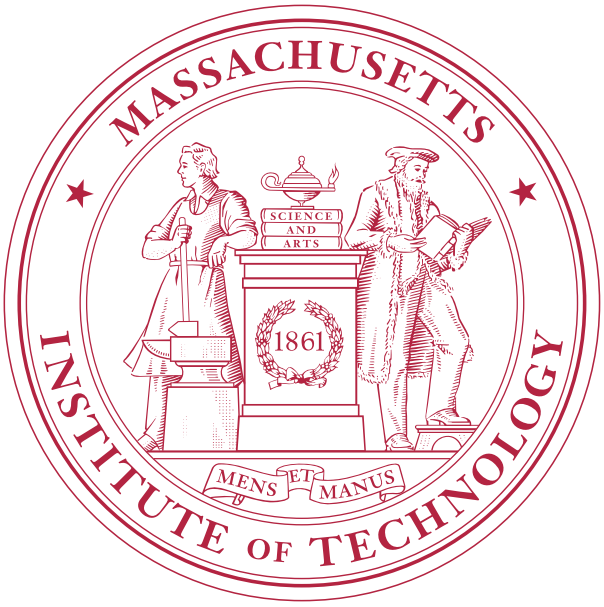
CP Violation in Flavor-Tagged $B_s \to J/\psi\phi$ Decays

Khaldoun Makhoul



Doctoral Dissertation

June 2009

CP Violation in Flavor-Tagged $B_s \to J/\psi\phi$ Decays

by

Khaldoun Makhoul

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2009

Author .....
Department of Physics
March 30, 2009

Certified by .....
Christoph M.E. Paus
Associate Professor
Thesis Supervisor

Accepted by .....
Thomas J. Greytak
Associate Department Head for Education, Professor

CP Violation in Flavor-Tagged $B_s \to J/\psi\phi$ Decays

by

Khaldoun Makhoul

Submitted to the Department of Physics
on March 30, 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

In this dissertation, we present the results of a time-dependent angular analysis of $B_s \to J/\psi\phi$ decays performed with the use of initial-state flavor tagging. *CP* violation is observed in this mode through the interference of decay without net mixing and decay with net mixing, that is, $B_s \to J/\psi\phi$ and $B_s \to \overline{B_s} \to J/\psi\phi$. The time-dependent angular analysis is used to extract the decay widths of the heavy and light B_s eigenstates and the difference between these decay widths $\Delta\Gamma_s \equiv \Gamma_s^L - \Gamma_s^H$. Initial-state flavor tagging is used to determine the matter-antimatter content of the B_s mesons at production time. We combine flavor tagging with the angular analysis, which statistically determines the contributions of the *CP*-even and *CP*-odd components at decay time, to measure the *CP*-violating phase β_s. The phase β_s is expressed in terms of elements of the Cabibbo-Kobayashi-Maskawa matrix as $\beta_s \equiv \arg\left(-V_{ts}V_{tb}^*/V_{cs}V_{cb}^*\right)$, and is predicted by the Standard Model to be close to zero, $\beta_s^{\rm SM} = 0.02$. In the measurement of $\Delta\Gamma_s$, we use a dataset corresponding to 1.7 fb^{-1} of luminosity, collected at the CDF experiment from proton-antiproton collisions at a center of mass energy $\sqrt{s} = 1.96$ TeV. In the measurement of β_s, we use a dataset corresponding to 1.3 fb^{-1} of collected luminosity. We measure $\Delta\Gamma_s = (0.071^{+0.064}_{-0.059} \pm 0.007)$ ps^{-1} using the time-dependent angular analysis. Combining the angular analysis with flavor-tagging, we find that assuming the Standard Model predictions of β_s and $\Delta\Gamma_s$, the probability of a deviation as large as the level of the observed data is 33%. We obtain a suite of associated results which are discussed in detail in this dissertation alongside the main results.

Thesis Supervisor: Christoph M.E. Paus
Title: Associate Professor

Acknowledgments

Experimental particle physics is a fundamentally collaborative effort, and the work presented in these pages would not have been possible without the guidance, assistance, and support of my colleagues at MIT and at the CDF experiment. My time as a Ph.D. student was also made infinitely more enjoyable by the personal relationships I have been lucky to forge during these years.

I am the eleventh student to have the privilege to call Christoph M.E. Paus his Ph.D. advisor. His intuition and skill as a physicist, his dedication, his indefatigability and leadership are all evident in the contributions he has made to our field. These qualities need not my reiteration; they form the natural basis for the admiration shared by all those who know him well. Eleven of us have been lucky to know the qualities that make Christoph stand out as an unmatched mentor and advisor. He manages to give his students generous support while simultaneously granting them the essential independence necessary to develop their own skills. He knew when to trust that my approach was the right one, or when to let me find out that it was the wrong one, and when to step in to correct my course. And finally, perhaps most importantly, he knew how to guide me through challenges — some of them the hardest of my student career — whose solutions lay not in equations, nor in analysis. I end my formal advisee role with Christoph richer in knowledge and in wisdom, but also incalculably wealthier in happy memories of a new and wonderful friendship.

Gian Piero Di Giovanni was my closest collaborator on the work presented in this dissertation. We learned in parallel, stumbled in parallel, often suffered in parallel, and laughed in parallel throughout our student years. We shared daily between the two of us more coffee than is healthy for twenty. We saw many sunrises together following nights of work that refused to end. His contributions to this analysis made it possible: his meticulous and careful tackling of the ever-frustrating B_d decay mode was essential. We worked in tandem on the implementation of much of the B_s analysis framework, especially the daunting incorporation of flavor-tagging. But Gian Piero's generous soul, ever-present optimism, and great capacity for laughter had as much to do with the success of our endeavor as any analytical contribution from either of us. Anna Sfyrla's charming smile, masterful cuisine, as well as her own scientific perspective were often the only things separating Gian Piero and me from insanity. I hope she forgives me for the evenings I stole Gian Piero away for emergency sessions of marathon coding.

Ivan K. Furić was a second advisor to Gian Piero and me. His clarity of vision and depth of understanding were crucial in such a multi-layered and complex analysis. I remember with fondness our numerous and obscenity-laden shouting matches in front of the white board in my office, arguing about the statistical subtleties and the details of the physical interactions in this analysis. Few academic interactions have been more fruitful, and our friendship was always strengthened by these discussions. And on one particular night, Ivan's heroic coding skills saved us from a potentially disastrous delay.

Guillelmo Gomez-Ceballos has the uncanny ability to meet any situation with calm, a smile, and a lightning-fast response. Be it in his response to questions over

e-mail or in his work to assist in a component of the analysis, he always managed to give the impression that he moves and thinks faster than a mere human should. His is a calming effect: no situation is ever too dire, no problem too intractable for Guillelmo. In the intense environment of CDF, Guillelmo was an oasis of measure and serenity. His inexhaustible knowledge of flavor-tagging helped Gian Piero and me to incorporate flavor-tagging into our analysis framework. Konstantin Anikeev gave me the initial push I needed to begin this analysis, and offered many warnings I sometimes heeded too late. His wisdom and his dark humor served me well in the later stages of my work. I am only sad that our interaction was so brief before he left CDF.

I followed in the footsteps of Jeffrey Miles in many ways. Jeff's explanations of many concepts were in my experience equal or superior to all other available explanations. He knows how to think like the person with whom he is conversing. In addition, he only expelled me from his office a few times when I came to shoot the breeze, and for that I am thankful. Our conversations, over coffee, beer, or the keyboard, were always enjoyable.

Diego Tonelli brought the best kind of scientific skepticism and rigor to his role as co-convener of the B Mixing and Lifetimes group at CDF. He was never satisfied with half-answers, never less than careful in his examination of our work, but also nothing but generous with suggestions, support, and assistance. Without a doubt, we owe much to his standards for the quality of the work presented here. His affable nature and charm went a long way toward making work a pleasure rather than a chore. Gavril Giurgiu, as the junior co-convener of our group for a significant period of our time in it, completed the team perfectly. My time working directly with him reinforced my view of him as a meticulous scientist and a reasoned man.

Markus Klute was my first office-mate at CDF. In the time that I have known him, he has come to be much more than that. I inherited his coffee machine, legendary for the strength of its coffee and the particular aroma its stained plastic managed to develop over the years. I hope that I have also inherited if only a little of his unimpeachable work ethic, his magnanimity, and his skill at balancing life with work. I like to think that Conor Henderson came to CDF in part because of my encouragement to do so, but I know that this is immodest. Regardless, I am delighted that he joined our group. His methodical approach to the handling of our Event Builder and Level 3 work turned it from mere service work to a pleasure. It is fair to say that his gregarious nature transformed our group into a social entity as much as a professional one. His skill at poker and fondness for the films of David Lynch cost me a bit financially, and a lot in lost hours of sleep, respectively. Julien Branlard, our favorite low-level RF engineer, brought a welcome bon-vivant perspective to our gatherings. His musical and artistic skills, not to mention his own terrifying poker bravado, added proverbial spice to our physicist's life. I am indebted to him for introducing me to many excellent bandes dessinées. Uros Mavric understood much of the stress of working on a deadline, and provided a kind and understanding ear when times were quite difficult for me. Artur Apresyan and Fabrizio Margaroli shared many coffees with me as well, and kept Gian Piero and me company on many a late night break. Daniel Whiteson was a sentitive and sensible confidant on several occasions, and brought

both a level-headed clarity and a riotous sense of humor to all our conversations.

Rob Roser and Jaco Konigsberg have displayed the best imaginable leadership of the CDF experiment during their time as spokespersons. Their door was always open to me and to all members of the collaboration, which went a long way toward making our very large family a bit more functional. Several other members of our collaboration helped make us more cohesive. Steve and Dee Dee Hahn, J.J. Schmidt and Farrukh Azfar provided wonderful professional as well as moral support long after my service work at the experiment or presentations in group meetings provided weekly interactions with them. Ray Culbertson proved himself indispensable and always available, despite having hundreds of windows open on his computer screen at any given time, each running a separate job. Ray was patient with me even when I tested the limits of human patience. I must also thank Steve Nahn for catching the most humorous mistake in this dissertation, and Hong Liu for providing helpful feedback on yet another thesis defense by one of Christoph's students.

I must confess that despite the several years I spent at Fermilab, my heart remained in Cambridge, in large part because of the irreplaceable memories and friendships I associate with it. Steve Pavlon, from his fort of an office at MIT, provided support with a personal touch. I spent many a late night one office away, working on page after page of at times endless problem sets, and his presence was great comfort. Travis Dunn and Frantisek Ricka enriched my first year of graduate school with the best roommate experience I can imagine. Vered Anzenberg and Julie Millane started out as my favorite homework partners and grew to become close and dependable friends. Rory Monaghan, Lauren Ashwell and Srini Turaga enlivened disproportionately every social event I attended with them. Francois Le Floch-Yin, arguably the most popular Teaching Assistant at MIT during his tenure, lifted my spirits over many discussions. Amandine Denis and Rachael McDermott brought a certain *je ne sais quoi* and a measure of style into our lives.

Katy Miller was the only one willing to take my side on a famous bet concerning my thesis completion date. Sadly, she lost. Meg Aycinena brought a much-needed feminine touch to our legendary 109 Windsor street apartment. As an early model of brilliance and determination in the Computer Science and Artificial Intelligence Lab, when many of us were still getting accustomed to the rigors of graduate student life, she inspired us to strive and succeed. No Sunday brunch was complete without Stephano Bartolucci, despite his inability to arrive less than ninety minutes late, supposedly due to his propensity to "walk slowly." Stephano is the very definition of a gentleman, whose charm and wit are unequaled. Mohamed Mesbah Mostagir Mostafa is one of the rare persons who understand exactly everything I want to express without my having to explain it. We have been through many of the same difficulties, share many of the same flaws, as well as many of the same aspirations. I will often reminisce about one particular summer when time appeared to stop and we seemed to be the only two people in Cambridge. I have rarely been happier than I was during those few months. Paolo Morra is what he coined: a "true man." Behind his bear-like strength — responsible for once cracking one of my ribs — lies the warmest heart I know. His faux light-sleeping notwithstanding, he exemplifies the integrity of true friendship. Théophane Weber combines the nobility of a philanthrope with a

brilliance *éblouissante* and a humility *nonpareil*. He is the center of our little social constellation, and rightly so. He has made the days brighter and the air lighter in our lives for six years, and I know he will continue to do so for many years to come. I owe him a great deal for an excellent credo that begins with an unassuming "On a...".

One particularly peripatetic person held my heart away from Fermilab more strongly than all others. With her, it traveled from Cambridge to Chicago to Washington to Baghdad and back. Missy Ryan has been by my side during the hardest moments of my studies and the happiest, in my lowest moods and at my most stratospheric. With her I have forged too many memories to list exhaustively, from our shared hot chocolates in the Chicago cold to our strolls in the warmth of Savannah to our lazy afternoons in the Cambridge summer heat on Harvard street. Each one is a jewel, our shared memories a veritable treasure chest of happiness. She is warmest joy and loveliest inspiration.

My parents and my brothers saw very little of me during much of my time at MIT and Fermilab. And yet, every moment we passed together replenished me in ways I cannot describe. My brothers have not only filled my days with laughter and joy, but also given me the motivation to aspire to greater heights since the moment they entered our lives. We are alike in many ways, and in this sometimes-mirror of our brotherhood, I have learned more about life than I ever could by looking only into myself. I hope I have made them as proud as they have made me. My parents believed in me long before I believed in myself, and they have never wavered for a moment in their support and in their willingness to sacrifice everything for my brothers and me. Their love has been the greatest gift I have received. There is no height that I can reach that would rob me of the luxury of looking up to them for inspiration or reaching out to them for comfort. I cannot and could not repay my debt to my family, save to say that I love them all.

In memory of Nadimeh Mouhanna,
whose love still warms my heart

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Particle physics is the study of the elementary constituents of the universe, their properties, and their interactions. Experiments in the field have so far revealed four fundamental forces, or interactions: gravity, electromagnetism, the weak interaction, and the strong interaction. Elementary particles fall into two categories, defined by the interactions they experience: quarks, which are subject to all four, and leptons, which are not subject to the strong interaction. All quarks and leptons are fermions, with spin of 1/2. Unlike leptons, quarks are never found singly; they combine in groups of two or three to form bound states called hadrons. Mesons (baryons) are hadrons composed of two (three) quarks. For each particle, there exists an anti-particle with identical properties except for opposite quantum numbers, e.g. same mass and intrinsic spin but opposite charge.

The Standard Model of particle physics (SM) is the collection of theoretical models that describe the fundamental interactions of particles. This collection comprises the electroweak theory, describing electromagnetic and weak interactions, and quantum chromodynamics, describing the strong interaction. These models are quantum field theories, providing a mathematical description of particles and interactions that incorporates quantum mechanics and special relativity by construction. The Standard Model does not include a quantum field theory of gravity, as no such model currently exists that is compatible with the SM and available experimental data. In the SM, interactions occur as the result of the exchange of bosonic (intrinsic spin 1) mediating particles, called intermediate gauge bosons. Experiments have confirmed the existence of the $W^{\pm}$ and Z bosons, the photon, and the gluons, mediating respectively charged weak, neutral weak, electromagnetic, and strong interactions. Table 1.1 lists the elementary particles, organized into three generations: the gauge bosons interact with all three generations. For visual clarity, the following particle symbols are used in lieu of their names: u, d, c, s, t, b, for up, down, charm, strange, top and bottom quarks; e, μ, τ for electron, muon and tau; $\nu_{e/\mu/\tau}$ for electron/muon/tau neutrinos. The up-type quarks (u, c, t) carry electric charge of $+2/3\ e$, while the down-type (d, s, b) quarks $-1/3\ e$, where e is the magnitude of the electron charge. The electron, muon and tau all have charge $-1\ e$; neutrinos carry no electric charge.

The overwhelming majority of the matter around us is made of just three of the fundamental particles listed in Table 1.1: up and down quarks, and electrons. These

	u	c	t
	$1.5 - 3.3$ MeV/c^2	$1.27^{+0.07}_{-0.11}$ GeV/c^2	171.2 ± 2.1 GeV/c^2
Quarks			
	d	s	b
	$3.5 - 6.0$ MeV/c^2	104^{+26}_{-34} MeV/c^2	$4.2^{+0.17}_{-0.07}$ GeV/c^2
	ν_e	ν_μ	ν_τ
	< 2 eV/c^2	< 0.19 MeV/c^2	< 18.2 MeV/c^2
Leptons			
	e	μ	τ
	511 KeV/c^2	106 MeV/c^2	1.78 GeV/c^2

Table 1.1: The elementary particles of the Standard Model listed by generation along the columns, with their masses. Masses are obtained from the latest Particle Data Group averages [1].

quarks form the two baryons that make up atomic nuclei, the proton ($|uud\rangle$ in Dirac notation) and the neutron ($|udd\rangle$), which combine with electrons to form atoms. Particle physics experiments make use of collisions of particles at very high energies in order to produce forms of matter that are not available in natural abundance, to allow the study of the laws that govern the interactions of all matter.

This dissertation focuses on the study of the B_s meson (and its antiparticle conjugate $\overline{B_s}$), an unstable bound state of an anti-bottom quark $\overline{b}$ and a strange quark s that can only be studied at high-energy particle collider experiments. The collision of protons and antiprotons yields, among other things, the necessary b quarks through the interaction $\overline{p}p \rightarrow \overline{b}b$. The outgoing b quarks undergo a process known as hadronization, whereby energy is converted into an associated s ($\overline{s}$) quark, which combines with the $\overline{b}$ (b) quark to form a B_s ($\overline{B_s}$) meson.

Symmetries and conservation laws have played an important role in the development of the Standard Model. This thesis focuses on the symmetry with respect to the charge-conjugation-parity operator, or CP, and its violation in the B_s system. As the combination of two individual operators, CP is fully defined by the actions of its constituent operators. Charge-conjugation takes a particle state into its anti-particle state: it leaves time, position, energy, momentum and spin intact, and takes each quantum number into its additive inverse. The parity operator induces space-reflection about the origin: it takes the momentum vectors $\vec{p}$ into $-\vec{p}$.

The weak interaction is the only force that violates the following symmetries that are otherwise respected by electromagnetic and strong interactions: charge-conjugation, parity, and their combination CP. The weak interaction violates parity maximally. On the other hand, the violation of CP symmetry is a small effect that only manifests itself in the quark sector, and is not readily observable without precision measurement of the physical observables affected by CP violation.

The violation of CP symmetry occurs in the interactions among quarks and the

charged carrier of the weak force, the W boson. These interactions are described in the Standard Model by the Cabibbo-Kobayashi-Maskawa (CKM) mechanism [2]. Several processes are available for study in order to test the predictions of this model. The most common of these have been the production and decay of K ($|\bar{s}q\rangle$) and B ($|\bar{b}q\rangle$) mesons. While the values of individual parameters governing these interactions are not predicted by the SM, it is possible to test the self-consistency of the description. In this way, we test whether there lies physics beyond the SM mechanism that describes these interactions. All of the experimental data up to this point have validated the self-consistency of the CKM picture. However, until recently, no experimental data was available on the effect of *CP* violation in the B_s system.

The study of the production, time-evolution, and decay of B_s mesons yields a unique set of physical observables that test this description. The B_s has been the subject of intense focus in particle physics research in recent years. B_s mesons undergo oscillations between their matter and antimatter states at very high frequency of 17.77 ± 0.12 ps^{-1} [1]. The recent direct observation of this oscillation and the measurement of its frequency provided one of the most important constraints of the CKM model [3]. In order to measure the degree of *CP* violation in the B_s system and determine whether it is consistent with the Standard Model expectation, we focus on the decay mode $B_s \to J/\psi\phi$, because of the *CP* properties of the final state.

This chapter begins by describing the theoretical framework of the weak interaction, focusing on the violation of *CP* symmetry in the B_s system. We then describe the decay mode $B_s \to J/\psi\phi$ and its physical observables, and review the current experimental status of the study of *CP* violation in the B_s system. Chapter 2 describes the experimental apparatus used to effect high-energy $p\bar{p}$ collisions and detect the product of those collisions, including a discussion of the main data acquisition system. Chapter 3 gives a detailed description of the data sample selection methods used in this dissertation. Chapter 4 discusses the measurement of lifetime and width difference in the B_s system. Chapter 5 explains the methods used for flavor-tagging, an essential tool in this analysis that allows us to infer the matter or antimatter content of B_s mesons at production time. Chapter 6 details the simultaneous measurement of the width difference and *CP*-violating phase in the decay mode $B_s \to J/\psi\phi$. We conclude in Chapter 7 with a summary of the results we obtain and a discussion of future prospects.

1.1 The Weak Interaction

Both leptons and quarks undergo processes involving the weak interaction. A very wide range of experimental avenues is consequently available to test the predictions of electroweak theory. Rather than providing a full theoretical description or an exhaustive catalogue of experimental tests of Standard Model weak processes, we describe in this section the main points of the chronological development of weak theory, and then focus on weak interactions in the quark sector, paying particular attention to the violation of *CP* symmetry in neutral mesons.

1.1.1 General Historical Background

The weak interaction was first proposed in 1934 by Fermi [4] to explain the phenomenology of β-decay of the neutron to a proton, electron and neutrino: $n \rightarrow p + e^- + \overline{\nu_e}$. The neutrino, whose existence had only recently been postulated by Pauli [5] but would not be observed for two more decades [6], played a key role in the theory: the presence of a massless and unobserved decay product was indispensable in the correct calculation of the energy spectrum of the electron. In terms of the quark model and electroweak theory, this process is the result of the underlying interaction $d \rightarrow u + e^- + \overline{\nu_e}$, shown in Figure 1-1. Written as such, it demonstrates the first of several symmetry violations by the weak interaction: non-conservation of quark flavor by charged weak processes.

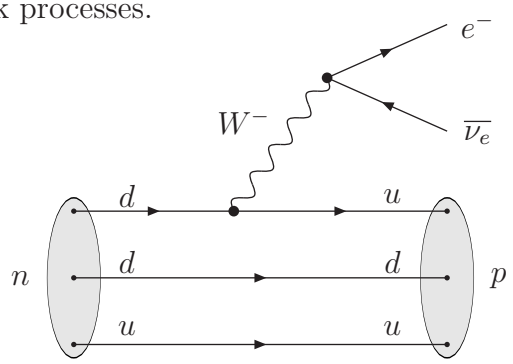


Figure 1-1: Beta decay in terms of constituent quarks of the proton and neutron.

Throughout its history, the study of the weak interaction has involved the study of the non-conservation of quantities preserved by the electromagnetic and strong interactions. In the 1950s, Gell-Mann and Nishijima invoked the weak interaction to explain the slow decays of the K meson, Σ^0 and Λ^0 baryons produced in the first modern particle accelerators [7]. All three were dubbed strange particles in light of their comparatively slow decays, which occur on a time scale of 10^{-10} seconds as opposed to the electromagnetic and strong decays which occur on a time scale of 10^{-21} and 10^{-23} seconds respectively. In addition, the electromagnetic and strong interactions were known to preserve strangeness, creating strange particles in particle-antiparticle pairs. However, the decays proceeding by the weak interaction produced final states without strange quarks. This formed the second example of non-conservation by the weak interaction: weak interactions do not respect quark generations.

In 1956, T.D. Lee and C.N. Yang suggested several experiments to test parity conservation in β decay and meson decays [8]. In 1957, C.S. Wu was the first to perform one of the suggested experiments, and showed that the weak interaction violates parity symmetry [9]. Wu prepared radioactive Co^{60} nuclei with all nuclear spins aligned. In the β-decay of the nuclei, she observed that electrons are emitted in the direction of nuclear spin, but not in the opposite direction. The discovery of the violation of parity was a particularly shocking development, as parity was considered to be a fundamental symmetry of the universe. Nevertheless several proposals quickly arose for a theoretical mechanism that leads to such an interaction [10, 11, 12].

The period 1963-1970 saw the development of a consistent framework to explain the non-conservation of quark flavor and generation by the weak interaction. In 1963, Cabibbo proposed an explanation [13] to another mystery involving strangeness: weak transitions that involve a change in strangeness number occur at a lower rate than those without such a change. He suggested that the $q_1\overline{q_2}W$ vertex factors carried a factor of $\sin\theta_C$ in the former case, and $\cos\theta_C$ in the latter, with $\theta_C \sim 0.257$. This addition to the theory of weak interactions was successful in predicting the relations between several decay rates. However, the calculated decay rate for $K^0 \rightarrow \mu\mu$ far exceeded the upper limit that had been set experimentally at the time. In 1970, Glashow, Iliopoulos, and Maiani (GIM) proposed the existence of a fourth quark, c, and predicted that weak transitions from d to c involve a factor of $-\sin\theta_C$ and from d to c a factor of $\cos\theta_C$ [14]. This proposal displayed enormous foresight and more than a little audacity, as the quark model itself would not gain widespread acceptance until the 1974 discovery of the J/ψ, a $c\overline{c}$ bound state [15, 16].

The GIM mechanism lends itself to an elegant interpretation, which incorporates Cabibbo's model and provides a logical extension. Rather than coupling to the mass states of the quarks d and s, the W boson couples to rotated states, defined in Equation 1.1. The rotation matrix incorporates all factors involving the Cabibbo angle θ_C. The choice of rotating the down-type quarks is by convention only.

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \tag{1.1}$$

In the meantime in 1964, Cronin, Fitch and Christenson discovered that charged weak interactions also violate *CP* symmetry [17]. They used neutral K mesons and their antiparticle conjugate $\overline{K}$ because they form quantum superposition that were expected to be eigenstates of *CP*. These eigenstates transform into one another via weak interaction, and their lifetimes differ by a large factor, allowing the separate observation of the decays from (expected) *CP*-even and *CP*-odd states. They found that *CP* symmetry is violated in the process of oscillating between matter and anti-matter, which manifested itself as a small fraction of *CP*-even decay products from the *CP*-odd K states. The violation of *CP* symmetry in the K system was found to be a much smaller effect than the violation of C and P symmetries. However, at the time no contemporary theoretical framework, including the Cabibbo and GIM descriptions, could explain the violation of *CP* at any level. After the discovery of parity violation, *CP* symmetry had come to be viewed as a good candidate for the "true" symmetry of the universe, providing a symmetry between matter and anti-matter that could be the replacement for the mirror symmetry provided by parity conservation [18].

In 1973, Kobayashi and Maskawa proposed an extension to the Cabibbo and GIM framework that included a mechanism for *CP* violation in the quark sector [2]. In the Cabibbo-Kobayashi-Maskawa (CKM) mechanism, the quark rotation matrix is 3×3, for three generations of quark doublets. In addition, the matrix elements are complex, rather than real numbers. The presence of *CP* violation is represented by a complex phase in one of the elements, which will be developed in coming sections.

The CKM mechanism does not predict the values of the individual matrix element, but several constraints apply on the relations between them, as is discussed in the following sections. In the three decades since its proposal, it has continued to fuel experimental investigation in the flavor sector. At the time of its proposal, the bottom and top quarks had not yet been observed, nor the intermediate gauge bosons on whose existence the entire theory of the weak interaction depended.

Evidence of the existence of the bottom quark first came in the form of the discovery of the $\Upsilon(1S)$ $b\bar{b}$ resonance in 1977 [19]. The Υ is a bound state with a net bottom quantum number of zero, and thus an example of "hidden beauty", in analogy to the J/ψ meson, the $c\bar{c}$ "hidden charm" bound state discovered three years earlier. The first B meson with non-zero bottom quantum number was observed shortly thereafter in 1981 [20]. The first observation of the heavy gauge bosons $W^{\pm}$ and Z^0 occurred in 1983 [21]. In 1994, the CDF and D$\emptyset$ experiments at Fermilab completed the SM quark table with the discovery of the top quark [22, 23].

The past decade has been one of intense experimental activity in the investigation of *CP* violation in the quark sector. In 1999 the KTeV collaboration reported the first evidence of another manifestation of *CP* violation in the kaon sector: the observation of direct *CP* violation in the decays of neutral kaons [24]. This manifestation of the effect reflects the presence of a *CP* asymmetry in the decay amplitudes themselves, rather than in the oscillation between matter and antimatter. Since then, the investigation of *CP* violation has largely shifted to the B meson sector. Of particular interest in this dissertation is the observation in 2001 of *CP* violation in the decay $B_d \to J/\psi K_S$, arising from the interference of matter-antimatter oscillation and decay [25, 26]. The manifestation of *CP* violation in this decay mode has the same theoretical origin as *CP* violation in the decay mode $B_s \to J/\psi\phi$ treated in this document, and many of the experimental features necessary for its detection. The investigation of *CP* violation in meson decays continues to be a very active field of study, with a rich phenomenology yielding no less than thirteen observable parameters for experimental verification [27]. However, there has been until now a marked paucity of experimental information on *CP* violation in the B_s sector. As we will see in the remainder of this chapter, a measurement of the *CP*-violating phase in B_s decays has not only been a sorely lacking piece in the verification of Standard Model predictions, but is also fertile ground in the search for new physics beyond the Standard Model.

1.1.2 The CKM Mechanism

The Lagrangian describing charged-current interactions for quarks is written:

$$\mathcal{L}_{W^{\pm}} = \frac{g}{\sqrt{2}} \, \overline{u_{Li}} \, \gamma^{\mu} \, (V_{\mathrm{CKM}})_{ij} \, d_{Lj} \, W^{+}_{\mu} + \text{hermitian conjugate}, \tag{1.2}$$

where g is the constant associated with the $\mathrm{SU(2)_L}$ gauge group indicating the strength of the coupling, and the indices i, j represent the quark generations. This is represented graphically by the vertex shown in Figure 1-2.

The set of complex vertex factors $(V_{\mathrm{CKM}})_{ij}$ in Equation 1.2, when arranged as a $3\times$ 3 matrix, form the Cabibbo-Kobayashi-Maskawa quark mixing matrix. The elements of the CKM matrix specify the coupling in each transition between quarks i, j. The matrix rotates the quark flavor eigenstates into the weak interaction eigenstates. By the same convention as the Cabibbo-GIM description, the down-type quarks are rotated:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{1.3}$$

where the primed quarks correspond to the weak eigenstates and the unprimed quarks correspond to the mass (strong) eigenstates. A 3×3 matrix of complex numbers has $2 \cdot 3^2 = 18$ real parameters. However, the CKM matrix has only 4 free parameters in the SM due to the following:

- To preserve gauge invariance in the electroweak interaction, it is required that V_{CKM} be unitary: $V_{\mathrm{CKM}}^{\dagger} V_{\mathrm{CKM}} = 1$. This reduces the number of free parameters to nine: three real angles and six phases.

- The freedom to redefine the phases of the six quark fields by a transformation $u_i \rightarrow e^{i\xi} u_i$ allows us to eliminate five of the six phases leaving only one physical phase.

The four free parameters are traditionally interpreted as three rotational angles and one phase, which leads to the following parametrization of the matrix [28]:

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

where $c_{ij} \equiv \cos(\theta_{ij}), s_{ij} \equiv \sin(\theta_{ij})$. This convention yields the three rotation angles $\theta_{12}, \theta_{23},$ and θ_{13} and one phase δ.

An alternate parametrization originally proposed by Wolfenstein [29] emphasizes the relative magnitudes of the individual elements of the matrix. In this parametrization, we redefine the parameters to be $A,\ \lambda,\ \rho$ and η, with η playing the role of the

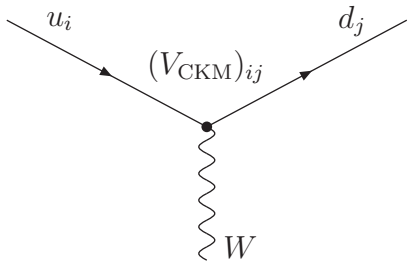


Figure 1-2: Feynman vertex for flavor-changing charged current interactions.

CP-violating phase and λ being used as an expansion parameter. The parametrization is valid at all orders. Here we expand the terms in $\lambda \equiv \sin(\theta_{12}) \approx 0.23$ ignoring terms of $\mathcal{O}(\lambda^6)$ or smaller, yielding:

$$\mathbf{V} \approx \begin{pmatrix} 1 - \frac{\lambda^2}{2} - \frac{\lambda^4}{8} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda + \frac{A^2\lambda^5}{2}\left(1 - 2(\rho + i\eta)\right) & 1 - \frac{\lambda^2}{2} - \frac{\lambda^4}{8}(1 + 4A^2) & A\lambda^2 \\ A\lambda^3\left(1 - (1 - \frac{\lambda^2}{2})(1 - \rho - i\eta)\right) & -A\lambda^2 + \frac{A\lambda^4}{2}\left(1 - 2(\rho + i\eta)\right) & 1 - \frac{A^2\lambda^4}{2} \end{pmatrix} \tag{1.5}$$

where $\eta \equiv \sin(\theta_{13})\sin\delta/A\lambda^3$, $\rho \equiv \sin(\theta_{13})cos(\delta)/A\lambda^3$, and $A \equiv \sin(\theta_{23})/\lambda^2$. The Wolfenstein parametrization illustrates the fact that transitions across two quark generations (e.g. $b \to u$), or more generally transitions that involve further off-diagonal elements of the CKM matrix are suppressed with respect to transitions involving diagonal elements. All physical quantities are naturally independent of the chosen parametrization.

Unitarity Condition

The free parameters of the CKM matrix are inputs to the Standard Model, meaning that they are not specified by theory and must be determined by experiments. However, if the Standard Model is the correct description of the universe, the unitary of the CKM matrix must hold. A common experimental approach is to test this condition. To explain the details of this approach, we begin by writing the unitarity condition as follows:

$$\sum_{i=1}^{3} V_{ij}V_{ik}^* = \sum_{i=1}^{3} V_{ji}V_{ki}^* = \delta_{ij} \tag{1.6}$$

for any $j, k \in \{1, 2, 3\}$. Expanding Equation 1.6 for any j, k yields nine equations, of which the six equations involving the off-diagonal elements of δ_{ij} describe triangles in the complex plane. These six triangles fall into two groups of three, differing only by their orientation in the complex plane. If we take the three triangles that arise out of the products of columns of V_{CKM}, we obtain the following equations:

$$(V^\dagger V)_{31} : V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0, \tag{1.7}$$

$$(V^\dagger V)_{32} : V_{ub}^* V_{us} + V_{cb}^* V_{cs} + V_{tb}^* V_{ts} = 0, \tag{1.8}$$

$$(V^\dagger V)_{21} : V_{us}^* V_{ud} + V_{cs}^* V_{cd} + V_{ts}^* V_{td} = 0. \tag{1.9}$$

The study of processes involving flavor-changing charged weak interactions, e.g. matter-antimatter oscillations of mesons or weak decays, allows for the measurements of physical observables (oscillation frequencies and decay rates in our examples) that depend on real quantities such as the moduli of elements $|V_{ij}|$ in various combinations. In turn, these measurements can be converted into measurements of the length of the sides and interior angles of the unitarity triangles. By measuring all sides and interior angles, we over-constrain the triangles and test whether unitarity holds.

In an experimental setting, large interior angles of such unitarity triangles lend

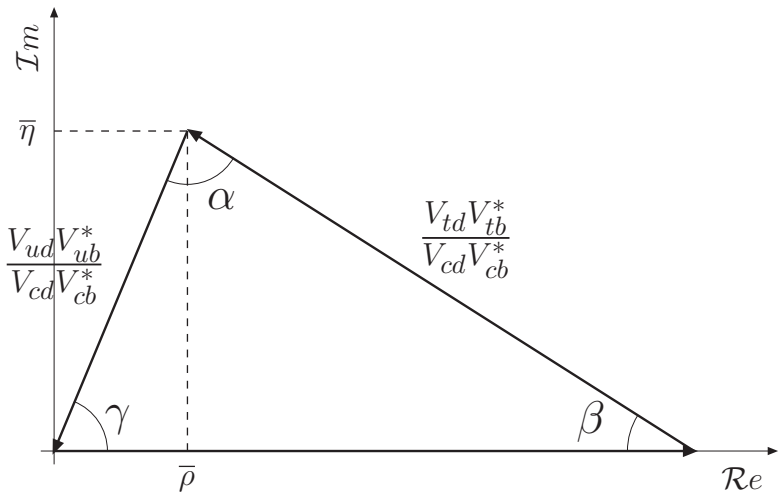


Figure 1-3: Normalized unitarity triangle obtained from $V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0$.

themselves to easier measurement. It is for this reason that the most well studied triangle is the one arising from Equation 1.7, although there is no *a priori* theoretical reason to favor it over others. The recent literature usually refers to this triangle as "The Unitarity Triangle." In the usual representation of the unitarity triangle, shown in Figure 1-3, one of the vertices is at $(0,0)$ in the complex plane, and the sides are normalized such that the other two vertices are placed at $(1,0)$ and $(\overline{\rho},\overline{\eta})$, where:

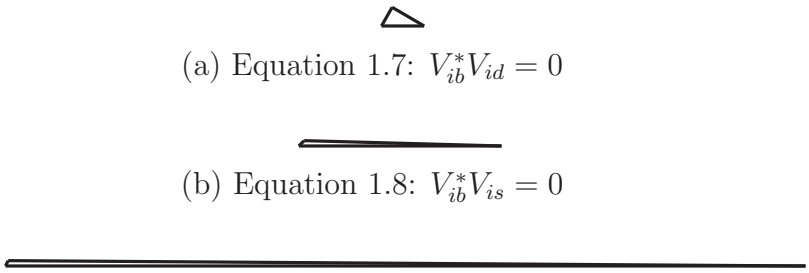
$$\overline{\rho} \equiv \left(1 - \frac{\lambda^2}{2}\right)\rho, \qquad \overline{\eta} \equiv \left(1 - \frac{\lambda^2}{2}\right)\eta.$$

The angles of the unitarity triangle are defined by:

$$\phi_1 \equiv \beta = arg\left(-\frac{V_{cb}^* V_{cd}}{V_{tb}^* V_{td}}\right)$$
$$\phi_2 \equiv \alpha = arg\left(-\frac{V_{tb}^* V_{td}}{V_{ub}^* V_{ud}}\right)$$
$$\phi_3 \equiv \gamma = arg\left(-\frac{V_{ub}^* V_{ud}}{V_{cb}^* V_{cd}}\right)$$

For comparison, the three triangles obtained from Equations 1.7-1.9 are depicted in a common scale in Figure 1-4. The triangle of interest in this dissertation is the one formed by Equation 1.8. The value of its smallest angle, which we call β_s by analogy to the canonical unitarity triangle, is proportional to the phase of the transition amplitude for $B_s - \overline{B_s}$ oscillation. It is defined as $\beta_s \equiv \arg\left(-V_{ts}V_{tb}^*/V_{cs}V_{cb}^*\right)$.

The magnitude of the transition amplitude for B_s oscillations is proportional to Δm_s. The oscillation frequency Δm_s is related to CKM parameters according to: $\Delta m_s \propto |V_{tb}^* V_{ts}|^2$, where the proportionality constant is a coefficient that includes hadronic matrix elements that must be calculated using lattice QCD.



(a) Equation 1.7: $V_{ib}^* V_{id} = 0$

(b) Equation 1.8: $V_{ib}^* V_{is} = 0$

(c) Equation 1.9: $V_{is}^* V_{id} = 0$

Figure 1-4: The three unitarity triangles obtained from the products of columns of $V_{\rm CKM}$ and the unitarity condition (Equation 1.6). The triangles are drawn on the complex plane, and to a common scale.

CP Violation in the CKM mechanism

Symmetry with respect to the CPT operator where the T operator represents time reversal is a basic assumption of the quantum field theories that make up the Standard Model. The Standard Model does not respect T symmetry, and therefore implies a violation of CP. Although a discussion of CP violation does not have to assume CPT invariance, we limit ourselves in this discussion to cases where CPT symmetry holds.

The presence of a CP violating phase or phases in the quark sector requires the existence of at least three generations of quarks. For two generations, the freedom to redefine the quark field permits the removal of all complex phases from a 2×2 quark rotation matrix. Furthermore, it requires that the up-type (M^u) and down-type (M^d) quark mass matrices be non-degenerate, i.e. that no up-type quark have the same mass as another up-type quark, and likewise for down-type quarks. If such a degeneracy existed, the remaining physical phase in the CKM matrix could be removed by the appropriate choice of a unitary transformation of the quark fields. A measure of the CP violation in the SM that is independent of the chosen phase convention and includes the previous condition is given by [30]:

$$\begin{aligned} \mathcal{I}m \, \det\left([M^u M^{u\dagger}, M^d M^{d\dagger}]\right) = 2J(m_t^2 - m_c^2)(m_t^2 - m_u^2)(m_c^2 - m_u^2) \\ \times(m_b^2 - m_s^2)(m_b^2 - m_d^2)(m_s^2 - m_d^2). \end{aligned} \tag{1.10}$$

The dependence on V_{ij} is contained in the Jarlskog invariant J:

$$\mathcal{I}m \left(V_{ij} V_{kl} V_{il}^\dagger V_{kj}^\dagger\right) = J \sum_{m,n=1}^{3} \varepsilon_{ikm} \varepsilon_{jln}, \tag{1.11}$$

where ε_{ikm} is the total antisymmetric tensor. In terms of the standard parametrization $J = c_{12} c_{23} c_{13}^2 s_{12} s_{23} s_{13} \sin\delta$.

The observation of CP violation arising from this description relies on the study of interference phenomena because it is the only way to detect effects that are caused by the presence of phases in the Lagrangian.

1.2 *B* Meson Lifetimes

The measurements of the B_s and B_d lifetimes play an important role in the analysis presented in this document. In the case of $B_s \to J/\psi\phi$, we measure the average lifetime τ_s and the width difference $\Delta\Gamma_s$ between the two mass eigenstates (see Section 1.3). This initial measurement is an important test of our analysis framework. In the subsequent measurement of *CP* violation in $B_s \to J/\psi\phi$, τ_s and $\Delta\Gamma_s$ must be measured simultaneously since they are correlated with $2\beta_s$. In the case of the $B_d \to J/\psi K^*$ decay, we measure the average lifetime τ_d. Since the $B_d \to J/\psi K^*$ mode acts as a control sample in our analysis, the measurement of lifetime, which is correlated with the angular amplitudes, provides an additional validation of our method. The B_d lifetime and angular amplitudes have been measured independently at several collider experiments, allowing detailed comparisons with our results. In this section, we provide a brief theoretical introduction to B meson lifetimes. The lifetime measurements we perform are not accompanied by the full suite of calculation of systematic effects characteristic of dedicated lifetime analyses, as our focus remains on *CP* violation effects. For the same reason, this theoretical introduction is kept minimal. In the Standard Model, a b quark decays to a c or u quark through a virtual W boson. The decay width is given by:

$$\Gamma_{q_1\bar{q}_2}(b \to q) = \frac{3G_F^2 m_b^5}{192\pi^2} \left|V_{q_1 q_2}\right|^2 \left|V_{qb}\right|^2 F(\epsilon_q), \tag{1.12}$$

where G_F is the Fermi coupling constant and $F(\epsilon_q)$ is the factor associated with phase space in the decay. Equation 1.12 provides the source of the connection between fundamental SM parameters and measurements of lifetime $\tau \equiv 1/\Gamma$.

B meson lifetime are described to first order by the Spectator Model, where a heavy quark (here the b) is bound to a lighter spectator quark. In this model, the heavy quark dominates the lifetime, and the lifetimes of various B mesons should be the same. This is contradicted by observations, which give the following hierarchy:

$$\tau_{B_c} < \tau_{B_s} \approx \tau_{B_d} < \tau_{B_u}. \tag{1.13}$$

More precise predictions of hadron lifetimes must take into account additional considerations beyond Equation 1.12, since the latter makes a prediction only for free quarks, not for bound states which will be affected by quark-quark interactions.

Heavy Quark Expansion theory (HQE) models the differences in B meson lifetimes. Using QCD, the decay rate is expressed as an expansion series in $1/m_b$. Leading order calculations in HQE reproduce the results of the Spectator Model. Terms of order $1/m_b^2$ are ignored because they refer to higher order corrections associated with the b quark alone. Terms of order $1/m_b^3$ arise from interactions among two of the constituent quarks in the hadron, which can be categorized as follows for B mesons.

Pauli Interference (PI), shown in Figure 1-5, affects the decays of B_u mesons, as the decay products of both the external and color-suppressed internal W decays are the same, which allows for interference. For B_d mesons, the two decays proceed with different decay products, and thus there is no interference. The interference is de-

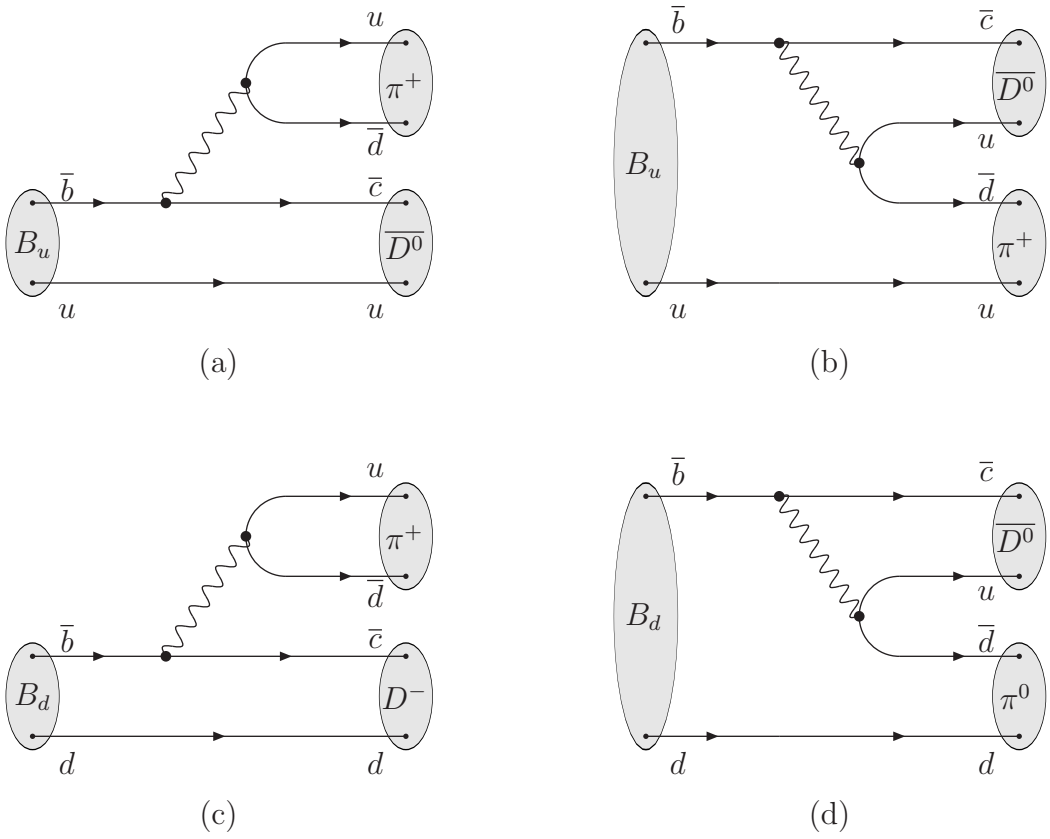


Figure 1-5: Pauli interference in the decay of B_u and B_d.

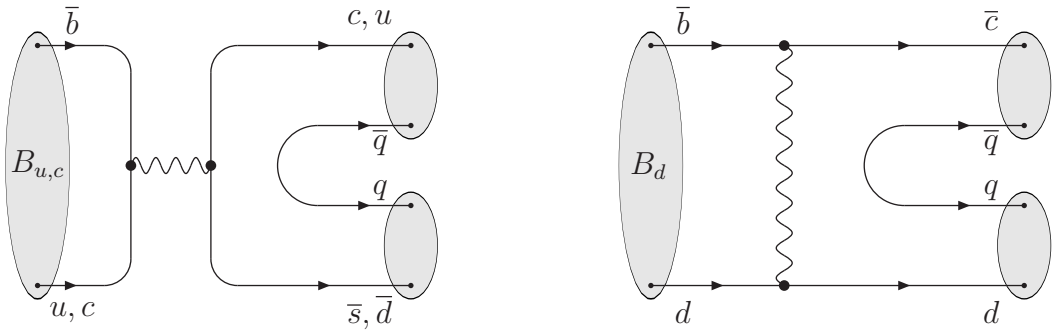


Figure 1-6: Weak Annihilation and Weak Exchange in the decay of B_u, B_c and B_d.

structive for B_u: its effect is to lengthen the lifetime in comparison to a decay without PI. Weak Annihilation (WA), shown is Figure 1-6 (left) is possible only for charged B mesons, and provides an additional decay mechanism, thus shortening the lifetimes of B_u and B_c mesons. Weak Exchange (WE), shown is Figure 1-6 (right) is possible only for neutral B mesons and baryons. However, this effect is helicity suppressed: the $cd(cs)$ spin is restricted by the fact that the $B_d(B_s)$ has zero spin. Therefore, in a meson, the q and $\overline{q}$ have opposite helicity, which suppresses this decay mode. For B_u decays, the effect of PI in lengthening the lifetime is stronger than the effect of WA in shortening it because of the relative magnitudes of the matrix elements involved, a full derivation of which is beyond the scope of this discussion.

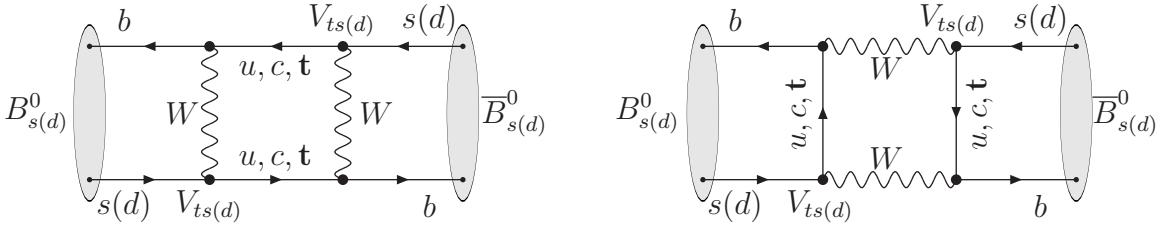


Figure 1-7: Leading order Feynman diagrams contributing to B_s flavor oscillations.

1.3 *CP* Violation in Mixing and Decay of B_s

A B meson is a bound state of an anti-b quark and another quark q, written in Dirac notation as $|\bar{b}q\rangle$, or simply $|B\rangle$. Neutral mesons such as the B_s, B_d and K mesons have the property that they transform into their own anti-particle and back, a phenomenon known as mixing. The leading order Feynman diagrams contributing to B_s mixing are shown in Figure 1-7. The study of the oscillation and decay of neutral mesons yields observables that are functions of the CKM matrix elements that appear at the vertices of the interactions in question. In this section, we first describe mixing and decay of neutral mesons with a general formalism using the properties of matter-antimatter oscillation. This formalism is largely independent of the details at the interaction vertices, describing equally well all neutral mesons. We subsequently refine our focus on B_s mixing, lifetime difference and *CP* violation in the decay $B_s \to J/\psi\phi$.

1.3.1 General Time Development

As a result of $B - \overline{B}$ oscillations, a pure B or $\overline{B}$ state at time 0, such as the meson states created as a result or $p\overline{p} \to \bar{b}b$ interactions at the Tevatron, will evolve to be a superposition of B and $\overline{B}$ after a time t:

$$|\psi(t)\rangle = a(t)|B\rangle + b(t)|\overline{B}\rangle. \tag{1.14}$$

We describe the time evolution of such a system, including oscillations and decay, using a 2×2 matrix formalism known as the Weisskopf-Wigner approximation [31]. In this approximation, the Hamiltonian is written as the sum of two matrices forming a non-hermitian complex matrix, in order to accommodate the description of decay:

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^* & M_{22} \end{pmatrix} - \frac{i}{2}\begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{22} \end{pmatrix}, \tag{1.15}$$

where $\mathbf{M}$ and $\mathbf{\Gamma}$ are themselves complex and hermitian. The eigenstates of this Hamiltonian, called the mass eigenstates, are written as:

$$|B_{L,H}\rangle = p|B\rangle \pm q|\overline{B}\rangle, \tag{1.16}$$

where the subscripts L and H stand for light and heavy.

In general, the mass eigenstates are not the same as the *CP* eigenstates, which

are defined as

$$\begin{aligned} |B^{\text{even}}\rangle &= \frac{1}{\sqrt{2}}\left(|B\rangle - |\overline{B}\rangle\right), \\ |B^{\text{odd}}\rangle &= \frac{1}{\sqrt{2}}\left(|B\rangle + |\overline{B}\rangle\right). \end{aligned} \tag{1.17}$$

Since $\mathbf{M}$ and $\mathbf{\Gamma}$ are Hermitian, their off-diagonal elements are related by $M_{ij}^* = M_{ji}$, $\Gamma_{ij}^* = \Gamma_{ji}$ for any i or j. Furthermore, the requirement that the Lagrangian respect CPT symmetry forces $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$. Using this information, we solve the eigenvalue problem and find that the eigenvalues $\alpha_{H,L}$ of the Hamiltonian are equal to:

$$\alpha_{H,L} = \left(M - \frac{i}{2}\Gamma\right) \pm \sqrt{\left(M_{12} - \frac{i}{2}\Gamma_{12}\right)\left(M_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)} \tag{1.18}$$

where $\Gamma \equiv \Gamma_{11} = \Gamma_{22}$ and $\mathrm{M} \equiv \mathrm{M}_{11} = \mathrm{M}_{22}$. We substitute these solutions back into the eigenvalue problem to obtain the ratio $(q/p)^2$, which defines the eigenvectors:

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

The mass difference and width difference of the heavy and light states are written:

$$\Delta m \equiv m_H - m_L = \mathcal{R}e(\alpha_H - \alpha_L) \tag{1.20}$$

$$\Delta\Gamma \equiv \Gamma_L - \Gamma_H = 2\,\mathcal{I}m(\alpha_H - \alpha_L). \tag{1.21}$$

We now write the time evolution of an initially pure B ($\overline{B}$) state after time t:

$$\begin{aligned} |B(t)\rangle &= g_+(t)|B\rangle - \frac{q}{p}g_-(t)|\overline{B}\rangle, \\ |\overline{B}(t)\rangle &= g_+(t)|\overline{B}\rangle - \frac{p}{q}g_-(t)|B\rangle, \end{aligned} \tag{1.22}$$

where

$$g_\pm \equiv \frac{1}{2}\left(e^{-im_H t - \frac{1}{2}\Gamma_H t} \pm e^{-im_L t - \frac{1}{2}\Gamma_L t}\right). \tag{1.23}$$

Before passing to a theoretical examination of CP violation in the B_s system, we note that the width difference and the mass difference between the heavy and light eigenstates play very different roles in this measurement, which is not obvious from the above general treatment. There is significant overlap between the methods used to extract $\Delta\Gamma$ and those used to extract the CP-violating phase in $B_s \to J/\psi\phi$ decays, including the angular analysis described in Section 1.4. In essence, the dynamics of this decay mode impose on us the need to account for and measure $\Delta\Gamma$ in order to measure the CP-violating phase. However, the measurement of the mass difference is done using fundamentally different methods employing a Fourier analysis and accounting for different systematic effect. Therefore, although Δm will enter into the equations we use in this analysis, it is treated as an input taken from the world

average of recent measurements, unlike $\Delta\Gamma$, which can be treated as a simultaneous measurement in this dissertation.

1.3.2 *CP* Violation Observables

We have stated that *CP* violation effects come from irreducible phases in the Lagrangian governing interactions. We now investigate how such phases are introduced in the decay of a B or $\overline{B}$ into a final state f, and how they manifest themselves as an interference phenomenon in the case of $B_s \to J/\psi\phi$. In the general case, we define the amplitudes coming from such decays as

$$A_f = \langle f|\mathcal{H}|B\rangle, \quad \overline{A}_f = \langle f|\mathcal{H}|\overline{B}\rangle, \tag{1.24}$$

where $\mathcal{H}$ is the Hamiltonian governing weak decays.

Combining Equations 1.22– 1.24, we write the decay rates $|\langle f|B(t)\rangle|^2$ and $\left|\langle f|\overline{B}(t)\rangle\right|^2$:

$$\begin{aligned}|\langle f|B(t)\rangle|^2 \cdot e^{\Gamma t} = \\ \left(|A_f|^2 + \left|(q/p)\overline{A}_f\right|^2\right)\cosh(\Delta\Gamma t/2) + \left(|A_f|^2 - \left|(q/p)\overline{A}_f\right|^2\right)\cos(\Delta m t) + \\ 2\mathcal{R}e((q/p)A_f^*\overline{A}_f)\sinh(\Delta\Gamma t/2) - 2\mathcal{I}m((q/p)A_f^*\overline{A}_f)\sin(\Delta m t).\end{aligned} \tag{1.25}$$

$$\begin{aligned}\left|\langle f|\overline{B}(t)\rangle\right|^2 \cdot e^{\Gamma t} = \\ \left(|(p/q)A_f|^2 + \left|\overline{A}_f\right|^2\right)\cosh(\Delta\Gamma t/2) - \left(|(p/q)A_f|^2 - \left|\overline{A}_f\right|^2\right)\cos(\Delta m t) + \\ 2\mathcal{R}e((p/q)A_f\overline{A}_f^*)\sinh(\Delta\Gamma t/2) - 2\mathcal{I}m((p/q)A_f\overline{A}_f^*)\sin(\Delta m t).\end{aligned} \tag{1.26}$$

Similar decay rates for $\left|\langle \overline{f}|B(t)\rangle\right|^2$, and $\left|\langle \overline{f}|\overline{B}(t)\rangle\right|^2$ are obtained by replacing A_f by $A_{\overline{f}}$ and $\overline{A}_f$ by $\overline{A}_{\overline{f}}$ where $A_{\overline{f}} = \langle \overline{f}|\mathcal{H}|B\rangle$ and $\overline{A}_{\overline{f}} = \langle \overline{f}|\mathcal{H}|\overline{B}\rangle$. We focus on equations 1.25 and 1.26 in anticipation of the discussion of decays of both B_s and $\overline{B_s}$ to $J/\psi\phi$.

From straightforward inspection of the relevant equations, we see that there are several contributions to Equations 1.25 and 1.26. These contributions are of two general categories, depending on whether the meson oscillates between matter and antimatter an even (no net oscillation) or an odd (net oscillation) number of times between creation of the meson bound state and its decay. Terms proportional to $|A_f|^2$, and $\left|\overline{A}_f\right|^2$ are associated with decays where no net $B - \overline{B}$ oscillations occurs. Terms proportional to $\left|(q/p)\overline{A}_f\right|^2$ and $|(p/q)A_f|^2$ come from decays where there is a net oscillation between production and decay. The sine and hyperbolic sine terms in each equation arise from the interference between these two cases.

We classify *CP* violating effects into three phenomenological types: *CP* violation in decay, in mixing, and in the interference between decays with net mixing and decays without net mixing. We focus here on the B_s sector in the discussion of the three classes of *CP* violation [27]:

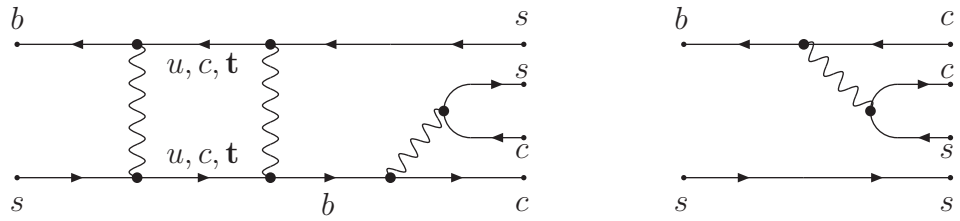


Figure 1-8: The $B_s \to J/\psi\phi$ decay mode allows for interference of mixing and decay.

1. *CP* violation in decay, also called *direct CP* violation, defined by

$$\left|A_f/\overline{A}_f\right| \neq 1. \tag{1.27}$$

In the Standard Model, the weak phase contributing to direct *CP* violation in B_s decays is Cabibbo-suppressed by λ^2, where λ is the Wolfenstein parameter [32]. Furthermore, new physics contributions to this phase have to compete with a tree diagram, and therefore are not expected to provide a sufficiently large experimental signature [33]. We therefore assume that $|A_f| = |\overline{A}_f|$ in B_s decays, ignoring direct *CP* violation.

2. *CP* violation in mixing, defined by

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

In the B_s system, $|q/p| - 1 < \mathcal{O}(10^{-2})$ [34]. The presence of *CP* violation in mixing results in a charge asymmetry in semileptonic B_s decays such as $B_s \to \mu^+ D_s^- \nu X$ and its conjugate $\overline{B_s}$ mode. The resultant time-integrated decay asymmetry is sensitive to deviations of $|q/p|$ from 1. The decay rate of $B_s \to J/\psi\phi$ does not isolate factors of $|q/p|$, and we therefore do not speak strictly of *CP* violation in mixing. Nevertheless, both sets of observables are related to the same fundamental phase and results can therefore be combined from both measurements.

3. *CP* violation in the interference between decay without net mixing and decay with net mixing, defined by

$$\mathcal{I}m(\lambda_f) \neq 0, \qquad \lambda_f \equiv \frac{q}{p}\frac{\overline{A}_f}{A_f}. \tag{1.29}$$

This type of *CP* violation is present only in modes for which the final state is available to both B_s and $\overline{B_s}$, therefore including $B_s \to J/\psi\phi$. Figure 1-8 shows the Feynman diagrams demonstrating the interference in this decay mode.

With these definitions, we rewrite Equation 1.25 in terms of λ_f for the decay of a

B_s meson into a final state f:

$$|\langle f|B_s(t)\rangle|^2 \propto e^{-\Gamma_s t}|A_f|^2 \cdot \big[\cosh(\Delta\Gamma_s t/2)+ \\ \mathcal{R}e(\lambda_f)\sinh(\Delta\Gamma_s t/2) - \mathcal{I}m(\lambda_f)\sin(\Delta m_s t)\big], \quad (1.30)$$

where we have made use of $|A_f| = |\overline{A}_f|$ to set $|\lambda_f| = 1$, ignoring the effect of *CP* violation in decay. This leaves as the only source of *CP* violation the interference of the two decays shown in Figure 1-8.

1.3.3 Standard Model Expectation and New Physics

With all the necessary theoretical preparation in place, we now examine the Standard Model and possible new physics contributions to *CP* violation in the B_s system. We recall Equation 1.15, the effective Hamiltonian used to describe neutral meson dynamics, and focus on Γ_{12} and M_{12}. As shown in Figures 1-7 and 1-8, contributions to the former are dominated by tree-level CKM-favoured decays, while the latter is induced primarily by short distance physics by way of a virtual top quark, making it much more sensitive to non-SM contributions [33]. Using experimental information that $\Delta\Gamma_s \gg \Delta m_s$, we know that $|\Gamma_{12}| \ll |M_{12}|$. We use the solutions to the eigenvalue problem in Equation 1.18 and an expansion in Γ_{12}/M_{12} to rewrite the mass difference, width difference, and q/p as

$$\Delta m_s = 2|M_{12}|, \quad \Delta\Gamma_s = 2|\Gamma_{12}|\cos\phi, \quad \frac{q}{p} = -e^{i\phi_M}\left[1 - \frac{1}{2}\left|\frac{\Gamma_{12}}{M_{12}}\right|\sin\phi\right], \quad (1.31)$$

where we have defined

$$\phi_M \equiv \arg(M_{12}), \quad \text{and} \quad \frac{M_{12}}{\Gamma_{12}} \equiv -\left|\frac{M_{12}}{\Gamma_{12}}\right| e^{i\phi}, \quad (1.32)$$

and where we have discarded terms of order Γ_{12}^2/M_{12}^2 or higher. In the Standard Model, the phase $\phi = \phi_M - \phi_\Gamma$ can be expressed in terms of the CKM matrix elements that contribute to the respective mixing and decay diagrams:

$$\begin{aligned} \phi^{\rm SM} &= \arg(V_{tb}V_{ts}^*)^2 - \arg(-\Gamma_{12}) \\ &\approx \arg(V_{tb}V_{ts}^*)^2 - \arg(V_{cb}V_{cs}^*)^2 \end{aligned} \quad (1.33)$$

Because $\arg(V_{tb}V_{ts}^*)$ and $\arg(V_{cb}V_{cs}^*)$ are very close and the remaining corrections to $\arg(\Gamma_{12})$ involving $V_{ub}V_{us}^*$ are small and suppressed by a factor of m_c^2/m_b^2, $\phi^{\rm SM}$ is predicted to be very small, $\phi^{\rm SM} = 0.0041 \pm 0.0008$ [35].

Ignoring the effect of direct *CP* violation, we can also write

$$\frac{\overline{A}_f}{A_f} = \mp e^{i\arg(V_{cb}V_{cs}^*)^2} \quad (1.34)$$

where the upper (lower) sign is for a *CP*-even (-odd) final state f. Combining this

with Equation 1.31, we find

$$\mathcal{R}e(\lambda_f) = \mathcal{R}e\left(\frac{q}{p}\frac{\overline{A}_f}{A_f}\right) = \sin\phi, \quad \text{and} \quad \mathcal{I}m(\lambda_f) = \mathcal{I}m\left(\frac{q}{p}\frac{\overline{A}_f}{A_f}\right) = -\cos\phi \tag{1.35}$$

for a *CP*-even final state, with the signs reversed for *CP*-odd. We thereby identify the phase ϕ that is responsible for *CP* violation in the interference of mixing and decay in $B_s \to J/\psi\phi$ with the phase of the B_s mixing amplitude.

Traditionally, the phase β_s, defined in analogy to the angle β as $\beta_s \equiv arg\left(-\frac{V_{cd}V_{cb}^*}{V_{tb}^*V_{ts}}\right)$, is associated with *CP* violation in the B_s system, as it is the analogous angle of the B_s unitarity triangle. The phase β_s is predicted by the Standard Model to be equal to $\beta_s^{\rm SM} = 0.04 \pm 0.01$ [36]. We can relate β_s and ϕ in the following manner. First, we define SM and new physics (NP) contributions to ϕ. With the phase conventions we have chosen, the NP contribution to ϕ and β_s is:

$$\phi = \phi^{\rm SM} + \phi^{\rm NP}, \quad \text{and} \quad 2\beta_s = 2\beta_s^{\rm SM} - \phi_s^{\rm NP}. \tag{1.36}$$

Since both $\beta_s^{\rm SM}$ and $\phi^{\rm SM}$ are expected to be negligibly small, in the presence of new physics, we neglect the SM contribution, and use the shorthand

$$2\beta_s \approx -\phi. \tag{1.37}$$

1.4 Angular Analysis of $P \to VV$ Decay of B_s and B_d

In this section, we discuss the final theoretical ingredient necessary for a measurement of the *CP*-violating phase ϕ. In the absence of direct *CP* violation, when the final state of the B decay is pure *CP* eigenstate, denoted f_{CP}, it is possible to measure the asymmetry $\mathcal{A}_{f_{CP}}$ of neutral meson decays into f_{CP}:

$$\mathcal{A}_{f_{CP}}(t) = \frac{d\Gamma/dt[\overline{B}(t) \to f_{CP}] - d\Gamma/dt[B(t) \to f_{CP}]}{d\Gamma/dt[\overline{B}(t) \to f_{CP}] + d\Gamma/dt[B(t) \to f_{CP}]} \tag{1.38}$$

If $\Delta\Gamma = 0$ and $|q/p| = 1$ are good approximations as in the case of $B_d \to J/\psi K_S$, this reduces to a simple expression, showing that $\mathcal{A}_{f_{CP}}$ is identically zero unless $\sin\phi \neq 0$:

$$\mathcal{A}_{f_{CP}}(t) \propto \sin(\Delta m t)\sin\phi. \tag{1.39}$$

However, we must use another method to measure ϕ for two reasons: $\Delta\Gamma_s$ is not equal to 0, and the final state $J/\psi\phi$ is an admixture of CP-even and CP-odd contributions rather than an eigenstate of *CP*. An observation of *CP* violation in $B_s \to J/\psi\phi$ requires a statistical determination of the odd and even contribution to final state decays. We perform this statistical separation using an angular analysis. The validation of this method is performed using the decay mode $B_d \to J/\psi K^*$, which lends itself to a similar angular analysis.

