$\mathcal{CP}$-Violation in Beautiful-Strange Oscillations at LHCb

Robert A. Currie

Doctor of Philosophy

The University of Edinburgh

August 2013

Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This thesis is written in accordance with the postgraduate assessment regulations for research degrees.

(*Robert A. Currie*)

Abstract

The LHCb experiment is an experiment based at the LHC in Geneva and is dedicated to the study of mesons containing bottom and charm quarks. One of the primary goals of the physics at LHCb is to measure $\mathcal{CP}$-violating effects which lead to a dominance of matter over anti-matter in the universe. This thesis presents the measurement of the $\mathcal{CP}$-violating phase ϕ_s which is one of the golden channels at LHCb. This phase is observed as the interference between mixing of $B_s^0 \leftrightarrow \overline{B}_s^0$ and decay of $B_s^0 \to J/\psi\, K^+K^-$. The results, based upon the $1.0\,\text{fb}^{-1}$ dataset collected by LHCb during 2011, are:

$$\begin{aligned} \phi_s &= 0.07 \pm 0.09 \pm 0.01\,\text{rad}\ , \\ \Delta\Gamma_s &= 0.100 \pm 0.016 \pm 0.002\,\text{ps}^{-1}\ , \\ \Gamma_s &= 0.663 \pm 0.005 \pm 0.006\,\text{ps}^{-1}\ . \end{aligned}$$

This analysis is also able to measure the mixing parameter $\Delta m_s = 17.71 \pm 0.10 \pm 0.01\,\text{ps}^{-1}$. To improve upon this measurement the $B_s^0 \to J/\psi\, K^+K^-$ analysis is combined with the $B_s^0 \to J/\psi\, \pi^+\pi^-$ decay channel to make the most accurate measurements to date of, $\phi_s = 0.01 \pm 0.07 \pm 0.01\,\text{rad}, \Delta\Gamma_s = 0.106 \pm 0.011 \pm 0.007\,\text{ps}^{-1}$ and $\Gamma_s = 0.661 \pm 0.004 \pm 0.006\,\text{ps}^{-1}$. As an integral part of this work a comprehensive software suite known as RapidFit was developed, which is used by many other physicists and this is described.

Lay Abstract

For the universe that we observe to exist it is required that there are processes which lead to an excess of matter over anti-matter in the early universe. Current theories include such processes but predict only enough to account for approximately one galaxy, and therefore there must be new physics phenomena which we have not yet discovered. The LHCb experiment, situated at the Large Hadron Collider in Geneva, in searching for such new phenomena. This thesis presents results obtained from the LHCb data to measure the properties of the decays of B_s^0 mesons which may signal new differences between matter and anti-matter. As an integral part of this work a comprehensive software suite known as RapidFit was developed, and which is used by many other physicists and this is described.

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Peter Clarke for his tutoring throughout the thesis writing and my PhD studies. Also, I would like to specifically thank Matthew Needham and Greig Cowan who both provided invaluable assistance during the writing of this thesis.

I would like to thank, Ailsa Sparkes and Conor Fitzpatrick for sharing their expert knowledge on all things LHCb. Ben Wynne, who laid the groundwork for what has become the ever evolving RapidFit fitting framework. As well as all of the other students and post-docs I've met and worked with throughout my PhD, specifically those in the Edinburgh PPE group. Although too numerous to name them all, they've helped keep me on track throughout my PhD and thesis writing, as well as providing useful and informative conversations in and out of the office.

Finally, but by no means least, I would like to thank my parents who have always supported me throughout my academic studies. They have been just a phone call away when needed and have always been a helping hand.

Table of Contents

List of figures

List of tables

"It's taken me seven years, and it's perfect... My magnum opus, Baldrick. Everybody has one novel in them, and this is mine."

Edmund Blackadder

1

Introduction

One of the observations of modern physics is that the visible universe is composed entirely of matter particles. However the universe is assumed to have initially been composed of equal amounts of matter and anti-matter. In order for the universe we observe today to exist a process leading to an excess of baryons over anti-baryons must have occurred.

In 1967 Andrei Sakharov [1] proposed three conditions that must be met for a process known as baryogenesis to occur. These conditions are:

- Interactions must occur out of thermal equilibrium.
- Baryon number must be violated.
- Processes violating the Charge-Parity ($\mathcal{CP}$) symmetry must be occurring.

The Standard Model, which describes the interaction of fundamental particles, allows for processes meeting all three of these conditions to occur. The subject of this thesis is the third of these conditions which requires $\mathcal{CP}$-violating processes to be possible in the Standard Model.

To understand if the Standard Model satisfies the third Sakharov condition, the amount of $\mathcal{CP}$ violation within this model can be compared to the amount of $\mathcal{CP}$ violation observed in the universe.

The amount of $\mathcal{CP}$ violation in the universe is related to ratio of the baryon (n_B) and photon (n_γ) numbers present in the cosmic microwave background. Recent measurements of this ratio (η) by the WMAP collaboration [2] determine this to be:

$$\eta = \frac{n_B}{n_\gamma} = (6.19 \pm 0.14) \times 10^{-10} \ . \tag{1.1}$$

Within the Standard Model the only observed $\mathcal{CP}$-violating processes are described by the CKM matrix [3, 4]. (This is discussed in more detail in Section 2.3.) A quantity (C) proportional to the amount of $\mathcal{CP}$ violation present in this matrix can be constructed using equation 1.2 [5].

$$\mathrm{C} = \left(m_t^2 - m_u^2\right)\left(m_t^2 - m_c^2\right)\left(m_c^2 - m_u^2\right)\left(m_b^2 - m_d^2\right)\left(m_b^2 - m_s^2\right)\left(m_s^2 - m_d^2\right) J \ . \tag{1.2}$$

Here the various mass parameters m_i correspond to the mass of the ith quark and J is a quantity known as the Jarlskog invariant [5]. The quantity C has units of $(\,\mathrm{GeV}/c^2)^{12}$ and so is typically normalised using the temperature (T) of the universe at the time these interactions were taking place, raised to the twelfth power (T^{12}). This temperature is normally taken to be the energy corresponding to the electro-weak phase transition $\mathcal{O}\,(100\,\mathrm{GeV}/c^2)$. Using the known quark masses (Table 2.1) and knowledge that $J \backsimeq 3 \times 10^{-5}$ [6, 7] the amount of $\mathcal{CP}$ violation in the Standard Model is estimated to be [6]:

$$\mathrm{C}\big/T^{12} \approx \mathcal{O}\left(10^{-20}\right) \ . \tag{1.3}$$

This asymmetry falls short of the astronomically observed quantity (equation 1.1) by a factor of 10^{-10} [6, 8]. This is equivalent to stating that the amount of $\mathcal{CP}$ violation in the Standard Model is enough to account for just one galaxy in the observable universe. This large difference of ten orders of magnitude is difficult to account for within the Standard Model and suggests the presence of new $\mathcal{CP}$-violating processes which have not yet been observed.

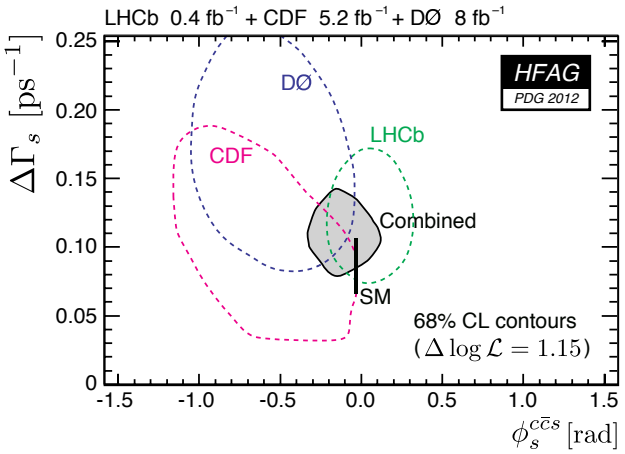


Figure 1.1: *Published measurements of ϕ_s in $B_s^0 \to J/\psi\, K^+K^-$ as of 2012 including results from; CDF [9], DØ [10] and LHCb [11], with this figure from reference [7]. The shaded grey area corresponds to the combined 68% confidence level.*

To search for possible new $\mathcal{CP}$-violating effects, the amount of $\mathcal{CP}$ violation within the Standard Model is being tested to high precision. New physics can potentially alter known $\mathcal{CP}$-violating processes through the interaction of new particles with those of the Standard Model.

One parameter in the Standard Model which is sensitive to the effects of new physics is the $\mathcal{CP}$-violating phase ϕ_s. This phase is a measurement of the amount of $\mathcal{CP}$ violation within the interference between the mixing of $B_s^0 \leftrightarrow \overline{B}_s^0$ and the decay of these mesons to the final state $J/\psi\, K^+K^-$. The lifetime of the different $\mathcal{CP}$ final states in this decay is separated by the difference $\Delta\Gamma_s$.

Figure 1.1 shows the current precision of measurements made of the phase ϕ_s and the lifetime difference $\Delta\Gamma_s$. The results presented here include the measurements from the CDF [9] and DØ [10] experiments at the Tevatron and the first measurement of these parameters at LHCb [11]. The first LHCb result corresponds to a measurement made with the first $337\,\text{pb}^{-1}$ of data collected in 2011 and is already the world's most precise measurement.

This thesis presents the updated measurement of ϕ_s at LHCb using the $1.0\,\text{fb}^{-1}$ of data collected in 2011. The results of this analysis have recently been published in Ref [12] and improve the precision by a factor of 2.

"Physicists are made of atoms.
A physicist is an attempt by an atom to understand itself."

Parallel Worlds – Michio Kaku

2

Theoretical Background

The known interactions of fundamental particles is described by the Standard Model of particle physics (Section 2.1). This model has many symmetries which are important in understanding the possible interactions (Section 2.2). The Standard Model allows for $\mathcal{CP}$ violation within certain interactions as described in Section 2.3. Experimentally $\mathcal{CP}$ violation can present itself in different forms, these are described in Section 2.4. The phenomenology of the decay of $B_s^0 \to J/\psi\, K^+K^-$ is fully described within Section 2.5. As described in Chapter 1 the main observables of interest in this decay are ϕ_s and $\Delta\Gamma_s$ with the Standard Model estimates for these parameters given in Section 2.6.

2.1 The Standard Model

Particles within the Standard Model exist in three generations and their properties are summarised in Table 2.1. A complete review of all known particles and decay modes is published by the Particle Data Group (PDG) [7]. The known particles of matter are grouped into quarks and leptons. Quarks always occur in bound states but leptons exist as free particles. The quarks are grouped such that the *up*, *charm* and *top* quarks are referred to as up-type quarks and the *down*, *strange* and *bottom* quarks are known as down-type quarks. The lepton generations are referred to as *electron*, *muon* and *tau*, each composed of one charged lepton and one neutrino.

The forces described by the Standard Model are the electromagnetic, weak and strong forces. The electromagnetic force governs the interaction of charged particles through the exchange of photons (γ). The weak force governs the weak interactions of both quarks and leptons. Weak charged interactions are mediated by the exchange of a $W^{\pm}$ boson and weak neutral interactions are mediated by the exchange of Z bosons. The weak force is responsible for both radioactive decays and interactions that change the flavour of quarks. The strong force only interacts with quarks and is responsible for quark confinement, it is mediated through the exchange of gluons (g).

A scalar Higgs field is responsible for all particles having mass in the Standard Model. The presence of this Higgs field also requires the existence of an additional scalar Higgs boson (H).

Quarks

Name		Charge	Spin	Mass
up	(u)	$+2/3$	1/2	2.4 MeV/c^2
charm	(c)	$+2/3$	1/2	1.27 GeV/c^2
top	(t)	$+2/3$	1/2	171.2 GeV/c^2
down	(d)	$-1/3$	1/2	4.8 MeV/c^2
strange	(s)	$-1/3$	1/2	104 MeV/c^2
bottom	(b)	$-1/3$	1/2	4.2 GeV/c^2

Leptons

Name		Charge	Spin	Mass
electron	(e)	-1	1/2	0.511 MeV/c^2
muon	(μ)	-1	1/2	105.7 MeV/c^2
tau	(τ)	-1	1/2	1.777 GeV/c^2
electron neutrino	(ν_e)	0	1/2	<2.2 eV/c^2
muon neutrino	(ν_μ)	0	1/2	<0.17 MeV/c^2
tau neutrino	(ν_τ)	0	1/2	<15.5 MeV/c^2

Gauge Mediating Bosons

Name		Charge	Spin	Mass
photon	(γ)	.	1	0
gluon	(g)	.	1	1
W Boson	$(\mathrm{W}^{\pm})$	± 1	1	80.4 GeV/c^2
Z Boson	(Z^0)	.	1	91.2 GeV/c^2
Higgs Boson	(H)	.	0	(*)125 GeV/c^2

Table 2.1: *Various properties of the known Fundamental Particles of matter (quarks and leptons) and Gauge mediating bosons within the Standard Model as given in Ref [7].*
(*) Mass of the Higgs boson candidate observed by LHC experiments.

2.2 Symmetries in the Standard Model

Symmetries are important in the development and understanding of the Standard Model. This model is invariant under the application of various different transforms, each corresponding to a different symmetry. These different symmetries are described as either continuous or discrete in nature.

2.2.1 Continuous symmetries

Continuous symmetries in the Standard Model correspond to conserved quantities in accordance with Noether's theorem [13]. These symmetries correspond to the invariance of the Lagrangian through an infinitesimal transform in a given dimension. Three examples of continuous symmetries within the Standard Model are:

Symmetry	Conserved Quantity
time symmetry	Energy
translational symmetry	Linear momentum
rotational symmetry	Angular momentum

2.2.2 Discrete symmetries

A discrete symmetry corresponds to the Lagrangian being invariant under the application of an operator applied to it. Below are examples of three important discrete symmetries:

	Operator	Effect
$\mathcal{C}$	Charge Conjugation	Inverts the sign of all internal charges of particles
$\mathcal{P}$	Parity Conjugation	Reverse of the spatial and momentum direction
$\mathcal{T}$	Time Conjugation	Reverse the direction of time for the particle

Due to the fact that the Standard Model is a Lorentz invariant quantum field theory it is required to be invariant under the application of the combined $\mathcal{CPT}$ operator [14, 15].

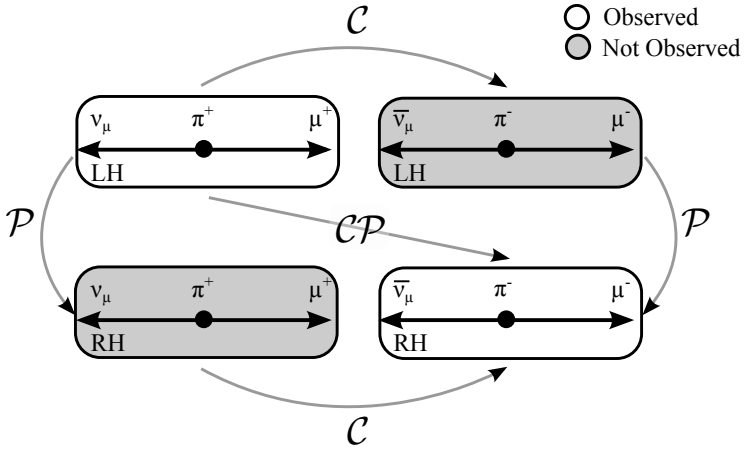


Figure 2.1: *This diagram shows the decay $\pi^+ \to \mu^+ \nu_\mu^{LH}$ and the corresponding processes anticipated due to $\mathcal{C}$, $\mathcal{P}$, and $\mathcal{CP}$ transforms.*
The decay modes involving right-handed neutrinos and left handed anti-neutrinos are not observed.

Originally it was assumed that the Standard Model was invariant under the application of each of these discrete operators. It was discovered by Wu et al. [16] in 1957 that the weak interaction maximally violates the parity symmetry. It is also observed that the charge symmetry is maximally violated within the Standard Model. Although both of the symmetries are broken in the Standard Model it was discovered that the theory was invariant under the combined $\mathcal{CP}$ operator.

An example of $\mathcal{C}$ and $\mathcal{P}$ violation may be shown by considering the weak decay of charged pions ($\pi^\pm$) as in Figure 2.1. Only left-handed neutrinos or right handed anti-neutrinos can interact with the weak force. The decay of $\pi^+ \to \mu^+ \nu_\mu^{LH}$ is allowed as the final state contains a left handed neutrino. This figure shows the processes which correspond to the separate $\mathcal{C}$ and $\mathcal{P}$ conjugates which are not observed. However, the process corresponding to the $\mathcal{CP}$ conjugate, $\pi^- \to \mu^- \nu_\mu^{RH}$, is observed.

Although the $\mathcal{C}$ and $\mathcal{P}$ symmetries are maximally violated the compound $\mathcal{CP}$ symmetry is conserved in many weak decays. Later in 1964 Cronin and Finch [17] discovered kaon decays violate this $\mathcal{CP}$ symmetry leaving the full Standard Model only invariant under the full $\mathcal{CPT}$ symmetry with both the $\mathcal{CP}$ and $\mathcal{T}$ symmetries being broken.

2.3 $\mathcal{CP}$ violation in the Standard Model

Within the Standard Model $\mathcal{CP}$ violation only occurs within the CKM matrix. This matrix is used within the Standard Model to describe the interaction of quarks with the weak force. The structure of the CKM matrix is discussed in more detail in the following subsection.

2.3.1 CKM matrix

The CKM matrix named after Cabibbo [3], Kobyashi and Maskawa [4], is required to describe the interaction of quarks with the weak force. In quantum mechanics the evolution of states with time is related to the mass eigenstates (Table 2.1), however, the the weak force acts upon the quark's weak eigenstates. The evolution of weak eigenstates with time can be determined by representing the weak eigenstates as a mixture of mass eigenstates. The convention normally chosen is such that the up-type mass and weak eigenstates are aligned, whilst the down-type mass eigenstates are rotated by the CKM to form the weak eigenstates. This is shown in equation 2.1, where the weak eigenstates are identified by "$_{\mathrm{W}}$".

$$\begin{pmatrix} d_{\mathrm{W}} \\ s_{\mathrm{W}} \\ t_{\mathrm{W}} \end{pmatrix} = \mathrm{V}_{CKM} \begin{pmatrix} d \\ s \\ t \end{pmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{pmatrix} d \\ s \\ t \end{pmatrix} . \tag{2.1}$$

To describe the mixing between the mass eigenstates the CKM matrix can be expressed in terms of three Euler angles and a $\mathcal{CP}$-violating phase δ. In this representation θ_{ij} represents the mixing angle between the generations i and j. To simplify the notation the the sin and cos of the different angles are denoted as c_{ij} and s_{ij} respectively. Expressing the CKM using Euler angles gives,

$$\begin{aligned}
\mathrm{V}_{CKM} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & s_{23}c_{13} \end{bmatrix} . \qquad (2.2)
\end{aligned}$$

One thing that is immediately apparent in this representation is that although $\mathcal{CP}$ violation is only present due to the single phase δ, this appears in multiple elements of the matrix with different coefficients. Another common representation of the CKM matrix is to use the Wolfenstein representation [18] using the definitions,

$$\begin{aligned}
&\lambda = s_{12} = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}} , \quad A\lambda^2 = s_{23} = \lambda \left| \frac{V_{cb}}{V_{us}} \right| , \\
&A\lambda^3 (\rho + i\eta) = s_{13}e^{i\delta} = V_{ub}^* = \left\{ \frac{A\lambda^3 (\overline{\rho} + i\overline{\eta}) \sqrt{1 - A^2\lambda^4}}{\sqrt{1 - \lambda^2} \left[1 - A^2\lambda^4 (\overline{\rho} + i\overline{\eta}) \right]} \right\} . \qquad (2.3)
\end{aligned}$$

These definitions allow the CKM matrix to be re-expressed as,

$$\begin{aligned}
\mathrm{V}_{CKM} &= \begin{bmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3 (\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3 (1 - \rho - i\eta) & -A\lambda^2 & 1 \end{bmatrix} + \Delta \left(\lambda^4 \right) . \qquad (2.4) \\
\Delta \left(\lambda^4 \right) &= \begin{bmatrix} -\frac{1}{8}\lambda^4 & 0 & 0 \\ A^2\lambda^5 \left(\frac{1}{2} - \rho - i\eta \right) & -\frac{1}{8}\lambda^4 (1 + 4A^2) & 0 \\ \frac{1}{2}A\lambda^5 (\rho + i\eta) & +A\lambda^4 \left(\frac{1}{2} - \rho - i\eta \right) & -\frac{1}{2}A^2\lambda^4 \end{bmatrix} + \mathcal{O} \left(\lambda^6 \right) .
\end{aligned}$$

Using this representation the relative size of the elements in terms of $\mathcal{O}(\lambda^n)$ is clear. Expanding the CKM matrix up to the order of $\mathcal{O}(\lambda^4)$ provides the additional useful representation of the matrix as,

$$\mathrm{V}_{CKM} = \begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}|e^{-i\gamma} \\ -|V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}|e^{-i\beta} & -|V_{ts}|e^{+i\beta_s} & |V_{tb}| \end{bmatrix} + \delta\left(\lambda^5\right) \ . \tag{2.5}$$

From this representation it is easily seen that to the precision of λ^4, only the phases of three elements within the CKM can be measured.

2.3.2 Unitarity triangles

The condition of unitarity on the CKM matrix gives nine separate requirements using the elements in equation 2.1 $\sum_k V_{ki}V_{kj}^* = \delta_{ij}$. Each of these quantities $V_{ik}V_{jk}^*$ can be represented by vectors in the complex plane. There are six conditions requiring that the sum of three complex numbers is 0. This is the same as requiring their path in the complex plane forms a closed triangle.

Two unitarity triangles are shown Figure 2.2 and are constructed from the unitarity relations,

$$\begin{aligned} &V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \ , \\ &V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0 \ . \end{aligned} \tag{2.6}$$

These two conditions are similar and are commonly referred to as, the B^0 unitarity triangle in Figure 2.2 (a) and, the B_s^0 triangle in Figure 2.2 (b). The B^0 triangle contains the angles β and γ from the CKM matrix as in equation 2.5. The B_s^0 triangle contains the angle β_s from this matrix.

The unitarity triangles are normally represented with one side normalised to lie on the real axis, making use of the parameters $(\overline{\rho}, \overline{\eta})$ in equation 2.3. In the B^0 unitarity triangle, all sides have the same scale in powers of λ. The B_s^0 unitarity triangle, however, has one side suppressed by a factor of λ^2. From the second triangle the definition of β_s is given in equation 2.7 as in Figure 2.2 (b).

$$\beta_s = -\arg\left(\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) \ . \tag{2.7}$$

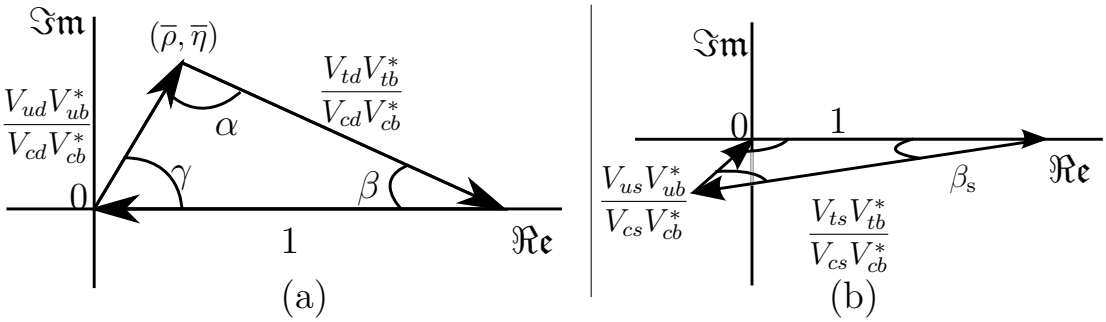


Figure 2.2: *Diagram showing (a) the* B^0 *unitarity triangle and (b) the* B_s^0 *unitarity triangle. These two triangles are not to the same scale, although both have been scaled with one side of length 1 on the real axis. The two triangles can be obtained from the unitarity relations in equation 2.6.*

Area of unitarity triangles

The area of all of the unitarity triangles, before normalisation of the real axis, is defined as the quantity A which is related to the amount of $\mathcal{CP}$ violation in the Standard Model. This area is defined as half of the phase-independent Jarlskog invariant J, in terms of the Euler parameters is,

$$2A = J = c_{12}c_{13}^2c_{23}s_{12}s_{23}\sin\delta = \mathcal{O}\left(10^{-5}\right) \ . \tag{2.8}$$

For the Jarlskog invariant to be non-zero it is required that there must be a global non-zero phase δ within the CKM.

2.4 Types of $\mathcal{CP}$ violation

The three different ways that $\mathcal{CP}$ violation may be experimentally present are (i) differences between two decay processes related by a $\mathcal{CP}$ transform, (ii) differences introduced in mixing between two $\mathcal{CP}$ states and (iii) differences due to the interference between these two processes.

2.4.1 $\mathcal{CP}$ violation through decay

$\mathcal{CP}$ violation in decay presents itself when there is a difference between one decay process and its $\mathcal{CP}$ conjugate. This requires that $\Gamma\left(\mathrm{B} \to f\right) \neq \Gamma\left(\overline{\mathrm{B}} \to \overline{f}\right)$. In this case the $\mathcal{CP}$-violating observable $A_{\mathcal{CP}}$ is defined as,

$$A_{\mathcal{CP}} = \frac{\Gamma\left(\mathrm{B} \to f\right) - \Gamma\left(\overline{\mathrm{B}} \to \overline{f}\right)}{\Gamma\left(\mathrm{B} \to f\right) + \Gamma\left(\overline{\mathrm{B}} \to \overline{f}\right)} \,. \tag{2.9}$$

The total amplitudes representing these processes are,

$$\begin{aligned} A_f &= |a_1|\, e^{i\left(\delta_1 + \phi_1\right)} + |a_2|\, e^{i\left(\delta_2 + \phi_2\right)} + \dots \,, \\ \overline{A}_{\overline{f}} &= |a_1|\, e^{i\left(\delta_1 - \phi_1\right)} + |a_2|\, e^{i\left(\delta_2 - \phi_2\right)} + \dots \,. \end{aligned} \tag{2.10}$$

where the a_i represents distinct contributions to the total decay amplitude. These amplitudes are complex quantities and have associated weak and strong phases $\phi_{1,2}$ and $\delta_{1,2}$ respectively. The total decay amplitude A_f corresponds to the decay of $\mathrm{B} \to f$ and $\overline{A}_{\overline{f}}$ corresponds to $\overline{\mathrm{B}} \to \overline{f}$. From equation 2.10, taking only two possible contributions, the $\mathcal{CP}$-violating observable is described by,

$$A_{\mathcal{CP}} = \frac{-2\,|a_1|\,|a_2| \sin\left(\delta_2 - \delta_1\right) \sin\left(\phi_2 - \phi_1\right)}{|a_1|^2 + |a_2|^2 + 2\,|a_1|\,|a_2| \cos\left(\delta_2 - \delta_1\right) \cos\left(\phi_2 - \phi_1\right)} \,. \tag{2.11}$$

This equation demonstrates that $\mathcal{CP}$ violation in decay is only observable if there are at least two decay amplitudes each with different strong and weak phases. If the distinct contributions in these amplitudes do not have different strong and different weak phases then it is not possible to measure $\mathcal{CP}$ violation. Even when $\mathcal{CP}$ violation is not measurable in decay it may still be present. $\mathcal{CP}$ violation in decay is often referred to as "direct" $\mathcal{CP}$ violation.

2.4.2 $\mathcal{CP}$ violation through mixing

Neutral meson oscillation leads to an additional source of $\mathcal{CP}$ violation through mixing between two states related by $\mathcal{CP}$, e.g. $\mathrm{B_s^0}$ and $\overline{\mathrm{B}}_\mathrm{s}^0$. The time evolution of this system is described in terms of the heavy (H) and light (L) mass eigenstates $\mathrm{B_H}$ and $\mathrm{B_L}$. These are constructed from different contributions of $\mathrm{B_s^0}$ and $\overline{\mathrm{B}}_\mathrm{s}^0$ mesons as described in equation 2.12. Here p and q are complex quantities.

$$\begin{aligned} |\mathrm{B_H}\,(t)\rangle &= p|\mathrm{B_s^0}\,(t)\rangle - q|\overline{\mathrm{B}}_\mathrm{s}^0\,(t)\rangle \ , \\ |\mathrm{B_L}\,(t)\rangle &= p|\mathrm{B_s^0}\,(t)\rangle + q|\overline{\mathrm{B}}_\mathrm{s}^0\,(t)\rangle \ . \end{aligned} \tag{2.12}$$

Details of the time evolution of this system are given in Appendix A.2. The time dependant probabilities for $\mathrm{B_s^0}\left(\overline{\mathrm{B}}_\mathrm{s}^0\right)$ meson to evolve into $\overline{\mathrm{B}}_\mathrm{s}^0(\mathrm{B_s^0})$ state is given in equation 2.13. The time evolution of this system is encapsulated in the function $g_-\,(t)$, defined in Appendix A.2.

$$\left|\left\langle \mathrm{B_s^0}|\overline{\mathrm{B}}_\mathrm{s}^0\,(t)\right\rangle\right|^2 = \left|\frac{p}{q}\right|^2 |g_-\,(t)|^2 \ , \quad \left|\left\langle \overline{\mathrm{B}}_\mathrm{s}^0|\mathrm{B_s^0}\,(t)\right\rangle\right|^2 = \left|\frac{q}{p}\right|^2 |g_-\,(t)|^2 \ . \tag{2.13}$$

The condition for $\mathcal{CP}$ violation in mixing requires that $\Gamma\left(\mathrm{B_s^0} \to \overline{\mathrm{B}}_\mathrm{s}^0\right) \neq \Gamma\left(\overline{\mathrm{B}}_\mathrm{s}^0 \to \mathrm{B_s^0}\right)$ which implies,

$$\left|\frac{q}{p}\right| \neq 1 \neq \left|\frac{p}{q}\right| \ . \tag{2.14}$$

This is often referred to as "indirect" $\mathcal{CP}$ violation. Even in the case of $\left|\frac{q}{p}\right| = 1$ there may still be a $\mathcal{CP}$-violating phase present.

2.4.3 $\mathcal{CP}$ violation through interference

$\mathcal{CP}$ violation also occurs as a result of the interference between the mixing and decay processes in the same channel. An important feature of this is that there can still be $\mathcal{CP}$ violation through interference even if it is not observed either in mixing or decay. This is only possible if the final state $f = X^{\mathcal{CP}}$ is a $\mathcal{CP}$-eigenstate accessible from both $\mathrm{B_s^0}$ and $\overline{\mathrm{B}}_\mathrm{s}^0$ meson decays as shown in Figure 2.3.

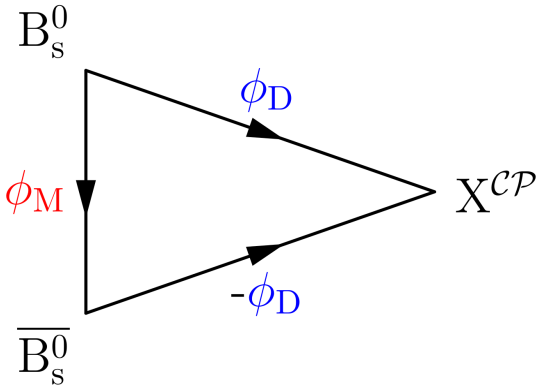


Figure 2.3: *This figure shows how $\mathcal{CP}$ violation can come about due to interference between direct decays and decays after neutral B_s^0 oscillations. The weak phase introduced through the decay of B_s^0, $\bar{B}_s^0$ is ϕ_D, $-\phi_D$ and the strong phase introduced through the mixing of the 2 mesons is ϕ_M.*

This figure shows the two possible decay paths for a B_s^0 meson to a final state $X^{\mathcal{CP}}$. Starting from a B_s^0 meson the decay can proceed directly or it may change flavour into a $\bar{B}_s^0$ meson and then decay. The difference in the phases between these decay paths leads to $\mathcal{CP}$ violation in interference.

$\mathcal{CP}$ violation in interference requires that there must be an unbalanced weak phase between the mixing and decay contributions which gives the definition,

$$\phi_s = \phi_M - 2\phi_D \ . \tag{2.15}$$

This phase ϕ_s can also defined in terms of the decay amplitudes A_f and $\overline{A}_{\bar{f}}$ and the complex parameters p and q as,

$$-\phi_s = \arg\left(\frac{q}{p}\frac{\overline{A}_{\bar{f}}}{A_f}\right) \neq 0 \ . \tag{2.16}$$

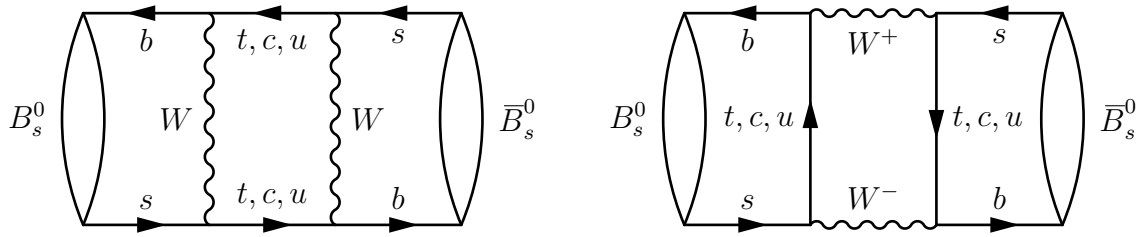


Figure 2.4: *Feynman diagrams for the box mixing diagrams which are responsible for the* $\mathrm{B_s^0} \leftrightarrow \overline{\mathrm{B_s^0}}$ *oscillations.*

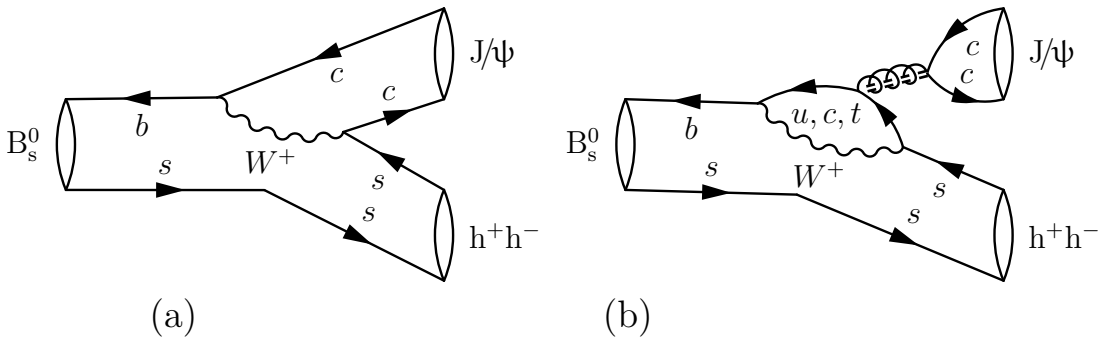


Figure 2.5: *Feynamn diagrams corresponding to the tree-level decay (a) and penguin-level decays (b) within the Standard Model.*

2.5 $B_s^0 \to J/\psi\ K^+K^-$ phenomenology

The decay of $B_s^0 \to J/\psi\ K^+K^-$ mainly proceeds through the resonance of $B_s^0 \to J/\psi\,\phi$ which is a pseudo-scalar to vector-vector decay. This section describes the construction of the differential decay rate and the physical observables that are measured for each event.

The B_s^0 meson is a neutral meson and may oscillate between the different flavour states before decaying into the $J/\psi\ K^+K^-$ daughters. The dominant Feynman diagrams describing the oscillation and decay of the B_s^0 mesons are shown in Figures 2.4 and 2.5.

$\mathcal{CP}$ violation in the $B_s^0 \to J/\psi\ K^+K^-$ decay presents itself in the form of interference between the mixing and decay contributions.

2.5.1 B meson time evolution

Before considering the full $B_s^0 \to J/\psi\ K^+K^-$ decay it is informative to consider the time evolution of the B meson system. Using the definition of the B mesons in equation 2.12 and starting from the evolution of the mass eigenstates in equation 2.17, Appendix A.2, describes the time evolution of the B and $\overline{\mathrm{B}}$ systems.

$$\begin{aligned} |\mathrm{B_H}\,(t)\rangle &= e^{-i\lambda_\mathrm{H} t}|\mathrm{B_H}\,(0)\rangle = e^{-i\mathrm{M_H}t-\frac{\Gamma_\mathrm{H}}{2}t}|\mathrm{B_H}\,(0)\rangle\ , \\ |\mathrm{B_L}\,(t)\rangle &= e^{-i\lambda_\mathrm{L} t}|\mathrm{B_L}\,(0)\rangle\ = e^{-i\mathrm{M_L}t-\frac{\Gamma_\mathrm{L}}{2}t}|\mathrm{B_L}\,(0)\rangle\ . \end{aligned} \tag{2.17}$$

From this formalism, in the case that there is only one final state, the decay of a B meson to the final state f is described by the differential decay rate,

$$\frac{d\Gamma\,(\mathrm{B}\to f)}{dt} = Ne^{-\overline{\Gamma}\,t}\left\{\begin{array}{l} a\ \cosh\left(\frac{\Delta\Gamma}{2}\,t\right) + b\ \sinh\left(\frac{\Delta\Gamma}{2}\,t\right) \\ c\ \cos\left(\Delta m\,t\right)\ + d\ \sin\left(\Delta m\,t\right) \end{array}\right\} = h\left(t\right)\ . \tag{2.18}$$

Where N is 1 in this case. Here, the difference in mass between the eigenstates is defined as $\Delta m = \mathrm{M_H} - \mathrm{M_L}$. The definitions of $\overline{\Gamma} = \frac{\Gamma_\mathrm{L}+\Gamma_\mathrm{H}}{2}$, and $\Delta\Gamma = \Gamma_\mathrm{L} - \Gamma_\mathrm{H}$, are also introduced where $\overline{\Gamma}$ is the average decay width and $\Delta\Gamma$ is the decay width splitting. The multiplication factors a, b, c and d vary depending on the configuration of the final state.

2.5.2 Full $B_s^0 \to J/\psi\ K^+K^-$ differential decay rate

In the case of $B_s^0 \to J/\psi\ K^+K^-$ there are four final states with different angular configurations. Three of these states originate from the P-wave $B_s^0 \to J/\psi\,\phi$ channel (where the K^+K^- proceed via the ϕ resonance) and the fourth corresponds to a K^+K^- S-wave contribution. The three amplitudes corresponding to the P-wave are A_0, $A_\parallel$ and $A_\perp$. These final states correspond to different $\mathcal{CP}$ configurations with the polarisation vectors of J/ψ and ϕ aligned as in Figure 2.6. The amplitude corresponding to the S-wave is defined as A_S.

The first P-wave amplitude A_0 is a $\mathcal{CP}$-even state corresponding to the configuration where the two daughters are longitudinally polarised. The other two P-wave final states ($A_\parallel$ and $A_\perp$) are transversally polarised. The $A_\perp$ amplitude is a $\mathcal{CP}$-odd amplitude which corresponds to the case where the daughters are aligned perpendicular to each other. The final P-wave amplitude $A_\parallel$ is $\mathcal{CP}$-even state with the daughters are aligned parallel to each other. The S-wave amplitude has a $\mathcal{CP}$-odd configuration similar to the $A_\perp$ P-wave state. Each of these decay amplitudes has an associated complex strong phase. These are δ_0, $\delta_\perp$, $\delta_\parallel$ and δ_S, for A_0, $A_\perp$, $A_\parallel$ and A_S respectively.

To separate these amplitudes the angular distribution of the final decay products, $\vec{\Omega}$, needs to be described. The choice of angular basis and the angular definitions are defined in Section 2.5.3.

