Observation of B_s^0-$\overline{B}_s^0$ Oscillations Using Partially Reconstructed Hadronic B_s Decays

by

Jeffrey Robert Miles

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2008

Author .....
Department of Physics
September 7, 2007

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

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

Observation of B_s^0-$\overline{B}_s^0$ Oscillations Using Partially Reconstructed Hadronic B_s Decays

by

Jeffrey Robert Miles

Submitted to the Department of Physics
on September 7, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis describes the contribution of partially reconstructed hadronic decays in the world's first observation of B_s^0-$\overline{B}_s^0$ oscillations. The analysis is a core member of a suite of closely related studies whose combined time-dependent measurement of the B_s^0-$\overline{B}_s^0$ oscillation frequency Δm_s is of historic significance. Using a data sample of 1 fb^{-1} of $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV collected with the CDF-II detector at the Fermilab Tevatron, we find signals of 3150 partially reconstructed hadronic B_s decays from the combined decay channels $B_s^0 \to D_s^{*-}\pi^+$ and $B_s^0 \to D_s^-\rho^+$ with $D_s^- \to \phi\pi^-$. These events are analyzed in parallel with 2000 fully reconstructed $B_s^0 \to D_s^-\pi^+$ ($D_s^- \to \phi\pi^-$) decays. The treatment of the data is developed in stages of progressive complexity, using high-statistics samples of hadronic B^0 and B^+ decays to study the attributes of partially reconstructed events. The analysis characterizes the data in mass and proper decay time, noting the potential of the partially reconstructed decays for precise measurement of B branching fractions and lifetimes, but consistently focusing on the effectiveness of the model for the oscillation measurement. We efficiently incorporate the measured quantities of each decay into a maximum likelihood fitting framework, from which we extract amplitude scans and a direct measurement of the oscillation frequency. The features of the amplitude scans are consistent with expected behavior, supporting the correctness of the calibrations for proper time uncertainty and flavor tagging dilution. The likelihood allows for the smooth combination of this analysis with results from other data samples, including 3500 fully reconstructed hadronic B_s events and 61 500 partially reconstructed semileptonic B_s events. The individual analyses show compelling evidence for B_s^0-$\overline{B}_s^0$ oscillations, and the combination yields a clear signal. The probability that random fluctuations could produce a comparable signature is 8×10^{-8}, which exceeds the 5 standard deviations threshold of significance for observation. The discovery threshold would not be achieved without inclusion of the partially reconstructed hadronic decays. We measure $\Delta m_s = 17.77 \pm 0.10$ (stat) ± 0.07 (syst) ps^{-1} and extract $|V_{td}/V_{ts}| = 0.2060 \pm 0.0007(\text{exp}) \,^{+0.0081}_{-0.0060}(\text{theory})$, consistent with the Standard Model expectation.

Thesis Supervisor: Christoph M. E. Paus

Title: Associate Professor

Acknowledgments

I am not proud that it's taken me three months since the formal completion of this thesis to write these acknowledgments. So I begin with this admission, not to diminish the importance of this section, but rather to illustrate my anxiety that I would be unable to adequately express my appreciation for the support that I've been privileged to receive from many people. If only because decency requires it, now seems like a good time to try.

Christoph Paus took me into the fold of his many students when I was young and in need of an advisor. With good reason, he is known as a determined leader who works at least as hard as anyone under his supervision. Christoph was accommodating when I needed time to pass MIT's general exams, and when I was finally free, he provided a malleable analysis framework with which I could quickly make meaningful contributions. He knew when to push, and when to give way. The knowledge and perspective I have gained from Christoph – items big and small, positive and negative – are too numerous to list, so I will primarily thank him for the infrastructure on which I have based this dissertation.

Ivan K. Furić was truly a second advisor to me during the final six months culminating in the B_s mixing observation, despite his being "just" a young postdoc. His rare blend of intensity and focus with genuine modesty and a fundamental respect for others is something I can only try to imitate. When I began writing this thesis, it was my intention to announce here that "someone should offer this young man a professorship", but unsurprisingly, several fine institutions beat me to it.

Other postdocs and research scientists were, by way of their example, central figures in my growth as a student. Ilya Kravchenko deployed our original opposite side taggers, helped to direct some of my earlier analysis efforts, and taught me about large-scale Linux computing. He also shared the MIT house at Fermilab with me, never complained when I was noisy or messy, and sometimes even came out to watch Futurama [1]. Guillelmo Gomez-Ceballos was an important resource on all aspects of partially reconstructed B decays – he was the first person to tell me what a κ-factor is – and he was a co-convener of the B Mixing and Lifetimes sub-group along with Sinéad Farrington. Stephanie Menzemer spearheaded development of the PID-based same side tagger, and Aart Heijboer set up the fast fitter and likelihood grids. Konstantin Anikeev made the neural network selection happen, and he was willing to explain it to me more than once.

I. Joseph Kroll was a leading faculty member who provided centralized guidance for our collective effort and who had a hand in my maturity as a student. Kevin Pitts and Matt Herndon served as conveners of the CDF B group during my most relevant contributions, and Marjorie Shapiro was a rigorous and benevolent investigator of my work.

Graduate school can be a trying time, but the experience is made better by sharing it with other students. Alberto Belloni and I followed a semi-parallel trajectory through the CDF experiment. He is an open and generous man who let me pester

[1] Good news everyone!

him about Monte Carlo and who asked only that I listen to his frequent updates about the Juventus soccer team. I thank Vivek Tiwari, Nuno Leonardo, and Jonatan Piedra for their camaraderie and their fundamental contributions to the analysis. Gian Piero Di Giovanni and Bruno Casal were also there at the end to get the work done. Boris Iyutin shared the MIT house and many beers, and he was a good partner for commiseration. Arkadiy Bolshov tolerated my significant presence in our tiny office, and he taught me about chess.

Khaldoun Makhoul mirrored my MIT/CDF experience for an impressive streak of years. He is a man of intellectual breadth, sensitivity, and style. I thank him for many cups of coffee and for being a good friend. Conor Henderson reintroduced me to poker and shared in our small pocket of Red Sox nation; I wish he had joined the group earlier. Markus Klute was another example of how accomplishment and drive do not preclude kindness toward all.

Gerry Bauer was generous with his time and knowledge when I was brand new to the group. I think Kostya said it best about Gerry: "He is one nice guy to talk to." Rob Roser is an outstanding spokesperson who was unfailingly supportive during my early years of service work and through my approach to graduation. JJ Schmidt was another consistently friendly face around the control room. Steve Pavlon at MIT deserves thanks for often asking "do you need anything?" and really meaning it.

Even grad students need activities away from work to make life tolerable. I have especially fond memories of Fermilab volleyball, particularly the "intense Wednesday" group. For good times on the court and in the sand, I thank Sergey Los, Maxim Grigoriev, Igor Terekhov, Dmitri Sidorov, Mikko Voutilainen, Mark Neubauer, and the Ratnikovs, among many others. Volleyball was a great place to blow off steam and to learn to count in Russian. Mike Vincent was my first friend outside work at Fermilab, improving my initial time there with frequent cheap sushi. I also thank the Boston Red Sox organization for providing a reliably entertaining diversion through the years, with special emphasis for the 2004 and 2007 teams.

My close personal relationships are few but very important. Although these acknowledgments are traditionally reserved for last, these people are the foundation of my life and therefore come first. Allyn Dullighan has been with me on this journey since the second day of grad school. She is as tolerant as anyone could be of my crazy schedules, and she was close to me even when I had to be 1000 miles away. My brothers, Kevin, Steven, and RJ always made it worth coming home. I have never found better partners/opponents for tennis, wiffleball, various games on ping pong tables, and Mario Kart.

There is no way that I could effectively condense my appreciation for my parents into only a few sentences. Since my earliest memories, they've told me I could make myself into whatever I wanted to be, and I believed them. It is wonderful to be unconditionally accepted for whomever you are. It would be wrong to dedicate this Ph.D. dissertation to them, because in truth, I did it for myself. (And really, what is a few years' work against the gift of life?) They know that I love them, and that is enough.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Motivation

The goal of high energy physics is to understand the physical universe at its most basic level. The story of "fundamental" science is marked by a chain of models with progressively more basic components, from the hypothetical atoms of antiquity, to chemical elements, to protons and electrons, to quarks and leptons. The modern *high energy* label for this pursuit is based on its requirement for energetic collisions in order to probe the smallest known structures. Moreover, many elementary particles are more massive than those of the familiar world and have not existed in significant concentrations since the universe was very young. Violent collisions of everyday matter provide the energy to manufacture these particles.

Such conditions are generated by particle accelerators. This dissertation is based on data from the Fermilab Tevatron, which has stood for two decades as the world's most powerful collider. The Tevatron directs counter-rotating beams of protons and antiprotons through an underground tunnel four miles in circumference until they are focused into head-on collisions. The CDF detector is centered on one such interaction point, where its multiple specialized layers detect the passage of outgoing remnant particles to provide a multifaceted picture of the underlying collision event.

The CDF physics program is broad, spanning the range of topics from precision tests of the Standard Model – the comprehensive theoretical framework for matter and its interactions – to searches for the signatures of new kinds of physics. A flagship analysis of the current Tevatron era is the collective effort to measure the B_s^0-$\overline{B}_s^0$ oscillation frequency. These mesons, which spontaneously transform back and forth between their matter and antimatter states, provide a direct window on the least known parameters of the weak interaction. The very rapid rate of these B_s oscillations makes their observation a technical challenge, but a successful measurement of the frequency yields a precise test of the completeness of the Standard Model. This dissertation describes the contribution of partially reconstructed hadronic decays to the world's first observation of B_s^0-$\overline{B}_s^0$ oscillations.

1.1 The Standard Model

The Standard Model [1] of particle physics is a quantum field theory which describes all types of matter and the forces through which they interact. The most elementary particles of matter, which are assumed to be point-like fermions, are quarks and leptons. The forces that act on matter include the electromagnetic (EM), weak, and strong interactions, all of which are transmitted by bosons. One deficiency of the Standard Model is that it neglects the force of gravity. This interaction does not yet have a quantum field theory and is so weak as to not affect current laboratory experiments.

Quarks and leptons are grouped in three major *generations*, each containing similar doublets of quarks and leptons but with different scales of mass. Table 1.1 depicts this structure along with important particle properties. Particles of the higher generations are unstable and therefore decay to less massive particles of a lower generation. All matter of the visible universe is composed of particles from the first, least-massive group. The apparently-redundant nature of these generations is not understood, nor is the hierarchy of particle masses, facts that comprise another shortcoming of the model. The Standard Model requires each particle to have an antiparticle with identical mass but opposite charge.

	I	**II**	**III**	Charge	Interactions
Quarks (spin-1/2)	Up (**u**) 1.5 - 3 MeV/c^2	Charm (**c**) 1.25 ± 0.09 GeV/c^2	Top (**t**) 174.3 ± 5.1 GeV/c^2	+2/3	EM, Weak, Strong
	Down (**d**) 3 - 7 MeV/c^2	Strange (**s**) 95 ± 25 MeV/c^2	Bottom (**b**) 4.1 - 4.8 GeV/c^2	−1/3	
Leptons (spin-1/2)	e Neutrino (ν_e) < 2 eV/c^2	μ Neutrino (ν_μ) < 0.19 MeV/c^2	τ Neutrino(ν_τ) < 18.2 MeV/c^2	0	Weak
	Electron (e) 0.511 MeV/c^2	Muon (μ) 105.7 MeV/c^2	Tau (τ) 1.777 GeV/c^2	−1	EM, Weak
Gauge Bosons (spin-1)	Photon (γ) 0 GeV/c^2			0	EM
	W Boson (W) 80.4 GeV/c^2			1	Weak
	Z Boson (Z) 91.2 GeV/c^2			0	Weak
	Gluon (g) 0 GeV/c^2			0	Strong

Table 1.1: Particles of the Standard Model and their important properties, including masses [2], electric charge in units of the magnitude of the electron charge, and the interactions in which the particles participate. The gauge bosons which mediate these forces are common to all three generations.

Table 1.1 also lists the force-carrying bosons. The massless photon carries the EM interaction with infinite range, while the massless gluon of the strong force is highly localized due to the color structure noted below. The $W^\pm$ and Z^0 bosons that mediate the weak force are massive, a fact that is ultimately responsible for the "weakness" of the interaction.

Leptons are grouped into doublets for the purposes of the weak interaction, where

each pair is composed of a neutrino and a charged particle. The charged leptons consist of the electron and its heavier muon and tau variations. In addition to the weak interaction, these particles undergo the EM interaction, as evidenced by the bound states of the stable electron within atoms. The neutrinos carry no charge and therefore do not couple to the photon. Their cross-sections for interacting with matter are vanishingly low, such that they cannot be directly reconstructed by the CDF-II detector.

The six varieties of quarks whose names label Table 1.1 are referred to as the different quark *flavors*. The doublet of each generation contains an "up-type" and a "down-type" quark, as named after the first-generation counterparts. Quarks participate in all forces and are the only particles to interact via the strong force. In addition to the electric charge of the EM interaction, they carry one of three "color" charges for the strong interaction. These colors are metaphorically labelled as red, green, and blue, or their antiparticle conjugates. The potential of the strong interaction does not allow colored particles to exist freely. Free particles must be color-neutral, a condition which is achieved by composite quark particles. These bound quark states are known collectively as *hadrons*, which figure prominently in this analysis. Hadrons exist in one of two groupings: a triplet of quarks $|qqq\rangle$ with one quark of each color, or a quark-antiquark pair $|q\overline{q}\rangle$ of opposite color values. Hadrons of the $|qqq\rangle$ variety are known as baryons, a group which includes common protons and neutrons as well as more massive variants. Quarks of the first generation are the underlying elements of familiar nuclear matter, and the the binding of nuclei is a by-product of the residual forces between strongly-bound quarks. Hadrons of the $|q\overline{q}\rangle$ type are called *mesons*, whose variations containing the b quark are the focus of this dissertation. The process by which bare quarks become confined into composite particles is called "hadronization", which is particularly relevant in the $p\overline{p} \to b\overline{b}$ interactions that provide the data for this analysis. The hadronization process of interest involves the kinetic energy of the separate outgoing b and $\overline{b}$ quarks being transferred via the strong potential into the rest energy of new quark-antiquark pairs. In the case of the b quark, the antiquark member of a new pair binds with it via the strong interaction to form a B meson.

B Mesons

B mesons are bound states of a $\overline{b}$ antiquark and any of the lighter quark flavors, $q = \{u, d, s, c\}$. States with quark content $|\overline{b}q\rangle$ are generally denoted as B_q, while the lightest B species are often denoted by their charges: $B^+ = |\overline{b}u\rangle$ and $B^0 = |\overline{b}d\rangle$. The $B_s^0 = |\overline{b}s\rangle$ meson is the focus of this dissertation, while similar properties of the light B species are leveraged in development of the analysis tools. As discussed below, the neutral B mesons undergo probabilistic oscillations between their particle and antiparticle states due to flavor-changing currents of the weak interaction. Measurement of these oscillation frequencies provides insight into the least-well-known parameters of the weak interaction.

1.2 Weak Interaction and the CKM Matrix

Both the strong and EM interactions are flavor-invariant, proceeding with equal probability for each quark flavor in the absence of phase space (mass) considerations. Additionally, these interactions conserve the quantum numbers of flavor. A quark cannot be created or annihilated without an associated antiquark of the same flavor.

In contrast, weak interactions cause transitions between up-type and down-type quarks. Specifically, processes which are mediated by the charged $W^{\pm}$ boson are termed flavor-changing weak currents because they connect differently flavored quarks. Moreover, the weak interaction connects flavors not only within the same generation but across generations as well.

Before the proposal and discovery of quarks, "strangeness" – the presence of a strange quark – was observed to be a quantum number violated by the weak interaction. To explain differences in decay rates between $\Delta S = 0$ decays (such as $n \to pe\bar{\nu}_e$) and $\Delta S = 1$ decays (such as $K^+ \to \mu^+\nu_\mu$), Cabibbo [3] proposed in 1964 a rotation of angle θ_c between the two currents. In Gell-Mann's subsequent three-quark model [4], the u quark couples to the superposition $d \cdot \cos\theta_c + s \cdot \sin\theta_c$ rather than simply to its doublet partner d. However, several decay rate predictions of this model were still inconsistent with data. The 1970 hypothesis of the charm quark by Glashow, Iliopulous, and Maiani [5] introduced new decay amplitudes that cancelled appropriately with the older model's (the GIM mechanism). The fourth quark expanded the rotation of the weak currents into a two-dimensional "mixing matrix" between the two proposed quark generations:

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

This matrix rotates the $\{d, s\}$ basis of flavor eigenstates to the $\{d', s'\}$ basis of the weak interaction, coupling them to the up-type quarks. Shortly after this proposal, the 1974 discovery of the J/ψ ($|c\bar{c}\rangle$) meson [6,7] demonstrated the existence of charm.

The 1964 observation of violation of the CP symmetry in K^0 decays [8] motivated the proposal of a third generation of quarks by Kobayashi and Maskawa [9] in 1973. Expansion of the unitary rotation matrix of Equation (1.1) to a three-dimensional transformation requires three real rotation angles and one irreducible complex phase. This complex component mandates the existence of CP violation in a natural way. The general form of the Cabibbo-Kobayashi-Maskawa (CKM) matrix is expressed as

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

By convention, the CKM matrix rotates the down-type flavor basis $\{d, s, b\}$ into the weak eigenstate basis $\{d', s', b'\}$. The subsequent discoveries of the bottom quark [10] in 1977 and the top quark [11] in 1993, both at Fermilab experiments, confirmed the existence of the third generation.

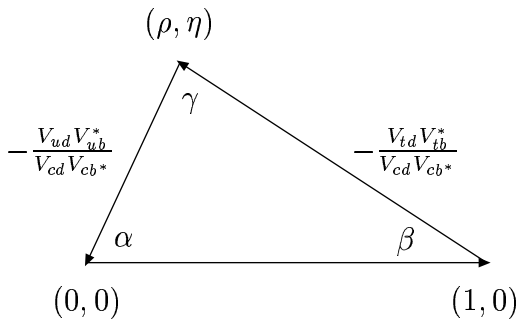

Figure 1-1: The unitarity triangle in the complex plane.

1.2.1 Wolfenstein Parameterization and Unitarity Triangles

A convenient parameterization of the CKM matrix was proposed by Wolfenstein [12]:

$$V = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{1.3}$$

in which the four independent parameters are labelled as

$$\lambda = \sin\theta_c \quad , \quad A \quad , \quad \rho \quad , \quad \eta.$$

This parameterization is accurate to third order in $\lambda = |V_{us}| \approx 0.22$ [2]. The magnitudes of these elements indicate the coupling of the weak interaction between various up-type and down-type flavors. The off-diagonal elements show that cross-generation weak interactions are suppressed by powers of λ. The diagonal elements are close to unity, indicating that interactions are most likely within each generation. In this parameterization, the CP-violating phase is contained in the relative values of the third-generation ρ and η parameters.

Conservation of probability requires that the matrix satisfy the unitarity condition $VV^\dagger = 1$, which yields the six independent equations

$$\sum_j V_{ij}V_{kj}^* = \delta_{ik}. \tag{1.4}$$

The off-diagonal ($i \neq k$) equations may be represented as triangles in the complex plane, for which each of the three $V_{ij}V_{kj}^*$ terms is a vector and the total sum is zero. Real CKM elements produce triangles that lie entirely on the real axis. Since the complex component of the CKM matrix is responsible for the CP violation of the weak interaction, the area of the triangles is proportional to the rate of its CP-violating effects.

The area of a triangle is easiest to measure when all of the angles are of approximately the same size. The triangle defined by the unitarity condition

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \tag{1.5}$$

may be written to leading order in λ as

$$A\lambda^3(\rho + i\eta) - A\lambda^3 + A\lambda^3(1 - \rho - i\eta) = 0. \tag{1.6}$$

This triangle is roughly equilateral, with each side proportional to λ^3. The Equation (1.5) condition is often simply called the unitarity triangle (UT) and is illustrated in Figure 1-1. In this common depiction, Equation (1.5) is divided by $V_{cd}V_{cb}^*$ such that the length of that side is unity and may be aligned with the real axis. This arrangement emphasizes the least-well-known parameters of the CKM matrix. Based on the Wolfenstein parameterization, the apex of the UT has coordinates $(\bar{\rho}, \bar{\eta}) = (1 - \lambda^2/2)(\rho, \eta)$. The three angles are given by

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{1.7}$$

The upper-right side of the UT in Figure 1-1 is of particular interest for this analysis; its length is

$$R_t \equiv \left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right| = \sqrt{(1-\bar{\rho})^2 + \bar{\eta}^2} = \frac{1}{\lambda}\left|\frac{V_{td}}{V_{ts}}\right|. \tag{1.8}$$

Section 1.2.2 discusses how the ratio $|V_{td}/V_{ts}|$ is related to the frequencies of neutral B oscillations.

As may be inferred from the Equation (1.4) unitarity condition, the number of sides of the UT represents the number of quark generations. The existence of four or more generations would cause the triangle to expand to a polygon of appropriate sides in the complex plane. If experimental data were to produce conflicting measurements of the $(\bar{\rho}, \bar{\eta})$ apex, the non-closed nature of the unitarity triangle would be direct evidence for some new kind of physics. Therefore, it is a priority of experimental physics to over-constrain the UT by measuring all of its sides and angles in order to determine whether it is closed.

1.2.2 Constraints of the CKM Matrix

The parameters of the CKM matrix are not predicted by the Standard Model and must be extracted from experimental data. However, there are a limited number of measurements which can be cleanly interpreted in terms of constraints on the UT [13]. The predictions for many measurements (such as branching fractions or lifetimes) are burdened with significant theoretical uncertainties. A common method for reducing such uncertainties for interpretation of measurements is to take a *ratio* of experimental results, as noted below.

Figure 1-2 shows a projection of the constraints on the CKM matrix parameters from the beginning of the year 2006 [14]. The UT apex position $(\bar{\rho}, \bar{\eta})$ is fitted with five constraints that are extracted cleanly from the following measurements. These constraints are labelled in the figure as $|Vub/V_{cb}|$, ϵ_K, $\sin(2\beta)$, Δm_d, and $\Delta m_d/\Delta m_s$, respectively.

1. The ratio of rates of $b \to u\ell\nu$ and $b \to c\ell\nu$ decays determines the $|V_{ub}/V_{cb}|$

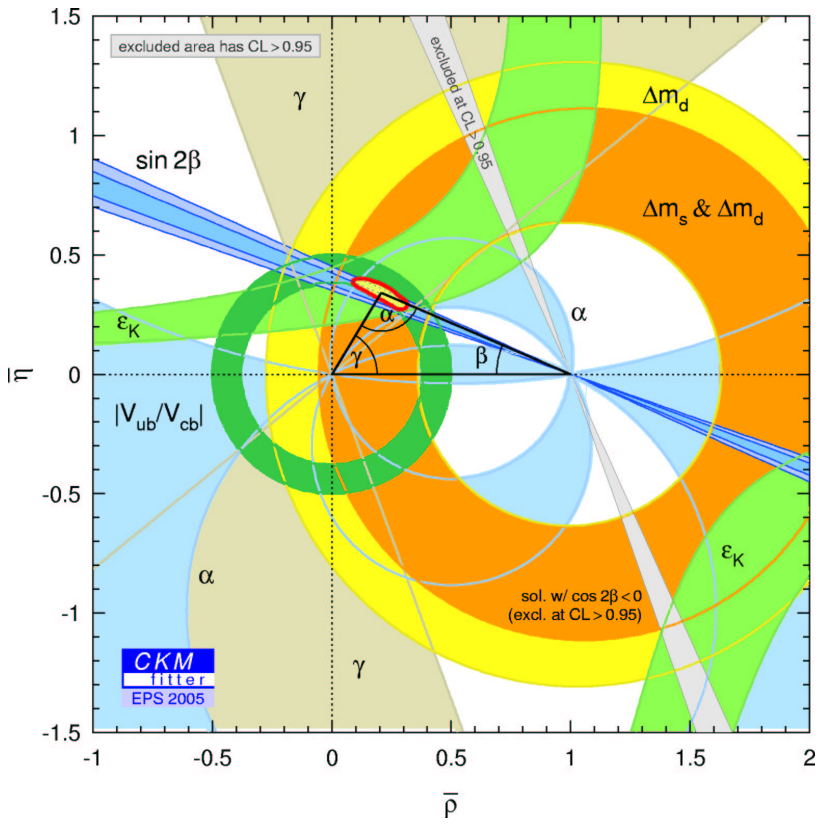

Figure 1-2: Fitted regions of 68% and 95% probability for the apex position $(\bar{\rho}, \bar{\eta})$ of the unitary triangle, using constraints available at the beginning of 2006 [14].

constraint via

$$\sqrt{\bar{\rho}^2 + \bar{\eta}^2} = \left(1 - \frac{\lambda^2}{2}\right) \frac{1}{\lambda} \left|\frac{V_{ub}}{V_{cb}}\right|. \tag{1.9}$$

Measurements of these types of B branching fractions constrain the length of the UT side between $(0,0)$ and $(\bar{\rho}, \bar{\eta})$. The best results for these measurements are obtained by the "B Factory" experiments, BABAR [15] and Belle [16].

2. The ratio

$$\epsilon_K = \frac{A(K_L \to (\pi\pi))_{I=0}}{A(K_S \to (\pi\pi))_{I=0}} \tag{1.10}$$

is a measure of the asymmetry of CP violation in the neutral kaon system, where I denotes isospin. Using input from QCD calculations, measurement of ϵ_K results in a roughly hyperbolic constraint on the UT [13].

3. Measurement of the time-dependent asymmetry of CP violation in $B^0(\overline{B}^0) \to J/\psi K_s^0$ decays results in a direct measurement of $\sin(2\beta)$:

$$\frac{N_{B^0 \to J/\psi K_s^0} - N_{\overline{B}^0 \to J/\psi K_s^0}}{N_{B^0 \to J/\psi K_s^0} + N_{\overline{B}^0 \to J/\psi K_s^0}}(t) = \sin(2\beta)\sin(\Delta m_d t), \tag{1.11}$$

where β is the UT angle shown in Figure 1-2 and Δm_d is the oscillation frequency of the B^0 system discussed below. These measurements are also dominated by BABAR and Belle.

4. The length of the UT side between $(1,0)$ and $(\bar{\rho},\bar{\eta})$ is constrained by the neutral B oscillation frequencies Δm_d and Δm_s. Equation (1.8) notes that this length R_t is defined by the simple ratio of matrix elements $|V_{td}/V_{ts}|$. Figure 1-2 clearly shows that the UT apex position is most poorly constrained along this side: these elements are the least well measured of the CKM matrix. The B Factory and Tevatron experiments are complementary in this and many other arenas of B physics, in that copious B^0 and B^+ mesons are collected by BABAR and Belle, while only the Tevatron detectors currently have regular access to B_s mesons [1] and heavier b hadrons. The B^0 frequency Δm_d is well-measured at the B Factories [2,18], but Δm_s analyses have historically resulted only in lower limits. Section 1.2.2 describes how measurement of the B_s oscillation frequency allows a direct extraction of the $|V_{td}/V_{ts}|$ ratio and a much tighter constraint on the UT.

1.3 B Flavor Oscillations in the Standard Model

The oscillation of neutral B mesons between their particle and antiparticle flavor states is an expected phenomenon of the Standard Model. The section discusses the origin and phenomenology of these oscillations and the relationship of their frequencies to CKM matrix parameters.

1.3.1 Phenomenology of B Mixing

The transformation of a B meson to its own antiparticle conjugate proceeds through the "box" diagrams shown in Figure 1-3. The process involves internal exchange of two W bosons and two up-type quarks, where the oscillation rates of the B_s^0 and B^0 systems are reflective of their s or d quark content, respectively. Section 1.3.2 below discusses the connection of physical observables to these internal couplings. However, the phenomenon of oscillations may be readily described independently of the underlying mechanism.

Due to the applicability of this discussion to both B_s^0 and B^0 mesons, the general notation $|B\rangle$ and $|\overline{B}\rangle$ is used for the neutral mesons. Additionally, the term *flavor* is used here and throughout this analysis to mean the charge conjugate states of *bottom* flavor, as opposed to the more general quark flavors of Table 1.1.

B_q mesons are created via strong or electromagnetic interactions as pure eigenstates of flavor $|B_q\rangle = |\overline{b}q\rangle$. Based on the process of Figure 1-3, and because the weak interaction eigenstates are different than those of flavor, the initially pure $|B\rangle$ or $|\overline{B}\rangle$ state evolves into a mixture of the flavor states. For the state vector $|B(t)\rangle$ at time t

[1] The Belle experiment has collected B_s mesons in test runs of the KEKB e^+e^- accelerator operating at the $\Upsilon(5S)$ resonance [17], but the detector lacks the proper time resolution to observe B_s mixing.

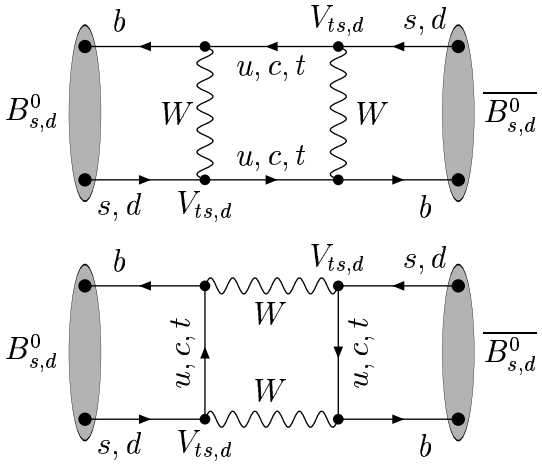

Figure 1-3: Lowest-order Feynman diagrams contributing to B^0_s and B^0 mixing.

for initial state $|B\rangle$, the time evolution is governed by the Schrödinger equation

$$i\frac{d}{dt}\begin{pmatrix} |B(t)\rangle \\ |\overline{B}(t)\rangle \end{pmatrix} = \left(\mathbf{M} - i\frac{\mathbf{\Gamma}}{2}\right)\begin{pmatrix} |B(t)\rangle \\ |\overline{B}(t)\rangle \end{pmatrix}, \tag{1.12}$$

using the Hamiltonian of free propagation [19]

$$\mathcal{H} = \left(\mathbf{M} - i\frac{\mathbf{\Gamma}}{2}\right) = \begin{pmatrix} m & M_{12} \\ M^*_{12} & m \end{pmatrix} + \frac{i}{2}\begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma^*_{12} & \Gamma \end{pmatrix}. \tag{1.13}$$

The diagonal elements of the Hamiltonian describe the mass m and decay width Γ of the flavor eigenstates, which are guaranteed by CPT invariance to be the same for both $|B\rangle$ and $|\bar{B}\rangle$. The off-diagonal terms of the Hamiltonian can induce transitions between the charge-conjugate states. The mass eigenstates are defined as the eigenvectors of Equation (1.13) and may be expressed in terms of the flavor eigenstates as

$$|B_{L,H} = p|B\rangle \pm q|\overline{B}\rangle \tag{1.14}$$

with $|p|^2 + |q|^2 = 1$. The labels L and H refer to the "light" and "heavy" mass eigenstates. The associated eigenvalues are given by

$$\lambda_{L,H} = (m - \frac{i}{2}\Gamma) \pm \frac{q}{p}(M_{12} - \frac{i}{2}\Gamma_{12}) \tag{1.15}$$

in combination with

$$\frac{q}{p} = \sqrt{\frac{M^*_{12} - \frac{i}{2}\Gamma^*_{12}}{M_{12} - \frac{i}{2}\Gamma_{12}}}. \tag{1.16}$$

The time evolution of the flavor states may be written as

$$|B(t)\rangle = g_+(t)|B(0)\rangle + \frac{q}{p}g_-(t)|\overline{B}(0)\rangle\,, \tag{1.17}$$

$$|\overline{B}(t)\rangle = \frac{p}{q}g_-(t)|B(0)\rangle + g_+(t)|\overline{B}(0)\rangle\,, \tag{1.18}$$

using

$$g_\pm(t) = \frac{1}{2}\left[e^{-\left(im_L+\frac{1}{2}\Gamma_L\right)t} \pm e^{-\left(im_H+\frac{1}{2}\Gamma_H\right)t}\right]\,. \tag{1.19}$$

Of specific interest for this analysis are the time-dependent probabilities for mesons in initial flavor eigenstates to decay at later time t with the same or opposite flavor. In the limit that $|q/p| = 1$ and $(\Gamma_L - \Gamma_H)/\Gamma$ is negligibly small, both of which are good approximations for the B^0 and B_s^0 systems [20], the probability densities are given by

$$\mathcal{P}_{B\to B}(t) = \mathcal{P}_{\overline{B}\to\overline{B}}(t) = \frac{\Gamma}{2}e^{-\Gamma t}\left[1+\cos\left(\Delta m t\right)\right]\,, \tag{1.20}$$

$$\mathcal{P}_{B\to\overline{B}}(t) = \mathcal{P}_{\overline{B}\to B}(t) = \frac{\Gamma}{2}e^{-\Gamma t}\left[1-\cos\left(\Delta m t\right)\right]\,, \tag{1.21}$$

where $\Gamma = 1/\tau$ is the inverse of the B lifetime and Δm is the difference between the mass eigenstates

$$\Delta m \equiv m_H - m_L. \tag{1.22}$$

Equations (1.20) and (1.21) describe how a freely propagating neutral B meson *oscillates* between being more probable to decay in the same or opposite flavor state as that in which it was created. The rate of this "mixing" between flavor states is governed by the mass difference Δm. This analysis uses the convention $\hbar = c = 1$, such that Δm is conveniently described as the oscillation frequency in units of inverse time, typically ps^{-1}. The probabilistic oscillation may be visualized by constructing the asymmetry of the flavor transitions as a function of time

$$A(t) = \frac{N_{\text{unmixed}} - N_{\text{mixed}}}{N_{\text{unmixed}} + N_{\text{mixed}}}(t) = \cos(\Delta m t), \tag{1.23}$$

where N_{unmixed} and N_{mixed} represent the number of B_s mesons that are observed to decay according to the probabilities of Equations (1.20) and (1.21), respectively. This idealized cosinusoidal asymmetry is illustrated in Figure 1-4.

The probability densities of Equations (1.20) and (1.21) are central to this analysis. With additional features to account for realistic detector considerations, they provide the core structure of proper time modeling for neutral B decays.

1.3.2 Relationship to CKM Matrix

The leading-order diagrams that contribute to B flavor oscillations are depicted in Figure 1-3. This second-order weak interaction involves the virtual exchange of two W bosons and two up-type quarks, where the contribution of the t quark dominates

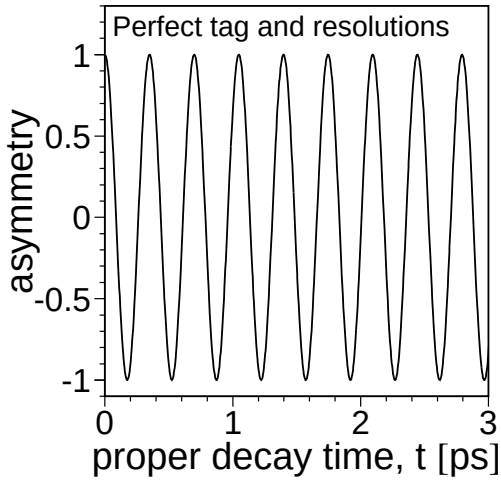

Figure 1-4: Time-dependent flavor asymmetry with perfect detector resolutions and b-flavor identification.

that of the lighter quarks by $(m_t/m_{u,c})^2$ [21]. Therefore, based on the masses of Table 1.1, the effective couplings of the transformation are proportional to $(V_{tb}^* V_{td}^*)^2$ and $(V_{tb}^* V_{ts}^*)^2$ for the B^0 and B_s^0 systems, respectively. The oscillation rate Δm may be expressed as

$$\Delta m_q \propto m_{B_q} f_{B_q}^2 \hat{B}_{B_q} |V_{tb}^* V_{tq}^*|^2, \quad \text{with} \quad q = \{d, s\}, \tag{1.24}$$

where the exact relation is completed by quantities common to both species [21]. The form factors f_{B_q} and bag factors $\hat{B}_{B_q}$ are calculated via application of lattice QCD, but the combined uncertainty of these factor is approximately 10%, which limits the utility of Equation (1.24) for individual species. Nevertheless, these theoretical uncertainties are significantly reduced in the ratio of frequencies

$$\frac{\Delta m_s}{\Delta m_d} = \xi^2 \frac{m_{B_s^0}}{m_{B^0}} \frac{|V_{ts}|^2}{|V_{td}|^2}, \tag{1.25}$$

with

$$\xi = \frac{f_{B_s^0}}{f_{B^0}} \sqrt{\frac{\hat{B}_{B_s^0}}{\hat{B}_{B^0}}} = 1.210 \, {}^{+0.047}_{-0.035} \text{ [22]}, \tag{1.26}$$

and the ratio of masses is precisely measured by CDF [23]. Thus, measurement of the ratio of oscillation frequencies for B_s^0 and B^0 yields a direct measurement of $|V_{td}/V_{ts}|$ with relatively small theoretical uncertainty. Equation (1.8) shows that this ratio of CKM matrix parameters defines the length of the least-constrained side of the unitarity triangle in Figure 1-2. The Δm_s constraint in this plot uses the late-2005 world-average lower limit of $\Delta m_s > 14.4$ ps^{-1} at 95% CL. The B^0 mixing frequency $\Delta m_d = 0.507 \pm 0.005$ ps^{-1} [2, 18] is precisely measured by BABAR and

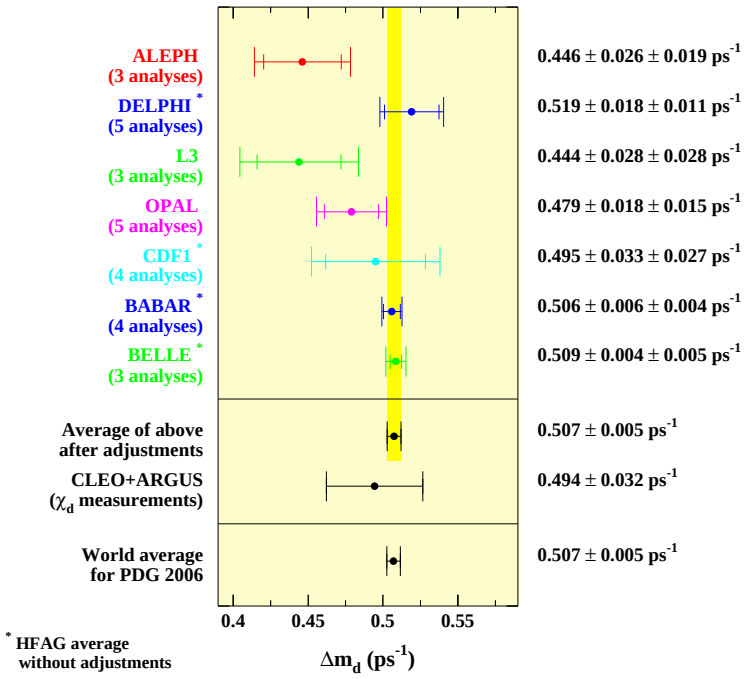

Figure 1-5: Summary of measurements of Δm_d.

Belle, but a definitive observation and measurement of Δm_s has not been achieved prior to this analysis. Under the assumption that the Standard Model is a complete and fundamental theory, the constraints of Figure 1-2 predict a value of $\Delta m_s = 18.3\ ^{+6.5}_{-1.5}$ ps^{-1}. The asymmetric uncertainty of this prediction arises from the lower limits on Δm_s placed by previous analyses. However, such a high frequency has made B_s oscillations too rapid to be directly resolved. A definitive and precise measurement of Δm_s would produce a $|V_{td}/V_{ts}|$ measurement with significantly smaller uncertainty than previously available, thus providing a stringent test of the unitarity of the CKM matrix.

1.4 Status of B Mixing Measurements

Evidence of neutral B oscillations was first reported in 1987 by the UA1 Collaboration using a time-integrated study of like-signed muon pairs from $p\overline{p}$ collisions [24]. Shortly thereafter, the ARGUS Collaboration reported the first observation of time-integrated B^0 mixing using like-signed leptons from $\Upsilon(4S) \to B^0\overline{B}^0$ [25]. This observation was confirmed in 1989 by the CLEO experiment [26]. Time-dependent measurement of B^0_d-$\overline{B}^0_d$ oscillations have since yielded precise measurements of Δm_d [2, 18] which are summarized in Figure 1-5.

In combination with the time-integrated results from the B^0 system, the UA1

results indicated that B_s^0-$\overline{B}_s^0$ oscillations also occur, although the mixing frequency was too high to resolve. Searches for B_s^0 mixing in the regime of low sensitivity to Δm_s are greatly assisted by aspects of Fourier analysis, as discussed below.

1.4.1 Fourier Analysis and Amplitude Scans

The PDF for the proper decay time of a neutral B meson is of the form

$$\mathcal{P}(t) \sim \frac{1}{2\tau} e^{-t/\tau}[1 \pm \cos(\Delta m t)]. \tag{1.27}$$

The most direct analysis approach would involve a fit of the B_s mixing frequency, searching the space of Δm_s for best agreement with the data. However, the fit routine will fail or produce inaccurate confidence level estimates if the hypothesized signal has inadequate significance. Historical mixing searches have been made possible by incorporating elements of Fourier analysis, for which the PDF is expanded by one parameter to read

$$\mathcal{P}(t) \sim \frac{1}{2\tau} e^{-t/\tau}[1 \pm \mathcal{A}\cos(\Delta m t)], \tag{1.28}$$

where the *amplitude* $\mathcal{A}$ serves to scale the oscillation term. In this method [27], a range of Δm_s values are assumed, one by one, and the amplitude parameter is fitted in the data for each. In the case that the probed Δm_s value is the true oscillation frequency, the fitted amplitude should be consistent with unity. For frequencies away from the true Δm_s, the fitted amplitude should be consistent with zero, suppressing the oscillation term. Figure 1-6 shows the Equation (1.27) PDF and the associated amplitude in frequency space. The shape of the amplitude peak is related to a Breit-Wigner distribution, depending in part on the B decay width and the oscillation frequency [27]. The amplitude method effectively transforms the fit of periodic oscillations in the time space into a *resonance* search in the frequency space of Δm_s. The features and interpretation of amplitude scans are discussed in greater detail in Chapter 7.

The fitted amplitude uncertainty $\sigma_{\mathcal{A}}$ is Gaussian-distributed. Therefore, a given Δm_s value can be excluded at 95% CL if its associated amplitude meets the condition

$$\mathcal{A}(\Delta m_s) + 1.645 \cdot \sigma_{\mathcal{A}}(\Delta m_s) \leq 1, \tag{1.29}$$

and the largest frequency for which this condition is true defines the lower limit on Δm_s. Statistical fluctuations of $\mathcal{A}$ can create more or less stringent exclusion regions of Δm_s. The "sensitivity" of a data sample is therefore defined to represent the largest value of Δm_s that would be excluded for the ideal non-oscillating $\mathcal{A}(\Delta m_s) = 0$:

$$1.645 \cdot \sigma_{\mathcal{A}}(\Delta m_s) = 1. \tag{1.30}$$

This quantity is also useful in assessing the expected power of a dataset.

In addition to allowing searches in which a direct observation is not possible, the resonance behavior of the amplitude method is advantageous because it provides

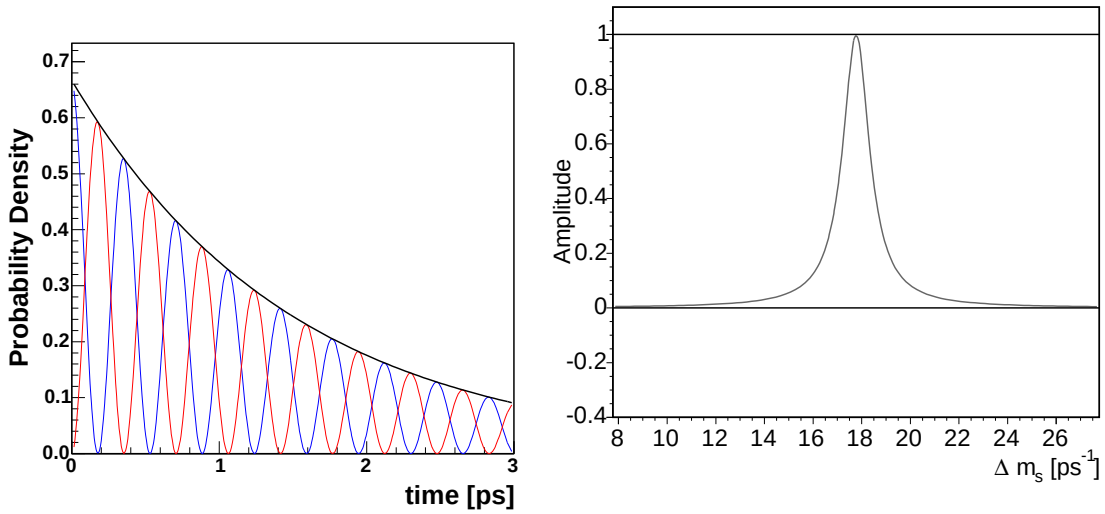

Figure 1-6: Left: the time PDFs of unmixed (blue) and mixed (red) B_s decays with Δm_s set to 17.75 ps^{-1}. Right: the associated amplitude in frequency space.

for simple combination of independent measurements. Multiple amplitude scans need only be averaged together, with appropriate weight from their uncertainties. Figure 1-7 compiles amplitude values from LEP, SLD [28], and CDF at the probed value $\Delta m_s = 17.5$ ps^{-1}. The corresponding world average amplitude scan [2, 29] is shown in Figure 1-8, which yields the limit $\Delta m_s > 14.4$ ps^{-1} used in the unitarity triangle constraint of Figure 1-2.

Very recently, the DØ Collaboration reported 17 ps^{-1} $< \Delta m_s <$ 21 ps^{-1} at 90% C.L. using a sample of semileptonic B_s decays [30]. However, this bound is based on an amplitude signature with a 5% probability of being a random fluctuation, equivalent to a Gaussian significance of about 2.2 standard deviations. Immediately thereafter, CDF produced the first precise measurement of $\Delta m_s = 17.31\ ^{+0.33}_{-0.18}$ (stat) $\pm$ 0.07 (syst) ps^{-1} from a combined analysis of fully hadronic and semileptonic B_s decays [31]. Nevertheless, that evidence for mixing had a 0.2% probability to be a random fluctuation, equivalent to 3.1 standard deviations, and remained inadequate to claim a discovery. This dissertation describes the upgrade to that CDF analysis, in an attempt to make a definitive observation of B_s oscillations.

1.5 Roadmap for Observation of B_s Oscillations

This section provides a synopsis of the principles which underlie a generic analysis of B_s mixing, concluding with an introduction to the original contribution on which this dissertation is based. In general, the outline neglects rigorous treatment of the concepts in favor of a qualitative overview. Specific chapters containing the detailed development are noted throughout.

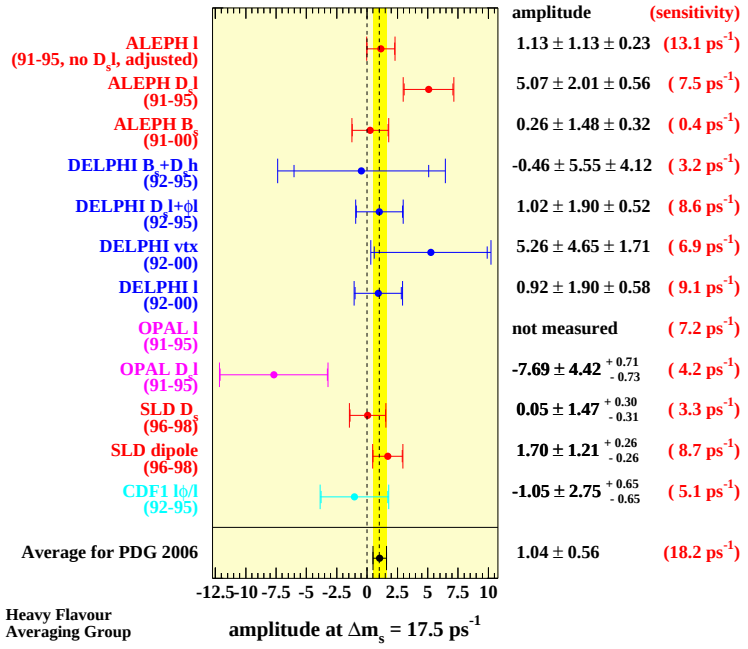

Figure 1-7: Amplitude measurements and sensitivities for various experiments at $\Delta m_s = 17.5$ ps^{-1} as of early 2006.

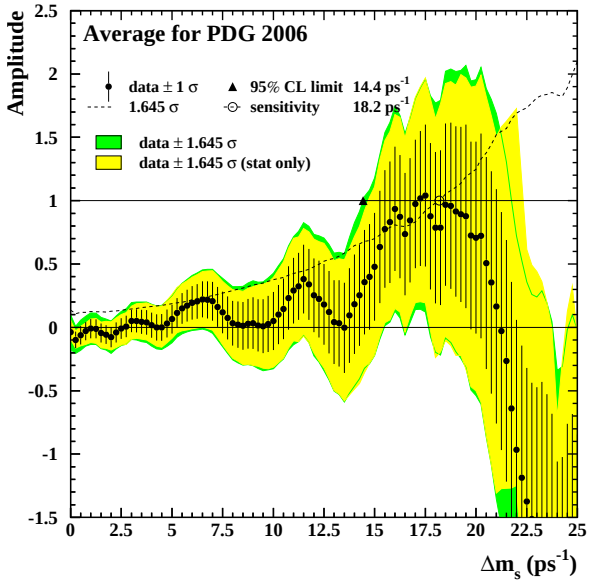

Figure 1-8: The world-average amplitude scan through early 2006, demonstrating low-significance behavior.

1.5.1 Essential Ingredients

The first requirement for a successful measurement is the data sample. This analysis is based on data collected at the Fermilab Tevatron, whose $p\overline{p}$ collisions produce B mesons among many other particle states. These B events are identified, recorded, and accumulated in different samples according to basic characteristics. This process is discussed in depth in Chapter 3.

Given a sample of B_s mesons, the set of measurements necessary for observation of flavor mixing is conceptually simple. Regardless of decay channel, there are only three properties which must be known for each event:

1. the *proper time* elapsed in the rest frame of the B_s between its production and decay,
2. the b-flavor of the B_s at the time of its *decay*: whether the particle decayed as a B_s^0 or as a $\overline{B}_s^0$,
3. the b-flavor of the B_s at the time of its *production*.

All of the quantities measured in the data contribute either to knowledge of these properties or to discrimination and modeling of signal and background.

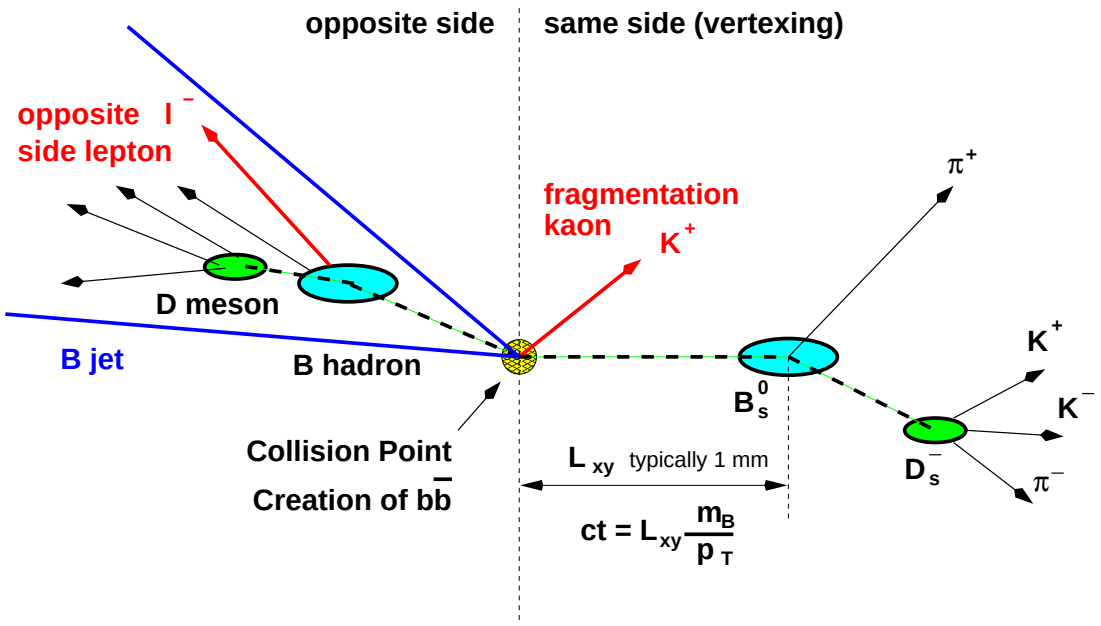

Figure 1-9: Simple sketch in the transverse plane of the hadronization and decay of a $b\overline{b}$ pair, producing a B_s^0 signal event with structures for same- and opposite-side flavor tags.

Figure 1-9 depicts an idealized sketch of an event in the transverse plane. In this simplified view, the $p\overline{p}$ collision produces a $b\overline{b}$ pair from which each quark hadronizes in opposite hemispheres. This example shows the $\overline{b}$ antiquark bound to a strange

quark, with the resulting B_s^0 meson traveling a measurable distance in the detector (typically about 1 mm) before decaying. As noted in Section 3.1, this displacement is an essential characteristic for identification of B signal samples. The charged particles resulting from this decay and its unstable daughters are reconstructed by the detector, and measurements of the B_s decay length $L_{\rm xy}$ and momentum p_T are used to infer the particle's proper time ct via the relation

$$ct = \frac{L_{\rm xy} \cdot M}{p_T}. \tag{1.31}$$

The reconstruction procedure and event variables used by the CDF detector are discussed in Section 2.2, while Chapter 5 describes the critical considerations for proper time. The signal channels used in the analysis are chosen such that the b-flavor of a signal meson at the time of its decay is given directly from the charge of the resulting tracks, allowing for unambiguous knowledge of whether it was a B_s^0 or a $\overline{B}_s^0$. However, while extraction of decay flavor is relatively easy, the measurement of production flavor is a technical challenge. The algorithms for inferring the initial flavor state are collectively referred to as *flavor tagging*, and they rely on reconstruction of other event features. Along with the signal decay, Figure 1-9 shows the fragmentation kaon, which is likely to result from signal hadronization, as well as a general depiction of the hadronization and decay of the opposite b quark. The production flavor tagging methods associated with these features are known as Same Side Tagging and Opposite Side Tagging, respectively, and they are described in Chapter 6.

Numerous other quantities are reconstructed for each event, including masses, momenta of all objects, estimates of the quality of a given flavor tag, and the estimated resolutions with which many of these properties are measured. These are a subset of many variables which, although not relevant in the idealized scenario presented thus far, are used for discrimination and modeling of signal and background components in the realistic analysis.

In addition to B_s decays, samples of the less massive B^+ and B^0 mesons are also collected. These datasets, for which much larger data samples are available than for the B_s, are critical as a testbed for the analysis framework and for calibration of all analysis components.

Despite the conceptual simplicity noted at the beginning of this section, the push to directly measure B_s oscillations represents an intensive multi-year effort. The large value of the mixing frequency, in combination with a small number of signal events, challenges the capabilities of the detector and established analysis tools. Many separate studies are devoted to thorough characterization of the data content, validation of the detector's measurement of proper decay time, and calibration of the b-flavor tagging algorithms.

The statistical significance of an oscillation signal depends on a combination of the precision of knowledge of the above concepts. While introduction of the expression's quantitative form is withheld until Section 1.5.3, the dependence is comprised of specific terms for resolution in proper decay time, the efficiency and accuracy of flavor tagging, and the size and cleanness of the B_s sample. Optimization of these

characteristics and a solid understanding of their values are central considerations for the development of this analysis.

1.5.2 Development of the Likelihood Model

With a pure sample of B_s events and perfect information on the short list of proper time and flavor quantities in Section 1.5.1, it would be a straightforward procedure to demonstrate the presence of an oscillation signal and to measure its frequency. As visualized in Section 1.2.2, the flavor attributes at production and decay could be used to construct the mixing asymmetry, and its time-dependence would yield a cosinusoidal behavior whose frequency is simply Δm_s. However, such an ideal sample does not exist, and the simplistic asymmetry picture is clouded by realistic considerations such as the presence of backgrounds and an imperfect ability to reconstruct proper time and initial flavor. These realistic effects combine to make the direct asymmetry method impossible without sample sizes far beyond what is currently available. Indeed, the previous searches for B_s oscillations noted in Section 1.3.2 are heavily limited by the size of the sample or the sensitivity of each event to mixing.

A fundamental flaw of the simple asymmetry approach is that, in its directness in illustrating the phenomenon of flavor oscillation, it neglects a significant set of information available from the detector. Several measurable quantities provide useful discrimination between signal and background events and the various subcomponents of each. The simplest and most powerful of these quantities is the mass of each B_s candidate, although other attributes of proper time and flavor tagging also contribute discrimination power. In an analysis which is starved for oscillation sensitivity, it is critical that a maximal amount of information is extracted from each event.

Toward that end, the method used for fitting the data in this analysis is the maximum likelihood method, a description of which is provided, for example, in Reference [2]. Given the assumption of a model, the likelihood method finds the values of its parameters that make this dataset most probable to have been observed. However, rather than simply performing a fit in one measurement space, such as flavor asymmetry versus time, the likelihood method involves a joint probability density of *multiple quantities* from each event. This approach combines probabilistic models of several facets of the data, generalizing a one-dimensional least-squares fit into a multi-dimensional space and using a more inclusive set of information. The basic formalism of the likelihood is outlined in Appendix C. The fits of the data to the likelihood are performed numerically via the minimization program Minuit [32, 33].

Model Development

Many features of the likelihood are motivated from first principles, such as the exponential decay time distribution for signal B events or the cosinusoidal B_s oscillation frequency. However, at least as many other modeling decisions must be based on empirical studies of various effects from the detector or event selection algorithms. The final result is a complex likelihood function with tens of probability density components and approximately one hundred parameters. To ensure that every facet of

the model is valid and well-understood, it is important that the likelihood be carefully built up from the simplest components. Once simple attributes of the data are thoroughly characterized, features of progressively increasing complexity are added to the likelihood. It is this cycle of validation and expansion to which the dominant effort of this analysis is directed.

Large samples of B^+ and B^0 mesons are collected alongside the B_s data. These samples are useful in testing the analysis framework before it is applied to B_s events because the light B decays are similar in topology and kinematics to the signal of interest. To facilitate this aspect of development, the likelihood is designed as a generalized model for application to any of the B meson species. As is noted in the context of the following chapters, relevant features of the model are activated or neglected according to the data under consideration.

Development of the likelihood model begins with the mass space, since it is a powerful signal/background discriminant and because fits of mass spectra are well-established at the CDF experiment. Chapter 4 describes how the mass model is constructed and how it is used to study the composition of the data. Of particular note is the use of the high-statistics B^+ and B^0 samples to develop and validate detailed features of the generic mass model. Chapter 5 incorporates a description of B decay time into the likelihood, again using the light B species as a test-bed. Chapter 6 outlines the flavor tagging tools and the process of including correctly calibrated tagging information in the likelihood. Chapter 7 applies the full machinery of the likelihood framework to the problem of observing a signal for B_s^0-$\overline{B}_s^0$ mixing and measurement of the oscillation frequency.

The capability for simple combination of the likelihoods of independent B_s samples is essential for the final observation of a mixing signal with current analyses. As of this writing, no single data sample has sufficient B_s statistics for a stand-alone signal of 5 standard deviations. It is only through a combination with similar samples that this goal can be achieved. Chapter 7 describes the results of closely related CDF analyses and the combination to produce an oscillation signal.

1.5.3 Signal Significance

Section 1.4.1 introduces the method of amplitude scans in Δm_s frequency space, which yields a resonance for a clear signal or allows extraction of a limit if Δm_s cannot be resolved. The strength of the limit depends on the fitted amplitude uncertainties $\sigma_{\mathcal{A}}$. Furthermore, $\sigma_{\mathcal{A}}$ is important for determining the significance of an apparent signal resonance. Accounting for detector resolutions, flavor identification, and statistical fluctuations [27, 34], the uncertainty on amplitude at a given oscillation frequency Δm is

$$\frac{1}{\sigma_{\mathcal{A}}} = \sqrt{\frac{\mathcal{S}}{\mathcal{S}+\mathcal{B}}} \cdot \exp(-\frac{\Delta m^2 \sigma_t^2}{2}) \cdot \sqrt{\mathcal{S}\frac{\epsilon \mathcal{D}^2}{2}} \tag{1.32}$$

where σ_t is the detector's proper time resolution, $\mathcal{S}$ and $\mathcal{B}$ are the numbers of signal and background events, and $\epsilon\mathcal{D}^2$ is a measure of the effectiveness of flavor identification. The expression is written as $1/\sigma_{\mathcal{A}}$ to emphasize the significance of a given

amplitude point, rather than its uncertainty. Of particular note is the fact that σ_t and Δm enter the expression as squared arguments in an exponent, indicating that the significance is rapidly degraded as these values grow. This relation underscores the extreme importance of achieving good proper time resolution in measuring decays, and it explains why large values of Δm are much more difficult to observe than small ones. The dependence may be qualitatively understood by visualizing the need to place each B_s event at its appropriate decay time in the asymmetry curve of Figure 1-4. Uncertainties in proper time "wash out" the measured asymmetry, and larger oscillation frequencies demand correspondingly better resolution. It is Equation (1.32) that defines which B_s decay channels are useful.

It should be emphasized that $\sigma_{\mathcal{A}}$ is not the final measure of the significance of an oscillation signature. The amplitude errors indicate the significance only of each scan point, rather than the complete frequency range. Furthermore, the *p-value*, or probability that random fluctuations of the data could produce a comparable likelihood maximum, should be calculated without any assumption that the true Δm_s is within a particular window around an apparent amplitude resonance. Chapter 7 describes how a likelihood *ratio* taken with amplitude values for the null and signal hypotheses is used to collectively consider the full range of frequency space. Nevertheless, it remains true that $\sigma_{\mathcal{A}}$ is a good approximation of the oscillation significance, and the overall p-value determined by the likelihood ratio is firmly tied to the terms of Equation (1.32).

1.5.4 Opportunity for New Signal Channels

Figure 1-10 depicts the most powerful individual B_s sample used in the April 2006 evidence of oscillations [31] found by the CDF experiment. The selection algorithms for this sample were tuned to identify events containing the decay channel [2] $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$, which is represented in the figure by the dominant peak centered on the B_s mass 5.3696 GeV/c^2. This channel has often been referred to as the "golden mode" because, based on its combination of good signal-to-background fraction with excellent proper time resolution, it provides the largest contribution of statistical significance from an individual sample. In order to restrict the sample to well-understood components of signal and background, only events from Figure 1-10 with mass *greater than 5.3* GeV/c^2 were included in the analysis of proper time and flavor oscillations. This hard cut served to retain nearly all of the targeted signal, while eliminating the difficulty of modeling the more complex "satellite" components by neglecting the lower-mass events altogether. However, given that the combined April 2006 analysis falls short of the 5 standard deviations threshold for definitive observation of an oscillation signal, a larger sample or a more inclusive usage of its events is desirable.

An important attribute of the $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$ mode is that its sta-

[2] The convention throughout this dissertation is that references to a specific decay chain such as $B_s^0 \to D_s^- \pi^+$ also implicitly refer to the charge conjugate decay $\overline{B}_s^0 \to D_s^+ \pi^-$, unless explicitly stated otherwise. This convention is occasionally abbreviated by a non-charge-specific notation like $B_s \to D_s \pi$.

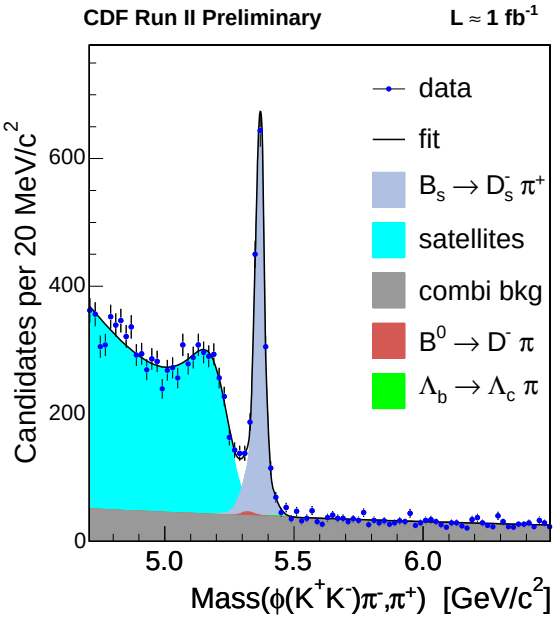

Figure 1-10: Previously used sample of $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$ decays, with backgrounds. For mixing analysis, only events within the mass range $[5.3, 6.0]$ GeV/c^2 were used.

ble decay products [3] are all charged particles. Because the CDF detector excels at recording the passage of charge, such a final state allows this signal channel to be *fully reconstructed*. None of its decay particles escapes detection, and momentum and position information is recorded for each. In contrast, the bump structure in the mass range below about 5.3 GeV/c^2 is caused by the presence of other B_s decay channels whose final states contain one or more particles which are not reconstructed. This scenario arises primarily for decay modes containing *neutral* stable particles, for which the CDF tracking system is not designed (Section 2.2). These decay channels are collectively referred to as satellites or *partially reconstructed* modes. Whereas the fully reconstructed signal peak of Figure 1-10 has a Gaussian shape, indicating only the detector's imperfect mass resolution, the spectra of the partially reconstructed modes are smeared in mass both by the detector resolution as well as by the kinematic effects of unmeasured momentum being carried away by lost particles. The non-trivial shape of the mass spectrum below 5.3 GeV/c^2 results from the sum of contributions from numerous partially reconstructed channels, each with its own smearing from decay kinematics, angular distributions, and topology.

Based on a review of relative branching fractions [2] and the above-mentioned decay characteristics, it can be estimated that the partially reconstructed region of the sample selected for $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$ contains significant contributions

[3] In addition to protons and electrons, CDF analyses typically treat pions, kaons, and muons as "stable" particles because their relativistic speed makes them unlikely to spontaneously decay within the detector volume.

from two similar channels:

- $B_s^0 \to D_s^{*-}\pi^+$, with $D_s^{*-} \to D_s^-\gamma$ or $D_s^{*-} \to D_s^-\pi^0$; and
- $B_s^0 \to D_s^-\rho^+$, with $\rho^+ \to \pi^+\pi^0$.

While the detailed study of these and other contributions is reserved for Chapter 4, several qualitative observations are made as to the prospects for inclusion of these signals in an oscillation analysis. Most obviously, the number of partially reconstructed events in the mass spectrum is large. If a meaningful fraction of these can be harnessed as usable signal, the yield would rival or exceed that of the fully reconstructed signal. Additionally, both of these prospective new channels involve only one neutral particle each, a π^0 or a γ, minimizing the number of particles whose momentum is lost. Moreover, the fact that the partially reconstructed structure in the mass spectrum appears in close proximity to the fully reconstructed peak implies that the amount of lost momentum is relatively small and narrowly distributed. As discussed in Chapter 5, this attribute is critical for achieving good precision in measurement of proper decay time. Equation (1.32) underscores the importance of proper time resolution for statistical significance in a fit of B_s oscillations. Finally, it is expected that these signal events are compatible with existing tools for flavor tagging. The presence of additional neutral particles should not be relevant for their applicability because the tagging methods rely on charged particles. One disadvantage for inclusion of partially reconstructed channels is the potential for introduction of increased background levels in the sample. The fraction of background events due to random combination of particles increases in the lower mass range, and the partially reconstructed B_s events which populate this range cannot be expected to be entirely comprised of usable signal components.

It should be noted that semileptonic analyses of B_s mixing [35], in which the signal decays are of the form $B_s \to D_s^{(*)}\ell\nu X$, also target partially reconstructed events. This is a necessity because the neutrino is never directly detected. The critical difference between those decays and the fully hadronic signals of this analysis is that, whereas the neutrino is likely to carry a large momentum in a semileptonic decay, the new partially reconstructed decays introduced above contain low-momentum neutral particles from subdecays of the B. This configuration is discussed in greater detail in Chapter 4, and it allows the new hadronic signal decays to be "almost fully reconstructed," which is central to its power in resolving a mixing signal.

This dissertation is based on the adaptation of partially reconstructed hadronic signal events for inclusion into analysis of B_s oscillations. The core chapters are devoted to the studies and validation necessary for full characterization of the sample and development of the likelihood model. This work also incorporates other upgraded analysis tools, including a more refined signal selection algorithm and more comprehensive flavor tagging methods. These tools are briefly described in context. The combination of these upgrades with the inclusion of the partially reconstructed B_s events is the foundation for pushing the combined CDF mixing analysis to the level of 5 standard deviations.

Chapter 2

Experimental Apparatus

The higher generations of matter do not occur naturally on Earth, discounting cosmic rays or other external effects. Creation of rare massive particles in a controlled fashion demands a facility which can accelerate available particles to high energies and then collide them together on-demand. The ability to generate a sufficiently energetic environment for these collisions and then reliably detect their by-products are the main criteria for the experimental apparatus.

2.1 The Tevatron Accelerator

The facility responsible for generation of the high-energy collisions from which the datasets are derived is the Fermi National Accelerator Laboratory in Batavia, Illinois, 35 miles west of Chicago. Fermilab, as it is commonly known, is the home of the Tevatron accelerator, which has been the highest energy particle collider in the world for two decades. The Tevatron provides for the acceleration, storage, and collision of two counter-rotating beams of protons and antiprotons. The final energy per particle in each beam is 980 GeV, such that the center-of-mass energy of the $p\overline{p}$ interactions is 1.96 TeV. The creation and isolation of protons and antiprotons and the acceleration of each beam to such energetic levels is a complex process which must be performed in multiple stages. In fact, the Tevatron is only the last leg of the particles' journey through various low- and intermediate-energy accelerators, a sketch of which is shown in Figure 2-1. This sequence is very broadly outlined below, while full discussions are provided in cited references. In addition to the CDF and DØ detectors which directly use the Tevatron collisions, Fermilab also supports several fixed-target experiments and the beams that supply them. While these beams are closely related to the accelerator chain that culminates in the Tevatron, the discussion focuses on the components that are central to generating collisions for this analysis.

Distinct periods of operation of the accelerator and detectors are historically referred to as "Runs". These time spans are delimited by significant upgrades to the facilities. As a loose guide for the occasional reference to the Tevatron's history, it is useful to know that Run I spanned the years 1992-1996, and Run II is a period continuing since 2001. These Runs are typically subdivided into "a" and "b" periods

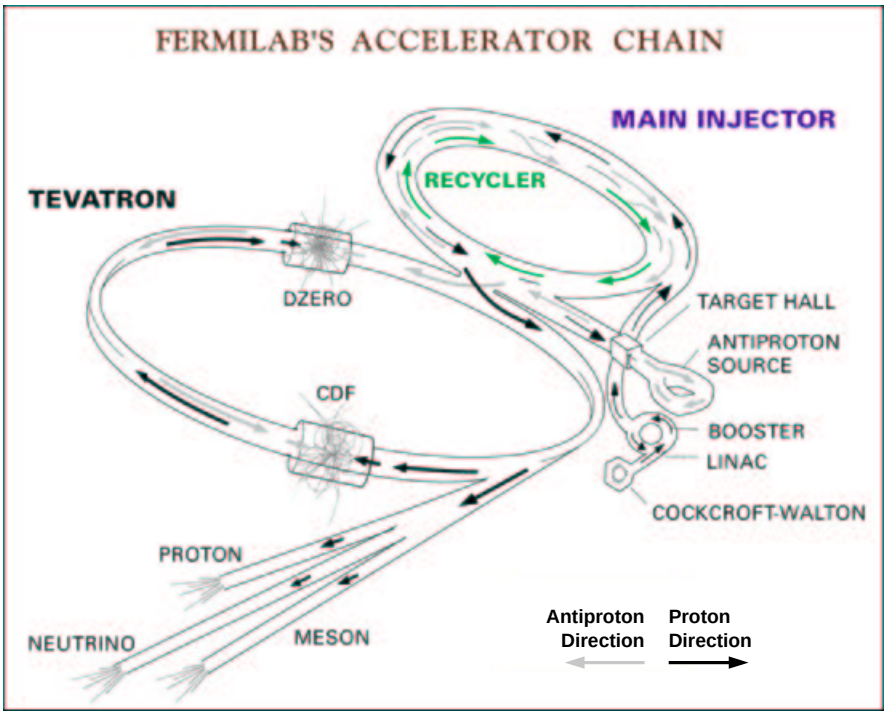

Figure 2-1: Schematic of the Fermilab accelerator complex.

to denote important detector and accelerator upgrades.

Preparation of Beams

The acceleration process must begin with readily available forms of matter. Hydrogen gas is a useful starting material because its nucleus is comprised solely of a proton, requiring no nuclear reactions for extraction. Electrodes ionize hydrogen to produce H^- in the dome of a Cockcroft-Walton [36] electrostatic accelerator, which then accelerates the ions to 750 keV. This beam is directed into the Linac [37, 38] linear accelerator, which brings the ions to 400 MeV. The Linac consists of a series of resonant RF cavities, each of which provides an incremental "kick" to the beam [39]. In order to keep particles in phase with the RF field, the beam is naturally segregated into packets in time which are known as *bunches*. Transverse focus of the beam is maintained by quadrupole magnets positioned between each cavity.

The 400 MeV H^- ions are then injected into the Booster [40], an intermediate-energy synchrotron 75 m in radius. A series of carbon foils strips the ions to bare protons, which are directed in their approximately circular orbit by dipole magnets [39]. Many Linac cycles are injected into the Booster to increase the proton beam intensity. Acceleration is achieved by increasing the RF frequency from 38 to 53 MHz over the course of about 33 ms, during which time the dipole field strength is also increased to maintain constant orbital radius. By passing repeatedly through the same RF cavities, the protons are accelerated to 8 GeV.

This beam is extracted from the Booster into the Main Injector [41], a synchrotron which serves as an important hub for coordinating beams in the accelerator complex. In addition to bringing both protons and antiprotons to 150 GeV as the penultimate

stage of the accelerator chain, the Main Injector provides protons to the Antiproton Source, handles the resulting antiprotons, and delivers protons to Fermilab's fixed-target experiments. As a five-year construction effort ending in 1999, the Main Injector stands as the major upgrade to the laboratory during the Tevatron era. Prior to Run II, the old Main Ring performed most of these tasks, but with reduced efficiency because of its physical location in the same tunnel with the Tevatron.

The creation and storage of antiprotons is a carefully balanced operation. The number of available antiprotons has always placed an upper limit on luminosity throughout the Tevatron era, and preparation of a collection large enough to justify colliding beams requires several hours. The Antiproton Source [42] consists of a fixed nickel target assembly and two small synchrotrons called the Debuncher and the Accumulator. The most time-consuming activity of the Main Injector is in support of antiproton production. Even as the Tevatron is generating collisions, the Main Injector and Antiproton Source are usually working together to prepare antiprotons for the next store. Antiprotons are created in the interactions of 120 GeV protons from the Main Injector with other protons in the nickel target. The spray of secondary particles is focused by a lithium lens and 8 GeV antiprotons are selected by a pulsed dipole field. The overall production efficiency is about 2×10^{-5} antiprotons per incident proton. The Debuncher and Accumulator synchrotrons serve to collect and "cool" the antiproton beam, reducing the spread of its kinetic energy distribution. These stages use methods of stochastic cooling [42, 43], in which the transverse or longitudinal motion of the beam is measured at one point on the ring and precisely timed negative feedback is applied at another. In addition to continual cooling, the Accumulator maintains the antiproton *stack* until it reaches $\sim \mathcal{O}(10^{12})$ particles and the Tevatron is ready for new beam.

The newer Recycler ring [44], inhabiting the same tunnel as the Main Injector, is an extension of these facilities. With construction beginning in 1997, it was designed to literally "recycle" the valuable antiprotons which would otherwise be discarded after circulating in the Tevatron. The Recycler is also equipped with additional electron cooling [45], where a beam of "cold" dense electrons is intermingled with the "hot" antiprotons in a 20 m straight section of the ring. Unlike stochastic cooling, whose effectiveness is inversely proportional to the linear particle density of the beam [42], electron cooling effectiveness is independent of beam density. Although the recycling function of the ring has not been implemented as of this writing, its cooling and storage attributes contributed greatly to the increased luminosity delivered for data-taking after August 2005.

Tevatron

The Fermilab accelerator complex culminates in the Tevatron [46], the highest energy collider in the world. This synchrotron accelerates counter-rotating beams of protons and antiprotons to 980 GeV and maintains them for periods of up to 36 hours. The Tevatron was constructed in the early 1980s to supersede the older Main Ring, a proton-only synchrotron for fixed-target experiments which was limited to $\sim$ 500 GeV by its conventional electromagnets. The new accelerator was viewed as

an opportunity to use superconducting magnets to handle higher beam energies while also incorporating antiprotons to annihilate in head-on collisions [47]. The Run II beam energy of 980 GeV represents an increase over the Run I level of 900 GeV and the 800 GeV used for fixed target experiments before the antiproton era. Given the advantage of counter-rotating particles whose collisions convert all of the beams' energy into the center-of-mass frame, the Tevatron provides the ability to probe physics up to $\sqrt{s} = 1.96$ TeV, which represents the frontier of high energy physics [1].

The Tevatron is built in a tunnel of mean radius 1.0 km and is surrounded by repeating sequences of cryogenically cooled superconducting magnets for bending and focusing of the beams. The ring is divided into 6 *sectors* of equal length, with the CDF-II detector located at the B0 position. The DØ detector is named for its position on the ring.

Initiation of a Tevatron *store* of collisions consists of a standardized sequence of operations, collectively known as *shot setup* [46]. The Tevatron receives 150 GeV protons and antiprotons from the Main Injector, each in 36 bunches, and accelerates them to 980 GeV. The beams are focused to interact at the collision points on which the two major detectors are centered, with a bunch crossing interval of 396 ns.

The critical quantity for determining the rate of collisions is the beam *luminosity*, which may be expressed as

$$L = \frac{f_B N_B N_p N_{\overline{p}}}{2\pi(\sigma_p^2 + \sigma_{\overline{p}}^2)} F(\frac{\sigma_l}{\beta^*}), \tag{2.1}$$

where f_B is the bunch revolution frequency, N_B is the number of bunches per beam, N_p ($N_{\overline{p}}$) is the number of protons (antiprotons) per bunch, and σ_p and $\sigma_{\overline{p}}$ are the Gaussian transverse widths of the bunches. The form factor F provides a small correction for the bunch shape and depends on the ratio of the bunch length σ_l and the beta function β^* at the interaction point [2, 46]. In discussions of instantaneous accelerator performance, luminosity is often given in units of [$\mathrm{cm}^{-2}\mathrm{s}^{-1}$]. For a given physical process with cross-section σ, luminosity L will yield that interaction at a rate

$$R = \sigma L \tag{2.2}$$

with σ in [cm^2] and R in [Hz]. This quantity gives rise to integrated luminosity, which better characterizes the amount of data delivered over time. With σ in the more natural units of picobarnes (or other subunits of barnes), integration of Equation (2.2) shows that L appears in units of inverse cross-section such as [pb^{-1}].

The stated goal of the Tevatron for Run II was to provide instantaneous luminosities of at least 2×10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Difficulties were encountered in the early years of the run, particularly with regard to the ability of the Accumulator to maintain a large enough stack, and initial luminosities were on the order of 1×10^{31} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Because of the incorporation of the Recycler as well as overall improvement in ac-

[1] The Large Hadron Collider is a proton-proton synchrotron under construction at CERN, and its $\sqrt{s} = 14$ TeV collisions will soon expand the high energy frontier. The LHC is scheduled to be commissioned with low-luminosity beam tests in late-2007 [48].

