3.3. The following

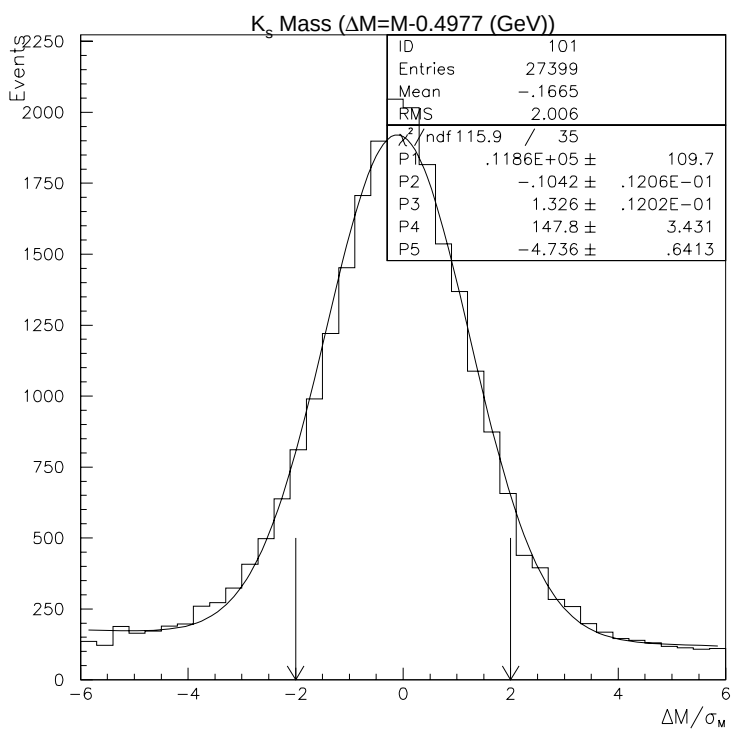


Figure 3.3: Reconstructed K_S mass. The parameters in the plot are the number of signal events (p_1), the center of the Gaussian (p_2), the width of the Gaussian (p_3), the mean height of the background (p_4) and the slope of the background (p_5).

requirements are applied:

- pion track cut:
  - SVX track: impact parameter $d > 1.5\,\sigma_d$
  - CTC track: no impact parameter cut
- two-track ($\pi^+\pi^-$) vertex fit:
  - $L_{xy} > 4\,\sigma_{L_{xy}}$ and $L_{xy} < 4.5$ cm
  - CTVMFT (CDF vertex fitting subroutine) $\chi^2 < 7$
- $|m_{\pi\pi} - m_{K_S}| < 2\,\sigma_{\pi\pi}$ and $\sigma_{\pi\pi} < 0.004\,\mathrm{GeV}/c^2$

3.2.3 Protons

We use protons from $\Lambda^0 \to p\pi^-$. Fig. 3.4 shows the reconstructed events. The following requirements are applied to optimize the signal/background ratio:

- two-track ($p\pi$) vertex fit: CTVMFT $\chi^2 < 7$
- Λ^0 $L_{xy} > 5\,\sigma_{L_{xy}}$ and $L_{xy} > 0.2$ cm, where L_{xy} is the distance the Λ^0 travels in the x-y plane and $\sigma_{L_{xy}}$ is its error
- $|m_{p\pi} - m_{\Lambda^0}| < 2\,\sigma_{p\pi}$ and $\sigma_{p\pi} < 0.002\,\mathrm{GeV}/c^2$
- The absolute values of the impact parameters for both proton and pion must be greater than 3 times their uncertainties
- The track with higher momentum is chosen as the proton

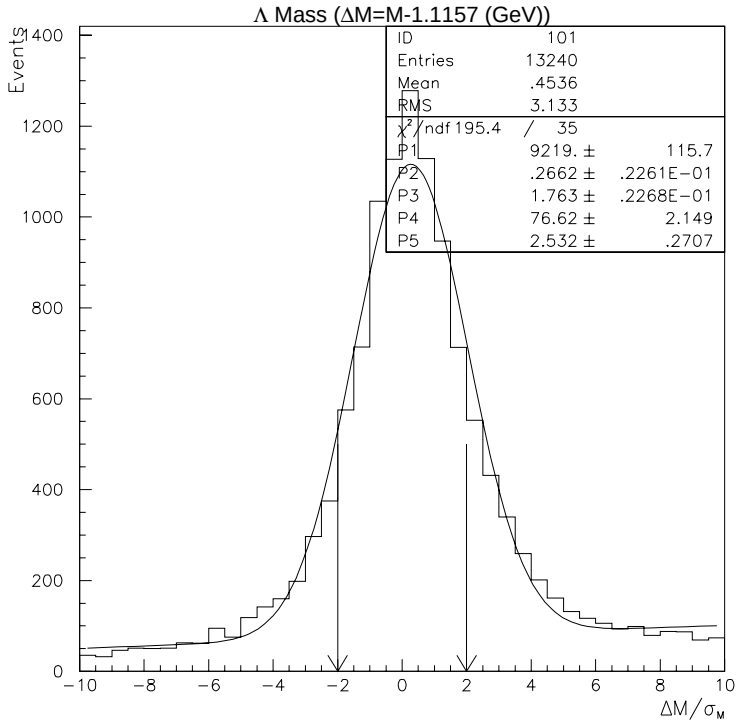


Figure 3.4: Reconstructed Λ mass. Events with $M_{p\pi}$ within $2\,\sigma_{M_{p\pi}}$ of M_{Λ^0} are selected. The same notation for parameters as in Fig. 3.3 is used.

3.2.4 Electrons and Positrons

We use conversion elections and positrons for the electron/positron sample. The following criteria are used to select pairs of electrons and positrons from photon conversions [11]. Fig. 3.5 shows the variables that are cut on.

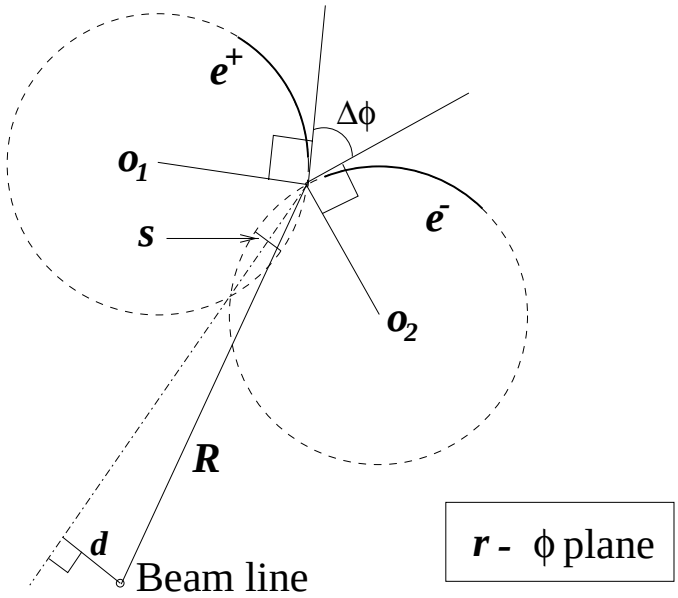


Figure 3.5: Schematic of electron conversions. The electron and positron tracks are represented as arcs that are part of two circles. $\Delta\phi$ is the azimuthal angle between the tracks at the intersection point. s is the distance between tracks at their tangential points. d is the photon impact parameter.

- separation between tangential points in the r-ϕ plane $s < 0.2$ cm;
- $\Delta \cot\theta$ between the electron and the positron < 0.03;
- Δz between the electron and the positron < 2.0 cm;

- conversion radius 22.0 cm $< R <$ 30.0 cm. This is the region where the CTC inner cylinder and the SVX outer cylinder reside. In this region, conversion electrons and positrons are produced with a higher purity;

- $\Delta\phi$ between the electron and the positron at the conversion radius < 0.01;

- pointing deviation from origin (beamline) $d < 1$ cm.

The reconstructed mass spectrum from the conversions is shown in Fig. 3.6.

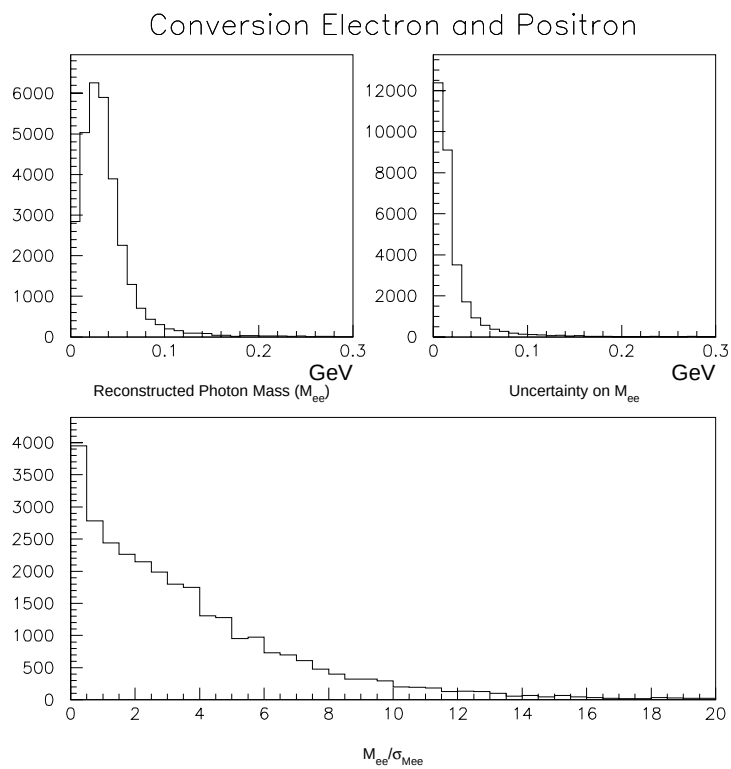


Figure 3.6: Reconstructed conversion electron-positron mass.

3.2.5 Low Momentum Tracks

In order to obtain a good fit to the universal curve, tracks at the low $\beta\gamma$ are needed. The fact that the low $\beta\gamma$ region is also the most rapidly changing region of the Bethe-Bloch function makes it more important to have data in this region. The protons from Λ^0 decays do not effectively cover the low momentum region due to the selection rule that always picks the higher p_T track to be the proton. The pions from Λ^0 decays have $p_T > 0.4\,\mathrm{GeV}/c$ because of the tracking system cutoff. Since the protons carry much of the momenta in Λ^0 decays, the result is that protons have even higher momenta. We need to look elsewhere for the low $\beta\gamma$ tracks. Fortunately we can collect protons from beam pipe interactions. These events have large impact parameters (they come from the beam pipe), and are easily selected with an impact parameter cut.

The requirement is:

- impact parameter $d > 4\,\sigma_d$, where σ_d is the uncertainty of the impact parameter. This enhances the proton fraction.

3.3 MINUIT Fit

We use MINUIT [12] to fit for the dE/dx universal curve (Q_0) of the form in Eqn. 3.2; c_0, c_1, b, a_1 and a_2 are the parameters to be fitted. We have also investigated the dE/dx resolution as a function of NCTC. A log-likelihood method is used and is given by

$$-\log\mathcal{L} = -\sum_i \log\mathcal{P}_i, \tag{3.3}$$

where the sum is over all tracks, $\mathcal{P}_i$ is the ith track's probability

$$\mathcal{P}_i = G(\frac{\mathrm{QCTC}_i - Q_0}{\sigma}), \tag{3.4}$$

where a Gaussian distribution is assumed and $G(x) = e^{-\frac{1}{2}x^2}$; QCTC_i is the track's QCTC value; Q_0 is the mean dE/dx that is determined by the fit parameters and σ is the width of the Gaussian distribution.

3.3.1 Low $\beta\gamma$ Region

In addition to the parameters in the pseudo Bethe-Bloch equation, we need other parameters to reflect the fact that in the low $\beta\gamma$ region deuterons, kaons and pions co-exist with protons, although protons are more numerous. The proton contamination can be seen in the low $\beta\gamma$ region of Fig. 3.8, in which we use the proton mass to calculate $\beta\gamma$. In this region, tracks should not be assumed to be protons only. Instead, their probabilities should each be a weighted sum of Gaussian distributions corresponding to proton, kaon, pion and deuteron. Additional parameters are needed to represent the percentages of proton, kaon, deuteron and pion (we need 3 parameters for 4 types of particles since the final p.d.f. should be normalized). We choose the three parameters prob_D, prob_p and prob_π as the corresponding percentages for deuterons, protons and pions. Thus the probability for such a track is

$$\mathcal{P}_i = \sum_j \mathrm{prob}_j G(\frac{\mathrm{QCTC}_i - Q_{0j}}{\sigma}), \tag{3.5}$$

where the index j runs over kaon, pion, proton and deuteron; and Q_{0j} is the mean dE/dx calculated assuming the track is of type j.

3.3.2 Resolution as a Function of NCTC

It is essential to look at the resolution's dependence on NCTC since NCTC varies in a wide range. To do this, we assign a parameter to represent the width of the Gaussian distribution. Then we fit for the universal curve only using events with a particular NCTC. We repeat this process for NCTC=11 to 30; tracks with NCTC < 11 are not used because 10 hits are not considered enough to provide a valid 80% truncated mean for QCTC. The relation between the resolution and NCTC is shown in Fig. 3.7. The first plot is a histogram of NCTC. The remaining plots show the dE/dx resolution vs. NCTC. Various simple functional forms are attempted. From the plots, we choose the simple and accurate form

$$\sigma = \frac{s}{\mathrm{NCTC}}. \tag{3.6}$$

The constant s will be determined in the overall unbinned MINUIT fit.

3.3.3 Overall Fit

We fit the universal dE/dx curve for all tracks using the relation $\sigma = \frac{s}{\mathrm{NCTC}}$ to reflect that the dE/dx width differs according to NCTC of each track. The results of the fit are shown in Table 3.1.

As the result indicates, Run 1B events with run number > 67390 differ significantly from Run 1A and earlier 1B events. The percentage of kaons can be obtained by subtracting the percentages of proton, pion and deuteron from unity. The universal curves are plotted in Fig. 3.8. We have also applied the fit to positive tracks and negative tracks separately and observed no difference.

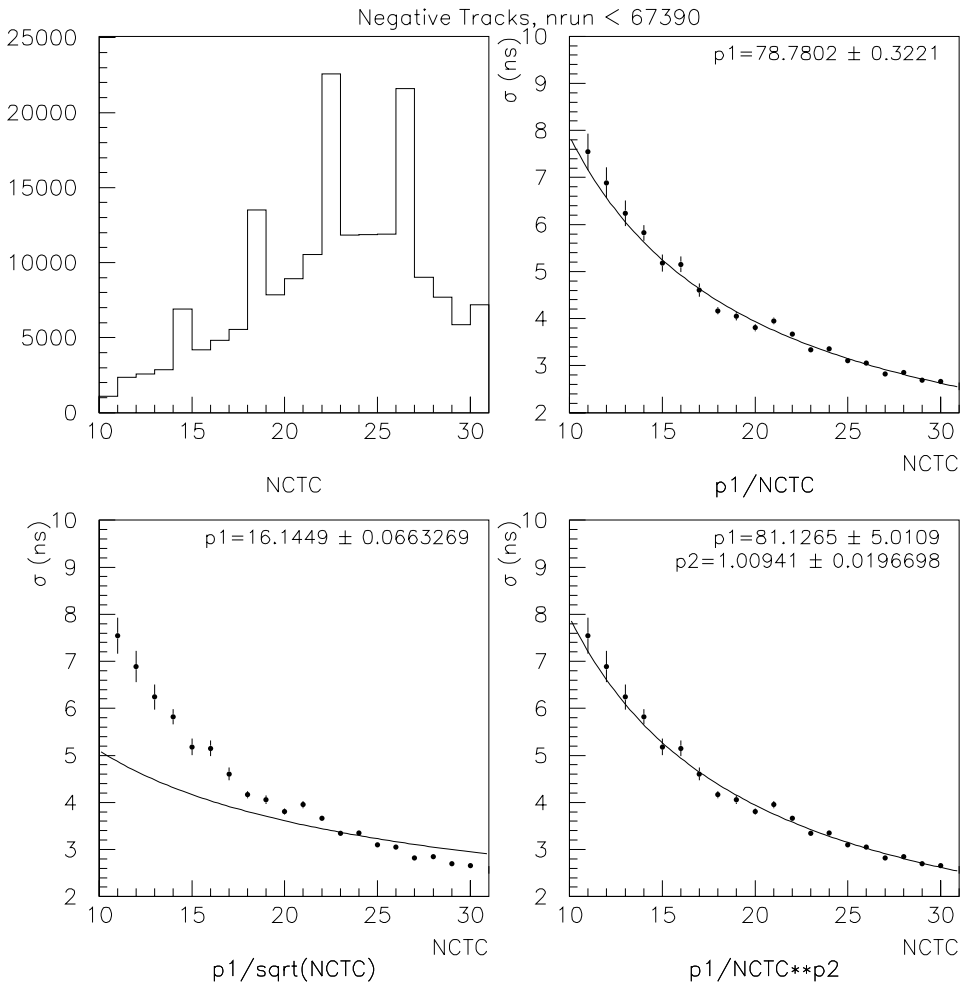


Figure 3.7: QCTC resolution and its dependence on NCTC. The first plot shows the histogram of NCTC, the peaks at a few NCTC values are caused by the truncated mean operation; the other plots show different fits to σ as a function of NCTC, where p_1 and p_2 are constants in relations shown in the plots. From these fits, $\sigma = \frac{p_1}{\mathrm{NCTC}}$ is a simple and accurate relation to use.

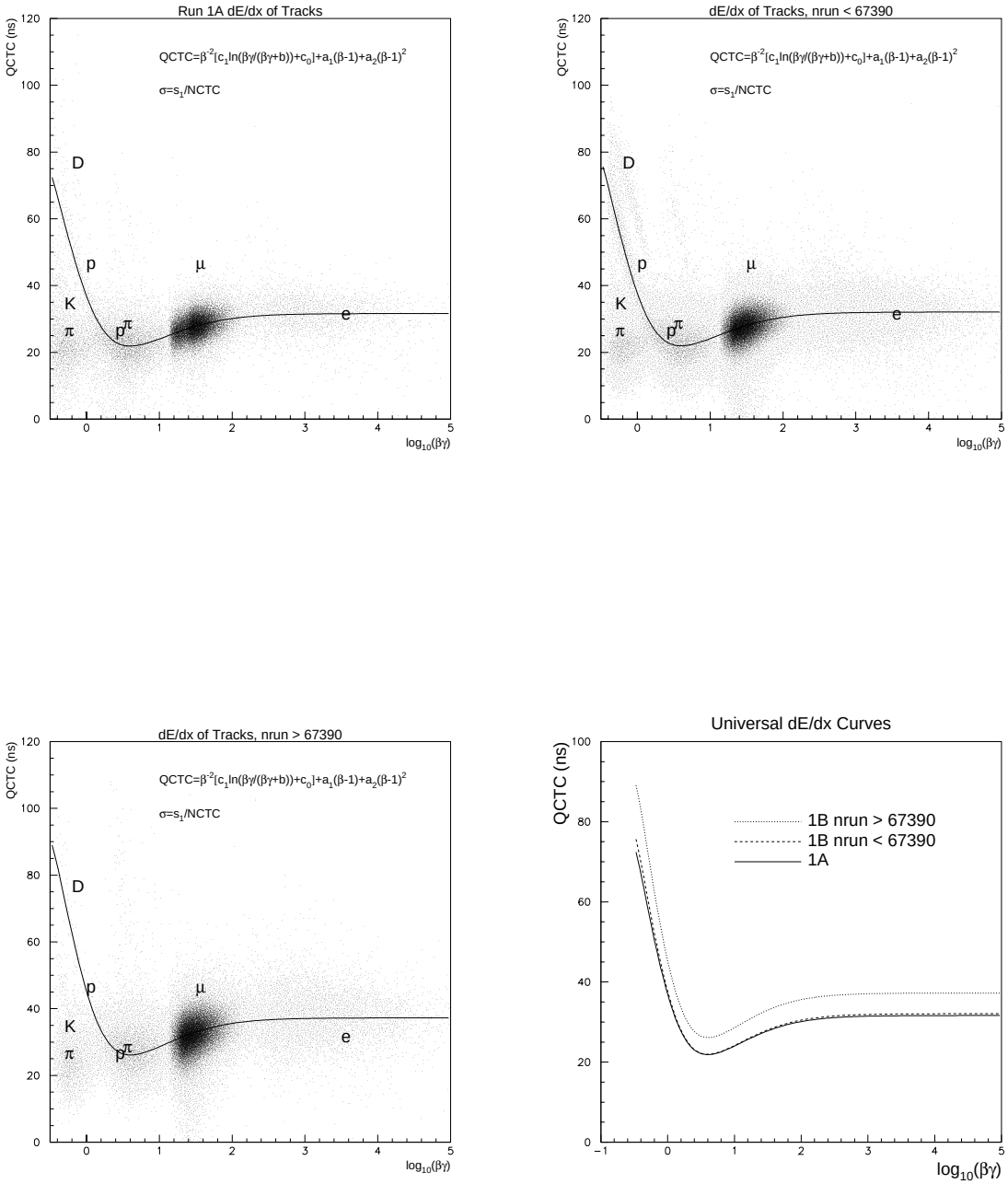


Figure 3.8: QCTC dE/dx universal curves. For $\beta\gamma < 1$, tracks marked as K, π and D have their $\beta\gamma$ calculated using the proton mass. Upper left: 1A; upper right: 1B, run number < 67390; lower left: 1B, run number > 67390; lower right: all three curves. The two Run 1B curves can be well related by a constant ratio of 1.2.

Table 3.1: Parameters of Run 1 universal dE/dx fit.

parameter	value		
	1A	1B (nrun< 67390)	1B (nrun> 67390)
c_0	$31.6 \pm .042$	$32.1 \pm .015$	$37.2 \pm .032$
c_1	$6.06 \pm .030$	$6.08 \pm .009$	$7.38 \pm .017$
b	$28.3 \pm .572$	$30.3 \pm .166$	$25.2 \pm .279$
a_1	$-58.5 \pm .752$	$-59.3 \pm .331$	$-73.0 \pm .498$
a_2	-29.0 ± 1.94	$-22.4 \pm .853$	-27.4 ± 1.24
s(ns)	$78.9 \pm .158$	$83.6 \pm .102$	$97.5 \pm .161$
prob_D	$.036 \pm .004$	$.050 \pm .002$	$.064 \pm .002$
prob_p	$.476 \pm .013$	$.528 \pm .004$	$.619 \pm .004$
prob_π	$.366 \pm .014$	$.326 \pm .004$	$.214 \pm .004$

3.4 dE/dx For Particle ID

The main use of dE/dx in B tagging is K-π separation and low energy electron identification. Fig. 3.9 shows the dE/dx separation in units of σ (resolution) for tracks with 31 hits (NCTC = 31). For tracks with momenta > 2 GeV/c, the separation is about $1.5\,\sigma$. Fig. 3.10 shows the separation power for tracks with 11 hits. The separation is about $0.5\,\sigma$ for tracks with momenta > 2 GeV/c. While Run I dE/dx does not show complete K-π separation, we can utilize this information on a probabilistic basis. For example, instead of returning a track's identity, dE/dx can return a track's probability of being a kaon in the opposite-side B tagging.

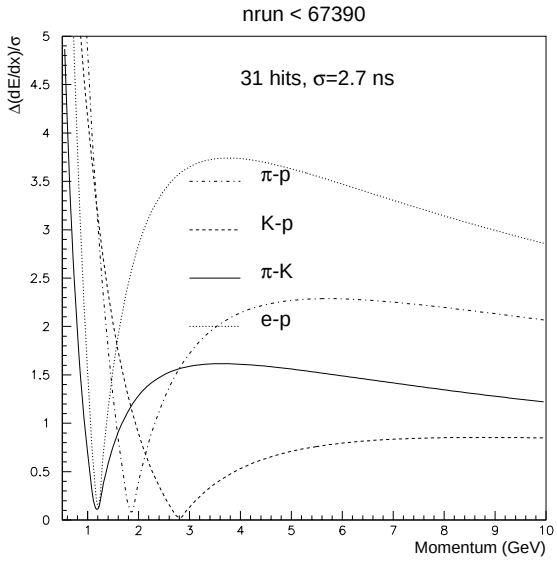


Figure 3.9: QCTC separation power for tracks of NCTC=31.

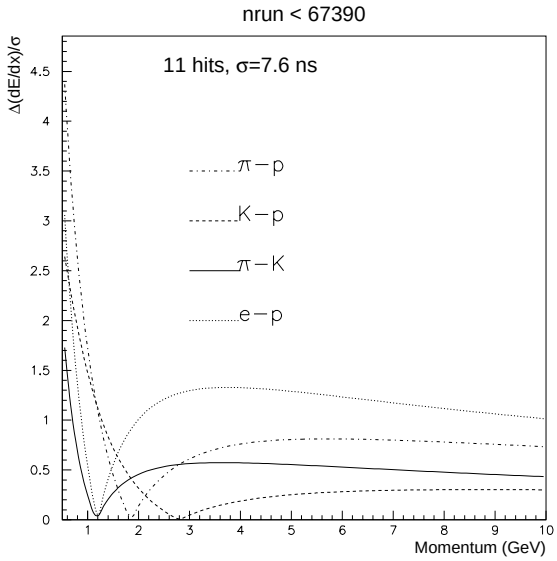


Figure 3.10: QCTC separation power for tracks of NCTC=11.

Chapter 4

Flavor Tagging

Flavor tagging is used to determine the b-quark content of a B meson when it is produced. A tag is a quantity that reveals such b quark content. To measure CP violation in $B^0/\bar{B}^0 \to J/\psi K_S$ or $B^0/\bar{B}^0 \to \psi(2S)K_S$ decays, we need to tag the flavor of the neutral B meson at production time. Generally, flavor tagging is done on a probabilistic basis, i.e., given a (neutral) B meson, only the probability (P) of it being b (or $\bar{b}$) quark can be estimated. Another aspect of a flavor tag is how often events are being tagged in a given sample. A typical scenario would be, for example, 40% of the events in a sample are tagged with 60% chance of being a b quark. Because of the probabilistic nature of the flavor tagging, the measured $\sin 2\beta$ value from a sample is the true $\sin 2\beta$ diluted, i.e., $\sin 2\beta^{\mathrm{M}} = D \sin 2\beta^{\mathrm{T}}$. The factor D, called dilution, is a characteristic of the tagging method in use. Its value is usually calibrated using an independent sample. This description is suitable for tags that have discrete values. In the case where a tag has a continuous spectrum of values, a probability density function is used to describe the frequency that certain tag values occur. Then for a given tag value, the probability of it being a b quark is also a function of the tag value.