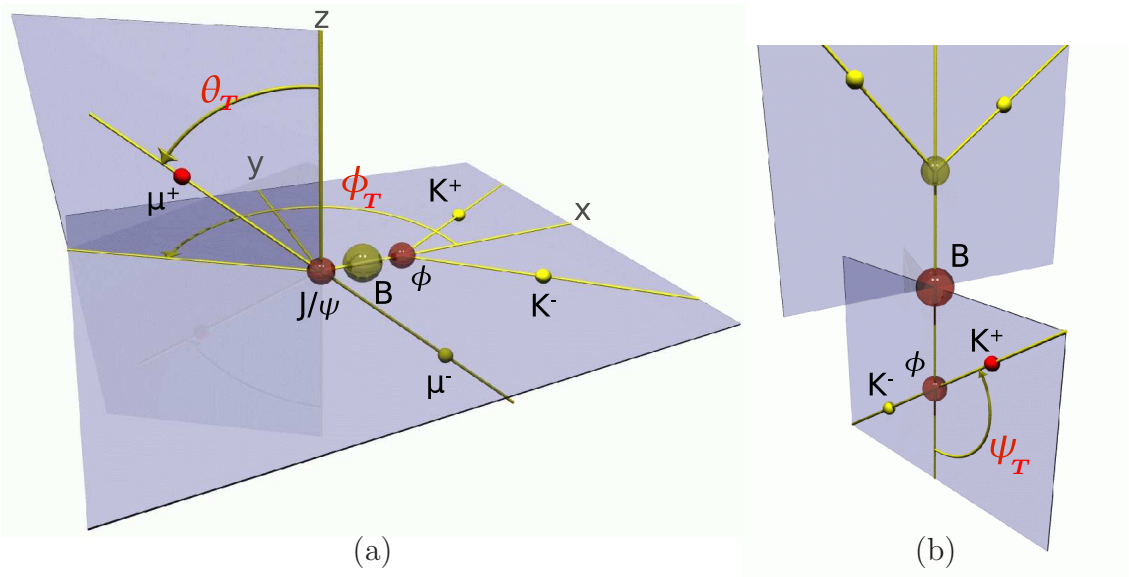
The decay modes $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$ are both pseudoscalar to vector vector ($P \to VV$) decays: the B_s (B_d) mesons each has intrinsic spin 0 while J/ψ and ϕ (K^*) have intrinsic spin of 1. We infer the parity of the final state by examining angular correlation in the decay products. In the case of $B_s \to J/\psi\phi$, because both J/ψ and ϕ are C-odd, inferring the CP state of the decay products reduces to inferring the parity of the final state. A straightforward application of conservation of momentum using the usual quantum mechanical rules shows that there are three possible values for the relative angular momentum L of the vector particles that respect conservation of total angular momentum $J \equiv S + L$: $L = 0, 1$ or 2. To each value of the relative angular momentum of the vector particles corresponds a decay amplitude. However, rather than working with states of relative angular momentum, it is mathematically more convenient to work with the relative polarization of the vector particles. This corresponds to a change to the "transversity" basis, which is convenient because it allows us to separate contributions to the decay by CP-odd and CP-even components using angular correlations [37].

The transversity basis, illustrated visually in Figure 1-9, requires a change of angular variables used to describe the decay, defined as follows. The first two angles are calculated in the rest frame of the J/ψ, and the third in the rest frame of the ϕ. In the rest frame of the J/ψ, the ϕ meson direction defines the x axis. The plane of K^+K^- defines the xy plane with $p_y(K^+) > 0$. From there:

- θ_T: in the J/ψ meson rest frame, the angle between $p(\mu^+)$ and the xy plane
- ϕ_T: in the J/ψ meson rest frame, the angle between the x axis and $p_{xy}(\mu^+)$
- ψ_T: in the ϕ meson rest frame, the angle between $p(K^+)$ and $-p(J/\psi)$

The angular variables of the $B_d \to J/\psi[\to \mu^+\mu^-]K^*[\to K\pi]$ decay can be expressed in the same transversity basis by substituting K^* for ϕ and K for K^+ in Figure 1-9. For brevity and convenience we use the symbol $\vec{\omega} = \{\cos\theta_T, \phi_T, \cos\psi_T\}$ to refer to the transversity variables together. The vector notation is simply shorthand, as $\vec{\omega}$ does not transform as a vector under rotations in space. Throughout this thesis, we drop the transversity subscripts on $\{\theta, \phi, \psi\}$ when there is no possible ambiguity.

In the transversity basis, the three decay amplitudes correspond to the different relative polarizations of the vector particles: A_0, $A_\parallel$ and $A_\perp$ for the polarizations being longitudinal, transverse and parallel, or transverse and perpendicular to one another, respectively. With each amplitude A_α is associated a phase $\delta_\alpha = \arg(A_\alpha)$. We have the freedom to redefine one the phases because only relative phases are physically detectable. We choose a convention where $\delta_0 \equiv 0$, $\delta_\parallel \equiv \arg(A_\parallel)$, and $\delta_\perp \equiv \arg(A_\perp)$. Throughout this document, the notation A_α ($|A_\alpha|$) refers to (the magnitude of) any of the three amplitudes A_0, $A_\parallel$ and $A_\perp$, while the notation $\{A_\alpha\}$ refers to all three amplitudes. Likewise, δ_α refers to any of the three phases δ_0, $\delta_\parallel$ or $\delta_\perp$, while $\{\delta_\alpha\}$ refers to all three.



(a) (b)

Figure 1-9: Definition of the transversity angles: (a) θ_T and ϕ_T are defined in the J/ψ meson rest frame, (b) ψ_T is defined in the ϕ meson rest frame. The transversity angles are marked in red to differentiate them from the particles in the decay.

Decay rate for $B_s \to J/\psi\phi$

The time and angle dependent $P \to VV$ decay rates for B_s and $\overline{B_s}$ expressed as a function of the amplitudes $\{A_\alpha\}$ are derived in Reference [38] in a manner analogous to our derivation of Equations 1.25 and 1.26. We forego the full algebraic details here, and present the resultant decay rate for $B_s \to J/\psi\phi$:

$$\begin{aligned}\frac{d^4\,\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\;dt} \propto\; & |A_0|^2 g_1(t) f_1(\vec{\omega}) + |A_\parallel|^2 g_2(t) f_2(\vec{\omega}) + |A_\perp|^2 g_3(t) f_3(\vec{\omega}) \\ & + |A_\parallel||A_\perp| g_4(t) f_4(\vec{\omega}) + |A_0||A_\parallel| g_5(t) f_5(\vec{\omega}) \\ & + |A_0||A_\perp| g_6(t) f_6(\vec{\omega}),\end{aligned} \tag{1.40}$$

where the functions $f_i(\vec{\omega} \equiv \{\cos\theta, \phi, \cos\psi\})$ contain the dependence on the transversity angles. We have dropped the subscript T for visual clarity:

$$\begin{aligned} f_1 &= \frac{9}{32\pi} 2\cos^2\psi(1-\sin^2\theta\cos^2\phi), & f_4 &= -\frac{9}{32\pi}\sin^2\psi\sin 2\theta\sin\phi, \\ f_2 &= \frac{9}{32\pi}\sin^2\psi(1-\sin^2\theta\sin^2\phi), & f_5 &= \frac{9}{32\pi}\frac{1}{\sqrt{2}}\sin 2\psi\sin^2\theta\sin 2\phi, \\ f_3 &= \frac{9}{32\pi}\sin^2\psi\sin^2\theta, & f_6 &= \frac{9}{32\pi}\frac{1}{\sqrt{2}}\sin 2\psi\sin 2\theta\cos\phi. \end{aligned} \tag{1.41}$$

The dependence on $2\beta_s$, Δm $\Delta\Gamma$, Γ, $\delta_\parallel$ and $\delta_\perp$ is contained in the g_i functions,

defined in Equation 1.42, with $\delta_1 \equiv \delta_\perp - \delta_\parallel$, $\delta_2 \equiv \delta_\perp$, and taking the upper (lower) sign for B_s ($\overline{B_s}$).

We note that Equation 1.40, with substitutions according to 1.42, is invariant under a simultaneous change of sign of $\cos 2\beta_s$, $\Delta\Gamma$, $\cos\delta_1$, and $\cos\delta_2$. We accommodate this in our analysis by removing the absolute value signs on $\cos 2\beta_s$ and $\Delta\Gamma$ and finding the two corresponding sets of values for the parameters of interest. This degeneracy in the equation for the decay rate is removed if we have external information on the sign of the $\cos\delta_1$ and $\cos\delta_2$, or alternatively the sign of $\delta_\parallel$ and $\delta_\perp$.

$$\begin{aligned}
\frac{g_1(t)}{e^{-\Gamma t}} &= \cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma|t}{2} \mp \sin 2\beta_s \sin\Delta m t \\
\frac{g_2(t)}{e^{-\Gamma t}} &= \cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma|t}{2} \mp \sin 2\beta_s \sin\Delta m t \\
\frac{g_3(t)}{e^{-\Gamma t}} &= \cosh\frac{\Delta\Gamma t}{2} + |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma|t}{2} \pm \sin 2\beta_s \sin\Delta m t \\
\frac{g_5(t)}{e^{-\Gamma t}} &= \left(\cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma|t}{2} \mp \sin 2\beta_s \sin\Delta m t\right)\cos(\delta_2 - \delta_1) \\
\frac{g_4(t)}{e^{-\Gamma t}} &= \pm\sin\delta_1\cos\Delta m t \mp \cos\delta_1\cos 2\beta_s \sin\Delta m t + \cos\delta_1 \sin 2\beta_s \sinh\frac{\Delta\Gamma t}{2} \\
\frac{g_6(t)}{e^{-\Gamma t}} &= \pm\sin\delta_2\cos\Delta m t \mp \cos\delta_2\cos 2\beta_s \sin\Delta m t + \cos\delta_2 \sin 2\beta_s \sinh\frac{\Delta\Gamma t}{2}.
\end{aligned} \tag{1.42}$$

Rearranging terms for brevity, and expressing the g_i functions in terms of Γ_H and Γ_L, we obtain the following expression for the decay rate of a B_s ($\overline{B_s}$) meson, taking the upper (lower) sign:

$$\begin{aligned}
\frac{d^4\,\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} \propto\; & |A_0|^2(\mathcal{T}_+ \mp \mathcal{T}_{ss})\,f_1 + |A_\parallel|^2(\mathcal{T}_+ \mp \mathcal{T}_{ss})\,f_2 + |A_\perp|^2(\mathcal{T}_- \pm \mathcal{T}_{ss})\,f_3 \\
& + |A_0||A_\parallel|\cos(\delta_2-\delta_1)(\mathcal{T}_+ \mp \mathcal{T}_s)\,f_5 \\
& - |A_\parallel||A_\perp|\left(\cos\delta_1\,\mathcal{T}_s \mp \sin\delta_1\,\mathcal{T}_c \pm \cos\delta_1\,\mathcal{T}_{sc}\right)\,f_4 \\
& - |A_0||A_\perp|\left(\cos\delta_2\,\mathcal{T}_s \mp \sin\delta_2\,\mathcal{T}_c \pm \cos\delta_2\,\mathcal{T}_{sc}\right)\,f_6,
\end{aligned} \tag{1.43}$$

where, for compactness, we have defined the following:

$$\begin{aligned}
\mathcal{T}_\pm &\equiv \frac{1}{2}\left((1\pm\cos 2\beta_s)e^{-\Gamma_L t} + (1\mp\cos 2\beta_s)e^{-\Gamma_H t}\right) \\
\mathcal{T}_{ss} &\equiv \sin(\Delta m t)\,e^{-\frac{\Gamma_L+\Gamma_H}{2}t}\sin 2\beta_s \\
\mathcal{T}_{sc} &\equiv \sin(\Delta m t)\,e^{-\frac{\Gamma_L+\Gamma_H}{2}t}\cos 2\beta_s \\
\mathcal{T}_c &\equiv \cos(\Delta m t)\,e^{-\frac{\Gamma_L+\Gamma_H}{2}t} \\
\mathcal{T}_s &\equiv \sin 2\beta_s \frac{e^{-\Gamma_H t} - e^{-\Gamma_L t}}{2}.
\end{aligned} \tag{1.44}$$

The g_i functions, and therefore Equation 1.43, require knowledge of the full time evolution of the B_s meson from production to decay, including the production flavor (B_s or $\overline{B_s}$), as well as the oscillation frequency Δm_s. This requires the use of flavor-tagging algorithms explained in Chapter 5. We obtain a shorter expression – retaining a dependence on $2\beta_s$ – even if we do not know the production flavor of the B meson. Since B mesons are produced as a result of $\overline{p}p \to \overline{b}b$ interactions, and the probability for hadronization into B_s is equal for matter and antimatter, an equal number of B_s and $\overline{B_s}$ are created at the Tevatron. Summing the decay rates for upper and lower signs in 1.40 cancels out the dependence on Δm_s:

$$\begin{aligned}\frac{d^4\ \mathcal{P}_{\rm unt}(\vec{\omega},t)}{d\vec{\omega}\ dt} \propto\ & |A_0|^2\,\mathcal{T}_+\,f_1 + |A_\parallel|^2\,\mathcal{T}_+\,f_2 + |A_\perp|^2\,\mathcal{T}_-\,f_3 + |A_0||A_\parallel|\cos(\delta_\parallel)\,\mathcal{T}_+\,f_5 \\ & - |A_\parallel||A_\perp|\sin(2\beta_s)\frac{e^{-\Gamma_H t} - e^{-\Gamma_L t}}{2}\cos(\delta_\perp - \delta_\parallel)\,f_4 \\ & - |A_0||A_\perp|\sin(2\beta_s)\frac{e^{-\Gamma_H t} - e^{-\Gamma_L t}}{2}\cos(\delta_\perp)\,f_6.\end{aligned} \tag{1.45}$$

In addition to the loss of time-evolution information that results from the absence of initial state flavor tagging, a measurement of $2\beta_s$ in an untagged sample of B_s decays exhibits a higher-order ambiguity in the solutions in the $\Delta\Gamma$–$2\beta_s$ plane. Equation 1.45 is invariant under a simultaneous change of sign of $\cos 2\beta_s$ and $\Delta\Gamma$, and also under an independent change of sign of $\cos\delta_1$ and $\cos\delta_2$, resulting in a four-fold ambiguity in the $\Delta\Gamma$–$2\beta_s$ plane. The expressions for the tagged and untagged decay rates are expanded in Appendix A for ease of inspection of the transformation properties.

Decay rate for $B_d \to J/\psi K^*$

The time and angle dependent $P \to VV$ decay rates for B_d and $\overline{B_d}$ expressed as a function of the amplitudes $\{A_\alpha\}$ are derived in Reference [39]. The decay rate for B_d differs from the one for B_s in two main ways: the width difference $\Delta\Gamma_d$ does not appear because we use the approximation $\Delta\Gamma_d = 0$, and the decay rate does not depend on a *CP* violating phase because this mode does not allow for interference of mixing and decay. Moreover, instead of two decay rates for $B_s \to J/\psi\phi$, there are four decay rates in the mode $B_d \to J/\psi K^*$, one for each combination of an initial state B_d or $\overline{B_d}$ and a final state K^* or $\overline{K^*}$. We can distinguish between the latter two final states because of their different decay products: $K^* \to K^+\pi^-$, and $\overline{K^*} \to K^-\pi^+$. Since we do not make use of initial state flavor tagging in the angular analysis of $B_d \to J/\psi K^*$, we sum the B_d and $\overline{B_d}$ rates in this treatment. Here again, we forego the full algebraic details, and present the resultant decay rate for an untagged sample of $B_d \to J/\psi K^*$:

$$\begin{aligned}\frac{d^4\mathcal{P}(t,\vec{\omega})}{d^4\vec{\omega}dt} \propto e^{\Gamma_d t}\cdot\Big\{ & |A_0|^2 f_1 + |A_\parallel|^2 f_2 + |A_\perp|^2 f_3 \\ & \pm \mathcal{I}m(A_\parallel^* A_\perp) f_4 + \mathcal{R}e(A_0^* A_\parallel) f_5 \pm \mathcal{I}m(A_0^* A_\perp) f_6\Big\},\end{aligned} \tag{1.46}$$

where the upper (lower) sign is for $K^+\pi^-$ ($K^-\pi^+$).

For our purposes, the $B_d \to J/\psi K^*$ decay serves as an excellent control sample to evaluate our methods in the angular analysis, as well as a good comparison for the lifetime measurement. The decay amplitudes $\{A_\alpha\}$ have been measured independently in both decay modes. However, the amplitudes in the $B_d \to J/\psi K^*$ decay have been measured by the B factories (the BaBar and Belle experiments) and at the Tevatron experiments to higher precision than those in the $B_s \to J/\psi\phi$ decay. We expect from SU(3) flavor symmetry that the decay amplitudes for B_s and B_d are of the same order. Finally, theory predicts that the decay width of the B_s and B_d mesons differ by less than 1% [40]:

$$\left|\frac{\Gamma_s}{\Gamma_d} - 1\right| < 0.01. \tag{1.47}$$

Both of these theoretical predictions are used in our analysis for complementary results.

1.5 Current Experimental Status

This dissertation describes the first measurement of the *CP*-violating phase $2\beta_s$ in $B_s \to J/\psi\phi$ decays that combines an angular analysis to infer the *CP* state of the decay products and initial state flavor tagging to determine the production flavor (B_s or $\overline{B_s}$) of the B_s mesons. In building up toward this final measurement, we present several associated measurements that stand on their own: the average lifetime τ_s, the width difference $\Delta\Gamma_s$ and the angular amplitudes in the B_s decay, as well as the average lifetime and angular amplitudes in the B_d decay. In this section we summarize the experimental status of the measurements performed prior to our analysis.

The average lifetimes of the B_s and B_d mesons have been measured both at the B factories and at the Tevatron. The precision of the current world average is approximately 1%, and all individual measurements are primarily limited by the statistical uncertainty rather than the uncertainty arising from systematic effects. Because the measurements of average lifetime have all been limited by the statistical uncertainty, the chance to update our current knowledge using a single measurement with a large sample of B decay candidates motivates our inclusion of it in this analysis. The world average values are summarized in Table 1.2

B_q	Mass [MeV/c^2]	Lifetime [ps]
B_s	5369.6 ± 2.4	1.466 ± 0.059
B_d	5279.3 ± 0.7	1.530 ± 0.009

Table 1.2: Current world averages of experimental results for the mass and lifetimes of the B_d and B_s mesons as of 2006, prior to the publication of the results presented in this dissertation [41].

Efforts to measure the angular amplitudes in $P \to VV$ decays have been under way

for the past ten years, with success in particular for B_d decays. The slower oscillation frequency of the B_d meson permits time-dependent (as opposed to time-integrated) measurements at the B factories as well as at CDF and DØ. The former are limited in time-dependent measurements by less powerful proper time resolution. We summarize in Table 1.3 the results of these collaborations in measurements of these amplitudes in both decays studied here. The most recent measurements made at the B factories of the B_d amplitudes serve as benchmarks we seek to match in the analysis of this decay mode. We limit the overview to the time-dependent measurements since these are currently the most common. A good overview of past time-integrated measurement of these amplitudes is provided in Reference [42].

	CDF (2004)	Belle (2005)	BaBar (2007)
$\|A_0\|^2$	$0.562 \pm 0.025 \pm 0.017$	$0.574 \pm 0.012 \pm 0.009$	$0.556 \pm 0.009 \pm 0.010$
$\|A_\parallel\|^2$	$0.223 \pm 0.032 \pm 0.007$	$0.231 \pm 0.012 \pm 0.008$	$0.211 \pm 0.010 \pm 0.006$
$\delta_\parallel$	$2.86 \pm 0.22 \pm 0.08$	$2.887 \pm 0.090 \pm 0.080$	$2.930 \pm 0.080 \pm 0.040$
$\delta_\perp$	$0.15 \pm 0.15 \pm 0.04$	$2.938 \pm 0.064 \pm 0.010$	$2.910 \pm 0.050 \pm 0.030$

Table 1.3: The results of previous measurements of the angular amplitudes in and $B_d \to J/\psi K^*$ decays. The first uncertainty listed for each parameter is statistical, and the second is from systematic effects. Collected from References [42, 43, 44].

The first measurement of the width difference between heavy and light eigenstates of the B_s was done at CDF with a data sample corresponding to 258 pb^{-1} of integrated luminosity*, using an identical angular analysis as the one used in our analysis, and making the assumption $2\beta_s = 0$ [42]. As the pioneering measurement of the width different in $B_s \to J/\psi\phi$, its largest contribution was confirming that such a difference was detectable and non-zero. The analysis also included a measurement of the angular amplitudes $\{A_\alpha\}$ in the $B_s \to J/\psi\phi$ decay. The results were:

$$\begin{aligned} \Delta\Gamma_s &= 0.47^{+0.19}_{-0.24} \pm 0.01 \text{ ps}^{-1}, \\ |A_0|^2 &= 0.615 \pm 0.064 \pm 0.011, \\ |A_\parallel|^2 &= 0.260 \pm 0.086 \pm 0.013, \\ \delta_\parallel &= 1.93 \pm 0.36 \pm 0.03, \end{aligned}$$

listing the statistical uncertainty first, and the systematic uncertainty second for each parameter. This dissertation presents the first update to this measurement using the CDF dataset.

Shortly prior to the completion of analysis presented in this document, the DØ collaboration published the results of a measurement of the width difference and *CP*-violating phase in the B_s system using a time-dependent angular analysis but without flavor tagging [45]. The data sample used corresponds to 1.1 fb^{-1} of luminosity. The analysis was performed under two separate scenarios. First, the measurement was

*This corresponds to approximately 15% of the size of the data sample used in the analogous analysis presented in this document. See Section 2.1.4 for a definition of luminosity.

done with the *CP*-violating phase $2\beta_s$ fixed to zero. Second the measurement was done letting $2\beta_s$ have any value. The first scenario corresponds exactly to the measurement in Reference [42]. The measurement conducted under the second scenario was the first of its kind. The results under the assumption that $2\beta_s \equiv 0$ were:

$$\begin{aligned} \Delta\Gamma_s &= 0.12^{+0.08}_{-0.10} \pm 0.02 \text{ ps}^{-1}, \\ |A_0|^2 - |A_\parallel|^2 &= 0.38 \pm 0.05, \\ |A_\perp|^2 &= 0.45 \pm 0.05, \\ \delta_\parallel &= 2.6 \pm 0.4, \end{aligned}$$

listing the statistical uncertainty first, and the systematic uncertainty second for, as we do throughout this section. No systematic uncertainties were calculated for the other parameters. The results of the measurement letting $2\beta_s$ float freely were:

$$\begin{aligned} \Delta\Gamma_s &= 0.17 \pm 0.09 \pm 0.02 \text{ ps}^{-1}, \\ 2\beta_s &= 0.79 \pm 0.56\ ^{+0.01}_{-0.14}, \\ |A_0|^2 - |A_\parallel|^2 &= 0.37 \pm 0.06, \\ |A_\perp|^2 &= 0.46 \pm 0.06, \\ \delta_\parallel &= 2.6 \pm 0.4, \\ \delta_\perp &= 0.7 \pm 1.1. \end{aligned}$$

The methods used to obtain the results in Reference [45] are very similar to the methods described this document. These include an identically motivated lifetime and angular analysis, similar treatment of detector effects and similar sources of systematic uncertainties. One important difference is the statistical methods used to extract the values of the parameters of interest. While the measurement in Reference [45] makes use of a straightforward maximum likelihood fit (see Section 4.1), we use the likelihood ratio method described in Chapter 6. This change is motivated by our conclusion that the maximum likelihood method is insufficiently robust for the measurement of $2\beta_s$ in $B_s \to J/\psi\phi$ decays. A more detailed discussion of these results from DØ, and a comparison to the results presented in this dissertation are included in Chapter 7.

Updated Results in $B_s \to J/\psi\phi$

The most recent measurements of the *CP*-violating phase $2\beta_s$ in $B_s \to J/\psi\phi$ decays were completed after the analysis described here, and during the writing of this document. We include them in our experimental status overview for completeness. The DØ collaboration released in early 2008 the results of a time-dependent angular analysis of flavor-tagged $B_s \to J/\psi\phi$ decays using a data sample corresponding to 2.8 fb^{-1} of integrated luminosity [46]. Besides the addition of flavor-tagging and the use of a larger sample of data, this measurement uses the same strategy as the one described in Reference [45], including the continued use of a straightforward maximum likelihood fit. The value of $\Delta\Gamma$ obtained by fixing the value of $2\beta_s$ to the Standard Model value of 0.04 is $\Delta\Gamma_s = 0.14 \pm 0.07\ ^{+0.02}_{-0.01}$ ps^{-1}. The results obtained by letting

$2\beta_s$ be determined by the fit are:

$$\Delta\Gamma_s = 0.19 \pm 0.07\ ^{+0.02}_{-0.01}\ \text{ps}^{-1},$$
$$2\beta_s = 0.57^{+0.24\ +0.02}_{-0.30\ -0.07}.$$

In August 2008, the CDF collaboration released an updated measurement of $2\beta_s$ using a sample of flavor-tagged $B_s \to J/\psi\phi$ decays corresponding to an integrated luminosity of 2.8 fb^{-1} [47]. This measurement uses the same methods as the one we describe in this document, including the likelihood ratio method to extract a confidence region in the $\Delta\Gamma$-$2\beta_s$ plane. The updated measurement features a new and improved flavor-tagging algorithm that increases the ability to determine the flavor of B_s meson at production time. The value of $\Delta\Gamma$ obtained by fixing the value of $2\beta_s$ to zero is $\Delta\Gamma_s = 0.02 \pm 0.05 \pm 0.01$ ps^{-1}. Allowing $2\beta_s$ to float freely, and assuming the Standard Model predictions of $2\beta_s$ and $\Delta\Gamma$, the probability of a deviation as large as the level of the observed data is 7%, corresponding to 1.8 Gaussian standard deviations.

As this section has made clear, there is currently tremendous experimental activity in the measurement of *CP* violation in the B_s system, coupled with great interest in the theoretical community. This interest is expected to remain high during the operating lifetimes of the Tevatron experiments and similar measurement are expected to feature prominently in the B physics program at the LHC accelerator.

1.6 Analysis Overview

Our ultimate goal is to present the first measurement of the *CP* violating phase $2\beta_s$ that makes use of initial-state flavor tagging in the $B_s \to J/\psi\phi$ mode. In doing so, we combine elements from several individual analyses, each of which presents its own set of challenges. We need to approach the task in a step-by-step fashion in order to build up a sound analysis framework, yielding a robust result. Along the way, we obtain and present several measurements which stand in their own right as important results.

The first task is to perform a time-dependent angular analysis in the $B_d \to J/\psi K^*$ mode. This accomplishes several important goals: it validates our lifetime fitting framework as well as our angular analysis tools, in particular our treatment of the angular acceptance of our detector and selection criteria. In addition, this analysis yields measurements of the average lifetime of the B_d meson, and of the angular amplitudes associated with the $P \to VV$ decay in this channel. Given the current size of our data sample, the measurements we obtain in this decay channel are not merely used to cross-check our methods and validate our framework. We aim to challenge the precision achieved at the B factories.

The second task is to perform a time-dependent angular analysis in the $B_s \to J/\psi\phi$ mode, measuring in the process the width difference $\Delta\Gamma_s$ as well as the average lifetime and angular amplitudes in this decay channel. At this stage, we assume no *CP* violation by setting $2\beta_s$ to zero. By doing so, we verify that the methods we have

used are applicable to this decay mode, and we gain the opportunity to compare these values to previously measured results while insulating ourselves from any systematic effects associated with the *CP* violation measurement. This provides a timely update to the 2004 measurements of these quantities at CDF.

Finally, we tackle the measurement of the *CP*-violating phase $2\beta_s$ in the $B_s \to J/\psi\phi$ decay. The investigation of $2\beta_s$ is a nascent field, and only one publication exists in the literature prior to our analysis, the recent publication by DØ documented in Reference [45]. We advance the techniques used to carry out this measurement in two ways. First, we detail an alternate method of simultaneously measuring $\Delta\Gamma_s$ and $2\beta_s$ using a variation on interval estimation. This method is used to perform the measurement first without, and then with knowledge of the initial-state flavor of the B_s meson. Second, we incorporate the use of flavor tagging to enhance our sensitivity to the parameter $2\beta_s$. We use the algorithms developed for the measurement of Δm_s at CDF and include them in our analysis framework. Following the strategy we outline here, we then perform the first measurement of $2\beta_s$ in a sample of flavor-tagged $B_s \to J/\psi\phi$ decays.

Chapter 2

Experimental Apparatus

The data used to perform the analysis described in this document were obtained at the Collider Detector Facility experiment with the eponymous CDF Run II detector (CDF-II), which records the products of proton-antiproton collisions in the Tevatron accelerator at Fermi National Accelerator Laboratory (FNAL or Fermilab). The B decays used for this measurement are among the outgoing collision products of the $p\overline{p}$ interactions effected at the Tevatron. The CDF-II detector, located at one of the two collision points of the Tevatron, measures the energy, momentum and position in space of the comparatively long-lived outgoing products of the collisions. This chapter first describes the basics of production, acceleration, and collision of protons and antiprotons at the Fermilab accelerator complex, which culminates in the Tevatron collider. In the next section, we describe the CDF-II detector, with an emphasis on the subsystems that are most relevant to this dissertation. Finally, we describe the data acquisition and real-time event selection systems at CDF.

2.1 Accelerator

The Tevatron is a synchrotron of radius 1 km. It accelerates and collides counter-rotating beams of protons and antiprotons, each with an energy of 980 GeV per hadron. It is the last and most energetic stage in a chain of individual particle accelerators at Fermilab, and the only one capable of operating in colliding beams mode. The accelerator complex begins by ionizing hydrogen gas to produce a source of H^- ions, and ends by delivering proton-antiproton collisions at a center of mass energy $\sqrt{s} = 1.96$ TeV at two positions along the beam. The CDF and DØ detectors are placed around the two interaction points to detect the results of those collisions. In addition to serving the two collider experiments, the accelerator complex also delivers proton beams of lower-than-maximum energy to the test beam facilities, fixed-target and neutrino experiments located at Fermilab. Table 2.1 shows the energy of a proton or antiproton in the beam after each of the five stages of acceleration. Figure 2.1 shows an overhead diagram of the chain of accelerators at Fermilab.

The Tevatron was the first accelerator to use cryogenically cooled superconducting magnets along its entire circumference. Built in the 1980s, it replaced the Main

Accelerator	Highest Energy
Cockroft-Walton	750 KeV
Linac	400 MeV
Booster	8 GeV
Main Injector	150 GeV
Tevatron	980 GeV

Table 2.1: The five stages of acceleration at the Fermilab accelerator complex in Run II of the Tevatron. The energies listed refer to an individual proton in the beam after each stage, and to an individual antiproton starting at the Main Injector.

Ring, a proton-only synchrotron that used conventional magnets to deliver a 500 GeV beam to fixed-target experiments. During the operating lifetime of the Tevatron, two major accelerator and detector upgrades took place to increase data-taking rates by

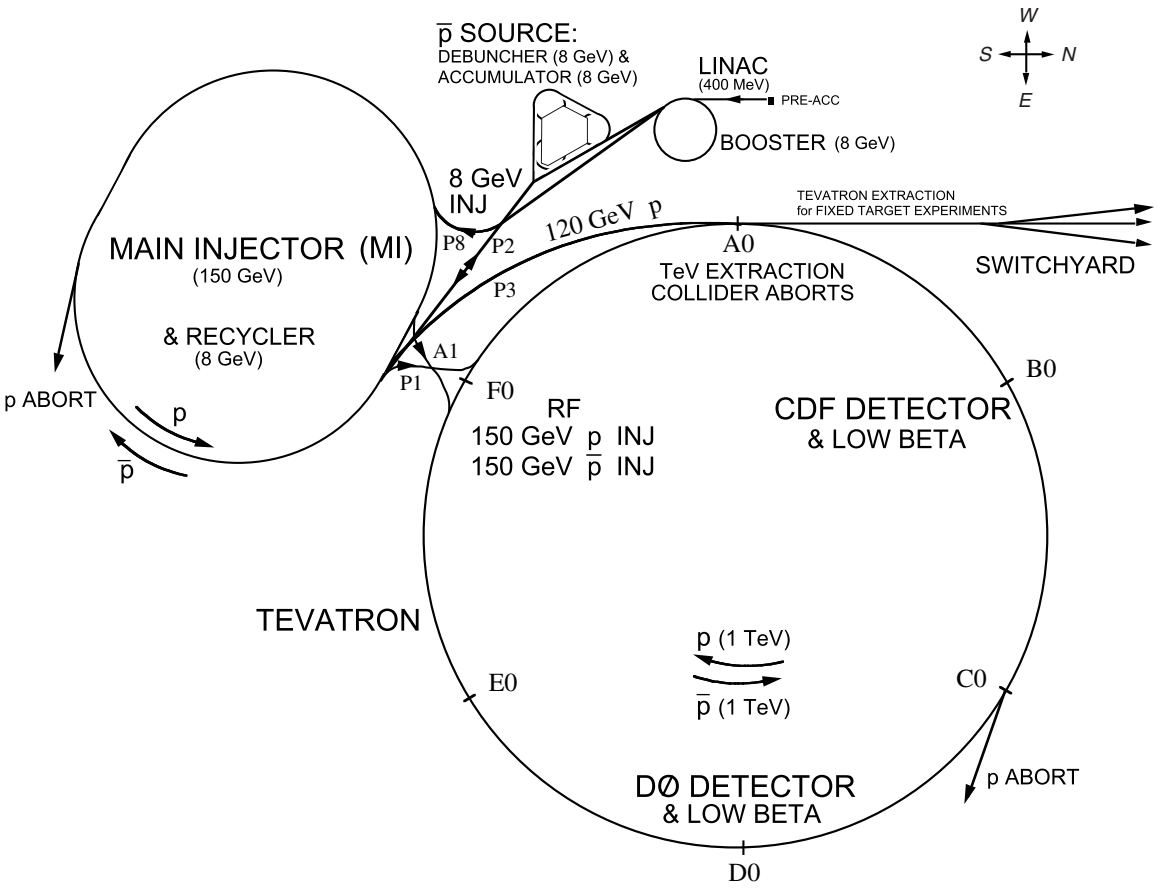


Figure 2-1: Overhead diagram showing the accelerator complex at Fermilab. The switchyard leads to the fixed target experiments and test beam areas. The individual accelerators are discussed in the text in Section 2.1, and the CDF detector is discussed in detail in Section 2.2.

increasing the luminosity of the collisions. During Run I, which ran from 1992 to 1996, the Tevatron accelerated each beam to 900 GeV. In Run II, the energy of each beam was increased to 980 GeV.

2.1.1 Initial Acceleration

Proton production begins with a source of H^-. Electrical pulses in hydrogen gas produce negative ions, which are then sent to the first accelerator. This first stage is an electrostatic Cockroft-Walton device, which accelerates the H^- ions in a linear path up to an energy of 750 KeV [48]. Because electrostatic acceleration works by creating a potential difference between a source of charged particles and a target point, the upper energy limit is set by electrostatic breakdown in air, and is approximately 10 MeV.

Following electrostatic acceleration, H^- ions enter the Linac [48], a two-stage, 130-meter linear accelerator, which uses radio-frequency (RF) accelerating cavities. RF acceleration eschews the problem of electrostatic breakdown by repeatedly passing charged particles through an acceleration gap, each time imparting additional energy, rather than attempting to reach the target energy in one pass. In the case of a linear accelerator, this is done with a series of colinear RF cavities, or drift tubes. In each cavity, the oscillating electric field component of the RF field is made to point in the accelerating direction at each passage of the beam. In the construction used at Fermilab, the drift tubes themselves are designed to fulfill this condition: the ith tube has length $l_i = \beta\lambda_{rf}$, where β is the desired particle speed and λ_{rf} is the wavelength associated with the RF field. In order to keep the particles in phase with the RF field, the beam is not continuous but divided in discrete bunches. The number of bunches and their separation varies as the beam makes its way through the various accelerators.

The first stage of the Linac is what remains of the original 1971 200 MeV machine [49]. It now consists of five drift tubes, and accelerates the ions to 116 MeV. The second stage consists of seven RF cavities operating at 805 MHz, and accelerates the ions to 400 MeV [50]. It is only after acceleration in the Linac that the ions are passed through a carbon fiber foil, which strips the electrons and leaves a beam of protons.

From the Linac, the proton beam is sent to the Booster [51], a synchrotron of radius 75 m, where they are accelerated up to an energy of 8 GeV, gaining 500 KeV per revolution. Acceleration is achieved by increasing the dipole field strength and simultaneously bringing the RF frequency from 38 to 53 MHz in an interval of 33 ms. In order to increase the intensity of the beam, several Linac cycles are injected into the Booster before passing to the next acceleration stage. One Booster batch consists of 84 bunches of protons, each consisting of approximately 6×10^{10} protons.

2.1.2 Main Injector

The Main Injector (MI) [52], a synchrotron of radius 525 m, plays three main roles:

- accelerate protons and antiprotons to 150 GeV for the Tevatron,

- accelerate protons to 120 GeV for antiproton production, fixed target experiments and neutrino production,
- inject protons and antiprotons into the Tevatron.

In order to serve these functions, the MI receives 8 GeV protons from the Booster and 8 GeV antiprotons from the Antiproton Source. It accelerates or decelerates particles between 8 and 150 GeV, and deliver beams to the Antiproton Source, the Recycler, the Switchyard, and the Tevatron.

In the MI, the proton beam is coalesced from 6 Booster batches of 84 bunches into 36 bunches in preparation for Tevatron injection.

2.1.3 Antiproton Production and Storage

The supply of antiprotons is the major bottleneck of a $p\overline{p}$ collider. Since they are not found in any abundance from natural sources, antiprotons are made at Fermilab by colliding protons with a fixed target, and capturing the antiprotons from the particle shower that ensues. This is still a fundamentally inefficient process, yielding 1-2 antiprotons for every 10^5 protons incident on the target. Therefore, it is necessary to effect these collisions repeatedly, storing the antiprotons that are captured each time until a sufficiently large supply is ready for $p\overline{p}$ collisions at the desired luminosity.

Antiproton Source

The 120 GeV protons from the Main Injector to be used for antiproton production are sent through a transport line to the target station of the Antiproton Source (AS) [53]. There, the protons are focused using a set of quadrupole magnets to a smaller beam cross section, then strike a nickel target, producing a shower of secondary particles. A Lithium lens renders all charged particle trajectories parallel, following which a dipole magnet sends all the negatively charged particles with energy near 8 GeV to the Debuncher, where cooling and storage begin.

Cooling and Storage

All else being equal, the smaller the phase space volume that the colliding beams occupy, the higher the luminosity. By reducing the spread of the particles in transverse and longitudinal momentum, we increase the probability of obtaining a collision. Because this condition thermodynamically corresponds to lower temperature, the process of preparing beams in this manner is called *cooling*.

In the case of protons, their abundance allows for a very simple method of reducing the phase space volume of the particles in the beam. We simply discard protons that lie outside the desired space. On the other hand, antiprotons are too rare to permit such cavalier exclusion of a portion of the available beam.

Two primary types of cooling are used to reduce the phase space occupied by antiproton bunches: stochastic and electron cooling. Stochastic cooling is a feedback-based mechanism developed by Simon van der Meer and first used at the CERN SPS

collider [54, 55]. This method of cooling has been used at the Fermilab since prior to Run I of the Tevatron [56]. The transverse or longitudinal motion of particles in the beam is sampled by pickup sensors at one point, and corrections are applied using kicker electrodes and magnets at a further point along the beam's trajectory. Its effectiveness is inversely proportional to the linear particle density in the beam, and is therefore less useful when the antiproton stack becomes larger. This is because higher linear density induces long-range interactions that interfere with the feedback-based correction process. On the other hand, the effectiveness of electron cooling is independent of the beam linear density. This cooling method works by sending a cold beam of electrons through a warm beam of antiprotons, which transfers additional unwanted motion from the antiprotons to the electrons, which are then discarded. In this discussion, the words "cold" and "warm" refer to narrower and wider spread in phase space, respectively.

Once antiprotons are transferred to the Debuncher, two types of stochastic cooling are applied: betatron cooling, which reduces the transverse size of the beam, and momentum cooling, which reduces the momentum spread of the antiprotons.

The antiprotons are then sent to the Accumulator, where they undergo further cooling as well as RF deceleration to ease storage. The Accumulator keeps the antiproton stack until it reaches $\mathcal{O}(10^{12})$ particles before they are sent to the Tevatron.

Recycler

The Recycler ring was built beginning in 1997 to serve two purposes [57]. The first, which gives the ring its name, is to retrieve and reuse the antiprotons that remain in the Tevatron at the end of a collision period, when the collision rate drops to a threshold below which it is necessary to inject a new set of proton and antiproton bunches. The second role of the Recycler is to store antiprotons from the Accumulator to improve cooling performance using electron cooling [58], which is limited in the Accumulator by the size of the antiproton stack. Since its commissioning, the Recycler has performed the second role, and at the time of writing, is not used to recycle Tevatron antiprotons. The Recycler employs both momentum cooling and electron cooling to provide a higher intensity beam with reduced tranverse and longitudinal spread.

2.1.4 Acceleration and Collisions in the Tevatron

The Tevatron is the final step in particle acceleration at Fermilab [59]. It has been for decades the highest energy particle accelerator in the world, with a center of mass energy ($\sqrt{s}$) of 1.96 TeV. In 2009, it is set to be overtaken by the Large Hadron Collider at the CERN laboratory in Switzerland, which is designed to operate in colliding proton-proton beams mode at a $\sqrt{s}$ of 14 TeV [60].

Aside from its radius of 1 km and its high energy, two other characteristics distinguish the Tevatron from the other synchrotrons at Fermilab: it uses superconducting magnets, and is capable of operating in colliding beams mode.

Superconducting coils have the advantage over standard resistive electromagnets that they can sustain a much higher current density, allowing a much larger magnetic field. This large magnetic field, and the strong bending of charged particles that comes with it, in turn make it practical to accelerate particles to very high energy without the need to build nearly impossibly large accelerators. All Tevatron magnet coils are made of niobium-titanium (NbTi) and are cryogenically cooled to temperature of 4.6 K, necessary to achieve superconductivity for this material at the desired current density. The cooling is achieved by immersing the coils in a stream of liquid helium, insulated with liquid nitrogen. Magnets serve three functions in the Tevatron: dipoles bend the beam into a circular trajectory, quadrupoles focus it for collisions, and corrective elements adjust for imperfections in the beam trajectory.

Protons are injected at 150 GeV into the Tevatron one bunch at a time, followed by antiprotons 4 bunches at a time. The particles are accelerated to 980 GeV, and collimated to prepare for collisions. Quadrupole magnets focus the beam at the two interaction points from a transverse size of approximately 1 mm to 25 μm to improve luminosity. In the Tevatron the 36 bunches correspond to a bunch spacing of 396 ns.

The highest energy achieved in a particle accelerator governs the kind of physics that are studied at the associated experiments, while the amount of data it generates increases the chances of observing rare phenomena and decreases the statistical uncertainty of measured physical parameters. Given that the highest beam energy achieved by an accelerator cannot be changed without large, expensive and thus rare upgrades, the figure of merit for the performance of a particle collider is the instantaneous luminosity, usually given in units of $\mathrm{cm}^{-1}\,\mathrm{s}^{-1}$, which is a measure of the instantaneous rate of collisions. Maximizing luminosity maximizes the interaction rate $R = \sigma L$, where σ is the interaction cross section of a process. This in turn maximized the chance of discovery of new phenomena, and increases the dataset of known phenomena under study. Efforts to increase instantaneous luminosity have been a constant and continuous undertaking at Fermilab, and the progress achieved since the beginning of Run II is a testament to the ingenuity and indefatigability of the Beams Division. Instantaneous luminosity at the Tevatron is defined as

$$L = \frac{f_{\mathrm{rf}}\ n\ N_p N_{\overline{p}}}{2\pi\left(\sigma_p^2 + \sigma_{\overline{p}}^2\right)} F\left(\frac{\sigma_l}{\beta^*}\right), \tag{2.1}$$

where f_{rf} is the (RF) bunch revolution frequency, n is the number of bunches, N_p and $N_{\overline{p}}$ are the number of protons and antiprotons per bunch, and σ_p, $\sigma_{\overline{p}}$ are the root-mean-squared widths of the beams at the interaction points. The form factor F corrects for deviations from the ideal bunch shape and depends on the bunch length σ_l and the beta function β^* at the interaction point. The instantaneous luminosity is highest at the beginning of each period of (24 – 36 hour) data-taking period, and then drops off approximately exponentially as the numbers of protons and antiprotons available for collision decrease when they collide, and as long-range interactions begin to make beam conditions less well suited for collisions. The integrated luminosity defined as $\mathcal{L} = \int L\,dt$ is a measure of the total amount of data collected, and is usually given in units of fb^{-1}, inverse femtobarns, where a barn is

equal to 10^{-24} cm. Efforts to increase the collected luminosity have been made each week, both by the Beams Division, by trying to keep beam time high, and by the CDF collaboration, by keeping the detector uptime and data-taking efficiency as high as possible. The record instantaneous and integrated luminosities over the length of Run II of the Tevatron are shown in Figure 2-2.

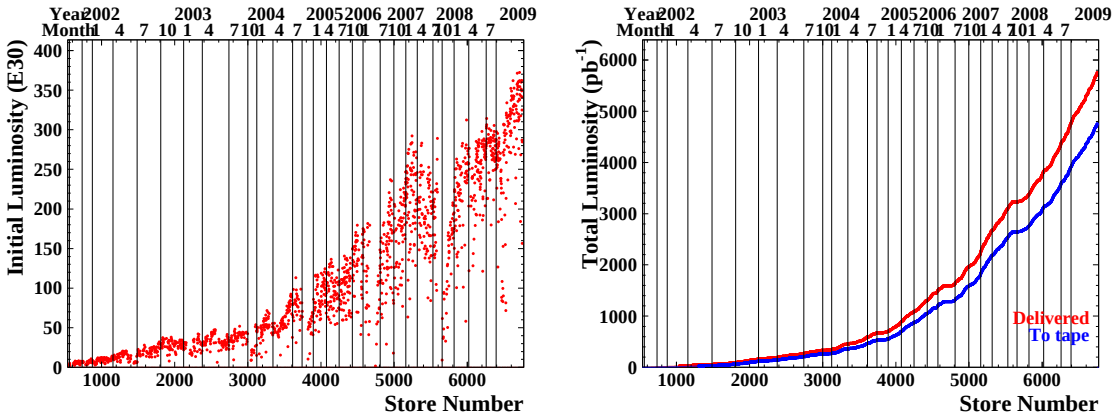


Figure 2-2: Instantaneous and integrated luminosity in Run II of the Tevatron. The current world record for instantaneous luminosity at a hadron collider was achieved at the Tevatron on December 26th, 2008, and is equal to 3.61×10^{32} cm^{-2} s^{-1}. As of this writing, the total integrated luminosity delivered by the Tevatron is 5.80 fb^{-1}, and the total luminosity acquired and written to tape by the CDF experiment is 4.79 fb^{-1}.

2.2 The CDF-II Detector

The CDF Run II detector [61] is an upgraded version of the CDF Run I and Run 0 detector, which was used between 1987 and 1996. In this discussion, unless specified, "CDF detector" refers to the current apparatus. The CDF detector is an azimuthally and forward-backward symmetric machine designed to measure the energy, momentum, and position of particles produced in $p\bar{p}$ collisions at the B0 interaction point, around which it is installed. It is of a general-purpose design, in order to permit the study of all phenomena of interest arising from the 1.96 TeV collisions at the Tevatron.

CDF is a 5000-ton assembly of detector subsystems of overall cylindrical shape, from innermost to outermost: thin beryllium beam pipe, precision charged-particle tracker and time-of-flight detector, superconducting solenoidal magnet, projective calorimeters, and finally muon detectors. The detector is shown in an elevation view in Figure 2-3. The Run II upgrade to CDF was designed to handle the increased luminosity and center-of-mass energy of the upgraded Tevatron, and included: new silicon tracking system and drift chamber, introduction of the time-of-flight detector, installation of preshower detectors in the calorimeters, better muon coverage, and an

upgrade of the triggering and data acquisition systems. The total number of channels read out in the CDF detector is $\mathcal{O}(10^6)$, with the majority ($\sim 750\,000$) coming from the silicon tracking system.

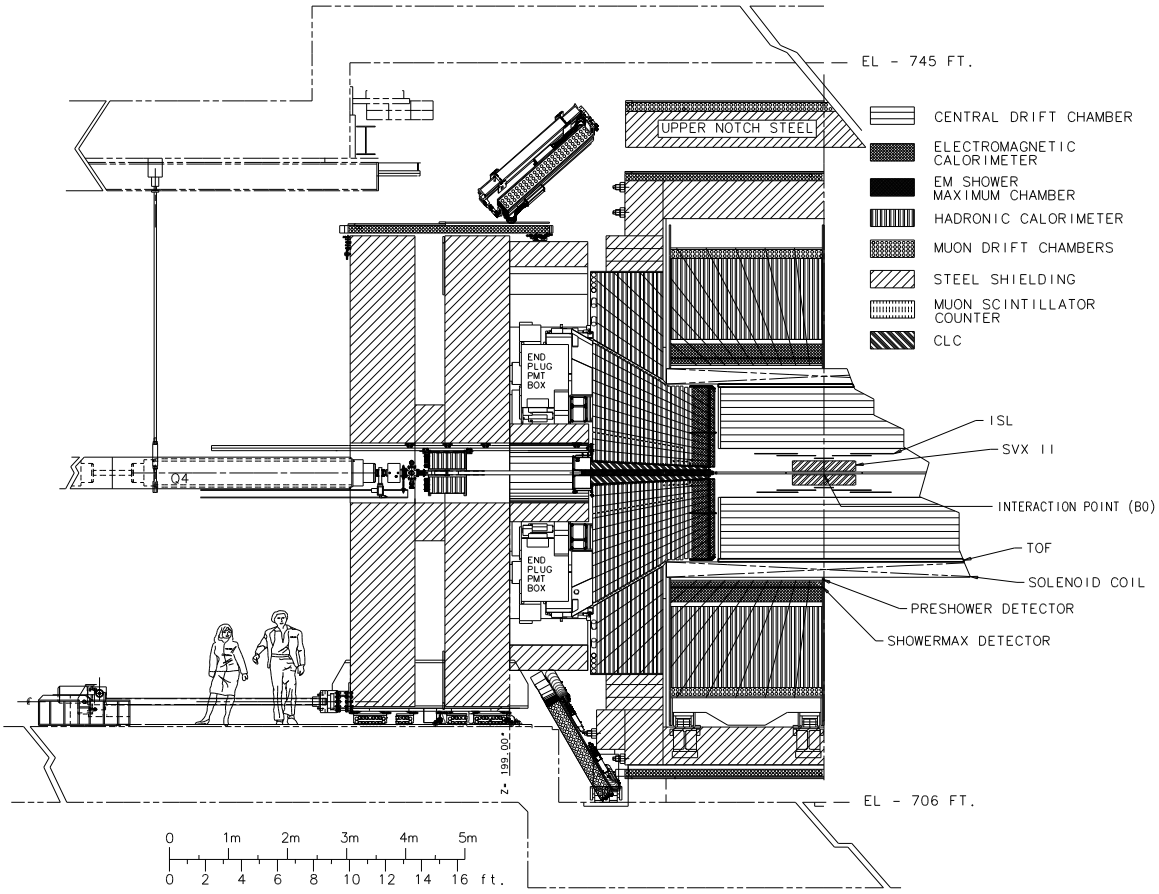


Figure 2-3: Elevation view of one half of the CDF detector. The other half is symmetric.

This section provides an overview of the various CDF subdetectors, with special emphasis on the tracking system and time-of-flight detector. Prior to any discussion of detector details, it is useful to define the standard variables in use at CDF to describe the detector geometry and the recorded collisions, in order to better characterize detector performance and design parameters.

2.2.1 Standard Definitions and Conventions

The origin in the CDF coordinate system is defined at the B0 interaction point, which is assumed to lie at the center of the drift chamber. The positive z direction points along the proton beam direction, and forms a tangent with it at the nominal interaction point. In rectangular coordinates, the y axis points vertically upward, and the x axis is defined by the right-hand rule, radially outward with respect to the Tevatron ring. In this dissertation, by convention *longitudinal* means parallel

to the z axis and *transverse* means perpendicular to the z axis, i.e. in the $x-y$ plane. Cylindrical coordinates (r, ϕ, z) and polar coordinates (ρ, ϕ, θ) are useful because of the detector's approximate axial symmetry and because the p and $\overline{p}$ beams are unpolarized, which makes interactions invariant under azimuthal rotation around the beam. The ρ coordinate is the distance from the origin, and r is the projection of ρ in the transverse plane. The angle ϕ is the azimuth, while θ is defined relative to the z axis.

The interesting interactions at a hadron collider occur between quarks and gluons which carry an unknown fraction of the parent hadron's momentum. In any such collision, the center of mass system is therefore boosted in the lab frame along z by an unknown amount. By comparison, the momentum of the center of mass system in the transverse plane is negligible. Therefore it is common practice in CDF analyses to use the projection of several useful quantities in the transverse plane. For the same reason, it is desirable to use a variable that is invariant under boosts along the z direction instead of θ. The rapidity Y fulfills this criterion and is defined as

$$Y \equiv \frac{1}{2} \ln \left[\frac{E + |\vec{p}\,| \, \cos\theta}{E - |\vec{p}\,| \, \cos\theta} \right], \tag{2.2}$$

where E, $\vec{p}$ are the particle's energy and momentum, respectively. Nevertheless, because Y depends on energy as well as momentum, it is impractical to use because it requires knowledge of each particle's mass, for which precise particle identification is indispensable but not always available. The pseudorapidity η is used instead, and is defined as

$$\eta \equiv -\ln \tan(\theta/2) \tag{2.3}$$

In the relativistic limit $pc \gg mc^2$, rapidity is well approximated by pseudorapidity: $Y \to \eta + \mathcal{O}(m^2/p^2)$. As desired, the pseudorapidity is only a function of the momentum because the polar angle θ is obtained from the ratio of the longitudinal and transverse momenta: $\cot\theta = \frac{p_z}{p_T}$. The region of the detector where $|\eta| \sim 0$ is called the central region; the regions of high $|\eta|$ are called the forward regions. The terms "central" and "forward," are context-dependent and refer only to the general position along the polar angle. Whenever a more detailed specification of location is useful, we give specific η ranges.

A charged particle with non-zero initial velocity passing through the region of constant magnetic field such as the one inside the CDF tracking volume follows a helical trajectory. Such a trajectory is described at CDF using five parameters:

- C: helix half-curvature, signed to match the particle's charge, defined as $C \equiv \frac{q}{2R}$, where R is the radius of the helix. This is related to transverse momentum by $p_T = \frac{cB}{2|C|}$ where lowercase c is the speed of light and B is the magnetic field strength.

- λ: helix pitch, defined by $\lambda \equiv \cot\theta$, where θ is the polar angle at the point of closest approach to the z axis. This is related to longitudinal momentum by $p_z = p_T \cot\theta$.

- ϕ_0: ϕ direction of trajectory at the point of closest approach to the z axis.
- d_0: signed impact parameter, i.e. the distance of closest approach to the z axis.
- z_0: the z coordinate at the point of closest approach to the z axis.

A particle's reconstructed trajectory is called a track, though no actual continuous tracks are made in the CDF detector. The full trajectory of particles is deduced from the set of discrete points at which charged particles interact with the detector, either in the layers of the silicon tracker or the cells of the drift chamber. A track at CDF is therefore obtained from a helical fit of the set of spatial measurements, or hits, reconstructed from charge deposition at various layers in the tracking subdetectors. For any two tracks, the opening angle ΔR between them is defined using the difference in azimuth and pseudorapidity:

$$\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.4}$$

Even relatively long-lived mesons such as the B_s do not "live" long enough to interact directly with the detector material. We deduce their presence as an outgoing product of $p\overline{p}$ collisions by reconstructing the tracks made by their decay products. The intersection of two or more reconstructed tracks near their origin (at r close to the beam) indicates that the particles that made those tracks came from the decay of an unstable particle. We refer to that intersection as a decay vertex.

The interaction point for the $\overline{p}p$ collision is called the primary vertex, and a large number of tracks, $\mathcal{O}(50)$ emanate from it. Any vertex that does not come from the interaction region is called a secondary vertex. For these vertices, L_{xy}, the transverse decay length in the lab frame is defined as

$$L_{xy} \equiv \vec{r} \cdot \frac{\vec{p_T}}{|\vec{p_T}|}. \tag{2.5}$$

We calculate the length of time that elapsed before an unstable particle that originated at the primary vertex decayed at a secondary vertex. The proper time elapsed in the particle's reference frame between creation and decay is defined as

$$ct = \frac{L_{xy} \cdot M}{p_T}, \tag{2.6}$$

where M is the mass of the particle, which is often assumed to be the world average value.

2.2.2 Silicon Tracking Detectors

The CDF silicon tracker [62] is comprised of three subdetectors, from innermost to outermost: Layer 00 (L00) [63], the Silicon Vertex Detector II (SVX-II) [64], and the Intermediate Silicon Layers (ISL) [65]. Layer 00 is installed directly on the beryllium beampipe, at an average radius r of 1.5 cm. The SVX-II subdetector has

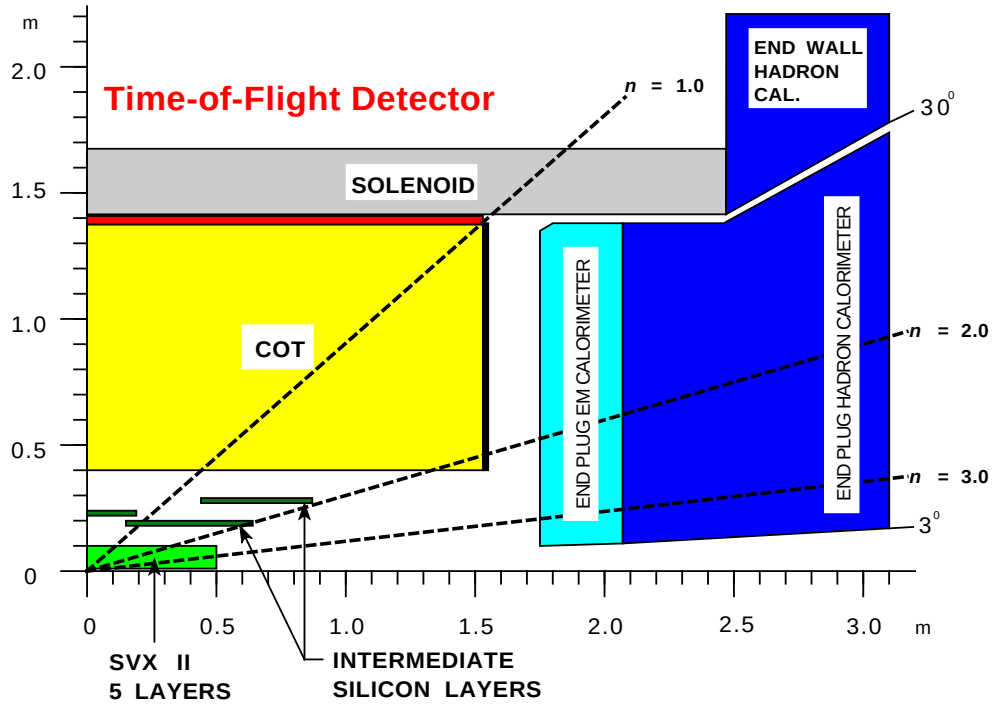


Figure 2-4: Side view of one quarter of the CDF tracking system and surrounding detectors. The remaining portions of the tracking system are essentially symmetric.

five layers, installed at equal radial distances from one another between $r = 2.5$ cm and $r = 10.6$ cm. The ISL has three layers: one central layer installed at $r = 22$ cm, and two forward layers at $r = 20$ cm and $r = 28$ cm.

This arrangement of silicon layers provides measurements of charged particle's trajectory at multiple r values as it passes through the inner tracker, with the exact number of measurements dependent on the centrality of the track. At values of pseudorapidity $|\eta| < 1$, the silicon tracking system provides measurements at seven values of r, six of which are (r, ϕ, z) measurements, and one of which is an (r, ϕ) measurement. For values of pseudorapidity $1 < |\eta| < 2$, the silicon tracking system provides measurements at eight values of r, seven of which are (r, ϕ, z) measurements, and one of which is an (r, ϕ) measurement.

Position measurements are made in the silicon detectors by gathering charge deposition resulting from the passage of charged particles in finely segmented silicon microstrips. A silicon detector is fundamentally a reverse-biased p-n junction. The passage of charged particles through a semiconductor knocks electrons from the valence to the conduction band, which creates an equal number of holes in the valence band. In a p-n junction the passage of charged particles creates electron-hole pairs in the depletion region. The charge is collected as the electrons then drift toward the positively biased anode and the holes drift toward the negatively charged cathode. The reverse bias extends the depletion region over the full thickness of the sensor and aids charge collection by reducing the flow of charge from background thermal

excitation and by increasing the strength of the field that sweeps out the signal ionization charge. When the p (and/or n) side of the junction is arranged into strips, and the charge deposition is read out for each strip separately, we obtain position information. Charge deposition in one microstrip is called a hit. Several hits in adjacent strips are called a hit cluster. The position is calculated as an average of the hit positions weighted by the amount of charge deposited in each strip.

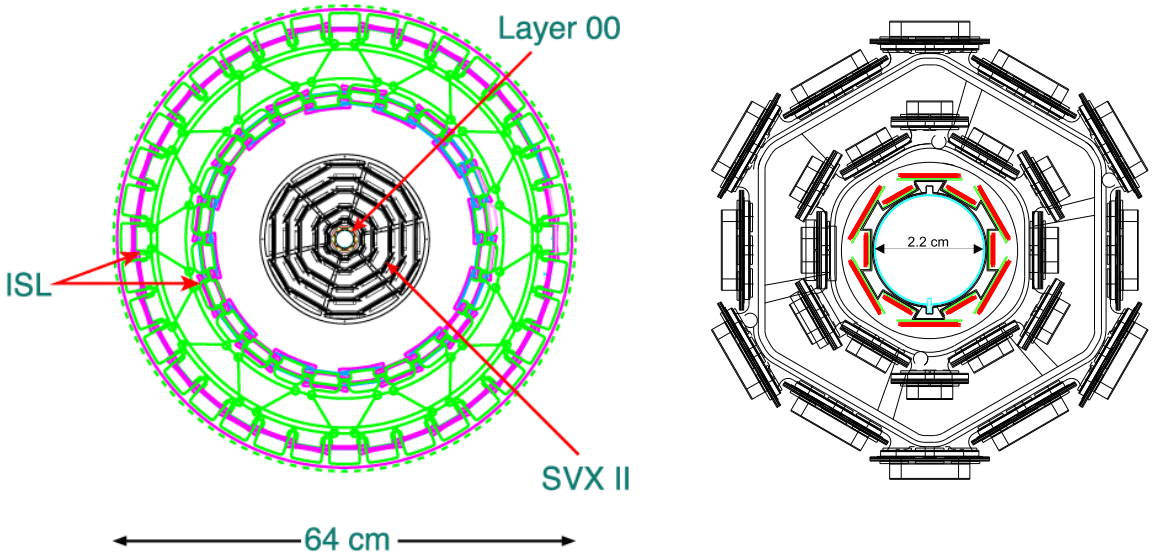


Figure 2-5: Left: End view of the CDF tracking detectors. Right: End view of Layer 00 and the first two layers of SVX-II. The two are not drawn to the same scale.

In a single-sided sensor, only the p side of the junction is segmented into strips. A double-sided sensor has strips on the p and n sides. The advantage of a double-sided detector is that one side can be rotated at an angle, called a stereo angle, in order to provide a full three dimensional position measurement. The advantage of a single-sided design, used for L00, is that it fits into a smaller space requiring fewer readout electronics. At CDF, the p side of the junction for all silicon sensors are arranged along the z direction, providing $r-\phi$ measurements.

The silicon detector makes use of a "deadtimeless" data acquisition (DAQ) system, which integrates a new analog signal while the previously collected charge is being digitized and processed. The silicon DAQ system can keep up with bunch crossings as low as 132 ns.

Layer 00

The Layer 00 sensors are placed directly on the beryllium beampipe at alternating radii or 1.35 and 1.62 cm, as shown in Figure 2-5, with single-sided strips of AC-coupled p-in-n silicon providing measurements in $r-\phi$. The distance between identical points on adjacent strips, or *pitch*, is 25 μm. The hit spatial resolution for L00 is 6 μm. L00 spans 80 cm in z, enough to cover the approximately Gaussian distribution

of primary vertex z values, centered at $z = 0$ with a width $\sigma_z \sim 30$ cm. The silicon sensors of L00 are radiation-hardened to survive the data-taking lifespan of CDF, and are expected to remain operational beyond 5 fb^{-1} of integrated luminosity. Layer 00 owe their longevity in part to the fact that they are capable of withstanding a higher bias voltage (~ 500 V), which helps to fully deplete the sensor for readout even after radiation damage [66, 67]. The total number of channels read out in Layer 00 is $\sim 14\,000$.

Silicon VerteX-II Detector

All five SVX-II layers are double-sided, with 300 μm-thick sensors. Three of the layers have a stereo angle of 90°. The other two layers have stereo angles of $-1.2°$ and $1.2°$ respectively. The pitch of the axial and small-angle stereo strips is 60 – 65 μm. The pitch of the large-angle stereo strips is 125 to 141 μm. The SVX-II provides full coverage for tracks with $|\eta| < 2$

The silicon sensors are positioned onto *ladders*, carbon fiber support assemblies. Each ladder supports four silicon sensors arranged along the z direction as well as the readout electronics, placed at the end of the ladder. Each SVX-II layer is made of 12 ladders, organized into wedges in ϕ, with the ladders slightly overlapping to guarantee coverage. The ladders are further arranged in three, 29 cm long *barrels*, mounted end-to-end along the z axis. Each barrel contains a total of 60 ladders, mounted between two beryllium bulkheads which provide mechanical support and which also carry the water cooling lines for the readout electronics. The total number of channels in SVX-II is $\sim 400\,000$.

