Measurement of the CP Violating Phase $\sin(2\beta_s)$ using $B_s^0 \to J/\psi\phi$ Decays at CDF

Elisa Pueschel

Department of Physics

Carnegie Mellon University

A thesis submitted for the degree of

Doctor of Philosophy

May 2010

Abstract

A B_s^0 meson can oscillate into its anti-particle, the $\bar{B_s^0}$ meson, before decaying. CP violation in this system is made possible by the presence of amplitudes from both mixed and unmixed B_s^0 meson decays. The CP violating phase β_s appears in the interference between the decay amplitudes. The quantity $\sin(2\beta_s)$ is expected to be small in the standard model. Thus, measuring a large value for $\sin(2\beta_s)$ would be an unequivocal sign of new physics participation in the B_s^0 mixing loop diagram.

In this thesis, we present a latest measurement of $\sin(2\beta_s)$, using 5.2 fb^{-1} of data collected at CDF from $p\bar{p}$ collisions at a center of mass energy of $\sqrt{s}$=1.96 TeV. A time-dependent angular analysis, with the production flavor of the B_s^0 meson identified with flavor tagging methods, is used to extract $\sin(2\beta_s)$ from $\sim$6500 $B_s^0 \to J/\psi\phi$ decays. Other parameters of interest, such as the B_s^0 lifetime and the decay width difference $\Delta\Gamma$ between the heavy and light B_s^0 mass eigenstates are determined to high precision. Also, the effect of potential contributions to the final state from $B_s^0 \to J/\psi f_0$ and $B_s^0 \to J/\psi K^+K^-$ decays is considered for the first time.

We present 68% and 95% confidence regions in the $\beta_s - \Delta\Gamma$ plane. The probability that the observed central value is a fluctuation of the data from the standard model expected value of β_s is calculated to be 44%. The observed confidence region shows better agreement with the standard model prediction than previous measurements.

Acknowledgements

There are a number of people whose work and interest made this thesis possible, and who have helped to make my time in graduate school both pleasant and productive. I've been helped by enough people that it is impossible to acknowledge everyone, but a few people deserve special thanks for their extraordinary efforts.

First and foremost, I'd like to thank my advisor, Manfred Paulini. I count myself very lucky to have had the opportunity to work with him. Throughout the time we've worked together, he has struck a balance between guiding me and allowing me independence. His good judgement, professionalism and sense of humor have been an invaluable example to me. I thank him for being involved and supportive.

I would like to thank my collaborators in the $\sin 2\beta_s$ group: Farrukh Azfar, Joe Boudreau, Karen Gibson, Gavril Giurgiu, Guillelmo Gomez-Ceballos, Thomas Kuhr, Michal Kreps, Louise Oakes, Jan Morlock and Andreas Schmidt. I especially thank Joe and Michal, from whom I have learned a great deal of physics. They have always been wonderful about responding to my questions. I thank Karen for her help and instruction when I was learning the analysis, and her friendship and advice. I thank Gavril for answering hundreds of middle of the night emails about the analysis, and setting an example of careful and thorough work that I strive to emulate. I thank Karen and Gavril for always being encouraging and supportive. Louise has been my partner in crime since we joined the analysis. She has been a huge source of support through the trials and tribulations of completing the work in this thesis. I thank her for innumerable helpful physics and non-physics conversations, for coaxing me off the academic ledge a time or two, and for indulging my fondness for dosai at India Harvest. I would

also like to thank Chunlei Lui, who worked on the first iteration of the analysis, and helped me get started.

I would like to thank the conveners of the B group and the conveners of the B Mixing, Lifetime, and CP violation subgroup. In my time at CDF, this has included Diego Tonelli and Michal Kreps for the B group (previously Giovanni Punzi and Manfred), and Michael Morello and Farrukh Azfar for BMLCPV (previously Diego and Gavril). Their work in trying to make this result the best it could be, and in proposing useful cross-checks has been much appreciated. I thank the members of the B group for their comments and attention to this work.

I appreciate the help of the members of my thesis committee: Manfred, Joe, Jim Russ, Roy Briere, and Adam Leibovich. I additionally thank Jim for helping to create the positive environment in the CMU HEP group, and for being a good source of advice.

On the personal side, I would like to thank the many friends who have dragged me away from what I was working on and forced me to enjoy life. In particular, I thank my officemate, Buck Thome, for always making me laugh, helping me keep my sense of perspective, and taking me to the mall. I thank Dean Hidas for making sure that I ate dinner and sometimes even lunch on a regular basis when I was living at Fermilab. I thank Bo Jayatilaka for fielding my endless questions about the logistics of conference registration and many other tedious subjects and for providing a place to crash in Chicago. From CMU, I thank Frankie Li for always being up for a cup of coffee or a stiff drink, as circumstances demanded. I thank my housemate, Peter LeFanu Lumsdaine, for looking after me this past year. I thank Vivian Ferry for sartorial support.

I thank my parents for everything they have done for me. I thank my siblings, Eric, Kristen and Norah, for being wonderful.

Lastly, I would like to thank my husband, John Bulava, for useful physics discussions and for always believing in me.

Contents

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

Chapter 1

Introduction

The standard model of particle physics, for the last several decades, has been immensely successful in describing all observed interactions between particles. Yet with its many input parameters, and inability to explain, for instance, the mass hierarchy of the fundamental particles, many physicists believe the standard model to be an effective theory that is only accurate in the energy regimes that have been probed thus far. By building particle accelerators that make accessible higher energy regimes, physicists hope to discover a fundamental theory of nature, of which the standard model is only a low energy approximation.

Signs of the structure of a higher energy theory may be accessible at currently available energies. Heavy, non-standard model particles could be off-shell contributors in non-tree level loop processes, altering decay rates from their standard model predictions. The measurement described in this thesis is of a phase that is sensitive to non-standard model contributions.

1.1 The standard model

The standard model (SM) is a quantum field theory which incorporates both quantum mechanics and special relativity [1]. It describes three of the four known forces: the strong force and the electroweak (electromagnetic and weak) force. The last force, gravity, is not included in the SM. The interactions of the building blocks of matter, the quarks and leptons (electron, muon, tau, neutrinos), as well as those of the force mediators (photon, W and Z bosons, gluons) are described by the SM. Photons are

	Generation 1	Generation 2	Generation 3
charged leptons	electron (0.000511)	muon (0.106)	tau (1.78)
neutrinos	ν_e ($\sim$0)	ν_μ ($\sim$0)	ν_τ ($\sim$0)
up-type quarks	up (0.0015-0.0033)	charm (1.27)	top (171.2)
down-type quarks	down (0.0035-0.0060)	strange (0.104)	bottom (4.2)

Table 1.1: The three generations of quarks and leptons in the standard model, with masses in GeV/c^2.

the force mediators for electromagnetism, W and Z for the weak force, and gluons for the strong force. In this thesis, we are mainly concerned with the weak interaction.

The quarks, leptons, and the leptons' associated neutrinos can be arranged into generations as shown in Table 1.1 including particle masses in GeV/c^2 [2]. The standard model does not explain why the fundamental particles' masses span such a broad range, from the nearly massless neutrinos to the the top quark, which is approximately as heavy as a gold nucleus.

The first generation particles constitute most of the matter than can be seen outside of particle accelerators. The particles in the second and third generations are highly unstable with heavier particles decaying into lighter ones. In order to study their properties, we must produce them via high energy interactions in the controlled environment of a particle detector, where they can be studied as they decay. The standard model does not explain why the quark and lepton masses span such a broad ranges, from nearly massless neutrinos to the the top quark, which is approximately as heavy as a gold nucleus.

Quarks and leptons are all *fermions*, particles with an intrinsic spin of 1/2. Quarks carry electric charge, weak isospin, and color charge, and interact via the electromagnetic, weak, and strong interactions. The strong interaction causes quarks to bind to each other in mesons or baryons, color neutral conglomerates of two or three quarks. The electric charge of the up-type quarks is $+2/3$ and $-1/3$ for the down-type quarks. The charged leptons have charged -1 and are colorless. They do not interact strongly but only via the electromagnetic or weak interaction. Neutrinos are electrically neural, and therefore interact only via the weak interaction, making them notoriously difficult to detect [1].

Each of the particles listed in Table 1.1 has an *antiparticle*, a particle with the same mass but opposite charge. Other quantum numbers, such as the color charge, can also be reversed when translating a particle into an antiparticle.

1.1.1 The Weak Interaction and the CKM Matrix

The W and Z gauge bosons, the mediators of the weak interaction, are both spin-1 particles. The W carries electric charge ± 1, and the Z is neutral. Unlike the other force-mediating gauge bosons, the photon and gluon, the W and Z are massive: 80.4 and 91.2 GeV/c^2. The amplitude for a tree level interaction via a gauge boson is proportional to $1/(M^2 - q^2)$, where M is the mass of the gauge boson and q^2 is the four momentum. A low q^2 or, equivalently, long range, the mass of the gauge boson makes amplitude and hence the force of the weak interaction small. This explains why the weak force is short-range, while the massless photon gives the electromagnetic force infinite range. The strong force is short-range in spite of the massless gluon, but the reasons for this are beyond the scope of this thesis [1].

The weak force has two special properties, with experimental effects that were inexplicable by the theoretical framework at the time of their observation. The first is that the weak force maximally violates parity conservation. As we will discuss in more detail in Sec. 1.2, the parity operator transforms a particle's momentum vector $\vec{p}$ to $-\vec{p}$. The parity operator does not affect the particle's spin, though, so the helicity of a particle ($\vec{p} \cdot \vec{s}$, where $\vec{s}$ is the spin direction) will flip under a parity transformation. Parity is conserved under the strong and electromagnetic interactions, but is maximally violated by the weak interaction. The effect of this is that the W will only couple to left handed particles and right handed antiparticles, where "left-handed" and "right-handed" refer to sign of the chirality. The Z, one the other hand, can couple to particles with left- or right-handed chirality, but not with equal strength [1].

Cronin and Fitch first demonstrated parity violation in the weak interaction in 1964 [3] in the neutral kaon system. In the kaon system, the long-lived CP odd state decays to three pions and the short-lived CP even state decays to two pions. In the absence of CP violation, at a sufficiently long distance from the point where the kaons are produced, only the decay products from the CP odd long-lived states should be present, as the CP even states will have already decayed. However, Cronin and Fitch's

experiment found a small fraction of CP even decay products far from the production point, indicating that CP violation occurs in the neutral kaon system.

Interactions involving a W are referred to as flavor-changing charged currents. A down-type quark can decay in a W and an up-type quark or vice versa. The vertex for this process is shown in Figure 1.1. Note that electric charge is conserved in this process. For example, a $+1/3$ down-type quark can have a vertex with a W^- and a $+2/3$ up-type quark. Since $(+1/3)+(-1)+(+2/3)=0$, no charge is lost or gained at the vertex. The neutral Z boson cannot form a flavor-changing vertex that conserves electric charge, resulting in the absence of flavor-changing neutral currents in the SM at tree-level. The Z boson could be exchanged, without violating charge conservation, in vertices linking up-type quarks to other up-type quarks, or down-type quarks to other down-type quarks. However, such vertices do not arise from the electroweak Lagrangian and thus do not appear in the standard model.

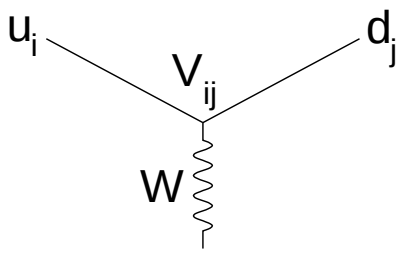

Figure 1.1: Vertex for up-type quark decaying to a down-type quark through emission of a W boson.

The interaction Lagrangian for the process represented in Fig. 1.1 is the following:

$$\mathcal{L} = -\frac{g}{2\sqrt{2}}[\bar{u}_i\gamma^\mu(1-\gamma^5)d_i'W_\mu^+ + h.c.], \tag{1.1}$$

where "h.c." is the Hermitian conjugate and summation over indices i and j is implied. The constant g sets the overall scale for the strength of the coupling. u_i and d_j are

Dirac spinors, and γ^μ and γ^5 are Dirac matrices [1].

One notices that the down-type quarks are primed, due to the fact that the quark mass eigenstates *d, s* and *b* are not the same as the quark weak eigenstates. A rotation must be performed in order to translate the mass eigenstates into the weak eigenstates. This rotation defines the Cabibbo-Kobayashi-Maskawa (CKM) matrix [4]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{1.2}$$

Using the CKM matrix, we can rewrite the interaction Lagrangian as:

$$\mathcal{L} = -\frac{g}{2\sqrt{2}}[\bar{u}_i \gamma^\mu (1-\gamma^5) V_{ij} d_j W_\mu^+ + h.c.]. \tag{1.3}$$

Thus, the coupling between two quarks $\bar{u}_i$ and d_j is proportion to the matrix element V_{ij}. The standard model makes no prediction for the size of the CKM matrix elements. Experimentally, it has been shown that the CKM matrix is approximately a unit matrix; the dominant quark transitions are along the diagonal. The fact that the off-diagonal elements are not exactly zero makes *CP* (charge-parity) violation possible, as we will see in Sec. 1.2.

The CKM matrix is unitary by construction, a requirement for conservation of probability. An $(n \times n)$ unitary matrix can be described by n^2 real parameters. Of these parameters, $(2n-1)$ can be absorbed into the phases of the quark fields, leaving $(n-1)^2$ physically meaningful parameters [5]. The 3×3 CKM matrix can be parameterized by four independent parameters: three rotational angles $(\theta_{12}, \theta_{23}, \theta_{13})$ and one phase (δ):

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.4}$$

where c_{ij} and s_{ij} are the sine and cosine of the respective rotational angles.

The Wolfenstein parameterization of the CKM matrix [6] uses a different set of parameters, emphasizing the hierarchy of the matrix. It uses the four parameters $A = \sin(\theta_{23})/\lambda^2$, $\lambda = \sin(\theta_{12})$, $\rho = \sin(\theta_{13})\cos(\delta)/A\lambda^3$ and $\eta = \sin(\theta_{13})\sin\delta/A\lambda^3$.

Matrix elements can be expanded in $\lambda \approx 0.23$, allowing a comparison of the relative magnitudes of the matrix elements. Dropping terms above $\mathcal{O}(\lambda^5)$, we have

$$V_{CKM} \approx \begin{pmatrix} 1-\frac{\lambda^2}{2}-\frac{\lambda^4}{8} & \lambda & A\lambda^3(\rho-i\eta) \\ -\lambda+\frac{A^2\lambda^5}{2}[1-2(\rho+i\eta)] & 1-\frac{\lambda^2}{2}-\frac{\lambda^4}{8}(1+4A^2) & A\lambda^2 \\ A\lambda^3[1-(1-\frac{\lambda^2}{2})(\rho+i\eta)] & -A\lambda^2+\frac{A\lambda^4}{2}[1-2(\rho+i\eta)] & 1-\frac{A^2\lambda^4}{2} \end{pmatrix} \quad (1.5)$$

It is clear that the diagonal elements of the matrix are approximately one, and the off-diagonal elements are small, suppressing decays between different quark generations.

We can use the unitarity property of the CKM matrix

$$\sum_{j=1}^{3} V_{ij}V_{kj}^* = \delta_{ik} \quad (1.6)$$

to derive a series of unitarity relations. For example, from the first and third columns of the CKM matrix we can derive

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \quad (1.7)$$

These relations can be represented graphically as a triangle in the complex plane, where $(V_{ij}V_{kj}^*)$ define the sides of the triangle. The graphical representation of Equation 1.7 is shown in Fig. 1.2. The triangle can be rotated and the sides scaled such that the longest side extends from the origin to (1, 0), and the apex's coordinates are given by $(\bar{\rho}, \bar{\eta})$, where $\bar{\rho}$ and $\bar{\eta}$ are related to ρ and η by a factor of $(1 - \lambda^2/2)$. The result is shown in Fig. 1.3.

The angles of the triangle can be defined by ratios of the lengths of the sides, which are given as products of CKM matrix elements. The goal of CKM physics is to precisely experimentally determine the size of all the angles and sides of a unitarity triangle in multiple ways. The unitarity relation will then be over-constrained. If a discrepancy is observed between two methods of measuring a CKM angle or side, it indicates that one of the processes used for the measurement is affected by non-standard model physics. Thus, consistency with the standard model for all the processes that yield the measurements can be tested.

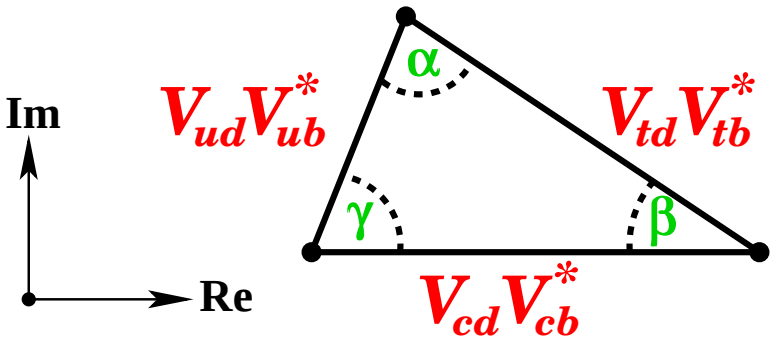

Figure 1.2: Unitarity triangle from the first and third columns of the CKM matrix.

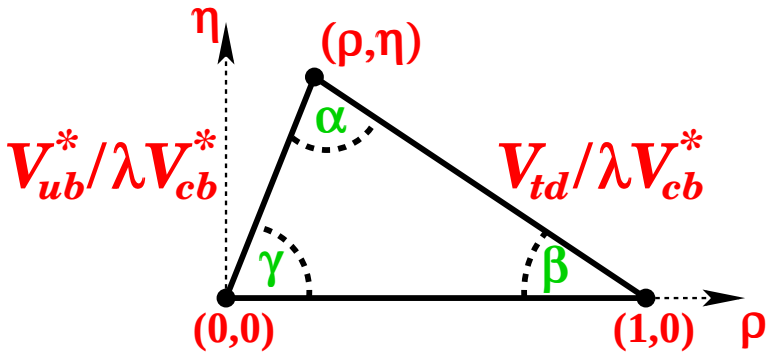

Figure 1.3: Unitarity triangle from the first and third columns of the CKM matrix, scaled in terms of Wolfenstein parameters.

Referring back to Eq. 1.6, we see that there are six unitarity triangles. The unitarity triangle shown in Fig. 1.7 has been well-studied experimentally, in part due to its conveniently large angles. However, in this thesis, we are interested in the unitarity triangle coming from the second and third columns of the CKM matrix:

$$V_{us}V_{ub}^{*} + V_{cs}V_{cb}^{*} + V_{ts}V_{tb}^{*} = 0. \tag{1.8}$$

In particular, we wish to measure the angle equivalent to β in Fig. 1.7, referred to as β_s. The unitary triangle containing β_s is shown in Fig. 1.4 (not to scale). This is a challenging experimental endeavor, because β_s is very small. We can get a rough idea of the size of β_s in two steps. First, we describe β_s in terms of the sides of the unitarity triangle where it appears:

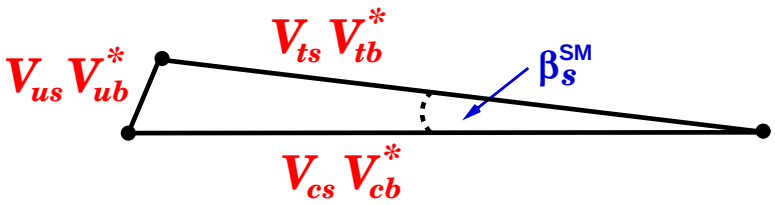

Figure 1.4: Unitarity triangle from the second and third columns of the CKM matrix.

$$\beta_s = arg\left(-\frac{V_{ts}V_{tb}^{*}}{V_{cs}V_{cb}^{*}}\right). \tag{1.9}$$

We can then examine the $\mathcal{O}(\lambda)$ of the CKM matrix elements that define β_s, using the Wolfenstein parameterization given Eq. 1.5. We find that

$$\beta_s = arg\left(\frac{-V_{ts}V_{tb}^{*}}{V_{cs}V_{cb}^{*}}\right) \sim \lambda^2 < 0.1. \tag{1.10}$$

The standard model expectation, based on indirect measurements that constrain the unitary relation, is β_s=0.02. We return to the identification of the angle β_s as a "*CP*-violating phase" in the next section.

1.2 CP Violation

CP violation refers to an absence of symmetry under a combined charge and parity transformation. The so-called charge conjugation C transforms a particle into its antiparticle, but leaves momentum unaffected. The parity transform P is a spatial inversion, transforming $\vec{x}$ to $-\vec{x}$. As stated previously, the weak interaction maximally violates parity conservation. The standard model respects CPT symmetry, where T is the time reversal transformation than maps t to -t. This implies that if CP is violated, then T must be as well in order to preserve the symmetry of the combined transformation.

CP violation can manifest itself in different ways in physical processes. A common thread exists, though: in order for CP to be violated, an interfering amplitude must exist in addition to the dominant decay amplitude, to provide a phase that flips signs under the CP transformation [5].

Another way of stating the condition for CP violation is to demand that the area of the unitarity triangle be non-zero, as we will now explain. A phase convention independent measure of CP violation, which is related to the area of the unitarity triangles, is the Jarlskog invariant J [7]:

$$J = |Im(V_{ij}V_{il}^*V_{kj}^*V_{kl})|, \quad i \neq k, j \neq l. \tag{1.11}$$

The condition that $i \neq k$ and $j \neq l$ (and the fact that there are only three generations) guarantees that for any choice of i, j, k, J contains at least two off-diagonal CKM elements. This corresponds to interference between the dominant amplitude and a generation-changing one, or between two generation-changing amplitudes [5].

Perhaps more intuitively, one can parameterize the Jarlskog invariant in terms of the Wolfenstein parameters and approximate it as:

$$J \approx A^2\eta\lambda^6. \tag{1.12}$$

In order for CP violation to occur, the parameter η, related to the "height" of the unitarity triangle in Fig. 1.7 must be non-zero. If η is zero, the size of the Jarlskog

invariant is also zero and CP violation does not occur. η is the magnitude of the imaginary components of off-diagonal elements of the CKM matrix, as shown in Eq. 1.5. As mentioned before, the small but non-zero off-diagonal elements are necessary for CP violation to occur, and a factor of i must be present to give the unitary triangles some "height" in the complex plane! Furthermore, we can determine the relative size of the CP asymmetries in different processes, based on the order of λ multiplying $(i\eta)$ in the CKM elements involved in a processes.

CP violation can occur in decay (also called direct CP violation), in neural meson mixing, or in the interference between direct decay and decays via mixing. Before describing the different classes of CP violation, we first give an overview of neutral meson mixing [8].

1.2.1 Phenomenology of Neutral Meson Mixing

Neural meson mixing is a quantum mechanical process where a particle transforms into its antiparticle, via a "box" diagram, a one loop process. The box diagrams for neutral B oscillation are shown in Fig. 1.5. Oscillation occurs via the exchange of W bosons and up-type quarks. The oscillation frequency goes as m_q^2/m_W^2. The top quark, by far the most massive quark, gives the dominant contribution to the oscillation amplitude. In some extensions to the standard model, though, non-SM heavy particles could be exchanged in addition to the up-type quarks in Fig. 1.5. This would alter the amplitude associated with $B_q^0 \bar{B}_q^0$ oscillation from the standard model prediction, and could also affect the participating weak CP violating phases [9, 10].

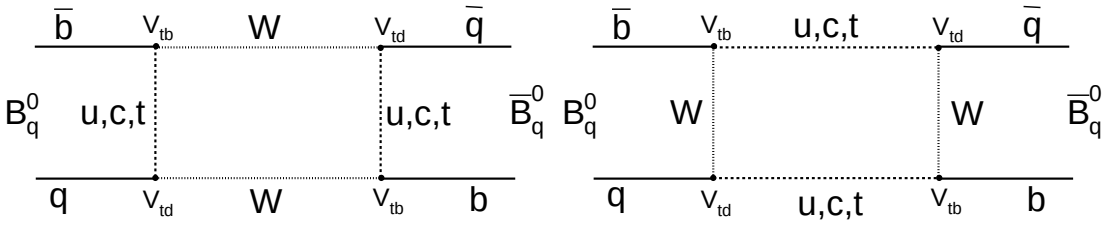

Figure 1.5: Leading order diagrams for B meson mixing.

For convenience, we describe the case of the B_s^0 system, but the following discussion could equally well apply to B^0, K^0, or D^0 mesons.

A bottom quark and a strange quark can bind together to form a neutral B_s^0 meson, or its antiparticle, $\bar{B}_s^0$

$$|B_s^0\rangle = |\bar{b}s\rangle, \qquad |\bar{B}_s^0\rangle = |b\bar{s}\rangle. \tag{1.13}$$

B_s^0 and $\bar{B}_s^0$ can both decay to a set of final states $|f\rangle$. The total Hamiltonian can be broken into two parts,

$$H = H_0 + H_{weak}. \tag{1.14}$$

$|B_s^0\rangle$ and $|\bar{B}_s^0\rangle$ are eigenstates of H_0, the strong and electromagnetic components of the total Hamiltonian. H_{weak} contains the weak component of the Hamiltonian, which is responsible for mixing and decay to the final state. After a time t, an initial B_s^0 or $\bar{B}_s^0$ will be in the state

$$|\psi(t)\rangle = a(t)|B_s^0\rangle + b(t)|\bar{B}_s^0\rangle + \sum_f c(t)|f\rangle, \tag{1.15}$$

a superposition of the two flavor eigenstates B_s^0 or $\bar{B}_s^0$ and the states $|f\rangle$ to which B_s^0 and $\bar{B}_s^0$ can decay.

The time-evolution can be equivalently described by the time-dependent Schrodinger equation, which removes explicit dependence on $|f\rangle$:

$$i\frac{d}{dt}\begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} = \left(\mathbf{M} - \frac{i}{2}\mathbf{\Gamma}\right)\begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} \tag{1.16}$$

where

$$\mathbf{H} = \left(\mathbf{M} - \frac{i}{2}\mathbf{\Gamma}\right) = \begin{pmatrix} M & M_{12} \\ M_{12}^* & M \end{pmatrix} - \frac{i}{2}\begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma \end{pmatrix} \tag{1.17}$$

and the mass matrix $\mathbf{M}$ and decay matrix $\mathbf{\Gamma}$ are hermitian. The diagonal elements of the mass matrix are related to the mass eigenvalues of H_0. In quantum mechanics, off-diagonal elements are always related to transitions between states. In this case, the off-diagonal elements are related to the meson oscillation via off-shell intermediate

states. The diagonal elements of the decay matrix are due to on-shell decays of $B_s^0 \to f$ and $\bar{B}_s^0 \to \bar{f}$, and the off-diagonal elements are due to the transitions $B_s^0 \to f \to \bar{B}_s^0$ and $\bar{B}_s^0 \to f \to B_s^0$.

Diagonalizing the Hamiltonian, one can extract the mass eigenstates of $\mathbf{H}$,

$$|B_s^{H,L}\rangle = \frac{1}{\sqrt{|p|^2 + |q|^2}}(p|B_s^0\rangle \mp q|\bar{B}_s^0\rangle), \tag{1.18}$$

where H is the heavy state, and L is the light state, and p and q describe the proportions of B_s^0 and $\bar{B}_s^0$ in the mass eigenstate. The experimentally observable mass difference is defined as $\Delta m_s = m_H - m_L$, the decay width difference as $\Delta\Gamma = \Gamma_H - \Gamma_L$, and the average decay width as $\Gamma = (\Gamma_H + \Gamma_L)/2$. One can also solve for the ratio q/p:

$$\left(\frac{q}{p}\right)^2 = \left|\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}\right| \tag{1.19}$$

Translating back from mass eigenstates to flavor eigenstates, we can now express the time dependence of the flavor eigenstates, using information gleaned from studying the mass eigenstates. After time t, an initially pure B_s^0 or $\bar{B}_s^0$ will be:

$$|B_s^0(t)\rangle = g_+(t)|B_s^0\rangle - \frac{q}{p}g_-(t)|\bar{B}_s^0\rangle, \tag{1.20}$$

$$|\bar{B}_s^0(t)\rangle = g_+(t)|\bar{B}_s^0\rangle - \frac{p}{q}g_-(t)|B_s^0\rangle \tag{1.21}$$

where

$$g_\pm = \frac{1}{2}(e^{-im_H t - \frac{1}{2}\Gamma_H t} \pm e^{-im_L t - \frac{1}{2}\Gamma_L t}). \tag{1.22}$$

With this information, we can calculate the probability of the initial B_s^0 or $\bar{B}_s^0$ being pure B_s^0 or $\bar{B}_s^0$ at a later time t:

$$Prob\langle B_s^0(t)|B_s^0(0)\rangle = \frac{1}{4}[e^{-\Gamma_H t} + e^{-\Gamma_L t} + 2e^{-\Gamma t}cos(\Delta m_s t)] \tag{1.23}$$

$$Prob\langle \bar{B}_s^0(t)|B_s^0(0)\rangle = \frac{1}{4}\left|\frac{p}{q}\right|[e^{-\Gamma_H t} + e^{-\Gamma_L t} - 2e^{-\Gamma t}cos(\Delta m_s t)] \tag{1.24}$$

$$Prob\langle B_s^0(t)|\bar{B}_s^0(0)\rangle = \frac{1}{4}\left|\frac{q}{p}\right|[e^{-\Gamma_H t} + e^{-\Gamma_L t} - 2e^{-\Gamma t}cos(\Delta m_s t)] \tag{1.25}$$

$$Prob\langle \bar{B}_s^0(t)|\bar{B}_s^0(0)\rangle = \frac{1}{4}[e^{-\Gamma_H t} + e^{-\Gamma_L t} + 2e^{-\Gamma t}cos(\Delta m_s t)]. \tag{1.26}$$

In the case of the B_s^0 system, the mixing frequency Δm_s is very rapid, $\Delta m_s = 17.77\pm0.12$ ps^{-1}, as opposed to $\Delta m_d \approx 0.5$ ps^{-1} for B^0 mixing. As a result, the probability that the initial particle will have oscillated to its antiparticle at decay time is $\sim$50% [5].

1.2.2 Direct *CP* Violation

Direct CP violation, or CP violation in decay occurs when the condition

$$|\bar{A}_{\bar{f}}/A_f| \neq 1 \tag{1.27}$$

is met, where $\bar{A}_{\bar{f}} = \langle \bar{f}|H|\bar{B}\rangle$ and $A_f = \langle f|H|B\rangle$. The standard model expectation for direct CP violation in the B_s^0 system is small, on order of λ^2, where λ is the Wolfenstein parameter [8].

1.2.3 *CP* Violation in Mixing

CP violation occurs in mixing when

$$|q/p| \neq 1 \tag{1.28}$$

From Eq. 1.19, $|p/q|$ is dependent on $M_{12}^{(*)}$ and $\Gamma_{12}^{(*)}$. It can be shown that CP violation in mixing occurs when

$$Im(M_{12}\Gamma_{12}^*) \neq 0. \tag{1.29}$$

In other words, M_{12} and Γ_{12} cannot be collinear in the complex plane [5]. This induces a time-dependent decay rate asymmetry

$$\mathcal{A} = \frac{d\Gamma/dt(B \to f) - d\Gamma/dt(\bar{B} \to \bar{f})}{d\Gamma/dt(B \to f) + d\Gamma/dt(\bar{B} \to \bar{f})} = \frac{1 - |q/p|^4}{1 + |q/p|^4}. \tag{1.30}$$

The phase associated with CP violation in mixing in the B_s^0 system has as observed upper limit of $\phi_s \sim$0.01 through the measurement of the asymmetry defined in Eq. 1.30 on semileptonic B_s^0 decays. The standard model expectation, ϕ_s^{SM}, is approximately 0.004 [8].

1.2.4 *CP* Violation in the Interference Between Mixing and Decay

CP violation in the interference between mixing and decay refers to the case when the particle and anti-particle can both decay to the same final state, and oscillation between particle and anti-particle can occur before the decay. A schematic of this process is shown in Fig. 1.6 for $B^0(\bar{B}^0) \to J/\psi K_S^0$ and $B_s^0(\bar{B}_s^0) \to J/\psi\phi$.

Figure 1.6: Schematic of the interfering amplitudes for direct decays and decays via mixing.

The Feynman diagrams for the interfering processes $B_s^0 \to J/\psi\phi$ and $B_s^0 \to \bar{B}_s^0$, which can also decay to $J/\psi\phi$, are shown in Fig. 1.7.

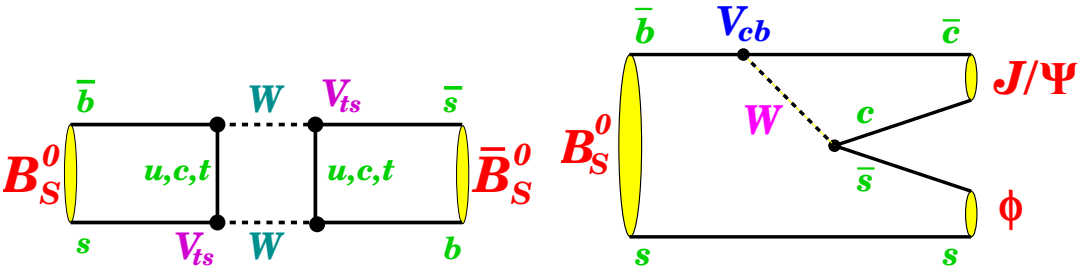

Figure 1.7: Leading order diagrams for $B_s^0 \to \bar{B}_s^0$, which can decay to $J/\psi\phi$, and $B_s^0 \to J/\psi\phi$. The interference of these processes can produce a CP asymmetry.

The CP violation in the interference of direct decays and decays via mixing is defined by

$$Im(\lambda_f) \neq 0 \tag{1.31}$$

where

$$\lambda_f = \frac{q}{p}\frac{\bar{A}_f}{A_f}. \tag{1.32}$$

The CP asymmetry in this case is

$$\mathcal{A} = \frac{d\Gamma/dt(B \to f_{CP}) - d\Gamma/dt(\bar{B} \to f_{CP})}{d\Gamma/dt(B \to f_{CP}) + d\Gamma/dt(\bar{B} \to f_{CP})}, \tag{1.33}$$

where f_{CP} is a CP pure final state. This reduces to a very simple form if direct CP violation and violation in mixing are small, and $\Delta\Gamma$ is close to zero:

$$\mathcal{A} = -Im\lambda_f \sin(\Delta m_s t) \tag{1.34}$$

We can, with great ease, define the CP violating phase β_s by defining

$$\lambda_f = \frac{q}{p}\frac{\bar{A}_f}{A_f} = e^{-2i\beta_s}. \tag{1.35}$$

We will now demonstrate that the CP violating phase that is produced by the interference of decays with and without mixing is, in fact, the same β_s defined by the CKM unitarity triangle.

Assuming that CP violation in mixing is small, we can make the following approximations:

$$\Delta m_s = m_H - m_L \approx 2M_{12}, \quad \Delta\Gamma = \Gamma_H - \Gamma_L \approx 2|\Gamma_{12}|\cos(\phi_s) \tag{1.36}$$

where ϕ_s, the CP violating phase in mixing, is defined by

$$\frac{M_{12}}{\Gamma_{12}} = -\left|\frac{M_{12}}{\Gamma_{12}}\right| e^{i\phi_s} \tag{1.37}$$

Eq. 1.19 can be expanded such that

$$\lambda_f = e^{-i\phi_M}\left(1 - \frac{1}{2}\left|\frac{\Gamma_{12}}{M_{12}}\right|\sin\phi_s\right)\frac{\bar{A}_f}{A_f} = e^{-2i\beta_s} \tag{1.38}$$

where ϕ_M is the (not CP violating) mixing phase. The quantity $e^{-i\phi_M}$ is given as $V_{tb}^* V_{ts}/V_{tb}V_{ts}^*$, as one can guess from looking at the vertices in the mixing box diagram. The phase ϕ_s is expected to be small, approximately 0.004, allowing us to drop the second term in Eq. 1.38. Note $\bar{A}_f/A_f$ can be expressed in terms of the contributing CKM matrix elements in the $\bar{b} \to \bar{c}c\bar{s}$ transition. The dominant CKM dependence for this transition is on $(V_{cb}^* V_{cs})$ (or $(V_{cb}V_{cs}^*)$ for the $b \to \bar{c}cs$ transition). λ_f becomes:

$$\lambda_f = \frac{V_{tb}^* V_{ts}}{V_{tb}V_{ts}^*}\frac{V_{cb}V_{cs}^*}{V_{cb}^* V_{cs}} = e^{-2i\beta_s}. \tag{1.39}$$

Taking the imaginary part of λ_f, we see that

$$Im(\lambda_f) = -\sin(2\beta_s). \tag{1.40}$$

The CKM angle β_s given in Eq. 1.10 is indeed equivalent to the CP violating phase from the interference of direct decays and decays via mixing!

The standard model expectation is $\beta_s \approx 0.02$. Measuring a large value of β_s would be a clear sign of new physics participation in the mixing box diagram.

It is worth noting that a large, non-standard model CP violating phase ϕ^{NP} would affect β_s and ϕ_s in the same way. We choose our sign conventions such that

$$\phi_s = \phi_s^{SM} + \phi^{NP} \qquad 2\beta_s = 2\beta_s^{SM} - \phi^{NP}. \tag{1.41}$$

Both β_s and ϕ_s are predicted to be small in the standard model, so the new physics phase will dominate, allowing us to equate the two CP violating phases, and to relate another important quantity, $\Delta\Gamma$, to β_s.

$$2\beta_s \approx -\phi_s, \qquad \Delta\Gamma \approx 2|\Gamma_{12}|\cos(2\beta_s). \tag{1.42}$$

1.3 Analysis of Scalar to Vector-Vector Decays

We now turn to the phenomenology of measuring β_s using $B_s^0 \to (\bar{B}_s^0) \to J/\psi\phi$ decays (branching fraction = 0.00093±0.00033), with the J/ψ decaying to $\mu^+\mu^-$ (branching fraction = 5.93±0.06%), and the ϕ decaying to K^+K^- (branching fraction = 48.9±0.5). Eq. 1.33 gives the form of a measurable CP asymmetry produced by CP violation in the interference between direct decays and decays via mixing. The asymmetry is accessible by measuring differential decay rates of initial B_s^0 and $\bar{B}_s^0$ to a CP pure final state.

Unfortunately, $B_s^0 \to J/\psi\phi$ decays do not result in a CP pure final state, but to an admixture of CP-odd and CP-even states. The B_s^0 meson, a spin-0 pseudoscalar, decays to J/ψ and ϕ, both spin-1 vector particles. The spins of the vector particles add to produce multiple orbital angular momentum final states, one CP-odd and two CP-even. An angular analysis is used to determine relative CP-odd and even proportions of the $J/\psi\phi$ final state [11, 12, 13].

Combining the angular momenta from two spin-1 particles, the angular momentum in the final state includes contributions from L=0, 1, and 2. These angular states are also referred to as S-wave (L=0), P-wave (L=1) and D-wave (L=2). The L=0 and L=2 states are CP even, and the L=1 state is CP odd.

The CP of the final angular momentum states can be determined using $CP_{tot} = CP_1CP_2(-1)^L$. J/ψ has C=−1 and P=−1, ϕ has C=−1 and P=−1, and f_0 and non-resonant K^+K^- have C=+1 and P=+1. We find that for the $J/\psi\phi$ final state, CP_{tot} is even for L=0, odd for L=1, and even for L=2.

An angular analysis can be used to measure the relative proportions of CP-odd and CP-even in the final state, and allow β_s to be extracted. We use flavor-tagging, discussed in the experimental methods section, to increase sensitivity to β_s by determining whether the decaying particle was a B_s^0 or $\bar{B}_s^0$ at production.

With the goal of determining the relative proportion of CP-even to CP-odd in the final state, it is convenient to choose a basis where the measured angular amplitudes can be directly related to CP states. The transversity basis satisfies this requirement [11]. It identifies three "transversity" amplitudes $A_0, A_{||}$ and $A_\perp$, according to the relative linear polarizations of the vector particles. For A_0, the vector particles are longitudinally polarized, for $A_{||}$, their polarization is transverse to their direction of motion, and parallel to each other, and for $A_\perp$, their polarization is transverse to their direction of

motion and perpendicular to each other. The transversity amplitudes can be related to the CP. The amplitude $A_\perp$ is CP odd, while $A_{||}$ and A_0 are CP even. We choose the normalization $|A_0|^2 + |A_{||}|^2 + |A_\perp|^2 = 1$.

Each amplitude has an associated phase. We have the freedom to set one phase to zero and retain two relative phases. We set $\delta_0 = arg(A_0)$=0, and retain $\delta_{||} = arg(A_{||})$ and $\delta_\perp = arg(A_\perp)$.

We also define a set of three transversity angles to describe the final state angular distribution, shown in Fig. 1.8 [11]. The $+x$ axis is aligned with the direction of motion of the ϕ in the J/ψ rest frame, and the K^+ and K^-, decaying from the ϕ, define the xy plane, with the direction of the K^+ in the $+y$ direction. θ and ϕ are transversity angles in the J/ψ rest frame. θ is the polar angle, drawn from the $+z$ axis to the μ^+ momentum vector. ϕ is the azimuthal angle, drawn from the $+x$ axis counter-clockwise to the projection of the μ^+ momentum vector in the xy plane. ψ is a helicity angle in the ϕ rest frame, drawn from the axis defined by the negative of the J/ψ momentum direction to the K^+ momentum direction. In the following sections, we often refer to the set of angles as $\vec{\rho} = (\cos(\theta), \phi, \cos(\psi))$.

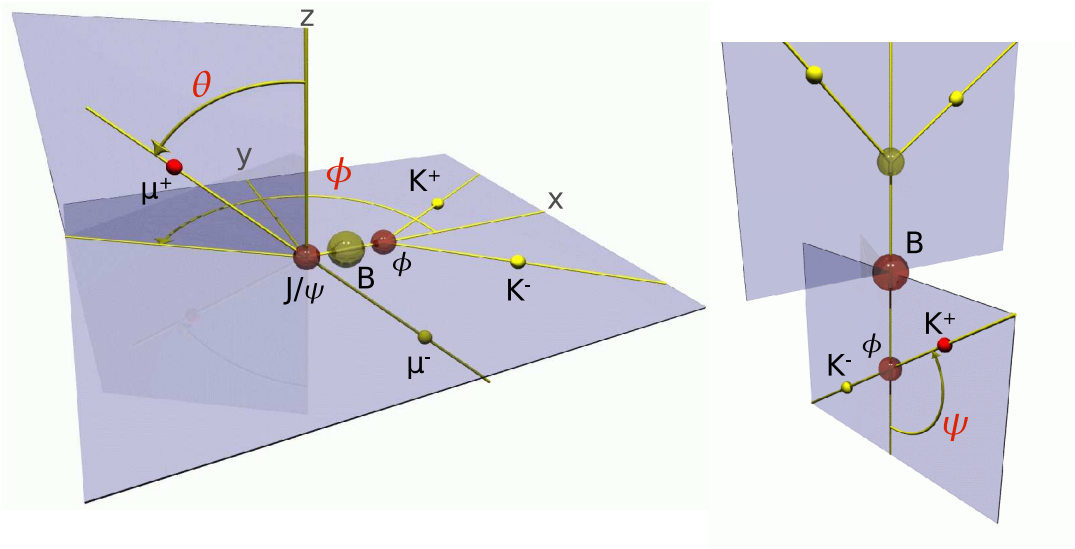

Figure 1.8: The chosen angular basis includes θ and ϕ, defined in the J/ψ rest frame, and ψ, defined in the ϕ rest frame.

The angular and time dependence of $B_s^0 \to J/\psi\phi$ decays is contained in the formulae

for the differential decay rate $d^4\Gamma(\vec{\rho},t)/d\vec{\rho}dt$. The differential decay rate is also dependent on β_s, and measuring the decay rate allows β_s to be measured as well [11, 12, 13].

The differential decay rate for B_s^0 and $\bar{B}_s^0$ is equivalent to a probability density (also time and angularly dependent) for B_s^0 and $\bar{B}_s^0$ [11]. The probability densities are easily expressed in a compact form, so we use this description here. The equivalence of $d^4\Gamma(\vec{\rho},t)/d\vec{\rho}dt$ to the probability density, with full expansion of the time and angular dependence, is given in Appendix A.

1.3.1 Decay rate

In the transversity basis, the probability density for B_s^0 or $\bar{B}_s^0$ is succinctly expressed as [14]

$$\begin{aligned} P_B(\theta,\phi,\psi,t) &= \frac{9}{16\pi}|\mathbf{A}(t)\times\hat{n}|^2 \\ P_{\bar{B}}(\theta,\phi,\psi,t) &= \frac{9}{16\pi}|\bar{\mathbf{A}}(\mathbf{t})\times\hat{n}|^2, \end{aligned} \tag{1.43}$$

where the dependence on two of the three transversity angles is contained in the unit vector $\hat{n}$

$$\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{1.44}$$

and the transversity amplitudes $(A_0, A_{||}, A_\perp)$ and the dependence on the remaining transversity angle ψ is contained in the complex vector $\mathbf{A}$:

$$\begin{aligned} \mathbf{A}(t) &= \left(A_0(t)\cos\psi, -\frac{A_\parallel(t)\sin\psi}{\sqrt{2}}, i\frac{A_\perp(t)\sin\psi}{\sqrt{2}}\right) \\ \bar{\mathbf{A}}(t) &= \left(\bar{A}_0(t)\cos\psi, -\frac{\bar{A}_\parallel(t)\sin\psi}{\sqrt{2}}, i\frac{\bar{A}_\perp(t)\sin\psi}{\sqrt{2}}\right). \end{aligned} \tag{1.45}$$

The probabilities, and thus the transversity amplitudes, are normalized such that

$$\int\int\int\int \sum_{j=B,\bar{B}} P_j(\psi,\theta,\phi,t)d(\cos\psi)d(\cos\theta)d\phi dt = 1. \tag{1.46}$$

The details of this normalization will be discussed in Sec. 4.2.1.