Several flavor tagging methods have been implemented in the CDF detector environment.

4.1 Same-Side Tag

The same-side tag uses the charge of the pion from the fragmentation of the same-side (trigger-side) B to tag its flavor [13]. That is, a π^+ is more likely to accompany a B^0 in the process.

Therefore, the charge of the pion that is "closest" to the neutral B meson may tag its flavor. The use of the same-side tag at CDF is detailed in [14] and [15]. We will use the existing same-side tag in this analysis.

4.2 Soft-Lepton Tag

The soft-lepton tag uses the charge of the lepton from the opposite-side B semileptonic decay to tag the flavor of the same-side B. In B semileptonic decay $b \to cl\nu$, the charge of the lepton l can be used to tag the b quark on the opposite side. Since the B's are produced in pairs and hence have opposite flavors initially, the trigger-side B flavor is tagged. The use of the soft-lepton tag at CDF is detailed in [16] and [17]. We will use the existing CDF soft-lepton tagging procedure in this analysis.

4.3 Jet-Charge Tag

The jet-charge tag uses the charge of the opposite-side B jet to tag its flavor, from which the flavor of the same-side B is inferred. If the charged jet is from a b quark, it tends to have a negative

charge. Since the b quarks are produced in pairs, the trigger-side B flavor at production is the opposite. The use of the jet-charge tag at CDF is detailed in [18][19]. The jet-charge tag involves clustering tracks into jets and finding the tag jet. In this analysis, we will improve the jet-charge tag by combining more tags with it. That is, the jet-charge tag itself remains unchanged.

4.4 Kaon Tag

The identification of the kaon track in a B decay can be used to tag the B flavor due to the decay chain $b \to c \to s$. That is, knowing the kaon charge one can reveal the opposite-side B flavor and hence the trigger-side B flavor. The kaon identification is achieved using CTC dE/dx information on a probabilistic basis, since such information does not provide enough precision to separate kaon tracks from other charged tracks on a track-by-track basis.

4.5 Neutral Tag

Some kinematical information, unable to tag the B flavor when used alone, may improve flavor tagging when combined with a flavor tag. In opposite-side tagging, such information is selected, refined and formulated into tags, then combined with the jet-charge tag. The result is a better tagging power ϵD^2. Embedded in the combining algorithm we will address in later sections, these neutral tags improve the overall tagging not by directly distinguishing b quark from $\bar{b}$ quark, but rather by distinguishing true opposite-side $b/\bar{b}$-jets from non-$b/\bar{b}$ jets.

Chapter 5

An Algorithm of Combining Tags

5.1 Introduction

In the previous CDF $\sin 2\beta$ analysis [23] three flavor tags were used (same-side pion tag, opposite-side soft-lepton tag, opposite-side jet-charge tag); they were defined to minimize any correlation between them. Jet charge used the opposite-side B-jet to tag the flavor of the same-side B meson. In this analysis, we improve the opposite-side-jet tag while leaving the same-side tag and the soft-lepton tag unchanged.

Here, the jet charge (Q_{jet}) is calculated exactly as specified in [18]. However, we will add information related to the opposite-side jet to improve its performance; this information does not necessarily tag the B flavor by itself. In particular, we look for information that reveals the quality of the jet (is it more or less like a B jet?) as well as the particle flavor information. Here, we formulate all such information into tags and combine them.

Unlike in the previous analysis, the tags being considered here are highly correlated since they

are all built upon the same clustering algorithm. For this reason, we have developed an algorithm that can cope with tag correlations to combine these tags.

In this analysis, tags that do not tag the B flavor by themselves are called neutral tags, and those which can tag the B flavor are called flavor tags. The latter includes not only the traditional flavor tags, such as the jet-charge tag and the soft lepton tag, but also any quantity that fits the definition. One example of the latter case is the signed dilution.

Given a sample of N events and a tagger, let $n(b)$ be the number of times an event is tagged as a b-quark event and $n(\bar{b})$ the number of times an event is tagged as a $\bar{b}$-quark event. The probability of tagging a b-quark event from this sample is then $P(b) = \frac{n(b)}{N}$. In flavor tagging, dilution is defined as $D = 2P - 1$. From its definition, $D \in [-1, 1]$. If the tagger is not biased, i.e., has the same chance of tagging b and $\bar{b}$ events, then $D = 0$. If the tagger is completely biased, i.e., tag only one type of events, then $D = 1$ for tagging b events only and $D = -1$ for tagging $\bar{b}$ events only. Generally, with more than one tag, the signed dilution

$$D = \frac{n(b) - n(\bar{b})}{n(b) + n(\bar{b})} \tag{5.1}$$

is a function of a set of tagging quantities (Q_i), i.e. $D = D(Q_1, Q_2, \ldots)$.

5.2 The Algorithm

In a sample of N events, suppose two tagging quantities, Q_1 and Q_2, are available. We will use $\oplus$ for the particular operation that combines two tags in this algorithm. It is convenient to make all tagging quantities in the range $[-1, 1]$. The central idea of this algorithm is to combine two tagging quantities into one using a dilution matrix.

The construction of a dilution matrix requires each event's flavor be known. Monte Carlo events are used for this purpose. It proceeds as follows:

1. Calculate Q_{1i} and Q_{2i} for each event i, $i = 1, \ldots, N$;

2. Populate events in the Q_1-Q_2 plane according to their coordinates (Q_1, Q_2);

3. Quantize the Q_1 and Q_2 axes into m finite bins to obtain a grid in Q_1-Q_2 plane;

4. Calculate the signed dilution in each bin (i, j): $D_{ij} = \frac{n_{ij}(b) - n_{ij}(\bar{b})}{n_{ij}(b) + n_{ij}(\bar{b})}$; D_{ij} is the (i, j) element of a matrix (dilution matrix M_{12});

5. Fit the dilution matrix M_{12} to a dilution function $f = D_{12}(Q_1, Q_2)$;

6. Now we have combined Q_1 and Q_2 into one combined tag $Q_{12} = D_{12}(Q_1, Q_2)$, the signed dilution.

To calculate a combined tag Q_{12} for an event is relatively easy:

1. Calculate Q_1 and Q_2 for this event;

2. $Q_{12} = D_{12}(Q_1, Q_2)$.

It is important to know that Q_1 and Q_2 can be combined tags themselves. By repeatedly using the procedure, one can combine an arbitrary number of tags. This algorithm takes two tags as input and returns one tag as output in each step. The output tag is a signed dilution, while the input tags can be either the usual tagging quantities like the jet charge or signed dilutions. Indeed, a signed dilution *is* a tagging quantity. Q_{jet} is a tagging quantity only because given its value a signed dilution can be derived, while a signed dilution is a tagging quantity for which this derivation

is trivial (has already been performed). Therefore we can combine a signed dilution Q_{12} found after the previous step $(Q_1 \oplus Q_2)$ and a third tagging quantity Q_3 to give another signed dilution Q_{123}.

The fact that at any given moment only two tags are being combined leaves flexibility in the order of tag association, e.g. $(Q_1 \oplus Q_2) \oplus Q_3$ or $Q_1 \oplus (Q_2 \oplus Q_3)$. Whether one particular order is more favorable will be addressed in Section 5.5.

The final signed dilution is obtained after all tagging quantities are combined and is denoted as Q_{final}.

Once dilution matrices are constructed and dilution functions fitted, given n individual tags $(\{Q_i\},\ i = 1, \ldots, n)$, Q_{final} for each data event can be calculated by stepping through the dilution functions in the same order the matrices are constructed. The following is the procedure that is used:

- M_{12}: $Q_1 \oplus Q_2 \rightarrow Q_{12}$
- $M_{12,3}$: $Q_{12} \oplus Q_3 \rightarrow Q_{12,3}$
- $M_{1\ldots n-1,n}$: $Q_{1\ldots n-1} \oplus Q_n \rightarrow Q_{1\ldots n}, \quad Q_{\text{final}} \equiv Q_{1\ldots n}$

An alternative tag-combination scheme would be to combine the n tags in one step using an n-dimensional dilution matrix. While clearly optimum, this scheme is impractical since it would require orders of magnitude more Monte Carlo events (and years of CPU time) to compute the n-dimensional dilution matrix (with $m^n \simeq 10^7$ bins) to sufficient accuracy.

5.3 The Tags

We will start with a brief overview of the jet-charge tag. Then we will add new tags to the jet-charge tag to increase the final ϵD^2 using this algorithm. There are two classes of tags that are of particular interest, jet-related neutral tags and dE/dx related tags.

5.3.1 Jet-Charge Tag

The jet-charge tag uses the charge of the opposite-side B-jet to tag its flavor, i.e. a positive charge indicates a $\bar{b}$ quark. The steps are the same as in [18].

In forming a candidate set of tracks, a cone in η-ϕ space centered around the reconstructed B event is excluded to eliminate non-opposite-side tracks. Then higher p_T (> 1.75 GeV) tracks are selected as seeds, which are merged to minimize pseudo jet mass (pseudo, because in each stage of merging, the jet and track masses are set to zero) until its upper limit ($\sqrt{24}$ GeV) is reached. When the seed-merging is over—a few jets may emerge—other qualified tracks are used in the same way if there are more. Tracks are merged with a jet until the pseudo jet mass upper limit is met—in each step of merging, the track that yields the minimal pseudo jet mass is chosen. This process continues until right before the upper limit is reached or all qualified tracks are used. After the jet clustering, if there is more than one jet formed, only one jet is picked. The jet that has the lowest jet probability [24] (probability of being associated with the primary vertex, defined using impact parameters of component tracks) below 0.2 is selected. Otherwise a jet with the highest p_T is selected. Then the jet charge is calculated as

$$Q_{\text{jet}} = \frac{\sum\limits_{\text{tracks}'} q_i p_{Ti}(2 - T_{pi})}{\sum\limits_{\text{tracks}'} p_{Ti}(2 - T_{pi})}, \tag{5.2}$$

where q_i is the charge of the ith track, T_{pi} the track probability, and the summation is over all tracks with $p_T > 0.75\,\text{GeV}/c$.

5.3.2 Jet-related Neutral Tags

In the jet-charge tag, once Q_{jet} is calculated, information for selecting the jet is not used further. Although this information may not reveal the B flavor, it may differentiate the quality of the opposite side jet.

In the jet-charge tag, the number of tracks used in the jet cannot tag the flavor of the B-jet and so we refer to that information as flavor-neutral. Among jets with the same jet charge, the ones with more tracks are likely to be better tagged (have greater $|D|$) because they are less affected by jet clustering defects or errors.

Similar arguments can be applied to other properties of the jet. A jet that is more back-to-back against the trigger B in the azimuthal direction is more likely to be the true opposite-side B-jet. A jet with a smaller jet probability is less likely to have come from the primary vertex and is thus more likely to be a B-jet. A jet with a smaller cone half-angle is more "jet-like" geometrically.

These jet properties are quantified and stretched so as to give an approximately flat distribution in the range of $[-1, 1]$. They are listed as follows:

- $Q_{\Delta\phi}$: calculate the azimuthal $\Delta\phi$ angle between the jet and trigger B, $\Delta\phi$ should be within

$[0, \pi]$, and the tag is defined as

$$Q_{\Delta\phi} = 2\Delta\phi/\pi - 1. \tag{5.3}$$

- $Q_{n_{\text{used}}}$: count the number of tracks used in the jet-charge calculation n_{used}, the tag is defined as

$$Q_{n_{\text{used}}} = \frac{2}{m}(\min(n_{\text{used}}, m-1) + \frac{1}{2}) - 1, \tag{5.4}$$

where m is the number of bins. This arrangement is only to fit $Q_{n_{\text{used}}}$ into $[-1, 1]$; the bin index still equals the value of n_{used}.

- Q_{jp}: jet probability [24] is the probability that if a jet comes from the primary vertex, that the jet will be found greater than that distance from the primary vertex.

The tag is defined as

$$Q_{jp} = \frac{2}{(1 - \frac{1}{2}\log(\text{jetprob}))} - 1. \tag{5.5}$$

In the jetprob distribution, a large number of jets (those containing tracks from B decay) pile up in the region near zero. The logarithm is used in the expression to stretch that particular segment over several bins.

- Q_{sprd}: calculate jet spread as

$$\text{jetsprd} = \frac{\sum_{i=1}^{n} \cos(\Delta\alpha_i)}{n}, \tag{5.6}$$

where α is the angle between the track and the jet direction and n is the total number of tracks in the jet. We define

$$Q_{\text{sprd}} = 2 \cdot \text{jetsprd} - 1. \tag{5.7}$$

5.3.3 dE/dx Tag

Besides jet-related neutral tags, we also make use of the CTC dE/dx information to implement a kaon tag. Kaons from B decays can be used to tag the B flavor due to the decay chain $b \to c \to s$. This requires that kaons be identified among other tracks. For CDF Run 1, CTC dE/dx is the designated tool for the charged particle identification [25]. We are able to calculate the probability of a track being a kaon (kprob) using the dE/dx information. The kaon tag is defined as follows:

The CTC dE/dx information is used to calculate the kaon probability, which is defined as

$$\text{kprob} = \frac{f_K \cdot G(QCTC - Q_0(\frac{p}{m_K}); \sigma)}{\sum_{i=1}^{5} f_i \cdot G(QCTC - Q_0(\frac{p}{m_i}); \sigma)}, \tag{5.8}$$

where the index i runs over proton, kaon, pion, muon and electron,

$$G = G(x; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x}{\sigma})^2}, \tag{5.9}$$

Q_0 is the predicted mean dE/dx and is a function of $\beta\gamma$ as shown in Fig. 3.8, σ is the dE/dx resolution, p is the momentum of the particle, f_i is the percentage of the ith species in J/ψ data, and f_K is the kaon percentage. We obtain the f_i's from our Monte Carlo samples. m_i is the mass of particle i. The tag is defined as

$$Q_K = \sum_j \text{kprob}_j^2 \cdot q_j \tag{5.10}$$

where index j runs over all tracks in the jet, and q_j is the charge of the jth track.

5.4 Optimization Using the Monte Carlo Sample

5.4.1 Monte Carlo Sample

There are several benefits of using Monte Carlo samples to optimize our algorithm:

- the true $b/\bar{b}$ flavor is known;
- sufficient number of events to minimize statistical fluctuations;
- no background events;
- the data can be reserved for calibration.

We use Monte Carlo $B^0/\bar{B}^0 \to J/\psi K_S$ events to optimize the tags and to construct the dilution matrices. Tuned PYTHIA 5.702 [26] is used for event generation and CLEOMC (QQ9.0) is used to decay the particles. Since the clustering is tracking based, we can simulate the detector behavior without using CDFSIM. We also simulate the SVX acceptance using simple geometrical parameters. The B lifetime is simulated and track impact parameters are smeared with the run 1 SVX resolution.

On average, only 3×10^{-5} of generated B events will decay to $\mu^+\mu^-\pi^+\pi^-$ through $B^0/\bar{B}^0 \to J/\psi K_S$ without forcing the decay. Therefore we force $B^0/\bar{B}^0 \to J/\psi K_S$ decays to speed up generation while at the same time being careful to keep the decay spectrum of the opposite-side B unaffected:

Events with at least one neutral B meson are candidates for forcing a $B^0/\bar{B}^0 \to J/\psi K_S$ decay, all others are discarded. Mixing is turned on for all neutral B mesons in CLEOMC: for $B^0/\bar{B}^0$ mixing, $x_d = 0.723$ is used; for $B^0_S/\bar{B}^0{}_S$ mixing, $x_s = 15.0$ is used. When there is only one $B^0/\bar{B}^0$ produced, we force it to decay to $\mu^+\mu^-\pi^+\pi^-$ through $J/\psi K_S$ and let the other B product decay naturally.

If an event has a B^0 and a $\bar{B}^0$, we have to choose one as the trigger and let the other one decay naturally. Thus a $B^0/\bar{B}^0$ meson has a 50% probability to be the trigger in double neutral B events, while in single $B^0/\bar{B}^0$ events it is always the trigger. This is quite different from simply picking out

$B^0/\bar{B}^0 \to J/\psi K_S$ events, which is more natural but too slow. To assign the same opportunity to the forced $B^0/\bar{B}^0 \to J/\psi K_S$ decay in both the single-B event and the double-B event, we throw away at random a half of the single-B events.

5.4.2 Binning

We collected $\sim 250,000$ Monte Carlo events. We use 11×11 binning to construct the dilution matrices. If, during the process of filling a matrix, a bin is found to have too few entries (< 20), it is absorbed by neighboring bins.

In this analysis we choose to add new tags to the jet-charge tag in the order of Q_{nused}, Q_{jetsprd}, $Q_{\Delta\phi}$, Q_{jetprb} and Q_K.

5.4.3 Dilution Matrices and Dilution Functions

In each step of the algorithm, once a dilution matrix is constructed, a dilution function is fitted and will be used by later steps. The forms of the dilution functions are chosen empirically. The fits are checked through both the χ^2 values and the randomness of the sign of the residuals.

In the following dilution functions, the parameter vector $\vec{p}$ is expressed through its components p_i. The errors of the parameters are used for estimating the overall error of ϵD^2 in Monte Carlo data. They are not used in the calibration fit or the $\sin 2\beta$ fit.

$\boldsymbol{Q_{\text{jet}} \oplus Q_{\text{nused}}}$

Fig. 5.1 shows the dilution matrix M_{12} constructed given an event's Q_{jet} and Q_{nused}. When an event has only one track used for the jet-charge calculation, its Q_{jet} is either 1 or -1. This is seen

in the row with the lowest Q_n-value in the matrix. For fixed Q_{jet}, there is a trend that larger n_{used} gives a larger $|D|$.

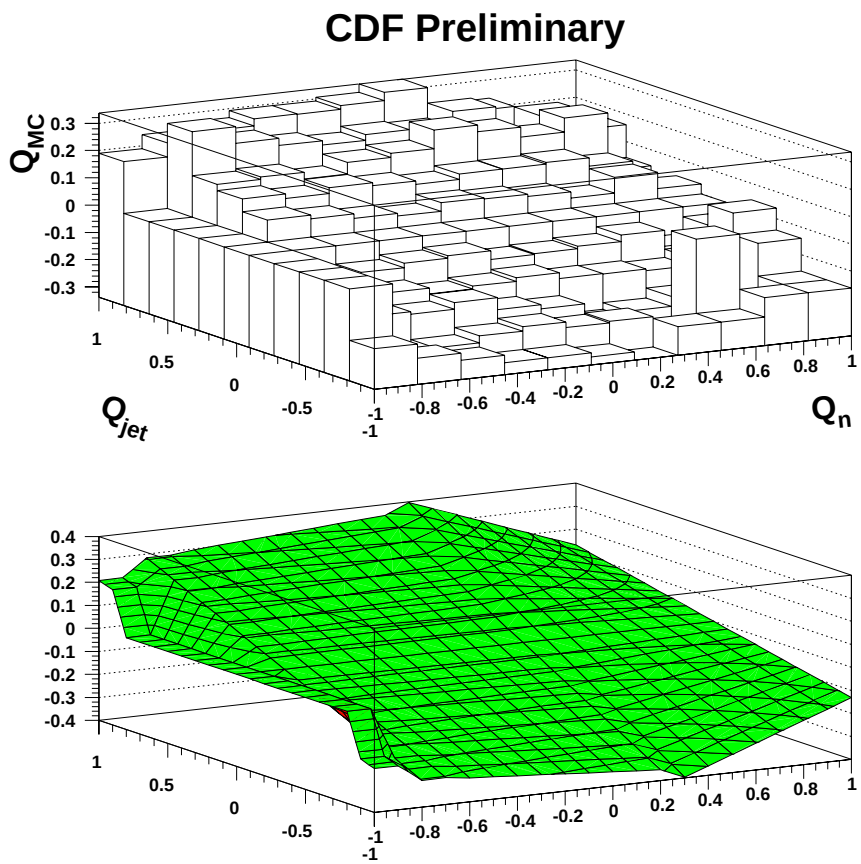


Figure 5.1: Top: The dilution matrix. Bottom: The fit.

$$Q_{\text{jet}.n_{\text{used}}} = \begin{cases} 0 & n_{\text{used}} = 1,\ |Q_{\text{jet}}| \neq 1 \\ p_4 Q_{\text{jet}} & n_{\text{used}} = 1,\ |Q_{\text{jet}}| = 1,\ p_4 = 0.208 \pm 0.011 \\ p_1 Q_{\text{jet}} & n_{\text{used}} = 2,\ p_1 = 0.283 \pm 0.009 \\ p_2 Q_{\text{jet}} + p_3 Q_{\text{jet}} Q_{n_{\text{used}}} & 3 \leq n_{\text{used}} \leq 7,\ p_2 = 0.339 \pm 0.008,\ p_3 = 0.072 \pm 0.022 \\ p_5 Q_{\text{jet}} + p_6 Q_{\text{jet}} Q_{n_{\text{used}}} & 8 \leq n_{\text{used}},\ p_5 = 0.510 \pm 0.062,\ p_6 = -0.379 \pm 0.118 \end{cases} \tag{5.11}$$

$Q_{\text{jet.nused}} \oplus Q_{\text{jetsprd}}$

The signed dilution produced from the previous step is rescaled to fully cover the range $[-1, 1]$. This new quantity is now the "base charge", ready to combine with Q_{jetsprd} to form a new matrix $M_{12,3}$, shown in Fig. 5.2. Rows with larger x-values are events with jets whose tracks cluster closer in the direction of the jet axis; they also have better dilutions.

$$Q_{\text{jet}.n.\text{jd}} = p_1 Q_{\text{jet}.n} + p_2 Q_{\text{jet}.n} Q_{\text{jd}} + p_3 (4Q_{\text{jet}.n}^3 - 3Q_{\text{jet}.n}) + p_4 Q_{\text{jet}.n} (2Q_{\text{jd}}^2 - 1). \tag{5.12}$$

The parameters are: $p_1 = 0.292 \pm 0.008$, $p_2 = 0.089 \pm 0.010$, $p_3 = -0.027 \pm 0.004$, $p_4 = 0.024 \pm 0.009$.

$Q_{\text{jet.nused.jetsprd}} \oplus Q_{\Delta\phi}$

Repeating the process, we obtain a matrix of $Q_{\text{jet.nused.jetsprd}}$ and $Q_{\Delta\phi}$, shown in Fig. 5.3. Rows with larger x-values are events with jets more back-to-back to the trigger B. They generally have larger $|D|$ as can be seen from the matrix.

$$Q_{\text{jq}.n.\text{jd}.\Delta\phi} = p_1 Q_{\text{jet}.n.\text{jd}} Q_{\Delta\phi} + p_2 Q_{\text{jet}.n.\text{jd}}. \tag{5.13}$$

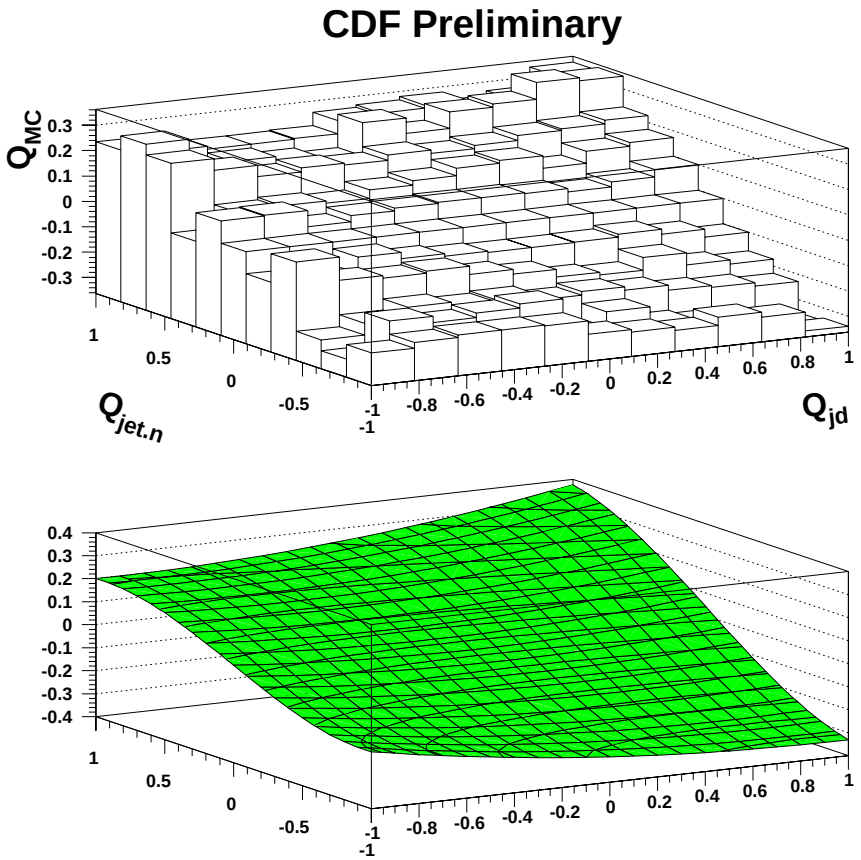


Figure 5.2: Top: The dilution matrix. Bottom: The fit.

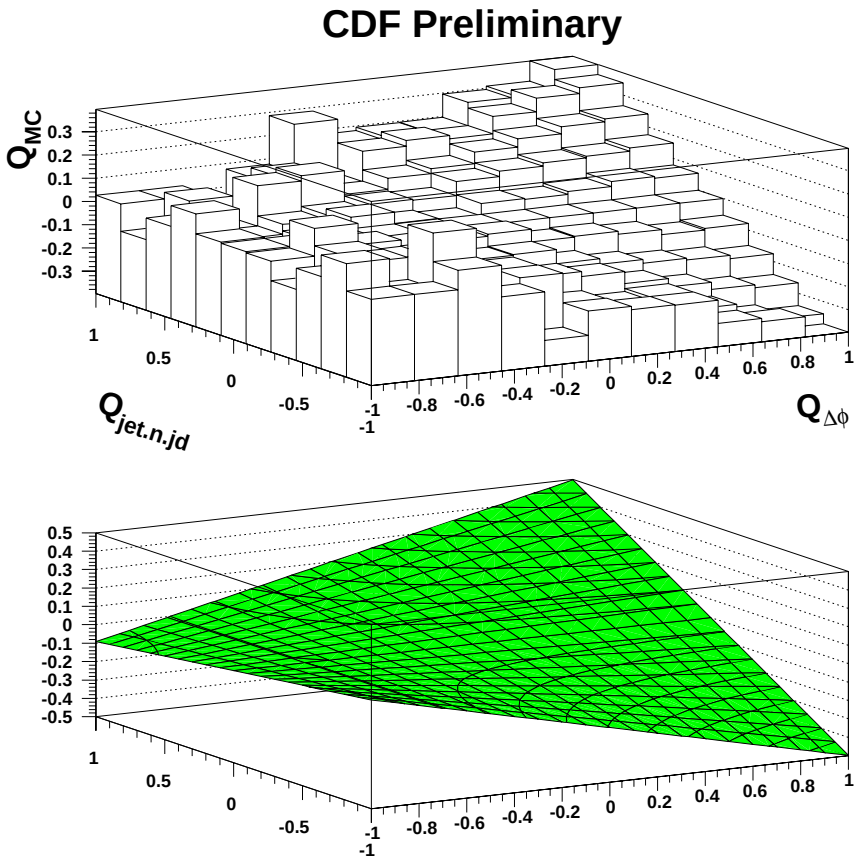


Figure 5.3: Top: The dilution matrix. Bottom: The fit.

