ABSTRACT

Title of dissertation:	SEMITAUONIC B_c^+ DECAYS AND QUARK FLAVOR IDENTIFICATION METHODS
	Jack Wimberley, Doctor of Philosophy, 2018
Dissertation directed by:	Professor Hassan Jawahery Department of Physics

The LHCb experiment at the Large Hadron collider is a unique laboratory for studying the properties of heavy quarks. The physics program of the experiment includes studies of CP violation, measurements of CKM matrix parameters, searches for rare decays, quarkonia studies, and other flavor physics, forward physics, and new physics topics. This thesis presents an analysis of the semitauonic branching fraction $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)$ of the doubly-heavy B_c^+ meson, which serves as a powerful probe of the universality of the couplings of leptons (e, μ and τ) in electroweak interactions. The ratio of this branching fraction to the semimuonic branching fraction $\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$ is measured to be $\mathcal{R}(J/\psi) = 0.71 \pm 0.17 \,(\text{stat}) \pm 0.18 \,(\text{syst})$. A second topic of the thesis is the creation of a new algorithm for tagging the flavor of neutral mesons in CP violation studies, and a powerful method for calibrating these flavor tagging algorithms via binomial regression.

SEMITAUONIC B_c^+ DECAYS AND QUARK FLAVOR IDENTIFICATION METHODS

by

Jack Wimberley

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Abolhassan Jawahery, Chair/Advisor
Professor Kaustubh Agashe
Professor Sarah Eno
Professor Nick Hadley
Professor Eric V. Slud

Dedication

To Jen, whose encouragement, advice, and love have supported me while I was finishing this dissertation, and who corrected this sentence's grammar.

Acknowledgments

There are many people and institutions to thank for their support during my pursuit of a doctoral degree in physics. The University of Maryland, with the help of the National Science Foundation, has supported me financially and has provided an excellent environment for my studies and research. Furthermore, many wonderful graduate classes led by the excellent professors of the Department of Physics have shaped the direction of my research. I have also had the great fortune to be a member of CERN, and of the LHCb collaboration in particular; it has been a pleasure to take part in this passionate community of scientists and engineers, and my work would have been impossible without their years of accumulated efforts.

My work has been in collaboration with the members of the University of Maryland's LHCb group and our colleagues elsewhere. In particular, I would like to thank Professor Hassan Jawahery for his leadership, advice, and interesting suggestions of research projects. My work on semitauonic B_c^+ decays was significantly influenced by the work of Brian Hamilton and Greg Ciezarek (CERN). Similarly, my effort in the development of the OS charm tagger for LHCb owes much to Riccardo Cenci (now at INFN Pisa), who began its development as a post-doctorate in our group. My development of new methods for calibrating flavor tagging algorithms benefitted greatly from meetings with Professor Eric V. Slud. Many others provided a helping hand or patient ear over the years, including my fellow graduate student and officemate Jason E. Andrews and many members of the LHCb collaboration. I also appreciate the efforts of the remaining members of my dissertation committee, Profs. Sarah Eno, Nick Hadley, and Kaustubh Agashe.

The path towards completing my research and earning a doctorate in physics did not begin once I entered graduate school, and the efforts required did not stop at the boundaries of the University's campus. My parents Darryl and Doris, and sister Morgan, encouraged my love of math and physics from a young age and sacrificed

much to put me through college as a full-time student. The professors of my alma mater Boston College (too many to name, but including Profs. Krzysztof Kempa, Michael Graf, and Rein Uritam) inspired and pushed me to pursue a graduate degree in physics. Finally, I owe the most to my girlfriend Jen, whose love, patience, and encouragement over the years of my research were key to the completion of this dissertation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

SM	Standard Model
QED	quantum electrodynamics
QCD	quantum chromodynamics
C	charge symmetry
P	parity symmetry
CP	charge-parity symmetry
CPT	charge-parity-time symmetry
CKM	Cabibbo-Kobayashi-Maskawa
FCCC	Flavor-changing charged current
FCNC	Flavor-changing neutral current
MSSM	Minimal Supersymmetric (extension of the) Standard Model
LHC	Large Hadron Collider
ECAL	electromagnetic calorimeter
VELO	VErtex LOcator
TT	Tracker Turicensis
IT	Inner Tracker
OT	Outer Tracker
HCAL	hadronic calorimeter
PID	particle identification
RICH	ring-imaging Cherenkov detector
IP	impact parameter
PV	primary vertex
MC	Monte Carlo
EFG	Ebert, Faustov, and Galkin
BCL	Bourelly, Caprini, and Lellouch
AUC	area under the curve (a classifier figure-of-merit)

Part I

Introduction

Chapter 1: The Standard Model

1.1 Particles and forces of the Standard Model

The Standard Model (SM) is the quantum field theory of three gauge forces (the electromagnetic, weak, and strong forces) that act on quarks and leptons, the particles that comprise matter. The matter particles of the SM are all spin-1/2 fermions, and come in three families (or generations) with increasingly higher masses. There are two classes of matter particles: quarks and leptons. In each generation, there are two quarks and two leptons, shown in Table 1.1. The vast bulk of everyday matter is made up of fermions from the first and lightest generation: the up and down quarks (u and d), which combine to form protons and neutrons, and the electron (e). The particles of the second and third generation are essentially identical, except that they are more massive (with the possible exception of the neutrinos, whose masses are very small and have not yet been directly measured). The top quark is nearly five orders of magnitude more massive than the up and down quarks, while the tau is over three orders of magnitude more massive than the electron [1]. These higher mass particles are generally unstable and decay into first-generation particles. There is, so far, no confirmed explanation for this pattern.

The SM also contains force carrying particles, which are all vector bosons. The particle that carries the electromagnetic force is the massless photon. The weak force is transmitted by massive particles called the $W^{\pm}$ and Z (which have about 80 and 90 times the mass of the proton). The strong force is transmitted by

Table 1.1: The matter particles of the SM

		First	Second	Third
Quarks	Up-type	u(p)	c(harm)	t(op)
	Down-type	d(own)	s(trange)	b(ottom)
Leptons	Massive	e	μ	τ
	Neutrinos	ν_e	ν_μ	ν_τ

massless particles called gluons. All these forces are described by non-abelian local gauge theories [2]. The gauge structure of the SM is $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $SU(3)_C$ component is responsible for the strong force, and its gauge theory is known as quantum chromodynamics (QCD). The $SU(2)_L \times U(1)_Y$ component for the electroweak force. The SM contains a scalar field called the Higgs that is responsible for breaking the electroweak gauge group $SU(2)_L \times U(1)_Y$ down to the electromagnetic abelian subgroup $U(1)_\gamma$ through a mechanism called spontaneous symmetry breaking [3]; the gauge theory of this abelian symmetry is known as quantum electrodynamics (QED). Symmetry breaking simultaneously generates the masses of all the elementary particles, including, in particular, the masses of the $W^\pm$ and Z bosons, which is responsible for the eponymous weakness of the weak force. The Higgs particle was the last particle of the SM to be discovered [4,5].

1.2 The electroweak force and CP violation

Historically, the theoretical development of the electroweak force tracked experimental findings regarding radioactive decay, flavor changing currents, and parity and charge–parity (CP) violation, in a complicated fashion. Parity violation was first discovered in the beta decay of cobalt-60, after predictions made by Lee and Yang [1] motivated by observations of kaon decays. In the beta decay process, a neutron changes to a proton via the process $d \to u\, e^- \,\bar{\nu}_e$. Eventually it was learned that this process was due to the coupling of a charged vector boson W^+ to the u, d and

$e^-, \bar{\nu}_e$ pairs. The parity violation observed in cobalt-60 beta decay was complete — the process involved only the left-handed components of fermions and right-handed components of antifermions [1]. In later experiments, such as the detection of certain neutrino scatterings in 1973 and forwards-backwards asymmetry in e^+e^- collisions, evidence was found for the neutral Z boson. In contrast to processes involving $W^\pm$, the parity violation arising from Z interactions was found to be only partial. A theoretical explanation for these effects, termed the weak force, was eventually found in a chiral $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge theory, which also encompasses electromagnetism.

1.2.1 The $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ electroweak force

Before spontaneous symmetry breaking, the coupling terms in the SM Lagrangian of fermions ψ to the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ electroweak force are

$$-g\left(\bar{\psi}\gamma^\mu T^i \psi\right) W^i_\mu - g'\left(\bar{\psi}\gamma^\mu Y \psi\right) B_\mu \tag{1.1}$$

where $T^i = \tau^i/2$ are the generators of the two-dimensional representation of $\mathrm{SU}(2)$. The operators T^i and Y act on the fields, and their eigenvectors and corresponding eigenvalues are shown in Table 1.2. Notably, the eigenvectors are not simply the SM particles, but are instead the left-handed and right-handed chiral components of the particles. The pairs (u_L, d_L) and (ν_{eL}, e_L) transform as doublets, while the right-handed chiral states transform as singlets (the same pattern and eigenvalues hold in the second and third generations of quarks and leptons). The handedness of these couplings accounts for the experimental observation of parity violation.

In this gauge theory, the parity violation from each of the the charged $W^\pm_\mu = W^1_\mu \mp W^2_\mu$ and neutral W^3_μ bosons is maximal. This contrasts with the experimental observations of maximal parity violation in processes involving weak charged currents but partial parity violation in ones involving weak neutral currents. This

discrepancy is explained by spontaneous symmetry breaking, after which the massless photon field A_μ and massive Z boson field Z_μ, eigenvectors of the mass matrix M_{W^3B}, are orthogonal linear combinations of W^3_μ and B_μ. The SO(2) rotation

$$\begin{pmatrix} Z \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix}$$

replaces the kinetic terms in the Lagrangian for Z and A with covariant ones for B and W^3. Defining $\tan\theta_W = g'/g$, called the Weinberg mixing angle, the coupling terms after the rotation become

$$- eJ^\mu_\gamma A_\mu - g_Z J^\mu_Z Z_\mu - \frac{g}{\sqrt{2}} \left\{ J^{\dagger\mu} W^+_\mu + J^\mu W^-_\mu \right\} \tag{1.2}$$

where $g_Z = \sqrt{g^2 + g'^2}$ and $e = gg'/\sqrt{g^2 + g'^2}$. The terms J^μ_γ, J^μ_Z, and J^μ are the electromagnetic current, the weak neutral current, and the weak charged current, given by

$$J^\mu_\gamma = \bar{\psi}\gamma^\mu Q \psi \tag{1.3}$$

$$J^\mu_Z = -\frac{1}{\sin^2\theta_W}\bar{\psi}\gamma^\mu Q \psi + \bar{\psi}\gamma^\mu T^3 P_L \psi \tag{1.4}$$

$$J^\mu = \Psi^\dagger \bar{\sigma}^\mu \tau^- \Psi \tag{1.5}$$

$$J^{\dagger\mu} = \Psi^\dagger \bar{\sigma}^\mu \tau^+ \Psi \tag{1.6}$$

where $Q = T^3 + Y$ is the electric charge operator (Table 1.2). The weak charged current is written in Weyl notation, where Ψ is a doublet wavefunction that stands for (u_L, d_L) or (ν_{eL}, e_L), and $\tau^\pm = (\tau^1 \pm i\tau^2)/2$ are the 2×2 projection matrices with a single non-zero entry 1 in the upper right $(+)$ or lower left $(-)$ corner. The coupling of A_μ to J^μ_γ reproduces QED, while Z_μ-to-J^μ_Z produces a partial parity violating effect. Because the Z is massive while the photon γ is massless, the QED

5

Table 1.2: Electroweak eigenvalues for first-generation fermions [6].

	T	T^3	Y	Q		T	T^3	Y	Q
u_L	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{2}{3}$	ν_{eL}	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	0
d_L	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{6}$	$-\frac{1}{3}$	e_L	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	-1
u_R	0	0	$\frac{2}{3}$	$\frac{2}{3}$	e_R	0	0	-1	-1
d_R	0	0	$-\frac{1}{3}$	$-\frac{1}{3}$					

Table 1.3: C, P, and CP transformations of operators appearing in the Lagrangian [6, 7]. Here $\psi^c_{L,R} = \mp i\sigma^2 \psi^*_{R,L}$ and $(-1)^\mu = 1$ for $\mu = 0$ and -1 for $\mu = 1, 2, 3$. Both operations also transform four-vectors as $x = (x^0, \vec{x}) \to x' = (x^0, -\vec{x})$. This table ignores phases that cancel in the Lagrangian.

Object	C	P	CP
$\psi_{L,R}$	$\psi^c_{L,R}$	$\psi_{R,L}$	$\psi^c_{R,L}$
$W^{\pm\mu}$	$W^{\mp\mu}$	$-(-1)^\mu W^{\pm\mu}$	$-(-1)^\mu W^{\mp\mu}$
A^μ	$A^{\dagger\mu}$	$-(-1)^\mu A^\mu$	$-(-1)^\mu A^{\dagger\mu}$

force is long-range and in general much stronger, while the weak neutral force is suppressed at low energies by a factor $1/M_Z^2$ and the charged weak force by a factor $1/M_W^2$.

1.2.2 CP symmetry

The simple one-generation electroweak force described in the previous paragraphs violates parity, but is invariant under the combined charge-parity, or CP, operation. Recall that the W^i bosons couple left-handed particles and right-handed antiparticles. The parity operator P transforms left-handed particles to right-handed particles. Since the latter do not couple to $W^\pm$, P is not a symmetry of the weak force. Subsequently performing the charge conjugation operator C, however, transforms these right-handed particles to right-handed antiparticles. Thus, the combination of these two operators is a symmetry of the weak force. The explicit transformations caused by C, P, and CP are given in Table 1.3.

For now, consider only the first and lightest generation of quarks and leptons,

ignoring their heavier cousins. Since CP takes $\psi_{L,R} \to \psi^c_{R,L}$, the left-handed weak charged current for $\Psi = (\psi_{1L}, \psi_{2L})$ transforms into

$$J^\mu = \psi^\dagger_{2L} \bar\sigma^\mu \psi_{1L} \xrightarrow[CP]{} \psi^{c\dagger}_{2R} \bar\sigma^\mu \psi^c_{1R} = -\psi^\dagger_{1L} \sigma^\mu \psi_{2L} = -(-1)^\mu J^{\dagger\mu}, \qquad (1.7)$$

where the penultimate equality is due to an identity derivable from the Fierz identities [6]. Similarly, $J^{\dagger\mu}$ is transformed to $-(-1)^\mu J^\mu$ by the CP operator. Simultaneously, the CP transformation also exchanges the charged bosons W^+ and W^- and flips their parity ($W^{\pm\mu} \to -(-1)^\mu W^{\mp\mu}$), so the coupling terms simply exchange:

$$J^{\dagger\mu} W^+_\mu \xleftrightarrow[CP]{} J^\mu W^-_\mu \qquad (1.8)$$

Meanwhile, the weak neutral current (1.4) and electromagnetic current (1.3) are both Hermitian, $i.e.$, $J^{\dagger\mu}_Z = J^\mu_Z$ and $J^{\dagger\mu}_\gamma = J^\mu_\gamma$. Consequently $J^\mu_Z \to -(-1)^\mu J^\mu_Z$ and $J^\mu_\gamma \to -(-1)^\mu J^\mu_\gamma$ under CP. Since the neutral vector bosons transform as $Z^\mu \to -(-1)^\mu Z^\mu$ and $A^\mu \to -(-1)^\mu A^\mu$, the weak neutral and electromagnetic coupling terms are unchanged under CP:

$$J^\mu_Z Z_\mu \xrightarrow[CP]{} J^\mu_Z Z_\mu, \ J^\mu_\gamma A_\mu \xrightarrow[CP]{} J^\mu_\gamma A_\mu \qquad (1.9)$$

The entire electroweak Lagrangian as presented so far is invariant under the CP operation. The neutral coupling terms (involving the γ and Z) are unchanged while the charged terms trade places.

1.2.3 Quark mixing and CP violation

The simplest incorporation of the two heavier generations of quarks and leptons into the theory would be achieved by triplicating all the first-generation terms

in the Lagrangian. The weak charged current would then be

$$J^{\dagger\mu} = \sum_{i=1}^{3} u_i^{\dagger} \bar{\sigma}^{\mu} d_i \qquad (1.10)$$

where u_1, u_2, and u_3 represent the left-handed up-type quarks (up u_L, charm c_L, top t_L); similarly d_1, d_2, and d_3 represent the left-handed down-type quarks (down d_L, strange s_L, and bottom b_L). Conservation laws would then imply the existence of a host of stable mesons and hadrons containing strange and bottom quarks. However, all matter containing second or third generation quarks decays quickly. The semileptonic decay of charged kaons, for example, is mediated by the process $s \to u\, e^- \, \bar{\nu}_e$, in which the s quark takes the place of a d quark in a flavor changing charged current (FCCC).

This quark mixing can be explained if the three eigenstates of the down-type quarks that couple to the electroweak force are slightly skewed from the three mass eigenstates. In principle, the weak eigenstates of the up-type quarks could be skewed as well, but a simultaneous unitary rotation of both the up-type and down-type families leaves all the weak coupling terms invariant, so there is always freedom to redefine the quark fields so that all mixing occurs in the down-type family. The mass eigenstate s would then not be completely orthogonal to the weak eigenstate d', allowing for a FCCC as in charged kaon decays. Cabibbo used this basic idea to explain strangeness-violating decays before it was known that there were three generations of quarks. In this model, the weak eigenstates d', s' are related to the mass eigenstates d, s by

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \qquad (1.11)$$

where θ_c is the Cabibbo angle and experimentally $\sin\theta_c \sim 0.23$ [6]. This mechanism

introduces a factor of $\sin \theta_c$ into the amplitude for the strangeness-violating decay $s \to u\,e^- \bar{\nu}_e$, so the probability of this decay is suppressed relative to $d \to u\,e^- \bar{\nu}_e$ by a factor of $1/\sin^2 \theta_c \sim 20$.

Denoting the down-type quark mixing matrix by V_{ij}, the skewed weak charged current becomes

$$J^{\dagger\mu} = \sum_{i,j} u_i^\dagger \bar{\sigma}^\mu V_{ij} d_j \tag{1.12}$$

Meanwhile, the electromagnetic and weak neutral currents are unchanged, since they do not couple up-type quarks to down-type quarks; because of the unitarity of V_{ij}, the down-down couplings are

$$J^\mu_{Z,\gamma} = \sum_{i,j,k} d_i^\dagger V_{ij}^\dagger \bar{\sigma}^\mu V_{jk} d_k = \sum_i d_i^\dagger \bar{\sigma}^\mu d_j. \tag{1.13}$$

The crucial consequence is that the weak neutral current is entirely flavor diagonal and cannot be responsible for tree-level flavor changing neutral currents (FCNCs). As an example, a decay of a neutral kaon $K^0 = |\bar{s}d\rangle$ like $K^0 \to \mu^+\mu^-$ or $K^0 \to \pi\pi(\pi)$ changes flavor ($\Delta S = 1$). If the Z boson could couple to $\bar{s}d$, the branching fraction of the decay $K^0 \to \mu^+\mu^-$ would be enhanced beyond experimental bounds (Fig. 1.1a). Instead, this FCNC must be mediated by loop-level box diagrams (Fig. 1.1b) with a much-suppressed amplitude. Furthermore, even this suppressed amplitude for $K^0 \to \mu^+\mu^-$ appeared to large originally, leading to the proposed existence of the c, which would resolve the discrepancy via the GIM mechanism.[1]

This quark mixing is a change of basis in the down-type family, and to preserve the normalization of the eigenstates it must be unitary. So, with N generations of quarks, it is accomplished by an $N \times N$ unitary matrix, called the Cabibbo-

---

[1]The amplitude of each box diagram is proportional to the mass of the quark in the loop. Still, the remnant FCNC is too large if the only quarks in the theory are the u, d, and s, because only the u can appear in the box loop due to charge conservation. The existence of the charm quark was proposed by Giorgi, Iliopoulos, and Maiani because it introduces a box diagram with opposite sign, partially cancelling the u amplitude [1,6]. This cancellation is referred to as the GIM mechanism after its creators.

(a) Hypothetical (vanishing) tree level FCNC

(b) Box diagram FCNC
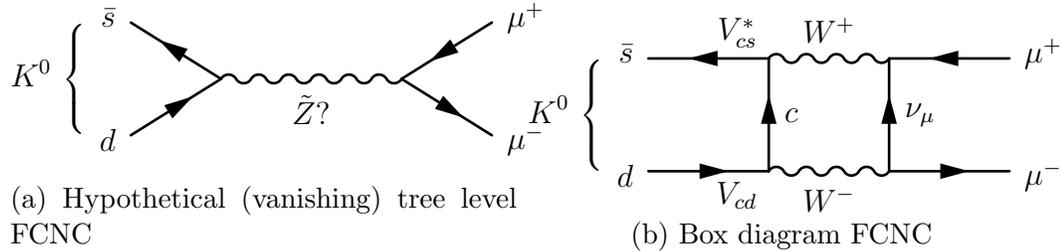
Figure 1.1: FCNCs in K^0 decay

Kobayashi-Maskawa (CKM) matrix V_{CKM} [8] when $N = 3$. In general, an $N \times N$ unitary matrix has N^2 real parameters, but some of these do not represent physical degrees of freedom. Without skewing flavor states, there is freedom to redefine the $2N$ quark flavors by $2N - 1$ relative phase differences. Eliminating these artificial degrees of freedom that can be absorbed by quark fields, just $(N-1)^2$ real parameters in left the CKM matrix [7]. For $N = 2$, this is just one — the Cabibbo angle θ_c. For $N = 3$, there four irreducible real parameters. The 3×3 CKM matrix is generally labelled

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.14}$$

There are various parameterizations that relate these coefficients to the four independent paramete.lkprs, *e.g.*, the Wolfenstein parametrization (see [6]). As shown by Kobayashi and Maskawa, the matrix can be parametrized by three real constants and one complex phase [8]. The weak charged current can be written as a sum over quark flavor pairs:

$$J^{\dagger\mu} = \sum_{\substack{i \in \{u,c,t\} \\ j \in \{d,s,b\}}} u_i^\dagger \bar{\sigma}^\mu V_{ij} d_j$$

Eq. (1.7) now only holds only if $V_{ij} = V_{ij}^*$. Only in that case does the CP transformation simply exchange $J^{\dagger\mu} W_\mu^+$ and $J^\mu W_\mu^-$. This was the case for the two-generation Cabibbo matrix. Unless the phase parameter is exactly zero, this is not the case

$$K^0 \left\{ \begin{array}{c} \bar{s} \xrightarrow{\quad V_{cs}^* \quad W^+ \quad} \bar{d} \\ c \qquad c \\ d \xrightarrow{\quad V_{cd} \quad W^- \quad} s \end{array} \right\} \bar{K}^0$$
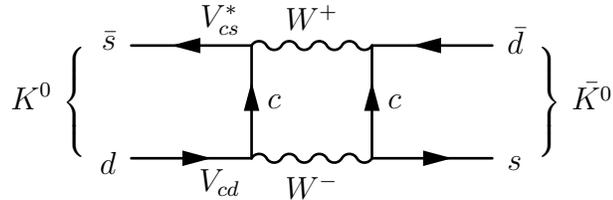
Figure 1.2: K^0–$\bar{K}^0$ mixing receives contributions from internal loops with the three up-type quarks ($q = u, c, t$) and shows CP violation due to the phase differences between $V_{qs}V_{qs}{}^*$ and $V_{qs}{}^*V_{qs}$.

for the three-generation CKM matrix. Each charge-raising vertex in a Feynman diagram for some process $M \to f$ is associated with a V_{ab} coefficient, while each charge lowering current comes with a V_{ab}^* coefficient. The amplitude has an overall factor that can be complex, and the argument of this factor is the weak phase of the diagram. In the CP conjugate Feynman diagram representing the process $\overline{M} \to \overline{f}$, the charge-raising vertices become charge-lowering vertices, and vice versa. Thus, the weak phase of $\overline{M} \to \overline{f}$ is the opposite of the weak phase of $M \to f$, and CP violation is introduced to the Standard Model.

This alone does not cause CP violation — though the amplitudes of the CP-conjugate processes differ by a phase, their magnitudes (and hence probabilities) are equal. Instead, there must be interference between different possible diagrams for a process $A \to B$ to have CP violation. Chapter 5 discusses several different mechanisms by which such interference might appear in B meson decays. In general, experimentally measured asymmetries in CP conjugate processes can be used to determine the complex coefficients of the CKM matrix.

1.2.4 CP violation in the neutral kaon system

The discovery of CP violation in decays of neutral kaons in 1964 [9] predated the proposal of the CKM matrix in 1973. Since the CKM mechanism requires three generations to produce CP violation, this discovery partly motivated the prediction of the bottom and top quarks, eventually discovered in 1977 and 1995 [10, 11]. The

study of neutral mesons has historically been very fruitful, and research continues today with the study of neutral D^0 mesons, B^0 mesons, and B_s^0 mesons.

A neutral kaon is a bound state $K^0 = |\bar{s}d\rangle$, and its antiparticle is $\overline{K}^0 = |s\bar{d}\rangle$. Both are uncharged under all SM forces and are free to mix due to couplings in the effective Hamiltonian generated by weak interaction diagrams such as in Fig. 1.2. Without CP violation, the mass eigenstates would be the orthogonal combinations

$$|K_{1,2}\rangle = \frac{1}{\sqrt{2}} \left(|K^0\rangle \mp |\overline{K}^0\rangle \right). \tag{1.15}$$

These were originally identified with the short-lived K_S^0 state and long-lived K_L^0 state that are observed experimentally. "Short" and "long" here are relative terms because both particles are quite long-lived when compared to other particles, with lifetimes of 90 ps and 51 ns respectively [12]. Assuming no CP violation, the two states would be CP eigenstates, with corresponding eigenvalues 1 for the K_S^0 state and -1 for the K_L^0 state.

Two final states that the K_S^0 and K_L^0 could potentially decay into are $\pi\pi$ (or 2π) and $\pi\pi\pi^0$ (or 3π), where $\pi\pi$ could be $\pi^+\pi^-$ or $\pi^0\pi^0$. The 2π state is a CP eigenstate with eigenvalue 1, while the 3π state has eigenvalue -1 (except when the pion system has internal angular momentum, in which case the decay is suppressed). According to CP invariance, the K_S^0 should decay predominantly to 2π (and occasionally to P-wave 3π) while the K_L^0 should decay exclusively to 3π. Since the K_L^0 is much longer-lived than the K_S^0, a pure K_L^0 beam can be created out of a mixed neutral kaon beam (produced by scattering pions on a beryllium target) by letting the beam propagate some sufficient distance — after 300 decay lengths of the K_S^0, only about 40% of the K_L^0 mesons in the beam will have decayed. In their seminal experiment, Christenson, Cronin, Fitch, and Turlay found unambiguous evidence of the CP violating decay $K_L^0 \to 2\pi$ [9].

This unexpected discovery inspired the development of the CKM mechanism and the proposal of a third generation of quarks. The phenomenology of CP violation in the neutral kaon system has been studied in great detail. The real K_S^0 and K_L^0 are not simply the K_1 and K_2 states but instead [6]

$$|K_{S,L}\rangle = \frac{|K_{1,2}\rangle + \tilde{\epsilon}|K_{2,1}\rangle}{\sqrt{1 + |\tilde{\epsilon}|^2}} \tag{1.16}$$

where $\tilde{\epsilon}$ is a parameter measuring CP violation that occurs in the mixing of neutral kaons. The source of this CP violation in mixing comes from complex phases in the box diagrams responsible for K^0–$\overline{K}^0$ mixing (Fig. 1.2). Now K_L^0 is allowed to decay to 2π because it contains a small K_1 admixture. This is the dominant cause of CP violation in neutral kaons; however, a smaller source of direct CP violation arises from interference between tree-level and loop-level decay amplitudes. These manifestations of CP violation — indirect (via mixing), direct (via decay), and via the interference between mixing and decay — define the broad types of CP violation for all neutral meson systems, including D^0–$\overline{D}^0$, B^0–$\overline{B}^0$, and B_s^0–$\overline{B}_s^0$. They are described in more detail in Chapter 5.

1.3   Lepton flavor universality

In the Standard Model, the couplings of all flavors of leptons to the gauge bosons are identical, with the three distinguished only by their masses (*i.e.*, by their couplings to the Higgs boson). Due to lepton universality, differences between decays with electrons, muons, or taus in the decay final state arise only because of the different masses of the three leptons. The effects of these masses are often considered to be reliably calculable, allowing experimental tests of lepton flavor universality. Many studies have looked for signs of lepton flavor non-universality, and several have turned up possible signs of tension. Most relevant to this thesis

Table 1.4: Measurements of semitauonic $B \to D^{(*)}\tau\nu$ branching fractions. The Belle measurement marked *sl* is performed with semileptonic tagging, and the other with hadronic tagging. The measurements marked † are performed using hadronic decay modes of the τ.

Experiment		$\mathcal{R}(D)$	$\mathcal{R}(D^*)$
BABAR	[16]	$0.440 \pm 0.058 \pm 0.042$	$0.332 \pm 0.024 \pm 0.018$
Belle	[17]	$0.375 \pm 0.064 \pm 0.026$	$0.293 \pm 0.038 \pm 0.015$
Belle^{sl}	[18]	–	$0.302 \pm 0.030 \pm 0.011$
Belle†	[19]	–	$0.270 \pm 0.035^{+0.028}_{-0.025}$
LHCb	[20]	–	$0.336 \pm 0.027 \pm 0.030$
LHCb†	[21]	––	$0.285 \pm 0.019 \pm 0.029$
Average	[13]	$0.407 \pm 0.039 \pm 0.024$	$0.304 \pm 0.013 \pm 0.007$
SM	[22, 23]	0.300 ± 0.008	0.252 ± 0.003

are the observables $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$, defined by

$$\mathcal{R}(D^{(*)}) = \frac{\mathcal{B}(B^{(0,-)} \to D^{(*-)}\tau^+\bar{\nu}_\tau)}{\mathcal{B}(B^{(0,-)}) \to D^{(*-)}\mu^+\bar{\nu}_\mu)}, \tag{1.17}$$

which have been measured by the BaBar, LHCb, and Belle collaborations. Accounting for their correlations, the combined measurements of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ differ from the SM predictions by 4.1σ according to the HFLAV group [13]. These measurements and predictions are shown in Table 1.4, and a graphical representation is shown in Fig. 1.3. Other recent LHCb measurements with hints of new flavor physics include comparisons of the branching fractions of the decays $B \to K^{(*)}\mu^+\mu^-$ and $B \to K^{(*)}e^+e^-$ [14, 15].

Many models of beyond the standard model (BSM) physics produce explicit lepton flavor non-universality. Extra Higgs doublets, present in BSM theories such as supersymmetry, increase the number of processes that involve the lepton Yukawa couplings and can lead to lepton flavor non-universality. Couplings to a charged Higgs boson H^+ have been discussed as a possible cause of deviations in $\mathcal{R}(D^{(*)})$.

A popular suggestion for sources of lepton flavor non-universality are leptoquarks, a moniker for any hypothetical bosons compatible with the SM gauge
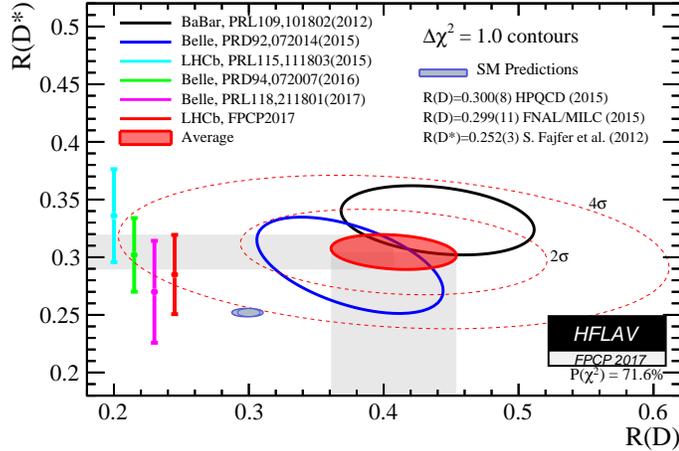
Figure 1.3: Measurements of semitauonic $B \to D^{(*)}\tau\nu$ branching fractions, reproduced from Ref. [13].

symmetries that carry both baryon and lepton number [2, 12]. Many BSM theories, in particular supersymmetry and grand unified theories, generate different variety of leptoquarks: scalar or vector, color singlet or color triplet, and weak singlet, doublet, or even triplet. Irrespective of their origin, leptoquarks would couple directly to lepton–antiquark anti-lepton–quark pairs, with different couplings for the three generations of quarks and leptons. If the leptoquarks couple directly to the quark mass eigenstates, no lepton flavor non-universality is created. However, if the couplings were to a different eigenbasis of quarks, in the same manner as the weak couplings, lepton flavor non-universality would be generated indirectly by leptoquark mediated four-fermion couplings. Leptoquark models that explain recent flavor anomalies without violating other experimental bounds are possible [24].

Another class of models turn $L_\mu - L_\tau$ into a gauge symmetry, directly violating universality [25, 26]. In the SM, the conserved quantities $L_e - L_\mu$, $L_e - L_\tau$, and $L_\mu - L_\tau$ correspond to anomaly-free global U(1) symmetries. It is possible to promote one of these three to a local symmetry mediated by a heavy neutral boson Z'. [2] Of the

---

[2]In fact it is a common philosophical position (originally promoted by Yang and Mills) that all possible anomaly-free symmetries should be local; one impetus for this belief is that the promotion of all possible U(1) anomaly-free global symmetries leads to hypercharge (and thus electric charge)

three possible choices, $L_\mu - L_\tau$ has the least stringent experimental bounds. The new Z' gauge boson would acquire mass through the spontaneous symmetry breaking of $U(1)_{L_\mu - L_\tau}$, via a new electroweak singlet Higgs field carrying an $L_\mu - L_\tau$ quantum number (preventing Z–Z' mixing). The breaking of this symmetry does not break the global symmetries responsible for the separate conservation of the lepton family numbers L_e, L_μ, and L_τ. Fermions of the SM would couple to the Z' with a charge $L_\mu - L_\tau$ (*i.e.*, the new particle would only couple μ, ν_μ, τ, and ν_τ). Each vertex conserves charge and the lepton family numbers (*e.g.*, it can only couple to $\mu^+\mu^-$ or $\tau^+\tau^-$). However, the new couplings available to μ and τ but not to e destroy lepton flavor universality.

---

quantization, which is accidental within the SM [25].

Chapter 2:   The LHC and LHCb detector

2.1   The Large Hadron Collider

The LHCb detector is located at the point 8 interaction area of the Large
Hadron Collider (LHC), a proton-proton collider at CERN [27]. CERN, founded
in 1952, is an intergovernmental organization dedicated to nuclear and high en-
ergy physics, located in Meyrin, Switzerland, a suburb of Geneva.[1] At 26.7 km in
circumference, the collider itself straddles the Franco-Swiss border, about a hun-
dred meters underground (Fig. 2.1). The LHC is a roughly circular synchrotron
that circulates two parallel beams of protons in opposite directions. It was built
in the preexisting tunnel dug for the Large Electron Positron (LEP) collider. Pro-
tons enter the LHC with an energy of $450\,\mathrm{GeV}$ from the Super Proton Synchrotron
(SPS), an older synchrotron that now serves as the final injector for the LHC.[2] The
magnetic field that guides the protons around the LHC ring is provided by super-
conducting dipole magnets placed around the LHC ring and cooled to $< 2\,\mathrm{K}$ by
liquid helium. Since the proton beams have the same charge and must be circulated
in opposite directions to collide, the LHC ring contains two separate beampipes in
which the dipole magnets produce oppositely oriented fields. As the protons are
injected, the magnets maintain about a $0.54\,\mathrm{T}$ field. The protons are accelerated

---

[1]CERN was originally an acronym for Conseil Européen pour la Recherche Nucléaire, but its
official long name is now European Organization for Nuclear Research.

[2]Built in 1976, the SPS was originally a particle collider, and the site of the UA1 and UA2
experiments where the first signals of the W and Z bosons were found. Protons are injected into
the SPS at $25\,\mathrm{GeV}$ from the Proton Synchrotron (PS), built in 1959, itself in turn originally a
final-stage collider.
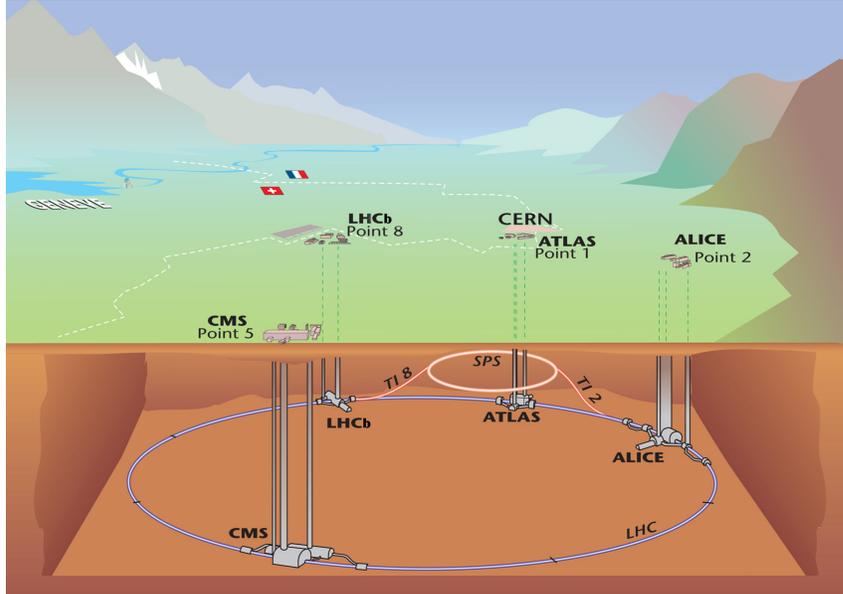
Figure 2.1: The LHC collider [28].

in several straight radio-frequency (RF) cavities, and the dipole magnet fields are correspondingly ramped up. The RF cavities also compensate for losses from synchrotron radiation. In the first two years of operation (2011 and 2012), known as Run I, the LHC dipole magnets operated at $4.1\,$T and $4.7\,$T, producing pp collisions at center-of-mass energies of $7\,$TeV and $8\,$TeV, respectively. After a brief pause for maintenance and upgrades, in 2015 the LHC reached its maximum collision energy, ramping the dipole magnetic fields up to $7.7\,$T and producing $13\,$TeV pp collisions (Run II).

The protons travel around the LHC in a few thousand bunches, separated by $50\,$ns during Run I and by $25\,$ns in the current Run II. The rate at which pp interactions occur is the product of the pp collision cross section and the luminosity of the collisions, which can vary for the different experiments located at the collider. The luminosity is proportional to the product of the number of protons in the two colliding bunches, and to the average collision rate of bunches (equal to the number of bunches per beam times the revolution frequency). The luminosity is also

inversely proportional to the square of the transverse profile (*i.e.*, width) of the beam; focusing the beam is critical to achieve a high luminosity. The LHC has reached a peak luminosity of $1 \times 10^{34} \, \text{cm}^{-2} \, \text{s}^{-1}$ for the CMS and ATLAS experiments and $4 \times 10^{32} \, \text{cm}^{-2} \, \text{s}^{-1}$ for the LHCb experiment. The luminosity at LHCb is intentionally lowered by focusing the beams to a wider transverse profile, and by separating the beams and reducing the area of their overlapping profiles. During the fill, while the number of circulating protons decreases due to collisions with each other and with the accelerator material, the beams at the LHCb interaction point are brought closer together to maintain a roughly constant luminosity.

The LHC's magnetic optics system also includes a large number of quadrupole and sextupole magnets used for focusing and controlling the proton beams. As the beams are injected into the LHC, their transverse profile has a width of a few centimeters. Before accelerating the protons to their maximum energy, the beams must be focused to a transverse profile of a few millimeters. At the interaction points, the beams are further focused to a transverse profile on the order of $50 \, \mu\text{m}$ (for CMS and ATLAS) to reach the target luminosity for physics data-taking. Furthermore, the counter-circulating beams must be precisely guided to intersect at several interaction points where the various detectors are located. In the interaction regions of the experiments, the beams share the same beampipe. To prevent bunches from colliding at locations other than the desired interaction point, they are displaced to the left and right of the beampipe's central axis except at the interaction point.

2.2   The LHCb Detector

The LHCb detector is a single-arm forward spectrometer located around the beampipe at interaction point 8 of the LHC (see Fig. 2.2). The acceptance of the detector covers the angular ranges 10 mrad to 250 mrad in the vertical direction and 10 mrad to 300 mrad in the horizontal direction, corresponding roughly to pseudo-
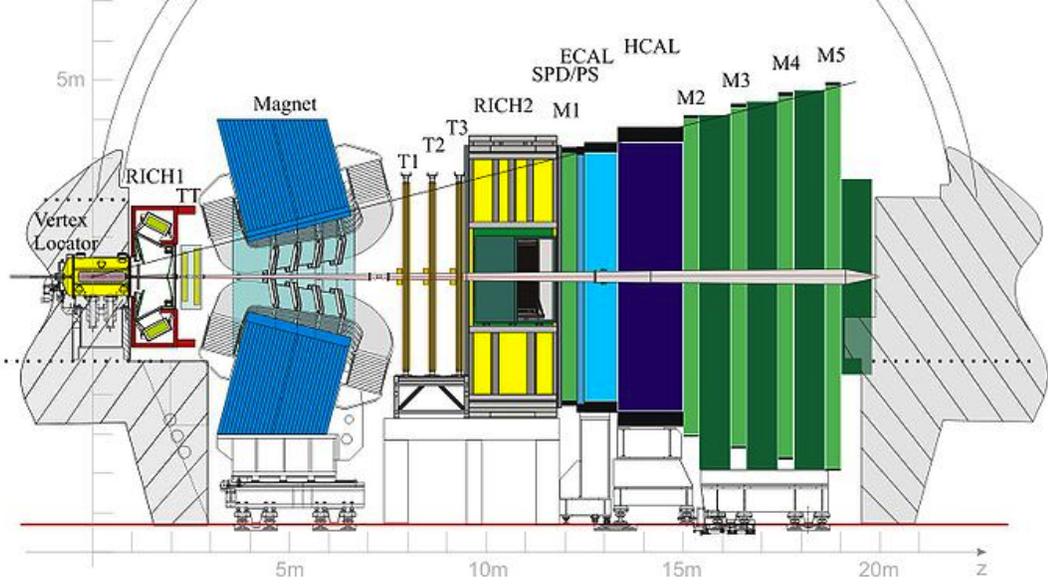
Figure 2.2: The LHCb detector [29]. In the center lies the dipole bending magnet (blue). Upstream (decreasing z) lie the Vertex Locator (VELO) and TT tracking stations, and the RICH1 station. Downstream (increasing z) lie the IT and OT trackers (in three layers labelled T1, T2, and T3), RICH2 station, ECAL, HCAL, and muon chamber.

rapidities in the range $1.6 < \eta < 4.9$. Though not a general purpose detector like its CMS and ATLAS counterparts, the LHCb detector operates during p–p, p–Pb, and Pb–Pb collisions and the data collected is used for several categories of physics studies. The primary focus of the LHCb collaboration is flavor physics: the study of the decays of b-hadrons and c-hadrons. The collaboration also conducts research in soft QCD and electroweak physics, for which the forward acceptance of LHCb is complementary to the acceptances of CMS and ATLAS. Finally, the collaboration has an active group studying heavy-ion collisions.

A complete overview of the entire LHCb detector is found in Ref. [30]. Presented here is a brief overview of the physics of $b\bar{b}$ production at the LHC and the various components of the detector. LHCb is primarily a tracking detector, with stations upstream and downstream of a dipole bending magnet. Behind these last tracking stations are an electromagnetic and a hadronic calorimeter, and finally

a muon detector, which figure prominently in LHCb's trigger system. For particle identification, particularly important for flavor physics, two ring-imaging Cherenkov (RICH) detectors (one upstream and one downstream of the magnet) help distinguish charged pions, kaons, and protons. The calorimeters also help distinguish electrons from other charged tracks and photons from neutral pions; the muon detector separates muons from hadrons. These components are shown in Fig. 2.2. The LHC beampipe is located along the central axis of the detector. The conventional coordinate system used to describe the geometry of the detector calls the direction of the beam z, while the vertical direction is y and the horizontal direction (perpendicular to the cross section in the figure) is x. Particular attention will be paid to how these components contribute to the analysis of the decay $B_c^+ \to J/\psi \tau^+ \nu_\tau$ and to flavor tagging and CP violation measurements.

2.2.1 $b\bar{b}$ production at a proton-proton collider

The production of b-hadrons in proton–proton collisions at the LHC (or, previously, in proton–antiproton collisions at the Tevatron), via the production of $b\bar{b}$ pairs, is complex. The momentum and energy of the proton are carried not just by the valence quarks but also by the sea of gluons and $q\bar{q}$ pairs that bind the valence quarks together. Each momentum-carrying element of a proton is termed a parton, and pp collision is typically a scattering between two partons (double parton interaction occurs more rarely).

Several tree level diagrams for $b\bar{b}$ production are shown in Fig. 2.3; these are heavily modified by soft QCD effects. For $\sqrt{s} = 7\,\mathrm{TeV}$ collisions, the total $\sigma(pp \to b\bar{b}X)$ cross section at the LHC is about $250\,\mathrm{\mu b}$, and approximately doubles at $13\,\mathrm{TeV}$ collisions. The prompt charm cross section $\sigma(pp \to c\bar{c}X)$ is about an order of magnitude larger, around $1.5\,\mathrm{mb}$ at $7\,\mathrm{TeV}$, and again about double that at

(a) $q\bar{q}$ fusion   (b) s-channel gg fusion   (c) t-channel gg fusion
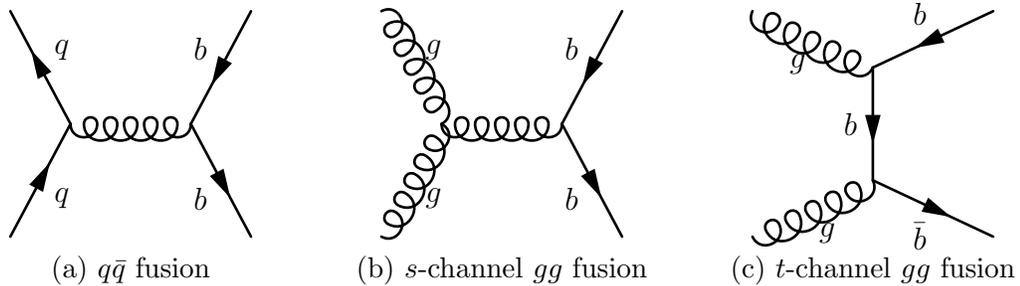
Figure 2.3: Tree-level $b\bar{b}$ pair production in pp collisions

13 TeV.[3] In comparison, the total inelastic cross section $\sigma(pp \to X)$ is approximately 60 mb at 7 TeV and [32, 33] and 75 mb at 13 TeV [34, 35].

The $b\bar{b}$ pairs produced at the LHC are concentrated in the region of high pseudorapidity [30]. Despite covering just 4% of the solid angle around the interaction point, LHCb's acceptance includes about 25% of all $b\bar{b}$ pairs produced. The LHCb collaboration has measured $b\bar{b}$ and prompt $c\bar{c}$ cross sections inside LHCb's acceptance of $(75 \pm 14)\,\mu b$ and $(1.4 \pm 0.1)\,mb$ at 7 TeV [36, 37]. Recent measurements show that both cross sections have approximately doubled at 13 TeV, to $(154 \pm 14)\,\mu b$ and 2.8 mb, respectively [38, 39].

2.2.2 Tracking system

The LHCb tracking system measures the paths and momenta of charged tracks emanating from the pp interaction point. A series of silicon and straw tube trackers detect the transverse positions of tracks at several locations along the beam axis. To enable measurement of the tracks' momenta, the detector features a non-superconducting dipole magnet with a bending power of 4 T m to 5 T m [30]. Momenta of particle tracks are calculated by measuring their deflection within the field. The geometry of the magnet is unique: two saddle-shaped aluminum coils are yoked inside an iron return, and face each other across a gap that serves as a window

_____

[3]These approximate calculations were made using the `MadGraph` simulation software [31], requiring the b and c quarks to have a $p_T > 1\,GeV$ and $\eta < 5$.

for the passage of the charged tracks downstream. In the cross section shown in Fig. 2.2, the aluminum coils are colored silver, and the iron return blue. During data taking, the polarity of the magnet can be flipped, exchanging the way positively and negatively charged tracks bend in the field. This allows the measurement of differences in the acceptance efficiencies for oppositely-charged tracks that are due to the geometry of the detector, an important control factor in measurements of CP violation.

Tracking technologies in LHCb

There are two basic tracking detector technologies used at LHCb: silicon microstrip detectors and several variants of proportional counters. A silicon microstrip detector is created from a silicon wafer doped in a p^+-on-n or n^+-on-p configuration, forming a semiconductor junction. A high reverse bias voltage, up to $500\,\mathrm{V}$, is applied across the wafer, accentuating the depletion region inside the junction and reducing leaking current. A high-energy charged particle passing through the silicon wafer liberates a large number of electron-hole pairs; the typical ionization energy required to create a single pair is just $3\,\mathrm{eV}$ [40]. In the influence of the high electric field, these electrons and holes travel to the opposite collection surfaces of the wafer. The collection time for a silicon sensor is typically on the order of $10\,\mathrm{ns}$. The silicon sensor is etched with microstrips of a given width, or pitch, typically on the order of $50\,\mu\mathrm{m}$ to $200\,\mu\mathrm{m}$. These are electrically insulated from each other and are read out separately. This provides a measurement of one coordinate of the hit's position. The baseline resolution of such a measurement is equal to the pitch of the sensor divided by $\sqrt{12}$. Charged particles often deposit energy into several adjacent strips; examining this charge-sharing can improve the baseline resolution. In short, silicon microstrip detectors take up a small volume in the detector (a benefit but also a limitation in terms of coverage) and have excellent position precision.

In a proportional counter, the role played by the solid silicon bulk in a silicon detector is played instead by a gas, and role of the electrons and holes are given to free electrons and ions. A charged particle passing through gas ionizes the gas molecules to produce electron-ion pairs. Because of the low density of the gas, and because the required ionization energy is typically $30\,\mathrm{eV}$ [40], the number of free electrons can be quite small. In the presence of a large electric field, as high as $1 \times 10^6\,\mathrm{V\,m^{-1}}$, however, these free electrons accelerate to high enough energies to ionize further gas molecules, creating an avalanche of charge carriers. With a prudent choice of gas and electric field strength, the signal created by the avalanche is amplified, but remains proportional to the original signal. The potential difference between an anode wire and a closely-located cathode creates a large enough electric field to produce an avalanche, while localizing the avalanche near the anode.[4] When instrumented and operated as a drift tube, the spatial resolution of the tube is determined by its measurement of the drift time — the elapsed time between the initial ionization and the arrival of the signal at the anode. This timing measurement is made in concert with a measurement of the expected arrival time of a charged particle at the tube, and with appropriate calibration can provide a measurement of the drift distance with good resolution.

This drift time varies considerably depending on the geometry of the counter, but is typically much longer than for a silicon sensor. Because they are lighter and cheaper, proportional counters can cover a larger volume than silicon detectors, but with poorer spatial resolution.

Several variants of proportional counters are used at LHCb: straw tube drift chambers, multi-wire proportional counters, and gas electron multipliers. In straw tube drift chambers, each anode wire is surrounded by its own cathode tube, created by winding a thin strip of foil like a tennis racket grip [40]. These straw tubes are

---

[4]The electric field created by this arrangement is typically inversely proportional to the distance from the anode.

then grouped into closely-packed arrays.

In a multi-wire proportional chamber (MWPC), an array or arrays of anode wires are sandwiched in a plane between two cathode plates. With no walls separating the separate anodes, the ionization from a charged track produces a signal on several anodes, with the largest signal at the wire closest to the track. The achievable spatial resolution is on the order of the distance between neighboring wires, usually a few millimeters. A MWPC is typically used in settings where superior resolution is not needed, such as in a muon detector where the path of the tracks is already smeared due to multiple scattering. Because it is constructed from less material and is cheaper, a MWPC can cover yet more volume than a straw tube drift chamber.

A gas electron multiplier (GEM) is sometimes used in place of a MWPC when the event rate is very high [40]. A GEM foil consists of two parallel conducting plates on either side of an insulator, with a lattice of holes through all three layers. Typically the thickness of this foil and the diameter of the holes are both on the order of $100\,\mu\text{m}$. When a high voltage is applied across the two plates, a very large electric field can be created inside the holes. A number of these foils (typically three, known as a triple-GEM arrangement) are placed in a sequence along the z axis inside a gas chamber with a smaller potential difference between the opposite ends. Free electrons created on the left end by ionization drift in the z direction to the right, and are squeezed through the high-field holes in the first GEM foil, producing charge avalanches. This process repeats at the remaining GEM foils, leading to large gains. Rather than grouping the charges together at the locations of the holes, the left-right symmetry of the electric fields ensures that after passing through the holes, the charges return to roughly their original xy position. The charge avalanches finally reach the readout plane on the far end, where their positions are recorded, with a spatial resolution on the order of the hole diameter. The large surface area of the

readout plane is better equipped to handle large charge depositions than the many thin wires in a MWPC.

An essential consideration for the design of the LHCb detector was to strike the appropriate balance between maximizing the tracking efficiency and precision and minimizing the amount of material between the interaction point and the calorimeters (and, of course, cost). Charged particles passing through matter lose energy and can be rescattered, changing their direction. This can result in a loss of energy, pointing resolution, and tracking efficiency (should the track become unrecognizably distorted). The LHCb's design, in particular the number of tracking stations and combination of detector technologies in use, reflects this balance. Two silicon tracking stations are located upstream of the magnet: the Vertex Locator (VELO) and the Tracker Turicensis (TT). Another silicon station, the Inner Tracker (IT), is placed downstream of the magnet close to the beampipe. The TT and the IT are together known as the Silicon Tracker (ST). The final tracking station is the Outer Tracker (OT), a straw tube drift chamber located downstream of the magnet, at the same z position as the IT, but further away from the beampipe. The muon system, located behind the calorimeters, is not considered part of the tracking system, but uses multi-wire proportional chambers.

VELO

The Vertex Locator (VELO) is a silicon microstrip detector located very near the interaction point. It plays a key role in resolving vertices and measuring track impact parameters (IPs) with high precision [30]. The shape of each VELO silicon microstrip sensor is roughly a semi-annulus, with an inner radius of 8 mm and an outer radius of about 4.2 cm in diameter. The opposite surfaces of each sensor are etched with microstrips along the r and ϕ coordinates, respectively, and so the sensor measures the polar coordinates of track hits. The inner gap of the annulus acts as

an interaction region of the colliding proton beams. The detector consists of 42 of these sensors staggered along alternating sides of the beampipe over a distance of about 1 m. The arc measure of each sensor is 182°, so that when viewing the xy profile of the VELO, the alternating sensors partially overlap each other [41]. Thus there is no gap in ϕ that a track can pass through without leaving hits. The geometry of the VELO ensures that any track in LHCb's angular acceptance window ($15\,\mathrm{mrad} < \theta < 300\,\mathrm{mrad}$) passes through at least three of its sensors.

A novel requirement of the VELO detector is that it must mechanically move into place after the injection of the LHC beams and before operation. When the collider is being filled with protons, the effective diameter of the proton beam, or aperture, is wider than at operating conditions, and is eventually focused by the LHC's optics system. The final position in xy of the beam center after focusing varies within a few millimeters. In operation, the innermost reach of the VELO sensors is just 8 mm from the beam center. During the fills, the two halves of the VELO must retract a distance of 3 cm from the beamline. Then, after the fill, the two halves converge to the new proper position to within a tolerance of 20 μm to 100 μm (depending on the degree of freedom). The VELO successfully meets this engineering challenge.

Silicon Tracker and OT

The second upstream tracking detector is the Tracker Turicensis (TT), not a physics term, but instead derived from the Latin name for Zurich. This detector consists of four planar layers of silicon microstrip sensors in x, u, v, and x configurations [30]. The x planes measure the x-coordinate of track hits, while the u and v planes measure coordinates along axes that are rotated $\pm5°$ with respect to the x axis. Each plane contains a large number of silicon microstrip sensors, grouped into 17 or 19 vertical modules (depending on the layer) containing 14 sensors each.

Each sensor is approximately $100\,cm^2$ in area and $500\,\mu m$ thick, and etched with 512 silicon microstrips along the (mostly) vertical axis, with a pitch of $183\,\mu m$. As a result, the TT is able to measure the x and y coordinates of a track with about $50\,\mu m$ and $600\,\mu m$ resolution, respectively [42].[5]

The IT and OT geometrically (though not electronically) make up one detector located downstream of the LHCb magnet. The IT/OT assembly is divided into three stations, T1, T2, and T3, each of which consists of x-u-v-x planar layers, like the TT. Each plane has an IT and an OT component. The silicon IT component is cross-shaped and placed nearest the beampipe, where the number of tracks and detector occupancy are highest. The OT component making up the rest of the plane is a straw tube drift chamber.

The cross shape of each IT layer is made of four panels, above and below and to either side of the beampipe. Each panel is divided vertically into seven modules; each module of the upper and lower panels contains one silicon sensor while each module of the left and right panels contains two silicon sensors. The design of the IT's silicon sensors and their performance is very similar to that of the TT, with a spatial resolution of about $50\,\mu m$. Each sensor is rectangular and approximately $80\,cm^2$ in area, giving the full IT panel an area of about $0.35\,m^2$. The angular acceptance of the IT extends from $10\,mrad$ to $13.3\,mrad$ in the vertical direction and $80\,mrad$ in the horizontal direction.

The OT layers surround the IT layers and extend away from the beampipe out towards LHCb's maximum angular acceptances ($250\,mrad$ and $300\,mrad$ in the vertical and horizontal directions). Each layer of the OT contains two arrays of straw tubes, about $5\,mm$ in diameter and staggered to eliminate gaps. The straw tubes contain a mixture of argon and carbon dioxide [30]. The low electronegativity of argon prevents the capture of free electrons, while carbon dioxide acts as a

---

[5]The y coordinate is calculated from the x, u, and v coordinates and its resolution is worse by a factor of $\sim \csc 5°$.

quench gas, absorbing photons produced during the avalanche that could otherwise propagate and create displaced avalanches. A 1700 V potential difference is applied between the cathodes and anodes. In these conditions, the drift time of free electrons in the straw tubes is less than 50 ns. The OT's drift time resolution provides an effective spatial resolution of around 200 µm.

Tracking performance

Tracks of charged particles at LHCb are divided into several categories. Tracks that pass through at least the VELO and the IT or OT are known as long tracks; those that pass through only the VELO and TT are upstream tracks; those that pass through only the TT and IT or OT are downstream tracks; those that only leave hits in the IT or OT are T tracks; and finally those that only leave VELO hits are VELO tracks. Preferably, all charged particles created at the interaction point should leave long tracks, while those created during the decay of long-lived K_S and Λ particles are typically downstream tracks at best. The hits from different stations are combined to produce tracks during track reconstruction. For charged particles in LHCb's acceptance with a momentum in the range 5 GeV to 200 GeV, the track-finding efficiency is 96% [42]. About 6.5% of long tracks are fakes, or ghosts, created by joining hits from the VELO and IT or OT that were not created by the same particle. A high-performing neural network classifier is used to decrease this fake rate.

The momentum of the track is measured by its deflection when passing through LHCb's magnetic field. The mass of a composite particle that decays into charged particles is given by the invariant mass of the sum of its decay products' four-momenta. A number of particles decay to a muon pair $\mu^+\mu^-$, including the J/ψ and $\psi(2S)$, the $\Upsilon(nS)$ states, and the much heavier Z boson, providing a clean probe of the momentum resolution, mass resolution, and other attributes of LHCb's

(a) J/ψ resonance

(b) $\psi(2S)$ resonance

(c) $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ resonances

(d) Z boson
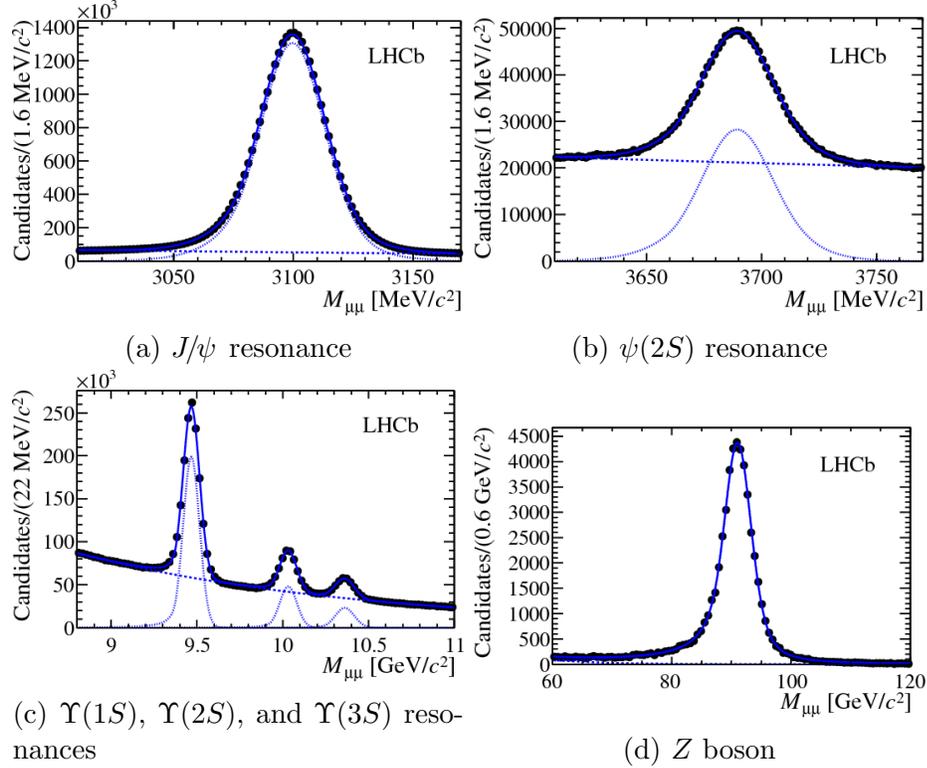
Figure 2.4: Fits to the invariant mass distributions of the J/ψ, $\psi(2S)$, $\Upsilon(nS)$, and Z dimuon resonances [42].

tracking system. Plots of the fits to the invariant mass distributions of these decays are shown in Fig. 2.4.

The momentum resolution, $\delta p/p$, is found to be about 0.5% for low momentum particles and degrades at higher momenta, plateauing at about 1.1% for momenta greater than 200 GeV. Similarly, the mass resolution is about 0.5% for the light and middleweight particles, and jumps to about 1.9% for the massive Z [42].