Due to the fact that $B_s^0 \to J/\psi\ K^+K^-$ has four possible decay paths the differential decay rate for this channel is described by ten terms, four corresponding to the decay amplitudes and six describing the interference between them,

$$\frac{d\Gamma_{(B_s^0 \to J/\psi\ K^+K^-)}\left(t, \vec{\Omega}\right)}{dt\, d\vec{\Omega}} = \sum_k^{10} h_k\left(t\right) f_k\left(\vec{\Omega}\right) \ . \tag{2.19}$$

The full form of this differential decay rate for all ten terms are given in Tables 2.2 and 2.3 for the $f_k\left(\vec{\Omega}\right)$ and $h_k\left(t\right)$ terms respectively.

k	$f_k\left(\vec{\Omega}=\{\theta_{tr},\psi_{tr},\phi_{tr}\}\right)$	$g_k\left(\vec{\Omega}=\{\theta_{\mu},\theta_K,\phi_h\}\right)$
1	$2\cos^2\psi_{tr}\left(1-\sin^2\theta_{tr}\cos^2\phi_{tr}\right)$	$2\cos^2\theta_K\sin^2\theta_{\mu}$
2	$\sin^2\psi_{tr}\left(1-\sin^2\theta_{tr}\sin^2\phi_{tr}\right)$	$\sin^2\theta_K\left(1-\sin^2\theta_{\mu}\cos^2\phi_h\right)$
3	$\sin^2\psi_{tr}\sin^2\theta_{tr}$	$\sin^2\theta_K\left(1-\sin^2\theta_{\mu}\sin^2\phi_h\right)$
4	$-\sin^2\psi_{tr}\sin 2\theta_{tr}\sin\phi_{tr}$	$\sin^2\theta_K\sin^2\theta_{\mu}\sin 2\phi_h$
5	$\frac{1}{2}\sqrt{2}\sin 2\psi_{tr}\sin^2\theta_{tr}\sin 2\phi_{tr}$	$\frac{1}{2}\sqrt{2}\sin 2\theta_K\sin 2\theta_{\mu}\cos\phi_h$
6	$\frac{1}{2}\sqrt{2}\sin 2\psi_{tr}\sin 2\theta_{tr}\cos\phi_{tr}$	$-\frac{1}{2}\sqrt{2}\sin 2\theta_K\sin 2\theta_{\mu}\sin\phi_h$
7	$\frac{2}{3}\left(1-\sin^2\theta_{tr}\cos^2\phi_{tr}\right)$	$\frac{2}{3}\sin^2\theta_{\mu}$
8	$\frac{1}{3}\sqrt{6}\sin\psi_{tr}\sin^2\theta_{tr}\sin 2\phi_{tr}$	$\frac{1}{3}\sqrt{6}\sin\theta_K\sin 2\theta_{\mu}\cos\phi_h$
9	$\frac{1}{3}\sqrt{6}\sin\psi_{tr}\sin 2\theta_{tr}\cos\phi_{tr}$	$-\frac{1}{3}\sqrt{6}\sin\theta_K\sin 2\theta_{\mu}\sin\phi_h$
10	$\frac{4}{3}\sqrt{3}\cos\psi_{tr}\left(1-\sin^2\theta_{tr}\cos^2\phi_{tr}\right)$	$\frac{4}{3}\sqrt{3}\cos\theta_K\sin^2\theta_{\mu}$

Table 2.2: *The angular dependant* $\mathrm{B_s^0}\rightarrow \mathrm{J/\psi\ K^+K^-}$ *decay rate coefficients used in fitting to data.* $f_k\left(\vec{\Omega}\right)$ *describes the transversity basis and* $g_k\left(\vec{\Omega}\right)$ *the helicity basis. Here* k *corresponds to* k*th term in the differential decay rate, equation 2.19.*

k	N_k	a_k	b_k	c_k	d_k
1	$\left\|\mathrm{A_0}\left(0\right)\right\|^2$	1	D	C	−S
2	$\left\|\mathrm{A_{\parallel}}\left(0\right)\right\|^2$	1	D	C	−S
3	$\left\|\mathrm{A_{\perp}}(0)\right\|^2$	1	−D	C	S
4	$\left\|\mathrm{A_{\parallel}}\left(0\right)\mathrm{A_{\perp}}(0)\right\|$	$\mathrm{C}\sin\left(\delta_{\perp}-\delta_{\parallel}\right)$	$\mathrm{S}\cos\left(\delta_{\perp}-\delta_{\parallel}\right)$	$\sin\left(\delta_{\perp}-\delta_{\parallel}\right)$	$\mathrm{D}\cos\left(\delta_{\perp}-\delta_{\parallel}\right)$
5	$\left\|\mathrm{A_0}\left(0\right)\mathrm{A_{\parallel}}\left(0\right)\right\|$	$\cos\left(\delta_{\parallel}-\delta_0\right)$	$\mathrm{D}\cos\left(\delta_{\parallel}-\delta_0\right)$	$\mathrm{C}\cos\left(\delta_{\parallel}-\delta_0\right)$	$-\mathrm{S}\cos\left(\delta_{\parallel}-\delta_0\right)$
6	$\left\|\mathrm{A_0}\left(0\right)\mathrm{A_{\perp}}(0)\right\|$	$\mathrm{C}\sin\left(\delta_{\perp}-\delta_0\right)$	$\mathrm{S}\cos\left(\delta_{\perp}-\delta_0\right)$	$\sin\left(\delta_{\perp}-\delta_0\right)$	$\mathrm{D}\cos\left(\delta_{\perp}-\delta_0\right)$
7	$\left\|\mathrm{A_S}\left(0\right)\right\|^2$	1	−D	C	S
8	$\left\|\mathrm{A_S}\left(0\right)\mathrm{A_{\parallel}}\left(0\right)\right\|$	$\mathrm{C}\cos\left(\delta_{\parallel}-\delta_{\mathrm{S}}\right)$	$\mathrm{S}\sin\left(\delta_{\parallel}-\delta_{\mathrm{S}}\right)$	$\cos\left(\delta_{\parallel}-\delta_{\mathrm{S}}\right)$	$\mathrm{D}\sin\left(\delta_{\parallel}-\delta_{\mathrm{S}}\right)$
9	$\left\|\mathrm{A_S}\left(0\right)\mathrm{A_{\perp}}(0)\right\|$	$\sin\left(\delta_{\perp}-\delta_{\mathrm{S}}\right)$	$-\mathrm{D}\sin\left(\delta_{\perp}-\delta_{\mathrm{S}}\right)$	$\mathrm{C}\sin\left(\delta_{\perp}-\delta_{\mathrm{S}}\right)$	$\mathrm{S}\sin\left(\delta_{\perp}-\delta_{\mathrm{S}}\right)$
10	$\left\|\mathrm{A_S}\left(0\right)\mathrm{A_0}\left(0\right)\right\|$	$\mathrm{C}\cos\left(\delta_0-\delta_{\mathrm{S}}\right)$	$\mathrm{S}\sin\left(\delta_0-\delta_{\mathrm{S}}\right)$	$\cos\left(\delta_0-\delta_{\mathrm{S}}\right)$	$\mathrm{D}\sin\left(\delta_0-\delta_{\mathrm{S}}\right)$

Table 2.3: *The time dependant* $\mathrm{B_s^0}\rightarrow \mathrm{J/\psi\ K^+K^-}$ *decay rate used in fitting to data.*
Using: $\mathrm{C}=\dfrac{1-\left|\lambda_s\right|^2}{1+\left|\lambda_s\right|^2}$, $\mathrm{D}=\dfrac{-2\,\Re\mathfrak{e}\left(\lambda_s\right)}{1+\left|\lambda_s\right|^2}$, $\mathrm{S}=\dfrac{2\,\Im\mathfrak{m}(\lambda_s)}{1+\left|\lambda_s\right|^2}$, $\left|\lambda_s\right|=\left|\dfrac{p}{q}\dfrac{\overline{A}}{A}\right|$.
Here k *corresponds to* k*th term in the differential decay rate, equation 2.19.*

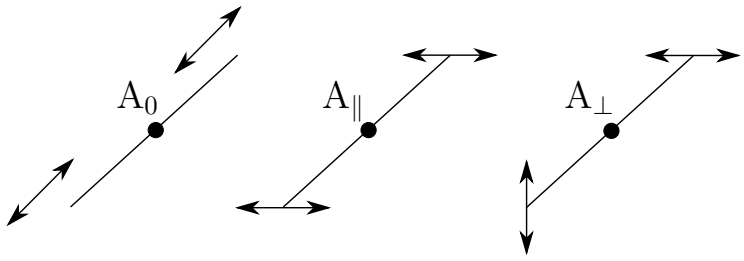


Figure 2.6: *Relative polarisation of P-wave final states.* A_0*: longitudinal polarisation,* $A_\parallel$*: parallel transverse polarisation,* $A_\perp$*: perpendicular transverse polarisation.*

For each of the possible amplitude configurations the time dependence is described by equation 2.18 with the full structure of all ten terms given in Table 2.3. Each term $h_k(t)$ describes the time evolution of the kth decay amplitude. The phases (δ_i) appearing in Table 2.3 are due to the strong phases of the different decay amplitudes as described in Section 2.5.3.

In Table 2.3 the parameter λ_s is the term containing all of the $\mathcal{CP}$ violation, with its complex phase given as $-\phi_s$. Using this notation the imaginary component of this is defined as $\Im\mathfrak{m}(\lambda_s) = |\lambda_s| \sin(-\phi_s)$ and the real component is $\Re\mathfrak{e}(\lambda_s) = |\lambda_s| \cos(-\phi_s)$.

The corresponding differential decay rate for $\overline{\mathrm{B}_\mathrm{s}^0} \to \mathrm{J/\psi\ K^+K^-}$ can be derived from equation 2.18 by changing the sign of terms c and d as described in Appendix A.2.

2.5.3 Choice of angular basis

The measurements of any underlying physics parameters are independent of the choice of angular basis used in measuring the distribution. Nevertheless, there are traditionally two different angular bases used for describing the data. These are the helicity and transversity bases as shown in Figure 2.7 (a) and (b) respectively.

The helicity basis defines two separate z-axes in the $\mathrm{J/\psi}$ and $\mathrm{K^+K^-}$ rest frames and defines the angles of θ_μ and θ_K between each z axis and the flight direction of the positively charged daughter. The angle ϕ_H is defined as the angle between the two planes of the di-muon and di-kaon system.

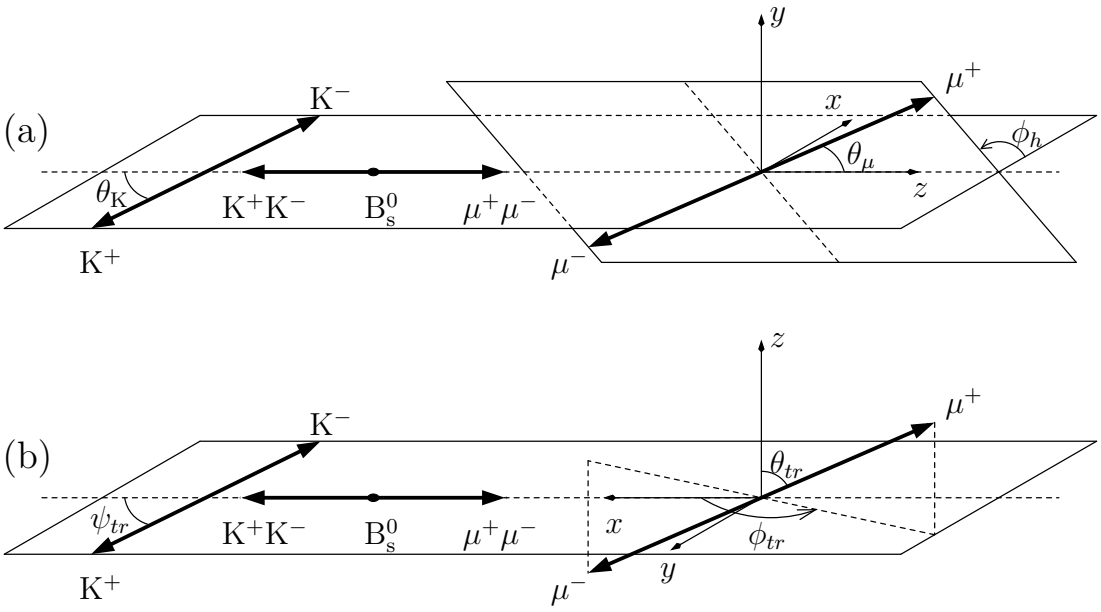


Figure 2.7: *Angular basis used within the* $B_s^0 \to J/\psi\ K^+K^-$ *analysis. (a) Helicity basis, (b) Transversity Basis.*

The transversity basis defines the angles θ_{tr} and ϕ_{tr} in terms of μ^+ in the J/ψ rest frame and ψ_{tr} is defined using K^+ in the in the K^+K^- rest frame. The full derivation of the helicity basis using the momenta of the decay particles is given in Ref [12].

2.6 Standard Model predictions

As shown in Figure 1.1 ϕ_s is known to be correlated with $\Delta\Gamma_s$ from previous analyses. The Standard Model predictions for these 2 values are summarised within this section.

2.6.1 ϕ_s in the Standard Model

In $B_s^0 \to J/\psi\, K^+K^-$ the angle ϕ_s is the $\mathcal{CP}$-violating phase introduced through the interference between decay and mixing as described in Section 2.4.3. The value of ϕ_s is therefore the sum of the weak phases introduced in mixing $\left(\frac{q}{p}\right)$ and decay $\left(\frac{\overline{A}}{A}\right)$,

$$-\phi_s = \arg\left(\lambda_s\right) = \arg\left\{\frac{q}{p}\frac{\overline{A}}{A}\right\} \ . \tag{2.20}$$

$\mathcal{CP}$ violation in mixing of $B_s^0 \to J/\psi\, K^+K^-$

The time evolution of the B_s^0 system is described through the Matrix ($\mathbf{R}$),

$$\mathbf{R} = \mathbf{M} - i\frac{\mathbf{\Gamma}}{2} = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{21} & \mathcal{M}_{22} \end{bmatrix} - \frac{i}{2}\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \ . \tag{2.21}$$

Considering the evolution of the heavy and light mass eigenstates as given in equation 2.12, the ratio between these two states (q/p) is given in equation 2.22.

$$\frac{q}{p} = \sqrt{\frac{\mathcal{M}_{12}^* - \frac{i}{2}\Gamma_{12}^*}{\mathcal{M}_{12} - \frac{i}{2}\Gamma_{12}}} = e^{-\phi_M} \ . \tag{2.22}$$

This solution is shown more completely in Appendix A. In the Standard Model the approximation $|\Gamma_{12}| \ll |\mathcal{M}_{12}|$ is valid, and hence $\frac{q}{p} \approx \mathcal{M}_{12}$.

The Feynman diagrams associated with neutral B_s^0 mixing in $B_s^0 \to J/\psi\, K^+K^-$ is shown in Figure 2.4. The main internal quark contribution to this diagram is a top quark and as such the transitions involved in mixing are described by the V_{tb} and V_{ts} elements of the CKM matrix 2.5.

The mixing phase (ϕ_M) introduced through $\mathcal{M}_{12}$ and is the phase corresponding to the leading virtual top contribution to the box diagram Figure 2.4. Hence the phase introduced through decay is given as:

$$\arg\left(\frac{q}{p}\right) = \phi_M = 2\left(V_{ts}V_{tb}^*\right) \ . \tag{2.23}$$

$\mathcal{CP}$ violation in decay in $B_s^0 \to J/\psi\, K^+K^-$

The tree level amplitude from Figure 2.5 (a) is taken to be A_T and the contribution from each penguin diagram in Figure 2.5 (b) is P_q where q is the internal u, c or t quark. Using this notation, and making use of the β_s unitarity relation 2.6 the total amplitude of the flavour changing decay can be defined as:

$$\begin{aligned} A\left(\overline{b} \to \overline{c}c\overline{s}\right) &= V_{cs}V_{cb}^*\left(A_T + P_c\right) + V_{us}V_{ub}^* P_u + V_{ts}V_{tb}^* P_t \\ &= V_{cs}V_{cb}^*\left(A_T + P_c - P_t\right) + V_{us}V_{ub}^*\left(P_u - P_t\right) \\ &= V_{cs}V_{cb}^*\left(A_T + P_c - P_t\right) + \mathcal{O}\left(\lambda^4\right) \ . \end{aligned} \tag{2.24}$$

The dominant contribution to this amplitude is the tree level decay A_T in Figure 2.5. This decay includes contributions from the V_{cs} and V_{cb} elements in the CKM matrix 2.5. Ignoring penguin terms suppressed by λ^4, $\dfrac{\overline{A}}{A}$ is approximated to be 2.25.

$$\begin{aligned} \frac{\overline{A}}{A} &= -\frac{|V_{cb}V_{cs}^*|\, e^{i\arg(V_{cb}V_{cs}^*)}}{|V_{cs}V_{cb}^*|\, e^{-i\arg(V_{cb}V_{cs}^*)}} = -\frac{|V_{cb}V_{cs}^*|}{|V_{cs}V_{cb}^*|} e^{2i\arg\left(V_{cb}V_{cs}^*\right)} \ . \\ \arg\left[\frac{\overline{A}}{A}\right] &= 2\arg\left(V_{cb}V_{cs}^*\right) = 2\phi_D \ . \end{aligned} \tag{2.25}$$

Standard Model value of $\phi_s^{\left(J/\psi\, K^+K^-\right)}$

From equations 2.23 and 2.25 it follows that the Standard Model expectation for the phase $\phi_s\left(J/\psi\ K^+K^-\right)$ is given in equation 2.26.

$$\begin{aligned} \lambda_s\left(J/\psi\ K^+K^-\right) &= \left|\lambda_s\left(J/\psi\ K^+K^-\right)\right| e^{-i\phi_s\left(J/\psi\ K^+K^-\right)} = \left|\frac{q}{p}\frac{\overline{A}}{A}\right| e^{-i\phi_M} e^{2i\phi_D} \ . \\ \phi_s\left(J/\psi\ K^+K^-\right) &= \left(\phi_M - 2\phi_D\right) \\ &= 2\arg\left(V_{ts}V_{tb}^*\right) - 2\arg\left(V_{cs}V_{cb}^*\right) \\ &= 2\arg\left[\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right] \ , \quad \beta_s = \arg\left\{-\left[\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right]\right\} \ , \\ \phi_s\left(J/\psi\ K^+K^-\right) &= -2\beta_s \ . \end{aligned} \tag{2.26}$$

From a global fit to the CKM matrix, this the Standard Model value of ϕ_s is calculated to be $\phi_s^{SM} = -0.036 \pm 0.002\,\text{rad}$ [19].

2.6.2 $\Delta\Gamma_s$ in the Standard Model

In $B_s^0 \to J/\psi\, K^+K^-$ final state is composed of $\mathcal{CP}$-even and $\mathcal{CP}$-odd components which have different lifetimes Γ_L and Γ_H separated by the lifetime difference $\Delta\Gamma_s = \Gamma_L - \Gamma_H$.

The time evolution of the B_s^0 meson is described using the matrix ($\mathbf{R}$), given in equation 2.21. Taking the eigenvectors of this matrix as described in Appendix A shows the splitting between these 2 states to be described by,

$$\Delta M + \frac{i}{2}\Delta\Gamma = 2\sqrt{\left(\mathcal{M}_{12} - \frac{i}{2}\Gamma_{12}\right)\left(\mathcal{M}_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}\,. \tag{2.27}$$

From this definition of the splitting and under the assumption of no $\mathcal{CP}$ violation, $\Delta\Gamma_s$ can be approximated to be,

$$\Delta\Gamma_s = -2\Gamma_{12}\,. \tag{2.28}$$

The value of $\Delta\Gamma_s$ can be calculated using heavy quark expansion theory (HQET) which is described more in Ref [20]. Using this approach the Standard Model prediction is, $\Delta\Gamma_s{}^{SM} = 0.082 \pm 0.021\,\mathrm{ps}^{-1}$ [21–23].

"For a long period of time there was much speculation and controversy about where the so-called 'missing matter' of the Universe had got to . . . but when eventually it was tracked down it turned out in fact to be all the stuff which the equipment had been packed in."

Mostly Harmless – Douglas Adams

3

The LHCb Detector

The LHCb experiment is one of the four major experiments performing physics measurements at the LHC accelerator in CERN, as outlined in Section 3.1. The detector is designed to make detailed measurements of $\mathrm{B}(\overline{\mathrm{B}})$ mesons.

The detector is briefly described in Section 3.2 and more detail can be found in Ref [24]. It has been designed such that it has the ability to track particles with high accuracy. This allows for the correct identification of the flavour of each particle.

The LHCb detector is ideally suited for measuring the decay channels of interest in this thesis. The B_s^0 meson has a relatively large flight distance that can be accurately measured and the $\mathrm{J}/\psi\,(\to \mu^+\mu^-)\,\mathrm{K}^+\mathrm{K}^-$ final state is composed of charged particles that can be accurately tracked through the LHCb detector as described in Section 3.3. The experiment is also able to accurately identify particles using information from various sub detectors as described in Section 3.4.

The selection of data by the LHCb trigger system is outlined in Section 3.5, and the software stack used for data reconstruction and analysis within the LHCb experiment is summarised in Section 3.6.

3.1 The LHC

The Large Hadron Collider (LHC), located across the Swiss-French border near Geneva, is the world's largest and highest energy particle accelerator. The circular tunnel housing the LHC is 27 km in diameter and approximately 100 m underground.

The protons supplied to the LHC originate from a source of pure Hydrogen which is ionised and then passed through a series of progressively higher energy accelerators. The series of accelerators traversed by the protons is: LINAC2 [50MeV] $\rightarrow$ Proton Synchrotron Booster (BOOSTER) [1.4GeV] $\rightarrow$ Proton Synchrotron (PS) [25GeV] $\rightarrow$ Super Proton Synchrotron (SPS) [450GeV]. Figure 3.1 show a digram of these different accelerators. Two beams are injected from the SPS into the main LHC ring, one carrying protons in a clockwise direction, and the other counter-clockwise. The LHC accelerator uses 400 MHz Radio-Frequency (RF) cavities to accelerate protons and is designed to reach a centre of mass energy of 14 TeV.

During 2011 the LHC machine delivered proton-proton collisions with a centre of mass energy of $\sqrt{s} = 7\,\text{TeV}$, with a luminosity of the order of $10^{33}\,\text{cm}^{-2}\text{s}^{-1}$.
To direct the proton beams around the circular accelerator 1,232 super-conducting dipole magnets are used each capable of generating a field of approximately 8.3 T.

Around the LHC there are four major experiments situated at different interaction points. These are ALICE[25], ATLAS[26], CMS[27] and LHCb [24]. The position of these four experiments on the LHC ring is shown in Figure 3.1.
The ATLAS and CMS experiments are general purpose detectors designed primarily for the discovery of new particles. ALICE is an experiment specialising in heavy ion collisions and LHCb is designed to make precision measurements of B mesons.

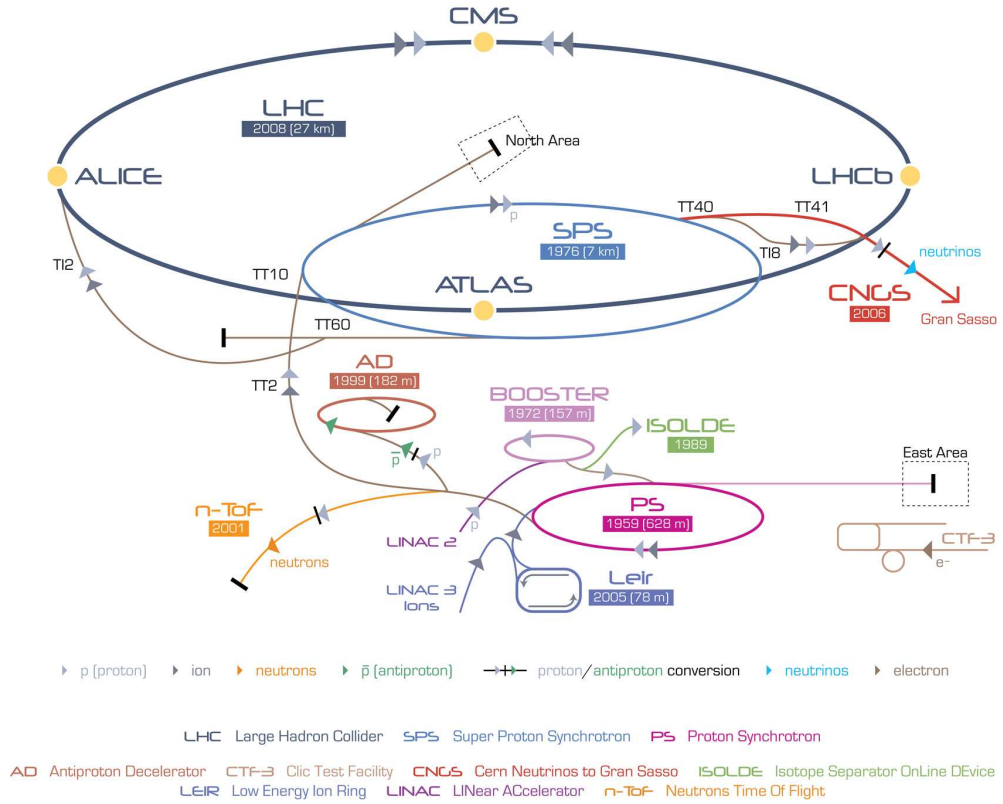


Figure 3.1: *Diagram outlining the accelerators at CERN (not to scale) [28].*

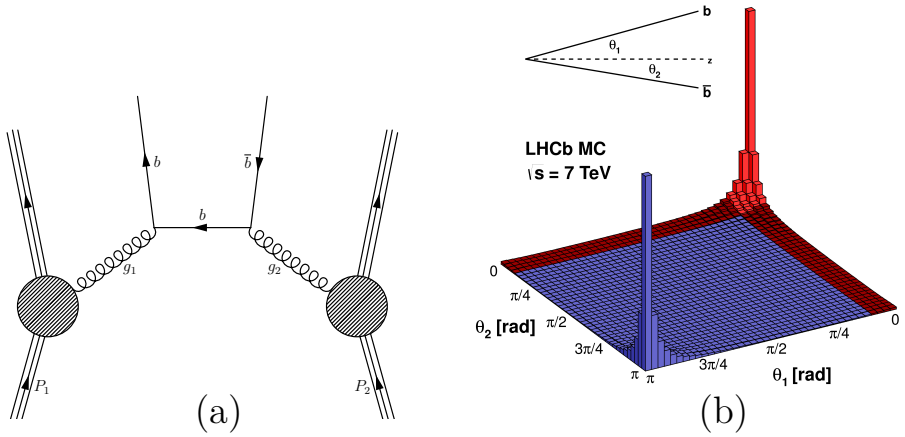


Figure 3.2: *(a) Dominant Feynman diagram responsible for the production of* $B/\overline{B}$ *mesons at the LHC involving gluons (*g_1 *and* g_2*) radiated from two colliding protons (*P_1 *and* P_2*).*
(b) Distribution of the $b/\bar{b}$ *quark pairs produced according to* Pythia *[29]. The red region indicates where both quarks are produced in the same forward direction. Quark pairs produced in a back to back configuration is dis-favoured [30].*

3.1.1 B meson production at the LHC

The main production process for B mesons at the LHC is through gluon-gluon fusion as shown in Figure 3.2 (a). The B mesons are produced such that they are highly boosted along the direction of the beam line. The total $b/\bar{b}$ production cross section has been measured to be $\sigma\left(pp \to b\bar{b}X\right) = 284 \pm 20 \pm 49\mu\text{b}$ at 7 TeV [31].

The distribution of $b/\bar{b}$ quark pairs as calculated by PYTHIA [29] is shown in Figure 3.2 (b). The direction favoured for producing these quark-pairs is in the forward/backward region with both quarks boosted in the same direction from the interaction point.

3.2 The LHCb experiment

The LHCb detector is a forward arm spectrometer. A cross section through the LHCb detector is shown in Figure 3.3, highlighting the position of important sub-detectors. Proton beams from the LHC collide at the interaction point on the left of the figure within the VELO sub-detector. The beam pipe can be seen running along the length of the detector.

The coordinate system of the experiment is defined such that positive z is along the beamline, positive y is in the upward direction and positive x points toward the centre of the LHC ring, in a right-handed coordinate system. The acceptance of the LHCb detector in the y-z plane extends to 250 mrad and in the x-z plane it extends to 300 mrad. The dipole magnet in the LHCb experiment deflects particles in the x-z plane.

The VELO and Tracking sub-detectors are important for tracking charged particles and are described in Section 3.3. To identify the individual particles passing through the LHCb detector information is combined from the RICH sub-detectors, the calorimeters and the muon stations as described in Section 3.4.

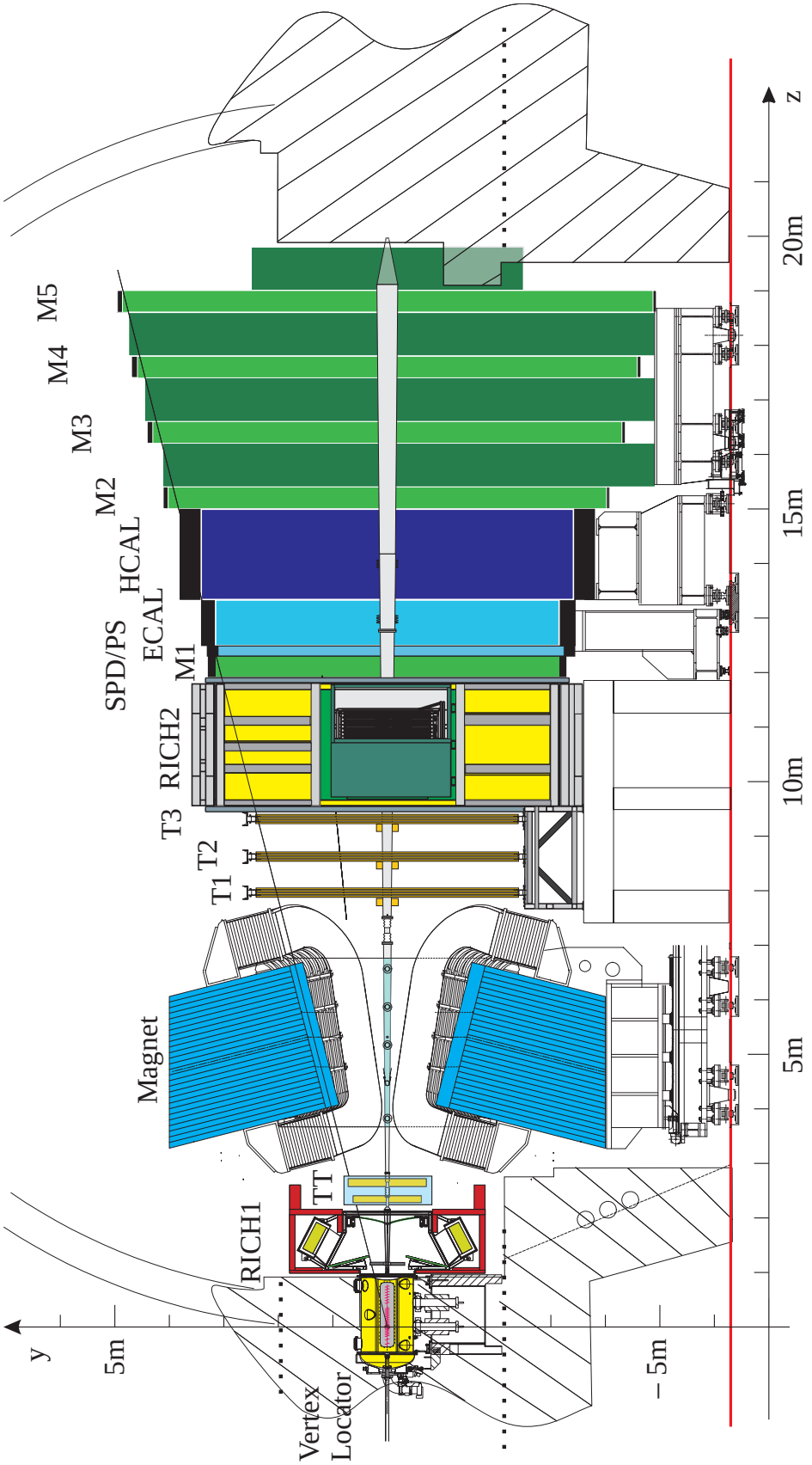


Figure 3.3: *Layout of the LHCb detector [32]. The y-z view is shown.*

3.3 Particle tracking

Reconstructing the trajectories of charged particles allows their momentum and decay vertices to be accurately determined. Particle tracking is performed using two sub-detectors in the LHCb detector. These are, the VELO, which is discussed in Section 3.3.1 and the four Tracking Stations, described in Section 3.3.2. Charged particles are deflected in the bending plane of the magnetic field allowing their curvature q/p where q is the charge and p the momentum to be determined. The dipole magnet generating this field is described in Section 3.3.3.

3.3.1 The VELO

The VErtex LOcator (VELO) is a critical sub-detector to the LHCb experiment. Its primary purpose is to measure the position of decay vertices with high accuracy. The VELO is constructed of two halves centred around the interaction point of the two proton beams as in Figure 3.3. At the injection energy of 450 GeV the beam is defocused. To protect the VELO from radiation damage due to particles in the beam the two halves are designed to open and close around the interaction point. When the proton beams are focussed and in collision the VELO is closed to a distance of 8 mm from the interaction point. When not collecting data the two halves are able to open to 3 cm from the interaction point [24].

The VELO is composed of a series of 21 modules that measure the position of charged particles as they pass through the active silicon sensor. Each module has two types of sensors, the first has silicon strips arranged in radial (r) paths, the second has strips with fixed axial (ϕ) paths. These sensors are orientated perpendicular to the beam line. The strip pitch is such that it is finest close to the interaction point in the region of the highest radiation flux and widens with increasing radius.

Within the VELO the modules are arranged as shown in Figure 3.4. This shows that the majority of the modules are closely spaced around the location of the interaction point. The layout of the modules has been designed such that particles within an acceptance of < 390 mrad will pass through at least six modules (three sensors). This is a larger acceptance than the rest of the LHCb detector.

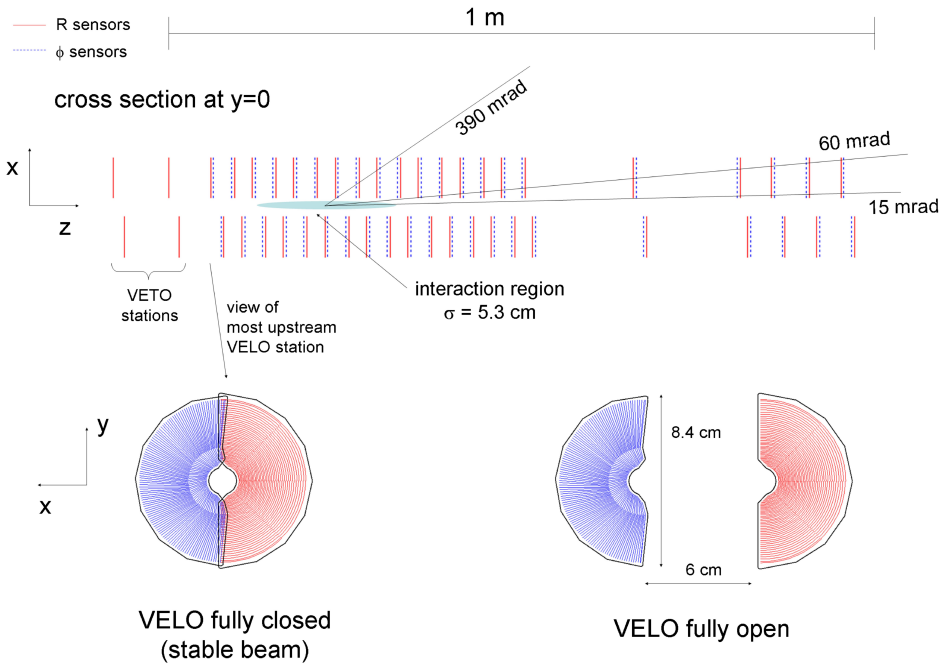


Figure 3.4: *A view of the VELO geometry showing the positions of the different r-ϕ sensors [24].*

In order to be close to the interaction point the VELO is integrated into the beam-pipe running through the experiment. To isolate the VELO from the vacuum of the beam-pipe the VELO modules are encased within a corrugated foil box. This foil protects the VELO from picking up radio frequency signals from the beams.

3.3.2 Tracking stations

The tracking system consists of four stations, the Tracker Turicensis (TT) and the three T-Stations as shown in Figure 3.3. The TT is a silicon tracker located after RICH1 and before the LHCb magnet. The T-Stations (T1, T2 and T3) are located downstream of the LHCb magnet and are composed of two detector technologies. The first of these is the silicon Inner Tracker (IT) which is placed around the beam pipe in regions of highest radiation flux and the second is the drift-chamber Outer Tracker (OT).

The silicon sensors in the TT and IT are able to measure the position of charged particles in a similar way to the VELO (Section 3.3.1). These silicon sensors have a high granularity and radiation hardness which is required in the region of high radiation flux around the beam pipe. The drift-chamber sensors in the OT measure the position of charged particles through gas-ionization. The OT has a larger granularity covering the rest of the LHCb acceptance around the IT.

Individual trackers are composed of four layers, the outer layers are referred to as the x layers and the middle two layers are referred to as u and v layers. The x layers are orientated vertically, parallel to the y axis. The u and v layers are rotated about the z axis of the beam pipe in stereo angles of $+5^\circ$ and -5° as shown for the v layer of the TT in Figure 3.5 (a).

Tracker Turicensis and Inner Tracker

Within the TT the xu and vx layers are separated by 27 cm to improve the transverse momentum (p_T) resolution for use in the trigger, whilst in the IT this separation is 4 cm. Sensors within the TT and IT are arranged as shown in Figures 3.5 (a) and (b). The IT detectors present in T1, T2 and T3 are composed of four boxes around the LHCb beam-pipe. The IT trackers are positioned in front of each of the OT tracking stations.

Outer Tracker

The OT is a drift chamber detector and is composed of staggered straw tubes made a Kapton foil and with an Aluminium coating. The straw-tubes are grounded and act as the cathode which encloses an anode wire held at high voltage. Inside the straw tubes there is a a mixture of Ar (70%) and CO_2 (30%) gas which has a drift time of 40 ns. The relative position and size of the modules within the OT trackers are shown in Figure 3.5 (c).

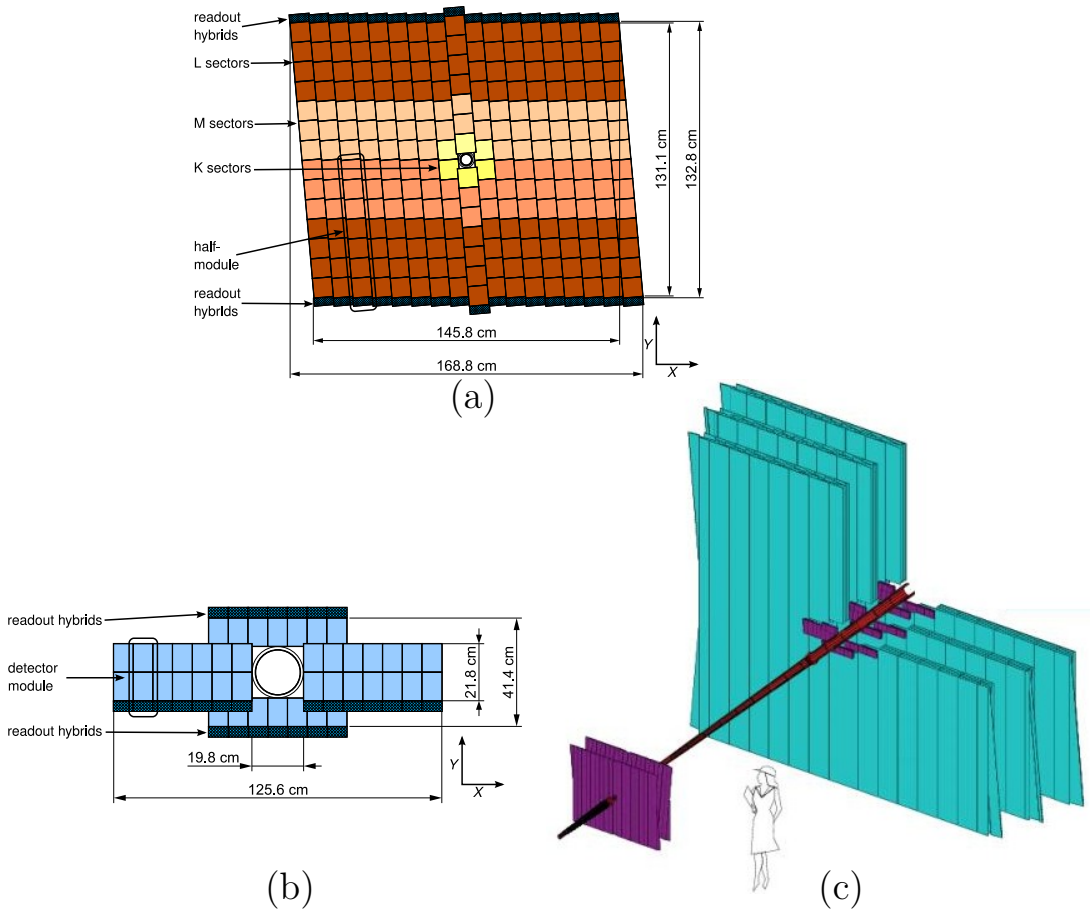


Figure 3.5: *(a) Layout of sensors in the v layer of the TT sub-detector. [24]*
(b) Layout of sensors in the IT sub-detector. [24]
(c) Layout of the Tracker sub-detectors. [24]

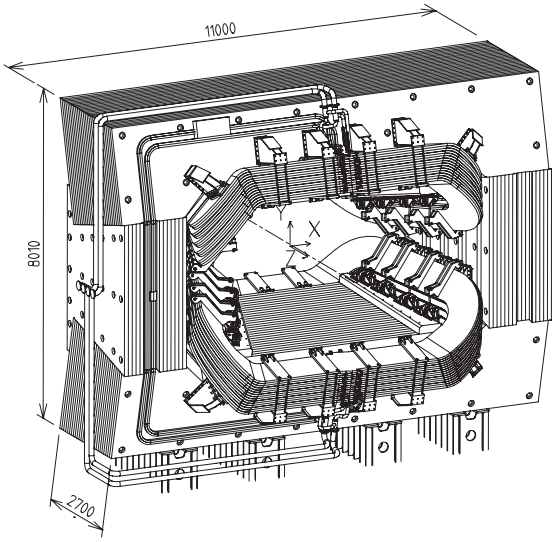


Figure 3.6: *The LHCb Magnet [24].*

3.3.3 The magnet

The LHCb detector includes a water cooled, warm dipole magnet between the TT and T1 stations as shown in Figure 3.3. This has an integrated $\int B\,dl$ of 4 Tm for particles originating from the interaction point and traversing the entire spectrometer to $z = 10\,\text{m}$. The magnet is composed of two large aluminium coils supported by a large iron yoke as in Figure 3.6. During data taking the polarity of the magnet is periodically reversed to minimise detector asymmetries.

3.4 Particle identification

An important requirement of the LHCb experiment is to be able to correctly identify the different types of particles. Information is used from the tracking system (Section 3.3.2) and the RICH sub-detector (Section 3.4.1) to determine the particle mass. The RICH sub-detector provides information which constrains the particle velocity using Cherenkov radiation. The calorimeters in (Section 3.4.2) provide information to correctly identify electrons, photons and hadrons. Finally, hits in the muon chambers (Section 3.4.3), identify high-energy muons.

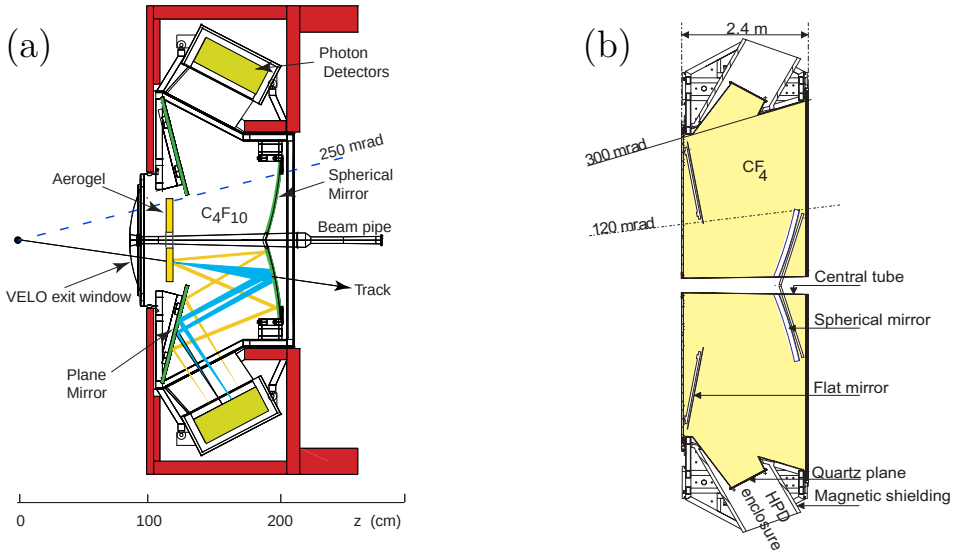


Figure 3.7: *Schematics of (a) RICH1 [24] and (b) RICH2 [24] sub-detectors. The view of the RICH1 sub-detector is side-on, whilst the view of the RICH2 sub-detector is top-down.*

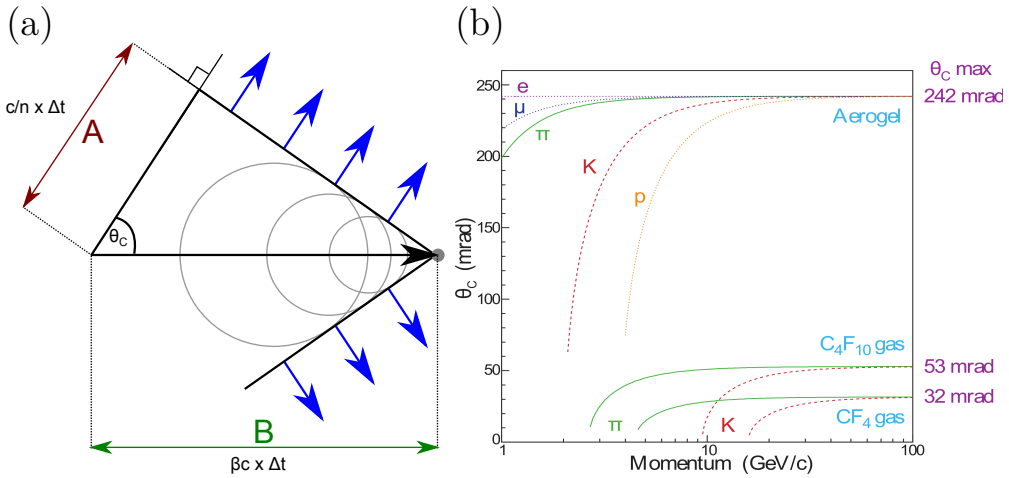


Figure 3.8: *(a) Definition of the Cherenkov angle θ_C for a particle with velocity v, ($\beta = \frac{v}{c}$) passing through a radiator with refractive index n. (b) Variation of the measured angle θ_C for different particle types with varying energies within the RICH sub-detectors at LHCb [24].*

3.4.1 The RICH system

LHCb incorporates two separate Ring Imaging CHerenkov (RICH) sub-detectors, RICH1 and RICH2. These are designed to measure the Cherenkov photons emitted by charged particles passing through a radiating material with a velocity greater than the local speed of light in that medium. The structure of the two sub-detectors is shown in Figure 3.7. The photon detectors used in both RICH detectors are pixelated Hybrid Photon Detectors (HPDs) that combine photo-cathode and silicon sensor technologies.