The parameters are: $p_1 = 0.295 \pm 0.012$, $p_2 = 0.203 \pm 0.008$

$\boldsymbol{Q_{\text{jet.nused.jetsprd.}\Delta\phi} \oplus Q_{\text{jetprb}}}$

We next add Q_{jetprb} to $Q_{\text{jet.nused.jetsprd.}\Delta\phi}$. In Fig. 5.4 Rows with larger x-values are events with jet probabilities closer to 1, and rows with smaller x-values are events with jet probabilities closer to 0. Events with smaller jet probability, i.e. far away from the primary vertex, tend to have a greater $|D|$.

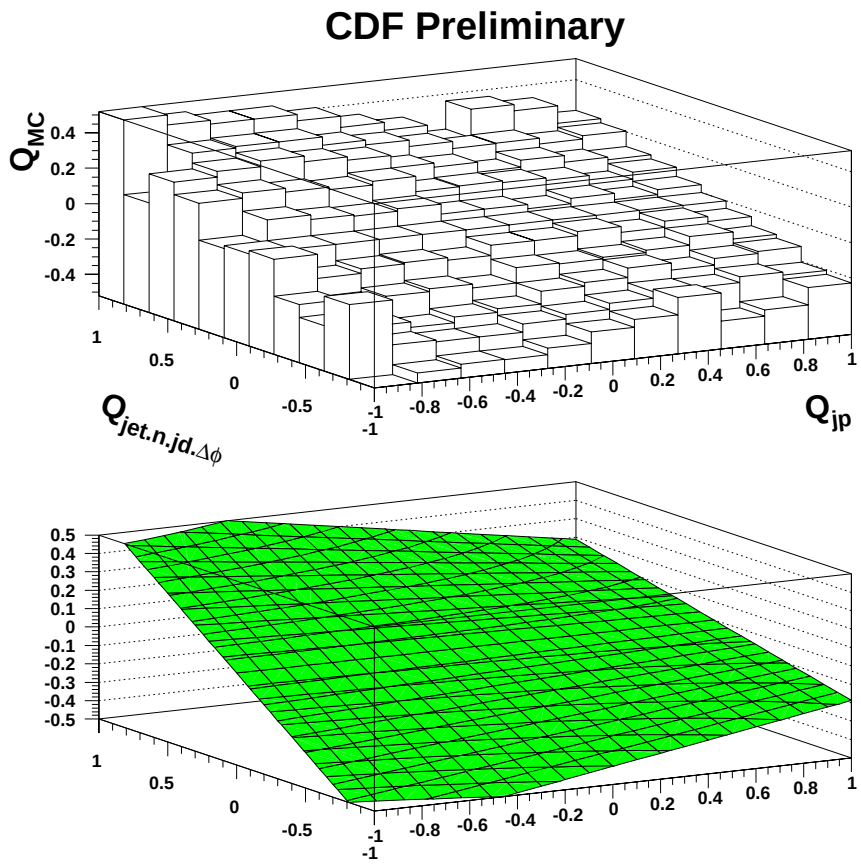


Figure 5.4: Top: The dilution matrix. Bottom: The fit.

$$Q_{\text{jq.}n\text{.jd.}\Delta\phi\text{.jp}} = p_1 Q_{\text{jet.}n\text{.jd.}\Delta\phi} Q_{\text{jp}} + p_2 Q_{\text{jet.}n\text{.jd.}\Delta\phi}. \tag{5.14}$$

The parameters are: $p_1 = -0.214 \pm 0.010$, $p_2 = 0.401 \pm 0.005$.

$\boldsymbol{Q_{\text{jet.nused.jetsprd.}\Delta\phi\text{.jetprb}} \oplus Q_K}$

The dilution matrix is shown in Fig. 5.5. Notice the matrix is diagonally symmetric instead of left-right symmetric as in the previous cases, because the kaon tag is a symmetric (anti-symmetric, $D(-Q) = -D(Q)$) tag just like the jet-charge tag. The dilutions achieve greater amplitudes when the two charges (Q_K and the base charge) more or less agree, i.e., around the primary diagonal line (from $(-1,-1)$ to $(1,1)$) of the matrix. The opposite happens when events fall along the other diagonal line. This is because the kaon tag uses the same flavor-charge relation as the jet-charge tag. When Q_K disagrees with $Q_{\text{jet.nused.jetsprd.}\Delta\phi\text{.jetprb}}$, the latter dominates the sign of the dilutions.

$$Q_{\text{jq.}n\text{.jd.}\Delta\phi\text{.jp.}K} = p_1 Q_K + p_2 Q_{\text{jq.}n\text{.jd.}\Delta\phi\text{.jp}}. \tag{5.15}$$

The parameters are: $p_1 = 0.158 \pm 0.007$, $p_2 = 0.506 \pm 0.006$.

Given the individual tags, the final quantity returned from calling the dilution functions is the Monte Carlo predicted dilution. Each data event with the individual tagging quantities calculated will traverse all the dilution functions in the same order to calculate the Monte Carlo predicted dilution Q_{final}.

5.5 Tag Association Order

Generally, the order in which the tags are combined is not associative. A more relevant question is by how much the resulting ϵD^2 from different orders differ, or if there exists a particular order that

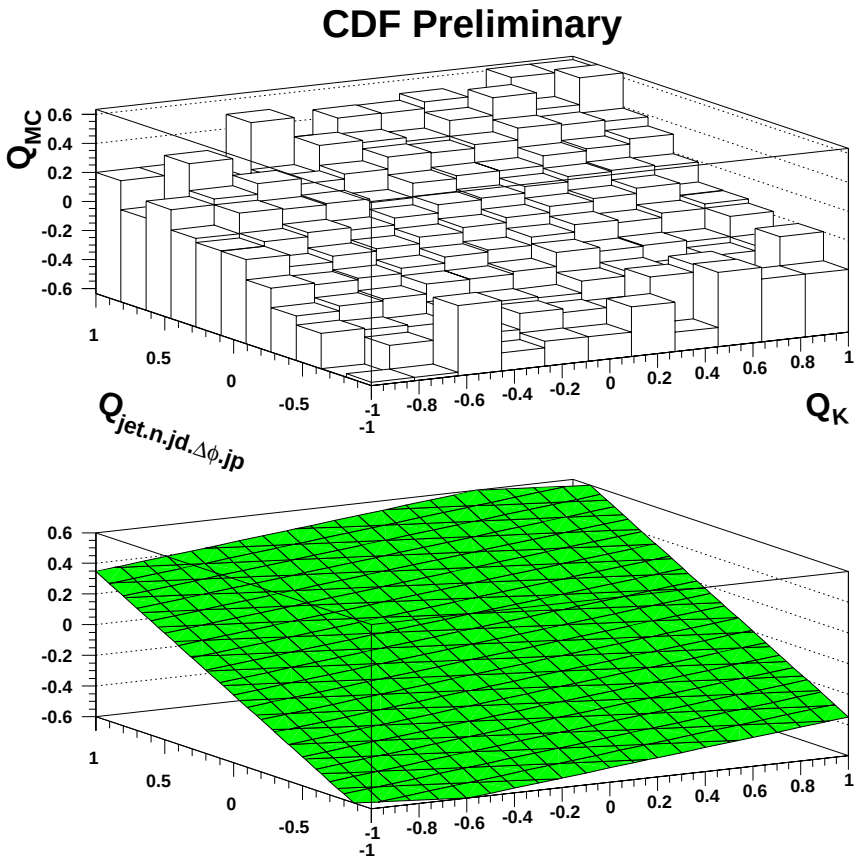


Figure 5.5: Top: The dilution matrix. Bottom: The fit.

maximizes ϵD^2. Before trying out all possible association orders, it helps that we categorize them and eliminate some that will never work.

Clearly we do not combine two neutral tags. Recall that the combined tag is a dilution quantity, by doing so we simply get a trivial dilution function $f(Q_1, Q_2) = 0$. We also do not combine two tags if the resulting global pattern (dilution function) empirically does not depend on both of them.

Given the above guideline, we compiled three association orders. The results are listed in Table 5.1.

Table 5.1: Tag Association Order

Tags	ϵD^2 (%)	Tags	ϵD^2 (%)
Q_{jet}	2.31 ± 0.09	Q_{jet}	2.31 ± 0.09
$\oplus Q_{n_{\text{used}}}$	2.54 ± 0.09	$\oplus Q_{n_{\text{used}}}$	2.54 ± 0.09
$\oplus Q_{\text{jd}}$	2.66 ± 0.09	$\oplus Q_{\text{jd}}$	2.66 ± 0.09
$\oplus Q_K$	2.75 ± 0.10	$\oplus Q_{\Delta\phi}$	2.81 ± 0.10
$\oplus Q_{\Delta\phi}$	2.82 ± 0.10	$\oplus Q_{\text{jp}}$	2.92 ± 0.10
$\oplus Q_{\text{jp}}$	3.04 ± 0.10	$\oplus Q_K$	3.22 ± 0.10

The ϵD^2s in Table 5.1 are calculated, very much like during the construction of the dilution matrices, as follows:

- step-1:for each event, a combined tag is calculated using the dilution function from fitting to the dilution matrix;

- step-2: group events according to this combined tag;

- step-3: for each group, calculate dilution using $D = \frac{n-\bar{n}}{n+\bar{n}}$;

- step-4: sum D^2 for all groups, with each group weighted by $(n + \bar{n})$.

We find no significant order dependence of this algorithm with our particular tags. The fact that we have used tags distinctive in their physical bases helps to decouple the overall ϵD^2 from the association order. In this analysis, we pick the tags and the association order in column two of Table 5.1.

5.6 Gory Details

All tagging quantities are scaled and stretched as flat as simplicity and monotonicity allow to fill in the range of $[-1, 1]$, because each bin in the dilution matrices should have enough events for calculating the dilution and its error. This effort will be seriously undermined if we do not do the same to the combined tags (the intermediate dilutions) since they are mostly in the range of $[-0.5, 0.5]$ thus leaving bins outside that range unoccupied. For this reason, the intermediate dilutions in each matrix are scaled by a factor that is set by the largest dilution in that matrix. The result is that in each matrix, the base charge always ranges from -1 to 1.

Some tags possess an intrinsic (anti)symmetry, for example, in jet charge tag, $D(q) = -D(-q)$. The kaon tag and all the intermediate dilutions also have this symmetry, while the neutral tags do not. This tagging symmetry should be reflected in the dilution matrices. Limited by the Monte Carlo sample size, the dilution matrices are not born symmetric. Therefore an explicit symmetrization is in order. In symmetrizing the dilution matrices we also reduce the statistical uncertainty in each bin. The symmetrization consists of finding the symmetric pairs of bins in the

matrix and making them satisfy the above relation. For a dilution matrix of one symmetric tag (Q_1) and one non-symmetric tag (Q_2), $D(Q_1, Q_2) = -D(-Q_1, Q_2)$. Hence the symmetric pair is (i, j) and $(m+1-i, j)$, where m is the dimension of the matrix. For a dilution matrix of two symmetric tags, $D(Q_1, Q_2) = -D(-Q_1, -Q_2)$. Hence the symmetric pair is (i, j) and $(m+1-i, m+1-j)$. The dilution for a pair is calculated as such: suppose there are n_1 b and $\bar{n_1}$ $\bar{b}$ events in bin (i, j), and n_2 b and $\bar{n_2}$ $\bar{b}$ events in its partner; define $n = n_1 + \bar{n_2}$ and $\bar{n} = \bar{n_1} + n_2$; then the dilution $D(i, j) = \frac{n-\bar{n}}{n+\bar{n}}$, the dilution of its partner is $-D(i, j)$.

5.7 Tag Correlation

We have experimented with a variety of tags to optimize the algorithm and to further our understanding. One feature of this algorithm is that it handles the tag correlation automatically—nothing is assumed about the relation of the tags being combined. Yet it is instructional to see if there exists a connection between the tag correlation and the tagging improvement.

To quantify the discussion, define the pseudo correlation between two tags as

$$\zeta = \frac{\epsilon_1 D_1^2 + \epsilon_2 D_2^2 - \epsilon_{12} D_{12}^2}{\min(\epsilon_1 D_1^2, \epsilon_2 D_2^2)}, \tag{5.16}$$

where the ϵD^2 terms are the tagging powers—they correspond to the ϵD^2 columns in Table 5.2. $\zeta = 1$ means complete overlap; $\zeta = 0$ means no correlation; and ζ being a large negative number indicates one of the tags is neutral because in case of a neutral tag the denominator is ~ 0 and the numerator is negative because the combined ϵD^2 works better than the originals (see Appendix A)..

Table 5.2 shows the result from various tags being added to the jet-charge tag. ϵD^2 terms are obtained by summing over all bins. The improvement from adding the second tag to jet-charge tag

is shown as the ratio of the combined tag ϵD^2 to that of the jet-charge tag. For one case (JETQ), we combine jet-charge tag with itself; the result, as expected, is that no improvement occurs. Due to correlations, two good tags do not necessarily produce a better combined tag, as demonstrated by the PT tag which very much resembles the jet-charge tag—with the jet probability information dropped. On the other hand, a neutral tag can improve the jet-charge tag, e.g. DEPHI ($Q_{\Delta\phi}$) boosts the ϵD^2 by 21%. As another category, KPRB2Q (Q_K) improves the jet-charge tag by about 57% while still having a 28% correlation because it is a flavor tag containing particle ID information that is not in the jet-charge tag.

5.8 Summary

This new generic algorithm combines arbitrary tags. In particular, the algorithm is not restricted to combining uncorrelated tags. Neutral tags are combined with flavor tags to improve ϵD^2. Using this algorithm, in later chapters we will improve the opposite-side-jet tag by adding the jet-kinematics tags and the kaon dE/dx tag.

Table 5.2: Adding a Second Tag to Jet Charge Tag (part 1 of 3)

2nd Tag	Formula	ϵD^2 (%)			Ratio	ζ
		Jet Q Tag	2nd Tag	Combined		
ALL1	$\frac{\sum k_p p_T e^{1-t_p} q}{\sum k_p p_T e^{1-t_p}}$	2.58 ± 0.16	1.70 ± 0.13	3.26 ± 0.18	1.26	0.60
DEALPHA	$\cos \alpha_B^{\rm jet}$	2.36 ± 0.15	0.02 ± 0.02	2.71 ± 0.16	1.15	-13.78
DEETA	$\frac{2}{1+\lvert\frac{\eta_{\rm jet}+\eta_B}{\eta_{\rm jet}-\eta_B}\rvert} - 1$	2.37 ± 0.15	0.02 ± 0.01	2.65 ± 0.16	1.12	-17.23
DEPHI	$2\frac{\Delta\phi}{\pi} - 1$	2.45 ± 0.15	0.02 ± 0.01	2.96 ± 0.17	1.21	-27.61
DR2	$\frac{2}{1+(\Delta\eta^2+\Delta\phi^2)} - 1$	2.49 ± 0.16	0.02 ± 0.01	2.76 ± 0.16	1.11	-16.01
JETETA	$\frac{2}{1+\lvert\eta_{\rm jet}\rvert} - 1$	2.35 ± 0.15	0.01 ± 0.01	2.63 ± 0.16	1.12	-17.40
JETKPRB	$2 \cdot j_p^K - 1$	2.38 ± 0.15	0.03 ± 0.01	2.43 ± 0.15	1.02	-0.95
JETM	$2(\frac{M_{\rm jet}^2}{24})^2 - 1$	2.48 ± 0.16	0.03 ± 0.02	2.77 ± 0.16	1.12	-8.64
JETPRB	$\frac{2}{1-\frac{1}{2}\log(j_p)} - 1$	2.52 ± 0.16	0.03 ± 0.02	3.01 ± 0.17	1.19	-13.60
JETPRB2	$2j_p - 1$	2.49 ± 0.16	0.03 ± 0.02	2.93 ± 0.17	1.17	-16.34
JETPT	$1.44\log(p_T^{\rm jet}/5) - 1$	2.33 ± 0.15	0.02 ± 0.02	2.50 ± 0.15	1.07	-5.72
JETQ	$\frac{\sum p_T(2-t_p)q}{\sum p_T(2-t_p)}$	2.36 ± 0.15	2.36 ± 0.15	2.36 ± 0.15	1.00	1.00
JETQLIKE	$\frac{\sum k_p p_T(2-t_p)q}{\sum k_p p_T(2-t_p)}$	2.49 ± 0.16	1.49 ± 0.12	3.11 ± 0.17	1.25	0.58
JETSPRD	$\frac{\sum \cos \alpha_i^{\rm jet}}{\sum 1}$	2.31 ± 0.15	0.01 ± 0.01	2.48 ± 0.16	1.07	-13.32

Table 5.2: Adding a Second Tag to Jet Charge Tag (part 2 of 3)

2nd Tag	Formula	ϵD^2 (%)			Ratio	ζ
		Jet Q Tag	2nd Tag	Combined		
KPRB	$2\max(\{k_p\}) - 1$	2.57 ± 0.16	0.03 ± 0.02	2.89 ± 0.17	1.12	−9.50
KPRB2Q	$\sum k_p^2 q$	2.45 ± 0.15	1.96 ± 0.14	3.86 ± 0.19	1.57	0.28
KPRBCUT	KPRBQ, $k_p > 0.2$	2.53 ± 0.16	1.83 ± 0.13	3.71 ± 0.19	1.47	0.36
KPRBPTQ	$\sum k_p p_T q$	2.46 ± 0.16	2.16 ± 0.14	3.58 ± 0.19	1.45	0.48
KPRBQ	$\sum k_p q$	2.55 ± 0.16	2.14 ± 0.14	3.98 ± 0.19	1.56	0.33
KPRBQRN	KPRBQ $\cdot$ randomsign	2.33 ± 0.15	0.03 ± 0.02	2.59 ± 0.16	1.11	−9.29
KPRBTRKP	$\sum k_p(2 - t_p)q$	2.44 ± 0.15	1.54 ± 0.12	3.16 ± 0.17	1.29	0.53
KPRBV	$\frac{\sum(1-k_p)q}{\sum(1-k_p)}$	2.49 ± 0.16	0.28 ± 0.05	2.83 ± 0.16	1.13	−0.18
NJET	number of jets	2.35 ± 0.15	0.01 ± 0.01	2.40 ± 0.15	1.02	−8.36
NKAON	n of kaons ($k_p > 0.999$)	2.58 ± 0.16	0.02 ± 0.01	2.63 ± 0.16	1.02	−2.25
NONE	$\sum q_i$	2.54 ± 0.16	1.59 ± 0.13	2.76 ± 0.16	1.09	0.86
NTRK	n of tracks in jet	2.42 ± 0.15	0.03 ± 0.02	2.64 ± 0.16	1.09	−7.42
NUSED	n of tracks used	2.45 ± 0.15	0.02 ± 0.01	2.68 ± 0.16	1.09	−11.12
ONEKAONQ	q of $\max(\{k_p\})$	2.45 ± 0.15	1.31 ± 0.11	3.09 ± 0.17	1.26	0.51

Table 5.2: Adding a Second Tag to Jet Charge Tag (part 3 of 3)

2nd Tag	Formula	ϵD^2 (%)			Ratio	ζ
		Jet Q Tag	2nd Tag	Combined		
PPRB2Q	$\sum p_p^2 q$	2.51 ± 0.16	0.30 ± 0.05	2.85 ± 0.17	1.13	-0.13
PPRBCUT	PPRBQ, $p_p > 0.2$	2.30 ± 0.15	0.17 ± 0.04	2.71 ± 0.16	1.18	-1.33
PPRBPTQ	$\sum p_p p_T q$	2.66 ± 0.16	0.20 ± 0.04	3.01 ± 0.17	1.13	-0.79
PPRBQ	$\sum p_p q$	2.59 ± 0.16	0.20 ± 0.04	2.98 ± 0.17	1.15	-0.95
PT	$\frac{\sum p_T q}{\sum p_T}$	2.49 ± 0.16	2.36 ± 0.15	2.59 ± 0.16	1.04	0.96
PT2	$\frac{\sum p_T^2 q}{\sum p_T^2}$	2.52 ± 0.16	2.30 ± 0.15	2.65 ± 0.16	1.05	0.94
REPT	$\frac{\sum q/p_T}{\sum 1/p_T}$	2.47 ± 0.16	0.87 ± 0.09	2.73 ± 0.16	1.11	0.70
SSPT	$1.44 \log(p_T^B/5) - 1$	2.35 ± 0.15	0.01 ± 0.01	2.58 ± 0.16	1.10	-20.68
TRKP	$\frac{\sum (2-t_p)q}{\sum (2-t_p)}$	2.41 ± 0.15	1.64 ± 0.13	2.62 ± 0.16	1.09	0.87
TRKP3	$\frac{\sum t_p q}{\sum t_p}$	2.38 ± 0.15	0.60 ± 0.08	2.70 ± 0.16	1.13	0.47
TRKP4	$\frac{\sum (1-t_p)q}{\sum (1-t_p)}$	2.51 ± 0.16	1.13 ± 0.11	2.83 ± 0.16	1.13	0.72
TRKP5	$\frac{\sum (1-\log(\max(t_p,\epsilon)))q}{\sum (1-\log(\max(t_p,\epsilon)))}$	2.47 ± 0.16	1.91 ± 0.14	2.67 ± 0.16	1.08	0.89
WORST	$\frac{\sum \mathrm{t_p} q/p_T}{\sum \mathrm{t_p}/p_T}$	2.34 ± 0.15	0.31 ± 0.06	2.71 ± 0.16	1.16	-0.19

Chapter 6

Measuring $\sin 2\beta$

6.1 Introduction

In a previous CDF analysis [21], the Standard Model CP-violation parameter $\sin 2\beta$ was measured in $B^0/\bar{B}^0 \to J/\psi K_S$ decay. Three tagging methods were used in that analysis: a same-side tag, a soft-lepton tag and a (opposite-side) jet-charge tag.

In this analysis, we have improved the opposite-side-jet tagging by combining the jet-charge tag with information from jet kinematics and a kaon dE/dx tag. The jet-kinematics information includes characteristics of a jet, such as the number of tracks used by the jet-charge tag, the azimuthal angle between the opposite-side jet and the trigger B, the jet probability and the half-angle of the jet. For generality, these quantities are treated as tags, referred to as "neutral" tags, which, by themselves, have no flavor tagging power. The kaon tag uses CTC dE/dx information to evaluate the weight (probability of being a kaon) of each track in the jet, returning a weighted sum of track charges as the result. To make use of all the tags in a manner that maximizes the overall

ϵD^2, we have developed a generic algorithm to combine them. The algorithm is detailed in [22]. Nonetheless, for the completeness of this note, an overview of the algorithm and some updates are addressed here.

In this analysis, similar to [23], tagging efficiency and dilution are both functions of the tag. However, instead of having discrete tag values of $+1$, 0 and -1, the tag Q_{final} is now a continuous variable between $[-1, 1]$. A good reference for a generalized discussion of tagging can be found in [20].

The same-side tag and the soft-lepton tag remain the same as in the previous analysis.

6.2 Tagging With Asymmetry

Symmetric tagging is at best an approximation in an experimental setup. Therefore, we naturally would like to deal with a tagging asymmetry. In the previous analysis [23], since all tags take discrete values, tagging asymmetry was reflected in the different efficiencies and dilutions for $+$, $-$, and 0 bins. Due to the tagging asymmetry, the 0 bin ($|Q_{\text{jet}}| < 0.2$) has non-zero dilutions. In this analysis, with continuous Q_{final}, tagging with asymmetry involves finding the proper forms to parameterize ε and D as functions of Q_{final}. The tagging asymmetry should be incorporated into the functions.

6.2.1 Efficiency and Dilution Functions

The Q_{final} distribution in the inclusive J/ψ events, $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^\pm \to J/\psi K^\pm$ events can be well-summarized by a concentric double Gaussian distribution—a symmetric function. To allow

the tagging efficiency to be asymmetric, an anti-symmetric component is needed. We use the odd Chebyshev polynomials for this anti-symmetric component. The efficiency function, the normalized Q_{final} distribution, takes the form

$$\varepsilon(Q) = \epsilon_0 \delta(Q) + (1-\epsilon_0)\left(\frac{f}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\frac{Q^2}{\sigma_1^2}} + \frac{1-f}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\frac{Q^2}{\sigma_2^2}}\right)\left[1 + c_1 T_1(2Q) + c_3 T_3(2Q)\right], \tag{6.1}$$

where $\delta(Q)$ is a δ-function representing events which do not have jets. The efficiency asymmetry is described by the two odd Chebyshev polynomials T_1 and T_3 (multiplied by the double Gaussian). Since Chebyshev polynomials are orthogonal over $[-1, 1]$ and the Monte Carlo predicted dilution Q_{final} is mostly in $[-0.5, 0.5]$, we scaled the variable Q by a factor of 2 in T_i's. Their coefficients c_1 and c_3 are determined from the inclusive J/ψ events.

For a symmetric dilution function, $D(-Q) = -D(Q)$. With the same logic as in the efficiency function, the dilution asymmetry is handled by adding even Chebyshev polynomials. Selecting the lowest order polynomial—the 0th order—we have the dilution function

$$D(Q) = \alpha Q + \beta, \tag{6.2}$$

where β is the coefficient of the Chebyshev polynomial $T_0 = 1$. Rather than being a free parameter, β is related to other tagging parameters in both the efficiency and the dilution functions, as shown in the next section.

6.2.2 A Relation Between Efficiency and Dilution Asymmetry

Note that equal numbers of b-quarks and $\bar{b}$-quarks are produced (associated-production), and that there is no charge bias in the J/ψ data set. That is, the events trigger on either di-muon or a single muon with very high p_T, neither presents a charge bias; Moreover, the triggers are not used by the

tagging algorithms. Therefore, the prior probability of starting with a b-quark is 50%. That is, the tagging efficiency and dilution functions satisfy the following relation

$$\int_{-1}^{1} \varepsilon(Q)P(Q)dQ = \frac{1}{2}, \tag{6.3}$$

where $\varepsilon(Q)$ is the normalized efficiency function (a probability density function). $\varepsilon(Q)dQ$ gives the probability of observing $Q_{\text{final}} \in [Q, Q+dQ)$. $P(Q)$ is the probability that the trigger-B in an event with $Q_{\text{final}} = Q$ contains a b-quark, and is related to the signed dilution via $D = 2P - 1$. We have the following relation after the substitution

$$\int_{-1}^{1} \varepsilon(Q)D(Q)dQ = 0. \tag{6.4}$$

This relation is referred to as the beauty conservation rule in the text that follows.