The time dependence of the transversity amplitudes can be calculated from the transition of the flavor eigenstates given in Eq. 1.22 to the final state f_i:

$$A_i(t) = \langle f_i|H|B_s^0\rangle\langle B_s^0|B_{s,phys}^0(t)\rangle + \langle f_i|H|\bar{B}_s^0\rangle\langle \bar{B}_s^0|B_{s,phys}^0(t)\rangle \tag{1.47}$$

and

$$\bar{A}_i(t) = \langle f_i|H|\bar{B}_s^0\rangle\langle B_s^0|B_{s,phys}^0(t)\rangle + \langle f_i|H|\bar{B}_s^0\rangle\langle \bar{B}_s^0|\bar{B}_{s,phys}^0(t)\rangle. \tag{1.48}$$

where $B_{s,phys}^0$ is the state of the particle at time t, in some superposition of mixed and unmixed. Explicitly expressing the time evolution, the amplitudes can be shown to be:

$$\begin{aligned} A_i(t) &= \frac{e^{-imt}e^{-\Gamma t/2}}{\sqrt{\tau_H + \tau_L \pm \cos 2\beta_s\,(\tau_L - \tau_H)}}\left[E_+(t) \pm e^{2i\beta_s}E_-(t)\right]A_i(t=0), \\ \bar{A}_i(t) &= \frac{e^{-imt}e^{-\Gamma t/2}}{\sqrt{\tau_H + \tau_L \pm \cos 2\beta_s\,(\tau_L - \tau_H)}}\left[\pm E_+(t) + e^{-2i\beta_s}E_-(t)\right]A_i(t=0), \end{aligned} \tag{1.49}$$

where

$$E_\pm(t) \equiv \frac{1}{2}\left[e^{+(\frac{-\Delta\Gamma}{4}+i\frac{\Delta m_s}{2})t} \pm e^{-(\frac{-\Delta\Gamma}{4}+i\frac{\Delta m_s}{2})t}\right]. \tag{1.50}$$

A useful way of expressing the time evolution of the probabilities is in terms of the time evolution of the CP-odd and even components. We start by separating the complex vector $\mathbf{A}$ into its CP-even $(+)$ and CP-odd $(-)$ components

$$\mathbf{A}(t) = \mathbf{A}_+(t) + \mathbf{A}_-(t), \quad \bar{\mathbf{A}}(t) = \bar{\mathbf{A}}_+(t) + \bar{\mathbf{A}}_-(t). \tag{1.51}$$

Then, we separate the time evolution, with the remaining term depending only on on the values of the transversity amplitudes at t=0. For the CP-even terms, we obtain

$$\begin{aligned} \mathbf{A}_+(t) = \mathbf{A}_+ f_+(t) &= \left(A_0\cos\psi, -\frac{A_\parallel \sin\psi}{\sqrt{2}}, 0\right)\cdot f_+(t) \\ \bar{\mathbf{A}}_+(t) = \bar{\mathbf{A}}_+ \bar{f}_+(t) &= \left(A_0\cos\psi, -\frac{A_\parallel \sin\psi}{\sqrt{2}}, 0\right)\cdot \bar{f}_+(t), \end{aligned} \tag{1.52}$$

and for the CP-odd terms,

$$\begin{aligned}\mathbf{A}_-(t) &= \mathbf{A}_- f_-(t) = \left(0, 0, i\frac{A_\perp \sin\psi}{\sqrt{2}}\right)\cdot f_-(t)\\ \bar{\mathbf{A}}_-(t) &= \bar{\mathbf{A}}_- \bar{f}_-(t) = \left(0, 0, i\frac{A_\perp \sin\psi}{\sqrt{2}}\right)\cdot \bar{f}_-(t).\end{aligned} \tag{1.53}$$

Expanding the probability density for the B_s^0 and $\bar{B}_s^0$ states gives

$$\begin{aligned}P_B(\theta,\psi,\phi,t) &= \frac{9}{16\pi}\left\{|\mathbf{A}_+(\mathbf{t})\times\hat{n}|^2 + |\mathbf{A}_-(\mathbf{t})\times\hat{n}|^2 + 2Re([\mathbf{A}_+(\mathbf{t})\times\hat{n}]\cdot[\mathbf{A}_-^*(\mathbf{t})\times\hat{n}])\right\}\\ &= \frac{9}{16\pi}\left\{|\mathbf{A}_+\times\hat{n}|^2|f_+(t)|^2 + |\mathbf{A}_-\times\hat{n}|^2|f_-(t)|^2\right\}\\ &+ \frac{9}{16\pi}\left\{2Re((\mathbf{A}_+\times\hat{n})\cdot(\mathbf{A}_-^*\times\hat{n})\cdot f_+(t)\cdot f_-^*(t))\right\}\end{aligned} \tag{1.54}$$

and

$$\begin{aligned}P_{\bar{B}}(\theta,\psi,\phi,t) &= \frac{9}{16\pi}\left\{|\bar{\mathbf{A}}_+(t)\times\hat{n}|^2 + |\bar{\mathbf{A}}_-(t)\times\hat{n}|^2 + 2Re[\bar{\mathbf{A}}_+(t)\times\hat{n}]\cdot[\bar{\mathbf{A}}_-^*(t)\times\hat{n}])\right\}\\ &= \frac{9}{16\pi}\left\{|\mathbf{A}_+\times\hat{n}|^2|\bar{f}_+(t)|^2 + |\mathbf{A}_-\times\hat{n}|^2|\bar{f}_-(t)|^2\right\}\\ &+ \frac{9}{16\pi}\left\{2Re((\mathbf{A}_+\times\hat{n})\cdot(\mathbf{A}_-^*\times\hat{n})\cdot\bar{f}_+(t)\cdot\bar{f}_-^*(t)\right\}.\end{aligned} \tag{1.55}$$

The non-interfering time-dependent terms are:

$$\begin{aligned}|\bar{f}_\pm(t)|^2 &= \frac{1}{2}\frac{(1\pm\cos 2\beta_s)e^{-\Gamma_L t} + (1\mp\cos 2\beta_s)e^{-\Gamma_H t} \pm 2\sin 2\beta_s e^{-\Gamma t}\sin\Delta m_s t}{\tau_L(1\pm\cos 2\beta_s) + \tau_H(1\mp\cos 2\beta_s)},\\ |f_\pm(t)|^2 &= \frac{1}{2}\frac{(1\pm\cos 2\beta_s)e^{-\Gamma_L t} + (1\mp\cos 2\beta_s)e^{-\Gamma_H t} \mp 2\sin 2\beta_s e^{-\Gamma t}\sin\Delta m_s t}{\tau_L(1\pm\cos 2\beta_s) + \tau_H(1\mp\cos 2\beta_s)}\end{aligned} \tag{1.56}$$

and the interference terms are:

$$\begin{aligned}\bar{f}_+(t)\bar{f}_-{}^*(t) &= \frac{-e^{-\Gamma t}\cos\Delta m_s t - i\cos 2\beta_s e^{-\Gamma t}\sin\Delta m_s t + i\sin 2\beta_s(e^{-\Gamma_L t} - e^{-\Gamma_H t})/2}{\sqrt{[(\tau_L-\tau_H)\sin 2\beta_s]^2 + 4\tau_L\tau_H}},\\ f_+(t)f_-^*(t) &= \frac{e^{-\Gamma t}\cos\Delta m_s t + i\cos 2\beta_s e^{-\Gamma t}\sin\Delta m_s t + i\sin 2\beta_s(e^{-\Gamma_L t} - e^{-\Gamma_H t})/2}{\sqrt{[(\tau_L-\tau_H)\sin 2\beta_s]^2 + 4\tau_L\tau_H}}.\end{aligned} \tag{1.57}$$

A simplification to the probability density that is sometimes utilized is the so-called "untagged" version, where we do not use flavor tagging to determine whether the B meson was a B_s^0 or $\bar{B}_s^0$ at production. This simplification removes sensitivity to any terms containing Δm_s. When the probability density function for the particle and the anti-particle cannot be distinguished from each other, we must use the summed probability of a B_s^0 or $\bar{B}_s^0$. In this case, all the terms containing Δm_s cancel, because they have the opposite signs for B_s^0 and $\bar{B}_s^0$.

1.3.2 Decay rate including a K^+K^-/f_0 S-wave contribution

An additional contribution to the final state comes from non-resonant decays of $B_s^0 \to (\bar{B}_s^0) \to J/\psi K^+K^-$ or $B_s^0 \to (\bar{B}_s^0) \to J/\psi f_0$. Both states are approximately flat in the region of the resonant ϕ invariant mass. Since they do not have an easily detectable shape, there is a danger of identifying decays to the non-resonant or f_0 state as $B_s^0 \to J/\psi\phi$ decays, contaminating the signal, and disturbing the measurement of the value of β_s. The contributions of f_0 and non-resonant K^+K^- in the measurement of β_s are considered for the first time in this thesis. Both the f_0 and non-resonant K^+K^- are spin 0, and so the angular momentum of the final state, combined with the spin of the J/ψ, is L=1. For the $J/\psi f_0$ or $J/\psi K^+K^-$ final state, CP_{tot} is odd.

We have a phase δ_S for the f_0 and non-resonant K^+K^- contribution, which can take any value with respect to the dominant decay amplitude. We also make use of the K^+K^-/f_0 S-wave fraction f_{swave}.

Including an f_0 or non-resonant K^+K^- in the final state state is a major change to the probability densities detailed in the previous section. These additional S-wave terms, as we will see, interfere with the dominant decay amplitudes, and break a symmetry inherent in the original probability density.

Based on estimates from analogous decays [15], the amount of predicted f_0 and non-resonant K^+K^- contribution in our data is small, less than 10%. The amount is expected to be reduced substantially by the selection requirements employed in this measurement, particularly requirements on the K^+K^- invariant mass, which is required to be in the range [1.009, 1.032] GeV/c^2. For a small fraction, we can approximate the combined f_0 and non-resonant K^+K^- contributions as a flat line shape. The phase δ_S is expected to vary quite slowly across the ϕ mass range. We approximate it as flat across the ϕ mass range, based on simulation.

We perform a mass-integration of the f_0/non-resonant K^+K^- probability density over of the ϕ window. The ϕ mass line shape is taken as a relativistic Breit-Wigner with a mass-dependent width. The full mathematical description of the line shape is given in Appendix B. Due to the complicated shape, a numerical integration is performed. A slightly simplified description of the probability density, using a non-relativistic Breit-Wigner and analytic integration is given here, as a more straight-forward illustration of the effect of additional S-wave contributions on the total probability density [14].

The normalized probability density for $B_s^0 \to (\bar{B}_s^0) \to J/\psi f_0/K^+K^-$ decays is analogous to Eq. 1.43

$$\begin{aligned} Q_B(\theta,\phi,\psi,t) &= \frac{3}{16\pi}|\mathbf{B}(t)\times\hat{n}|^2 \\ Q_{\bar{B}}(\theta,\phi,\psi,t) &= \frac{3}{16\pi}|\bar{\mathbf{B}}(\mathbf{t})\times\hat{n}|^2. \end{aligned} \tag{1.58}$$

$\mathbf{B}$ and $\bar{\mathbf{B}}$ contain the explicit time dependence of the additional amplitude.

$$\begin{aligned} \mathbf{B}(t) &= (B(t),0,0) \\ \bar{\mathbf{B}}(t) &= (\bar{B}(t),0,0) \end{aligned} \tag{1.59}$$

where

$$\begin{aligned} B(t) &= \frac{e^{-\Gamma t/2}}{\sqrt{\tau_H+\tau_L-\cos 2\beta_s\,(\tau_L-\tau_H)}}\left[E_+(t)-e^{2i\beta_s}E_-(t)\right], \\ \bar{B}(t) &= \frac{e^{-\Gamma t/2}}{\sqrt{\tau_H+\tau_L-\cos 2\beta_s\,(\tau_L-\tau_H)}}\left[-E_+(t)+e^{-2i\beta_s}E_-(t)\right]. \end{aligned} \tag{1.60}$$

The amplitude from f_0 and non-resonant K^+K^- contribution is added to the dominant decay amplitude, and their total amplitude squared. The resulting probability has a mass dependence, due to the fact that two amplitudes interfere in the ϕ mass region. Otherwise, we could simply sum the two squared amplitudes. We define a generic K^+K^- invariant mass range $[\mu_{lo},\,\mu_{hi}]$, in which the K^+K^- invariant mass μ lies. We now have a total probability density of

$$\begin{aligned} \rho_B(\theta,\phi,\psi,t,\mu) &= \frac{9}{16\pi}\left|\left[\sqrt{1-f_{swave}}g(\mu)\mathbf{A}(t)+e^{i\delta_S}\sqrt{f_{swave}}\frac{h(\mu)}{\sqrt{3}}\mathbf{B}(t)\right]\times\hat{n}\right|^2 \\ \rho_{\bar{B}}(\theta,\phi,\psi,t,\mu) &= \frac{9}{16\pi}\left|\left[\sqrt{1-f_{swave}}g(\mu)\bar{\mathbf{A}}(t)+e^{i\delta_S}\sqrt{f_{swave}}\frac{h(\mu)}{\sqrt{3}}\bar{\mathbf{B}}(\mathbf{t})\right]\times\hat{n}\right|^2. \end{aligned} \tag{1.61}$$

where

$$g(\mu) = \sqrt{\frac{\Gamma_\phi/2}{\Delta\omega}} \cdot \frac{1}{\mu - \mu_\phi + i\Gamma_\phi/2} \tag{1.62}$$

and

$$h(\mu) = \frac{1}{\sqrt{\Delta\mu}} \tag{1.63}$$

and

$$\Delta\mu = \mu_{hi} - \mu_{lo}. \tag{1.64}$$

The quantity μ_ϕ is the ϕ mass (1.019 GeV/c^2), and Γ_ϕ is the ϕ width (4.26 MeV/c^2).

As in the case of the dominant decay, we rewrite the probability density in a way that separates the *CP*-even and odd time evolutions from each other and from the transversity amplitudes. In the case of the f_0 and non-resonant K^+K^-, the time evolution of the final state is *CP*-odd.

$$\begin{aligned} Q_B(\theta,\psi,\phi,t) &= \frac{3}{16\pi}|\mathbf{B}(\mathbf{t}) \times \hat{n}|^2 \\ &= \frac{3}{16\pi}|\mathbf{B} \times \hat{n}|^2 |f_-(t)|^2 \end{aligned} \tag{1.65}$$

and

$$\begin{aligned} Q_{\bar{B}}(\theta,\psi,\phi,t) &= \frac{3}{16\pi}|\bar{\mathbf{B}}(t) \times \hat{n}|^2 \\ &= \frac{3}{16\pi}|\bar{\mathbf{B}} \times \hat{n}|^2 |\bar{f}_-(t)|^2, \end{aligned} \tag{1.66}$$

where $\mathbf{B}$ is the vector (1, 0, 0).

The total probability density is now:

$$\begin{aligned} \rho_B(\theta,\psi,\phi,t,\mu) &= (1 - f_{swave})\frac{\Gamma_\phi/2}{\Delta\omega} \cdot \frac{1}{(\mu-\mu_\phi)^2 + \Gamma_\phi^2/4} \cdot P_B(\theta,\psi,\phi,t) \\ &+ f_{swave}\frac{1}{\Delta\mu}Q_B(\theta,\psi,\phi,t) \\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[\mathcal{F}(\mu)\left((\mathbf{A}_- \times \hat{n}) \cdot (\mathbf{B} \times \hat{n}) \cdot |f_-(t)|^2\right)\right] \\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[\mathcal{F}(\mu)\left((\mathbf{A}_+ \times \hat{n}) \cdot (\mathbf{B} \times \hat{n}) \cdot f_+(t) \cdot f_-^*(t)\right)\right] \end{aligned} \tag{1.67}$$

and

$$\begin{aligned}\rho_{\bar{B}}(\theta,\psi,\phi,t,\mu) &= (1-f_{swave})\frac{\Gamma_\phi/2}{\Delta\omega}\cdot\frac{1}{(\mu-\mu_\phi)^2+\frac{1}{1+f_{swave}}\Gamma_\phi^2/4}\cdot P_{\bar{B}}(\theta,\psi,\phi,t)\\ &+ f_{swave}\frac{1}{\Delta\mu}Q_{\bar{B}}(\theta,\psi,\phi,t)\\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[\mathcal{F}(\mu)\left((\mathbf{A}_-\times\hat{n})\cdot(\mathbf{B}\times\hat{n})\cdot|\bar{f}_-(t)|^2\right)\right]\\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[\mathcal{F}(\mu)\left((\mathbf{A}_+\times\hat{n})\cdot(\mathbf{B}\times\hat{n})\cdot\bar{f}_+(t)\cdot\bar{f}_-^*(t)\right)\right] \qquad (1.68)\end{aligned}$$

where

$$\mathcal{F}(\mu)\equiv\sqrt{\frac{f_{swave}(1-f_{swave})\Gamma_\phi}{2\Delta\mu\Delta\omega}}\cdot\frac{e^{i\delta_S}}{\mu-\mu_\phi-i\Gamma_\phi/2}. \qquad (1.69)$$

In this thesis, we do not perform a mass dependent analysis. We integrate out the mass dependence, leaving the final probability densities

$$\begin{aligned}\rho_B(\theta,\psi,\phi,t) &= (1-f_{swave})\cdot P_B(\theta,\psi,\phi,t)+f_{swave}Q_B(\theta,\psi,\phi,t)\\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[\mathcal{I}_\mu\left((\mathbf{A}_-\times\hat{n})\cdot(\mathbf{B}\times\hat{n})\cdot|f_-(t)|^2\right)\right]\\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[(\mathbf{A}_+\times\hat{n})\cdot(\mathbf{B}\times\hat{n})\cdot f_+(t)\cdot f_-^*(t)\right] \qquad (1.70)\end{aligned}$$

and

$$\begin{aligned}\rho_{\bar{B}}(\theta,\psi,\phi,t) &= (1-f_{swave})\cdot P_{\bar{B}}(\theta,\psi,\phi,t)+f_{swave}Q_{\bar{B}}(\theta,\psi,\phi,t)\\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[\mathcal{I}_\mu\left((\mathbf{A}_-\times\hat{n})\cdot(\mathbf{B}\times\hat{n})\cdot|\bar{f}_-(t)|^2\right)\right]\\ &+ 2\frac{\sqrt{27}}{16\pi}Re\left[(\mathbf{A}_+\times\hat{n})\cdot(\mathbf{B}\times\hat{n})\cdot\bar{f}_+(t)\cdot\bar{f}_-^*(t)\right] \qquad (1.71)\end{aligned}$$

where

$$\mathcal{I}_\mu\equiv\int\mathcal{F}(\mu)d\mu=\sqrt{\frac{f_{swave}(1-f_{swave})\Gamma_\phi}{2\Delta\mu\Delta\omega}}\cdot e^{i\delta_S}\cdot\log\frac{\mu_{hi}-\mu_\phi-i\Gamma/2}{\mu_{lo}-\mu_\phi-i\Gamma/2}. \qquad (1.72)$$

Eq. 1.70 and Eq. 1.71 are the final probability densities used in an unbinned maxium likelihood fit to determine β_s.

1.3.3 Symmetries

The probability density for the case of $B_s^0 \to (\bar{B}_s^0) \to J/\psi\phi$ decays, without considering f_0 and non-resonant K^+K^- contributions, has an inherent symmetry under the following transformation:

$$(\beta_s, \Delta\Gamma, \delta_\perp, \delta_{||}) \Leftrightarrow (\pi/2 - \beta_s, -\Delta\Gamma, \pi - \delta_\perp, 2\pi - \delta_{||}). \tag{1.73}$$

This symmetry produces an ambiguity in the measured value of β_s. For any measured β_s, an equally valid solution exists under the above transformation of variables. Fundamentally, the symmetry and ambiguity are due to the fact that we can only measure relative, not absolute values for the strong phases, and β_s enters the probability densities as a sine or cosine.

The addition of the f_0/non-resonant K^+K^- contribution to the final state removes this symmetry, and the ambiguity in the solution for β_s [16]. This occurs via interference between the f_0/non-resonant K^+K^- partial wave and the dominant decay in the invariant mass region of the ϕ resonance. According to Wigner's causality principle, the strong phase from the dominant decay will increase rapidly across the ϕ resonance [17]. As we have noted before, the phase of the f_0/non-resonant K^+K^- contribution, δ_S, is nearly flat across the resonance. We expect the phase differences $\delta_S - \delta_0$, $\delta_S - \delta_\perp$, and $\delta_S - \delta_{||}$ to decrease across the ϕ resonance. If the phase difference were constant across the ϕ resonance, it would be possible to find a transformation of δ_S and the strong phase that left the probability density invariant. However, the value of $\delta_S - \delta_0$ is different at each value of the K^+K^- invariant mass. It is impossible to find a transformation of the phases that holds for every point along the ϕ resonance. The symmetry of the probability density is therefore broken, and only one β_s solution remains.

1.4 B Meson Lifetimes

The observables $\Delta\Gamma$ and Γ are key variables in the time-dependent component of the measurement of β_s. The decay width Γ is simply related to the lifetime τ by $\tau = 1/\Gamma$. The lifetime for B mesons is long, on order of 1 ps ($\sim$1.5 ps for the B_s^0). This is the case because the top quark is extremely heavy compared to the bottom, forcing the bottom to decay inter-generationally. Such decays are suppressed by the factor $|V_{cb}|^2 \approx 0.04$.

We also refer frequently to the related proper decay length $c\tau$, where c is the speed of light. For the B_s^0 meson, $c\tau$ is $\sim$450 μm.

The "lifetime" quantity most directly calculable from information collected by the CDF detector is the proper decay time t. It is defined by

$$t = \frac{LM^B}{p} = \frac{L_{xy}M^B}{p_T} \tag{1.74}$$

where M^B is the mass of the B meson, and p and p_T are the momentum and transverse momentum of the B. The transverse decay length L_{xy} is shown schematically in Fig. 1.9. L is the distance from the primary event vertex to the decay vertex, and L_{xy} is the transverse distance from the decay vertex to the primary vertex, projected onto the B transverse momentum.

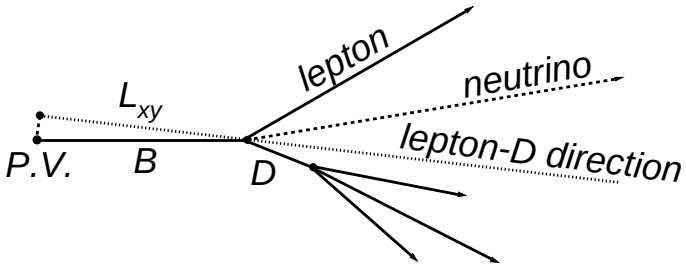

Figure 1.9: Schematic of a B decay, illustrating the transverse decay length L_{xy}.

The uncertainty on the proper time, σ_t, is directly related to the uncertainty on the transverse decay length in the case of decays for which all decay products are detected and used to reconstruct the decay:

$$\sigma_t = \frac{\sigma(L_{xy})M^B}{p_T}. \tag{1.75}$$

The proper time and lifetime both appear in the probability density function for the proper time for the B meson, which is exponentially decaying:

$$P(t) \sim \frac{1}{\tau}e^{-t/\tau}. \tag{1.76}$$

1.5 Current Experimental Status

The measurement of β_s has been carried out by both the CDF and DØ collaborations. Of course, a number of measurement were pre-cursors to the measurement of β_s, including measurements of the B_s^0 lifetime [18, 19, 20], the decay width $\Delta\Gamma$ [21], and the measurement of the $B_s^0/\bar{B}_s^0$ oscillation frequency Δm_s [22]. Also, both collaborations have performed measurements of the β_s without the use of flavor-tagging [23, 24].

The CDF collaboration has performed one flavor-tagged measurement of β_s using 1.35 fb^{-1} of data [25, 26, 27], and updated the measurement to a 2.8 fb^{-1} dataset in the summer of 2008 [28, 29]. The more recent result was compromised by the use of non-optimal flavor-tagging. DØ produced a flavor-tagged measurement of β_s, also on 2.8 fb^{-1}. A combined CDF/DØ result has been produced from the two 2.8 fb^{-1} results [30]. Neither CDF nor DØ includes the effect of decays to $J/\psi f_0$ or $J/\psi K^+K^-$.

Both CDF and DØ measure a positive deviation of β_s from the standard model prediction of β_s^{SM}=0.02. The deviations are not statistically significant: 1.7σ for the CDF-only measurement, and 2.3σ for the combined CDF/DØ result. Even so, the measurement has been a source of great excitement in the community as a realm for detecting possible non-standard model physics.

1.6 Analysis Strategy

This analysis relies on an angular analysis to separate the CP-odd and CP-even components of the $J/\psi\phi$ final state. The decay $B_s^0 \to J/\psi\phi$ is a pseudoscalar decay to two vector particles and, as such, includes $L = 0$, $L = 1$, and $L = 2$ components in the final state. The spin components of the final state can be related to the CP of the final state: $L = 0$ and $L = 2$ are CP even, while $L = 1$ is CP odd. We define the angular momentum distribution of the final state in terms of linear polarization of the vector particles using the transversity basis [31]. This allows us to describe the decay angular dependence in terms of transversity amplitudes A_0, $A_{||}$, and $A_\perp$. By definition, $|A_0|^2+|A_{||}|^2+|A_\perp|^2 = 1$. We also define the strong phases $\delta_0 = 0$, $\delta_{||} = arg(A_{||}(0))$ and $\delta_\perp = arg(A_\perp(0))$, which carry information about the initial relative polarizations of the vector particles. The strong phases and the transversity amplitudes are extracted by fitting for the transversity angles $cos(\theta)$, ϕ, and $cos(\psi)$.

We consider, for the first time, potential contamination from $B_s^0 \to J/\psi f_0$ and $B_s^0 \to J/\psi K^+K^-$. The f_0 and K^+K^- are S-wave partial waves whose angular momenta combine with that of the J/ψ to form a CP-odd final state. We perform a mass-integrated fit about the ϕ resonance to measure the fraction and phase of the f_0 and K^+K^- S-wave.

Sensitivity to β_s can be improved by separately tracking the time evolution of the B_s^0 and $\bar{B}_s^0$ mesons. A B_s^0 meson will mix to $\bar{B}_s^0$ many ($\sim$20) times before decaying to J/ψ and ϕ. Thus, to distinguish the time evolution of B_s^0 mesons the from time evolution of $\bar{B}_s^0$ mesons, we must employ flavor-tagging to determine the production flavor of the reconstructed meson. The flavor-tagging algorithms used in this analysis will be described in dedicated sections.

The angular analysis and flavor-tagged, time-dependent analysis are combined in an unbinned maximum likelihood fit. The fit is used to extract β_s, the decay width difference $\Delta\Gamma$ between the B_s^0 and $\bar{B}_s^0$ mass eigenstates, the average B_s^0 lifetime, the transversity amplitudes, the strong phases, and the K^+K^-/f_0 S-wave fraction and phase δ_S. The mass difference, $\Delta m_s = m_H - m_L$ features prominently in the likelihood. With the statistics of our current dataset, there is no sensitivity to Δm_s. The PDG average, Δm_s=17.77$\pm$0.12 ps^{-1} is taken as an input. We describe the derivation of the likelihood in detail in a dedicated section. A frequentist interpretations of the likelihood is presented. We quote a confidence region in the $\beta_s - \Delta\Gamma$ plane, in addition to a confidence interval on β_s, and point estimates for other parameters of interest in case of zero CP violation.

Chapter 2

Experimental Apparatus

2.1 Tevatron

The Tevatron was the world's highest energy accelerator until Fall 2009 when the Large Hadron Collider (LHC) turned on. The Tevatron is a proton-antiproton collider, located at Fermi National Accelerator Laboratory (Fermilab). The Tevatron, wherein the proton and antiproton beams rotate and collide, is 1 km in radius and is situated underground. Collisions between protons and antiprotons occur at a center of mass energy of 1.96 TeV. The proton and antiproton beams each contain 36 bunches, with each proton bunch containing typically 200 billion protons, and each antiproton bunch containing about 20 billion antiprotons, for 2009 operation. Bunch crossings occur every 396 ns. The beam trajectories in the Tevatron are guided such that collisions occur at two points along the main ring, corresponding to the location of two detectors, the CDF and DØ experiments. The data used in this thesis was collected by the CDF detector, also referred to as the CDF II detector to specify that is an upgraded version of the original CDF detector. We will describe the components of the CDF detector relevant to this analysis in the following sections, as well as the features of the accelerator complex that produces and accelerates the protons and antiprotons.

The Tevatron accelerator complex consists of proton and antiproton sources, and a series of accelerators used to raise the produced particles to from the 150 GeV injection energy to the energy necessary for 1.96 TeV center of matter collisions. A schematic of the accelerator complex is given in Fig. 2.1. Detailed information about the Tevatron is given in Ref. [32, 33]. We explain the chain of accelerators next.

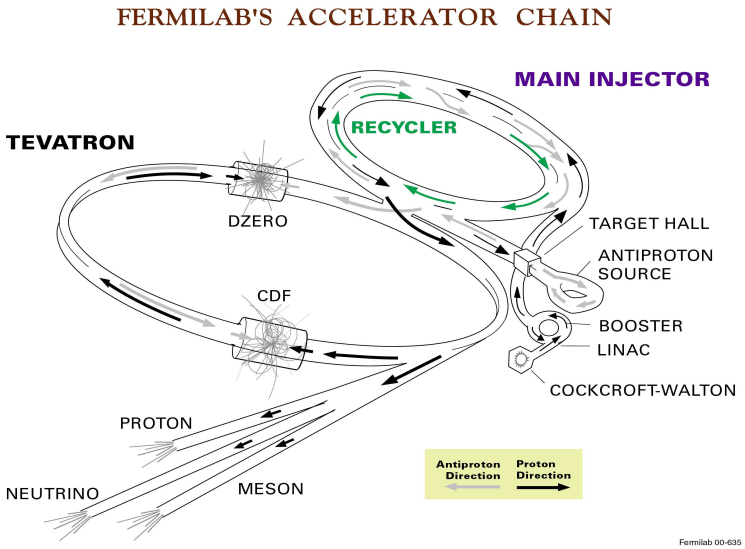

Figure 2.1: Schematic of the Tevatron accelerator complex.

2.1.1 Proton Production

Proton production begins with a Cockcroft-Walton electostatic pre-accelerator. Hydrogen gas is ionized to produce negatively charged H^{-1} ions with a kinetic energy of 750 keV. The ions move along a transport line and through an electostatic chopper. The chopper selects a small portion of the beam and allows it to continue into the Linac, a 150 m long linear accelerator. The Linac's drift tubes accelerate the beam to 400 MeV. Its series of radio-frequency (RF) stations separate the proton beam into longitudinal bunches. Then, both electrons are stripped from the H^{-1} ions with a carbon foil.

The resulting protons are injected into a 75 m radius circular synchrotron called the Booster. The Booster further accelerates the protons, and increases the intensity of the beam by injecting more protons into existing bunches. Its RF cavities do not have the same frequency as those in the Linac, and the proton bunches must come into phase with the new frequency. The proton beam exits the Booster with an energy of 8 GeV, separated into 84 bunches.

Protons are then transferred to another circular synchrotron, the Main Injector. The Main Injector has a much larger radius than the Booster (0.5 km radius). The beam energy is increased from 8 GeV to 150 GeV. The bunch structure is also collapsed from 84 to 36 bunches. The beam is then injected into the Tevatron, where it is accelerated

to the final energy of 980 GeV.

2.1.2 Antiproton Production

Antiprotons are produced by taking 150 GeV protons from the Main Injector and colliding them against a nickel target. Before colliding with the target, though, the proton beam is taken through the so-called "flattop" stage, during which the bunches are rotated 90° in longitudinal phase space. This process is performed so that the protons all collide with the target at approximately the same time. After flattop, the beam is focused by a quadrupole magnet, and collided against the nickel target. A shower of particles including a few protons and antiprotons is produced by the protons' interaction with the nickel nuclei. The antiprotons are focused with a lithium lens, and sent through a dipole magnet that selects antiprotons with an energy of around 8 GeV. The process is highly inefficient, with 20 antiprotons produced per million incident protons.

The surviving antiproton beam is transferred to the Debuncher, a triangular synchrotron. The beam must be consolidated both longitudinally and transversely to be prepared for efficient transfer through the next steps. This is accomplished through two types of cooling: stochastic cooling (transverse) and momentum (longitudinal) cooling, which employ electrodes and magnets to reduce the size and momentum spread of the bunches.

After cooling in the Debuncher, the beam is injected into another triangular synchrotron, the Accumulator. The antiproton beam undergoes further stochastic cooling, and is stored until an adequate number of antiprotons beam have been accumulated. This process is called "stacking", and can take around one day. The stacking rate and hence limited antiproton supply is a determines the Tevatron's maximum possible rate of $p\bar{p}$ collisions.

After the antiproton stack in the accumulator is deemed large enough for collisions (about 20 billion antiprotons per bunch), the antiproton beam is transferred to the Main Injector and accelerated to 150 GeV. At this point, the antiproton beam has the same bunch structure as the proton beam.

The Main Injector has an additional beam pipe built in its tunnel, the Recycler Ring. The recycler ring is used to store antiprotons during stacking.

2.1.3 Injection and Collision

After acceleration in the Main Injector, the antiproton beam is injected into the Tevatron in 36 bunches, and accelerated to 980 GeV. The protons are already present, circling the ring in the opposite direction to the anti-protons. Both protons and antiprotons circulate in 3 groups (*trains*) of 12 bunches. The gap between trains is large enough at 2.6μs that antiprotons can be injected into the Tevatron while protons are circulating, without affecting the proton orbit.

The same beam pipe and magnet system can be used for both beams, because the p's and $\bar{p}$'s are oppositely changed. The beam trajectories are helical, and controlled such that the proton and antiproton beams are separated everywhere but at the collision points at CDF and DØ. The basic circular trajectory of the beams is produced with a uniform magnetic field. The separation between the proton and antiproton beams and helical trajectory is accomplished with electrostatic separators. The beams are focused by a series of quadrupole magnets to ensure an intense beam at the collision points, maximizing the chance of proton-antiproton collisions. The beam spot at the interaction point is roughly circular and has a diameter of 30 μm.

The rate of proton-antiproton collisions is proportional to the instantaneous luminosity, which for a head-on $p\bar{p}$ collision is given as:

$$L = \frac{fBN_pN_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}^2})}F(\sigma_l/\beta^*) \tag{2.1}$$

where f is the bunch revolution frequency, B is the number of bunches per beam, N_p or $N_{\bar{p}}$ are the number of protons or antiprotons per bunch, σ_p^2 ($\sigma_{\bar{p}}^2$) is the rms of the p ($\bar{p}$) beam size at the interaction point, and F is a form factor that accounts for the effect of the bunch shape. F is dependent on the bunch length σ_l and the beta function β^*, which measures the beam width.

The instantaneous luminosity turns into the integrated luminosity by $\mathcal{L} = \int L dt$, where t is the length of time for which the Tevatron delivers collisions. $\mathcal{L}$ is the more important quantity for a physics analysis, as it determines the number of occurrences N of a particular physics process when combined with the cross section σ for the process:

$$N = \mathcal{L} \times \sigma. \tag{2.2}$$

The cross section σ for a particular physics process is dependent on the available center of mass energy $\sqrt{s}$, among other quantities. Thus, when building or upgrading an accelerator, the goal is to maximize the design $\sqrt{s}$ and L. Once an accelerator is built and in operation, the goal is to run the accelerator for as long as possible, maximizing the integrated luminosity. The Tevatron has performed remarkably well over the past few years. The initial instantaneous luminosity versus time is shown in Fig. 2.2, and integrated luminosity versus store number delivered by the Tevatron and recorded by CDF is displayed in Fig. 2.3.

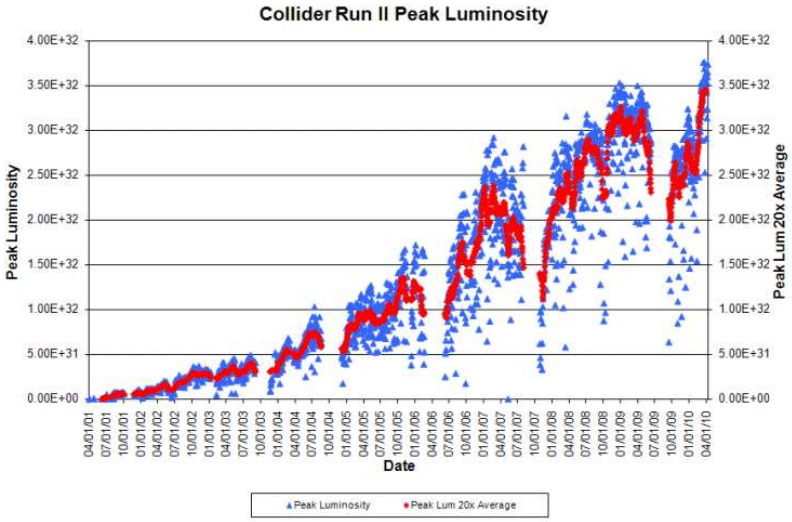

Figure 2.2: Tevatron instantaneous luminosity versus time.

2.2 CDF Detector

The data used in this analysis were collected by the CDF detector between February 2004 and June 2009. The CDF Detector is a multi-purpose detector that collects information relevant to a diverse array of physics analyses about the products of $p\bar{p}$ collisions. It is a cylindrical detector consisting of concentric layers of detector subsystems for measuring various properties of collision products. A schematic view of the CDF detector showing the different detector subsystems is given in Fig. 2.4 (cut-away view) and Fig. 2.5 (elevation view).

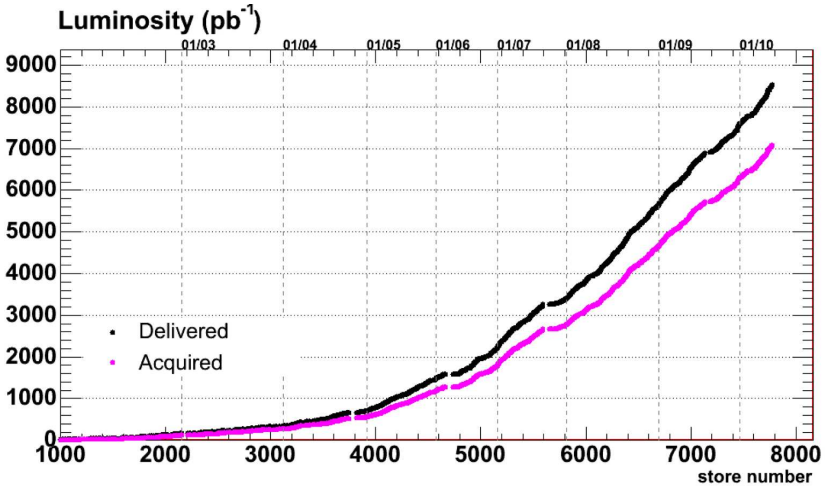

Figure 2.3: Tevatron delivered luminosity, and CDF acquired luminosity.

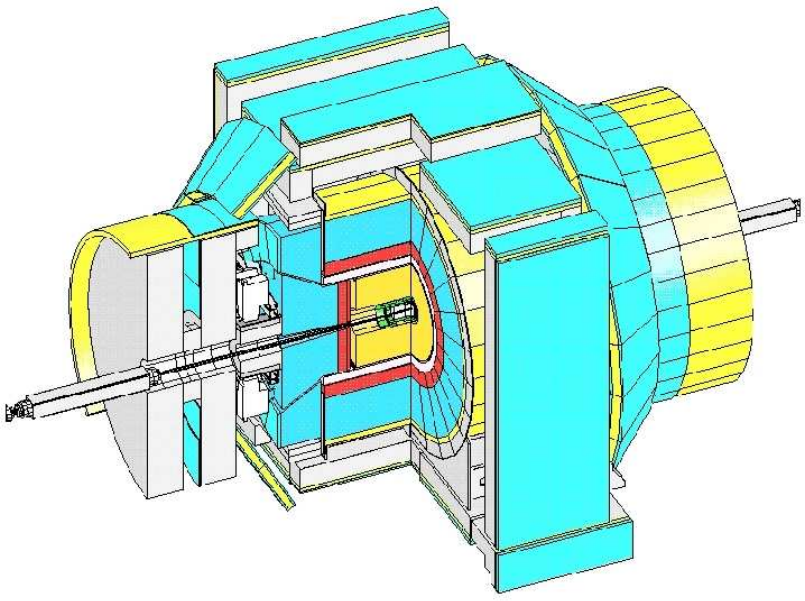

Figure 2.4: Cut-away view of the CDF detector.

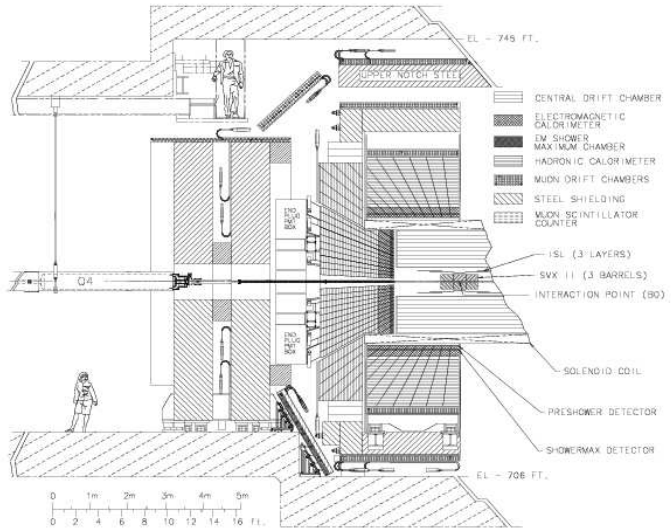

Figure 2.5: Elevation view of the CDF detector.

The CDF technical design report [34] contains a full and detailed description of the CDF detector. In the next sections, we will briefly describe the lay-out and function of the detector, and give a more detailed description of the subsystems necessary to this analysis.

A collision at the interaction point produces a spray of different particles, stable and unstable. The stable particles are electrons, protons, and neutron. They transverse all layers of the detector without decaying, but unstable particles, depending on their type, may decay in different parts of the detector, though most will decay within the beampipe. Each type of particle has characteristic behavior in the detector.

Two tracking systems are located in the innermost layer of the detector, closest to the $p\bar{p}$ interaction point. They record the trajectory of charged particles passing through them, referred to as *tracks*. The tracking system closest to the beam pipe is a silicon microstrip detector. It is surrounded by a wire drift chamber, which is in term surrounded by a solenoid coil with a 1.4 T magnetic field. Outside the solenoid is the time of flight (TOF) system, used for identifying the particle type of a track. The tracking systems facilitate the reconstruction of particle trajectories, based on the particle's interactions with the silicon strips and the gas of the drift chamber. Additionally, they are used for vertexing: tracing multiple tracks back to a common origin, either to the primary vertex at the interaction point (characteristic of unstable

particles that decay immediately to more stable daughter particles), or to a displaced or secondary vertex (characteristic of unstable particles that travel some distance before decaying).

Outside the tracking systems and the solenoid are the electromagnetic and hadronic calorimeters. Photons and electrons deposit energy in the electromagnetic calorimeters, pions, protons and neutrons deposit energy in the hadronic calorimeter. At the outer radius of the detector is the muon system, a combination of drift chambers, scintillators, and steel shielding. Muons are minimum ionizing particles and do not interact readily with the inner detector components. The behavior of different particles as they pass through the detector is illustrated in Fig. 2.6.

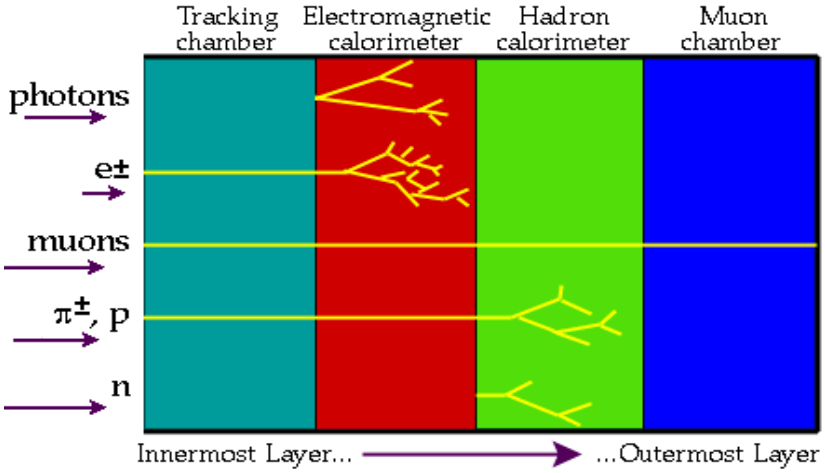

Figure 2.6: Schematic of how different particles pass through the subsystems of the CDF detector.