Cherenkov light emission is similar to a sonic shock-wave as shown in Figure 3.8 (a). The radiation forms of a ring of photons measured by the photo-detectors, from this the angle θ_C can be calculated. This angle relies on the velocity of the particle v as well as the refractive index of the medium n. Due to the reliance of θ_C on n different radiators can identify different particles as in Figure 3.8 (b).

RICH1

RICH1 is located upstream of the magnet immediately after the VELO and covers the the full acceptance region. RICH1 houses two radiating materials, aerogel tiles and C_4F_{10} gas. The aerogel tiles are positioned immediately after the VELO and are placed in their own gas-tight box. This is surrounded by a larger box encompassing the rest of the RICH1 detector which contains the C_4F_{10} gas. The Cherenkov photons are reflected by mirrors within the detector to HPD photo-detectors outside the LHCb acceptance. The HPDs are shielded from the magnetic field generated by the dipole magnet and are positioned above and below the detector acceptance.

RICH2

RICH2 is located after the T3 tracking station and covers a narrower acceptance of 120 mrad in the y-axis. This allows the detection of high momentum particles and extends the coverage of the RICH system to 100 GeV/c^2.

Using a similar design to RICH1 the mirrors in the RICH2 sub-detector are housed in a gas-tight box containing CF_4. The Chrenkov radiation is reflected to magnetically shielded HPDs positioned left and right of the bending plane.

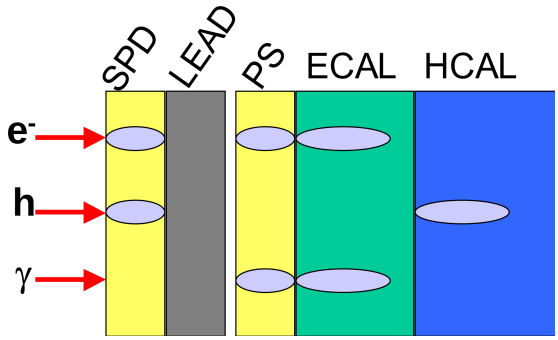


Figure 3.9: *The layout of the calorimeters within LHCb. It also illustrates the energy deposition from different particles throughout the detector. [33]*

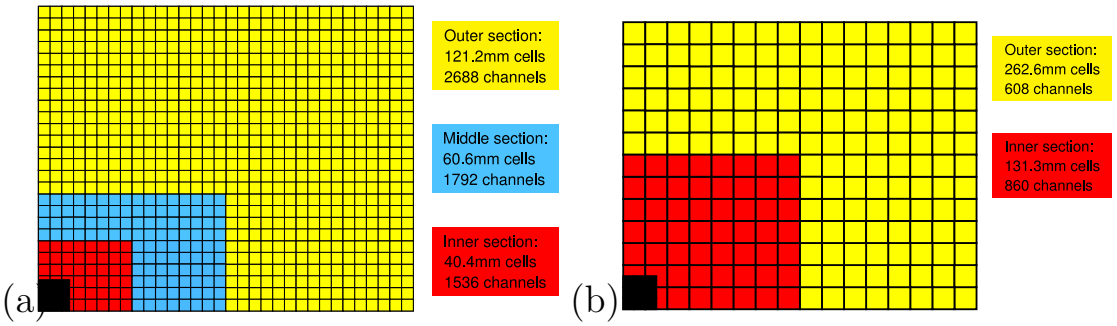


Figure 3.10: *This shows the segmentation of one quarter of (a) the SPD/PS/ECAL and (b) the HCAL detectors. The segmentation is chosen to give a more uniform occupancy across the whole detector. The tiles used in the different cells have different sizes as described in the figure. [24]*

3.4.2 The calorimeters

The calorimeter system is composed of multiple layers. These are the Scintillation Pad Detector (SPD), Pre-Shower detector (PS), Electromagnetic CALorimeter (ECAL) and the Hadronic CALorimeter (HCAL). This system allows for electron and photon identification and also provides information to the L0 Trigger (Section 3.5). The layout of these detectors are shown in Figure 3.9. All of the Calorimeters are constructed in two halves and are supported on rails to allow them to be opened and closed for access.

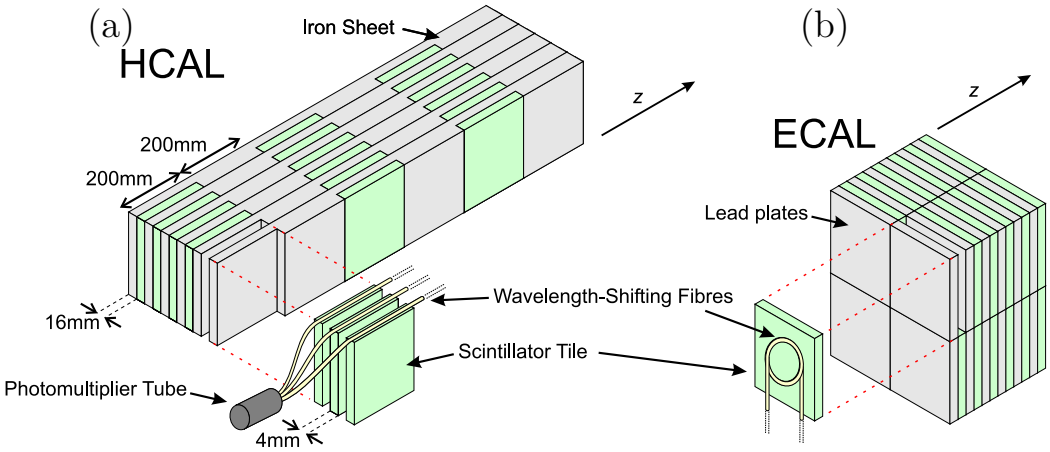


Figure 3.11: *Layout of the tiles (cells) within the panels of the (a) HCAL and (b) ECAL sub-detectors [34].*

SPD and PS sub-detectors

The Scintillation Pad Detector (SPD) and Pre-Shower (PS) detectors are composed of planes of scintillating plastic. These are separated by a layer of lead of approximately $2.5\,X_0$ or $0.1\,\lambda_I$ thickness[1]. The SPD and PS detector planes are composed of thin square plastic scintillator tiles (cells) composed of polystyrene containing small quantities of WaveLength Shifting (WLS) dopants. Tiles within the SPD and PS detector have the same segmentation as the ECAL in Figure 3.10 (a). This granularity is chosen to minimise the occupancy of the individual cells in the regions of high radiation flux close to the beam pipe as shown in the figure.

To detect the light emitted within the tiles WLS optical fibres are embedded within a circular groove for each of the panels as in Figure 3.11 to carry this light to photo-detectors outside the LHCb acceptance.

To distinguish between photons and electrons within the Calorimeter the readout electronics within the SPD and PS differ. The electronics for the SPD readout a time-integrated signal and perform a threshold cut. This allows for high energy photons to be discarded from the SPD output. The PS detector uses a 10-bit ADC to digitize the measured signal pulse. Between the PS and SPD the lead absorber is thick enough to allow electrons to interact electromagnetically but not for pions to interact hadronically.

[1] X_0 is the electromagnetic radiation length of a material, λ_I is the nuclear interaction length.

Electromagnetic Calorimeter

The Electromagnetic CALorimeter (ECAL) is a multi-layered sub-detector situated after the PS. This detector uses a layered shashlik[1] design which has a modest energy resolution. This design sandwitches 4 mm deep scintillator tiles between layers of 2 mm deep lead tiles as in Figure 3.11 (b).

The total ECAL consists of 66 layers of lead tiles and 67 layers of scintillators, which corresponds to approximately $25\,\mathrm{X}_0$, or $1.1\,\lambda_I$. The ECAL is designed to provide information to the L0 trigger to identify events containing electrons with high E_T.

Additionally, when selecting electrons the ECAL must reject π^0 candidates with large E_T. Similar to the SPD/PS, light from scintillators in the ECAL are collected by WLS fibres and is carried to photo-detectors outside the detector acceptance. The resolution of the ECAL is $\sigma(E)/E = 10\%/\sqrt{E} \oplus 1\%$ (E given in GeV/c^2)[24].

Hadronic Calorimeter

The Hadronic CALorimeter is another multi-layered sub-detector positioned downstream to the ECAL sub-detector. This calorimeter makes use of iron rather than lead to separate the scintillating tiles. The HCAL provides information to the L0 trigger to select hadronic events containing candidates with a large E_T.

The HCAL also uses a layered detector design. Unlike the ECAL the scintillators are orientated vertically rather than horizontally as in Figure 3.11 (a). The distance between the iron sheets is chosen to correspond to $1\,\mathrm{X}_0$ vertically, and $1\,\lambda_I$ along the z axis.

The HCAL detector is designed with the constraints of the LHCb cavern in mind and measures $5.6\,\lambda_I$ along the z axis. It is built as a wall downstream of from the ECAL. Unlike other calorimeters the readout electronics are located within the detector acceptance, behind the detector. The HCAL resolution is $\sigma(E)/E = (65 \pm 5)\,\%/\sqrt{E} \oplus (9 \pm 5)\,\%$ (E given in GeV/c^2)[24].

[1] 'This type of calorimeter has been nicknamed shashlik calorimeter from the Russian word for skewer...' [35], which is also the name for a popular type of eastern European kebab.

3.4.3 Muon chambers

The muon chambers in the LHCb detector consist of five separate stations M1-M5 shown in Figure 3.3. These are designed to measure the position of muons traversing all of the chambers through gas ionization. The muon stations M2-M5 are separated by large iron absorbers each of which corresponds to approximately $20\,\lambda_I$. The muon stations are also assembled in two halves and designed to open for access to the detectors.

The energy required for muons to pass through all five stations is 6 GeV. The M1 muon station is located before the calorimeters in Figure 3.12 (a). This provides information for the L0 trigger to select events containing muons with a large p_{T}. This adds $0.26\,\lambda_I$ of material before the calorimeters.

The muon chambers are segmented using four different regions R1-R4 as in Figure 3.12 (b). The size of each station increases incrementally between M1 and M5. The low occupancy segments R2-R4 are composed of multi wire proportional chamber (MWPC) detectors. The detector granularity varies to reduce the occupancy closer to the beam pipe.

The region closest to the beam pipe given by R1 in Figure 3.12 (b) is composed of triple gas electron multiplier (GEM) detectors which have a finer granularity. Triple GEM detectors are also more resilient to the high radiation flux close to the beam pipe. The gas used within both detector types is a mixture of Ar (45%) / CO_2 (15%) / CF_4 (40%), chosen to achieve good time resolution.

3.5 Trigger system

The role of the trigger in the LHCb detector is to reduce the rate of data from the detector to a rate which can be stored offline. To reduce the data rate, decisions are made on a per-event basis to keep interesting events and to reject background. There are multiple trigger lines which are designed to pick different types of events based on their characteristic properties. The LHCb trigger is composed of two main stages shown in Figure 3.13:

1. Hardware Trigger (Level 0)
2. Software Trigger (HLT1 & HLT2)

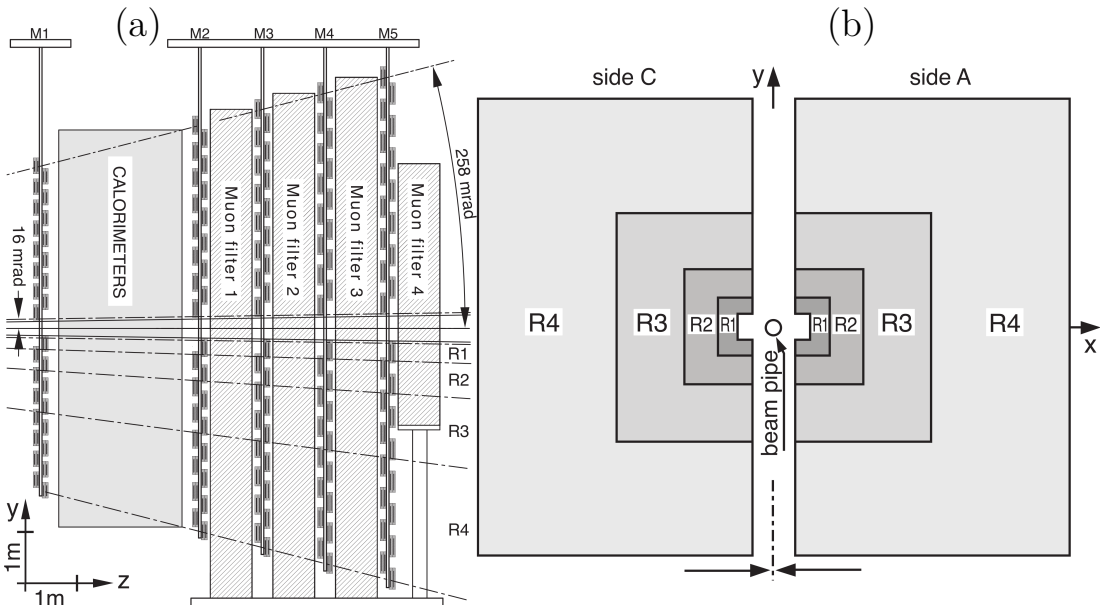


Figure 3.12: *This cross-section shows the relative size and position of the Muon sub-detectors M1-M5 (a) [24] as well as the R1-R4 regions (b) [36] in the cross-section of each Muon plane.*

Hardware trigger

The first level of the trigger L0 is based on hardware and reduces the total rate from the experiment from a maximum of 40 MHz to 1.1 MHz primarily by selecting events with high transverse momentum. The information used in the L0 trigger is limited to that from the calorimeter and the muon systems.

Within the L0 trigger, multiple lines are designed to identify particles with high transverse momentum (p_{T}) or high transverse energy ($\mathrm{E_T}$). During 2011 there were 23 L0 trigger lines [37].

Software trigger

After the data has been read out using the L0 trigger it passes to the High Level Trigger (HLT) which is implemented in software. The HLT runs on the Event Filter Farm (EFF) which is a cluster of CPUs running in the radiation shielded part of the LHCb cavern. The EFF runs 26,000 concurrent instances of the Moore program (Section 3.6) in order to reconstruct the particles in each event. The decision to accept or reject each event is made within a time window of 30 ms. Due to this time constraint only a certain amount of information can be used on a per-event basis.

The HLT trigger is sub-divided into two levels, HLT1 and HLT2. The HLT1 trigger is designed to verify the decision made by the L0 trigger and reduce the total rate to a final 50 kHz. HLT1 performs a full 3D reconstruction of tracks within the VELO to determine the best primary vertex of each event. Cuts are made based on the measured momentum (p) and transverse momentum (p_{T}) of various tracks as well as the quality of the impact parameter for each event. During the 2011 running there were 38 HLT1 trigger lines [37].

The HLT2 trigger is more complex and is designed to run over all of the events that pass the HLT1 trigger lines. Because the HLT2 trigger has to run over less events it is able to make a more complete reconstruction of each event. This allows the HLT2 trigger to reconstruct all tracks which are seeded by hits in the VELO.

The HLT2 triggers are separated into inclusive and exclusive. Inclusive triggers make cuts which are generic to all B meson decays whilst the exclusive triggers make cuts designed to improve the signal for specific decays. The HLT2 trigger is able to reduce the rate to 5 kHz of events being written to disk. The current trigger configuration is stored in a Trigger Configuration Key (TCK) which is stored with each selected event. Throughout 2011 running a typical TCK contained information on 131 HLT2 trigger lines [37].

3.6 LHCb software

The LHCb software stack uses multiple programs to perform different tasks at each stage of the analysis. These tools, used for simulating the LHCb detector as well as (re)processing the collected data, are built around the GAUDI framework [39]. A large amount of the software stack also makes use of the ROOT analysis framework [40, 41].

A list of the major tools used in the reconstruction and analysis of data is as follows [42]:

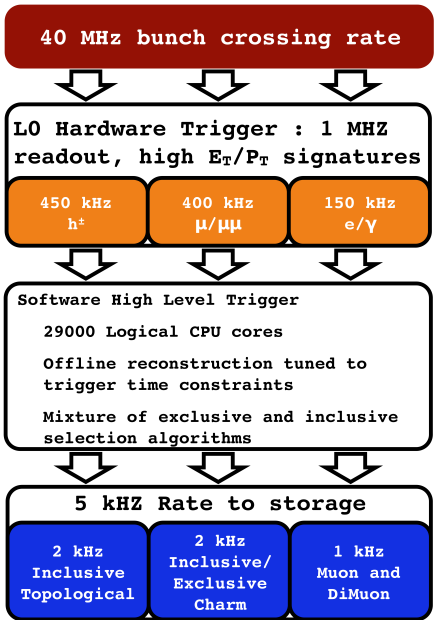


Figure 3.13: *This is diagram showing the flow of data through the LHCb Trigger System [38]. Here the rate of events stored to disk is only 5 kHz compared to the 40 MHz of the LHCb experiment.*

- GAUSS

  The GAUSS [42] software package is used to perform generation and simulation of Monte Carlo (MC) events. Internally, GAUSS uses the PYTHIA [29] program to simulate the underlying hard interactions, with the EVTGEN [43] package used to simulate the B meson decays, and the final state radiation simulated by PHOTOS [44]. The actual detector simulation uses the GEANT4 [45] toolkit.

- BOOLE

  The BOOLE [42] program is used to digitize the MC simulated output from GAUSS. This piece of software simulates the detector response.

- MOORE

  The Moore [42] program is used to apply HLT selections on the input data. This program can be run over raw data from the detector as well as the output from digitized MC from BOOLE. MOORE adds extra information to each event including a summary of which triggers selected each event and the particles used in the trigger decision.

- Brunel
  Brunel [42] is used to fully reconstruct events offline.
- DaVinci
  DaVinci [42] is used for Stripping data that has been fully reconstructed as well as for performing a large amount of analysis work.

ROOT Analysis Framework

The ROOT framework [40, 41] is heavily used within the LHCb software architecture. This framework comprises a large set of libraries useful for data analysis, simulation and presentation. The ROOT framework has been designed to allow for interactive use of most tools within the framework along with the ability to link against the framework with external executables.

To perform a physics analysis using the data collected by LHCb the RapidFit [46] program was developed in Edinburgh. This package makes extensive use of ROOT libraries to perform complex tasks and is discussed in more detail in Chapter 7.

3.6.1 Data reconstruction

Data reconstruction for LHCb makes use of the computing GRID [47, 48] in order to store and reconstruct data. The GRID consists of distributed resources across many computing sites. For LHCb, larger computing sites are used in reconstructing/storing data and are referred to as Tier1 sites. Smaller computing sites are used for generating Monte Carlo data and are referred to as Tier2. CERN is the only Tier0 site, but also functions as a Tier1 site.

The data from the LHCb detector is first stored in a raw format containing both information from the detector and the trigger decision. This data is then reconstructed in an initial processing, the output from this has the format of 'Data Summary Tape' (DST) files which are distributed across the Tier1 sites of the GRID. These files are then used in the Stripping process which selects events into different physics streams for different analyses. The output from this selection process is also in a DST format.

Datasets which are used in the main physics analysis are selected from events from these stripping lines. The selections relevant to this thesis are discussed in more detail in Chapter 4. The output from this selection process is stored in a standard ROOT file format.

"What is the most important symbol in physics? Is it this: $+$? Is is this: $\times$? Is it this $=$? No. I claim that it is this: $\sim$. Tell me the order of magnitude, the scaling.
***That** is the physics."*

Yuval Grossman

4

Data selection, acceptances and tagging

As introduced in Chapter 1, the focus of this thesis is to study the decay of $B_s^0 \to (J/\psi \to \mu^+\mu^-)\ K^+K^-$ (Section 2.5). This chapter describes the process by which the data is collected and the selection cuts that have been applied.

Candidate events are selected using the triggers outlined Section 4.1. These events are required to pass further selection cuts applied by the stripping in Section 4.2 and a final selection summarised in Section 4.3.

Fully simulated Monte-Carlo (MC) samples are used to optimise and understand the selection of candidate events. They are also used in understanding the behaviour of the LHCb detector. The MC datasets used in this analysis are described in Section 4.4 and a comparison between the data and MC samples is made in Section 4.5.

The effect of finite time resolution in the reconstruction of data is described in Section 4.6. In addition to this, variations in acceptance for both the decay time and angular distributions are discussed in Section 4.7.

Important input required for the full $B_s^0 \to J/\psi\ K^+K^-$ analysis is the flavour of the B_s^0 meson at production, determined through flavour tagging the meson. Flavour tagging uses additional information from each event and is described in more detail in Section 4.8.

4.1 Trigger selection

The selection of the data used in this analysis makes use of different combinations of triggers (Section 3.5) to select events of interest. The presence of high momentum muons in the $B_s^0 \to J/\psi\, K^+K^-$ final state provides a clean signature which can be exploited at the trigger level.

Two combinations of triggers are used in this analysis to construct independent datasets. These are referred to as decay time *almost unbiased* and decay time *exclusively biased* [12]. This naming scheme reflects the fact that different cuts in these triggers effects the decay time distribution of recorded events. The corresponding trigger decisions for these datasets are listed in Table 4.1.

Dataset	Trigger decision
almost unbiased	`Hlt2DiMuonDetachedJPsi TOS` and `Hlt1DiMuonHighMass Dec`
exclusively biased	`Hlt2DiMuonDetachedJPsi TOS` and `not Hlt1DiMuonHighMass Dec` and (`Hlt1TrackMuon` or `Hlt1TrackAllL0`)

Table 4.1: *Triggers used for selecting the data used in this analysis.*

The cuts applied by the L0 triggers used in this analysis are shown in Table 4.2 and are described in more detail in [37]. These triggers select events containing muons which have a large transverse momentum (p_T) as well as placing an upper limit on the number of recorded hits in the Scintillating Pad Detector (SPD) (Section 3.4.2) per event.

The HLT triggers used in selecting the *almost unbiased* dataset are described in Table 4.3. Both triggers for this dataset make use of measurements from the individual muon tracks.

The `Hlt1DiMuonHighMass` trigger selects only events containing two muons with large momenta (p) and transverse momenta (p_T). It also cuts on the quality of the muons by using the χ^2/nDoF of the reconstructed track trajectory. In addition, lose cuts are made on the reconstructed vertex, including the Distance Of Closest Approach (DOCA) of the two muon candidates, as well as cutting on the chi-squared of the vertex fit (χ^2_{vtx}) and the reconstructed mass.

The `Hlt2DiMuonDetachedJPsi` trigger performs a more complete candidate reconstruction and places a tighter constraint on the reconstructed track χ^2, the re-calculated χ^2_{vtx} and the J/ψ mass. The trigger also calculates the flight distance of particles; this is defined as the distance between the primary vertex and the decay vertex. Cuts on this flight distance χ^2 are made for the reconstructed J/ψ meson with respect to the primary vertex.

The additional triggers used in selecting the *exclusively biased* dataset are described in Table 4.4. The Impact Parameter (IP) of a particle is the Distance Of Closest Approach (DOCA) of the reconstructed track to the primary vertex. These triggers make use of the B^0_s IP and the associated χ^2. Additionally these triggers both have requirements for the muons to have high momentum and transverse momentum. Good quality of tracks are required by cutting on the χ^2/nDoF as well as the number of hits/missed-hits per track in the VELO and OT/IT (Sections 3.3.1 and 3.3.2).

4.2 Stripping selection

The stripping used for the $1.0\,\text{fb}^{-1}$ analysis is commonly referred to as Stripping17(b) [12] and the cuts applied during this level of the selection are listed in Table 4.5. The selection made at the stripping stage is tighter than any previous trigger selection.

For muons in the J/$\psi \to \mu^+\mu^-$ decay the difference between the log-likelihood functions under the different particle hypotheses of the muon and pion ($\Delta\text{LL}\,\mu\pi$[1]) is required to be > 0. This is a lose cut favouring the muon hypothesis. Also, the fit to the J/ψ vertex is required to have a low χ^2_{vtx}/nDoF and a selection is made on the di-muon invariant mass about the J/ψ mass [7].

For kaons in the $\phi \to K^+K^-$ decay the difference between the log-likelihood functions for the kaon and pion hypotheses ($\Delta\text{LL}\,K\pi$) is required to be > -2, another lose cut favouring the kaon hypothesis. The transverse momentum of the ϕ meson is required to be $> 1\,\text{GeV}/c$, and the invariant di-kaon mass is required to be within a narrow window around the ϕ mass [7]. Additionally the fit to the ϕ meson vertex is required to have a low χ^2_{vtx}/nDoF.

[1] $\Delta\text{LL}\,\mu\pi = \ln\{\text{L}(\mu)\} - \ln\{\text{L}(\pi)\} = \ln[\text{L}(\mu)/\text{L}(\pi)]$

Track Property	L0Muon	L0DiMuon
Threshold	$p_{\mathrm{T}} > 1.48$ GeV/c	$\sqrt{p_{\mathrm{T}}^{largest} \times p_{\mathrm{T}}^{2nd\ largest}} > 1.296$ GeV/c
SPD* multiplicity	< 600	< 900

Table 4.2: *The requirements for candidates to pass the* L0Muon *and* L0DiMuon *Triggers [49]. *Scintillating Pad Detector*

Track Property	Hlt1DiMuonHighMass	Hlt2DiMuonDetachedJPsi
Pre-Requirements	L0Muon or L0DiMuon	any Hlt1 trigger
Track p_{T} [GeV/c]	> 0.5	-
Track p [GeV/c]	> 6	-
Track χ^2/nDoF	< 4	< 5
DOCA* [mm]	< 0.2	-
χ^2_{vtx}	< 25	< 25
Mass [GeV/c^2]	> 2.7	$M_{\mathrm{J/\psi}} \pm 0.12$
Flight Distance** (J/ψ) χ^2	-	> 9

Table 4.3: *Requirements for events to pass the trigger selection associated with the decay time almost unbiased trigger decision [49]. (*) Distance Of Closest Approach. (**) Distance between primary and decay vertices.*

Track Property	Hlt1TrackMuon	Hlt1TrackAllL0
Track IP* [mm]	> 0.1	> 0.1
Track IP* χ^2	> 16	> 16
Track p_{T} [GeV/c]	> 1	> 1.7
Track p [GeV/c]	> 8	> 10
Track χ^2/nDoF	< 2	< 2.5
Number VELO hits/track	-	> 9
Number missed VELO hits/track	-	< 3
Number OT+IT $\times$ 2 hits/track	-	> 16

Table 4.4: *Requirements for events to pass the triggers only used as part of the decay time exclusively biased trigger decision [49]. (*) Impact Parameter, defined as the DOCA of the track to the primary vertex.*

Decay mode	Cut parameter	Stripping 17
all tracks	$\chi^2_{\text{track}}/\text{nDoF}$	< 5
$\text{J}/\psi \to \mu^+\mu^-$	$\Delta\text{LL}\ \mu\pi$	> 0
	$\chi^2_{\text{vtx}}/\text{nDoF}(\text{J}/\psi)$	< 16
	$\lvert M(\mu^+\mu^-) - M(\text{J}/\psi)\rvert$	$< 80\,\text{MeV}/c^2$
$\phi \to \text{K}^+\text{K}^-$	$\Delta\text{LL}\ \text{K}\pi$	> -2
	$p_{\text{T}}(\phi)$	$> 1\,\text{GeV}/c$
	$M(\phi)$	$\in [980,\ 1050]\,\text{MeV}/c^2$
	$\chi^2_{\text{vtx}}/\text{nDoF}(\phi)$	< 16
$\text{B}^0_{\text{s}} \to \text{J}/\psi\,\phi$	$M(\text{B}^0_{\text{s}})$	$\in [5200,\ 5550]\,\text{MeV}/c^2$
	$\chi^2_{\text{vtx}}/\text{nDoF}(\text{B}^0_{\text{s}})$	< 10
	t (*)	$> 0.2\,\text{ps}$

Table 4.5: *This table outlines the selection applied in the stripping. (*) Decay time as calculated using OfflineVertexFitter tool.*

Cuts are also made on the fully reconstructed $\text{B}^0_{\text{s}} \to \text{J}/\psi\ \text{K}^+\text{K}^-$ candidates. The final B^0_{s} vertex is required to have a small $\chi^2_{\text{vtx}}/\text{nDoF}$ and in addition to this the lifetime of the B^0_{s} is required to be $> 0.2\,\text{ps}$. Also, only events within the mass range $M(\text{B}^0_{\text{s}}) \in [5200, 5550]\ \text{MeV}/c^2$ are kept.

4.3 Final selection

The output from the stripping selection is the basis for a final selection which produces the dataset used in the main physics analysis. The final cuts used for this analysis are listed in Table 4.6.

The final selection tightens the requirement on the χ^2/nDoF for all tracks, as well as removing clone tracks. Clone tracks are candidates created when a single particle trajectory is split during the event reconstruction. To remove these clones a cut is made on the Kullback-Liebler distance [50]; this is a measure of the shared information between two tracks. These clone tracks are rejected by requiring this distance be > 5000.

Decay mode	Cut parameter	Final selection
all tracks	$\chi^2_{\text{track}}/\text{nDoF}$	< 4
	clone distance	> 5000
$\text{J}\!/\!\psi \to \mu^+\mu^-$	$\Delta\text{LL}\,\mu\pi$	> 0
	$\min\left(p_{\text{T}}\left(\mu^+\right), p_{\text{T}}\left(\mu^-\right)\right)$	$> 0.5\,\text{GeV}/c$
	$\chi^2_{\text{vtx}}/\text{nDoF}(\text{J}\!/\!\psi)$	< 16
	$\lvert M\left(\mu^+\mu^-\right) - M\left(\text{J}\!/\!\psi\right)\rvert$	$\in [3030,\, 3150]\,\text{MeV}/c^2$
$\phi \to \text{K}^+\text{K}^-$	$\Delta\text{LL}\,\text{K}\pi$	> 0
	$p_{\text{T}}(\phi)$	$> 1\,\text{GeV}/c$
	$M(\phi)$	$\in [990, 1050]\,\text{MeV}/c^2$
	$\chi^2_{\text{vtx}}/\text{nDoF}(\phi)$	< 16
$\text{B}^0_{\text{s}} \to \text{J}\!/\!\psi\,\phi$	$M(\text{B}^0_{\text{s}})$	$\in [5200,\, 5550]\,\text{MeV}/c^2$
	$\chi^2_{\text{vtx}}/\text{nDoF}(\text{B}^0_{\text{s}})$	< 10
	$\chi^2_{\text{DTF(B+PV)}}/\text{nDoF}(\text{B}^0_{\text{s}})$	< 5
	$\chi^2_{\text{IP}}(\text{B}^0_{\text{s}})$	< 25
	$\chi^2_{\text{IP,next}}(\text{B}^0_{\text{s}})$	> 50
	t (*)	$[0.3, 14.0]\,\text{ps}$

Table 4.6: *This table outlines the cuts applied by the final selection. (*) Decay time as calculated using DecayTreeFitter tool.*

The final selection for the decay of $J/\psi \rightarrow \mu^+\mu^-$ involves narrowing the mass window and making an additional cut on the minimum p_T for either of the muon tracks. The final selection for the $\phi \rightarrow K^+K^-$ decay is the same as used in the stripping selection, aside from the mass range of the ϕ candidates being reduced.

For the fully reconstructed decay of $B_s^0 \rightarrow J/\psi\, K^+K^-$ the final selection requires the decay time range to be tightened. This removes a significant amount of lower time background events and poorly reconstructed events with a high decay time.

The B_s^0 meson is also required to have originated from the best PV for each event. This selection is performed through requiring that the χ^2 of B_s^0 with respect to the best PV $\chi^2_{IP}(B_s^0)$ is small, while the χ^2 with respect to any other PV in the same event $\chi^2_{IP,next}(B_s^0)$ is large.

For the final selection the decay time of the B_s^0 meson is re-calculated by the DecayTreeFitter tool [51]. This package performs a global kinematic fit to the whole B_s^0 decay chain including the IP. In addition to cutting on this re-calculated decay time, it is required that the total χ^2/nDoF from this kinematic fit, $(\chi^2_{DTF(B+PV)}/nDoF)$ should be < 5.

The mass distributions for the reconstructed J/ψ, ϕ and B_s^0 candidates which pass all of the selections for this analysis are discussed in detail in Sections 5.1 and 5.2.

4.4 MC dataset

Two large datasets of fully simulated Monte Carlo (MC) events were generated using the LHCb software framework described in Section 3.6. Each of these datasets contain 5 million signal events and are identified as being reconstructed using the processings labelled `Sim05-Reco12-Stripping17` and `Sim05-Reco12a-Stripping17`. The conditions used for generating these datasets is representative of the detector and reconstruction conditions throughout 2011. The two datasets were selected using a trigger configuration which is representative of the running conditions during 2011[1]. The underlying physics parameters used for the generation of events in fully simulated MC data are given in Table 4.7.

[1]This configuration is identified by the Trigger Configuration Key (TCK), which for MC Simulation corresponds to using $0 \times$ `40760037`.

Parameter	Generation Value	
Γ_s	0.6793	[ps^{-1}]
$\Delta\Gamma_s$	0.060	[ps^{-1}]
$\lvert A_0 \rvert$	0.775	
$\lvert A_\perp \rvert$	0.40	
$\lvert A_\parallel \rvert$	0.49	
$\lvert A_S \rvert$	0	
δ_0	0	[rad]
$\delta_\perp$	-0.17	[rad]
$\delta_\parallel$	0	[rad]

Table 4.7: *Input physics parameters used for the generation of the fully simulated MC datasets. (These parameters are defined in Section 2.5.)*

4.5 Data and MC comparison

Some distributions of candidate properties from MC simulations are compared with data in Figure 4.1, with both data and MC distributions normalised based on their sample sizes. These distributions overlay the momentum and transverse momentum for kaons, muons and B_s^0 mesons. Each data distribution is background subtracted using data in the upper and lower B_s^0 mass side-bands ([5218,5293]MeV/c^2 and [5443,5518]MeV/c^2) to remove the background in the signal region ([5293,5443]MeV/c^2). The mass distribution of the signal events is discussed in Section 5.1.

Here, all candidates from both MC simulation and data have passed a modified selection which has no cuts applied on the J/ψ or ϕ reconstructed masses.
In Figure 4.1 the distributions are in good agreement. Nevertheless, the momentum distributions for the B_s^0 mesons and kaons are consistently harder in data than MC. Any impact this may have on fitting to data is discussed in Section 8.2.

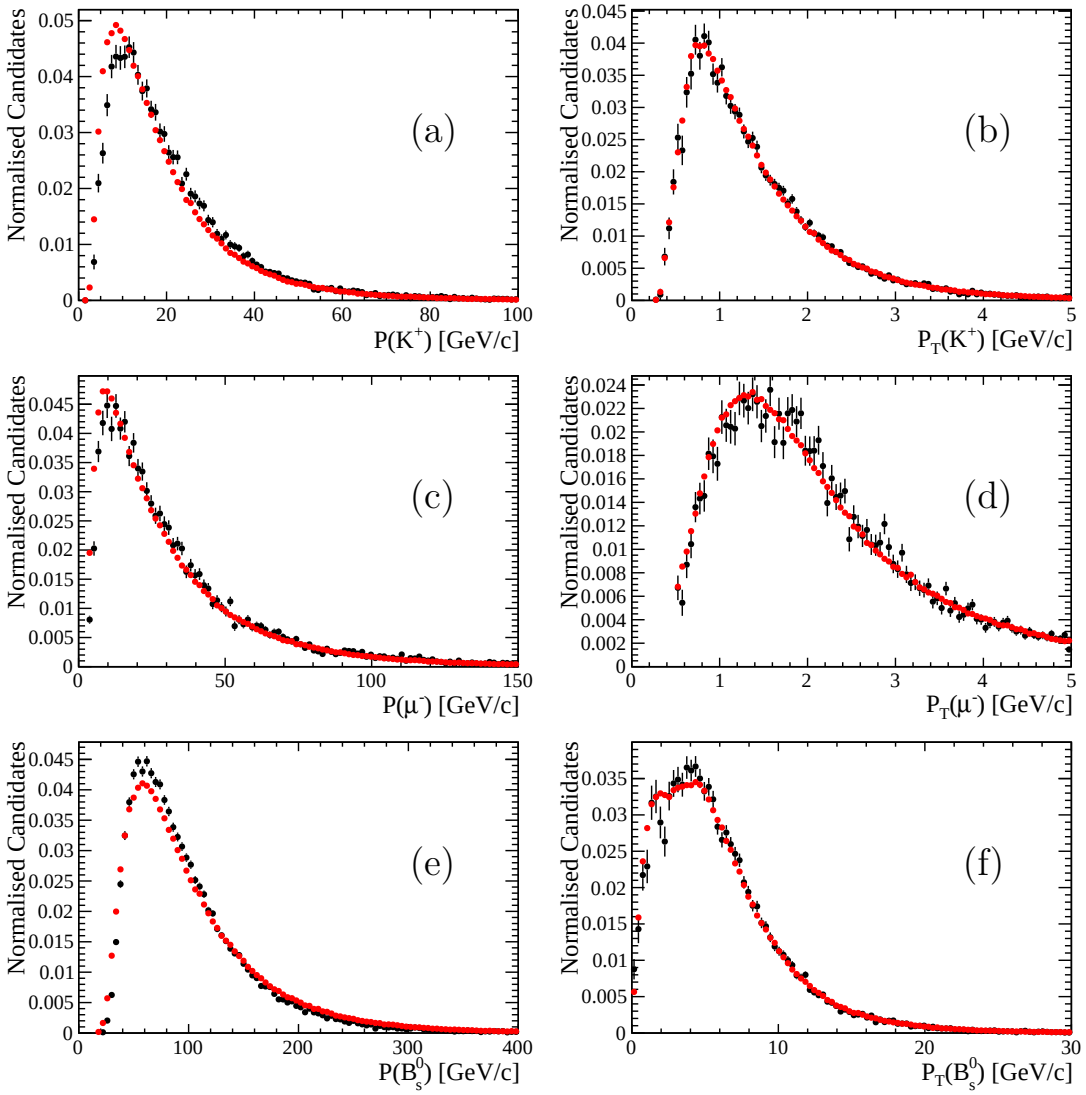


Figure 4.1: *Here the distributions corresponding to the momentum (p) and transverse momentum (p_{T}) for the kaon, muon and $\mathrm{B}_{\mathrm{s}}^{0}$ meson systems shown for MC and data. kaon: (a) p, (b) p_{T}; muon: (c) p, (d) p_{T}; $\mathrm{B}_{\mathrm{s}}^{0}$: (e) p, (f) p_{T} .*

4.6 Decay time resolution

The VELO detector (Section 3.3.1) makes high precision measurements of the beam interaction point and decay vertex of B_s^0 mesons. This allows for precision measurement of the decay time of each event. This is important to resolve the B_s^0 oscillation period of approximately 350 fs. Each measurement has an associated error which is calculated from the kinematic and vertex fits. To incorporate the effects of this in the fit to data it is important to understand the detector resolution.

When fitting to data a per-event model of time resolution is used to optimally make use of the information in the fit. This time resolution model is calibrated using the time resolution model of prompt events which are described below.

Time resolution of prompt events

The prompt time resolution is determined by the distribution of $J/\psi\ K^+K^-$ candidates reconstructed from a prompt J/ψ meson and two random kaon tracks. Here the data has passed a modified stripping and final selection which places no cut on the decay time. The decay time distribution of these events is shown in Figure 4.2 (a) and (b) using two different scales. Events in this distribution have been background subtracted with the *s*Plot technique [52] using the $\mu^+\mu^-$ mass distribution as a discriminating variable.

The distribution of events in Figure 4.2 is described through the convolution of a signal function with a resolution function R. The signal function corresponds to a delta function $\delta(t=0)$ describing prompt J/ψ events and two exponentials describing the $B_s^0 \to J/\psi\ K^+K^-$ decay as seen in the upper lifetime of Figure 4.2 (b). The resolution function R describing the distribution of prompt-events is described as the sum of 3 Gaussians:

$$R(t;\sigma_t) = \sum_{i=1}^{3} \frac{f_i}{\sqrt{2\pi}\ (s_i \cdot \sigma_t)} e^{\left(-\frac{(t-d)^2}{2(s_i\cdot\sigma_t)^2}\right)} . \tag{4.1}$$

Here the ith Gaussian function has a width of $s_i \cdot \sigma_t$ and a common offset d with the values of t and σ_t determined on a per-event basis.

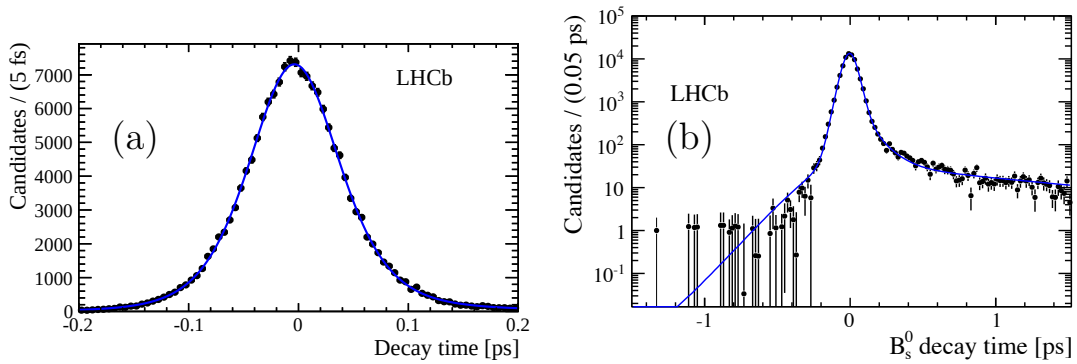


Figure 4.2: *Decay time distribution of prompt* $B_s^0 \to J/\psi\, K^+K^-$ *[12]. The fit result (blue line) is the result from fitting the signal model described in the main text, convolved with the time resolution function as described in the text. Both figures (a) and (b) are projections from the same fit and dataset over different scales and ranges. The distributions in these plots have been background subtracted as described in the text.*

Per-Event Decay Time Resolution

When fitting to data a different per-event decay time resolution model is used. In order to reduce the complexity of the analysis a simplified form of the resolution model described by equation 4.1 is used.

$$R_1\left(t;\sigma_t\right) = \frac{1}{\sqrt{2\pi}\ \left(s_t\cdot\sigma_t\right)} e^{\left(-\frac{t^2}{2(s_t\cdot\sigma_t)^2}\right)} \ . \tag{4.2}$$

The width of this single Gaussian function is defined as the per-event decay time uncertainty σ_t multiplied by the scale factor s_t. Using the distributions of prompt events the calibration factor is determined to be $s_t = 1.45 \pm 0.06$. The calibrated distribution of per-event decay time uncertainties is given in Figure 4.3, the fixed uncertainty of 45 fs is shown for comparison.