The parameters in $\varepsilon(Q)$ (ϵ_0, f, σ_1, σ_2, c_1 and c_3) and the parameters in $D(Q)$ (α and β) are constrained by the integral relation addressed above. Evaluating that integral, we have

$$\begin{aligned}
0 &= \int_{-1}^{1} \left\{ \epsilon_0\delta(q) + (1-\epsilon_0)\left(\frac{f}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\frac{q^2}{\sigma_1^2}} + \frac{1-f}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\frac{q^2}{\sigma_2^2}} \right) [1 + c_1T_1(2q) + c_3T_3(2q)] \right\} (\alpha q + \beta)dq \\
&= \int_{-1}^{1} \epsilon_0\delta(q)\beta dq + \int_{-1}^{1} (1-\epsilon_0)\left(\frac{f}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\frac{q^2}{\sigma_1^2}} + \frac{1-f}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\frac{q^2}{\sigma_2^2}} \right) [\beta + \alpha c_1 q T_1(2q) + \alpha c_3 q T_3(2q)]\, dq \\
&= \epsilon_0\beta + (1-\epsilon_0)\beta + (1-\epsilon_0)\alpha \int_{-1}^{1} \left(\frac{f}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\frac{q^2}{\sigma_1^2}} + \frac{1-f}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\frac{q^2}{\sigma_2^2}} \right) \left[(2c_1 - 6c_3)q^2 + 32c_3q^4 \right] dq
\end{aligned}$$

Now we are left with integrals of types $\int e^{-\frac{1}{2}x^2}x^2dx$ and $\int e^{-\frac{1}{2}x^2}x^4dx$.

$$\begin{aligned}
\int_a^b e^{-\frac{1}{2}x^2}x^2dx &= e^{-\frac{1}{2}x^2}(-x)\Big|_a^b - \int_a^b (-1)e^{-\frac{1}{2}x^2}dx \\
&= -be^{-\frac{1}{2}b^2} + ae^{-\frac{1}{2}a^2} + \sqrt{2\pi}
\end{aligned} \tag{6.5}$$

and

$$\int_a^b e^{-\frac{1}{2}x^2} x^4 dx \;=\; e^{-\frac{1}{2}x^2}(-x^3)\Big|_a^b - \int_a^b e^{-\frac{1}{2}x^2}(-3x^2)dx$$

Since σ_1, $\sigma_2 \ll 1$, the absolute values of the integral limits $a = -\frac{1}{\sigma}$ and $b = \frac{1}{\sigma}$ are both $\gg 1$. The integrals are simplified to be

$$\int_a^b e^{-\frac{1}{2}x^2} x^2 dx = \sqrt{2\pi} \tag{6.6}$$

and

$$\int_a^b e^{-\frac{1}{2}x^2} x^4 dx = 3\sqrt{2\pi} \tag{6.7}$$

Using these results and with some rescaling of the integral variable we obtain a relationship of the tagging parameters,

$$-\beta = (1-\epsilon_0)\alpha\left\{(2c_1 - 6c_3)\left[f\sigma_1^2 + (1-f)\sigma_2^2\right] + 96c_3\left[f\sigma_1^4 + (1-f)\sigma_2^4\right]\right\} \tag{6.8}$$

Both SVX and CTC events follow the above "beauty conservation rule", although they use different parameter sets when applicable.

6.2.3 Efficiency Asymmetry Using Inclusive J/ψ

The efficiency asymmetry is expected to be small. In [23] efficiency asymmetry was handled by constraining ϵ_+/ϵ_- using inclusive J/ψ data. Here we use the same data sample to constrain the asymmetry parameters c_1 and c_3.

The inclusive J/ψ sample provides a large data sample after a B lifetime cut removes J/ψ's that are not from B decays. The selection requirements of the inclusive J/ψ events are as follows:

- CTC covariance matrix scale factor = 2.0
- SVX alignment
  - Run 1A: 39101
  - Run 1B: 51110
- both muons SVX ($\geq$ 3 hits)
- CTVMFT 2-track vertex-constrained fit
  - $\chi^2 < 15$ (Prob(χ^2) $> 0.01\%$)
  - $L_{xy} > 0.02$ cm
  - J/ψ $p_T > 4.5\,\mathrm{GeV}/c$

The Q_{final} distribution of these events is shown in Fig. 6.1. Two odd Chebyshev polynomials (T_1 and T_3) are used to describe the efficiency asymmetry. The coefficients of the two terms are small:

$$c_1 = 0.153 \pm 0.039 \tag{6.9}$$

$$c_3 = -0.043 \pm 0.018. \tag{6.10}$$

The values and uncertainties (actually the full error matrix) of c_1 and c_3 will be used to constrain their counterparts in the $B^{\pm} \to J/\psi K^{\pm}$ calibration fit and the $\sin 2\beta$ fit on $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$.

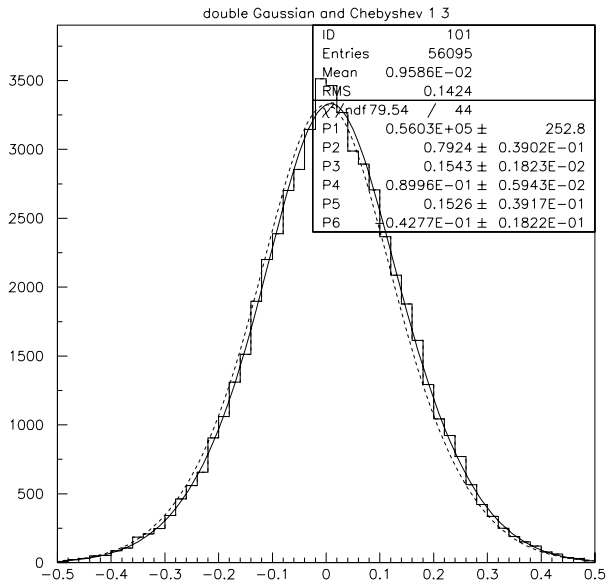


Figure 6.1: The histogram is the Q_{final} distribution for inclusive J/ψ events that have jets. The solid line is a fit to the efficiency function of the form described in the text. The dashed line is a fit to the efficiency function without the two odd Chebyshev polynomial terms.

6.2.4 Source Of Tagging Asymmetry

Sources of the tagging asymmetry include the tracking asymmetry that is present in the CTC and beam pipe protons in the kaon dE/dx tag.

The CTC super-layer cells are rotated by the Lorentz angle (45°) away from the radial direction. For low p_T tracks, positive and negative tracks curve in opposite ways in the magnetic field. This causes one type of track to encounter more super-layer cells but fewer sense wires per cell, and the other type of track to encounter fewer cells but more sense wires per cell. This built-in asymmetry in the CTC geometry results in the asymmetry in tracking efficiency.

In the kaon dE/dx tag—unlike in the jet-charge tag where only tracks with $p_T > 0.75\,\text{GeV}/c$

are used to calculate the jet charge—all tracks in the jet are used to calculate the weighted kaon charge of the jet. That is, low-p_T protons from beam pipe interactions are included if present in the jet since they have relatively large impact parameters just like tracks from the B jet. Thus the excessive positive tracks from beam pipe protons lead to more events with positive Q_{final}. In order to maximize the dE/dx particle identification capability, we keep the low-p_T tracks in the kaon tag. Since tagging asymmetry is handled throughout the analysis anyway, there is no need to cut off the low-p_T tracks just to make the tagging more symmetric.

6.3 Calibration Using $B^{\pm} \to J/\psi K^{\pm}$

We use Run 1 $B^{\pm} \to J/\psi K^{\pm}$ events from the J/ψ dataset to calibrate the D vs Q_{final} relation.

6.3.1 Event Selection

The $B^{\pm} \to J/\psi K^{\pm}$ selection cuts are listed as follows:

- CTC covariance matrix scale factor = 2.0
- SVX alignment:
  - Run 1A: 39101
  - Run 1B: 51110
- CTVMFT 2-track ($\mu^+\mu^-$) vertex-constrained fit:
  - $\chi^2 < 15.136$ (Prob(χ^2) > 0.01%)

- $|m_{\mu\mu} - 3.097| < 0.05\,\mathrm{GeV}/c^2$

- CTVMFT 3-track ($\mu^+\mu^-K^\pm$) vertex and J/ψ-mass constrained fit:

  - $\chi^2 < 16.811$ (Prob(χ^2) > 1.0%)

  - $5.15\,\mathrm{GeV}/c^2 < m_{\mu\mu K} < 5.40\,\mathrm{GeV}/c^2$ or $|(m_{\mu\mu K} - 5.279)/\sigma| < 20$

  - $p_T(K^\pm) > 1.25\,\mathrm{GeV}/c$

  - $p_T(B^\pm) > 6.0\,\mathrm{GeV}/c$

We collected 613 SVX and 390 CTC signal events, and 9355 SVX and 12647 CTC background events.

6.3.2 Q_{final} Tag

For each event, individual tags are first calculated, i.e. Q_{jet}, Q_{nused}, Q_{jetsprd}, $Q_{\Delta\phi}$, Q_{jetprb} and Q_K. They are then used as input to the combining algorithm to give Q_{final} as the output. The efficiency function $\varepsilon(Q)$ is used to describe this distribution, with its parameters to be determined by a fit where the tagging asymmetry parameters (c_1 and c_3) are constrained by inclusive J/ψ data and the rest of the parameters float freely.

6.3.3 Dilution vs. Q_{final}

For each event of a given Q_{final} value, we need to know the dilution D. Since Q_{final} is nothing but a dilution, albeit a dilution that is calculated using Monte Carlo data rather than measured directly, we expect $D \sim Q_{\mathrm{final}}$.

In order to have the dilution function $D(Q_{\text{final}})$ accommodate the tagging asymmetry, we have written $D(Q) = \alpha Q + \beta$, where α is the parameter we need to obtain for the $\sin 2\beta$ fit, and β, expressed through the beauty conservation rule, provides the mechanism of dilution asymmetry.

6.3.4 The Calibration Fit

A log-likelihood method is used to do the comprehensive fit that contains B mass, lifetime and flavor information. This calibration fit has the same structure as the $\sin 2\beta$ fit in [23]. However, instead of three taggers (SST, SLT, JCH), we have two taggers: the Q_{final} tag and the "perfect" tag, that is, the flavor of the trigger B in $B^{\pm} \to J/\psi K^{\pm}$.

In [23] dilutions and efficiencies have discrete values for $+$, $-$ and 0 tags. We keep this feature for the "perfect" tag since it is always either a $+$ tag or a $-$ tag. However, for the Q_{final} tag, since both dilution and efficiency are continuous functions of Q_{final}, corresponding changes are made in this part of the fitter.

Figure 6.2 and Fig. 6.3 show the Q_{final} distributions of $B^{\pm} \to J/\psi K^{\pm}$ 5σ mass region events, SVX and CTC, respectively. The smooth curves are overlays using the results (f, σ_1 and σ_2 for signal and background, asymmetry parameters c_1 and c_3) from the log-likelihood calibration fit. The number of signal and the number of background events are free fit parameters in both figures since in the log-likelihood fit a much wider mass window is used.

The fit parameters are summarized in Table 6.1.

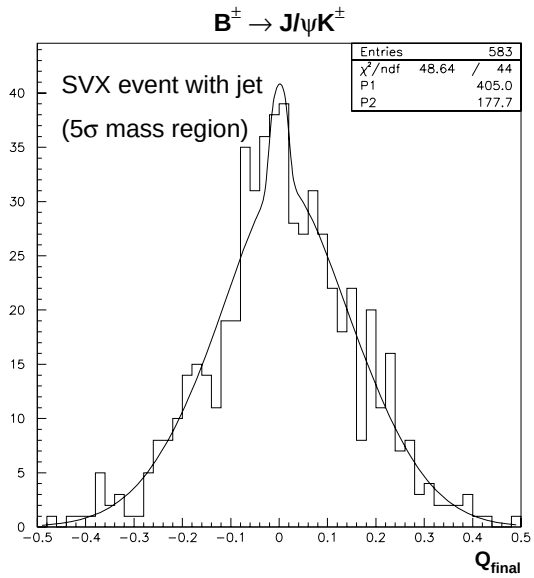


Figure 6.2: Q_{final} distribution for signal region $B^{\pm} \to J/\psi K^{\pm}$ SVX events that have jets. All events satisfy $|m_{\mu\mu} - 5.279| < 5\sigma$; $c\tau_B > 0.005$cm.

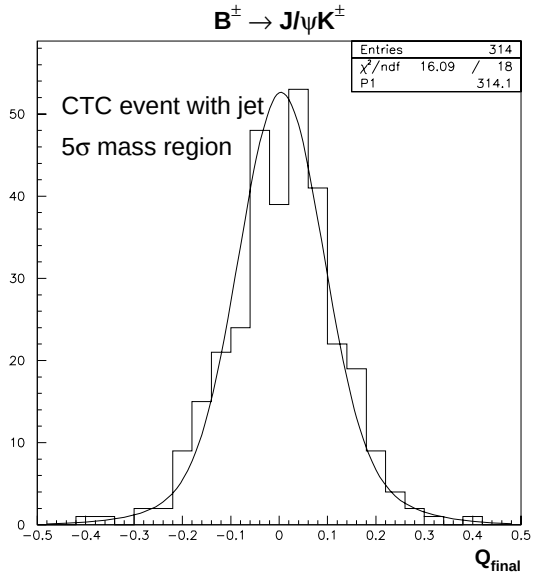


Figure 6.3: Q_{final} distribution for signal region $B^{\pm} \to J/\psi K^{\pm}$ CTC events that have jets. All events satisfy $|m_{\mu\mu} - 5.279| < 5\sigma$; $c\tau_B > 0.005$cm.

Table 6.1: $B^{\pm} \to J/\psi K^{\pm}$ calibration fit parameters and their definitions (part 1 of 3)

	parameter	symbol	description
1	`mscale`	S_m	B-mass error scale
2	`mslopep`	Σ_m^{P}	mass-slope of the prompt background
3	`mslopel`	Σ_m^{L}	mass-slope of the long-lived background
4	`tscale`	S_t^{SVX}	SVX lifetime error scale
5	`bklife`	τ_+^{SVX}	mean lifetime of long-lived SVX background
6	`tneg`	τ_-^{SVX}	$-$ SVX background lifetime
7	`nfracl`	F_-^{SVX}	$-$ lifetime frac. of long-lived SVX background
8	`mscalectc`	S_m^{CTC}	CTC B-mass error scale
9	`mslopepctc`	$\Sigma_m^{\mathrm{P_{CTC}}}$	CTC mass-slope of the prompt background
10	`mslopelctc`	$\Sigma_m^{\mathrm{L_{CTC}}}$	CTC mass-slope of the long-lived background
11	`tscalectc`	S_t^{CTC}	CTC lifetime error scale
12	`bklifectc`	τ_+^{CTC}	mean lifetime of CTC long-lived background
13	`tnegctc`	τ_-^{CTC}	$-$ CTC background lifetime
14	`nfraclctc`	F_-^{CTC}	$-$ lifetime frac. of long-lived CTC background
15	`bklifep`	$\tau_{+'}^{\mathrm{SVX}}$	2nd + SVX background lifetime
16	`fracp`	$F_{+'}^{\mathrm{SVX}}$	2nd + SVX background fraction
17	`nsigsvx`	$N_{\mathrm{S}}^{\mathrm{SVX}}$	number of SVX signal events
18	`nbgsvx`	$N_{\mathrm{B}}^{\mathrm{SVX}}$	number of SVX background events
19	`fraclsvx`	$F_{\mathrm{L}}^{\mathrm{SVX}}$	long-lived fraction of SVX background
20	`nsigctc`	$N_{\mathrm{S}}^{\mathrm{CTC}}$	number of CTC signal events

Table 6.1: $B^{\pm} \to J/\psi K^{\pm}$ calibration fit parameters and their definitions (part 2 of 3)

	parameter	symbol	description
21	`nbgctc`	$N_{\mathrm{B}}^{\mathrm{SVX}}$	number of CTC background events
22	`fraclctc`	$F_{\mathrm{L}}^{\mathrm{CTC}}$	long-lived fraction of CTC background
23	`dslope`	α^{SVX}	coefficient of Q_{final} in SVX dilution func.
24	`esjetf`	f_{S}	frac. of dominant Gaussian in SVX signal Q_{final} distribution
25	`esjets1`	$\sigma_{1\mathrm{S}}$	width of dominant Gaussian for SVX signal
26	`esjets2`	$\sigma_{2\mathrm{S}}$	width of other Gaussian for SVX signal
27	`epjetf`	f_{P}	frac. of dominant Gaussian in SVX prompt bg Q_{final} distribution
28	`epjets1`	$\sigma_{1\mathrm{P}}$	width of dominant Gaussian for SVX prompt background
29	`epjets2`	$\sigma_{2\mathrm{P}}$	width of other Gaussian for SVX prompt background
30	`eljetf`	f_{L}	frac. of dominant Gaussian in SVX long-lived bg Q_{final} distribution
31	`eljets1`	$\sigma_{1\mathrm{L}}$	width of dominant Gaussian for SVX long-lived bg
32	`eljets2`	$\sigma_{2\mathrm{L}}$	width of other Gaussian for SVX long-lived bg
33	`b0life`	τ_{B^0}	mean B^0 lifetime
34	`kplusf`	$f_{\mathrm{S}}^{K^+}$	fraction of K^+ in SVX signal
35	`kpluspf`	$f_{\mathrm{P}}^{K^+}$	fraction of K^+ in SVX prompt bg
36	`kpluslf`	$f_{\mathrm{L}}^{K^+}$	fraction of K^+ in SVX long-lived bg
37	`effs`	$\epsilon_{0\mathrm{S}}^{\mathrm{SVX}}$	jet-clustering efficiency for SVX signal
38	`effp`	$\epsilon_{0\mathrm{P}}^{\mathrm{SVX}}$	jet-clustering efficiency for SVX prompt bg
39	`effl`	$\epsilon_{0\mathrm{L}}^{\mathrm{SVX}}$	jet-clustering efficiency for SVX long-lived bg
40	`bgslope`	$\alpha_{\mathrm{B}}^{\mathrm{SVX}}$	coeff. of Q_{final} in SVX bg dilution func.

Table 6.1: $B^{\pm} \to J/\psi K^{\pm}$ calibration fit parameters and their definitions (part 3 of 3)

	parameter	symbol	description
41	`bgslopec`	$\alpha_{\mathrm{B}}^{\mathrm{CTC}}$	coeff. of Q_{final} in CTC bg dilution func.
42	`esjetfc`	$f_{\mathrm{S}}^{\mathrm{CTC}}$	fract. of dominant Gaussian in CTC signal Q_{final} distribution
43	`esjets1c`	$\sigma_{1\mathrm{S}}^{\mathrm{CTC}}$	width of dominant Gaussian for CTC signal
44	`esjets2c`	$\sigma_{2\mathrm{S}}^{\mathrm{CTC}}$	width of other Gaussian for CTC signal
45	`epjetfc`	$f_{\mathrm{P}}^{\mathrm{CTC}}$	frac. of dominant Gaussian in CTC prompt bg Q_{final} distribution
46	`epjets1c`	$\sigma_{1\mathrm{P}}^{\mathrm{CTC}}$	width of dominant Gaussian for CTC prompt bg
47	`epjets2c`	$\sigma_{2\mathrm{P}}^{\mathrm{CTC}}$	width of other Gaussian for CTC prompt bg
48	`eljetfc`	$f_{\mathrm{L}}^{\mathrm{CTC}}$	frac. of dominant Gaussian in CTC long-lived Q_{final} distribution
49	`eljets1c`	$\sigma_{1\mathrm{L}}^{\mathrm{CTC}}$	width of dominant Gaussian for CTC long-lived bg
50	`eljets2c`	$\sigma_{2\mathrm{L}}^{\mathrm{CTC}}$	width of other Gaussian for CTC long-lived bg
51	`effsc`	$\epsilon_{0\mathrm{S}}^{\mathrm{CTC}}$	jet-clustering efficiency for CTC signal
52	`effpc`	$\epsilon_{0\mathrm{P}}^{\mathrm{CTC}}$	jet-clustering efficiency for CTC prompt bg
53	`efflc`	$\epsilon_{0\mathrm{L}}^{\mathrm{CTC}}$	jet-clustering efficiency for CTC long-lived bg
54	`chebpar1`	c_1	coeff. of first Chebyshev Polynomial in efficiency func.
55	`chebpar2`	c_3	coeff. of third Chebyshev Polynomial in efficiency func.
56	`bgdoff`	β'^{SVX}	dilution offset for SVX background
57	`bgdoffc`	β'^{CTC}	dilution offset for CTC background

The signal dilution D is related to Q_{final} by the coefficient α. Such a relation is also used for background events in the fit, where a residual correlation between the B flavor and the charge of X in $B \to J/\psi X$ may exist, thus leading to a small positive coefficient. Two such background coefficients are supplied in the fit, one for SVX ($\alpha_{\text{B}}^{\text{SVX}}$) and one for CTC ($\alpha_{\text{B}}^{\text{CTC}}$), due to the different cuts that are applied to SVX and CTC events.

The calibration fit returns the coefficient relating Q_{final} to D for signal events as

$$\alpha = 1.277 \pm 0.380, \tag{6.11}$$

which is consistent with 1.0, indicating the Monte Carlo models the data well. The fit results are shown in Table 6.2. Figure 6.4 shows the calibration fit result of D vs. Q_{final} relation. A direct χ^2 fit is also shown for comparison. The apparent large negative offset from the latter fit is partially due to more $B^+ \to J/\psi K^+$ in the calibration sample than $B^- \to J/\psi K^-$. In the log-likelihood fit, the tagging asymmetry caused by this apparent charge bias, purely statistical, is constrained to result from the inclusive J/ψ sample.

6.3.5 The Improved Tagging Power

The tagging power of the Q_{final} tag is

$$[\epsilon D^2] = \int_{-1}^{1} \varepsilon(Q)[D(Q)]^2 dQ. \tag{6.12}$$

Substituting the efficiency function $\varepsilon(Q)$ and dilution function $D(Q)$ and calculating the integral, we have

$$[\epsilon D^2] = (1-\epsilon_0)\alpha^2[f\sigma_1^2 + (1-f)\sigma_2^2] - \beta^2, \tag{6.13}$$

where the parameters are defined in 6.2.1. The error on $[\epsilon D^2]$ is

$$\Delta[\epsilon D^2] = \sum_{i,j} \frac{\partial[\epsilon D^2]}{\partial x_i} \Delta x_i \Delta x_j \frac{\partial[\epsilon D^2]}{\partial x_j} \tag{6.14}$$

where x_i, x_j are parameters in $[\epsilon D^2]$ expression, and the term $\Delta x_i \Delta x_j$ is the error matrix element relating x_i and x_j. Using the parameter values from the $\sin 2\beta$ fit (see Section 6.5), for SVX signal events, we have

$$[\epsilon D^2] = 0.025 \pm 0.014. \tag{6.15}$$

For CTC signal events, we have

$$[\epsilon D^2] = 0.019 \pm 0.011. \tag{6.16}$$

This is the tagging power of the Q_{final} tag in the context of two other tags (SLT and SST).