Finally, the precise location of the primary vertex (PV) where the protons collide and any secondary vertex (SV) where an unstable particle decays are important observables in many analyses performed at LHCb. These pieces of information are used to separate tracks coming from an SV from background tracks coming from the PV, and to calculate the flight distance and decay time of unstable particles. These latter quantities are especially crucial to analyses of time-dependent CP violation and the study of the decay $B_c^+ \to J/\psi\tau^+\nu_\tau$, both discussed in this thesis.
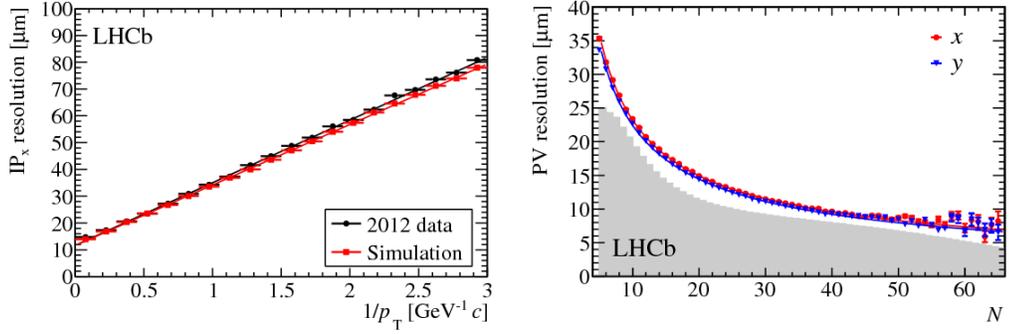
(a) IP resolution dependence on trans-
verse momentum.

(b) PV resolution dependence on the
number of tracks used in reconstruction.

Figure 2.5: IP and PV resolutions.

The resolution of the PV location, impact parameter (IP) of charged tracks — which is their distance of closest approach to the PV — and decay time are important metrics of the performance of the tracking system. The resolution on the x and y coordinates of the PV depends strongly on the number of tracks used to reconstruct the PV, and ranges between $7\,\mu\mathrm{m}$ to $35\,\mu\mathrm{m}$, with the worst performance for PVs created from only 5 to 10 tracks and much better resolution ($15\,\mu\mathrm{m}$ or better) when more than 20 tracks are used [42]. The resolution of the z coordinate is worse by a factor of 5 or 6. The IP resolution is nearly inversely proportional to the p_T of the track, and plateaus at $13\,\mu\mathrm{m}$ for the highest transverse momentum tracks. For the lowest p_T tracks reconstructed by LHCb, the IP resolution is about $80\,\mu\mathrm{m}$. A plot of the IP resolution is shown in Fig. 2.5a. The decay time of an unstable particle is calculated from its separation from the PV and its momentum, and so uncertainties in both the PV and SV positions and momenta of its decay products contribute to decay time uncertainty. For decays of neutral B^0 and B^0_s mesons studied in time-dependent CP violation analyses, the decay time resolution is typically about $50\,\mathrm{fs}$. Relative points of comparison are the lifetime of B mesons (around $1.5\,\mathrm{ps}$), the B^+_c meson lifetime (around $0.5\,\mathrm{ps}$), the B^0 oscillation period (about $12\,\mathrm{ps}$), and the B^0_s oscillation period (about $350\,\mathrm{fs}$) [12].

2.2.3 Particle identification

LHCb's particle identification (PID) systems consist of the ring-imaging Cherenkov (RICH) detectors, calorimeter system, and muon detector. These systems work together to discriminate various species of charged and neutral particles. Moreover, the calorimeter system and muon detector are important inputs to LHCb's hardware level trigger. The performance of the PID systems depends critically on the amount of material inside the LHCb detector. In particular, the energy resolution of the calorimeter systems depends on their ability to completely contain the electromagnetic and hadronic showers produced by interactions, and so it is critical to maximize their instrumented material while minimizing interfering material upstream. Conversely, the muon system identifies $\mu^\pm$ by maximizing the shielding of all other charged particles; for the trigger, though, it also must measure p_T with some acceptable resolution, which requires limiting the amount of shielding and the consequent multiple scattering it produces. The proper balance of these various concerns is considered in the material budget of each subdetector. An important figure of merit is the size of the component in radiation lengths. The radiation length X_0 of a material, which is the characteristic length scale of high energy electron and photon energy loss in the material, and also of the resulting electromagnetic (EM) showers as they traverse matter [12]. A average path through the tracking system traverses about 0.25 radiation lengths [30].

The RICH detectors are designed to distinguish charged tracks, primarily pions and kaons, by exploiting the Cherenkov effect. When a charged particle travels through a medium with refractive index n at a speed β faster than the speed of light $1/n$, it radiates at an angle $\cos\theta_c = 1/\beta n$ with respect to its velocity [12, 40]. This light cone shines onto a ring on a transverse detector, which allows measurement of the velocity β of the particle. Together with the momentum p measured by

the tracking system, the mass m of the particle can be inferred. There are two RICH detectors at LHCb, one located before the magnet (RICH1) and one after (RICH2). RICH1, located between the VELO and the TT, is optimized for lower momentum tracks ($1\,\mathrm{GeV}$ to $60\,\mathrm{GeV}$) and RICH2, between the IT/OT and the first muon station, for higher momentum tracks ($15\,\mathrm{GeV}$ to $100\,\mathrm{GeV}$) [30]. Lower momentum tracks can be bent out of the LHCb cone, so measuring their velocities must be done upstream of the magnet.

The range of speeds β at which a radiator is sensitive is limited. At the low end, a particle's velocity β must be above the threshold value $\beta_t = 1/n$ to produce Cherenkov radiation at all. At the high end, θ_c quickly approaches a maximum value $\cos^{-1}(1/n)$ as $\beta \to c$ and discriminating power is lost. To increase its sensitivity range, RICH1 uses both solid aerogel and gaseous C_4F_{10} radiators, with $n = 1.03$ and $n = 1.0014$ respectively for $\lambda = 400\,\mathrm{nm}$. The typical θ_c produced in the aerogel radiator is $50\,\mathrm{mrad}$ to $242\,\mathrm{mrad}$, while that in C_4F_{10} is $20\,\mathrm{mrad}$ to $50\,\mathrm{mrad}$. RICH2 uses as its readiator CF_4, with $n = 1.0005$ at the same wavelength. These different choices reflect the requirement that the RICH1 have greater sensitivity at lower momenta and the RICH2 at higher momenta.

The geometries of RICH1 and RICH2 are similar and are designed to minimize their material budget, resulting in radiation lengths of $0.08X_0$ and $0.15X_0$ respectively. The acceptances of the detectors are $25\,\mathrm{mrad}$ to $300\,\mathrm{mrad}$ in the horizontal plane and $25\,\mathrm{mrad}$ to $250\,\mathrm{mrad}$ in the vertical plane for RICH1 (limited on the low end by the beampipe) and $15\,\mathrm{mrad}$ to $120\,\mathrm{mrad}$ in the horizontal plane and $15\,\mathrm{mrad}$ to $100\,\mathrm{mrad}$ in the vertical plane for RICH2. Inside their acceptances, tracks encounter only the radiators, spherical mirrors, and gas-tight entrance and exit windows, in that order. Cherenkov radiation produced in the radiators travels upstream to the spherical mirrors placed immediately around the beampipe, where it is reflected at an angle backwards out of the acceptance region. In this outer re-
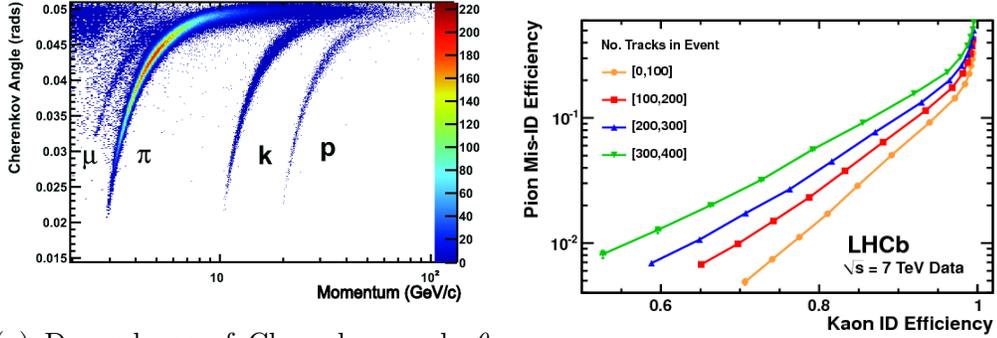
(a) Dependence of Cherenkov angle θ_c on particle species and momentum, for C_4F_{10}

(b) K^+–π^+ classification curve, showing dependence on track multiplicity

Figure 2.6: RICH kaon identification performance [43]

gion, the radiation is reflected forwards by flat mirrors onto hybrid photon detector (HPD) enclosures.[6] The custom HPDs used at LHCb are similar to photomultipliers, consisting of a photocathode followed by a vacuum acceleration region under a potential difference of 20 keV, and finally by a pixelated silicon detector. The HPDs require that the RICH1 and RICH2 be encased in substantial magnetic shielding, except at the entrance and exit windows.

The calorimeter system, located behind RICH2 (and the first muon station M1), consists mainly of the electromagnetic calorimeter (ECAL) followed by the hadronic calorimeter (HCAL). Several auxiliary components are placed in front of the ECAL: the scintillator pad detector (SPD), a thin lead converter, and the pre-shower (PS).

With momenta greater than about 10 MeV, electrons (and positrons) traversing through matter lose the great majority of their energy via brehmsstrahlung, while photons are converted to e^+e^- pairs. Brehmsstrahlung is not an important affect for the more massive $\mu^\pm$ and $\pi^\pm$ until energies over 100 GeV, and even higher energies for $K^\pm$ and $p^\pm$. So, the electromagnetic interactions in the ECAL are most relevant for photons and electrons, while other particles passing through suffer only

---

[6]This double-reflection allows the detector to be narrower in the z-direction and taller in the less constrained y-direction.

from ionization deposits and multiple scattering.[7] The characteristic scale of the transverse spread of an EM shower is its Molière radius, which is related to the radiation length and to the atomic number of the material. Typical Molière radii are on the order of several centimeters, and the bulk of the shower's energy is deposited within two Molière radii.

Of the remaining types of tracks, all but $\mu^{\pm}$ are hadrons and have potentially large hadron-nucleon cross-sections $\sigma(hN)$. The nuclear interaction length λ_i measures the characteristic length scale of hadronic showers. Typically λ_i is much greater than X_0, making it much more difficult to completely contain hadronic showers than EM showers, given the same space constraints. In the transverse plane, the bulk of the energy of the shower is deposited within a radius of a single interaction length, again larger than the transverse size of an EM shower. Many photons and electrons are produced in the hadron shower and produce subsidiary EM showers, carrying a large fraction of the shower's total energy. Additionally, much of the shower's energy is spent towards nuclear binding energies in inelastic collisions and is not observable. These effects can make it more difficult to achieve good energy resolution in a hadron calorimeter.

The ECAL detects EM showers produced by photons and electrons, information which is used for the L0 hardware trigger and for offline analysis. It is a sampling, shashlik type calorimeter, consisting of alternating planes of lead absorbers $2\,\mathrm{mm}$ thick and scintillating tiles that are segmented into a grid in the xy planes and run through with transverse fibers that carry scintillation light to photomultiplier tubes. There are in total 66 lead/scintillator layers. The total radiation length of the ECAL, due mostly to the lead absorbers, is $25X_0$ (about $42\,\mathrm{cm}$). The nuclear interaction length of the ECAL should be small to limit hadronic backgrounds in the ECAL and to avoid impeding the performance of the HCAL; it is $1.2\lambda_i$.

---

[7]Neutral pions decay via $\pi^0 \to \gamma\gamma$ before the ECAL and are reconstructed as resolved candidates, with two distinct photon showers, or merged candidates, where the two showers overlap.

The scintillating tiles in the ECAL are mainly made of polystyrene, with small concentrations of fluors that shift the wavelength of the UV light that is initially produced down into the blue spectrum. The emitted scintillation light is collected by wavelength-shifting (WLS) fibers that further shift its wavelength down from the blue spectrum to the green spectrum and carry it downstream through the ECAL to an array of photomultiplier tubes (PMTs) that collect the light.

The ECAL provides information for the hardware-level triggers on high transverse energy (E_T) electrons and photons. The trigger must separate electron showers from photon showers and also must suppress a few backgrounds. The main backgrounds for the electron trigger are high energy $\pi^0 \to \gamma\gamma$ decays with merged photon showers and $\pi^\pm$ interactions; the former is also a background for the photon trigger. The auxiliary SPD and PS detectors in front of the ECAL provide electron–photon separation and suppress these backgrounds.[8] The SPD and the PS are two scintillator pads located on either side, respectively, of a lead converter 2.5 radiation lengths thick (about 15 mm). The SPD identifies an ECAL deposit as an electron shower by detecting charged particles that produce scintillation light in the SPD pad upstream of an ECAL shower. Similarly, SPD hits are used to reject π^0 background for the electron trigger. Meanwhile, the PS rejects $\pi^\pm$ background by detecting EM showers produced in the lead converter; in the same way, it also partially rejects π^0 background for the hardware-level photon trigger. The readout of the SPD and PS is similar to that of the ECAL: they are segmented into an xy grid and transverse WLS fibers bring the signal downstream to PMTs. The principle difference is that multi-anode PMTs (MAPMTs) are used, in which multiple PMTs with relatively smaller anodes are enclosed in the same housing for increased spatial resolution.

Charged pions, kaons, protons, and neutrons are absorbed by the HCAL. The technology of the HCAL is similar to that of the ECAL. The HCAL uses iron as the

---

[8]A residual background $\gamma \to e^+e^-$ pair conversions induced by material upstream of the ECAL cannot be rejected at the hardware level.

absorber material. The scintillating tiles, WLS fibers, and PMTs used in the HCAL are similar to those of the ECAL. The geometric structure of the HCAL, however, is quite different from the ECAL. While the ECAL tiles and absorbers are perpendicular to the beampipe, the HCAL layers are parallel to the beampipe. A longitudinal path through the HCAL traverses just three absorber/scintillator layers, largely because of space limits in the LHCb chamber; the total nuclear interaction length of the HCAL is $5.6\lambda_i$. However, the large transverse spread of a hadronic shower passes through many such layers. This layout optimizes the HCAL's sensitivity to hadronic shower shape given its size constraints.

The ECAL and HCAL extract several types of information from EM and hadron showers. Primarily, the amount of the scintillation light collected is proportional to the energy of the shower; since the number of photons (in the ECAL case) produced follows Poisson statistics, up to Fano corrections, this means that the calorimeters have energy resolutions that scale like $\sigma(E)/E \sim 1/\sqrt{E}$, plus constant systematic factors. The total energy E of the shower and its location in η on the calorimeter determine its transverse energy E_T, defined as

$$E_T = \sqrt{m^2 + p_T^2} \tag{2.1}$$

where some standard mass assumption is made. The quantity E_T is very useful in the hardware-level trigger. Finally, the overall shape of the shower can reveal information on the species of the incident particle. In particular, in offline analysis the shape can be analyzed to separate neutral pion showers from photon showers in the ECAL. Figure 2.7 shows that data from the calorimeter system provide moderate-to-good discrimination between neutral pions and photons; eliminating 95% of pions requires a cut that keeps only 60% of photons.

The muon system consists of five stations M1 through M5, the first of which
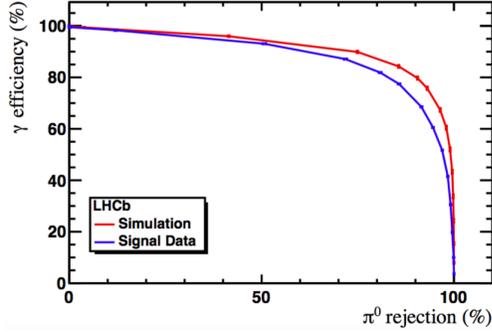
Figure 2.7: Discrimination of γ and merged π^0 showers in the ECAL

is located in front of the calorimeters and the remainder behind, separated from each other by iron absorbers [30]. Because they are largely located behind the calorimeters and iron filters, few charged electrons or hadrons penetrate into the muon stations. The ECAL and HCAL material in front of the M2-M5 muon stations corresponds to 25 electromagnetic radiation lengths and 5.6 nuclear interaction lengths, and the total material of muon stations contributes another 13 interaction lengths. Therefore, muon detector hits are a powerful way to distinguish muons from hadrons. A muon with a momentum of at least 6 GeV is able to penetrate all the way to M5 and leave hits at each station. Because M4 and M5 have the fewest number of non-muon hits out of the muon stations, and because muons reaching them are most subject to energy loss and multiple scattering, they are designed primarily to identify penetrating muons and have relative relaxed spatial resolution. M1 through M3, on the other hand, provide better spatial resolution for the muon detector hits. Their spatial resolution is good enough to measure the p_T of a muon track with a resolution of around 20%, quickly and accurately enough for use in LHCb's trigger system. Each station uses multi-wire proportional chambers (MW-PCs) to record muon track hits, though M1, located in front of the calorimeters, uses triple-GEM chambers in the high-occupancy region nearest the beampipe. The MWPC and GEM chambers are filled with a mixture of argon, carbon dioxide, and

methane gas, where the latter two serve as quenchers. The muon system provides an L0 hardware level trigger on high transverse momentum muons and dimuons and higher level triggers on reconstructed J/ψ mesons, used in the $B_c^+ \rightarrow J/\psi \tau^+ \nu_\tau$ analysis.

The large number of components in the PID subsystem provide a wealth of information for each charged track passing through the LHCb detector. To simplify the use of this information by analysts, a high-performing neural network assigns a score to each track for each particle hypothesis; for example, a purified kaon sample may be selecting by requiring a high kaon score and low pion or proton score. These PID scores are used frequently in LHCb analyses and are referenced throughout this thesis.

2.2.4 Trigger system

The trigger system at LHCb is divided into a hardware level (L0) and two software levels — the High Level Trigger 1 (HLT1) and High Level Trigger 2 (HLT2). The L0 hardware trigger processes information directly from the VELO, calorimeter system, and muon system to produce a real-time trigger decision. Given the 50 ns bunch crossing time in Run I, and 25 ns in Run II, this requires very fast data collection and processing via the electronics system — the L0 trigger must operate at 40 MHz. After making its trigger decisions with a 2.5% retention rate, it outputs data at 1 MHz, which is the maximum rate possible due to the time required to read out the detector.

The L0 trigger has five channels: muon, dimuon, photon, electron, and hadron [42]. The muon trigger identifies the highest p_T muon and requires $p_T > 1.48$ GeV, while the dimuon trigger identifies a pair of muons with a geometric mean p_T greater than 1.30 GeV. These muon triggers lean on M4 and M5 to identity the muons and M1 through M3 to provide a p_T measurement. The photon and electron triggers

identify the highest E_T ECAL electron and photon showers (with crucial background rejection provided by the SPD and PS), and require their E_T to be greater than 2.50 GeV. Finally, the hadron trigger requires the E_T of hadron showers to be greater than 3.50 GeV (adding in the E_T of the matching deposit in the ECAL). The VELO system provides a pile-up veto when there are multiple detectable PVs; the r–ϕ geometry of the VELO microstrips is crucial to the quick counting of PVs [30].

With a 40-fold increase in available processing time, the high level triggers process additional information from the tracking system and RICH detectors, retaining just 0.2% of events and outputting data at 5 kHz in Run 1 [44]. The first level, HLT1, largely refines the L0 trigger lines, defining several streams for different event types. For each stream defined in HLT1, there is a large number of HLT2 trigger lines, which range from broad general purpose lines to analysis-specific lines. Taking advantage of tracking information requires performing track reconstruction. Because of time constraints during Run I, simplified reconstruction software, in comparison to the full offline reconstruction software, was used in the HLT. For LHC's Run II, LHCb has switched to a novel online trigger alignment and calibration system [45]. While the bunch crossing time in Run II has decreased to 25 ns, the LHC collider is still in downtime between fills about 70% of the time. A large data buffer (about 5.5 PB) allows the HLT to run during this downtime, and with the extra processing time available it can use the same reconstruction software as is used offline. This also requires calibration and alignment to be performed online. Interpretation of the data collected at LHCb depends on the detector's runtime conditions, including the beam conditions and detector positions; recall that the beam's xy center varies on the order of 1 mm from fill to fill and that the VELO's position varies in response to this. Other variable conditions such as temperature and pressure affect the gases in the RICH detectors, OT, and muon system. The calibration and alignment procedure corrects for these affects. During Run I, this was performed

offline using unbiased data collected via calibration trigger lines, but in the current run is performed automatically between the HLT1 and HLT2.

2.2.5 Simulation

Particle physics processes are simulated by a chain of software collectively referred to as Monte Carlo (MC) generators. The simulated data produced by this software is called Monte Carlo data, as well. The first stage of the simulation generation is the simulation of pp collisions, performed using the PYTHIA generator [46, 47]. PYTHIA uses measured parton distribution functions to simulate the quark and gluon contents of colliding protons. It then simulates $2 \rightarrow 1$, $2 \rightarrow 2$, and $2 \rightarrow 3$ QCD and QED scattering processes and models the showers of the resulting quarks and gluons, producing detector-stable final particles (*e.g.*, various hadrons, mostly). While the parton collision itself is modeled using leading order and next-to-leading order matrix elements, the simulated hadronization relies heavily on measured properties of pp collisions rather than directly simulating soft QCD. A special configuration of PYTHIA tunes this experimental input to optimize the event generation for the soft, high pseudorapidity region studied by LHCb [48]. Several plugin tools exist to improve PYTHIA's modeling of the production of rare, exotic, or hypothetical particles in showers. In particular, BCVEGPY modifies the production of the relatively rare $B_c^{\pm}$ mesons to better explain more recent experimental findings. PYTHIA does not model the decays of the hadrons produced as a result of the parton showers. This is achieved by the EVTGEN package [49], which models hadron decays using analytic expressions derived from QFT, but with numerical factors like masses, branching fractions, and some distribution shape parameters (*e.g.*, form factors) taken from experimental measurements. EVTGEN also uses the PHOTOS [50] tool to model final-state radiation generated during decays. After the decay process, the final state particles are a mixture of leptons and some long-lived

hadrons (*e.g.*, pions, kaons, Λ). These pass through and interact with the material of detector, leaving signals in their wake, but also being affected in turn through loss of energy, multiple scattering, neutral kaon regeneration, $\gamma \to e^+e^-$ conversion, decay in flight, *etc.* The interaction with matter in the detector system is modeled by the GEANT4 toolkit [51, 52]. GEANT4 requires a detailed model of the LHCb detector, described in Ref. [53].

The output of the simulation after GEANT4 is a physical model of *pp* collisions and the resulting response of the LHCb detector. This is not yet analogous to the raw electronic data collected by the LHCb subdetectors. For this purpose, the BOOLE software simulates the electronic response of the LHCb subdetectors to an simulation event. After processing by BOOLE, simulation data is stored in the same format as real data and can be processed using the same tools.

2.2.6 Computing

The LHCb collaboration participates in a globally distributed computing network used for the storage and analysis of particle physics data [54, 55]. Several tiers of clusters store and replicate data collected by the LHC experiments, as well as simulated simulation datasets, and provide computational nodes for the analysis of the data. The raw data collected by LHCb (or produced during simulation) is first stored at CERN's Tier-0 storage center. This raw data is replicated to a number of Tier-1 storage centers located across Europe. The full offline reconstruction of particle tracks is eventually performed on this raw data, a procedure known as stripping. The computational resources for this stripping are provided by the Tier-0 and Tier-1 centers, as well as a number of Tier-2 computing-only centers. The stripped data is also replicated across the grid. The grid's storage and computing resources are also used for simulation and for user analyses that process the stripped data, searching for the signatures of various particle decays and event types.

Part II

Semitauonic B_c^+ decays

Chapter 3:   Theoretical background

The goal of this analysis is the first measurement of the branching fraction
ratio

$$\mathcal{R}(J/\psi) = \frac{\mathcal{B}(B_c^+ \to J/\psi\,\tau^+\nu_\tau)}{\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu_\mu)}, \tag{3.1}$$

which is sensitive to new physics.[1] In constructing the Standard Model (SM), the electroweak couplings of all flavors of leptons are taken to be identical, with the three distinguished only by their couplings to the Higgs sector. This feature is known as lepton flavor universality, and implies that differences between branching fractions involving the three leptons differ only due to their separate masses (see Section 1.3). The direct effect of these masses in electroweak processes is reliably calculable, and the semileptonic decay is well-approximated by the simplest tree level diagram. Moreover, factors with experimental uncertainties, such as $|V_{cb}|^2$, cancel in the ratios of branching fractions such as Eq. (3.1). However, the ratio $\mathcal{R}(J/\psi)$ also depends on non-perturbative QCD effects encoded in the $B_c^+ \to J/\psi$ hadronic form factors. Current models predict $\mathcal{R}(J/\psi)$ to be in the range 0.25–28 [56–59]. The observable $\mathcal{R}(J/\psi)$ is analogous to the ratios

$$\mathcal{R}(D^{(*)}) = \frac{\mathcal{B}(B^{(0,-)} \to D^{(*-)}\tau^+\bar{\nu}_\tau)}{\mathcal{B}(B^{(0,-)}) \to D^{(*-)}\mu^+\bar{\nu}_\mu)}, \tag{3.2}$$

discussed in the introduction.

The chapter gives an overview of the necessary theoretical background for the

---

[1]Here and throughout, charge conjugate modes are implied.

description of the analysis presented in the next chapter. This begins with a brief summary of the properties of the B_c^+ meson and of semileptonic decay branching fractions to tree-level in the weak force. This is followed by a discussion of theoretical models for the form factors of semileptonic B_c^+ decays and of generic parametrized models for these form factors. Finally, there is a discussion of the mathematical method used to approximation the rest frame of the B_c^+ meson, relying on the fact that this spinless meson is not produced in a polarized state.

3.1 The B_c^+ meson and its decays

The B_c^+ meson consists of valence $\bar{b}$ and c quarks, together with the typical sea of $q\bar{q}$ pairs and gluons. With a mass of 6.277 GeV/c^2 [12], it is too heavy to be produced in $\Upsilon(nS)$ decays at the e^+e^- B factory experiments. Evidence for the B_c^+ was seen at OPAL, and subsequently it was first observed by the CDF experiment at the Tevatron collider in the decay $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ [60, 61]. Because the b quark is more stable than the c quark, the B_c^+ is expected to decay to $B_s X$ around 70% of time [58]; one such decay has been observed at LHCb [62]. To a good approximation, when either the $\bar{b}$ or c decays via a weak transition, the other acts as a spectator quark, and thus the B_c^+ width should be roughly the sum of D^+ and B^+ meson widths (plus the partial width for $\bar{b}c$ annihilation, which is suppressed by the decay constant f_{B_c} and helicity factors). Neglecting annihilation, this approximation predicts a lifetime of 0.637 ps, which is not far off the world average lifetime 0.507(9) ps [12].

3.2 $B_c^+ \to c\bar{c}$ form factors

The matrix element of the semileptonic decay of a particle of mass M to a three-body $X\ell\nu$ state is the product of two tensors describing the leptonic and
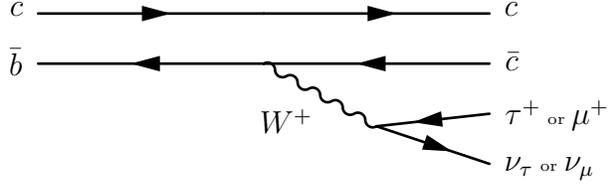
Figure 3.1: Semileptonic B_c^+ decays in the SM
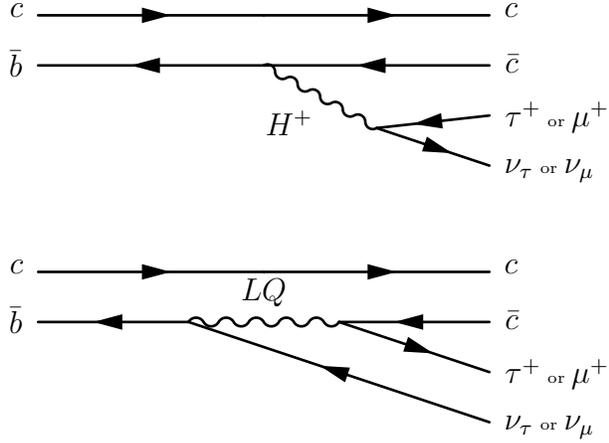
Figure 3.2: Semileptonic B_c^+ decays beyond the SM

hadronic portions of the process. It is most easily computed by summing over the helicity states of the virtual $W^\pm$ mediating the decay. These are

$$\mathcal{M}_{\lambda_X}^{\lambda_\ell} = \frac{1}{\sqrt{2}} G_F V_{cb} \sum_{\lambda_W} \eta_{\lambda_W} L_{\lambda_W}^{\lambda_\ell} H_{\lambda_W}^{\lambda_X}$$

where η_{λ_W} is 1 for the ± 1 and 0 helicity states and -1 for the timelike scalar (s) helicity state. The helicity components of the tensors are:

$$L_{\lambda_W}^{\lambda_\ell} = \epsilon_\mu(\lambda_W) \, \langle \ell \bar{\nu} | \bar{\ell} \gamma^\mu \left(1 - \gamma^5\right) \nu | 0 \rangle, \tag{3.3}$$

$$H_{\lambda_W}^{\lambda_X} = \epsilon_\mu^*(\lambda_W) \, \langle X | \bar{c} \gamma^\mu \left(1 - \gamma^5\right) b | B \rangle. \tag{3.4}$$

The components of the leptonic tensor are quite simple. Because of the weakness of the electroweak force, $b \to c$ decays can be treated perturbatively, and the tree-level

46

(a) B_c^+ rest frame   (b) Leptonic rest frame
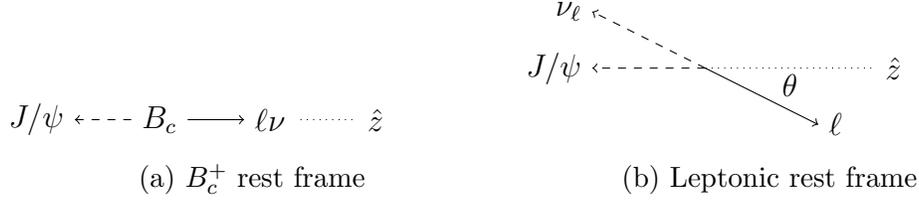
Figure 3.3: Diagram of the semileptonic decay of a B_c^+ into a hadronic part (J/ψ) and leptonic part (ℓ and ν_ℓ). As shown, θ is the angle between the lepton and the leptonic (*i.e.*, lepton-neutrino) flight direction in the leptonic rest frame.

decay through a single off-shell $W^\pm$ is a good approximation. Thus, the leptonic tensor is described entirely by the lepton mass, the off-shell $W^\pm$ mass-squared q^2, and the angular parameter $\cos\theta$, shown in Fig. 3.3. However, the initial and final state bottom and charm mesons are bound states of the strong force, and the long range details of their destruction and creation depend non-perturbatively on strong dynamics. Conservation of azimuthal spin does mean that $H_{\lambda_W}^{\lambda_X}$ is largely "diagonal." For a spin-0 decay product X, whose only helicity state is 0, only H_0^0 and H_s^0 are non-zero. For a spin-1 decay product, with additional ± 1 helicity states, H_+^+ and H_-^- are also non-zero. A spin-2 decay product has ± 2 helicity states, but these are not accessible from any W helicity state. So, without any ambiguity, the non-zero hadronic tensor helicity components can be denoted $H_\pm$, H_0, and H_s, where only the last two are non-zero for scalar decay products.

After integrating over the available phase space, the doubly differential decay rate with respect to q^2 and the helicity angle θ is proportional to the square of the matrix element:

$$
\begin{aligned}
\frac{d\Gamma}{dq^2\, d\cos\theta} &= \frac{|\vec{p}_X|\,(q^2)}{256\pi^3 M^2}\,|\mathcal{M}|^2 \\
&= \frac{G_F^2 |V_{cb}|^2\,|\vec{p}_X|\,(q^2)q^2}{256\pi^3 M^2}\left(1-\frac{m_\ell}{q^2}\right)^2 \\
&\quad \times \Big[(1-\cos\theta)^2 H_+^2 + (1+\cos\theta)^2 H_-^2 + 2\sin^2\theta H_0^2 \\
&\qquad +\frac{m_\ell^2}{q^2}\left(\sin^2\theta\left(H_+^2 + H_-^2\right) + 2\left(H_s - H_0\cos\theta\right)^2\right)\Big],
\end{aligned}
\tag{3.5}
$$

47

where $|\vec{p}_X|\,(q^2)$ is the momentum of the decay product X in the rest frame of the B parent B, which is equal to

$$|\vec{p}_X|\,(q^2) = \sqrt{\left(\frac{M^2 + M_X^2 - q^2}{2M}\right)^2 - M_X^2}. \tag{3.6}$$

Integrating the doubly differential decay rate over $\cos\theta$ gives the differential decay rate as a function of q^2:

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2 |V_{cb}|^2\, |\vec{p}_X|\,(q^2) q^2}{96\pi^3 M^2}\left(1 - \frac{m_\ell}{q^2}\right)^2 \times \\ \left[\left(1 + \frac{m_\ell^2}{2q^2}\right)\left[H_+^2 + H_-^2 + H_0^2\right] + \frac{3m_\ell^2}{2q^2}H_s^2\right] \tag{3.7}$$

Various symmetries constrain the tensor form of the matrix element $\langle X|\bar{c}\gamma^\mu\,(1-\gamma^5)\,b|B\rangle$ (and thereby the hadronic tensor helicity components) to a sum of several distinct components, each of which has a scalar magnitude with functional dependence on q^2, known as a form factor. This tensor form depends on the spin (and parity) of the decay product X. In the following, let the momentum of the B be p^μ, that of its decay product X be k^μ, and let $P = p+k$ and $Q = p-k$.

Scalar X: For a scalar or pseudoscalar decay product, the tensor form of the matrix element must be of the form

$$\langle X|\bar{c}\gamma^\mu\left(1-\gamma^5\right)b|B\rangle = A(q^2)P^\mu + B(q^2)Q^\mu. \tag{3.8}$$

When X is a scalar, this transition is actually entirely through the vector channel $(c\gamma^\mu b)$ — since Bs are pseudoscalars — and conversely is entirely through the axial vector channel $(c\gamma^\mu\gamma^5 b)$ when X is a pseudoscalar. Projecting this form into its helicity components according to Eq. 3.4, the helicity components

are

$$H_0 = \frac{2M\,|\vec{p}_X|\,(q^2)}{\sqrt{q^2}}A(q^2)$$

$$H_s = \frac{M^2 - M_X^2}{\sqrt{q^2}}A(q^2) + \sqrt{q^2}B(q^2).$$

The form factors are typically recast as

$$\langle X|\bar{c}\gamma^\mu\left(1 - \gamma^5\right)b|B\rangle = \left(P^\mu - \frac{M^2 - M_X^2}{q^2}Q_\mu\right)f_+(q^2)$$
$$+ \left(\frac{M^2 - M_X^2}{q^2}Q_\mu\right)f_0(q^2)$$

(3.9)

so that H_0 and H_s are proportional to $f_+(q^2)$ and $f_0(q^2)$, respectively:

$$H_0 = \frac{2M\,|\vec{p}_X|\,(q^2)}{\sqrt{q^2}}f_+(q^2), \tag{3.10}$$

$$H_s = \frac{M^2 - M_X^2}{\sqrt{q^2}}f_0(q^2). \tag{3.11}$$

Vector X: For a vector or pseudovector X, the tensor form must be

$$\langle X|\bar{c}\gamma^\mu\left(1 - \gamma^5\right)b|B\rangle = iA(q^2)\varepsilon^{\mu\nu\rho\sigma}\epsilon^*(\lambda_X)_\nu P_\rho Q_\sigma$$
$$- B(q^2)\epsilon^*(\lambda_X)^\mu$$
$$- C(q^2)\left(\epsilon^*(\lambda_X)\cdot Q\right)P^\mu$$
$$- D(q^2)\left(\epsilon^*(\lambda_X)\cdot Q\right)Q^\mu.$$

When X is a vector, A corresponds to the vector channel $((c\gamma^\mu b)$ and B, C, D to the axial vector channel $((c\gamma^\mu\gamma^5 b)$, and vice versa when X is a pseudovector. To simplify their relationship to the helicity components, and also to give them the same units as the scalar form factors f_+ and f_0, the standard

49

parameterization of the vector form factors is

$$
\begin{aligned}
\langle X | \bar{c} \gamma^\mu \left(1 - \gamma^5 \right) b | B \rangle = {} & \frac{2iV(q^2)}{M + M_X} \varepsilon^{\mu\nu\rho\sigma} \epsilon_\nu^* p_\rho k_\sigma \\
& - 2 M_X A_0(q^2) \frac{\epsilon^* \cdot q}{q^2} Q^\mu \\
& - (M + M_X) A_1(q^2) \left(\epsilon^{*\mu} - \frac{\epsilon^* \cdot Q}{q^2} Q^\mu \right) \\
& + \frac{1}{M + M_X} A_2(q^2)(\epsilon^* \cdot Q) \left(P^\mu - \frac{M^2 - M_X^2}{q^2} Q^\mu \right).
\end{aligned}
$$

$$(3.12)$$

With minor differences, this convention is followed by many theory papers [16, 49,63]. In terms of these form factors, the hadronic tensor helicity components are: [16]

$$
H_\pm(q^2) = (M + M_X) A_1(q^2) \mp \frac{2M}{M + M_X} |\vec{p}_X|(q^2) V(q^2), \tag{3.13}
$$

$$
\begin{aligned}
H_0(q^2) = {} & -\frac{1}{2 M_X \sqrt{q^2}} \left[\frac{4 M^2 |\vec{p}_X|(q^2)^2}{M + M_X} A_2(q^2) \right. \\
& \left. - (M + M_X) \left(M^2 - M_X^2 - q^2 \right) A_1(q^2) \right],
\end{aligned} \tag{3.14}
$$

$$
H_s(q^2) = \frac{2M |\vec{p}_X|(q^2)}{\sqrt{q^2}} A_0(q^2). \tag{3.15}
$$

Tensor X: For a tensor X, the tensor form is quite similar to that of a vector X, with the substitution

$$
\epsilon^*(\lambda_X)_\mu \to \frac{\epsilon^*(\lambda_X)_{\mu\nu} P^\nu}{M}. \tag{3.16}
$$

Viewing the spin-2 tensor particle as a superposition of two spin-1 vector particles, the polarization tensors can be built up out of the polarization vectors via Clebsch-Gordan coefficients:

$$
\epsilon(\lambda_X)_{\mu\nu} = \sum_{\lambda_i, \lambda_j} \langle 2\lambda_X | 1\lambda_i 1\lambda_j \rangle \epsilon(\lambda_i)_\mu \epsilon(\lambda_j)_\nu. \tag{3.17}
$$

Explicit calculation reveals that the polarization tensors of interest are:

$$\epsilon(\pm 1)_{\mu\nu}P^\nu = \sqrt{\frac{1}{2}}\frac{|\vec{p}_X|\,(q^2)}{M_X}\epsilon(\pm 1)_\mu, \tag{3.18}$$

$$\epsilon(0)_{\mu\nu}P^\nu = \sqrt{\frac{2}{3}}\frac{|\vec{p}_X|\,(q^2)}{M_X}\epsilon(0)_\mu. \tag{3.19}$$

This extra factor of $|\vec{p}_X|\,(q^2)/M_X$ can be interpreted as a consequence of the P-wave decay that is required for a scalar (B) decaying to a tensor (X) and a vector (W). Thus, given an otherwise identical parameterization of the matrix element in terms of form factors A, B, C, and D, the effect of spin-2 final state particle is accounted for by the substitution

$$H_\pm|_{spin\text{-}2} = \sqrt{\frac{1}{2}}\frac{|\vec{p}_X|\,(q^2)}{M_X}\,H_\pm|_{spin\text{-}1}, \tag{3.20}$$

$$H_{0,s}|_{spin\text{-}2} = \sqrt{\frac{2}{3}}\frac{|\vec{p}_X|\,(q^2)}{M_X}\,H_{0,s}|_{spin\text{-}1} \tag{3.21}$$

in the doubly differential decay rate formula (Eq. (3.5)).

3.2.1 B_c^+ form factor models

Measurements of decay rates can be used to determine semileptonic form factors. Though naively these are arbitrary functions of q^2, theoretical models taking advantage of exact and approximate symmetries in QCD (for example the CLN parameterization [64]) determine them up to a finite number of parameters; these can be extracted in a fit to experimental data. This has been done for the $B \to D^{(*)}$ transitions by the ALEPH, CLEO, OPAL, DELPHI, Belle, and BaBar collaborations, whose results were rescaled and combined by the HFLAV group [13]. It has not yet been possible to do this for the $B_c^+ \to J/\psi$ transitions because of limited experimental data. Therefore, Monte Carlo simulations of these decays must rely on entirely theoretical models of the form factors. Many authors have produced models

of the $B_c^+ \to J/\psi$ form factors, in a large array of parameterizations and theoretical models. Among these are models by Kiselev [57] and Ebert, Faustov, and Galkin (EFG) [63] models of $B_c^+ \to \psi$ transitions, which are among the most recent and have been adopted as the "default" parameterizations by LHCb. However, neither of these papers derives form factors for the three χ_c final states (scalar, vector, and tensor). One model of the form factors for all the $B_c^+ \to \chi_c$ transitions is given by Wang, Wang, and Lu (WWL) [65], and was chosen to model these decays for this analysis. The study of B_c^+ semileptonic decay form factors, to the above final states and others, has also been considered in various formalisms: relativistic and non-relativistic Bethe-Salpeter equation based models [66–68]; light-front quark models [56, 69]; and various other models [58, 70].

The Kiselev model of form factors is derived from a non-relativistic QCD (NRQCD) treatment, in which the QCD Lagrangian is expanded in powers of Λ/m_b, as in heavy quark effective theory, and also in v/c. The model uses Shifman-Vainshtein-Zakhorov QCD sum rules to evaluate the $B_c^+ \to J/\psi$ form factors, accounting for Coulomb-like corrections. The QCD sum rule technique evaluates QCD correlation functions by complex analysis techniques that give their value in terms of their poles; in turn these poles are identified as hadronic resonances, allowing input from experimental data on the hadron properties [71]. Kiselev's form factors have 5 free parameters, shown in Table 3.1. The paper computes the values of the form factors as a function of q^2 numerically, but fits them with a parameterization function proportional to

$$k(q^2) = \frac{1}{1 - q^2/M_{\text{pole}}^2}. \tag{3.22}$$

The four form factors $F_V(q^2)$, $F_A^+(q^2)$, $F_A^0(q^2)$, and $F_A^-(q^2)$, which are closely related to the standard parametrization $V(q^2)$ and $A_n(q^2)$, are all proportional to $k(q^2)$, and their values at $q^2 = 0$ are parameters of the model. In the helicity components

Table 3.1: Parameters of the Kiselev form factor model for both $B_c^+ \to J/\psi$ and $B_c^+ \to \psi(2S)$ transitions.

Quantity	J/ψ value	$\psi(2S)$ value
M_{pole}	$4.5\,\text{GeV}/c^2$	$4.5\,\text{GeV}/c^2$
$F_V(0)$	0.11	0.030
$F_A^+(0)$	-0.074	-0.013
$F_A^0(0)$	5.9	1.4
$F_A^-(0)$	0.12	0.044

parameterization, the Kiselev model is:

$$H_\pm(q^2) = \left[\mp F_A^0(0) + F_V(0)\left(2m_{B_c}\,|\vec{p}_X|\,(q^2)\right) \right] k(q^2), \tag{3.23}$$

$$H_0(q^2) = \left[F_A^0(0)\left(\frac{q_{\text{max}}^2 - q^2}{2}\right) + F_A^+(0)\left(2m_{B_c}^2\,|\vec{p}_X|\,(q^2)^2\right) \right] \frac{k(q^2)}{m_\psi\sqrt{q^2}}, \tag{3.24}$$

$$H_S(q^2) = \left[F_A^0(0)\left(\frac{2m_{B_c}^2}{m_{B_c} + m_\psi}\,|\vec{p}_X|\,(q^2)\right) + F_A^-(0)\left(m_{B_c}\,|\vec{p}_X|\,(q^2)q^2\right) \right] \frac{k(q^2)}{m_\psi\sqrt{q^2}}. \tag{3.25}$$

Examining the parameterization, it is clear that $F_A^0(0)$ (the largest of the parameters) is the main parameter controlling the normalization of the Kiselev form factors. The small parameter $F_V^0(0)$ controls the difference between H_+ and H_-, which would otherwise be oppositely signed but have identical magnitude. Similarly, the small $F_A^+(0)$ and $F_A^-(0)$ are responsible for adjusting H_0 and H_S, whose magnitudes would otherwise be fixed relative to $H_\pm$.

The Kiselev model can also be mapped onto a CLN-like parameterization [64], with an Isgur-Wise function $\xi(w)$ and three functions $R_0(w)$, $R_1(w)$, and $R_2(w)$. This allows the extraction of the slope at zero recoil and curvature implied by the model, related to the parameters ρ^2 and σ. The Isgur-Wise function implied by the

Kiselev form factor model is

$$
\begin{aligned}
\frac{\xi(w)}{\xi(1)} &= \frac{2}{w+1} \times \frac{A_1(q^2(w))}{A_1(q^2_{\text{max}})} \\
&= \frac{4}{(1+w)(2+(2\rho^2-1)(w-1))} \\
&= 1 - \rho^2(w-1) + \left(\frac{1-2\rho^2+4\rho^4}{4}\right)(w-1)^2 + \mathcal{O}((w-1)^3),
\end{aligned} \tag{3.26}
$$

where $\xi(1)$ and ρ^2 are parameters related to those of the Kiselev model. Interestingly, the slope parameter ρ^2 is directly related to the Kiselev pole mass:

$$
1 - 2\rho^2 = \frac{4m_{B_c}m_\psi}{q^2_{\text{max}} - M^2_{\text{pole}}}. \tag{3.27}
$$

Given the nominal values of the Kiselev parameters, this slope is $\rho^2 = 4.336$ for the $B_c^+ \to \mathcal{R}(J/\psi)$ transition. Since the Kiselev pole mass was taken to be $4.5\,\text{GeV}/c^2$ by fiat, the slope parameter is not particularly constrained by the Kiselev model (nor, by extension, the nearly identical EFG model, at least in the $B_c^+ \to J/\psi$ case). The rest of the CLN-like form factors are quite simple:

$$
R_0(w) = \frac{RA_0(q^2(w))}{\xi(w)} = R_0(1) \times \frac{1+w}{2} + \left[R^2 - R_0(1)\right]\frac{m_{B_c}m_\psi}{q^2_{\text{max}}}, (w^2 - 1) \tag{3.28}
$$

$$
R_1(w) = \frac{RV(q^2(w))}{\xi(w)} = R_1(1) \times \frac{1+w}{2}, \tag{3.29}
$$

$$
R_2(w) = \frac{RA_2(q^2(w))}{\xi(w)} = R_2(1) \times \frac{1+w}{2}, \tag{3.30}
$$

where $R_{0,1,2}(1)$ are the three remaining parameters of the model and R is the ratio of the charmonium mass to the B_c^+ mass. The values of these parameters are shown in Table 3.2.

The Ebert-Faustov-Galkin (EFG) model of Ref. [63] is derived from a relativistic constituent quark model, following the work of Logunov and Tavkhelidze [72] and Martynenko and Faustov [73]. The meson system (for both B_c^+ and the decay prod-

Table 3.2: CLN-like parameters of the Kiselev form factors for both $B_c^+ \to J/\psi$ and $B_c^+ \to \psi(2S)$ transitions.

Quantity	J/ψ	$\psi(2S)$
ρ^2	4.336	4.336
σ^2	33.77	33.77
R	0.493	0.587
$\xi(1)$	2.548	0.362
$R_0(1)$	0.168	0.248
$R_1(1)$	0.399	0.721
$R_2(1)$	0.268	0.307

ucts J/ψ and $\psi(2S)$) is treated as a bound state in the center-of-mass frame with a relativistic reduced mass μ_r and relative momentum $p_{\text{on-shell}}$ (equal in magnitude and opposite in direction for each quark), both functions of the phenomenological quark masses and total meson mass. In the off-shell relative momentum basis, the wavefunction for the meson M is governed by the Schrödinger like equation

$$\left(\frac{p_{\text{on-shell}}^2}{2\mu_r} - \frac{\mathbf{p}^2}{2\mu_r} \right) \psi_M(p) = \int \frac{d^3\mathbf{q}}{(2\pi)^3} V(\mathbf{p}, \mathbf{q}; M) \psi_M(\mathbf{q}). \qquad (3.31)$$

The quasipotential $V(\mathbf{p}, \mathbf{q})$ is derived from the QCD scattering amplitude, and contains a one-gluon exchange term and a phenomenological linear confining potential; it also contains retardation and one-loop radiative corrections, along with other small adjustments as required. While Eq. 3.31 resembles the Schrödinger equation, it is derived by algebraically manipulating a Green's function relationship, together with some approximations; the formalism somewhat resembles the Bethe-Salpeter equation. The parameters of this model are determined by fits to experimental data (chiefly regarding the charmonium and bottomonium states) and by comparison to HQET [74]. The form factors of the EFT model are extracted from the matrix elements

$$\langle J/\psi(\mathbf{k}) | \bar{c}\gamma_\mu(1 - \gamma^5) b | B_c^+(\mathbf{k}') \rangle = \int \frac{d^3\mathbf{p}\, d^3\mathbf{q}}{(2\pi)^6} \psi_{J/\psi}(\mathbf{p}; \mathbf{k}) \Gamma_\mu(\mathbf{p}, \mathbf{q}) \psi_{B_c^+}(\mathbf{q}; \mathbf{k}'), \quad (3.32)$$

where the wavefunction $\psi_M(\mathbf{p}; \mathbf{k})$ derived from the Schrödinger-like quasipotential equation is boosted out of the center-of-mass-frame by the momentum $\mathbf{k}$ and Γ_μ is a two-particle vertex function containing the $V - A$ coupling term and corrections from the quasipotential (essentially subleading Feynman diagrams involving a single scattering between the quarks before or after the $V - A$ vertex). The EFG model uses the standard form factor parameterization given in Eq. 3.12. Unfortunately, however, the paper does not present a simple analytic approximation for the form factors, instead only providing plots and an unwieldy integral formula that are impractical to fully implement. This means that the only obvious way to implement the model is by scanning and numerically fitting their plots[2].

For $B_c^+ \to J/\psi$, the biggest difference between the Kiselev and EFG form factors are their magnitudes, which only control the partial widths of the decays and therefore their different branching fractions. Normalizing away their magnitudes, the form factors are quite similar. Fig. 3.4 shows the differential decay rates vs. q^2 for $B_c^+ \to J/\psi \mu^+ \nu_\mu$ and $B_c^+ \to J/\psi \tau^+ \nu_\tau$. There is encouraging agreement between the two models' q^2 distributions. Larger differences are found in the angular distributions predicted by the models. Fig. 3.5 shows the differential decay rates vs. E_ℓ^* for $B_c^+ \to J/\psi \mu^+ \nu_\mu$ and $B_c^+ \to J/\psi \tau^+ \nu_\tau$, which is affected by the predicted angular distribution. There is still good agreement between the Kiselev and EFG models, but they are noticeably different, with the Kiselev parameter tending to predict smaller values of E_ℓ^*. Additional plots of the individual Kiselev and EFG form factors are shown in Figs. A.1 and A.2 in Appendix A.

Kiselev and EFG also give form factor models for the $B_c^+ \to \psi(2S)$ transition,

---

[2] A typo in Fig. 8 of Ref. [63] implies that the maximum q^2 available in $B_c^+ \to \psi(2S)$ is 5.8, while in reality it is 6.7, a mistake likely due to using 3.86 for the $\psi(2S)$ mass in rather than the correct value 3.686. This was not caught and was absorbed by LHCb's implementation of the EFG model. Strange behavior of the form factors in the region $5.8\,\mathrm{GeV}^2 < q^2 < 6.7\,\mathrm{GeV}^2$ (particular evident in A_2 and H_0) is attributable to this bug. The effect of the mistake on the q^2 spectrum was noticeable, but is much smaller than the difference between the true EFG model and the Kiselev model.

(a) μ mode   (b) τ mode
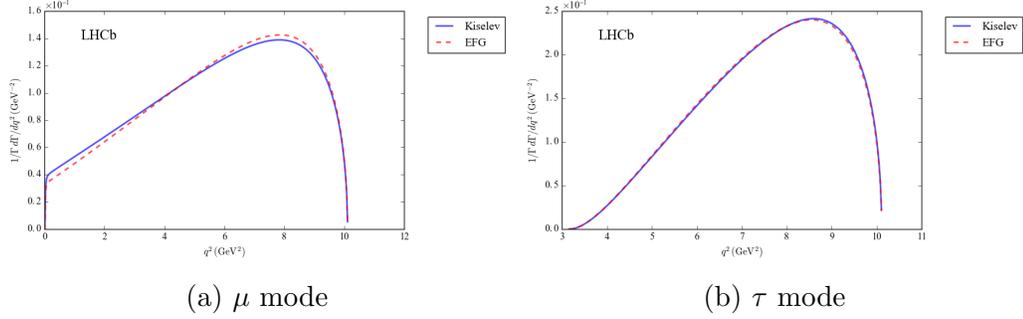
Figure 3.4: The Kiselev and EFG q^2 distribution for $B_c^+ \to J/\psi\, \ell^+ \nu_\ell$ are nearly identical, for both the muonic and tauonic channels.

(a) μ mode   (b) τ mode
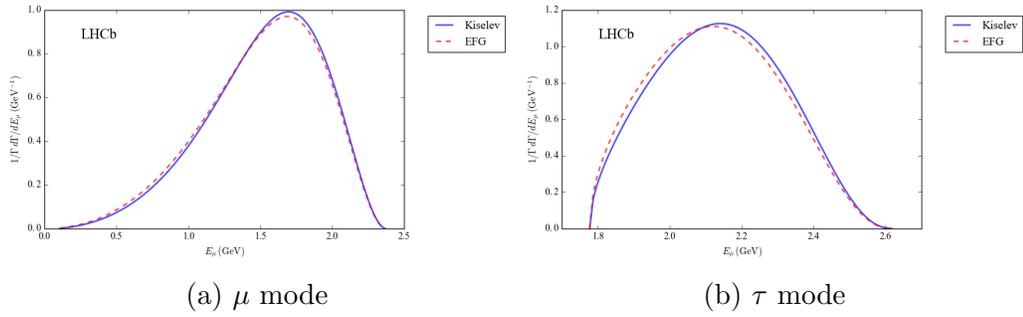
Figure 3.5: The Kiselev and EFG E_ℓ^* distribution for $B_c^+ \to J/\psi\, \ell^+ \nu_\ell$ are also very similar; the Kiselev form factors predict a slightly softer E_ℓ^*.

|(a) μ mode | (b) τ mode |
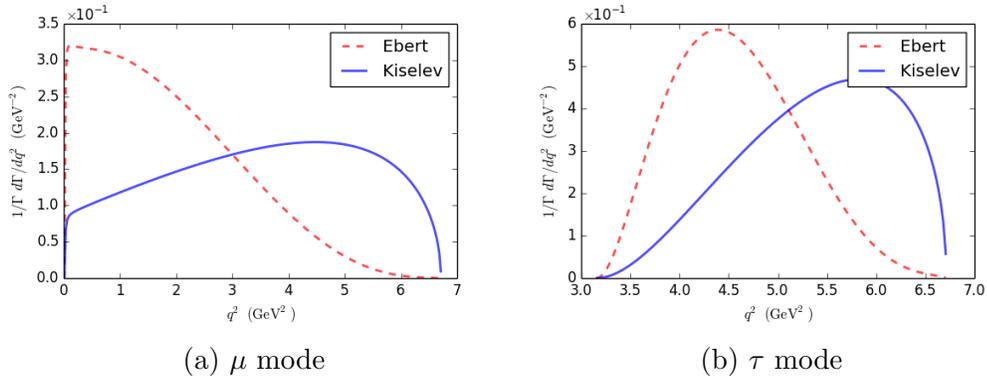
Figure 3.6: The Kiselev and EFG q^2 distribution for $B_c^+ \to \psi(2S)\ell^+\nu_\ell$ disagree completely.

following the same conventions used for the J/ψ transition. Plots of the individual Kiselev and EFG form factors and q^2 distributions are shown in Figs. 3.6, A.3 and A.4. While the two models are in good agreement for the $B_c^+ \to J/\psi$ transition, the corresponding models of $B_c^+ \to \psi(2S)$ transitions differ dramatically. The Kiselev model is parametrically identical to that for the J/ψ final state, with only changes in the values of the numerical parameters[3]. On the other hand, one important consequence of the EFG model is that the $\psi(2S)$ meson, being an $n = 2$ radially excited charmonium state, has a node in its wavefunction, causing a dramatic change in the q^2 spectrum of the $B_c^+ \to \psi(2S)$ decays in comparison to the $B_c^+ \to J/\psi$ decays.

The simulation for this analysis was generated using the EFG model for both $B_c^+ \to J/\psi$ and $B_c^+ \to \psi(2S)$. However, the form factors for the former are modified using corrections derived from data following the BCL parametrization. This parameterization is described below in Section 3.2.2, and the data-driven correction is described in Section 4.3.1. The form factor model for $B_c^+ \to \psi(2S)$ is given a small amount of freedom in the fit, corresponding to variations in the CLN-like ρ^2 param-

---

[3]This is, in fact, an assumption of the model: the transition to the excited 2S state is difficult to extract from the QCD sum rules, and so instead finite energy QCD sum rules and the "principal of stability" are used to determine scale factors to transform the $B_c^+ \to J/\psi$ form factors into $B_c^+ \to \psi(2S)$ form factors

eter representing the slope of the Isgur-Wise function, using a method described in Section 4.2.1

3.2.2 BCL parameterization

An increasingly popular set of parameterizations for semileptonic form factors are known as z-expansions. These parameterizations are defined via the aid of the conformal mapping

$$z(q^2) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_-}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_-}}, \tag{3.33}$$

where $t_\pm = (M_{B_c} \pm M_{J/\psi})^2$. Figure 3.7 representing the mapping of the q^2 plane into the z plane. There are several useful properties of this mapping related to the analytic properties of form factors. In the z plane, this branch cut is mapped to the unit circle; the path just below the branch to the upper semicircle and the path just above the branch to the lower semicircle. The threshold value t_+ is mapped to -1 and t_- to 0, and the regions with $\Re(q^2)$ in-between or outside $t_\pm$ are mapped to various regions within the unit circle, as shown in Fig. 3.7. In a z-expansion, a form factor $F(q^2)$ is expanded into

$$F(q^2) = P(q^2) \sum_{k=0}^{K} b_k z(q^2)^k. \tag{3.34}$$

The form of the initial term $P(q^2)$ depends on the particular z-expansion framework.

One particular such parameterization has been proposed by Bourrely, Caprini, and Lellouch (BCL) [75] and recently used for the lattice computation of $B \to D\ell\nu$ form factors [76]. The motivation of this parameterization is to naturally encode constraints on semileptonic decay form factors arising from general physical principles. For building a fit model, the principal feature of the BCL expansion is that $P(q^2)$ is a pole mass term $1/(1 - q^2/M_{\text{pole}}^2)$, for a pole mass M_{pole} that corresponds to an excited resonance whose quantum numbers correspond to the

form factor. For $A_0(q^2)$, whose quantum numbers are 0^-, the pole mass is the B_c^+ mass itself. For $A_1(q^2)$ and $A_2(q^2)$, whose quantum numbers are 1^+, the B_c^+ resonance has not been measured but is predicted to have a mass of approximately $6.73\,\mathrm{GeV}$. For $V(q^2)$ (1^-) the predicted mass of the resonance is $6.33\,\mathrm{GeV}$ [74].

A form factor $F(q^2)$ is required to be analytic in the complex q^2 plane, save for a branch cut along the real axis for q^2 greater than the threshold value t_+. Below this threshold, and on the real q^2 axis, the form factor must be real and analytic, save for a few possible isolated poles. Above this threshold, the scattering process $\ell\nu \to W^* \to B_c^+\, J/\psi$, related by crossing symmetry to $B_c^+ \to J/\psi\, \ell^+\nu_\ell$, is kinematically possible. Its decay cross section is a function of q^2 and the W^* polarization. For the scalar polarization, the decay must be p-wave but above the threshold picks up an imaginary value. For $q^2 > t_+$, there are extra constraints on some of the form factors due to angular momentum conservation. The optical theorem requires that the imaginary part of the form factor scales as $\Im(F(q^2)) \sim (q^2 - t_+)^{3/2}$. The BCL parameterization uses this constraint to reduce the number of coefficients from $K+1$ to K. The parameterization is only completely specified for scalar-to-scalar transitions (*i.e.*, $B \to D$), and it is not completely clear how to apply this result to a scalar-to-vector transition. Consequently, we do not use this aspect of the BCL parameterization.

The four form factors $A_{0,1,2}(q^2)$ and $V(q^2)$ can be represented by BCL expansions to linear order in z. The pole corresponding to the $A_0(q^2)$ form factor is that from a $\bar{b}c$ state with $J^P = 0^-$, which is just the B_c^+ meson itself. The $A_1(q^2)$ and $A_2(q^2)$ form factors have a pole contribution from $J^p = 1^+$ states, the lowest lying of which is predicted to have a heavy mass near $6.73\,\mathrm{GeV}$ [57, 63]. Finally, the pole mass for $V(q^2)$ is a $J^P = 1^-$ state, with a predicted pole mass of about $6.332\,\mathrm{GeV}$. Each form factor has an zero-th parameter b_0 and a linear parameter b_1. The Ebert model can be approximated with this BCL expansion. The b_0 and b_1 parameters of

(a) q^2 plane (b) z plane
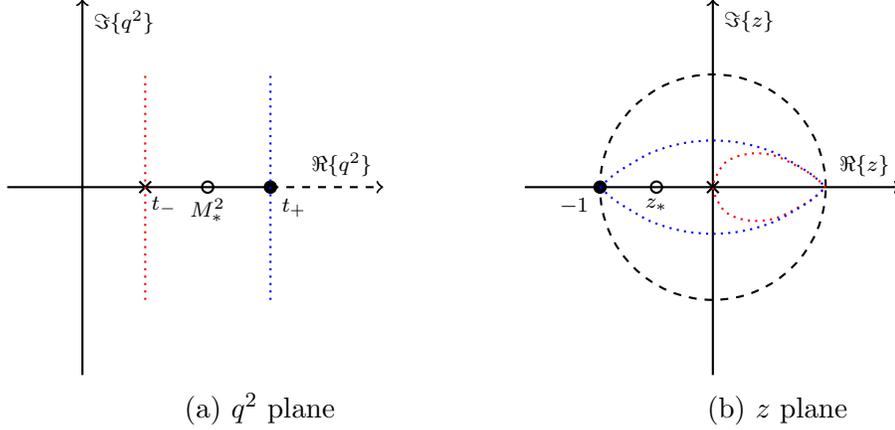
Figure 3.7: Analytic properties of semileptonic form factors in the q^2 and z planes. The branch cuts are represented by dashed lines and the threshold points by solid circles. The pole is represented by an open circle, and the reference point $q^2 = t_-$ by a cross.

this approximation are shown in Table 4.4.

