SEARCH FOR CP VIOLATION IN $\mathbf{B_S^0 \to J/\psi\phi}$

by

Chunlei Liu

B.S. in Physics, 2004, University of Science and Technology of China

M.S. in Physics, 2005, University of Pittsburgh

Submitted to the Graduate Faculty of

the Department of Physics and Astronomy in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF PHYSICS AND ASTRONOMY

This dissertation was presented

by

Chunlei Liu

It was defended on

Sep 19, 2008

and approved by

Joseph Boudreau, PhD, Professor

Adam Leibovich, PhD, Professor

James Mueller, PhD, Professor

Manfred Paulini, PhD, Professor (CMU)

Paul Shepard, PhD, Professor

Dissertation Director: Joseph Boudreau, PhD, Professor

SEARCH FOR CP VIOLATION IN $\mathbf{B_S^0 \to J/\psi\phi}$

Chunlei Liu , PhD

University of Pittsburgh, 2008

The CDF experiment, which uses $p\bar{p}$ collisions at $\sqrt{s} = 1.9$ TeV produced at the Fermilab Tevatron, is unique in its ability to observe all flavors of B hadrons and measure their properties. Among them, CP violation is of fundamental interest. In the B^+ and B^0 systems, measurements of CP violation parameters performed essentially at B factories have borne out the predictions of CKM model. Little is known, experimentally, about CP violation in the B_s^0 system. The standard model predicts very little CP violation there, thus any nonzero measurement could be an indication of new physics. In this thesis, we will report on the very first flavor-tagged analysis of approximately 2000 $B_s^0 \to J/\psi\phi$ decays reconstructed in a 1.35 fb^{-1} data sample collected at CDF. This channel is sensitive not only to the width difference $\Delta\Gamma$ in the B_s^0 system but also to the CP violation parameter β_s. The final result we obtain is a confidence region in the two dimensional space of β_s and $\Delta\Gamma$. Assuming the standard model predictions of β_s and $\Delta\Gamma$, the probability is 15%, corresponding to 1.5 Gaussian standard deviations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis cannot possibly be finished without the guidance from my advisor Professor Joseph Boudreau. I want to thank him for introducing particle physics to me and bringing me into this fascinating field. He has given me great freedom and much help throughout the past two years when I have been doing my research at Fermilab. I have also enjoyed working with many great colleagues as well. They are Dr. Karen Gibson from University of Pittsburgh, Dr. Gavril Giurgiu and Professor Petar Maksimovic from The Johns Hopkins University, Professor Manfred Paulini from Carnegie Mellon University, Dr. Juan Fernandez from C.I.E.M.A.T., Física experimental de Partículas, Dr. Luis Labarga from Dpto. F. Teórica, Universidad Autónoma de Madrid, Dr. Guillelmo Gomez-Ceballos from Massachusetts Institute of Technology. This thesis is also dedicated to my parents in China and my girl friend who have always been there for me. Finally, I want to thank all of the members who served in my thesis committee.

1.0 INTRODUCTION

When one looks into the universe, one of the biggest questions is, "Where is all the anti-matter." The matter and antimatter asymmetry can be obtained by measuring the ratios of baryons and antibaryons to photons [1]. The baryon over photon ratio turns out to be at the order of 10^{-11}, while the antibaryon over photon ratio is practically zero. This means the matter and antimatter asymmetry is 100%. in 1967, Sakharov postitulated that [2], the matter-antimatter asymmetry could be achieved with the three following ingredients: Baryon number violation, C and CP violation in the reaction, and deviation from thermal equilibrium. Thermal equilibrium can be lost as the universe expands and cools, and electroweak baryogenesis says the standard model may allow baryon number violation in exotic ways. While CP violation has been observed in the quark sector of the standard model, the problem is that it is far from enough to explain the observed matter-antimatter asymmetry in the universe. Thus, the search for CP violation beyond the standard model is key to understand our own universe.

A CP transformation is a combination of a charge conjugation transformation C and a parity transformation P. A C transformation simply changes a particle to its anti-particle, while a P transformation reverses the direction of motion of a particle and preserves the spin of that particle. Thus CP together changes a left-handed particle to a right-handed anti-particle. In quantum mechanics, a symmetry is a transformation that does not alter physical observables. Both C and P were previously believed to be symmetries of all the fundamental interactions. A C symmetry means the invariance of physical laws under a conjugation of all internal quantum numbers, while the P symmetry implies that the rate of a mirror image process, such as a chemical process or radiative process, is the same as the original one.

In 1956, Tsung-Dao Lee and Cheng Ning Yang made a critical review of all the experiments, and found parity symmetry was verified in strong interaction and electromagnetism, but never tested in weak decay [3]. At the same year, a group led by Chien-Shiung Wu found the parity is violated in the beta decay of Cobalt-60 [4]. The discovery of parity violation led to the 1957 Nobel Prize for Lee and Yang. In the standard model, both C and P are maximally violated in the weak interaction. For example, the neutrino is left-handed, but no left-handed anti-neutrino or right-handed neutrinos exist. But the combined CP will transform a left-handed neutrino into right-handed anti-neutrino. The existence of both left-handed neutrino and right-handed anti-neutrino makes it natural to propose CP symmetry. In this situation, a neutrino couples to the charged W^+ in the same way as the interaction of anti-neutrino to charged W^-.

In 1964, Cronin, Fitch and co-workers discovered for the first time that CP symmetry is violated in the kaon system [5]. $K_L \to \pi^+\pi^-$ was observed at the level of a few permil. This type of CP violation manifested itself in the decay of a CP odd eigenstate to a CP even final state. This work led to the Nobel Prize in 1980 for Cronin and Fitch.

In 1973, Kobayashi and Maskawa introduced the CKM mechanism to explain CP violation for three generations of quarks [6], which won them the Nobel Prize in 2008. In the Standard Model, with SU(2)×U(1) as the gauge group of electroweak interactions, the quark mass eigenstates are not the same as the weak flavor eigenstates. The matrix which relates the mass eigenstates to the weak eigenstates is called the Cabibbo-Kobayashi-Maskawa (CKM) quark mixing matrix $V_{\rm CKM}$, which operates by convention on the charge $-1/3$ quarks d, s and b,

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

There are only four independent variables associated with the CKM matrix, including three mixing angles and one phase. The phase is the only source of CP violation in the CKM mechanism, which successfully accommodates the small CP violation observed in kaon system. This model, which predicted the b quark before its discovery in 1977, also predicts

large CP violation in the B system. In the 1990s, the Tevatron experiment at Fermilab [7] and the boosted B factories, Babar [8] and Belle [9] began producing large numbers of b hadrons. These experiments (especially the B factories) were largely motivated by CP violation studies.

CP violation effects in the B system can be represented graphically in terms of a *unitarity triangle*, which is due to the unitarity of the CKM matrix. By multiplying the first and the third columns of the matrix, one can obtain a relation involving the two smallest elements, V_{ub} and V_{td},

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0.$$

This is a triangle relation in the complex plane, as shown in Figure 1. The magnitudes of the CKM elements determine the size of the legs of this triangle. The angles α, β and γ are related to CP violating asymmetries in B decays. If there is no CP violation, the triangle will collapse to a straight line with zero height. So by measuring the size of the angles, one can determine quantitatively the extent of CP violation in the B systems.

In 2001, CP violation was established successfully in the B systems for the first time by measuring the angle β at BaBar and Belle experiments [10, 11]. The definition of this angle is

$$\beta \equiv \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right).$$

One way to observe the CP violation is to use the interference between the direct decay of a B^0 into a CP final state f_{CP} and the decay which proceeds via $B^0 - \bar{B}^0$ mixing and interferes with the direct decay $B^0 \to f_{CP}$. One measures β by reconstructing $B^0/\bar{B}^0$ decays to the CP eigenstate $J/\psi K_S^0$, the interference of direct decays, $B^0 \to J/\psi K_s^0$, with those that undergo mixing, $B^0 \to \bar{B}^0 \to J/\psi K_s^0$. This produces a time-dependent CP asymmetry,

$$\mathcal{A}_{CP}(t) \equiv \frac{\bar{B}^0(t) - B^0(t)}{\bar{B}^0(t) + B^0(t)} = \sin 2\beta \cdot \sin \Delta m_d t,$$

where $B^0(t)$ $[\bar{B}^0(t)]$ is the number of decays to $J/\psi K_s^0$ at proper time t, given that the produced meson was a B^0 $(\bar{B}^0)$ at $t = 0$. The CP phase difference between the two decay paths appears via the factor $\sin 2\beta$, and the $B\bar{B}$ flavor oscillation is driven by the mass difference Δm_d between the two B^0 mass eigenstates.

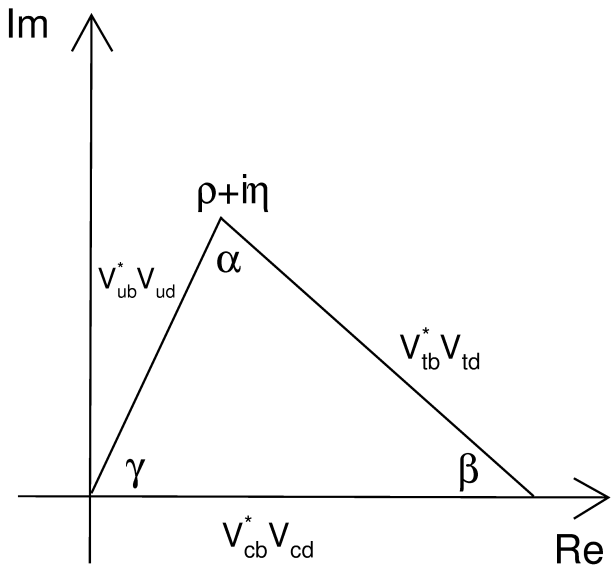

Figure 1: The CKM unitarity triangle in the complex plane.

A similar approach can be used to pursue CP violation in B_s^0 system. By contracting the second and third columns of the CKM matrix, another unitarity triangle can be obtained, and the β_s angle is defined as

$$\beta_s \equiv \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right).$$

The golden channel to measure β_s is $B_s^0 \to J/\psi\phi$. However, unlike the β measurement in $B^0 \to J/\psi K_s^0$, β_s is predicted to be nearly zero in the Standard Model. Another difference is that the non-zero decay width difference of B_s^0 system and CP eigenstates admixtures make it more complicated to measure.

The measurement of the largely unknown parameter β_s is the primary goal of this thesis. Any observation of a CP asymmetry that is significantly larger than zero will provide an unambiguous signal for new physics. Specifically, it is likely to be related to new, CP violating contributions to B_s^0-$\bar{B}_s^0$ mixing.

This thesis is organized as follows. The theory of CP asymmetries of neutral B mesons and a derivation of the time development of an admixture of CP eigenstates are discussed in chapter 2. The accelerator and detector are described in chapter 3. The data sample and event reconstruction is the subject of chapter 4, while the corresponding experimental strategies are described in chapter 5. A measurement of $\sin 2\beta$ using $B^0 \to J/\psi K_s^0$ decays used as a cross check of the opposite flavor tagging algorithms is the subject of chapter 6. The Toy Monte Carlo used to test the fitter and evaluate the likelihood function is described in chapter 7. The determination of β_s from $B_s^0 \to J/\psi\phi$ is discussed in chapter 8.

2.0 THEORY AND MOTIVATION

In this chapter, CP violation related theory will be discussed first. Then, formulas which we are going to use to measure the CP violation in the $B_s^0 \to J/\psi\phi$ channel are derived.

2.1 THE STANDARD MODEL

At the beginning of the last century, most of physics was believed to be understood. The world was constructed by solid atoms, and motion of objects was governed by Newtonian Mechanics; charged and massive objects interacted with classical fields like gravity and electromagnetism. The first subatomic particle to be discovered was the electron by Thompson in 1897. Rutherford's discovery of the nucleus in 1911 demonstrated that the positive charges were concentrated in a very small fraction of the atoms' volume. The components of the nucleus, protons, and neutrons were discovered later. The first sign of subatomic reactions came in the form of radiation discovered in 1896 by Henri Becquerel. The birth of special relativity and quantum mechanics carried modern physics into a new era. Weak interactions, and neutrinos, were proposed to explain nuclear beta decay, while the discovery of the muon in cosmic rays became the first indication of a second generation. As more and more new particles were found, theoretical and experimental evidence began to accumulate that the so called hadrons (proton, neutron, other baryons, and mesons) were not fundamental, but composite. In 1964, Murray Gell-Mann and George Zweig proposed quarks in a framework called the Eight-fold Way. The first evidence for quarks as the constituents of the proton came from deep inelastic scattering experiments at the Stanford Linear Accelerator Complex (SLAC) in 1969. A theoretical framework incorporating all of the known particles and their

	Generation 1		Generation 2		Generation 3	
Quarks	Up	u	Charm	c	Top	t
	Down	d	Strange	s	Bottom	b
Leptons	Electron	e	Muon	μ	Tau	τ
	Electron neutrino	ν_e	Muon neutrino	ν_μ	Tau neutrino	ν_τ

Table 1: Three generations of elementary particles.

interactions, called the Standard Model (SM), first by John Lliopoulos, was developed during the period 1970-1973. Early predictions of the standard model, including the c quark, b quark, W and Z bosons, and the top quark were verified one by one. In 2000, the tau neutrino, the last piece of the standard model (besides the Higgs boson) was observed at Fermilab.

In the standard model, all the elementary particles are fermions with spin 1/2 and they interact with each other by exchanging bosons which transmit gauge forces. Table 1 shows all the twelve elementary particles organized in three generations and Table 2 shows the known fundamental forces in nature.

The standard model describes three of the four known fundamental interactions between the elementary particles. It unifies the electroweak theory and quantum chromodynamics

Interaction filed	Boson	Spin
Gravitation field	Gravitons	2
Weak field	W^+, W^-, Z particles	1
Electromagnetic field	Photons	1
Strong filed	Gluons	1

Table 2: Four types of know fundamental interactions.

into a structure which can be described by gauge groups SU(3)× SU(2) × U(1). The group SU(3) is the gauge group of the theory of strong interactions known as QCD, and there are $3^2 - 1 = 8$ massless gauge bosons called "gluons", corresponding to the eight generators of the SU(3) group. The group SU(2) has three generators of gauge symmetry and would normally give three massless gauge bosons. However the gauge symmetry is broken through a mechanism called spontaneous symmetry breaking, which occurs because of a SU(2) doublet scalar field called the Higgs field. The three gauge bosons (W^+, W^- and Z^0) acquire mass by coupling to the Higgs field. These bosons mediate the weak force, while the massless photon arising from the SU(2) × U(1) group mediates the electromagnetic force. The large masses of the gauge bosons are responsible for the short range of the weak force and the "weakness" of the weak interactions. The Higgs Field, with four degrees of freedom, loses one degree of freedom to each of the massive bosons (the new "longitudinal" degree of freedom characteristic of massive spin-1 bosons). The other degree of freedom is the Higgs Boson. Direct searches for the Higgs at the e^+e^- collider LEP, at CERN, have excluded it below a mass of 114 GeV (at the 95% Confidence Level). It is currently being sought at Fermilab. As of August 2008, combined result from CDF and D0 experiments at the Tevatron excluded the Higgs boson at 170 GeV/c^2 (at 95% confidence level) [12]. If the Higgs boson is not found at Tevatron, the Large Hadron Colider (LHC) at CERN, a high luminosity machine with a center of mass energy of 14 TeV, will have a much better chance.

Although the standard model has achieved great success since its birth, it is still not a perfect theory for several reasons. The primary reason is that it does not include the gravity, one of the four known fundamental interactions. There are many free parameters in the standard model, such as the fermion masses and coupling constants, which are not predicted by the theory. There are no explanation for the three generations of elementary particles. For these reasons, there is both theoretical interest in a more complete theory and experimental interest in the search for phenomenon beyond the standard model. The first real sign of new physics came from the discovery of neutrino oscillations [13], which implies that the neutrino has non-zero mass. In this thesis we pursue a search for new physics in the quark sector, through measurements of a precisely predicted CP violation observable, which we explain in the next section.

2.2 CP VIOLATION IN THE STANDARD MODEL

In the standard model, there are only two ways to break the CP symmetry. The first of these is in the QCD Lagrangian. However, this has not been found experimentally, and one would expect either no CP violation or a CP violation that is many, many orders of magnitude too large. On the other hand, CP symmetry can be also broken in the weak interaction through the CKM mechanism, which has been verified by the experiments. Further from a theoretical point of view, the CKM model could exist along with other sources of CP violation. We therefore will discuss how the CKM mechanism explains the CP violation in the standard model.

One of the striking features of charged current weak interactions is that they couple to both vector and axial vector currents $V - A$, unlike QED and QCD which couple solely to a vector current. Consequently, the left- and right-handed fermions transform differently under the electroweak gauge group SU(2)×U(1). The right-handed fermions are singlets under SU(2), and do not couple to charge W bosons:

$$\begin{aligned} E_R &= (e_R, \mu_R, \tau_R),\ Y_E = -2; \\ U_R &= (u_R, c_R, t_R),\ Y_U = \frac{4}{3}; \\ D_R &= (d_R, s_R, b_R),\ Y_D = -\frac{2}{3}. \end{aligned} \tag{2.1}$$

On the other hand, left-handed fermions do couple to $W^{\pm}$, and are doublets under SU(2):

$$\begin{aligned} L_L &= \left(\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \right),\ Y_L = -1; \\ Q_L &= \left(\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L \right),\ Y_Q = \frac{1}{3}. \end{aligned} \tag{2.2}$$

In the above definitions, $Y = 2(Q - I_3)$ is the hypercharge, where Q is charge, and I_3 is the weak isospin. In the standard model, fermion masses, flavor violation and CP violation all arise from Yukawa interactions among the fermion fields and the Higgs field. Let us consider the simple situation where we only have the first generation leptons:

$$\mathcal{L}_Y = -ye^{i\delta}\bar{l}_L\phi e_R - ye^{-i\delta}\bar{e}_R\phi^\dagger l_L, \tag{2.3}$$

where $\bar{l}_L = (\bar{\nu}_L, \bar{e}_L)$, and y is real. To preserve Lorentz invariance ϕ must have spin 0. To preserve invariance under U(1), ϕ must have hypercharge $Y_\phi = Y_L - Y_E = +1$. To preserve invariance under SU(2), ϕ must be a doublet,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{2.4}$$

where the superscripts denote the electric charge using relationship $Q = I_3 + Y/2$. This interaction is called a Yukawa interaction, because similar interaction was introduced by Yukawa to describe decay $\pi^+ \to \mu^+\nu_\mu$, and y is called the Yukawa coupling. At first look, the interaction in Eq. 2.3 seems to be CP violated with phase δ. However, one can change the phases of l_L and e_R to remove it. Thus, the Yukawa interaction conserves CP.

The renormalizable interaction for the potential energy of the ϕ field is

$$V(\phi) = -\lambda v^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \tag{2.5}$$

with two parameters, v and λ. The vacuum state with no propagating particles is realized when one minimize $V(\phi)$. Parameter λ must be positive to obtain stable vacuum. If v^2 is negative, there is a single minimum of the potential with vacuum expectation value $\langle\phi\rangle = 0$, which does not interest us here. If v^2 is greater than 0, then $V(\phi)$ takes the shape of a saddle with three-dimensional family of minima:

$$\langle\phi\rangle = e^{i\langle\epsilon^a\rangle\sigma_a/2v} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}. \tag{2.6}$$

One can set $\langle\epsilon^a\rangle$ to 0 by proper SU(2) transformation. Although the full Lagrangian respects SU(2) $\times$ U(1) gauge symmetry, this vacuum solution does not. This is called spontaneous symmetry breaking.

If fluctuation around the expectation value is introduced, physical particles can be obtained, so one writes

$$\phi(x) = e^{i\epsilon^a(x)\sigma_a/2v} \begin{pmatrix} 0 \\ [v + h(x)]/\sqrt{2} \end{pmatrix}. \tag{2.7}$$

Masses of the physical particles are found by inserting the above equation into the interactions in the Lagrangian and examining the corresponding quadratic terms. For example, one

can obtain the electron masses $m_e = yv/\sqrt{2}$ by comparing the $\bar{e}_R e_L$ terms in L_Y. Similarly, field h (the Higgs field) is found to have a mass $m_h = \sqrt{2\lambda}v$ from $V(\phi)$. By examining the kinetic energy of the scalar filed, one obtain masses for three of the gauge bosons: $W^{\pm}$, Z^0, while photon remains massless.

The same doublet as for leptons is used to repeat the construction with $(\bar{u}_L, \bar{d}_L)$ and d_R with hypercharge $Y_Q - Y_D = +1$. But for $(\bar{u}_L, \bar{d}_L)$ and u_R, a doublet with hypercharge $Y_Q - Y_U = -1$ is necessary, so the standard model uses the charge-conjugate of the doublet used for leptons:

$$\tilde{\phi} = i\sigma_2\phi^* = \begin{pmatrix} \bar{\phi^0} \\ -\phi^- \end{pmatrix}. \tag{2.8}$$

With three generations of fermions, the full set of Yukawa interaction becomes complicated with different generations interacting with each other as in

$$\mathcal{L}_Y = -\sum_{i,j=1}^{3} \left[\hat{y}^e_{ij} \bar{L}^i_L \phi E^j_R + \hat{y}^d_{ij} \bar{Q}^i_L \phi D^j_R + \hat{y}^u_{ij} \bar{Q}^i_L \tilde{\phi} U^j_R \right]. \tag{2.9}$$

The Yukawa couplings y_{ij} could be complex in principle. For the leptons, however, as mentioned above, one can transform the lepton fields while keep the non-Yukawa part of the Lagrangian invariant:

$$\begin{aligned} E_R \to R E_R, \bar{E}_R \to \bar{E}_R R^\dagger; \\ L_L \to S L_L, \bar{L}_L \to \bar{L}_L S^\dagger, \end{aligned} \tag{2.10}$$

where $R \in U(3)_{E_R}$ and $S \in U(3)_{L_L}$. With these transformations, the new Yukawa matrix will be $y^e = S\hat{y}^e R^\dagger$. It is easy to see that, with the proper choice, y^e can be made diagonal and real. So the leptons' Yukawa interactions will conserve CP.

For the quarks, the non-Yukawa part of the Lagrangian is invariant with transformations:

$$\begin{aligned} D_R \to R_d D_R, \bar{D}_R \to \bar{U}_R R^\dagger_d; \\ U_R \to R_u U_R, \bar{U}_R \to \bar{U}_R R^\dagger_u; \\ Q_L \to S_u Q_L, \bar{Q}_L \to \bar{Q}_L S^\dagger_u. \end{aligned} \tag{2.11}$$

Although we may exploit S_u and R_u to transform Yukawa coupling matrix $\tilde{y}^u$ into a diagonal and real one y^u, coupling matrix $\tilde{y}^d$ cannot be transformed into a diagonal or real one simultaneously in general. Instead, it has the following form:

$$S_u \tilde{y}^d R_d^\dagger = S_u S_d^\dagger S_d \tilde{y}^d R_d^\dagger = V y^d, \tag{2.12}$$

where y^d is real and diagonal with proper choice of S_d, and the matrix

$$V = S_u S_d^\dagger \tag{2.13}$$

is just the Cabibbo-Kobayashi-Maskawa (CKM) matrix.

With this basis, the CKM matrix appears in the the charged current interactions (that is the $W^\pm$ interactions) for quarks

$$-\mathcal{L}_{W^\pm} = \frac{g}{\sqrt{2}} \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_L \gamma^\mu V_{\mathrm{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L W_\mu^+ + h.c. \tag{2.14}$$

To emphasize the physical transitions associated with the CKM matrix, one can write the explicit form of the CKM matrix as

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2.15}$$

In principle, magnitudes of all the matrix elements can be measured through the decay with corresponding quarks. For example $|V_{ud}|$ can be determined from nuclear β decay, $|V_{us}|$ can be measured from semileptonic kaon decays, $|V_{ub}|$ and $|V_{cb}|$ can be measured from inclusive or exclusive b decay, $|V_{cd}|$ and $|V_{cs}|$ can be extracted from semileptonic or leptonic charm decays, $|V_{td}|$ and $|V_{ts}|$ can be only accessed through box diagrams or through second order weak interactions such as QCD penguin diagrams, and $|V_{tb}|$ can be determined from top decays.

To find how many observable parameters the CKM matrix contains, one starts from $2N^2$ free parameters of an arbitrary $N \times N$ complex matrix and reduce that number by applying constraints. Because the CKM matrix is unitary, $|V_{\mathrm{CKM}}|^2 = 1$ gives N^2 constraints, so there

are N^2 parameters left. Mathematically, with the unitarity property, one can prove that of the N^2 independent parameters, there are $N(N-1)/2$ real and $N(N+1)/2$ imaginary components [14]. One can further change the phase of each of the $2N$ quark states independently without altering the physics and absorb imaginary parts, but an overall phase change still leaves $V_{\rm CKM}$ invariant, so the number of remaining independent imaginary components is $N(N+1)/2-(2N-1) = (N-1)(N-2)/2$ and the total number of independent parameters is $N^2-(2N-1) = (N-1)^2$.

For three generation of quarks with $N = 3$, the CKM matrix has four observable parameters; of them three are real and one is a phase factor. These four numbers are fundamental constants of nature, just like the Fermi coupling constant, and need to be determined from experiments. The CKM matrix can be parameterized by four parameters. One parameterization favored by the Particle Data Group uses three rotation angles, θ_{12}, θ_{23}, θ_{12} and one complex phase δ_{13}, such that

$$V_{\rm CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix}, \tag{2.16}$$

where $s_{ij} = \sin(\theta_{ij})$,$c_{ij} = \cos(\theta_{ij})$. Another popular approximation of the CKM matrix which emphasizes the hierarchy in the size of the matrix elements is due to Wolfenstein [15]. By introducing

$$\lambda \equiv s_{12}, A \equiv \frac{s_{23}}{s_{12}^2}, \rho \equiv \frac{s_{13}\cos(\delta_{13})}{s_{12}s_{23}}, \eta \equiv \frac{s_{13}\sin(\delta_{13})}{s_{12}s_{23}}, \tag{2.17}$$

the matrix can be expanded in terms of λ

$$V = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda + \frac{1}{2}A^2\lambda^5 - A^2\lambda^5(\rho + i\eta) & 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4(1+4A^2) & A\lambda^2 \\ A\lambda^3(1-\bar{\rho} - i\bar{\eta}) & -A\lambda^2 + A\lambda^4(\frac{1}{2} - \rho - i\eta) & 1 - \frac{1}{2}A^2\lambda^4 \end{pmatrix} + \mathcal{O}(\lambda^6). \tag{2.18}$$

Here, λ is also the sine of the Cabibbo angle: $\lambda = \sin\theta_{\rm c} = |V_{us}|$, and we used the shorthand notation $\bar{\rho} = \rho(1-\lambda^2/2)$, $\bar{\eta} = \eta(1-\lambda^2/2)$. From experiments [16], $\lambda \sim 0.23$, $A \sim 0.81$, $\bar{\rho} \sim 0.14$, $\bar{\eta} \sim 0.34$.

CP violation arises from the complex phase in the CKM matrix. For example, the $b \to uW$ transition in the standard model has coupling

$$\mathcal{L} = -\frac{g}{\sqrt{2}}[V_{ub}\bar{u}_L\gamma^\mu b_L W_\mu^+ + V_{ub}^*\bar{b}_L\gamma^\mu u_L W_\mu^-]. \tag{2.19}$$

With a CP transformation, one obtains

$$CP\mathcal{L}(CP)^{-1} = -\frac{g}{\sqrt{2}}[V_{ub}\bar{b}_L\gamma^\mu u_L W_\mu^- + V_{ub}^*\bar{u}_L\gamma^\mu b_L W_\mu^+]. \tag{2.20}$$

Clearly, this is equivalent to the previous equation only when $V_{ub} = V_{ub}^*$. As shown already, there is in general no way one can rotate the quark fields to remove the phase in the coupling, so CP is violated in this interaction.

One can gauge the size of the CP violation [17] with a quantity defined as

$$J_{CP} = \mathrm{Im}(V_{i\alpha}V_{j\beta}V_{i\beta}^*V_{j\alpha}^*),\ (i \neq \alpha, j \neq \beta). \tag{2.21}$$

J_{CP} has nine possible combinations arising from the unitarity of the CKM matrix, and all of them are equal up to an overall sign. Using the parameterization of the CKM matrix, J_{CP} can be calculated as $\lambda^6 A^2 \eta = \mathcal{O}(10^{-5})$, meaning that the CP violation is small in the standard model.

All the physics from the CKM matrix can be reflected in six "unitarity" triangles by applying the unitarity constraint to different columns and rows of CKM matrix. For example, if one apply unitarity to the first and third columns, one obtains

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{2.22}$$

The unitarity triangle is just a geometric representation of this equation, which is a triangle in the complex plane known as the "db" triangle. One can choose to orient the side $V_{cd}V_{cb}^*$ along the horizontal direction and to normalize it to unit length dividing by the magnitude.

In this geometrical representation (Figure 2) the angles of the triangle are related to the CP phase according to

$$\alpha \equiv \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \tag{2.23}$$

$$\beta \equiv \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \tag{2.24}$$

$$\gamma \equiv \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{2.25}$$

These angles can be measured through time-dependent CP asymmetries in exclusive channels; for example, the angle β can be determined from $B^0 \to J/\psi K_s^0$ decays.

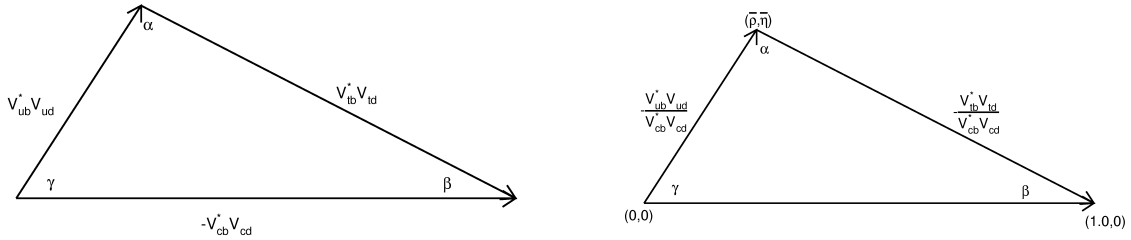

Figure 2: Illustration of *db unitarity* triangle (left), and rescaled *db unitarity* triangle (right) with $\bar{\rho}$ and $\bar{\eta}$ defined.

If one replaces all the d quarks with s quarks in Eq. 2.22, one obtains another "squashed" triangle, which is usually called the "sb" triangle. As shown in Figure 3, all the sides of "db" triangle are of order of λ^3, while for the "sb" triangle, two sides are of order of λ^2, the third one is of order of λ^4. This makes the "db" triangle much higher than the "sb" triangle, but one can easily prove, by using unitarity property of the CKM matrix, all the unitarity triangles have the same area, $J_{CP}/2$. One small angle arises from the "sb" triangle, defined as

$$\beta_s \equiv \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right). \tag{2.26}$$

Using the parametrization in Eq.2.18, β_s is calculated to be $\beta_s = \eta\lambda^2 \sim 0.018$. In principle, it can be measured from $B_s^0 \to J/\psi\phi$ decays, as we will see later.

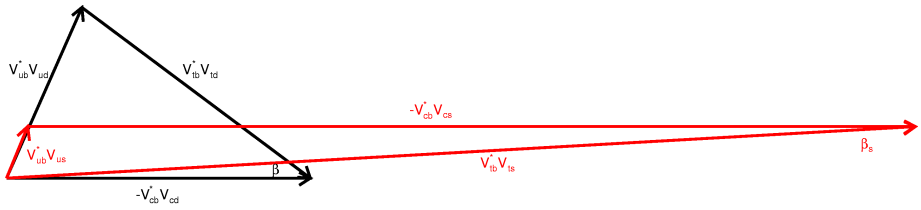

Figure 3: Comparison of *sb unitarity* triangle (red) and *db unitarity* triangle. One side of the *sb* triangle is about 4 times smaller than the side of *db* triangle, but the other two sides of *sb* triangle are 4 times bigger than the sides in *db* triangle.

2.3 MIXING IN NEUTRAL MESONS

Neutral mesons such as K^0, D^0, B_d^0 and B_s^0 which we will refer as X^0 all oscillate between the particle and its own antiparticle. The system can be described in its rest frame by a two component wave function

$$|\psi(t)\rangle = \psi_1(t)|X^0\rangle + \psi_2(t)|\bar{X}^0\rangle. \tag{2.27}$$

The evolution of the wave function can be calculated by a Schrödinger like equation

$$i\frac{d}{dt}\begin{pmatrix}\psi_1(t)\\ \psi_2(t)\end{pmatrix} = \begin{pmatrix}H_{11} & H_{12}\\ H_{21} & H_{22}\end{pmatrix}\begin{pmatrix}\psi_1(t)\\ \psi_2(t)\end{pmatrix}. \tag{2.28}$$

The matrix H is usually written as $H = M - i\Gamma/2$, where both M and Γ are 2×2 Hermitian matrices, called the mass and the decay matrices. The solution of this equation leads to two mass eigenstates, whose mass difference and decay width difference can be measured experimentally. The latest results according to the PDG [18] and HFAG [19] are shown in Table 3.

From Table 3, one notices the different scales of mixing parameters for different neutral mesons. The mass differences can be obtained from the box diagrams which are similar for all the mesons as shown in figure 4. The much smaller mass differences for B^0 and K^0 compared with B_s^0 are due to the CKM suppression. For the D^0 system, both the GIM suppression and

species	$\lvert\Delta m\rvert/\Gamma$	$\lvert\Delta\Gamma\rvert/\Gamma$	striking features
K^0	~ 0.47	~ 0.99	large decay width difference
D^0	$\sim 9.7 \times 10^{-3}$	$\sim 1.6 \times 10^{-2}$	none
B^0	~ 0.78	~ 0.01	large CP violation
B_s^0	~ 26.1	~ 0.15	large oscillation frequency

Table 3: Experimental results for mass and decay width differences of neutral mesons over average decay width, where CP conservation is assumed.

CKM suppression contribute. The K^0 meson has an especially large decay width difference, because the dominant kaon decay $K^0 \to \pi\pi$ has CP even final eigenstate which mostly K_s^0 can decay into. Similarly, the B_s^0 decay $B_s^0 \to D_s^+ D_s^-$ are Cabibbo favored and the final state is CP even, so only CP even eigenstate will decay like that (CP conservation assummed). While for $B^0 \to D^+ D^-$, although the final state is also CP even, it is Cabibbo suppressed, so CP even eigenstate does not have a larger decay width than the CP odd eigenstate. The same reasoning applies to the D^0 system.

In the following sections, we are going to focus on the general mixing property of neutral B systems, especially B_s^0 mesons. Their decay rates into certain final states will be derived and the calculation of the CP asymmetry will follow in specific decay channels, focusing on the $B_s^0 \to J/\psi\phi$ which is more complex than the $B^0 \to J/\psi K_s^0$ case.

States which we shall denote as $|B_q^0\rangle$ and $|\bar{B}_q^0\rangle$ ($q = d,\ s$) are eigenstates of the strong interaction but are mixed by the weak interaction. We use the following convention for CP transformations of meson states and quark currents [20]:

$$CP|B_q^0\rangle = -|\bar{B}_q^0\rangle, \quad CP\bar{q}_L\gamma_u b_L(CP)^{-1} = -\bar{b}_L\gamma_u q_L, \tag{2.29}$$

for quark $q = d$ or s. We can construct two CP eigenstates as follows

$$|B_q^{even}\rangle = \frac{1}{\sqrt{2}}(|B_q^0\rangle - |\bar{B}_q^0\rangle, \quad |B_q^{odd}\rangle = \frac{1}{\sqrt{2}}(|B_q^0\rangle + |\bar{B}_q^0\rangle. \tag{2.30}$$

The time development of an arbitrary system $a(t)|B^0\rangle + b(t)|\bar{B}^0\rangle$ is governed by Schrödinger's equation [21],

$$i\frac{d}{dt}\begin{pmatrix} a \\ b \end{pmatrix} = H\begin{pmatrix} a \\ b \end{pmatrix}, \tag{2.31}$$

where $H = M - \frac{i}{2}\Gamma$, both M and Γ being Hermitian 2×2 matrices, so $M_{12} = M_{21}^*$ and $\Gamma_{12} = \Gamma_{21}^*$. CPT invariance requires that $M_{11} = M_{22} \equiv M$ and $\Gamma_{11} = \Gamma_{22} \equiv \Gamma$. The off-diagonal elements of M and Γ are generated by the dispersive and absorptive parts of the box diagram as shown in Figure 4. In the standard model, short range interactions described

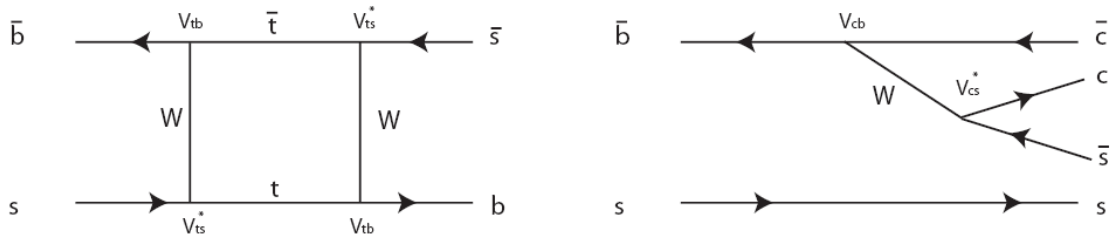

Figure 4: Feynman diagrams for B_s^0 - $\bar{B}_s^0$ mixing (left) and the decay to $B^0 \to J/\psi\phi$ (right). These two diagrams generate a small standard model CP violation due to the interference of mixing and decay.

by this box diagram dominates B mixing. For the dispersive part of the B^0-$\bar{B}^0$ oscillation amplitude M_{12}, the top quark contribution is dominant due to the large top quark mass and CKM matrix structure, and is calculated to be [22, 23]

$$M_{12} = \frac{G_F^2 M_W^2 \eta_{B_q^0} m_{B_q^0} B_{B_q^0} f_{B_q^0}^2}{12\pi^2} S_0(\frac{M_t^2}{M_W^2})(V_{tq}^* V_{tb})^2 \tag{2.32}$$

for quark $q = d$ or s. G_F is the Fermi coupling constant, M_W and $M_{B_q^0}$ are the W boson and B meson masses. The function $S_0(\frac{M_t^2}{M_W^2})$ is a known function calculated in the ratio of the top and W boson masses. $\eta_{B_q^0}$ is the QCD correction factor. The B meson decay constant $f_{B_q^0}$ and bag parameter $B_{B_q^0}$ are usually obtained through Lattice QCD calculations.

For the absorptive part of the oscillation amplitude, we need to consider the contribution for c or u quarks in the loop, where both B_q^0 and $\bar{B}_q^0$ decay to the same on-shell intermediate states. The off-diagonal element is calculated to be [22, 23]

$$\Gamma_{12} = -\frac{G_F^2 m_{B_q^0} m_b^2 \eta_{B_q^0} B_{B_q^0} f_{B_q^0}^2}{8\pi}\left[(V_{tq}^* V_{tb})^2 + \frac{8}{3} V_{tq}^* V_{tb} V_{cq}^* V_{tb} (\frac{m_c^2}{m_b^2})^2 + \mathcal{O}(\frac{m_c^2}{m_b^2})^4\right], \tag{2.33}$$

where $V_{tq}^* V_{tb}$ is used to replace $V_{cq}^* V_{cb}$ in order to facilitate the comparison of Γ_{12} and M_{12}, since $|V_{tq}^* V_{tb}| \approx |V_{cq}^* V_{cb}|$.

In the case of B_d^0 and B_s^0, one can compute

$$\left|\frac{\Gamma_{12}}{M_{12}}\right| \sim \frac{3\pi}{2}\frac{m_b^2}{M_W^2}\frac{1}{S_0(\frac{m_t^2}{M_W^2})} \sim 5 \times 10^{-3}, \tag{2.34}$$

so $|\Gamma_{12}|$ is negligible compared with $|M_{12}|$.

The mass eigenstates of the system are found by diagonalizing the Hamiltonian, such that

$$\begin{aligned} |B_L^0\rangle &= p|B_q^0\rangle + q|\bar{B}_q^0\rangle, \\ |B_H^0\rangle &= p|B_q^0\rangle - q|\bar{B}_q^0\rangle, \end{aligned} \tag{2.35}$$

where q and p govern the $B_q^0 - \bar{B}_q^0$ mixing and are complex numbers. They satisfy the normalization condition $|p|^2 + |q|^2 = 1$. The eigenvalues of the Hamiltonian are

$$\begin{aligned} E_{L,H} &\equiv M_{L,H} - \frac{i}{2}\Gamma_{L,H} \\ &= M - \frac{i}{2}\Gamma \mp \sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)} \\ &= (m \mp \frac{\Delta m}{2}) - \frac{i}{2}(\Gamma \pm \frac{\Delta\Gamma}{2}), \end{aligned} \tag{2.36}$$

where some quantities are expressed as $m = M = 1/2(M_L + M_H)$, $\Gamma = 1/2(\Gamma_L + \Gamma_H)$ and

$$\Delta m = M_H - M_L = 2\mathrm{Re}\sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)}, \tag{2.37}$$

$$\Delta\Gamma = \Gamma_L - \Gamma_H = 4\mathrm{Im}\sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)}. \tag{2.38}$$

In the case of $|\Gamma_{12}| \ll |M_{12}|$, Δm and $\Delta\Gamma$ are related to matrix elements as

$$\Delta m = 2|M_{12}|, \tag{2.39}$$

$$\Delta\Gamma = \frac{-2Re(M_{12}\Gamma_{12}^*)}{|M_{12}|} = 2|\Gamma_{12}|\cos\phi_s, \tag{2.40}$$

where $\phi_s = \arg(-M_{12}/\Gamma_{12})$.

With eigenvalues obtained, the relation of q and p is

$$\frac{q}{p} = -\sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}}, \tag{2.41}$$

and with the approximation $|\Gamma_{12}| \ll |M_{12}|$, one obtains

$$\frac{q}{p} = -\frac{M_{12}^*}{|M_{12}|} = -\frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}}. \tag{2.42}$$

The system is now completely characterized by the quantities m, Γ, Δm, $\Delta\Gamma$, q, and p. If q/p is real and equal to one, one can find that the light eigenstate is exactly a CP even eigenstate, and the heavy eigenstate is CP odd.

Solutions of the time dependent Schrödinger equation, Eq. 2.31, are linear superpositions of

$$\begin{aligned} |B_L^0(t)\rangle &= e^{-(im+\Gamma/2)t}e^{+(-\Delta\Gamma/4+i\Delta m/2)t}|B_L^0\rangle, \\ |B_H^0(t)\rangle &= e^{-(im+\Gamma/2)t}e^{-(-\Delta\Gamma/4+i\Delta m/2)t}|B_H^0\rangle, \end{aligned} \tag{2.43}$$

for the light and heavy eigenstates, respectively. In particular, by arranging Eq. 2.35, and substituting the mass eigenstates B_L^0 and B_H^0, the states produced as an initially pure $|B^0\rangle$ or $|\bar{B}^0\rangle$ can be written as

$$|B^0(t)\rangle = e^{-imt}e^{-\Gamma t/2}\left[E_+(t)|B^0\rangle + \frac{q}{p}E_-(t)|\bar{B}^0\rangle\right], \tag{2.44}$$

$$|\bar{B}^0(t)\rangle = e^{-imt}e^{-\Gamma t/2}\left[\frac{p}{q}E_-(t)|B^0\rangle + E_+(t)|\bar{B}^0\rangle\right], \tag{2.45}$$

where the functions $E_\pm(t)$ are defined as

$$E_\pm(t) = \frac{1}{2}\left[e^{+\left(\frac{-\Delta\Gamma}{4}+i\frac{\Delta m}{2}\right)t} \pm e^{-\left(\frac{-\Delta\Gamma}{4}+i\frac{\Delta m}{2}\right)t}\right], \tag{2.46}$$

For the B^0 system, where the width difference $\Delta\Gamma$ can be neglected, $E_+(t) = \cos(\Delta mt)$ and $E_-(t) = i\sin(\Delta mt)$, and the above reduce to simpler formulas. For the B_s^0 system, however, the more general expression is required because of sizable $\Delta\Gamma$.

Let $|f\rangle$ be a final state and the decay amplitude $\langle f|B^0\rangle = A_f$, $\langle f|\bar{B}^0\rangle = \bar{A}_f$. The time dependent transition amplitudes from an initially pure B^0 or $\bar{B}^0$ are

$$\begin{aligned}
\langle f|B^0(t)\rangle &= e^{-imt}e^{-\Gamma t/2}\left[E_+(t) + \frac{q}{p}\frac{\bar{A}_f}{A_f}E_-(t)\right]A_f, \\
\langle f|\bar{B}^0(t)\rangle &= e^{-imt}e^{-\Gamma t/2}\left[\frac{p}{q}\frac{A_f}{\bar{A}_f}E_-(t) + E_+(t)\right]\bar{A}_f.
\end{aligned} \tag{2.47}$$

Defining $\lambda_f = \frac{q\bar{A}_f}{pA_f}$, one can write

$$\begin{aligned}
\langle f|B^0(t)\rangle &= e^{-imt}e^{-\Gamma t/2}\left[E_+(t) + \lambda E_-(t)\right]A_f, \\
\langle f|\bar{B}^0(t)\rangle &= e^{-imt}e^{-\Gamma t/2}\left[E_+(t) + \lambda^{-1}E_-(t)\right]\bar{A}_f.
\end{aligned} \tag{2.48}$$

We can then define the time-dependent decay rates

$$\begin{aligned}
R &= |\langle f|B^0(t)\rangle|^2, \\
\bar{R} &= |\langle f|\bar{B}^0(t)\rangle|^2,
\end{aligned} \tag{2.49}$$

and note that

$$R = |A_f|^2\left[|E_+(t)|^2 + \left(\frac{q}{p}\frac{\bar{A}_f}{A_f}\right)^2|E_-(t)|^2 + 2Re\left\{\frac{q}{p}\frac{\bar{A}_f}{A_f}E_+^*(t)E_-(t)\right\}\right], \tag{2.50}$$

$$\bar{R} = |A_f|^2\left[\left|\frac{\bar{A}_f}{A_f}\right|^2|E_+(t)|^2 + \left(\frac{p}{q}\right)^2|E_-(t)|^2 + 2\left(\frac{p}{q}\right)^2 Re\left\{\left(\frac{q}{p}\frac{\bar{A}_f}{A_f}\right)^* E_+^*(t)E_-(t)\right\}\right], \tag{2.51}$$

or in terms of λ_f and more explicitly:

$$\begin{aligned}
R = \frac{|A_f|^2}{2}e^{-\Gamma t}\Bigg[&\left((1+|\lambda_f|^2)\cosh\left(\frac{\Delta\Gamma t}{2}\right) - 2Re\{\lambda_f\}\sinh\left(\frac{\Delta\Gamma t}{2}\right)\right) \\
&+ \left((1-|\lambda_f|^2)\cos(\Delta mt) - 2Im\{\lambda_f\}\sin(\Delta mt)\right)\Bigg],
\end{aligned} \tag{2.52}$$

$$\begin{aligned}
\bar{R} = \frac{|\bar{A}_f|^2}{2|\lambda_f|^2}e^{-\Gamma t}\Bigg[&\left((1+|\lambda_f|^2)\cosh\left(\frac{\Delta\Gamma t}{2}\right) - 2Re\{\lambda_f\}\sinh\left(\frac{\Delta\Gamma t}{2}\right)\right) \\
&- \left((1-|\lambda_f|^2)\cos(\Delta mt) + 2Im\{\lambda_f\}\sin(\Delta mt)\right)\Bigg].
\end{aligned} \tag{2.53}$$

2.4 CP ASYMMETRIES IN NEUTRAL B MESONS

The CP asymmetries arise when the decay rates are different for B^0 and $\bar{B}^0$, and the time dependent asymmetry is defined as

$$A_{CP}(t) = \left(\bar{R} - R\right) / \left(\bar{R} + R\right). \tag{2.54}$$

From Eqs. 2.50 and 2.51, one realizes that several scenarios can lead to a CP asymmetry:

- if $|A_f| \neq |\bar{A}_f|$, this is called CP *violation in the decay amplitudes*,
- if $|q/p| \neq 1$, even without CP violation in the decay amplitudes, the admixture of of B_q^0 and $\bar{B}_q^0$ are not equal in the mass eigenstates. This type of CP violation is called CP *violation in the mixing*, which results in two physical eigenstates which are different from CP eigenstates,
- CP violation can still arise even when there is no CP violation in the mixing or decay amplitudes. From the third term of Eq. 2.50 and 2.51, it is clear that when $\lambda_f = \frac{q}{p}\frac{\bar{A}_f}{A_f}$ is complex, namely when $Im(\lambda_f) \neq 0$, the decay rates of B_q^0 and $\bar{B}_q^0$ are different. This type of CP violation is usually referred to as CP *violation in the interference between the mixing and decay amplitudes.*

For B_q^0 ($q = d,\ s$) systems, the decay amplitude for B^0 or $\bar{B}^0$ to decay into final state $|f\rangle$ is [24]:

$$A_f, \bar{A}_f \propto \left[1 + \left(\frac{\lambda^2}{1-\lambda^2}\right) a_p e^{i\theta} e^{\pm i\gamma}\right]. \tag{2.55}$$

Here the weak phase factor $e^{i\gamma}$ which is associated with the "penguin parameter" $a_p e^{i\theta}$, is strongly Cabibbo suppressed by two powers of the Wolfenstein parameter $\lambda \sim 0.23$. So we account only for the leading order of λ and set $|A_f| = |\bar{A}_f|$, which means we neglect CP *violation in the decay amplitudes* (direct CP violation). According to Eq. 2.34, one can also set $|q/p| = 1$. So we can treat $|\lambda_f| = 1$ in Eqs. 2.52 and 2.53, and the equations become:

$$R = |A_f|^2 e^{-\Gamma t}\left[\cosh\left(\frac{\Delta\Gamma t}{2}\right) - Re\{\lambda_f\}\sinh\left(\frac{\Delta\Gamma t}{2}\right) - Im\{\lambda_f\}\sin(\Delta m t)\right], \tag{2.56}$$

$$\bar{R} = |A_f|^2 e^{-\Gamma t}\left[\cosh\left(\frac{\Delta\Gamma t}{2}\right) - Re\{\lambda_f\}\sinh\left(\frac{\Delta\Gamma t}{2}\right) + Im\{\lambda_f\}\sin(\Delta m t)\right]. \tag{2.57}$$

For B_d^0 systems, where $\Delta\Gamma \approx 0$, the decay rates reduce to:

$$R = |A_f|^2 e^{-\Gamma t} \left[1 - Im\{\lambda_f\} \sin(\Delta m t)\right], \tag{2.58}$$

$$\bar{R} = |A_f|^2 e^{-\Gamma t} \left[1 + Im\{\lambda_f\} \sin(\Delta m t)\right], \tag{2.59}$$

giving the familiar expression for the CP asymmetry of $B^0 \to J/\psi K_s^0$ decays, which is a CP eigenstate

$$A_{CP}(t) \longrightarrow Im(\lambda_f) \sin \Delta m t. \tag{2.60}$$

The overall rate of B_q^0 or $\bar{B}_q^0$ decays gives the very familiar expression

$$\bar{R} + R \longrightarrow e^{-t/\tau}. \tag{2.61}$$

2.5 PHENOMENOLOGY OF $B_S^0 \to J/\psi\phi$ TO A SINGLE CP EIGENSTATE

Let us assume the final state of $B_s^0 \to J/\psi\phi$ is a single CP eigenstate, with CP quantum number η_{CP}. The diagram for this process is shown in Figure 4. The ratio of decay amplitudes is related to the CKM matrix elements by

$$\frac{\bar{A}_f}{A_f} = -\eta_{CP} \frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*} = \mp e^{i2\phi_{c\bar{c}s}}, \tag{2.62}$$

where $\eta_{CP} = \pm 1$ are for CP even and CP odd final states, respectively, and the phase $\phi_{c\bar{c}s} = arg(V_{cs}^* V_{cb})$. Combining this with q/p from Eq. 2.42 gives

$$\lambda_f = \pm \frac{V_{ts} V_{tb}^*}{V_{ts}^* V_{tb}} \frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*} = \pm e^{i2(\beta_s - \pi)} = \pm e^{i2\beta_s}, \tag{2.63}$$

where $\beta_s = arg(-\frac{V_{ts} V_{tb}^*}{V_{cs} V_{cb}^*})$. The real and imaginary parts of this are $\pm \cos 2\beta_s$ and $\pm \sin 2\beta_s$. Making those substitutions in Eqs. 2.56 and 2.57 gives

$$\begin{aligned} R_+ &= |A_f|^2 e^{-\Gamma t} \left[\cosh\left(\frac{\Delta\Gamma t}{2}\right) - \cos 2\beta_s \sinh\left(\frac{\Delta\Gamma t}{2}\right) - \sin 2\beta_s \sin(\Delta m t)\right], \\ \bar{R}_+ &= |A_f|^2 e^{-\Gamma t} \left[\cosh\left(\frac{\Delta\Gamma t}{2}\right) - \cos 2\beta_s \sinh\left(\frac{\Delta\Gamma t}{2}\right) + \sin 2\beta_s \sin(\Delta m t)\right], \end{aligned} \tag{2.64}$$

for a CP even final eigenstate, and

$$R_- = |A_f|^2 e^{-\Gamma t}\left[\cosh\left(\frac{\Delta\Gamma t}{2}\right) + \cos 2\beta_s \sinh\left(\frac{\Delta\Gamma t}{2}\right) + \sin 2\beta_s \sin(\Delta m t)\right],$$
$$\bar{R}_- = |A_f|^2 e^{-\Gamma t}\left[\cosh\left(\frac{\Delta\Gamma t}{2}\right) + \cos 2\beta_s \sinh\left(\frac{\Delta\Gamma t}{2}\right) - \sin 2\beta_s \sin(\Delta m t)\right], \quad (2.65)$$

for a CP odd final eigenstate.