6.3.6 Validity of Dilution vs. Q_{final} Relation

To verify that $D = \alpha Q + \beta$ is sufficient to describe the relation between the dilution and Q_{final}, (Chebyshev) polynomials of higher orders are added. Note that any addition of polynomials of even

Table 6.2: Calibration fit result with $B^{\pm} \to J/\psi K^{\pm}$ (part 1 of 3)

	parameter	symbol	value	parabolic error
1	`mscale`	S_m	1.375	0.066
2	`mslopep`	Σ_m^{P}	0.014	0.001
3	`mslopel`	Σ_m^{L}	−0.005	0.003
4	`tscale`	S_t^{SVX}	1.002	0.010
5	`bklife`	τ_+^{SVX}	1.125	0.046
6	`tneg`	τ_-^{SVX}	0.780	0.116
7	`nfracl`	F_-^{SVX}	0.068	0.011
8	`mscalectc`	S_m^{CTC}	1.107	0.096
9	`mslopepctc`	$\Sigma_m^{\mathrm{P_{CTC}}}$	0.018	0.001
10	`mslopelctc`	$\Sigma_m^{\mathrm{L_{CTC}}}$	0.010	0.003
11	`tscalectc`	S_t^{CTC}	1.326	0.013
12	`bklifectc`	τ_+^{CTC}	2.463	0.079
13	`tnegctc`	τ_-^{CTC}	3.428	0.251
14	`nfraclctc`	F_-^{CTC}	0.170	0.015
15	`bklifep`	$\tau_{+'}^{\mathrm{SVX}}$	0.100	constant
16	`fracp`	$F_{+'}^{\mathrm{SVX}}$	0.000	constant
17	`nsigsvx`	$N_{\mathrm{S}}^{\mathrm{SVX}}$	612.56	31.528
18	`nbgsvx`	$N_{\mathrm{B}}^{\mathrm{SVX}}$	9355.4	98.672
19	`fraclsvx`	$F_{\mathrm{L}}^{\mathrm{SVX}}$	0.119	0.005
20	`nsigctc`	$N_{\mathrm{S}}^{\mathrm{CTC}}$	395.28	34.974

Table 6.2: Calibration fit result with $B^{\pm} \to J/\psi K^{\pm}$ (part 2 of 3)

	parameter	symbol	value	parabolic error
21	`nbgctc`	$N_{\mathrm{B}}^{\mathrm{SVX}}$	12642.	116.04
22	`fraclctc`	$F_{\mathrm{L}}^{\mathrm{CTC}}$	0.154	0.006
23	`dslope`	α^{SVX}	1.277	0.380
24	`esjetf`	f_{S}	0.911	0.071
25	`esjets1`	$\sigma_{1\mathrm{S}}$	0.159	0.009
26	`esjets2`	$\sigma_{2\mathrm{S}}$	0.045	0.026
27	`epjetf`	f_{P}	0.778	0.058
28	`epjets1`	$\sigma_{1\mathrm{P}}$	0.152	0.004
29	`epjets2`	$\sigma_{2\mathrm{P}}$	0.079	0.009
30	`eljetf`	f_{L}	0.865	0.066
31	`eljets1`	$\sigma_{1\mathrm{L}}$	0.165	0.007
32	`eljets2`	$\sigma_{2\mathrm{L}}$	0.051	0.019
33	`b0life`	τ_{B^0}	1.655	0.068
34	`kplusf`	$f_{\mathrm{S}}^{K^+}$	0.520	0.021
35	`kpluspf`	$f_{\mathrm{P}}^{K^+}$	0.509	0.004
36	`kpluslf`	$f_{\mathrm{L}}^{K^+}$	0.498	0.013
37	`effs`	$\epsilon_{0\mathrm{S}}^{\mathrm{SVX}}$	0.603	0.024
38	`effp`	$\epsilon_{0\mathrm{P}}^{\mathrm{SVX}}$	0.558	0.006
39	`effl`	$\epsilon_{0\mathrm{L}}^{\mathrm{SVX}}$	0.645	0.018
40	`bgslope`	$\alpha_{\mathrm{B}}^{\mathrm{SVX}}$	0.010	0.099

Table 6.2: Calibration fit result with $B^{\pm} \to J/\psi K^{\pm}$ (part 3 of 3)

	parameter	symbol	value	parabolic error
41	`bgslopec`	$\alpha_{\mathrm{B}}^{\mathrm{CTC}}$	0.258	0.104
42	`esjetfc`	$f_{\mathrm{S}}^{\mathrm{CTC}}$	0.207	0.336
43	`esjets1c`	$\sigma_{1\mathrm{S}}^{\mathrm{CTC}}$	0.169	0.103
44	`esjets2c`	$\sigma_{2\mathrm{S}}^{\mathrm{CTC}}$	0.085	0.020
45	`epjetfc`	$f_{\mathrm{P}}^{\mathrm{CTC}}$	0.264	0.076
46	`epjets1c`	$\sigma_{1\mathrm{P}}^{\mathrm{CTC}}$	0.160	0.010
47	`epjets2c`	$\sigma_{2\mathrm{P}}^{\mathrm{CTC}}$	0.096	0.004
48	`eljetfc`	$f_{\mathrm{L}}^{\mathrm{CTC}}$	0.445	0.193
49	`eljets1c`	$\sigma_{1\mathrm{L}}^{\mathrm{CTC}}$	0.146	0.016
50	`eljets2c`	$\sigma_{2\mathrm{L}}^{\mathrm{CTC}}$	0.082	0.012
51	`effsc`	$\epsilon_{0\mathrm{S}}^{\mathrm{CTC}}$	0.535	0.040
52	`effpc`	$\epsilon_{0\mathrm{P}}^{\mathrm{CTC}}$	0.548	0.005
53	`efflc`	$\epsilon_{0\mathrm{L}}^{\mathrm{CTC}}$	0.617	0.017
54	`chebpar1`	c_1	0.135	0.036
55	`chebpar2`	c_3	-0.052	0.017
56	`bgdoff`	β'^{SVX}	-0.033	0.010
57	`bgdoffc`	β'^{CTC}	-0.003	0.009

order to the right-hand side of the equation implies a dilution asymmetry, which is a small effect already described by β—there is no need to complicate it by adding high order terms. Given this, it is straight forward to add the next order Chebyshev polynomial T_3 in the equation:

$$D = \alpha Q + \alpha_1 T_3(2Q) + \beta, \tag{6.17}$$

where α_1 is the coefficient of the new term.

A straightforward calculation of the associated production (beauty conservation) of $b/\bar{b}$ gives

$$\begin{aligned} -\beta &= (1-\epsilon_0)\alpha \left\{ (2c_1 - 6c_3) \left[f\sigma_1^2 + (1-f)\sigma_2^2 \right] + 96c_3 \left[f\sigma_1^4 + (1-f)\sigma_2^4 \right] \right\} + \\ & (1-\epsilon_0)\alpha_1 \left\{ 12(3c_3 - c_1) \left[f\sigma_1^2 + (1-f)\sigma_2^2 \right] + 192(c_1 - 6c_3) \left[f\sigma_1^4 + (1-f)\sigma_2^4 \right] + \right. \\ & \left. 15360c_3 \left[f\sigma_1^6 + (1-f)\sigma_2^6 \right] \right\}. \end{aligned} \tag{6.18}$$

With the additional term the calibration fit yields

$$\alpha_1 = 0.06 \pm 0.09. \tag{6.19}$$

α_1 being a small number and consistent with zero within 1σ uncertainty indicates that our choice of $D = \alpha Q + \beta$ is sufficient in describing the relation between the dilution and Q_{final}.

6.3.7 No-jet Events

One question arises when considering events with no jet and events with $Q_{\text{final}} = 0$: do they have the same dilution? For an unbiased sample, naturally both would have a dilution of zero. In case of asymmetric tagging, we believe both types are subject to the same bias, i.e., $D = \beta$. Just as there is no need to distinguish dilutions of jets with different number of tracks provided they have the same Q_{final} value, there is no need to single out the case where the number of tracks is zero.

To check and verify this argument, a fit with the two types of events having different dilutions is conducted. For events with no jets found, a new parameter D_{notag} is introduced into the fit. The beauty conservation expression is adjusted accordingly to

$$-\beta = \alpha \left\{ (2c_1 - 6c_3) \left[f\sigma_1^2 + (1-f)\sigma_2^2 \right] + 96c_3 \left[f\sigma_1^4 + (1-f)\sigma_2^4 \right] \right\} + \frac{\epsilon_0}{(1-\epsilon_0)} D_{\text{notag}} \tag{6.20}$$

The resulting dilution parameters are $\alpha = 1.29 \pm 0.38$ and $D_{\text{notag}} = 0.031 \pm 0.047$, both are consistent with the fit used in this analysis.

6.4 Tags Revisited

6.4.1 Neutral Tags in *b*-jets vs. non-*b*-jets

A direct piece of evidence that a neutral tag does distinguish $b(\bar{b})$-jets from non-$b(\bar{b})$ jets is the difference in its distributions in the two categories. For this purpose, the b content of a jet in the Monte Carlo must be known. There are two types of tracks in the jet—the ones that originated from the opposite-side b quark, i.e., the b-tracks that the jet-charge algorithm targeted; and the rest, i.e., the noise tracks. Whether a track is a b-track or a noise track is known in the Monte Carlo since track parent information is available. Figure 6.5 shows a histogram of the percentage of the target (opposite-side-b) tracks in the jet (the dotted line); as a comparison, the percentage of tracks from the same-side-b is also plotted (the solid line)—the low percentage is due to the fact that these tracks are effectively disqualified during the clustering process. From Fig. 6.5, it is reasonable to categorize the jets with the opposite-side-b percentage ≥ 0.6 as $b(\bar{b})$-jets, and jets

with the opposite-side-b percentage ≤ 0.4 as non-$b(\bar{b})$-jets.

With b-jets and non-b-jets defined, the neutral tags are ready for display in these two categories. In Fig. 6.6, four pieces of different neutral information are shown, i.e. the neutral tags $Q_{\rm jd}$, $Q_{\Delta\phi}$ and $Q_{\rm jp}$, and the jet p_T.

Clearly the neutral tags have different distributions between b-jets and non-b-jets.

6.4.2 Tags in Monte Carlo vs. data

As presented in [22] the combining algorithm that maximizes the overall ϵD^2 is optimized (or trained) with the Monte Carlo sample. Naturally some degree of resemblance between the Monte Carlo sample and the data is required to retain the algorithm's behavior in the data. A slight mismatch should not introduce a systematic uncertainty since the true tagging power in data is determined by the calibration procedure. Nonetheless, comparing the tags between Monte Carlo and data ensures that the calibration relation is justified. That is, a linear dilution vs. $Q_{\rm final}$ (the Monte Carlo predicted dilution) relation is acceptable if the Monte Carlo resembles the data. Verifications of this calibration relation are discussed in sections 6.3.6 and 6.7.3. Figure 6.7 shows the tags in the Monte Carlo vs. in the calibration sample. The degree of their resemblance gives confidence in our method.

6.4.3 $Q_{\rm final}$ sensitivity to samples with different $\Delta\phi$ characteristics

There have been concerns that the PYTHIA/QQ Monte Carlo simulation does not yield the correct azimuthal distributions of tracks. Such concern leads to questions on whether $Q_{\rm final}$ behavior in the Monte Carlo resembles that in the data. Figure 6.7 confirms that the tags do behave the same

in Monte Carlo and the data, yet it is worthwhile to know that Q_{final} is not drastically sensitive to sample selections. For such an investigation, a sample with different azimuthal characteristics is in order. Such a different sample can be easily produced. In the original sample, each event is assigned a weight $\in [0, 1]$. A random number generator is used to decide if an event is going to be discarded. This will effectively produce a sample with different characteristics. The weight is defined as $w = \Delta\phi/\pi$, where $\Delta\phi$ is the azimuthal angle between the b-$\bar{b}$ quarks.

With two samples of different azimuthal characteristics, two different sets of dilution matrices (functions) are constructed. Then a third sample is produced to have the same characteristics as one of the previous two samples. By using both sets to calculate the Q_{final}, for each event in this sample, there are two Q_{final}'s produced—one using the correct set of dilution functions, and the other using the incorrect set. Figure 6.8 shows the comparison of the two Q_{final}'s. A linear relation between the two is evident in the plot, which can be correctly handled by the linear $D(Q)$ calibration process.

6.5 Measuring $\sin 2\beta$

The CP asymmetry parameter $\sin 2\beta$ is measured in CDF run 1 $B^0/\bar{B}^0 \to J/\psi K_S$ events and $B^0/\bar{B}^0 \to \psi(2S)K_S$ events. Since these two decay modes have the same CP property, they can be combined into a single sample. The difference in the error-scale of their normalized mass spectra is recognized ($\alpha_{\psi(2S)}$).

6.5.1 $B^0/\bar{B}^0 \to J/\psi K_S$ Event Selection

For $B^0/\bar{B}^0 \to J/\psi K_S$, the same selection criteria as in [23] are used. We reiterate the selection criteria here:

- CTC covariance matrix scale factor = 2.0
- Material scale factor = 1.00
- SVX alignment:
  - Run 1A: 39101
  - Run 1B: 51110
- muon matching cuts (CMP/CMX and CMU/CMP/CMX muons not allowed)
  - CMU: $\chi^2_x < 9$, $\chi^2_z < 12$
  - CMP: $\chi^2_x < 9$
  - CMX: $\chi^2_x < 9$
- SVX/CTC track selection:
  - **muon:** use CTC-only fit if **any** of the following is true:
    - Prob(χ^2(SVX)/hit) < 1%
    - less than 3 SVX hits
  - **pion:** use CTC-only track fit if **any** of the following is true:
    - Prob(χ^2(SVX)/hit) < 1%

    * less than 3 SVX hits

- CTVMFT 2-track ($\mu^+\mu^-$) vertex-constrained fit:
  - $|(m_{\mu\mu} - m_{J/\psi})/\sigma_{\mu\mu}| < 5$
- CTVMFT 2-track ($\pi^+\pi^-$) vertex-constrained fit:
  - $|(m_{\pi\pi} - m_{K^0_S})/\sigma_{\pi\pi}| < 5$
  - Prob(χ^2) $> 0.1\%$
  - We originally allow the tracks from the K^0_S to be SVX tracks. If the fitted vertex has a radius greater than 4.5 cm, then we use the CTC fit for both tracks.
  - We accept mixed K^0_S track pairs (*i.e.* one SVX track and one CTC track).
  - $L_{xy}/\sigma_{L_{xy}} > 5$
- CTVMFT 4-track fit: muon pair is vertex and J/ψ-mass constrained; pion pair is vertex and K^0_S mass constrained; K^0_S is required to point to the dimuon vertex and the B is required to point to the primary vertex:
  - $|(m_{\mu\mu\pi\pi} - m_{B^0})/\sigma| < 20$
  - $p_T(B) > 4.5\,\mathrm{GeV/c}$
  - $p_T(K^0_S) > 700\,\mathrm{MeV/c}$
  - Prob(χ^2) $> 0.01\%$ (Standard 3-D χ^2 returned from CTVMFT.)

We collected 201 SVX and 193 CTC signal events, and 1619 SVX and 2185 CTC background events. Figure 6.9 shows the normalized mass distribution of the sample.

6.5.2 $B^0/\bar{B}^0 \to \psi(2S)K_S$ Event Selection

For $B^0/\bar{B}^0 \to \psi(2S)K_S$, there are two categories of events: $\psi(2S) \to \mu^+\mu^-$ and $\psi(2S) \to \mu^+\mu^-\pi^+\pi^-$, both with $K_S^0 \to \pi^+\pi^-$. We only list the selection criteria that are different from $B^0/\bar{B}^0 \to J/\psi K_S$:

- $\psi(2S) \to \mu^+\mu^-$
  - $|(m_{\mu\mu} - m_{\psi(2S)})/\sigma_{\mu\mu}| < 5$.
- $\psi(2S) \to J/\psi\pi^+\pi^-$ and subsequently $J/\psi \to \mu^+\mu^-$:
  - CTVMFT 4-track fit to $\psi(2S)$ ($\mu^+\mu^-\pi^+\pi^-$): Prob(χ^2) > 0.1%;
  - CTVMFT 6-track fit to B.

We collected 18 SVX and 35 CTC signal events, and 195 SVX and 372 CTC background events. Fig. 6.10 shows the $\psi(2S)$ events.

6.5.3 Calculate Q_{final}

Q_{final} is calculated for each event the same way as in the $B^\pm \to J/\psi K^\pm$ calibration.

6.5.4 Fit for $\sin 2\beta$

A log-likelihood method is designed to simultaneously fit multiple parameters of the samples involved. $\sin 2\beta$ is one of the parameters. For a detailed discussion of the likelihood function, one should refer to [23]. The likelihood function consists of contributions of each event's probability and constraints on the parameters. The probability of an event has three components—signal, prompt

background and long-lived background probabilities. Each component in turn, is the product of its probabilities of lifetime, mass and tagging-efficiency. This tagging-efficiency is really the combined efficiency of the three tags. The procedure to combine un-correlated tags based on their dilutions and efficiencies is based on the procedure used in the previous measurement [23]:

1. Initialize: set $n = 1$, $D = 0$, and $\varepsilon = 1$
2. Update ε: set $\varepsilon = \varepsilon\varepsilon_n(1 + DD_n)$
3. Update D: set $D = (D + D_n)/(1 + DD_n)$
4. Increment: set $n = n + 1$ and go to step 2 if $n \leq$ number of tags
5. Terminate: D and ε now have their final values

Now that the Q_{final} tag is a continuous tag, whose dilution and efficiency are continuous functions of Q_{final}, the above steps are still valid. We simply need to replace its ε and D with $\varepsilon(Q_{\text{final}})$ and $D(Q_{\text{final}})$.

Now we can focus on our newly-improved opposite-side jet-tag. The same form of efficiency functions are used for the $B^0/\bar{B}^0 \to J/\psi K_S$ and the $B^0/\bar{B}^0 \to \psi(2S)K_S$ samples as in $B^\pm \to J/\psi K^\pm$, that is, the double-Gaussian with fractions of asymmetries and the δ-function. Events of different categories, i.e. SVX or CTC, signal or background, all have the same widths of the Gaussian components, but the relative weights of the Gaussians differ among categories.

The D vs. Q_{final} relation from the calibration sample is supplied as a constraint to the $\sin 2\beta$ fit. Unlike the background in the calibration fit, where the charge of X in $B \to J/\psi X$ may have residual correlations to Q_{final}, here background dilutions are zero because K_S^0 is charge symmetric.

The same-side tag and the soft-lepton tag are unchanged. The greedy feature of the soft-lepton tag is turned on—for an event in which both SLT and Q_{final} tags exist, only SLT tag is used [23] due to its large dilution. The reason for using only one tag (SLT) when both are present in the event is the possible correlation between the two since they are both opposite-side tags. This is a conservative step. The free parameters are listed in Table 6.3.

The parameters are returned via a simultaneous fit with all tags using all events. The fit result is listed in Table 6.4. The log-likelihood as a function of $\sin 2\beta$ is shown in Fig. 6.11. The fit is illustrated in Fig. 6.13.

Figure 6.12 shows the log-likelihood as a function of $\sin 2\beta$ when only Q_{final} tag is used in the fit. It has a well-determined minimum.

A summary of a variety of $\sin 2\beta$ fits are shown in Table 6.5.

A simultaneous fit to both $\sin 2\beta$ and Δm_{B^0} with the latter not constrained to the world average yields $\sin 2\beta = 1.06^{+0.39}_{-0.37}$ and $\Delta m_{B^0} = 0.66 \pm 0.12\text{ps}^{-1}$. The Δm_{B^0} result agrees with the PDG value at the level of $\sim 1.6\sigma$.

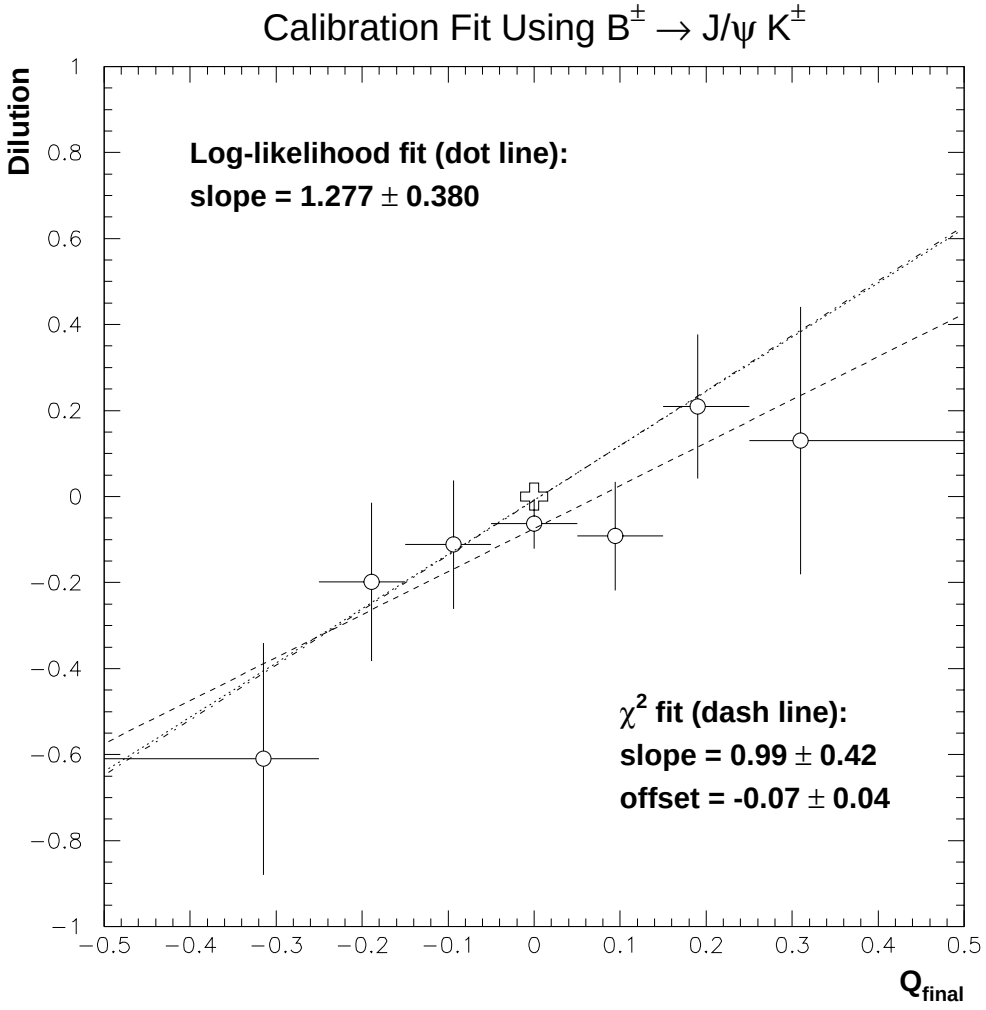


Figure 6.4: Dilution vs. Q_{final} in $B^{\pm} \to J/\psi K^{\pm}$ sample.

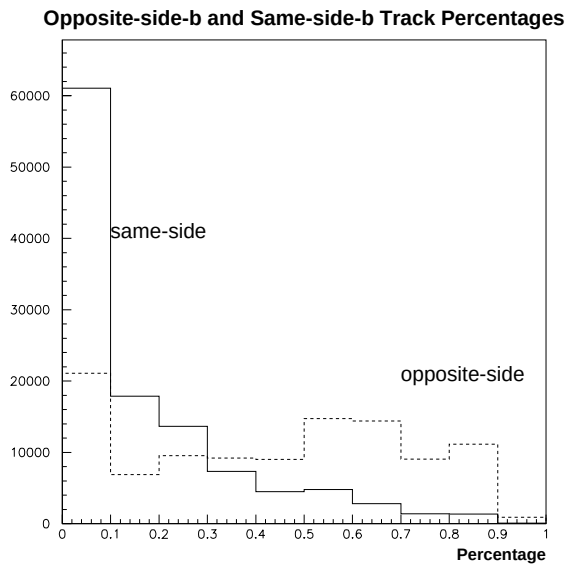


Figure 6.5: Percentages of tracks from opposite-side b and same-side b. The jet clustering algorithm has removed much of the same-side tracks, resulting in their low percentage in the jet.

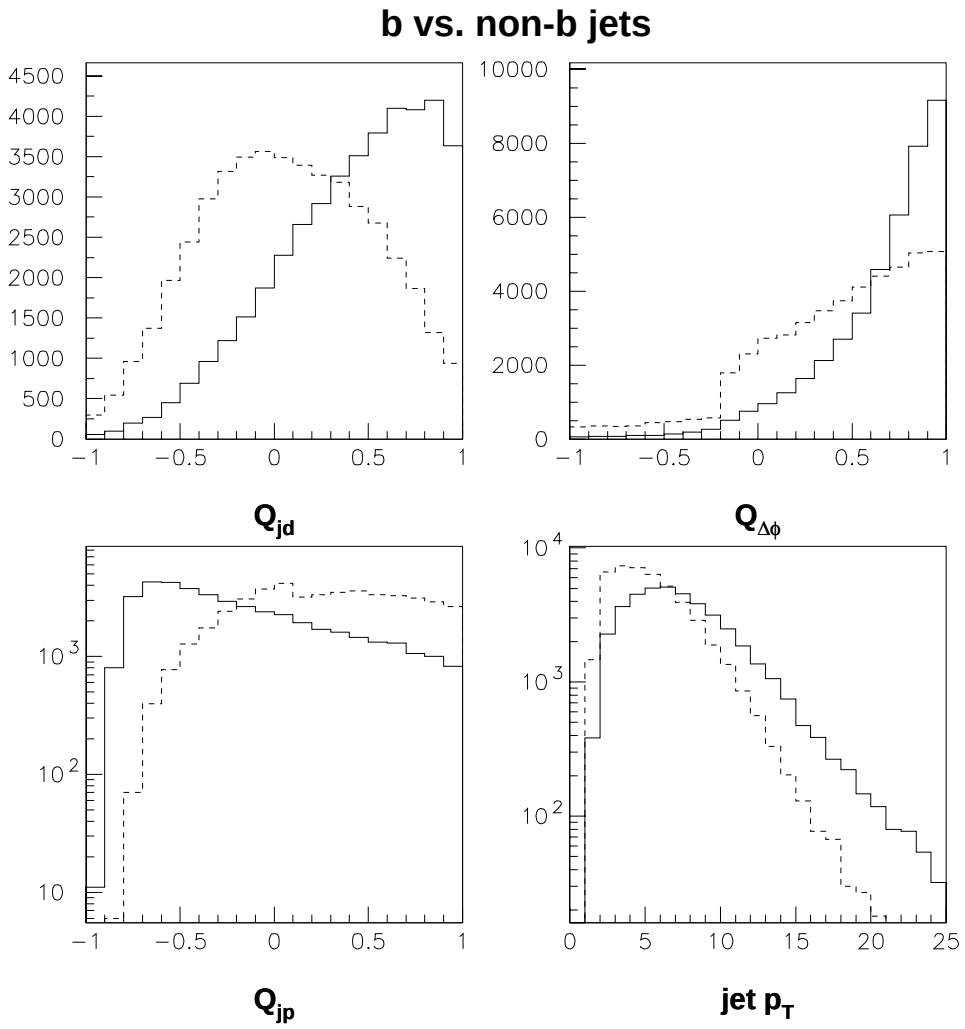


Figure 6.6: Solid: b-jet; Dash: non-b jet. Plots show that the neutral variables differentiate b jets from non-b jets.

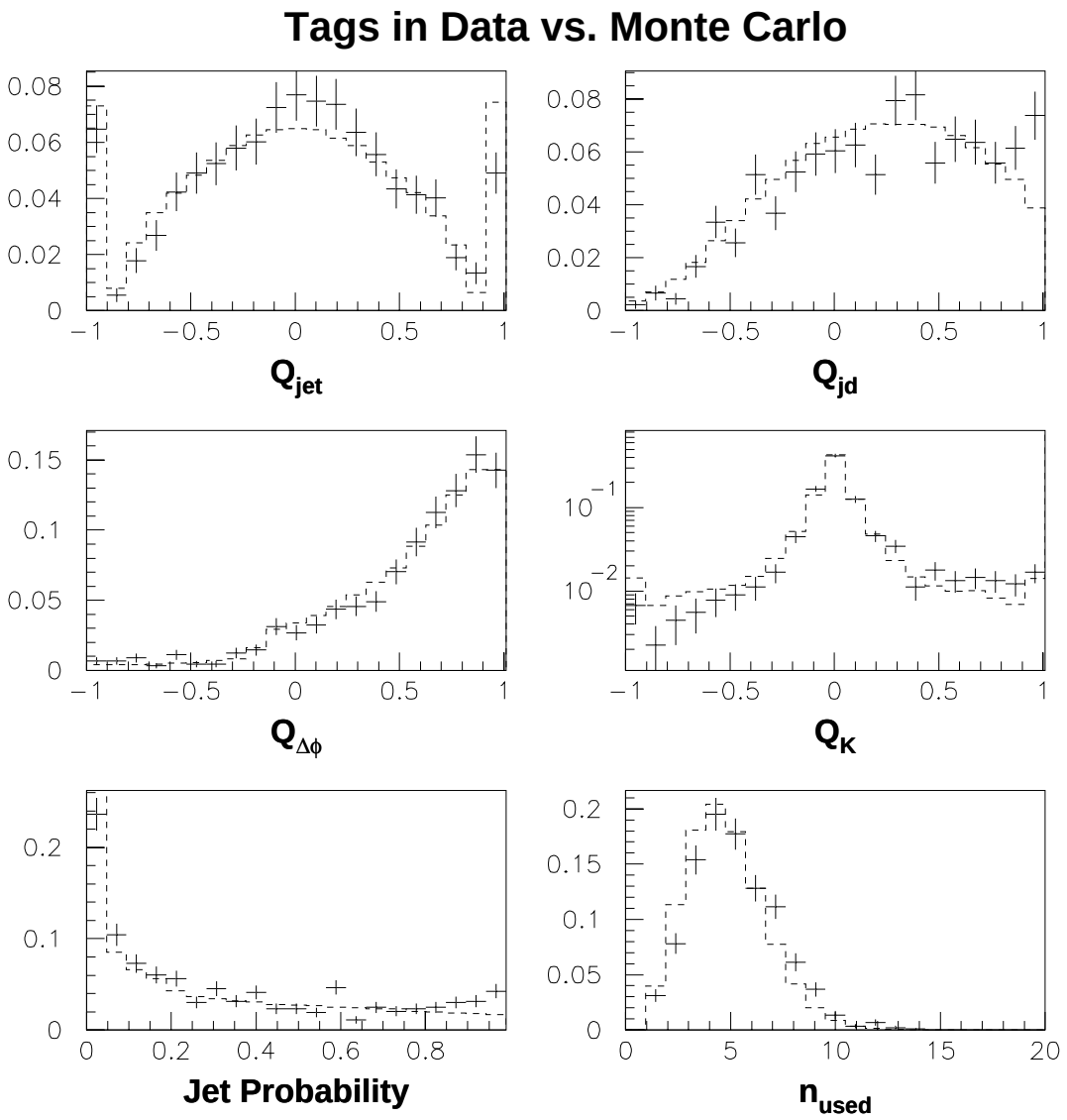


Figure 6.7: The dash line shows the tag distribution in the Monte Carlo. The points with error bars show the tag distribution in the $B^{\pm} \to J/\psi K^{\pm}$ data. In the n_{used} tag plot, a small disagreement can be seen. The algorithm is verified to be not sensitive to this difference.