3.3 B_c^+ rest frame approximation

We wish to measure the momentum vector of the B_c^+ candidate. With this, we can boost into the rest frame, in which there are several variables which discriminate $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ and $B_c^+ \to J/\psi\,\tau^+\nu_\tau$, $\tau \to \mu\nu\nu$ decays from each other and from background. At LHCb, the observables in the semileptonic B_c^+ decays that might allow a measurement of the B_c^+ momentum are the flight vector of the B_c^+ meson and the energy and momentum of the visible part of the final state, $Y = J/\psi\,\mu$. The flight vector, which is the physical vector pointing from the B_c^+'s production vertex (at a PV) to its decay vertex, is measured by information from LHCb's tracking system, especially the VELO. The decay vertex is determined by a fit to the three muon tracks, and the B_c^+ candidate is matched to its best PV (the one where the χ^2 of a fit requiring the B_c^+ to come from the PV is minimized).

These observables and the known B_c^+ mass do not uniquely determine the four-momentum of the B_c^+ candidate. Consider the boosted frame whose z axis is

aligned in the B_c^+'s lab flight direction and in which the momentum of the $Y = J/\psi\,\mu$ (*i.e.*, the visible subsystem of the final state) is perpendicular to that of the B_c^+. For the direct semimuonic decay, where there is one massless neutrino, there are two solutions for the B_c^+ momentum in this frame, one pointing parallel to the z axis and the other antiparallel. For the semitauonic decay, whose three-neutrino subsystem can have any invariant mass in the range $[0, m_{B_c} - m_{J/\psi} - m_\mu]$, there is a continuum of possible solutions. Instead, the B_c^+ momentum must be estimated from the available observables. This can be done with a technique developed originally to study the $B \to D^{(*)}\tau\nu$ decays at LHCb [20]. Consider the decay of the B_c^+ into a visible system Y and invisible neutrino system, as shown in Fig. 3.8 in both the B_c^+ rest frame (or the center of mass (CM) frame) and lab frame. ϑ_{CM} is the angle in the rest frame between the visible momentum and the B_c^+ flight direction, and $p_{Y,\mathrm{CM}}$ is the four-momentum of the Y system in the rest frame. The four-momenta of the B_c^+ and Y system in the lab frame, meanwhile, are denoted simply p_{B_c} and p_Y, and ϑ is the angle between them. The Lorentz transformation between these two frames shows that

$$p_{B_c} = \frac{m_{B_c}}{E_{Y,\mathrm{CM}}} p_Y^{\parallel} - E_{B_c}\beta_{Y,\mathrm{CM}}\cos\vartheta_{\mathrm{CM}}, \tag{3.35a}$$

$$p_Y^{\perp} = p_{Y,\mathrm{CM}}\sin\vartheta_{\mathrm{CM}}, \tag{3.35b}$$

where the parallel and perpendicular components of p_Y are relative to the flight direction of the B_c^+ (*i.e.*, $p_Y\cos\vartheta$ and $p_Y\sin\vartheta$, respectively). Conservation of energy and momentum in the decay limits the speed $\beta_{Y,\mathrm{CM}}$ to be less than

$$\beta_{\max} = \frac{m_{B_c}^2 - (m_\psi + M_\mu)^2}{m_{B_c}^2 + (m_\psi + M_\mu)^2} \approx 0.587.$$

The distribution of $\beta_{Y,\mathrm{CM}}$ peaks between 0 and $\beta_{\max}$, and dies towards $\beta_{Y,\mathrm{CM}} = 0$

(a) B_c^+ rest frame   (b) Lab frame
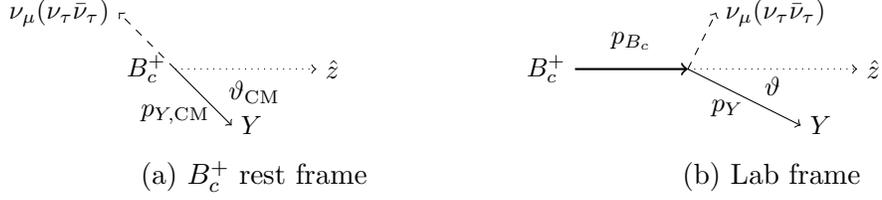
Figure 3.8: Semileptonic B_c^+ decay into a visible $J/\psi\,\mu$ system and invisible neutrino system

and $\beta_{Y,\mathrm{CM}} = \beta_{\max}$ where there is less available phase space.

Since the B_c^+ meson is spinless, the Y direction in the rest frame is isotropically distributed and uncorrelated with both $\beta_{Y,\mathrm{CM}}$ and the lab energy of the B_c^+ (ignoring acceptance and selection effects). Therefore, $\cos\vartheta_{\mathrm{CM}}$ will be zero on average, and

$$\frac{m_{B_c}}{E_{Y,\mathrm{CM}}}\, p_Y^{\parallel}$$

would be an unbiased estimator of p_{B_c} if $E_{Y,\mathrm{CM}}$ were a measured quantity. Note that taking the expectation $\cos\vartheta_{\mathrm{CM}} \to 0$ is equivalent to assuming that $\vartheta_{\mathrm{CM}} = \pi/2$. With this assumption, Eq. 3.35 allows an estimate of the B_c^+ momentum from the measured momentum and energy of the Y and B_c^+ flight vector:

$$p_{\mathrm{est}} = p_{B_c}\big|_{\vartheta_{\mathrm{CM}}=\frac{\pi}{2}} = \frac{m_{B_c} p_Y^{\parallel}}{\sqrt{M_Y^2 + \left(p_Y^{\perp}\right)^2}}. \tag{3.36}$$

It follows from Eq. 3.35 that the fractional error on this estimate is

$$\frac{p_{\mathrm{est}} - p_{B_c}}{p_{B_c}} = \frac{\beta_{Y,\mathrm{CM}}}{\beta_{B_c}}\cos\vartheta_{\mathrm{CM}} + A\left(1 + \frac{\beta_{Y,\mathrm{CM}}}{\beta_{B_c}}\cos\vartheta_{\mathrm{CM}}\right), \tag{3.37}$$

where the quantity A, which is small at typical values of $\beta_{Y,\mathrm{CM}}$, is given by

$$A = \frac{1}{\sqrt{1 - \beta_{Y,\mathrm{CM}}^2 \cos^2\vartheta_{\mathrm{CM}}}} - 1 \approx \frac{1}{2}\beta_{Y,\mathrm{CM}}^2 \cos^2\vartheta_{\mathrm{CM}} + \mathcal{O}\left(\beta_{Y,\mathrm{CM}}^4\right). \tag{3.38}$$

63
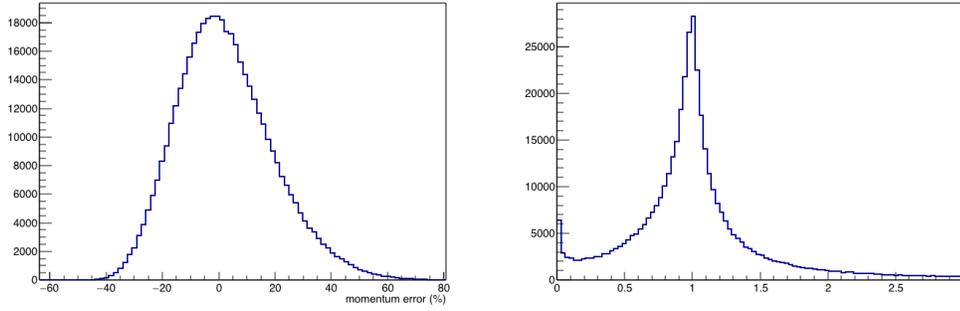
(a) Fractional error on the B_c^+ momentum

(b) Ratio of squared kinematic error (from Eq. (3.37)) to the total squared error

Figure 3.9: The B_c^+ momentum resolution, determined in a Monte Carlo simulation and including effects from detector resolution, is approximately 17%, and is dominated by the kinematic approximation rather than by detector resolution effects.

The first term on the right-hand side of Eq. 3.37 is zero on average and is roughly bounded by $\beta_{\max} \approx 0.587$, since $\beta_{B_c} \sim 1$ for the highly boosted B_c^+ mesons at the LHC. The second term is non-zero on average, producing a positive bias in the estimated momentum. However, it is a higher order correction in the relatively small parameter $\beta_{Y,\mathrm{CM}}$.

This analysis of the estimation method does not take into account detector resolution or acceptance effects. However, simulation verifies that Eq. (3.37) represents the fractional error well. Fig. 3.9 plots the distribution of the fractional error of the estimated B_c^+ momentum. This error, which includes resolution effects as well as kinematic effects, is dominated by the kinematic terms in Eq. 3.37. In fact, while the standard deviation of the complete fractional error is $\sigma(p_{B_c})/p_{B_c} = 17.09\%$, the standard deviation of the dominant kinematic terms in Eq. (3.37) is actually marginally larger — $\sigma(p_{B_c})/p_{B_c} = 17.22\%$. This small difference indicates that detector effects contribute negligibly to the momentum resolution.

The momentum estimate given by Eq. (3.36) is nearly equivalent to a simpler

estimation technique:

$$(p_{\text{est}})_z = \frac{m_{B_c}}{M_Y} (p_Y)_z.$$

(3.39)

The x- and y-components of the momentum are fixed by the B_c^+ flight direction. Since the B_c^+ flight direction is close to parallel to the z axis in LHCb's acceptance, and $p_Y^{\perp}$ is small, this is very similar to rescaling $p_Y^{\parallel}$. Simulation shows that the difference in resolution between Eq. (3.36) and Eq. (3.39) is negligible. For the record, this analysis uses Eq. (3.36).

Chapter 4: Measurement of $\mathcal{R}(J/\psi)$, the ratio of $\mathcal{B}(B_c^+ \to J/\psi\,\tau^+\nu_\tau)$

to $\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu_\mu)$

This analysis measures the value $\mathcal{R}(J/\psi)$ through an analysis of data corresponding to $3\,\mathrm{fb}^{-1}$ collected by LHCb in 2011 and 2012. The strategy of the analysis is to select from the data a sample of events containing $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ and $B_c^+ \to J/\psi\,\tau^+\nu_\tau$ decays (as well as several backgrounds) and determine their relative yield by fitting the distribution of the m_{miss}^2, q^2, E_μ^*, and decay time using templates derived from simulation and data control samples. The method results of this analysis, published in Ref. [77], are described in greater detail here.

Candidates from the signal decay mode $B_c^+ \to J/\psi\,\tau^+\nu_\tau$ and normalization decay mode $B_c^+ \to J/\psi\,\mu^+\nu_\mu$, where $J/\psi \to \mu^+\mu^-$ and $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$, have the same trimuon final state. A sample of trimuon candidates containing these two modes, as well as a number of backgrounds, is selected from the LHCb data, as described in Section 4.1.

For these selected candidates, the technique described in Section 3.3 is used to to approximate the B_c^+ momentum; Fig. 3.3 shows a schematic diagram of a $B_c^+ \to J/\psi\,\ell^+\nu_\ell$ decay in the B_c^+ in the center-of-mass and lab frames. Using this momentum p_{B_c}, several useful variables can be computed that discriminate between signal and background. The *missing mass squared*, m_{miss}^2, is the norm of the four-vector difference between the approximated B_c^+ four-momentum and the
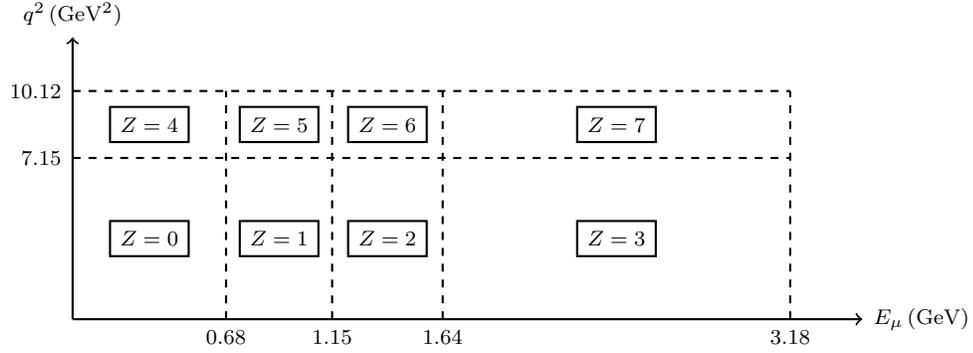
Figure 4.1: Definition of the categorical variable Z in the q^2, E_μ plane.

four-momentum of its visible decay products, or

$$m^2_{\text{miss}} = \left(p_{B_c^+} - p_{J/\psi} - p_\mu\right)^2. \tag{4.1}$$

This represents the invariant mass of the set of decay products of the B_c^+ that are not included in the candidate. For the normalization decay, this is a single neutrino, while for the signal decay it is a system of three neutrinos, and in general various backgrounds may have decay products missing from the final state (or may not correspond to the decay of B_c^+, or any other, meson). Another useful variable is the *squared four-momentum transfer* q^2, defined as $(p_{B_c^+} - p_{J/\psi})^2$, which is equal to the invariant mass squared of the off-shell $W^\pm$; this variable also plays an important role in the $B_c^+ \to J/\psi\,\ell\nu$ form factors, as described in Section 3.2. The *helicity angle* θ between the lepton momentum and the axis anti-parallel to the J/ψ momentum in the $W^\pm$ rest frame is also related to the form factors. Another variable is the energy of the muon in the B_c^+ rest frame, or *muon rest-frame energy*, denoted E_μ^*. These four variables are not algebraically independent: there are three underlying degrees of freedom, and E_μ^* can be determined as a function of m^2_{miss}, q^2, and θ, as described in Appendix B. Moreover, even an algebraically independent set of these variables (in particular, m^2_{miss}, q^2, and E_μ^*, the set used in the fit) are not statistically independent, but instead exhibit non-trivial correlations.

The parameter $\mathcal{R}(J/\psi)$ is measured in a three-dimensional binned maximum likelihood fit. The first variable is the missing mass squared of the B_c^+ candidate, assuming a transition to $J/\psi\,\mu X$, with m_{miss}^2 bins in the range $-5\,(\mathrm{GeV})^2$ to $10\,(\mathrm{GeV})^2$. The second is the decay time of the B_c^+ candidate, with 5 bins in the range $0.3\,\mathrm{ps}$ to $2.18\,\mathrm{ps}$. The third is a quasi-categorical variable, denoted Z for convenience, with 8 values covering the allowed space of E_μ^* and q^2 values for the decay.[1] The first four values correspond to E_μ^* between the bin edges $[0, 0.68, 1.15, 1.64, \infty)\,\mathrm{GeV}$ for $q^2 < 7.15\,(\mathrm{GeV})^2$. The next four values correspond to E_μ^* in the same regions, but for $q^2 \geq 7.15\,(\mathrm{GeV})^2$. This is represented graphically in Fig. 4.1. This set of variables provides good discrimination between the signal, normalization, and backgrounds. The categorical variable Z, and to a lesser extent the missing mass, is also sensitive to the $B_c^+ \to J/\psi$ form factor parameters. Other details of the fit model itself are described in Section 4.2.

The three-dimensional distributions of the various components of the sample in the fit variables have no simple analytic descriptions. Therefore, histogram templates are created for the fit for the signal, normalization, and each background contribution, and are constructed from data, Monte Carlo, and data-driven control samples. The two principal components of the fit are templates for the signal mode $B_c^+ \to J/\psi\,\tau^+\nu_\tau$ and the normalization mode $B_c^+ \to J/\psi\,\mu^+\nu_\mu$. The modeling of these components is described in Section 4.3.

Other templates represent the important background modes that have sufficiently similar signatures to pass the selection. These are briefly listed here, summarized in Table 4.1, and discussed in more detail throughout this chapter. The first of these are the feed-down backgrounds, described in Section 4.4, in which a B_c^+ decays semileptonically to a heavy charmonium state that in turn decays down to a J/ψ. Other backgrounds coming from B_c^+ decays are $B_c^+ \to J/\psi\,DX$ decays in

---

[1]This procedure is equivalent to a four-dimensional fit with two bins in q^2 and four in E_μ^*, but this cannot be done directly because of limitations in ROOT.

which the D decays leptonically or semileptonically.[2]; these are described in more detail in Section 4.5.

The largest background arises from inclusive decays of B^+, B^0, and B_s mesons to J/ψ + hadrons in which a pion or kaon (or, less frequently, proton or electron) from the decay is misidentified as a muon. Though the muon misidentification rate at LHCb is low, the cross sections of the B mesons are very large relative to that of the B_c^+. The primary signature of this background is that the B_c^+ candidates from this background source follow a B-like lifetime distribution and peak high in missing mass and q^2. The techniques involved in modeling this background, using a data-driven control sample made by combining J/ψ mesons with tracks fail to penetrate the muon chamber, is described in Section 4.8

Additionally, there are several other non-negligible backgrounds. The largest of these is combinatorial background, wherein a J/ψ from a b-hadron decay is combined with true muon from the rest of the event. Since b-hadrons are typically produced in pairs, a b-hadron or its decay products usually has a partner b-hadron from the "opposite-side" of the diagram of the collision; because of the PV separation requirements, the muon in the combinatorial background generally originates from the decay of the opposite-side b-hadron. This background is modeled by a Monte Carlo cocktail of $B_{u,d,s} \rightarrow J/\psi\,X$ decays, described in detail in Section 4.6.

There is also a small background from combinatorial J/ψ candidates. This is modeled using data from the $m[J/\psi]$ sidebands, as described in Section 4.7. This assumes that the fit variables are uncorrelated with the J/ψ mass, which is true to a good approximation across the J/ψ mass window (which is only a few percent of the total J/ψ mass).

Many of the above backgrounds are produced using simulation. A number of corrections are required to improve the agreement between simulation and data.

---

[2]Here and in the following, D refers to any of D^+, D^0, or D_s^+.

Table 4.1: Characteristics of major signal and physics background channels

Mode	Topology	Signature
Signal	$B_c^+ \to J/\psi\, \tau^+ \nu_\tau$ $\Rightarrow J/\psi \to \mu\mu$ $\Rightarrow \tau \to \mu\nu\nu$	J/ψ dimuon, unpaired μ Large missing mass (3ν) Short B_c^+ lifetime
Normalization	$B_c^+ \to J/\psi\, \mu^+ \nu_\mu$ $\Rightarrow J/\psi \to \mu\mu$	J/ψ dimuon, unpaired μ Smaller missing mass (1ν) Short B_c^+ lifetime
feed-down	$B_c^+ \to (c\bar{c})^* \mu\nu$ $\Rightarrow (c\bar{c})^* \to J/\psi\, X$	J/ψ dimuon, unpaired μ Large missing mass $(1\nu + X)$ Short B_c^+ lifetime
$J/\psi\, DX$	$B_c^+ \to (c\bar{c})DX$ $\Rightarrow (c\bar{c}) \to J/\psi\,(X)$ $\Rightarrow D \to (\mu,\tau)\nu(Y)$	J/ψ dimuon, unpaired μ Small to large missing mass $(1\nu, (X), (Y))$ Short B_c^+ lifetime
Combinatorial	$B \to J/\psi\, X,\ B \to \mu X$ $\Rightarrow J/\psi \to \mu\mu$	J/ψ dimuon, 1 real OS μ Broad missing mass distribution Mimics B_c^+ lifetime
Mis-ID	$B \to J/\psi\, \mu_{(fake)}X$ $\Rightarrow J/\psi \to \mu\mu$	J/ψ dimuon, 1 fake μ Large missing mass (X) Long B lifetime

These are described in Section 4.9.

4.1   Selection of $B_c^+ \to J/\psi\, \mu^+ X$ candidates

This analysis uses the full LHCb dataset from Run 1, which occurred during the years 2011 and 2012., consisting of runs in 2011 at $\sqrt{s} = 7\,\mathrm{TeV}$ and 2012 at $\sqrt{s} = 8\,\mathrm{TeV}$. The 2012 data corresponds roughly $2\,\mathrm{fb}^{-1}$ collected at $\sqrt{s} = 8\,\mathrm{TeV}$, while the 2011 data corresponds to roughly $1\,\mathrm{fb}^{-1}$ collected at $\sqrt{s} = 7\,\mathrm{TeV}$.

Candidate $B_c^+ \to J/\psi\,(\to \mu^+\mu^-)\,\mu^+$ decays are reconstructed from a subset of events flagged for the presence of detached J/ψ mesons during the stripping of LHCb's data (see Section 2.2.6). This stripped data requires the presence of two oppositely charged μ candidates with $p_T > 550\,\mathrm{MeV}$. These tracks should not originate at the primary vertex (PV) where the proton collision occurs. A typical discriminant for this is the impact parameter (IP) χ^2, defined as the increase in the primary

vertex fit χ^2 when this track is included[3]; this IP χ^2 is required to be greater than 4 for each muon with respect to any PV. The distance of closest approach (DOCA) of the two muons is determined in a fit, and must have a $\chi^2 < 30$, and a vertex is also fit from the track parameters, which must have $\chi^2 < 20$; the vertex location must also be significantly separated from the PV, by three standard deviations given their uncertainties. The measured invariant mass of the J/ψ candidate must be within $55 \, \mathrm{MeV}/c^2$ of its true mass, and the candidate must have a significant decay length (more than 5 standard deviations away from the nearest PV given its tracking uncertainties) to remove prompt background It is additionally required that this J/ψ candidate be the signature that triggered the LHCb detector to save the event, at the hardware and both software (HLT1 and HLT2) levels; this is usually the case, but is important to enforce strictly to remove effects on the measured efficiency ratio.

From this sample of events, the J/ψ meson candidates are paired with muon tracks to form B_c^+ candidates, with a set of selection requirements summarized in Table 4.2). Some of these are further individual cuts on the allowed J/ψ and unpaired muon candidates: the J/ψ is required to have a high p_T ($> 2 \, \mathrm{GeV}/c$); the unpaired muon μ_B must be moderately detached from its PV (with an IP χ^2 greater than 4.8) and have p_T greater than 750 MeV/c. This muon candidate must also pass strict PID requirements. The resulting vertex fit must have a χ^2 significance less than 75, while the fit determining the J/ψ and unpaired muon distance of closet approach (DOCA) must be less than $0.15 \, \mathrm{mm}$ and have a χ^2 significance less than 30. A selection is made to reduce partially reconstructed backgrounds using a boosted decision tree examining all the tracks in the event, described in Section 4.1.1.

The B_c^+ candidate formed from combining the J/ψ and unpaired μ^- must

---

[3]The LHCb detector fits various vertices using a Kalman-filter algorithm [78]. The figure of merit for this fit is denoted $\chi^2 = r^T V^{-1} r$, where r are the residuals of the measured data in the fit model and V is the covariance matrix of the measured data.

Table 4.2: Main selection requirements. If more than one candidate is present in the event after this selection, one is chosen randomly.

Object	Selection
Event	number of identified tracks < 600
$\mu^{\pm}$	$p > 3\,\text{GeV}/c$, $p_T > 550\,\text{MeV}$
—	IP $\chi^2 > 4$
—	moderate PID criteria
J/ψ	$p_T > 2\,\text{GeV}/c$
—	DOCA $\chi^2 < 30$
—	vertex fit $\chi^2 < 20$
—	mass within $55\,\text{MeV}$ of $m_{J/\psi}$ ($3095\,\text{MeV}$)
unpaired μ^+	$p > 3\,\text{GeV}/c$, $p_T > 750$ MeV/c
—	$\eta < 3.6$ or $p \geq 15\,\text{GeV}$
—	IP $\chi^2 > 4.8$
—	strict PID criteria
$J/\psi + \mu^+$	vertex $\chi^2 < 75$
—	DOCA $< 0.15\,\text{mm}$
—	DOCA $\chi^2 < 30$
B_c^+	invariant mass less than the B_c^+ mass
—	$\cos(\theta_{XY})$ between J/ψ, $\mu^+ > -0.8$
—	maximum isolation BDT < 0.2
—	missing mass squared in range -5–$10\,\text{GeV}^2/c^4$
—	decay time in range 0.3–$2.18\,\text{ps}$
unpaired $\mu^+ + \mu^-$	mass not within $50\,\text{MeV}$ of $m_{J/\psi}$
unpaired $\mu^+ + \mu^+$	$\cos(\theta_{\text{lab}}) > 1 - e^{-8} \approx 0.9997$

have an invariant mass less than the B_c^+ mass, since energy is lost to the neutrino system for the signal and normalization decays. The approximation method is used to calculate the rest frame variables, which are required to be in the ranges used in the fit; in particular, -5–$10\,\text{GeV}^2/c^4$ for the missing mass squared and 0.3–$2.18\,\text{ps}$ for the decay time.

A few selection requirements targeted specific backgrounds without significantly affecting the signal efficiency. Two cuts target combinatorial background by limiting the DOCA of the J/ψ and μ^+ and requiring that the J/ψ and μ^+ not travel in nearly opposite directions in the XY plane (characteristic of decays where they are produced by boosted opposite-sign b-hadrons). Two small backgrounds are pos-

sible because the trimuon final state can contain a double counted muon track or two swapped same-sign muons. In the first of these, one of the muon tracks from the J/ψ is accidentally identified as two separate particles by LHCb's reconstruction software, and together these three tracks form a candidate. This background is excluded by vetoing candidates where the angle between the unpaired muon and the same-sign muon in the lab frame is sufficiently close to zero. The latter background occurs in an ambiguous region of phase space where both pairs of opposite-sign muons have an invariant mass in the J/ψ window, and when the wrong pair has been chosen as the J/ψ candidate; this is vetoed by a veto in a window around the J/ψ for these combinations. A veto is made for candidates whose unpaired muon has $\eta \geq 3.6$ and $p \leq 15$ GeV/c. This criterion is over 99% efficient for signal and normalization decays, with respect to the rest of the selection, and reduces the amount of mis-ID background while significantly improving overall data–simulation agreement.

In events with at least one candidate passing the final selection, there is such a single candidate greater than 98% of the time. The primary source of multiple candidates in an event is a single J/ψ meson combined with several bachelor muon candidates. In the rare events containing multiple candidates, a single candidate is randomly retained.

4.1.1 Isolation BDT

Many of the backgrounds in this analysis (particularly the mis-ID and double charm backgrounds) are partially reconstructed from b-hadron decays that produce more than three charged tracks. The presence of extra charged tracks that are associated with the B_c^+ candidate's vertex differentiates these backgrounds from the signal and normalization modes.

This analysis makes use of the isolation boosted decision tree (BDT) devel-
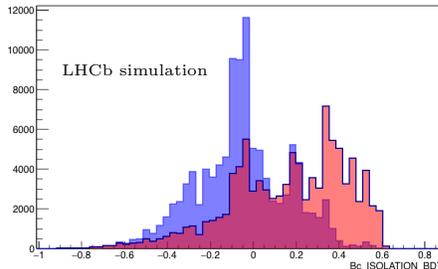
Figure 4.2: The distributions of the isolation BDT score for $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ (blue) and for mis-identified background template, before applying the PID selection (red)

oped in Ref. [20]. This BDT is trained using simulation to distinguish associated tracks, *i.e.*, those originating from the same B meson as the candidate tracks, from unassociated tracks, *e.g.*, those originating from PVs or other secondary decays. Its input variables include the transverse momentum of the track, its IP χ^2 with respect to both the PV and the $J/\psi\,\mu^+$ vertex, the angle between its flight direction and the $J/\psi\,\mu^+$ momentum, and the significances of the separation of the $J/\psi\,\mu^+$ and $J/\psi\,\mu^+$ + track vertices from the associated PV. The output of the isolation BDT is used to rank every track in the detector. Figure 4.2 shows the distributions of the maximum BDT score for the $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ normalization mode and the misidentified background. The selection requires the BDT score of the most associated-like track (*i.e.*, the highest BDT score of all the tracks) to be less than 0.2.

4.1.2 Efficiency ratio

The raw ratio of the number of observed $B_c^+ \to J/\psi\,\tau^+\nu_\tau$ events to $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ events that is measured by the fit does not represent the actual ratio of branching fractions. This ratio must be corrected by the overall selection efficiency for the two decay modes and the $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$ branching fraction. This selection efficiency includes contribution from the detector acceptance efficiency (measured at the generator level and labelled GEN); the reconstruction and preliminary selection efficiency (REC); the efficiency of the final selection proper, without isolation or

bachelor muon PID cuts (SEL); the efficiency of the isolation cut (ISO); the efficiency of the PID requirement for the bachelor muon (PID); and the L0 hardware and first level HLT1 software triggers (both labelled L0) and second level HLT2 software trigger (HLT2). The overall ratio of efficiencies for the two modes is broken down into

$$\frac{\epsilon_\tau}{\epsilon_\mu} = \frac{\epsilon_\tau^{\text{GEN}}}{\epsilon_\mu^{\text{GEN}}} \times \frac{\epsilon_\tau^{\text{REC}}}{\epsilon_\mu^{\text{REC}}} \times \frac{\epsilon_\tau^{\text{SEL}}}{\epsilon_\mu^{\text{SEL}}} \times \frac{\epsilon_\tau^{\text{ISO}}}{\epsilon_\mu^{\text{ISO}}} \times \frac{\epsilon_\tau^{\text{PID}}}{\epsilon_\mu^{\text{PID}}} \times \frac{\epsilon_\tau^{\text{L0}}}{\epsilon_\mu^{\text{L0}}} \times \frac{\epsilon_\tau^{\text{HLT2}}}{\epsilon_\mu^{\text{HLT2}}}. \tag{4.2}$$

The relative efficiency ratios are evaluated using the simulation samples for the different years of data taking and magnet polarities. There is good consistency between the various samples. Most of the relative efficiency ratios are found to be very close to 100%. The exceptions are the reconstruction efficiency, selection efficiency, and PID efficiency, which are found to be roughly 74%, 72%, and 93%, respectively. The total efficiency ratio is found to be $(52.4 \pm 0.4)\,\%$.

4.2   Fit strategy for the measurement of $\mathcal{R}(J/\psi)$

The parameter $\mathcal{R}(J/\psi)$ is measured in a maximum likelihood fit to the binned distributions of the missing mass squared (m_{miss}^2), decay time, and rest frame variables q^2 and E_μ^* (through the variable Z) in the data sample passing the selection described in the above selection. The fit model assumes that these distributions are a mixture of the distributions of the signal and normalization decay modes and various background processes. Their separate distributions bear no simple parametrized descriptions, and so are described using histogram templates derived from simulation and data, as described in the following sections. The relative fractions of the various components are allowed to vary in the fit, with some constraints, to find the fractions maximizing the likelihood of the observed number of events in each event under a Poisson distribution.

The fit is performed using the `HistFactory` package, part of the `RooStats`

Table 4.3: Table of selection efficiency ratios

Sample	$\frac{\epsilon_\tau^{\mathrm{GEN}}}{\epsilon_\mu^{\mathrm{GEN}}}$	$\frac{\epsilon_\tau^{\mathrm{REC}}}{\epsilon_\mu^{\mathrm{REC}}}$	$\frac{\epsilon_\tau^{\mathrm{SEL}}}{\epsilon_\mu^{\mathrm{SEL}}}$	$\frac{\epsilon_\tau^{\mathrm{ISO}}}{\epsilon_\mu^{\mathrm{ISO}}}$	$\frac{\epsilon_\tau^{\mathrm{PID}}}{\epsilon_\mu^{\mathrm{PID}}}$	$\frac{\epsilon_\tau^{\mathrm{L0}}}{\epsilon_\mu^{\mathrm{L0}}}$	$\frac{\epsilon_\tau^{\mathrm{HLT2}}}{\epsilon_\mu^{\mathrm{HLT2}}}$	$\frac{\epsilon_\tau}{\epsilon_\mu}$
2011 Up	1.016(7)	0.733(4)	0.724(4)	1.009(2)	0.933(4)	0.997(9)	1.002(2)	0.507(7)
2011 Down	1.029(10)	0.738(4)	0.721(4)	0.998(2)	0.935(4)	1.012(9)	0.998(2)	0.516(9)
2012 Up	1.015(7)	0.755(4)	0.726(4)	1.005(2)	0.935(4)	0.991(10)	1.007(2)	0.522(9)
2012 Down	1.001(7)	0.755(4)	0.728(4)	1.001(2)	0.930(4)	1.001(9)	1.005(2)	0.515(8)

project [79]. This framework allows several specializations that improve the fit model. In addition to the contribution fraction of each component, the fit framework allows extra degrees of freedom that parametrize changes in the histogram shapes, known as shape systematics, described in Section 4.2.1. The tool package also implements the Beeston-Barlow method [80] to evaluate the effect on the fit of the statistical uncertainty in the template distributions, which are derived from simulation and data samples of finite size. This is described in Section 4.2.2 and is used to determine the nominal value of $\mathcal{R}(J/\psi)$ and to evaluate an important systematic uncertainty.

Finally, the maximum likelihood analysis used for the fit is not guaranteed to be unbiased. In fact, a sizeable bias is found to exist due to the presence of empty and nearly-empty bins, and their interplay with statistical fluctuations in the fit templates. This bias is evaluated using a new toy study method utilizing kernel density estimation for resampling data and fit templates, and is described in Section 4.2.3.

4.2.1 Shape systematics

Using the `HistFactory` framework, alternative templates for the various B_c^+ decay modes, evaluated with certain parameters adjusted higher and lower, provide the basis for shape systematic parameters that float in the fit. For a given value of a shape systematic, `HistFactory` creates an interpolated template based on the bin contents of the nominal template and the templates with higher and lower values of the corresponding parameter. Typically this is done to assign systematic uncertainty to a measured value due to the imprecisely known value of some parameter affecting the template shapes. It can also be used to measure the shape parameter values themselves.

Several shape systematics are introduced in the fit. In particular, the B_c^+

lifetime is not yet a precisely measured quantity. The default value of the B_c^+ lifetime used in simulation is set to the out-of-date value $0.453\,\mathrm{ps}$, reported by CDF and D0. Recent measurements by LHCb have indicated a slightly higher value, closer to $0.510\,\mathrm{ps}$, and the new PDG average value is $0.507(9)\,\mathrm{ps}$. The various B_c^+ decay simulation samples are reweighted to have a lifetime equal to $0.5113\,\mathrm{ps}$, a value picked before the recent PDG update but compatible with it. A shape systematic, corresponding to an uncertainty in the lifetime of $0.0094\,\mathrm{ps}$, is included in the fit.

As described in the sections below, several form factor parameters for $B_c^+ \to J/\psi\, \ell^+ \nu_\ell$, for $V(q^2)$, $A_1(q^2)$, and $A_2(q^2)$, are determined in an auxiliary fit normalization-rich region using shape systematic parameters (see Section 4.3.1). Several remaining form factor parameters are allowed to vary in the main fit, including those for $A_0(q^2)$, which affects the signal decay $B_c^+ \to J/\psi\, \tau^+ \nu_\tau$ but not the normalization decay $B_c^+ \to J/\psi\, \mu^+ \nu_\mu$, and some for $B_c^+ \to \psi(2S)\ell^+ \nu_\ell$ and $B_c^+ \to \chi_c \mu^+ \nu_\mu$, described below in Section 4.4.1.

4.2.2 The Beeston-Barlow algorithm

Since no simple analytical shapes are expected for the distributions of these variables, it is prudent to use a binned maximum template likelihood fit with signal and background histograms taken from Monte Carlo and data driven samples. It is also necessary to evaluate the statistical error resulting from the finite statistics used to produce these templates with the Beeston-Barlow method. This can be easily done with the `HistFactory` package.

4.2.3 Assessing bias in the fit model through pseudo-experiments

A maximum likelihood estimate (MLE) of a physical parameter produced using a dataset of a finite number of events is not guaranteed to be unbiased. The canonical example is the MLE estimate of the variance of a normal distribution fit to a set

of N events; its $1/N$ factor must be replaced with $1/(N-1)$ (known as Bessel's correction) to produce an unbiased estimate. Similarly, a maximum likelihood fit has no mathematical guarantee that it's estimates of event counts are not biased.

The canonical procedure for estimating this potential bias is the bootstrap. In an unbinned maximum likelihood fit, where analytic PDFs for the various fit components are available, the parametric bootstrap is applied. Values are sampled from the various PDFs (typically with parameters near the best-fit point) to produce a series of psuedo-datasets. The fit procedure is applied to each pseudo-dataset to produce estimates of the parameter(s) of interest. Comparing these estimates to the generated values produces a measurement of the bias in the MLE estimator. In the most general case, the bias may depend on the value of the parameter of interest (or on the multi-dimensional space of values of all fit parameters).

On the other hand, a binned maximum likelihood fit often makes use of discrete histogram templates (technically, probability mass functions or PMFs) in place of analytic PDFs, such as in this analysis. The total template, constructed from a sum of the histogram templates for the various signal and background modes, is typically quasi-parametric, containing parameters representing the fractions of its component modes (as well as others corresponding to template shape systematics). The above procedure can be extended by sampling binned pseudo-datasets from the total histogram template. Indeed, this approach has been used in assessing bias for several analyses using binned maximum likelihood template fits.

However, there is a subtle but crucial difference in the case of binned likelihood fits. While sampling from an analytic PDF can produce pseudo-data in any physically allowed region, sampling from histogram templates cannot impute pseudo-data of some component type in regions where the histogram template of that component is empty. When the true underlying distributions have sparse regions (especially the case for multi-dimensional fits), the histogram templates are likely to have certain
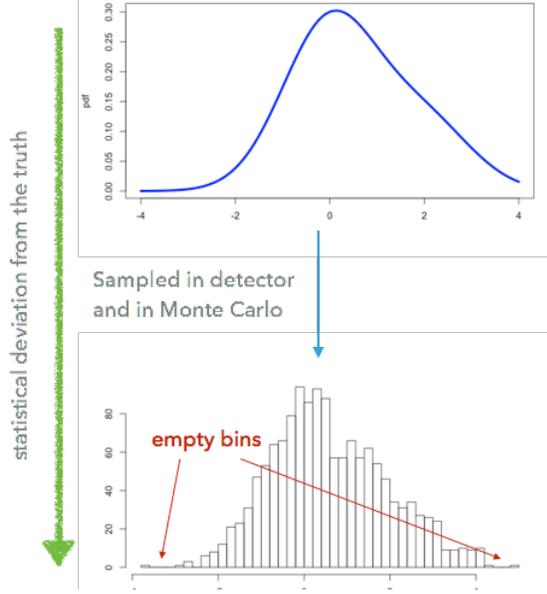
Figure 4.3: The true distributions $f_{i,1}(x)$ (top) has support everywhere, but the high statistics histogram template (bottom) has sparse bins in the tail regions.

patterns of empty and nearly-empty bins in these regions; moreover, these patterns are bound to differ by component. Since these patterns are preserved when sampling pseudo-data from the total template, they serve as artificial features identifying the various components. The space of pseudo-datasets is thus quite constrained, and the toy studies can both overestimate the ability to measure certain components (because they preserve artificial features) and ignore possible biases induced by the nature of the histogram templates in these sparse regions.

An ideal toy study would instead sample pseudo-datasets directly from the true distributions underlying the various component modes, thereby generating realistic pseudo-data in sparse regions and preventing the introduction of artificial features. However, the very reason that these analyses use binned fits with histogram templates is that the true underlying distributions are unknown; or rather that they are un-parametrizable: the fastest method available for sampling from them is simulation generation. So, in cases where the available simulation statistics are limited and the number of bins large, allowing sparse regions, this "typical"

80

straight-forward toy study method can be insufficient.

KDE pseudo-experiments

The problem can be solved by performing the toy study with an approximated set of underlying distributions obtained via kernel density estimation (KDE). These are similar to the simulation distributions on which they are based, but have support where these templates are empty and permit fast sampling via random number generation (RNG).

Each component i in the fit has a true underlying distribution, or first generation distribution $f_{i,1}(x)$, which describes both the generated simulation sample and the corresponding class of events in data (assuming no important data/MC differences). The available simulation sample for i, sampled from the first generation distribution $f_{i,1}(x)$, can be used to produce an estimated distribution through kernel density estimation. In this technique, the model for the true distribution $\hat{f}_{i,1}(x)$ is assumed to be a mixture of gaussians, each centered around one of the simulation events. The width of each gaussian is known as the bandwidth of the kernel density estimator. A popular implementation is the adaptive kernel density estimate, in which this width is narrower in highly populated regions and wider in sparser regions of the simulation dataset.

The resulting estimated distribution $\hat{f}_{i,1}(x)$ is not an unbiased estimate of the true distribution in any sense — in particular, it tends to broaden sharp features. However, it is similar enough to the true distribution to be useful for toy studies. So, it is denoted the second-generation distribution $f_{i,2}(x) = \hat{f}_{i,1}(x)$. New simulation histogram templates can be sampled from $f_{i,2}(x)$ with equivalent statistics to the original simulation templates, as well as a series of pseudo-datasets (see Fig. 4.4). Fitting these second-generation pseudo-datasets with the second-generation histogram templates is analogous to the ideal but impractical toy study in which pseudo-datasets
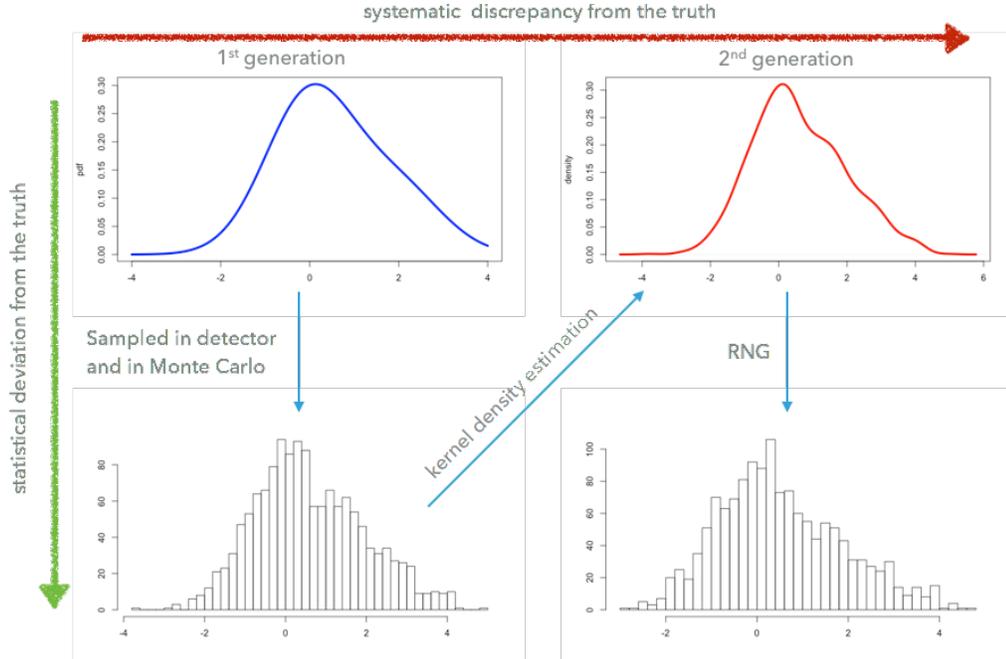
Figure 4.4: Applying a kernel smoothing procedure to sampled simulation data produces second-generation distributions $f_{i,2}(x)$ (top right), which can quickly sampled via RNG (bottom right).

sampled directly from the true first-generation distributions $f_{i,1}(x)$ is fit with the nominal histogram templates sampled from the same.

KDE pseudo-experiment validation

As simulation statistics increase, the histogram templates become better representations of the underlying distributions, with fewer and fewer sparse or empty regions. In the limit of infinite simulation statistics, in fact, the histogram templates become exact PMF shapes. In this case, sampling from the templates itself (as in a "typical" toy study) is equivalent to sampling directly from the underlying distribution.

This correspondence suggests a closure test that can be applied to the KDE toy-study method. On one hand, extremely high statistics histogram templates can be sampled from the second-generation distributions $f_{i,2}(x)$ and used to fit
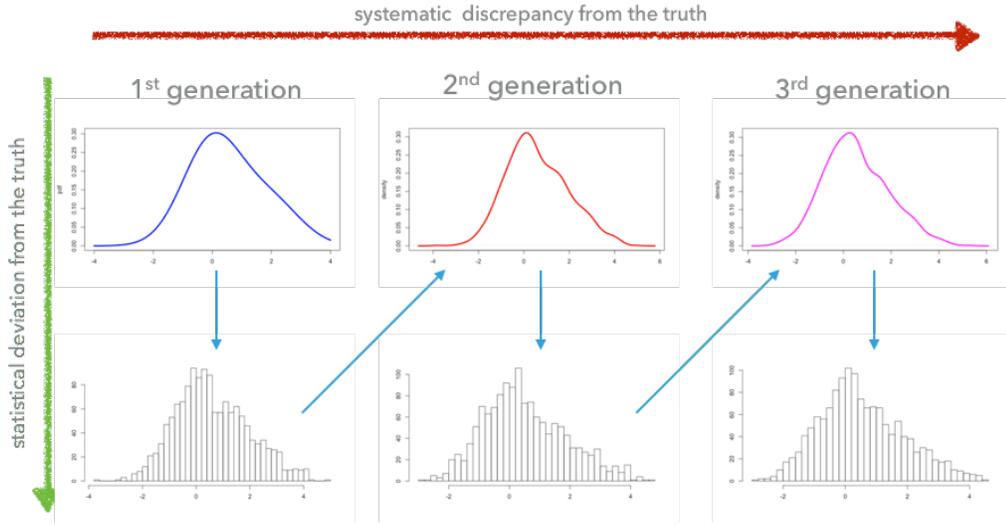
Figure 4.5: Nesting the kernel-smoothing procedure produces third generation distributions $f_{i,3}(x)$ (top right) — indexed by the set of the second-generation histogram templates used to produce them.

second-generation pseudo-datasets (which are nearly equivalent to being resampled directly from the high-statistics templates). On the other, a "typical" toy study can be performed, where the first-generation histogram templates are resampled to produce first-generation pseudo-datasets (in effect, assuming them to be exact representations of $f_{i,1}(x)$). These separate toy studies are analogous, with the crucial difference that one takes place in the space of kernel-smoothed distributions $f_{i,2}(x)$, while the other takes place in a space of empirical distributions determined by the nominal histogram templates (which are statistically representative of $f_{i,1}(x)$). In essence, a comparison of the two tests for biases introduced by the kernel smoothing procedure itself.

KDE pseudo-experiment systematics

The closure test validates the KDE toy study method by exploiting the infinite statistics limit, where biases are typically very small. A second procedure validates

the method at finite simulation statistics levels by extending the KDE procedure to yet one more generation, in effect nesting it within itself. Since the method exploits performing toy studies with second-generation distributions $f_{i,2}(x)$ to represent impractical toy studies with the first-generation distributions $f_{i,1}(x)$, a new set of toy studies can be performed using third-generation distributions $f_{i,3}(x)$. Such third-generation distributions can be created for a number of histogram templates sampled from $f_{i,2}(x)$ via kernel-smoothing (see Fig. 4.5).[4] Just as the second-generation distributions are systematically smeared from the true first-generation distributions, so are the third from the second. Based on this analogy, the discrepancy between the second- and third- generation distributions approximates the discrepancy between the first- and second-generation distributions, motivating a test to measure the effect of this discrepancy on the bias estimation.

The exact procedure of the test is to produce a number (say 25) of sets of second-generation histogram templates by sampling from $f_{i,2}(x)$. Each of these can be used to fit a pseudo-datasets sampled from $f_{i,2}(x)$ as well. Simultaneously, each set is used as the basis for the creation of kernel-smoothed third-generation distributions $f_{i,3}(x)$. From each of these a number of third-generation histogram templates and pseudo-datasets are created (say 25 again). The net result is a collection of 25 second-generation fits, each with an associated 25 third-generation fits. To explore possible dependence of any systematic on $\mathcal{R}(J/\psi)$, it is convenient to use a different value of $\mathcal{R}(J/\psi)$ for each of the 25 groups (1 second-generation and 25 third-generation) of pseudo-datasets. Then the bias of the 25 second-generation fits can be compared to the bias of the 625 third-generation fits, both as a function of $\mathcal{R}(J/\psi)$. The agreement (or lack thereof) of these bias estimates measures the accuracy of the second-generation toy studies in estimating the true bias, and can

---

[4]Since the third-generation distributions depend on the pseudodata sampled from the second-generation distributions, they could be more precisely labeled $f_{i,3}^k(x)$ where k labels the second-generation pseudo-dataset.

be used to assign a systematic uncertainty to the bias correction.

Application of KDE pseudo-experiment procedure

As mentioned above, an unexpected bias was discovered in the fit after the first unblinding (first noticed as a dependence of fit yields on the number of bins used in the fit). As "typical" toy studies had not anticipated this bias, we now apply our KDE toy study technique. For each histogram template i in the fit, we produce a second generation density $f_{i,2}(x)$ via KDE. For computational efficiency, and to maximize the support of the distribution in the multidimensional space, we perform one-dimensional KDE along m_{miss}^2 in the 40 separate bins of decay time and Z. We then proceed to sample second-generation templates and 1500 pseudo-datasets from these distributions. Before unblinding the value of $\mathcal{R}(J/\psi)$, we are interested in the relationship between its measured value and its true value, and therefore we generate these datasets with a range of values for $\mathcal{R}(J/\psi)$ (ranging from zero to absurdly large values). Figure 4.6 shows the results of this toy study. In these tests, the same second-generation templates are used in each fit; subsequent tests involving alternate template sets do not show any increased variance or change in the overall bias.

Given a true value θ for a parameter (*i.e.*, $\mathcal{R}(J/\psi)$) measured to be $\hat{\theta}$ via likelihood maximization, the bias is the quantity $E[\hat{\theta} \mid \theta] - \theta$, where E is the statistical expectation operator. In the measurement scenario, however, θ is an unknown quantity and instead only $\hat{\theta}$ is known. The desired quantity is $E[\theta \mid \hat{\theta}]$, which by Bayes' theorem is equal to

$$E[\theta \mid \hat{\theta} = \hat{\vartheta}] = \frac{\int \vartheta f_{\hat{\theta},\theta}(\hat{\vartheta}, \vartheta) \, d\vartheta}{\int f_{\hat{\theta},\theta}(\hat{\vartheta}, \vartheta) \, d\vartheta} = \frac{\int \vartheta f_{\hat{\theta}|\theta}(\hat{\vartheta} \mid \vartheta) f_{\theta}(\vartheta) \, d\vartheta}{\int f_{\hat{\theta}|\theta}(\hat{\vartheta} \mid \vartheta) f_{\theta}(\vartheta) \, d\vartheta}. \qquad (4.3)$$

where the quantities f are joint or conditional densities of the variables θ and $\hat{\theta}$. The
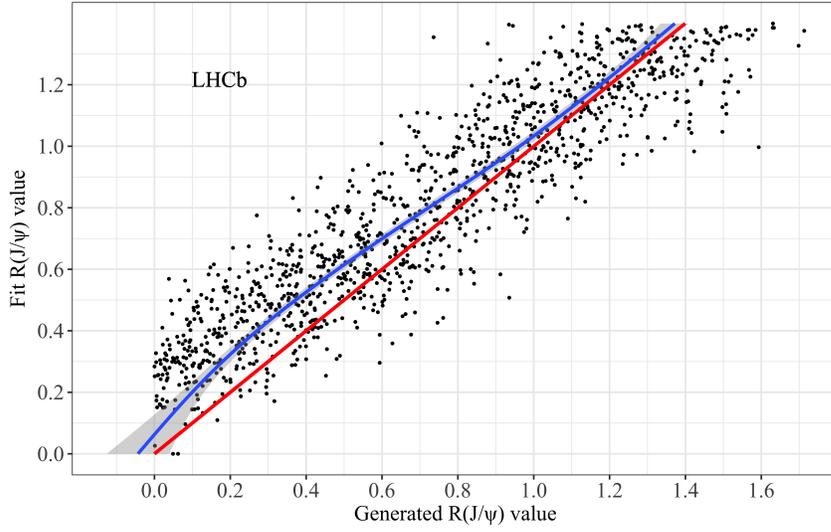
Figure 4.6: A significant bias exists in the fit for a range of value of $\mathcal{R}(J/\psi)$. The generated and measured values of $\mathcal{R}(J/\psi)$ are shown in the x- and y-axes. The blue curve shows the conditional mean generated value as a function of the fit value. The sharp increase at large values is an artifact of the cut-off value used in the prior distribution of $\mathcal{R}(J/\psi)$.

KDE toy studies provide a direct evaluation of this bias (under the assumption that the kernel smoothing does not significantly alter the bias, which is a addressed by a systematic uncertainty). The dataset of toy study results contains pairs $(\theta, \hat{\theta})$ of generated and measured values, with the generated values sampled uniformly from a large non-negative range. Assuming that any positive value of θ is equally likely a priori, the quantity in Eq. (4.3) is equal to the conditional mean of θ (the generated, x-axis quantity) at fixed $\hat{\theta}$ (the measured, y-axis quantity). This conditional mean is computed by producing a lowess curve (a locally linear regression) representing θ as a function of $\hat{\theta}$.

We perform the closure test described above, which compares the typical (pure bootstrap) toy study to a KDE study using extremely high statistics histogram templates. Its results are shown in Fig. 4.7. As expected, there is no discrepancy in the relationship between the generated and fit values between the two toy studies.

A systematic uncertainty on the bias correction is derived using the procedure

86

(a) Typical toy study   (b) High statistics KDE toy study
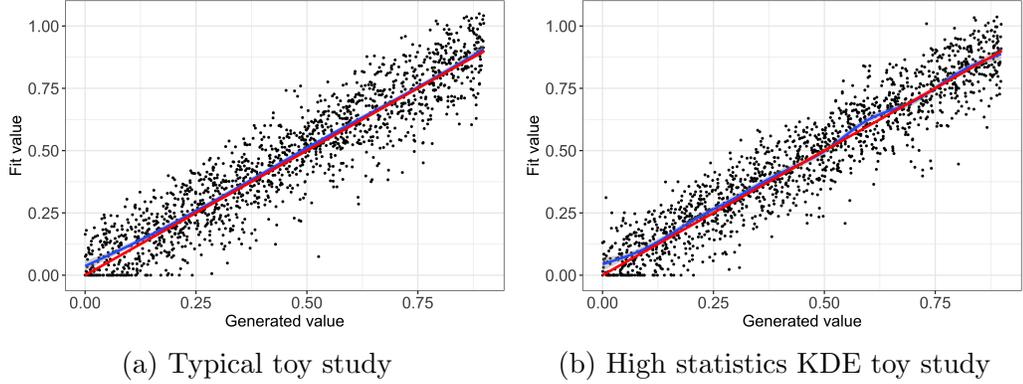
Figure 4.7: There is good agreement in the closure test comparing the high-statistics KDE toy study to the typical toy study. For ease of comparison, the blue curve shows the conditional mean fit value as a function of the generated value.

described above. In brief summary, from the second-generation distributions $f_{i,2}(x)$ a small number (25) of sets of second-generation templates and pseudo-datasets are sampled, equal in size to the nominal templates and dataset, respectively. From each of these sets, indexed by k, KDE is used to create corresponding third-generation distributions $f_{i,3}^k(x)$, and a number (25) of third-generation templates and pseudo-datasets are sampled (again equal in size to the corresponding nominal templates and dataset). For each k, a unique value of $\mathcal{R}(J/\psi)$ is used in the creation of the single second-generation pseudo-dataset and 25 associated third generation pseudo-datasets. A fit is performed using each of the 650 sets of templates and pseudo-datasets, with results shown in Fig. 4.8. The mean bias of the 625 third-generation fits differs from that of the 25 second-generation fits, in a manner largely independent of $\mathcal{R}(J/\psi)$. The net difference, equal to 5.4% after correcting for the efficiency ratio, is assigned as a systematic uncertainty.

4.3   Signal and normalization modeling

The projections of the templates for the signal and normalization mode are shown in Fig. 4.9. The normalization mode $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ is characterized by

(a) Tinker toy study results

(b) Difference between 2nd-gen. and 3rd-gen. measurements
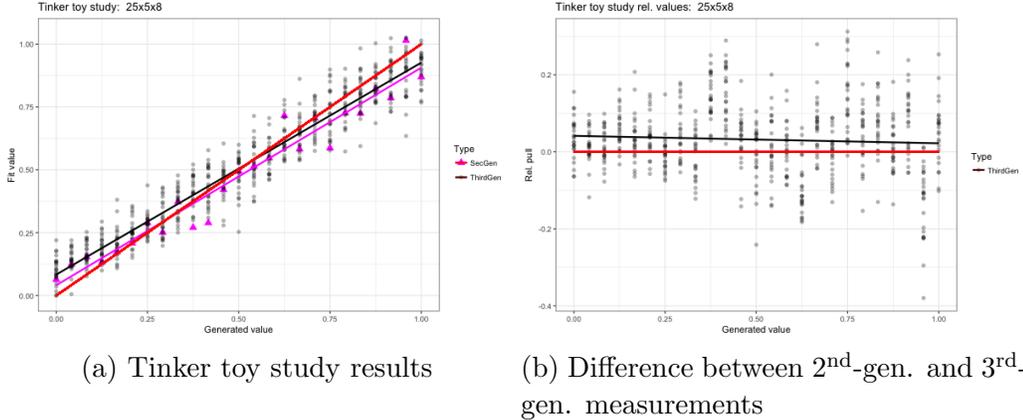
Figure 4.8: The tinker toy test compares a second-generation KDE toy study (purple) to a third-generation study (black). The fits in the third-generation toys generally produce slightly higher estimates than their corresponding second-generation fits, leading to a systematic uncertainty.

a small missing mass squared, which is only non-zero because of the resolution of the estimated B_c^+ momentum. The missing mass squared of the signal mode has a broader distribution, reflecting the invariant mass lost to the neutrino system. The decay time distributions of the two modes are similar, following a roughly exponential distribution with the B_c^+ decay time modified by lifetime acceptance parameters. The q^2 distribution of the normalization mode tends toward lower values than the signal, while its E_μ^* distribution tends toward higher values.

The signal decay $B_c^+ \to J/\psi \tau^+ \nu_\tau$, where $\tau^+ \to \mu^+ \nu_\mu \bar{\nu}_\tau$, is characterized by a higher missing mass. Its distribution in Z differs from the normalization because the τ^+ is produced through the scalar polarization of the W^* much more frequently than the μ^+, and this affects the E_μ^* and q^2 distributions of the final state μ^+. The expected ratio of signal to normalization events is $\mathcal{R}(J/\psi)\mathcal{B}(\tau^+ \to \mu^+ \nu_\mu \bar{\nu}_\tau) \approx 0.044\%$, taking the lower range of SM predictions $\mathcal{R}(J/\psi) = 0.25$ [57].

4.3.1 Form factor correction

The distribution of the fit variables for these modes are sensitive to the semileptonic decay form factors for the $B_c^+ \to J/\psi$ transition, which have not yet been
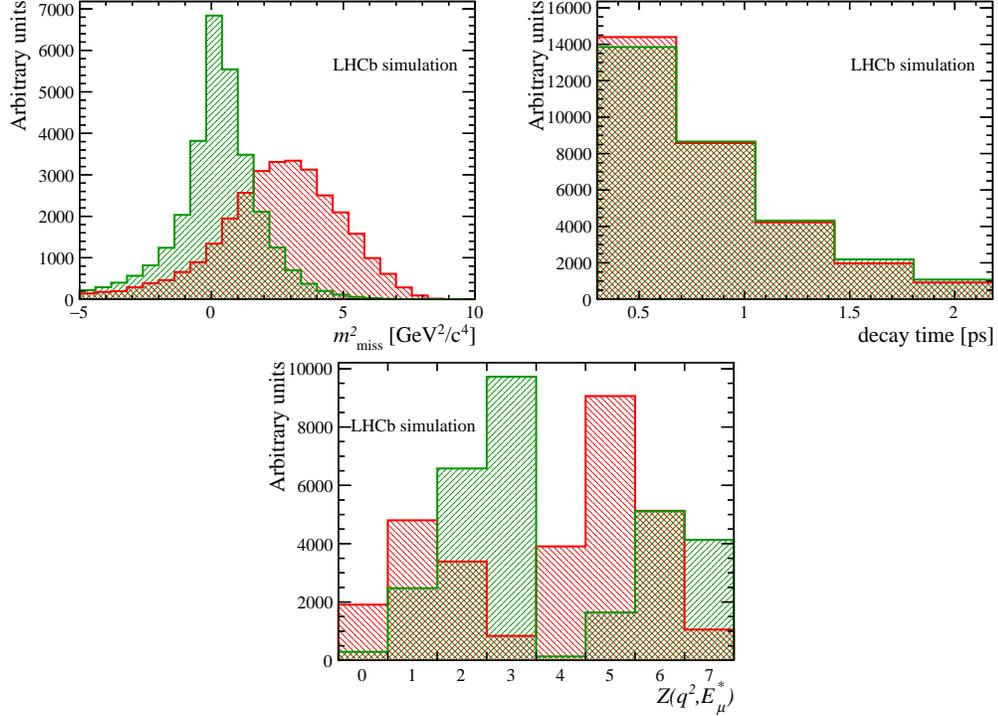
Figure 4.9: Fit templates for normalization $B_c^+ \to J/\psi \, \mu^+ \nu_\mu$; the signal template is shown in red for comparison.

measured and whose theoretical models cannot be trusted completely. A general overview of the treatment of form factors in this analysis is presented above in Section 3.2.

To account for the unknown true form factors of the $B_c^+ \to J/\psi \ell\nu$ decays, the signal and normalization templates are reweighted using a BCL expansion (see Section 3.2.2) with parameters determined in two stages within the fit. In the first stage, the parameters for the form factors $A_1(q^2)$, $A_2(q^2)$, and $V(q^2)$ (which describe the $B_c^+ \to J/\psi \, \mu^+ \nu_\mu$ normalization decay) are determined by a fit to a region of the phase space that is dominated by the normalization channel, hereafter referred to as normalization-rich region.[5] The normalization mode is not appreciably sensitive to the scalar form factor $A_0(q^2)$, which predominantly affects the semitauonic decay and whose parameters are determined in the main fit (described in Section 4.2.1).

---

[5]The definition of this region can be found later, in Section 4.9.1.

The fit procedure itself borrows largely from the fit used in the main analysis, using this as the basis for an iterative procedure to determine the form factor parameters. Given a set of BCL z-expansion parameters b_i for each form factor, the differential decay rate $d\Gamma/dq^2\,d\cos\theta$ is calculated as a function of the MC-truth momenta of the B_c^+ and its decay products via Eq. (3.5). The ratio of this quantity to the differential decay rate of the EFG model (with which the simulation was generated) is used to weight the MC. This ignores the effect of the form factors on angular correlations involving the J/ψ decay (which would be corrected by a quadruply-differential decay rate $d^4\Gamma/dq^2\,d\cos\theta\,d\chi\,d\cos\phi$), but is an acceptable approximation because this analysis does not rely on precise angular variables and correlations.