In the case of $B_d^0 \to J/\psi K_s^0$ which has a CP odd final eigenstate, with $\Delta\Gamma = 0$ and $\mathrm{Im}\lambda_f = \sin(2\beta)$, one can obtain

$$R = |A_f|^2 e^{-t/\tau}\left[1 - \sin 2\beta \sin(\Delta m t)\right],$$
$$\bar{R} = |A_f|^2 e^{-t/\tau}\left[1 + \sin 2\beta \sin(\Delta m t)\right], \quad (2.66)$$

so the time dependent asymmetry is

$$A_{CP}(t) = \sin 2\beta \sin(\Delta m t). \quad (2.67)$$

2.6 PHENOMENOLOGY OF $B_S^0 \to J/\psi\phi$ TO AN ADMIXTURE OF CP EIGENSTATES

2.6.1 CP Eigenstates Separation in Transversity Basis

Both J/ψ and ϕ are vector mesons with properties $J^{PC} = 1^{--}$. Since both of them are C odd eigenstates, the total C is even and the total CP only depends on the total P. To get the total $P = P(J/\psi) * P(\phi) * (-1)^L$, one needs to know the angular momentum L. Because B_s^0 is a spin 0 particle, to conserve the total spin, the angular momentum could be $L = 0, 1, 2$. It Is clear that for S ($L = 0$) and D ($L = 2$) partial waves, the final states is CP even, while for P ($L = 1$) partial wave, the final state is CP odd.

Alternatively, one can decompose the decay amplitudes into three independent components, corresponding to linear polarization states of the vector mesons which are either

longitudinal (A_0), or transverse to their directions of motion and parallel ($A_\parallel$) or perpendicular ($A_\perp$) to each other [25]. The states 0 and $\parallel$ are CP-even, while the state $\perp$ is CP-odd. We let the $A's$ be normalized such that $|A_0(0)|^2 + |A_\parallel(0)|^2 + |A_\perp(0)|^2 = 1$.

The polarization now determines the angular distribution of the decay products, which are characterized by three *transversity angles* [26], see Figure 5. In the J/ψ rest frame, the x-axis is chosen to lie along the momentum of the ϕ and the z-axis is perpendicular to the decay plane of the ϕ. The coordinates (θ, ϕ) describe the decay direction of the μ^+. In the ϕ rest frame, the angle ψ is chosen to be the angle between the K^+ direction and the x-axis.

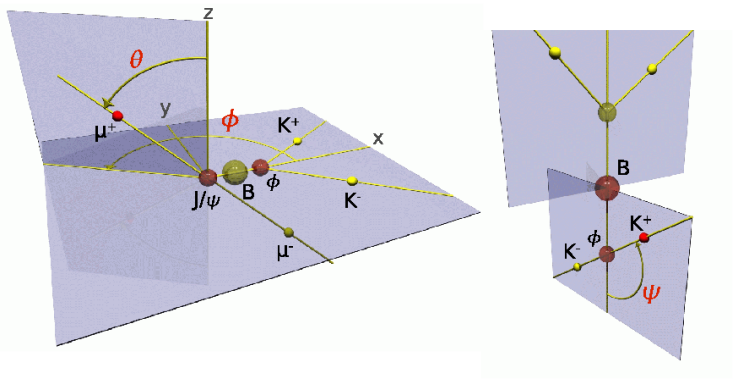

Figure 5: The J/ψ rest frame (left), where the direction of ϕ defines x axis and the plane of K^+K^- defines y axis with $p_y(K^+) > 0$. The ϕ rest frame (right), where ψ is the angle between K^+ and the negative of the direction of J/ψ in that rest frame.

The predicted angular distributions can be found from the following prescriptions [25][26]. Let $\hat{n}$ be the unit vector in the direction of the l^+ (J/ψ rest frame),

$$\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{2.68}$$

and let A be a complex vector defined as

$$\mathbf{A} = (A_0\cos\psi, -\frac{A_\parallel\sin\psi}{\sqrt{2}}, i\frac{A_\perp\sin\psi}{\sqrt{2}}). \tag{2.69}$$

The angular distributions are governed by the probability density

$$P(\theta,\phi,\psi) = \frac{9}{16\pi}|\mathbf{A}\times\hat{n}|^2, \tag{2.70}$$

where the coefficient comes from geometric normalization via integration over all the angles. The amplitudes themselves evolve with a time-dependence whose explicit form is that of Eq. 2.48 for B_s^0 and $\bar{B}_s^0$, with $\lambda_f = \pm e^{2i\beta_s}$ for CP even and odd eigenstates,

$$\begin{aligned} A_i &\equiv \langle f_i | B_s^0(t) \rangle = e^{-imt} e^{-\Gamma t/2} \left[E_+(t) \pm e^{2i\beta_s} E_-(t) \right] A_i(0), \\ \bar{A}_i &\equiv \langle f_i | \bar{B}_s^0(t) \rangle = e^{-imt} e^{-\Gamma t/2} \left[E_+(t) \pm e^{-2i\beta_s} E_-(t) \right] \bar{A}_i(0), \end{aligned} \tag{2.71}$$

where $i = 0, \|$ correspond to CP even (+), $i = \perp$ corresponds to CP odd (-). The vector $\mathbf{A}$ in Eq. 2.69 now has to be replaced by $\mathbf{A}(t)$. With Eq. 2.62, one can substitute $\bar{A}_i(0)$ with $\pm e^{i2\phi_{c\bar{c}s}} A_i(0)$ and get:

$$\begin{aligned} A_i &= e^{-imt} e^{-\Gamma t/2} \left[E_+(t) \pm e^{2i\beta_s} E_-(t) \right] A_i(0), \\ \bar{A}_i &= e^{-imt} e^{-\Gamma t/2} \left[\pm E_+(t) + e^{-2i\beta_s} E_-(t) \right] e^{i2\phi_{c\bar{c}s}} A_i(0). \end{aligned} \tag{2.72}$$

With definitions of $f_\pm(t)$ and $\bar{f}_\pm(t)$,

$$\begin{aligned} f_\pm(t) &= e^{-imt} e^{-\Gamma t/2} \left[E_+(t) \pm e^{2i\beta_s} E_-(t) \right], \\ \bar{f}_\pm(t) &= e^{-imt} e^{-\Gamma t/2} \left[\pm E_+(t) + e^{-2i\beta_s} E_-(t) \right] e^{i2\phi_{c\bar{c}s}}, \end{aligned} \tag{2.73}$$

one can write down A_i and $\bar{A}_i$ in this form:

$$\begin{aligned} A_i &= f_\pm(t) A_i(0), \\ \bar{A}_i &= \bar{f}_\pm(t) A_i(0), \end{aligned} \tag{2.74}$$

where "+" is for CP even, "-" is for CP odd.

2.6.2 Time Development

Eqs. 2.70 and 2.74, together with the definitions 2.46, are all we need to implement a model for the differential rates of $B_s^0 \to J/\psi\phi$ parametrized by β_s, Γ, $\Delta\Gamma$, and the magnitudes and phases of the $A's$. No further expansion of the probability densities is needed, since complex vectors and complex functions are all implemented in standard `HEP` libraries or standard runtime libraries. However, to extract these parameters from data requires an unbinned maximum likelihood fit. Any reasonable likelihood fit will include resolution effects, which can be computed analytically, both for exponential functions and exponential sine or cosine functions. Eq 2.70 contains, all mixed up together, several components for the time-development of the rate. We can not "smear" amplitudes only probability densities, so we separate these so that they can each be convolved analytically with a resolution function. One simplification is that the time development of $A_0(t)$ and $A_\parallel(t)$ amplitudes is identical, although it differs from that of $A_\perp(t)$. We begin by decomposing

$$\mathbf{A}(t) = \mathbf{A}_+(t) + \mathbf{A}_-(t), \tag{2.75}$$

where

$$\mathbf{A}_+(t) = \mathbf{A}_+ f_+(t) = (A_0 \cos\psi, -\frac{A_\parallel \sin\psi}{\sqrt{2}}, 0) \cdot f_+(t), \tag{2.76}$$

and

$$\mathbf{A}_-(t) = \mathbf{A}_- f_-(t) = (0, 0, i\frac{A_\perp \sin\psi}{\sqrt{2}}) \cdot f_-(t). \tag{2.77}$$

We then have in place of Eq. 2.70 for B_s^0 and $\bar{B}_s^0$

$$\begin{aligned} P(\theta,\psi,\phi,t) &= \left\{|\mathbf{A}_+(\mathbf{t}) \times \hat{n}|^2 + |\mathbf{A}_-(\mathbf{t}) \times \hat{n}|^2 + 2Re((\mathbf{A}_+(\mathbf{t}) \times \hat{n}) \cdot (\mathbf{A}_-^*(\mathbf{t}) \times \hat{n}))\right\} \\ &= \left\{|\mathbf{A}_+ \times \hat{n}|^2 |f_+(t)|^2 + |\mathbf{A}_- \times \hat{n}|^2 |f_-(t)|^2 + 2Re((\mathbf{A}_+ \times \hat{n}) \cdot (\mathbf{A}_-^* \times \hat{n}) \cdot f_+(t) \cdot f_-^*(t))\right\}, \end{aligned} \tag{2.78}$$

and

$$\begin{aligned} \bar{P}(\theta,\psi,\phi,t) &= \left\{|\bar{\mathbf{A}}_+(t) \times \hat{n}|^2 + |\bar{\mathbf{A}}_-(t) \times \hat{n}|^2 + 2Re(\bar{\mathbf{A}}_+(t) \times \hat{n}) \cdot (\bar{\mathbf{A}}_-^*(t) \times \hat{n}))\right\} \\ &= \left\{|\mathbf{A}_+ \times \hat{n}|^2 |\bar{f}_+(t)|^2 + |\mathbf{A}_- \times \hat{n}|^2 |\bar{f}_-(t)|^2 + 2Re((\mathbf{A}_+ \times \hat{n}) \cdot (\mathbf{A}_-^* \times \hat{n}) \cdot \bar{f}_+(t) \cdot \bar{f}_-^*(t)\right\}. \end{aligned} \tag{2.79}$$

Notice that the above expressions are not normalized at this time, and the calculation of a new normalization term N will come later.

Computing $|f_+(t)|^2$ and $|f_-(t)^2|$ is almost same as what has already been done in the context of the decay to a single CP eigenstate (Eq. 2.64) and (Eq. 2.65), which are:

$$\begin{aligned}
|f_\pm(t)|^2 &= e^{-\Gamma t}\left\{\cosh\frac{\Delta\Gamma}{2}t \mp \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t \mp \sin 2\beta_s \sin\Delta m t\right\}, \\
|\bar{f}_\pm(t)|^2 &= e^{-\Gamma t}\left\{\cosh\frac{\Delta\Gamma}{2}t \mp \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t \pm \sin 2\beta_s \sin\Delta m t\right\},
\end{aligned} \tag{2.80}$$

or in terms of Γ_L and Γ_H:

$$\begin{aligned}
|f_\pm(t)|^2 &= \frac{1}{2}\{(1 \pm \cos 2\beta_s)e^{-\Gamma_L t} + (1 \mp \cos 2\beta_s)e^{-\Gamma_H t} \mp 2\sin 2\beta_s e^{-\Gamma t}\sin\Delta m t\}, \\
|\bar{f}_\pm(t)|^2 &= \frac{1}{2}\{(1 \pm \cos 2\beta_s)e^{-\Gamma_L t} + (1 \mp \cos 2\beta_s)e^{-\Gamma_H t} \pm 2\sin 2\beta_s e^{-\Gamma t}\sin\Delta m t\}.
\end{aligned} \tag{2.81}$$

Next, we need to do the same thing for the interference term, $f_+(t)f_-^*(t)$. For B_s^0 and $\bar{B}_s^0$, those terms are

$$\begin{aligned}
f_+(t)f_-^*(t) &= e^{-\Gamma t}\cos\Delta m t + i\cos 2\beta_s e^{-\Gamma t}\sin\Delta m t + i\sin 2\beta_s(e^{-\Gamma_L t} - e^{-\Gamma_H t})/2, \\
\bar{f}_+(t)\bar{f}_-^{\,*}(t) &= -e^{-\Gamma t}\cos\Delta m t - i\cos 2\beta_s e^{-\Gamma t}\sin\Delta m t + i\sin 2\beta_s(e^{-\Gamma_L t} - e^{-\Gamma_H t})/2.
\end{aligned} \tag{2.82}$$

This accomplishes the desired separation. Eq. 2.78 and Eq. 2.79 are further expanded in detail in APPENDIX A.

2.6.3 Normalization Calculation

To normalize the probability function, an overall normalization factor needs to be calculated in this way:

$$N \equiv \int\int\int\int \sum_{i=B,\bar{B}} P_i(\psi,\theta,\phi,t)d(\cos\psi)d(\cos\theta)d\phi dt. \tag{2.83}$$

We first carry out the summation and integrate over t, giving rise to the following factors

$$\begin{aligned}
\int |f_+(t)|^2 + |\bar{f}_+(t)|^2 dt &= \tau_L + \tau_H + \cos 2\beta_s(\tau_L - \tau_H), \\
\int |f_-(t)|^2 + |\bar{f}_-(t)|^2 dt &= \tau_L + \tau_H - \cos 2\beta_s(\tau_L - \tau_H), \\
\int f_+(t)f_-^*(t) + \bar{f}_+(t)\bar{f}_-^*(t) dt &= i \sin 2\beta_s(\tau_L - \tau_H),
\end{aligned}$$

so that, using Eqs. 2.78 and 2.79, we can write N as

$$\begin{aligned}
N &= \int\int\int d(\cos\psi)d(\cos\theta)d\phi \\
&\quad \big[|\mathbf{A}_+ \times \hat{n}|^2 \times (\tau_L + \tau_H + \cos 2\beta_s(\tau_L - \tau_H)) \\
&\quad + |\mathbf{A}_- \times \hat{n}|^2 \times (\tau_L + \tau_H - \cos 2\beta_s(\tau_L - \tau_H)) \\
&\quad + 2Re((\mathbf{A}_+ \times \hat{n}) \cdot (\mathbf{A}_-^* \times \hat{n}) \times (i \sin 2\beta_s(\tau_L - \tau_H)))\big] \\
&= \frac{16\pi}{9}[(|A_0|^2 + |A_\parallel|^2)(\tau_L + \tau_H + \cos 2\beta_s(\tau_L - \tau_H)) + \\
&\quad |A_\perp|^2(\tau_L + \tau_H - \cos 2\beta_s(\tau_L - \tau_H))].
\end{aligned} \tag{2.84}$$

Now the normalized probability density functions for B_s^0 and $\bar{B}_s^0$ look like:

$$\begin{aligned}
P'(\psi,\theta,\phi,t) &= \frac{1}{N} P(\psi,\theta,\phi,t), \\
\bar{P}'(\psi,\theta,\phi,t) &= \frac{1}{N} \bar{P}(\psi,\theta,\phi,t).
\end{aligned} \tag{2.85}$$

2.7 GENERAL ANALYSIS STRATEGY

Our goal is to measure the CP violation phase β_s. As discussed in the previous sections, the phase arises in the $B_s^0 \to J/\psi\phi$ channel through

$$2\beta_s = \arg(\eta_{CP}\frac{q}{p}\frac{\bar{A}_f}{A_f}).$$

In the standard model, this angle is almost zero, meaning CP conservation. If there is any new physics associated with the B_s^0 mixing box diagram, the relative phase of the mass eigenstate coefficients q/p will be changed, thus large CP violation can occur through the interference of mixing and decay. With large CP violation, the mass eigenstates of B_s^0 are not CP eigenstates anymore. This can be seen clearly through the following equations (where the phase of $\bar{A}_f/A_f$ is neglected)

$$\begin{aligned}
|B_L^0\rangle &= p|B_s^0\rangle + q|\bar{B}_s^0\rangle \\
&= \frac{e^{-i\beta_s}}{\sqrt{2}}|B_s^0\rangle - \frac{e^{+i\beta_s}}{\sqrt{2}}|\bar{B}_s^0\rangle \\
&= \cos\beta_s(\frac{1}{\sqrt{2}}|B_s^0\rangle - \frac{1}{\sqrt{2}}|\bar{B}_s^0\rangle) - i\sin\beta_s(\frac{1}{\sqrt{2}}|B_s^0\rangle + \frac{1}{\sqrt{2}}|\bar{B}_s^0\rangle) \\
&= \cos\beta_s|B_s^{even}\rangle - i\sin\beta_s|B_s^{odd}\rangle, \qquad (2.86) \\
|B_H^0\rangle &= p|B_s^0\rangle - q|\bar{B}_s^0\rangle, \\
&= \cos\beta_s|B_s^{odd}\rangle - i\sin\beta_s|B_s^{even}\rangle. \qquad (2.87)
\end{aligned}$$

To measure the phase β_s, we will start from decay rate functions (Eq. 2.78 and Eq. 2.79) for B_s^0 and $\bar{B}_s^0$. There are four physical parameters associated with the decay rate functions, including the decay width Γ, the decay width difference $\Delta\Gamma$, the CP violation phase β_s and the oscillation frequency Δm_s. The observation of $B_s^0 - \bar{B}_s^0$ oscillation at CDF in 2006 [28] makes this analysis possible with Δm_s as an input value. $B_s^0 \to J/\psi\phi$ channel will be reconstructed and analyzed. To identify the flavor of B_s^0 meson at production time, flavor tagging algorithms are developed, and the $B^+ \to J/\psi K^+$ channel will be reconstructed to calibrate one of the tagging algorithms. As a crosscheck, the CP violation phase β is also measured from the reconstructed $B^0 \to J/\psi K_s^0$ decays.

3.0 THE TEVATRON COLLIDER AND CDF EXPERIMENT

The Collider Detector at Fermilab (CDF) experiment is one of the two main experiments located at the Tevatron collider of Fermilab in Batavia, IL. The Tevatron is a circular proton-antiproton synchrotron 1 km in radius. The ring collides counter-rotating bunches of protons and antiprotons at a center-of-mass energy of 1,960 GeV in two interaction regions along the ring: B0, which is the site of the CDF experiment, and D0, where the D0 experiment is located. Each proton bunch contains approximately two hundred billion protons, while each antiproton bunch contains approximately twenty billion antiprotons, with collisions occuring at a rate of about 2.5 million Hz. In this chapter, the production and acceleration of particle and anti-particle beams will be described first (a more detailed description can be found in Refs. [29, 30]). A description of the CDF experiment, with a focus on the parts of the detector which are relevant for B physics, will follow.

3.1 THE TEVATRON COLLIDER

The Tevatron at Fermilab is the last and highest energy stage of the large accelerator complex. This accelerator complex was first established in 1969. The physics program began in 1972 with proton beam energies of 200 GeV for fixed target experiments. The first $p\bar{p}$ collisions started in 1986. Since then, several extensive upgrades have been undertaken leading to a major improvement of the overall performance. While operating in a collider mode, the Tevatron collides 36 bunches of protons and antiprotons every 396 ns with each beam at energy of 980 GeV, currently the highest in the world. Radio-frequency (RF) buckets are used to accelerate the particles and define the bunches. Once the beams are injected

and accelerated, collisions are made to occur at two points along the ring, inside the CDF and D0 experiments. The schematic plot of the Fermilab Accelerator complex is shown in Figure 6.

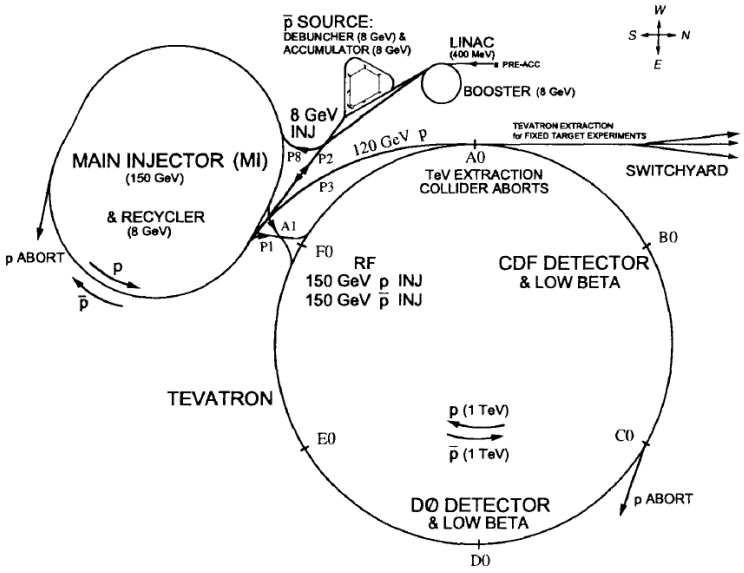

Figure 6: Schematic plot of the accelerator complex.

3.1.1 Proton and Antiproton Production

The first stage in the acceleration of protons is the commercial Cockcroft-Walton accelerator, which will boost H^- ions produced by ionization of gaseous hydrogen to 750 KeV. The ions are then injected into a 150 m long linear accelerator ("Linac") which increases their energy to 400 MeV. The Linac is made up of two types of RF stations, which are electromagnetically resonant cavities with natural resonant frequencies lying within the radio frequencies of the electromagnetic spectrum. A carbon foil is used to strip the two electrons from H^- before the resulting protons are injected into the "Booster". The Booster is a circular synchrotron with 18 RF cavities distributed about a ring with a 75 m radius. The 201 MHz frequencies of the bunches from Linac do not match the 37.8 MHz frequencies of the RF cavities in the Booster. After all the beam has been injected, the protons eventually come into phase with the cavities, and a new 37.8 MHz bunch structure is formed and accelerated to 8 GeV.

The protons are then transfered to the "Main Injector" which brings their energies to 150 GeV. The Main Injector is a circular synchrotron with 18 accelerating RF cavities and a circumference of almost 2 miles (completed in 1999), approximately half the circumference of the Tevatron. The final step of the process is the transfer to the Tevatron, a synchrotron which employs superconducting Nb-Ti alloy filaments embedded in copper as magnet coils. The magnetic field of 5.7 T keeps the protons on an approximately circular orbit while they reach the final energy of 980 GeV.

Once the protons bunches circulating in the Main Injector reach an energy of 120 GeV, some are slammed into a rotating 7 cm thick nickel or copper target to produce antiprotons through the reaction

$$p + \begin{pmatrix} n \\ p \end{pmatrix} \rightarrow p + p + \begin{pmatrix} n \\ p \end{pmatrix} + \bar{p}. \tag{3.1}$$

Before colliding, the protons bunches are rotated by 90° in phase space, so that they have a large spread in energy but a small lag in arrival time at the target. A spatially broad shower of particles is produced and then focused into a beam via cylindrical lithium lens. This beam, which has a bunch structure similar to the the incident proton beam, is passed through a pulsed dipole magnet. The magnetic field separates the negatively charge antiprotons with about 8 GeV of kinetic energy. About 20 antiprotons are produced for every 10^6 protons on target and then stored into the "Debuncher". The Debuncher is a triangular-shaped synchrotron with mean radius of 90 m. The beam is stochastically cooled by picking up signal from antiproton circulating on one side of the ring, amplifying the signal, and sending it to the opposite side of the ring. After cooling, the antiprotons are then transferred to the "Accumulator", which is another triangular-shaped synchrotron with a mean radius of 75 m. The Accumulator is a storage ring for the antiprotons; there they are stored at 8 GeV and cooled until needed for acceleration in the Main Injector, where they are accelerated to 150 GeV. Finally, the antiprotons are also transferred to the Tevatron, where 36 previously injected bunches of protons are already circulating in opposite direction. Since 2004, an additional Recycler Ring has been added in the same tunnel as the Main Injector and provides additional storage of antiprotons. By limiting the stack size, the Debuncher allows an optimization of antiprotons accumulation rate, and this rate is the largest limiting

factor in Tevatron running.

3.1.2 The Collision and Luminosity

In order to create collisions, 36 bunches of protons are injected into Tevatron first. Twelve bunches each separated by 21 RF (396 ns) are grouped together into three trains of bunches. The trains are have a larger separation of 140 RF buckets and the gaps provide enough space to insert the next 36 bunches of antiprotons without disturbing the protons. The bunch structure is illustrated in Figure 7

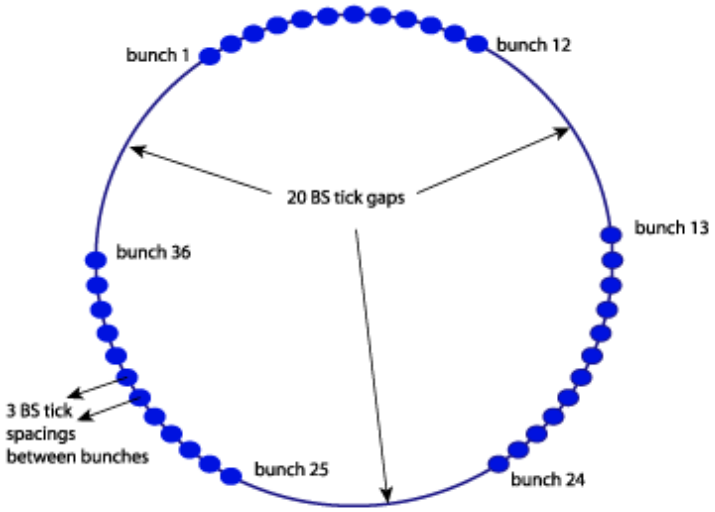

Figure 7: The Tevatron timing is based on 53 MHz rate, so the beam in RF buckets is spaced by about 19 ns. The beam is split into 3 trains of 12 bunches each. Each bunch is separated by the next by 21 buckets or 396 ns, or 3 BS (beam sync) ticks.

The antiproton bunch pattern is a mirror image of the proton pattern and circulates along the Tevatron in the opposite direction within the same magnet and vacuum systems. The energy of the machine is increased in about 10 seconds from 150 to 980 GeV.

Special quadrupole magnets (low-β squeezers) located at CDF and D0 experiments along the beam pipe squeeze the beam in order to maximize luminosity inside the detectors. A roughly Gaussian distribution of the interaction region along the beam axis is achieved

($\sigma_z \approx 30$ cm). The transverse shape of the interaction region has an almost circular spatial distribution with a diameter of 30 μm.

The Tevatron performance can be evaluated in terms of two parameters, the center-of-mass energy, $\sqrt{s}$, and the instantaneous luminosity, $\mathcal{L}$. $\sqrt{s}$ defines the accessible phase space for the production of particles in the final states and $\mathcal{L}$ relates the production rate of a given process and its cross section σ in this way: $rate\ [events\ s^{-1}] = \mathcal{L} \times \sigma$. With ideal head-on $p\bar{p}$ collision, the instantaneous luminosity is given by:

$$\mathcal{L} = \frac{fBN_pN_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} F(\sigma_l/\beta^*), \tag{3.2}$$

where f is the revolution frequency, B is the number of bunches in each beam, N_p ($N_{\bar{p}}$) is the number of protons (antiprotons) in each bunch, σ_p ($\sigma_{\bar{p}}$) is the rms proton (antiproton) beam size at the interaction point, and F is a form factor which depends on the ratio of longitudinal rms length σ_l to the beta function at the interaction point, β^*.

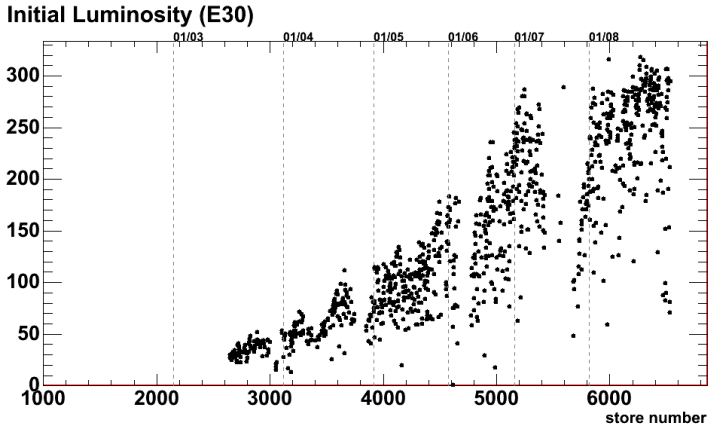

Figure 8: Initial Luminosity recorded at CDF as a function of time up to October 2008, which shows the improvement of Tevatron over the past years.

Due to beam-beam interactions and collisions, the instantaneous luminosity decreases exponentially over time. The beam is usually dumped intentionally after 15-20 hours of recording collisions and replaced with a new store.

3.2 THE CDF EXPERIMENT

The CDF detector to which we refer throughout this thesis is in fact the "CDF-II" detector which was upgraded in many ways over the CDF detector in Run I during the 1992-1995 data taking period. A full description of the detector can be found in the CDF technical design report [31]. The start date of the CDF-II experiment was June 2001. In this section a brief overview of the detector will be given followed by a description of the subsystems relevant to B physics.

3.2.1 Detector Overview and Coordinate System

The CDF detector is a large multi-purpose detector designed for reconstruction of many types of $p\bar{p}$ collisions. As shown in the cut-away view of the CDF detector in Figure 9, the detector consists of various tracking systems immersed in a solenoidal magnetic field, and positioned within the electromagnetic (EM) and hadronic calorimeters, and a muon detection systems which includes several drift chambers and steel shielding. A simple view of the passage of particles through the detector is shown in Figure 10. The momentum and impact parameter of charged particles can be measured by the tracking systems, which are not sensitive to the neutral particles. The EM calorimeter can measure the energy of photons as well as electrons. Hadronic energy is measured in the hadronic calorimeter. The muons which are minimum ionizing will be detected by the outer muon drift chambers. Neutrinos are not detected but can be inferred indirectly through missing transverse energy and momentum.

A natural coordinate system for the detector is a cylindrical system since the detector design is azimuthally symmetric. The z-axis is defined to be parallel to beamline along the proton direction through the detector, while x-axis is defined as the horizontal pointing away from the Tevatron ring, as shown in Figure 11. The polar angle θ is defined from positive z-axis, while the azimuthal angle ϕ is defined from the positive x-axis. The direction parallel to the z-axis is usually called "longitudinal" and the direction perpendicular to z-axis is called "transverse", i.e., in the $(x, y) \equiv (r, \phi)$ plane. In hadron collisions, a particularly

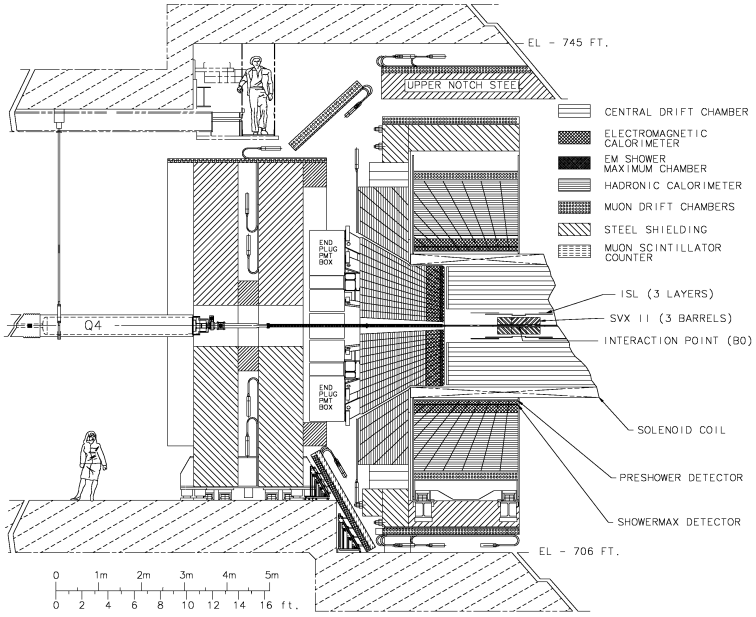

Figure 9: The CDF RunII detector.

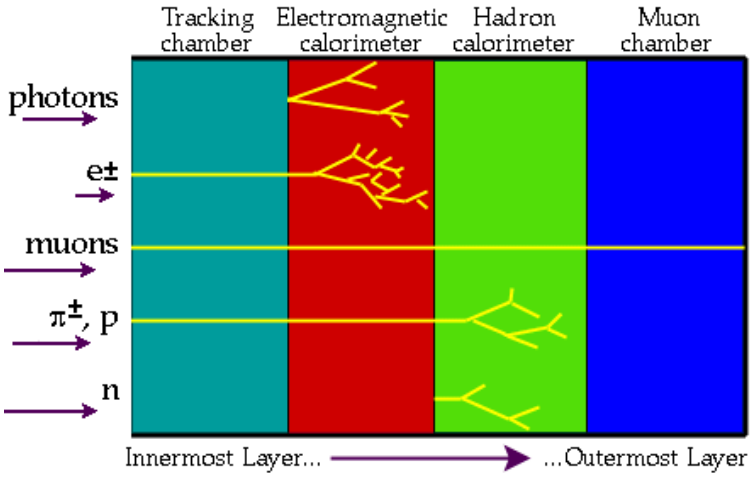

Figure 10: The passage of particles through the CDF detector.

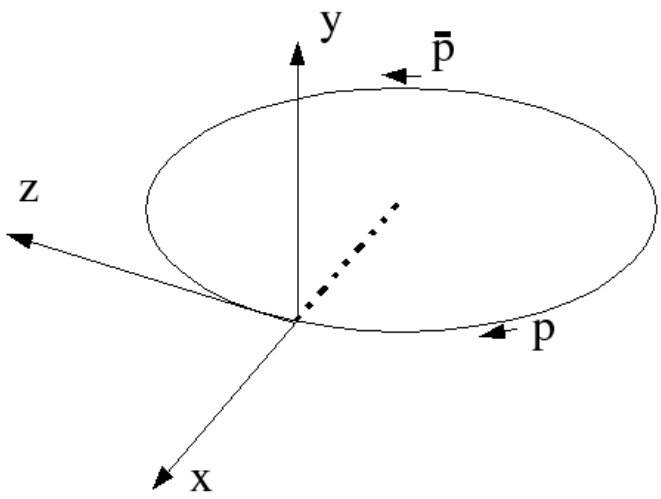

Figure 11: The coordinate system applied at CDF detector.

useful kinematic variable, rapidity, is used

$$Y = \frac{1}{2} ln \left[\frac{E + p\cos\theta}{E - p\cos\theta} \right], \tag{3.3}$$

where $(E, \vec{p})$ is the energy-momentum four-vector of the particle. This variable has the virtue of invariance to z boosts. In the ultrarelativistic limit, the rapidity Y is closely approximated by the pseudorapidity η, related only to the track angle θ according to

$$\eta = \frac{1}{2} ln \left[\frac{p + p\cos\theta}{p - p\cos\theta} \right] = -ln[\tan(\theta/2)]. \tag{3.4}$$

One can show that $Y \to \eta + \mathcal{O}(m^2/p^2)$. Other convenient variables include the transverse component of the momentum (p_T), the "transverse energy" E_T, and the approximately Lorentz-invariant angular distance ΔR, defined as:

$$p_T \equiv p\sin\theta, \;\; E_T \equiv E\sin\theta, \;\; \Delta R \equiv \sqrt{\eta^2 + \phi^2}. \tag{3.5}$$

3.2.2 Tracking System

Charged particles are observed in tracking detectors through the ionization they leave behind. The particle's trajectory can be reconstructed electronically from local clusters of ionization called "hits". The CDF tracking system is an integrated system including silicon tracking detectors and a large outer drift-chamber. Together, they provide three-dimensional particle tracking with excellent transverse momentum resolution and precise impact parameter measurement. All the tracking systems up to 1.4 m are immersed in a 1.4 T magnetic field which curves the charged particle, providing momentum information. As shown in Figure 12, the inner silicon detector consists of three subdetectors that extend from radius of $r = 1.35$ cm to $r = 28$ cm and cover the range range of $|\eta| < 2$. Surrounding the silicon detector is the Central Outer Tracker (COT) which covers the radius from 44 cm to 132 cm and 310 cm along z direction.

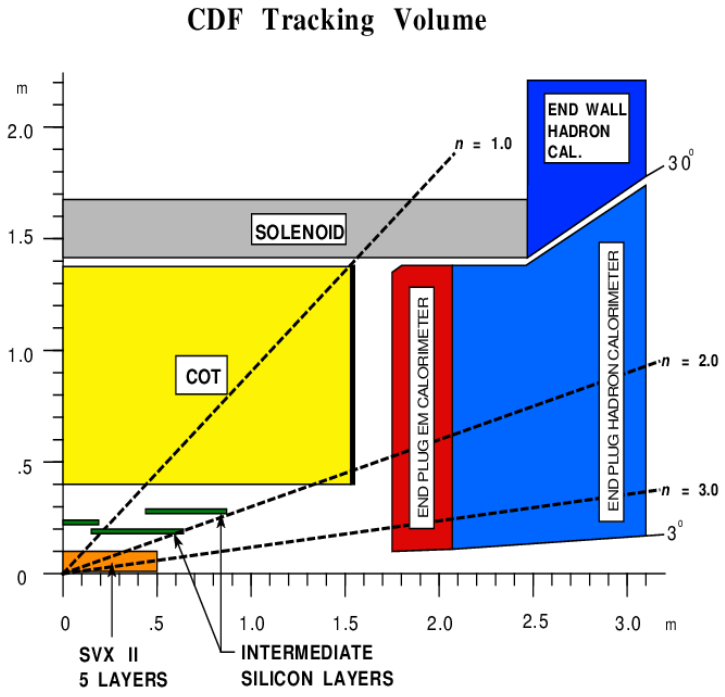

Figure 12: One quadrant of the CDF detector tracking layout.

The accurate measurement of tracks close to the beamline is essential for CDF physics analyses including B physics. The silicon detectors used at CDF were introduced into the

experiment for the first time in Run I, and significantly upgraded for Run II. Silicon detectors close to the beamline are ideal for precision impact parameter measurement for several reasons. The first reason is the excellent spatial resolution. The width of each silicon strip is about 60 μm, which gives a resolution 60 μm/$\sqrt{12} \sim 17\,\mu$m. Also, the silicon is close to the beam pipe and gives small error propagation. The main silicon detector is the Silicon Vertex Detector II (SVXII). The outer extension, the Intermediate Silicon Layers (ISL), link the tracks reconstructed in the chamber and hits detected in the SVXII, and also extends the track range to $1 < |\eta| < 2$ where the COT coverage is marginal. The inner extension, Layer 00 (L00), is a light-weight silicon layer placed on the beam-pipe. It recovers the resolution degradation of the reconstructed vertex position due to multiple scattering particularly in the SVXII read-out electronics and cooling systems.

Layer 00

The innermost layer of the silicon detector at a radius of 1.35 cm consists of one layer of single sided AC-coupled silicon sensors which cover the beryllium beam pipe over 80 cm longitudinal distance. The 7.85 cm long silicon sensors can be biased to very high voltages allowing a good signal-to-noise ratio. The strips are parallel to the beam axis giving the first sampling of a track in the $r - \phi$ plane. Signals from more than 13 K channels are fed via special optical fiber cables placed in a region separated from sensors and less exposed to the radiation. A flux of gas through tiny aluminum pipes installed between the sensor and the beam pipe keeps the operation temperature around 0 °C. The pressure of the gas is below atmospheric pressure to avoid leaks of fluid in case of damaged cooling pipe.

Silicon Vertex Detector II (SVXII)

The Silicon Vertex Detector is built in three cylindrical barrels each 29 cm long (see Figure 13(left)), at radii from 2.5 cm to 10.6 cm. Each barrel is made of five concentric layers of double-sided silicon sensors and divided into twelve wedges called "ladders" (see Figure 13(right)). Each ladder consists of two double sided rectangular 7.5 cm long sensors and the "hybrid", a multilayer board containing the frond end electronics, biasing circuits and fanouts. The ladders of each barrel are longitudinally juxtaposed head-to-head within a barrel's layer to leave the two hybrids at the two outside extremities of the barrel. The AC-coupled silicon sensors consist of microstrips implanted on a 300 μm n-type(n^-) bulk.

Bias is applied through integrated polysilicon resistors. Three types of strip orientations are used: $r-\phi$ (axial) strips oriented parallel to the beam axis, small angle stereo (SAS) strips tilted by 1.2° w.r.t the beam axis and the 90° stereo strips which lie in the transverse plane. All of the five layers have axial strips on one side, three of the other sides have 90° stereo and two have SAS strips. axial strips are p-type (p^+) and stereo strips are n-type(n^-). The measured average signal-to-noise ratio S/N$\geq$10, with a single hit efficiency greater than 99%. A water-glycol systems cools the whole SVXII system at roughly temperature of 10-15 °C.

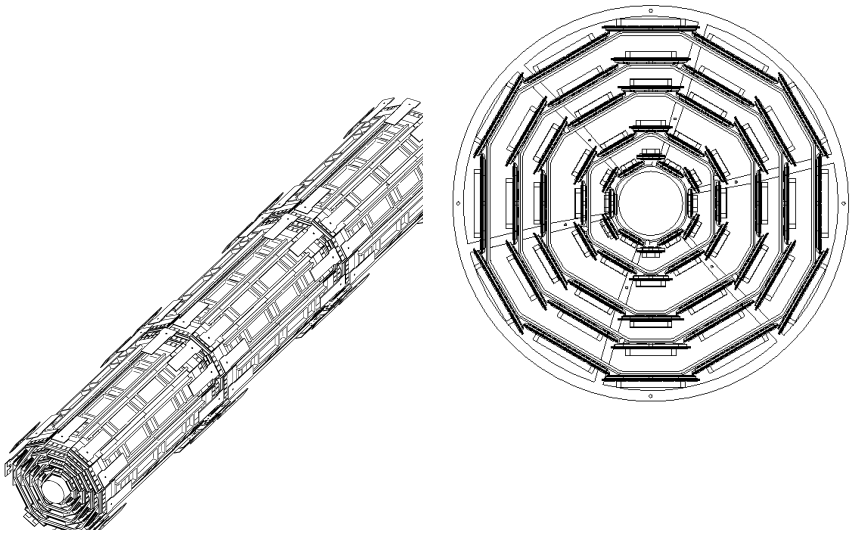

Figure 13: Three instrumental mechanical barrels of SVXII (left) and the cross section of a SVXII barrel (right).

Intermediate Silicon Layers (ISL)

The Intermediate Silicon Layer is a double-sided silicon detector segmented into twelve wedges like SVXII. It has a single layer of silicon in the central region at radius of 22 cm and two layers of silicon in the forward region at radius of 20 cm and 28 cm. ISL employs double sided AC-coupled 300 μm thick sensors, each sensor has axial strips on one side and SAS strips on the other.

Central Outer Tracker (COT)

The Central Outer Tracker is a cylindrical open-cell drift chamber. It contains 96 sense wire layers in the radial direction which are arranged into eight superlayers as shown in Figure 14(left). The maximum drift distance is approximately the same for all superlayers. Superlayers 1, 3, 5, 7 have their constituent sense wires oriented parallel to the beam axis ("axial") to measure hit positions in $r - \phi$ plane, the other superlayers have their sense wires tilted +3° or -3° w.r.t the beam axis ("stereo"). Combined readout of axial and stereo superlayers give $r - z$ position information. Each superlayer is divided in ϕ into open drift cells. Figure 14 (right) shows the transverse view of 3 cells. Each cell has 12 sense wires and 17 potential wires and are closed by the Mylar gold cathode of the detection circuit called "field panel"' along the azimuthal direction. The potential wires help to shape the electric field near the sense wires.

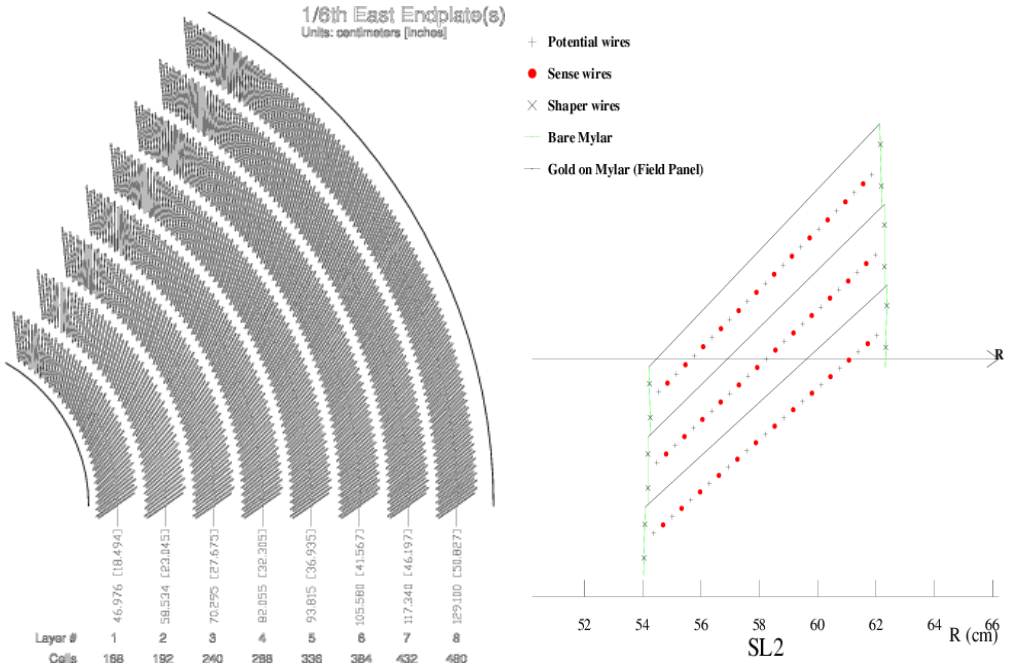

Figure 14: 1/6 section of the COT end-plate (left) and the transverse view of three COT cells (right).

The COT is filled with a 50:50 Argon-Ethane gas mixture which functions as an active medium. Charged particles that travel through the chamber will leave a trail of ionization in

the gas. Electrons drift, in crossed electric and magnetic fields, toward the sense wires at a Lorentz angle of 35°. So, the cells in each superlayer are not aligned along $\hat{r}$, a 35° azimuthal tilt is provided instead. The tilt also helps the high p_T (radial) tracks go through the full range of the drift distance in each superlayer. Electrons travel with a drift speed around $100\,\mu$m/ns, so the maximum drift time is about 100ns with a distance $\sim$ 1 cm between wires. The drift time is small compared with the inter-bunch spacing time of 396 ns to provide enough time for processing data from COT. Trigger signals derived from the COT are particularly relevant to this analysis.

Principle of Tracking

The trajectory of a charged particle which moves in a homogeneous solenoidal magnetic field ($\vec{B} = (0, 0, -1)$) is described by a helix. The projection of the helix on the $r-\phi$ plane is a circle. To parametrize this helix uniquely, five parameters are used : C, $\cot\theta$, d_0, ϕ_0 and z_0. C is the signed curvature of the circle defined as $C \equiv \frac{1}{2Q\rho}$, where ρ is the radius of the circle and Q is sign of the charge (± 1). So the positive charged tracks curve counterclockwise in the $r-\phi$ plane when looking in the negative z direction, while negative charge tracks curve the other way. θ is the polar direction of the particle at the point of its closest approach to z axis, it is also the angle between the momentum and the z axis. $\cot\theta$ is the helix pitch, which is related to the longitudinal component of the momentum: $p_Z = p_T \cot\theta$. ϕ_0 is the ϕ direction of the particle at the point of closest approach to the z axis. z_0 is the z coordinate of the point of closest approach to the z axis. d_0 is the signed impact parameter, i.e. the distance of closest approach to z axis defined as $d_0 \equiv Q(\sqrt{x_0^2 + y_0^2} - \rho)$, where (x_0, y_0) is the center of the helix circle. The transverse momentum of the track can be calculated in this way:

$$p_T = Q\frac{1.49896 \cdot 10^{-4} \cdot B_{magnet}}{C}, \tag{3.6}$$

where the unit of the magnetic field is Gauss, and the constant 1.49896 comes from half of the light speed $c/2$.

To reconstruct the tracks, only COT information is used first. Initially the algorithm looks for a circular path in the axial superlayers of the COT. Four or more hits in each axial superlayer are used to form a straight line or "segments". Then two approaches can be used for finding tracks. One way is to link together the segments which are consistent with the

tangent to a common circle. The other way is to constrain its circular fit to the beamline, then add hits which are consistent with this path. Once a circular track in the $r-\phi$ plane is found, segments in the stereo superlayers can be added in a three dimensional track fit. Once a track is constructed in the COT, it is extrapolated into the SVX detector. A three dimensional "road map" is formed based on the estimated errors on the track parameters. Starting from the outermost layer and working inward, silicon clusters found along the road are added to the track. For every added cluster, the track fit is updated, modifying the estimated track parameters and their errors, and generally narrowing the search "road". In the first pass, only $r-\phi$ clusters are used. Then stereo clusters are added. If there is more than one track with different combinations of SVX hits associated with the same COT track, the track with maximum number of SVX hits will be chosen. Transverse momentum resolution of the reconstructed track is very good. The COT alone can provide resolution at $\sigma_{p_T^2}/p_T \sim 0.16\%/\text{GeV/c}$. With SVX information added, the resolution can even reach $\sigma_{p_T}/p_T^2 \sim 0.07\%/\text{GeV/c}$. A more complete description of the COT and the SVX tracking can be found in [32].

3.2.3 Particle Identification System

In this analysis it is important to separate charged kaons from charged pions and other backgrounds, both for signal selection and for tagging. Two sub-detectors at CDF provide such information. The first one is based on ionization energy loss dE/dx measured in the COT. When a charged particle traverses the gas volume of the drift chamber, it leaves a trail of ionization along its path. The energy loss is proportional to the amount of ionization. The energy loss for relativistic charge particles other than electrons can be described by the Bethe-Bloch equation [33]:

$$\frac{dE}{dx} = \frac{4\pi N e^4}{mc^2\beta^2} z^2 (ln\frac{2mc^2\beta^2\gamma^2}{I} - \beta^2 - \frac{\delta(\beta)}{2}), \tag{3.7}$$

where N is the number density of electrons in the material, m is the electron mass, z is the incident particle charge, I is the mean excitation energy of the material atoms, $\beta = v/c$ and $\gamma^2 = 1/(1-\beta^2)$, and $\delta(\beta)$ is a correction of the density effect at high β. According to this

equation, for a given drift chamber, dE/dx only depends on particle velocity, so it can be used to distinguish particles of different mass when combined with momentum measurement. Different particles have quite different dE/dx in certain momentum regions, as can be seen in Figure 15. The difference between two particle hypotheses divided by the dE/dx resolutions gives the separation power. To make use of dE/dx from COT, the mean dE/dx as a function of $\beta\gamma$ is calibrated according to data with a so called "universal" curve [34]:

$$\frac{dE}{dx} = \frac{c_1 log\frac{\beta\gamma}{\beta\gamma+b} + c_0}{\beta^2} + a_1(\beta - 1) + a_2(\beta - 1)^2, \tag{3.8}$$

with all the coefficients are derived from fitting data. The dE/dx measurement provides one standard deviation separation power between charged kaons and charged pions for momentum greater than 2 GeV.

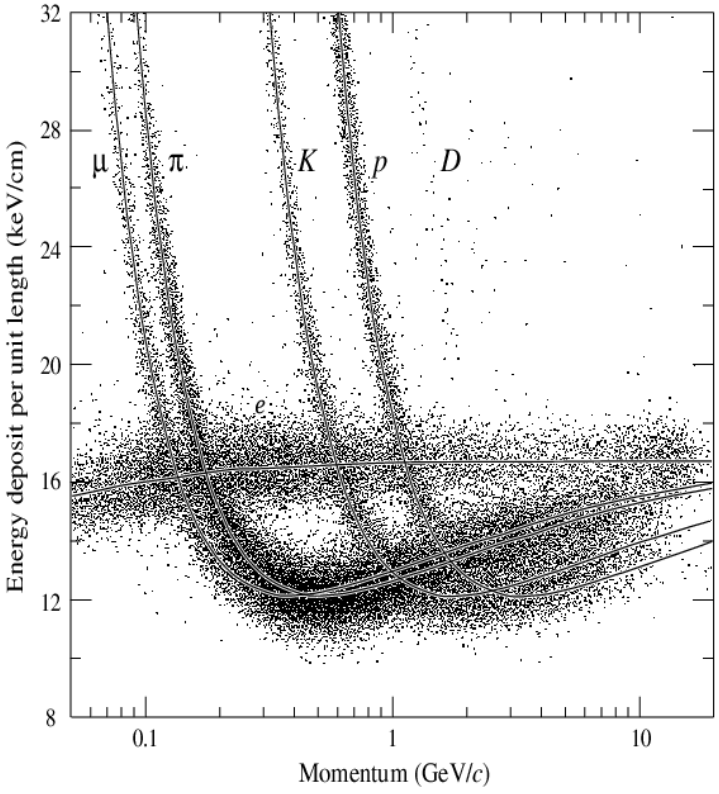

Figure 15: An example of dE/dx for different particles from the PDG data book.

The Time of Flight (TOF) detector [35] was added in 2001 to improve the particle identification capability. It is a cylindrical array made of 216 bars of Bicron BC-408 scintillator with dimensions $4 \times 4 \times 279$ cm located between the external surface of the COT and the cryostat of the super-conducting solenoid at a mean radius of 140 cm. The pseudo-rapidity coverage of the system is about $|\eta| < 1$. Both ends of each bar collect the light pulse using photomultiplier tubes (PMTs) and measure accurately the timing of the two pulses. The time of flight t is obtained by measuring the time of arrival of a particle at the scintillator with respect to the collision time. With momentum information from the track, the particle mass m can be determined:

$$m = \frac{p}{c}\sqrt{\frac{c^2t^2}{L^2} - 1}, \tag{3.9}$$

where L is the path length measured by the tracking system. With a time-of-flight resolution around 110 ps, the TOF system can provide at least a two standard deviation separation between charged kaons and charged pions for momentum $p < 1.6$ GeV, complementary to the dE/dx measurement from COT. A separation power plot for TOF is shown in Figure 16 together with dE/dx separation power superimposed.