CDF uses coordinate systems with the following conventions. The origin is at CDF's center, roughly the interaction point:

- +x: away from the center of the Tevatron main ring
- +y: up
- +z: proton direction along the beamline
- $r = \sqrt{x^2 + y^2 + z^2}$
- $\phi = \tan^{-1}(y/x)$

- $\theta = \cos^{-1}(z/r)$

The coordinate r ($\rho = x^2+y^2$ in cylindrical coordinates) describes the radial distance from the origin. The angle θ is the polar angle from the +z axis and ϕ is the azimuthal angle from the +x axis. Note that the (ρ, ϕ) plane is the same as the (x, y) plane. We denote the direction parallel to the +z axis as the *longitudinal* direction, and the direction perpendicular to the +z axis as the *transverse* direction. For example, transverse energy and momentum are often relevant quantities for an analysis: $E_T = E\sin\theta$ and $p_T = p\sin\theta$. Another useful quantity is the pseudorapidity

$$\eta = -\log\tan\left(\frac{\theta}{2}\right), \tag{2.3}$$

which is the ultra-relativistic limit of the rapidity

$$y = \frac{1}{2}\log\frac{E+p_z}{E-p_z}. \tag{2.4}$$

A particle direction is usually described in terms of η and ϕ, which is convenient because η is invariant under Lorentz boost in the z direction. The pseudorapidity η is also used to describe the detector coverage, because distributions of particles tend to be flat in η. We refer to high η regions of the detectors as the *forward* region ($\eta >\sim 1$), and low η regions as the *central* region ($\eta <\sim 1$).

2.2.1 Tracking

The tracking subsystems, consisting of the silicon vertex detectors and the central outer tracker (COT), are critical to this particular analysis. These detectors record the ionization energy produced by charged particles interacting with the detector material. The clusters of ionization energy are called *hits*, and a series of hits in concentric layers of the detector can be reconstructed to form a *track*. The tracking detectors are situated in a uniform magnetic field, thus the radius of curvature of the charged particle track can be used to determine the particle's momentum. A cut-away view of the CDF tracking system is show in Fig. 2.7.

Silicon Vertex Detector

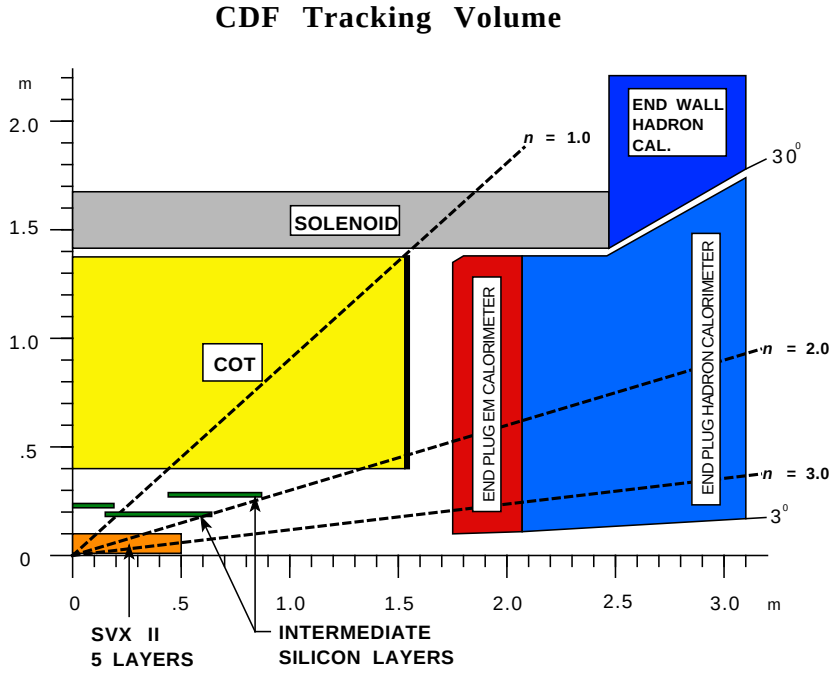

Figure 2.7: Schematic of a quandrant of the CDF detector, showing the layout of tracking subsystems.

The silicon vertex detector [35, 36] is made up of layers of silicon strip sensors. The sensors act as a p-n junction, and ionization from a charged particle will leave a cluster of electron-hole pairs. The sensors can be single sided, segmented into strips that lie along the z direction (axial). In this case, only $r-\phi$ information about the charged particle track is available. Or, the sensors can be double sided, with the segments on the back side tilted at an angle relative to the z direction (stereo). The tilt allows double sided sensors to provide z information in addition to the $r-\phi$ information provided by the strips on the front side. A charged particle will usually produce electron-hole pairs in a number of strips. During reconstruction, the magnitude of the charge deposition in adjacent strips is examined to determine precisely the location through which the charged particle passed.

The CDF silicon detector has three subcomponents: L00, SVX, and ISL. An end-view of the silicon system is shown in Fig. 2.8.

The L00 detector is a single layer of one-sided strips, set directly against the beam pipe, $\sim$1.7 cm from the beam axis. The main part of the silicon detector, the SVXII, has five concentric cylinders of silicon sensors, located between 2.5 cm and 10.6 cm from

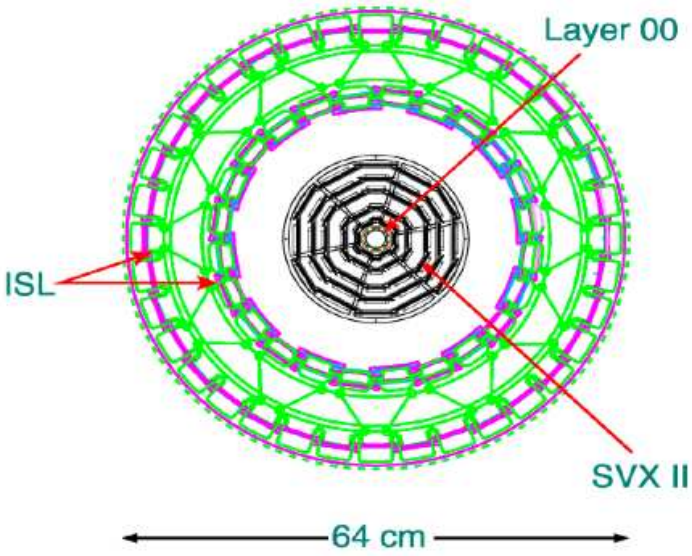

Figure 2.8: Cross sectional view of CDF silicon vertex detector.

the beam axis. Each layer of the cylinder is made of 12 *ladders* of sensors (one ladder contains 4 sensors), forming a ring in ϕ. The sensors are double sided, and stereo layers make an angle of either 1.2° or 90° relative to the z direction. All the cylinders are three ladders long in the z direction. The three barrels are stacked end to end, bringing the total length of the SVX to 87 cm. The Intermediate Silicon Layers (ISL) are three doubled sided layers, located between the SVX and the COT, at ρ=20, 22, and 28 cm. They provide coverage in both the central region and the critical forward region.

Central Outer Tracker

The Central Outer Tracker (COT) [37] is a multi-wire, open-cell drift chamber, with coverage for tracks with $|\eta|$ less than 1. The volume of the COT is segmented into 8 concentric superlayers, made up of open cells in ϕ, as shown in Fig. 2.9.

The lay-out of the open cells is shown in Fig. 2.10. Each cell has 12 gold-tungsten sense wires and 17 potential wires that shape the electric field. The cells are separated from each other with grounded, gold-Mylar plates. Wires are aligned to be either axial or stereo (with stereo angles of ±3°), in order to provide r, ϕ and z information.

The volume of the COT is filled with an 50/50 argon-ethane mixture, which is ionized as a charged particle passes through. The resulting shower of electrons drifts to the sense wires, where the magnitude of deposited charge is recorded. The maximum

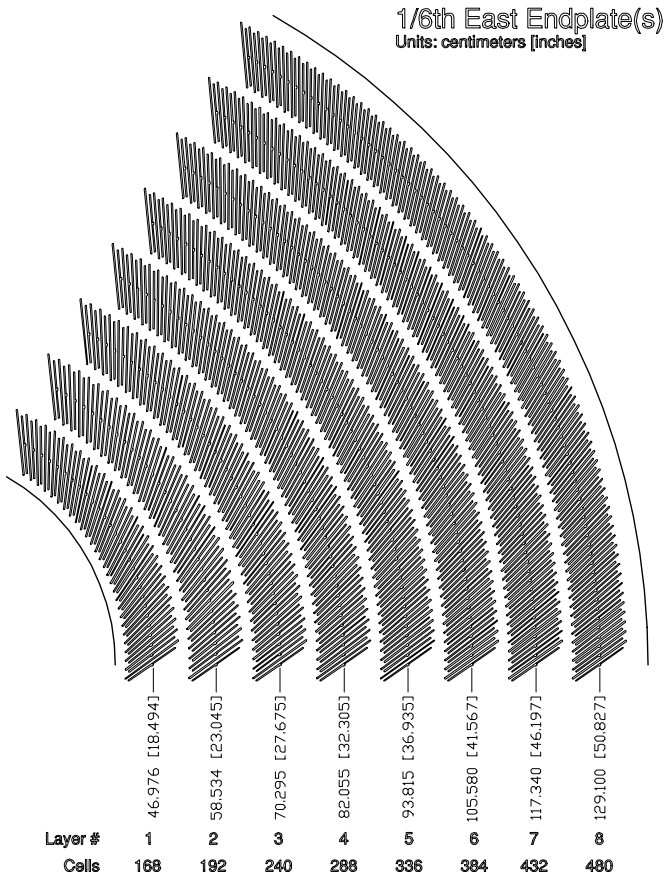

Figure 2.9: Partial cross section of the CDF Central Outer Tracker.

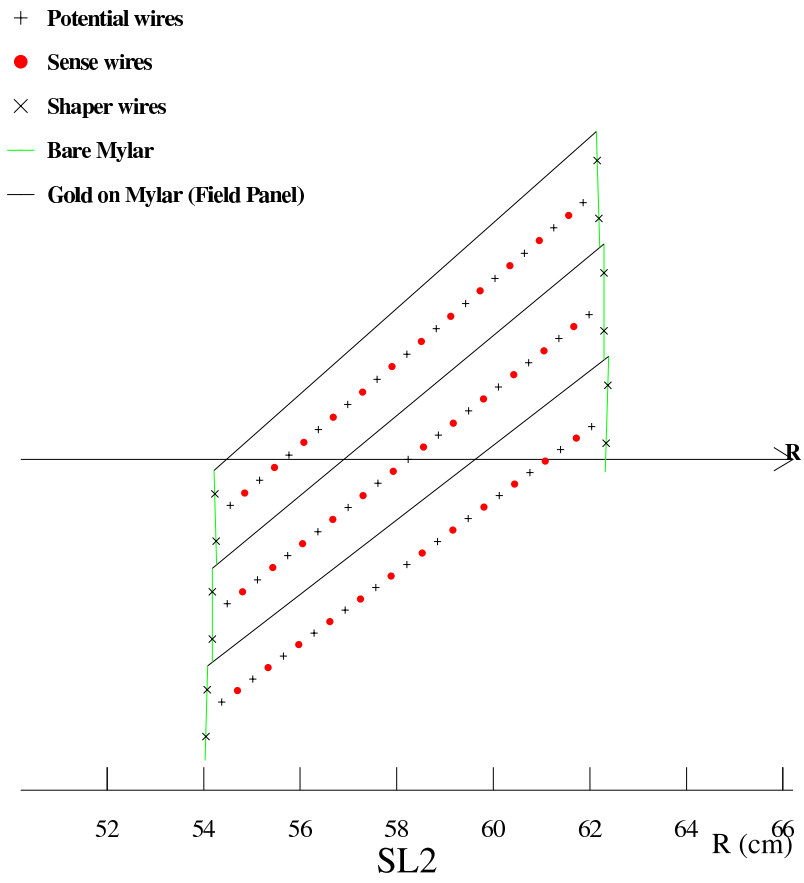

Figure 2.10: Cell layout for superlayer 2.

drift time is 100 ns, which is small compared to the bunch crossing rate.

As stated earlier, the tracking systems are situated inside a uniform magnetic field, thus the trajectory of a charged particle will be helical. Track reconstruction involves several layers of pattern recognition to construct a full helical track from disjoint hits in the silicon detector and COT. In the first step, only COT hits are used. Line segments are reconstructed from hits within a single superlayer. Then, a pattern recognition algorithm searches for line segments that are tangent to the same circle (the circle being the helical trajectory's projection in the r, θ plane). This association of line segments is done using on hits in the axial superlayers. In order to construct the 3D helical path, adding the z dimension, stereo information about the line segments is added, and a helical fit is performed.

The helix is described by five parameters [38]: C, θ, ϕ_0, d_0 and z_0. The latter three are associated with the point of closest approach of the helix to the z axis: z_0 and ϕ_0 are simply the z and ϕ coordinates of that point, and d_0, the impact parameter, is the transverse distance to the CDF origin. The former two parameters are C, the curvature of the helix, and θ, the direction of the momentum vector at the point of closest of approach to the z axis, projected into the $r - z$ plane.

After the COT information has been used to reconstruct the helical trajectory of the charged particle in the COT, silicon hit information is added. Hits in the outermost silicon layer consistent with the helical track in the COT are added. If no consistent hits are found, the search is performed in the next layer. When a consistent hit is found, the track is refit to include the new hit, decreasing the errors on the five track parameters described above. The process is repeated until all of the silicon layers have been searched for hits consistent with the increasingly well-defined track.

The momentum resolution of the CDF detector is excellent. If both the silicon detectors and COT are used, the resolution can be as precise as σ_{p_T}/p_T 0.07%/GeV/c, where σ_{p_T} is the error on the p_T measurement.

2.2.2 Time of flight

The time of flight (TOF) system [39] tracks particle flight time from the interaction point to through the tracking system. The system is located between the COT and the solenoid, and is a cylindrical layer of scintillators with coverage $|\eta|$ <1. The arrival

time of particles is measured by a light pulse in the scintillator, and combined with knowledge of the collision time, the time of of flight can be determined.

2.2.3 Calorimeters

After passing through the COT, a charged particle will enter the electromagnetic calorimeter, and possibly the hadronic calorimeter, depending on the particle's type. The purpose of the calorimeters is to capture and measure all of the passing particle's energy. The calorimeters are composed of alternating layers of absorber and scintillator material. If the particle is an electron or photon, it will lose most of its energy in the electromagnetic calorimeter through bremsstrahlung. The characteristic behavior of an electron or a photon radiating in the electromagnetic calorimeter is the production of a cascade of photons starting immediately upon entering the calorimeter.

Other charged particles are too massive to emit bremsstrahlung radiation, and will pass to the hadronic calorimeters before losing their energy. They will interact with the nuclei of the absorbing materials via the strong interaction and produce a shower of daughter particles. This behavior results in a different shower shape than showers in the electromagnetic calorimeter, as the chances of immediately interacting with a nucleus upon entering the hadronic calorimeter is small. Hadronic showers will begin after the particle has traveled some distance through the calorimeter material. Minimum ionizing particles, such as muons, will not lose much energy in the electromagnetic or hadronic calorimeters, and will pass to the muon chambers, which will be discussed in Section 2.2.4.

The electromagnetic and hadronic calorimeters [40] are made up of several sections, the plug hadronic and electromagnetic calorimeters (PHA and PEM), central hadronic and electromagnetic calorimeters (CHA and CEM) and wall hadronic calorimeter (WHA). The different sections afford broad coverage, $0 < |\eta| < 3.6$, as shown in Fig. 2.11. The PHA and PEM cover the forward regions, and CHA and CEM the central region, and WHA the intermediate region between the plug and central calorimeters.

The absorbing materials differ for the different calorimeter types: iron for the electromagnetic calorimeters, iron for the PHA, and steel for the CHA and WHA. Plastic scintillators are used in all calorimeters. The electromagnetic calorimeters are ∼19-23 radiation lengths thick, and the hadronic calorimeters are ∼4.5-7 interaction lengths

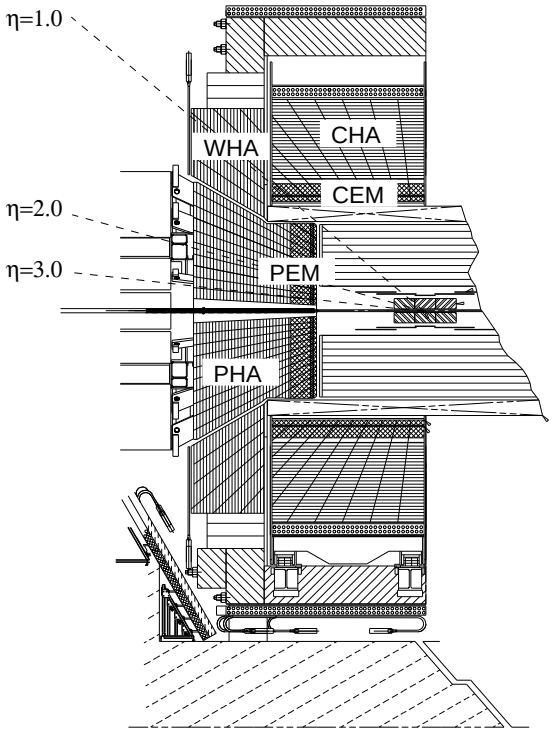

Figure 2.11: Side view of CDF calorimeter system.

thick, depending on whether the particle passes through the central or the plug region. Signals are read out by photomultiplier tubes that pick up light from the scintillators.

The calorimeters are segmented into towers in ϕ, with segments of 15° for the central region and the four towers in the most forward region ($|\eta| > 2.1$), and segments of 7.5° for the other towers in the forward region. As shown in Fig. 2.11, the towers are projective and point towards the interaction point.

Additional information is provided by the shower maximum detector [41]. The "shower max" detector is composed of scintillators and wires, and is located inside the CEM in the area where bremsstrahlung showers reach their maximum intensity. It provides information about the coordinates of the shower that can be useful in separating the signals of photons, electrons, and pions. The shower position information can be matched to tracks in the COT and silicon trackers.

2.2.4 Muon chambers

The muon chambers [42] are located outside of the calorimeters, far from the interaction point. Muons are too heavy to emit significant bremsstrahlung radiation in the electron magnetic calorimeters, and do not interact with nuclei in the hadronic calorimeters, and will thus pass through to the muon chambers. Layers of steel shielding are placed in front of the muon chambers in order to exclude absorb other particles and absorb part of the muons' energy. In general, only muons will reach the muon chambers, and all other particles (excluding neutrinos) will deposit all their energy in either the electromagnetic or hadronic calorimeter. Occasionally, though, an energetic hadronic particle will decay towards the back of the hadronic calorimeter, and its decay products will "punch through" to the muon chambers. These particles are called "fakes" or "punch throughs".

The muon system has several subsystems, the central muon detector (CMU), the central muon upgrade detector (CMP), the central muon extension detector (CMX), and the intermediate muon upgrade (IMU). In this analysis, the IMU is not used. The coverage of the muon systems in $\eta\phi$ space is shown in Fig. 2.12. The muon subsystems are composed of a combination of drift chambers, scintillators, and shielding.

CMU

The CMU detector is located outside of the CHA, with a layer of steel shielding between the two to help decrease punch through. The CMU provides coverage for

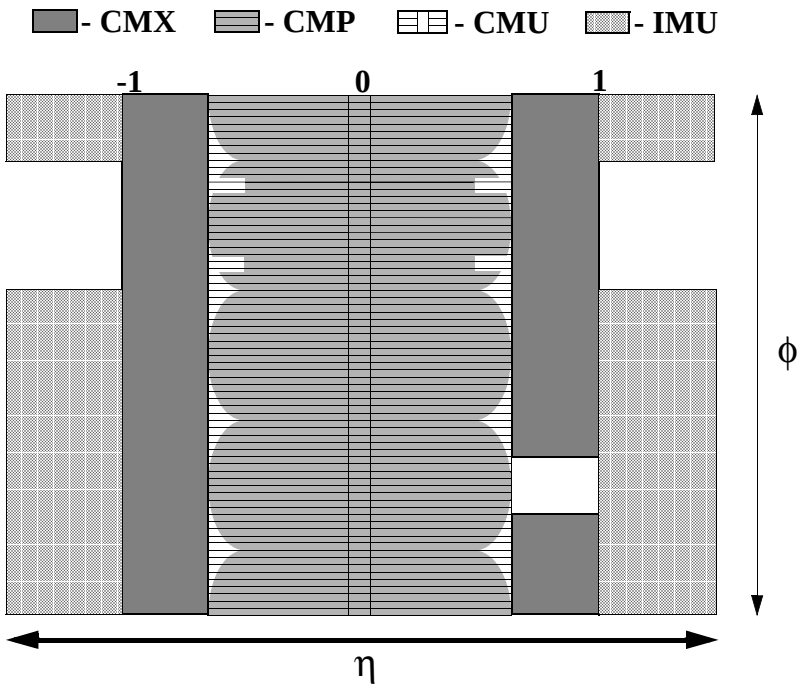

Figure 2.12: $\eta\phi$ coverage of the CDF muon systems.

$|\eta|$ <0.6. It is made up of 2880 axially positioned single wire proportional chambers. It is segmented into 15° wedges, but gaps between the wedges decrease the total ϕ acceptance to 84% The wedges are segmented into three smaller wedges containing an array of drift cells which are 4 cells wide and 4 cells deep. A muon must have p_T >1.4 GeV/c to reach the CMU.

CMP

The CMP is located beyond the CMU, and is separated from in by 60 cm of steel. It has the same $|\eta|$ coverage as the CMU, but it is built around the return yoke of the solenoid, which intrudes on the ϕ coverage in places. It is composed of 1076 axially positioned single wire proportional chambers. A muon must have p_T >2.2 GeV to reach the CMP. Muons that leave a signal in both the CMU and CMP are called CMUP muons. Such a sample has a very low fake rate, due to the extra shielding of the systems.

CMX

The CMX offers coverage in the more central region, 0.6< $|\eta|$ <1.0. It is composed

of 2208 drift tubes, arranged into conical sections at either end of the detector. The sections of the CMX are 8 layers deep, and are tilted slightly with respect to the beamline. A muon must have p_T >1.4 GeV to reach the CMX.

Scintillating tiles are present on the outside surfaces of the CMP and CMX. They are used to compensate for the slow drift time of the muon drift tubes, which is greater than the bunch crossing rate. A signal in the scintillators allows the muon to be matched to the bunch crossing which produced it.

Muons are reconstructed by first searching for a line segment produced by hits in several layers of the drift chamber. This line segment is called a *stub*, and must contain hits from at least three layers. The stub is then matched to a track in the COT. A minimum momentum is required on the track, depending on the type of muon stub to which it is matched. Restrictions are placed on the matching in the $r\phi$ plane between the muon stub and the COT track to ensure that the matching between stub and COT track is correct. An upper limit is set on the difference between the extrapolated track location and the stub location, again based on the type of muon stub.

2.2.5 Particle identification

The ability to distinguish final state decay products from each other is important for any analysis with a fully reconstructed, hadronic final state. In this analysis, it is critical to distinguish kaons from charged pions. This goal is aided by the use of particle identification, using the specific ionization of particle species in the COT (dE/dx) and the mass-dependent time of flight of particles through the COT (TOF).

A charged particle traveling through the COT deposits energy via ionization. The amount of ionization energy deposited is dependent on the particle velocity, in accordance with the Bethe-Bloch equation. Therefore, dE/dx information can be used to determine a particle's velocity and, when combined with a measurement of the particle's momentum, to determine the particle's mass.

Low momentum tracks are identified using information from the time of flight system. When time of flight is combined with momentum measurement, and the length of the track L, the particle's mass can be calculated using

$$m = \frac{p}{c}\sqrt{\frac{c^2t^2}{L^2} - 1}. \tag{2.5}$$

The two methods of particle identification are complementary, in that the separation power between charged pions and kaons is greatest for for particle momentum less than 1.6 GeV/c, while dE/dx continues to provide good separation above 1.6 GeV/c. Using both TOF and *dE/dx* information provides particle separation of 1-2σ across the range of possible particle momenta. It should be noted that the current calibration of *dE/dx* does not use tracks with a transverse momentum under 2.0 GeV/c, somewhat limiting the particle identification power.

2.2.6 Trigger

Collisions in the CDF detector occur at a rate of 2.5 million interactions per second, far too high a rate for every event to be recorded, not to mention the unreasonable amount of storage space required for recording every event. Thus, it is critical to make real-time decisions during data-taking of which events to store and which to discard. This multi-tiered selection process is called *triggering*. It is of paramount importance that the trigger does not simply discard events randomly, but retains events that contain interesting physics.

The trigger must reduce the rate from 2.5 MHz (rate of collisions) to ∼100 Hz (rate at which CDF is capable of writing out data). This reduction is achieved through several levels of triggering, including hardware and software triggers which utilize increasingly detailed information about different parts of the detector. There are three levels of triggering. An event must satisfy requirements at Level 1, 2 and 3 to be kept permanently. The particular set of selection criteria is called a *trigger path*. There are ∼140 different trigger paths at CDF, designed to select different types of physics events.

Level 1

Level 1 is a hardware trigger that takes ∼6 μs to accept or reject an event, equivalent to a 25 kHz readout rate. It uses information from the COT, calorimeters and muon chambers. It only uses specific information from these sub-systems: total energy in the calorimeter, hits in the axial superlayers of the COT, and muon stubs in the CMU, CMP and CMX. This information is used to reconstruct *trigger primitives*: muon and electron objects, and tracks.

Primitives are reconstructed using the Extremely Fast Tracker (XFT) [43]. The XFT classifies hits as prompt or delayed, depending on whether they have a short or

long drift time. After classifying hits, the XFT reconstructs track segments, and links them together to make tracks in the COT. The XFT tracks can then be matched to muon stubs and calorimeter towers by the Extrapolator unit (XRTP) to make muon, photon, electron and jet objects.

Level 2

The Level 2 trigger processes events selected by the Level 1 trigger. It has two levels, one hardware, one software, and a 400 Hz readout rate. It uses more detector information than the Level 1 trigger, including shower max information and axial silicon hits, as well as more precise information from the detector subsystems used at Level 1. A schematic of the detector subsystems used by Level 1 and 2 is given in Fig. 2.13.

A component of the Level 2 trigger is the Silicon Vertex Trigger (SVT) [44] which extrapolates XFT tracks into the silicon detector. By utilizing silicon hits and the location of the interaction point, the SVT is able to measure the track impact parameter d_0. Mesons containing b quarks fly $\sim$1 mm in the laboratory frame before decaying, therefore have large impact parameters. The ability of the SVT to measure and trigger on events with large impact parameters is critical to selecting a sample rich in heavy flavor decays. The SVT's impressive d_0 resolution, 47 μm for tracks with p_T >2 GeV/c, aids in this goal.

Level 3

Level 3 is a software trigger, running on CPU farm. It has a $\sim$100 Hz readout rate. Decisions at Level 3 take on order of one second. Level 3 uses fully reconstructed events. Three dimensional tracks and vertices are reconstructed. A candidate's L_{xy} and invariant mass can be calculated and triggered on.

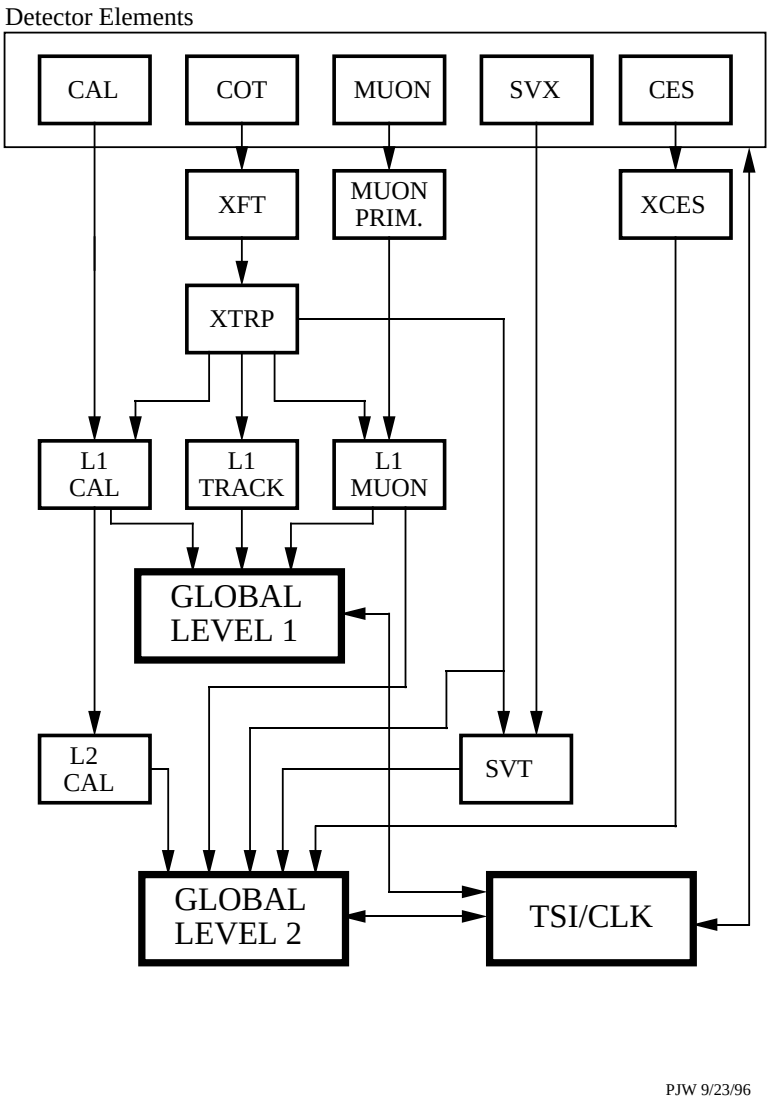

Figure 2.13: Detector components used by the CDF trigger system.

Chapter 3

Data Selection and Reconstruction

The data sample used for this study was collected with the CDF detector between February 2002 and June 2009, corresponding to an integrated luminosity of 5.2 fb^{-1}. The luminosity corresponds not to the raw CDF recorded luminosity during this time period, but to the luminosity after applying a "good run list", which discards all data taking periods during which a detector subsystem was malfunctioning or off.

This measurement uses data collected with the di-muon trigger. The two-track and lepton plus displaced track triggers were used for studies that are utilized by our measurement. All three triggers will be described in the following sections.

Two decays mode are reconstructed,

- $B_s^0 \to J/\psi(\to \mu^+\mu^-)\phi(\to K^+K^-)$, used to extract the phase β_s.
- $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$, used to calibrate the opposite-side flavor tagger, as detailed in a later section.

Charge conjugates are assumed whenever B_s^0 or B^+ are referred to, unless otherwise stated. The decay modes are fully reconstructed, meaning that all final state muon and kaon tracks have been measured to reconstruct the B meson. This is in contrast to partially reconstructed decays, such as semi-leptonic decays, where not all of the final state particles are measured, for instance, if a neutrino is present in the final state.

3.1 Trigger Requirements

The cross section for b hadron production at CDF is small compared to the inelastic $p\bar{p}$ cross section: $\sim$18 μb compared with 60 mb. The cross section for b hadron $\rightarrow$ J/ψ modes is even smaller, $\sim$0.3 μb. CDF's di-muon, semileptonic, and two-track trigger [45] are designed to use the distinctive properties of b decays to distinguish them from a seemingly overwhelming background.

The di-muon trigger [46] is used to collect the data sample for this thesis. The two muons in the final state provide an excellent trigger handle. The requirements of the di-muon trigger are as follows:

- Level 1
  - Two oppositely charged XFT tracks,
  - Two muon stubs matched to XFT tracks, either CMU-CMU or CMU-CMX,
  - Muon has p_T >1.5 GeV/c (CMU) or p_T >2.2 GeV/c (CMX),
  - Opening angle in COT superlayer 6 smaller than 135°
- Level 3
  - $2.7 < M_{\mu\mu} < 4$ GeV/c^2

The di-muon trigger data is an ideal sample to measure β_s, because the trigger does make any requirement on the impact parameter of the XFT tracks. An impact parameter cut would distort the B_s^0 proper decay time distribution and complicate the time-dependent portion of the measurement.

The two-track trigger sample was used for the measurement of $B_s^0\bar{B}_s^0$ oscillations. This measurement is a prerequisite to the measurement of β_s for two reasons. First, the oscillation frequency Δm_s is used in our likelihood, but the statistics in the $B_s^0 \rightarrow J/\psi\phi$ mode are not sufficient to determine Δm_s from our data sample. Secondly, the measurement of $B_s^0\bar{B}_s^0$ oscillations is used to calibrate the same-side kaon tagger. This will be discussed in detail in Sec. 4.3.2. The two-track trigger requires two displaced tracks ($d_0 > 100$ μm) and a displaced vertex.

The lepton plus displaced-track trigger sample was used for the development of the same and opposite side flavor tagging algorithms, and the calibration of the opposite

side flavor tagger. The lepton plus displaced-track trigger requires one displaced track ($d_0 > 120$ μm) and one lepton with $p_T > 4$ GeV/c.

3.2 Event Reconstruction

After data has been selected by triggers and written out, it is subjected to an offline production process. This includes refitting tracks, fitting vertices, reconstructing decay chains, and packaging the information necessary to allow for physics measurements in a user-friendly format.

3.2.1 Track Refitting

Tracks are refit offline with the most accurate available detector information and calibration. Energy loss corrections are applied depending on a track's particle type. If L00 hit information is available, it is added to the track refitting. Requirements on matching between muon stubs and COT tracks are made more stringent. Trigger confirmation is required for muons: the muons coming from the B decay of interest must pass the di-muon trigger requirement. In other words, the event must have been selected by the di-muon trigger based on the final state muons from a B decay, not from random muons in the event that happened to meet the trigger requirements.

3.2.2 Vertex Fitting

Vertex fitting is central to reconstructing a decay chain. Tracks observed in the detector are matched together as coming from a common origin, where the parent particle decayed. The parent particle may itself be a decay product of a heavier particle. The decay chain is traced backwards from the stable tracks as a series of intermediate decay vertices, leading to the primary vertex at the interaction point. Momenta can be assigned to the decay particles, based on the measured momenta of the final state tracks.

The decays used in this thesis were reconstructed using the CTVMFT vertex fitting program [47]. Vertices can be fit either in 3D or in 2D, providing the user with a handle on the goodness of fit for the vertices in the form of a χ^2 probability.

Knowledge of the primary vertex is critical to the measurement of the B meson lifetime. Multiple methods of determining the primary vertex exist. For this analysis,

the primary vertex location is recalculated for each event, using tracks in the event that were not related to the B meson decay of interest. This is referred to as the "event-by-event primary vertex method", and it provides an accurate assessment of the primary vertex position, with a resolution of $\sim$20 μm.

3.2.3 BStNtuple

The data used in this thesis was stored in the BSTNTUPLE framework [48]. The framework is based on the standard STNTUPLE framework [49]. It stores information about stable and decaying particle objects, together with information for flavor tagging and particle identification. It allows the user the flexibility to access different decay modes without requiring excessive CPU or storage space.

3.2.4 Preselection

In the next section, the process of separating signal from background using an artificial neural network is described. Before training the network, preselection requirements are applied to remove events from the data sample that almost certainly do not contain a B_s^0 decay. An important aspect of these preselection requirements are cuts on track quality. The number of fake or mis-measured tracks can be reduced by requiring a minimum number of hits in the COT and silicon layers.

The preselection requirements for the $B_s^0 \to J/\psi\phi$ mode are as follows:

- Track quality
  - $\geq$10 axial COT hits for each track
  - $\geq$10 stereo COT hits for each track
  - $\geq$3 axial silicon hits for each track
  - $p_T > 400$ MeV/c for each kaon track
- Vertex quality
  - $\chi^2_{r\phi}$ <50, where $\chi^2_{r\phi}$ is the χ^2 of the 2D $r\phi$ vertex fit
- Decay particle kinematic cuts
  - $5.1 < M_{J/\psi\phi} < 5.6$ GeV/c^2

- $p_T \geq 4.0$ GeV/c for B_s^0 candidate
- $1.009 < M_{KK} < 1.028$ GeV/c^2
- $p_T \geq 1$ GeV/c for ϕ candidate
- $3.04 < M_{\mu\mu} < 3.14$ GeV/c^2

3.3 MC Sample

Numerical simulation of $B_s^0 \to J/\psi\phi$ and other relevant B decays is employed at several points in this thesis, when an analytic or data-based approach is not feasible. Realistic Monte Carlo events are used for modeling detector angular analysis and for training an artificial neural network to distinguish signal from background.

A series of programs are used to trace all the steps of a $B_s^0 \to J/\psi\phi$ decay, beginning with b quark generation in a $p\bar{p}$ collision, fragmentation of the b quark to form a hadron, and the decay of the hadron in the CDF detector. We simulate the generation and fragmentation of b quarks using the BGENERATOR program [50]. BGENERATOR is based on next-to-leading-order QCD calculations and the Peterson fragmentation function [51]. The BGENERATOR package massively simplifies the collision environment, generating one b quark which fragments only to the specified B hadron. We use a phase-space model for the B_s^0 decay. The phase-space model ensures that all transversity angles are generated flat for B_s^0 decays. We examine how the initially flat distributions are sculpted by detector acceptance to determine the detector efficiency, as described Sec. 4.2.1.

The hadron's decay is simulated with the EVTGEN package [52]. The interaction of the particles in the decay chain with the CDF detector is simulated with the GEANT [53] based CDFSIM package [54]. Monte Carlo events simulating detector and data-taking conditions through about 1 fb^{-1} of data was used in this thesis. Monte Carlo events were not used for studies where a strong time dependence was expected, thus it was not necessary to simulate the conditions of the entire dataset. After passing through the detector, simulated events are then subjected to the same trigger requirements and reconstruction process as a data event, and stored in the BSTNTUPLE format.

The p_T spectrum of the B_s^0 is correlated with the final state angular distribution. Because Monte Carlo is used to model the angular acceptance of the data, we must demand good Monte Carlo/data agreement for the B_s^0 p_T distribution. The agreement

is improved by re-weighting the Monte Carlo events based on quantities that affect the B_s^0 p_T spectrum, namely track p_T distributions and trigger composition.The Monte Carlo data is split into several categories based on muon p_T and whether the event is CMU-CMU or CMU-CMX. Then the relative proportion of each of the categories is weighted to produce the best possible agreement between data and Monte Carlo.

3.4 Toy MC Pseudo-experiments

We use toy Monte Carlo pseudo-experiments, called "toys" or "pseudo-experiements" at several points in this measurement. They are extremely useful for testing the properties of the likelihood function, and the behavior of a fit. They are particularly used for studies where a full realistic Monte Carlo simulation is not necessary, and a simplified simulation of the data is all that is required to test the fit behaviors. The pseudo-experiment events are created by randomly sampling a probability density. The events are generated according to a set of inputs for each parameter in the likelihood. These inputs are defined by the user. The sample of pseudo-experiment events can then be fit in the manner as a sample of data events.

As we will see in Sec. 4.1, the likelihood contains both signal and background. Therefore, when generating pseudo-experiments, we can include background events, parameterized in the same manner as the data.

Also, simplified versions of the likelihood can be used to generate pseudo-experiments. For instance, signal-only toys may be used, or toys that do not account for detector efficiency.

3.5 Neural Network Training

The removal of events for which a random combination of tracks happens to have an invariant mass similar to the decay particle of interest is an important task. The goal is to decrease combinatoric background as much as possible without removing signal events. This is achieved by selecting events based on kinematic properties that are different for signal and background. A standard figure of merit to optimize is $S/\sqrt{S+B}$ where S is the number of signal events and B is the number of background events. One might optimize a selection by cutting on each selection variable with varying tightness,

and choosing the cut value for each selection variable that optimizes the figure of merit. This method, attractive in its transparency, does not account for correlations between selection variables. It also gives each variable full power to accept or reject an event, so an event which fails the cut for one variable will be discarded, even if it clearly passes for all other variables, indicating that it is most likely a signal event.

A more sophisticated method of distinguishing signal from background is to use an artificial neural network. Selection on all of the variables is combined into a single discriminant, such that the value of a single variable does not decide with an event is discarded. Based on the discriminant, an event is classified as signal-like or background-like on a scale of +1.0 to -1.0. Correlations between variables are taken into account by the neural network, and the weight of each variable in the overall discriminant is affected by its correlation with other variables.

We use the Neurobayes package [55]. It is a three-layer network with an input layer, a hidden layer, and an output layer. A schematic is shown in Fig. 3.1. In the input layer, there is one node for each input, in the hidden layer there are approximately as many nodes as in the input layer, and in the output layer, there are as many nodes as are demanded by the discriminant type. We require one output node.

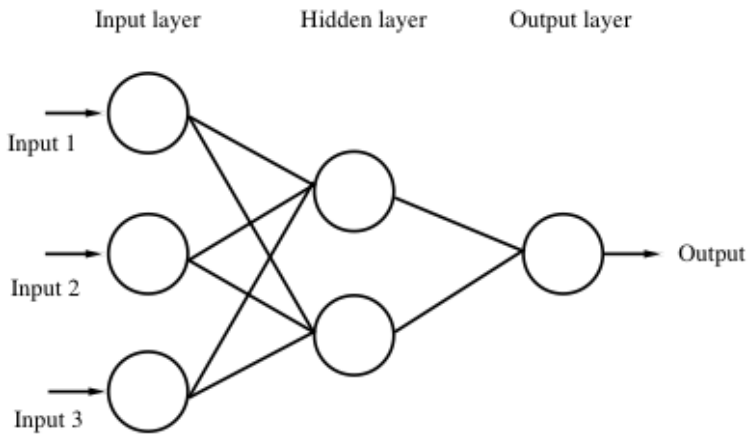

Figure 3.1: Schematic of the three layers of the Neurobayes network.

Inputs to each node are combined together in a weighted sum, and then mapped to the output interval [-1, +1] using a sigmoid function. The user defines the input variables and the topology of the network. The weights with which the inputs are

combined are determined by *training* the neural network. By training the network on a known sample, the weights between the nodes can be optimized to give a discriminant that makes the correct decision for a maximal number of events. Training is accomplished by minimizing an entropy function that compares the network output with the true result. This is an iterative process, and the user must avoid over-training. If too many iterations are made, the network begins to make selections based on statistical fluctuations in the training sample.

Our neural network training signal sample is 350,000 $B_s^0 \to J/\psi\phi$ Monte Carlo events, generated as described previously. The training background sample is 300,000 data events, taken from the $J/\psi\phi$ invariant mass sideband region. We use 3.9 fb^{-1} of data for the training of the background sample.

We use the following variables for selection, listed by rank. The better the rank (the best rank is 1), the more important the variable is to the total discriminant. The identification number is also listed for each variable:

- (1) p_T of ϕ (5)
- (2) Kaon likelihood for K_2 (11)
- (3) Kaon likelihood for K_1 (10)
- (4) Muon likelihood for μ_2 (9)
- (5) $\chi^2_{r\phi}$ for B_s^0 (2)
- (6) p_T of B_s^0 (3)
- (7) Muon likelihood for μ_1 (8)
- (8) vertex probability for B_s (4)
- (9) vertex probability for ϕ (6)
- (10) vertex probability for J/ψ (7).

The vertex probabilities are χ^2 probabilities for the 3D vertex fit. The muon and kaon likelihoods are quantities used for particle identification. The algorithm used to determine the muon likelihood is described in [56]. The kaon likelihood is a discriminant constructed from dE/dx and TOF information.

The correlations between the variables is shown in Fig. 3.2. There is no sign of large, unexplained correlations between training variables.

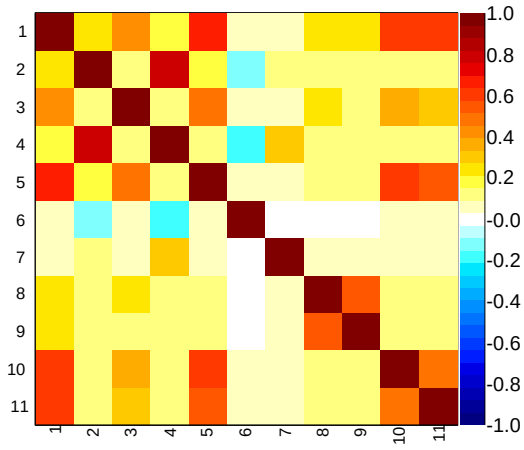

Figure 3.2: Correlations between variables used for signal selection. The eleventh row and column show truth (signal or background). The other columns and rows are are the selection variables, labeled by identification number.

The results of the neural network training are shown in Fig. 3.3. The neural network output distributions for signal (red) and background (black) of the training sample are shown in the left plot. The separation between signal (+1) and background (−1) is good. The right plot shows purity of the test sample as a function of network output cut. This is a linear relation, as it should be if the network has been trained properly: the tighter the cut, the greater the background rejection. Fig. 3.4 shows the network efficiency in classifying the signal training sample. The black curve is the actual efficiency, the red curve is the maximum possible efficiency, and the blue curve corresponds to classifying events randomly. Our network performs with nearly optimal efficiency on the training sample.