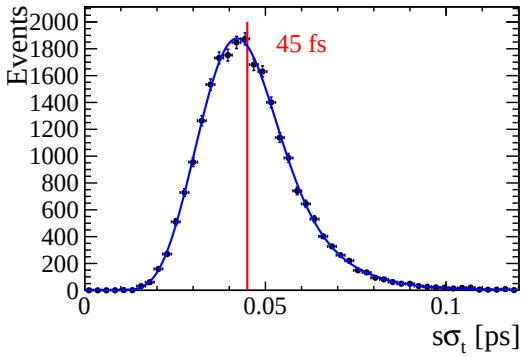


Figure 4.3: *The distribution per-event values of $s \cdot \sigma_t$ with 45 fs highlighted which corresponds to the value for a single fixed width Gaussian time resolution model.*

4.7 Detector acceptance

The geometry of the LHCb detector and cuts applied in the data selection modifies the true distribution of events. A description of detector acceptance in general relies on simulated Monte Carlo (MC) events. The detector efficiency (ε) is the ratio of the number of generated B_s^0 mesons compared to the number which were selected,

$$\varepsilon = \frac{\text{Selected } B_s^0}{\text{Generated } B_s^0} \ . \tag{4.3}$$

The efficiency is known to vary as a function of both the decay time and angles and is described by the function $\varepsilon\left(t, \vec{\Omega}\right)$. This efficiency function is separable into angular and decay time efficiencies, $\varepsilon^a\left(\vec{\Omega}\right)$ and $\varepsilon^t\left(t\right)$ as in equation 4.4. These efficiencies are discussed separately in Sections 4.7.1 and 4.7.2 respectively.

$$\varepsilon\left(t, \vec{\Omega}\right) = \varepsilon^t\left(t\right) \times \varepsilon^a\left(\vec{\Omega}\right) \ . \tag{4.4}$$

When fitting to data an NLL function is used (Section 6.2) which is constructed from a probability density function (PDF). This function is constructed to be the normalised form of the differential decay rate, $\frac{d\Gamma_{(B_s^0 \to J/\psi\, K^+K^-)}\left(t, \vec{\Omega}\right)}{dt\, d\vec{\Omega}}$, described

in Section 2.5 and represented as $X\left(t,\vec{\Omega}\right)$. The detector acceptance, $\varepsilon\left(t,\vec{\Omega}\right)$, is incorporated into this PDF as a product of the decay rate,

$$\text{PDF}\left(t,\vec{\Omega}\right) = \frac{X\left(t,\vec{\Omega}\right) \times \varepsilon\left(t,\vec{\Omega}\right)}{\iint dt d\vec{\Omega}\; X\left(t,\vec{\Omega}\right) \times \varepsilon\left(t,\vec{\Omega}\right)} \; . \tag{4.5}$$

Here both the decay rate $\Gamma_{\mathrm{B_s^0}}\left(t,\vec{\Omega}\right)$ and the efficiency function ε appear explicitly in the numerator and denominator. When the decay rate and acceptance functions are separable into independent time and angular functions, equation 4.5 may be re-written as,

$$\text{PDF}\left(t,\vec{\Omega}\right) = \frac{\sum_{k}^{10} \varepsilon_k^t\left(t\right) \varepsilon_k^a\left(\vec{\Omega}\right) f_k\left(\vec{\Omega}\right) h_k\left(t\right)}{\sum_{k}^{10} \int d\vec{\Omega}\; \varepsilon_k^a\left(\vec{\Omega}\right) f_k\left(\vec{\Omega}\right) \int dt\; \varepsilon_k^t\left(t\right) h_k\left(t\right)} \; . \tag{4.6}$$

Here h_k contains the decay-time dependent component and f_k the angular component from the decay rate equation 2.19. The description of the acceptance can be incorporated into the PDF through the use of an analytic function or using a histogram describing the variation of the acceptance.

4.7.1 Angular acceptance

The knowledge of angular acceptance relies on information from fully simulated Monte Carlo (MC) events. The acceptance function is used differently within the numerator and denominator of the PDF in equation 4.5. The reason for this is due to the way that the fit to data is constructed as described in Section 6.2.1.

In the numerator the acceptance function is only used when constructing PDF projections and for performing consistency checks. Because of this, the detector acceptance is described by a three-dimensional histogram.

The angular acceptance within the denominator however is included within an integral over the whole phase-space. In order to incorporate this acceptance, a different method is used to efficiently pre-calculate the result of this integral. This results in a set of angular acceptance weights, which are incorporated into the fit.

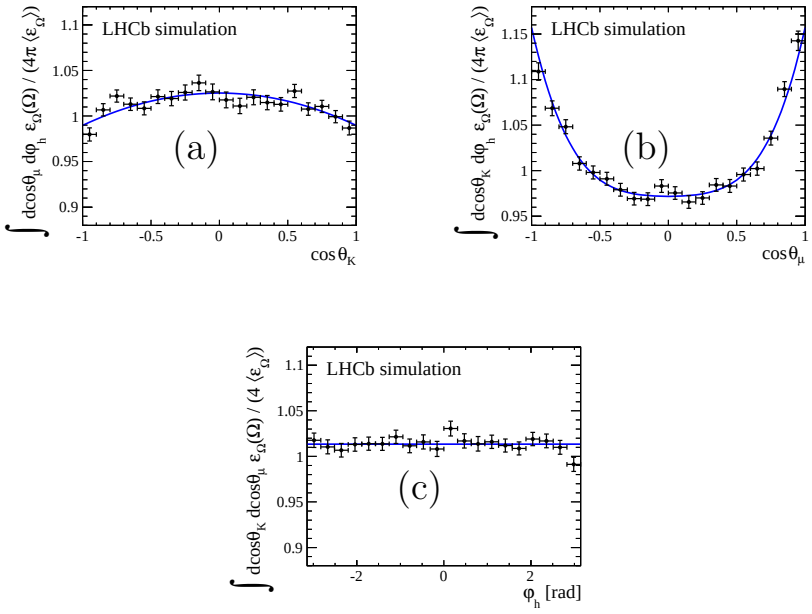


Figure 4.4: *Projections of the efficiency histogram extracted from fully simulated MC [12]. (a) Projection in* $\cos(\theta_{\mathrm{K}})$*, (b) Projection in* $\cos(\theta_{\mu})$*, (c) Projection in* ϕ_{h}

Angular acceptance in PDF numerator

The efficiency histogram used in the numerator for the fit function is a three-dimensional histogram constructed from the comparison of the distribution of MC events and the true angular distribution. The MC events used within this histogram have been fully reconstructed and have passed all cuts applied in the final selection of data. The true angular distribution is calculated using the signal differential decay rate (equation 2.19) and knowledge of the physics conditions used to generate the MC (Table 4.7).

The full angular histogram is calculated across all three angles using the complete MC dataset. The efficiency of the selection is defined by comparing the true distribution of expected events to the distribution of observed MC events as in equation 4.3. Figure 4.4 shows the one-dimensional projections from the full efficiency histogram. These projections show that the variation of angular efficiency at most approximately 15–20% at the edge of the angular acceptance.

As previously mentioned these acceptance histograms are used for making PDF projections, however they are not required when fitting to data.

Angular acceptance in PDF denominator

The angular acceptance weights used in the denominator are also determined from fully simulated MC samples. In the denominator of the PDF in equation 4.6 the angular efficiencies are included within an integral over the whole phase-space. Due to this, the angular acceptance in the denominator is described through the use of acceptance weights. The process by which the weights are determined is summarised in this section and discussed in more detail in Appendix B as well as in Ref [53].

From equation 4.6, the angular components of the PDF are defined as $f_k\left(\vec{\Omega}\right)$ hence the corresponding weights are ξ_k and are calculated as,

$$\xi_k = \int d\vec{\Omega}\ \varepsilon_k^a\left(\vec{\Omega}\right) f_k\left(\vec{\Omega}\right) \quad . \tag{4.7}$$

$$\mathrm{PDF}\left(t,\vec{\Omega}\right) = \frac{\sum_k^{10} \varepsilon_k^t\left(t\right) \varepsilon_k^a\left(\vec{\Omega}\right) f_k\left(\vec{\Omega}\right)}{\sum_k^{10} \xi_k \int dt\ \varepsilon_k^t\left(t\right) f_k\left(t\right)} \quad .$$

When calculating the acceptance weights using MC, only events passing all of the cuts in the final selection are used. The distribution of observed MC events is defined as $\mathcal{S}_{\mathrm{obs}}\left(\vec{\Omega}\,|\,t\right)$ and the distribution of the true physical distribution is $\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}\,|\,t\right)$. Using these distributions ξ_k can be expressed as,

$$\xi_k = \int d\vec{\Omega}\ \mathcal{S}_{\mathrm{obs}}\left(\vec{\Omega}\,|\,t\right) \frac{f_k(\vec{\Omega})}{\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}\,|\,t\right)} \quad . \tag{4.8}$$

This integral is approximated using a large MC dataset of $\mathrm{N_{obs}}$ events to integrate over the MC phase-space to produce,

$$\begin{aligned}\xi_k &= \frac{1}{\mathrm{N_{obs}}}\sum_{\mathrm{e}=1}^{\mathrm{N_{obs}}}\left[\frac{f_k\left(\vec{\Omega}_{\mathrm{e}}\right)}{\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}_e\,|\,t_e\right)}\right] \quad , \\ &= \frac{1}{\mathrm{N_{obs}}}\sum_{\mathrm{e}=1}^{\mathrm{N_{obs}}}\left[f_k\left(\vec{\Omega}_{\mathrm{e}}\right)\Big/\frac{X\left(\vec{\Omega}_{\mathrm{e}}|t_{\mathrm{e}}\right)}{\int d\vec{\Omega}\,X\left(\vec{\Omega}|t_{\mathrm{e}}\right)}\right] \quad . \end{aligned} \tag{4.9}$$

This final equation 4.9 is used to calculate the angular acceptance weights from MC. The angular acceptance weights calculated using the full MC dataset (Section 4.4) are shown in Table 4.8. The errors shown in this table correspond to statistical uncertainties from the size of the Monte-Carlo dataset.

For a detector with perfect angular acceptance, the angular acceptance weights in Table 4.8 reduce to $\xi_{1,2,3,7} = 1,\ \xi_{4,5,6,8,9,10} = 0$. This particular case is often referred to as "flat" angular acceptance. This table shows that the final acceptance weights calculated from MC are close to the case of a "flat" angular acceptance and as such the effect of angular acceptance is expected to be small.

k	ξ_k	
	From MC	Flat Acceptance
1	$+0.9800 \pm 0.0009$	1
2	$+1.0287 \pm 0.0014$	1
3	$+1.0283 \pm 0.0016$	1
4	-0.0010 ± 0.0015	0
5	-0.0004 ± 0.0010	0
6	-0.0004 ± 0.0009	0
7	$+0.9916 \pm 0.0009$	1
8	$+0.0010 \pm 0.0014$	0
9	$+0.0007 \pm 0.0013$	0
10	-0.0072 ± 0.0028	0

Table 4.8: *Angular Normalisation weights calculated using equation 4.9. These weights were calculated using the full MC datasets described in Section 4.4 using events passing the final selection. The acceptance weights defined for the case of "Flat Acceptance" corresponds to the case of a perfect acceptance. [12]*

4.7.2 Decay time acceptance

Unlike the angular acceptance, the decay time acceptance can be determined using data driven methods. There are two different effects which alter the decay time efficiency. The lower decay time acceptance effect is a selection efficiency introduced by the triggers used to select events. The upper decay time efficiency is an effect which reduces the efficiency of events with large decay time. Both of these acceptance effects can be described through the use of an efficiency histogram in decay time. This efficiency histogram is used in both the numerator and denominator of the fitting function PDF $\left(t, \vec{\Omega}\right)$ (equation 4.5).

Lower decay time acceptance

The lower decay time acceptance effect is due to the use of time biasing triggers in the selection of events. The trigger requirements which produce a time biasing effect are cuts on parameters correlated to the decay time such as the cuts on the impact parameter (IP) or the decay length. There are two different categories of triggers which are used to construct the *almost unbiased* and *exclusively biased* datasets. These triggers are defined in Table 4.1. The two different triggers also require different strategies to determine the biasing effect that they have on the decay time.

For the *almost unbiased* trigger line a subset of events are additionally selected by the *completely unbiased* trigger line. This additional trigger is unbiased at the HLT1 level and uses the pre-scaled `Hlt2DiMuonJPsi` trigger line. It is possible to use this additional trigger line due to the fact that the selection it performs on data is the same as the `Hlt2DiMuonDetachedJPsi` trigger aside from the cuts on the impact parameter known to bias the lifetime. Making use of the fact that some of these events pass through both triggers it is possible to calculate the efficiency (ϵ_1) using[1],

$$\epsilon_1(t) = \frac{\#\text{ events selected by (}\textit{almost unbiased}\text{) AND (}\textit{completely unbiased}\text{) triggers}}{\#\text{ events selected by (}\textit{completely unbiased}\text{) trigger}} . \tag{4.10}$$

[1] Whilst this equation correctly defines the trigger efficiency the actual calculation uses knowledge of the pre-scaling applied to the `Hlt2DiMuonJPsi` trigger line. This corrects for the fact this trigger line was pre-scaled by a factor of 5 during the second half of running in 2011.

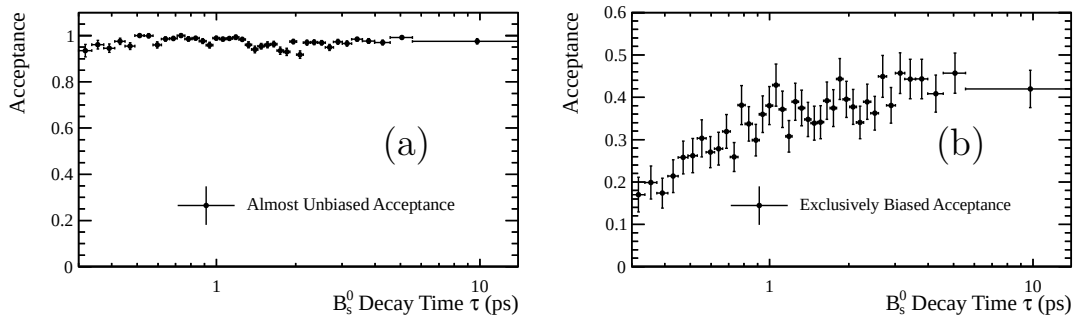


Figure 4.5: *Lower decay time acceptance histograms for the (a) almost unbiased [12] and (b) exclusively biased [12] datasets.*

In order to calculate the number of events selected by each trigger line, the number of B_s^0 signal events is extracted from a fit to the $m(J/\psi\ K^+K^-)$ distribution. This fit is the same as described in Section 5.1. The resulting acceptance histogram for the *almost unbiased* dataset is shown in Figure 4.5 (a). This method is referred to as the "overlap" method.

For the *exclusively biased* dataset the events are selected such that they are independent of the *almost unbiased* dataset. To calculate the decay acceptance for this trigger the decay time distribution of events selected by this trigger can be compared to the distribution of events selected by the *completely unbiased* trigger. The decay time acceptance for these events (ϵ_2) is therefore calculated using,

$$\epsilon_2(t) = \frac{\#\text{ events selected by (}\textit{exclusively biased}\text{) trigger}}{\#\text{ events selected by (}\textit{completely unbiased}\text{) trigger}} \ . \tag{4.11}$$

The yield in each bin in time is calculated using fits to $m(J/\psi\ K^+K^-)$ as described previously and the resulting acceptance histogram is shown in Figure 4.5 (b). This method is referred to as the "ratio" method.

Upper decay time acceptance

In addition to the lower decay time acceptance effects there is a drop in efficiency in selecting B_s^0 mesons with large decay times [54]. This drop in upper decay time efficiency is described by a linear function,

$$\epsilon(t) = (1 + \beta) \ . \tag{4.12}$$

The drop in upper decay time efficiency has been determined to be related to various effects including track reconstruction efficiency, VELO geometry and cuts performed in selecting events. Using a combination of data driven methods and fully simulated events the value of the β-factor for $B_s^0 \to J/\psi\, K^+K^-$ is determined to be [55],

$$\beta = (-8.3 \pm 4.0) \times 10^{-3}\,\mathrm{ps}^{-1} \tag{4.13}$$

Using simulated datasets it has been determined that the only effect of this efficiency is to apply an offset to the measured lifetime Γ_s. This allows for this efficiency to be applied to the measured value of Γ_s after fitting.

4.8 Flavour tagging at production

To study the oscillation of B_s^0 mesons it is important to determine the initial flavour of the b-quark at production. This process is referred to as flavour tagging. Several flavour tagging algorithms are used in the LHCb experiment, the results from each of these are grouped into two tagging decisions depending on the strategy used. These are referred to as the Opposite Side Tagger (OST) and the Same Side Tagger (SST) as illustrated in Figure 4.6.

The result of these two tagging algorithms and the use of the individual taggers are described in more detail in Refs [56] and [57] respectively and are summarised in Section 4.8.1.

Each tagged event has an associated tag decision (d) and a mistag (η) which describes the probability of the tag decision being incorrect. The final mistag calculation and calibration is discussed in more detail in Sections 4.8.2 and 4.8.3.

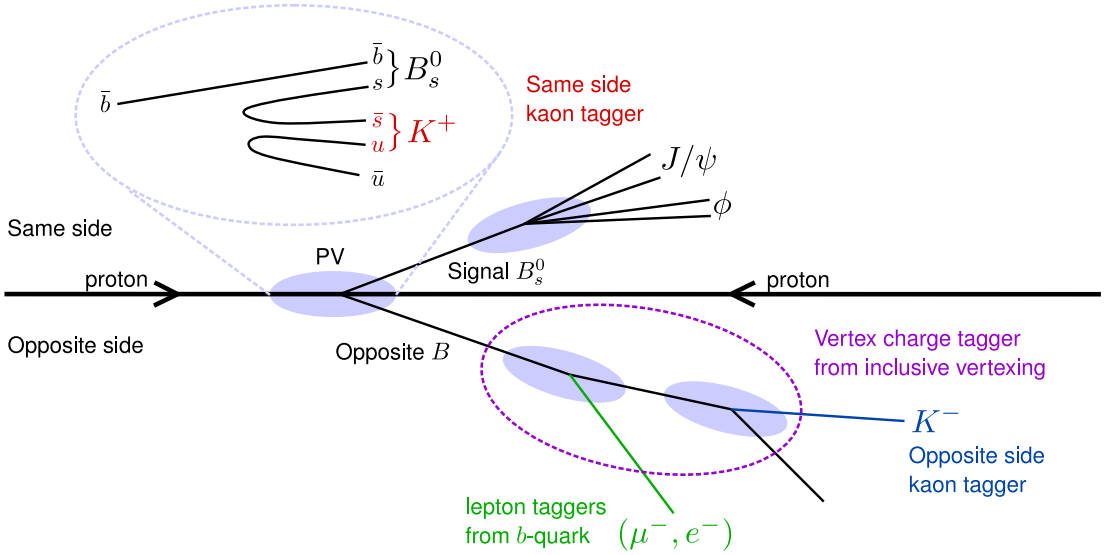


Figure 4.6: *The same side and opposite side taggers illustrated for the signal and opposite* B *mesons [34]. In this figure the signal and opposite* B *mesons are well separated, nevertheless in the actual detector the events may be geometrically close. The taggers are described in more detail in Section 4.8.1.*

4.8.1 Description of flavour taggers

B mesons are primarily produced in pairs through the gluon fusion mechanism (Section 3.1.1). In this analysis the B_s^0 meson is regarded as the signal meson with any secondary B hadron defined as the opposite side meson. Because of the two mesons in each event there are two different types of taggers used in identifying the initial flavour of the B_s^0 meson. The Opposite Side Tagger (OST) uses information from the decay of this opposite side B meson. This information is collected from four different taggers shown in Figure 4.6. The Same Side Tagger (SST) uses a single kaon tagger using a possible kaon that is formed through hadronisation of the signal B_s^0 meson.

Opposite side tagger

The four different taggers used in the OST are, the vertex charge tagger, two leptonic taggers (e and μ) and the opposite side kaon tagger [56]. The electron and muon taggers tag events based on the semi-leptonic decays of the opposite B meson ($B \to X\, l\overline{\nu}_l$). The opposite side kaon tagger assigns a tag to the event based on the kaon produced in the decay chain $b \to c \to s$, Ref [56].The e, μ and K taggers are commonly referred to as single particle taggers.

Tagger	p_{T} (GeV/c)	p (GeV/c)	IP/σ_{IP}	PID cuts	$\mathrm{IP_{PU}}/\sigma_{\mathrm{IP_{PU}}}$
μ	> 1.2	> 2.0	–	$\Delta\mathrm{LL}\ \mu\pi > 2.5$	> 3.0
e	> 1.0	> 2.0	> 2.0	$\Delta\mathrm{LL}\ e\pi > 2.5$	> 3.0
K	> 0.8	> 5.9	> 4.0	$\Delta\mathrm{LL\ K}\pi > 6.5$ $\Delta\mathrm{LL\ Kp} > -3.5$	> 4.7

Table 4.9: *Particle selection criteria used in the selection of candidates for the various single particle taggers used in the OST [56]. The various parameters used to apply selection cuts are described in more detail in the text.*

The candidates used by the single particle tagging algorithms are required to have large transverse momentum (p_{T}) and a large impact parameter significance (IP/σ_{IP}). For events containing multiple candidates the the candidate with the highest transverse momentum is chosen as the opposite B meson to tag the flavour of the signal $\mathrm{B_s^0}$.

Each event used by the single particle tagger must pass a PID requirement on the difference of the log-likelihood (ΔLL) for each particle hypothesis. Only B candidates corresponding to the best primary vertex are used. They must also have a large significance with respect to any other Pile-Up vertices ($\mathrm{IP_{PU}}/\sigma_{\mathrm{IP_{PU}}}$) in the same event. The requirements for the events used by these taggers is summarised in Table 4.9.

Lepton taggers also place additional requirements on the candidates they tag. The electron tagger requires these candidates to deposit a certain amount of ionising charge in the VELO. They also require that the ratio of energy (E) over momentum (p), as measured in the ECAL, to be $\mathrm{E}/p > 0.6$. The muon tagger also requires that each muon track not share any hits in the muon chambers with tracks from other events.

In addition to the single particle taggers a vertex charge tagger is used which is able to determine the sign of the charge for a secondary decay vertex of the opposite side B meson decay.

The vertex charge tagging process includes performing an inclusive reconstruction of the decay vertex of the opposite side B meson. From this reconstructed vertex a weighted charge is calculated using,

$$Q_{\text{vtx}} = \frac{\sum_i p_{\text{T}}^{\kappa}(i) Q_i}{\sum_i p_{\text{T}}^{\kappa}(i)} \tag{4.14}$$

In this equation the parameter κ is optimised for the best tagging power resulting in $\kappa = 0.4$. Only charged final vertices are tagged and events with vertices having a charge of $|Q_{\text{vtx}}| < 0.275$ remaining untagged.

The final decision of the OST relies on the combination of the results from each of the four taggers into one decision for each event. This decision is made using the method outlined in Section 4.8.2, where the results of all four individual taggers are combined.

Same side tagger

The SST makes a decision on the flavour of each event based on the same side kaon tagger as in Figure 4.6. This makes a decision on the tag of each event based on a charged kaon produced as part of the hadronisation of the signal B_s^0 meson. Based on simulations using the hadronisation in the PYTHIA [29] event generator (Section 3.6) approximately 50% of events produce a charged kaon candidate alongside the production of the signal B_s^0 meson [57].

Candidates used by the same side kaon tagger require that an additional kaon is consistent with the B_s^0 production vertex. This requirement is made through an IP significance cut of $\text{IP}/\sigma_{\text{IP}} < 2.5$. Kaon candidates are also required to have a high momentum p and transverse momentum p_{T} to increase the efficiency of the reconstructed candidate. As with the OST, in the case of events having multiple candidate kaons, the candidate with the highest p_{T} is chosen and determines both the event tag and mistag.

The kaons are required to be outside of a conical volume containing the B_s^0 decay which is defined through a cut on the azimuthal angle about the B_s^0 flight direction. Backgrounds from misidentified particles are mainly suppressed through cuts on the PID information for each candidate. Backgrounds introduced though combining kaons from other sources with the signal B_s^0 meson are also reduced. This supression is obtained by cutting on the difference in mass between the signal B_s^0 meson and the mass from the reconstructed $\text{B}_\text{s}^0\text{K}^\pm$ vertex.

4.8.2 Combining tagging decisions

For each tagging decision, a corresponding mistag probability is estimated using knowledge of the mistag algorithm as well as the kinematics of the event. This mistag probability is calculated using a neural network. The final output mistag is calibrated using data as discussed in Section 4.8.3.

When candidates pass the selection required for multiple tagging algorithms, multiple tag decisions are made with associated uncertainties. The result from multiple taggers must be combined into a single tag decision.

For each event the tag decision is given by $d = 1(-1)$ when the flavour of the quark is $\overline{b}(b)$ in a $\mathrm{B}(\overline{\mathrm{B}})$ meson. For each event the predicted probability of an incorrect tag is η. Using the information available on a per-event basis the tags are combined using equation 4.15. Here the per-event probability that the B meson contains a $b/(\overline{b})$-quark is given as $P(b)/P\left(\overline{b}\right)$. The probability is calculated using the combined tagger decisions $p(b)$ and $p\left(\overline{b}\right)$. The combined tagger decision is calculated as,

$$P(b) = \frac{p(b)}{p(b) + p\left(\overline{b}\right)} \qquad P\left(\overline{b}\right) = 1 - P(b) \tag{4.15}$$

where:

$$p(b) = \prod_i \left(\frac{1 + d_i}{2} - d_i \left(1 - \eta_i\right) \right), \quad p\left(\overline{b}\right) = \prod_i \left(\frac{1 + d_i}{2} + d_i \left(1 - \eta_i\right) \right)$$

Here the individual tagger decisions are d and associated mistags, η. To reduce the contribution of taggers with poor tagging power only tags from the kaon or vertex taggers with $\eta < 0.46$ are tagged, otherwise the event remains untagged.

The combination of taggers using equation 4.15 doesn't take into account any correlation between the tagging algorithms. Due to this, the combined mistag for the OST is potentially overestimated. To correct for this the tagging algorithm has to be calibrated using data. This allows for the correct mistag ω for each event to be calculated.

4.8.3 Mistag calibration

The per-event mistag probability calculated for a given tagging algorithm (η) is calibrated to the true mistag ω using self-tagging control channels. For the OST the control channel used is $\mathrm{B}^{\pm} \to \mathrm{J/\psi\,K}^{\pm}$ and for the SST the control channel is $\mathrm{B_s^0} \to \mathrm{D_s^{\pm}}\pi^{\mp}$. To perform this calibration, events in these channels are tagged and binned according to their associated mistag (η). The true value of the mistag (ω) is then calculated for all events in each bin.

From this, the variation of of η vs ω can be plotted and a calibration equation is fit to the distribution. The resulting distribution for the control channel $\mathrm{B}^{\pm} \to \mathrm{J/\psi\,K}^{\pm}$ is presented in Figure 4.7 (a), and the calibration equations are,

$$\begin{aligned} \omega(\eta) &= p_0 + \frac{\Delta p_0}{2} + p_1(\eta - \langle\eta\rangle) \quad , \\ \overline{\omega}(\eta) &= p_0 - \frac{\Delta p_0}{2} + p_1(\eta - \langle\eta\rangle) \quad . \end{aligned} \tag{4.16}$$

These equations give the true mistag values $\omega(\eta)$ and $\overline{\omega}(\eta)$ for B and $\overline{\mathrm{B}}$ mesons. The parameters p_0 and p_1 are obtained through fitting equation 4.16 to the binned distribution of η vs ω.

The parameter $\langle\eta\rangle$ is the mean calculated mistag probability, and $\frac{\Delta p_0}{2}$ is used to account for any asymmetry in the tagging efficiency between B and $\overline{\mathrm{B}}$. This is defined as $\Delta p_0 = p_0^{\mathrm{B}} - p_0^{\overline{\mathrm{B}}}$ and is calculated by performing the calibration using the different self tagged final states separately. In the case that the different tagging algorithms are uncorrelated and optimal it is anticipated that $p_0 = \langle\eta\rangle$ and $p_1 = 1$. The final mistag calibration for the two tagging decisions are listed in Table 4.10.

The per-event distributions of η for the combined decision as well as the separate OST and SST decisions are shown in Figures 4.7 (b), (c) and (d) respectively. These distributions have been background subtracted using the $_s$Plot [52] technique with $m(\mathrm{J/\psi\ K^+K^-})$ used as the discriminating variable. When fitting to data the per-event value of η is used in combination with the calibration equation 4.16.

The effective tagging power for a given tagger is defined as $\epsilon_{\mathrm{tag}}\mathcal{D}^2$. Here ϵ_{tag} is the efficiency of the tagger and the dilution factor is given as $\mathcal{D} = 1 - 2\omega$. The power of a tagger is defined such that itprovides the same information as $\mathrm{N} \times \epsilon_{\mathrm{tag}}\mathcal{D}^2$ perfectly tagged events, for a dataset of size N.

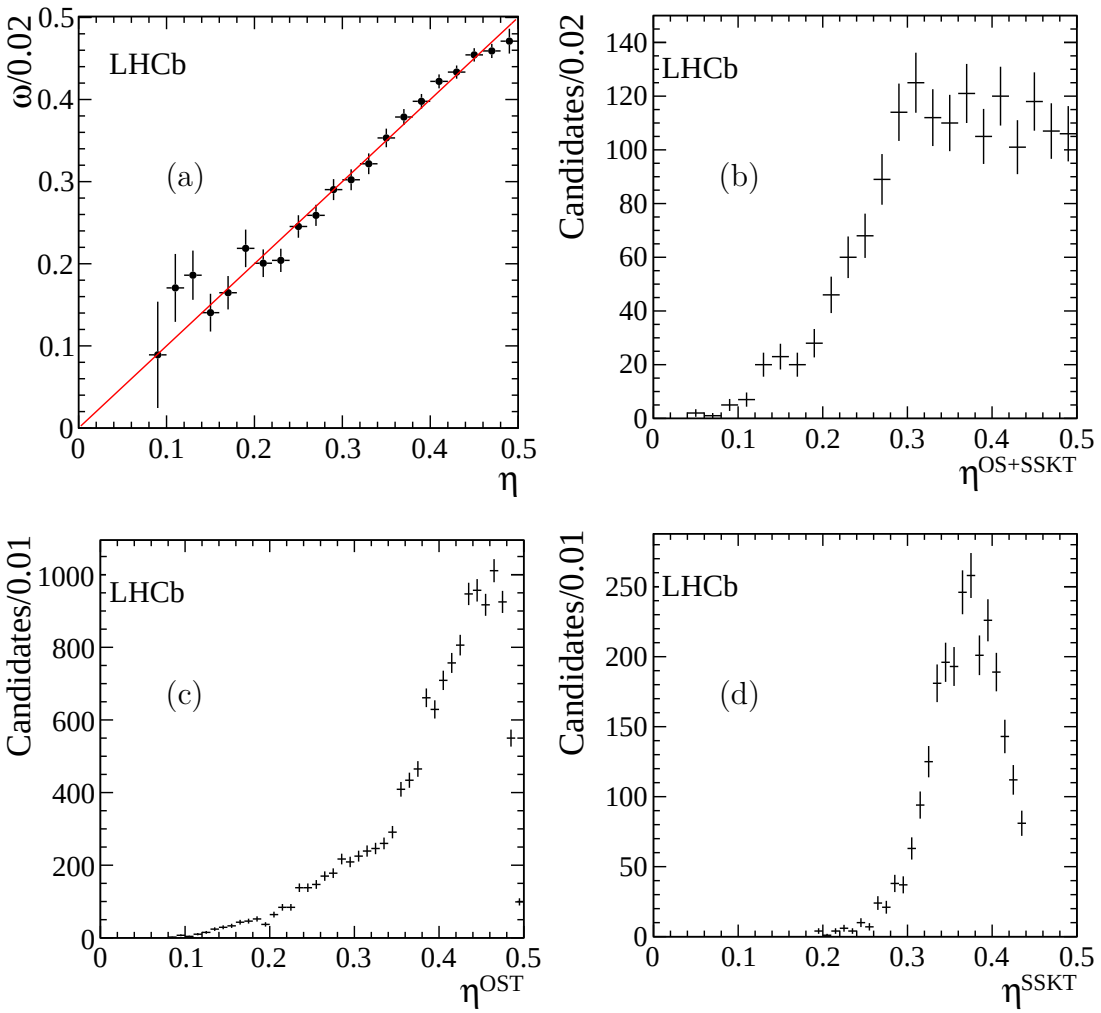


Figure 4.7: *(a): The calibration fit for the OST, η vs ω for the control channel $B^+ \to J/\psi K^+$. Fit result are projected as the red line and reviewed in Table 4.10 [12].*
For $B_s^0 \to J/\psi\, K^+K^-$ candidates, (b): the combined mistag probability η for OST and SST [12], (c): η for just the OST tagged events [12] and (d): η for just the SST tagged events [12]. These distributions have been background subtracted using the sPlot [52] technique with $m(J/\psi\, K^+K^-)$ used as the discriminating variable.

Calibration	p_0	p_1	$\langle\eta\rangle$	Δp_0
OST	$0.392 \pm 0.002 \pm 0.008$	$1.000 \pm 0.020 \pm 0.012$	0.392	0.011 ± 0.003
SST	$0.350 \pm 0.015 \pm 0.007$	$1.000 \pm 0.160 \pm 0.020$	0.350	-0.019 ± 0.005
OST+SST	0.000 ± 0.025	0.0	0.0	-0.011 ± 0.004

Table 4.10: *This are the final calibration values obtained in fitting the mistag calibration (eq 4.16) to both the OST and SST taggers and the result from the combination of tag decisions.*

SubSet	Tagging Efficiency ϵ_{tag} (%)	Tagging Power $\epsilon_{\text{tag}}\mathcal{D}^2$ (%)
OST Tagged Events	33.00 ± 0.28	2.29 ± 0.06
SST Tagged Events	10.26 ± 0.18	0.89 ± 0.17
OST+SST Tagged Events	0.51 ± 0.03	3.90 ± 0.11
All Data	39.36 ± 0.32	3.13 ± 0.12*

Table 4.11: *Tagging Power and Efficiency in* $B_s^0 \to J/\psi\ K^+K^-$ *for various subsets of the whole dataset [12]. * All quoted errors are statistical with an additional systematic error of* $\pm 0.20\%$ *being assigned to the tagging power calculated for the whole dataset.*

For $B_s^0 \to J/\psi\ K^+K^-$ the tagging efficiencies and tagging power are listed in Table 4.11 [12]. The errors quoted in this table are statistical with an additional error of $\pm 0.20\%$ associated with the total tagging power of the whole dataset.

"Everything should be as simple as it can be, but not simpler."

Albert Einstein

5

Signal Yield and Background

The $B_s^0 \to J/\psi\, K^+K^-$ dataset selected in Chapter 4 is discussed in more detail in this chapter. First a fit to the mass distribution and the signal yield of this dataset is described in Section 5.1. The analysis makes use of the *s*Plot technique [52] to statistically remove combinatorial background from the dataset and this process is described in Section 5.2. Peaking backgrounds present in this dataset are reviewed in Section 5.3.

5.1 Signal yield

The mass distribution for all events passing the final selection is shown in Figure 5.1 (a). The distributions corresponding to the *almost unbiased* and *exclusively biased* datasets are shown separately in Figures 5.1 (b) and (c). From these figures it is clear that the background in this analysis is small and corresponds to $\approx 8\%$ under the B_s^0 signal peak ($m(J/\psi\, K^+K^-) \in [5338.2, 5398.2]\,\mathrm{MeV}/c^2$).

In Figure 5.1 the result from a fit to the mass distribution using the PDF given in equation 5.1 is shown. The signal component of the PDF, $\mathcal{S}(m)$ is the sum of two Gaussian functions with different widths and a common mean. The background component $\mathcal{B}(m)$ is described by an exponential function. The signal fraction is defined as f_{sig}. In the fit to the $m(J/\psi\, K^+K^-)$ distribution all of the parame-

ters described are varied to obtain their optimal values. The result of this fit is summarised in Table 5.1.

$$\mathrm{PDF}\,(m) = f_{sig}\frac{\mathcal{S}\,(m)}{\int dm\,\mathcal{S}\,(m)} + (1 - f_{sig})\,\frac{\mathcal{B}\,(m)}{\int dm\,\mathcal{B}\,(m)} \tag{5.1}$$

$$\mathcal{S}\,(m) = f_{\sigma 1}\left[\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(m-\overline{m})^2}{\sigma_1^2}}\right] + (1 - f_{\sigma 1})\left[\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{(m-\overline{m})^2}{\sigma_2^2}}\right]$$

$$\mathcal{B}\,(m) = e^{-\alpha\, m}$$

Parameter	Final Value
σ_1	$6.1 \pm 0.1\,\mathrm{MeV}/c^2$
σ_2	$12.6 \pm 0.6\,\mathrm{MeV}/c^2$
$f_{\sigma 1}$	0.76 ± 0.03
$\bar{m}$	$5368.20 \pm 0.05\,\mathrm{MeV}/c^2$

Table 5.1: *Values for mass parameters in equation 5.1 with the result from this fit projected in Figure 5.1 (a).*

The average mass $\bar{m}$ from this fit is high compared to the B_s^0 value in Ref [7] ($5366.77\,\mathrm{MeV}/c^2$) due to the fact the momentum scale has not been fully calibrated.

The total signal yield is $\mathrm{N_{sig}} = 27617 \pm 117$, with the dataset composed of $\mathrm{N_{sig}^{UB}} = 23502 \pm 109$ events selected by the *almost unbiased* trigger, and $\mathrm{N_{sig}^{B}} = 4115 \pm 43$ events selected by the *exclusively biased* trigger. These signal yields and errors are calculated using the value and error of the signal fraction f_{sig} obtained from fitting equation 5.1 to the $m(\mathrm{J/\psi\ K^+K^-})$ distribution.

From the distributions in Figure 5.1 it is seen that the mass fit is well behaved and the signal and background are well described.

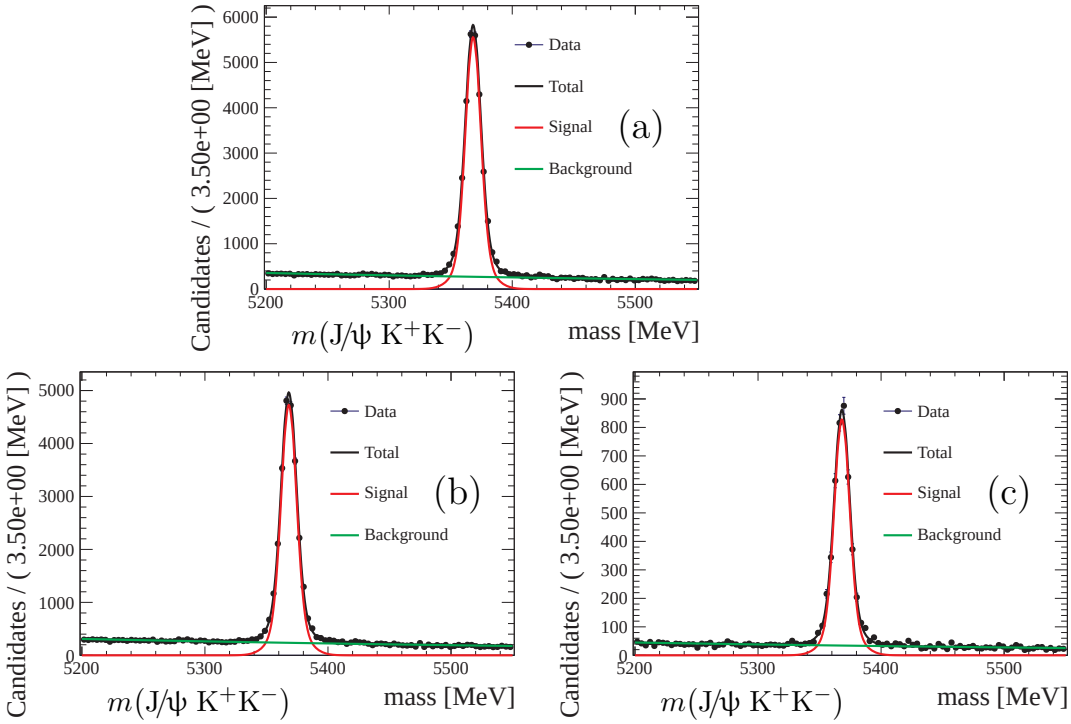


Figure 5.1: *The fit results to the $m(\mathrm{J/\psi\ K^+K^-})$ distribution of events passing the full offline selection. (a) Projected fit result to the complete dataset, (b) Projected fit result to the almost unbiased data subset, (c) Projected fit result to the exclusively biased data subset.*

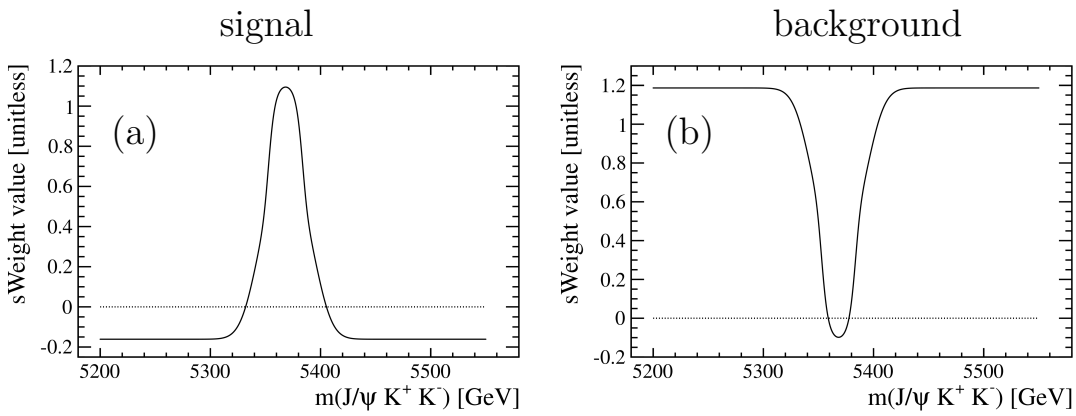


Figure 5.2: *Distribution of per-event weights vs $m(\mathrm{J/\psi\ K^+K^-})$ for all events selected using the almost unbiased trigger. (a) sWeights for extracting Signal from the dataset, (b) sWeights for extracting Background from the dataset.*

5.2 Combinatorial background subtraction

Almost all of the background in the $B_s^0 \to J/\psi\, K^+K^-$ analysis is combinatorial in origin and is statistically subtracted from data using the sPlot [52] technique. This technique involves using a physical observable, termed the discriminant, which is uncorrelated to the other quantities being analysed within the dataset. This discriminant is used to distinguish between the signal and background species. Using this, a per-event sWeight is calculated for the signal or background.