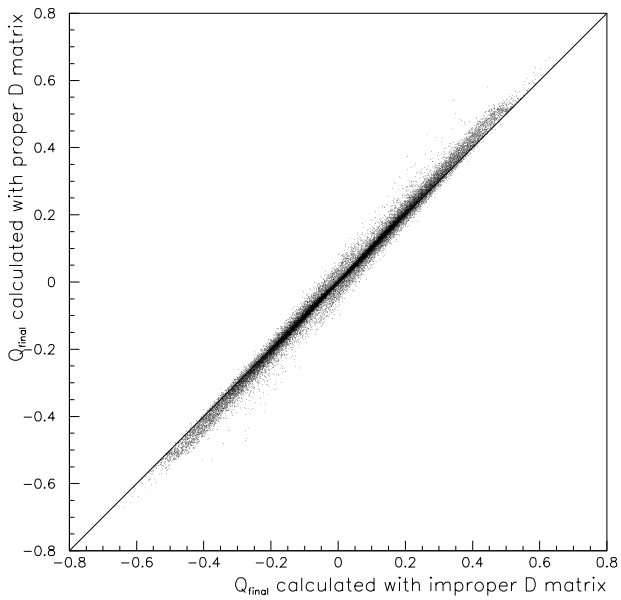


Figure 6.8: Q_{final}'s calculated using two sets of dilution matrices are plotted as 2D histogram. The diagonal line is $y = x$. The Q_{final} calculated using the proper dilution matrices have slightly higher amplitudes. A linear relation is retained between the two Q_{final}'s, i.e., we can still use the linear $D(Q)$ calibration relation.

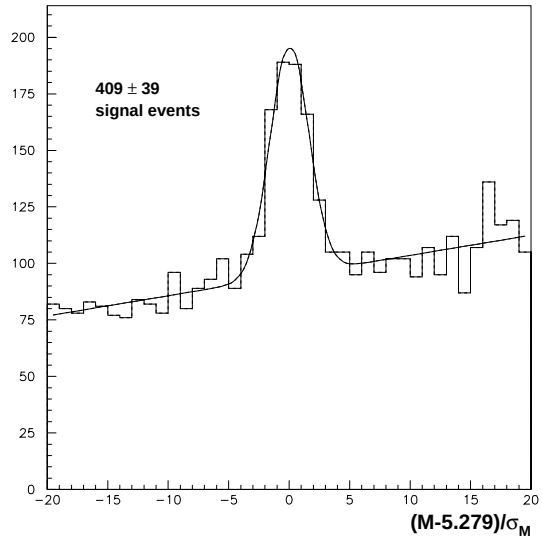


Figure 6.9: The normalized mass of $B^0/\bar{B}^0 \to J/\psi K_S$.

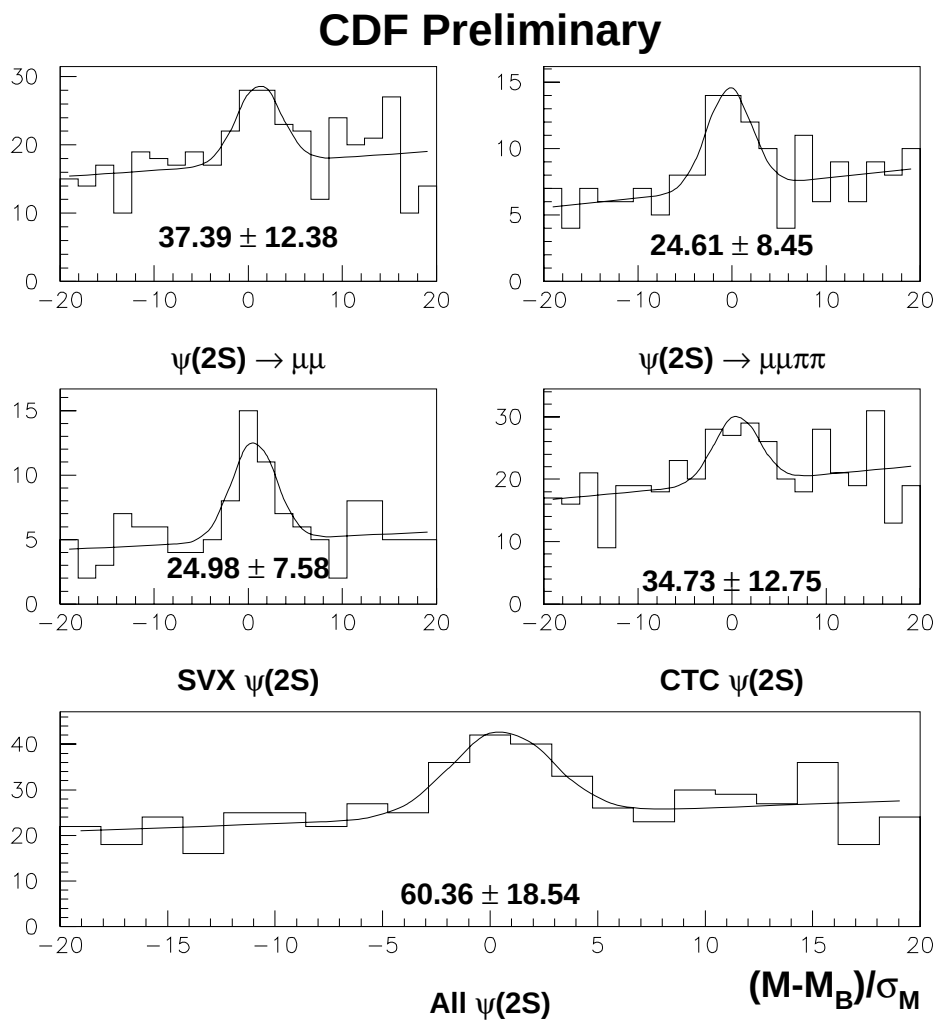


Figure 6.10: $B^0/\bar{B}^0 \to \psi(2S)K_S$ mass.

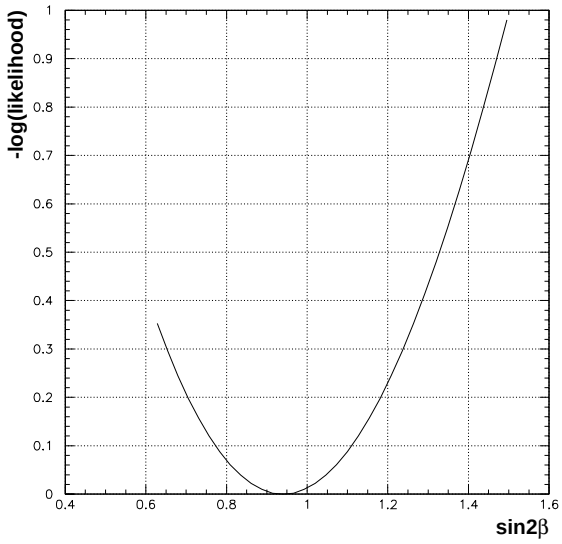


Figure 6.11: The negative log-likelihood function near the $\sin 2\beta$ minimum.

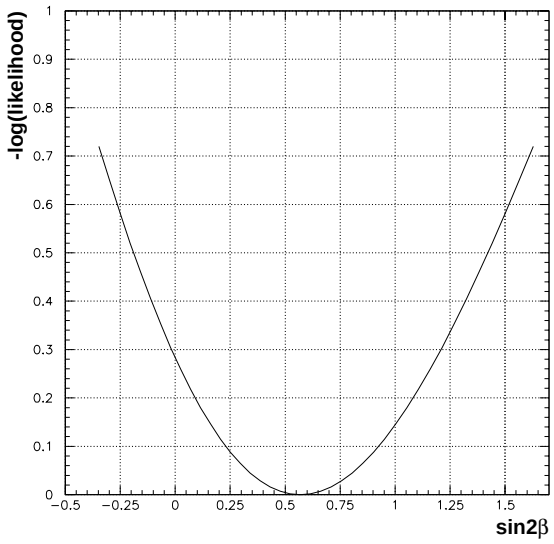


Figure 6.12: The negative log-likelihood function near the $\sin 2\beta$ minimum. Only the Q_{final} tag is used in the fit.

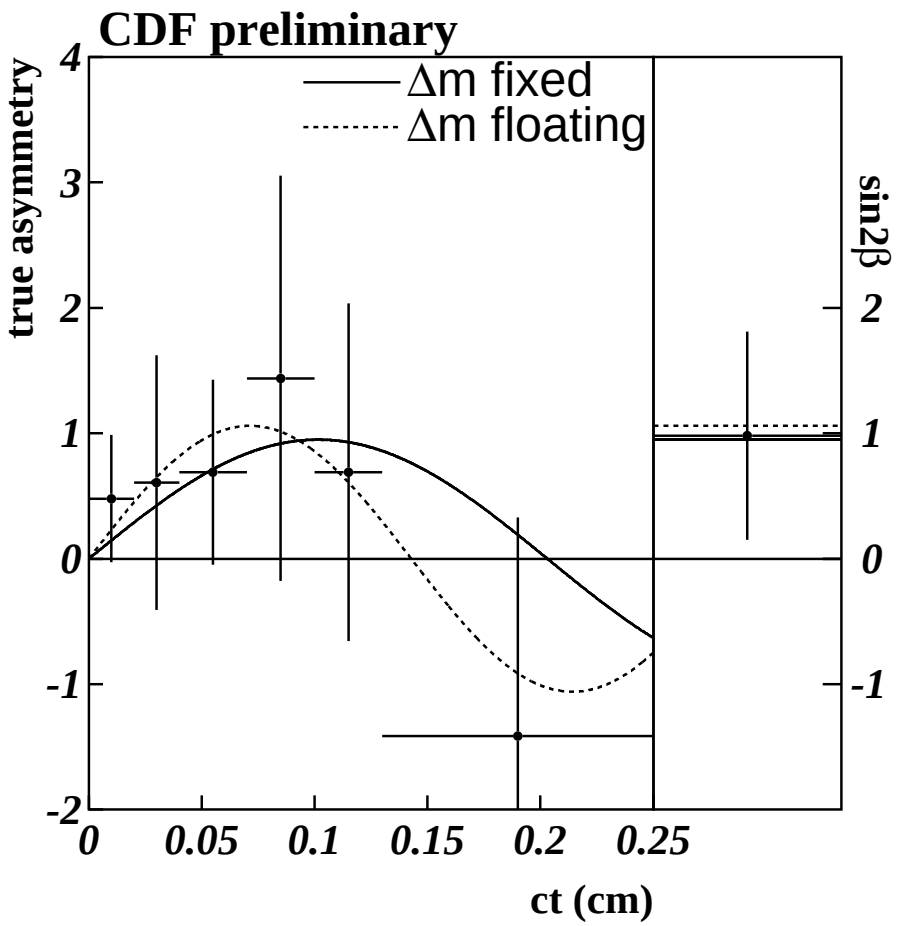


Figure 6.13: The true asymmetry ($\sin 2\beta \sin \Delta m_d t$) as a function of the lifetime for $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ events. The data points are sideband-subtracted. The non-SVX events are shown on the right indicating a time-integrated measurement of $\sin 2\beta$. Courtesy of Kevin Pitts, UIUC.

Table 6.3: Free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 1 of 4)

	parameter	symbol	description
1	`mscale`	S_m	B-mass error scale
2	`mslopep`	Σ_m^{P}	mass-slope of the prompt background
3	`mslopel`	Σ_m^{L}	mass-slope of the long-lived background
4	`tscale`	S_t^{SVX}	SVX lifetime error scale
5	`bklife`	τ_+^{SVX}	mean lifetime of long-lived SVX background
6	`sin2beta`	$\sin 2\beta$	asymmetry due to CP violation
7	`nsig`	$N_{\mathrm{S}}^{\mathrm{SVX}}$	number of SVX signal events
8	`nbck`	$N_{\mathrm{B}}^{\mathrm{SVX}}$	number of SVX background events
9	`nlfrac`	$F_{\mathrm{L}}^{\mathrm{SVX}}$	long-lived fraction of SVX background
10	`tneg`	τ_-^{SVX}	$-$ SVX background lifetime
11	`nfrac`	F_-^{SVX}	$-$ lifetime frac. of long-lived SVX background
12	`ap`	$A_{\mathrm{P}}^{\mathrm{SST_{SVX}}}$	asymmetry of prompt $\mathrm{SST_{SVX}}$ background
13	`al`	$A_{\mathrm{L}}^{\mathrm{SST_{SVX}}}$	asymmetry of long-lived $\mathrm{SST_{SVX}}$ background
14	`epsp`	$\epsilon_{\mathrm{P}}^{\mathrm{SST_{SVX}}}$	tagging eff. for prompt $\mathrm{SST_{SVX}}$ background
15	`epsl`	$\epsilon_{\mathrm{L}}^{\mathrm{SST_{SVX}}}$	tagging eff. for long-lived $\mathrm{SST_{SVX}}$ background
16	`apctc`	$A_{\mathrm{P}}^{\mathrm{SST_{CTC}}}$	asymmetry of prompt $\mathrm{SST_{CTC}}$ background
17	`alctc`	$A_{\mathrm{L}}^{\mathrm{SST_{CTC}}}$	asymmetry of long-lived $\mathrm{SST_{CTC}}$ background

Table 6.3: Additional free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 2 of 4)

	parameter	symbol	description
18	`epspctc`	$\epsilon_{\mathrm{P}}^{\mathrm{SST_{CTC}}}$	tagging eff. for prompt $\mathrm{SST_{CTC}}$ background
19	`epslctc`	$\epsilon_{\mathrm{L}}^{\mathrm{SST_{CTC}}}$	tagging eff. for long-lived $\mathrm{SST_{CTC}}$ background
20	`apsel`	$A_{\mathrm{P}}^{\mathrm{SLT}}$	asymmetry of prompt SLT background
21	`alsel`	$A_{\mathrm{L}}^{\mathrm{SLT}}$	asymmetry of long-lived SLT background
22	`epspsel`	$\epsilon_{\mathrm{P}}^{\mathrm{SLT}}$	tagging eff. for prompt SLT background
23	`epslsel`	$\epsilon_{\mathrm{L}}^{\mathrm{SLT}}$	tagging eff. for long-lived SLT background
24	`tscale2`	S_t^{CTC}	CTC lifetime error scale
25	`bklife2`	τ_+^{CTC}	mean lifetime of long-lived CTC background
26	`tneg2`	τ_-^{CTC}	$-$ CTC background lifetime
27	`nlfrac2`	$F_{\mathrm{L}}^{\mathrm{CTC}}$	long-lived fraction of CTC background
28	`nfrac2`	F_-^{CTC}	$-$ lifetime frac. of long-lived CTC background
29	`d0sstp`	$D_+^{\mathrm{SST_{SVX}}}$	$\mathrm{SST_{SVX}}$ + dilution
30	`d0sstn`	$D_-^{\mathrm{SST_{SVX}}}$	$\mathrm{SST_{SVX}}$ $-$ dilution
31	`d0selp`	D_+^{SLT}	SLT + dilution
32	`d0seln`	D_-^{SLT}	SLT $-$ dilution
33	`d0ctcp`	$D_+^{\mathrm{SST_{CTC}}}$	$\mathrm{SST_{CTC}}$ + dilution
34	`d0ctcn`	$D_-^{\mathrm{SST_{CTC}}}$	$\mathrm{SST_{CTC}}$ $-$ dilution
35	`epssstp`	$\epsilon_+^{\mathrm{SST_{SVX}}}$	$\mathrm{SST_{SVX}}$ + tag efficiency
36	`epssstn`	$\epsilon_-^{\mathrm{SST_{SVX}}}$	$\mathrm{SST_{SVX}}$ $-$ tag efficiency

Table 6.3: Additional free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 3 of 4)

	parameter	symbol	description
37	`epsselp`	$\epsilon_+^{\mathrm{SLT}}$	SLT + tag efficiency
38	`epsseln`	$\epsilon_-^{\mathrm{SLT}}$	SLT − tag efficiency
39	`epsctcp`	$\epsilon_+^{\mathrm{SST_{CTC}}}$	$\mathrm{SST_{CTC}}$ + tag efficiency
40	`epsctcn`	$\epsilon_-^{\mathrm{SST_{CTC}}}$	$\mathrm{SST_{CTC}}$ − tag efficiency
41	`nsigctc`	$N_{\mathrm{S}}^{\mathrm{CTC}}$	number of CTC signal events
42	`nbckctc`	$N_{\mathrm{B}}^{\mathrm{CTC}}$	number of CTC background events
43	`b0life`	τ_{B^0}	mean B^0 lifetime
44	`deltamb0`	Δm_{B^0}	B^0-$\bar{B}^0$ mixing frequency
45	`nppk`	N_{OST}^{+}	no. $J/\psi K^+$ constraining opposite-side tags
46	`nmpk`	N_{OST}^{-}	no. $J/\psi K^-$ constraining opposite-side tags
47	`empkslt`	$\epsilon_-^{\mathrm{SLT'}}$	$J/\psi K^\pm$ SLT − tag efficiency
48	`bklifep`	$\tau_{+'}^{\mathrm{SVX}}$	2nd + SVX background lifetime
49	`nfracp`	$F_{+'}^{\mathrm{SVX}}$	2nd + SVX background fraction
50	`mscalec`	S_m^{CTC}	CTC B-mass error scale
51	`mslopepc`	$\Sigma_m^{\mathrm{P_{CTC}}}$	CTC mass-slope of the prompt background
52	`mslopelc`	$\Sigma_m^{\mathrm{L_{CTC}}}$	CTC mass-slope of the long-lived background
53	`nppksst`	N_{SST}^{+}	no. $J/\psi K^+$ constraining SST
54	`nmpksst`	N_{SST}^{-}	no. $J/\psi K^-$ constraining SST
55	`emls`	$\epsilon_-^{\mathrm{SST'_{SVX}}}$	$J/\psi K^\pm$ $\mathrm{SST_{SVX}}$ − tag efficiency
56	`dmls`	$D_-^{\mathrm{SST'_{SVX}}}$	$J/\psi K^\pm$ $\mathrm{SST_{SVX}}$ − tag dilution
57	`emlc`	$\epsilon_-^{\mathrm{SST'_{CTC}}}$	$J/\psi K^\pm$ $\mathrm{SST_{CTC}}$ − tag efficiency
58	`dmlc`	$D_-^{\mathrm{SST'_{CTC}}}$	$J/\psi K^\pm$ $\mathrm{SST_{CTC}}$ − tag dilution

Table 6.3: Additional free parameters in mass-time-tag fit for $\sin 2\beta$ with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 4 of 4)

	parameter	symbol	description
59	`dslope`	α	signal coefficient of $Q_{\rm final}$ in dilution func.
60	`gwidth1`	σ_1	width of wide Gaussian component of efficiency func.
61	`gwidth2`	σ_2	width of narrow Gaussian component of efficiency func.
62	`esjetf`	$f_{\rm S}$	frac. of wide Gaussian in SVX signal $Q_{\rm final}$ distribution
63	`epjetf`	$f_{\rm P}$	frac. of wide Gaussian in SVX prompt bg $Q_{\rm final}$ distribution
64	`eljetf`	$f_{\rm L}$	frac. of wide Gaussian in SVX long-lived bg $Q_{\rm final}$ distribution
65	`effs`	$\epsilon_{\rm 0S}^{\rm SVX}$	jet-clustering efficiency for SVX signal
66	`effp`	$\epsilon_{\rm 0P}^{\rm SVX}$	jet-clustering efficiency for SVX prompt bg
67	`effl`	$\epsilon_{\rm 0L}^{\rm SVX}$	jet-clustering efficiency for SVX long-lived bg
68	`esjetfc`	$f_{\rm S}^{\rm CTC}$	frac. of wide Gaussian in CTC signal $Q_{\rm final}$ distribution
69	`epjetfc`	$f_{\rm P}^{\rm CTC}$	frac. of wide Gaussian in CTC prompt bg $Q_{\rm final}$ distribution
70	`eljetfc`	$f_{\rm L}^{\rm CTC}$	frac. of wide Gaussian in CTC long-lived bg $Q_{\rm final}$ distribution
71	`effsc`	$\epsilon_{\rm 0S}^{\rm CTC}$	jet-clustering efficiency for CTC signal
72	`effpc`	$\epsilon_{\rm 0P}^{\rm CTC}$	jet-clustering efficiency for CTC prompt bg
73	`efflc`	$\epsilon_{\rm 0L}^{\rm CTC}$	jet-clustering efficiency for CTC long-lived bg
74	`chebpar1`	c_1	coeff. of first Chebyshev Polynomial in efficiency func.
75	`chebpar2`	c_3	coeff. of third Chebyshev Polynomial in efficiency func.
76	`mspsip`	$\alpha_{\psi(2S)}$	multiplier for $B^0/\bar{B}^0 \to \psi(2S)K_S$ mass error scales w.r.t. $B^0/\bar{B}^0 \to J/\psi K_S$

Table 6.4: $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 1 of 4)

	parameter	symbol	value	parabolic error
1	`mscale`	S_m	1.388	0.107
2	`mslopep`	Σ_m^{P}	0.011	0.002
3	`mslopel`	Σ_m^{L}	0.003	0.006
4	`tscale`	S_t^{SVX}	1.024	0.026
5	`bklife`	τ_+^{SVX}	1.109	0.082
6	`sin2beta`	$\sin 2\beta$	0.943	0.378
7	`nsig`	$N_{\mathrm{S}}^{\mathrm{SVX}}$	221.01	18.683
8	`nbck`	$N_{\mathrm{B}}^{\mathrm{SVX}}$	1812.0	44.044
9	`nlfrac`	$F_{\mathrm{L}}^{\mathrm{SVX}}$	0.197	0.015
10	`tneg`	τ_-^{SVX}	0.760	0.143
11	`nfrac`	F_-^{SVX}	0.147	0.030
12	`ap`	$A_{\mathrm{P}}^{\mathrm{SST_{SVX}}}$	0.019	0.035
13	`al`	$A_{\mathrm{L}}^{\mathrm{SST_{SVX}}}$	0.077	0.077
14	`epsp`	$\epsilon_{\mathrm{P}}^{\mathrm{SST_{SVX}}}$	0.669	0.013
15	`epsl`	$\epsilon_{\mathrm{L}}^{\mathrm{SST_{SVX}}}$	0.797	0.029
16	`apctc`	$A_{\mathrm{P}}^{\mathrm{SST_{CTC}}}$	0.013	0.032
17	`alctc`	$A_{\mathrm{L}}^{\mathrm{SST_{CTC}}}$	0.018	0.063

Table 6.4: $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 2 of 4)

	parameter	symbol	value	parabolic error
18	`epspctc`	$\epsilon_{\mathrm{P}}^{\mathrm{SST_{CTC}}}$	0.756	0.012
19	`epslctc`	$\epsilon_{\mathrm{L}}^{\mathrm{SST_{CTC}}}$	0.854	0.021
20	`apsel`	$A_{\mathrm{P}}^{\mathrm{SLT}}$	0.002	0.095
21	`alsel`	$A_{\mathrm{L}}^{\mathrm{SLT}}$	−0.290	0.144
22	`epspsel`	$\epsilon_{\mathrm{P}}^{\mathrm{SLT}}$	0.048	0.004
23	`epslsel`	$\epsilon_{\mathrm{L}}^{\mathrm{SLT}}$	0.086	0.013
24	`tscale2`	S_t^{CTC}	1.692	0.047
25	`bklife2`	τ_+^{CTC}	4.095	0.262
26	`tneg2`	τ_-^{CTC}	4.342	0.375
27	`nlfrac2`	$F_{\mathrm{L}}^{\mathrm{CTC}}$	0.250	0.017
28	`nfrac2`	F_-^{CTC}	0.354	0.028
29	`d0sstp`	$D_+^{\mathrm{SST_{SVX}}}$	0.190	0.045
30	`d0sstn`	$D_-^{\mathrm{SST_{SVX}}}$	0.158	0.043
31	`d0selp`	D_+^{SLT}	0.688	0.188
32	`d0seln`	D_-^{SLT}	0.431	0.200
33	`d0ctcp`	$D_+^{\mathrm{SST_{CTC}}}$	0.190	0.051
34	`d0ctcn`	$D_-^{\mathrm{SST_{CTC}}}$	0.179	0.049
35	`epssstp`	$\epsilon_+^{\mathrm{SST_{SVX}}}$	0.360	0.020
36	`epssstn`	$\epsilon_-^{\mathrm{SST_{SVX}}}$	0.349	0.020
37	`epsselp`	$\epsilon_+^{\mathrm{SLT}}$	0.034	0.009
38	`epsseln`	$\epsilon_-^{\mathrm{SLT}}$	0.035	0.009
39	`epsctcp`	$\epsilon_+^{\mathrm{SST_{CTC}}}$	0.403	0.022
40	`epsctcn`	$\epsilon_-^{\mathrm{SST_{CTC}}}$	0.390	0.021