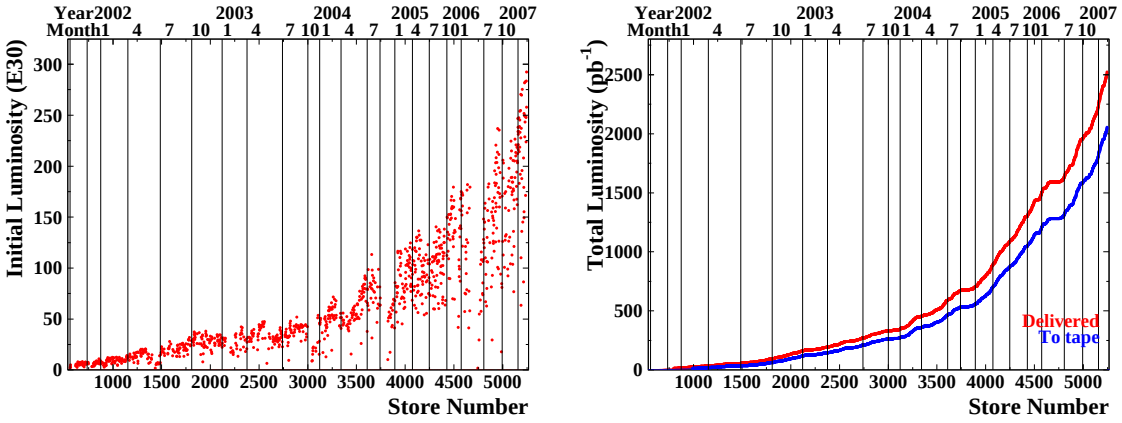

Figure 2-2: Tevatron Run II performance in terms of peak instantaneous (left) and integrated luminosity (right). The figure indicates both the integrated luminosity delivered by the accelerator and the amount recorded to tape by the CDF detector.

celerator operations, recent performance has the Tevatron regularly achieving peak luminosities of $2.25 - 2.75 \times 10^{32}$ cm^{-2}s^{-1}. Luminosity is at its peak when the beams initially begin colliding, followed by an approximately exponential decrease with time. Instantaneous luminosity generally decreases due to direct $p\overline{p}$ annihilation as well as long-range interaction between the beams which leads to larger emittances and lost particles. A Tevatron store is discarded once the luminosity has decreased by a factor of approximately 5, as long as enough antiprotons have been collected to begin anew. The total lifetime for a store covers a wide range, with 10 to 30 hours being common. Figure 2-2 depicts the history of Tevatron Run II performance with respect to both instantaneous and integrated luminosity. This analysis is performed with 1.0 fb^{-1} of B_s data, which represents the total data collected by the CDF detector between February 2002 and January 2006. Chapter 3 discusses the dataset in detail.

2.2 The CDF-II Detector

The datasets for this analysis are collected by the CDF-II detector [49], one of the two major particle detectors located on the Tevatron ring. The generic nature of CDF's full name, "Collider Detector at Fermilab" reflects the fact that it was the first detector built on the Tevatron for the $p\overline{p}$ era. The CDF-II nomenclature refers to the current incarnation of the detector [2] , which underwent major upgrades in many of its components to coincide with Tevatron Run II.

The detector's general task is to record the passage of particles resulting from $p\overline{p}$ collisions with as much information as possible on momentum, energy, and position. Its organization is generally axially symmetric about the beamline, covering almost all solid angles except for the regions closest in angle to the beam. Rather than being

[2]While the name of the detector is technically CDF-II, it is usually referred to as simply "CDF".

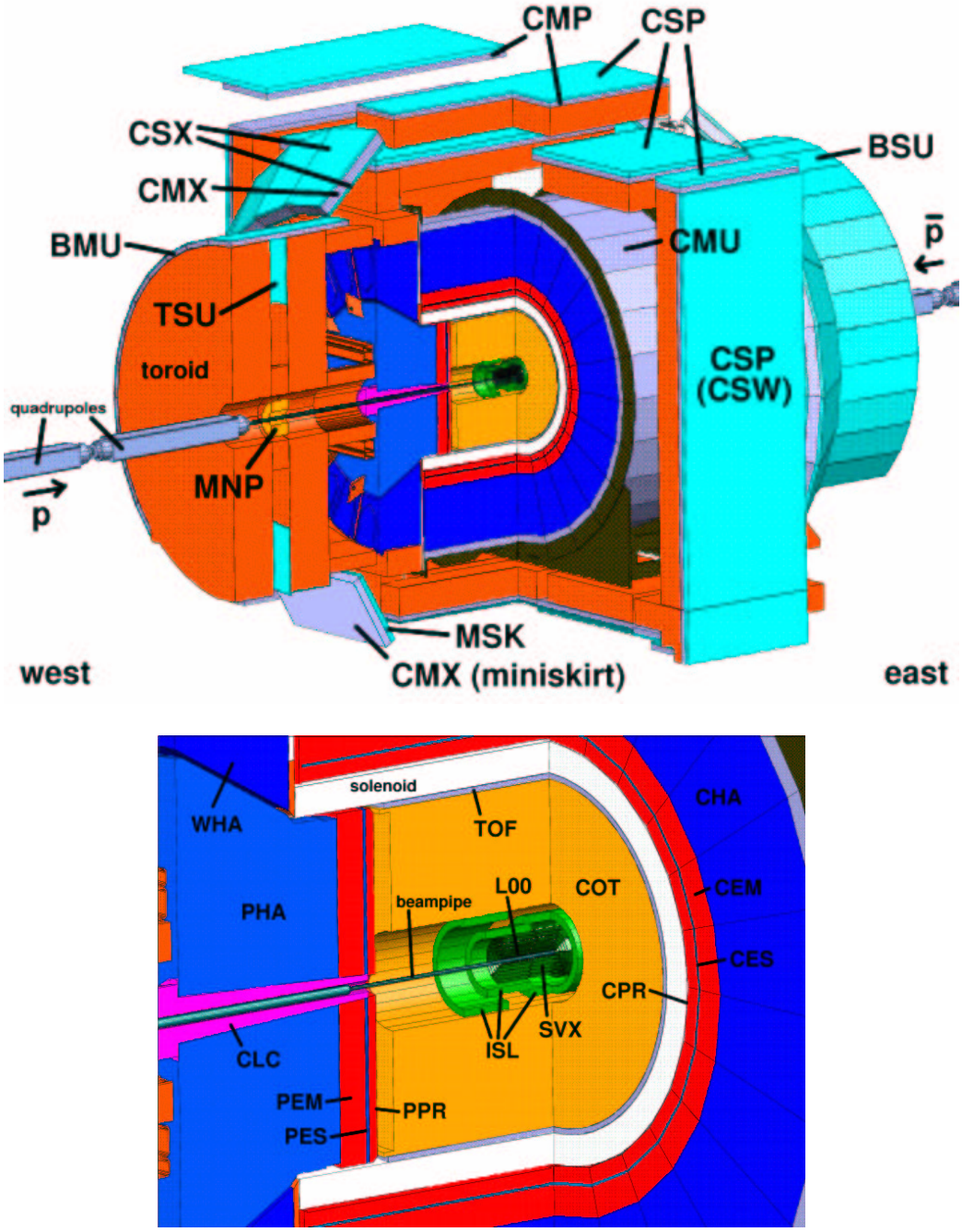

Figure 2-3: The full CDF-II detector with a quadrant removed (top) and a zoomed-in view of the inner subdetectors (bottom). Component labels are explained in the text.

designed for one specific class of high energy physics measurement, CDF has a versatility that qualifies it as a "multi-purpose" detector. A diagram of the CDF-II detector is shown in Figure 2-3, with a quadrant removed to expose the different components. These subdetectors fall into a few general categories. At the center of the detector

is the beam pipe, for which beryllium is used because it combines good mechanical properties with the lowest nuclear interaction cross section of all materials. The innermost subdetectors comprise the "integrated tracking system", which has cylindrical components centered concentrically about the beam pipe. Silicon microstrip layers occupy the smallest radii and are surrounded by a barrel-shaped drift chamber. As charged particles traverse this material, they leave ionization *tracks* which are reconstructed in the detector electronics. These tracking components perform the critical task of precisely measuring the particles' trajectories and momenta, knowledge which is indispensable to B physics analyses. The tracking system is surrounded by the Time-of-Flight detector, which is composed of bars of scintillating material. The timing information from this system, combined with the momentum from the tracking system, provides particle identification for low-momentum charged particles. All of the above components are encompassed by a superconducting solenoidal coil which generates a 1.4 T magnetic field. The superconducting nature of this magnet is as essential as it is for the Tevatron magnets, since it must continually carry approximately 4400 A of current without a large volume of the detector being devoted to cooling. This field bends the path of charged particles as they traverse the tracking system, allowing for inference of their momenta. Immediately outside the solenoid are calorimetry systems for measurement of the energy deposited by both charged and neutral particles. Each calorimetry chamber uses absorbing metal layers coupled with scintillators or proportional counters. The inner calorimeters are specialized to record energy deposition by electrons and photons via their electromagnetic interactions, while the outer calorimeters are designed to collect the energy of hadronic particles as they are absorbed by atomic nuclei. The outermost components of the CDF detector are the muon chambers, since these minimum ionizing particles are unlikely to be stopped by the matter at smaller radii. The isolation of the muons in this outer layer is accentuated by layers of steel shielding immediately inside the muon chambers. These chambers are largely composed of drift cells which serve to record the creation of muons and their positions as they exit the detector.

Each of these subdetectors is described in greater detail in following sections. While almost every component of the CDF detector contributes to the data in this analysis, the discussion of each is weighted by its importance for the observation of B_s^0-$\overline{B}_s^0$ oscillations. Specifically, the tracking system is absolutely critical, since the physical trajectories and momenta of tracks are the basis for signal identification as well as measurement of proper time in B_s decays. As discussed in Section 1.5.3, sensitivity to the oscillation signal is strongly dependent on proper time resolution. Additionally, timing information from the Time-of-Flight system is complemented by track energy loss in the drift chamber to provide particle identification (Appendix E). Reliable particle identification information is the basis for the powerful Same Side Kaon Tagger (Section 6.3) and is also incorporated into the Artificial Neural Network algorithm for signal selection (Section 3.3). The muon chambers are important for identifying muons for the Muon Tagger, which is the most powerful of the Opposite Side Taggers (Section 6.2.1). The calorimeters are the least-used components of the detector for B physics analysis, and the brevity of their description reflects that fact.

However, the electromagnetic calorimeters are useful in identifying electron tracks for application in the Electron Tagger (Section 6.2.1).

Another similarly important aspect of the CDF detector is the trigger system. CDF uses dedicated triggers which examine incoming data in real-time to collect a sample enriched in heavy-flavor signal. Section 2.3 describes the trigger system architecture.

2.2.1 Standard Definitions at CDF

Definitions for terms and conventions used in CDF analyses are compiled here in order to streamline discussion throughout this dissertation.

Because of its axial symmetry about the beam pipe, CDF uses a cylindrical coordinate system (r, φ, z) with the origin at the center of the detector and the z axis along the nominal direction of the proton beam. In cases where a rectangular system is simpler, the y axis points upward, which also defines the direction of the x axis by right-handedness. The plane perpendicular to the beam (that is, the x-y plane) is referred to as the *transverse plane*, while the projection of a particle's momentum vector onto this plane is called the transverse momentum p_T. Projections of various quantities onto the transverse plane are commonly used in CDF analyses for multiple reasons. In the relatively clean environment of e^+-e^- colliders, the point-like leptons annihilate completely such that the initial momentum of the resulting system along the longitudinal (z) direction is known. By contrast, the $p\overline{p}$ collisions in the Tevatron are the result of interactions between the constituent quarks or gluons which carry an indeterminate fraction of the particles' total momentum. The center-of-mass frame of the parton collisions is boosted along the beam direction by an unknown amount. However, because the transverse momenta of the partons are completely negligible, the center of mass can be considered at rest in this plane. Additionally, based on the design of the COT drift chamber, momenta in the transverse plane are measured much more precisely than in the longitudinal direction.

Charged particles moving through the homogeneous solenoidal magnetic field follow helical trajectories which are referred to as *tracks* when reconstructed. Although this nomenclature is derived from the visible tracks in early bubble and cloud chambers, these tracks are reconstructed entirely by electronics from the position of *hits* of charge deposition in the various detector layers. Five parameters are necessary to uniquely parametrize a helix in three dimensions. The CDF data applies three of these parameters to describe position and two others to describe the momentum vector at that position. This special position is the point of closest approach of the track helix to the beamline. The associated parameters are d_0, φ_0, and z_0, which represent the magnitude, azimuthal angle, and z coordinate of the position vector, respectively. The momentum vector is parameterized by the track curvature $C = \frac{q}{2R}$ (the reciprocal of the circular diameter of curvature) and the angle of the momentum in the r-z plane $\cot\theta$. The curvature is signed to match the sign of the particle's

charge and it directly yields the transverse momentum by

$$p_T = \frac{cB_{solenoid}}{2C} \tag{2.3}$$

The polar angle θ is defined with respect to the positive z axis and covers the domain $[0, \pi]$. It yields the ratio between the transverse and longitudinal momentum components via $\cot\theta = \frac{p_z}{p_T}$. The value φ_0 implicitly defines the direction of the transverse momentum vector at the point of closest approach to the beam, while d_0 is simply the particle's impact parameter (distance of closest approach) with respect to that point. The value of d_0 also carries a sign according to the convention $\text{sign}([\hat{p} \times \hat{d}] \cdot \hat{z})$, where $\hat{p}$, $\hat{d}$ and $\hat{z}$ are unit vectors in the direction of p_T, d_0 and z, respectively. In describing the angle between the particle momentum and the z axis, an alternate variable is the pseudorapidity η which is defined as

$$\eta \equiv -\ln\tan(\theta/2) \tag{2.4}$$

Pseudorapidity is equivalent to rapidity $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) = \tanh^{-1}\beta$ in the relativistic limit $pc \gg mc^2$, which is often but not always true for the tracks used in this analysis. However, η has the advantage that it can be determined directly from measurement of a track. Areas of the detector traversed by tracks with large positive or negative η are commonly known as forward regions, while the volume around $\eta \approx 0$ is called the central region. Pseudorapidity is often combined with azimuthal angle to define the angular distance or *opening angle* between two tracks:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2} \tag{2.5}$$

Reconstructed tracks are used as the building blocks for vertices which indicate the position of the decays of unstable particles. Vertices are assigned transverse momentum $\vec{p_T} = \sum \vec{p_T}^{\text{trk}}$, the vector sum over the momenta of their constituent tracks. The mass of the decay particle may be calculated from the momentum and angular measurements from the observed tracks. The position of the $p\overline{p}$ collision, from which many tracks originate (typically $\sim \mathcal{O}(50)$), is referred to as the *primary vertex*. The displacement of other vertices with respect to the primary vertex is often described by L_{xy}, the transverse decay length in the laboratory frame, which is defined as

$$L_{\text{xy}} = \vec{r} \cdot \hat{p_T} \tag{2.6}$$

where $\vec{r}$ is the displacement of the decay vertex in the transverse plane, and $\hat{p_T}$ is the unit vector in the direction of $\vec{p_T}$. That is, L_{xy} is the projection of the vertex's position vector onto the direction of its transverse momentum. Each vertex is also assigned an impact parameter d_0 under the same convention as tracks, with the total vector momentum extrapolated back to the point of closest approach to the primary vertex. The vertex d_0 and L_{xy} are related as legs of a triangle whose hypotenuse describes the transverse displacement between the primary and decay vertex.

Given a vertex which describes the decay of an unstable particle originating at

the primary vertex, these measurable quantities may be used to infer the *proper time* elapsed in the particle's reference frame between its creation and decay:

$$ct = \frac{L_{\mathrm{xy}} \cdot M}{p_T}, \tag{2.7}$$

where M is the mass of the particle (often assumed to be the world average value). Proper time and the resolution with which it can be measured are critical to this analysis, and these quantities are discussed at length in Chapter 5.

2.2.2 Tracking Systems

The CDF detector has a cylindrical tracking system for the primary purpose of precisely and efficiently measuring charged-particle momenta and trajectories. All components are immersed in a 1.4 T solenoidal magnetic field to allow for inference of the momenta. The innermost tracking component is a silicon microstrip detector which consists of three subdetectors. A layer of silicon sensors called Layer 00 (L00) [50] is installed directly onto the beryllium beam pipe at a radius of about 1.5 cm from the colliding beams. This inner layer is followed by five concentric layers of silicon sensors known as the Silicon Vertex detector (SVX-II) [51], located at radii between 2.5 and 10.6 cm. The Intermediate Silicon Layers (ISL) [52] are the outermost silicon systems, consisting of one layer at a radius of 22 cm in the central region and layers at radii 20 and 28 cm in the forward regions. The silicon detector is surrounded by the Central Outer Tracker (COT) [53], a 3.1-m-long cylindrical open-cell drift chamber covering radii from 40 to 137 cm. Figure 2-4 depicts the components of the tracking system in the r-z plane embedded within other structures of the CDF detector. The total system is referred to as the "integrated" tracking system because information from each component is combined, joining the superior momentum resolution of the COT to the excellent impact parameter resolution of the silicon. These subdetectors and their usage are detailed below.

Silicon Detectors

Silicon tracking detectors are used to obtain precise position measurements along the trajectory of a charged particle. Fundamentally, a silicon detector is a reverse-biased p-n junction. As in any material, a charged particle passing through the silicon wafer causes ionization. In the case of a semiconductor material, this excitation produces electron-hole pairs in the junction's depletion region. The charge is gathered as electrons drift toward the positively biased anode while holes drift toward the cathode. The presence of the reverse-bias increases the inherent potential across the junction, increasing the strength of the field that sweeps out the signal ionization charge while decreasing the flow of charge from background thermal excitations. The amount of charge collected is, to first order, proportional to the path length traversed in the detector material by the charged particle.

The p or n sides of the junction can be separately segmented into strips between which charge cannot flow. Reading out the charge deposition on every strip yields

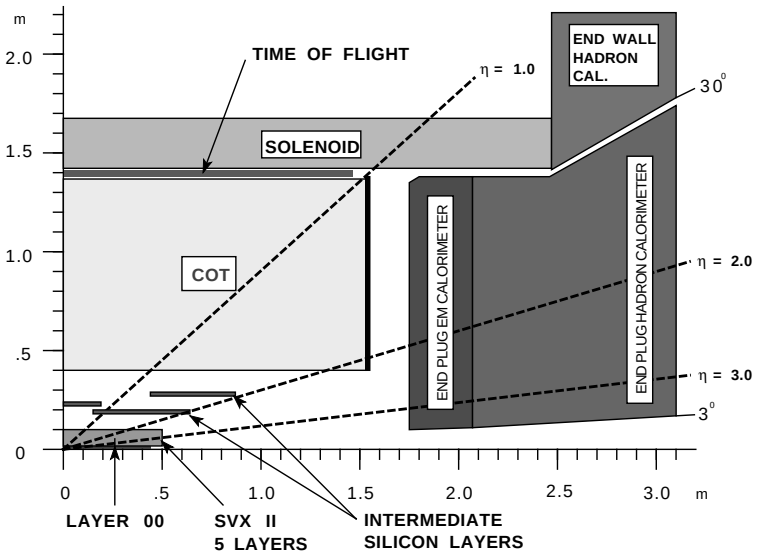

Figure 2-4: r-z view of the CDF tracking system and surrounding subdetectors.

sensitivity to the position of the charged particle. All of the CDF silicon tracking components are implemented as microstrip detectors, in which the typical distance between two strips is of order 60 μm and each layer has a nominal thickness of 300 μm. The charge deposited by an ionizing particle is referred to as a *hit*. A single particle passing through a silicon layer leaves a hit in one or more adjacent strips, creating a hit *cluster*. The nominal hit position is weighted by the charge collected in each affected strip.

CDF silicon detectors are of a single- or double-sided design. The innermost layer, L00, is made of single-sided silicon sensors which provide only r-φ measurements, while the SVX-II and ISL are made of double-sided sensors. In each layer, the p side of the junction has strips parallel to the z direction for optimal position measurements in the r-φ plane. The double-sided layers have n-side strips rotated with respect to the p side by a so-called *stereo angle*, which provides the important benefit of two-dimensional information on the location of its hit clusters.

The silicon sensors are supported by carbon rails in assemblies called *ladders*. The general configuration of each ladder has four sensors stacked end-to-end along the z direction, with readout electronics mounted onto the ladder ends. These units are combined as overlapping plates in φ in an approximately cylindrical configuration, creating *barrels* for each layer. The use of discrete ladders causes the barrels to be naturally segmented into 12 wedges, each covering 30° in φ with a small overlap at the edges. A cross-sectional r-φ view of this organization is depicted in Figure 2-5. Each SVX-II layer consists of three barrels which are positioned end-to-end along the z axis, separated by beryllium bulkheads through which the sensors' electronics and cooling equipment are routed. The three-barrel SVX-II span of approximately 100 cm

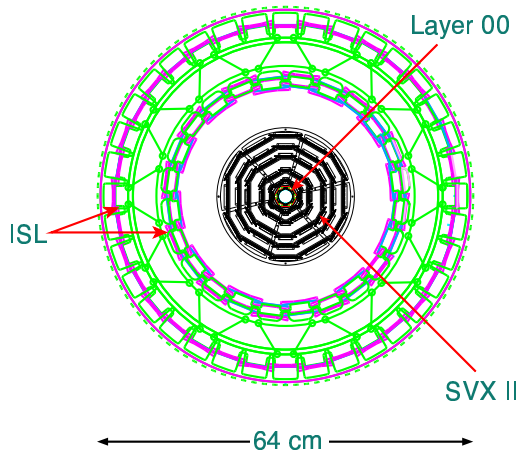

Figure 2-5: Concentric configuration of the silicon layers in the r-φ plane.

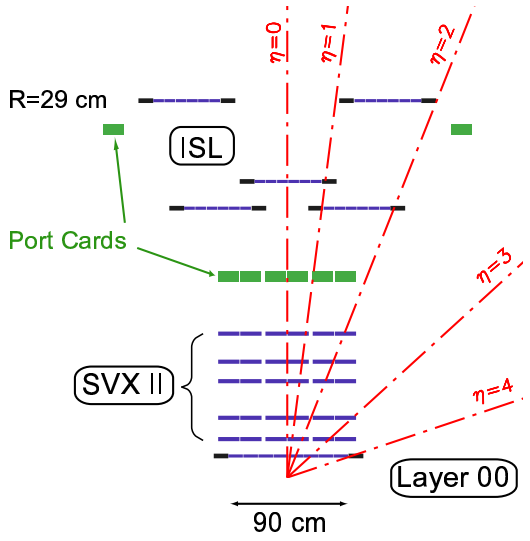

Figure 2-6: Coverage of the silicon subdetector systems in the r-z plane. The r and z axes have different scales.

covers the nominal interaction point in the center of the CDF detector. Figure 2-6 shows a simplified sketch of the coverage of the silicon detector subsystems in the r-z plane. The use of multiple layers with complete coverage in φ ensures that a properly functioning silicon system will register several hits for each charged particle leaving the interaction region.

The overlapping wedges of the Layer 00 detector [50] are installed directly onto the beam pipe, alternating between radii of 1.35 and 1.62 cm. Such a small radius is desirable for the precision it allows in extracting impact parameter measurements: fitted tracks with a L00 hit need only be extrapolated a minimal distance from their last known position toward the primary vertex. This advantage is supported by the

use of sensors with even narrower microstrips than those of the other layers. The distance between identical points on successive strips (the *pitch*) is 25 μm, with only alternate strips being read out to increase the signal-to-noise ratio. L00 hits have a typical spatial resolution of 6 μm. Since L00 is installed within a limited volume and its primary contribution to tracking is improved d_0 resolution, this layer is designed of only single-sided sensors for r-φ information. L00 is slightly shorter than the SVX-II, spanning 80 cm in z. The layer's 13,824 channels are read out by electronics located outside the tracking volume at larger z values. Because of its close proximity to the $p\overline{p}$ collisions and associated beam debris, L00 is fabricated from radiation-hardened silicon with an expected lifespan beyond 5 fb^{-1}. Additionally, the active areas of all layers in the silicon subdetectors are cooled to approximately 0° C to minimize damage from radiation. The improvement in impact parameter resolution due to use of L00 is noted in Section 3.2.

The SVX-II [51] layers are based on the original silicon subdetector of the previous CDF-I incarnation. Spanning the radii between 2.5 and 10.6 cm, its five layers are double-sided. As shown in Table 2.1, two layers have a small (1.2°) stereo angle and three have a 90° stereo angle. The pitch of the axial strips ranges from 60 to 65 μm, as does that of the small-angle stereo strips. In order to limit the number of channels in the 90° stereo layers, these strips have pitches of about 125 or 141 μm, as indicated. The entire SVX-II subdetector consists of approximately 400,000 channels.

Property	Layer 0	Layer 1	Layer 2	Layer 3	Layer 4
number of φ strips	256	384	640	768	869
number of Z strips	256	576	640	512	869
stereo angle [degrees]	90	90	+1.2	90	-1.2
φ strip pitch [μm]	60	62	60	60	65
Z strip pitch [μm]	141	125.5	60	141	65
active width [mm]	15.30	23.75	38.34	46.02	58.18
active length [mm]	72.43	72.43	72.38	72.43	72.43

Table 2.1: Specifications for each SVX-II layer.

In contrast to the inner silicon subdetectors, the layers of the ISL [52] cover different ranges in z. Its central layer at radius 22 cm detects tracks with $|\eta| < 1$ with respect to $z = 0$ and is useful for extrapolating COT tracks into the SVX. The forward layers at radii 20 and 28 cm provide measurement of tracks with $1 < |\eta| < 2$, where coverage of the drift chamber is incomplete. The ISL layers are double-sided, with each providing axial and small-angle (1.2°) stereo information. The strips have pitches of 55 μm the axial direction and 73 μm in the stereo direction. Due to the surface area associated with its larger radii, the ISL contains approximately 300,000 total channels. The full subdetector spans 190 cm in z.

Each layer of the silicon detectors requires $\sim \mathcal{O}(10\,\mu s)$ for full readout of the analog integrated charge signal. The electronics employ a "deadtimeless" data acquisition system [54], in which previously collected charge is digitized and processed as new analog signal is integrated. These parallel operations allow the detector to keep pace

with bunch-crossing intervals as low as 132 ns. Additionally, individual silicon strips are only read out when signal is detected nearby, decreasing the requirements on timing and data capacity.

Central Outer Tracker

The COT drift chamber [53] provides precise measurements of transverse momentum in the r-φ plane. Based on the orientation of its sense wires, measurements in the r-z plane for determination of longitudinal momentum are less accurate. The COT contains 96 layers of sense wires which are radially grouped into eight *superlayers*, as depicted in the end plate cross-section of Figure 2-7. The superlayers are divided in φ into *cells*, each of which has 12 sense wires and a maximum drift distance of 0.88 cm. To keep this maximum distance constant across superlayers, the number of cells per superlayer scales approximately with radius. The entire COT contains 30,240 sense wires. Even-numbered superlayers contain *axial* sense wires strung along the z direction, while odd-numbered superlayers use *stereo* wires strung at a small angle (2°) with respect to the z direction. The active volume of the COT begins at radius 43 cm from the nominal beamline and extends out to radius 133 cm. The chamber is 310 cm long. Particles originating from the nominal interaction point at $z = 0$ which have $|\eta| < 1$ pass through all 8 superlayers of the COT, while those with $|\eta| < 1.3$ pass through at least 4 superlayers.

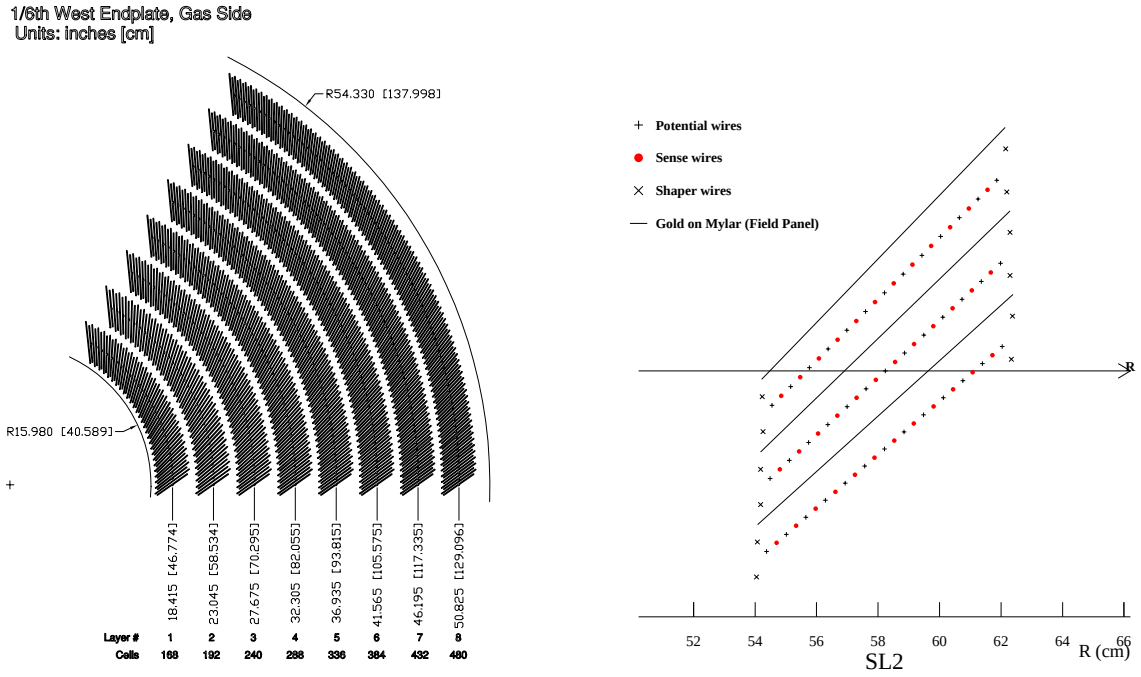

Figure 2-7: Cross-sectional views of COT superlayers (left) and individual wires in three COT cells (right). The Lorentz angle of 35° is clearly visible in the cells.

The organization of wires within a cell, as shown in cross-section in Figure 2-7 for

superlayer 2, consists of sense wires for collecting liberated electrons and potential wires for additional field shaping, all of which are surrounded by a field (or cathode) sheet on either side. Both the sense and potential wires are made of gold-plated Tungsten and are each 40 μm in diameter. The field sheets are 6.35 μm thick Mylar with vapor-deposited gold on both sides, and each sheet is shared by neighboring cells. The sense wires are held at potentials of 2.6 - 3.0 kV, while the potential wires are held at 1.0 - 2.0 kV and the field sheets are grounded. The wires are each held at slightly different potentials to provide a uniform drift field.The maximum drift time is less than 200 ns, well under the bunch-crossing interval of 396 ns.

The COT is filled with an argon-ethane gas and isopropyl alcohol mixture in the ratio 49.5:49.5:1. This mixture is chosen to have a constant drift velocity across the width of a cell, which simplifies calculations of timing. When a charged particle passes through a cell, the gas in its vicinity is ionized, allowing freed electrons to drift toward the sense wires. The $1/r$ dependence of the electric field in a cylindrical system causes the field to grow very large close to the sense wires. The high field strength results in an avalanche discharge as the originally ionized electrons acquire enough energy to free other electrons, which are in turn accelerated to high energy. This effect provides a gain of $\sim 10^4$ for the collected charge. Because the COT is immersed in a magnetic field, the trajectories of the drifting electrons are deflected off of the simple electric field lines. To compensate for this bending, the supercells are tilted by a Lorentz angle of 35° with respect to the $\hat{r}$ vector such that the drift trajectories remain azimuthal. This configuration is also depicted in Figure 2-7.

Signals on the sense wires are processed by the ASDQ (Amplifier, Shaper, Discriminator with charge encoding) chip, which provides input protection, amplification, pulse shaping, baseline restoration, discrimination, and charge measurement [55]. The width of the discriminator output encodes the measurement of integrated charge in each hit, which is directly related to specific ionization of the particle species traversing the chamber. This $\mathrm{d}E/\mathrm{d}x$ information is stored for particle identification (Appendix E) once tracks have been reconstructed. Hit times are later processed by pattern recognition software to form helical tracks. The hit resolution of the COT is approximately 140 μm. As described in Section 3.2, the final track reconstruction efficiency ranges from 95 - 99% for tracks which pass through all 8 COT superlayers [56] and the transverse momentum resolution is $\sigma_{p_T}/p_T = 0.0015\ p_T/(\mathrm{GeV}/c)$.

2.2.3 Time of Flight

The Time of Flight (TOF) [57] subdetector is designed to distinguish pions, kaons and protons by measuring the time required for these particles to travel from the primary vertex of the $p\overline{p}$ collision to the TOF radius. The system is most effective for low-momentum particles ($p_T \lesssim 1.4$ GeV/c), as the species are well-separated by their speeds in this regime. The primary motivation for inclusion of TOF in the CDF-II detector is the application of particle identification information in same-side b-flavor tagging tools, as discussed in Section 6.3.

The TOF system is located in the 4.7 cm radial gap between the tracking system and the cryostat of the superconducting magnetic coil. It consists of 216 bars of scin-

tillating material, each of which is 280 cm in length and 4 cm $\times$ 4 cm in cross-section. The bars are arranged in a cylindrical configuration at radius 138 cm, providing pseudorapidity coverage of approximately $|\eta| < 1$. Charged particles traversing the scintillator bars deposit energy which is converted by the material to flashes of light at visible wavelengths. These photons are transmitted down the bars to photomultiplier tubes (PMTs) on each end which provide a gain of $\sim 30,000$ on the signal. This analog signal is processed by a pre-amplifier circuit mounted on each PMT before being passed to readout electronics, where the signal time and amplitude are digitized. The time-to-digital conversion (TDC) information is taken from the time when the signal pulse reaches a fixed discriminator threshold. Since the threshold is reached sooner for a large-amplitude pulse than for a simultaneous small pulse, the TDC time is dependent on signal amplitude. This so-called "time-walk" effect is corrected for by the digitized amplitude information and the combined readings of the PMTs on each end. The best time resolution is achieved for large pulses. Because the light is attenuated by the scintillator material, signals originating in close proximity to a PMT end yield optimal resolution, while the timing precision is degraded with increasing axial distance.

TOF information contributes significantly to CDF's particle identification capability. The TOF detector measures the time t_{flight} between a particle's generation in the primary vertex at time t_0 and its arrival at the scintillator material. Given the three-dimensional track momentum p and helical path length L, the particle mass is calculated as

$$m = \frac{p}{c}\sqrt{\frac{c^2 t_{\text{flight}}^2}{L^2} - 1} \tag{2.8}$$

The timing resolution relies on calibration of the response in each TOF channel and achieves optimal values of about 110 ps. The system's utility in resolving pions, kaons, and protons in the low-momentum regime is complemented by $\mathrm{d}E/\mathrm{d}x$ information extracted from the COT. Appendix E provides a full discussion of CDF particle identification capabilities.

2.2.4 Calorimeters

Calorimetry at CDF is used to directly measure particle energy deposition and to distinguish electrons and photons from hadronic particles. The calorimeter subdetectors are divided into two categories: electromagnetic calorimeters are optimized to absorb the energy of electrons and photons via EM interactions, while hadronic calorimeters are designed to stop long-lived mesons and baryons primary via strong interactions with nuclei. Calorimetry is particularly important in top quark analysis and in direct searches for Higgs and new physics. In B_s mixing analyses, however, these detector components are used only in identification of electrons for application in opposite-side flavor tagging and classification of semileptonic B decays [35]. As such, the following discussion is limited in scope, while detailed descriptions can be found in the CDF-II Technical Design Report [49] and component-specific references.

The calorimetry subdetectors surround the solenoid coil and tracking volume with

good central and forward coverage. Both categories of calorimeter are designed as sampling devices, in which layers of an absorbing metal material are alternated with signal-producing scintillator layers. After exiting the tracking system, particles first encounter a layer of EM calorimeter systems, in which lead is used as the absorber to induce bremsstrahlung in energetic electrons or the electron-positron pairs produced by photon conversions. The resulting shower of secondary electrons and photons is detected by PMTs attached to successive layers of scintillator, whose signal strength is interpreted to correspond with the initial particle energy. The remaining spray of outgoing particles then traverses the hadronic calorimetry layers, in which iron is used to induce hadronic showers via a variety of nuclear interactions. The charged secondary particles resulting from the hadronic interactions produce signal in the corresponding scintillators. As minimum ionizing particles, most muons pass through the calorimetry systems without significant energy loss.

Each subdetector is segmented into *towers* in η and φ which point to the $p\overline{p}$ interaction region. The overall calorimeter system is divided into *central*, *wall*, and *plug* regions. The components are denoted as the central electromagnetic (CEM), central hadronic (CHA), wall hadronic (WHA), plug electromagnetic (PEM), and plug hadronic (PHA) subsystems. The central calorimeters cover the pseudorapidity region $|\eta| < 1.1$ with respect to $z = 0$, with each tower spanning about 0.1 in η and 15° in azimuthal angle. Plug calorimeters extend the coverage to $|\eta| = 3.6$ and use variable segmentations of 0.1 to 0.6 in $|\eta|$ and 7.5° to 15° in φ. The WHA system bridges the gap between the central and plug hadronic components. Hadronic towers are located directly behind corresponding EM towers and have matching segmentation.

The CEM [58] and PEM [59] systems consist of lead sheets interspersed with scintillator as the active detector medium. Lead is chosen for its large EM cross section, and its thicknesses ensure that punch-through of electrons and photons into the hadronic towers is minimal. Both electromagnetic calorimeters are embedded with shower maximum detectors [60] as well as a pre-shower detector [61] to provide enhanced electron identification. The EM calorimeters achieve energy resolutions of approximately $\sigma_E/E \sim 15\%/\sqrt{E}$ for particles at normal incidence [49, 59]. Table 2.2 lists details for the major subsystems.

Hadronic calorimeters [62] rely on strong interactions of the absorber nuclei with traversing particles to induce showers. Because this cross section is lower than that of EM interactions, the hadronic layers of the calorimeter system must be thicker than the EM components to reach a useful depth of interaction lengths. Iron is used because its nuclear cross section is comparable to that of lead while being lighter and less expensive. Energy resolution in hadronic calorimetry is degraded by the loss of secondary neutrinos and muons and by the energy expended in exciting and breaking up the absorber nuclei. The hadronic resolution is approximately $\sigma_E/E \sim 50 - 80\%/\sqrt{E}$ for particles at normal incidence [62]. Table 2.2 lists details for each subsystem.

System	$\|\eta\|$ coverage	Thickness	Energy Resolution
CEM	$\|\eta\| < 1.1$	19 X_0	$13.5\%/\sqrt{E \sin\theta} \oplus 2\%$
PEM	$1.1 < \|\eta\| < 3.6$	21 X_0	$16\%/\sqrt{E \sin\theta} \oplus 1\%$
CHA	$\|\eta\| < 1.1$	4.5 λ_0	$50\%/\sqrt{E \sin\theta} \oplus 3\%$
WHA	$0.7 < \|\eta\| < 1.3$	4.5 λ_0	$75\%/\sqrt{E \sin\theta} \oplus 3\%$
PHA	$1.3 < \|\eta\| < 3.6$	7 λ_0	$80\%/\sqrt{E \sin\theta} \oplus 5\%$

Table 2.2: Pseudorapidity coverage, thickness, and energy resolution $\frac{\sigma_E}{E}$ for the CDF calorimeter subdetectors [49, 59, 62]. Thicknesses are listed with respect to radiation length X_0 or interaction length λ_0. Hadronic interaction lengths are averaged over particle multiplicities. Energy is in units of GeV, while the $\oplus$ symbol indicates a constant resolution term to be added in quadrature to the energy-dependent term. θ represents angle of incidence with respect to a tower's normal vector.

2.2.5 Muon Systems

Muons are minimum ionizing particles and do not create showers in the calorimetry systems at the high energies produced in Tevatron collisions. Whereas other long-lived charged particles are well contained by the calorimeters, energetic muons typically traverse and exit the CDF detector before decaying. The role of the muon systems is to reconstruct short tracks at the outermost radii of the detector such that they can be matched to those of the integrated tracking system to identify muons. Because a muon can only produce a few hits in these subdetectors, and because of the relative coarseness of the hit resolutions, these short tracks are known as muon *stubs*.

The CDF detector contains four muon systems [63, 64]: the central muon (CMU), the central muon upgrade (CMP), the central muon extension (CMX), and the intermediate muon (IMU) detectors. The coverage of each subdetector is depicted in Figure 2-8.

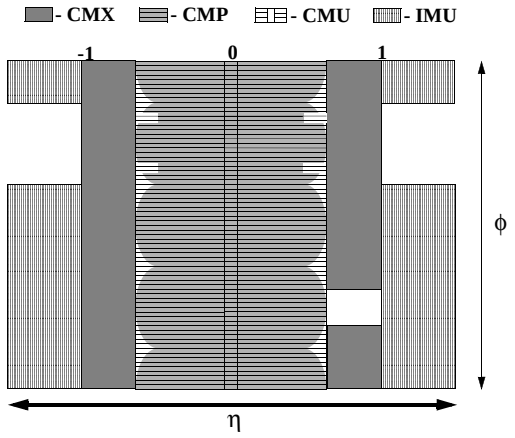

Figure 2-8: Coverage of the muon systems in η-φ.

The CMU detector [65] is located around the outside of the central hadronic calorimeter at radius 347 cm. It is segmented in φ into 144 modules, each of which contains 4 layers of 4 rectangular drift cells, and is divided into east (positive z) and west (negative z) halves centered in $z = 0$. Its sense wires are immersed in the same argon-ethane-alcohol gas mixture used in the COT. The CMP is an additional central muon detector positioned behind 60 cm thick steel shielding. In combination with the CMU, the CMP further improves the purity of muon identification. Muon candidates with stubs in both the CMU and the CMP are referred to as CMUP muons. The CMX detector provides a pseudorapidity extension to the central muon systems, covering $0.6 < |\eta| < 1.0$. It consists of a conical arrangement of drift cells within a sandwich of scintillators. The CMX is azimuthally segmented in 15° wedges, each comprised of 8 layers of rectangular drift cells. The IMU consists of a barrel of drift chambers and scintillator counters mounted on the outer radius of two steel toroids with additional counters between them. Its east and west halves cover $1.0 < |\eta| < 1.5$. The drift chambers and scintillators are similar to those of the other muon systems.

Steel shielding reduces the rate of hadrons reaching the muon systems. Although hadronic particles must traverse approximately 5 interaction lengths of material before exiting the central hadronic calorimeter, a non-negligible fraction will succeed in doing so. The presence of steel absorbers increases the purity of signal in shielded muon systems. However, the dense material also causes muons to lose energy, increasing the minimum threshold of initial momentum required for a muon to reach the outer detectors. Table 2.3 lists the number of pion interaction lengths and the muon momentum threshold associated with each muon system. Additional steel also increases the rate of Coulomb scattering, in which muons are randomly deflected as they move through the electromagnetic potential of the absorber atoms. Multiple scattering causes particles to progressively diverge from the path which is extrapolated from the integrated tracking volume. The thicknesses of the shielding used for the CDF muon systems is chosen to achieve a favorable compromise between increased muon purity and detrimental effects.

	η coverage	Pion interaction lengths	Minimum muon p_T
CMU	$\lvert\eta\rvert < 0.6$	5.5	1.4 GeV/c
CMP	$\lvert\eta\rvert < 0.6$	7.8	2.2 GeV/c
CMX	$0.6 < \lvert\eta\rvert < 1.0$	6.2	1.4 GeV/c
IMU	$1.0 < \lvert\eta\rvert < 1.5$	6.2 - 20	1.4 - 2.0 GeV/c

Table 2.3: Pseudorapidity coverage, shielded thickness in pion interaction lengths, and minimum muon p_T for the muon subdetectors [63–65].

Muon candidates are identified by extrapolating the reconstructed tracks from the inner tracking volume to the muon chambers, where they are tested for matches with stubs. Muon stubs are reconstructed by fitting sets of hits to line segments using a least-square method. The hit positions in the the wire chambers of the muon systems are identified using the same methods as the more finely instrumented COT. A minimum of three muon hits is required to be fitted as a stub. COT tracks are required

to exceed the minimum p_T threshold for a given muon subdetector to be considered as a match for its stubs. The tracks are extrapolated using a a simplified geometry model of the particles' motion in the non-uniform magnetic field in the calorimeters. The matching procedure compares the position and direction of a candidate track and stub, incorporating the stub χ^2 and the covariance matrix of the extrapolated track. Other discrimination variables include the Δz separation and opening angle $\Delta\varphi$ between the track and stub. The matching requirements are chosen to maximize muon reconstruction efficiency while maintaining high muon purity.

Although the B_s signal used in this analysis is not directly reconstructed from muons, these particles remain relevant. The Opposite Side Muon Tagger, described in Section 6.2.1, is the most powerful opposite-side flavor tagging algorithm. Muons also contribute to B decays in semileptonic and J/ψ channels, whose datasets support this analysis.

2.3 Data Acquisition and Triggering

The RF structure of the Tevatron is configured such that proton and antiproton bunches cross every 396 ns at the center of the CDF detector. Given the total inelastic cross section of ~ 60 mb for $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV, an instantaneous luminosity of at least $\sim \mathcal{O}(40 \times 10^{30})$ cm^{-2}s^{-1} yields a minimum of one interaction per crossing, creating a 2.5 MHz collision rate. Recent data-taking has been performed at luminosities in excess of 200×10^{30} cm^{-2}s^{-1}, which may yield as many as 5 inelastic interactions per bunch crossing. Moreover, CDF is designed to operate with the Tevatron in a 132 ns bunch crossing mode, which would require the ability to handle a 7.5 MHz collision rate. Since full readout of the CDF detector creates about 250 KB of data, the continuous recording of every event in the current 396 ns mode would generate data at a rate of about 600 GB/s. This rate of information storage is prohibitively large, and the computing time needed to examine such immense datasets would render analysis impractical. However, the processes of greatest physical interest occur at much lower cross sections than the overall rate of $p\overline{p}$ interactions. Most pertinent to this analysis, the cross section for Tevatron production of $b\overline{b}$ pairs is $\sim 100\ \mu b$. This situation presents an opportunity for the reduction of the data rate by at least three orders of magnitude without loss of useful events. The framework for selection or rejection of specific types of events as they are read from the detector is known as a *trigger*. The CDF data acquisition (DAQ) system and triggers are designed to efficiently identify interesting events and record data from relevant channels of the detector. Such manipulations are said to occur at the "online" level, in contrast to the "offline" analysis of data which has already been stored.

Because of the high rate at which collisions occur, a primary concern for the DAQ and trigger system is that it be *deadtimeless*, in that there is a negligible fraction of time during which new collisions are occurring but the detector is unable to record them. The most simplistic way to achieve this condition would be to have electronics which are so fast that the detector readout and trigger decision are made in the interval between bunch crossings. However, limitations of modern computing do not

allow this option, requiring the use of a *buffered* system. While the properties of one event are being surveyed, the information from all subsequent events must be temporarily stored until they can be considered in turn.

The CDF trigger employs a tiered system in which events are examined in increasing detail as they pass through three trigger stages. Events are passed to the next trigger level only as long as they are accepted by the preceding one. As unwanted events are discarded, more time is available to perform progressively more complex and accurate tests on the accepted events. At the first level, new events are loaded into a buffer with a depth of 42 slots. For each Tevatron clock cycle, the events are moved up one slot, implying that an accept/reject decision must be reached within about 5 μs for each event. Only the most rudimentary pattern matching and filtering algorithms are applied within each subdetector. These simple cuts remove a large majority of the background, reducing the accepted event rate to about 30 kHz. The second trigger level loads events into a buffer with a depth of 4 events, creating about 20 μs for the accept/reject decision. This stage's algorithms make combinations of information from related components within the tracking and calorimetry systems, cutting the rate to about 300 Hz. For the third level, data read out from all of the subdetectors are combined by the *Event Builder* system for a more computationally intensive examination of complete events. Distribution of the events across a network of almost 1000 processors allows for approximately 1 second per decision. With application of the full trigger, the 2.5 MHz event rate is reduced to a level of approximately 80 Hz, resulting in roughly $\sim$20 MB/s of data being delivered to storage for offline analysis. In order to optimize for speed, the Level 1 and 2 triggers are implemented with custom-designed electronics and are intrinsically tied to the DAQ system, while Level 3 triggering is performed by a farm of consumer-grade PCs. Figure 2-9 depicts the conceptual flow of data through the trigger system.

Data-taking is divided into variable periods of time labelled as *runs*. In contrast to Tevatron "Run I" and "Run II" that denote multi-year ranges of accelerator and detector capability, the data runs are an every-day unit of data-taking during which the detector is known to function in some stable configuration. A single run typically spans several hours in length. The 1 fb^{-1} of data used for this analysis is comprised of several hundred such individual runs.

2.3.1 Data Acquisition and Flow

The DAQ system [66] is responsible for collecting data fragments from the front-end electronics associated with each detector component and transferring the appropriate events to permanent storage as they are accepted by the trigger.

The front-end and trigger electronics are packaged as VME modules distributed across 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 running the VxWorks operating system for hardware initialization and event readout. Each crate also contains a *controller* module which distributes, via the VME backplane, the timing synchronization signals it receives. The timing signals originate from the Tevatron clock and are in coherence

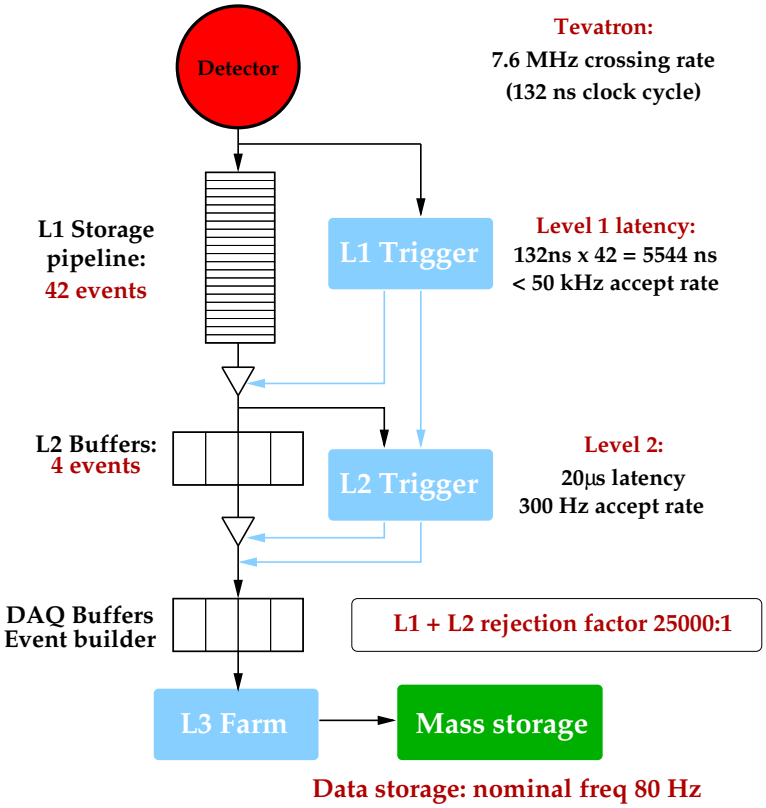

Figure 2-9: Functional block diagram of the CDF trigger system. The bunch crossing interval is 396 ns, although the Tevatron and CDF are designed to handle 132 ns.

with the $p\overline{p}$ bunch crossings to maintain global synchronization across the readout electronics and trigger. The controller module also provides the interface of the VME modules to the Trigger System Interface (TSI) and to the Event Builder (EVB). The TSI is responsible for receiving decisions from the hardware-based triggers (Level 1 and Level 2), communicating the decisions to the front-end crates, and supervising data flow until transfer to the EVB. The EVB and Level 3 systems [67] comprise a critical network of fast data transmission and control whose structure is outlined in Figure 2-10.

The data are first received in the EVB by the VME Readout Boards (VRBs). Each VRB is linked by fiber optics to a group of front-end crates. The VRBs themselves are distributed across 15 EVB crates, each of which is controlled by a single-board VxWorks processing unit known as a Scanner CPU (SCPU). The EVB crates transmit data to 16 *converter* node PCs of the Level 3 trigger farm through an Asynchronous Transfer Mode (ATM) network switch. This data flow is controlled by the Scanner Manager (SM), which is a task running in a processing unit seated in an additional EVB crate. The Scanner Manager also provides the interface of the EVB to the trigger decisions from the TSI. Communication among the SCPUs, converter nodes, and TSI is performed over a serial-ring reflective-memory control network called the SCRAMNet ("shared common random access memory network").

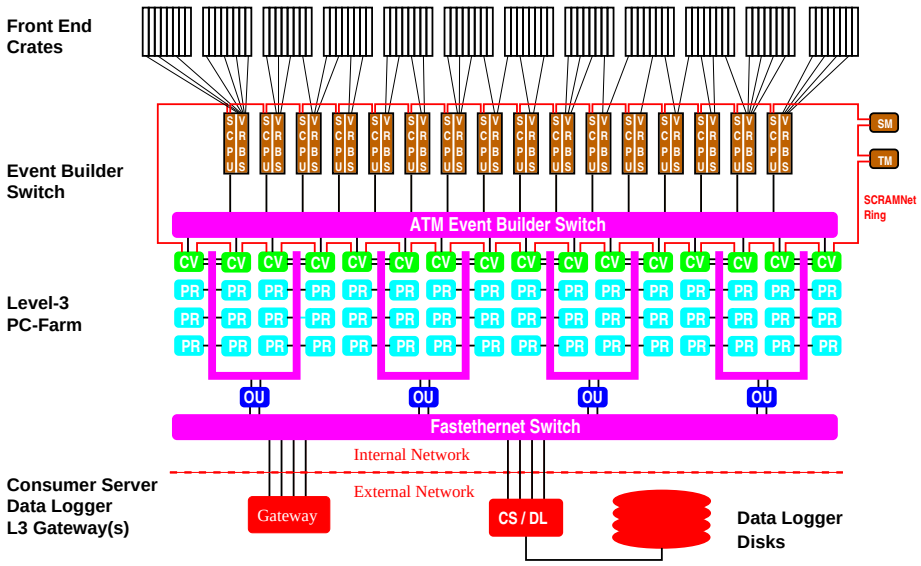

Figure 2-10: Block diagram of the DAQ system as it existed for most the analysis data. A recent upgrade replaced the SCRAMNet ring with Gigabit ethernet.

The Level 3 trigger farm is comprised of roughly 300 dual-processor PCs which run the Linux operating system and are networked via Ethernet. The farm is divided into 16 subfarms, each of which contains a converter node and 12 - 16 processor nodes. There are also 8 *output* nodes, each shared by two subfarms. This structure is loosely represented in Figure 2-10. The converter nodes function as the heads of their respective subfarms, distributing complete event records to the processor nodes. Once delivered to a processor node, events are passed through a trigger "filter" executable which performs high-quality reconstruction and applies a Level 3 decision. If the event is accepted by the trigger, it is passed to the associated output node. The output nodes funnel the accepted events into the Consumer Server Logger (CSL) system, where the data are first written to disk and later transferred to tape.

This description of the EVB and Level 3 systems reflects their configuration for the period during which most of the data for this analysis was collected. In an August 2005 upgrade to allow for larger throughput of the Level 3 trigger, the SCRAMNet ring and ATM switch responsible for data transfer were replaced by a Gigabit Ethernet network. As a scalable cluster of inexpensive PCs, the Level 3 farm has undergone numerous incremental upgrades in size and individual processor speed throughout Run II. The farm has recently been expanded from 16 to 18 subfarms, each of which contains 21 - 22 processor nodes. These upgrades in networking and processing capacity have increased the limit on Level 3 input rate from 300 Hz to 1 kHz, with the resulting output rate now exceeding 130 Hz. The EVB and trigger are also equipped to handle 500 KB of data per event, up from 250 KB. To account for the increased data rate, the CSL has been upgraded to store as much as 40 MB/s to disk.

Once in mass storage, the data are subjected to an offline stage of processing known as "production" to create the final versions of the most basic structures used

in analysis. These so-called physics objects, which include the final versions of tracks, leptons, and jets, are generated with more elaborate algorithms than those of the online trigger and use the most accurate and updated detector calibrations. Production is performed with another large farm of consumer-grade PCs [68].

2.3.2 Trigger Capabilities

The set of criteria necessary for an event to pass Levels 1, 2, and 3 constitutes a trigger path. The CDF experiment simultaneously uses $\sim \mathcal{O}(100)$ different trigger paths, each of which specifies requirements for the three stages. Accepted events must satisfy a well-defined trigger path, meeting a single path's criteria at each successive level. For offline storage of the data, groups of related trigger paths are combined into streams from which future analysis datasets are extracted.

This analysis relies primarily on data collected with triggers which use the unprecedented ability of the CDF electronics to rapidly identify and accept events containing displaced tracks. Based on the relatively long lifetimes of weakly decaying charm and bottom hadrons, these *displaced track triggers* are essential in compiling samples which are enriched in heavy-flavor events. The specific criteria for these trigger paths are described in Section 3.1, while the general routines and capabilities of the CDF trigger levels are outlined below.

Level 1

The Level 1 trigger is a synchronous system whose decisions are based on rudimentary information from the COT, calorimeters, and muon chambers. This input is used to form coarse versions of physics objects, including generic tracks, electrons, and muons, all of which are known as *trigger primitives*. The most important primitives for B physics analysis are tracks.

Track primitives in the transverse plane are identified by the Extremely Fast Tracker (XFT) [69]. Hit information from the four axial superlayers of the COT is used to make a rough measurement of the transverse momentum and azimuthal direction of tracks. The hits within each superlayer are compared to pre-loaded patterns, and the resulting track segments are matched against a list of template tracks which originate from the beamline. The XFT reports the p_T, φ position at superlayer 6, and charge of identified tracks. The system achieves track resolutions of $\sigma_{p_T}/p_T \sim 0.017 p_T/(\mathrm{GeV}/c)$ and $\sigma_{\varphi_6} = 5$ mrad, which are only a factor of about 10 less precise than the final offline reconstruction described in Section 3.2. The XFT efficiency of finding tracks is greater than 96% with respect to all tracks with $p_T > 1.5$ GeV/c in the final offline reconstruction, and the incidence of fake tracks is less than 3%. If more than 6 tracks are found, an automatic Level 1 accept decision is generated. For fewer tracks, the decision is based on specific trigger path requirements for p_T and φ_6, as noted in Section 3.1.

XFT tracks are extrapolated to the calorimeter and muon detector systems via the Extrapolator unit (XTRP) such that they can be matched to electromagnetic calorimeter towers and muon stubs. The calorimeter trigger is based on both the

identification of primitives such as electrons, photons, and jets, and on global event variables such as total transverse energy $\sum E_T$ and missing transverse energy $\not\!\!E_T$. The primitives are formed by applying signal thresholds for individual calorimeter towers. Transverse energy is estimated by summing the digitized calorimeter data into tower triggers, weighted by $\sin\theta$. Muon and di-muon primitives are derived from hits in the muon chambers or coincidences of hits with the scintillators.

Level 2

The Level 2 trigger is an asynchronous system which processes events accepted by Level 1 using better precision and more complete detector information. The most important of the Level 2 capabilities is the Silicon Vertex Trigger (SVT) [70], a powerful device which integrates the r-φ information of the SVX-II and COT subdetectors to provide precise measurements of track impact parameter d_0. The CDF detector is the pioneer of such measurements at the online level, which allow for triggering on the presence of events with displaced secondary decay vertices. This topology is characteristic of events containing bottom and charm quarks and is central to the SVT's ability to perform real-time collection of large samples enriched in heavy-flavor decays. As discussed further in Section 3.1, this attribute of the CDF trigger is central to the data samples for this analysis.

The configuration of the SVT electronics reflects the SVX detector's 12-wedge azimuthal symmetry and 3-barrel segmentation. SVT calculations are performed in parallel for each φ-sector and barrel. SVT track candidates require coincidence of an XFT track with hits in at least four of the five SVX layers. The XFT tracks are extrapolated into the SVX region, forming "roads" within which clusters of silicon charge deposition must be found. The SVT uses a list of $\sim \mathcal{O}(500)$ pre-loaded patterns for finding the best coincidences. The parameterizations of the XFT tracks and their associated SVX axial hit positions are passed to a linearized fitter which returns measurements of d_0, p_T, and φ_0. In addition to providing the critical measurement of impact parameter, the combination of SVX and XFT information improves on the initial Level 1 determination of p_T and φ_0. Information from the L00 and ISL silicon detectors is unavailable at the Level 2 stage and is neglected by the SVT.

The SVT impact parameter resolution for tracks with $p_T > 2.0$ GeV/c is approximately 35 μm, which is comparable to the quality of offline reconstruction without L00 hits. Figure 2-11 shows the d_0 distribution measured by the SVT. Since the sample is overwhelmingly composed of *prompt* tracks from the primary vertex, the distribution's fitted Gaussian width is representative of the overall d_0 resolution. The value of 47 μm includes the contribution of $\sim$ 30 μm from the transverse profile of the $p\overline{p}$ beam envelope.

Calorimeter trigger primitives are also refined at Level 2. Energy deposition information from Level 1 is used to form clusters of towers with appropriately varying signal levels [71]. The shower maximum chambers are used to further refine track matching and reduce fake electron and photon rates.

The time required for the Level 2 reconstruction of tracks or energy clusters in the separate systems is approximately 10 μs, after which the information is passed to

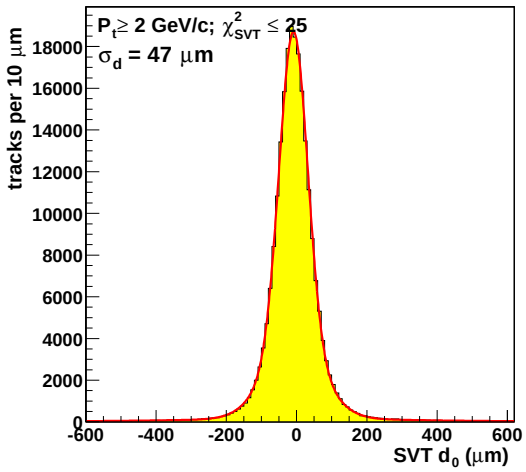

Figure 2-11: Distribution of track impact parameter measured by the SVT. The 47 μm width includes contributions from the 35 μm SVT d_0 resolution and the ~ 30 μm $p\overline{p}$ beam profile.

programmable Alpha processors for an overall accept or reject decision. The global decision requires roughly 10 μs, resulting in an average total SVT latency of about 19 μs.

Level 3

The Level 3 trigger takes advantage of its access to the complete event record and a longer per-event time allotment to make a more comprehensive trigger decision than those of the previous stages. Level 3 is organized as a farm of approximately 300 dual-processor PCs, which applies about one second of computing time to each incoming event. Reconstruction is performed at a level of quality near that of the offline data processing, making numerous variables available as criteria for the trigger paths. Values from the lower-level trigger systems may be used to drive the algorithms which subsequently refine them.

The Level 3 executable provides the earliest three-dimensional track reconstruction, allowing discrimination based on pseudorapidity and z position. SVT track parameterizations are matched to the full record of COT hits, combining the more precise aspects of each. Three-dimensional tracks are precisely matched to information from the calorimeter and muon systems, improving identification of leptons and jets. Tracks are also fitted to vertices for calculation of the $L_{\rm xy}$ and mass of decaying particles. Many Level 3 trigger bits incorporate explicit confirmation of lower-level trigger path criteria. The full set of criteria for the data used in analysis of B_s mixing is described in Section 3.1.

Chapter 3

Analysis Samples

Data is the foundation of any measurement. Equation (1.32) shows that the significance of the oscillation signal in frequency space depends on the relative fraction and absolute number of B_s signal events within the overall sample. Given that this analysis is part of an overall effort which is the first to cross a 5 standard deviation threshold of observation, the selection of the data sample is fundamentally important.

The data used for this analysis was recorded by the CDF detector during the period spanning February 2002 through January 2006. After application of requirements that all relevant detector systems were functioning properly, the B_s samples represent an integrated luminosity of 1 fb^{-1}. Because of the large statistics available for the B^0 and B^+ species, these development samples are primarily based on the first 355 pb^{-1} of the data.

This dissertation is devoted to the inclusion of partially reconstructed hadronic events in B_s mixing analysis. The presence of these decay channels in the samples is a natural consequence of the algorithms that are optimized to reconstruct and select fully reconstructed decays. Due to the inability of the detector to track neutral particles, or because the vertex reconstruction procedure neglects some charged particles, samples designed to identify fully reconstructed decays also contain B mesons whose reconstruction is incomplete. These components are a promising source of additional signal for flavor oscillations.

The new signal channels for this analysis are those reconstructed alongside the dominant fully reconstructed decay [1]

- $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$, $\phi \to K^+K^-$.

Within this sample, Chapter 4 shows that the partially reconstructed signal decays of greatest interest are

- $B_s^0 \to D_s^- \rho^+$, with $\rho^+ \to \pi^+\pi^0$,
- $B_s^0 \to D_s^{*-} \pi^+$, with $D_s^{*-} \to D_s^- \gamma$ (94.2%) or $D_s^{*-} \to D_s^- \pi^0$ (5.8%) ,

where the $D_s^- \to \phi\pi^-$ chain is common to all.

[1] Charge conjugate decays are always implied.

The analysis makes frequent use of the analogous partially reconstructed decays of the lighter B species as high-statistics development samples. Particularly with respect to B^0 events, both fully and partially reconstructed decays are very similar in topology and kinematics to the targeted B_s signals. Grouped by sample, with the fully reconstructed channel listed first, these decay channels are

- $B^0 \to D^- \pi^+$,
- $B^0 \to D^- \rho^+$, with $\rho^+ \to \pi^+ \pi^0$,
- $B^0 \to D^{*-} \pi^+$, with $D^{*-} \to D^- \pi^0$,

where the $D^- \to K^+ \pi^- \pi^-$decay is common to all, and

- $B^+ \to \overline{D}^0 \pi^+$,
- $B^+ \to \overline{D}^0 \rho^+$, with $\rho^+ \to \pi^+ \pi^0$,
- $B^+ \to \overline{D}^{*0} \pi^+$, with $\overline{D}^{*0} \to \overline{D}^0 \pi^0 / \gamma$,
- $B^0 \to D^{*-} \pi^+$, with $D^{*-} \to \overline{D}^0 \pi^-$,

where the $\overline{D}^0 \to K^+ \pi^-$decay is common to all. The presence of the $B^0 \to D^{*-} \pi^+$ component in the B^+ sample is a complication that ultimately limits its utility as a development sample. This issue is discussed in detail in later chapters.

It deserves to be reiterated that, despite the length of the preceding list of channels, the fact that the partially reconstructed decays are collected by algorithms designed to isolate *individual* fully reconstructed decay modes implies that all of the above channels are grouped within only three event samples. These are the samples whose reconstruction procedure is tuned to identify the decays

- $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi \pi^-$,
- $B^0 \to D^- \pi^+$, $D^- \to K^+ \pi^- \pi^-$,
- $B^+ \to \overline{D}^0 \pi^+$, $\overline{D}^0 \to K^+ \pi^-$,

where the D meson decay is the unifying attribute of each sample.

Although the above decay channels are the focus of this analysis, the resulting likelihood function is ultimately combined with the results of two other major B_s mixing analyses to produce a definitive signal. The two classes of signal used by these additional analyses are, respectively:

- five other fully reconstructed hadronic decay channels of the form $B_s \to D_s(3)\pi$, with D_s^- decays being reconstructed in three different modes; and
- semileptonic decays of the form $B_s \to D_s^{(*)} \ell \nu$, which are necessarily partially reconstructed but which have less statistical power per event than those of this analysis.

Unless otherwise noted, the procedures and techniques described in the following sections are also applied in the creation of those analysis datasets.

Furthermore, the ability to resolve B_s oscillations is the product of several complex tools which must be validated in supporting datasets. Much like the light B species are used for development of partially reconstructed signal models, various B^0 and B^+ samples are used for calibration of the measurement of proper time and flavor tagging. These samples include fully reconstructed decays of the type $B_{u,d} \to D^{(*)}_{u,d}(3)\pi$, $B_{u,d}$ decays containing J/ψ, and the special modified B topologies discussed in Section 5.1. The qualitative aspects of the procedures described below also apply to these samples.

This chapter describes the explicit trigger criteria, reconstruction methods, and final selection algorithms used for the data samples most central to this analysis. The selection of the other supporting datasets is broadly outlined, while detailed descriptions are available in the cited documentation.

3.1 Triggers

The function of the trigger system is to efficiently identify interesting events embedded within the very large rate of $p\overline{p}$ collisions as they occur in real-time. Section 2.3 introduces the capabilities of the three-tiered CDF trigger with emphasis on the Level 2 Silicon Vertex Trigger (SVT), a revolutionary tool for recognizing events with displaced tracks and compiling rich samples of B and D mesons. This analysis relies primarily on data collected with the associated "two-track trigger", which is detailed below. The analysis is also supported by related triggers, including the clean di-muon signature of $B \to J/\psi X$ decays and a hybrid trigger of lepton and displaced track requirements for semileptonic B decays. The discussion of this section makes frequent use of event variables whose definitions are compiled in Section 2.2.1.

3.1.1 The Two-Track Trigger

The relatively long lifetimes of weakly decaying heavy-flavor hadrons [2] implies that their decay vertices are displaced by a measurable amount from the primary $p\overline{p}$ interaction point. The tracks left by the charged daughter particles are therefore characterized by impact parameters which are inconsistent with zero. These displaced tracks may include the immediate decay particles of a B meson, such as a pion or a lepton, or they may arise from the daughters of subsequent D meson decays. This topology is in contrast to the appearance of more common interactions such as inelastic scattering, in which all tracks are "prompt", originating directly from the primary vertex. Figure 3-1 shows a sketch of the displaced tracks from a B meson decay.

Section 2.3.2 introduces the SVT system, which rapidly reconstructs tracks in the transverse plane of the silicon tracking layers. The precise silicon hit information and SVT electronics provide the trigger with impact parameter resolutions of approximately 35 μm for high-momentum tracks. This capability, which is the first of its kind among collider experiments and is unique at the Tevatron, allows for track displacement to be recognized at the online level and used in trigger decisions. Trig-

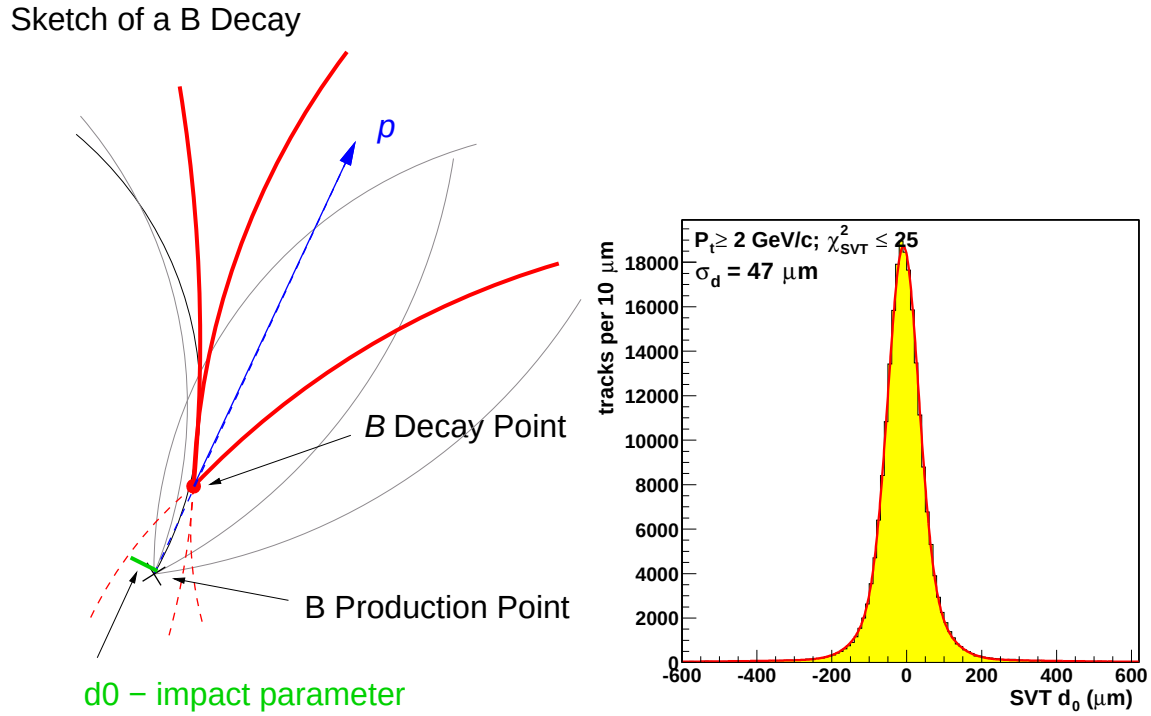

Figure 3-1: Left: Sketch of the displaced tracks from a secondary decay vertex. Right: Distribution of track impact parameter measured by the SVT, illustrating the 47 μm d_0 resolution.

gering on such a clear signature of long-lived events creates data samples which are enriched in events containing bottom and charm hadrons.

As the Tevatron luminosity decreases over the lifetime of a store, the rate of events which meet a given trigger's requirements decreases correspondingly. However, the CDF detector's ability to read out and store events remains constant. If one were to enforce very strict requirements on trigger paths such that they pass events at acceptable rates for high luminosities, the trigger bandwidth would become progressively under-utilized as the luminosity decreases. Conversely, loose trigger requirements would overwhelm the DAQ system during high luminosity running, such that high quality events could be missed as the system struggles to record less interesting events. A solution for efficient allocation of fixed trigger bandwidth is *prescaling*, in which the DAQ system may record only every N^{th} event that passes a trigger path. The prescale N can be varied dynamically to achieve a given acceptance rate, such that N may initially be set high ($\sim \mathcal{O}(5)$) and eventually decreased to unity later in the store.

The two-track trigger (TTT) encompasses a set of distinct trigger paths or "scenarios" [72] which aim to achieve efficient use of the trigger bandwidth at all Tevatron luminosities. The paths are applied simultaneously; that is, an event is tested against all scenarios rather than allowing only one as an option. A feature common to all scenarios is that the presence of at least two displaced tracks is required in order to make

a simple vertex in the transverse plane. In addition to impact parameter bounds, the paths make various requirements on the opening angle $\Delta\varphi$ and transverse momentum p_T of the displaced tracks and on the transverse displacement $L_{\rm xy}$ of the associated vertex. "Scenario A" establishes the baseline TTT path and is prescaled at high luminosity. The unprescaled "Scenario C" imposes stricter requirements and is designed to ensure acceptance of high quality events which might otherwise be discarded by the Scenario A prescale. The presence of an unprescaled trigger is also useful in that it simplifies potential cross-section measurements. "Scenario Low" is a more relaxed variation of Scenario A and is intended to use all available bandwidth by accepting a more inclusive sample at low luminosity. A final TTT variation was added during Run-II data-taking in response to the great value of cleanly flavor-tagged B_s events. The "Scenario Low plus Muon" path adds requirements for a muon stub in addition to the nominal Scenario Low configuration. Because muons originating from the decay of the opposite-hemisphere b quark are the basis for the most powerful opposite-side flavor tagging algorithm, TTT events containing a muon are not prescaled.

Table 3.1 compiles the specific requirements for these paths at each trigger level. Common notations include:

- $p_T^{\rm XFT}$ and $p_T^{\rm SVT}$: the transverse momenta as measured by either the XFT or SVT devices,
- $\sum p_T^{\rm XFT}$ and $\sum p_T^{\rm SVT}$: the scalar sum of the transverse momenta of the two displaced tracks,
- $\Delta\varphi_6^{\rm XFT}$: the separation in azimuthal angle φ between two tracks, as measured at COT superlayer 6 by the XFT system,
- $\Delta\varphi_0^{\rm SVT}$: the separation in azimuthal angle φ between two tracks, as extrapolated to the primary vertex position by the SVT system,
- Δz_0: the separation between two tracks along the beam axis as extrapolated to their points of closest approach.

3.1.2 Related Analysis Triggers

Di-muon Trigger

Data samples for calibrations in support of this analysis include the channels $B^0 \to J/\psi K^{*0}$ and $B^+ \to J/\psi K^+$ which are collected by the di-muon trigger. This class of trigger paths relies on the clean signature of the muons from $J/\psi \to \mu\mu$ decays. Because the primary inputs come from XFT tracking and the muon systems, both of which are available at Level 1, the Level 2 trigger does not play an important role for the di-muon trigger. Detailed trigger requirements are available in publications of B analyses devoted to J/ψ samples [73].

		Scenario			
	Requirement	A	C	Low	Low + Muon
L1	Two XFT tracks with:	opp. charge	opp. charge	any charge	any charge
	$p_T^{\rm XFT} > [{\rm GeV}/c^2]$	2.0	2.5	2.0	2.0
	$\sum p_T^{\rm XFT} > [{\rm GeV}/c^2]$	5.5	6.5	0	0
	$\Delta\varphi_6^{\rm XFT} \in$	[0,135]°	[0,120]°	[0,120]°	[0,90]°
	CMU or CMX muon matched to XFT trk	—	—	—	$\surd$
	$p_T^{\rm XFT}(\mu) > [{\rm GeV}/c^2]$	—	—	—	CMU:1.5/CMX:2.0
L2	XFT tracks matched to SVT tracks	$\surd$	$\surd$	$\surd$	$\surd$ (non-muon)
	$p_T^{\rm SVT} > [{\rm GeV}/c^2]$	2.0	2.5	2.0	2.0
	$\sum p_T^{\rm SVT} > [{\rm GeV}/c^2]$	5.5	6.5	0	0
	$\Delta\varphi_0^{\rm SVT} \in$	[2,90]°	[2,90]°	[2,90]°	[2,90]°
	$\lvert d_0^{\rm SVT}\rvert \in [\mu{\rm m}]$	[120,1000]	[120,1000]	[120,1000]	[120,1000]
	$L_{\rm xy}^{\rm SVT} > [\mu{\rm m}]$	200	200	200	200
	$\Delta\varphi_6^{\rm XFT}({\rm trk},\mu)$	—	—	—	[0,90]°
L3	SVT tracks matched to COT+silicon trks	$\surd$	$\surd$	$\surd$	$\surd$ (non-muon)
	$p_T > [{\rm GeV}/c^2]$	2.0	2.5	2.0	2.0
	$\sum p_T > [{\rm GeV}/c^2]$	5.5	6.5	0	0
	$\Delta\varphi_0 \in$	[2,90]°	[2,90]°	[2,90]°	[2,90]°
	$\lvert d_0\rvert \in [\mu{\rm m}]$	[80,1000]	[80,1000]	[80,1000]	[80,1000]
	$L_{\rm xy} > [\mu{\rm m}]$	200	200	200	200
	$\lvert\Delta z_0\rvert < [{\rm cm}]$	5.0	5.0	5.0	5.0
Prescaled?		$\surd$	—	$\surd$	—

Table 3.1: Specific requirements for the two-track trigger scenarios at each of the three trigger levels (L1/L2/L3). The notations are defined in the text.

ℓ+SVT Trigger

This analysis is ultimately combined with the results of an oscillation analysis of semileptonic $B \to D^{(*)}\ell\nu X$ decays. In addition to use of the two-track trigger described above, semileptonic decay modes are also collected with the ℓ+SVT trigger. This trigger is similar to the TTT except that it requires a lepton candidate from the EM calorimeter or muon systems in combination with only one displaced track from the SVT. Detailed trigger requirements are listed in publications devoted to the semileptonic analysis [35].