In this analysis the full process to subtract the background from a dataset using the sPlot method is:

1. Identify an uncorrelated physical observable which can be used to discriminate between signal and background. This is chosen to be the mass of each event.
2. Perform a fit to data to determine the optimal set of physics parameters, $\vec{\lambda}$ which describe the distribution of the chosen observable.
3. Using the physics parameters $\vec{\lambda}$, calculate the per-event sWeights according to equation 5.2.
4. Using these values a histogram is constructed populated by the per-event sWeights calculated for each event.

$$ {}_s\mathcal{P}_i\left(\vec{x}\right) = \frac{\sum_{j}^{n_{SP}} \mathbf{V}_{ij}\, F_j\left(\vec{x};\vec{\lambda}\right)}{\sum_{k}^{n_{SP}} N_k F_k\left(\vec{x};\vec{\lambda}\right)} \tag{5.2} $$

Where: $\vec{x}$ = Observables for each event x

${}_s\mathcal{P}_i$ = Per-event sWeight for species i $\quad$ $\mathbf{V}_{ij}$ = covariance matrix

N_k = Number events of species k $\quad$ n_{SP} = number of species

F_j = Normalised PDF describing distribution of species j

The elements of the covariance matrix $\mathbf{V}$ are calculated using the dataset as in equation 5.3.

$$\mathbf{V}_{ij}{}^{-1} = \sum_{y}^{events} \frac{F_i\left(\vec{x}_y;\vec{\lambda}\right) F_j\left(\vec{x}_y;\vec{\lambda}\right)}{\left[\sum_{k}^{n_{SP}} N_k F_k\left(\vec{x}_y;\vec{\lambda}\right)\right]^2} \tag{5.3}$$

In this analysis the PDF used in calculating the sWeights is that given by equation 5.1.

The distribution of per-event weights vs $m(\mathrm{J/\psi\ K^+K^-})$ for events selected using the *almost unbiased* trigger are shown in Figure 5.2. Figure 5.2 (a) shows the distribution of signal sWeights. Here the sWeights are > 0 in the central $m(\mathrm{J/\psi\ K^+K^-})$ region dominated by the signal species and < 0 in upper and lower sidebands which are dominated the background species. Figure 5.2 (b) shows the distribution of background sWeights. These weights are > 0 in the background dominated region and < 0 in the signal dominated region.

Using these sWeights the background subtracted distributions of the J/ψ and $\mathrm{K^+K^-}$ invariant masses are projected in Figure 5.3. In this figure the J/ψ and ϕ resonances are clearly visible and well defined. The $m(\mathrm{K^+K^-})$ distribution in Figure 5.3 (a) is described by a relativistic Breit-Wigner function convolved with a Gaussian function. The $m(\mu^+\mu^-)$ distribution in Figure 5.3 (b) is modelled using a Crystal-Ball function.

The time and angular distributions of the background subtracted $\mathrm{B_s^0}$ signal events are shown in Figure 5.4 (a). The result from using the same technique, but instead removing the signal from the background distributions, is shown in Figure 5.4 (b). From these figures the signal and background distributions are seen to have different structures.

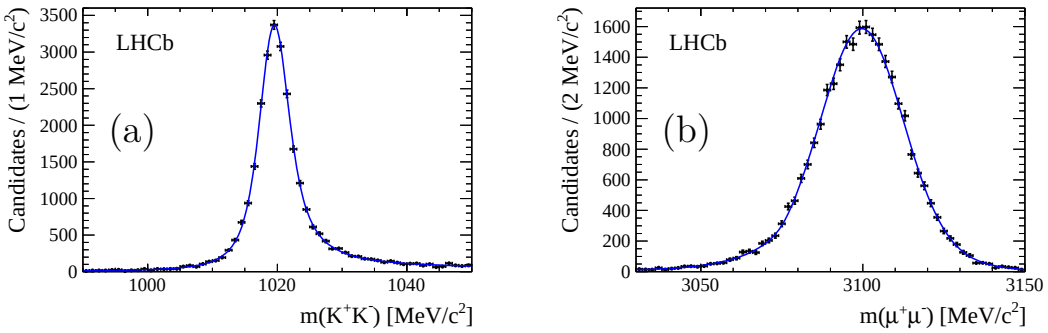


Figure 5.3: *Background-Subtracted masses of all events passing the final selection. The blue line shows the projected fit result to the mass distributions. (a) $m(\mathrm{K}^+\mathrm{K}^-)$ resonance modelled using a relativistic Breit-Wigner function, (b) $m(\mu^+\mu^-)$ resonance modelled using a Crystal-Ball function.*

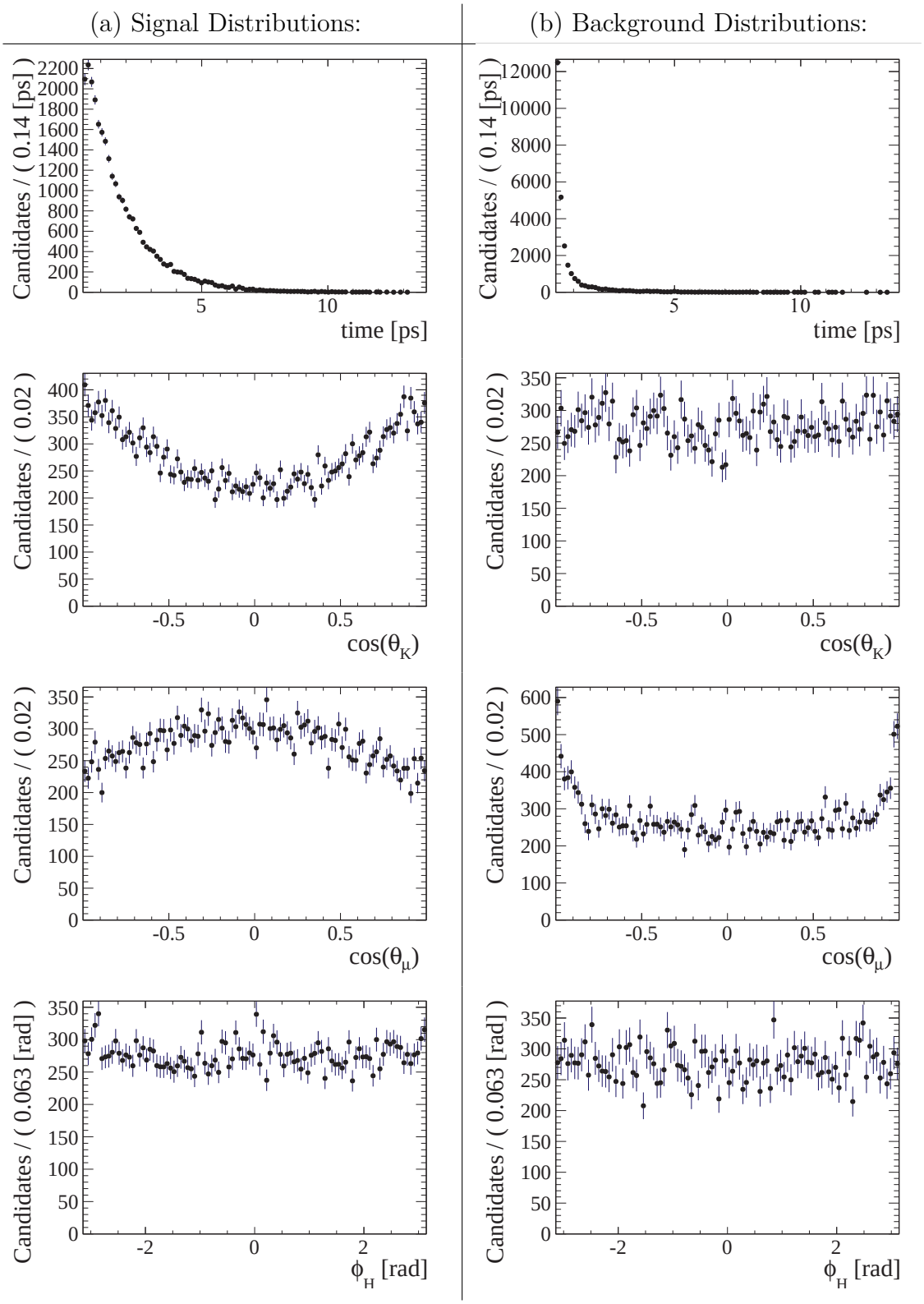


Figure 5.4: *Distributions of the decay time, and helicity angles for all events. Both sets of plots use the sPlot technique to remove one species with the $m(\mathrm{J/\psi\ K^+K^-})$ mass used to discriminate between signal and background. (a) Signal distributions with background subtracted, (b) Background distributions with signal subtracted.*

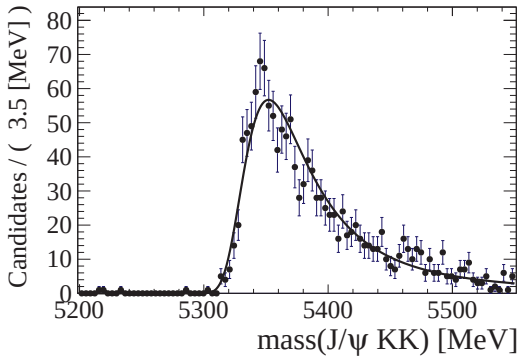


Figure 5.5: *Mass distribution of fully simulated* $\mathrm{B}^0 \to \mathrm{J/\psi\,K^*(892)^0}$ *MC events reconstructed under the* $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ *hypothesis.*

5.3 Peaking backgrounds

The main sources of peaking background in the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ analysis are considered to be from $\mathrm{B^0} \to \mathrm{J/\psi\,K^*(892)^0}$ and $\mathrm{B_c^\pm} \to \mathrm{B_s^0}X^\pm$ decays. These are discussed in more detail the following sections.

5.3.1 $\mathrm{B^0} \to \mathrm{J/\psi\,K^*(892)^0}$ background

One possible source of background is introduced through incorrectly reconstructing $\mathrm{B^0} \to \mathrm{J/\psi\,K^*(892)^0}$ events under the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ hypothesis in data. The mass distribution of fully simulated $\mathrm{B^0} \to \mathrm{J/\psi\,K^*(892)^0}$ MC events reconstructed under the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ hypothesis is shown in Figure 5.5. This distribution has a peak in the $\mathrm{B_s^0}$ signal region and a small contribution in the upper mass side-band of $m(\mathrm{J/\psi\ K^+K^-}) \in [5400, 5550]$ MeV/c^2.

In order to determine how many of these background events are present in data, events in the upper mass side-band, $m(\mathrm{J/\psi\ K^+K^-}) \in [5400, 5550]$ MeV/c^2, are reconstructed under the $\mathrm{J/\psi\,K\pi}$ hypothesis. The resulting mass distribution, $m\,(\mathrm{J/\psi\,K\pi}) \in [5200, 5320]$ MeV/c^2, is plotted in Figure 5.6.

In this figure the combinatorial background is described by a second order polynomial and the $\mathrm{B^0}$ component is described by a Gaussian function.This figure shows a small resonant component at the $\mathrm{B^0}$ mass ($m_{\mathrm{B^0}} \approx 5280\,\mathrm{MeV}/c^2$).

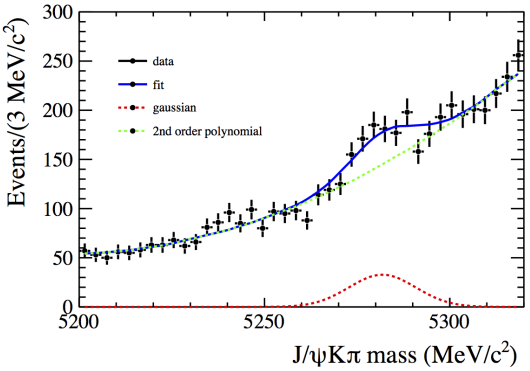


Figure 5.6: *This mass distribution shows* J/ψ KK *events in the upper mass sideband reconstructed under the* J/ψ Kπ *hypothesis.*

From this fit 245 ± 55 B^0 background events are found in the upper mass sideband of $m(J/\psi\ K^+K^-) \in [5400, 5550]$ MeV/c^2. Using Figure 5.5 the number of B^0 candidates in the total B_s^0 dataset is estimated to be 780 ± 170 in the mass range $m(J/\psi\ K^+K^-) \in [5200, 5550]$ MeV/c^2.

The decay time and angular distributions of the $B^0 \to J/\psi\, K^*(892)^0$ MC events reconstructed under the $B_s^0 \to J/\psi\ K^+K^-$ hypothesis are shown in Figure 5.7. From this figure the angular distribution of these events appears to be similar to the distribution of data signal events in Figure 5.4 (a). These events have a similar average lifetime to the signal, however they will not have the fast B_s^0 modulation.

As the total number of events from this background contribution is small compared to the signal yield this background is ignored in any further fits in this thesis. A systematic error introduced through not describing this background in the fit to data is discussed in Section 8.2.

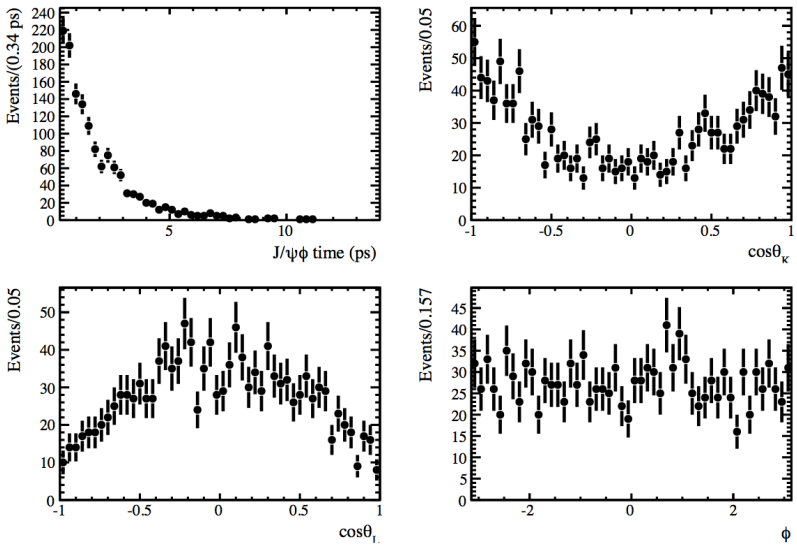


Figure 5.7: *The decay time and angular distributions of* $\mathrm{B}^0 \to \mathrm{J/\psi\,K^*(892)^0}$ *MC events reconstructed under the* $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ *hypothesis.*

5.3.2 Backgrounds from $\mathrm{B}_c^{\pm}$

A possible source of backgrounds in this analysis are from the decays of $\mathrm{B}_c^{\pm}$ mesons. A search for events from the decay of $\mathrm{B}_c^{\pm} \to \mathrm{B_s^0}\pi^{\pm}$ is performed using the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ dataset. Additionally an upper limit on the rate of $\mathrm{B}_c^{\pm} \to \mathrm{B_s^0}X^{\pm}$ is also estimated.

$\mathrm{B}_c^{\pm} \to \mathrm{B_s^0}\pi^{\pm}$ background

One possible source of this background is a contribution of non-prompt $\mathrm{B_s^0}$ mesons from the decay $\mathrm{B}_c^{\pm} \to \mathrm{B_s^0}X^{\pm}$. This background introduces an additional signal component with a different decay time distribution compared to the prompt because of the additional $\mathrm{B}_c^{\pm}$ decay.

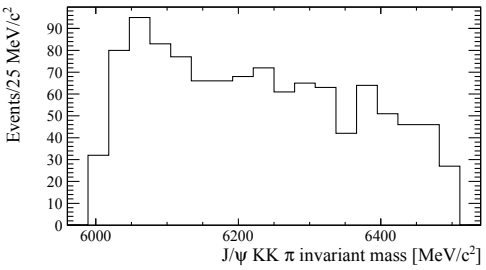


Figure 5.8: *Invariant mass distribution formed through adding an additional pion to the primary vertex from* $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ *events.*

To explore any contribution from this background the mass spectrum of $\mathrm{J/\psi\, K^+K^-\pi^\pm}$ is plotted in Figure 5.8. The mass for each event in this distribution is formed by adding an additional pion to the primary vertex of a $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ candidate. The additional pion must pass the criteria that $p_\mathrm{T} > 250\,\mathrm{MeV}/c$ and that its χ^2 about the best primary vertex is > 4. From this figure no resonance is observed at the $\mathrm{B_c^\pm}$ mass, $m(\mathrm{B_c^\pm}) \approx 6277\,\mathrm{MeV}/c^2$ [7]. There is no significant peaking contribution observed from this channel.

The analysis of $\mathrm{B_c^\pm} \to \mathrm{B_s^0}X^\pm$ in reference [58] finds 35 ± 8 $\mathrm{B_c^+} \to \mathrm{B_s^0}\,(\to \mathrm{J/\psi\ \phi})\,\pi^+$ events in $3\,\mathrm{fb}^{-1}$ of events collected by LHCb during 2011 and 2012. The selection in this analysis is loser than that for the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ analysis and as such this serves as an upper limit of the number of events expected from this decay in the dataset. Assuming that 12 ± 3 events were collected in the $1.0\,\mathrm{fb}^{-1}$ from 2011, this corresponds to $(0.04 \pm 0.01)\,\%$ of the dataset. The tights cut on the quality of the $\mathrm{B_s^0}$ impact parameter ($\chi^2_{\mathrm{IP,next}}(\mathrm{B_s^0}) > 50$ Table 4.6) reduces the number of $\mathrm{B_s^0}$ mesons originating from a $\mathrm{B_c^\pm}$ selected within the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ dataset.

As no significant peaking background is observed in this channel it is ignored in any further fits in this thesis. Any systematic uncertainty introduced through this decision is discussed in Section 8.2.

$\mathbf{B_c^\pm \to B_s^0 X^\pm}$ background

As $B_c^\pm \to B_s^0 \pi^\pm$ is only one possible decay mode of $B_c^\pm \to B_s^0 X^\pm$ the total number of B_s^0 mesons originating from a B_c^+ has to be estimated. The ratio of the number of B_s^0 mesons from B_c^+ compared to the number of prompt B_s^0 mesons is estimated to be:

$$\frac{N\left(B_c^\pm \to B_s^0 X^\pm\right)}{N\left(B_s^0\right)} = 0.6\% \ . \tag{5.4}$$

This quantity is estimated making use of the hadronisation fractions f_q and knowledge of cross sections and branching fractions,

$$\begin{aligned}
\frac{f_c}{f_s} &= \frac{\sigma\left(B_c\right)}{\sigma\left(B_s^0\right)} \\
&= \frac{f_c}{f_u} \Big/ \frac{f_s}{f_u} \\
&= \left[\left(\frac{\sigma\left(B_c\right) \times Br(B_c \to J/\psi\, \pi)}{\sigma\left(B^+\right) \times Br(B^+ \to J/\psi\, K^+)} \right) \times \frac{Br(B^+ \to J/\psi\, K^+)}{Br(B_c \to J/\psi\, \pi)} \right] \Big/ \frac{f_s}{f_u} \\
&= \left[0.68\% \times \frac{0.1016\%}{0.291\%} \right] \Big/ 0.27 \\
&= 0.0088 \ .
\end{aligned} \tag{5.5}$$

Using $\left(\dfrac{\sigma\left(B_c\right) \times Br(B_c \to J/\psi\, \pi)}{\sigma\left(B^+\right) \times Br(B^+ \to J/\psi\, K^+)} \right) = 0.68\%$ from Ref [59], $Br(B^+ \to J/\psi\, K^+) = 0.1016\%$ from Ref [7], $Br(B_c \to J/\psi\, \pi) = 0.291\%$ from Ref [60] and $\dfrac{f_s}{f_u} = 0.27$ from Ref [61].

From this, and taking the conservative branching fraction $Br(B_c^\pm \to B_s^0 X^\pm) = 70\%$, the amount of B_s^0 mesons from B_c^+ is expected to be:

$$\begin{aligned}
\frac{N\left(B_c^\pm \to B_s^0 X^\pm\right)}{N\left(B_s^0\right)} &= \frac{\sigma\left(B_c\right)}{\sigma\left(B_s^0\right)} \times Br\left(B_c^\pm \to B_s^0 X^\pm\right) \\
&= 0.0088 \times 70\% \\
&= 0.6\%
\end{aligned} \tag{5.6}$$

Fitting to simulated datasets with 0.6% of the signal events originating from $B_c^{\pm} \to B_s^0 X^{\pm}$ shows that no significant bias is introduced through ignoring this in fitting the $B_s^0 \to J/\psi\, K^+K^-$ PDF.

"Measure with a micrometer, mark with chalk, cut with an axe."

Ray's Rule of Precision

6

Fitting Methodology

This chapter describes the fitting process which is used to extract a set of physics parameters which most optimally describe a dataset.

A PDF describing all of the effects modifying the distribution of the observables is constructed in Section 6.1.

The physics parameters describing data are extracted using a Negative Log Likelihood (NLL) function. The construction of this function is described in Section 6.2. In the analysis of $B_s^0 \to J/\psi\, K^+K^-$, sWeights from the sPlot [52] method (Section 5.2) are combined with the NLL function to form an sFit as described in Section 6.3.

External measurements are used in the full analysis to constrain several free parameters within the fit. This is used to propagate the uncertainties from the external measurement of various parameters into the fit. The full set of constraints and how they are applied is reviewed in Section 6.4.

The behaviour of the S-wave and P-wave amplitudes across the $m(K^+K^-)$ range differ and so the analysis must be performed in multiple bins across this range. This is described in more detail in Section 6.5.

6.1 PDF for fitting to $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$

In order to describe the distribution of $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ signal events in data a probability density function (PDF) describing all of the observable distributions needs to be constructed. A PDF describes the probability of an event being in a given range within the phase space. These are defined to return a finite probability p in the range of $[0, 1]$, with a normalised integral of unity.

The tagging calibration is included in the PDF describing data through the prescription given in equation 6.1. The differential decay rate for $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$, $(\frac{d\Gamma_{\mathrm{B_s^0}}\left(t,\vec{\Omega}\right)}{dt\, d\vec{\Omega}})$ in equation 2.19, is denoted by $X\left(t,\vec{\Omega}\right)$ and for $\overline{\mathrm{B_s^0}} \to \mathrm{J/\psi\ K^+K^-}$ by $\overline{X}\left(t,\vec{\Omega}\right)$. Using these definitions the PDF describing an event with tag d is denoted $\mathcal{S}\left(t,\vec{\Omega}|d,\eta\right)$ and is constructed as,

$$\begin{aligned}
\mathcal{S}\left(t,\vec{\Omega}|+1,\eta;\vec{\mathcal{P}}\right) &= \left[\left(1-\omega\left(\eta\right)\right)X\left(t,\vec{\Omega}\right)+\overline{\omega}\left(\eta\right)\overline{X}\left(t,\vec{\Omega}\right)\right]\ , \\
\mathcal{S}\left(t,\vec{\Omega}|-1,\eta;\vec{\mathcal{P}}\right) &= \left[\omega\left(\eta\right)X\left(\vec{\Omega},t\right)+\left(1-\overline{\omega}\left(\eta\right)\right)\overline{X}\left(t,\vec{\Omega}\right)\right]\ , \\
\mathcal{S}\left(t,\vec{\Omega}|\quad 0,\eta;\vec{\mathcal{P}}\right) &= \frac{1}{2}\left[X\left(t,\vec{\Omega}\right)+\overline{X}\left(t,\vec{\Omega}\right)\right]\ .
\end{aligned} \tag{6.1}$$

Here the mistag calibrations $\omega\left(\eta\right)$ and $\overline{\omega}\left(\eta\right)$ are described by equation 4.16 and are defined by different tagging parameters for the different taggers used. The parameter $\vec{\mathcal{P}}$ contains all of the parameters required to define both the differential decay rate and the mistag calibrations. Using these functions the PDF describing the differential decay rate as well as detector acceptance and decay time resolution is defined as,

$$\mathrm{PDF}\left(t,\vec{\Omega}|d,\eta;\vec{\mathcal{P}},s_t\right) = \frac{\left[\mathcal{S}\left(t,\vec{\Omega}|d,\eta;\vec{\mathcal{P}}\right)\otimes R\left(t|\sigma_t;s_t\right)\right]\times\varepsilon\left(t,\vec{\Omega}\right)}{\iint dt\, d\vec{\Omega}\ \left[\mathcal{S}\left(t,\vec{\Omega}|d,\eta;\vec{\mathcal{P}}\right)\otimes R\left(t|\sigma_t;s_t\right)\right]\times\varepsilon\left(t,\vec{\Omega}\right)} \tag{6.2}$$

The decay time resolution function is described in Section 4.6, and the detector acceptance function is described in Section 4.7.

The full PDF describing data is invariant under the transform in equation 6.3. This invariance leads to two degenerate minimum values for the PDF function.

$$\left(\phi_{\rm s},\, \Delta\Gamma_{\rm s}\,,\, \delta_{\parallel},\, \delta_{\perp},\, \delta_{\rm S},\, \delta_0\right) \rightarrow \left(\pi - \phi_{\rm s},\, -\Delta\Gamma_{\rm s}\,,\, -\delta_{\parallel},\, \pi - \delta_{\perp},\, -\delta_{\rm S},\, -\delta_0\right)\ . \tag{6.3}$$

6.2 Performing a fit to a dataset

The likelihood ($\mathcal{L}$) of a given dataset is described as the product of the PDF values for each event. This is a function of the physics parameters $\vec{\mathcal{P}}$,

$$\mathcal{L}\left(\vec{\mathcal{P}}\right) = \prod_{i}^{N} \text{PDF}\left(\vec{x}_i; \vec{\mathcal{P}}\right)\ . \tag{6.4}$$

The aim of performing a fit is to extract the set of physics parameters $\vec{\mathcal{P}}$ corresponding to the maximum likelihood for a given dataset. Here $\vec{x}_i$ contains all of the observable quantities required by the PDF for event i in a dataset of N events.

6.2.1 Maximum likelihood fitting

The physics parameters ($\vec{\mathcal{P}}$) which best describe a dataset correspond to the global maximum of the likelihood function in equation 6.4. Although it is possible to determine the physics parameters using 6.4 a Negative Log Likelihood (NLL) function is more commonly used,

$$-\ln\left[\mathcal{L}\left(\vec{\mathcal{P}}\right)\right] = -\sum_{i}^{\rm N} \ln\left[\text{PDF}\left(\vec{x}_i; \vec{\mathcal{P}}\right)\right]\ . \tag{6.5}$$

This is a one to one transform, hence a minimum of the NLL corresponds to a maximum of equation 6.4. The best fit is then obtained by varying the physics parameters $\vec{\mathcal{P}}$ to determine the global minima of the NLL. With a minima of the function satisfying the condition:

$$-\frac{d}{d\vec{\mathcal{P}}} \sum_{i}^{\rm N} \ln\left[\text{PDF}\left(\vec{x}_i; \vec{\mathcal{P}}\right)\right] = 0\ . \tag{6.6}$$

The MINUIT [62] minimisation package is commonly used, with the search for the minimum performed by the MIGRAD minimiser [62]. This package explores the behaviour of the likelihood within the phase space described by the vector $\vec{\mathcal{P}}$. This attempts to find the best global minima corresponding to the condition in equation 6.6.

When constructing an NLL function using the PDF is described by equation 6.2, the numerator becomes factorisable,

$$-\ln\left[\mathcal{L}\left(\vec{\mathcal{P}}\right)\right] = -\sum_{i}^{\mathrm{N}} \ln\left[\frac{\left[\mathcal{S}\left(t,\vec{\Omega}|d,\eta;\vec{\mathcal{P}}\right)\otimes R\left(t|\sigma_t;s_t\right)\right]}{\iint dt\, d\vec{x}\ \left[\mathcal{S}\left(t,\vec{\Omega}|d,\eta;\vec{\mathcal{P}}\right)\otimes R\left(t|\sigma_t;s_t\right)\right]\times\varepsilon\left(t,\vec{\Omega}\right)}\right] - \sum_{i}^{\mathrm{N}} \ln\left[\varepsilon\left(t,\vec{\Omega}\right)\right] \tag{6.7}$$

Due to this and the fact the function $\varepsilon\left(t,\vec{\Omega}\right)$ is independent of any of the physics parameters $\vec{\mathcal{P}}$, the term describing detector acceptance in the numerator adds a constant offset to the NLL function. As a constant value does not affect the function minima (equation 6.6) this term can be ignored in constructing the PDF. This means that the effect of detector acceptance only has to be included in the denominator of equation 6.2 when determining the function minima. It is still necessary to include this term in the numerator when projecting the PDF over a dataset.

6.2.2 Error estimation

In addition to finding the optimal set of physics parameters it is important to determine the uncertainty of any values which are extracted from the fit. Taking the NLL function to be well behaved and continuous it can be described using a Taylor expansion. This expansion about the function minima $\left(\vec{\mathcal{P}}_{min}\right)$ is,

$$-\ln\mathcal{L}\left(\vec{\mathcal{P}}\right) = f\left(\vec{\mathcal{P}}\right) \approx \begin{bmatrix} f\left(\vec{\mathcal{P}}_{\mathrm{min}}\right) & + & \left.\dfrac{d\,f\left(\vec{\mathcal{P}}\right)}{d\vec{\mathcal{P}}}\right|_{\vec{\mathcal{P}}_{\mathrm{min}}}\left(\vec{\mathcal{P}}-\vec{\mathcal{P}}_{\mathrm{min}}\right) \\ & + \dfrac{1}{2!}\left.\dfrac{d^2 f\left(\vec{\mathcal{P}}\right)}{d\vec{\mathcal{P}}^2}\right|_{\vec{\mathcal{P}}_{\mathrm{min}}}\left(\vec{\mathcal{P}}-\vec{\mathcal{P}}_{\mathrm{min}}\right)^2 \\ & + & \mathcal{O}\left(\vec{\mathcal{P}}-\vec{\mathcal{P}}_{\mathrm{min}}\right)^3 \end{bmatrix},$$

$$\approx f\left(\vec{\mathcal{P}}_{\min}\right) + 0 + \frac{1}{2}\left.\frac{d^2 f\left(\vec{\mathcal{P}}\right)}{d\vec{\mathcal{P}}^2}\right|_{\vec{\mathcal{P}}_{\min}} \left(\vec{\mathcal{P}} - \vec{\mathcal{P}}_{\min}\right)^2 \ . \tag{6.8}$$

Equation 6.8 makes use of the fact that the first order derivative is defined to be 0 at the minima (equation 6.6). From this approximation the likelihood function for one free parameter a is shown in equation 6.9. Here $a_{\min}$ is the global minima for this free parameter and $f(a)$ the NLL function for a.

$$\begin{aligned}
\mathcal{L}(a) &\approx \exp\left[-f\left(a_{\min}\right)\right] \exp\left[-\frac{1}{2}\left(a - a_{\min}\right)^2 \Big/ \left(\left.\frac{d^2 f(a)}{da^2}\right|_{a_{\min}}\right)^{-1}\right] \ . \\
\mathcal{L}(a) &\approx A \exp\left[-\left(a - a_{\min}\right)^2 \Big/ 2\sigma^2(a)\right] \ . \\
\therefore \sigma^2(a) &= \left(\left.\frac{d^2 f(a)}{da^2}\right|_{a_{\min}}\right)^{-1} \ .
\end{aligned} \tag{6.9}$$

From this equation it can be seen that the likelihood function for a well behaved parameter is expected to approach a Gaussian, with the uncertainty on a, $\sigma(a)$, given by the second order derivative of the NLL function about the minimum $a_{\min}$.

When performing a fit to a complex PDF it is common for multiple physics parameters to be correlated. This is not accounted for in equation 6.9 since this only refers to a single parameter. In the case of multiple parameters the second order variance of the NLL function is described by a (Hessian) matrix, evaluated about the minima. The uncertainty on a parameter is therefore calculated using the inverse of this matrix. In the case of a fit extracting n physics parameters this is a Hessian matrix $\mathbf{H}$ of dimension $n \times n$. This matrix is composed of the second order partial derivatives of the NLL function for each combination of parameters. From this the error on the ith parameter, $\mathcal{P}_i$, is calculated from the (i,i)-th element of the inverted Hessian matrix.

$$\sigma\left(\mathcal{P}_i\right) = \sqrt{\left(\mathbf{H}^{-1}\right)_{i,i}} \ . \tag{6.10}$$

When calculating the error from data the tool most commonly used is called Hesse [62]. This tool calculates the Hessian matrix ($\mathbf{H}$) about the function minima. Hesse explicitly relies upon the assumption that the calculated matrix

must be positive definite. When this is not the case HESSE will attempt to force the Hessian matrix to become positive definite which often leads to inaccurate results.

There is an alternate definition for the error on a parameter. If the parameter a is translated by one unit of its uncertainty, $\sigma(a)$, then NLL function rises by $\delta = 1/2$,

$$
\begin{aligned}
-\ln\left[\mathcal{L}(a)\right] &= -\ln\left[\exp\left(f\left(a_{\min}\right)\right)\exp\left(-\frac{1}{2}\frac{\left(a-a_{\min}\right)^2}{\sigma(a)^2}\right)\right] \ . \\
a &\to a_{\min} + \sigma(a) \ , \quad \ln\mathcal{L}(a) \to \ln\mathcal{L}\left(a_{\min}\right) - \delta \ , \\
-\left[\cancel{\ln\mathcal{L}\left(a_{\min}\right)} - \delta\right] &= -\left[\cancel{f\left(a_{\min}\right)} - \frac{1}{2}\frac{\left(a_{\min}+\sigma(a)-a_{\min}\right)^2}{\sigma(a)^2}\right] \ , \quad \delta = \frac{1}{2} \ . \qquad (6.11)
\end{aligned}
$$

This alternative definition of the error from the NLL function is used by the MINOS [62] tool to determine the error on each parameter. The MINOS tool explores possible values in each parameter separately. As it does this it re-minimises the full NLL function to find the point where the function rises by $1/2$ due to varying a single parameter.

The two techniques to determine the error on a parameter are therefore:

- Invert the Hessian matrix and take the square root of the correct matrix element, as in equation 6.10.
- Find the point where the NLL function rises by $1/2$ by varying the given parameter, as in equation 6.11.

When a parameter has a Gaussian likelihood function the two definitions of the error are identical. Throughout the rest of this thesis (unless expressly stated otherwise) the errors quoted have been determined using the HESSE tool.

6.3 sWeighted fitting

The process used to calculate sWeights for a given dataset is described in Section 5.2. The result of this is a set of per-event weights w_i which are used to statistically remove the background from a dataset. This removes the need to describe the background species in the distributions of the observables.

In the case of $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ the $m(\mathrm{J/\psi\ K^+K^-})$ observable is used to calculate the per-event weights, with the distribution of weights shown in Figure 5.2.

When performing an sWeighted fit these weights are used to modify the calculation of the NLL. This modified form of the NLL (equation 6.5),

$$-\ln\left[\mathcal{L}\left(\vec{\mathcal{P}}\right)\right] = -\alpha \sum_{i}^{\mathrm{N}} w_i \ln\left[F\left(\vec{x}_i;\vec{\mathcal{P}}\right)\right] \ . \tag{6.12}$$

Performing a fit using this weighted function is commonly referred to as performing an sFit [63]. In the case that the dataset is un-weighted ($\omega_i \equiv 1$ and $\alpha \equiv 1$) equation 6.12 reduces to equation 6.6. Here, the per-event sWeights are incorporated in the NLL as w_i along with a correction factor α.

When sWeights are included in the definition of the NLL they may modify the corresponding Hessian matrix. The effect of per-event weights on the Hessian matrix are discussed in Ref [64]. This α parameter in equation 6.12 is used to correct the Hessian matrix. Within the RooFit [40, 41] framework α is a matrix which is used to scale the Hessian matrix $\mathbf{H}$ to extract the correct errors. The calculation in RooFit which defines α is given by,

$$\alpha = \mathbf{H}\,\mathbf{G}^{-1} \ . \tag{6.13}$$

Here $\mathbf{H}$ is the normal Hessian matrix from the sFit NLL function and $\mathbf{G}$ is the Hessian matrix derived from a modified form of equation 6.12, replacing w_i with w_i^2 and defining $\alpha = 1$. The calculation of $\mathbf{G}$ is non-trivial and can introduce technical problems. A simpler approach to calculate α is to use,

$$\alpha \approx \sum_{i}^{N} w_i \Big/ \sum_{i}^{N} w_i^2 \ . \tag{6.14}$$

Both of these methods for calculating α yield the same results when the sWeights are able to separate well the signal and background species.

It is possible to fit to multiple datasets with independent sWeights within a fit. As such a different values of α must be calculated for each set of weights. For simplicity the latter approach (equation 6.14) is used to calculate α in the full fit for $B_s^0 \to J/\psi\, K^+K^-$.

6.4 External constraints

When performing a fit some parameters are constrained from external measurements. This information is introduced in the fit to data through the use of an external constraint functions.

As the NLL function allows independent fit functions to be linearly added each of these external constraints are included as extra terms added to the total NLL function. These terms take the form shown in equation 6.15. This equation represents a constraint on a given parameter a with the central value $a_{\min}$ and error $\sigma(a)$,

$$-\ln \mathcal{L}(a) = -\left[\frac{(a - a_{\min})^2}{\sigma(a)^2}\right] . \tag{6.15}$$

The main use of external constraints is to propagate the uncertainties of externally measured parameters into the statistical uncertainties of other parameters which are determined in the fit. The full list of external constraints which are used in fitting to $B_s^0 \to J/\psi\, K^+K^-$ are given in Table 6.1.

Parameter	External Measurement
Δm_s	$17.63 \pm 0.11\,\mathrm{ps}^{-1}$ [65]
s_t	1.45 ± 0.06
$p_0\,(OS)$	0.392 ± 0.008
$p_1\,(OS)$	1.00 ± 0.023
$\Delta p_0\,(OS)$	0.011 ± 0.0034
$p_0\,(SS)$	0.350 ± 0.017
$p_1\,(SS)$	1.00 ± 0.16
$\Delta p_0\,(SS)$	-0.019 ± 0.005
$p_0\,(OS+SS)$	0.000 ± 0.0025
$\Delta p_0\,(OS+SS)$	-0.011 ± 0.004

Table 6.1: *External Constraints applied in the fit to* $\mathrm{B_s^0} \to \mathrm{J/\psi\,K^+K^-}$. *The constraint on* Δm_s *comes from another analysis [65] whilst the constraints on time resolution scale factor* s_t *and the tagging calibrations are described in Sections 4.6 and 4.8, respectively.*

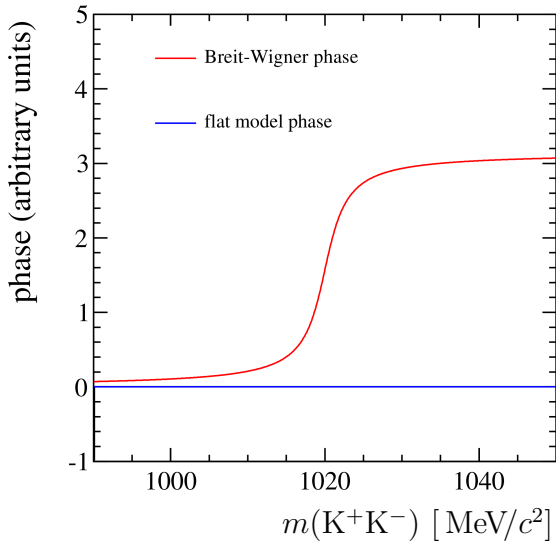


Figure 6.1: *Strong Phase behaviour of a Breit-Wigner function (red) and a flat amplitude (blue) across the $m(\mathrm{K^+K^-})$ range [55].*

6.5 Correction of relative S-P wave phase difference

When constructing a fit to data, the strong phases of the P-wave and S-wave amplitudes have different behaviours across the range $m(\mathrm{K^+K^-}) \in [990, 1050]$ MeV/c^2. The strong phase functions are shown in Figure 6.1. Across the $m(\mathrm{K^+K^-})$ range the S-wave strong phase is relatively flat while the P-wave varies dramatically across the ϕ resonance ($m(\phi) = 1020\,\mathrm{MeV}/c^2$ [7]). The physical solution of the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ analysis corresponds to a negative trend in the S-P strong phase difference across the $m(\mathrm{K^+K^-})$ range. The behaviour of the strong phases across the $m(\mathrm{K^+K^-})$ range was first examined in the $\cos 2\beta$ analysis at BarBar as described in Ref [66] and was applied to the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ analysis in Ref [67].

The behaviour of the strong phases as a function of $m(\mathrm{K^+K^-})$ is not described the in the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ decay rate. This $m(\mathrm{K^+K^-})$ dependence can be expressed as $g\left(m_{\mathrm{K^+K^-}}\right)$ for the P-wave amplitudes and $v\left(m_{\mathrm{K^+K^-}}\right)$ for the S-wave. Integrating over a finite range of $m(\mathrm{K^+K^-})$ for the different strong phases terms yields:

$$\text{where: } i,j \in (0,\|,\perp) \ ,$$

$$\int_{m^1_{\mathrm{K^+K^-}}}^{m^2_{\mathrm{K^+K^-}}} dm_{\mathrm{K^+K^-}}\ \mathrm{A}_i\mathrm{A}_j^*\left|g\left(m_{\mathrm{K^+K^-}}\right)\right|^2 = \left|\mathrm{A}_i\right|\left|\mathrm{A}_j^*\right| \ ,$$

$$\int_{m^1_{\mathrm{K^+K^-}}}^{m^2_{\mathrm{K^+K^-}}} dm_{\mathrm{K^+K^-}}\ \mathrm{A_S}\mathrm{A_S}^*\left|v\left(m_{\mathrm{K^+K^-}}\right)\right|^2 = \left|\mathrm{A_S}\right|^2 \ ,$$

$$\int_{m^1_{\mathrm{K^+K^-}}}^{m^2_{\mathrm{K^+K^-}}} dm_{\mathrm{K^+K^-}}\ \mathrm{A}_i\mathrm{A_S}^* g\left(m_{\mathrm{K^+K^-}}\right) v^*\left(m_{\mathrm{K^+K^-}}\right) = \left|\mathrm{A}_i\right|\left|\mathrm{A_S^*}\right| C_{\mathrm{S-P}} e^{-i\theta_{\mathrm{SP}}} \ . \quad (6.16)$$

From these integrals Terms 1-7 in Table 2.3 are unaffected, however an additional factor of $C_{\mathrm{S-P}}e^{-i\theta_{\mathrm{SP}}}$ is present in terms 8-10. These 3 terms of the differential decay rate are multiplied by the $C_{\mathrm{S-P}}$ factor and the complex phase $-\theta_{\mathrm{SP}}$ is absorbed into the difference of the strong phases in the decay rate.

The P-wave function $g\left(m_{\mathrm{K^+K^-}}\right)$ is taken to be a relativistic Breit-Wigner function and the S-wave is taken to be constant in $m(\mathrm{K^+K^-})$. The calculation of these $C_{\mathrm{S-P}}$ factors is described in Ref [12], for a single bin of $m(\mathrm{K^+K^-}) \in [990, 1050]$ MeV/c^2, $C_{\mathrm{S-P}} \approx 0.5$. To keep $C_{\mathrm{S-P}}$ close to 1, the full $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ range is split into six-bins in $m(\mathrm{K^+K^-})$. The list of the $m(\mathrm{K^+K^-})$ bin ranges and the corresponding $C_{\mathrm{S-P}}$ factors are given in Table 6.2.

Figure 6.2 shows these bin ranges overlaid on the $m(\mathrm{K^+K^-})$ distribution for all events passing the final selection (Section 4.3). The choice of binning is centred around the nominal ϕ mass (1020 MeV/c^2 [7]). The binning in the centre of the $m(\mathrm{K^+K^-})$ range is finest due to the large number events and the rapidly changing strong phase across the resonance. At the outer edges of the $m(\mathrm{K^+K^-})$ range the bins are largest due to the lower number of events and the slowly varying strong phases.