Uncertainties in the form factor values are added as degrees of freedom in the fit using the shape systematic framework of `HistFactory`, described in Section 4.2.1. To avoid higher-order effects induced by the interpolation procedure 's interpolation procedure, we employ an iterative procedure where in each step the fit templates are regenerated at the new values and $\pm n\sigma$ values of the shape parameters. The scaling n starts at 2 for the first iterations, so that the fit does not over-converge, and decreases to 1 at the end of the iterations. Table 4.4 shows the measured values of the parameters. Figure 4.12 compares the Kiselev, Ebert, and fitted BCL form factors to several preliminary lattice QCD results. The fitted BCL results are uncertain up to an overall scaling that cannot be determined in the fit. The fitted form of the $A_0(q^2)$ is included here for convenience, although these parameters are not determined until the nominal fit to the full dataset. Based on the limited information available, the fitted values agree better with the lattice predictions than either the Ebert or Kiselev models.

The parameters for the form factor $A_0(q^2)$, specific to the $B_c^+ \to J/\psi\,\tau^+\nu_\tau$ signal decay, are determined in the main fit, as nuisance parameters alongside $\mathcal{R}(J/\psi)$.
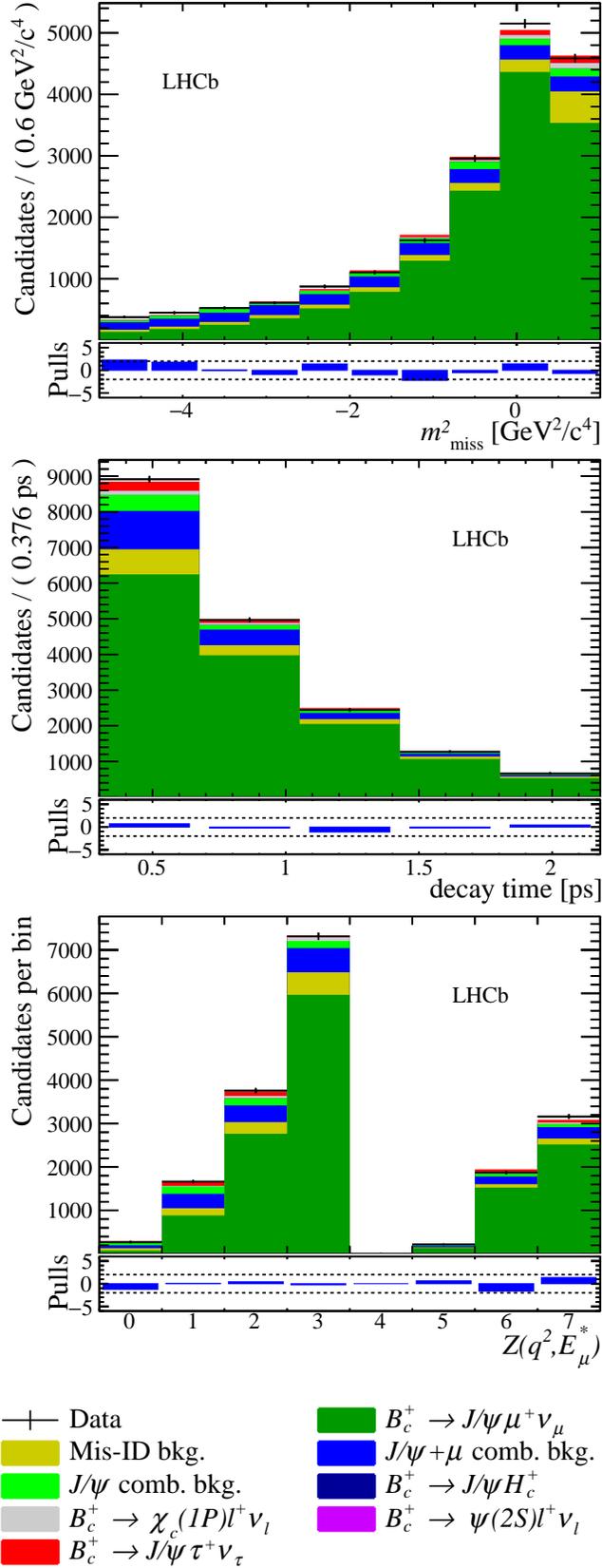
Figure 4.10: Projections of fit to normalization-rich region (including data and the non-signal component templates) onto m^2_{miss} (top), t (middle), and Z (bottom)

Table 4.4: The updated values of the BCL parameters after an iterative fit procedure to the normalization-rich region.

Form factor	Parameter	Starting value	Ending value
$A_1(q^2)$	b_0	0.652	fixed
	b_1	-5.631	-2.36 ± 1.56
$A_2(q^2)$	b_0	1.01	1.40 ± 0.18
	b_1	0	fixed
$V(q^2)$	b_0	0.884	1.29 ± 0.18
	b_1	-14.5	-15.7 ± 9.5
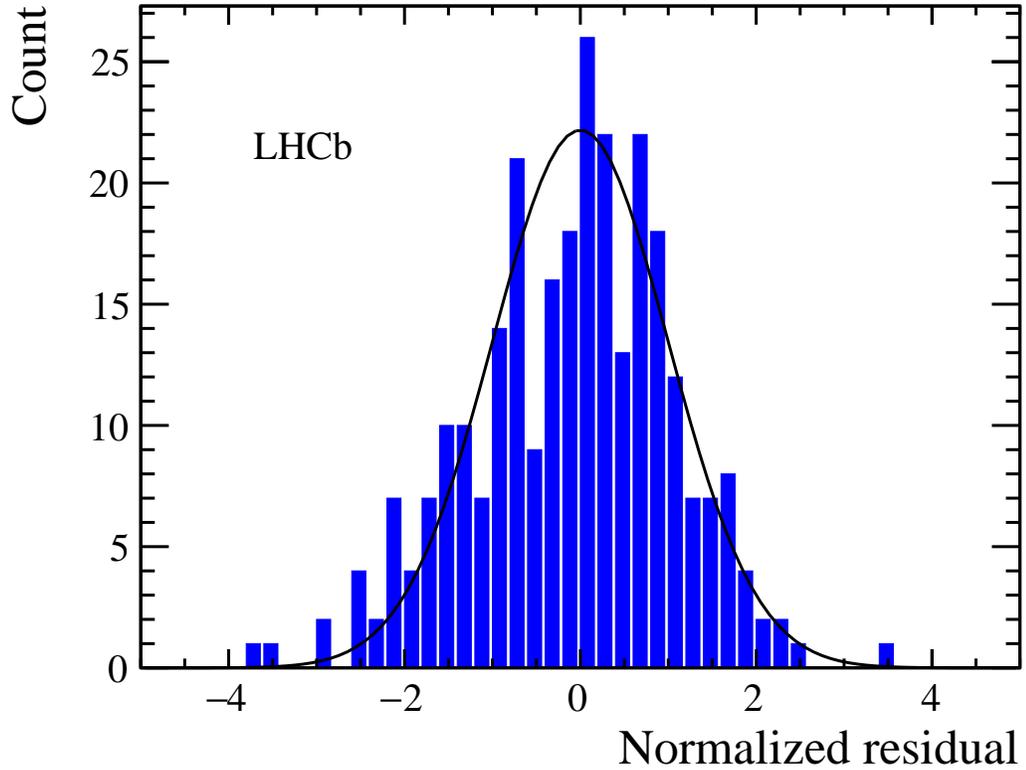
Figure 4.11: Histogram of the normalized residuals in each un-projected bin of the fit to the normalization-rich region. The red curve plots a standard normal distribution.

(a) $A_0(q^2)$ (b) $A_1(q^2)$

(c) $A_2(q^2)$ (d) $V(q^2)$
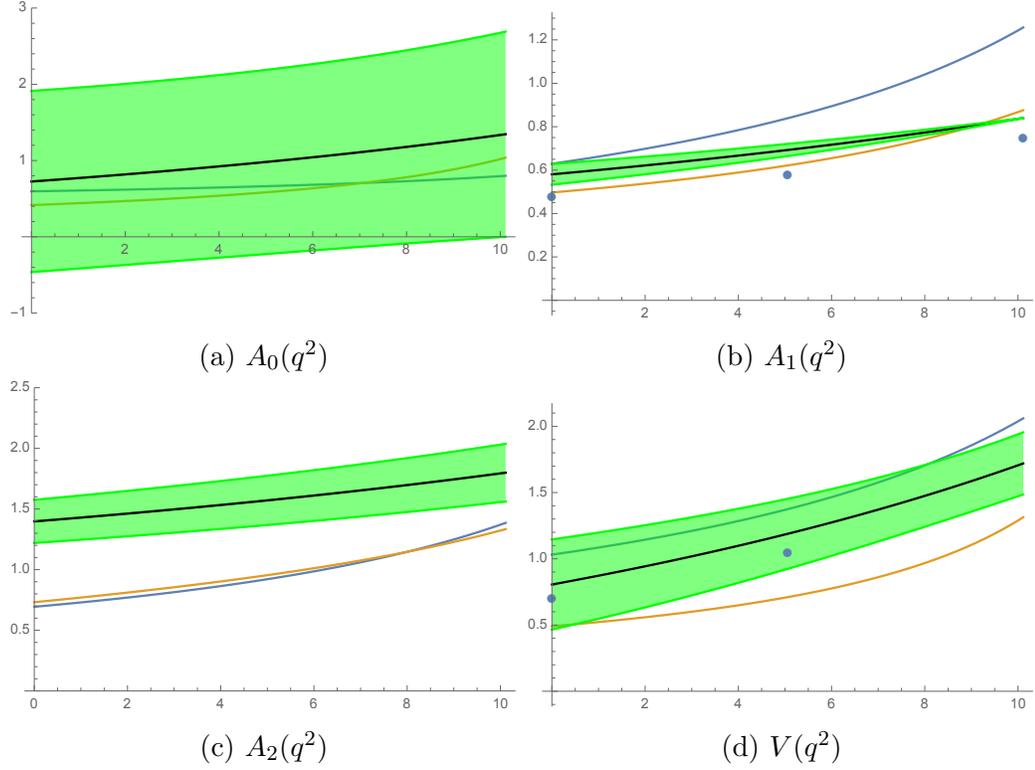
Figure 4.12: Plots of Kiselev (blue), EFG (orange), and fitted BCL form factors (black with green error bars); the black points show preliminary lattice QCD results.

(a) μ mode   (b) τ mode
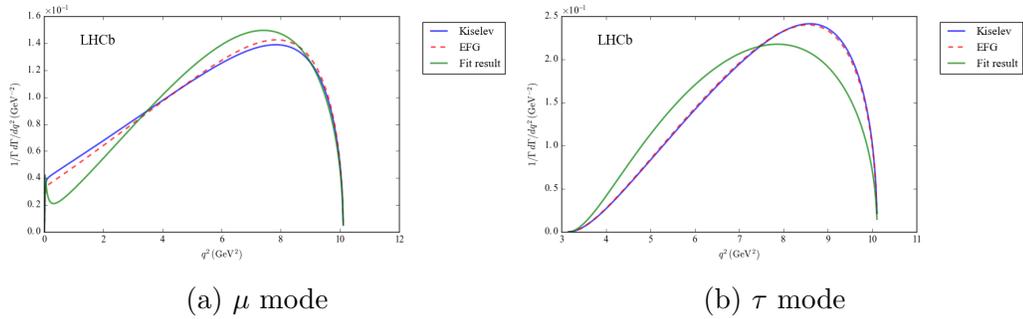
Figure 4.13: The Kiselev, EFG, and BCL q^2 distribution for $B_c^+ \to J/\psi\,\ell^+\nu_\ell$ show that the fit to the data prefers a slightly softer q^2 spectrum.

(a) μ mode
(b) τ mode
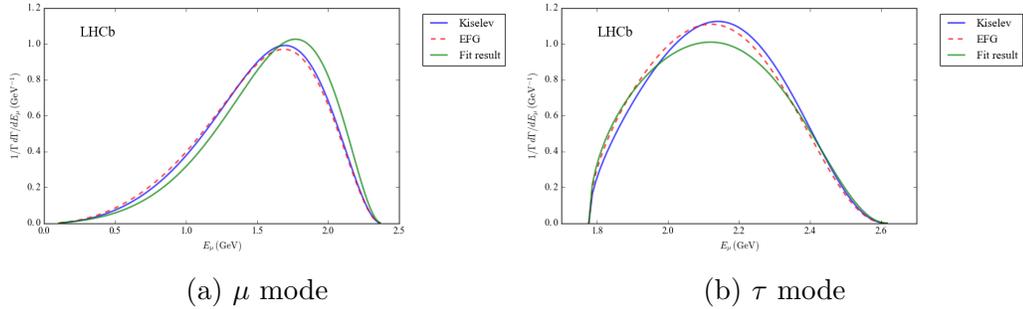
Figure 4.14: The Kiselev, EFG, and BCL E_ℓ^* distribution for $B_c^+ \to J/\psi\,\ell^+\nu_\ell$ show that the data prefers a slightly harder E_μ^*.

These signal and normalization templates are also subject to more generic data–MC differences discussed in Section 4.9, *e.g.*, corrections to the lifetime acceptance efficiency and the track multiplicity and impact parameter χ^2 distributions.

4.4 Feed-down background

Decays of B_c^+ to charmonia states heavier than the J/ψ, which in turn produce a J/ψ in their decay chains, are an important, though as it turns out small, background source. Little is known about these decays experimentally; on the other hand, the charmonium spectrum is well-understood and constrains the structure of possible feed-down backgrounds. In particular, only charmonium states with a mass below the DD threshold have an appreciable branching fraction to states including the J/ψ. The predominant feed-down channels are thus $\psi(2S) \to J/\psi\,\pi\pi$, considered in Section 4.4.1, and $\chi_{c\{0,1,2\}}(1P) \to J/\psi\,\gamma$, considered in Section 4.4.2. Possible smaller feed-down backgrounds involving the $\psi(3770)$ state and $X(3872)$ state are considered in Section 4.4.3. A fit to a control sample enriched in feed-down decays is described in Section 4.4.4.
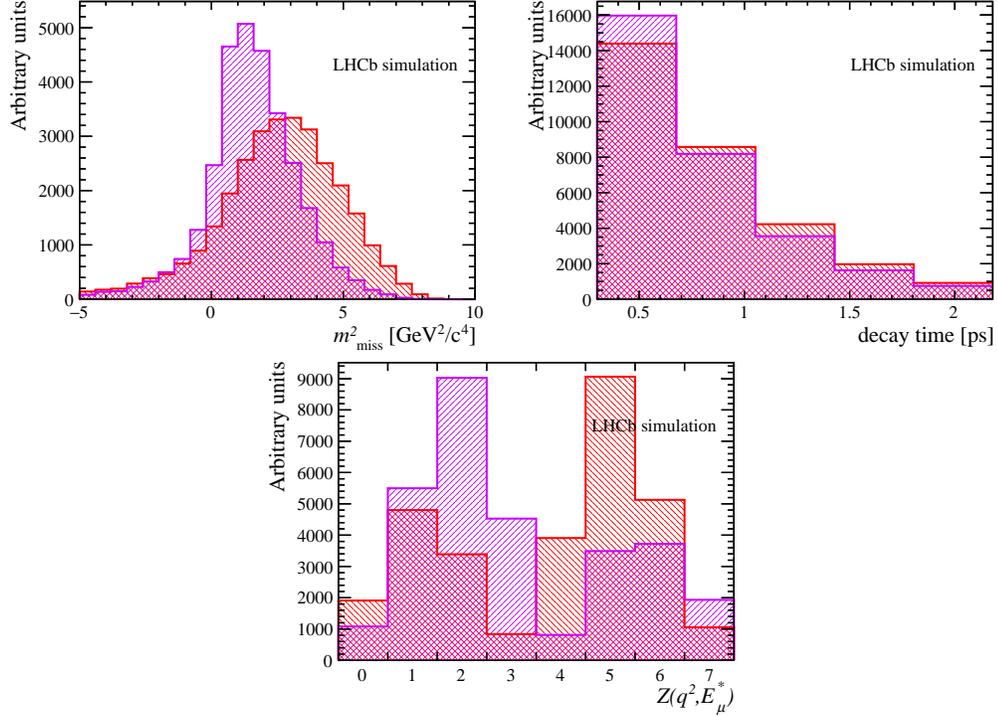
Figure 4.15: Fit templates for feed-down $B_c^+ \to \psi(2S)\mu^+\nu_\mu$ and $B_c^+ \to \psi(2S)\tau^+\nu_\tau$; the signal template is shown in red for comparison.

4.4.1 Feed-down from $B_c^+ \to \psi(2S)\ell\nu$

The available phase space for the pions in the principle feed-down chain $\psi(2S) \to J/\psi\,\pi\pi$ is constrained; consequently, the missing mass squared distribution of this process rises steeply at at the $\pi^+\pi^-$ threshold. As a B_c^+ decay, its decay time distribution mirrors that of the signal and distribution quite closely. Its q^2 distribution is heavily biased towards lower values than the signal. The projections of the template for this background are shown in Figure 4.15.

Predictions for the branching fraction ratio

$$\frac{\mathcal{B}(B_c^+ \to \psi(2S)\mu^+\nu_\mu)}{\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu_\mu)} \tag{4.4}$$

vary significantly. The EFG model predicts a ratio of 2.5% [63], while the Kiselev model predicts 4.9% [57]. Accounting for the 61% $\psi(2S) \to J/\psi\,X$ branching frac-

tion, the expected contribution is in the range $1.5 - 3\%$, but with large uncertainty. There is not much literature on the ratio $\mathcal{R}(\psi(2S))$; the Kiselev model predicts it to be 8.5% [57]. We conservatively take $\mathcal{R}(\psi(2S)) = 8.5\%$ in the fit, and assign a corresponding systematic uncertainty (see Section 4.11.11).

There is not enough statistical sensitivity to these background modes to warrant a full BCL parameterization a la $B_c^+ \to J/\psi \, \ell^+ \nu_\ell$. Instead, the slope of the effective Isgur-Wise function (corresponding to a CLN-like ρ^2) is allowed to vary. We place a generous constraint on this value, allowing it to increase or decrease by 2.3 at a 1σ penalty (this allows a change in sign at little penalty). This is the width of a uniform distribution between ± 4, roughly $\pm 100\%$ variations in the parameter. Similarly, ρ^2-like parameters for the χ_{c1} and χ_{c2} feed-down backgrounds are allowed to vary within the same constraints.

4.4.2   Feed-down from $B_c^+ \to \chi_c \ell \nu$

The projections of the combined template for the χ_c decay modes is shown in Fig. 4.16. Its distributions in all the physical fit variables it quite similar to those of the $\psi(2S)$ feed-down background, including a sharp rise in missing mass squared and a q^2 distribution biased towards lower values than the signal.

There are few theoretical treatments of the branching fractions to χ_c states. Hernandez *et al.* [59] predict

$$\frac{\mathcal{B}(B_c^+ \to \chi_{c(0,1,2)}(1P)\mu^+\nu_\mu)}{\mathcal{B}(B_c^+ \to J/\psi \, \mu^+\nu_\mu)} = (7.1\%, 4.3\%, 8.6\%), \tag{4.5}$$

while the WWL model [65] predicts

$$\frac{\mathcal{B}(B_c^+ \to \chi_{c(0,1,2)}(1P)\mu^+\nu_\mu)}{\mathcal{B}(B_c^+ \to J/\psi \, \mu^+\nu_\mu)} = (13.6\%, 9.1\%, 11.2\%). \tag{4.6}$$

The inclusive $\chi_{c(0,1,2)}(1P) \to J/\psi \, X$ branching fractions are $(1.27 \pm 0.06)\%$, $(33.9 \pm$
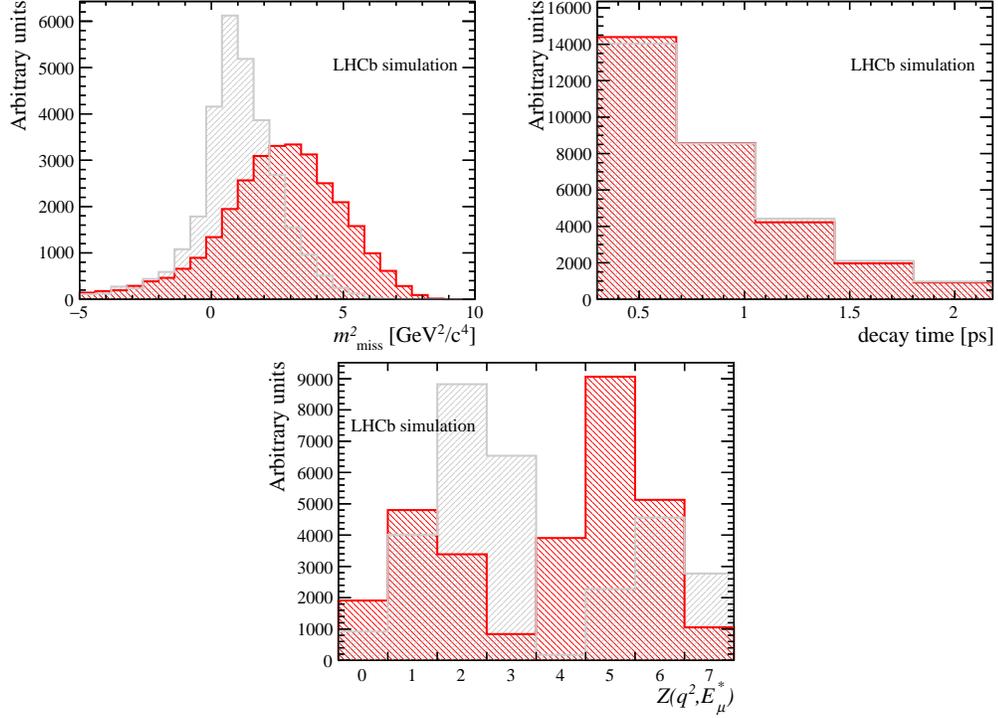
Figure 4.16: Fit templates for muonic feed-down $B_c^+ \to \chi_{c1}\mu^+\nu_\mu$ and $B_c^+ \to \chi_{c2}\mu^+\nu_\mu$; the signal template is shown in red for comparison.

1.2)%, and $(19.2 \pm 0.7)\%$, respectively [12]. Thus, the χ_{c0} feed-down mode can be safely ignored, with a branching fraction relative to the normalization of 0.1%–0.2%. The expected relative branching fractions of the χ_{c1} and χ_{c2} feed-down modes are in the ranges 1.5%–3.1% and 1.7%–2.2%, respectively, for a total contribution of 3.2%–5.3%. Given their kinematic similarity and similar kinematic distributions, the template representing this background is constructed from an equal proportion of these two decay modes.

Hernandez *et al.* also predict $\mathcal{R}(\chi_{c(0,1,2)}(1P)) = (12\%, 11\%, 7\%)$, while the WWL model predicts $(11\%, 11\%, 6\%)$. Combined with the 17.7% muonic branching fraction of the τ, a reasonable expectation for the fraction of such tauonic feed-down decays relative to the normalization is 0.05%–0.09%. This is entirely negligible, and so this semitauonic background is omitted from the fit; a corresponding systematic uncertainty is assigned in Section 4.11.11.

97

4.4.3 Exotic feed-down background

In general, charmonium states above the DD threshold decay very rarely to $J/\psi X$ and present a negligible background. A typical case is given by the $\psi(3770)$. The branching fraction $\psi(3770) \to J/\psi X$ is approximately ~ 0.002 level. Conservatively, the B_c^+ to charmonium branching fractions are (roughly) inversely proportional to the width of the charmonium state; *e.g.*,

$$\frac{\mathcal{B}(B_c^+ \to \psi(2S)X)}{\mathcal{B}(B_c^+ \to J/\psi X)} \sim \frac{\Gamma(J/\psi)}{\Gamma(\psi(2S))} \sim \frac{1}{3}.$$

This holds reasonably well in the B system. For the $\psi(3770)$, this ratio is $\sim 1/300$. If the branching fraction to $\psi(3770)\mu\nu$ were $1/300$ of the branching fraction to $J/\psi\,\mu\nu$ (in comparison, a theoretical estimate of this ratio is 3×10^{-5} [63]), and every $\psi(3770)$ event were identified as a $J/\psi\tau\nu$ event, this would lead to just a relative 0.02% increase in the measured value of R(J/ψ).

An atypical case is the $X(3872)$, with an anomalously small width and large branching fraction to $J/\psi X$. Estimating $\mathcal{B}(B_c^+ \to X(3872)\mu\nu)$ is highly dependent on whether the $X(3872)$ is a charmonium state or some DD threshold effect or molecule. This meson is quite similar to the $\psi(2S)$, both in its mass and its inclusive decay modes to $J/\psi X$. So, $B_c^+ \to X(3872)$ feed-down should have a similar missing mass distribution to $B_c^+ \to \psi(2S)$ feed-down, and is potentially absorbed by the same fit template. We have explicitly searched for $X(3872) \to J/\psi\pi^+\pi^-$ feed-down decays in our dataset and set a limit on its contribution; this is described in Section 4.4.4.

4.4.4 $\psi(2S)$ enriched control sample

To better understand these feed-down backgrounds, a cross-check fit is performed on a control sample enriched in $\psi(2S)$ feed-down, relative to other back-

grounds, to determine if the $\psi(2S)$ yield is compatible with the yield from the nominal data sample. In this control region, we require two oppositely charged tracks that look like secondary vertex (SV) tracks with an isolation BDT score > 0.2, with the third most SV track having a BDT score < 0.1, to reconstruct $\psi(2S) \to J/\psi \pi^+ \pi^-$ fee-down decays. We fit this sample with corresponding templates from our simulation and data-driven samples, defined in the same manner — information from the two charged pions does not enter into the fit. Plots of the fit projections and residuals are shown in Figs. 4.17 and 4.18. This fit quality is not spectacular but is acceptable for this control region.

The number of $\psi(2S)$ events is 383 ± 772 in the nominal sample and 45 ± 71 in the (relatively) enriched region. A data–MC comparison plot of the different in invariant mass between the $J/\psi \pi^+ \pi^-$ and J/ψ (Fig. 4.19) shows a clear peak from $\psi(2S)$ feed-down that is well modeled by the simulation and mis-ID templates. Most of the $\psi(2S)$ candidates are due to mis-ID background and not from B_c^+ decays. The selection efficiency of the $\psi(2S)$ feed-down decays is a factor of about 10 lower in the "enriched" control region. So, the $\psi(2S)$ yields in the two samples are compatible, though the large uncertainties make this a weak test. There is no indication that the fit to the signal sample is underestimating the contribution of feed-down from $B_c^+ \to \psi(2S)\mu^+\nu_\mu$ (the small tension would correspond to an overestimate).

Potential background from $X(3872) \to J/\psi X$ should also shows up in this control sample. The mass of this state is 775 MeV greater than the J/ψ. No visible peak at this value is evident in Fig. 4.19; however, one is revealed with finer binning. Figure 4.20 shows the fits to the $\psi(2S)$ and $X(3872)$ yields in this isolated region. The number of $X(3872)$ decays is about $8.0\% \pm 2.4\%$ of the number of $\psi(2S)$ decays. Performing the same fits in a larger sample without the strict unpaired muon PID criteria and anti-isolation requirements, we find a ratio of about $4.6\% \pm 2.2\%$ Figure 4.21 shows the fits to the $\psi(2S)$ and $X(3872)$ yields in this large dataset.
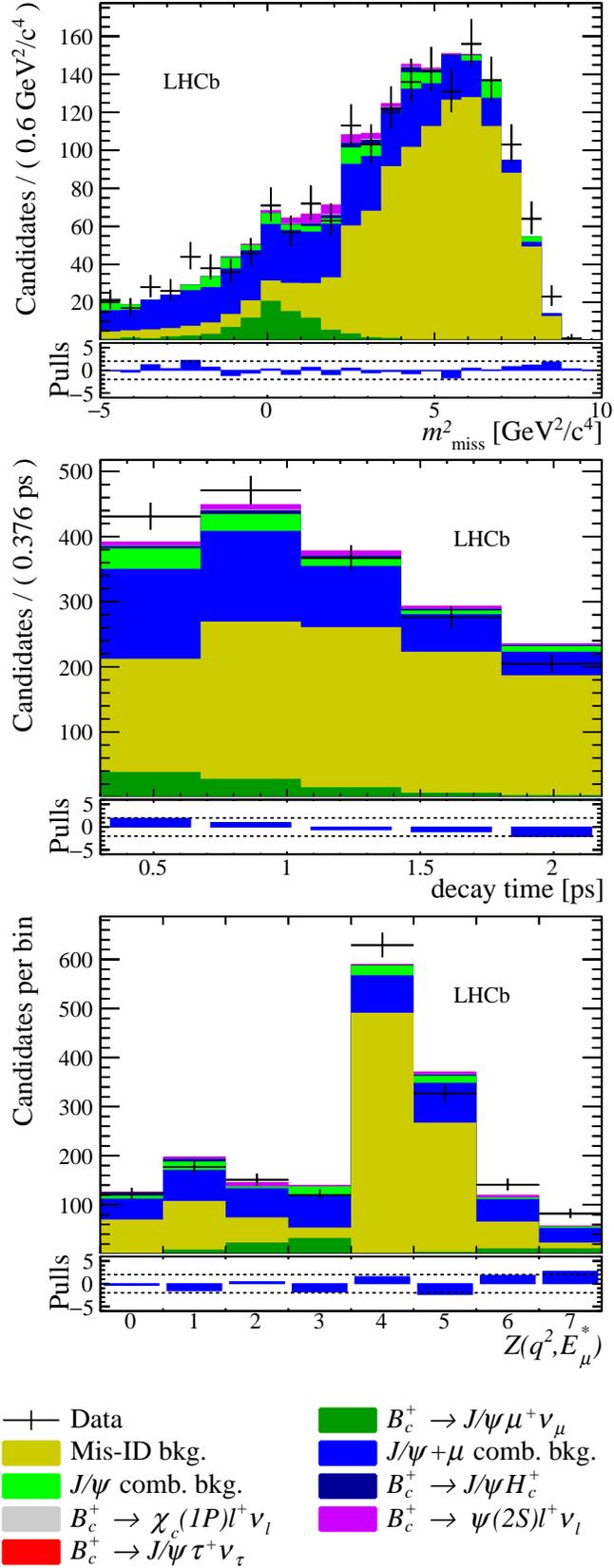
Figure 4.17: Projections of fit to the $\psi(2S)$ enriched sample (including data and the non-signal component templates) onto m^2_{miss} (top), t (middle), and Z (bottom).
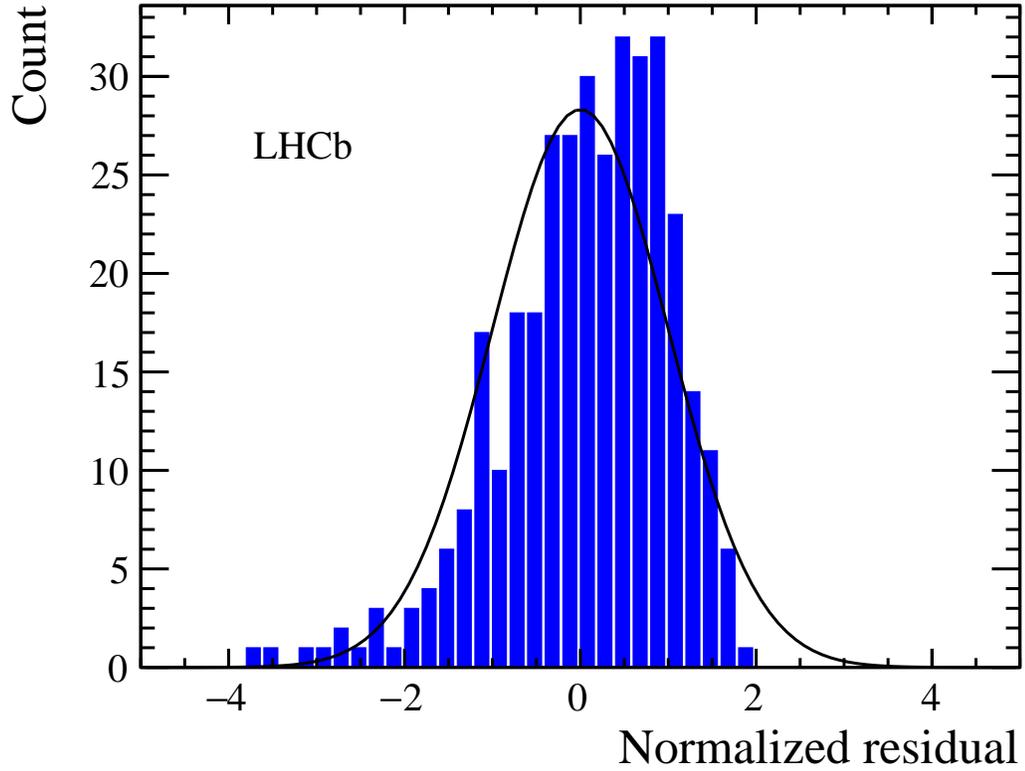
Figure 4.18: Histogram of the normalized residuals in each un-projected bin of the fit to the anti-isolated data. The red curve plots a standard normal distribution.
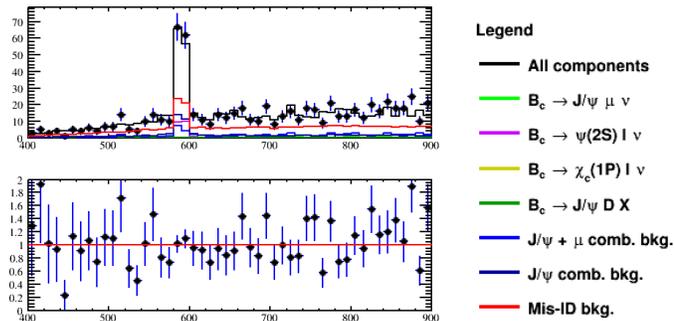
Figure 4.19: Comparison of the invariant mass in the $\psi(2S)$ window in data and in the simulation model corresponding to the best fit point, for the $\psi(2S)$ enriched control sample. The invariant mass of the J/ψ candidate is subtracted to sharpen the $\psi(2S) \to J/\psi \pi^+ \pi^-$ peak.
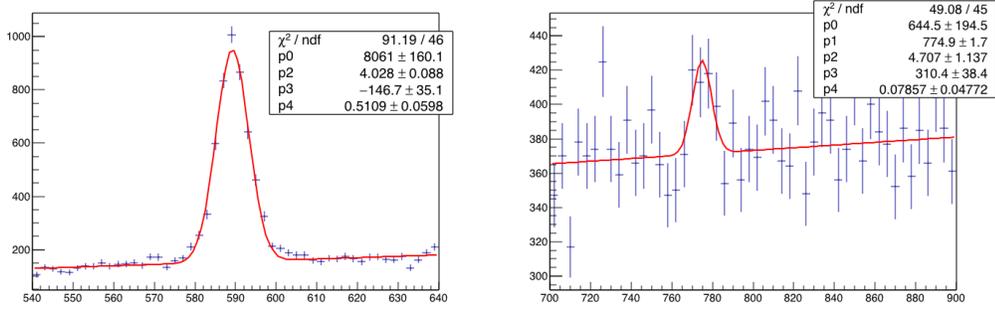
(a) Fit to the mass difference in the $\psi(2S)$ mass window

(b) Fit to the mass difference in the $X(3872)$ mass window

Figure 4.20: Fits to the $m(J/\psi\,\pi^+\pi^-) - m(J/\psi)$ distribution in the $\psi(2S)$ and $X(3872)$ regions in the anti-isolated dataset.
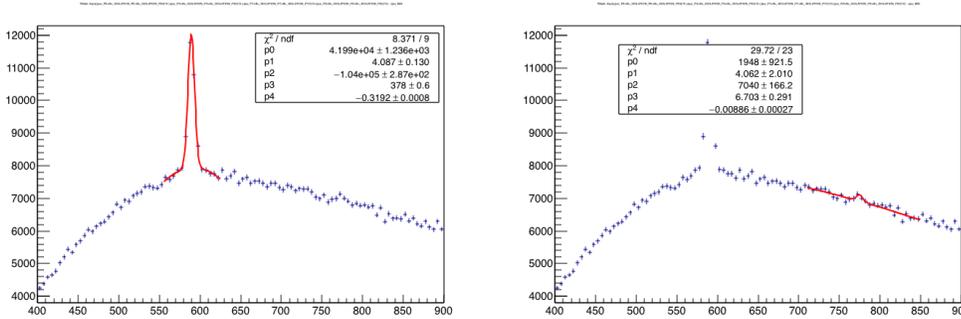
(a) Fit to the mass difference in the $\psi(2S)$ mass window

(b) Fit to the mass difference in the $X(3872)$ mass window

Figure 4.21: Fits to the $m(J/\psi\,\pi^+\pi^-) - m(J/\psi)$ distribution in the $\psi(2S)$ and $X(3872)$ regions in the entire (no-PID or isolation) dataset. The $X(3872)$ yield is still a small fraction of the $\psi(2S)$ yield in this expanded dataset.

Both of these charmonium populations originate from Bc decays and (mis-IDed) B decays. If the $X(3872)$ is a charmonium state, a reasonable assumption is that

$$\frac{\mathcal{B}(B_c^+ \to X(3872))}{\mathcal{B}(B_c^+ \to \psi(2S))} \sim \frac{\mathcal{B}(B^+ \to X(3872))}{\mathcal{B}(B^+ \to \psi(2S))} \sim 4.6\% \pm 2.2\%.$$

This is well within the statistical uncertainty in the amount of $\psi(2S)$ feed-down background (25%, relatively), and is expected to be kinematically similar. Therefore, the uncertainty in the treatment of the $\psi(2S)$ already covers the presence of any $X(3872)$ component.

4.5   Double-charm $B_c^+ \to J/\psi\, DX$ backgrounds

Potentially important backgrounds from $B_c^+ \to J/\psi\, DX$ decays, mediated by $b \to c\bar{c}s$ transitions, are currently not well studied in the literature nor well measured experimentally, and must be carefully considered in this analysis. In particular, $B_c^+ \to J/\psi\, D_s^{(*)}$ followed by $D_s \to \tau\nu$ or $D_s \to \mu\nu$ have very similar final states to the signal and normalization events. In analogy with B decays, there may be contributions from three-body $B_c^+ \to J/\psi\, DK$ decays, though theoretical predictions have not been made for these decays.

Recent measurements of the $\mathcal{B}(B_c^+ \to J/\psi\, D_s^{(*)})$ branching fractions relative to $\mathcal{B}(B_c^+ \to J/\psi\, \pi^+)$ by LHCb and ATLAS, shown in Table 4.5, help to clarify the situation [81, 82]. However, these are just one of a class of decays that presumably also includes $B_c^+ \to J/\psi\, D_s^{**}$ decays and three-body $B_c^+ \to J/\psi\, DK$ decays. Turning to the B system as a model, these decays are analogous to the relatively better measured $B \to D^* DX$ decays. Using data from the PDG, the relative branching fractions of these cocktail modes are shown in Table 4.6. The measured branching fractions of the D_{s1}^+ meson have not been normalized and are only known relative to each other, so we assume here that the only relevant branching fractions are 50% each to $D^{*+}K^0$ and $D^{*0}K^+$, ignoring the small non-resonant $D^+\pi^-K^+$ mode. A few decays are not completely analogous for the $B^\pm$, B^0, and B_c^+ parents, because B decays to charmonium states above the DD threshold can produce these final states. For example, $B^0 \to (\psi(3770) \to D^-D^+)K^0$ contributes to the $D^-D^+K^0$ final state, and no analogous resonance exists for the corresponding final states of $B^\pm$ and B_c^+ decays. The small affect of these differences on the cocktail composition is ignored.

Several simulation cocktails have been prepared and used to generate simulation for this analysis. The first cocktail represents two-body decay modes to $J/\psi\, D_{(s)}^{(*)}$.

Table 4.5: Measurements of relative $\mathcal{B}(B_c^+ \to M D_s^{(*)})$ branching fractions.

Experiment	$\dfrac{\mathcal{B}(B_c^+ \to J/\psi D_s^+)}{\mathcal{B}(B_c^+ \to J/\psi \pi^+)}$	$\dfrac{\mathcal{B}(B_c^+ \to J/\psi D_s^{*+})}{\mathcal{B}(B_c^+ \to J/\psi D_s^+)}$
LHCb	$2.90 \pm 0.57 \pm 0.24$	$2.37 \pm 0.56 \pm 0.10$
ATLAS	$3.8 \pm 1.1 \pm 0.4 \pm 0.2$	$2.8^{+1.2}_{-0.8} \pm 0.3$
Average	3.35 ± 0.67	2.59 ± 0.59

Table 4.6: Relative branching fractions of $b \to c\bar{c}s$ decays of B mesons. M is a $\bar{D}^{*0}$, D^{*-}, or J/ψ meson for the three species. $\mathcal{B}(\mu X)$ is the semimuonic branching fraction of D meson produced in the B_c^+ decay. The final column is the branching fraction of the B_c^+ decay mode followed by the semimuonic decay of the D, relative to that for the $M D_s^+$ mode.

Decay	$\mathcal{B}(\mu X)$	$B^{\pm}$	B^0	Relative BF
$M D_s^+$	$(8.0 \pm 0.4\%)$	1	1	1
$M D_s^{*+}$	$(8.0 \pm 0.4\%)$	2.09 ± 0.52	2.21 ± 0.35	2.17
$M D^+$	$(17.6 \pm 3.2)\%$	0.08 ± 0.03	0.08 ± 0.01	0.08
$M D^{*+}$	$(10.2 \pm 1.1)\%$	0.10 ± 0.03	0.10 ± 0.02	0.10
$M D^+ K^0$	$(17.6 \pm 3.2)\%$	0.26 ± 0.08	0.40 ± 0.06	0.35
$M D^0 K^+$	$(6.7 \pm 0.6)\%$	0.28 ± 0.06	0.31 ± 0.05	0.30
$M D^{*+} K^0$	$(10.2 \pm 1.1)\%$	1.12 ± 0.27	1.01 ± 0.16	1.04
$M D^{*0} K^+$	$(6.7 \pm 0.6)\%$	1.37 ± 0.32	1.33 ± 0.21	1.34
$M D_{s0}^{*+}$	$(8.0 \pm 0.4\%)$	0.11 ± 0.09	0.19 ± 0.08	0.15
$M D_{s1}^{+}$	$(8.0 \pm 0.4\%)$	1.28 ± 0.39	1.23 ± 0.30	1.25
$M D_{s1}^{\prime +}$	$(8.0 \pm 0.4\%)$	0.12 ± 0.04	0.09 ± 0.03	0.10
$M D_{s2}^{*+}$	$(8.0 \pm 0.4\%)$	< 0.06	< 0.03	0.03
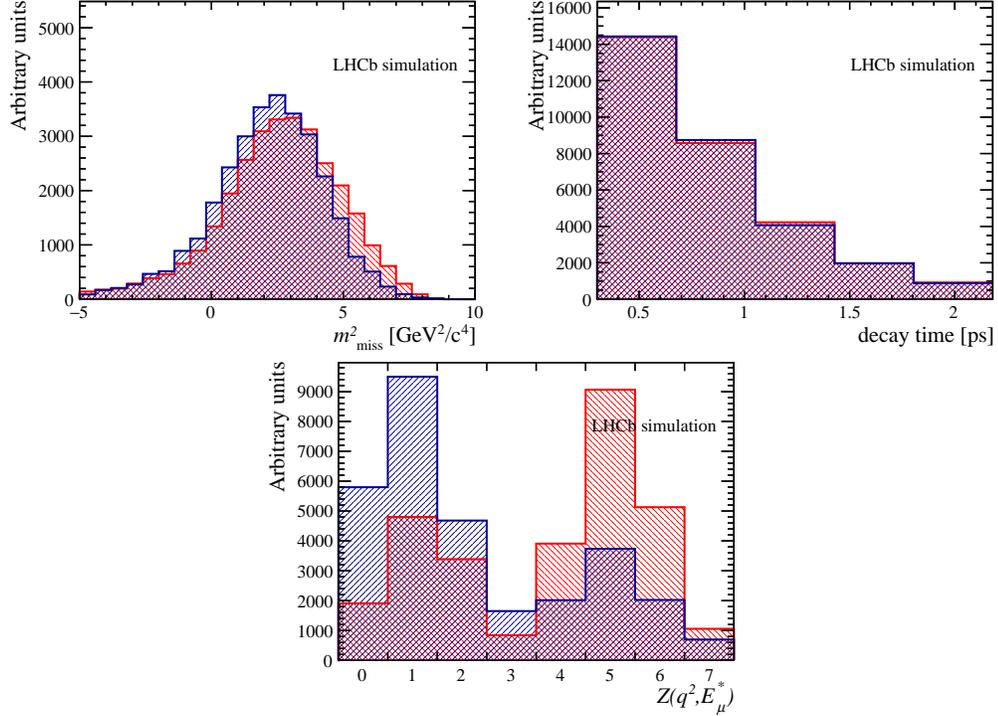
Figure 4.22: Fit templates for $B_c^+ \to J/\psi\, DX$ cocktail; the signal template is shown in red for comparison.

A second cocktail represents the quasi-two-body decays that proceed through a D_s^{**}, while a third cocktail represents potential non-resonant three-body decay. By the analogy to $B^\pm$ and B^0 decays, the expected contributions of the two-body and quasi-two-body decays are nearly identical, Thus, the template for this background is a combination of the two-body and quasi-two-body templates, in equal proportion. Its projections are shown in Fig. 4.22. Notably, its missing mass and decay time distributions are both quite similar to the signal; only the distinctly different distribution in q^2 (and to some extent E_μ^*) allows differentiation of this component and the signal.

4.6 Combinatorial $J/\psi + \mu^+$ background

The combinatorial background present in the reconstructed data comes predominantly from events containing a $B_{u,d,s} \to J/\psi X$ decay, where the secondary

J/ψ is paired with a muon coming from the rest of the event. Unfortunately, there is no simple data-driven way to obtain a sample of combinatorial background — for example, the three muon final state precludes the formation of a "wrong-sign" control sample.

Instead, the combinatorial background is modeled by simulation. The $B_{u,d,s}$ decay cocktails include many possible decays to J/ψ, $\psi(2S)$, χ_{c1}, and χ_{c2} Where possible, measured branching fractions reported by the PDG group are used, either from the B meson species in question or analogous decays of other meson species. This still leaves out many possible B decays with a J/ψ in the final state. Heuristic $SU(3)$ flavor symmetry guidelines, together with phase space ratios, were used to produce rough estimates for the branching fractions of the missing decays.

The total normalization of the templates representing the $B^{\pm}$, B^0, and B_s^0 decays is allowed to vary in the fit, while the fraction belonging to each are constrained to $0.4353 : 0.4446 : 0.1200$ within 20%, in accordance with the production cross sections of each species. The projections of the combined template for this background in shown in Fig. 4.23.

4.7 Combinatorial $(\mu^+ + \mu^-)\mu^+$ background

A second source of combinatorial background comes from fake J/ψ candidates created by combining random muons that happen to have an invariant mass in the narrow J/ψ window. This background can be modeled using data from the J/ψ invariant mass sidebands. While the nominal selection requires the J/ψ mass to be within $55\,\mathrm{MeV}$ of $3095\,\mathrm{MeV}$, the upper sidebands from $3150\,\mathrm{MeV}$ to $3190\,\mathrm{MeV}$ is kept to model the combinatorial J/ψ background (the lower sideband has a much larger contamination of real J/ψ decays due to final state radiation). The projections of the resulting template are shown in Fig. 4.24, and are broadly similar to those of the combinatorial background involving true J/ψ mesons combined with opposite-
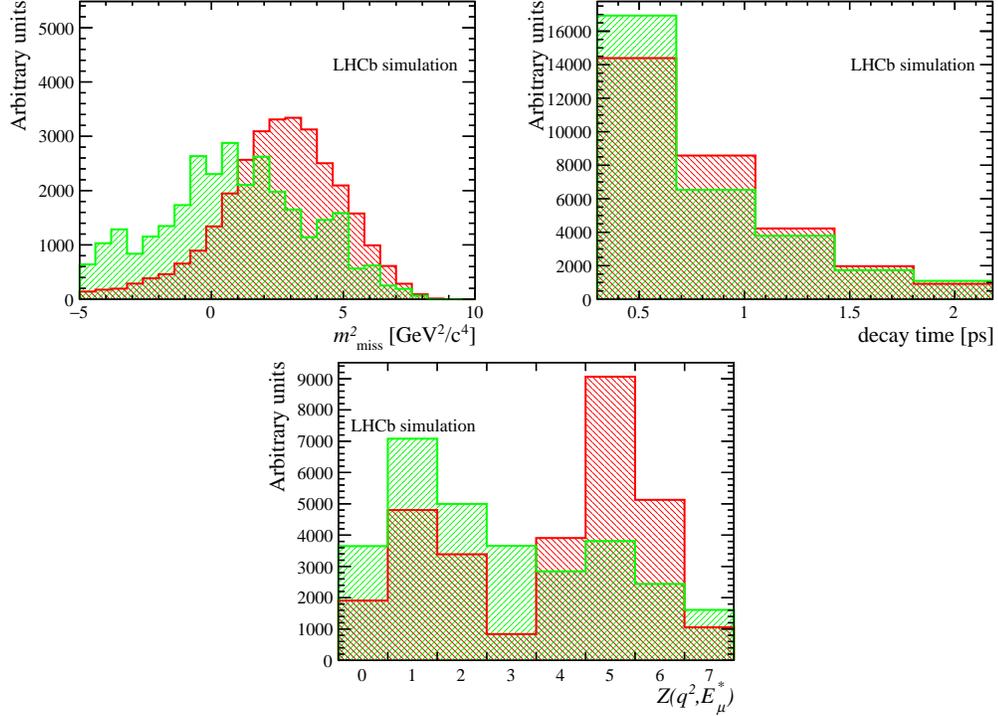
Figure 4.23: Fit templates for combinatorial background from $(B^0, B^+, B_s^0) \rightarrow J/\psi\, X$; the signal template is shown in red for comparison.

side muons.

To determine the expected amount of this background, a fit is performed to the J/ψ mass spectrum of the data sample using a double Crystal Ball shape for the true $J/\psi \rightarrow \mu\mu$ component and an exponential shape for the combinatorial background. The exponential slope of the background is found to be $-2.59(89) \times 10^{-4}\, \mathrm{MeV}^{-1}$. The corresponding ratio of combinatorial events within the signal region to combinatorial events in the sideband is 2.90, very nearly the simple ratio of widths $110/40 = 2.75$. An alternative fit modeling the signal distribution by a Cruijff shape (a type of bifurcated gaussian with wide tails) yields a slightly different result. In this case ratio of events within the signal region to those in the sideband is 3.360 91. Other modifications to the fit model (*e.g.*, a wide signal gaussian) could conceivably cause similar decreases in this ratio. There is thus a large systematic uncertainty on this scaling factor. The nominal value of this scaling is set to 2.9.
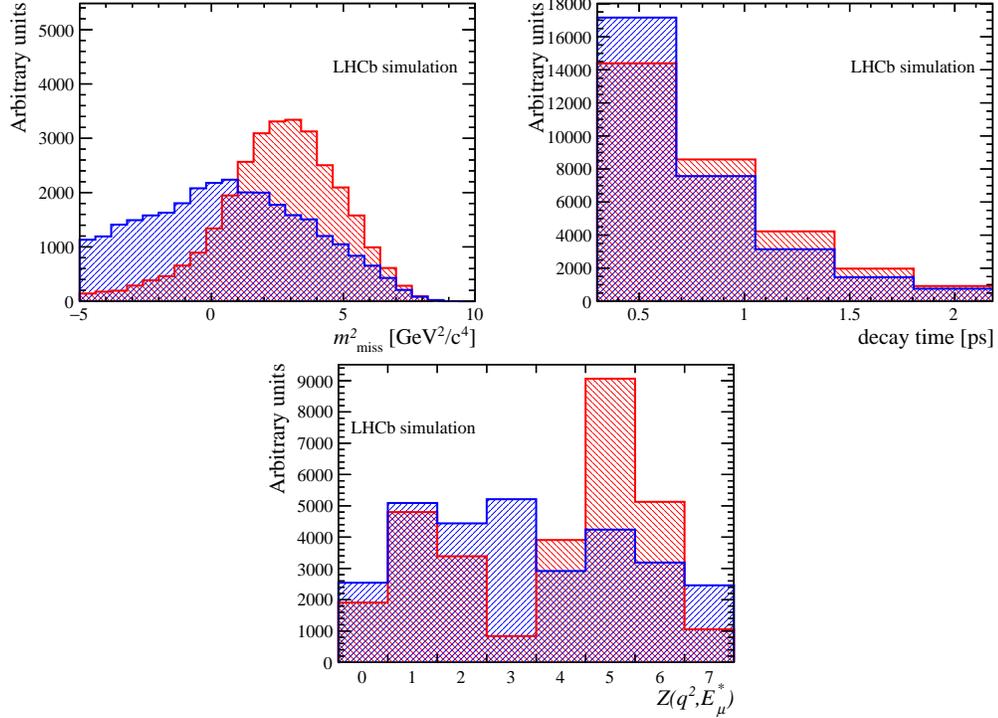
Figure 4.24: Fit templates for combinatorial J/ψ candidates; the signal template is shown in red for comparison.

The systematic uncertainty on this scaling is thus 16%. The effect of this systematic uncertainty on the value of $\mathcal{R}(J/\psi)$ is considered in Section 4.11.9.

4.8 Misidentification background

The predominant background in this analysis arises from b-hadron decays to $J/\psi X$ in which some hadron is misidentified as a muon. These decays occur frequently enough to compensate for the low rate at which non-muons pass muon PID criteria. This is referred to as the misidentification, or mis-ID, background.

A data-driven model for the background is created by repeating the signal reconstruction, but with the unpaired muons replaced with tracks not identified as muons. This control sample is enriched in various hadron and lepton species — pions, kaons, protons, and electrons. With proper weights, representing the probabilities for hadron and electron tracks to pass muon PID criteria, this sample

would be representative of the mis-ID background in the nominal data sample. These probabilities differ for the separate species of particles, and also depend on the kinematic properties of the tracks themselves and to some extant on the global properties of the events in which they are present.

To this end, the LHCb collaboration produces and maintains high purity samples of each particle species that leaves observable tracks (*e.g.*, muons, electrons, pions, kaons, and protons) [83]. For example, muons are tagged in $J/\psi \to \mu^+ \mu^-$ decays, and pions and kaons are tagged in $D^0 \to K^- \pi^+$ decays. Hereafter referred to as the particle identification (PID) samples, these are used to measure the efficiencies of PID selection criteria and misidentification rates.

Additionally, it was discovered that presence of the dimuon pair from the J/ψ in the final state of the decays studied in this analysis has an effect on the muon PID criteria of the third track in the final state that is not reproduced in the collaboration's PID samples of pions and kaons. This effect is greatest at high momenta, when the three particles are collimated. Then, the third track can "steal" muon chamber hits from the two real muons, and is more readily identified as a muon. This issue is circumvented by using an alternate PID sample of pions and kaons tagged in $B^0 \to J/\psi K^*(892)^0$ events, where the $K^*(892)^0$ undergoes the decay $K^*(892)^0 \to K^+ \pi^-$. This sample does not have the statistical power of the large collaboration samples from $D^0 \to K^- \pi^+$. Fortunately, the muon PID efficiency can be factored into a primary selection in which the above effect is pronounced and a conditional secondary selection in which this effect cancels out; the custom PID sample is used to evaluate the former probability while the collaboration samples are used to evaluate the conditional latter probability.

Another necessary set of information is not provided by the collaboration PID samples. The $J/\psi h^+$ control sample contains some fraction of ghost tracks — artifacts of reconstruction that do not correspond to real particles. This class

of tracks must be accounted for during the unfolding and weighting procedure. Unfortunately, accurate models of the PID distributions for ghost tracks, to obtain the probabilities $P(\hat{h} \mid g)$, are difficult to produce or acquire, since there are no pure samples of ghost tracks that can be taken from data. The best that can be done is to take PID efficiencies for ghost tracks by fiat or from simulation. A large sample of simulated $J/\psi\, h^+$ candidates, about 60 million such simulation events are used across a representative range of detector simulation conditions, is collected, combined from various sources, and processed with the nominal selection, save (for now) for the PID criteria on the unpaired muon. Candidates where the h^+ is a ghost tracks are selected from this sample. The sample is split into a signal-like component, where the ghost track is identified as a muon, as in the nominal candidate sample, and a control-like component, where it is not. The rates at which ghost tracks pass the muon PID criteria are obtained from the signal-like component, while the rates at which they pass the loose criteria used to sort the h^+ tracks into separate species are obtained from the control-like component. Typically, about half the ghosts fall into the ghost-tagged category $\hat{g}$, while the rest are split among mainly $\hat{\pi}$ and $\hat{K}$. The rates at which they pass the muon PID criteria are comparable to the pion fake rates $P(\pi \rightarrow \hat{\mu})$, though slightly smaller. Since the fake rates from simulation are certainly somewhat mismodeled, we make an alternative assumption for the purpose of assigning a systematic uncertainty — that all ghost tracks fall into the $\hat{g}$ category rather than simply half, and that none pass the muon PID criteria. A systematic uncertainty is assigned from examining the results of using these alternatives in the fit (see Section 4.11.7).

From the efficiencies tabulated from these various sources of PID information, weights are calculated for candidates in the control sample that appropriately represent the probability that a similar candidate's unpaired hadron would fake the PID requirements placed on the muon track. This procedure is not as straight-forward

as directly applying probabilities for tracks to pass muon PID criteria taken from the PID samples: the unpaired hadrons in the control sample belong to different species, each of which have different mis-ID probabilities, and cannot be cleanly decomposed. Instead, hadron PID criteria are used to break the control sample into six categories: pion-enriched, kaon-enriched, proton-enriched, electron-enriched, muon-enriched (contamination in the control sample), and ghost-enriched (any track failing all other criteria). The PID criteria for this decomposition are disjoint, and produce relatively pure enriched subsamples with small cross-contamination (save for the ghost-enriched category). The matrix of probabilities for tracks from the six species to fall into these six categories are obtained from the PID samples. A Bayesian unfolding technique is then used to decompose the control sample of $J/\psi\, h^+$ candidates into components representing each hadron species; only then are the muon mis-ID probabilities for these hadrons applied. Another correction factor accounts for the relative probabilities for the hadron tracks to end up in either the nominal fit sample or the control sample. The mathematical details of the calculation of these weights is described in detail in Appendix C.

The projection of the template for the mis-ID background derived from these control samples is shown in Fig. 4.25. One important feature of this background is that it contains the fully reconstructed $B^+ \rightarrow J/\psi\, K^+$ decay, whose distribution peaks in invariant mass, missing mass, and E_μ^*. Another is that, since it is made up predominantly of J/ψ mesons from light b-meson decays, its decay time distribution has a much longer tail than the B_c^+ decay modes, allowing discrimination in the fit.

No constraint is applied to the normalization of the mis-ID sample in the fit; the number of mis-ID events is determined entirely by the distribution's unique profile. To account for systematic uncertainties that depend on the particle species (due to both the PID samples and the fit selection, where particle species is correlated with flight distance and thereby with decay time), the separate categories of tagged
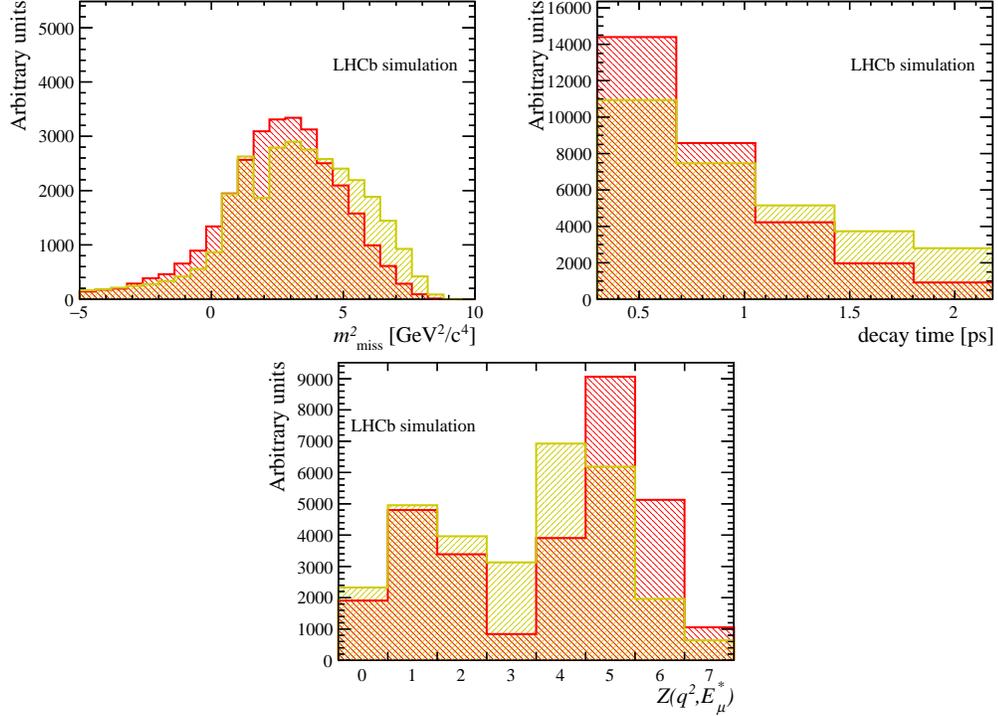
Figure 4.25: Fit templates for misidentification background; the signal template is shown in red for comparison. The peaking structure evident in the missing mass squared is $B^+ \to J/\psi\, K^+$ background.

species hypothesis (*e.g.*, pion-like, kaon-like, etc.) are allowed to vary separately in the fit. A common scale factor with respect to the expected number of events is assumed for all categories. The scale factors for the non-pion-like categories are allowed to deviate from this scale factor, under a Gaussian penalty with width 20%.

4.9 Corrections to Simulation

The simulated data produced for this analysis is corrected to better describe real data in several different ways.

4.9.1 Empirical correction in normalization-rich region of data

The Monte Carlo generator `BCVEGPY`, an extension of `Pythia6`, does not accurately model the event multiplicity of real pp collisions. An important factor in

the performance of the muon PID selection is the event multiplicity, defined as the number of reconstructed long tracks in the event. Events with high multiplicity are noisier, and in general have poorer PID performance; an example of this effect is shown in Fig. 2.6. The simulation must be weighted to match the event multiplicity distribution in data so that the effects of the PID selection on the fit variables are accurately modeled.

The analysis that led to the first measurement of $R(D^*)$ [20] found that simulation mis-modeled the D^0 impact parameter significance and transverse flight distance, and that these variables were correlated with the measured q^2 of the candidate. For this analysis, the J/ψ impact parameter significance is highly correlated with the B_c^+ decay time. The unpaired muon impact parameter significance, also correlated with the decay time, is expected to be similarly mis-modeled.