3.2 Reconstruction

The final reconstruction of particle decays is performed offline, once the events have been written to tape. The organizing principle is similar to that of the trigger, in that simple objects are constructed initially and then combined into structures of progressively greater complexity. However, this final procedure aims to use the most accurate available models of the detector and event attributes. The following discussion refers primarily to the central decay channels of this analysis, but all of the concepts apply equally to the supporting datasets.

3.2.1 Track Preparation

All particles are reconstructed on the basis of tracks. The precision of track measurements, specifically with respect to impact parameter, is directly responsible for the resulting resolution on vertex position in the detector and therefore on B proper decay time. Section 1.4.1 emphasizes the critical importance of proper time resolution to the ability to measure flavor oscillations. The most precise tracking methods at the CDF experiment have been developed for analysis of B_s mixing.

The initial offline production of tracks is performed in a generic fashion, where all particles are assumed to be pions. These tracks are then refitted with precision corrections and hypotheses of various particle species.

Pattern Recognition and Generic Production

Hits in the integrated tracking system of Section 2.2.2 are used to reconstruct the trajectories of charged particles. Tracks are fitted by several pattern recognition algorithms, the most important of which is the "Outside-In" method.

Reconstruction begins with COT information alone. The data for every COT sense wire contains hit times and integrated charge measurements of the ionization avalanches induced by passing particles. After correction for variations in the COT drift field and for the approximate time-of-flight from the $p\overline{p}$ interaction region, the timing information yields hit positions for tracks. Potential multiple interactions within each bunch crossing are assumed to occur simultaneously, and the resulting COT hit timing resolution is on the order of a few ns, roughly corresponding to this spread of collision times. The spatial resolution of COT hits is about 140 μm [53].

Since helical tracks in the detector form circles in the two-dimensional r-φ plane, pattern recognition begins with a search for circular track stubs in the COT. Sets of 4 or more hits are sought within each axial superlayer to form straight-line *segments*. Two approaches may be applied for finding tracks from a set of segments [74]. One method links together segments whose positions and φ_0 measurements lie tangent on a common circle, while the other method constrains the fit of segments in the outermost superlayer to originate from the beamline, followed by the addition of other axial hits that are consistent with this path. Regardless of the initial approach, once a circular path is found in the r-φ plane, segments and hits in the stereo superlayers are included based on their agreement with the circular fit. The resulting track is

a three-dimensional reconstruction of a particle's path with very precise transverse information and less precise longitudinal information. Typically, at least one of the algorithms succeeds in reconstructing a track from a real particle.

Once a track is reconstructed in the COT, it is extrapolated into the silicon layers [74]. Based on the estimated errors of the fitted track parameters, a three-dimensional "road" is formed around the extrapolated path. Beginning with the outermost ISL layer and working inward through the SVX, silicon clusters within the road are added to the track and the fit is repeated. As each cluster is added, the road is narrowed to reflect the updated track parameters and error matrix. Only r-φ silicon clusters are added in the first pass of this algorithm, while clusters with stereo information are added in a second pass. L00 hits are not added in this generic production stage.

The track reconstruction efficiency is about 95% for tracks which pass through all 8 COT superlayers with $p_T \geq 400$ MeV/c and about 99% for those with $p_T \geq 2$ GeV/c [56]. This efficiency is largely dependent on the number of tracks in the chamber at a given time. Numerous closely-spaced tracks can produce combinatorial ambiguity among hits, resulting in failed fits. The efficiency for successfully adding silicon information is about 93% in tracks with at least 3 r-φ silicon hits.

Track Refitting

Tracks for this analysis are selected from the standard set of production tracks and refitted to create the most accurate available parameterizations. In order to minimize low-quality reconstructions, accepted tracks are required to have at least 10 axial hits and 10 stereo hits in the COT and at least 3 r-φ hits in the silicon layers. Using the most accurate models of the detector material, this set of tracks is refitted [75] to account for energy loss in matter. Because specific ionization is dependent on particle mass, each track fit is assigned the hypothesis of a pion, kaon, *etc.* in turn, with the resulting track versions assigned to different "collections". Moreover, the raw uncertainties on the COT hit positions do not fully account for the effects of multiple scattering, whose importance increases for the longer path lengths to the outer layers. To correct for these underestimates as well as other resolution effects, the covariance matrix elements of each fitted track are inflated by empirically determined scale factors [76].

An especially important technical feature of this analysis is the use of L00 hits whenever available. These small-radius silicon hits are not included in the generic production of tracks and must be added during the refitting sequence. With typical spatial resolutions of about 6 μm and a radius of ~ 1.5 cm, L00 hits provide a superior lever arm for measurement of track impact parameter. Figure 3-2 shows the stark difference in impact parameter resolution for tracks fitted with and without a L00 hit. Within the COT and silicon hit requirements noted above, the efficiency for adding a L00 hit is approximately 65% [77], which results in an average reduction in proper time uncertainty of $\sim 10\%$. This modest improvement becomes essential in the context of the sensitivity of the analysis to B_s mixing. Equation (1.32) indicates that the oscillation significance is proportional to $1/\sigma_{\mathcal{A}} \propto \exp(-\frac{\Delta m^2 \sigma_t^2}{2})$, with a squared exponential loss due to poor time resolution. Issues of proper time resolution are

discussed further in Section 5.1.

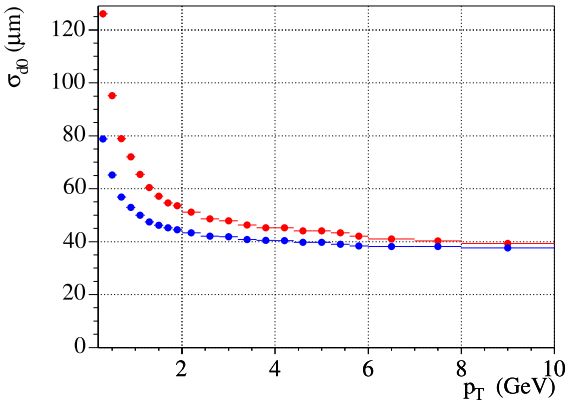

Figure 3-2: Comparison of track impact parameter resolution achieved with (blue) and without (red) a L00 hit. Tracks from B meson decays are most commonly in the p_T range below 2 GeV/c, where the greatest improvement is observed.

Tracks are observed for the five species of charged particle which are "stable" within the detector volume and typical momentum boosts: pions, kaons, protons, muons, and electrons. The large majority of reconstructed tracks are pions. Following the track refitting procedure, electrons and muons are matched to their respective signatures in the EM calorimeters and muon chambers for better identification. Additionally, time-of-flight information is available to help distinguish among the hadronic particles, as outlined in Appendix E.

As an additional step to select only well-measured tracks, particles are required to have a fitted transverse momentum of at least 350 MeV/c for consideration in this analysis. The resulting set of tracks typically ranges in number between 10 and 100 entries per $p\overline{p}$ collision.

3.2.2 Vertex Fitting and Preselection

Having prepared a set of tracks for every event, these traces of quasi-stable particles must be combined into various levels of heavy unstable parent particles. The main tool for this task is the CTVMFT vertex fitting package [78], which performs a three-dimensional fit of two or more track helices to a common origin. Requirements for acceptable vertices are imposed on the fit χ^2, spatial displacement, or kinematic variables such as momentum and mass.

The reconstruction of vertices begins with the most basic unstable particles which decay only to stable tracks. A few of the many possible examples include $\phi \to K^+K^-$, $\overline{D}^0 \to K^+\pi^-$, and $D^- \to K^+\pi^-\pi^-$. Tracks are combined from the collections of pions and kaons as appropriate, where care is taken not to combine duplicate tracks. The charge of the tracks is often considered, so as not to expend CPU time making

vertices with nonsensical charge correlations. After screening these candidate vertices with loose requirements, they are combined by the fitting package with an additional track to produce more massive unstable particles. The decays of $D_s^- \to \phi\pi^-$ and $B^0 \to D^-\pi^+$ are examples of this level. The fits may also include pointing constraints, such that the total vertex momentum is assumed to originate from the primary vertex, or mass constraints, such that constituent unstable particles are assumed to have a certain mass when fitted with another track. Additionally, a fit may be attempted between two vertices to create a mutual parent particle; in practice, this is how the $B_s^0 \to D_s^-\pi^+\pi^+\pi^-$ decay is implemented, where the three pions are first fitted as a hypothetical vertex. Such procedures may be repeated indefinitely to build decays of any complexity, with intermediate requirements on fit quality, position, and kinematics.

For illustrative purposes, the detailed reconstruction procedure for the $B_s^0 \to D_s^-\pi^+$, $D_s^- \to \phi\pi^-$decay chain is outlined here. CTVMFT attempts fits of all pairs of oppositely charged tracks from the kaon collection, and only vertices that meet the following conditions are accepted for the $\phi \to K^+K^-$ collection:

- $\chi^2_{xy} < 15$,
- $1.005 < M(K^+K^-) < 1.035$ GeV/c^2,
- $|\Delta z_0| < 1.5$ cm between the tracks.

Tracks from the pion collection are then fitted in turn to each ϕ candidate vertex, provided that they are not one of the two tracks with the kaon hypothesis. Only resultant vertices that meet the following conditions are accepted for the $D_s^- \to \phi\pi^-$ collection:

- $\chi^2_{xy} < 40$,
- $1.870 < M(K^+K^-\pi^\pm) < 2.070$ GeV/c^2.

Finally, the momentum vectors of $D_s^\pm$ candidates are combined with a pion track. The pion and $D_s^\pm$ candidates are required to have opposite charge and to represent four distinct tracks. The mass of the D_s vertex is constrained to the world average value [2] of M_{D_s}, because this mass is well established and any deviation for a real D meson is due only to detector resolution. In the case that the vertex does not represent a true D_s, the B_s candidate is likely to eliminated because the constraint drives up the fit χ^2. These vertices are retained for the $B_s^0 \to D_s^-\pi^+$, $D_s^- \to \phi\pi^-$ collection if they meet the following requirements:

- $\chi^2_{xy} < 25$,
- $4.4 < M(K^+K^-\pi^\pm\pi_B^\mp) < 6.6$ GeV/c^2.

The final pion is often referred to as the "bachelor pion" π_B because it is often separate from the other tracks in both momentum and space.

The tracks comprising each candidate are additionally subjected to *trigger confirmation*, which verifies that the SVT was triggered by two displaced tracks in this candidate and not some other part of the event. The final parameterization of each

track is compared to the online tracks fitted by the SVT, and a threshold of agreement is applied to the χ^2 of the curvature and φ_0 parameters in order to identify a match. The SVT requirements on impact parameter and p_T are verified for the final versions of matched tracks, and only B candidates which contain both tracks are accepted for analysis.

Partially Reconstructed Candidates

It is worth pausing here to revisit a central concept of this analysis. Because the CDF tracking system is not designed to detect neutral particles, only channels that decay completely to charged particles may be fully reconstructed. As one example, decays of $\phi \to \pi^+\pi^-\pi^0$ are *not* admitted to the $\phi \to K^+K^-$ collection because the neutral π^0 is likely to carry away a large fraction of the total ϕ energy, causing the reconstructed mass to fall outside the selected window. Additionally, the combined efficiency of charged particle tracking and vertex reconstruction is high, but not perfect. It is possible for a charged particle track to be missed or for a vertex fit to fail where a true decay occurred, causing a charged particle to be lost in the reconstruction. While the outcome of the ϕ example above may be obvious, what to expect is less apparent when only a small amount of the total phase space is available to a neutral particle or when the reconstructed mass window is wide as in the B_s case just illustrated. In fact, as introduced in Section 1.5.4, such decay modes *are* accepted by the reconstruction procedure and eventually enter into the analysis sample. A sample whose reconstruction is optimized to contain one fully reconstructed B decay channel also contains partially reconstructed decays, where the various modes share a single species of fully reconstructed D meson as a common attribute. Figure 3-3 shows a representative reconstruction topology in which a neutral particle is present. Subsequent chapters demonstrate that the $B_s^0 \to D_s^-\rho^+$ and $B_s^0 \to D_s^{*-}\pi^+$ modes are partially reconstructed configurations in which the lost momentum is small and narrowly distributed, allowing the associated signal components to be of significant use.

3.2.3 Primary Vertex Fitting

The measurement uncertainty on the distance between B production and decay is affected by the resolution on both the B decay vertex, whose reconstruction is described above, and the *primary vertex* of the $p\overline{p}$ collision. This decay length is sketched by the dotted line of Figure 3-3. Because resolution of this distance is of such critical importance to observation of a B_s oscillation signal, it is advantageous to minimize the uncertainty on both of the vertices whenever possible.

The position of the $p\overline{p}$ beam envelope is well known because it is stable over short periods of data-taking, during which time many billions of collisions occur. This *beamline* is modeled as an extended hourglass-shaped envelope whose central position in the x-y plane is a linear function of the nominal beam direction z. The beam's hourglass contour reflects its focusing by the Tevatron's quadrupole magnets in order to minimize the transverse profile at the collision point, thus maximizing instantaneous

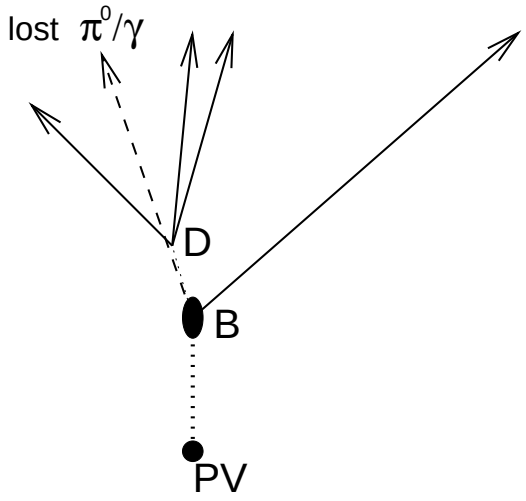

Figure 3-3: Representative topology of a partially reconstructed $B_s^0 \to D_s^{*-}\pi^+$ decay. The dashed line indicates a neutral particle which leaves no track.

luminosity. The PV of the collision may be taken as the central beamline position at the z_0 coordinate extrapolated from the B decay. However, the beam profile is approximately 35 μm in diameter at its narrowest point at $z \approx 0$ cm and ranges up to 50 - 60 μm at $|z| \approx 40$ cm, implying that a B meson is likely to be produced at any x-y position within that uncertainty. These values are comparable to the spatial uncertainty on the secondary B vertex and therefore contribute significantly to the overall decay length resolution.

Because the $p\overline{p}$ collision produces many tracks in addition to the decaying B meson, another viable method for measuring the PV position is to fit the common origin of most of the tracks in an event. This "event-by-event primary vertex" [79] stands as a second measurement of the PV position in addition to the beamline estimate. In fact, because both measurements are valuable, the most precise estimate of PV position comes from an average of the beamline measurement and the fitted vertex, weighted by the uncertainties from each. Figure 3-4 shows a representative sketch of this combination in the transverse plane. The PV fit is performed separately for each B candidate, where tracks that are used in the B reconstruction are explicitly excluded. In this sense, the fitted track origin is better described as a "candidate-by-candidate" PV, although the event-by-event name is most common.

The CTVMFT fit is seeded by the candidate's z_0 position and begins with all tracks within a fixed window of z and d_0/σ_{d_0} significance from the beamline estimate. The fit is performed iteratively, with the track of worst χ^2 agreement with the vertex being eliminated at each repetition until all tracks match the fitted vertex with a χ^2 below a given threshold. If too few tracks remain to make a good vertex or if the fit fails, then only the beamline estimate is used for the PV position. In the data samples used for this analysis, a beamline-constrained event-by-event primary vertex is available for approximately 90% of B candidates.

The procedure for PV fitting is validated as unbiased using test samples in which

the set of tracks is divided in two and the same PV is fitted from each sub-set. Comparison of the displacement between the vertices and their CTVMFT error estimates reveals that the fitted uncertainties are systematically underestimated by a factor 1.38. The uncertainties of primary vertices used in this analysis are corrected by this factor. The resulting beamline-constrained PV uncertainty in the transverse plane is typically ~ 20 μm.

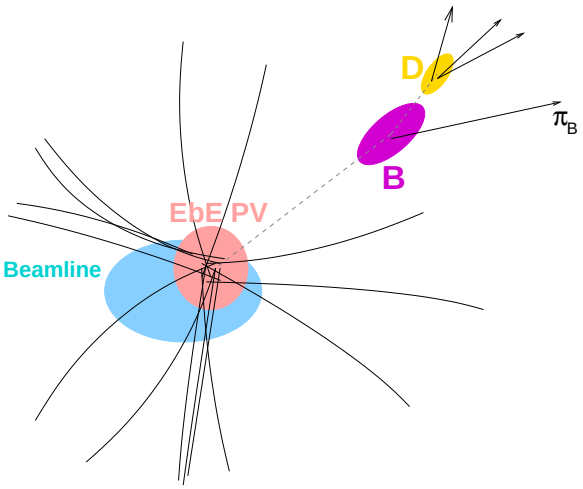

Figure 3-4: Sketch of the fitted event-by-event primary vertex within the beam profile. The transverse uncertainty of the weighted-average PV position is typically ~ 20 μm.

3.2.4 Compact Data Format

The volume of disk space devoted to storage of the CDF data is very large. Even the 1 fb^{-1} datasets resulting from the two-displaced-track trigger, which is only part of the CDF bandwidth, require 27.3 Terabytes of magnetic tape for the full event records. For more efficient access to the data samples on a frequent basis, it is useful to store candidates in a simple format containing only information which might be necessary for analysis. Moreover, the process of preparing tracks and fitting them into vertices is computationally intensive, requiring multiple weeks with the full processing power available to the CDF experiment. This analysis is made possible by performing the basic calculations only once and then storing the essential tracks and vertices in a "mother ntuple". Reconstructed candidates are stored in the BStntuple format [80], which is an extension of the Stntuple [81] structure for linking together all levels of parent and daughter candidates and their associated tracks. A full set of kinematic and spatial variables is accessible for each stable or unstable candidate, as is information about the vertex fits. An arbitrarily large number of decay modes may be combined in the same BStntuple file, such that entries for possible decays of B_s, B^0, or any other mesons which are reconstructed in the same collision event may share links to the same data blocks. The BStntuple format also contains links

to objects for flavor tagging, lepton identification, and particle identification tools, which are collectively discussed in Chapter 6 and Appendix E.

3.3 Candidate Selection

3.3.1 Decay Channel Skims

Despite the loose preselection requirements applied to reconstructed vertices and the relatively lightweight format of the BStntuple, the resulting samples are still too unwieldy to work with on an every-day basis. The vertex preselection is chosen to be conservative, so as to ensure that no potential signal events are discarded at that initial stage. However, the selection of candidates for the final analysis is much more stringent than these loose cuts. Additionally, the mother ntuples above contain a large number of decay channels and their associated reconstruction objects, most of which are not relevant to the decay modes of interest. In order to reduce the samples to yet more manageable sizes, they are "skimmed" by an intermediate selection step based on a few powerful variables. To streamline data access, the B candidate collections are also isolated into separate files for each decay channel. Identical skim cuts are applied to each of the three data samples of this analysis: the reconstructed collections of $B_s^0 \to D_s^- \pi^+$, $B^0 \to D^- \pi^+$, and $B^+ \to \overline{D}^0 \pi^+$, where each sample contains both the fully reconstructed modes and their partially reconstructed complements. The generic nature of the cuts is made possible by the topological and kinematic similarities common to the decay channels. Of particular note is the requirement on the displacement of the B candidate, which leverages the same principle as the two track trigger. The skim cuts are

- $\chi^2_{xy}(D) < 20$,
- $\chi^2_{xy}(B) < 20$,
- $L_{\rm xy}/\sigma_{L_{\rm xy}}(B) > 2.0$ and $L_{\rm xy}/\sigma_{L_{\rm xy}}(D) > 2.0$, requiring that the B and D mesons are at least 2σ displaced from the primary vertex, using the CTVMFT uncertainty,
- $d_0(B) < 200$ μm, removing probable background events in which the B candidate is measured to have originated away from the PV,
- $p_T(B) > 5.5$ GeV/c, removing the low-momentum phase space which is more heavily populated with random low-momentum track combinations.

A final preselection requirement is made using lepton identification information. The discrimination of electrons and muons is performed with a multivariate likelihood calibrated for each lepton type, based on quantities from the tracking system, calorimetry systems, and muon chambers [35]. These standards for identification of leptons are applied in selection of the complementary semileptonic analysis [35] sample. Because the results of this dissertation are eventually combined with the semileptonic analysis, it is important that the two partially reconstructed B_s datasets do not share any of the same candidates. Such an overlap could occur easily for $B_s \to D_s \ell \nu X$

events in which only a small fraction of the total B_s momentum is lost but where the lepton ℓ is hypothesized to be the bachelor pion. To ensure that the samples are mutually exclusive, the bachelor track is simply *anti-selected* with the aforementioned lepton requirements. In addition to maintaining independent datasets from the semileptonic analysis, this "lepton veto" also serves to strongly reduce the levels of semileptonic decays in the partially reconstructed mass region; in this analysis of only fully hadronic decays, any type of semileptonic decay is considered to be a background source. This selection is efficient at discriminating leptons from hadrons [35], thus maintaining all hadronic signals while removing the leptonic components.

Excluding the lepton veto, similar skim procedures are applied to other datasets which support or complement this analysis, including fully reconstructed hadronic channels of the form $B_{u,d,s} \to D_{u,d,s}(3)\pi$ and partially reconstructed semileptonic channels of the form $B_{u,d,s} \to D^{(*)}_{u,d,s}\ell\nu X$. Variations on the cuts listed above include more stringent displacement or p_T requirements for the hadronic modes with large track multiplicities or different vertex requirements for semileptonic modes. Channels whose decay chains include $D_s^- \to K^{*0}K^-$ or $D_s^- \to \pi^-\pi^-\pi^+$ also apply skim cuts on particle identification information and mass to remove the presence of $D^{(*)-}$ decays. This contamination is not present for this analysis because of the requirement of a ϕ resonance.

3.3.2 Neural Network Selection

Final selection of the main data samples for analysis is performed with Artificial Neural Networks (ANN, or simply "NN selection"). Along with the inclusion of the partially reconstructed hadronic decays described in this dissertation, the incorporation of ANNs into CDF mixing analyses represents one of the major upgrades that are designed to bring the early-2006 mixing evidence [31] from the 3.1 standard deviation level to a 5 standard deviation observation. Previous versions of the analysis employed a cut-based selection like that of the skimming procedure in order to identify the final samples. This selection method remains in use for the B^+ sample, whose correspondence to the B_s sample is less complete than that of the B^0 sample and so receives less attention. In addition to the NN-selected sample of B^0 candidates, the older cut-based B^0 sample is also occasionally used. This section describes both the NN selection of the B_s and B^0 samples and the simpler cut-based requirements for B^0 and B^+ candidates.

Selection Principles and Training

A common problem in physics analysis is that of distinguishing signal events from backgrounds, which are defined here as any B candidates that are not from the targeted decay modes. Equation (1.32) indicates that the significance of a flavor oscillation signal is proportional to $\mathcal{S}/\sqrt{\mathcal{S}+\mathcal{B}}$, where $\mathcal{S}$ and $\mathcal{B}$ denote the collective signal and background content of the analysis sample. The goal of the selection algorithm is to evaluate each B candidate and make an accept/reject decision so as to maximize the overall sample $\mathcal{S}/\sqrt{\mathcal{S}+\mathcal{B}}$. Neural networks provide a practical solution to

making a binary decision based on the information contained in numerous correlated event variables. Figure 3-5 depicts a simple sketch of a general NN structure. In this method, the list of event variables such as mass, p_T, L_{xy}, *etc.*, are conceptualized as *input nodes* whose values are weighted and combined in subsequent *hidden layers.* The resulting values from the hidden nodes are also weighted before being normalized by the *output node(s)* to produce a single discriminant whose value is between zero and one. The final selection decision for a given event is made by requiring that its NN output value be greater than some minimum threshold, where the threshold is chosen to optimize sample $\mathcal{S}/\sqrt{\mathcal{S}+\mathcal{B}}$.

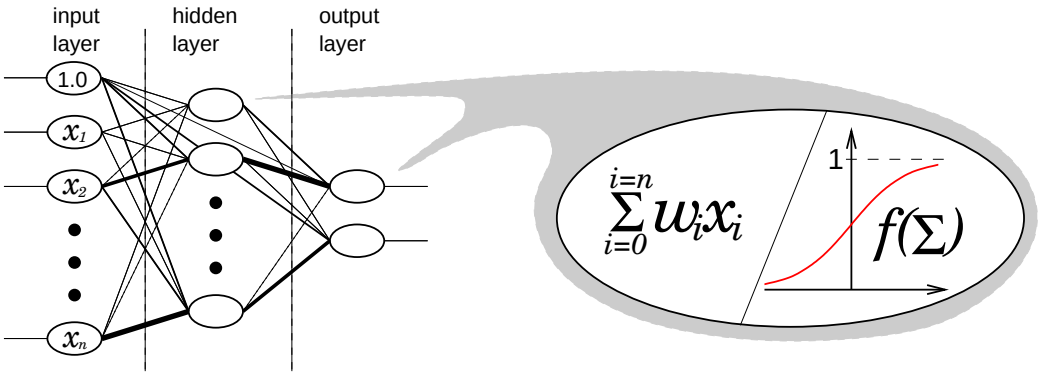

Figure 3-5: A sketch of a general neural network structure for selection of events. The lines represent weights applied to the event variables, which are normalized to produce a single output discriminant between 0 and 1. The neural networks in this analysis employ a single hidden layer and one output node.

It is worthwhile to note that selection of events based on hard boundaries to various quantities is simply a special case of the neural network framework described above. Whereas the general NN contains hidden nodes that apply a smooth weighting function to the input values, the special case may be achieved by using step functions for the variable weights. In this case, event variables which fall within the allowed bounds receive equal, non-zero weights in the selection decision ("yes"), while variables outside the bounds receive zero weight ("no"). The weights are combined multiplicatively to produce the final decision. Due to the sharp-edged step that defines the weighting functions in this case, such cut-based selection methods are sometimes referred to as *rectangular cuts.*

In contrast to the hard demands of a rectangular cut selection, the general neural network can make a collective judgment of an event's quality based on the values of all input quantities. For example, whereas a cut-based selection would discard a candidate which displays very signal-like properties in several variables but fails one cut by a small margin, the NN selection would produce a single output discriminant between zero and one that attempts to encapsulate this information with appropriate weights. This hypothetical candidate is likely to receive an output value closer to one and be above the threshold for event acceptance.

To reliably produce discriminant values, a network must be *trained* with samples

whose content is well understood. Using the ROOT-SNNS simulation package [33, 82], a separate neural network is trained for each of the data samples to which the selection is to be applied. Just as in the reconstruction methods, the selection is directed toward the fully reconstructed decay channels, while the kinematically similar partially reconstructed modes appear in parallel as a natural by-product. The training uses Monte Carlo samples to simulate signal, while data events from the upper mass sidebands provide the background behavior. The use of signal MC prevents contamination from various types of background that would be unavoidable in any sample of data events. The MC samples are of the type introduced in Section 3.4. To validate their use in modeling characteristics of the signal, the MC distributions of the variables on which the selection is based are carefully compared for agreement with those of sideband-subtracted signal data. The signal training is performed on MC events within about 3 standard deviations width of the fully reconstructed B mass for a given species, based on the CDF mass resolution. Table 3.2 lists the training ranges for the B_s and B^0 networks. The use of data events to represent the background relies on the assumption that the combinatorial background events (Chapter 4) of the sideband are very similar to those directly within the signal range. The sideband mass range is chosen to be close enough to the signal for this assumption to be valid while isolated enough to contain only background events.

Channel	Signal [GeV/c^2]	Background [GeV/c^2]
$B_s^0 \to D_s^- \pi^+$	5.31 - 5.42	5.5 - 5.8
$B^0 \to D^- \pi^+$	5.22 - 5.33	5.5 - 5.8

Table 3.2: Mass ranges of the MC and sideband data events used for training of the neural networks. The signal selection is directed toward these fully reconstructed channels, and the presence of kinematically similar partially reconstructed modes is a natural consequence.

In practice, the training consists of minimization of the distance between a vector of target decisions $\vec{t}$ (for example, the set of values $\{1\}$ for a sample of MC events which are known to represent signal) and the vector $\vec{o}$ of actual NN output values. This distance may be expressed as

$$\Delta = \sqrt{\sum_{i=1}^{N} \frac{(t_i - o_i)^2}{N}}, \tag{3.1}$$

where N is the total number of signal and background events in the training samples. The training procedure is iterative, with the internal weights being adjusted from their initially random values to produce progressively smaller values of Δ. To reduce the possibility of the network being trained toward statistical fluctuations within the samples, independent test samples are also used for evaluation of Δ.

Once a network is trained, it is applied to an independent test sample to find the output discriminant threshold that yields a global maximum in sample $\mathcal{S}/\sqrt{\mathcal{S}+\mathcal{B}}$.

The test samples use signal MC from the ranges of Table 3.2 for evaluation of $\mathcal{S}$, while the corresponding background levels are extrapolated from the data sidebands.

NN Variables

Table 3.3 lists the candidate variables used for the NN selection of the $B_s^0 \to D_s^- \pi^+$ and $B^0 \to D^- \pi^+$ modes and their optimized discriminant thresholds. The set of quantities includes the basic ones on which rectangular cut selections are typically based, but it also incorporates many other correlated variables which would be too complex to handle in selection frameworks other than a neural network. The B and D candidates are each represented by transverse-plane fit quality χ^2_{xy}, impact parameter d_0, transverse momentum p_T, and transverse decay length from the primary vertex and associated significance, $L_{\rm xy}$ and $L_{\rm xy}/\sigma_{L_{\rm xy}}$. The transverse decay length of the D candidate with respect to the B vertex $L_{\rm xy}(D \to B)$ is also used. Although the D mass is constrained to the world-average value in the B reconstruction, its unconstrained mass value is supplied to the network. The mass and p_T of the fitted ϕ vertex are used for the B_s selection. Decay angles are also useful for weighting signal and background, particularly in decays to a vector resonance such as $D_s^- \to \phi\pi^-$. Figure 3-6 illustrates the definitions of angles for the ϕ system. For a general two-body decay, θ^* represents the lesser angle in the decay's center-of-mass frame between one of its daughter products and the decay's momentum direction in the laboratory frame. The B_s selection incorporates the $\cos\theta^*$ values measured for the decays of the D_s and ϕ candidates. The networks for both B_s and B^0 also use the angle $\Theta(B,\pi_B)$ between the bachelor pion momentum in the decaying B rest frame and the direction of the B momentum. The networks for both modes use the p_T of each track, including the maximum and minimum values. The extremum values of impact parameter significance and the sum over all tracks are also included, as are similar track values for the combined particle identification likelihood CLL defined in Equation (E.4). The B_s selection incorporates the maximum separation of tracks along the nominal beamline and the scalar difference between the transverse momenta of the kaons from the ϕ resonance. The B^0 selection includes the Dalitz masses of the $K^+\pi_1^-$ and $K^+\pi_2^-$ daughter combinations of the $D^- \to K^+\pi_1^-\pi_2^-$ candidate.

As a final cut applied after the NN selection, B candidates are required to have proper time resolution $\sigma_{ct} < 200$ μm. At this stage, the rejected events are outliers on the upper tail of the σ_{ct} distribution and represent much less than 1% of either sample. These events may be discarded because, as indicated by Equation (1.32), their poor time resolution causes their contribution to the sample significance to be virtually negligible. Moreover, very large σ_{ct} values can cause the analytical functions of the likelihood to become numerically unstable, further decreasing their practical value.

Figure 3-7 overlays the mass spectra of the samples selected by the neural network and the older rectangular cuts for the $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$ decay chain. The samples are shown as both inclusive spectra for each algorithm as well as *exclusive* spectra; that is, the distributions of events chosen by one selection method and not the other are made clear. The increased inclusive signal size of the NN sample is

Variable	$B_s^0 \to D_s^- \pi^+$	$B^0 \to D^- \pi^+$
$\chi^2_{xy}(B)$	$\surd$	$\surd$
$d_0(B)$	$\surd$	$\surd$
$L_{\rm xy}(B)$	$\surd$	$\surd$
$L_{\rm xy}/\sigma_{L_{\rm xy}}(B)$	$\surd$	$\surd$
$p_T(B)$	$\surd$	$\surd$
$L_{\rm xy}(D \to B)$	$\surd$	$\surd$
$\chi^2_{xy}(D)$	$\surd$	–
$d_0(D)$	$\surd$	$\surd$
$L_{\rm xy}(D)$	$\surd$	$\surd$
$L_{\rm xy}/\sigma_{L_{\rm xy}}(D)$	$\surd$	$\surd$
$M(D)$	$\surd$	$\surd$
$p_T(D)$	$\surd$	$\surd$
$\cos\Theta(B,\pi_B)$	$\surd$	$\surd$
$\cos\theta^*(D_s)$	$\surd$	–
$\cos\theta^*(\phi)$	$\surd$	–
$M(\phi)$	$\surd$	–
$p_T(\phi)$	$\surd$	–
p_T(tracks)	all	all
$\min(p_T)$	$\surd$	$\surd$
$\max(p_T)$	$\surd$	$\surd$
$\min(d_0/\sigma_{d_0})$	$\surd$	$\surd$
$\max(d_0/\sigma_{d_0})$	$\surd$	$\surd$
$\sum(d_0/\sigma_{d_0}(\text{tracks}))$	$\surd$	$\surd$
$\max(CLL)$	$\surd$	$\surd$
$\min(CLL)$	$\surd$	$\surd$
$\sum(CLL(\text{tracks}))$	$\surd$	$\surd$
$\max\lvert\Delta z_0\rvert$	$\surd$	–
$\lvert p_T(K^+) - p_T(K^-)\rvert$	$\surd$	$\surd$
$M_{D^- \to K^+\pi_1^-\pi_2^-}(K^+\pi_1^-)$	–	$\surd$
$M_{D^- \to K^+\pi_1^-\pi_2^-}(K^+\pi_2^-)$	–	$\surd$
NN cut	0.954	0.828

Table 3.3: Candidate variables used in the training of the neural networks for B_s and B^0 selection.

obvious. The candidates selected by the NN but not the rectangular cuts emphasize the fully reconstructed peak, but some partially reconstructed content is also added. This result is not surprising, given the kinematic similarity of these closely positioned structures. It is also reassuring to note that the candidates that would be accepted by the rectangular cuts but are rejected by the new NN form a smooth and continuous background-like spectrum, implying that no signal events are rejected.

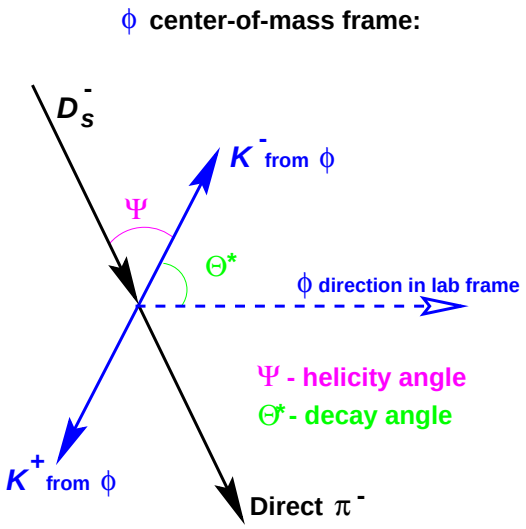

Figure 3-6: Illustration of angles in a two-body decay using the example of ϕ in $D_s^- \to \phi\pi^-$.

Rectangular Cut Selection

As noted above, the neural network selection represents an upgrade of the analysis from previous versions. While the network is the only algorithm applied for the B_s sample, the cuts for the older B^0 sample are listed here because that sample version is occasionally used for investigation of possible effects from the NN selection. Additionally, the selection of the B^+ sample is performed exclusively with rectangular cuts.

The optimization of a set of rectangular cuts is a less precise generalization of the process of choosing the selection threshold for a trained neural network. Ideally, the entire space of possible cut combinations is explored to select the cut set which gives a global maximum in sample $\mathcal{S}/\sqrt{\mathcal{S}+\mathcal{B}}$. Whereas one only has to scan a single dimension for the NN output, the cut optimization procedure becomes lengthy for a long list of cuts. As such, the allowed variable ranges and increments of change must be chosen with care.

Table 3.4 compiles the rectangular cuts optimized for the B^+ and B^0 samples. The variables are a subset of those introduced for the NN selection except for $\Delta R(D, \pi_B)$, the simple opening angle in the laboratory frame between the D momentum and the bachelor pion.

Supporting Sample Selection

This analysis is ultimately combined with two other major CDF B_s mixing analyses. Furthermore, it relies on calibrations performed with several supporting light B samples in addition to the B^0 and B^+ modes described above. While the details of each sample are unique, all general aspects of prior discussions apply equally to their

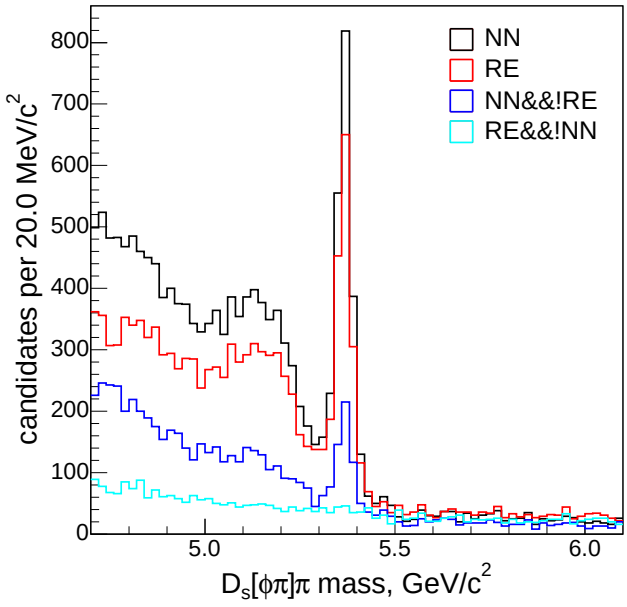

Figure 3-7: Comparison of samples selected for the $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$ chain by the neural network and by the older rectangular cuts.

Variable		$B^+ \to \overline{D}^0 \pi^+$	$B^0 \to D^- \pi^+$
$\chi^2_{xy}(B)$	$<$	15	15
$\chi^2_{xy}(D)$	$<$	15	15
$L_{\rm xy}/\sigma_{L_{\rm xy}}(B)$ [μm]	$>$	8	11
$L_{\rm xy}(D \to B)$ [μm]	$>$	-150	-300
$\lvert d_0(B)\rvert$ [μm]	$<$	80	110
$p_T(B)$ [GeV/c]	$>$	5.5	5.5
$p_T(\pi_B)$ [GeV/c]	$>$	1.0	1.2
$\Delta R(D, \pi_B)$	$<$	2.0	1.5

Table 3.4: Rectangular cuts for B^+ and B^0 candidates. These requirements are applied to the B^+ sample in this analysis, while the main neural network-selected B^0 sample is occasionally compared to this rectangular B^0 selection. The B_s sample is selected exclusively with a NN using the quantities of Table 3.3.

reconstruction and selection.

In the mixing analysis of other fully reconstructed hadronic B_s channels, neural networks with variables similar to those of this $B_s^0 \to D_s^- \pi^+$ mode are used for final candidate selection, resulting in comparable or greater improvement over rectangular cuts. These neural networks are documented in Reference [83]. The samples of semileptonic B_s decays are selected with rectangular cuts documented in Reference [35]. Although NN selections are explored for these modes, the potential for improvement in $\mathcal{S}/\sqrt{\mathcal{S}+\mathcal{B}}$ has been found to be more modest than in the fully

hadronic modes.

Calibration of the opposite-side flavor tagger is performed with a combination of light B decays to various fully reconstructed hadronic and J/ψ modes, as is discussed further in Chapter 6. For this study, the B candidate samples are selected by rectangular cuts documented in Reference [84]. The cuts for fully hadronic decays are similar to those listed in Table 3.4, while those of the J/ψ modes are similar except for a lack of requirements on displacement.

3.4 Monte Carlo Simulation

The reconstructed B candidate properties that enter the likelihood fit are the final result of a complex chain of physical processes and logical decisions. These discrete steps include the $p\overline{p}$ collision, the hadronization and decay of short-lived heavy flavor mesons, and the interaction of more stable daughter particles with the detector and trigger systems. Predictions for the distribution of certain B quantities are critical for modeling the expected behavior of the data. However, the cumulative result of the processes underlying the data are far too complex to be accurately predicted from first principles or by any analytical method. A viable solution to this complexity is a numerical simulation of the sequential, individual steps in the evolution of events. A representative ensemble of such simulated events constitutes a *Monte Carlo sample.*

3.4.1 Applications

Monte Carlo simulation plays an important role in almost every aspect of this analysis. The most general application of a MC sample is to model the expected behavior of a certain class of event in some measurement space, such as mass or proper decay time. This approach may be applied equally well to signal or background events that originate from specific mixtures of B decays. Models derived from MC simulation are combined in the overall likelihood with models motivated from first principles, where the end goal is an unbiased extraction of physical parameters from the signal components of the data.

The modeling of partially reconstructed hadronic decays in this analysis uses MC simulation in four major applications which are summarized below, while full detail is described in later chapters. This list of techniques generally employs ensembles of MC events which are uniquely generated for each decay channel, where a given channel's dedicated sample is the basis for all of its MC modeling. Section 3.4.2 discusses the tools used to generate the samples. As is the standard paradigm in this analysis, these techniques are tested and refined in the high-statistics development samples of B^0 and B^+, with emphasis on the former, before being applied to the B_s data.

- *Mass spectra* of partially reconstructed decay channels: Whereas the mass spectra of fully reconstructed channels are influenced primarily by the resolution of the detector, the mass distributions of decays on which this analysis is based are smeared by the loss of unreconstructed momentum. The final mass spectra are a complex result of the decay kinematics and their effect on the triggering,

reconstruction, and selection algorithms. Chapter 4 details the derivation and incorporation of these models into the likelihood.

- *Reconstruction efficiencies*: The relative efficiency with which both fully and partially reconstructed decays enter the selected samples is another result of the origins listed for the mass spectra above. While relative fractions of the signal decays may be fitted using only these mass spectra, Chapter 4 discusses how the likelihood is constrained by the expected values derived from MC simulation.

- *Bias in proper decay time distributions*: Both the two-track trigger and selection requirements rely on the long-lived nature of B mesons to identify a clean data sample. By definition, this approach preferentially selects candidates with longer decay times, such that the resulting proper time distributions do not follow the simple exponential behavior that would otherwise be expected for an unstable state. The "sculpting" of each channel's distribution depends on its kinematics and individual selection algorithm. To account for this bias, the proper time distributions of the pure MC samples are used to extract a proper time-dependent efficiency which parameterizes the relative probability of candidates with different decay times to enter the analysis sample. Chapter 5 describes the derivation of this efficiency term and how it modifies the decay time components in the likelihood.

- *Reconstructed momentum fractions*: Measurement of the proper decay time for each B candidate requires knowledge of its total transverse momentum. While this information is readily available for fully reconstructed decays , the missing particles of partially reconstructed decays carry away unmeasured momentum, thereby inducing uncertainty on the total candidate p_T. This uncertainty propagates into uncertainty on proper decay time measurement, whose precision is of critical importance for this analysis. The momentum uncertainty may be quantified and accounted for by modeling the fraction of the total reconstructed B p_T for each partially reconstructed signal component. The variation of fractional momenta observed over a representative sample of MC events produces the so-called "κ-factor" distribution. Chapter 5 details the extraction of these fractional reconstructed p_T distributions and their convolution into the decay time components of the likelihood.

In addition to these four major applications, Monte Carlo simulation is also used for initial studies of sample composition to identify the partially reconstructed decay channels of greatest interest. Whereas the above techniques are best performed with dedicated MC samples for each decay channel, this preliminary study requires a generic simulation with full representation of B decay modes. Chapter 4 describes these studies, while Section 3.4.2 again discusses the tools and assumptions involved in generation of the generic samples.

Finally, MC samples play a fundamental role in the development of b-flavor tagging tools. Whereas simulation of only the portion of an event that results directly from the B decay is necessary for the above applications, tagging methods require

examination of the entire collision event. Generation of these more complete MC samples is described below, and their application to tagger development is outlined in Chapter 6.

3.4.2 Simulation

The procedure for generation of Monte Carlo is designed to follow the realistic evolution of $p\overline{p}$ collision events, within the bounds of detail necessary for analysis.

Hard Scattering and Hadronization

Event simulation begins with the hard $p\overline{p}$ interaction which defines the primary vertex. In a fully realistic collision, energetic quark-antiquark pairs and/or gluons are emitted in all directions and move apart until the potential of their strong interaction grows large enough to create new $q\overline{q}$ pairs. This process of fragmentation and hadronization results in the flow of numerous pions, kaons, and heavier particles from the primary vertex. In events in which a $b\overline{b}$ pair is created in the initial collision, hadronization is likely to create B mesons or baryons in opposite hemispheres, each in close proximity to many lighter particles.

Simulation of these processes is performed by one of two packages. PYTHIA [85] is a generator which includes all details of the collision and hadronization, resulting in a full description of the event. By contrast, the BGENERATOR [86] package generates only a single b hadron per event, neglecting the underlying details and associated particles. Because the MC samples most central to this analysis are used to model single B decays, the BGENERATOR approach is almost always adequate. This simplification is desirable because the complexity of PYTHIA events requires approximately an order of magnitude more computing time and storage space than what is necessary for BGENERATOR. The critical application which demands complete events from PYTHIA simulation is that of the Same Side Kaon Tagger described in Chapter 6. Both PYTHIA and BGENERATOR contain routines for simulating the decay of heavy flavor hadrons, but their methods are less accurate than other specialized packages. Regardless of the hard-scatter generator used, the eventual decay of B mesons is assigned to another package described in a following section.

BGENERATOR begins event simulation with a single b hadron whose kinematic properties are representative of those observed in data. The package can create samples comprised of exclusive species such as B^+, B^0, B_s, or Λ_b, or it can produce admixtures in any ratios. The kinematic tuning consists of assigning a momentum with respect to the primary vertex, where each B momentum is based on an η-p_T distribution measured in inclusive CDF B data [87]. The single-channel simulations used for the four major Monte Carlo applications are initiated with BGENERATOR. For computational efficiency, the package creates only B mesons of the species of interest for each dedicated sample. For example, BGENERATOR creates only B_s mesons to initiate events of the MC sample used to derive the behavior of $B_s^0 \to D_s^- \rho^+$ decays, while later steps of the simulation sequence ensure that the B_s decays via the appropriate channel.

The exception to this most simplistic case is the generation of "generic" MC simulation for sample composition studies. In this case, BGENERATOR creates b hadrons in the relative fragmentation fractions from the world average measurements [2]. The hadrons $B^+ : B^0 : B_s : \Lambda_b$ are created with probabilities $0.397 : 0.397 : 0.107 : 0.099$, where the contributions of the B_c meson and heavier b baryons are negligible. Each species is also allowed to decay in a generalized manner later in the simulation sequence.

Whereas BGENERATOR neglects the underlying physics of the $p\overline{p}$ interaction and produces B momenta from an empirical model, PYTHIA attempts to simulate the full complexity of the collision. The package uses the *string fragmentation* model to simulate the creation of quarks and their hadronization. Events include $b\overline{b}$ pairs created via the leading- and next-to-leading-order mechanisms of flavor creation, flavor excitation and gluon splitting [88]. 20% of all B^+ and B^0 mesons originate from B^{**} decays. PYTHIA is tuned to match the distributions observed in data for the leading charged jets, including their angular orientation, momentum, and multiplicity of tracks. Reference [89], which focuses on the development of the Same Side Kaon Tagger and therefore relies heavily on this MC generator, contains a more detailed discussion of PYTHIA capabilities.

Heavy Flavor Decay

Regardless of whether BGENERATOR or PYTHIA is used to generate a b hadron, its subsequent decay chain is simulated with the EvtGen package [90]. This specialized framework for modeling heavy flavor decays is based on quantum mechanical *amplitudes*, as opposed to simple probabilities, such that interferences and phases are correctly modeled to produce accurate decay distributions. The package has been extensively tuned by the BABAR [15] and Belle [16] experiments. EvtGen's decay settings are fully customizable, which allows for generation of the dedicated samples for each decay mode. For example, the EvtGen configuration for the $B_s^0 \to D_s^- \rho^+$ sample is modified to force the B_s meson from BGENERATOR to always decay via that mode, with the unstable D_s^-, ϕ, and ρ^+ daughter particles also decaying only via the reconstructed chain. This procedure follows from the goal that the MC samples demonstrate how a sample of ideal decays appears after passing through the detector and requirements for selection.

For generic samples, EvtGen allows the b hadrons to decay according to the full range of world average branching ratios [2]. The sole deviation from the generality of these samples is that events which do not contain a D meson of the appropriate species are filtered out. The filtered simulation therefore matches the selection of the data. In the example of the $B_s^0 \to D_s^- \pi^+$ data, the selection for the D_s^- and its ϕ resonance is such that D^- and D^0 mesons are not present. As noted in the beginning of this chapter, the common attribute of both fully and partially reconstructed decays in the B_s, B^0, and B^+ samples are the exclusive presence of D_s^-, D^-, and D^0 mesons, respectively. Because of this filtering on the D meson, these samples are more commonly referred to as "semi-generic" Monte Carlo simulation.

Detector Simulation

The product of the packages for hard scattering and decay is a collection of four-vectors. This abstract information represents the remnants of states that are short-lived enough to decay within the vacuum of the beam pipe, a group which includes all B and D mesons of interest for this analysis. However, many longer-lived particles are likely to decay within the volume of the detector, either spontaneously or through direct interaction with matter, and these decays must be simulated in the context of the detector material. More generally, the Monte Carlo framework must simulate the interaction of all outgoing particles with the detector and the response of its electronics. This complex functionality is performed by the GEometry ANd Tracking (GEANT [91]) package. GEANT takes responsibility for the decay of relatively long-lived states such as K^0 and Λ and for the interactions of charged stable and "quasi-stable" (pions, kaons, muons) particles with the tracking system and calorimeters. The detector response is modeled in terms of hits and energy deposition in the subdetectors, and the resulting data banks are identical to those produced in real data-taking.

Just as the data is segmented into "runs", each typically spanning several hours of stable data-taking, MC samples are also subdivided into units of integrated luminosity for each run in the analysis data. Run-specific conditions as detailed as the electrical status and physical alignment of each ladder in the silicon layers are included in the GEANT model.

Triggering and Event Selection

To complete the simulation of events, MC data banks are treated with the same trigger logic and reconstruction methods applied to real data. Events that fail the two-track trigger requirements of Section 3.1 are discarded. The run-dependent simulation extends to the trigger configuration, including the prescales for each trigger path. Accurate prescale settings in the trigger logic assists the simulation in achieving relative acceptances of various decay channels which are representative of the data. In fact, the major MC samples in this analysis focus on modeling the first 355 pb^{-1} of data, and systematic uncertainties are evaluated to estimate possible biases.

The simulated events are subjected to the standard routines of production, track refitting, and candidate reconstruction. One complication from the use of BGENERATOR is that the full set of tracks from $p\overline{p}$ fragmentation is not available for reconstruction of an event-by-event primary vertex. This limitation is circumvented by modifying the run-dependent beamline of Monte Carlo events to reflect the distribution of primary vertex properties observed in the data. Representative PVs are achieved by smearing the true $p\overline{p}$ interaction point of the simulation and assigning a vertex resolution consistent with the improvement observed in data.

Finally, the selection algorithms of Section 3.3 are applied to simulated events to produce the Monte Carlo samples for model development. The associated samples for the B_s and B^0 data are selected by the trained neural networks, while the rectangular cut set is applied in selecting the MC sample for modeling of B^+ data.

Chapter 4

Mass Spectra and Sample Composition

Development of the likelihood begins with the mass space. Candidate mass is a powerful discriminant among different classes of events, and fits of B mass spectra are well-established in CDF B physics analyses. This chapter describes the initial studies to identify the most prominent partially reconstructed decay modes for inclusion in the analysis, the methods for modeling various components of the mass spectra, and the application of the mass likelihood to the data.

4.1 Sample Composition

All prior CDF analyses of flavor oscillations with hadronic B decays have treated only fully reconstructed decay channels as signal candidates. Figure 4-1 depicts an older sample of $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$ decays from the April 2006 CDF evidence for B_s mixing [31]. The large number of partially reconstructed candidates located just below the fully reconstructed mass peak were excluded from the analysis by the requirement that events entering the final likelihood fit have mass in the range $[5.3, 6.0]$ GeV/c^2.

An attempt at inclusion of these unused candidates is well-motivated by the statistical limitations of current B_s mixing analyses. It is expected that this low-mass region contains contributions from various overlapping partially reconstructed decay channels. An initial exploration of this region with simulated events is useful for understanding its dominant contributions and in deciding an appropriate new lower mass bound on the analysis sample. Sample composition studies are performed with the semi-generic Monte Carlo simulation outlined in Section 3.4. As in every phase of this dissertation, high-statistics samples of B^0 and B^+ decays are used for development and validation of the procedure whenever possible. Each of the B_s, B^0, and B^+ data samples has a corresponding semi-generic MC sample. These MC samples are generated with the best present knowledge of b hadron mixtures and branching fractions. The single constraint on the generality of these samples is that events are required to contain the D decay chain that corresponds to the reconstruction

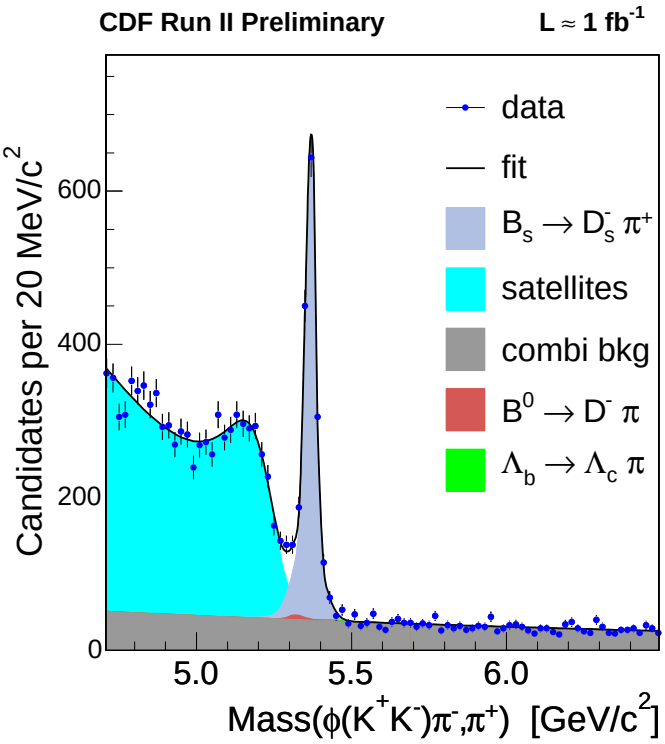

Figure 4-1: Previously used sample of $B_s^0 \to D_s^- \pi^+$ ($D_s^- \to \phi\pi^-$) decays. For mixing analysis, only candidates within the mass range $[5.3, 6.0]$ GeV/c^2 were used, with the partially reconstructed "satellites" being neglected.

procedure. In this sense, the semi-generic MC samples are given the short-hand descriptions listed in Table 4.1. By accessing the generator-level "truth" information of the Monte Carlo events, the partially reconstructed mass region can be decomposed into its numerous contributing decay channels.

Sample	Generic MC Content
B_s	$b \to D_s^- X$
B^0	$b \to D^- X$
B^+	$b \to \overline{D}^0 X$

Table 4.1: Short-hand descriptions of decays allowed for the semi-generic MC simulation used in sample composition studies, where b stands for the appropriate mixture of b hadrons.

The B^+ and B^0 samples are generated with 1.2×10^9 events from BGENERATOR, while the B_s sample is based on 6.0×10^8 events. In all cases, B mesons are generated within the pseudorapidity range $|\eta| < 10$ and with the η-p_T dependence measured in inclusive CDF B data [87]. This sample size is roughly equivalent to 100 pb^{-1}. It is important that the Monte Carlo sample be comparable in size to the data to

allow for adequate population of the numerous decay channels and to account for a significant fraction of events being discarded by the trigger and selection requirements. Figure 4-2 shows the mass spectra of the data and the MC sample of $b \to D^- X$ decays after candidate selection. Figure 4-3 shows the corresponding mass spectra for B^+ data and its generic $b \to \overline{D}^0 X$ MC sample. It is clear that both data and MC samples exhibit the same structure of partially reconstructed candidates below the main fully reconstructed peak, while the data contains a smooth and continuous component across the full mass range; this component is known to arise from inclusion of random tracks, and it is not observed in the MC samples because BGENERATOR creates a single b hadron without the full complement of associated particles from the $p\overline{p}$ collision. It should be noted that the mass spectra shown in these figures are selected with the rectangular cuts of Section 3.3.2, as the neural network selection had not yet been developed for the first pass at these studies. However, the sample decompositions that follow are based on the final selection algorithms used in the analysis. Regardless, the figures are primarily intended to illustrate the discussion.

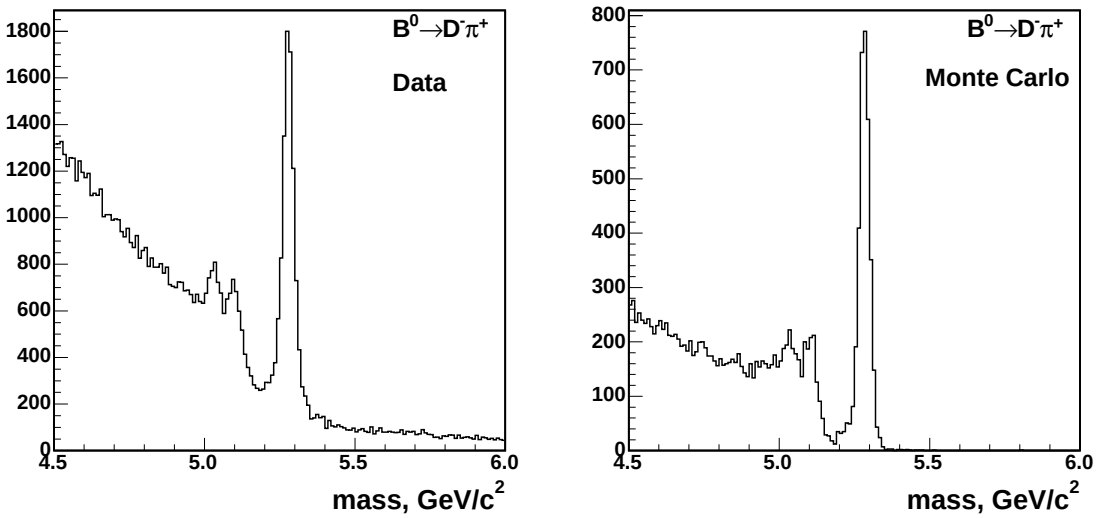

Figure 4-2: Mass spectra of the B^0 data (left) and its semi-generic MC sample (right) after selection with rectangular cuts.

To ensure that a model with only fully reconstructed decays was sufficient, previous analyses excluded all B_s candidates below 5.3 GeV/c^2 in mass. Similarly, the B^0 and B^+ samples were cut off at 5.2 GeV/c^2. One goal of this preliminary examination is to decide what new lower bounds should be applied. The criteria for this decision are that the boundary should be low enough to admit a large number of partially reconstructed decays for incorporation as signal, while not allowing it to be so low that an excessive amount of new background content is added. It is expected that partially reconstructed decays with the topologies $B \to D\rho$ and $B \to D^*\pi$ dominate the main "bump" structure of Figures 4-2 - 4-3. The boundary is placed such that most of the decays of these types are admitted, while other subleading partially

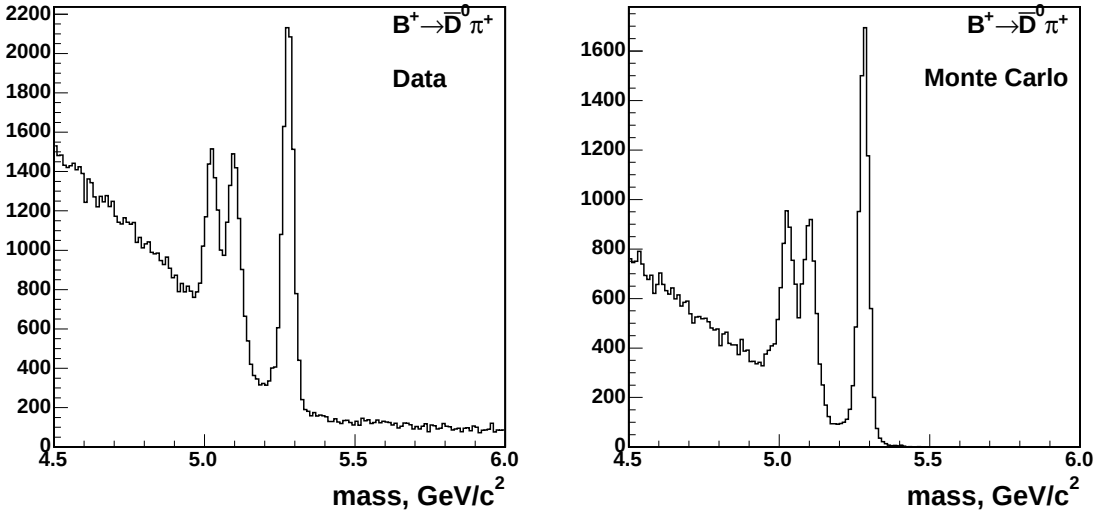

Figure 4-3: Mass spectra of the B^+ data (left) and its semi-generic MC sample (right) after selection with rectangular cuts.

reconstructed contributions, which are to be relegated to "background" status, are minimally present. Examination of distinct decay modes in the MC samples indicates that these goals are satisfied by a lower bound of 4.95 GeV/c^2 for the light B samples and a bound of 5.0 GeV/c^2 for the more massive B_s. Tables 4.2 - 4.3 compile the contributions of all significant channels above the respective cut-offs of the B^0, B^+, and B_s samples. It should be noted that the relative fractions of these tables are not the values applied in the final analysis, where fractions are driven by the data. Separate mass spectra for the partially reconstructed decay modes are withheld until Section 4.2, in which detailed templates are extracted from the large statistics of the dedicated MC samples for these channels. Several observations may be made from the decomposition:

- The largest contribution to each sample comes from the fully reconstructed decay of the form $B \to D\pi$, for which the selection algorithms are optimized.
- The Cabibbo-suppressed channel $B \to DK$ is present at the level of a few percent. Although it is fully reconstructed, this mode arises from the misassignment of the "bachelor kaon" as a pion, resulting in a small miscalculation of its mass and a characteristic bump just below the main peak. Modeling of this component is discussed in Section 4.2.
- The dominant partially reconstructed decay channels are indeed those of the form $B \to D\rho$ and $B \to D^*\pi$. Cabibbo-suppressed decays of $B \to D^*K$ are also present at the percent level via the same mechanism of track misassignment. The $D^*\pi/K$ modes can be treated together as one signal component, both because their overlap in mass makes them indistinguishable in data and because the Cabibbo-suppressed events are also useful as signal.

Channel	N_{events}	Frac [%]
$B^0 \to D^- \pi^+$	2162	51.3
$B^0 \to D^- K^+$	146	3.5
$B^0 \to D^- \rho^+$	1166	27.6
$B^0 \to D^{*-} \pi^+ / K^+$	512	12.1
$B^0 \to D^{(*)-} \ell^+ \nu(\gamma)$	74	1.8
$B^+ \to \overline{D}^{**0} \pi^+$	80	1.9
$B^0 \to D^{**-} \pi^+$	23	0.5
Other	55	1.3
Total	4218	100.0

Channel	N_{events}	Frac [%]
$B^+ \to \overline{D}^0 \pi^+$	9930	36.1
$B^+ \to \overline{D}^0 K^+$	595	2.2
$B^+ \to \overline{D}^0 \rho^+$	4732	17.2
$B^+ \to \overline{D}^{*0} \pi^+ / K^+$	8258	30.1
$B^0 \to D^{*-} \pi^+ / K^+$	3122	11.4
$B^{+/-} \to D \ell \nu(\gamma)$	211	0.8
Other	622	2.3
Total	27470	100.0

Table 4.2: Contributions of various classes of decay modes to the semi-generic MC samples of $b \to D^- X$ (left) and $b \to \overline{D}^0 X$ (right) decays. Events are counted within the mass range for the light B samples, [4.95 , 5.6] GeV/c^2. The symbol ℓ refers only to e and μ leptons. The significant presence of the additional B^0 decay in the generic B^+ sample is notable. The final analysis uses fractions driven by the data.

Channel	N_{events}	Frac [%]
$B_s^0 \to D_s^- \pi^+$	255	36.5
$B_s^0 \to D_s^- K^+$	19	2.7
$B_s^0 \to D_s^- \rho^+$	145	20.8
$B_s^0 \to D_s^{*-} \pi^+ / K^+$	218	31.2
$B_s^0 \to D_s^{(*,**)-} \ell^+ \nu(\gamma)$	14	2.0
Other	47	6.7
Total	698	100.0

Table 4.3: Contributions of various classes of decay modes to the semi-generic MC sample of $b \to D_s^- X$ decays. Events are counted within the mass range for the B_s sample, [5.0 , 6.0] GeV/c^2. The symbol ℓ refers only to e and μ leptons. The final analysis uses fractions driven by the data.

- The B^+ sample contains an additional contribution from $B^0 \to D^{*-} \pi^+$ which has no parallel in the B^0 and B_s samples. This B^0 channel enters the $B^+ \to \overline{D}^0 \pi^+$ reconstruction via the decay of $D^{*-} \to \overline{D}^0 \pi^-$, while an analogous B^+ decay does *not* enter the B^0 sample because it is kinematically forbidden. The presence of this channel limits the utility of the B^+ sample in model development.
- The B_s sample contains no significant contributions from light B mesons of the form $B^{+/0} \to D_s \pi X$.
- The dominant contributions to the B^0 and B_s samples are roughly analogous by $d \leftrightarrow s$ SU(3) symmetry. Although the relative fractions of the components are expected to be different, the structure of the high-statistics B^0 data can be

used to develop the framework for B_s analysis.

Using these new mass ranges, the semi-generic MC samples indicate that approximately 90% of the partially reconstructed sample content is incorporated as signal by inclusion of only the $B \to D\rho$ and $B \to D^*\pi/K$ topologies. In this analysis, the term "signal events" is expanded to include these channels as well as the main fully reconstructed peak. The various small contributions of other $b \to D\pi X$ channels are still present in the sample and must be treated as background events. These "other" partially reconstructed decays are collectively referred to as *physics background*, in contrast with the *combinatorial* background of candidates which include random tracks.

Two other observations may be made regarding the overall semi-generic Monte Carlo samples. First, the selected B^+ sample is larger than the final B^0 sample, despite their both having been generated from the same number of BGENERATOR events. This result is largely due to the fact that B^+ candidates are constructed from only three tracks, as opposed to the four tracks that make up B^0 and B_s candidates. Distributing the same B momentum over fewer tracks implies that the trigger is more likely to accept the event. As a related observation, the final samples of semi-generic MC are not especially large, having their equivalent luminosity distributed over the full range of b hadron decays. The resulting sample sizes for the decays of interest are large enough to draw preliminary conclusions on relative abundance, but they are not large enough to produce statistically significant models of mass and proper time behavior. Such modeling requires generation of dedicated MC samples for the dominant partially reconstructed decay channels to be used as signal.

4.2 Mass Templates

Having identified the dominant components of the partially reconstructed mass region, templates must be derived to model their mass spectra. Modeling of structures in this range is important for the correct representation of signal fractions in the likelihood. Large dedicated Monte Carlo samples are used to derive the templates for individual signal components, while the physics background is modeled collectively by the generic MC simulation. The dedicated MC samples are summarized below, followed by the parameterization of each sample component model.

4.2.1 Dedicated Monte Carlo

Dedicated Monte Carlo samples are generated for each of the prominent partially reconstructed channels which are to be incorporated as signal decays. The samples originate with 6×10^7 BGENERATOR events of the same kinematic distributions as the semi-generic samples. However, because EvtGen is restricted to decay the B mesons only via the decay chains of interest, the effective luminosity of the dedicated samples is orders of magnitude larger than the semi-generic simulation. The D^* species are configured to decay with the world average branching fractions [2], a fact which is

important for insuring that the $B_s^0 \to D_s^{*-}\pi^+$ mass templates receive the correct contributions from the $D_s^{*-} \to D_s^-\gamma$ and $D_s^{*-} \to D_s^-\pi^0$ decay modes. The Cabibbo-favored and suppressed $B \to D^*\pi/K$ modes are generated in parallel with the best available EvtGen relative ratios and are always treated together.

The dedicated MC samples are used throughout this analysis to model the behavior of distinct decay channels in various measurement spaces. In this chapter, they are applied for derivation of mass spectra, but the samples are equally applicable to later stages of the likelihood development.

Other Dedicated MC Samples

Dedicated samples are also simulated for the fully reconstructed $B \to D\pi$ decay channels for which the event selection is optimized. The simulation is performed exactly as above, with 6×10^7 BGENERATOR events and a restricted EvtGen decay chain. These samples are primarily used for modeling of proper decay time, with one notable exception for the mass space discussed below.

Similarly configured dedicated MC samples of 2×10^7 events are also generated for the fully reconstructed Cabibbo-suppressed $B \to DK$ channels. These decays are also treated as signal events, and the MC simulation is used to model their mass spectra. The relative efficiency with which these simulated events are triggered and selected is used to derive the Cabibbo-suppressed contribution, as noted below. Other MC samples, such as for $\Lambda_b \to \Lambda_c^+\pi^-$, are generated for application in modeling very small contributions to the data. These component classes are discussed in the following sections.

4.2.2 Function Definitions

This section compiles the definitions of functions which appear repeatedly in the model parameterizations of the following sections.

The Gaussian function is useful as a likelihood component, and its PDF is unit-normalized within the mass bounds of the fitted data. Because it appears repeatedly, the function is abbreviated and defined as

$$\mathcal{G}(m|M,\sigma_m,M_{\min},M_{\max}) = \frac{\frac{1}{\sqrt{2\pi}\sigma_m}e^{-\frac{1}{2}\left(\frac{m-M}{\sigma_m}\right)^2}}{\frac{1}{2}\left[\mathrm{Erf}\left(\frac{M_{\max}-M}{\sqrt{2}\sigma_m}\right)+\mathrm{Erf}\left(\frac{M-M_{\min}}{\sqrt{2}\sigma_m}\right)\right]}, \tag{4.1}$$

where m is an event mass and the Gaussian mean and width are given by the fit parameters M and σ_m. The error function, $\mathrm{Erf}(z) \equiv \frac{2}{\sqrt{\pi}}\int_0^z e^{-u^2}\,du$, is used to express the normalization factor. The fit boundaries $M_{\min}$ and $M_{\max}$ are usually dropped from the notation but remain implied.

Exponential shapes are often useful for modeling the presence of a component with a relatively sharp turn-on in the mass space. In this case, the standard bounded

exponential function

$$\mathrm{Exp}(m|M,\mu,M_{\mathrm{min}},M_{\mathrm{max}}) = \frac{\frac{1}{\mu}e^{-(m-M)/\mu}}{e^{-(M_{\mathrm{min}}-M)/\mu} - e^{-(M_{\mathrm{max}}-M)/\mu}}, \tag{4.2}$$

with characteristic decay interval μ and turn-on position M is often analytically convolved with a Gaussian function to account for kinematic effects. An important model for background uses a pure exponential which is continuous across the full mass range, with $M = (M_{\mathrm{max}} + M_{\mathrm{min}})/2$.

Models of backgrounds also involve linear functions. In the case that the component extends across the full mass range, the normalized function is simply

$$\mathrm{Lin}(m|a,M_{\mathrm{min}},M_{\mathrm{max}}) = am + \frac{1}{M_{\mathrm{max}} - M_{\mathrm{min}}} \cdot \left[1 - \frac{a}{2}(M_{\mathrm{max}}^2 - M_{\mathrm{min}}^2)\right], \tag{4.3}$$

with slope a. When the component ends at some upper cut-off within the bounds of the mass range, the normalized function is

$$\begin{aligned}\mathrm{LinCut}(m|C,M_{\mathrm{min}},M_{\mathrm{max}}) = {} & \frac{-2}{(M_{\mathrm{max}} - M_{\mathrm{min}})(2C - M_{\mathrm{max}} - M_{\mathrm{min}})}m \\ & + \frac{1}{M_{\mathrm{max}} - M_{\mathrm{min}}} \cdot \left[1 + \frac{(M_{\mathrm{max}}^2 - M_{\mathrm{min}}^2)}{(M_{\mathrm{max}} - M_{\mathrm{min}})(2C - M_{\mathrm{max}} - M_{\mathrm{min}})}\right],\end{aligned} \tag{4.4}$$

where the cut-off position C fully determines the slope.

4.2.3 Signal Components

This section introduces the templates used to model the mass spectra of all signal components. The features of the dominant partially reconstructed channels can be qualitatively estimated, based on fundamental properties of the decays. This basic understanding motivates the functional form of the models that are applied to the MC distributions, where they are fitted to extract each component's parameterization. These models make frequent use of the functions defined in Section 4.2.2.

$B \to D\pi$ Parameterization

Using a perfect detector, the observed mass distribution of a $B \to D\pi$ channel would be a delta function of negligible width. Because these decays are fully reconstructed, it can be said that their mass spectra contain no "kinematic smearing" from lost momentum. However, every detector has finite mass resolution, so the observed spectrum takes the approximate form of a Gaussian distribution. Given a large enough sample of decays, the fully reconstructed peak may exhibit multiple classes of detector resolution and be better modeled by more than one Gaussian function. In this analysis, fully reconstructed mass peaks are parameterized by a unit-normalized

double-Gaussian function which enters the likelihood as

$$\mathcal{P}^{\rm m}_{D\pi} = f_1 \cdot \mathcal{G}(m|M,\sigma_1) + (1-f_1)\cdot\mathcal{G}(m|M,r_2\sigma_1), \tag{4.5}$$

where both terms share the same central mass but have independent widths and the second subcomponent's width is larger by ratio r_2.

Because the fully reconstructed peaks are very prominent in the data, there is no need to extract the parameterizations from MC simulation, and the above parameters float freely in the fits to data. However, the dedicated MC samples for $B \to D\pi$ modes have an important application in modeling of the mass space for this analysis.

Despite the GEANT detector simulation's being extensively tuned, it cannot be assumed that the central values of mass structures are identically the same in simulation and in data. In fact, it is known that the mass of reconstructed MC events are $3-7$ MeV/c^2 higher than those of data. This 0.1% difference arises from the momentum scale of MC tracks. Because the parameterizations of partially reconstructed mass components, including their offsets with respect to the fully reconstructed peak, *are* extracted from MC for application to the data, it is important that the mass scale of the simulation is taken into account. This is simply accomplished by fitting the central mass values M in the dedicated MC samples of the fully reconstructed channels. The structures of their associated partially reconstructed channels are then parameterized with respect to that fixed fully reconstructed mass value. The relative offsets that are derived from MC simulation in this way can be applied directly to the data. Table 4.4 lists the values of M fitted in the dedicated $B \to D\pi$ MC samples and which are applied in the following template fits.

Channel	MC Mass [MeV/c^2]	
$B^0_s \to D^-_s\pi^+$	5374.24 $\pm$	0.04
$B^0 \to D^-\pi^+$	5283.74 $\pm$	0.04
$B^+ \to \overline{D}^0\pi^+$	5281.30 $\pm$	0.04

Table 4.4: Central mass values of the fully reconstructed $B \to D\pi$ dedicated MC samples used for the relative mass offset of partially reconstructed component templates.

$B \to D^*\pi$ Parameterization

Decays of $B^0 \to D^{*-}\pi^+$ and $B^+ \to \overline{D}^{*0}\pi^+$ are characterized by the presence of the vector D^* meson, which undergoes a polarized decay to the scalar D and π mesons. This polarization preferentially aligns the neutral pion along the forward or backward directions of the D^* momentum, causing the mass of the reconstructed charged particles to take on a double-peaked shape. This "double-horn" is readily observed in the spectra of the semi-generic MC samples. Additionally, the phase space available to the D^* decay is small. The example of $D^{*-} \to D^-\pi^0$ has Q-value

equal to

$$Q_{D^{*-}\to D^-\pi^0} = [M_{D^{*-}} - (M_{D^-} + M_{\pi^0})]\, c^2 = 5.7 \text{ MeV}. \tag{4.6}$$

Indeed, the fortunate parallelism of the dominant partially reconstructed content in the B^0 and B_s samples arises from the fact that the $\overline{D}^{*0} \not\to D^-\pi^+$ decay is kinematically forbidden, preventing any significant B^+ contribution to the B^0 sample. The highly constrained phase space implies that the D and π daughters have small momentum in the D^* rest frame. The "soft" neutral pion therefore causes the loss of very little momentum from the reconstruction, allowing the measured mass distribution to be narrow and well-defined. The well-defined momentum is also the basis for the expectation that these events are valuable in the oscillation analysis, due to the resulting good proper time resolution.

The above description applies equally well to the $D^{*-} \to \overline{D}^0\pi^-$ content of the B^+ sample, in which the soft pion is charged but was not reconstructed.

Another consideration for $B \to D^*\pi$ decays is that all D^* species have some branching fraction for $D^* \to D\gamma$. While this fraction is sub-dominant for the light D^* decays ($\sim$2% for D^{*-}, $\sim$38% for D^{*0}), the D_s^{*-} decay has opposite emphasis [2] due to an even tighter $D_s^-\pi^0$ phase space:

$$\begin{array}{ll} D_s^{*-} \to D_s^-\gamma & (94.2 \pm 0.7)\% \\ D_s^{*-} \to D_s^-\pi^0 & (\ 5.8 \pm 0.7)\% \end{array}$$

Regardless of D^* species, the mode that includes a spin-1 photon is a P-wave decay which does not induce the forward-backward recoil of the π^0 channel. The zero rest mass of the photon also creates more phase space for the decay. Therefore, the γ channel is expected to contribute a single kinematically smeared mass structure which is somewhat wider than that of the π^0 double-horn. In the case of $B_s^0 \to D_s^{*-}\pi^+$, this single peak dominates its overall contribution. It is noted that the B_s data in Figure 4-1 does not exhibit the same obvious double-peak structure of the light B samples.

The kinematic effects outlined in these simple arguments are present concurrently with the detector resolution effects common to all mass components. This empirical model represents the convolution of both kinematic and resolution effects. The double-horn structure of the $B \to D^*\pi$ decay topology is parameterized by two Gaussian distributions, which model the peaks, centered on a wider Gaussian. The normalized PDF that enters the likelihood may be expressed as

$$\begin{aligned} \mathcal{P}^{\rm m}_{D^*\pi} = (1 - f_{\rm horns}) &\cdot \mathcal{G}(m|M - \Delta_{D^*\pi}, \sigma_3) \\ + \quad f_{\rm horns} &\cdot [(1 - f_{\rm horn2}) \cdot \mathcal{G}(m|M - \Delta_{D^*\pi} - \delta_{\rm horns}, \sigma_4) \\ + \quad f_{\rm horn2} &\cdot \mathcal{G}(m|M - \Delta_{D^*\pi} + \delta_{\rm horns}, \sigma_4)], \end{aligned} \tag{4.7}$$

where $\Delta_{D^*\pi}$ is the offset of the distribution center with respect to the main fully reconstructed peak at M, and the $\delta_{\rm horns}$ parameter defines the splitting of the horns. Both horns share the same width, while the central Gaussian width is independently varied. Two fraction parameters define the relative content of the three subcomponents,

where $f_{\rm horns}$ is expected to be small for the B_s case.