Intermediate Silicon Layers

The ISL covers different ranges in z depending on the layer. The central layer at $r = 22$ cm has a z span corresponding to coverage for $|\eta| < 1$. This layer is useful for extrapolating tracks from the COT into the SVX. The forward layers at $r = 20$ and 29 cm provide measurements for $1 < |\eta| < 2$, where the COT coverage is incomplete. The ISL layers are double-sided with a stereo angle of 1.2°. The sensor pitch is 55 μm for the axial side and 73 μm for the stereo side. The ISL ladders overlap slightly in z to ensure proper coverage, which is made easier because of the increased available space at larger radius. The total number of channels from the ISL is $\sim 300\,000$. The full z span, taking into account all three layers is 190 cm.

Figure 2-6 shows an $r - z$ view of the η coverage provided by the three silicon tracking detectors. Table 2.2 summarizes some of the design parameters of the three silicon subsystems [61, 62].

2.2.3 Central Outer Tracker

The COT is an open-cell cylindrical drift chamber [68] providing tracking at radii between $r = 43.4$ cm to $r = 132.3$ cm. It covers a total z span of 310 cm. It is comprised of 96 layers of sense-wires, grouped into 8 superlayers (numbered 1 –

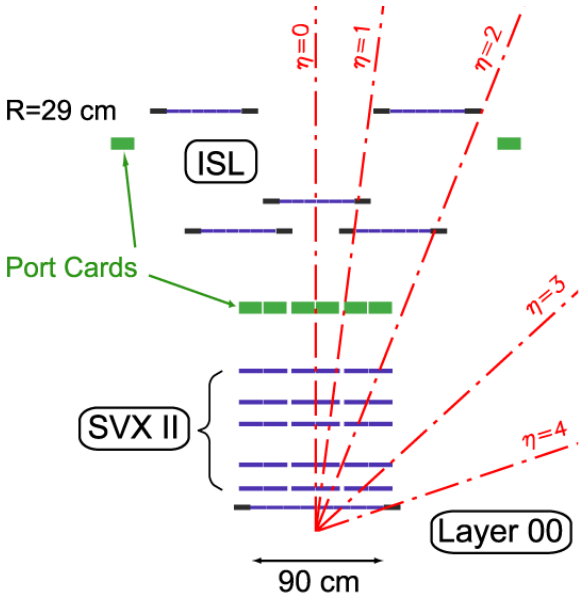


Figure 2-6: Side view of the top of the silicon tracking detectors. The r scale is stretched as compared to z to better display the separate layers of the silicon detectors.

Layer	Strip Orientation	Radius (cm)	Pitch (μm)	Max. Bias (V)
Layer 00	axial	1.35	25	500
Layer 00	axial	1.62	25	500
SVX L0	axial/90°	2.54	60/141	160
SVX L1	axial/90°	4.12	60/126	160
SVX L2	axial/1.2°	6.52	60/60	60
SVX L3	axial/90°	8.22	60/141	160
SVX L4	axial/−1.2°	10.1	60/65	60
ISL L0	axial/1.2°	19.9	112/112	120
ISL L1	axial/1.2°	22.8	112/112	120
ISL L2	axial/1.2°	28.8	112/112	120

Table 2.2: Design parameters of the silicon subsystems. Pitch refers to the implant pitch, with the axial side listed first and stereo second when two numbers are shown. ISL ladders are placed alternating radii. The r values shown is the average r values.

8), for a total of 30 240 sense-wires. The sense-wires in even numbered superlayers are oriented along the axial direction, running along the length in z of the drift chamber, while the sense-wires in the odd-numbered superlayers are tilted by a small stereo angle of 2°. The $r-\phi$ momentum measurements provided by the COT are considerably more precise than its $r-z$ measurements because of the orientation of the sense-wires. The superlayers are divided in ϕ into cells, each of which has 12 sense wires and a maximum drift distance for ionized charge of 0.88 cm. In order to keep this drift distance constant, the number of cells scales approximately with

radius. Keeping the drift distance as low as possible is desirable in order to keep up with the bunch crossing time. The corresponding drift time for the maximum drift distance is $\sim$ 175 ns, which is suitable for the 396 ns bunch separation of current Tevatron operation. Tracks with $|\eta| < 1$ go through all eight superlayers. Tracks with $|\eta| < 1.3$ go through at least four superlayers. The total amount of material in the COT corresponds to 0.017 radiation lengths X_0, for tracks at normal incidence. The hit resolution of the COT is approximately 140 μm, and the track p_T resolution is $\sigma_{p_T}/p_T = 0.0015 \cdot \frac{p_T}{\mathrm{GeV}/c}$. The track reconstruction efficiency is $95 - 99$ % for all 8 superlayers.

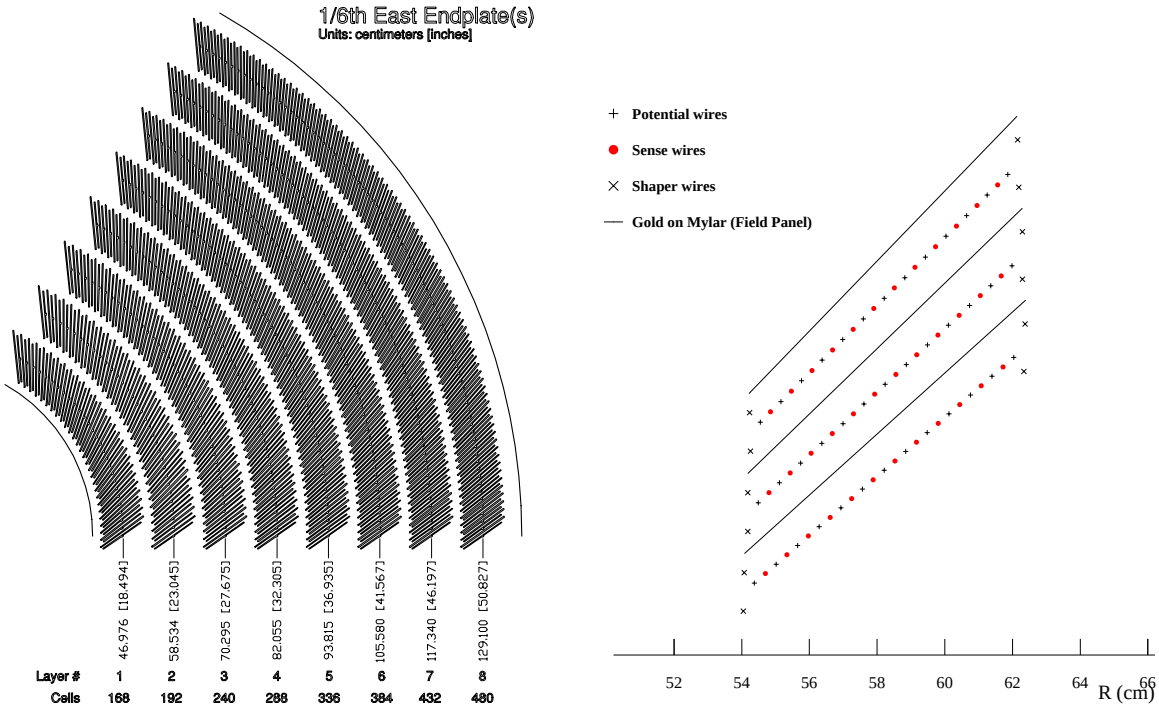


Figure 2-7: End view of a 1/6 slice of the 8 COT superlayers (left) and a cross section of three COT cells in superlayer 2 (right). Note the increase in the number of cells per unit ϕ, roughly scaling with radius.

The physical layout of the superlayers and the organization of wires within a cell are shown in Figure 2-7. Sense wires collect liberated electrons. Potential wires provide additional field shaping. Each cell is surrounded by a cathode sheet, shared between adjacent cells. The sense and potential wires lie in the same plane, are made of gold-plated tungsten and are 40 μm in diameter. The field sheets are made of 6.35 μm-thick Mylar with vapor-deposited gold on both sides. The drift field in normal operation is 1.9 kV/cm. The field sheets are grounded. The wires are each held at a slightly different potential in order to ensure a uniform drift field.

The COT is filled with an argon-ethane gas and isopropyl alcohol mixture in a ratio of 49.5:49.5:1, specified as such in order to have a constant drift velocity across each cell sufficient to keep up with bunch crossing, and in order to simplify timing

calculations. The mixture can be changed to argon-ethane-CF_4 in ratios of 50:35:15. Along with an increase of the drift field to ~ 2.4 kV/cm, this allows a reduction in drift time to 100 ns if the Tevatron were to operate with a bunch separation of 132 ns.

When charged particles pass through a cell, they ionize the gas to yield free electrons, which then drift toward the drift wires. As they approach the wires the free electrons experience an electric field that increases with a radial dependence well approximated by $1/r$, which causes an avalanche of discharge as the original freed electrons gain enough energy to become ionizing themselves. The total gain from this process is $\mathcal{O}(10^4)$. The trajectories of drifting electrons are deflected from the electric field lines by the presence of the solenoidal magnetic field. To compensate for this, the cells are tilted by a Lorentz angle of 35° with respect to the $\hat{r}$ direction also visible in Figure 2-7 such that drift trajectories remain azimuthal.

In addition to tracking, the COT provides a measurement of charged-particle energy loss, $\mathrm{d}E/\mathrm{d}x$, useful for particle ID. The electronics chip that provides readout of the sense-wire signals performs several functions: amplification, shaping, discrimination, and charge measurement. The measurement of charge above a baseline threshold is encoded into the width of the output pulse. This width is proportional to the logarithm of the charge deposited, and is therefore related to $\mathrm{d}E/\mathrm{d}x$ by $\Delta t \propto \log Q \propto \mathrm{d}E/\mathrm{d}x$.

2.2.4 Time-of-Flight

The Time-of-Flight detector (TOF) [69] uses plastic scintillators and fine-mesh photomultipliers to distinguish pions, kaons, and protons by measuring their travel time from the primary vertex to the TOF radius. The primary motivation for its addition to the CDF detector for Run II was to make use of kaon identification to discriminate B meson flavors. This identification works best for low-p_T particles ($p_T \sim 1.6$ GeV/c), because the particle species are well separated in this regime in the momentum-flight time plane. In this dissertation, particle ID information is useful both for signal selection and for the same-side tagging algorithm, where kaon identification is crucial.

The TOF is located in the 4.7 cm radial gap between the tracking system and the cryostat for the superconducting solenoidal magnet. It is comprised of 216 bars of scintillating material, each 279 cm in length and 4×4 cm in cross section. Each bar covers 1.7° in ϕ. They are arranged in a cylindrical configuration covering $|\eta| < 1.0$ at $r = 138$ cm. This radius corresponds to a flight time of ~ 5 ns for the fastest particles coming from the $p\overline{p}$ collisions. Because the photomultiplier tubes (PMT) are placed inside the 1.4 T magnetic field, they suffer an average gain reduction of 500 from the nominal gain of 10^6 for an effective gain of $\sim 30\,000$ on the signal.

Charged particles traversing the scintillator bars deposit energy which is converted to light emitted in the visible range. The photons are transmitted down the length of the bars to photomultiplier tubes on each end. The analog signal is sent through a pre-amplifier in each PMT, then passed to readout electronics, the signal time and amplitude are digitized. The signal follows two paths, performing a timing and charge measurement, respectively. In the first path, the signal is sent to a leading edge discriminator and then through a Time-to-Amplitude (TAC) circuit. In the

second path, an Analog-to-Digital Converter (ADC) integrates the charge in the pulse. The time-to-digital (TDC) information is taken from the time when the signal pulse reaches the fixed discriminator threshold. Because the TDC is dependent on signal amplitude, the effect is corrected by digitized amplitude information from the second path and combined PMT readings. The best time resolution is achieved for large amplitude pulses. The data are also corrected for non-linear TAC response, speed of light in the respective bar, and time of primary interaction. A schematic of the electronics is presented in Figure 2-8.

The TOF contributes significantly to particle identification. Given the flight time t_{flight}, the momentum p and the helical path length L, the mass of a particle is given by

$$m = \frac{p}{c}\sqrt{\frac{c^2 t_{\text{flight}}^2 - 1}{L}}. \tag{2.7}$$

The timing resolution relies on the calibration of responses in each TOF channel, and is ~ 110 ps. This allows for a two standard deviation separation between kaons and pions with momenta $p < 1.6$ GeV/c.

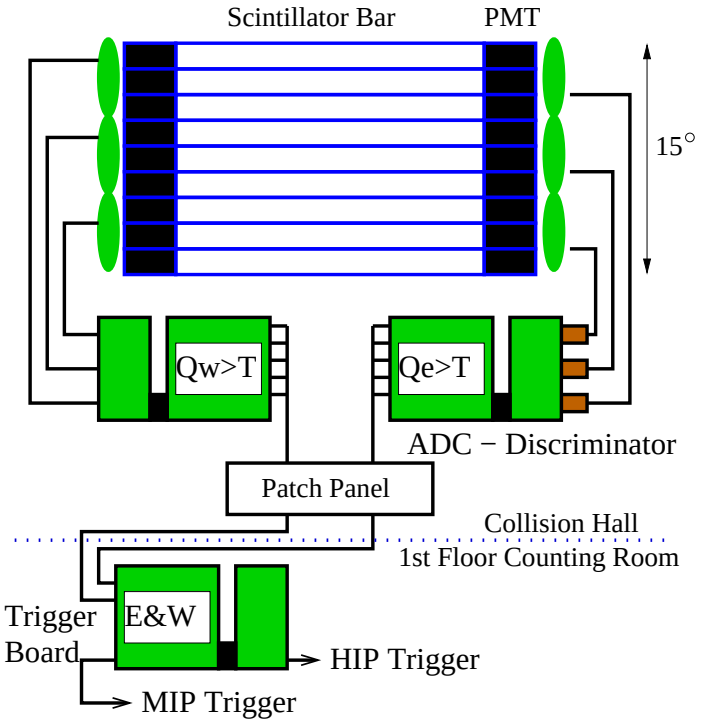


Figure 2-8: A schematic drawing of the TOF detector including scintillators, photomultiplier tubes, electronics and the trigger hardware.

2.2.5 Calorimeters

Calorimeters directly measure the energy deposited by passing particles, and make it possible to distinguish electrons and photons from hadrons. Two categories of calorimeters are used at CDF [70]: electromagnetic (EM) calorimeters are designed to primarily absorb energy from electrons and photons via electromagnetic interactions; hadronic calorimeters are designed to stop long-lived hadrons via strong interactions. Calorimetry is more important in the study of the top quark and in searches for new physics than it is in the study of B mesons. Nevertheless, calorimetry information is used in this analysis in order to perform electron identification for the opposite-side tagging algorithm and in order to build lepton likelihoods used in selection. The CDF calorimeters surround the solenoid with good central and forward coverage. Both EM and hadronic calorimeters are of a sampling design, alternating layers of passive metal absorber with layers of signal producing scintillators.

The CDF EM calorimeters use lead as the absorber to induce Bremsstrahlung in energetic electrons or to produce e^+e^- pairs from photon conversions. Through these two mechanisms, a passing electron or photon will create a shower of secondary particles. For example, an energetic photon converts to e^+e^-, which both undergo Bremsstrahlung, from which more photons pair produce, and so on until the energy of secondary particles dips below the minimum energy threshold of either reaction. Lead is used because of its large cross section for electromagnetic interaction. The energy from the shower is detected by PMTs attached to the scintillator material. The signal strength is related to the energy of the initiator particle. Any remaining spray of particles then goes on to hadronic calorimeters. The energy resolution of the EM calorimeters is $\sigma_E/E \sim \frac{15\%}{E}$ for normal incidence.

The hadronic calorimeters used iron to induce showers via nuclear interactions. Because the nuclear interaction cross section is lower than the EM cross section, a thicker detector must used to reach useful number of interaction lengths. Iron, with a similar cross section as lead, but lighter and cheaper, partially mitigates this difficulty. Charged secondary particles produce signal in the scintillators. The energy resolution in the hadronic calorimeters is lowered by the loss of secondary photons and muons and by the energy lost to excitation and break-up of the absorber nuclei. The energy resolution is $\sigma_E/E \sim \frac{50-80\%}{E}$ for normal incidence. Muons pass through both calorimeters, being minimum ionizing particles.

The combined calorimeters cover a total range of pseudorapidity $|\eta| < 3.6$. The calorimeters are segmented into towers in η and ϕ, shaped and oriented to point back to the $p\overline{p}$ interaction region. They are divided into central (CEM [71], CHA [72]), wall (WHA [72]) and plug region (PEM [73], PHA) detectors. In the central region, the detectors cover $|\eta| < 1.1$, with each tower having (η, ϕ) dimension of $(0.1, 15\ ^\circ)$. The plug region detectors extend η coverage to $|\eta| = 3.6$. The towers cover intervals in η that range between 0.1 and 0.6, and intervals in ϕ that range between 7.5 and 15 °. In the wall region, the WHA bridges the gap between the central and plug detectors, with towers matching the EM calorimeter segmentation. Figure 2-9 shows an elevation view of the plug and wall calorimeters and their position relative to the central calorimeters.

The EM calorimeters are supplemented by the showermax detectors, for example the CES in the central region [74], for maximum energy resolution at the depth of maximum energy deposition. In addition pre-shower detectors, such as the CPR in the central region [75] are used to assist in electron identification. The CES is a strip and wire chamber used to measure the energy distribution at a normal depth corresponding to the maximum of the transverse shower profile. The CPR used in CDF was originally a wire chamber but was upgraded to a new design using scintillator tiles with wavelength-shifting fibers read out by PMTs. It is placed in front of the EM calorimeter, and used to detect early particle showers.

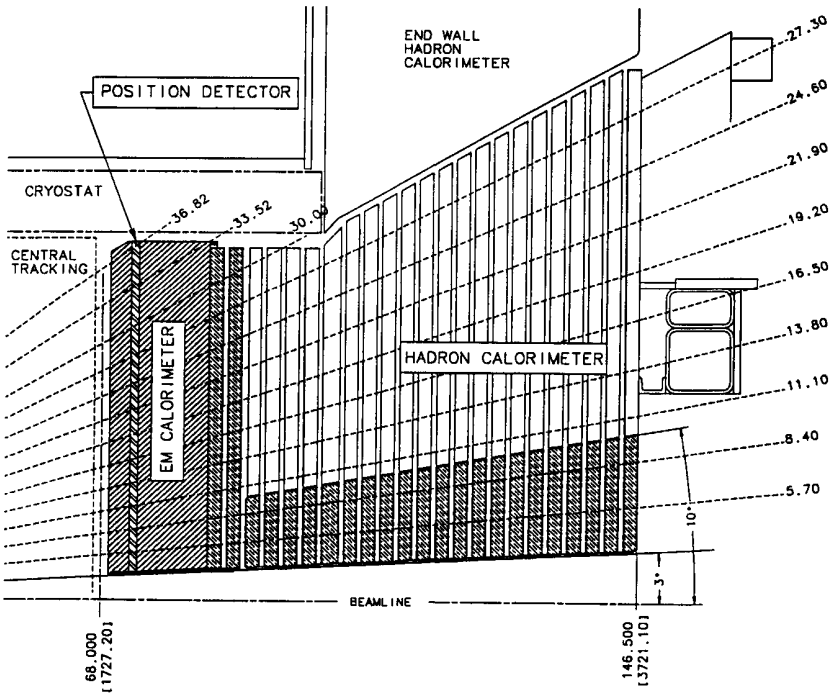


Figure 2-9: Elevation view showing the plug EM and hadronic calorimeters, and end wall hadronic calorimeter in relation to the central calorimeters.

2.2.6 Muon Detectors

In this analysis, muon identification and triggering is of particularly high importance. The $B_s \to J/\psi\phi$, $B_d \to J/\psi K^*$ and $B_u \to J/\psi K^+$ candidates that we use come from the di-muon trigger. The reconstruction of muon tracks in three dimensions is also of central importance because of the use of and angular analysis to separate *CP*-odd and -even decays.

Muons do not interact with atomic nuclei in the hadronic calorimeters. Furthermore, muons are minimum ionizing particles, meaning that their mean energy loss as they pass through matter is very low, and thus they do not produce large showers

in the EM calorimeters either. Therefore muon detectors are placed on the outside of the CDF detector, behind layers of steel shielding. Muons are detected by reconstructing short tracks at the outermost radii of the detector such that they can be matched to tracks formed in the integrated tracking system. These few short hits in the muon detectors with coarse hit resolution are called stubs. The placement of the muon detectors at the outer radii also means that few other particles reach these detectors, as most are absorbed either in the calorimeters or in the steel shielding. It also means that low energy muons, with p_T less than ~ 1.5 GeV/c, are absorbed in the inner detectors and do not reach the muon chambers. Finally, the additional shielding results in multiple Coulomb scattering, which deflects the muon trajectory and must be taken into account in track-to-stub matching.

There are four muon detectors in use at CDF in Run II, each consisting of a drift chamber and matching scintillator used for triggering [76]. The four drift chambers are: the Central Muon Detector (CMU), Central Muon uPgrade (CMP), Central Muon eXtension (CMX), and Intermediate Muon system (IMU) [77, 78]. The scintillators (CSP, CSX, and BSU) are placed near each muon detectors to measure and minimize backgrounds coming from out-of-time interactions in the beampipe. The Run II muon system is a significant upgrade to the Run I system, including but not limited to an extension to full azimuthal coverage for the central systems, and the entirely new IMU. Figure 2-10 shows the location in $\eta - \phi$ of the four muon systems in use at CDF in Run II. Table 2.3 summarizes a few of the relevant design parameters of these detectors.

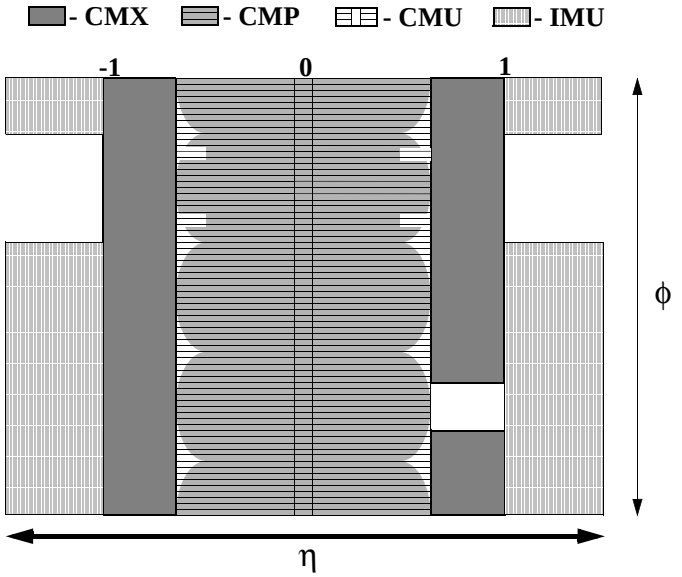


Figure 2-10: The location of the CMU, CMX, CMP, and IMU in the $\eta - \phi$ plane in CDF-II.

The CMU, which is the original Run I central muon system, is positioned around the CHA at $r = 347$ cm, and provides coverage $|\eta| < 0.6$. It is made of 144 ϕ

segments, or modules. Each module has four layers of four rectangular drift cells divided into East and West portions, with each cell containing a 50 μm stainless steel wire at the center, in the same argon-ethane-alcohol mix as in the COT. The sense-wires for the first and third (and second and fourth) radial layer are ganged together in readout. Each wire pair is instrumented with a TDC to measure the muon's ϕ location and and ADC to measure the muon's z location. The maximum drift time in the CMU is $\sim$ 800 ns. The minimum muon p_T required to reach the CMU is $\sim$ 1.4 GeV/c. The hit position resolution in the CMU is 250 μm in $r-\phi$ and $\sim$ 1.2 mm in z.

The CMP provides the same $|\eta|$ coverage as the CMU, behind an additional 60 cm of steel, which helps reduce misidentification rate by absorbing penetrating hadrons. The set of muons with stubs in both the CMU and CMP is of particularly high purity. The CMP chambers are rectangular, single-wire drift tubes configured in four layers with alternate half-cell staggering. Signals are readout by a single TDC per wire. The maximum drift time in the CMP is $\sim$ 1400 ns. The minimum muon p_T required to reach the CMU is $\sim$ 2.2 GeV/c. Both CMU and CMP provide full ϕ coverage.

The CMX provides an extension in pseudorapidity coverage to $0.6 < |\eta| < 1.0$. Although no additional steel shielding precedes the CMX, additional absorbing material is present because of the longer path through the the hadronic calorimeter, magnet yoke and original steel shielding at the larger angle of tracks that reach the CMX. The CMX originally had two gaps in ϕ adding up to 90°, which were filled in with an upgrade performed in 2004. The CMX has four layers of twelve drift tubes for each 15° ϕ section. The maximum drift time in the CMX tubes is $\sim$ 1400 ns.

The IMU provides coverage for $1.0 < |\eta| < 1.5$. It consists of a barrel-shaped array of muon chambers (BMU), and two arrays of scintillators, one barrel-shaped mounted parallel to the beamline (BSU), and one ring-shaped mounted perpendicular to it (TSU). The BMU consists of single-wire rectangular drift tubes mounted in four half-cell-staggered stacks, each covering 1.25° in ϕ. The maximum drift time in the CMX tubes is $\sim$ 800 ns. The BSU scintillators each cover 2.5° in ϕ and a range of 1.25 in η. The TSU scintillators each cover 5° in ϕ.

	CMU	CMP/CSP	CMX/CSX	IMU
Pseudorapidity range	$\|\eta\| < 0.6$	$\|\eta\| < 0.6$	$0.6 < \|\eta\| < 1.0$	$1.0 < \|\eta\| < 1.5$
Azimuthal coverage (°)	360	360	360	360
Max. drift time (ns)	800	1400	1400	800
Number of channels	2304	1076/269	2208/324	1728
Pion interaction length	5.5	7.8	6.2	6.2–20
Min. μ p_T (GeV/c)	1.4	2.2	1.4	1.4–2.0

Table 2.3: Design parameters of the muon detectors. The IMU is new to Run II of CDF. The CMX provides extended muon coverage in Run II. Assembled from References [76, 77, 78].

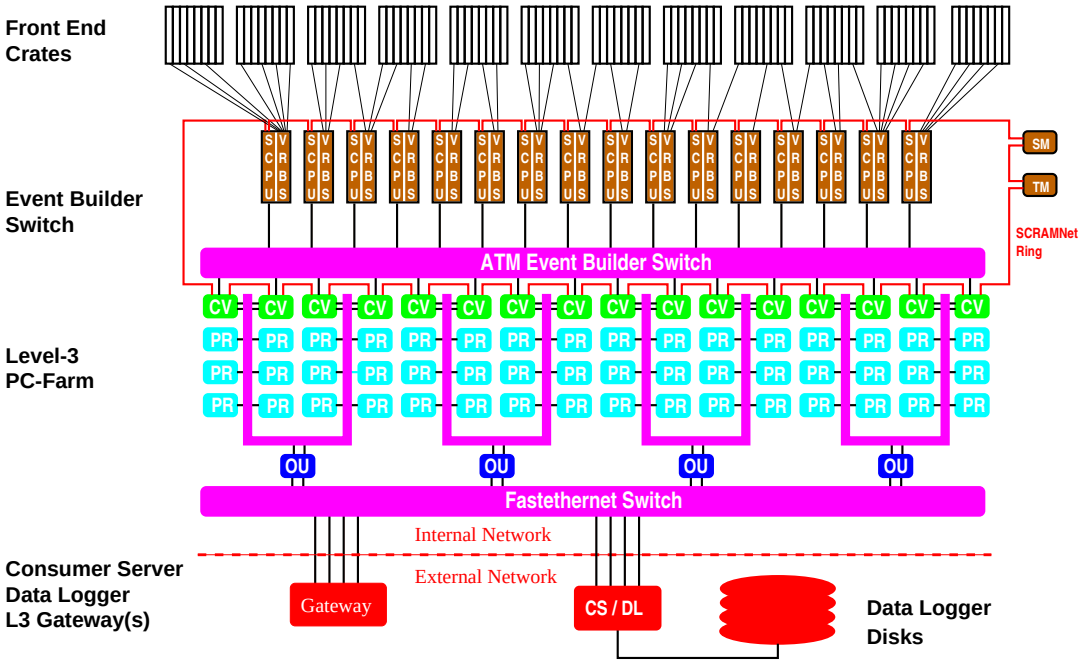


Figure 2-11: Block diagram of the DAQ system in use for much of CDF Run II. The EVB upgrade replaces the ATM switch with a Gigabit Ethernet switch and the SCRAMnet ring with an ethernet network.

2.3 Triggers and Data Acquisition

Proton and antiproton bunches cross every 396 ns at the B0 interaction point. The total $p\overline{p}$ inelastic cross section is approximately 75 mb at $\sqrt{s} = 1.96$ TeV. Together these conditions yield an interaction rate of 2.5 MHz for typical Tevatron instantaneous luminosities. Reading out the data from every detector subsystem for one event takes approximately 2 ms, which prohibits the recording of every event, ignoring even all other constraints. Furthermore, the average data size for one event is 250 KB, which would require a prohibitively large data recording rate in excess of 600 MB/s if every event were recorded. However, the large majority of events resulting from the collisions constitute background to the much rarer events of interest for physics study at the Tevatron. For instance, four orders of magnitude separate the total inelastic cross section from the much smaller $b\overline{b}$ cross section, and nine order of magnitude separate the former from the total $\overline{t}t$ cross section. The CDF trigger system circumvents the timing and storage problems that arise from all-inclusive event recording by making real-time decisions to accept or reject each event based on whether it contains any signatures of interest.

CDF uses a three-level trigger system, with each successive level using increasingly more complete information from the detector and more complex (slower) decision algorithms. An event must receive an accept decision at each level before being considered at the next level, which allows dramatic reductions in input rate at each trigger level.

2.3.1 DAQ System

Figure 2-11 shows an overview of the CDF Data Acquisition (DAQ) system. The CDF DAQ system [79, 80] collects data fragments from the front-end electronics of each detector subsystem and transfers the data for trigger-selected events to permanent storage. The DAQ system is tightly integrated with the triggering system to allow for optimal data flow. Figure 2-12 shows a functional block diagram of the CDF Run II readout electronics.

The front-end and trigger electronics are packaged as VME modules that reside in about 120 crates in the system. These modules process detector signals and make the data available to the DAQ system through the VME bus. Each front-end crate contains at least one processor board for hardware initialization and event readout, running under the VxWorks operating system. The crates contain also a controller module which distributes, through the VME backplane, received timing synchronization signals. These timing signals are dictated by the Tevatron clock, coherently with the bunch crossings, and ensure global synchronization of readout electronics and trigger.

The controller module further provides the interface of the VME modules to the Trigger Supervisor Interface (TSI), as well as the data interface to the EVB. The TSI is responsible for receiving the decisions from the hardware triggers (Level 1 and Level 2), communicating them to the front-end crates, and supervising the data flow up until it is transferred to the Event Builder.

Until data fragments reach the Event Builder (EVB) [81], no system has under its control the full data for an event. The data fragments from all the front-end crates are sent to the EVB upon Level 2 accept. The EVB then assembles all the fragments for each event, does some basic data integrity checks, and passes each event to the Level 3 processing farm. In 2005, the EVB received an upgrade to allow an increase in the Level 2 accept rate from 300 Hz to 1000 Hz [82].

The data are received on the EVB side by VME Readout Boards (VRB), which are linked by optic fibers to the front-end crates. The VRBs reside in 22 EVB crates, each of which is controlled by a Scanner CPU (SCPU) running a standard Linux distribution. The EVB crates are connected to the Level 3 farm by 18 *converter* nodes, standard rack computers which pass the events to the 20-21 *processor* nodes in each of the corresponding 18 subfarms in the Level 3 computing farm. Data flows from the EVB to the Level 3 farm through a Gigabit Ethernet switch. From the Level 3 farm, events are passed through 9 *output* nodes to the Consumer Server Logger, for permanent storage.

The last processing stage required before data is used in analysis is called *production*, in reference to the `ProductionExe` executable that carries out this task. Production is not performed in real time but offline, in order to take advantage of more elaborate algorithms and more accurate, up-to-date detector calibration and alignment information. Aside from this difference, this processing stage is very similar to the Level 3 stage in that it involves the full reconstruction of tracks, vertices, leptons, photons, jets, and other physics objects used in analysis. The computing structure used in this processing is discussed in Reference [83].

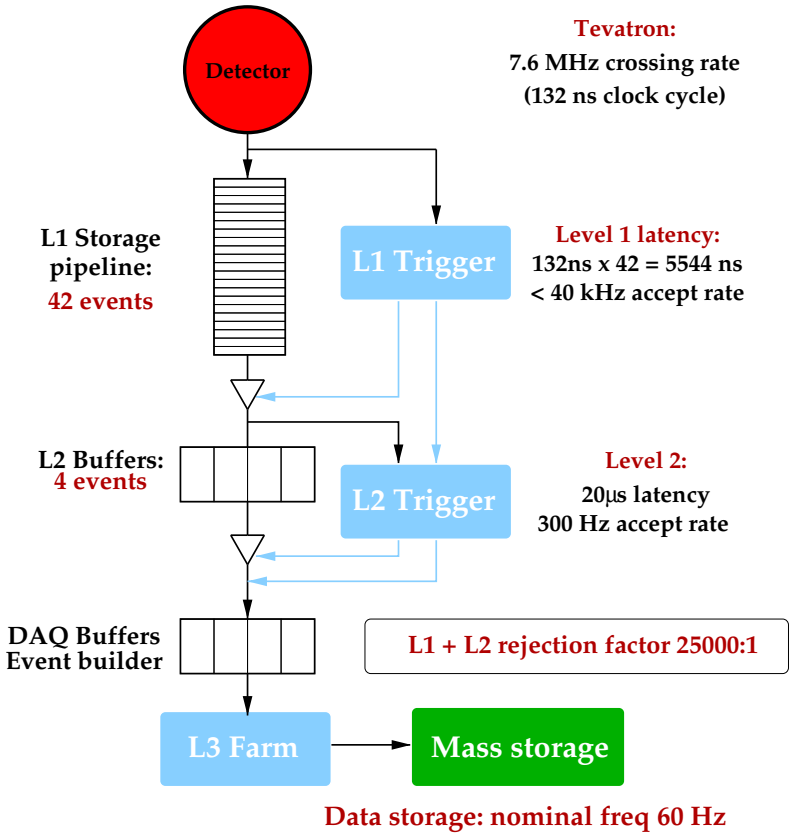


Figure 2-12: Block diagram of the CDF Run II readout electronics. The Tevatron was designed to be capable of running with 132 ns bunch separation, even though in actuality the bunch separation has remained 396 ns in Run II. Note also that the Level 2 accept rate was increased to 1000 Hz after the EVB upgrade.

2.3.2 Trigger System

Figure 2-13 shows a block diagram of the CDF Run II Trigger system, showing the three levels of the system, and indicating links between trigger levels and detector information [61]. We give an overview of the trigger system, with particular focus on the elements of interest for this analysis.

Level 1

The Level 1 trigger reaches a decision to accept or reject an event in approximately 5.5 μs. In order to run as a synchronous system, it uses an event pipeline with buffers that are 42 clock cycles deep, corresponding to 3 $\times$ 14 bunch crossings for 396 ns bunch spacing. The Tevatron was designed to allow bunch spacings as low as 132 ns. An event moves forward through the buffer at each clock cycle, and Level 1 reaches a decision by the time the event reaches the end of the buffer. The Level 1 accept rate

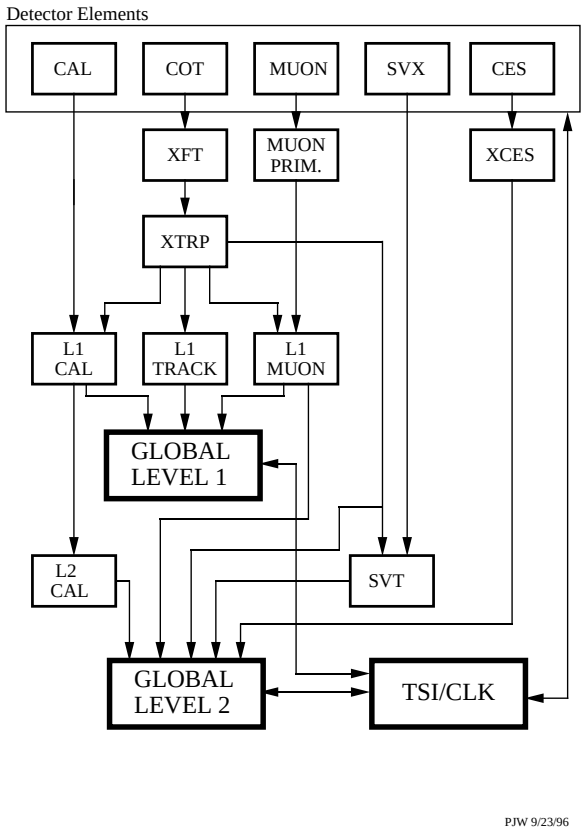


Figure 2-13: CDF Run II Trigger system block diagram. The XFT and XTRP are discussed in detail in the text. The Level 1 system is the most important to the analysis presented in this document, as it includes the di-muon triggers. The SVT plays a crucial role in the development of tagging algorithms discussed in Chapter 5

is approximately 50 KHz.

Level 1 uses event fragments from the COT, the calorimeters and the muon systems to form coarse versions of physics objects (tracks, electrons, muons) called trigger primitives in order to make decisions. Several algorithms are in place at Level 1 to identify events of interest. The calorimeter trigger is used to find electrons and jets and to calculate the summed transverse energy ($\sum E_T$) and missing transverse energy ($\not\!\!E_T$). The di-muon trigger uses information from the COT and the muon systems to identify muon signatures. This information is processed by two systems, the eXtremely Fast Tracker (XFT) and the eXTRaPolation unit (XTRP).

Track primitives with $p_T \sim 1.5$ GeV/c or above are identified by the XFT. The XFT determines the rough p_T and ϕ_6 (ϕ at COT superlayer 6) of the track using the four axial superlayers of the COT. The XFT reports track p_T, ϕ and charge with an uncertainty of $\sigma_{p_T} \sim 1.7\%/\text{GeV}/c$, $\sigma_\phi \sim 5$ mrad, or about 10 times larger than the uncertainty on those parameters in offline reconstruction. The XFT proceeds in two steps in order to identify a track: segment finding and segment linking. In segment

finding, wire hits are classified as "prompt" (drift time < 66 ns) or "delayed" (drift time between 66 and 220 ns). Adjacent COT cells are grouped together four at a time and the hit information is compared to a set of predefined patterns to find track segments. Track segments are then grouped together in sets of four (each from a different superlayer) to give a rough estimate of track parameters. The predefined patterns ensure that all found segments originate close to the beamline and have a high enough p_T given the pattern.

The XTRP receives track information from the XFT and extrapolates the tracks to the calorimeters and muon systems to look for matches in energy deposition or muon stubs. Muon and di-muon primitives are derived from hits in the muon chambers or coincidences of hits with the scintillators. This provides information which can be used immediately at Level 1 for triggering decisions.

The Di-Muon Trigger

All analysis samples used in this dissertation come from the CDF di-muon trigger. The di-muon trigger relies on a clear signature of two muons coming from $J/\psi \rightarrow \mu^+\mu^-$ decays. In order to make trigger decisions, it uses the XFT tracking and muon system information available at Level 1. Levels 2 and 3 play a small role in the event selection decision. Level 2 is used to tighten any existing requirements of Level 1, e.g. on the transverse momentum, and Level 3 uses more precise determination of several event variables, such as the transverse momentum of tracks, better track-stub matching, di-muon mass, etc. Because the di-muon trigger plays such an important role in this analysis, we isolate it for a detailed description here. It is important to note that it is one among $\mathcal{O}(100)$ separate triggers used at CDF for analysis.

Although we refer to it as a single entity, the di-muon trigger is in fact a combination of two triggers: CMU-CMU, where both muons are found in the most central muon chamber, and CMU-CMX, where one muons is found in the CMU and one in CMX. We describe the CMU-CMU trigger, and then comment on the differences in the CMU-CMX.

The following terminology is specific to triggering on CMU muons. A *stack* is a set of four drift cells stacked on top of each other. The CMU has 288 stacks in each of the East and West sides of the detector. A *Level 1 stub* is a track segment in a stack such that cells 1 and 3 or cells 2 and 4 have hits separated by no more than 396 ns ("or" is the mathematical $\vee$: the statement is true when one or both are true). A *tower* is a set of two neighboring stacks. A tower has *fired* when one or both stacks have a Level 1 stub, and is *empty* otherwise. A *muon tower* is a fired tower matched with an XFT track.

In order to keep the Level 1 decision time short enough to remain synchronous, only information about which towers have fired is used in triggering, rather than detailed hit positions and direction. The XFT reports the p_T and ϕ_6, as well as the charge of the track to the XTRP. The XTRP extrapolates this track to the CMU radius and creates a *footprint*, a 3 σ window in ϕ (wide enough to account for multiple Coulomb scattering). If a tower is found within that footprint, it is a muon tower. The CMU-CMU trigger requires that at least two muon towers be found such that

they are either on opposite sides of the detector or are separated by at least two other towers.

The CMU-CMX trigger uses much of the same algorithm. The changes to the decision algorithm arise from the differences between the CMU and CMX detectors. In the CMU-CMX case, only XFT tracks with $p_T > 2.2$ GeV/c are used to match to the CMX as the extra material that muons pass through to reach the CMX limits further the momentum requirements on the muon, and no azimuthal separation is required because the muons are by definition in different subdetector volumes.

Trigger algorithms are among the few elements of the experimental apparatus which are continuously improved and optimized, as this process does not require performing expensive and time-consuming hardware upgrades to the detector. The dimuon trigger has undergone constant revision in order to carry out such optimization. While the core logic outlined above is more or less constant, other parameters have been changed often to improve the trigger. Such parameters include requirements on the p_T of the XFT tracks, the difference in ϕ between the two muons $\Delta\phi$, and their transverse mass M_T. In addition, some of the triggers are prescaled, which means that only one our of N events are kept. This is done in order to deal with periods of high luminosity when triggering on every event that passes the nominal requirements would overwhelm the DAQ system. An improvement of static prescaling is dynamic prescaling, where the trigger automatically adjusts its prescale over the data taking period as luminosity decreases, in order to accept a larger and larger fraction of the fewer and fewer number of events that are passing though the DAQ as the luminosity drops. The various combinations of these requirements result in slightly different trigger requirements. The list of all the specific requirements used for the data in this analysis is given in Appendix B.

Level 2

Level 2 is an asynchronous decision system, using better precision and additional event information, reaching a decision about each event in ~ 20 μs. The Secondary Vertex Trigger processor uses $r - \phi$ hits in the silicon detector to extend the XFT tracks into the SVX. This provides d_0 information, as well as improved p_T and ϕ determination, allowing the trigger to distinguish between primary and secondary particles.

The SVT design reflects the SVX's 12-fold azimuthal symmetry and 3-barrel structure: calculations are performed in parallel for each ϕ-sector and barrel. SVT candidates require a coincidence of XFT tracks and hits in four out of the five SVX layers.

The ℓ+SVT Trigger

The ℓ+SVT trigger (ℓ is for lepton) uses the SVT to collect events coming from semileptonic decays of B mesons, $B \to D^{(*)}\ell\nu X$. In addition to requiring a displaced vertex, it also requires a lepton candidate in the EM calorimeters or in the muon chambers with one displaced track from the SVT. This trigger is used to develop the

opposite side tagging algorithms discussed in Section 5.2. The algorithms that make up this trigger are detailed in Reference [84].

Level 3

At Level 3, the event data is nearly at the same quality as in offline reconstruction. Level 3 uses the complete event record and makes a decision in approximately one second. It uses tight matching of tracks to calorimeter towers and muon stubs. As an event passes through the trigger system, it is only at Level 3 that 3-dimensional tracks and decay vertices are reconstructed, and L_{xy} is calculated. The mass of the particle is calculated and is used in trigger decisions.

Chapter 3

Sample Selection

The data used in this dissertation was collected at the CDF experiment between February 2002 and January 2007. The following decay modes are reconstructed:

- $B_s \to J/\psi\phi, \quad J/\psi \to \mu^+\mu^-, \quad \phi \to K^+K^-,$
- $B_d \to J/\psi K^*, \quad J/\psi \to \mu^+\mu^-, \quad K^* \to K^+\pi^-,$
- $B_u \to J/\psi K^+, \quad J/\psi \to \mu^+\mu^-,$

where the charge conjugate modes are implied. After applying basic detector and data quality requirements to ensure that all relevant subsystems are functioning properly, the total integrated luminosity of the B_s data set used for the lifetime and angular analysis corresponds to 1.7 fb^{-1}. The data set used for the tagged measurement of CP violation in B_s and the B_d data set correspond to 1.3 fb^{-1} collected*. The B_u data set used for flavor-tagging studies corresponds to an integrated luminosity of 1 fb^{-1}, which is sufficient for development studies because there are sufficient numbers of events available in this dataset.

The data are collected using the di-muon triggers. The set of all events accepted by the di-muon triggers is itself a superset of the analysis data used here, as nearly 80% of the events they select result from "prompt" J/ψ production from $p\overline{p}$ collisions. This brings up a common feature of analyses in experimental particle physics: the need to distinguish events arising from the physics processes we seek to describe or discover, i.e. signal events, from events that pass our data acquisition and triggering filters but in fact arise from a variety of other processes, which we collectively refer to as background events. This discrimination usually cannot be carried out to the total exclusion of all backgrounds because of the irreducible limitations of our detector and decision algorithms in distinguishing signal from background. An auxiliary goal of our sample selection is therefore to study the irremovable background events in the variable space we use to study our signal events.

*The energy loss per unit length $\mathrm{d}E/\mathrm{d}x$ measured in the COT drift chamber is a crucial variable for the same side tagging algorithms used to perform the tagged measurement. The necessary calibrations of this variable over the data-taking period beyond the first 1.3 fb^{-1} collected at CDF were a separate and significant undertaking. These calibrations became available after this analysis was complete, and therefore we excluded this portion of the dataset from the tagged analysis.

The decay mode of primary interest is $B_s \to J/\psi\phi$. The $B_d \to J/\psi K^*$ decay mode is used to fill two roles: validation of our method of modeling the angular acceptance of the CDF detector in the transversity variables, and validation of the tagging algorithms used for the measurement of *CP* violation in $B_s \to J/\psi\phi$ using flavor tagging. The $B_u \to J/\psi K^+$ decay mode is used in conjunction with the $B_d \to J/\psi K^*$ decay mode in the tagging validation. Simulated data is used to study signal and background distributions, and to study the detector acceptance in the transversity variables for the angular analyses. For the latter case, the same selection criteria are applied to data and simulated data.

In this chapter we describe the requirements placed on events and the various components of each event for inclusion in the analysis samples. We reiterate briefly the di-muon trigger requirements discussed in Section 2.3.2. We then outline the additional requirements placed on tracks and physics objects in event reconstruction, followed by our loose pre-selection cuts. Finally we explore in detail the neural networks used for final candidate selection in each analysis.

3.1 Trigger Requirements

The first set of requirements imposed on events for inclusion in our analysis is enforced by the trigger system in real time as data is being collected at CDF. The di-muon triggers are used to collect the B_s, B_d, and B_u decays we use to develop this analysis. The ℓ+SVT trigger introduced in Section 2.3.2 is used extensively in the development of the flavor tagging algorithms discussed in Chapter 5. However, this development happened prior to and independently of the work presented here. The ℓ+SVT trigger was not used for the collection of data in this analysis. We therefore forego an in-depth discussion of this trigger, and refer the reader to References [84] for more details on the matter.

As noted earlier, the di-muon triggers are a collection of trigger paths which have evolved over time. The requirements of a basic di-muon trigger are summarized by the following:

A. Level 1

- two XFT tracks with opposite charge,
- each track is matched with two muon stubs,
- each CMU (CMX) muon has $p_T^{\rm XFT} > 1.5(2.2)$ GeV/c^2
- $\Delta\phi_6$(CMU, CMU) $< 135^\circ$ for some paths, no cut in $\Delta\phi_6$(CMU,CMX),

B. Level 3

- $2.7 < M_{\mu\mu} < 4$ GeV/c^2

For detailed additional criteria for all the trigger paths used in this analysis, we refer the reader to Appendix B.

3.2 Event Reconstruction

3.2.1 Track Preparation and Refitting

All the physics objects in this analysis are reconstructed starting with tracks (as opposed to, for example, an analysis of di-photon events). In order to maximize the precision and accuracy of the track parameters, the tracks are refitted after the generic offline production to use the most accurate available information [85]. As an additional requirement on track quality, we only use tracks with at least 10 axial hits and 10 stereo hits in the COT drift chamber, and r–ϕ hits in at least 3 separate SVX layers.

Beside this track quality requirement, the track refitting algorithms make several improvements. In the generic offline production, all tracks are assumed to be pions. However, particle type will affect the amount of energy loss. Tracks are therefore refitted using appropriate particle type hypotheses. Moreover, the uncertainties on COT hit positions do not initially take into account multiple scattering, and are therefore underestimated. These uncertainties are corrected by empirical scale factors in the refitting.

Layer 00 hits are added in track refitting whenever available, as they are not present in the generic production. In this measurement, where information about the full time evolution of the B meson from production to decay is being used, the addition of Layer 00 is of great importance in improving the impact parameter resolution. In generic production, the track reconstruction efficiency is 95% for tracks passing through all 8 COT superlayers with $p_T > 400$ MeV/c and 99% for tracks with $p_T > 2$ GeV/c. The efficiency for adding silicon information from SVX and ISL is 93%. In track refitting, the efficiency for adding L00 hits is 65% [86].

3.2.2 Additional Muon Requirements

Muons are subject to additional requirements because they are used by the di-muon trigger to make an accept/reject decision. Muons in the CMU are subject to a requirement that the track and muon stub are compatible by requiring a $\chi^2 < 9.0$ for the match between the track and the stub. To ensure that the p_T determined by the XFT did not subsequently dip in track refitting below the minimum p_T to reach the muon chambers, we reimpose a minimum p_T requirement of 1.5 GeV/c.

Finally, only muons passing through the CMU or the CMX pass the Level 1 di-muon trigger requirements. In several parts of this analysis, we make use of simulation, including simulation of detector response and triggering systems. In order to make sure that the features of the events in our data samples can be reflected as closely as possible in simulation, we require trigger confirmation for the muons, meaning that the muons that are reconstructed to have come from the B decays are the ones that were used by the di-muon trigger, rather than other muons in the event.

3.2.3 Vertex Fitting

The next step in event reconstruction after track (re)fitting is vertex fitting. Vertex fitting allows us to determine whether tracks observed in the tracking system originate directly from the $p\overline{p}$ interaction or from the secondary decay of a relatively long lived particle happening prior to interaction with the tracking system. The reconstruction and fitting of decay vertices is carried out by the `CTVMFT` vertex fitting package [87]. This algorithm performs a three-dimensional fit of two or more helices obtained from the reconstruction of tracks to a common origin. The `CTVMFT` algorithm determines the coordinates of the decay vertex, the uncertainty on those coordinates. The algorithm also calculates an overall χ^2 probability $P(\chi^2, n)$, taking into account the number of degrees of freedom n in the track fit, as an indication of the goodness of fit. Vertices are assigned a transverse momentum $\vec{p_T} = \sum \vec{p_T}^{trk}$, the vector sum over the transverse momenta of their constituent tracks. The mass of the decay particle is calculated from the four-vectors of the decay products. The following quantities are used in the selection of a signal sample: $P(\chi^2, n)$ of the vertex fit, spatial coordinates and kinematic properties of the vertex.

Vertex fitting begins with stable tracks (e.g. pions, kaons, muons) and works step by step backward up the decay chain, reconstructing intermediate vertices until the original "mother" decay vertex is found. In this way, reconstruction of a decay such as $B_s \to J/\psi\phi, J/\psi \to \mu^+\mu^-, \phi \to K^+K^-$ begins by reconstructing $J/\psi \to \mu^+\mu^-$ and $\phi \to K^+K^-$, and then matches the J/ψ and ϕ to a common vertex. In combining the stable tracks, tracks already used for another vertex in the event are rejected, and only physically meaningful charge combinations are considered so as to avoid spending CPU time on the reconstruction of impossible decay chains.

3.2.4 Primary Vertex Fitting

The calculation of the proper decay length of an unstable particle requires the measurement of two spatial coordinates: the decay vertex and the primary vertex (PV) of the $p\overline{p}$ collision, where the unstable particle is produced. Optimal determination (i.e. with smallest possible uncertainty) of both these coordinates contributes to a more precise measurement of the distance between production and decay.

The distribution of primary vertices over the entire data-taking lifetime of the CDF experiment is well-known and well-studied. The structure of the beam envelope, or *beamline* comes mostly from the beam shaping effected by the Tevatron's quadrupole magnets, whose purpose is to minimize the transverse beam profile at the collision point to maximize the instantaneous luminosity. This structure is well-modeled by an hourglass shape whose central position in the $x - y$ plane is a linear function of the beam direction z. Using this information in combination with the information on the B decay vertex from `CTVMFT`, the PV for an event is determined as the beamline position at the z_0 coordinate extrapolated from the B vertex. This is called the beamline method of determining the primary vertex. The beamline method suffers from a large beam profile for most z positions, ranging from 35 μm in diameter at $z \sim 0$ to 60 μm at $|z| \sim 40$ cm, which results in a corresponding uncertainty on the

position of the PV as determined by the beamline method.

Because each $p\overline{p}$ event produces tracks other than the tracks used to reconstruct a B candidate (if one is present), we have additional information available to determine the position of the primary vertex by fitting the additional tracks in an event to a common origin. The estimate obtained by this method is referred to as the "event-by-event primary vertex" [88, 89]. In this method, the PV fit is performed separately for each B candidate, excluding the tracks used for reconstruction of the candidate.

The two methods are used in conjunction, since they form independent measurements of the primary vertex position, in order to improve the precision of the PV estimate. The beamline and separately fitted PV measurements are combined in an average weighted by their respective uncertainties. We obtained an uncertainty on the position of the beamline-constrained event-by-event primary vertex of ~ 20 μm.

3.2.5 BStntuple

To minimize the computation time and storage space needed to store and analyze the data in our analysis, we use the `BStntuple` framework [90], which is an extension of the `Stntuple` [91] framework. The `BStntuple` contains structures to hold the reconstructed candidates information, stable and decaying objects, as well as information needed for flavor-tagging (decision, raw dilution), and particle identification information (TOF, $\mathrm{d}E/\mathrm{d}x$, muon and electron quantities). Data is stored in an efficient format, allowing several potential decay candidates in the same event to share links to common data blocks.

3.2.6 Monte Carlo Simulation

In this analysis, simulation of B production and decay processes and of the subsequent detector response is used for two primary purposes: to model the acceptance of the combined detector and trigger system in the space of transversity angular variables described in Section 1.4, and to train the neural network used in candidate selection to recognize and differentiate between signal and background events. An analytical treatment of all the interactions involved from the $p\overline{p}$ collisions to B production, decay, and interaction with matter in the detector is simply impossible with current computational methods and resources. Such a treatment is in any case unnecessarily detailed for the tasks described above. The alternative is to use numerical simulation to carry out these tasks. The algorithms used all involve some type of random sampling to simulate processes, and are collectively called *Monte Carlo* simulation. The simulation is divided into several steps, which echo in order the physical processes involved in collecting data.

The first step in simulation is the treatment of the $p\overline{p}$ hard scattering, and the outgoing quark and gluon collision products, followed by simulation of the fragmentation and hadronization processes which yield hadrons and associated jets. The simulation package `Pythia` [92] carries out this simulation comprehensively for each event, producing outgoing quark-antiquark pairs and gluons in all directions, with associate initial- and final-state radiation and full fragmentation with multi-hadronic

final states. By contrast, the BGenerator package [93] concentrates on producing only one $b\bar{b}$ pair per event, which yields a great advantage in computational speed, but by design does not mimic the full collision environment. The former is useful in particular in studies of same-side kaon tagging, where the topology of the entire event is crucial to developing the flavor tagging algorithm. For our purposes, BGenerator is preferable and sufficient since we wish to model single B decay samples. We use BGenerator to produce samples where each event includes a single B_s or single B_d.

The second step is the simulation of the full decay chain of the B mesons under study. For this task, we use the EvtGen package [94]. EvtGen is specialized for heavy flavor decays and accounts correctly for quantum mechanical interference effects. In order to model the detector angular acceptance for both $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$, we use the phase-space decay model of EvtGen, which yields flat distributions in the angular variables whose acceptance we wish to study.

The third step in simulation incorporates the interaction of the decay products ($\mu\mu KK$ and $\mu\mu K\pi$) with the detector material. For this task we use the cdfSim package [95], which is a CDF-specific full detector simulation based on the GEANT simulator [96]. The final step is the simulation of the triggering and event reconstruction that data events pass through. Because cdfSim outputs simulated events with the same data banks as actual events, the triggering and reconstruction is performed in precisely the same manner as for data events. All Monte Carlo samples are then passed through the same candidate selection as the data, the details of which are provided below.

3.3 Candidate Selection

Signal and background events are distinguished using kinematic variables. We study the distributions of kinematic variables using events in the mass sidebands in data and events in the signal region in simulated data to determine which variables have discriminating power. The greater the difference between the signal and background distributions in a particular kinematic variable, the greater the discriminating power of that variable is.

Rectangular cuts are a method of candidate selection which establishes an accepted range in each variable. This method requires that the variable values for a signal event falls within these ranges, rejecting events as background otherwise. All the accept/reject decisions described so far, in the trigger specifications and requirements imposed in offline reconstruction fall in this category. The difficulty in optimizing such a method lies in the rigidity of the requirement: if simply one variable falls outside the accepted range even by a small amount, the event will be rejected.

The use of an artificial neural network (ANN or NN) alleviates some of the associated difficulties. A neural network combines the information from all kinematic distributions into a single output variable that denotes whether an event is signal-like or background-like. The kinematic variables are assigned weights that determine their contribution to the NN output variable. These weights are in proportion to their discriminating power. In doing so, a neural network does not allow any single

variable to veto an accept decision. This makes it possible to simultaneously improve the efficiency of both background rejection and signal acceptance.

Prior to the application of a neural network selection, we begin with a set of pre-selection rectangular cuts. The guiding principle of the pre-selection cuts is to remove the most background-like events from the sample while keeping the cuts loose enough to avoid rejecting any signal events. The goal is to increase the purity while keeping the efficiency the same. An auxiliary motivation is the resultant gain in computational speed achieved by a smaller sample. The neural network preparation, in which we determine what constitutes signal-like and background-like events, is performed with the sample of decay candidates that pass the pre-selection cuts.

Most of the variables discussed in this section are already familiar from the preceding discussion: momenta, masses and vertex fit probabilities. In addition to these, we make use of a particle identification variable that relies on the energy loss measured in the COT drift chamber ($\mathrm{d}E/\mathrm{d}x$) and time-of-flight measured in the TOF detector to distinguish between pions and kaons. The $\mathrm{d}E/\mathrm{d}x$ and TOF information is used to calculate a single discriminating variable: the combined log likelihood (CLL). We also use a muon identification variable, the muon likelihood (l_μ) to distinguish between real and fake muons. The muon likelihood combines five observables to perform this discrimination: the distance in the r–ϕ plane, in ϕ, and in the longitudinal direction between the extrapolated COT track and the location of the muon stub, as well as the energy deposition in the electromagnetic and hadronic calorimeters. The details of the construction of these particle ID variables and their discriminating power are discussed in Appendix C.

3.3.1 Pre-Selection

We first apply loose rectangular pre-selection cuts before performing the neural network training that is used for final candidate selection. These cuts are summarized in Table 3.1. The main pre-selection cuts are motivated by the following:

- We limit the B mass window to ± 100 MeV/c^2 around the world average value of the B mass. No fully-reconstructed signal events are expected so far from the B mass peak. The mass sidebands within this window provide us a clean sample of background events whose angular distributions are expected to be representative of the angular distributions of background events that fall in the mass peak.

- We exclude the most pion-like kaon tracks using the combined log likelihood for particle ID.

- Certain regions of the event variable space are not well-modeled in Monte Carlo simulation. This is particularly true of tracks with low transverse momenta (p_T). We cut on these variables to ensure good data-Monte Carlo agreement.

Common variables	Value of cut	
$P(\chi^2(B))$	$> 10^{-50}$	
$P(\chi^2(J/\psi))$	$> 10^{-50}$	
$p_T(B)$ [GeV/c]	> 4.0	
$p_T(\mu)$ [GeV/c]	> 1.5	
$p_T(K)$ [GeV/c]	> 0.4	
$\lvert M_{J/\psi\phi} - M_B^{PDG}\rvert$ [MeV/c^2]	< 100	
$\lvert M_{\mu\mu} - M_{J/\psi}^{PDG}\rvert$ [MeV/c^2]	< 80	
$\sigma_{c\tau}$ [μm]	< 150	
Mode-specific	$B_s \to J/\psi\phi$	$B_d \to J/\psi K^*$
$P(\chi^2(\phi))$	$> 10^{-50}$	–
$P(\chi^2(K^*))$	–	$> 10^{-50}$
$p_T(\phi)$ [GeV/c]	> 1.0	–
$p_T(K^*)$ [GeV/c]	–	> 2.0
$\lvert M_{KK} - M_\phi^{PDG}\rvert$ [MeV/c^2]	< 14	–
$\lvert M_{K\pi} - M_{K^*}^{PDG}\rvert$ [MeV/c^2]	–	< 80
$CLL(K_1)$	> -5.0	> -50
$CLL(K_2)$	> -5.0	–
$CLL(\pi)$		> -50

Table 3.1: Pre-selection cuts for the $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$ samples. The cut on the particle ID variable $CLL(K)$ is kept loose for the B_d mode to allow for cases where the kaon and pion were mistakenly swapped in the mass assignment done in reconstruction.

3.3.2 Neural Network for $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$

General Principles

The goal of a candidate selection neural network is to make the best possible use of the discriminating power of the various kinematic variables at our disposal. We have already stated that the first improvement a neural network makes over a rectangular selection is that the kinematic variables are assigned different weights in proportion to their discriminating power, with no single variable given veto power on the accept decision. In addition, neural networks take into account the correlations among kinematic variables in assigning weights, making use of them to refine our understanding of what constitutes signal- and background-like events.

A neural network is comprised of an input layer, a hidden layer and an output layer. Each variable such as mass, p_T, etc, corresponds to an *input node*. Each input node is assigned a different weight in the network's *hidden layers* to produce an output variable $\mathcal{O}_{\rm NN}$ to be used as a discriminant. The output variable $\mathcal{O}_{\rm NN}$ ranges between between 0 (background-like) and 1 (signal-like) .

We use the neural network package `ROOTSNNS` [97], which is an interface for the

ROOT analysis framework to the Stuttgart Neural Network Simulator (SNNS) [98]. Figure 3-1 shows a conceptual sketch of a neural network, illustrating the process of combining inputs in a hidden or output layer. A node receives inputs $(x_0, \ldots, x_n)$ from the previous layer, and adds them in a weighted sum to generate an output $f(x_0, \ldots, x_n)$. A standard form for $f(x)$ is:

$$f(x_0, \ldots, x_n) = \frac{1}{1 + \exp\left[-\sum_{i=0}^{i=n} w_i x_i\right]}. \tag{3.1}$$

In general other monotonic functions that satisfy $\lim_{x \to -\infty} f(x) = 0$ and $\lim_{x \to \infty} f(x) = 1$ can be used. In the case that a step function is used, the neural network reduces to the special case of rectangular cuts.

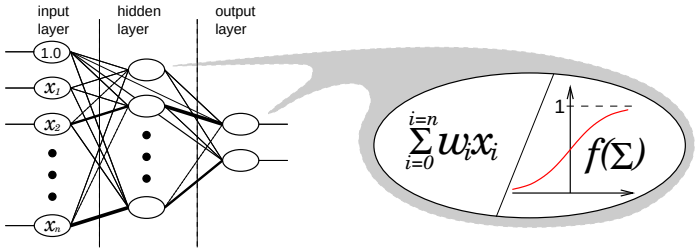


Figure 3-1: Conceptual sketch of a neural network with one input, one hidden and one output layer. The training of the neural network consists in assigning weights w_i to each node such that they can be combined to form an output (decision) node that discriminates between signal and background.

In order to assign weights to the input nodes, the neural network must have a pattern from which to establish quantitative definitions of what is *background-like* and *signal-like*. This is done by *training* the neural network with signal and background samples. For the signal training sample, we use Monte Carlo samples consisting solely of simulated signal events, while the background samples are comprised of the data events in the B meson mass sidebands. The use of a Monte Carlo sample as a signal training sample is the only way to avoid contamination from backgrounds that are inevitable in any data sample. Naturally, were the total removal of such contamination possible to begin with, this entire discussion would be moot.

The utility of using Monte Carlo to define the attributes of signal-like events relies on a simulation that closely models data in the space of variables which will be used as input nodes. The use of a Monte Carlo sample whose kinematic distributions do not match the signal distributions in the data leads to a sub-optimal selection. The discriminating power of the kinematic variables that are improperly modeled in Monte Carlo is partially wasted in such a case. Therefore, data – MC agreement is carefully checked for all variable used as input nodes to the NN.

Furthermore, the use of sideband data as a background training sample implicitly assumes that events under the mass peak behave similarly and have comparable topology as those in the sideband. Because the large majority of the background in

these decay modes is the result of the random combination of non-B_s (B_d) tracks, whose mass distribution is essentially flat, this assumption is justified.

The training process consists of an iterative optimization to minimize the distance between a target vector $\vec{t}$ of decisions the NN should make and an output vector $\vec{o}$ of decisions the NN does make at each stage of the optimization. This quantity to be minimized is expressed as:

$$\Delta = \left[\sum_{i=1}^{N} \frac{(t_i - o_i)^2}{N} \right]^{1/2}, \tag{3.2}$$

where N is the total number of background plus signal training events given to the neural network during the optimization procedure for learning purposes. In addition to the training set of events, we prepare two additional sets of events, or equivalently we divide the total set of potential signal and background events into three independent sets: the training set, the validation set, and the testing set. The training set is used to feed patterns to the neural network for learning purposes. The validation set is used to ensure that the reweighting procedure does not iterate too far by focusing on the particular statistical fluctuations of the training set. This potential *overtraining* is less than ideal because we seek to train the NN on the general features of any hypothetical set of signal events, not specifically the one we have chosen to use for training. Overtraining is avoided by comparing the value of Δ for the training and validation set at each iteration. If the training is proceeding as intended, both values should decrease with further iterations. If the training has gone too far, Δ_{train} will at some point continue to decrease while Δ_{valid} will suddenly level off and begin to increase, a sign that the training is now focusing on statistical fluctuations of the training set, and that iterations must stop. Finally, we wish to test the performance of the neural network algorithm we train. We do this using the testing set. We use the neural network to produce an output value $\mathcal{O}_{\text{NN}}$ for each background and signal event in the testing set, and calculate a *generalization error* Δ_{test} to see how well the neural network performs on another statistically independent set. We expect $\Delta_{\text{test}} \sim \Delta_{\text{train}}$, which indicates that the neural network performs equally well in a statistically independent set, and that the optimization is generally applicable.