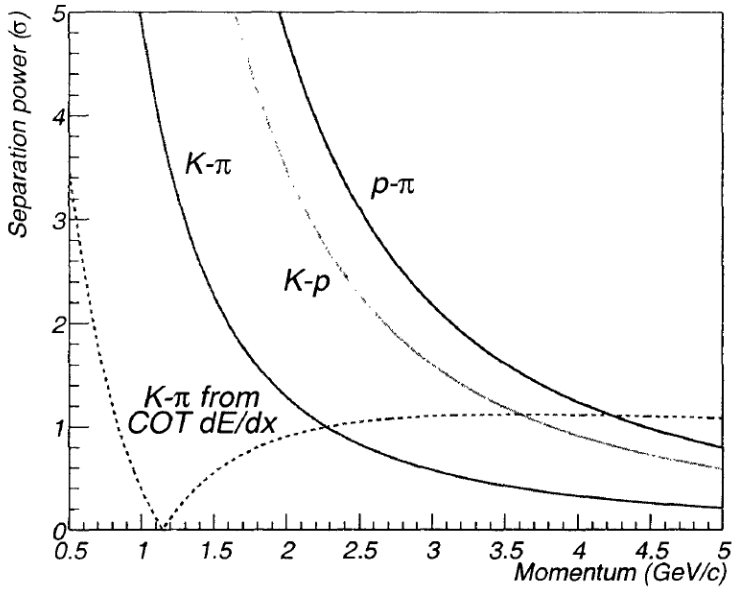

Figure 16: Separation power of TOF for different particles at CDF, with dE/dx separation power for kaon and pion from COT superimposed.

3.2.4 Calorimetry System

The CDF calorimetry is located outside the solenoid and composed of several systems of electromagnetic (EM) and hadronic scintillator sampling calorimeters which are segmented in a uniform pattern of projective towers. As shown in Figure 17, the calorimetry includes several different systems: Central Electromagnetic (CEM), Central Hadron (CHA), Wall Hadron (WHA), Plug Electromagnetic (PEM) and Plug Hadron (PHA). Together they cover the pseudorapidity region $|\eta| < 3.6$. The main task of the the calorimetry system is the measurement of energy deposition of photons, electrons and hadrons. The basic structure of the calorimeter is altering layers of passive absorbers and plastic scintillators. CEM and PEM use lead sheets for absorber material, while the CHA and WHA use steel, and the PHA uses iron. Particles with transverse momentum greater than about 350 MeV and $|\eta| < 1$ can reach the central calorimeters, where they will undergo energy loss, striking the absorber materials and producing daughter particles which interact in a cascade process, giving rise to a "shower" of particles. The showers propagate through many layers of absorbers and generate a detectable signal before they exhaust their energy. The sum of the signals collected by all the sampling active layers is proportional to the energy of the incident particle.

A charged particle with a track from the tracking system matching the electromagnetic shower is identified as an electron or a positron, with the sign of the charge obtained from the track curvature in the COT. An electromagnetic shower that does not have a matched charged particle track is identified as a high energy photon. Although muons have the same interactions as electrons, they do not create showers in the calorimeter, because the large muon mass (about 200 times that of the electron) suppress the bremsstrahlung. Hadrons are also much more massive than electrons, however they interact hadronically with the absorber to produce showers of a different shape: wider, more irregular, and further back in the calorimeter. The probability for an initial high-energy hadron to pass close enough to the a nucleus to initiate a shower is small enough that hadronic showers start at a random distance into the calorimeters, whereas EM showering is a continuous process that starts right at the the front face.

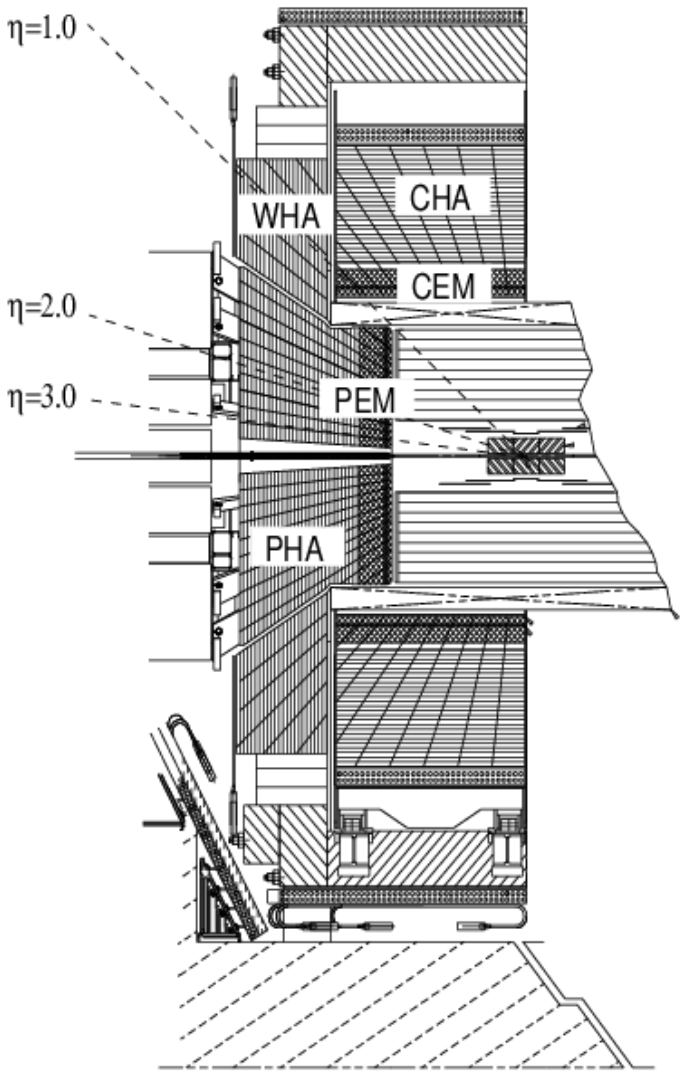

Figure 17: The CDFII calorimetry systems.

3.2.5 Muon System

Muons are over 200 times more massive than electrons, so bremsstrahlung radiation, inversely proportional to the mass squared of the incident particle, is suppressed by a factor of 4×10^4 w.r.t electrons. Muons do not interact strongly with nuclei in matter like pions and kaons either. Therefore, a muon with enough energy will pass through the calorimeter systems with minimum ionizing interactions. At CDF the minimum energy required to reach the muon chambers is 1.4 GeV. The Muon chambers are placed radially outside the calorimetry. In addition to the calorimeter, steel absorbers are placed upstream of the muon systems to reduce punch-through hadrons. The CDF muon system [36] consists of 4 subsystems: Central Muon (CMU), Central Muon uPgrade (CMP), Central Muon eXtension (CMX) and Intermediate MUon chambers (IMU). Figure 18 shows the coverage of the muon systems. The muon chambers employ sense wires parallel to the beamline and are filled with a 50:50 Argon-Ethane gas mixture. Muon candidates identified as track segments in the chambers are called muon stubs. A muon stub is matched with a track measured by the COT to reduce background from noise in the electronics and punch-through hadrons.

CMU

The Central MUon detector (CMU) is located outside the hadronic calorimeter at a radius of 347 cm from the beamline with coverage $|\eta| < 0.6$. The CMU is segmented into 24 wedges of 15° in ϕ, but only 12.6° of each wedge is active, so that the overall azimuthal acceptance is 84%. Each wedge is further segmented into three 4.2° modules each containing four layers of four drift cells as shown in Figure 19. The sense wires in these drift cells are made of stainless steel and held at potential of +2325 V. The timing information from the drift cells is used together with a time-to-distance relation, to reconstruct a muon stub. Muons with $p_T > 1.4$ GeV can reach the CMU.

CMP

The Central Muon uPgrade (CMP) is a second set of muon drift chambers outside of CMU with an additional 60 cm of steel absorbers between them. The material further reduces the probability of hadronic punch-through to the CMP. Muons need a transverse momentum about 2.2 GeV to reach the CMP. The CMP system is arranged in a box shape

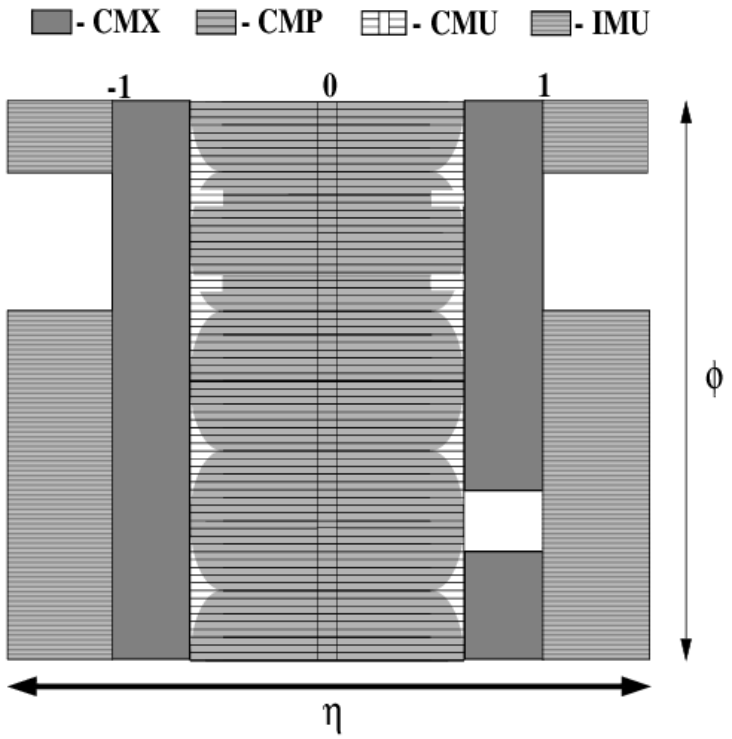

Figure 18: CDF muon chambers coverage in η and ϕ.

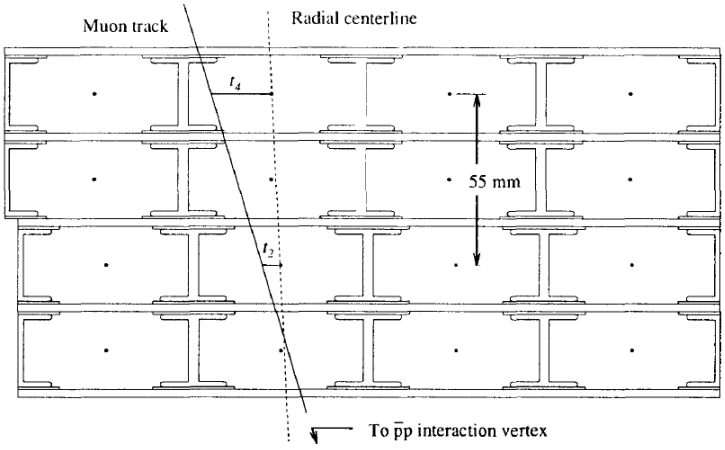

Figure 19: CMU module shown in cross-section, the layers offset are used to remove left-right ambiguity of hit reconstruction.

of similar acceptance as the CMU and conventionally serves as a confirmation of CMU for higher momentum muons. A layer of scintillation counters (CSP) is mounted on the outer surfaces of the CMP to improve light transmission in aging counters. The CMP and CMU have a large overlap in coverage and are often used together. CMP helps to cover CMU ϕ gaps and the CMU covers the CMP η gaps. A CMUP muon which has both CMU and CMP stubs is the least contaminated by fake muons.

CMX

The Central Muon eXtension (CMX) consists of drift tubes and scintillation counters (CSX) assembled in conically arranged sections. The CMX extends the pseudo-rapidity coverage to $0.6 < |\eta| < 1.0$. There are 8 layers in total and a slight stereo angle which provides z-coordinate information.

IMU

The Intermediate MUon (IMU) extends the pseudo-rapidity coverage even further to $1.0 < |\eta| < 1.5$. The IMU is mounted on the toroid magnets which provide shielding and consists of Barrel MUon chambers (BMU), Barrel (BSU) and Toroid (TSU) scintillation counters. The IMU is not used in this analysis, because no di-muon trigger is available for this system.

3.2.6 Trigger System

With all the CDF subdetectors functioning well, the CDF trigger system [37] separates signal from background at the hardware level. The overwhelming background in a hadron collider environment requires the trigger system to be able to extract the tiny fraction of interesting events. For example, the total $b\bar{b}$ cross section over the total inelastic cross section is of order $\sim$1/1000. The Tevatron is running at a luminosity around $10^{32}\mathrm{s}^{-1}\mathrm{cm}^{-2}$, and there are millions of interactions per second. But the maximum capacity for CDF to write out events is about 150 Hz. In order to achieve the required reduction in rate and record only the events of physical interest, the CDF trigger has been designed as a three-level system. Each level receives the data event from the previous level and, using more accurate detector information and more time for processing, makes a decision to reject or accept the event.

Level 1 is a hardware trigger taking input from only a subset of the detector components. Signals from the front end electronics go directly into Level 1, which has 5.5 μs to make its decision. The rate of events passed to Level 2 is typically about 25 KHz and limited to about 50 KHz. At Level 2 additional information including $r - \phi$ hits from the SVXII is incorporated. The events rate is reduced to 600-900 Hz. Level 1 and 2 are hardware based systems that use custom electronics. Level 3 is a software based trigger system implemented on a farm of about 500 commercial computers which accepts events at a rate of 100-150 Hz. A schematic diagram of the trigger system is shown in Figure 20.

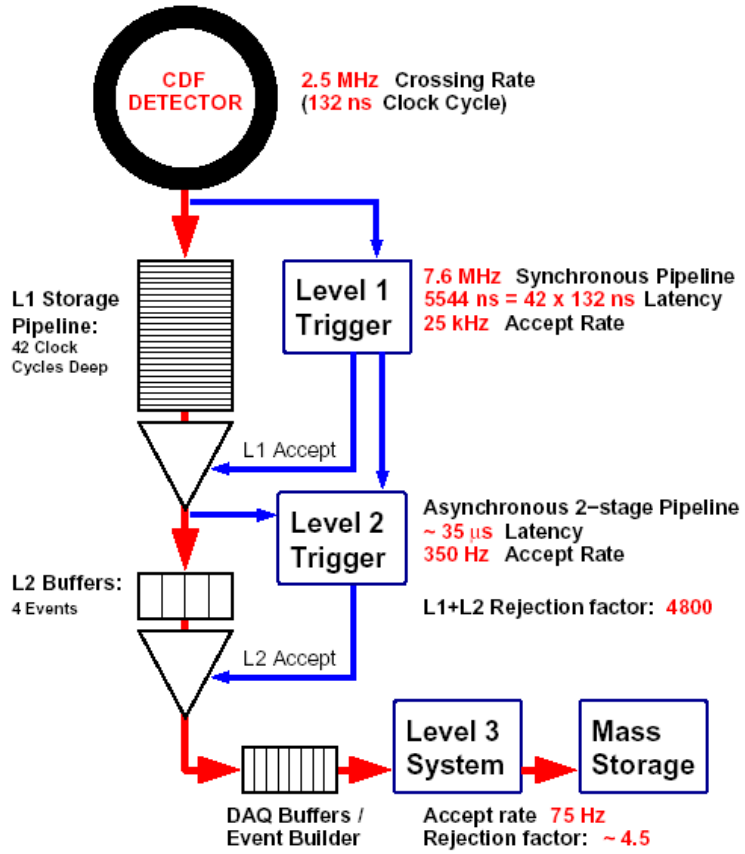

Figure 20: The CDFII trigger and data acquisition system.

Level 1

Level 1 is a synchronous system of custom-designed hardware which reconstructs information from three parallel streams: the calorimeter (total energy and some single tower

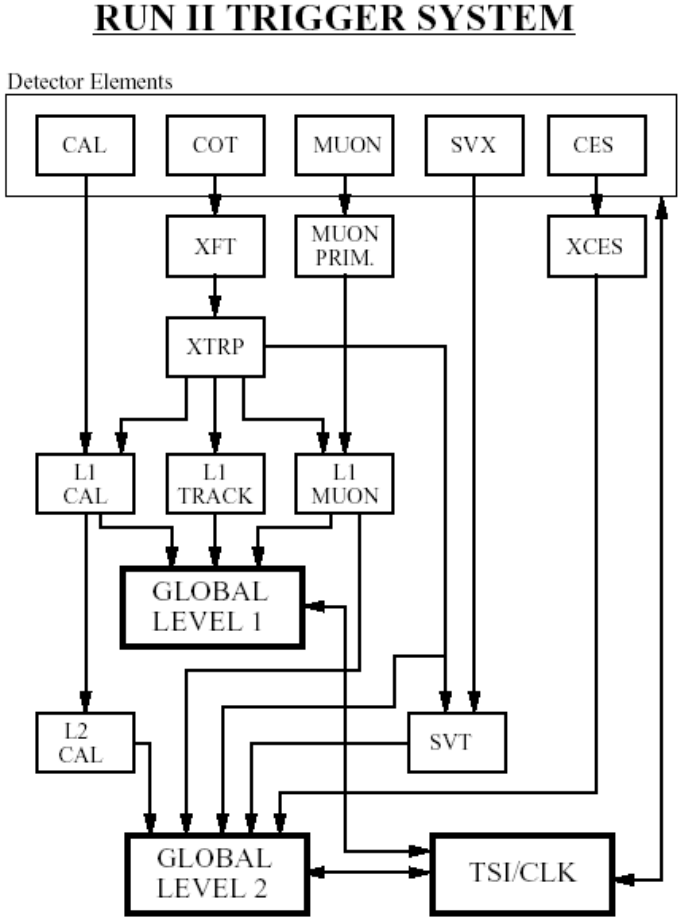

Figure 21: CDFII Level 1 and Level 2 systems with the major detector components.

information), the COT (only 4 axial superlayers are used for two-dimensional tracks) and the muon systems (stubs in the CMU, CMP and CMX). The eXtremely Fast Tracker (XFT) is custom processor used to identify two-dimensional tracks in the (r, ϕ) view of the COT. The XFT is capable of reconstructing tracks with $p_T > 1.5\ GeV$ with an efficiency of around 95% and a fake rate of a few percent. The XFT has an angular segmentation 1.25°, and an angular resolution of 0.3°. The momentum resolution is $\sigma_{p_T}/p_T^2 \sim 1.8\%/\text{GeV/c}$. The XFT passes the tracks to the extrapolation unit (XTRP) which feeds three L1 subprocesses: L1 CAL, L1 TRACK, and L1 MUON. L1 CAL and L1 MUON use extrapolated tracks and information from the calorimetry and muon systems respectively to search for possible electron, photon, jets and muon candidates. All three subprocesses report decisions to the Global Level 1 system and the accepted events will be buffered for Level 2 analysis.

Level 2

Level 2 is a combination of hardware and software triggers which processes events from Level 1 in the order they are accepted. It also incorporates additional information from the shower-max drift (CES) chambers in the central EM calorimeter and axial hits from the SVXII detector. An especially powerful Level 2 trigger processor is the Silicon Vertex Trigger (SVT). The SVT combines data from the XTRP and the SVX detector to identify displaced tracks indicative of B hadrons decays. The Level 2 output rate is reduced to around 900 Hz. A diagram of the decision process from the detector to Level 2 is shown in Figure 21.

Level 3

The final level of the CDF trigger is implemented exclusively by software on 500 commercial processors running in parallel. The output for each event passing the Level 2 trigger is read via optical fibers from all the sub detectors and sent to parallel processors for full event reconstruction. About 140 trigger paths are implemented at this level and the available output rate is around 150 Hz.

In the next chapter we describe how the CDF detector is used to select, reconstruct, and measure the decays of B^0, B^+, and B_s^0 that are of interest in this analysis.

4.0 DATA SAMPLE AND RECONSTRUCTION

This chapter discusses the procedure to reconstruct $B_s^0 \to J/\psi\phi$ decays, which will be used to measure the CP violation phase β_s. In addition to the $J/\psi\phi$ signal channel, the following decay channels involving different mesons are also reconstructed:

- B^+ mesons in the decay mode $B^+ \to J/\psi\, K^+$ are used to calibrate the flavor tagging algorithm, which will be described in the next chapter.
- B^0 mesons in $B^0 \to J/\psi\, K_s^0$ decays are used to measure $\sin 2\beta$ as a cross check of the implementation of the flavor tagging algorithm.

4.1 $B\bar{B}$ PRODUCTION AND TRIGGER

At the Tevatron, $b\bar{b}$ pairs are produced dominantly through QCD processes. Contributions from electroweak processes, such as $W^+ \to c\bar{b}$ and $Z \to b\bar{b}$ are negligible. The main production mechanism for $b\bar{b}$ is the gluon-gluon fusion process $g + g \to b + \bar{b}$ shown in Figure 22. At the lowest order there are 3 processes contributing in QCD $b\bar{b}$ production [38]:

1. Flavor Creation (direct production): Two gluons from beam particles interact through hard scattering resulting in two outgoing b quarks (gluon version). At the same lowest order a quark and antiquark form a $b\bar{b}$ pair through $q\bar{q}$ annihilation.
2. Flavor Excitation: One b quark from the sea of one of the beam particles is scattered with a parton from another beam particle.
3. Gluon Splitting: The $b\bar{b}$ quarks are created from a gluon after the hard scattering.

Figure 22 shows some of the lowest order Feynman diagrams from these three categories.

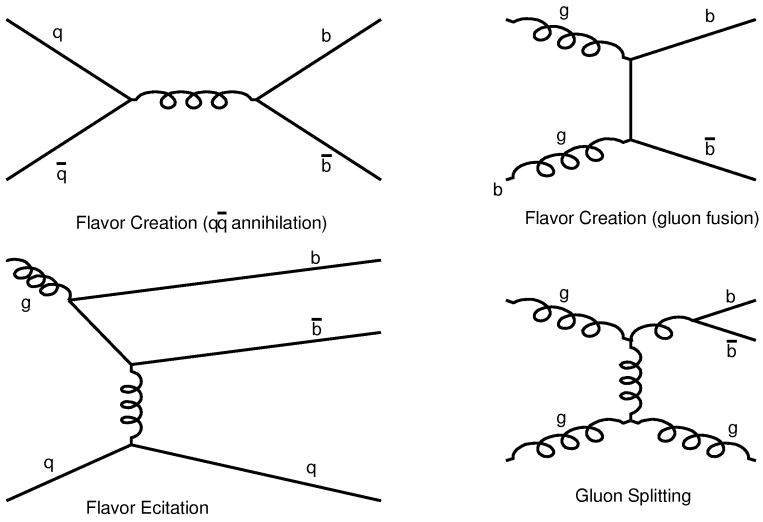

Figure 22: Lowest order diagrams of $b\bar{b}$ production at the Tevatron.

Once the $b\bar{b}$ quark pairs are produced, they will radiate gluons through the strong interaction. This process can be calculated perturbatively, because of the high virtuality scale $Q^2 \gg \Lambda^2_{QCD}$ indicating $\alpha_s \ll 1$. When the b and $\bar{b}$ quarks separate, the energy scale will decrease and the color interaction between the quarks will become stronger. At some point, the increasing potential energy between the quarks will be strong enough to create another $q\bar{q}$ pair from the vacuum. This process will repeat until the system creates a cluster of quarks and gluons with zero color and low internal momentum. A property of the strong interaction called color confinement binds the quarks to color-neutral hadronic final states which emerge from the interaction with a characteristic length below 1 fm. This process is known as hadronization. The probabilities for a b quark to hadronize into a B^-, $\bar{B}^0$, $\bar{B}^0_s$ or b-baryons are called the fragmentation fractions f_u, f_d, f_s and $f_{b-baryon}$. The contribution from excited B hadrons decaying into final states containing a B^-, $\bar{B}^0$, $\bar{B}^0_s$ or a b-baryon is included in this definition. The latest combined result from the Heavy Flavor Averaging

Group (HFAG) is [39]:

$$f_u = f_d = 0.401 \pm 0.010,\ \ f_s = 0.106 \pm 0.013,\ \ f_{b-baryon} = 0.092 \pm 0.018.$$

The B hadron production at the Tevatron is copious due to its large $b\bar{b}$ production cross section. QCD calculations determine the single B hadron cross section as [40]

$$\sigma_{|y|<0.6} = 16.8^{+7.0}_{-5.0}\ \mu\text{b}. \tag{4.1}$$

CDF measured in 2005 the b production cross section as [41]

$$\sigma[p\bar{p} \to H_b X,\ |y| < 0.6] = 17.6 \pm 0.4(stat)^{+2.5}_{-2.3}(syst)\ \mu\text{b}, \tag{4.2}$$

which agrees well with the theoretical prediction. For b production into J/ψ modes, which we are interested in, the cross section is much smaller, it is measured to be

$$\begin{aligned}\sigma[p\bar{p} \to H_b,\ H_b \to J/\psi X,\ p_T(J/\psi) > 1.25 GeV/c,\ |y(J/\psi)| < 0.6] \\ = 0.330 \pm 0.005(stat)^{+0.036}_{-0.033}(syst)\ \mu\text{b}.\end{aligned} \tag{4.3}$$

Although the $b\bar{b}$ production is large, it is only a few permil compared with the total inelastic $p\bar{p}$ cross section which is around 60 mb. Three types of dedicated triggers select B events from QCD background events at CDF: the di-muon trigger, the semileptonic trigger and the hadronic track trigger. The di-muon trigger selects muon pairs with a minimum transverse momentum cut of 1.5 GeV/c per muon. The semileptonic trigger selects events with a lepton (e or μ) with $p_T >$ 4 GeV/c and an SVT track with $p_T >$ 2 GeV/c, in addition the impact parameter of the SVT track must be greater than 200 μm. The hadronic trigger is also usually called two-track trigger. It selects two SVT tracks which satisfy the following requirements: sum of transverse momentum $p_{T1} + p_{T2} >$ 5.5 GeV/c, opening angle $2^\circ < \Delta\phi < 90^\circ$, impact parameter d_0 w.r.t. the primary vertex 120 μm $< d_0 <$ 1 mm, and decay length greater than 200 μm. In this analysis, the di-muon trigger is used to select the signal and calibration channels. The semileptonic trigger was used to develop the flavor tagging algorithms which will be described later. The hadronic trigger was used to observe $B_s^0 - \bar{B}_s^0$ oscillation and measure the oscillation frequency Δm_s, which is needed as input to this analysis. The measurement of CP violation in $B_s^0 \to J/\psi\phi$ was essentially not possible prior to the measurement of Δm_s.

4.2 DATA COLLECTION

To collect $B_s^0 \to J/\psi\phi$ decays, the di-muon trigger is used, which preferentially selects events containing $J/\psi \to \mu^+\mu^-$ decays [42]. The data were taken between March 2002 and October 2006 with integrated luminosity of 1.35 fb^{-1}. The basic data requirements are listed in Appendix C.

The J/ψ sample is defined by Level 1 and Level 3 trigger requirements. The trigger does not apply cuts on quantities related to the proper decay time of the B hadron, unlike the SVT triggers which severely bias the proper decay time distribution, making this an ideal sample in which to measure time dependent CP violation.

A J/ψ candidate at the Level 1 trigger must have either two muon stubs in the CMU, or one muon stub in the CMU and one in the CMX. The muon stubs must be matched to an XFT track; the matching requirements in track-stub displacement and angle are adjusted to be efficient for low-p_T muons (which undergo significant multiple-scattering in the calorimeters). XFT tracks associated with CMU stubs are required to have $p_T > 1.5$ GeV/c, while those associated with CMX stubs must have $p_T > 2.0$ GeV/c.

At the Level 3 trigger, the muon pairs are required to have opposite charge. The track-stub matching requirements are again enforced. The two muon tracks are also required to have $\Delta z < 5$ cm. Finally, the invariant mass of the di-muon pair is required to be $2.7 < m_{\mu\mu} < 4$ GeV/c^2.

4.3 OFFLINE RECONSTRUCTION

After a B_s^0 meson is produced, it will decay with an average lifetime $\tau \sim 1.5$ ps in the rest frame of the B_s^0 meson. Instead of using the proper decay time τ, we often refer to the proper decay length $c\tau \sim 450$ μm, where c is the speed of light. The topology of the decay chain $B_s^0 \to J/\psi\phi$, $J/\psi \to \mu^+\mu^-$, $\phi \to K^+K^-$ is shown in Figure 23.

A basic input quantity used in this analysis is the decay length L of a final-state B meson, which is the distance between its production and decay point in the laboratory frame.

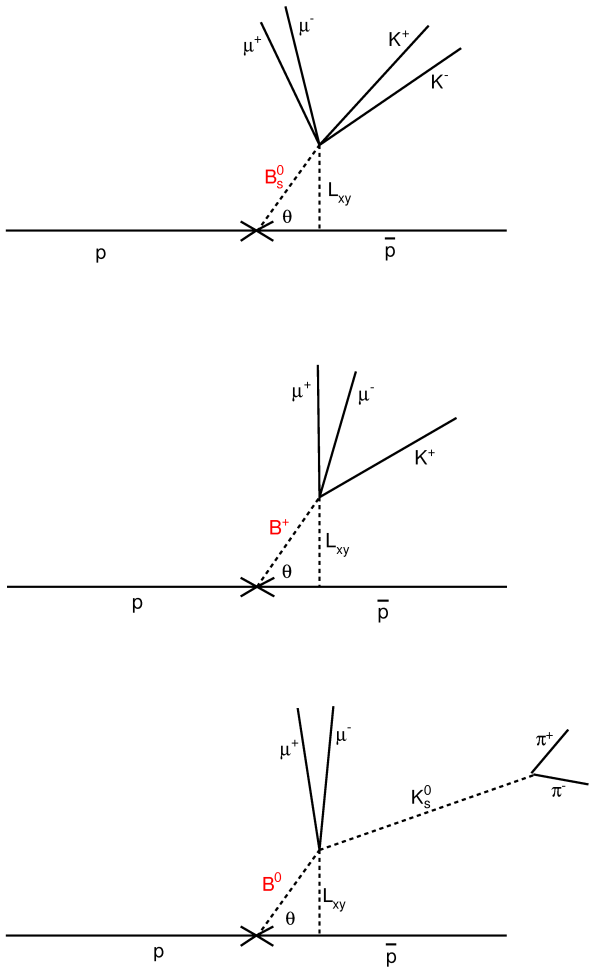

Figure 23: Illustrations of B_s^0 decays to J/ψ and ϕ, B^+ decays to J/ψ and K^+, B^0 decays to J/ψ and K_s^0.

Daughter tracks of the B meson and the primary vertex (where the B meson is produced) are needed to measure this quantity. The former are used to reconstruct the B meson, its four-momentum, and its decay point. The latter gives its production point.

The proper decay time is determined in the following manner:

$$\begin{aligned} ct &= ct_{lab}/\gamma \\ &= c\frac{L_{xy}/\sin\theta/v}{\gamma} \\ &= \frac{L_{xy}M_B}{p_T}, \end{aligned} \tag{4.4}$$

where L_{xy} is the transverse decay length defined as

$$L_{xy} = \frac{\vec{V}\cdot\vec{p}_T}{|\vec{p}_T|},$$

and $\vec{V}$ is the vector pointing from the primary to the secondary vertex position. Both vectors $\vec{V}$ and $\vec{p}_T$ are two-dimensional vectors, defined in the $r\phi$-plane.

4.3.1 Primary Vertex Reconstruction

Two methods of estimating the primary vertex are commonly used. The first method uses the time-dependent average beamline, which determines the beamspot with a lateral error of about 30 microns [43]. The second method uses the event-by-event primary vertex reconstructed from tracks in the event which do not belong to the reconstructed B mesons [44, 45]. The event-by-event primary vertex obtains a more accurate primary vertex position, and thus a more accurate estimate of the proper decay length determination. In this thesis, event-by-event primary vertex method is used.

4.3.2 Track Quality Cuts

To ensure the use of high quality tracks, a candidate track is required to have at least 5 COT hits from at least 2 axial and 2 stereo COT super-layers. For the tracks corresponding to the muons from $J/\psi \to \mu^+\mu^-$, the kaon in the $B^+ \to J/\psi\, K^+$ decay and the kaons of the $\phi\to K^+K^-$ decay from $B_s^0 \to J/\psi\,\phi$, at least 3 $r-\phi$ silicon hits, which can include ISL and L00 hits, are also required.

4.3.3 Muon Identification and Selection

A muon candidate is formed by a muon hit cluster (stub) in the muon chamber position-matched to a single track in the tracking system. The χ^2 of the position match between track segments in the CMU or CMX and the extrapolated track is required to be less than 9 in the $r\phi$-plane. Should a muon stub be matched to more than one track, the ensemble with the minimum χ^2 is selected as the muon candidate. The muon tracks used are of any of the following types: CMU, CMP, CMUP and CMX and are all to have required $p_T > 1.5$ GeV/c.

4.3.4 $\mathbf{J/\psi \to \mu^+\mu^-}$ Reconstruction

J/ψ candidates are formed from pairs of oppositely charged muons within the events of the sample. The two muons are fitted to a common vertex using the kinematic fitting program called `CTVMFT` [46] via the C++ wrapper `VertexFit` at CDF. The fit returns a χ^2 as well as an estimated vertex position, and refitted vertex-constrained tracks. The refitted tracks are used to estimate the relevant quantities like p_T and fitted J/ψ invariant mass. The $\mu^+\mu^-$ pair is considered as a J/ψ candidate if it has a vertex χ^2 probability > 0.001 and the fitted mass is within the range $3.014 < \mathrm{m}(\mu\mu) < 3.174$ GeV/c^2.

4.3.5 $\mathbf{K^0_s \to \pi^+\pi^-}$ Reconstruction

To search for K^0_s candidates we examine pairs of oppositely charged, non-muon, tracks in the J/ψ sample which intersect when projected into the transverse plane. Both tracks are considered to be pions. The invariant mass of a selected track pair is calculated at their intersection point. If it is inside a 20 MeV/c^2 window around the PDG K^0 mass value [47]: $0.473 < \mathrm{m}(\pi\pi) < 0.523$ GeV/c^2, the pair is kept.

Since K^0_s mesons have a large decay distance ($c\tau = 2.7$ cm), they can decay inside or even outside the SVX II detector. For this reason, no silicon requirement is made for K^0_s daugther tracks.

Some variables can be useful to separate a K^0_s meson from combinatorial background and Λ^0 background: 1) The transverse decay length of the K^0_s candidate, L_{xy}, which is defined

as the distance between the J/ψ decay vertex ($\mu^+\mu^-$ intersection point) and the K_s^0 decay vertex ($\pi^+\pi^-$ intersection point) in the transverse plane. 2) To remove the Λ^0 contamination in the K_s^0 sample, an active veto can be used assigning the proton/pion mass hypothesis to the track candidate forming the K_s^0. 3) The reconstructed K_s^0 line-of-flight (J/ψ to K_s^0 vertices) can be used to constrain the multi-track vertex fit in the reconstruction of the decay $B^0 \to J/\psi\, K_s^0$. This has been shown to improve significantly the reconstruction of B^0 mesons in the above channel (see the study done with an inclusive K_s^0 sample in Ref. [48], Appendix B).

4.3.6 $\phi \to \mathbf{K^+K^-}$ Reconstruction

To search for a ϕ candidate, we examine pairs of oppositely charged, non-muon tracks in J/ψ events that originate form a common displaced vertex. For the ϕ reconstruction, the two tracks are assumed to be kaons. The pair is kept if its invariant mass is within $1.008 < \mathrm{m}(K^+K^-) < 1.032$ GeV/c^2.

4.4 RECONSTRUCTION OF EXCLUSIVE MODES

Besides the common data selection procedures, an artificial neural network (NN) selection is utilized for each particular channel. We describe here the channel specific cuts used for the pre-selection, as well as the variables input to the NN optimization procedure. The pre-selection cuts are simply loose requirements to reduce the combinatorial background.

Errors in primary and secondary vertex positions are propagated to errors σ_{ct} in proper decay time according Eq. 4.4, while the error from the p_T determination is negligible when calculating the ct error. All masses and mass errors are returned from the `CTVMFT/VertexFit` package.

4.4.1 Channel-Specific Pre-Selection Cuts

- B^+, B^0, B_s^0

  - Vertex fit includes a J/ψ mass constraint to 3.096916 GeV/c^2,
  - No K_s^0 or ϕ mass constraint.
- $B^+ \to J/\psi\, K^+$
  - $p_T\,(K^+) > 1.0$ GeV/c
  - $p_T\,(B^+) > 4.0$ GeV/c
- $B^0 \to J/\psi\, K_s^0$
  - $p_T\,(K_s^0) > 0.5$ GeV/c
  - $p_T\,(B^0) > 4.0$ GeV/c
- $B_s^0 \to J/\psi\, \phi$
  - $p_T\,(\phi) > 1.0$ GeV/c
  - $p_T\,(B^0) > 4.0$ GeV/c

  After pre-selection the B_s^0 invariant mass distribution is shown in Figure 24. The total number of events is around 174,000 indicating a B_s^0 signal on a large background.

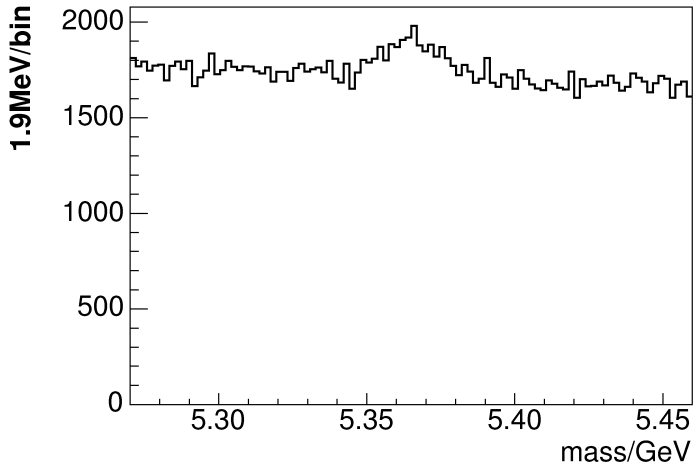

Figure 24: The B_s^0 invariant mass distribution of $B_s^0 \to J/\psi\phi$ after pre-selection.

4.4.2 Neural Network Training

As can seen in Figure 24, the background is huge compared to the B_s^0 signal. To suppress the combinatorial background, we need to further optimize the selection. One way to do

this is called the rectangular cuts. By applying cuts on different variables at the same time, such as transverse momentum and vertex fit probability, to a training sample with signal coming from Monte Carlo simulation and background coming from sidebands of data, one can find the best combination of cuts to maximize a figure of merit, e.g., $S/\sqrt{S+B}$, and apply the cuts on the data sample. However, this method requires managing different variables (discriminants) together, and the correlation between different variables is not always apparent. To overcome these difficulties, one can train an artificial neural network (NN) to optimize the selection, where all the discriminants can be combined into one neural network output value and the correlations are taken care of automatically.

The final selection of the B meson candidates is done using the three layer feed-forward network [49]. A training of the neural network is necessary at first. The neural network is given two different samples, one is pure signal, and the other is pure background. The background events can be taken from the sidebands of the data mass distribution. The signal events are produced with a full simulation (see Section 4.5), which includes trigger emulation and the complete `GEANT` representation of CDF II detector.

The variables used to train the neural network for $B^+ \to J/\psi K^+$ decay as input are, in this order: 1) transverse momentum of B^+, 2) B^+ vertex fit probability, 3) J/ψ vertex fit probability, 4) J/ψ transverse momentum , 5) K^+ transverse momentum, 6) K^+ particle identification likelihood.

For the $B^0 \to J/\psi\, K^0_s$ decay, the variables used as input are, in this order: 1) B^0 vertex probability , 2) B^0 transverse momentum , 3) K^0_s L_{xy}, 4) K^0_s $L_{xy}/\sigma(L_{xy})$, 5) K^0_s vertex probability, 6) J/ψ vertex probability, 7)K^0_s invariant mass, 8) Λ-veto-mass, 9) K^0_s transverse momentum, 10) π^+ transverse momentum, 11) π^- transverse momentum. The correlations between the different input variables are shown in Figure 25 in the order written above. The results of the training for $B^0 \to J/\psi\, K^0_s$ can be seen in Figure 26.

For the $B^0_s \to J/\psi\, \phi$ decay, the variables used as input are, in this order: 1) B^0_s vertex χ^2 in $r-\phi$, 2) B^0_s vertex probability, 3) B^0_s transverse momentum 4) J/ψ transverse momentum, 5) J/ψ vertex probability, 6) ϕ invariant mass , 7) ϕ transverse momentum, 8) K^+ transverse momentum, 9) K^+ particle identification likelihood, 10) K^- transverse momentum, 11) K^- particle identification likelihood. The correlations between the different input

variables are shown in Figure 27 in the order written here and the results of the training for $B_s^0 \to J/\psi\,\phi$ can be seen in Figure 28.

Of all the trainning variables used for the training, transverse momemtum of the tracks are found to have the dominant weights for all the decay channels.

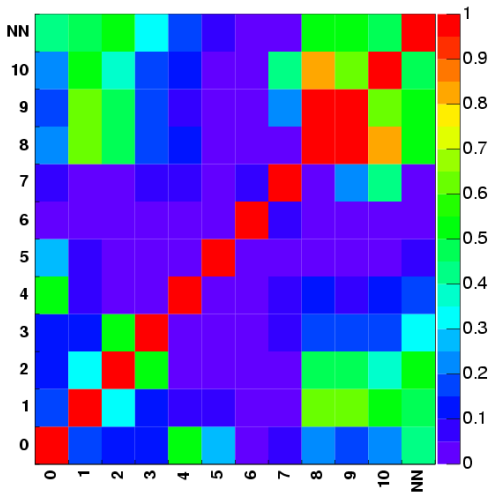

Figure 25: Correlation between different input variables used in the $B^0 \to J/\psi\,K_s^0$ selection.

As shown in Figures 26 and 28, with different cuts on the NN output variable, different signal-to-background ratios are obtained. Cutting higher on the NN output variable, a purer sample is obtained, but the total signal yield decreases.

The neural network training can be checked by applying it to a pure background sample. For an unbiased neural network, one expects no enhancement in the signal mass region for any NN output cut. Figure 29 shows different NN output cuts applied to a same sign kaon sample where the $\phi \to K^+K^-$ is reconstructed with two same sign kaons. As one can see, even with a very tight NN output cut (where a higher signal purity is expected), there is still no sign of a signal enhancement in B_s^0 mass region.

4.4.3 Signal Optimization

After the neural network is trained and checked, it is applied to the data. All the events in the data sample are assigned different NN output values. To optimize the NN output cut,

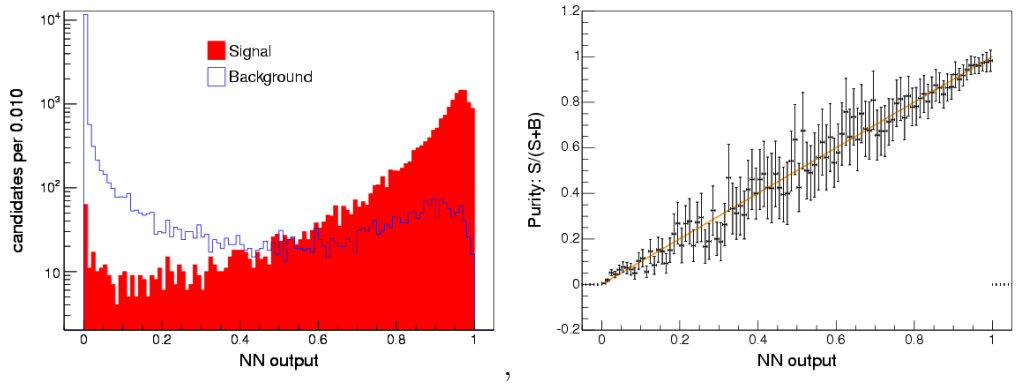

Figure 26: Distribution of neural network output variable for simulated $B^0 \to J/\psi\, K_s^0$ decays (red) and background events from the sidebands (blue) (left), and purity of the trainings sample as a function of the network output (right).

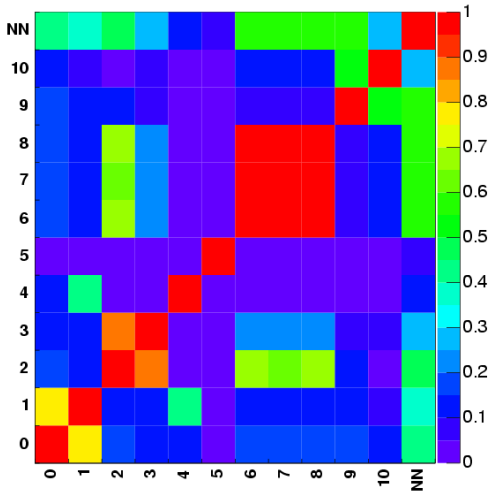

Figure 27: Correlation between the different input variables for $B_s^0 \to J/\psi\, \phi$ as defined in the text.

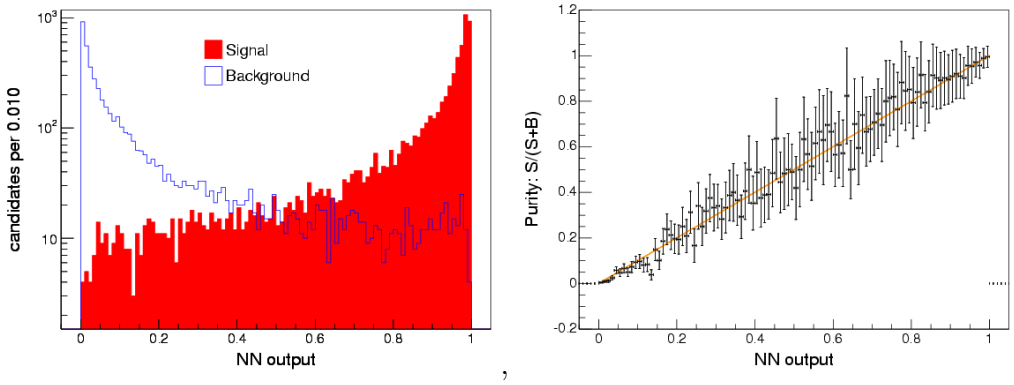

Figure 28: Example of neural network output variable distribution for the simulated $B_s^0 \to J/\psi\,\phi$ mesons (red) and background events from the sidebands (blue) (left), and purity of the trainings sample as a function of the network output (right).

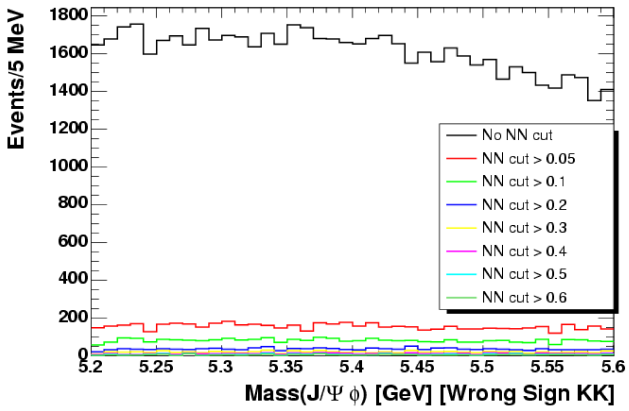

Figure 29: Mass distribution for different NN output cuts for a sample of same sign kaons.

the figure of merit $S/\sqrt{S+B}$ is used. Here, S and B represent the number of signal and background events respectively in the signal mass region defined between 5.3426 GeV/c^2 and 5.3966 GeV/c^2. Different NN output cuts varying between 0 and 1, with a step of 0.05, are tested. For each step, the number of signal and background events in the signal mass region are both calculated. The signal-to-background ratio, and the figure of merit as a function of the NN output cut for $B_s^0 \to J/\psi\phi$ channel are shown in Figure 30. The figure of merit is almost flat between 0.2 and 0.7. In order to obtain more signal events, the final cut on the NN output is chosen to be 0.26.

The signal yield after applying the NN output cut is estimated using two methods. In the first way, the number of events in the signal mass region and the number of events in the two equally sized sideband mass regions defined as (5.4236 GeV/c^2, 5.4506 GeV/c^2) $\bigcup$ (5.2886 GeV/c^2, 5.3156 GeV/c^2) are counted. Taking the difference in events, $1,921 \pm 79$ events are left. The signal-over-background ratio is around 1.13 and the value for the figure of merit is 32.8. For the second method, the number of signal events is estimated by fitting the mass distribution with a Gaussian describing the signal and a linear function describing the background. Figure 31 shows the mass distribution with the result of the fit superimposed. The estimated number of events is $1,939 \pm 62$, which is consistent with the result obtained using the counting method.

Same procedures are used in the case of $B^+ \to J/\psi\, K^+$ and $B^0 \to J/\psi\, K_s^0$ decays. Mass region definitions and results are all shown in Table 4, and the mass fits for the two channels can be seen in Figure 32.

4.5 FULLY SIMULATED MONTE CARLO

Fully simulated Monte Carlo is used in this analysis in two ways. First, for the neural network training discussed in section 4.4.2, second, to obtain the angular efficiencies or "detector sculpting" (see section 5.1). A Monte Carlo sample is produced with the `HeavyQuarkGenerator` package [50] for the corresponding expected signals. B mesons are generated above a p_T threshold of 4 GeV/c and within a pseudo-rapidity of $|\eta| < 1.3$. The

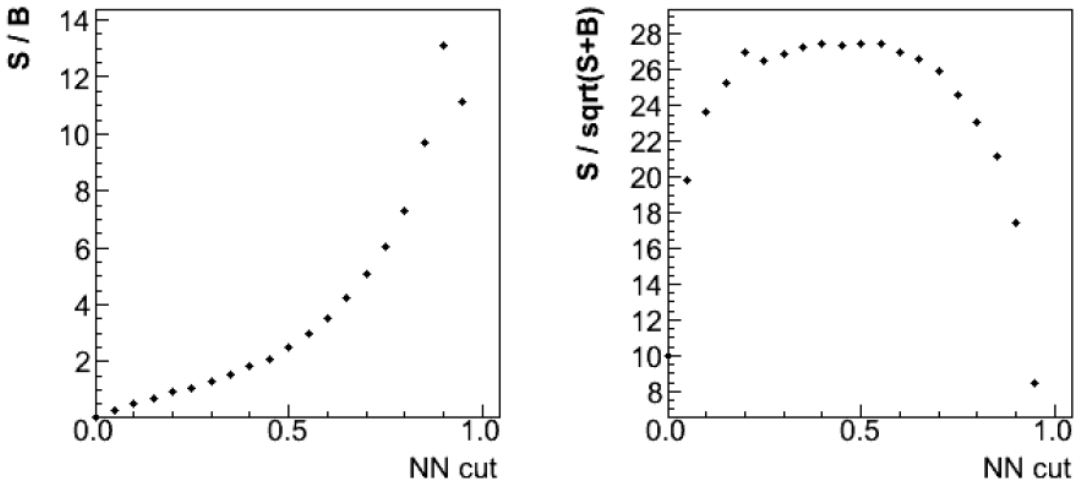

Figure 30: Signal over background S/B (left), and figure of merit $S/\sqrt{S+B}$ (right) as a function of the NN output cut.

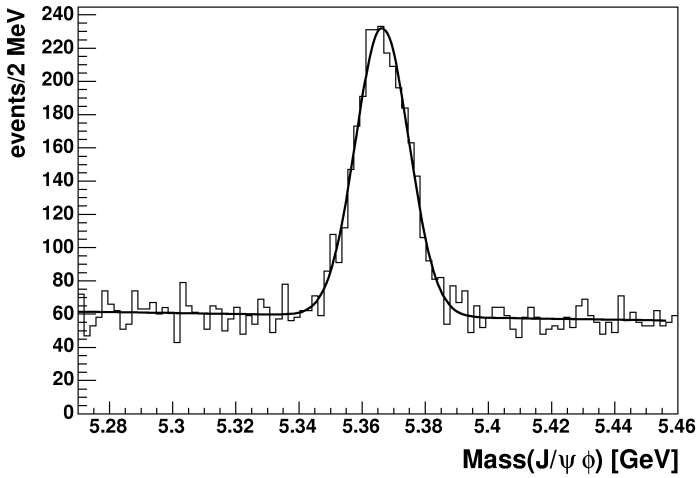

Figure 31: Invariant mass distribution of $\mu^+\mu^- K^+K^-$ with the result of the fit superimposed.

Decay	Signal Region (GeV/c^2)	Left Sideband (GeV/c^2)	Right Sideband (GeV/c^2)	Yield	S/B	$S/\sqrt{S+B}$
$B_s^0 \to J/\psi\,\phi$	(5.3426, 5.3966)	(5.2886, 5.3156)	(5.4236, 5.4506)	$1,939 \pm 62$	1.1	32.8
$B^+ \to J/\psi\,K^+$	(5.24645, 5.31085)	(5.18205, 5.21425)	(5.34305, 5.37525)	$18,748 \pm 292$	1.1	99.4
$B^0 \to J/\psi\,K_s^0$	(5.25031, 5.31048)	(5.19014, 5.22022)	(5.34056, 5.37065)	$5,724 \pm 130$	0.95	52.0

Table 4: Signal and sideband region definitions, signal yields, signal over background ratio S/B and figure of merit $S/\sqrt{S+B}$ for $B_s^0 \to J/\psi\,\phi$, $B^+ \to J/\psi\,K^+$, and $B^0 \to J/\psi\,K_s^0$.

B mesons are generated according to the transverse momentum spectrum measured in the inclusive J/ψ cross-section measurement [51] and are decayed using the `EVTGEN` package [52]. The events are filtered post-generation requiring the p_T of the muons to be greater than 1.4 GeV/c. The surviving events are reconstructed using the full CDF II detector simulation and offline reconstruction. For the training of the neural network, ten million events are generated in each of the three decay modes. The offline version of the B group Monte Carlo used for generation and reconstruction is `6.1.4mc`. The number of events generated corresponds to the first 1.35 fb^{-1} of data. The masses and lifetimes of the particles are taken from the `mcProduction/Bgroup/decay/pdt` − `nomix.table`. Finally, the Monte Carlo events are reconstructed with the `ProductionExe` executable and passed through the same analysis procedure as applied to data.