3.6 Signal Optimization

After the neural network has been trained, the next step is to choose a value of network output at which to cut. Traditionally, this is accomplished by optimizing the figure of merit $S/\sqrt{S+B}$. This figure of merit optimizes the B_s^0 signal significance. It is not

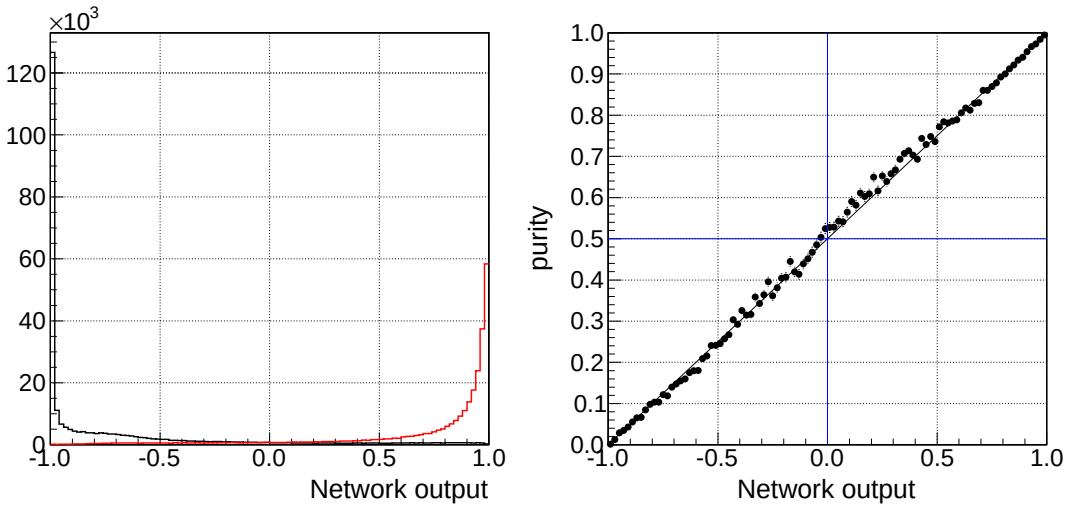

Figure 3.3: Signal (red) and background (black) distributions for training samples versus network output (left). Purity of test sample versus network output cut (right).

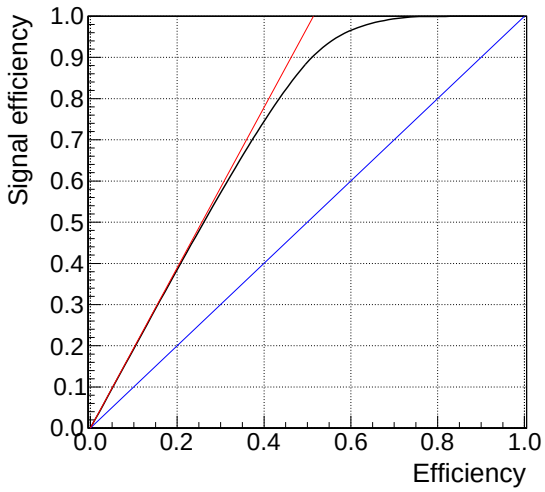

Figure 3.4: Signal efficiency versus efficiency for the training sample. The red curve is optimal performance, the black curve is actual performance, and the blue curve corresponds to random decision-making.

clear, though, that is the optimal figure of merit for this analysis. We want to optimize sensitivity to β_s, not to the B_s^0 signal. Our sensitivity to β_s has been shown on previous rounds of the analysis to be limited by statistics. As such, sensitivity to β_s may increase using a looser cut than that suggested by $S/\sqrt{S+B}$ optimization, adding more signal events, even at the expense of adding more background. We study the sensitivity to β_s as a function of neural network cut on toy Monte Carlo pseudo-experiments. This is a new feature of the analysis that was not used in previous measurements.

The effect of tightening the cut on the neural network output is simulated in pseudo-experiments generated at different values of S/N_{TOT}, where N_{TOT} is the total number of events in the $J/\psi\phi$ invariant mass window, and S is the number of B_s signal events. N_{TOT} and S are determined from mass fits to 3.9 fb^{-1} of data. Mass histograms are filled for a variety of neural network output cuts between -1.0 and 1.0. An invariant $J/\psi\phi$ mass window between 5.27 and 5.46 GeV/c^2 is used. The resulting distributions are fit with a single Gaussian for the signal and a first order polynomial for the background, and S and N_{TOT} are extracted.

The input values of all other parameters are kept the same for pseudo-experiments corresponding to each neural network cut; only N_{TOT} and S/N_{TOT} are varied. The parameter input values are taken from the results of an un-binned maximum likelihood fit on 2.8 fb^{-1} of data [57]. The only exceptions are the parameters describing the tagging power, which was compromised for the 2.8 fb^{-1} result. We generate toys with a tagging power equivalent to using uncompromised tagging in the full data set. We generate at β_s=0.5 and $\Delta\Gamma$=0.12, corresponding to fitted values from the analysis with 2.8 fb^{-1} [57].

We generate and fit 700 pseudo-experiments at each neural network output cut value, and then examine for each case the β_s error distribution. The error distributions were fit with Landau functions to extract their most probable values. An example is shown in Fig. 3.5, indicating how well the Landau function describes the error distributions.

The fitted most probable value is shown as a function of network output cut in Fig. 3.6. We verify that the width of the error distributions do not vary as a function of network output cut. Such a behavior could introduce an apparent trend in the most probable value versus network output cut that does not, in fact, exist.

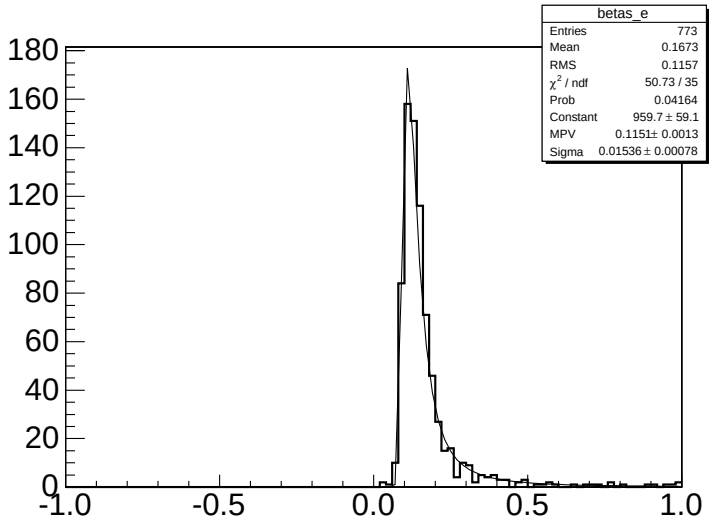

Figure 3.5: Example of a β_s error distribution fit with a Landau function to extract the most probable value.

Fig. 3.6 shows that tight cuts on the neural network output cut correspond to larger β_s errors. A looser cut, for example at a network output of 0.2, minimizes the errors on β_s, and thus maximizes our sensitivity. It should be emphasized that this technique is not intended to guarantee a particular error on β_s, but is merely used to identify the trend in β_s error size as a function of neural network output cut.

In order to verify that the trend in the sensitivity to β_s as a function of network output cut is independent of the true values of our parameters of interest, β_s and $\Delta\Gamma$, we repeat the same study at different generated values of β_s and $\Delta\Gamma$. The first alternative input we choose is the standard model point, β_s=0.02, $\Delta\Gamma$=0.1. About 800 toys were generated per cut for this case. The second alternative input was β_s=0.3, $\Delta\Gamma$=0.09. The number of toys generated per cut was ∼600 for this set of $\beta_s/\Delta\Gamma$ input values.

The most probable value is plotted versus network output cut for β_s=0.02 and $\Delta\Gamma$=0.1 and for β_s=0.3 and $\Delta\Gamma$=0.09, shown in Fig. 3.7 and Fig. 3.8. The trend in size of β_s errors is similar, regardless of the values of β_s and $\Delta\Gamma$ at which the toys are generated. β_s errors are large for tight cuts on the neural network output, but for looser cuts, the errors are smaller and appear to be insensitive to the amount of background. Based on this observation, we chose a cut at a network output of 0.2, where the β_s

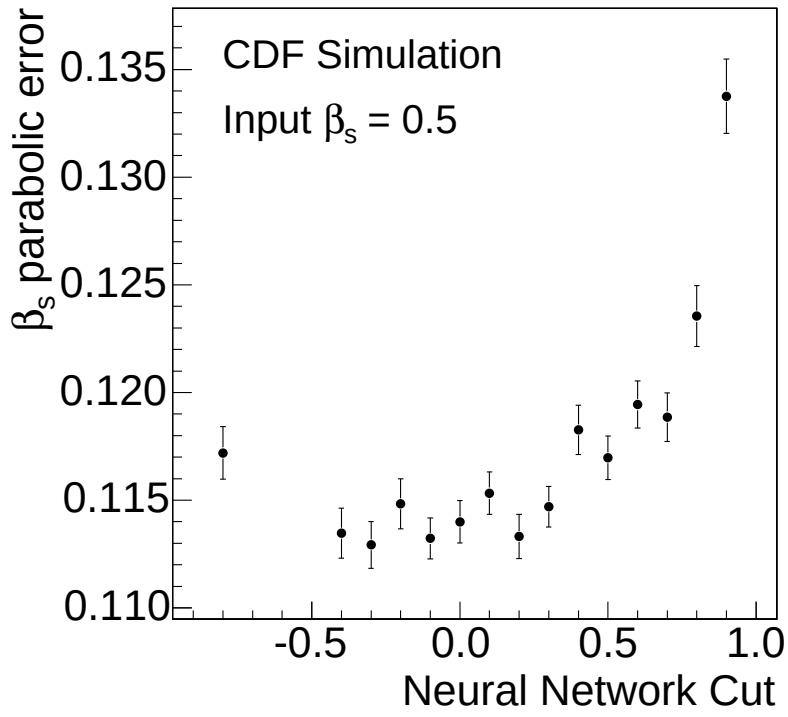

Figure 3.6: Magnitude of error on β_s as a function of neural network output cut, for β_s=0.5 and $\Delta\Gamma$=0.12.

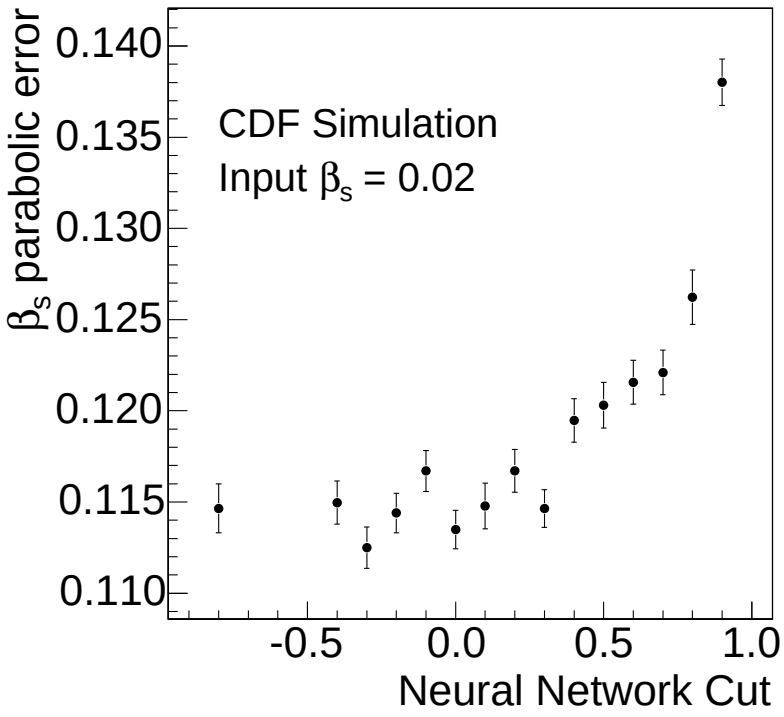

Figure 3.7: Magnitude of error on β_s as a function of neural network output cut, for β_s=0.02 and $\Delta\Gamma$=0.1.

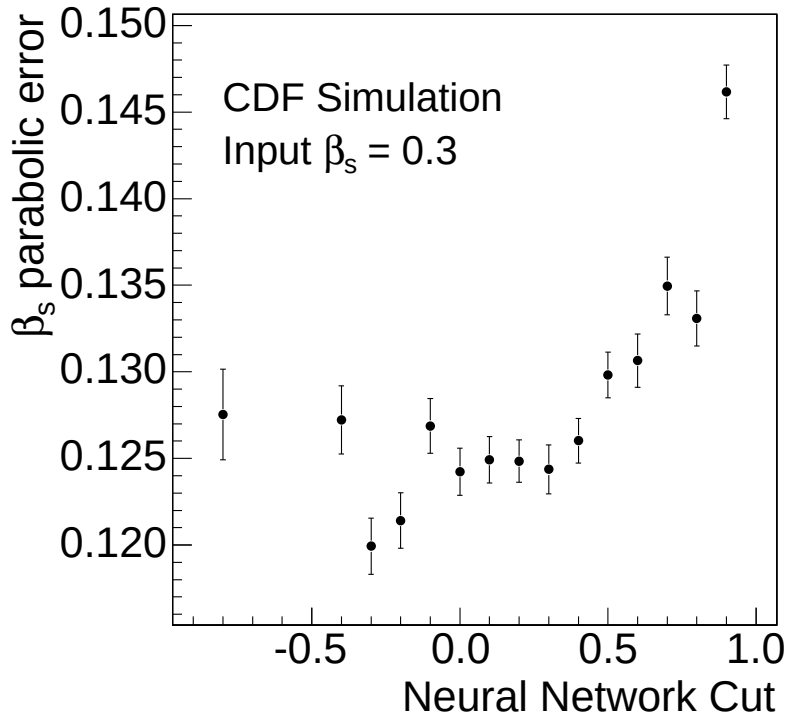

Figure 3.8: Magnitude of error on β_s as a function of neural network output cut, for β_s=0.3 and $\Delta\Gamma$=0.09.

errors are small, but we avoid adding unnecessary amounts of background. Comparing this optimization procedure with the standard optimization of $S/\sqrt{S+B}$ shown in Fig. 3.9, the cut that we choose based on studying the sensitivity to β_s is somewhat looser than the one we would choose based on optimizing $S/\sqrt{S+B}$ (0.2 versus 0.6).

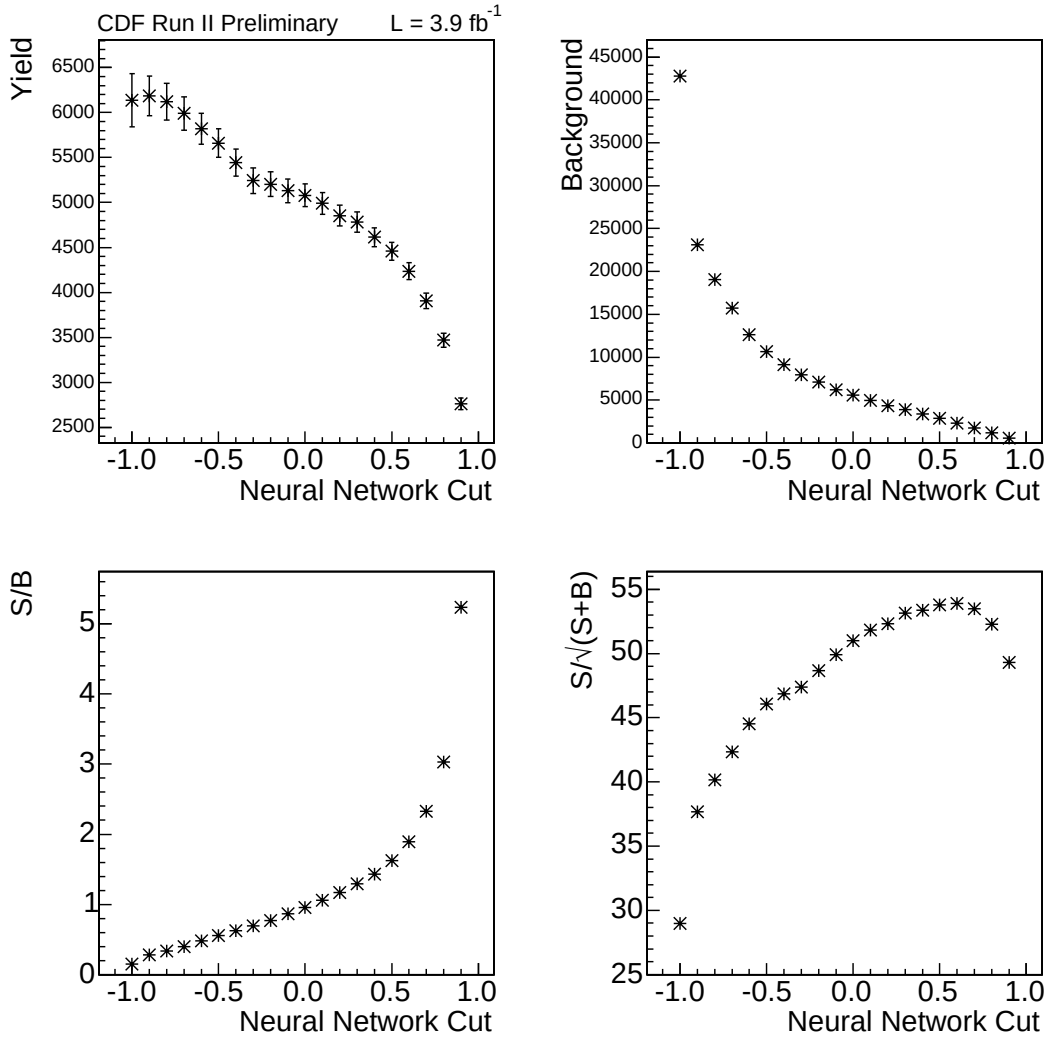

Figure 3.9: Yield, number of background events, S/B, and $S/\sqrt{S+B}$ as a function of neural network cut.

The $J/\psi\phi$ invariant mass distribution with the chosen cut applied is shown in Fig. 3.10. A cut on the neural network output of 0.2 yields 6504±85 $B_s \to J/\psi\phi$ signal events, as extracted by a fit with a single Gaussian with flat background. The signal

and sideband regions are indicated by the blue and red lines, respectively.

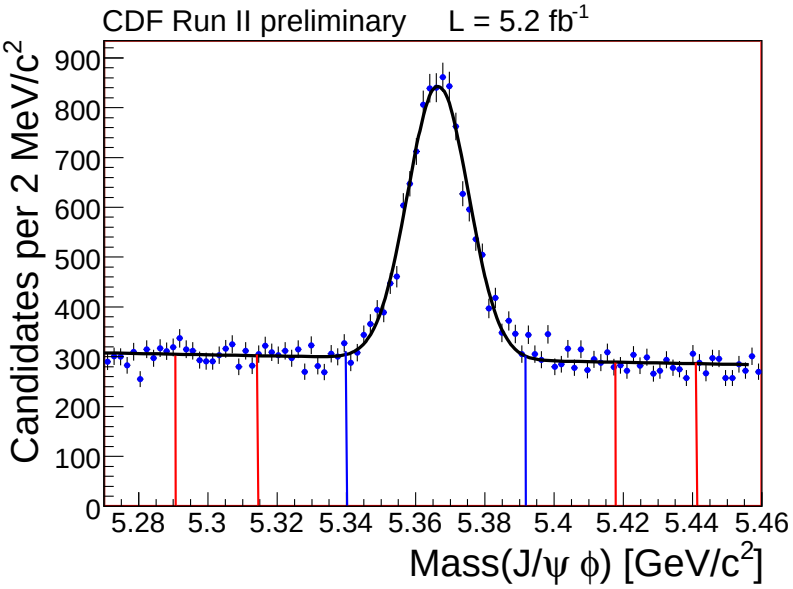

Figure 3.10: $J/\psi\phi$ invariant mass distribution with a cut of 0.2 on the neural network output. Blue and red lines denote the signal and sideband regions, respectively.

Chapter 4

Analysis Strategy

4.1 General Form of the Likelihood

The time and angular dependent probability density functions (PDFs) for B_s^0 and $\bar{B}_s^0$, presented in Chapter 1, cannot be used alone to extract β_s from the data. Detector effects, such as imperfect angular efficiency and resolution smearing must be taken into account. In addition to describing the B_s^0 and $\bar{B}_s^0$ events found in the data, we must also describe another major component of the data, the background. We construct a larger PDF for each event that includes, for instance, the candidate's mass and proper time PDFs. Often, probability densities for the same quantities are different for signal and background, so we have

$$\mathcal{P}_{tot} = f_s \mathcal{P}_s(\vec{q}) + (1 - f_s)\mathcal{P}_b(\vec{q}), \tag{4.1}$$

where $\mathcal{P}_s$ and $\mathcal{P}_b$ are the PDFs for signal and background, f_s is the signal fraction, and $\vec{q}$ is the ensemble of variables that parameterizes an event.

The value of $\mathcal{P}_{tot}$ is multiplied for all $B_s^0/\bar{B}_s^0$ candidates in the data sample to produce the likelihood function $\mathcal{L}$. By finding the value of each variable in the likelihood that maximizes $\mathcal{L}$, we maximize the related probabilities. We use the MINUIT minimization program [58] to maximize $\mathcal{L}$, or rather, to minimize -2 ln$\mathcal{L}$, MINUIT's score function. The technique used is referred to as an *unbinned maximum likelihood fit*, indicating that the PDFs are maximized event-by-event; we do not fit histograms containing the distribution of all events.

In this chapter, we will describe in detail all of the probability density functions that combine into the total PDF, and present the likelihood. This will include an explanation how detector effects are handled in the angular and lifetime parts of the measurement. Also, the flavor tagging algorithms and their calibrations are discussed.

4.2 Angular analysis

The final state distributions of transversity angles are modeled using the probability densities of Eq. 1.70 and Eq. 1.71 for signal events. The effect of the angular efficiency of the CDF detector on the angular distribution must be considered in the likelihood. The angular acceptance of the detector is not uniform, so a flat angular distribution will be sculpted by the angular acceptance, such that its measured distribution is no longer flat.

The background is modeled empirically based on events from the B_s^0 mass sidebands (shown in Fig. 3.10). Since we model real background events empirically, an efficiency correction for the background is not necessary, as the effects of the efficiency are already included in the angular distributions.

4.2.1 Normalization and angular efficiency

The probability density functions given in Chapter 1 were normalized to one. The CDF detector, however, does not have perfect efficiency as a function of angle. The PDFs must be multiplied by an angular efficiency term, $\epsilon(\psi, \theta, \phi)$. This necessitates updating the normalization to account for the new efficiency term.

Normalization in the absence of a non-resonant S-wave contribution

We first describe the normalization of the probability density functions given in Eq. 1.54 and Eq. 1.55, not accounting for contributions from f_0 or non-resonant K^+K^-. The extension of the normalization to include the additional contributions is relatively minor, and is given in Sec. 4.2.1.

First, we define the normalization. We demand that the probability density for the B_s^0 and $\bar{B}_s^0$, including detector efficiency, be normalized to one:

$$\frac{1}{N}\int\int\int\int \sum_{i=B,\bar{B}} P_i(\psi,\theta,\phi,t)\epsilon(\psi,\theta,\phi)d(\cos\psi)d(\cos\theta)d\phi dt = 1. \tag{4.2}$$

The normalization is therefore

$$N = \int\int\int\int \sum_{i=B,\bar{B}} P_i(\psi,\theta,\phi,t)\epsilon(\psi,\theta,\phi)d(\cos\psi)d(\cos\theta)d\phi dt. \tag{4.3}$$

We can integrate over time, and expand the efficiency as an (infinite) sum of spherical harmonics Y_l^m, with expansion coefficients a_l^m:

$$\epsilon(\psi,\theta,\phi) = \sum_{l,m} a_l^m(\psi)Y_l^m(\theta,\phi). \tag{4.4}$$

The angular terms in the PDFs can also be expanded as non-infinite sums of spherical harmonics, including harmonics up to $l = 2$. The property

$$\int\int Y_l^m(\theta,\phi)(-1)^{-m}Y_{l'}^{-m'}(\theta,\phi)\sin\theta d\theta d\phi = \delta_{mm'}\delta_{ll'} \tag{4.5}$$

guarantees that N will be a finite sum after integrating over the angles θ and ϕ, with only harmonics up to $l = 2$ contributing. We use the following relations

$$\begin{aligned} Y_{lm} &= Y_l^m \quad (m=0), \\ Y_{lm} &= \frac{1}{\sqrt{2}}(Y_l^m + (-1)^m Y_l^{-m}) \quad (m>0), \\ Y_{lm} &= \frac{1}{i\sqrt{2}}(Y_l^{|m|}) - (-1)^{|m|}Y_l^{-|m|}) \quad (m<0) \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} a_{lm} &= a_l^m \quad (m=0), \\ a_{lm} &= \frac{1}{\sqrt{2}}(a_l^m + (-1)^m a_l^{-m}) \quad (m>0), \\ a_{lm} &= \frac{i}{\sqrt{2}}(a_l^{|m|}) - (-1)^{|m|}a_l^{-|m|}) \quad (m<0) \end{aligned} \tag{4.7}$$

to re-express N in terms of real quantities:

$$
\begin{aligned}
N &= \frac{3}{8\sqrt{\pi}}\left[(|A_\perp|^2+|A_{||}|^2)\int a_{00}(\psi)sin^2\psi d(\cos\phi)+2|A_0|^2\int a_{00}(\psi)\cos^2\psi d(\cos\psi)\right] \\
&+ \frac{3}{8\sqrt{5\pi}}\left[(-|A_\perp|^2+\frac{1}{2}|A_{||}|^2)\int a_{20}(\psi)\sin^2\psi d(\cos\psi)+|A_0|^2\int a_{20}(\psi)\cos^2\psi d(\cos\psi)\right] \\
&- \frac{9}{16\sqrt{5\pi}}\frac{1}{\sqrt{1+\frac{4\tau_L\tau_H}{((\tau_L-\tau_H\sin 2\beta_s)^2)}}}\left[(A_{||}^*A_\perp+A_{||}A_\perp^*)\int a_{2-1}(\psi)\sin^2\psi d(\cos\psi)\right] \\
&+ \frac{9}{16}\frac{\sqrt{2}}{\sqrt{15\pi}}\frac{1}{\sqrt{1+\frac{4\tau_L\tau_H}{((\tau_L-\tau_H\sin 2\beta_s)^2)}}}\left[(A_0^*A_\perp+A_0A_\perp^*)\int a_{21}(\psi)\sin\psi\cos\psi d(\cos\psi)\right] \\
&+ \frac{9}{8\sqrt{15\pi}}\left[\frac{|A_{||}|^2}{2}\int a_{22}(\psi)\sin^2\psi d(\cos\psi)-|A_0|^2\int a_{22}(\psi)\cos^2\psi d(\cos\psi)\right] \\
&+ \frac{9}{16}\frac{\sqrt{2}}{\sqrt{15\pi}}\left[(A_0^*A_{||}+A_0A_{||}^*)\int a_{2-2}(\psi)\sin\psi\cos\psi d(\cos\psi)\right].
\end{aligned}
\tag{4.8}
$$

The coefficients $a_{lm}(\psi)$ can be expressed as a Legendre series,

$$
a_{lm}(\psi)=a_{lm}^k P_k(\cos\psi), \tag{4.9}
$$

in order to isolate their dependence on ψ for the final integration. After integration over ψ, we are left with

$$
\begin{aligned}
N &= \frac{3}{8\sqrt{\pi}}\left[\frac{4a_{00}^0}{3}(|A_0|^2+|A_{||}|^2+|A_\perp|^2)+\frac{4a_{00}^2}{15}(2|A_0|^2-|A_{||}|^2-|A_\perp|^2)\right] \\
&+ \frac{3}{8\sqrt{5\pi}}\left[\frac{2a_{20}^0}{3}(|A_0|^2+|A_{||}|^2-2|A_\perp|^2)+\frac{4a_{20}^2}{15}(|A_0|^2-\frac{1}{2}|A_{||}|^2+|A_\perp|^2)\right] \\
&- \frac{9}{16\sqrt{5\pi}}\frac{1}{\sqrt{1+\frac{4\tau_L\tau_H}{((\tau_L-\tau_H\sin 2\beta_s)^2)}}}\left[(A_{||}^*A_\perp+A_{||}A_\perp^*)(\frac{4}{3}a_{2-1}^0-\frac{4}{15}a_{2-1}^2)\right] \\
&+ \frac{9}{16}\frac{\sqrt{2}}{\sqrt{15\pi}}\frac{1}{\sqrt{1+\frac{4\tau_L\tau_H}{((\tau_L-\tau_H\sin 2\beta_s)^2)}}}\left[(A_0^*A_\perp+A_0A_\perp^*)(\frac{\pi a_{21}^1}{8}-\frac{\pi a_{21}^3}{32}+\ldots)\right] \\
&+ \frac{9}{8\sqrt{15\pi}}\left[\frac{2a_{22}^0}{3}(-|A_0|^2+|A_{||}|^2)-\frac{4a_{22}^2}{15}(|A_0|^2+\frac{1}{2}|A_{||}|^2)\right] \\
&+ \frac{9}{16}\frac{\sqrt{2}}{\sqrt{15\pi}}\left[(A_0^*A_{||}+A_0A_{||}^*)(\frac{\pi a_{2-2}^1}{8}-\frac{\pi a_{2-2}^3}{32}+\ldots)\right].
\end{aligned}
\tag{4.10}
$$

Normalization including a non-resonant S-wave contribution

The addition of terms in the PDF for an f_0/non-resonant K^+K^- contribution spoils the normalization just described. To reset the normalization of the new PDF to one, we must add two terms to N to find the new normalization, N_{swave}.

$$N_{swave} = (1 - f_{swave}) \cdot N + 2Re[\mathfrak{I}_\mu \cdot N'] + f_{swave} \cdot N'' \tag{4.11}$$

with

$$\begin{aligned} N' &= \sqrt{3}A_0^*(\frac{1}{6\sqrt{\pi}}a_{00}^1 + \frac{1}{12\sqrt{5\pi}}a_{20}^1 - \frac{1}{4\sqrt{15\pi}}a_{22}^1) \\ &+ \frac{3}{16}\sqrt{\frac{2}{5\pi}}A_\parallel^*(\frac{\pi}{2}a_{2-2}^0 - \frac{\pi}{8}a_{2-2}^2 + ...) \\ &+ \frac{3}{16}\sqrt{\frac{2}{5\pi}}A_\perp^*\frac{\sin 2\beta_s(\tau_L - \tau_H)}{\sqrt{((\tau_L - \tau_H)\sin 2\beta_s)^2 + 4\tau_L\tau_H}}(\frac{\pi}{2}a_{21}^0 - \frac{\pi}{8}a_{21}^2 + ...) \end{aligned} \tag{4.12}$$

and

$$N'' = \frac{1}{2\sqrt{\pi}}a_{00}^0 + \frac{1}{4\sqrt{5\pi}}a_{20}^0 - \frac{3}{4\sqrt{15\pi}}a_{22}^0, \tag{4.13}$$

where f_{swave} is the fraction of f_0/non-resonant K^+K^-, and $\mathfrak{I}_\mu$ was defined in Eq. 1.72 as an integral over the ϕ resonance, containing the dependence on the f_0/non-resonant K^+K^- phase δ_S. Eq. 4.11 is our final form of the normalization. The coefficients a_{lm}^k can be determined from data to fully parameterize the normalization.

Fitting the angular efficiency

We have introduced an analytic normalization for the PDF which accounts for the angular efficiency of the detector. We must now turn to handling the angular efficiency itself, which appears in the PDF.

We use the realistic $B_s^0 \to J/\psi\phi$ Monte Carlo described in the previous chapter to fit the angular efficiency. The Monte Carlo is generated with a phase-space model which yields flat distributions in the transversity angles before going through reconstruction. After reconstruction, the angles are sculpted by the detector efficiency in exactly the

same way that the data is sculpted. Thus, fitting the distribution of transversity angles for the Monte Carlo allows us to parameterize the detector efficiency. The Monte Carlo has been re-weighted according to track p_T and trigger composition, as described in Sec. 3.3, to ensure optimal data/Monte Carlo agreement in the transversity angles.

We can expand the efficiency function from Eq. 4.4 as a sum of Legendre polynomials:

$$\epsilon(\psi,\theta,\phi) = \sum_{lmk} a^k_{lm} P_k(\cos\psi) Y_{lm}(\theta,\phi), \tag{4.14}$$

where we equate the Monte Carlo (ψ,θ,ϕ) distribution with the detector efficiency. We can further expand the (θ,ϕ) dependence in a sum of real spherical harmonics:

$$Y_{lm}(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=0}^{l} [C_{lm}\cos(m\phi) + S_{lm}\sin(m\phi)] P_l^m(cos\theta), \tag{4.15}$$

and include the dependence on ψ by expanding C_{lm} and S_{lm} in Legendre polynomials

$$C_{lm} = \sum_{k=0}^{\infty} C^k_{lm}\sqrt{\frac{(2k+1)}{2}} P_k(\cos\psi), \tag{4.16}$$

$$S_{lm} = \sum_{k=0}^{\infty} S^k_{lm}\sqrt{\frac{(2k+1)}{2}} P_k(\cos\psi). \tag{4.17}$$

The expansion coefficients in Eq. 4.14, a^k_{lm} can be described in terms of C^k_{lm} and S^k_{lm}. We fit the distribution of transversity angles in the Monte Carlo to extract the expansion coefficients. We find that expanding to order 5 in l and m and order 2 in k is sufficient to make a good fit. The list of fitted coefficients is shown in Table 4.1.

A two dimensional projection of the fit in θ and ϕ, integrating over ψ, is shown in Fig. 4.1. The fit residual, which is the difference between the fit function and the Monte Carlo distribution, is shown in Fig. 4.2. The differences between the fit function and the distribution being fit are small, indicating that the fit is good.

4.2.2 Modeling of background angles

The final state angular distributions of signal events are predicted by the expanded version of Eqs. 1.70 and 1.71, but we do not have an *a priori* prediction for the PDF of

Parameter	Value	Parameter	Value
C_{00}^{0}	843.960 ± 0.779	C_{43}^{0}	-0.738 ± 0.784
C_{00}^{2}	-9.004 ± 0.780	C_{43}^{2}	0.608 ± 0.786
C_{11}^{0}	-3.733 ± 0.797	S_{43}^{0}	0.698 ± 0.786
C_{11}^{2}	-0.526 ± 0.797	S_{43}^{2}	0.768 ± 0.788
S_{11}^{0}	0.631 ± 0.770	C_{42}^{0}	3.80 ± 7.98
S_{11}^{2}	-0.435 ± 0.772	C_{42}^{2}	0.997 ± 7.999
C_{10}^{0}	0.226 ± 0.770	S_{42}^{0}	-0.580 ± 8.089
C_{10}^{2}	1.097 ± 0.771	S_{42}^{2}	-0.547 ± 8.101
C_{22}^{0}	40.294 ± 6.925	C_{41}^{0}	-0.444 ± 0.769
C_{22}^{2}	-2.710 ± 6.938	C_{41}^{2}	0.761 ± 0.770
S_{22}^{0}	0.468 ± 7.018	S_{41}^{0}	-0.279 ± 0.746
S_{22}^{2}	0.504 ± 7.028	S_{41}^{2}	-0.538 ± 0.748
C_{21}^{0}	0.784 ± 0.792	C_{40}^{0}	-13.408 ± 0.800
C_{21}^{2}	1.267 ± 0.793	C_{40}^{2}	-1.166 ± 0.801
S_{21}^{0}	-0.303 ± 0.761	C_{55}^{0}	0.214 ± 0.784
S_{21}^{2}	-1.301 ± 0.764	C_{55}^{2}	-0.674 ± 0.786
C_{20}^{0}	-32.327 ± 0.771	S_{55}^{0}	6.078 ± 0.784
C_{20}^{2}	0.227 ± 0.772	S_{55}^{2}	-0.946 ± 0.785
C_{33}^{0}	-0.094 ± 0.451	C_{54}^{0}	2.029 ± 0.786
C_{33}^{2}	-0.188 ± 0.452	C_{54}^{2}	-0.187 ± 0.787
S_{33}^{0}	2.599 ± 0.455	S_{54}^{0}	0.357 ± 0.787
S_{33}^{2}	-1.092 ± 0.456	S_{54}^{2}	0.253 ± 0.789
C_{32}^{0}	11.298 ± 6.119	C_{53}^{0}	-0.790 ± 0.777
C_{32}^{3}	-1.264 ± 6.131	C_{53}^{2}	1.179 ± 0.778
S_{32}^{0}	-0.026 ± 6.199	S_{53}^{0}	-2.598 ± 0.782
S_{32}^{2}	0.523 ± 6.208	S_{53}^{2}	0.244 ± 0.783
C_{31}^{0}	0.384 ± 0.776	C_{52}^{0}	6.060 ± 3.667
C_{31}^{2}	-0.708 ± 0.777	C_{52}^{2}	0.001 ± 3.674
S_{31}^{0}	-0.716 ± 0.755	S_{52}^{0}	0.045 ± 3.715
S_{31}^{2}	-1.186 ± 0.757	S_{52}^{2}	-0.024 ± 3.721
C_{30}^{0}	-0.280 ± 0.786	C_{51}^{0}	-0.913 ± 0.758
C_{30}^{2}	1.042 ± 0.786	C_{51}^{2}	-0.095 ± 0.758
C_{44}^{0}	-21.416 ± 0.782	S_{51}^{0}	1.197 ± 0.738
C_{44}^{2}	-0.690 ± 0.784	S_{51}^{2}	-1.135 ± 0.739
S_{44}^{0}	-0.110 ± 0.786	C_{50}^{0}	-1.891 ± 0.828
S_{44}^{2}	-0.036 ± 0.788	C_{50}^{2}	0.545 ± 0.830

Table 4.1: Fitted values of expansion coefficients for the angular efficiency function.

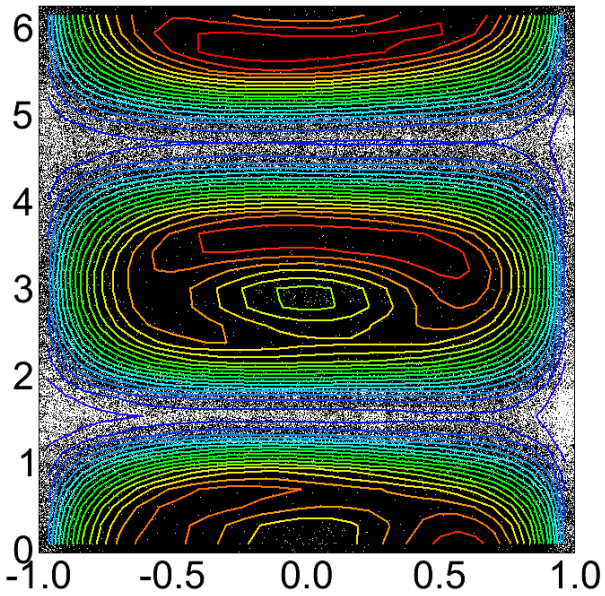

Figure 4.1: Projection in $\cos\theta$ and ϕ space of the fit of the Monte Carlo transversity angles.

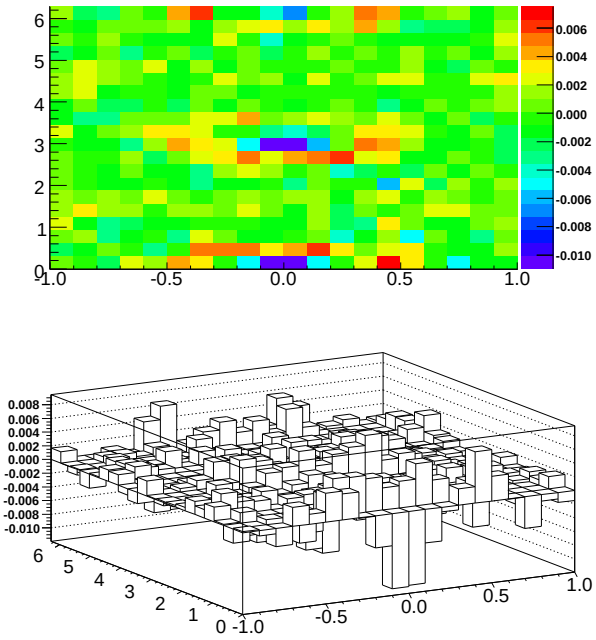

Figure 4.2: Residual for the fit in $\cos\theta$ and ϕ space of the Monte Carlo transversity angles.

the background angular distributions. We describe the background transversity angles $\cos\theta, \phi$ and $\cos\psi$ using events from the B_s^0 mass sideband regions.

We assume that correlations between the angles are small, and that we can describe each angle with an independent function:

$$f(\psi, \theta, \phi) = f(\psi) f(\theta) f(\phi). \tag{4.18}$$

The validity of this assumption is tested, and a systematic uncertainty is assigned to the small correlation found between the angles. This is described in Sec. 5.2.2.

We parameterize the distributions empirically, using the following expressions:

$$f(\cos\theta) = \frac{a_1 cos^2(\theta)}{2a_1/3} \tag{4.19}$$

$$f(\phi) = \frac{b_1 cos(2\phi)}{2\pi} \tag{4.20}$$

$$f(\cos\psi) = \frac{c_1 cos^2(\psi)}{2c_1/3}. \tag{4.21}$$

The background angular distributions are fit using our empirical functions to obtain starting values for the coefficients a_1, b_1 and c_1, and to ensure that the empirical description of the angles is reasonable. The coefficients are allowed to float freely in the unbinned maximum likelihood fit used to extract β_s. The agreement between the fit and the background angular distributions is shown in Fig. 4.3.

4.3 Flavor tagging

To maximize sensitivity to the *CP*-violating phase β_s, we must determine whether the measured B meson was a B_s^0 or $\bar{B}_s^0$ at production. This is referred to as the production flavor: whether the initial B meson contained a b or $\bar{b}$ quark. As we determined in Chapter 1, if we do not distinguish whether the meson was initially a B_s^0 or $\bar{B}_s^0$, we lose access to many terms in the probability density function that contain β_s. It is worth noting that some decays are self-tagging, or in other words, decayed to a state

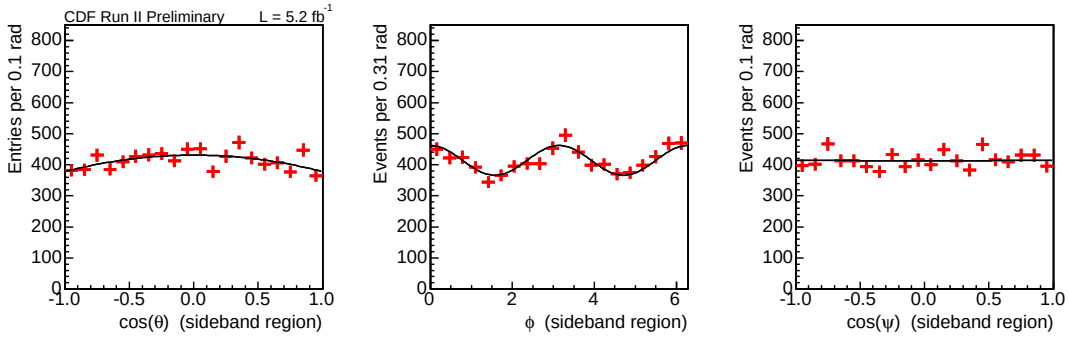

Figure 4.3: Distribution of transversity angles in the B_s^0 mass sideband region.

with charged particles that allowed the determination of the flavor at decay time, with the external knowledge of Δm_s and measurement of the proper decay time, we could determine the initial meson flavor. Because the J/ψ and ϕ are neutral, we cannot identify the flavor at decay, and therefore must rely on flavor tagging to identify the production flavor. Flavor tagging algorithms assign a tagging decision of $\xi = +1$, -1, or 0, corresponding to B_s^0 at production, $\bar{B}_s^0$ at production, or no tag assigned.

There are two general classes of flavor tagging algorithms, or flavor taggers: opposite side and same side. A complete description of the algorithm is beyond the scope of this thesis. Only a brief overview of the methods is given. Many references exist for both the opposite side tagger [59, 60, 61] and the same side tagger [62, 63, 64].

The schematic in Fig. 4.4 illustrates the two types of tagging. The same side tagger uses a charged track produced in association with the $b/\bar{b}$ quark in the $\bar{B}_s^0/B_s^0$. Its charge tags the production flavor of the B meson. The opposite side taggers capitalize on the fact that most b quarks produced in $p\bar{p}$ collisions are produced as $b\bar{b}$ pairs. The b or $\bar{b}$ quark on the opposite side of the B_s^0 candidate hadronizes into a B meson. The B meson can be tagged by its decay products, although the situation is somewhat complicated if it is a B^0 meson, and and can mix before decaying. If the opposite side quark is tagged as a b, the B_s^0 candidate must contain a $\bar{b}$, and if the opposite side quark is tagged as a $\bar{b}$, the $\bar{B}_s^0$ candidate must contain a b. Opposite tagging is based on several algorithms that tag on the type of opposite side decay product, including the soft lepton tagger, the soft muon tagger and the jet charge tagger.

In a real detector environment, taggers suffer from inefficiencies and mis-tags. We introduce several variables to quantify the tagging power. The first consideration is

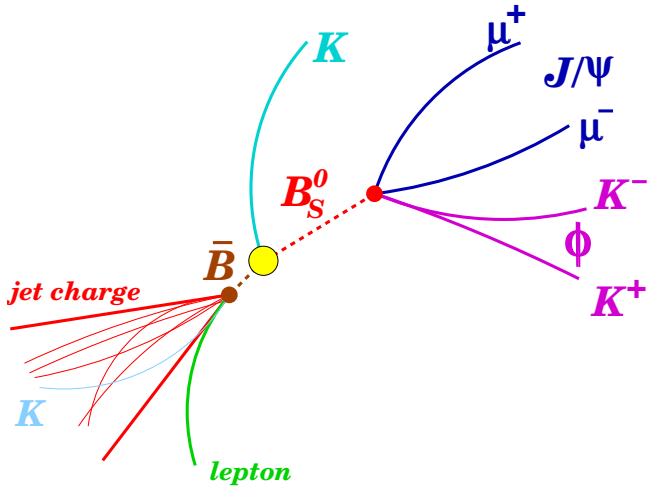

Figure 4.4: Schematic demonstrating same and opposite side flavor tagging algorithms.

the inefficiency of the tagger. Not every event receives a tag, as there is not always adequate information to make a tag decision. The efficiency is defined as the number of events that receive a tag, divided by the total number of events considered.