All of these potential mis-modelings are empirically corrected by comparing $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ simulation and a normalization-rich region of data, defined by $m_{\mathrm{miss}}^2 < 0\,\mathrm{GeV}^2$, decay time $\tau < 1.24\,\mathrm{ps}$, and an isolation BDT score < 0. The data sample contains a non-negligible amount of mis-ID background; this is subtracted using the mis-ID template described in Section 4.8.[6] The joint distributions of these three variables and the B_c^+ decay time in both simulation and the normalization-rich region of data are shown in Fig. D.1. As expected, there are strong correlations between the impact parameter significances and the B_c^+ decay time, as well as a slight correlation between the J/ψ and μ^+ significances. In order to correct the track multiplicity and IP significance distributions while not altering the decay time distribution, all four variables must be considered in the reweighting.

Several issues must be considered when generating four-dimensional corrections from the same datasets. While the track multiplicity is mostly uncorrelated

---

[6]The subtraction is performed with an overall scaling of 1.5, corresponding roughly to the amount that the mis-ID template is scaled up in the fit; separate weightings are generated with the mis-ID subtracted with scalings of 1 and 2 to assign a systematic uncertainty.

with the other fit variables, creating separate weights from the same dataset runs the risk of unwanted interference. Conversely, a simultaneous four-dimensional correction using binned data can suffer from low statistics. Instead, we use a boosted decision tree tool for reweighting [84]. Reweighting a dataset with a single decision tree is identical to a binned reweighting with an optimal binning scheme chosen by binary partitions. Reweighting via a boosted decision tree generates a weight through a superposition of overlapping, coarse binary partitions, producing a smoother and more robust response.

A three-fold cross validation strategy is employed to test the output of the gradient-boosted reweighter. The $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ simulation and normalization-rich data are each randomly partitioned into three parts. Using one third of the data, the simulation decay time distribution is reweighted to the observed data distribution. In the second third, a gradient-boosted reweighter is trained between simulation and data using the decay time, event multiplicity, and IP significances; the inclusion of the decay time is a constraint that forces the reweighting to be unbiased. Then, the final third of the data is used as a test sample on which the reweighting is applied. This process is repeated twice with the three data/MC partitions permuted. Figures D.2 to D.5 in Appendix D show, for one fold, the distributions of the four variables on the test sample before and after applying the weights determined by the BDT (in each case with the one-dimensional decay time reweighting applied first). The gradient-boosted reweighting successfully corrects the three target distributions while leaving the decay time unaffected; no excessively large weights are produced. The outputs of the three BDT algorithms are applied to all the generated simulation and averaged.

4.9.2 Lifetime acceptance correction

The simulation models not only the lifetime of the B_c^+ meson but also the more complicated lifetime acceptance of the LHCb detector. The reconstruction efficiency of B_c^+ meson decays is time-dependent, with separate effects at short lifetimes and long lifetimes. At short lifetimes, selections requiring minimum IP χ^2 for decay product tracks and the challenge of separating secondary vertex tracks from primary vertex tracks leads to an efficiency that increases from essentially 0 at $t = 0$ to near unity after $\sim 1\,\mathrm{ps}$. At longer lifetimes, secondary vertices are well separated from the PV, but the reconstruction efficiency has a small linear trend.

To incorporate possible correlation of the decay time acceptance with the missing momentum in the partially reconstructed decay, the correction is made using partially reconstructed $B^0 \to J/\psi K^*$ decays. First the $J/\psi K$ pair is reconstructed with the same kinematic and vertexing requirements as in the $B^+ \to J/\psi K^+$ case. The selection requires that there be a soft pion that puts the decay into the B^0 mass window, with no vertexing requirement; loose p_T and IP χ^2 cuts on the pion are unfortunate but necessary to reduce background. An offline selection cutting on the angle between the momentum of the B^0 and its flight direction (two largely independent measurements by the tracking and vertexing systems, whose concordance is characteristic of good signal candidates) and tightening the K^* mass window reduces background further. The invariant mass of the B^0 is fit using a double Crystal Ball shape, while the background is modeled using a second order polynomial, shown in Fig. 4.26. Weights are extracted from the fit to separate signal and background, using the *sPlot* technique [85]. The lifetime acceptance correction is performed in bins of the momentum fraction $p(\pi)/p(B^+)$, which is bound between 0 and 0.5. This quantity is analogous to $p_{\mathrm{miss}}/p(B_c^+)$ in the simulation B_c^+ decays. The acceptance function depends on $p(\pi)/p(B^+)$ in both data and MC, and the dependence is well
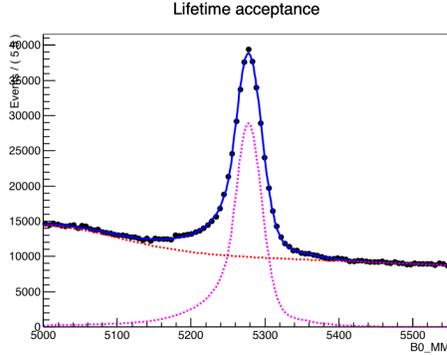
Figure 4.26: Fit to the invariant mass distribution of the $B^0 \to J/\psi K^*$ sample selected from data.

represented by the MC. Projections of the acceptance onto the lifetime in the four bins of missing momentum fraction are shown in Fig. 4.27. These samples are used as the lifetime acceptance correction.

4.9.3 PID correction

The efficiencies of PID criteria for muons in simulation are known to be mis-modeled. Rather than use these simulated rates, the efficiency of the PID requirement on the unpaired muon is computed using control samples produced by the LHCb collaboration of muons tagged in J/ψ decays, and are applied as weights to the simulation datasets.

4.10 Fit results

The results of the fit are presented in Fig. 4.28, showing the projections of the nominal fit result onto the quantities m^2_{miss}, decay time, and Z. The fit yields 1400 ± 300 signal and 19140 ± 340 normalization decays, where the errors are statistical and correlated. Accounting for the $\tau^+ \to \mu^+ \nu_\mu \overline{\nu}_\tau$ branching fraction and the ratio of efficiencies gives an uncorrected value of 0.79 for $\mathcal{R}(J/\psi)$. Correcting
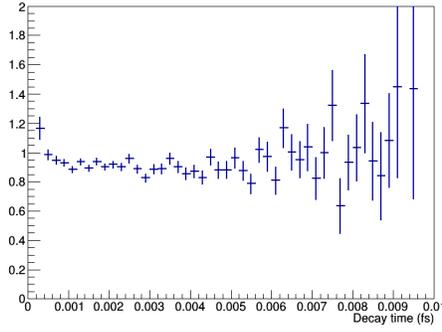
(a) $p_\pi/p_{B_c} < 12.5\%$
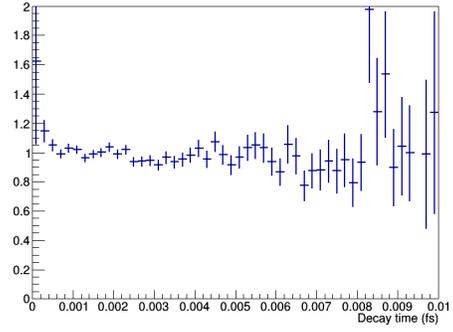
(b) p_π/p_{B_c} between 12.5% and 25%
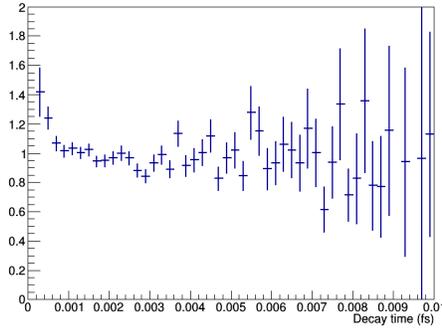
(c) p_π/p_{B_c} between 25% and 37.5%
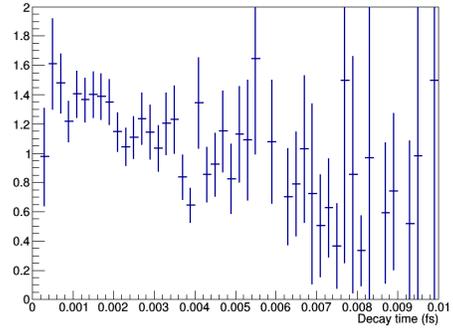
(d) p_π/p_{B_c} between 37.5% and 50%

Figure 4.27: The acceptance of $B^0 \to J/\psi\, K^*$ decays in simulation differ at short lifetimes and as a function of the missing momentum.

for the mean expected bias at this value, the measured value of $\mathcal{R}(J/\psi)$ is

$$\mathcal{R}(J/\psi) = 0.71 \pm 0.17 \,(\text{stat}) \pm 0.18 \,(\text{syst}), \qquad (4.7)$$

where the sources of systematic uncertainty are overviewed in the next section. The significance of the signal, determined from a likelihood scan procedure and corrected for the systematic uncertainty, is found to be 3 standard deviations.

Figures 4.30 and 4.31 show the projections of the nominal fit result onto $m^2_{\text{miss}}$and decay time in the eight bins of the observable Z. A histogram of the residuals is shown in Fig. 4.29 against a superimposed standard normal distribution. A table of fit results and the correlation matrix (excluding fixed variables) are shown in Tables 4.7 and 4.9.

4.11   Systematic uncertainties

The systematic uncertainty on the measured value of $\mathcal{R}(J/\psi)$ is broken down by source in Table 6.5. A description of these sources and the methodologies used to assess their corresponding uncertainties is presented in this section.

4.11.1   Simulation template statistical uncertainty

The finite simulation statistics leads to uncertainty in the template shapes and thereby the resulting fit. The effect of this systematic uncertainty is folded into the fit likelihood using the Beeston-Barlow-light procedure described in Section 4.2.2. We separate it out by comparing the quadrature difference in the uncertainty in $\mathcal{R}(J/\psi)$ from the fits with and without the Beeston-Barlow procedure enabled. After correcting for the relative selection efficiencies of signal and normalization, the Beeston-Barlow estimate of the uncertainty due to finite simulation statistics is 8.0%.
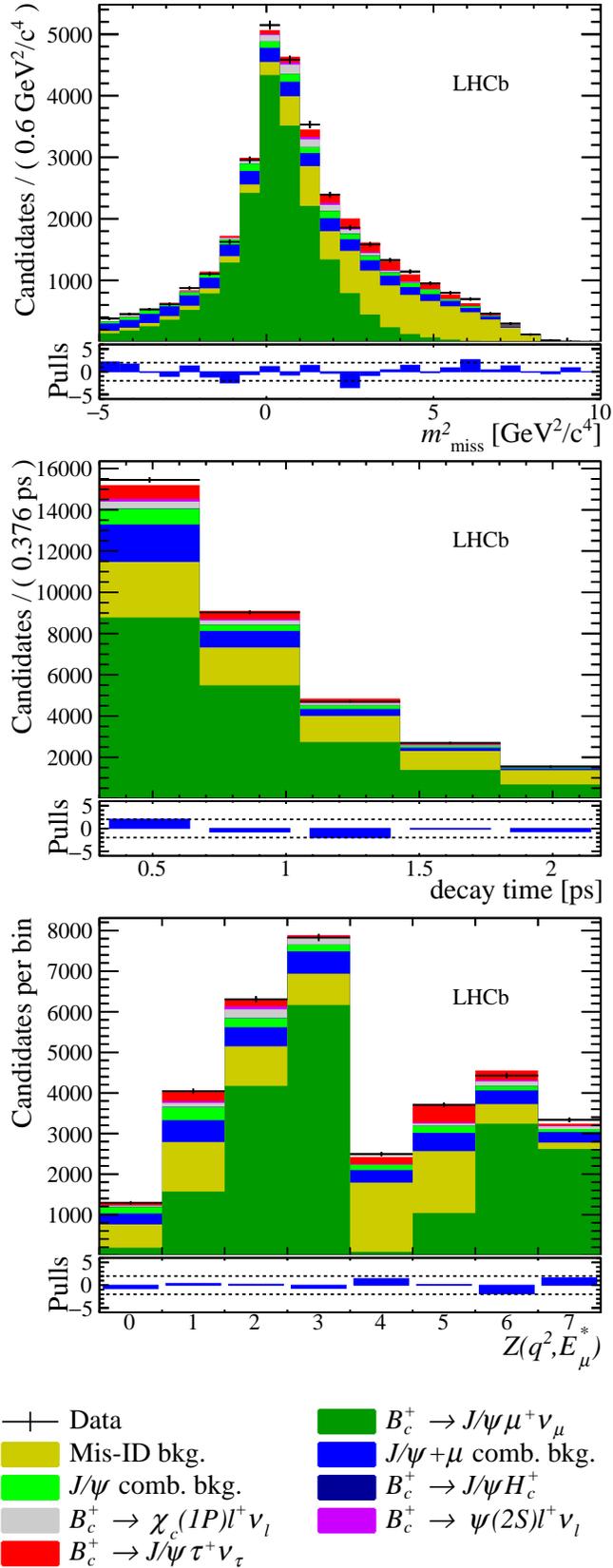
Figure 4.28: Projections of nominal fit onto m^2_{miss} (top), t (middle), and Z (bottom).

Table 4.7: Fit parameters, with statistical uncertainties that include the Beeston-Barlow uncertainty

Parameter		Value	Constraint
Num. of $J/\psi\,\mu^+\nu_\mu$		19140 ± 371	None
Num. of $J/\psi\,\tau^+\nu_\tau$		1400 ± 300	None
Raw $\psi(2S)\mu\nu$-to-$J/\psi\,\mu\nu$ fraction		0.020 ± 0.040	$\geq 10^{-6}$
Raw $\chi_c\mu\nu$-to-J/ψ fraction		0.111 ± 0.082	$\geq 10^{-6}$
Raw $J/\psi DX$-to-$J/\psi\,\mu\nu$ fraction		0.002 ± 0.023	$\geq 10^{-6}$
$\mathcal{R}(\psi(2S))$		0.085	Fixed
Num. comb. bkg.		3135 ± 451	None
Comb. J/ψ scaling		2.9	Fixed
Multiplicative B^+ correction		0.24 ± 0.92	Gaussian
Multiplicative B^0 correction		0.07 ± 0.93	Gaussian
Multiplicative B_s^0 correction		-0.37 ± 0.97	Gaussian
Mis-ID scale		1.862 ± 0.093	None
Kaon-like scale correction		-3.27 ± 0.58	Gaussian
Proton-like scale correction		0.63 ± 1.01	Gaussian
Electron-like scale correction		-0.20 ± 1.00	Gaussian
Muon-like scale correction		0.20 ± 0.98	Gaussian
Uncategorized scale correction		1.05 ± 1.03	Gaussian
B_c^+ lifetime		$(0.5087 \pm 0.0034)\,\text{ps}$	$\mathcal{N}(0.5113\,\text{ps}, (0.0094\,\text{ps})^2)$ [12]
$\psi(2S)$ ρ^2 correction factor		-0.10 ± 0.95	$\mathcal{N}(0, 1^2)$
χ_{c1} ρ^2 correction factor		-0.45 ± 0.97	$\mathcal{N}(0, 1^2)$
χ_{c2} ρ^2 correction factor		-0.25 ± 0.93	$\mathcal{N}(0, 1^2)$
$A_0(q^2)$	b_0	1.0 ± 1.0	None
	b_1	-9 ± 21	None
$A_1(q^2)$	b_0	0.652	Fixed
	b_1	-2.36 ± 1.56	Fixed from norm. rich region
$A_2(q^2)$	b_0	1.40 ± 0.18	Fixed from norm. rich region
	b_1	0	Fixed
$V(q^2)$	b_0	1.29 ± 0.18	Fixed from norm. rich region
	b_1	-15.7 ± 9.5	Fixed from norm. rich region
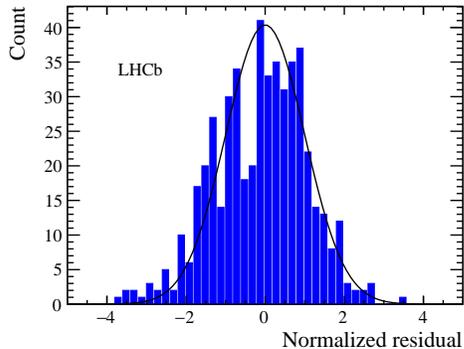
Figure 4.29: Histogram of the normalized residuals in each un-projected bin of the nominal fit. The red curve plots a standard normal distribution.

Table 4.8: Raw numbers of events of each category from the fit, with statistical uncertainties that include the Beeston-Barlow uncertainty

Component	Value	Details
$J/\psi\,\tau\nu$	1398 ± 332	
$J/\psi\,\mu\nu$	19140 ± 371	
$\psi(2S)\mu\nu$	231 ± 464	
$\psi(2S)\tau\nu$	3 ± 7	Constrained to 0.085×0.177 of above
$\chi_c\mu\nu$	731 ± 537	Half χ_{c1}, half χ_{c2}
$J/\psi H_c X$ bkg.	10 ± 148	Half two-body and half quasi-two-body
Comb. bkg.	3135 ± 451	
Comb. J/ψ bkg.	1210	Fixed
Mis-ID events	7437 ± 371	

Table 4.9: Correlation matrix in nominal fit

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 : $\mathcal{R}(J/\psi)$	1.00	-0.03	-0.26	-0.00	-0.11	-0.09	-0.03	0.02	0.01	-0.43	-0.12	-0.05	0.05	-0.18	-0.03	-0.05
2 : $\#(J/\psi\mu\nu)$	-0.03	1.00	-0.36	-0.36	0.17	-0.20	-0.01	0.02	-0.01	0.16	-0.21	0.04	-0.03	-0.02	0.03	-0.09
3 : $f(\psi(2S))$	-0.26	-0.36	1.00	-0.38	-0.34	0.08	0.01	-0.01	0.00	-0.11	0.28	0.08	-0.07	0.07	0.03	-0.08
4 : $f(\chi_c)$	-0.00	-0.36	-0.38	1.00	0.04	0.04	-0.01	-0.00	0.01	-0.02	-0.02	0.02	-0.01	0.01	-0.07	0.25
5 : $f(J/\psi DX)$	-0.11	0.17	-0.34	0.04	1.00	-0.16	0.03	0.00	-0.04	-0.14	-0.18	-0.05	0.05	0.11	0.01	0.04
6 : $\#$(Comb)	-0.09	-0.20	0.08	0.04	-0.16	1.00	-0.73	-0.73	-0.31	-0.10	0.13	-0.04	0.03	-0.02	-0.00	0.01
7 : B^0 fraction	-0.03	-0.01	0.01	-0.01	0.03	-0.73	1.00	0.38	0.15	-0.02	0.03	-0.00	0.00	-0.00	-0.00	-0.00
8 : B^+ fraction	0.02	0.02	-0.01	-0.00	0.00	-0.73	0.38	1.00	0.15	0.01	-0.02	0.00	0.00	0.00	0.00	0.00
9 : B_s fraction	0.01	-0.01	0.00	0.01	-0.04	-0.31	0.15	0.15	1.00	0.02	-0.01	0.00	-0.00	0.00	0.00	0.00
10 : Mis-ID scale	-0.43	0.16	-0.11	-0.02	-0.14	-0.10	-0.02	0.01	0.02	1.00	-0.11	-0.04	-0.02	0.08	0.01	0.00
11 : $\tau(B_c^+)$	-0.12	-0.21	0.28	-0.02	-0.18	0.13	0.03	-0.02	-0.01	-0.11	1.00	-0.04	0.03	0.02	-0.00	0.01
12 : A_0 int.	-0.05	0.04	0.08	0.02	-0.05	-0.04	-0.00	0.00	0.02	0.03	-0.04	1.00	0.10	-0.01	0.00	0.01
13 : A_0 slope	0.05	-0.03	-0.07	-0.01	0.05	0.03	0.00	0.00	-0.00	-0.02	0.03	0.10	1.00	0.01	-0.00	-0.01
14 : $\psi(2S)\ \rho^2$	-0.18	-0.02	0.07	0.01	0.11	-0.02	-0.00	0.00	0.00	0.08	0.02	-0.01	0.01	1.00	-0.02	-0.02
15 : $\chi_{c1}(1P)\ \rho^2$	-0.03	0.03	0.03	-0.07	0.01	-0.00	-0.00	0.00	0.00	0.01	-0.00	0.00	-0.00	-0.02	1.00	-0.02
16 : $\chi_{c2}(1P)\ \rho^2$	-0.05	-0.09	-0.08	0.25	0.04	0.01	-0.00	0.00	0.00	0.00	0.01	0.01	-0.01	-0.02	-0.02	1.00
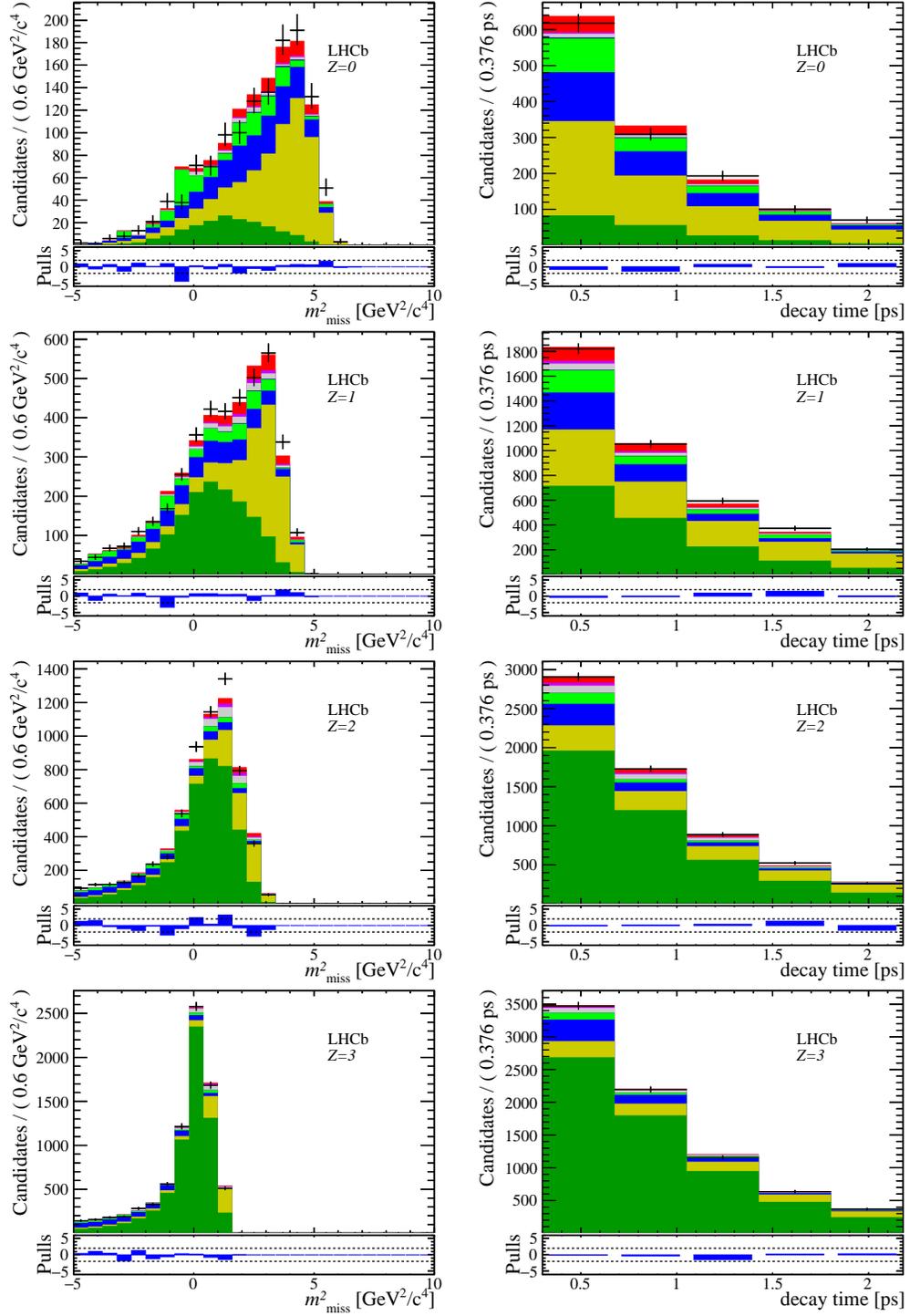
Figure 4.30: Projections of the nominal fit in bins 0–3 of Z, *i.e.*, individual bins of q^2 and E_μ^*.

Table 4.10: Systematic uncertainties in the extraction of $\mathcal{R}(J/\psi)$. The errors in the multiplicative uncertainties on the signal-to-normalization efficiency ratio are evaluated at the benchmark value $\mathcal{R}(J/\psi) = 0.25$.

Model uncertainties	Size (eff. corrected) $(\times 10^{-2})$
MC stat. uncertainty	8.0
$B_c^+ \to J/\psi$ form factors	12.1
$B_c^+ \to \psi(2S)$ form factors	3.2
Bias correction	5.4
$B_c^+ \to J/\psi \, DX$ cocktail composition	3.6
Z binning strategy	5.6
Misidentification background strategy	5.4
Combinatorial background cocktail	4.5
Combinatorial J/ψ sideband scaling	0.9
Empirical reweighting	1.6
Semitauonic $\psi(2S)$ and χ_c feed-down	0.9
Fixing $A_2(q^2)$ slope to zero	0.3
Efficiency ratio	0.6
$\mathcal{B}(\tau \to \mu\nu\nu)$w	0.2
B_c^+ lifetime	included in stat.
Total systematic uncertainty	**17.7**
Stat. uncertainty	**17.3**

While the Beeston-Barlow-light procedure will be used in producing the final unblinded fit result, it may underestimate the uncertainty. Since the light procedure does not keep track of separate fluctuations for the individual templates, but only the summed template, it accounts mainly for uncertainties in the largest fit components, *e.g.*, the normalization and mis-ID backgrounds. To this end, we perform a toy study in which we create bootstrapped fit templates — sampling the individual templates for the various modes with replacement to create alternative templates with equal statistics but with Poisson fluctuations. Fitting the data using a sequence of these bootstrapped templates, we find an uncertainty compatible with the Beeston-Barlow estimate.

The finite simulation statistics also introduces a systematic uncertainty in the signal-to-normalization efficiency ratio, since this ratio is evaluated using the simulation samples. The statistical uncertainty on the absolute efficiency ratio is

0.4%, which corresponds to a relative multiplicative uncertainty of 0.7%. At a "SM" value for $\mathcal{R}(J/\psi)$ of 25%, this would be an absolute uncertainty of 0.2%.

We investigated whether targeted simulation requests could reduce this systematic uncertainty. We fit our model to toy data generated from our fit templates with five times the data statistics. In addition to fits with and without Beeston-Barlow, which had consistent results, we performed fits in which various templates were given simulated infinite statistics (by turning Beeston-Barlow off for this component). The motivation was to see if more statistics in any one template would have an outsize effect. However, the main effects of these modifications were changes in the estimate of $\mathcal{R}(J/\psi)$ (much greater than the change between fits with Beeston-Barlow completely off and on) which obscured any pattern of improvements in the Beeston-Barlow uncertainty. One possible explanation is that when templates are given Beeston-Barlow uncertainties while others are not, the former are favored in the likelihood maximization because of their extra degrees of freedom, warping the result. In short, this study provided no insight into how a targeted simulation request could be made.

4.11.2 $B_c^+ \to J/\psi$ form factors

The $B_c^+ \to J/\psi$ form factors that strictly govern the normalization $B_c^+ \to J/\psi \mu^+ \nu_\mu$ decay, those pertaining to $V(q^2)$, $A_1(q^2)$, and $A_2(q^2)$, are determined in a fit to a normalization-rich region, as described above in Section 4.3.1, and then fixed in the main fit. (Two other parameters, related to the form factor $A_0(q^2)$, govern only the kinematics of the semitauonic signal decays and float in the main fit). The systematic uncertainty is assessed by performing an alternate fit in which all the form factor parameters are allowed to vary. In this alternate fit, the value of $\mathcal{R}(J/\psi)$ shifts by 6.3%, while the extra uncertainty in the parameter (in quadrature) is 7.3%. We take the maximum of these two values as the systematic uncertainty.

Another systematic uncertainty is due to the choice to fix the BCL slope parameter of $A_2(q^2)$ to zero, as discussed in Section 4.3.1. An alternate fit is performed, where this parameter is allowed to float in the full fit. The resulting shift in $\mathcal{R}(J/\psi)$ is 0.6%, and so a systematic uncertainty of 0.3% is applied.

4.11.3 $B_c^+ \to \psi(2S)$ form factors

The theoretical predictions for the ratio of $\mathcal{B}(B_c^+ \to \psi(2S)\mu^+\nu_\mu))$ to $\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu_\mu)$ are in the range 2.5 % to 5 % [57, 63]. Accounting for the relative selection efficiency (65%) and inclusive $\psi(2S) \to J/\psi\,X$ branching fraction (60.8%), the fraction of $\psi(2S)$ feed-down events relative to normalization events is 1 % to 2 %. In the nominal fit, using the EFG model, the measured $\psi(2S)$ fraction is 2.0(40) %, consistent with the above predictions. The Kiselev model for the $B_c^+ \to \psi(2S)\mu^+\nu_\mu$ feed-down decay, however, is quite distinct. Since simply reweighting the simulation cannot account for the difference, we perform resample generator-level simulation to account for detector effects, as described in Section E.1. With the resulting Kiselev model, the $\psi(2S)$ fraction in the fit is measured to be 2.8(31) %. The corresponding shift in the measured value of $\mathcal{R}(J/\psi)$ is 6.3%, substantial though much less than the statistical uncertainty. Half this value is assigned as a systematic uncertainty (see Section 4.11).

4.11.4 Bias correction systematic

As described in Section 4.8, a "tinker toy" study compares second- and third-generation fits. In particular, Fig. 4.8 shows the difference between third-generation toy-study measurements with respect to parent second-generation toy-studies from which their templates are derived. This difference is on average 5.4% and is independently of $\mathcal{R}(J/\psi)$, and so 5.4% is assigned as a corresponding systematic uncertainty.

4.11.5 $B_c^+ \to J/\psi DX$ background

The $B_c^+ \to J/\psi DX$ cocktail has several possible sub-components: two-body $J/\psi D^{(*)}$ final states, three-body $J/\psi DK$ final states, and quasi-two body $J/\psi D^{(**)}$ final states. The final two sub-components are not exclusive, and the relative fraction of three-body final states are non-resonant or proceed through a $D^{(**)}$ resonance is not precisely determined, though in analogy with decays of lighter B mesons the quasi-two-body decays is strongly expected to dominate. In the fit, 50% of the contribution is assumed to come from quasi-two-body decays and the remainder from direct two-body decays.

An earlier iteration assigned a very conservative systematic by replacing the quasi-two-body component with a three-body component created with the unphysical phase space decay model. However, with the coarser bins in the new fit model, the fit cannot easily distinguish the three-body decays from mis-ID or other components, and pushes their contribution up to produce an unrealistic fit. Instead, previous LHCb and ATLAS measurements constrain the expected amount of $B_c^+ \to J/\psi DX$ background as a fraction of the number of normalization events to $2.1(3)\%$ [81, 82, 86]. The fit value of this fraction is near 0 but with large uncertainties that make it compatible with the expectation. We now assign the systematic uncertainty by this nominal fit to an alternative with the $B_c^+ \to J/\psi DX$ fraction constrained around the expected fraction. Half the shift in $\mathcal{R}(J/\psi)$, or 3.6%, is assigned as a systematic uncertainty.

4.11.6 Z binning strategy

The fit variable Z, which corresponds to four bins in E_μ^* and two bins in q^2, is very coarse, and the value of $\mathcal{R}(J/\psi)$ could be sensitive to the binning strategy. Replacing the four bins in E_μ^* with only two bins leads to a 4.1% shift in the value

of $\mathcal{R}(J/\psi)$, while replacing it with three bins leads to a 9.3% shift. Doubling the number of E^*_μ bins leads to an 11.1% shift. On the other hand, keeping the number of bins constant but moving the bin boundaries, so that one E^*_μ bin straddles the lowest values and highest values, and ditto for one q^2 bin, the shift in $\mathcal{R}(J/\psi)$ is 0.8%. We assign half the larger shift as a systematic uncertainty of 5.6%.

4.11.7 Misidentification background strategy

Section 4.8 describes two ways of modeling the effect of ghost tracks. The fit procedure is performed with templates from each of these methods, and a systematic uncertainty is assigned from half the difference between the two fit minima.

4.11.8 Combinatorial background cocktail

The combinatorial background is produced entirely with simulation because no data-driven model can be made. A cross-check can be made by performing a fit in the high mass sideband. A combinatorial background enriched sideband region can be made by requiring the invariant mass of the J/ψ and μ^+ to be greater than 6400 MeV and eliminating several cuts in the normal selection designed to veto combinatorial background (the cuts on the angle between the J/ψ and the μ^+ in the transverse plane and the DOCA between the two particles, and the requirement that the missing mass squared be greater than -5 GeV$^2/c^4$). A fit to the invariant mass distribution in this region is shown in Fig. 4.32. In the fit, a shape systematic is included that corresponds to weighting the simulation invariant mass (M) distribution by

$$1 + \alpha \frac{M - 6277}{4(10000 - 6277)}.$$
(4.8)

When the Beeston-Barlow algorithm is enabled, α is measured to be -1.3 ± 0.6. This correction is applied to the nominal fit, and half the difference with the fit

with no correction (4.5% after the efficiency correction) is assigned as a systematic uncertainty.

4.11.9   Combinatorial J/ψ background

The combinatorial J/ψ background is modeled by data taken from the sidebands of the J/ψ invariant mass distribution, as described in Section 4.7. A systematic uncertainty is assigned by varying this scaling around its central value by its systematic uncertainty (16%). The average difference in the value of $\mathcal{R}(J/\psi)$ between the nominal fit and fits with the number of combinatorial J/ψ events scaled up and down by this uncertainty is 0.9% (absolute) after the efficiency correction. This is assigned as the corresponding systematic uncertainty.

4.11.10   Simulation corrections

The joint distribution of the event multiplicity and J/ψ and unpaired muon IP significances in simulation is reweighted using a data-driven correction derived from a normalization-rich region of the signal sample (Section 4.9.1). There are several arbitrary choices in the definition of this normalization-rich region. First, the expected mis-ID contribution is subtracted from data using a scaling factor of 1.5, chosen by peaking ahead at the fit results, rather than using the a priori expected normalization (*i.e.*, no scaling factor). Subtracting the mis-ID with no scaling factor leads to an absolute 0.25% shift in the value of $\mathcal{R}(J/\psi)$. Other arbitrary choices are the criteria used to define the normalization-rich region: the upper isolation BDT value (0.0), missing mass squared (0.0 GeV$^2/c^4$), and lifetime (1.24 ps). Relaxing the BDT cut to the cut used in the nominal fit (0.2) leads to an absolute 1.2% shift in the value of $\mathcal{R}(J/\psi)$; while relaxing the missing mass squared upper bound to 0.4 GeV$^2/c^4$ leads to a 2.6% shift and relaxing the decay time upper bound to 2.18 ps leads to a 1.3% shift. Taking the half the sum in quadrature of these shifts,

we assign an absolute 1.6% systematic uncertainty on $\mathcal{R}(J/\psi)$.

4.11.11 Semitauonic $\psi(2S)$ and χ_c feed-down

In the fit, we take $\mathcal{R}(\psi(2S)) = 8.5\%$, as described above 4.4.1. If this ratio were increased by half and all the excess were included in the signal component of the fit, the number of raw signal events would increase by at most $0.5 \times 8.5\% \times 3\%$ of the number of normalization events. The effect on $\mathcal{R}(J/\psi)$, accounting for the $\tau \to \mu\nu\nu$ branching fraction, would be an absolute increase of 0.7%.

Similarly, for the χ_c states, whose semitauonic decays are ignored in the fit, the total number of semitauonic decays as a fraction of the number of normalization decays is less than 0.09%. The corresponding absolute increase in $\mathcal{R}(J/\psi)$ were all this to be included in the signal component would be 0.5%.

The combination in quadrature of these two possible increases, 0.9%, is added as a systematic uncertainty.

4.11.12 B_c^+ lifetime

The B_c^+ lifetime is allowed to vary in the fit through the use of template shape systematics. The simulation model is reweighted to the PDG average of the B_c^+ lifetime and varies within its uncertainty. The systematic uncertainty could be assessed by the difference in quadrature between a fit with the lifetime floating and a fit with the lifetime fixed to its converged value. Currently, this effect is simply included in the statistical uncertainty from the fit.

4.11.13 Toy studies

To investigate a possible bias or under/overinflation of the statistical uncertainty from the maximum likelihood fit, we perform a study in which 500 toy datasets

are generated from the nominal (*i.e.*, best-fit) fit parameters, but with the blinded $\mathcal{R}(J/\psi)$ value replaced with the SM value 0.25 (or 0.15 after accounting for the efficiency ratio). The fit, without the Beeston-Barlow procedure, is performed for each of these toys. We record the pulls of $\mathcal{R}(J/\psi)$ with respect to the SM value, both for the symmetric Minuit uncertainty and the asymmetric MINOS uncertainty. Histograms of these pulls and their quantile-quantile plots are shown in Figs. 4.33 and 4.34. These pulls show no significant bias. Their distributions are non-normal, having thinner tails and standard deviations < 1, but this is to be expected. In particular, $\mathcal{R}(J/\psi)$ must be a positive quantity, severely constraining the lower tail. These pulls indicate that a likelihood scan will possibly yield a narrower confidence interval on $\mathcal{R}(J/\psi)$ than the naive extrapolation from the statistical uncertainty.

4.12   Conclusion

A measurement of the ratio

$$\mathcal{R}(J/\psi) = \frac{\mathcal{B}(B_c^+ \to J/\psi\,\tau^+\nu_\tau)}{\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu_\mu)}, \tag{4.9}$$

is performed on the $3\,\mathrm{fb}^{-1}$ Run 1 LHCb dataset. The analysis uses a template fit method, with templates taken from MC and data for the signal and backgrounds. The measured value is $0.71 \pm 0.17\,(\mathrm{stat}) \pm 0.18\,(\mathrm{syst})$. This is the first measurement of the decay $B_c^+ \to J/\psi\,\tau^+\nu_\tau$, and the significance of the measurement is about 3σ, including the effects of the slightly non-quadratic likelihood profile and systematic uncertainties. This value is excess of the range of SM predictions, 0.25–0.28, by 1.7 standard deviations. Though the strength of evidence that the SM prediction is low is weak, the deviation is in the same direction as stronger deviations measured in $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$. More definitive results may be found in analyses of Run 2 data collected by the LHCb detector in 2015–2017.
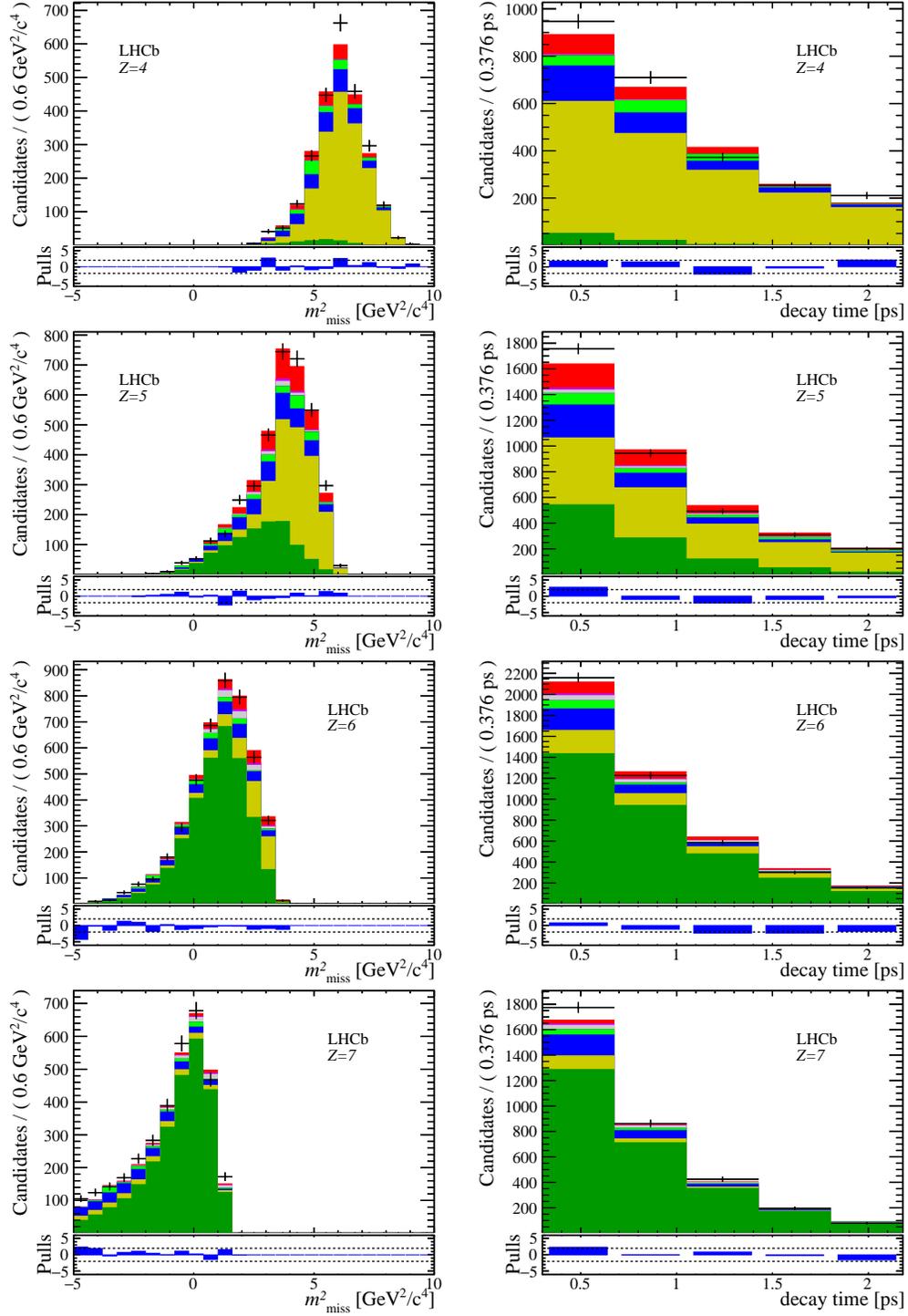
Figure 4.31: Projections of the nominal fit in bins 4–7 of Z, *i.e.*, individual bins of q^2 and E_μ^*.
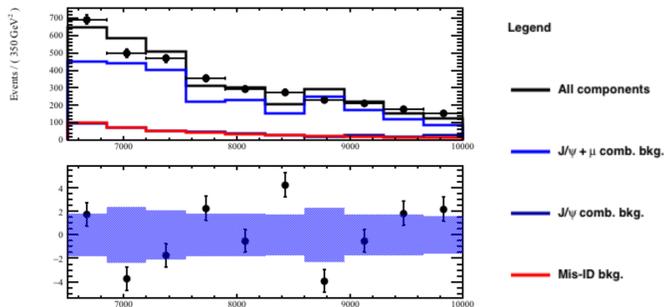
Figure 4.32: Fit to combinatorial-rich high mass sideband region

(a) Histogram

(b) Quantile-quantile plot
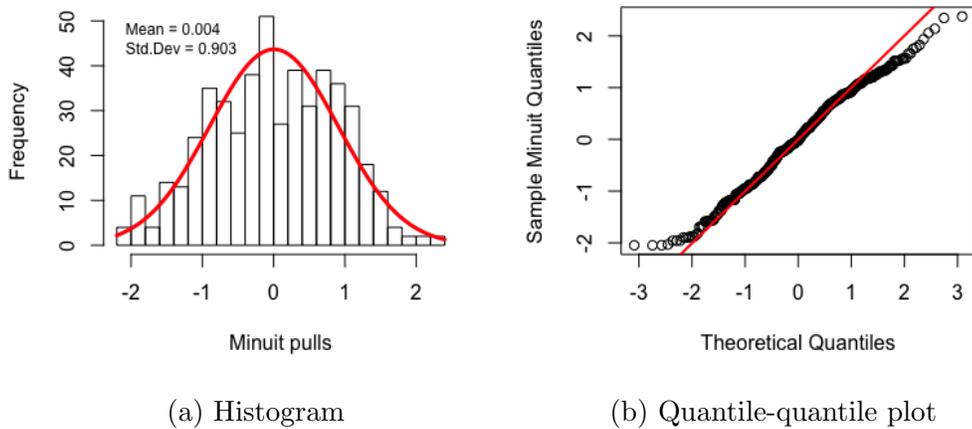
Figure 4.33: The Minuit pull distribution is unbiased and is close to normal between -2 and 1. It has thinner tails than a standard normal distribution.
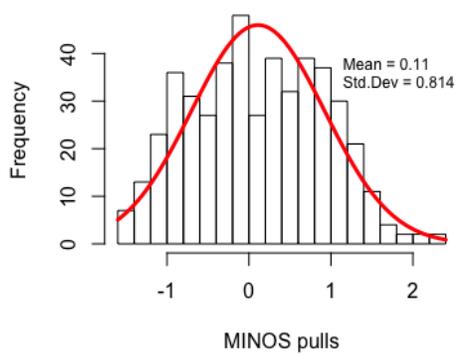
(a) Histogram
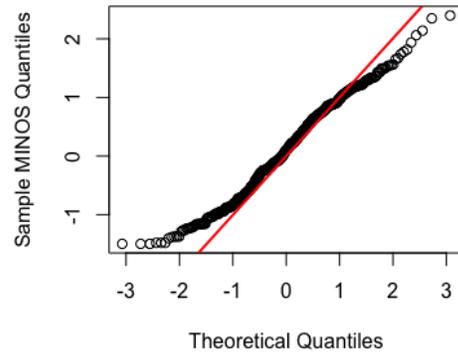
(b) Quantile-quantile plot

Figure 4.34: The MINOS pull distribution has a very small positive bias and is somewhat close to normal between -1 and 1. It has thinner tails than a standard normal distribution.

Part III

CP violation and flavor tagging

Chapter 5:   Measuring *CP* violation at LHCb

Since the theoretical development of the CKM mechanism, a major experimental effort has been undertaken to measure that parameters V_{ij} of the CKM matrix. Since the CKM matrix is unitary, its 9 measurable coefficients are over-defined and are a function of four real quantities. This can be checked by forming a unitarity triangle — three complex quantities that, if the CKM matrix is unitary, form the vertices of a triangle. The two most common unitarity triangles are shown in Fig. 5.1, and include a number of angles, most importantly α, β, γ, and β_s, and lengths. If the measured values are inconsistent with unitarity, this could indicate physics beyond the Standard Model that predicts new sources of *CP* violation.[1]

---

[1] The simplest example is extra quark generation(s), in which case the CKM matrix would be a 3×3 submatrix of a unitary $N \times N$ matrix
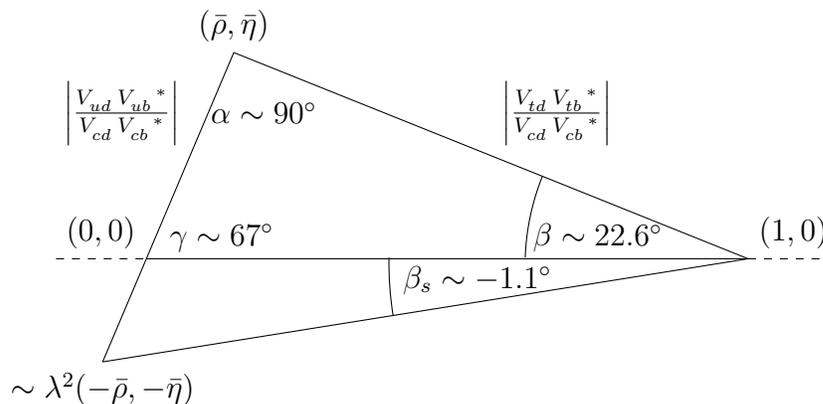
Figure 5.1: Standard unitarity triangles [87]
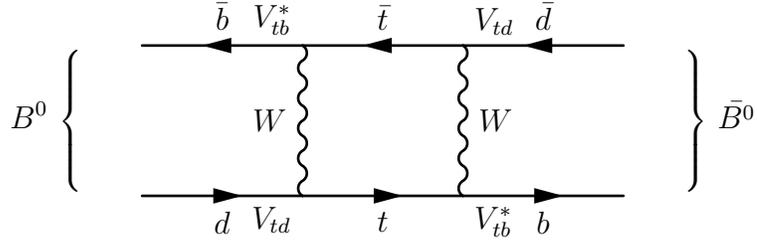
Figure 5.2: Diagram for B^0–$\bar{B}^0$ mixing. A similar diagram exists wherein the top quark in the box is replaced with an up or charm quark.

5.1 CP violation in B decays

The CKM mechanism only allows CP violation in processes where quarks couple to charged weak bosons. These effects can be studied using specific decays of neutral and charged mesons and baryons. The main focus of the LHCb experiment's CP violation program is the study of b- and c-hadrons.

There are three main categories of CP violation: direct CP violation in decay amplitudes, indirect CP violation in mixing, and CP violation in the interference between mixing and decay amplitudes [12]. Measurements of some of these effects in neutral B or charm hadron decays require flavor tagging algorithms to provide external measurements of the flavors of neutral B^0 and B_s^0 mesons at production time.

Neutral mesons have the interesting property that they can oscillate into their own antiparticles in a process called neutral meson mixing. Neutral meson oscillation is an excellent tool for probing CP violation; as recounted in Chapter 1, CP violation was discovered in neutral kaon oscillation. In addition to neutral kaons $K^0 = |\bar{s}d\rangle$, $D^0 = |c\bar{u}\rangle$, $B^0 = |\bar{b}d\rangle$, and $B_s = |\bar{b}s\rangle$ mesons undergo oscillation. Figure 5.2 shows the so-called box diagram, a fourth-order charged weak process, that causes a B^0 meson to evolve into a $\bar{B}^0$ meson and vice versa. Due to mixing, the mass eigenstates of the Hamiltonian are combinations

$$|B_{L,H}\rangle = p|B^0\rangle \pm q|\bar{B}^0\rangle, \tag{5.1}$$

137

where the complex parameters p and q are normalized so that $\langle B_{L,H}| B_{L,H}\rangle = 1.$[2] The neutral B^0 mesons are produced by strong interactions in the flavor definite states B^0 and $\bar{B}^0$, and thus as superpositions of the B_L and B_H states. Since these have differing masses (and hence oscillation frequencies), a meson that is produced as a B^0 (or a $\bar{B}^0$) evolves into a superposition of a B^0 and a $\bar{B}^0$. The evolution of these physical states can be derived from the evolution of the mixing eigenstates B_L and B_H [7]:

$$|B^0_{phys}(t)\rangle = e^{-imt}e^{-\frac{1}{2}\Gamma t}\left\{\cos\left(\frac{\Delta m}{2}t\right)|B^0\rangle + i\frac{q}{p}\sin\left(\frac{\Delta m}{2}t\right)|\bar{B}^0\rangle\right\}, \qquad (5.2)$$

$$|\bar{B}^0_{phys}(t)\rangle = e^{-imt}e^{-\frac{1}{2}\Gamma t}\left\{\cos\left(\frac{\Delta m}{2}t\right)|\bar{B}^0\rangle + i\frac{p}{q}\sin\left(\frac{\Delta m}{2}t\right)|B^0\rangle\right\}. \qquad (5.3)$$

Just as the B-mesons are produced in flavor-definite states B^0 and $\bar{B}^0$, they can decay in flavor-definite states as well. An important category of B decays is the set of those whose final state is common to both the B^0 and $\bar{B}^0$, and which can be used to measure CP violation in the interference between mixing and decay. Some of these are CP eigenstates f_{CP}, much like the two pion and three pion decays of the K^0_S and K^0_L. There are thus two interfering amplitudes for the decay to f_{CP}: $B^0 \to f_{CP}$, in which the B^0 decays directly to f_{CP}, and $B^0 \to \bar{B}^0 \to f_{CP}$, in which its flavor oscillates before it decays [88].[3] CP violation from interference can be

---

[2]The states $B_{L,H}$ are analogous to K^0_S and K^0_L, respectively (unfortunately the L subscript has switched places). The K_S and K_L states have vastly different lifetimes, by a factor of about 600, because the $K^0_L \to \pi\pi\pi$ decay allowed without CP violation has very little phase space and is suppressed. The much heavier D and B mesons are protected against this sort of effect because they are much more massive than their typical decay products.

[3]In the mass eigenstate basis, a B^0 meson is produced as a superposition of B_L and B_H, either of which can decay to f_{CP}. If p and q were both equal to unity, compatible with current measurements [13], the B_L and B_H states would be CP-even and CP-odd eigenstates, respectively. In this case, an alternative point of view is that CP violation is manifest the CP-even B_L component can decay to an odd final state and the CP-odd B_H component to a CP-even final state. In this case, an alternative point of view is that CP violation arises from the interference between the decay amplitudes $B_L \to f_{CP}$ and $B_H \to f_{CP}$ for the neutral B^0 with B_L and B_H components. Were instead $p \neq 1$ or $q \neq 1$, the B^0–$\bar{B}^0$ system would also exhibit CP violation in mixing.

measured from the asymmetry

$$\mathcal{A}_{f_{CP}} = \frac{d\Gamma/dt(\bar{B}^0_{phys} \to f_{CP}) - d\Gamma/dt(B^0_{phys} \to f_{CP})}{d\Gamma/dt(\bar{B}^0_{phys} \to f_{CP}) + d\Gamma/dt(B^0_{phys} \to f_{CP})}. \tag{5.4}$$

When $|q/p| = 1$, which is compatible with experimental evidence, this measurable quantity is related to the decay amplitudes by [12]

$$\mathcal{A}_{f_{CP}} = \frac{2\mathcal{I}m\lambda_f}{1 + |\lambda_f|^2} \sin(\Delta m\, t) - \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \cos(\Delta m\, t), \tag{5.5}$$

where λ_f is short for $\lambda_{f_{CP}}$, defined as

$$\lambda_{f_{CP}} = \frac{q}{p} \frac{\langle f_{CP}|H|\bar{B}^0\rangle}{\langle f_{CP}|H|B^0\rangle}. \tag{5.6}$$

The B factory experiments, BaBar and Belle, have extensively studied CP violation in B^0 decays, helping to verify the CKM mechanism, which was celebrated by the award of the 2008 Nobel prize in physics to Kobayashi and Maskawa. The hadron collider experiments have extended this research program to the B^0_s system. The LHCb collaboration has measured CP violation in several decays: $B^0 \to J/\psi\, K^0_{\rm S}$ [89] and $B^0 \to J/\psi\, \pi^+\pi^-$ [90], which provide information on the CKM angle β; $B^0 \to \pi^+\pi^-$, which provides information on the angle α [91]; and also $B^0_s \to J/\psi\, K^+K^-$, $B^0_s \to J/\psi\, \pi^+\pi^-$ [90,92], and $B^0 \to \psi(2S)K^+K^-$ [93], which provide information on β_s.

5.2   Flavor tagging

The time-dependent partial decay widths $d\Gamma/dt(B^0_{phys})$ and $d\Gamma/dt(\bar{B}^0_{phys})$ for states produced as B^0 and $\bar{B}^0$ respectively can only be measured if the production flavors of B^0 (or B^0_s) mesons are known. For example, an experimental analysis must determine whether a decay with the final state $J/\psi\, K_S$ originated from a B^0_{phys}

or $\bar{B}^0_{phys}$. The methods used to provide this knowledge, which typically rely on information in the event correlated with the flavor of the signal B^0 or B^0_s, are known as flavor tagging algorithms. Flavor tagging has been used by the B-factory experiments, notably BaBar [94–96] and Belle [97–100], as well as the pp collider experiments CDF [101] and D0 [102, 103] at the Tevatron and ATLAS [104, 105], CMS [106], and LHCb [89, 107–110] at the LHC.

In the LHCb experiment, bottom quarks are predominantly produced in $b\bar{b}$ pairs, producing a pair of hadrons with opposite bottom flavor. So, the parent B^0_{phys} or $\bar{B}^0_{phys}$ has a partner b hadron with opposite flavor that is typically within LHCb's acceptance. The flavor of this hadron can be determined if it undergoes a flavor specific decay, and so the flavor of the neutral B meson may be inferred indirectly. This strategy is known as opposite-side (OS) flavor tagging. OS muon and electron taggers look for leptons originating from semileptonic $b \to cW$ transitions of the b hadron, and an OS kaon tagger looks for kaons coming from $b \to c \to s$ transitions [107]. A vertex-charge tagger reconstructs the decay vertex of the OS b hadron and predicts its charge by weighting the charges of its decay products according to their transverse momentum. Finally, the charm tagger, described below, reconstructs secondary charm hadrons produced in b-hadron decays [108]. Alternatively, the flavor of the neutral B can be determined from the particles produced alongside it during the hadronization of its b-quark, a strategy called same-side (SS) flavor tagging [111]. The d or s partner of the light valence quark of the signal B has a roughly 50% chance of hadronizing into a charged pion or kaon. The LHCb experiment uses several SS tagging algorithms examining same-side pions, kaons, and protons [89, 110, 112].

The performance of the simplest algorithms is described by their tagging efficiency ε_{tag}, the fraction of events for which they produce a decision, and their mistag rate ω, the fraction of tagged events for which their decision is incorrect.

These separate performance metrics are reflected in the tagging power, or effective tagging efficiency,

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}(1 - 2\omega)^2, \qquad (5.7)$$

which measures the statistical sensitivity of an asymmetry measurement that uses the tagging algorithm. It has long been recognized that the tagging power can be increased by the use of a flavor tagging algorithm with a variable response. With the range of responses indexed by k, with corresponding mistag probabilities ω_k, each event is weighted by the corresponding dilution factor $1 - 2\omega$. The tagging power then becomes

$$\varepsilon_{\text{eff}} = \sum_k \varepsilon_{\text{tag}_k}(1 - 2\omega_k)^2 \qquad (5.8)$$

This power is greater than the simple tagging power of Eq. (5.7) by an amount proportional to the sample variance of ω_k; a more in-depth discussion is provided in Appendix F. Some implementations of this scheme split the tag decision into a number of discrete, ranked categories, ranging from weak decisions with ω near 0.5 to stronger tags with a smaller ω [99, 105]. Others instead produce a continuous output, parametrized by a predicted mistag probability η [107–110]. Frequently this output is the response of a multivariate algorithm (MVA), or statistical classifier, trained using real or simulated data to distinguish between correct and incorrect decisions [107].

In practice, no single flavor tagging algorithm has acceptable tagging power on its own. To maximize the statistical power of an analysis, flavor tagging algorithms are combined probabilistically, assuming that the taggers are uncorrelated. A further calibration is applied to this combination, since the assumption is unlikely to be completely correct. The typical combined tagging power of the current set of OS tagging algorithms used by LHCb is approximately 2.5% [89, 90, 92, 113]. Including SS taggers, recent measurements have had tagging powers as high 5% or more [114].

5.3 Flavor tagging calibration

The mistag probabilities ω_k of the flavor tagging algorithm are measured using data from a flavor specific B decay, whose decay mode identifies the flavor of the B meson at decay time. For charged $B^\pm$ decays this is identical to the flavor at production time, and can be directly compared to the production flavor predicted by the tagging algorithm. In the case of B^0 and B_s^0 decays, the decay flavor is related to the unobserved (latent) production flavor through the dynamics of the effective Hamiltonian of the B^0–$\overline{B}^0$ of B_s^0–$\overline{B}_s^0$ system, and the analysis becomes more difficult. When the number of discrete categories is few, the tagging efficiency and mistag rate can be evaluated separately in each independently; however, when the number of categories is high or the output is continuous, there may not be enough statistical power to perform these independent measurements. In this latter case, the mistag probability must instead be determined as a function $\omega(\eta)$ fit to the data using a calibration procedure.

A variety of calibration procedures are currently used in physics analyses. Different methods are in use for calibrations to decays of $B^\pm$ mesons and to decays of neutral B^0 and B_s^0 mesons. These methods typically suffer from two kinds of problems: they may introduce an arbitrary binning procedure that introduces unnecessary systematic effects, or they may use an unnecessary (and resource intensive) time-dependent likelihood maximization fit.

The problem of calibrating flavor tagging algorithms can instead be solved with binomial regression. This well-established statistical procedure is the canonical method for fitting functions $\pi(x)$ to data (X, Y) in which the bivalued response variable y follows a binomial distribution with probability $\pi(x)$ [115,116]; it is a close cousin of linear (*e.g.*, least squares) regression. For calibrations to decays of charged $B^\pm$ mesons the application of binomial regression is straight-forward: simply regress

$\omega(\eta)$ to the dataset of predicted mistag probabilities η and actual mistag outcomes. An extension to the usual binomial regression procedure, described in described in Section G.1, is required to calibrate to neutral B^0 and B_s^0 decay modes because of neutral meson oscillation. The ideal response variable, whether the algorithm's prediction is equal to the production flavor of the B^0 or B_s^0, is latent, and the regression must instead compare the predicted flavor to the oscillated decay flavor. Because of the precisely understood nature of neutral B meson mixing, the decay flavor can be probabilistically related to the production flavor with negligible ambiguity, as detailed in Section G.2. With this extension, binomial regression can be used for both charged mode and neutral calibration. The procedure does not suffer from arbitrary binning choices, and is less resource intensive than full time-dependent likelihood maximization procedures.

Appendix G describes a new flavor tagging calibration method using binomial regression. This procedure treats calibrations to B^+, B^0, and B_s^0 on an equal footing, and involves no binning procedure that would introduce additional systematic uncertainties in CP violation measurements. The implementation of the tool in a convenient command-line program called the `EspressoPerformanceMonitor` dramatically reduces the time required to perform calibrations to B^0 and B_s^0 decays down to a matter of hours or even minutes. Tables of calibration parameters and performance metrics and publication quality diagnostic plots are created automatically. Finally, the tool provides a broad set of generalized linear models (GLMs) that can be used in the calibration procedure and introduces goodness-of-fit tests for binomial regression described in the statistical literature, superior to the several assessments made before.

Chapter 6:   Charm tagger

This chapter describes a new flavor tagging algorithm for the LHCb experiment that reconstructs secondary charm hadrons produced in the decay of OS b hadrons.

6.1   Basic structure of the OS charm tagger

In events containing a signal B decay, opposite-side D^+, D^0, and Λ_c^+ charm hadrons are primarily produced through the quark-level $b \to c$ transition, and the charge of the D^+ or Λ_c^+ tags the flavor of the b hadron parent. For Cabibbo-favored $D^0 \to K^- X$ decays, the kaon charge tags the flavor of the charm hadron, and thereby that of the parent B hadron. The OS charm tagging algorithm uses charm meson candidates reconstructed in a number of decay modes, chosen for their relatively large branching fractions. These include fully reconstructed (or exclusive) hadronic modes with a single charged kaon in the final state, $D^0 \to K^- \pi^+$, $D^0 \to K^- \pi^+ \pi^+ \pi^-$, $D^+ \to K^- \pi^+ \pi^+$, and $\Lambda_c^+ \to p^+ K^- \pi^+$; a partially reconstructed (or inclusive) hadronic mode with an unobserved neutral pion, $H_c \to K^- \pi^+ X$; and partially reconstructed semileptonic modes, $H_c \to K^- e^+ X$ and $H_c \to K^- \mu^+ X$.[1] The selection and reconstruction criteria of these charm modes are described below in Section 6.2.