Table 6.4: $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 3 of 4)

	parameter	symbol	value	parabolic error
41	`nsigctc`	$N_{\mathrm{S}}^{\mathrm{CTC}}$	233.33	28.184
42	`nbckctc`	$N_{\mathrm{B}}^{\mathrm{CTC}}$	2561.7	55.876
43	`b0life`	τ_{B^0}	1.530	0.038
44	`deltamb0`	Δm_{B^0}	0.467	0.018
45	`nppk`	N_{OST}^{+}	544.71	31.341
46	`nmpk`	N_{OST}^{-}	456.31	30.058
47	`empkslt`	$\epsilon_{-}^{\mathrm{SLT}'}$	0.034	0.005
48	`bklifep`	$\tau_{+'}^{\mathrm{SVX}}$	0.100	constant
49	`nfracp`	$F_{+'}^{\mathrm{SVX}}$	0.000	constant
50	`mscalec`	S_{m}^{CTC}	1.667	0.255
51	`mslopepc`	$\Sigma_{m}^{\mathrm{P}_{\mathrm{CTC}}}$	0.008	0.002
52	`mslopelc`	$\Sigma_{m}^{\mathrm{L}_{\mathrm{CTC}}}$	0.010	0.005
53	`nppksst`	N_{SST}^{+}	294.65	22.996
54	`nmpksst`	N_{SST}^{-}	359.09	25.807
55	`emls`	$\epsilon_{-}^{\mathrm{SST}'_{\mathrm{SVX}}}$	0.300	0.012
56	`dmls`	$D_{-}^{\mathrm{SST}'_{\mathrm{SVX}}}$	0.171	0.085
57	`emlc`	$\epsilon_{-}^{\mathrm{SST}'_{\mathrm{CTC}}}$	0.324	0.012
58	`dmlc`	$D_{-}^{\mathrm{SST}'_{\mathrm{CTC}}}$	0.197	0.078

Table 6.4: $\sin 2\beta$ fit result with $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ (part 4 of 4)

	parameter	symbol	value	parabolic error
59	`dslope`	α	1.330	0.365
60	`gwidth1`	σ_1	0.150	0.004
61	`gwidth2`	σ_2	0.086	0.009
62	`esjetf`	f_{S}	1.000	constant
63	`epjetf`	f_{P}	0.846	0.091
64	`eljetf`	f_{L}	1.000	constant
65	`effs`	$\epsilon_{0\mathrm{S}}^{\mathrm{SVX}}$	0.587	0.041
66	`effp`	$\epsilon_{0\mathrm{P}}^{\mathrm{SVX}}$	0.572	0.014
67	`effl`	$\epsilon_{0\mathrm{L}}^{\mathrm{SVX}}$	0.620	0.036
68	`esjetfc`	$f_{\mathrm{S}}^{\mathrm{CTC}}$	1.000	constant
69	`epjetfc`	$f_{\mathrm{P}}^{\mathrm{CTC}}$	0.658	0.098
70	`eljetfc`	$f_{\mathrm{L}}^{\mathrm{CTC}}$	0.445	0.137
71	`effsc`	$\epsilon_{0\mathrm{S}}^{\mathrm{CTC}}$	0.427	0.053
72	`effpc`	$\epsilon_{0\mathrm{P}}^{\mathrm{CTC}}$	0.562	0.014
73	`efflc`	$\epsilon_{0\mathrm{L}}^{\mathrm{CTC}}$	0.699	0.028
74	`chebpar1`	c_1	0.157	0.038
75	`chebpar2`	c_3	-0.041	0.018
76	`mspsip`	$\alpha_{\psi(2S)}$	1.675	0.377

Table 6.5: Fit $\sin 2\beta$ results

data	tag(s)	$\sin 2\beta$	+ error	− error
$B^0/\bar{B}^0 \to J/\psi K_S$	all	0.94	0.39	0.37
and	SST	2.25	0.81	0.73
$B^0/\bar{B}^0 \to \psi(2S)K_S$	SLT	0.31	0.57	0.72
	Q_{final}	0.57	0.85	0.76
$B^0/\bar{B}^0 \to J/\psi K_S$	all	0.92	0.37	0.38
	SST	2.03	0.84	0.77
	SLT	0.52	0.61	0.75
	Q_{final}	0.57	0.88	0.79

6.6 The Goodness-of-fit

The goodness of the $\sin 2\beta$ fit should indicate the level of agreement between the theoretical event distribution and the one experimentally observed. The statistics section of Particle Data Book [27] gives a formula for evaluating the goodness-of-fit for Poisson-distributed events in a counting experiment. The complete sample of events is binned and for each bin the counts are compared between prediction and observation.

6.6.1 Event binning in 5 dimensions

The log-likelihood function has 5 dimensions—SST, SLT, Q_{final}, B normalized mass, and B lifetime, that is, each event's probability is the product of the five component probabilities. This 5-D space is binned into $3 \times 3 \times 5 \times 5 \times 5$ bins, in the above order. Each dimension is binned between its lower and upper limits. The SST and SLT naturally have $-$, 0 and $+$ three bins. The Q_{final} tag ranges $\in [-0.5, 0.5]$; the normalized B mass ranges $\in [-20, 20]$; the B lifetime ranges $\in [-10, 10]$ ps. For convenience, in each dimension the central bin is indexed as 0. For each bin, the observed number of events is simply counted; the predicted number of events is the sum of all events' p.d.f. integrated in that bin. Table 6.6 shows the comparison for leading bins.

Table 6.6: Event Distribution Comparison (part 1 of 2)

		Bin Index			Observation	Prediction	$\pounds$	Expectation	variance
SST	SLT	Q	$c\tau$	m					
1	0	0	0	0	239	252.124	0.695	1.001	2.003
-1	0	0	0	0	258	242.188	1.011	1.001	2.003
1	0	0	0	2	211	208.517	0.029	1.001	2.003
1	0	0	0	1	163	200.351	7.441	1.001	2.003
-1	0	0	0	2	186	200.106	1.019	1.001	2.003
-1	0	0	0	1	184	192.228	0.357	1.001	2.004
0	0	0	0	0	208	177.733	4.884	1.001	2.004
1	0	0	0	-1	168	174.618	0.254	1.001	2.004
-1	0	0	0	-1	149	167.399	2.101	1.001	2.004
1	0	0	0	-2	179	157.051	2.934	1.001	2.004
0	0	0	0	2	176	152.410	3.476	1.001	2.004
-1	0	0	0	-2	137	150.450	1.240	1.001	2.004
0	0	0	0	1	156	145.444	0.748	1.001	2.005
0	0	0	0	-1	133	125.845	0.399	1.001	2.005
0	0	0	0	-2	121	113.211	0.524	1.001	2.006

Table 6.6: Event Distribution Comparison (part 2 of 2)

		Bin Index			Observation	Prediction	$\pounds$	Expectation	variance
SST	SLT	Q	$c\tau$	m					
1	0	0	1	0	30	49.041	8.595	1.003	2.014
-1	0	0	1	0	62	47.406	4.092	1.004	2.015
1	0	1	0	0	53	44.268	1.619	1.004	2.016
-1	0	1	0	0	51	42.349	1.658	1.004	2.017
1	0	1	0	2	41	35.789	0.725	1.005	2.020
1	0	-1	0	0	41	35.782	0.726	1.005	2.020
1	0	1	0	1	35	34.414	0.010	1.005	2.021
-1	0	-1	0	0	31	34.334	0.335	1.005	2.021
-1	0	1	0	2	39	34.228	0.636	1.005	2.021
-1	0	1	0	1	31	32.901	0.112	1.005	2.022
0	0	1	0	0	30	31.136	0.042	1.006	2.023
1	0	-1	0	2	22	30.088	2.400	1.006	2.024
1	0	1	0	-1	34	30.039	0.501	1.006	2.024
0	0	0	1	0	41	29.396	4.075	1.006	2.024
1	0	-1	0	1	42	28.842	5.254	1.006	2.025
-1	0	-1	0	2	31	28.777	0.167	1.006	2.025
-1	0	1	0	-1	29	28.687	0.003	1.006	2.025
-1	0	-1	0	1	26	27.581	0.092	1.006	2.026

6.6.2 Goodness-of-fit $\pounds$

If the number of events in each bin is Poisson distributed, the quantity of merit indicating the goodness-of-fit for that bin is [27]

$$\pounds_i = [2(N_i^{\rm th} - N_i^{\rm obs}) + 2N_i^{\rm obs} \ln(N_i^{\rm obs}/N_i^{\rm th})], \tag{6.21}$$

where $N_i^{\rm obs}$ is the number of events observed in the ith bin and $N_i^{\rm th}$ is the number of events predicted by theory (p.d.f.). The expectation value and variance for $\pounds$ can be calculated [28] and are listed in Table 6.6.

Extra caution should be taken in interpreting the quantity $\pounds$, since it is only a good approximation when $N^{\rm th}$ is large [28]. In any case, the goodness-of-fit should be judged using a bin-by-bin comparison between observed and predicted values. For an overall evaluation of the goodness-of-fit, the summations should be used. We have $\pounds = \sum \pounds_i = 485.09$, with its expectation value $\langle \pounds \rangle = 448.78$ and its variance $\sigma_{\pounds}^2 = 735.14$, corresponding to an uncertainty of $\sqrt{\sigma_{\pounds}} = 27.11$. This is approximately an 1.3σ difference.

6.7 Fits to Toy Monte Carlo Data

To ensure the fit returns a valid result, that is, parameters fitted are consistent with parameters generated, Monte Carlo simulations of Run 1 $B^0/\bar{B}^0 \to J/\psi K_S$ are generated and fitted. The main focus is the $\sin 2\beta$ parameter, whose pull $((\sin 2\beta^{\rm Fit} - \sin 2\beta^{\rm Gen})/\sigma_{\sin 2\beta})$ distribution is investigated to check the behavior of the fitter.

6.7.1 Generation of tags

Each sample is generated in such a way that it resembles the data in every aspect from the fitter's standpoint [23], e.g. mass, lifetime and tags, signal, background, except that we assign an arbitrary value to $\sin 2\beta$. As an example, the generation of the Q_{final} tag is shown here.

```
float penn_tag(float eps,float f, float s1, float s2,
 float ch1, float ch3, int btyp,
double (*ru)(void), float slope){


  float offset;
  offset=(f*s1*s1+(1-f)*s2*s2)*(2*ch1-6*ch3);
  offset=offset + (f*s1*s1*s1*s1+(1-f)*s2*s2*s2*s2)*96*ch3;
  offset=-eps*slope*offset;


  if( ru()<(1.0-eps)*(1.0-btyp*offset) ) return 0.0;


  float q,prob;


  do {
    if( ru()<f ){
      q = s1*gauss_gen(ru);
    }else{
```

```
      q = s2*gauss_gen(ru);
    }
    //Chebyshev odd terms for asymmetric tagging
    const float ch = (2*q)*(ch1+ch3*(16*q*q-3));
    if(ch>=0){
      ;
    }else if(ru()+ch>0){
      ;
    }else{
      q=-q;
    }
    const float qmax = 0.5;
    if (q > qmax) q =  qmax;
    if (q <-qmax) q = -qmax;




    prob = 0.5*(1.0 - btyp*( slope*q + offset)) ;
  } while ( ru() > prob );


  return q;


}
```

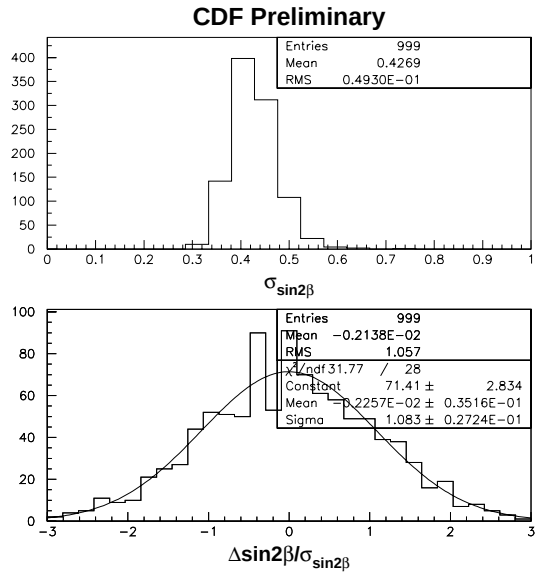


Figure 6.14: Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 0.25$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 0.25$)/$\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is -0.002 ± 0.035, which is consistent with zero.

This function takes a b quark type as well as the tagging parameters $(\epsilon_0, f, \sigma_1, \sigma_2, c_1, c_3)$ and returns a Q_{final} value accordingly based on the underlying probability.

The number of events that are generated corresponds to the total number of $B^0/\bar{B}^0 \to J/\psi K_S$ and $B^0/\bar{B}^0 \to \psi(2S)K_S$ events we collected. Figures 6.14-6.16 show the distribution of the 1σ MINOS errors and the pull of the $\sin 2\beta$ at different generated values. The MINOS error is calculated as the distance that $\sin 2\beta$ is allowed to move around its central value while the negative log-likelihood value increases by 0.5. The widths of the pull distribution are all close to 1. A breakdown of the pull distribution according to $\sigma_{\sin 2\beta}$, low, medium and high $\sigma_{\sin 2\beta}$, gives the widths of pull distributions consistent with each other [23]. This indicates the MINOS errors are correct estimates of the true error.

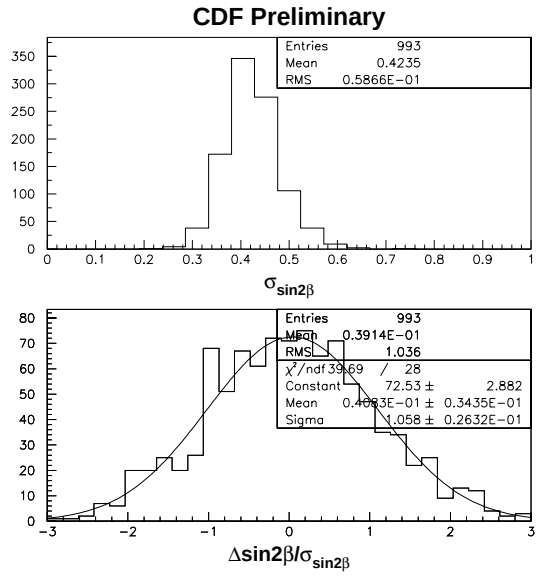


Figure 6.15: Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 0.8$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 0.8)/\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is 0.041 ± 0.034.

6.7.2 Fitter Bias of $\sin 2\beta$

Slight shifts from zero in the pull at large generated $\sin 2\beta$ values are observed in these toy Monte Carlo studies. The shift on average moves $\sin 2\beta$ away from zero and is approximately linear to the $\sin 2\beta$ value. Figure 6.17 shows the pattern and a linear fit. The bias only exists when there are multiple tags in the fitter and they have different opinions on the $\sin 2\beta$ values. There is no shift when the fit is done with each tag alone. When using the Q_{final} tag alone, the tagging parameter α returned is very close to its generated (= constrained) value; when another tag is activated, its return value fluctuates further around the constrained value. The same happens to SVX-SST and CTC-SST: individually either SVX or CTC returns its dilutions close to their generated values, but when

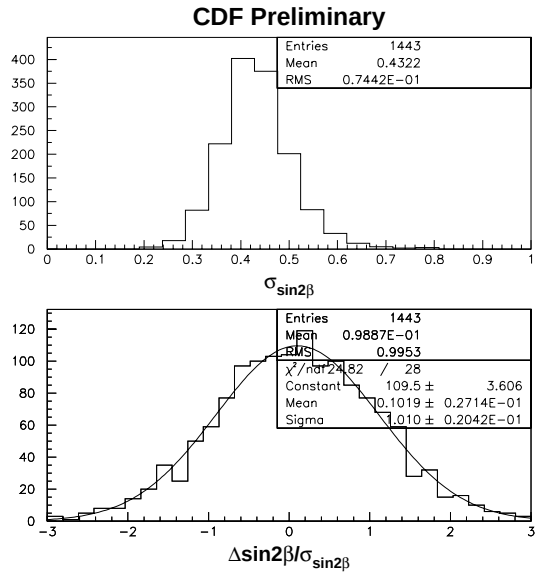


Figure 6.16: Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 1.0$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 1.0)/\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is 0.102 ± 0.027.

both are used, the dilutions fluctuate more. This communication between tags is possible because they are connected by the $\sin 2\beta$ parameter. This slight bias in the $\sin 2\beta$ result is confirmed to be associated with fits that use multiple tags in a Monte Carlo study of a simpler scenario (Appendix B), where the linear dependence of the pull on $\sin 2\beta$ is supported. Therefore we adjust the measured $\sin 2\beta$ value to remove the bias.

In order to correct for the bias in the fitter, the dependence of the bias on the true $\sin 2\beta$ must be known. This is achieved by fitting the bias vs. generated $\sin 2\beta$ value in the toy Monte Carlo as shown in Fig. 6.17. For a generalized discussion, let the relation be pull $= f(\sin 2\beta)$. If $\sin 2\beta$ is precisely measured, the bias would be determined by $f(\sin 2\beta)$. However, in reality $\sin 2\beta$ is measured with uncertainties, which we can approximate with a Gaussian distribution centered at

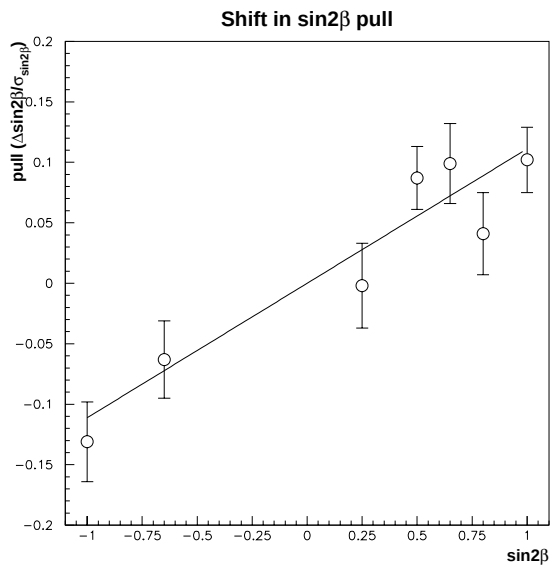


Figure 6.17: The average pull shift at different generated $\sin 2\beta$ values. A linear fit is returns a slope of 0.111 ± 0.016. If the shift were zero, the points would be aligned around $y = 0$.

$\sin 2\beta$ with a width of $\sigma_{\sin 2\beta}$. Given this p.d.f. of the $\sin 2\beta$ the p.d.f. of the pull can be derived using $p(y) = p(x)dx/dy$, where x will be replaced by $\sin 2\beta$ and y by the pull, and dx/dy is calculated from the empirical function f. A linear fit to Fig. 6.17 gives pull $= (0.111 \pm 0.016) \sin 2\beta^{\mathrm{T}}$, i.e. $dx/dy = 1/0.111$ is a constant. Therefore the p.d.f. of the pull is the p.d.f. of $\sin 2\beta$ rescaled. That is, the most probable pull value corresponds to the most probable $\sin 2\beta$ value, i.e., the measured $\sin 2\beta$.

From the pull definition

$$\text{pull} = \frac{\sin 2\beta^{\mathrm{M}} - \sin 2\beta^{\mathrm{T}}}{\sigma^{\mathrm{M}}_{\sin 2\beta}} \tag{6.22}$$

and its dependence on $\sin 2\beta^{\mathrm{T}}$

$$\text{pull} = 0.111 \times \sin 2\beta^{\mathrm{T}} \tag{6.23}$$

the adjustment should be

$$\sin 2\beta^{\mathrm{T}} = \frac{\sin 2\beta^{\mathrm{M}}}{1 + 0.111 \times \sigma^{\mathrm{M}}_{\sin 2\beta}}, \tag{6.24}$$

where the superscript "T" denotes "true" value, and "M" denotes "measured" values. It is convenient to define $\alpha_P = 1 + 0.111 \times \sigma^{\mathrm{M}}_{\sin 2\beta}$.

It is straightforward to correct the $\sigma_{\sin 2\beta}$ caused by the bias in pull. Just remember, however, there is one more piece to it— the linear function f's own uncertainty. Since the two sources of uncertainties are orthogonal, the total error is obtained through the quadratic sum of components,

$$\sigma^{\mathrm{T}}_{\sin 2\beta} = [(\frac{\sigma^{\mathrm{M}}_{\sin 2\beta}}{\alpha_P})^2 + (\frac{\sigma_f \sin 2\beta^{\mathrm{M}} \sigma^{\mathrm{M}}_{\sin 2\beta}}{\alpha_P^2})^2]^{\frac{1}{2}}. \tag{6.25}$$

where $\sigma_f = 0.016$ is the uncertainty of the slope of the pull function. The second term in the above expression is orders of magnitude smaller than the first term. A second approach using Neyman's construction of confidence interval gives the same result [30]. After correcting for the bias (a down shift of 0.03), the final result becomes

$$\sin 2\beta = 0.91 \begin{smallmatrix} +0.37 \\ -0.36 \end{smallmatrix}. \tag{6.26}$$

6.7.3 Deviation from linear calibration relation

As mentioned in Section 6.3.6, where a non-linear addition to the D vs. Q_{final} relation is shown to be insignificant, it is instructive to investigate the effect of such addition from another angle, i.e., what if the additional term in $D(Q)$ exists and we fit for $\sin 2\beta$ without (or pretending not) knowing it? To explore the possibility that some degree of deviation from the linear relation of D vs Q_{final} exists in the calibration data, we generate in the toy Monte Carlo sample the alternative D vs. Q_{final} relation as in section 6.3.6, i.e., $D = \alpha Q + \beta + \alpha_1 T_3(2Q)$ with $\alpha_1 = 0.1$ to approximately simulate

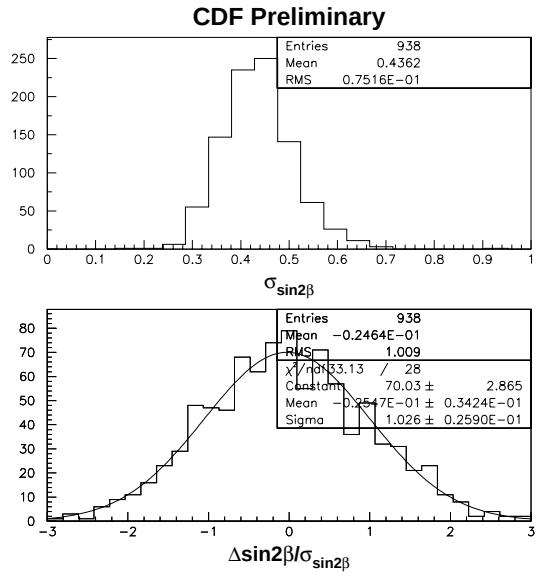


Figure 6.18: Top: Distribution of $\sigma_{\sin 2\beta}$ from multiple fits to Monte Carlo data generated with $\sin 2\beta = 1.0$ and the non-linear $D(Q)$. Bottom: Distribution of $\sin 2\beta$ pull (fit-$\sin 2\beta - 1.0)/\sigma_{\sin 2\beta}$ and a Gaussian fit. The mean pull is -0.025 ± 0.034.

the pattern in the calibration relation (Fig. 6.4). The $D(Q)$'s are shown in Fig. 6.19. Then still using the linear calibration relation without the additional term we fit to the generated samples. The resulting $\sin 2\beta$ uncertainty and pull of $\sim$ 1000 fits are shown in Fig. 6.18 with generated $\sin 2\beta = 1.0$.

The amount of change in both $\sin 2\beta$ value and its error caused by the non-linear assumption is not sufficient to justify the existence of such a term.

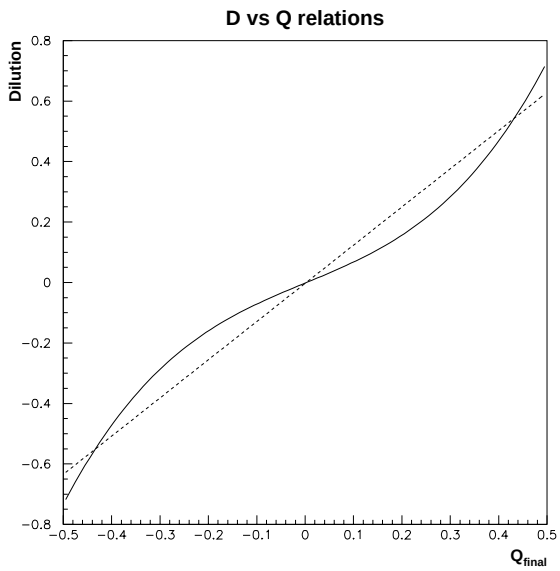


Figure 6.19: A non-linear calibration relation of D vs. Q_{final}.

Chapter 7

Conclusion

We have improved the opposite-side jet-charge tag by adding jet-related tags and a kaon dE/dx tag. These tags are combined by a generic algorithm we developed. The improvement on the tagging power ϵD^2 is seen in the Monte Carlo sample as from $(2.31 \pm 0.09)\%$ to $(3.22 \pm 0.10)\%$. Even though this improvement is not shown in the data due to limited sample size, the Q_{final} tag returns a different $\sin 2\beta = 0.57$ from the jet-charge tag. Moreover, the Q_{final} tag $\sin 2\beta$ value is much closer to the final result than the jet-charge tag was. Because the log-likelihood function is not parabolic far away from the minimum, the jet-charge tag could not contribute to reducing the overall uncertainties of $\sin 2\beta$ as Q_{final} tag does. We also increased the sample size by adding Run 1 $B^0/\bar{B}^0 \to \psi(2S)K_S$ data to $B^0/\bar{B}^0 \to J/\psi K_S$ data. The result is

$$\sin 2\beta = 0.91 \begin{array}{l} +0.37 \\ -0.36 \end{array} \tag{7.1}$$

The errors returned by the fit include both the statistical error on $\sin 2\beta$ and the systematic error from uncertainties of the tagging parameters. By fixing the dilution parameters to their constrained

values in the fit, we obtain the statistical error of $\sin 2\beta$, from which the systematic error can be calculated. The result with errors separated is

$$\sin 2\beta = 0.91 \pm 0.32(\text{stat}) \pm 0.18(\text{syst}). \tag{7.2}$$

The world average is shown in Fig. 7.1.