Neural Network Input Variables

The variables used as inputs for the neural network are listed in Table 3.2. We use the χ^2 of the three vertex fits in three dimensions in each decay mode: B_s, J/ψ and ϕ/K^* for B_s and B_d, respectively. In addition, we use $\chi^2_{r\phi}$, the measure of goodness of the transverse-plane fit of the B_s/B_d vertex. We include the reconstructed masses of the vector mesons J/ψ, ϕ and K^* as an additional check against random combinations of tracks not originating from these decays. The transverse momenta of the three decay particles in each mode are included. The decay products of the ϕ and K^* contribute additional variables: the maximum and the minimum of the p_T of the two kaons in the first case, and the p_T of the kaon and pion in the second case. We use the combined log likelihood ratio (CLL) for K/π separation (see Appendix C, Sections C.1–C.3);

the maximum and minimum of the values for the two kaons in the B_s decay and the kaon and pion values in B_d decay. Finally, we use the maximum and the minimum of the two muon likelihood values $\mathcal{L}_\mu$ (see Appendix C, Section C.4).

Variables	$B_s \to J/\psi\phi$	$B_d \to J/\psi K^*$	#
Vertex Fit Probability	$\chi^2(B_s)$	$\chi^2(B_d)$	2
	$\chi^2(J/\psi)$	$\chi^2(J/\psi)$	5
	$\chi^2(\phi)$	$\chi^2(K^*)$	8
	$\chi^2_{r\phi}(B_s)$	$\chi^2_{r\phi}(B_d)$	1
Mass	$\lvert M_{\mu\mu} - M^{PDG}_{J/\psi}\rvert$	$\lvert M_{\mu\mu} - M^{PDG}_{J/\psi}\rvert$	4
	$\lvert M_{KK} - M^{PDG}_{\phi}\rvert$	$\lvert M_{K\pi} - M^{PDG}_{K^*}\rvert$	7
Transverse Momentum	$p_T(B_s)$	$p_T(B_d)$	0
	$p_T(J/\psi)$	$p_T(J/\psi)$	3
	$p_T(\phi)$	$p_T(K^*)$	6
	$p_T(K_1)$	$p_T(K)$	11
	$p_T(K_2)$	$p_T(\pi)$	12
Particle ID	$\max(\mathcal{L}_{\mu^+}, \mathcal{L}_{\mu^-})$	$\max(\mathcal{L}_{\mu^+}, \mathcal{L}_{\mu^-})$	9
	$\min(\mathcal{L}_{\mu^+}, \mathcal{L}_{\mu^-})$	$\min(\mathcal{L}_{\mu^+}, \mathcal{L}_{\mu^-})$	10
	$\max(CLL(K^+), CLL(K^-))$	$CLL(K)$	13
	$\min(CLL(K^+), CLL(K^-))$	$CLL(\pi)$	14

Table 3.2: Variables used to train the background-suppression neural network for $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$. The numbers in the last column explain the legend of Figure 3-2, plots (c) and (f).

Neural Network Training Validation

We perform several tests to verify that the NN training proceeded as intended, based on the caveats discussed above. These tests are sanity checks on the training process, not an optimization of the NN selection, which we discussed next. They consist of the following verifications:

- **Generalization error in the training and validation samples**. We compare the generalization error in the training and validation sets at each iteration of the training procedure. We stop training when we have achieved the smallest possible Δ_{train} that is still equal to or larger than $\Delta_{\text{validation}}$.

- **Output values $\mathcal{O}_{\text{NN}}$ for background and signal events in the test sample**. We show the neural network output values for all events in the test sample in Figure 3-2 for B_s (a) and for B_d (d). The distribution of output values for the signal and background events are found to be smooth and featureless as expected, with background tending toward 0 and signal toward 1.

- **Purity of the test sample (P) versus neural network output cut (C_{NN})**. By definition, plotting P versus C_{NN} must yield a monotonically increasing

distribution. If successively tightening the cut on $\mathcal{O}_{\rm NN}$ does not yield a larger and larger percentage of signal events in our sample, the NN training has not properly learned signal- and background-like patterns. In addition, when we plot the distribution in a sample with the same signal-to-background ratio as the training samples, we expect it to be linear, which is what we find. These distributions are shown in Figure 3-2 for B_s (b) and for B_d (e).

- **Correlations**. We check correlations among the training variables, in order to see that they match well-motivated physical expectations. A two-dimensional histogram of the correlation matrix (displayed as a color map) is shown in Figure 3-2 for B_s (c) and for B_d (f).

This procedure ensures that the neural network is performing as intended.

Neural Network Cut Optimization

Once the neural network training is complete, a decision has to be made about the minimum NN output value used for final candidate selection. We use as a criterion the statistical significance, defined as $S/\sqrt{S+B}$, and choose the NN output value cut which gives the data sample with the greatest statistical significance. In order to obtain a good estimate of the signal and background fractions in our sample after the cut, we look at the mass distribution in the sample, which is a good determinant of the background fraction. We use a simple Gaussian fit model for the signal and a $1^{\rm st}$ order polynomial for the background, and perform a binned χ^2 fit. We estimate the number of signal and background events from the fitted parameters of the corresponding distributions.

In order to avoid biases when making the neural network cut, we do not compute the significance in the signal region using the number of data events in the Gaussian peak, but rather the number of Monte Carlo events in the Gaussian peak, divided by a factor to reflect the larger size of the Monte Carlo sample. We integrate the fitted signal Gaussian in a window of two standard deviations, σ_m, around the PDG mass where $\sigma_m = 0.01$ GeV/c^2 for B_s and for B_d. A graph of the significance versus neural network output is shown in Figure 3-3.

We see that the significance does not change very much for a range of the neural network output cut in $\{0.2, 0.5\}$. This leaves us with two possibilities: we can choose a tighter cut to reduce both signal and background, or a looser cut to enhance both signal and background. Motivated by the fact that our background is well modeled in the fit likelihood, we choose a looser cut of NN output equal to 0.3 to increase the signal event numbers and improve sensitivity.

In Figure 3-4, we show the mass distribution of B_s candidates after pre-selection and after a cut at $\mathcal{O}_{\rm NN} > 0.30$. We obtain an expected signal yield of approximately 2000 signal events with 1.3 fb^{-1} of data, and approximately 2450 events with 1.7 fb^{-1}.

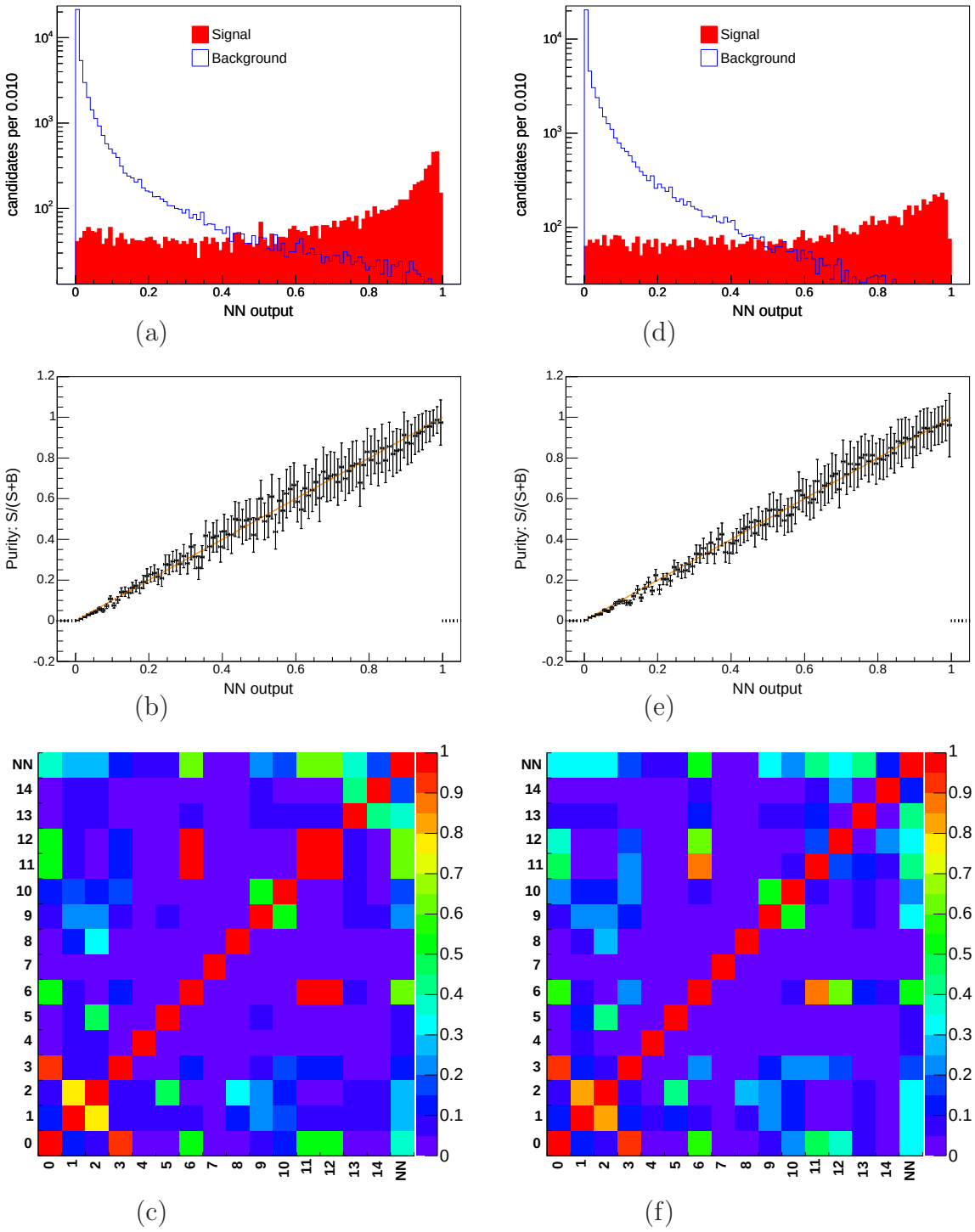


Figure 3-2: Neural network training validation plots: output values for background and signal, purity versus output, and correlations among neural network training variables (plots (a), (b), and (c) for B_s and (d), (e), and (f) for B_d, respectively). See Table 3.2 for the legend relating variable number and name in plots (c) and (f).

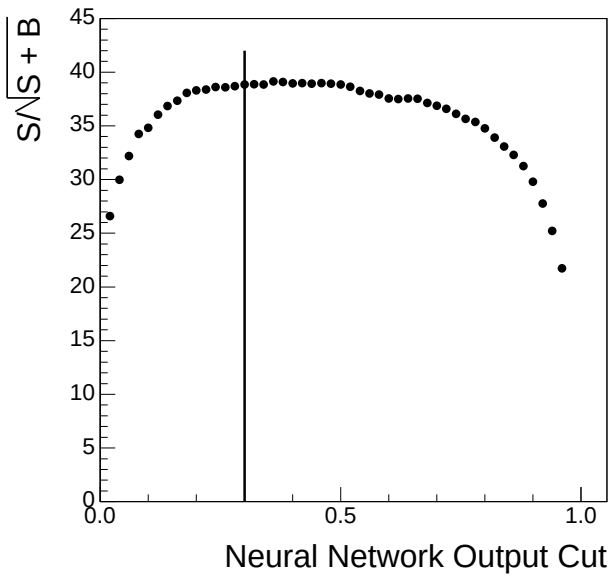


Figure 3-3: Significance plotted versus neural network output and mass distribution at chosen neural network cut of 0.3 for B_s.

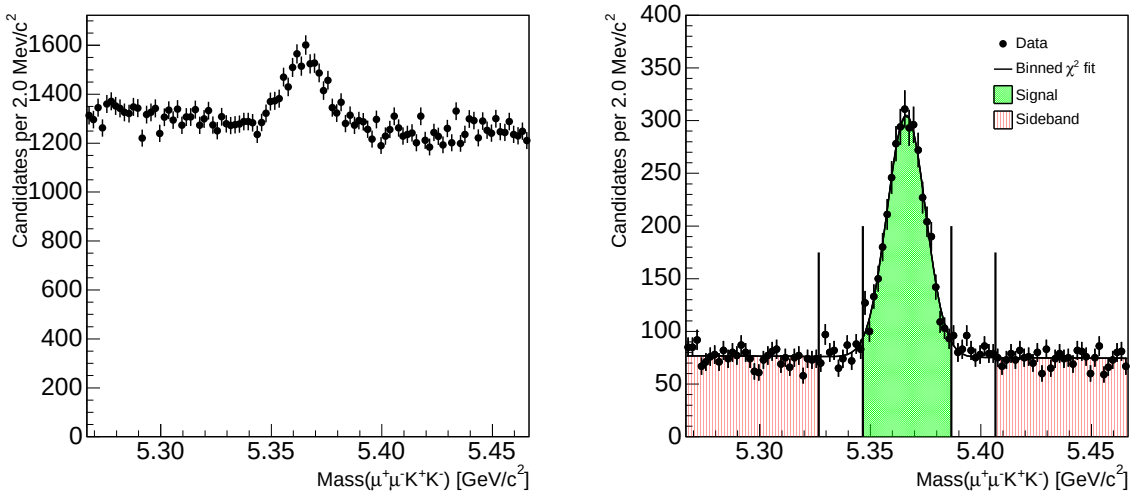


Figure 3-4: Mass spectrum of the B_s sample after pre-selection and final cuts.

3.3.3 Swap Suppression Neural Network for $B_d \to J/\psi K^*$

In addition to suppressing background, we seek to eliminate from our sample events in which the K and π are misidentified and swapped but reconstructed with a mass which allows them to fit inside the K^* mass window. For this task, we train a second artificial neural network. The signal training sample consists of correctly reconstructed $B_d \to J/\psi K^*$ events, and the background training sample consists of $B_d \to J/\psi K^*$ events for which the K and π were misidentified and swapped. In this case, the NN training is done using simulated data for both signal (correct mass assignments) and background (swapped mass assignments), because this is the only way to ascertain whether a swap occurs in reconstruction. We use the following variables to train the artificial neural network for the swap suppression selection:

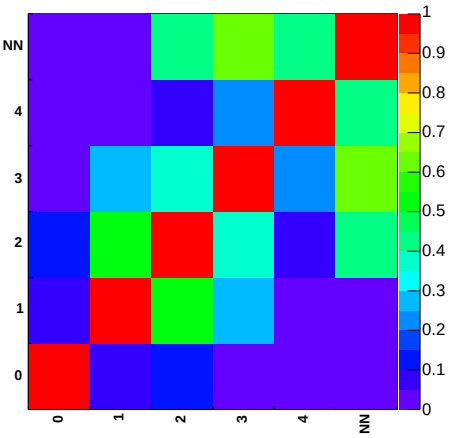
- Reconstructed mass of the vector particle: K^*
- Transverse momentum p_T: K, π
- Combined log likelihood CLL for the K and π

These variables are shown in Table 3.3, with the correlation to the NN output. These correlations indicate how important each variable is to determining whether or not a candidate is swapped. In Figure 3-5, we show the correlation of each variable to the neural network output value and to one another.

Variable	Correlation to $\mathcal{O}_{\rm NN}$
$CLL(K)$	0.62698
$p_T(\pi)$	0.43178
$CLL(\pi)$	0.43000
$M_{K\pi}$	0.03231
$p_T(K)$	0.00451

Table 3.3: Variables used to train the swap-suppression neural network, with correlation to the neural network output value $\mathcal{O}_{\rm NN}$.

We perform the same checks with the swap-suppression NN as we do with the background-suppression NN. The results of these checks are shown in Fig. 3.3.3. However, the same optimization method is not suitable here. Using simply a fit in the mass subspace to perform an optimization of the variable $S/\sqrt{S+B}$ pushes the swap NN cut to zero. This is because, by definition, swapped B_d candidates are assigned a B mass that is in the signal region and is considered by this method to contribute to signal significance. This means that the maximum $S/\sqrt{S+B}$ calculated by a fit in the mass subspace is achieved without the swap neural network. As a compromise, we perform the optimization of $S/\sqrt{S+B}$ using the background suppression NN alone, and then seek to find the swap NN cut that brings the swap fraction below 1%. This makes the swap fraction consistent with zero, as the uncertainty on the swap fraction is $\sim$ 1%. The final cuts are: 0.35 for the background suppression neural network output, and 0.30 for the swap suppression neural network output. The mass spectra after pre-selection and after the two neural network cuts for the B_d decay mode are



0: $M_{K\pi}$
1: $p_T(K)$
2: $p_T(\pi)$
3: $CLL(K)$
4: $CLL(\pi)$

Figure 3-5: Correlations among input variables and the neural network output for the swap suppression neural network. A legend matching variable name to variable number in the correlation table is provided on the right of the figure.

shown in Figure 3.3.3. We obtain an expected signal yield of approximately 7600 signal events with 1.3 fb^{-1} of data.

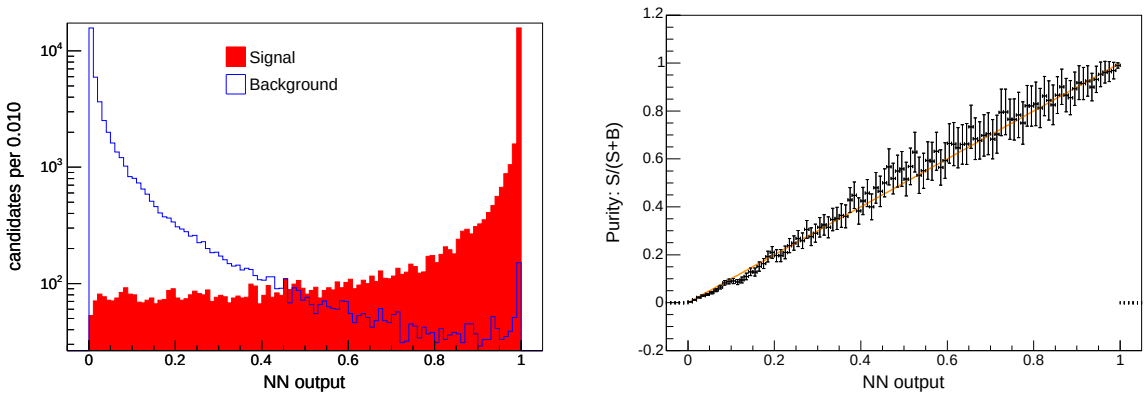


Figure 3-6: Output plots for the swap-suppression neural network.

3.3.4 Selection for $B_u \rightarrow J/\psi K$

The $B_d \rightarrow J/\psi K^*$ and $B_u \rightarrow J/\psi K^+$ decay modes are used for the calibration of the opposite tagger, as described in Chapter 5. For this task, we use the same neural network cuts for the B_d mode as described above, and a set of rectangular cuts for the B_u mode similar to the ones described in Reference [99].

Table 3.4 summarizes the rectangular cuts used. Figure 3-8 shows the mass spectrum of selected B_u candidates.

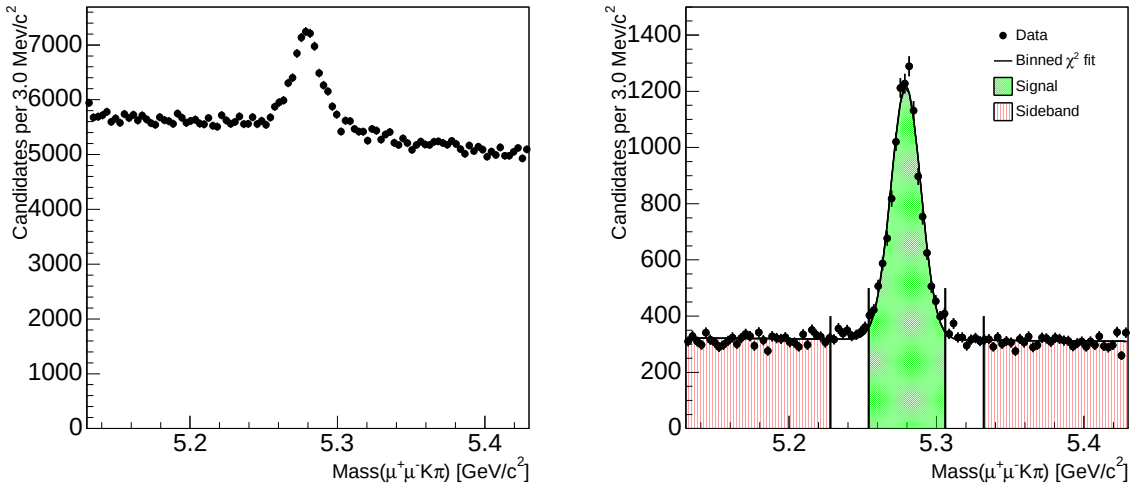


Figure 3-7: Pre-selection and final selection mass plots for B_d.

Cut Variable	Value of cut
$P(\chi^2(B))$	$> 10^{-4}$
$\lvert M_{J/\psi K^+} - M_B^{PDG} \rvert$ [MeV/c^2]	< 100
$p_T(B)$ [GeV/c]	> 6.0
$p_T(K)$ [GeV/c]	> 1.8
$p_T(\mu)$ [GeV/c]	> 1.5
$\lvert M_{\mu\mu} - M_{J/\psi}^{PDG} \rvert$ [MeV/c^2]	< 80
$\sigma_{c\tau}$ [μm]	< 150
$CLL(K)$	> -5.0

Table 3.4: Selection cuts on the B_u sample.

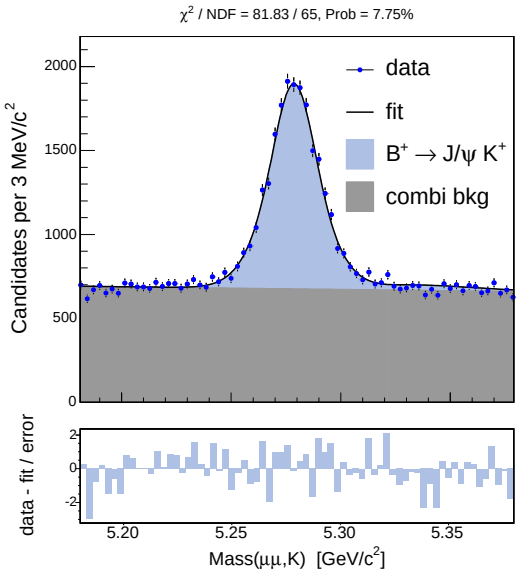


Figure 3-8: Mass spectrum of B_u candidates selected with the rectangular cuts detailed in Table 3.4.

Chapter 4

Lifetime, Width Difference and Angular Amplitudes

We present in this chapter the time-dependent angular analyses of $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$, yielding measurements of average lifetimes and angular amplitudes for both B meson decay modes, and the width difference of the two mass eigenstates of the B_s meson. In presenting these measurements, we cover much of the material necessary for the measurement of CP violation in $B_s \to J/\psi\phi$ (Chapter 6): general fitting technique, likelihood construction, development of the angular analysis, and the fit models for the background components in each sample.

Throughout this chapter, we assume that $2\beta_s = 0$. In so doing, we insulate ourselves from the complications that arise from letting $2\beta_s$ be determined by our likelihood fit, and focus on establishing a base framework and resolving all other analysis issues. We treat the case of letting $2\beta_s$ be determined by the likelihood fit in Chapter 6. Naturally, the measurements presented in this chapter constitute experimental results on their own, and are either competitive with or considerable improvements on previous results.

The B_s and B_d modes are treated in parallel to avoid superfluous repetition of analysis components they have in common. After a brief general introduction to the parameter estimation method, a large portion of this chapter is devoted to presenting and justifying the models used to describe signal and background components in each of the variable subspaces (mass, proper decay length, transversity angles). This is followed by a presentation of fitter cross-checks, nominal fit results, and estimation of systematic uncertainties. We end by presenting the results of each measurement with both statistical and systematic uncertainties.

4.1 Fitting Method

We use the maximum likelihood method (ML) to determine the unknown theoretical parameters θ. With a sample of N decay candidates, each is described by a vector of event variables $\vec{x}_i$. We denote by $p(\vec{x}\,|\,\theta)$ the probability density function (PDF) that describes the expected distribution of events in the space of event variables $\vec{x}$, given

the vector of parameters θ. We construct the likelihood function as

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} p(\vec{x}_i, \theta) \tag{4.1}$$

The probability of observing the data events in the order in which they were observed is given by $\mathcal{L}(\theta)$. The expression is divided by an additional factor of $n!$ to express the probability of observing this set of data events regardless of their order. However, since the likelihood expression is to be used to estimate the parameters θ by maximization with respect to θ, this constant factor does not affect the result, and is omitted here. Note also that while $p(\vec{x}_i \,|\, \theta)$ denotes the probability of observing the data points $\{\vec{x}\}$ given the theoretical parameters θ, $\mathcal{L}(\theta')$ is the likelihood of the theory parameters having been equal to θ', given the observed data $\{\vec{x}_i\}$.

In practice we perform a minimization of $-\log \mathcal{L}(\theta)$ numerically:

$$\log \mathcal{L} = \sum_{i=1}^{N} \log p(\vec{x}_i, \theta). \tag{4.2}$$

This is done in order to avoid problems with numerical precision that arise when multiplying many small numbers together. The numerical minimization is carried out using the `MINUIT` package [100] and the `ROOT` analysis framework [101].

4.1.1 Fit Variables

Given a B_s (B_d) candidate i for which we do not know the production b flavor, we have at our disposal seven (ten) variables to construct the PDFs used in the likelihood fit. A likelihood may be constructed that is a function of all of these, or we may limit ourselves to certain subspaces. We list the variables by the subspaces they define:

- m_i, σ_{m_i}: mass of the candidate and its uncertainty. These quantities are calculated for each decay vertex by the CTVMFT algorithm described in Section 2.2.1 using the parameters (with uncertainties and correlations) of the decay daughter tracks.

- ct_i, σ_{ct_i}: proper decay length (PDL) and its uncertainty. The PDL and its uncertainty are given by:

$$ct = \frac{\vec{L_{xy}} \cdot \vec{p_T}}{|\vec{p_T}|^2} m^{\mathrm{PDG}}, \tag{4.3}$$

$$\sigma_{ct} = \left(\frac{\sigma_{L_{xy}}}{p_T} m^{\mathrm{PDG}} \right) \oplus \left(ct \frac{\sigma_{p_T}}{p_T} \right), \tag{4.4}$$

  where $\oplus$ denotes the usual rule for addition of uncorrelated uncertainties: $a \oplus b \equiv \sqrt{a^2 + b^2}$. The candidate p_T is calculated from the vector sum in the $x - y$ plane of the p_T of the daughter tracks. The momentum uncertainty arises from the uncertainty on the position of individual track hits and is calculated

in the track fitting algorithm. At CDF, the momentum uncertainty is equal to $\sigma_{p_T}/p_T \approx 0.0015\ p_T/(\mathrm{GeV}/c)$, and contributes negligibly to σ_{ct} for most analyses. We calculate $\vec{L_{xy}}$ and $\sigma_{L_{xy}}$ using:

  - The coordinates of the primary vertex in the $x-y$ plane. We use the event-by-event method of determining the position of the primary vertex, as described in Section 3.2.4.
  - The coordinates of the secondary vertex in the $x-y$ plane, and the associated error matrix, both given by CTVMFT.

- $\vec{\omega} \equiv (\cos\theta, \phi, \cos\psi)$: angular variables characterizing the decay in the transversity basis, calculated as described in Section 1.4. In the case of B_d, we also calculate the quantities $\vec{\omega'} \equiv (\cos\theta', \phi', \cos\psi')$. The primed variables are calculating by swapping the K and π four-momenta in the procedure used to derive $\vec{\omega}$. This is used to construct a transversity PDF for the case when the K and π are misidentified as one another while retaining an mass for the parent K^* decay vertex that passes all selection criteria. The details of this construction are given below in Section 4.1.5.

In addition to the event variables outlined above, Chapter 5 introduces variables used to determine the production flavor (B or $\overline{B}$) of B meson candidates using several flavor-tagging algorithms. Each one of these algorithms provides a tag decision for each candidate as well as a variable quantifying confidence in the tag decision. The treatment of these additional event variables is discussed in Section 5.4 and 6.2.

4.1.2 PDF Construction

In order to construct a likelihood, we need to define the probability density function that defines the expected distribution of events in our sample in the space of variables described in Section 4.1.1. For one variable, the PDF $f_X(x)$ is defined such that the probability of observing $x \in [a, b]$ is given by

$$P(x \in [a,b]) = \int_a^b dx\ f_X(x). \tag{4.5}$$

We build the joint PDF $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ as a function of all event variables. It is defined such that the probability of observing variables within the n-dimensional volume V is given by

$$P(x_1, \ldots, x_n \in V) = \int_V dx_1 \ldots dx_n\ f_{X_1,\ldots,X_n}(x_1, \ldots, x_n). \tag{4.6}$$

We say that the joint PDF is separable if we can write it as a product of individual PDFs in the variable subspaces, i.e.

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n). \tag{4.7}$$

In this analysis, we encounter joint PDFs of the form

$$f_{Y,Z}(y,z) = g_Y(y) \cdot g_Z(z) + h_Y(y) \cdot h_Z(z), \tag{4.8}$$

a clear example of a non-separable joint PDF. By definition, a PDF must be unit-normalizable and positive definite in order to give meaningful probabilities. We note that the individual distributions g, h may not be positive-definite, as long as the sum $f_{Y,Z}(y,z)$ is.

We discuss the construction of the joint PDF and its constituent parts from the physics distributions in the variable subspaces and our knowledge of detector effects. Because such a discussion is more convenient if we consider each variable subspace by itself first, we use the following strategy in presenting the material in this chapter:

- **Mass** (Section 4.1.3): the mass distribution is separable from lifetime and angular portions of the joint PDF in the case of B_s. In the case of B_d, we construct a distribution for candidates with the correct $K-\pi$ mass assignment, and another for the case where there is a $K \leftrightarrow \pi$ swap of the mass assignment. We treat the B_s case and note the differences for B_d.

- **Proper Decay Length** (Section 4.1.4): the proper decay length PDF is separable in the case of $B_d \to J/\psi K^*$, but not in a time-dependent analysis of $B_s \to J/\psi\phi$. The PDF presented in that section is the one to use in a measurement of *average* lifetime, with no knowledge of angular variables. Alternatively, it describes the distribution (not probability density function) to be used in a PDL+Transversity PDF for the B_s time-dependent analysis.

- **Transversity** (Section 4.1.5): Like in the case of the PDL, the transversity distribution is separable into its own PDF for $B_d \to J/\psi K^*$ but not $B_s \to J/\psi\phi$. For the latter decay mode, this section presents the *time-integrated* distribution in the transversity space, with an appropriate redefinition of the angular amplitudes.

We treat the background distributions for all three subspaces as fully separable. For the signal, from the building blocks presented in Sections 4.1.3 – 4.1.5, we then build the full joint probability density function in Section 4.2.

4.1.3 PDF for Mass

Although the mass subspace is not used for separating *CP*-odd and *CP*-even decays of the B_s meson, the inclusion of a mass model in the likelihood expression is useful in this analysis primarily for its power of discrimination between signal and background events. The mass PDF factorizes out of the joint signal PDF as well as the joint background PDF for both B_s and B_d, for tagged and untagged samples.

Signal

The signal mass peak is approximately Gaussian, with a width that is dominated by detector resolution effects. We find that a better model of the detector resolution effects is achieved with a mass model described by two Gaussians with identical mean and different widths σ_1 and σ_2. These widths are determined in the maximum likelihood fit from the data sample, and vary independently. This description more closely matches the observed data than a model with a single Gaussian peak. We construct the PDF by normalizing the Gaussians over the mass window dictated by our preselection requirements. The normalized function is defined as:

$$\mathcal{G}(m \,|\, M, \sigma, M_{\min}, M_{\max}) = \frac{\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{m-M}{\sigma}\right)^2\right]}{\frac{1}{2}\left[\mathrm{Erf}\left(\frac{M_{\max}-M}{\sqrt{2}\sigma}\right) + \mathrm{Erf}\left(\frac{M-M_{\min}}{\sqrt{2}\sigma}\right)\right]}. \tag{4.9}$$

where Erf(z) is the error function, and $M_{\min}$ and $M_{\max}$ define the preselection mass window. While the minimum and maximum of the mass window must be used for proper normalization, we do not write $M_{\min}$ and $M_{\max}$ explicitly as parameters in the likelihood function because they are fixed. Therefore they are not subject to the minimization procedure used to find the best values of the other parameters. Using the definitions above, the signal mass PDF is then given by:

$$\begin{aligned} X_{\mathrm{sig}}(m_j \,|\, M, \sigma_1, \sigma_2, f_m) = (1-f_m) \;& \mathcal{G}(m_j \,|\, M, \sigma_1) + \\ f_m \;& \mathcal{G}(m_j \,|\, M, \sigma_2), \end{aligned} \tag{4.10}$$

where m_j is the mass of the B candidate, and f_m is defined as the fraction of the Gaussian with the larger width. We use Equation 4.10 to describe the signal mass distribution in the $B_s \to J/\psi\phi$ decay mode.

K^* Swap

The mis-assignment of masses to the daughter K and π in the $B_d \to J/\psi K^* \to \mu\mu K\pi$ decay mode (a.k.a. K^* swap) in principle leads to an additional component to the mass PDF for the B_d. The treatment of this component in Reference [42] consists of replacing the one-Gaussian mass model with a two-Gaussian model. We also use two Gaussians in the B_d mode; however, in contrast to Equation 4.10, where both Gaussians are part of the signal portion of the likelihood, one is used for the signal model, and one is used for the swap portion of the likelihood.

Background

We determine the shape of the background distribution in the mass subspace empirically. From examination of Figure 3-3, we conclude that the background is well described by a first-degree polynomial in the mass sector:

$$X_{\mathrm{bkg}}(m_j \,|\, A) = A \cdot m_j + \frac{1}{M_{\max} - M_{\min}}[1 - \frac{A}{2}(M_{max}^2 - M_{min}^2)] \tag{4.11}$$

where A is the slope of the polynomial, and $M_{\rm min}$, $M_{\rm max}$ are the boundaries of the mass window within which the function is normalized.

4.1.4 PDF for Proper Decay Length

The description of the distributions in PDL of the signal portion of our data sample must take into account the exponential decay of the B mesons as well as the detector resolution in proper decay length. The background distributions are physically motivated empirical descriptions, which also include resolution effects.

Signal

The lifetime of the signal component is described by an exponential decay convolved with a detector resolution function. We use a single Gaussian as a resolution function, whose width is given for each event j by the uncertainty on the proper decay length σ_{ct_j}. Here, as in the description of the mass distribution, the uncertainty σ_{ct_j} may be incorrectly estimated, and we use a global scale factor S_{ct} to correct for this effect on average. The distribution of the signal in proper decay length is given by:

$$Y_{\rm sig}(ct_j, \sigma_{ct_j} \,|\, c\tau, S_{ct}) = E(ct_j \,|\, c\tau) \otimes G(ct_j, \sigma_{ct_j} \,|\, S_{ct}), \tag{4.12}$$

where E and G are defined as follows:

$$E(ct \,|\, c\tau) = \begin{cases} 0, & ct < 0 \\ \dfrac{1}{c\tau} e^{-\frac{ct}{c\tau}}, & ct \geq 0, \end{cases} \tag{4.13}$$

and

$$G(ct, \sigma \,|\, S) = \frac{1}{\sqrt{2\pi} S \sigma} e^{-\frac{c^2 t^2}{2 S^2 \sigma^2}}. \tag{4.14}$$

The convolution of two functions denoted by $f(t) \otimes G(t, \sigma)$ is defined as

$$f(t) \otimes G(t, \sigma) \equiv \int_{-\infty}^{\infty} dt' \, f(t') G(t - t'; \sigma), \tag{4.15}$$

and is carried out by analytic integration.

Background

The background distribution in PDL is dominated by a large peak around $ct = 0$, from prompt J/ψ production coming directly from the $p\overline{p}$ interactions, and not a B decay. The muons from the J/ψ are mistakenly associated with two other tracks in the event reconstruction to give a fake $B_s \to J/\psi\phi$ decay. The prompt J/ψ distribution is modeled by a delta function $\delta(ct)$ convolved with a Gaussian for detector resolution as in the signal case. We use the same resolution function to smear the other components of the background PDL model. The remaining portions of the background are modeled by one short-lived exponential with positive lifetime, one short-lived exponential

with negative lifetime, and one long-lived exponential with positive lifetime. These are motivated in detail in Reference [42]. The long-lived background exponential is physically motivated for example by contributions from sequential semileptonic decays $b \to c \to s$ yielding two muons which fake a J/ψ decay. In the case that a truly displaced J/ψ is wrongly paired with a random track, this will also contribute to a long-lived background component. The short-lived positive and negative exponentials receive contributions from tracks that were reconstructed with an erroneous hit and from tracks that belong to a different displaced vertex, possibly the "other" b hadron from the $\bar{b}b$ pair. The background lifetime parametrization is thus be written as

$$\begin{aligned} Y_{\rm bkg}(ct_j, \sigma_{ct_j} \,|\, f_-, f_+, f_{++}, \lambda_-, \lambda_+, \lambda_{++}, S_{ct}) = \\ (1 - f_- - f_+ - f_{++})\, G(ct, \sigma_{ct_j} \,|\, S_{ct}) + \\ f_- E(-ct_j \,|\, \lambda_-) \otimes G(ct, \sigma_{ct_j} \,|\, S_{ct}) + \\ f_+ E(ct_j \,|\, \lambda_+) \otimes G(ct, \sigma_{ct_j} \,|\, S_{ct}) + \\ f_{++} E(ct_j \,|\, \lambda_{++}) \otimes G(ct, \sigma_{ct_j} \,|\, S_{ct}), \end{aligned} \tag{4.16}$$

where λ_-, f_- are the decay constant and fraction of the negative-lifetime tail, λ_+, f_+ are the decay constant and fraction of the short-live positive-lifetime tail, and λ_{++}, f_{++} are the decay constant and fraction of the long-lived positive-lifetime tail.

PDL Uncertainty Distribution

A joint probability density function that includes a measurement of proper decay length (PDL) and its uncertainty must include a term for the probability density of PDL uncertainties. Signal and background components have different intrinsic probabilities of having a given uncertainty σ_{ct} that arises from the properties of different properties of the secondary vertices of each class of events. Candidates arising from real B decays tend to have better-fitted tracks and vertices than those arising from combinatorial background, which is the dominant contribution to the background for the B_s and B_d decays treated here. Candidates from combinatorial background tend to have poorer ct resolution because they are is composed of random tracks that are incorrectly associated to form 4-track vertices.

In order to account for this correctly, the likelihoods for signal and for background components must be separately multiplied by the terms $P_{\rm sig}(\sigma_{ct})$ and $P_{\rm bkg}(\sigma_{ct})$ respectively. In this analysis, these probabilities are implemented as distributions of σ_{ct} for each class of event taken from the data. For the background distributions, we use a histogram of σ_{ct} taken from events in the upper and lower mass sidebands. For the signal distributions, we use use a histogram of σ_{ct} from the central mass region that is sideband-subtracted.

4.1.5 PDF for Transversity

In the case of the $B_d \to J/\psi K^*$ decay, the decay rate is separable into a product of proper decay length PDF and transversity PDF. In the case of $B_s \to J/\psi\phi$, we cannot speak of a PDF for transversity since the joint PDF is not separable in the

$ct - \vec{\omega}$ space. We treat the case of $B_d \to J/\psi K^*$, and of the background angular distribution, and discuss the necessary angular acceptance function.

Signal

In the case of $B_d \to J/\psi K^*$, the PDF for transversity is written as:

$$\begin{aligned} Z(\vec{\omega} \,|\, \{A_\alpha\}) = & \frac{1}{|A_0|^2 + |A_\parallel|^2 + |A_\perp|^2} \Big[|A_0|^2 f_1(\vec{\omega}) + |A_\parallel|^2 f_2(\vec{\omega}) + |A_\perp|^2 f_3(\vec{\omega}) \\ & \pm \mathcal{I}m(A_\parallel^* A_\perp) f_4(\vec{\omega}) + \mathcal{R}e(A_0^* A_\parallel) f_5(\vec{\omega}) \pm \mathcal{I}m(A_0^* A_\perp) f_6(\vec{\omega}) \Big], \end{aligned} \tag{4.17}$$

where the upper (lower) sign refers to the decay to $J/\psi[\to \mu^+\mu^-]K^*[\to K^+\pi^-]$ ($J/\psi[\to \mu^+\mu^-]\overline{K^*}[\to K^-\pi^+]$).

This PDF must be modified to account for the fact that the transversity acceptance is not uniform. The CDF detector itself, the di-muon trigger, the preselection requirements and the neural network selection algorithms all combine to sculpt the acceptance in the transversity subspace. This sculpting effect is modeled in Monte Carlo in order to construct an overall three-dimensional efficiency function $\epsilon(\vec{\omega})$. We generate Monte Carlo with a flat (i.e. uniform) distribution in $\vec{\omega}$. We then pass the Monte Carlo events through a full simulation of the CDF detector and trigger, and apply our preselection and selection criteria on the MC sample. The efficiency is defined as

$$\epsilon(\vec{\omega}) \equiv \frac{n_{\rm acc}(\vec{\omega}) d\vec{\omega}}{n_{\rm gen}(\vec{\omega}) d\vec{\omega}}, \tag{4.18}$$

where $n_{\rm gen}$ is the event density generated in a small phase space volume $d\vec{\omega}$ centered at $\vec{\omega}$, and $n_{\rm acc}$ is the event density in the same volume after the detector simulation, trigger requirements, and selection requirements have been applied in the Monte Carlo sample. In practice, we construct $\epsilon(\vec{\omega})$ as a three dimensional histogram, and the efficiency is calculated as a ratio of generated and accepted events in each three-dimensional bin in $\vec{\omega}$ divided by the total number of generated events such that the sum of the weights in all the bins in the histogram is 1. Two Monte Carlo samples are used, one for $B_s \to J/\psi\phi$ and one for $B_d \to J/\psi K^*$, to define an efficiency function for the transversity distribution in each case.

While the PDF Z is by definition unit-normalized, and $\epsilon(\vec{\omega})$ is normalized by construction, their product is not normalized; i.e. the integral

$$\int_{\vec{\omega}} Z_{\rm sig} \cdot \epsilon(\vec{\omega}) \, d\vec{\omega} \tag{4.19}$$

is not equal to 1. We calculate a normalization factor N_Z using analytical integration of $Z \cdot \epsilon$ over each bin of the histogram. The factor N_Z is a function of the physical parameters, whose fitted values are adjusted at each step of the minimization. To avoid recalculating these normalization integrals at each step of the minimization, we perform some algebraic simplification of the normalization task, the details of which are given in Appendix D.

In principle, the efficiency in transversity is also a function of kinematic variables as well as detailed trigger criteria. Therefore, we reweight the MC events used for the construction of the efficiency histogram so as to provide as close a match as possible to the kinematics and trigger composition of the data sample. We proceed as follows: we take the most discrepant distributions, i.e. the ones with the largest difference between data and Monte Carlo as measured by a χ^2 test, and apply a reweighting of the events used in the sculpting histogram iteratively. Figure 4-1 shows the transversity distributions of Monte Carlo candidates passed through our nominal B_s selection, projected onto the three axes of the transversity basis.

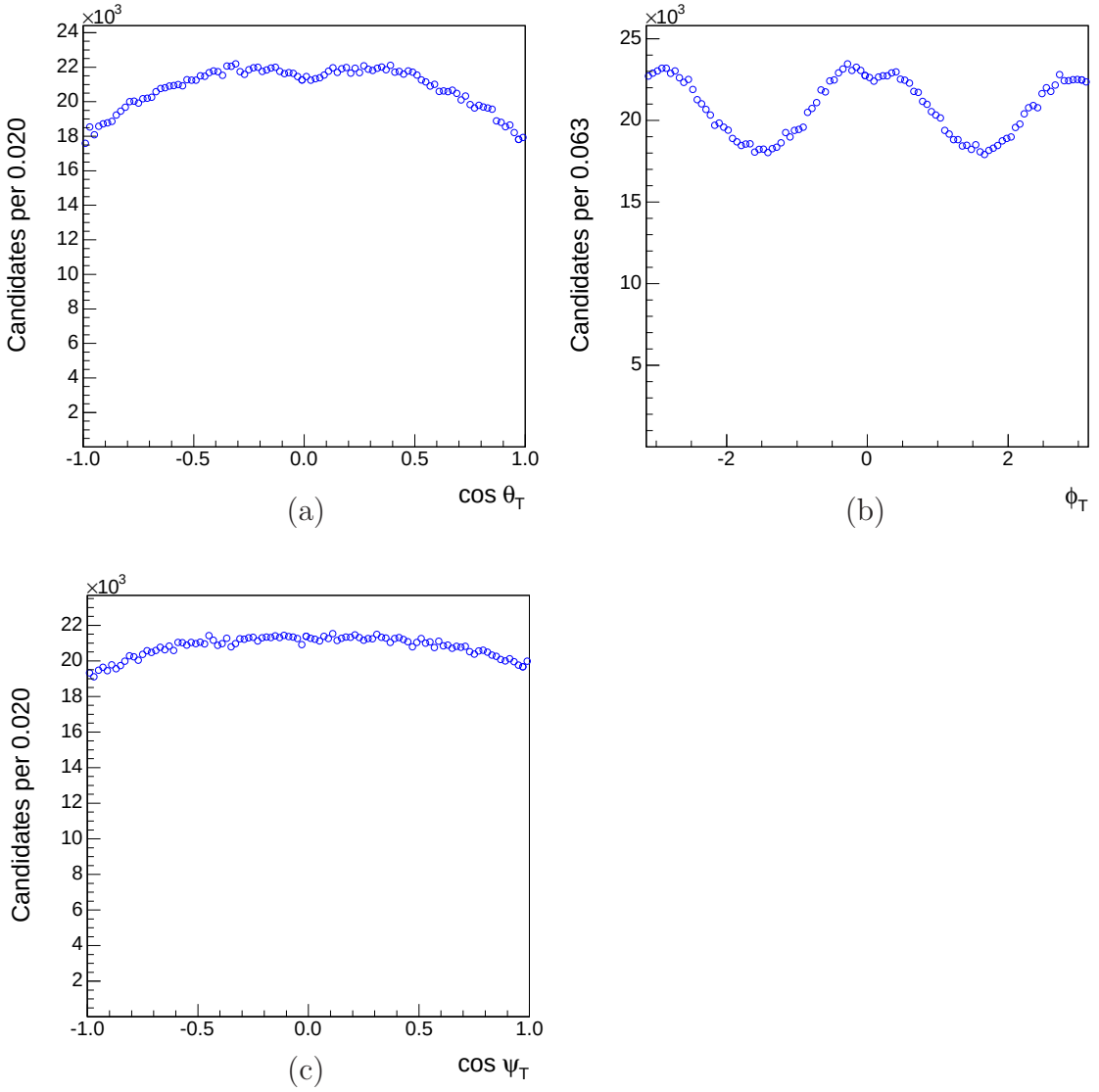


Figure 4-1: Sculpting of the transversity angles modeled in Monte Carlo simulation for the decay mode $B_s \to J/\psi\phi$: (a) $\cos\theta_T$, (b) ϕ_T, and (c) $\cos\psi_T$.

S-wave contribution from $B_d \to J/\psi K\pi$

The angular distribution of the $B_d \to J/\psi K^*$ decay $K^* \to K\pi$ is described entirely in terms of P-wave $K^* \to K\pi$ decays, in reference to the $L = 1$ initial orbital momentum of the K^*. However, in the mass range of the K^* allowed by our selection criteria, there is a sizable $K\pi$ S-wave contribution (initial $L = 0$) from the decay $B_d \to J/\psi K\pi$ [102]. We take this into account by introducing an amplitude A_S, with magnitude $|A_S|$ and phase δ_S, for the S-wave component. The introduction of the S-wave component into the decay model results in additional terms in the transversity PDF, including an interference term, as shown below. To describe the relative strength of the P- and S-wave contributions, we introduce in addition the real and positive parameter A_P, whose square A_P^2 denotes the overall strength of the P-wave portion of the decay. For easier comparison of results, we adopt the convention of Reference [102], and parametrize A_P and A_S as follows:

$$\cos\lambda \equiv \frac{A_P}{\sqrt{A_P^2 + |A_S|^2}}, \qquad \sin\lambda \equiv \frac{A_S}{\sqrt{A_P^2 + |A_S|^2}}. \tag{4.20}$$

To describe the additional dependence on $\vec{\omega}$ of the combined $P + S$ wave decay, we introduce the functions $f_7(\vec{\omega})$–$f_{10}(\vec{\omega})$:

$$\begin{aligned} f_7 =& \frac{3}{32\pi} 2(1 - \sin^2\theta_T \cos^2\phi_T), \qquad f_8 = \frac{-3}{32\pi}\sqrt{6}\sin\psi_T \sin^2\theta_T \sin 2\phi_T, \\ f_9 =& \frac{3}{32\pi}\sqrt{6}\sin\psi_T \sin 2\theta_T \cos\phi_T, \quad f_{10} = \frac{3}{32\pi} 4\sqrt{3}\cos\psi_T (1 - \sin^2\theta_T \cos^2\phi_T). \end{aligned} \tag{4.21}$$

The transversity PDF is modified as follows:

$$\begin{aligned} Z_{P+S} = & \cos^2\lambda \cdot Z(\vec{\omega}) + \sin^2\lambda \cdot f_7 + \\ & \frac{1}{2}\sin 2\lambda \cdot \left[f_8 \cos(\delta_\parallel - \delta_S)|A_\parallel| + f_9 \sin(\delta_\perp - \delta_S)|A_\perp| + f_{10}\cos(\delta_S)|A_0| \right], \end{aligned} \tag{4.22}$$

where the S-wave interference results in the $\sin 2\lambda$ term. We refer to the transversity Z_{P+S} as Z_{sig} in the discussion that follows, to differentiate it from Z_{swp}, the PDF for $K \leftrightarrow \pi$ swapped B_d candidates.

K^* Swap

In the $B_d \to J/\psi K^*$ decay, we need to treat the case when both the $K - \pi$ and $\pi - K$ mass assignments result in masses that fall into the specific K^* mass window with special care when we model the transversity distributions. Unfortunately we cannot treat the $K \leftrightarrow \pi$ swapped events as pure background because the likelihood to observe a swapped $K - \pi$ fully depends on the decay amplitudes $\{A_\alpha\}$. In the event that a $K \leftrightarrow \pi$ swap occurs in reconstruction, the angular distributions are not functions of the measured angles $\vec{\omega}$, but its real angles $\vec{\omega}'$, which can be computed by reassigning the mass. It is important to note that in the signal angular component

of the likelihood the sign in front of f_4 and f_6 depends on the flavour of the B_d assigned by looking at the sign of the kaon (see Equation 4.17). It is evident that this misidentification is reflected in the angular distributions and so in order to take this effect into account we split the angular likelihood in two terms: one for the correctly reconstructed candidates and one for the swapped ones. The angular PDF is rewritten as

$$Z_{\text{sig+swp}}(\vec{\omega}, \vec{\omega'} \,|\, \{A_\alpha\}) = (1 - f_{\text{swp}}) Z_{\text{sig}}(\vec{\omega}) \epsilon_{\text{sig}}(\vec{\omega}) + f_{\text{swp}} Z_{\text{swp}}(\vec{\omega'}) \epsilon_{\text{swp}}(\vec{\omega}), \qquad (4.23)$$

where f_{swp} is the fraction of candidates with the swapped mass assignment, and a new efficiency function ϵ_{swp} is computed from the Monte Carlo simulation using the swapped events (see Figure 4-2). The term $Z_{\text{swp}}(\vec{\omega'})$ has the signs of f_4 and f_6 flipped compared to $Z_{\text{sig}}(\vec{\omega})$.

It is observed in Figure 4-3 there is a strong but not complete correlation between the angles $\vec{\omega}$ of the real $B_d \to J/\psi K^*$ decay candidate and the angles $\vec{\omega'}$ of the candidate in which the daughters of the K^* are swapped. The entire kinematic information available to the fit is contained in the measured mass of the decay, m_i, and the six angles $\vec{\omega}$, $\vec{\omega'}$. The reconstructed decay time of the swapped candidate is the same as the decay time of the true candidate, because the measurements of the displacement (L_{xy}) and candidate momentum do not depend on the mass assignment of the tracks. Therefore the final signal probability density function for a B_d candidate is given as

$$\begin{aligned} P(m, \vec{\omega}, \vec{\omega'}, ct \,|\, M, S_m, S_m^{\text{swp}}, c\tau, S_{ct}, \{A_\alpha\}) = Y_{\text{sig}}(ct) \cdot \\ [(1 - f_{\text{swp}}) \cdot X_{\text{sig}}(m) Z_{\text{sig}}(\vec{\omega}) P(\vec{\omega'} \,|\, \vec{\omega}) + f_{\text{swp}} \cdot X_{\text{swp}}(m) Z_{\text{swp}}(\vec{\omega'}) P(\vec{\omega} \,|\, \vec{\omega'})], \end{aligned} \qquad (4.24)$$

where f_{swp} is the fraction of swapped candidates, Z_{sig} and Z_{swp} are the matter and anti-matter angular probability density fractions described above , X_{sig} and X_{swp} are the signal and swap Gaussian mass distributions, while $P(\vec{\omega'} \,|\, \vec{\omega})$ and $P(\vec{\omega} \,|\, \vec{\omega'})$ are the probabilities that a given combination of the vectors $\vec{\omega}$ and $\vec{\omega'}$ are observed for the same candidate. The probability densities $P(\vec{\omega'} \,|\, \vec{\omega})$ and $P(\vec{\omega} \,|\, \vec{\omega'})$ are closely related to the histograms in figure 4-3, but the normalization differ slightly between the two. The normalization of these functions is calculated from the requirement that the individual PDFs have unit integral across the entire six-angles phase space:

$$\int\int Z_{\text{sig}}(\vec{\omega}) P(\vec{\omega'} \,|\, \vec{\omega}) \, d^3\vec{\omega} d^3\vec{\omega'} = 1 \text{ and } \int\int Z_{\text{swp}}(\vec{\omega'}) P(\vec{\omega} \,|\, \vec{\omega'}) \, d^3\vec{\omega} d^3\vec{\omega'} = 1.$$

We rewrite the previous equations as:

$$\int Z_{\text{sig}}(\vec{\omega}) \left[\int P(\vec{\omega'} \,|\, \vec{\omega}) \, d^3\vec{\omega'} \right] d^3\vec{\omega} = 1$$
$$\int Z_{\text{swp}}(\vec{\omega'}) \left[\int P(\vec{\omega} \,|\, \vec{\omega'}) \, d^3\vec{\omega} \right] d^3\vec{\omega'} = 1.$$

In this form it is apparent that, in order to fulfill the requirements of normalization it is sufficient that:

$$\int P(\vec{\omega'} \,|\, \vec{\omega})\, d^3\vec{\omega'} = 1,\ \forall\, \vec{\omega}, \qquad \text{and} \qquad \int P(\vec{\omega} \,|\, \vec{\omega'})\, d^3\vec{\omega} = 1,\ \forall\, \vec{\omega'}.$$

This tells us how to calculate the functions $P(\vec{\omega'} \,|\, \vec{\omega})$ and $P(\vec{\omega} \,|\, \vec{\omega'})$ from the histograms shown in figure 4-3. To calculate $P(\vec{\omega'} \,|\, \vec{\omega})$, the histograms have to be normalized so that for a given set of true angles $\vec{\omega}$, the integral over all possible swapped angles $\vec{\omega'}$ is equal to one (column-wise normalization). To calculate $P(\vec{\omega} \,|\, \vec{\omega'})$, the corresponding histograms have to be normalized so that for a given set of swapped angles $\vec{\omega'}$, the integral over all possible true angles $\vec{\omega}$ is equal to one (row-wise normalization).

Background

We use a purely empirical model determined by examining the angular distributions in the mass sidebands. The use of an empirical model is motivated by the fact that the transversity angles do not have any intrinsic meaning for combinatorial background. The background angular probability density function is written as

$$Z_{\text{bkg}}(\vec{\omega}) = P^{\theta} \cdot P^{\phi} \cdot P^{\psi}. \tag{4.25}$$

For the $B_s \to J/\psi\phi$ mode, the individual PDFs P^{θ}, P^{ϕ}, and P^{ψ} are given by

$$\begin{aligned}
P^{\theta} &= \frac{1}{N_{\theta}} \Big(1 + p_2^{\theta} \cos^2\theta + p_4^{\theta} \cos^4\theta\Big), \\
P^{\phi} &= \frac{1}{N_{\phi}} \Big(1 + \frac{1 - p_1^{\phi} \cos\phi}{2} + \frac{1 + p_2^{\phi} \cos 2\phi}{2} + \frac{1 + p_4^{\phi} \cos 4\phi}{2}\Big), \\
P^{\psi} &= \frac{1}{N_{\psi}} \Big(1 + p_1^{\psi} \cos\psi\Big).
\end{aligned} \tag{4.26}$$

For the $B_d \to J/\psi K^*$ mode, the individual PDFs P^{θ}, P^{ϕ}, and P^{ψ} are given by

$$\begin{aligned}
P^{\theta} &= \frac{1}{N_{\theta}} \Big(1 + p_2^{\theta} \cos^2\theta + p_4^{\theta} \cos^4\theta\Big), \\
P^{\phi} &= \frac{1}{N_{\phi}} \Big(1 + p_1^{\phi} \cos(2\phi + \alpha) + p_2^{\phi} \cos^2(2\phi + \alpha)\Big), \\
P^{\psi} &= \frac{1}{N_{\psi}} \Big(1 + p_1^{\psi}(1 - \cos\psi) + p_2^{\psi}(1 - \cos^2\psi) + \\
&\qquad\qquad p_3^{\psi}(1 - \cos^3\psi) + p_4^{\psi}(1 - \cos^4\psi) + p_5^{\psi}(1 - \cos^5\psi) + \Big).
\end{aligned} \tag{4.27}$$

The coefficients in each individual PDF P^{θ}, P^{ϕ}, and P^{ψ} enter the fit as background parameters. Though this is a simple description, we find that it is adequate to describe the angular distributions in the sidebands. Figure 4-4 shows a binned fit to the sideband angular distributions using the above equations. The fit χ^2 probabilities

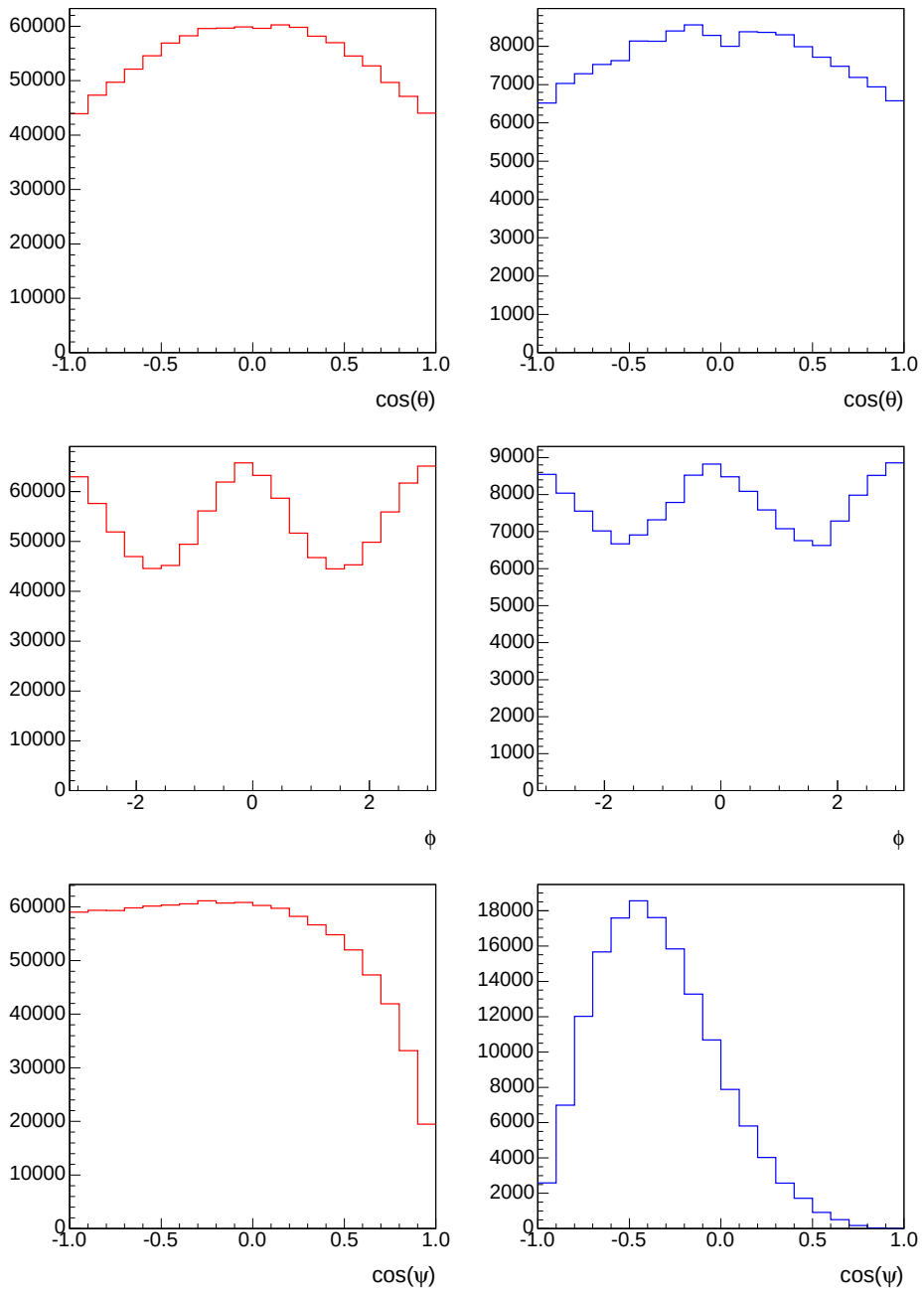


Figure 4-2: Sculpting of $\cos\theta_T$, ϕ_T, and $\cos\psi_T$ in the B_d decay after reconstruction with the correct (red) and the swapped (blue) hypotheses. The three distribution are generated flat in both cases.

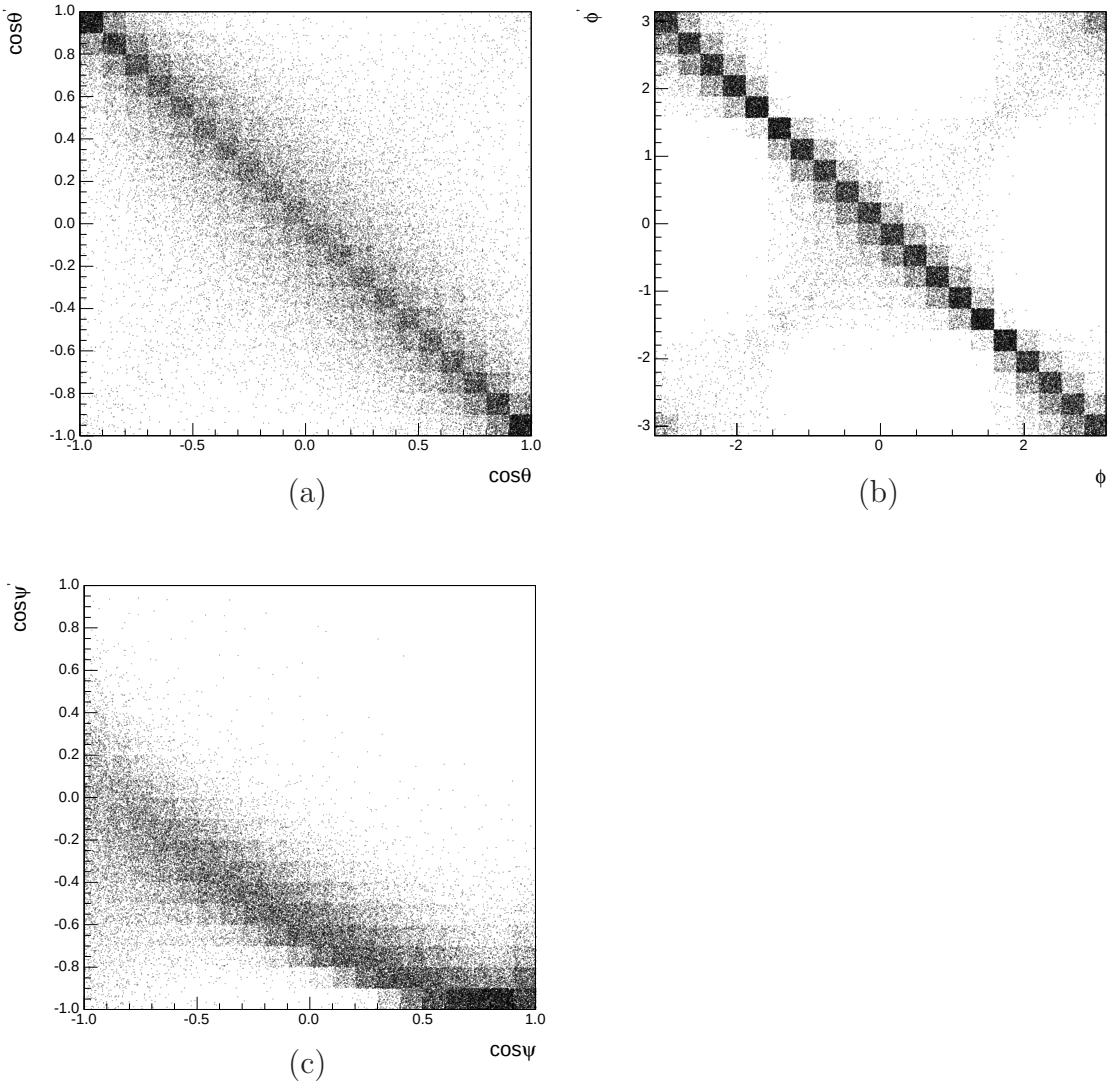


Figure 4-3: Histograms showing the relation between the transversity angles calculated with the correct K and π mass assignments, $\vec{\omega}$, and the transversity angles calculated with the swapped mass assignments, $\vec{\omega'}$, in the $B_d \to J/\psi K^*$ decay: (a) $\cos\theta'$ versus $\cos\theta$, (b) ϕ' versus ϕ, (c) $\cos\psi'$ versus $\cos\psi$.

are between 10% and 95%.

We make the assumption that the transversity distribution for the background component is separable into a product of functions of $\cos\theta$, ϕ, and $\cos\psi$. There is no a priori physical reason to expect such terms to appear, but we verify this assumption nevertheless by introducing one-by-one first-order terms in the products of the each of the transversity angles (terms proportional to $\cos\theta \cdot \phi$, $\cos\theta \cdot \cos\psi$, etc). The χ^2 probability of the fit performed in the 3D transversity histograms varies by no more than 3% when we add these terms. We thus exclude them from our likelihood model.