This parameterization is extracted from the dedicated MC simulation of $B \to D^*\pi$ and $B \to D^*K$.Although this combined component is often referred to as simply "$D^*\pi$", the presence of the small Cabibbo-suppressed contribution is always implied.

Figure 4-4 depicts the fitted templates for application in the B^0 and B_s data. The qualitative difference between the sharp double-horn structure of the $B^0 \to D^{*-}\pi^+$ template and the single central peak of the $B_s^0 \to D_s^{*-}\pi^+$ template is clear and is explained by the dominance of the $D_s^{*-} \to D_s^-\gamma$ subdecay. Figure 4-5 shows the templates for the two $D^*\pi$ contributions to the B^+ sample, including the $B^0 \to D^{*-}\pi^+$ cross-talk which has no analogue in the B^0 and B_s samples. The overall features of the B^+ templates are similar to those of the B^0 template, because both are double-horn shapes driven by polarized $D^* \to D\pi$ decays with similar kinematics.

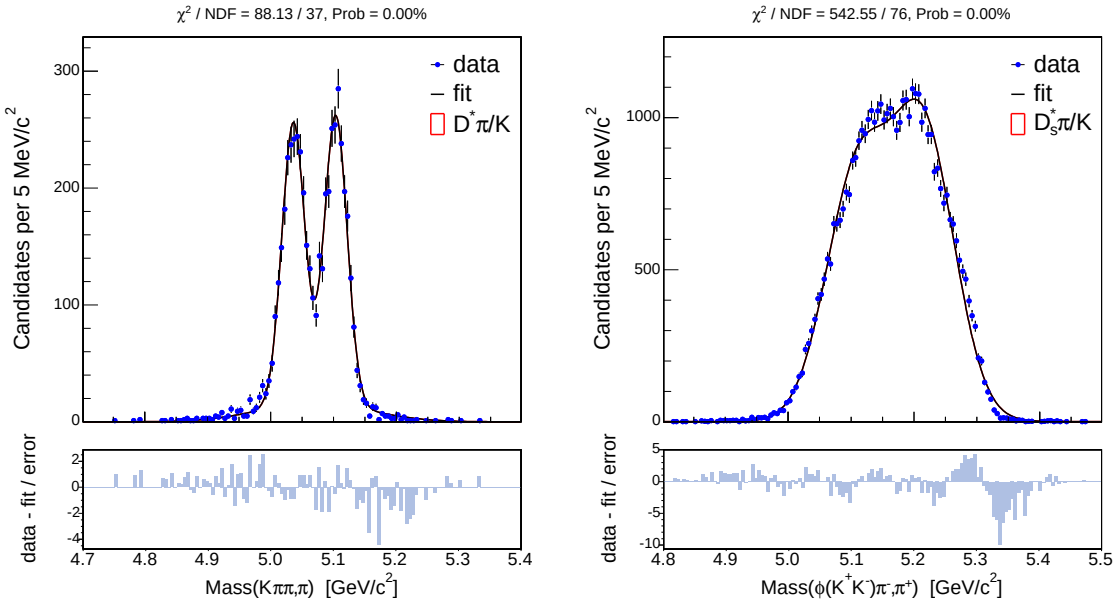

Figure 4-4: Mass templates for $B^0 \to D^{*-}\pi^+$ decays reconstructed as $B^0 \to D^-\pi^+$ (left) and $B_s^0 \to D_s^{*-}\pi^+$ decays reconstructed as $B_s^0 \to D_s^-\pi^+$ (right).

The χ^2 agreement between the fitted parameterizations and the simulation spectra is not universally good. These templates are composed of empirical parameterizations which are motivated by physics but which are extracted from MC samples that reflect the full complexity of kinematics and detector interactions. It is possible to increase the complexity of the parameterization to an arbitrary degree, such that these MC distributions are very well described. However, it must be noted that the B_s MC samples are approximately 20 times larger than the corresponding data. These distributions exhibit features which are completely unobservable in the data. Essentially, it is concluded that this three-component parameterization is an acceptable compromise between template complexity and perfect modeling of the B_s data. Most importantly, conservative systematic uncertainties are evaluated in Section 7.2.3 and Appendix A

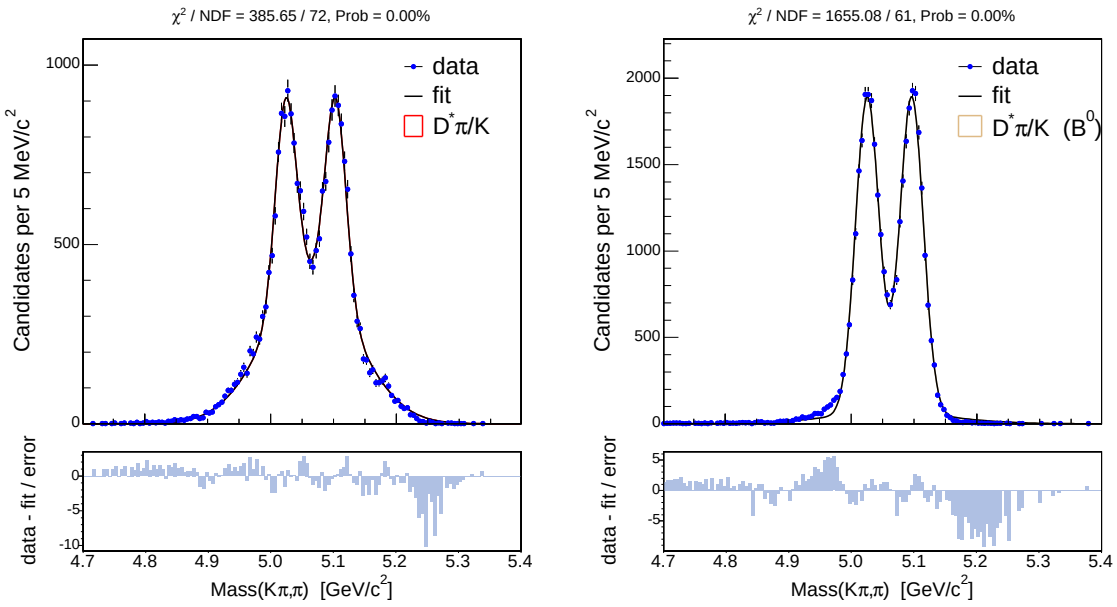

Figure 4-5: Mass templates for $B^+ \to \overline{D}^{*0}\pi^+$, $\overline{D}^{*0} \to \overline{D}^0\pi^0/\gamma$ (left) and $B^0 \to D^{*-}\pi^+$, $D^{*-} \to \overline{D}^0\pi^-$ (right) decays reconstructed as $B^+ \to \overline{D}^0\pi^+$.

to estimate the effects of mismodeled mass template. In practice, an incorrect mass template is equivalent to using slightly incorrect weights in the likelihood (as a function of candidate mass) for the proper time measurements of each decay class. The derivation of systematic uncertainties examines the effect of confusing the partially reconstructed signal components, and the results support the assertion that these templates are adequate.

The overall normalization of this component with respect to the other signal content is fitted in data.

$B \to D\rho$ Parameterization

The $B \to D\rho$ decay topology is characterized by greater available phase space than that of $D^*\pi$, which results in a stronger smearing of the reconstructed mass distribution. In this case, the lost neutral particle is from the $\rho^+ \to \pi^+\pi^0$ decay (~100% branching fraction), with $Q_{\rho^+\to\pi^+\pi^0} \sim 500$ MeV. The B mass spectrum is expected to be asymmetrical, with its upper side constrained by the fully reconstructed mass and with a long tail on the lower side. Additionally, the vector meson ρ^+ undergoes a similarly polarized decay as that of the D^*. The difference for this case is that the large Q-value provides a stronger smearing of the double-peaks, causing them to appear less prominently.

This empirical model again represents the convolution of these kinematic effects with the detector resolution. The $B \to D\rho$ topology is parameterized by a Gaussian-smeared exponential function toward low mass and a simple displaced Gaussian. This

combination models the weak double-horn structure with a long tail. The normalized PDF that enters the likelihood is expressed as

$$\begin{aligned}\mathcal{P}^{\mathrm{m}}_{D\rho} = \quad & f_{\mathrm{exp}} \cdot \mathrm{Exp}(-m|M - \Delta_{D\rho}, \mu_{D\rho}) \otimes \mathcal{G}(m|\sigma_5) \\ & + (1 - f_{\mathrm{exp}}) \cdot \mathcal{G}(m|M - \Delta_{\mathrm{D}\rho} + \delta_{\mathrm{gaus}}, \sigma_6),\end{aligned} \tag{4.8}$$

where the overall offset $\Delta_{D\rho}$ is fitted with respect to the centroid M from fully reconstructed MC simulation, and δ_{gaus} defines the offset of the small Gaussian peak from the smeared exponential. The parameter $\mu_{D\rho}$ defines the tail of the structure toward low mass. The σ width parameters are varied independently for each subcomponent.

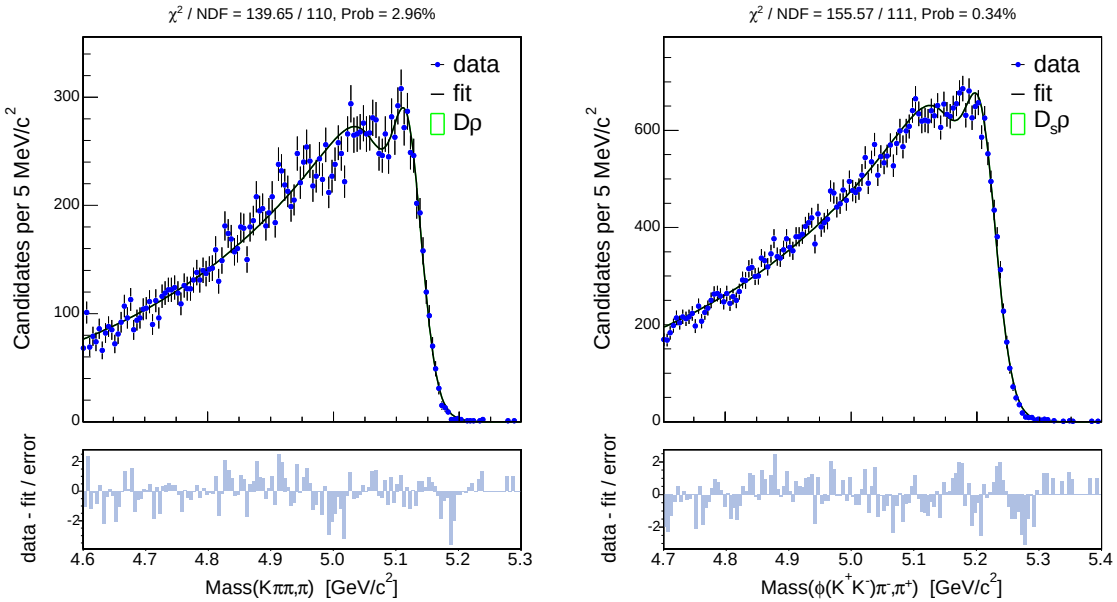

Figure 4-6: Mass templates for $B^0 \to D^-\rho^+$ decays reconstructed as $B^0 \to D^-\pi^+$ (left) and $B^0_s \to D^-_s\rho^+$ decays reconstructed as $B^0_s \to D^-_s\pi^+$ (right).

Figure 4-6 shows the fitted templates for the B^0 and B_s samples, while the B^+ template is shown in Figure 4-7. Because the partially reconstructed ρ^+ resonance is independent of the species of reconstructed B, the similarity of these mass distributions is reasonable. The same arguments from the $B \to D^*\pi$ discussion regarding fit quality and the large number of simulated events also apply here. As above, the overall normalization of this component is fitted in data.

Cabibbo-Suppressed $B \to DK$ Parameterization

In addition to partially reconstructed decay channels, the sample composition study notes the expected presence of Cabibbo-suppressed $B \to DK$ modes in the samples reconstructed for $B \to D\pi$. This contribution arises from the misinterpretation of the "bachelor kaon" as a pion. The selection algorithms do not apply explicit

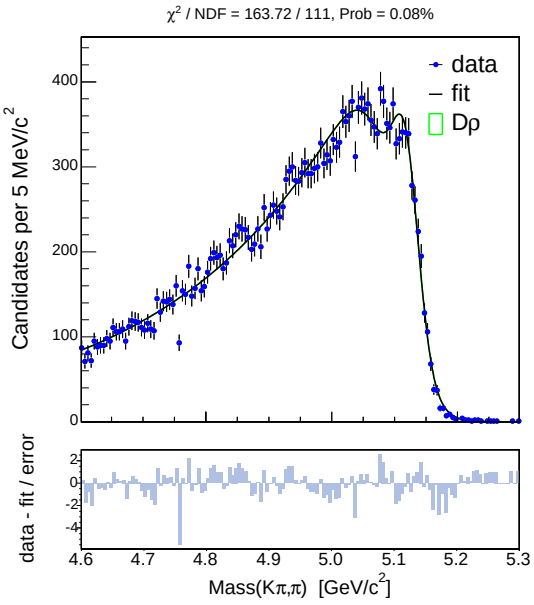

Figure 4-7: Mass templates for $B^+ \to \overline{D}^0 \rho^+$ decays reconstructed as $B^+ \to \overline{D}^0 \pi^+$.

particle identification information to this track, so its particle assignment is simply a hypothesis. The $B \to DK$ decays are fully reconstructed, in that no particles are lost, but the measured mass spectrum is made asymmetrical and offset to a lower central value than the true B mass because of the track mass misassignment. This component is treated as usable signal, modulo the associated systematic uncertainty discussed in Appendix A. The empirical model, which is extracted from dedicated MC samples, consists of an offset Gaussian function overlaid with a smeared exponential whose tail points toward low mass. The normalized PDF that enters the likelihood is expressed as

$$\begin{aligned} \mathcal{P}^{\rm m}_{\rm Cab} = \quad & f_{\rm gaus} \cdot \mathcal{G}(m|M-\Delta_{\rm Cab},\sigma_{\rm Cab}), \\ & + (1-f_{\rm gaus}) \cdot {\rm Exp}(-m|M-\Delta_{\rm Cab},\mu_{\rm Cab}) \otimes \mathcal{G}(m|\sigma_{\rm Cab}), \end{aligned} \tag{4.9}$$

where the subcomponents share a common offset $\Delta_{\rm Cab}$ with respect to the fully reconstructed mass M and a common Gaussian smearing $\sigma_{\rm Cab}$. The parameter $\mu_{\rm Cab}$ defines the tail of the structure toward low mass.

The overall normalization of this Cabibbo-suppressed contribution is also extracted from MC simulation in combination with branching fraction information. The $B \to DK$ normalization in the likelihood is fixed with respect to the floating normalization of the Cabibbo-favored fully reconstructed peak. This ratio is based on the ratio of branching fractions $\mathcal{B}(B \to DK)/\mathcal{B}(B \to D\pi)$ and corrected by the relative efficiency of triggering and reconstruction from the MC samples. For the B^0 and B^+ channel, the branching fractions are those of the world averages [2]. A value of 5% is used for the $\mathcal{B}(B_s^0 \to D_s^- K^+)/\mathcal{B}(B_s^0 \to D_s^- \pi^+)$ ratio, and a systematic uncer-

tainty is assigned in Appendix A for this assumption. The reconstruction efficiencies are calculated by simply counting the number of simulated events of $B \to DK$ and $B \to D\pi$ accepted by the $B \to D\pi$ selection within the mass range for the fit. This calculation is closely related to the procedure for deriving the constraints on relative branching fractions of the main signal components in Section 4.3, except that these $B \to DK$ ratios are permanently fixed. Table 4.5 lists the normalization ratio $r_{\rm Cab}$ and template parameters used for each sample's fully reconstructed Cabibbo-suppressed contribution.

Parameter	$B_s^0 \to D_s^- K^+$	$B^0 \to D^- K^+$	$B^+ \to \overline{D}^0 K^+$
$r_{\rm Cab}$	0.0609	0.0580	0.0616
$\Delta_{\rm Cab}$ [MeV/c^2]	52.7	61.4	57.2
$\sigma_{\rm Cab}$ [MeV/c^2]	26.9	28.2	27.5
$f_{\rm gaus}$	0.666	1.000	1.000
$\mu_{\rm Cab}$ [MeV/c^2]	43.1	—	—

Table 4.5: Template parameters and normalization ratios of Cabibbo-suppressed $B \to DK$ contributions with respect to the fully reconstructed $B \to D\pi$ peak.

4.2.4 Background Components

Whereas many aspects of the signal models introduced in the previous section are physically motivated, the parameterizations of background components are completely empirical. The primary consideration in modeling of background is that it receives the correct weight in the likelihood, such that the extraction of signal parameters from overlapping components is not biased. The template models are based on functions defined in Section 4.2.2.

$b \to DX$ "Physics" Background

The sample composition study of Section 4.1 shows that the mass region of partially reconstructed decays is dominated by the $B \to D^*\pi$ and $B \to D\rho$ channels, but that there are also small contributions of other $b \to D\pi X$ channels. The collective mass spectra of this physics background are parameterized based on semi-generic Monte Carlo simulation.

The signal channels of the previous section must be treated individually in the likelihood because the oscillation signature is extracted from their proper time models. By contrast, since details of the proper time behavior of partially reconstructed background decays are not of interest, they can be grouped together with empirical models in both mass and time. Additionally, since the contributing decays (for example, $D^{**}\pi$, Da_1, *etc.*) have larger Q-values and/or more than one unreconstructed particle, the collective mass spectrum should be smooth. It is shown in fits to the data that the contribution of physics background to the B_s mixing sample is small, at

the level of a few percent. Most importantly, the ability to resolve flavor oscillations is insensitive to the model for this small component.

The collective mass spectrum of partially reconstructed physics background decays is parameterized by linear "wedge" shapes which are more prominent at low mass and which fall to zero at an upper cut-off point. The normalized PDF that enters the likelihood is expressed as

$$\mathcal{P}^{\mathrm{m}}_{DX} = (1 - f_{\mathrm{wedge2}}) \cdot \mathrm{LinCut}(m|M - \Delta_{\mathrm{wedge1}}) + f_{\mathrm{wedge2}} \cdot \mathrm{LinCut}(m|M), \qquad (4.10)$$

where the cut-off point fully constrains the wedge slope. The parameter Δ_{wedge1} determines the offset of the first wedge's cut-off position with respect to the fully reconstructed mass M, while the (small) second wedge extends up to M by construction. Just as for partially reconstructed signal templates, the fixed value of M is taken from the MC simulation of the fully reconstructed mass centroid.

Figure 4-8 shows the fitted templates for the B^0 and B_s samples. In both cases, the linear wedge model is in reasonable agreement with the semi-generic MC distributions. The secondary wedges are assigned small fractions ($< 10\%$) to account for the few events that approach the fully reconstructed mass.

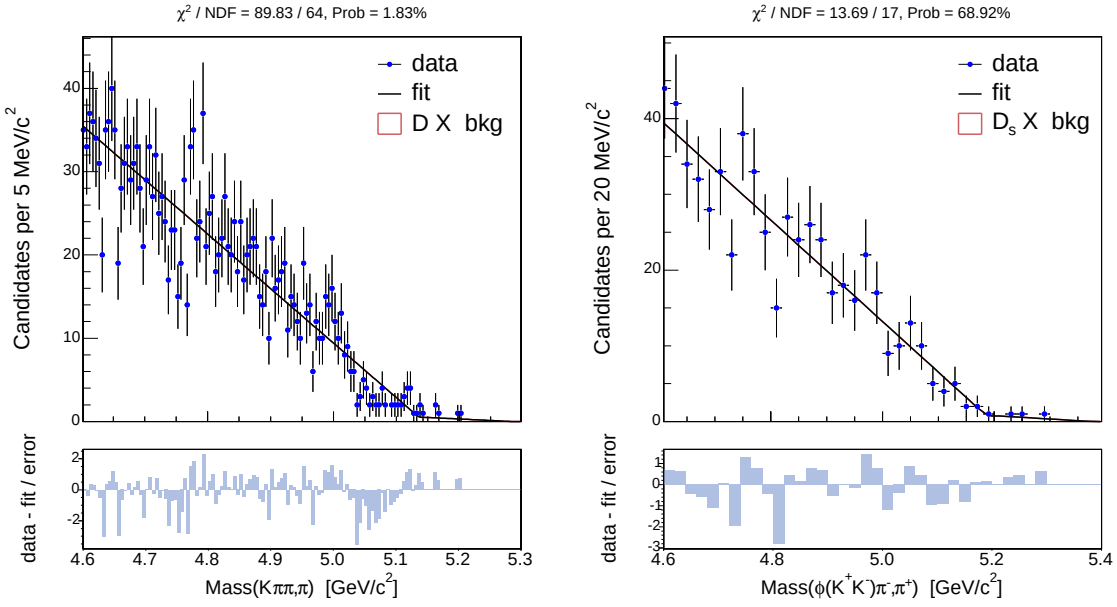

Figure 4-8: Mass templates for the collective $b \to D^- \pi^+ X$ physics background reconstructed as $B^0 \to D^- \pi^+$ (left) and $b \to D_s^- \pi^+ X$ background reconstructed as $B_s^0 \to D_s^- \pi^+$ (right).

The template for the $b \to \overline{D}^0 \pi^+ X$ physics background of the B^+ sample is a minor exception to this model, as shown in Figure 4-9. Its parameterization exchanges the first simple wedge for a linear cut-off function which is modulated by an exponential factor to hasten the fall-off. The second wedge is exchanged for a simple exponential.

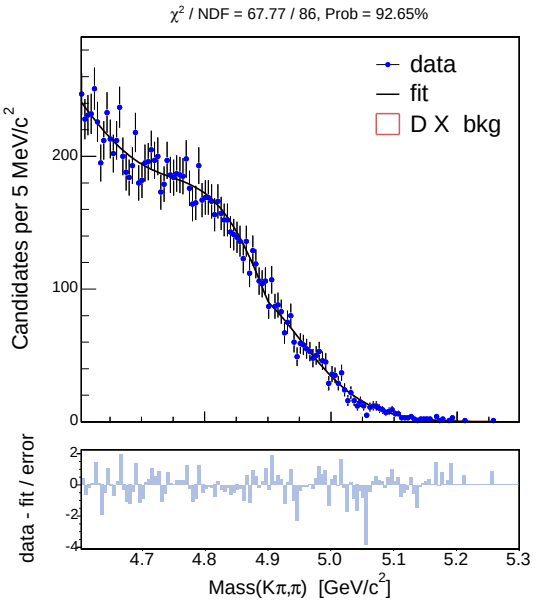

Figure 4-9: Mass templates for the collective $b \to \overline{D}^0 \pi^+ X$ physics background reconstructed as $B^+ \to \overline{D}^0 \pi^+$.

These modifications are made necessary by the presence of slightly more prominent $D^* \rho$ decays (in which two pions are lost), whose visible bump structure is not fully smoothed by other background channels. The MC distribution is well-described by the modified template. As is the signal templates, the overall normalization of this component is fitted in data.

Reflections

In a similar fashion to the Cabibbo-suppressed $B \to DK$ contribution, small components of other b hadrons enter the samples due to misassignment of track identity. The relevant examples of these "reflections" in the B_s sample are:

- True decays of $B^0 \to D^- \pi^+$, $D^- \to K^+ \pi^- \pi^-$, in which a pion track from the D^- vertex is hypothesized to be a kaon and the resulting K"K"mass falls within the accepted window of the ϕ resonance (and the general neural network selection).

- True decays of $\Lambda_b \to \Lambda_c^+ \pi^-$, $\Lambda_c^+ \to p^+ K^- \pi^+$, in which the proton track is hypothesized to be a kaon and the resulting candidate passes selection.

The B^0 sample also contains such reflection candidates, except that the first case has a kaon from a true $B_s^0 \to D_s^- \pi^+$ decay being misinterpreted as a pion, while the second case has the proton track being assigned as a pion.

The B^+ sample does not contain such contributions because its analogous charged reflections, such as B_c^+ and heavy b baryons, are heavily suppressed at production.

Whereas Cabibbo-suppressed $B \to DK$ decays are of the same B species as the targeted sample decay and so are treated as signal, the reflection candidates are a background source [1] . However, the approach to modeling the reflections in the mass space is very similar to that of the Cabibbo component. The misidentification of a track causes the measured candidate mass to be smeared and offset from the true b hadron mass. Depending on whether it is a Λ_b, B^0, or B_s being reconstructed as a B_s or B^0, the resulting mass distribution may be centered at a higher or lower mass than the main fully reconstructed peak. This mass template and its relative normalization is extracted from dedicated MC samples generated for each of the misreconstructed b hadron decay chains.

The empirical model of the reflection of B^0 decays in the B_s data sample is parameterized by a smeared exponential distribution. The normalized PDF that enters the B_s likelihood is

$$\mathcal{P}^{\mathrm{m}}_{B^0\mathrm{ref}} = \mathrm{Exp}(m|M - \Delta_{B^0\mathrm{ref}}, \mu_{B^0\mathrm{ref}}) \otimes \mathcal{G}(m|\sigma_{B^0\mathrm{ref}}), \tag{4.11}$$

where the offset from the fully reconstructed peak, characteristic decay interval, and Gaussian smearing are all denoted by the usual Δ, μ, and σ parameters. The shape of the mass distribution for the reflection of B_s decays in the B^0 data is qualitatively different and is better parameterized by a double-Gaussian function. The normalized PDF that enters the B^0 likelihood is

$$\mathcal{P}^{\mathrm{m}}_{B_s\mathrm{ref}} = (1 - f_2) \cdot \mathcal{G}(m|M - \Delta_{B_s\mathrm{ref}}, \sigma_{B_s\mathrm{ref}}) + f_2 \cdot \mathcal{G}(m|M - \Delta_{B_s\mathrm{ref}}, r_2\sigma_{B_s\mathrm{ref}}), \tag{4.12}$$

where the two distributions share the same offset from the main peak but the second subcomponent has width larger than the first by the ratio of parameter r_2.

The empirical model for the reflection of Λ_b decays is the same for both the B_s and B^0 reconstructed samples, consisting of a Gaussian distribution and a smeared exponential with its tail toward low mass. The normalized PDF is

$$\begin{aligned}\mathcal{P}^{\mathrm{m}}_{\Lambda_b\mathrm{ref}} = &(1 - f_{\mathrm{exp}}) \cdot \mathcal{G}(m|M - \Delta_{\Lambda_b\mathrm{ref}}, \sigma_{\Lambda_b\mathrm{ref}}) \\ + \quad &f_{\mathrm{exp}} \cdot \mathrm{Exp}(-m|M - \Delta_{\Lambda_b\mathrm{ref}}, \mu_{\Lambda_b\mathrm{ref}}) \otimes \mathcal{G}(m|\sigma_{\Lambda_b\mathrm{ref}}),\end{aligned} \tag{4.13}$$

where both subcomponents share the same mass offset and Gaussian smearing.

The overall normalization of these components in the likelihood is derived analogously to that of the Cabibbo-suppressed contribution. The fixed ratios of normalization with respect to the floating fully reconstructed peak are extracted from dedicated MC samples in combination with branching fraction information. In this case, the ratios of branching fractions and production cross-sections are taken from various CDF measurements [92–94]. Systematic uncertainties are assigned in Appendix A for these calculations.

Table 4.6 lists the normalization ratios r_{ref} and template parameters used for the B^0 and B_s reflection components, while Table 4.7 lists those of the Λ_b reflection.

[1] Indeed, reflections could be considered to be part of the "physics background", but because the reflections occupy distinct regions of the mass space, the practical convention is that they are not.

$B_s^0 \to D_s^- \pi^+$		$B^0 \to D^- \pi^+$	
Parameter	Value	Parameter	Value
$r_{B^0\text{ref}}$	$1.32 \pm 0.42\%$	$r_{B_s\text{ref}}$	$0.64 \pm 0.64\%$
$\Delta_{B^0\text{ref}}$	61.9 MeV/c^2	$\Delta_{B_s\text{ref}}$	-37.0 MeV/c^2
$\sigma_{B^0\text{ref}}$	19.0 MeV/c^2	$\sigma_{B_s\text{ref}}$	21.7 MeV/c^2
$\mu_{B^0\text{ref}}$	24.0 MeV/c^2	r_2	2.05
		f_2	19.2%

Table 4.6: Template parameters and normalization ratios of B^0 and B_s reflection components with respect to the fully reconstructed $B_s^0 \to D_s^- \pi^+$ and $B^0 \to D^- \pi^+$ peaks, respectively. The charged B^+ sample contains no neutral B reflections.

Parameter	$B_s^0 \to D_s^- \pi^+$	$B^0 \to D^- \pi^+$
$r_{\Lambda_b\text{ref}}$	$0.62 \pm 0.21\%$	2.65±??%
$\Delta_{\Lambda_b\text{ref}}$ [MeV/c^2]	-70.5	-160.5
$\sigma_{\Lambda_b\text{ref}}$ [MeV/c^2]	33.9	35.6
f_{exp}	0.596	0.0006
$\mu_{\Lambda_b\text{ref}}$ [MeV/c^2]	77.1	75.8

Table 4.7: Template parameters and normalization ratios of Λ_b reflection components with respect to the fully reconstructed $B \to D\pi$ peaks. The charged B^+ sample contains no neutral B reflections.

Combinatorial Background

In addition to backgrounds which arise from misinterpretation of real b hadrons, a significant source of fake candidates is associated with random track combinations. A typical event of this *combinatorial* background class is comprised of a real D meson whose momentum vector forms a vertex with a random track. The secondary D candidates themselves are not completely excluded from being formed by unrelated tracks, but the preselection requirements and neural network variables strongly diminish such a contribution.

The p_T distribution of the numerous tracks present in each event is concentrated toward low momentum. Furthermore, the likely scenario for reconstruction of these candidates is that the random fragmentation track and real D meson originate promptly from the primary vertex, creating an opening angle that tends to be small. The combination of these attributes implies that combinatorial background is more prominent in the samples at lower mass values than at high ones.

The upper mass boundary of the fitted sample is chosen to provide a range in which events are almost exclusively combinatorial background. These sideband candidates are useful in that they establish the mass dependence of the combinatorial component directly from the data. The fitted model is extrapolated across the full mass range, through the region of the signal events in which this background cannot be readily distinguished. The combinatorial background is parameterized in the data

by superposition of a smooth exponential distribution and a flat linear component. The normalized PDF that enters the likelihood is

$$\begin{aligned}\mathcal{P}^{\mathrm{m}}_{\mathrm{com}} &= (1 - f_{\mathrm{lin}}) \cdot \mathrm{Exp}(m|(M_{\mathrm{max}} + M_{\mathrm{min}})/2, 1/\lambda_{\mathrm{com}}) \\ &+ \quad f_{\mathrm{lin}} \cdot \mathrm{Lin}(m|0),\end{aligned} \tag{4.14}$$

where the simple exponential is centered in the middle of the fitted mass range and is parameterized in terms of shape λ. The general linear form is simplified to have zero slope. For the B_s sample, the nominal model uses only the exponential component.

The model parameters and normalization are fitted in data simultaneously with free parameters of the signal, as discussed in Section 4.4. This parameterization is shown to fit the data well, using the assumption that the form of the upper sideband mass is completely applicable to the low-mass region. The conservative systematic uncertainties which are explored in Appendices A and B support that the ability to resolve flavor oscillations is insensitive to this assumption.

4.3 Sample Fraction Constraints

The previous section describes the derivation of fixed template shapes to model the partially reconstructed signals and the physics background. While the separate template parameterizations are documented, their overall normalizations and combination in the likelihood remain unspecified. The expected values of these relative normalizations may be inferred from Monte Carlo simulation in combination with branching fractions. Since the mass space is a powerful tool for distinguishing various kinds of signal and background, we wish to leverage as much information as possible in the data mass fit. This section describes how the fractions of the various model components are constrained, based on branching fractions and efficiencies of triggering and reconstruction.

4.3.1 Method

With the exception of combinatorial background, the various signal and background contributions of Section 4.2 share a unifying attribute: each reconstructed candidate originates completely from hadronization of a b quark (or antiquark). Within the non-combinatorial portion of a sample, the fractional contributions of these components sum to 1. If the samples were composed of an unbiased selection of these b components – that is, if the selection algorithms were totally ineffective – then their expected fractional contributions could be calculated directly from b fragmentation fractions and B branching ratios. However, the sequences of vertex reconstruction and candidate selection are each specifically designed to isolate one channel, the fully reconstructed signal peak, and the other components are selected alongside it according to their similarity with that main mode. Each model component therefore has a different efficiency of triggering and reconstruction which must be combined with the underlying rates of production to produce fraction estimates.

Reconstruction Efficiencies

The method of deriving efficiencies is based on Monte Carlo simulation. The same MC samples are used as in the prior template derivations: dedicated simulation for the signal components and semi-generic simulation for the collective physics background. Each MC sample is expected to contain a certain amount of its generated decay channel(s), based on the initial number of b events and its configuration of branching fractions. The number of these decays that are finally selected after simulation of the detector, the trigger, and vertex reconstruction defines an absolute efficiency ϵ. Given the numerous stages involved in selection of a candidate, it may be anticipated that these absolute probabilities are well under 100%. Table 4.8 lists ϵ values for B_s and B^0 signal components. At the 0.1% level, the dedicated MC samples yield a minimum of several thousand events for each channel, yielding relative statistical uncertainties of $\sigma_\epsilon/\epsilon \lesssim 10^{-2}$. The calculation of the efficiencies uses the wide mass boundaries of each B species in the Section 4.4 data fits, modulo the few MeV/c^2 correction for the mass scale of MC simulation.

Channel	B_s	B^0
$B \to D\pi$	0.297	0.301
$B \to D^*\pi/\mathrm{K}$	0.259	0.189
$B \to D\rho$	0.110	0.081

Table 4.8: Reconstruction efficiencies ϵ [%] for the signal channels of the B_s and B^0 samples optimized for selection of the $B \to D\pi$ topology. The efficiencies are calculated from counting of MC events within the appropriate wide mass ranges. Statistical uncertainties are at the 10^{-4}% level.

The collective treatment of the physics background presents a slightly more involved case, in that the counting of its semi-generic MC events is affected by both the underlying fractions with which the channels are generated and the efficiency for selecting each one. The ratio of accepted and generated events for this component yields the product $\epsilon * \mathcal{B}$ for the collective physics background.

Branching Fractions and SU(3) Symmetry

Branching fractions are used in combination with the efficiencies above to calculate expected values for the fractions of each sample component. Using the world average branching measurements from the Particle Data Group (PDG) [2], the expected fraction f_i of any component may be calculated as

$$f_i = \frac{\epsilon_i * \mathcal{B}_i^{\mathrm{PDG}}}{(\sum_j \epsilon_j * \mathcal{B}_j^{\mathrm{PDG}}) + (\epsilon * \mathcal{B})_{DX}^{\mathrm{MC}}}, \tag{4.15}$$

where the sum over j includes all signal components and the term $(\epsilon * \mathcal{B})_{DX}^{\mathrm{MC}}$ refers to the combined efficiency and branching fractions of the collective physics background

from semi-generic simulation. Because these sample components are all assumed to contain the appropriate D meson species, the branching fractions of the D decay chain cancel out. However, the relevant D^* decay fraction(s) must be factored in to each $B \to D^*\pi$ channel. As in the calculation of reconstruction efficiencies, the $B \to D^*K$ fractions are combined with $B \to D^*\pi$. The sum over j contains an extra term in the B^+ sample fractions due to the $B^0 \to D^{*-}\pi^+$ content. For all samples, the uncertainties on all branching measurements $\mathcal{B}$ are readily propagated to produce uncertainties σ_{f_i}.

In the fraction calculations, the contributions from Cabibbo-suppressed $B \to DK$ decays and reflections of Λ_b and B^0 or B_s are factored directly into the $B \to D\pi$ component. Their normalizations r relative to the main peak, which are listed in Section 4.2, are also derived using world average branching fraction information and reconstruction efficiencies from MC simulation. This is a simplified two-component version of the overall calculation of Equation (4.15). Effectively, the $B \to D\pi$ branching fraction is modified in the equation by

$$\mathcal{B}^{\rm PDG}_{B\to D\pi} \to \mathcal{B}^{\rm PDG}_{B\to D\pi} \cdot (1 + r_{\rm Cab} + r_{\Lambda_b {\rm ref}} + r_{B_q {\rm ref}}).$$

An important decision for this analysis is that of what B_s branching fractions to use. While the decays of the light B species have been well-explored by BABAR [15] and Belle [16], the lack of high-statistics B_s samples has left a general unavailability of precision B_s branching measurements. Specifically, no measurements of the exclusively reconstructed $B_s^0 \to D_s^{*-}\pi^+$ and $B_s^0 \to D_s^-\rho^+$ channels have yet been published. In general, reasonable branching fraction estimates may be made by assuming SU(3) symmetry under $d \leftrightarrow s$ quark exchange. Therefore, the calculation of expected B_s sample fractions uses the world average B^0 branching ratios. This assumption is part of the motivation for evaluating conservative systematic uncertainties for the relative contributions between signal components and between signal and background components. Given that the resulting bias to flavor oscillation parameters is shown to be small in Section 7.2.3 and Appendix A, this decision is allowed to stand.

4.3.2 Gaussian Constraints

The expected sample fractions and their uncertainties are used to constrain the associated parameters of the likelihood. Neglecting the combinatorial background, the normalized PDF of the mass space is expressed as

$$\begin{aligned}\mathcal{P}^{\rm m}_{b-{\rm hadrons}} = {} & f_{D\pi} \cdot \frac{1}{1 + r_{\rm Cab} + r_{\Lambda_b {\rm ref}} + r_{B_q {\rm ref}}} \left[\mathcal{P}^{\rm m}_{D\pi} + r_{\rm Cab}\mathcal{P}^{\rm m}_{\rm Cab} + r_{B_q {\rm ref}}\mathcal{P}^{\rm m}_{B_q {\rm ref}} + r_{\Lambda_b {\rm ref}}\mathcal{P}^{\rm m}_{\Lambda_b {\rm ref}}\right] \\ & + f_{D^*\pi} \cdot \mathcal{P}^{\rm m}_{D^*\pi} \\ & + f_{D\rho} \cdot \mathcal{P}^{\rm m}_{D\rho} \\ & + (1 - f_{D\pi} - f_{D^*\pi} - f_{D\rho}) \cdot \mathcal{P}^{\rm m}_{DX}, \end{aligned} \tag{4.16}$$

where the fractions f_i are the constrained parameters and the components $\mathcal{P}$ are introduced in Section 4.2. The small Cabibbo-suppressed and reflection components

are included as fixed ratios of the main $D\pi$ peak, as noted previously. The model for the B^+ sample contains an extra term for its $B^0 \to D^{*-}\pi^+$ component.

Partitioning the sample into N components requires $N-1$ fraction parameters, since the fractions are constrained to sum to unity. The Gaussian constraints are applied to the f_i fractions of the signal components, and their uncertainties reflect the uncertainty on physics background content as well.

The constraints are implemented as χ^2-style terms added to the total likelihood of the sample. Given the set of fraction constraints $\{f_i^0\}$ and their uncertainties $\{\sigma_{f_i}^0\}$, the likelihood may be expressed as a function of the floating fraction parameters $\{f_i\}$:

$$-2\ln\mathcal{L}_{\rm constr}(\{f_i\}) \to -2\sum_{\rm events}\ln\mathcal{P}_{\rm evt} + \sum_i\left(\frac{f_i - f_i^0}{\sigma_{f_i}^0}\right)^2. \qquad (4.17)$$

The constraints serve to "pull" the log-likelihood minimum toward the expected fraction values in the parameter space. Table 4.9 lists the constraints and their uncertainties calculated via Equation (4.15) for the B_s and B^0 sample. Table 4.10 lists the constraints for the B^+ sample.

Channel	B_s	B^0
$B \to D\pi/K+$ refl.	28.3 ± 2.3	43.0 ± 3.1
$B \to D^*\pi/K$	24.5 ± 1.9	8.6 ± 0.8
$B \to D\rho$	26.9 ± 3.4	29.5 ± 3.7
$b \to DX$	20.4 ± —	18.9 ± —

Table 4.9: Fraction constraints in % applied to the non-combinatorial background portion of the B_s and B^0 samples, as derived from the Table 4.8 reconstruction efficiencies, and world average branching fractions (B^0), and SU(3) symmetry (B_s).

Channel	B^+
$B^+ \to \overline{D}^0\pi^+/K$	23.6 ± 1.3
$B^+ \to \overline{D}^{*0}\pi^+/K$	19.1 ± 1.4
$B^+ \to \overline{D}^0\rho^+$	21.0 ± 2.3
$B^0 \to D^{*-}\pi^+/K$	6.0 ± 0.5
$b \to \overline{D}^0X$	29.8 ± —

Table 4.10: Fraction constraints in % applied to the non-combinatorial background portion of the B^+ sample, as derived from reconstruction efficiencies and world average branching fractions.

4.4 Mass Fits in Data

Previous sections derive templates and their expected relative normalizations for partially reconstructed and misreconstructed signal and background components of the samples. This section combines these models in fits to the data and describes how information is extracted for use in more advances stages of the analysis.

4.4.1 Complete Mass Likelihood

The complete likelihood combines the parameterization for the combinatorial background spectrum with the constrained model for the "b-hadrons" signal and background contributions. Because the combinatorial background level is easily measured in the upper mass sideband, its overall fraction parameter is not constrained. The normalized per-event PDF for the mass space is expressed as

$$\begin{aligned}
\mathcal{P}_{\rm evt} &= (1 - f_{\rm com}) \cdot \mathcal{P}^{\rm m}_{b-{\rm hadrons}} \\
&+ \quad f_{\rm com} \cdot \mathcal{P}^{\rm m}_{\rm com} \\
&= (1 - f_{\rm com}) \cdot \Big[\; f_{D\pi} \cdot \frac{\mathcal{P}^{\rm m}_{D\pi} + r_{\rm Cab}\mathcal{P}^{\rm m}_{\rm Cab} + r_{B_q{\rm ref}}\mathcal{P}^{\rm m}_{B_q{\rm ref}} + r_{\Lambda_b{\rm ref}}\mathcal{P}^{\rm m}_{\Lambda_b{\rm ref}}}{1 + r_{\rm Cab} + r_{\Lambda_b{\rm ref}} + r_{B_q{\rm ref}}} \\
&\qquad\qquad + f_{D^*\pi} \cdot \mathcal{P}^{\rm m}_{D^*\pi} \\
&\qquad\qquad + f_{D\rho} \cdot \mathcal{P}^{\rm m}_{D\rho} \\
&\qquad\qquad + (1 - f_{D\pi} - f_{D^*\pi} - f_{D\rho}) \cdot \mathcal{P}^{\rm m}_{DX} \Big] \\
&+ \quad f_{\rm com} \cdot \mathcal{P}^{\rm m}_{\rm com},
\end{aligned} \tag{4.18}$$

where all individual components $\mathcal{P}$ are documented in Section 4.2. The mass spectra of components for which the data provides a ready handle float freely. These are the fully reconstructed $B \to D\pi$ peak and the combinatorial background. For the set of individual components which overlap strongly in mass across their entire range, the template shapes extracted from Monte Carlo simulation are fixed. This set is comprised of the partially reconstructed signal, the partially reconstructed physics background, and the small contributions of misreconstructed Cabibbo-suppressed and reflection decays. Furthermore, the offsets of these templates with respect to the fully reconstructed peak are also extracted from simulation and fixed. Effectively, the entire $\mathcal{P}^{\rm m}_{b-{\rm hadrons}}$ component is a semi-fixed spectrum whose position in mass is "anchored" by the data position of the main $B \to D\pi$ peak. The only free parameters within this super-component are the Gaussian constrained signal fractions and the parameters of the fully reconstructed peak. Table 4.11 summarizes the floating parameters of the likelihood for fits of the mass space.

Wide and Narrow Mass Bounds

Fits of the mass space alone are performed in a wider mass range than the fits that include proper time information. This procedure accepts a much larger fraction of

Group/notes	Parameters
Signal fractions – constrained	 $f_{D\pi}$, $f_{D^*\pi}$, $f_{D\rho}$
Fully reconstructed peak – anchors translation of fixed templates	 M , σ_1 , f_1 , r_2
Combinatorial background – extrapolated from upper sideband	 $f_{\rm com}$, $f_{\rm lin}$, $\lambda_{\rm com}$

Table 4.11: Free parameters of the fit to mass data. The B_s sample has $f_{\rm lin} \equiv 0$.

physics background than is allowed in for the final analysis. The motivation for this choice is to extract a maximal amount of sample composition information from the data mass space. Usage of a lower minimum mass provides the likelihood with a longer "lever arm" on the fixed templates in the partially reconstructed region, such that it may more easily resolve their relative fractions. Fits of proper time use mass bounds which are tightened to those derived in Section 4.1, maintaining most of the partially reconstructed signal content while excluding many background events. For proper time and flavor asymmetry fits, the normalizations of the signal and physics background components are adjusted and fixed based on the values extracted from the mass space. This treatment is discussed further in Section 4.4.3. Table 4.12 lists the mass bounds for the fits of mass alone and for fits including proper time.

Sample	Mass only	Proper time
B_s	[4.7 , 6.5]	[5.0 , 6.0]
B^0	[4.6 , 6.0]	[4.95 , 5.6]
B^+	[4.6 , 6.0]	[4.95 , 5.6]

Table 4.12: Mass bounds [GeV/c^2] of the data fits in the mass space alone and in fits including proper decay time.

Fitter Validation

Even at the stage in which only the mass space of the sample is modeled, the likelihood has considerable complexity. To justify confidence in measurements extracted with the fitter, it is essential to demonstrated that it produces reliable, unbiased results.

Throughout this analysis, validation of the fitter is performed with large ensembles of "toy" Monte Carlo samples. These samples are generated from parent distributions that exactly reflect the composition and structure of the nominal likelihood model. This method is in contrast to the realistic MC samples used thus far, which simulates the full interaction of particles with the detector and the trigger and reconstruction logic. The toy samples neglect all such underlying effects and reproduce the data at only the highest level. In the parent distributions, representative parameter values are used for all components and their normalizations. If the generated toy sample is

large enough, an unbiased likelihood should be minimized by a parameter set that is consistent with the generating set, within statistical fluctuations. From an ensemble of $\sim \mathcal{O}(1000)$ toy MC samples, each fitted parameter compiles a "pull" distribution of its Gaussian deviations from the generating value. An unbiased fitter produces unit-Gaussian pulls.

Studies with toy MC ensembles are performed at each level of increased likelihood complexity to verify that the fitter remains unbiased and able to resolve key parameters. Appendix D documents representative pull distributions from several key stages of the analysis, including the complete mass likelihood. In development of this mass model, the individual fits of all components and their sequential combinations have been validated as unbiased.

4.4.2 Results

Using the fixed templates for the partially reconstructed signals in the validated fitter, the complete likelihood of the mass space is applied to the data. In preparation of this analysis, the high-statistics light B samples from 355 pb^{-1} were examined first to evaluate the adequacy of the model. Figure 4-10 shows the fitted samples of 60621 B^0 candidates and 70327 B^+ candidates in the mass range $[4.6, 6.0]$ GeV/c^2. Figure 4-11 shows the fitted samples of 12535 B_s candidates from 1 fb^{-1} in the mass range $[4.7, 6.5]$ GeV/c^2.

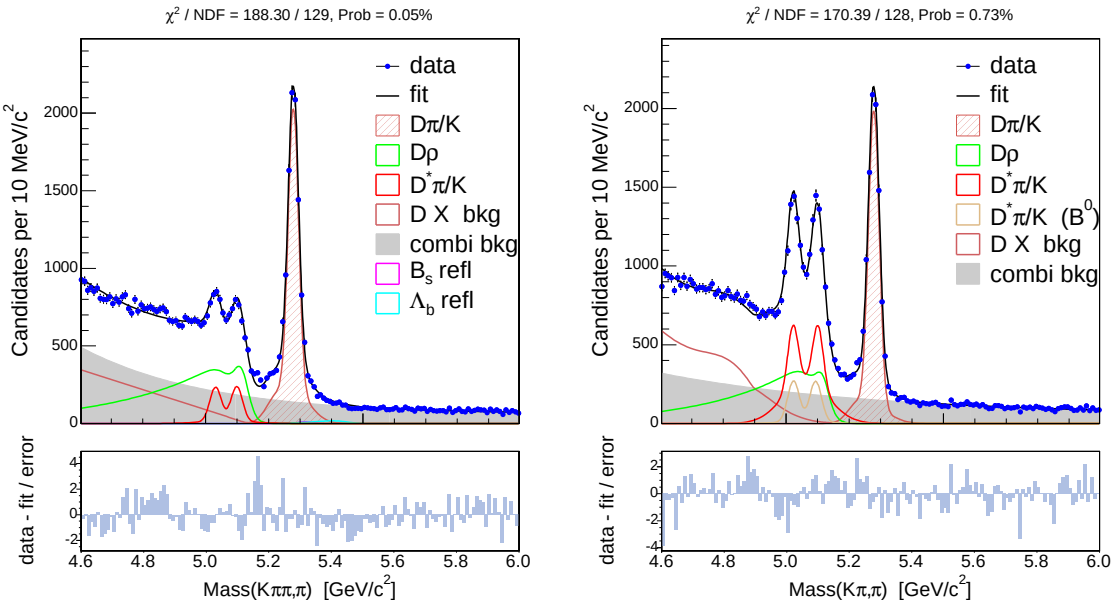

Figure 4-10: Fitted mass spectra of B^0 (left) and B^+ (right) candidates in 355 pb^{-1}. The "wire-frame" drawing style is chosen to clearly depict the individual component shapes.

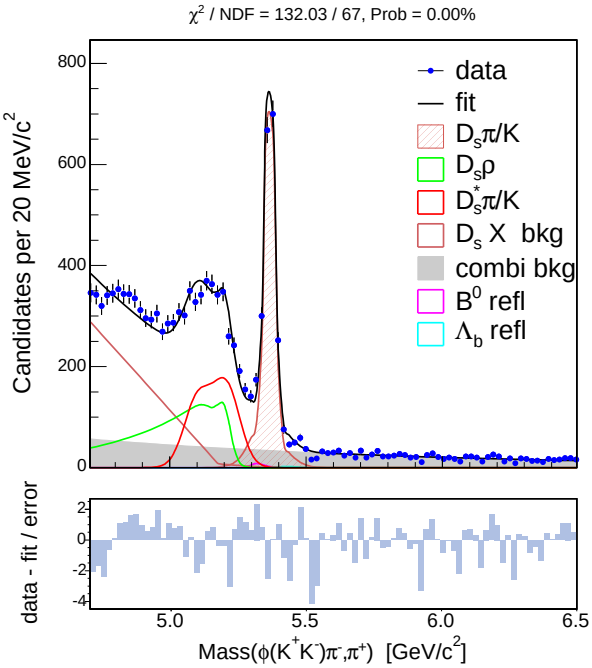

Figure 4-11: Fitted mass spectrum of B_s candidates in 1 fb^{-1}. The "wire-frame" drawing style is chosen to clearly depict the individual component shapes.

Tables 4.13 and 4.14 compile the fitted parameter values of the mass likelihood. A comparison of the constraint values and fitted values of the signal fractions is made below. Because the fits of proper time and flavor use only the data within the narrower mass ranges of Table 4.12, discussion of final signal yields is also withheld until Section 4.4.3.

The "wire-frame"-style figures clearly show the combination of the templates in modeling the data, including the relative contributions attributed to the partially reconstructed signals and physics background. The presence of the additional $B^0 \to D^{*-}\pi^+$ component is visible in the B^+ sample. The small Cabibbo-suppressed $B \to DK$ component is combined within the low-mass edge of each fully reconstructed peak, and the very small contributions of the Λ_b and B^0/B_s reflections are drawn separately in the same vicinity.

The resulting χ^2 probabilities are at the 0.1% level for the light B samples, with residuals distributed relatively evenly across the mass range. The large B^0 and B^+ sample sizes challenge the highly structured model, but not overwhelmingly in any single component. The B_s fit quality is reasonable throughout the narrower range of the proper time fit. However, the overall B_s fit quality is damaged by possible structure in the low-mass region of the $b \to D_s^- X$ physics background, for which the wedge-shaped template may be a simplification. As noted previously, the fixed templates of Section 4.2 are derived from Monte Carlo simulation which is tuned for the exact trigger mixtures of only the first 355 pb^{-1} of data. As a rudimentary check

Parameter		Fitted Value
M_{B^0}	[MeV/c^2]	5279.0 ± 0.2
σ_1	[MeV/c^2]	16.3 ± 0.4
f_1		0.261 ± 0.023
r_2		3.07 ± 0.26
$f_{B^0\to D^-\pi^+}$		0.313 ± 0.009
$f_{B^0\to D^-\rho^+}$		0.356 ± 0.017
$f_{B^0\to D^{*-}\pi^+}$		0.067 ± 0.004
$f_{\rm com}$		0.416 ± 0.023
$\lambda_{\rm com}$	[(GeV/c^2)$^{-1}$]	2.52 ± 0.17
$f_{\rm lin}$		0.333 ± 0.033

Parameter		Fitted Value
M_{B^+}	[MeV/c^2]	5278.9 ± 0.2
σ_1	[MeV/c^2]	17.1 ± 0.4
f_1		0.128 ± 0.024
r_2		3.05 ± 0.687
$f_{B^+\to\overline{D}^0\pi^+}$		0.210 ± 0.004
$f_{B^+\to\overline{D}^0\rho^+}$		0.238 ± 0.010
$f_{B^+\to\overline{D}^{*0}\pi^+}$		0.174 ± 0.008
$f_{B^0\to D^{*-}\pi^+}$		0.057 ± 0.004
$f_{\rm com}$		0.331 ± 0.019
$\lambda_{\rm com}$	[(GeV/c^2)$^{-1}$]	1.45 ± 0.29
$f_{\rm lin}$		0.302 ± 0.075

Table 4.13: Fitted parameter values from the mass space of 60621 B^0 (left) and 70327 B^+ (right) candidates in the range $[4.6, 6.0]$ GeV/c^2 from 355 pb^{-1} of data, as visualized in Figure 4-10.

Parameter		Fitted Value
M_{B_s}	[MeV/c^2]	5366.9 ± 0.6
σ_1	[MeV/c^2]	17.3 ± 0.7
f_1		0.220 ± 0.037
r_2		3.50 ± 0.42
$f_{B_s^0\to D_s^-\pi^+}$		0.221 ± 0.005
$f_{B_s^0\to D_s^-\rho^+}$		0.228 ± 0.017
$f_{B_s^0\to D_s^{*-}\pi^+}$		0.185 ± 0.010
$f_{\rm com}$		0.217 ± 0.016
$\lambda_{\rm com}$	[(GeV/c^2)$^{-1}$]	0.788 ± 0.118
$f_{\rm lin}$		0.000 ± (const)

Table 4.14: Fitted parameter values from the mass space of the 12535 B_s candidates in the range $[4.7, 6.5]$ GeV/c^2 from the full 1 fb^{-1} of data, as visualized in Figure 4-11.

of the applicability of the templates to the full 1 fb^{-1}, the B_s mass fits are examined for each of three data-taking periods which sum to the full dataset. The separate mass spectra of these B_s subsamples in Figure 4-12 show that the fit quality is indeed better for the first dataset than the later periods – a shift from 9.3% probability to 0.2% probabilities. It is likely that the physics background content of the later data is somewhat different than the simple wedge model applied here, resulting in this component being slightly misweighted in the mass likelihood of the overall data.

The results of these fits are relevant only to the extent that they determine the signal fractions within the narrow mass bounds of the final analysis. Specifically, the physics background content of the B_s sample above 5.0 GeV/c^2 is modeled to be only 3%. The comprehensive philosophy in evaluation of all fit results is that possible

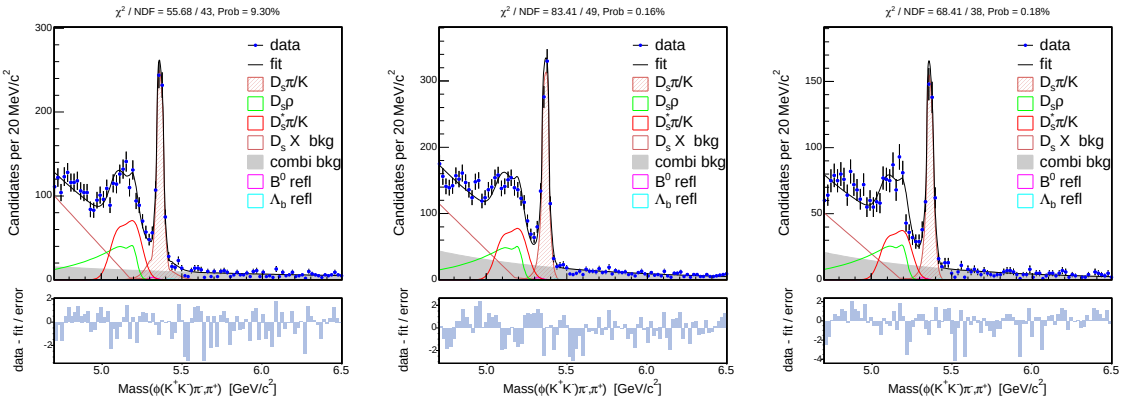

Figure 4-12: Fitted mass spectra of B_s candidates in three data-taking periods, totaling to the 1 fb^{-1} shown in Figure 4-11. Identical templates are applied to each B_s dataset.

deficiencies in modeling of relative component fractions must be accounted for by *conservative systematic uncertainty estimates*. These pessimistic estimates cover a wide range of mismodeling possibilities, including incorrect template shapes from skewed trigger mixtures, badly fitted signal fractions, and inadequate background parameterizations. Such studies make it possible to avoid the protracted development of flawless mass models if it is ultimately irrelevant to the flavor oscillation parameters. The discussion of Section 7.2.3 shows that large variations in modeling decisions result in only very small biases to the observed value of Δm_s.

Constraints vs. Fitted Signal Fractions

The Gaussian constraint values of the b hadron signal and background fractions are compared with their final fitted values in Tables 4.15 - 4.17. Each table shows the residual fraction assigned to the physics background, whose uncertainty is subsumed within those of the signal fractions. Minuit [32] extracts uncertainties on the fitted parameters by the covariance matrix evaluation with the modified likelihood function of Equation (4.17).

Component	Constraint	Fit
$B^0 \to D^-\pi^+/K^+$ + refl.	43.0 ± 3.1	31.3 ± 0.9
$B^0 \to D^-\rho^+$	29.5 ± 3.7	35.6 ± 1.7
$B^0 \to D^{*-}\pi^+/K^+$	8.6 ± 0.8	6.7 ± 0.4
$b \to D^- X$	$18.9 \pm$ —	$26.4 \pm$ —

Table 4.15: Comparison of constraints and fitted values for b hadron signals and background in the B^0 data. The constraints are derived from world average branching fractions and Monte Carlo simulation efficiencies.

Component	Constraint	Fit
$B^+ \to \overline{D}^0\pi^+/K^+$	23.6 ± 1.3	19.7 ± 0.2
$B^+ \to \overline{D}^0\rho^+$	21.0 ± 2.3	21.6 ± 0.6
$B^+ \to \overline{D}^{*0}\pi^+/K^+$	19.1 ± 1.4	17.4 ± 0.5
$B^0 \to D^{*-}\pi^+/K^+$	6.0 ± 0.5	5.9 ± 0.4
$b \to \overline{D}^0 X$	$29.8 \pm$ —	$35.4 \pm$ —

Table 4.16: Comparison of constraints and fitted values for b hadron signals and background in the B^+ data. The constraints are derived from world average branching fractions and Monte Carlo simulation efficiencies.

Component	Constraint	Fit
$B_s^0 \to D_s^-\pi^+/K^+ +$ refl.	28.3 ± 2.3	22.1 ± 0.5
$B_s^0 \to D_s^-\rho^+$	26.9 ± 3.4	22.8 ± 1.7
$B_s^0 \to D_s^{*-}\pi^+/K^+$	24.5 ± 1.9	18.5 ± 1.0
$b \to D_s^- X$	$20.4 \pm$ —	$36.6 \pm$ —

Table 4.17: Comparison of constraints and fitted values for b hadron signals and background in 1 fb^{-1} of B_s data. The constraints are derived from the assumption of SU(3) symmetry with world average branching fractions from B^0, in combination with Monte Carlo simulation efficiencies.

The expected fractions do not agree well with the data in any sample. In general, the data pulls the fractions strongly from their constraints, which may be due to the simple model used for the physics background template. Although this analysis is emphatically *not* intended as a branching fraction measurement, the 1 fb^{-1} of CDF data does appear to have the statistical ability to support one. Regardless, this analysis is firmly oriented toward observation of B_s mixing. Ultimately, questions about sample fractions fall under the same umbrella of systematic uncertainty studies noted above. Since the physical parameters of flavor oscillation are shown to be robust to large variations in the partially reconstructed model content, the input of the expected fractions is retained in the nominal model.

4.4.3 Extraction of Sample Fractions

A primary purpose of the fit in the wide mass space is to extract and propagate information of sample composition. In order to apply this information to more complex forms of the likelihood, the fitted signal fractions are *fixed* for the remainder of the analysis. However, these fractions must be adjusted to reflect the fact that the fits are performed on the data within the narrower mass ranges of Table 4.12. The rescaling of signal fractions is efficiently implemented within the machinery of the likelihood framework, such that adjustments to the model in any space are accounted for "on the fly".

Examination of the fitted mass spectra of Figures 4-10 and 4-11 reveals several obvious effects of these adjustments in the new narrow ranges. The relevance of the "DX" physics background template is drastically reduced within the 5.0 (4.95) GeV/c^2 cut-off for B_s (B^0, B^+). Additionally, much of the long tail of the $D\rho$ component is truncated, leaving only its main shoulder and a short tail. However, the well-defined $D^*\pi$ component is almost fully contained within the new bounds, increasing its relative contribution. Similarly, the fully reconstructed $D\pi$ peak is unaffected by the narrowed window. Overall, this situation merely reflects the choices made in Section 4.1, in which the selection of the mass bounds for the final analysis is designed to accept a large portion of the new partially reconstructed signal while avoiding all but a small amount of new background content.

Signal Yields

Signal yields for the final analysis samples are extracted from the fitted mass model within the narrow ranges of Table 4.12. Tables 4.18 and 4.19 compile the yields for each component, including fully correlated statistical uncertainties. While the fraction of physics background is implied by the sum of the other fractions and therefore has no explicit uncertainty, its associated uncertainty arises from full propagation of the signal yield uncertainties. In these tables, the Cabibbo-suppressed $B \to DK$ component is not included in the fully reconstructed $B \to D\pi$ yield.

Component	Yield
$B^0 \to D^-\pi^+$	10148 ± 179
$B^0 \to D^-\rho^+$	6568 ± 226
$B^0 \to D^{*-}\pi^+/K^+$	2332 ± 153
$b \to D^-X$	1164 ± 154

Component	Yield
$B^+ \to \overline{D}^0\pi^+$	9331 ± 192
$B^+ \to \overline{D}^0\rho^+$	6023 ± 259
$B^+ \to \overline{D}^{*0}\pi^+/K^+$	7874 ± 322
$B^0 \to D^{*-}\pi^+/K^+$	2641 ± 185
$b \to \overline{D}^0X$	1020 ± 68

Table 4.18: Yields of signal and physics background components in 355 pb^{-1} of B^0 (left) and B^+ (right) data in the mass range [4.95, 5.6] GeV/c^2. The statistical uncertainties include full correlations.

Component	Yield
$B_s^0 \to D_s^-\pi^+$	2004 ± 57
$B_s^0 \to D_s^-\rho^+$	1350 ± 103
$B_s^0 \to D_s^{*-}\pi^+/K^+$	1795 ± 96
$b \to D_s^-X$	394 ± 17

Table 4.19: Yields of B_s signal and physics background in 1 fb^{-1} of data in the mass range [5.0, 6.0] GeV/c^2. The statistical uncertainties include full correlations.

As was hoped for in motivating the inclusion of the partially reconstructed candidate region, use of the new signal channels more than doubles the overall signal yield

of the B_s sample. However, an equally relevant consideration is that of the statistical power per signal event, which is quantified in terms of the oscillation amplitude uncertainty $1/\sigma_{\mathcal{A}}$ in Equation (1.32). The background fraction of this new region is somewhat higher, and given that $1/\sigma_{\mathcal{A}} \propto \sqrt{\frac{\mathcal{S}}{\mathcal{S}+\mathcal{B}}}$, the added signal significance is slightly reduced. Moreover, the added significance is heavily dependent on the proper time and flavor tagging of the new signal events. The following chapters are devoted to inclusion of these attributes in the likelihood model. A comprehensive estimate of the B_s mixing significance due to partially reconstructed signal events is discussed in Chapter 7.

Chapter 5

Proper Decay Time

Measurement of proper decay time is one of the few indispensable ingredients for observation of flavor oscillations. Whereas modeling of the mass space is important for its ability to distinguish signal and background candidates, the proper time model is the framework from which the mixing signature is eventually extracted. The precision of proper time measurement is the most critical critical factor in the resulting oscillation significance, and the time resolution associated with the new partially reconstructed signal events determines their relative contribution. This chapter introduces the method of proper time measurement, the important calibration of its uncertainty, the expansion of the likelihood into proper time space, and the application of the extended fitter to the data samples.

5.1 Measurement of Proper Time

The proper time t between the production and decay of each B candidate is inferred from observable quantities in the laboratory frame. It is reconstructed via the expression

$$ct = \frac{L_{\rm xy} \cdot M_B}{p_T}, \tag{5.1}$$

where p_T is the transverse B momentum and $L_{\rm xy}$ is the transverse displacement of the B decay vertex from the $p\overline{p}$ collision. Because the long-lived B mesons have negligible resonance widths, the world average masses M_B [2] are used. Although the quantity ct is universally referred to as proper time, it is more accurately labelled proper decay *length*, and its units reflect this additional factor c. The variables t and ct are used interchangeably in this chapter to suit the context.

The uncertainty on measurement of this quantity is a fundamental consideration of all B_s mixing analyses. Equation (1.32) introduces that the statistical significance of the oscillation amplitude goes as $1/\sigma_{\mathcal{A}} \propto \exp(-\frac{\Delta m^2 \sigma_t^2}{2})$, with an exponential loss due to poor time resolution. As derived from Equation (5.1), the uncertainty on

measurement of ct takes the form

$$\sigma_{ct} = \left(\sigma_{L_{\rm xy}} \cdot \frac{M_B}{p_T}\right) \oplus \left(ct \cdot \frac{\sigma_{p_T}}{p_T}\right), \tag{5.2}$$

where $\oplus$ implies that the terms are added in quadrature and the uncertainty on M_B is neglected. Therefore, the proper time resolution for each candidate receives two conceptually simple contributions: the first term arises from uncertainty on the spatial position of decay vertices, while the second term is due to the relative uncertainty on the momentum of the complete B decay. Because of the importance of achieving good σ_{ct} in this analysis, both of these underlying contributions are treated with care.

$L_{\rm xy}$ Measurement

The transverse decay length $L_{\rm xy}$ is defined as the distance between the primary vertex position $\vec{r}_{\rm PV}$ of the $p\overline{p}$ collision and the secondary vertex position $\vec{r}_{\rm SV}$ of the subsequent B decay, as projected onto the B p_T direction:

$$L_{\rm xy} = (\vec{r}_{\rm SV} - \vec{r}_{\rm PV}) \cdot \hat{p_T}. \tag{5.3}$$

As such, its uncertainty $\sigma_{L_{\rm xy}}$ comes from the uncertainty on the spatial positions of both the primary and secondary vertices.

The contribution of the primary vertex uncertainty is minimized by use of the "event-by-event" PV whose reconstruction is detailed in Section 3.2.2. The spatial uncertainty on the secondary B decay vertex depends on various characteristics of the decay, including its kinematics and fit quality. Although the fitted vertex is already the best available from existing software, it is important that the resulting uncertainty for each candidate be estimated as accurately as possible. Section 5.1.1 below describes the procedure for calibration of the overall $L_{\rm xy}$ uncertainty using representative samples of secondary decay vertices. The average proper time uncertainty for fully reconstructed decays is about $\sigma_{ct} \sim 29$ μm.

Momentum Uncertainty

In analysis of fully reconstructed decay channels, uncertainty on candidate momentum arises only from measurement uncertainty on the momenta of constituent tracks. With the excellent tracking system of the CDF detector, this $\sigma_{p_T}/p_T \approx 0.0015\, p_T/({\rm GeV}/c)$ contribution is negligible. This case underscores the importance of fully reconstructed decays to B_s mixing analyses: the proper time uncertainty, on which the oscillation signature strongly depends, receives essentially no contribution from the second term of Equation (5.2). The precision of vertex positions is solely responsible for σ_{ct}.

In contrast to this simple scenario, partially reconstructed decays involve the unmeasured momentum of one or more particles. The proper time $ct^{\rm reco}$ reconstructed via Equation (5.1) is not the decay time ct^B which would be measured for a fully reconstructed B candidate. The fully reconstructed time is proportional to the observed

value by the relation

$$ct^B = \kappa \cdot ct^{\text{reco}}, \tag{5.4}$$

where the "κ-factor" is defined for each event as

$$\kappa = \frac{p_T^{\text{reco}} L_{\text{xy}}^B}{p_T^B L_{\text{xy}}^{\text{reco}}}. \tag{5.5}$$

Neglecting the difference in L_{xy}, the κ-factor is conceptually understood as the fractional B momentum reconstructed by the detector. Since the lost particle(s) may be distributed across a range of decay angles and phase space, its unmeasured value introduces an uncertainty on the true momentum of the complete B decay. That is, the value of κ for individual events cannot be known and must be treated as a *probability density distribution* for the entire class of events. The use of partially reconstructed decays as signal events requires that the κ-factor distribution $F(\kappa)$ be modeled and incorporated into the proper time space of the likelihood. The more broadly distributed the κ-factor distribution is, the greater relative uncertainty σ_{p_T}/p_T is present for each signal candidate. Furthermore, Equation (5.2) shows that the contribution of momentum uncertainty to the overall proper time uncertainty scales with ct itself, implying that longer-lived events have progressively worse resolution. Therefore, well-defined momentum distributions are essential to the statistical significance of partially reconstructed signal channels. One of the primary assets of the partially reconstructed hadronic channels of this dissertation is that their κ-factor distributions are relatively narrow, maintaining good proper time resolution.

The incorporation of these concepts into the likelihood model and visualization of their effects are discussed in full in Section 5.3 below.

5.1.1 Calibration of Vertex Resolution

Correct estimates of proper time uncertainty are important for extracting the best sensitivity to flavor oscillations from each signal decay. In historical mixing searches, a correct calibration of σ_{ct} is critical for derivation of accurate lower limits on Δm_s. Additionally, even once the required statistical power is available for a mixing observation, a correctly measured oscillation amplitude demands that the proper time uncertainty be well-understood. In this analysis, estimates of spatial resolution returned by the vertex fit are calibrated to produce accurate and unbiased uncertainties.

The need for an overall calibration of vertex resolution stems from the most basic units of tracking. Individual hits in the COT and especially in the silicon layers have complex resolution functions which are approximated as Gaussian. Although the track preparation of Section 3.2.1 makes low-level corrections to the track covariance matrices, the fitted vertex uncertainties remain systematically incorrect. An expedient approach is to make a correction on the estimated uncertainty of the overall L_{xy} displacement between the primary and secondary B vertices.

The general procedure for calibration of uncertainty estimates requires comparison of a fitted vertex and its uncertainty with a known true vertex position. In the case of the primary vertex validation noted in Section 3.2.2, the large number of prompt

tracks from each $p\overline{p}$ collision makes it possible to simply fit the same vertex with two sets of tracks. However, real B mesons are long-lived and the secondary decay vertex of each event occurs at a unique point in space. Since the true vertex position cannot be known, the analysis samples themselves cannot be used for $\sigma_{L_{\rm xy}}$ calibration.

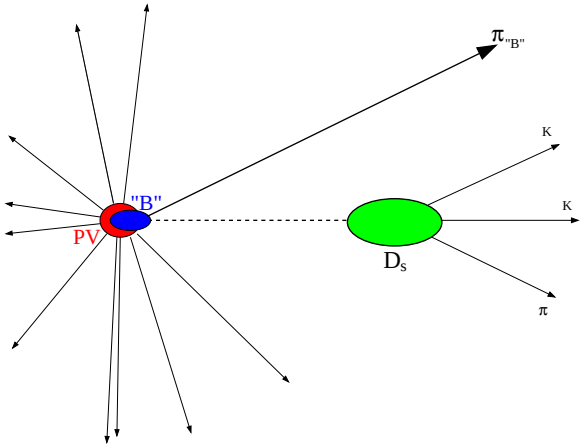

Figure 5-1: Sketch of a D + random track candidate for the $\sigma_{L_{\rm xy}}$ calibration sample, constructed to coincide with the primary vertex.

The solution employed here [95] is to create samples of "pseudo-B" candidate vertices which are known to coincide with the PV and therefore have true value $L_{\rm xy} = ct = 0$. These vertices are formed by selecting *prompt* D mesons and fitting the momentum vector of each with a *random* track. Figure 5-1 shows a cartoon of the vertex topology. Since the real D and the random track are both assumed to originate from the PV, the resulting $L_{\rm xy}$ distribution of the calibration sample should be centered on $L_{\rm xy} = 0$ and have Gaussian width corresponding to the estimated $\sigma_{L_{\rm xy}}$ of each pseudo-B candidate. The sample of vertices is constructed to be kinematically and topologically similar to the real analysis data by requiring that they meet the same selection requirements on p_T, opening angle, and other quantities. Because the neural network selection considers all event variables collectively and relies strongly on B displacement, it is unsuitable for these studies. The requirements of the *rectangular cuts*, excluding those on $L_{\rm xy}(B)$, are used for calibration sample selection.

In hadronic analyses, each B species has a corresponding pseudo-B sample which is made from the appropriate species of prompt D meson and associated selection requirements. Separate calibration samples are constructed for the $B \to D\pi$ decay topologies of this analysis and the $B \to D3\pi$ decays used in other fully reconstructed B_s mixing efforts [83]. Because of the importance of a precise $\sigma_{L_{\rm xy}}$ calibration, the larger statistics of the pseudo-B^0 sample is leveraged by applying that calibration to the B_s data. This choice is justified by the similarity of the B^0 and B_s decay topologies.

Fit Method

The calibration of vertex resolution is implemented by performing a likelihood fit of a sample's prompt proper time distribution with a Gaussian PDF for each candidate:

$$\mathcal{L} = \prod_i \mathcal{G}(ct_i, \sigma_{ct,i} | \delta_{\text{offset}}), \tag{5.6}$$

with

$$\mathcal{G}(ct, \sigma_{ct} | \delta) = \frac{1}{\sigma_{ct}\sqrt{2\pi}} \exp(\frac{-(ct-\delta)^2}{2\sigma_{ct}^2}), \tag{5.7}$$

and the ideal value of δ_{offset} is zero but it is allowed to float. In effect, each candidate has its own resolution function which smears the ideal delta function distribution at the primary vertex. If the estimates of $\sigma_{L_{\text{xy}}}$ (which is directly proportional to σ_{ct} in these "fully reconstructed" candidates) returned by the vertex fit were systematically unbiased, then this simple Gaussian would produce a good fit for the prompt peak. However, because of the low-level tracking considerations noted above, the estimated uncertainties require a correction factor s on each candidate, for which the model is modified via

$$\sigma_{ct} \rightarrow s \cdot \sigma_{ct}. \tag{5.8}$$

The scale factor s may therefore be fitted in the calibration samples as an additional parameter in Equation (5.6) and then *transferred* to the real data samples to correct their spatial resolution values. Typical values of s in CDF measurements range from 1.10 - 1.50, implying that $\sigma_{L_{\text{xy}}}$ is underestimated.

An additional level of complexity is introduced by the fact that the precision of vertex fitting is dependent on many candidate characteristics. These effects are dominated by a few variables, including

- vertex χ^2, for which lower values produce more precise spatial measurements,
- the opening angle between the D meson and random track, for which small angles are associated with poorer resolution, and
- the isolation of the vertex from surrounding tracks, in which more isolated vertices have better measurements.

A full list of quantities is included below. The overall scale factor is a *parameterization* with respect to these vertex quantities. By this method, the transferred scale factor provides a more accurate correction to the uncertainty of each vertex. Rather than a constant average scale factor for the whole sample, the correction for each vertex is customized according to its most relevant attributes. Moreover, the parameterized scale factor accounts for differences in phase space occupied by the real data samples and the calibration samples. While each calibration sample is selected to be as similar to its corresponding data as possible, modest differences are still incurred. The transfer of a correction factor which is a function of candidate attributes therefore does not rely on the two samples being distributed identically. The data and calibration samples are compared to verify that any extrapolation in the parameter space is justified.

Representative Results

The scale factor parameterization which is applied to the B_s data is derived from the pseudo-B^0 (that is, D^- + random track) calibration sample. This correction function is described here, along with representative figures of the fitted parameterization and the phase space of the samples. Additionally, the calibration is extracted for samples from each of three data-taking periods used for internal study. Since the three parameterizations are qualitatively similar, only the sample from the first 355 pb^{-1} is discussed here.

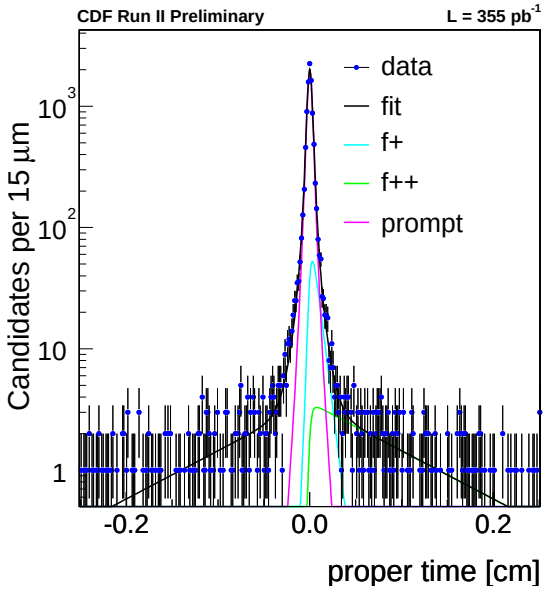

Figure 5-2: Representative proper time distribution of 10,000 events from the D^- + track calibration sample.

Because Tevatron $p\overline{p}$ collisions produce copious $c\overline{c}$ pairs, the number of available prompt D mesons is large. Reconstruction of D^-+ track vertices yields 340K candidates, and the overall scale factor is fitted to be $s = 1.406 \pm 0.003$ (stat). Figure 5-2 shows the proper time fit of a representative sample of 10 000 candidates. The logarithmic scale shows the prominent prompt peak of Equation (5.6), which is augmented by small exponential tails to account for misreconstructed vertices.

The overall calibration sample is subdivided by various event attributes to measure the dependences of the resolution scale factor. Figure 5-3 shows the variation of the fitted correction with respect to the most relevant quantities, including: opening angle ΔR of the pion momentum with respect to the D^-momentum, vertex χ^2, isolation, defined as the fraction of the total track momentum within a cone of half-angle 0.7 about the vertex momentum, pseudorapidity η, beamline-axis position z, and transverse momentum of the prompt pion. In general, parabolic functions are sufficient to parameterize the scale factor dependences.