---

[1] The easily reconstructible charm decay modes all contain kaons. The pre-existing OS kaon tagger also selects opposite-side kaons. However, the two algorithms have different selection strategies and select different (though somewhat overlapping) sets of kaon tracks. Moreover, the algorithms examine different features when producing their predicted mistag rates η: the OS kaon tagger selects and makes its predictions based entirely on the individual kinematic properties of the kaon, while the charm tagger's decision is based on the properties of at least a pair of tracks.

Even with perfectly reconstructed charm hadrons and no background, several effects produce an inherent mistag probability for the OS charm tagging algorithm. Chief among these are B^0–$\bar{B}^0$ or B_s^0–$\bar{B}_s^0$ oscillation on the opposite-side and "wrong sign" charm hadrons produced in $b \to c\bar{c}q$ transitions. The impact of D^0–$\bar{D}^0$ mixing and doubly Cabibbo-suppressed decays such as $D^0 \to K^+\pi^-$ is negligible. Accounting for relative production cross sections of b hadrons, neutral B^0 and B_s^0 oscillation rates, and b-hadron to c-hadron branching fractions, the inherent mistag probabilities for D^0, D^+ and Λ_c^+ modes are estimated to be 23%, 19%, and 6%, respectively. The input values used in these estimates, taken primarily from PDG and LHCb measurements, are given in Table 6.1.

In addition to the inherent mistag probability arising from these physics effects, the reconstructed charm hadron candidates are contaminated with combinatorial and partially reconstructed b and c hadron backgrounds that can produce incorrect flavor tags. Suppressing these backgrounds is crucial in optimizing the performance of the OS charm tagger. For each mode, the charm tagger uses a multivariate algorithm that combines geometric and kinematic properties of the c hadron candidate and its decay products. The resulting discriminating variable is used both to suppress the combinatorial background and to predict the mistag probability of the best surviving candidate. These multivariate algorithms are described in detail in Section 6.3.

6.2 Reconstruction of charm candidates

The decay modes used by the charm tagger have been selected based on their relatively large branching fractions and potentially low combinatorial backgrounds. The most effective modes for flavor tagging are those with a single kaon and a few pions. Modes containing a neutral pion, K_S^0, or Λ suffer from low reconstruction efficiencies and would not contribute significantly to the tagger.

145

Table 6.1: Factors determining the inherent mistag probability of tagging by the charm-hadron decay products of b-hadron decays. These include the relative production cross sections of the b-hadrons, the average probability χ that they oscillate before decaying, and their inclusive branching fractions to studied c-hadrons; † indicates estimates.

Particle	$\sigma/\sigma(B^+)$	χ	$\mathcal{B}(D^0)$	$\mathcal{B}(\overline{D}^0)$	$\mathcal{B}(D^+)$	$\mathcal{B}(D^-)$	$\mathcal{B}(\Lambda_c^+)$	$\mathcal{B}(\overline{\Lambda}_c^-)$
B^+	1	0	8.6%	79%	2.5%	9.9%	2.1%	2.8%
B^0	~ 1	0.186	8.1%	47.4%	2.5% †	36.9%	2.1% †	5.0%
B_s^0	$\sim 1/4$	0.499	8.2% †	30% †	2.5% †	9.9% †	2.1% †	2.4% †
Λ_b^0	$\sim 3/4$	0	0% †	0% †	0% †	0% †	100% †	0% †

All of the current OS taggers use one or more tracks from a track list developed for the tagging algorithms (see, $e.g.$, [117, 118]). The kinematic requirements on these lists are too stringent for the reconstruction of the charm hadrons with sufficient efficiency. To produce a clean set of charm hadrons without sacrificing tagging efficiency, candidates are created by combining long tracks with relatively loose kinematic requirements but tighter particle identification (PID) requirements. Tracks corresponding to decay products of the reconstructed signal B are removed.

Charm hadron candidates are created from tracks passing a number of selections designed to efficiently reduce background. In particular, they must have at least a momentum of 1000 MeV/c and transverse momentum of 100 MeV/c, or 150 MeV/c for the especially background heavy $K\pi\pi$ mode. The tracks also must have an IP chi^2 greater than 4 for each PV; this is raised for the $K\pi\pi$ mode and inclusive modes. Each track must also pass particle identification (PID) criteria, using the PID scores provided for it by a high-performing neural network, to reduce cross-contamination between the sets of selected pions, kaon, protons, and leptons.

The candidates built out of these tracks are required to pass further selection requirements. These include requirements on vertex quality: a maximum distance of closest approach between each pair of tracks and a maximum χ^2 per degree of freedom for the decay vertex fit. Each candidate is required to be well separated

from its best associated primary vertex (BPV), with a minimum χ^2 of the distance separating the decay vertex from the BPV. The candidate is required to have a trajectory that leads back to the PV, by imposing a minimum cosine of the angle between the momentum and the flight direction from the BPV to the decay vertex. The invariant mass of the charm hadron candidate is required to be consistent with the known mass of the corresponding charm hadron, within $100\,\mathrm{MeV}/c^2$ for the Λ_c^+ channel and $50\,\mathrm{MeV}/c^2$ for all other fully reconstructed D decay modes. For the partially reconstructed $D \to K^-\pi^+X$ mode, the $K^-\pi^-$ mass is required to be no less than $400\,\mathrm{MeV}/c^2$ below the known D^0 mass This invariant mass region is favored by $K^-\pi^+$ pairs from the quasi-two body decay $D^0 \to K^-\rho^+$. Alternatively, the $K^-\pi^+$ pair can be within $50\,\mathrm{MeV}/c^2$ window of the mass of the K^{*0} resonance, selecting $D \to K^{*0}X$ decays.

Fig. 6.1 shows the invariant mass distributions of the fully-reconstructed opposite-side charm hadrons, in a control sample of $B^+ \to J/\psi K^+$ events taken from 2011 and 2012 data (described in Section 6.4). These distributions show significant remaining background, which must be reduced in order to lower the mistag probability of the algorithm. A machine learning algorithm is employed to achieve this.

6.3   Optimization of the flavor tagging algorithm

Charm hadron candidates contain several sources of backgrounds that can dilute flavor tagging information. These include combinatorial background from particles that did not originate from the same parent; partially reconstructed background, where a subset of decay products from a heavier parent particle have been combined; mis-identified background, where the particle species of one or more decay products has been incorrectly assigned; and ghost background, where one of the tracks is created from hits that belong to several different tracks and/or electronic
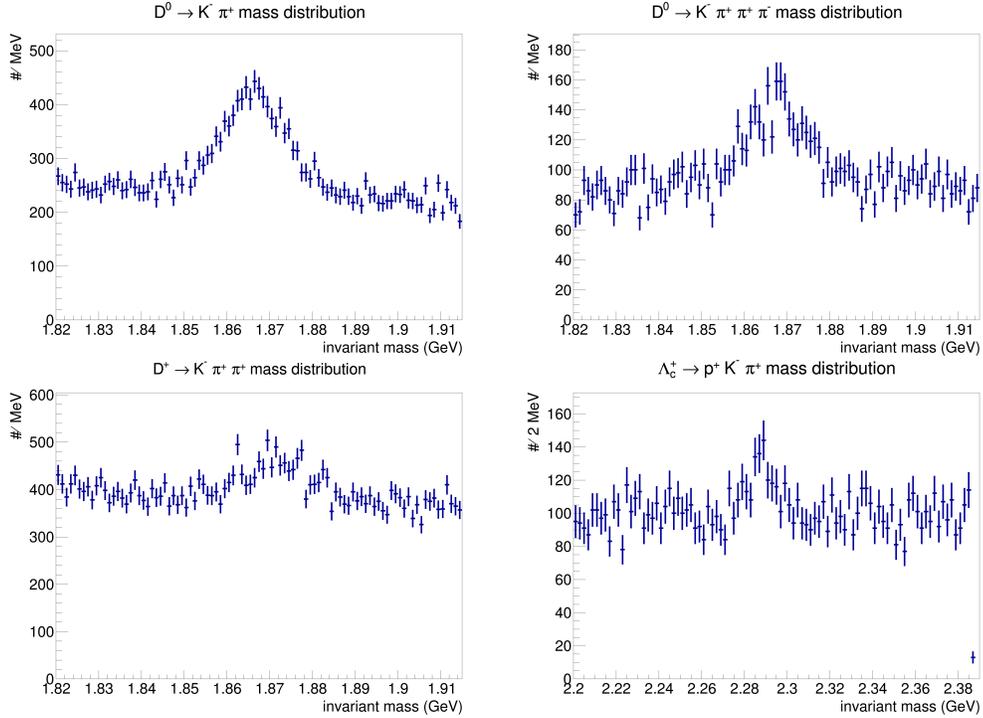
Figure 6.1: Charm hadron mass distributions in data.

noise.

The charm tagger employs a selection designed to reduce this background and to calculate an estimated mistag probability for the tagging decision. First, a machine learning algorithm calculates a score for each candidate that discriminates between real charm hadron signals and background processes. This score is turned into a predicted mistag probability for the decision. A final selection picks the best candidate for each event with a predicted mistag probability below some maximum threshold.

6.3.1   Signal and background discrimination

For each mode, a boosted decision tree (BDT) [119], adaptively boosted using `AdaBoost` [120], is used to discriminate signal from background and thereby to estimate mistag probabilities. The BDT algorithms are trained using simulations of $b\bar{b}$ events containing $B^+ \to J/\psi \, K^+$, $B^0 \to J/\psi \, K^{*0}$, and $B_s^0 \to J/\psi \, \phi$ decays on

the signal side and inclusive decays of the b hadron on the opposite-side. These B decays are used to model the various sources and relative amounts of background when reconstructing OS charm hadrons recoiling against signal B decays. In total, 6.5×10^6 such simulated events are available for training and optimization.

The input variables used in training the BDT are variables describing the kinematics, vertex quality and location, and PID scores of the charm hadron decay and its decay products:

Charm hadron candidate: its invariant mass, momentum, transverse momentum, a pseudo decay-time,[2] χ^2 probability of the vertex fit, flight distance, flight distance χ^2, angle between its flight direction and momentum, and the maximum ghost score of all its decay products;

Any pion, kaon, or proton candidate: its p_{T}, impact parameter χ^2 with respect to the BPV, and minimum impact parameter χ^2 with respect to any PV, kaon score, and proton score;

Electron or muon candidate: its electron score or muon score, respectively.

For modes where more than one pion is present, the minimum or maximum of each variable is taken, as appropriate

For each mode, the BDT algorithm is trained using that mode's simulation truth-matched charm candidates as the signal sample, and the remaining candidates as the background sample. The simulation truth-matching includes the requirement that the candidate originate from a B decay. Each sample is randomly divided into two equally size parts. One part is used for the algorithm training, while the other is used to cross-check the final result.

The trained BDT provides a score corresponding to the probability for the charm hadron candidate to be signal. This information is used to predict the mistag

---

[2]This is calculated from the distance between the c hadron's decay vertex and the corresponding best PV, and approximates the sum of the decay-times of the c hadron and its parent b hadron.

for the candidate in a two-step process. First, an "estimated" mistag is computed in ideal conditions, using a simplified model of the mistag due to the signal and background components. Second, the mistag observed in Monte Carlo is parameterized as a function of this estimated mistag.

6.3.2   Final selection

Multiple tagging candidates may be present in the same event; in the Monte Carlo dataset there are about 1.5 candidates per event, on average. In this case, the candidate with the lowest predicted mistag is retained. This is typically equivalent to picking the candidate most likely to be signal. The BDT algorithms are trained to separate signal from background, with a higher BDT score corresponding to a higher probability to be signal; the translation of this BDT score into the mistag probability is usually a monotonically decreasing function. Keeping only one candidate is by definition 100 % efficient, since the number of events tagged is unchanged. This choice throws out 34% of the tagging charm hadron candidates in the Monte Carlo dataset.

Applying this method to the Monte Carlo events, we find that $\sim 40\%$ of events have no candidate that has predicted mistag $\eta < 45\%$, as shown in Figure 6.2. The contribution of these candidates to the tagging power, which is proportional to their predicted dilution squared, is small. Fig. 6.3 shows that the total tagging power of the events vetoed by this cut is around 5% of the total tagging power in Monte Carlo. Therefore, events with predicted mistag greater than 45% are vetoed and removed from further consideration in tagging decisions. Removing candidates that fail this criterion significantly reduces the computing time of the algorithm at little cost to tagging performance.

The efficiencies of various selection criteria are evaluated on Monte Carlo and data. It is found that the requirement $\eta < 45\%$ removes $(46.6 \pm 0.4)\%$ and
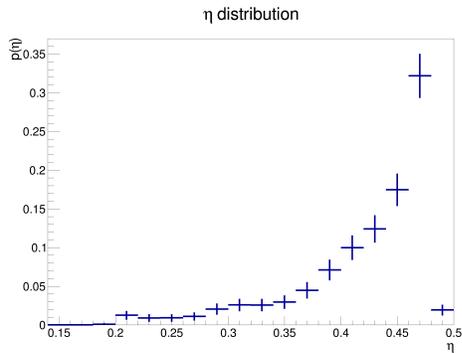
Figure 6.2: Distribution of predicted mistag η for the control sample in Monte Carlo events; the requirement that $\eta < 45\%$ has been removed for illustrative purposes.
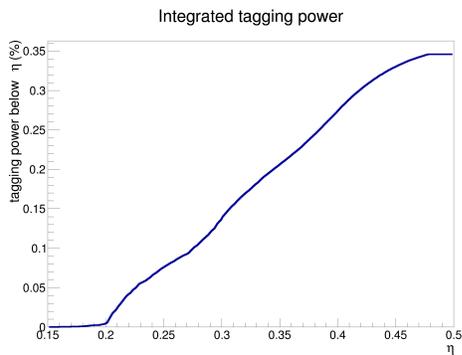
Figure 6.3: The y-axis shows the tagging power (evaluated on MC) as a function of the upper bound on the estimated mistag (η); the requirement that $\eta < 45\%$ has been removed for illustrative purposes.

Table 6.2: Selection efficiencies for the best candidate requirement and the mistag requirement $\eta < 45\%$, measured in terms of tagging candidates retained and tagged events retained, respectively.

Sample	Best candidate cut	$\eta < 45\%$ cut
2012 simulation	$(65.79 \pm 0.04)\,\%$	$(58.99 \pm 0.07)\,\%$
2011	$(61.9 \pm 0.2)\,\%$	$(53.4 \pm 0.4)\,\%$
2012	$(60.4 \pm 0.2)\,\%$	$(52.0 \pm 0.3)\,\%$

$(48.0 \pm 0.3)\,\%$ of events in 2011 and 2012 data, respectively. Table 6.2 tabulates the efficiencies of these selections in the Monte Carlo, 2011, and 2012 datasets. The relative Monte Carlo background compositions of the charm hadron candidates at various stages of the selection process are detailed in Section 6.3.3.

6.3.3 Background composition

In summary, there are two steps of candidate selection in the charm tagger. The first is the reconstruction and pre-selection of the charm candidates from the basic particle lists. The second is the final selection of a single candidate using the BDT algorithms. Table 6.3 shows the background composition after each these two steps of the charm hadron candidates belonging to the exclusive and inclusive modes. For the inclusive modes, the K^{*0} resonance and $K\ell\nu$ semileptonic decays are counted as signal. After the preselection, there are large partially reconstructed, reflection, ghost, primary vertex, and pileup backgrounds for the exclusive and inclusive modes, with a smaller amount of low mass and combinatorial backgrounds. The final selection significantly increases the signal to background ratio by reducing all the major backgrounds, with the exception of partially reconstructed background.

Table 6.3: Fractions of signal and significant backgrounds for the exclusive modes.

Category	Pre-selection		Final selection	
	Exclusive	Inclusive	Exclusive	Inclusive
Signal	11.7%	14.1%	41.8%	29.6%
Partially rec.	27.2%	38.4%	29.7%	44.2%
Low mass		3.0%		5.2%
Reflection		11.7%		1.4%
Ghost	10.2%	6.8%	6.0%	4.1%
PV	28.9%	15.4%	13.1%	9.3%
Pileup	18.5%	10.3%	1.9%	4.0%
Combinatorial	2.6%	4.9%	1.9%	2.2%

6.4 Calibration and performance

In order to evaluate its performance, the charm tagger must be calibrated to collision data by comparing the algorithm's predictions to the known flavors of signal B candidates, according to the procedure detailed in . The calibration parameters δp_0, p_1, Δp_0, and Δp_1 are defined by

$$\omega = \langle \eta \rangle + \delta p_0 + p_1 \left(\eta - \langle \eta \rangle \right)$$

$$\Delta \omega = \Delta p_0 + \Delta p_1 \left(\eta - \langle \eta \rangle \right)$$

where $\langle \eta \rangle$ is the average predicted mistag probability, ω is the actual mistag probability averaged over B^+ and B^- signal mesons, and $\Delta \omega$ is the excess mistag probability for B^+ mesons with respect to B^- mesons; equivalent definitions hold for $B^0/\overline{B}^0$ signal.

Several calibrations of the charm tagger using older methods described in Ref. [107] are detailed in the paper on the charm tagger, with corresponding estimates of systematic uncertainties [108]. Presented here instead are several calibrations of the algorithm performed using binomial regression via the EspressoPerformanceMonitor to the flavor self-tagged modes $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$,

(a) $B^+ \to J/\psi\, K^+$ data
(b) $B^0 \to J/\psi\, K^{*0}$ data
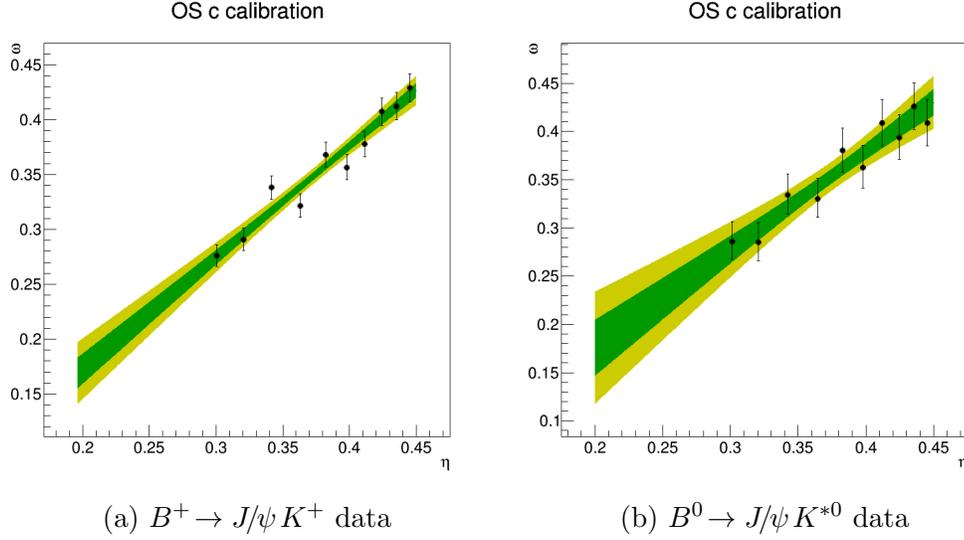
Figure 6.4: Calibration of charm tagger on charmonium B decays.

$B^+ \to \overline{D}{}^0 \pi^+$, and $B^0 \to D^- \pi^+$. These calibrations use samples of these decays created for common use in flavor tagging calibration studies. The signal candidates have been selected from about $1.1\,\mathrm{fb}^{-1}$ and $2\,\mathrm{fb}^{-1}$ of data collected in 2011 and 2012. Fits to the reconstructed B mass distribution have been used to separate signal and background via the *sPlot* procedure, which computes signal and background weights for each candidate [85]. Plots of the resulting calibrations are shown in Figs. 6.4 and 6.5.

The resulting calibration parameters are given in Table 6.4. These calibration parameters show good agreement in general between the magnet up and magnet down datasets and between the 2011 and 2012 datasets, with the exception of out-lying discrepancies in the measurements of Δp_1 in the 2011 magnet up and magnet down datasets. These are the two poorest determined calibration parameters, and differ from each other by 3.0σ. The overall $\chi^2/n_{\mathrm{d.o.f.}}$ between the eight parameters in the magnet up and magnet down datasets (incorporating the correlation between fit variables) is 2.23. Meanwhile, the $\chi^2/n_{\mathrm{d.o.f.}}$ between the eight parameters in the 2011 and 2012 datasets is 2.20. These high χ^2 values are mostly due to the Δp_1 outliers in 2011; neglecting Δp_1, they would be 0.71 and 1.25.

(a) $B^+ \to D^0 \pi^+$ data

(b) $B^0 \to D^- \pi^+$ data
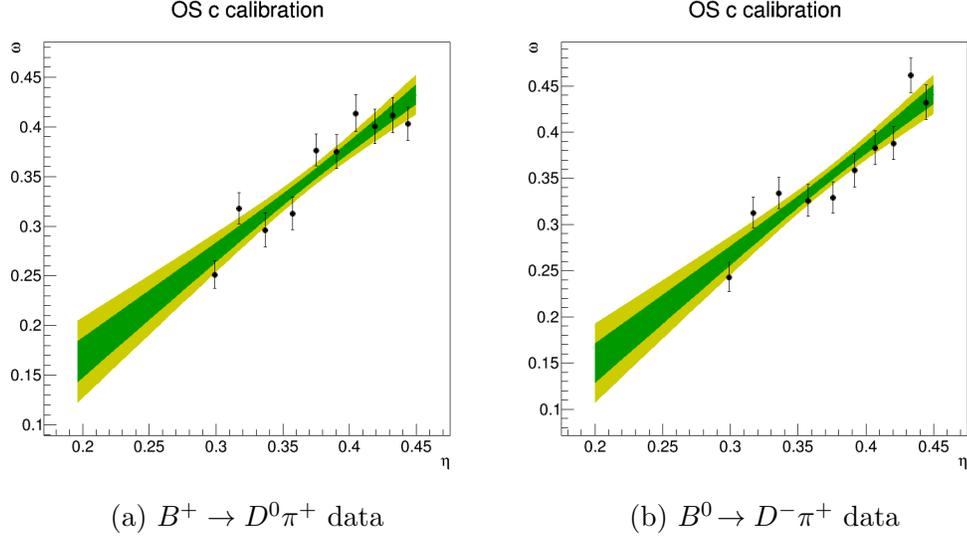
Figure 6.5: Calibration of charm tagger on open charm B decays.

Table 6.4: Calibration parameters as determined from the various control sample; these parameters are defined in Section 6.4.

Sample	$\langle \eta \rangle$	δp_0	p_1	Δp_0	Δp_1
$B^+ \to J/\psi\, K^+$	0.3795	$-0.0247(37)$	1.015(76)	0.0102(73)	0.033(152)
$B^0 \to J/\psi\, K^{*0}$	0.3794	$-0.0213(76)$	1.019(159)	0.0228(110)	0.322(232)
$B^+ \to \overline{D}^0 \pi^+$	0.3764	$-0.0221(54)$	1.061(112)	$-0.0065(108)$	0.110(224)
$B^0 \to D^- \pi^+$	0.3777	$-0.0212(58)$	1.166(119)	0.0133(79)	0.073(165)

Parameter	Systematic
p_0	0.0030
p_1	0.017
Δp_0	0.0043
Δp_1	0.040

Table 6.5: Systematic uncertainties from simulation studies.

6.4.1 Systematic uncertainty

A calibration to the channel $B_s^0 \to D_s^- \pi^+$ was attempted; however, the smaller production cross section of this decay, together with the low tagging rate of the algorithm, made this channel statistically limited, meant that there were insufficient data to perform the required lifetime analysis. Instead, in order to ensure that the algorithm performs similarly for B_s^0 channels as well as B^+ and B^0 channels, separate calibrations to simulated $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$, and $B_s^0 \to J/\psi \phi$ events have been performed. Where statistically significant differences between the calibration parameters in the three channels were found, a systematic uncertainty, corresponding to half of the maximum difference, have been assigned to the parameter. These systematic uncertainties are roughly comparable to typical statistical uncertainties for the parameters δp_0 and Δp_0, but are negligible for p_1 and Δp_1. The propagation of these uncertainties results in a 0.011% absolute systematic uncertainty on the tagging power, comparable to its statistical uncertainty.

6.4.2 Performance

The distribution of η after calibration for the $B^+ \to J/\psi K^+$ control sample is shown in Fig. 6.6. Table 6.6 reports the breakdown of the charm tagger's performance by decay mode. The relative rate and power of each mode are the amounts that it contributes to the algorithm's total tagging rate ε_{tag} and tagging power ε_{eff}. These figures indicate room for future improvements. The rate of the

Table 6.6: Decay modes used in the OS charm tagger. The symbol H_c stands for any c hadron. The definition of the two right-most columns is given in the text.

Decay mode	Relative rate	Relative power
$D^0 \to K^- \pi^+$	10.0%	24.0%
$D^0 \to K^- \pi^+ \pi^+ \pi^-$	5.9%	8.4%
$D^+ \to K^- \pi^+ \pi^+$	10.3%	2.6%
$H_c \to K^- \pi^+ X$	69.7%	61.5%
$H_c \to K^- e^+ X$	0.5%	0.2%
$H_c \to K^- \mu^+ X$	3.4%	0.3%
$\Lambda_c^+ \to p^+ K^- \pi^+$	0.2%	2.4%

$D^+ \to K^- \pi^+ \pi^+$ is outsized in proportion to its tagging power, and the semileptonic modes make a negligible contribution. These could be addressed by reexamining the selections and multivariate algorithms for these modes. In particular, the semileptonic modes might benefit by adding further variables describing the leptons into the corresponding BDTs.

The tagging efficiency, mistag fraction, and the tagging power of the charm tagger are reported in Table 6.7 for the training sample of simulated $B \to J/\psi X$ decays and for each calibration channel. The propagated statistical uncertainty of the calibration parameters dominates the statistical uncertainty of the tagging power. As expected, the overall tagging power is slightly higher in simulation than in data, due to differences in the distributions of input variables. The tagging powers in the two $B \to J/\psi X$ calibration channels are consistent. There is a significant difference in the tagging efficiency, higher in the $B^0 \to J/\psi K^{*0}$ channel, which is offset by a slightly higher mistag rate. These differences are likely due to the different signal selections of the $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$ channels, which can shape the kinematic distributions of the opposite-side b hadrons and secondary charm hadrons.

The tagging efficiency for these samples is found to be higher than for the samples of $B \to J/\psi X$ decays, due to correlations between the kinematics of the

(a) $B^+ \to J/\psi\, K^+$

(b) $B^0 \to J/\psi\, K^{*0}$
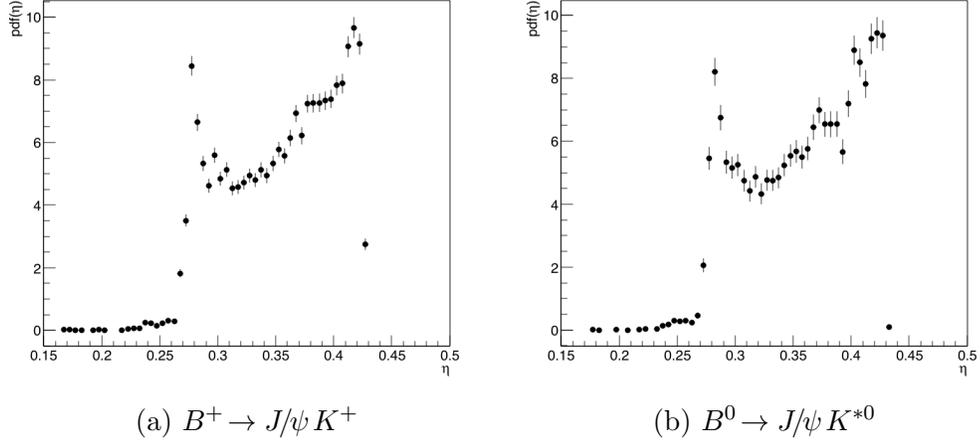
Figure 6.6: Distributions of the predicted mistag η (after calibration) for the $B \to J/\psi\, X$ data samples.

Table 6.7: Tagging efficiencies (ε_{tag}), effective mistag fractions (ω_{eff}), and tagging powers (ε_{eff}) in the various data samples studied. The first uncertainties are statistical and the second are systematic. The sample labeled Simulation is the training sample of simulated $B^+ \to J/\psi\, K^+$, $B^0 \to J/\psi\, K^{*0}$, and $B_s^0 \to J/\psi\, \phi$ decays. The first uncertainties are statistical and the second are uncertainties due to the flavor tagging calibration.

Sample	ε_{tag}	ω_{eff}	ε_{eff}
Simulated	$(4.88 \pm 0.01)\%$	$(37.0 \pm 0.0)\%$	$(0.33 \pm 0.0)\%$
$B^+ \to J/\psi\, K^+$	$(3.08 \pm 0.02)\%$	$(34.70 \pm 0.03 \pm 0.355)\%$	$(0.288 \pm 0.003 \pm 0.013)\%$
$B^0 \to J/\psi\, K^{*0}$	$(3.26 \pm 0.04)\%$	$(35.03 \pm 0.05 \pm 0.75)\%$	$(0.292 \pm 0.004 \pm 0.029)\%$
$B^+ \to \overline{D}^0 \pi^+$	$(4.71 \pm 0.05)\%$	$(34.60 \pm 0.03 \pm 0.54)\%$	$(0.448 \pm 0.005 \pm 0.031)\%$
$B^0 \to D^- \pi^+$	$(4.85 \pm 0.11)\%$	$(35.45 \pm 0.12 \pm 1.57)\%$	$(0.411 \pm 0.011 \pm 0.089)\%$

signal B and the opposite-side charm hadrons. The effective mistag fraction for these samples is consistent with that on the $B \to J/\psi\, X$ samples. The net effect is an increased tagging power for these $B \to DX$ decays, similar to that observed for other opposite-side tagging algorithms [92, 121].

6.5 Combination with current OS taggers

To use the charm tagger in a physics analysis, the flavor tagging information from the charm tagger can be combined with information from other tagging

(a) $B^+ \to \overline{D}^0 \pi^+$ (b) $B^0 \to D^- \pi^+$
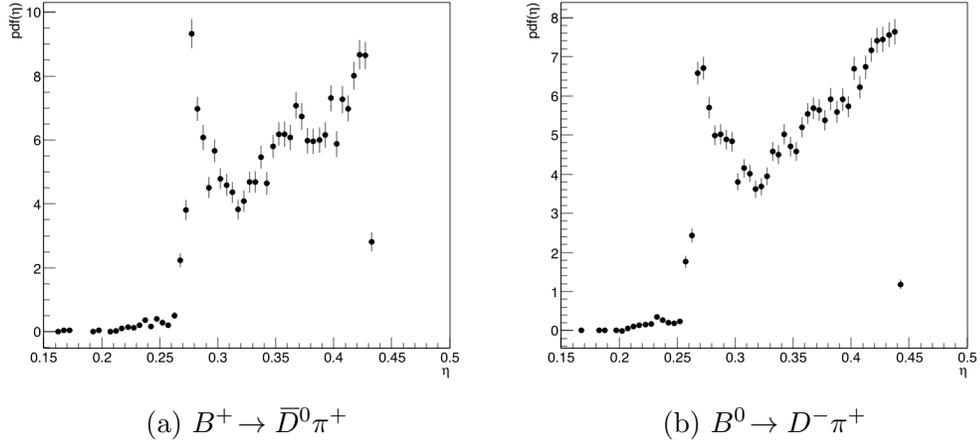
Figure 6.7: Distributions of the predicted mistag η (after calibration) for the $B \to D\pi$ data samples.

algorithms. Assessing the actual gain in performance depends on the method of combination and calibration, as well as on the set of tagging algorithms being combined. Due to correlations with other tagging algorithms, in particular the OS kaon and vertex-charge taggers, the maximum possible increase in tagging power after the addition of the charm tagging algorithm is less than its individual tagging power. However, beyond the added tagging power, the presence of an additional tagging algorithm provides useful and necessary redundancy, which is important for robustness of the flavor tagging performance against changes in detector conditions and reconstruction algorithms.

The overall correlation of the charm tagger with the other OS taggers can be determined by assigning each tagger's decision a value 0 (if no decision is made) or ± 1 (based on the predicted flavor). This procedure shows that the charm tagger is most correlated with the OS kaon tagger, at about $\sim 20\%$, and has small correlations with the other OS taggers on the order of 5% total. This correlation is non-zero because the two taggers can pick up the same kaon track to make their decisions; however, it is still much less than 100%, and the greater part of the information provided by the charm tagger is new.

In the standard combination algorithm, the relative probability for either charge hypothesis is calculated by multiplying the probabilities assigned by the tagging algorithms. This combined tagger requires its own calibration, on top of the calibration of each individual tagger. Furthermore, it is known to be sub-optimal, because it does not take into account correlations between separate tagging algorithms.

Assessing the actual gain in performance depends on the method of combination and calibration, as well as on the set of tagging algorithms being combined. We have compared the performance of a calibrated combination of the OS muon, electron, neural-net kaon, and vertex charge taggers to a calibrated combination with the addition of the charm tagger. Without the charm tagger, the tagging power of the combination is about 2.8%, which is typical for the current set of OS tagging algorithms. Using the standard combination algorithm leads to a net gain in tagging power of 0.066%. The poor performance of the combination algorithm can be partially compensated for by performing separate calibrations in three disjoint sets of events. Given sets of events E_{comb} and E_{charm} where the combination tagger and charm tagger give a tag decision, respectively, these three sets are $E_{\text{comb}} \cap E_{\text{charm}}$, $E_{\text{comb}} \setminus E_{\text{charm}}$, and $E_{\text{charm}} \setminus E_{\text{comb}}$. Then, the gain in tagging power becomes 0.113%.

6.6 Future improvements to the charm tagger

There are many possible ways in which the charm tagger could be improved in the future. The most potential lies in improvements to the selections for the charm decay modes, which might be improved to limit background contamination, and in the discriminating variables supplied to the BDT, which could be augmented. There could also be improvements to the training and machinery of the BDT and the calculation of the predicted mistag η, that besides possibly increasing the tagging power would streamline the algorithm. For example, the charm tagger currently uses
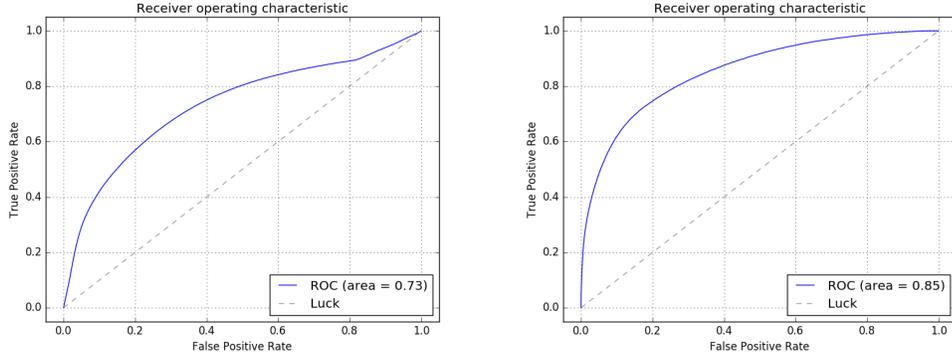
(a) ROC curve for nominal BDTs   (b) ROC curve using `xgboost` prototype

Figure 6.8: Possible improvement in charm tagger AUC using `xgboost`

seven different BDTs trained for each decay mode. To make better use of limited simulation resources, a single pooled BDT could be created in which the decay mode of the candidate is represented by a set of seven binary dummy variables. Furthermore, the current parameterization procedure used to calculate the mistag probability η from the BDT score is not very robust, and could be replaced with a more standard technique, *i.e.*, a logistic regression for each separate decay mode.

These potential improvements have been prototyped using the `xgboost` boosted decision tree tool in place of `TMVA` [122]. The same input variables used in the nominal charm tagger BDTs are used to train an `AdaBoost`ed BDT using `xgboost`. A significant increase in the AUC score is found over the effective AUC score from the nominal BDTs.[3]

6.7 Conclusion

An algorithm has been developed that determines the flavor of a signal b hadron at production time by reconstructing opposite-side charm hadrons from a number of decay channels. The flavor tagger uses boosted decision tree algorithms trained on simulated data, and has been calibrated and evaluated on data using

---

[3]It is possible that the raw outputs of the seven nominal BDTs are on different scales, so each is converted into a corresponding signal purity fraction.

the self-tagged decay $B^+ \to J/\psi K^+$. Its calibration and performance are found to be consistent between magnet up and magnet down running conditions, as well as between the 2011 and 2012 datasets. Its tagging power for data in this channel is found to be $(0.288 \pm 0.003(\text{stat}) \pm 0.013(\text{cal}))\%$. The calibration has been cross-checked using the decay $B^0 \to J/\psi K^{*0}$, giving consistent results. The tagging power is found to be higher for the decays $B^0 \to D^- \pi^+$ and $B_s^0 \to D_s^- \pi^+$, at $(0.411 \pm 0.011(\text{stat}) \pm 0.089(\text{cal}))\%$ and $(0.39 \pm 0.03)\%$, respectively.

Part IV

Appendices

Appendix A: More detailed form factor plots
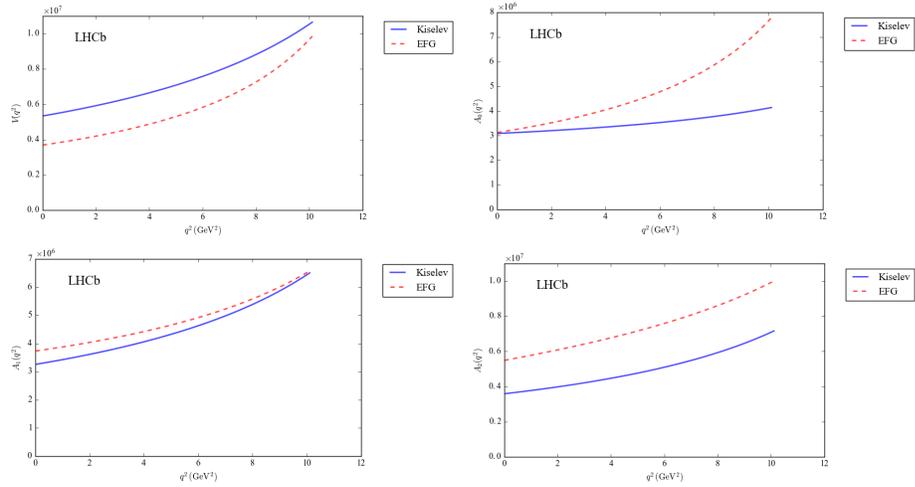
Figure A.1: The Kiselev and EFG form factors do not seem to agree well in the standard parameterization.
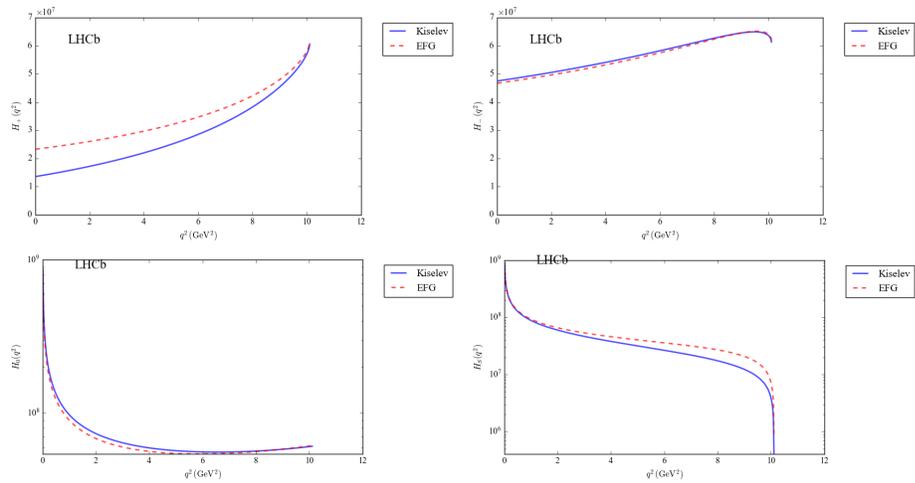
Figure A.2: In the helicity parameterization, closest to the differential decay rate, the Kiselev and EFG hadronic tensor helicity components are in broad agreement, with the biggest difference in H_+.
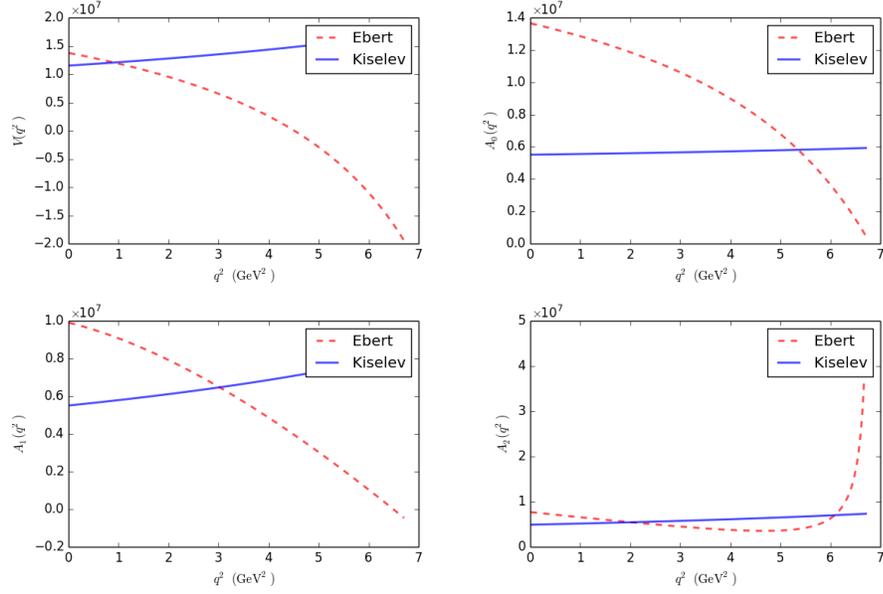
Figure A.3: None of the standard form factors of the Kiselev and EFG models of the $B_c^+ \to \psi(2S)$ transition agree well.
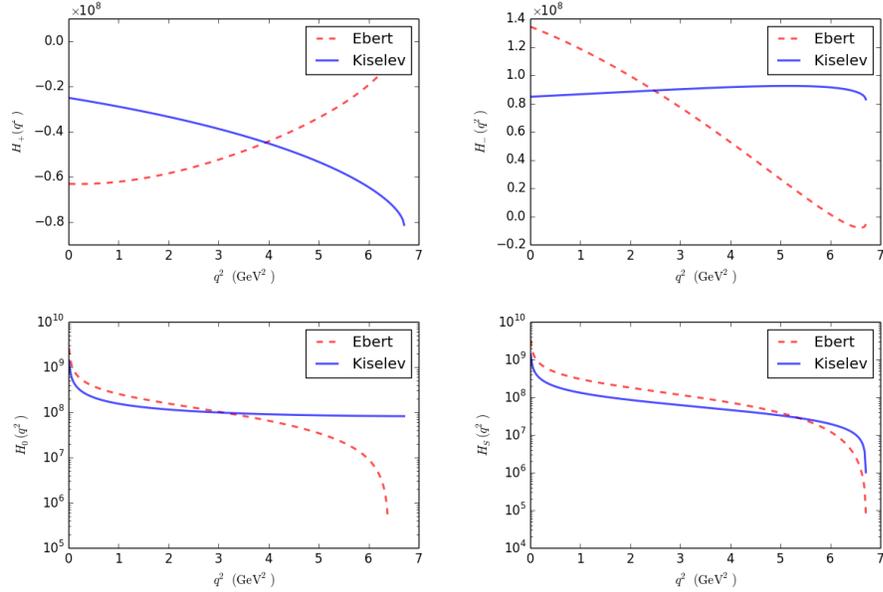
Figure A.4: The huge difference in the Kiselev and EFG models of the $B_c^+ \to \psi(2S)$ transition is due to their H_+ and H_- hadronic tensor helicity components.
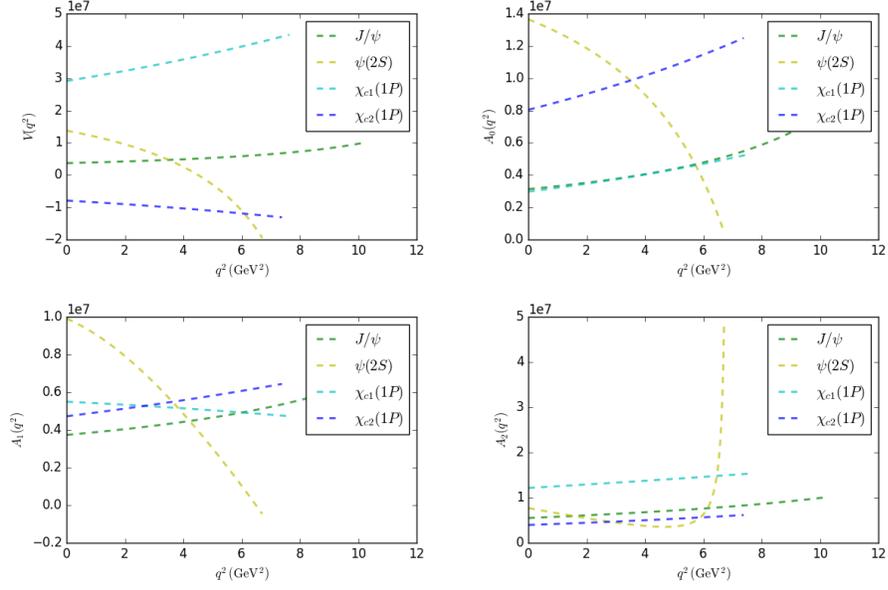
Figure A.5: Form factors for J/ψ and $\psi(2S)$ (from EFG) and χ_{c1} and χ_{c2} (from WWL).
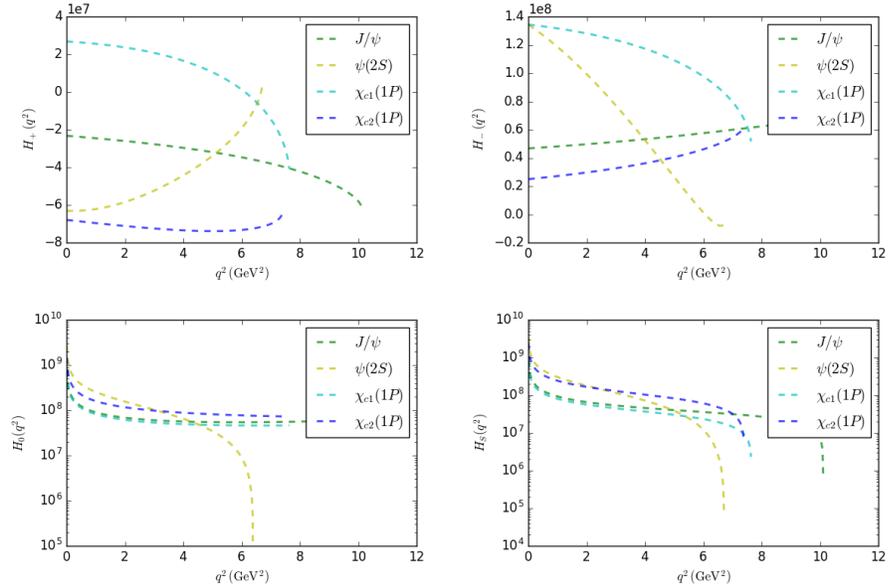
Figure A.6: Helicity form factors for J/ψ and $\psi(2S)$ (from EFG) and χ_{c1} and χ_{c2} (from WWL).

Appendix B: Rest frame variables

Using the (approximated) B_c^+ momentum, several rest frame variables are used in the fit, notably the missing mass squared (Eq. (4.1)), the squared four-momentum transfer q^2, the lepton helicity angle θ, and the muon energy in the rest frame E_μ^*. These are not all independent observables. When there is no missing mass, E_μ^* is a function of q^2 and $\cos(\theta)$:

$$E_\ell^* = \frac{M_{B_c}}{2} \left[(1 - Rw) \left(1 + \frac{m_\mu^2}{q^2} \right) - R\sqrt{w^2 - 1} \left(1 - \frac{m_\mu^2}{q^2} \right) \cos\theta \right] \tag{B.1}$$

where $w = v_{B_c} \cdot v_{J/\psi} = (M_{B_c}^2 + M_{J/\psi}^2 - q^2)/(2M_{B_c}M_{J/\psi})$ and $R = M_{J/\psi}/M_{B_c}$. Neglecting the muon mass, this is approximately

$$E_\ell^* = \frac{M_{B_c}}{2} \left(1 - Rw - R\sqrt{w^2 - 1} \cos\theta \right). \tag{B.2}$$

Given a joint distribution for q^2 and $\cos(\theta)$, the distribution of E_ℓ^* can be evaluated numerically, which is useful when investigating the effect of the B_c^+ form factors. For a given value of E_ℓ^*, the integration region covering the allowed values of q^2 and $\cos\theta$ is difficult to define analytically, except in the approximation $m_\mu \to 0$. However, the range of $\cos\theta$ is unrestricted, and the extreme values of q^2 lie along the curves $\cos\theta = \pm 1$, and so the integration region can be evaluated numerically.

When instead the missing mass is nonzero, such as in the signal decay $B_c^+ \to$

$J/\psi \tau^+ \nu_\tau$, E_μ^* instead depends on the missing mass as well:

$$E_\ell^* = \frac{M_{B_c}}{2} \left[(1 - Rw) \left(1 + \frac{m_\ell^2 - m_{\text{miss}}^2}{q^2} \right) - R\sqrt{w^2 - 1} \frac{\sqrt{X(q^2)}}{q^2} \cos\theta \right], \qquad \text{(B.3)}$$

where $X(q^2)$ is a completely symmetric polynomial of m_ℓ^2, m_{miss}, and q:

$$X(q^2) = (m_\ell + m_{\text{miss}} + q)(m_\ell + m_{\text{miss}} - q)(m_\ell - m_{\text{miss}} + q)(m_\ell - m_{\text{miss}} - q)$$
$$= \left(q^2 - m_\ell^2 + m_{\text{miss}}^2 \right)^2 - 4m_\ell^2. \qquad \text{(B.4)}$$

For tauonic decays, the energy E_μ^* of the muon decay product of the tau is the actual quantity measured in the analysis, together with a non-zero missing mass, and neglecting experimental resolution they obey the above relationship. The true tau energy E_τ in $B_c^+ \to J/\psi \tau^+ \nu_\tau$, which is studied below in relation to the $B_c^+ \to J/\psi$ hadronic form factors, is not measurable because the τ is not reconstructed. Its distribution affects the observed m_{miss}^2 and E_μ^* distributions.

The above can be derived from Lorentz algebra. Alternatively, we observe from Fig. 3.3 that the rest frame energy of a lepton with mass m_ℓ when the missing mass is m_ν is equal to the energy of the neutrino system E_ν when the missing mass is m_ℓ, the lepton mass is m_ν, and $\theta \mapsto \pi/2 - \theta$. Moreover, conservation of energy in the rest frame requires that $E_\ell^* + E_\nu = M_{B_c}(1 - Rw)$. Together, these two observations imply that the quantity

$$F(m_\ell, m_\nu, \cos\theta) = E_\ell^* - \frac{M_{B_c}}{2}(1 - Rw)$$

has a particular antisymmetric form, where

$$F(m_\ell, m_\nu, \cos\theta) = -F(m_\nu, m_\ell, -\cos\theta).$$

Letting $F = A(m_\ell, m_\nu) + B(m_\ell, m_\nu)\cos\theta$, it follows that $A(m_\ell, m_\nu) = -A(m_\nu, m_\ell)$

and that $B(m_\ell, m_\nu) = A(m_\nu, m_\ell)$. That is, A is an antisymmetric function and B a symmetric function of the two masses. Knowing their values at $m_\nu = 0$, we can obtain Eq. (B.3) by ansatz, ignoring higher order terms in $m_\ell m_\nu$.

Appendix C: Mathematical details of misidentification unfolding

As described in Section 4.8, a control sample of $J/\psi\, h^+$ candidates disjoint to the nominal data sample is created using hadron tracks that are not identified as muons and then split into 5 categories enriched in pions, kaons, protons, electrons, and contaminant muons, and one category of all remaining tracks which is relatively rich in ghost tracks. A track that falls into the pion-enriched subsample is referred to as a tagged pion; a similar convention is used for the other hadron species. The index h refers to a generic true hadron or lepton species, while $\hat{h}$ refers to a tagged hadron or lepton species. Thus, the categorization sorts tracks of a number of species $\{\pi, K, p, \mu, e, g\}$ into categories $\{\hat{\pi}, \hat{K}, \hat{p}, \hat{\mu}, \hat{e}, \hat{g}\}$, where g stands for ghost tracks.

The PID samples (including samples produced by the collaboration as well as several custom sources described in Section 4.8) are used to measure the probability that a track of a given true species h that is not identified as a muon candidate (typically by failing to penetrate the muon chamber and leave a muon track) will be tagged as a member of the $\hat{h}'$ category (*i.e.*, be an h' candidate). This probability depends on kinematic properties of the track and global properties of the event and is calculated in several bins of a few relevant observables — the momentum p, pseudorapidity η, and the number of tracks in the event, or track multiplicity. The probability for a track of species h that fails muon identification to pass the selection for $\hat{h}'$ is referred to as $P(\hat{h}' \mid h)$. The PID samples are also used to obtain the probabilities that a track of the species h which is identified as a muon candidate

passes the strict PID criteria used in the analysis, referred to as $P(\hat{\mu}_{\text{strict}} \mid h)$.[1]

In a particular region of momentum, pseudorapidity, and track multiplicity phase where the various classification probabilities and mis-ID probabilities are known, with N_h candidates of each hadron species h in the control sample, the average numbers of tracks in the tagged categories, disregarding statistical fluctuations, are related to the numbers in the true hadron categories through:

$$
\begin{pmatrix} N_{\hat{\pi}} \\ N_{\hat{K}} \\ \vdots \\ N_{\hat{g}} \end{pmatrix} = \begin{pmatrix} P(\hat{\pi} \mid \pi) & P(\hat{\pi} \mid K) & \cdots & P(\hat{\pi} \mid g) \\ P(\hat{K} \mid \pi) & P(\hat{K} \mid K) & \cdots & P(\hat{K} \mid g) \\ \vdots & \vdots & \ddots & \vdots \\ P(\hat{g} \mid \pi) & P(\hat{g} \mid K) & \cdots & P(\hat{g} \mid g) \end{pmatrix} \begin{pmatrix} N_{\pi} \\ N_K \\ \vdots \\ N_g \end{pmatrix}, \tag{C.1}
$$

or more concisely

$$
N_{\hat{h}'} = \sum_h P(\hat{h}' \mid h) N_h. \tag{C.2}
$$

This linear relationship in C.2 must be inverted to obtain the counts N_h given the observed counts $N_{\hat{h}}$. The exact solution is not appropriate because Eq. (C.1) is a statistical, not exact, equality: matrix inversion in the presence of statistical fluctuations can produce nonsensical results, where some very large positive N_h is offset by a very large negative $N_{h'}$.

A more stable solution makes use of the iterative Bayesian procedure, implemented in the `RooBayesUnfold` tool provided by the `RooUnfold` package [123]. The relationship in Eq. (C.2) is interpreted as an unfolding problem, with the matrix of classification probabilities as the response matrix. In the first step, the prior distribution of the species h of the tracks in the bin is taken to be uniform, *i.e.*, it is assumed that $N_h/N = 1/6$ for each species h. Then, using the observed counts and the conditional probabilities $P(\hat{h}' \mid h)$, the posterior distribution is calculated. In

---

[1]By definition, $P(\hat{\mu}_{\text{strict}} \mid \mu) = 0$, because the counterpart of the muon tracks that fail muon identification are not fake muons but actual muons passing muon identification.

this subsequent step this posterior distribution is used as the new prior distribution, and this is repeated for several iterations (seeking exact convergence in this iterative procedure is known to produce overfit results; a small number of iterations around 5 is suggested).

The posterior distribution from the final iteration provides the inferred fractions $P_{\text{unfold}}(h) = \hat{N}_h/N$. One simple way to use these would be to argue that, since the effective number of fake muons N_{fake} is related to the fake rates $P(\hat{\mu}_{\text{strict}} \mid h)$, the weight

$$w_{\text{fake,average}} = \sum_h P(\hat{\mu}_{\text{strict}} \mid h) P_{\text{unfold}}(h), \tag{C.3}$$

should be applied to all N tracks. However, this would assign equal weights to tracks from different hadron species $\hat{h}'$, and so if the distribution of the fit variables differs between the hadron species, this method would not accurately reproduce the distribution of the misidentified background. This is in fact the case for this analysis: a large subcomponent of the mis-ID background comes from fully reconstructed $B^+ \to J/\psi K^+$ decays, which peak in missing mass squared, and so in this region (which is not necessarily in correspondence with one of the kinematic bins in which the unfolding is performed) the kaon component is a larger fraction than elsewhere. Giving the tracks a weight that is the average fake rate for different species would not reproduce the peak in missing mass well.

Instead, this posterior result is then one more time used as a prior distribution to compute the inverted probabilities $P(h \mid \hat{h}')$ via Bayes' rule:

$$P(h \mid \hat{h}') = \frac{P(\hat{h}' \mid h) P_{\text{unfold}}(h)}{\sum_{h''} P(\hat{h} \mid h'') P_{\text{unfold}}(h'')} \tag{C.4}$$

A track tagged as hadron species $\hat{h}'$ is then given the weight

$$w_{\text{fake}}(\hat{h}') = \sum_h P(\hat{\mu}_{\text{strict}} \mid h) P(h \mid \hat{h}'). \tag{C.5}$$

172

Appendix D: Extra figures for data–simulation correction

(a) MC

(b) Data
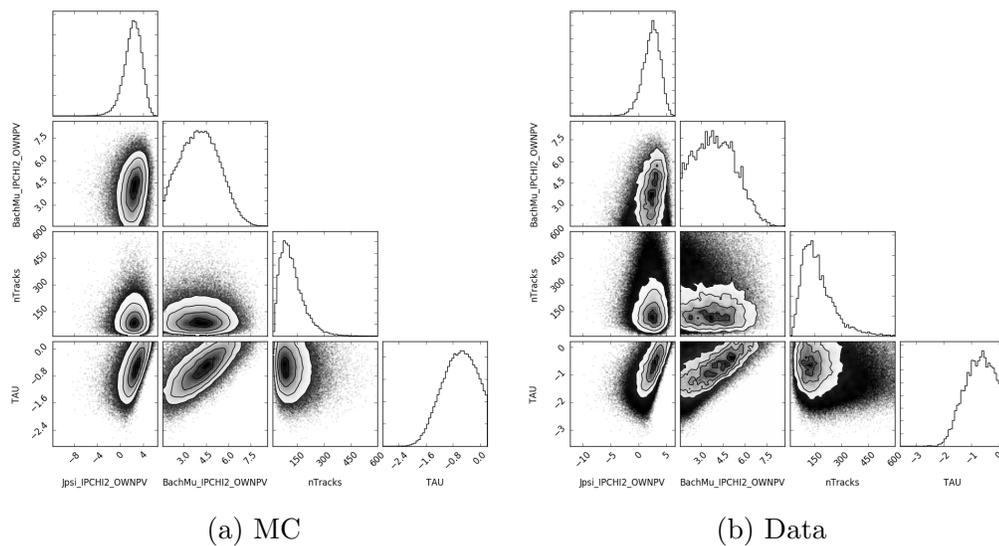
Figure D.1: Distribution of the track multiplicity and impact parameter significances in simulation and in normalization-rich data. These are also compared to the B_c^+ decay time. The distributions of the significances and decay time are shown in log scale.

(a) Before

(b) Data
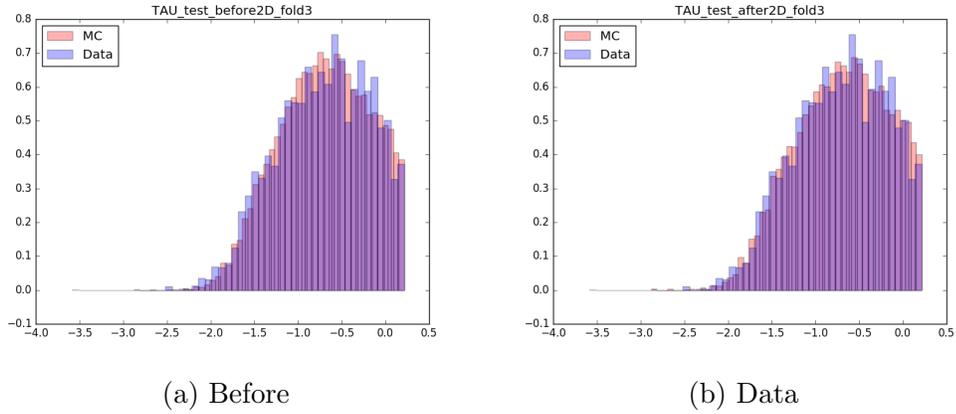
Figure D.2: Before training the gradient-boosted reweighter, the simulation and data samples are weighted to have equivalent decay time distributions. Then the decay time is included as a constraint in the reweighter so that its distribution is not altered.

(a) Before

(b) After
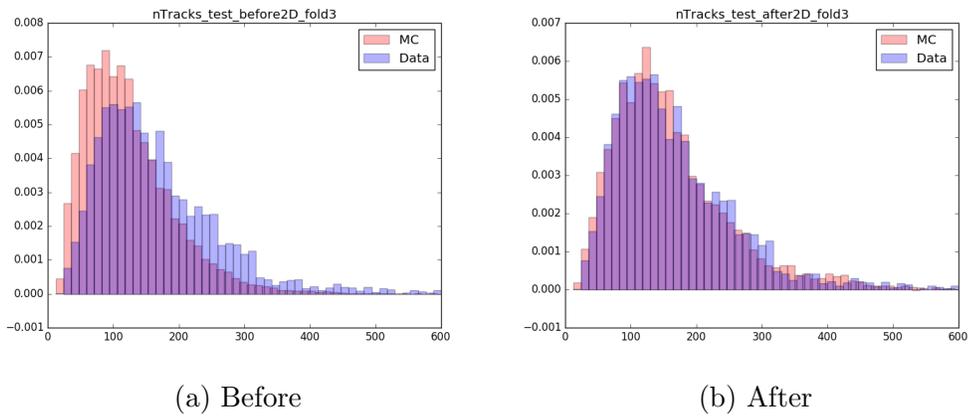
Figure D.3: The track multiplicity is corrected to match the higher activity distribution observed in data.