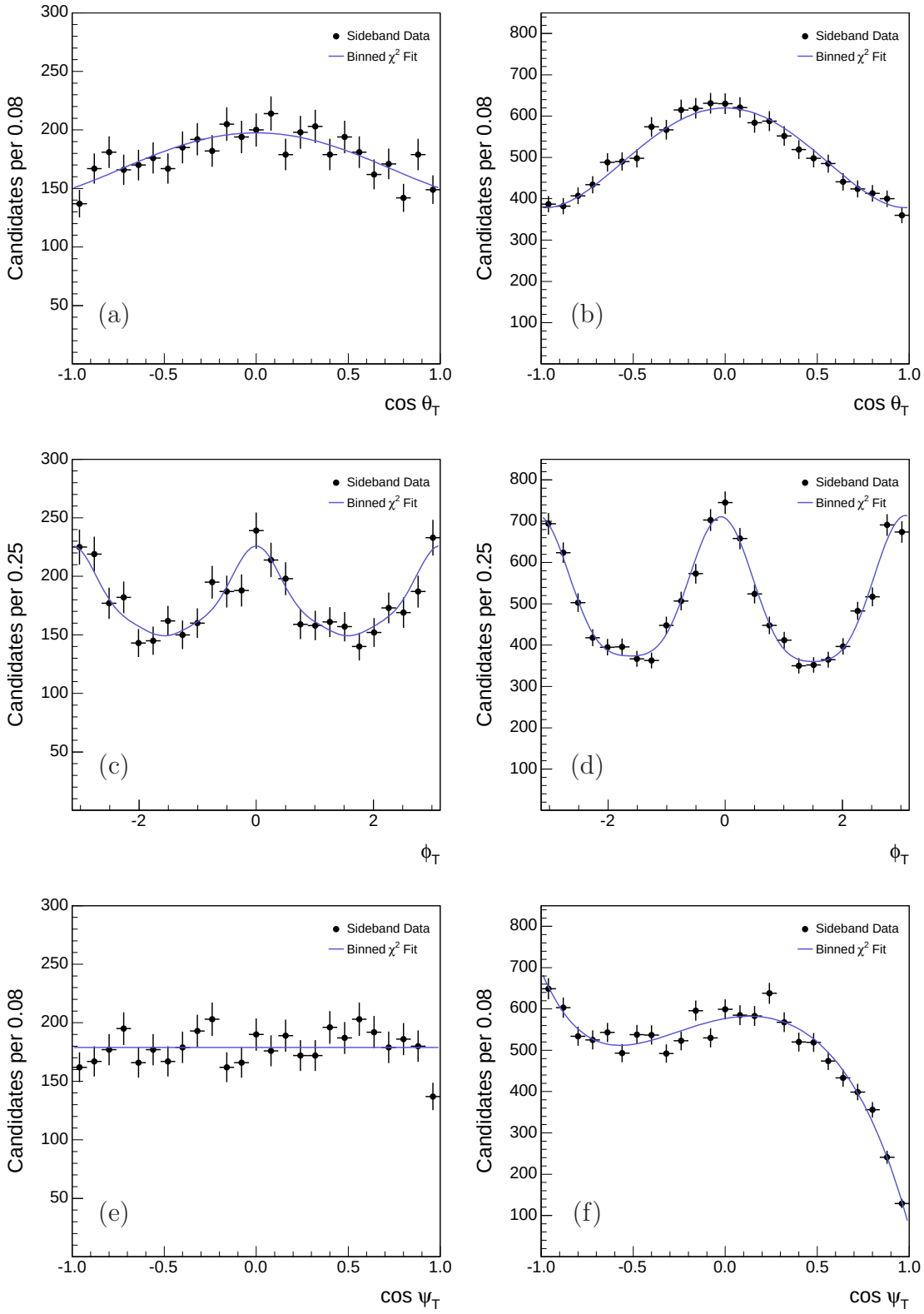


Figure 4-4: Angular distributions in the B_s (left) and B_d (right) mass sidebands after neural network selection: (a)-(b) $\cos\theta_T$, (c)-(d) ϕ_T, (e)-(f) $\cos\psi_T$.

4.1.6 PDF for Proper Decay Length and Transversity

We now address the construction of a PDF in proper decay length and transversity for $B_s \to J/\psi\phi$ decays. We reiterate that we are assuming $2\beta_s = 0$ throughout this treatment.

We begin by rewriting the time-dependent decay rate for $B_s \to J/\psi\phi$ summing contributions from B_s and $\overline{B_s}$ as done in Equation 1.45, but this time setting $2\beta_s$ to zero:

$$\frac{d^4\,\mathcal{P}}{d\vec{\omega}\;dt} \propto \left(|A_0|^2 f_1 + |A_\parallel|^2 f_2 + |A_0||A_\parallel|\cos\delta_\parallel f_5\right) e^{-\Gamma_L t} + |A_\perp|^2 f_3\, e^{-\Gamma_H t}, \tag{4.28}$$

where the dependence on $\vec{\omega}$ is implicit in the f_i functions defined in Chapter 1 and we have used the relations

$$\begin{aligned} \Delta\Gamma &\equiv \Gamma_L - \Gamma_H \\ \Gamma &\equiv \frac{\Gamma_L + \Gamma_H}{2} \end{aligned} \tag{4.29}$$

From there, the construction of joint PDF $W(ct, \sigma_{ct}, \vec{\omega})$ is matter of taking into account the following factors already discussed in previous sections: angular acceptance, PDL resolution function, and normalization. The joint PDF is then given by

$$W(ct, \sigma_{ct}, \vec{\omega}\,|\,c\tau_{L,H}, S_{ct}, \{|A_\alpha|\}, \delta_\parallel) = \frac{1}{N}\,\epsilon(\vec{\omega})\,\frac{d^4\,\mathcal{P}}{d\vec{\omega}\;dt} \otimes G(ct, \sigma_{ct}\,|\,S_c t), \tag{4.30}$$

where N is calculated using the method described in Section 4.1.5. For illustrative purposes, we note that this expression lends itself to a reorganization of terms to emphasize the fact that it is a sum of heavy and light contributions in lifetime and angular amplitudes:

$$\begin{aligned} W(ct, \sigma_{ct}, \vec{\omega}) = \frac{1}{N}\epsilon(\vec{\omega})\big[c\tau_L Y^L(ct, \sigma_{ct}) \cdot Z^L(\vec{\omega}\,|\,A_{0,\parallel}, \delta_\parallel) + \\ c\tau_H Y^H(ct, \sigma_{ct}) \cdot Z^H(\vec{\omega}\,|\,A_\perp)\;\big], \end{aligned} \tag{4.31}$$

where $Y(ct, \sigma_{ct}\,|\,c\tau, S_{ct})$ is defined as in Section 4.1.4, and

$$\begin{aligned} Z^L &= |A_0|^2 f_1 + |A_\parallel|^2 f_2 + |A_0||A_\parallel| f_5 \cos(\delta_\parallel) \\ Z^H &= |A_\perp|^2 f_3 \end{aligned}$$

This rearrangement emphasizes the fact that the two mass eigenstates decay with different angular distributions, which makes their statistical separation possible. It is functionally equivalent in every way to the expression in Equation 4.28.

4.2 Full Fit Likelihoods

We have now established the necessary foundation to construct the full joint PDF, or alternatively likelihood function for the maximum likelihood fits of both decay modes $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$.

Time-dependent angular analysis in $B_d \to J/\psi K^*$: The main caveat in constructing the likelihood for the B_d decay mode is the proper handling of the candidates with $K \leftrightarrow \pi$ mass mis-assignment. Besides this, the joint PDF is formed from the product of the mass, lifetime and transversity PDFs in each case for swapped and unswapped mass assignments, and background. The lifetime PDF is the same for both because the $K \leftrightarrow \pi$ swap does not affect the lifetime measurement.

$$\begin{aligned}\mathcal{L}_j = f_s(1-f_{\rm swp})X_{\rm sig}(m_j)Y_{\rm sig}(ct_j,\sigma_{ct_j})Z_{\rm sig}(\vec{\omega}_j)P_{\rm sig}(\sigma_{ct_j})+ \\ f_s f_{\rm swp}X_{\rm swp}(m_j)Y_{\rm sig}(ct_j,\sigma_{ct_j})Z_{\rm swp}(\vec{\omega'}_j)P_{\rm sig}(\sigma_{ct_j})+ \\ (1-f_s)X_{\rm bkg}(m_j)Y_{\rm bkg}(ct_j,\sigma_{ct_j})Z_{\rm bkg}(\vec{\omega}_j)P_{\rm bkg}(\sigma_{ct_j}). \quad (4.32)\end{aligned}$$

The likelihood $\mathcal{L}_j$ is a function of the following parameters:

- Signal: M, σ_m, S_{ct}, $c\tau$, $|A_0|$, $|A_\parallel|$, δ_0, $\delta_\parallel$, $\delta_\perp$, f_s,
- Swapped: $f_{\rm swp}$, $\sigma_m^{\rm swp}$,
- Background: A, f_-, f_+, f_{++}, $\lambda_-, \lambda_+, \lambda_{++}$, p_2^θ, p_4^θ, $p_1^\phi, p_2^\phi, p_4^\phi$, p_1^ψ.

Time-dependent angular analysis in $B_s \to J/\psi\phi$: The joint PDF is given by the multiplication of the mass PDF and the combined PDF for proper decay length and transversity. The normalization constant is implicit in $W_{\rm sig}$.

$$\begin{aligned}\mathcal{L}_j = f_s X_{\rm sig}(m_j)W_{\rm sig}(ct_j,\sigma_{ct_j},\vec{\omega}_j)P_{\rm sig}(\sigma_{ct_j})+ \\ (1-f_s)X_{\rm bkg}(m_j)Y_{\rm bkg}(ct_j,\sigma_{ct_j})Z_{\rm bkg}(\vec{\omega}_j)P_{\rm bkg}(\sigma_{ct_j}). \quad (4.33)\end{aligned}$$

The likelihood $\mathcal{L}_j$ is a function of the following parameters:

- Signal: M, σ_1, σ_2, f_m S_{ct}, $c\tau_L$, $c\tau_H$, $|A_0|$, $|A_\parallel|$, $\delta_\parallel$, f_s
- Background: A, f_-, f_+, f_{++}, $\lambda_-, \lambda_+, \lambda_{++}$, p_2^θ, p_4^θ, $p_1^\phi, p_2^\phi, p_4^\phi$, p_1^ψ

We note that $\delta_\perp$ is not a fit parameter in the B_s fit presented in this chapter, since setting β_s to zero eliminates $\delta_\perp$ from the likelihood expression.

4.3 Fitter Tests

We perform several tests of the maximum likelihood fitter in order to accomplish the following goals: to validate the correctness of its implementation, to test the

statistical limits of our sensitivity to the fit parameters, to investigate the behavior of the likelihood under various circumstances, and to detect any potential fit biases.

Pull distributions are a commonly used method to answer several of these questions. The procedure is as follows: we generate a large set of pseudo-experiments, randomly polling the probability density function in each variable subspace to assign event variables. As a consequence, each pseudo-experiment yields a different random sample of events. The event variables are assigned by randomly sampling the theoretical PDFs. For each of these pseudo-experiments, we perform a fit as we do on data. For each parameter θ in the fit that is allowed to float, the pull distribution P is defined as

$$P = \frac{\theta^{\text{fit}} - \theta^{\text{input}}}{\sigma_\theta} \tag{4.34}$$

We expect P to follow a Gaussian distribution with a mean of zero and a width of one. Complications arise in some scenarios, for example in edge cases. When there is a hard physical limit on one side or both of the allowed parameter range, we use asymmetric uncertainty evaluations in our data fit to account for this constraint. In the pull studies, we adopt the convention that the pull for a given pseudo-experiment is calculated using the positive (negative) uncertainty if the fit value is lower (higher) than the input value.

A degeneracy in the likelihood expression itself must be treated with care. An illustrative and simple example of this is a likelihood in which a theoretical parameter θ only enters the expression as $|\theta|$. In such a situation, the theoretical prediction does not allow us to make a measurement on data that distinguishes between θ and $-\theta$. We take this into account when calculating the pull value using Equation 4.34.

We generate 500 pseudo-experiments for each of the B_s and B_d decay modes, and study the distributions of pull values for each parameter. In the B_d pseudo-experiments, we find that the fitter returns unbiased estimates and consistent uncertainties for all parameters. We show representative values of pull distribution means and widths for signal parameters in Table 4.1. In the B_s pseudo-experiments, we find

Parameter	Mean	RMS
m_d	-0.079 ± 0.030	0.949 ± 0.021
$c\tau_d$	0.012 ± 0.034	1.059 ± 0.024
$\|A_0\|^2$	0.003 ± 0.033	1.027 ± 0.023
$\|A_\parallel\|^2$	0.020 ± 0.031	0.986 ± 0.022
$\delta_\parallel - \delta_0$	-0.001 ± 0.034	1.069 ± 0.024
$\delta_\perp - \delta_0$	0.033 ± 0.032	0.999 ± 0.022

Table 4.1: Pull distributions for the signal parameters in the $B_d \to J/\psi K^*$ mode obtained from generating 500 pseudo-experiments.

a small bias on the order of 20% of the statistical uncertainty in the estimation of $\delta_\parallel$. This is understood to be caused by the shape of the likelihood profile in $\delta_\parallel$, which has two minima. As a result, a pseudo-experiment of a toy sample with a true value of $\delta_\parallel$ at one of the minima in some cases returns the other minimum as the fit result. The

effect is dependent on the true value of $\delta_\parallel$, which dictates the distance between the two minima and consequently the probability of the wrong minimum to be selected by the minimization procedure. The bias observed does not affect the pull distributions of other parameters, and is left uncorrected in the final result. Table 4.2 shows the summary of pull distribution results for two initial values of $\delta_\parallel$.

Param.	Input	Pull Mean	Pull Width	Input	Pull Mean	Pull Width
m_s	5.36*	0.045 ± 0.032	1.016 ± 0.023	5.36*	0.0269 ± 0.0319	0.995 ± 0.0225
$c\tau_s$	455†	-0.004 ± 0.031	0.995 ± 0.022	440†	-0.1030 ± 0.0328	1.020 ± 0.0232
$\Delta\Gamma_s$	0.08‡	0.087 ± 0.032	1.017 ± 0.023	0.12‡	0.0619 ± 0.0328	1.020 ± 0.0232
A_0^2	0.52	0.016 ± 0.030	0.959 ± 0.021	0.57	-0.0337 ± 0.0328	1.020 ± 0.0232
$A_\parallel^2$	0.23	0.072 ± 0.031	0.976 ± 0.022	0.20	-0.0099 ± 0.0321	1.000 ± 0.0227
$\delta_\parallel$	-2.8	0.228 ± 0.031	0.994 ± 0.022	-2.0	0.0306 ± 0.0302	0.943 ± 0.0214

Table 4.2: Pull distributions for the signal parameters in the $B_s \to J/\psi\phi$ mode obtained for two sets of input values used to define the theoretical PDFs from which we randomly sample event variables. We generate 500 pseudo-experiments for each set of input values. (* Input in GeV/c^2. † Input in μm. ‡ Input in ps^{-1}.)

4.4 Fit Details and Results

In this section, we present and discuss the results of the maximum likelihood fit performed for the time-dependent angular analyses in the $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$ decay modes. In the case of $B_s \to J/\psi\phi$, motivated by theoretical predictions, we perform additional fits in which we apply constraints on the two fit parameters $c\tau$ and $\delta_\parallel$ obtained from the results of previous experiments. We supplement the discussion with additional information on how each fit was performed. The results presented in this section include only statistical uncertainties. Uncertainties arising from systematic effects are presented in Section 4.5.

In addition to numerical results, we present *projections* of the results of the likelihood fits in mass, proper decay length, and transversity variables, and overlay them on the data distributions in each of these variables. A projection in a variable x_i is a one-dimensional distribution obtained by integrating over all x_j for $j \neq i$ the likelihood expression evaluated at the best-fit values of the theoretical parameters $\vec{\theta}$ obtained from the minimization procedure.

4.4.1 $B_s \to J/\psi\phi$

In the B_s fits, we allow for the *CP*-odd and *CP*-even components to have different lifetimes. We use as fit parameters $c\tau$ and $\Delta\Gamma$, where $\tau = 1/\Gamma$ is the average lifetime and Γ and $\Delta\Gamma$ are defined in Equation 4.29. Only two of the three angular amplitudes $A_{\{0,\parallel,\perp\}}$ are used as parameters to be determined by the fit; we obtain $|A_\perp|^2$ from $|A_0|^2 + |A_\parallel|^2 + |A_\perp|^2 = 1$. We use the parameters $|A_0|^2$ and $|A_1|^2 \equiv |A_\parallel|^2/(1 - |A_0|^2)$ instead of $|A_0|^2$ and $|A_\parallel|^2$ in the minimization procedure. Both $|A_0|^2$ and $|A_1|^2$ are allowed to float in the range $[0, 1]$. This change of variables prevents the fit from

allowing the amplitudes to have unphysical values arising from $|A_0|^2 + |A_\parallel|^2 > 1$, without having to impose arbitrary limits on any individual amplitude. As discussed in section 4.2, the phase $\delta_\perp$ drops out of the likelihood expression when setting β_s to zero, and therefore cannot be determined in the likelihood fit.

We perform four fits with the $B_s \to J/\psi\phi$ sample: one unconstrained fit, and three constrained fits, in which we constrain respectively $\delta_\parallel$ only, $c\tau$ only and finally both $\delta_\parallel$ and $c\tau$ at the same time. The details of and motivations for these constraints are given below. In all cases, we use the same sample of candidates that pass the final selection described in Section 3.3.2. This sample consists of a total of 10 059 candidates, both signal and background. Aside from the changes of variables discussed above, we use the 24-parameter likelihood expression in Equation 4.33. In the constrained fits, this expression is supplemented by constraining terms as detailed below.

Unconstrained Fit

The results of the nominal unconstrained fit in the $B_s \to J/\psi\phi$ sample are shown in Table 4.3. We perform a separate determination of the positive and negative uncertainties for all parameters. We find that the parameters $c\tau$, $\Delta\Gamma$ and $\delta_\parallel$ warrant that we quote asymmetric uncertainties. We quote symmetric uncertainties for the other parameters. The matrix of correlation coefficients returned by the likelihood fit is shown in Tables 4.5 and 4.6. Finally, the projections of the fit results onto mass and proper decay length are shown along with data distributions in Figures 4-5. The projections onto the angular variables are shown in Figure 4-6.

In Figure 4-6, we show two types of angular projections. For each transversity angle, we first present the results of the likelihood fit for both signal and background

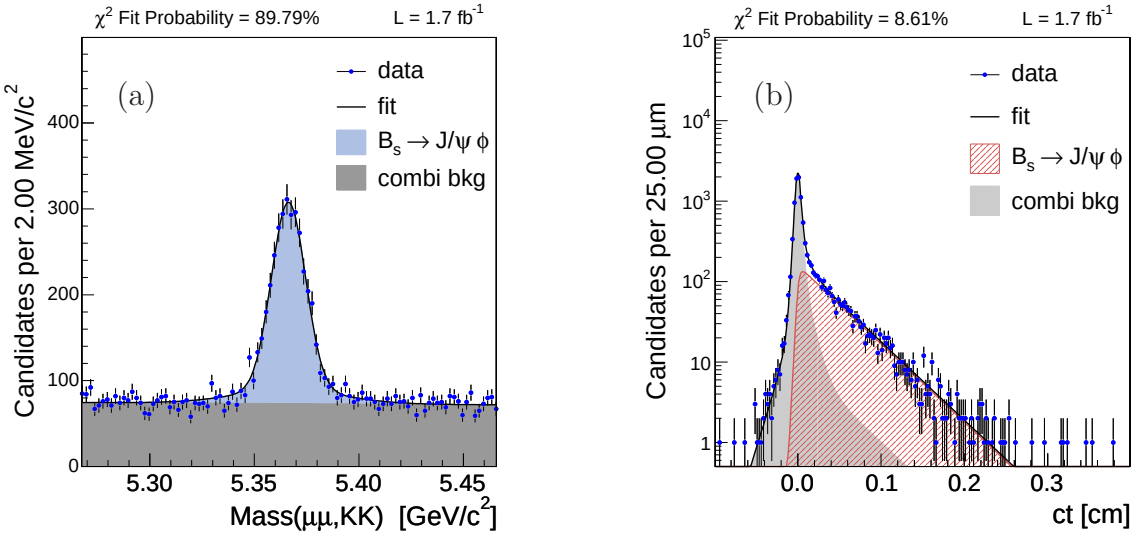


Figure 4-5: Projections of the results of the unconstrained maximum likelihood fit performed on the $B_s \to J/\psi\phi$ sample overlaid on data: (a) Mass, (b) PDL.

Parameter	Fit Result	Unit
M_s	5366.32 ± 0.21	MeV/c^2
σ_1	8.14 ± 0.36	MeV/c^2
σ_2	25.84 ± 4.85	MeV/c^2
f_m	18.26 ± 3.71	%
A	-0.95 ± 1.01	(GeV/c^2)$^{-1}$
$1 - f_s$	72.8 ± 0.6	%
$c\tau_s$	$455.5^{+13.1}_{-12.9}$	μm
$\Delta\Gamma_s$	$0.071^{+0.064}_{-0.059}$	ps^{-1}
S_{ct}	1.34 ± 0.02	
f_+	9.9 ± 1.0	%
f_{++}	2.3 ± 0.9	%
f_-	2.5 ± 0.4	%
$c\tau_+$	82.4 ± 14.8	μm
$c\tau_{++}$	448.4 ± 90.9	μm
$c\tau_-$	132.4 ± 15.1	μm
$\delta_\parallel$	$2.63^{+0.18}_{-0.29}$	
$\|A_0\|^2$	0.529 ± 0.020	
$\|A_1\|^2$	0.511 ± 0.053	
p_2^θ	-0.517 ± 0.112	
p_4^θ	0.266 ± 0.129	
p_1^ϕ	0.017 ± 0.046	
p_2^ϕ	0.537 ± 0.060	
p_4^ϕ	0.048 ± 0.045	
p_1^ψ	0.049 ± 0.021	

Table 4.3: Results of the unconstrained fit of the $B_s \to J/\psi\phi$ sample. All uncertainties quoted are statistical only.

components overlaid on the data distributions in each angle. Second, we show the results of the fit for the signal component only, and overlay it on the sideband-subtracted and sculpting-corrected distributions of the data candidates in the central mass region only. The sideband subtraction is performed by subtracting the angular distribution of sideband candidates from the angular distribution of central mass candidates, after scaling the distributions to account for the difference in the number of candidates in each region. The sculpting correction is performed in the three-dimensional transversity space by dividing the sideband-subtracted angular distribution of central mass candidates by the sculpting distribution in the transversity angles.

Constrained Fits

In addition to the nominal fit in the $B_s \to J/\psi\phi$ mode, we also perform constrained fits in which we provide the fitter with additional information gained from other experiments and motivated by theoretical predictions. Constraints are applied in practice by multiplying the likelihood expression by a Gaussian factor. For a given parameter θ in our likelihood expression, and its independently obtained value and uncertainty $\{\Theta, \sigma_\Theta\}$, the likelihood is modified in the following way:

$$\mathcal{L} \to \frac{1}{\sqrt{2\pi}\sigma_\Theta} e^{\frac{-(\theta-\Theta)^2}{2\sigma_\Theta^2}} \cdot \mathcal{L}, \tag{4.35}$$

where Θ is the latest world average of experimental measurements of θ, and σ_Θ is calculated from both its experimental and theoretical uncertainties using $\sigma_\Theta^2 = \sigma_{\rm exp}^2 + \sigma_{\rm th}^2$, whenever available. Multiple constraints can be combined, with each constraint contributing another Gaussian multiplicative factor to the fit likelihood.

We apply two constraints, first separately, then combined. First we use the theoretical prediction of Equation 1.47 that the values Γ_d and Γ_s are within 1% of one another. We constrain $c\tau_s$ using $c\tau_d = 458.7$ μm and $\sigma_{c\tau_d} = 5.3$ μm, which incorporates both theoretical and experimental uncertainties [41]. Second we use the predictions of SU(3) flavor symmetry mentioned in Section 1.4 to constrain the value of $\delta_\parallel$ to its measured value in the $B_d \to J/\psi K^*$ decay. We use for the constraint $\delta_\parallel = 2.36$ and $\sigma_{\delta_\parallel} = 0.17$, which includes only the experimental uncertainty [41]. The second constraint is additionally motivated by the anticipation of an simultaneous ambiguity in the sign of $\Delta\Gamma$ and the sign of $\cos\delta_\parallel$ and $\cos\delta_\perp$, which will be discussed in Chapter 6.

The results of all three constrained fits are presented in Table 4.4. In addition, we show the likelihood projections in mass, PDL and transversity angles in Figures 4-7–4-9. For the angular projections of the constrained fit results, we show only the sideband-subtracted and sculpting-corrected distributions in the interest of brevity.

4.4.2 $B_d \to J/\psi K^*$

The main goal of the analysis in this decay channel is to extract the lifetime and angular amplitudes for comparison with independently obtained measurements of these parameters. An additional goal is to obtain results whose precision rivals those from the B factories. The $B_d \to J/\psi K^*$ fit is performed with the addition of the S-wave component arising from the interference of non-resonant $K\pi$ decays. Both $\delta_\parallel$ and $\delta_\perp$ are measured using the time-dependent angular analysis, in contrast to the $B_s \to J/\psi\phi$ fit.

The results of the nominal likelihood fit are summarized in Table 4.7. Figure 4-10 shows the projections of the likelihood in the mass, PDL, and angular sub-spaces, with the data distributions overlaid. We show the angular projections with the sculpting correction and sideband-subtraction. In addition to the nominal projections, we show the projection in $\cos\psi$ obtained from a fit without the S-wave component included.

We see that the inclusion of the S-wave component leads to considerably better agreement between fit and data, as expected.

The results obtained from the $B_d \to J/\psi K^*$ fit conclusively validate the time-dependent angular analysis framework. Furthermore, the results obtained are indeed competitive with the B factory results, as discussed later in this document (Chapter 7).

Parameter	Constrain $c\tau$	Constrain $\delta_\parallel$	Constrain $c\tau$, $\delta_\parallel$	Unit
M_s	5366.32 ± 0.21	5366.32 ± 0.21	5366.32 ± 0.21	MeV/c^2
σ_1	8.14 ± 0.35	8.14 ± 0.36	8.14 ± 0.35	MeV/c^2
σ_2	25.82 ± 4.82	25.85 ± 4.86	25.8 ± 4.8	MeV/c^2
f_m	18.25 ± 3.70	18.28 ± 3.71	18.27 ± 3.70	%
A	-0.95 ± 1.01	-0.95 ± 1.01	-0.95 ± 1.01	$(\mathrm{GeV}/c^2)^{-1}$
$1 - f_s$	72.8 ± 0.6	72.8 ± 0.6	72.82 ± 0.59	%
$c\tau_s$	$458.2^{+4.9}_{-4.8}$	$455.7^{+13.6}_{-13.2}$	$458.3^{+4.9}_{-4.8}$	μm
$\Delta\Gamma_s$	$0.079^{+0.047}_{-0.048}$	$0.070^{+0.066}_{-0.060}$	$0.078^{+0.048}_{-0.049}$	ps^{-1}
S_{ct}	1.34 ± 0.02	1.34 ± 0.02	1.34 ± 0.02	
f_+	9.9 ± 1.0	9.8 ± 1.0	9.9 ± 1.0	%
f_{++}	2.3 ± 0.9	2.3 ± 0.9	2.3 ± 0.9	%
f_-	2.5 ± 0.5	2.5 ± 0.4	2.5 ± 0.4	%
$c\tau_+$	82.4 ± 14.7	82.4 ± 14.7	82.3 ± 14.7	μm
$c\tau_{++}$	446.6 ± 90.2	448.6 ± 91.1	447.0 ± 90.3	μm
$c\tau_-$	132.4 ± 15.1	132.4 ± 15.1	132.4 ± 15.1	μm
$\delta_\parallel$	$2.63^{+0.18}_{-0.29}$	$2.49^{+0.11}_{-0.12}$	$2.49^{+0.11}_{-0.12}$	
$\lvert A_0\rvert^2$	0.531 ± 0.018	0.529 ± 0.020	0.531 ± 0.018	
$\lvert A_1\rvert^2$	0.516 ± 0.049	0.524 ± 0.051	0.529 ± 0.046	
p_2^θ	-0.517 ± 0.110	-0.518 ± 0.110	-0.518 ± 0.110	
p_4^θ	0.266 ± 0.127	0.266 ± 0.127	0.266 ± 0.126	
p_1^ϕ	0.017 ± 0.046	0.017 ± 0.046	0.017 ± 0.046	
p_2^ϕ	0.537 ± 0.060	0.536 ± 0.060	0.536 ± 0.060	
p_4^ϕ	0.048 ± 0.045	0.049 ± 0.045	0.049 ± 0.045	
p_1^ψ	0.049 ± 0.021	0.049 ± 0.021	0.049 ± 0.021	

Table 4.4: Results of the constrained fits of the $B_s \to J/\psi\phi$ sample. All uncertainties quoted are statistical only.

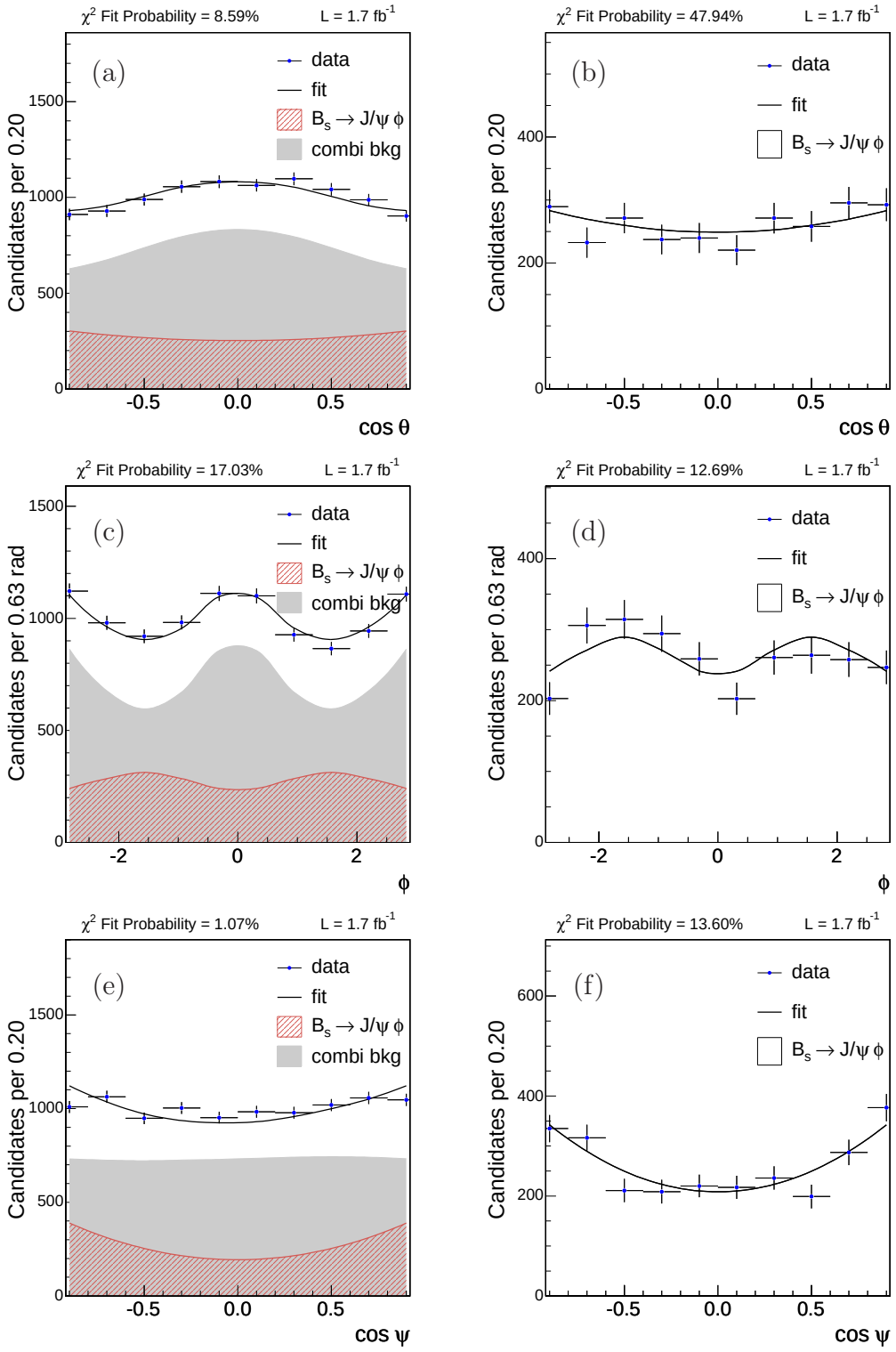


Figure 4-6: Projections of the results of the unconstrained maximum likelihood fit performed on the $B_s \to J/\psi\phi$ sample overlaid on data: (a)-(b) $\cos\theta$, (c)-(d) ϕ, (e)-(f) $\cos\psi$. On the left we show the signal and background distributions, and on the right, we show sideband-subtracted and sculpting-corrected signal distributions.

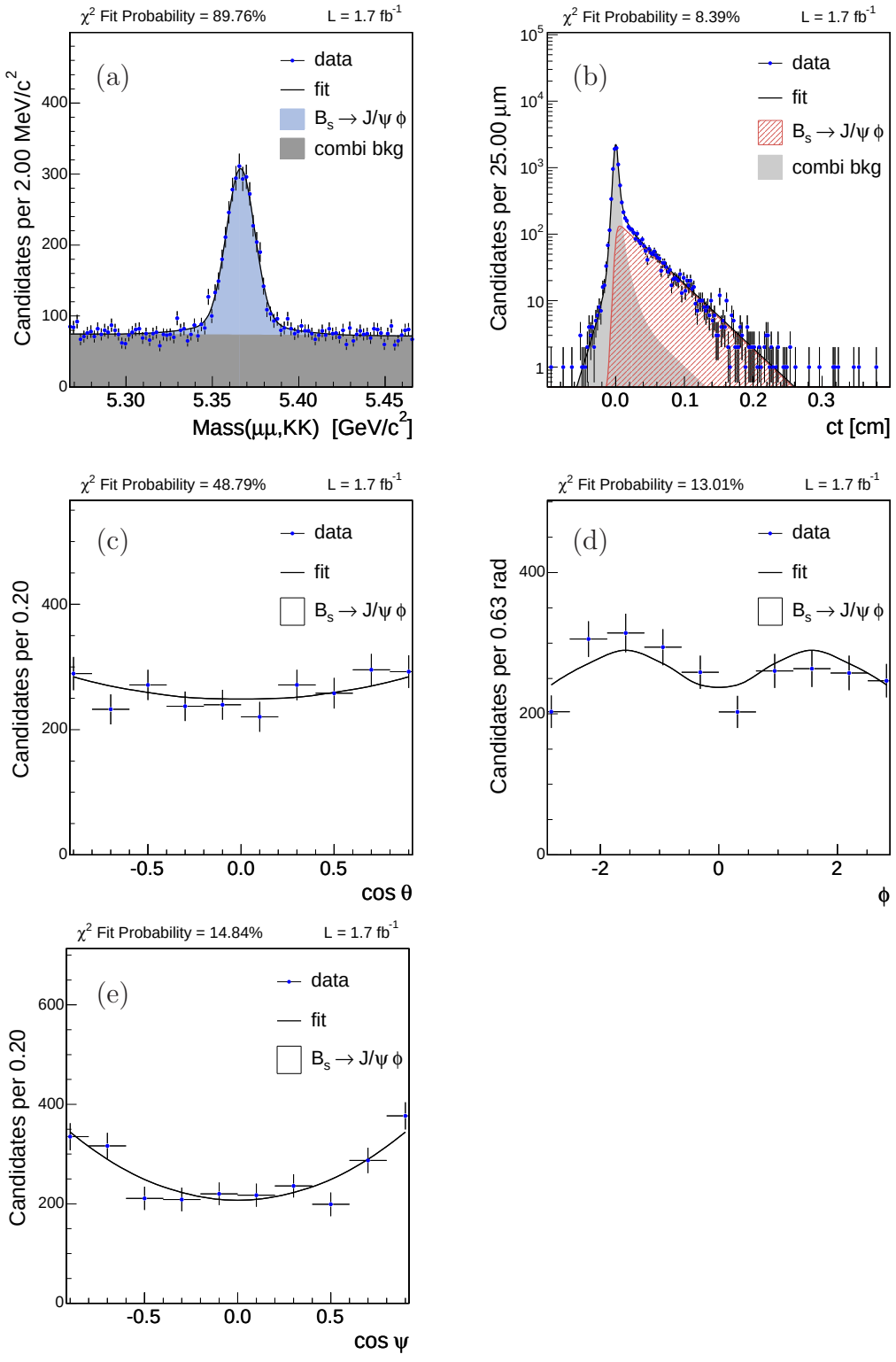


Figure 4-7: Projections of the results of the constrained maximum likelihood fit with $c\tau_s$ constrained to 458.7 ± 5.3 μm performed on the $B_s \to J/\psi\phi$ sample, overlaid on data: (a) Mass (b) PDL (c)† $\cos\theta$, (d)† ϕ, (e)† $\cos\psi$. [† Sideband-subtracted and sculpting-corrected signal distributions.]

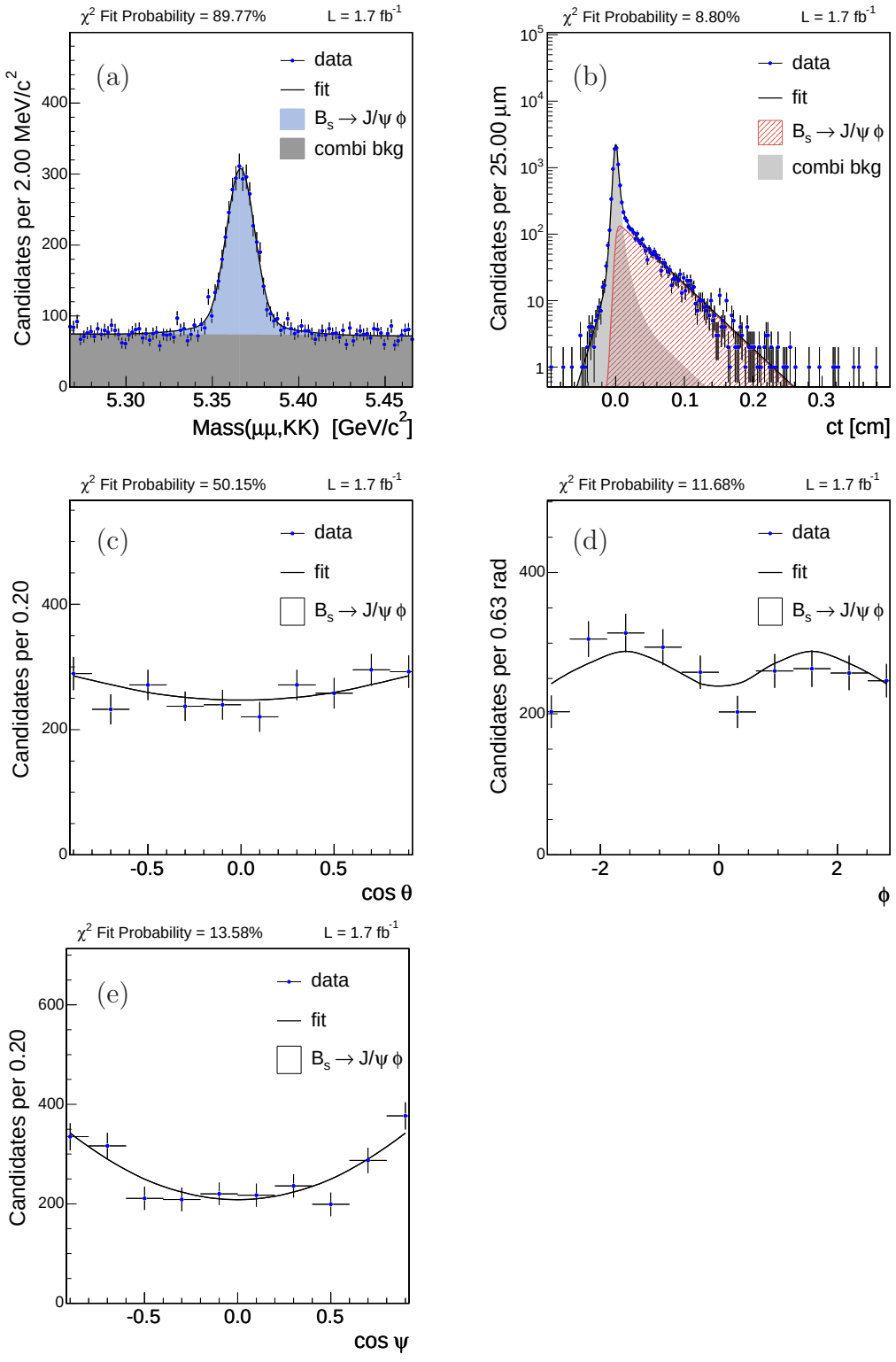


Figure 4-8: Projections of the results of the constrained maximum likelihood fit with $\delta_{\|}$ constrained to 2.36 ± 0.17 μm performed on the $B_s \to J/\psi\phi$ sample overlaid on data: (a) Mass (b) PDL (c)† $\cos\theta$, (d)† ϕ, (e)† $\cos\psi$. [† Sideband-subtracted and sculpting-corrected signal distributions.]

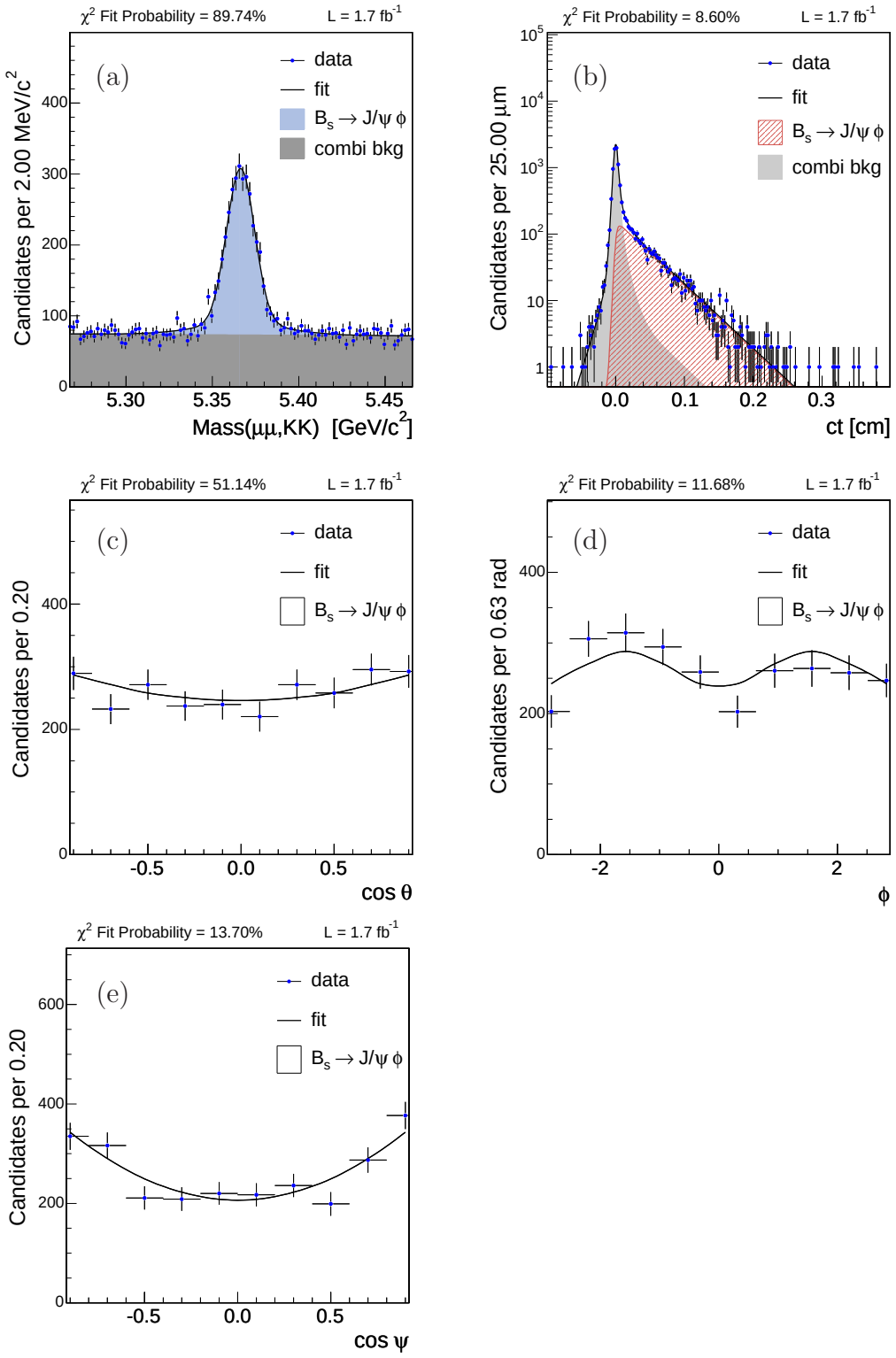


Figure 4-9: Projections of the results of the constrained maximum likelihood fit with $\delta_{\parallel}$ constrained to 2.36 ± 0.17 μm and $c\tau_s$ constrained to 458.7 ± 5.3 μm performed on the $B_s \to J/\psi\phi$ sample, overlaid on data: (a) Mass (b) PDL (c)† $\cos\theta$, (d)† ϕ, (e)† $\cos\psi$. [† Sideband-subtracted and sculpting-corrected signal distributions.]

Par.	M_s	σ_1	σ_2	f_m	A	f_s	$c\tau_s$	$\Delta\Gamma$	S_{ct}	f_+	f_{++}	f_-
M_s	1	-0.122	-0.130	0.131	-0.027	0.049	0.002	0.000	-0.016	0.000	0.063	0.006
σ_1	-0.122	1	0.779	-0.845	0.005	-0.314	-0.013	0.003	0.053	-0.014	-0.347	-0.012
σ_2	-0.130	0.779	1	-0.833	0.002	-0.445	-0.017	0.002	0.076	-0.025	-0.496	-0.018
f_m	0.131	-0.845	-0.833	1	-0.006	0.253	-0.013	-0.006	-0.050	-0.022	0.325	0.014
A	-0.027	0.005	0.002	-0.006	1	0.002	-0.001	-0.001	0.001	0.001	0.000	0.000
f_s	0.049	-0.314	-0.445	0.253	0.002	1	0.115	0.006	-0.045	0.101	0.312	0.000
$c\tau_s$	0.002	-0.013	-0.017	-0.013	-0.001	0.115	1	0.652	-0.006	0.058	0.029	-0.004
$\Delta\Gamma$	0.000	0.003	0.002	-0.006	-0.001	0.006	0.652	1	-0.001	0.000	0.026	0.001
S_{ct}	-0.016	0.053	0.076	-0.050	0.001	-0.045	-0.006	-0.001	1	-0.361	-0.158	-0.455
f_+	0.000	-0.014	-0.025	-0.022	0.001	0.101	0.058	0.000	-0.361	1	-0.118	0.189
f_{++}	0.063	-0.347	-0.496	0.325	0.000	0.312	0.029	0.026	-0.158	-0.118	1	0.082
f_-	0.006	-0.012	-0.018	0.014	0.000	0.000	-0.004	0.001	-0.455	0.189	0.082	1
$c\tau_+$	-0.024	0.106	0.152	-0.108	0.001	-0.087	-0.018	-0.019	0.242	-0.214	-0.753	-0.153
$c\tau_{++}$	-0.023	0.095	0.135	-0.088	0.002	-0.092	-0.090	-0.050	0.117	0.178	-0.748	-0.073
$c\tau_-$	-0.006	0.021	0.031	-0.020	0.000	-0.017	-0.001	-0.001	0.322	-0.148	-0.075	-0.628
$\lvert A_0\rvert^2$	0.000	-0.017	-0.028	0.003	0.000	0.053	0.456	0.654	-0.001	0.017	0.043	-0.001
$\lvert A_1\rvert^2$	0.005	-0.010	-0.017	0.009	-0.001	0.016	0.423	0.525	-0.002	0.009	0.021	0.001
p_2^θ	0.001	-0.018	-0.024	0.014	0.000	0.018	-0.012	-0.017	-0.003	0.007	0.009	0.001
p_4^θ	0.000	0.012	0.015	-0.008	0.000	-0.013	0.002	0.004	0.002	-0.007	-0.005	-0.001
p_1^ϕ	-0.002	0.010	0.013	-0.012	0.000	-0.003	0.000	-0.003	0.003	0.000	-0.013	-0.002
p_2^ϕ	-0.010	0.057	0.083	-0.056	0.000	-0.054	-0.009	0.003	0.010	-0.011	-0.050	-0.002
p_4^ϕ	0.000	0.000	0.005	0.001	0.000	-0.011	-0.002	0.003	-0.001	-0.002	-0.004	0.001
p_1^ψ	0.005	-0.009	-0.010	0.011	-0.001	0.000	-0.001	0.001	-0.002	0.002	0.002	0.001
$\delta_\parallel$	-0.007	-0.004	-0.004	-0.005	0.000	0.018	-0.027	0.017	-0.001	0.005	0.012	0.000
Glob.	0.140	0.859	0.917	0.909	0.028	0.544	0.681	0.813	0.563	0.687	0.932	0.684

Table 4.5: Correlation matrix for the parameters in the $B_s \to J/\psi\phi$ likelihood returned by the unconstrained fit (Part 1 of 2).

Par.	$c\tau_+$	$c\tau_{++}$	$c\tau_-$	$\|A_0\|^2$	$\|A_1\|^2$	p_2^θ	p_4^θ	p_1^ϕ	p_2^ϕ	p_4^ϕ	p_1^ψ	$\delta_\parallel$
M_s	-0.024	-0.023	-0.006	0.000	0.005	0.001	0.000	-0.002	-0.010	0.000	0.005	-0.007
σ_1	0.106	0.095	0.021	-0.017	-0.010	-0.018	0.012	0.010	0.057	0.000	-0.009	-0.004
σ_2	0.152	0.135	0.031	-0.028	-0.017	-0.024	0.015	0.013	0.083	0.005	-0.010	-0.004
f_m	-0.108	-0.088	-0.020	0.003	0.009	0.014	-0.008	-0.012	-0.056	0.001	0.011	-0.005
A	0.001	0.002	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	0.000	-0.001	0.000
f_s	-0.087	-0.092	-0.017	0.053	0.016	0.018	-0.013	-0.003	-0.054	-0.011	0.000	0.018
$c\tau_s$	-0.018	-0.090	-0.001	0.456	0.423	-0.012	0.002	0.000	-0.009	-0.002	-0.001	-0.027
$\Delta\Gamma$	-0.019	-0.050	-0.001	0.654	0.525	-0.017	0.004	-0.003	0.003	0.003	0.001	0.017
S_{ct}	0.242	0.117	0.322	-0.001	-0.002	-0.003	0.002	0.003	0.010	-0.001	-0.002	-0.001
f_+	-0.214	0.178	-0.148	0.017	0.009	0.007	-0.007	0.000	-0.011	-0.002	0.002	0.005
f_{++}	-0.753	-0.748	-0.075	0.043	0.021	0.009	-0.005	-0.013	-0.050	-0.004	0.002	0.012
f_-	-0.153	-0.073	-0.628	-0.001	0.001	0.001	-0.001	-0.002	-0.002	0.001	0.001	0.000
$c\tau_+$	1	0.663	0.116	-0.015	-0.014	0.002	-0.003	0.012	0.019	0.000	-0.001	0.002
$c\tau_{++}$	0.663	1	0.057	-0.037	-0.026	0.006	-0.007	0.008	0.013	0.000	0.001	-0.004
$c\tau_-$	0.116	0.057	1	-0.001	-0.001	-0.001	0.001	0.002	0.004	-0.001	-0.001	-0.001
$\|A_0\|^2$	-0.015	-0.037	-0.001	1	0.277	-0.014	0.002	-0.004	0.023	0.007	0.000	0.049
$\|A_1\|^2$	-0.014	-0.026	-0.001	0.277	1	-0.026	0.003	-0.002	-0.033	-0.003	0.001	-0.382
p_2^θ	0.002	0.006	-0.001	-0.014	-0.026	1	-0.961	-0.001	-0.002	-0.001	0.004	0.009
p_4^θ	-0.003	-0.007	0.001	0.002	0.003	-0.961	1	0.001	0.001	0.001	-0.004	-0.001
p_1^ϕ	0.012	0.008	0.002	-0.004	-0.002	-0.001	0.001	1	0.225	0.048	-0.002	-0.001
p_2^ϕ	0.019	0.013	0.004	0.023	-0.033	-0.002	0.001	0.225	1	0.321	-0.001	0.016
p_4^ϕ	0.000	0.000	-0.001	0.007	-0.003	-0.001	0.001	0.048	0.321	1	0.001	-0.002
p_1^ψ	-0.001	0.001	-0.001	0.000	0.001	0.004	-0.004	-0.002	-0.001	0.001	1	0.001
$\delta_\parallel$	0.002	-0.004	-0.001	0.049	-0.382	0.009	-0.001	-0.001	0.016	-0.002	0.001	1
Glob.	0.881	0.814	0.630	0.664	0.667	0.961	0.961	0.227	0.395	0.323	0.013	0.461

Table 4.6: Correlation matrix for the parameters in the $B_s \to J/\psi\phi$ likelihood returned by the unconstrained fit (Part 2 of 2).

Parameter	Fit Result	Unit
M_s	5279.40 ± 0.20	MeV/c^2
σ_1	7.5 ± 0.20	MeV/c^2
σ_2	18.2 ± 0.4	MeV/c^2
f_m	0.40 ± 0.08	%
A	-2.16 ± 0.54	(GeV/c^2)$^{-1}$
$1 - f_s$	0.750 ± 0.003	%
$c\tau_d$	455.3 ± 6.4	μm
S_{ct}	1.36 ± 0.01	
f_+	2.4 ± 0.2	%
f_{++}	5.3 ± 0.5	%
f_-	2.4 ± 0.2	%
$c\tau_+$	63 ± 8	μm
$c\tau_{++}$	382 ± 26	μm
$c\tau_-$	104 ± 9	μm
$\lvert A_0\rvert^2$	0.560 ± 0.009	
$\lvert A_\parallel\rvert^2$	0.210 ± 0.012	
$\delta_\parallel$	-2.946 ± 0.089	
$\delta_\perp$	2.949 ± 0.055	
$\cos\lambda$	0.994 ± 0.003	
δ_S	2.200 ± 0.217	
p_2^θ	-0.600 ± 0.055	
p_4^θ	0.223 ± 0.050	
p_1^ϕ	0.334 ± 0.010	
p_2^ϕ	0.142 ± 0.014	
α	0.219 ± 0.021	
p_1^ψ	0.845 ± 0.064	
p_2^ψ	0.502 ± 0.059	
p_3^ψ	-0.167 ± 0.130	
p_4^ψ	-0.888 ± 0.069	
p_5^ψ	-1.076 ± 0.122	

Table 4.7: Results of the fit of the $B_d \to J/\psi K^*$ sample. All uncertainties quoted are statistical only.

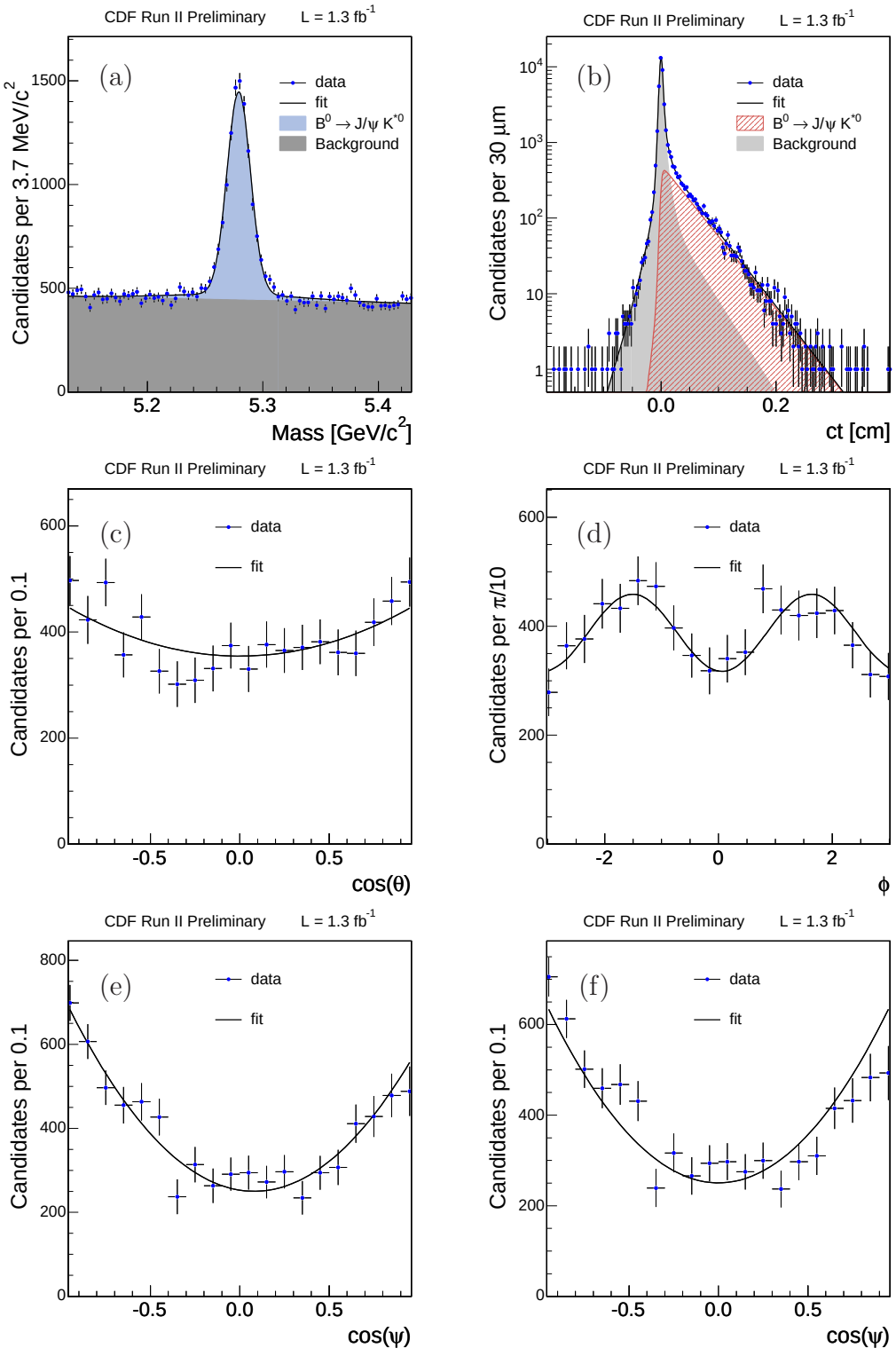


Figure 4-10: Projections of the results of the unconstrained maximum likelihood fit performed on the $B_d \to J/\psi K^*$ sample overlaid on data: (a) Mass, (b) PDL, (c) $\cos\theta$, (d) ϕ, (e) $\cos\psi$, (f) $\cos\psi$ without the inclusion of the S-wave component.

4.5 Systematic Uncertainties

Every analysis and measurement suffers from biases and systematic effects which cannot be removed or corrected. Naturally, we remove biases which are fully understood. However, we are limited by our finite knowledge of various potential sources of systematic effects. The solution is to estimate uncertainties arising from our inability to pinpoint all these effects*. The estimation of systematic uncertainties stemming from these effects is an important part of quoting a reliable and representative overall uncertainty on all quantities measured.

In this section, we discuss the sources of systematic uncertainty that we identify and deem important to this analysis. We begin by discussing our procedure for estimating each uncertainty, along with the reasoning behind the estimation of each systematic effect. Each individual systematic effect induces additional uncertainties in the measurement of the signal parameters. We assess their quantitative impact on the key signal parameters: the average lifetime $c\tau$, the width difference $\Delta\Gamma$, and the angular amplitudes $\{A_\alpha\}$. We report the impact of these effects on fitted parameters. All of the systematic effects except for the B_s crossfeed are relevant to both decay modes. Tables 4.8 and 4.9 list the systematic uncertainty associated with each effect for these signal parameters. The total systematic uncertainty is calculated by adding the individual uncertainties in quadrature.

4.5.1 SVX Alignment

The calculation of the proper decay length requires precise vertexing. This vertexing is reliant on our knowledge of the positions in space of all elements of the tracking detectors, e.g. the silicon wafers of the SVX tracker, which are known up to a finite precision only. Lifetime measurements are affected by this limited knowledge. Recent studies at CDF [103] suggest that the effect is a 2 μm systematic uncertainty for $c\tau$ in single-lifetime measurements.

We expect that an additional uncertainty on the lifetime measurement causes a systematic uncertainty in the measurement of the other signal parameters, because the signal parameters we measure are correlated with the average lifetime of the B meson in each decay mode. In order to assess the effect of this systematic uncertainty on other parameters in our fit, we conduct pseudo-experiments in which we generate samples with an average lifetime ($c\tau$) shifted from the measured value by $\pm$ 2 μm. We note that in the case of $B_s \to J/\psi\phi$, another strategy is to generate pseudo-experiments in which the lifetimes of the light and heavy eigenstates, $c\tau_L$ and $c\tau_H$, are shifted by $\pm$ 2 μm. However, the evaluation of the systematic uncertainty in Reference [103] is done using the average lifetime, including in the case of $B_s \to J/\psi\phi$. Therefore we choose to apply this shift to $c\tau$ in both the B_d and B_s decay modes. In addition to preserving the conditions under which the shift of $\pm$ 2 μm is originally

*By definition, systematic effects arising from our limited knowledge do not change the central value of a parameter obtained from our likelihood fit. Rather, they result in an additional uncertainty on each parameter they affect.

evaluated, this allows us to measure the effect on $\Delta\Gamma$ more effectively. This is because shifting $c\tau_L$ and $c\tau_H$ by $\pm$ 2 μm effectively cancels out the effect on $\Delta\Gamma$ on average.

We also assess the effect that this shift has on the remaining fit parameters. We conclude that the effect of the ± 2 μm shift in the average lifetime in the $B_s \to J/\psi\phi$ mode is a systematic uncertainty of ± 1 μm for $c\tau$, ± 0.002 ps^{-1} for $\Delta\Gamma$, and ± 0.001 for $\delta_\parallel$. The ± 1 μm uncertainty for $c\tau$ is smaller than the original shift. This is understood as a result of two factors: the fact that the value of $\Delta\Gamma$ is unchanged in the pseudo-experiments we perform to asses this systematic effect, and the correlation between $c\tau$ and $\Delta\Gamma$. In the $B_d \to J/\psi K^*$ fit, the effect of the ± 2 μm shift in the average lifetime is a systematic uncertainty of ± 4 μm for $c\tau$, ± 0.001 for $\delta_\parallel$, and ± 0.001 for $\delta_\perp$. This systematic effect is the biggest contributor to the overall systematic uncertainty on the B_d lifetime.

4.5.2 $B_s \to J/\psi\phi$ and $B_d \to J/\psi K^*$ Crossfeed

We have already discussed the possibility of a $K \leftrightarrow \pi$ mass mis-assignment in the case of real $B_d \to J/\psi K^*$ decays. In that discussion, both particles are misidentified and the mass assignments are swapped. A related effect also occurs in reconstruction when π and K are at times simply misidentified singly, without a swap of the other daughter particle of the vector meson. In this case, KK is reconstructed as $K\pi$, and $K\pi$ as KK.

This effect causes the mis-reconstruction of a small percentage of $B_d \to J/\psi K^*$ decays as $B_s \to J/\psi\phi$, in the case that the π is misidentified as a K. It also causes the mis-reconstruction of a portion of $B_s \to J/\psi\phi$ decays as $B_d \to J/\psi K^*$, in the case that one of the kaons is misidentified as a π. The probability of misidentification is enhanced for tracks with $p_T > 1.5$ GeV/c because the particle identification algorithms lose their separation power in this momentum range. If the mass assignments made in these erroneous reconstructions fall within the respective requirements of the "faked" decay, the error results in a pollution of each data sample from the other, which we refer to as a *crossfeed*.

To assign a systematic uncertainty on measured parameters, we first estimate the size of this effect in each decay mode. Therefore we determine the fraction in composition of each sample that is a result of misidentification. Taking a B_d decay mis-reconstructed as B_s decay as an example, this fraction is equal to

$$f(B_d \text{ in } \{B_s\}) = \frac{R_d}{R_s} \cdot \frac{Br(B_d \to J/\psi K^*) Br(K^* \to K\pi)}{Br(B_s \to J/\psi\phi) Br(\phi \to KK)} \cdot \epsilon_{B_s}(B_d), \tag{4.36}$$

where $\frac{R_d}{R_s}$ is the relative production rate of d and s quarks at the Tevatron, $Br(\dots)$ indicates a branching fraction, and $\epsilon_{B_s}(B_d)$ is the efficiency of misidentifying a B_d meson as a B_s meson, which takes into account online and offline reconstruction as well as our selection criteria. The parameter ϵ is estimated using Monte Carlo simulation, including detector and trigger simulation. The branching fractions are taken from Reference [1], while the ratio of quark production rates is taken from Reference [34]. Using Equation 4.36, we estimate that approximately 2.5% of the

B_s sample is composed of $B_d \to J/\psi K^*$ decays, which are not accounted for in our likelihood framework. The fraction of B_s decays in the B_d sample is estimated to be more than one order of magnitude smaller. As a result, we consider its effect for the B_d signal sample negligible and do not assign a systematic uncertainty for crossfeed in this direction.

Having estimated the B_d pollution in the B_s sample, we proceed to determine its effect on measured parameters. We generate $B_s \to J/\psi\phi$ pseudo-experiments and in each case, replace 2.5% of the events with events that are assigned mass, proper decay length and transversity angles from the $B_d \to J/\psi K^*$ PDFs. We then perform a fit using the B_s likelihood and compare the results with the fit in the pseudo-experiments without B_d pollution.

We observe that the B_d crossfeed in the $B_s \to J/\psi\phi$ sample results in a systematic uncertainty of -1 μm for the average B_s lifetime $c\tau$, and ± 0.004 ps^{-1} for $\Delta\Gamma$. The effect for the angular amplitudes is ± 0.003 for $|A_0|^2$, ± 0.003 for $|A_\parallel|^2$, and ± 0.070 for $\delta_\parallel$.

4.5.3 Mass Model for Signal Components

As we discuss in Section 4.1.3, our model for the mass distribution in the signal portion of our analysis sample is driven by a desire to model detector resolution effects. The two-Gaussian modeling is empirically motivated, as it gives a close match to observed data. However, there is no *a priori* reason to believe that it is the perfect choice to model the detector resolution. We reason that any hypothetical model that provides a better description of the detector resolution effects in the mass subspace has to include at the least two Gaussians in conjunction with another function, e.g. to model the Gaussian tails better. We further reason that our exclusion of the additional component of this hypothetical better model has a smaller effect than the use of one Gaussian as compared with the use of two.

We therefore estimate an upper bound for the systematic uncertainty caused by the mass model by generating pseudo-experiments from a two-Gaussian distribution in mass, and fitting the sample from each pseudo-experiment using only one Gaussian. The effect on the measured parameters is reported, and treated as a conservative (i.e. upper limit) estimate of the systematic uncertainty associated with our choice of mass model.

This systematic effect results in an uncertainty of ± 2 μm for each of the B_d and B_s average lifetimes, and of ± 0.001 ps^{-1} for $\Delta\Gamma$. In the $B_s \to J/\psi\phi$ fit, we observe a systematic uncertainty of ± 0.001 for $|A_0|^2$, ± 0.002 for $|A_\parallel|^2$, and ± 0.004 for $\delta_\parallel$. In the $B_d \to J/\psi K^*$ fit, we observe a systematic uncertainty of ± 0.001 for $|A_0|^2$, ± 0.001 on $|A_\parallel|^2$, ± 0.001 for $\delta_\parallel$, and ± 0.002 for $\delta_\perp$.