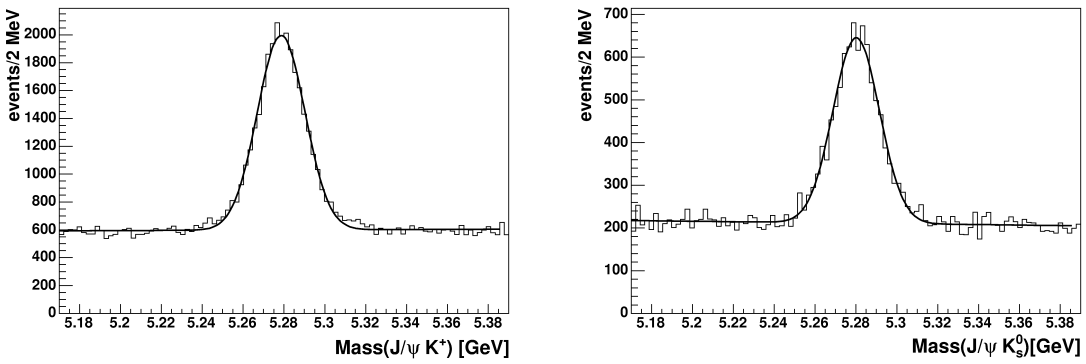

Figure 32: Invariant mass of $\mu^+\mu^- K^+$ with fit superimposed (left), and invariant mass of $\mu^+\mu^-\pi^+\pi^-$ with fit superimposed (right).

5.0 EXPERIMENTAL STRATEGY

The measurement of β_s requires an experimental strategy with several components. Since the final state $J/\psi\phi$ is CP-mixed, the CP even and odd components must be separated using an analysis of the angular decay variables. To measure a CP asymmetry, the flavor of the B_s^0 meson at production time is also needed. Ultimately, the extraction of the CP mixture (via the amplitudes A_0, $A_\parallel$, $A_\perp$) and the CP phase is done using an unbinned maximum likelihood method that combines elements of the angular analysis with elements of a flavor-tagged, time-dependent CP asymmetry measurement. This is a simultaneous fit to mass, proper decay time, proper decay time uncertainty, tagging decision, and angular variables. In this chapter we first describe the way that the angular variables are reconstructed and their distributions are modeled. Then, we describe how the tagging decision is produced and modeled. Finally, we describe the likelihood function used to extract a confidence interval in the space of physical parameters, especially β_s and $\Delta\Gamma$.

5.1 ANGULAR DISTRIBUTIONS IN THE TRANSVERSITY BASIS

As discussed in section 2.6.1, three angles are needed to specify the decays $B_s^0 \to J/\psi\phi$: the polar angle θ and the azimuthal angle ϕ of the positive muon in the J/ψ rest frame, and the helicity angle ψ of the positive kaon in the ϕ rest frame. The four-momentum of all tracks in the laboratory frame is reconstructed by the tracking system. From these quantities the four-momentum of the B_s^0, the J/ψ, and the ϕ is computed. The Lorentz boost of these (magnitude, and direction) are derived from these quantities. With the boost one can compute all of the needed four-momentum, and three-momentum, in the rest frame

of any reconstructed particle.

To get the helicity angle ψ of the K^+ in the ϕ rest frame, one boosts the four-momentum of K^+ and J/ψ with the velocity of ϕ meson $\vec{v}_\phi$ measured in the lab frame:

$$(E_{K^+}, \vec{p}_{K^+}) \rightarrow (E'_{K^+}, \vec{p'}_{K^+}), \tag{5.1}$$

$$(E_{J/\psi}, \vec{p}_{J/\psi}) \rightarrow (E'_{J/\psi}, \vec{p'}_{J/\psi}). \tag{5.2}$$

Then,

$$\cos\psi = -\frac{\vec{p'}_{K^+} \cdot \vec{p'}_{J/\psi}}{|\vec{p'}_{K^+}||\vec{p'}_{J/\psi}|}. \tag{5.3}$$

To get the polar angle θ and the azimuthal angle ϕ of the μ^+ in the J/ψ rest frame, the four-momentum of the ϕ meson, the K^+ meson, and the μ^+ are boosted with the velocity of the J/ψ $\vec{v}_{J/\psi}$ measured in the lab frame:

$$(E_{K^+}, \vec{p}_{K^+}) \rightarrow (E'_{K^+}, \vec{p'}_{K^+}), \tag{5.4}$$

$$(E_{\phi}, \vec{p}_{\phi}) \rightarrow (E'_{\phi}, \vec{p'}_{\phi}), \tag{5.5}$$

$$(E_{\mu^+}, \vec{p}_{\mu^+}) \rightarrow (E'_{\mu^+}, \vec{p'}_{\mu^+}). \tag{5.6}$$

The axes $\vec{x'}$, $\vec{y'}$ and $\vec{z'}$ are given by:

$$\hat{x'} = \frac{\vec{p'}_{\phi}}{|\vec{p'}_{\phi}|}, \tag{5.7}$$

$$\hat{y'} = \frac{\vec{p'}_{K^+} - (\vec{p'}_{K^+} \cdot \hat{x'})\hat{x'}}{|\vec{p'}_{K^+} - (\vec{p'}_{K^+} \cdot \hat{x'})\hat{x'}|}, \tag{5.8}$$

$$\hat{z'} = \hat{x'} \times \hat{y'}. \tag{5.9}$$

Then, $\cos\theta$ and ϕ are obtained through:

$$\cos\theta = \frac{\vec{p'}_{\mu^+}}{|\vec{p'}_{\mu^+}|} \cdot \hat{z'}, \tag{5.10}$$

$$\phi = \tan^{-1}\left((\frac{\vec{p'}_{\mu^+}}{|\vec{p'}_{\mu^+}|} \cdot \hat{y'}) / (\frac{\vec{p'}_{\mu^+}}{|\vec{p'}_{\mu^+}|} \cdot \hat{x'}) \right), \tag{5.11}$$

where the ambiguity of the angle ϕ can be resolved with the signs of $\vec{p'}_{\mu^+} \cdot \hat{x'}$ and $\vec{p'}_{\mu^+} \cdot \hat{y'}$.

5.1.1 Detector Sculpting of the Signal Angular Distributions

The detector acceptance and the transverse momentum cuts (see chapter 4) affect the observed distribution of the kinematic variables through an efficiency function $\epsilon(\psi, \theta, \phi)$. With this detector sculpting of the angular distribution, the normalized probability density function discussed in section 2.6.3 will be spoiled. A new normalization of the function N' must be obtained from

$$\begin{aligned} N' &= \int\int\int \epsilon(\psi,\theta,\phi) d(\cos\psi) d(\cos\theta) d\phi \\ &\quad \left[|\mathbf{A}_+ \times \hat{n}|^2 \times (\tau_L + \tau_H + \cos 2\beta_s (\tau_L - \tau_H)) \right. \\ &\quad + |\mathbf{A}_- \times \hat{n}|^2 \times (\tau_L + \tau_H - \cos 2\beta_s (\tau_L - \tau_H)) \\ &\quad \left. + 2Re((\mathbf{A}_+ \times \hat{n}) \cdot (\mathbf{A}_-^* \times \hat{n}) \times (i \sin 2\beta_s (\tau_L - \tau_H))) \right]. \end{aligned} \tag{5.12}$$

With this new normalization, the renormalized probability density function for the signal is:

$$\begin{aligned} P'(\psi,\theta,\phi,t) &= \frac{1}{N'} P(\psi,\theta,\phi,t)\epsilon(\psi,\theta,\phi), \\ \bar{P}'(\psi,\theta,\phi,t) &= \frac{1}{N'} \bar{P}(\psi,\theta,\phi,t)\epsilon(\psi,\theta,\phi). \end{aligned} \tag{5.13}$$

We parameterize the detector efficiency using a set of basis functions in three dimensions in the ranges $0 < \psi < \pi$, $0 < \theta < \pi$, and $0 < \phi < 2\pi$:

$$\epsilon(\psi,\theta,\phi) = \sum_{lmk} a_{lm}^k P_k(\cos\psi) Y_{lm}(\theta,\phi). \tag{5.14}$$

The orthonormality of these basis functions can be used to derive an analytic expression for the normalization constant N' (see details in APPENDIX B):

$$
\begin{aligned}
N' &= \frac{2\sqrt{\pi}}{3}\left[\frac{4a^0_{00}}{3}(|A_0|^2n_1+|A_\parallel|^2n_1+|A_\perp|^2n_2)+\frac{4a^2_{00}}{15}(2|A_0|^2n_1-|A_\parallel|^2n_1-|A_\perp|^2n_2)\right]\\
&+ \frac{2\sqrt{\pi}}{3\sqrt{5}}\left[\frac{2a^0_{20}}{3}(|A_0|^2n_1+|A_\parallel|^2n_1-2|A_\perp|^2n_2)+\frac{4a^2_{20}}{15}(|A_0|^2n_1-\frac{1}{2}|A_\parallel|^2n_1+|A_\perp|^2n_2)\right]\\
&- \frac{\sqrt{\pi}}{\sqrt{15}}\left[(A^*_\parallel A_\perp+A_\parallel A^*_\perp)n_3(\frac{4}{3}a^0_{2-1}-\frac{4}{15}a^2_{2-1})\right]\\
&+ \frac{\sqrt{2\pi}}{\sqrt{15}}\left[(A^*_0 A_\perp+A_0 A^*_\perp)n_3(\frac{\pi a^1_{21}}{8}-\frac{\pi a^3_{21}}{32}+...)\right]\\
&+ \frac{2\sqrt{\pi}}{\sqrt{15}}\left[\frac{2a^0_{22}}{3}(-|A_0|^2+|A_\parallel|^2)n_1-\frac{4a^2_{22}}{15}(|A_0|^2+\frac{1}{2}|A_\parallel|^2)n_1\right]\\
&+ \frac{\sqrt{2\pi}}{\sqrt{15}}\left[(A^*_0 A_\parallel+A_0 A^*_\parallel)n_1(\frac{\pi a^1_{2-2}}{8}-\frac{\pi a^3_{2-2}}{32}+...)\right],
\end{aligned}
$$

which, while not particularly lovely on the face of it, represents an analytic normalization of the fitting function and provides an efficient way to compute the likelihood during a maximum log likelihood fit.

The detector efficiency in the above expression enters through the coefficients a^k_{lm}. To fit for these efficiency coefficients, a phase space decay model is used to generate Monte Carlo events. A sample of 100 million $B^0_s \to J/\psi\,\phi$ events is obtained according to the prescription in section 4.5. Since the angular distribution should be flat for the Monte Carlo events, after the reconstruction procedure (the same as applied to the data), the distributions reflect the efficiencies. Projections of these distributions after reconstruction and all analysis selections are shown in Figure 33.

In order to obtain the coefficients a^k_{lm} in Eq. 5.14, the angles are re-binned in each dimension in twenty bins. Then, the three-dimensional histogram is fitted with an expansion of real spherical harmonics for the (θ,ϕ) angles, where each term of the spherical harmonic series is further expanded as a function of Legendre polynomials used to fit the ϕ angle. The fit to (θ,ϕ) is expanded in terms of real spherical harmonics according to the Laplace series

$$
f(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=0}^{l}\left[C_{lm}\cos(m\phi)+S_{lm}\sin(m\phi)\right]P^m_l(\cos\theta). \tag{5.15}
$$

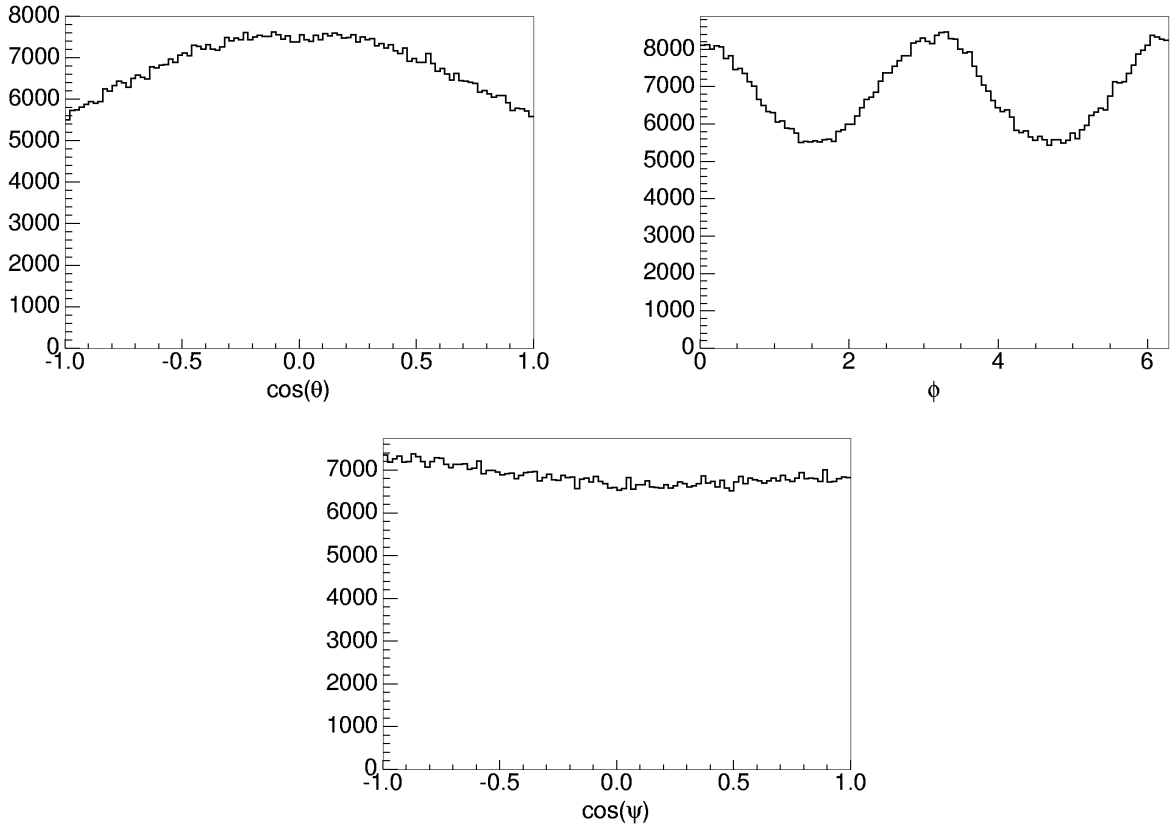

Figure 33: Angular distributions of Monte Carlo $B_s^0 \to J/\psi\,\phi$ phase space events. The non-flat distributions indicate the detector sculpting effect.

It is noted that this series can be easily related to Eq. B.4 by

$$Y_{lm} = \begin{cases} Y_l^0 & \text{if } m = 0 \\ \frac{1}{\sqrt{2}}\left(Y_l^m + (-1)^m Y_l^{-m}\right) = \sqrt{2} C_{lm} \cos(m\phi) P_l^m(\cos\theta) & \text{if } m > 0 \\ \frac{1}{i\sqrt{2}}\left(Y_l^{|m|} - (-1)^{|m|} Y_l^{-|m|}\right) = \sqrt{2} S_{l|m|} \sin(|m|\phi) P_l^m(\cos\theta) & \text{if } m < 0 \end{cases}, \quad (5.16)$$

where the coefficients C_{lm} and S_{lm} are expanded as orthogonal Legendre polynomials functions such that

$$C_{lm} = \sum_{k=0}^{\infty} C_{lm}^k \sqrt{\frac{(2k+1)}{2}} P_k(\cos\psi), \quad (5.17)$$

$$S_{lm} = \sum_{k=0}^{\infty} S_{lm}^k \sqrt{\frac{(2k+1)}{2}} P_k(\cos\psi), \quad (5.18)$$

where C_{lm}^k and S_{lm}^k are the final coefficients that can be easily transferred to a_{lm}^k as in Eq. 5.14.

A two dimensional fit to the (θ, ϕ) distribution and the resulting residuals for all $\cos\psi$ is shown in Figure 80 in APPENDIX D. The one-dimensional fit projections for the three angles are shown in Figure 81. The full three-dimensional fit is difficult to present graphically, but the fit values are listed in Table 23 in APPENDIX D.

5.1.2 Modeling the Background Angular Distributions

For the purpose of constructing the likelihood function to fit data, the background angular distributions should also be modeled. The background distributions can be determined from the mass sidebands of the $B_s^0 \to J/\psi\phi$ decay. To enhance the number of events, the left and right side sideband mass regions are chosen to be be 6 times wider than the Gaussian width of the mass fit, $\sigma = 8.6$ MeV/c^2. Each sideband is 3σ away from the signal region. The signal region is defined as $\pm 3\sigma$ from the fitted mass value $M(B_s^0) = 5.366$ GeV/c^2.

The parameterization of the function used to fit the background angular distribution can be simplified if the three angular angles can be treated independently, assuming no correlations among them. The probability density function can be written as:

$$P_{bkg}(\cos\theta, \phi, \cos\psi) = P_{bkg}(\cos\theta) \cdot P_{bkg}(\phi) \cdot P_{bkg}(\cos\psi).$$

This assumption is tested by calculating the correlation coefficients between any two angular variables. The correlation coefficient for two random variables X and Y with expected values μ_x and μ_y and standard deviations σ_x and σ_y is defined as:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y}, \tag{5.19}$$

where *cov* means covariance, and E is the expected value operator. With this definition, the coefficients are calculated to be:

$$\rho(\theta,\phi) = 0.5\%,\ \ \rho(\theta,\psi) = 0.3\%,\ \ \rho(\phi,\psi) = 2.5\%,$$

which are all negligible.

The independence of the three angles with respect to each other is investigated further by plotting the distributions of one angular variable while another variable is in different ranges. For example, we plot the distribution of $\cos\theta$, while ϕ is in (0, $2\pi/3$), ($2\pi/3$,$4\pi/3$) and ($4\pi/3$, 2π) respectively. We also plot the the distribution of $\cos\theta$, while $\cos\psi$ is in (-1, -0.33), (-0.33, 0.33) and (0.33, 1) respectively. If $\cos\theta$ is independent of the other two angular variables, we should see similar distributions. Similar comparisons for the ϕ and $\cos\psi$ are also implemented. As shown in Figures 82 - 84 in APPENDIX D, these angular variables are independent of each other.

The probability density function for each angular variable is now parameterized independently with some empirical functions in the following forms:

$$P_{bkg}(\cos\theta) = \frac{1 - a_1\cos^2\theta + a_2\cos^4\theta}{2 - 2a_1/3 + 2a_2/5}, \tag{5.20}$$

$$P_{bkg}(\phi) = \frac{1 + b_1\cos(2\phi) + b_2\cos^2(2\phi)}{\pi(2+b_2)}, \tag{5.21}$$

$$P_{bkg}(\cos\psi) = \frac{1 + c_1\cos^2\psi + c_2\cos^4\psi}{2 + 2c_1/3 + 2c_2/5}. \tag{5.22}$$

Figure 34 shows the background angular distributions of $\cos\theta$, ϕ, and $\cos\psi$ with fit projections.

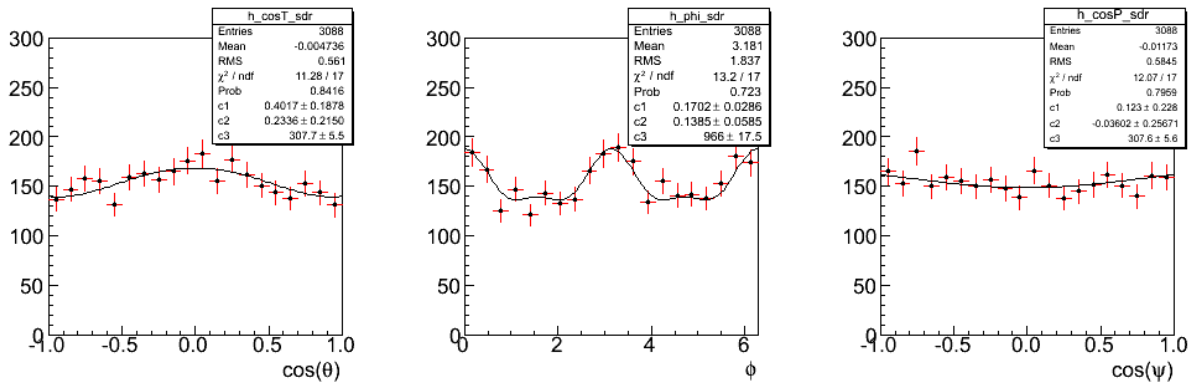

Figure 34: Angular distributions of background events from the sideband region.

It is interesting to note the following: If the angular distribution of the background events can be assumed to be flat before detector acceptance effects, the observed background angular distributions (after detector acceptance effects) should be very similar to the acceptance distributions determined from MC studies (Figure 33). This does seem to be the case, although the statistics in the background angular distributions is quite limited.

5.2 FLAVOR IDENTIFICATION AT PRODUCTION TIME

A crucial input to this analysis is to determine the flavor of the neutral B meson at production time in order to separate the time development of B_s^0 from that of $\bar{B}_s^0$. At the Tevatron, $b\bar{b}$ pairs are generally produced in association, and two independent types of flavor identification algorithms, i.e. flavor tagging, can be used to exploit the specific features of the production of each of the b quarks (see Figure 35). The first type of flavor tagging infers the production flavor of the B_s^0 or $\bar{B}_s^0$ meson from the decay products of the b hadron produced by the other b quark in the event. This is known as the opposite side flavor tagging (OST). The second type of flavor tagging identifies the flavor of the reconstructed B_s^0 or $\bar{B}_s^0$ at production by correlating it with the charge of an associated kaon arising from fragmentation processes. This is usually referred to as the same side kaon tagging (SSKT).

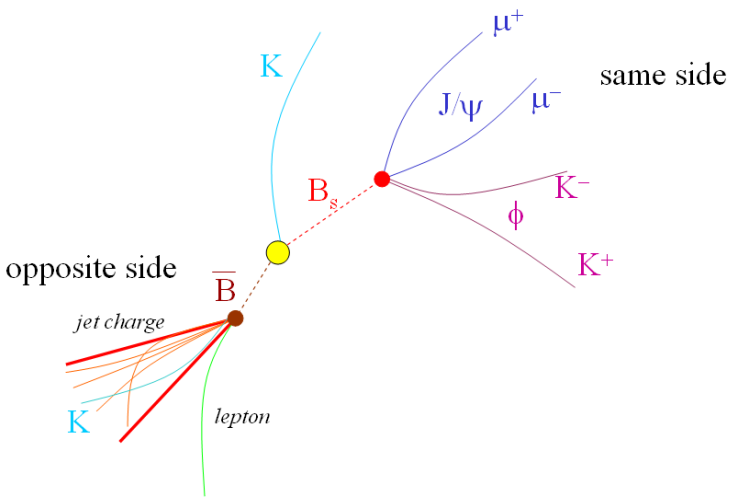

Figure 35: Two types of flavor tagging algorithms used at CDF: OST and SST.

5.2.1 Flavor Tagging Principle

The OST used in this analysis is actually a combination of three different OST algorithms:

- The semileptonic muon tagging (SMT) [53] infers the B_s^0 flavor by identifying the muon from semileptonic decay of the B hadron on the opposite side, and, hence, the flavor of the other b quark through the muon charge.
- The semileptonic electron tagging (SET) [54] infers the flavor of the other b quark by identifying the electron from semileptonic decay of the B hadron.
- The jet charge tagging (JQT) [55] infers the flavor of the other b quark from the p_T weighted charge of all the tracks inside the b jet on the opposite side.

However, none of the above tagging algorithms are fully efficient, nor do they give perfect tagging decisions. Due to low branching ratios, tracking efficiency, and detector acceptance, only a fraction of opposite side b quarks can be associated with leptons from semileptonic decays. If the b quark hadronizes into a neutral B meson, the probability of mixing ($\sim 18\%$ for B^0, $\sim 50\%$ for B_s^0) causes wrong flavor identification. Also, leptons from the charm quark in the sequential decay of a b quark gives wrong flavor identification and, thus, dilutes SMT

and SET tagging power. For the jet charge tagging algorithm, the main challenge is how to identify the b jet. Taking advantage of the long lifetime of b hadrons, one can identify the b jet by requiring the existence of a secondary vertex inside the jet. This type of jet charge tagging algorithm is known as JQT1. If only displaced tracks are identified in the jet, the jet is also a good candidate, and the type is called JQT2. The rest of the situations can be treated as a JQT3 type which has neither secondary vertices nor displaced tracks. It is easy to understand that JQT3 has a larger efficiency but a lower tagging power.

The three opposite side tagging algorithms are combined using the following hierarchical procedure. If an event is tagged by the SMT and the muon candidate satisfies quality cuts, the SMT is used. If there is no SMT, but an SET satisfying quality requirements for the electron, the SET is used. If there is no lepton tagging available, then the JQT is used in the following order: JQT1, JQT2, and JQT3. This ordering is determined by the expected efficiency and tagging power of the tags. Lepton tags have lower efficiencies but higher power. Jet charge tags have larger efficiencies but lower power.

Same side tagging (SST) depends on the properties of the reconstructed meson. As illustrated in Figure 36, the leading particle during the fragmentation associated with a B^+ is a π^-, for a B^0 it is a π^+. For B_s^0 the leading fragmentation particle is a K^+, and the tagging algorithm is also usually referred to the same side kaon tagging (SSKT) [56]. So, by identifying the charge of kaon, one can identify the flavor of the B_s^0 mesons. SSKT, which uses particle identification information, has a higher tagging power than OST. However, the tagging efficiency is only around 50% at CDF because the positive kaon from fragmentation is not always available. For example a neutral K^{*0} could accompany the B_s^0 meson; in such events SSKT has no tagging power. On the other hand, the misidentification of the kaon (e.g. a pion from the K^{*0}) will also give a wrong flavor tagging decision, and the SSKT tagging power will be diluted.

Both OST and SSKT assign two quantities to each reconstructed event:

- A decision ξ which gives the flavor of the meson. $\xi = +1$ identifies B_s^0; $\xi = -1$ identifies $\bar{B}_s^0$; $\xi = 0$ means the tagging algorithm does not give a reasonable decision. The fraction of tagged events ($|\xi| = 1$) is called the efficiency of the flavor tagging algorithm expressed as ϵ.

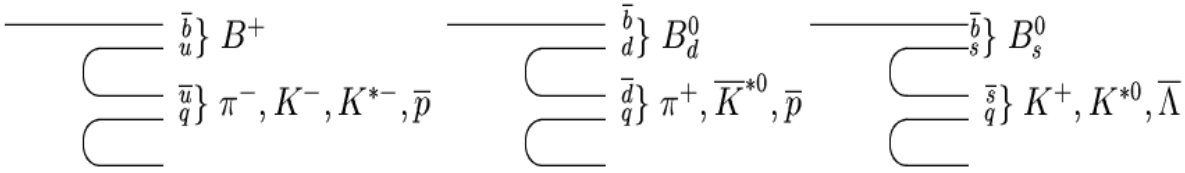

Figure 36: Illustration of SST due to fragmentation processes.

- A predicted dilution $\mathcal{D}$ which quantifies the quality of the decision. $P_{tag} = (1 + \mathcal{D})/2$ gives the correct tagging probability.

The distributions of predicted dilution for both tagging algorithms are shown in Figure 37 for tagged events ($|\xi| = 1$). In these plots the background consists of events from the sideband region of the mass plots in Figure 31, while the signal consists of events from background-subtracted signal region.

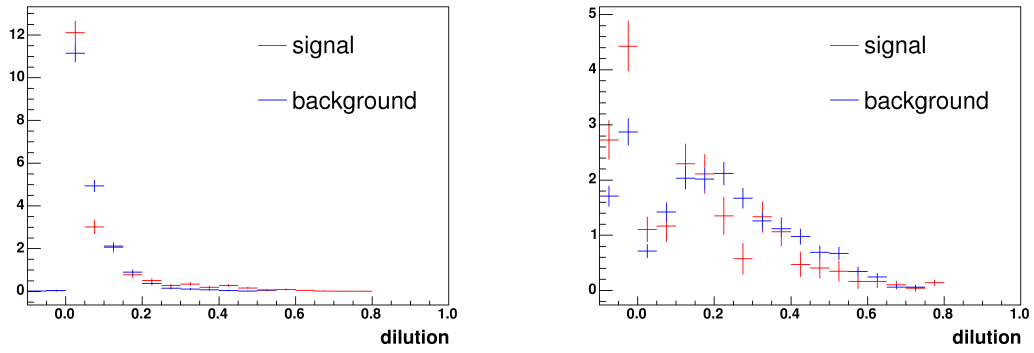

Figure 37: Normalized predicted dilution distributions for the combined OST (left) and SSKT (right) in B^0_s sample. Sideband subtracted signal distributions are shown in red while the background distributions from sidebands are shown in blue.

5.2.2 Tagging Calibration

Since opposite side tagging does not depend on the identity of the B meson reconstructed on the near side, it can be checked and calibrated on the high statistics $B^+ \to J/\psi\, K^+$ sample. In such decays the flavor of the B meson is un-ambiguously determined by the charge of the kaon.

We check that the dilution algorithm correctly predicts its dilution. Since the true flavor of the B^+ is known from the charge of the K^+, one can define the measured (true) dilution as $\mathcal{D}_m = (N^R - N^W)/(N^R + N^W)$ where N^R and N^W are the number of events tagged correctly or incorrectly. In Figures 38 - 40, we plot measured dilution versus predicted dilution for real B^+ decay, real B^- decay, and B^+ and B^- combined, respectively. The fits to these data are superimposed. The relationship between measured dilution and predicted dilution is well modeled by a straight line passing through the origin. When applying these algorithms to $B_s^0 \to J/\psi\phi$ decay, we accommodate the possibility that the dilutions are over-predicted or under-predicted by allowing for either possibility via a collective "scale factor" called the dilution scale factor applied to the dilution. The scale factor is measured using the B^+ sample by the slope of the linear fit. The slopes of the fits are summarized in Table 5. However, the scale factors actually used for B_s^0 are determined through an unbinned maximum likelihood fit (see section 6.1). Since the scale factors for B^+ and B^- agree within statistical uncertainty, the average of the two scale factors is used to re-calibrate the OST when that algorithm is applied to the B_s^0 sample. The final value obtained from the likelihood fit is listed in Table 6.

The measured dilution versus the predicted dilution for background events in the B^+ sample is also investigated. Since the background events should not depend on flavor, one expects that the measured dilution is flat and close to zero over all predicted ranges. Indeed, the expected behavior is observed, as shown in Figures 41 and 42, where plots for B^+ and B^- separately, as well as for the combined B^+ and B^-, are shown. We therefore exclude event-by-event dilution in the background and only fit for the fraction of positively tagged events.

The SSKT for this analysis is effectively calibrated in a previous B_s^0 mixing analysis

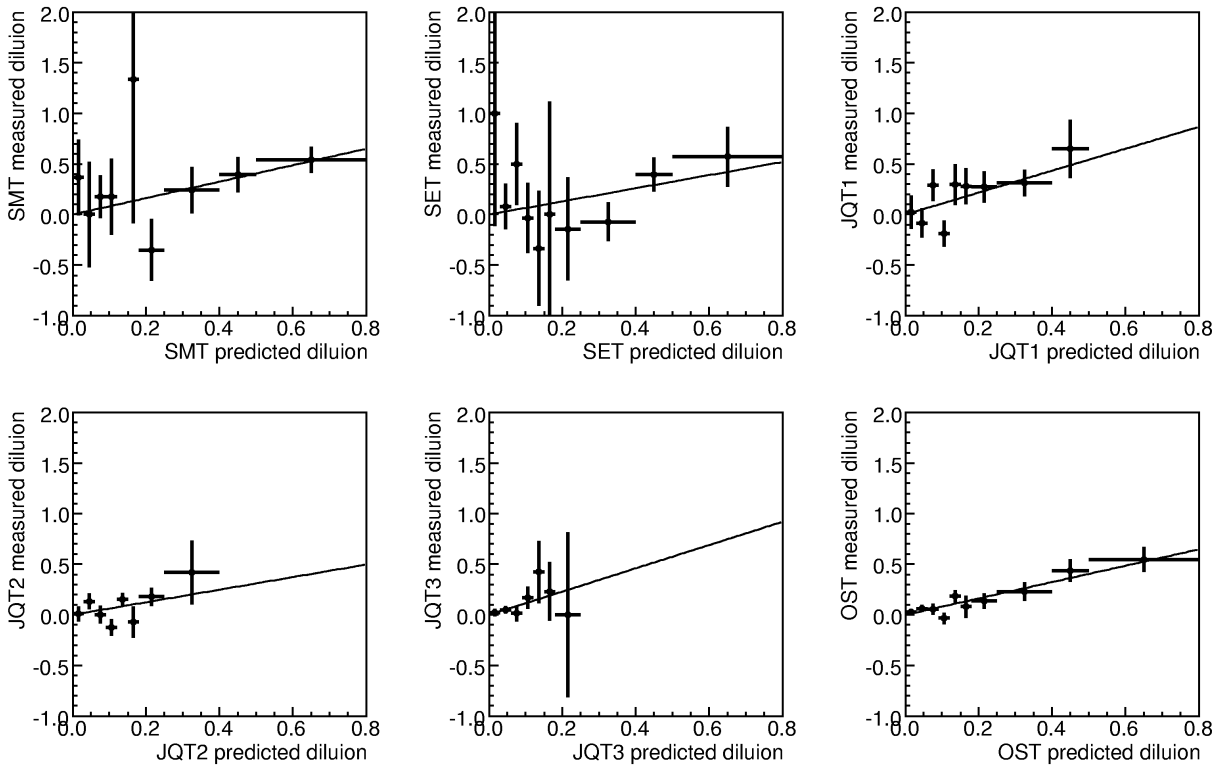

Figure 38: Measured dilution versus the predicted dilution for signal B^+ mesons. Distributions are shown separately for the SMT, SET, JQT1, JQT2, JQT3, and the combined OST.

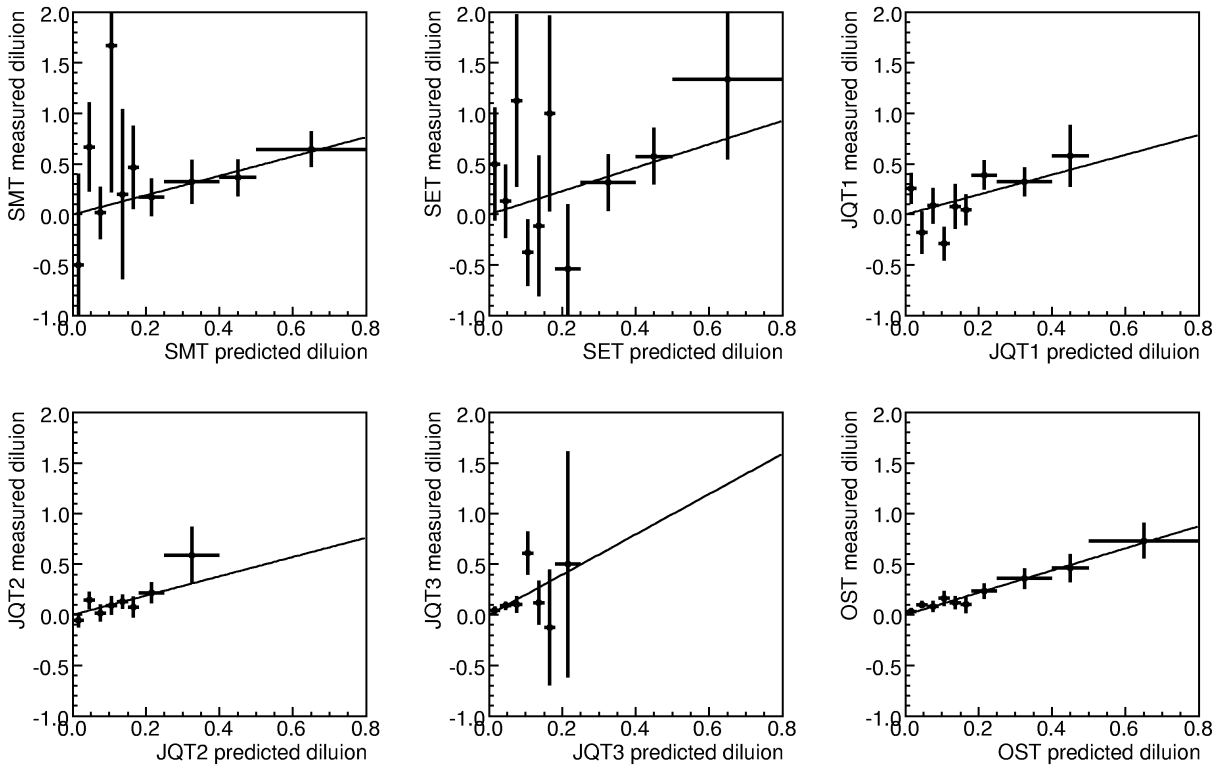

Figure 39: Measured dilution versus the predicted dilution for signal B^- mesons. Distributions are shown separately for the SMT, SET, JQT1, JQT2, JQT3, and the combined OST.

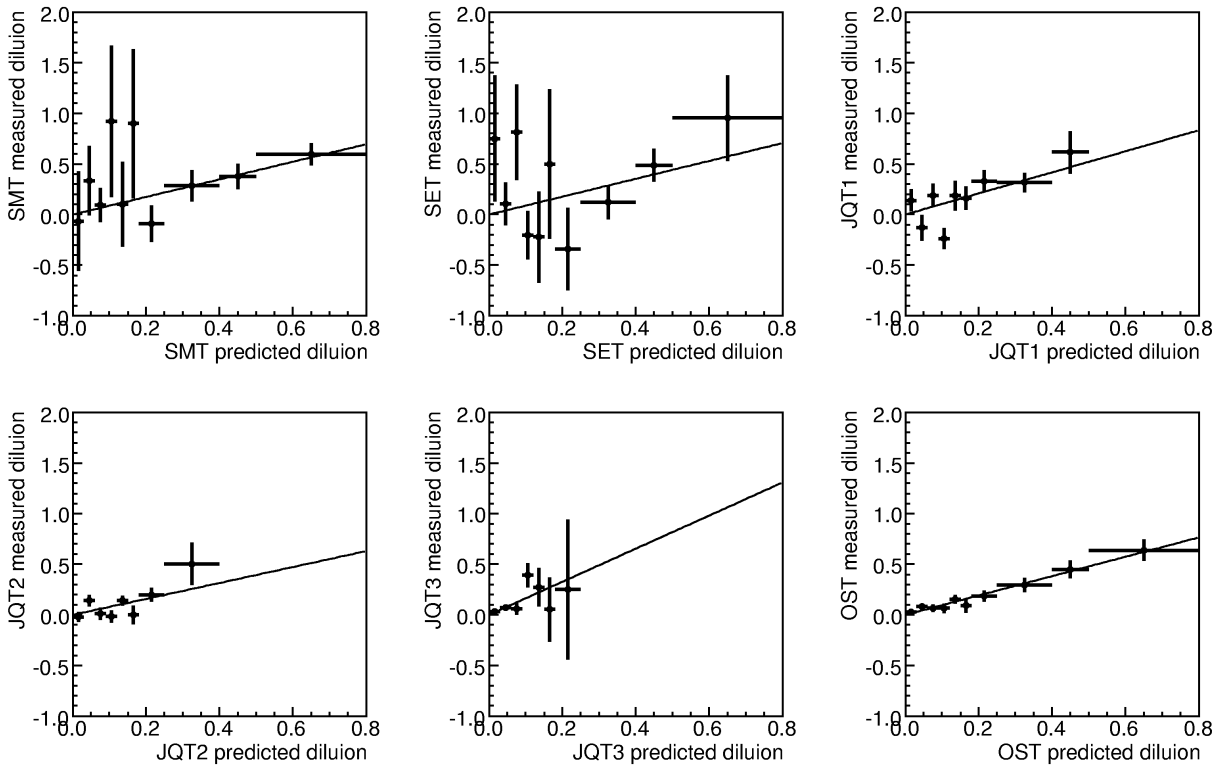

Figure 40: Measured dilution versus the predicted dilution for signal B^+ and B^- mesons. Distributions are shown separately for the SMT, SET, JQT1, JQT2,JQT3, and the combined OST.

	SMT	SET	JQT1	JQT2	JQT3	combined OST
B^+	0.81 ± 0.17	0.64 ± 0.26	1.08 ± 0.29	0.62 ± 0.26	1.14 ± 0.50	0.81 ± 0.11
B^-	0.96 ± 0.21	1.16 ± 0.45	0.98 ± 0.29	0.95 ± 0.27	1.99 ± 0.57	1.10 ± 0.13
$B^{+/-}$	0.87 ± 0.14	0.88 ± 0.26	1.04 ± 0.20	0.79 ± 0.19	1.63 ± 0.40	0.96 ± 0.09

Table 5: Fitted slopes (dilution scale factors) of the measured versus predicted dilution as measured in the B^+ sample.

at CDF [28]. In that analysis, the SSKT calibration and systematic errors were carefully studied using Monte Carlo methods described in Ref. [56]. SSKT dilution scale factors were obtained for different CDF data periods and they are listed in Table 6. A brief description of the mixing measurement follows since it is relevant to the present analysis in many ways.

The probability for a B_s^0 meson to stay the same or oscillate as a function of time is proportional to

$$P(t)_{B_s^0 \to B_s^0, \bar{B}_s^0} \propto (1 \pm \mathcal{D} \cos \Delta m_s t). \tag{5.23}$$

One can introduce an amplitude A as in the following [57]

$$P(t)_{B_s^0 \to B_s^0, \bar{B}_s^0} \propto (1 \pm A\mathcal{D} \cos \Delta m_s t). \tag{5.24}$$

If one fixes Δm_s to different probe values, one can fit for the amplitude at each point. If the same side tagging is correctly calibrated, the amplitude will be consistent with unity when Δm_s is close to the true value. As shown in Figure 43, the amplitude is consistent with unity when Δm_s is fixed to 17.75 ps^{-1}. This value is consistent with the final value obtained through a likelihood fit in the time domain, which is 17.77 ps$^{-1} \pm 0.10(stat) \pm 0.07(syst)$ [28].

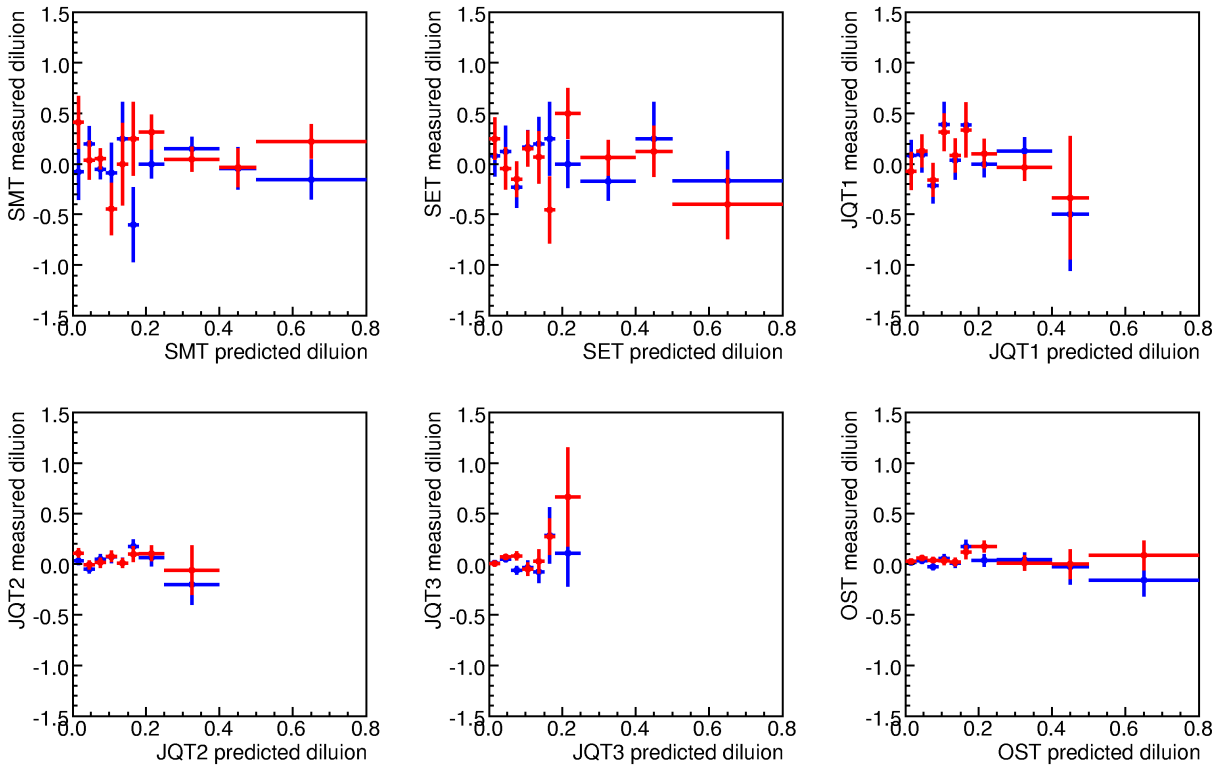

Figure 41: Measured dilution versus the predicted dilution for background B^+ (blue) and B^- (red) mesons separately. Distributions are shown separately for the SMT, SET, JQT1, JQT2, JQT3, and the combined OST.

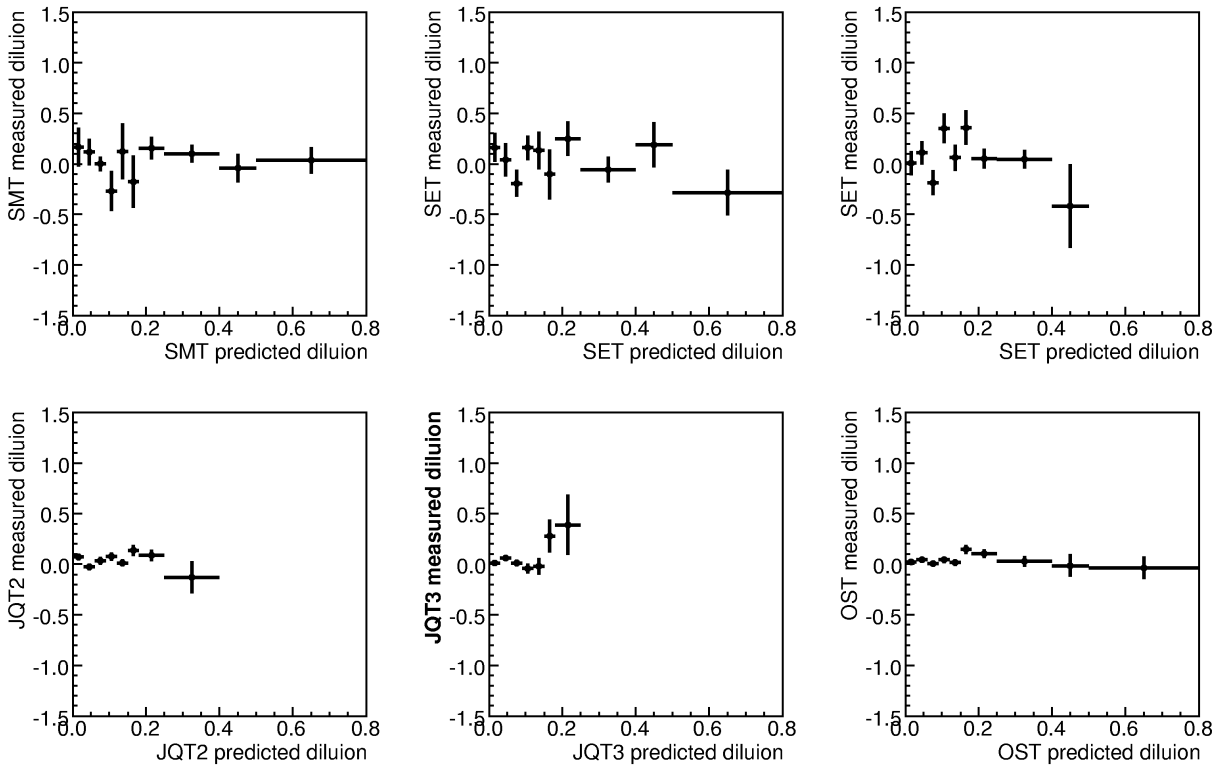

Figure 42: Measured dilution versus the predicted dilution for background B^+ and B^- mesons combined. Distributions are shown separately for the SMT, SET, JQT1, JQT2, JQT3, and the combined OST.

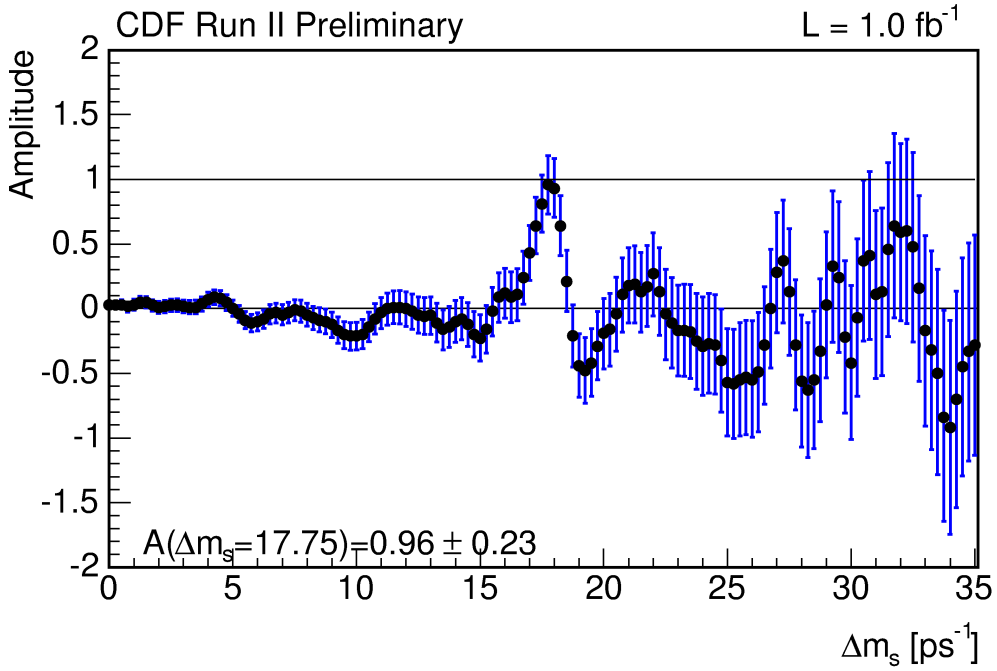

Figure 43: B_s^0 oscillation amplitude scan plot with same side kaon flavor tagging only from Ref. [58]. The amplitude is consistent with unity when $\Delta m_s = 17.75$ ps^{-1}.

	OST	SSKT (0d)	SSKT (0h)	SSKT (0i)
$S_{\mathcal{D}}$	$0.89 \pm 0.08(stat + syst)$	$0.992^{+0.107}_{-0.143}(stat + syst)$	$0.959^{+0.108}_{-0.144}(stat + syst)$	$0.950^{+0.108}_{-0.144}(stat + syst)$

Table 6: Dilution scale factors for OST and SSKT. $0d$, $0h$ and $0i$ are different data periods at CDF.

5.2.3 Tagging Performance

The tagging power of each tagging algorithm is quantified using the tagging efficiency ϵ as well as the average dilution defined as:

$$\epsilon_s = \frac{N(|\xi| = 1)_{sig} - S \cdot N(|\xi| = 1)_{sb}}{N_{sig} - SN_{sb}}, \tag{5.25}$$

$$\langle \mathcal{D} \rangle_s = \sqrt{\frac{\sum_{i,sig}^{N_{sig}} \mathcal{D}_i^2 - S \sum_{i,sb}^{N_{sb}} \mathcal{D}_i^2}{N_{sig} - SN_{sb}}}, \tag{5.26}$$

where sig means summing over all the events in the defined mass signal region, sb means summing over all the events in the defined mass sideband regions, and S is the ratio of the number of background events in the signal region over the number of background events in the sideband region. By choosing equal area, S is simply 1.0. Only efficiency and the average dilution for the signal events are used to determine the tagging power, but the efficiency for the background events is also needed for modeling the background. The efficiency and the average dilution for combined OST and SSKT of the B_s^0 sample are shown in Table 7. The efficiency obtained here comes from counting data directly, but it will be re-determined later in the actual likelihood fit.

From Table 7 the combined OST efficiency is $\sim 97\%$ with an average predicted dilution of $\sim$ 12%. The SSKT efficiency is lower ($\sim$ 51%), but the average predicted dilution is $\sim 26\%$. The total tagging power is calculated to be

$$\epsilon D^2 = \epsilon_{OST} D_{OST}^2 + \epsilon_{SSKT} D_{SSKT}^2 \sim 4.8\%.$$

	OST	SSKT
$\epsilon_{sig}(\%)$	96.5 ± 5.1	50.5 ± 3.5
$\langle \mathcal{D}_{sig} \rangle (\%)$	13.3 ± 0.3	26.6 ± 0.8
$S_D \langle \mathcal{D}_{sig} \rangle (\%)$	11.8 ± 1.1	25.7 ± 3.0

Table 7: OST and SST efficiency and the average predicted dilution for the signal in the B_s^0 sample. The average dilution has been corrected with a dilution scale factor.

5.3 LIKELIHOOD FUNCTION CONSTRUCTION

An unbinned maximum likelihood fit will be used in chapter 8 to extract parameters of interest from the $J/\psi\phi$ data sample. For a fixed set of data with some underlying probability model (likelihood function), this method can give the values of the parameters that make the data more likely than any other values of the parameters would make them. In chapter 7 we validate the maximum likelihood estimator using Monte Carlo. In this section we construct a likelihood function which describes the data sample and includes both signal and background. The function must also be normalized either analytically or numerically. Our likelihood consists of two components, one for the signal and one for the background, and has the general form

$$\mathcal{L}_i = f_s \cdot P_s(\vec{q}) + (1 - f_s) \cdot P_b(\vec{q}), \tag{5.27}$$

where $\vec{q}$ represents all the variables describing an event in the sample; f_s is the fraction of signal events in the sample, and P_s and P_b are normalized probability density functions for signal and background, respectively. For the $B_s^0 \to J/\psi\phi$ decay, $\vec{q}$ includes: the angular variables $\cos\psi$, $\cos\theta$, and ϕ, the proper decay time t, the proper decay time uncertainty σ_t, the flavor tagging decision ξ and dilution $\mathcal{D}$, the mass m of the meson, and the mass uncertainty σ_m. The construction of the likelihood function will be discussed in three subsections. First we describe the likelihood function used for the signal only, then we describe the likelihood function used for the background only, and finally the full likelihood function will be given.