The dilution $\mathcal{D}$ quantifies the probability of an event receiving an incorrect tag. The dilution is defined as

$$\mathcal{D} = 1 - 2p_W, \tag{4.22}$$

where p_W is the probability of a mis-tag. Thus, the probability of a mis-tag is $p_W = (1 - \mathcal{D})/2$, while the probability of a correct tag is $p_R = (1 + \mathcal{D})/2$. The dilution is obtained by counting the number of incorrectly assigned tags,

$$\mathcal{D} = \frac{N_R - N_W}{N_R + N_W}, \tag{4.23}$$

where N_R is the number of correct tags and N_W is the number of incorrect tags. Clearly, the optimal dilution has a value of one (zero mistags). A dilution of zero corresponds to random tagging assignments, which have a 50% chance of tagging correctly or incorrectly. A dilution of -1, is in fact, as powerful as a dilution of 1. If the mis-tag rate is known to be 100%, one can merely flip the tag decision to obtain perfect dilution.

In addition to the tagging decision ξ, tagging algorithms assign an event-by-event predicted dilution, based on the event variables on which the mis-tag probability depends. The dilution prediction is based on the detector environment for a particular event. The predicted dilutions are important because they assign a confidence to the tag decision and make it possible to transfer a tagging algorithm from the data sample on which it was developed to a different data sample. The predicted dilution distributions for our data sample are shown in Figs. 4.5 and 4.6. Fig. 4.5 shows the predicted dilution for the opposite side tagger, for sideband subtracted signal events (signal), and for events from the sideband region (background). Fig. 4.6 shows the predicted dilution distribution for the same side tagger. Note that for both the opposite and same side taggers, the predicted dilution distributions are slightly different for signal and background events. We account for this effect in the unbinned maximum likelihood fit.

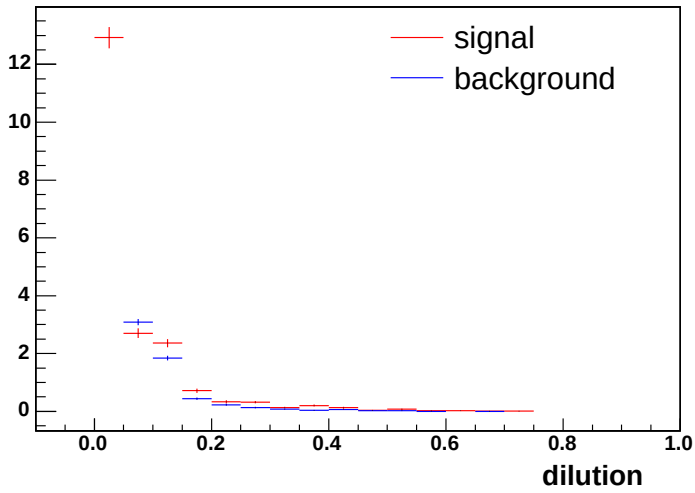

Figure 4.5: Predicted dilution for opposite side tagger.

The tagging power depends on both dilution and efficiency. The most commonly used figure of merit to characterize the tagging power is $\epsilon\langle\mathcal{D}^2\rangle$, where the average dilution is calculated by summing over the event by event dilution.

We use multiple tagging algorithms in this analysis, which necessitates a combination of the dilutions for an overall measure of the tagging power. Assuming that two

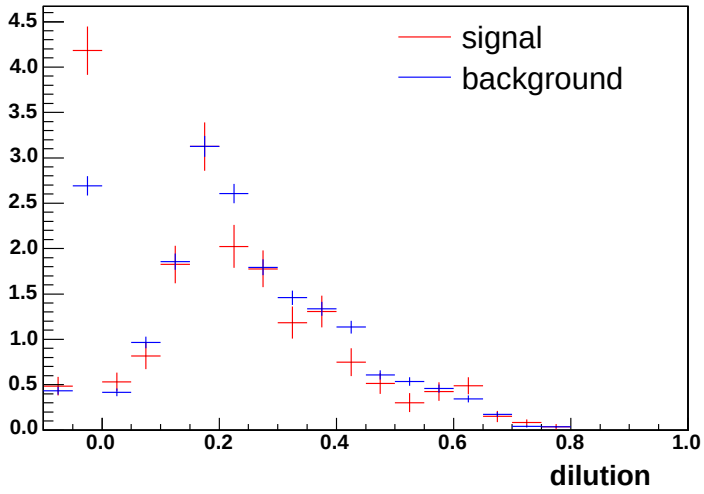

Figure 4.6: Predicted dilution for same side tagger.

tagging algorithms are independent, the dilutions can be combined as follows [65]. If the tagging decisions agree, the combined dilution is

$$\mathcal{D}_c = \frac{\mathcal{D}_1 + \mathcal{D}_2}{1 + \mathcal{D}_1\mathcal{D}_2}. \tag{4.24}$$

If the decisions disagree, the combined dilution is

$$\mathcal{D}_c = \frac{|\mathcal{D}_1 - \mathcal{D}_2|}{1 - \mathcal{D}_1\mathcal{D}_2}. \tag{4.25}$$

The probability of a particular combined tag decision is dependent on the efficiency of each of the taggers, in the following way:

$$P(\vec{\xi}) = \begin{cases} (1-\epsilon_1)(1-\epsilon_2) & (\xi_1 = 0,\ \xi_2 = 0) \\ \epsilon_1(1-\epsilon_2) & (\xi_1 = \pm 1,\ \xi_2 = 0) \\ (1-\epsilon_1)\epsilon_2 & (\xi_1 = 0,\ \xi_2 = \pm 1) \\ \epsilon_1\epsilon_2 & (\xi_1 = \pm 1,\ \xi_2 = \pm 1) \end{cases} \tag{4.26}$$

In the case of combining an opposite side tag with a same side tag, the algorithms are indeed independent, since after production of the initial $b\bar{b}$ pair, the opposite and same side hadronization and decay processes are uncorrelated.

4.3.1 Opposite side tagging

The opposite side flavor tagger is a combination of several algorithms: the soft muon tagger (SMT), the soft electron tagger (SET), and the jet charge tagger (JQT). The names refer to the charged object used to determine the flavor of the opposite side b quark. Each of the algorithms will be discussed individually.

The outcomes of the three tagging algorithms are combined to give a single opposite side tag decision and predicted dilution. This can be done hierarchically, based on the relative dilutions of the taggers. The SMT has the highest dilution, followed by the SET and then the jet charge tagger. Thus, if there is a soft muon tag decision for an event, it will be taken as the final tag decision. If there is no soft muon tag decision, but a soft electron tag decision exists, the soft electron tag decision will be taken as the final tag decision. If no soft muon or soft electron tag decision exists, the jet charge tag decision is used.

In this thesis, we optimize the total dilution by combining the taggers using a neural network. The neural network has the benefit of automatically handling correlations between the tagging algorithms. This is helpful for combining the various opposite side taggers, which are not guaranteed to be independent. It is trained on data from the SVT+l trigger.

Lepton Tagging

The lepton taggers, SMT and SET, utilize the charge of a muon or electron to determine the production flavor of their parent B meson. As they are the products of semi-leptonic decays $b \to cl^-\bar{\nu}_l X$ and $\bar{b} \to \bar{c}l^+X$, one can infer that a positive muon or positron comes from a $\bar{b}$ at production, and a negative muon or electron from a b.

The efficiency of the lepton taggers is low, due to the low branching fractions for semi-leptonic decays ($\mathcal{B} \to lX$). Also, detector and tracking inefficiencies lower the efficiency of the lepton tagging algorithms. The mis-tag rate is increased by the possibility of mixing, if the lepton came from a B^0 decay. Mis-tags are also caused by the presence of $b \to c \to lX$ decays, which can be misidentified as $b \to lX$ decays. In this case, the lepton will have the opposite sign as if it came directly from a b transition, and the tag decision will be the negative of the correct decision. Compared to the jet charge tagging algorithms, the lepton tagging algorithms have a high dilution but a low efficiency.

The leptons used in the tagging algorithms are selected based on their electron or muon likelihoods. Track quality cuts are imposed, ensuring that low p_T, poorly measured tracks are not used. The predicted dilution is determined using the lepton likelihoods, and the lepton's p_T^{rel}, the transverse momentum of the track relative to the axis of the cone in which it was found. The axis of the cone is determined by the B momentum direction. To be selected as a track for tagging, a track must fall within a cone of a selected radius. A schematic illustrating the cone and p_T^{rel} is shown in Fig. 4.7.

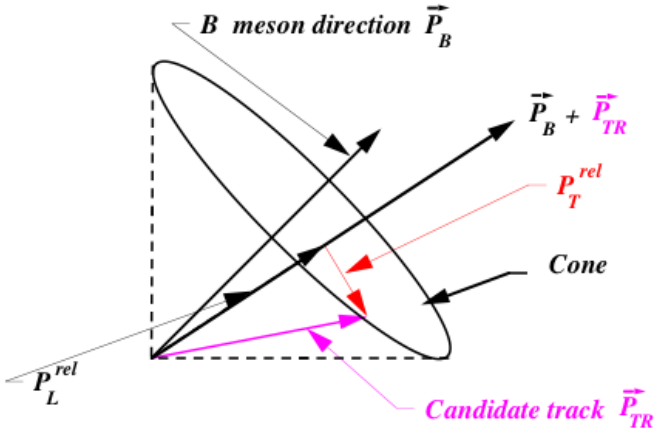

Figure 4.7: Illustration of p_T^{rel} and cone.

Jet Charge Tagging

The jet charge tagging algorithms use the fact that if a b quark fragments into a jet, the jet should have the same charge as the b quark the produced it. The charge of the jet is determined by taking the momentum-weighted sum of the tracks in the jet, as follows:

$$Q_{jet} = \frac{\sum_i Q^i p_T^i (1 + P_{trk}^i)}{\sum_i p_T^i (1 + P_{trk}^i)}, \tag{4.27}$$

where P_{trk}^i is the probability of the track being part of the b jet. If Q_{jet} is approximately $-1/3$, the jet is presumed to have come from a b quark, if it is $+1/3$, the jet is presumed to have come from a $\bar{b}$ quark.

The jet is identified by defining a cone around the momentum direction of the reconstructed B meson. Tracks belonging to the jet must fall within the cone. Various

selection requirements based on track p_T and impact parameter are applied to further isolate the tracks in the jet. If a displaced vertex can be identified within the cone, the chances that the jet was produced by b fragmentation are increased.

Jets are separated into three classes, based on their ability to meet the b jet selection requirements. A Class 1 jet has an displaced vertex within the jet isolation code, and $L_{xy}/\sigma(L_{xy}) > 3$. A Class 2 jet is any jet containing a track with $P_{trk} > 0.5$ GeV/c. A Class 3 jet is a jet that passes the selection requirements for identifying a jet. A Class 3 jet can be identified for nearly 100% of events. Due to the loose restrictions on the jets used for the jet charge tagging algorithm, the tagger has an very high efficiency, but poor purity. The average dilution for the JQT is significantly lower than for the semi-leptonic taggers.

Calibration of opposite side tagging

The opposite side taggers are calibrated on data from the SVT+l trigger, while this analysis uses data from the di-muon trigger. Thus, it is necessary to verify that the predicted dilution is not sample dependent, and that the measured dilution on di-muon data matches the predicted dilution from calibration on the SVT+l data.

We study the behavior of the opposite side taggers on the control channel $B^+ \to J/\psi K^+$. It is assumed that the opposite side hadronization has identical behavior on the $B^+ \to J/\psi K^+$ sample as on $B_s^0 \to J/\psi\phi$. Since the $B^+ \to J/\psi K^+$ sample is self-tagging, the predicted dilution can be compared to the true, observed dilution to calculate a dilution scale factor. The $B^+ \to J/\psi K^+$ channel also has the benefit of higher statistics than the B_s^0 samples: 52,000 signal events. $B^+ \to J/\psi K^+$ decays are selected from the di-muon data sample using a similar neural network to the one used for selecting $B_s^0 \to J/\psi\phi$ decays. The cut on the neural network output is a chosen by optimizing $S/\sqrt{S+B}$. The resulting $J/\psi K^+$ invariant mass distribution is shown in Fig. 4.8.

The production flavor of the B^+ is easily identified by checking the charge of the kaon daughter track, since a K^+ must be the daughter of a B^+, and a K^- must be the daughter of a B^-. Using the kaon charge, we can determine the measured dilution in bins of the predicted dilution. If the shape of the measured versus predicted dilution histogram is a straight line with a slope consistent with unity, we can conclude that the tagger calibration from the SVT+l trigger data is appropriate for di-muon data. The slope of this line gives us the value for a scale factor associated with the opposite

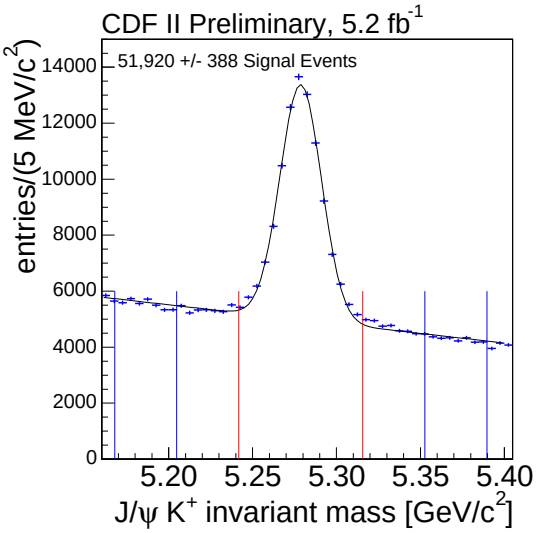

Figure 4.8: $J/\psi K^+$ invariant mass distribution.

side tagger. The scale factor accounts for the effect of applying the OST on a new data sample, and takes any unexpected inefficiencies into consideration that prevent the real dilution from matching the predicted one.

We use two scale factors for the opposite side tagger, one for B^+ mesons and one for the B^- mesons, to allow for any asymmetry in the tagging algorithms. For both B^+ and B^-, we use one scale factor to describe the entire dataset. As a cross-check, though, we determine the scale factors in different periods of the data, to ensure that the scale factors are stable and consistent with one throughout all parts of the data. The measured versus predicted dilution histograms for B^+ and B^- are shown in Fig. 4.9. The corresponding histograms for different times of data taking are shown in Figs. 4.10-4.12. Table 4.2 reports the OST dilution scale factors in different parts of the data. Additionally, we measure a tagging efficiency of 94.3±0.3 and an average predicted dilution on the $B^+ \to J/\psi K^+$ signal of (6.9±0.1)%.

Scale Factor	Period 0-9	Period 10-16	Period 17-25	Period 0-25
$S_{\mathcal{D}}^+$	0.83 ± 0.13	0.89 ± 0.16	0.98 ± 0.16	0.93 ± 0.09
$S_{\mathcal{D}}^-$	1.03 ± 0.17	1.10 ± 0.18	1.13 ± 0.15	1.12 ± 0.10

Table 4.2: OST dilution scale factors for B^+ and B^- in different parts of the data.

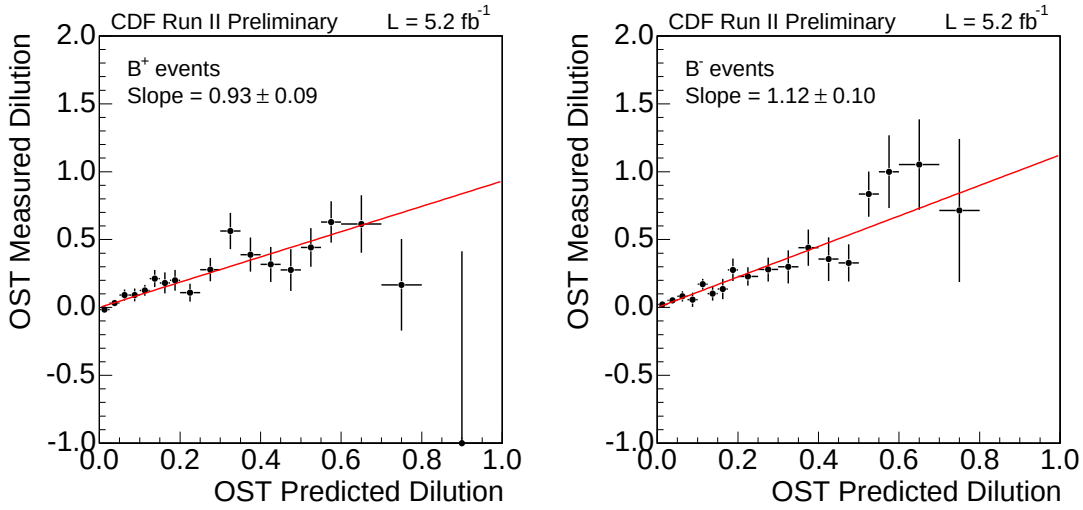

Figure 4.9: Measured versus predicted dilution for B^+ (left) and B^- (right) for the entire data sample.

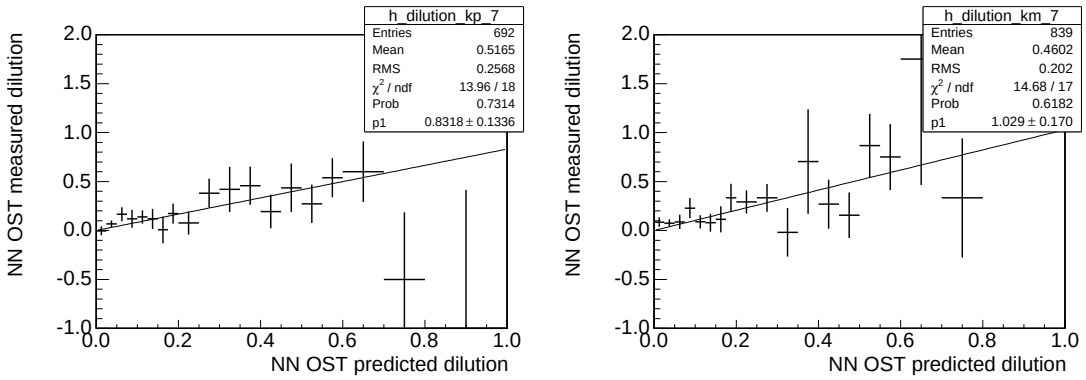

Figure 4.10: Measured versus predicted dilution for B^+ (left) and B^- (right) for data taken between February 2002 and November 2006.

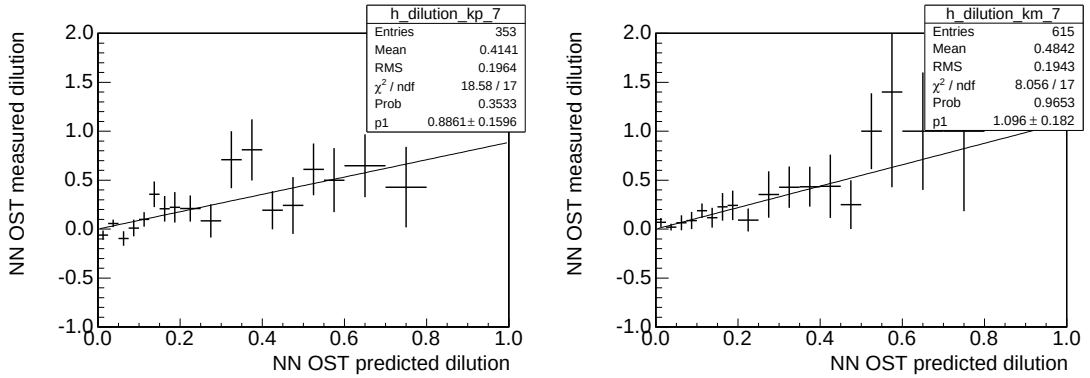

Figure 4.11: Measured versus predicted dilution for B^+ (left) and B^- (right) for data taken between November 2006 and February 2008.

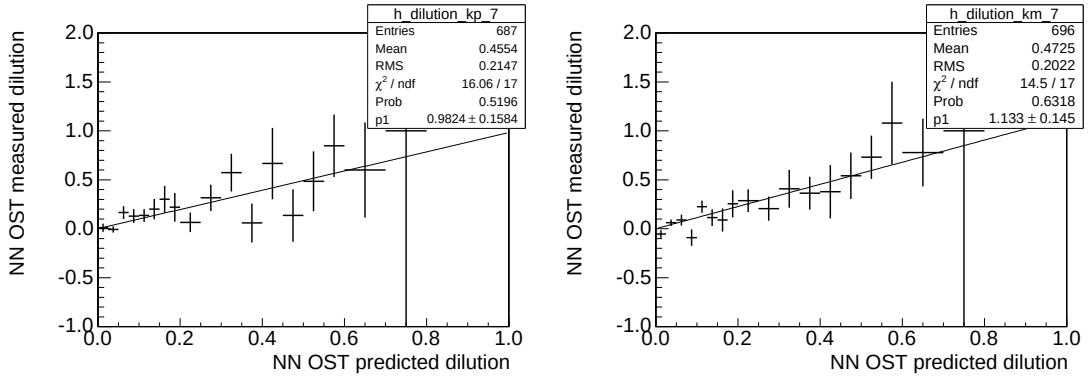

Figure 4.12: Measured versus predicted dilution for B^+ (left) and B^- (right) for data taken between February 2008 and June 2009.

4.3.2 Same side tagging

The principle of the same side tagging method is significantly different from that of opposite side tagging, where the hadronization and fragmentation of the opposite side b quark is independent of the species of the B meson of interest. Therefore, we can develop an opposite side tagging algorithm on high statistics B^+ samples.

In the case of same side tagging, we tag with tracks that are produced in association with the reconstructed B meson. The type of track produced in association with the B meson varies depending on the species of the reconstructed B meson, as shown in Fig. 4.13. The associated track is a combination of the non-b quark in the reconstructed B meson, and a matching anti-quark from a vacuum-produced $q\bar{q}$ pair. Thus, the other quark in the B meson determines the type of associated track produced, which will be different for a B^+, a B^0, or a B_s^0 meson.

Thus, the SST algorithm was developed on a high statistics B_s^0 Monte Carlo sample, and is calibrated on B_s^0 data. In the case of the B_s^0 meson, the leading associated track is a kaon. A K^+ will tag a B_s^0 meson, while a K^- tags a $\bar{B}_s^0$ meson. However, the B_s^0 could be accompanied by an associated track from a neutral K^{*0} or Λ, which are not able to tag the B_s^0 flavor and therefore lower the tagging efficiency.

The selection on associated tracks used for same side tagging is similar to the selection for opposite side tracks. Tracks must be within a cone around the B_s^0 momentum to be considered tag candidates, and tracks must have $p_T > 0.4$ GeV/c. The same side tagging power benefits from the use of particle identification (dE/dx and time of flight) to identify the associated track as a kaon. This decreases the mis-tag rate, which occurs if a pion is mistakenly identified as an associated kaon. Particle identification, as well as information about track momentum, is used to decide which track to choose as the tagging track, if there are multiple possible tagging tracks in an event.

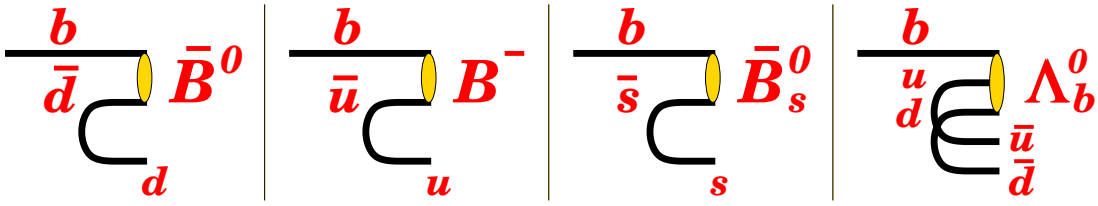

Figure 4.13: Tracks produced in association with B mesons, used for same side tagging.

Calibration of same-side tagging

The same side kaon tagger (SSKT) has been calibrated using the most recent measurement of B_s^0 mixing [66], made on 5.2 fb^{-1} of data. This is the first time that the dilution scale factor for the same side kaon tagger has been determined from data, rather than Monte Carlo. The calibration is achieved in the following way.

The probability for observing the B_s^0 in a flavor eigenstate as a function of time is

$$P(t)_{B_s^0,\bar{B}_s^0} \propto |1 \pm \cos \Delta m_s t|. \tag{4.28}$$

Adding in the effect of the measured dilution, our measurement probability becomes

$$P(t)_{B_s^0,\bar{B}_s^0} \propto |1 \pm \mathcal{D} \cos \Delta m_s t|. \tag{4.29}$$

This can be re-expressed in terms of an amplitude,

$$P(t)_{B_s^0,\bar{B}_s^0} \propto |1 \pm A\mathcal{D}_p \cos \Delta m_s t|, \tag{4.30}$$

where $\mathcal{D}_p$ is the predicted dilution. The likelihood that uses this probability is normalized such that at the correct value of the mixing frequency Δm_s, the amplitude is one. If we fix Δm_s and measure the amplitude at each value of Δm_s, the true value of Δm_s should have the maximum amplitude. If the measured and predicted dilutions are equal, the maximum amplitude will be exactly one, by normalization. The amplitude is a scale factor for the dilution, relating the predicted dilution to the measured dilution.

The B_s^0 mixing measurement was made on 5.2 fb^{-1} of data collected with the two-track trigger. B_s^0 events were fully reconstructed in four modes: $B_s^0 \to D_s\pi$ with $D_s \to \phi\pi$, $B_s^0 \to D_s\pi$ with $D_s \to K^*K$, $B_s^0 \to D_s\pi$ with $D_s \to 3\pi$, and $B_s^0 \to D_s 3\pi$ with $D_s \to \phi\pi$.

The Δm_s amplitude scan on 5.2 fb^{-1} of data is shown in Fig. 4.14. The maximum amplitude in Δm_s occurs at Δm_s=17.79±0.07(stat.) ps^{-1}, a value consistent with the PDG average. We use the PDG average as an input in the unbinned maximum likelihood fit, because the errors on Δm_s are smaller. The size of the amplitude at maximum and the measured dilution scale factor for the same side kaon tagger is $\mathcal{S}_\mathcal{D}$=0.94±0.15 (stat)±0.13 (syst). We measure a tagging efficiency of 52.2±0.7, and an average predicted dilution on B_s^0 signal of 21.8±0.3%.

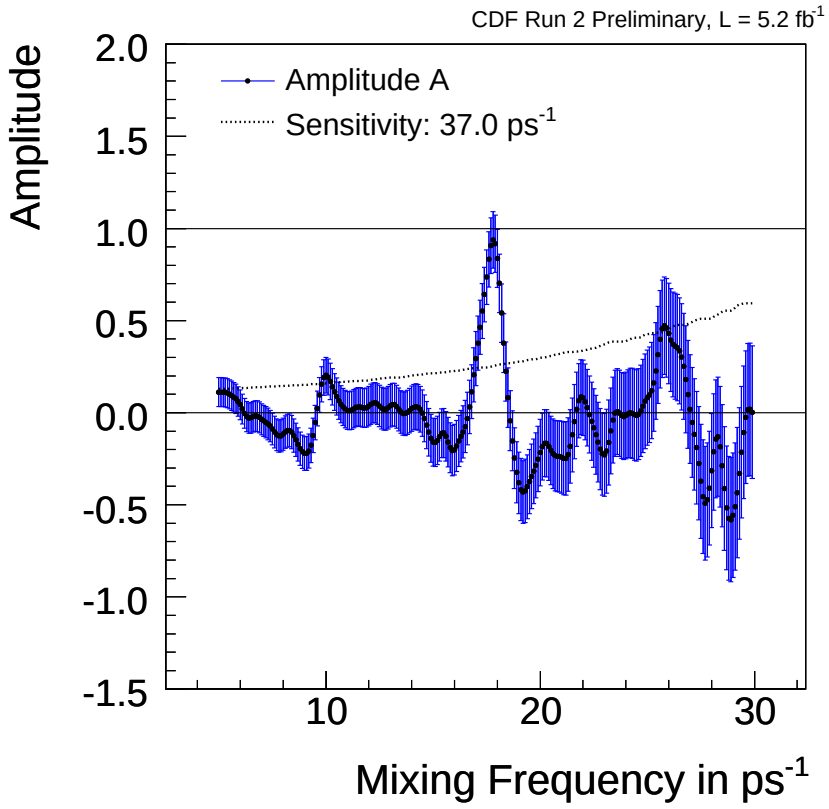

Figure 4.14: Δm_s amplitude scan.

4.3.3 Tagging power

The tagging power $\epsilon\langle\mathcal{D}^2\rangle$ is scaled by the dilution scale factor to be $\epsilon\mathcal{S}\langle\mathcal{D}^2\rangle$. The tagging power is 1.2±0.2% for OST, and 3.5±1.4% for SSKT. The tagging power is quite low at the Tevatron, compared to the tagging power at the B factories or even the predicted tagging power at LCHb. As a result, the main effect of tagging in this analysis is to break the $\beta_s \to -\beta_s$ symmetry that exists without tagging, reducing the number of valid solutions for β_s from four to two. The tagging power is not large enough to substantially reduce the errors on the remaining solutions. Each of the four untagged β_s solutions has comparable errors to those on the tagged β_s solutions. An example of this using a pseudo-experiment is shown in Fig. 4.15. The toy was fit with and without considering tagging information for a grid of β_s and $\Delta\Gamma$ values. Including tagging removes one of the symmetries of the probability density, removing half of the allowed region in $\beta_s - \Delta\Gamma$ space. The errors on the remaining allowed region, though, are not noticeably reduced. If the tagging power were greater, we would expect our sensitivity to β_s to be substantially better in the tagged case, and the errors on β_s to be smaller.

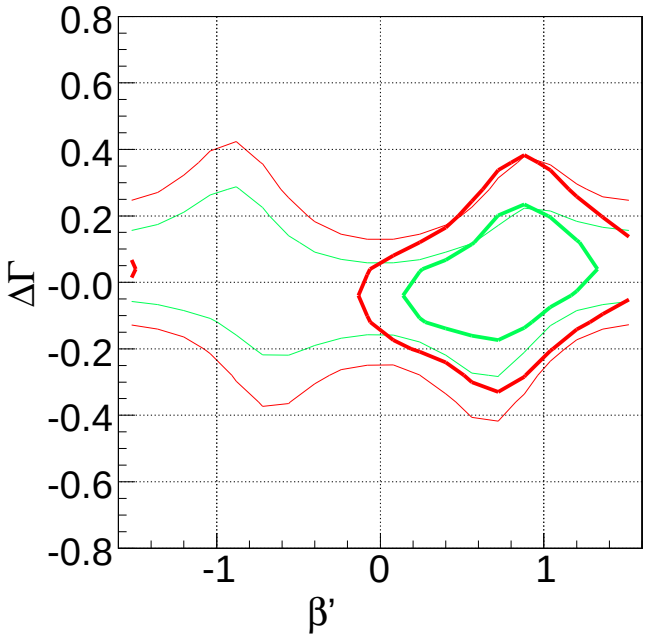

Figure 4.15: A pseudo-experiment fit with and without considering tagging information.

4.4 Unbinned maximum likelihood fit

We are now ready to introduce the likelihood function, which is a function of mass, flavor tagging information, proper time, transversity angles, and the proper time error distributions. The likelihood is different for signal and background events. It contains the probability densities of Eq. 1.70 and 1.71 after including the detector efficiency and the analytic normalization after the inclusion of the detector efficiency, such that

$$P'(\psi,\theta,\phi,t) = \frac{1}{N}P(\psi,\theta,\phi,t)\epsilon(\psi,\theta,\phi) \tag{4.31}$$

and

$$\bar{P}'(\psi,\theta,\phi,t) = \frac{1}{N}\bar{P}(\psi,\theta,\phi,t)\epsilon(\psi,\theta,\phi). \tag{4.32}$$

The likelihood function is as follows:

$$\begin{aligned}\mathcal{L} &= f_s \cdot P_s(m) \cdot P_s(\xi) \cdot T(t,\psi,\theta,\phi,\mathcal{D},\xi) \cdot P_s(\sigma_t) \cdot P_s(\mathcal{D}) \\ &+ (1-f_s) \cdot P_b(m) \cdot P_b(\xi) \cdot P_b(t,\sigma_t) \cdot P_b(\psi) \cdot P_b(\theta) \cdot P_b(\phi) \cdot P_b(\sigma_t) \cdot P_b(\mathcal{D}).\end{aligned} \tag{4.33}$$

4.4.1 Signal likelihood

We present here a description of all the components of the signal likelihood function.

Mass ($P_s(m)$)

The B_s^0 invariant mass distribution is modeled with a single Gaussian. Errors on the mass distribution are assessed using event-by-event errors. The error distributions are assumed to be the same for signal and background. We later assign a systematic uncertainty to handle this assumption. A mass scale factor is floated in the fit to handle for our imperfect knowledge of the mass resolution. The mass scale factor multiplies each event-by-event mass error.

Proper time, angles, tagging ($T(t,\psi,\theta,\phi,\mathcal{D}_1,\mathcal{D}_2,\xi_1,\xi_2)$)

The core of the likelihood function for signal, containing the proper time, angular, and tagging dependence is given by

$$T(t,\psi,\theta,\phi,\mathcal{D}_1,\mathcal{D}_2,\xi_1,\xi_2) = \frac{1+\xi_1 s_1 \mathcal{D}_1}{1+|\xi_1|}\frac{1+\xi_2 s_2 \mathcal{D}_2}{1+|\xi_2|}P'(t,\psi,\theta,\phi)\otimes G_1(\sigma_t)G_2(\sigma_t)$$
$$+\frac{1-\xi_1 s_1 \mathcal{D}_1}{1+|\xi_1|}\frac{1-\xi_2 s_2 \mathcal{D}_2}{1+|\xi_2|}\bar{P}'(t,\psi,\theta,\phi)\otimes G_1(\sigma_t)G_2(\sigma_t). \tag{4.34}$$

The probability densities P' and $\bar{P}'$ are smeared with the functions $G_1(\sigma_t)$ and $G_2(\sigma_t)$ to account for the proper time resolution ($\sigma_t \sim$25 μm). We model the proper time resolution with two Gaussians, one broad, one narrow. This model was chosen because it best describes the shape of the resolution function in the data. The Gaussians each have an associated scale factor which floats in the fit, to handle possible mis-modeling of the resolution. The fraction of the narrow to the broad Gaussian also floats in the fit.

The coefficients on P' and $\bar{P}'$ can be understood by returning to the expression of the probability of correct tag, $(1+\mathcal{D})/2$ for a B_s^0, and $(1-\mathcal{D})/2$ for a $\bar{B}_s^0$. The probability density for the tag decision is then

$$P_{B_s^0,\bar{B}_s^0} = \frac{(1+\mathcal{D})}{2}P + \frac{(1-\mathcal{D})}{2}\bar{P}, \tag{4.35}$$

or, including the possibility of mis-tag or no tag,

$$P_{B_s^0,\bar{B}_s^0} = \frac{1+\xi\mathcal{D}}{1+|\xi|}P + \frac{1-\xi\mathcal{D}}{1+|\xi|}\bar{P}. \tag{4.36}$$

For two independent taggers, the coefficients to the probability densities combine multiplicatively to

$$P_{B_s^0,\bar{B}_s^0} = \frac{1+\xi_1\mathcal{D}_1}{1+|\xi_1|}\frac{1+\xi_2\mathcal{D}_2}{1+|\xi_2|}P + \frac{1-\xi_1\mathcal{D}_2}{1+|\xi_1|}\frac{1-\xi_2\mathcal{D}_2}{1+|\xi_2|}\bar{P}. \tag{4.37}$$

Including the dilution scale factors s_1 and s_2, we have the coefficients shown in Eq. 4.34. The dilution scale factors, as well as Δm_s, are determined from independent measurements, not from our likelihood fit. Consequently, they are not allowed to float freely in the fit, but are Gaussian-constrained within their errors in the likelihood fit.

Tagging decision and dilution ($P_s(\xi), P_s(\mathcal{D})$)

The probability density $P_s(\xi)$ is the probability density of a combined tagging decision, using two taggers (opposite side and same side). The form of this PDF is given in Eq. 4.26.

Tagging dilution distributions are taken from data. In the case of the signal, they are the side-band subtracted distributions taken from the B_s^0 invariant mass signal region and stored in histograms. The histograms are normalized to translate from dilution distributions to probability densities for the dilution. The fit accesses one histogram for the opposite side tagger and another for the same side tagger.

Proper decay time errors ($P_s(\sigma_t)$)

The errors on the proper decay time are modeled using Gamma functions, such as:

$$\Gamma(x) = \frac{x^a e^{-x/b}}{b^{a+1}\Gamma(a+1)}. \tag{4.38}$$

Two Gamma functions are used to describe the σ_t distribution for signal.

4.4.2 Background likelihood

We present here a description of all the components of the background likelihood function.

Mass ($P_b(m)$)

The probability density for the background invariant mass is linear, modeled with a degree one polynomial.

Tagging decision and dilution ($P_b(\xi), P_b(\mathcal{D})$)

In the case of background, we are not concerned with the event-by-event dilution, as we do not wish to determine the production flavor of background events. Knowledge of the fractions of positively and negatively charged events is sufficient to describe the behavior of the background. Therefore, the probability density $P_b(\xi)$ is equal to the fraction of events with positive tags for ξ=+1, and equal to the fraction of events with negative tags for ξ=-1.

The background dilution distributions are handled analogously to the signal dilution. The dilution distributions are taken from the B_s^0 invariant mass sideband region, normalized to form a probability density, and stored in histograms.

Proper decay time and errors ($P_b(t|\sigma_t), P_b(\sigma_t)$)

For the background proper decay time distribution, we follow CDF convention and parameterize with a prompt peak, two positive exponentials, and one negative exponential. The prompt peak comes mostly from J/ψ's that are not the decay products of a B meson. The prompt peak, originally a delta function, is smeared with two Gaussians in the same manner as the signal, to account for resolution effects. The exponentials include long lived backgrounds from non B_s^0 decays and mis-reconstructed B decays. The form of the probability density is

$$\begin{aligned} P_b(t|\sigma_t) &= \left(f_g + (1-f_g)\left(\frac{f_{++}}{\lambda_{++}}e^{-\frac{t}{\lambda_{++}}} + (1-f_{++})\left(\frac{f_-}{\lambda_-}e^{\frac{t}{\lambda_-}} + \frac{(1-f_-)}{\lambda_+}e^{-\frac{t}{\lambda_+}}\right)\right)\right) \\ &\otimes G_1(\sigma_t)G_2(\sigma_t), \end{aligned} \tag{4.39}$$

where f_g is the fraction of prompt background, f_{++} is the fraction of background in the longer positive exponential tail, and f_- is the fraction of background in the negative exponential tail.

The errors on the background proper decay time are parameterized with two Gamma distributions, as described for the errors on the signal proper decay time.

Background angles ($P_b(\psi), P_b(\theta), P_b(\phi)$)

The PDFs for the background transversity angles are parameterized using empirical functions, based on cos or cos^2 functions of the angles. The coefficients in the functions are allowed to float in the fit.

4.4.3 Fit Variables

We list here all of the variables floating in the unbinned maximum likelihood fit, to simplify the presentation of the fit results. The fit parameters are divided into the physics parameters of interest, and nuisance parameters for which we are not interested in quoting a value.

Physics Parameters

- β_s
- $\Delta\Gamma$
- $\phi_{||}$
- $\phi_\perp$

- α_{CPodd}: related to $|A_\perp|$ (described below)
- $\alpha_{||}$: related to A_0 and $A_{||}$ (described below)
- f_{swave}: fraction of f_0/non-resonant K^+K^-
- δ_S: phase of f_0/non-resonant K^+K^-

Nuisance Parameters

- m: B_s^0 mass
- m_s: mass error scale factor
- N: number of B_s^0 signal events
- B_m: slope of line fitting background mass distribution
- f_s: signal fraction
- $S_{\mathcal{D}}^+$: OST dilution scale factor for B_s^0
- $S_{\mathcal{D}}^-$: OST dilution scale factor for $\bar{B}_s^0$
- $S_{\mathcal{D}}$: SSKT dilution scale factor for B_s^0 and $\bar{B}_s^0$
- $\epsilon_s(OST)$: signal OST tagging efficiency
- $\epsilon_s(SSKT)$: signal SSKT tagging efficiency
- $\epsilon_b(OST)$: background OST tagging efficiency
- $\epsilon_b(SSKT)$: background SSKT tagging efficiency
- $A^+(OST)$: OST background positive tag asymmetry
- $A^+(SSKT)$: SSKT background positive tag asymmetry
- $c\tau$: B_s^0 proper decay time
- Δm_s
- f_p: fraction of background in prompt lifetime peak

- f_-: fraction of background in negative lifetime tail
- f_+: fraction of background in positive lifetime tails
- λ_-: background lifetime for negative lifetime tail
- λ_+: background lifetime for first positive lifetime tail
- λ_{++}: background lifetime for second positive lifetime tail
- sf_1: scale factor for first Gaussian in lifetime resolution
- sf_2: scale factor for second Gaussian in lifetime resolution
- res_{frac}: fraction of first to second Gaussian in lifetime resolution
- ϕ: coefficient for background angle ϕ parameterization
- $\cos\psi$: coefficient for background angle ψ parameterization
- $\cos\theta$: coefficient for background angle θ parameterization

4.4.4 Definition of α_{CPodd} and $\alpha_{||}$

Instead of using the transversity amplitudes in the fit, we use the CP odd and even fractions, α_{CPodd} and $\alpha_{||}$. The relationship between α_{CPodd} and $\alpha_{||}$ and the transversity amplitudes is as follows:

$$a_{||} = \alpha_{||}(1 - \alpha_{CPodd}), \tag{4.40}$$

$$a_{\perp} = \alpha_{CPodd} \tag{4.41}$$

and

$$a_{\perp}^2 + a_{||}^2 + a_0^2 = 0. \tag{4.42}$$

The quantities $a_{||}$, $a_{\perp}$ and a_0 are related to the amplitudes $A_{||}$, $A_{\perp}$ and A_0 by

$$
\begin{aligned}
|A_\perp|^2 &= \frac{|a_\perp|^2 y}{1+(y-1)|a_\perp|^2} & |a_\perp|^2 &= \frac{|A_\perp|^2 y}{y+(1-y)|A_\perp|^2} \\
|A_{||}|^2 &= \frac{|a_{||}|^2}{1+(y-1)|a_\perp|^2} & |a_{||}|^2 &= \frac{|A_{||}|^2 y}{y+(1-y)|A_\perp|^2} \\
|A_0|^2 &= \frac{|a_0|^2}{1+(y-1)|a_\perp|^2} & |a_0|^2 &= \frac{|A_0|^2 y}{y+(1-y)|A_\perp|^2},
\end{aligned}
\tag{4.43}
$$

where y and z are defined as $y = (1-z)/(1+z)$ and $z = \cos 2\beta_s \Delta\Gamma/(2\Gamma)$.

4.4.5 σ_t for signal and background

The coefficients for the Gamma functions describing the signal and background σ_t error distributions are determined using a simplified, lifetime only fit, removing flavor tagging and angular dependences. The parameterization of the Gamma functions is fixed in the main fit, based on the simplified fit results. The fitted values of the free coefficients a and b in the Gamma functions are given in Table. 4.3.

parameter	value
a, signal, 1st Gamma function	9.96 ± 0.56
a, signal, 2nd Gamma function	6.07 ± 0.87
b, signal, 1st Gamma function	0.00018 ± 0.00001
b, signal, 2nd Gamma function	0.00044 ± 0.00004
a, background, 1st Gamma function	10.38 ± 0.16
a, background, 2nd Gamma function	3.70 ± 0.18
b, background, 1st Gamma function	0.000237 ± 0.000004
b, background, 2nd Gamma function	0.00099 ± 0.00003
fraction of 1st to 2nd Gamma function in signal	0.76 ± 0.05
fraction of 1st to 2nd Gamma function in background	0.89 ± 0.01

Table 4.3: Fitted parameters for Gamma functions describing $\sigma(c_t)$

4.4.6 Fitter Validation

The properties of the likelihood function are tested using pseudo-experiments. Pseudo-experiments are generated by sampling the input probability density functions. They

are a useful tool for testing the implementation of the likelihood. Since the pseudo-events are generated by polling the input PDFs, at sufficiently high statistics, statistical fluctuations in the fit should be insignificant and fitted parameters should match the generated parameters. At lower statistics, the difference between the generated and fitted value of a parameter quantifies the fit's ability to measure a parameter's true value. The agreement between the generated and fitted values of parameters is tested by studying *pull* distributions. Each entry in the pull distribution is the record of a pseudo-experiment's value of

$$P = \frac{p_{fit} - p_{input}}{\sigma_p} \tag{4.44}$$

where σ_p is error on the fitted value.

In the high statistics limit, the pull distribution for each parameter in the likelihood function is expected to have a mean of zero (unbiased), and to have a Gaussian distribution with unit width.

The unbiased, Gaussian nature of pulls distributions in the asymptotic limit was exhaustively tested for the fit in previous iterations of this analysis [25]. We do not perform a full re-validation, but we do extensively test the novel extensions to the likelihood, namely the f_0/non-resononant K^+K^- S-wave contribution.

We generate ensembles of 500 pseudo-experiments, each with 100,000 signal events. All of the fit parameters are generated at their fitted value on data, with the exception of the K^+K^- S-wave fraction and phase. These are generated for an array of values: S-wave fractions of 5%, 10%, 25% and 50%, and S-wave phases of 0, $\pi/4$, $\pi/2$ and $3\pi/4$. β_s and $\Delta\Gamma$ were generated at the standard model values of 0.02 rad and 0.096 ps^{-1}, respectively.

An example of the pull distributions for the fit parameters of interest is shown in Fig. 4.16-4.18. The input values for these pseudo-experiments were β_s=0.02 and $\Delta\Gamma$=0.096, S-wave fraction of 25%, and S-wave phase of 0.