When performing a fit to data using these six bins, only the S-wave parameters are independent for each bin, with all other free parameters taken to be common to all bins. This results in six different S-wave phases and amplitudes measured in the full fit to data. These will be reported in Chapter 8.

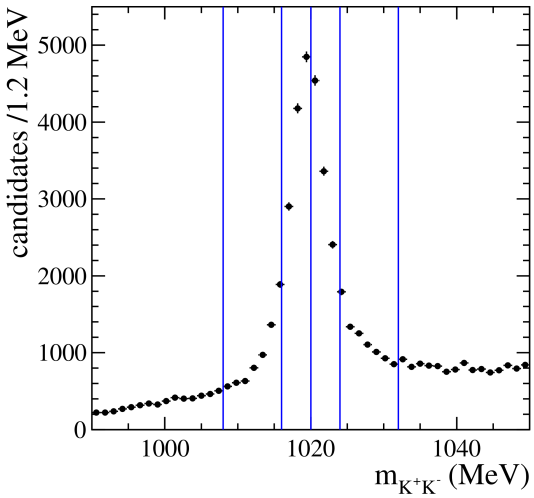


Figure 6.2: *The bin edges of the six bins in $m(\mathrm{K^+K^-})$ chosen for the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ analysis [55]. The $m(\mathrm{K^+K^-})$ range has been divided into six bins. These are symmetrically arranged around the nominal ϕ mass ($1020\,\mathrm{MeV}/c^2$ [7]). This histogram contains all of the events passing the final selection described in Section 4.3.*

	$m(\mathrm{K^+K^-})$ range (MeV/c^2)	$C_{\mathrm{S-P}}$
bin 1	990–1008	0.966
bin 2	1008–1016	0.956
bin 3	1016–1020	0.926
bin 4	1020–1024	0.926
bin 5	1024–1032	0.956
bin 6	1032–1050	0.966

Table 6.2: *List of S-wave bin edges and $C_{\mathrm{S-P}}$ factors.*

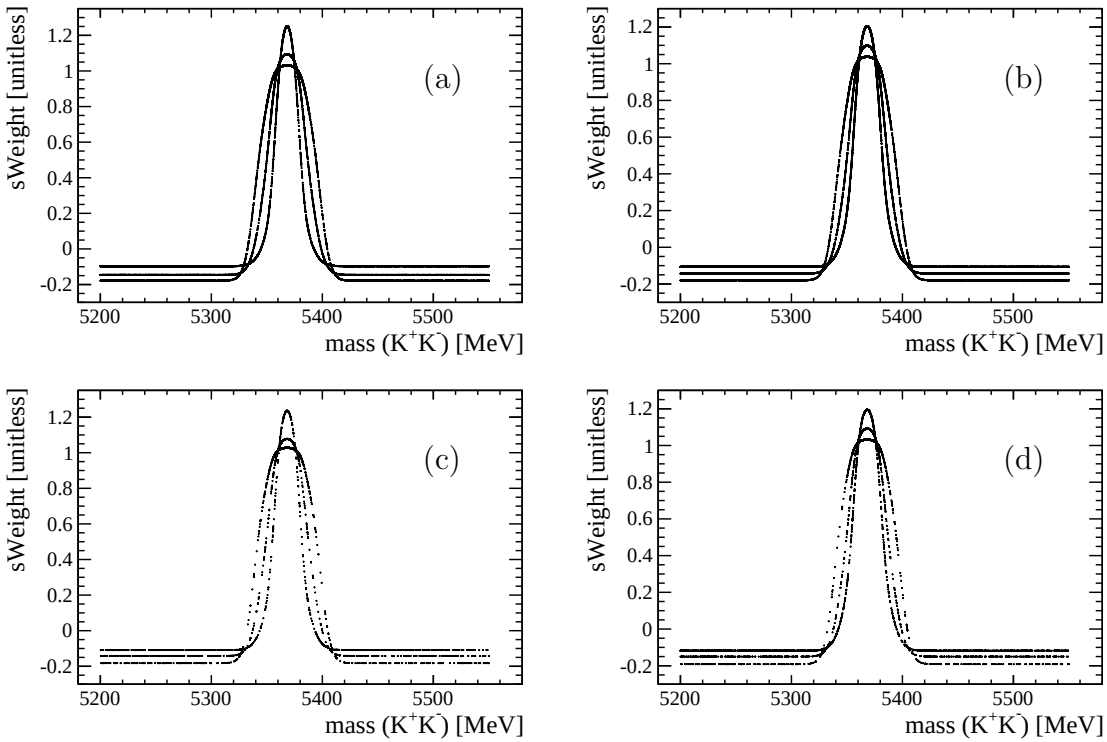


Figure 6.3: *Distribution of sWeights vs* $m(\mathrm{J/\psi\,K^+K^-})$ *for: (a)* $990MeV < m(\mathrm{K^+K^-}) < 1020MeV$ *for decay time unbiased events, (b)* $1020MeV < m(\mathrm{K^+K^-}) < 1050MeV$ *for decay time unbiased events, (c)* $990MeV < m(\mathrm{K^+K^-}) < 1020MeV$ *for decay time biased events, and (d)* $1020MeV < m(\mathrm{K^+K^-}) < 1050MeV$ *for decay time biased events. The three separate lines are from the different bins of* $m(\mathrm{K^+K^-})$ *described in the text.*

	Signal Yield	α from sWeights
bin 1	545 ± 34	0.5658
bin 2	2772 ± 46	0.8632
bin 3	9815 ± 80	0.9564
bin 4	8118 ± 73	0.9478
bin 5	3301 ± 53	0.8547
bin 6	2016 ± 57	0.6617
total	26567 ± 117	n/a

Table 6.3: *Summary of the yield and corresponding α values from the six-bin mass fits.*

6.5.1 Six bin $m\left(\mathrm{K^+K^-}\right)$ mass fit

In order to correctly perform the analysis in six $m(\mathrm{K^+K^-})$ bins, the $m(\mathrm{J/\psi\ K^+K^-})$ fit, described in Section 5.1, is repeated with a different signal fraction and background model in each bin. The fit is constructed with an independent background description and signal fraction in each bin and a common signal description across all six bins. The results of this fit are described in more detail in Appendix C. The calculated yield in agreement with the result quoted in Section 5.1.

The set of calculated sWeights in each bin are independent. Due to this, a separate α correction factor is calculated for each bin using equation 6.14. The distribution of sWeights vs $m(\mathrm{J/\psi\ K^+K^-})$ for all events is shown in Figure 6.3. Here, the *almost unbiased* dataset is split between Figures 6.3 (a) and (b) and the *exlusively biased* dataset between Figure 6.3 (c) and (d). The choice of binning allows for the distributions of sWeights for each bin in $m(\mathrm{K^+K^-})$ to be clearly viewed. The 3 separate distributions of sWeights are visible as 3 lines in each subplot. From this the distribution of sWeights in each bin is similar to the distribution calculated using a single bin of $m(\mathrm{K^+K^-})$ in Figure 5.2.

A summary of the signal yield in each $m(\mathrm{K^+K^-})$ bin and the corresponding α values calculated from the sWeight are given in Table 6.3.

6.6 $\mathbf{B_s^0 \to J/\psi\, K^+K^-}$ fitting review

The fit to $B_s^0 \to J/\psi\, K^+K^-$ involves minimising the NLL function to determine the set of physics parameters $\vec{\mathcal{P}}$ corresponding to the best description the $1.0\,\text{fb}^{-1}$ dataset. Background is removed from this analysis through the use of an *s*Fit [63] as described in Section 6.3. This fitting approach removes the need to describe the background species in the NLL. In addition a set of external constraints (Section 6.4) are also applied in the fit; this propagates the uncertainties from the mistag and proper time, and Δm_s calibrations into the fit.

For reference later the full set of parameters used in the fit for $B_s^0 \to J/\psi\, K^+K^-$ are described below.

Of the three P-wave amplitudes only two of them are independent, this allows for $\left|A_\parallel(0)\right|^2$ to be defined as,

$$\left|A_\parallel(0)\right|^2 = 1 - \left(|A_0(0)|^2 + |A_\perp(0)|^2\right) \ . \tag{6.17}$$

The strong phases between the decay amplitudes are such that only the differences between the phases are measurable. By convention $\delta_0 = 0$, allowing the other phases to be measured relative to this.

Instead of fitting directly for the S-wave amplitude $|A_S(0)|^2$ the S-wave fraction F_S is used,

$$F_S = \frac{|A_S(0)|^2}{1 + |A_S(0)|^2} \ . \tag{6.18}$$

From fitting to simulated datasets, the strong phase $\delta_\perp$ and δ_S are known to be highly correlated parameters at $\approx 90\%$. Hence, the parameter δ_S is replaced by $\delta_{S-\perp}$ which is defined as,

$$\delta_{S-\perp} = \delta_S - \delta_\perp - \theta_{SP} \ . \tag{6.19}$$

The physical observables described by the $B_s^0 \to J/\psi\, K^+K^-$ differential decay rate are,

$$\vec{x} = \left\{t, \vec{\Omega}\right\} \ . \tag{6.20}$$

The full set of physics parameters,

$$\vec{\mathcal{P}} = \left\{ \Gamma_{\rm s}\,, \Delta\Gamma_{\rm s}\,, |{\rm A}_0\,(0)|^2, |{\rm A}_\perp(0)|^2, F_{\rm S},\, \delta_\perp,\, \delta_\parallel,\, \delta_{{\rm S}-\perp},\, \phi_s,\, |\lambda_s|\,,\, \Delta m_s \right\}\ . \quad (6.21)$$

Where, there are six S-wave fractions $F_{\rm S}$ and phases $\delta_{{\rm S}-\perp}$, one for each bin in $m({\rm K}^+{\rm K}^-)$. The nuisance parameters are,

$$\vec{\mathcal{P}}\,(nuisance) = \begin{array}{r} \{p_0\,(OS)\,, p_1\,(OS)\,, \Delta p_0\,(SS)\,, p_0\,(SS)\,, p_1\,(SS)\,, \Delta p_0\,(SS)\,, \\ p_0\,(OS+SS)\,, \Delta p_0\,(OS+SS)\,, s_t\} \end{array}\ . \quad (6.22)$$

"The last bug isn't fixed until the last user is dead."

Sidney Markowitz

7

RapidFit Framework

RapidFit is a general purpose XML driven fitting framework developed and used within the LHCb groups at Edinburgh and EPFL Universities. The framework was originally developed by another Edinburgh PhD student Ben Wynne before he moved to the ATLAS experiment. Throughout my PhD I have worked extensively on the RapidFit fitting framework assuming the role of lead developer. My contributions toward RapidFit have been varied and most notably include:

- Implementing component projections for user PDFs
- Parallelising the existing technologies
- Development of the RapidPlot tool for post-processing and analysis

In this chapter I will discuss the general structure of the RapidFit fitting framework. I will also describe the information which a user must provide in order to use this framework to perform a fit to extract a set of physics parameters.

I will also describe the RapidPlot tool used alongside the fitting framework. This tool is used in the analysis and post-processing of results from fits.

7.1 Introduction to RapidFit

RapidFit is a fitting framework developed within Edinburgh University primarily around the needs of the $B_s^0 \to J/\psi\, K^+K^-$ analysis. The design of RapidFit is heavily influenced by the Object Orientated [68] approach to computing. The RapidFit architecture is composed of two independent binaries named fitting and RapidPlot. The fitting binary is the executable which contains all of the code which performs the complex fitting and contains the users' PDF. The RapidPlot binary is the executable which is responsible for the post-processing and presentation and complex formatting of the results from fitting. However, RapidFit is not intended to be a full drop-in replacement for the more comprehensive RooFit fitting framework [69] which is distributed as part of the ROOT [40, 41] architecture.

As well as the fitting binary being able to be built locally it is also possible to build RapidFit as an external ROOT library. This allows for large scale analyses to be distributed across computing clusters where ROOT is supported.

An overview of the key elements within RapidFit framework is given in Section 7.1.1.

Due to its modular design this framework supports multiple complex analyses. Each analysis involves two user provided components.

1. User written PDF
   This is a user written PDF component describing the actual calculation to be performed for each DataPoint. Information on the requirements of a users' PDF are described in Section 7.1.2.

2. Fitting XML
   This is a user written XML file which drives the actual fit which RapidFit is to perform. This contains the information required to configure RapidFit to perform an analysis using a dataset. The information that must be provided by the XML is described in Section 7.1.3.

One of the main advantages of this design is that multiple fit configurations can be captured and developed independently to the PDF development. It also allows the same fit to be replicated between users sharing the same XML.

7.1.1 Overview of RapidFit

In discussing RapidFit a list of important C++ objects and their descriptions are given in Table 7.1. The flow of information between these main objects within RapidFit is shown in Figure 7.1.

When fitting to a dataset a "FitFunction" needs to be constructed which can be used with a chosen "Minimiser" to extract a set of physics parameters. This FitFunction used in this thesis is the negative log likelihood function (NLL) discussed in Chapter 6. This FitFunction knows how to evaluate a set of given external constraints as well as how to evaluate a PDF for a each event in a dataset, based on a given "ParameterSet". The design of the RapidFit architecture allows for multiple PDFs and datasets to be fit in parallel, this allows for complex analyses involving multiple channels to be constructed by adding multiple fit functions together.

The requirements of the FitFunction are grouped such that the "PhysicsBottle" contains all of the information required to evaluate a given set of functions based on datasets and the "ParameterSet" which describes that function. The functions included within the PhysicsBottle include both external constraints and C++ PDF functions written by the user. The data sets which are evaluated over each of the user written PDFs are pointed to as IDataSet objects which contain both the measured events and the phase space which contains them.

In RapidFit the functions which are evaluated correspond to either external constraints or PDFs with associated datasets. The external constraints applied in $B_s^0 \to J/\psi\, K^+K^-$ are described in Section 6.4, and the combined PDFs and datasets are described in Sections 6.1 and 4.3 respectively.

The requirements of the C++ PDF provided by the user are outlined in Section 7.1.2 and the requirements of the user written XML are described in Section 7.1.3.

Term	Definition
PDF	User Written C++ class inheriting from BasePDF object in RapidFit
Observable	A per-event physical observable of some quantity, e.g. mass/time
DataPoint	A collection of Observable objects for each event in a dataset
IDataSet	An interface class for a container for DataPoint objects in a dataset. This is typically a MemoryDataSet container for many DataPoints.
IConstraint	An interface class for the range of possible values for a given physical observable eg $m_{\mathrm{B_s^0}} \in [5200, 5550]$ MeV/c^2, $d = \pm 1$.
PhaseSpaceBoundary	A collection of IConstraint objects describing the entire Phase-Space of the dataset
PhysicsBottle	Contains all of the Information required to construct a FitFunction
PhysicsParameter	The Physics Parameter a PDF depends upon e.g. Γ_s , $\Delta\Gamma_\mathrm{s}$
ParameterSet	A collection of PhysicsParameter objects
FitFunction	The FitFunction which is being evaluated by a minimiser i.e. χ^2 or NLL Function
Minimiser	The tool being used to minimise a given FitFunction over a dataset typically Minuit or Minuit2 toolset
FitResult	A collection of all of the Output from a fit, very similar to RooFitFitResult
ResultParameter	Contains the information on the range, central value and uncertainty of a physics parameter after performing a fit
ResultParameterSet	A collection of ResultParameter objects

Table 7.1: *Common RapidFit C++ objects*

In addition to these high level descriptions a minimalist analysis is described in Appendix E which provides an example C++ PDF and XML configuration file.

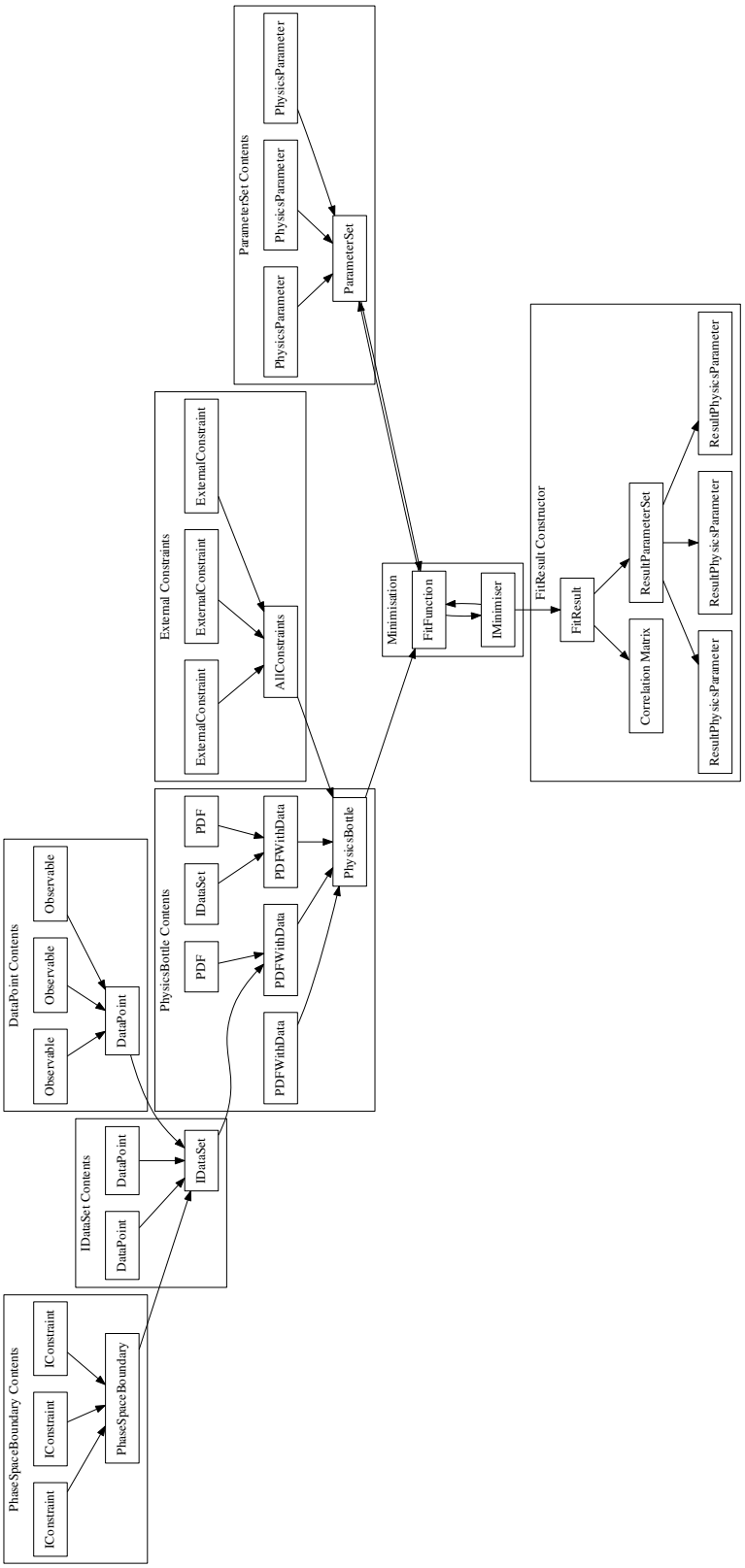


Figure 7.1: *Flow of information through* RapidFit *during a fit.*

7.1.2 User written PDFs

The user written PDF in RapidFit is a C++ class which is intended to inherit from the BasePDF object. These user written PDFs are required to provide a certain amount of information to the fitting framework in order for it to be used in RapidFit.

This information is:

- List of observables described by the PDF. e.g. mass, time
- The required PhysicsParameters. e.g. Γ_s , $\Delta\Gamma_s$
- A function to evaluate.
- How to analytically normalise the function. (Optional)

The first two pieces of this information must be provided by the PDF class upon construction. This allows the fitting framework to check that the appropriate physics parameters and observables are provided by the user at runtime. In addition to the list of observables that the PDF describes it must also inform the framework of observables which it is conditional upon. An example of these are the mistag calibration parameters in the $B_s^0 \to J/\psi\ K^+K^-$ analysis.

The third piece of information provided by the PDF is what to calculate for a given DataPoint. This is provided through an implementation of an Evaluate method within the PDF which returns a *double* value. This function is expected to be well behaved, however, the framework is designed to provide minimal checking of this output should this method return an invalid value or throw an exception.

The fourth piece of information a PDF can provide is a method which is able to calculate the normalisation factor for a given DataPoint and PhaseSpaceBoundary. This information is provided by the user implementing a Normalise method within their PDF. As it it not always possible to analytically integrate complex functions this method is strictly optional. Providing a method to normalise a PDF has two advantages. Firstly it is less computationally intensive to analytically integrate a function and secondly this avoids problems with computational accuracy often associated with numerical integration. Should a users PDF not provide a normalisation method then the numerical integral is calculated automatically within RapidFit.

Although the PDF is required to provide these four pieces of information there are many other virtual functions in the BasePDF which the users PDF may overload to provide additional functionality, however discussion of these is beyond the scope of this thesis.

More complex PDF

In addition to the user providing their own PDFs some common PDFs are distributed as part of the RapidFit. In addition to these PDFs the framework is also able to combine these functions using multiplication and addition.

This is achieved through the ProdPDF and NormalisedSumPDF classes which are constructed from other PDFs in RapidFit. These PDF take the form shown in equation 7.1.

$$\begin{aligned} \text{ProdPDF}\left(\vec{y},\vec{z}\right) &= \frac{F\left(\vec{y}\right) \times F'\left(\vec{z}\right)}{\int d\vec{y}\, d\vec{z}\;\; F\left(\vec{y}\right) \times F'\left(\vec{z}\right)} \\ \text{NormalisedSumPDF}\left(\vec{y},\vec{z}\right) &= f_1 \left[\frac{F\left(\vec{y}\right)}{\int d\vec{y}\, F\left(\vec{y}\right)}\right] + \left(1-f_1\right)\left[\frac{F'\left(\vec{z}\right)}{\int d\vec{z}\, F'\left(\vec{z}\right)}\right] \end{aligned} \tag{7.1}$$

Here the functions $F\left(\vec{y}\right)$ and $F'\left(\vec{z}\right)$ can be any of the PDF evaluate functions known to RapidFit.

The ProdPDF as an example here can be used to perform a combined fit to the mass distribution in combination with the decay time and angular PDF. The NormalisedSumPDF can be used to construct a fit to data which contains both signal and background species.

7.1.3 User written XML files

As RapidFit is configured primarily through the use of a fitting XML there is a minimal set of information that it must provide.

- What type of FitFunction to use
- Which Minimiser to use
- Where to start the fit
- What Dataset(s) and PDF(s) to use
- Which studies are to be performed (optional)

When fitting to data it is important to know what FitFunction is being used. Within RapidFit there are multiple FitFunctions, however there are mostly different implementations of the Negative Log Likelihood (NLL) function (Chapter 6) which have been modified to address specific technical challenges.

The user is able to request the minimiser used to find the minimum of the fit function. Normally the Minuit fitting suite [62] is used when fitting to data, however an interface to the Minuit2 suite [70] is also implemented within the framework.

When performing a fit to data, choosing the starting position for the physics parameters can be important. Choosing starting values close to the expected true minima reduces the computational requirements of the fit. It is also important to choose the parameters to be in a region where the FitFunction is well behaved.

The users XML also provides information on which PDF to use in an analysis and which dataset the PDF describes. Typically the data is provided in a ROOT TTree format with the relevant information for fitting read into memory. The XML interface to configure reading in from a TTree allows for a final selection to be placed on the dataset. In addition to this the XML interface has been designed to allow for the stored TTree to be transparently converted as the data is read from disk. This allows the users PDF class to be written in a way that it doesn't depend on the format of the actual dataset.

RapidFit is also capable of performing complex studies. These are configurable through extra options which are added to the users XML file. The list of typical studies which a user may perform are described in Sections 7.3 and 7.4.

7.2 PDF projections

Part of my work on RapidFit was to build in the capability of constructing projections from a users' PDF and overlaying it over a full dataset. These projections can represent the complete PDF or individual components of the function. In order for a PDF to project individual components it must provide knowledge of the components it knows about and a method to evaluate them.

The components a PDF can calculate is communicated to RapidFit by the PDFComponents method and the function which returns the values of each component is EvaluateComponent. This EvaluateComponent function accepts two arguments which are the requested component name and the DataPoint that the component is to be evaluated for. In the case that the user does not implement this method the BasePDF implementation provides knowledge of the 0-th PDF component (PDF total) and wraps the EvaluateComponent method to the Evaluate method.

Generally a users' PDF in RapidFit describes n observable quantities. To construct a projection of the PDF over one of its observables requires the framework to perform an integrations over the remaining $(n-1)$ dimensions. To perform this projection, RapidFit makes use of numerical integration. The complex ProdPDF and NormalisedSumPDF objects provided by the fitting framework fully support the projection of individual components allowing the user to construct complex component plots from simple PDFs and an XML.

In order to integrate a PDF the phase-space which the data occupies has to be known. It is possible to construct a phase space in RapidFit which contains both discrete and continuous observables. An example of a discrete observable is the flavour tagging decision in $B_s^0 \to J/\psi\, K^+K^-$ which has three possible values for the initial B_s^0 flavour. An example of a continuous observable is the decay time observable which may take any value between the pre-defined minimum and maximum bounds used in the analysis.

When calculating these projections RapidFit calculates a separate projection for each discrete combination which is described by the PDF. The PDF being projected may also be conditional on a given set of observables which are not described. When this is the case the numerical integration is performed using the fixed mean value calculated from the distribution of a given observable.

In the case of $B_s^0 \to J/\psi\ K^+K^-$ the PDF is conditional upon the per-event mistag (η) and time resolution (σ_t) observables. In this case the integration is performed using the values of $\eta \approx 0.37$ and $\sigma_t \approx 45\,\text{fs}$. The result of projections for the $B_s^0 \to J/\psi\ K^+K^-$ analysis are shown in Section 8.1.1 .

7.3 Fast-MC studies in RapidFit

To perform fast Monte Carlo (fast-MC) studies RapidFit makes use of the Foam [71] MC event generator which is distributed as part of the ROOT framework. This generator is able to adapt to any PDF written by the user and has excellent performance. The toy datasets are transient in nature and are generated and stored when required. Fitting to many datasets generated this way is commonly referred to as a fast-MC study.

Analysing the results from a such a study provides two important pieces of information. Firstly, a MC study provides a powerful cross-check to search for any potential fit biases. The fit bias is defined as the difference between the generated value and the optimal value extracted from a given dataset. Secondly, a MC study provides a reliable estimate of the sensitivity of a fit to the various physics parameters.

One of the important features of the implementation of Foam within RapidFit is the reproducibility of the fit results. This is due to the fact that Foam makes use of a pseudo-random number generator from within the ROOT framework.

Using these random number generators allows RapidFit to be able to reproduce the results from a fast-MC fit exactly which serves as a powerful diagnostic tool. As random numbers are only used in the generation of fast-MC events, different studies can be performed which can fit a different PDF to the same transient datasets. The results from a fast-MC study using the full $B_s^0 \to J/\psi\ K^+K^-$ PDF are shown in Appendix F.

7.4 Likelihood profiles

An important part in understanding the results from fitting to data is to determine the behaviour of the NLL function about the extracted minima. This function is defined to be well behaved when it has a single minima and rises parabolically about that minima. The parameters the likelihood profile is calculated for are referred to as the 'control parameters'.

The process to construct a likelihood profile is as follows:

1. Perform a global fit to data.
2. Change the control physics parameter(s) and fix their values.
3. Re-minimise the fit function to obtain the minimum NLL.
4. Repeat steps 2-3 until the function has been defined over a given range.

The choice of where to start the fits in step 3 depends on the implementation within the framework, this generally takes one of two approaches. The first approach is to start with parameter values close to the global minima to determine the best likelihood any other given point. (This is the approach that is taken by the Minuit fitting package.) This approach is susceptible to becoming stuck in a local minima. The second approach is to perform a full minimisation starting from a position far from the global/local minima. The second approach consumes more computing resources, but is far less susceptible to becoming stuck in spurious local minimas. Likelihood scans in RapidFit are performed with the minimisation at each point started from the same starting coordinates in the user written XML. This allows the user to control the behaviour of the scans for a faster or more stable result.

One important requirement of all of the fits is that the result should be reliable. In order to guarantee that this is the case RapidFit requires that the fit has correctly converged at each point with a well defined correlation matrix. Should the fit not reach this criteria the framework will attempt to refit at different coordinates close to where the fit failed. This approach of attempting to recover from a failed fit gives the best coverage possible for unstable likelihood functions.

The scanning technology in RapidFit is user configurable and the fit is always restarted from the location defined in the user written XML. Both 1D and 2D Scans in RapidFit are performed by the '*fitting*' binary which returns a ROOT file containing the result from re-minimising the NLL at each point in the control parameter. The result from this is then passed to the '*RapidPlot*' binary which performs the post-analysis and plotting of the likelihood function.

The result of constructing 1D likelihood profiles for the physics parameters in $B_s^0 \to J/\psi\ K^+K^-$ are shown in Appendix D.

7.5 Work on RapidFit performance

Throughout my PhD I have worked extensively on improving the performance of the RapidFit fitting framework. This work involves many different aspects with the most notable gains in performance being described in this section.

7.5.1 Optimising fit performance

A full angular time dependent $B_s^0 \to J/\psi\ K^+K^-$ fit to the $1.0\,fb^{-1}$ dataset takes approximately 4-5 hours on a modern single CPU core ($\approx$ 3GHz clock-speed). The vast majority of the calculations performed during this time are related to the per-event time resolution. When performing the full $B_s^0 \to J/\psi\ K^+K^-$ analysis this became a significant technical hurdle which had to be addressed.

The simplest way that a FitFunction can evaluate a dataset is to use the user written C++ PDF to evaluate each DataPoint in series. This is normally adequate for simple analyses, however, for analyses with large amounts of data and/or complex PDFs this can often become limited by the speed of the CPU core the program is running on. One way to overcome this barrier is to exploit the fact that modern computers have access to multiple CPU cores. In order for RapidFit to make use of these additional resources a large part of my work has been to allow for the users PDF to be distributed across many CPU cores using multiple processing threads to evaluate large datasets in parallel.

The effect of this work has been to bring the length of time required to perform a fit to data down from 4-5 hours on one CPU core to 30 Min with the work spread across 8 cores. Although this performance improvement is most noticeable in the analysis of $B_s^0 \to J/\psi\ K^+K^-$ the framework has been written in a generic enough way to allow for all RapidFit analyses to benefit from these improvements.

7.5.2 Numerical Integration

To numerically integrate a PDF the function has to be evaluated for many different coordinates within a given phase-space.

The simplest approach to optimising numerical integration is to optimise the sampling of the phase-space to reduce the number of coordinates that the PDF must be evaluated at. This is the approach taken by the AdaptiveIntegratorMultiDim tool distributed with the ROOT framework. This is a single threaded numerical integration tool which samples the phase-space depending upon the behaviour of the user's function. One major advantage of this is that it does not oversample in regions where the PDF is slowly varying, but this comes at the cost of being constrained to running on one CPU core.

An alternative approach to improve the performance of numerical integration is to distribute the computational load of sampling the phase-space. To do this RapidFit samples the phase-space using many coordinates calculated in a pre-determined way using the GSL libraries [72]. This sampling of the phase-space is performed using the Sobol [73] random number sequence which adapts to the number of dimensions its sampling. This sequence allows for the phase-space to be sampled uniformly in n dimensions, allowing an integral to be calculated in a brute-force way. The coordinates generated using this algorithm are uniformly distributed within each observable.

Once these coordinates have been pre-calculated the fitting framework is able to use multiple threads to sample the phase-space using the PDF. This multi-threaded numerical integration has been implemented in a transparent way such that user written code does not have to worry about how many threads are being used. This brings a significant performance boost to complex operations such as PDF projections which rely on numerical integration to project PDFs over one Observable.

7.5.3 Highly distributed analyses

Although the performance of RapidFit running individual fits can be improved locally using techniques such as multi-threading, ultimately larger analyses need to be distributed across systems with a large number of CPU cores. One of the advantages of RapidFit being build upon the ROOT framework is that it allows it to be compiled as an external ROOT library. Using this ROOT library in combination with the LHCb software framework (most notably the Ganga [74] grid frontend) allows RapidFit to be built once and run across many systems supporting the ROOT framework. This is advantageous in the case of large scale distributed analyses such as running many complex MC studies for the $1.0\,\mathrm{fb}^{-1}$ $\mathrm{B}_\mathrm{s}^0 \to \mathrm{J}/\psi\ \mathrm{K}^+\mathrm{K}^-$ analysis.

Taking the $\mathrm{B}_\mathrm{s}^0 \to \mathrm{J}/\psi\ \mathrm{K}^+\mathrm{K}^-$ as an example, as a single fit takes approximately 5 hours on a single CPU, producing a 2D likelihood contour from sampling the PDF at 20×20 points would take approximately three months on a single CPU. Making use of Ganga [74] and the computing GRID [47, 48], these calculations can be performed in about 10-12 hours. In this case running on a highly distributed platform represents a significant performance improvement by a factor of 150.

7.6 Post Processing with RapidPlot

When performing a single fit to a dataset RapidFit creates a folder with a unique time-stamp. This folder contains a summary of the fit result in LaTeX format and a ROOT file. This ROOT file contains the fit result and the runtime configuration which produced it, including a copy of the user-written XML. If projections of the fit results onto the dataset are requested then these are also stored in this output folder.

When a more complex analysis is performed using the RapidFit framework an additional ROOT output file is produced. This file contains the results from the many fits involved in a complex study and additional information allowing the results to be post-processed by the RapidPlot tool. An important feature of this output format is that the results from various stages of these analyses can be stored in separate files which can be merged before post-processing.

The RapidPlot tool is designed to inspect the output ROOT file from RapidFit. This tool is able to determine the appropriate post-processing which should be performed.

7.6.1 Post-processing toy study results

The output from performing a toy study contains (i) the measured central values, (ii) the estimated sensitivities and (iii) per-dataset pulls, where the pull is defined as,

$$\text{pull} = \frac{\text{Measured Value} - \text{Generated Value}}{\text{Estimated Uncertainty}} \quad . \tag{7.2}$$

For a toy study large numbers of each of these three values are measured and their distributions be plotted. It is assumed that they follow a Gaussian distribution. Using this assumption the data is plotted in a binned histogram with the optimal bin number determined according to D. Scott's method [75]. Using a histogram with this number of bins a fit is performed to extract the mean and width of each of these three distributions. The behaviour of the central value and pull distributions are important for understanding how a parameter is behaving in a fit, however, the pull distribution is the most useful parameter for checking that a fit is consistent.

It is expected that parameters which are well behaved in fits will have pull distributions which are centred at 0 with a width of 1. Should the pull distribution not match these criteria then different statements about the fit can be made.

- Pull distribution central value $\neq 0$
  When the distribution is centred significantly far from zero the parameter is biased.

- Width of pull distribution > 1
  When the width of the distribution is wider than 1 the uncertainty of the parameter in the fit has likely been underestimated in individual fits to datsets.

- Width of pull distribution < 1
  When the width of the distribution is narrower than 1 the uncertainty of the parameter in the fit has likely been overestimated in the fits to individual datasets.

When extracting the mean and width of the distribution of results the tool attempts to fit a Poisson, Gaussian and Gamma function to the dataset. The function which returns the lowest χ^2/nDoF is taken to be the best fit function and is used to define the mean and width of the distribution. For well behaved parameters the distributions of value, estimated error and pull are described by a Gaussian function. However, parameters with a true value close to a boundary condition have distributions best described by an asymmetric Poisson or Gamma function.

7.6.2 Likelihood Profiles

Once RapidFit has explored the behaviour of the likelihood function by repeatedly fitting with different values the results are saved in a ROOT output file. This output file contains enough information for the RapidPlot tool to determine if a 1D or 2D likelihood profile was constructed and from this the appropriate post-processing is applied to the data in the file. The chosen parameters for which the likelihood function is being explored are referred to as control parameters.

To explore the likelihood for given control parameter the FitFunction is minimised at various pre-defined coordinates for the control parameter. The information stored for each of these fits includes the full set of fit results in addition to the absolute NLL at each point. From this the ΔLL is defined as the difference between the local NLL and the global best NLL for each coordinate.

1D likelihood functions

For a well behaved fit the a 1D likelihood function is defined to be parabola with a clearly defined minima with ΔLL values of 0.5 at $\pm\sigma$, as defined in Section 6.2.2.

A likelihood function is taken to be well behaved and continuous around the function minima. Hence when plotting the likelihood function based on sampled

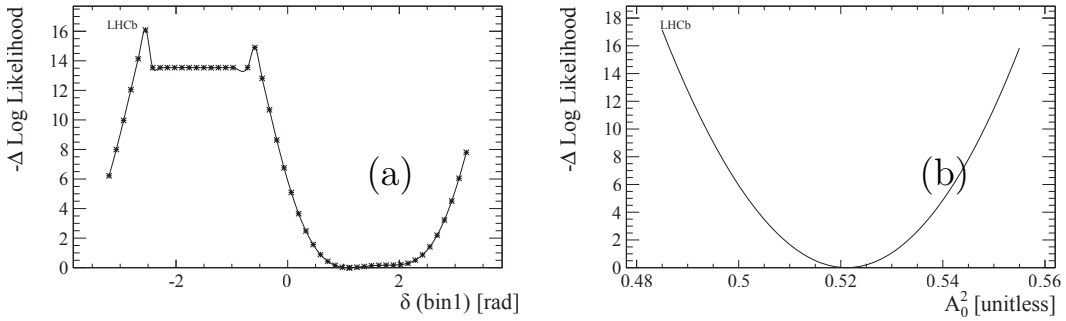


Figure 7.2: *1D likelihood profiles for (a) $\delta_{S-\perp}$ (bin1) and (b) $|A_0(0)|^2$ in the 1.0 fb^{-1} $B_s^0 \to J/\psi\, K^+K^-$ analysis. The profile for $\delta_{S-\perp}$ is not not parabolic with a clear minima, however the profile for $|A_0(0)|^2$ is parabolic with a single global minima.*

DataPoints it is approximated in ROOT using a third order spline function. This function connects the measured ΔLL values with a continuous smooth line.

For example, the likelihood functions corresponding to $\delta_{S-\perp}$ and $|A_0(0)|^2$ from a simulated $B_s^0 \to J/\psi\, K^+K^-$ dataset are shown in Figure 7.2. This figure shows that for well behaved parameters such as $|A_0(0)|^2$ the fit function is indeed well behaved and parabolic, whilst for poorly behaved parameters such as $\delta_{S-\perp}$ the function may not be parabolic in nature.

Although not shown here, the projection tool also plots the variation of the nuisance parameters which correspond to the NLL as a function of the control parameter. The plots produced this way are useful for understanding the complex behaviour of fits.

2D likelihood contours

2D Likelihood Contours are constructed from minimising over a pre-defined surface for two physics parameters. To construct the contours which correspond to the confidence limits from the likelihood function, the data points must be interpolated to produce a smooth plane. This step is performed within ROOT using a process known as Delaunay[1] triangulation to construct a 2D surface based on the measured ΔLL values. The 2D profiles are then constructed as contours of constant value lying on this surface.

[1]Named after B. Delaunay for his work on the subject [76].

1D ΔLL	2D ΔLL	Confidence Limit
0.5	1.15	68%
2.0	2.36	90%
4.5	3.00	95%

Table 7.2: *ΔLL values which are required to calculate the 1D uncertainties or to construct the corresponding 2D Confidence Limits to form 2D likelihood projections.*

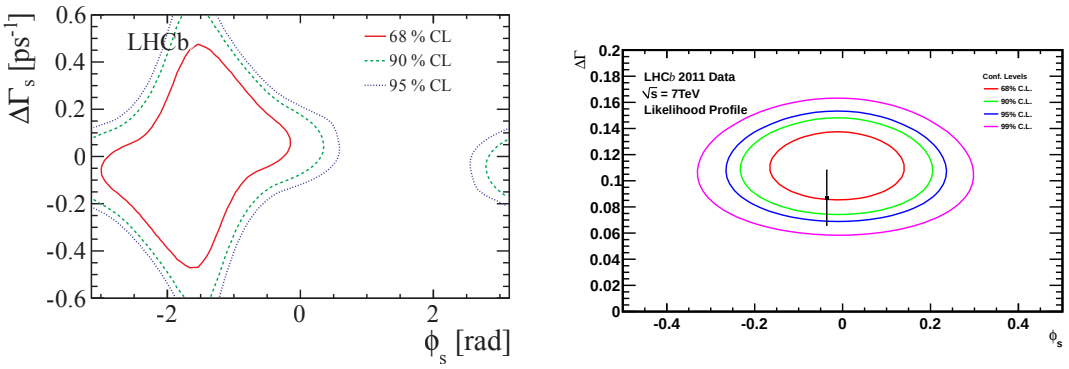


Figure 7.3: *2D likelihood profiles constructed using* RapidFit *of ϕ_s vs $\Delta\Gamma_s$ for datasets of (a) 36 pb^{-1} [77] and (b) 337 pb^{-1} [11] in size.*

Unlike the case of the 1D profile likelihood the contour lines correspond to different changes in the NLL function. The values of ΔLL and the confidence bands they represent are listed in Table 7.2.

Figure 7.3 shows 2D likelihood contours on the ϕ_s vs $\Delta\Gamma_s$ plane for $B_s^0 \to J/\psi\, K^+K^-$ analyses using (a) 36 ps^{-1} [77] and (b) 337 pb^{-1} [11] of data. The correlation between the two parameters was found to be, (a) $\rho(\Delta\Gamma_s, \phi_s) = 0.20$, (b) $\rho(\Delta\Gamma_s, \phi_s) = -0.08$. In the case of a small dataset the contours may not be well defined as in Figure 7.3 (a), however, when the parameters are well defined the contours will appear similar to Figure 7.3 (b).