5.3.1 Likelihood Function for the Signal

The likelihood function for the signal is based on formulas derived in chapter 2, and the main result is shown in Eq. 5.13 and Eq. 5.14. The decay probability density function for B_s^0 is expressed as $P'(\psi, \theta, \phi, t)$, while for $\bar{B}_s^0$ it is described by $\bar{P}'(\psi, \theta, \phi, t)$. To accommodate the the vertex resolution from B_s^0 decay reconstruction, which is propagated to the proper decay time ($\sigma_{ct} \sim 25\mu m$), the decay probability function is convoluted with a resolution function. The resolution function is modeled with a single component Gaussian function. The smeared decay probability functions for both B_s^0 and $\bar{B}_s^0$ now look like:

$$\begin{aligned} P(\psi, \theta, \phi, t|\sigma_t) &= P'(\psi, \theta, \phi, t) \otimes G(\sigma_t), \\ \bar{P}(\psi, \theta, \phi, t|\sigma_t) &= \bar{P}'(\psi, \theta, \phi, t) \otimes G(\sigma_t). \end{aligned} \tag{5.28}$$

With flavor tagging, we have a tagging decision ξ and some predicted dilution $\mathcal{D}$ for each event. The tagging decision variable ξ is a discrete variable that can only take values $-1, 0,$ or 1. If $\xi = -1$, the initial state is determined to be $\bar{B}_s^0$, if $\xi = 1$, the initial state is determined to be B_s^0, and if the tagging variable is zero, the initial state is undermined. $(1 + \mathcal{D})/2$ gives the probability for a right decision. For untagged events ($\xi = 0$), $\mathcal{D}$ is simply set to zero. If an event has a positive tag, it means the event is a B_s^0 candidate with probability $(1 + \mathcal{D})/2$, or a $\bar{B}_s^0$ candidate with probability $(1 - \mathcal{D})/2$. An event with tagging decision $\xi = +1$ can then be described with a probability density function

$$\begin{aligned} P_s(\psi, \theta, \phi, t|\sigma_t, \xi = +1, \mathcal{D}) &= P^{RS} \cdot P_{B_s^0} + P^{WS} \cdot P_{\bar{B}_s^0} \\ &= \frac{1 + \mathcal{D}}{2} \cdot P(\psi, \theta, \phi, t|\sigma_t) + \frac{1 - \mathcal{D}}{2} \cdot \bar{P}(\psi, \theta, \phi, t|\sigma_t), \end{aligned} \tag{5.29}$$

where "RS" means right sign, "WS" means wrong sign. A more general expression with an arbitrary tagging decision is

$$P_s(\psi, \theta, \phi, t|\sigma_t, \xi, \mathcal{D}) = \frac{1 + \xi\mathcal{D}}{1 + |\xi|} \cdot P(\psi, \theta, \phi, t|\sigma_t) + \frac{1 - \xi\mathcal{D}}{1 + |\xi|} \cdot \bar{P}(\psi, \theta, \phi, t|\sigma_t). \tag{5.30}$$

Since we have two independent taggers, the likelihood function is actually written as

$$\begin{aligned} P_s(\psi, \theta, \phi, t|\sigma_t, \vec{\xi}, \vec{\mathcal{D}}) &= \frac{1 + \xi_1\mathcal{D}_1}{1 + |\xi_1|} \frac{1 + \xi_2\mathcal{D}_2}{1 + |\xi_2|} \cdot P(\psi, \theta, \phi, t|\sigma_t) \\ &\quad + \frac{1 - \xi_1\mathcal{D}_1}{1 + |\xi_1|} \frac{1 - \xi_2\mathcal{D}_2}{1 + |\xi_2|} \cdot \bar{P}(\psi, \theta, \phi, t|\sigma_t). \end{aligned} \tag{5.31}$$

However, this is only a conditional probability density function since it depends on an event-by-event σ_t, ξ and $\mathcal{D}$. The full signal probability density function should be written in this way

$$P_s(\psi,\theta,\phi,t,\sigma_t,\vec{\xi},\vec{\mathcal{D}}) = P_s(\psi,\theta,\phi,t|\sigma_t,\vec{\xi},\vec{\mathcal{D}}) \cdot P_s(\sigma_t) \prod_{i=1}^{2} P_s(\xi_i) \cdot P_s(\mathcal{D}_i), \tag{5.32}$$

where $P_s(\sigma_t)$ describes the σ_t distribution for signal events, $P_s(\mathcal{D}_i)$ describes the dilution distribution of signal events for each tagger, while $P_s(\xi_i)$ has the following form:

$$P(\xi_i) = (1-\epsilon_i)\cdot\delta(\xi_i - 0) + \epsilon_i \cdot \delta(|\xi_i| - 1), \tag{5.33}$$

where ϵ_i is the tagging efficiency of each tagging algorithm. With two taggers, we can write $P_s(\vec{\xi}) \equiv P_s(\xi_1)P_s(\xi_2)$ explicitly as

$$P_s(\vec{\xi}) = \begin{cases} (1-\epsilon_1)(1-\epsilon_2) & (\xi_1 = 0, \xi_2 = 0) \\ \epsilon_1(1-\epsilon_2) & (\xi_1 = \pm 1, \xi_2 = 0) \\ (1-\epsilon_1)\epsilon_2 & (\xi_1 = 0, \xi_2 = \pm 1) \\ \epsilon_1\epsilon_2 & (\xi_1 = \pm 1, \xi_2 = \pm 1) \end{cases} . \tag{5.34}$$

5.3.2 Likelihood Function for the Background

The likelihood function for the background is constructed in an analogous fashion to that of the signal, with the same proper time resolution effects included. However, the background model does not depend on the event-by-event tagging decision and dilution. Only the fraction of positive tagged events is a necessary part of the background model. For the background angular distributions, each angle is modeled independently, since they do not have correlations among them. The general form of the conditional background probability density function expression is:

$$P_b(\psi,\theta,\phi,t|\sigma_t,\vec{\xi},\vec{\mathcal{D}}) = \begin{cases} P_b(t|\sigma_t)\cdot P_b(\psi)\cdot P_b(\theta)\cdot P_b(\phi) & (\xi = 0) \\ \epsilon_{bg}^{+}\cdot P_b(t|\sigma_t)\cdot P_b(\psi)\cdot P_b(\theta)\cdot P_b(\phi) & (\xi = +1) \\ (1-\epsilon_{bg}^{+})\cdot P_b(t|\sigma_t)\cdot P_b(\psi)\cdot P_b(\theta)\cdot P_b(\phi) & (\xi = -1) \end{cases} . \tag{5.35}$$

Proper decay time for the background $P_b(t|\sigma_t)$: The modeling is the same as the CDF lifetime measurement analysis [59] which includes one prompt peak (δ function), two positive exponential tails, and one negative exponential tail. All the components are convoluted with the same resolution function used for signal:

$$P_b(t|\sigma_t) = \left(f_g + (1-f_g)\left(f_{++}\frac{1}{\lambda_{++}}e^{-\frac{t}{\lambda_{++}}} + (1-f_{++})\left(f_-\frac{1}{\lambda_-}e^{\frac{t}{\lambda_-}} + (1-f_-)\frac{1}{\lambda_+}e^{-\frac{t}{\lambda_+}}\right)\right)\right) \otimes G(\sigma_t). \quad (5.36)$$

Angular distributions for the background $P_b(\psi), P_b(\theta), P_b(\phi)$: The distributions of angular variables are parameterized from the sidebands of the $B_s^0 \to J/\psi\,\phi$ (see section 5.1.2) with all parameters floating in the final fit.

The full probability density function for the background is then

$$P_b(\psi,\theta,\phi,t,\sigma_t,\vec{\xi},\vec{\mathcal{D}}) = P_b(\psi,\theta,\phi,t|\sigma_t,\vec{\xi},\vec{\mathcal{D}})\cdot P_b(\sigma_t)\prod_{i=1}^{2} P_b(\xi_i)\cdot P_b(\mathcal{D}_i), \quad (5.37)$$

where $P_b(\sigma_t)$ describes the σ_t distribution for background events, $P_b(\mathcal{D}_i)$ describes the background dilution distribution for each tagger, while $P_b(\xi_i)$ has same form as Eq. 5.33, but does not necessarily have same tagging efficiency as the signal, so they are fitted separately in the final fit.

5.3.3 Overview of the Full Likelihood

To construct the full likelihood function, probability density functions for mass signal and background are also included. This is important, because the mass distribution in fully reconstructed B_s^0 is crucial in separating signal and background events. The full likelihood including the mass distribution is:

$$\begin{aligned}\mathcal{L}_i &= f_s\cdot P_s(m)\cdot P_s(\psi,\theta,\phi,t,\sigma_t,\vec{\xi},\vec{\mathcal{D}}) + (1-f_s)\cdot P_b(m)\cdot P_b(\psi,\theta,\phi,t,\sigma_t,\vec{\xi},\vec{\mathcal{D}}) \\ &= f_s\cdot P_s(m)P_s(\psi,\theta,\phi,t|\sigma_t,\vec{\xi},\vec{\mathcal{D}})\cdot P_s(\sigma_t)\prod_{i=1}^{2} P_s(\xi_i)\cdot P_s(\mathcal{D}_i) \\ &\quad +(1-f_s)\cdot P_b(m)\cdot P_b(\psi,\theta,\phi,t|\sigma_t,\vec{\xi},\vec{\mathcal{D}})\cdot P_b(\sigma_t)\prod_{i=1}^{2} P_b(\xi_i)\cdot P_b(\mathcal{D}_i). \quad (5.38)\end{aligned}$$

$P_s(\sigma_t)$ **and** $P_b(\sigma_t)$: As described in previous sections, these terms are needed to model the proper decay time uncertainties needed for the full likelihood functions. However, they can be dropped from the likelihood function if the signal and background have similar distributions. As one can see from Figure 44, the distributions are different for signal and background, so they have to be included in the final fit. A sum of two components of the probability density functions of the Gamma distribution are used to model the proper decay time error distribution with the probability density functions defined as

$$f(x;a,b) \equiv \frac{x^a e^{-x/b}}{b^{a+1}\Gamma(a+1)}, \tag{5.39}$$

which is valid only for $x > 0$.

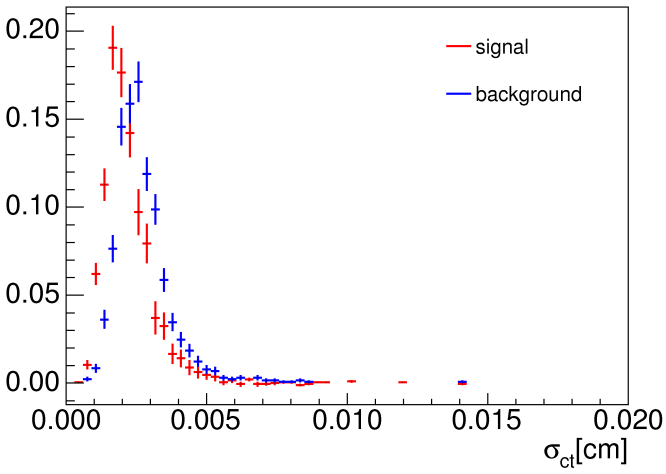

Figure 44: Normalized proper decay time uncertainty distributions for signal and background events.

$P_s(\mathcal{D}_i)$ **and** $P_b(\mathcal{D}_i)$: As can be seen in Figure 37 in chapter 5.2,the distribution are not exactly the same for signal and background. So we use normalized histograms to describe the effect in the likelihood function. This is done by taking the probability density from a histogram. The signal histogram comes from the sideband-subtracted signal region, while for the background it is taken from the sideband region. The histograms are divided into 22 bins between -0.1 and 1.0 for both signal and background events.

Mass Signal $P_s(m)$: The mass distribution for signal events is usually modeled with a single Gaussian or double-Gaussian curve with a universal mass uncertainty or an event-by-event mass uncertainty, depending on which model describes the shape better. In $B_s^0 \to J/\psi\phi$ channel, the signal is modeled with a single Gaussian with event-by-event mass uncertainty. In principle, we should also consider mass uncertainty distributions $P(\sigma_m)$ for both signal and background events, but as already observed in a previous CDF lifetime analysis [59], the distributions are very similar for the signal and the background events, so they are not included in the final likelihood fit.

Mass Background $P_b(m)$: The mass model for the background is modeled with a first order polynomial.

Scale Factor $s_{\sigma_t}, s_{\sigma_m}, s_{\mathcal{D}}$: To account for a collective mis-estimation of time and mass resolutions, a scale factor s_{σ_t} for event-by-event proper decay time uncertainty and s_{σ_m} for event-by-event mass uncertainty are introduced. Both of them are allowed to float in the fit. As discussed in chapter 5.2, to correct the predicted dilution, the event-by-event dilution is adjusted with the scale factor $s_{\mathcal{D}}$ for both tagging algorithms. However, the dilution scale factors are Gaussian-constrained within uncertainties in the final fit.

5.3.4 Invariance Property of the Likelihood Function

The likelihood function introduced in the previous section is invariant under simultaneous transformations of several parameters. The invariance can be seen clearly if one checks the explicit form of the probability density function Eqs. A.11 and A.12 derived in APPENDIX A. It will be useful to write the decay probability density function for B_s^0 and $\bar{B}_s^0$ in the following way

$$\begin{aligned} P(t,\vec{\omega}) \;\propto\; & |A_0|^2\mathcal{T}_+ g_1(\vec{\omega}) + |A_\parallel|^2\mathcal{T}_+ g_2(\vec{\omega}) \\ + \; & |A_\perp|^2\mathcal{T}_- g_3(\vec{\omega}) + |A_0||A_\parallel|\cos(\phi_\parallel)\mathcal{T}_+ g_4(\vec{\omega}) \\ + \; & |A_\parallel||A_\perp|\mathcal{U}_+ g_5(\vec{\omega}) \\ + \; & |A_0||A_\perp|\mathcal{V}_+ g_6(\vec{\omega}), \end{aligned} \tag{5.40}$$

and

$$\begin{aligned}
\bar{P}(t,\vec{\omega}) &\propto |A_0|^2\bar{\mathcal{T}}_+ g_1(\vec{\omega}) + |A_\parallel|^2\bar{\mathcal{T}}_+ g_2(\vec{\omega}) \\
&+ |A_\perp|^2\bar{\mathcal{T}}_- g_3(\vec{\omega}) + |A_0||A_\parallel|\cos(\phi_\parallel)\bar{\mathcal{T}}_+ g_4(\vec{\omega}) \\
&+ |A_\parallel||A_\perp|\mathcal{U}_- g_5(\vec{\omega}) \\
&+ |A_0||A_\perp|\mathcal{V}_- g_6(\vec{\omega}),
\end{aligned} \tag{5.41}$$

where the functions $g_1(\vec{\omega}) \ldots g_6(\vec{\omega})$ are associated with the angular distributions and defined in APPENDIX A. A_0, $A_\perp$ and $A_\parallel$ are decay amplitudes decomposed according to the linear polarizations of J/ψ and ϕ at time zero. The time dependent terms are:

$$\begin{aligned}
\mathcal{T}_\pm = e^{-\Gamma t} &\times \left[\cosh\frac{\Delta\Gamma}{2}t \mp \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t \mp \sin 2\beta_s \sin\Delta m t\right], \\
\bar{\mathcal{T}}_\pm = e^{-\Gamma t} &\times \left[\cosh\frac{\Delta\Gamma}{2}t \mp \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t \pm \sin 2\beta_s \sin\Delta m t\right], \\
\mathcal{U}_\pm = \pm e^{-\Gamma t} &\times \left[\sin(\phi_\perp - \phi_\parallel)\cos(\Delta m t) - \cos(\phi_\perp - \phi_\parallel)\cos(2\beta_s)\sin(\Delta m t)\right. \\
&\pm \left.\cos(\phi_\perp - \phi_\parallel)\sin(2\beta_s)\sinh(\frac{\Delta\Gamma t}{2})\right], \\
\mathcal{V}_\pm = \pm e^{-\Gamma t} &\times \left[\sin(\phi_\perp)\cos(\Delta m t) - \cos(\phi_\perp)\cos(2\beta_s)\sin(\Delta m t)\right. \\
&\pm \left.\cos(\phi_\perp)\sin(2\beta_s)\sinh(\frac{\Delta\Gamma t}{2})\right],
\end{aligned} \tag{5.42}$$

where $\phi_\parallel$ and $\phi_\perp$ are phases of amplitudes $A_\parallel$ and $A_\perp$ with respect to amplitude A_0. The probability density functions are found to be invariant under the following simultaneous transformations

$$\begin{aligned}
\beta_s &\rightarrow \pi/2 - \beta_s, \\
\Delta\Gamma &\rightarrow -\Delta\Gamma, \\
\phi_\parallel &\rightarrow 2\pi - \phi_\parallel, \\
\phi_\perp &\rightarrow \pi - \phi_\perp.
\end{aligned} \tag{5.43}$$

We further consider situation when no flavor tagging information is available. In this case, the likelihood function can be obtained by simply setting tagging decisions and dilutions

to be zero in Eq. 5.38. Terms $P(\xi_i)$ and $P(\mathcal{D}_i)$ also disappear because the likelihood does not depend depend on flavor tagging any more. The likelihood takes the the following form:

$$\begin{aligned}\mathcal{L}_i = \quad & f_s \cdot P_s(m) \cdot P_s(t,\psi,\theta,\phi|\sigma_t) \cdot P_s(\sigma_t) + \\ & (1-f_s) \cdot P_b(m) \cdot P_b(t|\sigma_t) \cdot P_b(\psi) \cdot P_b(\theta) \cdot P_b(\phi) \cdot P_b(\sigma_t),\end{aligned} \tag{5.44}$$

where the signal likelihood is the sum of the smeared decay probabilities of B_s^0 and $\bar{B}_s^0$ as in Eq. 5.28:

$$P_s(t,\psi,\theta,\phi|\sigma_t) = P(t,\psi,\theta,\phi|\sigma_t) + \bar{P}(t,\psi,\theta,\phi|\sigma_t). \tag{5.45}$$

With this likelihood function, one realizes that there are still terms with β_s unlike the B^0 system which has practically no decay width difference. However, the sensitivity of the measurement to β_s is reduced greatly, because of the lack of flavor identification. Furthermore, a four-fold ambiguity arises due to simultaneous transformations:

$$\begin{aligned}\beta_s \quad &\rightarrow \quad -\beta_s \\ \phi_\perp \quad &\rightarrow \quad \pi + \phi_\perp,\end{aligned} \tag{5.46}$$

or

$$\begin{aligned}\beta_s \quad &\rightarrow \quad \pi/2 + \beta_s \\ \Delta\Gamma \quad &\rightarrow \quad -\Delta\Gamma.\end{aligned} \tag{5.47}$$

However, if one assumes CP conservation (i.e. $\beta_s = 0$), the decay probability functions will be simplified. The mass eigenstates are also CP eigenstates with this assumption, and one can measure the lifetime and decay width difference to check standard model predictions. The simplified likelihood function is invariant with the transformation of a single parameter, which is

$$\phi_\parallel \rightarrow 2\pi - \phi_\parallel.$$

For this thesis, this measurement without using flavor tagging information will not be the main result, but will serve as a cross-check.

6.0 OST RE-CALIBRATION USING $B^+ \to J/\psi K^+$ AND CROSSCHECK FROM CP VIOLATION MEASUREMENT IN $B^0 \to J/\psi K^0_S$

All the opposite side tagging algorithms (SMT, SET and JQT) are developed based on lepton + SVT track triggered events collected with a semileptonic trigger (section 4.1). To apply these tagging algorithms to events with fully reconstructed B mesons, the dilution needs to be increased or decreased by a collective scale factor to accommodate potential differences in the tagging dilution arising from kinetic differences in the two trigger streams. As anticipated in section 5.2, the OST dilution scale factor can be extracted, along with other parameters, in an unbinned maximum likelihood fit from the $B^+ \to J/\psi\, K^+$ sample. Since there is no flavor mixing involved in charged B^+ decays, the true flavor is known from the charged kaon. The scale factors for B^+ and B^- are treated separately in the fit to check for a charge dependence of the tagging algorithms. After the scale factor is obtained, it is used in a measurement of $\sin 2\beta$ from the time-dependent CP asymmetry in the $B^0 \to J/\psi K^0_s$ channel as a crosscheck. The final state of $B^0 \to J/\psi K^0_s$ is pure CP odd, and no angular analysis is needed to separate CP eigenstates. Both fits to the $B^0 \to J/\psi K^0_s$ and $B^+ \to J/\psi K^+$ data use likelihood functions similar to, but vastly simpler than, that used for $B^0_s \to J/\psi\phi$ as discussed in the previous chapter.

6.1 OPPOSITE SIDE TAGGING CALIBRATION IN $B^+ \to J/\psi K^+$ CHANNEL

To determine the OST scale factors from $B^+ \to J/\psi\, K^+$, the likelihood function uses the following input variables for each data event:

- q: charge of B^+,
- m: B^+ reconstructed invariant mass,
- t and σ_t: B^+ proper decay time and its error,
- ξ and $\mathcal{D}$: flavor tagging decision and predicted dilution.

The likelihood function is constructed as the following, with the fit parameters as listed in Table 9:

$$\begin{aligned}\mathcal{L}_i = \quad & f_s \cdot P_s(m) \cdot P_s(t|\sigma_t, \xi, \mathcal{D}) \cdot P_s(\sigma_t) \cdot P_s(\xi) \cdot P_s(\mathcal{D}) + \\ & (1 - f_s) \cdot P_b(m) \cdot P_b(t|\sigma_t, \xi, \mathcal{D}) \cdot P_b(\sigma_t) \cdot P_b(\xi) \cdot P_b(\mathcal{D}),\end{aligned} \tag{6.1}$$

where

- $P_s(m)$: model for B^+ mass signal, described by two components of Gaussian function,
- $P_b(m)$: model for B^+ mass background, described by linear polynomial,
- $P_s(\sigma_t)$ and $P_b(\sigma_t)$: distributions of proper decay time uncertainty for signal and background modeled with probability density function of Gamma function,
- $P_s(\xi)$ and $P_b(\xi)$: same as Eq. 5.33,
- $P_s(\mathcal{D})$ and $P_b(\mathcal{D})$: dilution distribution for signal and background modeled with histogram as shown in Figure 45.

The sample can be divided into two parts according to the charge of the B^+ meson. The dilution scale factor for B^+ is fitted as s^+, while the dilution scale factor for B^- is s^-. The time dependent parts of the likelihood function are

$$P_s(t|\sigma_t, \xi, \mathcal{D}) = \begin{cases} \frac{1+s^+\mathcal{D}}{2} \cdot \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes G(\sigma_t) & (q = +1, \xi = +1) \\ \frac{1-s^+\mathcal{D}}{2} \cdot \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes G(\sigma_t) & (q = +1, \xi = -1) \\ \frac{1-s^-\mathcal{D}}{2} \cdot \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes G(\sigma_t) & (q = -1, \xi = +1) \\ \frac{1+s^-\mathcal{D}}{2} \cdot \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes G(\sigma_t) & (q = -1, \xi = -1) \end{cases}, \tag{6.2}$$

and

$$P_b(t|\sigma_t, \xi, \mathcal{D}) = \begin{cases} \left[\delta(\xi) + f_B\delta(\xi - 1) + (1 - f_B)\delta(\xi + 1)\right] P_b(t|\sigma_t) & (q = +1) \\ \left[\delta(\xi) + f_B\delta(\xi + 1) + (1 - f_B)\delta(\xi - 1)\right] P_b(t|\sigma_t) & (q = -1) \end{cases}, \tag{6.3}$$

where δ is the delta function, and $P_b(t|\sigma_t)$ is the same as in Eq. 5.36.

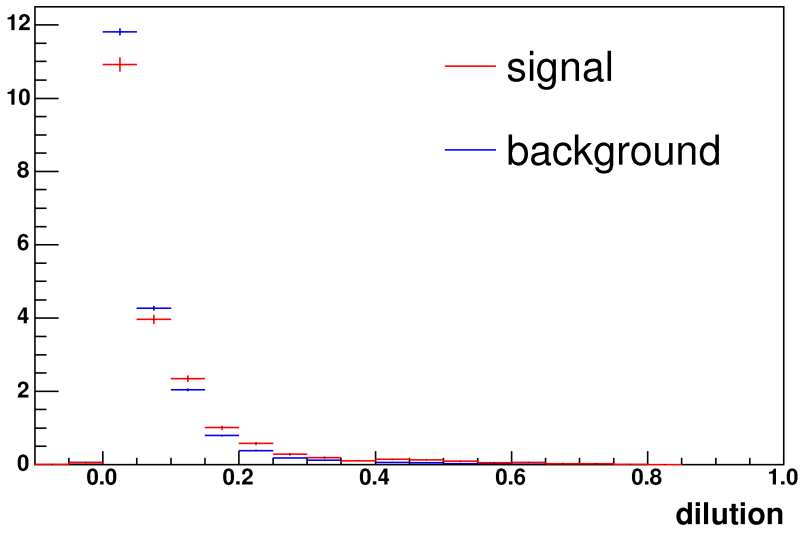

Figure 45: Normalized OST dilution distributions for signal (red) and background (blue) in the $B^+ \to J/\psi\, K^+$ channel.

6.1.1 Fitter Validation for the B^+ sample

The fitter which implements the likelihood function is validated using a toy Monte Carlo technique (pseudo-experiments). A sample of pseudo-experiments are generated according to some input values. Within each pseudo-experiment, we record the value

$$(p_{fit} - p_{input})/\sigma_p$$

for all the parameters of interest. Here p_{input} is the input value for that parameter, p_{fit} is the fitted value for the same parameter, and σ_p is the corresponding uncertainty as returned by the fit. If the distribution of such quantity, which is usually called the "pull" distribution, follows a normal distribution, it is generally a sign of an unbiased fit.

The fitter is tested by running 500 toy MC experiments. Each experiment has 50,000 events, with a signal fraction of ~0.25. The means and widths of the Gaussian fits to the pull distributions of the signal parameters are listed in Table 8. We find that the pull distributions of all the fit parameters are consistent with normal distributions.

Parameter	Mean	Width
$c\tau$	-0.003 ± 0.046	1.034 ± 0.033
f_s	-0.09 ± 0.05	0.98 ± 0.04
s^+	-0.08 ± 0.05	1.02 ± 0.03
s^-	-0.05 ± 0.04	0.97 ± 0.03
ϵ_s	-0.004 ± 0.044	0.991 ± 0.031
m	-0.03 ± 0.04	0.99 ± 0.03

Table 8: Mean and width of Gaussian fits to 500 toy MC pull distributions for the $B^+ \to J/\psi K^+$ channel.

6.1.2 Fit Results

The validated fit function is then applied to the $B^+ \to J/\psi K^+$ data, which are reconstructed using the procedure described in section 4.4.3. The full set of fit parameters is listed in Table 9, along with the fit results. The fitted scale factors, s^+ and s^- are consistent with each other within a one sigma uncertainty. We therefore use the average value as the dilution scale factor

$$s = \frac{s^+ + s^-}{2} = 0.89 \pm 0.05(stat).$$

The systematic uncertainty on the average OST dilution scale factor is determined as follows. The main systematic uncertainty comes from the possible asymmetry between the B^+ and B^- samples. The sample is fitted again with one overall scale factor for both B^+ and B^-, and the difference between this fitted value and the previous average value is treated as systematic uncertainty, which results into an uncertainty of 0.06.

The next source of systematic uncertainty is due to the treatment of the dilution probability distributions for signal and background events using histograms. The systematic uncertainty comes from the fact that the sideband subtracted signal distribution has a few negative bins in the large dilution tails where the statistics is poor. In the default fit, for any event that has negative signal dilution probability, the difference between signal and back-

ground is ignored by setting both $P_s(\mathcal{D})$ and $P_b(\mathcal{D})$ to be one. To evaluate the systematics, only $P_s(\mathcal{D})$ is set to be one, while $P_b(\mathcal{D})$ remains the same. The average OST scale factors for both situations are determined, and we take the difference as a systematic error, which gives 0.02.

The total systematic uncertainty is calculated to be still 0.06, and the average scale factor becomes

$$s = 0.89 \pm 0.08(stat + syst).$$

6.2 MEASUREMENT OF SIN 2β

After the OST scale factor is determined from $B^+ \to J/\psi K^+$ decays, we measure the time dependent asymmetry $\sin 2\beta$ in the $B^0 \to J/\psi\, K^0_s$ channel. The input variables include:

- m and σ_m: B^0 reconstructed mass and its error,
- t and σ_t: B^0 proper decay time and its error,
- ξ and $\mathcal{D}$: flavor tagging decisions and dilution.

The likelihood function is written in the following, with parameters listed in Table 11.

$$\begin{aligned} \mathcal{L}_i = \quad & f_s \cdot P_s(m|\sigma_m) \cdot P_s(t|\sigma_t, \xi, \mathcal{D}) \cdot P_s(\sigma_t) \cdot P_s(\xi) \cdot P_s(\mathcal{D}) + \\ & (1 - f_s) \cdot P_b(m|\sigma_m) \cdot P_b(t|\sigma_t, \xi, \mathcal{D}) \cdot P_b(\sigma_t) \cdot P_b(\xi) \cdot P_b(\mathcal{D}), \end{aligned} \tag{6.4}$$

where

- $P_s(m|\sigma_m)$: B^0 signal mass model described by one Gaussian with event-by-event mass uncertainty,
- $P_b(m|\sigma_m)$: background mass model described by a linear polynomial,
- $P_s(\sigma_t)$ and $P_b(\sigma_t)$: distributions of proper decay time uncertainty, for signal and background, described by two or three components of Gamma functions, respectively,
- $P_s(\xi)$ and $P_b(\xi)$: same as Eq. 5.33,
- $P_s(\mathcal{D})$ and $P_b(\mathcal{D})$: dilution distribution for signal and background, using histograms from the data (see Fig 46).

Parameter	Description	Fit Result
s^+	Positive dilution scale factor	0.83 ± 0.07
s^-	Negative dilution scale factor	0.95 ± 0.07
ϵ_s	Tagging probability in signal	0.962 ± 0.002
ϵ_b	Tagging probability in background	0.982 ± 0.001
f_B	Positive/negative tagging fraction in background	0.520 ± 0.002
m	B hadron mass [MeV/c^2]	$5,278.8 \pm 0.1$
f_s	Signal Fraction	0.254 ± 0.002
s_{m1}	Mass Sigma of first Gaussian [MeV/c^2]	11.1 ± 0.1
s_{m2}	Mass Sigma of second Gaussian [MeV/c^2]	30.6 ± 1.6
f_{m1}	Fraction of first mass Gaussian	0.819 ± 0.013
P_1	Background slope	0.329 ± 0.297
$c\tau$	B hadron lifetime	488.1 ± 3.9 μm
s_t	Proper decay time scale factor	1.217 ± 0.005
f_p	Fraction of remainder which is prompt	0.872 ± 0.004
λ_+	Effective background lifetime, pos. component 1	590 ± 59 μm
f_{++}	Fraction of remainder which is in component 2	0.734 ± 0.016
λ_{++}	Effective background lifetime, pos. component 2	91.2 ± 4.5 μm
f_-	Fraction of remainder which is in neg. tail	0.611 ± 0.043
λ_-	Effective background lifetime, neg. component	157.6 ± 7.2 μm
f_1^s	Fraction of 1^{st} Gamma dist. in signal	0.094 ± 0.006
a_1^s	a-parameter, signal, 1^{st} Gamma dist.	1.81 ± 0.16
b_1^s	b-parameter, signal, 1^{st} Gamma dist.	23.4 ± 1.7 μm
a_2^s	a-parameter, signal, 2^{nd} Gamma dist.	7.96 ± 0.15
b_2^s	b-parameter, signal, 2^{nd} Gamma dist.	2.74 ± 0.05 μm
f_1^b	Fraction of 1^{st} Gamma dist. in background	0.047 ± 0.002
a_1^b	a-parameter, background, 1^{st} Gamma dist.	0.55 ± 0.05
b_1^b	b-parameter, background, 1^{st} Gamma dist.	106.7 ± 3.4 μm
f_2^b	Fraction of 2^{nd} Gamma dist. in background	0.682 ± 0.016
a_2^b	a-parameter, background, 2^{nd} Gamma dist.	16.40 ± 0.40
b_2^b	b-parameter, background, 2^{nd} Gamma dist.	1.50 ± 0.04 μm
a_3^b	a-parameter, background, 3^{rd} Gamma dist.	6.60 ± 0.19
b_3^b	b-parameter, background, 3^{rd} Gamma dist.	5.1 ± 0.1 μm

Table 9: Fit parameters and fit results for $B^+ \to J/\psi\, K^+$.

The differential rates for B^0 decays shown in Eq. 2.66 are much simpler than those for $B_s^0 \to J/\psi\phi$ decays, because the decay width difference $\Delta\Gamma$ is negligible for the B^0 system and the final state is only CP odd. With tagging information, the normalized probability density function for the signal $P_s(t|\sigma_t, \xi, \mathcal{D})$ of $B^0 \to J/\psi K_s^0$ is expressed as

$$P_s(t|\sigma_t, \xi, \mathcal{D}, \xi) = \frac{1 - \xi \cdot s\mathcal{D} \sin 2\beta \sin(\Delta m_d t)}{1 + |\xi|} \cdot \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes G(\sigma_t), \tag{6.5}$$

where the oscillation frequency Δm_d will be constrained to the PDG 2006 value $\Delta m_d = 0.507 \pm 0.005$ ps^{-1} in the fit. The time dependent part for the background $P_b(t|\sigma_t, \xi, \mathcal{D})$ can be written as

$$P_b(t|\sigma_t, \xi, \mathcal{D}) = \left[\delta(\xi) + \epsilon_{bg}^+ \delta(\xi - 1) + (1 - \epsilon_{bg}^+)\delta(\xi + 1)\right] P_b(t|\sigma_t), \tag{6.6}$$

where ϵ_{bg}^+ is the fraction of positive tagged background events, and $P_b(t|\sigma_t)$ is defined in Eq. 5.36.

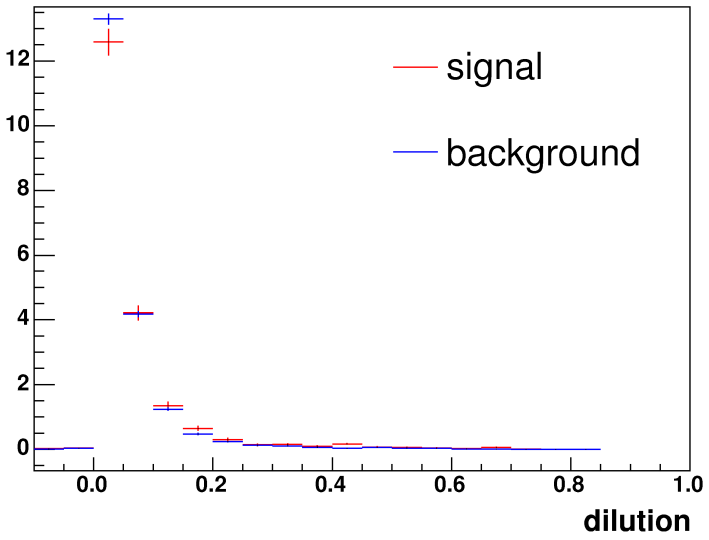

Figure 46: Normalized OST dilution distributions for signal (red) and background (blue) in the $B^0 \to J/\psi\, K_s^0$ channel.

The fitter is validated first by running 900 toy MC experiments with a $\sin 2\beta$ input value 0.73. In each experiment, 30,000 events are generated with a signal fraction of 0.224,

Parameter	Mean	Width
$\sin 2\beta$	-0.04 $\pm$ 0.03	1.04 $\pm$ 0.03
$c\tau$	+0.01 $\pm$ 0.03	0.98 $\pm$ 0.02
f_s	-0.05 $\pm$ 0.03	1.00 $\pm$ 0.02
ϵ_s	+0.01 $\pm$ 0.04	1.01 $\pm$ 0.02
m	+0.04 $\pm$ 0.03	1.00 $\pm$ 0.02

Table 10: Mean and width of Gaussian fits to 900 toy MC experiments for the $B^0 \to J/\psi K^0_s$ channel.

resulting in 6,720 signal events. Event-by-event dilution is generated according to the OST dilution histograms obtained from data. The means and widths of the Gaussian fits to the pull distribution of the signal parameters are listed in Table 10.

The B^0 sample is then fitted using the validated fitter. The full set of fit parameters and fit results from the fit to the B^0 sample are given in Table 11. The CP violation parameter is measured as

$$\sin 2\beta = 0.85 \pm 0.22(stat),$$

which agrees well with the PDG 2006 value (0.725 ± 0.037). Since the measurement of $\sin 2\beta$ is only a crosscheck for the CP violation measurement in the $B^0_s \to J/\psi\phi$ channel, and since it agrees well with the current PDG value combing results mainly from the BaBar and Belle experiments, we do not perform a full systematic error study. However, various systematical errors associated with the oscillation frequency and the tagging dilution scale factor have already been included in the statistical error, since they are Gaussian constrained in the fit.

To check the fit quality, projections of the likelihood function on different variables are also plotted. The mass projection is shown in Figure 47, the proper decay time projection is displayed in Figure 48, and the proper decay time uncertainty projection can be found in Figure 49.

Parameter	Description	Fit result
$\sin 2\beta$	CP asymmetry parameter	0.85 ± 0.22
ϵ_{0s}	Tagging probability of signal	0.957 ± 0.003
ϵ_{0b}	Tagging probability of background	0.977 ± 0.001
ϵ_{bg}^{+}	Positive tagging fraction of background	0.501 ± 0.004
m	B hadron mass [MeV/c^2]	5280.1 ± 0.2
f_s	Signal fraction	0.215 ± 0.003
s_m	Mass error scale factor	1.78 ± 0.03
P_1	Background Slope	-1.23 ± 0.50
$c\tau$	B hadron lifetime	465.7 ± 7.2 μm
s_t	Proper decay time scale factor	1.268 ± 0.010
f_p	Prompt fraction of background	0.790 ± 0.005
λ_+	Effective background lifetime, pos. component 1	526 ± 20 μm
f_{++}	Fraction of bkg which decays w/λ_{++}	0.430 ± 0.014
λ_{++}	Effective background lifetime, pos. component 2	234 ± 15 μm
f_-	Fraction of bkg which decays w/λ_-	0.33 ± 0.01
λ_-	Effective background lifetime, neg. component	382 ± 17 μm
f_1^s	Fraction of 1^{st} Gamma dist. in signal	0.197 ± 0.009
a_1^s	a-parameter, signal, 1^{st} Gamma dist.	1.03 ± 0.16
b_1^s	b-parameter, signal, 1^{st} Gamma dist.	64.1 ± 6.7 μm
a_2^s	a-parameter, signal, 2^{nd} Gamma dist.	7.73 ± 0.26
b_2^s	b-parameter, signal, 2^{nd} Gamma dist.	3.10 ± 0.10 μm
f_1^b	Fraction of 1^{st} Gamma dist. in background	0.163 ± 0.005
a_1^b	a-parameter, background, 1^{st} Gamma dist.	0.8832 ± 0.072
b_1^b	b-parameter, background, 1^{st} Gamma dist.	125.7 ± 3.9 μm
f_2^b	Fraction of 2^{nd} Gamma dist. in background	0.732 ± 0.024
a_2^b	a-parameter, background, 2^{nd} Gamma dist.	15.0 ± 0.5
b_2^b	b-parameter, background, 2^{nd} Gamma dist.	1.91 ± 0.07 μm
a_3^b	a-parameter, background, 3^{rd} Gamma dist.	6.08 ± 0.50
b_3^b	b-parameter, background, 3^{rd} Gamma dist.	6.79 ± 0.44 μm

Table 11: Fit parameters and fit results for $B^0 \to J/\psi\, K_s^0$.

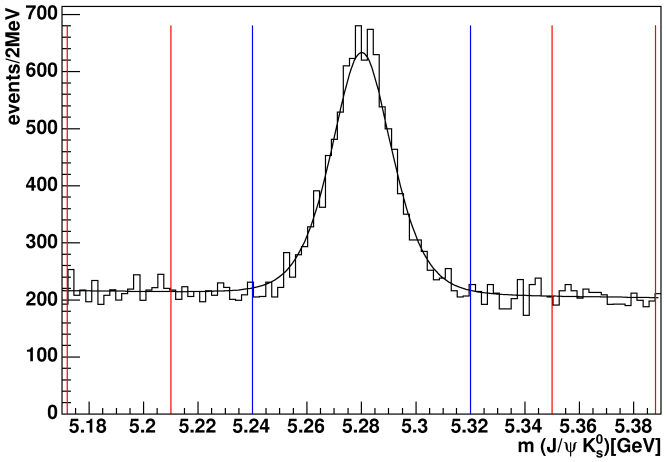

Figure 47: The mass projection of the $B^0 \to J/\psi\, K_s^0$ fit with signal region defined between blue lines and sideband region defined between red lines.

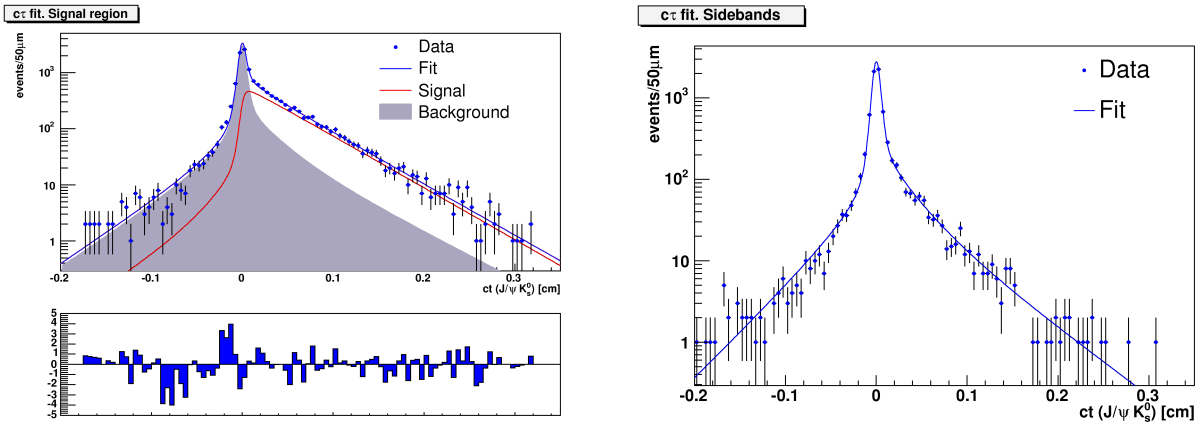

Figure 48: The proper decay time projection of the $B^0 \to J/\psi\, K_s^0$ fit. The projection in mass signal region (left), where the bottom plot shows residual in each bin divided by the estimated error. The projection in the mass sideband region (right).

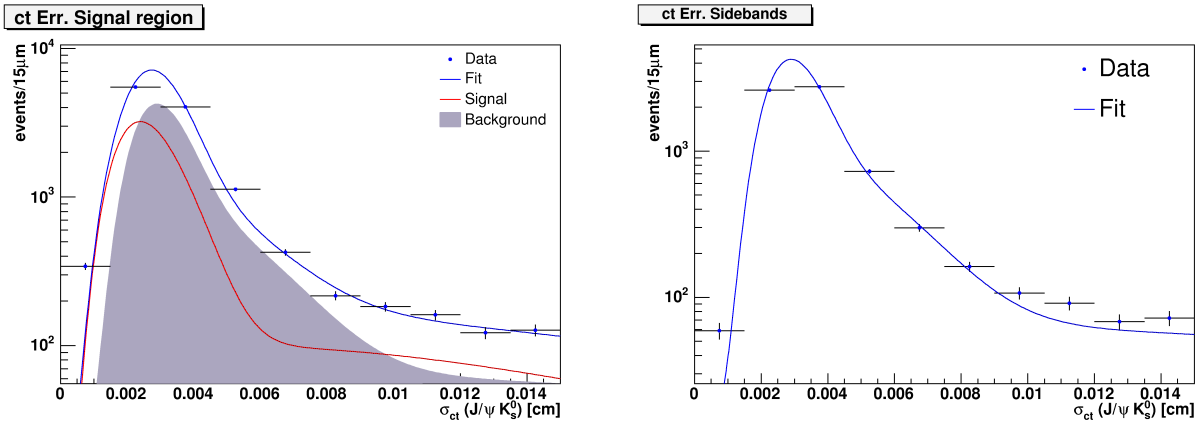

Figure 49: The proper decay time uncertainty projection of the $B^0 \to J/\psi\, K_s^0$ fit. The projection in mass signal region (left), and the projection in the mass sideband region (right).

7.0 FITTER VALIDATION FOR THE $B_S^0 \to J/\psi\phi$ ANALYSIS

Before fitting the data sample for the CP violating phase β_s we test the fitter extensively with toy Monte Carlo experiments. In the case of the fit for $\sin 2\beta$, the fit is simple and displays no pathologies. In the case of the fit to $B_s^0 \to J/\psi\phi$. The two main conclusions of this chapter are the following:

- the likelihood function (described in chapter 5) is rigorously validated,
- the maximum likelihood method cannot be used to derive an unbiased point estimate (meaning, central value and error estimate) for the main physics parameters of interest (namely $\Delta\Gamma$ and β_s), so heavier statistical machinery must be used.

It will help to discuss at this stage several important properties of the maximum likelihood method used to derive point estimates. First, under very general conditions, the maximum likelihood function can be shown to be consistent, namely unbiased in the limit of infinite statistics [60]. It is not, in general, unbiased for a finite data sample.

One of the conditions for the consistency of the maximum likelihood estimate is, of course, that the estimate exists and be unique. The symmetry discussed in the end of chapter 5 in fact violates this condition. It can be restored by restrictions on the parameter space, which effectively eliminate one half of the space. The second symmetric solution can in principle then be obtained from the first one. However in real data the parameter space can obviously never be so restricted.

Pull distributions are used to establish the expected consistency of the estimator, as a validation of the fitting function and procedure. This requires a data sample larger than the one actually collected, and a large amount of CPU time.

To reduce this CPU time we can use several tricks to increase the *effective* sample size without actually generating more events. For CP asymmetries to a single CP eigenstate, in the absence of a decay width difference, the effective sample size of a measurement with tagging efficiency ϵ, tagging dilution $\mathcal{D}$, and proper time resolution σ_t, is

$$N_{eff} = S\epsilon\mathcal{D}^2 e^{-(\Delta m_s \sigma_t)^2} \frac{S}{S+B}. \tag{7.1}$$

This means: an experiment with finite resolution ($\sigma_t > 0$), limited tagging probability ($\epsilon < 1$ and $\mathcal{D} < 1$), with S signal events and B background events has a sensitivity to CP asymmetry equivalent to a perfectly tagged background-free sample of N_{eff} events. This formula is derived from Fourier methods [61] and applicable, for example, to the decay $B^0 \to J/\psi K^0_s$. No equivalent formula is available, to our knowledge, for the decay $B^0_s \to J/\psi\phi$; however it is clear that even in this case high dilution samples with good resolution are a cheap way to achieve an increase in the effective statistical power of the sample. We employ such samples, generated with Toy Monte Carlo, during the validation of the fitting procedure.

Finally, when we return to realistic sample sizes, dilutions, resolutions and background levels after establishing the accuracy of our probability distributions, we observe irregularities in the likelihood function, exacerbated by all of the above effects, in addition to symmetries and periodicities in the likelihood function. In the following chapter we develop the machinery to deal with that.

In the following sections, we check the fitter in several different cases:

- Signal-only toy Monte Carlo with 2,000 events per toy, where the tagging power is perfect, detector resolution smearing is not included, and the detector acceptance is perfect. This checks the implementation of the pure physics.
- Signal-only toy Monte Carlo with 2,000 events per toy, with perfect tagging power and detector resolution, but using realistic detector acceptance. This checks the implementation of detector acceptance function.
- Signal-only toy Monte Carlo with 10,000 events per toy, with perfect tagging power, but using realistic detector resolution and acceptance. This checks the implementation of detector resolution smearing.

- Full toy Monte Carlo with both signal and background, with 50,000 events per toy, with realistic tagging power, detector resolution and acceptance, but high statistics. This checks the background modeling.

The signal generation of toy Monte Carlo experiments includes several pieces:

- **Lifetime and decay angles**

  The proper decay time and angular distribution angles are generated according to the four dimensional decay probability density functions for both B_s^0 and $\bar{B}_s^0$ as in Eq. 2.85.

- **Detector acceptance**

  The detector acceptance is added by multiplying the probability density functions with the acceptance functions obtained from Monte Carlo sample.

- **Detector resolution**

  After lifetime and decay angles for each event are generated according to theoretical formulas with or without acceptance correction, the detector resolution effect is added by smearing the proper decay time with an event-by-event error which varies according to a distribution obtained from the data.

- **Flavor tagging**

  Tagging efficiencies, event-by-event tagging decisions and dilution estimates are generated according to distributions obtained from the data.

- **Mass and mass error**

  The mass for each event is generated according to a Gaussian distribution with event-by-event mass errors following a distribution obtained from data.

The background of the toy Monte Carlo is generated according to the background model discussed in section 5.3.2.

Unless specified all the toy Monte Carlo experiments are generated using same signal parameters given in Table 12. Lifetime τ_s is chosen to be typical, $\Delta\Gamma$ and β_s values are around SM predictions. The initial values of the angular amplitudes and their phases are taken roughly from the first CDF Run II measurement of the $B^0 \to J/\psi\, K^{*0}$ [62]. The oscillation frequency Δm_s is generated according to CDF measurement and fixed during the fit.

Fit Name	Parameter	Input Value
Tau	$c\tau_s$	450 μm
DeltaGamma	$\Delta\Gamma$	0.1 ps^{-1}
Betas	β_s	0.02
PhiPara	$\arg(A_\parallel A_0^*)$	2.4
PhiPerp	$\arg(A_\perp A_0^*)$	0.5
AlphaPerp	$\lvert A_\perp\rvert^2$	0.215
AlphaPara	$\lvert A_\parallel\rvert^2/(\lvert A_0\rvert^2 + \lvert A_\parallel\rvert^2)$	0.284
DeltaM	Δm_s	17.77 ps^{-1}

Table 12: Parameters used in generation of toy Monte Carlo experiments.

7.1 TOY MONTE CARLO STUDY WITH FLAVOR TAGGING

In this section we first validate the fitter using large sample of toy Monte Carlo, adding one effect after another. After the fitter is validated, two-dimensional likelihood profiles of β_s and $\Delta\Gamma$ for random toy Monte Carlo are plotted to study the two-fold ambiguity in the likelihood function, and demonstrate the irregularities in the contours for low statistics.

7.1.1 Pull Study of Toy Monte Carlo

To check the implementation of different components of the likelihood function, different kinds of toy Monte Carlo experiments are generated. Pull distributions of signal parameters and fit results are shown in each case.

- **Differential rates of B_s^0 and $\bar{B}_s^0$, no detector effects**

  Signal only toy Monte Carlo experiments are generated according to Eq. 2.85 and maximum likelihood fits are performed in order to obtain estimates of the seven physics parameters in Table 13. The pull distributions of the signal parameters are shown in Figure 50, and the means and widths of Gaussian fits to the pulls are given in Table 13.

All are consistent with normal distributions, giving confidence in the implementation of the pure signal likelihood and normalization in the fit.

- **Differential rates with detector acceptance**

  After checking the "perfect" scenario, we include our implementation of detector acceptance in the Monte Carlo generator and also in the likelihood fit. The toy Monte Carlo experiments are generated with the angular distributions sculpted by the acceptance function, and in the fit, the differential rates are multiplied by the acceptance function and renormalized according to section 5.1.1. The pull distributions of the signal parameters are shown in Figure 51, and the means and widths of Gaussian fits to the pulls are given in Table 14. All are consistent with normal distributions, giving confidence in the implementation of detector sculpting effect in the fit.

- **Detector resolution**

  We check the implementation of detector resolution effects by generating proper decay time with event-by-event uncertainties distributed as in the data, and fitting the toy with a probability density function smeared with a Gaussian resolution function. High statistics toy Monte Carlo experiments (10,000 events per toy) are generated in order to compensate for the decrease of statistical power due to limited detector resolution. The pull distributions of the signal parameters are shown in Figure 52, and the means and widths of Gaussian fits to the pulls are listed in Table 15. The normal distribution of all the signal parameters indicates the correct implementation of the resolution function.

- **Full Monte Carlo tests with realistic tagging power**

  At this point, the signal likelihood has been validated. We continue to study the most realistic situation, where both signal and background are included. The toy Monte Carlo experiments are generated with both detector resolution and sculpting effects. Event-by-event dilution estimates are generated from histograms of the predicted dilution for OST and SSKT, rescaled by the dilution scale factors discussed in section 5.2. All the toys are generated with high statistics to check the implementation of the full likelihood function.