The distributions are generally Gaussian and unbiased, with the exception of the distributions for the phases $\phi_{||}, \phi_{\perp}$ and δ_S. However, the deviations from unit width, unbiased pulls is small enough that this effect is considered to be a statistical effect that should vanish for pseudo-experiments with a suitably high number of signal events.

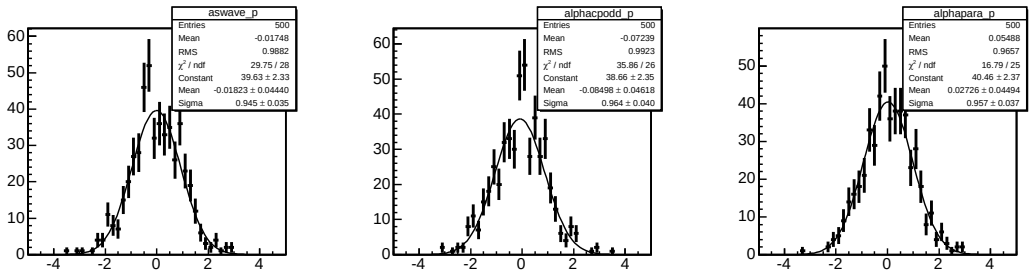

Figure 4.16: Pull distributions for 500 high statistics pseudo-experiments for the S-wave fraction (left), α_{CPodd} (center), and $\alpha_{||}$ (right).

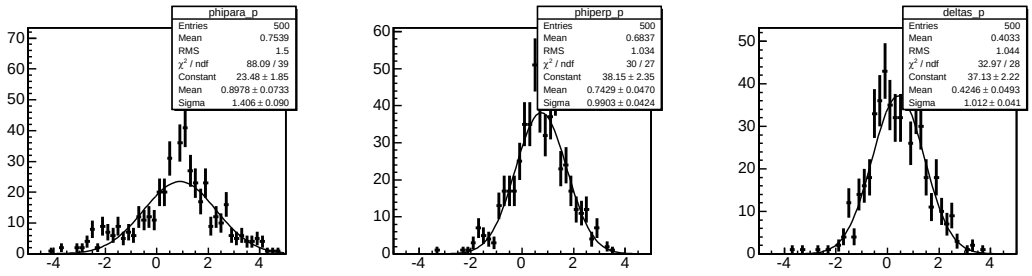

Figure 4.17: Pull distributions for 500 high statistics pseudo-experiments for β_s (left), $\Delta\Gamma$ (center), and ct (right).

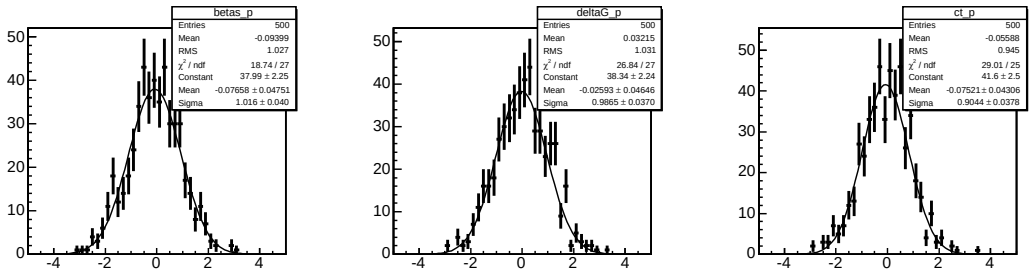

Figure 4.18: Pull distributions for 500 high statistics pseudo-experiments for $\phi_{||}$ (left), $\phi_{\perp}$ (center), and δ_S (right).

It is not the sort of behavior indicative of a problem with the implementation of the likelihood.

We record the pulls means and widths for the parameters of greatest interest, the S-wave fraction and phase, in Tables 4.4-4.7. The pulls for the S-wave fraction are unbiased and of unit width for all input values of the fraction and phase. The pulls for the S-wave phase are biased for some input values. However, the size of the biases, though statistically significant, are small in their absolute magnitude, on order of 0.005.

f_{sw} Mean	5%	10%	25%	50%
0	0.22 ± 0.06	0.13 ± 0.05	-0.02 ± 0.04	-0.05 ± 0.05
$\pi/4$	-0.09 ± 0.11	-0.09 ± 0.06	-0.02 ± 0.05	-0.06 ± 0.05
$\pi/2$	-0.11 ± 0.08	-0.08 ± 0.05	-0.08 ± 0.04	-0.10 ± 0.05
$3\pi/4$	0.02 ± 0.11	-0.07 ± 0.04	-0.15 ± 0.05	-0.09 ± 0.05

Table 4.4: Mean of the S-wave fraction pull.

f_{sw} Width	5%	10%	25%	50%
0	0.83 ± 0.05	0.87 ± 0.04	0.95 ± 0.03	0.99 ± 0.04
$\pi/4$	1.01 ± 0.13	0.93 ± 0.06	0.97 ± 0.04	1.02 ± 0.04
$\pi/2$	0.97 ± 0.09	0.91 ± 0.04	0.95 ± 0.03	0.99 ± 0.04
$3\pi/4$	0.91 ± 0.11	0.89 ± 0.03	0.95 ± 0.04	1.01 ± 0.04

Table 4.5: Width of the S-wave fraction pull.

δ_S Mean	5%	10%	25%	50%
0	0.10 ± 0.06	0.18 ± 0.05	0.42 ± 0.05	-0.47 ± 0.04
$\pi/4$	0.14 ± 0.10	0.24 ± 0.05	0.25 ± 0.04	0.27 ± 0.04
$\pi/2$	-0.06 ± 0.06	0.04 ± 0.04	-0.03 ± 0.04	-0.05 ± 0.05
$3\pi/4$	-0.19 ± 0.11	-0.21 ± 0.05	-0.31 ± 0.04	-0.28 ± 0.04

Table 4.6: Mean of the S-wave relative phase pull.

The fit was also tested for statistics corresponding to the 5.2 fb^{-1} data set, ~6500 signal events, and non-standard model β_s and $\Delta\Gamma$. Toys were generated with β_s of 0.25 and 0.50. $\Delta\Gamma$ was generated according to $\Delta\Gamma = \Gamma_{12}\cos(2\beta_s)$. Input values of the S-wave phase and fraction for these pseudo-experiments were $\pi/2$ and 25%. The pull

δ_S Width	5%	10%	25%	50%
0	0.87 ± 0.06	0.91 ± 0.04	1.01 ± 0.04	0.94 ± 0.03
$\pi/4$	0.96 ± 0.09	0.86 ± 0.04	0.92 ± 0.03	0.93 ± 0.04
$\pi/2$	0.80 ± 0.05	0.87 ± 0.03	0.92 ± 0.04	0.97 ± 0.04
$3\pi/4$	0.98 ± 0.11	0.89 ± 0.04	0.86 ± 0.03	0.82 ± 0.03

Table 4.7: Width of the S-wave relative phase pull.

distributions are given in Appendix C. Their behavior is similar to the that of the pulls generated with 100,000 signal events at the standard model point.

4.4.7 Fit Projections

Fit projections for the proper decay time and proper decay time errors are shown in Figs. 4.19 and 4.20. All fit parameters, other than the proper time or proper time errors, have been integrated over, in order to examine agreement between the data and the fit function for the parameters of interest. The good agreement between the data and the fit projections validates our parameterization of both the signal and background distributions and their errors.

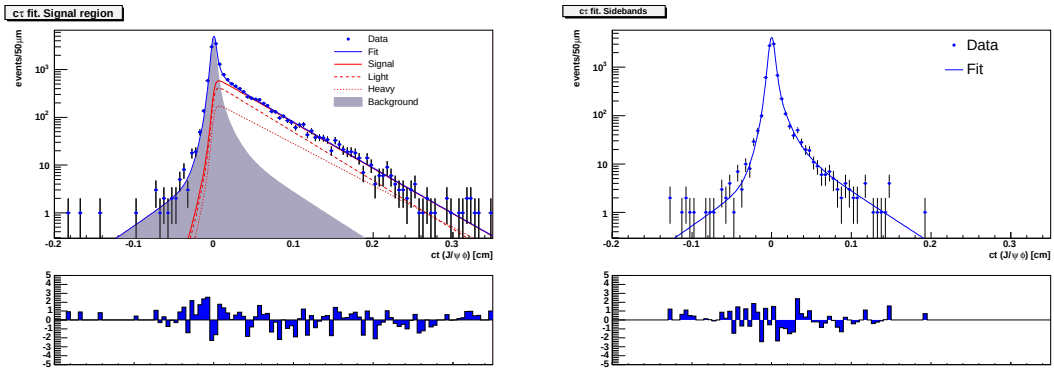

Figure 4.19: Proper decay time fit projections for B_s^0 signal and background regions.

Fit projections for the transversity angles $\cos\theta$, ϕ, and $\cos\psi$ in the signal and side-band regions are shown in Fig. 4.21 and Fig. 4.22.

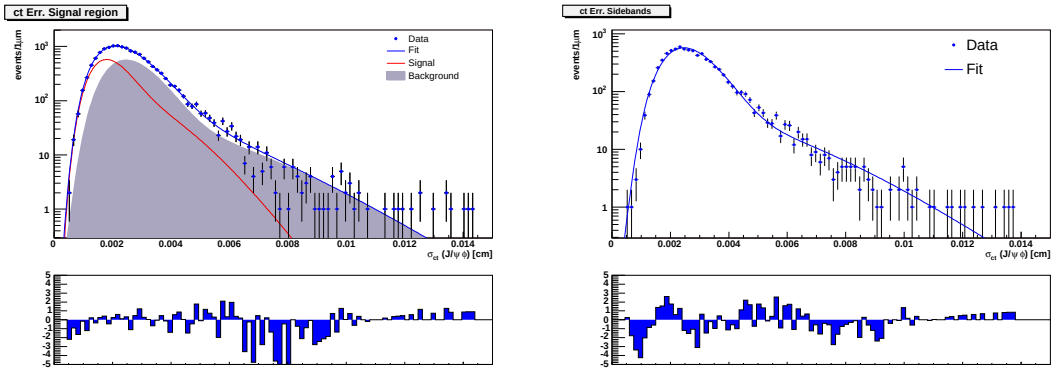

Figure 4.20: Proper decay time error fit projections for B_s^0 signal and background regions.

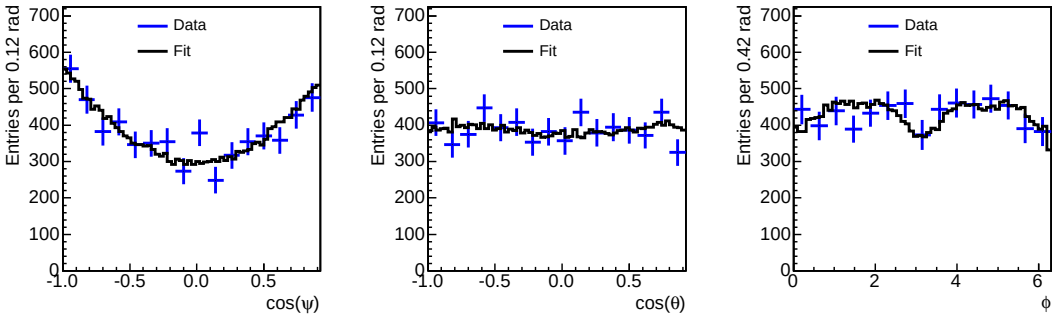

Figure 4.21: Fit projections for transversity angles for side-band subtracted signal.

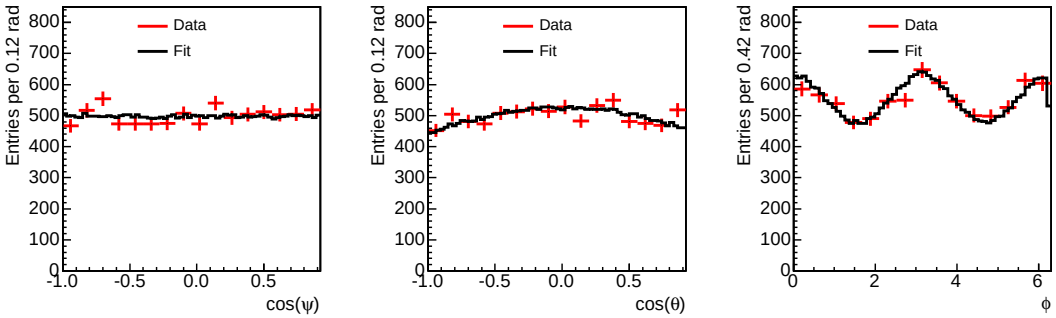

Figure 4.22: Fit projections for transversity angles in background region.

Chapter 5

Results

5.1 KK invariant mass fit

In order to increase confidence in the ability of the unbinned maximum likelihood fit to determine the fraction of f_0/non-resonant K^+K^-, we perform a binned mass fit of the K^+K^- invariant mass to extract f_{swave}. To accomplish this, we use BSTNTUPLEs with a loose selection on the ϕ mass window. These ntuples were only produced for 3.8 fb^{-1} of data. The ϕ mass window is [0.98, 1.20] GeV/c^2, compared to the window [1.009, 1.028] GeV/c^2 typically used in this analysis. The loose selection allows much more f_0/non-resonant K^+K^- to be retained. Since the expected contribution is small, it is advantageous to make the selection as loose as possible to retain a contribution that is large enough to fit.

The wide ϕ mass window introduces a complication in that it increases the amount of non-combinatorial physics background retained, in the form of a B^0 reflection. B^0 reflections refer to $B^0 \to J/\psi K^{*0}$ events where the pion from the K^{*0} decay is misidentified as a kaon in the reconstruction of the $B_s^0 \to J/\psi\phi$ final state. The event is misreconstructed as a B_s^0, but the invariant mass distribution of the B_s^0 and ϕ is distorted by assigning the kaon mass to the pion. Several other reflections are also allowed, but their fractions are negligible for this selection.

Fig. 5.1 shows the fit to the B_s^0 mass peak, using the wide KK mass window. The B^0 reflection is modeled using inclusive $B^0 \to J/\psi X$ Monte Carlo. The fraction of B^0 reflection in the KK mass distribution is fixed based on the fraction of the B^0 reflection relative to the B_s^0 peak as determined from the fit in Fig. 5.1.

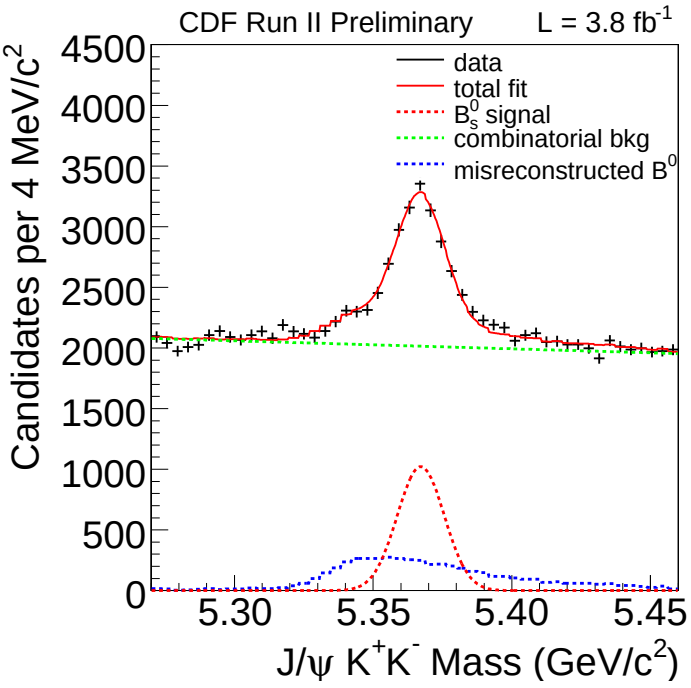

Figure 5.1: Binned mass fit of B_s^0 for wide KK mass window, with B^0 reflection.

The ϕ mass peak is modeled using EVTGEN fully reconstructed Monte Carlo. The Monte Carlo includes the effect of mass resolution. The background is modeled using the B_s^0 mass sidebands. The f_0 and non-resonant K^+K^- states are modeled with a flat line shape. This assumption can be justified in both the wide KK mass window, and the narrow mass window used in the main fit. The non-resonant K^+K^- state has a flat shape by definition, and the shape of the f_0 can be estimated from an analytic function, as shown in Fig. 5.2 and 5.3 for the wide and narrow KK mass window. Since only the tail of the f_0 resonance appears under the ϕ peak, the assumption of a flat f_0 shape is reasonable as well. The variation of the phase across the ϕ mass range is also shown. For the narrow KK mass window, the phase very slowly increasing. This justifies our treatment of δ_S as flat across the ϕ resonance in the main fit.

The fit of the K^+K^- invariant mass distribution, including B^0 reflection, is shown in Fig. 5.4. The f_0/non-resonant K^+K^- fraction was floated in the fit, but no contribution was found. From this study, we conclude that the f_0/non-resonant K^+K^- contributions under the ϕ mass peak are small. The distribution can be modeled successfully using only resonant K^+K^-, combinatorial background, and B^0 reflection.

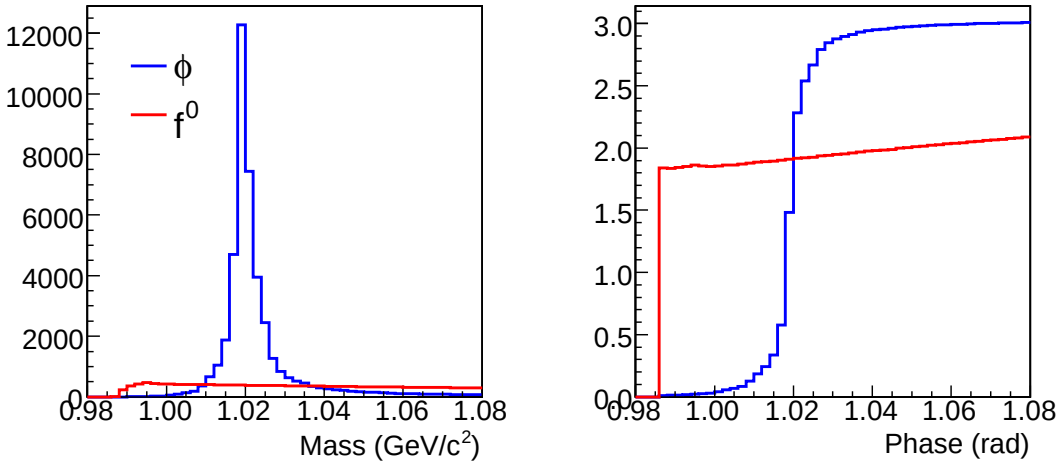

Figure 5.2: In red, f_0 line shape (left) and phase shape (right) for wide KK mass window. In blue, ϕ line shape and phase shape.

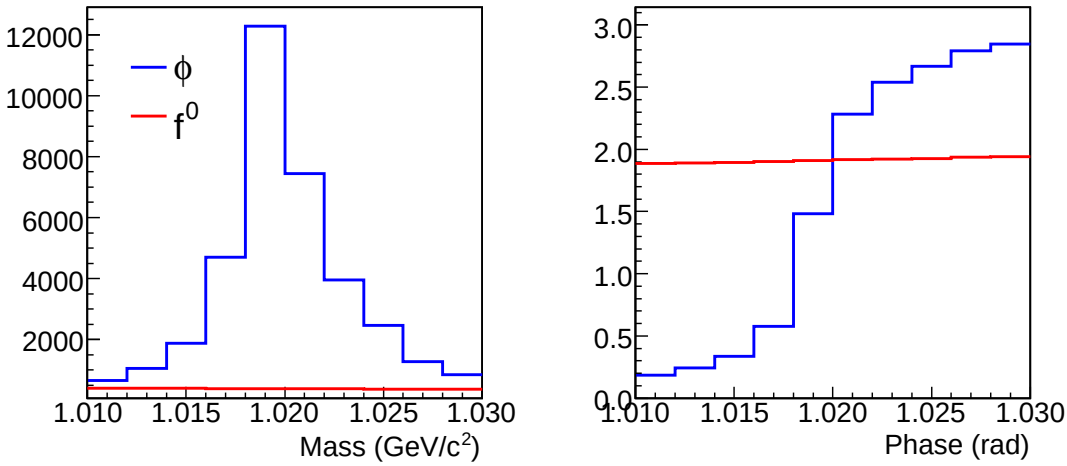

Figure 5.3: f_0 line shape (left) and phase shape (right) for narrow KK mass window. In blue, ϕ line shape and phase shape.

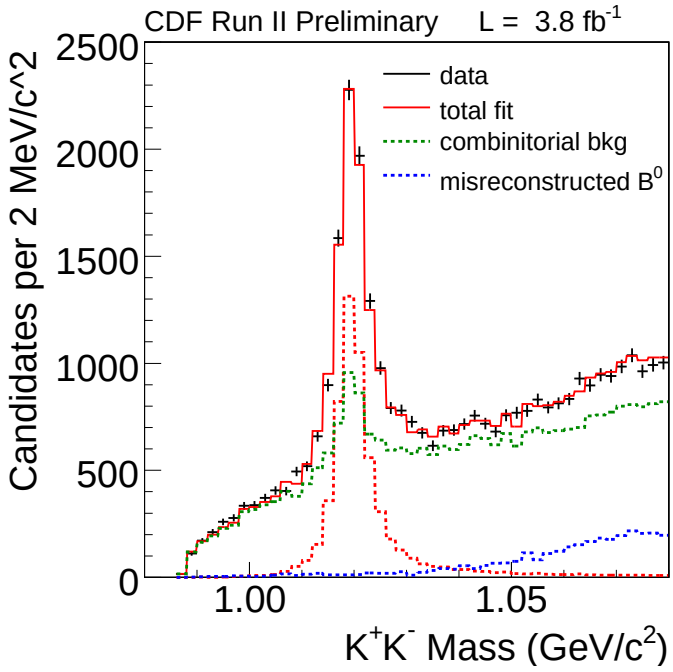

Figure 5.4: Binned mass fit of K^+K^- invariant mass distribution for wide KK mass window, including B^0 reflection.

5.2 Point estimates

In an ideal scenario, we would be able to use the unbinned maximum likelihood fit to determine a precise value for β_s and any other fit parameters of interest. The statistical power of our data sample, though, is limited, and pseudo-experiments generated with statistics corresponding to our data sample show substantial biases in the pull distributions for β_s and other parameters. The pull distributions and residuals ($p_{fit} - p_{input}$) are shown in Fig. 5.5-5.7. These biases are not well understood, and thus we choose to quote confidence regions rather than point estimates (specific values with errors for parameters of interest).

However, the tagged fit with the simplifying assumption of CP violation fixed to zero has also been studied. The fit results on data, with β_s fixed to the zero, are shown in Table 5.1. For this fit, the pulls for many of the fit parameters are well-behaved, making it possible to quote point estimates. We review the pull studies that show which parameters can be assigned a point estimate, and the determination of systematic errors.

5.2.1 Pull studies for tagged fit at SM point

To test the behavior of the likelihood in our statistical regime, we generate pseudo-experiments and study the pull distributions of fit parameters for which we would like to obtain point estimates. For the sake of simplicity, we could use the fitted values of parameters from the data fit to generate the pseudo-experiments. This method has the draw-back that it does not probe the full parameter space. To conservatively determine the behavior of a parameter's pull, we must probe its full parameter space, including possibly problematic extreme values of the parameter. We accomplish this by randomizing the generated values for the parameters within a broad range. The ranges the parameters are randomized within are shown in Table 5.2.

Randomization is only done for parameters that have well-behaved non-randomized pulls. In the case of the strong phases, the pulls are biased when the pseudo-experiments are generated with the fitted values for data, but not at other points in their range. Randomizing the inputs in this case could make the biases less apparent, so we only generated at the fitted data values for the strong phases. The other parameters in the

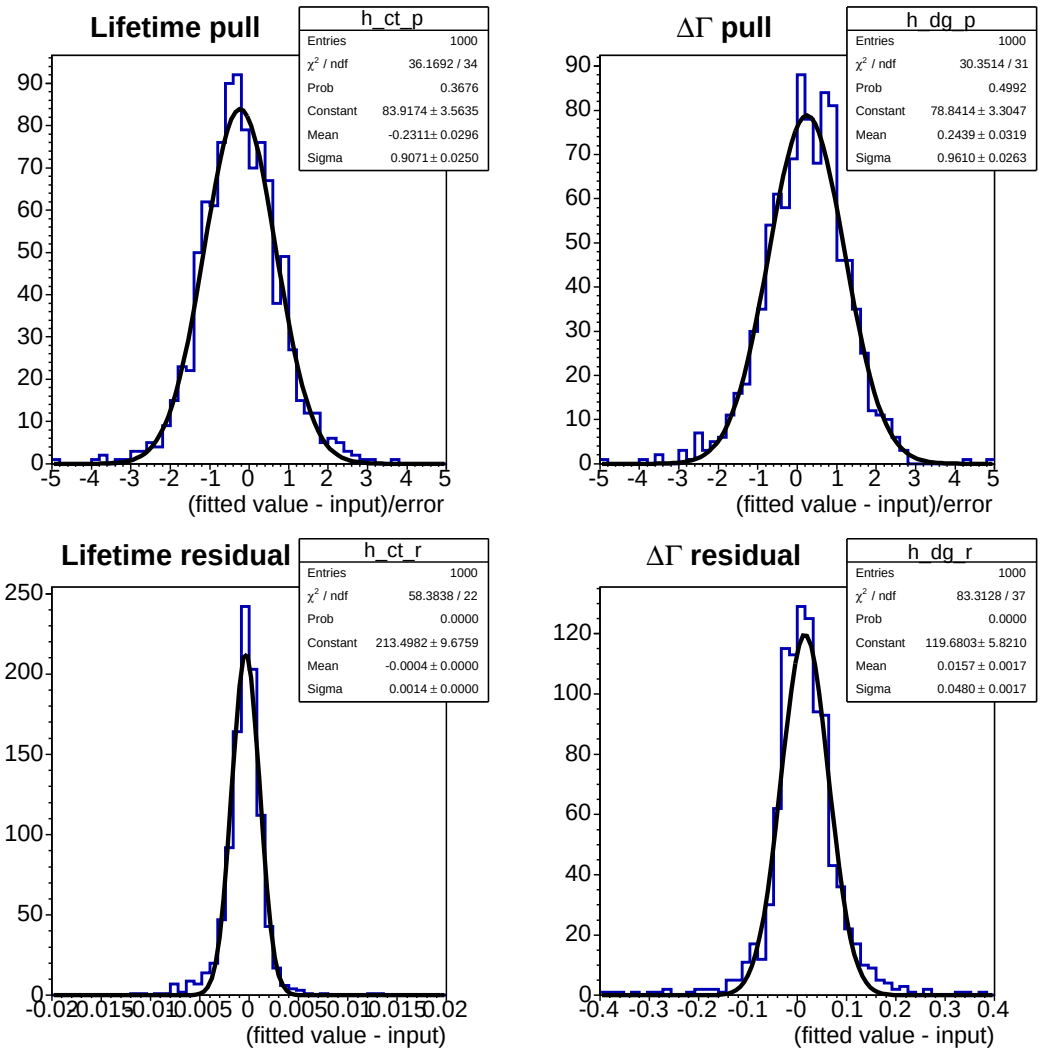

Figure 5.5: Pulls (top) and residuals (bottom) for cτ and $\Delta\Gamma$, with β_s floating in fit.

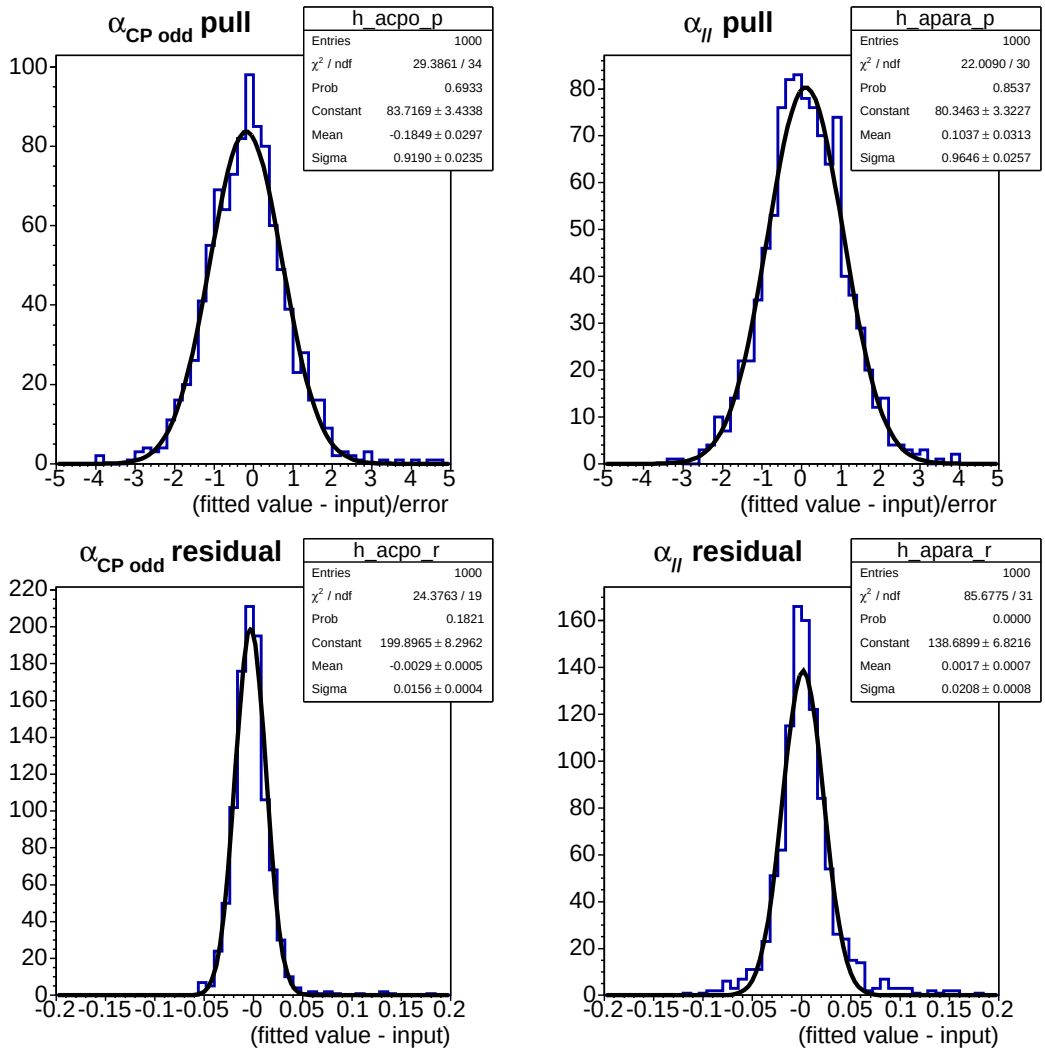

Figure 5.6: Pulls (top) and residuals (bottom) for α_{CPodd} and $\alpha_{||}$, with β_s floating in fit.

Parameter	Value
$\Delta\Gamma$	0.075 ± 0.035
$\phi_{\|\|}$	3.08 ± 0.63
$\phi_{\perp}$	2.95 ± 0.64
α_{CPodd}	0.266 ± 0.014
$\alpha_{\|\|}$	0.306 ± 0.015
f_{swave}	0.019 ± 0.027
δ_S	1.37 ± 0.77
m	5.37 ± 0.00
m_s	1.73 ± 0.02
N	6.5e+03 ± 85
B_m	-2.3 ± 0.6
f_s	0.181 ± 0.002
$S^+_{\mathcal{D}}$	0.92 ± 0.09
$S^-_{\mathcal{D}}$	1.12 ± 0.09
$S_{\mathcal{D}}$	0.901 ± 0.175
$\epsilon_s(OST)$	0.943 ± 0.003
$\epsilon_s(SSKT)$	0.522 ± 0.007
$\epsilon_b(OST)$	0.87 ± 0.00
$\epsilon_b(SSKT)$	0.719 ± 0.003
$A^+(OST)$	0.495 ± 0.003
$A^+(SSKT)$	0.496 ± 0.003
$c\tau$	0.0459 ± 0.0008
Δm_s	17.7 ± 0.1
f_p	0.884 ± 0.005
f_-	0.173 ± 0.034
f_+	0.662 ± 0.045
λ_-	0.038 ± 0.004
λ_+	0.041 ± 0.004
λ_{++}	0.011 ± 0.001
sf_1	1.27 ± 0.01
sf_2	3.32 ± 0.19
res_{frac}	0.882 ± 0.012
ϕ	0.139 ± 0.008
$\cos\psi$	0.004 ± 0.020
$\cos\theta$	0.161 ± 0.018

Table 5.1: Fit results for unbinned maximum likelihood fit with β_s fixed to 0.0 (the SM point).

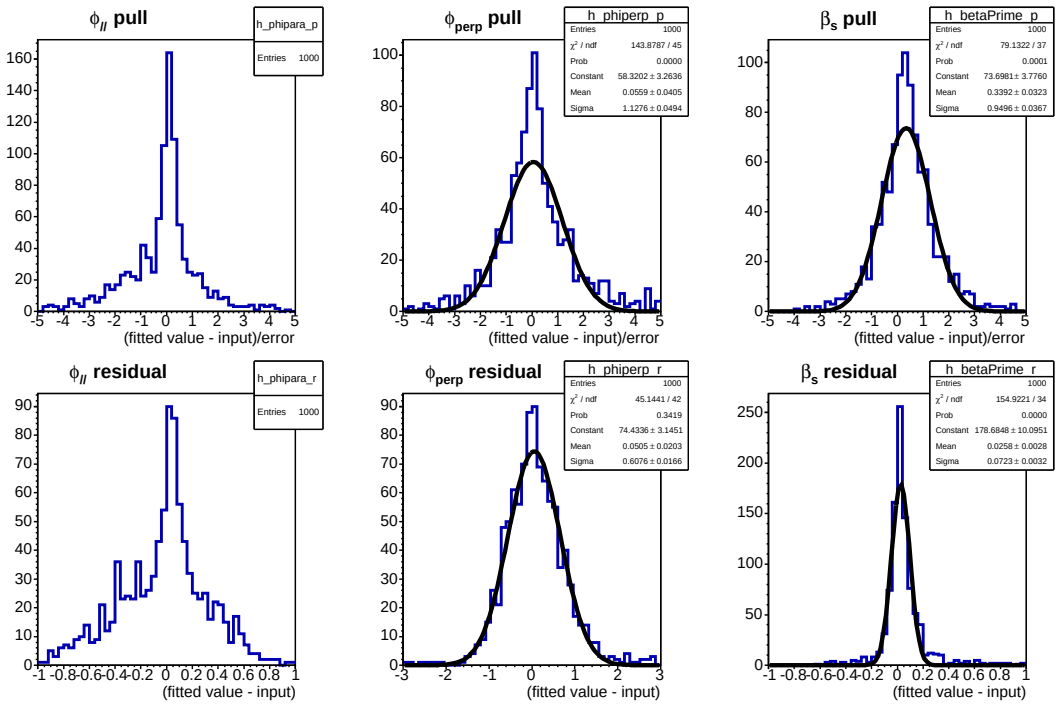

Figure 5.7: Pulls (top) and residuals (bottom) for $\phi_{||}$, $\phi_{\perp}$ and β_s with β_s floating in fit.

fit, for which we are not interested in quoting point estimates, are generated with the fitted values from data.

The pulls from ∼1000 pseudo-experiments with the number of signal events fitted from the data, and randomized inputs for $c\tau$, $\Delta\Gamma$, and the transversity amplitudes are shown in Fig. 5.8-5.10. The residuals, $p_{fit} - p_{input}$ are also shown. The pull means and widths are shown in Table 5.3. The pulls are Gaussian distributions with unit width, except in the case of $\phi_{||}$. The pathology of the $\phi_{||}$ pull distribution has been investigated. It is due to the fact that $\phi_{||}$ has a symmetry axis at π, and the fitted value

Parameter	Minimum	Maximum
$c\tau$	$400\mu m$	$500\mu m$
$\Delta\Gamma$	0.0ps^{-1}	0.6ps^{-1}
α_{CPodd}	0.0%	90.0%
$\alpha_{||}$	0.0%	90.0%

Table 5.2: Ranges for randomized inputs for generating pseudo-experiments.

Parameter	Pull mean	Pull σ
$c\tau$	0.05 ± 0.03	0.99 ± 0.03
$\Delta\Gamma$	-0.08 ± 0.03	1.00 ± 0.03
α_{CPodd}	-0.15 ± 0.03	0.90 ± 0.02
$\alpha_{\|\|}$	0.15 ± 0.03	0.88 ± 0.03
$\phi_{\|\|}$	non-Gaussian	
$\phi_{\perp}$	0.01 ± 0.02	0.87 ± 0.03

Table 5.3: Pulls means and width for fit with β_s fixed to zero.

in the data is close to the symmetry axis. Pseudo-experiments are therefore generated close to π, and the fit tends to have difficulty distinguishing between the $\phi_{||}$ solutions above and below π, and will in many cases simply return a fitted value of exactly π, the median value between the two solutions. This causes a large a spike in the $p_{fit} - p_{input}$ distribution, which in turn makes the pull non-Gaussian.

Due to the non-Gaussian nature of the pull, we do not quote a point estimate for $\phi_{||}$. The pulls for $c\tau$, $\Delta\Gamma$, the transversity amplitudes, and $\phi_{\perp}$ are well-behaved Gaussians, allowing us to quote point estimates for these parameters. From Table 5.3, it is evident that $\Delta\Gamma$ and the transversity amplitudes have greater than 2σ biases. These biases must be accounted for by subtracting them off the quoted value of the parameters.

5.2.2 Systematics

After assessing which parameters we can quote point estimates for, it is necessary to assess the relevant systematic uncertainties. This is again done using pseudo-experiments. We fit 1000 pseudo-experiments generated with β_s fixed to zero, with and without systematic variations. The pseudo-experiments are fit using the default fit, without modifications. The systematic variations are made one by one. The difference between the mean of the residual $(p_{fit} - p_{input})$ without systematics and the mean of the residual with systematics is calculated, and assigned as the systematic.

- Signal angular efficiency

  The systematic uncertainty associated with the modeling of angular efficiency for signal events is calculated using histograms of realistic Monte Carlo to generate the angular efficiency in pseudo-experiments, rather than using the standard fit

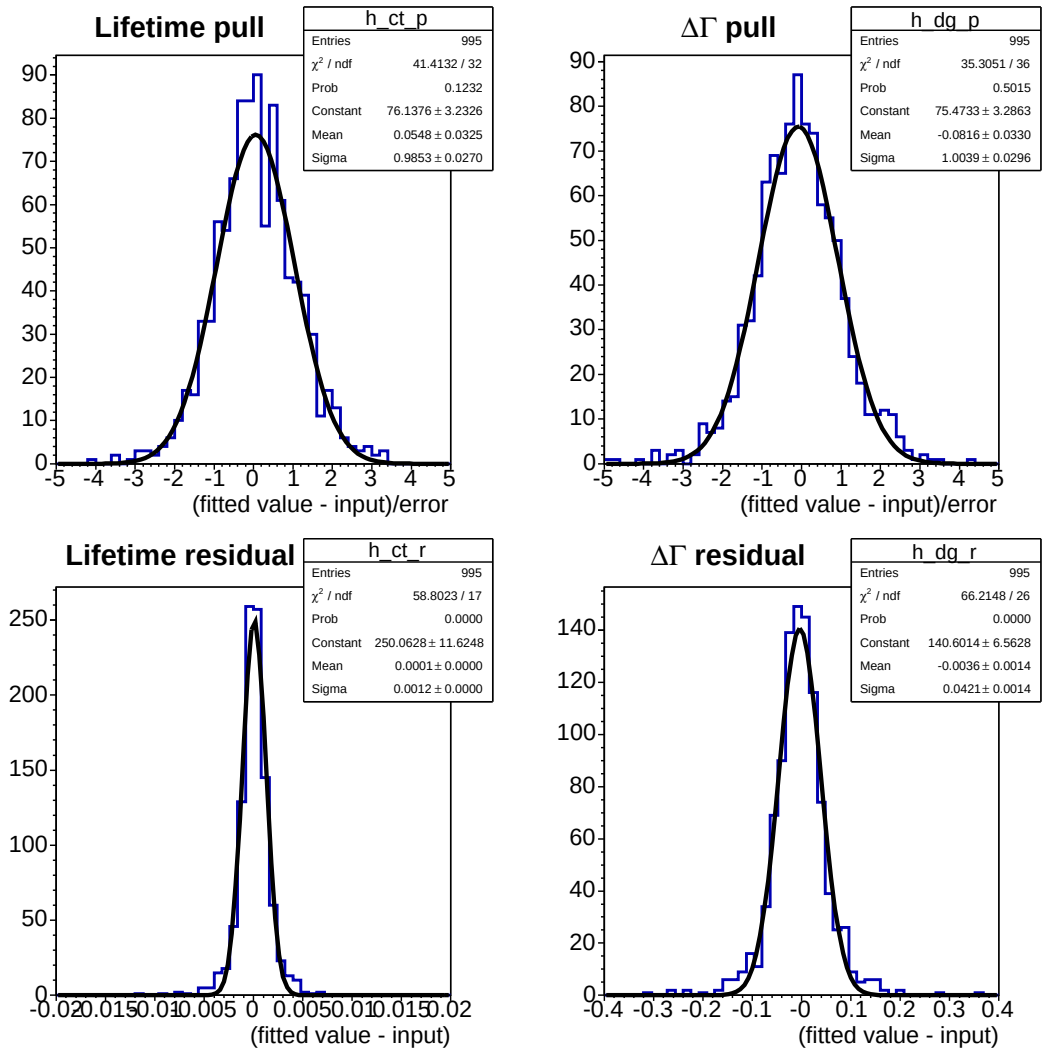

Figure 5.8: Pulls (top) and residuals (bottom) for $c\tau$ and $\Delta\Gamma$, with β_s fixed to zero.

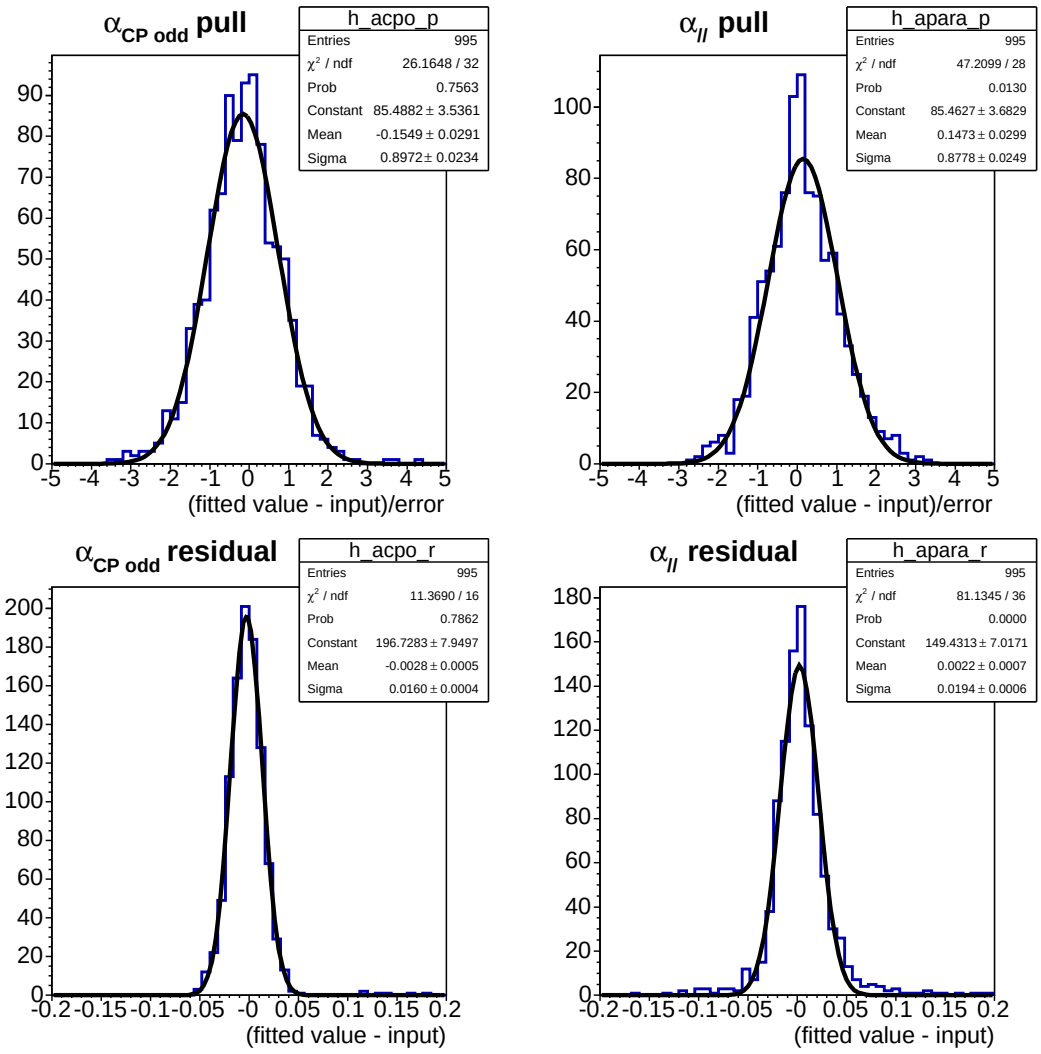

Figure 5.9: Pulls (top) and residuals (bottom) for α_{CPodd} and $\alpha_{||}$, with β_s fixed to zero.