174

(a) Before   (b) After
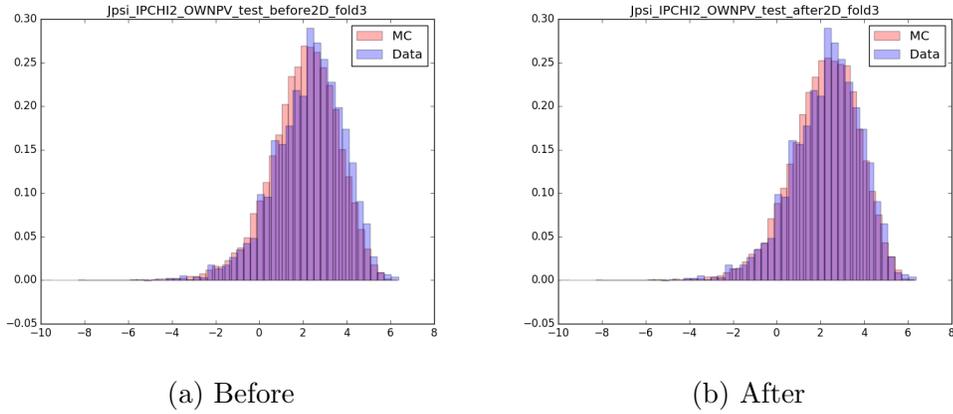
Figure D.4: The J/ψ impact parameter significance is shifted slightly higher to match the data.

(a) Before   (b) After
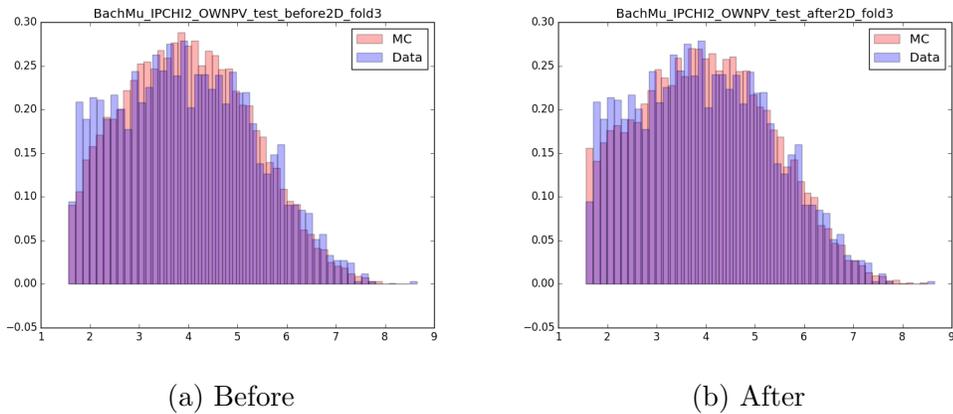
Figure D.5: The μ^+ impact parameter significance distribution is weighted up slightly at small values.
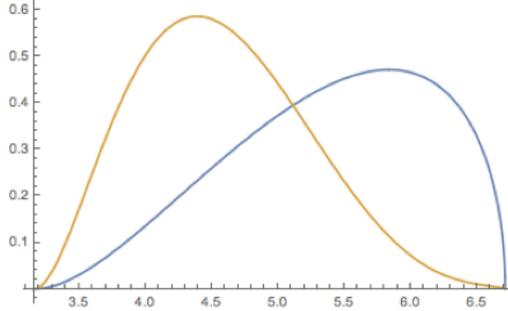
Figure E.6: The q^2 spectra of $B_c^+ \to \psi(2S)\mu^+\nu_\mu$ are completely distinct in the EFG model (orange) and Kiselev model (blue).

E.1 $\psi(2S)$ form-factor resampling

For the $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decay, the Kiselev and EFG form factor models are quite similar. Our BCL parameterization represents a compromise between them (with freedom to vary in the fit to the normalization-rich region), and event-by-event weights can easily model the difference between the two form factor models. For the $B_c^+ \to \psi(2S)\mu^+\nu_\mu$ decay, however, the predictions of the two models are very different, as shown in Fig. E.6. In the Kiselev model, unlike the EFG model, the $\psi(2S)$ decays populate the high q^2 and moderate E_μ^* region of phase space. A simple reweighting of events generated with the EFG model (as our MC events are) cannot accurately represent the distribution under the Kiselev model, because a small number of events will be given very large weights.

Instead, to generate a representative sample of the $\psi(2S)$ decay, we use a fast MC technique. The distributions of the fit variables m_{miss}^2, q^2, and E_μ^* are shaped by selection and detector resolution effects. Experience indicates that, apart from the most important kinematic cuts, the latter dominate. With this in mind, we apply a loose selection (including principally the p_{T} cuts on the muons) to a sample of $10k$ generator-level MC events produced with the Kiselev form factors. This dataset contains the true values of m_{miss}^2, q^2, E_μ^*, and decay time under the B_c^+ rest frame approximation. Meanwhile, our large sample of EFG-model $\psi(2S)$

(a) m_{miss}^2

(b) q^2

(c) E_μ^*
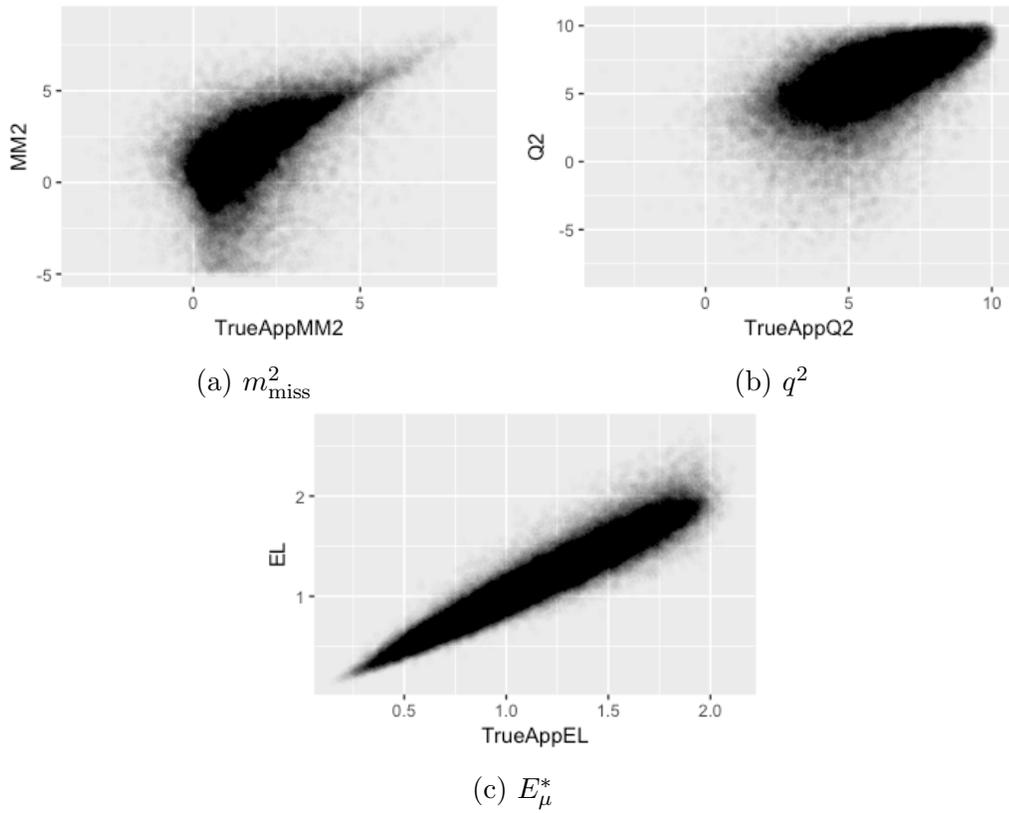
Figure E.7: The distributions of the reconstructed parameters vs. their true values in MC generated $\psi(2S)$ events in the EFT model.

(a) m_{miss}^2 (b) q^2

(c) E_μ^*
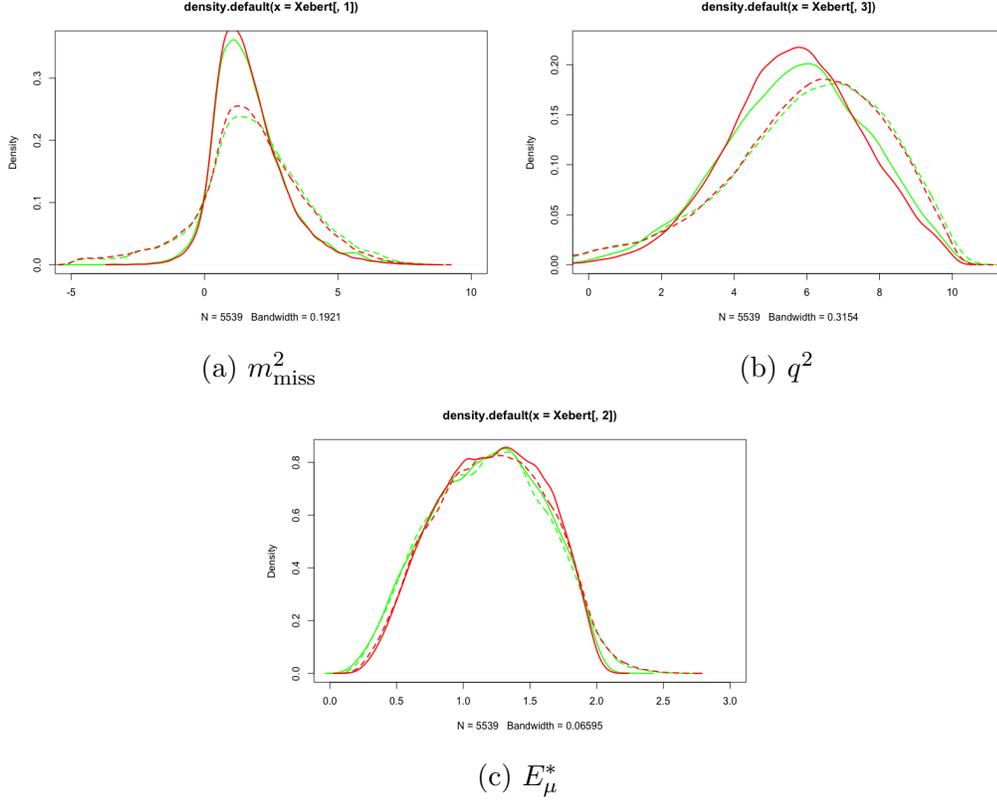
Figure E.8: The results of the kNN resampling closure test. Shown, from both the EFG generator-level MC (green) and independent full MC (red), are the distributions of the true variables (solid line) and the resampled or directly reconstructed distributions, respectively (dashed).

MC contains not only the true values of these variables but their reconstructed values (whose relationship is partially shown in Fig. E.7), as well as the PID weight and isolation BDT score that are used in the final selection and are correlated with the fit variables. For each set of values of of m_{miss}^2, q^2, E_μ^* and decay time in the Kiselev model generator-level MC dataset, we find the k events in the Ebert model full-MC sample with the closest MC-truth values, as determined by k-nearest-neighbors (kNN). Then, we randomly pick one event from these k events and assign its reconstructed values and PID weight and isolation BDT score to the Kiselev MC. The number k was chosen to be 5 through cross validation (maximizing the product of one-dimensional Kolmogorov-Smirnov tests).

As a closure test, the above procedure was tested with an independent sample

density.default(x = Y[, 2])   density.default(x = Y[, 4])

(a) E_μ^* (b) E_μ^*
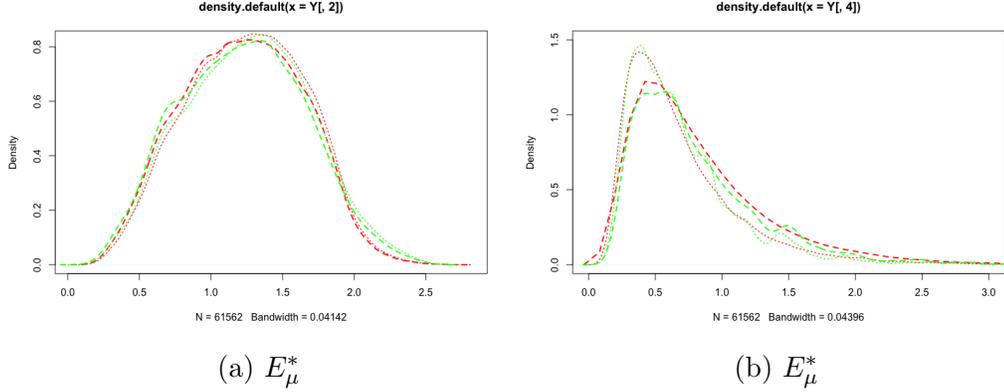
Figure E.9: The results of the kNN resampling closure test, continued. Shown, from both the EFG generator-level MC (green) and independent full MC (red), are the resampled or directly reconstructed distributions, respectively, before (dashed) and after (dotted) applying the PID and isolation BDT correction/selection.

of generator-level EFG model MC containing $10k$ events. The results of this are shown in Figs. E.8 and E.9. From this closure test, we conclude that the kNN sampling technique adequately smears the generator-level MC variables for the purpose of evaluating the systematic uncertainty due to the unknown $\psi(2S)$ form factors. There are some remaining discrepancies due to acceptance effects, the largest of which is the lifetime acceptance and which is manually corrected after-the-fact. The results of the kNN smearing technique for the generator-level Kiselev MC are shown in Figs. E.10 and E.11. As expected, the Kiselev model is still weighted towards a higher q^2 region than the Ebert model after resampling.

(a) m_{miss}^2

(b) q^2

(c) E_μ^*
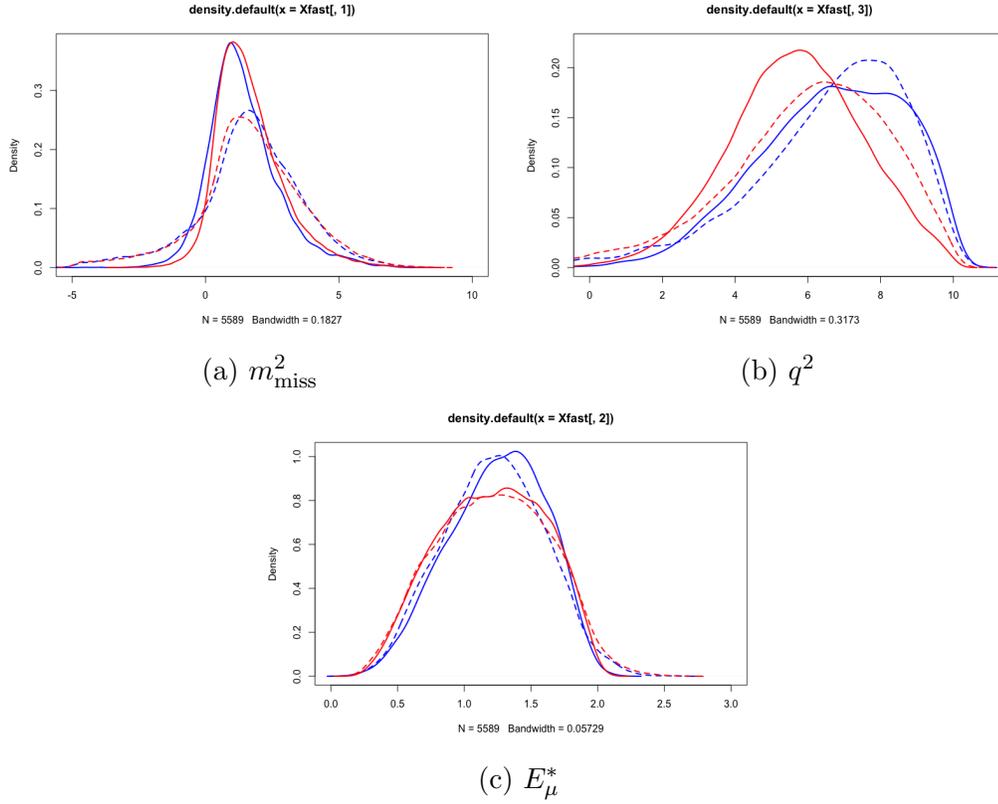
Figure E.10: The results of the kNN resampling. Shown, from both the Kiselev generator-level MC (green) and EFG full MC (red), are the distributions of the true variables (solid line) and the resampled or directly reconstructed distributions, respectively (dashed).

(a) E_μ^*

(b) E_μ^*
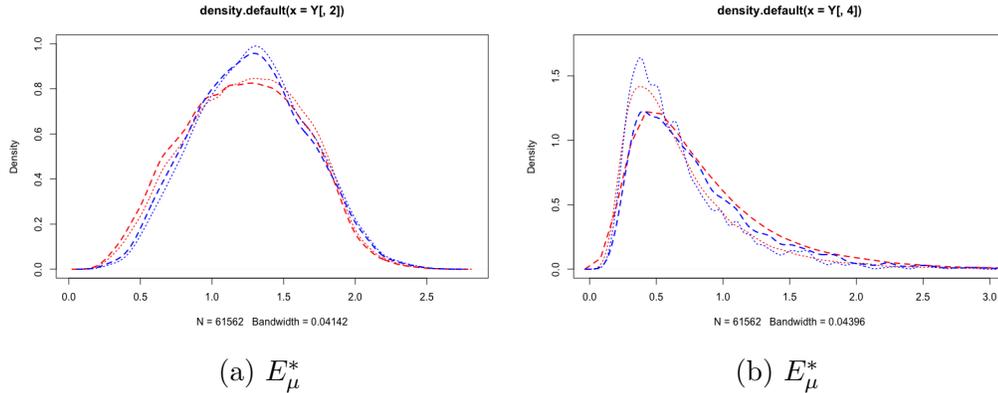
Figure E.11: The results of the kNN resampling, continued. Shown, from both the Kiselev generator-level MC (green) and EFG full MC (red), are the resampled or directly reconstructed distributions, respectively, before (dashed) and after (dotted) applying the PID and isolation BDT correction/selection.

Appendix F: Flavor tagging performance metrics

The usual goal of a CP analysis is often to measure asymmetries in the rates of two CP-conjugate processes. For example, the processes might be the rates of the decays $B^0 \to J/\psi\,K_S$ and $\overline{B}^0 \to J/\psi\,\phi$, decays subject to CP violation in the interference between mixing and decay and whose different probabilities are related to the CP-violating parameters $\sin 2\beta$ and ϕ_s. In most analyses at LHCb, these processes involve decays of b-hadrons (especially B^+, B^0, and B_s^0 mesons), and the flavor of the parent b-quark marks the population to which the process belongs. The flavor of the process is labeled by $d'$, equal to $+1$ for $\bar{b}$ ($e.g.$, B^+, B^0, and B_s^0) and -1 for b ($e.g.$, B^-, $\overline{B}^0$, and $\overline{B}_s^0$). If the process exhibits mixing ($e.g.$, for B^0 and B_s^0 decays) and there is a separate decay flavor, this is labeled by $d'' = \pm 1$. If the conjugate processes have different probabilities $p_{d'}$, the quantity measured in the analysis is an asymmetry of the form

$$\mathcal{A} = \frac{p_{+1} - p_{-1}}{p_{+1} + p_{-1}} \tag{F.1}$$

Observed processes must be sorted into one of two CP-conjugate populations $\mathcal{P}_{d'}$. In simple cases, such as studies of direct CP violation in $B^\pm$ decays, d' is an observable, but frequently it is not, as in the $\sin 2\beta$ example presented above where the final state $J/\psi\,K_S^0$ does not tag the flavor d', and flavor tagging algorithms must be used that produce a tagged ($i.e.$, predicted) flavor d.

181

F.1 Likelihood analysis of an asymmetry measurement

In practice, physics analyses measuring CP-violating asymmetries involve complicated models describing the distributions of the processes in several variables (such as invariant mass and decay time, typically) with many additional floating parameters (Δm, $\Delta\Gamma$, particle masses and resolutions, $etc.$). However, a simplified model that ignores these complications can shed light on how the properties of the flavor tagging algorithm impact the measurement of the asymmetry.

With either first-hand knowledge of d' or a perfectly accurate tagging algorithm (that always produce $d = d'$), the asymmetry in Eq. (F.1) can be measured directly via the estimator

$$\hat{\mathcal{A}} = \frac{\mathcal{N}'_+ - \mathcal{N}'_-}{\mathcal{N}'_+ + \mathcal{N}'_-}, \tag{F.2}$$

where $\mathcal{N}'_{d'}$ count the number of events in the populations $\mathcal{P}'_{d'}$. Assuming uncorrelated Poisson uncertainties on $\mathcal{N}'_{d'}$ and defining the total number of events $\mathcal{N}' = \mathcal{N}'_+ + \mathcal{N}'_-$, the uncertainty on the measured asymmetry is

$$\hat{\sigma}_{\mathcal{A}} = \sqrt{\frac{1 - \hat{\mathcal{A}}^2}{\mathcal{N}'}} \tag{F.3}$$

Alternatively, the likelihood of obtaining the measured counts given an a priori asymmetry is given by the binomial distribution pmf:

$$\mathcal{L}(\mathcal{N}'_+, \mathcal{N}'_- \mid \mathcal{A}) \propto \left(\frac{1+\mathcal{A}}{2}\right)^{\mathcal{N}'_+} \left(\frac{1-\mathcal{A}}{2}\right)^{\mathcal{N}'_-} \tag{F.4}$$

The maximum likelihood estimate is found by setting the derivative of the log-likelihood

$$\frac{\partial \log \mathcal{L}(\mathcal{N}'_+, \mathcal{N}'_- \mid \mathcal{A})}{\partial \mathcal{A}} = \frac{\mathcal{N}'_+}{1+\mathcal{A}} - \frac{\mathcal{N}'_-}{1-\mathcal{A}} \tag{F.5}$$

to zero, recovering Eq. (F.2)) as expected. The canonical variance on the estimate

is given by the second derivative

$$\frac{1}{\hat{\sigma}_{\mathcal{A}}^2} = -\left.\frac{\partial^2 \log \mathcal{L}(\mathcal{N}'_+, \mathcal{N}'_- \mid \mathcal{A})}{\partial^2 \mathcal{A}}\right|_{\hat{\mathcal{A}}} = \left.\frac{\mathcal{N}'_+}{(1+\mathcal{A})^2} + \frac{\mathcal{N}'_-}{(1-\mathcal{A})^2}\right|_{\hat{\mathcal{A}}} = \frac{\mathcal{N}'}{1-\hat{\mathcal{A}}^2}, \qquad \text{(F.6)}$$

in agreement with Eq. (F.3)). Likelihood maximization is perhaps more formal than required to obtain these results, but it is easier to apply to generalized situations that will be encountered later.

It is more often the case that flavor tagging algorithms cannot tag observed processes as $d' = +1$ or $d' = -1$ with perfect accuracy. A generic tagging algorithm, besides producing a predicted flavor category d that an event belongs to, also provides a per-event probability ω, called the mistag probability, that its tag is incorrect. First take the simple case that the mistag probability is a constant Ω. Call the numbers of events belonging to the population $\mathcal{P}_d$ of events tagged (possibly incorrectly) as type d $\mathcal{N}_d$. Then, the probability for the process to be tagged as $d = +1$ is not $(1 + \mathcal{A})/2$ but rather

$$(1 - \Omega) \times \frac{1 + \mathcal{A}}{2} + \Omega \times \frac{1 - \mathcal{A}}{2} = \frac{1 + D\mathcal{A}}{2} \qquad \text{(F.7)}$$

where $D = 1 - 2\Omega$ is called the dilution of the tagger.[1] Since Ω is the mistag probability and $1 - \Omega$ the success rate, the dilution is the difference between the success and mistag probabilities, and counts the asymmetry between the numbers of correctly and incorrectly tagged events. Likewise, the probability to be tagged as $d = -1$ is $(1 - D\mathcal{A})/2$. So, the log-likelihood function becomes

$$\log \mathcal{L}(\mathcal{N}_+, \mathcal{N}_- \mid \Omega, \mathcal{A}) = C + \mathcal{N}_+ \log\left(\frac{1 + D\mathcal{A}}{2}\right) + \mathcal{N}_- \log\left(\frac{1 - D\mathcal{A}}{2}\right). \qquad \text{(F.8)}$$

---

[1]This quantity is unfortunately named, because a tagger with large dilution (and thus low performance) corresponds to a small value of D, and vice versa.

The estimator for $\mathcal{A}$ and its variance become

$$\hat{\mathcal{A}}_\Omega = \frac{1}{D}\frac{\mathcal{N}_+ - \mathcal{N}_-}{\mathcal{N}_+ + \mathcal{N}_-} \tag{F.9}$$

and

$$\hat{\sigma}_{\mathcal{A}_\Omega} = \sqrt{\frac{1 - D^2\hat{\mathcal{A}}^2}{D^2\mathcal{N}}} \tag{F.10}$$

To zero-th order in $\mathcal{A}$ (since $D\mathcal{A}$ is typically quite small), the effective statistical size of the sample is reduced from $\mathcal{N}$ by a factor D^2. This quantity is called the tagging power of the tagging algorithm, and is usually denoted ϵ_{eff}.[2]

F.1.1 Incorporating per-event mistag probability

More generally, the predicted mistag probability varies from event to event; some events are relatively clean and have a mistag probability as low as 10% to 15%, but most are much more chaotic and have a mistag probability not much better than 50%. For a given event, there are probabilities $p(d = +1 \mid d' = +1)$ ($p(+ \mid +)$ for short) for the tag decision (*i.e.*, the predicted flavor) to be $d = +1$ given that the actual CP flavor is $d' = +1$ and $p(d = +1 \mid d' = -1)$ ($p(- \mid -)$ for short) for the tag decision to be $d = +1$ given that the actual flavor is $d' = -1$. Given an asymmetry $\mathcal{A}$, the likelihood of a tag decision $d = +1$ is thus

$$p(d = +1) = p(+ \mid +)\frac{1 + \mathcal{A}}{2} + p(+ \mid -)\frac{1 - \mathcal{A}}{2} \tag{F.11}$$

The two quantities $p(+ \mid +)$ and $p(+ \mid -)$ are clearly related to the tag flavor decision and predicted mistag ω: $p(+ \mid +) = 1 - \omega$, since it is a correct tag, and

---

[2]In practice, a flavor tagging algorithm is not able to tag all signal decays. The fraction of events tagged by an algorithm is called its tagging efficiency ϵ. For simplicity, in this document ϵ is assumed to be 100%, or equivalently that the number of events N is the number of tagged events.

$p(+ \mid -) = \omega$, since it is incorrect. In general, the likelihood of the tag d is

$$p(d) = \frac{1}{2}\left(1 + dD\mathcal{A}\right) \tag{F.12}$$

where the dilution $D = 1 - 2\omega$ now varies from event to event. Ignoring some constants, the complete log-likelihood function[3] is then

$$\log(\mathcal{L}(\mathcal{A})) = -N\log(2) + \sum_i \log\left(1 + d_i D_i \mathcal{A}\right). \tag{F.13}$$

The estimator $\mathcal{A}$ can be solved by maximizing this log-likelihood, whose derivative is

$$\frac{\partial \log(\mathcal{L}(\mathcal{A}))}{\partial \mathcal{A}} = \sum_i \frac{d_i D_i}{1 + d_i D_i \mathcal{A}}. \tag{F.14}$$

Given an arbitrary number of events with arbitrary dilutions D_i, there is no apparent exact analytic solution to the above equation. Since CP-violating asymmetries are typically quite small, an expansion of the solution in powers of $\mathcal{A}$ can be useful. To first order, the score function of Eq. (F.14) is

$$\frac{\partial \log(\mathcal{L}(\mathcal{A}))}{\partial \mathcal{A}} = \sum_i d_i D_i - \mathcal{A} \sum_i D_i^2 + \mathcal{A}^2 \sum_i d_i D_i^3 + \mathcal{O}\left(\left\langle D^4 \right\rangle \mathcal{A}^3\right) \tag{F.15}$$

, and the corresponding approximate solution is

$$\hat{\mathcal{A}} = \frac{\langle dD \rangle}{\langle D^2 \rangle} + \mathcal{O}\left(\left\langle D^3 \right\rangle \mathcal{A}^2\right) \tag{F.16}$$

---

[3]Here, the mistag probability ω is treated as fixed and known; in practice, however, the calibration of the tagging algorithm provides a standard error that can be incorporated into the likelihood.

Defining $\mathcal{N}_+^w$ and $\mathcal{N}_-^w$ to be the weighted number of events tagged as $d = +1$ and $d = -1$, respectively, this can be recast as

$$\hat{\mathcal{A}} = \frac{\langle D \rangle}{\langle D^2 \rangle} \frac{\mathcal{N}_+^w + \mathcal{N}_-^w}{\mathcal{N}_+^w - \mathcal{N}_-^w} + \mathcal{O}\left(\left\langle D^3 \right\rangle \mathcal{A}^2\right) \tag{F.17}$$

In this form the comparison to Eq. (F.2) and Eq. (F.9) is more obvious. The first order estimate of $\hat{\mathcal{A}}$ is essentially a weighted asymmetry of the two counts, multiplied by an overall factor accounting for the dilution. Given that the achievable dilution D with LHCb's flavor tagging algorithms is on average quite small, in the 0.2 to 0.4 range, and at most 0.5, the corrections to Eq. (F.16) are typically quite small.

To first order in $\mathcal{A}$, the second derivative of the log-likelihood is

$$-\frac{\partial^2 \log(\mathcal{L}(\mathcal{A}))}{\partial \mathcal{A}^2} = \sum_i D_i^2 - 2\mathcal{A} \sum_i d_i D_i^3 + \mathcal{O}\left(\left\langle D^3 \right\rangle \mathcal{A}^2\right)$$
$$= \left\langle D^2 \right\rangle \mathcal{N} + 2\mathcal{A} \left\langle dD^3 \right\rangle \mathcal{N} + \mathcal{O}\left(\left\langle D^3 \right\rangle \mathcal{A}^2\right).$$

The standard error on the estimator is the value of this second derivative at the estimated value. To zero-th order, the standard error of the estimate is

$$\hat{\sigma}_{\hat{\mathcal{A}}}^2 \approx \frac{1}{\langle D^2 \rangle \mathcal{N}} + \mathcal{O}\left(\mathcal{A}\right), \tag{F.18}$$

making the tagging power of the algorithm equal to the RMS value of the dilution:

$$\epsilon_{\text{eff}} = \left\langle D^2 \right\rangle + \mathcal{O}\left(\mathcal{A}\right). \tag{F.19}$$

F.1.2   A toy example of tagging power

To measure $\mathcal{A}$ exactly, the likelihood must be maximized numerically. Take as a relatively simple example a tagger whose mistag probability is uniformly distributed between 0 and 0.5, and so whose dilution is uniformly distributed between

0 and 1. The canonical tagging power in this case is simply $\int_0^1 D^2 \, dD = 1/3$. For convenience define the continuous classifier $\delta = dD$ (not to be confused with the integration element). To produce an observed asymmetry $\mathcal{A}_{\mathrm{obs}}$ in the tagged response, the pdf of δ must be $p(\delta) = 1/2 + \mathcal{A}_{\mathrm{obs}}\delta$. Correspondingly, the quasi-moment generating function for the distribution is

$$h_\delta(t) = \int \frac{\frac{1}{2} + \mathcal{A}_{\mathrm{obs}}\delta}{1 - t\delta} \, d\delta = \frac{(2\mathcal{A}_{\mathrm{obs}} + t)\operatorname{arctanh}(t) - 2\mathcal{A}_{\mathrm{obs}}t}{t^2}$$

The solution to $h_\delta(-\mathcal{A}) = 1$ and the associated variance in the estimator are simply

$$\hat{\mathcal{A}} = 2\mathcal{A}_{\mathrm{obs}} \tag{F.20}$$

$$\hat{\sigma}_{\hat{\mathcal{A}}}^2 = \frac{\hat{\mathcal{A}}^3}{\mathcal{N}\left(\operatorname{arctanh}(\hat{\mathcal{A}}) - \hat{\mathcal{A}}\right)} \tag{F.21}$$

The limiting value of the variance as $\hat{\mathcal{A}} \to 0$ is $3/\mathcal{N}$, making the canonical tagging power $1/3$. In general, the real tagging power at finite values of the asymmetry is

$$\epsilon_{\mathrm{eff}}(\hat{\mathcal{A}}) = \frac{1}{\mathcal{N}\hat{\sigma}_{\hat{\mathcal{A}}}^2} = \frac{\operatorname{arctanh}(\hat{\mathcal{A}}) - \hat{\mathcal{A}}}{\hat{\mathcal{A}}^3} \tag{F.22}$$

Figure F.1 shows the results of a toy study in which toy datasets of 10k events were generated with a real asymmetry ranging between 0% and 95% and a uniform dilution between 0 and 1. The asymmetry is estimated by numerical likelihood maximization, and the estimated asymmetry and corresponding tagging power agree with Eqs. (F.20) and (F.22). Still, the canonical tagging power $1/3$ is a good approximation up through 50% asymmetries.

The range of asymmetries in this toy study is much wider than the typical asymmetries measured in CP violation studies are usually less than 1% and certainly much less than 10%. In this range, the canonical tagging power (Eq. (F.19)) evaluated at $\hat{\mathcal{A}} = 0$ is an excellent approximation for the exact tagging power eval-

187

(a) Estimated asymmetry

(b) Tagging power of estimator
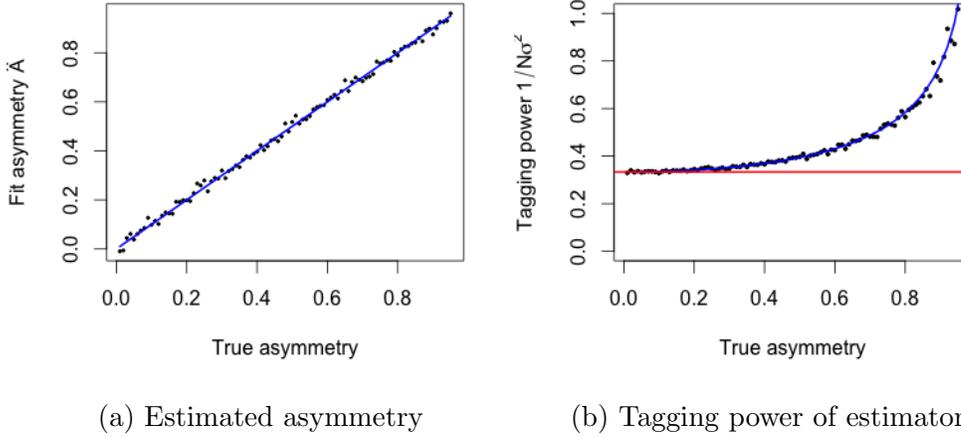
Figure F.1: The asymmetry and estimated numerically from samples of 10k events and associated tagging power (black points) agree very well with the analytic estimates in the asymptotic limit (blue). For small asymmetries the tagging power tends to the canonical tagging power, $1/3$ (red).

Table F.1: Calibrated tag probabilities in the presence of tagging asymmetries. Treating the calibrated mistag values $\omega_\pm$ as givens, the probabilities for each of the tag decisions $d = \pm$ depend on these probabilities and the true flavor d'.

	$d = +1$	$d = -1$
$d' = +1$	$p(+ \mid +) = 1 - \omega_+$	$p(- \mid +) = \omega_+$
$d' = -1$	$p(+ \mid -) = \omega_-$	$p(- \mid -) = 1 - \omega_-$

uated at finite $\hat{\mathcal{A}}$ This has only been demonstrated for the uniform dilution distribution considered in the toy study, but it seems reasonable that this approximation continues to hold even for more realistic dilution distributions.

F.1.3   Incorporating mistag probability asymmetries $\Delta\omega$

In the final analysis, asymmetries in the tagging algorithm itself make the mistag probability different for processes with real flavors $d' = +1$ and $d' = -1$. This means that there are separate mistag values ω_+ and ω_- when $d' = +1$ and $d' = -1$, respectively. So, $p(+ \mid +)$ becomes $1 - \omega_+$, $p(+ \mid -)$ becomes ω_-, and so on. A table of these probabilities is shown in Table F.1. Defining $\Delta\omega = \omega_+ - \omega_-$

and $\omega = (\omega_+ + \omega_-)/2$, the overall probability that the tag decision is d is

$$p(d) = p(d \mid d = +1)\frac{1+\mathcal{A}}{2} + p(d \mid d = -1)\frac{1-\mathcal{A}}{2} = 1 - d\Delta\omega + dD\mathcal{A}. \quad \text{(F.23)}$$

This slightly modifies the log-likelihood, making it

$$\log(\mathcal{L}(\mathcal{A})) = -N\log(2) + \sum_i \log\left(1 - d_i\Delta\omega_i + d_iD_i\mathcal{A}\right) \quad \text{(F.24)}$$

To first order in $\mathcal{A}$, the derivative of the log-likelihood is

$$\frac{\partial \log(\mathcal{L}(\mathcal{A}))}{\partial \mathcal{A}} = \sum_i \frac{d_iD_i}{1 - d_i\Delta\omega_i} - \mathcal{A}\sum_i \frac{D_i^2}{(1 - d_i\Delta\omega_i)^2} + \mathcal{O}(\langle D^3 \rangle \mathcal{A}^2). \quad \text{(F.25)}$$

This makes the corresponding first-order estimate of the asymmetry

$$\hat{\mathcal{A}} = \left\langle \frac{dD}{1 - d\Delta\omega} \right\rangle \Big/ \left\langle \left(\frac{D}{1 - d\Delta\omega}\right)^2 \right\rangle + \mathcal{O}(\langle D^3 \rangle \mathcal{A}^2) \quad \text{(F.26)}$$

and the zero-th order estimate of the tagging power

$$\epsilon_{\text{eff}} = \left\langle \left(\frac{D}{1 - d\Delta\omega}\right)^2 \right\rangle + \mathcal{O}(\langle D^2 \rangle \mathcal{A}). \quad \text{(F.27)}$$

F.2   Relationship between tagging power and AUC score

A tagging algorithm's decision is a discrete classifier between two flavor categories, and its predicted mistag probability is a continuous classifier between two tag categories (correctly tagged and incorrectly tagged). These can be put together into a continuous classifer between the flavor categories: the signed dilution $\delta \equiv dD$, where $d = \pm1$ is the tagged flavor ($+1$ for B and -1 for $\bar{B}$) and $D = 1 - 2\omega$ is the predicted dilution.

A very common metric used to evaluate continuous classification algorithms in

the AUC, which is the area under the receiver operating characteristic (ROC) curve. The AUC score has a straight-forward interpretation. In the context of a tagging algorithm's mistag probability, it is the probability that a random B event will have a larger δ than a random $\bar{B}$ event. Letting $g(\delta)$ be the pdf of the algorithm's signed dilution δ, and g_+ and g_- the densities for $+$ and $-$ events, this modified AUC is

$$AUC \equiv \int g_+(\delta_1)d\delta_1 \int g_-(\delta_2)d\delta_2 \, I(\delta_1 > \delta_2),$$

where $I(\delta_1 > \delta_2)$ is the indicator function equal to 1 when $\delta_1 > \delta_2$ and 0 otherwise. Now we must rewrite g_+ and g_- in terms of g. Ignore differences $\Delta\omega$ in the mistag probability between $+$ and $-$ events. For an event with dilution $D = \|\delta\|$, there is a probability $(1 + \|\delta\|)/2$ that it is tagged correctly and $(1 - \|\delta\|)/2$ that it is tagged incorrectly. The density g_+ corresponds to events that are tagged correctly for $\delta > 0$ and incorrectly for $\delta < 0$, and so its density over the whole range is

$$g_+(\delta) = \frac{\frac{1+\delta}{2}g(\delta)}{\int \frac{1+\delta_1}{2}g(\delta_1)d\delta_1}$$

In the denominator, $\int \delta_1 g(\delta_1)d\delta_1$ is just the average value $\langle \delta \rangle \propto \mathcal{A}$. Since we only need to work to zero-th order in $\mathcal{A}$, we can assume $\mathcal{A} = 0$, and so the above simplifies to

$$g_\pm(\delta) = (1 \pm \delta)g(\delta).$$

The AUC score becomes

$$AUC = \int g(\delta_1)d\delta_1 \int g(\delta_2)d\delta_2 \, (1 + \delta_1)(1 - \delta_2)I(\delta_1 > \delta_2).$$

Exchanging variables $\delta_2 \leftrightarrow \delta_1$ leads to an alternate expression for AUC over the rest of the integration region, where $\delta_1 < \delta_2$, and taking the average of the two integrals

190

yields

$$AUC = \frac{1}{2} \int g(\delta_1)d\delta_1 \int g(\delta_2)d\delta_2 \left(1 + \|\delta_1 - \delta_2\| - \delta_1\delta_2\right).$$

The first term is 1 and the last $\langle\delta\rangle^2 = \mathcal{A}^2$, so to zero-th order in $\mathcal{A}$

$$AUC = \frac{1 + \langle\|\delta_1 - \delta_2\|\rangle}{2}.$$

This is not directly related to the tagging power, unfortunately. However, if we replace the mean absolute value with the RMS value, which should be similar, we get

$$AUC_{\text{mod}} = \frac{1 + \sqrt{\langle(\delta_1 - \delta_2)^2\rangle}}{2} = \frac{1 + \sqrt{2\langle D^2\rangle}}{2}$$

In other words, the AUC is not directly related to the tagging power, but there is a sense in which they are complementary metrics, with the only difference being the choice of an L_1 or L_2 norm. This does not guarantee that maximizing the ROC score of a tagger is equivalent to maximizing its tagging power, but it does give reason to be optimistic that this is usually the case.

Appendix G: Flavor tagging calibration

G.1   Misclassified response within binomial regression

In the simplest flavor tagging calibrations to decays of charged $B^\pm$ mesons, the production flavor of the meson is directly identified by the charges of one or more of its decay products. In this case, applying binomial regression is entirely straightforward: if the predicted flavor agrees (disagrees) with the observed flavor, the tag is correct (incorrect); this response is regressed as a function of the predicted mistag probability. In general, however, a binomial regression tool for calibrating flavor tagging algorithms must work with decay modes of neutral B^0 and B_s^0 mesons where the production flavor is latent and only the decay flavor is observed. The following Section G.2 describes the probabilistic relationship between the production and decay flavors; for describing the regression procedure, it is enough to note that there is a fixed probability $\varpi^c(t)$ that a B^0 or B_s^0 meson that decays at time t has not oscillated from its original production flavor.

G.1.1   Definitions and formalism

As noted above, flavor tagging algorithms are often described in terms of the predicted mistag probability η and true/calibrated mistag probability ω. Because this definition in terms of negatives is somewhat awkward, this section will refer instead to the predicted correct-tag probability $\varpi = 1 - \eta$. A number of quantities (some observable and others latent) are important in the calibration of flavor tagging

192

algorithms. For each event tagged by a flavor tagging algorithm, these relevant pieces of information (given the subscript k for the k-th event) are:

R (latent for B^0 and B_s^0 modes): equal to 1 if the tagged flavor d equals the production flavor d', $i.e.$, the decision made by the algorithm is correct; otherwise equal to 0;

ϖ: the uncalibrated probability that $d = d'$, $i.e.$, $1 - \eta$;

$\varpi^*(\boldsymbol{\theta})$ (to be regressed): the calibrated probability that $d = d'$, a function of the uncalibrated probability ϖ and some calibration parameters $\boldsymbol{\theta}$;

R^c (latent for B^0 and B_s^0 modes): equal to 1 if the production flavor d' equals the decay flavor d'', $i.e.$, if either the signal is a baryon or a charged meson, or, if it is a B^0 or B_s^0, it has not oscillated; otherwise equal to 0;

ϖ^c: the probability that $d' = d''$, which is 1 for a baryon or a charged meson and can be calculated from the decay time for a neutral B^0 or B_s^0 meson;

R^{obs}: equal to 1 if the tagged flavor d equals the observed decay flavor d''; otherwise equal to 0; since this can happen if either $R = 1$ (with probability ϖ^*) and $R^c = 1$ (with probability ϖ^c) or if $R = 0$ (with probability $1 - \varpi^*$) and $R^c = 0$ (with probability $1 - \varpi^c$), it is equal to $RR^c + (1 - R)(1 - R^c)$. Interpreting these as boolean variables, this is equivalent to $R \leftrightarrow R^c$, a symbol meaning $(R \wedge R^c) \vee (\neg R \wedge \neg R^c)$;

π: the probability that $d = d''$; following the same logic as above this is equal to $\varpi^* \varpi^c + (1 - \varpi^*)(1 - \varpi^c)$, or $\varpi^* \leftrightarrow \varpi^c$ for short.

The calibration of the flavor tagging algorithm is a binomial regression problem where $\pi = \varpi^* \leftrightarrow \varpi^c$ is the predictor of the observable response R^{obs}, wherein ϖ^* is a function of the predicted value ϖ.

G.1.2   Generalized linear models for binomial regression

One particular class of binomial regression model is the generalized linear model [116]. In such a model, the calibrated probability ϖ^* is related to the prediction ϖ by

$$g\left(\varpi^*(\boldsymbol{\theta})\right) = \sum_{k=0}^{M} \theta_k P_k(\varpi) \tag{G.1}$$

where $P_k(\varpi)$ are a set of basis functions, θ_k are a set of calibration parameters expected to be small, and g is called a link function. This is a very flexible class of models. The set of basis functions $P_k(\varpi)$ can include not just simple constant or linear terms but also polynomial bases or various types of basis splines, especially b-splines and natural cubic splines [124]. The link function g is required to be a map from the range $[0, 1]$ to the entire real axis, and so is typically the inverse of the cumulative distribution function of a distribution defined over $\mathbb{R}$. The most popular link function is the logit link (in which case binomial regression is typically called logistic regression).

G.1.3   Likelihood maximization

It is algebraically convenient to define $C = 2R - 1$, equal to ± 1, and the "dilution" $D = 2\varpi - 1$. In this basis the composition of probabilities is simplified:

$$R^{\text{obs}} = R \leftrightarrow R^c \mapsto C^{\text{obs}} = CC^c, \tag{G.2}$$

$$\pi = \varpi^* \leftrightarrow \varpi^c \mapsto D = D^*D^c. \tag{G.3}$$

For each event in the dataset, the probability that $R^{\text{obs}} = 1$ is

$$\pi(\varpi; \boldsymbol{\theta}) = \varpi^*(\boldsymbol{\theta}) \leftrightarrow \varpi^c = \frac{1}{2}\left(1 + D^*(\varpi; \boldsymbol{\theta})D^c\right) \tag{G.4}$$

194

Ignoring the likelihood of getting a particular distribution of dilutions $D^*(\varpi; \boldsymbol{\theta})$ and D^c, which are fixed *a priori* information, the relevant likelihood function is

$$\mathcal{L}\left(\boldsymbol{\theta}; \vec{R}^{\text{obs}}\right) = \prod_{\text{events } e} \pi_e(\boldsymbol{\theta})^{R_e^{\text{obs}}} \left(1 - \pi_e(\boldsymbol{\theta})\right)^{1 - R_e^{\text{obs}}} \tag{G.5}$$

The log-likelihood is given by

$$\ell\left(\boldsymbol{\theta}; \vec{R}^{\text{obs}}\right) = \sum_{\text{events } e} R_e^{\text{obs}} \log\left(\pi_e(\boldsymbol{\theta})\right) + \left(1 - R_e^{\text{obs}}\right) \log\left(1 - \pi_e(\boldsymbol{\theta})\right) \tag{G.6}$$

and typically notated simply $\ell(\boldsymbol{\theta})$. Taking advantage of the notation $C_e^{\text{obs}} = 2R_e^{\text{obs}} - 1$, the log-likelihood is given by

$$\ell(\boldsymbol{\theta}) = \sum_{\text{events } e} \log\left(\frac{1 + C_e^{\text{obs}} D_e^*(\boldsymbol{\theta}) D_e^c}{2}\right). \tag{G.7}$$

The canonical maximum likelihood estimate (MLE) of the calibration parameters is given by the root of the score function, which is the first derivative of the log-likelihood:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} = \sum_{\text{events } e} \frac{2 C_e^{\text{obs}} P_i(\varpi_e) D_e^c}{g'(\varpi^*(\boldsymbol{\theta}))} \times \frac{1}{1 + C_e^{\text{obs}} D_e^*(\boldsymbol{\theta}) D_e^c} \tag{G.8}$$

The root of the system of equations can be found numerically using a Newton-Raphson type algorithm.

G.1.4 Tagging asymmetry parameters

Another requirement for the flavor tagging calibration procedure is that it measure parameters quantifying the dependence of the calibration parameters on the actual flavor of the signal B^0 or B_s^0 meson. There is no guarantee that the out-of-the-box prediction from a flavor tagging algorithm equally well predicts B^0 and $\bar{B}^0$ mesons. For example, the algorithm examining opposite side kaons might

be trained on simulation or data where K^+–K^- detection asymmetries as a function of p_T are different from for the signal sample, so that when the signal is a $\overline{B}^0$, and opposite side kaons are preferentially K^+, the multivariate algorithm that predicts the flavor and η is biased in a different way than if the signal were a B^0. In practice these tagging asymmetry parameters are quite consistent with 0, but it is important to quantify them to form confidence intervals and to propagate their uncertainty to an asymmetry measured in a flavor-tagged physics analysis.

Given that the two flavor states of the signal B^0 or B^0_s meson are labelled by $d' = +1$ and $d' = -1$, there are two calibrated probabilities $\varpi^*(\boldsymbol{\theta}; d')$. A tag decision d can agree with the decay flavor d'' in two ways: it can be correctly tag the production flavor d' (*e.g.*, $d = d'$ and $R = 1$) while the B-meson has not oscillated (*e.g.*, $R^c = 1$), or it can incorrectly tag the production flavor but the B meson has oscillated (*e.g.*, R and R^c are both 0). If the tagged flavor is $d = +1$, the sum of these probabilities (suppressing the dependence on $\boldsymbol{\theta}$) is

$$\pi(d = +1) = \varpi^*(d' = +1)\varpi^c + (1 - \varpi^*(d' = -1))(1 - \omega^c) \tag{G.9}$$

while if the tagged flavor is $d = -1$ the sum is

$$\pi(d = -1) = \varpi^*(d' = -1)\varpi^c + (1 - \varpi^*(d' = +1))(1 - \omega^c) \tag{G.10}$$

In general for a tagged flavor d, the probability that this is equal to the decay flavor d'' is

$$\pi(d) = \varpi^* \leftrightarrow \varpi^c + \frac{d\Delta\varpi^*}{2} \tag{G.11}$$

where ϖ^* is the average of $\varpi^*(d' = \pm)$ and $\Delta\varpi^*$ is the difference $\varpi^*(d' = +) - \varpi^*(d' = -)$.

A simple way to parameterize the separate calibration functions $\varpi^*(\boldsymbol{\theta}; d')$ in

196

a GLM is to simply add splitting terms to the parameters:

$$g\left(\varpi_k^*(\boldsymbol{\theta}; d')\right) = \sum_{i=0}^{M} \left(\theta_i + \frac{d'\Delta\theta_i}{2}\right) P_i(\varpi_k) \tag{G.12}$$

Even though d' is latent, these new asymmetry parameters can be maximized in the same likelihood maximization, since the probability π_d that the tag is correct depends only on the known quantities d and ϖ and the calibration parameters. Note that unless g is the identity link, $\Delta\varpi^*$ is not exactly equal to $d' \sum \Delta\theta_k P_k(\varpi)$. The modification is to the algebra in the above and below sections is straight-forward.

G.1.5   Fisher Information and Cramer-Rao bounds

The variance of a statistical estimator has a minimum limit, given by the Cramer-Rao bound [125] (as long as the estimator meets several regularity conditions). For a set of quantities $\boldsymbol{\theta}$, with a corresponding likelihood function $\mathcal{L}\left(\boldsymbol{\theta}; \vec{R}^{\text{obs}}\right)$, the Fisher information matrix of the parameters $\boldsymbol{\theta}$ is

$$i(\boldsymbol{\theta})_{ij} = E_{\boldsymbol{\theta}}\left[-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right]. \tag{G.13}$$

The Cramer-Rao theorem then states that the covariance matrix of an unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ that meets the necessary regularity conditions is bounded by

$$\text{cov}(\hat{\boldsymbol{\theta}}) \geq i(\boldsymbol{\theta})^{-1} \tag{G.14}$$

(where $\geq$ means that the matrix difference is positive semidefinite) and in particular that the variance of any individual parameter is bounded by

$$\text{var}(\hat{\theta}_i) \geq i(\boldsymbol{\theta})_{ii}^{-1}. \tag{G.15}$$

This can be applied directly to the problem of extracting calibration parameters from data. Ignoring the tagging asymmetry parameters, the Fisher information matrix of the log-likelihood function Eq. (G.7) for the set of calibration parameters $\boldsymbol{\theta}$ is

$$i(\boldsymbol{\theta})_{ij} = 4 \sum_{\text{events } e} \frac{P_i(\varpi_e)P_j(\varpi_e)}{g'\left(\varpi_e^*(\boldsymbol{\theta})\right)^2} \frac{(D_e^c)^2}{1 - (D_e^*(\boldsymbol{\theta})D_e^c)^2} \qquad (G.16)$$

When $D^c = 1$ for each event and there is no chance of misclassification, this matrix is especially simple, and reduces to the standard form for binomial regression:

$$i(\boldsymbol{\theta})_{ij} = \sum_k \frac{P_i(\varpi_k)P_j(\varpi_k)}{g'\left(\varpi_k^*(\boldsymbol{\theta})\right)^2} \frac{1}{\varpi_k^*(\boldsymbol{\theta})(1 - \varpi_k^*(\boldsymbol{\theta}))}. \qquad (G.17)$$

In the case of flavor tagging algorithms, an approximate set of values for the calibration parameters is typically known, for instance from a previous calibration to a simulation sample or an independent, separate dataset. The Fisher information matrix can be evaluated for this set of values, providing a useful estimate of the covariance matrix that will be obtained during calibration. When the off-diagonal terms of the Fisher information matrix are small, the Cramer-Rao bounds on the covariances between the parameters are suppressed. This estimated information matrix can be used to transform the set of basis functions to minimize the correlation between the calibration parameters during the MLE.

G.1.6 GLM link functions

Up to now no specification has been made for the link function $\eta = g(\varpi)$, other than that it is the inverse of a cdf $\varpi = h(\eta)$ of a probability distribution over the real axis. However, the link function implicit in previous LHCb flavor tagging calibration is the identity $g(\varpi) = \varpi$. This link has the benefit of being very simple, but care must be taken that $\eta(\boldsymbol{\theta})$ does not stray outside the range $[0, 1]$ — otherwise there will be numerical issues in the likelihood maximization. In this case $g' = 1$,

making the Fisher information matrix (when there is no misclassification)

$$i(\boldsymbol{\theta})_{ij} = \sum_k \frac{P_i(\varpi_k)P_j(\varpi_k)}{\varpi_k^*(\boldsymbol{\theta})(1 - \varpi_k^*(\boldsymbol{\theta}))}. \tag{G.18}$$

The most widespread link function in the statistical literature is the logistic link, whose link function is variously called the logit or log-odds function:

$$g(\varpi) = \text{logit}\,\varpi = \log(\varpi) - \log(1 - \varpi) \tag{G.19}$$

This has an appealing interpretation as the logarithm of the odds ratio $\varpi/(1 - \varpi)$. The inverse of the link is the logistic or sigmoid function:

$$h(\eta) = \frac{1}{1 + e^{-\eta}} \tag{G.20}$$

The derivative of the link $g'(\varpi) = 1/\varpi(1 - \varpi)$ takes the same simple form as other terms in the likelihood function and its derivatives. Consequently the Fisher information matrix is equally simple for the logit link:

$$i(\boldsymbol{\theta})_{ij} = \sum_k \varpi_k^*(\boldsymbol{\theta})(1 - \varpi_k^*(\boldsymbol{\theta}))P_i(\varpi_k)P_j(\varpi_k). \tag{G.21}$$

Other link functions are less commonly used. These include the probit link, which is the inverse of the cdf of the standard normal distribution, and the cauchit link, which is the inverse of the cdf of the Cauchy distribution. It is possible that these alternative links might fit a certain dataset better; however, their analytic expressions are messier than the two standard links detailed above.

G.1.7 Binomial regression goodness-of-fit tests

Current methods of flavor tagging calibration do not have well studied goodness-of-fit tests. In the binned procedure for calibrations to $B^\pm$ decays, the reduced chi-squared statistic is often used as a goodness-of-fit test, but like the result of regression this test depends on the arbitrarily chosen binning scheme. The time-dependent likelihood maximization procedure has related goodness-of-fit tests, but these are sensitive to other sources of disagreement in this larger likelihood function, such as the modeling of the lifetime acceptance efficiency.

Binomial regression, on the other hand, has well studied goodness-of-fit tests, though none is considered is canonical [115, 116]. The binned Hosmer-Lemeshow decile of risk tests are popular, though they as well depend on an arbitrary binning procedure. A number of tests appropriate for continuous data are described in [126] and its references.

G.2 Physical and experimental effects in the regression framework

In the mathematical formalism above, the quantity ϖ^c represents the probability that the observed decay flavor d of the b-hadron is the same as its production flavor d'. This probability is only less than unity when the b-hadron is a B^0 or B_s^0 meson and oscillates. Then, several effects contribute to the value of ϖ^c: the oscillation of the neutral meson, the decay time resolution of the LHCb detector for the given mode, and possible b–$\bar{b}$ production asymmetries, which are constrained to be small but are likely not nonzero.

G.2.1 Oscillation probability of a neutral B meson

To start with, the probability that a B with production flavor d' will be decay at time t with flavor $d'' = -d'$ must be determined. The initial state of such a B

meson is

$$|\psi(0)\rangle = |d'\rangle = \frac{|B_H\rangle + d'|B_L\rangle}{\sqrt{2}} \tag{G.22}$$

The evolution the B_H and B_L states are specified by the mean decay width Γ and mass m and mean decay width difference $\Delta\Gamma$ and mass difference Δm, and are

$$e^{-iHt}|B_H\rangle = e^{-\frac{1}{2}\left(\Gamma + \frac{1}{2}\Delta\Gamma\right)t + i\left(m + \frac{i}{2}\Delta m\right)t}$$

$$e^{-iHt}|B_L\rangle = e^{-\frac{1}{2}\left(\Gamma - \frac{1}{2}\Delta\Gamma\right)t + i\left(m - \frac{i}{2}\Delta m\right)t}$$

Therefore, the state evolves into

$$|\psi(t)\rangle = e^{-\frac{1}{2}\Gamma t + imt}\left[\cos\left(\frac{\Delta mt + \frac{1}{2}i\Delta\Gamma t}{2}\right)|d'\rangle \right.$$
$$\left. + i\sin\left(\frac{\Delta mt + \frac{1}{2}i\Delta\Gamma t}{2}\right)|-d'\rangle\right] \tag{G.23}$$

Discounting the overall decay probability $\exp(-\Gamma t)$, the probability that a B meson that decays at time t will not oscillate into a different flavor is

$$P_{d''=d'}(t) \equiv |\langle d'|\psi(t)\rangle|^2 = \frac{1}{2}\left(1 + \cos\Delta mt \operatorname{sech}\frac{1}{2}\Delta\Gamma t\right). \tag{G.24}$$

This probability is independent of d', as it should be, and from here on this d' dependence will be dropped. The corresponding dilution due to oscillation is

$$D_{\mathrm{osc}}(t) = 2P_{d''=d'}(t) - 1 = \cos\Delta mt \operatorname{sech}\frac{1}{2}\Delta\Gamma t. \tag{G.25}$$

It might be naively expected that $\varpi^c(t) = P_{d''=d'}(t)$, and this is close to the truth. In particular, ϖ^c is the probability that that B^0 or B_s^0 meson observed with decay flavor d'' was really produced with flavor $d' = d$ — symbolically $P(d' \mid d)$. The quantum mechanical probability $P_{d''=d'}(t)$ is $P(d'' \mid d')$, and a straightforward application of

Bayes' theorem shows that

$$
\begin{aligned}
\varpi^c(t) &\equiv P\left(d'', 0 \mid d'', t\right) \\
&= \frac{P\left(d'', t \mid d'', 0\right) P(d'', 0)}{P(d'', t)} \\
&= \frac{P_{d''=d'}(t)(1/2)}{(1/2)} \\
&= P_{d''=d'}(t) \tag{G.26}
\end{aligned}
$$

This confirms the naive expectation, and the total tagging dilution $D^c(t) = 2\varpi^c(t) - 1$ is equal to $D_{\mathrm{osc}}(t)$. In reality, however, experimental effects from decay time resolution and production flavor asymmetries modify this simple result.

G.2.2 Decay time resolution effects

The finite decay time resolution at a particle detector can lead to extra dilution. Its effect is suppressed in the limit where the resolution δt is $\ll 1/\Delta m$; however, in the opposite limit where $\delta t \gg 1/\Delta m$, the ability to tag the oscillation correctly is completely washed out by resolution effects. In practice, B^0 decays belong to the former case, while B_s decays are intermediate — decay time resolution is an important effect, but does not dominate.

In general, the resolution function $\mathcal{R}(t, t')$, which usually is equivalent to a function $\mathcal{R}(t - t')$, modifies the observed decay time distribution of the B mesons. Because the true distribution is not uniform in the region around the observed time (being biased towards shorter decay times), the effect of the resolution function cannot be exactly disentangled from the shape of the decay time distribution. However, in practice, the width of the resolution function is much smaller than the timescales over which the decay time distribution varies appreciably, with the exception of decay times near zero where the distribution is discontinuous. So, to a

good approximation an *a priori* uniform decay time distribution can be assumed in examining the effect of the decay time resolution.