  The pull distributions of all signal parameters are shown in Fig 53. Means and widths of Gaussian fits to those parameters are listed in Table 16. For all the parameters we

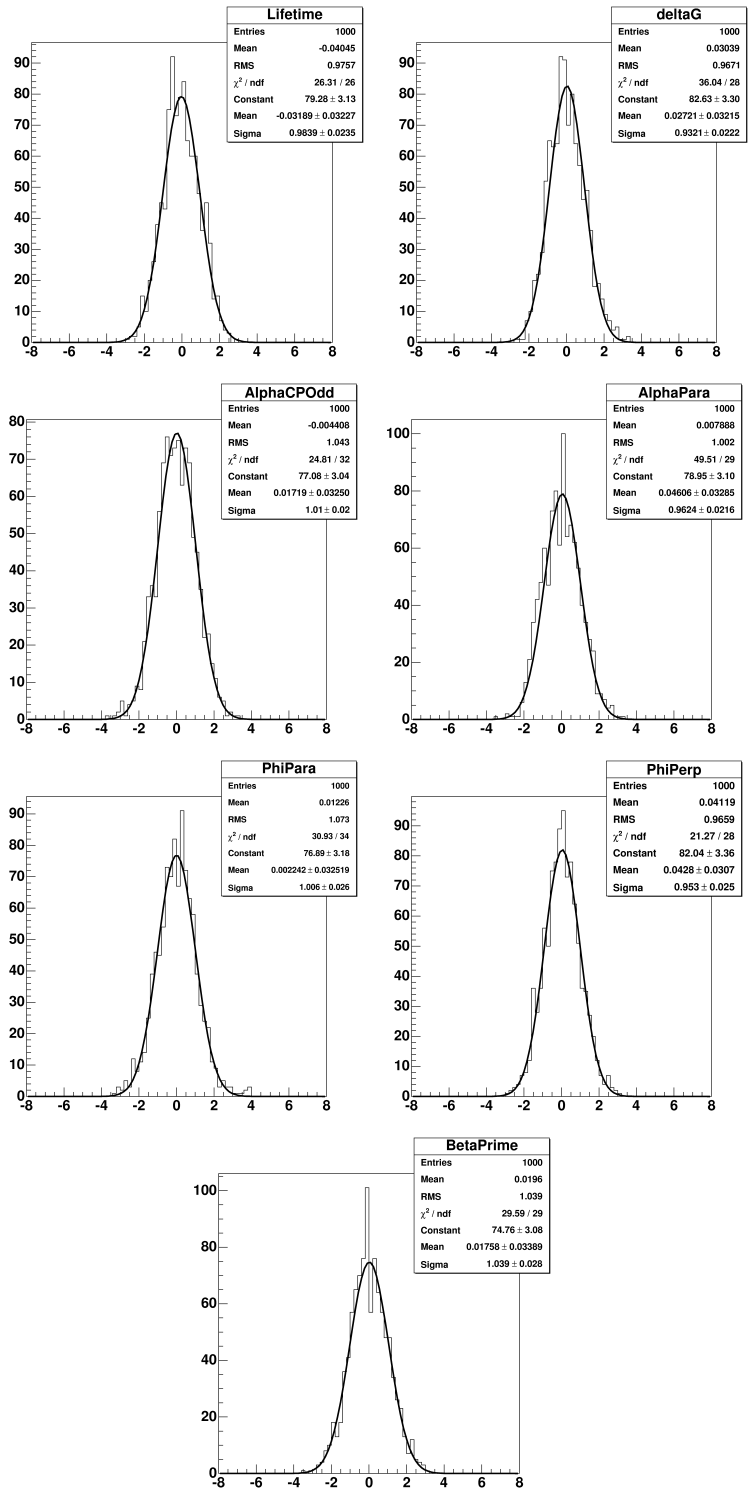

Figure 50: Pull distributions for 1,000 unsmeared signal only toy Monte Carlo experiments with 2,000 events per toy, perfect tagging and detector acceptance.

Parameter	Mean	Width
$c\tau$	-0.03 ± 0.03	0.98 ± 0.02
$\alpha_\perp$	0.02 ± 0.03	1.01 ± 0.02
$\alpha_\parallel$	0.05 ± 0.03	0.96 ± 0.02
$\phi_\parallel$	0.002 ± 0.030	1.006 ± 0.026
$\phi_\perp$	0.04 ± 0.03	0.95 ± 0.03
$\Delta\Gamma$	0.03 ± 0.03	0.93 ± 0.03
β_s	0.02 ± 0.03	1.04 ± 0.03

Table 13: Mean and width of Gaussian fits to 1,000 unsmeared signal only toy Monte Carlo experiments with 2,000 events per toy with perfect tagging and detector acceptance.

observe unit pull distributions which demonstrate correct modeling of both signal and background.

7.1.2 Likelihood Profile of β_s and $\Delta\Gamma$

In this section we define a likelihood profile, which in the context of Bayesian statistics is equivalent to a confidence region. Suppose we have a likelihood function with n parameters

$$\mathcal{L} = \mathcal{L}(p_1, p_2 ... p_n).$$

The maximum likelihood fit will return an overall best (maximum likelihood or minimum negative likelihood) value,

$$\mathcal{L}_{best} = \mathcal{L}(\hat{p}_1, \hat{p}_2 ... \hat{p}_n).$$

Now let us fix m ($m \leq n$) parameters to some specific values and maximize the likelihood function with respect to all the other parameters. The new likelihood value obtained at

$$\mathcal{L}_{rest} = \mathcal{L}(p_1^* ... p_m^*, \hat{p}_{m+1} ... \hat{p}_n),$$

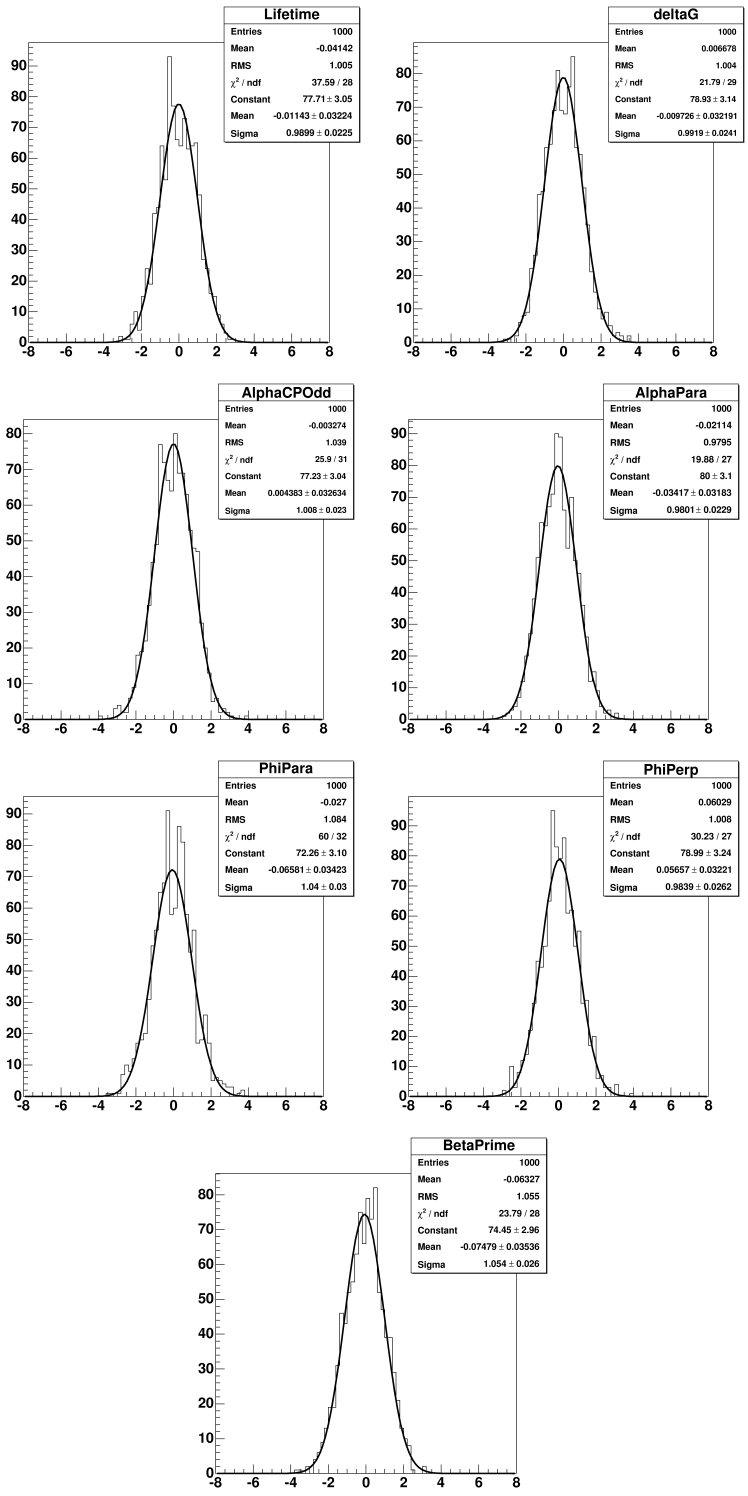

Figure 51: Pull distributions for 1,000 unsmeared signal only toy Monte Carlo experiments with 2,000 events per toy, perfect tagging, and detector sculpting included.

Parameter	Mean	Width
$c\tau$	-0.01 ± 0.03	0.99 ± 0.02
$\alpha_\perp$	0.004 ± 0.030	1.008 ± 0.023
$\alpha_\parallel$	-0.03 ± 0.03	0.98 ± 0.02
$\phi_\parallel$	-0.07 ± 0.03	1.04 ± 0.03
$\phi_\perp$	0.06 ± 0.03	0.98 ± 0.03
$\Delta\Gamma$	-0.01 ± 0.03	0.99 ± 0.02
β_s	-0.07 ± 0.03	1.06 ± 0.03

Table 14: Mean and width of Gaussian fits to the 1,000 unsmeared signal only toy Monte Carlo experiments with 2,000 events per toy with perfect tagging. Detector sculpting effect is included.

with these m parameters fixed, define the quantity

$$-2log(\mathcal{L}_{rest}/\mathcal{L}_{best}) = -2(log\mathcal{L}_{rest} - log\mathcal{L}_{best}) \equiv -2\Delta log\mathcal{L},$$

which in a typical experiment is small for likely values of the m parameters and large for unlikely values. For ensembles of psedudo-experiment, $-2\Delta log\mathcal{L}$ follows a χ^2 distribution. With only one degree of freedom ($m = 1$), the 95% confidence region is obtained at $-2\Delta log\mathcal{L}$ $\sim$ 3.84. For a two-dimensional likelihood profile of two parameters, the same confidence region is obtained with $-2\Delta log\mathcal{L} \sim 5.99$.

As described in the previous chapter, we observe a two-fold ambiguity in the likelihood function with simultaneous transformations of the following parameters: β_s, $\Delta\Gamma$, $\phi_\parallel$ and $\phi_\perp$. To visualize the invariance, we plot the two-dimensional likelihood profiles of β_s and $\Delta\Gamma$ for different toy Monte Carlo experiments. The toy Monte Carlo is generated with same statistics as the data and all kinds of realistic effects included.

The $\beta_s - \Delta\Gamma$ space is divided into $N \times M$ (20×20 here) grids within the ranges: -0.7 to +0.7 for $\Delta\Gamma$, -$\pi/2$ to +$\pi/2$ for β_s. $\Delta\Gamma$ and β_s are fixed to the center value of each grid and the likelihood function is maximized with respect to all other parameters to get the

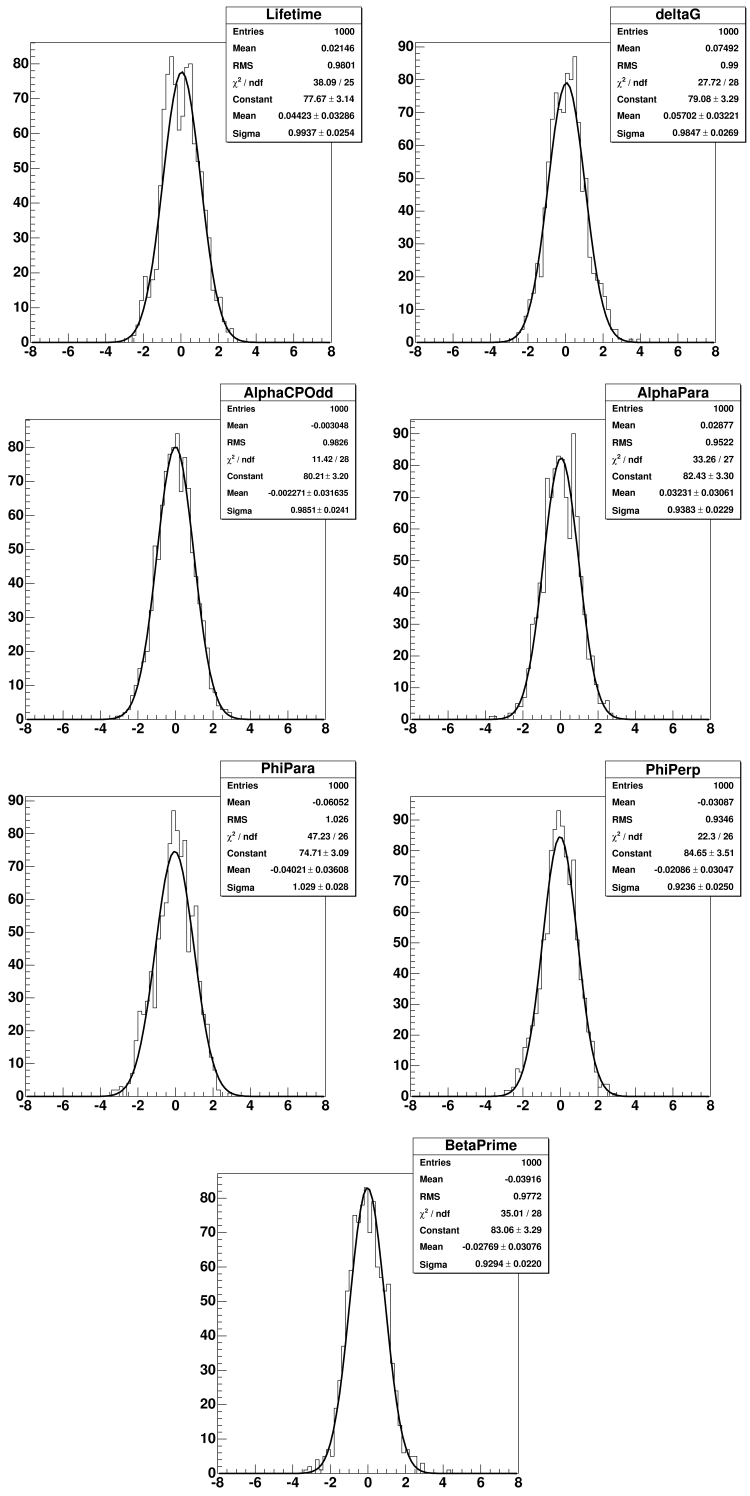

Figure 52: Pull distributions for 1,000 smeared signal only toy Monte Carlo experiments with 2,000 events per toy with perfect tagging, and detector sculpting effect is included.

Parameter	Mean	Width
$c\tau$	0.04 ± 0.03	0.99 ± 0.03
$\alpha_\perp$	-0.003 ± 0.030	0.985 ± 0.024
$\alpha_\parallel$	0.03 ± 0.03	0.94 ± 0.03
$\phi_\parallel$	-0.04 ± 0.04	1.03 ± 0.03
$\phi_\perp$	-0.02 ± 0.03	0.92 ± 0.03
$\Delta\Gamma$	0.06 ± 0.03	0.98 ± 0.03
β_s	-0.03 ± 0.03	0.93 ± 0.02

Table 15: Mean and width of Gaussian fits to 1,000 smeared toy Monte Carlo with 10,000 events per toy and perfect dilution. Detector sculpting effect is included.

likelihood value. These likelihood values are then compared to the best value obtained when all the parameters including β_s and $\Delta\Gamma$ are allowed to float. The likelihood profile is thus obtained, and we connect all the grids with $-2\Delta log\mathcal{L} \sim 2.30$ and $-2\Delta log\mathcal{L} \sim 5.99$ to get a Bayesian confidence region at 68% and 95% confidence levels respectively.

In generating the contours we vary the input values of β_s at 0.02, 0.42 and 0.78 respectively, but input value for $\Delta\Gamma$ is kept at +0.1. Several Monte Carlo experiments are generated to study the two-dimensional likelihood profiles. Here we show some typical likelihood profiles for different β_s input values. The likelihood profiles with β_s generated at 0.02 are shown in Figure 54, the likelihood files with β_s generated at 0.42 are shown in Figure 55, and the likelihood files with β_s generated at 0.78 are shown in Figure 56. From these likelihood profiles, one does observe the expected ambiguity. However, for a small β_s input value, the 68% confidence regions are more separated. With a larger β_s input value, one realizes that β_s cannot be resolved. The likelihood profiles shown in this series of plots are highly irregular, and their shapes are far from parabolic. In such situations, "point estimates" (the point with maximum likelihood value, plus the covariance matrix), are not meaningful. The likelihood profile itself is needed to obtain a Bayesian confidence region. Our final result, a frequentist confidence region, is explained in Chap. 8. For the moment, however, we can

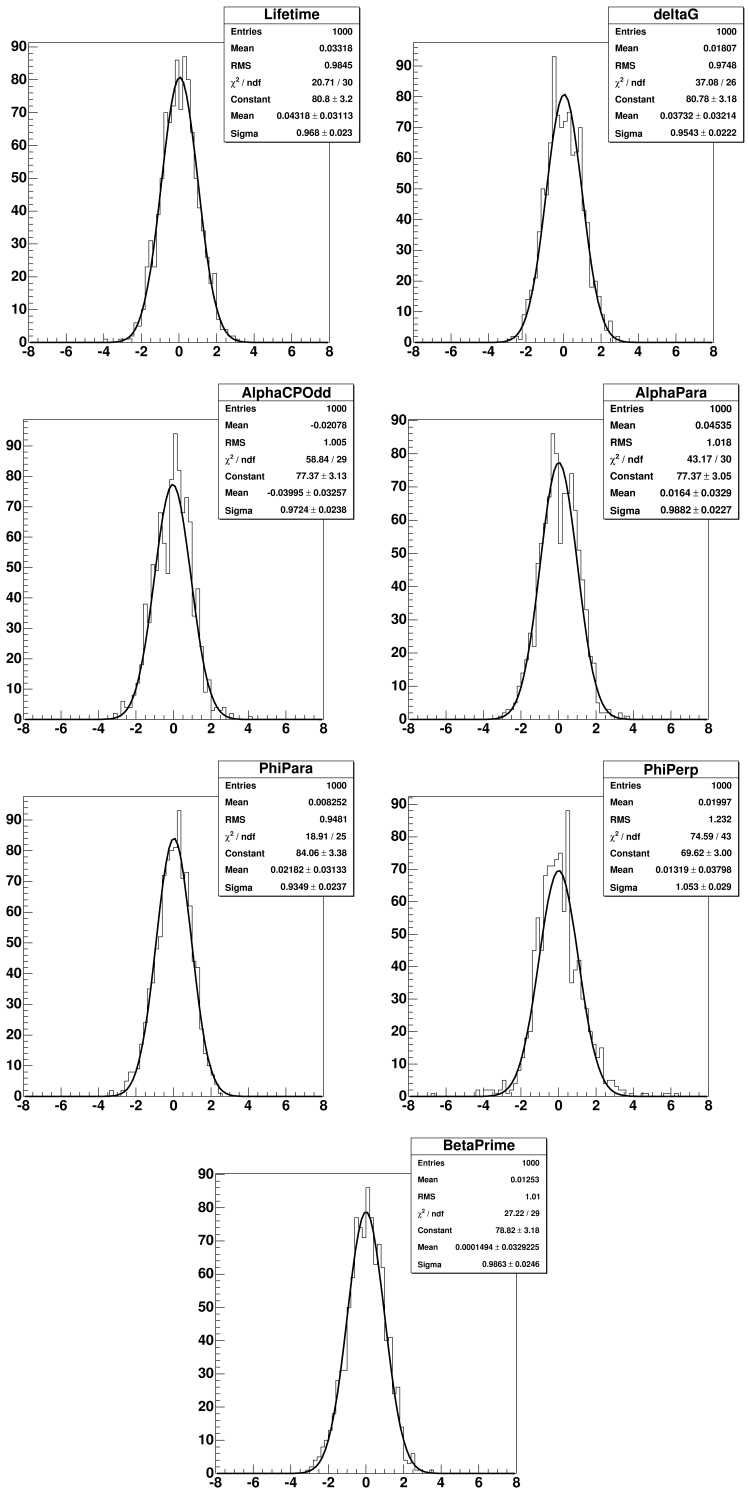

Figure 53: Pull distributions of signal parameters for 1,000 fully simulated toy Monte Carlo experiments tagged with the combined OST and SSKT as measured in data, generated with 50,000 events per toy, where the signal fraction is 0.26.

Parameter	Mean	Width
$c\tau$	0.04 ± 0.03	0.97 ± 0.02
$\alpha_\perp$	-0.04 ± 0.03	0.97 ± 0.02
$\alpha_\parallel$	0.02 ± 0.03	0.99 ± 0.02
$\phi_\parallel$	0.02 ± 0.03	0.94 ± 0.02
$\phi_\perp$	0.01 ± 0.04	1.05 ± 0.03
$\Delta\Gamma$	0.04 ± 0.03	0.95 ± 0.02
β_s	0.0002 ± 0.0329	0.9863 ± 0.025

Table 16: Mean and width of Gaussian fits to 1,000 toy Monte Carlo experiments tagged with the combined OST and SSKT as measured in data, generated with 50,000 events per toy, where the signal fraction is 0.26.

conclude that a point estimate for this analysis is precluded.

7.2 TOY MONTE CARLO STUDY WITHOUT FLAVOR TAGGING

We also consider the situation when flavor tagging is not available, which is referred as the untagged analysis. The lack of flavor identification simplifies the likelihood function, but also introduces a higher level of symmetry. As discussed in section 5.3.4, however, if one assumes CP conservation ($\beta_s = 0$), the likelihood function is further simplified, and most of the ambiguities are absent, except $\phi_\parallel$ and $2\pi - \phi_\parallel$ cannot be distinguished. The lifetime τ and decay width difference $\Delta\Gamma$ can be extracted easily and compared to the standard model predictions. For this study, the fitter is validated with 1,000 high statistics and full toy Monte Carlo with both signal and background in. The pull distributions for the signal parameters are shown in Figure 57, and the Gaussian fits of those pulls are listed in Table 17, which are good in general.

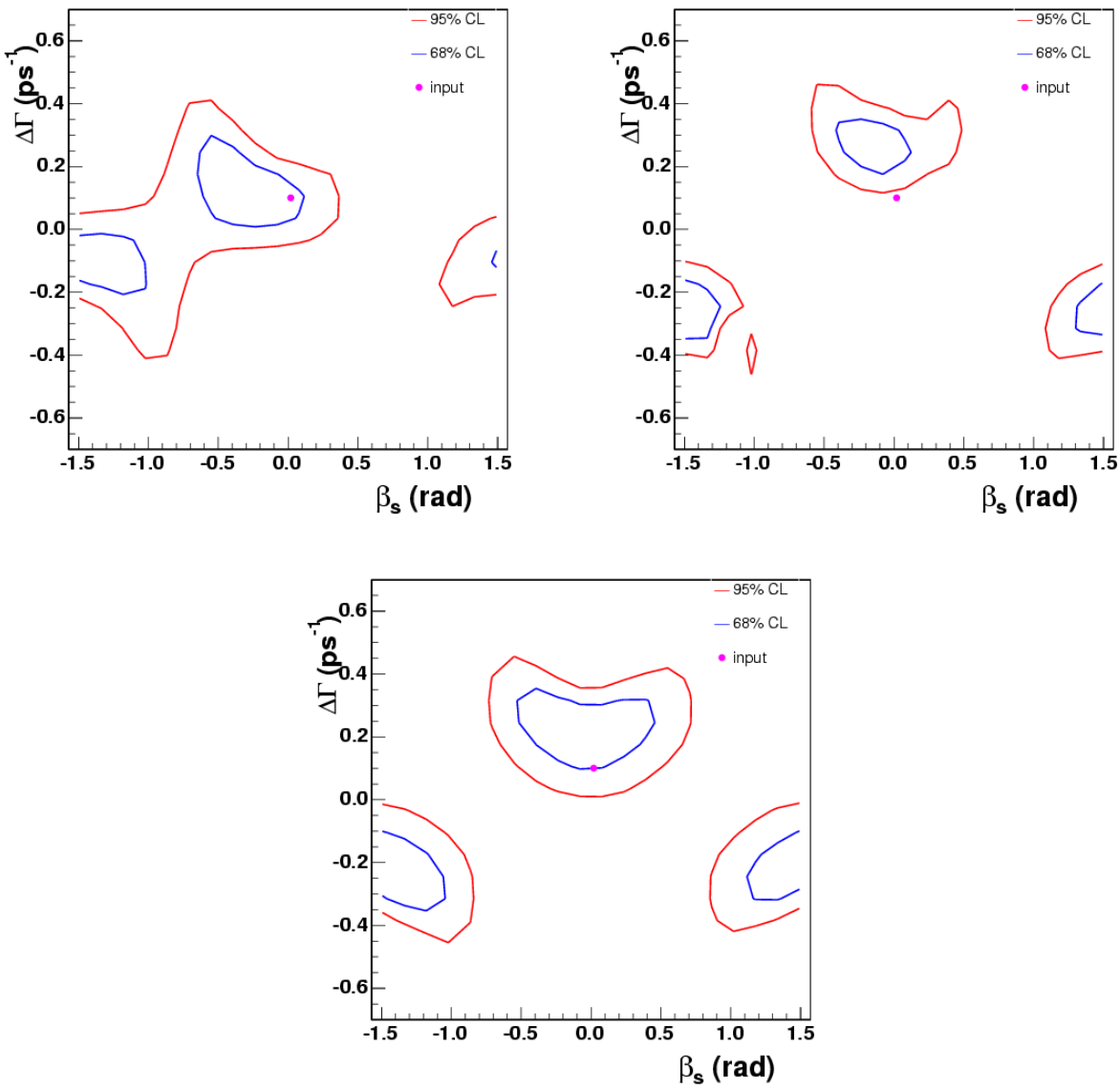

Figure 54: Likelihood profiles of random toys with an input value of $\beta_s = 0.02$.

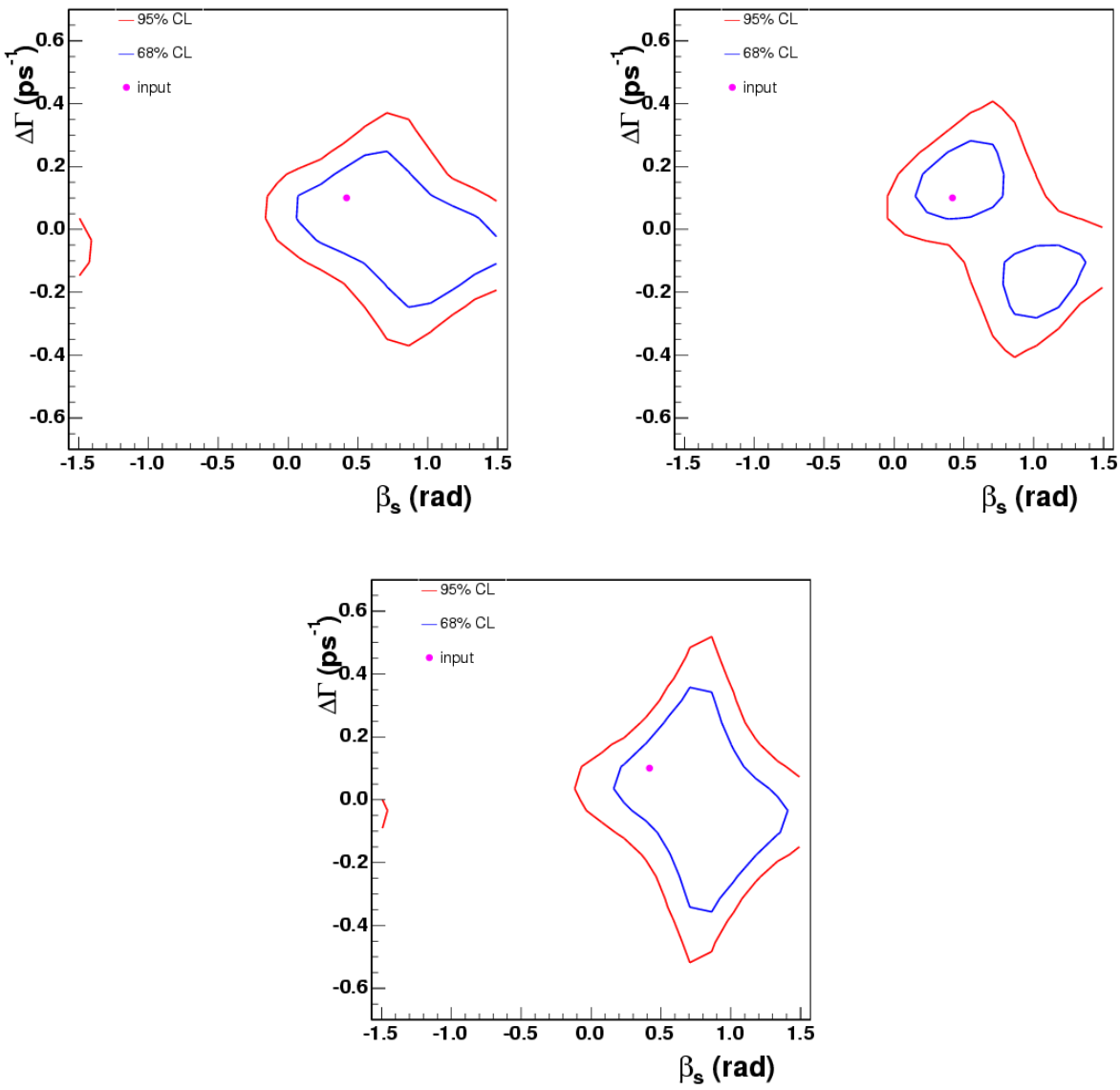

Figure 55: Likelihood profiles of random toys with an input value of $\beta_s = 0.42$.

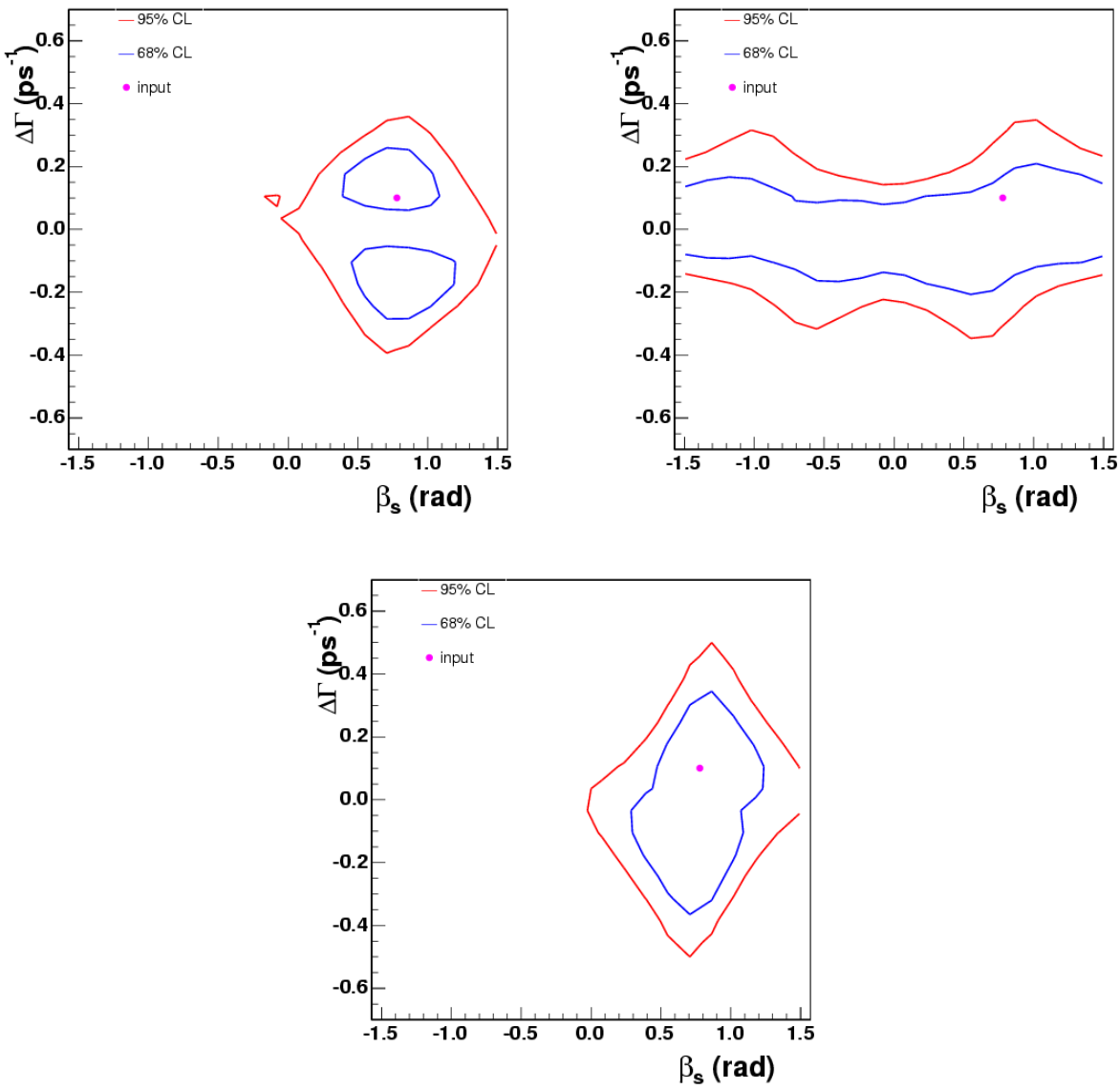

Figure 56: Likelihood profiles of random toys with an input value of $\beta_s = 0.78$.

Parameter	Mean	Width
$c\tau$	-0.06 $\pm$ 0.03	0.99 $\pm$ 0.03
$\alpha_\perp$	-0.002 $\pm$ 0.030	0.976 $\pm$ 0.023
$\alpha_\parallel$	0.07 $\pm$ 0.03	0.99 $\pm$ 0.02
$\phi_\parallel$	-0.002 $\pm$ 0.030	0.938 $\pm$ 0.024
$\Delta\Gamma$	-0.04 $\pm$ 0.03	0.97 $\pm$ 0.02

Table 17: Mean and width of Gaussian fits to 1,000 untagged toy Monte Carlo with 50,000 events per toy, where the signal fraction is 0.26 and β_s is fixed to zero. The parameter $\phi_\perp$ does not appear in the likelihood function.

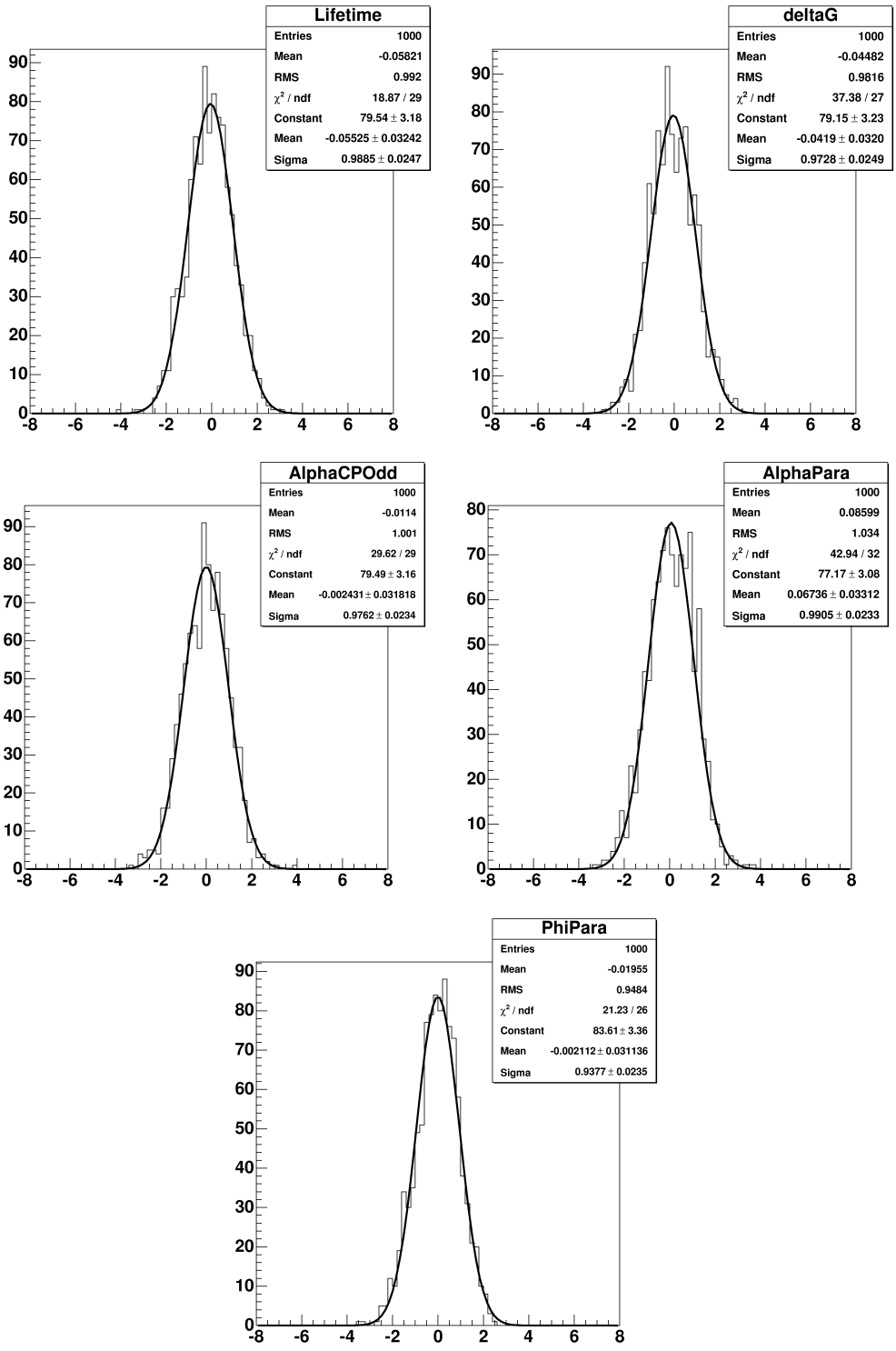

Figure 57: Pull distributions for 1,000 fully simulated untagged toy Monte Carlo with 50,000 events per toy, where the signal fraction is 0.26 and β_s is fixed to zero.

8.0 DETERMINATION OF β_S IN $B_S^0 \to J/\psi\phi$

With the likelihood function validated using toy Monte Carlo we are now ready to fit the data. First, we extract the lifetime and decay width difference without flavor tagging information, where CP conservation is assumed, obtaining "standard model" fit results and likelihood projections. Then we add flavor tagging information and fit for CP violation. The fit results and projections associated with different minima are given. We present then a frequentist statistical procedure used to obtain the final confidence region, taking into account both systematic uncertainties and the non-Gaussian behavior of the uncertainty.

8.1 UNTAGGED MEASUREMENT RESULTS WITH β_S FIXED TO ZERO

We fit the 1.35 fb^{-1} of data without using tagging information. The results are compared to CDF untagged analysis [63] as a crosscheck.

The fit parameters and results are given in Table 18. The mass likelihood projection is shown in Figure 58. The proper decay time likelihood projections are shown in Figures 59 and 60, and the angular likelihood projections are shown in Figure 61.

We can compare the result we obtained:

$$\Delta\Gamma = -0.017 \pm 0.088(\text{stat}) \text{ ps}^{-1}$$

with the result from Ref. [63]:

$$\Delta\Gamma = 0.076^{+0.059}_{-0.063}(\text{stat}) \pm 0.006(\text{syst}) \text{ ps}^{-1}$$

Parameter	Description	Fit result
$\Delta\Gamma$	CP asymmetry parameter [ps^{-1}]	-0.017 ± 0.088
$\alpha_\perp$	CP odd fraction	0.243 ± 0.037
$\alpha_\parallel$	$A_\parallel$ fraction in CP even states	0.313 ± 0.027
$\phi_\parallel$	$\arg(A_\parallel A_0^*)$	2.43 ± 0.18
$N(B_s^0)$	B_s^0 yield	$2,024 \pm 45$
m	B hadron mass [MeV/c^2]	5365.9 ± 0.2
f_s	Signal Fraction	0.259 ± 0.006
s_m	Mass error scale factor	1.578 ± 0.036
p_1	Background slope	-2.44 ± 1.25
$c\tau$	Proper decay time	445.8 ± 18.8 μm
s_t	Proper decay time scale factor	1.26 ± 0.02
f_p	Prompt fraction of background	0.792 ± 0.015
λ_+	Effective background lifetime, pos. component 1	366 ± 47 μm
f_{++}	Fraction of bkg which decays w/λ_{++}	0.655 ± 0.041
λ_{++}	Effective background lifetime, pos. component 2	70 ± 9 μm
f_-	Fraction of bkg which decays w/λ_-	0.382 ± 0.059
λ_-	Effective background lifetime, neg. component	125 ± 15 μm
$\cos(\theta)_1$	First parameter in bkg fit to $\cos(\theta)$	0.501 ± 0.131
$\cos(\theta)_2$	Second parameter in bkg fit to $\cos(\theta)$	0.341 ± 0.149
ϕ_1	First parameter in bkg fit to ϕ	0.176 ± 0.020
ϕ_2	Second parameter in bkg fit to ϕ	0.083 ± 0.041
$\cos(\psi)_1$	First parameter in bkg fit to $\cos(\psi)$	0.243 ± 0.168
$\cos(\psi)_2$	Second parameter in bkg fit to $\cos(\psi)$	-0.122 ± 0.185
f_1^s	Fraction of 1^{st} Gamma dist. in signal	0.744 ± 0.157
a_1^s	a-parameter, signal, 1^{st} Gamma dist.	10.5 ± 1.1
b_1^s	b-parameter, signal, 1^{st} Gamma dist.	1.68 ± 0.20 μm
a_2^s	a-parameter, signal, 2^{nd} Gamma dist.	7.38 ± 2.82
b_2^s	b-parameter, signal, 2^{nd} Gamma dist.	3.38 ± 0.75 μm
f_1^b	Fraction of 1^{st} Gamma dist. in bkg.	0.909 ± 0.014
a_1^b	a-parameter, bkg., 1^{st} Gamma dist.	12.0 ± 0.4
b_1^b	b-parameter, bkg., 1^{st} Gamma dist.	1.97 ± 0.07 μm
a_2^b	a-parameter, bkg., 2^{nd} Gamma dist.	2.94 ± 0.35
b_2^b	b-parameter, bkg., 2^{nd} Gamma dist.	11.1 ± 1.0 μm

Table 18: Fit parameters and results for the untagged measurement of $\Delta\Gamma$ in 1.35 fb^{-1} of data.

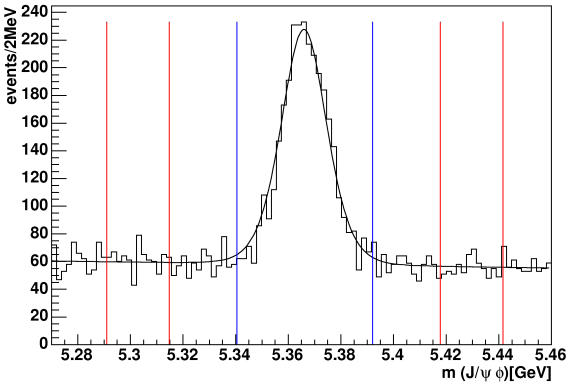

Figure 58: Mass likelihood projection, where region between blue lines is signal region, regions between red lines are two sideband regions.

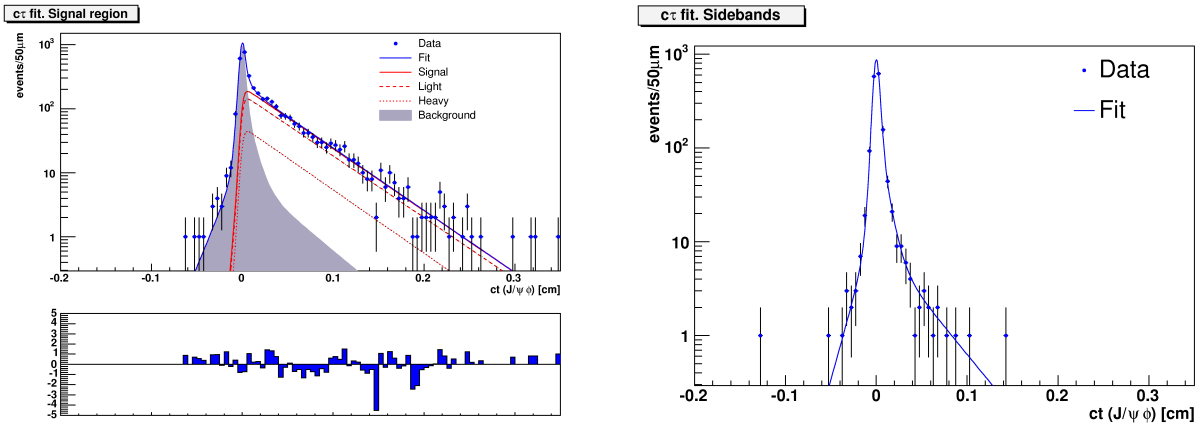

Figure 59: Proper decay time likelihood projection in signal region (left). The light eigenstate (CP even) is dominant, and the slight slope difference between light and heavy eigenstates shows the decay width difference. The bottom plot shows residual in each bin divided by the estimated error. Proper decay time likelihood projection in sideband region (right), which is basically all background.

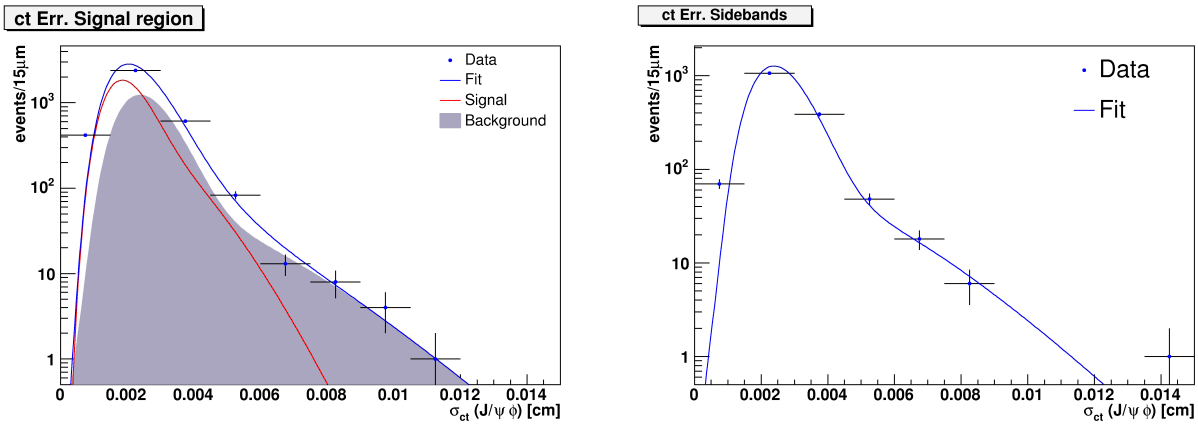

Figure 60: Proper decay time uncertainty likelihood projection in signal region (left). Proper decay time uncertainty likelihood projection in sideband region (right).

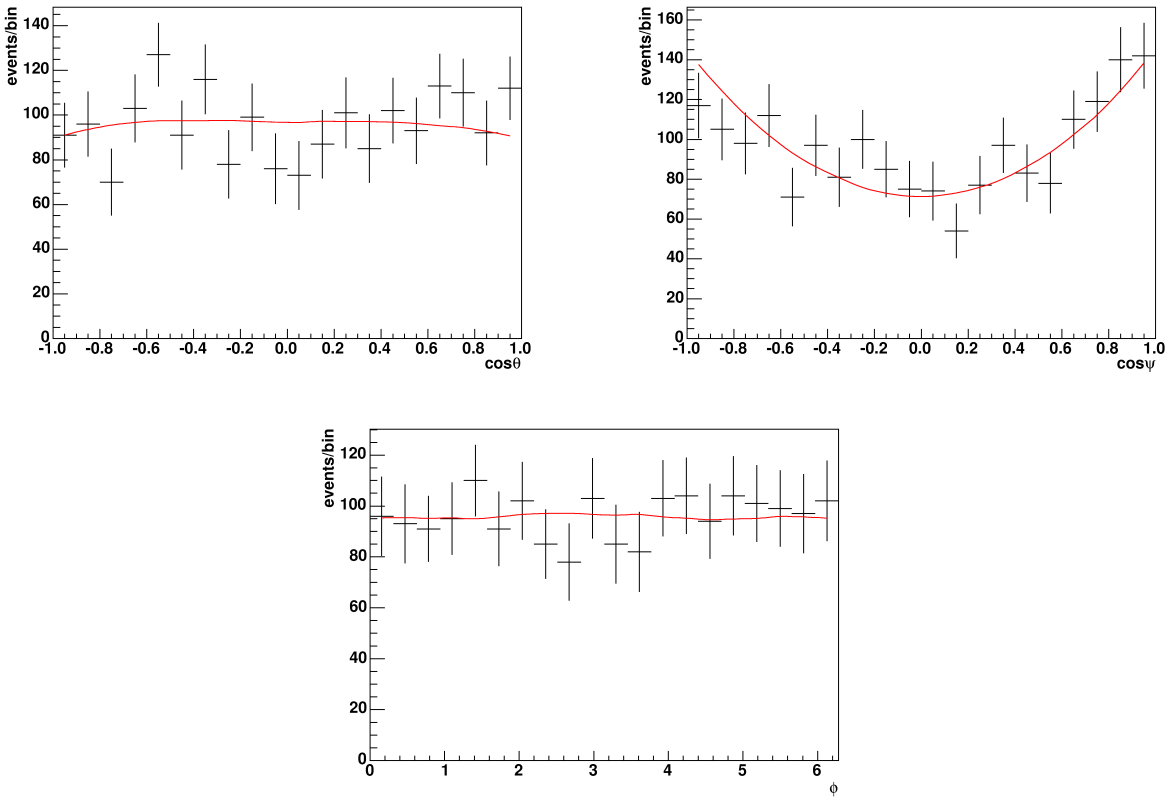

Figure 61: Sideband subtracted signal angular likelihood projection

which is measured with 1.7 fb^{-1} of data. These two results are consistent with each other within statistic uncertainties, which all agree with the standard model prediction [22]:

$$\Delta\Gamma_{SM} = 0.096 \pm 0.039 \text{ ps}^{-1}.$$

It is obvious that the measurement of $\Delta\Gamma$ is statistics dominated at this time.

To verify that the difference comes from statistical fluctuations, we add another 0.2 fb^{-1} of data in which flavor tagging was unavailable and fit all the data up to 1.55 fb^{-1} of luminosity, which is more comparable to the integrated luminosity used in the untagged analysis [63]. The fit results for this data are given in Table 19. The new measurement shows

$$\Delta\Gamma = 0.042 \pm 0.069 \text{ ps}^{-1},$$

which is in better agreement with Ref. [63].

8.2 FIT β_S WITH FLAVOR TAGGING

With the inclusion of tagging information, the likelihood function described in section 5.3 and validated in section 7.1 is used. The B_s^0 oscillation frequency Δm_s is constrained within its uncertainty to the measured value from the CDF B_s mixing analysis [28]. Dilution scale factors discussed in section 5.2 are also constrained within their errors.

We first demonstrate the two-fold ambiguity discussed in previous chapters by fitting the data with different starting values for those corresponding parameters. The maximum likelihood value $-2ln\mathcal{L}$ returned by the fit is used to make sure we get two equal mimima. First, the likelihood fit is started with positive initial value for decay width difference $\Delta\Gamma$, and the fit results are listed in Table 20. Second, the fit is started with negative initial value for $\Delta\Gamma$, and the fit results are listed in Table 21. Comparing the results in those tables, one notices that parameters β_s, Δ_Γ, $\phi_\parallel$ and $\phi_\perp$ transform roughly according to the predictions as in Eq. 5.43 with statistical error.

To illustrate the fit quality, we show likelihood projections for positive $\Delta\Gamma$ as an example. The mass likelihood projections shown in Figure 62, while the proper decay time likelihood

Parameter	Description	Fit result
$\Delta\Gamma$	CP asymmetry parameter [ps^{-1}]	0.042 ± 0.069
$\alpha_\perp$	CP odd fraction	0.247 ± 0.032
$\alpha_\parallel$	$A_\parallel$ fraction in CP even states	0.309 ± 0.026
$\phi_\parallel$	$\arg(A_\parallel A_0^*)$	2.65 ± 0.24
$N(B_s^0)$	B_s^0 yield	$2,303 \pm 45$
m	B hadron mass [MeV/c^2]	5365.9 ± 0.2
f_s	Signal Fraction	0.261 ± 0.005
s_m	Mass error scale factor	1.591 ± 0.034
p_1	Background slope	-2.55 ± 1.18
$c\tau$	Proper decay time	452.9 ± 15.3 μm
s_t	Proper decay time scale factor	1.26 ± 0.02
f_p	Prompt fraction of background	0.792 ± 0.016
λ_+	Effective background lifetime, pos. component 1	342 ± 42 μm
f_{++}	Fraction of bkg which decays w/λ_{++}	0.635 ± 0.041
λ_{++}	Effective background lifetime, pos. component 2	65 ± 10 μm
f_-	Fraction of bkg which decays w/λ_-	0.361 ± 0.055
λ_-	Effective background lifetime, neg. component	121 ± 14 μm
$\cos(\theta)_1$	First parameter in bkg fit to $\cos(\theta)$	0.527 ± 0.123
$\cos(\theta)_2$	Second parameter in bkg fit to $\cos(\theta)$	0.366 ± 0.140
ϕ_1	First parameter in bkg fit to ϕ	0.183 ± 0.019
ϕ_2	Second parameter in bkg fit to ϕ	0.092 ± 0.039
$\cos(\psi)_1$	First parameter in bkg fit to $\cos(\psi)$	0.276 ± 0.150
$\cos(\psi)_2$	Second parameter in bkg fit to $\cos(\psi)$	-0.159 ± 0.165
f_1^s	Fraction of 1^{st} Gamma dist. in signal	0.783 ± 0.100
a_1^s	a-parameter, signal, 1^{st} Gamma dist.	10.5 ± 0.9
b_1^s	b-parameter, signal, 1^{st} Gamma dist.	1.70 ± 0.17 μm
a_2^s	a-parameter, signal, 2^{nd} Gamma dist.	6.71 ± 1.73
b_2^s	b-parameter, signal, 2^{nd} Gamma dist.	3.80 ± 0.64 μm
f_1^b	Fraction of 1^{st} Gamma dist. in bkg.	0.915 ± 0.013
a_1^b	a-parameter, bkg., 1^{st} Gamma dist.	11.8 ± 0.4
b_1^b	b-parameter, bkg., 1^{st} Gamma dist.	2.02 ± 0.07 μm
a_2^b	a-parameter, bkg., 2^{nd} Gamma dist.	2.72 ± 0.31
b_2^b	b-parameter, bkg., 2^{nd} Gamma dist.	11.9 ± 1.1 μm

Table 19: Fit parameters and results for the untagged measurement of $\Delta\Gamma$ in 1.55 fb^{-1} of data.