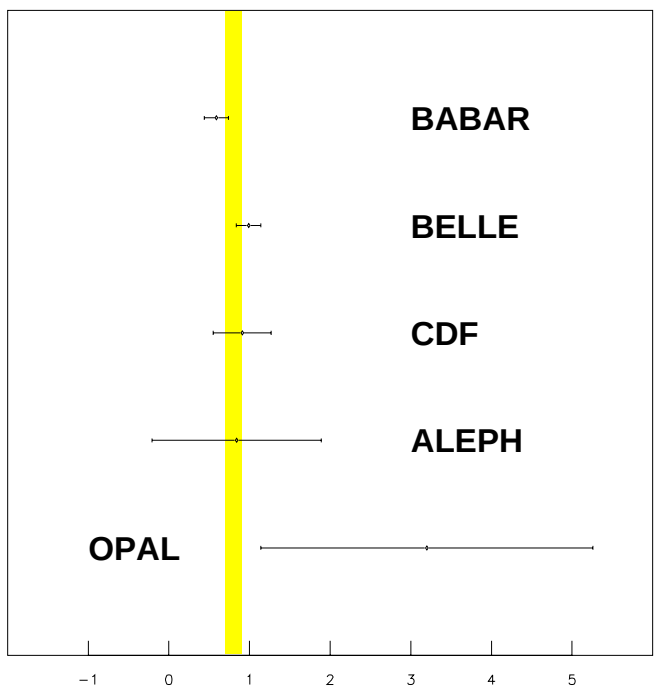


Figure 7.1: The world average is calculated using $\sin 2\beta$ results from five experiments: OPAL, ALEPH, BELLE, BABAR and CDF.

The tag-combining algorithm showed how both tagging and non-tagging information can be combined to maximize ϵD^2. This algorithm can be further generalized to include the soft-lepton tag and the same-side tag in the future flavor tagging effort. With the improved CDF Run 2 particle identification, the kaon tag will have a greater contribution in the frame of this algorithm.

Appendix A

Combining Tags Does Not Reduce ϵD^2

In flavor tagging, the tagging power is expressed in ϵD^2, where ϵ is the probability the event is tagged as Q, and D is defined through $D = 2P - 1$ where P is the probability that an event with tag Q is a b quark. We want to prove that by combining two tags ϵD^2 is no worse than the original tags.

A.1 Background

For a set of N events and a given tagging variable with value Q, assume n events are b quark events, $\bar{n}$ events are $\bar{b}$ quark events. The efficiency of obtaining this tag value Q is given by

$$\epsilon = \frac{n + \bar{n}}{N},$$

the probability of an event with tag Q being a b quark event is

$$P = \frac{n}{n + \bar{n}}$$

and the dilution of tag Q is

$$D = \frac{n - \bar{n}}{n + \bar{n}}.$$

The power of this tag is described by ϵD^2.

Question: By adding flavor-neutral information to the existing tagger, can ϵD^2 be improved?

A.2 Preparation

In the following text, $\sum$ is the sum over i, $\sum'$ is the sum over i and j where $i \neq j$. x_i is a set of measured data.

Let

$$\sum x_i = n\bar{x},$$

we have

$$\begin{aligned} n\bar{x}^2 &= \frac{1}{n}(\sum x_i)^2 \\ &= \frac{1}{n}(\sum x_i^2 + \sum{}' x_i x_j) \\ &\leq \frac{1}{n}(\sum x_i^2 + \frac{1}{2}\sum{}'(x_i^2 + x_j^2)) \end{aligned}$$

where $x^2 + y^2 \geq 2xy$ is used. But

$$\begin{aligned} \sum{}'(x_i^2 + x_j^2) &= \sum{}' x_i^2 + \sum{}' x_j^2 \\ &= 2\sum{}' x_i^2 \\ &= 2\sum_i \sum_{j, j\neq i} x_i^2 \\ &= 2\sum_i (n-1) x_i^2 \end{aligned}$$

$$= \quad 2(n-1)\sum_i x_i^2,$$

insert this relation into the previous equation, we have

$$n\bar{x}^2 \quad \leq \quad \frac{1}{n}(\sum x_i^2 + (n-1)\sum x_i^2)$$

the r.h.s. is simply $\sum x_i^2$. Hence,

$$n\bar{x}^2 \leq \sum x_i^2.$$

A.3 Flavor-neutral Information

Suppose some additional information leads to the $n + \bar{n}$ tagged events being divided into smaller groups, i.e., each with efficiency ϵ_i and dilution D_i,

$$\sum \epsilon_i = \epsilon,$$

$$\sum n_i = n,$$

$$\sum \bar{n}_i = \bar{n},$$

$$D_i = \frac{n_i - \bar{n}_i}{n_i + \bar{n}_i}.$$

Define the relation

$$\sum \epsilon_i D_i = \epsilon \bar{D}$$

we have, from the previous discussion,

$$\sum \epsilon_i D_i^2 \geq \epsilon \bar{D}^2.$$

but

$$
\begin{aligned}
\bar{D} &= \frac{1}{\epsilon}\sum \epsilon_i D_i \\
&= \frac{1}{n+\bar{n}}\sum (n_i + \bar{n_i})\frac{n_i - \bar{n_i}}{n_i + \bar{n_i}} \\
&= \frac{1}{n+\bar{n}}\sum (n_i - \bar{n_i}) \\
&= \frac{n-\bar{n}}{n+\bar{n}} \\
&= D.
\end{aligned}
$$

Thus we proved by adding new information to a tag, ϵD^2 cannot get worse.

Appendix B

A Model for Bias in $\sin 2\beta$ Toy MC

This appendix, authored by Joel Heinrich of University of Pennsylvania, is included in this thesis because it is not yet publicly accessible.

B.1 Introduction

During the course of running the toy MC to check the $\sin 2\beta$ analysis we uncovered a bias that occurs when we generate large $\sin 2\beta$. However, in general, this bias occurs only when we have two or more independent sets of dilutions floating in the fit. For example, the likelihood function has separate dilutions for SST-SVX and SST-CTC-only. If we only produce SVX events in the toy MC, the $[\sin 2\beta(\text{fit}) - \sin 2\beta(\text{generated})]/\sigma_{\sin 2\beta}$, or pull, distribution is well centered on zero. The same is true when we fit CTC-only toy MC samples (see lines 1 and 2 in the table below).

The bias does happen when we fit to a combination of SVX and CTC-only type events—the mean pull is then 0.106 ± 0.031 (line 3), which is 3σ above zero. This effect is repeatable; it is not a

statistical fluctuation. It seems to happen whenever we simulate multiple tags, the SST-SVX and SST-CTC-only case is only one example.

In lines 1 and 2 of the table, the dilutions, although floating, always stay very close to their constrained values. This is because the dilutions appear in the likelihood as $D \sin 2\beta$, and $\sin 2\beta$ is unknown. In multiple tag fits (like line 3), the fit's final values for the dilutions vary from fit to fit. This happens because the data for one tag, together with the dilution constraint for that tag, determine $\sin 2\beta$. Using that $\sin 2\beta$, the fitter can get a better estimate of second tag's dilution, based on the second tag's dilution constraint and the estimated $\sin 2\beta$ (If you know $\sin 2\beta$ you can calculate the dilution of a tagging algorithm from $J/\psi K_S^0$ data.) The result is that each tag pulls on the other tag's dilutions.

Same Side Tag 1000 Fit Toy MC Runs

All cases generated $\sin 2\beta = 1.0$

Mean pull is from fit to Gaussian

	Description	Mean Pull
1	SVX	-0.001 ± 0.033
2	CTC-only	0.026 ± 0.031
3	SVX & CTC-only	0.106 ± 0.031
4	SVX & CTC-only, all dilutions fixed	-0.020 ± 0.033
5	SVX & CTC-only, only SVX dilutions float	0.036 ± 0.033
6	SVX & CTC-only, only CTC-only dilutions float	-0.016 ± 0.032

In line 4, we do not permit the dilutions for the SVX and CTC-only tags to float: they are fixed rigidly to their constrained values (the constrained values of the dilutions are also the generated

values). In this case the shift (or bias) in the pull vanishes. The same is true if we only fix one of the tag's dilutions, while letting the other's dilutions float (lines 5 and 6).

The bias in the pull (in the 2-tag case) is ~ 0.1 when we generate $\sin 2\beta = 1.0$, consistent with zero when we generate $\sin 2\beta = 0.0$, and ~ -0.1 when we generate $\sin 2\beta = -1.0$. Having exhausted other explanations, we are faced with the possibility that the bias observed in the toy MC is intrinsic to the form of the likelihood itself. This would not be totally unexpected, but in the past we have not seen clear evidence of this. The next section will show that such biases do exist in a very simple model of the $\sin 2\beta$ fit.

B.2 Model $\sin 2\beta$ Fit

The purpose of the model $\sin 2\beta$ fit is to preserve the essential features of the real fit, while simplifying as much a possible. This allows us to see more clearly what is going on, and enables us to run many more fits than is possible in the case of the toy MC. We run 1000 toy MC fits in about 24 hours; the model fit runs at a rate of $\sim 10^6$ fits per minute.

The following expression gives the negative log likelihood for the model fit:

$$\frac{1}{2}\left(\frac{D_1 \sin 2\beta - A_1}{\sigma_{A_1}}\right)^2 + \frac{1}{2}\left(\frac{D_2 \sin 2\beta - A_2}{\sigma_{A_2}}\right)^2 + \frac{1}{2}\left(\frac{D_1 - C_1}{\sigma_{C_1}}\right)^2 + \frac{1}{2}\left(\frac{D_2 - C_2}{\sigma_{C_2}}\right)^2$$

The case we are addressing has two mutually exclusive tags with dilutions D_1 and D_2 (similar to the SST-SVX and SST-CTC-only case). The dilutions are constrained to the values $C_1 \pm \sigma_{C_1}$ and $C_2 \sigma_{C_2}$, which in the real fit would come from calibration data. The two values A_1 and A_2 represent the $J/\psi K_S^0$ data. That is, $A_1 \pm \sigma_{A_1}$ is the best estimate of $D_1 \sin 2\beta$ using the data. For the model fits we will generate A_1 and A_2 directly. The model fit differs from the real fit mainly in the shape

of the log likelihood: the real fit's log likelihood as a function of $D \sin 2\beta$ is not parabolic in general. It will be seen, however, that even this simple model will lead to a bias in the pull distribution.

One could minimize the above expression for $\sin 2\beta$, D_1, and D_2 simultaneously, get the 3×3 error matrix, etc., but, because we are in fact interested only in $\sin 2\beta$ it is convenient to eliminate D_1 and D_2 first. For a fixed value of $\sin 2\beta$ we can trivially minimize the log likelihood above with respect to D_1 and D_2, giving:

$$D_1 = \frac{A_1 \sin 2\beta\, \sigma_{C_1}^2 + C_1 \sigma_{A_1}^2}{\sin^2 2\beta\, \sigma_{C_1}^2 + \sigma_{A_1}^2} \qquad D_2 = \frac{A_2 \sin 2\beta\, \sigma_{C_2}^2 + C_2 \sigma_{A_2}^2}{\sin^2 2\beta\, \sigma_{C_2}^2 + \sigma_{A_2}^2}$$

Substituting these equations back into the negative log likelihood given above yields

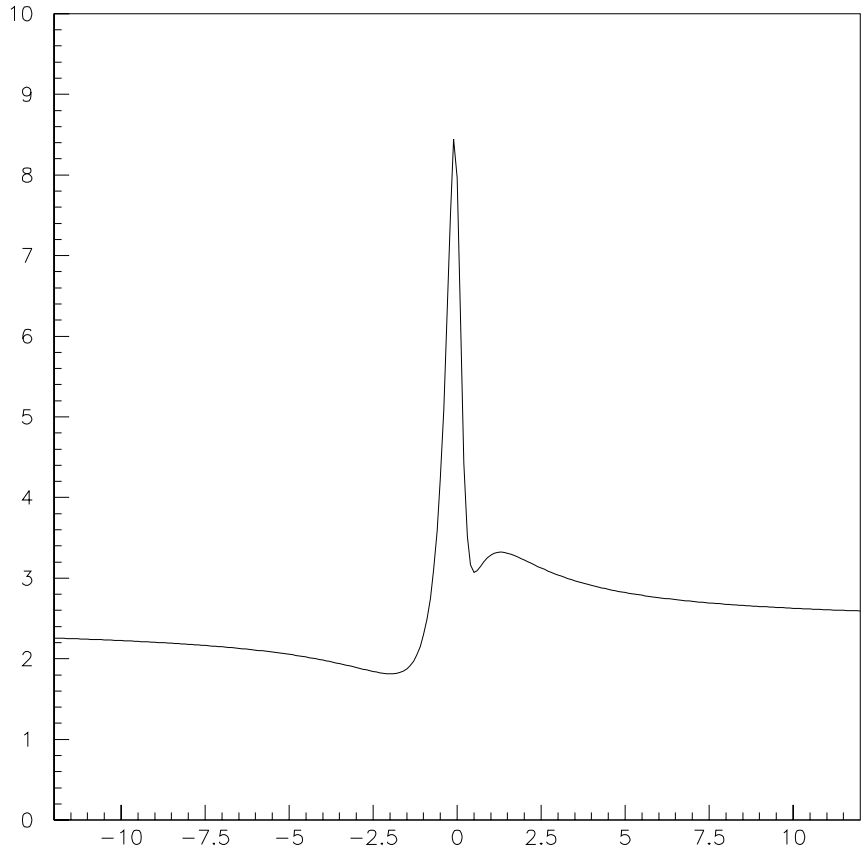
$$\ell(\sin 2\beta) = \frac{1}{2}\left[\frac{(C_1 \sin 2\beta - A_1)^2}{\sin^2 2\beta\, \sigma_{C_1}^2 + \sigma_{A_1}^2} + \frac{(C_2 \sin 2\beta - A_2)^2}{\sin^2 2\beta\, \sigma_{C_2}^2 + \sigma_{A_2}^2}\right]$$

Thus, instead of minimizing a function of many variables, as in the real fit, the model fit reduces to a numerical minimization of a function of only one variable. In spite of the apparently simple form assumed initially, the model $\ell(\sin 2\beta)$ is quite non-parabolic unless $\sigma_{C_1} \ll \sigma_{A_1}$ and $\sigma_{C_2} \ll \sigma_{A_2}$. For example, as $\sin 2\beta \to \pm\infty$, $\ell(\sin 2\beta) \to \frac{1}{2}[(C_1/\sigma_{C_1})^2 + (C_2/\sigma_{C_2})^2]$ (naively one might have expected $\ell(\sin 2\beta) \to \infty$), and $\ell(\sin 2\beta)$ will in general have local maxima as well as minima. For reference, the derivative $\ell'(\sin 2\beta)$ is given by

$$\frac{(C_1 \sin 2\beta - A_1)(A_1 \sin 2\beta\, \sigma_{C_1}^2 + C_1 \sigma_{A_1}^2)}{(\sin^2 2\beta\, \sigma_{C_1}^2 + \sigma_{A_1}^2)^2} + \frac{(C_2 \sin 2\beta - A_2)(A_2 \sin 2\beta\, \sigma_{C_2}^2 + C_2 \sigma_{A_2}^2)}{(\sin^2 2\beta\, \sigma_{C_2}^2 + \sigma_{A_2}^2)^2}$$

and the expression for $\ell''(\sin 2\beta)$, while straightforward to calculate, is too big to conveniently reproduce here.

The following plot shows a very pathological case of $\ell(\sin 2\beta)$ that has two local minima and two local maxima:



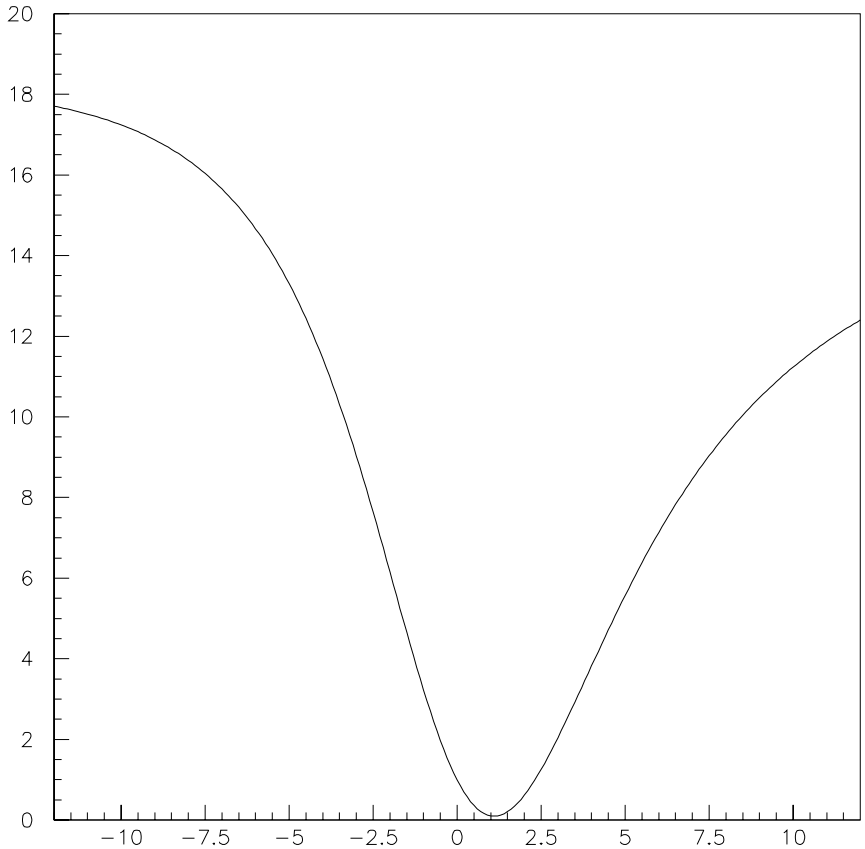
$\ell(\sin 2\beta)$ vs $\sin 2\beta$.

$$A_1 = -0.18 \pm 0.10,\ A_2 = 0.1425 \pm 0.0400$$

$$C_1 = 0.18 \pm 0.10,\ C_2 = 0.19 \pm 0.15$$

What allows this case to be so pathological is that σ_{C_1} and σ_{C_2} are rather large. Were MINUIT used to find the minimum in this example, it might fail to locate the lowest minimum, and would not be able to successfully calculate MINOS errors.

A nicely behaved (and more typical) case is shown next:



$\ell(\sin 2\beta)$ vs $\sin 2\beta$.

$$A_1 = 0.135 \pm 0.200,\ A_2 = 0.285 \pm 0.230$$

$$C_1 = 0.18 \pm 0.04,\ C_2 = 0.19 \pm 0.05$$

The only complaint one might encounter in this example is a "non-positive definite error matrix" message if MINUIT strayed too far from the minimum.

B.3 Model $\sin 2\beta$ Fit Results

To obtain the pull distribution for fits to the model, one needs to generate cases corresponding to a fixed value of $\sin 2\beta$, obtain the minimum of $\ell(\sin 2\beta)$ and the MINOS errors, and plot the resulting pulls. To generate A_1 and A_2 (which replace the $J/\psi K_S^0$ data) is trivial: A_1 is generated as a Gaussian centered on $C_1 \sin 2\beta$ with RMS equal to σ_{A_2}, for example. The following table shows the mean of the pull distribution for a series of runs with different values of generated $\sin 2\beta$.

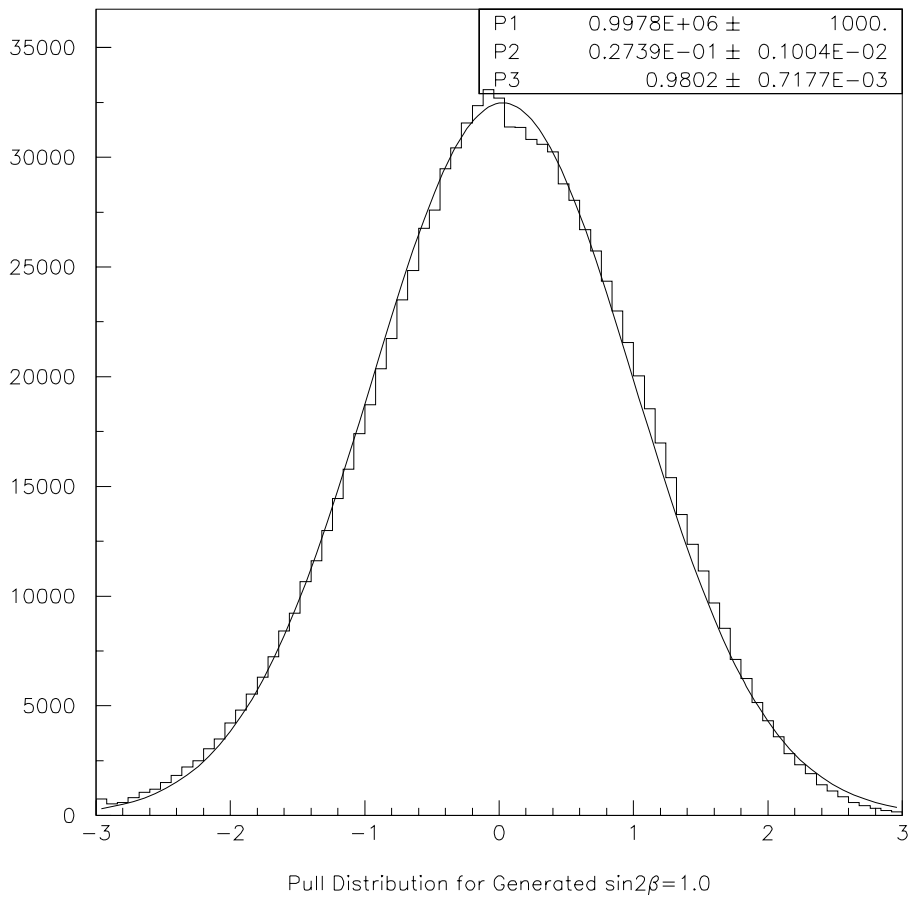
$\sin 2\beta$ (generated)	C_1	σ_{C_1}	C_2	σ_{C_2}	σ_{A_1}	σ_{A_2}	mean pull
1.00	0.18	0.04	0.19	0.05	0.20	0.23	0.027 ± 0.001
0.75	0.18	0.04	0.19	0.05	0.20	0.23	0.021 ± 0.001
0.50	0.18	0.04	0.19	0.05	0.20	0.23	0.014 ± 0.001
0.25	0.18	0.04	0.19	0.05	0.20	0.23	0.007 ± 0.001
0.00	0.18	0.04	0.19	0.05	0.20	0.23	0.000 ± 0.001
−0.25	0.18	0.04	0.19	0.05	0.20	0.23	-0.007 ± 0.001
−0.50	0.18	0.04	0.19	0.05	0.20	0.23	-0.014 ± 0.001
−0.75	0.18	0.04	0.19	0.05	0.20	0.23	-0.021 ± 0.001
−1.00	0.18	0.04	0.19	0.05	0.20	0.23	-0.028 ± 0.001

Model Fit: 10^6 Fit Runs

The size of the bias in the pull distribution for the model fits is linearly proportional to the generated $\sin 2\beta$. Because of the high statistics, one can be more certain of one's conclusions here—the 1000 fit toy MC runs can leave some doubt about whether a bias is present or not in any

particular case, due to statistical fluctuations.

The pull distribution and a Gaussian fit for the generated $\sin 2\beta = 1.0$ case are shown in the following figure.



With $\sim 10^6$ fits in the plot, one can easily see that the distribution is nearly Gaussian, but with a small discontinuity at zero pull where the switch is made from negative to positive MINOS errors.

B.4 Conclusions

Qualitatively, the model fits show a bias that behaves similarly to the toy MC fits. In the case of the model fits, the source of this shift in the pull distribution can only be due to the non-parabolic shape of the negative log likelihood functions. This suggests that we should expect a similar bias in the toy MC, and in the fit to the real data.

Bibliography

[1] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[2] L. Wolfenstein, Phys. Rev. Lett. **51**, 1945 (1983).

[3] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).

[4] Section 12, Particle Data Book, 2000.

[5] Abe *et al*, Nuclear Instruments and Methods in Physics Research A271 (1988) 387-403.

[6] D. E. Groom, Particle Data Book, The European Physical Journal C, vol 3 (1-4), p 1-794, 1998.

[7] A. B. Wicklund and K. Byrum, CDFNOTE 2624, Jan. 1995.

[8] H. Keutelian, CTC dE/dx Applications, Transparencies from Bottom Group Meeting, Apr. 1995.

[9] M. Peters, CDFNOTE 3807, Aug. 1996.

[10] W. Blum and L. Rolandi, Particle detection with drift chambers, Chapter 1, Springer-Verlag, 1994.

[11] O. Long, Chapter on Conversion Electron Cuts in SLT Analysis, Ph.D. thesis, University of Pennsylvania, 1997.

[12] Minimization package, CERN program library.

[13] M. Gronau, A. Nippe and J. L. Rosner, Phys. Rev. D **47** 1988 (1993).

[14] G. Bauer *et al.*, CDFNOTE 4267.

[15] P. Maksimovic *et al.*, CDFNOTE 4106; F. Abe *et al.*, Phys. Rev. Lett. **80:** 2057, 1998.

[16] M. Peters *et al.*, CDFNOTE 3809; M. Peters *et al.*, CDFNOTE 3808.

[17] N. Lockyer *et al.*, CDFNOTE 4745, 1998.

[18] K. Pitts, "Jet Charge Flavor Tagging in Fully Reconstructed B Decays", CDFNOTE 4744, Oct. 1998.

[19] F. Abe *et al.*, Phys. Rev. D **60**, 072003 (1999).

[20] J. Heinrich, "Flavor Tagging for $\sin 2\beta$: a Statistical Primer", CDFNOTE 5711, Aug. 2001.

[21] CDF Collaboration, T. Affolder et al., Phys. Rev. D 61, 072005 (2000).

[22] T. Gao, J. Heinrich, N. Lockyer, "An Algorithm For Combining Flavor Tags", CDFNOTE 5480, Feb. 2001.

[23] J. Heinrich, N. Lockyer, K. Pitts, "A Measurement of $\sin 2\beta$ Using Multiple Flavor Tags", CDFNOTE 4747, Jan. 1999. Phys. Rev. D 61, 072005 (2000)

[24] D. Gerdes, "B Identification Using Jet Probability", CDFNOTE 2023, Apr. 1993.

[25] T. Gao, J. Heinrich, N. Lockyer, "Run I dE/dx For Particle ID In B Tagging", CDFNOTE 5479, Nov. 2000.

[26] P. Sphicas, D. Vucinic, CDFNOTE 4097, Mar. 1998.

[27] Section 28.3.4, Particle Data Book, 2000.

[28] J. Heinrich, "The Log Likelihood Ratio of the Poisson Distribution for Small μ", CDFNOTE 5718, Aug. 2001.

[29] T. Gao, J. Heinrich, N. Lockyer, K. Pitts, CDFNOTE 5620, Oct. 2001.

[30] Section 28.6, Particle Data Book, 2000.