Given these scale factor variations, the goal of the calibration is to extract their

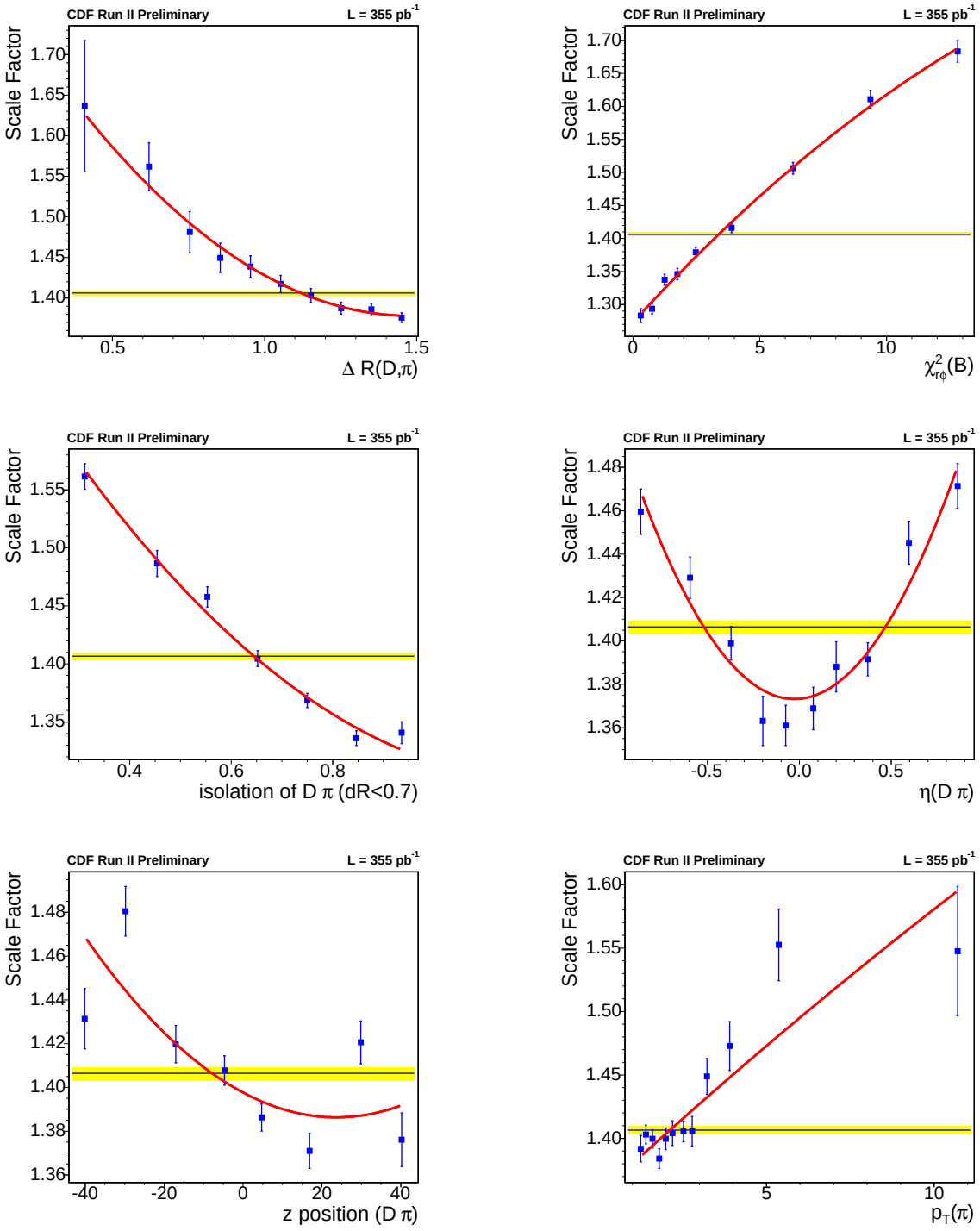

Figure 5-3: Dominant $\sigma_{L_{xy}}$ scale factor dependences in the D^- + track calibration sample. The yellow band indicates the average scale factor value.

functional dependence and apply the resulting parameterization to the data. Modeling of the dependences of Figure 5-3 is performed under the assumption that they

factor completely. That is, scale factor variation with respect to one quantity is first fitted and corrected for in the calibration sample, and then the correction for a second variable is extracted and applied, *etc.* The pattern is continued until deviation from a constant residual scale factor value of 1.00 is observed to be within a reasonable threshold. The iterative procedure for parameterizing these dependences is as follows:

1. **$\Delta R(D,\pi)$ Tuning:** Fit the variation of scale factor versus opening angle $\Delta R(D,\pi)$ with a parabola. Applying this parabolic correction to $\sigma_{L_{\rm xy}}$, all fits are repeated. By construction, the newly fitted scale factor is independent of $\Delta R(D,\pi)$ at this stage.

2. **Isolation Tuning:** Fit the *new* variation of scale factor versus isolation i with a parabola. Applying this second correction to $\sigma_{L_{\rm xy}}$, all fits are repeated. By construction, the scale factor is independent of isolation at this stage.

3. **Pseudorapidity Tuning:** Fit the new variation of scale factor versus pseudorapidity η with a parabola. Applying this third correction to $\sigma_{L_{\rm xy}}$, all fits are repeated. The scale factor is independent of pseudorapidity.

4. **z-Position Tuning:** Fit the new variation of scale factor versus beamline axis position z with a parabola. Applying this fourth correction to $\sigma_{L_{\rm xy}}$, all fits are repeated. The scale factor is independent of z.

5. **χ^2 Tuning:** Fit the new variation of scale factor versus χ^2 with a parabola. Applying this fifth and final correction to $\sigma_{L_{\rm xy}}$, all fits are repeated. The scale factor is independent of χ^2.

Having applied these fully factored corrections to the calibration sample, the subsamples are fitted again for the residual scale factor value and its variations. Figure 5-4 illustrates that the overall scale factor is consistent with 1.0 and that the deviations across the range of the relevant event attributes are within a few percent. The fact that the residual dependences remain small for all variables indicates that the parameterization factors successfully.

A systematic uncertainty on the oscillation amplitude is evaluated to account for possible deficiencies of this method. Appendix A describes how the residual variation of $\pm \sim 4\%$ from unity is applied as a gross shift to the overall scale factor in the data.

The parameterization of the $\sigma_{L_{\rm xy}}$ scale factor which is transferred to the first 355 pb^{-1} of B_s data is of the form

$$\begin{aligned} s \;=\; & \left[p_0^R + p_1^R \Delta R + p_2^R \Delta R^2\right] \\ & \cdot \left[p_0^i + p_1^i i + p_2^i i^2\right] \\ & \cdot \left[p_0^\eta + p_1^\eta \eta + p_2^\eta \eta^2\right] \\ & \cdot \left[p_0^z + p_1^z z + p_2^z z^2\right] \\ & \cdot \left[p_0^\chi + p_1^\chi \chi^2 + p_2^\chi (\chi^2)^2\right], \end{aligned} \tag{5.9}$$

with parameter values listed in Table 5.1. The other B_s data periods receive very similar corrections.

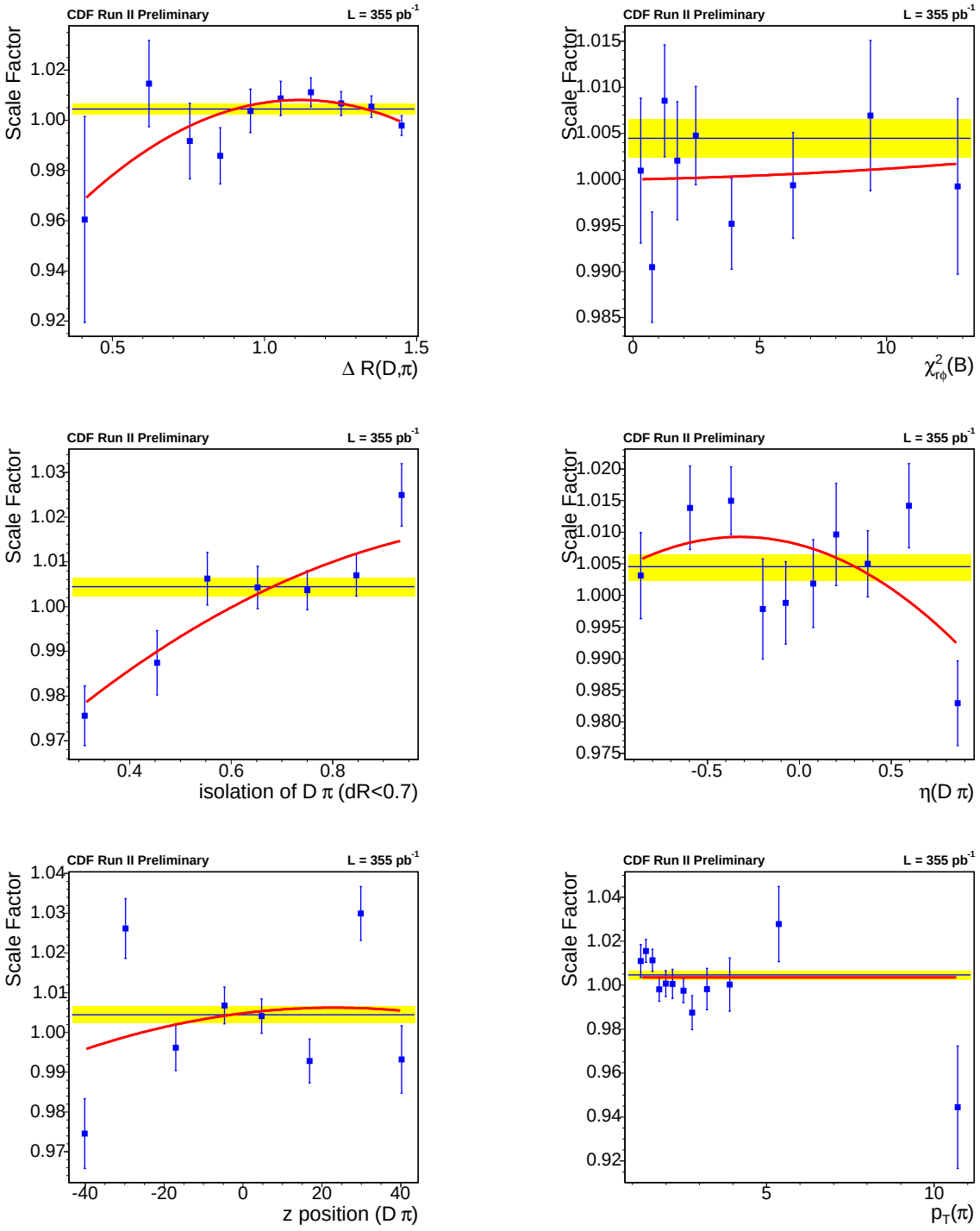

Figure 5-4: Dependences of residual $\sigma_{L_{xy}}$ scale factors in the D^- + track calibration sample after application of the fully factored correction. The overall scale factor is consistent with 1.0 and the variations are acceptable. Note the differences in vertical scale with respect to Figure 5-3.

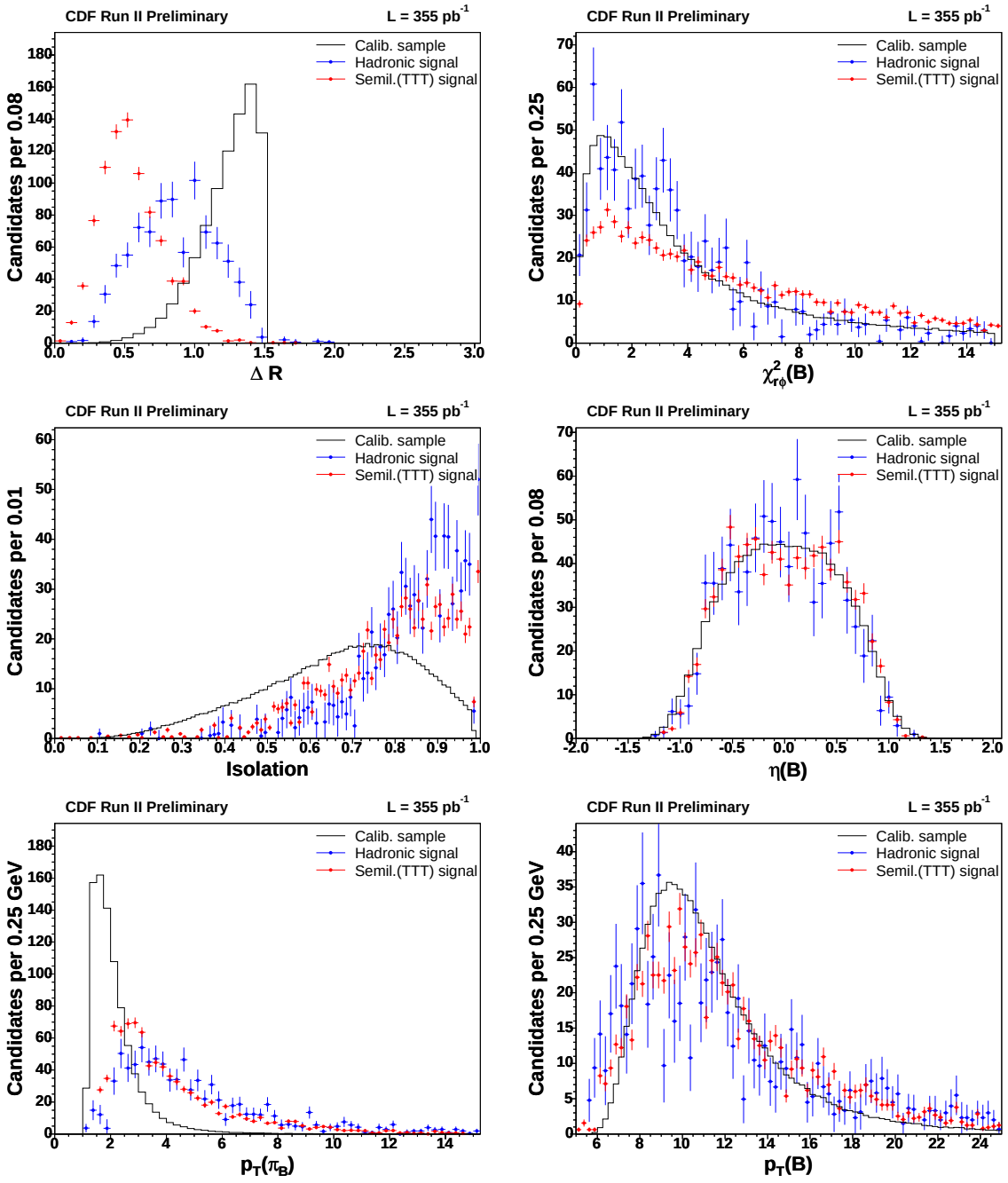

Figure 5-5: Comparison of the D^- + track calibration sample (black line) with sideband-subtracted B_s data distributions. The hadronic data of this analysis is drawn in blue, while the red lines indicate semileptonic B_s data [35], to which this calibration is also applied. The distributions are shown normalized to the scale of the hadronic data, by factor $\sim \frac{1}{500}$ for the calibration sample.

The fact that the scale factor dependences are parameterized implies that the calibration sample and actual data are not required to be identically distributed in

	$\Delta R(D,\pi)$	i	η	z	χ^2
p_0	1.853	1.259	0.9790	0.99469	0.91398
p_1	-0.642	-0.519	0.0046	-0.00062	0.02723
p_2	0.217	-0.200	0.0928	0.00001	-0.00055

Table 5.1: Parameter values of the Equation (5.9) parameterization for application to the first 355 pb^{-1} of B_s data.

phase space. However, in order for the correction to be deemed reliable, it should still be verified that the parameter ranges of the data are also populated by the calibration sample. Figure 5-5 overlays the distributions of the samples, showing that the calibration sample is well-populated in the large majority of the data space. The calibration sample distributions are normalized by a factor of $\sim 1/500$ to match the size of the B_s data, indicating that many events are still present where its content appears small. Therefore, for $p_T(\pi_B)$, the only variable in which the calibration sample does not fully cover the range of data, the necessary extrapolation of the resolution correction is modest.

Vertex Resolution in Data

Figure 5-6 shows the distribution of the vertex resolution component of proper time uncertainty in hadronic B_s data after application of the above scale factor parameterization on an event-by-event basis. The average proper time spatial resolution of the B_s data in this analysis is about 29 μm. The Figure 5-6 distribution includes $B_s^0 \to D_s^-\pi^+\pi^+\pi^-$ decays which are used in other analyses [83] and whose uncertainty averages a few μm lower because of their additional B vertex tracks, but the qualitative distributions of the $B \to D\pi$ and $B \to D3\pi$ topologies are very similar.

Because fully reconstructed decays have negligible uncertainty on momentum, their entire proper time uncertainty stems from this vertex resolution contribution. For partially reconstructed decays, the momentum uncertainty characterized by the κ-factor distributions adds an additional component to the mean baseline of $\sigma_{ct} \sim$ 29 μm. Section 5.3 explores the proper time resolution that results from the momentum uncertainty contribution.

5.2 Motivation and Formalism of Signal Modeling

The likelihood function developed for the mass space must be expanded to include the properly calibrated measurement of proper decay time and its uncertainty for each candidate. As in the case of mass, signal models tend to be based on considerations of fundamental physics, while background models are empirically motivated.

The practical implementation of signal modeling in the proper time space involves the mingling of two fundamental concepts. Rather than introducing these principles in sequence only to provide incomplete derivations of their treatment, this discussion

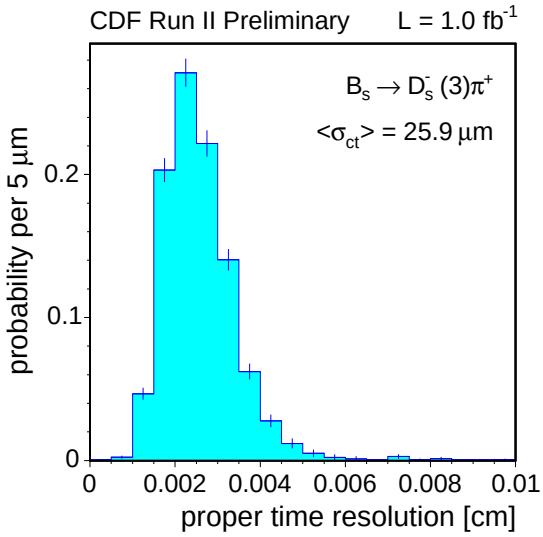

Figure 5-6: Distribution of proper time uncertainty due to vertex spatial resolution for hadronic B_s decays. The mean resolution for $B_s^0 \to D_s^- \pi^+$ decays is $\sigma_{ct} \sim 29$ μm.

seeks to outline the basic concepts before their collective treatment in Section 5.3. For purposes of brevity, the formalism is presented in terms of time t. Empirical background models are motivated and derived in Section 5.4.

5.2.1 Selection Bias

The true decay time of an unstable state is distributed as an exponential function

$$\mathrm{Exp}(t|\tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \theta(t) = \begin{cases} 0, & t < 0 \\ \frac{1}{\tau} e^{-\frac{t}{\tau}}, & t \geq 0 \end{cases} \tag{5.10}$$

with characteristic lifetime τ. The measurement of this distribution is smeared by the vertex resolution of the detector, where the Gaussian

$$\mathcal{G}(t|\sigma_t) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{t^2}{2\sigma_t^2}} \tag{5.11}$$

is a distinct resolution function for each candidate, characterized by the estimated proper time resolution σ_t. The resulting normalized probability density for measure-

ment of proper decay time is represented by the convolution of the two distributions

$$\begin{aligned}\mathcal{P}(t|\sigma_t,\tau) &= \mathrm{Exp}(t|\tau) \otimes \mathcal{G}(t|\sigma_t) \\ &= \int_{-\infty}^{\infty} \frac{1}{\tau} e^{-\frac{t'}{\tau}} \theta(t') \cdot \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t-t')^2}{2\sigma_t^2}}\, dt' \\ &= \frac{1}{2\tau} e^{-\frac{1}{\tau}(t-\frac{\sigma_t^2}{2\tau})} \cdot \mathrm{Erfc}\left(\frac{\sigma_t^2 - t\tau}{\sqrt{2}\sigma_t}\tau\right), \end{aligned} \tag{5.12}$$

using the complementary error function $\mathrm{Erfc}(z) \equiv \frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-u^2}\, du$.

In analyses whose samples are collected without regard for vertex displacement, such as recent B lifetime measurements in J/ψ decay channels [73], Equation (5.12) completely describes the decay time distribution of signal events. However, the effectiveness of the two-track trigger, on which the data samples of this analysis are based, is rooted in the fact that it is designed to accept events with decay vertices that are displaced from the primary vertex. By construction, the resulting time distribution of signal events is biased toward non-zero decay times. Therefore, the PDF used to model signal components in proper time must contain a modification to account for this additional "sculpting".

The average probability for a candidate with a given decay time to be accepted into the analysis sample may be viewed as an efficiency factor. Because the selection methods rely on candidate displacement, the probability must be treated as a *proper-time-dependent* efficiency $\varepsilon(t)$. If this acceptance function is derived with respect to the reconstructed proper time, then it models the biased time distribution by acting as a simple multiplicative factor in the overall proper time PDF

$$\mathcal{P}(t|\sigma_t,\tau) = \frac{1}{\mathcal{N}} \left[\mathrm{Exp}(t|\tau) \otimes \mathcal{G}(t|\sigma_t)\right] \cdot \varepsilon(t), \tag{5.13}$$

where the convolution is performed as in Equation (5.12). The additional normalization factor is constructed as

$$\mathcal{N}(\sigma_t,\tau) = \int_{-\infty}^{\infty} \frac{1}{2\tau} e^{-\frac{1}{\tau}(t-\frac{\sigma_t^2}{2\tau})} \cdot \mathrm{Erfc}\left(\frac{\sigma_t^2 - t\tau}{\sqrt{2}\sigma_t}\tau\right) \cdot \varepsilon(t)\, dt, \tag{5.14}$$

where the σ_t-dependence requires that it be separately evaluated for each candidate. Section 5.3 describes the extraction of selection bias curves $\varepsilon(t)$ from Monte Carlo simulation and the choice of parameterization that allows for analytical integration of Equation (5.14).

5.2.2 Unreconstructed B Momentum

In addition to effects of the displacement-dependent event selection, the measured proper decay time distribution may also be sculpted by the loss of momentum information due to unreconstructed particles. The κ-factor, defined in Equation (5.5) as the fraction of the true B momentum reconstructed for each candidate, must be incorporated into the measured decay time PDF. Because the κ-factor for any single

partially reconstructed event can only be known as a distribution, the normalized probability density $F(\kappa)$ is convolved into the PDF

$$E'(t|\tau) = \mathrm{Exp}(t|\tau) \otimes F(\kappa) = \int \frac{\kappa}{\tau} e^{-\frac{\kappa t}{\tau}} \cdot F(\kappa)\, d\kappa. \tag{5.15}$$

Just as in the case where no $F(\kappa)$ momentum smearing is present, the PDF must also include the spatial smearing of vertex reconstruction and the proper-time-dependent selection efficiency $\varepsilon(t)$. The complete PDF is represented by

$$\begin{aligned}\mathcal{P}(t|\sigma_t,\tau) &= \frac{1}{\mathcal{N}} \left[\mathrm{Exp}(t|\tau) \otimes \mathcal{G}(t|\sigma_t) \otimes F(\kappa)\right] \cdot \varepsilon(t) \\ &= \frac{1}{\mathcal{N}} \int \frac{\kappa}{2\tau} \exp\left(-\frac{\kappa}{\tau}(t - \frac{\kappa\sigma_t^2}{2\tau})\right) \cdot \mathrm{Erfc}\left(\frac{\kappa\sigma_t^2 - t\tau}{\sqrt{2}\sigma_t\tau}\right) \cdot F(\kappa)\, d\kappa \cdot \varepsilon(t), \end{aligned} \tag{5.16}$$

using the same Gaussian smearing convolution as in Section 5.2.1. Since $\varepsilon(t)$ is defined in terms of the reconstructed proper time (as opposed to the true t), it remains a multiplicative factor outside the integrals. The new normalization factor $\mathcal{N}$ must be evaluated for each event as $\mathcal{N} = \int \mathcal{N}(\sigma_t|\tau) \cdot F(\kappa)\, d\kappa$, using the definition of Equation (5.14) and the usual κ-factor transformation $\tau \to \tau/\kappa$.

It is worthy of note that the κ-factor distribution of fully reconstructed decays is simply the delta function

$$F(\kappa) = \delta(\kappa - 1). \tag{5.17}$$

In that case, the formalism of $F(\kappa)$ convolution reduces immediately to the fully reconstructed-decay PDF of Section 5.2.1.

The likelihood models for all signal components are explicitly described in Section 5.3 below.

5.3 Signal Model Components

The effort of Chapter 4 to individually model the partially reconstructed decay modes is continued here. Because proper time is the space from which the oscillation parameters are eventually extracted, it is important that it be described as accurately as possible for each signal channel. For each B species, the fully reconstructed $B \to D\pi$ and the partially reconstructed $B \to D^*\pi$ and $B \to D\rho$ decay topologies that constitute the analysis signal are modeled separately.

Just as in the case of the mass space, modeling of signal components in proper time requires the use of dedicated Monte Carlo samples. The same signal samples of Section 4.2.1 are used here. Since the κ-factor distributions are necessary for the derivation of selection bias curves, the κ-factors are extracted first.

5.3.1 κ-Factor Distributions

Extraction of κ-factor distributions from Monte Carlo simulation is straightforward. Given the generator-level truth information and the fully simulated sample reconstruction and selection, each MC event provides an entry in a histogram of the

fractional reconstructed B momentum

$$\kappa = \frac{p_T^{\rm reco} L_{\rm xy}^B}{p_T^B L_{\rm xy}^{\rm reco}}. \tag{5.18}$$

It is important that κ-factors are modeled only from MC events that inhabit the mass region of the final oscillation analysis. Whereas the mass templates and subsequent data fits draw on a wide mass range, the proper time analysis is performed within a narrower range to limit the presence of backgrounds. Only MC events within these narrow ranges are accepted for proper time templates.

In theory, it would be possible to parameterize the resulting κ-factor distributions and attempt to analytically integrate the convolutions of Section 5.2.1. However, it is shown below that the κ-factors are distributed narrowly enough that use of fewer than 20 histogram bins provides an adequate description. The probability densities $F(\kappa)$ in this analysis are modeled as normalized histograms of κ for each signal MC sample, which are numerically convolved in the likelihood framework.

Figure 5-7 shows the normalized κ-factor distributions of the partially reconstructed signal components in the B_s and B^0 samples. As is done throughout the analysis, the $B \to D^*\pi$ and its small Cabibbo-suppressed contribution are treated together within the $D^*\pi$/K component. Comparison of the κ-factors shows that the $B \to D\rho$ distribution is slightly higher that the $D^*\pi$/K component, despite having a larger Q-value. This result is explained by the low-mass boundary of the samples, which truncates the long tail of the $B \to D\rho$ and effectively imposes a smaller Q-value for the accepted decays. The $D^*\pi$/K decays are well within the bounds and are less affected. The mass spectrum of the $B \to D\rho$ component also cuts off more sharply than $D^*\pi$/K at the upper mass side due to the phase space constraint of the π^0 rest mass.

Overall, the $B \to D^*\pi$/K and $B \to D\rho$ distributions are qualitatively similar, with both B_s components averaging 96% of the total B momentum reconstructed by the charged tracks. Correspondingly, their $F(\kappa)$ **widths** are both narrow, as necessitated by the fact that the average $\langle\kappa\rangle$ is nearly fully reconstructed. Because these distributions are convolved into the proper time PDFs, thus smearing the measured decay time for each event, the well-defined momentum distributions translate directly into better proper time resolution. The relative momentum uncertainty of $\sigma_{p_T}/p_T \sim 3-5\%$ implies via Equation (5.2) that the partially reconstructed nature of these new signal events causes only a small increase in their proper time uncertainty. Since the significance of the final oscillation signature goes as $1/\sigma_{\mathcal{A}} \propto \exp(-\frac{\Delta m^2 \sigma_t^2}{2})$, the narrow κ-factor distributions of these channels are the *critical determinant* of their contribution to B_s mixing.

This discussion has thus far omitted the fully reconstructed channels because, by definition, their κ-factor distribution is the delta function centered at $\kappa = 1$. This attribute of "infinite" narrowness is the origin of their value in proper time resolution and it leads to the simpler PDF formalism. Only partially reconstructed signal channels require extraction of κ-factors.

For a relative comparison of the potential value of the new signals that are the focus

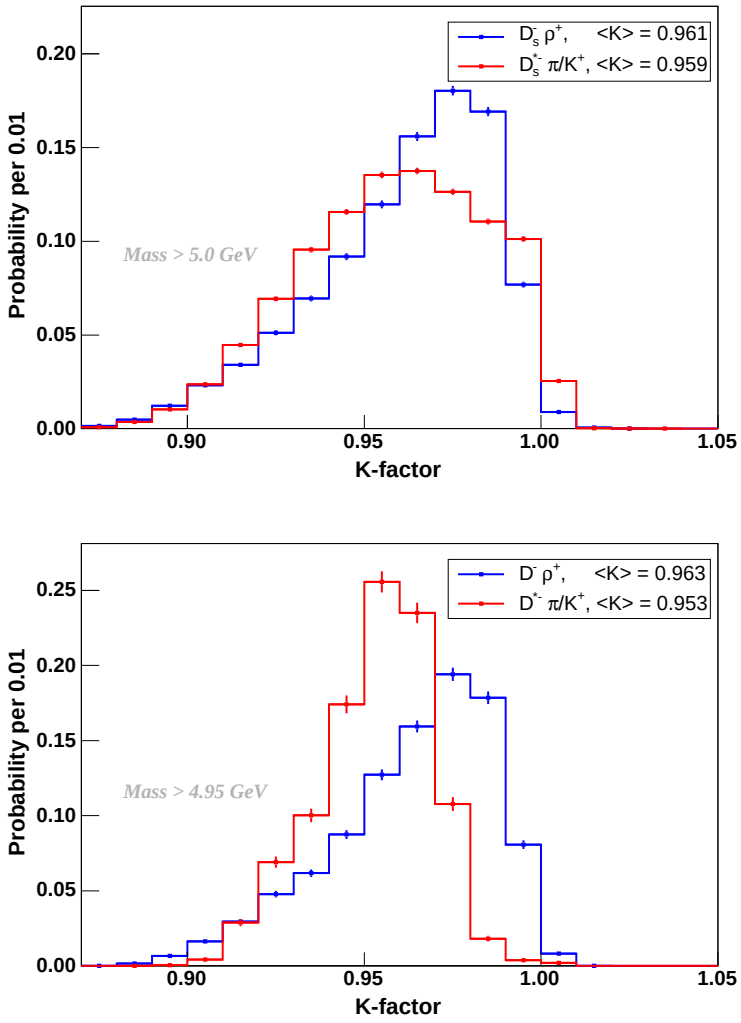

Figure 5-7: κ-factor distributions for partially reconstructed components of the B_s (top) and B^0 (bottom) samples.

of this dissertation, Figure 5-8 overlays the κ-factor distributions of these partially reconstructed hadronic channels with those of various semileptonic $B_s \to D_s^{(*)}\ell\nu$ decays [35]. The hadronic $F(\kappa)$ are narrower than that of even the best class of semileptonic decays.

Figure 5-8 also depicts the resulting average proper time resolution of these channels with respect to decay time ct. The heavy horizontal line marks the oscillation period corresponding to $\Delta m_s = 18$ ps^{-1}, which presents an approximate upper bound for the proper time resolution of events that contribute any significance. The Equation (5.2) dependence of σ_{ct} on ct is readily visible. Fully reconstructed hadronic decays, with negligible momentum uncertainty, have a constant mean σ_{ct} which arises from spatial resolution alone. The narrow κ-factors of the partially reconstructed hadronic channels implies that their time resolution remains valuable, well below the

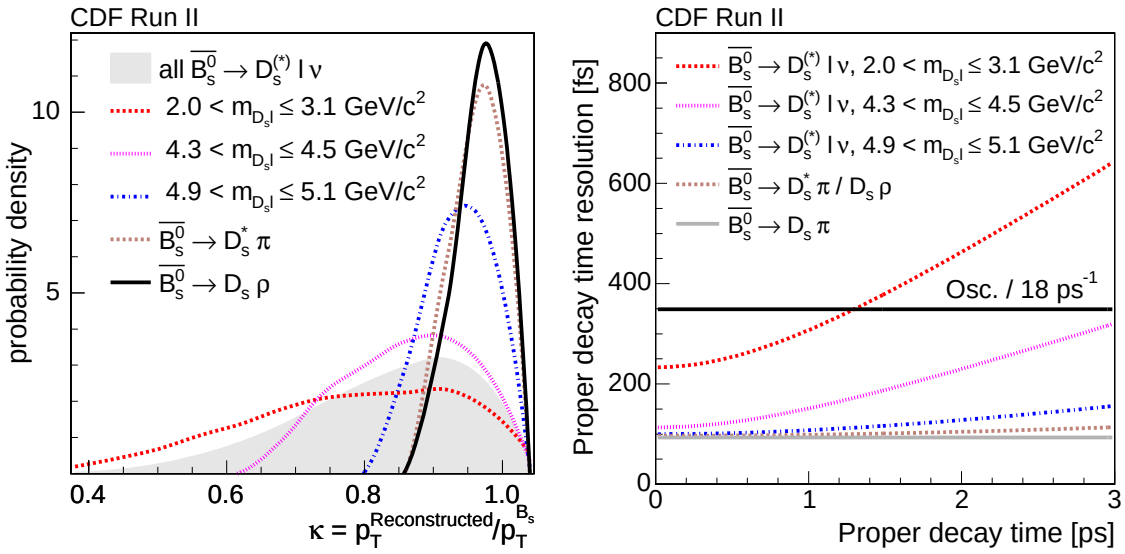

Figure 5-8: Comparison of κ-factor distributions (left) of the partially reconstructed hadronic decays in this analysis with those of semileptonic decays, and the resulting average dependence of σ_{ct} on ct (right). The narrow hadronic κ distributions translate directly into better proper time resolution. The horizontal line indicates the oscillation period corresponding to $\Delta m_s = 18$ ps^{-1}, beyond which σ_{ct} cannot resolve mixing.

approximate oscillation period, even at the upper range of decay times. The wide semileptonic κ-factor distributions lead to progressively worse proper time resolution.

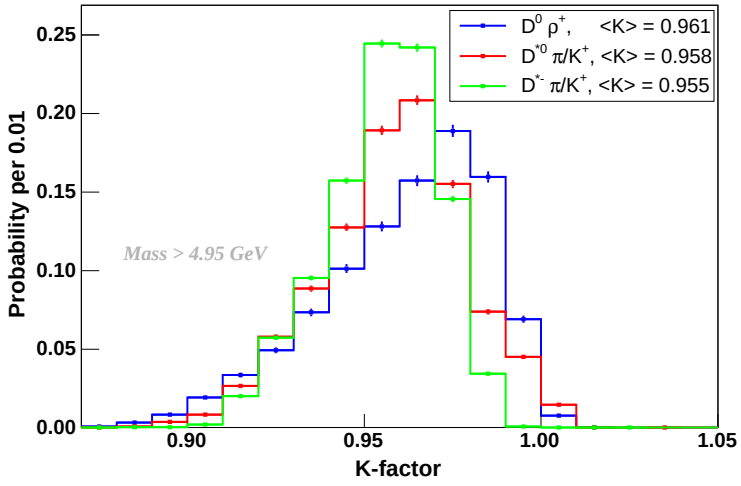

Figure 5-9: κ-factor distributions for partially reconstructed components of the B^+ sample, including its $B^0 \to D^{*-}\pi^+$ contribution.

Figure 5-9 shows the κ-factor distributions of partially reconstructed components

in the B^+ sample. The histograms are similar to those of the other B species, in terms of both widths and mean values, with the addition of the $B^0 \to D^{*-}\pi^+$ component.

The normalized histograms of Figures 5-7 and 5-9 are used as the $F(\kappa)$ distributions in the proper time likelihood components of the following sections.

5.3.2 Selection Bias Curves $\varepsilon(t)$

The proper-time-dependent selection efficiency for each signal channel is also derived from dedicated Monte Carlo simulation. The definition of $\varepsilon(t)$ is tied to its treatment as an efficiency, in that it represents the ratio, as a function of time, of the number of *accepted* events divided by the number of events *expected* from an unbiased distribution. For fully reconstructed decays, the expression is

$$\varepsilon(t) = \frac{t\text{-distribution after selection}}{\sum_{\{\sigma_t\}} \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t, \sigma_t)}, \tag{5.19}$$

where the sum over σ_t must be performed to account for the distinct resolution function of each MC event. The Gaussian smearing is performed analytically as in Section 5.2, and the lifetime τ of the expected time distribution must match that of the generated MC sample [2]. The case of partially reconstructed decays has an additional degree of complexity, as the expected distribution must contain the appropriate κ-factor:

$$\varepsilon(t) = \frac{t\text{-distribution after selection}}{\sum_{\{\sigma_t\}} \frac{1}{\tau} e^{-\frac{t}{\tau}} \otimes \mathcal{G}(t, \sigma_t) \otimes F(\kappa)}, \tag{5.20}$$

where the $F(\kappa)$ convolution is performed numerically using the histograms of Section 5.3.

As in the case of κ-factor distributions, individual selection bias curves are extracted from dedicated MC for each of the signal channels. Because all modes are sculpted by the trigger selection, both fully and partially reconstructed decay topologies must be modeled by an $\varepsilon(t)$ function.

It is seen below that a precise histogram-representation of $\varepsilon(t)$ would require many bins. Therefore, because the normalization integral of Equation (5.14) is evaluated for every candidate, computational efficiency demands an $\varepsilon(t)$ parameterization that is analytically integrable with the smeared exponential distribution. Functions of the form

$$\varepsilon(t) = \sum_{i=1}^{4} f_i (t - \delta_i)^2 \cdot e^{-\frac{t}{\tau_i}} \cdot \theta(t - \delta_i) \tag{5.21}$$

are observed to accurately model the bias curves while meeting the integrability requirement. The lengthy normalization integral is derived in Reference [34].

Figure 5-10 shows the $\varepsilon(t)$ distributions and fitted parameterizations for the

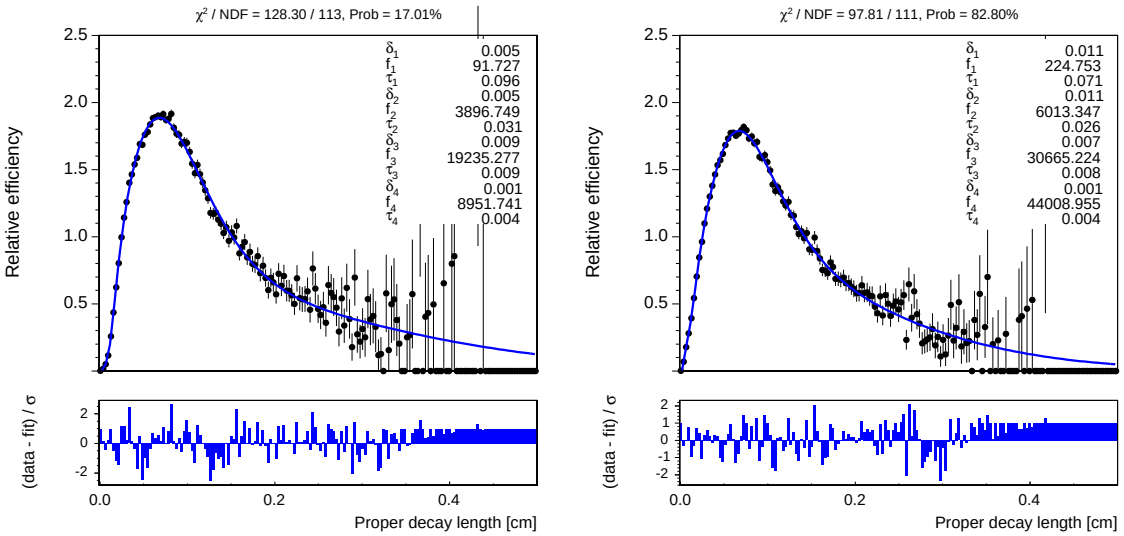

Figure 5-10: Selection bias curves $\varepsilon(t)$ for fully reconstructed $B_s^0 \to D_s^- \pi^+$ (left) and $B^0 \to D^- \pi^+$ (right) decays.

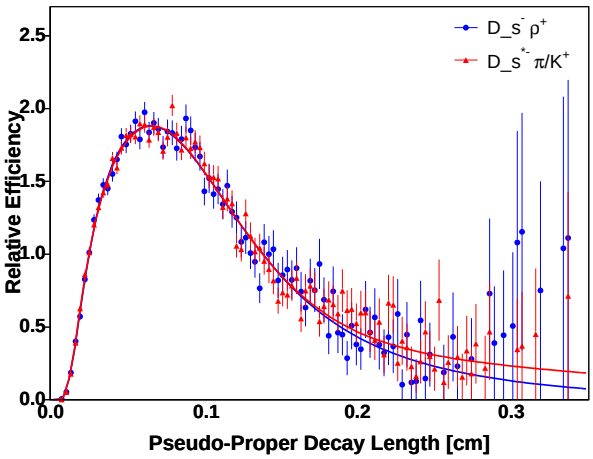

Figure 5-11: Comparison of the bias curves for partially reconstructed B_s channels.

fully reconstructed B_s and B^0 signal components. The curves for partially reconstructed signals are qualitatively very similar, both to one another and to the fully reconstructed channels. Figure 5-11 overlays the fitted $\varepsilon(t)$ distributions of the $B_s^0 \to D_s^{*-}\pi^+/K^+$and $B_s^0 \to D_s^- \rho^+$ components. In each case, the minimum displacement requirements of the trigger and selection algorithms produce the efficiency turn-on at low proper time, while the efficiency drop at large ct is due to the upper cuts on trigger track displacement. Only the relative values of points on $\varepsilon(t)$ are important, as the normalization accounts for the absolute scale.

Cabibbo-Suppressed $B \to DK$

Chapter 4 discusses how Cabibbo-suppressed $B \to DK$ decays are misreconstructed as fully reconstructed $B \to D\pi$ events and enter the sample as a small displaced bump in the mass spectrum. Because these are real B decays, they are treated as signal events for the eventual extraction of oscillation parameters. The associated mass contribution is displaced, and $B \to DK$ events are modeled by a separate component in the mass space. However, the collective proper time behavior of the Cabibbo-suppressed component is negligibly different from that of the main fully reconstructed signal, specifically with regard to the selection bias curves. As is noted below in Section 5.5.1, where the proper time likelihood is assembled, the $B \to D\pi$ and $B \to DK$ components share an identical proper time model.

5.3.3 Complete Signal PDFs

Using the κ-factor distributions and selection bias curves derived above, the complete signal PDFs may be explicitly defined. The fully reconstructed $B \to D\pi$ channels are distributed in the form of Equation (5.13):

$$\mathcal{P}^{\rm ct}_{D\pi} = \frac{1}{\mathcal{N}} \left[{\rm Exp}(ct|c\tau_B) \otimes \mathcal{G}(ct|\sigma_{ct}) \right] \cdot \varepsilon(ct), \tag{5.22}$$

where the $\varepsilon(ct)$ functions are represented in Figure 5-10. The free parameter $c\tau_B$ is the real lifetime of the appropriate B species, and the resolution function uses the calibrated proper time resolution σ_{ct} for each candidate.

The partially reconstructed decay modes require the inclusion of κ-factors and are therefore distributed as Equation (5.16). The time distributions of all partially reconstructed signals in this analysis are of identical form, differing only in their constituent $F(\kappa)$ and $\varepsilon(ct)$ templates:

$$\begin{aligned} \mathcal{P}^{\rm ct}_{D^*\pi, D\rho} &= \frac{1}{\mathcal{N}} \left[{\rm Exp}(ct|c\tau_B) \otimes \mathcal{G}(ct|\sigma_{ct}) \otimes F(\kappa) \right] \cdot \varepsilon(ct) \\ &= \frac{1}{\mathcal{N}} \left[\sum_{\kappa} \left[{\rm Exp}(ct|c\tau_B) \otimes \mathcal{G}(ct|\sigma_{ct}) \right] \cdot F(\kappa) \right] \cdot \varepsilon(ct), \end{aligned} \tag{5.23}$$

where the integration over $F(\kappa)$ is performed numerically. The $\varepsilon(ct)$ functions are represented in Figure 5-11 (for the partially reconstructed B_s signals, while the other species are very similar), and the κ-factor distributions are shown in Figures 5-7 and 5-9. Because the $c\tau_B$ lifetime is a physically meaningful quantity, the same parameter is *shared* among the fully and partially reconstructed signal components. In the case of the $B^0 \to D^{*-}\pi^+$ contribution to the B^+ sample, the B^0 lifetime receives a separate parameter.

5.4 Background Templates

Modeling of background components is empirically driven, with the goal of describing the time distribution of each class of events so as not to bias the extraction of physical parameters from the signal models. Because B_s oscillations occur rapidly with respect to any other variations in the proper time distribution, the systematic uncertainty studies of Appendix B show that the measured value of Δm_s is robust to the choice of background descriptions.

In a similar fashion to the associated templates for mass, the proper time model for combinatorial background may be fitted directly in data, while the "physics background" classes of decays are entangled with other components and require templates from Monte Carlo simulation.

Function Definition

The empirical approach to background modeling, in combination with selection and detector effects that are common to all components, allows for a universally applicable template form. The candidate displacement requirements and detector resolution consistently produce smeared exponentials with turn-ons that are offset from $ct = 0$. The normalized function which forms the basis of all background decay time models is expressed as

$$\mathrm{ExpGaus}(ct|\delta, c\tau, \sigma) = \left[\frac{1}{c\tau} e^{-\frac{(ct-\delta)}{c\tau}} \theta(ct-\delta)\right] \otimes \mathcal{G}(ct, \sigma), \tag{5.24}$$

where the convolution is performed as in Section 5.2. The Gaussian width parameter σ is unrelated to the estimated σ_{ct} for each candidate, because the calibration of signal resolution cannot be assumed to apply to background candidates and because the same level of precision is not needed. The bounds of the analysis proper time space are wide enough that, unlike the mass space, no explicit boundary-dependent normalization is necessary. ExpGaus is used alone or in linear combinations to model the background templates below.

$b \to DX$ Physics Background

Although the wide mass range of the initial data fits in Chapter 4 causes the $b \to DX$ wedge component to appear prominently, its relative contribution is reduced drastically by the truncation of the sample to narrower mass ranges. Therefore, the physics background template for proper time receives a weight of only $\lesssim 3\%$ of the total sample.

The DX proper time template is extracted from the semi-generic MC samples in the narrow mass range. The templates for each B species are parameterized by a single ExpGaus smeared exponential

$$\mathcal{P}^{\mathrm{ct}}_{DX} = \mathrm{ExpGaus}(ct|\delta_{DX}, c\tau_{DX}, \sigma_{DX}). \tag{5.25}$$

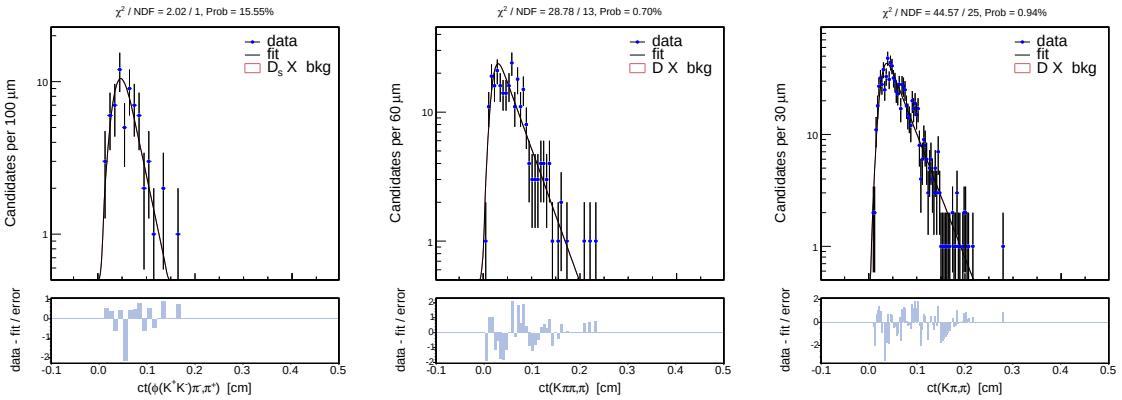

Figure 5-12: Proper time templates for $b \to DX$ physics background in the B_s (left), B^0 (center), and B^+ (right) samples.

Figure 5-12 shows the fitted templates of the physics background components for the three B species, which are very similar. If this background were to play a larger role in the analysis, than the limited number of Monte Carlo events would be of concern. However, the presence of this template has minimal impact on the parameters of interest, as shown in Appendices A and B.

Reflections

The considerations for modeling of the misreconstructed b hadron reflections are similar to those of the physics background, except that the reflection contributions to the samples are even smaller. These proper time templates are extracted from the same dedicated MC samples used in Chapter 4 for the associated mass templates.

The reflected component of B^0 decays in the B_s sample is parameterized with one smeared exponential, while the B_s component in the B^0 sample requires two such functions. The general form of the template is

$$\begin{aligned}\mathcal{P}^{\rm ct}_{B_q\,{\rm ref}} = (1 - f_{B_q\,{\rm ref}2}) \cdot {\rm ExpGaus}(ct|\delta_{B_q\,{\rm ref}1}, c\tau_{B_q\,{\rm ref}1}, \sigma_{B_q\,{\rm ref}1}) \\ + \quad f_{B_q\,{\rm ref}2} \cdot {\rm ExpGaus}(ct|\delta_{B_q\,{\rm ref}2}, c\tau_{B_q\,{\rm ref}2}, \sigma_{B_q\,{\rm ref}2}),\end{aligned} \tag{5.26}$$

with independent parameters. The reflected Λ_b component in the B_s and B^0 samples is modeled by a single smeared exponential

$$\mathcal{P}^{\rm ct}_{\Lambda_b\,{\rm ref}} = {\rm ExpGaus}(ct|\delta_{\Lambda_b\,{\rm ref}}, c\tau_{\Lambda_b\,{\rm ref}}, \sigma_{\Lambda_b\,{\rm ref}}). \tag{5.27}$$

The charged B^+ sample contains no neutral b hadron reflections. The template parameters are compiled in Table 5.2. Using the relative normalizations derived in Chapter 4, the contribution of these components is very small.

Parameter		$B_s^0 \to D_s^- \pi^+$	$B^0 \to D^- \pi^+$
$c\tau_{B_q \mathrm{ref}1}$	[μm]	413.8	355.3
$\delta_{B_q \mathrm{ref}1}$	[μm]	218.5	243.3
$\sigma_{B_q \mathrm{ref}1}$	[μm]	66.3	66.7
$f_{B_q \mathrm{ref}2}$		0.0	0.176
$c\tau_{B_q \mathrm{ref}1}$	[μm]	—	191.4
$\delta_{B_q \mathrm{ref}1}$	[μm]	—	529.3
$\sigma_{B_q \mathrm{ref}1}$	[μm]	—	154.8
$c\tau_{\Lambda_b \mathrm{ref}}$	[μm]	282.7	254.4
$\delta_{\Lambda_b \mathrm{ref}}$	[μm]	251.0	282.2
$\sigma_{\Lambda_b \mathrm{ref}}$	[μm]	76.4	89.5

Table 5.2: Proper time template parameters for reflection components.

Combinatorial Background

The presence of a nearly pure combinatorial background sideband in the upper mass region allows for its model to be extracted directly from the data. This component is again parameterized with two smeared exponentials, where only the first function is necessary for the lower statistics of the B_s data:

$$\begin{aligned} \mathcal{P}_{\mathrm{com}}^{\mathrm{ct}} &= (1 - f_2) \cdot \mathrm{ExpGaus}(ct | \delta_{\mathrm{com}1}, c\tau_{\mathrm{com}1}, \sigma_{\mathrm{com}1}) \\ &+ \quad f_2 \cdot \mathrm{ExpGaus}(ct | \delta_{\mathrm{com}2}, c\tau_{\mathrm{com}2}, \sigma_{\mathrm{com}2}). \end{aligned} \tag{5.28}$$

Calculation of proper time via Equation (5.1) with a world average M_B is reasonable for signal decays, where this is the true mass. In contrast, M_B is not meaningful for the random track vertices of combinatorial background candidates. The resulting proper time distribution of this background component has a mass dependence, such that the fitted model of Equation (5.28) becomes a weighted average over the total mass range. Appendix F.2 discusses this effect in greater detail with representative distributions. Since the flavor oscillations of the signal events occur much more rapidly than any other variations of the proper time distribution, the mixing parameters are minimally affected by the background ct model. Specifically, the bias to Δm_s from use of a mass-invariant distribution is negligible, as discussed in Section 7.2.3 and Appendix B.

The measurement where this combinatorial background variation has a strong effect is in the extraction of B lifetimes from the partially reconstructed signals. An unbiased lifetime fit requires the likelihood function to distinguish the proper time distributions of signal and background, and an imprecise background model in the partially reconstructed region skews the fitted signal model. While the partially reconstructed signals have high statistical precision in lifetime measurement, the associated systematic uncertainties are large. As discussed below, the systematic uncertainties on lifetime measurement are not quantitatively studied in this analysis.

The nominal model of Equation (5.28) is fitted in data for the full combinatorial

background component.

5.5 Lifetime Fits in the Data

Previous sections describe the measurement of proper decay time and its modeling for individual sample components. This section incorporates these models into the likelihood framework and applies it to the data.

5.5.1 Complete Likelihood Without Flavor Tagging

The likelihood developed for the mass space in Chapter 4 must be expanded to contain the PDFs for proper time. Each sample component becomes the product of the separately normalized probability densities for its mass and time distributions, $\mathcal{P} = \mathcal{P}^{\rm m} \cdot \mathcal{P}^{\rm ct}$, as generalized in Equation (C.3) [1]. Using the PDFs defined for each signal and background component in Sections 5.3 and 5.4, the likelihood is expressed as

$$\begin{aligned}\mathcal{P}_{\rm evt} = (1-f_{\rm com}) \cdot \Bigg[\quad & f_{D\pi} \cdot \frac{(\mathcal{P}^{\rm m}_{D\pi} + r_{\rm Cab}\mathcal{P}^{\rm m}_{\rm Cab})\mathcal{P}^{\rm ct}_{D\pi} + r_{B_q{\rm ref}}\mathcal{P}^{\rm m}_{B_q{\rm ref}}\mathcal{P}^{\rm ct}_{B_q{\rm ref}} + r_{\Lambda_b{\rm ref}}\mathcal{P}^{\rm m}_{\Lambda_b{\rm ref}}\mathcal{P}^{\rm ct}_{\Lambda_b{\rm ref}}}{1 + r_{\rm Cab} + r_{\Lambda_b{\rm ref}} + r_{B_q{\rm ref}}} \\ & + f_{D^*\pi} \cdot \mathcal{P}^{\rm m}_{D^*\pi}\mathcal{P}^{\rm ct}_{D^*\pi} \\ & + f_{D\rho} \cdot \mathcal{P}^{\rm m}_{D\rho}\mathcal{P}^{\rm ct}_{D\rho} \\ & + (1 - f_{D\pi} - f_{D^*\pi} - f_{D\rho}) \cdot \mathcal{P}^{\rm m}_{DX}\mathcal{P}^{\rm ct}_{DX} \Bigg] \\ + \qquad & f_{\rm com} \cdot \mathcal{P}^{\rm m}_{\rm com}\mathcal{P}^{\rm ct}_{\rm com}. \end{aligned} \tag{5.29}$$

The complexity of this function underscores the value of having extracted and fixed the parameters of the mass space before including any proper time information. In particular, the signal fractions are fixed to those measured from the wide mass range, after adjusting them to the narrow mass range. The fits to data before inclusion of flavor tags involve relatively few free parameters which are summarized in Table 5.3.

The likelihood for the B^+ sample also contains an additional partially reconstructed component for the $B^0 \to D^{*-}\pi^+$ contribution. The associated $c\tau_{B^0}$ lifetime parameter is separate from the one shared by the main signals and is fixed to the world average value [2].

Fitter Validation

As is done for the fits in the mass space alone, the expanded likelihood is validated with large ensembles of toy Monte Carlo samples. Under the assumption that the

[1] As an aside, it should be noted that the signal PDFs $\mathcal{P}^{\rm ct}(ct, \sigma_{ct})$ of Section 5.2 are technically defined as conditional probability densities of ct with respect to σ_{ct}: $\mathcal{P}(ct, \sigma_{ct}) = \mathcal{P}(ct|\sigma_{ct})\mathcal{P}(\sigma_{ct})$. Rigorous construction of the likelihood would therefore require the latter probability distribution of σ_{ct} for each sample component [96]. However, these distributions are observed to be so similar as to effectively cancel out of the expression for the hadronic data. The same simplification may not be made in, for example, $B \to J/\psi$ decays [34]. This concept is revisited for the likelihood of Chapter 6.

Group/notes	Parameters
Signal lifetime	
– shared by components	$c\tau_B$
Combinatorial background	
– fraction of sample	$f_{\rm com}$
– ct distribution	$\{c\tau_i\ ,\ \delta_i\ ,\ \sigma_i\} \otimes \{i = 1, 2\};\ f_2$

Table 5.3: Free parameters in fits of proper time without tagging information. All other parameters of Equation (5.29) are fixed from the mass space.

κ-factor distributions and selection bias curves are modeled correctly, each of the separate pure signal components and their combinations are verified to produce unbiased measurements of the shared lifetime. Toy samples are constructed for the full Equation (5.29) model using parameter values and sample sizes which are representative of the data, and the fitter framework is demonstrated to extract unbiased lifetime and background fraction measurements. Appendix D documents representative results.

5.5.2 Results

The complete likelihood of proper time without flavor tagging is applied to the data in the narrow mass ranges Figures 5-13 - 5-15 show the fitted proper time and mass space of 30478 B^0 candidates, 37360 B^+ candidates, and 7339 B_s candidates, respectively. The ct distributions are depicted on a logarithmic scale.

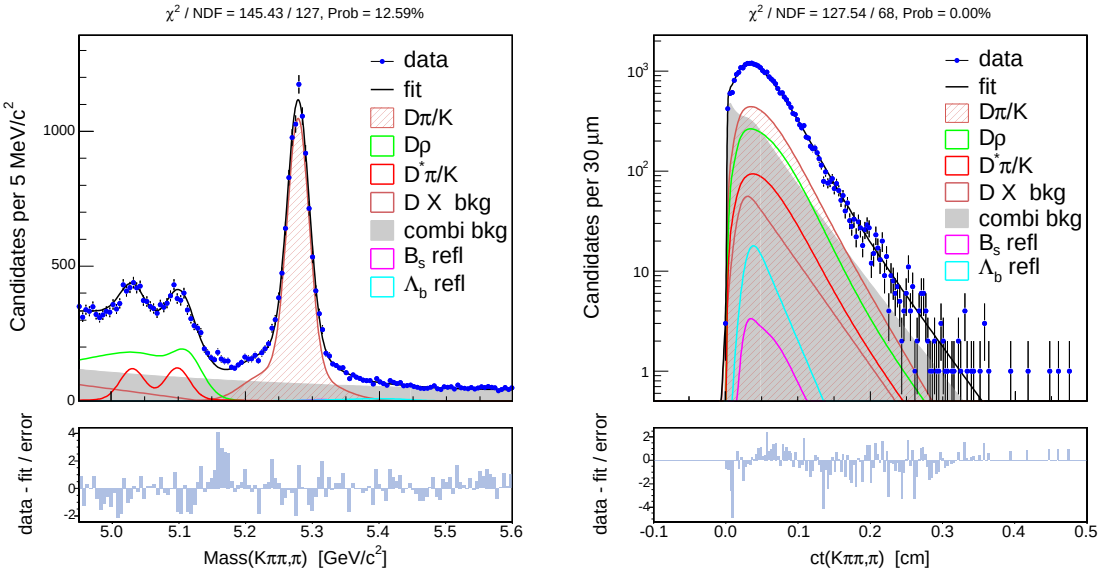

Figure 5-13: Fitted mass and proper time spaces of B^0 candidates in 355 pb^{-1}.

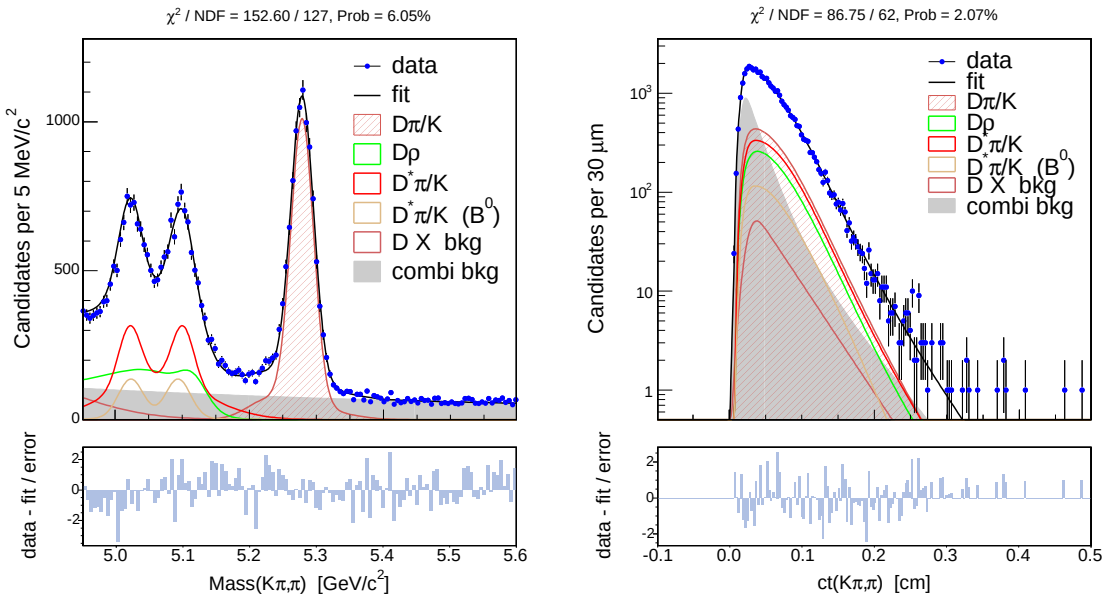

Figure 5-14: Fitted mass and proper time spaces of B^+ candidates in 355 pb^{-1}.

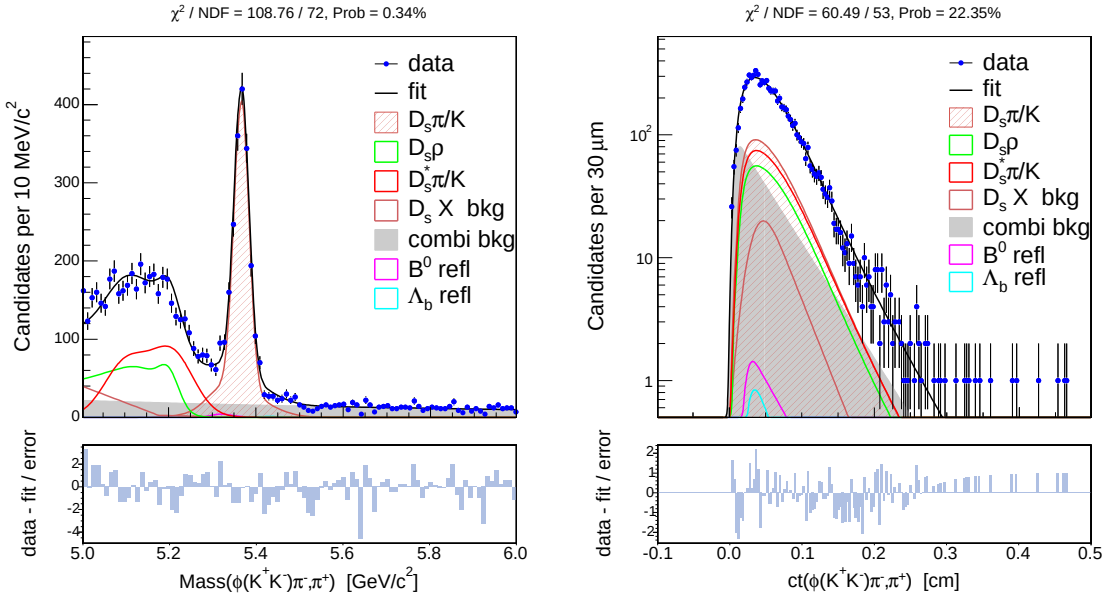

Figure 5-15: Fitted mass and proper time spaces of B_s candidates in 1 fb^{-1}.

Inspection of the mass spectra indicates that the signal fractions extracted from Chapter 4 model the data well and confirms that the presence of the physics background is greatly reduced in the narrowed range.

In general, and especially for the B^+ and B_s samples, the proper time distributions appear well-modeled, lending further support for the correctness of the fixed

fractions. The contributions of the fully and partially reconstructed signals are separately visible, with their shapes differing only by the sculpting of the individual $\varepsilon(ct)$ selection efficiencies and the κ-factor distributions for the $D\rho$ and $D^*\pi$ components. The fixed templates for the physics background and reflections are visible in the log-scale of the ct figures but have very small contributions. The steep turn-on of the distribution at low ct is due to the combinatorial background.

The only apparent flaw in the proper time model is in the turn-on of the B^0 combinatorial background. The B^0 sample has a modest "kink" in the first few bins of the Figure 5-13 likelihood projection which, in combination with the large sample statistics, the nominal background model struggles to account for. This mild deficiency is not observed in the B_s data, and further studies of proper time distributions in the upper mass sidebands indicate that the B_s background does not exhibit such a sharp turn-on. Regardless, it should be noted that this turn-on effect is, at most, weakly coupled to the signal lifetime, which is largely determined by the portion of the distribution at higher ct. Most importantly, as discussed in greater depth below, any such effects are irrelevant for measurement of B_s oscillations, which are very rapid and are extracted from the ct region with greater signal content.

Parameter		B_s	B^0	B^+
$c\tau_B$	[μm]	xxx.x $\pm$ 10.3	469.1 $\pm$ 5.7	486.1 $\pm$ 5.9
$f_{\rm com}$		0.203 $\pm$ 0.007	0.296 $\pm$ 0.005	0.264 $\pm$ 0.004
$c\tau_1$	[μm]	463.7 $\pm$ 20.7	353.6 $\pm$ 33.0	188.9 $\pm$ 9.9
δ_1	[μm]	75.4 $\pm$ 7.1	28.0 $\pm$ 0.9	166.7 $\pm$ 4.4
σ_1	[μm]	37.1 $\pm$ 6.0	5.4 $\pm$ 0.8	44.9 $\pm$ 2.3
f_2		0.0 $\pm$ —	0.319 $\pm$ 0.058	0.215 $\pm$ 0.052
$c\tau_2$	[μm]		509.7 $\pm$ 20.0	471.2 $\pm$ 42.8
δ_2	[μm]		251.5 $\pm$ 25.7	97.1 $\pm$ 5.1
σ_2	[μm]		90.5 $\pm$ 12.6	16.1 $\pm$ —

Table 5.4: Fitted parameters of the proper time space without flavor tagging information. All uncertainties are statistical only. The B_s lifetime is not reported, as discussed in the text.

Table 5.4 compiles the fitted parameters of the three samples. While the lifetimes of the light B mesons are reported, it must be emphasized that the focus of this analysis is on flavor oscillations and *not* the extraction of lifetimes. The fitted lifetime values are very sensitive to the fraction of the background and its proper time model, to a degree that is not matched for parameters of flavor oscillations. Qualitatively, it may be readily understood how the relative weight and lifetime of the background component in the likelihood has a directly proportional effect on the signal lifetime. Rigorous lifetime measurements would require evaluation of systematic uncertainties on $c\tau_B$ to account for different possible background models [97]. In contrast, Chapter 6 introduces how the signal ct distributions are expanded to include cosinusoidal oscillation terms, which remain visible even while the non-oscillating background model is varied. The systematic uncertainty discussion of Section 7.2.3 and Appendix A

support this assertion, which is the justification for the lack of concern about fitted B lifetime values. In this sense, the lifetime values are only important in this mixing analysis to the extent that they are free parameters that allow the nominal model to best fit the data.

As such, it is reassuring to note that the fitted values of $c\tau_{B^0}$ and $c\tau_{B^+}$ in Table 5.4 are consistent with the world average values, but they are incomplete measurements without systematic uncertainty estimates. The data also produces a reasonable lifetime for the B_s signal. However, its value is not reported because a dedicated CDF analysis of lifetimes in the partially reconstructed hadronic decays is ongoing [97] and the B_s lifetime for this data is blinded. Regardless, despite the lack of systematic uncertainties, the statistical precision of $c\tau$ measurement including the additional partially reconstructed signal events is impressive.

Rigorous measurement of the fully reconstructed B lifetimes in the first 355 pb^{-1} of this data is documented elsewhere [98].

Extraction of Parameters

The potential complexity of the lifetime fits has been reduced by the use of fixed parameter values which are extracted from the mass space in Chapter 4. In an analogous strategy, the fitted parameters of the lifetime analysis are also set constant upon expansion of the likelihood to incorporate flavor tagging and asymmetry. Specifically, the proper time parameterization of the combinatorial background component and its overall sample fraction are fixed to the values of Table 5.4 for the following chapters. The B_s amplitude scans also use the lifetime value fitted above, because the ability of the nominal model to resolve lifetimes is fully represented here.

Chapter 6

Flavor Tagging and Asymmetry

In addition to the sample composition and proper time measurements characterized in previous chapters, the third critical ingredient for observation of flavor oscillations is the ability to identify B production flavor. This chapter outlines the associated techniques and the considerations for their use for partially reconstructed hadronic decays. The likelihood framework is expanded to describe flavor asymmetry and is tested by application to the data for the light B species.

6.1 Principles of Flavor Tagging

Although the analysis of mass and proper time has thus far been flavor-invariant, measurement of time-dependent oscillations between B_s^0 and $\overline{B}_s^0$ conjugate states requires knowledge of the B_s flavor at production and decay. The decay flavor is readily inferred from the charge of daughter particles, but measurement of the production flavor presents a technical challenge. The associated "flavor tagging" algorithms rely on related attributes of the $p\overline{p}$ collision event to infer the identity of the hadronized b or $\overline{b}$ quark. The techniques fall into two categories: *Opposite Side Tagging*, which examines the decay products of the "other" member of the $b\overline{b}$ pair, and *Same Side Tagging*, which looks for fragmentation tracks associated with the hadronization of the triggered B meson. Figure 6-1 sketches a $b\overline{b}$ event with the structures for flavor tagging. These algorithms and their implementation at the CDF experiment are outlined below.

It should be understood that this dissertation is devoted to inclusion of the partially reconstructed hadronic decay channels, for which much of the associated effort is focused on characterization of mass and proper time and on the ramifications of the model for the flavor oscillation signatures. Therefore, although the tagging techniques themselves are of exceptional importance to observation of B_s mixing, the discussion of their underlying details in the following sections is disproportionately brief. The technical development of the tagging tools is documented extensively elsewhere [89, 99–103], while this chapter aims primarily to introduce the methods and motivate their application to the signal channels of this analysis.

The two categories of flavor tagger share a common formalism for quantifying

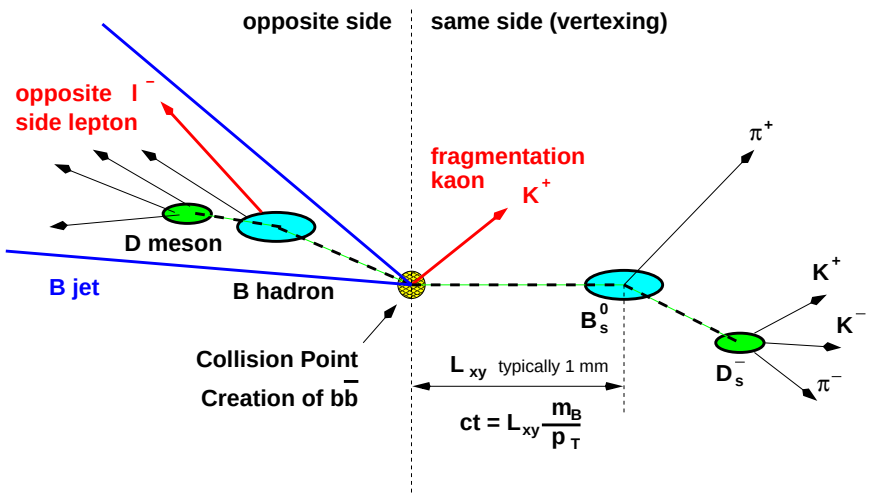

Figure 6-1: Idealized sketch of the hadronization and decay of a $b\bar{b}$ pair, producing a B_s^0 signal event with same side and opposite side tagging attributes.

their effectiveness. These principles are discussed first below, followed by an overview of each method.

Tagging Formalism

The introductory theory of Chapter 1 describes how a neutral B_s^0 meson, upon its creation in the flavor eigenstate $|\bar{b}s\rangle$, evolves in time to produce non-zero probabilities of decay in the $\overline{B}_s^0$ flavor state $|b\bar{s}\rangle$. The same time evolution occurs for $\overline{B}_s^0$ states, such that B_s mesons probabilistically oscillate in time between the flavor state in which they are created and the opposite state. The flavor state term of the decay probability is

$$\mathcal{P}(t)_{\substack{\text{unmixed}\\\text{mixed}}} \sim \frac{1}{2}\left[1 \pm \cos(\Delta m t)\right], \tag{6.1}$$

where the $+$ $(-)$ sign refers to decay at proper time t in the same (opposite) state as that of production. By comparison of the tagged production flavor and the observed decay state, each B_s candidate is classified as being "unmixed" or "mixed". These decay types are often referred to as "right sign" (RS) and "wrong sign" (WS), respectively, because the tagging methods are all ultimately based on comparison of the signs of particle charges. Tagged events may be used to construct the flavor asymmetry

$$A = \frac{N_{\text{unmixed}} - N_{\text{mixed}}}{N_{\text{unmixed}} + N_{\text{mixed}}}, \tag{6.2}$$

where N_{unmixed} and N_{mixed} represent the number of B_s mesons of each class. Use of Equation (6.1) reveals that the time-dependence of the flavor asymmetry is simply

$$A(t) = \cos(\Delta m t), \tag{6.3}$$

with cosinusoidal amplitude equal to 1. This unit amplitude would result from tagging algorithms which are always correct. In reality, the methods are imperfect for various reasons. Each tag has some probability p_w of being incorrect, so that the measured asymmetry is actually

$$\begin{aligned} A &= \frac{N_{\rm RS} - N_{\rm WS}}{N_{\rm total}} \\ &= \frac{[(1-p_w)N_{\rm unmixed} + p_w N_{\rm mixed}] - [(1-p_w)N_{\rm mixed} + p_w N_{\rm unmixed}]}{N_{\rm unmixed} + N_{\rm mixed}} \\ &= (1-2p_w)\frac{N_{\rm unmixed} - N_{\rm mixed}}{N_{\rm total}} \;=\; \mathcal{D}\frac{N_{\rm unmixed} - N_{\rm mixed}}{N_{\rm total}}. \end{aligned} \tag{6.4}$$

where $N_{\rm RS}$ and $N_{\rm WS}$ are the measured numbers of unmixed and mixed decays. The *dilution* $\mathcal{D} \equiv (1-2p_w)$ is the standard measure of flavor tagger correctness, and it acts as a simple damping factor on the measured flavor asymmetry

$$A(t) \to \mathcal{D}\cos(\Delta m t), \tag{6.5}$$

$$\mathcal{P}(t)_{\substack{\rm unmixed \\ \rm mixed}} \to \frac{1}{2}\left[1 \pm \mathcal{D}\cos(\Delta m t)\right]. \tag{6.6}$$

Therefore, a flavor tagger which always yields the correct tag has a dilution of 1, while a completely random tagger gives a correct decision with 50% probability and has a dilution of zero. Obviously, the analysis needs the tagger dilution to be as large as possible in order to resolve the oscillation term of the PDF.

In addition to limitations in correctness, each flavor tagging algorithm has only a certain probability to find a tag with either decision for a given event. For example, the lepton candidate on which certain opposite side methods rely may not be available. The tagging efficiency is determined by the availability of the event attribute(s) on which each method is based, and it is defined as

$$\epsilon = \frac{N_{\rm RS} + N_{\rm WS}}{N_{\rm untagged} + N_{\rm RS} + N_{\rm WS}}, \tag{6.7}$$

with $N_{\rm untagged}$ as the number of candidates for which no tag decision is available. While all B candidates in the sample are used for analysis of the mass and proper time spaces, only tagged candidates contribute to the oscillation measurement. As such, the analysis depends on the tagger efficiencies being as high as possible.

Tagger Effectiveness

Equation (1.32) indicates that the tagger-dependence of the oscillation signal significance goes as $1/\sigma_{\mathcal{A}} \propto \sqrt{\mathcal{S} \cdot \epsilon\mathcal{D}^2}$, with $\mathcal{S}$ signal events. Therefore, in weighing the effectiveness of the tagging methods throughout this section, the figure of merit for each is its $\epsilon\mathcal{D}^2$ value. Typical values at CDF are $\epsilon\mathcal{D}^2 \sim \mathcal{O}(1\%)$ for each algorithm. The "effective statistics" of the data for the oscillation significance is given by the product $\mathcal{S}\epsilon\mathcal{D}^2$, which is the equivalent number of perfectly tagged signal events.

The dilution associated with each flavor tag is dependent on several variables, such as p_T or discriminants for particle identification. Event-by-event estimates of tag dilution are made in order to assign appropriate weight to each candidate in the likelihood fit, thus making the best statistical use of flavor information. The oscillation term of the likelihood remains in the form of Equation (6.6), but each tagged candidate contributes a different $\mathcal{D}$ value. With this additional input, the overall tagging effectiveness is quantified by the averaged value $\epsilon\langle\mathcal{D}^2\rangle$, although it is commonly referred to as simply $\epsilon\mathcal{D}^2$.

This analysis applies the two flavor tagging methods to each event of the data, whenever possible. Section 6.5 describes the combination of the tagging decisions and event-by-event dilutions in the PDF for each event. Given that each method arrives at its tag decisions independently, their separate $\epsilon\mathcal{D}^2$ values combine linearly.

6.2 Opposite Side Tagging

The strategy used by the opposite side tagger (OST) algorithms is to infer the initial flavor of the reconstructed B meson by examination of the hadronization products of the other quark of the original $b\bar{b}$ pair. The ideal scenario for this technique would involve full reconstruction of the opposite side B, whose decay products would identify its flavor as clearly as is available for the trigger-side B. However, the fraction of events for which such a complete reconstruction is possible on both decay sides is much too small to be of practical use. Instead, the OST methods examine the inclusive attributes of the b decay opposite the triggered B in order to make a statistical judgment about its initial flavor.

The situation is complicated by the need to identify the remnants of the opposite side b hadron decay among numerous unrelated tracks. Additionally, the opposite side jet is often not even within the fiducial acceptance of the detector. The inclusive nature of the jet reconstruction implies that the opposite side method has a high efficiency for returning a tag decision but with low dilution. Finally, the tagger must contend with the time-integrated probability for an opposite side neutral B^0 or B_s meson to have mixed b flavors itself.

A distinctive feature of the OST strategy is that its decisions rely on event attributes which are independent of the species of reconstructed B meson. That is, whether a $\bar{b}$ quark hadronizes on the triggered side of the event to produce a B_s^0, B^0, or B^+ meson (or the conjugate case), the behavior of the opposite b hadronization remains the same. Therefore, development of the OST method may be performed in the data of the light B samples, whose flavor asymmetry is well understood. The resulting models for dilution are then applied without change to the B_s data. This scenario is in contrast to the development of the Same Side Kaon Tagger of Section 6.3 below, which relies on event information which depends on the reconstructed b hadron and which must be developed in Monte Carlo simulation.

The application of the OST method to partially reconstructed decays is straightforward because the opposite-hemisphere event structures are independent of the reconstructed B topology. This analysis uses three major OST techniques. The sepa-

rate output values of these algorithms are combined by a neural network into a single decision and dilution estimate. The three approaches are outlined below, followed by their results in combination.