Parameter	Description	Fit result
β_s	CP asymmetry parameter	0.55 ± 0.28
$\Delta\Gamma$	CP asymmetry parameter [ps^{-1}]	0.12 ± 0.10
$\alpha_\perp$	CP odd fraction	0.215 ± 0.028
$\alpha_\parallel$	$A_\parallel$ fraction in CP even states	0.314 ± 0.026
$\phi_\perp$	arg$(A_\perp A_0^*)$ asymmetry parameter	3.11 ± 0.81
$\phi_\parallel$	arg$(A_\parallel A_0^*)$ asymmetry parameter	3.84 ± 0.18
Δm_s	B_s^0 mixing frequency(constrained)	17.73 ± 0.11
$S_\mathcal{D}(OST)$	OST dilution scale factor(constrained)	0.894 ± 0.079
$S_\mathcal{D}(SST1)$	SST dilution scale factor, period 1(constrained)	0.981 ± 0.125
$S_\mathcal{D}(SST2)$	SST dilution scale factor, period 2(constrained)	0.954 ± 0.124
$S_\mathcal{D}(SST3)$	SST dilution scale factor, period 3(constrained)	0.946 ± 0.124
$\epsilon_s(OST)$	OST tagging efficiency for signal	0.963 ± 0.004
$\epsilon_b(OST)$	OST tagging efficiency for background	0.980 ± 0.002
$\mathcal{A}^+(OST)$	OST background positive tag asymmetry	0.490 ± 0.007
$\epsilon_s(SST)$	SST tagging efficiency for signal	0.515 ± 0.012
$\epsilon_b(SST)$	SST tagging efficiency for background	0.746 ± 0.006
$\mathcal{A}^+(SST)$	SST background positive tag asymmetry	0.507 ± 0.008
m	B hadron mass [MeV/c^2]	5365.9 ± 0.2
f_s	Signal Fraction	0.259 ± 0.006
s_m	Mass error scale factor	1.571 ± 0.035
p_1	Background Slope	-2.44 ± 1.25
$c\tau$	Proper decay time	450.56 ± 18.3 μm
s_t	Proper decay time scale factor	1.262 ± 0.197
f_p	Prompt fraction of background	0.789 ± 0.015
λ_+	Effective background lifetime, pos. component 1	369 ± 47 μm
f_{++}	Fraction of bkg which decays w/λ_{++}	0.655 ± 0.040
λ_{++}	Effective background lifetime, pos. component 2	70 ± 9 μm
f_-	Fraction of bkg which decays w/λ_-	0.381 ± 0.058
λ_-	Effective background lifetime, neg. component	124 ± 15 μm
$\cos(\theta)_1$	First parameter in bkg fit to cos(θ)	0.523 ± 0.130
$\cos(\theta)_2$	Second parameter in bkg fit to cos(θ)	0.358 ± 0.148
ϕ_1	First parameter in bkg fit to ϕ	0.178 ± 0.020
ϕ_2	Second parameter in bkg fit to ϕ	0.088 ± 0.041
$\cos(\psi)_1$	First parameter in bkg fit to cos(ψ)	0.240 ± 0.158
$\cos(\psi)_2$	Second parameter in bkg fit to cos(ψ)	-0.122 ± 0.174
f_1^s	Fraction of 1^{st} Gamma dist. in signal	0.748 ± 0.145
a_1^s	a-parameter, signal, 1^{st} Gamma dist.	10.4 ± 1.1
b_1^s	b-parameter, signal, 1^{st} Gamma dist.	1.71 ± 0.21 μm
a_2^s	a-parameter, signal, 2^{nd} Gamma dist.	7.2 ± 2.2
b_2^s	b-parameter, signal, 2^{nd} Gamma dist.	3.46 ± 0.64 μm
f_1^b	Fraction of 1^{st} Gamma dist. in bkg.	0.909 ± 0.014
a_1^b	a-parameter, bkg., 1^{st} Gamma dist.	12.0 ± 0.4
b_1^b	b-parameter, bkg., 1^{st} Gamma dist.	1.98 ± 0.07 μm
a_2^b	a-parameter, bkg., 2^{nd} Gamma dist.	2.9 ± 0.3
b_2^b	b-parameter, bkg., 2^{nd} Gamma dist.	11.2 ± 1.0 μm

Table 20: Tagged fit results in $B_s^0 \to J/\psi\,\phi$ when the initial value of $\Delta\Gamma > 0$ in the fit, where Δm_s and the dilution scale factors are Gaussian constrained.

Parameter	Description	Fit result
β_s	CP asymmetry parameter	1.02 ± 0.30
$\Delta\Gamma$	CP asymmetry parameter [ps^{-1}]	-0.12 ± 0.11
$\alpha_\perp$	CP odd fraction	0.215 ± 0.028
$\alpha_\parallel$	$A_\parallel$ fraction in CP even states	0.314 ± 0.026
$\phi_\perp$	arg$(A_\perp A_0^*)$ asymmetry parameter	0.04 ± 0.86
$\phi_\parallel$	arg$(A_\parallel A_0^*)$ asymmetry parameter	2.45 ± 0.18
Δm_s	B_s^0 mixing frequency(constrained)	17.73 ± 0.12
$S_\mathcal{D}(OST)$	OST dilution scale factor(constrained)	0.894 ± 0.080
$S_\mathcal{D}(SST1)$	SST dilution scale factor, period 1(constrained)	0.981 ± 0.125
$S_\mathcal{D}(SST2)$	SST dilution scale factor, period 2(constrained)	0.954 ± 0.125
$S_\mathcal{D}(SST3)$	SST dilution scale factor, period 3(constrained)	0.946 ± 0.125
$\epsilon_s(OST)$	OST tagging efficiency for signal	0.963 ± 0.004
$\epsilon_b(OST)$	OST tagging efficiency for background	0.980 ± 0.002
$\mathcal{A}^+(OST)$	OST background positive tag asymmetry	0.490 ± 0.007
$\epsilon_s(SST)$	SST tagging efficiency for signal	0.515 ± 0.012
$\epsilon_b(SST)$	SST tagging efficiency for background	0.746 ± 0.006
$\mathcal{A}^+(SST)$	SST background positive tag asymmetry	0.507 ± 0.007
m	B hadron mass [MeV/c^2]	53659.3 ± 0.2
f_s	Signal Fraction	0.259 ± 0.006
s_m	Mass error scale factor	1.571 ± 0.035
p_1	Background Slope	-2.44 ± 1.25
$c\tau$	Proper decay time	450.5 ± 19.6 μm
s_t	Proper decay time scale factor	1.262 ± 0.020
f_p	Prompt fraction of background	0.789 ± 0.015
λ_+	Effective background lifetime, pos. component 1	369 ± 47 μm
f_{++}	Fraction of bkg which decays w/λ_{++}	0.655 ± 0.040
λ_{++}	Effective background lifetime, pos. component 2	70 ± 9 μm
f_-	Fraction of bkg which decays w/λ_-	0.381 ± 0.058
λ_-	Effective background lifetime, neg. component	124 ± 15 μm
$\cos(\theta)_1$	First parameter in bkg fit to $\cos(\theta)$	0.523 ± 0.130
$\cos(\theta)_2$	Second parameter in bkg fit to $\cos(\theta)$	0.358 ± 0.148
ϕ_1	First parameter in bkg fit to ϕ	0.178 ± 0.021
ϕ_2	Second parameter in bkg fit to ϕ	0.088 ± 0.041
$\cos(\psi)_1$	First parameter in bkg fit to $\cos(\psi)$	0.240 ± 0.158
$\cos(\psi)_2$	Second parameter in bkg fit to $\cos(\psi)$	-0.122 ± 0.174
f_1^s	Fraction of 1^{st} Gamma dist. in signal	0.748 ± 0.184
a_1^s	a-parameter, signal, 1^{st} Gamma dist.	10.4 ± 1.4
b_1^s	b-parameter, signal, 1^{st} Gamma dist.	1.71 ± 0.27 μm
a_2^s	a-parameter, signal, 2^{nd} Gamma dist.	7.2 ± 2.5
b_2^s	b-parameter, signal, 2^{nd} Gamma dist.	3.46 ± 0.65 μm
f_1^b	Fraction of 1^{st} Gamma dist. in bkg.	0.909 ± 0.014
a_1^b	a-parameter, bkg., 1^{st} Gamma dist.	12.0 ± 0.4
b_1^b	b-parameter, bkg., 1^{st} Gamma dist.	1.98 ± 0.07 μm
a_2^b	a-parameter, bkg., 2^{nd} Gamma dist.	2.9 ± 0.3
b_2^b	b-parameter, bkg., 2^{nd} Gamma dist.	11.2 ± 1.0 μm

Table 21: Tagged fit results in $B_s^0 \to J/\psi\,\phi$ when the initial value of $\Delta\Gamma < 0$ in the fit, where Δm_s and the dilution scale factors are Gaussian constrained.

projections are shown in Figures 63 and 64. The signal and background angular likelihood projections are shown in Figure 65 and Figure 66.

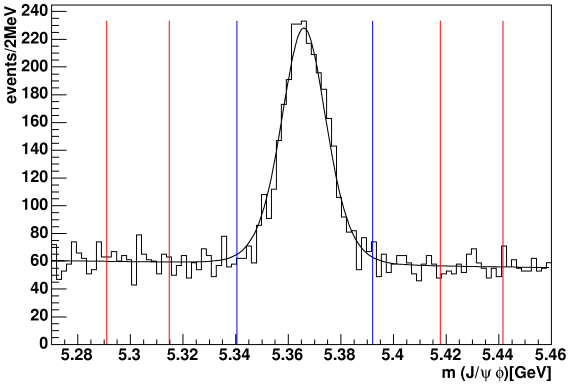

Figure 62: Mass likelihood projection with signal and sideband region marked when the initial value of $\Delta\Gamma > 0$ in the fit.

8.2.1 Likelihood Profile

As we concluded in chapter 7, we do not quote the numbers in Table 20 and Table 20 as our final result. Instead we study the likelihood profile in an effort to get the confidence regions following the procedure discussed in section 7.1.2. Since the ambiguity comes from the invariance of the likelihood function, which says in the four-dimensional space of β_s, $\Delta\Gamma$, $\phi_\parallel$ and $\phi_\perp$, two equivalent positions cannot be distinguished. However, we can try to break the symmetry by bounding one of the parameters. We cannot bound parameters β_s and $\Delta\Gamma$, which are our parameters of interest. This leaves us only the two strong phases to choose. As one can see in Table 20 and Table 21, parameter $\phi_\parallel$ is the best candidate because of its small error. We then try to bound $\phi_\parallel$ within the range $(0, \pi)$ first, which should in principle remove the invariance of the likelihood function. According to results in Table 21, only the minimum around $(\beta_s, \Delta\Gamma) = (1.02, -0.12)$ should be found. If $\phi_\parallel$ are bounded within $(\pi, 2\pi)$, we should find the other minimum.

To get the likelihood profile, the parameter β_s is divided into 40 bins between $-\pi/2$ and $+\pi/2$, and $\Delta\Gamma$ is divided into 20 bins between -0.7 ps^{-1} and +0.7 ps^{-1}, giving a total of 800

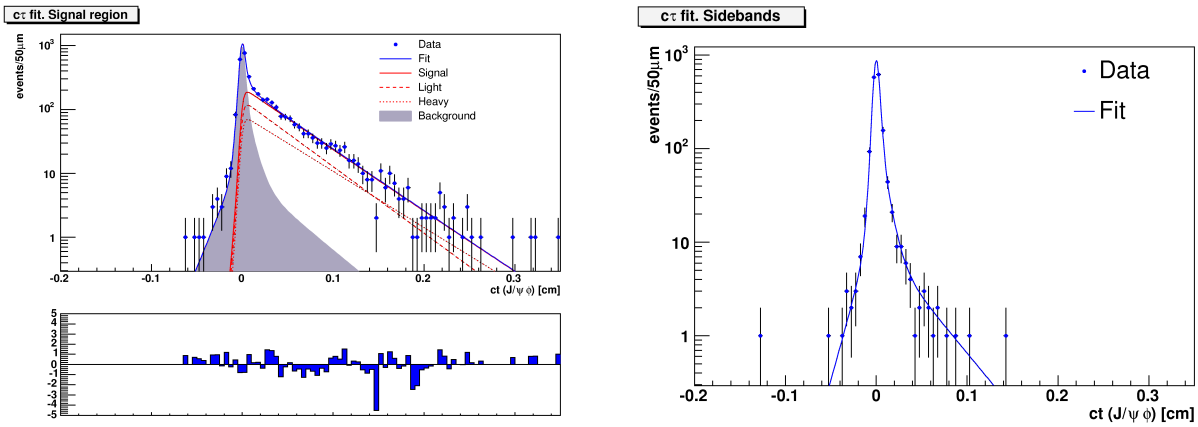

Figure 63: Proper decay time likelihood projection in signal region (left). The mass eigenstates are not same as CP eigenstates due to CP violation, the bottom plot shows residual in each bin divided by the estimated error. Proper decay time likelihood projection in sideband region (right).

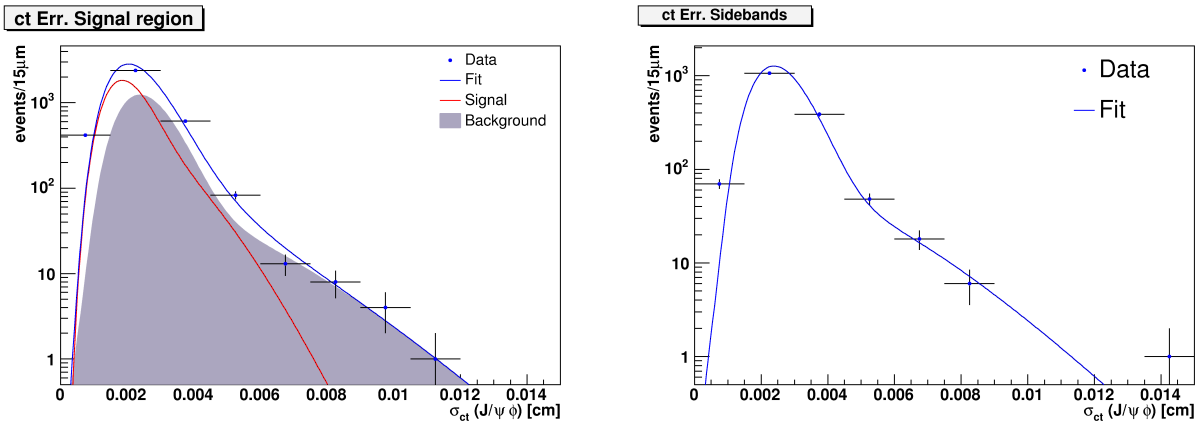

Figure 64: Proper decay time uncertainty likelihood projection in signal region (left). Proper decay time uncertainty likelihood projection in sideband region (right).

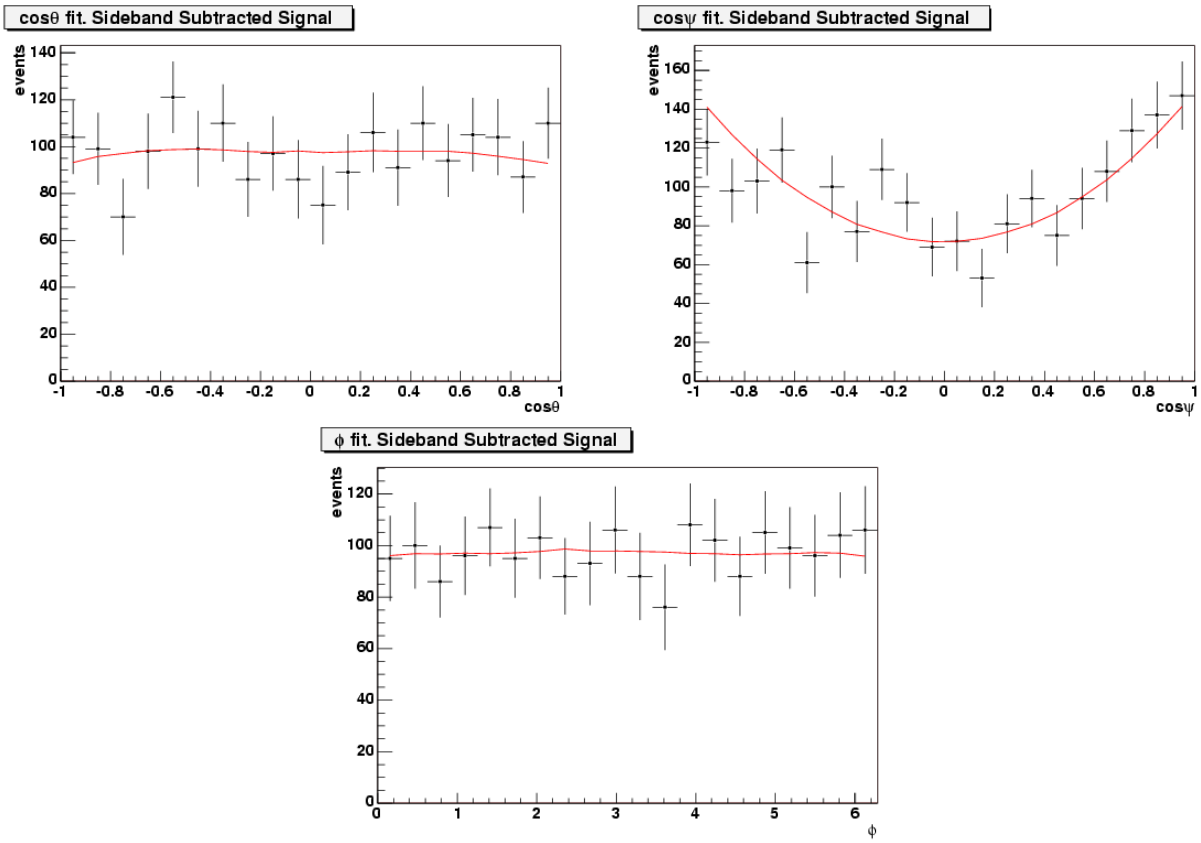

Figure 65: Sideband subtracted signal angular likelihood projections.

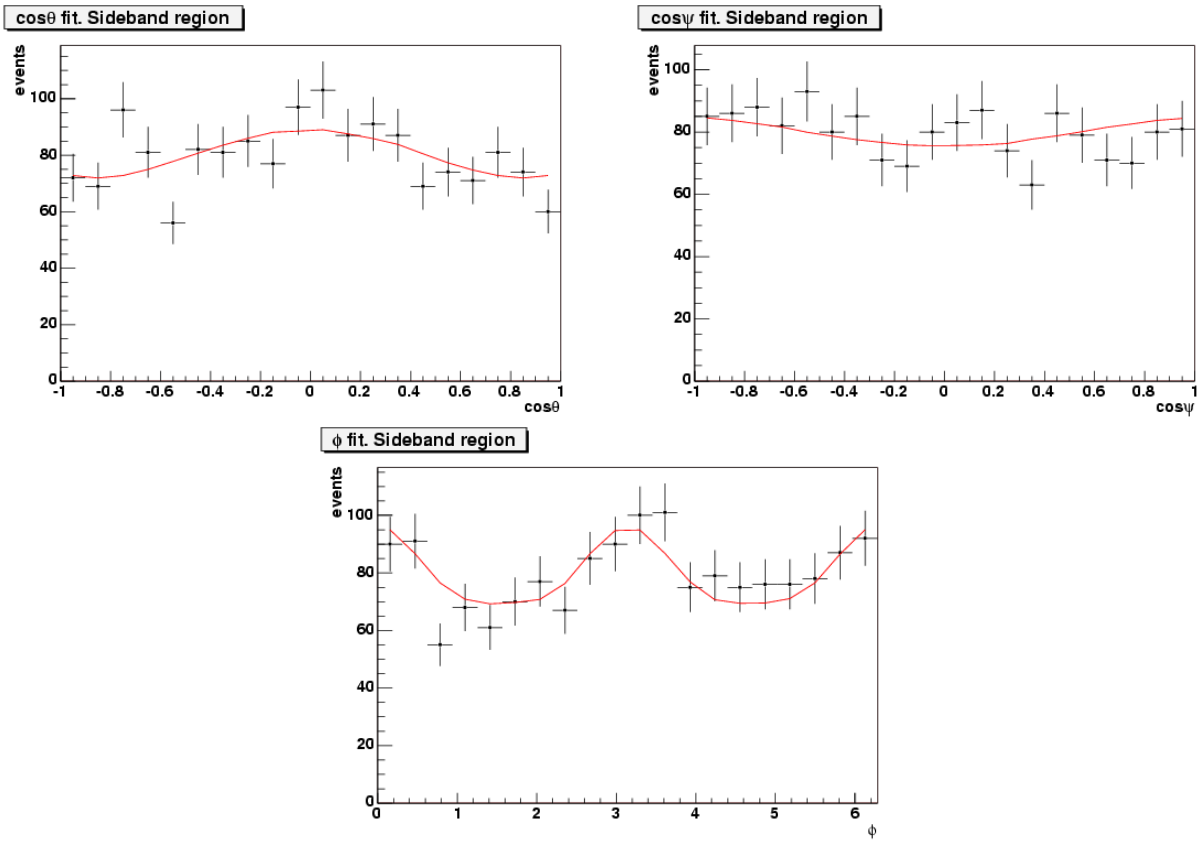

Figure 66: Background angular likelihood projections.

grid points in the two-dimensional space. The likelihood function is then minimized with respect to $N-2$ parameters at each grid point in order to obtain $-2\Delta log\mathcal{L}$ value. In the end grid points with likelihood ratios at $-2\Delta log\mathcal{L} = 2.30$ and $-2\Delta log\mathcal{L} = 5.99$ are connected to give 68% and 95% confidence levels (CL).

The likelihood profile with $\phi_\parallel$ bounded between 0 and π is shown in Figure 67. With $\phi_\parallel$ restricted between π and 2π, the likelihood profile is shown in Figure 68. Surprisingly, we still see two minima in those likelihood profiles clearly, but the depths are different. The explanation is that the likelihood function is approximately invariant with a simultaneous transformation of a subset of those four parameters. A closer examination of Eq. A.7 illustrates our limited ability to resolve the quadrant of β_s, or the sign of $\Delta\Gamma$. Even when restrictions are applied to $\phi_\parallel$ which lift the exact symmetry, resolution of the residual approximate symmetry relies crucially on the interference terms between the decay amplitudes. As the statistical power of our data is still fairly limited, we are not able to resolve those terms well enough to distinguish the two sectors. We have studied the expected behavior of these approximate invariances by comparing the numerical likelihood values with different combinations of the four separate transformations using toy Monte Carlo. The results show two kinds of general remaining approximate invariances : (1) an approximate ambiguity remains when $\beta_s \to \frac{\pi}{2} - \beta_s$ and $\Delta\Gamma \to -\Delta\Gamma$, provided β_s is not large, and (2) $\phi_\perp \to \pi - \phi_\perp$. For large values of β_s, approximate invariances also involve the above transformations in β_s, $\Delta\Gamma$, and some modulation of one of the strong phases, generally $\phi_\perp$.

Since restrictions on $\phi_\parallel$ do not help resolve the ambiguity, and approximate symmetries arise because of low statistics, we combine the two likelihood profiles in a way that at each grid point we always pick the deeper likelihood value. By doing this, we remove dependence on the strong phases, and the final likelihood profile shown in Figure 69 is just a confidence region of β_s and $\Delta\Gamma$, without conditions on $\phi_\parallel$.

A one-dimensional confidence interval for β_s alone can also be obtained by maximizing the likelihood function over all the parameters including $\Delta\Gamma$ except β_s. The scan range is $(-\pi/2, +\pi/2)$ with 100 bins between them. Two likelihood profiles are plotted with the parameter $\phi_\parallel$ restricted to different ranges respectively. The final profile is obtained by choosing larger likelihood value at every β_s grid point. All the one-dimensional likelihood

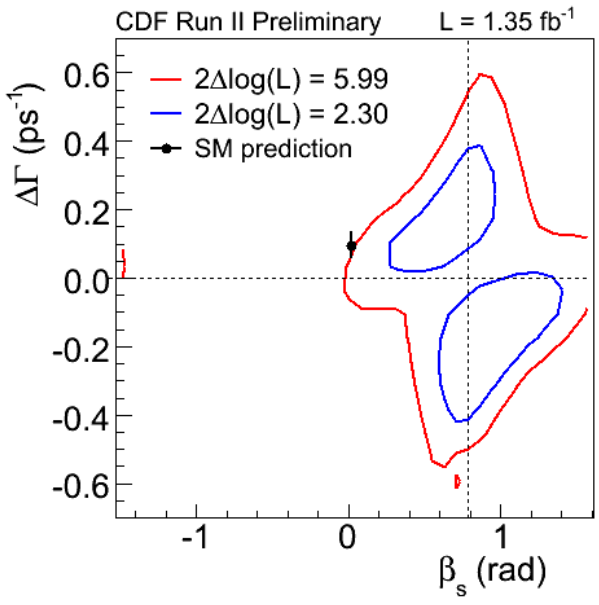

Figure 67: Two-dimensional likelihood profile of β_s and $\Delta\Gamma$, with parameter $\phi_{\|}$ is bounded in the range $(0, \pi)$.

profiles are shown in Figure 70.

For well-behaved parabolic log-likelihood functions, the confidence level can be determined directly from the likelihood value. The standard model point $(\beta_s, \Delta\Gamma) = (0.02, 0.096)$ is at the edge of 95% confidence level in the two-dimensional space, corresponding to a probability of 5% or 2 Gaussian standard deviations. However, since the likelihood profiles in this case have a highly non-parabolic shape and multiple minima, we do not expect the likelihood contour to give the correct confidence regions, which we therefore calculate with Monte Carlo pseudo-experiments.

8.2.2 Confidence Region with Feldman Cousins Method

A more robust frequentist approach originally suggested by Feldman and Cousins [64] is applied to obtain the correct confidence region. Numbers of pseudo-experiments are generated at each grid point in the $\beta_s - \Delta\Gamma$ space. Every pseudo-experiment is fitted twice with β_s

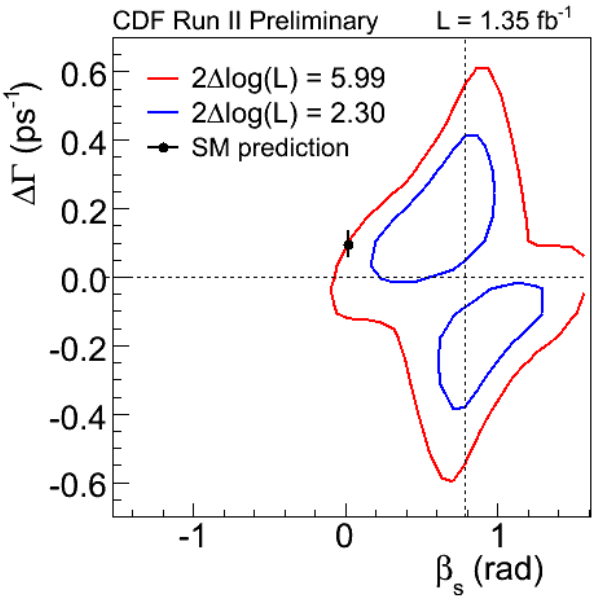

Figure 68: Two-dimensional likelihood profile of β_s and $\Delta\Gamma$, with parameter $\phi_{\|}$ is bounded in the range $(\pi,2\pi)$.

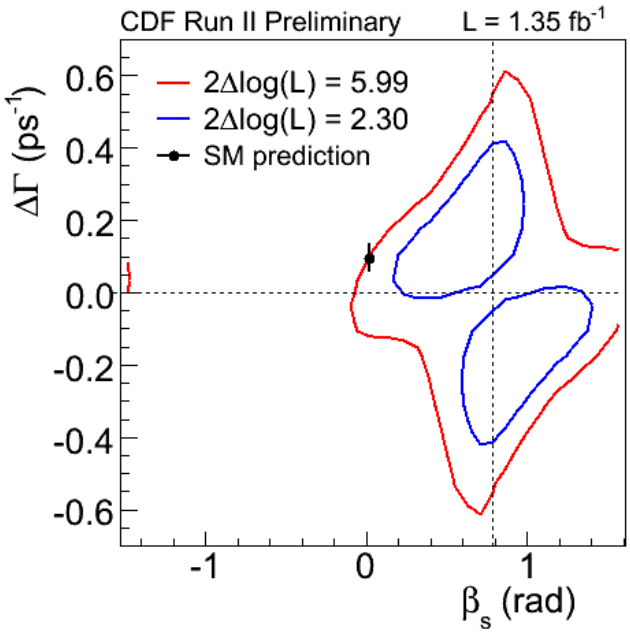

Figure 69: Symmetrized two-dimensional likelihood profile of β_s and $\Delta\Gamma$.

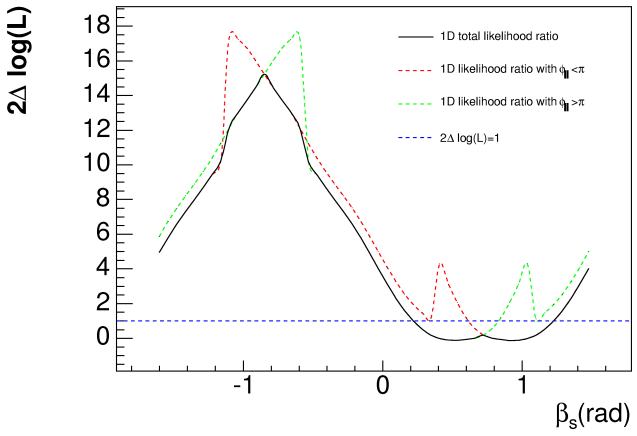

Figure 70: One-dimensional likelihood profile, where the $2\Delta log(L) = 1$ line gives 68% CL.

and $\Delta\Gamma$ floating in the fit, or fixed to the input values respectively, to get a likelihood ratio, which is defined as:

$$\mathcal{LR} \equiv -2log\frac{\mathcal{L}(\beta_s{}^*, \Delta\Gamma^*, \vec{\xi})}{\mathcal{L}(\hat{\beta_s}, \hat{\Delta\Gamma}, \vec{\xi})} = -2(log\mathcal{L}(\beta_s{}^*, \Delta\Gamma^*, \vec{\xi}) - log\mathcal{L}(\hat{\beta_s}, \hat{\Delta\Gamma}, \vec{\xi})), \tag{8.1}$$

where $\hat{\beta_s}$ and $\hat{\Delta\Gamma}$ indicate the values of β_s and $\Delta\Gamma$ that minimize the likelihood, while $\beta_s{}^*$ and $\Delta\Gamma^*$ indicate fixed to input values. $\vec{\xi}$ represents all other "nuisance" parameters which are always allowed to float in the fit.

At the end of this procedure, we obtain likelihood ratio distributions for all the grid points. Figure 71 shows the likelihood ratio distribution when $(\beta_s, \Delta\Gamma) = (0.550, -0.245)$ as an example. The likelihood ratio obtained by fitting data at the same grid point is also shown in the figure. By counting how many times the likelihood ratio of pseudo-experiments exceeds the one obtained from data, we get a p-value. In this specific case, the p-value is calculated to be 0.284. The corresponding confidence level is then

$$C.L. = 1 - p.$$

All the grid points which have p-values at $p = 32\%, 5\%$ are connected respectively to show the 68% and 95% confidence regions. The final result is shown in Figure 72, which indicates the standard model point has p-value at 12.7%.

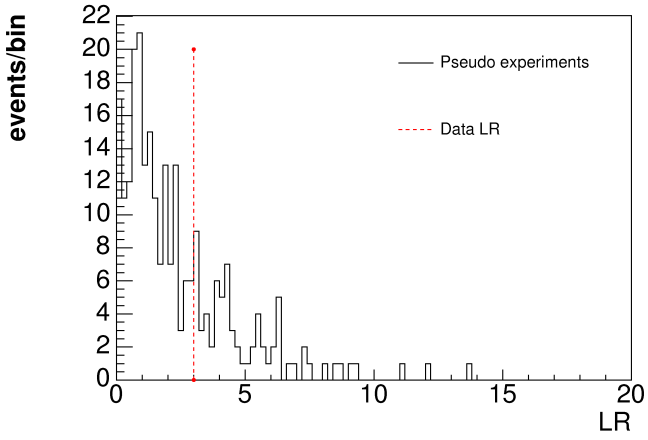

Figure 71: Likelihood ratio distribution example at one grid point. About 250 pseudo experiments are generated. The fraction of experiments with a larger likelihood ratio than the data sample (red line) is the p-value.

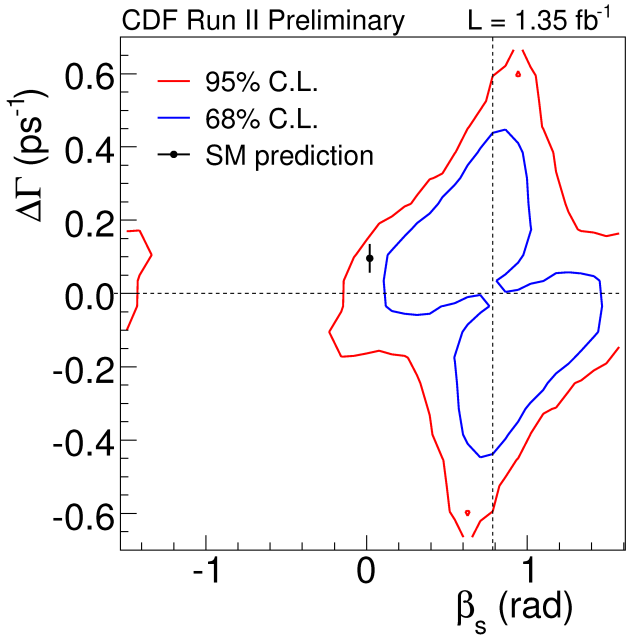

Figure 72: Confidence region obtained by Feldman Cousins method.

8.2.3 Systematic Studies

Although we are not quoting a single number as the final result, we can still try to evaluate the possible systematic uncertainties associated with the result from fitting the data, which are expected to be small compared with the dominant statistical uncertainties at this time. Several sources of systematic uncertainties are evaluated by running pseudo-experiments with β_s and $\Delta\Gamma$ generated at standard model values $\sim (0.02, 0.1)$. In all cases the systematic uncertainties are significantly smaller than the statistical error. The details for different kinds of systematic uncertainties are listed below, and the results are shown in Table 22.

- **Alignment of SVX**

  The alignment of the SVX affects the measurement of β_s by altering the length scale of the SVX detector. A different length scale affects the analysis through the mixing frequency Δm_s. We note that the Gaussian constraint on the mixing frequency [28], which comes from CDF exclusively, already includes this uncertainty.

- **Dilution scale factors**

  The dilution scale factors are allowed to float within their errors as Gaussian constraints in the fit. Therefore, any systematic effect due to imperfect knowledge of the scale factors is already included in the final errors assigned to the fit parameters from the fit package `Minuit`. In order to assess the level of systematic uncertainty, we fix all parameters which were floated within Gaussian constraints and observe almost no change in the final errors. To double check this, we perform the fit with the scale factors fixed to their upper and lower limits and determine the shift in the parameters. In all cases, the observed shifts are negligible.

- **Resolution scale factor**

  We consider the systematic uncertainty due to alternative models of the resolution. In order to do so, we fit the data with a two-Gaussian resolution function for signal and the prompt background, with a single Gaussian resolution function for all other background tails. Although both Gaussians are allowed to float in the fit, we observe that the fit strongly prefers a single Gaussian for the signal and prompt background resolution as well, and the fraction of the second Gaussian returned by a fit to data is consistent with

zero. In order to assess a systematic effect, we generate with a two-Gaussian resolution function in toy MC, where the fraction of the second Gaussian is 5%, which corresponds to the uncertainty in the fit to the data. The MC experiments are generated with a second scale factor of 2.8, which is also determined from the fit to data, while the first scale factor is 1.25. We then fit the toy MC experiments with the default single Gaussian resolution model to determine the systematic uncertainty.

- **Signal angular efficiency**

  Systematics due to the modeling of the signal angular efficiencies are determined by generating toy MC from the 3-dimensional histogram we find in the Monte Carlo and fitting with the default model, in which the efficiencies are parameterized by spherical harmonics and Legendre polynomials.

- **Background angular distributions**

  We assign a systematic uncertainty for the background angular distributions by generating toy MC with histogram from sideband region of data and fitting with the parameterization we use in data, described in section 5.1.2.

- B^0 **reconstructed as** B_s^0

  We find from the Monte Carlo sample of 10M events that 451 $B^0 \to J/\psi\, K^{*0}$ events generated according to phase space and reconstructed as $B_s^0 \to J/\psi\, \phi$ survive the selection. Since this corresponds to 0.4% of the phase space $B_s^0 \to J/\psi\, \phi$ events that pass this selection, we do not assign a systematic uncertainty for these remaining events.

- **Signal mass fit model**

  We examine the effect of the mass fit model by generating toy MC with a double Gaussian model with two mass error scale factors. The parameters for generating the toy are determined by fitting data with a double Gaussian and two mass error scale factors. To determine the systematic effect, we then fit the toy with our default single Gaussian model.

- **Background lifetime fit model**

  Our background lifetime fit model is evaluated by generating toy MC according to the lifetime distribution from the sidebands. The default fit model is used to fit the toy and assign the systematic uncertainty.

Systematics	β_s	$\Delta\Gamma$ [ps^{-1}]
Signal efficiency	0.0127	0.0046
Mass model	0.0273	0.0037
Resolution model	0.0117	-0.0038
Background lifetime	0.0015	-0.0227
Background angular distribution	0.0037	0.0031

Table 22: List of systematic uncertainties.

8.2.4 Frequentist Incorporation of Systematic errors

Since the final result is a two-dimensional confidence region, the systematic errors we obtained as listed in Table 22 cannot be applied directly, instead, a different method is needed to evaluate the systematic uncertainty.

When we obtain the confidence region, we also have the likelihood ratio distributions for different grid points from running pseudo-experiments. If all the distributions are plotted together, one notices that the shapes are very similar and nearly independent of the generating values of β_s and $\Delta\Gamma$. For a regular likelihood function, it is known that the likelihood ratio distribution should follow a χ^2 distribution with two degrees of freedom $\chi^2(2)$ for two parameters of interest. However, if the tail integral distribution of a $\chi^2(2)$ and the likelihood ratio distributions of all the grid points are put together, one finds that the likelihood ratio distributions we obtained have longer tail as shown in Figure 73. The longer tail explains why the confidence region shown in Figure 72 expands with respect to the corresponding likelihood profile as shown in Figure 69 which is obtained assuming $\chi^2(2)$ distribution.

The confidence region in $\beta_s - \Delta\Gamma$ space is actually a projection of multidimensional confidence region in a larger space that includes all the nuisance parameters. To exclude any specific $(\beta_s, \Delta\Gamma)$ region at some confidence level in a two-dimensional space, one should make sure that it can be also excluded with different values of all the other nuisance parameters. To check if the confidence region we obtained satisfies such a requirement, 25 grid points

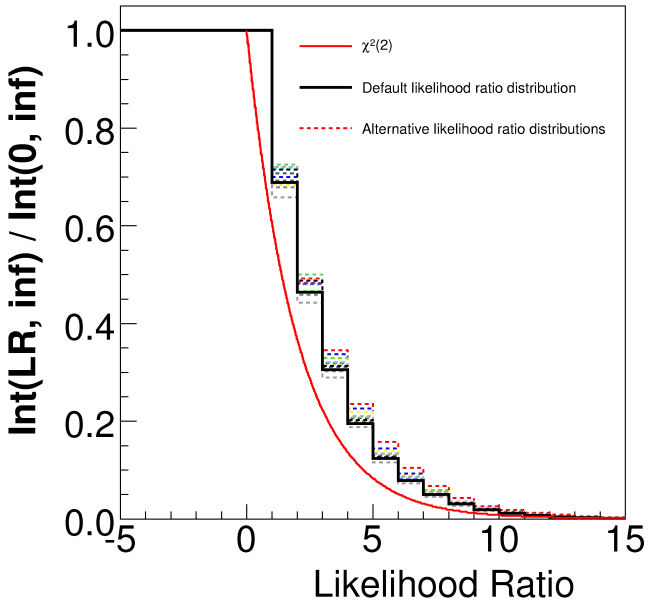

Figure 73: Tail integral of normalized likelihood ratio distribution, where x axis is the likelihood ratio, y axis gives the p-value of corresponding likelihood ratio. The red line represents a nominal likelihood ratio distribution following $\chi(2)$. The black histogram shows the behavior of the likelihood function used in this analysis. All the other 16 histograms come from varying nuisance parameters within 5σ's range.

distributed uniformly in the $\beta_s - \Delta\Gamma$ space are chosen to generate 16 more likelihood ratio distributions at each grid point. The number of points picked here are simply limited by CPU power. For each of these 16 alternative "universes", all the nuisance parameters are varied randomly within 5σ's range of the default values. The 16 likelihood ratio distributions of all the 25 grid points are plotted in Figure 73. The small variation of these histograms from alternative "universes" can be explained as a systematic effect. To further evaluate this effect, grid points along the default 95% confidence level in Figure 72 are used to do the same exercise. For each grid point along the line, the variance (RMS) of the p-values coming from 16 alternative "universes" is obtained. Then, the average value of these variances of all the grid points is calculated to be the systematic error, which is

$$< \sigma_p >= 2.3\%.$$

8.2.5 Final Confidence Region

With systematic error evaluated, the final confidence region is reconstructed by expanding the default confidence region with a universal 2.3% increase. In this way, the final 68% CL and 95% CL are just the 70.3% CL and 97.3% CL from the default confidence region, respectively. The final confidence region is shown in Figure 74, where the standard model point has a p-value 15% corresponding to 1.5 Gaussian standard deviations. The theoretical relationship between ϕ_s and $\Delta\Gamma$ from Eq. 2.40,

$$\Delta\Gamma = 2|\Gamma_{12}| \cos\phi_s,$$

which comes from the assumption of mixing-induced CP violation, is used to constrain space available for new physics. Since both phases ϕ_s and β_s are small, and are affected by new physics via q/p in Eq. 2.41, to a close approximation [65], we have

$$\phi_s = -2\beta_s. \tag{8.2}$$

This relationship also predicts $\Delta\Gamma$ for different β_s values, because Γ_{12} is precisely predicted in HQET (Eq. 2.33). As shown in Figure 75, the confidence region is very consistent with the assumption of mixing-induced CP violation.

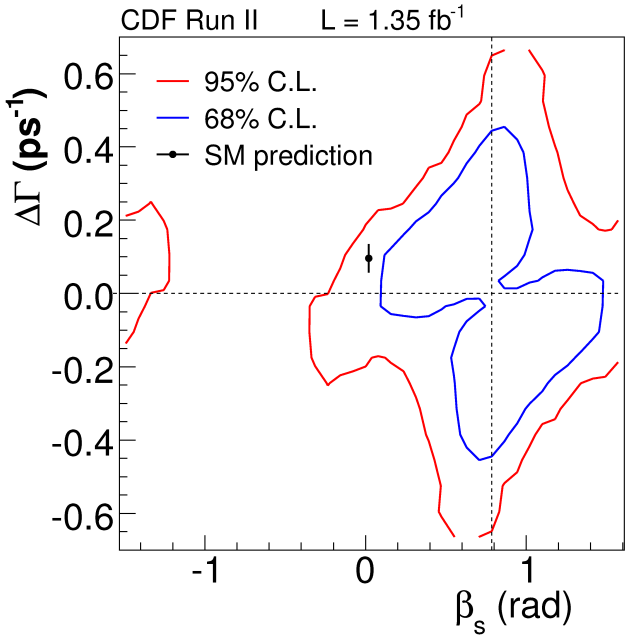

Figure 74: Confidence region with systematics included. The standard model prediction of $(\beta_s, \Delta\Gamma)$ has a p-value 15%.

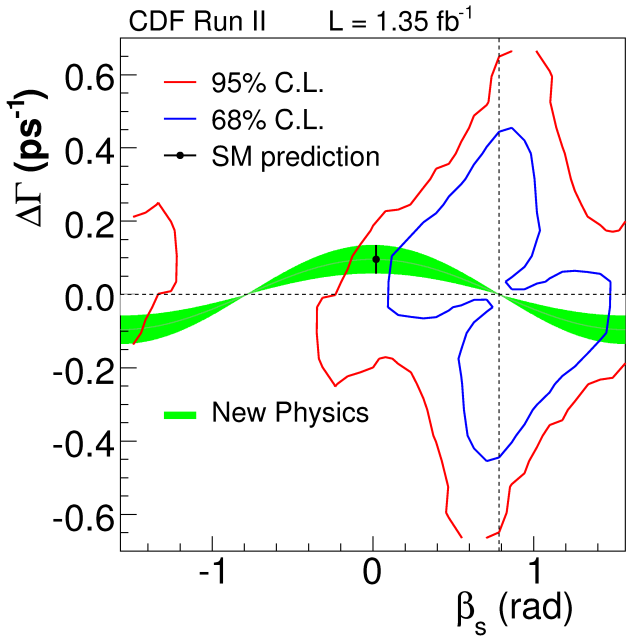

Figure 75: Confidence region with theoretical curve superimposed.

The one-dimensional confidence region for β_s alone is also obtained. The procedure is basically the same as the one uses to obtain the two-dimensional confidence region. The 68% confidence interval for β_s is obtained as $(0.16, 1.41)$.

8.2.6 Studies With External Constraints

It is interesting to apply some theoretical and experimental constraints when we fit the data. For these studies, results are mainly shown as likelihood profiles, which is good enough to give one some general idea about the possible improvements with external information.

The first constraint one can apply is the B_s^0 lifetime using the B^0 lifetime, according to the theoretical prediction $\tau_s/\tau_d = 1.00 \pm 0.01$ [65]. The lifetime is Gaussian constrained to the PDG 2006 B^0 lifetime value. The likelihood profile with this constraint is shown in Figure 76, where one can see that confidence region reduces especially on the $\Delta\Gamma$ direction. This is due to the strong correlation between the lifetime and the decay width difference.

If one assumes the two strong phases $\phi_{\|}$ and $\phi_{\perp}$ in $B_s^0 \to J/\psi\phi$ decay are the same as in the $B^0 \to J/\psi K^{*0}$ decay [1], the phases can be also constrained. With these constraints, the two-fold ambiguity can be resolved, because the likelihood function is not invariant anymore. The result is shown in Figure 77 with the strong phases constrained to the measured values as in Ref. [66]. As expected, there is only one minimum is in the profile.

If both lifetime and strong phases are constrained, one gets the likelihood profile as shown in Figure 78.

If $\Delta\Gamma$ is substituted according to the the theoretical relation $\Delta\Gamma = 2|\Gamma_{12}|\cos(2\beta_s)$ in the likelihood function with constraint $\Gamma_{12} = 0.048 \pm 0.018$ [22], we get one-dimensional confidence region using Feldman Cousins method for $\beta_s \in [0.12, 0.68] \cup [0.89, 1.45]$ at 68% confidence level. By applying additional lifetime and strong phase constraints, we get $\beta_s \in [0.2, 0.6]$ at 68% confidence level.

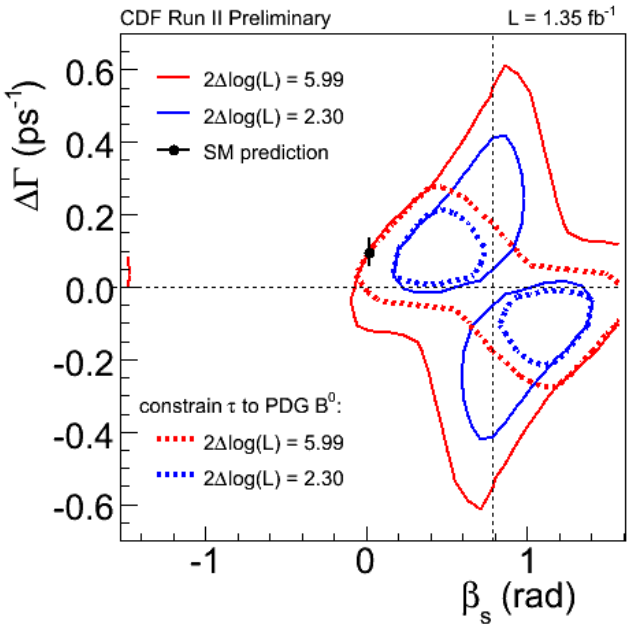

Figure 76: Likelihood profile with lifetime constrained to PDG B^0 lifetime.

[1] Questionable, since the K^{*0} is part of the $J^{PC} = 1^{++}$ octet, while the ϕ is an admixture of octet and nonet states.

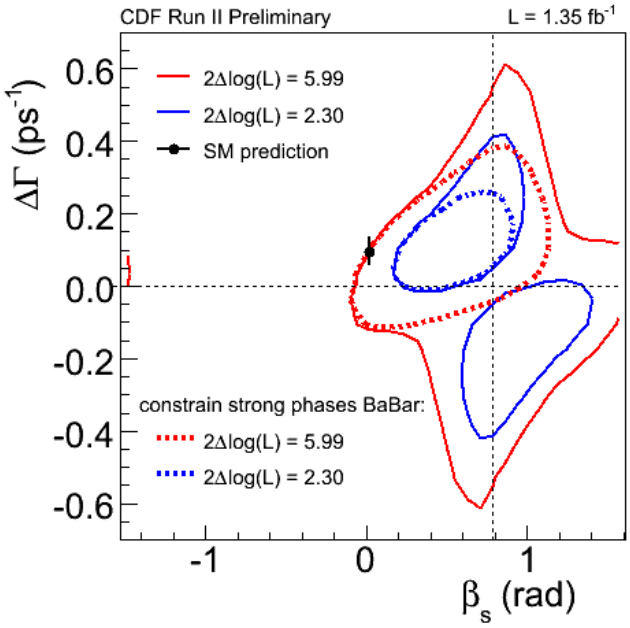

Figure 77: Likelihood profile with strong phases $\phi_{\parallel}$ and $\phi_{\perp}$ constrained to BaBar measurement result.

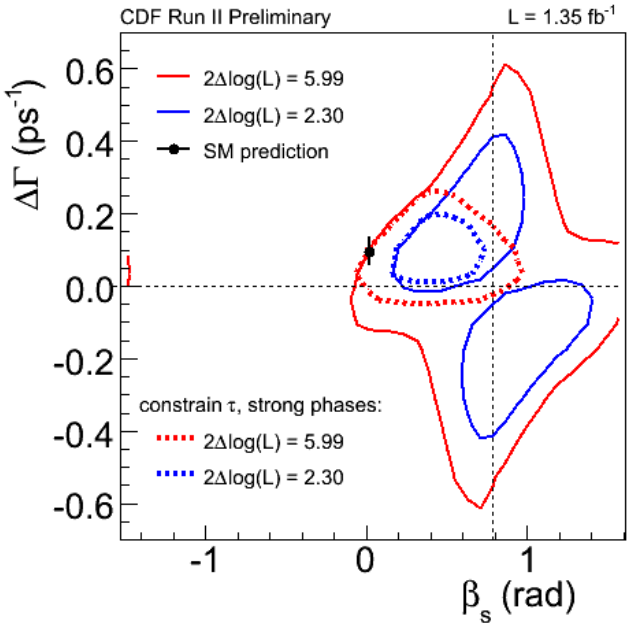

Figure 78: Likelihood profile with both lifetime and strong phases constrained.

8.3 CONCLUSION

The CP violation analysis carried out in this thesis is an indirect search for physics beyond the standard model. This is interesting in its own right, and could also have an impact beyond particle physics. In cosmology, for example, the CKM mechanism is said to fall short, by over 10 orders of magnitude, of producing the observed matter-antimatter imbalance of the universe. The chief result of this thesis is an indication of possible CP violation beyond the standard model, which is, therefore, of fundamental interest.

The constraint on β_s from other CKM matrix elements is very tight, and gives a precise prediction of β_s, in the standard model [16]

$$2\beta_s = 0.0368^{+0.0018}_{-0.0017}.$$

Measuring this angle is an ideal way to observe sources of CP violation beyond the standard model. The $B_s^0 \to J/\psi\phi$ channel we use to measure this angle is quite clean, i.e., free of

pollution from penguin diagrams. Because of its cleanliness, as well as the accuracy of the theoretical prediction, the measurement of β_s stands out from many other searches for CP violation. This work (also described in "First Flavor-Tagged Determination of Bounds on Mixing-Induced CP Violation in $B_s^0 \to J/\psi\phi$ Decays", published in PRL, April 2008) is the first time flavor-tagging has been applied to an analysis of the CP phase in the B_s^0 system.