4.5.4 Transversity Model for the Background

In estimating the systematic uncertainty associated with our choice of background transversity model, we follow the same logic used to estimate the uncertainty associated with the mass model. We generate pseudo-experiments using our background

transversity model, then perform fits with a model that removes the highest order component in each background transversity PDF (P^θ, P^ϕ, P^ψ). This is considered an upper limit on the effect of higher order terms not included in our likelihood fit. We report on the effect on the measured parameters.

In the $B_s \to J/\psi\phi$ fit, we observe a systematic uncertainty of ± 2 μm for the average lifetime $c\tau$, ± 0.004 ps^{-1} for $\Delta\Gamma$, ± 0.006 for $|A_0|^2$, ± 0.002 for $|A_\parallel|^2$, and ± 0.072 for $\delta_\parallel$. In the $B_d \to J/\psi K^*$ fit, we observe a systematic uncertainty of ± 2 μm for the average lifetime $c\tau$, ± 0.004 for $|A_0|^2$, and ± 0.006 for $|A_\parallel|^2$.

4.5.5 Model for the Proper Decay Length Resolution

Our model for the ct resolution is a single Gaussian whose width is primarily determined by the peak in the ct distribution associated with prompt J/ψ production. Dedicated studies indicate that a better estimate of the detector smearing in ct is obtained by a 3-Gaussian model [104]. We generate pseudo experiments using a 3-Gaussian resolution model, and perform a likelihood fit in each case using a single Gaussian model. The parameters of the 3-Gaussian model are taken from a fit on the data sample. We use the difference in measured parameters as an estimate of the systematic shift associated with our choice of resolution model.

In the $B_s \to J/\psi\phi$ fit, we observe a systematic uncertainty of ± 4 μm for the average lifetime $c\tau$, ± 0.004 ps^{-1} for $\Delta\Gamma$, ± 0.001 for $|A_\parallel|^2$, and ± 0.002 for $\delta_\parallel$. In the $B_d \to J/\psi K^*$ fit, we observe a systematic uncertainty of ± 2 μm for the average lifetime $c\tau$, ± 0.001 for $|A_\parallel|^2$, ± 0.004 for $\delta_\parallel$, and ± 0.001 for $\delta_\perp$. This effect is the largest contributor to the overall systematic uncertainty on the B_s lifetime.

	$c\tau[\mu\text{m}]$	$\Delta\Gamma[\text{ps}^{-1}]$	$\lvert A_0\rvert^2$	$\lvert A_\parallel\rvert^2$	$\delta_\parallel$
SVX	± 1	± 0.002	—	—	± 0.001
Crossfeed	-1	± 0.004	± 0.003	± 0.003	± 0.070
Mass Model	± 2	± 0.001	± 0.001	± 0.002	± 0.004
Angular Model	± 2	± 0.004	± 0.006	± 0.002	± 0.072
ct Resolution	± 4	± 0.004	—	± 0.001	± 0.002
Total	$^{+5.0}_{-5.1}$	± 0.007	± 0.007	± 0.004	± 0.101

Table 4.8: Summary of systematic uncertainties assigned to the quantities measured in the $B_s \to J/\psi\phi$ likelihood fit.

4.5.6 Impact of Systematic Effects

The systematic uncertainties listed in Tables 4.8 and 4.9 are all smaller than the respective statistical uncertainties for the signal parameters. Several improvements contribute to the reduction of systematic effects compared to the 2004 measurement of $\Delta\Gamma$ at CDF [42]. The SVX alignment has been studied extensively, and its position is better known as a result of dedicated analyses in B physics and other domains. The

	$c\tau[\mu\text{m}]$	$\lvert A_0\rvert^2$	$\lvert A_\parallel\rvert^2$	$\delta_\parallel$	$\delta_\perp$
SVX	± 4	—	—	± 0.001	± 0.001
Mass Model	± 2	± 0.001	± 0.001	± 0.001	± 0.002
Angular Model	± 2	± 0.004	± 0.006	—	—
ct Resolution	± 2	—	± 0.001	± 0.004	± 0.001
Total	± 5.3	± 0.004	± 0.006	± 0.004	± 0.002

Table 4.9: Summary of systematic uncertainties assigned to the quantities measured in the $B_d \to J/\psi K^*$ likelihood fit

B_d crossfeed is reduced in part because of the use of particle identification as part of the neural network selection algorithm. The angular model for the background component is robust due to the large sample of clean background events.

This bodes well for future updates to the measurements presented here. Future measurements of these parameters performed at CDF (and D$\emptyset$) will include a larger data sample, with a larger number of signal events, which will further reduce the statistical uncertainty on the measured parameters. The systematic uncertainties are not expected to be a limiting factor to the precision of these measurements in the operating lifetime of the Tevatron experiments.

Chapter 5

Flavor Tagging

The analysis described up to this point has not made use of information concerning the production flavor of the B mesons under study. We have been able to experimentally determine several physical quantities whose measurement does not require this knowledge: mass, lifetime, lifetime difference, angular amplitudes and phases, and several associated experimental parameters. In order to measure the CP violating phase in the $B_s \to J/\psi\phi$ decay mode, information about the B_s production flavor, i.e. whether it hadronized as a B_s or $\overline{B_s}$, adds the crucial element. The determination of the production flavor of B mesons is referred to as flavor tagging, and the individual algorithms are called flavor taggers. Determining the flavor of a candidate means that we establish whether it is matter or antimatter, and does not refer to the species of B meson (B_s, B_d, B_u). The flavor tagging algorithms described here all provide information about the flavor of B mesons at the moment of hadronization. Determining the flavor at the time of decay is done by examining the charge of the decay products for decay modes which proceed with different charge for different decay flavor, e.g. $B_s \to D_s^- \pi^+$ with $D_s^- \to K^+K^-\pi^-$. Such a decay mode is referred to as *self-tagging*. In the case of $B_s \to J/\psi\phi$, information about the decay flavor is not available because J/ψ and ϕ are both C-odd states. Such information is not needed in any case, and we focus on the description of flavor tagging at production time.

Flavor tagging techniques, albeit essential to the measurement, were not developed by the author. This chapter should not be considered as an exhaustive description of the development of flavor tagging techniques, a topic that is itself the subject of a multitude of dissertations, journal articles, and internal CDF documents. In this discussion, we describe the techniques used for flavor tagging, and the calibrations required to use them in our analysis sample. The likelihood expression is modified in order to include this additional information.

5.1 Concepts

Given a B meson candidate, we use our knowledge of the $p\overline{p} \to \overline{b}b$ process, as well as knowledge about the hadronization and fragmentation processes to deduce the production flavor of the B candidate. Two categories of flavor taggers exist. *Same*

side tagging (SST) examines the fragmentation tracks of the B meson candidate selected by our trigger. *Opposite side tagging* (OST) relies on information from the other member of the $\bar{b}b$ pair and its decay products. Figure 5-1 shows a sketch of two $\bar{b}b$ events, and the associated B decays as an example of same and opposite side information available for flavor tagging. This sketch provides visual guide to the physical processes used by the taggers described in this chapter: the fragmentation track used by the same side tagger, and the lepton and jet charge used by the opposite side tagger.

We begin by describing the formalism common to all taggers. As in the case of detecting any physical process in a particle collider experiment, the detection of the physical processes used to determine the B production flavor is not a perfect procedure. The event information used for tag decisions, for instance well detected

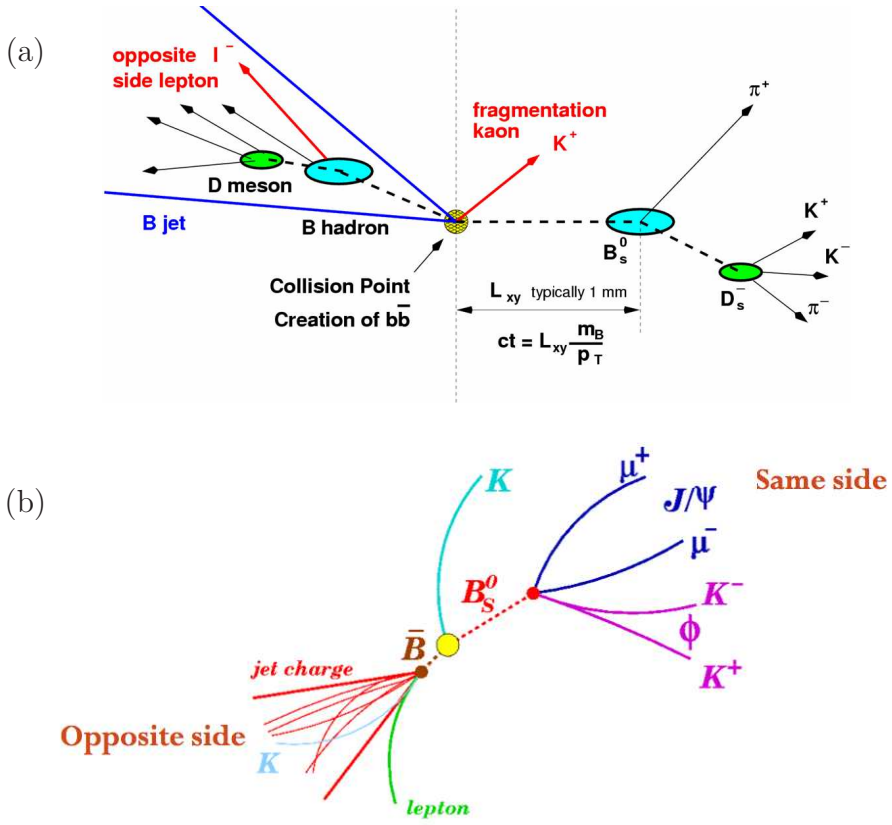


Figure 5-1: Sketch of same and opposite side tagging information in the case of (a) the hadronic decay $B_s \to D_s^-[K^+K^-\pi^-]\ \pi^+$, and (b) the decay under study in this dissertation, $B_s \to J/\psi[\mu^+\mu^-]\ \phi[K^+K^-]$. The two figures show that flavor tagging makes use of the same information for both types of decays: we use the fragmentation track (here a kaon) in the same side tagger, and the jet charge and lepton from the semileptonic B decay in the opposite side tagger.

opposite side decays used by the OST, is not always available. As a result, a tagger is characterized by an efficiency ϵ that is defined as the fraction of times that a tagger makes a tag decision:

$$\epsilon \equiv \frac{N_{\text{tagged}}}{N_{\text{untagged}} + N_{\text{tagged}}}, \tag{5.1}$$

where N_{tagged} is the number of B candidates for which a tag decision is available and N_{untagged} is the number of B candidates for which no such decision is available. A reasonably accurate estimate of the efficiency is determined in data by simple counting. More elaborate determinations are made through fits to all data using the efficiency as a fit parameter.

In addition, the decision made by a tagging algorithm is at times wrong. The probability of giving a wrong tag decision is denoted by p_w. For convenience in several expressions, we use the variable dilution ($\mathcal{D}$), defined as

$$\mathcal{D} \equiv 1 - 2p_w. \tag{5.2}$$

Although the term dilution might be interpreted as having a negative connotation from its uses in other contexts, high dilution is in fact what we desire in a tagger. A hypothetical perfect tagger that always makes the correct decisions has dilution of one corresponding to a probability of being wrong $p_w = 0$. A tagger that makes exactly random decisions has $p_w = 0.5$, and therefore a dilution equal to zero.

While this dissertation is not devoted to the observation and measurement of the frequency of B_s oscillations, it is illustrative to examine the effect of flavor tagger characteristics on such a measurement. We have described in Section 1.3.1 the oscillatory behavior of B_s eigenstates. The B_s meson, after creation as a $|\overline{b}s\rangle$ ($|b\overline{s}\rangle$) state, undergoes oscillations such that the probability of observing it in the antimatter conjugate or the same initial state is time dependent and sinusoidal. Putting aside other variables, we express the probability of observing one of the flavor eigenstates as a function of proper time t as:

$$\mathcal{P}(t)_{\substack{\text{unmixed}\\\text{mixed}}} = \frac{1}{2}[1 \pm \cos\omega t], \tag{5.3}$$

where ω is the oscillation frequency Δm_s, and the $+$ $(-)$ sign refers to observing the same (conjugate) state at time t as was created at time 0. Expressing the tag decision as ξ, where $\xi = +1$ denotes no mixing, $\xi = -1$ mixing, and $\xi = 0$ no tag decision, we can write the probability for a flavor change to be observed or not as

$$\mathcal{P}(\xi|t) = \frac{1}{2}[1 + \xi\cos\omega t], \tag{5.4}$$

where we are assuming perfect tagging performance with $\epsilon = 1$ and $\mathcal{D} = 1$. Adding the effect of dilution $\mathcal{D} < 1$, we obtain:

$$\mathcal{P}(\xi|t,\mathcal{D}) = \frac{1+\xi\cos\omega t}{2}\frac{1+\mathcal{D}}{2} + \frac{1-\xi\cos\omega t}{2}\frac{1-\mathcal{D}}{2} = \frac{1+\xi\mathcal{D}\cos\omega t}{2}. \tag{5.5}$$

We note that $\mathcal{D}$ acts as a damping term on the sinusoidal oscillation, which decreases the sensitivity to observation of the full time evolution by reducing the amplitude of the flavor oscillation signal.

The probability of making a wrong tag decision is dependent on several event variables, and varies from one event to the next. In order to assign a larger (smaller) weight to events for which there is higher (lower) confidence in the tag decision, the dilution for each tagger is parametrized as a function of several event variables, and predicted dilution is assigned to each event as a result. This assignment of predicted dilution has two main beneficial effects: it enhances the overall tagger performance, and it allows us to transport tagger calibrations from one sample to another. The details of these parametrizations are given in the discussion of individual taggers.

The significance of the B_s oscillation signal is proportional to $\sqrt{\epsilon\mathcal{D}^2}$, which justifies the use of $\epsilon\mathcal{D}^2$ as the figure of merit for quantifying tagging power. When quoting the tagging power with a sample of events which have been assigned individual predicted dilution, we calculate the tagging power as $\epsilon\langle\mathcal{D}^2\rangle$, where the average is taken over the ensemble of selected events.

5.2 Opposite Side Tagging

Given a reconstructed B meson hadronized from one of the b quarks in a $\bar{b}b$ pair, opposite side tagging algorithms infer the meson's initial production flavor by examining the hadronization products of the other b quark in the pair. In practice, this means determining the charge of leptons coming from semileptonic B decays or of b jets on the opposite side. The hadronization and fragmentation processes of the b and $\bar{b}$ are independent of one another, and so the opposite side tag decisions uses information that is itself independent of what happens on the trigger side. The main advantage of such a method is that the tagger can be developed using light B data samples, and then applied directly to B_s tagging. The light B data samples are already well understood from precious measurements. Furthermore, the light B mesons are more copiously produced than the B_s, by several orders of magnitude.

Several factors complicate the task of opposite side tagging. In the case of determining the jet charge, the b tracks must be identified among several other tracks in the event, and there is probability that the b jet is not within the fiducial acceptance of the detector. The time-dependent oscillations of B_d mesons on the opposite side must also be accounted for, lest we infer the wrong charge for the B meson at production time.

The ℓ+SVT trigger event sample is used for the development of all opposite side taggers. The parametrization of the predicted dilution as a function of several variables is facilitated by the large sample size of the data set from this trigger. The parametrization is applied to the di-muon trigger sample after the application of a global dilution scale factor, which is used to account for the kinematic differences between events selected by these two triggers. The calibration of the opposite side taggers for use in the di-muon sample is done on B_d and B_u data in a simultaneous

fit of the masses, lifetimes, tagging efficiencies, and scale factors.

The development of both the opposite side and same side taggers is done within a cone of $\Delta R = 0.7$, ensuring isolation of the two sets of algorithms and independence of the tagger decisions for easier combination of the two.

5.2.1 Lepton Taggers

Lepton taggers infer the production flavor of the b quark using the semileptonic decay of the opposite side b through the transitions $b \to cl^-\overline{\nu_l}X$ and $\overline{b} \to \overline{c}l^-\nu_l X$. A positive lepton (muon or electron) indicates a $\overline{b}$, while a negative lepton indicates a b quark. Two algorithms are developed for tagging using these semileptonic decays, one for each lepton: the soft muon tagger (SMT), and the soft electron tagger (SET). The development of the muon tagger is discussed in detail in Reference [105], and the electron tagging framework is explored in Reference [106].

Several considerations must be taken into account regarding the efficiency for and accuracy of a tag decision from the lepton taggers. The comparatively low semileptonic branching fraction $\mathcal{B}(B \to lX) \sim 20\%$ means that the efficiency is lower for lepton taggers than for jet charge taggers. The efficiency is further lowered by the fact that the lepton may not be within the fiducial volume of the tag acceptance. The accuracy of the tag decision is affected by several factors. In the event that the semileptonic decay comes from a B_d, the latter has a probability of having oscillated to the antimatter conjugate state before decay. In the case of a semileptonic B_s decay, the fast oscillations render the lepton tag from such a decay unusable. If the lepton comes from a $b \to c \to lX$ transition, the inferred flavor will be opposite to the true flavor state. In addition, the lepton itself may be misidentified. Nevertheless, the good purity of lepton identification at CDF means that dilution is comparatively high for lepton taggers.

Lepton identification is done using likelihood discriminants that combine information from several detector systems, formed to have a value near 1 for real leptons and near 0 for other particles misidentified as leptons. The muon likelihood is described in Appendix C and is also used as a neural network selection variable in this analysis (Section 3.3.2). The electron likelihood combines information from the EM and hadronic calorimeters, including the showermax and preshower detectors, with energy loss in the tracking system. Several selection criteria are applied to lepton candidates before use in the lepton taggers. These criteria include track quality requirements: a minimum of 10 axial and 10 stereo COT hits and at least 2 $r-\phi$ hits in the SVX for electrons. Lepton likelihood requirements are chosen from examining clean lepton samples from $J/\psi \to \mu^+\mu^-$ and $\gamma \to e^+e^-$. In addition, minimum p_T requirements of 1.5 and 2.0 GeV/c are used for electrons and muons respectively. The impact parameter $|d_0|$ is required to be smaller than 0.3 cm in order to reject muons coming from hadron decays.

The performance of the lepton taggers shows dependencies that are used to parametrize the dilution as a function of two variables: the lepton likelihood and the transverse momentum of the lepton with respect to the axis of the jet in which the former was found, denoted by p_T^{rel}. The dilution increases with higher lepton likelihood, as

expected since a higher purity of the lepton sample should yield better b flavor determination. The performance also increases with higher $p_T^{\rm rel}$. This is due to the fact that leptons coming from b transitions tend to have a larger spread of momentum values in the plane transverse to the b jet than in jets coming from lighter quarks because of the larger available phase space. The dependence on $p_T^{\rm rel}$ is described by:

$$\mathcal{D}(p_T^{\rm rel}) = A_\ell \left(1 - e^{p_T^{\rm rel}+\delta_\ell}\right), \tag{5.6}$$

where A_ℓ and δ_ℓ are determined empirically. The muon and electron taggers are treated separately in this procedure.

5.2.2 Jet Charge Taggers

Jet charge tagging makes use of the fact that a b quark has a charge of $-1/3$, while a $\overline{b}$ has a charge of $+1/3$. Jet charge taggers exploit this fact by examining the jets on the opposite side to infer the charge, and thus the flavor, of the b quark on the trigger-side. The development of jet charge taggers at CDF is documented in References [107] and [108].

These taggers attempt to find the jet that originates from the b quark on the opposite side, because we expect the charge of the b to be given on average by the momentum-weighted sum of the charges of tracks within the jet.

This process begins by the reconstruction of jets on the opposite side using an isolation cone of $\Delta R < 0.7$ with respect to the trigger B momentum. Within this cone, tracks which were used for the reconstruction of the trigger-side B are excluded. Selection requirements on $|d_0|$, z_0 and p_T are imposed on the tracks to be included in jet reconstruction. The jets are reconstructed using a track based cone clustering algorithm, merging tracks iteratively starting from seed tracks that must have $p_T > 1.0$ GeV/c.

Artificial neural networks (NN) are used to identify from the set of reconstructed jets the jet with the highest probability to originate from a b quark. The neural networks are trained and optimized on a large sample of simulated events which include fragmentation tracks in addition to the main B decay. The identification of the highest probability b jet is done in two steps. A first NN, using as input several kinematic variables, is used to assign to each track a probability $P_{\rm trk}$ of having originated from the decay of a B meson. A second NN assigns to each jet a probability $P_{\rm jet}$ of having come from a b quark, using kinematic variables as well as the probability $P_{\rm trk}$ for each track. Once the highest probability b jet is found, the jet charge is computed as

$$Q_{\rm jet} \equiv \frac{\sum_i Q^i \, p_T^i \, (1 + P_{\rm trk}^i)}{\sum_i p_T^i \, (1 + P_{\rm trk}^i)}, \tag{5.7}$$

where the summation index runs over all the tracks of the chosen jet, and Q^i, p_T^i, and $P_{\rm trk}^i$ are respectively the charge, transverse momentum and probability of having originated from a B decay for the $i^{\rm th}$ track.

In order to take advantage of the greater statistical power available from jets with

different characteristics, this procedure is optimized for three exclusive classes of jets. Class 1 jets contain an identified secondary vertex and a decay length significance $L_{xy}/\sigma_{L_{xy}} > 3$. Class 2 jets do not fulfill the criteria for Class 1 and have at least one track with $P_{\text{trk}} > 0.5$. Class 3 jets are the remaining reconstructed jets that pass jet selection criteria.

The dilution for jet charge tagging is parametrized as a function of the quantity $|Q_{\text{jet}}| \cdot P_{\text{jet}}$. The parametrization is done separately for the three jet classes. The dependence is observed to be linear in all cases.

5.2.3 Combination of Opposite Side Taggers

The separate OST algorithms make use of some common events attributes in order to make a tag decision, which means that their tag decisions are not necessarily independent. Combining the separate algorithms into one simplifies the task of handling multiple taggers. Furthermore, an optimal combination reduces the number of parameters in the ultimate likelihood fit used for our *CP* violation measurement, which decreases computation time and reduces the number of correlated statistical uncertainties. Two combination strategies are considered, based on previous work developed for the measurements of Δm and Δm_s oscillation frequencies. A hierarchical exclusive combination or all OST algorithms akin to the one already used to choose among the three jet charge taggers [109], and a neural network combination of the tag decisions [110].

The neural network combination of opposite side taggers exploits the correlations among their tag decisions by design. The neural network is trained on $\ell+$ SVT data, using as input the tag decisions, and predicted dilutions of the separate taggers as well as several kinematic quantities of the event.

The hierarchical combination uses an exclusive decision algorithm, ordering the taggers in the hierarchy based on their respective tagging power. A tag decision are taken from the highest ranked OST algorithm if it is available; otherwise the decision is taken from the next highest ranked tagger, and so on until we reach the bottom of the decision ladder. The taggers are ranked as follows, from first to last choice

- Soft Muon Tagger,
- Soft Electron Tagger,
- Jet Charge Taggers.

In choosing the method for combination of the OST algorithms, there are additional considerations beside the tagging power. In the measurement of *CP* violation in $B_s \to J/\psi\phi$, it is desirable to minimize any tagging asymmetry effects between matter and antimatter. The presence of such an asymmetry complicates the task of measuring the inherent and physical *CP* asymmetry in the decay. Furthermore, our calibration process makes a fundamental assumption when correcting for kinematics differences between the development sample of the tagger, i.e. the ℓ+SVT sample, and the analysis sample, i.e. the di-muon sample. We make the assumption that a sufficient calibration of the taggers is provided by the addition of a dilution scale

factor such that $\mathcal{D}' = S_{\mathcal{D}} \cdot \mathcal{D}$. This linear relation must hold in order to be able to properly perform the calibration. When the parametrization of the dilution is a good description for our data sample, we expect the scale factor $S_{\mathcal{D}}$ to be consistent with unity, as measured by the likelihood fit.

We find that the linear relation $\mathcal{D}' = S_{\mathcal{D}} \cdot \mathcal{D}$ does not hold well separately for the case of B^+ and B^- in the case of the neural network combined OST. This effect is not fully understood and is still under investigation at the time of writing. In the case of the hierarchically combined OST, we find that there is a better match to expectation. Figure 5-2 shows the effect for both combinations. Although we suffer a loss tagging power, we therefore choose to use the hierarchical combination.

5.2.4 OST Calibration in Di-Muon Data

In order to use the opposite side tagging algorithms in the analysis data set, we need to correct for effects arising from the difference between the kinematics of events accepted by the ℓ+SVT trigger and those accepted by the di-muon triggers. This calibration is performed on data using B_u and B_d samples, making use of the fact that OST information can be transferred from a sample with one B meson species to samples with different species due to the independence of the algorithms from the trigger side hadronization and fragmentation. The correction consists of multiplying the predicted dilution for each event by a scale factor:

$$\mathcal{D}_j \rightarrow S_{\mathcal{D}} \cdot \mathcal{D}_j \tag{5.8}$$

The dilution scale factor is extracted from a simultaneous fit of of the $B_u \rightarrow J/\psi K^+$ and $B_d \rightarrow J/\psi K^*$ samples. We start with mass and lifetime fits, and then perform fits of the scale factors and Δm_d. In the simultaneous fit, we use a single efficiency ϵ and a single dilution scale factor $S_{\mathcal{D}}$ for both data samples. The results of this fit are summarized in Table 5.1. Mass and lifetime projections of the fit are presented in Figure 5-3.

5.3 Same Side Tagging

Same side tagging exploits the correlation between the b flavor of the trigger side B meson and the charge of associated tracks. Different fragmentation tracks are expected for different B mesons. In this case, we mean different species *and* matter/antimatter content. When a b quark hadronizes, quark pairs must be pulled from the vacuum in order to form hadrons. The fact that hadronization creates quark *pairs* forms the correlation between the b flavor and the associated leading track. As shown in Figure 5-4, B_d mesons are often produced in conjunction with π^- from the formation of a $\overline{d}d$ pair, B^- with π^+ from a $\overline{u}u$, and B_s with K^- from an $\overline{s}s$. We discuss here the Same Side Kaon Tagger (SSKT) for use in distinguishing B_s from $\overline{B_s}$ at production time. Reference [111] provides a detailed treatment of the development of the SSKT and its performance in fully reconstructed hadronic B_s decays.

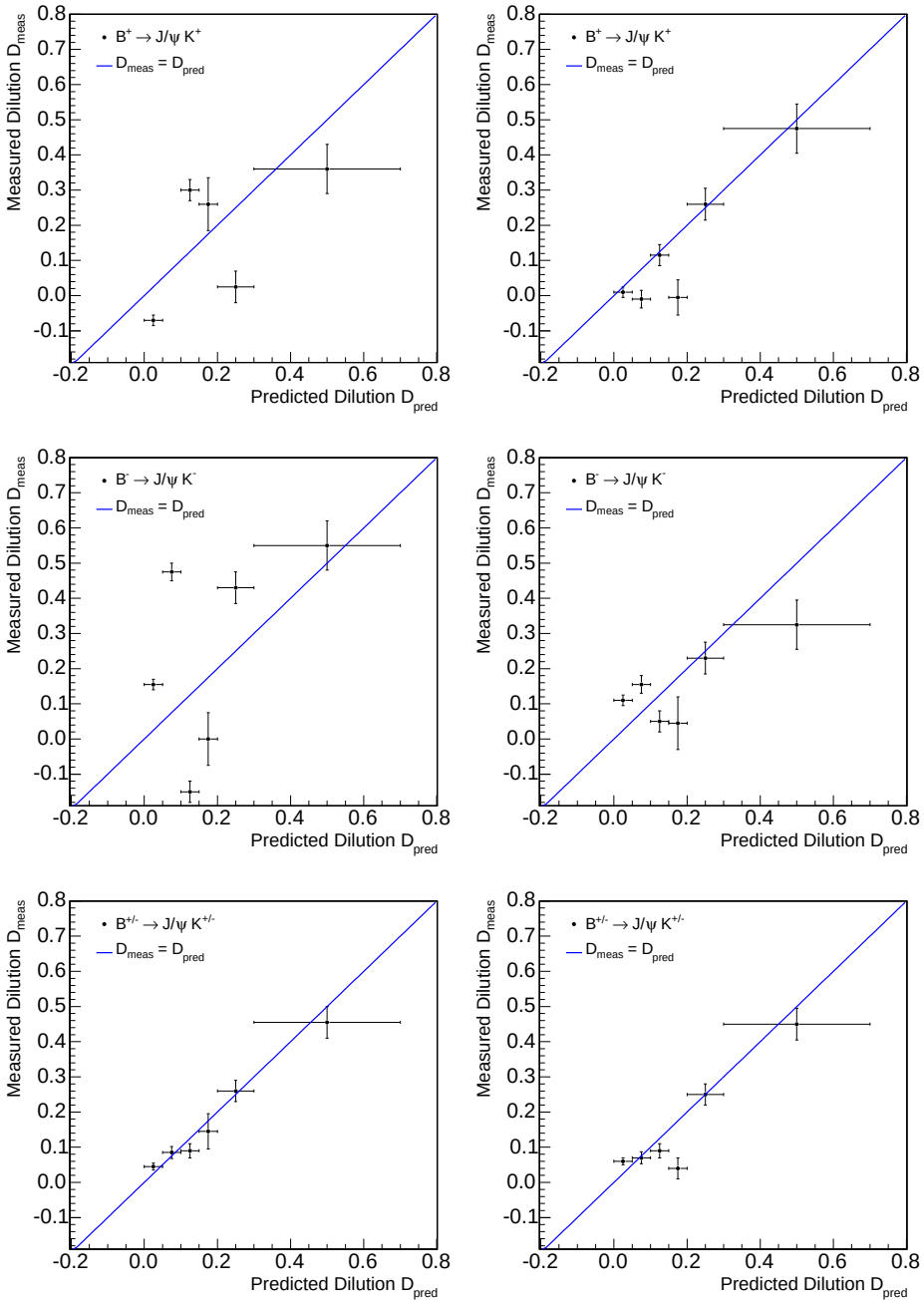


Figure 5-2: Measured versus predicted dilution in the case of the neural network combined OST (left) and hierarchically combined OST (right), for B^+, B^- and both.

The fact that we examine the fragmentation tracks of the trigger side B for same side tagging presents a problem that motivates a fundamental difference in strategy. The information used by the tagger is no longer independent of the specific B meson on the trigger side, and thus we cannot develop and calibrate the algorithms for use in B_s tagging using B_d and B_u sets from the large ℓ+SVT data sample. The

Parameter	Fit Result	World Avg. (2006)	Unit
M_{B_u}	5278.6 ± 0.1	5279.0 ± 0.5	MeV/c^2
$c\tau_{B_u}$	486.6 ± 4.7	491.1 ± 3.3	μm
M_{B_d}	5279.3 ± 0.2	5279.4 ± 0.7	MeV/c^2
$c\tau_{B_d}$	455.9 ± 6.0	458.7 ± 2.7	μm
Δm_d	0.44 ± 0.07	0.507 ± 0.005	ps^{-1}
ϵ	0.9589 ± 0.0015		
$S_{\mathcal{D}}$	0.853 ± 0.055		

Table 5.1: Mass, lifetime, and tagger performance results of the simultaneous likelihood fit of the B_u and B_d data samples using the combined hierarchical OST. We include for comparison the world average values of selected parameters prior to the publication of our results [41]. We find that our fit results are compatible with these world averages. Note that our fit results include only statistical uncertainties, while the world averages incorporate statistical and systematic uncertainties.

performance of such an algorithm is expected to differ depending on the B meson species. Instead, the same side tagging algorithms are developed in large samples of fully reconstructed decays for the B_u and B_d samples. The fast oscillations of the B_s forbid accurate assessment of the SST algorithm in data. Therefore, the SST is developed for the B_s in Monte Carlo samples in which we simulated the full $p\overline{p}$ event include both the outgoing $\overline{b}b$ and fragmentation tracks. A data – Monte Carlo comparison for relevant variables samples is used to ensure the correctness of the fragmentation model and its applicability to the identification of $B_s/\overline{B_s}$ mesons.

Preselection requirements are applied on tracks to be used for the SST. Each track must have $p_T > 0.4$ GeV/c, because the track reconstruction efficiency displays a charge asymmetry for tracks below this threshold. The track must be within an angular cone of $\Delta R < 0.7$ centered around the B momentum. Tracks used for the reconstruction of the B candidate are explicitly removed from consideration. A maximum impact parameter significance of $d_0/\sigma_{d_0} < 4$ is required to ensure that we select fragmentation tracks from the $p\overline{p}$ collision and not tracks from a displaced vertex. A data – Monte Carlo comparison is performed for the sample of preselected tracks, which includes the following variables: the B transverse momentum, the ΔR between the track and the B momentum, the multiplicity of tag candidates, the combined kaon identification discriminant $CLL(K)$. Comparisons are also carried out for the kinematic variables of the track p_T, $p_T^{\rm rel}$, and $p_L^{\rm rel}$, which are defined in Figure 5-5. Good agreement is found for samples of B_s, B_d, and B_u decays.

The preselection requirements sometimes leave more than one track as a tagging candidate. In the case that the tag decision does not match for all tracks, which is true for $\sim$ 65 % of events with multiple tracks, a decision algorithm must be developed to choose among them. This algorithm must provide a predicted dilution for the track. The final selection of tracks for same side kaon tagging is done using a neural network [112]. The neural network combines particle identification information from

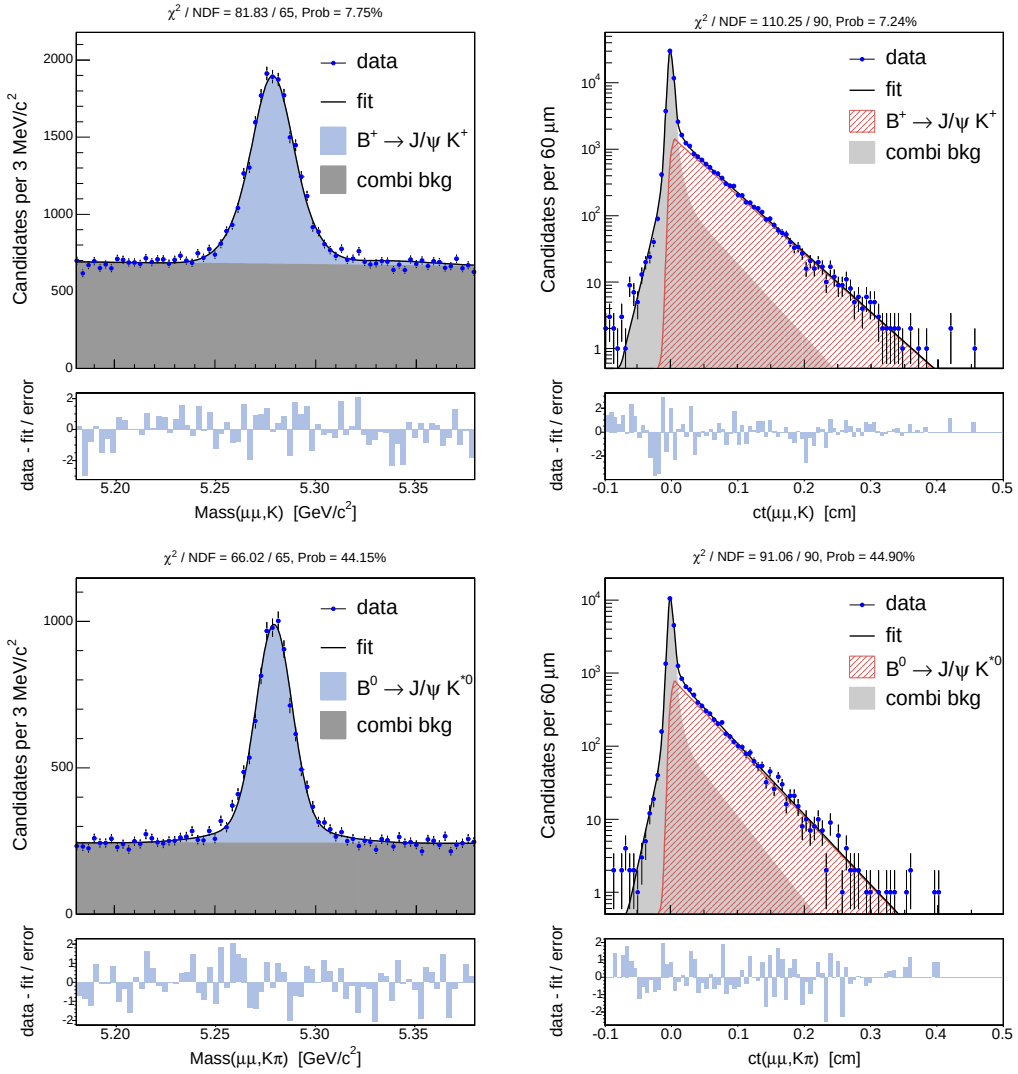


Figure 5-3: Mass and Lifetime projections for B_u (top) and B_d (bottom) from the fits used to calibrate the OST on di-muon data. We find that the fit projections indicate good agreement between the fit results and the data.

the Time-of-Flight and the $\mathrm{d}E/\mathrm{d}x$ in the COT, encapsulated into $CLL(K)$ with kinematic variables p_T, p_T^{rel}, and p_L^{rel} and ΔR. The neural network is trained with samples of pure B_s Monte Carlo events generated with `Pythia`. The signal training sample is comprised of kaons with the correct charge correlation, and the background training sample is comprised of tracks from other particle species as well as tracks misidentified as kaons. The track with the highest value of neural network output, i.e. most *signal-like*, is chosen as the tagging track.

The dilution in the SSKT is parametrized as a function of the neural network output value n using an empirically chosen cubic form, reflecting the observed dilution

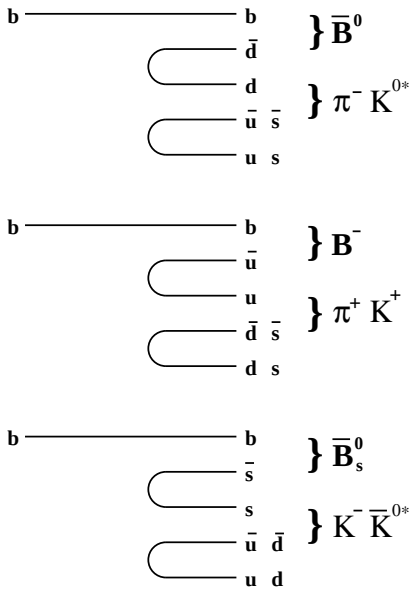


Figure 5-4: Different B mesons and their associated leading fragmentation tracks

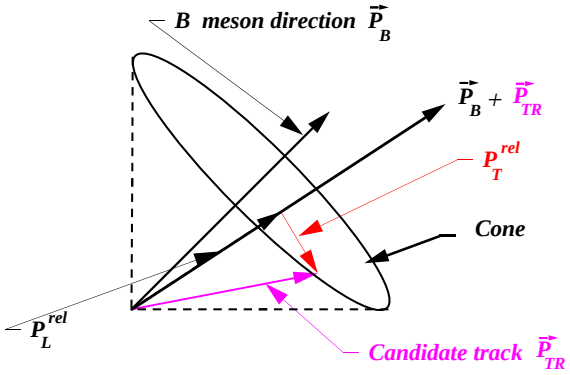


Figure 5-5: Diagram showing the definitions of the quantities $p_T^{\rm rel}$ and $p_L^{\rm rel}$ with respect to the B meson momentum $\vec{p}_B$ and candidate track momentum $\vec{p}_{\rm TR}$.)

dependence:

$$\mathcal{D}(n) = c_0 + c_1 n + c_2 n^2 + c_3 n^3. \tag{5.9}$$

SSKT Scale Factor. As in the case of the OST, we introduce a dilution scale factor in order to use same side kaon tagger on the di-muon dataset. We extract this scale factor from a modified B_s MC sample where the oscillation frequency has been set to zero. The dilution scale factor $S_{\mathcal{D}}$ is extracted from a fit of the flavor asymmetry, which in the case of a sample with $\Delta m_s = 0$ give the true sample dilution. We fit for this scale factor in subsamples reflecting three different data taking periods at CDF, and in the whole sample combined.

Performance. As a final step, we assess the performance of the SST in our $B_s \rightarrow$

$J/\psi\phi$ sample. We observe an efficiency of $\epsilon_s = 50.0\%$, average dilution $\langle\mathcal{D}\rangle = 25.4\%$, and $\epsilon_s S_{\mathcal{D}_s}\langle\mathcal{D}^2\rangle = 3.0 \pm 0.9\%$.

5.4 General Tagging Likelihood Construction

We must modify the likelihood expression in Equation 4.33 for the analysis of the $B_s \to J/\psi\phi$ decay to incorporate flavor tagging. We present the general procedure for expanding the likelihood expression to include these variables. We treat the general cases of non-oscillating and oscillating components. In this section we restrict the discussion to flavor oscillations common to all neutral mesons. This introduces the necessary concepts and the general procedure for modifying a fit likelihood to take into account flavor tagging. A comprehensive overview of the likelihood construction for tagging measurements is provided in Reference [113]. We leave to Chapter 6 the specific discussion of the likelihood expression used in the $2\beta_s$ measurement in a tagged sample of $B_s \to J/\psi\phi$ decays.

When flavor tagging is used, each event is assigned the following additional variables:

- a tag decision from each tagger (OST and SST): ξ_o, ξ_s, where $\xi = 1$ corresponds to a B_s at production time, $\xi = -1$ corresponds to a $\overline{B_s}$, and $\xi = 0$ means that no tag decision was made,
- a predicted dilution from each tagger: $\mathcal{D}_o$ and $\mathcal{D}_s$, indicating the degree of confidence in the tag decision.

In the following sections we detail the expansion of the following likelihood that depends on mass (via P_m) and proper decay length (via P_{ct} and $P_{\sigma_{ct}}$):

$$\mathcal{L} = P_m \cdot P_{ct} \cdot P_{\sigma_{ct}} \tag{5.10}$$

to incorporate the variables ξ_o, ξ_s, $\mathcal{D}_o$ and $\mathcal{D}_s$. We begin the expansion using one tagger, and then detail the changes needed to combine independent tagger decisions.

5.4.1 Dilution Probability Density Function

The first necessary addition to Equation 5.10 that we introduce is the PDF for the predicted dilution. A normalized joint probability density function that takes into account tagging variables must include a PDF for the dilution $P(\mathcal{D})$ because the latter is dependent on the sample component. This portion of the likelihood is essentially analogous to the PDF for the ct uncertainty described in Section 4.1.4, and differs for the signal component and the background component. Just as in the untagged lifetime measurement, we deal with terms of the form $P(ct|\sigma_{ct}) \cdot P(\sigma_{ct})$, in this case we deal with terms of the form $P(ct|\mathcal{D}) \cdot P(\mathcal{D})$. We implement $P(\mathcal{D})$ as normalized histograms extracted from the data distributions. For the background component, we use upper and lower sidebands. For the signal component, we use the sideband-subtracted central region. The distributions of predicted dilutions for OST and SST are shown in Figure 5-6

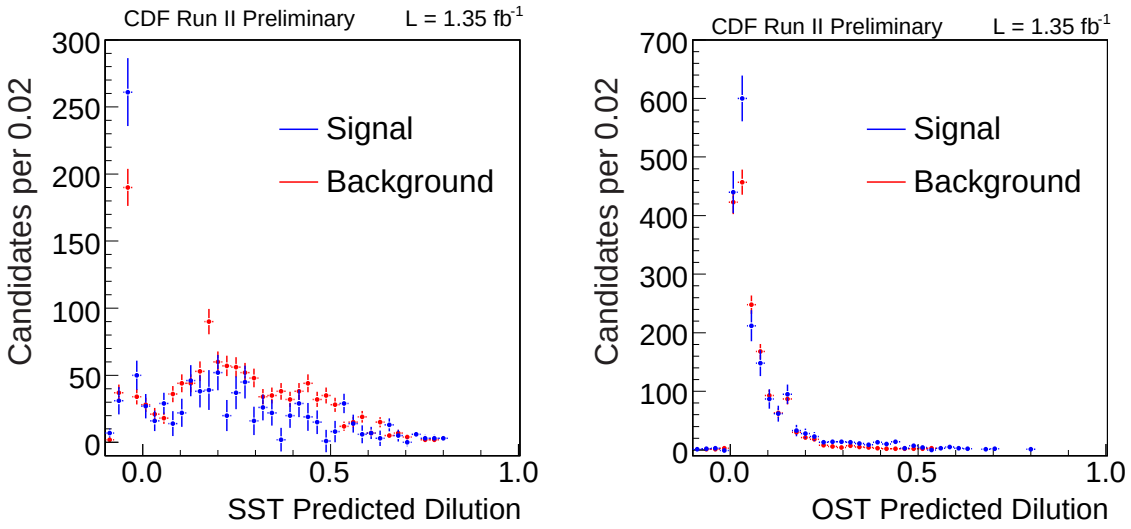


Figure 5-6: Predicted dilution distributions for opposite side and same side taggers in the $B_s \to J/\psi\phi$ sample. We obtain the PDF $P(\mathcal{D})$ by normalizing these distributions.

5.4.2 Construction with One Tagger

We expand the likelihood to incorporate a decision and predicted dilution from one tagger. For a given sample component, i.e. either signal or background, the likelihood 5.10 is now expanded to

$$\mathcal{L}(m, ct, \sigma_{ct}, \xi, \mathcal{D}) = P_m \cdot P_{ct} \cdot P_\xi \cdot P_\mathcal{D} \cdot P_{\sigma_{ct}}, \tag{5.11}$$

The mass portion of this likelihood remains decoupled from the tag decisions, as it is decoupled from the proper decay length portion. The remaining PDFs are rewritten to reflect the relationship between tag decisions, dilution, and the probability density in t:

$$\mathcal{L}(m, ct, \sigma_{ct}, \xi, \mathcal{D}) = P(m) \cdot P(ct|\xi, \mathcal{D}, \sigma_{ct}) \cdot P(\xi) \cdot P(\mathcal{D}) \cdot P(\sigma_{ct}). \tag{5.12}$$

The mass probability density $P(m)$ is familiar from Chapter 4, and corresponds to X_{sig} (X_{bkg}) for the signal (background) component. The probability density $P(\sigma_{ct})$ is calculated from the mass sideband-subtracted and sideband distributions of σ_{ct}, as previously described, and likewise for $P(\mathcal{D})$ using the distribution of predicted dilutions. The distribution $P(ct|\xi, \mathcal{D}, \sigma_{ct})$ is the probability density in t given $\xi, \mathcal{D}, \sigma_{ct}$, and varies depending on the component. Examples are given below for an oscillating signal and non-oscillating background. Finally, $P(\xi)$ is the probability of obtaining a tag decision for a given event, and depend uniquely on the efficiency of the tagger. For a single tagger, $P(\xi)$ is given by:

$$P(\xi) = (1 - \epsilon) \cdot \delta(\xi - 0) + \epsilon \cdot \delta(|\xi| - 1), \tag{5.13}$$

where ϵ is the efficiency of the tagger. Equation 5.13 can be understood by substituting in values of ξ depending on whether or not the event is assigned a tag decision. For $\xi = \pm 1$, $P(\xi) = \epsilon$, and for $\xi = 0$, $P(\xi) = 1 - \epsilon$. In other words Equation 5.13 states that the probability of (not) being assigned a tag decision is by definition (one minus) the efficiency of the tagger.

We now discuss two specific cases of likelihood construction with one tagger: a non-oscillating background, and an oscillating signal. This amounts to describing the form of $P(ct|\xi, \mathcal{D}, \sigma_{ct})$ in each of these cases.

Non-Oscillating Background

The likelihood expression must be modified to account for the tagging decisions and dilutions assigned to all events, including background events. This is true even though tag decisions do not have the physical meaning for the background component as they do for the signal component, given that the former is largely composed of random combinations of tracks. We construct $P(ct|\xi, \mathcal{D}, \sigma_{ct})$ using a non-oscillating time distribution for the background, in this case Y_{bkg} obtained from Section 4.1.4:

$$\begin{aligned} P(ct|\xi, \mathcal{D}, \sigma_{ct}) &= \frac{1}{N}\left[\frac{1+\xi\mathcal{D}}{1+|\xi|} Y_{\text{bkg}}(ct)\right] \otimes G(ct|\sigma) \\ &= \frac{1}{N} \cdot \begin{cases} Y_{\text{bkg}}(ct) \otimes G(\sigma_{ct}), & (\xi = 0) \\ \dfrac{1+\mathcal{D}}{2} Y_{\text{bkg}}(ct) \otimes G(\sigma_{ct}), & (\xi = +1) \\ \dfrac{1-\mathcal{D}}{2} Y_{\text{bkg}}(ct) \otimes G(\sigma_{ct}), & (\xi = -1), \end{cases} \end{aligned} \tag{5.14}$$

where

$$N = \int Y_{\text{bkg}}(ct) \otimes G(\sigma_{ct})\, dt.$$

Single-Lifetime Oscillating Signal

For an oscillating signal with a single lifetime, the probability density $P(ct|\xi, \mathcal{D}, \sigma_{ct})$ is constructed by accounting for the exponential decay, the flavor oscillation, and the detector smearing in ct as follows:

$$\begin{aligned} P(ct|\xi, \mathcal{D}, \sigma_{ct}) &= \frac{1}{N}\left[\frac{1}{c\tau} e^{-ct/c\tau}\theta(ct) \frac{1+\xi\mathcal{D}\cos(\omega t)}{1+|\xi|}\right] \otimes G(\sigma). \\ &= \frac{1}{N} \cdot \begin{cases} \dfrac{1}{c\tau} e^{-\frac{ct}{c\tau}}\theta(ct) \otimes G(\sigma_{ct}), & (\xi = 0) \\ \dfrac{1+\mathcal{D}\cos\omega t}{2} \dfrac{1}{c\tau} e^{-\frac{ct}{c\tau}}\theta(ct) \otimes G(\sigma_{ct}), & (\xi = +1) \\ \dfrac{1-\mathcal{D}\cos\omega t}{2} \dfrac{1}{c\tau} e^{-\frac{ct}{c\tau}}\theta(ct) \otimes G(\sigma_{ct}), & (\xi = -1). \end{cases} \end{aligned} \tag{5.15}$$

The oscillating term differs from Equation 5.1 in that we account for the possibility of no tag decision being taken ($\xi = 0$). When this occurs, Equation 5.15 reduces to the case of single lifetime exponential described in Section 4.1.4. We have used the following definitions:

$$N = \int \frac{1}{c\tau} e^{-\frac{ct}{c\tau}} \theta(ct) \otimes G(\sigma_{ct})\, dt,$$

$$G(\sigma_{ct}) \equiv G(ct^{'}; \sigma_{ct}) = \frac{1}{\sqrt{2\pi}\sigma_{ct}} e^{-\frac{(ct-ct^{'})^2}{2\sigma_{ct}^2}},$$

$$f(ct) \cdot g(ct) \otimes G(\sigma) \equiv \left(\int_{-\infty}^{\infty} dt'\, f(ct^{'}) g(ct^{'}) G(ct - ct^{'}; \sigma) \right).$$

5.4.3 Construction with Two Taggers

With two taggers, Equation (5.13) changes to accommodate independent decisions by the various taggers, each with a decision ξ_j and a dilution $\mathcal{D}_j$. We use the subscripts o and s in reference to the *opposite* and *same-side* taggers. The general expression is written for the probability of obtaining a vector of tag decisions $\vec{\xi}$:

$$P(\vec{\xi}) = (1 - \sum_{j=1}^{2} \epsilon_j) \cdot \delta(\sum_{j=1}^{2} \xi^j - 0) + \sum_{j=1}^{2} \epsilon_j \cdot \delta(|\xi^j| - 1)$$

$$= \begin{cases} (1-\epsilon_s)(1-\epsilon_o) & (\xi_s = 0, \xi_o = 0) \\ \epsilon_s(1-\epsilon_o) & (\xi_s = \pm 1, \xi_o = 0) \\ (1-\epsilon_s)\epsilon_o & (\xi_s = 0, \xi_o = \pm 1) \\ \epsilon_s \epsilon_o & (\xi_s = \pm 1, \xi_o = \pm 1). \end{cases} \tag{5.16}$$

Moreover, we rewrite Equation (5.15) to accommodate independent decisions made by two taggers:

$$P(ct|\sigma_{ct}, \vec{\xi}, \vec{\mathcal{D}}) = \frac{1}{N} \left[\frac{1}{c\tau} e^{-ct/c\tau} \theta(ct) \frac{(1 + \xi_s \xi_o \mathcal{D}_s \mathcal{D}_o) + \xi_s \mathcal{D}_s + \xi_o \mathcal{D}_o \cos(\omega t)}{1 + |\xi_s| \cdot 1 + |\xi_o|} \right] \otimes G(\sigma). \tag{5.17}$$

Note that Equation (5.17) reduces to Equation (5.15) if only one tagger is applied. It is general, and accommodates the case of correct decisions by both taggers, incorrect decisions by both or a combination of correct and incorrect decisions.

Chapter 6

CP Violation in $B_s \to J/\psi\phi$

In this chapter, we tackle our final analysis and the highlight of this dissertation, the measurement of the *CP*-violating phase $2\beta_s$ in $B_s \to J/\psi\phi$ decays. This measurement relies simultaneously on all the analysis techniques presented up to this point: the selection method presented in Chapter 3, the framework of the lifetime, width difference and angular analysis presented in Chapter 4, and the flavor-tagging techniques described in Chapter 5. As was demonstrated in Section 1.3.2, even in the absence of information about the production flavor of the B meson (B_s or $\overline{B_s}$) — the *untagged* case — a time-dependent angular analysis has some sensitivity to $2\beta_s$. The untagged analysis is comparatively straightforward, which is not to say without challenges. In principle, a measurement of the *CP*-violating phase $2\beta_s$ performed in this manner differs from the lifetime and angular analysis presented in Chapter 4 only by the addition of one free parameter in the maximum likelihood fit. In practice, this simple change necessitates that we adopt a wholly different method — a variation on interval estimation — to determine the value of $2\beta_s$ and the degree of (dis)agreement of our results with the Standard Model. We use this variation on interval estimation to make a measurement of $2\beta_s$ with both a tagged and untagged analysis of $B_s \to J/\psi\phi$ decays. We begin the discussion by recalling and expanding the probability density functions and likelihood expressions introduced for our measurements of average lifetime, width difference and angular parameters. We then make the case for the interval estimation method by discussing the shortcomings of the straightforward maximum likelihood estimator for this measurement, and present the alternative used in this analysis. We conclude this chapter by presenting the results of the untagged and tagged measurements.

6.1 Selection

The selection algorithm used in the $B_s \to J/\psi\phi$ sample was described in Section 3.3.2. The sample used for the untagged measurement of $2\beta_s$ corresponds to 1.7 fb^{-1}. The sample used for the tagged measurement of $2\beta_s$ corresponds to 1.3 fb^{-1}. The samples are different because as the first analysis of its kind at CDF, the tagged measurement of $2\beta_s$ we present here relies on the tagging development work used in the Δm_s

measurement shortly prior to this analysis. Future updates to this measurement are expected to include the calibrations of taggers for the full sample of CDF data.

6.2 Fitting Method

In this section, we detail the modifications to the likelihood presented in Chapter 4 necessary for a measurement of $2\beta_s$ in a sample of B_s decays without, then with flavor tagging. We recall the portions of the likelihood which remain unchanged from Equation 4.33. The following distributions are taken directly from Chapter 4:

- signal mass PDF (Section 4.1.3):

$$X_{\rm sig}(m_j \,|\, M, \sigma_1, \sigma_2, f_m) = (1-f_m)\;\; \mathcal{G}(m_j \,|\, M, \sigma_1) + \\ f_m \quad \mathcal{G}(m_j \,|\, M, \sigma_2),$$

  with the definition

$$\mathcal{G}(m \,|\, M, \sigma, M_{\rm min}, M_{\rm max}) = \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{m-M}{\sigma}\right)^2\right]}{\frac{1}{2}\left[{\rm Erf}\left(\frac{M_{\rm max}-M}{\sqrt{2}\sigma}\right)+{\rm Erf}\left(\frac{M-M_{\rm min}}{\sqrt{2}\sigma}\right)\right]},$$

- background mass PDF (Section 4.1.3):

$$X_{\rm bkg}(m_j \,|\, A) = A\cdot m_j + \frac{1}{M_{\rm max}-M_{\rm min}}[1-\frac{A}{2}(M_{max}^2 - M_{min}^2)],$$

- background proper decay length PDF (Section 4.1.4):

$$\begin{aligned} Y_{\rm bkg}(ct_j, \sigma_{ct_j} \,|\, f_-, f_+, f_{++}, \lambda_-, \lambda_+, \lambda_{++}, S_{ct}) = \\ (1-f_- - f_+ - f_{++})\, G(ct, \sigma_{ct_j} \,|\, S_{ct}) + \\ f_- E(-ct_j \,|\, \lambda_-) \otimes G(ct, \sigma_{ct_j} \,|\, S_{ct}) + \\ f_+ E(ct_j \,|\, \lambda_+) \otimes G(ct, \sigma_{ct_j} \,|\, S_{ct}) + \\ f_{++} E(ct_j \,|\, \lambda_{++}) \otimes G(ct, \sigma_{ct_j} \,|\, S_{ct}), \end{aligned}$$

- background transversity PDF (Section 4.1.5):

$$Z_{\rm bkg}(\vec{\omega}) = P^\theta \cdot P^\phi \cdot P^\psi,$$

  where

$$\begin{aligned} P^\theta &= \frac{1}{N_\theta}\;\; \left(1 + p_2^\theta \cos^2\theta + p_4^\theta \cos^4\theta\right) \\ P^\phi &= \frac{1}{N_\phi}\;\; \left(1 + \frac{1-p_1^\phi\cos\phi}{2} + \frac{1+p_2^\phi\cos 2\phi}{2} + \frac{1+p_4^\phi\cos 4\phi}{2}\right) \\ P^\psi &= \frac{1}{N_\psi}\;\; \left(1 + p_1^\psi \cos\psi\right), \end{aligned}$$

- PDF for the proper decay length uncertainty implemented as one-dimensional histograms (Section 4.1.4), and
- detector sculpting of angular variables (Section 4.1.5).

This leaves only one change for the untagged analysis: the signal PDF for the proper decay length and transversity is modified to include the possibility of non-zero *CP* violation in $B_s \to J/\psi\phi$ according to the theoretical discussion in Section 1.4. In the case of the tagged analysis, the modifications to the fitting framework are more elaborate. The dependence on $2\beta_s$ is different in the decay rates of B_s and $\overline{B_s}$. Furthermore, the fit likelihood is constructed to incorporate tagging variables, following the general procedure of Section 5.4.

6.2.1 Likelihood without Tagging

In the $2\beta_s$ measurement performed without flavor tagging, the likelihood differs very little from the one used for the measurement of the B_s lifetime difference and angular amplitudes. Without the need to incorporate tagging information, the only change is in fact in the signal PDF for the proper decay length and transversity. We start again from Equation 1.45. Unlike the PDF derived in Section 4.1.6, we do not set $2\beta_s$ to zero. We thus define a decay rate $\frac{d^4\mathcal{P}_{\rm unt}(\vec{\omega},t)}{d\vec{\omega}\ dt}$ for an untagged sample of B_s and $\overline{B_s}$, allowing for the possibility of *CP* violation:

$$\begin{aligned}\frac{d^4\ \mathcal{P}_{\rm unt}(\vec{\omega},t)}{d\vec{\omega}\ dt} \propto\ & |A_0|^2\,\mathcal{T}_+\,f_1 + |A_\parallel|^2\,\mathcal{T}_+\,f_2 + |A_\perp|^2\,\mathcal{T}_-\,f_3 + |A_0||A_\parallel|\cos(\delta_\parallel)\,\mathcal{T}_+\,f_5 \\ & - |A_\parallel||A_\perp|\sin(2\beta_s)\frac{e^{-\Gamma_H t} - e^{-\Gamma_L t}}{2}\cos(\delta_\perp - \delta_\parallel)\,f_4 \\ & - |A_0||A_\perp|\sin(2\beta_s)\frac{e^{-\Gamma_H t} - e^{-\Gamma_L t}}{2}\cos(\delta_\perp)\,f_6,\end{aligned} \tag{1.45}$$

where the functions $\mathcal{T}_\pm$ are defined in Equation 1.44, and the subscript unt refers to the untagged decay rate. The decay rate in Equation 1.45 describes the physics of the decay. We incorporate the various detector effects relevant to this measurement to obtain a likelihood expression. The likelihood construction follows the procedure we outlined in Chapter 4:

- we include the same term $\epsilon(\vec{\omega})$ described in Section 4.1.6 to account for the sculpting of transversity angles by the combined action of the detector, the trigger, and the selection algorithms,
- we convolve all functions of ct with a Gaussian distribution to describe the effect of detector smearing in the ct subspace,
- we calculate a normalization factor $N_{\rm unt}$ that takes into account the detector sculpting in transversity. The normalization factor $N_{\rm unt}$ differs from N because of the different dependence of Equation 1.45 on the theoretical parameters, including $2\beta_s$ (see Appendix D).

The joint PDF in the proper decay length and transversity subspace is then described by

$$W_{\rm unt}(ct,\sigma_{ct},\vec{\omega}\,|\,c\tau_{L,H},S_{ct},\{A_\alpha\},\beta_s) = \frac{1}{N_{\rm unt}}\epsilon(\vec{\omega})\,\frac{d^4\mathcal{P}_{\rm unt}(\vec{\omega},t)}{d\vec{\omega}\;dt}\otimes G(ct,\sigma_{ct}\,|\,S_{ct}). \quad (6.1)$$

The PDF in Equation 6.1 depends on β_s as expected. We also point out that when β_s is not set to zero as it is for the analysis presented in Chapter 4, we restore the dependence of the likelihood expression on $\delta_\perp$.

As is the case when ignoring the effect of *CP* violation, the PDF for proper decay length and transversity presented here lends itself to a reorganization of terms to emphasize the contributions from light and heavy decays:

$$W_{\rm unt}(ct,\sigma_{ct},\vec{\omega}) = \frac{1}{N_{\rm unt}}\epsilon(\vec{\omega})\big[c\tau_L Y^L(ct,\sigma_{ct})\cdot Z^L_{\rm unt}(\vec{\omega}\,|\,\{A_\alpha\})+ \\ c\tau_H Y^H(ct,\sigma_{ct})\cdot Z^H_{\rm unt}(\vec{\omega}\,|\,\{A_\alpha\})\;\big]. \quad (6.2)$$

Equation 6.1 (6.2) is the analogue of Equation 4.30 (4.31) in the case of non-zero *CP* violation. Note that in this case, both light and heavy contributions depend on all three angular amplitudes. As a result of *CP* violation, the light decay is no longer exclusively associated with the A_0 and $A_\parallel$ angular components, and the heavy decay is no longer exclusively associated with the $A_\perp$ component. This is visible in the definitions of $Z^{L/H}_{\rm unt}$:

$$\begin{aligned}
Z^L_{\rm unt} =& |A_0|^2\; f_1(\vec{\omega})\;\frac{1+\cos(2\beta_s)}{2} && + |A_\parallel||A_\perp|\; f_4(\vec{\omega})\;\frac{\cos(\delta_\parallel)\sin(2\beta_s)}{2} \;+\\
& |A_\parallel|^2\; f_2(\vec{\omega})\;\frac{1+\cos(2\beta_s)}{2} && + |A_0||A_\parallel|\; f_5(\vec{\omega})\;\frac{1+\cos(2\beta_s)}{2}\cos(\delta_\perp-\delta_\parallel) \;+\\
& |A_\perp|^2\; f_3(\vec{\omega})\;\frac{1-\cos(2\beta_s)}{2} && + |A_0||A_\perp|\; f_6(\vec{\omega})\;\frac{\cos(\delta_\perp)\sin(2\beta_s)}{2},\\
Z^H_{\rm unt} =& |A_0|^2\; f_1(\vec{\omega})\;\frac{1-\cos(2\beta_s)}{2} && - |A_\parallel||A_\perp|\; f_4(\vec{\omega})\;\frac{\cos(\delta_\parallel)\sin(2\beta_s)}{2} \;+\\
& |A_\parallel|^2\; f_2(\vec{\omega})\;\frac{1-\cos(2\beta_s)}{2} && + |A_0||A_\parallel|\; f_5(\vec{\omega})\;\frac{1-\cos(2\beta_s)}{2}\cos(\delta_\perp-\delta_\parallel) \;+\\
& |A_\perp|^2\; f_3(\vec{\omega})\;\frac{1+\cos(2\beta_s)}{2} && - |A_0||A_\perp|\; f_6(\vec{\omega})\;\frac{\cos(\delta_\perp)\sin(2\beta_s)}{2}.
\end{aligned} \quad (6.3)$$

With Equation 6.1 in hand, the full fit likelihood for the measurement of $2\beta_s$ in the untagged sample is constructed. As was the case for the measurements of lifetime, width difference and angular amplitudes, the likelihood is a product of a mass PDF and a PDF for the proper decay length (PDL) and the transversity for the signal component and a product of a mass PDF, a PDL PDF, and a transversity PDF for the background component. Written in this form, the likelihood is identical in general structure to Equation 4.33, with the modifications contained in $W_{\rm unt}$:

$$\mathcal{L}_j = f_s \cdot X_{\rm sig}(m_j) \cdot W_{\rm unt}(ct_j, \sigma_{ct_j}, \vec{\omega}_j) \cdot P_{\rm sig}(\sigma_{ct_j}) + (1 - f_s) \cdot X_{\rm bkg}(m_j, \sigma_{m_j}) \cdot Y_{\rm bkg}(ct_j, \sigma_{ct_j}) \cdot Z_{\rm bkg}(\vec{\omega}_j) \cdot P_{\rm bkg}(\sigma_{ct_j}). \quad (6.4)$$

The likelihood $\mathcal{L}_j$ is a function of the following parameters:

- Signal: M, S_1, S_2, f_m S_{ct}, $c\tau_L$, $c\tau_H$ $|A_0|$, $|A_\parallel|$, $\delta_\parallel$, $\delta_\perp$, $2\beta_s$, and f_s,
- Background: A, f_-, f_+, f_{++}, $\lambda_-, \lambda_+, \lambda_{++}$, p_2^θ, p_4^θ, $p_1^\phi, p_2^\phi, p_4^\phi$, p_1^ψ.

6.2.2 Likelihood with Tagging

The likelihood expression for the tagged measurement of $2\beta_s$ in a sample of B_s mesons differs substantially from the Equation 6.4, both because of the different decay rates for B_s and $\overline{B_s}$ and because of the incorporation of tagging dilutions and efficiencies. The likelihood construction is comprised of the following steps, some of which are combined in the treatment that follows:

- we revert to the separate B_s and $\overline{B_s}$ decay rates (Equation 1.43),
- an overall normalization is calculated accordingly as described in Appendix D,
- we convolve all functions of ct with a Gaussian accounting for detector smearing,
- tag decisions, dilutions (with scale factors) and efficiencies are incorporated as per the procedure described in Section 5.4,
- we include the likelihood terms for the dilution probability distributions (signal and background) described in Section 5.4.1,
- finally, the full likelihood is constructed with the remaining unchanged PDFs from Sections 4.1.3–4.1.5.