In this simplification, the resolution function $\mathcal{R}(t, t')$ modifies the oscillation probability via the convolution

$$\omega^c(t) = \int P_{d''=d'}(t)\mathcal{R}(t, t')\, dt' \tag{G.27}$$

Because convolution is a linear operation, this is equivalent to a convolution of $D_{\mathrm{osc}}(t)$. In the simple case with no production asymmetry, this dilution is a product of a cosine term and a hyperbolic secant term (Eq. (G.25)), but the latter varies quite slowly within the decay time resolution ($\Delta\Gamma\sigma \ll 1$, typically). When the resolution function is a single gaussian, the only contribution to the convolution is

$$\frac{1}{\sigma\sqrt{2\pi}} \int \cos\left(\Delta m t'\right) \exp\left(-\frac{(t - t')^2}{2\sigma^2}\right) dt' = \cos\left(\Delta m t\right) e^{-\frac{1}{2}\Delta m^2\sigma^2}$$

Thus there is an effective dilution

$$D_{\mathrm{res}} = \exp\left(-\frac{1}{2}\Delta m^2\sigma^2\right). \tag{G.28}$$

due to the finite decay time resolution. For a more general resolution function $\mathcal{R}(t, t')$, the effective dilution is the real part of the Fourier transform of G with frequency Δm with respect to t'. The total observed dilution due to oscillation is therefore

$$D^c(t) = 2\varpi^c(t) - 1 = D_{\mathrm{res}}D_{\mathrm{osc}}(t). \tag{G.29}$$

A more exact treatment must account for the slight non-uniformity of the decay time distribution over the support of the resolution function. The observed decay time distribution is given by the a priori decay time distribution $p(t)$ (including

acceptance effects) convolved with the resolution function:

$$p_{\text{obs}}(t_{\text{obs}}) = \int p(t) \mathcal{R}(t_{\text{obs}}, t) \, dt. \tag{G.30}$$

The resolution function is just the conditional probability distribution, since

$$p(t_{\text{obs}} \mid t) = \mathcal{R}(t_{\text{obs}}, t) \tag{G.31}$$

Typically the resolution function $\mathcal{R}(t_{\text{obs}}, t) = \mathcal{R}(t_{\text{obs}} - t)$, making the observed time distribution simply the convolution $p_{\text{obs}} = \mathcal{R} * p$. By Bayes' theorem,

$$p(t \mid t_{\text{obs}}) = \frac{\mathcal{R}(t_{\text{obs}}, t) p(t)}{p_{\text{obs}}(t_{\text{obs}})} \tag{G.32}$$

If the decay time distribution $p(t)$ varied on a time scale much greater than the width of the resolution function, the approximation made previously, then it would cancel out in this fraction. No longer neglecting this effect, the expectation value of a function $f(t)$ of the true decay time is

$$E\left[f(t) \mid t_{\text{obs}}\right] = \int f(t) p(t \mid t_{\text{obs}}) \, dt = \frac{1}{p_{\text{obs}}(t_{\text{obs}})} \int f(t) p(t) \mathcal{R}(t_{\text{obs}}, t) \, dt \tag{G.33}$$

Less compactly,

$$E\left[f(t) \mid t_{\text{obs}}\right] = \frac{\int f(t) p(t) \mathcal{R}(t_{\text{obs}}, t) \, dt}{\int p(t) \mathcal{R}(t_{\text{obs}}, t) \, dt} \tag{G.34}$$

There are two possible effects to consider. The first is the possibility that $p(t)$ does appreciably vary over the support of resolution function. The second is that the range of t is restricted to positive values. Assuming an exponential decay time distribution $p(t) = \Gamma \exp(-\Gamma t)$ with no acceptance effects and a gaussian resolution function,

$$p_{\text{obs}}(t_{\text{obs}}) = \Gamma e^{-\Gamma t_{\text{obs}}} \times \Phi\left(\frac{t_{\text{obs}}}{\sigma} - \Gamma \sigma\right) e^{\frac{1}{2}\Gamma^2 \sigma^2}, \tag{G.35}$$

where $\Phi(x) = (1 + \mathrm{erf}(x/\sqrt{2}))/2$ is the cdf of the standard normal distribution. The sharp discontinuity of the distribution at $t = 0$ is rounded off and distribution extends into negative observed times. There are separate contributions to this expression from the restriction to positive values of t in the integral, the exponential shape of the decay time distribution, and the interaction of these two properties. The first can be identified in the limit $\Gamma \to 0$ and the second in the limit $\sigma \to 0$. The effect of the interaction is present in the form of the term $\Gamma\sigma$, which shifts the center of the cdf.

On closer examination, this correction to the decay time distribution is only relevant for observed times t_{obs} within several resolution widths σ of 0. Not only is this precisely where acceptance effects modify the posterior decay time distribution, selections often veto candidates below a minimum decay time equal to many multiples of the resolution width. Even with perfect acceptance and a selection that includes candidates at small decay times, much of the effect on the observed decay time distribution cancels out in $E[f(t) \mid t_{\mathrm{obs}}]$. The quantity of most interest is $D^c(t)$, which (ignoring $\Delta\Gamma$) is given by the expectation of $f(t) = \cos \Delta m t$. In this case, the modification to the native convolution $f * \mathcal{R}$ is quite negligible for positive observed times. The exact result is

$$D^c(t) = E\left[\cos(\Delta m t) \mid t_{\mathrm{obs}}\right] = (\Re(\mathcal{X})\cos(\mathcal{Y}) - \Im(\mathcal{X})\sin(\mathcal{Y})) \times e^{-\frac{1}{2}\Delta m^2 \sigma^2} \quad \text{(G.36)}$$

where

$$\mathcal{X} = \frac{\Phi(t/\sigma - \Gamma\sigma + i\Delta m\sigma)}{\Phi(t/\sigma - \Gamma\sigma)}, \quad \mathcal{Y} = \Delta m t - \Gamma\Delta m\sigma^2. \quad \text{(G.37)}$$

The exponential term simply reproduces the effective dilution due to resolution D_{res} derived previously.

Both $\Gamma\sigma$ (~ 0.03) and $\Delta m\sigma$ (~ 0.025) are small for B^0 decays. In particular, the exponential factor $\exp(-\Delta m^2 \sigma^2/2)$ is indistinguishable from unity, and the

(a) B^0 (b) B_s
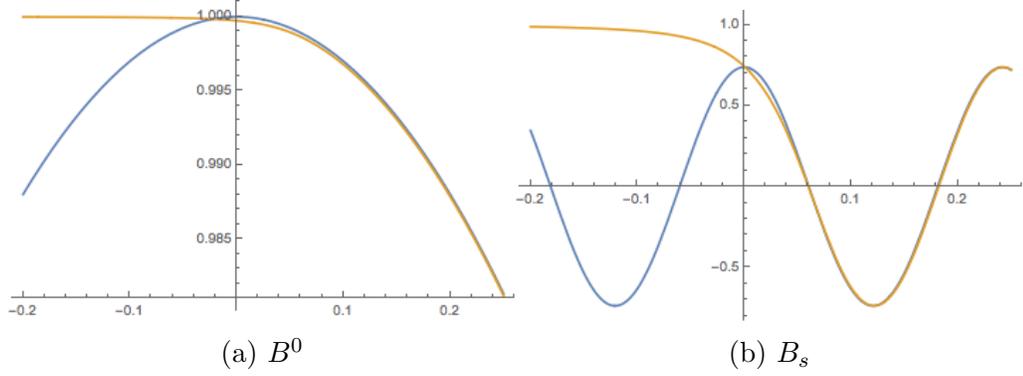
Figure G.1: A plot of the approximation of $D_{\mathrm{res}} \cos \Delta m t$ from Eq. (G.29) (blue) is compared to the analytic result $E\left[\cos(\Delta m t) \mid t_{\mathrm{obs}}\right]$ from Eq. (G.36) (orange) for B^0 mesons (left) and B_s mesons (right). For each plot, the scale of the time axis is chosen such that $\Gamma = 1$ and $\sigma = 0.03$ (corresponding to 50 fs decay time resolution), while $\Delta m / \Gamma = 0.775$ for B^0 and 26 for B_s.

phase shift $\Gamma \Delta m \sigma^2$ is less than 0.001. Consequently, Eq. (G.36) is essentially equal to $\cos \Delta m t$, and the decay time distribution has a negligible effect in calibrations to B^0 decay modes and can be ignored. In B^0_s decays, on the other hand, $\Gamma \sigma$ is again about 0.03 but $\Delta m \sigma$ is quite significant (~ 0.9). The phase shift $\Gamma \Delta m \sigma^2 \sim 0.03$ is still fairly small, and to a good approximation can be ignored. Fig. G.1 shows plots of the exact value of Eqs. (G.29) and (G.36) for both B^0 and B_s decays. The only non-negligible effect of the exact analysis is that $E\left[\cos(\Delta m t) \mid t_{\mathrm{obs}}\right]$ quickly approaches 1 for small and negative values of t_{obs} (a kinematic region with typically very low detector acceptance efficiency), rather than continuing to oscillate.

G.2.3 Production asymmetries

An initial flavor asymmetry $\mathcal{A}_0$ between B^0 (B^0_s) and $\overline{B}^0$ ($\overline{B}^0_s$) mesons at production evolves into an asymmetry

$$\mathcal{A}(t) = \mathcal{A}_0 D_{\mathrm{osc}}(t) \tag{G.38}$$

for mesons decaying at time t. This changes the *a priori* probability for each flavor at production and decay. Repeating the Bayesian logic above which demonstrated that $\varpi^c(t) = P_{d''=d'}(t)$ with this new prior information instead shows that

$$
\begin{aligned}
\varpi^c(t) &\equiv P\left(d'', 0 \mid d'', t\right) \\
&= \frac{P\left(d'', t \mid d'', 0\right) P(d'', 0)}{P(d'', t)} \\
&= P_{d''=d'}(t) \times \frac{(1 + d'' \mathcal{A}_0)/2}{(1 + d'' \mathcal{A}(t))/2} \\
&= P_{d''=d'}(t) \times \frac{1 + d'' \mathcal{A}_0}{1 + d'' \mathcal{A}_0 D_{\mathrm{osc}}(t)}
\end{aligned}
\tag{G.39}
$$

As a sanity check, consider how this probability behaves in limiting cases of the production flavor asymmetry. Without loss of generality, take the flavor at decay to be positive. When $\mathcal{A}_0 = 0$ this reduces to Eq. (G.26) as it should. It becomes 0 when $\mathcal{A}_0 = -1$, since a positive flavor at decay is always oscillated if the flavor is always negative at production. Finally, it becomes 1 when $\mathcal{A}_0 = 1$, since a positive flavor at decay is always correct if the flavor is always positive at production.

Calculating the effective probability that the B candidate has oscillated requires taking the convolution of Eq. (G.39), which would be analytically difficult. It is much more practical, since the production asymmetry $\mathcal{A}_0$ is certainly quite small, to approximate $\omega(d', t)$ to first order in the asymmetry,

$$
\varpi^c_{d'}(t) = P_{d''=d'}(t) + \frac{1}{2} d' \mathcal{A}_0 \left(1 - D_{\mathrm{osc}}(t)^2\right) + \mathcal{O}\left(\mathcal{A}_0^2\right),
\tag{G.40}
$$

introducing a slight dependence of $\varpi^c(t)$ on the production flavor d'. The correction term is proportional to $1 - D_{\mathrm{osc}}(t)^2$, whose only non-negligible time dependence comes through the term $\cos^2 \Delta m t$, which has a simple convolution

$$
\int dt' \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(t - t')^2}{2\sigma^2}\right) \cos^2\left(\Delta m t'\right) = \frac{1}{2}\left(1 + \cos\left(2\Delta m t\right) e^{-2\Delta m^2 \sigma^2}\right).
$$
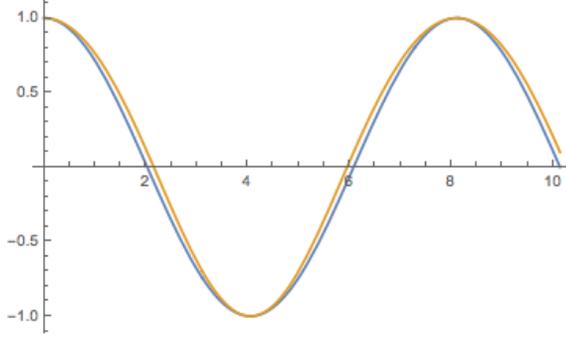
Figure G.2: A plot of the approximation Eq. (G.29) (blue) is compared to Eq. (G.41) (orange) for B^0 oscillation, when $\mathcal{A}_0 = 20\%$ and the observed flavor is B^0. When the observed flavor is $\bar{B}^0$ the sign of the correction flips.

When the secant term in $P_{\text{osc}}(t)$ can be ignored for simplicity (for B^0 decays), the total dilution becomes

$$D^c_{d'}(t) = D_{\text{res}} D_{\text{osc}}(t) + \frac{1}{4} d' \mathcal{A}_0 \left(1 - D^4_{\text{res}} \cos \left(2 \Delta m t \right) \right) + \mathcal{O}\left(\mathcal{A}_0^2 \right). \tag{G.41}$$

Figure G.2 compares Eq. (G.29) with the approximation Eq. (G.29), for a B^0 decay mode with a B^0 production asymmetry to be 20% (much larger than experimental bounds). Evidently the approximate result is still accurate to about 5% even in this extreme case; in practice production asymmetries are constrained to be on the order of 1% or less, and should have negligible effect. Detection asymmetries, on the other hand, can be much larger, but since they are a function of the decay flavor and not the production flavor they have no effect on the *a priori* flavor probabilities and thus do not affect $D^c(t)$.

G.3 Implementation in the `EspressoPerformanceMonitor`

Several popular tools are freely available for performing binomial regression. The R programming language, for example, contains the tools `glm` and `lrm`, which can perform binomial regression using logit, probit, and cauchit link functions (among others). The `scikit-learn` package in Python, meanwhile, has an im-

plementation of logistic regression. However, for a number of reasons, neither of these tools is adequate for use in flavor tagging. Neither implementation provides a mechanism to handle misclassification of response, necessary for calibrating to neutral B^0 and B_s^0 decay modes, and incorporating this requires large modifications to the likelihood function being maximized. Calculation of the tagging asymmetry parameters also requires small modifications to the likelihood function. The data used in flavor tagging calibration is primarily stored in the ROOT binary data format, which has native interfaces in C++ and Python but not R, making use of the superior `glm` difficult without using an intermediate storage format. Finally, the datasets used to calibrate flavor tagging algorithms usually require weights to separate signal from background [85], and these have the unusual property that they are sometimes negative; since this is usually ill-defined generic regression implementations like `glm` require weights to be positive-definite.

A custom implementation of binomial regression, on the other hand, can read flavor tagging data from its native ROOT file format, automatically account for neutral meson oscillation and tagging asymmetry, and make use of other optimizations pertinent to flavor tagging. The `EspressoPerformanceMonitor` tool is written in C++ and hosted on LHCb's GitLab code repository. The tool uses the ROOT libraries to read data stored in ROOT format [127]. Binomial regression is implemented using a Newton-Raphson like algorithm provided by `GSL` [128], or, alternatively, using the `Minuit` numerical maximization algorithm via ROOT [129]. The tool performs calibrations using one of a number of GLM models (polynomials, b-splines, and n-splines) and link functions (identity, logit, probit, or cauchit). It reports several goodness of fit metrics, including the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) [116, 124] and scores corresponding to the Pearson X^2 and le Cessie–van Houwelingen–Copas–Hosmer S test statistics [115, 126]. The tool also produces a number of publication quality plots using

ROOT's graphics libraries.

G.3.1   Special GLM models for flavor tagging

The `EspressoPerformanceMonitor` implements three classes of GLM models: polynomial models, b-spline models, and n-spline models. There are several unique features of these classes that are tailored to the specific requirements of calibrating flavor tagging algorithms.

Spline models

Two popular types of GLM spline bases are basis-splines, or b-splines, and natural splines, or n-splines [124]. All splines are piecewise functions defined between two boundary knots with several interior knots at which they are only C^2 continuous. In short, b-splines have no boundary constraints and are similar to bump functions, defined in a way that limits the overlap of their supports, while n-splines are similar to polynomials and are constrained to be linear at their boundaries knots. The `EspressoPerformanceMonitor` chooses the knot locations for the b-spline and n-spline models are via quantiles of the η distribution.

When extrapolating the calibration to values outside of the boundary knots (a rare occurence), different choices must be made for b-splines and n-splines. Because a b-spline is discontinuous at the boundary knots, no extrapolation can be made beyond the boundary, and any value of η outside the boundary knots is pulled back to the nearby boundary knot. For n-splines, constrained to be linear at the boundaries, the linearly extrapolated value is taken.

η-subtracted GLMs

Typically LHCb studies have dealt with the "mistag probability" of a tagger, which is the probability that its tag decision is incorrect. The uncalibrated mistag probability is denoted $\eta = 1 - \varpi$, while the calibrated mistag probability is $\omega = 1 - \varpi^*$. The probability of misclassification due to B–$\bar{B}$ oscillation, $1 - \omega^c$, is $\sin^2(\Delta mt/2)$, as described above, making $\delta^c = \cos(\Delta mt)$. The real mistag probability ω tends to be approximately equal to the uncalibrated value η produced by flavor tagging algorithms, typically because the flavor tagging algorithm is given a default calibration using some Monte Carlo or actual dataset. This is quite unlike the typical scenario in the literature of GLMs, and can be taken advantage of by using an "η-subtracted" GLM in which

$$g(\omega) = g(\eta) + \sum_k \theta_k P_k(\eta). \tag{G.42}$$

There are several advantages to this modification. First, if the tagger is already perfectly calibrated, all the coefficients θ should be 0, and even if it is slightly miscalibrated they should at least be close to 0. This provides a uniform starting point for likelihood maximization, no matter what model or link function is used. It also makes the coefficients themselves more meaningful — any non-zero coefficient is a sign of mis-calibration. This is especially useful when the basis functions $P_k(\eta)$ are splines, in which case (unlike for simple polynomials) the linear combination corresponding to $\omega = \eta$ is not obvious.

The fact that $\omega(\eta) \approx \eta$ also has implications for polynomial and n-spline regression models when a non-identity link is used. Rather than using polynomials $P_k(\eta) = \eta^k$, the `EspressoPerformanceMonitor` uses polynomials $P_k(\eta) = g(\eta)^k$, so

that the calibration model is (before the decorrelation described below)

$$g(\omega) = g(\eta) + \sum_i \theta_k g(\eta)^k. \tag{G.43}$$

This make the model polynomial in the transformed space $g(\eta)$; in particular the linear term proportional to θ_1 is indeed a slope correction to the default term $g(\eta)$. Much the same is true for n-spline models, which the `EspressoPerformanceMonitor` creates in the space $x = g(\eta)$, with knots at values $g(\eta_q)$ for quantiles η_q of the predicted mistag distribution. This way, the first two n-spline basis functions are identical to the first two polynomials, and he term θ_1 is again a slope correction. Moreover, beyond the boundary knots the n-spline model has the form $g(\omega) = a + bg(\eta)$. So, the tails of the calibration function are linear in the transformed space $(g(\eta), g(\omega))$.

Orthogonal GLM bases

With the good initial guess that the calibration values $\boldsymbol{\theta}$ are approximately zero, an estimate of the Fisher information matrix can be computed and an orthogonal basis can be chosen in which the calibration parameters are decorrelated. This is particularly applicable to the polynomial and n-spline models, both of which feature a progression of increasingly higher curvature functions, beginning with 1 and $g(\eta)$. Because the supports of the basis functions overlap greatly, the calibration parameters can be very highly correlated. Instead, a basis in which Eq. (G.16) is diagonalized can be chosen via the Gram-Schmidt process. This can be done exactly or, for convenience, by making some approximation to the information matrix.

One approximation that is made is ignoring the misclassification of flavor due to B–$\bar{B}$ oscillation and diagonalizing Eq. (G.17). Flavor tagging algorithms generally produce a mistag probability η whose distribution depends at most slightly on

the species of B meson and decay mode being observed. The oscillation probability, however, depends entirely on the species of B meson. It is convenient to use a basis that is more-or-less independent of the B species, up to statistical fluctuations. Since calibrations to charged B decay modes are simpler, more common, and more precise, calibrations to neutral B mesons should share the same basis.

Another approximation is made when using the identity link function to preserve backwards compatibility with older models used in the calibration of LHCb's flavor tagging algorithms. Since flavor tagging algorithms have a fairly high mistag rate η, $D = 1 - 2\eta$ is typically small. Moreover, η and the oscillation probability (related to the decay time) are typically uncorrelated. So, the Fisher information matrix is approximately

$$i(\boldsymbol{\theta})_{ij} \approx 4N \left\langle P_i(\varpi)P_j(\varpi) \right\rangle \left\langle (D^c)^2 \right\rangle. \tag{G.44}$$

and thus the Fisher information matrix is approximately diagonalized by the orthogonal polynomials $\left\langle P_i(\varpi)P_j(\varpi) \right\rangle \propto \delta_{ij}$. The first two such polynomials are $P_0(\eta) = 1$ and $P_1(\eta) = \eta - \langle\eta\rangle$. These are the basis polynomial historically used in LHCb's calibrations and analyses, and $\omega(\eta) = \eta + \theta_0 + \theta_1(\eta - \langle\eta\rangle)$. Calibrations are often presented using the parameterization $\omega(\eta) = p_0 + p_1(\eta - \langle\eta\rangle)$, which is related via $\theta_0 = \delta p_0$ and $\theta_1 = p_1$ The single non-unity coefficient is $\langle\eta\rangle$, which is simple to calculate and interpret.

With these approximations applied, an inner product on the basis functions is defined by $\langle P_i, P_j \rangle \equiv i(\theta_0)_{ij}$. This inner product extends naturally to linear combinations $Q_k = \sum q_{k\ell} P_\ell$. Starting with initial values $Q_k = P_k$, i.e., $q_{k\ell} = \delta_{k\ell}$, the stabilized Gram-Schmidt procedure iteratively updates the values $q_{k\ell}$ until $\langle Q_k Q_\ell \rangle = \delta_{k\ell}$. The only nonzero coefficients $q_{k\ell}$ are for $\ell \leq k$, so that Q_k is a linear combination of $P_0, \ldots, P_k$. The orthonormality is not needed for flavor tagging and

can be discarded by dividing Q_k by q_{kk}. When $P_0(x) = 1$ and $P_1(x) = x$, the first two basis functions created this procedure are always $Q_0(x) = 1$ and $Q_1(x) = x - x_0$.

G.3.2 Goodness-of-fit tests

There is less consensus in the literature on the best goodness-of-fit tests for logistic regression than there is for linear regression. One metric for comparing separate models obtained via likelihood maximization is to compare the maximum likelihoods that they achieve. This is typically expressed in terms of the deviance $G^2 = -2\log(\mathcal{L})$, which smaller for a model with a greater likelihood. Introducing more parameters into a model can always improve the deviance, however, and so the Akaike information criterion (AIC) penalizes the deviance according to the number of parameters in the model [116]:

$$AIC = G^2 + 2k = -2\left(\log(\mathcal{L}) - k\right). \tag{G.45}$$

An alternative to the AIC is the Bayesian information criterion (BIC), defined by

$$BIC = G^2 + k\log(N) = -2\log\left(\frac{\mathcal{L}}{\sqrt{N}^k}\right). \tag{G.46}$$

For large samples, this criterion more strongly punishes complex models. Assuming a uniform prior probability for all calibration models, the BIC is related to the posterior probability of the model: given two BIC values BIC_1 and BIC_2, the log-odds of the two models is proportional to $BIC_1 - BIC_2$, motivating the choice of the model with the minimum BIC [124]. The BIC is asymptotically consistent, picking the correct model in the infinite statistics limit. Studies have shown that the AIC selection criterion tends to select too complex models, while the BIC criterion sometimes picks too simple models. These information criteria can be useful for comparing separate models, but offer little guidance for the intrinsic quality of a

single model.

The significance of extra parameters in nested models provides a similar comparison test: a model with a parameter θ provides a better fit than a smaller model with this parameter fixed to θ_0 if the estimated value $\hat{\theta}$ is significantly different from θ_0 , *i.e.*, $(\hat{\theta} - \theta_0)^2/\hat{\sigma}^2$ is past some accepted threshold[1] [116]. Concretely, if one or several of the parameters in the fit model are not statistically distinct from 0, the fit model might be too complex.

Goodness-of-fit tests that measure the intrinsic quality of models rely on examining the distribution of residuals. Three types of residuals are commonly used for binomial regression: Pearson-like residuals, deviance residuals, and simple unweighted residuals [126]. These residuals are functions of the predicted probability for the tag to be correct (a quantity that incorporates the predicted mistag, calibration parameters and delta parameters, and B oscillation)

$$\pi^{(*)} = \omega^{(*)} \leftrightarrow \omega^c = \frac{1}{2}\left(1 + \delta^{(*)}\delta^c\right) \tag{G.47}$$

and also the boolean R^{obs}, which is 1 (0) if the tag is correct (incorrect) at decay time.

The deviance residual is based on the actual likelihood function for the data given a model, and is theoretically appealing —but has some numerical problems. The residual for a particular tag is defined by:

$$r_{G^2} = \sqrt{-2\left(R^{\text{obs}}\log(\pi) + (1 - R^{\text{obs}})\log(1 - \pi)\right)} \tag{G.48}$$

Not surprisingly, the deviance residual is defined such that the sum in quadrature

---

[1]This is known as the Wald test. There are two common alternatives. The likelihood ratio test examines the decrease in deviance ΔG^2, which should follow a χ^2 distribution with a degree freedom for each extra parameter. The score test, based entirely on the likelihood function at θ_0, is a third test for the same situation. All three tests are asymptotically equivalent [116].

of the deviance residuals is simply the deviance G^2.

The Pearson residual is the direct analog of the standard residuals for a least-squares regression:

$$r_{X^2} = \frac{R^{\text{obs}} - \pi}{\sqrt{\pi(1-\pi)}} \tag{G.49}$$

The denominator is the standard deviation of a binomial distribution with probability π. The Pearson residual has a few properties typically associated with residuals, *e.g.*, that the average value of the residuals is 0 and their variance is 1. Of course, the distribution is not at all normal, with separate clusters below 0 and 1 for correctly and incorrectly tagged events. The sum in quadrature of the squared Pearson residuals is known as the Pearson X^2.

The unweighted residual is an unweighted version of the Pearson residual:

$$r_S = R^{\text{obs}} - \pi \tag{G.50}$$

On the surface, there is not much theoretical justification for this residual; however, it is quite popular. The sum in quadrature of these residuals is sometimes called the le Cessie-van Houwelingen-Copas-Hosmer statistic S [126]. The mean-square of the residuals S/N is known as the mean squared error (MSE) or, especially in meteorological and machine learning contexts, the Brier score [130]. If the model is perfectly specified, the expected value of the score is

$$E\left[MSE\right] = \frac{1}{N}\sum E\left[(R^{\text{obs}} - \pi)^2\right] = \frac{1}{N}\sum \pi\left(1-\pi\right) = \frac{1}{4}\left(1 - \langle D^2 \rangle\right) \tag{G.51}$$

Thus the MSE for a correctly specified model is directly related to its tagging power.

There are multiple methods of turning residuals into a test statistics. These fall into two major categories: tests which divide the data into groups and examine the agreement between the model and the average outcome in each, and those

which examine the agreement between the model and the outcomes at a per-event level. The former, the canonical example of which is the Hosmer-Lemeshow procedure [115, 126], are simpler and generally produce statistics that follow a χ^2 distribution. However, they have some interpretational issues that make them ill suited for applying to the calibration of flavor tagging algorithms. For one, their results depend on the number of groups that the data is binned into. Moreover, they have only been studied in the context of simple linear fit models where the number of bins g (typically 10) is much greater than the number of degrees of freedom in the fit (2), in which case the test statistic asymptotically follows a χ^2 distribution with $g-2$ degrees of freedom. For polynomial or spline calibrations, where the number of degrees of freedom might approach 4 or 5 (or double this if the tagging asymmetry parameters are counted), it is unclear what the asymptotic distribution of these test statistics would be.

Instead, the `EspressoPerformanceMonitor` applies unbinned tests based on the sums of residuals G^2, X^2, and S. All three quantities belong to the family of power-divergence statistics [116, 131].[2] These statistics were developed primarily with discrete data in mind, in which case the metrics follow χ^2 distributions. When the data is continuous, as is the case for flavor tagging via the predicted mistag η, the power divergence statistics asymptotically follow normal distributions $N(\mu, \sigma)$, and can be rescaled to get a test statistics following a standard normal distributions. Their means and variances depend on whether the test is non-parametric, applied to a data sample where the tagging algorithm is assumed to be well calibrated already and no fit is performed, or parametric, performed after the tagging algorithm has been calibrated on the dataset (in which case the agreement should be better than random). A procedure for calculating these values is established in [132] for X^2 and G^2 and in [133] for the entire family of power divergence statistics. In

---

[2]The Neyman modified X^2 metric, defined for a general multinomial distribution, reduces to the S metric in the binomial case.

the non-parametric case, there are a fairly simple expressions for the means and variances, while in parametric case each variance must be reduced by a quadratic form involving the information matrix of the log-likelihood maximization and the covariance between the metric and the score vector (the first derivative of the log-likelihood).

The `EspressoPerformanceMonitor` calculates the AIC, BIC, and Brier score, and implements the ungrouped tests corresponding to the G^2, X^2 and S statistics (as well as the Cressie-Read test, a compromise between G^2 and X^2). In certain contexts, the G^2 and X^2 tests do not work — in particular, the former does not work when using the logit link and the latter when using the identity link. In these cases, the tool prints a message stating that the test cannot be computed.

One final note is that the `EspressoPerformanceMonitor` uses weights generated by the *sPlot* technique [85], when calculating these tests. They are added as multiplicative weights in any sum over all events, and where possible mean values and their standard errors are replaced with weighted means and their standard errors. There is no guarantee that this procedure is exactly correct, and it likely is not given the negative weights produced by the *sPlot* technique. Pseudoexperiments indicate that moderate background levels (up to 10% to 20%) do not significantly affect the goodness-of-fit test scores.

G.3.3   Example output

Complete documentation for the `EspressoPerformanceMonitor` is provided on its GitLab page. Here is a brief portfolio of output from the tool from various calibrations applied to the OS Combination tagger in a very large $B^{\pm} \rightarrow J/\psi K^{\pm}$ dataset.

The simplest type of calibration that can be performed is a linear calibration with the identity link function, the same model used in previous LHCb calibrations.
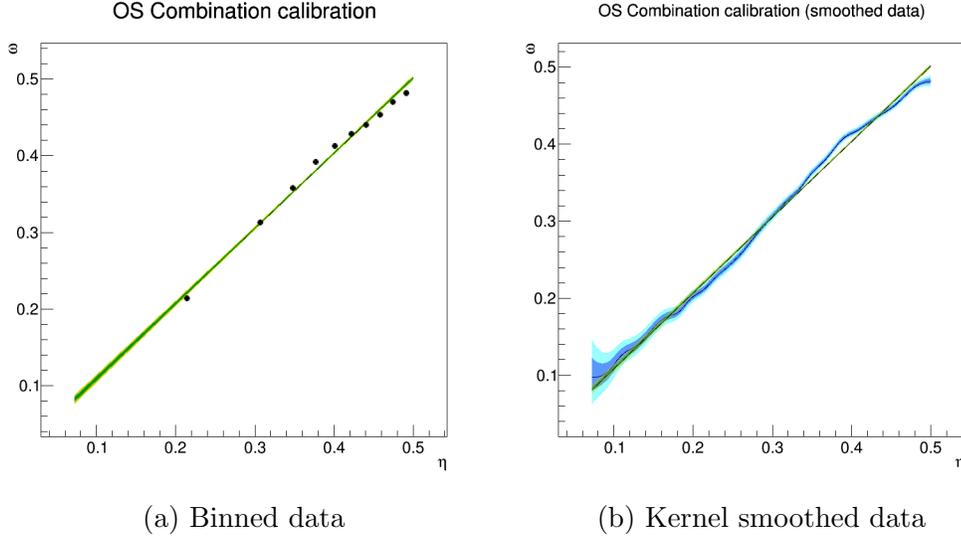
(a) Binned data   (b) Kernel smoothed data

Figure G.3: Linear calibration to $B^\pm \to J/\psi\, K^\pm$; the green and yellow areas represent 1σ and 2σ bands for the calibration, while the blue and light blue areas in the right-hand plot represent 1σ and 2σ confidence intervals in the kernel-smoothed data.

Plots of the calibration are shown in Fig. G.3. Since the outcome variables in logistic regression are binary (0 or 1), plain scatter plots are not a good representation of the fit. Instead, the `EspressoPerformanceMonitor` produces two kinds of plots. In the first kind, the data is binned into deciles by the predicted mistag η, and in each bin the average η and the fraction ω of incorrectly tagged candidates are calculated. When the calibration mode is a B^0 or B^0_s decay, calculating the latter requires factoring effects from misclassification due to oscillation. In the second kind, a gaussian kernel smoother is used to create a continuous representation of the data: instead of deciles, a value of ω is calculated for each η_0 by considering all candidates with η close to η_0, weighting each according to the distance $\|\eta - \eta_0\|$. These two plots for the linear calibration are shown in Fig. G.3.

It is visually quite apparent that the linear calibration is insufficient to describe the data. The output `EspressoPerformanceMonitor` also includes a table of goodness-of-fit test results, whose values are included in the first column in Table G.1. The X^2 and S scores in the two bottom rows are related to the intrinsic
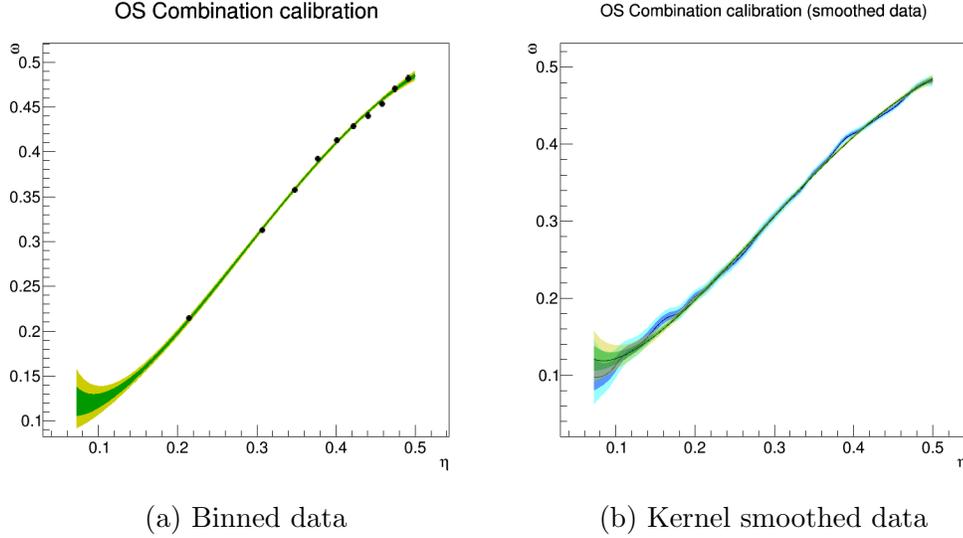
(a) Binned data	(b) Kernel smoothed data

Figure G.4: Cubic calibration to $B^{\pm} \to J/\psi\,K^{\pm}$; the green and yellow areas represent 1σ and 2σ bands for the calibration, while the blue and light blue areas in the right-hand plot represent 1σ and 2σ confidence intervals in the kernel-smoothed data.

quality of the fit and, if the fit model is good, should follow standard normal distributions. The le Cessie-van Houwelingen-Copas-Hosmer score 3.5885 stands out, since this corresponds to a p-value of 0.0003, and marks this as a poor fit. It is important to note that the Pearson score, nearly 0, does not reject the fit. No one goodness-of-fit test can reject all kinds of model mis-specifications in binomial regression.

A more complicated cubic polynomial calibration, with 4 degrees of freedom, can also be performed on the $B^{\pm} \to J/\psi\,K^{\pm}$ sample. To keep the calibrated probability in the range $0 \leq \omega \leq 1$, the logistic link function is used, and the results are shown in Fig. G.4. Visually, this model tracks the data much more closely than the linear model. Its goodness-of-fit tests scores are also improved, equal to 2.2σ and -1.7σ. Still, the Pearson score shows some evidence that the cubic model is mis-specified, with a p-values equal to 0.028.

An even more complicated calibration is a b-spline model with 1 internal knot, which as 5 degrees of freedom. Visually inspecting the fits, shown in Fig. G.5, this
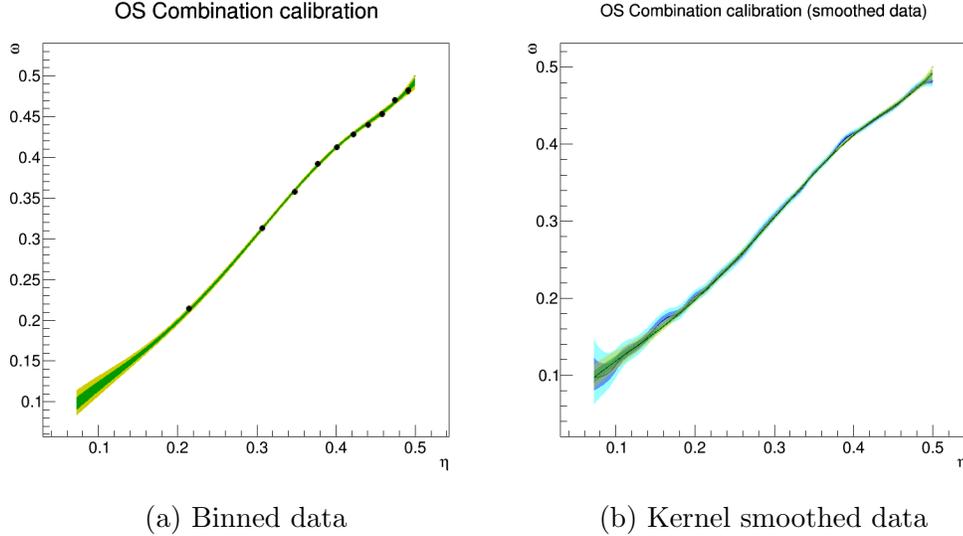
(a) Binned data　　　　　　(b) Kernel smoothed data

Figure G.5: B-spline calibration to $B^{\pm} \rightarrow J/\psi K^{\pm}$; the green and yellow areas represent 1σ and 2σ bands for the calibration, while the blue and light blue areas in the right-hand plot represent 1σ and 2σ confidence intervals in the kernel-smoothed data.

model tracks the data much more closely than the linear model. With goodness-of-fit tests scores of -0.9 and -1.2, there is no evidence of any tension between the fit model and the data. Visually, this fit looks quite similar to the cubic polynomial fit, possibly with better agreement in the range $0.35 < \eta < 0.45$, where there is a noticeable kink in the data that the cubic polynomial cannot quite match.

An alternative n-spline model can also be fit to the data. Because of its boundary conditions, an n-spline model with 5 degrees of freedom has three internal knots (to the b-spline model's one). This might allow better fit agreement for central values of η, at the cost of worse agreement in the tails. Its goodness-of-fit tests scores are also acceptable, and like the b-spline model it appears to do well in the "kink" region between $0.35 < \eta < 0.45$ (Fig. G.6).

Finally, a yet more intricate n-spline model with four internal knots and six degrees of freedom yields the fits shown in Fig. G.7. The visual difference between this fit and the n-spline fit with 5 degrees of freedom is nearly imperceptible.

Selecting the best fit using the Akaike information criterion (AIC) or Bayesian
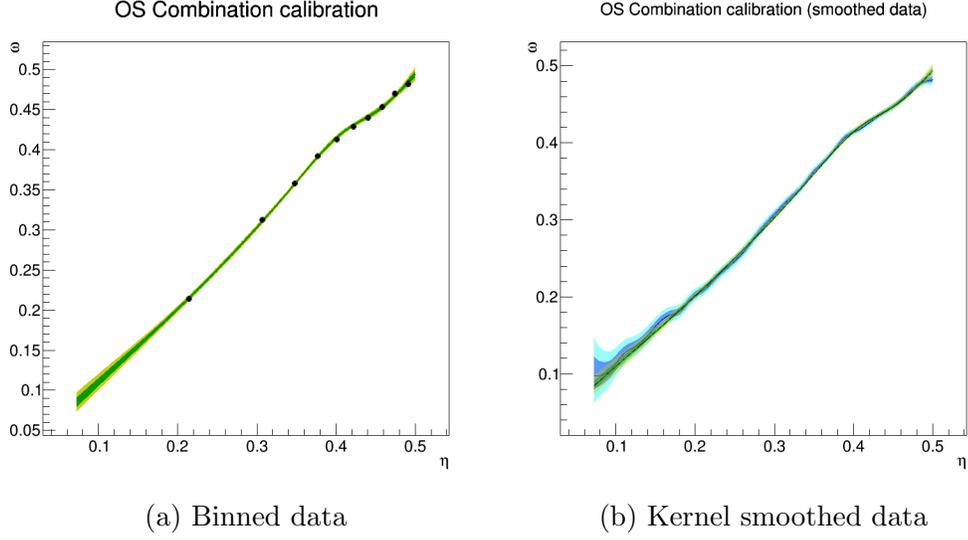
(a) Binned data   (b) Kernel smoothed data

Figure G.6: N-spline calibration to $B^{\pm} \rightarrow J/\psi K^{\pm}$; the green and yellow areas represent 1σ and 2σ bands for the calibration, while the blue and light blue areas in the right-hand plot represent 1σ and 2σ confidence intervals in the kernel-smoothed data.

(a) Binned data   (b) Kernel smoothed data

Figure G.7: Enhanced n-spline calibration to $B^{\pm} \rightarrow J/\psi K^{\pm}$; the green and yellow areas represent 1σ and 2σ bands for the calibration, while the blue and light blue areas in the right-hand plot represent 1σ and 2σ confidence intervals in the kernel-smoothed data.
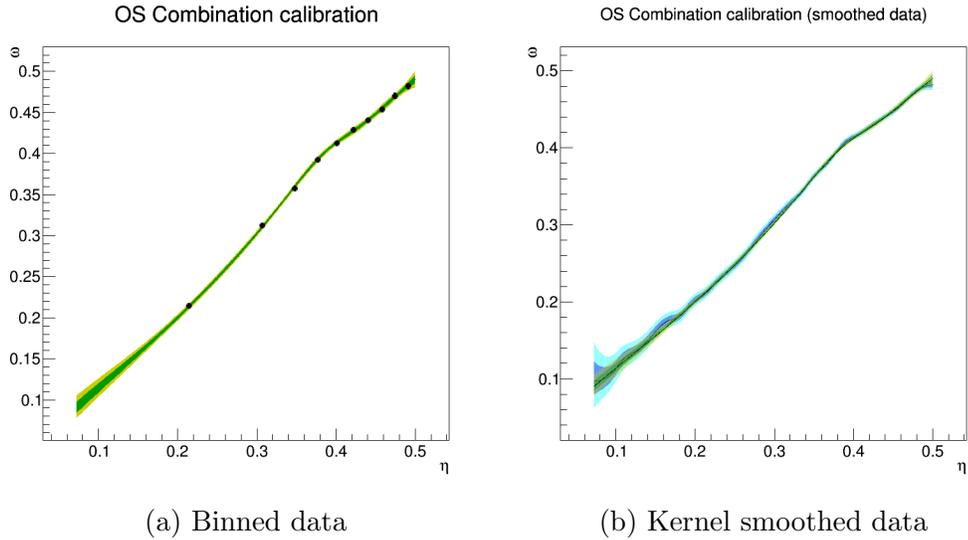
Table G.1: Values of goodness-of-fit tests for various types of calibrations

Calibration	Deviance	AIC	BIC	X^2	S
Linear polynomial	377322.84	377326.84	377350.36	0.0σ	3.6σ
Cubic polynomial	377264.22	377272.22	377319.27	2.2σ	-1.7σ
B-spline	377253.02	377263.02	377321.83	-0.9σ	-1.2σ
N-spline	377253.63	377263.63	377322.44	0.6σ	0.3σ
Enhanced N-spline	377248.06	377260.06	377330.63	-0.4σ	-0.1σ

information criterion (BIC) is subjective. Each criterion penalizes the deviance according to the number of fit parameters. The BIC penalty scales with the logarithm of the number of events, reflecting the fact that a good extra parameter should improve the likelihood across the board, and for large data samples the BIC is perhaps a superior metric. However, it is not clear whether the number of parameters to use should include the asymmetry parameters, which are not expected to differ significantly from zero but are still needed for assigning systematic uncertainty. By fiat, the `EspressoPerformanceMonitor` does not include penalties from the asymmetry parameters. Table G.1 reports the goodness-of-fit test scores for the various calibrations shown here. According the AIC score, the enhanced n-spline model is the best model. According to the BIC score, on the other hand, the cubic model is the best, by a slim margin. The extra degrees of freedom in the spline models, especially the enhanced n-spline model, do not reduce the deviance enough to counter the penalty that the BIC scores applies to them.

Bibliography

[1] Frances Halzen and Alan D. Martin. *Quarks and Leptons.* John Wiley & Sons, Inc., 1984.

[2] Cheng and Li. *Gauge theory of elementary particle physics.* Oxford Science Publications, 1984.

[3] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, 1995.

[4] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[5] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[6] Paul Langacker. *The Standard Model and Beyond.* Series in High Energy Physics, Cosmology, and Gravitation. CRC Press - Taylor & Francis Group, 2010.

[7] Report of the BABAR Physics Workshop. The BABAR physics book: Physics at an asymmetric B factory. Technical report, The BABAR Collaboration, 1998.

[8] Makoto Kobayashi and Toshihide Maskawa. *CP*-violation in the renormalizable theory of weak interaction. *Progress of Theoretical Physics*, 49(2):652–657, 1973.

[9] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. Evidence for the 2 pi Decay of the k(2)0 Meson. *Phys. Rev. Lett.*, 13:138–140, 1964.

[10] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.

[11] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[12] K. A. Olive et al. Review of particle physics. *Chin. Phys.*, C38:090001, 2014.

[13] Y. Amhis et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016. 2016.

[14] R. Aaij et al. Test of lepton universality using $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays. *Phys. Rev. Lett.*, 113:151601, 2014.

[15] R. Aaij et al. Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using $3\,\text{fb}^{-1}$ of integrated luminosity. *JHEP*, 02:104, 2016.

[16] J.P. Lees et al. Measurement of an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays and implications for charged Higgs bosons. *Phys. Rev. D*, 88:072012, 2013.

[17] M. Huschle et al. Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle. 2015.

[18] A. Abdesselam et al. Measurement of the branching ratio of $\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method. 2016.

[19] S. Hirose et al. Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$. *Phys. Rev. Lett.*, 118(21):211801, 2017.

[20] R. Aaij et al. Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$. *Phys. Rev. Lett.*, 115:111803, 2015.

[21] R. Aaij et al. Measurement of the $\mathcal{B}(B^0 \to D^{*-}\tau^+\nu_\tau)$ branching fraction using three-prong τ decays. 2017. in preparation.

[22] Jernej F. Kamenik and Federico Mescia. $B \to D\tau\nu$ Branching Ratios: Opportunity for Lattice QCD and Hadron Colliders. *Phys. Rev.*, D78:014003, 2008.

[23] Svjetlana Fajfer, Jernej F Kamenik, and Ivan Nišandžić. On the $B \to D^*\tau\nu_\tau$ sensitivity to new physics. *Phys. Rev. D*, 85(9):094025, May 2012.

[24] Marat Freytsis, Zoltan Ligeti, and Joshua T. Ruderman. Flavor models for $\bar{B} \to D^{(*)}\tau\bar{\nu}$. *Phys. Rev.*, D92(5):054018, 2015.

[25] Xiao-Gang He, G. C. Joshi, H. Lew, and R. R. Volkas. Simplest Z' model. *Phys. Rev. D*, 44:2118–2132, Oct 1991.

[26] Andreas Crivellin, Giancarlo D'Ambrosio, and Julian Heeck. Addressing the LHC flavor anomalies with horizontal gauge symmetries. *Phys. Rev.*, D91(7):075006, 2015.

[27] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[28] Some LHC milestones.... Quelques dates cls du LHC... (BUL-NA-2008-132. 38/2008. 38/2008):1, Sep 2008. Cover article.

[29] LHCb reoptimized detector design and performance: Technical Design Report, 2003. LHCb-TDR-009.

[30] A. A. Alves Jr. et al. The LHCb detector at the LHC. *JINST*, 3:S08005, 2008.

[31] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[32] Serguei Chatrchyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV. *Phys. Lett.*, B722:5–27, 2013.

[33] R. Aaij et al. Measurement of the inelastic pp cross-section at a centre-of-mass energy of $\sqrt{s} = 7$ TeV. *JHEP*, 02:029, 2015.

[34] H. Van Haevermaet. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. In *Proceedings, 24th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2016): Hamburg, Germany, April 11-25, 2016*, 2016.

[35] Morad Aaboud et al. Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC. 2016.

[36] R. Aaij et al. Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region. *Phys. Lett.*, B694:209, 2010.

[37] R. Aaij et al. Prompt charm production in pp collisions at $\sqrt{s} = 7$ TeV. *Nucl. Phys.*, B871:1, 2013.

[38] R. Aaij et al. Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP*, 03:159, 2016.

[39] R. Aaij et al. Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions. *Phys. Rev. Lett.*, 118:052002, 2017. Erratum in preparation.

[40] Glenn F. Knoll. *Radiation Detection and Measurement*. John Wiley & Sons, Inc., 2010.

[41] LHCb VELO (VErtex LOcator): Technical Design Report, 2001. LHCb-TDR-005.

[42] R. Aaij et al. LHCb detector performance. *Int. J. Mod. Phys.*, A30:1530022, 2015.

[43] M. Adinolfi et al. Performance of the LHCb RICH detector at the LHC. *Eur. Phys. J.*, C73:2431, 2013.

[44] Albert Puig. The LHCb trigger in 2011 and 2012. Technical Report LHCb-PUB-2014-046. CERN-LHCb-PUB-2014-046, CERN, Geneva, Nov 2014.

[45] Giulio Dujany and Barbara Storaci. Real-time alignment and calibration of the LHCb Detector in Run II. Apr 2015.

[46] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[47] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. A brief introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178:852–867, 2008.

[48] I. Belyaev et al. Handling of the generation of primary events in Gauss, the LHCb simulation framework. *J. Phys. Conf. Ser.*, 331:032047, 2011.

[49] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[50] Piotr Golonka and Zbigniew Was. PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays. *Eur.Phys.J.*, C45:97–107, 2006.

[51] John Allison, K. Amako, J. Apostolakis, H. Araujo, P.A. Dubois, et al. Geant4 developments and applications. *IEEE Trans.Nucl.Sci.*, 53:270, 2006.

[52] S. Agostinelli et al. Geant4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250, 2003.

[53] Clemencic, M. and others. The LHCb simulation application, Gauss: Design, evolution and experience. *J. Phys. Conf. Ser.*, 331:032023, 2011.

[54] A. Tsaregorodtsev, N. Brook, A. Casajus Ramo, P. Charpentier, J. Closier, et al. DIRAC3: The new generation of the LHCb grid software. *J. Phys. Conf. Ser.*, 219:062029, 2010.

[55] Ricardo Graciani Diaz, Adria Casajus Ramo, Ana Carmona uero, Thomas Fifield, and Martin Sevior. Belle-DIRAC setup for using Amazon Elastic Compute Cloud. *Journal of Grid Computing*, 9:65–79, 2011.

[56] A Yu Anisimov, I M Narodetskii, C Semay, and B Silvestre-Brac. The B_c meson lifetime in the light-front constituent quark model. *Phys. Lett. B*, 452(1-2):129–136, April 1999.

[57] V.V. Kiselev. Exclusive decays and lifetime of B_c meson in QCD sum rules.

[58] Mikhail A. Ivanov, Jurgen G. Korner, and Pietro Santorelli. Exclusive semileptonic and nonleptonic decays of the B_c meson. *Phys. Rev. D*, 73:054024, 2006.

[59] E. Hernandez, J. Nieves, and J.M. Verde-Velasco. Study of exclusive semileptonic and non-leptonic decays of B_c - in a nonrelativistic quark model. *Phys. Rev. D*, 74:074008, 2006.

[60] F. Abe et al. Observation of B_c mesons in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. D*, 58:112004, 1998.

[61] F. Abe et al. Observation of the B_c meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett.*, 81:2432–2437, 1998.

[62] R. Aaij et al. Observation of the decay $B_c^+ \rightarrow B_s^0 \pi^+$. *Phys. Rev. Lett.*, 111:181801, 2013.

[63] D Ebert, R Faustov, and V Galkin. Weak decays of the B_c meson to charmonium and D mesons in the relativistic quark model. *Phys. Rev. D*, 68(9):094020, November 2003.

[64] Irinel Caprini, Laurent Lellouch, and Matthias Neubert. Dispersive bounds on the shape of $\bar{B} \rightarrow D^{(*)} \ell \bar{\nu}$ form-factors. *Nucl. Phys. B*, 530:153–181, 1998.

[65] Xiao Xia Wang, Wei Wang, and Cai Dian Lü. B_c to p-wave charmonia transitions in covariant light-front approach. *Phys. Rev. D*, 79:114018, Jun 2009.

[66] Chao-Hsi Chang and Yu-Qi Chen. Decays of the B_c meson. *Phys. Rev. D*, 49(7):3399–3411, April 1994.

[67] A Abd El-Hady, J Muñoz, and J Vary. Semileptonic and nonleptonic B_c decays. *Phys. Rev. D*, 62(1):014019, June 2000.

[68] Chao Hsi Chang, Yu Qi Chen, Guo Li Wang, and Hong Shi Zong. Decays of the meson B_c to a P wave charmonium state χ_c or h_c. *Phys. Rev. D*, 65:014017, 2002.

[69] Ho-Meoyng Choi and Chueng-Ryong Ji. Semileptonic and radiative decays of the B(c) meson in light-front quark model. *Phys. Rev.*, D80:054016, 2009.

[70] P Colangelo and F De Fazio. Using heavy quark spin symmetry in semileptonic B-c decays. *Phys. Rev. D*, 61(3), 2000.

[71] Pietro Colangelo and Alexander Khodjamirian. QCD sum rules, a modern perspective. 2000.

[72] Logunov, A. A. and Tavkhelidze, A. N. Quasioptical approach in quantum field theory. *Nuovo Cim.*, 29:380–399, 1963.

[73] Martynenko, A. P. and Faustov, R. N. Relativistic Reduced Mass and Quasipotential Equation. *Theor. Math. Phys.*, 64:765–770, 1985. [Teor. Mat. Fiz.64,179(1985)].

[74] D Ebert, R Faustov, and V Galkin. Properties of heavy quarkonia and B_c mesons in the relativistic quark model. *Physical Review D*, 67(1):014027, January 2003.

[75] Claude Bourrely, Irinel Caprini, and Laurent Lellouch. Model-independent description of B πlν decays and a determination of $|V_{ub}|$. *arXiv.org*, (1):013008, July 2008.

[76] Heechang Na, Chris M Bouchard, G Peter Lepage, Chris Monahan, and Junko Shigemitsu. $B \rightarrow D\ell\nu$ Form Factors at Non-Zero Recoil and Extraction of $|V_{cb}|$. *arXiv.org*, (5):054510, May 2015.

[77] R. Aaij et al. Bose-Einstein correlations of same-sign charged pions in the forward region in pp collisions at $\sqrt{s} = 7$ TeV. 2017. in preparation.

[78] Wouter D. Hulsbergen. Decay chain fitting with a Kalman filter. *Nucl.Instrum.Meth.*, A552:566–575, 2005.

[79] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical Report CERN-OPEN-2012-016, New York U., New York, Jan 2012.

[80] Roger J. Barlow and Christine Beeston. Fitting using finite Monte Carlo samples. *Comput. Phys. Commun.*, 77:219–228, 1993.

[81] R. Aaij et al. Observation of $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays. *Phys. Rev.*, D87:112012, 2013.

[82] Georges Aad et al. Study of the $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays with the ATLAS detector. *Eur. Phys. J.*, C76(1):4, 2016.

[83] Lucio Anderlini, Andrea Contu, Christopher Rob Jones, Sneha Sirirshkumar Malde, Dominik Muller, Stephen Ogilvy, Juan Martin Otalora Goicochea, Alex Pearce, Ivan Polyakov, Wenbin Qian, Barbara Sciascia, Ricardo Vazquez Gomez, and Yanxi Zhang. The PIDCalib package. Technical Report LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul 2016.

[84] A. Rogozhnikov. Reweighting with Boosted Decision Trees. *J. Phys. Conf. Ser.*, 762(1):012036, 2016.

[85] Muriel Pivk and Francois R. Le Diberder. sPlot: A statistical tool to unfold data distributions. *Nucl.Instrum.Meth.*, A555:356–369, 2005.

[86] R. Aaij et al. Observation of $B_c^+ \to J/\psi D^{(*)} K^{(*)}$ decays. *Phys. Rev.*, D95:032005, 2017.

[87] J. Charles, Andreas Hocker, H. Lacker, S. Laplace, F. R. Le Diberder, J. Malcles, J. Ocariz, M. Pivk, and L. Roos. CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories. *Eur. Phys. J.*, C41(1):1–131, 2005.

[88] Ikaros I. Y. Bigi and A. I. Sanda. Notes on the Observability of CP Violations in B Decays. *Nucl. Phys.*, B193:85–108, 1981.

[89] R. Aaij et al. Measurement of *CP* violation in $B^0 \to J/\psi K_{\mathrm{s}}^0$ decays. *Phys. Rev. Lett.*, 115:031601, 2015.

[90] R. Aaij et al. Measurement of the *CP*-violating phase β in $\overline{B}^0 \to J/\psi \pi^+ \pi^-$ decays and limits on penguin effects. *Phys. Lett.*, B742:38, 2015.

[91] R Aaij et al. Measurement of b-hadron branching fractions for two-body decays into charmless charged hadrons. *JHEP*, 10:037, 2012.

[92] R. Aaij et al. Precision measurement of *CP* violation in $B_s^0 \to J/\psi K^+ K^-$ decays. *Phys. Rev. Lett.*, 114:041801, 2015.

[93] R. Aaij et al. Measurement of the *CP* violating phase and decay-width difference in $B_s^0 \to \psi(2S)\phi$ decays. *Phys. Lett.*, B762:253, 2016.

[94] Bernard Aubert et al. Measurement of $B^0 - \overline{B}^0$ flavor oscillations in hadronic B^0 decays. *Phys. Rev. Lett.*, 88:221802, 2002.

[95] Bernard Aubert et al. A study of time dependent CP-violating asymmetries and flavor oscillations in neutral B decays at the $\Upsilon(4S)$. *Phys. Rev.*, D66:032003, 2002.

[96] Bernard Aubert et al. Measurements of time-dependent CP asymmetries in $B^0 \to D^{(*)} + D^{(*)}$ - decays. *Phys. Rev.*, D79:032002, 2009.

[97] Kazuo Abe et al. An Improved measurement of mixing induced CP violation in the neutral B meson system. *Phys. Rev.*, D66:071102, 2002.

[98] T. Tomura et al. Measurement of the oscillation frequency for B0 - anti-B0 mixing using hadronic B0 decays. *Phys. Lett.*, B542:207–215, 2002.

[99] H. Kakuno et al. Neutral B flavor tagging for the measurement of mixing induced CP violation at Belle. *Nucl. Instrum. Meth.*, A533:516–531, 2004.

[100] M. Rohrken et al. Measurements of Branching Fractions and Time-dependent CP Violating Asymmetries in $B^0 \to D^{(*)\pm}D^{\mp}$ Decays. *Phys. Rev.*, D85:091106, 2012.

[101] T. Aaltonen et al. First Flavor-Tagged Determination of Bounds on Mixing-Induced CP Violation in $B_s^0 \to J/\psi\phi$ Decays. *Phys. Rev. Lett.*, 100:161802, 2008.

[102] V. M. Abazov et al. Measurement of B_d mixing using opposite-side flavor tagging. *Phys. Rev.*, D74:112002, 2006.

[103] V. M. Abazov et al. Measurement of B_s^0 mixing parameters from the flavor-tagged decay $B_s^0 \to J/\psi\phi$. *Phys. Rev. Lett.*, 101:241801, 2008.

[104] Georges Aad et al. Flavor tagged time-dependent angular analysis of the $B_s \to J/\psi\phi$ decay and extraction of $\Delta\Gamma_s$ and the weak phase ϕ_s in ATLAS. *Phys. Rev.*, D90(5):052007, 2014.

[105] Georges Aad et al. Measurement of the CP-violating phase ϕ_s and the B_s^0 meson decay width difference with $B_s^0 \to J/\psi\phi$ decays in ATLAS. *JHEP*, 08:147, 2016.

[106] Vardan Khachatryan et al. Measurement of the CP-violating weak phase ϕ_s and the decay width difference $\Delta\Gamma_s$ using the $B_s^0 \to J/\psi\phi(1020)$ decay channel in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett.*, B757:97–120, 2016.

[107] R. Aaij et al. Opposite-side flavour tagging of B mesons at the LHCb experiment. *Eur. Phys. J.*, C72:2022, 2012.

[108] R. Aaij et al. B flavour tagging using charm decays at the LHCb experiment. *JINST*, 10:P10005, 2015.

[109] R. Aaij et al. Neural-network-based same side kaon tagging algorithm calibrated with $B_s^0 \to D_s^-\pi^+$ and $B_{s2}^*(5840)^0 \to B^+K^-$ decays. *JINST*, 11:P05010, 2016.

[110] R. Aaij et al. New algorithms for identifying the flavour of B^0 mesons using pions and protons. *Eur. Phys. J.*, C77:238, 2017.

[111] Michael Gronau, Alex Nippe, and Jonathan L. Rosner. Method for flavour tagging in neutral B meson decays. *Phys.Rev.*, D47:1988–1993, 1993.

[112] Optimization and calibration of the same-side kaon tagging algorithm using hadronic B_s^0 decays in 2011 data, Nov 2012.

[113] R. Aaij et al. Measurement of the time-dependent CP asymmetries in $B_s^0 \to J/\psi\,K_S^0$. *JHEP*, 06:131, 2015.

[114] R. Aaij et al. Measurement of CP violation in $B \to D^+D^-$ decays. *Phys. Rev. Lett.*, 117:261801, 2016.

[115] David W. Hosmer, Jr., Stanley Lemeshow, and Rodney X Sturdivant. *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley, 3rd edition, 2013.

[116] Alan Agresti. *Categorial Data Analysis*. Wiley Series in Probability and Statistics. Wiley, 3rd edition, 2013.

[117] K Kreplin, G Krocker, and S Hansmann-Menzemer. The opposite-side kaon tagger: Datasimulationperformance comparison and optimization using neural networks. LHCb-INT-2013-014, July 2013.

[118] Marc Grabalosa. *Flavour Tagging developments within the LHCb experiment*. PhD thesis, Barcelona U., Mar 2012. Presented 15 May 2012.

[119] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and regression trees*. Wadsworth international group, Belmont, California, USA, 1984.

[120] Robert E. Schapire and Yoav Freund. A decision-theoretic generalization of on-line learning and an application to boosting. *Jour. Comp. and Syst. Sc.*, 55:119, 1997.

[121] R. Aaij et al. Measurement of the *CP*-violating phase ϕ_s in $\overline{B}_s^0 \to D_s^+ D_s^-$ decays. *Phys. Rev. Lett.*, 113:211801, 2014.

[122] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[123] Tim Adye. Unfolding algorithms and tests using RooUnfold. (arXiv:1105.1160):6 p, May 2011. Comments: 6 pages, 5 figures, presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report.

[124] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, 2001.

[125] Frederick James. *Statistical Methods in Experimental Physics*. Springer Series in Statistics. Springer, 2nd edition, 2001.

[126] David W Hosmer, Trina Hosmer, Saskia Le Cessie, Stanley Lemeshow, et al. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, 16(9):965–980, 1997.

[127] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[128] M. Galassi et al. *GNU Scientific Library Reference Manual*. Network Theory Ltd., third edition, 2009.

[129] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.

[130] José Hernández-Orallo, Peter Flach, and Cesar Ferri. A unified view of performance metrics: translating threshold choice into expected classification loss. *The Journal of Machine Learning Research*, 13(1):2813–2869, 2012.

[131] Noel Cressie and Timothy RC Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 440–464, 1984.

[132] Jocelyn R. Dale. Asymptotic normality of goodness-of-fit statistics for sparse product multinomials. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(1):48–59, 1986.

[133] Gerhard Osius and Dieter Rojek. Normal goodness-of-fit tests for multinomial models with large degrees of freedom. *Journal of the American Statistical Association*, 87(420):1145–1152, 1992.