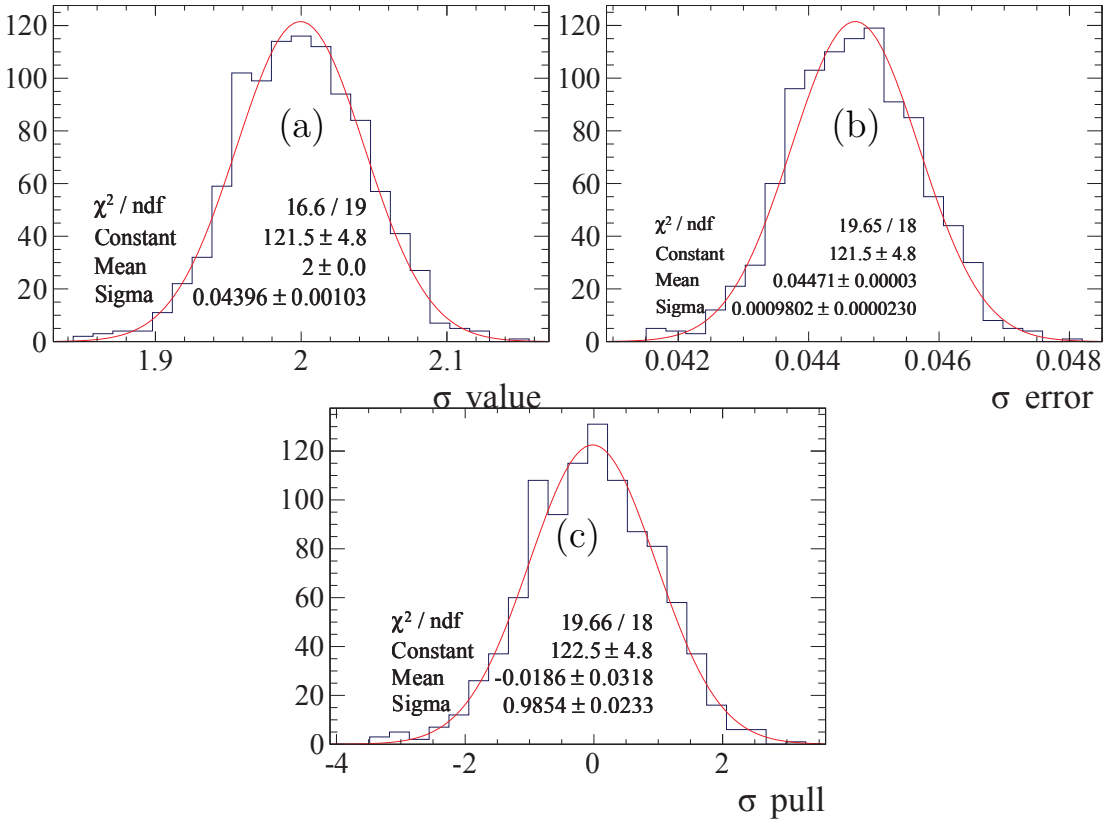


Figure 7.4: *The resulting distribution of the measured (a) value and (b) from fitting to width of a Gaussian distribution. From these fits the pull (equation 7.2) distribution is centred at 0 with a width of 1 indicating that the fits are unbiased.*

7.7 Validation of RapidFit architecture

A simple PDF composed of a Gaussian distribution with a measurable width is used to verify the RapidFit architecture. Using this PDF a fast-MC study fitting to 1,000 datasets each containing 1,000 events was performed and the results are shown in Figure 7.4.

From Figure 7.4, performing a fast-MC study using a simple PDF using the RapidFit architecture gives results which are unbiased with a pull distribution centred at 0 with a width of 1. This gives confidence that the fitting framework works as expected and introduces no bias when fitting to a dataset.

"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be."

Lord Kelvin

8

$\mathbf{B_s^0 \to J/\psi\ K^+K^-}$ Results

The results for all physics parameters measured in the $B_s^0 \to J/\psi\ K^+K^-$ analysis are summarised in Section 8.1. The systematic uncertainties are discussed in Section 8.2 and are summarised in Section 8.3. As this analysis is performed in six separate bins of $m(K^+K^-)$ the two-fold ambiguity of the PDF (Section 6.1) can be resolved, this is discussed in Section 8.4.

In addition to $B_s^0 \to J/\psi\ K^+K^-$ the $\mathcal{CP}$-violating phase ϕ_s can also be measured through the decay of $B_s^0 \to J/\psi\ \pi^+\pi^-$ as described in Ref [78]. By performing a combined fit to data from both the $B_s^0 \to J/\psi\ K^+K^-$ and $B_s^0 \to J/\psi\ \pi^+\pi^-$ decay channels it is possible to improve on the statistical sensitivity of ϕ_s and other parameters, this is discussed in Section 8.5.

8.1 Fit results for $B_s^0 \to J/\psi\ K^+K^-$

The central values and statistical uncertainties extracted from the likelihood fit for $B_s^0 \to J/\psi\ K^+K^-$ are shown in Table 8.1. The physics parameters at the top of table are the key physics parameters. The six bins correspond to the different ranges in $m(K^+K^-)$ as defined in Table 6.2. Each of these bins has an independent set of S-wave parameters but all other parameters are common to all bins in $m(K^+K^-)$. The nuisance parameters within the fit are allowed to vary with external constraints applied as described in Section 6.4. The correlation matrix from this fit is shown in Figure 8.1 with a reduced correlation matrix of the key physics parameters in Table 8.2.

The likelihood profiles corresponding to the key physics parameters are reviewed in Appendix D. Most of the likelihood profiles are parabolic with a single well defined minima, indicating that their uncertainty is Gaussian.

The likelihood profiles for $\delta_{\parallel}$ and the S-wave parameters F_S and $\delta_{S-\perp}$ are not parabolic. Due to this, the statistical uncertainties quoted for these parameters are asymmetric and are extracted using MINOS as described in Section 6.2.2.

The results presented in Tables 8.1 and 8.2 have been published in Ref [12]. From Table 8.2 it can be seen that there is only a small correlation between $\Delta\Gamma_s$ and ϕ_s, however, there are larger correlations between Γ_s , $\Delta\Gamma_s$ and the decay amplitudes.

8.1.1 $B_s^0 \to J/\psi\ K^+K^-$ projections

The projections of the signal PDF onto the data are shown in Figure 8.2. This figure shows the different decay time and angular distributions for the total PDF and the sub-component $\mathcal{CP}$-odd, $\mathcal{CP}$-even and S-wave decay amplitudes.

In this figure the histograms corresponding to the signal dataset are constructed using sWeights as described by the sPlot method in Section 5.2. The projected PDF is then normalised to the signal yield of this weighted dataset.

Although the analysis is performed in the helicity basis rather than the transversity (Section 2.5.3) the projections of the full PDF in both angular bases are shown.

	Parameter	Fit result and statistical uncertainty	
Key Physics Parameters	Γ_s	0.6631 ± 0.0048	$[\,\mathrm{ps}^{-1}]$
	$\Delta\Gamma_s$	0.100 ± 0.016	$[\,\mathrm{ps}^{-1}]$
	$\lvert A_\perp(0)\rvert^2$	0.24865 ± 0.0086	
	$\lvert A_0(0)\rvert^2$	0.5211 ± 0.0061	
	$\delta_\parallel$	$3.30\ ^{+0.13}_{-0.21}$	[rad]
	$\delta_\perp$	3.07 ± 0.22	[rad]
	Δm_s	17.669 ± 0.077	$[\,\mathrm{ps}^{-1}]$
	ϕ_s	0.067 ± 0.091	[rad]
	$\lvert\lambda_s\rvert$	0.943 ± 0.031	
bin 1 S-wave Parameters	F_S	$0.227\ ^{+0.081}_{-0.073}$	
	$\delta_{S-\perp}$	$1.31\ ^{+0.78}_{-0.49}$	[rad]
bin 2 S-wave Parameters	F_S	$0.067\ ^{+0.030}_{-0.027}$	
	$\delta_{S-\perp}$	$0.77\ ^{+0.38}_{-0.23}$	[rad]
bin 3 S-wave Parameters	F_S	$0.008\ ^{+0.014}_{-0.007}$	
	$\delta_{S-\perp}$	$0.51\ ^{+1.40}_{-0.30}$	[rad]
bin 4 S-wave Parameters	F_S	$0.016\ ^{+0.012}_{-0.009}$	
	$\delta_{S-\perp}$	$-0.51\ ^{+0.21}_{-0.35}$	[rad]
bin 5 S-wave Parameters	F_S	$0.055\ ^{+0.027}_{-0.025}$	
	$\delta_{S-\perp}$	$-0.46\ ^{+0.18}_{-0.26}$	[rad]
bin 6 S-wave Parameters	F_S	$0.167\ ^{+0.043}_{-0.042}$	
	$\delta_{S-\perp}$	$-0.65\ ^{+0.18}_{-0.22}$	[rad]
Nuisance Parameters (Section 6.4)	$p_0\,(OS)$	0.392 ± 0.008	
	$p_1\,(OS)$	1.000 ± 0.023	
	$\Delta p_0\,(OS)$	0.0111 ± 0.0034	
	$p_0\,(SS)$	0.355 ± 0.016	
	$p_1\,(SS)$	1.030 ± 0.16	
	$\Delta p_0\,(SS)$	-0.019 ± 0.005	
	$p_0\,(OS+SS)$	0.005 ± 0.024	
	$\Delta p_0\,(OS+SS)$	-0.011 ± 0.004	
	s_t	1.45 ± 0.06	

Table 8.1: *Central values and statistical uncertainties of the fitted parameters in the $B_s^0 \to J/\psi\, K^+K^-$ analysis. Here the value of Δm_s is constrained according to [65]. The central values of the Nuisance parameters are consistent with the external constraints described in Section 6.4.*

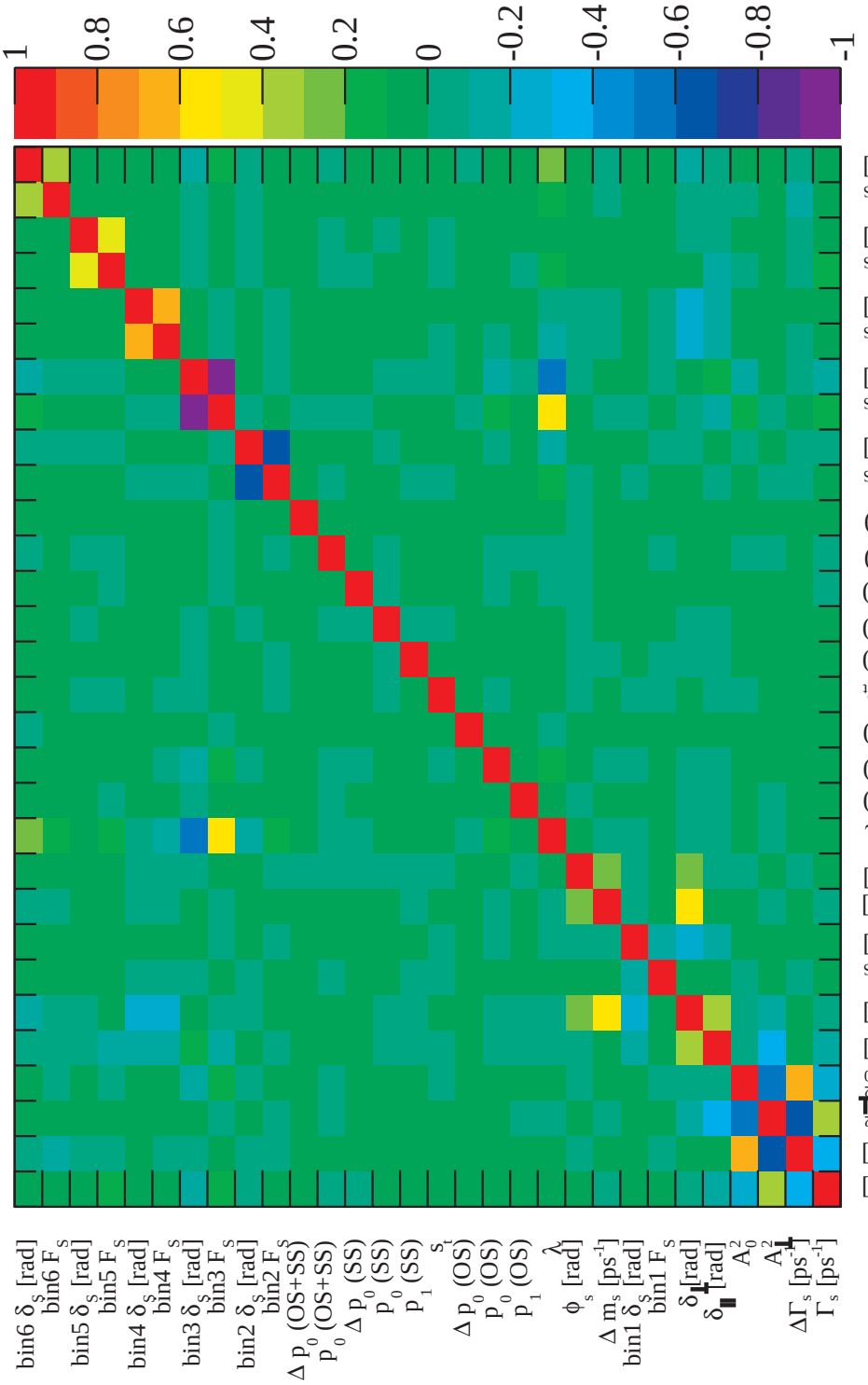


Figure 8.1: *Correlation matrix from* $B_s^0 \to J/\psi\, K^+K^-$ *likelihood fit.*

	Γ_s $[\mathrm{ps}^{-1}]$	$\Delta\Gamma_s$ $[\mathrm{ps}^{-1}]$	$\lvert A_\perp(0)\rvert^2$	$\lvert A_0(0)\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	ϕ_s [rad]	$\lvert\lambda_s\rvert$
Γ_s $[\mathrm{ps}^{-1}]$	1.00	−0.39	0.37	−0.27	−0.09	−0.03	0.06	0.03
$\Delta\Gamma_s$ $[\mathrm{ps}^{-1}]$		1.00	−0.68	0.63	0.03	0.04	−0.04	0.00
$\lvert A_\perp(0)\rvert^2$			1.00	−0.58	−0.28	−0.09	0.08	−0.04
$\lvert A_0(0)\rvert^2$				1.00	−0.02	−0.00	−0.05	0.02
$\delta_\parallel$ [rad]					1.00	0.32	−0.03	0.05
$\delta_\perp$ [rad]						1.00	0.28	0.00
ϕ_s [rad]							1.00	0.04
$\lvert\lambda_s\rvert$								1.00

Table 8.2: *Correlation matrix for key physics parameters.*

Note that θ_K in the helicity basis is the same as ψ_{tr} in the transversity basis.

The discontinuities in the PDF function are related to the bin edges of the angular acceptance histogram discussed in Section 4.7.1. These histograms are only used when projecting the PDF over the dataset and don't effect the numerical results in Table 8.1 as described in Section 6.3.

In these projections, the total PDF, shown by the black line, agrees well with the total distribution of decay time and angles. The red and green lines describe the $\mathcal{CP}$-even and $\mathcal{CP}$-odd components of the full PDF and clearly have different decay time and angular distributions. The small S-wave component in the dataset is also shown within these projections as the blue line.

In order to asses the goodness of fit the data was binned into a 4D with a binning of $(4 \times 4 \times 4 \times 4)$ (time $\times$ $\cos\theta_L \times \cos\theta_K \times \phi_H$) with a total of 256 bins across the whole phase-space. The result of projecting the full PDF across these bins yields $\chi^2/\mathrm{nDoF} = 1.03$ with a probability of $= 49.4\%$.

8.1.2 Fitting with Δm_s free

Although Δm_s is constrained within the full analysis using the result from Ref [65], it is possible to independently measure this parameter using this decay channel. In the expression for the decay rate of $B_s^0 \to J/\psi\, K^+K^-$ most terms including Δm_s are multiplied by $\sin\phi_s$ (Section 2.5). For small values of ϕ_s, determining Δm_s

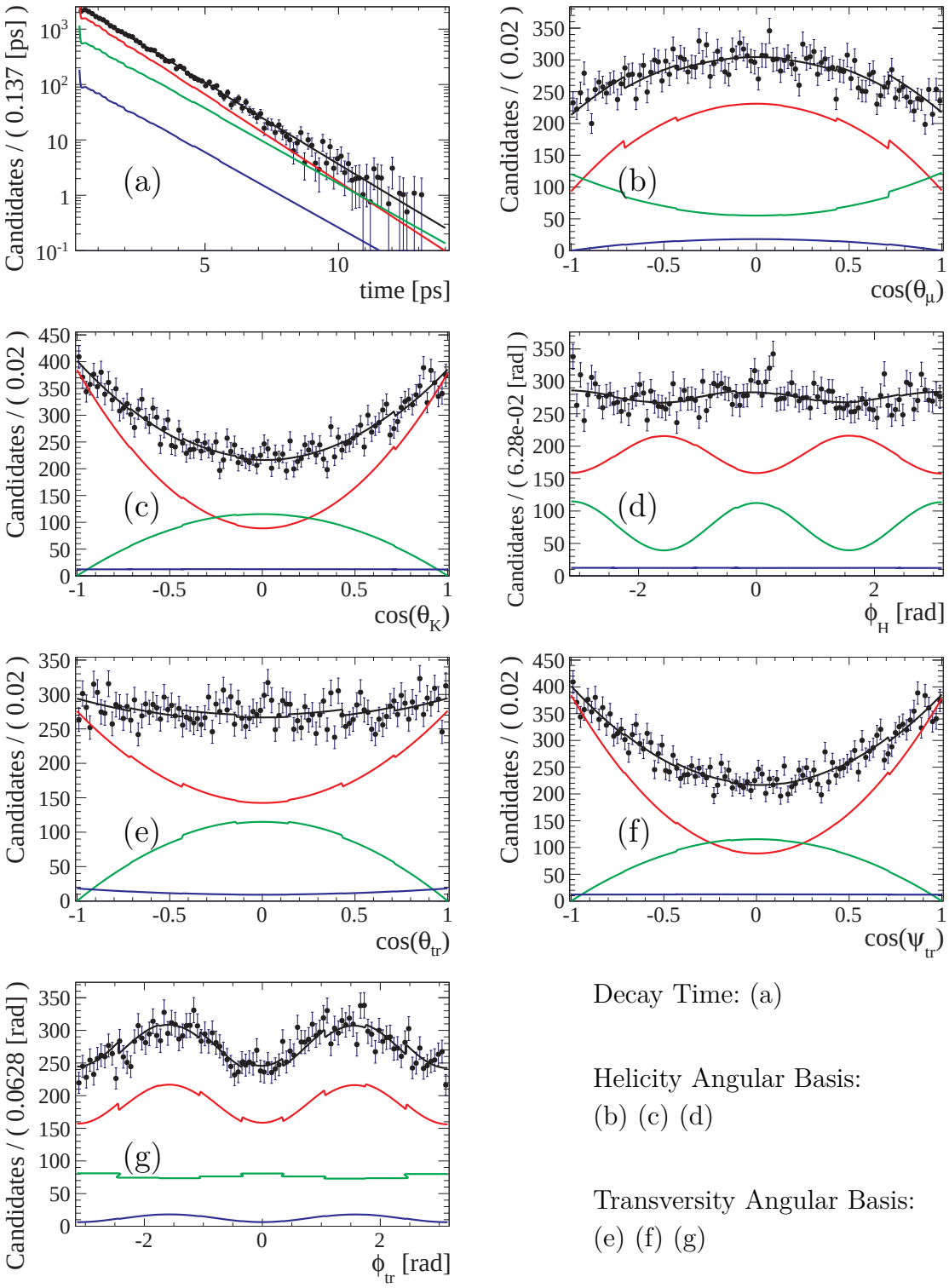


Figure 8.2: *Projections of the Fit Results for* $B_s^0 \to J/\psi\, K^+K^-$ *showing the various* $\mathcal{CP}$ *components of the fit. Black Line: Total PDF, Red Line:* $\mathcal{CP}$*-even P-wave component, Green Line:* $\mathcal{CP}$*-odd P-wave component, Blue Line: S-wave component. Discontinuities in the PDF functions are due to the angular acceptance histograms (Section 4.7.1).*

Parameter	Fit result and statistical uncertainty	
Γ_s	0.6631 ± 0.0048	$[\mathrm{ps}^{-1}]$
$\Delta\Gamma_s$	0.100 ± 0.016	$[\mathrm{ps}^{-1}]$
$\lvert A_\perp(0)\rvert^2$	0.2488 ± 0.0088	
$\lvert A_0(0)\rvert^2$	0.5211 ± 0.0061	
$\delta_\parallel$	$3.29\ ^{+0.13}_{-0.21}$	[rad]
$\delta_\perp$	$\mathbf{3.13} \pm 0.25$	[rad]
Δm_s	$\mathbf{17.71} \pm 0.10$	$[\mathrm{ps}^{-1}]$
ϕ_s	$\mathbf{0.080} \pm 0.094$	[rad]
$\lvert\lambda_s\rvert$	0.945 ± 0.034	

Table 8.3: *Parameters that are significantly changed when performing a fit to* $B_s^0 \to J/\psi\, K^+K^-$ *without an external constraint on* Δm_s.

can therefore be difficult. However, terms describing the interference between the decay amplitudes also contain factors of Δm_s which allows for this parameter to be measured even when $\phi_s \approx 0$.

Removing the external constraint on Δm_s and repeating the analysis yields an additional set of results. There is a small correlation between Δm_s and most parameters in the canonical fit. For brevity, only the key physics parameters of this fit are listed in Table 8.1. Only the values which differ by more than 10% of the statistical uncertainty compared to the nominal fit (and Δm_s) appear in bold in Table 8.3. The likelihood function for Δm_s is presented in Appendix D.2 and is found to have a single well defined minima.

8.2 Systematic studies

Individual systematic uncertainties in this analysis are discussed in Sections 8.2.1–8.2.11 and a summary of all uncertainties is given in Table 8.5.

8.2.1 Combinatorial background

As previously described in Sections 5.2 and 6.3 the combinatorial backgrounds are statistically removed from the fit to data through the use of sWeights.

These sWeights are calculated based on a fit to the $m(J/\psi\, K^+K^-)$ distribution. As such an additional uncertainty comes from propagating the statistical uncertainties from this fit. To propagate this uncertainty the physics parameters in the fit to $m(J/\psi\, K^+K^-)$ are adjusted by one sigma and the sWeights are re-calculated. Using these new sWeights the full analysis is repeated and the new physics results are compared to the nominal fit results in Table 8.1. The systematic uncertainty due to this effect is defined as the difference between the repeated and nominal fit results.

In order to use the $m(J/\psi\, K^+K^-)$ observable to calculate the sWeights it is assumed that it is independent of the decay time and angular observable distributions. To explore the effect of this assumption, the fit to the mass distribution is repeated in different bins of decay time and angles. New sets of sWeights are then calculated using these fit results and the full $B_s^0 \to J/\psi\, K^+K^-$ analysis is re-performed for each new set of weights. The variance of the fit results due to the new set of weights is defined as the systematic uncertainty and these are summarised in Table 8.4.

8.2.2 Peaking backgrounds

Two possible sources of peaking backgrounds are considered in Section 5.3. The contributions from these are found to be small compared to the signal dataset and so these are ignored in the construction of likelihood fit.

Approximately 780 ± 170 $B^0 \to J/\psi\, K^*(892)^0$ events are found in the mass range $m(J/\psi\, K^+K^-) \in [5200, 5550]$ MeV/c^2 as described in Section 5.3.1. This corresponds to approximately 1.5% of the signal in the analysis as described in Section 5.3.1. The uncertainty associated with ignoring this background is determined by re-performing the fit to data with this $B^0 \to J/\psi\, K^{*0}$ background statistically removed. This background is removed by introducing $B^0 \to J/\psi\, K^{*0}$ MC events with negative weights such that they statistically cancel the background in the

	Parameter	Decay time uncertainty	$\cos\theta_K$ uncertainty	$\cos\theta_\mu$ uncertainty	ϕ_{hel} uncertainty	Total uncertainty
	Γ_s $[ps^{-1}]$	0.0013	0.0006	0.0038	0.0003	0.0041
	$\Delta\Gamma_s$ $[ps^{-1}]$	0.001	0.001	0.001	0.000	0.002
	$\lvert A_\perp(0)\rvert^2$	0.00	0.00	0.00	0.00	0.00
	$\lvert A_0(0)\rvert^2$	0.0009	0.0005	0.0029	0.0002	0.0031
	$\delta_\parallel$ [rad]	0.01	0.01	0.03	0.00	0.03
	$\delta_\perp$ [rad]	0.01	0.01	0.02	0.00	0.02
	ϕ_s [rad]	0.001	0.001	0.002	0.000	0.003
	$\lvert\lambda_s\rvert$	0.001	0.001	0.002	0.001	0.003
bin 1	F_S	0.004	0.002	0.013	0.001	0.014
	$\delta_{S-\perp}$ [rad]	0.01	0.01	0.02	0.01	0.03
bin 2	F_S	0.001	0.001	0.003	0.001	0.003
	$\delta_{S-\perp}$ [rad]	0.01	0.01	0.01	0.01	0.02
bin 3	F_S	0.000	0.000	0.001	0.000	0.001
	$\delta_{S-\perp}$ [rad]	0.0	0.0	0.0	0.0	0.0
bin 4	F_S	0.000	0.001	0.001	0.000	0.002
	$\delta_{S-\perp}$ [rad]	0.01	0.01	0.02	0.01	0.03
bin 5	F_S	0.001	0.001	0.004	0.000	0.004
	$\delta_{S-\perp}$ [rad]	0.00	0.00	0.01	0.00	0.01
bin 6	F_S	0.001	0.002	0.006	0.001	0.006
	$\delta_{S-\perp}$ [rad]	0.00	0.00	0.01	0.00	0.01

Table 8.4: *Systematic uncertainties due to the possible correlation of $m(J/\psi\, K^+K^-)$ with the decay time and angular observables.*

dataset. The difference between these refitted results and the nominal results are taken to be the associated uncertainty.

Using Ref [58], Section 5.3.2 predicts 12 ± 3 events from the $B_c^\pm \to B_s^0 X^\pm$ decay contribute to this final state. Therefore no systematic uncertainty is assigned for the $B_c^\pm \to B_s^0 X^\pm$ background contribution.

8.2.3 Angular acceptance

The inclusion of angular acceptance in the fit is discussed in Section 4.7.1. This involves the use of angular acceptance weights in the PDF. The nominal angular acceptance weights are calculated using the fully simulated MC as described in Section 4.7.1. However, there is a known difference between the momenta distributions of MC and data as shown in Figure 4.1. To determine the uncertainty associated with this difference the angular acceptance weights are re-calculated using re-weighted MC such that the kaon and B_s^0 momenta distributions agree. (Re-weighting the MC such that the muon momenta agrees with data has no observable effect.) The difference introduced by fitting with these new angular acceptance weights is taken to be the systematic uncertainty.

Due to the fact that the full MC sample is only six times larger than the full $J/\psi\, K^+K^-$ dataset there is an associated statistical uncertainty on the angular acceptance weights. This uncertainty is propagated to the full fit by generating new angular acceptance weights with values fluctuated within their statistical uncertainties. This fluctuation is done in a correlated way using a weight error matrix. The fit is is then repeated using these new acceptance weights and the differences between these values and the nominal fit results are taken as the systematic uncertainty.

8.2.4 Decay time acceptance

The lower decay time acceptance is modelled using a histogram which describes the acceptance function. This is described in detail in Section 4.7.2. Each bin in this histogram has an associated statistical uncertainty as shown in Figure 4.5. To propagate this uncertainty in the fit to data additional acceptance histograms

are generated. The acceptance within each bin of the new histograms is floated within the statistical uncertainty of the values from the nominal histogram. The full analysis is repeated for each histogram and the systematic uncertainty is extracted from the variance of the fit results. This systematic uncertainty in the lower decay time acceptance only affects Γ_s and $\Delta\Gamma_s$.

The upper decay time acceptance is incorporated in the fit through the use of the β factor as described in Section 4.7.2. There is an uncertainty with this parameter which propagates to a systematic uncertainty only on Γ_s of $4.0 \times 10^{-3}\,\mathrm{ps}^{-1}$.

8.2.5 Nuisance $\mathcal{CP}$ asymmetries

This analysis is potentially susceptible to the effects of nuisance $\mathcal{CP}$-asymmetries. By construction, any possible asymmetry between the decay of B_s^0 and $\overline{B_s^0}$ is modelled in the likelihood fit through the parameter $|\lambda_s|$. The interpretation of $|\lambda_s|$ as the magnitude of the $\mathcal{CP}$ violation relies on the description of all other asymmetries within the fit to data. Other possible asymmetries include differences in the production or tagging of the two flavours of B_s^0 mesons.

Separate tagging calibrations are used for B_s^0 and $\overline{B_s^0}$ as described in Section 4.8.3. The asymmetry between the tagging efficiencies has been found to be small. In addition to this a different normalisation is used for each tagging decision in the fit. The mistag calibration is floated with a Gaussian constraint incorporating its systematic uncertainty. Because of this, any systematic uncertainty from this source is already included within the statistical uncertainties. Simulations show that the effect of a small asymmetry in production has a negligible effect on the fit and so no systematic uncertainty is assigned.

8.2.6 C_{S-P} factors

The C_{S-P} factors used in the analysis are described in Section 6.5. These are calculated under the assumption the line-shape of the S-wave across the $m(K^+K^-)$ range is flat. As a cross-check these parameters are re-calculated using a Flatté function [79] to describe the S-wave strong phase. Repeating the fit to data with these new factors is found to only effect the measured values of $\delta_{S-\perp}$ by approximately 10% of the statistical uncertainty in each of the six $m(K^+K^-)$ bins, which is taken as the systematic uncertainty.

8.2.7 Fit bias

To explore any possible biases in the fit MC datasets are generated using the same PDF[1] function which is fit to data. These datasets can then be fit to and the distribution of extracted fit results and errors in each parameter can be explored. Each dataset is constructed to contain 24,500 events and the sensitivities are scaled to the sensitivities expected for 28,000 events to make the comparisons to data easier. The resulting distributions from performing 1,000 fits to MC data are presented in Appendix F. The majority of these parameters are well behaved with Gaussian distributions centred at their generated value and with uncertainties matching the results from data. However, both $|A_\perp(0)|^2$ and $\delta_\parallel$ are found to have an approximate 10% bias, compared to their generated value. A systematic uncertainty of 0.0010 is also conservatively assigned to $|A_\perp(0)|^2$, based on the result of this study. The $\delta_\parallel$ parameter has a central value close to a secondary minima, due to this the uncertainty is asymmetric and is described using a range. As $\delta_\parallel$ is known to have a non parabolic likelihood no systematic bias is assigned to this parameter. The S-wave fractions (F_S) and phases ($\delta_{S-\perp}$) are also observed to have significant fit biases and so these are taken to be a conservative systematic uncertainty on these parameters. All of these biases have been observed to vanish when fitting with much larger MC datasets.

8.2.8 Momentum and length scale

The measurements of the momentum scale and decay length of candidates at LHCb both have a small associated uncertainty of 0.020%. The uncertainty in the decay length propagates to a systematic uncertainty on Γ_s and $\Delta\Gamma_s$ of $0.0009\,\mathrm{ps}^{-1}$ and $0.0002\,\mathrm{ps}^{-1}$ respectively. This is a relatively small uncertainty on Γ_s and is negligible for $\Delta\Gamma_s$. The uncertainty in the momentum scale effects both the reconstructed B_s^0 momentum and mass values and the effect largely cancels when calculating the decay time. Due to this there is no systematic effect associated with the momentum uncertainty in the fit.

[1]The PDF is evaluated using the same physics parameters obtained from fitting to data.

8.2.9 Decay time resolution offset

In Section 4.6 the decay time resolution of prompt events is described by a Gaussian function, this Gaussian has an associated offset of $\approx -4\,\text{fs}$. When performing a per-event fit to data the fit is performed using an offset of 0 fs. The effect of this difference is calculated by re-performing the full analysis with all events having a decay time offset of $+4\,\text{fs}$ applied. The difference between this fit result and the nominal fit is taken as the systematic uncertainty. A series of MC fits were performed to verify this uncertainty. These MC datasets were generated using nominal fit conditions and the same physics parameters found in data. These datasets were then fit to before and after an offset of $-4\,\text{fs}$ was applied. The bias introduced in fitting with and without these offsets is compatible with the observed bias introduced when fitting to data with a $+4\,\text{fs}$ correction applied.

8.2.10 Total systematic uncertainties

The final set of all of the systematic uncertainties are listed in Tables 8.5 (a), (b) and (c). Table 8.5 (a) contains the breakdown and total systematic uncertainties on the key physics parameters. Table 8.5 (b) contains the systematic uncertainties on the measurement of F_S in each bin of $m(K^+K^-)$. Table 8.5 (c) contains the systematic uncertainties on $\delta_{S-\perp}$ in each of the $m(K^+K^-)$ bins.

8.2.11 Free Δm_s systematics

The dominant source of systematic uncertainty for Δm_s comes from the uncertainty in length and momentum scales. Adding these contributions in quadrature produces a systematic uncertainty of $\pm\,0.005\,\text{ps}^{-1}$. Repeating the full fit to data ignoring the effects of decay time and angular acceptances as well as the finite time resolution yields a maximum systematic uncertainty of $\pm\,0.01\,\text{ps}^{-1}$. Half of this value is taken as a conservative systematic uncertainty. Adding both of these sources in quadrature yields a systematic uncertainty of $\pm\,0.007\,\text{ps}^{-1}$.

Source	Γ_s $[\mathrm{ps}^{-1}]$	$\Delta\Gamma_s$ $[\mathrm{ps}^{-1}]$	$\lvert A_\perp(0)\rvert^2$	$\lvert A_0(0)\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	ϕ_s [rad]	$\lvert\lambda_s\rvert$
Stat. uncertainty	0.0048	0.016	0.0086	0.0061	$^{+0.13}_{-0.21}$	0.22	0.091	0.031
Background subtraction	0.0041	0.002	–	0.0031	0.03	0.02	0.003	0.003
$B^0 \to J/\psi\, K^{*0}$ background	–	0.001	0.0030	0.0001	0.01	0.02	0.004	0.005
Ang. acc. re-weighting	0.0007	–	0.0052	0.0091	0.07	0.05	0.003	0.020
Ang. acc. statistical	0.0002	–	0.0020	0.0010	0.03	0.04	0.007	0.006
Lower decay time acc. model	0.0023	0.002	–	–	–	–	–	–
Upper decay time acc. model	0.0040	–	–	–	–	–	–	–
Length and mom. scales	0.0002	–	–	–	–	–	–	–
Fit bias	–	–	0.0010	–	–	–	–	–
Decay time resolution offset	–	–	–	–	–	0.04	0.006	–
Quadratic sum of syst.	0.0063	0.003	0.0064	0.0097	0.08	0.08	0.011	0.022
Total uncertainties	0.0079	0.016	0.0107	0.0114	$^{+0.15}_{-0.23}$	0.23	0.092	0.038

(a) *Statistical and systematic uncertainties for key physics parameters.*

Source	bin 1 F_S	bin 2 F_S	bin 3 F_S	bin 4 F_S	bin 5 F_S	bin 6 F_S
Stat. uncertainty	$^{+0.081}_{-0.073}$	$^{+0.030}_{-0.027}$	$^{+0.014}_{-0.007}$	$^{+0.012}_{-0.009}$	$^{+0.027}_{-0.025}$	$^{+0.043}_{-0.042}$
Background subtraction	0.014	0.003	0.001	0.002	0.004	0.006
$B^0 \to J/\psi\, K^{*0}$ background	0.010	0.006	0.001	0.001	0.002	0.018
Angular acc. re-weighting	0.004	0.006	0.004	0.005	0.006	0.007
Angular acc. statistical	0.003	0.003	0.002	0.001	0.003	0.004
Fit bias	0.009	–	0.002	0.002	0.001	0.001
Quadratic sum of syst.	0.020	0.009	0.005	0.006	0.008	0.021
Total uncertainties	$^{+0.083}_{-0.076}$	$^{+0.031}_{-0.029}$	$^{+0.015}_{-0.009}$	$^{+0.013}_{-0.011}$	$^{+0.028}_{-0.026}$	$^{+0.048}_{-0.047}$

(b) *Statistical and systematic uncertainties for F_S for each $m(K^+K^-)$ bin.*

Source	bin 1 $\delta_{S-\perp}$ [rad]	bin 2 $\delta_{S-\perp}$ [rad]	bin 3 $\delta_{S-\perp}$ [rad]	bin 4 $\delta_{S-\perp}$ [rad]	bin 5 $\delta_{S-\perp}$ [rad]	bin 6 $\delta_{S-\perp}$ [rad]
Stat. uncertainty	$^{+0.78}_{-0.49}$	$^{+0.38}_{-0.23}$	$^{+1.40}_{-0.30}$	$^{+0.21}_{-0.35}$	$^{+0.18}_{-0.26}$	$^{+0.18}_{-0.22}$
Background subtraction	0.03	0.02	–	0.03	0.01	0.01
$B^0 \to J/\psi\, K^{*0}$ background	0.08	0.04	0.08	0.01	0.01	0.05
Angular acc. re-weighting	0.02	0.03	0.12	0.13	0.03	0.01
Angular acc. statistical	0.033	0.023	0.067	0.036	0.019	0.015
Fit bias	0.005	0.043	0.112	0.049	0.022	0.016
C_{S-P} factors	0.007	0.028	0.049	0.025	0.021	0.020
Quadratic sum of syst.	0.09	0.08	0.20	0.15	0.05	0.06
Total uncertainties	$^{+0.79}_{-0.50}$	$^{+0.39}_{-0.24}$	$^{+1.41}_{-0.36}$	$^{+0.26}_{-0.38}$	$^{+0.19}_{-0.26}$	$^{+0.19}_{-0.23}$

(c) *Statistical and systematic uncertainties for $\delta_{S-\perp}$ in each $m(K^+K^-)$ bin.*

Table 8.5: *Table of systematic uncertainties for: (a) $B_s^0 \to J/\psi\, K^+K^-$ P-wave physics parameters (b) S-wave fractions F_S (c) Strong phase differences $\delta_{S-\perp}$.*

Parameter	Value $\pm$ stat $\pm$ syst
Γ_s $[\mathrm{ps}^{-1}]$	$0.663 \pm 0.005 \pm 0.006$
$\Delta\Gamma_s$ $[\mathrm{ps}^{-1}]$	$0.100 \pm 0.016 \pm 0.003$
$\lvert A_\perp(0)\rvert^2$	$0.249 \pm 0.009 \pm 0.006$
$\lvert A_0(0)\rvert^2$	$0.521 \pm 0.006 \pm 0.010$
$\delta_\parallel$ [rad]	$3.30\,^{+0.13}_{-0.21} \pm 0.08$
$\delta_\perp$ [rad]	$3.07 \pm 0.22 \pm 0.07$
ϕ_s [rad]	$0.07 \pm 0.09 \pm 0.01$
$\lvert\lambda_s\rvert$	$0.94 \pm 0.03 \pm 0.02$

Table 8.6: *Final set of results for* $B_s^0 \to J/\psi\ K^+K^-$. *The central values and statistical uncertainties in this table are the same as in Table 8.1.*

8.3 Final $B_s^0 \to J/\psi\ K^+K^-$ results

The final measurement of the key physics parameters in the fit to $B_s^0 \to J/\psi\ K^+K^-$, including the statistical and systematic uncertainties are presented in Table 8.6.

Previous analyses have shown that $\Delta\Gamma_s$ and ϕ_s are highly correlated parameters for analyses using small datasets. This is not the case in this analysis as shown in the correlation matrix in Table 8.2. Figure 8.3 shows the likelihood profile corresponding to the solution $\Delta\Gamma_s > 0$. This figure shows the contours corresponding to confidence limits of 68%, 90% and 95% with the standard model expected value shown as a black point with 1σ uncertainties. The contours in here only include the statistical uncertainties on each parameter. It can be seen from Figure 8.3 that the minima from this analysis is consistent with the standard model expectation of ${\phi_s}^{SM} = -0.036 \pm 0.002\,\mathrm{rad}$ [19] ${\Delta\Gamma_s}^{SM} = 0.082 \pm 0.021\,\mathrm{ps}^{-1}$ [21–23].

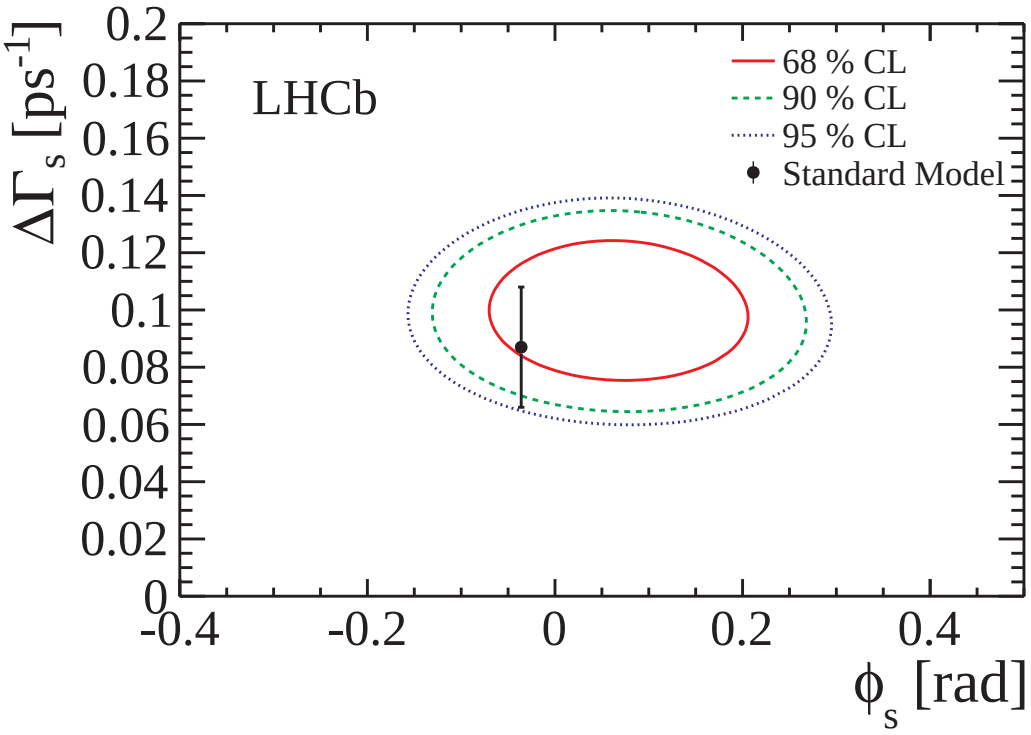


Figure 8.3: *Two-dimensional profile likelihood in* $(\Delta\Gamma_s, \phi_s)$ *plane [12]. The Standard Model expectation is shown as a black point with uncertainties and corresponds to* ${\phi_s}^{SM} = -0.036 \pm 0.002\,\text{rad}$ *[19],* ${\Delta\Gamma_s}^{SM} = 0.082 \pm 0.021\,\text{ps}^{-1}$ *[21–23].*

8.4 $\Delta\Gamma_s$ sign resolution

Section 6.1 describes how the $B_s^0 \to J/\psi\, K^+K^-$ PDF is invariant under the transform of equation 6.3 giving rise to two degenerate results which have different signs of $\Delta\Gamma_s$. As previously discussed in Section 6.5 the nominal fit to data is performed in six bins of $m(K^+K^-)$. Performing this analysis in six bins allows the variation of the strong phase difference $\delta_{S-\perp}$[1] to be determined. This was first implemented in Ref [67].

The numerical fit results for the second minima in the $B_s^0 \to J/\psi\, K^+K^-$ analysis correspond to the results in Table 8.1 with the application of the transform in equation 6.3.

The results of the strong phase difference in the six $m(K^+K^-)$ bins are shown in Figure 8.4. The total uncertainties in this plot are defined as the quadrature sum of both the statistical and systematic uncertainties in Tables 8.1 and 8.5.

The expected behaviour of the S-wave and P-wave strong phases are shown in Figure 6.1. From this the strong phase difference of the physical solution is expected to decrease across the $m(K^+K^-)$ range. In Figure 8.4 the blue points correspond to the solution where $\Delta\Gamma_s > 0$ and the red points where $\Delta\Gamma_s < 0$. With the physical solution corresponds to the set of points with a decreasing trend across $m(K^+K^-)$, the ambiguity is resolved such that $\Delta\Gamma_s$ is positive. This is in agreement with the observed result from the previous analysis in Ref [67].