The final result, which takes the form of a confidence region, reduces the space available for β_s to about half of the physically allowed region, and shows that the standard model predictions of β_s and $\Delta\Gamma$ together are consistent with what we observe in data *only at the 15% or 1.5 σ level*. The 68% confidence interval for β_s alone lies in the range $(+0.16, +1.41)$. With external information (see 8.2.6) the range can shrink to $(+0.2, +0.6)$ depending upon the constraints that are applied.

The same qualitative features are also observed by the D0 experiment at Fermilab in a similar analysis [67]. The combined CDF and D0 result is obtained by the Heavy Flavor Averaging Group [68], and is consistent with the standard model prediction for β_s and $\Delta\Gamma$ together at the level of 2.2 σ, or 3%. The combined 90% confidence level for β_s alone is within the range $(+0.14, +0.73) \bigcup (+0.82, +1.43)$. The combined two-dimensional contour is shown in Figure 79. It is interesting that both experiments observe a deviation in the same direction, although neither result is conclusive because of limited statistics. With the Tevatron running into the year 2010, CDF expects to get data up to 8 fb^{-1}. The sensitivity to β_s will be very much enhanced by the new data, even as analysis techniques, such as flavor tagging, continue to improve. In addition, various techniques to resolve the ambiguities have recently been proposed [69, 70].

The CDF and D0 searches for CP violation in $B_s^0 \to J/\psi\phi$ decays have generated interest in many new theoretical models. Some of these include new sources of CP violation coming from heavy particles not yet directly produced. One interesting hypothesis is a fourth generation quark, described in Ref. [71]. The measured CP violation in the B_s^0 system can be also used to constrain lepton flavor violation in a supersymmetric SU(5) theory [72]. A supersymmetric extension of the standard model might contain new sources of flavor and/or CP violation in its soft supersymmetry breaking terms. Such a theory could first reveal itself through large CP violation. Other possible models with large CP violation include

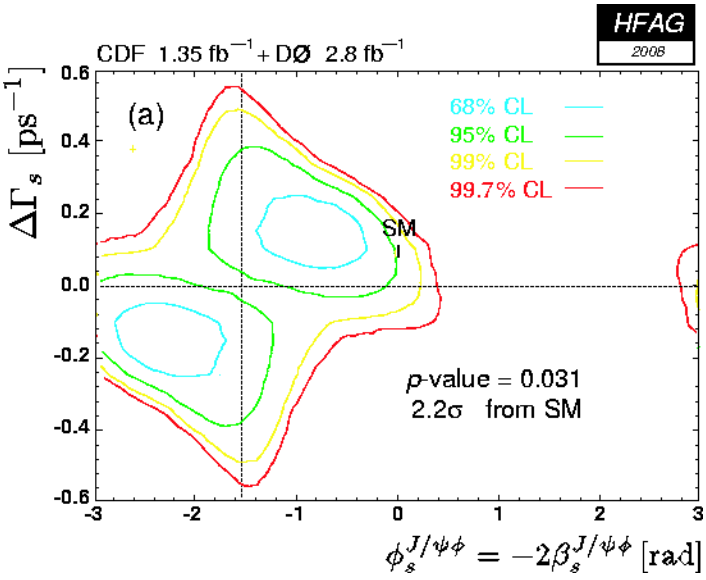

Figure 79: Combined two-dimensional contour from CDF and D0 results by HFAG.

extra dimensions [73], littlest Higgs model [74], and unparticle physics [75].

Still, it is too early to draw a definite conclusion from the current data samples. If the deviation survives more data at Fermilab or at the LHC, it would indicate at least one new source of CP violation.

BIBLIOGRAPHY

[1] C. L. Bennett et al. (WMAP Collaboration), Astrophys. J. Suppl. **148**, 1 (2003).

[2] A. D. Sakharov, "Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe," Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967) [JETP Lett. **5**, 24 (1967 SOPUA,34,392-393.1991 UFNAA,161,61-64.1991)].

[3] T. D. Lee and C. N. Yang, "Question Of Parity Conservation In Weak Interactions," Phys. Rev. **104**, 254 (1956).

[4] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, "EXPERIMENTAL TEST OF PARITY CONSERVATION IN BETA DECAY," Phys. Rev. **105**, 1413 (1957).

[5] J. H. Christenson *et al.*, Phys. Rev. Lett. **13**, 138 (1964).

[6] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973);
N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).

[7] D. A. Edwards, "The Tevatron: Introduction," *In *Batavia 1987/Ithaca 1988, Proceedings, Physics of particle accelerators* 1830-1844. (see Conference Index)*

[8] P. F. Harrison, "An Introduction to the BABAR experiment," Nucl. Instrum. Meth. A **368**, 81 (1995).

[9] S. L. Olsen [BELLE Collaboration], "The Belle Experiment At Kekb," *Prepared for 1994 Meeting of the American Physical Society, Division of Particles and Fields (DPF 94), Albuquerque, New Mexico, 2-6 Aug 1994.*

[10] B. Aubert *et al.*, [BABAR Collab.] Phys. Rev. Lett. **87**, 091801 (2001).

[11] K. Abe *et al.*, [Belle Collab.] Phys. Rev. Lett. **87**, 091802 (2001).

[12] G. Bernardi *et al.* [Tevatron New Phenomena Higgs Working Group and CDF Collaboration and D], arXiv:0808.0534 [hep-ex].

[13] Y. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **81**, 1562 (1998) [arXiv:hep-ex/9807003].

[14] P. Dita, "Parametrisation of unitary matrices", J. Phys. A: Math.Gen. **15** (1982) 3465-3473.

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

[16] http://www.slac.stanford.edu/xorg/ckmfitter/plots_fpcp07/ckmEval_results_fpcp07.html

[17] C. Jarlskog,"commutator of the quark Mass matrices in the Standard Electroweak Model and a measure of Maximal CP Nonconservation", Phys. Rev. Lett. 55, 1039-1042 (1985).

[18] C. Amsler et al., Physics Letters B **667**, 1 (2008).

[19] http://www.slac.stanford.edu/xorg/hfag/osc/end_2007/

[20] I. Dunietz, R. Fleischer, U. Nierste, Phys. Rev. D **63**, 114015 (2001) [arXiv:hep-ph/0012219].

[21] Colin Gay, hep-ex/0103016.

[22] A. Lenz and U. Nierste, J. High Energy Phys. **06**, 072 (2007) [arXiv:hep-ph/0612167].

[23] T.E.Browder and S.Pakvasa, Phys. Rev. D **52**, 3123 (1995).

[24] R.Fleischer, Eur. Phys. J. C**10** 299 (1999).

[25] A. Dighe, I. Dunietz, H.Lipkin, J.Rosner, hep-ph/9511363.

[26] A. Dighe, I. Dunietz, R. Fleisher, hep-ph/9804253.

[27] Bs Mixing Group, "Updated Measurement of Δm_d and Calibration of Flavor Tagging in Fully Reconstructed Decays", CDF note 7920.

[28] A. Abulencia *et al.* [CDF Collaboration], Phys. Rev. Lett. 97, 242003 (2006). FERMILAB-PUB-06-344-E.

[29] D. P. McGinnis, "Fermilab Tevatron status", C. Horak (ed.), proccedings of the Particle Accelerator Conference (PAC 05), (2005).

[30] http://www-bd.fnal.gov/runII/index.html
http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html

[31] R. Blair,*et. al.*, CDF-II Collaboration, FERMILAB-PUB-96-390-E, (1996).

[32] C. Hays, P. Tamburello, et al., "The COT Pattern Recognition Algorithm and Offline Code", CDF note 6992.

[33] W.R. Leo, "Techniques for Nuclear and Particle Physics Experiments", Springer, 1994

[34] S. Yu, J. Heinrich, et al., "COT dE/dx Measurement and Corrections", CDF note 6361.

[35] Ch.Paus, *et. al.*, "Design and Performance Tests of the CDF Time of Flight System", Nucl. Instrumen. Meth., vol. A461, pp.579-581, 2001.

[36] A. Artikov, et al., "Design and construction of new central and forward muon counters for CDF II", Nuclear Instruments and Methods in Physics Research A 538 (2005) 358-371.

[37] R.Downing, *et. al.*, "Track Extrapolation and Distribution for the CDF-II Trigger System", Nucl. Instrumen. Meth., vol. A570, pp.36-50, 2007.

[38] E. Ben-Haim, "The b quark fragmentation function, from LEP to TeVatron," FERMILAB-THESIS-2004-50.

[39] http://www.slac.stanford.edu/xorg/hfag/results/index.html

[40] J. Pumlin, D. R. Stump, and W. L. Tung, Phys. Rev. D **65**, 014011 (2002); D. R. Stump *et al.*, Phys. Rev. D **65**, 014012 (2002); J. Pumplin *et al.*, Phys. Rev. D **65**, 014013 (2002).
P. Nason, S. Dawson, and R.K. Ellis, Nucl. Phys. B**303**, 607(1988); Nucl. Phys. **B327**, 49 (1989); Nucl. Phys. B**335**, 260(1990).

[41] D. Acosta *et al.* [CDF Collaboration], " Measurement of the J/psi meson and b-hadron production cross sections in p[overline p] collisions at sqrt(s)=1960 GeV", Phys. Rev. D**71**, 032001 (2005).

[42] D. Acosta et al. (CDF Collaboration), "Measurement of the J/ψ meson and b-hadron production cross sections in $p\bar{p}$ collisions at sqrt(s)=1960 GeV", Phys. Rev. D**71**, 032001 (2005).

[43] S. Lai, J. Guimaraes da Costa, W. Wagner, P. K. Sinervo, R. Culbertson, "An updated measurement of the beam width at CDF", CDF note 6492.

[44] Kevin Burkett, Joao Guimaraes da Costa, "Development of a CTVMFT-based primary vertex finder: PrimeVtxFinder" CDF note 6430.

[45] B_s mixng group, "Performance of the event-by-event primary vertex finder on B physics data at CDF", CDF note 7512.

[46] J. Marriner, "Secondary Vertex Fit with Mass and Pointing Constraints (CTVMFT)", CDF internal Note 2724 (1994).

[47] W.-M. Yao *et al*, J. Phys. G**33**, 1 (2006).

[48] J.Boudreau,J.P.Fernandez, L.R.Flores-Castillo,L.Labarga, M.Spezziga, "Systematic studies on lifetimes of B mesons using the exclusive J/ψ decay modes", CDF note 7796.

[49] K. Anikeev, D. Smirnov, ROOT interface to SNNS, CDF note 7999.

[50] S. Gromoll, Ch. Paus, "A Heavy Quark Generator", CDF note 5985.

[51] CDF Collaboration (D. Acosta et al.), "Measurement of the J/psi meson and b-hadron production cross sections in p anti-p collisions at s(NN)**(1/2) = 1960-GeV", Phys.Rev.D71:032001,2005.

[52] William Bell, J.Pablo Fernandez, Luis Flores, Frank Wuerthwein, R.J. Tesarek,"User Guide For EvtGen at CDF", CDF/DOC/BOTTOM/CDFR/5618, April 2001.

[53] D. Usynin, M. Jones, J. Kroll, B. Wicklund, "B Flavor Identification Using Opposite Side Muons", CDF note 6483;
G. Giurgiu, V. Tiwari, M. Paulini, J. Russ, "Muon B Flavor Tagging - A Likelihood Approach", CDF note 7043;
G. Giurgiu, V. Tiwari, M. Paulini, J. Russ, M. Tanaka, "Updated Likelihood Muon Tagger", CDF note 7644.

[54] V. Tiwari, G. Giurgiu, M. Paulini, J. Russ, B. Wicklund, "Likelihood Based Electron Tagging", CDF note 7121.

[55] G. Bauer *et al.*, "Improved Jet Charge Tagger for summer conferences 2004", CDF note 7131.

[56] A. Belloni *et al.*, "Same Side Kaon Tagging Studies in Fully Reconstructed Decays", CDF note 7979;
Bs Mixing Group, "Same side Kaon Tagging Studies in Fully Reconstructed Decays on the New Data", CDF note 8182;
K. Anikeev *et al.*, "Optimization of the Same Side Kaon Tagging Algorithm combining PID and kinematic variables", CDF note 8344.

[57] C. Liu [CDF Collaboration and D0 Collaboration], "Recent Lifetime and Mixing Measurements at the Tevatron," arXiv:0806.4786 [hep-ex].

[58] http://www-cdf.fnal.gov/physics/new/bottom/060831.blessed-Bsmix/plots/all_sst_unblind_ampscan.gif

[59] J. Boudreau, J.P. Fernandez, L. Labarga, M. Neubauer, R. Madrak. C. Liu, "Measurement of Exclusive B Lifetimes in the modes $B^+ \to J/\psi\, K^+$, $B^0 \to J/\psi\, K^{*0}$, $B^0 \to J/\psi\, K^0_s$, $B^0_s \to J/\psi\, \phi$ and $\Lambda_b \to J/\psi\, \Lambda$, and measurement of the B^+/B^0 $B^0_s/B^0$and Λ_b/B^0 lifetime ratios.", CDF note 8524.

[60] Cramér H., "Mathematical methods of statistics", Princeton, NJ: Princeton University Press, 1946. 500p.

[61] H. G. Moser and A. Roussarie, Nucl. Instrum. Methods **384**, 491 (1997).

[62] D. Acosta *et al.* [CDF Collaboration], Phys. Rev. Lett. **94**, 101803 (2005) [arXiv:hep-ex/0412057].

[63] T. Aaltonen *et al.* [CDF collaboration], "Measurement of Lifetime and Decay-Width Difference in B0s -¿ J/psi phi Decays," Phys. Rev. Lett. **100**, 121803 (2008) [arXiv:0712.2348 [hep-ex]].

[64] G. J. Feldman and R. D. Cousins, *Phys. Rev. D* **57**,3873 (1998).

[65] A. Lenz, "Theoretical status of B_s-mixing and lifetimes of heavy hadrons," Nucl. Phys. Proc. Suppl. **177-178**, 81 (2008) [arXiv:0705.3802 [hep-ph]].

[66] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. D **71**, 032005 (2005) [arXiv:hep-ex/0411016].

[67] V. M. Abazov *et al.* [D0 Collaboration], "Measurement of B_s^0 mixing parameters from the flavor-tagged decay $B_s^0 \to J/\psi\phi$," arXiv:0802.2255 [hep-ex].

[68] http://www.slac.stanford.edu/xorg/hfag/osc/end_2007/#BETAS

[69] J. K. Parry, "The large CP phase in B(s) - anti-B(s) mixing and Unparticle Physics," arXiv:0806.4350 [hep-ph].

[70] M. Gronau and J. L. Rosner, "Flavor symmetry for strong phases and determination of $\beta_s, \Delta\Gamma$ in $B_s \to J/\psi\phi$," arXiv:0808.3761 [hep-ph].

[71] Wei-Shu Hou, "Source of CP Violation for Baryon Asymmetry of the Universe", arXiv:hep-ex/0803.1234.

[72] J. h. Park and M. Yamaguchi, "B_s mixing phase and lepton flavor violation in supersymmetric SU(5)," arXiv:0809.2614 [hep-ph].

[73] J. K. Parry, arXiv:0806.4350 [hep-ph].

[74] M. Blanke, A. J. Buras, S. Recksiegel and C. Tarantino, "The Littlest Higgs Model with T-Parity Facing CP-Violation in B_s - anti-B_s Mixing," arXiv:0805.4393 [hep-ph]

[75] J. K. Parry, "The large CP phase in B(s) - anti-B(s) mixing and Unparticle Physics," arXiv:0806.4350 [hep-ph].

[76] http://www-cdf.fnal.gov/internal/physics/bottom/bpak/goodrun/goodrun.sql

APPENDIX A

EXPLICIT TIME AND ANGULAR DEPENDENCE

We have expanded the relations in section 2.6.2 to provide the explicit time and angular dependence for ease of comparison with other work. The example of geting probability function for B_s^0 is given, which is similar to get $\bar{B}_s^0$. Starting with the relations for B_s^0, Eq. 2.78, we find

$$|\mathbf{A}_+(0)\times\hat{n}|^2 = |A_0(0)|^2\cos^2\psi(\cos^2\theta+\sin^2\theta\sin^2\phi)+\frac{1}{2}|A_\parallel(0)|^2\sin^2\psi(\cos^2\theta+\sin^2\theta\cos^2\phi) \\ +\frac{1}{2\sqrt{2}}|A_0(0)||A_\parallel(0)|\cos(\phi_\parallel)\sin 2\psi\sin^2\theta\sin 2\phi), \tag{A.1}$$

for the CP even states, where we take the real part of $|A_0(0)||A_\parallel(0)|$, and

$$|\mathbf{A}_-(0)\times\hat{n}|^2 = \frac{1}{2}|A_\perp(0)|^2\sin^2\psi\sin^2\theta, \tag{A.2}$$

for the CP odd state. The expression for the intereference term between CP even and odd states is

$$\begin{aligned}(\mathbf{A}_+\times\hat{n})\cdot(\mathbf{A}_-^*\times\hat{n}) &= \frac{i}{4}|A_\parallel(0)||A_\perp(0)|e^{i(\phi_\parallel-\phi_\perp)}\sin^2\psi\sin 2\theta\sin\phi \\ &\quad +\frac{i}{4\sqrt{2}}|A_0(0)||A_\perp(0)|e^{-i\phi_\perp}\sin 2\psi\sin 2\theta\cos\phi. \end{aligned}\tag{A.3}$$

Note that the angular part of Eq. A.1 can be written, using the relations

$$\begin{aligned}\cos^2\theta + \sin^2\theta\sin^2\phi &= 1 - \sin^2\theta + \sin^2\theta(1-\cos^2\phi) \\ &= 1 - \sin^2\theta\cos^2\phi, \qquad \text{(A.4)} \\ \cos^2\theta + \sin^2\theta\cos^2\phi &= 1 - \sin^2\theta + \sin^2\theta(1-\sin^2\phi) \\ &= 1 - \sin^2\theta\sin^2\phi, \qquad \text{(A.5)}\end{aligned}$$

so we can define the following angular functions

$$\begin{aligned} g_1(\vec{\omega}) &= \cos^2\psi(1-\sin^2\theta\cos^2\phi) \\ g_2(\vec{\omega}) &= \frac{1}{2}\sin^2\psi(1-\sin^2\theta\sin^2\phi) \\ g_3(\vec{\omega}) &= \frac{1}{2}\sin^2\psi\sin^2\theta \\ g_4(\vec{\omega}) &= \frac{1}{2\sqrt{2}}\sin 2\psi\sin^2\theta\sin 2\phi \\ g_5(\vec{\omega}) &= -\frac{1}{2}\sin^2\psi\sin 2\theta\sin\phi \\ g_6(\vec{\omega}) &= \frac{1}{2\sqrt{2}}\sin 2\psi\sin 2\theta\cos\phi, \end{aligned}$$

where $\vec{\omega} \equiv (\theta, \phi, \psi)$. We can then write the time-dependent probability as

$$\begin{aligned} P(t,\vec{\omega}) \propto\ & \{|A_0(0)|^2|f_+(t)|^2 g_1(\vec{\omega}) + |A_\parallel(0)|^2|f_+(t)|^2 g_2(\vec{\omega}) + |A_\perp(0)|^2|f_-(t)|^2 g_3(\vec{\omega}) \\ & + |A_0(0)||A_\parallel(0)|\cos(\phi_\parallel)|f_+(t)|^2 g_4(\vec{\omega}) \\ & + \mathcal{R}e\{i|A_\parallel(0)||A_\perp(0)|e^{i(\phi_\parallel-\phi_\perp)}f_+(t)f_-^*(t)\}g_5(\vec{\omega}) \\ & + \mathcal{R}e\{i|A_0(0)||A_\perp(0)|e^{i\phi_\perp}f_+(t)f_-^*(t)\}g_6(\vec{\omega})\}. \qquad \text{(A.6)}\end{aligned}$$

The time dependence of the last two terms in the above expression need to be explicitly written out with Eq. 2.82 in order to extract the real part of the above expression. Doing

so gives

$$\begin{aligned}
P(t,\vec{\omega}) \;\propto\; & |A_0(0)|^2|f_+(t)|^2 g_1(\vec{\omega}) + |A_\parallel(0)|^2|f_+(t)|^2 g_2(\vec{\omega}) + |A_\perp(0)|^2|f_-(t)|^2 g_3(\vec{\omega}) \\
& + |A_0(0)||A_\parallel(0)|\cos(\phi_\parallel)|f_+(t)|^2 g_4(\vec{\omega}) \\
& + |A_\parallel(0)||A_\perp(0)|[\sin(\phi_\parallel - \phi_\perp)e^{-\Gamma t}\cos(\Delta m t) \\
& \qquad + \cos(\phi_\parallel - \phi_\perp)(\cos(2\beta_s)e^{-\Gamma t}\sin(\Delta m t) + \sin(2\beta_s)(e^{-\Gamma_L t} - e^{-\Gamma_H t})/2)]g_5(\vec{\omega}) \\
& + |A_0(0)||A_\perp(0)|[\sin(\phi_\perp)e^{-\Gamma t}\cos(\Delta m t) \\
& \qquad - \cos(\phi_\perp)(\cos(2\beta_s)e^{-\Gamma t}\sin(\Delta m t) + \sin(2\beta_s)(e^{-\Gamma_L t} - e^{-\Gamma_H t})/2)]g_6(\vec{\omega}),
\end{aligned} \tag{A.7}$$

In order to facillitate comparison between these relations and [20], note that

$$\begin{aligned}
\frac{e^{-\Gamma_L t} - e^{-\Gamma_H t}}{2} &= e^{-\Gamma t}\frac{e^{-\frac{\Delta\Gamma t}{2}} - e^{+\frac{\Delta\Gamma t}{2}}}{2} \\
&= -e^{-\Gamma t}\sinh(\frac{\Delta\Gamma t}{2}),
\end{aligned} \tag{A.8}$$

and that [20] defines

$$\delta_1 \equiv Arg(A_\parallel^*(0)A_\perp(0)) = \phi_\perp - \phi_\parallel \tag{A.9}$$

$$\delta_2 \equiv Arg(A_0^*(0)A_\perp(0)) = \phi_\perp. \tag{A.10}$$

Making these substitutions and replace $|f_\pm(t)|^2$ with Eq. 2.80, we find:

$$\begin{aligned}
P(t,\vec{\omega}) \;\propto\; & |A_0(0)|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m t\right] g_1(\vec{\omega}) \\
& + |A_\parallel(0)|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m t\right] g_2(\vec{\omega}) \\
& + |A_\perp(0)|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t + \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m t\right] g_3(\vec{\omega}) \\
& + |A_0(0)||A_\parallel(0)|\cos(\delta_2 - \delta_1) \times \\
& e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m t\right] g_4(\vec{\omega}) \\
& + |A_\parallel(0)||A_\perp(0)|e^{-\Gamma t} \times \\
& \left[\sin(\delta_1)\cos(\Delta m t) - \cos(\delta_1)\cos(2\beta_s)\sin(\Delta m t) + \cos(\delta_1)\sin(2\beta_s)\sinh(\frac{\Delta\Gamma t}{2})\right] g_5(\vec{\omega}) \\
& + |A_0(0)||A_\perp(0)|e^{-\Gamma t} \times \\
& \left[\sin(\delta_2)\cos(\Delta m t) - \cos(\delta_2)\cos(2\beta_s)\sin(\Delta m t) + \cos(\delta_2)\sin(2\beta_s)\sinh(\frac{\Delta\Gamma t}{2})\right] g_6(\vec{\omega}).
\end{aligned} \tag{A.11}$$

With same proceduce, probability density function for $\bar{B}_s^0$ can be obtained:

$$\begin{aligned}
\bar{P}(t,\vec{\omega}) \quad \propto \quad & |A_0(0)|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m t\right] g_1(\vec{\omega}) \\
& + |A_\parallel(0)|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m t\right] g_2(\vec{\omega}) \\
& + |A_\perp(0)|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t + \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m t\right] g_3(\vec{\omega}) \\
& + |A_0(0)||A_\parallel(0)|\cos(\delta_2-\delta_1)\times \\
& e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m t\right] g_4(\vec{\omega}) \\
& + |A_\parallel(0)||A_\perp(0)| e^{-\Gamma t}\times \\
& \left[-\sin(\delta_1)\cos(\Delta m t) + \cos(\delta_1)\cos(2\beta_s)\sin(\Delta m t) + \cos(\delta_1)\sin(2\beta_s)\sinh(\frac{\Delta\Gamma t}{2})\right] g_5(\vec{\omega}) \\
& + |A_0(0)||A_\perp(0)| e^{-\Gamma t}\times \\
& \left[-\sin(\delta_2)\cos(\Delta m t) + \cos(\delta_2)\cos(2\beta_s)\sin(\Delta m t) + \cos(\delta_2)\sin(2\beta_s)\sinh(\frac{\Delta\Gamma t}{2})\right] g_6(\vec{\omega}).
\end{aligned} \tag{A.12}$$

Both of the PDFs agree with [20] up to the transformation $2\beta_s \to -\phi_s$.

APPENDIX B

ANALYTICAL NORMALIZATION

In order to simplify the integration over the angles, we observe that

$$\sin\theta\cos\phi = \sqrt{\frac{2\pi}{3}}(Y_1^{-1} - Y_1^1),$$
$$\sin\theta\sin\phi = \sqrt{\frac{2\pi}{3}}i(Y_1^{-1} + Y_1^1),$$
$$\cos\theta = \sqrt{\frac{4\pi}{3}}Y_1^0.$$

These terms do not appear directly in the probability density, however their products do

$$\begin{aligned}
\sin^2\theta\cos^2\phi = \frac{2\pi}{3}((Y_1^{-1})^2 + (Y_1^1)^2 - 2Y_1^{-1}Y_1^1) &= \sqrt{\frac{4\pi}{9}}Y_0^0 - \sqrt{\frac{4\pi}{45}}Y_2^0 + \sqrt{\frac{2\pi}{15}}(Y_2^{-2} + Y_2^2),\\
\sin^2\theta\sin^2\phi = -\frac{2\pi}{3}((Y_1^{-1})^2 + (Y_1^1)^2 + 2Y_1^{-1}Y_1^1) &= \sqrt{\frac{4\pi}{9}}Y_0^0 - \sqrt{\frac{4\pi}{45}}Y_2^0 - \sqrt{\frac{2\pi}{15}}(Y_2^{-2} + Y_2^2),\\
\cos^2\theta = \frac{4\pi}{3}(Y_1^0)^2 &= \sqrt{\frac{4\pi}{9}}Y_0^0 + \sqrt{\frac{16\pi}{45}}Y_2^0,\\
\sin^2\theta\cos\phi\sin\phi = i\frac{2\pi}{3}((Y_1^{-1})^2 - (Y_1^1)^2) &= i\sqrt{\frac{2\pi}{15}}(Y_2^{-2} - Y_2^2),\\
\sin\theta\cos\theta\cos\phi = \sqrt{(}2)\frac{2\pi}{3}(Y_1^0Y_1^{-1} - Y_1^0Y_1^1) &= \sqrt{\frac{2\pi}{15}}(Y_2^{-1} - Y_2^1),\\
\sin\theta\cos\theta\sin\phi = \sqrt{(}2)i\frac{2\pi}{3}(Y_1^0Y_1^{-1} + Y_1^0Y_1^1) &= i\sqrt{\frac{2\pi}{15}}(Y_2^{-1} + Y_2^1).
\end{aligned}$$

Bringing the above expressions into Eq. 5.12, and define

$$n_1 \equiv \tau_L + \tau_H + \cos 2\beta_s(\tau_L - \tau_H)$$
$$n_2 \equiv \tau_L + \tau_H - \cos 2\beta_s(\tau_L - \tau_H)$$
$$n_3 \equiv \sin 2\beta_s(\tau_L - \tau_H),$$

we find

$$\begin{aligned}
N' \;=\; & \int\int\int \epsilon(\psi,\theta,\phi)d(\cos\psi)d(\cos\theta)d\phi \\
& \frac{|A_\perp|^2}{2}n_2\left[\sqrt{\frac{16\pi}{9}}Y_0^0 - \sqrt{\frac{16\pi}{45}}Y_2^0\right]\sin^2\psi \\
& +\frac{|A_\parallel|^2}{2}n_1\left[\sqrt{\frac{4\pi}{9}}Y_0^0 + \sqrt{\frac{16\pi}{45}}Y_2^0\right]\sin^2\psi + |A_0|^2 n_1\left[\sqrt{\frac{4\pi}{9}}Y_0^0 + \sqrt{\frac{16\pi}{45}}Y_2^0\right]\cos^2\psi \\
& +\frac{|A_\parallel|^2}{2}n_1\left[\sqrt{\frac{4\pi}{9}}Y_0^0 - \sqrt{\frac{4\pi}{45}}Y_2^0 + \sqrt{\frac{2\pi}{15}}(Y_2^{-2} + Y_2^2)\right]\sin^2\psi \\
& +|A_0|^2 n_1\left[\sqrt{\frac{4\pi}{9}}Y_0^0 - \sqrt{\frac{4\pi}{45}}Y_2^0 - \sqrt{\frac{2\pi}{15}}(Y_2^{-2} + Y_2^2)\right]\cos^2\psi \\
& +i\frac{A_\parallel A_0^* + A_\parallel^* A_0}{\sqrt{2}}n_1\left[\sqrt{\frac{2\pi}{15}}(Y_2^{-2} - Y_2^2)\right]\sin\psi\cos\psi \\
& +2Re\left[i\times n_3\left\{\frac{A_\parallel A_\perp^*}{2}\sqrt{\frac{2\pi}{15}}(Y_2^{-1} + Y_2^1)\sin^2\psi + i\frac{A_0 A_\perp^*}{\sqrt{2}}\sqrt{\frac{2\pi}{15}}(Y_2^{-1} - Y_2^1)\sin\psi\cos\psi\right\}\right].
\end{aligned} \tag{B.1}$$

We note that it is perfectly general to parameterize the efficiency function $\epsilon(\psi,\theta,\phi)$ as

$$\epsilon(\psi,\theta,\phi) = a_l^m(\psi)Y_l^m(\theta,\phi). \tag{B.2}$$

While this series may be infinite, only the terms through $l = 2$ are important, because they are the only ones to survive the integration in Eq. 5.12. Utilizing the fact that

$$\begin{aligned}
\int\int Y_l^m(\theta,\phi)\bar{Y}_{l'}^{m'}(\theta,\phi)\sin\theta d\theta d\phi &= \delta_{mm'}\delta_{ll'}, \\
Y_l^{-m}(\theta,\phi) &= (-1)^m\bar{Y}_l^m(\theta,\phi),
\end{aligned}$$

the normalization factor, N', can then be written

$$
\begin{aligned}
N' \;=\; & \int d(\cos\psi) \\
& \frac{|A_\perp|^2}{2} n_2 \left[\sqrt{\frac{16\pi}{9}} a_0^0(\psi) - \sqrt{\frac{16\pi}{45}} a_2^0(\psi) \right] \sin^2\psi \\
& + \frac{|A_\parallel|^2}{2} n_1 \left[\sqrt{\frac{4\pi}{9}} a_0^0(\psi) + \sqrt{\frac{16\pi}{45}} a_2^0(\psi) \right] \sin^2\psi + |A_0|^2 n_1 \left[\sqrt{\frac{4\pi}{9}} a_0^0(\psi) + \sqrt{\frac{16\pi}{45}} a_2^0(\psi) \right] \cos^2\psi \\
& + \frac{|A_\parallel|^2}{2} n_1 \left[\sqrt{\frac{4\pi}{9}} a_0^0(\psi) - \sqrt{\frac{4\pi}{45}} a_2^0(\psi) + \sqrt{\frac{2\pi}{15}} (a_2^{-2}(\psi) + a_2^2(\psi)) \right] \sin^2\psi \\
& + |A_0|^2 n_1 \left[\sqrt{\frac{4\pi}{9}} a_0^0(\psi) - \sqrt{\frac{4\pi}{45}} a_2^0(\psi) - \sqrt{\frac{2\pi}{15}} (a_2^{-2}(\psi) + a_2^2(\psi)) \right] \cos^2\psi \\
& - i \frac{A_\parallel A_0^* + A_\parallel^* A_0}{\sqrt{2}} n_1 \left[\sqrt{\frac{2\pi}{15}} (a_2^{-2}(\psi) - a_2^2(\psi)) \right] \sin\psi\cos\psi + 2Re\,[i \times n_3 \\
& \times \left\{ -\frac{A_\parallel A_\perp^*}{2} \sqrt{\frac{2\pi}{15}} (a_2^{-1}(\psi)) + a_2^1(\psi) \sin^2\psi + i \frac{A_0 A_\perp^*}{\sqrt{2}} \sqrt{\frac{2\pi}{15}} (a_2^{-1}(\psi) - a_2^1(\psi)) \sin\psi\cos\psi \right\} \Bigg] .
\end{aligned}
\tag{B.3}
$$

As the coefficients a_l^m are complex, it is advantageous to adopt the following set of orthonormal basis functions related to the spherical harmonics,

$$
\begin{aligned}
Y_{lm} &= Y_l^m \quad (m = 0), \\
Y_{lm} &= \frac{1}{\sqrt{2}} (Y_l^m + (-1)^m Y_l^{-m}) \quad m > 0, \\
Y_{lm} &= \frac{1}{i\sqrt{2}} (Y_l^{|m|} - (-1)^{|m|} Y_l^{-|m|}) \quad m < 0.
\end{aligned}
\tag{B.4}
$$

We then transcribe the coefficients in the following manner

$$
\begin{aligned}
a_{lm} &= a_l^m \quad (m = 0), \\
a_{lm} &= \frac{1}{\sqrt{2}} (a_l^m + (-1)^{|m|} a_l^{-m}) \quad m > 0, \\
a_{lm} &= \frac{i}{\sqrt{2}} (a_l^{|m|} - (-1)^{|m|} a_l^{-|m|}) \quad m < 0,
\end{aligned}
\tag{B.5}
$$

and, in particular,

$$
\begin{aligned}
a_{21} &= \frac{1}{\sqrt{2}}(a_2^1 - a_2^{-1}),\\
a_{2-1} &= \frac{i}{\sqrt{2}}(a_2^1 + a_2^{-1}),\\
a_{22} &= \frac{1}{\sqrt{2}}(a_2^2 + a_2^{-2}),\\
a_{2-2} &= \frac{i}{\sqrt{2}}(a_2^2 - a_2^{-2}).
\end{aligned}
$$

Making those substitutions and simplifying terms leads us to the following expression, which depends only on ψ,

$$
\begin{aligned}
N' &= \frac{2\sqrt{\pi}}{3}\left[(|A_\perp|^2 n_2 + |A_\parallel|^2 n_1)\int a_{00}(\psi)\sin^2\psi d(\cos\psi) + 2|A_0|^2 n_1 \int a_{00}(\psi)\cos^2\psi d(\cos\psi)\right]\\
&+ \frac{2\sqrt{\pi}}{3\sqrt{5}}\left[(-|A_\perp|^2 n_2 + \frac{1}{2}|A_\parallel|^2 n_1)\int a_{20}(\psi)\sin^2\psi d(\cos\psi) + |A_0|^2 n_1 \int a_{20}(\psi)\cos^2\psi d(\cos\psi)\right]\\
&- \frac{\sqrt{\pi}}{\sqrt{15}}\left[(A_\parallel^* A_\perp + A_\parallel A_\perp^*)n_3 \int a_{2-1}(\psi)\sin^2\psi d(\cos\psi)\right]\\
&+ \frac{\sqrt{2\pi}}{\sqrt{15}}\left[(A_0^* A_\perp + A_0 A_\perp^*)n_3 \int a_{21}(\psi)\sin\psi\cos\psi d(\cos\psi)\right]\\
&+ \frac{2\sqrt{\pi}}{\sqrt{15}}\left[\frac{|A_\parallel|^2}{2} n_1 \int a_{22}(\psi)\sin^2\psi d(\cos\psi) - |A_0|^2 n_1 \int a_{22}(\psi)\cos^2\psi d(\cos\psi)\right]\\
&+ \frac{\sqrt{2\pi}}{\sqrt{15}}\left[(A_0^* A_\parallel + A_0 A_\parallel^*)n_1 \int a_{2-2}(\psi)\sin\psi\cos\psi d(\cos\psi)\right].
\end{aligned}
$$

In the final step of the evaluation of N', we carry out the integration over ψ. To do this, we first express the functions $a_{lm}(\psi)$ as a Fourier-Legendre series,

$$a_{lm}(\psi) = a_{lm}^k P_k(\cos\psi).$$

Fortuitously, few of the a_{lm}^k's survive integration,

$$
\begin{aligned}
\int a_{lm}(\cos\psi)\sin^2(\psi)d(\cos\psi) &= \frac{4}{3}a_{lm}^0 - \frac{4}{15}a_{lm}^2,\\
\int a_{lm}(\cos\psi)\cos^2(\psi)d(\cos\psi) &= \frac{2}{3}a_{lm}^0 + \frac{4}{15}a_{lm}^2,\\
\int a_{lm}(\cos\psi)\cos(\psi)\sin\psi d(\cos\psi) &= \frac{\pi}{8}a_{lm}^1 - \frac{\pi}{32}a_{lm}^3 + ...
\end{aligned}
$$

Note, the peculiarities of the un-scuplted probability densities make the first two expansions finite, while the third one is in fact infinite. However, in practical terms, the third expansion does terminate since the efficiencies will be fit with a finite number of polynomials. The final expression for the normalization constant is then given by

$$
\begin{aligned}
N' &= \frac{2\sqrt{\pi}}{3}\left[\frac{4a^0_{00}}{3}(|A_0|^2n_1+|A_\parallel|^2n_1+|A_\perp|^2n_2)+\frac{4a^2_{00}}{15}(2|A_0|^2n_1-|A_\parallel|^2n_1-|A_\perp|^2n_2)\right] \\
&+ \frac{2\sqrt{\pi}}{3\sqrt{5}}\left[\frac{2a^0_{20}}{3}(|A_0|^2n_1+|A_\parallel|^2n_1-2|A_\perp|^2n_2)+\frac{4a^2_{20}}{15}(|A_0|^2n_1-\frac{1}{2}|A_\parallel|^2n_1+|A_\perp|^2n_2)\right] \\
&- \frac{\sqrt{\pi}}{\sqrt{15}}\left[(A^*_\parallel A_\perp+A_\parallel A^*_\perp)n_3(\frac{4}{3}a^0_{2-1}-\frac{4}{15}a^2_{2-1})\right] \\
&+ \frac{\sqrt{2\pi}}{\sqrt{15}}\left[(A^*_0 A_\perp+A_0 A^*_\perp)n_3(\frac{\pi a^1_{21}}{8}-\frac{\pi a^3_{21}}{32}+...)\right] \\
&+ \frac{2\sqrt{\pi}}{\sqrt{15}}\left[\frac{2a^0_{22}}{3}(-|A_0|^2+|A_\parallel|^2)n_1-\frac{4a^2_{22}}{15}(|A_0|^2+\frac{1}{2}|A_\parallel|^2)n_1\right] \\
&+ \frac{\sqrt{2\pi}}{\sqrt{15}}\left[(A^*_0 A_\parallel+A_0 A^*_\parallel)n_1(\frac{\pi a^1_{2-2}}{8}-\frac{\pi a^3_{2-2}}{32}+...)\right].
\end{aligned}
$$

APPENDIX C

GOOD RUN SELECTION

The data periods used are taken from the **xpmm0d**, **xpmm0h** and **xpmm0i** datasets at CDF. They consist of runs between 138342 and 224521.

The datasets **xpmm0d** and **xpmm0h** were reconstructed with the CDF offline version 5.3.1 and the **xpmm0i** dataset with the 6.1.2. Track refitting and vertex reconstruction was done using versions 5.3.4 and 6.1.2. Alignment and calibration constants are obtained from the database **used sets**, `PassName 17`.

Only runs satisfying the following quality criteria are used:

1. the run must have a valid SVX beam information based on **used sets**, `PassName 17`,
2. it should meet the modified "BPAK goodrun.sql" requirements (see the following)
3. it should not have been marked bad because known detector problems (e.g. COT high voltage, 179056 $<$ run $<$ 182627), trigger problems or with the SVX II not fully functioning.

For good run selection, a slightly modified version of the sql script provided by the BPAK group[76] is used. The only difference between our selection and the standard selection is that we remove the requirement on `CAL_OFFLINE` and `SVT_STATUS`. With this good run selection, we obtain an offline integrated luminosity of ~ 1.35 fb^{-1}.

The modified requirements are listed here:

```
SELECT RUNNUMBER, sum(LUM_INTEGRAL_OFFLINE),  sum(LUM_INTEGRAL_ONLINE)
FROM Run_Status, FILECATALOG.CDF2_RUNSECTIONS
WHERE
Run_Status.RUNNUMBER = FILECATALOG.CDF2_RUNSECTIONS.RUN_NUMBER
```

```
AND
Run_Status.RUNCONTROL_STATUS = 1
AND
Run_Status.SHIFTCREW_STATUS = 1
AND
Run_Status.CLC_STATUS = 1
AND
Run_Status.L1T_STATUS = 1
AND
Run_Status.L2T_STATUS = 1
AND
Run_Status.L3T_STATUS = 1
AND
Run_Status.COT_OFFLINE = 1
AND
(Run_Status.SVX_OFFLINE = 1 OR
((Run_Status.SVX_OFFLINE Is Null) AND Run_Status.SVX_STATUS = 1))
AND
(Run_Status.CMU_OFFLINE = 1 OR
((Run_Status.CMU_OFFLINE Is Null) AND Run_Status.CMU_STATUS = 1))
AND
(Run_Status.CMP_OFFLINE = 1 OR
((Run_Status.CMP_OFFLINE Is Null) AND Run_Status.CMP_STATUS = 1))
AND
(RUNNUMBER<=179056 OR RUNNUMBER>=182843
OR (RUNNUMBER>=180954 AND RUNNUMBER<=181190))
AND
(RUNNUMBER<184062 OR RUNNUMBER>184208)
AND
RUNNUMBER <= 210011
GROUP BY RUNNUMBER
ORDER BY RUNNUMBER ASC
/
QUIT
```

APPENDIX D

ANGULAR DISTRIBUITON PLOTS

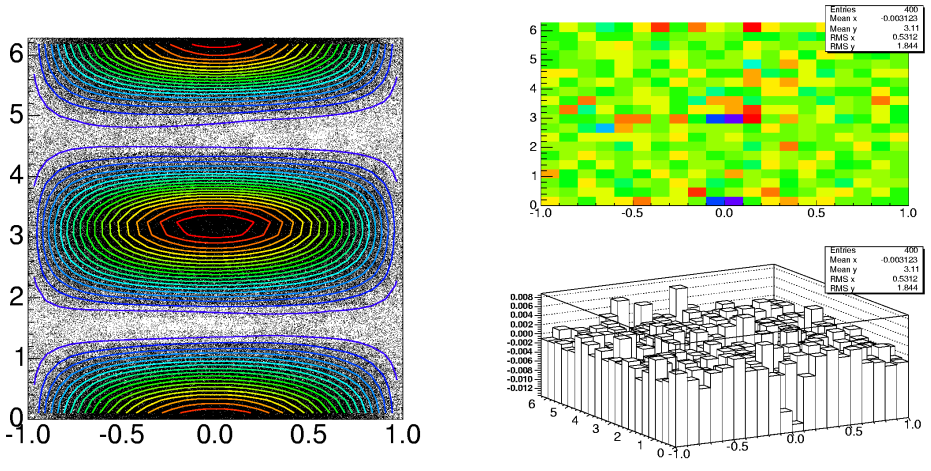

Figure 80: Two dimensional fit to (θ, ϕ) averaging over $\cos\psi$.

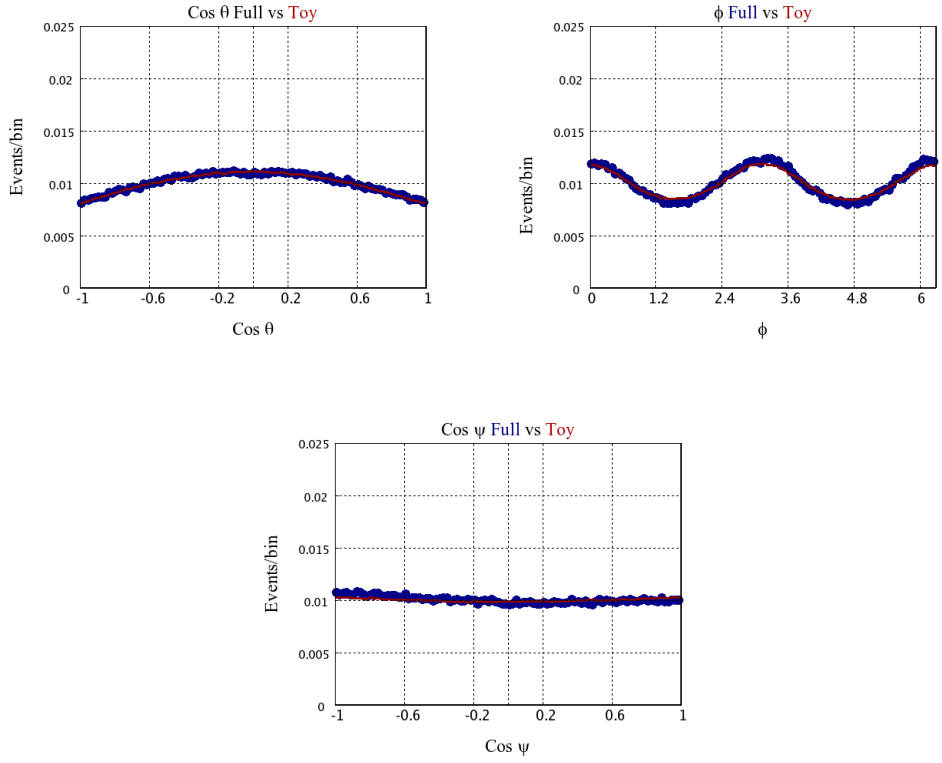

Figure 81: Fit projection of angular distributions on realistic Monte Carlo $B_s^0 \to J/\psi\,\phi$ phase space events.

Coefficient	Fit Result	Coefficient	Fit Result	Coefficient	Fit Result
C_{00}^{0}	$4.754e+02 \pm 5.4e-01$	C_{22}^{0}	$6.657e+01 \pm 5.4e-01$	C_{33}^{0}	$2.677e-01 \pm 3.2e-01$
C_{00}^{2}	$6.061e+00 \pm 5.5e-01$	C_{22}^{2}	$1.920e+00 \pm 5.5e-01$	C_{33}^{2}	$1.181e-01 \pm 3.2e-01$
C_{11}^{0}	$2.395e+00 \pm 5.8e-01$	S_{22}^{0}	$-9.986e-01 \pm 5.4e-01$	S_{33}^{0}	$5.606e-01 \pm 3.1e-01$
C_{11}^{2}	$4.067e-01 \pm 5.8e-01$	S_{22}^{2}	$-6.166e-01 \pm 5.3e-01$	S_{33}^{2}	$-9.437e-02 \pm 3.2e-01$
S_{11}^{0}	$-1.049e+00 \pm 5.1e-01$	C_{21}^{0}	$-8.325e-02 \pm 5.5e-01$	C_{32}^{0}	$1.007e+01 \pm 5.1e-01$
S_{11}^{2}	$6.664e-01 \pm 5.2e-01$	C_{21}^{2}	$7.574e-01 \pm 5.5e-01$	C_{32}^{3}	$-2.453e-01 \pm 5.2e-01$
C_{10}^{0}	$-8.355e-01 \pm 5.1e-01$	S_{21}^{0}	$-7.834e-01 \pm 5.0e-01$	S_{32}^{0}	$4.977e-01 \pm 5.3e-01$
C_{10}^{2}	$-6.286e-02 \pm 5.2e-01$	S_{21}^{2}	$-5.981e-01 \pm 5.1e-01$	S_{32}^{2}	$-3.666e-01 \pm 5.3e-01$
		C_{20}^{0}	$-4.415e+01 \pm 5.3e-01$	C_{31}^{0}	$4.784e-01 \pm 5.4e-01$
		C_{20}^{2}	$-2.996e+00 \pm 5.4e-01$	C_{31}^{2}	$1.085e+00 \pm 5.4e-01$
				S_{31}^{0}	$5.170e-02 \pm 5.0e-01$
				S_{31}^{2}	$-4.321e-01 \pm 5.0e-01$
				C_{30}^{0}	$3.028e-01 \pm 5.3e-01$
				C_{30}^{2}	$5.726e-01 \pm 5.3e-01$

Coefficient	Fit Result	Coefficient	Fit Result
C_{44}^{0}	$4.972e+00 \pm 5.5e-01$	C_{55}^{0}	$-3.825e-01 \pm 5.5e-01$
C_{44}^{2}	$2.886e-01 \pm 5.6e-01$	C_{55}^{2}	$3.554e-01 \pm 5.6e-01$
S_{44}^{0}	$4.011e-01 \pm 5.5e-01$	S_{55}^{0}	$3.778e+00 \pm 5.5e-01$
S_{44}^{2}	$-5.199e-01 \pm 5.6e-01$	S_{55}^{2}	$3.374e-01 \pm 5.6e-01$
C_{43}^{0}	$-3.773e-01 \pm 5.4e-01$	C_{54}^{0}	$2.655e-01 \pm 5.4e-01$
C_{43}^{2}	$-1.089e+00 \pm 5.4e-01$	C_{54}^{2}	$-3.293e-01 \pm 5.4e-01$
S_{43}^{0}	$6.010e-01 \pm 5.4e-01$	S_{54}^{0}	$-3.364e-01 \pm 5.4e-01$
S_{43}^{2}	$-6.842e-01 \pm 5.4e-01$	S_{54}^{2}	$-4.727e-01 \pm 5.4e-01$
C_{42}^{0}	$-1.590e+01 \pm 5.1e-01$	C_{53}^{0}	$1.168e-02 \pm 5.3e-01$
C_{42}^{2}	$-1.103e+00 \pm 5.2e-01$	C_{53}^{2}	$1.450e+00 \pm 5.4e-01$
S_{42}^{0}	$1.226e-01 \pm 5.2e-01$	S_{53}^{0}	$-9.011e-01 \pm 5.3e-01$
S_{42}^{2}	$2.145e-01 \pm 5.1e-01$	S_{53}^{2}	$-1.524e-01 \pm 5.4e-01$
C_{41}^{0}	$-5.957e-02 \pm 5.3e-01$	C_{52}^{0}	$6.080e+00 \pm 1.3e-01$
C_{41}^{2}	$2.692e-01 \pm 5.4e-01$	C_{52}^{2}	$2.026e-01 \pm 1.4e-01$
S_{41}^{0}	$3.128e-01 \pm 4.9e-01$	S_{52}^{0}	$-1.281e-01 \pm 1.5e-01$
S_{41}^{2}	$3.424e-01 \pm 5.0e-01$	S_{52}^{2}	$7.805e-02 \pm 1.4e-01$
C_{40}^{0}	$1.481e+00 \pm 5.5e-01$	C_{51}^{0}	$-1.446e-01 \pm 5.2e-01$
C_{40}^{2}	$6.310e-01 \pm 5.5e-01$	C_{51}^{2}	$-1.458e-01 \pm 5.3e-01$
		S_{51}^{0}	$-6.549e-02 \pm 4.9e-01$
		S_{51}^{0}	$-2.483e-02 \pm 4.9e-01$
		C_{50}^{0}	$3.233e-01 \pm 5.6e-01$
		C_{50}^{2}	$1.366e-01 \pm 5.7e-01$

Table 23: Fit results for angular efficiencies.

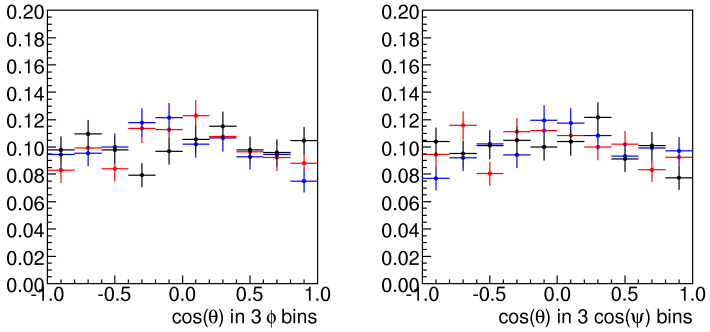

Figure 82: Background $\cos(\theta)$ distributions in different ϕ and $\cos(\psi)$ bins.

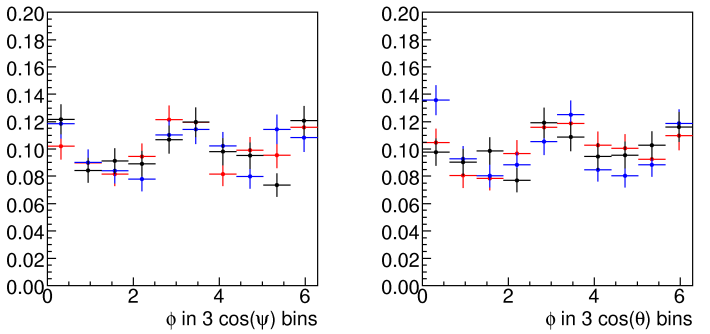

Figure 83: Background ϕ distributions in different $\cos(\theta)$ and $\cos(\psi)$ bins.

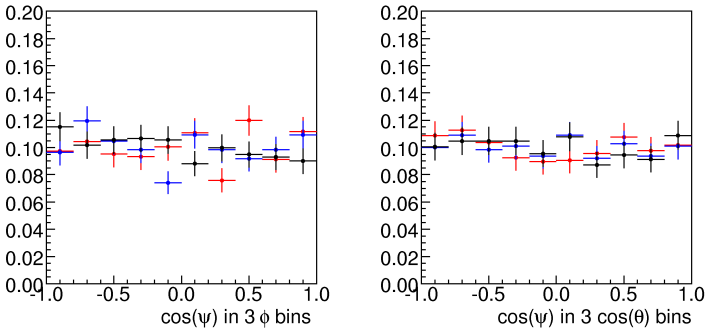

Figure 84: Background $\cos(\psi)$ distributions in different ϕ (left) and $\cos(\theta)$ bins.