We begin by recalling Equation 1.43:

$$\begin{aligned}\frac{d^4\,\mathcal{P}(\vec{\omega},t)}{d\vec{\omega}\,dt} \propto\ & |A_0|^2(\mathcal{T}_+ \mp \mathcal{T}_{ss})\,f_1 + |A_\parallel|^2(\mathcal{T}_+ \mp \mathcal{T}_{ss})\,f_2 + |A_\perp|^2(\mathcal{T}_- \pm \mathcal{T}_{ss})\,f_3 \\ & + |A_0||A_\parallel|\cos(\delta_2-\delta_1)(\mathcal{T}_+ \mp \mathcal{T}_s)\,f_5 \\ & - |A_\parallel||A_\perp|\left(\cos\delta_1\,\mathcal{T}_s \mp \sin\delta_1\,\mathcal{T}_c \pm \cos\delta_1\,\mathcal{T}_{sc}\right)\,f_4 \\ & - |A_0||A_\perp|\left(\cos\delta_2\,\mathcal{T}_s \mp \sin\delta_2\,\mathcal{T}_c \pm \cos\delta_2\,\mathcal{T}_{sc}\right)\,f_6,\end{aligned} \quad (1.43)$$

with $\mathcal{T}_\pm, \mathcal{T}_{ss}, \mathcal{T}_{sc}, \mathcal{T}_c,$ and $\mathcal{T}_s$ defined in Equation 1.44. The upper sign in Equation 1.43 is for the decay of B_s, and the lower sign for the decay of $\overline{B_s}$. When the terms are separated, we denote the B_s decay rate by $\mathcal{P}_{\rm tag}$, and the $\overline{B_s}$ decay rate by $\overline{\mathcal{P}}_{\rm tag}$.

We define a PDF for the proper decay length and transversity that incorporates the tag decisions and predicted dilutions of the two taggers we use (same side and opposite side). In the following PDF, the terms accounting for angular sculpting,

normalization, and Gaussian convolution in ct are included simultaneously:

$$T(ct_j, \sigma_{ct}, \vec{\omega}, \mathcal{D}_{o,s}, \xi_{o,s}) = \epsilon(\vec{\omega}) \cdot \left[\frac{1}{N_{\rm tag}} \frac{1+\xi_o S_{\mathcal{D}_o} \mathcal{D}_o}{1+|\xi_o|} \cdot \frac{1+\xi_s S_{\mathcal{D}_s} \mathcal{D}_s}{1+|\xi_s|} \cdot \frac{d^4 \mathcal{P}_{\rm tag}(\vec{\omega}, t)}{d\vec{\omega}\ dt} + \frac{1}{\overline{N}_{\rm tag}} \frac{1-\xi_o S_{\mathcal{D}_o} \mathcal{D}_o}{1+|\xi_o|} \cdot \frac{1-\xi_s S_{\mathcal{D}_s} \mathcal{D}_s}{1+|\xi_s|} \cdot \frac{d^4 \overline{\mathcal{P}}_{\rm tag}(\vec{\omega}, t)}{d\vec{\omega}\ dt} \right] \otimes G(ct, \sigma_{ct}). \quad (6.5)$$

Equation 6.5 is used to construct the full likelihood:

$$\mathcal{L}_j = f_s \cdot X_{\rm sig}(m_j, \sigma_{m_j}) \cdot P_{\rm sig}(\vec{\xi_j}) \cdot T(ct, \sigma_{ct_j}, \vec{\omega}_j, \mathcal{D}_{o,s}, \xi_{o,s}) \cdot P_{\rm sig}(\sigma_{ct_j}) \cdot P_{\rm sig}(\mathcal{D}_j) + (1-f_s) \cdot X_{\rm bkg}(m_j, \sigma_{m_j}) \cdot P_{\rm bkg}(\vec{\xi_j}) \cdot Y_{\rm bkg}(ct_j, \sigma_{ct_j}) \cdot Z_{\rm bkg}(\vec{\omega}_j) \cdot P_{\rm bkg}(\sigma_{ct_j}) \cdot P_{\rm bkg}(\mathcal{D}_j). \quad (6.6)$$

The likelihood $\mathcal{L}_j$ is a function of the following parameters:

- Signal: M, S_1, S_2, f_m S_{ct}, $c\tau_L$, $c\tau_H$ $|A_0|$, $|A_\parallel|$, $\delta_\parallel$, $\delta_\perp$, $2\beta_s$, Δm_s and f_s,
- Background: A, f_-, f_+, f_{++}, $\lambda_-, \lambda_+, \lambda_{++}$, p_2^θ, p_4^θ, $p_1^\phi, p_2^\phi, p_4^\phi$, p_1^ψ,
- Tagging parameters: ϵ_o, ϵ_s, $S_{\mathcal{D}_o}$, $S_{\mathcal{D}_s}$.

Parameter Constraints in the Fit for the Tagged Measurement

The parameters Δm_s, $S_{\mathcal{D}_o}$ and $S_{\mathcal{D}_s}$ enter the likelihood in the fit performed in the tagged sample, but they are not determined by that fit. For Δm_s, the reason is that this measurement uses a fundamentally different method than a dedicated Δm_s measurement. In the latter, a Fourier analysis is used, and different systematic effects must be accounted for. As a result, we constrain Δm_s in our fit to the world average value of 17.77 ± 0.12 ps^{-1}.

The dilution scale factors $S_{\mathcal{D}_o}$ and $S_{\mathcal{D}_s}$ enter the likelihood in several terms as multiplicative factors of $2\beta_s$. Using a constraint on these terms therefore enhances our sensitivity $2\beta_s$. These constraints are appropriate because the scale factors are determined to good precision and confidence by the tagging calibration procedure. We constrain them in the fit to the values obtained from the calibrations discussed in Chapter 5.

6.3 Limitations of the ML Method

The maximum likelihood (ML) method is a very common choice of procedure for parameter estimation, but it is not without its limitations. In particular, it is only in the limit of infinite statistics that it is guaranteed to be an unbiased estimator returning the smallest possible statistical uncertainty on the estimated parameter. The size of the data sample required to achieve an unbiased result with optimal

uncertainty is dependent on the likelihood function used, as the ML estimator is not immune to certain pathologies of the probability density functions used to construct the likelihood.

The most direct way to test the robustness of the parameter estimation method is to generate pseudo-experiments at known values of the parameters we wish to estimate, and then perform fits as we do on data to see how well the estimator performs. In Figure 6-1, we show the result of generating and subsequently analyzing 300 such pseudo-experiments at $\Delta\Gamma = 0.1, 2\beta_s = 0$ without the use of flavor tagging. The color map shows the number of fits that return a fitted value of $\Delta\Gamma, 2\beta_s$ at each point in the plane. We expect the distribution of fitted values to be centered at the generated values, and symmetric around them. We observe a large shift of $2\beta_s$ away from zero, as well as a less pronounced, but still significant bias in $\Delta\Gamma$, on the order of the statistical uncertainty on $\Delta\Gamma$.

The explanation for this effect lies in the structure of the PDF for proper decay length and transversity angles. Equation 1.45 is significantly altered in cases where $2\beta_s = 0$: the only terms that contain $\delta_\perp$ go to zero, and the number of degrees of freedom available to the estimator is therefore reduced by one. In a maximization procedure, regions of phase space with an effectively reduced number of degrees of freedom are heavily disfavored. In essence, this means that regardless of the true value of $2\beta_s$, a maximum likelihood estimator will systematically move away from small values of $2\beta_s$. Analogously, for $\Delta\Gamma = 0$, the two lifetime exponentials collapse into one, which removes both $2\beta_s$ and $\delta_\perp$ from the likelihood expression, an effective decrease of 2 in the number of degrees of freedom.

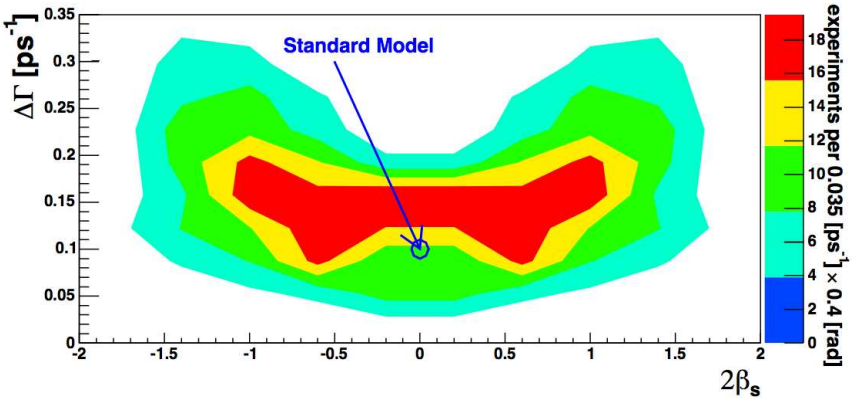


Figure 6-1: The results of 300 pseudo-experiments performed at input values $\Delta\Gamma = 0.1, 2\beta_s = 0$.

We also observe large shifts in the fitted values of $\Delta\Gamma$ and $2\beta_s$ in pseudo-experiments that include flavor tagging when the tagging parameters (efficiency ϵ, dilution $\mathcal{D}$) are set to be commensurate to the actual performance we see in data. Although with perfect tagging ($\epsilon = 1$, $\mathcal{D} = 1$), the likelihood expression is shielded from the pathologies of the tagged analysis, in a realistic scenario with less than perfect efficiency and dilution, a majority of events in the tagged sample effectively behave as untagged events, either because a tag decision is in fact not reached, or because dilution is low.

In the maximization procedure, the ML method here again shifts away from regions of phase space where the number of degrees of freedom is reduced for these effectively untagged events.

6.4 Confidence Region in the $\Delta\Gamma$–$2\beta_s$ Plane

6.4.1 Likelihood Ratio Method

We have established in Section 6.3 that the maximum likelihood method is not an adequate framework for estimating the values of $\Delta\Gamma$ and $2\beta_s$ with which the data in our sample are most compatible. We opt instead to test the compatibility of our data directly with a set of points on the $\Delta\Gamma$–$2\beta_s$ plane using the likelihood ratio (LR) method suggested in Reference [114]. We define a ratio R as

$$R(\Delta\Gamma, 2\beta_s) = \log \frac{\mathcal{L}(\hat{\Delta\Gamma}, 2\hat{\beta}_s, \hat{\theta})}{\mathcal{L}(\Delta\Gamma, 2\beta_s, \hat{\theta}')}, \tag{6.7}$$

where θ represents all fit parameters other than the ones for which we want to estimate data-model agreement, in this case all parameters beside $\Delta\Gamma$ and $2\beta_s$. The hat on a parameter denotes that the likelihood is evaluated at the value of that parameter that maximizes $\mathcal{L}$. The numerator in Equation 6.7, $\mathcal{L}(\hat{\Delta\Gamma}, 2\hat{\beta}_s, \hat{\theta})$ is thus simply the maximum likelihood obtained from a fit on our data sample, allowing all parameters to float freely. The denominator is the maximum likelihood evaluated at the fixed point $(\Delta\Gamma, 2\beta_s)$, letting all others parameters θ float freely. We note that $\hat{\theta}$ and $\hat{\theta}'$ differ because different values of these parameters will maximize $\mathcal{L}$ for the points $(\hat{\Delta\Gamma}, 2\hat{\beta}_s)$ and $(\Delta\Gamma, 2\beta_s)$. We evaluate R on a fixed grid in the $\Delta\Gamma$–$2\beta_s$ plane for $\Delta\Gamma \in [-0.7, 0.7]$, $2\beta_s \in [-\pi, \pi]$.

Having obtained the ratios $R_{\text{data}}(\Delta\Gamma_i, \beta_{s_j})$ for the theoretical hypotheses $(\Delta\Gamma_i, \beta_{s_j})$, we need to evaluate the probability that statistical fluctuations in the data assuming each of these hypotheses would produce the value of R observed. We generate a set of 300 pseudo-experiments for each hypothesis $(\Delta\Gamma_i, \beta_{s_j})$. The value of the parameters θ used to generate these experiments is $\hat{\theta}'$, as this is considered to be the best estimate of θ for a model that assumes $(\Delta\Gamma_i, \beta_{s_j})$. We define $H(R, \Delta\Gamma_i, \beta_{s_j})$ as the normalized distribution of the ratio R as each point $(\Delta\Gamma_i, \beta_{s_j})$.

Finally, we evaluate a p-value for each point $(\Delta\Gamma_i, \beta_{s_j})$, which is the *fraction of hypothetical experiments conducted in a physical world in which* $(\Delta\Gamma, \phi_s) = (\Delta\Gamma_i, \beta_{s_j})$ *for which we would observe* $R(\Delta\Gamma_i, \beta_{s_j}) \geq R_{\text{data}}(\Delta\Gamma_i, \beta_{s_j})$. The value $p(\Delta\Gamma_i, \beta_{s_j})$ is thus defined as

$$p(\Delta\Gamma_i, \beta_{s_j}) = \int_{R_{\text{data}}(\Delta\Gamma_i, \beta_{s_j})}^{\infty} H(R, \Delta\Gamma_i, \beta_{s_j})\, dR. \tag{6.8}$$

These p-values are results in and of themselves, quantifying the agreement of the data with any given model with $(\Delta\Gamma_i, \beta_{s_j}, \theta')$. In order to obtain a more practical summary result, we use the p-values thus obtained to define a confidence region in the $\Delta\Gamma$–$2\beta_s$

plane.

6.4.2 Confidence Region Estimation

The likelihood ratio method described above lends itself naturally to the construction of an $X\%$ confidence level (C.L.) region which has a probability of $X\%$ of containing the true values of $\Delta\Gamma$ and $2\beta_s$. In this method, the probabilities we report are defined in the frequentist sense: the parameters of interest have probability 1 of being at their true value, and probability zero of being anywhere else, and confidence regions constructed with this method have probability $X\%$ of enclosing those values.

Using these definitions, the set of points for which $p(\Delta\Gamma_i, \beta_{s_j}) > X\%$ define the $X\%$ C.L. region. We define 68% and 95% C.L. regions.

6.4.3 Systematic Effects with the LR Method

In performing the confidence region estimation, we make one important assumption that must be tested: when we calculate the denominator of the likelihood ratio R at a given point $\Delta\Gamma_i$, β_{s_i}, we use the values of the parameters $\hat{\theta}'$ that maximize the likelihood at $\Delta\Gamma_i$, β_{s_i}. In particular, we choose the values $\hat{\theta}'$ as inputs when we generate pseudo-experiments to evaluate the likelihood ratio. While this is a reasonable assumption because it is supported by the data (in maximizing $\mathcal{L}$), it is not known whether these are the true values of the parameters θ for a physical world in which $\Delta\Gamma = \Delta\Gamma_i$ and $2\beta_s = \phi_{s_i}$.

We test the effect of this assumption on the confidence region estimation by replacing the data sample in our likelihood ratio calculation with a pseudo-experiment that we generate with values of $\theta \neq \hat{\theta}'$. We focus on the parameters most highly correlated with $\Delta\Gamma$ and β_s because the likelihood ratio method is computationally prohibitive: $c\tau$, $|A_0|$, $|A_\parallel|$, $\delta_\parallel$, $\delta_\perp$ and S_{ct}.

We perform this test for 25 points on the $\Delta\Gamma$–β_s plane for $\Delta\Gamma \in [-0.8, 0.8]$. First, the nominal fit is performed at each $\Delta\Gamma$–β_s point. Second, at each point, we generate 16 pseudo-experiments. We choose the input parameters for $c\tau$, $|A_0|$, $|A_\parallel|$, $\delta_\parallel$, $\delta_\perp$ and S_{ct} randomly from a uniform distribution centered at the value returned by the nominal fit (at this $\Delta\Gamma$–β_s point) and extending to $\pm 5\sigma$ for each parameter, with σ equal to the symmetric uncertainty returned by the nominal fit. We observe some fluctuations in the values of R obtained in this manner, but the p-values obtained are comparable to the ones we obtain in the standard evaluation of the confidence region. The confidence region is not modified by these changes.

6.4.4 Comparison with 2D Likelihood Profile

The confidence region calculated using the likelihood ratio method should not be confused with another common method used to present the measurements of two parameters simultaneously in two-dimensional space, the 2D likelihood profile. A likelihood profile for two parameters θ, δ is evaluated in the following manner.

- First, we perform an ordinary likelihood fit and note the value of the likelihood function $\mathcal{L}^{\max}$ for which the likelihood is maximized. In this fit, all parameters, including θ and δ, are allowed to float freely. Their best values are determined by the fit.
- Second, we perform likelihood fits at several points $(\theta_i,\, \delta_j)$ in the 2D plane θ–δ, each time fixing the value of the two parameters to $(\theta_i,\, \delta_j)$ and letting other fit parameters float. All parameters *except* θ *and* δ are allowed to vary in order to maximize $\mathcal{L}$. We retain the value of the likelihood function $\mathcal{L}_{ij}^{\max}$ for each of these points.
- Finally, we define likelihood contours corresponding to $n\sigma$ uncertainty. For instance, the 1σ contour corresponds to points for which $2\Delta\log\mathcal{L} \equiv 2(\mathcal{L}^{\max} - \log\mathcal{L}_{ij}^{\max}) = 2.30$, and the 2σ contour corresponds to points for which $2\Delta\log\mathcal{L} = 5.99$*.

The steps outlined here constitute a portion of the procedure that is used to evaluate a confidence region using the likelihood ratio method. They do not include the calculation of a p-value quantifying the betting odds that correspond to each θ–δ point. The fundamental difference between the two methods lies in the statistical statement that is made regarding the result. Using the likelihood ratio method, the confidence region ensures proper statistical *coverage* by construction: the $X\%$ confidence region has a probability $X\%$ of enclosing the true values of the parameters θ, δ. The likelihood contour is not guaranteed to do so without explicit verification using a large sample of pseudo-experiments or simulated data.

As a matter of practical consideration however, the likelihood ratio method is substantially more demanding in computation time. For each point in the parameter space for which we wish to construct a confidence region, we perform $\mathcal{O}(300)$ likelihood fits, one for each pseudo-experiment we generate. On the other hand, for each point in the parameter space for which we wish to construct a likelihood profile, we perform only one likelihood fit.

6.5 Results

In this section, we present and discuss the results of the $2\beta_s$ measurement using the likelihood ratio method. We begin with the confidence region obtained using the untagged analysis with a sample of $B_s \to J/\psi\phi$ decays corresponding to 1.7 fb^{-1} of integrated luminosity. We then present the results obtained using the tagged analysis with a data sample corresponding to 1.3 fb^{-1} of integrated luminosity.

The focus of this dissertation is on the tagged analysis. Therefore we perform additional studies yielding several associated results using the tagging framework: likelihood profiles obtained by constraining the fit parameters $c\tau$, $\delta_\parallel$ and $\delta_\perp$. We choose likelihood profiles for the constrained fit results because the likelihood ratio method is two orders of magnitude more demanding in computing time. For each

*These values depend on the dimensionality of the likelihood profile, i.e. 1D, 2D, etc. They are quoted here for the 2D case.

confidence region, we report the p-value of the Standard Model expectation for $\Delta\Gamma$ and $2\beta_s$. In addition to numerical results, we present projections of the results of the likelihood fits in mass, proper decay length, and transversity variables, and overlay them on the data distributions in each of these variables.

6.5.1 Confidence Region in the Untagged Analysis

The confidence region in the $\Delta\Gamma$–$2\beta_s$ plane is calculated with the untagged analysis for $\Delta\Gamma \in [-0.7, 0.7]$ and $2\beta_s \in [-\pi, \pi]$. We scan this plane with a grid of 20×20 points, and evaluate the likelihood ratio at each point using between 100 and 300 pseudo-experiments, with fewer pseudo-experiments used at points where the p-value is low. The data fits are performed with the same sample as in the measurements of $c\tau$, $\Delta\Gamma$ and $\{A_\alpha\}$ performed with the assumption $2\beta_s = 0$. This sample is comprised of 10 059 candidates, both signal and background. Figure 6-2 shows the confidence region in the $\Delta\Gamma$–$2\beta_s$ plane calculated using the likelihood ratio method. Assuming the Standard Model predictions of $2\beta_s$ and $\Delta\Gamma$, shown as a point in Figure 6-2, the probability of a deviation as large as the level of the observed data is 22%. The projections of the fit results onto mass and proper decay length are shown along with data distributions in Figures 6-3. The projections onto the angular variables are shown in Figure 6-4. The correlation matrix of parameters returned by the nominal fit is shown in Tables 6.1 and 6.2 at the end of this chapter.

We see in Figure 6-2 that the untagged analysis does not limit significantly the range of $2\beta_s$ with which our data is compatible. This is in part due to the additional degeneracy in the likelihood of the untagged measurement. There are four equivalent minima in the $\Delta\Gamma$–$2\beta_s$ plane with exactly the same p-value. In fact, the entire confidence region displays the symmetry predicted in Section 1.4. The likelihood, and therefore the ratios R and the p-value, are all invariant under the following transformation:

$$\begin{aligned} \beta_s &\to \frac{\pi}{2} - \beta_s, \\ \Delta\Gamma &\to -\Delta\Gamma, \end{aligned}$$

and independently under the transformation

$$\begin{aligned} \delta_\parallel &\to 2\pi - \delta_\parallel, \\ \delta_\perp &\to 2\pi - \delta_\perp, \end{aligned}$$

resulting in a four-fold ambiguity in the $\Delta\Gamma$–$2\beta_s$ plane. The four points are close enough that the contours around them overlap in our confidence region. As a result, no portion of $2\beta_s$ axis is excluded by our confidence region.

6.5.2 Confidence Region in the Unconstrained Tagged Fit

The confidence region in the $\Delta\Gamma$–$2\beta_s$ plane is calculated with the tagged analysis for $\Delta\Gamma \in [-0.7, 0.7]$ and $2\beta_s \in [-\pi, \pi]$. We scan this plane with a grid of 20×20 points,

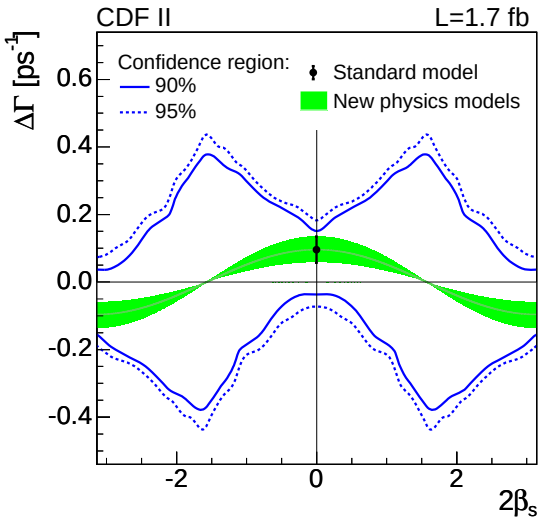


Figure 6-2: The 90 and 95% confidence level regions in the $\Delta\Gamma - 2\beta_s$ plane obtained using the likelihood ratio method in the untagged analysis. The green band shows the set of points in the plane that correspond to new physics models which are compatible with the relation $\Delta\Gamma = 2|\Gamma_{12}|\cos 2\beta_s$.

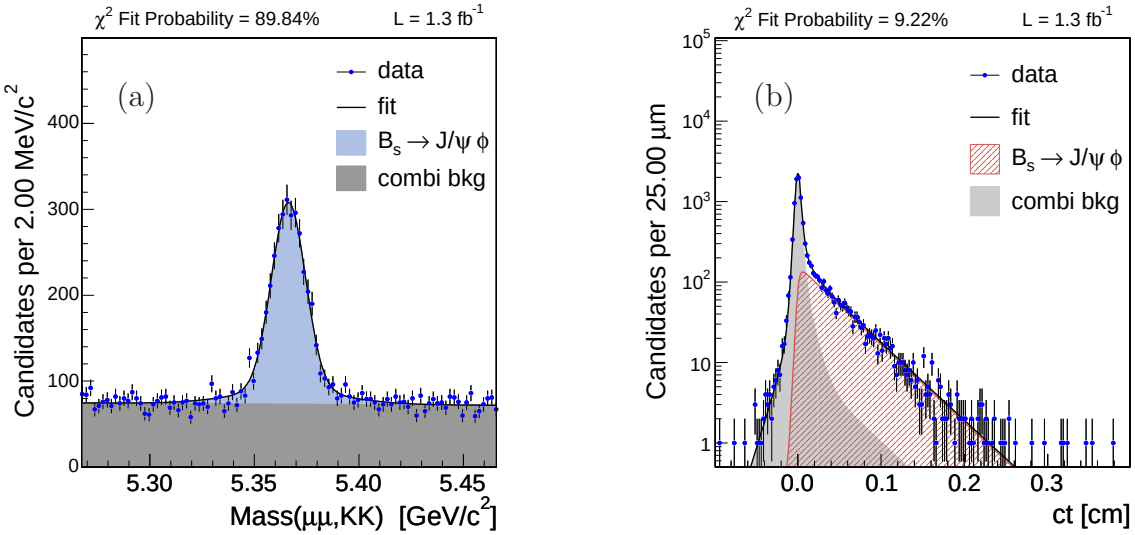


Figure 6-3: Projections of the results of the untagged maximum likelihood fit for $2\beta_s$ performed on the $B_s \to J/\psi\phi$ sample overlaid on data: (a) Mass, (b) PDL.

and evaluate the likelihood ratio at each point using 350 pseudo-experiments. The data fits are performed with a data sample that is comprised of 6 946 candidates, both signal and background. The confidence region in the $\Delta\Gamma$–β_s plane is shown in Figure 6-5, and represents the main result using the likelihood ratio method. For comparison, the 2D likelihood profile is shown alongside the main result. The 68 and

95% confidence level contours are shown. Assuming the Standard Model predictions of $2\beta_s$ and $\Delta\Gamma$ (shown as a point in both figures), the probability of a deviation as large as the level of the observed data is 33%, corresponding to one standard deviation. We see that the addition of tagging reduces by approximately 50% the allowed range of $2\beta_s$ in our confidence region. The precision in the estimation of $\Delta\Gamma$ is reduced by the fact that we are using a data sample that is approximately 25% smaller than in the untagged measurement. The projections of the fit results onto mass and proper decay length are shown along with data distributions in Figures 6-6. The projections onto the angular variables are shown in Figure 6-7. Finally, the correlation matrix of parameters returned by the nominal fit is shown in Tables 6.3 and 6.4 at the end of this chapter.

6.5.3 Likelihood Profiles in the Constrained Tagged Fits

We perform constrained likelihood fits and report the likelihood profile in each case in addition to the nominal result. While we remove the four-fold degeneracy of the untagged likelihood by making use of flavor tagging in the fit, it is visible in Figure 6-5 that there are still two equivalent points in the $\Delta\Gamma$–β_s plane, and that the likelihood cannot distinguish between them. As noted in Section 1.4, the tagged likelihood, and therefore the ratio R and the p-value, are all invariant under the following transformation:

$$\begin{aligned} \beta_s &\to \frac{\pi}{2} - \beta_s, & \delta_\parallel &\to 2\pi - \delta_\parallel, \\ \Delta\Gamma &\to -\Delta\Gamma, & \delta_\perp &\to 2\pi - \delta_\perp. \end{aligned}$$

However, the inclusion in the fit of additional knowledge on these parameters limits the allowed range in the confidence region. Constraining the phases $\delta_\parallel$ and $\delta_\perp$ removes the degeneracy in the solution. Therefore we perform three additional fits, in which we constrain $c\tau$, then the phases $\delta_\parallel$ and $\delta_\perp$, then both $c\tau$ and the phases. These are shown in Figures 6-8 and 6-9.

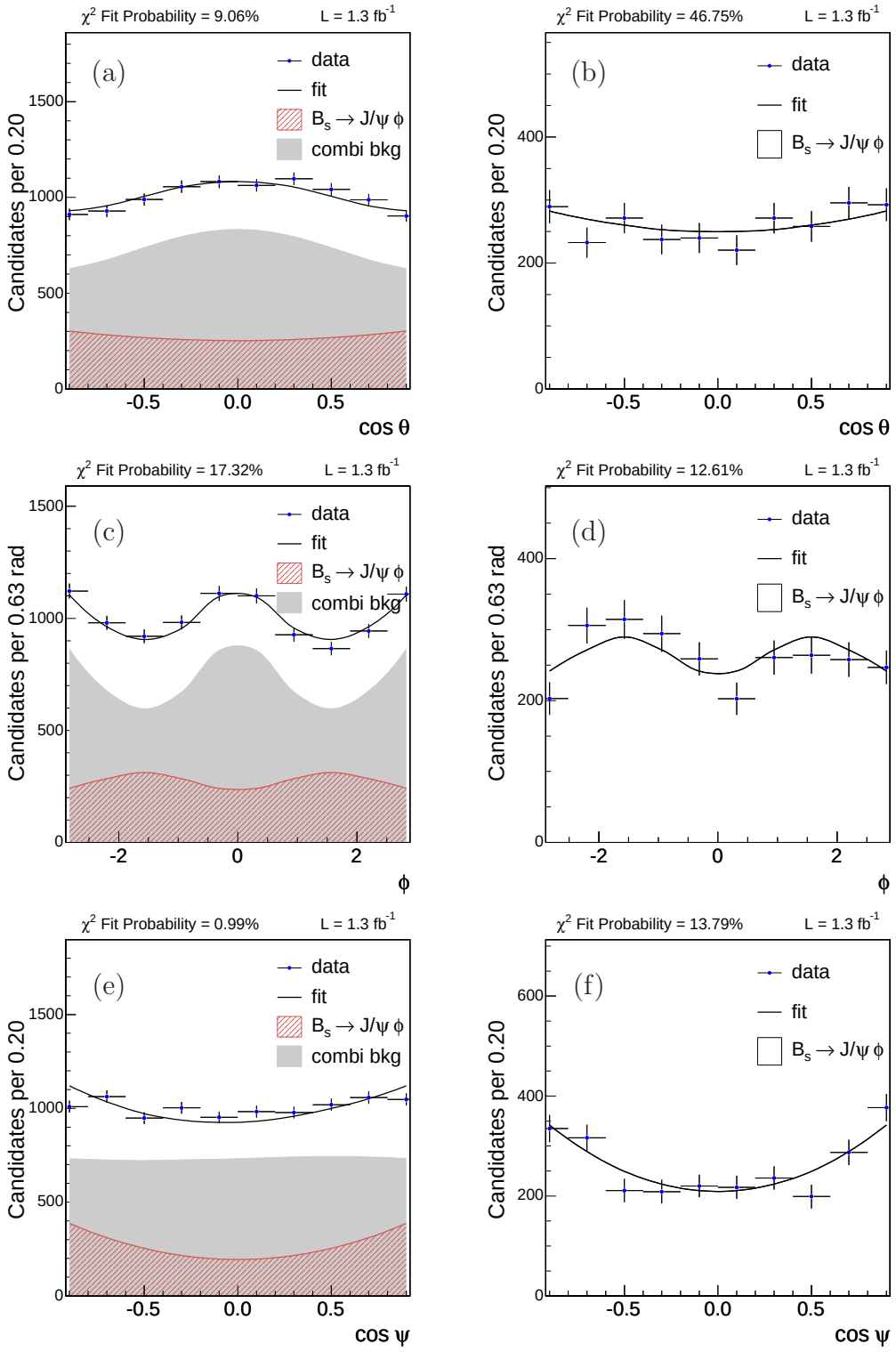


Figure 6-4: Projections of the results of the untagged maximum likelihood fit for $2\beta_s$ performed on the $B_s \to J/\psi\phi$ sample overlaid on data: (a)-(b) $\cos\theta$, (c)-(d) ϕ, (e)-(f) $\cos\psi$. On the left we show the signal and background distributions, and on the right, we show sideband-subtracted and sculpting-corrected signal distributions.

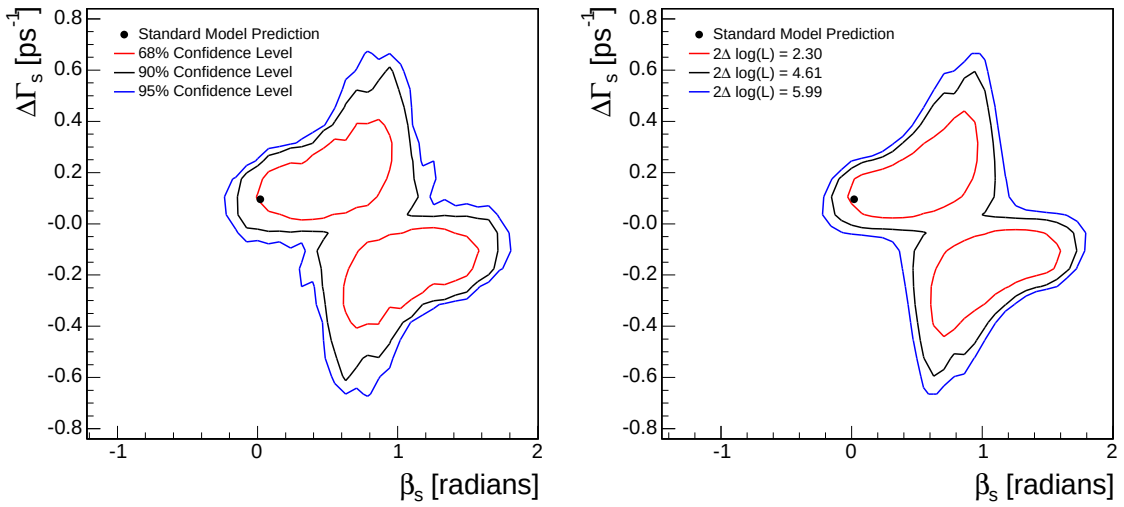


Figure 6-5: The 68, 90 and 95% confidence level regions in the $\Delta\Gamma - 2\beta_s$ plane obtained using the likelihood ratio method (left) and the 2D likelihood profile (right).

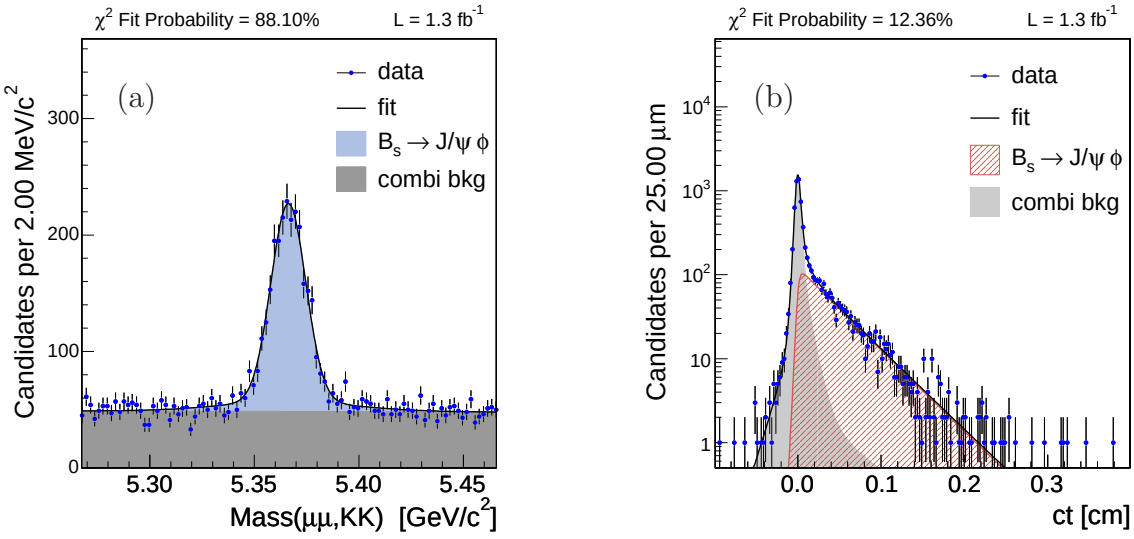


Figure 6-6: Projections of the results of the unconstrained maximum likelihood fit performed on the $B_s \to J/\psi\phi$ sample using flavor tagging overlaid on data: (a) Mass, (b) PDL.

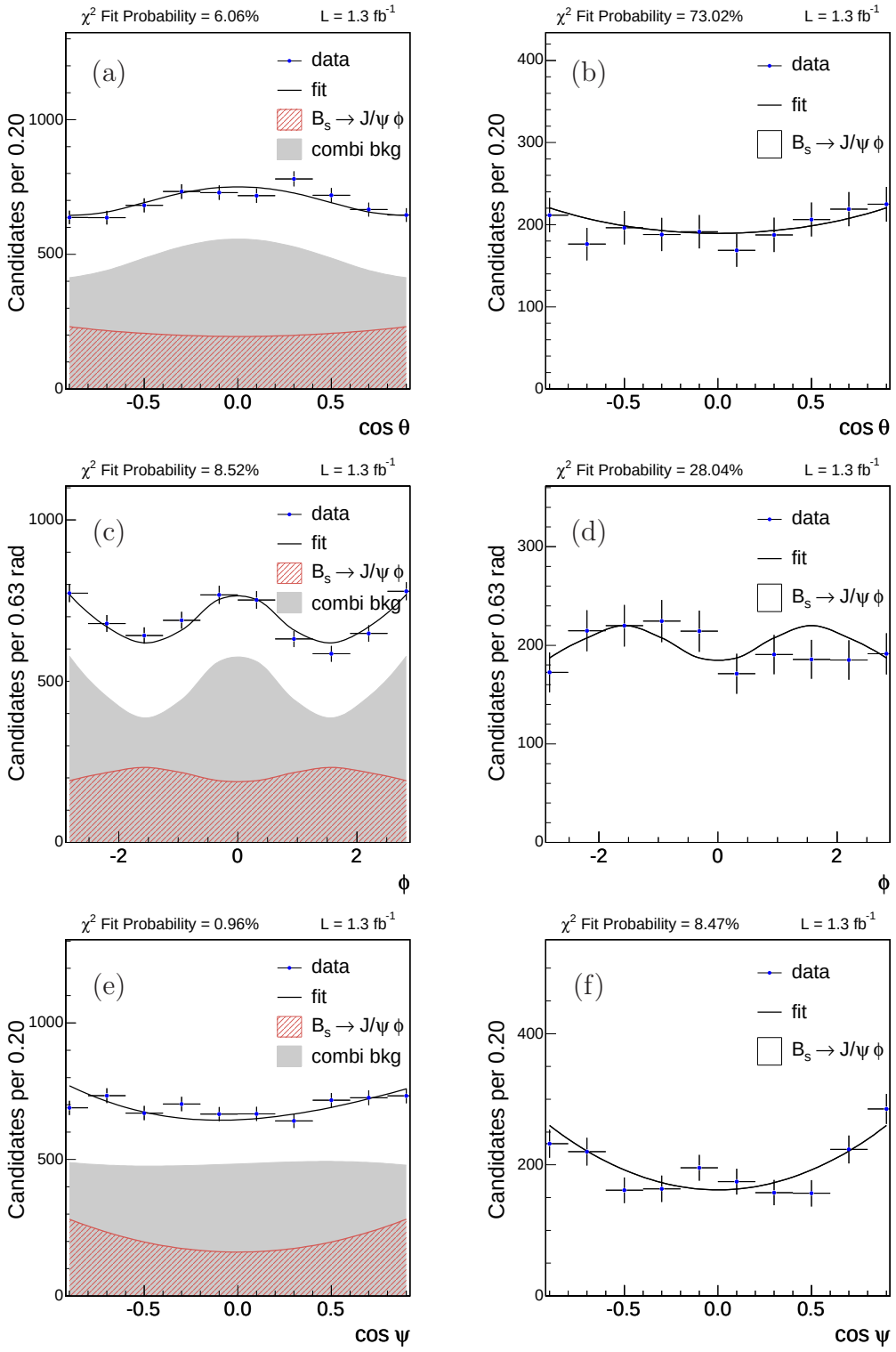


Figure 6-7: Projections of the results of the unconstrained maximum likelihood fit performed on the $B_s \to J/\psi\phi$ sample using flavor tagging overlaid on data: (a)-(b) $\cos\theta$, (c)-(d) ϕ, (e)-(f) $\cos\psi$. On the left we show the signal and background distributions, and on the right, we show sideband-subtracted and sculpting-corrected signal distributions.

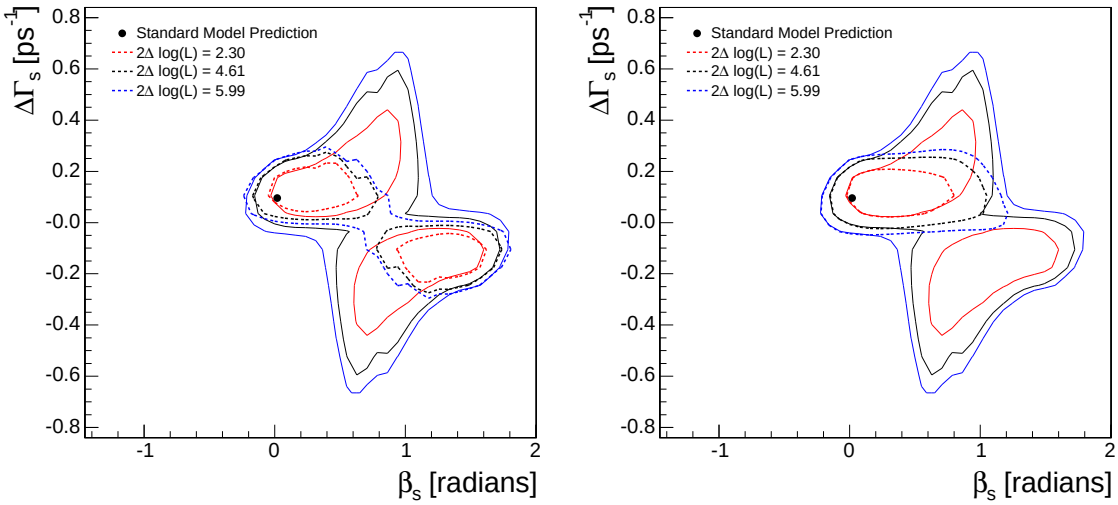


Figure 6-8: Likelihood profiles from the tagged fit with $c\tau$ (left) or δ_α (right) constrained.

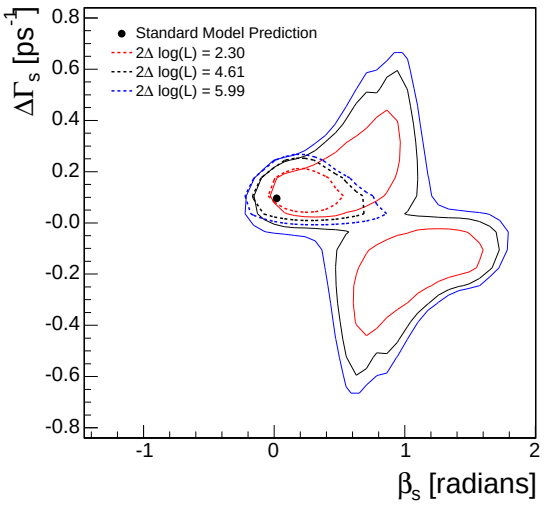


Figure 6-9: Likelihood profiles from the tagged fit with $c\tau$ and δ_α simultaneously constrained.

Par.	M_s	σ_1	σ_2	f_m	A	f_s	$c\tau_s$	$\Delta\Gamma$	S_{ct}	f_+	f_{++}	f_-	$c\tau_+$
M_s	1	-0.127	-0.137	0.135	-0.027	0.056	0.001	0.001	-0.018	-0.002	0.075	0.007	-0.033
σ_1	-0.127	1	0.784	-0.841	0.007	-0.344	-0.013	0.004	0.066	-0.008	-0.398	-0.017	0.146
σ_2	-0.137	0.784	1	-0.823	0.006	-0.478	-0.013	0.002	0.091	-0.017	-0.555	-0.025	0.204
f_m	0.135	-0.841	-0.823	1	-0.009	0.269	-0.022	0.020	-0.061	-0.033	0.366	0.019	-0.146
A	-0.027	0.007	0.006	-0.009	1	-0.000	-0.001	0.000	0.001	0.001	-0.002	-0.000	0.002
f_s	0.056	-0.344	-0.478	0.269	-0.000	1	0.171	-0.127	-0.057	0.110	0.347	0.003	-0.115
$c\tau_s$	0.001	-0.013	-0.013	-0.022	-0.001	0.171	1	-0.830	0.001	0.129	-0.053	-0.014	0.034
$\Delta\Gamma$	0.001	0.004	0.002	0.020	0.000	-0.127	-0.830	1	-0.005	-0.112	0.080	0.014	-0.051
S_{ct}	-0.018	0.066	0.091	-0.061	0.001	-0.057	0.001	-0.005	1	-0.358	-0.170	-0.455	0.251
f_+	-0.002	-0.008	-0.017	-0.033	0.001	0.110	0.129	-0.112	-0.358	1	-0.122	0.186	-0.202
f_{++}	0.075	-0.398	-0.555	0.366	-0.002	0.347	-0.053	0.080	-0.170	-0.122	1	0.086	-0.759
f_-	0.007	-0.017	-0.025	0.019	-0.000	0.003	-0.014	0.014	-0.455	0.186	0.086	1	-0.156
$c\tau_+$	-0.033	0.146	0.204	-0.146	0.002	-0.115	0.034	-0.051	0.251	-0.202	-0.759	-0.156	1
$c\tau_{++}$	-0.034	0.151	0.207	-0.142	0.003	-0.127	0.024	-0.082	0.129	0.181	-0.764	-0.077	0.678
$c\tau_-$	-0.007	0.027	0.037	-0.024	0.000	-0.021	0.003	-0.005	0.322	-0.146	-0.079	-0.627	0.120
$\lvert A_0\rvert^2$	0.002	-0.027	-0.039	0.007	-0.000	0.084	0.359	-0.078	-0.002	0.038	0.035	-0.003	-0.010
$\lvert A_1\rvert^2$	0.006	-0.015	-0.024	0.009	-0.001	0.043	0.321	-0.076	-0.002	0.029	0.014	-0.001	-0.009
p_2^θ	0.002	-0.020	-0.026	0.015	0.000	0.020	-0.005	-0.002	-0.004	0.007	0.012	0.001	0.001
p_4^θ	-0.000	0.012	0.016	-0.007	-0.000	-0.014	-0.002	0.003	0.002	-0.007	-0.006	-0.001	-0.002
p_1^ϕ	-0.003	0.011	0.014	-0.013	0.000	-0.004	0.002	-0.003	0.004	0.000	-0.015	-0.002	0.013
p_2^ϕ	-0.013	0.068	0.096	-0.064	0.000	-0.067	-0.018	0.014	0.013	-0.012	-0.062	-0.003	0.027
p_4^ϕ	-0.000	0.001	0.005	0.001	0.000	-0.012	-0.005	0.004	-0.001	-0.003	-0.005	0.001	0.001
p_1^ψ	0.004	-0.009	-0.009	0.011	-0.001	-0.000	-0.003	0.002	-0.002	0.001	0.003	0.001	-0.001
β_s	-0.001	-0.002	-0.000	-0.022	-0.001	0.118	0.919	-0.791	0.004	0.101	-0.057	-0.011	0.034
$\delta_\parallel$	-0.006	-0.012	-0.013	0.005	0.000	0.010	-0.080	0.089	-0.002	-0.006	0.024	0.001	-0.004
$\delta_\perp$	-0.003	0.023	0.032	-0.000	-0.000	-0.129	-0.757	0.852	-0.003	-0.102	0.056	0.012	-0.044
Glob.	0.146	0.860	0.921	0.903	0.028	0.585	0.939	0.933	0.565	0.692	0.940	0.683	0.886

Table 6.1: Correlation matrix for the parameters in the $B_s \to J/\psi\phi$ likelihood returned by the unconstrained untagged fit (Part 1 of 2).

Par.	$c\tau_{++}$	$c\tau_{-}$	$\|A_0\|^2$	$\|A_1\|^2$	p_2^θ	p_4^θ	p_1^ϕ	p_2^ϕ	p_4^ϕ	p_1^ψ	β_s	$\delta_\parallel$	$\delta_\perp$
M_s	-0.034	-0.007	0.002	0.006	0.002	-0.000	-0.003	-0.013	-0.000	0.004	-0.001	-0.006	-0.003
σ_1	0.151	0.027	-0.027	-0.015	-0.020	0.012	0.011	0.068	0.001	-0.009	-0.002	-0.012	0.023
σ_2	0.207	0.037	-0.039	-0.024	-0.026	0.016	0.014	0.096	0.005	-0.009	-0.000	-0.013	0.032
f_m	-0.142	-0.024	0.007	0.009	0.015	-0.007	-0.013	-0.064	0.001	0.011	-0.022	0.005	-0.000
A	0.003	0.000	-0.000	-0.001	0.000	-0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	-0.000
f_s	-0.127	-0.021	0.084	0.043	0.020	-0.014	-0.004	-0.067	-0.012	-0.000	0.118	0.010	-0.129
$c\tau_s$	0.024	0.003	0.359	0.321	-0.005	-0.002	0.002	-0.018	-0.005	-0.003	0.919	-0.080	-0.757
$\Delta\Gamma$	-0.082	-0.005	-0.078	-0.076	-0.002	0.003	-0.003	0.014	0.004	0.002	-0.791	0.089	0.852
S_{ct}	0.129	0.322	-0.002	-0.002	-0.004	0.002	0.004	0.013	-0.001	-0.002	0.004	-0.002	-0.003
f_+	0.181	-0.146	0.038	0.029	0.007	-0.007	0.000	-0.012	-0.003	0.001	0.101	-0.006	-0.102
f_{++}	-0.764	-0.079	0.035	0.014	0.012	-0.006	-0.015	-0.062	-0.005	0.003	-0.057	0.024	0.056
f_-	-0.077	-0.627	-0.003	-0.001	0.001	-0.001	-0.002	-0.003	0.001	0.001	-0.011	0.001	0.012
$c\tau_+$	0.678	0.120	-0.010	-0.009	0.001	-0.002	0.013	0.027	0.001	-0.001	0.034	-0.004	-0.044
$c\tau_{++}$	1	0.062	-0.026	-0.017	0.004	-0.005	0.009	0.022	0.000	-0.000	0.042	-0.013	-0.083
$c\tau_-$	0.062	1	-0.001	-0.001	-0.001	0.001	0.002	0.005	-0.001	-0.001	0.004	-0.001	-0.004
$\|A_0\|^2$	-0.026	-0.001	1	0.296	-0.012	0.001	-0.003	0.016	0.005	-0.000	0.462	0.039	-0.137
$\|A_1\|^2$	-0.017	-0.001	0.296	1	-0.024	0.000	-0.002	-0.037	-0.003	0.000	0.383	-0.386	-0.217
p_2^θ	0.004	-0.001	-0.012	-0.024	1	-0.960	-0.001	-0.002	-0.001	0.004	-0.008	0.008	-0.000
p_4^θ	-0.005	0.001	0.001	0.000	-0.960	1	0.001	0.002	0.001	-0.004	-0.001	0.000	0.004
p_1^ϕ	0.009	0.002	-0.003	-0.002	-0.001	0.001	1	0.231	0.048	-0.001	0.001	-0.000	-0.002
p_2^ϕ	0.022	0.005	0.016	-0.037	-0.002	0.002	0.231	1	0.326	-0.001	-0.011	0.016	0.020
p_4^ϕ	0.000	-0.001	0.005	-0.003	-0.001	0.001	0.048	0.326	1	0.001	-0.002	-0.002	0.004
p_1^ψ	-0.000	-0.001	-0.000	0.000	0.004	-0.004	-0.001	-0.001	0.001	1	-0.002	0.001	0.002
β_s	0.042	0.004	0.462	0.383	-0.008	-0.001	0.001	-0.011	-0.002	-0.002	1	-0.058	-0.742
$\delta_\parallel$	-0.013	-0.001	0.039	-0.386	0.008	0.000	-0.000	0.016	-0.002	0.001	-0.058	1	0.263
$\delta_\perp$	-0.083	-0.004	-0.137	-0.217	-0.000	0.004	-0.002	0.020	0.004	0.002	-0.742	0.263	1
Glob.	0.825	0.629	0.667	0.672	0.959	0.959	0.233	0.405	0.328	0.013	0.947	0.549	0.882

Table 6.2: Correlation matrix for the parameters in the $B_s \to J/\psi\phi$ likelihood returned by the unconstrained untagged fit (Part 2 of 2).

Par.	M_s	σ_1	σ_2	f_m	A	f_s	$c\tau_s$	$\Delta\Gamma$	S_{ct}	f_+	f_{++}	f_-	$c\tau_+$	$c\tau_{++}$	$c\tau_-$	$\lvert A_0\rvert^2$	$\lvert A_1\rvert^2$	p_2^θ
M_s	1	-0.092	-0.091	0.094	-0.026	0.022	0.006	0.002	-0.012	0.001	0.040	0.004	-0.009	-0.027	-0.004	-0.000	0.007	0.001
σ_1	-0.092	1	0.625	-0.701	0.006	-0.202	-0.040	-0.011	0.025	0.002	-0.271	0.001	0.034	0.168	0.008	-0.024	-0.002	-0.026
σ_2	-0.091	0.625	1	-0.646	0.009	-0.377	-0.064	-0.032	0.047	-0.007	-0.521	-0.000	0.076	0.311	0.016	-0.048	-0.013	-0.045
f_m	0.094	-0.701	-0.646	1	-0.007	0.072	0.003	0.010	-0.014	-0.062	0.182	0.002	-0.036	-0.128	-0.004	0.009	0.002	0.012
A	-0.026	0.006	0.009	-0.007	1	-0.003	-0.002	-0.000	0.002	-0.002	-0.003	-0.001	0.001	0.001	0.001	0.000	-0.001	0.000
f_s	0.022	-0.202	-0.377	0.072	-0.003	1	0.133	0.001	-0.031	0.102	0.312	-0.010	-0.038	-0.176	-0.009	0.051	0.010	0.032
$c\tau_s$	0.006	-0.040	-0.064	0.003	-0.002	0.133	1	0.171	-0.015	0.072	0.028	-0.003	0.000	-0.081	-0.004	0.330	0.357	-0.003
$\Delta\Gamma$	0.002	-0.011	-0.032	0.010	-0.000	0.001	0.171	1	0.003	-0.011	0.037	-0.000	-0.006	-0.013	0.000	0.444	0.416	-0.009
S_{ct}	-0.012	0.025	0.047	-0.014	0.002	-0.031	-0.015	0.003	1	-0.385	-0.118	-0.415	0.224	0.096	0.290	-0.003	-0.004	-0.004
f_+	0.001	0.002	-0.007	-0.062	-0.002	0.102	0.072	-0.011	-0.385	1	-0.004	0.186	-0.401	0.036	-0.146	0.015	0.011	0.008
f_{++}	0.040	-0.271	-0.521	0.182	-0.003	0.312	0.028	0.037	-0.118	-0.004	1	0.048	-0.616	-0.760	-0.050	0.047	0.016	0.028
f_-	0.004	0.001	-0.000	0.002	-0.001	-0.010	-0.003	-0.000	-0.415	0.186	0.048	1	-0.133	-0.047	-0.591	-0.002	0.002	-0.000
$c\tau_+$	-0.009	0.034	0.076	-0.036	0.001	-0.038	0.000	-0.006	0.224	-0.401	-0.616	-0.133	1	0.555	0.099	-0.003	-0.010	0.003
$c\tau_{++}$	-0.027	0.168	0.311	-0.128	0.001	-0.176	-0.081	-0.013	0.096	0.036	-0.760	-0.047	0.555	1	0.042	-0.031	-0.019	-0.011
$c\tau_-$	-0.004	0.008	0.016	-0.004	0.001	-0.009	-0.004	0.000	0.290	-0.146	-0.050	-0.591	0.099	0.042	1	-0.001	-0.002	-0.001
$\lvert A_0\rvert^2$	-0.000	-0.024	-0.048	0.009	0.000	0.051	0.330	0.444	-0.003	0.015	0.047	-0.002	-0.003	-0.031	-0.001	1	0.225	-0.008
$\lvert A_1\rvert^2$	0.007	-0.002	-0.013	0.002	-0.001	0.010	0.357	0.416	-0.004	0.011	0.016	0.002	-0.010	-0.019	-0.002	0.225	1	-0.023
p_2^θ	0.001	-0.026	-0.045	0.012	0.000	0.032	-0.003	-0.009	-0.004	0.008	0.028	-0.000	0.003	-0.011	-0.001	-0.008	-0.023	1
p_4^θ	-0.001	0.021	0.036	-0.008	-0.000	-0.027	-0.003	-0.001	0.003	-0.008	-0.022	0.000	-0.004	0.008	0.001	-0.001	0.002	-0.960
p_1^ϕ	-0.003	0.013	0.023	-0.013	0.000	-0.010	-0.002	-0.002	0.005	-0.002	-0.021	-0.002	0.010	0.014	0.002	-0.004	-0.001	-0.003
p_2^ϕ	-0.009	0.037	0.077	-0.028	0.002	-0.050	-0.018	-0.003	0.010	-0.012	-0.056	-0.002	0.017	0.033	0.003	0.019	-0.033	-0.004
p_4^ϕ	-0.001	0.005	0.018	-0.002	0.001	-0.017	-0.006	0.002	0.001	-0.003	-0.012	0.001	-0.001	0.006	0.000	0.004	-0.002	-0.001
p_1^ψ	0.002	0.001	0.005	0.003	-0.001	-0.011	-0.004	-0.001	-0.002	-0.003	-0.008	0.001	0.002	0.001	-0.000	-0.001	-0.002	-0.001
β_s	0.008	-0.018	-0.029	0.019	-0.002	0.021	0.619	0.111	-0.015	0.016	0.023	0.006	-0.013	-0.046	-0.005	0.415	0.366	-0.010
$\delta_\parallel$	-0.006	-0.021	-0.030	0.011	-0.001	0.025	-0.015	-0.015	-0.001	0.003	0.021	-0.001	0.010	-0.009	-0.000	0.044	-0.336	0.008
$\delta_\perp$	0.004	0.009	0.015	-0.004	-0.000	-0.007	0.091	-0.050	-0.006	0.003	-0.008	0.003	-0.004	0.001	-0.002	0.028	0.044	-0.004
ϵ_s	-0.007	0.049	0.094	-0.022	0.002	-0.089	-0.026	-0.004	0.009	-0.024	-0.088	0.001	0.016	0.050	0.003	-0.011	-0.009	-0.005
$S_{\mathcal{D}_s}$	-0.002	0.005	0.008	-0.006	0.001	-0.000	0.025	-0.007	0.002	0.001	-0.002	-0.001	-0.001	-0.001	0.001	0.017	0.008	-0.001
ϵ_s^{bkg}	-0.001	0.024	0.043	-0.005	-0.001	-0.049	-0.020	0.003	0.003	-0.013	-0.027	0.002	-0.001	0.015	0.001	-0.007	0.004	-0.006
$S_{\mathcal{D}_s}^{\text{bkg}}$	0.000	0.009	0.016	-0.010	0.000	-0.008	-0.013	-0.001	-0.001	-0.004	-0.010	0.000	0.006	0.006	0.000	-0.009	-0.003	-0.002
ϵ_o	0.004	-0.023	-0.049	0.027	-0.001	0.014	-0.000	0.003	-0.002	-0.006	0.032	0.000	-0.009	-0.018	-0.001	0.006	0.000	0.006
$S_{\mathcal{D}_o}$	-0.000	-0.000	-0.001	0.000	-0.000	0.000	0.012	0.004	0.001	0.000	0.000	-0.001	0.000	-0.001	0.000	0.010	0.005	-0.000
ϵ_o^{bkg}	-0.006	0.039	0.078	-0.031	0.001	-0.048	-0.012	-0.003	0.005	-0.004	-0.058	0.001	0.011	0.033	0.002	-0.010	-0.001	-0.008
$S_{\mathcal{D}_o}^{\text{bkg}}$	-0.001	0.002	0.002	0.000	-0.000	-0.003	-0.003	0.000	0.001	0.003	-0.010	-0.000	0.005	0.008	0.000	-0.000	-0.001	-0.000
Δm_s	0.001	0.007	0.015	-0.004	0.000	-0.007	0.145	0.008	0.016	-0.004	-0.014	-0.006	0.006	-0.000	0.005	0.088	0.093	-0.004
Glob.	0.109	0.742	0.848	0.801	0.027	0.499	0.659	0.597	0.536	0.679	0.896	0.645	0.826	0.781	0.593	0.589	0.644	0.959

Table 6.3: Correlation matrix for the parameters in the $B_s \to J/\psi\phi$ likelihood returned by the unconstrained tagged fit (Part 1 of 2).

Par.	p_4^θ	p_1^ϕ	p_2^ϕ	p_4^ϕ	p_1^ψ	β_s	$\delta_\parallel$	$\delta_\perp$	ϵ_s	$S_{\mathcal{D}_s}$	$\epsilon_s^{\rm bkg}$	$S_{\mathcal{D}_s}^{\rm bkg}$	ϵ_o	$S_{\mathcal{D}_o}$	$\epsilon_o^{\rm bkg}$	$S_{\mathcal{D}_o}^{\rm bkg}$	Δm_s
M_s	-0.001	-0.003	-0.009	-0.001	0.002	0.008	-0.006	0.004	-0.007	-0.002	-0.001	0.000	0.004	-0.000	-0.006	-0.001	0.001
σ_1	0.021	0.013	0.037	0.005	0.001	-0.018	-0.021	0.009	0.049	0.005	0.024	0.009	-0.023	-0.000	0.039	0.002	0.007
σ_2	0.036	0.023	0.077	0.018	0.005	-0.029	-0.030	0.015	0.094	0.008	0.043	0.016	-0.049	-0.001	0.078	0.002	0.015
f_m	-0.008	-0.013	-0.028	-0.002	0.003	0.019	0.011	-0.004	-0.022	-0.006	-0.005	-0.010	0.027	0.000	-0.031	0.000	-0.004
A	-0.000	0.000	0.002	0.001	-0.001	-0.002	-0.001	-0.000	0.002	0.001	-0.001	0.000	-0.001	-0.000	0.001	-0.000	0.000
f_s	-0.027	-0.010	-0.050	-0.017	-0.011	0.021	0.025	-0.007	-0.089	-0.000	-0.049	-0.008	0.014	0.000	-0.048	-0.003	-0.007
$c\tau_s$	-0.003	-0.002	-0.018	-0.006	-0.004	0.619	-0.015	0.091	-0.026	0.025	-0.020	-0.013	-0.000	0.012	-0.012	-0.003	0.145
$\Delta\Gamma$	-0.001	-0.002	-0.003	0.002	-0.001	0.111	-0.015	-0.050	-0.004	-0.007	0.003	-0.001	0.003	0.004	-0.003	0.000	0.008
S_{ct}	0.003	0.005	0.010	0.001	-0.002	-0.015	-0.001	-0.006	0.009	0.002	0.003	-0.001	-0.002	0.001	0.005	0.001	0.016
f_+	-0.008	-0.002	-0.012	-0.003	-0.003	0.016	0.003	0.003	-0.024	0.001	-0.013	-0.004	-0.006	0.000	-0.004	0.003	-0.004
f_{++}	-0.022	-0.021	-0.056	-0.012	-0.008	0.023	0.021	-0.008	-0.088	-0.002	-0.027	-0.010	0.032	0.000	-0.058	-0.010	-0.014
f_-	0.000	-0.002	-0.002	0.001	0.001	0.006	-0.001	0.003	0.001	-0.001	0.002	0.000	0.000	-0.001	0.001	-0.000	-0.006
$c\tau_+$	-0.004	0.010	0.017	-0.001	0.002	-0.013	0.010	-0.004	0.016	-0.001	-0.001	0.006	-0.009	0.000	0.011	0.005	0.006
$c\tau_{++}$	0.008	0.014	0.033	0.006	0.001	-0.046	-0.009	0.001	0.050	-0.001	0.015	0.006	-0.018	-0.001	0.033	0.008	-0.000
$c\tau_-$	0.001	0.002	0.003	0.000	-0.000	-0.005	-0.000	-0.002	0.003	0.001	0.001	0.000	-0.001	0.000	0.002	0.000	0.005
$\|A_0\|^2$	-0.001	-0.004	0.019	0.004	-0.001	0.415	0.044	0.028	-0.011	0.017	-0.007	-0.009	0.006	0.010	-0.010	-0.000	0.088
$\|A_1\|^2$	0.002	-0.001	-0.033	-0.002	-0.002	0.366	-0.336	0.044	-0.009	0.008	0.004	-0.003	0.000	0.005	-0.001	-0.001	0.093
p_2^θ	-0.960	-0.003	-0.004	-0.001	-0.001	-0.010	0.008	-0.004	-0.005	-0.001	-0.006	-0.002	0.006	-0.000	-0.008	-0.000	-0.004
p_4^θ	1	0.003	0.003	0.001	0.000	0.002	-0.002	0.003	0.005	-0.000	0.004	0.001	-0.005	-0.000	0.007	-0.000	0.002
p_1^ϕ	0.003	1	0.255	0.057	-0.002	-0.003	-0.002	0.003	0.005	0.000	-0.001	0.000	-0.002	0.000	0.003	0.001	0.000
p_2^ϕ	0.003	0.255	1	0.325	0.002	-0.004	0.012	0.001	0.012	0.002	0.006	0.002	0.001	0.000	0.003	0.001	0.001
p_4^ϕ	0.001	0.057	0.325	1	0.000	-0.001	-0.002	-0.000	-0.000	0.001	0.005	0.000	0.002	0.000	-0.001	0.000	-0.000
p_1^ψ	0.000	-0.002	0.002	0.000	1	-0.002	-0.000	0.000	0.004	0.000	0.000	-0.001	0.001	-0.000	0.000	0.001	-0.000
β_s	0.002	-0.003	-0.004	-0.001	-0.002	1	0.040	0.161	-0.005	0.041	-0.002	-0.019	0.004	0.020	-0.005	-0.005	0.237
$\delta_\parallel$	-0.002	-0.002	0.012	-0.002	-0.000	0.040	1	0.000	-0.009	-0.012	-0.001	-0.002	0.001	0.005	-0.004	-0.003	-0.029
$\delta_\perp$	0.003	0.003	0.001	-0.000	0.000	0.161	0.000	1	0.003	0.031	-0.000	-0.004	0.002	-0.020	-0.001	-0.002	0.246
ϵ_s	0.005	0.005	0.012	-0.000	0.004	-0.005	-0.009	0.003	1	-0.001	-0.085	0.003	0.004	-0.000	0.005	0.002	0.003
$S_{\mathcal{D}_s}$	-0.000	0.000	0.002	0.001	0.000	0.041	-0.012	0.031	-0.001	1	0.001	0.005	-0.000	-0.018	0.000	0.000	-0.068
$\epsilon_s^{\rm bkg}$	0.004	-0.001	0.006	0.005	0.000	-0.002	-0.001	-0.000	-0.085	0.001	1	0.000	-0.007	0.000	0.011	-0.000	-0.000
$S_{\mathcal{D}_s}^{\rm bkg}$	0.001	0.000	0.002	0.000	-0.001	-0.019	-0.002	-0.004	0.003	0.005	0.000	1	-0.000	-0.001	0.001	0.001	-0.005
ϵ_o	-0.005	-0.002	0.001	0.002	0.001	0.004	0.001	0.002	0.004	-0.000	-0.007	-0.000	1	0.000	-0.112	0.000	0.000
$S_{\mathcal{D}_o}$	-0.000	0.000	0.000	0.000	-0.000	0.020	0.005	-0.020	-0.000	-0.018	0.000	-0.001	0.000	1	0.000	0.001	0.004
$\epsilon_o^{\rm bkg}$	0.007	0.003	0.003	-0.001	0.000	-0.005	-0.004	-0.001	0.005	0.000	0.011	0.001	-0.112	0.000	1	0.000	0.000
$S_{\mathcal{D}_o}^{\rm bkg}$	-0.000	0.001	0.001	0.000	0.001	-0.005	-0.003	-0.002	0.002	0.000	-0.000	0.001	0.000	0.001	0.000	1	-0.001
Δm_s	0.002	0.000	0.001	-0.000	-0.000	0.237	-0.029	0.246	0.003	-0.068	-0.000	-0.005	0.000	0.004	0.000	-0.001	1
Glob.	0.959	0.256	0.411	0.326	0.913	0.715	0.409	0.280	0.158	0.104	0.114	0.027	0.120	0.036	0.138	0.013	0.330

Table 6.4: Correlation matrix for the parameters in the $B_s \to J/\psi\phi$ likelihood returned by the unconstrained tagged fit (Part 2 of 2).