Development Samples

The opposite side taggers are developed in a large sample of semileptonic B decays which are collected by the ℓ+SVT trigger noted in Section 3.1. The large size of the sample is important for parameterizing the dependence of the tagging dilution on various quantities. The dilution parameterization allows the method to be transferred directly to the hadronic B samples, where a global scale factor is applied to account for kinematic differences between the hadronic two-track trigger and the ℓ+SVT sample. This final calibration is performed using measurements of flavor asymmetry in hadronic B^0 and B^+ samples, as described below after the neural network combination.

All OST methods use jets which are reconstructed with a track-based cone clustering algorithm [99]. Reconstruction of the b hadron decay point is also attempted in order to identify the track-based jets that are most likely to originate from an opposite-side b quark. Jets with successfully reconstructed secondary vertices are known as `SECVTX` jets [104].

6.2.1 Lepton Taggers

The cleanest flavor signature among OST methods comes from leptons originating in semileptonic decays of the opposite side b hadron. The lepton charge is correlated with the opposite b flavor in a simple manner: ℓ^- tracks for muons and electrons are associated with $b \to c\,\ell^-\bar{\nu}X$ transitions, while ℓ^+ tracks originate from $\overline{b}$ quarks. Lepton tags have low efficiency because the inclusive semileptonic B branching fraction $\mathcal{B}(B \to \ell X) \sim 20\%$ is relatively small. Additionally, the lepton is often not within the fiducial volume of tag acceptance. However, because of the good purity of lepton identification with the CDF detector, the associated dilution is high.

Leptons in opposite side jets are identified via likelihood discriminants that incorporate information from several detector systems [35]. The muon likelihood is constructed from variables including the agreement between COT tracks and muon chamber stubs as well as energy deposition in the calorimeters. The electron likelihood is formed from many EM and hadronic calorimeter variables, in addition to energy loss information from the tracking system. Minimum likelihood requirements are derived from clean samples of each lepton type, such as $J/\psi \to \mu^+\mu^-$ decays and $\gamma \to e^+e^-$ conversions, in combination with fake lepton samples. Electrons and muons are also required to have $p_T > 2.0$ GeV/c and $p_T > 1.5$ GeV/c, respectively.

Lepton tagger dilution is parameterized in terms of lepton likelihood and the quantity $p_T^{\rm rel}$, which is defined for the tagging lepton as the magnitude of its momentum component perpendicular to the axis of its associated jet, where the jet momentum axis is recalculated without the contribution of the lepton. The large semileptonic data sample is divided into bins of lepton likelihood and $p_T^{\rm rel}$, and the resulting vari-

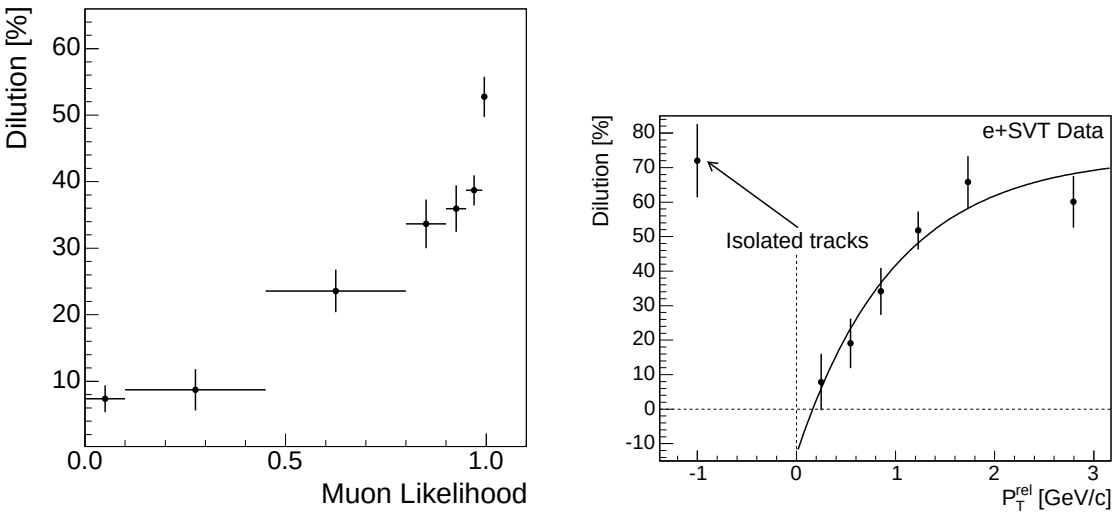

Figure 6-2: Representative dilution dependences in the lepton taggers. Left: muon tag dilution with respect to muon likelihood. Right: dilution of high-quality electron tags with respect to $p_T^{\rm rel}$.

ation of the measured dilution is fitted. The muon and electron taggers are treated as different algorithms, where muon tags have higher dilution because of the purity of the muon chamber signatures. Figure 6-2 shows representative variations of the dilution for the lepton taggers, while the overall $\epsilon\mathcal{D}^2$ values are compiled with those of the other tagging algorithms in Table 6.1.

6.2.2 Jet Charge Taggers

The ideal charge of an opposite side jet is directly related to the flavor of its underlying b quark. The jet charge tagger uses the momentum-weighted sum of charges within jet tracks to estimate the overall jet charge. Artificial neural networks (NN) [99, 105] are used to identify the most probable b jets from the set of jets reconstructed by the cone-clustering algorithm. The network is trained with a large PYTHIA Monte Carlo sample. The NN training is performed in two stages. The first stage involves construction of a NN to estimate the probability of each track to have originated from a b hadron decay. For each track, the resulting probability $\mathcal{P}_{\rm trk}$ and other kinematic variables are input to a second NN to determine the probability $\mathcal{P}_{\rm nn}$ for the associated jets to be that of a b quark. For each event with one or more available jets, the jet with highest $\mathcal{P}_{\rm nn}$ is selected for the jet charge tag.

The estimated opposite side flavor is given by the sign of the jet charge Q_{jet}, which is defined as

$$Q_{\rm jet} = \frac{\sum_i Q_i \cdot p_T^i \cdot (1+\mathcal{P}_{\rm trk}^i)}{\sum_i p_T^i \cdot (1+\mathcal{P}_{\rm trk}^i)}, \tag{6.8}$$

where the index i runs over all the tracks in the selected jet, and Q_i, p_T^i, and $\mathcal{P}_{\rm trk}^i$ are

the charge, p_T, and NN probability of each track. Due to the inclusive nature of jet reconstruction, the jet charge tagger has very high efficiency but lower dilution.

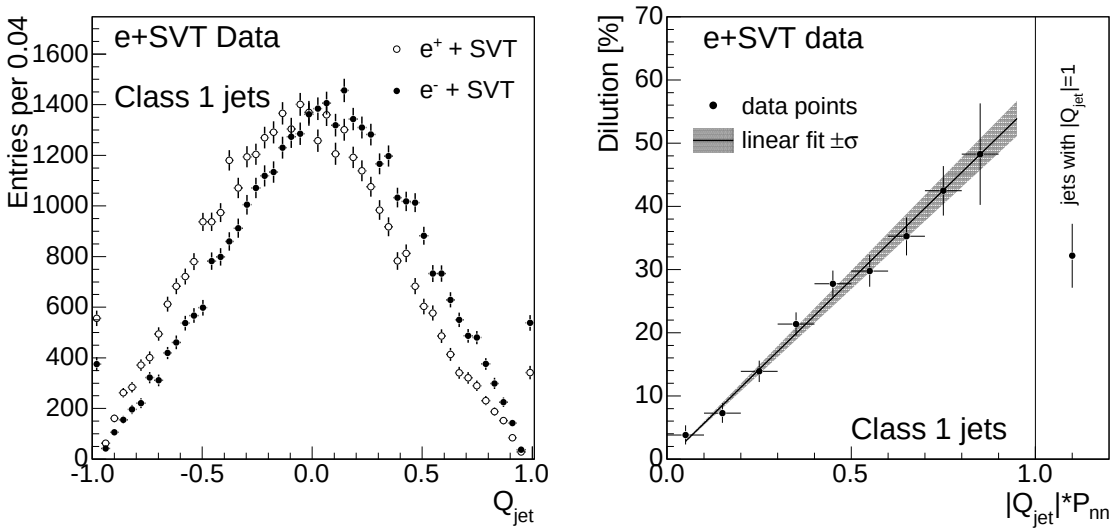

Figure 6-3: Representative jet charge tagger results in the e+SVT data. Left: distributions of measured jet charge for B (open) and $\overline{B}$ (solid) mesons. Right: the associated variation of measured dilution.

Tagged jets are split into three groups to better utilize the statistical power of the algorithm. Class 1 tags are `SECVTX` jets whose secondary vertex has a minimum displacement significance of $L_{\rm xy}/\sigma_{L_{\rm xy}} > 3$. These jets have the highest purity and best tagging dilution. Class 2 includes all jets not contained in Class 1 but which have at least one track in the jet with track probability $\mathcal{P}_{\rm trk} > 50\%$. Class 3 jets are the remainder of tagged jets which are not contained in Class 1 but which have $\mathcal{P}_{\rm trk} < 50\%$ for all tracks. The jet charge tagging classes are mutually exclusive.

Tagger dilution is parameterized with respect to the quantity $|Q_{\rm jet}| \cdot \mathcal{P}_{\rm nn}$. Figure 6-3 shows representative results in the e+SVT data, including separate distributions of measured jet charge for B and $\overline{B}$ mesons and the associated parameterization of dilution. The overall tagger performance quantities are compiled with those of the other tagging algorithms in Table 6.1.

6.2.3 Opposite Side Kaon Tagger

Decays of b quarks are likely to contain kaons via the Cabibbo-favored decay chain $b \to c \to s$. The opposite side kaon tagger (OSKT) [100] attempts to find such charged kaons and thereby deduce the flavor of the original b. The algorithm uses particle identification from the TOF detector and COT $\mathrm{d}E/\mathrm{d}x$ to discriminate kaons from the large background of pion tracks. The technique relies on a likelihood ratio similar to that of the same side kaon tagger of Section 6.3, in combination with impact parameter measurements to identify displaced tracks.

Because of the difficulty in identifying kaons within the large hadronic background, the dilution and efficiency of the OSKT method are both modest. The resulting performance is compiled with that of the other tagging algorithms in Table 6.1.

6.2.4 Neural Network Combination of OST

The opposite side tagger algorithms discussed above may use common attributes from the same events, such that their tag decisions are not mutually exclusive. Previous versions of CDF mixing analyses [31] used an exclusive OST combination based on a fixed hierarchy of average tag dilution. If multiple tags were available for an event, muon tags were accepted over electron tags, which were in turn preferred over jet charge tags.

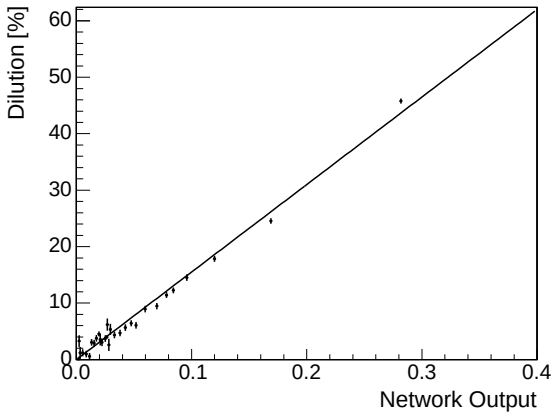

Figure 6-4: Linear dependence of the combined opposite side tagger dilution on neural network output.

However, multiple tag decisions may be combined into one by use of a neural network. In addition to simplifying the framework for handling multiple taggers, this technique is able to exploit the correlations of various decisions to produce a single tag of higher average dilution than the exclusive hierarchy [101]. The NN is again trained on ℓ+SVT data, with the set of input variables including the separate tagging decisions, predicted dilutions, and various kinematic quantities. The dilution of the combined tag has a linear dependence on the NN output discriminant, as depicted in Figure 6-4.

The performance of the separate OST algorithms and their neural network combination in the ℓ+SVT data are summarized in Table 6.1. The $\epsilon\mathcal{D}^2$ effectiveness of the NN-combined OST represents an approximately 20% relative improvement over than of the exclusive combination [101].

Opposite Side Tag	Efficiency ε [%]	Dilution $\langle\mathcal{D}\rangle$ [%]	Tagging Power $\epsilon\mathcal{D}^2$ [%]
Muon	4.6 ± 0.0	34.7 ± 3.5	0.58 ± 0.02
Electron	3.2 ± 0.0	30.3 ± 0.7	0.29 ± 0.01
Jet Charge	95.5 ± 0.1	9.7 ± 0.2	0.90 ± 0.03
Kaon	18.1 ± 0.1	11.1 ± 0.9	0.23 ± 0.02
Combined NN	95.8 ± 0.1	12.7 ± 0.2	1.55 ± 0.04

Table 6.1: Performance of various opposite side tagging algorithms on the ℓ+SVT sample, including the neural network combination. All uncertainties are statistical only.

6.2.5 OST Calibration in Hadronic Data

One advantage of opposite side tagging is that the physical processes on which it relies are independent of the details of the trigger-side B meson. The combined tagger may therefore be applied to either semileptonic or hadronic decays of any B species. However, kinematic differences between the ℓ+SVT sample and the hadronic data may lead to different overall dilutions than are predicted by the development sample. Such differences are accounted for by application of a global scale factor for the tagging algorithm, so that the predicted dilution $\mathcal{D}_i$ for each event i is transformed by

$$\mathcal{D}_i \to S_{\mathcal{D}} \cdot \mathcal{D}_i. \tag{6.9}$$

The scale factor $S_{\mathcal{D}}$ is derived for the hadronic data by analysis of flavor asymmetry in light B decay channels [84,106]. The constant flavor asymmetry of the non-mixing B^+ data is a direct indication of the tag dilution and is particularly valuable. Measurement of the B^0 oscillation frequency Δm_d, which is slow in comparison with the Δm_s, also allows the likelihood framework to compare the predicted dilution values against the best fit of the data.

The dilution calibration is extracted from a simultaneous fit of as many as 10 fully reconstructed B^0 and B^+ modes, including $B^{0,+} \to D(3)\pi$, $B^{0,+} \to J/\psi K^{*0,+}$, and $B^0 \to D^{*-}(3)\pi$ with $\overline{D}^0 \to K(3)\pi$. The analysis involves mass and lifetime fits similar to those of Chapters 4 and 5 but which do not include partially reconstructed decays, followed by the fit of dilution scale factors and Δm_d. The likelihood for flavor asymmetry is conceptually identical to the fully reconstructed case discussed below in Section 6.5. The fit also uses a kinematic same side tagging algorithm, as introduced in Section 6.3, to better resolve the true Δm_d of the data.

For transfer of scale factors to the full 1 fb^{-1} of B_s data, the calibration is separately performed in three internal data-taking periods to account for minor differences in detector configuration. The corresponding scale factor is applied to each period of B_s data.

The combined measurement of the B^0 oscillation frequency is $\Delta m_d = 0.523 \pm 0.024$ (stat) ps^{-1}, which is consistent with the world average value [2]. The overall OST dilution scale factor is $S_{\mathcal{D}} = 1.10 \pm 0.02$ (stat), indicating that the event-by-event dilution of the combined tagger is slightly underestimated. Values of $S_{\mathcal{D}}$ very close

to this one are applied in the B_s analysis of Chapter 7.

6.3 Same Side Kaon Tagging

The same side tagging strategy is based on correlations between b quark flavor and the charge of associated particles [107]. In contrast to opposite side methods, which are independent of the trigger-side hadronization, the same side approach must be customized for each B species and the different particles associated with their production. Given a b quark that hadronizes to form a $\overline{B}^0$ meson, the remaining d quark may combine with a $\overline{u}$ to produce a π^-. Analogously, B^- mesons are often produced alongside π^+, while K^- mesons are associated with $\overline{B}_s^0$. Figure 6-5 illustrates the simplest examples of particles produced in conjunction with $\overline{B}^0$, B^-, and $\overline{B}_s^0$. An additional source of B mesons and correlated particles is the P-wave decay of orbitally excited B^{**} mesons.

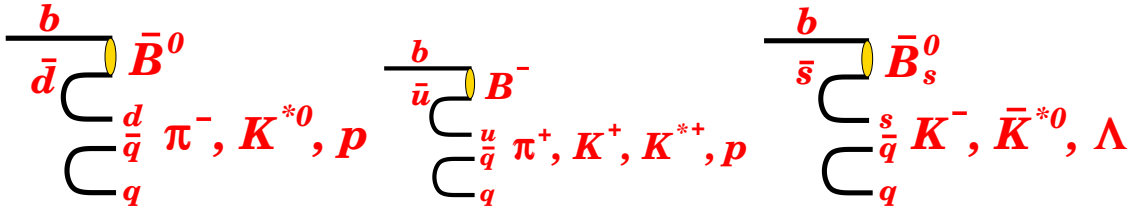

Figure 6-5: Charge correlation of B mesons with fragmentation particles. Same side tagging of B_s requires kaons, while the light B species are associated with differently charged pions.

Same side tagging (SST) has several advantages over the opposite side approach. The leading fragmentation track occupies the same η-ϕ region of the detector as the triggered B, making it likely to be reconstructed as well and yielding a high tagging efficiency. This is in contrast with the OST method, whose high efficiency stems from very inclusive jet reconstruction with low dilution. Additionally, since the search region for fragmentation tracks is restricted to the vicinity of the trigger B, the algorithm is robust against background tracks from the underlying event or multiple $p\overline{p}$ interactions. Low background content translates directly into higher dilution. Finally, SST does not suffer from the possibility of opposite side neutral B mixing, which is inherent in all OST algorithms. Because of these beneficial attributes, the effectiveness of the same side tagger for B_s decays is significantly higher than that of the combined OST. Same side tagging dominates the overall flavor tagging power, making it critical for the significance of the B_s oscillation signal.

Despite its numerous advantages, same side tagging has a fundamental weakness for its application to the B_s data. Because the fragmentation process depends on the B species, the tagger cannot simply be developed and calibrated in B^0 and B^+ data. Development of the Same Side Kaon Tagger (SSKT) for measurement of B_s flavor must be performed in samples of PYTHIA Monte Carlo [89, 103], which simulates the

entire $p\overline{p}$ collision event. The validity of the fragmentation model is supported by comparison of the MC samples with high-statistics B^0 and B^+ data for the most relevant variables.

Reliable selection of kaons for B_s tagging is largely dependent on the particle identification abilities of the Time-of-Flight (TOF) detector and $\mathrm{d}E/\mathrm{d}x$ information from the COT. The SSKT algorithm for this analysis uses a neural network to combine this essential information with kinematic variables and thereby produce high tag dilutions [89]. The case of same side tagging for B^0 and B^+ mesons presents a simpler case, because the leading fragmentation particle is most often a pion. The leading tracks may be selected by kinematic requirements alone due to the predominance of pions among track particles. Application of kinematic SST methods to the light B species is documented elsewhere [102], and the discussion below is primarily focused on the use of SSKT for B_s decays.

6.3.1 Preselection of Same Side Tag Candidates

For each event, the set of potential tagging tracks is selected from the tracks defined in Section 3.2.1.

- Tracks must have $p_T > 450$ MeV/c to limit the charge asymmetry of track reconstruction efficiency, and pseudorapidity $|\eta| < 1.0$ to maintain good momentum and impact parameter resolution. Tracks which are used in the B candidate reconstruction are explicitly excluded.

- Tracks must inhabit the angular cone within $\Delta R \leq 0.7$ of the reconstructed B momentum. This requirement removes tracks which are unlikely to be of interest, and also maintains an exclusive set of tracks from those eligible for use in an opposite side tag algorithm.

- Tracks must have impact parameter significance $d_0/\sigma_{d_0} < 4$, because only prompt particles from the $p\overline{p}$ collision are useful.

- The presence of particles from other $p\overline{p}$ interactions is diminished by requiring tracks to be within $|\Delta z| < 1.2$ cm of the B meson.

- Lepton tracks are rejected via the same *anti*-selection of the lepton likelihoods [35] used for hadronic B reconstruction in Section 3.3. Photon conversions are rejected by their associated criteria.

The resulting set of tracks for each event forms the pool from which the tagging track must be selected, as discussed below.

Based on the preselected tag candidates, the agreement between the PYTHIA Monte Carlo simulation and data for the light B species is studied. The set of variables used for the comparison is intended to test the validity of the MC fragmentation model as thoroughly as possible. These variables include B meson p_T, the multiplicity of preselected tag candidates, opening angle ΔR between the track and B momentum, and the tracks' combined kaon identification discriminant $CLL(K)$ described

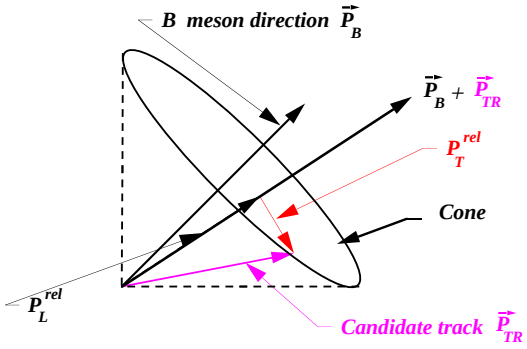

Figure 6-6: Definition of kinematic quantities for same side tagging.

in Appendix E. Comparisons are also made for kinematic variables of the tag tracks such as p_T, $p_L^{\rm rel}$, and $p_T^{\rm rel}$, whose definitions are constructed in Figure 6-6. In general, good data-MC agreement is observed for fully reconstructed B_s candidates and for the large control samples of fully reconstructed B^0 and B^+ decays. Figure 6-7 shows prominent examples of the comparisons for B_s decays.

6.3.2 SSKT Algorithms

As indicated in the upper-left plot of Figure 6-7, the preselection requirements for tag candidates often accept more than one track. This situation encompasses $\sim 40\%$ of B events that have at least one tag track. In the case that the charges of all tagging tracks agree, the flavor decision is obvious. However, for about 65% of events with multiple tracks, the charges disagree. This ambiguity requires the use of an algorithm to select the best track. Moreover, regardless of track agreement, an algorithm is necessary to provide the desired estimate of dilution. In the development of the final same side kaon tagger, the following algorithms are explored:

1. The "maximum $p_L^{\rm rel}$" algorithm selects the tagging track with the largest component of longitudinal momentum along the B direction.

2. The "maximum $CLL(K)$" algorithm selects the tagging track which is measured as most likely to be a kaon.

3. The neural network (NN) algorithm combines the particle identification information encapsulated in $CLL(K)$ with various kinematic quantities, including $p_L^{\rm rel}$, $p_T^{\rm rel}$, p_T, and ΔR. The NN is trained with pure samples of PYTHIA B_s Monte Carlo events. Kaons of the correct charge correlation are used for the signal training, and other particle species and incorrect kaons are used for the background training. The tagging track with the largest NN output discriminant is selected for the tag. This algorithm is used for the final B_s oscillation analysis.

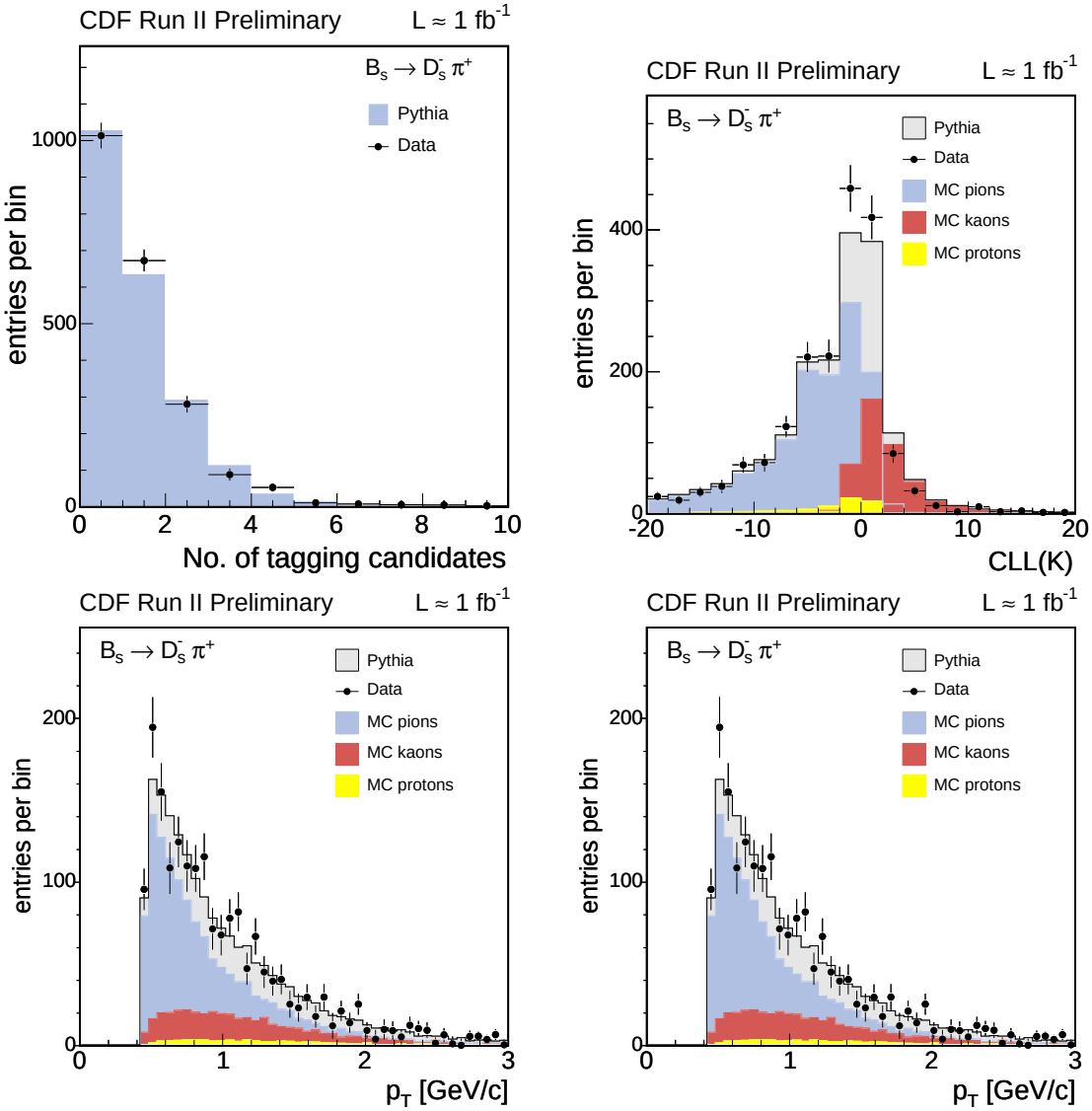

Figure 6-7: Comparison of PYTHIA Monte Carlo simulation with data for fully reconstructed B_s decays. Top-left: total number of tagging candidates. Top-right: Kaon discriminant $CLL(K)$ for all candidate tracks. Bottom-left: p_T for all candidates. Bottom-right: p_T for candidate tracks with $CLL(K) > 1$, dominated by kaons. MC distributions are normalized to match the number of sideband-subtracted data entries.

The dilution associated with each method may be parameterized with respect to any variable by dividing the Monte Carlo events into appropriate bins and examining the dependences. The maximum $p_L^{\rm rel}$ method parameterizes dilution in terms of the p_T of the selected track, while the kaon identification method uses $CLL(K)$. The NN combination measures dilution with respect to its NN discriminant n using the cubic

form

$$\mathcal{D}(n) = c_0 + c_1 n + c_2 n^2 + c_3 n^3. \tag{6.10}$$

The dilution is parameterized separately for the "agreement" case, in which multiple tag candidates agree *or* only one track is found, and the "disagreement" case of multiple tracks with different charge. The dilution of the disagreement case is observed to be systematically lower by ~ 0.2 than the case of agreement [89].

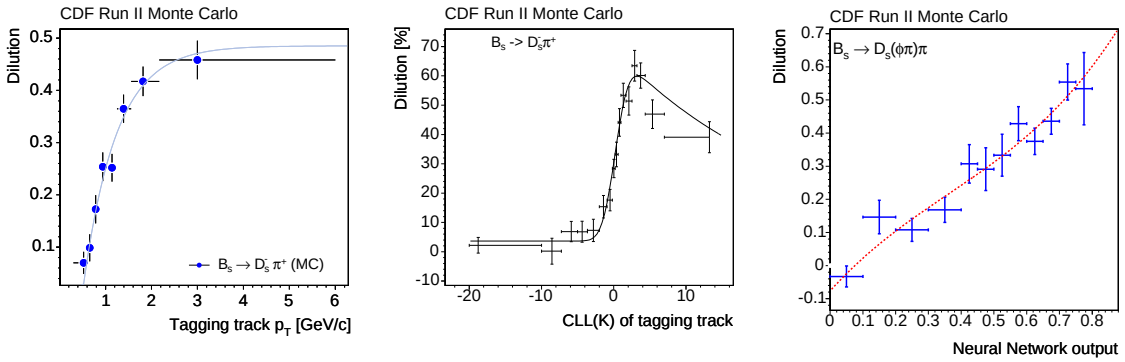

Figure 6-8: Dilution dependences of various same side tagging algorithms for events with one of more tag candidates of agreeing charge. Left: Dilution of the maximum p_L^{rel} algorithm with respect to tagging track p_T. Center: Dilution of the maximum $CLL(K)$ algorithm with respect to $CLL(K)$. Right: Dilution of the combined neural network algorithm with respect to the NN discriminant.

Figure 6-8 shows the parameterization of the dilution for the agreement cases of the three algorithms. The dilution of the p_L^{rel} algorithm increases sharply as track p_T rises to a few GeV/c, but its maximum dilution is less than 50%. The particle identification algorithm has dilution near zero for tracks with low $CLL(K)$, which are overwhelmingly pions and protons, but it rises sharply at $CLL(K) > 0$, for which the tracks are likely to be kaons. The dilution suffers at very high $CLL(K)$ because, as noted in Appendix E, particle ID performs best at low track momenta. Although these tracks are very likely to be kaons, the low-momentum regime is more heavily contaminated with background tracks. The NN algorithm combines the performance of these separate methods and produces dilution that rises steadily with the output discriminant.

The performance of the three algorithms is evaluated in both data and Monte Carlo events for B^0,B^+, and B_s decays. Both dilution and efficiency can be measured in the light B data, while only efficiency is compared for the B_s data. The agreement between the MC predictions and the data is good [89], which demonstrates that processes relevant to same side tagging are well-simulated in MC samples across all B species.

Scale Factor for SSKT Dilution

As a final calibration, the same side tagger also receives a global scale factor on the dilution $\mathcal{D}_i$ predicted for each event i. The modification is of the form $\mathcal{D}_i \to S_{\mathcal{D}} \cdot \mathcal{D}_i$, which is analogous to the scale factor of the combined opposite side tagger discussed in Section 6.2. The SSKT scale factor is derived from a modified B_s MC sample in which the oscillation frequency has been fixed to zero. Application of the parameterized dilution results in a constant flavor asymmetry which directly indicates the true sample dilution, and the scale factor $S_{\mathcal{D}}$ is extracted as a fitted parameter in the likelihood term of Equation (6.6). Like other calibrations in the analysis, the scale factor is extracted separately for the three internal data-taking periods. The fitted values range from 0.95 - 1.0 [89].

Extensive systematic uncertainties are evaluated for SSKT dilution to account for the fact that its calibration relies on MC simulation [89]. These uncertainties are important for the oscillation amplitude but not the final Δm_s measurement.

6.3.3 SSKT Effectiveness

Table 6.2 lists the overall $\epsilon\mathcal{D}^2$ effectiveness of the three same side tagging algorithms, as evaluated in PYTHIA Monte Carlo samples of fully reconstructed $B_s^0 \to D_s^- \pi^+$ decays. The neural network combination provides the best performance and is selected for application in the analysis.

Algorithm	$\epsilon\mathcal{D}^2$
Max. p_L^{rel}	$2.8\ ^{+0.6}_{-0.8}\%$
Max. $CLL(K)$	$2.9 \leftrightarrow 4.0\ ^{+0.9}_{-1.2}\%$
NN combination	$3.1 \leftrightarrow 4.3\ ^{+1.0}_{-1.4}\%$

Table 6.2: Effectiveness of the same side tagging algorithms in PYTHIA Monte Carlo samples of fully reconstructed $B_s^0 \to D_s^- \pi^+$ decays, including systematic uncertainties.The variation in effectiveness of the $CLL(K)$ and NN methods is due to changes in the performance of the TOF and $\mathrm{d}E/\mathrm{d}x$ particle identification tools. The NN combination algorithm is selected for the analysis.

6.4 Tagging for Partially Reconstructed Signals

Previous chapters devote much attention to inclusion of the new partially reconstructed signal channels. Due to the momentum lost to their unreconstructed particles, these decays require customized models in the mass and proper time spaces. Because flavor tagging is critical to the data's ability to resolve an oscillation signal, it is important to consider the applicability of the tagging tools to these signal channels.

All tagging methods are contingent on the idea that the B flavor at its *decay* may be directly observed in the charges of the resulting particles, and the tagger algorithm

need only be concerned with the initial-state flavor. This approach applies equally well to partially reconstructed signals. In each of the partially reconstructed B decay topologies, the daughter D meson remains fully reconstructed, and its charged decay particles completely specify the decaying B flavor.

The only issue for application of the taggers is with regard to potential interference of the unreconstructed particles with the tagging algorithms. The true concern is not that the modeling of the tagger may need to be adjusted for the partially reconstructed signals, but rather that unforeseen effects could bias the resulting dilution, which would cause the calibrated dilutions of Sections 6.2 and 6.3 to be systematically inaccurate and would bias the eventual oscillation amplitude.

Opposite Side Tagging

The applicability of the opposite side tagger is fairly transparent. Preselection of tracks for this algorithm requires that they be outside the cone of opening angle $\Delta R < 0.7$ around the reconstructed B momentum. Since the B system is boosted by mean $\langle\beta\gamma\rangle \sim 2$ and the lost particles are demonstrated in Chapter 5 to be relatively soft within the system, these tracks are very unlikely to participate in any OST decision. Moreover, the unreconstructed particles for the dominant partially reconstructed B_s and B^0 are **neutral**, implying that the very attribute which prevents them from being measured also protects their B decays from tagger interference. This attribute is essential for the discussions below. As formalized in the construction of the signal likelihood in Section 6.5, the neural network OST algorithm is applied directly to both fully and partially reconstructed signal channels with shared dilution and efficiency parameters.

Same Side Tagging

The same side tagger presents a greater potential challenge to the equivalence of the signal channels because the unreconstructed particles coincide with the region of tagging tracks. However, a great advantage is again provided by the fact that the missing particles of the B_s and B^0 decays are neutral. With the assumption that the lost π^0/γ cannot interfere with the SSKT algorithm, the externally-calibrated dilution may be applied directly to these components. In contrast, the "physics background" of partially reconstructed $b \to D\pi X$ decays encompasses numerous subdominant channels, some of which involve the loss of *charged* particles that may be incorrectly selected by the tagger. It is expected that the physics background component measures a different dilution and/or efficiency than those of the signal, and the likelihood must be constructed such that these parameters are independent. A similar consideration is made for the special case of the B^+ data, as noted below. Nevertheless, given that the signal channels with lost *neutral* particles are assigned correct fractions in the likelihood, the SSKT dilution remains correctly calibrated for the partially reconstructed signals and produces an unbiased oscillation signature. As formalized in Section 6.5, the neural network same side kaon tagger is applied directly to both fully and partially reconstructed B_s signal channels with shared dilution and

efficiency parameters.

Special Case for B^+ Data

As has been noted at each stage of this analysis, the data sample that is selected for $B^+ \to \overline{D}^0\pi^+$ contains the off-species contribution of partially reconstructed $B^0 \to D^{*-}\pi^+$ decays, which has no analogous component in the B^0 and B_s samples. The treatment of this component in the measurement space of mass and untagged proper time is qualitatively identical to the modeling of the other partially reconstructed signals: it has a similar mass template, constrained fraction, κ-factor distribution, etc.

However, application of flavor tagging requires that the treatment of the extra $B^0 \to D^{*-}\pi^+$ component diverges significantly. Unlike the "standard" partially reconstructed $B \to D^*\pi$ and $B \to D\rho$ topologies which are of the same B species as the main fully reconstructed decay, this off-species contribution arises because of the loss of a *charged* particle. Therefore, while simple application of the opposite side tagger is justified due to its isolation from the B, it must be anticipated that the same side tagger performance will be modified by the additional charged particle. The situation is similar to that of the physics background noted above, except that the modifications to the expected efficiency and dilution arise from a single decay mode. Specifically, the lost pion from the D^{*-} decay is very soft and therefore collinear with the D^{*-}, which in turn is close to the B momentum. This configuration is sketched in Figure 3-3. The SST algorithm optimized for light B decays uses only kinematic information, as opposed to the kaon identification methods of the tagger for B_s decays, so it is expected that this soft pion may be misidentified as the tagging track in a non-negligible fraction of events. Based on this pool of additional tracks, the B^0 component of the B^+ data should measure an increased SST efficiency and modified dilution. In order to avoid such complications, the same side tagger is not applied to B^+ data in the asymmetry fits of Section 6.5.5.

It must be emphasized that this component is a special case for the B^+ data and is *completely irrelevant* to the B_s analysis or the parallel B^0 development sample. The results from the B^+ data are reported below, but only as a supplement to the B^0 fits. Because of its lack of parallelism with the simpler B^0 and B_s flavor-tagging behavior, the relevance of the B^+ data as a development sample is diminished at this stage of the analysis.

6.5 Likelihood Expansion for Flavor Tagging

The flavor tagging information returned by each tagging algorithm is encapsulated in two variables:

- the tag *decision* ξ, where $\xi = \pm 1$ corresponds to the unmixed/mixed (or right sign/wrong sign) decision, and $\xi = 0$ indicates that no tag was found;
- the tag *dilution* $\mathcal{D}$, which the algorithms estimate on an event-by-event basis to reflect the confidence of a tag decision.

Therefore, the two major tagging algorithms of this analysis produce four observables for every B candidate: ξ_o and $\mathcal{D}_o$, from the opposite side tagger; and ξ_s and $\mathcal{D}_s$, from the side side tagger. The likelihood terms of each sample component must be expanded to include these variables. The signal and background contributions are most generally divided into "oscillating" and "non-oscillating" components. The formalism of these cases is developed immediately below, followed by explicit specification of the sample components in Sections 6.5.3 and 6.5.4. Additionally, the decisions of the two taggers are combined whenever both are available; the case of a single tagger is developed first, followed by the generalization to multiple taggers. Application of the taggers is summarized in a table in Section 6.5.5.

6.5.1 Single Tagger Case

The analysis framework readily incorporates new measurement spaces via expansion of the likelihood. Given the additional observables ξ and $\mathcal{D}$, the most general form of each signal or background component is the joint PDF

$$\mathcal{P} = \mathcal{P}^{\text{mass}} \cdot \mathcal{P}^{\text{time}} \cdot \mathcal{P}^{\xi} \cdot \mathcal{P}^{\mathcal{D}} \cdot \mathcal{P}^{\sigma_t}, \tag{6.11}$$

where the experimental time variable ct is again abbreviated to t for brevity. Given the connections between t, ξ, and $\mathcal{D}$, the PDF for each sample component is best expressed as

$$\begin{aligned} \mathcal{P}(m, t, \sigma_t, \xi, \mathcal{D}) &= \mathcal{P}(m) \cdot \mathcal{P}(t, \xi | \mathcal{D}, \sigma_t) \cdot \mathcal{P}(\mathcal{D}) \cdot \cancelto{1}{\mathcal{P}(\sigma_t)} \\ &= \mathcal{P}(m) \cdot \mathcal{P}(t | \xi, \mathcal{D}, \sigma_t) \cdot \mathcal{P}(\xi | \mathcal{D}, \sigma_t) \cdot \mathcal{P}(\mathcal{D}), \end{aligned} \tag{6.12}$$

where, as noted in Section 5.5, the $\mathcal{P}^{\sigma_t}$ distributions of proper time resolution specific to each sample component are so similar as to cancel out for these datasets. The mass space remains decoupled from the other observables. The final component factors are individually normalized probability densities which have definite conceptual interpretations:

- $\mathcal{P}(m)$ describes how the sample component is distributed in the mass space. Effectively, it weights the component's proper time and flavor contributions as a function of candidate mass.
- $\mathcal{P}(t|\xi, \mathcal{D}, \sigma_t)$ is the component's probability density in time, given each candidate's tag decision, associated dilution, and proper time smearing.
- $\mathcal{P}(\xi|\mathcal{D}, \sigma_t) \to \mathcal{P}(\xi)$ is the probability for an event of this component to receive tag decision ξ, which is modeled independently of $\mathcal{D}$ and σ_t. This factor is readily interpreted in terms of *tagger efficiency* ϵ below.
- $\mathcal{P}(\mathcal{D})$ is the distribution of dilutions estimated by the tagger, specifically compiled for each sample component. This factor is required by conditional probability [96]. Unlike proper time resolution, which is similar enough among the signal and background components that $\mathcal{P}(\sigma_t)$ can be cancelled from the overall

likelihood, the dilution distributions differ between some components and would bias the oscillation amplitude if not modeled.

Each signal and background component discussed in previous chapters has a joint PDF with factors corresponding to these terms. The decoupled mass term remains the same as developed in Chapter 4, while the untagged proper time characterization of Chapter 5 is expanded to account for the remaining terms. The sample components may be divided into the oscillating category, which are neutral B mesons that undergo flavor mixing, and the non-oscillating category, which includes B^+ decays and most backgrounds. The oscillating case may be reduced to the non-oscillating cases and so is discussed first.

Oscillating Signals

Section 6.1 introduces that the flavor state probability of a neutral B meson with imperfect tagging is described by

$$\mathcal{P}(t|\mathcal{D}) = \frac{1 \pm \mathcal{D}\cos(\Delta m t)}{2}, \tag{6.13}$$

where Δm is the appropriate oscillation frequency and t is the exact decay time without detector effects. Incorporating this relation into the model of decay time $\mathcal{P}^{\rm ct}(t|\sigma_t)$ from Chapter 5, and using the tag decision variable ξ, the time-flavor PDF is expressed as

$$\mathcal{P}^{\rm osc}(t|\xi,\mathcal{D},\sigma_t) = \frac{1}{\mathcal{N}}\left[\left(\frac{1}{\tau}e^{-\frac{t}{\tau}}\theta(t)\cdot\frac{1+\xi\mathcal{D}\cos(\Delta m t)}{(1+|\xi|)}\right)\otimes\mathcal{G}(t|\sigma_t)\otimes F(\kappa)\right]\cdot\varepsilon(t). \tag{6.14}$$

In the $\xi = 0$ case that no tag decision is available, the expression reduces to the non-flavor-specific PDF of the lifetime fits, while the $\xi = \pm 1$ cases produce the cosinusoidal flavor mixing. The ideal distributions of exponential decay and flavor oscillation are both smeared by the Gaussian detector resolution, and the time-dependent efficiency applies the "sculpting" of the selection bias. The convolution of the Gaussian $\mathcal{G}(t|\sigma_t)$ with the term proportional to $(\exp(-t/\tau)\cdot\cos(\Delta m t))$ is performed analytically [34]. The $F(\kappa)$ convolution applies only to the partially reconstructed signal components and is again performed numerically using the histograms of Section 5.3. A fundamental concept for this convolution is that both the lifetime τ *and* the oscillation frequency Δm participate in the momentum-smearing:

$$\sum_{\kappa} \quad \tau \to \tau/\kappa \quad \Delta m \to \kappa\Delta m,$$

which demonstrates that the widths of the κ-factor distributions are directly tied to sensitivity to the Δm parameter.

The time-flavor PDF of Equation (6.14) must be weighted to properly account for the fractions of the sample component which are tagged or untagged. The probability to measure decision ξ is simply the overall tagger efficiency ϵ, which enters the joint

PDF as

$$\mathcal{P}(\xi) = \epsilon|\xi| + (1-\epsilon)(1-|\xi|) = \begin{cases} \epsilon & \xi = \pm 1 \\ (1-\epsilon) & \xi = 0 \end{cases} \tag{6.15}$$

No additional normalization factors are necessary. The factor $\mathcal{N}$ in Equation (6.14) is the same normalization of Section 5.2.

The final factor $\mathcal{P}(\mathcal{D})$ for the probability density of measured dilutions must be extracted from the data. This distribution enters the model as normalized histogram whose derivation via sideband-subtraction is discussed for the signal components in Section 6.5.3.

Non-oscillating Components

Sample components which do not mix, either because they represent charged B^+ mesons or because they are backgrounds, are characterized by a simplification of the oscillation formalism. The non-oscillating condition is achieved by setting $\Delta m = 0$, such that the cosinusoidal time dependence is removed. The time-flavor PDF for signals follows directly from Equation (6.14):

$$\mathcal{P}_{\rm sig}^{\rm non-osc}(t|\xi,\mathcal{D},\sigma_t) = \frac{1}{\mathcal{N}}\left[\left(\frac{1}{\tau}e^{-\frac{t}{\tau}}\theta(t)\cdot\frac{1+\xi\mathcal{D}}{(1+|\xi|)}\right)\otimes\mathcal{G}(t|\sigma_t)\otimes F(\kappa)\right]\cdot\varepsilon(t), \tag{6.16}$$

which represents a constant flavor asymmetry which arises only because of the imperfect tagger dilution.

In the case of non-oscillating background components, the flavor asymmetry remains as above but the time-dependent aspects of the model are simplified to the empirical models of Section 5.4. In the general background case of a smeared-exponential, the time-flavor PDF takes the form of

$$\mathcal{P}_{\rm bkg}^{\rm non-osc}(t|\xi,\mathcal{D}) = \left[\left(\frac{1}{\tau}e^{-\frac{(t-\delta)}{\tau}}\theta(t-\delta)\right)\otimes\mathcal{G}(t,\sigma)\right]\left[\frac{1+\xi\mathcal{D}}{(1+|\xi|)}\right], \tag{6.17}$$

where the constant flavor asymmetry again acts as a simple multiplicative factor. The τ, δ, and σ background parameters are determined in Chapter 5 and are expanded to include multiple smeared-exponentials, as appropriate. In contrast to the signal case, where dilution is estimated on an event-by-event basis, most background components are empirically modeled with a simple parameter for average dilution.

Even in a component for which this simplification is made, the probability density $\mathcal{P}(\mathcal{D})$ of its measured dilutions is still required to satisfy the overall likelihood normalization. The dilution PDFs are again extracted from the data and modeled as normalized histograms. The combinatorial background has different dilution distributions than the signal components, as shown in Section 6.5.3 below.

The structure of the tagging efficiency factors is identical to Equation (6.15), which maintains the overall normalization between the tagged and untagged PDFs. However, the tagger efficiencies for background components are always represented by parameters which are independent of the signal efficiencies.

6.5.2 Combination of Taggers

For simplicity, the above discussion assumes the application of a single tagging algorithm. However, it is advantageous to use the independent information of both the same-side and opposite-side taggers, whenever possible. This scenario is described by the same joint PDF of Equation (6.12), except that the flavor-specific terms must be expanded to include the two pairs of $\{\xi, \mathcal{D}\} \otimes \{\mathrm{OST}, \mathrm{SST}\}$ observables. The flavor state probability for oscillations generalizes to

$$\mathcal{P}^{\mathrm{osc}}(t|\xi_o, \xi_s, \mathcal{D}_o, \mathcal{D}_s) \propto \frac{(1 + \xi_o\mathcal{D}_o\xi_s\mathcal{D}_s) + (\xi_o\mathcal{D}_o + \xi_s\mathcal{D}_s)\cos(\Delta m t)}{(1 + |\xi_o|)\,(1 + |\xi_s|)}, \tag{6.18}$$

which is substituted into the product with the exponential decay and the convolutions of Equation (6.14). The expression reduces to the single-tagger case if $\xi_o = 0$ or $\xi_s = 0$, and it again reduces to the non-flavor-specific proper time model if both tags are unavailable. Equation (6.18) is structured such that the dilutions of the taggers give weight to their decisions appropriately: two tags which are in agreement add their weight constructively, while disagreeing tags destructively cancel the associated B candidate's contribution in measuring the oscillation.

The simplification to non-oscillating cases is equally intuitive. The mixing frequency is again fixed to $\Delta m = 0$ in Equation (6.18), and the expression factors to

$$\begin{aligned}\mathcal{P}^{\mathrm{non-osc}}(t|\xi_o, \xi_s, \mathcal{D}_o, \mathcal{D}_s) &\propto \frac{(1 + \xi_o\mathcal{D}_o\xi_s\mathcal{D}_s) + (\xi_o\mathcal{D}_o + \xi_s\mathcal{D}_s)}{(1 + |\xi_o|)\,(1 + |\xi_s|)} \\ &= \left(\frac{1 + \xi_o\mathcal{D}_o}{1 + |\xi_o|}\right) \cdot \left(\frac{1 + \xi_s\mathcal{D}_s}{1 + |\xi_s|}\right),\end{aligned} \tag{6.19}$$

which are simply the uncorrelated probabilities for measuring the correct or incorrect flavor, given a decision from each tagger.

In both the mixing and non-mixing cases, the $\mathcal{P}(\xi)$ term for tagging efficiency is a generalization of the single-tag case:

$$\mathcal{P}(\xi_o, \xi_s) = \begin{cases} \epsilon_o\epsilon_s & \text{for } \xi_o = \pm 1, \quad \xi_s = \pm 1 \\ \epsilon_o(1 - \epsilon_s) & \text{for } \xi_o = \pm 1, \quad \xi_s = 0 \\ (1 - \epsilon_o)\epsilon_s & \text{for } \xi_o = 0, \quad \xi_s = \pm 1 \\ (1 - \epsilon_o)(1 - \epsilon_s) & \text{for } \xi_o = 0, \quad \xi_s = 0. \end{cases} \tag{6.20}$$

The probability density $\mathcal{P}(\mathcal{D}_o, \mathcal{D}_s)$ of measured dilutions factors to $\mathcal{P}(\mathcal{D}_o)\mathcal{P}(\mathcal{D}_s)$. As in the single-tagger case, this PDF is implemented as the product of two normalized histograms extracted from data, with each tagger providing its dilution distribution.

The final multiple-tagger expansion of the likelihood requires that the appropriate proper time PDF of Chapter 5 be augmented by the substitution

$$\mathcal{P}^{\mathrm{ct}}(t|\sigma_t) \to \mathcal{P}^{\mathrm{osc/non}}(t|\xi_o, \mathcal{D}_o, \xi_s, \mathcal{D}_s, \sigma_t) \cdot \mathcal{P}(\xi_o, \xi_s) \cdot \mathcal{P}(\mathcal{D}_o, \mathcal{D}_s). \tag{6.21}$$

6.5.3 Signal Components

As noted above, the flavor form of the oscillating and non-oscillating components is similar, being separated only by a non-zero or zero Δm value. Aside from issues of shared tagging parameters, the signal PDFs of the neutral B mesons and charged B^+ mesons are divided in this fashion.

The non-flavor-specific attributes of the joint PDF remain unchanged for each signal component, including the mass templates of Chapter 4 and the selection bias $\varepsilon(ct)$ and κ-factor distributions of Chapter 5, as applicable.

Application of the taggers to signal components is summarized in a table in Section 6.5.5.

B^0 and B_s Signals

As argued in Section 6.4, the fact that the lost particles of the partially reconstructed signal channels are neutral allows both tagging algorithms to be applied to all signals. Both the predicted dilution value for each event and the overall tagger efficiencies are expected to be consistent between the signals. The joint PDF for each of the neutral B signal components is modeled as in Equations (6.21) and (6.14), with the same event-by-event dilution and the same Δm parameter shared among the decay channels. Additionally, a global scale factor parameter for each tagger is applied to its predicted dilution via the substitution $\mathcal{D} \rightarrow S_{\mathcal{D}} \cdot \mathcal{D}$. For the B_s analysis of Chapter 7, these scale factors are fixed to the values from the final tagger calibrations, while fits in the light B data of Section 6.5.5 allow the parameters to float. In either case, the scale factor is applied uniformly across all signal components for all B candidates. The efficiency parameters ϵ_o and ϵ_s of Equation (6.20) are also shared among the signal components.

The normalized histograms of dilution that satisfy the $\mathcal{P}(\mathcal{D})$ requirement in the joint PDF are common to all signal components of each sample as well. The derivation of these distributions and comparison between their signal and background behavior is discussed below.

B^+ Signals

The charged B^+ signals are differentiated from the neutral B mesons by the fact that they do not undergo flavor oscillations. The "standard" set of fully reconstructed $D\pi$ and partially reconstructed $D\rho$ and $D^*\pi$ signal components of the B^+ sample are therefore described exactly as for the signals discussed above, with the exception that the $\Delta m = 0$ condition reduces the time-flavor PDFs to the constant asymmetry of Equations (6.16) and (6.19). For each tagger, these signal components share the same event-by-event dilutions, dilution scale factors, and tagger efficiency parameters.

The additional contribution of the $B^0 \rightarrow D^{*-}\pi^+$ decays in the B^+ data creates a tagging challenge because the missing particle is charged. Since the presence of this track may interfere with the same side tagger, the use of shared efficiency and dilution parameters with the other signal components would not be appropriate. Although careful studies of the B^+ data would involve independent versions of these

parameters, a simpler solution for inspection of B^+ flavor asymmetry is to neglect its SST information altogether.

In contrast to the special demands of the same side tagger, the fact that the opposite side algorithm is independent of decay channels allows it to be applied unchanged to the additional component. Since it represents B^0 decays, this component is fixed to oscillate with the world average value of Δm_d [2].

Although less parallel to the B_s sample in its partially reconstructed content than the B^0 data, the B^+ sample still provides a useful indication of OST behavior across signal components.

Cabibbo-Suppressed $B \to DK$ Mixing

Misidentified $B \to DK$ decays are treated exactly as signal events, with tagged decays contributing equally to neutral B oscillations or B^+ flavor asymmetry. A systematic uncertainty on the resulting B_s oscillation amplitude is evaluated for this choice in Appendix A, while no systematic bias is observed for the measurement of Δm_s. The Cabibbo-suppressed decays of the $B \to D^*\pi/K$ topology are also treated together, as is the case in both mass and proper time.

Dilution Distributions

The conditional probability densities of the joint PDF for each sample component require inclusion of a dilution PDF $\mathcal{P}(\mathcal{D})$. These distributions are implemented as normalized histograms which are extracted from the data. The combinatorial background dilution is readily modeled via the upper mass sideband, and the studies of Appendix F.1 show that background dilution is invariant of mass. The dilution distributions of signal components must be deduced via sideband-subtraction, in which the distribution in the signal mass region is constructed and the distribution of the pure-background sideband is subtracted from it. The weight of the subtraction is derived from the extrapolated mass model of the combinatorial background.

Rather than providing the likelihood with individual SSKT dilution templates for each signal component, all signal decays of each B species share the same normalized dilution histogram. In light of the fact that the dilution and efficiency values are also shared among the signals, this approach ensures that the distribution of dilutions is exactly as weighted in the nominal model and is properly represented in the overall likelihood. Figure 6-9 shows the normalized distributions of same-side tagger dilution for the combined signal components and for the combinatorial background of the B_s data.

The approach for providing the likelihood with OST dilution templates is simpler because of the independence of the opposite side algorithm among B species and decay channels. The sideband and sideband-subtracted distributions are extracted from the high-statistics B^0 data and applied to the B_s data as well. The signal distribution is taken only from the fully reconstructed mass region, because the OST dilution is invariant across the fully and partially reconstructed signal range.

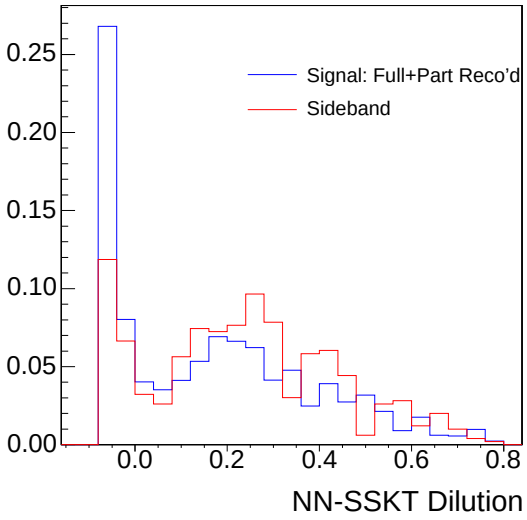

Figure 6-9: Normalized distributions of same-side tagger dilution used for the $\mathcal{P}(\mathcal{D})$ term of the joint PDFs. The combined fully and partially reconstructed B_s signal components are overlaid against combinatorial background from the upper mass sideband.

6.5.4 Background Components

Expansion of the background components to the flavor space follows the same principles as for the signal sources. Attributes of mass and proper time modeling from previous chapters remain unchanged. The general form of background flavor modeling follows from Equation (6.17) and the multiple-tagger expansion of Equation (6.19), in which the asymmetry is constant. Furthermore, the general treatment of dilution and efficiency parameters is that they are fitted empirically and independently of other background components. As in the simple proper time space, the primary consideration for background sources is that their models describe the data well enough so as not to bias the extraction of physical parameters from the signal. Application of the taggers to backgrounds is summarized in Table 6.3 below.

$b \to DX$ Physics Background

Regardless of B species, the physics background is modeled as a non-mixing component. Even though some decay channels within this component are likely to be neutral B mesons (whether B^0 or B_s), the fit to a constant flavor asymmetry collectively accounts for the time-integrated probability of their oscillations. The implementation of Equations (6.19) and (6.20) uses independent ϵ and $\mathcal{D}$ parameters for both taggers which are not shared with other sample components. The parameter $\mathcal{D}$ represents a single average dilution value which is applied to all candidates for the physics background component, as opposed to the event-by-event dilution estimates of the signals. The appropriate proper time template of Section 5.4 is again applied.

The corresponding dilution templates are modeled using those of the signal distributions for both tagger algorithms. While this assumption is safe for the OST dilution, the debris from various partially reconstructed channels in the physics background is expected to skew its SST dilution distribution to some extent. Any bias to the observed oscillation amplitude is expected to be very small because this component is responsible for less than $\sim 3\%$ of the B_s sample. This choice is supported by systematic uncertainty studies of amplitude in Appendix A, which examine the effects of neglecting the physics background component altogether. More importantly, this simplification of the dilution templates is observed to produce no bias to Δm_s in Appendix B.

Combinatorial Background

The combinatorial background is also modeled as a non-oscillating component. Like the physics background, its efficiency and dilution parameters are independent of any other component, and the dilutions represent averages for the taggers rather than using any per-event estimates. The constant flavor asymmetry term of Equation (6.20) multiplies the double smeared-exponential proper time template who parameters are extracted in Section 5.5. The dilution template in the background's joint PDF is extracted from the upper mass sideband for each species, as depicted with the B_s signal distribution in Figure 6-9.

Reflections

The structure of flavor modeling for the misreconstructed Λ_b and B^0/B_s reflection components is similar to that of the other backgrounds. These components comprise a very small fraction of the sample and are of little importance to the signal B_s mixing parameters. Regardless, their treatment aims to describe them as accurately as is reasonable.

The baryonic Λ_b component does not undergo flavor mixing and is fitted with a constant asymmetry. Because it is a fully reconstructed decay, it is reasonable to expect that the SST efficiency is the same as that of the signal components. However, the signal SST dilutions are not calibrated for Λ_b decays, so this component uses a fully independent parameter for average SST dilution. In contrast, the use of the uncorrelated OST information mimics that the signals exactly, including the shared signal efficiency, predicted dilutions for each event, and dilution scale factor.

The B^0 and B_s reflections are also fitted with a constant flavor asymmetry, despite the fact that they do oscillate. For the case of the B_s reflection, the small component has no sensitivity to Δm_s and so is best modeled by the integrated asymmetry. The small component for the B^0 reflection may have some sensitivity to mixing because of the low value of Δm_d, but the PDF is modeled with a constant asymmetry for simplicity. This choice is demonstrated to cause no bias to the fitted Δm_s value in the B_s data. In both of these fully reconstructed cases, the SST efficiency is again taken to be that of the signal, while the SST dilution is fitted with an independent parameter for the average. The OST application again shares the signal tagging

efficiency and predicted event-by-event dilutions, but an independent dilution scale factor is applied as a free parameter in order to insulate the signal dilution from effects of oscillations in the neutral B reflection.

In applying OST dilution templates, the same distributions of the signal region are used for both Λ_b and neutral B reflections. The signal histogram of the SST dilution is also used for the Λ_b component, while the B^0 signal histogram is used for the SST dilution templates of both the B_s and B^0 reflections.

6.5.5 Complete Likelihood and Tagger Application

After expansion into flavor space, the complete likelihood remains of the same overall structure as outlined in Equation (5.29). The non-flavor-specific terms $\mathcal{P}^{\rm ct}$ for the lifetime fits are augmented with the new flavor terms of Equation (6.11), taking the information of both taggers into account.

The most physically meaningful parameters of the flavor space, particularly Δm for the neutral B species as well as tagger efficiencies and dilutions, are shared among the signal components. This discussion does not yet introduce the Fourier oscillation amplitude $\mathcal{A}$, which is withheld until the B_s analysis of Chapter 7. However, $\mathcal{A}$ is also shared among the signal parameters.

The multiple sets of independent efficiency and dilution parameters introduced for the various backgrounds are intended to empirically model the flavor asymmetry of the non-signal components. In combination with correct modeling in the mass and simple proper time spaces, these empirical models minimize the possibility of background components biasing the physically meaningful parameters. Because these empirical values are not of specific interest themselves, they are not reported with the fits to data for any B species.

As an overview of the flavor space, Table 6.3 summarizes the application of the taggers to the various model components which are developed throughout this dissertation. The common treatment of the fully and partially reconstructed signal decays is emphasized, with shared parameters of flavor asymmetry and oscillation.

It must be made clear that the complications induced by the admixture of B^+ and B^0 species in the B^+ data is a special case for that sample. The B^0 and B_s samples remain highly parallel in their content, qualitatively differing only in the value of the oscillation frequency and the need to fix the dilution parameters of the B_s likelihood. In the fits of the light B data below, the attention paid to the B^+ results is minimal, as the utility of the B^+ data as a development sample for partially reconstructed decays is limited at this stage of the analysis.

6.6 Asymmetry Fits in B^0 and B^+ Data

The method of previous chapters has been to simultaneously report the results of the fitted data for each of the B species. This approach is reasonable in the mass and simple proper time spaces because the attributes of the likelihood components are qualitatively similar across the samples. However, because of the need to introduce

Component	Osc.?	ϵ,$\mathcal{D}$ Usage					
		OST			SST		
full.reco $D\pi$	$\surd$	pred	sig ϵ	sig $S_{\mathcal{D}}$	pred	sig ϵ	sig $S_{\mathcal{D}}$
part.reco $D\rho$	$\surd$	pred	sig ϵ	sig $S_{\mathcal{D}}$	pred	sig ϵ	sig $S_{\mathcal{D}}$
part.reco $D^*\pi$	$\surd$	pred	sig ϵ	sig $S_{\mathcal{D}}$	pred	sig ϵ	sig $S_{\mathcal{D}}$
part.reco $D^*\pi(B^+)$	$\surd$	pred	sig ϵ	sig $S_{\mathcal{D}}$	avg	own ϵ	own $\mathcal{D}$
part.reco DX	—	avg	own ϵ	own $\mathcal{D}$	avg	own ϵ	own $\mathcal{D}$
refl.Λ_b	—	pred	sig ϵ	sig $S_{\mathcal{D}}$	avg	sig ϵ	own $\mathcal{D}$
refl.$B_{d,s}$	—	pred	sig ϵ	own $S_{\mathcal{D}}$	avg	sig ϵ	own $\mathcal{D}$
combi.bkg	—	avg	own ϵ	own $\mathcal{D}$	avg	own ϵ	own $\mathcal{D}$

Table 6.3: Overview of the use of tagger parameters for all fully ("full.reco") and partially reconstructed ("part.reco") sample components, including whether each component undergoes flavor oscillations and uses the signal efficiencies ϵ and dilution scale factors $S_{\mathcal{D}}$. The "pred" notation refers to the use of predicted event-by-event dilutions from the tagger calibrations, as opposed to a dilution parameter $\mathcal{D}$ whose value the fitter averages ("avg") over the component. As discussed in the text, background components generally use their "own" independent tagger ϵ and $\mathcal{D}$ parameters. The "$D^*\pi(B^+)$" component applies only to the B^+ sample.

the oscillation amplitude for B_s data in Chapter 7, the exactly parallel nature of the analysis ends at these flavor asymmetry fits.

This section applies the likelihood to the B^0 and B^+ data only. As in simpler stages of the analysis, this step provides reassurance that the fitter produces reasonable values for known physical quantities – in this case, Δm_d – and generally demonstrates reasonable behavior of the tagger parameters. Of specific interest is the stable behavior of the tagger parameters in the new partially reconstructed signal components, which cannot be directly tested for B_s data.

The mass range of the asymmetry fits is the same as the narrow range of the lifetime fits. In all fits of flavor asymmetry, the previously fixed parameters from the mass space remain fixed, including the mass templates and the relative signal fractions. Additionally, the fitted parameters of the simple proper time space are extracted and fixed, including the ct distribution of the combinatorial background and the overall background fraction. As noted in the discussion of signal components, the application of the established selection bias curves and κ-factor distributions remains the same. The signal lifetime parameter is allowed to float, although it is only weakly correlated with the flavor parameters.

Fitter Validation

As is done with every new expansion of the likelihood, the fitting framework is validated with large ensembles of toy Monte Carlo samples. The MC distributions are generated with physically realistic values from the B^0 data, with specific attention to the oscillation frequency (at low values such as $\Delta m_d = 0.5$ ps^{-1}; the high frequency regime is not explored yet) and tagger parameters for the signal. Appendix D

illustrates that the likelihood produces unbiased measurements of these physically important parameters.

6.6.1 Results

The results of greatest interest are the oscillation and tagging parameters of the fitted B^0 data and the consistency of these values as separately measured in fully and partially reconstructed signal components. Table 6.4 compiles the signal asymmetry parameters for the full sample and for restricted mass regions corresponding to the two signal types.

Parameter	**[4.95,5.6]** GeV/c^2	**[4.95,5.2]** GeV/c^2	**[5.2,5.6]** GeV/c^2
Δm_d [ps^{-1}]	0.500 ± 0.034	0.448 ± 0.043	0.549 ± 0.048
$S_{\mathcal{D}}^{\mathrm{SSKT}}$	0.258 ± 0.067	0.213 ± 0.102	0.293 ± 0.089
$S_{\mathcal{D}}^{\mathrm{OST}}$	1.092 ± 0.089	1.298 ± 0.138	1.004 ± 0.118
ϵ^{SSKT} [%]	59.2 ± 0.4	62.7 ± 0.9	57.0 ± 0.5
ϵ^{OST} [%]	97.7 ± 0.1	100 ± 0.03	96.8 ± 0.2

Table 6.4: Fitted parameters of flavor asymmetry in 355 pb^{-1} of B^0 data, including the complete mass range (left) and restricted regions corresponding to the partially (center) and fully reconstructed (right) signal decays. The non-unity SSKT scale factor is discussed in the text. Only statistical uncertainties are reported.

The fitted values of Δm_d are consistent with the world average value [2] and combine well as independent measurements from the two data regions. The overall OST dilution scale factor is consistent with unity, which is expected for a well-calibrated OST algorithm. For technical reasons, this data sample is tagged with the same side *kaon* tagger of Section 6.3, rather than a purely kinematic SST as has been used previously at CDF [102]. The non-unity value of the fitted SSKT scale factor is readily explained by the observation that charged fragmentation kaons associated with B^0 are *anti*-correlated with its flavor, as inferred by considering the result of $K^{*0} \to K^+\pi^-$ decay in Figure 6-5. Therefore, this negative contribution to the tagger dilution is accounted for in the fitted $S_{\mathcal{D}}^{\mathrm{SSKT}}$ value. Regardless, the oscillation frequency is clearly resolved. This observation is relevant for the B_s analysis in that it supports the concept that dilution effects are not strongly correlated with Δm effects.

The tagging efficiencies for the separate signal regions are similar but somewhat inconsistent. This result may point to the signal fractions being slightly incorrect in the partially reconstructed region, allowing the extra tracks associated with background events to pull the tagger efficiencies. The systematic uncertainty studies conservatively account for the possibility of confused application of the taggers, as is discussed below.

As a secondary examination of the B^0 data, and as an indication of the kinematic SST dilution, fit results from the corresponding B^0 sample of the older rectangular cut selection are compiled in Table 6.5. This fit measures a single average SST dilution

from all signal candidates. Additionally, due to the fast pace of upgrades for the CDF B_s mixing analysis, the neural network combination of the OST algorithms is not available for this sample. Separate parameters are listed for the five distinct OST methods, which are combined in the hierarchical order of dilution noted in Section 6.2.4. This sample uses fixed parameters of mass and proper time which are extracted in the same manner as the main NN-selected samples.

Parameter	**[4.95,5.6]** GeV	**[4.95,5.2]** GeV	**[5.2,5.6]** GeV
Δm_d [ps^{-1}]	0.555 ± 0.031	0.508 ± 0.049	0.594 ± 0.042
$\mathcal{D}^{\mathrm{SST}}$ [%]	11.9 ± 2.3	10.5 ± 4.4	12.5 ± 2.7
$S_{\mathcal{D}}^{\mathrm{SMT}}$	1.361 ± 0.175	1.573 ± 0.346	1.327 ± 0.217
$S_{\mathcal{D}}^{\mathrm{SET}}$	1.228 ± 0.237	1.091 ± 0.355	1.306 ± 0.305
$S_{\mathcal{D}}^{\mathrm{JVX}}$	1.055 ± 0.329	1.416 ± 0.586	0.714 ± 0.428
$S_{\mathcal{D}}^{\mathrm{JJP}}$	0.825 ± 0.248	0.789 ± 0.412	0.913 ± 0.317
$S_{\mathcal{D}}^{\mathrm{JPT}}$	0.819 ± 0.480	1.692 ± 0.820	0.612 ± 0.602
ϵ^{SST} [%]	62.3 ± 0.5	63.7 ± 1.2	62.0 ± 0.6
ϵ^{SMT} [%]	4.4 ± 0.2	4.6 ± 0.5	4.4 ± 0.3
ϵ^{SET} [%]	2.8 ± 0.2	3.2 ± 0.4	2.7 ± 0.2
ϵ^{JVX} [%]	7.1 ± 0.3	7.3 ± 0.5	7.1 ± 0.3
ϵ^{JJP} [%]	28.1 ± 0.5	28.4 ± 1.1	28.2 ± 0.6
ϵ^{JPT} [%]	52.1 ± 0.6	51.6 ± 1.2	52.1 ± 0.6

Table 6.5: Fitted parameters of flavor asymmetry in 355 pb^{-1} of B^0 data selected with rectangular cuts; these results are supplementary to the main B^0 oscillation fit of Table 6.4. The sample uses a kinematic SST algorithm without per-event dilution and a hierarchical combination of OST methods without the NN. Only statistical uncertainties are reported.

The oscillation frequency is again compatible with the world average, although it is different from the main B^0 fit of Table 6.4 because the data samples only partially overlap and because the taggers are notably different. The behavior of the taggers across the independent signal regions is consistent.

The B^+ data also has available only the hierarchical OST combination. The sample is fitted without same side tagging, to reduce the probability for the extra charged tracks to interfere with tag decisions. Table 6.6 compiles the fitted signal asymmetry for the complete sample and for restricted mass regions corresponding to the fully and partially reconstructed signal types. The performance of opposite side tagging is consistent between the independent regions.

Summary of Tagging in Data

The fitted asymmetry of the light B species indicates that the tagger performance in fully and partially reconstructed signal channels is largely consistent. However, the comparison raises enough questions that further studies are warranted. This analysis affirms the simple application of identical tagger information to all signal types, as

Parameter		[**4.95,5.6**] GeV	[**4.95,5.2**] GeV	[**5.2,5.6**] GeV
$S_{\mathcal{D}}^{\mathrm{SMT}}$		1.179 ± 0.079	1.259 ± 0.125	1.121 ± 0.108
$S_{\mathcal{D}}^{\mathrm{SET}}$		1.027 ± 0.123	1.162 ± 0.227	0.971 ± 0.154
$S_{\mathcal{D}}^{\mathrm{JVX}}$		1.132 ± 0.155	1.007 ± 0.237	1.205 ± 0.201
$S_{\mathcal{D}}^{\mathrm{JJP}}$		1.041 ± 0.127	1.054 ± 0.184	1.051 ± 0.174
$S_{\mathcal{D}}^{\mathrm{JPT}}$		1.421 ± 0.228	0.999 ± 0.334	1.622 ± 0.310
ϵ^{SMT}	[%]	4.3 ± 0.2	4.3 ± 0.2	4.3 ± 0.2
ϵ^{SET}	[%]	2.4 ± 0.1	2.2 ± 0.2	2.5 ± 0.1
ϵ^{JVX}	[%]	6.6 ± 0.2	6.5 ± 0.3	6.6 ± 0.2
ϵ^{JJP}	[%]	28.0 ± 0.4	28.0 ± 0.5	28.0 ± 0.2
ϵ^{JPT}	[%]	52.6 ± 0.4	52.8 ± 0.6	52.3 ± 0.3

Table 6.6: Fitted parameters of flavor asymmetry in 355 pb^{-1} of B^+ data, using the hierarchical OST combination. Only statistical uncertainties are reported.

argued in Section 6.4, but conservative systematic uncertainties are evaluated to provide confidence in the final results from B_s data. The systematic studies described in Section 7.2.3 and Appendix A account for the possibility of severe confusion between different classes of signal and between signal and background events, such as might produce more extreme variations in efficiency and dilution than are observed in the data. These uncertainties include a modest bias to the oscillation amplitude, but the effect on Δm_s or the significance of the B_s mixing signature is negligible.

Chapter 7

Observation of B_s^0-$\overline{B}_s^0$ Oscillations

In this chapter, the fully validated analysis framework is applied to the flavor space of B_s data. The sample of fully and partially reconstructed hadronic decays is studied with amplitude scans of the frequency domain, which is then combined with the results of related CDF mixing analyses. The significance of the overall oscillation signature is evaluated, followed by a precise measurement of Δm_s.

7.1 Amplitude Scans

Previous analyses have lacked the statistical significance to directly measure the B_s oscillation frequency. Section 1.4.1 introduces the amplitude parameter, which probes the frequency space in search of the resonance corresponding to Δm_s. Amplitude scans combine the utility of a low-statistics resonance search with a simple method of combining independent analysis results. This analysis uses amplitude scans to identify an oscillation signal.

7.1.1 Amplitude Formalism

The oscillation amplitude $\mathcal{A}$ represents the final addition to the likelihood. This parameter enters the time-flavor PDF as a multiplicative factor on the cosinusoidal oscillation of signal components. The signal expression of Equation (6.18) is expanded to read

$$\mathcal{P}^{\text{osc}}(t|\xi_o,\xi_s,\mathcal{D}_o,\mathcal{D}_s) \propto \frac{(1+\xi_o\mathcal{D}_o\xi_s\mathcal{D}_s)+\mathcal{A}\cdot(\xi_o\mathcal{D}_o+\xi_s\mathcal{D}_s)\cos(\Delta mt)}{(1+|\xi_o|)\,(1+|\xi_s|)}, \qquad (7.1)$$

where a common amplitude is shared across all fully and partially reconstructed signal components. The free parameter $\mathcal{A}$ is fitted at regular intervals of fixed Δm and the resulting amplitude spectrum plots the response of the likelihood for various oscillation frequencies. Effectively, the model is "tuned" in frequency space to reveal the Δm_s value preferred by the data. As introduced in Section 1.4.1, the likelihood should favor the presence of the oscillation term at the true mixing frequency and

therefore produce $\mathcal{A} = 1$, while the cosine term should be *suppressed* with $\mathcal{A} = 0$ away from Δm_s.

Equation (7.1) highlights the importance of using fixed dilution values for the taggers: the amplitude cannot be determined if $\mathcal{D}$ is not known. Furthermore, properly calibrated dilutions are required for the likelihood maximization to produce the expected amplitude peak at $\mathcal{A} = 1$. Since the oscillation term is smeared by proper time uncertainty, as explicitly written in Equation (6.14), the peak height also depends on the correct calibration of ct resolution.

The statistical significance of the data sample is manifested in the precision with which the amplitude is measured. Section 1.4.1 discusses how the dependences of the amplitude significance

$$\frac{1}{\sigma_{\mathcal{A}}} = \sqrt{\frac{\mathcal{S}}{\mathcal{S}+\mathcal{B}}} \cdot \exp(-\frac{\Delta m^2 \sigma_t^2}{2}) \cdot \sqrt{\mathcal{S}\frac{\epsilon \mathcal{D}^2}{2}} \tag{7.2}$$

dictate the value of the hadronic decay channels in this analysis. Much of the historical utility of the amplitude scan method lies in its ability to set lower limits, based on the frequency value above which Δm_s cannot be excluded. More importantly, the significance of $\mathcal{A}/\sigma_{\mathcal{A}}$ is closely related to the overall significance of the potential B_s mixing observation, as detailed in Section 7.2.

Fitter Validation

As in previous chapters, the expansion of the likelihood requires validation of the fitting framework. Unbiased measurement of the amplitude is verified with large ensembles of toy Monte Carlo samples which use physically realistic values from the B_s data and which assume various representative values of Δm_s.

Systematic Uncertainties on Amplitude

In analyses whose result is an amplitude-based lower limit, the systematic uncertainties on amplitude are fundamentally important. As may be inferred from Equation (7.1), uncertainties on the fixed dilution values are prominent error sources for the amplitude. The detector resolution model and the possibility of correlated tagging algorithms are also potential bias sources. However, if a true mixing signature is present, these amplitude uncertainties are unrelated to to the Δm_s measurement. The amplitude scans of this section incorporate several sources of systematic uncertainty, which are described for completeness in Appendix A. The subset of relevant effects for Δm_s are detailed in Section 7.2.3 below.

7.1.2 Blind Analysis

In order to maintain an unbiased search for an amplitude resonance, the analysis is developed in a "blinded" manner. The decisions $\{\xi\}$ of the flavor taggers are randomized, such that the predictive power of the data may be studied without explicit revelation of where a potential signal may lie. The decisions are deterministically

scrambled via the transformation $\xi \to (-1)^n \xi$, where n is an event number assigned by the CDF DAQ system. If decisions from both taggers are available, both values undergo the same transformation, thus retaining their combined contribution. Because the randomization leaves intact the various attributes on which $\sigma_{\mathcal{A}}$ depends, including tag efficiency and dilution, proper time resolution, and signal fractions, the precision of the measured amplitudes in blinded data are a valuable indicator of the expected Δm_s sensitivity.

Extraction of Fixed Parameters

Although the tagger dilution for rapidly oscillating signal components cannot be assumed to be directly measurable, the empirical parameters for efficiency and dilution of the constant background asymmetry may be extracted. Additionally, by fixing the amplitude to $\mathcal{A} = 0$, the signal tagging *efficiency* may be measured without interference from the oscillation terms. In performing any amplitude scan, whether blinded or not, the only parameter which remains free as various frequency values are probed is $\mathcal{A}$ itself. The flavor parameters of the background components as well as the signal efficiencies are fitted once with amplitude fixed to zero, after which these values are extracted and fixed in the amplitude scan. Additionally, all parameters of the mass space and non-flavor-specific proper time are fixed to the values measured in previous chapters. For completeness, the signal efficiencies and empirical background parameters are listed with the unblinded scans of Section 7.1.3.