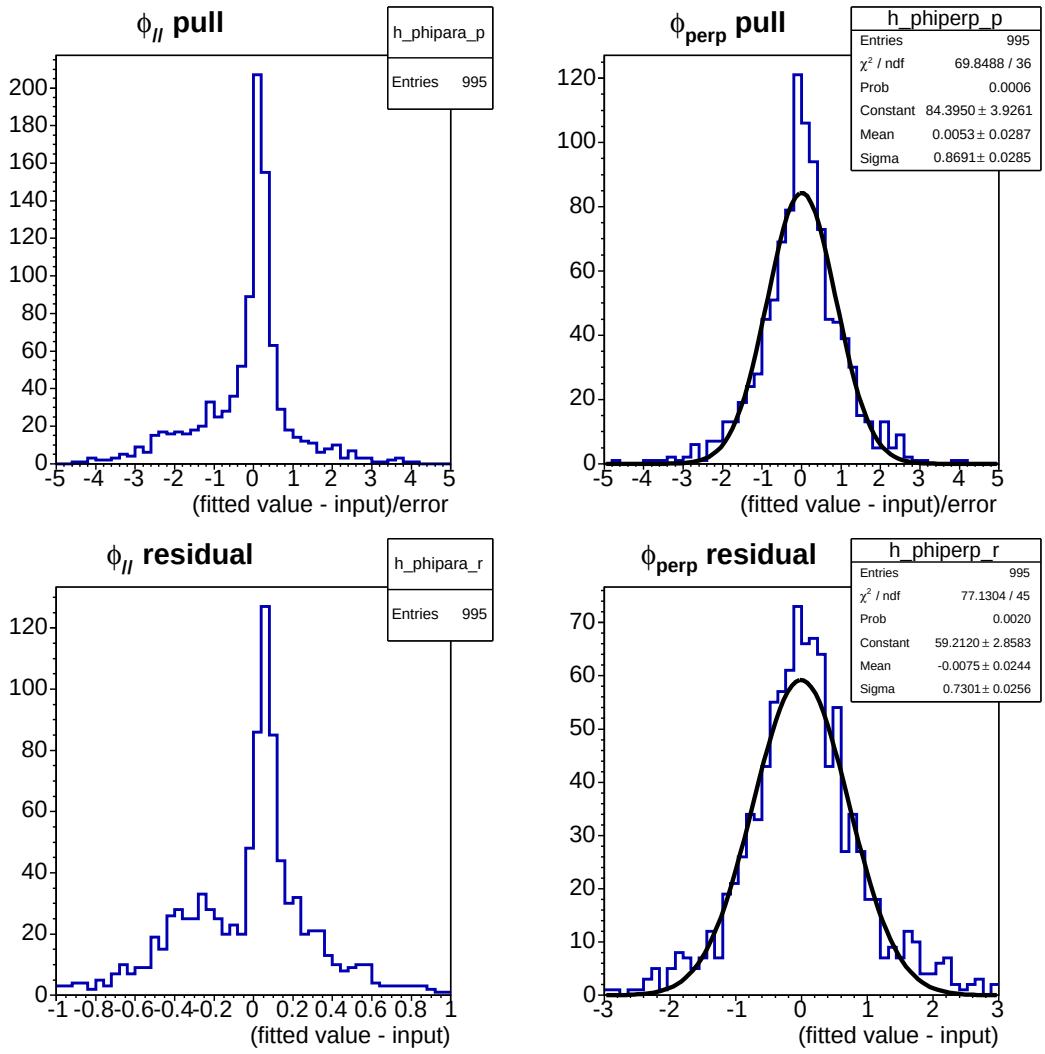

Figure 5.10: Pulls (top) and residuals (bottom) for $\phi_{||}$ and $\phi_{\perp}$, with β_s fixed to zero.

parameterization as an expansion of spherical harmonics. The effect is evaluated both for re-weighted and non-reweighted Monte Carlo.

- Signal mass model

  The systematic uncertainty associated with the signal mass model is calculated by fitting the B_s^0 signal in the data with two Gaussians, and then generating pseudo-experiments with the Gaussian widths and scale factor between the two Gaussians found in the data. The pseudo-experiments are then fit using the usual single Gaussian mass model.

- Background mass model

  The systematic uncertainty associated with the background mass model is calculated by fitting the B_s^0 sideband invariant mass distribution in data with a second order polynomial, and then generating pseudo-experiments with according to the fitted second order polynomial. The pseudo-experiments are then fit with the usual first order polynomial.

- Lifetime resolution model

  The systematic uncertainty associated with the lifetime resolution model (for the signal) is calculated by generating pseudo-experiments with a three Gaussian resolution model. Two of the Gaussian widths and scale factors are taken from the data, and the third is assigned an intermediate width between the two. The pseudo-experiments are then fit with the usual two Gaussian resolution model.

- Background lifetime model

  The systematic uncertainty associated with the background lifetime model is assessed by generating pseudo-experiments using lifetime distributions from the B_s^0 invariant mass side bands.

- Angular background model

  The systematic uncertainty associated with the background angular model is assessed by generating pseudo-experiments using angular distributions from the B_s^0 invariant mass side bands. Additionally, the background angles are assumed to be independent in the likelihood, but small correlations between $\cos(\theta)$ and ϕ were

observed. In addition, correlations between the angles and $\sigma(c\tau)$ were observed. The effect of the correlation between $\cos(\theta)$ and ϕ is assessed by binning ϕ in three bins of $\cos(\theta)$. Then ϕ is generated from the histograms corresponding to the generated value of $\cos(\theta)$. The correlation between the angles and $\sigma(c\tau)$ are handled similarly, with the angles binned in $\sigma(c\tau)$. The generated values of the angles are sample from the bin dictated by the generated value of $\sigma(c\tau)$.

- B^0 crossfeed

  The systematic effect of B^0 reflections from mis-reconstructed $B^0 \to J/\psi K^{*0}$ decays is handled by generating pseudo-experiments with some B^0 reflection included. The fraction of B^0 reflection is calculated to be 2.2% from $B^0 \to J/\psi K^{*0}$ Monte Carlo with our selection requirements applied.

- SVX alignment

  In accordance with other lifetime measurements at CDF, we assign a 2μm systematic uncertainty to the B_s^0 lifetime measurement to account for SVX misalignment. We generate pseudo-experiments with the lifetime varied by $\pm 2\mu$m to assign an SVX alignment uncertainty to the other parameters.

- Mass errors

  The systematic uncertainty associated with the modeling of mass errors is assessed by generating pseudo-experiments with the mass errors taken from the sideband subtracted B_s^0 signal region for signal events, and the B_s^0 sideband region for background events. The pseudo-experiments are fit with the default assumption that the signal and background mass error distributions are identical.

- cτ errors

  The systematic uncertainty associated with the modeling of cτ errors is assessed by generating pseudo-experiments with the cτ errors taken from the sideband subtracted B_s^0 signal region for signal events, and the B_s^0 sideband region for background events. A small observed correlated between $\sigma(c\tau)$ and the mass is assigned a systematic uncertainty by using two cτ histograms for generation, one for the upper and one for the lower B_s^0 sideband. If the generated mass falls in

the lower sideband, the latter histogram is sampled, and if the generated mass falls in the upper sideband, the former is sampled.

The total systematic uncertainty for each parameter for which a point estimate is being quoted is shown in Table 5.4. The final results, with systematic uncertainties assigned, are

$$\begin{aligned} c\tau_s &= 458.6 \pm 7.5 \text{ (stat.)} \pm 3.6 \text{ (syst.) } \mu\text{m} \\ \Delta\Gamma &= 0.075 \pm 0.035 \text{ (stat.)} \pm 0.010 \text{ (syst.) } ps^{-1} \\ |A_{||}(0)|^2 &= 0.231 \pm 0.014 \text{ (stat)} \pm 0.015 \text{ (syst.)} \\ |A_0(0)|^2 &= 0.524 \pm 0.013 \text{ (stat)} \pm 0.015 \text{ (syst.)} \\ \phi_\perp &= 2.95 \pm 0.64 \text{ (stat)} \pm 0.07 \text{ (syst.)}. \end{aligned} \tag{5.1}$$

The measurements of $c\tau$ and $\Delta\Gamma$ are precise and in good agreement with the world average. The transversity amplitudes are in good agreement with those measured in the B^0 system.

5.3 Likelihood scans

We now move to fitting with β_s floating in the fit, rather than being fixed to zero. The fit results on data for the unbinned maximum likelihood fit with all parameters floating is shown are Table 5.5.

Biases are observed in the pulls for pseudo-experiments fit with β_s floating. Rather than quoting point estimates, we produce a confidence region in the $\beta_s - \Delta\Gamma$ plane. A likelihood scan is produced by fixing β_s and $\Delta\Gamma$ for each point in a grid, and minimizing the remaining fit parameters. In previous iterations of the analysis, an exact symmetry existed in the likelihood that we attempted to remove by restricting the strong phase $\phi_{||}$ to be between 0 and π, or between π and 2π. A local minimum would still exist in addition to the global one in both these cases. This was due to an approximate symmetry in the likelihood which produced the local minimum. We accounted for this effect by symmetrizing the contour, reflecting the contour over the symmetry axes and taking the deeper of the two minima at each point on the grid.

Systematic	$\Delta\Gamma$	$c\tau_s$	$\lvert A_{\lvert\lvert}(0)\rvert^2$	$\lvert A_0(0)\rvert^2$	$\phi_\perp$
Signal efficiency:					
Parameterization	0.0024	0.96	0.0076	0.008	0.016
MC reweighting	0.0008	0.94	0.0129	0.0129	0.022
Signal mass model	0.0013	0.26	0.0009	0.0011	0.009
Background mass model	0.0009	1.4	0.0004	0.0005	0.004
Resolution model	0.0004	0.69	0.0002	0.0003	0.022
Background lifetime model	0.0036	2.0	0.0007	0.0011	0.058
Background angular distribution:					
Parameterization	0.0002	0.02	0.0001	0.0001	0.001
$\sigma(c\tau)$ correlation	0.0002	0.14	0.0007	0.0007	0.006
Non-factorization	0.0001	0.06	0.0004	0.0004	0.003
$B^0 \to J\psi K^*$ crossfeed	0.0014	0.24	0.0007	0.0010	0.006
SVX alignment	0.0006	2.0	0.0001	0.0002	0.002
Mass error	0.0001	0.58	0.0004	0.0004	0.002
$c\tau$ error	0.0012	0.17	0.0005	0.0007	0.013
Pull bias	0.0028		0.0013	0.0021	
Totals	0.01	3.6	0.015	0.015	0.07

Table 5.4: Total systematic uncertainties for each parameter where a point estimate is determined.

With the inclusion of the S-wave in the fit, the exact symmetry under $\Delta\Gamma$ to -$\Delta\Gamma$ and $\phi_{||}$ to 2π-$\phi_{||}$ is broken. We find a slightly deeper minimum for the case where $\Delta\Gamma$ is negative and $\phi_{||}$ is greater than π. The difference between the two minima is not significant (0.04 units in the $-2\log\mathcal{L}$), so we cannot claim to resolve the ambiguity between the minima.

With the exact symmetry broken, we are no longer able to symmetrize the contour by hand. Instead we run the likelihood scan with $\phi_{||}$ restricted between 0 and π and between π and 2π. At each point in the $\beta_s - \Delta\Gamma$ plane, we take the deeper of the two $-2\log\mathcal{L}$'s corresponding to the different $\phi_{||}$ ranges. This procedure guarantees that we use the global, not the local minimum at each grid point. The contour is shown in Fig. 5.11.

One dimensional likelihood scans for β_s, $\Delta\Gamma$, $c\tau$, $\phi_{||}$, $\phi_\perp$ and the f_0/non-resonant

Parameter	Value
β_s	0.244 ± 0.132
$\Delta\Gamma$	0.097 ± 0.035
$\phi_{\|\|}$	3.02 ± 0.47
$\phi_\perp$	3.03 ± 0.52
α_{CPodd}	0.264 ± 0.014
$\alpha_{\|\|}$	0.307 ± 0.015
f_{swave}	0.018 ± 0.023
δ_S	1.5 ± 0.6
m	5.37 ± 0.00
m_s	1.73 ± 0.02
N	6.5e+03 ± 85
B_m	-2.3 ± 0.6
f_s	0.181 ± 0.002
$S^+_{\mathcal{D}}$	0.92 ± 0.08
$S^-_{\mathcal{D}}$	1.12 ± 0.09
$S_{\mathcal{D}}$	0.886 ± 0.174
$\epsilon_s(OST)$	0.943 ± 0.003
$\epsilon_s(SSKT)$	0.522 ± 0.007
$\epsilon_b(OST)$	0.870 ± 0.002
$\epsilon_b(SSKT)$	0.719 ± 0.003
$A^+(OST)$	0.495 ± 0.003
$A^+(SSKT)$	0.496 ± 0.003
$c\tau$	0.0459 ± 0.0007
Δm_s	17.7 ± 0.1
f_p	0.884 ± 0.005
f_-	0.172 ± 0.033
f_+	0.661 ± 0.045
λ_-	0.038 ± 0.004
λ_+	0.041 ± 0.004
λ_{++}	0.011 ± 0.001
sf_1	1.27 ± 0.01
sf_2	3.32 ± 0.19
res_{frac}	0.882 ± 0.012
ϕ	0.139 ± 0.008
$\cos\psi$	0.004 ± 0.020
$\cos\theta$	0.161 ± 0.018

Table 5.5: Fit results for unbinned maximum likelihood fit with β_s floating.

K^+K^- fraction fraction and phase are shown in Figs. 5.11-5.19. Fig. 5.18 zooms in on the area around the minimum in the f_0/non-resonant K^+K^- fraction likelihood scan. The S-wave fraction is observed to be small; consistent with zero. Its range is [0, 6.7] at the 95% confidence level. It is clear from the likelihood scan for δ_S that the fit does not have sensitivity to this parameter; the minimum for δ_S is extremely shallow.

In the likelihood scan, the 68% and 95% confidence levels (CL) are always shown in blue and red, respectively. *Confidence level* refers to the number of trials, out of one hundred, for which one expects a condition to be satisfied. In our case, we expect the region inside the 95% CL contour to have at least a 95% chance of enclosing the true β_s and $\Delta\Gamma$. However, our likelihood is complicated by pathological behaviors. As stated before, pull distributions for the fit with β_s floating are not well-behaved. As one can see from the β_s likelihood scan in Fig. 5.12, the minima are not Gaussian. Thus, the canonical values that one must go up from the minimum in $-2\log\mathcal{L}$ for the 68% and 95% confidence levels in fact no longer guarantee 68% and 95% confidence. The non-Gaussianity must be accounted for using a coverage adjustment procedure described in the next section, to find the true 68% and 95% confidence levels.

The contour for the fit for direct CP violation is shown in Fig. 5.20. The likelihood profile is consistent with $|\lambda|$=1 within 2σ. The results are consistent with no direct CP violation, or, in the case of a more plausible source of asymmetry, no asymmetry is evident in the flavoring tagging. The $\cos 2\beta_s$-$\sin 2\beta_s$ contours are also in good agreement with the contours in Fig. 5.11.

5.4 Fit for Direct *CP* violation

We check the robustness of our result with an independent fit which allows $\cos(2\beta_s)$ and $\sin(2\beta_s)$ to float separately. We do not impose the requirement that $|\lambda_f| = 1$, which allows for the possibility of measuring non-zero direct CP violation, either with $|\bar{A}_f/A_f| \neq 1$, or $|p/q| \neq 1$. Although CP is not expected to be violated directly in this system, the fit serves as a useful cross check of the default fit, and also as a cross check of the tagging calibration. Mis-calibration of the tagging algorithms would mimic an asymmetry in the B_s^0 decay rate that would cause $|\lambda_f| \neq 1$. If we observe signs of direct CP violation, the most probable scenario is that there is an asymmetry in a tagger that has not been accounted for properly.

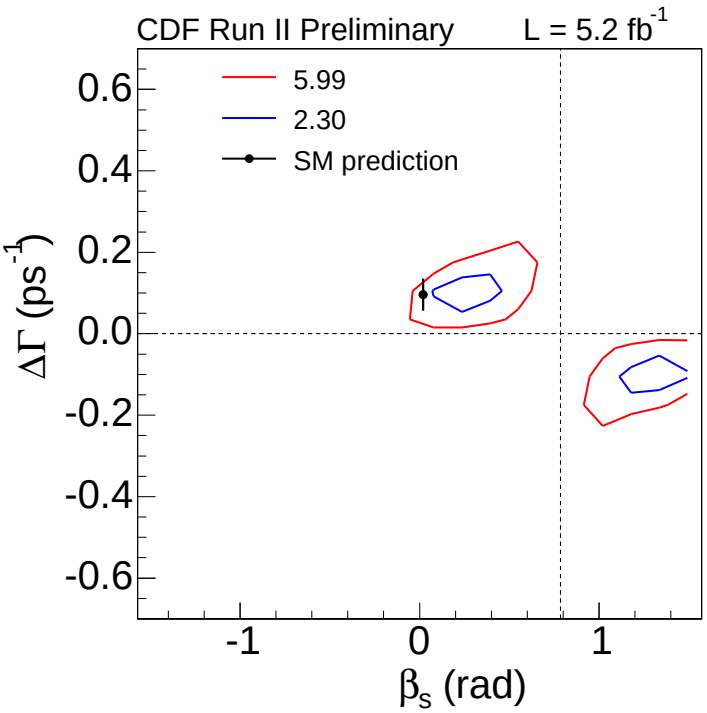

Figure 5.11: Likelihood scan for β_s and $\Delta\Gamma$. The dotted lines are the symmetry axes for the strong phases.

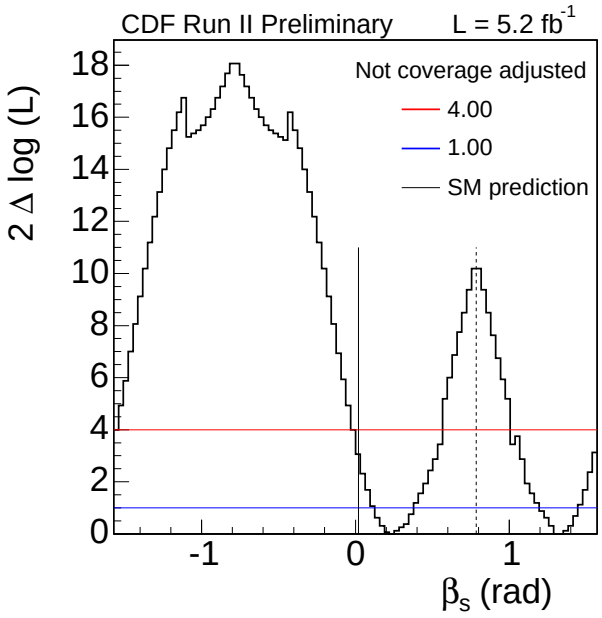

Figure 5.12: Likelihood scan for β_s.

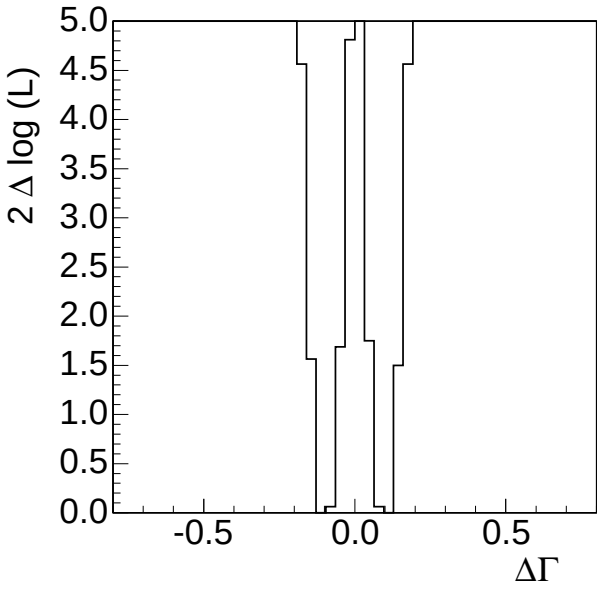

Figure 5.13: Likelihood scan for $\Delta\Gamma$.

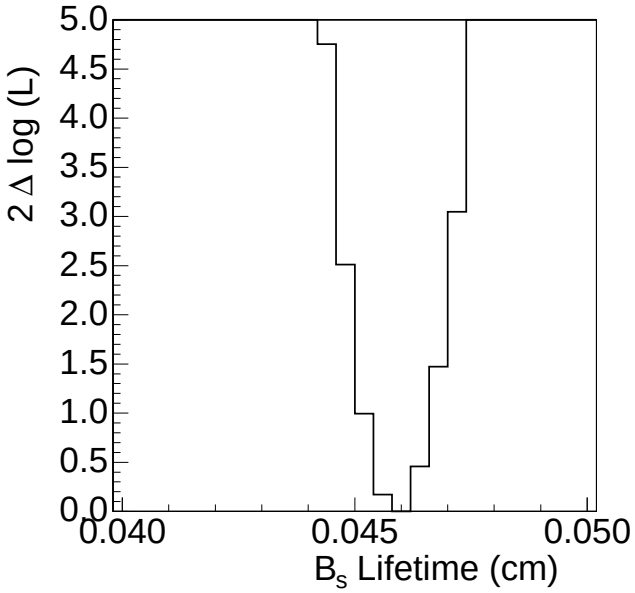

Figure 5.14: Likelihood scan for ct.

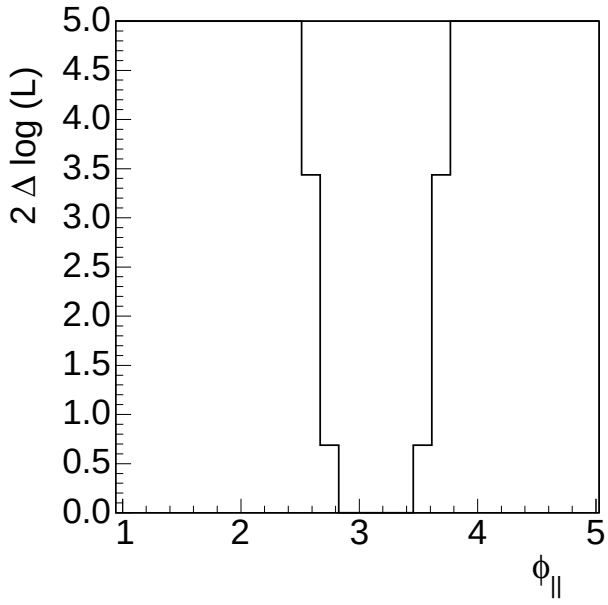

Figure 5.15: Likelihood scan for $\phi_{||}$.

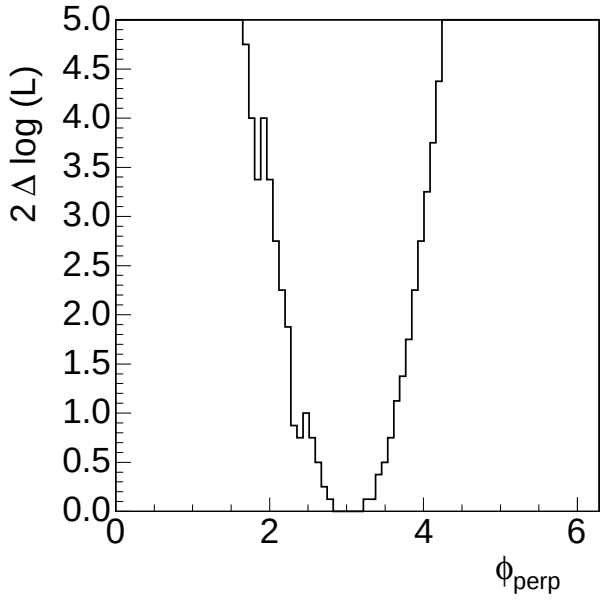

Figure 5.16: Likelihood scan for $\phi_{\perp}$.

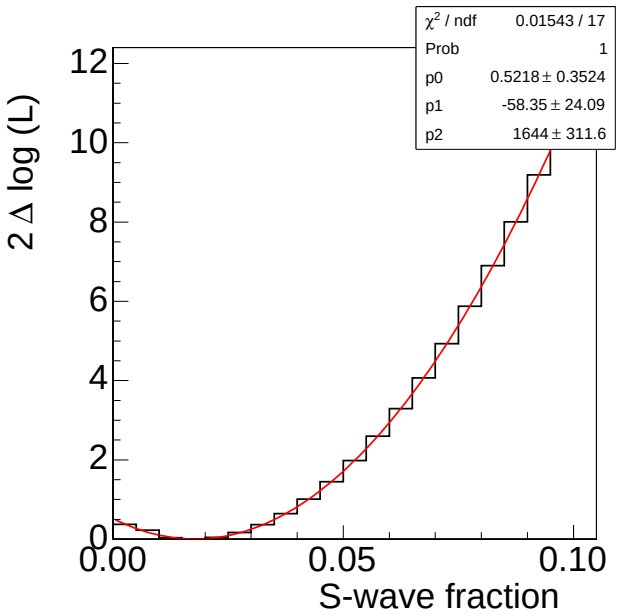

Figure 5.17: Likelihood scan for f_0/non-resonant K^+K^- fraction.

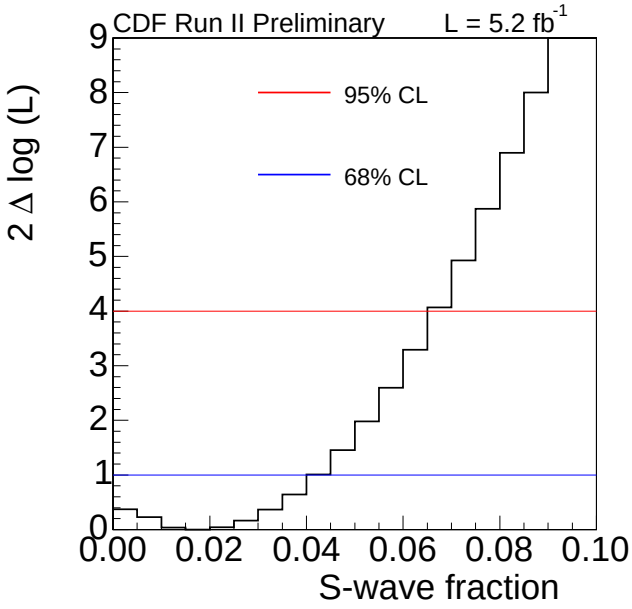

Figure 5.18: Magnified version of f_0/non-resonant K^+K^- fraction likelihood scan. The blue line is the 68% confidence level, the green line is the 95% confidence level.

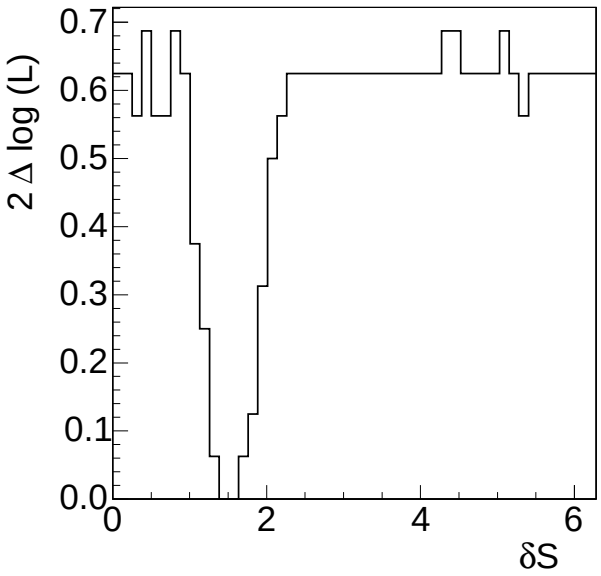

Figure 5.19: Likelihood scan for δ_S.

Allowing $\sin(2\beta_s)$ and $\cos(2\beta_s)$ independence from each other changes both the time-dependent terms and the analytic normalization of the B_s^0 and $\bar{B}_s^0$ PDFs. Note that we do not use the version of the PDF that includes the contributions of f_0 and non-resonant K^+K^-. The normalization and time-dependence allowing direct CP violation can be obtained from a straight-forward expansion of terms containing β_s to use $\sin(2\beta_s)$ and $\cos(2\beta_s)$ instead. We test the implementation of the likelihood using pseudo-experiments. The tests are described in Appendix D, and the implementation was found to be valid.

In order to address concerns that the inclusion of the f_0/non-resonant K^+K^- contribution substantially alters the shape of the $\beta_s - \Delta\Gamma$ contour, rendering previous measurements invalid, we revert to the original likelihood. The overlay between the $\beta_s - \Delta\Gamma$ contours with the original and updated likelihoods is shown in Fig. 5.21. Clearly, the effect of adding the f_0/non-resonant K^+K^- contribution is marginal, as expected, given that the fitted fraction f_{swave} is consistent with zero.

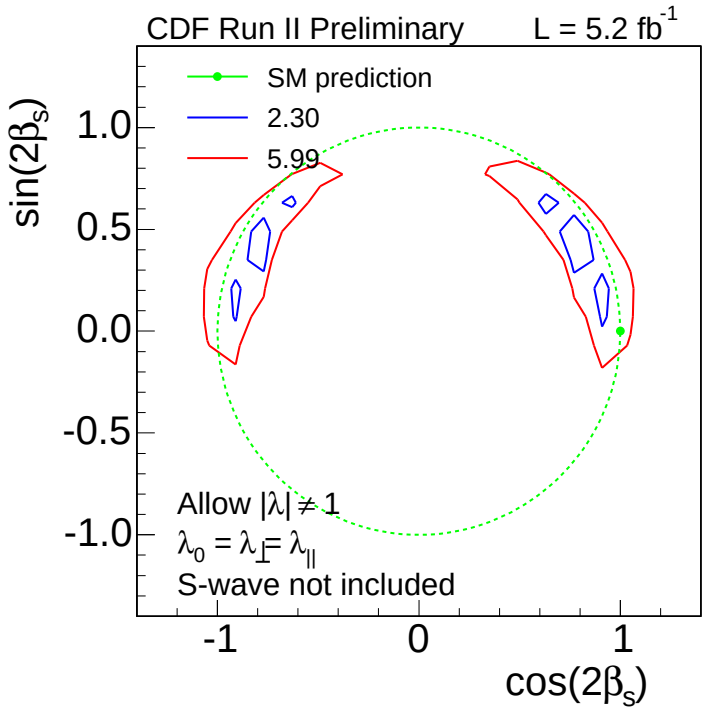

Figure 5.20: $\cos 2\beta_s$-$\sin 2\beta_s$ likelihood profile for fit for direct CP violation.

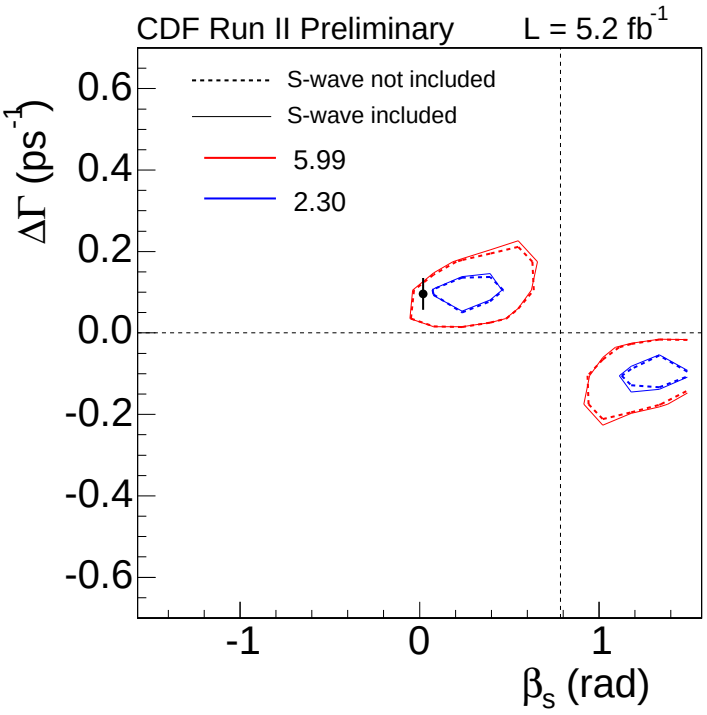

Figure 5.21: $\beta_s - \Delta\Gamma$ likelihood scan, with and without f_0/non-resonant K^+K^- contribution.

5.4.1 Coverage adjustment

We use a likelihood ratio ordering procedure to ensure that our $\beta_s - \Delta\Gamma$ contours do not undercover and account for the non-Gaussian errors that we observe with the present statistics [67]. We also calculate a p-value to determine the probability that our observed measurement of β_s is a fluctuation from the standard model expectation, β_s=0.02.

We generate 1000 toys at the standard model point in the $\beta_s - \Delta\Gamma$ plane (β_s=0.02, $\Delta\Gamma$=0.096). The toys are fit with all fit parameters including β_s and $\Delta\Gamma$ floating, and then fit a second time with β_s and $\Delta\Gamma$ fixed to their standard model values. For each toy we form a likelihood ratio by dividing the $-2\log\mathcal{L}$ returned by the first fit by the one returned by the second fit, resulting in

$$\mathcal{LR} = 2\log\frac{\mathcal{L}(\beta_s, \Delta\Gamma, \vec{\xi})}{\mathcal{L}(\vec{\xi})}, \tag{5.2}$$

where $\vec{\xi}$ denotes the set of fit parameters, except β_s and $\Delta\Gamma$. The fit parameters in $\vec{\xi}$ are often referred to as "nuisance" parameters.

The distribution of likelihood ratios for the toys is integrated and normalized. Taking the log of the integrated likelihood ratio distribution, we obtain the distribution of 1 - (Confidence Level), as shown in the solid black line in Fig. 5.22. From this distribution, it is possible to determine the "up" values necessary to guarantee correct coverage in a non-Gaussian regime at the 68 and 95% confidence levels. This is achieved by finding the $\mathcal{LR}$ values that map to $1 - 0.68 = 0.32$ and $1 - 0.95 = 0.05$ in Fig. 5.22. These $\mathcal{LR}$s are the up values, which give the number of units one must go up from the minimum at zero in the $\beta_s - \Delta\Gamma$ likelihood profile to find the 68% and 95% confidence regions. The ideal case of Gaussian errors is shown in green.

The p-value is the probability that our measurement of β_s is a fluctuation from a standard model true value. It is calculated by comparing the $\mathcal{LR}$ for data to the $\mathcal{LR}$ for each toy. The p-value is the number of toys whose $\mathcal{LR}$ exceeds the data value of $\mathcal{LR}$ divided by the total number of toys. We calculate a p-value of 0.27. This corresponds to "up" values of 2.85 (68% CL) and 7.34 (95% CL).

In order to guarantee coverage over a range of possible values of nuisance parameters, an ensemble of sixteen "alternate" universes is generated. Each universe is pro-

duced by generating toys with nuisance parameters randomized uniformly within $\pm 5\sigma$ of their measured value on data. The only exception is the phase δ_S, which is generated flat between 0 and 2π. Since the f_0/non-resonant K^+K^- fraction is consistent with zero, we lack sensitivity to the associated phase and choose to vary it over the full range possible.

To determine the final coverage adjustment, we generate 1000 toys for each of the sixteen alternative universes. The same likelihood ratio procedure described above is performed on each universe, and the most conservative p-value and "up" values are taken. The colored, dashed lines in Fig. 5.22 show the log of the integrated likelihood ratio distribution for the sixteen alternative universes. After adjusting the coverage to take into account the "worst" (most conservative) universes, we obtain a p-value of 0.44 and "up" values of 4.27 (68% CL)and 9.10 (95% CL).

The same coverage adjustment procedure is carried out for the 1D β_s likelihood profile. The β_s profile coverage adjustment has the modification that $\Delta\Gamma$ is randomized in pseudo-experiment generation and treated analogously to any other fit parameter.

For the 1D β_s case, using 1000 toys, we find a p-value of 0.14 for the non-Gaussian error adjustment. The "up" values are 1.68 (68% CL) and 5.37 (95% CL). The solid black line in Fig. 5.23 shows the (1 - CL) distribution.

We use 1000 toys per alternative universe for the coverage adjustment in the 1D case. Considering the worst case alternative universe from Fig. 5.23, we obtain a p-value of 0.31 and "up" values of 2.93 (68% CL) and 7.73 (95% CL).

5.4.2 Coverage adjusted contours

We present the coverage adjusted contours in this section. The $\beta_s - \Delta\Gamma$ scan after the first coverage adjustment is shown in Fig. 5.24. The final, fully adjusted contour is shown in Fig. 5.25. In both plots, the standard model prediction is shown as black marker. For the final contour, the standard model prediction is partially included, partially excluded at the 68% confidence level.

The one dimensional β_s confidence interval after the initial coverage adjustment is shown in Fig. 5.26. The final β_s confidence interval, after full adjustment, is shown in Fig. 5.27.

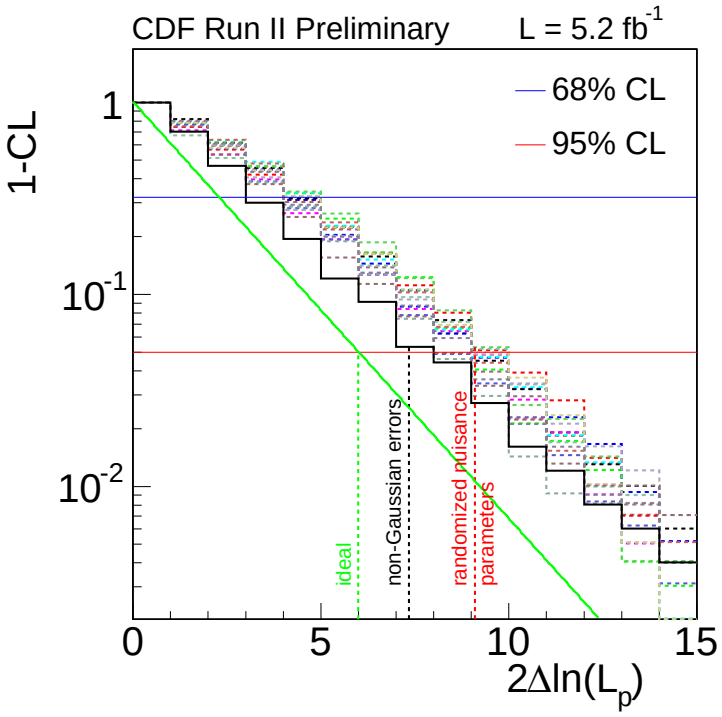

Figure 5.22: 1-CL distribution for β_s-$\Delta\Gamma$ scan.

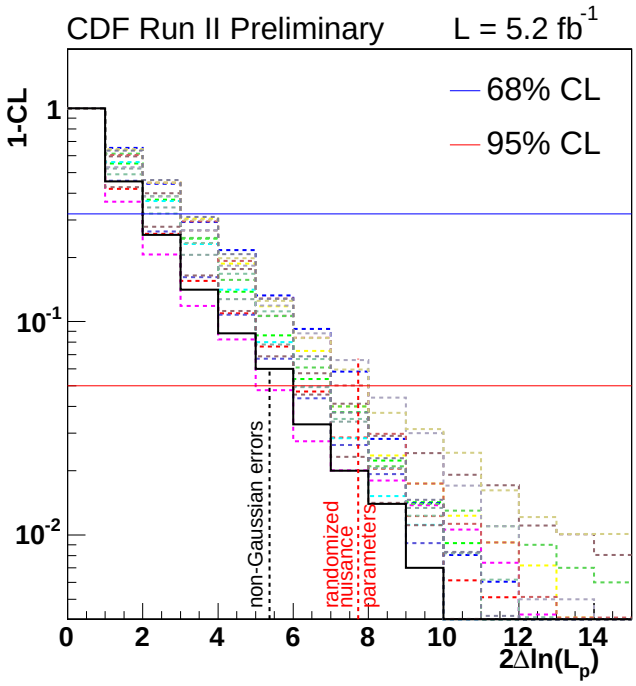

Figure 5.23: 1-CL distribution for β_s scan.

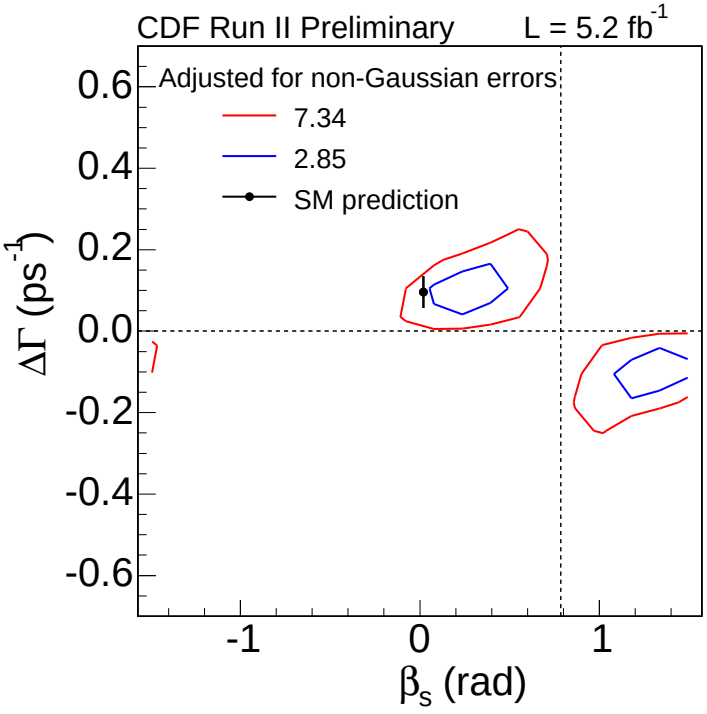

Figure 5.24: $\beta_s - \Delta\Gamma$ confidence region after initial coverage adjustment.

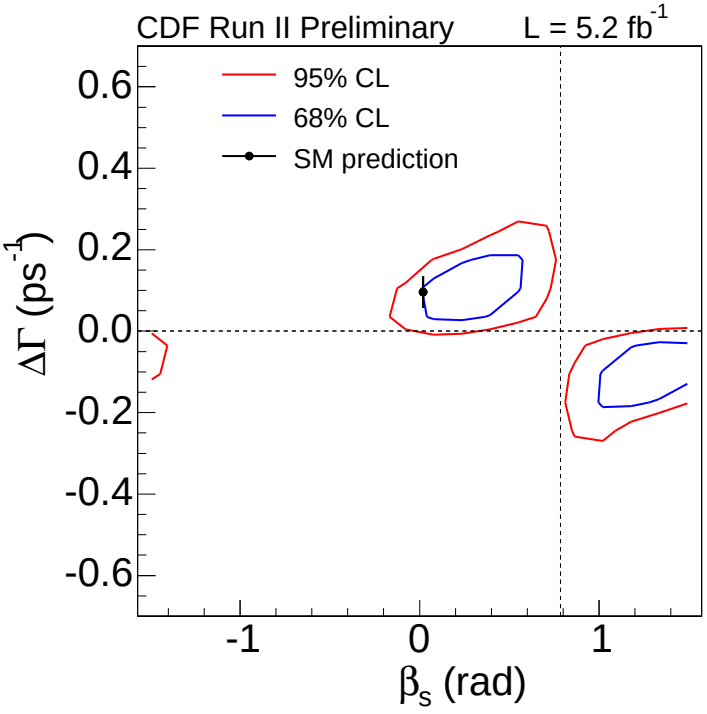

Figure 5.25: $\beta_s - \Delta\Gamma$ confidence region after final coverage adjustment.

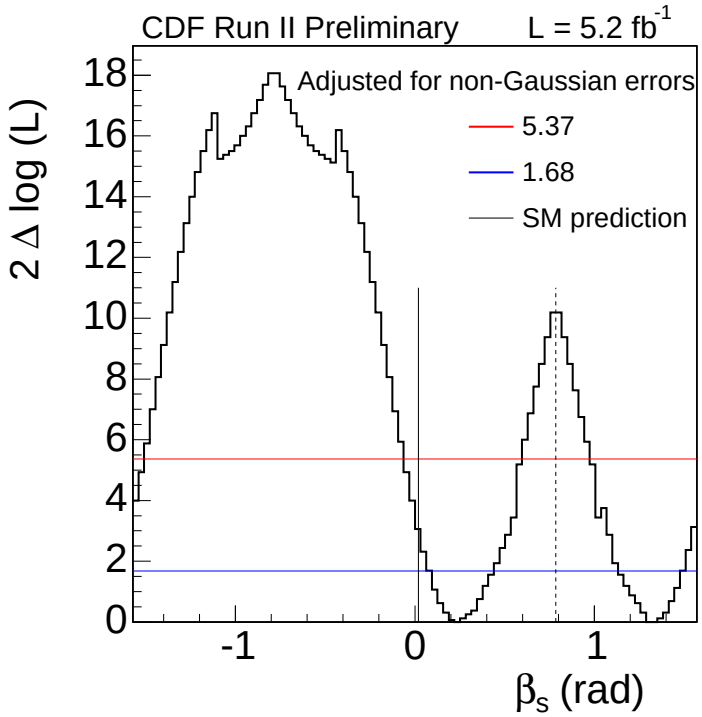

Figure 5.26: β_s confidence interval after initial coverage adjustment.