Chapter 7

Summary and Discussion

We conclude this dissertation by presenting a summary of our final results. When presenting numerical results, we quote first the central value extracted from the likelihood fit followed by the statistical uncertainty, followed by the systematic uncertainty, when the latter is calculated. The results are followed by a discussion of the impact of these measurements, comparisons with other experimental results, and a brief examination of the future outlook and prospects of these measurements.

We measure the width difference $\Delta\Gamma$ and the *CP*-violating phase $2\beta_s$ in $B_s \to J/\psi\phi$ decays. We find that our result is compatible with the Standard Model prediction at the at the level of one standard deviation. The use of the $B_d \to J/\psi K^*$ decay mode as a cross-check, and of the likelihood ratio method as a robust framework give us confidence in our result.

Our result itself is not the only accomplishment of the work presented here. The analysis techniques developed for the measurement are expected to be a great boon to future endeavors in this field at the Tevatron and at the LHC. Finally, we expect that these results are of great interest in the theoretical community, which has had no experimental input on $2\beta_s$ until very recently.

7.1 Final Results

In the time-dependent analysis of $B_d \to J/\psi K^*$ decays, we obtain the following results:

$$\begin{aligned} c\tau_d &= (455.3 \pm 6.4 \pm 5.3)\ \mu\text{m}, \\ |A_0|^2 &= 0.560 \pm 0.009 \pm 0.004, \\ |A_\parallel|^2 &= 0.210 \pm 0.012 \pm 0.006, \\ \delta_\parallel &= -2.946 \pm 0.083 \pm 0.004, \\ \delta_\perp &= 2.941 \pm 0.056 \pm 0.002. \end{aligned}$$

In the time-dependent analysis of $B_s \to J/\psi\phi$ decays assuming $2\beta_s = 0$, we obtain

the following results:

$$
\begin{aligned}
c\tau_s &= (455.5^{+13.1\ +5.0}_{-12.9\ -5.1})\ \mu\text{m}, \\
\Delta\Gamma_s &= (0.071^{+0.064}_{-0.059}\ \pm 0.007)\ \text{ps}^{-1}, \\
|A_0|^2 &= 0.529 \pm 0.020 \pm 0.007, \\
|A_\parallel|^2 &= 0.241 \pm 0.029 \pm 0.006, \\
\delta_\parallel &= 2.63^{+0.18}_{-0.29}\ \pm 0.10.
\end{aligned}
$$

In the measurement of the *CP*-violating phase without the use of flavor tagging, we find that, assuming the Standard Model predictions of $2\beta_s$ and $\Delta\Gamma$, the probability of a deviation as large as the level of the observed data is 22%.

In the measurement of the *CP*-violating phase with the use of flavor tagging, we find that, assuming the Standard Model predictions of $2\beta_s$ and $\Delta\Gamma$, the probability of a deviation as large as the level of the observed data is 33%.

7.2 Comparisons with Other Results

7.2.1 Time-dependent Angular Analysis of $B_d \to J/\psi K^*$

The analysis in the $B_d \to J/\psi K^*$ mode yields a result that is competitive with other measurements at the Tevatron and at B factories. While not the best single-measurement performance in terms of the overall uncertainty on each parameter, it rivals the best measurements to date. One must keep in mind that the B factories perform this analysis with a data sample corresponding to an integrated luminosity that is *two orders of magnitude* larger than the sample we use in our analysis. In light of this fact, our measurement is undoubtedly a success. Table 7.1 lists the results of our analysis alongside the latest results from B factories at the time of writing.

	CDF (2007)	BaBar (2007)
$\lvert A_0\rvert^2$	$0.569 \pm 0.009 \pm 0.004$	$0.556 \pm 0.009 \pm 0.010$
$\lvert A_\parallel\rvert^2$	$0.211 \pm 0.012 \pm 0.006$	$0.211 \pm 0.010 \pm 0.006$
$\delta_\parallel$	$-2.956 \pm 0.083 \pm 0.004$	$-2.930 \pm 0.080 \pm 0.040$
$\delta_\perp$	$2.971 \pm 0.056 \pm 0.002$	$2.910 \pm 0.050 \pm 0.030$

Table 7.1: The results of the most recent measurements of the angular amplitudes in and $B_d \to J/\psi K^*$ decays. The first uncertainty listed for each parameter is statistical, and the second is from systematic effects. Collected from Reference [44].

As for the distinguishing characteristics of the measurement itself, this analysis marks two firsts in the history of this measurement at CDF. This is the first time that the swapped $K \leftrightarrow \pi$ candidates are explicitly targeted for removal from our analysis sample, with good success. Second, this is the first time that the S-wave contribution of $B_d \to J/\psi K\pi$ decays is taken into account directly in the fitting model. We see the improvement in the model of the angular space most readily.

Naturally, in addition to this, the consequence of this measurement is to increase our confidence in our single-lifetime fitting model, and in our treatment of the angular acceptance of the CDF detector in the transversity basis. These are crucial elements for the analysis performed in the $B_s \to J/\psi\phi$ mode, and the B_d sample acts as a good control sample for these tasks.

7.2.2 Time-dependent Angular Analysis of $B_s \to J/\psi\phi$

We compare the results of the time-dependent angular analysis of $B_s \to J/\psi\phi$ to the latest results from the DØ collaboration, which we list here again [45]:

$$\begin{aligned} \Delta\Gamma_s &= 0.12^{+0.08}_{-0.10} \pm 0.02 \text{ ps}^{-1}, \\ |A_0|^2 - |A_\parallel|^2 &= 0.38 \pm 0.05, \\ |A_\perp|^2 &= 0.45 \pm 0.05, \\ \delta_\parallel &= 2.6 \pm 0.4, \end{aligned}$$

We see that the uncertainties extracted from our likelihood fit are an improvement, no doubt in large part because of the larger data sample used to perform the analysis (1.7 fb^{-1} versus 1.1 fb^{-1}), but also in part due to our improved neural network selection algorithm.

7.2.3 Measurement of the *CP* Phase $2\beta_s$ in $B_s \to J/\psi\phi$

The comparison of our result in the measurement of $2\beta_s$ to other experimental results is complicated by our choice of the likelihood ratio method to perform the measurement. At first glance, the result of the DØ collaboration described in Reference [45] seems to be considerably more precise, with a much smaller uncertainty:

$$\begin{aligned} \Delta\Gamma_s &= 0.17 \pm 0.09 \pm 0.02 \text{ ps}^{-1}, \\ 2\beta_s &= 0.79 \pm 0.56 \, ^{+0.01}_{-0.14}. \end{aligned}$$

In particular, the DØ result excludes a significant portion of the $2\beta_s$ axis at the 90% (2σ) level. On the other hand, the confidence region we show in Section 6.5.1 does no such thing. However, the comparison is misleading because the LR method takes into account the failure of the maximum likelihood method to return reliable, unbiased estimates of the physical parameters we measure. In fact, the uncertainties returned by the nominal ML fit in our analysis are comparable in size to the uncertainties quoted in the DØ result. However, we do not quote these uncertainties because they are equally unreliable and subject to biases in the ML fit which are dependent on the true values of $\Delta\Gamma$ and $2\beta_s$.

It is important to note that the differences between the CDF and DØ detectors and data collection method, however real they are, have nothing to do with the apparent discrepancy between the performance of the analyses performed at the two collaborations. The use of the LR method is motivated by the observed failure of the maximum likelihood estimator to return reliable unbiased results in the case of the

CP violation analysis in $B_s \to J/\psi\phi$. This failure is observed by an examination of the likelihood function itself. This is essentially irremovable, except with a very large sample of signal events. There is no reason to suspect that this observed failure of the ML method only affects the CDF measurement because the numbers of signal events are comparable (per unit fb^{-1}) between the two measurements.

7.3 Discussion

The final results quoted in this chapter have at least one thing in common: the statistical uncertainty dominates the systematic uncertainty. This was already the case in the first analysis of the B_s width difference performed at CDF [42]. This situation has remained the same in part because the systematic uncertainties associated with these measurements have been reduced at the same time that the statistical uncertainties have decreased thanks to the larger number of signal events in our data sample. This is no doubt a positive outcome.

The 2004 measurement of $\Delta\Gamma$ at CDF concluded that assuming the Standard Model predictions of $2\beta_s$ and $\Delta\Gamma$, the probability of a deviation as large as the level of the observed data is approximately 0.3%. This is not the case of our measurement. Nevertheless, the result of our tagged analysis excludes approximately half of the $\Delta\Gamma$–$2\beta_s$ plane.

Future prospects for this measurement are promising. This measurement is not only of interest at the Tevatron, but is also slated to be a highlight of the LHCb B physics program at CERN. Figure 7-1 shows the likelihood profile expected in a data sample with integrated luminosity of 10 fb^{-1}, using a pseudo-experiment generated with an input value of $2\beta_s$ equal to 0.4.

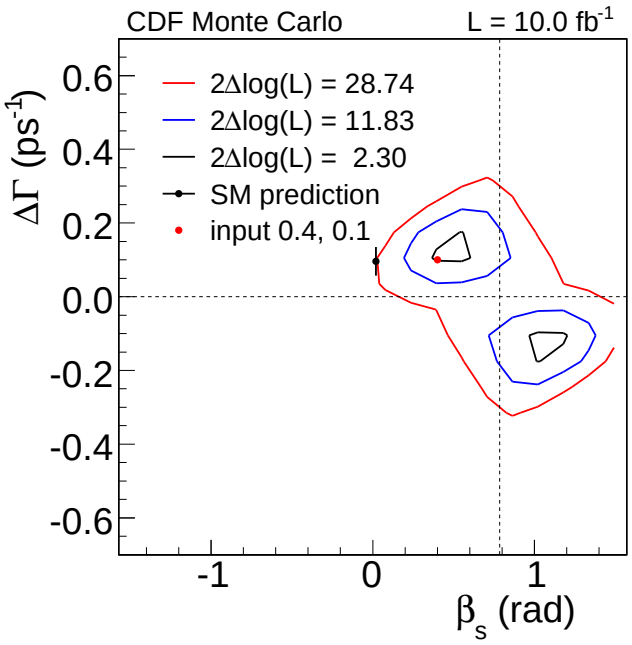


Figure 7-1: Likelihood profiles in the $\Delta\Gamma$–$2\beta_s$ plane obtained from a pseudo-experiment with a sample size corresponding to 10 fb^{-1}. The input value of $\Delta\Gamma$ is 0.1, and of $2\beta_s$ is 0.4.

Appendix A

Fully Expanded Decay Rates

The differential decay rates for $B_s \to J/\psi\phi$ and $\overline{B_s} \to J/\psi\phi$ expressed in terms of angular and time dependence have several peculiarities that warrant close examination. First, in the individual expressions for the B_s and $\overline{B_s}$ decay are invariant under a simultaneous change of sign of $\cos 2\beta_s$, $\Delta\Gamma$, $\cos(\delta_\perp - \delta_\parallel)$, and $\cos\delta_\perp$. The combined expression for equal proportions of B_s and $\overline{B_s}$ is invariant under a simultaneous change of sign of $\cos 2\beta_s$ and $\Delta\Gamma$, and also under an independent change of sign of $\cos(\delta_\perp - \delta_\parallel)$ and $\cos\delta_\perp$, resulting in a four-fold ambiguity in the $\Delta\Gamma$–$2\beta_s$ plane. Second, when we combine the two decay rates as in the case of an untagged measurement, $\delta_\perp$ drops out of the expression if $\beta_s = 0$, and $\delta_\perp$ and β_s drop out of the expression if $\Delta\Gamma = 0$.

This examination is easier to carry out when the decay rates are written in fully expanded form. For the sake of clarity in the main text, this expansion is done in this appendix. We begin with the decay rate for the tagged case, when we determine the initial state flavor of the B_s ($\overline{B_s}$) meson, taking the upper (lower) sign:

$$\begin{aligned}\frac{d^4\,\mathcal{P}_{\rm tag}(\vec{\omega},t)}{d\vec{\omega}\,dt} \propto\ & |A_0|^2(\mathcal{T}_+ \mp \mathcal{T}_{ss})\,f_1 + |A_\parallel|^2(\mathcal{T}_+ \mp \mathcal{T}_{ss})\,f_2 + |A_\perp|^2(\mathcal{T}_- \pm \mathcal{T}_{ss})\,f_3 \\ & + |A_0||A_\parallel|\cos(\delta_2-\delta_1)(\mathcal{T}_+ \mp \mathcal{T}_s)\,f_5 \\ & - |A_\parallel||A_\perp|\,(\cos\delta_1\,\mathcal{T}_s \mp \sin\delta_1\,\mathcal{T}_c \pm \cos\delta_1\,\mathcal{T}_{sc})\ f_4 \\ & - |A_0||A_\perp|\,(\cos\delta_2\,\mathcal{T}_s \mp \sin\delta_2\,\mathcal{T}_c \pm \cos\delta_2\,\mathcal{T}_{sc})\ f_6,\end{aligned} \tag{A.1}$$

where the $\mathcal{T}$ and f_i functions are defined in Section 1.4, and $\delta_1 \equiv \delta_\perp - \delta_\parallel$, $\delta_2 \equiv \delta_\perp$. If we combine upper and lower sign, we obtain:

$$\begin{aligned}\frac{d^4\,\mathcal{P}_{\rm unt}(\vec{\omega},t)}{d\vec{\omega}\,dt} \propto\ & |A_0|^2\,\mathcal{T}_+\,f_1 + |A_\parallel|^2\,\mathcal{T}_+\,f_2 + |A_\perp|^2\,\mathcal{T}_-\,f_3 + |A_0||A_\parallel|\cos(\delta_\parallel)\,\mathcal{T}_+\,f_5 \\ & - |A_\parallel||A_\perp|\sin(2\beta_s)\frac{e^{-\Gamma_H t} - e^{-\Gamma_L t}}{2}\cos(\delta_\perp-\delta_\parallel)\,f_4 \\ & - |A_0||A_\perp|\sin(2\beta_s)\frac{e^{-\Gamma_H t} - e^{-\Gamma_L t}}{2}\ \cos(\delta_\perp)\,f_6.\end{aligned} \tag{A.2}$$

Equation A.1 expands to:

$$\begin{aligned}
\frac{d^4\ \mathcal{P}_{\rm tag}(\vec{\omega},t)}{d\vec{\omega}\ dt} \propto \\
|A_0|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2} \mp \sin 2\beta_s \sin\Delta m t\right] f_1 + \\
|A_\parallel|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2} \mp \sin 2\beta_s \sin\Delta m t\right] f_2 + \\
|A_\perp|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} + |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2} \pm \sin 2\beta_s \sin\Delta m t\right] f_3 + \\
|A_0||A_\parallel| e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2} \mp \sin 2\beta_s \sin\Delta m t\right] \cos(\delta_2-\delta_1) f_5 + \\
|A_\parallel||A_\perp| e^{-\Gamma t}\left[\pm\sin\delta_1\cos\Delta m t \mp \cos\delta_1\cos 2\beta_s\sin\Delta m t + \cos\delta_1\sin 2\beta_s\sinh\frac{\Delta\Gamma t}{2}\right] f_4 + \\
|A_0||A_\perp| e^{-\Gamma t}\left[\pm\sin\delta_2\cos\Delta m t \mp \cos\delta_2\cos 2\beta_s\sin\Delta m t + \cos\delta_2\sin 2\beta_s\sinh\frac{\Delta\Gamma t}{2}\right] f_6.
\end{aligned} \tag{A.3}$$

Equation A.2 expands to:

$$\begin{aligned}
\frac{d^4\ \mathcal{P}_{\rm unt}(\vec{\omega},t)}{d\vec{\omega}\ dt} \propto \\
|A_0|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2}\right] f_1 + \\
|A_\parallel|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2}\right] f_2 + \\
|A_\perp|^2 e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} + |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2}\right] f_3 + \\
|A_0||A_\parallel| e^{-\Gamma t}\left[\cosh\frac{\Delta\Gamma t}{2} - |\cos 2\beta_s|\sinh\frac{|\Delta\Gamma| t}{2}\right] \cos(\delta_2-\delta_1) f_5 + \\
|A_\parallel||A_\perp| e^{-\Gamma t}\left[\cos\delta_1\sin 2\beta_s\sinh\frac{\Delta\Gamma t}{2}\right] f_4 + \\
|A_0||A_\perp| e^{-\Gamma t}\left[\cos\delta_2\sin 2\beta_s\sinh\frac{\Delta\Gamma t}{2}\right] f_6.
\end{aligned} \tag{A.4}$$

In these forms, it is simpler to visualize the effect of setting β_s to zero, or $\Delta\Gamma$ to zero. Likewise, it is a simple matter to visualize the effect of the transformations on β_s, $\Delta\Gamma$, δ_1 and δ_2 as they all appear explicitly in the expanded expressions.

Appendix B

Di-Muon Trigger

The trigger paths used to collect the data analyzed in this dissertation fall into two main groups: CMU-CMU paths, where both muons are incident on the CMU subdetector (in the most central region), and CMU-CMX paths, where one of the two muons is incident on the CMU and the other on the CMX, which extends muon coverage in η. The trigger name encodes the other criteria when they are used to accept or reject an event, including the following:

- **PTx**: p_T of the XFT track needs to be greater than x GeV
- **DPHIx**: The difference in φ of the two muons must be smaller than x degrees.
- **OPPQ**: Both muons need to have opposite charge.
- **xMTy**: The transverse mass for the two muons must be between x and y GeV/c^2.
- **PSx, DPS**: The trigger is either pre-scaled by the factor x or dynamically pre-scaled.

Triggers that are prescaled by a factor x only send to Level 2 one out of x events that pass Level 1 requirements. This is done lest we overwhelm the DAQ system when luminosity is high and event rate is higher than the rate that can be processed by the trigger without incurring dead time. Dynamically prescaled triggers can adjust the factor x in real-time, depending on whether luminosity is high enough to warrant this protective measure. They constitute a more optimal version of the basic prescaled triggers.

Table B.1 lists all the trigger paths that were used in this analysis.

Level 2 Trigger	Level 3 Trigger
L2_AUTO_L1_CMU1.5_PT1.5_&_CMX1.5_PT2	JPSI_CMU1.5_CMX20
L2_AUTO_L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX	JPSI_CMU1.5_CMX22
L2_AUTO_L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX_PS0	JPSI_CMU1.5_CMX23
L2_AUTO_L1_TWO_CMU1.5_PT1.5	JPSI_CMU1.5_CMX2_ALLPHI
L2_PS100_L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX	JPSI_CMU1.5_CMX2_NOL2
L2_PS100_L1_TWO_CMU1.5_PT1.5	JPSI_CMU1.5_CMX2_NOL2
L2_PS10_L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX	JPSI_CMU1.5_CMX2
L2_PS10_L1_TWO_CMU1.5_PT1.5	JPSI_CMU2_CMX2_PS10
L2_TWO_CMU1.5_PT1.5_DPHI120_OPPQ	JPSI_CMU2_CMX2_PS2
L2_TWO_CMU1.5_PT1.5_DPHI120_OPPQ_DPS	JPSI_CMU2_CMX2_PS50
L2_TWO_CMU1.5_PT2_DPHI120_OPPQ	JPSI_CMUCMU1.5
L2_TWO_CMU1.5_PT2_DPHI120_OPPQ_PS10	JPSI_CMUCMU1.5_ALLPHI
L2_AUTO_L1_CMUP6_PT4	JPSI_CMUCMU1.5_NOL2
L2_CMU1.5_PT1.5_&_CMX1.5_PT2_DPHI120_OPPQ	JPSI_CMUCMU2
L2_CMU1.5_PT1.5_&_CMX1.5_PT2_DPHI120_OPPQ_DPS	JPSI_CMUCMU2_PS10
L2_CMU1.5_PT1.7_&_CMU1.5_PT3_1.7MT7_DPS	JPSI_CMUCMU2_PS2
L2_CMU1.5_PT1.7_&_CMU1.5_PT3_1.7MT7_LUMI_185	JPSI_CMUCMU2_PS50
L2_CMU1.5_PT1.7_&_CMX1.5_PT3_1.7MT7_DPS	JPSI_CMUCMU_MT_DPS
L2_CMU1.5_PT1.7_&_CMX1.5_PT3_1.7MT7_LUMI_185	JPSI_CMUCMU_MT_LUMI_185
L2_CMU1.5_PT2_&_CMX1.5_PT2_DPHI120_OPPQ_PS10	JPSI_CMUCMX3_MT_DPS
L2_CMU1.5_PT2_&_CMX1.5_PT2_DPHI120_OPPQ_PS2	JPSI_CMUCMX3_MT_LUMI_185
L2_CMU1.5_PT2_&_CMX1.5_PT2_DPHI120_OPPQ_PS50	JPSI_CMUP4_CMU1.5
L2_CMUP6_PT8	JPSI_CMUP4_CMU1.5_L2_DPS
L2_CMUP6_PT8_DPS	JPSI_CMUP4_CMU_L2_DPS
L2_CMX1.5_PT2_&_CMU1.5_PT3_1.7MT7_DPS	JPSI_CMUP4_CMX2
L2_CMX1.5_PT2_&_CMU1.5_PT3_1.7MT7_LUMI_185	JPSI_CMUP4_CMX2_L2_DPS
L2_DPS_L1_CMUP6_PT4	JPSI_CMUP4_CMX_L2_DPS
L2_PS200_L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX	JPSI_CMUPCMU_HIGHPT
L2_PS200_L1_TWO_CMU1.5_PT1.5	JPSI_CMUPCMU_HIGHPT_DPS
L2_PS500_L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX	JPSI_CMUPCMX_HIGHPT
L2_RL20HZ_L1_CMUP6_PT4	JPSI_CMUPCMX_HIGHPT_DPS
L2_TRK8_L1_CMUP6_PT4	JPSI_CMXCMU3_MT_DPS
L2_TWO_CMU1.5_PT2_DPHI120_OPPQ_PS2	JPSI_CMXCMU3_MT_LUMI_185
L2_TWO_CMU1.5_PT2_DPHI120_OPPQ_PS50	EXPRESS_JPSI_CMUCMU

Table B.1: Summary of trigger paths used for data collection in this analysis. The nomenclature is explained in the text.

Appendix C

Particle Identification

Particle identification, or PID, plays an important role in several elements of the analysis presented in this dissertation. Separation of kaons from pions and of leptons from lepton fakes is of particular importance. Distinguishing kaons from pions is important for the selection algorithms that rely in part on PID to exclude misreconstructed $B_s \to J/\psi\phi$ or $B_d \to J/\psi K^*$ decays, and it is further essential in the same-side kaon tagging algorithms. Lepton identification, or lepton ID, is also used in candidate selection, and in the opposite-side lepton taggers. Hadron PID is possible via measurement of the specific ionization per unit track length in the CDF drift chamber ($\mathrm{d}E/\mathrm{d}x$ from the COT) as well as the Time-of-Flight measured in the TOF subdetector. Lepton ID uses measurements from several subdetectors combined in a likelihood expression representing the probability that the lepton is real rather than a fake. We focus the discussion of lepton ID on muon identification because of the additional importance it has for decay modes involving a $J/\psi \to \mu\mu$ final state, and refer the reader to Reference [115] for additional details on electron identification.

C.1 Particle ID using $\mathrm{d}E/\mathrm{d}x$ in the COT

As a charged particle passes through a gas, it loses energy through ionization. The energy loss per unit length is described by the Bethe-Bloch formula:

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi N e^4}{m_e c^2 \beta^2} q^2 \left[\ln \frac{2 m_e c^2 (\beta\gamma)^2}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{C.1}$$

where N is the number density of electrons in the material, e, m_e, q and βc are the electron charge and mass, and the particle's charge and speed respectively, I is the mean excitation energy of the atoms in the medium, and $\delta(\beta\gamma)$ is a relativistic density effect correction. The parameters N, e, m_e and I are known for a gas of given properties, such as the mixture present in the CDF drift chamber. Combining a measurement of $\mathrm{d}E/\mathrm{d}x$ with a measurement of momentum (yielding q by observation of curvature) permits a determination of the particle's mass, and thus identity.

The COT, built for speed and accuracy of tracking, was not designed primarily for particle identification. For example, the drift cells are made to be small to improve

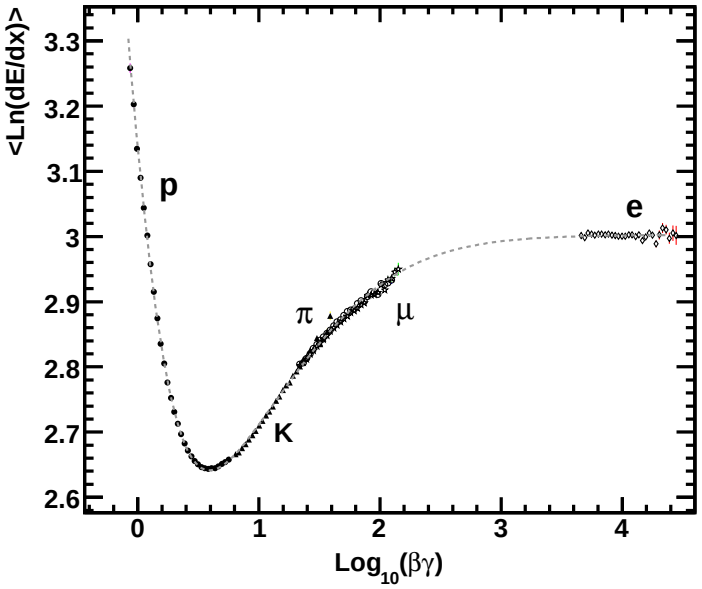


Figure C-1: Universal curve showing the relation between $\ln \mathrm{d}E/\mathrm{d}x$ and $\log \beta\gamma$ of reconstructed particles at CDF. Different ranges in $\beta\gamma$ are dominated by different particle species.

tracking accuracy, but this limits the amount of charge collected at sense wires, which in turns leads to larger statistical fluctuations in the measurement of energy loss by the passing particle. Nevertheless, the PID performance is improved by careful calibration using pure samples of pions, kaons, protons, electrons and muons coming from decays including $D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi^+$, $\Lambda^0 \to p\pi^-$, $J/\psi \to \mu^+\mu^-$, and conversion electrons.

A *universal curve* of $\mathrm{d}E/\mathrm{d}x$ versus $\beta\gamma$ is obtained by using a variant of Equation C.1 that models more accurately the energy loss in the COT, and by substituting appropriate values for the mass of each particle species. Different ranges of $\beta\gamma$ dominated by different particles. The universal curve as measured for various particles at CDF is shown in Figure C-1.

For each reconstructed track, we define the following variable:

$$Z(i) = \ln\left(\frac{\mathrm{d}E/\mathrm{d}x_{\mathrm{meas}}}{\mathrm{d}E/\mathrm{d}x_{\mathrm{exp}}(i)}\right), \quad i = e, \mu, \pi, K, p, \tag{C.2}$$

where $\mathrm{d}E/\mathrm{d}x_{\mathrm{meas}}$ is the measured energy loss per length for the track, and $\mathrm{d}E/\mathrm{d}x_{\mathrm{exp}}(i)$ is the expected energy loss per length from the universal curve for each particle hypothesis i. For a pure sample containing only one particle type, Z is modeled by a single Gaussian distribution. The width σ_Z of the distribution $Z(i)$ is an indication of the resolution in $\mathrm{d}E/\mathrm{d}x$ that is achieved at CDF. We observe a width σ_Z corresponding to a resolution in $\mathrm{d}E/\mathrm{d}x$ of 4–5 %, depending on the particle type. The distribution of $Z(i)$ for muons with $p_T > 1.5$ GeV/c is shown in Figure C-2.

To calculate $\mathrm{d}E/\mathrm{d}x$ resolution for tracks measured in data, an estimate is taken

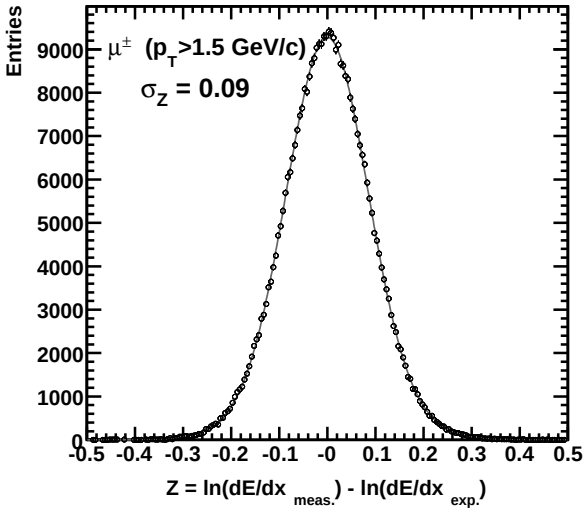


Figure C-2: Distribution of $Z(i)$ for muons with $p_T > 1.5$ GeV. The width of the distribution Z corresponds to a $\mathrm{d}E/\mathrm{d}x$ resolution of ~ 4 %.

from the $\mathrm{d}E/\mathrm{d}x$ calibration samples, taking into account the dependence of σ_Z on track parameters such as the number of COT sense wire contributing to the measurement and the track p_T.

The $Z(i)$ distributions are individually normalized to yield probability density functions $\mathcal{P}_{\mathrm{d}E/\mathrm{d}x}(i)$. These probability density functions are used to calculate the probability of a given track to be caused by each of the particle species i.

Measurement of $\mathrm{d}E/\mathrm{d}x$ in the COT provides a statistical separation of pions and kaons equivalent to ~ 1.0 standard deviation for tracks with momemtum $p > 1.8$ GeV/c, and of pions and electrons equivalent to 3 standard deviations for tracks with momemtum $p = 1.5$ GeV/c. The separation power for several particle species is shown in Figure C-3. The efficiency for associating $\mathrm{d}E/\mathrm{d}x$ information with a charged track reconstructed in the COT in the pseudorapidity range $|\eta| < 1$ and with transverse momentum $p_T > 400$ MeV/c is almost 100%.

C.2 Particle ID with the TOF Detector

The Time-of-Flight system has particle identification capabilities that are complementary to those of the COT. For K/π separation, whereas the latter performs best for tracks with $p > 2$ GeV/c, the TOF yields separation of more than 2 standard deviations for tracks with $p < 1.5$ GeV/c. As its name implies, the TOF measures the arrival time t_{TOF} of particles as they reach the scintillator bars. The arrival time t_{TOF} is combined with the time t_0 of the $p\overline{p}$ interaction and the measurement of both momentum p and track length L from the COT. This information is used to identify

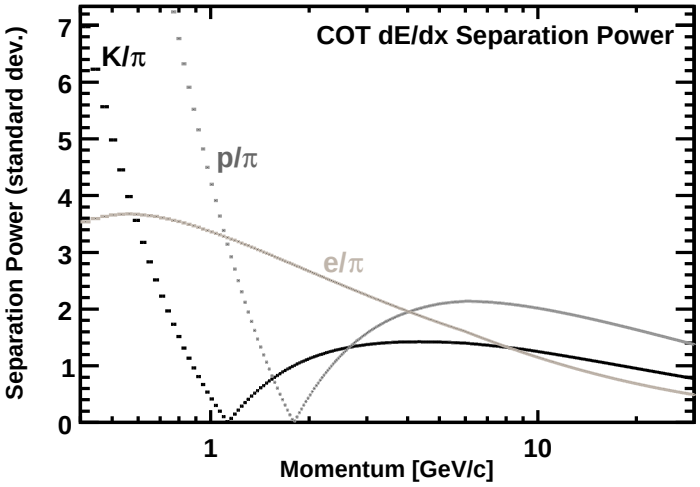


Figure C-3: Separation power of kaons from pions, protons from pions and electron from pions as a function of track momentum, expressed in units of standard deviation, using only the COT for particle identification.

particles by their mass m using

$$m = \frac{p}{c}\sqrt{\frac{c^2 t_{\text{flight}}^2 - 1}{L}}, \tag{C.3}$$

where $t_{\text{flight}} = t_{\text{TOF}} - t_0$.

The measurements of t_{TOF} and of t_0 are the primary contributions to the TOF resolution. Because the $p\overline{p}$ interaction region has a spread of ~ 30 cm, the t_0 measured for any given track has a resolution of several nanoseconds. At CDF we improve the precision on the t_0 measurement by combining TOF measurements for all particles in an event, making the assumption that the relative composition in terms of stable particles is known. This procedure yields a t_0 resolution of 50 ps for $b\bar{b}$ events. The resolution in t_{TOF} is estimated using a sample of $J/\psi \to \mu\mu$ decays and measuring the difference Δ_{TOF} in the time-of-flight of the two muons. The interaction time t_0 cancels out in such a measurement, and kinematic difference between the two tracks in this decay make negligible contributions to Δ_{TOF}, thereby isolating the effect of the TOF instrumental resolution. The distribution of Δ_{TOF} is shown in Figure C-4, with an overlaid fit done using two Gaussians of identical mean and different widths. Depending on the scintillator bar and data-taking period, the narrow Gaussian has widths between 140 and 170 ps, and accounts for 85 % of the area of the distribution. The wider Gaussian primarily accounts for tracks with erroneous TOF information and has a width on the order of several hundred picoseconds. The instrumental TOF time resolution, using the narrow Gaussian to estimate it, is $\sigma_{\text{TOF}} = \sigma_{\Delta_{\text{TOF}}}/\sqrt{2} =$ 110 ps.

The probability distributions for the time-of-flight of different particle species, denoted by $\mathcal{P}_{\text{TOF}}(i)$ are computed using the difference of measured and expected

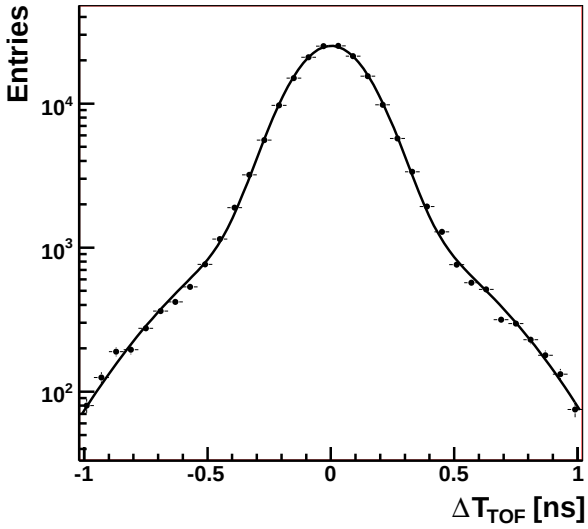


Figure C-4: Distribution of the difference Δ_{TOF} in the time-of-flight of the two muons from $J/\psi \to \mu\mu$ decays, used to determine the resolution of the TOF detector.

time-of-flight, using the same calibration samples as in the case of $\mathrm{d}E/\mathrm{d}x$. Because the TOF measurements are distorted by multiple particles hitting the detector at the same time, and because this is a rather common occurence, the efficiency for associating TOF information with a given track is lower than the analogous efficiency for $\mathrm{d}E/\mathrm{d}x$, and is between 50 and 60 % for $b\bar{b}$ events. This efficiency decreases with increasing instantaneous luminosity because an increase in the latter causes higher track occupancy and a higher likelihood of simultaneous TOF hits.

The separation power of the TOF measurement for several particle hypothesis pairs is shown in Figure C-5. For comparison, the separation power for K/π separation from the $\mathrm{d}E/\mathrm{d}x$ measurement is shown on the same plot.

C.3 Combined Particle ID with $\mathrm{d}E/\mathrm{d}x$ and TOF

In the case of K/π separation, we combine the independently obtained information from the COT and the TOF detectors to enhance our particle identification capability. This combination into a single discriminating variable yield K/π separation of at least 1.5σ for tracks with momentum p up to 5 GeV/c, with higher statistical separation at lower momentum. We define the new *combined log likelihood* (CLL) variable as

$$CLL(K) = \log\left(\frac{\mathcal{P}(K)}{f_p\mathcal{P}(p) + f_\pi\mathcal{P}(\pi)}\right), \quad \mathcal{P}(i) \equiv \mathcal{P}_{\mathrm{TOF}}(i) \cdot \mathcal{P}_{\mathrm{d}E/\mathrm{d}x}(i), \tag{C.4}$$

where $f_p = 0.1$ and $f_\pi = 0.9$ are the proton and pion fractions in the background composition, and $\mathcal{P}_{\mathrm{TOF}}$ and $\mathcal{P}_{\mathrm{d}E/\mathrm{d}x}$ are the probability density functions developed in sections C.1 and C.2.

The combined log likelihood gives an improvement of K/π separation power of

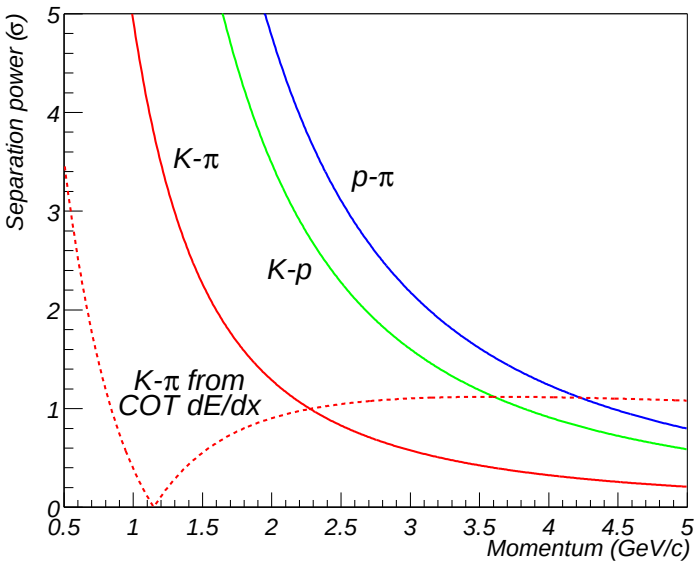


Figure C-5: Separation power of kaons from pions, kaons from protons and protons from pions as a function of track momentum, expressed in units of standard deviation, using only the TOF for particle identification. The separation power of kaons from pions using the COT for particle identification is shown for comparison.

10–20 %, depending on the momentum. Figure C-6 shows the performance of the combination using a sample of $D^{*+} \to D^0\pi^+, D^0 \to K^-\pi^+$.

C.4 Muon ID

We use muon identification to distinguish muons coming from $J/\psi \to \mu^+\mu^-$ from other particles which leave stubs in the muon chambers. We use a likelihood framework developed for the use of the opposite side lepton tagger for the analysis of B_d mixing [117]. This framework was originally used to distinguish between muons coming from semileptonic B decays and backgrounds coming from other sources, which are discussed below.

Muon reconstruction is done by matching COT tracks to stubs in the muon chambers, as discussed in Section 2.3.2. The muon chambers are located on the outside of the CDF detector and separated from the calorimeters by steel shielding. As a result, particles other than muons have a very low probability of reaching the muon detectors. Nevertheless, this leaves two main sources of what we call muon "fakes": punch-through hadrons, and decay-in-flight fakes.

High-energy hadrons at times interact late in the calorimeter volume. A punch-through occurs in this case when one or several outgoing (non-muon) interaction products have a large fraction of the original hadron momentum as well as low transverse momentum with respect to the original hadron track. These interaction products

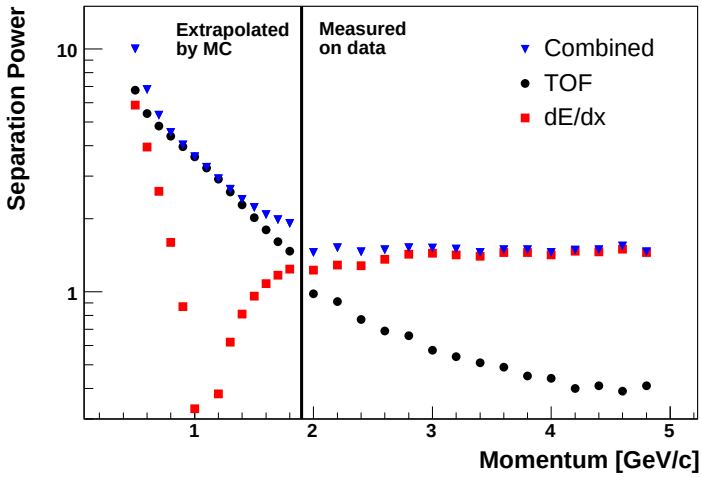


Figure C-6: Separation power of kaons and pions using the combined log likelihood. The performance is evaluated on data for tracks with $p_T > 1.9$ GeV/c and extrapolated using Monte Carlo for tracks with $p_T < 1.9$ GeV/c [116].

leave hits in the muon chambers, and the event reconstruction algorithms match the associated stub(s) to the original hadron track, producing a muon fake.

Decay-in-flight occurs when a pion or a kaon decays to a real muon after leaving the tracking volume, and this muon in turns reaches the muon chambers. Although an actual muon causes the interaction with the chambers in this case, we refer to this phenomenon as a muon fake because the reconstructed track belongs to a hadron.

The use of particle identification using additional detector information is helpful in discriminating between real muons, in our case those from $J/\psi \to \mu^+\mu^-$ decays, and fakes coming from either punch-through or decay-in-flight.

Muon ID is done using a likelihood constructed for signal (real muons) and background (fakes) in five observables: the distance in the r–ϕ plane (ΔX), in ϕ ($\Delta\phi$), and in the longitudinal direction (ΔZ) between the extrapolated COT track and the location of the muon stub, as well as the energy deposition in the electromagnetic and hadronic calorimeters ($E_{\rm EM}$ and $E_{\rm HAD}$). Probability density functions (PDFs) are defined for each variable using empirical functions whose parameters are fitted using calibration samples. To obtain the empirical distributions for real muons, a large sample of $J/\psi \to \mu^+\mu^-$ decays is used. To obtain empirical distributions for the background, hadrons from several decays with associated stubs in the muon chambers are used: pions from $K_S \to \pi^+\pi^-$, kaons from $D^0 \to K^+\pi^-$ and protons from $\Lambda^0 \to p\pi^-$. Figure C-7 shows the distributions of real and fake muons in ΔX and in $E_{\rm HAD}$. In general, real muons have narrower distributions in the the track-to-stub matching observable than fakes. The energy deposition in the calorimeters is centered at the muon's minimum ionizing peak for real muons and is flat for hadronic particles.

The joint PDF S describing the probability that a reconstructed muon is real is

written as a product of PDFs S_i in the variables $\Delta X, \Delta\phi, \Delta Z, E_{\rm EM},$ and $E_{\rm HAD}$:

$$S = \prod_i S_i, \tag{C.5}$$

where S is the joint signal PDF for a real muon, and $i = \Delta X, \Delta\phi, \Delta Z, E_{\rm EM}, E_{\rm HAD}$. The joint PDF B describing the probability that a reconstructed muon is a fake is written as

$$B = \prod_i B_i, \tag{C.6}$$

where $i = \Delta X, \Delta\phi, \Delta Z, E_{\rm EM}, E_{\rm HAD}$. The likelihood estimator is defined as $\mathcal{L}_\mu = S/(S+B)$. With these definitions, real muons have $\mathcal{L}_\mu$ values skewed toward 1 and fakes have $\mathcal{L}_\mu$ values skewed toward 0.

The distributions for real and fake muons in ΔX and $E_{\rm HAD}$ are shown in Figure C-7. The distribution of likelihood values for real and fake muons is shown in Figure C-8.

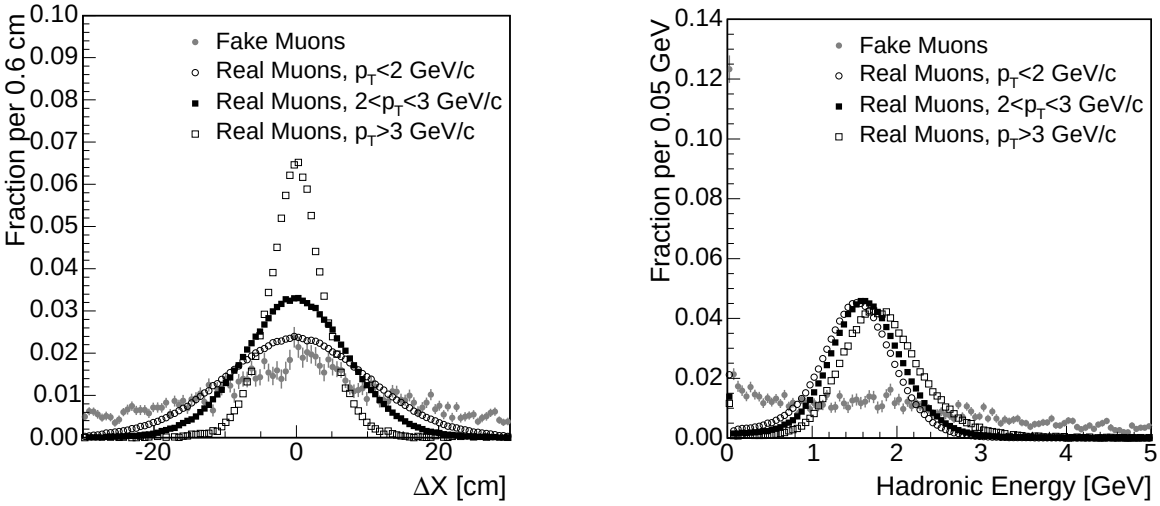


Figure C-7: Distributions for real and fake muons in ΔX and $E_{\rm HAD}$

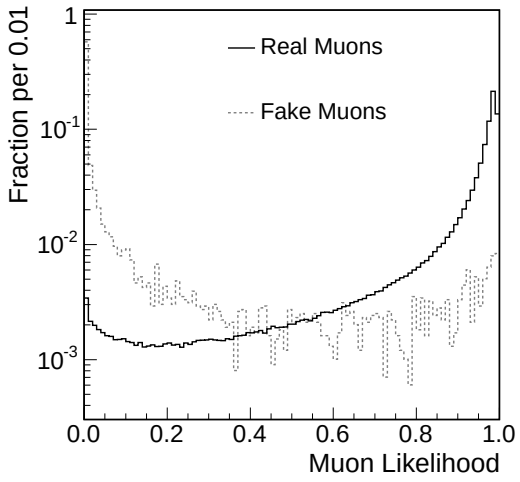


Figure C-8: Muon likelihood values for real and fake muons tested in Monte Carlo

Appendix D

Normalized Angular PDF

In the angular analyses of the $B_s \to J/\psi\phi$ and the $B_d \to J/\psi K^*$ decay modes, the dependence on the transversity angles $\vec{\omega} = \{\cos\theta, \phi, \cos\psi\}$ of the likelihood expression is contained in the f_i functions:

$$\begin{aligned} f_1 &= \frac{9}{32\pi} 2\cos^2\psi(1-\sin^2\theta\cos^2\phi), & f_4 &= -\frac{9}{32\pi}\sin^2\psi\sin 2\theta\sin\phi, \\ f_2 &= \frac{9}{32\pi}\sin^2\psi(1-\sin^2\theta\sin^2\phi), & f_5 &= \frac{9}{32\pi}\frac{1}{\sqrt{2}}\sin 2\psi\sin^2\theta\sin 2\phi, \\ f_3 &= \frac{9}{32\pi}\sin^2\psi\sin^2\theta, & f_6 &= \frac{9}{32\pi}\frac{1}{\sqrt{2}}\sin 2\psi\sin 2\theta\cos\phi, \end{aligned}$$

which have the following convenient property:

$$\int_{-1}^{1}\int_{-\pi}^{\pi}\int_{-1}^{1} f_i(\vec{\omega})\; d\cos\theta\; d\phi\; d\cos\psi = \begin{cases} 1 & , \quad i = 1,2,3 \\ 0 & , \quad i = 4,5,6 \end{cases}$$

Unfortunately, this convenient normalization is spoiled by the introduction of an angular efficiency term $\epsilon(\vec{\omega})$. Indeed, while the integral of ϵ over $\vec{\omega}$ is itself equal to 1, the integral of its product with any of the f_i functions

$$\int_{-1}^{1}\int_{-\pi}^{\pi}\int_{-1}^{1} f_i(\vec{\omega})\cdot\epsilon(\vec{\omega})\; d\cos\theta\; d\phi\; d\cos\psi \tag{D.1}$$

is not equal to 1 ($i = 1,2,3$) or 0 ($i = 4,5,6$). We normalize likelihood expressions with such terms by first calculating the integrals in Equation D.1. Because ϵ is implemented as a normalized 3-D histogram, this amounts to calculating the following sum over the bins of the histogram:

$$N_i = \sum_j \int\int\int f_i(\vec{\omega})\cdot\epsilon_j(\vec{\omega})\; d\cos\theta\; d\phi\; d\cos\psi, \tag{D.2}$$

where the integration limits correspond to the upper and lower limit in each dimension of the bin j.

Bibliography

[1] C. Amsler *et al.*, (Particle Data Group), Review of Particle Physics, *Phys. Lett.*, B667:1, 2008.

[2] M. Kobayashi and T. Maskawa, *CP* Violation in the Renormalizable Theory of Weak Interaction, *Prog. Theor. Phys.*, 49:652–657, 1973.

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

[4] E. Fermi, An Attempt of a Theory of Beta Radiation, *Z. Phys.*, 88:161–177, 1934.

[5] W. Pauli, Liebe Radioaktive Damen und Herren, Letter to the participants of the nuclear physics workshop at Tubingen, 1930.

[6] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Detection of the Free Neutrino: A Confirmation, *Science*, 124:103–104, 1956.

[7] K. Nishijima and T. Nakano, Charge Independence for V Particles, *Prog. Theor. Phys.*, 10:581, 1953.

[8] T. D. Lee and C. N. Yang, Question of parity conservation in weak interactions, *Phys. Rev.*, 104(1):254, 1956.

[9] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Experimental Test Of Parity Conservation In Beta Decay, *Phys. Rev.*, 105:1413–1414, 1957.

[10] J.J. Sakurai, Mass Reversal and Weak Interactions, *Nuovo Cimento*, 7:649, 1958.

[11] E. C. G. Sudarshan and R. E. Marshak, Chirality Invariance and the Universal Fermi Interaction, *Phys. Rev.*, 109:1860–1862, 1958.

[12] R. P. Feynman and M. Gell-Mann, Theory of the Fermi interaction, *Phys. Rev.*, 109:193–198, 1958.

[13] N. Cabibbo, Unitary Symmetry and Leptonic Decays, *Phys. Rev. Lett.*, 10:531–532, 1963.

[14] S. L. Glashow, J. Iliopoulos, and L. Maiani, Weak Interactions with Lepton-Hadron Symmetry, *Phys. Rev.*, D2:1285–1292, 1970.

[15] J. J. Aubert *et al.*, Experimental Observation of a Heavy Particle J, *Phys. Rev. Lett.*, 33:1404–1406, 1974.

[16] J. E. Augustin *et al.*, Discovery of a Narrow Resonance in e^+e^- Annihilation, *Phys. Rev. Lett.*, 33:1406–1408, 1974.

[17] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, Evidence for the 2π Decay of the K_2^0 Meson, *Phys. Rev. Lett.*, 13:138–140, 1964.

[18] L. D. Landau, On the Conservation Laws for Weak Interactions, *Nucl. Phys.*, 3:127–131, 1957.

[19] S. W. Herb *et al.*, Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions, *Phys. Rev. Lett.*, 39:252–255, 1977.

[20] C. Bebek *et al.*, Evidence for New Flavor Production at the $\Upsilon(4S)$, *Phys. Rev. Lett.*, 46:84, 1981.

[21] C. Rubbia, (UA1 Collaboration), Experimental Observation of the Intermediate Vector Bosons W^+, W^-, and Z^0, *Rev. Mod. Phys.*, 57:699–722, 1985.

[22] F. Abe *et al.*, (CDF Collaboration), Observation of Top Quark Production in $\bar{p}p$ Collisions, *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[23] S. Abachi *et al.*, (D$\emptyset$ Collaboration), Observation of the Top Quark, *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[24] A. Alavi-Harati *et al.*, (KTeV Collaboration), Observation of Direct *CP* Violation in $K_{S,L} \to \pi\pi$ Decays, *Phys. Rev. Lett.*, 83:22–27, 1999.

[25] B. Aubert *et al.*, (BaBar Collaboration), Observation of *CP* Violation in the B^0 Meson System, *Phys. Rev. Lett.*, 87:091801, 2001.

[26] K. Abe *et al.*, (Belle Collaboration), Observation of Large *CP* Violation in the Neutral B Meson System, *Phys. Rev. Lett.*, 87:091802, 2001.

[27] C. Amsler *et al.*, (Particle Data Group), Review of Particle Physics: *CP* Violation in Meson Decays, *Phys. Lett.*, B667:153–181, 2008.

[28] L. L. Chau and W. Y. Keung, Comments on the Parametrization of the Kobayashi-Maskawa Matrix, *Phys. Rev. Lett.*, 53:1802, 1984.

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

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

[31] V. Weisskopf and E. Wigner, Over the Natural Line Width in the Radiation of the Harmonious Oscillator, *Z. Phys.*, 65:18–29, 1930.

[32] P. Ball and R. Fleischer, An analysis of B_s Decays in the Left-Right Symmetric Model with Spontaneous *CP* Violation, *Phys. Lett.*, B475:111–119, 2000.

[33] I. Dunietz, R. Fleischer, and U. Nierste, In Pursuit of New Physics with B_s Decays, *Phys. Rev.*, D63:114015, 2001.

[34] K. Anikeev *et al.*, (CDF Collaboration), *B* Physics at the Tevatron: Run II and Beyond, 2001.

[35] A. Lenz and U. Nierste, Theoretical Update of B_s-$\overline{B}_s$ Mixing, *JHEP*, 06:072, 2007.

[36] E. Barberio *et al.*, (Heavy Flavor Averaging Group), Averages of *b*-Hadron Properties at the End of 2005, `arXiv.org:hep-ex/0603003`, 2006.

[37] I. Dunietz, H. Quinn, A. Snyder, W. Toki, and H. J. Lipkin, How to Extract *CP*-Violating Asymmetries from Angular Correlations, *Phys. Rev. D*, 43(7):2193–2208, 1991.

[38] A. S. Dighe, I. Dunietz, H. J. Lipkin, and J. L. Rosner, Angular Distributions and Lifetime Differences in $B_s \to J/\psi\phi$ Decays, *Phys. Lett.*, B369:144–150, 1996.

[39] A. S. Dighe, I. Dunietz, and R. Fleischer, Extracting CKM Phases and B_s-$\overline{B}_s$ Mixing Parameters from Angular Distributions of Non-Leptonic *B* Decays, *Eur. Phys. J.*, C6:647–662, 1999.

[40] M. Beneke, G. Buchalla, and I. Dunietz, Width Difference in the B_s-$\overline{B}_s$ System, *Phys. Rev.*, D54:4419–4431, 1996.

[41] W. M. Yao *et al.*, (Particle Data Group), Review of Particle Physics, *J. Phys.*, G33:1–1232, 2006.

[42] K. Anikeev, *Measurement of the Lifetimes of B Meson Mass Eigenstates*, Ph.D. Thesis, Massachusetts Institute of Technology, 2004.

[43] R. Itoh *et al.*, (Belle Collaboration), Studies of *CP* violation in $B \to J/\psi K^*$ decays, *Phys. Rev. Lett.*, 95:091601, 2005.

[44] B. Aubert *et al.*, (BaBar Collaboration), Measurement of Decay Amplitudes of $B \to J/\psi K^*$, $B \to \psi(2S)K^*$, and $B \to \chi_{c1}K^*$ with an Angular Analysis, *Phys. Rev.*, D76:031102, 2007.

[45] V. M. Abazov *et al.*, (DØ Collaboration), Lifetime Difference and *CP*-Violating Phase in the B_s^0 System, *Phys. Rev. Lett.*, 98:121801, 2007.

[46] V. M. Abazov *et al.*, (DØ Collaboration), Measurement of B_s^0 mixing parameters from the flavor-tagged decay $B_s^0 \to J/\psi\phi$, *Phys. Rev. Lett.*, 101:241801, 2008.

[47] CDF Collaboration, An Updated Measurement of the CP Violating Phase $\beta_s^{J/\psi\phi}$, CDF Public Note 9458, 2008.

[48] B. Worthel *et al.*, *Linac Rookie Book*, 2004, `www-bdnew.fnal.gov/operations/rookie_books/LINAC_v2.pdf`.

[49] C. D. Curtis and C. W. Owen, Operation of the Fermilab 200 MeV Proton Linac, Proceedings of the Fourth All-Union National Conference on Particle Accelerators, Moscow, USSR, 1974.

[50] J. A. MacLachlan, 400 MeV Upgrade for the Fermilab LINAC, Presented at 1989 IEEE Particle Accelerator Conf., Chicago, IL, March 20-23, 1989.

[51] B. Worthel *et al.*, *Booster Rookie Book*, 1998, `www-bdnew.fnal.gov/operations/rookie_books/Booster_V3_1.pdf`.

[52] T. Asher *et al.*, *Main Injector Rookie Book*, 2003, `www-bdnew.fnal.gov/operations/rookie_books/Main_Injector_v1.pdf`.

[53] J. Morgan *et al.*, *Antiproton Source Rookie Book*, 1999, `www-bdnew.fnal.gov/operations/rookie_books/Pbar_V1_1.pdf`.

[54] S. van der Meer, Stochastic Damping of Betatron Oscillations in the ISR, CERN-ISR-PO-72-31, 1972.

[55] S. van der Meer, Stochastic Cooling In The CERN Anti-Proton Accumulator, *IEEE Trans. Nucl. Sci.*, 28:1994–1998, 1981.

[56] J. Marriner, Stochastic Cooling at Fermilab, Presented at 13th International Conference on High Energy Accelerators, Novosibirsk, USSR, Aug 7-11, 1986.

[57] G. Jackson, The Fermilab Recycler Ring Technical Design Report. Rev. 1.2, 1991, FERMILAB-TM-1991.

[58] S. Nagaitsev, Electron Cooling for the Fermilab Recycler, *Nucl. Instrum. Meth.*, A391:142–146, 1997.

[59] B. Worthel *et al.*, *Tevatron Rookie Book*, 2006, `www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v1.pdf`.

[60] LHC Re-Start Scheduled for 2009. CERN Press Release PR10.08. `press.web.cern.ch/press/PressReleases/Releases2008/PR10.08E.html`.

[61] R. Blair *et al.*, (CDF Collaboration), The CDF-II detector: Technical design report, FERMILAB-PUB-96-390-E, 1996.

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

[63] C. S. Hill, (CDF Collaboration), Initial Experience with the CDF Layer 00 Silicon Detector, *Nucl. Instrum. Meth.*, A511:118–120, 2003.

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

[65] A. A. Affolder *et al.*, (CDF Collaboration), Status Report of the Intermediate Silicon Layers Detector at CDFII, *Nucl. Instrum. Meth.*, A485:6–9, 2002.

[66] R. J. Tesarek *et al.*, (for the CDF Collaboration), A Measurement of the Radiation Environment in the CDF Tracking Volume, *Nucl. Instrum. Meth.*, A514:188–193, 2003.

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

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

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

[70] S. Kuhlmann *et al.*, (for the CDF Collaboration), The CDF Calorimeter Upgrade for Run IIB, *Nucl. Instrum. Meth.*, A518:39–41, 2004.

[71] S. R. Hahn *et al.*, (CDF Collaboration), Calibration Systems for the CDF Central Electromagnetic Calorimeter, *Nucl. Instr. Meth.*, A267:351, 1988.

[72] S. Bertolucci *et al.*, (CDF Collaboration), The CDF Central and Endwall Hadron Calorimeter, *Nucl. Instrum. Meth.*, A267:301, 1988.

[73] M. G. Albrow *et al.*, (CDF Collaboration), The CDF Plug Upgrade Electromagnetic Calorimeter: Test Beam Results, *Nucl. Instrum. Meth.*, A480:524–546, 2002.

[74] G. Apollinari, K. Goulianos, P. Melese, and M. Lindgren, (for the CDF Collaboration), Shower Maximum Detector for the CDF Plug Upgrade Calorimeter, *Nucl. Instrum. Meth.*, A412:515–526, 1998.

[75] M. G. Albrow *et al.*, (CDF Collaboration), A Preshower Detector for the CDF Plug Upgrade: Test Beam Results, *Nucl. Instrum. Meth.*, A431:104–111, 1999.

[76] C. M. Ginsburg, (for the CDF Collaboration), CDF Run 2 Muon System, *Eur. Phys. J.*, C33:s1002–s1004, 2004.

[77] G. Ascoli *et al.*, (CDF Collaboration), CDF Central Muon Detector, *Nucl. Instrum. Meth.*, A268:33, 1988.

[78] T. Dorigo, (for the CDF Collaboration), The Muon System Upgrade for the CDF II Experiment, *Nucl. Instrum. Meth.*, A461:560–562, 2001.

[79] A. Meyer, (for the CDF Collaboration), The CDF Data Acquisition System for Tevatron Run II, Presented at CHEP'01: Computing in High-Energy Physics and Nuclear, Beijing, China, September 3-7 2001.

[80] W. F. Badgett, (for the CDF Collaboration), Design, Performance and Control of the CDF Run II Data Acquisition System, Proceedings of 2004 IEEE Nuclear Science Symposium and Medical Imaging Conference, Rome, Italy, October 2004.

[81] N. Leonardo *et al.*, (for the CDF Collaboration), Event Builder and Level 3 Manual for Experts, CDF Internal Note 6138, 2003.

[82] W. F. Badgett, G. Gomez-Ceballos, S. Nahn, J. Patrick, and M. Shochet, (for the CDF Collaboration), Review of the Run2b EVB Upgrade, CDF Internal Note 7422, 2005.

[83] J. Antos *et al.*, (for the CDF Collaboration), Data Production of a Large Linux PC Farm for the CDF Experiment, `arXiv.org:hep-ex/0603008`, 2006.

[84] M.S. Martin, Run II BVtx (Lepton SVT) Algorithms, CDF Internal Note 6147, 2002.

[85] S. Menzemer, *TrackingCal - A Tracking and Alignment Software Package for the CDF II Silicon Detector*, Ph.D. Thesis, Universität Karlsruhe, 2002.

[86] J. R. Miles, L00 in B Lifetimes, Presentation to the BPAK Group, March 5, 2004.

[87] J. Marriner, Secondary Vertex Fit with Mass and Pointing Constraints (CTVMFT), CDF Internal Note 1996, 1993.

[88] D. E. Acosta *et al.*, (CDF Collaboration), Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using Lepton + Jets Events with Secondary Vertex $b-$Tagging, *Phys. Rev.*, D71:052003, 2005.

[89] K. Burkett, J Guimaraes da Costa, and D. Sherman, The Iterative Seed in *PrimeVtx*, CDF Internal Note 7859, 2005.

[90] Ch. Paus *et al.*, BottomMods, `cdfcodebrowser.fnal.gov/CdfCode/source/BottomMods/`.

[91] P. Murat *et al.*, Stntuple, `cdfcodebrowser.fnal.gov/CdfCode/source/Stntuple/`.

[92] T. Sjostrand, L. Lonnblad, and S. Mrenna, PYTHIA 6.2: Physics and manual, 2001.

[93] K. Anikeev, P. Murat, and Ch. Paus, Description of BGenerator II, CDF Internal Note 5092, 1999.

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

[95] CDF Simulation Group, CDF Detector Simulation, 2000, `www-cdf.fnal.gov/cdfsim/cdfsim_main.html`.

[96] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, GEANT3, CERN-DD/EE/84-1, 1987.

[97] K. Anikeev and D. Smirnov, ROOT interface to SNNS, CDF Internal Note 7999, 2005.

[98] A. Zell *et al.*, Stuttgart Neural Network Simulator, Univ. of Stuttgart, 1993.

[99] CDF B_s Mixing Group, Updated Measurement of Δm_d and Calibration of Flavor Tagging in Fully Reconstructed Decays, CDF Public Note 7920, 2005.

[100] F. James and M. Roos, Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations, *Comput. Phys. Commun.*, 10:343–367, 1975.

[101] R. Brun and F. Rademakers, ROOT: An Object Oriented Data Analysis Framework, *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[102] B. Aubert *et al.*, (BaBar Collaboration), Ambiguity-free Measurement of $\cos(2\beta)$: Time-integrated and Time-dependent Angular Analyses of $B \to J/\psi K\pi$, *Phys. Rev.*, D71:032005, 2005.

[103] J. Boudreau *et al.*, (CDF Collaboration), Measurement of Exclusive B Lifetimes in the Modes: $J/\psi K^+$, $J/\psi K^{*+}$, $J/\psi K^{*0}$, and $J/\psi K_s$, CDF Internal Note 6387, 2004.

[104] CDF Collaboration, Measurement of Exclusive B Lifetimes, CDF Public Note 8524, 2006.

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

[106] V. Tiwari *et al.*, Likelihood Based Electron Tagging, CDF Internal Note 7121, 2004.

[107] G. Bauer *et al.*, B Flavor Tagging Using Opposite Side Jet Charge, CDF Internal Note 6951, 2004.

[108] G. Bauer *et al.*, Improved Jet Charge Tagger for Summer Conferences 2004, CDF Internal Note 7131, 2004.

[109] A. Abulencia *et al.*, (CDF Collaboration), Measurement of the $B_s^0 - \bar{B}_s^0$ Oscillation Frequency, *Phys. Rev. Lett.*, 97:062003, 2006.

[110] CDF B_s Mixing Group, Combined Opposite Side Flavor Tagger, CDF Public Note 8460, 2006.

[111] CDF B_s Mixing Group, Same Side Kaon Tagging in Fully Reconstructed Decays, CDF Public Note 8206, 2006.

[112] A. Belloni, *Observation of B_s^0-$\bar{B}_s^0$ Oscillations and the Development and Application of Same-Side-Kaon Flavor Tagging*, Ph.D. Thesis, Massachusetts Institute of Technology, 2007.

[113] N. T. Leonardo, *Analysis of B_s Flavor Oscillations at CDF*, Ph.D. Thesis, Massachusetts Institute of Technology, 2006.

[114] G. J. Feldman and R. D. Cousins, A Unified Approach to the Classical Statistical Analysis of Small Signals, *Phys. Rev.*, D57:3873–3889, 1998.

[115] V. Tiwari, *Measurement of the B_s $\bar{B}_s$ Oscillation Frequency Using Semileptonic Decays*, Ph.D. Thesis, Carnegie Mellon University, 2007.

[116] P. Squillacioti *et al.*, Particle Identification by combining TOF and $\mathrm{d}E/\mathrm{d}x$ information, CDF Internal Note 7488, 2005.

[117] G. A. Giurgiu, *B Flavor Tagging Calibration and Search for B_s Oscillations in Semileptonic Decays with the CDF Detector at Fermilab*, Ph.D. Thesis, Carnegie Mellon University, 2005.