Blinded Scan

Figure 7-1 shows the amplitude scan of fully and partially reconstructed hadronic B_s decays in 1 fb^{-1} of blinded data. Using the randomized tag decisions, the fitted amplitude is consistent with the $\mathcal{A} = 0$ hypothesis of no oscillations. It should be noted that consecutive amplitude points are highly correlated, since they fit the same data with otherwise-identical parameters, so the point-to-point variations do not correspond directly to the size of the uncertainties. The 1-σ statistical uncertainties are plotted as error bars, while the colored bands illustrate the 95% uncertainties with and without systematic uncertainties. Uncertainty sources are combined in quadrature as $\sigma_{\mathcal{A}}^{\text{total}} = \sqrt{(\sigma_{\mathcal{A}}^{\text{stat}})^2 + (\sigma_{\mathcal{A}}^{\text{syst}})^2}$, with the total uncertainty plotted in gray. As indicated by Equation (7.2), the statistical uncertainty grows rapidly toward large Δm_s. The constant vertical offset at $\Delta m_s = 0$ is dictated by the number of signal events and by the tagger performance, while the proper time resolution determines the rate of growth in Δm_s.

The intersection of the combined 95% uncertainty with the $\mathcal{A} = 1$ level provides the definition of *sensitivity* – the frequency value below which the data is expected to exclude Δm_s with 95% confidence if no resonance is present, or the value below which an existing resonance should begin to appear. Of course, the lower limit derived from the blinded data is not meaningful. The limit is based on the incompatibility of the unblinded data points with $\mathcal{A} = 1$, a relationship which relies on the fitted amplitude

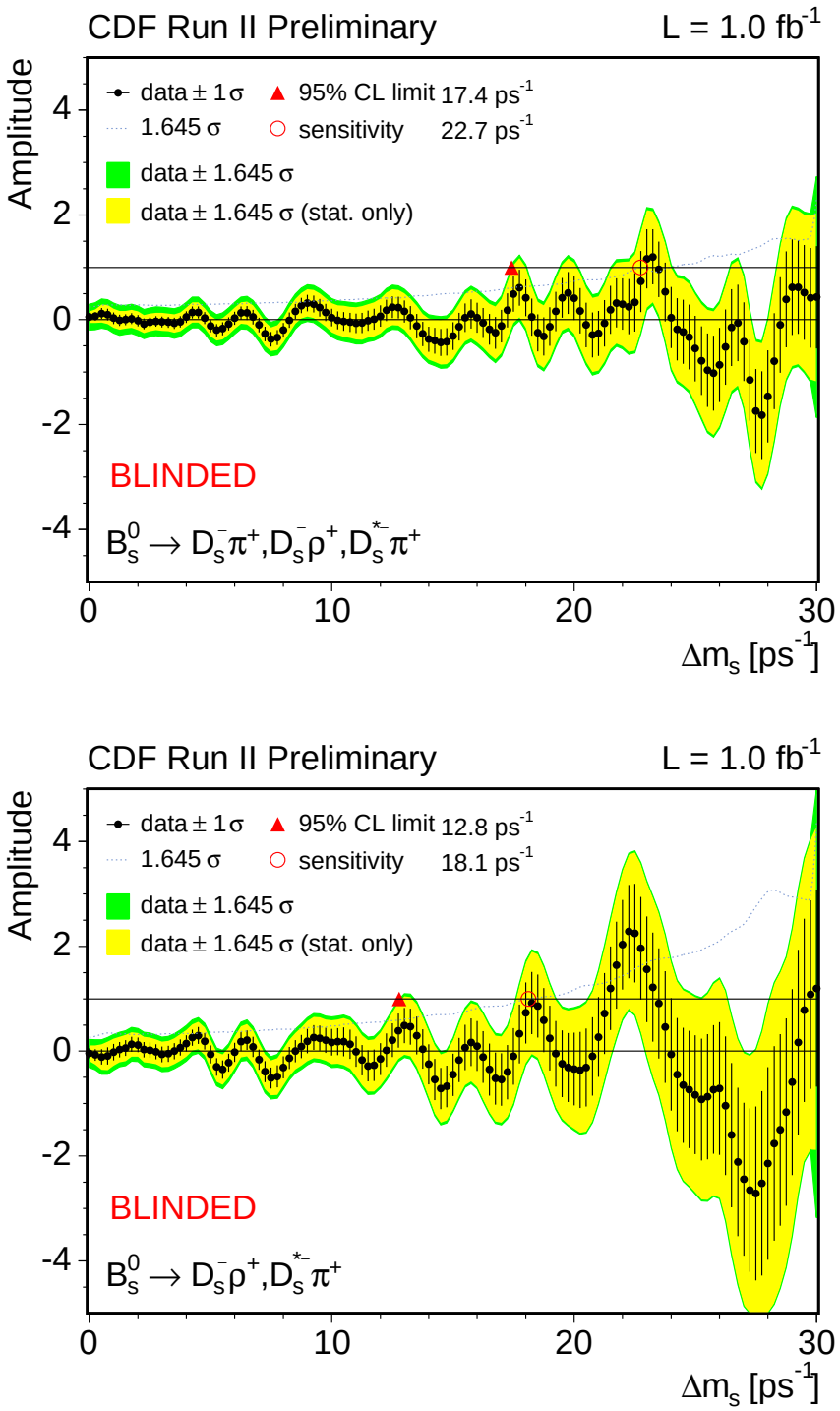

Figure 7-1: Blinded amplitude scans in 1 fb^{-1} of B_s data. Top: the combined scan of fully and partially reconstructed hadronic decays. Bottom: the contribution of partially reconstructed decays, using only the mass range [5.0, 5.3] GeV/c^2.

values as well as their errors. Assuming symmetric fluctuations of the amplitude points, the sensitivity defines the expected median position of the unblinded limit.

However, the goal of this analysis is not to provide a limit but rather to contribute to the first definitive observation of oscillations. The recent CDF evidence for B_s mixing and measurement of Δm_s [31] was based on a combined sample of 3.1 standard deviation significance with 95% CL sensitivity to 25.8 ps^{-1}. That analysis combined semileptonic decays with five fully reconstructed hadronic modes, including the $B_s^0 \to D_s^- \pi^+$ channel of this sample, using less powerful event selection and flavor tagging. As depicted in Figure 7-1, the expected sensitivity of this analysis alone is 22.7 ps^{-1}, including the fully reconstructed $B_s^0 \to D_s^- \pi^+$ channel as well as the new partially reconstructed hadronic modes. Figure 7-1 also shows the amplitude scan contribution of these $B_s^0 \to D_s^- \rho^+$ and $B_s^0 \to D_s^{*-} \pi^+$ decays alone, affirming the expected behavior of blinded data and indicating a sensitivity of 18.1 ps^{-1}. The addition of the other fully reconstructed modes and semileptonic data is expected to raise the combined sensitivity to an unprecedented level. The final significance of a new signal, if any is observed, must be evaluated after removal of the randomization of tag decisions, the unblinded data.

7.1.3 Unblinded Scans

On unblinding the B_s tagging information, the parameters of background flavor asymmetry and signal efficiency are again extracted from the fit with amplitude fixed to $\mathcal{A} = 0$. The amplitude scans are performed separately for the three internal divisions of data-taking for which distinct calibrations of various analysis tools are done. Table 7.1 compiles representative flavor parameters from the first 355 pb^{-1} of B_s data. The overall tagging efficiencies for the signal decays are $\epsilon^{\mathrm{SSKT}} = (51.5 \pm 0.9)\%$ for the Same Side Kaon Tagger and $\epsilon^{\mathrm{OST}} = (97.6 \pm 0.2)\%$ for the Opposite Side Tagger.

Efficiency [%]	Dilution [%]
$\epsilon_{\mathrm{sig}}^{\mathrm{SSKT}} = 50.9 \pm 1.4$	
$\epsilon_{\mathrm{sig}}^{\mathrm{OST}} = 97.5 \pm 0.4$	
$\epsilon_{\mathrm{com}}^{\mathrm{SSKT}} = 84.6 \pm 2.4$	$\mathcal{D}_{\mathrm{com}}^{\mathrm{SSKT}} = 0.6 \pm 6.6$
$\epsilon_{\mathrm{com}}^{\mathrm{OST}} = 99.1 \pm 0.5$	$\mathcal{D}_{\mathrm{com}}^{\mathrm{OST}} = 6.8 \pm 6.2$
$\epsilon_{DX}^{\mathrm{SSKT}} = 69.3 \pm 10.2$	$\mathcal{D}_{DX}^{\mathrm{SSKT}} = 13.8 \pm 20.8$
$\epsilon_{DX}^{\mathrm{OST}} = 100.0 \pm 3.4$	$\mathcal{D}_{DX}^{\mathrm{OST}} = 10.9 \pm 19.0$

Table 7.1: Representative flavor parameters of signal and background fitted with $\mathcal{A} \equiv 0$. The parameters are fixed in the subsequent amplitude scan. These values correspond to the first 355 pb^{-1} of B_s data. Signal efficiencies are fitted directly in the data, while the dilutions are calibrated in Chapter 6. Empirical parameters are fitted for the constant asymmetry of combinatorial background ("com") and partially reconstructed physics background ("DX").

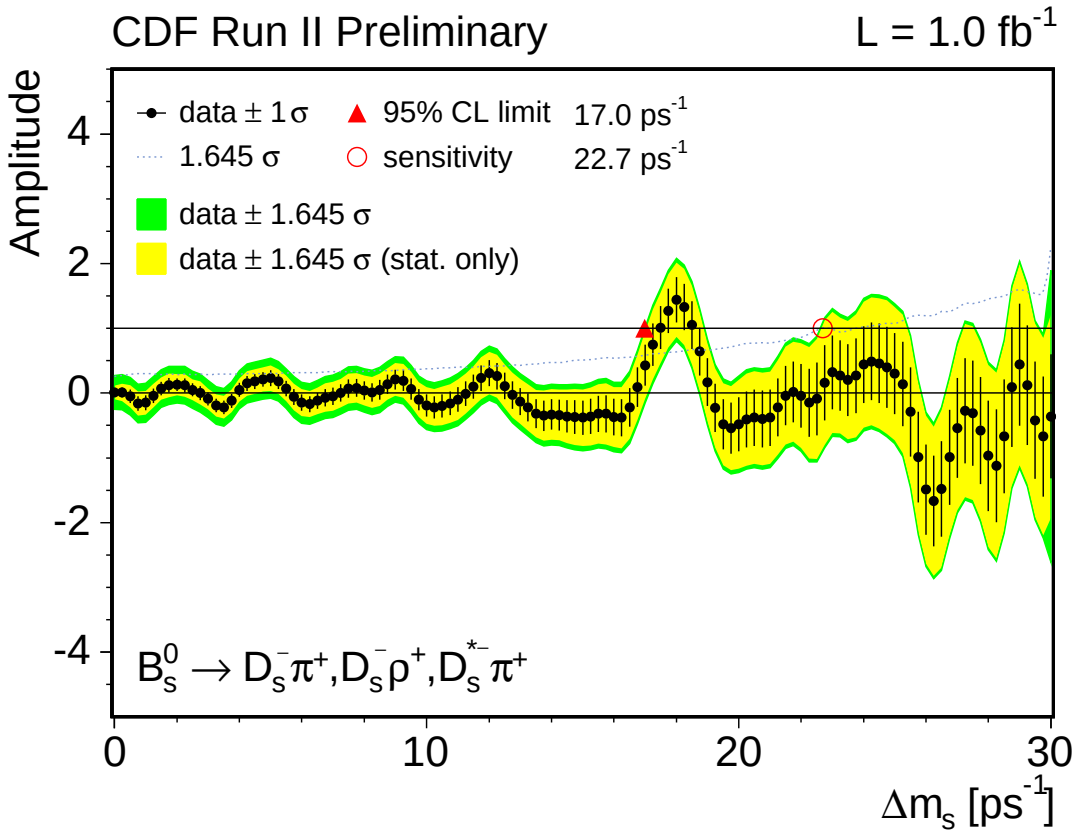

Figure 7-2: Unblinded amplitude scan in 1 fb^{-1} of the fully and partially reconstructed hadronic B_s decay channels of this analysis. The resonance has a maximum amplitude significance of $\mathcal{A}/\sigma_{\mathcal{A}} = 4.1$. The colored bands indicate 95% uncertainties, while the smaller error bars provide the 1σ precision.

Figure 7-2 shows the unblinded amplitude scan for the combined fully and partially reconstructed signal components of this analysis. The spectrum reveals an apparent resonance in the vicinity of 18 ps^{-1} with a maximum statistical significance of $\mathcal{A}/\sigma_{\mathcal{A}} = 4.1$. The final measurements of Δm_s are independent of the amplitude scan binning and are discussed in Section 7.3.2. The precision of the amplitude measurements are dominated by statistical uncertainty in the region of interest. The small systematic component shown here and derived in Appendix A is relevant only for amplitude, while separate uncertainties are evaluated for Δm_s.

The sensitivity of the sample is 22.7 ps^{-1}, as predicted by the blinded data, while the nominal lower limit on Δm_s is 17.0 ps^{-1}, where the amplitude rises sharply to 1. The maximum amplitude value of 1.44 ± 0.39 (stat.+ syst.) is consistent with unity, indicating that the tagger dilutions and proper time resolutions are correctly calibrated. The behavior of the unblinded scan is exactly as expected, producing amplitudes consistent with zero away from the resonance and consistent with unity

at the peak.

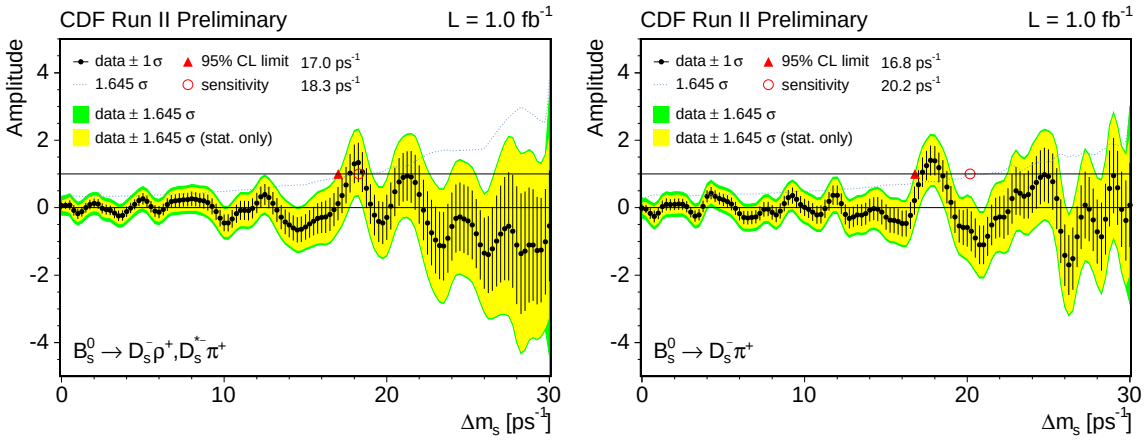

Figure 7-3: Unblinded amplitude scans of separate B_s signal classes in 1 fb^{-1}. Left: partially reconstructed $B_s^0 \to D_s^- \rho^+$ and $B_s^0 \to D_s^{*-} \pi^+$ decays. Right: fully reconstructed decays of $B_s^0 \to D_s^- \pi^+$ ($D_s^- \to \phi \pi^-$).

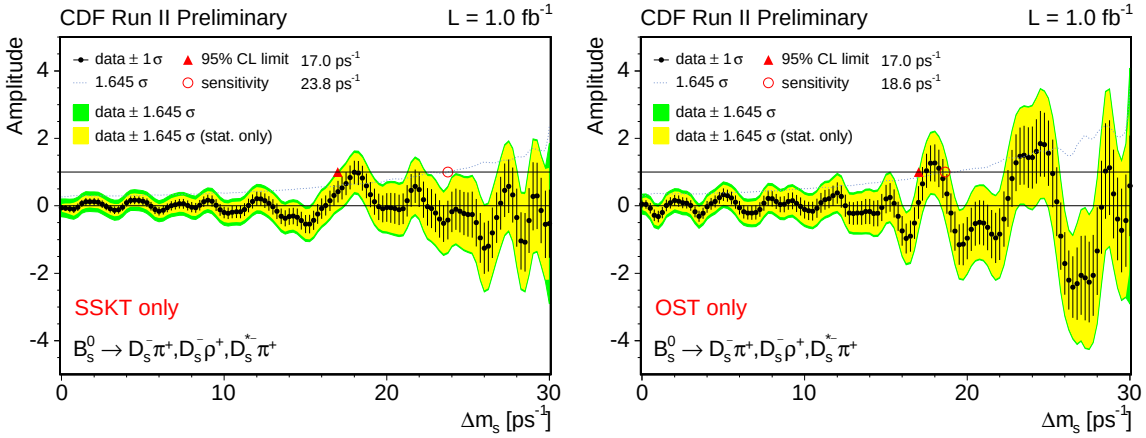

Figure 7-4: Unblinded amplitude scans of the combined fully and partially reconstructed B_s signal classes using separate tagging algorithms. Left: SSKT only. Right: OST only.

Figure 7-3 shows the separate contributions of the fully and partially reconstructed decays as divided by a mass cut at 5.3 GeV/c^2. Both scans diverge from zero and are consistent with unity at the resonance position. The partially reconstructed hadronic channels have statistical precision of $\sigma_{\mathcal{A}} = 0.567$ at the 17.75 ps^{-1} frequency point, which is close to the value of Δm_s measured below, and maximum amplitude significance of $\mathcal{A}/\sigma_{\mathcal{A}} = 2.3$. The fully reconstructed channel has $\sigma_{\mathcal{A}} = 0.443$ at 17.75 ps^{-1}

and maximum $\mathcal{A}/\sigma_{\mathcal{A}} = 3.2$. These uncertainties may be naively compared to estimate that the contribution of the ~ 3150 partially reconstructed hadronic B_s decays incorporated by this analysis is equivalent to a $\left(\sigma_{\mathcal{A}}^{\rm full}/\sigma_{\mathcal{A}}^{\rm part}\right)^2 \sim 61\%$ increase in yield over the ~ 2000 fully reconstructed decays. Comparison of $\sigma_{\mathcal{A}}^{\rm part}$ with the total sample's statistical uncertainty $\sigma_{\mathcal{A}} = 0.343$ at 17.75 ps^{-1} indicates an equivalent yield increase of $\sim 67\%$. The smaller statistical power per event for the partially reconstructed decay channels is explained by the minor degradation of proper time resolution due to the κ-factor smearing of Section 5.3 and the increased background levels of the low-mass region.

Figure 7-4 shows the amplitude scans for the complete sample using separate application of the flavor tagging algorithms. Both scans shows signs of a resonance at the appropriate position. The statistical precision of the amplitudes at 17.75 ps^{-1} for the SSKT scan and the OST scan are $\sigma_{\mathcal{A}} = 0.352$ and $\sigma_{\mathcal{A}} = 0.533$, respectively, indicating the relative importance of same-side tagging in the analysis.

7.1.4 Related Analyses

This dissertation represents a central component of a suite of related B_s oscillation analyses at the CDF experiment. In addition to the $B_s^0 \to D_s^- \pi^+$ and associated partially reconstructed decays of this analysis, five other fully reconstructed hadronic channels of the form $B_s \to D_s(3)\pi$ are analyzed [83,89]. These channels are comprised of the three D_s decay modes $D_s^- \to \phi\pi^-$, $D_s^- \to K^{*0}K^-$, and $D_s^- \to \pi^-\pi^-\pi^+$, in combination with the B_s topologies $B_s^0 \to D_s^-\pi^+$ and $B_s^0 \to D_s^-\pi^+\pi^+\pi^-$, where the $B_s^0 \to D_s^-\pi^+$ ($D_s^- \to \phi\pi^-$) mode is the fully reconstructed channel of this analysis. As in previous incarnations of the $B_s^0 \to D_s^-\pi^+$ ($D_s^- \to \phi\pi^-$) analysis, the other hadronic samples neglect partially reconstructed decays by excluding all candidates below 5.3 GeV/c^2 in mass. The combined yield of these other decays is 3500 signal events which maintain the advantages of good proper time resolution inherent in fully reconstructed decays while having somewhat lower signal-to-background ratios than those shown in Chapter 4. Identical tagging algorithms are used for all hadronic decays. Another CDF analysis uses large samples of semileptonic $B_s \to D_s^{(*)}\ell X$ decays [35], where ℓ stands for an electron or a muon and the D_s decays via the same three modes as above. Although the semileptonic channels benefit from higher branching fractions, producing a total yield of 61 500 signal events, the momentum distributions of these partially reconstructed decays are much less well-defined than in the hadronic $B_s^0 \to D_s^-\rho^+$ and $B_s^0 \to D_s^{*-}\pi^+$ channels, causing a loss of significance for oscillation measurements. These decays use very similar taggers which have different final dilution calibrations.

Figure 7-5 shows the unblinded amplitude scans from these related analyses. The hadronic sample exhibits all the expected traits of an amplitude scan with a resonance. The statistical uncertainty $\sigma_{\mathcal{A}} = 0.29$ and $\mathcal{A}/\sigma_{\mathcal{A}} = 4.4$ at 17.75 ps^{-1} indicates that the combination of the other five hadronic channels has somewhat more significance than the total fully and partially reconstructed decays of this dissertation's analysis sample alone. The scan of the semileptonic sample provides an example of

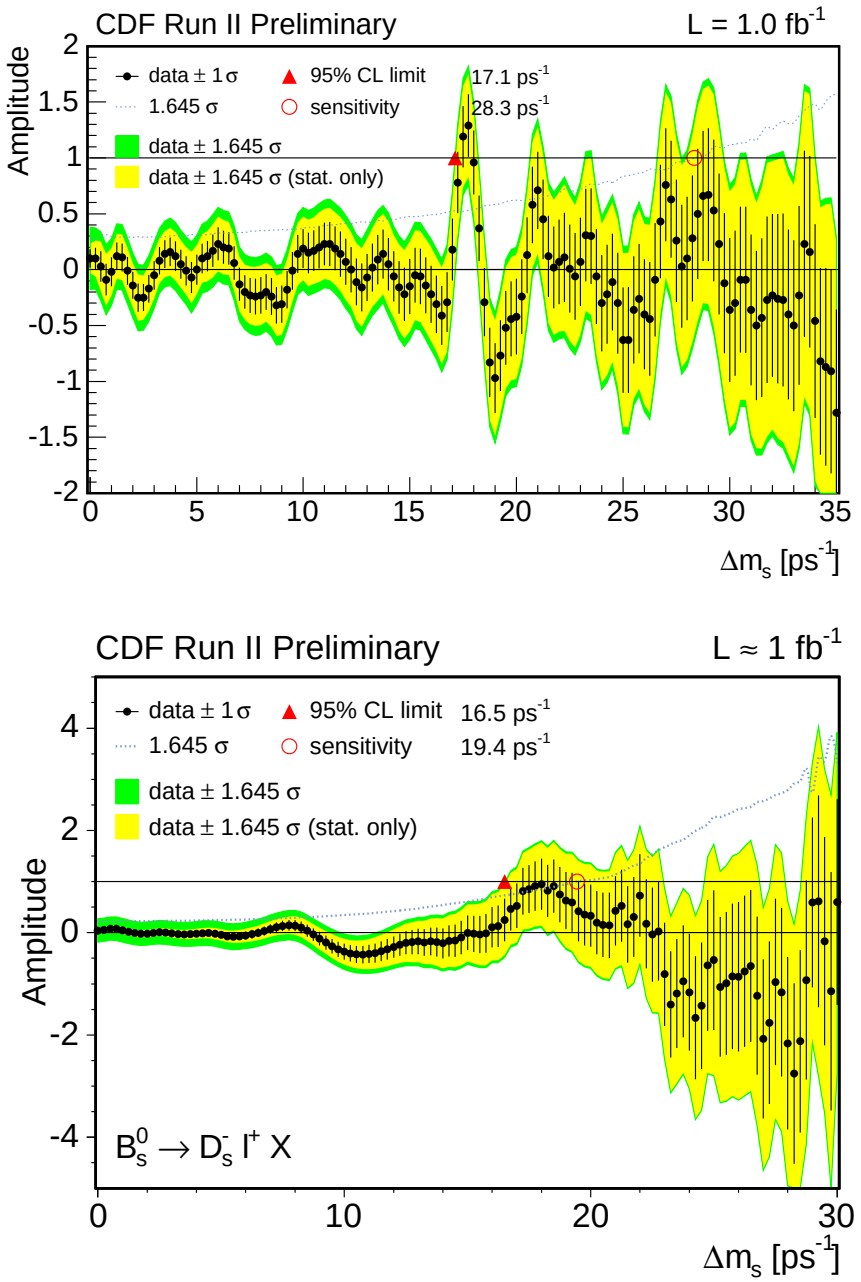

Figure 7-5: Amplitude scans of related CDF B_s oscillation analyses in 1 fb^{-1}. Top: Five fully reconstructed hadronic $B_s \to D_s(3)\pi$ channels [83, 89]. Bottom: partially reconstructed semileptonic $B_s \to D_s^{(*)}\ell X$ channels [35].

amplitude behavior at lower statistical significance. The fitted amplitudes rise to good consistency with unity at the apparent resonance position, but the $\mathcal{A}/\sigma_{\mathcal{A}} \sim 2$ displacement from zero is not conclusive. Although such an indication of B_s mixing would have been exciting as recently as late-2005, the semileptonic significance is

largely overshadowed by the hadronic samples. However, the measured sensitivity of 19.4 ps^{-1} represents the most powerful semileptonic sample for B_s mixing analysis in the world [1] . The semileptonic amplitudes are very precise at low frequency values, illustrating the strong Δm_s-dependence of Equation (7.2). An additional notable feature of the semileptonic amplitude scan is its relative smoothness in comparison with the hadronic scans. This behavior is closely related to the width of the associated κ-factor distributions. As depicted in Figure 5-8, the large uncertainty in momentum-smearing for semileptonic decays results in worse proper time resolution than in the partially reconstructed hadronic channels. This smearing, in addition to degrading the amplitude precision at high Δm_s, also causes a greater point-to-point correlation between the fitted amplitudes and a general smoothness of variation across the scan. The addition of semileptonic data to the final combination provides a smoother scan than the more abrupt variations of the hadronic channels alone, particularly at low frequencies where the semileptonic sample is very sensitive.

7.1.5 Combined Amplitude Scan

Part of the traditional utility of the amplitude method lies in the simplicity of combination between independent scans. With the assumption of symmetric Gaussian uncertainties on each amplitude point, the fitted amplitudes maybe combined in a weighted average at each point in the frequency space.

The slightly more general approach used here is to combine the likelihoods of the independent analyses and examine the variation of the total likelihood with amplitude:

$$\ln \mathcal{L}_{\text{total}}(\mathcal{A}, \Delta m_s) = \sum_{samples} \ln \mathcal{L}_s(\mathcal{A}, \Delta m_s). \tag{7.3}$$

Using the fixed parameters that characterize all but the oscillation terms of the analyses, the resulting likelihood may be represented as a two-dimensional function of the parameters $\mathcal{A}$ and Δm_s, which is illustrated in Section 7.2.1 below. At each fixed point in frequency space, the combined $-2 \ln \mathcal{L}_{\text{total}}$ is minimized with respect to $\mathcal{A}$ to produce the overall amplitude and its uncertainty. This process is, of course, identical to the procedure of the fitter in the case of a single analysis sample. The difference for separate scans is that it smoothly incorporates the possibility of asymmetric amplitude uncertainties, which would otherwise be assumed symmetric in the simple combination method. Systematic errors are added in quadrature as before, although given clear evidence for a resonance, they are primarily shown for display purposes and do not contribute to the final significance of Section 7.2. This same method of likelihood combination is used for construction of likelihood ratios and measurement of Δm_s in later sections.

Figure 7-6 shows the amplitude scan of the combined CDF B_s oscillation analyses in 1 fb^{-1}. The two plots depict the same scan, with the lower scan emphasizing the statistical precision. The combined resonance is prominent, with a maximum

[1] During final preparation of this dissertation, the DØ experiment released an analysis of 2.4 fb^{-1} of semileptonic B_s decays with a sensitivity of 27.4 ps^{-1} (statistical uncertainties only).

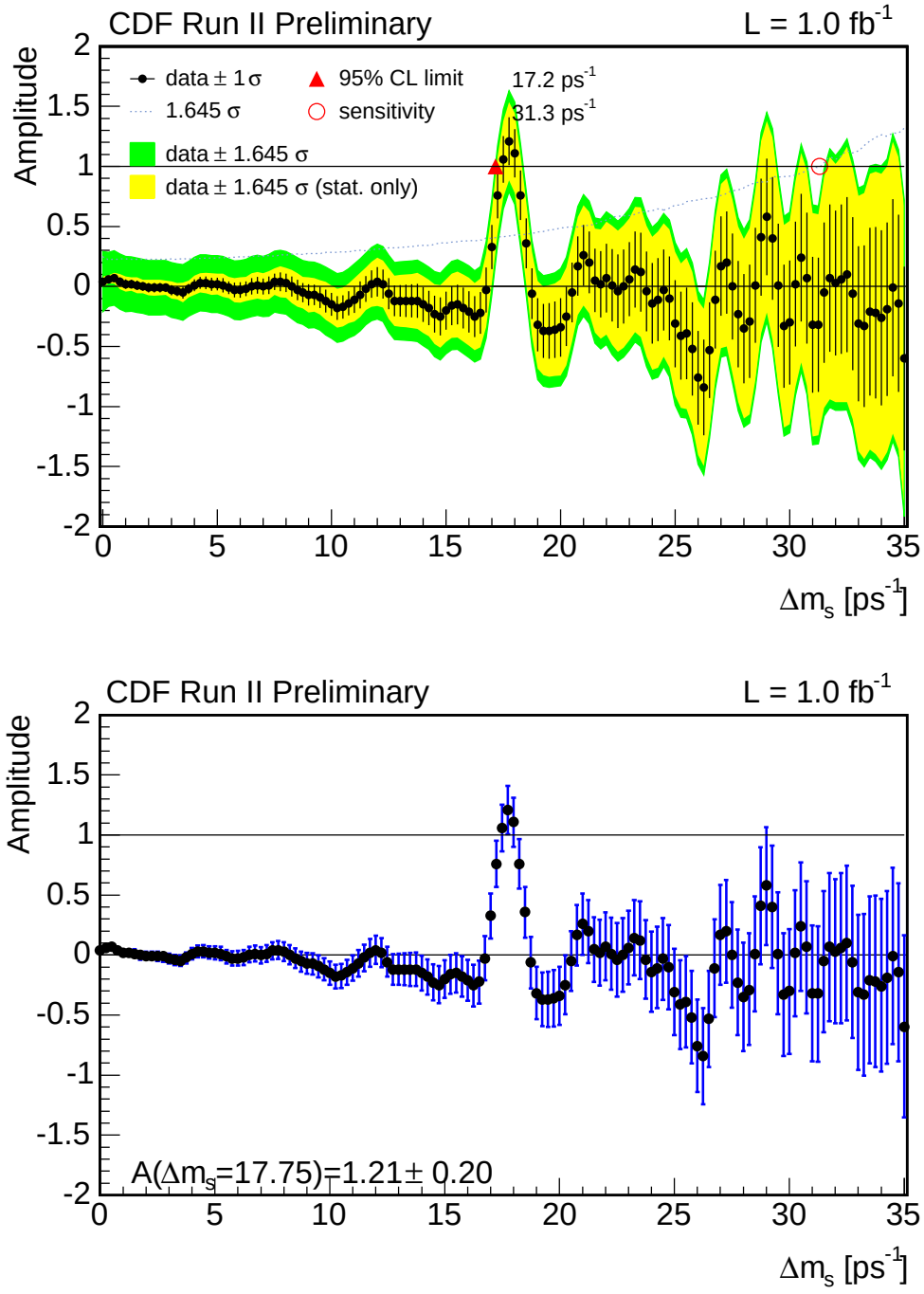

Figure 7-6: Combined amplitude scan of CDF B_s oscillation analyses in 1 fb^{-1}, including full systematic errors (top) and the statistical uncertainties alone (bottom), the latter of which are the relevant errors for the total significance. The data includes the fully and partially reconstructed hadronic decays of this analysis, five other fully reconstructed hadronic channels, and the large semileptonic samples.

amplitude of $\mathcal{A} = 1.21 \pm 0.20(\text{stat.})$ at 17.75 ps^{-1}. It should be noted that the overall statistical significance of the oscillation signal is not simply the $\mathcal{A}/\sigma_{\mathcal{A}} \sim 6$ value of the peak point, but rather must account for the possibility of the true resonance appearing at *any* Δm_s value, whether within the window of the scan or otherwise. The p-value estimate of Section 7.2 provides a comprehensive indication of the signal significance.

The features of the combined scan are well-behaved. The amplitudes are consistent with zero away from the resonance and consistent with unity at the peak, thus indicating that the overall calibration of dilution and proper time resolution are performed correctly. The combined scan leverages the complementary power of the hadronic and semileptonic samples. It incorporates the smoothness and precision at lower frequency values from the high-statistics semileptonic data with the sensitivity at high Δm_s of the lower-statistics hadronic data. The total sensitivity of the CDF samples is 31.3 ps^{-1}, the highest in the world [2]. The lower limit of $\Delta m_s > 17.2$ ps^{-1} at the 95% CL remains meaningful, but the apparent significance of the oscillation signal justifies a direct measurement of the frequency. The following sections confirm the significance of the combined observation and discuss the precise extraction of Δm_s.

7.2 Combined Observation and Signal Significance

To claim an observation of B_s mixing, the signature for flavor oscillations must first be quantitatively defined. The significance of the analysis is then determined by the p-value, the probability for a random fluctuation to produce the observed signature. This analysis uses a likelihood ratio as the comprehensive signature of flavor oscillations.

7.2.1 Likelihood Ratio

Estimation of the p-value is performed with use of the ratio

$$\Lambda = -\log \frac{\mathcal{L}^{\mathcal{A}=1}(\Delta m_s)}{\mathcal{L}^{\mathcal{A}=0}}, \tag{7.4}$$

where $\mathcal{L}^{\mathcal{A}=1}$ is the likelihood of the data under the hypothesis that Δm_s is the true mixing frequency. The constant $\mathcal{L}^{\mathcal{A}=0}$ is independent of Δm_s and represents the likelihood for $\mathcal{A} = 0$, which is equivalent to oscillations with $\Delta m_s \to \infty$. The log-ratio Λ defines a "likelihood profile" that maps out the oscillation response of the data in frequency space. The definition of Equation (7.4) implies that the unit-resonance of the amplitude scans corresponds to a minimum in the likelihood profile. The minimum value of Λ is related to the significance of the signal, while the position of the minimum in frequency space determines the mean value and uncertainty of Δm_s.

Figure 7-7 compares the likelihood profile of the combined CDF data with the expected values of an equivalent sample with and without actual B_s oscillations. The figure also depicts the likelihood in the two-dimensional space of amplitude and

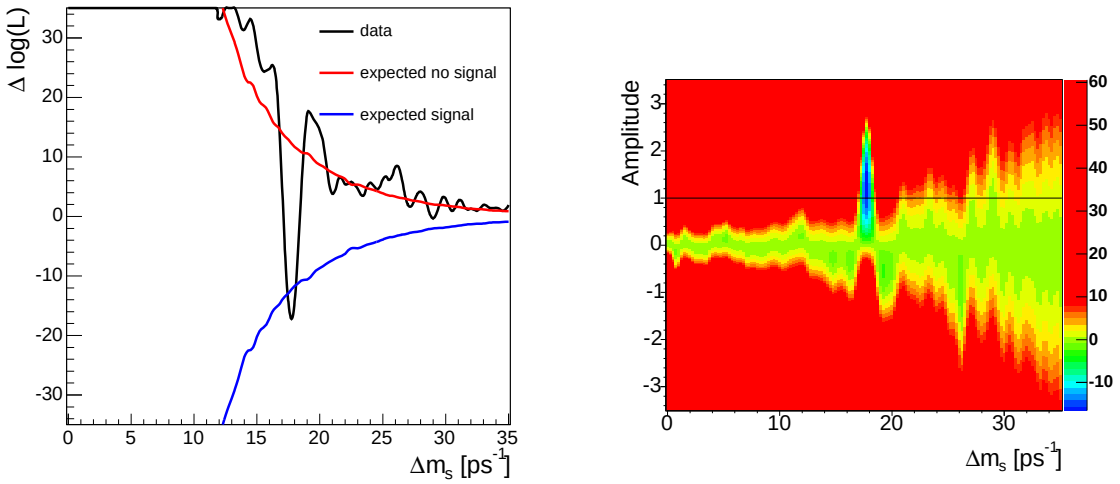

Figure 7-7: Left: Likelihood profile of the combined CDF sample, compared to the values expected from equivalent data with and without B_s oscillations. The blue line indicates the expected likelihood depth were the oscillation signal to be located at that frequency. Right: Two-dimensional variation of the likelihood with respect to $\mathcal{A}$ and Δm_s, which is the common source of both the likelihood profile and the amplitude scan.

frequency, whose values at $\mathcal{A} = 1$ and $\mathcal{A} = 0$ determine the profile in Δm_s. The expected profiles are derived from the amplitude precision of the unblinded data using Equation (7.5) below. The deep minimum of $\Lambda = -17.26$ at $\Delta m_s = 17.77$ ps^{-1} in the real data is close to the value expected for mixing at that Δm_s frequency, while the profile is close to the non-mixing prediction away from the minimum. Details of the final Δm_s measurement are discussed in Section 7.3.2. The minimum Λ value is tied to the signal significance because it is directly related to the amplitude features. Neglecting statistical fluctuations, the simple relation [27]

$$\Lambda(\Delta m_s) = \frac{\frac{1}{2} - \mathcal{A}}{\sigma_{\mathcal{A}}^2}, \tag{7.5}$$

predicts the likelihood depth based on the measured amplitude at frequency Δm_s. However, since it is constructed from data, the likelihood ratio contains similar fluctuations to the amplitude scan. One may initially question why values of $\mathcal{A}/\sigma_{\mathcal{A}}$ for individual amplitude points are not a sufficient measure of significance. The useful difference between Λ and $\mathcal{A}/\sigma_{\mathcal{A}}$ as significance indicators lies in their behavior at high Δm_s, away from the oscillation frequency: the fluctuations in an amplitude scan cause $\mathcal{A}/\sigma_{\mathcal{A}}$ to be distributed as a unit-Gaussian about zero, including some large values. The use of amplitude scans for p-value estimation requires the inherent assumption of a fixed Δm_s search window. In contrast, the statistical fluctuations of the likelihood profile at high Δm_s are increasingly small about zero and never approach the

depth of the signal minimum. The data profile of Figure 7-7 illustrates this behavior. Therefore, the minimum depth Λ represents a comprehensive indicator of significance for an infinite frequency range which is not available from an amplitude scan alone.

Systematic uncertainties on amplitude are not important for the value of Λ. Sources of bias for amplitude act to shift the fitted value away from the ideal signal $\mathcal{A} = 1$. As may be visualized in Figure 7-7, the resulting effect for the likelihood ratio would be that the depth of the signal minimum is *diminished*. Therefore, while the apparent significance of the analysis is punished for incorrect or non-optimal data modeling, systematic biases cannot "fake" a more significant signal in the likelihood profile.

The form of the expected Λ values in Figure 7-7 provide an alternative representation of the data's power. Given the statistical sensitivity of the CDF data, the expected difference between the likelihood minima of the mixing/no mixing cases is large at low Δm_s: slow oscillations are easy for the data to resolve. The expected values of Λ converge at very high Δm_s, where the sensitivity of the data becomes inadequate. As a final note regarding Figure 7-7, the depth of the observed minimum is slightly greater than expected from the statistical power of the dataset. Based on Equation (7.5), this behavior is explained by the fact that the peak amplitude turns out to be $\mathcal{A} = 1.21 \pm 0.20$, on the high side of unity.

The following discussions of p-values and measurement of Δm_s include the contributions to the overall likelihood profile from separate analysis samples, including the fully and partially reconstructed hadronic channels on which this dissertation is based.

7.2.2 *p*-value Estimation

The attributes of the likelihood profile makes its minimum value Λ an ideal signature of oscillation significance. The p-value is defined as the probability for a dataset of equivalent statistical power with no mixing signal present to observe a likelihood minimum at least as deep as Λ at any Δm_s value.

The most straightforward method of reproducing the statistical power of the data while simulating the null hypothesis is to simply randomize the tags in the data itself. Much like the blinded amplitude scans, in which the tagger decisions are deterministically randomized, the p-value trials involve each tag being flipped with 50% probability according to the output of a random number generator. The randomized data is equivalent to that which would be measured for B_s mesons which do not undergo flavor oscillations at all. The resulting likelihood profile is inspected in the range from 0 to 35 ps^{-1} and the minimum value Λ is recorded. Examination of likelihood profiles up to very high ranges in Δm_s support that the probability of a global minimum outside this window is completely negligible, in contrast to the simple $\mathcal{A}/\sigma_{\mathcal{A}}$ case noted above.

The randomization is performed in 3.5×10^8 trials, resulting in 28 cases with $\Lambda < -17.26$. Figure 7-8 shows the distribution of Λ for the randomized data, illustrating the disparity between the actually observed value and the distribution of values for data without a mixing signal. The probability of an equivalent dataset containing no

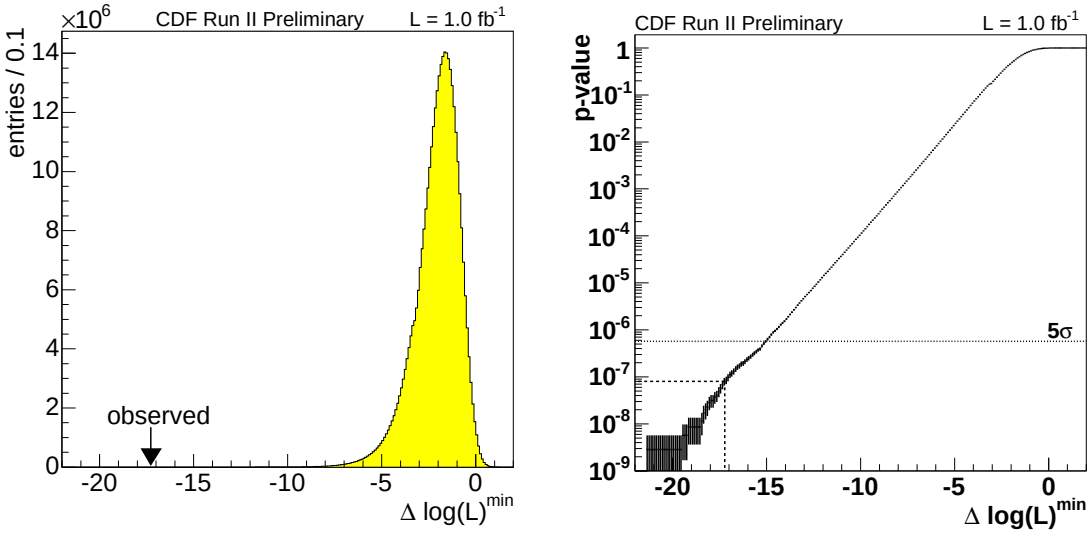

Figure 7-8: Distribution of Λ for fluctuations of randomized data (left), and the dependence of the extracted p-value on the observed signal Λ (right).

mixing signal to produce Λ at least as significant as the one observed in CDF data is 8.0×10^{-8}, with negligible statistical uncertainty from the finite number of trials. This p-value corresponds to a Gaussian significance of 5.4 standard deviations. Figure 7-8 also shows that this p-value is well beyond the 5 standard deviations threshold of 5.7×10^{-7}. The combined CDF dataset therefore represents the world's first definitive observation of B_s^0-$\overline{B}_s^0$ oscillations.

7.2.3 Contributions from Separate B_s Channels

This observation is the product of a suite of interrelated CDF analyses whose samples are combined and which includes the results of this dissertation.

Figure 7-9 compares the likelihood profiles of the combined hadronic samples and semileptonic sample. This hadronic profile represents the fully and partially reconstructed channels of this analysis as well as the five other fully reconstructed channels. The dominance of the hadronic contribution to the overall signal significance is obvious, while both profiles converge to the same non-signal behavior at high frequencies.

Figure 7-10 compares the contributions of signal channels from this dissertation, overlaying the profiles of the fully reconstructed $B_s^0 \to D_s^- \pi^+$ and the combined partially reconstructed hadronic decay channels. As noted in the examination of the amplitude scans, the fully reconstructed signal is indeed more significant than the partially reconstructed components. Both decay types have well-behaved likelihood minima with consistent measurements of Δm_s, as discussed in Section 7.3.2. The likelihood ratio has $\Lambda = -7.4$. Using the randomized trials of only this sample, this depth corresponds to a p-value of 0.15%. The corresponding significance of 3.2 standard deviations is greater than that of the entire CDF analysis of early 2006 [31], which was published with 3.1 standard deviations evidence.

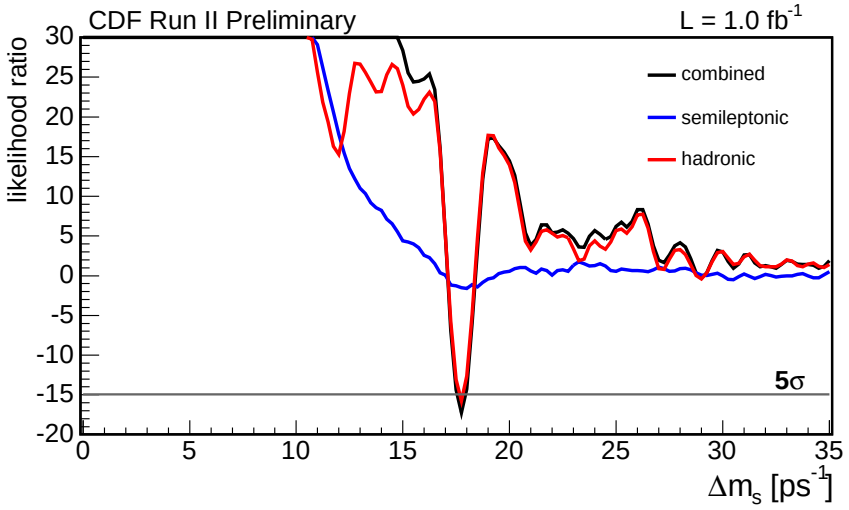

Figure 7-9: Separate likelihood profiles for the combined hadronic and semileptonic decay samples.

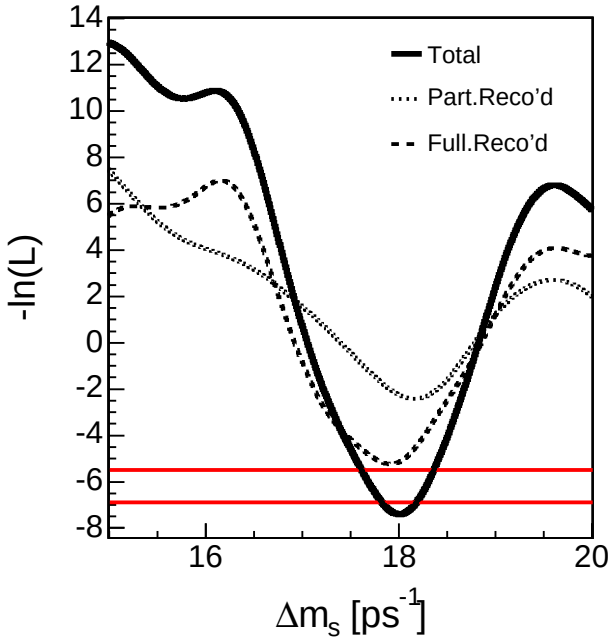

Figure 7-10: Separate likelihood profiles of the fully reconstructed $B_s^0 \to D_s^- \pi^+$ and partially reconstructed hadronic decay channels of this analysis, as extracted from the independent mass ranges about 5.3 GeV/c^2. The red lines indicate the 1σ and 95% CL bounds for the Δm_s measurement.

The fully reconstructed $B_s^0 \to D_s^- \pi^+$, $D_s^- \to \phi\pi^-$ channel produces $\Lambda = -5.2$ and an approximate p-value of 1.5%. The likelihood depth of the combined $B_s^0 \to D_s^- \rho^+$ and $B_s^0 \to D_s^{*-} \pi^+$ signals is $\Lambda = -2.4$, with an approximate p-value of $\sim 28\%$. This significance alone is unimpressive. However, it should be noted that the overall sample does not quite meet the 5 standard deviations threshold without the contribution of these partially reconstructed hadronic channels.

7.3 Systematic Uncertainties on Δm_s

Given confirmation of a genuine oscillation signal, a measurement of Δm_s may be readily extracted from the likelihood profiles. However, whereas systematic uncertainties on amplitude become unimportant in the likelihood ratio, potential sources of bias in the frequency space must still be addressed. The analysis has relatively few systematic uncertainty sources for Δm_s, all of which relate to the measurement of proper decay time.

The total Δm_s uncertainty is driven by statistical errors, as extracted in Section 7.3.2. This section describes the evaluation of non-negligible systematic sources. The possibility of other biases related to modeling of partially reconstructed hadronic signals is excluded in Appendix B. Three uncertainty sources are applied uniformly to all hadronic and semileptonic B_s decay modes, while the partially reconstructed hadronic decays incur one additional source.

Evaluation Procedure

Potential systematic biases to the measurement of Δm_s may be inferred from other data measurements in proper time or they may be evaluated using large ensembles of toy Monte Carlo experiments. Two of the major effects are estimated using other lifetime measurements in CDF data, as discussed below, while others are tested with Monte Carlo simulation.

The toy MC samples are constructed to reflect the statistical power of the real data. Events are generated using the precise parameter values of the nominal likelihood model and following the observed distributions of proper time resolution and tagger dilution. Each sample contains the same number of total events as the data, with signal and background candidates generated according to the measured fractions.

This method examines the effects on the fitted Δm_s value from the nominal likelihood model due to changes in the toy simulated "data" samples. For each uncertainty source, a distinct ensemble of "biased" toy experiments is generated to simulate a potentially mismodeled aspect of the data. For example, the toy simulation may introduce incorrect signal fractions that are not in the nominal model. Each element of the biased sample set is generated with a sequence of random numbers which corresponds exactly to that of an element of the "default" or nominal set. Fitted Δm_s values are extracted from the biased samples using the nominal likelihood model, thus mimicking the scenario of an unknown bias in data. The mean shift in the measured Δm_s between the biased and unbiased samples provides the systematic uncertainty estimate. Using large ensemble sizes of at least several hundred experiment pairs,

this method of evaluation allows for the systematic effects to be decoupled from the statistical effects of the toy Monte Carlo samples.

7.3.1 Universal Uncertainties

Using the above toy Monte Carlo procedure for each source, the complete set of systematic effects evaluated for amplitude scans in Appendix A is tested for an associated bias in Δm_s. Excluding one source related to the partially reconstructed hadronic decays which is noted below, these uncertainties are found to be negligible, as expected. Since the B_s oscillation frequency is large and has a period which is a small fraction of the B_s lifetime [2], the measurement of Δm_s is decoupled from other issues of proper time modeling. This assertion is supported by studies discussed in Appendix B. The three primary sources of systematic uncertainty on Δm_s are related to the absolute proper time scale of the CDF detector. These effects may be summarized as follows:

- **Silicon detector alignment:** A common source of systematic uncertainty in B lifetime measurements with the CDF detector stems from imperfect knowledge of the positions of silicon wafers [73]. Since lifetime measurements are effectively averages of the decay times of signal events, the associated biases from distortion of the silicon detector may also be used as estimates of the relative bias to proper time for the Δm_s measurement. Shifts in B lifetimes are measured from simulated distortions on the order of $\sim$50 μm applied to radial displacement, bowing, and rotation of silicon sensors. The distortion tolerances are based on a physical survey of the detector. The maximum lifetime bias is found to be 1.0 μm, which corresponds to a 0.2% uncertainty on the proper time scale. A systematic error contribution of 0.04 ps^{-1} is therefore assigned to the measured Δm_s value.

- **Track-fit bias:** Mismeasurements of track curvature may be propagated into biased calculations of proper time, either via the transverse decay length $L_{\rm xy}$ of track vertices or via the measured B momentum. The positive or negative direction of the bias depends on whether the affected tracks curve toward or away from each other. Like the above distortions of the silicon detector, these mismeasurements introduce an overall systematic shift in measured B lifetimes which may be used as indicators of the relative proper time bias. Based on simulated variations to COT and silicon detector performance from surveys of the CDF tracking system, the bias is found to be 1.3 μm, corresponding to 0.3% uncertainty in the proper time scale or 0.05 ps^{-1} for this measurement.

- **Primary vertex bias.** Systematic mismeasurements of the primary vertex position lead to biases in $L_{\rm xy}$ and therefore in proper decay time. The main origin of this bias is events in which displaced tracks from the opposite side b-hadron are incorporated into the primary vertex. This effect may be exacerbated in events with opposite side flavor tags, which are enriched in such displaced tracks. The bias is studied in data samples of fully reconstructed B^+ and B^0 events by comparing the primary vertex position to that of the average beam

position. The maximum bias is found to be 1.0 μm in the reference frame of the detector, which is simulated in toy Monte Carlo samples to determine a systematic shift of 0.02 ps^{-1} in the measured Δm_s value.

7.3.2 Uncertainties in Partially Reconstructed Decays

The development of modeling for the partially reconstructed channels points to several areas that would demand more study for other analyses, with notable examples being measurements of branching fractions or lifetimes. However, this analysis is oriented toward flavor oscillations, and the overriding consideration for all model decisions is how the Δm_s observation and measurement might be affected. Potential areas of concern include uncertainties on

- the relative fractions of partially reconstructed signal components,
- the relative fractions of signal and background,
- small details in the shape of partially reconstructed mass templates,
- the precision of proper time templates for combinatorial background.

Regardless of the details, the dominant effect from any of these modeling decisions is variation in the number of partially reconstructed signal events modeled by the likelihood fit. Such issues are comprehensively addressed by evaluation of conservative systematic uncertainties in which the signal weighting is grossly modified. The ability of the data to resolve Δm_s is robust to such variations, particularly with regard to background modeling. In this sense, the rapid frequency of oscillations which make them so difficult to observe also serves to insulate their signature from slower time variations once they are resolved. The only non-negligible Δm_s uncertainty arises from the relative signal fractions: the partially reconstructed signals are associated with different κ-factor distributions, which are convolved with the Δm_s parameter and therefore directly affect its fitted value. This source is discussed here, while the exploration of negligible uncertainties is documented for completeness in Appendix B. Other issues of partially reconstructed modeling are more important for amplitude measurement. These uncertainties are summarized in Appendix A.

- **Hadronic κ-factors and signal fractions:** The dominant partially reconstructed $B_s^0 \to D_s^- \rho^+$ and $B_s^0 \to D_s^{*-} \pi^+$ channels have similar kinematics and qualitatively similar models. For the purposes of Δm_s measurement, the only significant difference in the modeling of their proper time space is in their κ-factor distributions, which have slightly different widths and mean values. These distributions are compared in Figure 5-7. The $F(\kappa)$ width is primarily important for a channel's sensitivity to oscillations and therefore its likelihood depth, while the κ-factor mean is directly proportional to the fitted Δm_s value. Uncertainty about the signal fractions assigned to partially reconstructed hadronic components is equivalent to uncertainty on the likelihood weights applied for each κ-factor distribution. Whatever the deviation of the nominal model from the true signal fractions, it is certain to be no worse than 100% confusion of

the signal components. The effect of this confusion is simulated by generating toy MC samples which are entirely composed of only one of the partially reconstructed components and then fitting it as though it were the other. Treatment of the $B_s^0 \to D_s^- \rho^+$ channel as $B_s^0 \to D_s^{*-} \pi^+$ is simulated, as is the reverse scenario. In practice, there is no basis for belief that the nominal model might be so incorrect. Therefore, 50% of the maximum induced Δm_s bias is taken as a very conservative systematic uncertainty, producing a final uncertainty contribution of 0.03 ps^{-1}. This value is assigned only for the Δm_s measurement from the partially reconstructed hadronic mass region and does not contribute significantly to the systematic uncertainty for the overall measurement.

Table 7.2 summarizes the estimated systematic uncertainties for the Δm_s measurement, which combine to 0.07 ps^{-1}.

Source	Uncertainty [ps^{-1}]
Amplitude scan effects	< 0.01
Silicon detector alignment	0.04
Track fit bias	0.05
Primary vertex bias	0.02
Hadronic κ-factors	0.03
Total	0.07

Table 7.2: Systematic uncertainties for the measurement of Δm_s. The uncertainty related to partially reconstructed hadronic κ-factors is applied only to that Δm_s contribution, as discussed further in Section 7.3.2.

7.4 Measurement of Δm_s

Measurement of Δm_s is extracted from the likelihood profile. Applying a Gaussian interpretation, the 1 standard deviation statistical uncertainty may be inferred from the Δm_s range over which the log-likelihood ratio deviates from the minimum Λ by 1/2. One-sided confidence intervals of 95% (90%) correspond to a deviation of 1.90 (1.64).

The likelihood fitter has been validated in toy Monte Carlo ensembles to produce statistically unbiased measurements of Δm_s for various frequency values across the sensitivity range of the combined data. Additionally, the measured confidence level bounds are verified to contain the true toy MC Δm_s value in the appropriate percentage of experiments (95%, 90%, *etc*). Finally, in order to verify that the uncertainty extracted from the likelihood is probable for datasets of equivalent significance, the 1 standard deviation uncertainties are compared with the distributions of uncertainties measured in large ensembles of equivalent toy MC experiments, as shown below.

Combination with Systematic Uncertainties

If every source of systematic uncertainty described in Section 7.2.3 were uniformly applicable to all data samples, the overall systematic error could simply be added in quadrature to the combined measurement. However, the partially reconstructed hadronic sample has an additional systematic source which should not be applied to the combined data. The solution is to convolve the likelihood profiles of the separate datasets with Gaussian distributions of the appropriate systematic uncertainty width for each. These "smeared" likelihood profiles are then combined and used to extract the overall uncertainty. Specifically, the additional uncertainty is applied only to the likelihood from the mass region of the partially reconstructed decays. As is noted in the measurements below, the systematic uncertainties are smaller than the statistical uncertainty, so their addition to the likelihood profiles causes only a small increase in the widths of the minima. Once the smeared profiles are combined, the overall systematic uncertainty may be inferred by comparing the systematically smeared combined measurement with the statistical uncertainty for the unsmeared combination. Subtraction in quadrature of the uncertainties indicates the overall systematic effect.

In practice, this treatment is only necessary for procedural rigor. Because the sole "separate" systematic source is a modest additional contribution for the partially reconstructed hadronic channels, and because the overall depth of the likelihood profile is dominated by fully reconstructed decay channels, the additional error source does not change the overall systematic uncertainty within the significant digits quoted here.

7.4.1 Combined Measurement

Using the combined likelihood profile depicted in Figure 7-9 and convolving the systematic uncertainty into its various contributions, the final measurement of the B_s oscillation frequency is

$$\Delta m_s = 17.77 \ \pm 0.10 \ (\text{stat}) \ \pm 0.07 \ (\text{syst}) \, \text{ps}^{-1}.$$

The likelihood also determines the confidence intervals

$$17.56 < \Delta m_s < 17.96 \ \text{ps}^{-1} \text{ at } 90\% \text{ CL}$$

and

$$16.51 < \Delta m_s < 18.00 \ \text{ps}^{-1} \text{ at } 95\% \text{ CL},$$

which include both statistical and systematic uncertainties.

These statistical uncertainties are compared with the distribution of errors fitted in equivalent toy Monte Carlo samples. Figure 7-11 shows that the measured data values are consistent with the expected range. The probabilities of measuring positive and negative errors at least as large as those of this data are 18% and 37%, respectively.

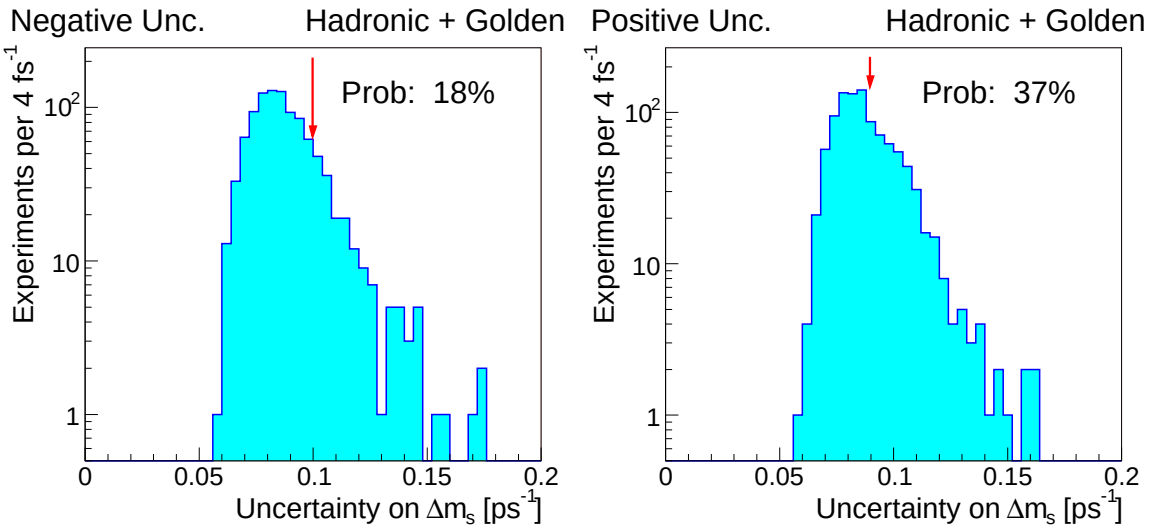

Figure 7-11: Distributions of negative (left) and positive (right) Δm_s statistical uncertainties measured in toy Monte Carlo samples of equivalent statistical power as the combined CDF data. The values actually measured in data are indicated.

7.4.2 Comparison of Hadronic Measurements

The B_s oscillation frequency can be measured in the primary dataset of this dissertation alone. Using the likelihood profile depicted in Figure 7-10, the measured frequency of the complete fully and partially reconstructed B_s signal is

$$\Delta m_s = 18.01 \ \pm 0.18 \ (\text{stat}) \ \pm 0.07 \ (\text{syst}) \, \text{ps}^{-1}.$$

This statistical uncertainty is shown to be well within the distribution of expected errors in Figure 7-12, repeating the exercise with equivalent toy MC samples. The probabilities of measuring positive and negative errors at least as large as those of this data are 37% and 43%, respectively.

Although less statistically significant, the separate measurements of the fully and partially reconstructed hadronic decay classes may be compared. The values are, respectively,

$$\begin{aligned} \Delta m_s &= 17.91 \ \pm 0.24 \ (\text{stat}) \ \pm 0.07 \ (\text{syst}) \, \text{ps}^{-1} \qquad & B_s^0 \to D_s^- \pi^+, \\ \Delta m_s &= 18.14 \ \pm 0.27 \ (\text{stat}) \ \pm 0.08 \ (\text{syst}) \, \text{ps}^{-1} \qquad & B_s^0 \to D_s^- \rho^+, B_s^0 \to D_s^{*-} \pi^+, \end{aligned}$$

which are completely consistent within their independent statistical uncertainties.

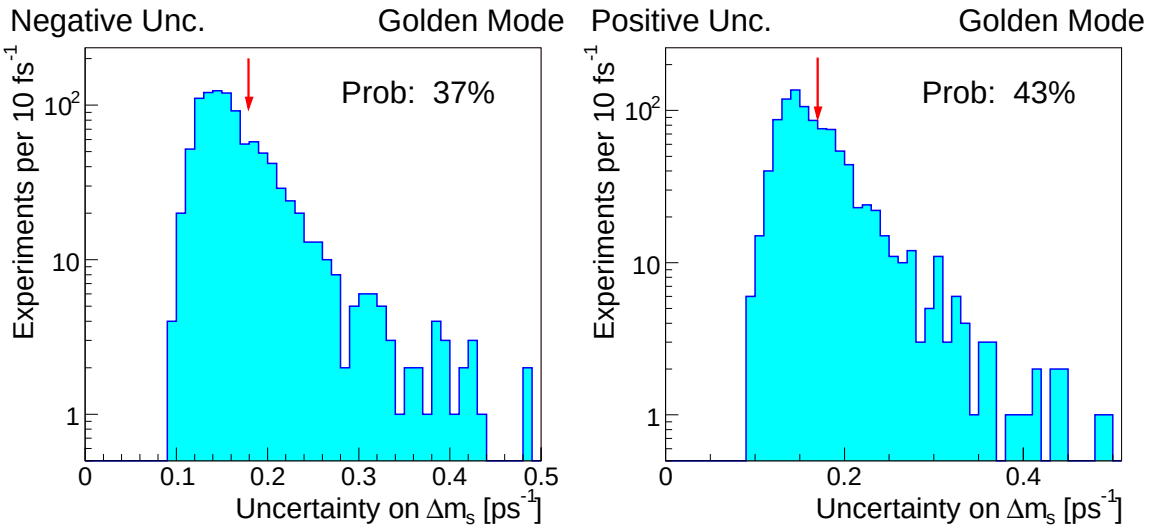

Figure 7-12: Distributions of negative (left) and positive (right) Δm_s statistical uncertainties measured in toy Monte Carlo samples of equivalent statistical power as the dataset of this dissertation. The values actually measured in data are indicated.

7.5 Impact on the Standard Model

The precise measurement of the B_s^0-$\overline{B}_s^0$ oscillation frequency is used to directly constrain the apex of the unitarity triangle. Section 1.2.2 introduces that the ratio of neutral B frequencies is proportional to the least known CKM matrix elements by

$$\frac{\Delta m_s}{\Delta m_d} = \xi^2 \frac{m_{B_s}}{m_{B^0}} \frac{|V_{ts}|^2}{|V_{td}|^2} \qquad \text{with} \quad \xi = 1.210\ ^{+0.047}_{-0.035}\ [22]. \tag{7.6}$$

Combining the CDF measurement of $\Delta m_s = 17.77\ \pm 0.10\ (\text{stat})\ \pm 0.07\ (\text{syst})\,\text{ps}^{-1}$ with the world average $\Delta m_d = 0.507\ \pm 0.005\,\text{ps}^{-1}$ [2] and the CDF measurement of $m_{B^0}/m_{B_s} = 0.98390$ [23], which has negligible uncertainty, the extracted ratio is

$$|V_{td}/V_{ts}| = 0.2060\ \pm 0.0007(\text{exp})\ ^{+0.0081}_{-0.0060}(\text{theory}).$$

The uncertainty denoted "exp" refers to the experimental uncertainty associated with the Δm_s measurement, while the "theory" uncertainty is dominated by ξ of lattice QCD but also includes a contribution from Δm_d. This convention emphasizes that the contribution of the Δm_s uncertainty is an order of magnitude smaller than the combined other sources for $|V_{td}/V_{ts}|$. With enough improvement of the dominant QCD theory uncertainty, the $|V_{td}/V_{ts}|$ constraint would be limited by the 1% uncertainty of the Δm_d value, while this Δm_s measurement has 0.7% precision.

Figure 7-13 shows the effect of the new constraint on the unitarity triangle, comparing the late-2005 figure with the result of this observation of B_s oscillations. The improvement in the $\Delta m_s/\Delta m_d$ constraint is visually striking, directly "squeezing"

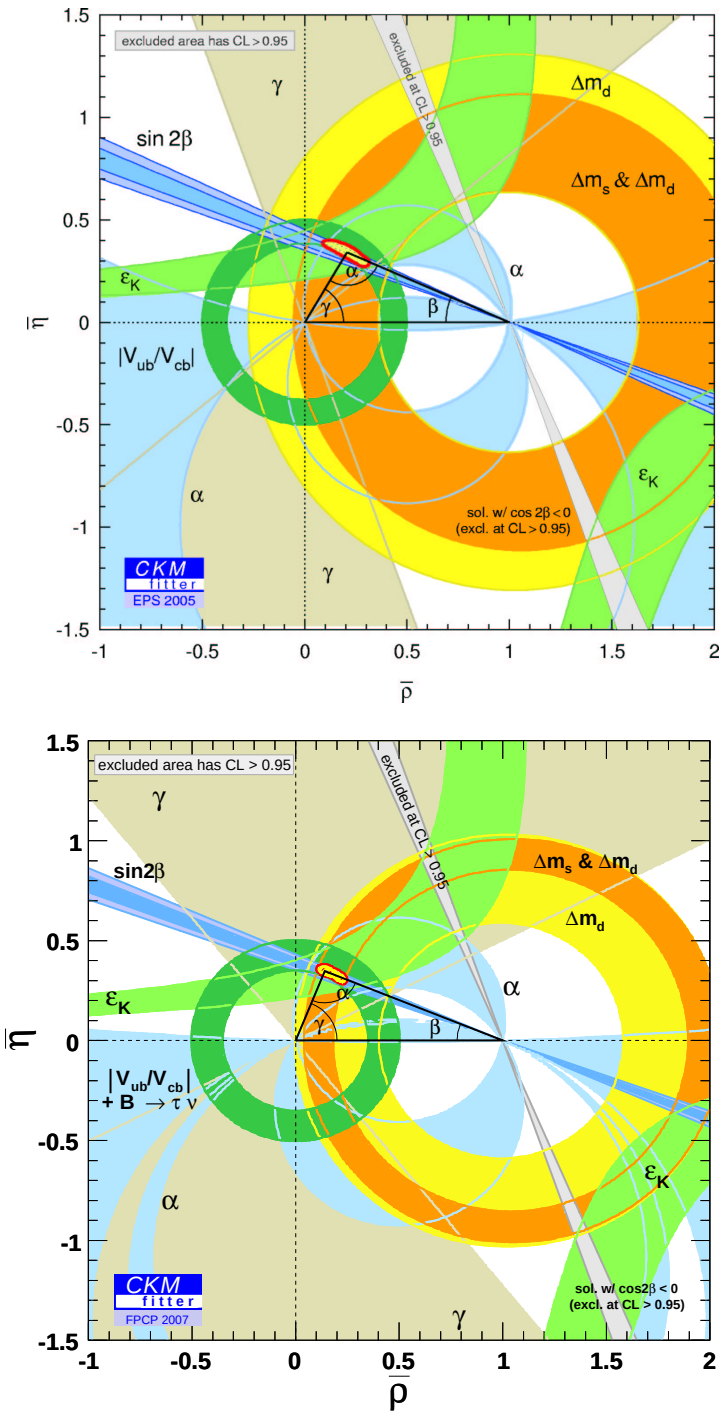

Figure 7-13: Comparison of the unitarity triangle constraints of late-2005 (top) using $\Delta m_s > 14.4$ ps^{-1}, and after definitive observation of B_s oscillations (bottom). The strength of the new $|V_{td}/V_{ts}|$ constraint is limited by QCD theory.

the $(\overline{\rho}, \overline{\eta})$ 95% CL ellipse beyond the contribution of the ϵ_K constraint of the neutral kaon system. The upper-right side R_t of the triangle remains the less-well-constrained

of the ellipse axes, but the change is significant. It should be noted that the recent DØ limit and CDF evidence and measurement also provided significant improvements over the 2005 constraint of Figure 7-13. However, the direct comparison of these plots is justified by the definitive nature of this Δm_s measurement. The central Δm_s value of 17.77 ps^{-1} is consistent with the pre-observation Standard Model expectation of $18.3\ ^{+6.5}_{-1.5}\ \text{ps}^{-1}$. This observation is consistent with the unitarity of the CKM matrix.

7.6 Conclusions

This dissertation demonstrates the contribution of partially reconstructed hadronic decays to the CDF experiment's historic observation of B_s^0-$\overline{B}_s^0$ oscillations [108]. The new signal channels $B_s^0 \to D_s^{*-}\pi^+$ ($D_s^- \to \phi\pi^-$) and $B_s^0 \to D_s^-\rho^+$ ($D_s^- \to \phi\pi^-$) are incorporated into a maximum likelihood fit alongside fully reconstructed $B_s^0 \to D_s^-\pi^+$ ($D_s^- \to \phi\pi^-$) decays, resulting in a robust framework for analysis of flavor oscillations. Development of mass and proper decay time models for the partially reconstructed channels also indicates potential for statistically precise determination of B branching fractions or lifetimes, although the sole focus of the analysis is the oscillation measurement.

Based on the amplitude method, the 3150 partially reconstructed signal events provide an equivalent yield increase of $\sim 65\%$ over the clean sample of 2000 fully reconstructed $B_s^0 \to D_s^-\pi^+$ ($D_s^- \to \phi\pi^-$) decays alone. The fitted amplitude of the oscillation signal peak is consistent with unity for the total sample and for all relevant subsamples, showing that the event-by-event estimates of proper time resolution and tagger dilution are calibrated correctly.

The likelihood depth of this dissertation's analysis sample is $\Lambda = -7.4$, corresponding to a p-value of 0.15% or 3.2 standard deviations. This sample is combined with 3500 fully reconstructed hadronic decays from the five other channels $B_s^0 \to D_s^-\pi^+$ and $B_s^0 \to D_s^-\pi^+\pi^+\pi^-$ with $D_s^- \to \phi\pi^-$, $D_s^- \to K^{*0}K^-$, and $D_s^- \to \pi^-\pi^-\pi^+$, as well as 61 500 semileptonic decays of $B_s \to D_s^{(*)}\ell X$, where ℓ stands for an electron or a muon and the D_s meson decays via the same three modes as the hadronic channels. The combined CDF B_s mixing analysis has a p-value of 8×10^{-8}, corresponding to a Gaussian significance of 5.4 standard deviations. The discovery threshold would not be achieved without inclusion of the partially reconstructed hadronic decays.

The fully and partially reconstructed channels produce consistent measurements of the B_s oscillation frequency Δm_s, and the systematic uncertainties incurred by the new signal decays are small. The combined CDF measurement is

$$\Delta m_s = 17.77 \ \pm 0.10 \ (\text{stat}) \ \pm 0.07 \ (\text{syst}) \, \text{ps}^{-1},$$

where the precise statistical uncertainty arises from the deep and well-behaved minimum of the likelihood profile. The Δm_s measurement allows for extraction of

$$|V_{td}/V_{ts}| = 0.2060 \ \pm 0.0007(\text{exp}) \ {}^{+0.0081}_{-0.0060}(\text{theory}),$$

with uncertainty dominated by lattice QCD theory. The uncertainty contribution of the Δm_s measurement is an order of magnitude smaller than other sources, and the 0.7% precision of Δm_s is better than the 1% level of Δm_d. The measured B_s oscillation frequency is consistent with the pre-observation prediction of the Standard Model and with the unitarity of the CKM matrix.

Appendix A

Systematic Uncertainties on Amplitude Scans

Although systematic uncertainties on amplitude are historically important for setting lower limits on Δm_s, they become largely irrelevant once a statistically significant oscillation signal is available. Specifically, these uncertainties do not affect the measurement of Δm_s or evaluation of the p-value. Moreover, systematic uncertainties are of only minor importance even for the amplitude scans because, in the Δm_s region of sensitivity relevant for this analysis, the results are limited by statistics. For completeness, this appendix describes the evaluation of the systematic uncertainties applied in the amplitude scans of Section 7.1.

All systematic biases for amplitude are evaluated using the toy Monte Carlo approach of Section 7.2.3. For each effect, a large ensemble of MC samples is generated to simulate the potentially mismodeled attribute. Each sample in this biased ensemble is mirrored by a corresponding element of an ensemble generated by the nominal model. The mean shift in fitted amplitude between the pairs is used to calculate the systematic uncertainty contribution via the expression [27]

$$\sigma^{\text{syst}} = \Delta\mathcal{A} + (1 - \mathcal{A})\frac{\Delta\sigma_{\mathcal{A}}}{\sigma_{\mathcal{A}}}. \tag{A.1}$$

As in Section 7.2.3, this method of evaluation allows for the systematic effects to be decoupled from the statistical effects of the toy Monte Carlo samples. The uncertainties for each source are calculated in Δm_s intervals of 5 ps^{-1} across the full frequency range and then interpolated for all other points. To ensure conservatism, each systematic uncertainty is assigned symmetrically even though individual biases are typically in either the positive or negative direction only.

Many possible amplitude biases are considered. The discussion below first lists the sources of uncertainty that are common to all hadronic and semileptonic amplitude scans, followed by the sources that are unique to the partially reconstructed hadronic signals. Specific semileptonic uncertainties are described elsewhere [35]. The dependence of the systematic uncertainties on Δm_s is shown graphically in Figure A-1 for the hadronic analyses. The statistical uncertainty is the dominant contribution to

the overall uncertainty on the amplitude.

Universal Systematic Uncertainties

The following systematic uncertainties are common to all hadronic and semileptonic analyses.

- **Dilution scale factors:** The uncertainty on the magnitude of the predicted dilution for each event is the dominant source of systematic uncertainty for the amplitude. The potential variation of dilution is taken from the respective scale factor calibration analyses for OST [84] and SSKT [89]. These variations are propagated into the toy MC samples, which are fitted under the nominal dilution modeling.

- **Correlated taggers:** In this analysis, the opposite- and same-side tag decisions are combined under the well-studied assumption that the tagger algorithms are uncorrelated. The bias that would result from taggers whose decisions are correlated by as much as 10% are studied with toy MC samples.

- **Detector resolution function:** In addition to dilution values, a well-calibrated amplitude is also dependent on use of the correct proper time resolution. The nominal fit model assumes that the detector resolution function for proper decay time is a Gaussian distribution. While this resolution model is adequate for the B_s samples, the high-statistics vertex calibration samples of Section 5.1.1 indicate that a double-Gaussian plus symmetric exponential tails provides a more precise resolution model. The induced amplitude bias is studied by generating toy MC samples with the more complex model and fitting them with the nominal model.

- **σ_{ct} scale factor:** The σ_{ct} calibration of Section 5.1.1 derives a scale factor parameterization that has residuals of only a few percent. Any resulting over- or under-estimate of the uncertainty on proper decay time of the B_s signal is accounted for by fitting nominal toy MC samples with $\pm 4\%$ variations in the final σ_{ct} scale factor.

- **Potentially large $\Delta\Gamma_s/\Gamma_s$:** The value of $\Delta\Gamma_s/\Gamma_s$ for B_s mesons affects the proper decay time fit of the data and can introduce a bias on the amplitude. The toy MC samples are generated with a conservative value of $\Delta\Gamma_s/\Gamma_s$, which is then neglected to estimate its effect on the fit.

- **Selection bias curves:** The selection bias curves $\varepsilon(ct)$ of Section 5 are important for measurements of lifetime [98]. However, B_s oscillations occur on such a short time scale that this analysis is largely insensitive to the $\varepsilon(ct)$ parameterization. Biases from this aspect of the model are ruled out by generating toy MC samples with different variations of the efficiency curves and fitting them with the nominal set.

Systematic Uncertainties in Hadronic Amplitude Scans

The following systematic uncertainties are unique to the hadronic channels, including the fully and partially reconstructed samples:

- **Dilution templates:** The probability distributions for dilution which are used in the fit model for signal and background are known with finite statistical precision. The toy Monte Carlo is generated with dilution distributions which are varied according to this statistical uncertainty while the nominal templates are applied in the amplitude fit.

- **Fractions of various components:** Several systematic uncertainties are assigned for sample composition. Various toy MC ensembles are generated to address uncertainty on the relative fractions of the Cabibbo-suppressed $B_s^0 \to D_s^- K^+$ and the "reflection" components of Λ_b and B^0. An independent systematic uncertainty is assigned for incomplete knowledge of the fraction of partially reconstructed $\overline{B}_s^0 \to D_s^+ X$ events, as discussed below.

- **Dilution of Cabibbo-suppressed component:** While the Cabibbo-favored decays of $B_s^0 \to D_s^- \pi^+ (\pi^+\pi^-)$ are completely self-tagging for final b-quark flavor, the Cabibbo-suppressed $D_s^- K^+ (\pi^+\pi^-)$ decays receive contributions from two different tree-level amplitudes which have opposite charges in the final state particles. This component is therefore not exactly self-tagging in decay flavor, resulting in an uncertainty on its dilution. The nominal model has the dilution of this small component equal to that of the Cabibbo-favored signal. In fact, these fully reconstructed components use the same time-flavor PDF. To conservatively estimate the possible bias from this choice, the dilution of the Cabibbo-suppressed components is varied strongly. A biased MC ensemble is generated where each experiment has this dilution damped by a random factor between zero and one, simulating a potentially large interference between the contribution amplitudes. These samples are fitted with the nominal dilution model, and the average bias is taken as the systematic uncertainty estimate.