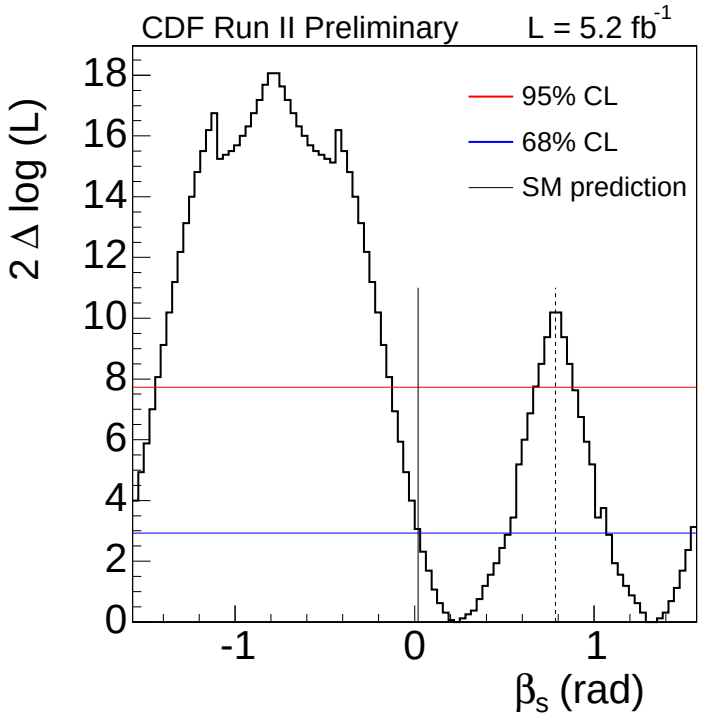

Figure 5.27: β_s confidence interval after final coverage adjustment.

Based on the one-dimensional confidence interval, β_s has the range [0.02, 0.52] U [1.08, 1.55] at the 68% confidence level, and [-0.13, 0.68] U [0.89, pi/2] U [-pi/2, -1.44] at the 95% confidence level.

Chapter 6

Conclusions

In this thesis, we presented the latest measurement of the CP violating phase β_s, using $B_s \to J/\psi\phi$ decays. We performed a time-dependent, flavor-tagged angular analysis, and extract the parameters of interest with an unbinned likelihood fit. The signal selection was optimized in a novel way which used pseudo-experiments to maximize the sensitivity to β_s. We measured several parameters related to the CP violating phase, including the B_s^0 lifetime, the decay width difference $\Delta\Gamma$ between the heavy and light B_s^0 mass eigenstates, the transversity amplitudes $|A_{||}|$ and $|A_\perp|$, and the strong phase $\phi_\perp$. The measurements for $c\tau$ and $\Delta\Gamma$ are in good agreement with the PDG world averages, and are the world's most precise measurements on a single decay mode. The measurements of the transversity amplitudes are consistent with previous measurements in the $B_s^0 \to J/\psi\phi$ system. The transversity amplitudes are also consistent with those measured using $B^0 \to J/\psi K_S^0$ decays, whose final state angular behavior is expected to be similar to the $J/\psi\phi$ system [25]. For the first time, we incorporate in the likelihood the possibility of contributions to the final state from $B_s^0 \to J/\psi f_0$ and $B_s^0 \to J/\psi K^+K^-$ decays. We measure this contribution to be less than 6.7% at the 95% confidence level.

The probability that the observed confidence regions in the $\beta_s - \Delta\Gamma$ plane are fluctuations from the standard model point is 44%, compared to 7% for the most recent iteration of the analysis. Our consistency with the standard model prediction is much better than it has been for previous iterations of the analysis [68]. This effect does not appear to be due to the inclusion of the f_0/non-resonant K^+K^- contributions, but simply the effect of increasing statistics and decreasing the errors on β_s.

We also presented a complementary, independent fit that allowed for the possibility of direct CP violation. The results of this fit were consistent with the theoretical prediction of no direct CP violation in the $B_s^0 \to J/\psi\phi$ system. The confidence levels also were well-consistent with the main fit, increasing our confidence in the robustness of our result.

The measurements presented in this thesis are still statistically dominated. Thus, further measurements on larger data samples are warranted in the future. This measurement will continue to play an important role in the study of heavy flavor physics, both in the final years of Tevatron operation, and in the first years of running at the Large Hadron Collider.

Appendix A

Decay Rates, With Full Angular Dependence

We expand Eqs. 1.54 and 1.55 to show the full angular dependence of the B_s^0 and $\bar{B}_s^0$ probability densities. We first expand the time-independent cross products for the transversity amplitudes at $t = 0$ with the unit vector $\hat{n}$ to make the angular dependence explicit. The expanded terms are:

$$\begin{aligned} |\mathbf{A}_+ \times \hat{n}|^2 &= |A_0(0)|^2 \cos^2\psi(\cos^2\theta + \sin^2\theta\sin^2\phi) + \frac{1}{2}|A_{||}(0)|^2 \sin^2\psi(\cos^2\theta + \sin^2\theta\cos^2\phi) \\ &+ \frac{1}{2\sqrt{2}}|A_0(0)||A_{||}(0)|\cos\phi_{||}\sin 2\psi\sin^2\theta\sin 2\phi, \end{aligned} \tag{A.1}$$

$$|\mathbf{A}_- \times \hat{n}|^2 = \frac{1}{2}|A_\perp(0)|^2 \sin^2\psi\sin^2\theta, \tag{A.2}$$

and

$$\begin{aligned} (\mathbf{A}_+ \times \hat{n}) \cdot (\mathbf{A}_-^* \times \hat{n}) &= \frac{i}{4}|A_{||}(0)||A_\perp(0)|e^{i(\phi_{||}-\phi_\perp)}\sin^2\psi\sin 2\theta\sin\phi \\ &+ \frac{i}{4\sqrt{2}}|A_0(0)||A_\perp(0)|e^{-i\phi_\perp}\sin 2\psi\sin 2\theta\cos\phi. \end{aligned} \tag{A.3}$$

We define six angular functions

A. DECAY RATES, WITH FULL ANGULAR DEPENDENCE

$$
\begin{aligned}
g_1(\vec{\rho}) &= \cos^2\psi(1-\sin^2\theta\cos^2\phi) \\
g_2(\vec{\rho}) &= \frac{1}{2}\sin^2\psi(1-\sin^2\theta\sin^2\phi) \\
g_3(\vec{\rho}) &= \frac{1}{2}\sin^2\psi\sin^2\theta \\
g_4(\vec{\rho}) &= \frac{1}{2\sqrt{2}}\sin 2\psi\sin^2\theta\sin 2\phi \\
g_5(\vec{\rho}) &= -\frac{1}{2}\sin^2\psi\sin 2\theta\sin\phi \\
g_6(\vec{\rho}) &= \frac{1}{2\sqrt{2}}\sin 2\psi\sin 2\theta\cos\phi.
\end{aligned} \tag{A.4}
$$

The expanded cross products given above, in addition to the angular functions, can be used to rewrite Eq. 1.54 as

$$
\begin{aligned}
P(t,\vec{\rho}) \;\propto\;& |A_0(0)|^2|f_+(t)|^2 g_1(\vec{\rho}) + |A_{||}(0)|^2|f_+(t)|^2 g_2(\vec{\rho}) + |A_\perp(0)|^2|f_-(t)|^2 g_3(\vec{\rho}) \\
+\;& |A_0(0)||A_{||}(0)|\cos\phi_{||}|f_+(t)|^2 g_4(\vec{\rho}) \\
+\;& Re\{i|A_{||}(0)||A_\perp(0)|e^{i(\phi_{||}-\phi_\perp)}f_+(t)f_-^*(t)\}g_5(\vec{\rho}) \\
+\;& Re\{i|A_0(0)||A_\perp(0)|e^{i(\phi_\perp)}f_+(t)f_-^*(t)\}g_6(\vec{\rho}).
\end{aligned} \tag{A.5}
$$

We can expand the terms that express the time dependence, and utilize the strong phases

$$\delta_1 = Arg(A_{||}^*(0)A_\perp(0)) = \phi_\perp - \phi_{||} \tag{A.6}$$

$$\delta_2 = Arg(A_0^*(0)A_\perp(0)) = \phi_\perp - \phi_0 \tag{A.7}$$

to write the B_s^0 probability density with both the time and angular dependence made explicit, and only real, measurable components remaining:

$$
\begin{aligned}
P(t,\vec{\rho}) \;\propto\;& |A_0(0)|^2 e^{-\Gamma t}[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m_s t] g_1(\vec{\rho}) \\
+\;& |A_{||}(0)|^2 e^{-\Gamma t}[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m_s t] g_2(\vec{\rho}) \\
+\;& |A_{\perp}(0)|^2 e^{-\Gamma t}[\cosh\frac{\Delta\Gamma}{2}t + \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m_s t] g_3(\vec{\rho}) \\
+\;& |A_0(0)||A_{||}(0)| \cos(\delta_2-\delta_1) e^{-\Gamma t} \times \\
& [\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m_s t] g_4(\vec{\rho}) \\
+\;& |A_{||}(0)||A_{\perp}(0)| e^{-\Gamma t} \times \\
& [\sin\delta_1 \cos\Delta m_s t - \cos\delta_1 \cos 2\beta_s \sin\Delta m_s t + \cos\delta_1 \sin 2\beta_s \sinh\frac{\Delta\Gamma}{2}t] g_5(\vec{\rho}) \\
+\;& |A_0(0)||A_{\perp}(0)| e^{-\Delta\Gamma t} \times \\
& [\sin\delta_2 \cos\Delta m_s t - \cos\delta_2 \cos 2\beta_s \sin\Delta m_s t + \cos\delta_2 \sin 2\beta_s \sinh\frac{\Delta\Gamma}{2}t] g_6(\vec{\rho})
\end{aligned}
\tag{A.8}
$$

The same algebra can be repeated to arrive at the explicit form of the $\bar{B}_s^0$ probability density:

$$
\begin{aligned}
\bar{P}(t,\vec{\rho}) \;\propto\;& |A_0(0)|^2 e^{-\Gamma t}[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m_s t] g_1(\vec{\rho}) \\
+\;& |A_{||}(0)|^2 e^{-\Gamma t}[\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m_s t] g_2(\vec{\rho}) \\
+\;& |A_{\perp}(0)|^2 e^{-\Gamma t}[\cosh\frac{\Delta\Gamma}{2}t + \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t - \sin 2\beta_s \sin\Delta m_s t] g_3(\vec{\rho}) \\
+\;& |A_0(0)||A_{||}(0)| \cos(\delta_2-\delta_1) e^{-\Gamma t} \times \\
& [\cosh\frac{\Delta\Gamma}{2}t - \cos 2\beta_s \sinh\frac{\Delta\Gamma}{2}t + \sin 2\beta_s \sin\Delta m_s t] g_4(\vec{\rho}) \\
+\;& |A_{||}(0)||A_{\perp}(0)| e^{-\Gamma t} \times \\
& [-\sin\delta_1 \cos\Delta m_s t + \cos\delta_1 \cos 2\beta_s \sin\Delta m_s t + \cos\delta_1 \sin 2\beta_s \sinh\frac{\Delta\Gamma}{2}t] g_5(\vec{\rho}) \\
+\;& |A_0(0)||A_{\perp}(0)| e^{-\Gamma t} \times \\
& [-\sin\delta_2 \cos\Delta m_s t + \cos\delta_2 \cos 2\beta_s \sin\Delta m_s t + \cos\delta_2 \sin 2\beta_s \sinh\frac{\Delta\Gamma}{2}t] g_6(\vec{\rho}).
\end{aligned}
\tag{A.9}
$$

These fully expanded forms of probability densities match the standard differential decay rates for decay of a pseudo-scalar particle to two vector particles [11].

A. DECAY RATES, WITH FULL ANGULAR DEPENDENCE

Appendix B

Explicit Description of ϕ Mass Line Shape

The function that we use to parameterize the shape of the ϕ resonance is an asymmetric relativistic Breit-Wigner with a mass dependent width. The function is as follows:

$$h(\mu) = \frac{\mu}{\mu_\phi} \cdot \Gamma_1 \cdot \frac{k^*(B_s, \mu, J/\psi)}{k^*(B_s, \mu_\phi, J/\psi)} \cdot \frac{1}{(\mu_\phi^2 - \mu^2)^2 + \mu_\phi^2 \cdot \Gamma_{tot}^2}. \tag{B.1}$$

The k^* terms give the momentum of a particle with mass μ in the rest frame of the B_s^0, assuming a two body decay in which the other daughter particle is a J/ψ. The total decay width Γ_{tot} is

$$\Gamma_{tot} = \Gamma_1 + \Gamma_2 + \Gamma_3 \tag{B.2}$$

where Γ_1, Γ_2 and Γ_3 are the three dominant ϕ partial decay widths: $\phi \to K^+K^-$, $\phi \to K_L^0 K_s^0$, and $\phi \to \rho\pi + \pi^+\pi^-\pi^0$, with branching ratios 0.489, 0.342, and 0.153, respectively. The partial decays widths are defined by

$$\begin{aligned}
\Gamma_1 &= \Gamma_\phi \cdot 0.489 \cdot \left[\frac{k^*(\mu, \mu_{K^+}, \mu_{K^+})}{k^*(\mu_\phi, \mu_{K^+}, \mu_{K^+})}\right]^3 \cdot \frac{\mu_\phi}{\mu} \cdot \left[\frac{1 + R^2 k^{*2}(\mu_\phi, \mu_{K^+}, \mu_{K^+})}{1 + R^2 k^{*2}(\mu, \mu_{K^+}, \mu_{K^+})}\right] \\
\Gamma_2 &= \Gamma_\phi \cdot 0.342 \cdot \left[\frac{k^*(\mu, \mu_{K^0}, \mu_{K^0})}{k^*(\mu_\phi, \mu_{K^0}, \mu_{K^0})}\right]^3 \cdot \frac{\mu_\phi}{\mu} \cdot \left[\frac{1 + R^2 k^{*2}(\mu_\phi, \mu_{K^0}, \mu_{K^0})}{1 + R^2 k^{*2}(\mu, \mu_{K^0}, \mu_{K^0})}\right] \\
\Gamma_3 &= \Gamma_\phi \cdot (1 - 0.489 - 0.342),
\end{aligned} \tag{B.3}$$

where Γ_ϕ is the ϕ width (4.26 MeV/c^2), R is the Blatt-Weisskopf radius of $\sim 3\text{GeV}^{-1}$, which defines the barrier factor associated with a resonance's potential well. The barrier factor is determined by the orbital angular momentum of the state.

The analytic function for the ϕ line shape agrees well with realistic Monte Carlo, as shown in Fig. B.1.

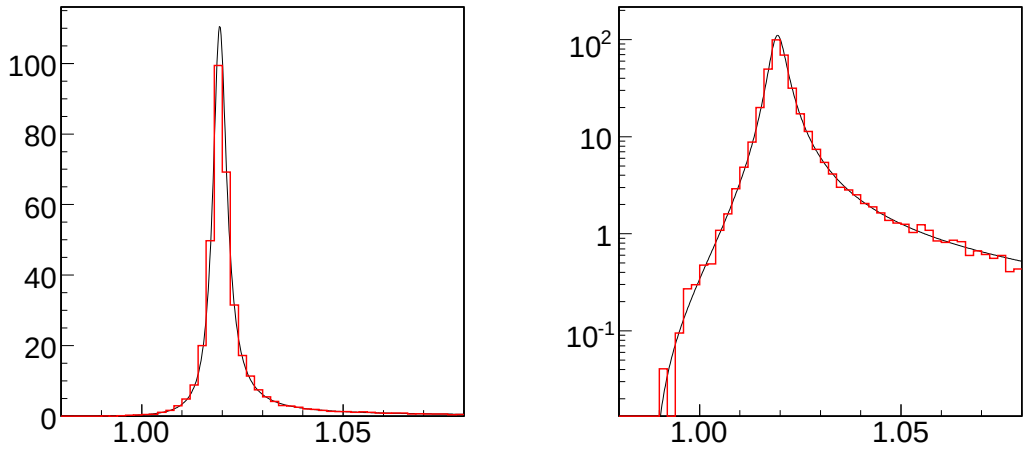

Figure B.1: Agreement between analytic description of ϕ line shape (black) and Monte Carlo (red).

Appendix C

Pulls for non-standard model β_S and $\Delta\Gamma$

We present the pull distributions for pseudo-experiments with realistic statistics (6,500 signal events), and generated at non-standard model values of β_s and $\Delta\Gamma$. All pull distributions are generated with an f_0/non-resonant K^+K^- fraction of 25%, and δ_S of $\pi/2$. Figs. C.1-C.3 give the pulls for pseudo-experiments generated with β_s of 0.25 and $\Delta\Gamma$ of 0.08. Figs. C.4-C.6 give the pulls for pseudo-experiments generated with β_s of 0.50 and $\Delta\Gamma$ of 0.05.

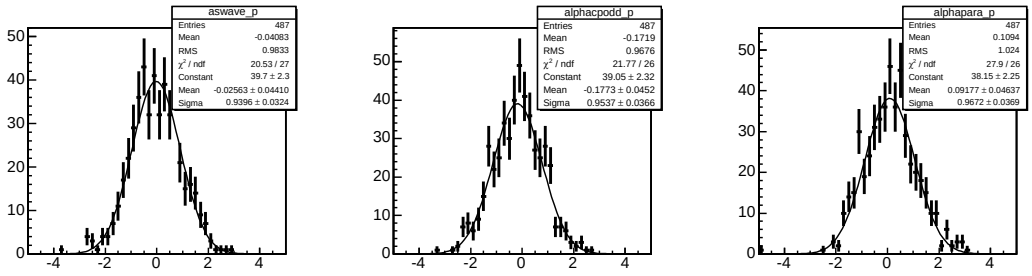

Figure C.1: Pulls for f_{swave} (left), α_{CPodd} (center) and $\alpha_{||}$ (right), generated with β_s=0.25 and $\Delta\Gamma$=0.08.

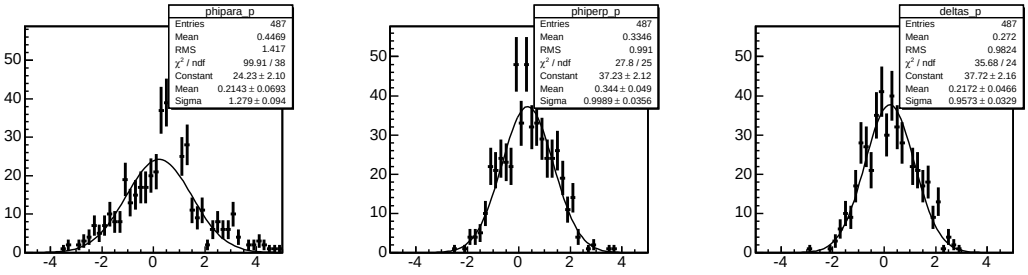

Figure C.2: Pulls for $\phi_{||}$ (left), $\phi_{\perp}$ (center) and δ_S (right), generated with β_s=0.25 and $\Delta\Gamma$=0.08.

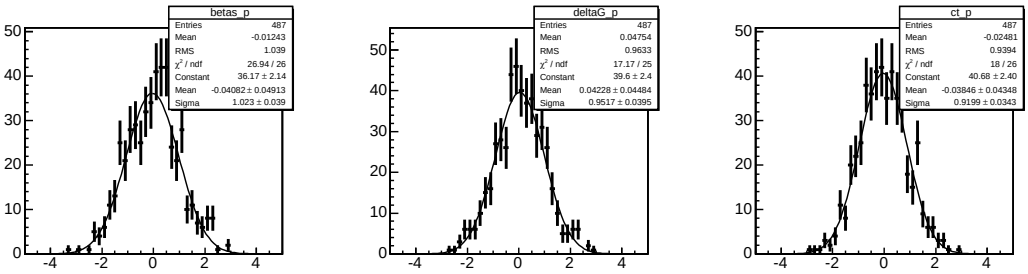

Figure C.3: Pulls for β_s (left), $\Delta\Gamma$ (center) and ct (right), generated with β_s=0.25 and $\Delta\Gamma$=0.08.

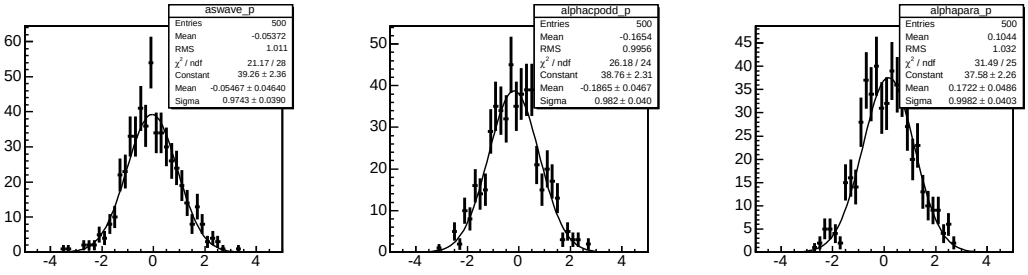

Figure C.4: Pulls for f_{swave} (left), α_{CPodd} (center) and $\alpha_{||}$ (right), generated with β_s=0.50 and $\Delta\Gamma$=0.05.

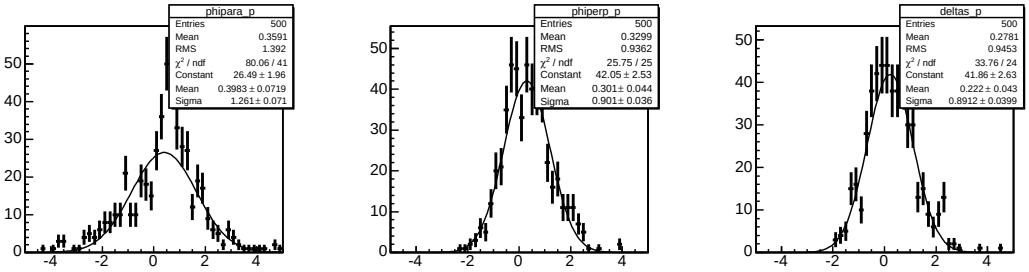

Figure C.5: Pulls for $\phi_{||}$ (left), $\phi_{\perp}$ (center) and δ_S (right), generated with β_s=0.50 and $\Delta\Gamma$=0.05.

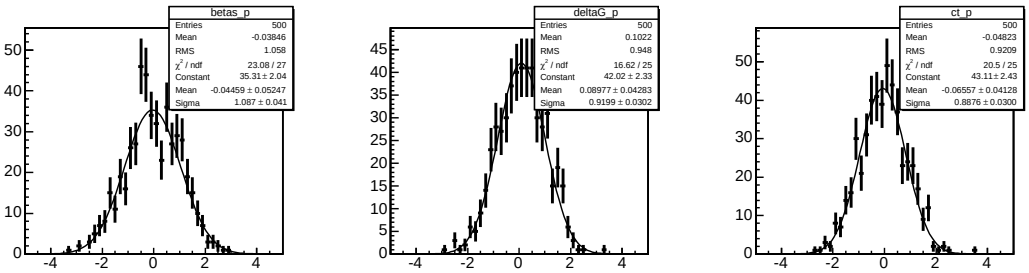

Figure C.6: Pulls for β_s (left), $\Delta\Gamma$ (center) and ct (right), generated with β_s=0.50 and $\Delta\Gamma$=0.05.

C. PULLS FOR NON-STANDARD MODEL β_S AND $\Delta\Gamma$

Appendix D

Direct CP Violation fit validation

We test the fit for direct CP violation in the usual way, using toy Monte Carlo pseudo-experiments to verify that pull distributions for the fit parameters of interest are Gaussian, with mean of zero and unit width. We check that the fit is unbiased for several generated values of β_s: $\beta_s = 0.0$, $\beta_s = 0.25$, $\beta_s = \pi/2$. We test the high statistics limit, using 150,000 signal events per toy, and 1,500 toys for each β_s scenario. Pulls are shown is Figs. D.1-D.9. The pulls are Gaussian, unbiased and unitary in the high statistics limit, which validates our implementation of the likelihood function.

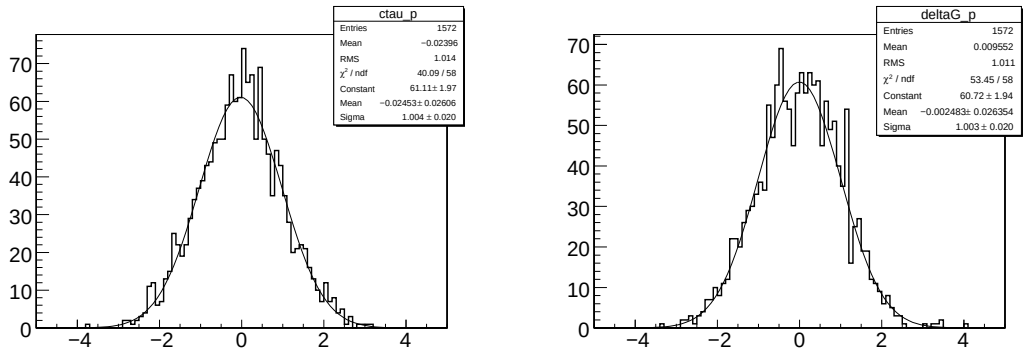

Figure D.1: High statistics DCPV fit pulls for ct (left) and $\Delta\Gamma$ (right), generated β_s=0.

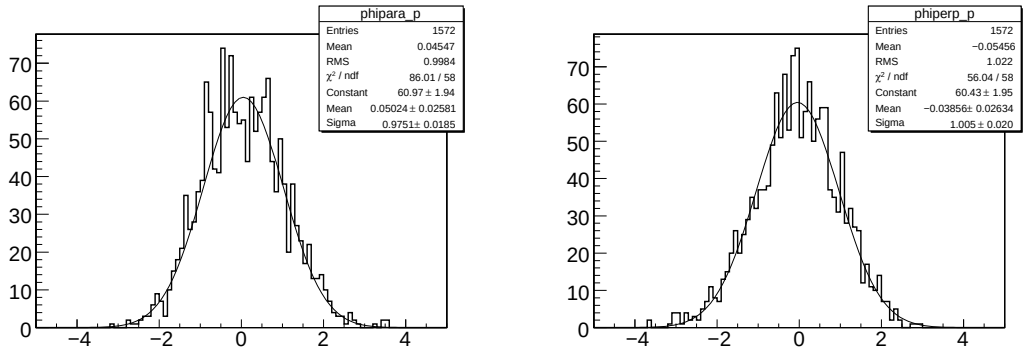

Figure D.2: High statistics DCPV fit pulls for $\phi_{||}$ (left) and $\phi_{\perp}$ (right), generated β_s=0.

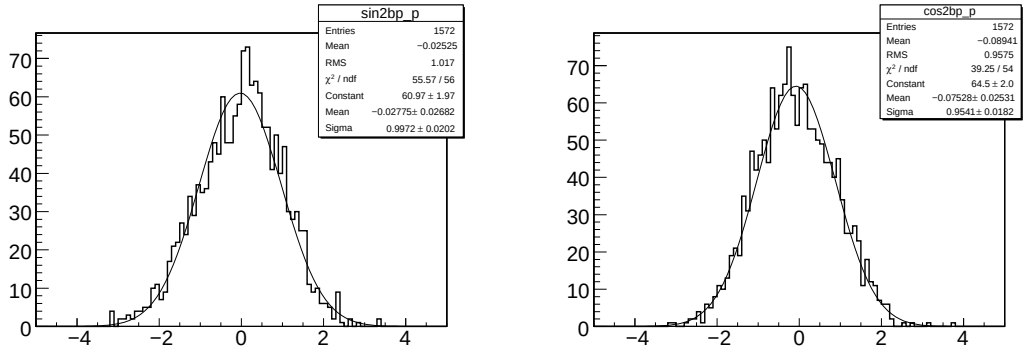

Figure D.3: High statistics DCPV fit pulls for $\sin(2\beta_s)$ (left) and $\cos(2\beta_s)$ (right), generated β_s=0.

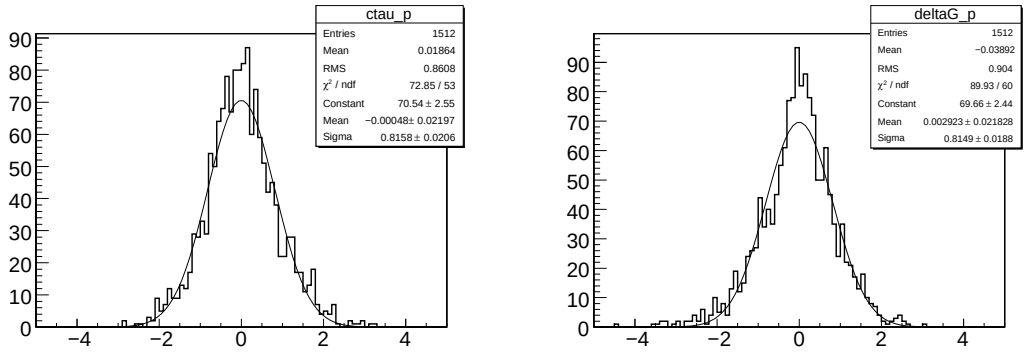

Figure D.4: High statistics DCPV fit pulls for ct (left) and $\Delta\Gamma$ (right), generated β_s=0.

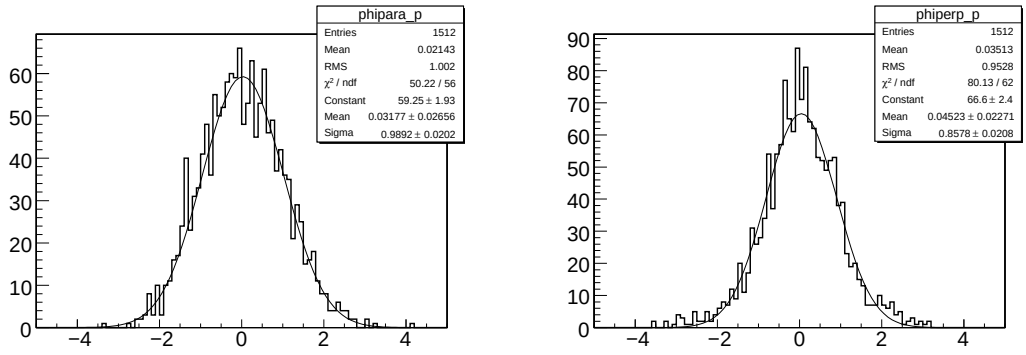

Figure D.5: High statistics DCPV fit pulls for $\phi_{||}$ (left) and $\phi_{\perp}$ (right), generated β_s=0.

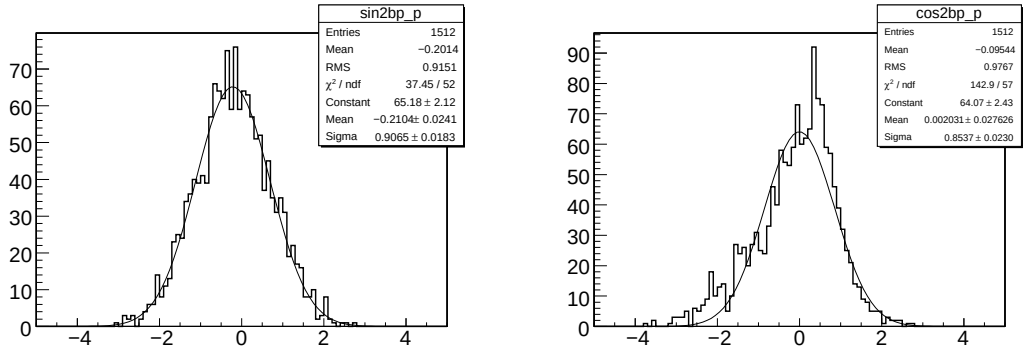

Figure D.6: High statistics DCPV fit pulls for $\sin(2\beta_s)$ (left) and $\cos(2\beta_s)$ (right), generated β_s=0.

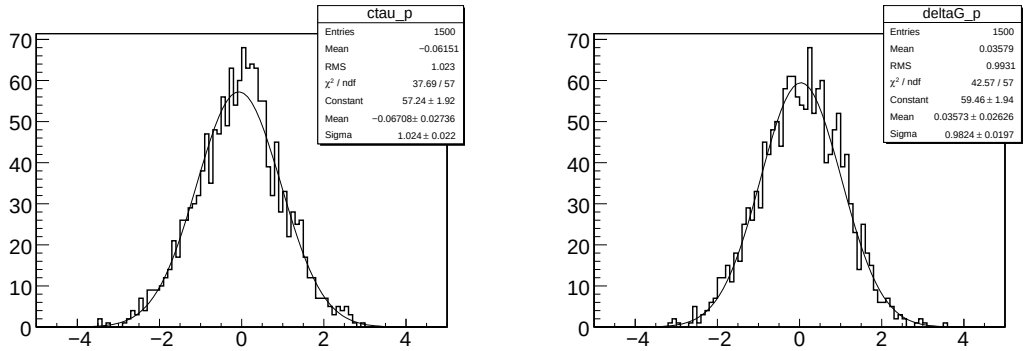

Figure D.7: High statistics DCPV fit pulls for ct (left) and $\Delta\Gamma$ (right), generated β_s=0.

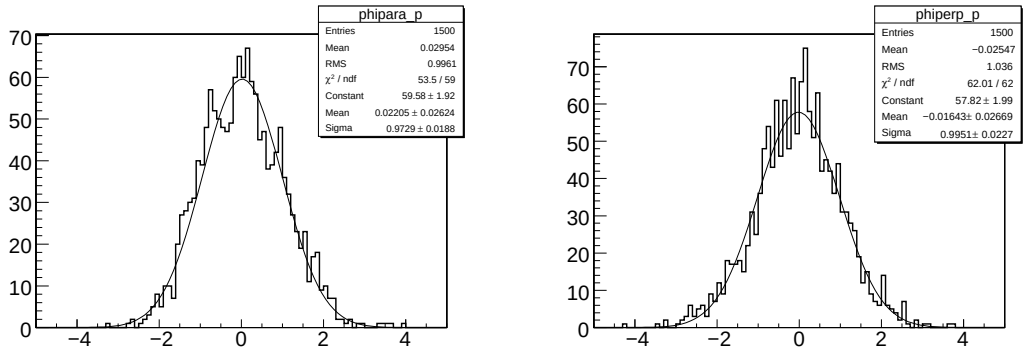

Figure D.8: High statistics DCPV fit pulls for $\phi_{||}$ (left) and $\phi_{\perp}$ (right), generated β_s=0.

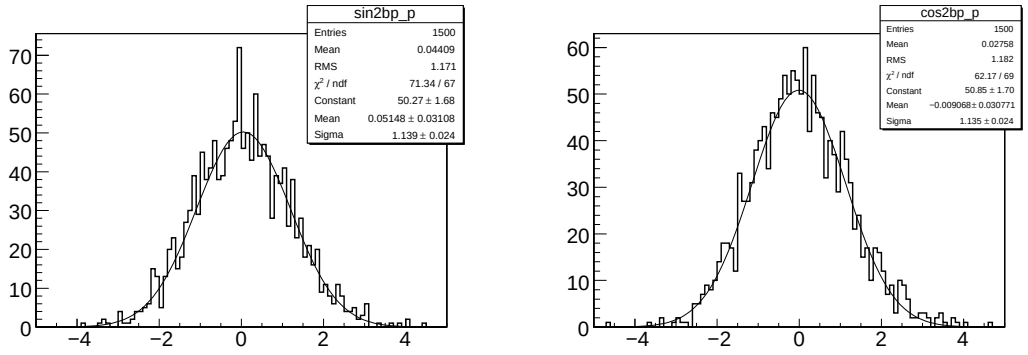

Figure D.9: High statistics DCPV fit pulls for $\sin(2\beta_s)$ (left) and $\cos(2\beta_s)$ (right), generated β_s=0.

References

[1] F. Halzen and A. Martin. *Quarks and Leptons, An Introductory Course in Modern Particle Physics*. John Wiley, 1984.

[2] Particle Data Group.
http://pdg.lbl.gov/

[3] J.H. Cristenson et al. **Evidence for the 2π Decay of the K_2^0 Meson**. *Phys. Rev. Lett.*, **13**:138–140, 1964.

[4] M. Kobayashi and T. Maskawa. **CP Violation in the Renormalizable Theory of Weak Interaction**. *Prog. Theor. Phys.*, **49**:652, 1973.

[5] J. Richman. *Heavy-Quark Physics and CP Violation*. Elsevier Science, 2009.

[6] L. Wolfenstein. **Parametrization of the Kobayashi-Maskawa Matrix**. *Phys. Rev. Lett.*, **51**:1945, 1983.

[7] C. Jarlskog. **Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP**. *Phys. Rev. Lett.*, **55**:1039, 1985.

[8] D. Kirkby and Y. Nir. **CP Violation in Meson Decays**.
http://pdg.lbl.gov/2009/reviews/rpp2009-rev-cp-violation.pdf

[9] W. Hou et al. **Large Time-dependent CP Violation in B_s^0 System and Finite $D^0 - \bar{D}^0$ Mass Difference in Four Generation Standard Mode**. *Phys. Rev. D*, **76**:016004, 2007.

[10] W. Hou et al. **Difference in B^+ and B^0 direct CP asymmetry as effect of a fourth generation**. *Phys. Rev. Lett.*, **95**:141601, 2005.

[11] A. Dighe et al. **Angular Distributions and Lifetime Differences in $B_s \to J/\psi\phi$ Decays**. 1995.

[12] I. Dunietz et al. **In pursuit of new physics with B_s^0 decays**. *Phys. Rev. D*, **63**:114015, 2001.

[13] M. Gronau and J. Rosner. **Flavor symmetry for strong phases and determination of β_s, $\Delta\Gamma$ in $B_s \to J/\psi\phi$.** *Phys. Lett.*, **B669**:321–326, 2008.

[14] F. Azfar et al. **Formulae for the Analysis of the Flavor-Tagged Decay $B_s^0 \to J/\psi\phi$.** CDF Note 10032, 2010.

[15] S. Stone and L. Zhang. **S-waves and the Measurement of CP Violating Phases in B_s Decays**. *Phys. Rev. D*, **79**:074024, 2009.

[16] B. Aubert and et al (BABAR Collaboration). **Ambiguity-Free Measurement of cos2beta: Time-Integrated and Time-Dependent Angular Analyses of $B \to J/\psi K\pi$.** 2004

[17] E.P. Wigner. **Lower Limit for the Energy Derivative of the Scattering Phase Shift**. *Phys. Rev.*, **98**:145, 1955.

[18] The CDF Collaboration. **Measurement of the $\bar{B}_s^0$ Meson Lifetime Using Semileptonic Decays with Single Lepton Datasets in CDF Run**——. CDF Note 7757, 2005.

[19] The CDF Collaboration. **Measurement of the B_s^0 Lifetime in Hadronic Decays Using Partially and Fully Reconstructed Modes**. CDF Note 9203, 2008.

[20] The CDF Collaboration. **Measurement of b hadron Lifetimes in the modes $B^+ \to J/\psi K^+, B^0 \to J/\psi K^{*0}, B^0 \to J/\psi K_s^0$, and $\Lambda_b^0 \to J/\psi\Lambda^0$, and measurement of the B^+/B^0 and Λ_b^0/B^0 lifetime ratios**. CDF Note 10071, 2010.

[21] D. Acosta et al. (CDF Collaboration). **Measurement of the Lifetime Difference Between Bs Mass Eigenstates**. 2004.

[22] A. Abulencia et al. (CDF Collaboration). **Observation of $B_s^0 - \bar{B}_s^0$ Oscillations**. *Phys. Rev. Lett.*, **97**:242003, 2006.

[23] V.M. Abazov et al. (DØ Collaboration). **Lifetime Difference and CP-Violating Phase in the B_s^0 System**. 2007.

[24] M. Feindt et al. **Measurement of $\Delta\Gamma$ and ϕ_s in untagged $B_s^0 \to J/\psi\phi$ and $B_d \to J/\psi K^*$ Decays**. CDF note 8753, 2007.

[25] J. Boudreau et al. **Measurement of the CP Violation Parameter β_s in Bs to Jpsi Phi**. CDF note 8960, 2007.

[26] J. Boudreau et al. **Study of the time evolution of flavor-tagged Bs mesons flavor-tagged Bs mesons**. CDF note 9090, 2007.

[27] T. Aaltonen and others (CDF Collaboration). **First Flavor-Tagged Determination of Bounds on Mixing- Induced CP Violation in $B_s \to J/\psi\phi$ Decays**. *Phys. Rev. Lett.*, **100**:161802, 2008.

[28] F. Azfar et al. **An Updated Measurement of the CP Violating Phase β_s using $B_s^0 \to J/\psi\,\phi$ Decays**. CDF Note 9395, 2008.

[29] The CDF Collaboration. CDF Public Note 9458, 2008. An Updated Measurement of the CP Violating Phase β_s using $B_s^0 \to J/\psi\,\phi$ Decays

[30] CDF and DØ Combination Working Group. **Combination of DØ and CDF Results $\Delta\Gamma_s$ and the CP-Violating $\beta_s^{J/\psi\phi}$**. CDF Note 9787, 2009.

[31] A. Dighe et al. **Extracting CKM phases and B/s anti-B/s mixing parameters from angular distributions of non-leptonic B decays**. *Eur. Phys. J. C*, **6**:647–662, 1999.

[32] Fermilab Beams Division.
http://www-bd.fnal.gov/runII/index.html

[33] Fermilab Beams Division.
http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html

[34] R. Blair et al. (CDF Collaboration). **The CDF II Detector: Technical Design Report**. Fermilab-Pub-96-390E, 1996.

[35] A. Sill (for the CDF Collaboration). **CDF Run II Silicon Tracking Projects**. *Nucl. Instrum. Method*, **A447**:1–8, 2000.

[36] C.S. Hill (for the CDF Collaboration). **Operational Experience and Performance of the CDFII Silicon Detector**. *Nucl. Instrum. Method*, **A530**:1–6, 2004.

[37] A. A. Affolder et al. (CDF Collaboration). **CDF Central Outer Tracker**. *Nucl. Instrum. Method*, **A526**:249–299, 2004.

[38] H. Wenzel. **Tracking in SVX**. CDF note 1790, 1992.

[39] D. Acosta et al. (CDF Collaboration). **A Time-of-Flight Detector in CDF-II**. *Nucl. Instrum. Method*, **A518**:605–608, 2004.

[40] S. Kuhlmann et al. (for the CDF Collaboration). **The CDF Calorimeter Upgrade for Run IIB**. *Nucl. Instrum. Method*, **A518**:39–41, 2000.

[41] G. Apollinari et al (for the CDF Collaboration). **Shower Maximum Detector for the CDF Plug Upgrade Calorimeter**. *Nucl. Instrum. Method*, **A412**:515–526, 1998.

REFERENCES

[42] C. M. Ginsberg (for the CDF Collaboration). **CDF Run II Muon System**. *Eur. Phys. J.*, **C33**:1002–1004, 2004.

[43] E.J. Thomson et al. *IEEE Trans. Nucl. Sci.*, **49**:1063, 2002.

[44] A. Bardi et al. *Nucl. Instrum. Method*, **A409**:658–661, 1998.

[45] S. Donati. **CDF Detector and Trigger for B Physics**. CDF Note 8391, 2006.

[46] M. Bishai et al. **Run-II Dimuon Trigger Optimization and Efficiency Measurement**. CDF Note 6004, 2002.

[47] J. Marriner. **Secondary Vertex Fit with Mass at Pointing Constraints (CTVMFT)**. CDF Note 1996, 1993.

[48] C. Paus et al. http://cdfcodebrowser.fnal.gov/CdfCode/source/BottomMods/

[49] P. Murat et al. http://cdfcodebrowser.fnal.gov/CdfCode/source/Stntuple/

[50] K. Anikeev et al. **Description of Bgenerator II**. CDF note 5092, 1999.

[51] C. Peterson et al. *Phys. Rev. D*, **27**:105, 1983.

[52] D.J. Lange. **The EvtGen Particle Decay Simulation Package**. *Nucl. Instrum. Method*, **A462**:152–155, 2001.

[53] R. Brun et al., 1987. CERN-DD/EE/84-1

[54] CDF Simulation Group. http://www-cdf.fnal.gov/cdfsim/cdfsim_main.html

[55] M. Feindt. **A Neural Bayesian Estimator for Conditional Probability Densities**. *arxiv:physics/0402093*, 2004.

[56] G. Giurgiu et al. **Muon B Flavor Tagging - A Likelihood Approach**. CDF Note 7043, 2004.

[57] F. Azfar et al. **An Updated Measurement of the CP Violating Phase β_s using $B_s \to J/\psi\phi$ Decays**. CDF Note 9395, 2008.

[58] F. James. **Minuit**. http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html

[59] Bs Mixing Group. **Combined opposite side flavor tagger**. CDF note 8314, 2006.

[60] G. Giurgiu et al. **Muon B Flavor Tagging - A Likelihood Approach**. CDF note 7043, 2004.

[61] A. Schmidt. **Combined B Flavour Tagging with Artificial Neural Networks**. http://www-ekp.physik.uni-karlsruhe.de/pub/web/thesis/iekp-ka2006-11.pdf

[62] Bs Mixing Group. **Same Side Kaon Tagging Studies in Fully Reconstructed Decays**. CDF note 7979, 2006.

[63] Bs Mixing Group. **Same side Kaon Tagging Studies in Fully Reconstructed Decays on the New Data**. CDF note 8182, 2006.

[64] Bs Mixing Group. **Optimization of the Same Side Kaon tagging algorithm combining PID and kinematic variables**. CDF note 8344, 2006.

[65] P. Sphicas. **Combining Flavor-Taggers**. CDF note 3425, 1995.

[66] M. Feindt et al. **Measurement of Mixing Frequency and Amplitude of Neutral B-Mesons**. CDF note 10025, 2010.

[67] G. Punzi. **Ordering Algorithms and Confidence Intervals in the Presence of Nuisance Parameters**. arXiv:physics/0511202v1

[68] The CDF Collaboration. **An Updated Measurement of the CP Violating Phase β_s using $B_s \to J/\psi\phi$ Decays**. CDF Public Note 9458, 2008.