- **Probability templates for σ_{ct}.** The signal components of the likelihood are conditional PDFs which are dependent on event-by-event proper time resolution. However, the hadronic samples are not modeled with additional probability densities for σ_{ct} because the signal and background σ_{ct} distributions are very similar. Minor biases are possible and are estimated by comparing fits of the standard toy MC samples with and without the complete $\mathcal{P}(\sigma_{ct})$ terms in the model.

- **Mixing of B^0 and non-signal satellites:** The nominal fit model assumes that the small contributions of B^0 and partially reconstructed candidates which are not explicitly treated as signal do not oscillate. Based on toy MC samples generated such that these components do mix, this systematic uncertainty takes into account biases that could be caused by neglecting their time-dependent asymmetry.

Systematic Uncertainty in Partially Reconstructed Hadronic Amplitude

The following systematic uncertainties are unique to the partially reconstructed hadronic channels, with values compiled in Table A.1:

- **Relative signal fraction:** Uncertainty on either the relative fractions of the partially reconstructed hadronic signals **or** on their mass template shapes is equivalent to uncertainty on the weight that each signal component receives in the likelihood. Just as in the Δm_s systematic study for the hadronic κ-factors in Section 7.2.3, this uncertainty is conservatively addressed by examining the effect of generating a toy Monte Carlo ensemble composed entirely of one partially reconstructed signal component while performing the amplitude fit as though it were the other. This procedure simulates 100% confusion between the components.

- **Background shape and level:** The nominal model assumes that the combinatorial background is described in mass by smooth exponential behavior across the entire mass range. Appendix F explores the possibility that the low-mass region of the partially reconstructed signals contains more combinatorial background than is accounted for. To evaluate the effect of an additional background component, an ensemble of MC samples is generated from the skewed mass distribution of Figure F-5, in which the background rises much more rapidly under the partially reconstructed signal region than the model predicts. These samples are then fitted for amplitude using the nominal background model, effectively treating the background events as signal.

- **Combinatorial background proper time template:** Based in part on the studies of Appendix F, it is known that the distribution of proper decay time for the combinatorial background changes slowly as a function of mass. The nominal model fits a single ct template for the background across the full 1 GeV/c^2 mass range, effectively averaging over the small variations of the distribution. To conservatively estimate the effect of this choice, toy MC samples are generated with the abruptly mass dependent ct template for background shown in Figure F-7, which are then fitted using the nominal model. This uncertainty is not assigned to the fully reconstructed hadronic decays because the associated mass range is narrower and because of the excellent separation of signal and background.

Source	**0** ps^{-1}	**5** ps^{-1}	**10** ps^{-1}	**15** ps^{-1}	**20** ps^{-1}	**25** ps^{-1}
All hadronic	12.54	13.16	14.15	15.30	16.90	19.11
Relative signal fraction	0.00	0.50	0.90	0.77	0.77	1.37
Background mass shape	5.39	4.55	4.74	4.66	4.29	4.90
Background ct template	4.55	3.46	2.32	1.16	0.28	0.27
Total:	14.4	14.4	15.1	16.1	17.5	20.0

Table A.1: Summary of systematic uncertainties [%] on amplitude for partially reconstructed hadronic channels at various values of Δm_s. The results are displayed graphically in Figure A-1.

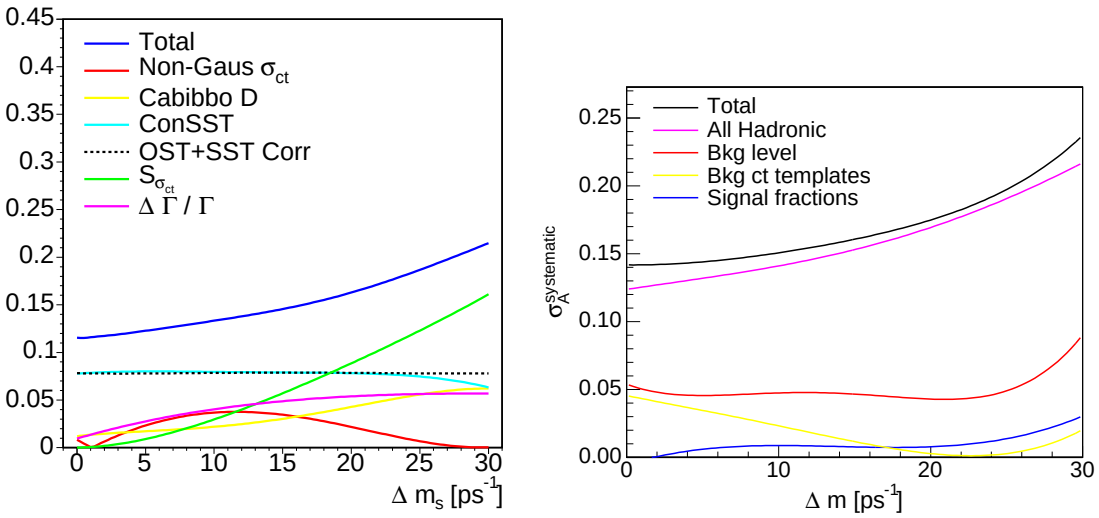

Figure A-1: Graphical summary of systematic uncertainties for all hadronic amplitude scans (left) and for partially reconstructed hadronic channels only (right).

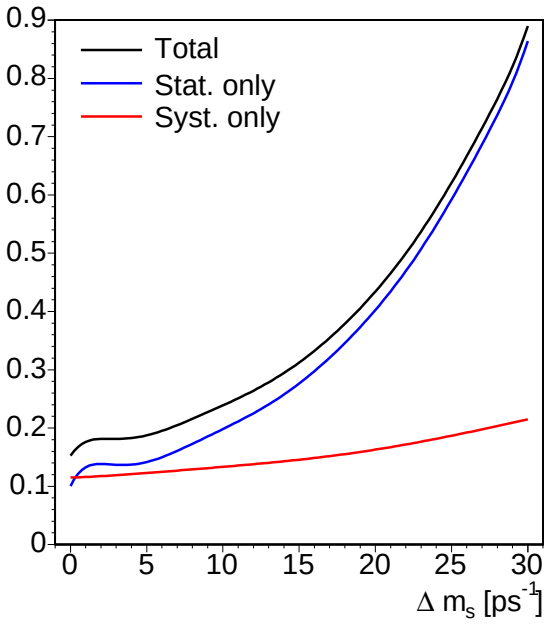

Figure A-2: Comparison of statistical and systematic uncertainties for amplitude scans (fully reconstructed hadronic only). The statistical uncertainty is the dominant uncertainty across the sensitivity region.

Appendix B

Exclusion of Other Δm_s Biases

All non-negligible contributions to the systematic uncertainty on Δm_s are discussed in Section 7.2.3. Because this analysis represents the first use of partially reconstructed hadronic decays in B_s mixing analysis, numerous related sources of possible bias to Δm_s measurement are considered, only one of which is found to be significant. For completeness, this appendix describes the most prominent excluded bias sources. The method of bias investigation uses the toy Monte Carlo approach of Section 7.2.3.

Mean Hadronic κ-Factor Values

As noted in Section 7.2.3, the mean value of a signal component's κ-factor distribution is directly related to its measurement of Δm_s. The final systematic uncertainty estimate contains a contribution from uncertainty about the relative weight applied to each partially reconstructed signal model, but none is assigned for the $F(\kappa)$ distributions themselves. Inaccuracies in the mean $\langle\kappa\rangle$ will translate directly to bias in Δm_s, while incorrectly modeled widths or tails will be propagated into decreased signal significance (given that mismodeling is rarely able to fake a signal and is much more likely to simply damage the likelihood depth). At first glance, this seems a formidable source of systematic uncertainty. However, it is crucial to remember that the models on which the EvtGen [90] decays of $B_s^0 \to D_s^- \rho^+$ and $B_s^0 \to D_s^{*-} \pi^+$ are based are well tested by numerous experiments and are grounded in fundamental physical principles. The spherical wavefunctions through which the Vector $\to$ Scalar Scalar or Vector $\to$ Vector Scalar decays of $\rho^+ \to \pi^+ \pi^0$ and $D_s^{*-} \to D_s^- \gamma$ occur are not subject to doubt. The only basis for questioning the derived κ-factors may stem from the parameters that are input to those models in the Monte Carlo generation. Under the reasonable assumption that the EvtGen models for these modes are reliable, the remaining cause for potential bias is if the phase space that were made available to a decay were incorrectly described. For example, if the D_s^{*-} mass were too large, then the phase space available to the photon or soft pion would be too great, creating a κ-factor which would be wider and with a possibly lower mean than appropriate. Therefore, the settings of the masses and widths of constituent decay particles in the MC simulation are the primary cause for concern in the nominal $F(\kappa)$ distributions.

The EvtGen simulation generally uses the best available world-average [2] masses

for relevant particles, including B_s, D_s^-, and ρ^+. To evaluate the effect of such parameters, these masses are varied in EvtGen according to their uncertainties and new κ-factors are generated. The toy MC approach is used to create a "biased" ensemble of samples, followed by fitting them with the nominal model, thus simulating the effect of incorrect $F(\kappa)$ distributions. The ρ^+ particle must also be characterized by its width, which is varied in the same procedure. These studies reveal completely negligible systematic shifts in Δm_s.

Combinatorial Background Mass Shape and Level

Appendix F explores the possibility that the low-mass region of the partially reconstructed signals contains more combinatorial background than is nominally modeled. In order to estimate the effect of confusing this background for signal content, the analogous toy MC studies of Appendix A are repeated to look for a bias in Δm_s. An ensemble of MC samples is generated from the skewed mass distribution of Figure F-5, which is then fitted according to the nominal model. In addition to mismodeling the background events as signal in proper time space, the fits also simulate incorrect treatment of the dilution templates. Although this effect is a prominent source of bias in amplitude fitting, the result for Δm_s measurement is very small. The systematic Δm_s shift is observed to be 0.002 ± 0.001 ps^{-1}, which does not contribute significantly to the overall uncertainty. This result is reasonable in light of the observation that the rapid oscillations of the signal PDFs are readily visible above the background, whatever its modeled amount.

Combinatorial Background Proper Time Template

Appendix F also explores the effect of the choice to use a single proper time template for the combinatorial background across the total mass region, rather than a more accurate model that accounts for its mass dependence. The analogous toy MC studies of Appendix A are repeated to look for a bias in Δm_s. An ensemble of MC samples is generated from the abruptly mass dependent ct template for background shown in Figure F-7, which is then fitted using the nominal model. Although this effect is a modest source of bias in amplitude fitting, the result for Δm_s measurement is completely negligible.

Physics Background Content

The "wedge" of the $b \to D_s X$ background in the B_s mass space comprises only about $\sim 3\%$ of the total sample, but the shape and content of this partially reconstructed component is less well-determined than for the signals. Additionally, its empirical proper time template in Figure 5-12 is modeled somewhat crudely. The importance of this component in measuring Δm_s is assessed by generating toy MC samples with the nominal model and then completely neglecting its presence in the resulting Δm_s fits. The fractions of Equation (5.29) are renormalized without any contribution from the joint PDF of the physics background, including its mass shape, ct template, or

flavor asymmetry. The effect on measurement of the rapid Δm_s oscillations of the signal components is completely negligible.

Appendix C

Likelihood Formalism

Data for this analysis is fitted using the unbinned maximum likelihood method. For each event, a set of measurements $\vec{x}$ is reconstructed in the detector, while the model contains a vector of parameters $\vec{\vartheta}$ (of dimensionality independent of $\vec{x}$) for which the likelihood is to be maximized. Each event is modeled by a normalized joint PDF of the form

$$\mathcal{P}(\vec{\vartheta}|\vec{x}) = f_{\rm sig} \cdot \mathcal{P}_{\rm sig} + (1 - f_{\rm sig}) \cdot \mathcal{P}_{\rm bkg} \tag{C.1}$$

where the signal and background components are functions of subsets of the parameter vector $\vec{\vartheta}$ and of the fixed measured values $\vec{x}$. The components are weighted by a signal/background fraction which is itself an element of $\vec{\vartheta}$. Additionally, the total signal and background components may themselves be the sums of subcomponents describing different classes of physics events:

$$\mathcal{P}_{\rm sig} = \sum_{\alpha} f_{\alpha} \cdot \mathcal{P}_{{\rm sig},\alpha} \tag{C.2}$$

the details of which are developed in the core chapters of this dissertation.

The multi-dimensionality of the per-event PDF is manifested in the fact that the overall signal and background components are each the product of separately normalized densities for various event attributes, including mass, proper time, and flavor information:

$$\mathcal{P}_{\rm sig,bkg} = \mathcal{P}^{\rm mass}_{\rm sig,bkg} \cdot \mathcal{P}^{\rm time}_{\rm sig,bkg} \cdot \mathcal{P}^{\rm flav}_{\rm sig,bkg} \cdots \tag{C.3}$$

In this framework, an event that exhibits strongly signal-like or background-like properties in any of its modeled attributes is assigned an appropriately strong weight in the overall PDF. More typically, the small indications that an event is slightly more likely to be of a given type of signal or background are considered collectively, generating a more accurate final discriminant. In a one-dimensional fit, which is equivalent to integrating over all subspaces except one in Equation (C.3), such multi-faceted modeling information is neglected, and thus the data is not optimally used.

The total likelihood for a data sample is formed from the product of the joint PDF from each event i:

$$\mathcal{L}(\vec{\vartheta}) = \prod_{i} \mathcal{P}_i(\vec{\vartheta}|\vec{x}_i). \tag{C.4}$$

In practice, because it is computationally simpler to work with sums rather than products, we choose to calculate the natural logarithm of the sample likelihood

$$\ln \mathcal{L}(\vec{\vartheta}) = \sum_i \ln \mathcal{P}_i(\vec{\vartheta}|\vec{x}_i). \tag{C.5}$$

The logarithm of a function increases monotonically with the function, so maximization of $\ln \mathcal{L}$ is equivalent to maximization of $\mathcal{L}$. The problem is turned into one of *minimization* by application of a factor of -2, where the 2 is necessary for correct uncertainty estimates. The solution to the most probable parameter values may be extracted by extremizing the sample likelihood function with respect to $\vec{\vartheta}$:

$$\frac{\partial(-2\ln \mathcal{L})}{\partial \vec{\vartheta}} = 0 \tag{C.6}$$

which is performed numerically via the minimization program Minuit [32, 33].

Another advantage of the likelihood method is that the combination of data samples is relatively straight-forward, regardless of how differently they may be modeled. For each sample s, an appropriate per-event PDF is constructed and a total log-likelihood $\ln \mathcal{L}_s$ is calculated as in Equation (C.5). The combined log-likelihood is simply

$$\ln \mathcal{L}_{\text{total}}(\vec{\Theta}) = \sum_s \ln \mathcal{P}_s(\vec{\vartheta}_s) \tag{C.7}$$

where $\vec{\Theta}$ denotes the combined set of fitted parameters used in any or all of the sample-specific parameter sets $\vec{\vartheta}_s$. In this framework, it is important to note that parameters of $\vec{\Theta}$ which represent physical constants being fitted, such as B lifetimes or mixing frequencies, must be *common* to all sample-specific sets $\vec{\vartheta}_s$. In contrast, parameters of the sets $\vec{\vartheta}_s$ which are used purely for empirical modeling of backgrounds or other non-universal effects may be varied independently. Finally, in direct analogy with the single-sample fit in Equation (C.6), the most probable combined parameter set may be extracted by minimizing $-2\ln \mathcal{L}_{\text{total}}$ with respect to $\vec{\Theta}$.

Implicit in this discussion is the fact that this framework is implemented for *unbinned* fitting. That is, rather than integrating information in discrete intervals of mass, time, and any other space, each event is taken as an entry in calculation of the total likelihood. This approach allows for extraction of maximal information from a given data sample and its likelihood model.

A final advantage of the likelihood method is that it provides a comprehensive test for evaluation of the signal significance. The *likelihood ratio* between the function evaluated under the assumption of an oscillation signal and with that of the null hypothesis is a definitive measure of the signal strength. As introduced in Section 1.4.1, the hypothesis of flavor oscillations is ultimately dependent on the single parameter of amplitude. Although the ratio test would be valid for any degree of modeling complexity, Chapter 7 demonstrates that the evidence for the oscillation signal becomes particularly simple.

Appendix D

Likelihood Framework Validation

The complete analysis likelihood is a complex function of about one hundred parameters, describing several B candidate classes in multiple measurement spaces. As charted in this dissertation, the development of the likelihood begins with the simplest components of the mass spectra, followed by the addition of progressively more complex terms. At each stage of development, the likelihood framework is verified to produce reliable, unbiased results for the physically important parameters.

The validation procedure is performed with large ensembles of "toy" Monte Carlo samples. These samples are generated from parent distributions that exactly reflect the composition and structure of the nominal likelihood model. This method is in contrast to the realistic MC samples, which simulate the full interaction of particles with the detector and the trigger and reconstruction logic. The toy samples neglect all such underlying effects and reproduce the data at only the highest level. Representative parameter values are used for all components and their normalizations. The unbiased likelihood should be minimized by a parameter set that is consistent with the generating set, within statistical fluctuations. From an ensemble of $\sim \mathcal{O}(1000)$ toy MC samples, a "pull" distribution of the Gaussian deviation between fitted and generating values

$$Z_i = \frac{x_i^{\rm fit} - x_i^{\rm gen}}{\sigma_{x_i}^{\rm fit}} \tag{D.1}$$

is compiled for each free parameter. The distribution of pulls Z_i is unit-Gaussian for unbiased parameters and uncertainty estimates.

Pull studies serve two major purposes for the analysis. First, on a practical level, they are effective at revealing mistakes in the likelihood software. The code that generates each attribute of the toy samples (for example, the lost momentum of a partially reconstructed signal decay) is completely distinct from the code that implements the corresponding term in the likelihood. Potential mistakes are therefore discovered by pull studies before they might bias the fitted data. The second purpose of fitter validation is to test whether it should be *expected* that the likelihood can fit a parameter in an unbiased manner, given the existing sample composition and model. For example, it would not be reasonable to expect that the nominal fit of the signal lifetime in Section 5.5 could also resolve the proper time distribution of a small

component such as the physics background. Without the "assistance" of its fixed ct template, the $\sim 3\%$ weight of the $b \to DX$ component, which overlaps strongly in the mass space and is not well distinguished, would probably be pulled towards a fluctuation. The resulting pull distribution would not be unit-Gaussian, and such a situation is part of the reasoning for extracting templates from Monte Carlo simulation. In constrast, it is critical that physically important parameters *can* be resolved from the nominal sample composition. Obvious examples include sample fractions, Δm, amplitude, and flavor tagging parameters for signal components. Were these parameters not measurable in the data with the current sample composition, a different solution would be required for extracting their values. Good likelihood behavior requires that important parameters produce unit-Gaussian pulls, while modest biases to empirically oriented parameters are of little concern.

This appendix documents the most relevant results of the numerous pull studies conducted for this analysis. As noted in the main body of the text, the fitter is validated for every discrete component of the likelihood, including individual mass template shapes, sculpted and smeared signal lifetime distributions, background proper time templates, flavor asymmetries of separate components, *etc.* Only the combined likelihood pulls are documented here.

Figure D-1 shows the pull distributions of an ensemble of 1500 fits to the wide mass space, such as is studied in Chapter 4. The toy MC samples are generated with parameter values of the B^+ data in order to test the fully generalized mass likelihood, including the additional $B^0 \to D^{*-}\pi^+$ component. Table D.1 lists the Gaussian fits of these pull distributions, including mean, width, and χ^2 probability. The important signal fraction parameters exhibit unbiased fits, producing unit-Gaussian means and widths.

Parameter	Mean	Width	Prob [%]
M_{B^+}	-0.052 ± 0.027	1.010 ± 0.021	77.5
σ_1	-0.069 ± 0.028	0.988 ± 0.020	0.5
f_1	0.189 ± 0.026	0.913 ± 0.020	0.0
r_2	0.018 ± 0.033	0.994 ± 0.024	0.0
$f_{B^+\to\overline{D}^0\pi^+}$	0.010 ± 0.026	0.979 ± 0.018	68.9
$f_{B^+\to\overline{D}^0\rho^+}$	0.026 ± 0.026	0.978 ± 0.019	65.1
$f_{B^+\to\overline{D}^{*0}\pi^+}$	-0.081 ± 0.027	1.010 ± 0.020	65.3
$f_{B^0\to D^{*-}\pi^+}$	0.000 ± 0.026	0.979 ± 0.018	45.5
$f_{\rm com}$	-0.062 ± 0.026	0.967 ± 0.018	35.0
$\lambda_{\rm com}$	0.069 ± 0.026	0.970 ± 0.017	29.1
$f_{\rm lin}$	0.057 ± 0.027	0.978 ± 0.019	0.7

Table D.1: Gaussian pull results from 1500 fits of the mass space, using parameter values representative of the B^+ data. The important signal fraction parameters exhibit unit-Gaussian means and widths.

Table D.2 lists the Gaussian pulls of 600 fits of the proper time space without flavor identification, such as is studied in Chapter 5. Representative values for the

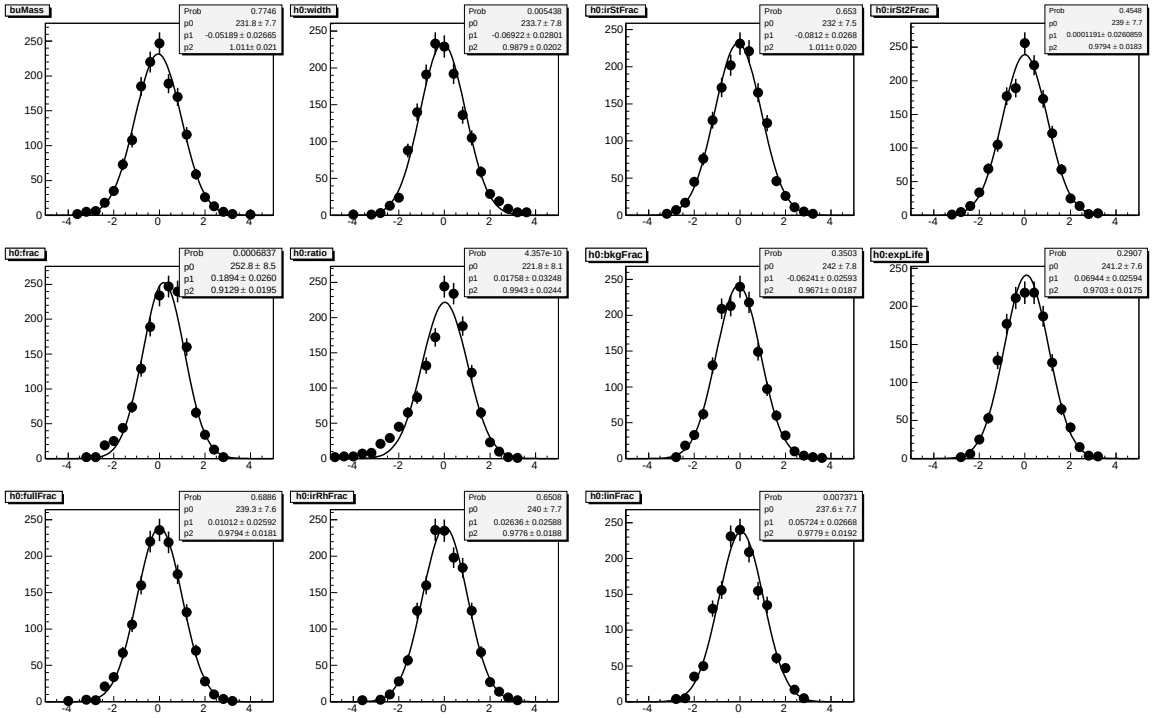

Figure D-1: Pull distributions from 1500 fits of the wide mass space, using parameter values representative of the B^+ data.

B^+ data are again used for full likelihood generality. The most important result is that the overall background fraction is unbiased, while the common signal lifetime also has unit pulls.

Parameter	Mean	Width	Prob [%]
$c\tau_B$	0.027 ± 0.042	0.992 ± 0.032	59.5
$f_{\rm com}$	0.048 ± 0.043	1.030 ± 0.031	90.6
$c\tau_1$	-0.121 ± 0.047	1.120 ± 0.038	56.4
δ_1	0.121 ± 0.043	1.010 ± 0.029	24.8
σ_1	-0.038 ± 0.040	0.959 ± 0.028	72.2
f_2	0.167 ± 0.049	1.130 ± 0.046	7.7
$c\tau_2$	-0.209 ± 0.047	1.100 ± 0.038	53.7
δ_2	0.059 ± 0.047	1.100 ± 0.037	27.5
σ_2	-0.045 ± 0.049	1.150 ± 0.045	3.0

Table D.2: Gaussian pull results from 600 fits of the proper time space without flavor identification, using parameter values representative of the B^+ data. The fits of overall background fraction and signal lifetime are unbiased.

Table D.3 lists the Gaussian pulls of 900 fits of flavor asymmetry, using parameter values representative of the B^0 data in Chapter 6. The oscillation frequency and signal

tagging values are unbiased. Similar validation is performed with $\Delta m_s \approx 17$ ps^{-1}.

Parameter	Mean	Width	Prob [%]
Δm_d	-0.039 $\pm$ 0.034	0.972 $\pm$ 0.024	34.7
$S_{\mathcal{D}}^{\mathrm{SSKT}}$	0.064 $\pm$ 0.034	0.981 $\pm$ 0.027	65.7
$S_{\mathcal{D}}^{\mathrm{OST}}$	0.010 $\pm$ 0.036	1.030 $\pm$ 0.028	49.9
ϵ^{SSKT}	-0.017 $\pm$ 0.035	1.000 $\pm$ 0.026	50.8
ϵ^{OST}	-0.004 $\pm$ 0.033	0.942 $\pm$ 0.025	74.3

Table D.3: Gaussian pull results from 900 fits of flavor asymmetry, using parameter values representative of the B^0 data.

Table D.4 shows the Gaussian pulls from 5000 fits of amplitude, using parameter values representative of the B_s data with $\Delta m_s = 17$ ps^{-1}. Figure D-2 shows the associated pull distribution. Measurement of amplitude from realistic samples is unbiased.

Parameter	Mean	Width	Prob [%]
$\mathcal{A}$	0.021 $\pm$ 0.014	1.016 $\pm$ 0.010	35.8

Table D.4: Gaussian pulls from 5000 fits of amplitude using parameter values representative of the B_s data and $\Delta m_s = 17$ ps^{-1}.

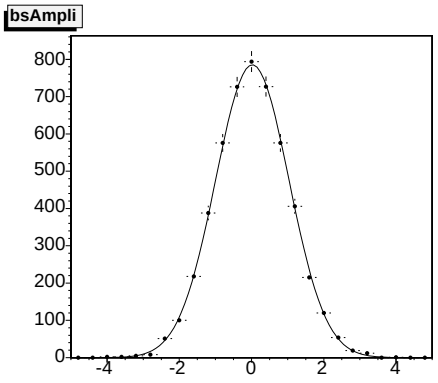

Figure D-2: Pull distribution from 5000 fits of amplitude using parameter values representative of the B_s data and $\Delta m_s = 17$ ps^{-1}.

Appendix E

Particle Identification

Particle identification (PID) tools play an important role in many facets of this analysis. The Same Side Kaon Tagger for B_s flavor tagging (Section 6.3) is critically dependent on its ability to reliably select fragmentation kaons. Kaon identification is also used for the Opposite Side Kaon Tagger (Section 6.2.3) and for final selection of some fully reconstructed [83] and semileptonic [35] B decay channels in analyses complementary to this one. Furthermore, identification of leptons is used at several stages of this analysis, including rejection of semileptonic decays during hadronic candidate selection (Section 3.3), selection of opposite side lepton tags (Section 6.2), and rejection of leptons from same side kaon tag candidates (Section 6.3).

Since the same side tagger is so essential for the B_s oscillation sensitivity of any CDF analysis, this appendix presents a basic outline of the tools for kaon identification. However, because lepton ID is most central to implementation of the semileptonic mixing analysis, the complete description of the lepton likelihoods and their development is available elsewhere [35].

Charged hadronic particles can be distinguished via measurement of the specific ionization per unit track length ($\mathrm{d}E/\mathrm{d}x$) in the COT and by the time-of-flight measured in the TOF system. These two techniques are described below, followed by the method of their combination.

E.1 Particle ID Using $\mathrm{d}E/\mathrm{d}x$

The design of drift chambers is based on the fact that charged particles traversing the gas volume leave ionization trails. The Bethe-Bloch formula [2] describes the mean energy loss per particle due to ionization:

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{4\pi N e^4}{mc^2\beta^2} q^2 \left[\ln \frac{2mc^2(\beta\gamma)^2}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \tag{E.1}$$

where the particle has charge q and speed βc, N is the number density of electrons in the medium, m and e are the electron mass and charge, I is the mean excitation energy of the atoms in the medium, and $\delta(\beta\gamma)$ is the relativistic correction of the density effect at high speeds. The formula indicates that $\mathrm{d}E/\mathrm{d}x$ depends only on

the velocity factor $\beta\gamma$ for a given particle in a gas of fixed properties. Therefore, when combined with measurement of particle momentum, $\mathrm{d}E/\mathrm{d}x$ may be used to distinguish particles of differing mass.

Although the COT drift chamber is optimized for fast and precise tracking rather than $\mathrm{d}E/\mathrm{d}x$ measurement, careful calibration of the $\mathrm{d}E/\mathrm{d}x$ response results in good statistical power of separation between particle species. As noted in Section 2.2.2, the width of the discriminator output for each COT hit encodes its integrated charge, which is directly proportional to specific ionization. This $\mathrm{d}E/\mathrm{d}x$ response and and its resolutions are studied in pure samples of pions, kaons, protons, electrons, and muons. These calibration samples include decays of $D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi^+$, $\Lambda^0 \to p^+\pi^-$, $J/\psi \to \ell^+\ell^-$, and conversion electrons. Substitution of appropriate values into Equation (E.1) produces a so-called "universal curve" of Figure E-1 (left), with different regions of the $\mathrm{d}E/\mathrm{d}x$ versus $\beta\gamma$ curve dominated by different species.

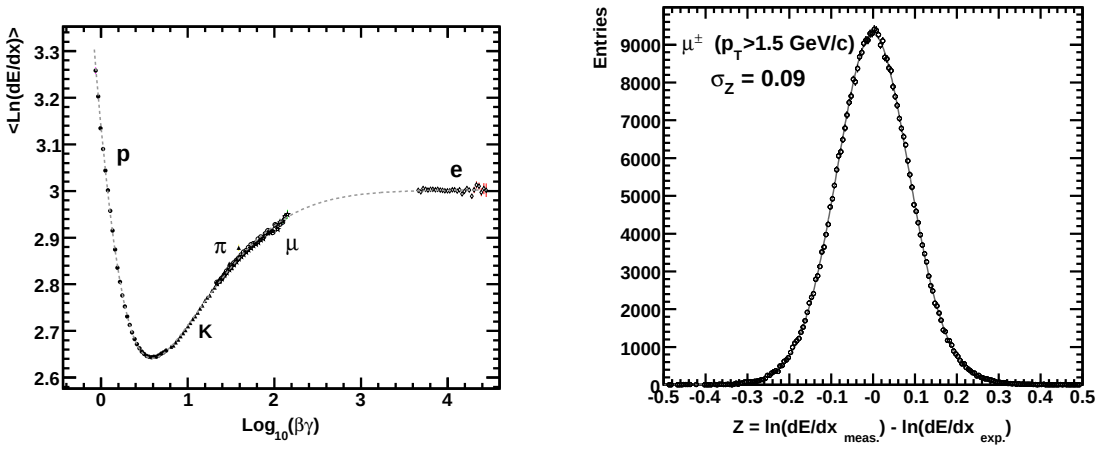

Figure E-1: Left: Average $\ln(\mathrm{d}E/\mathrm{d}x)$ versus particle $\beta\gamma$ as measured for various particle species in CDF. Right: $\mathrm{d}E/\mathrm{d}x$ resolution measured for muons with $p_T > 1.5$ GeV/c. The measured width of the $Z(\mu) = \ln(\mathrm{d}E/\mathrm{d}x_{\mathrm{meas}}) - \ln(\mathrm{d}E/\mathrm{d}x_{\mathrm{expected}})$ distribution corresponds to a resolution of about 4%.

The probability density and resolution of $\mathrm{d}E/\mathrm{d}x$ measurements are estimated by distributions of the variable

$$Z(i) = \ln\left(\frac{\mathrm{d}E/\mathrm{d}x_{\mathrm{meas}}}{\mathrm{d}E/\mathrm{d}x_{\mathrm{exp}}}\right), \quad i = \pi, K, p, e, \mu \tag{E.2}$$

which is evaluated for each track using the measured value of $\mathrm{d}E/\mathrm{d}x_{\mathrm{meas}}$ in comparison with the expected value $\mathrm{d}E/\mathrm{d}x_{\mathrm{exp}}(i)$ for the hypothesis of species i. For pure particle samples, Z is described by a single Gaussian distribution with typical resolution width of $4-5\%$. Figure E-1 shows the Z distribution of muons with $p_T > 1.5$ GeV/c. Resolution estimates for transfer to the data are extracted from the pure calibration

samples, using the dependence of σ_Z on tracking variables such as the number of contributing COT sense wires and track p_T. Much like the calibration of scale factors for proper time resolution, the $\mathrm{d}E/\mathrm{d}x$ calibration is performed for each of the three major data-taking periods which sum to 1 fb^{-1}.

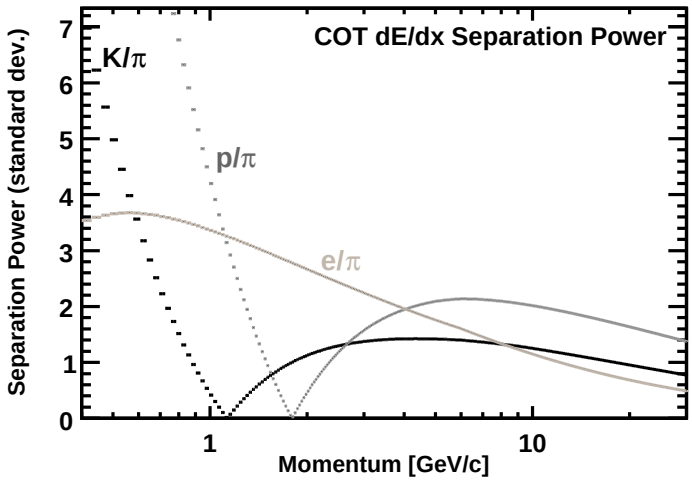

Figure E-2: Measured performance of $\mathrm{d}E/\mathrm{d}x$ particle identification, showing separation power in units of standard deviations versus particle momentum. The K-π separation is shown in black, with p-π in gray and e-π in light gray.

Figure E-2 shows the final $\mathrm{d}E/\mathrm{d}x$ discriminating power between particle species. The statistical separation of pions and kaons is equivalent to 1.4 standard deviations for tracks with $p > 2$ GeV/c. The efficiency for associating $\mathrm{d}E/\mathrm{d}x$ information with a charged COT track of $p_T > 400$ MeV/c in the pseudorapidity range $|\eta| < 1$ is nearly 100%. The $Z(i)$ distributions from the calibration samples are normalized to produce the $\mathrm{d}E/\mathrm{d}x$ probability density $\mathcal{P}_{\mathrm{d}E/\mathrm{d}x}(i)$ for each particle species i.

E.2 Particle ID Using TOF

Whereas $\mathrm{d}E/\mathrm{d}x$ information generally yields the best K/π separation at momenta above 2 GeV/c, the TOF detector provides excellent complementary separation for $p < 1.5$ GeV/c.

As described in Section 2.2.3, the TOF detector measures the arrival time t of particles at its scintillator bars with respect to the $p\overline{p}$ interaction at time t_0. Combined with the COT measurement of momentum p and track path length L, a particle's mass m is determined by

$$m = \frac{p}{c}\sqrt{\frac{c^2 t_{\mathrm{flight}}^2}{L^2} - 1}. \tag{E.3}$$

The uncertainty on $t_{\rm flight}$ receives contributions from both instrumental resolution (i.e. the detector and calibration) and the determination of the particle production time t_0. The latter component of the resolution arises from that fact that the $p\overline{p}$ interaction region within the detector is longitudinally distributed over $\sim$ 30 cm, creating an inherent time uncertainty of a few nanoseconds per track. However, by averaging the time measurements for every particle in an event and assuming that their species composition follows a predicted distribution, the typical t_0 resolution is reduced to about 50 ps for $b\overline{b}$ events.

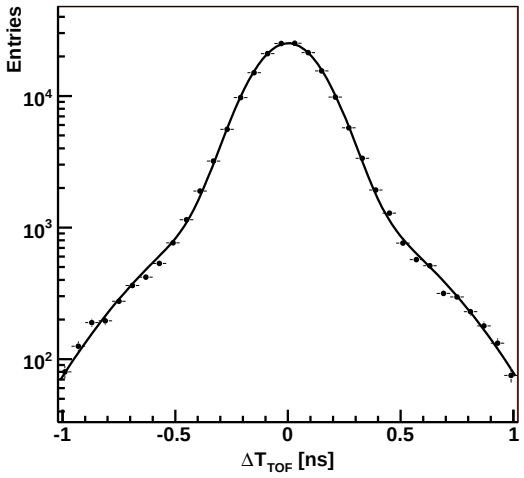

Figure E-3: Distribution of the difference Δ_{TOF} between the time-of-flight measured for the two muons in $J/\psi \to \mu^+\mu^-$ decays. The width is used to determine the instrumental TOF time resolution $\sigma_{\rm TOF} = \sigma_{\Delta_{\rm TOF}}/\sqrt{2} = 110$ ps.

The contribution of instrumental resolution to $\sigma_{t_{\rm flight}}$ is studied with a large sample of $J/\psi \to \mu^+\mu^-$ decays. The measured time-of-flight difference $\Delta_{\rm TOF}$ between the two muons is dominated by the detector resolution, as opposed to real kinematic differences, and is therefore a direct indication of the $t_{\rm flight}$ resolution. An additional advantage for the method is that the collision time t_0 cancels in the difference. Figure E-3 shows a distribution of $\Delta_{\rm TOF}$ for the J/ψ sample, which is fitted with a double-Gaussian. The narrow component has widths ranging from $140 - 170$ ps, depending on the scintillator bar and period of data-taking, and it accounts for $\sim 85\%$ of the distribution area. The wide Gaussian arises from tracks associated with incorrect TOF information and has widths of several hundred picoseconds. Based on the narrow Gaussian width, the instrumental TOF time resolution is estimated to be $\sigma_{\rm TOF} = \sigma_{\Delta_{\rm TOF}}/\sqrt{2} = 110$ ps.

Similar to the method of the dE/dx calibration, the TOF probability densities $\mathcal{P}_{\rm TOF}(i)$ are constructed using the difference between the measured and expected time-of-flight for particle hypothesis i. These calibrations are again performed with large samples of pure pions, kaons, and muons from $D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi^+$, $K^0_s \to \pi^+\pi^-$, and $J/\psi \to \mu^+\mu^-$ decays. Multiple particles hitting the same TOF bar

tend to distort the measurement, creating significant inefficiency in associating time information with such tracks. The overall efficiency for associating TOF information with charged particles in typical $b\bar{b}$ events varies from $50-65\%$ and decreases with instantaneous luminosity due to increased track occupancy in the detector.

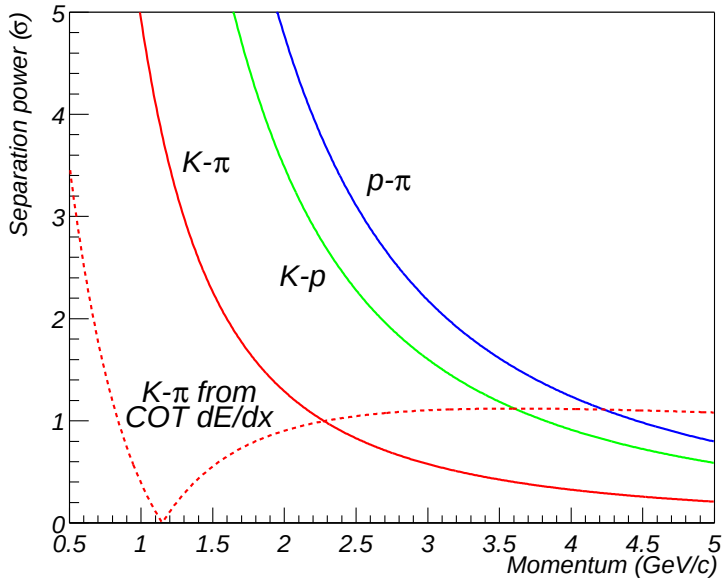

Figure E-4: Performance of the TOF system, showing separation power in units of standard deviations versus particle momentum. The dashed line of the COT $\mathrm{d}E/\mathrm{d}x$ K/π separation power is overlaid for comparison of the systems' complementary performance.

The separation power of the TOF system is depicted in Figure E-4. The K/π separation is better than 2 standard deviations for track momenta in the regime $p < 1.4$ GeV/c, which is heavily populated in the data samples of this analysis. The low-momentum separation power of the TOF system is well-complemented by the high-momentum power of $\mathrm{d}E/\mathrm{d}x$, whose K/π separation is superimposed on Figure E-4.

E.3 Combined Likelihood for Kaon ID

By combining the low-momentum separation power of the TOF system with the complementary high-momentum power of $\mathrm{d}E/\mathrm{d}x$ measurements, good particle identification significance may be achieved for a large majority of CDF tracks.

For identification of kaons, the calibrated measurements from the two systems are merged into a single "combined log-likelihood" (CLL) variable which is defined as

$$CLL(K) = \log\left(\frac{\mathcal{P}(K)}{f_p\mathcal{P}(p) + f_\pi\mathcal{P}(\pi)}\right), \text{ with } \mathcal{P}(i) = \mathcal{P}_{\mathrm{TOF}} \cdot \mathcal{P}_{\mathrm{d}E/\mathrm{d}x}, \tag{E.4}$$

where $f_p = 0.1$ and $f_\pi = 0.9$ are assumed as the a priori probabilities for the background composition. $\mathcal{P}_{\mathrm{TOF}}$ and $\mathcal{P}_{\mathrm{d}E/\mathrm{d}x}$ represent the probability densities developed for the separate systems. In the case that a track has no information from one system or the other, the associated probability is fixed to 1 for the CLL calculation.

Using this combined quantity, a K/π separation of at least 1.5 standard deviations is obtained for track momenta up to 5 GeV/c, with the significance increasing toward low momentum. Figure E-5 shows the CLL separation power as a function of track momentum in comparison with the individual power of the TOF and dE/dx measurements. The performance is measured in a sample of kaons and pions from the clean decay chain $D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi^+$. The combined variable provides an improvement in K/π separation power of $10 - 25\%$, depending on momentum.

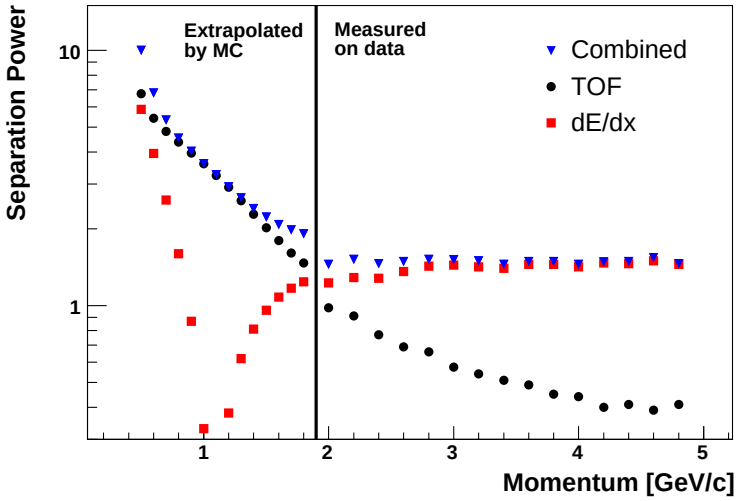

Figure E-5: K/π separation power of the combined particle identification quantity $CLL(K)$ in standard deviations, as compared to the individual TOF and dE/dx systems. The track sample is obtained from the clean decay chain $D^{*+} \to D^0\pi^+$ ($D^0 \to K^-\pi^+$).

Appendix F

Combinatorial Background Studies

The expansion of hadronic B_s mixing analysis into the low-mass regions of partially reconstructed decays involves several decisions regarding combinatorial background modeling. This appendix compiles the discussion of issues and effects that inform the nominal model and evaluation of systematic uncertainties.

F.1 Background Content in the Partially Reconstructed Region

In analyses of only fully reconstructed decays, modeling of combinatorial background in the mass space is simple because its exponential shape does not need to be extrapolated very far. Use of the partially reconstructed signals adds an element of complexity because the background cannot be directly fitted in the low-mass region. Although the fixed mass templates of the partially reconstructed signals provide good models against which to compare the combinatorial level, the shape of the combinatorial background remains largely extrapolated from the upper sideband. The fitted mass spectra of Figures 5-13 - 5-15 illustrate the nominal background modeling for analysis of proper time and flavor. Although the B_s and B^+ mass spectra provide no cause for concern regarding this procedure, the high-statistics B^0 sample displays a small unexpected step. Moreover, the use of neural networks for B candidate selection is not yet well-established at CDF, which warrants a generally cautious approach. Therefore, two studies are performed using candidate isolation and SSKT dilution in an attempt to independently measure the combinatorial background content of the partially reconstructed mass region.

Isolation

Candidate isolation, defined as the fractional momentum carried by the B candidate among the momentum of all tracks within a cone of $\Delta R < 0.7$, is known to be an excellent discriminant of signal and background. Figure F-1 shows the normalized distributions of signal and background for $B_s^0 \to D_s^- \pi^+$ where the signal distribution is extracted from the sideband-subtracted fully reconstructed peak and

the background is from the pure upper sideband. The background distribution is clearly shifted toward low isolation, as is reasonable for a sample component made of random tracks.

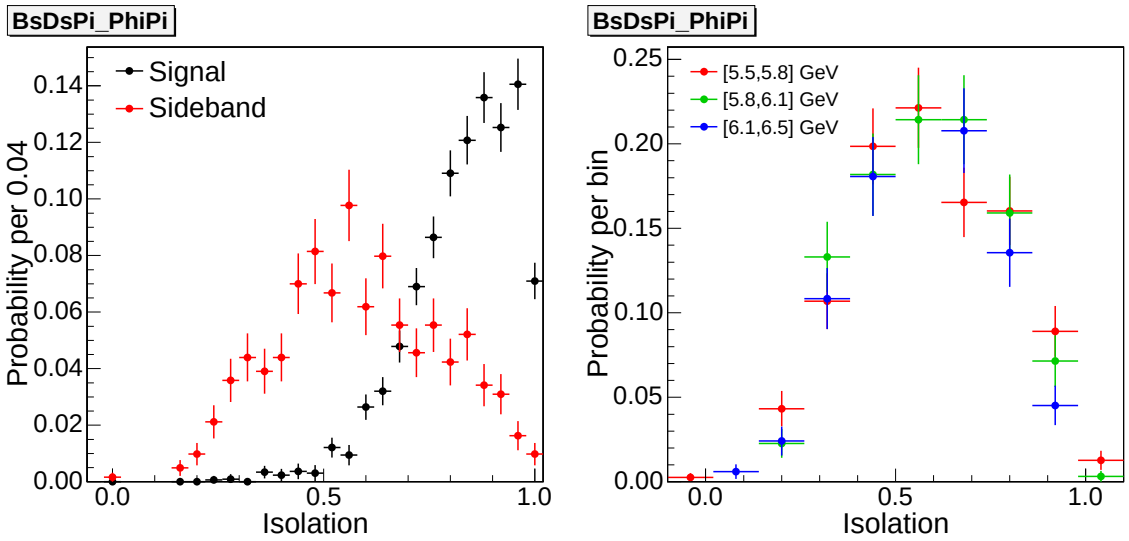

Figure F-1: Normalized isolation distributions for background and fully reconstructed signal in B_s (left), and the statistically consistent distributions of consecutive sidebands, demonstrating the mass-independence of background isolation (right).

By compiling the isolation distributions of data candidates in coarse mass bins across the partially reconstructed region, it should be possible to make a linear combination of the depicted signal and background isolation histograms and thereby deduce the best fits for background fraction. The clean separation of signal and background in isolation space provides an independent measure of background content from that of the mass fit. This procedure is dependent on the assumption that the isolation distribution of background candidates is invariant of mass. Figure F-1 also shows that consecutive sidebands have statistically identical distributions, in support of this assumption. The mass space is sliced into bins of 25 MeV/c^2 each, and the associated isolation distributions are fitted with the normalized signal and background templates using a simple χ^2 minimization. The resulting background fractions are scaled by the total candidate population per bin and the appropriate bin widths to produce the expected "spectrum" of background content in Figure F-2. The data points are fitted to a simple linear curve to illustrate the expected background level.

This projected background level may be compared to the nominal B_s mass fit of Figure 5-15, which is reproduced in the left plot of Figure F-5 below. The projected level of about 25 candidates per bin at the low mass edge of 5.0 GeV/c^2 is only moderately higher than the nominal level of 20 candidates.

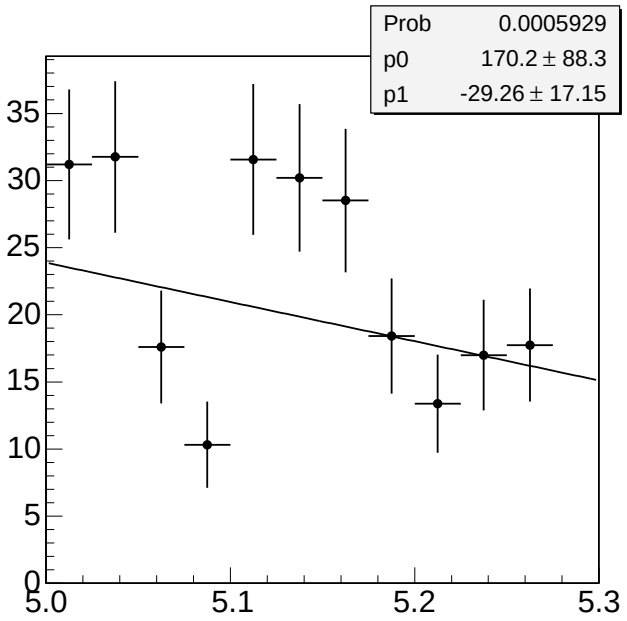

Figure F-2: Combinatorial background level as fitted in isolation space, scaled for comparison to the 10 MeV/c^2 binning in the left plot of Figure F-5.

SSKT Dilution

The same exercise as above is performed with SSKT dilution. Dilution distributions are expected to be less powerful than isolation as a signal/background discriminant, but it is known that the distributions are different between the two. (Indeed, the dilution templates are required in the likelihood because of this fact, as noted in Section 6.5.) Figure F-3 shows the normalized distributions of signal and background for $B_s^0 \to D_s^- \pi^+$ where the signal distribution is the result of sideband-subtraction of the fully reconstructed peak and the background is from the pure upper sideband. This signal distribution is different from the one in the nominal model, which accounts for the dilution properties of the total partially reconstructed region. However, the goal here is to make a prediction about the background level in that region. Figure F-3 also shows various sideband distributions, illustrating that dilution is independent of mass and may be used to fit the background content in the partially reconstructed region.

The mass space is again sliced into bins of 25 MeV/c^2 each, and the associated dilution distributions are fitted with the normalized signal and background templates using a simple χ^2 minimization. The resulting background fractions are scaled by the total candidate population per bin and the appropriate bin widths to produce the expected spectrum of background content in Figure F-4. The data points are fitted to a simple linear curve to illustrate the expected background level.

This projected background level may be compared again to the nominal B_s mass fit reproduced in the left plot of Figure F-5. In contrast to the level projected from

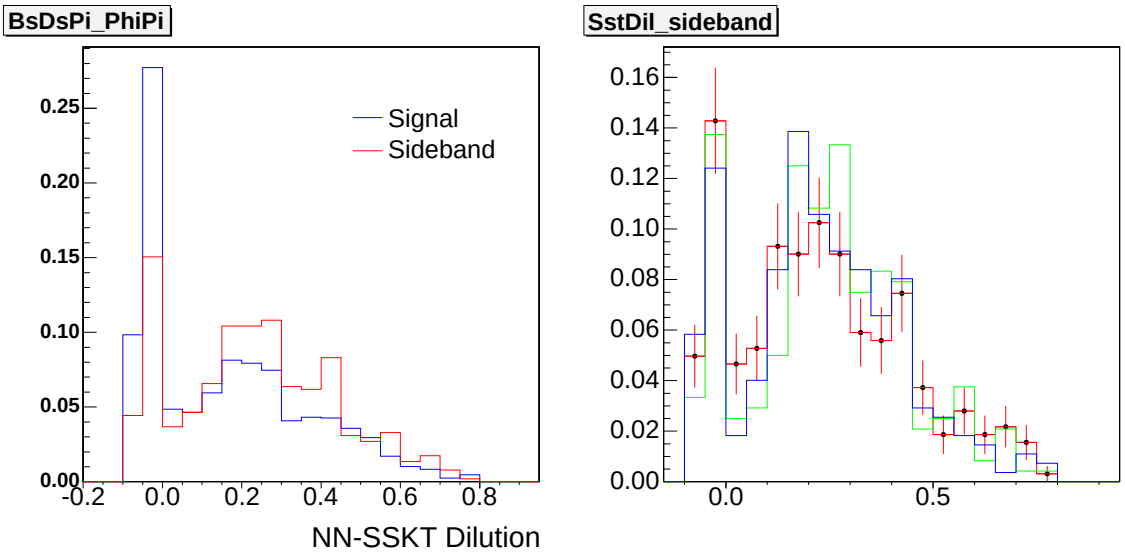

Figure F-3: SSKT dilution templates for B_s combinatorial background and fully reconstructed signal (left), and the statistically consistent distributions of consecutive sidebands, demonstrating the mass-independence of background dilution (right).

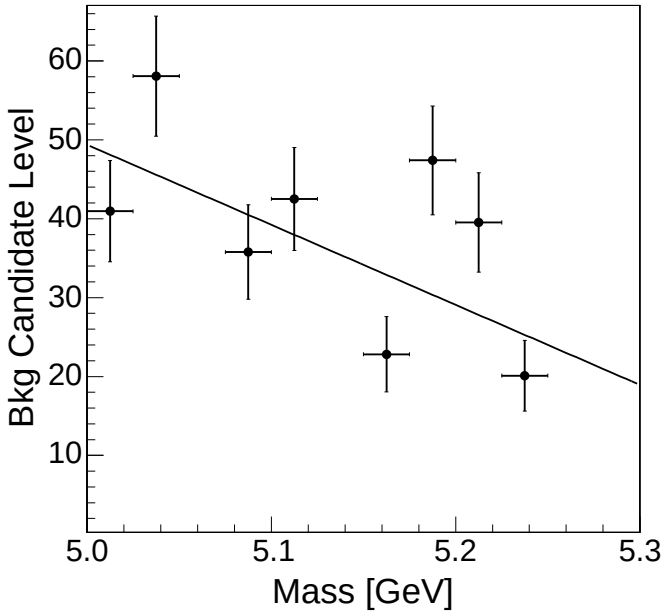

Figure F-4: Combinatorial background level fitted in SSKT dilution space, scaled for comparison to the 10 MeV/c^2 binning in the left plot of Figure F-5.

isolation space, which did not disagree strongly with the nominal model, the dilution predicts a level of about 50 candidates per bin at the low mass edge of 5.0 GeV/c^2.

The dilution space therefore indicates that the background content at the lowest end of the mass range is underestimated by a factor of about 2.

Background Variation for Systematic Uncertainties

The exercises above provide conflicting evidence for the accuracy of the nominal combinatorial background model. The isolation variable, which provides very good signal/background discrimination, indicates that the nominal model may only slightly underestimate the background level in the partially reconstructed region. The SSKT dilution, which weakly separates the candidate classes, indicates a more severe underestimate. So as to avoid any unreported biases in this first-ever observation of B_s oscillations, the systematic uncertainties are evaluated conservatively with the more pessimistic of the two predictions. In systematic studies of both Δm_s and amplitude measurement, as reported in Section 7.2.3 and Appendix A, the effect of an error in combinatorial background level is evaluated by generating toy Monte Carlo samples with additional background content similar to what is fitted in dilution space. These skewed samples are then fitted with the nominal model. Figure F-5 depicts the nominal B_s mass model and the skewed spectrum for systematic studies. For the skewed samples, the relative fractions of signal and physics background components is kept unchanged. The resulting systematic uncertainties are modest for amplitude and completely negligible for Δm_s.

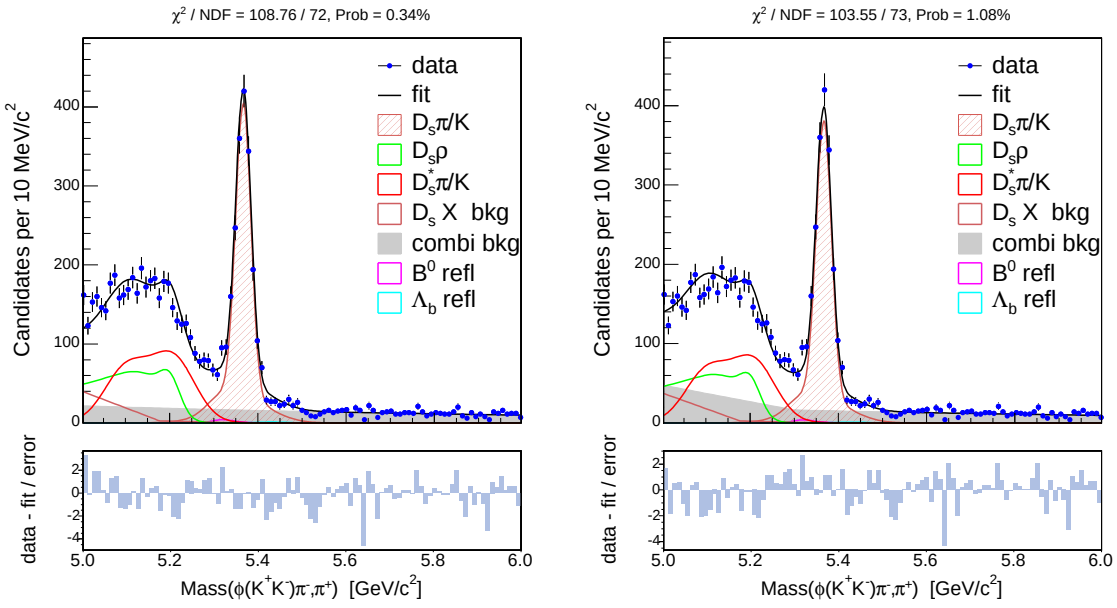

Figure F-5: Left: nominal mass model for 1 fb^{-1} of B_s candidates. Right: skewed combinatorial background model for systematic uncertainty evaluation in toy MC samples. Relative signal fractions are held constant.

F.2 Proper Time Distributions of Combinatorial Background

The shape of the ct distribution for combinatorial background has a dependence on candidate mass, which presents a difficulty for rigorous lifetime measurements in the partially reconstructed mass region. Lifetime fits of only fully reconstructed signal decays, which are performed in the mass range [5.3,6.0] GeV/c^2 for B_s, produce systematically different background ct distributions than those of lifetime fits of the total signal range [5.0,6.0] GeV/c^2. The distributions vary further still in fits of only the partially reconstructed region [5.0,5.3] GeV/c^2. Figure F-6 clearly shows the variation of the ct distributions of consecutive mass sidebands in B^0 data.

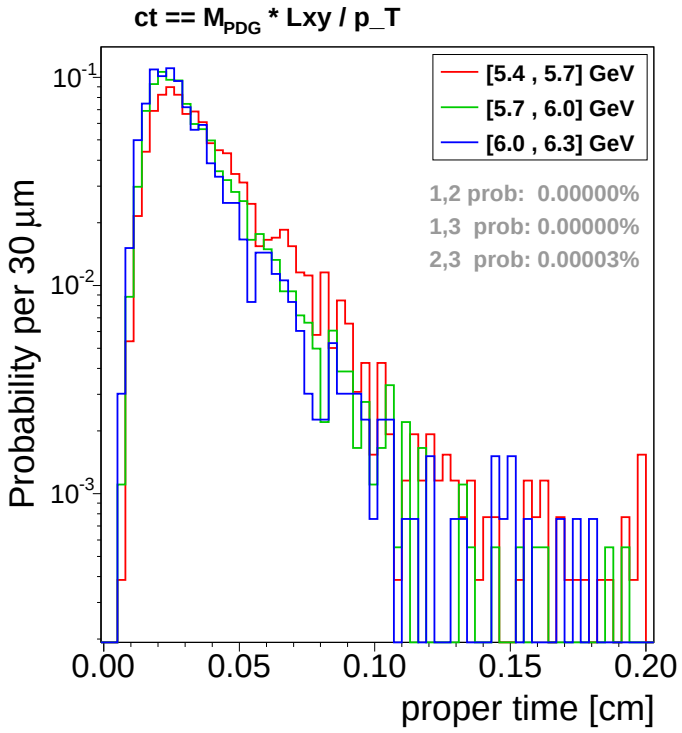

Figure F-6: Variation of the proper time distributions in consecutive mass sidebands of B^0 data, including χ^2 probabilities.

Such an effect is relatively unimportant for analyses of fully reconstructed decays, in which the mass window is narrow and the background component is easily resolved over nearly its full range. The fitted background ct distribution is a weighted average of the true distributions across the mass range. The same is true of lifetime fits in a wider range including partially reconstructed signal channels. However two differences stand out: the combinatorial background is more difficult to resolve in the overlapping components of the low-mass region, thus possibly giving the high-mass side of the spectrum more weight, and the window is simply wider, implying that the

mass-dependent behavior of the background ct is being averaged over a greater span. The contribution of the proper time space to signal/background weighting, although clearly subordinate to the stronger discrimination of the mass space, may therefore be slightly misweighted for candidates at the extremes of mass.

It should be understood that such difficulties may present a significant source of systematic bias for measurements of B *lifetimes*, but are largely irrelevant for B_s mixing. Systematic uncertainties are evaluated in Appendices A and B to support this assertion. The procedure for these estimates involves generation of toy Monte Carlo experiments in which the combinatorial background ct is sampled from a different distribution for the upper and lower mass regions. These experiments are subsequently fitted with a single average ct template to study the effects on the important flavor parameters. The generator ct distributions are extracted from fits of the high-statistics B^0 data, using the most extreme variations observed in an upper mass sideband and from a fit of only the partially reconstructed low-mass region. Figure F-7 shows the parameterizations of these templates. Even while forcing the fitter to average a single background ct model over such large mass-dependent variation, the observed bias to amplitude is modest while the systematic shift in measured Δm_s is completely negligible.

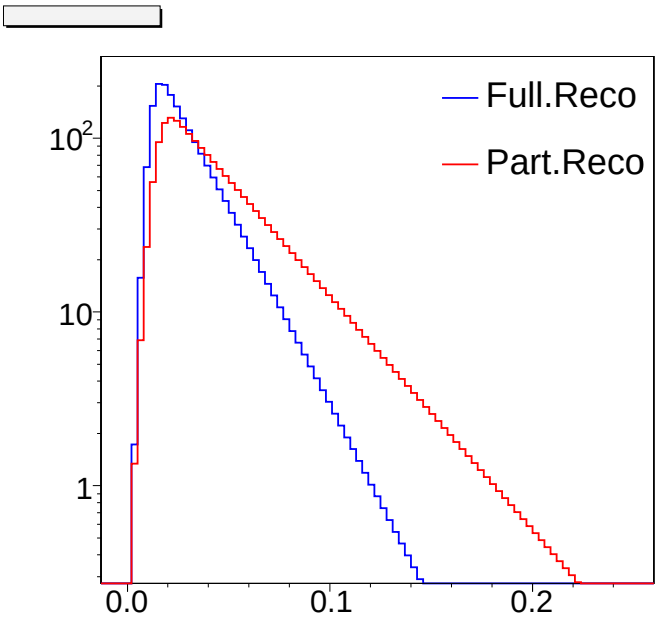

Figure F-7: Extreme variation of combinatorial background proper time distributions used for evaluation of systematic uncertainties.

Bibliography

[1] S.L. Glashow, Nucl. Phys. **22**, 579 (1961);
S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967);
A. Salam, *Elementary Particle Theory*, ed. N. Svartholm, Stockholm, Almquist & Wiksell, 367 (1968).

[2] Particle Data Group, W.M. Yao *et al.*, J. Phys. G **33**, 1 (2006).

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

[4] M. Gell-Mann, Phys. Rev. Lett. **8**, 214 (1964).

[5] S.L. Glashow, J. Iliopoulos, L. Maiani, Phys. Rev. **D2**, 1585 (1970).

[6] J.J. Aubert *et al.*, Phys. Rev. Lett. **33**, 1404-1406 (1974).

[7] J.E. Augustin *et al.*, Phys. Rev. Lett. **33**, 1406-1408 (1974).

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

[9] M. Kobayashi, T. Maskawa, Progr. Theor. Phys. **49**, 652 (1973).

[10] S.W. Herb *et al.*, Phys. Rev. Lett. **39**, 252 (1977).

[11] F. Abe *et al.*, Phys. Rev. Lett. **74**, 2626 (1995).

[12] L. Wolfenstein, Phys. Rev. Lett. **51**, 1915 (1983).

[13] M. Battaglia *et al.*, *The CKM Matrix and the Unitarity Triangle*, `hep-ph/0304132`.

[14] J. Charles *et al.* (CKMfitter Group), Eur. Phys. J. C **41**, 1 (2005) `hep-ph/0406184`, see http://ckmfitter.in2p3.fr/.

[15] D. Boutigny *et al.*, BABAR *Technical Design Report*, SLAC-R-0457 (1995).

[16] M.T. Cheng *et al.*, *Belle Technical Design Report*, BELLE-TDR-3-95 (1995).

[17] A. Drutskoy *et al.*, *Results from the* $\Upsilon(5S)$ *Engineering Run (Belle)*, `hep-ex/0605110` (2006).

[18] K. Abe *et al.* (BELLE Collaboration), Phys. Rev. D **71**, 072003 (2005);
N. C. Hastings *et al.* (BELLE Collaboration), Phys. Rev. D **67**, 052004 (2003);
B. Aubert *et al.* (BABAR Collaboration), Phys. Rev. Lett. **88**, 221803 (2002).

[19] V. Weisskopf, E.P. Wigner, Z. Phys. **63**, 54 (1930).

[20] K. Anikeev *et al.*, *B Physics at the Tevatron: Run II and Beyond*, `hep-ph/0201071` (2001).

[21] A.J. Buras, W. Slominski, H. Steger, Nucl. Phys. B **245**, 369 (1984).

[22] M. Okamoto, *Full Determination of the CKM Matrix Using Recent Results from Lattice QCD*, PoS LAT2005 013 `[hep-lat/0510113]` (2005).

[23] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **96**, 202001 (2006).

[24] C. Albajar *et al.* (UA1 Collaboration), Phys. Lett. **B**186, 247 (1987).

[25] H. Albrecht *et al.* (ARGUS Collaboration), Phys. Lett. **B**192, 245 (1987).

[26] M. Artuso *et al.* (CLEO Collaboration), Phys. Rev. Lett. **62**, 2233 (1989).

[27] H.G. Moser, A. Roussarie, Nucl. Instr. Meth. A **384**, 491 (1997).

[28] J. Abdallah *et al.* (DELPHI Collaboration), Eur. Phys. J. C **35**, 35 (2004);
K. Abe *et al.* (SLD Collaboration), Phys. Rev. D **67**, 012006 (2003);
A. Heister *et al.* (ALEPH Collaboration), Eur. Phys. J. C **29**, 143 (2003).

[29] Heavy Flavor Averaging Group, http://www.slac.stanford.edu/xorg/hfag/.

[30] V.M. Abazov *et al.* (DØ Collaboration), Phys. Rev. Lett. **97**, 021802 (2006).

[31] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **97**, 062003 (2006).

[32] F. James, **MINUIT***: Function Minimization and Error Analysis*, CERN Program Library Long Writeup D506, http://consult.cern.ch/writeup/minuit/ (1994).

[33] R. Brun *et al.*, Nucl. Instr. Meth. A **389**, 81-86 (1997). http://root.cern.ch/ .

[34] N. Leonardo, Ph. D. thesis, Massachusetts Institute of Technology,
Analysis of B_s Flavor Oscillations at CDF,
FERMILAB-THESIS-2006-18 (2006).

[35] V. Tiwari, Ph. D. thesis, Carnegie Mellon University,
Measurement of the B_s Oscillation Frequency with Semileptonic Decays,
FERMILAB-THESIS-2007-09 (2007).

[36] G. Aubrecht *et al.*, *A Teachers Guide to the Nuclear Science Wall Chart*, Contemporary Physics Education Project (2003). http://www.lbl.gov/abc/wallchart/teachersguide/pdf/Chap11.pdf.

[37] B. Worthel *et al.*, *Linac Rookie Book* (2006).
http://www-bdnew.fnal.gov/operations/rookie_books/LINAC_v2.pdf.

[38] C.W. Schmidt, *The Fermilab 400-MeV Linac upgrade*,
FERMILAB-CONF-93-111 (1993).

[39] B. Evanger *et al.*, *Accelerator Concepts Rookie Book* (2006).
http://www-bdnew.fnal.gov/operations/rookie_books/Concepts_v3.1.pdf.

[40] B. Worthel *et al.*, *Booster Rookie Book* (2006).
http://www-bdnew.fnal.gov/operations/rookie_books/Booster_V3_1.pdf.

[41] T. Asher *et al.*, *Main Injector Rookie Book* (2006).
http://www-bdnew.fnal.gov/operations/rookie_books/Main_Injector_v1.pdf.

[42] E. Harms, J. Morgan *et al.*, *The Antiproton Source Rookie Book* (2006).
http://www-bdnew.fnal.gov/operations/rookie_books/Pbar_V1_1.pdf.

[43] J. Marriner, *Stochastic Cooling Overview*, FERMILAB-CONF-03-158 (2003).

[44] P. Karns *et al.*, *Recycler Rookie Book* (2006).
http://www-bdnew.fnal.gov/operations/rookie_books/Recycler_RB_v1.pdf.

[45] S. Nagaitsev, *Antiproton Cooling in the Fermilab Recycler Ring*,
FERMILAB-CONF-05-550-AD (2005).

[46] Fermilab Accelerator Division, *Tevatron Rookie Book* (2006).
http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v1.pdf.

[47] A. Tollestrup, *The Tevatron Hadron Collider: A Short History*,
FERMILAB-CONF-94-378-E (1994).

[48] O.S. Bruning, *LHC progress and commissioning plans*,
CERN-LHC-PROJECT-REPORT-922 (2006).

[49] R. Blair *et al.*, *The CDF-II detector: Technical design report*,
FERMILAB-PUB-96/390-E (1996).

[50] T.K. Nelson *et al.*, *The CDF Layer 00 Detector*, FERMILAB-CONF-01/357-E
(2001).

[51] A. Sill *et al.*, Nucl. Instr. Meth. A **447**, 1 (2000).

[52] T. Affolder *et al.*, Nucl. Instr. Meth. A **485** , 6 (2002).

[53] K.T. Pitts *et al.*, *CDF Central Outer Tracker*, FERMILAB-CONF-96-443-E
(1996).

[54] A. Affolder *et al.*, *The Effect of Dead-timeless Silicon Strip Readout at CDF*,
FERMILAB-CONF-02/026-E (2002).

[55] R. Wagner *et al.*, *CDF Central Outer Tracker*, FERMILAB-PUB-03-355-E (2003).

[56] D. Acosta *et al.* (CDF Collaboration), *Measurement of Prompt Charm Meson Production Cross Sections in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, FERMILAB-PUB-03-355-E (2003).

[57] G. Gomez-Ceballos *et al.*, *The CDF-II Time of Flight Detector*, CDF Public Note 6258 (2003).

[58] L. Balka *et al.*, Nucl. Instr. Meth. A **267**, 272 (1988).

[59] M.G. Albrow *et al.*, Nucl. Instr. Meth. A **480**, 524 (2002).

[60] G. Apollinari, K. Goulianos, P. Melese, M. Lindgren, Nucl. Instr. Meth. A **412**, 515 (1998).

[61] M. Gallinaro *et al.*, *A New Scintillator Tile-Fiber Preshower Detector for the CDF Central Calorimeter*, FERMILAB-CONF-04-325-E (2004).

[62] S. Bertolucci *et al.*, Nucl. Instr. Meth. A **267**, 301 (1988).

[63] C.M. Ginsburg *et al.*, Eur. Phys. J. C **33**, s1002 (2004).

[64] CDF Collaboration, *The 1992 CDF Muon System Upgrade*, CDF Public Note 2858;
P. Schlabach, *CDF Detector Muon Upgrades for Run II*, http://www-cdf.fnal.gov/internal/upgrades/muon/muon_upgrades.html.

[65] G. Ascoli *et al.*, Nucl. Instr. Meth. A **268**, 33 (1988).

[66] A. Meyer *et al.*, *The CDF Data Acquisition System for Tevatron Run II*, FERMILAB-CONF-01-242-E (2001).
W. Badgett *et al.*, *Design, Performance and Control of the CDF Run II Data Acquisition System*, FERMILAB-CONF-05-233-E (2005).

[67] N. Leonardo *et al.*, *Event Builder and Level 3 Manual for Experts*, CDF Note 6138 (2002).

[68] J. Antos *et al.*, *Data Processing Model for the CDF Experiment*, FERMILAB-PUB-06-169-CD-E (2006).

[69] E.J. Thomson *et al.*, IEEE Trans. Nucl. Sci **49**, 1063 (2002).

[70] W. Ashmanskas *et al.*, Nucl. Instr. Meth. Phys Res. A **447**, 218 (2000);
A. Bardi *et al.*, Nucl. Instr. Meth. Phys Res. A **485**, 6 (2001);
W. Ashmanskas *et al.*, IEEE Trans. Nucl. Sci **49**, 1177 (2002).

[71] H. Ray *et al.*, *Level 2 Electron Trigger Efficiencies for Run II*, CDF Public Note 6098 (2002).

[72] I.K. Furić, A. Belloni, Ch. Paus, *Multibody Trigger Paths in the Two-Track Trigger Data*, CDF Note 6526 (2003).

[73] K. Anikeev, Ph. D. thesis, Massachusetts Institute of Technology, *Measurement of the Lifetimes of B Meson Mass Eigenstates*, FERMILAB-THESIS-2004-12 (2004).

[74] F.D. Snider *et al.*, *Tracking at CDF: Algorithms and Experience from Run I and Run II*, FERMILAB-CONF-05-573-CD (2005).

[75] M. Feindt, S. Menzemer, K. Rinnert, P. Schemitz, A. Skiba, *A Fast and Versatile Kalman Track-Fitter for CDF II*, CDF Note 5388 (2000); M. Campanelli, E. Gerchtein, *Calibration of the Momentum Scale for Kalman Refitter Using J/psi Events*, CDF Note 6905 (2004).

[76] A. Korn, R. Snyder, J. Boudreau, http://cdfkits.fnal.gov/CdfCode/source/TrackingUserHL/.

[77] J. Miles, Presentation to the CDF BPAK Group, *L00 in B Lifetimes*, March 5, 2004, http://www-cdf.fnal.gov/internal/physics/bottom/bpak.

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

[79] D. Acosta *et al.* (CDF Collabboration), Phys. Rev. D **71**, 052003 (2005); K. Burkett, J. Guimaraes da Costa, D. Sherman, *The Iterative Seed in PrimeVtx*, CDF Note 7859 (2005).

[80] Ch. Paus *et al.*, http://cdfkits.fnal.gov/CdfCode/source/BottomMods/.

[81] P. Murat *et al.*, http://cdfkits.fnal.gov/CdfCode/source/Stntuple/.

[82] A. Zell *et al.*, *Stuttgart Neural Network Simulation*, Univ. of Stuttgart (1993).

[83] A. Abulencia *et al.* (CDF Collaboration), *Observation of B_s^0-$\overline{B}_s^0$ Oscillations*, Article in preparation for Phys. Rev. D.

[84] CDF B_s Mixing Group, *Updated Measurement of Δm_d and Calibration of Flavor Tagging in Fully Reconstructed Decays*, CDF Public Note 7920 (2005).

[85] T. Sjostrand, L. Lonnblad, S. Mrenna, *PYTHIA 6.2: Physics and manual*, `hep-ph/0108264` (2001).

[86] K. Anikeev, P. Murat, Ch. Paus, *Description of Bgenerator II*, CDF Note 5092 (1999).

[87] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 032001 (2005).

[88] R. Field, Phys. Rev. D **65**, 094006 (2002).

[89] A. Belloni, Ph. D. thesis, Massachusetts Institute of Technology, *Observation of B_s-$\overline{B}_s$ Oscillations and the Development and Application of Same-Side-Kaon Flavor Tagging*, CDF Public Note 8988 (2007).

[90] D.J. Lange, Nucl. Instr. Meth. A **462**, 152-155 (2001); J. Beringer *et al.*, *User Guide for the New EvtGen at CDF*, CDF Public Note 7917 (2005).

[91] R. Brun *et al.*, *GEANT3*, CERN-DD/EE/84-1 (1987).

[92] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **98**, 122002 (2007).

[93] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **96**, 191801 (2006).

[94] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **98**, 061802 (2007).

[95] CDF B_s Mixing Group, *Scale Factors for Proper Time Uncertainties*, CDF Note 7818, CDF Public Note 7944 (2005).

[96] G. Punzi, *Comments on Likelihood Fits with Variable Resolution*, `hep-ex/0401045` (2004).

[97] A. Deisher *et al.*, *Determination of B^0 and B^+ Lifetimes Using Partially and Fully Reconstructed Modes*, CDF Note 8778 (2007).

[98] CDF B_s Mixing Group, *B Mesons Lifetime Determination in Fully Hadronic Decays*, CDF Public Note 7386 (2005).

[99] C. Lecci, Ph. D. thesis, University of Karlsruhe, *A Neural Jet Charge Tagger for the Measurement of the B_s^0-$\overline{B}_s^0$ Oscillation Frequency at CDF*, FERMILAB-THESIS-2005-89 (2005).

[100] CDF B_s Mixing Group, *Opposite Side Kaon Tagger*, CDF Public Note 8179 (2006).

[101] CDF B_s Mixing Group, *Combined Opposite Side Flavor Tagger*, CDF Public Note 8460 (2006).

[102] CDF B_s Mixing Group, *Measurement of B^0 Oscillations Using Same-Side Tagging in Semileptonic B Decays*, CDF Public Note 7081 (2004).

[103] CDF B_s Mixing Group, *Same Side Kaon Tagging in Fully Reconstructed Decays*, CDF Public Note 8206 (2006).

[104] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **97**, 082004 (2006).

[105] M. Feindt, U. Kerzel, Nucl. Instr. Meth. A **559**, 190-194 (2006).

[106] J. Piedra, Ph. D. thesis, University of Cantabria, *Determination of Δm_d and Absolute Calibration of Flavor Taggers for the Δm_s Analysis, in Fully Reconstructed Decays at the CDF Experiment*, FERMILAB-THESIS-2005-27 (2005).

[107] A. Ali, F. Barreiro, Z. Phys. C **30**, 635 (1986);
M. Gronau, A. Nippe, J.L. Rosner, Phys. Rev. D **47**, 1988 (1993);
M. Gronau, J.L. Rosner, Phys. Rev. D **49**, 254 (1994).

[108] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **97**, 242003 (2006).