As described in Section 6.6, the difference between the strong phases shown in Figure 8.4. Using the values of $\delta_\perp$ from the fit to data (Table 8.1) and averaging over the six bins the S-wave strong phase is calculated to be $\delta_S = 1.7^{+0.3}_{-0.2}\,[\,\text{rad},]$ under the approximation that the phase is constant across $m(K^+K^-)$. The difference between the phases in bins 1 and 6 in Figure 8.4 is approximately $\pi\,[\,\text{rad}]$ when the variation of θ_{SP} in included. This corresponds to the phase change of the P-wave across the ϕ resonance.

[1] It is also possible to study the $\delta_S - \delta_\parallel$ phase difference, however, the phase difference $\delta_S - \delta_\perp$ has a smaller uncertainty.

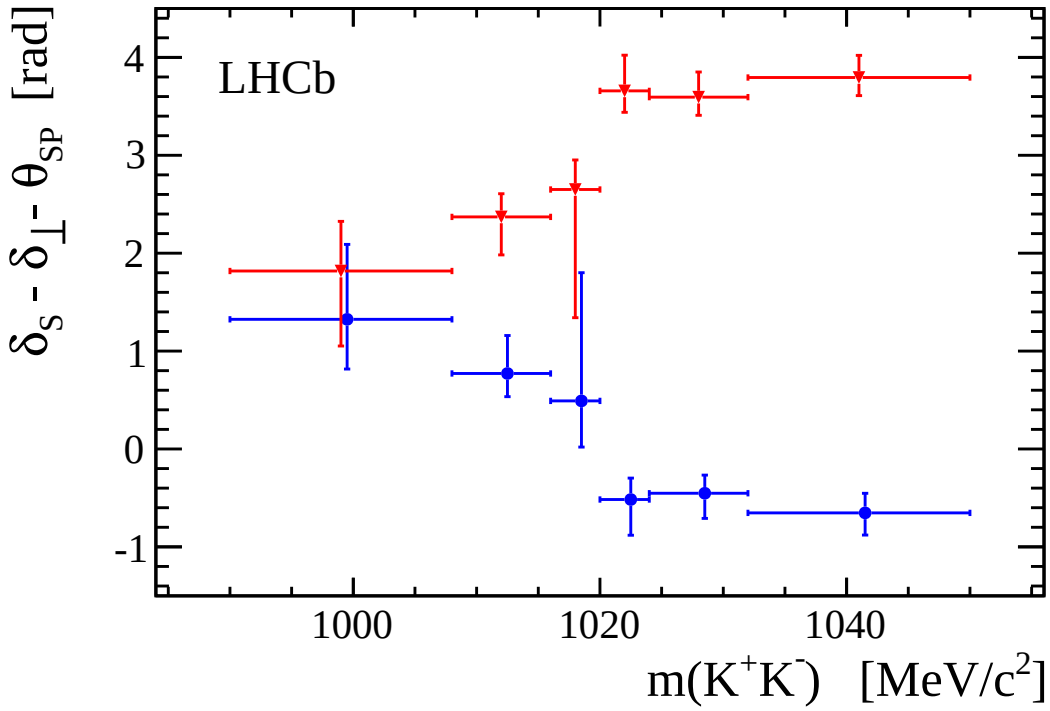


Figure 8.4: *Variation of strong phase difference across $m(K^+K^-)$ for the two possible solutions from the analysis of $B_s^0 \to J/\psi\, K^+K^-$ [12]. Blue Points: Solution corresponding to $\Delta\Gamma_s > 0$. Red Points: Solution corresponding to $\Delta\Gamma_s < 0$. Total uncertainties include the quadrature sum of the statistical and systematic uncertainties $\delta_{S-\perp}$ results.*

8.5 Inclusion of $B_s^0 \to J/\psi\, \pi^+\pi^-$

The decay of B_s^0 mesons to the final state of $J/\psi\, h^+h^-$ is described in Section 2.5 with the leading order Feynman diagrams given in Figure 2.5. In addition to the final state of $J/\psi\, K^+K^-$, which has been discussed earlier, the final state $J/\psi\, \pi^+\pi^-$ is also possible. A measurement of ϕ_s in the channel of of $B_s^0 \to J/\psi\, \pi^+\pi^-$ using $1.0\,fb^{-1}$ of data at LHCb has previously been published in Ref [78]. The results from this analysis have been superseded with more updated results published in Ref [12]. Using a combination of the $B_s^0 \to J/\psi\, K^+K^-$ and $B_s^0 \to J/\psi\, \pi^+\pi^-$ analyses it is possible to make a more precise measurement of ϕ_s, Γ_s and $\Delta\Gamma_s$, as will be described below.

Although the $B_s^0 \to J/\psi\, \pi^+\pi^-$ analysis was originally performed by other groups, the results have been reproduced as part of this thesis as required to perform the combined measurement between the decay channels.

8.5.1 Fit to only $B_s^0 \to J/\psi\, \pi^+\pi^-$

One of the important results from the analysis of Ref [78] is that, in the range $m(\pi^+\pi^-) \in [775, 1550]$ MeV/c^2, 97.7% of the signal is in a $\mathcal{CP}$-odd final state at 95% C.L. From this result the $B_s^0 \to J/\psi\, \pi^+\pi^-$ dataset is taken to be composed of an entirely $\mathcal{CP}$-odd component. Because of this, there is no need to perform an angular analysis to separate the final states and the $B_s^0 \to J/\psi\, \pi^+\pi^-$ cross-section is described only by term $h_7(t)$ in Table 2.3.

The analysis of $B_s^0 \to J/\psi\, \pi^+\pi^-$ is unable to simultaneously determine the values of Γ_s and $\Delta\Gamma_s$ by itself. Therefore, the measurements and correlation of Γ_s and $\Delta\Gamma_s$ are used from from Tables 8.1 and 8.2. The value of ϕ_s from this updated $B_s^0 \to J/\psi\, \pi^+\pi^-$ analysis alone is,

$$\phi_s = -0.14^{+0.17}_{-0.16} \pm 0.01\,\text{rad}\ . \tag{8.1}$$

The systematic uncertainty on this measurement of ϕ_s is calculated in the same way as described in the published $B_s^0 \to J/\psi\, \pi^+\pi^-$ analysis [78].

Parameter	Value $\pm$ stat $\pm$ syst
Γ_s $[\,ps^{-1}\,]$	$0.661 \pm 0.004 \pm 0.006$
$\Delta\Gamma_s$ $[\,ps^{-1}\,]$	$0.106 \pm 0.011 \pm 0.007$
$\lvert A_\perp(0)\rvert^2$	$0.246 \pm 0.007 \pm 0.006$
$\lvert A_0\,(0)\rvert^2$	$0.523 \pm 0.005 \pm 0.010$
$\delta_\parallel$ [rad]	$3.32\,^{+0.13}_{-0.21} \pm 0.08$
$\delta_\perp$ [rad]	$3.04 \pm 0.20 \pm 0.07$
ϕ_s [rad]	$0.01 \pm 0.07 \pm 0.01$
$\lvert\lambda_s\rvert$	$0.93 \pm 0.03 \pm 0.02$

Table 8.7: *Key physics parameters measured for the combined analysis of* $B_s^0 \to J/\psi\ K^+K^-$ *and* $B_s^0 \to J/\psi\ \pi^+\pi^-$ *decay channels.*

8.5.2 Combined results for $B_s^0 \to J/\psi\ h^+h^-$

As these two independent analyses provide independent constraints on ϕ_s it is possible to use the two channels to provide a combined constraint on these parameters. This is performed by adding the NLL fit functions from the $B_s^0 \to J/\psi\ K^+K^-$ and $B_s^0 \to J/\psi\ \pi^+\pi^-$ analyses and performing a simultaneous fit of the combined likelihood to the two datasets.

The final combined results from performing a simultaneous fit to to both $B_s^0 \to J/\psi\ K^+K^-$ and $B_s^0 \to J/\psi\ \pi^+\pi^-$ are shown in Table 8.7, with the correlation matrix given in Table 8.8. Due to the background model in $B_s^0 \to J/\psi\ \pi^+\pi^-$ and variations in the decay time acceptance, the combined analysis includes additional systematic uncertainties of $0.001\,ps^{-1}$ on Γ_s and $0.006\,ps^{-1}$ on $\Delta\Gamma_s$. The systematic uncertainties on all other parameters are unaffected.

In Table 8.8 it can be seen that the correlation between Γ_s and $\Delta\Gamma_s$ is reduced compared to just the $B_s^0 \to J/\psi\ K^+K^-$ analysis. This is largely due to the addition of the $\mathcal{CP}$-odd component from $B_s^0 \to J/\psi\ \pi^+\pi^-$ which allows for the different $\mathcal{CP}$ eigenstates to be more clearly resolved.

	Γ_s [ps^{-1}]	$\Delta\Gamma_s$ [ps^{-1}]	$\lvert A_0(0)\rvert^2$	$\lvert A_\perp(0)\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	ϕ_s [rad]	$\lvert\lambda_s\rvert$
Γ_s [ps^{-1}]	1.00	0.10	0.08	0.03	−0.08	−0.04	0.01	0.00
$\Delta\Gamma_s$ [ps^{-1}]		1.00	−0.49	0.47	0.00	0.00	0.00	−0.01
$\lvert A_\perp(0)\rvert^2$			1.00	−0.40	−0.37	−0.14	0.02	−0.05
$\lvert A_0(0)\rvert^2$				1.00	−0.05	−0.03	−0.01	0.01
$\delta_\parallel$ [rad]					1.00	0.39	−0.01	0.13
$\delta_\perp$ [rad]						1.00	0.21	0.03
ϕ_s [rad]							1.00	0.06
$\lvert\lambda_s\rvert$								1.00

Table 8.8: *Correlation matrix for the combined analysis of* $B_s^0 \to J/\psi\ K^+K^-$ *and* $B_s^0 \to J/\psi\ \pi^+\pi^-$ *decay channels.*

"An expert is a person who has made all the mistakes that can be made in a very narrow field."

Niels Bohr

9

Complimentary Decay Time Studies

In Chapter 8 the results of the main analysis of this thesis, the fit to $B_s^0 \to J/\psi\, K^+K^-$, have been presented. The results in Chapter 8 include, among other physics parameters, the average B_s^0 decay width (Γ_s) and width splitting between the mass eigenstates ($\Delta\Gamma_s$) using the full $B_s^0 \to J/\psi\, K^+K^-$ decay rate (Section 2.5).

In this final chapter two complementary and more "visual" ways to view the lifetime information in the data are used. One such piece of information is the effective decay time $\tau_{\rm eff}$. This is defined as the decay time obtained by fitting a single exponential function to the whole dataset. Because the final state is composed of different short and long lived components the effective decay time is related to the average decay width, Γ_s, and splitting, $\Delta\Gamma_s$. A fit to measure this quantity is described in Section 9.1, and forms a simple qualitative cross-check of the full $B_s^0 \to J/\psi\, K^+K^-$ analysis.

The values of Γ_s and $\Delta\Gamma_s$ are related to the decay widths of the individual mass eigenstates, Γ_H and Γ_L, as described in Appendix A.2. Using an angular moment analysis (Ref [80]) is it also possible to measure these decay widths separately. Another complementary analysis using this technique is described in Section 9.3. The values of Γ_s and $\Delta\Gamma_s$ can then be calculated, which provides an additional cross check of the full $B_s^0 \to J/\psi\, K^+K^-$ analysis.

9.1 Effective decay time study

Using the total $1.0\,\mathrm{fb}^{-1}$ dataset from the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ selection it is possible to fit a single exponential function to the decay time distribution. The result from this fit is to measure the effective decay time $\tau_{\mathrm{eff}} = 1/\Gamma_{\mathrm{eff}}$.

The expected value for τ_{eff} of the dataset relies on knowledge of the average decay width, Γ_{s} , the decay width splitting, $\Delta\Gamma_{\mathrm{s}}$, and the amplitudes of the different $\mathcal{CP}$ eigenstates. This relation is described in Ref [81] and given below,

$$\frac{1}{\Gamma_{\mathrm{eff}}} = \tau_{\mathrm{eff}} = \tau_{\mathrm{B_s}} \frac{1}{1-\mathrm{y_s^2}} \left(\frac{1 + 2\mathrm{A}^{\mathrm{f}}_{\Delta\Gamma_{\mathrm{s}}} \mathrm{y_s} + \mathrm{y_s^2}}{1 + \mathrm{A}^{\mathrm{f}}_{\Delta\Gamma_{\mathrm{s}}} \mathrm{y_s}} \right) , \tag{9.1}$$

$$\text{Where: } \mathrm{y_s} = \frac{\Delta\Gamma_{\mathrm{s}}}{2\Gamma_{\mathrm{s}}} \text{ , and, } \mathrm{A}^{\mathrm{f}}_{\Delta\Gamma_{\mathrm{s}}} = 2\left|\mathrm{A}_{\mathcal{CP}-odd}\left(0\right)\right|^2 - 1 \text{ .}$$

Here, $\left|\mathrm{A}_{\mathcal{CP}-odd}\left(t=0\right)\right|^2$, is the amplitude of the $\mathcal{CP}$-odd component within the dataset. This is defined as,

$$\left|\mathrm{A}_{\mathcal{CP}-odd}\left(0\right)\right|^2 = \frac{\left|\mathrm{A}_{\perp}(0)\right|^2 + \left|\mathrm{A_S}\left(0\right)\right|^2}{1 + \left|\mathrm{A_S}\left(0\right)\right|^2} \text{ .} \tag{9.2}$$

The effective lifetime is extracted using an $_s$Fit strategy (Section 6.3) using the same $_s$Weights as in the $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ analysis. The PDF used to extract the effective decay time from the dataset is defined as,

$$\mathrm{PDF}\left(t|\sigma_t;\tau_{\mathrm{eff}}\,,s_t\right) = \frac{\left[e^{-\left(t/\tau_{\mathrm{eff}}\right)} \otimes R\left(t|\sigma_t;s_t\right)\right] \times \varepsilon\left(t\right)}{\int dt\ \left[e^{-\left(t/\tau_{\mathrm{eff}}\right)} \otimes R\left(t|\sigma_t;s_t\right)\right] \times \varepsilon\left(t\right)} \text{ .} \tag{9.3}$$

The decay time efficiency function $\varepsilon\left(t\right)$ incorporates the effect of lower decay time acceptance into the fit as described in Section 4.7.2. This study ignores the effect of the upper decay time acceptance.

The numerical results from the fit to the effective decay time are summarised in Table 9.1. The final projection of the fit of the single exponential to data is shown in Figure 9.1. The measured and expected values for τ_{eff} are in good agreement.

	τ_{eff} [ps]	Γ_{eff} [ps^{-1}]
Measured	1.457 ± 0.009	0.686 ± 0.004
Expected*	1.456 ± 0.006	0.685 ± 0.003

Table 9.1: *A comparison between expected and measured effective decay time τ_{eff} ($\Gamma_{eff} = 1/\tau_{eff}$). The expected effective decay time is calculated using equation 9.1 with: $\Gamma_s = 0.6715 \pm 0.0048\,\text{ps}^{-1}$(Ignoring upper decay time acceptance.), $\Delta\Gamma_s = 0.100 \pm 0.016\,\text{ps}^{-1}$, $|A_{\mathcal{CP}-odd}(0)|^2 = 0.285 \pm 0.001$. * These are conservative errors which don't account for parameter correlations.*

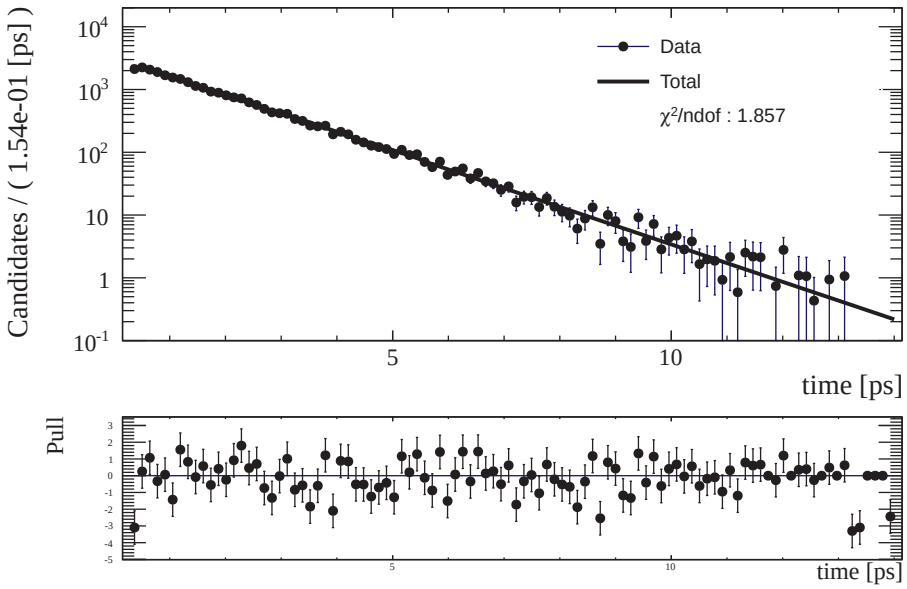


Figure 9.1: *Projection of the single exponential effective decay time fit result from the sWeighted fit to the $B_s^0 \to J/\psi\, K^+K^-$ data.*

This is a useful qualitative cross check that a simple fit is consistent with the information found in the more complex fits of the thesis. The $\chi^2/\text{nDoF} = 1.9$ with a probability of $< 1\%$ this is believed to be in part due to approximating the distribution containing 2 lifetimes with a single effective lifetime.

9.2 Angular moment weights

The individual time distributions that correspond to the high and low mas eigenstates can be constructed using orthonormal functions of the angular distribution of the data to extract the different decay widths $\Gamma_{\rm H}$ and $\Gamma_{\rm L}$.

For convenience, and to be consistent with the method described in Ref [80], this analysis ignores the contribution of the small S-wave component[1] in the fit and all calculations are performed in the transversity angular basis (Section 2.5.3).

The simplest form of the full $\rm B_s^0 \to J/\psi\ K^+K^-$ decay rate which allows for angular moments to be used is the one-angular differential decay rate [80, 82]. This decay rate only describes the distribution of events in θ_{tr} and decay time,

$$\frac{d\Gamma'(t,\theta_{tr})}{dtd\theta_{tr}} = \sum_k^2 h_k{}'(t) f_k{}'(\theta_{tr}) \ , \tag{9.4}$$
$$\text{where: } h_k{}'(t) = N_k{}' e^{-\Gamma_k t} \ .$$

Here the decay time distributions are described by $h_k{}'(t)$ and the angular distributions are described by $f_k{}'(\theta_{tr})$, both of these terms fully expanded in Table 9.2. This form of the decay rate is constructed under the assumption of no $\mathcal{CP}$ violation being present in the dataset, i.e. $\phi_{\rm s} = 0$ and $|\lambda_s| = 1$. This table shows that the decay is therefore approximated by two different mass eigenstates with different angular distributions and different decay widths.

Using angular moment functions, $\omega_j(\theta_{tr})$, defined in Ref [80] and presented in Table 9.3, it is possible to integrate over the angular distribution and extract each of the individual terms in Table 9.2. The weights $\omega_j(\theta_{tr})$ in Table 9.3 are orthonormal to the angular distributions of the $\Gamma_{\rm H}$ and $\Gamma_{\rm L}$ components shown in

[1] The S-wave across the whole dataset corresponds to $F_{\rm S} \approx 4\%$. This is calculated using a weighted average over all six bins in $m(\rm K^+K^-)$.

k	$f_k{}'(\theta_{tr})$	$N_k{}'$	$e^{-\Gamma_k t}$
1	$\frac{3}{8}(1+\cos^2\theta_{tr})$	$\lvert \mathrm{A}_0(0)\rvert^2 + \lvert \mathrm{A}_\parallel(0)\rvert^2$	$e^{-\Gamma_\mathrm{H} t}$
2	$\frac{3}{4}\sin^2\theta_{tr}$	$\lvert \mathrm{A}_\perp(0)\rvert^2$	$e^{-\Gamma_\mathrm{L} t}$

Table 9.2: *1-angular form of the* $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ *decay rate [80], [82], where* $f_k{}'(\theta_{tr})$ *and* $h_k{}'(t) = N_k{}' e^{-\Gamma_k t}$ *are terms in equation 9.4.*

k	j	$f_k{}'(\theta_{tr})$	$\omega_j(\theta_{tr})$	I
1	1	$\frac{3}{8}(1+\cos^2\theta_{tr})$	$(5\cos^2\theta_{tr}-1)$	1
1	2	$\frac{3}{8}(1+\cos^2\theta_{tr})$	$(2-5\cos^2\theta_{tr})$	0
2	1	$\frac{3}{4}\sin^2\theta_{tr}$	$(5\cos^2\theta_{tr}-1)$	0
2	2	$\frac{3}{4}\sin^2\theta_{tr}$	$(2-5\cos^2\theta_{tr})$	1

Table 9.3: *The appropriate projection operators* $\omega_j(\theta_{tr})$ *which satisfy the condition:* $I = \int_{-1}^{+1} d\cos\theta_{tr}\ \omega_j(\theta_{tr})\ f_k{}'(\theta_{tr}) = \delta_{jk}$ *[80].*

Table 9.2. The integral I, over the product of the angular moment functions and the angular distributions extracts only one of the decay time components for each weight function,

$$I = \int_{-1}^{+1} d\cos\theta_{tr}\ \omega_j(\theta_{tr})\, f_k{}'(\theta_{tr})\, h_k{}'(t) = \delta_{jk}\, h_k{}'(t)\ . \tag{9.5}$$

Figures 9.2 and 9.3 show the decay time projections where the datapoints correspond to the datasets weighted using $\omega_1(\theta_{tr})$ and $\omega_2(\theta_{tr})$ respectively. The long lived signal components are clearly visible in addition to the short lived background events.

Figures 9.4 and 9.5 show the mass projections where again the datapoints correspond to the weighted datasets using the angular moment functions $\omega_1(\theta_{tr})$ and $\omega_2(\theta_{tr})$ respectively.

When fitting to data, the integral I is approximated by summing over the per-event angular moment weights for all events in the dataset.

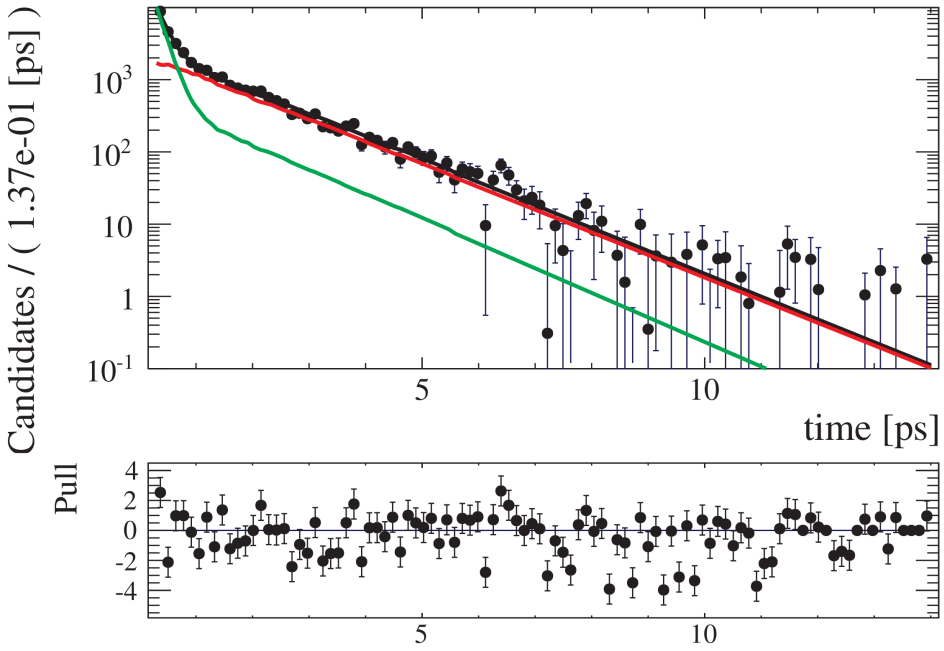


Figure 9.2: *Decay time projection of the fit to the dataset weighted using* $\omega_1\,(\theta_{tr})$ *to determine* Γ_{H}*. The Red line corresponds to the signal component, the Green line the background and the black line the PDF total.*

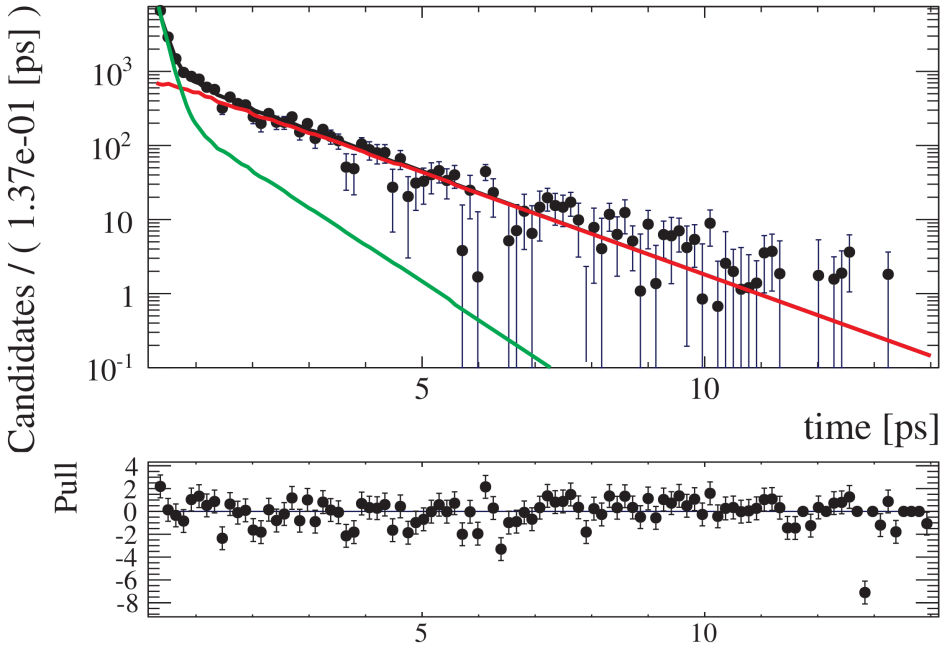


Figure 9.3: *Decay time projection of the fit to the dataset weighted using* $\omega_2\,(\theta_{tr})$ *to determine* Γ_{L}*. The Red line corresponds to the signal component, the Green line the background and the black line the PDF total.*

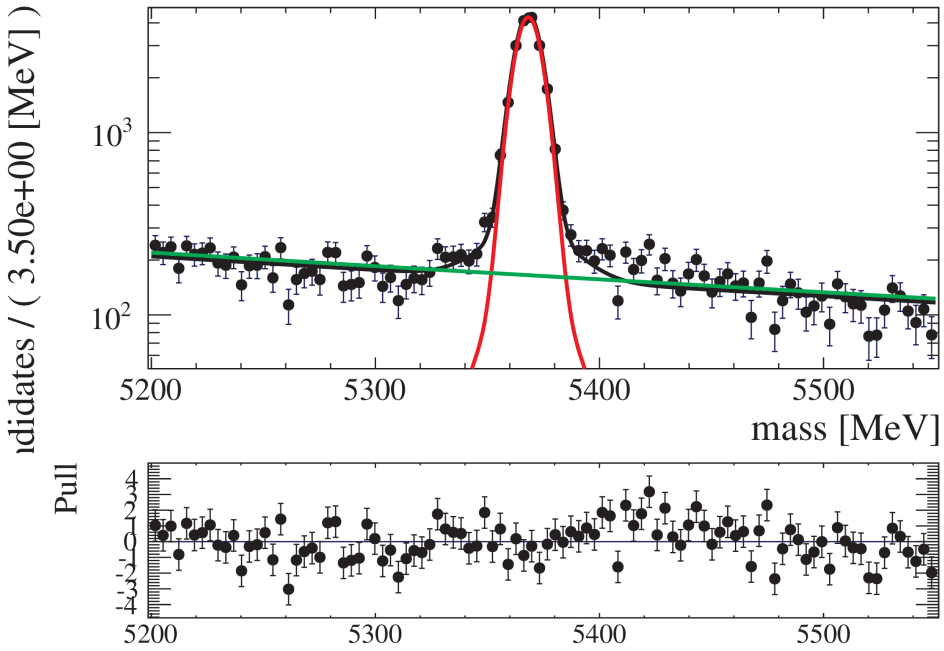


Figure 9.4: *Mass projection of the fit to the dataset weighted using $\omega_1(\theta_{tr})$ to determine Γ_{H}. The Red line corresponds to the signal component, the Green line the background and the black line the PDF total.*

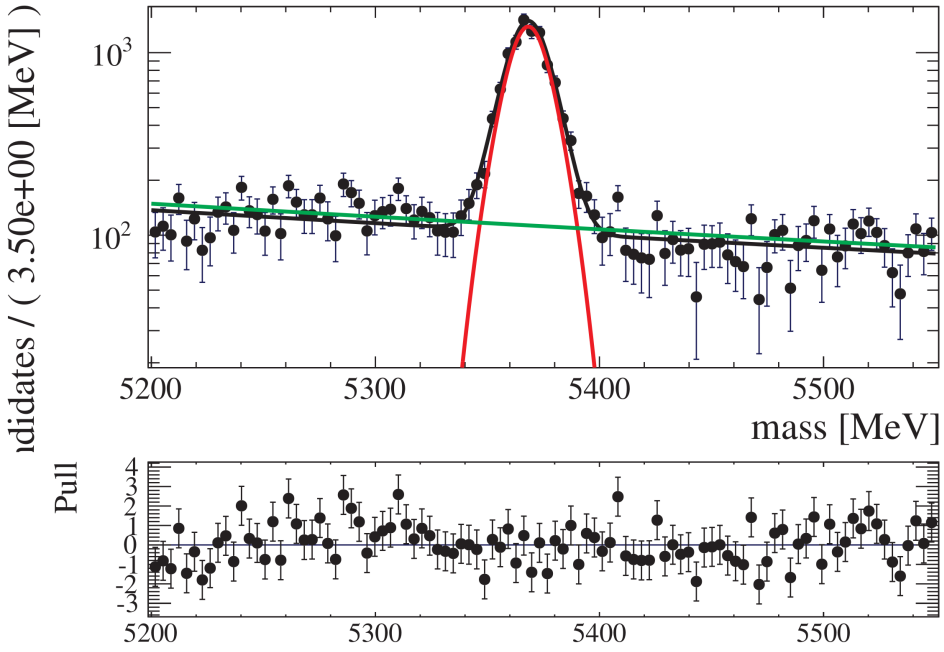


Figure 9.5: *Mass projection of the fit to the dataset weighted using $\omega_2(\theta_{tr})$ to determine Γ_{L}. The Red line corresponds to the signal component, the Green line the background and the black line the PDF total.*

9.3 Fitting to the Projections

The PDF used to fit to the projected data in Figures 9.2, 9.3, 9.4 and 9.5 describes both the signal and background components in decay time and mass,

$$\text{PDF}\left(t, m; \vec{\mathcal{P}}\right) = f_{\mathcal{S}} \frac{\mathcal{S}\left(t, m; \vec{\mathcal{P}}\right)}{\int dt\, dm\; \mathcal{S}\left(t, m; \vec{\mathcal{P}}\right)} + (1 - f_{\mathcal{S}}) \frac{\mathcal{B}\left(t, m; \vec{\mathcal{P}}\right)}{\int dt\, dm\; \mathcal{B}\left(t, m; \vec{\mathcal{P}}\right)} \;, \tag{9.6}$$

where:

$f_{\mathcal{S}} = \text{Signal Fraction}$,

$\mathcal{S}(t, m) = \left[e^{-(t/\tau)} \otimes R(t|\sigma_t; s_t) \times \varepsilon^t\left(t\right)\right] \times \mathcal{S}(m)$,

$\mathcal{B}(t, m) = \left(\left[f_{LL1}\, e^{-\Gamma_{LL1} t} + (1 - f_{LL1})\, e^{-\Gamma_{LL2} t}\right] \otimes R(t|\sigma_t; s_t) \times \varepsilon^t\left(t\right)\right) \times \mathcal{B}(m)$.

Here the PDF $\left(t, m; \vec{\mathcal{P}}\right)$ is composed of the sum of two PDFs describing the signal, $\mathcal{S}(t, m)$, and background, $\mathcal{B}(t, m)$, species within the dataset. The signal species is described by a single exponential function in time multiplied by the mass signal PDF $\mathcal{S}(m)$ (a double Gaussian function). The signal decay width is either $\Gamma_{\rm H}$ or $\Gamma_{\rm L}$ depending upon which projection is being fit. The background species is described by two decay time exponentials (for the short and long lived background components), multiplied by the exponential background in mass $\mathcal{B}(m)$ (Section 5.1).

The time resolution function $R(t, \sigma_t)$ and decay time acceptance $\varepsilon^t(t)$ are the same as used in the $\rm B_s^0 \to J/\psi\, K^+K^-$ analysis (Section 6.1). To be compatible with the other results in this thesis there are six separate mass background models, one for each of the 6 bins in $m(\rm K^+K^-)$ as described in Section 6.5.1.

The numerical results from the weighted fits to equation 9.6 are given in Table 9.4. Here, the precision of the expected values from the $\rm B_s^0 \to J/\psi\, K^+K^-$ analysis is expected to be better as this analysis makes use of more information when fitting to data. The expected values are calculated using the values, uncertainties and correlations for $\Gamma_{\rm s}$ and $\Delta\Gamma_{\rm s}$ quoted in Chapter 8. The corresponding decay time projections for $\Gamma_{\rm H}$ and $\Gamma_{\rm L}$ are shown in Figures 9.2 and 9.3 respectively.

$B_s^0 \to J/\psi\, K^+K^-$ Analysis:

$\Gamma_H\ [\mathrm{ps}^{-1}]$	$\Gamma_L\ [\mathrm{ps}^{-1}]$
$\Gamma_H^* = 0.622 \pm 0.010$	$\Gamma_L^* = 0.722 \pm 0.009$

Angular Moment Analysis:

$\Gamma_H\ [\mathrm{ps}^{-1}]$	$\Gamma_L\ [\mathrm{ps}^{-1}]$
$\Gamma_H^* = 0.627 \pm 0.018$	$\Gamma_L^* = 0.721 \pm 0.011$

Table 9.4: *Table comparing the expected vs the measured decay time of the different mass eigenstates from data. The expected decay times are calculated using:* $\Gamma_s{}^* = 0.6715 \pm 0.0048\,\mathrm{ps}^{-1}$, $\Delta\Gamma_s = 0.100 \pm 0.016\,\mathrm{ps}^{-1}$, $\rho(\Gamma_s, \Delta\Gamma_s) = -0.36$. *(* Here the measured decay width has not been corrected for the effect of upper decay time acceptance.)*

The χ^2/nDoF values from these fits are $\chi^2/\mathrm{nDoF}\,(\Gamma_H) = 2.1$ and $\chi^2/\mathrm{nDoF}\,(\Gamma_L) = 1.9$, both with probabilities of $< 1 \times 10^{-5}$. One of the reasons for the poor fit quality of fit is due to the unknown effect of angular weights on the backgrounds in the dataset. Additionally the poor fit quality is partially due to the fact that the angular moments are used to statistically subtract a large number of events from the dataset. This removal of a large number of events potantially amplifies the effect of statistical fluctuations. These values for χ^2/nDoF suggest that the errors presented in Figures 9.2 and 9.3 are potentially underestimated, as well as the results presented in Table 9.4.

The equivalent measurements of $\Delta\Gamma_s$ and Γ_s using the angular moment analysis are quoted in Table G.3. Here the uncertainties are calculated using the values in Table 9.4 in combination with the correlation factor $\rho(\Gamma_H, \Gamma_L) = -0.789$ calculated from MC datasets as described in Appendix G.2. There is a very good agreement between the angular moment analysis and the $B_s^0 \to J/\psi\, K^+K^-$ analysis which is further qualitative evidence that the measurements from the complex analysis in this thesis are correct.

$B_s^0 \to J/\psi\, K^+K^-$ Analysis:

Γ_s $[\,ps^{-1}]$	$\Delta\Gamma_s$ $[\,ps^{-1}]$
$\Gamma_s = 0.6631 \pm 0.0048$	$\Delta\Gamma_s = 0.100 \pm 0.016$

Angular Moment Analysis:

Γ_s $[\,ps^{-1}]$	$\Delta\Gamma_s$ $[\,ps^{-1}]$
$\Gamma_s = 0.6661 \pm 0.0050$	$\Delta\Gamma_s = 0.094 \pm 0.025$

Table 9.5: *This is a comparison between the results from the full analysis and the measurements of Γ_s and $\Delta\Gamma_s$ extracted from angular moment fits. The uncertainties quoted here are statistical only.*

"If you torture the data enough, nature will always confess."

Ronald Coase

10

Summary of Results and Outlook

The analysis of $B_s^0 \to J/\psi\, K^+K^-$ makes use of $1.0\,\text{fb}^{-1}$ of data collected by LHCb during 2011 and contains $\approx 27,600$ signal events. The results from this full analysis represent the single most precise measurement ϕ_s, Γ_s and $\Delta\Gamma_s$,

$$\begin{aligned}
\phi_s &= 0.07 \pm 0.09\ (\text{stat}) \pm 0.01\ (\text{syst})\ \text{rad} \\
\Gamma_s &= 0.663 \pm 0.005\,(\text{stat}) \pm 0.006\,(\text{syst})\ \text{ps}^{-1}, \\
\Delta\Gamma_s &= 0.100 \pm 0.016\,(\text{stat}) \pm 0.003\,(\text{syst})\ \text{ps}^{-1}.
\end{aligned} \tag{10.1}$$

The $B_s^0 \to J/\psi\, K^+K^-$ analysis is also able to perform an independent measurement of $\Delta m_s = 17.70 \pm 0.10 \pm 0.01\,\text{ps}^{-1}$. Combining this analysis with that of $B_s^0 \to J/\psi\, \pi^+\pi^-$ improves the statistical precision of the results,

$$\begin{aligned}
\phi_s &= 0.01 \pm 0.07\ (\text{stat}) \pm 0.01\ (\text{syst})\ \text{rad} \\
\Gamma_s &= 0.661 \pm 0.004\,(\text{stat}) \pm 0.006\,(\text{syst})\ \text{ps}^{-1}, \\
\Delta\Gamma_s &= 0.106 \pm 0.011\,(\text{stat}) \pm 0.007\,(\text{syst})\ \text{ps}^{-1}.
\end{aligned} \tag{10.2}$$

These results are the most precise measurement to date and are in agreement with Standard Model predictions. This result and other independent measurements of ϕ_s and $\Delta\Gamma_s$ made by different analyses are shown in Figure 10.1. In this figure the global average (shown in grey) calculated by the HFAG [83] is in agreement with the Standard Model predictions. In this figure it can be seen that the global average is dominated by the LHCb measurement described in this thesis.

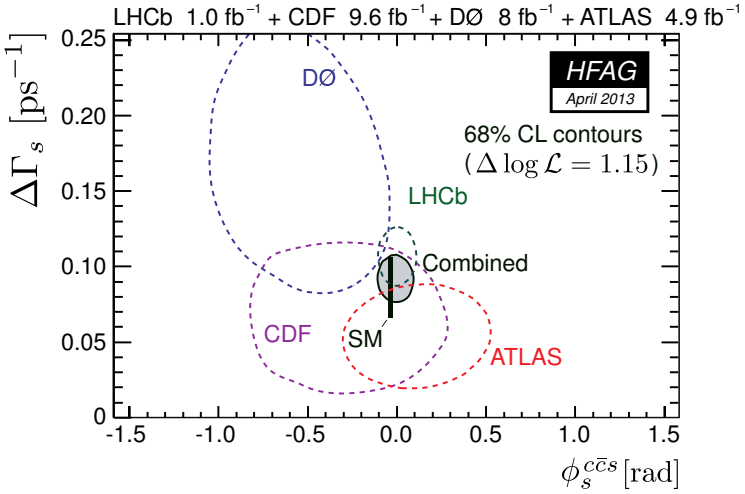


Figure 10.1: *HFAG [83] combined measurement of $\Delta\Gamma_s$ and ϕ_s as of April 2013 [84]. This result includes independent measurements from DØ [10], CDF [85], ATLAS [86] and LHCb [12]. The outlined regions correspond to the 68% Confidence Limit (CL) from each measurement and the grey region corresponds to the combined 68% CL.*

In addition to the measurements from the $B_s^0 \to J/\psi\, K^+K^-$ analysis, measurements of Γ_s and $\Delta\Gamma_s$ have been presented from an angular moment analysis and these results are shown below,

$$\begin{aligned} \Gamma_s\ (Angular Moment) &= 0.666 \pm 0.005\,\text{ps}^{-1} \\ \Delta\Gamma_s\ (Angular Moment) &= 0.094 \pm 0.025\,\text{ps}^{-1} \end{aligned} \tag{10.3}$$

These results are in good agreement with those from the full $B_s^0 \to J/\psi\, K^+K^-$ analysis and give qualitative support that the results from the more complex analysis are correct.

10.1 Constraints on new physics

New physics beyond the Standard Model can potentially modify the measured value of ϕ_s from the prediction in Section 2.3. It is expected that new physics will most likely modify ϕ_s through modifications to the weak mixing phase ϕ_M. This could occur due to new particles contributing to the B_s^0 mixing Feynman diagram in Figure 2.4. The matrix, $\mathbf{R}$ describing the evolution of these states is defined as,

$$\mathbf{R} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{12}^{*} & \mathcal{M}_{22} \end{bmatrix} - \frac{i}{2}\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^{*} & \Gamma_{22} \end{bmatrix} . \tag{10.4}$$

Here, the weak mixing phase is defined as $\phi_{\mathrm{M}} = \arg(\mathcal{M}_{12})$. Modifications to ϕ_M can be described in a model independent way using a complex scale factor Δ_s,

$$\mathcal{M}_{12} = \mathcal{M}_{12}^{SM} \cdot \Delta_s = |\Delta_s| \, \mathcal{M}_{12}^{SM} e^{-\phi_{\Delta_s}} . \tag{10.5}$$

Here Δ_s is related to the scale of the new physics effects expressed in terms of its magnitude $|\Delta_s|$ and associated complex phase ϕ_{Δ_s}. The Standard Model prediction for this quantity is therefore $|\Delta_s| = 1$, $\phi_{\Delta_s} = 0$. The recent measurement places strong constraints on new physics contributions to ϕ_{s}. The constraints on Δ_s as calculated by the CKMFitter group [87] are shown in Figure 10.2 which covers the complex plane of $\Re\mathfrak{e}\,(\Delta_s)$ vs $\Im\mathfrak{m}\,(\Delta_s)$.

Here the solid shaded ranges correspond to 68% Confidence Limits (CL) from various analyses, with the constraint from the recent ϕ_{s} measurements indicated in solid blue. The red dashed region indicates the combined 68% CL region from all of the analyses, this is compatible with the Standard Model expectation. The contribution of any new physics to $\mathrm{B_s^0}$ mixing must be small in order to be compatible with the measurements made of $\mathcal{CP}$ violation within this process.

10.2 Outlook

Throughout 2012 the LHCb experiment collected an additional $2.2\,\mathrm{fb}^{-1}$ of data which is currently being analysed. From this, a combined fit using the 2011+2012 datasets is expected to have a statistical uncertainty on ϕ_{s} of $\approx 0.04\,\mathrm{rad}$, improving on the current best measurement. The LHCb experiment is to be upgraded in 2018 and is expected to collect $50\,\mathrm{fb}^{-1}$ of data throughout the upgraded run. The analysis of $\mathrm{B_s^0} \to \mathrm{J/\psi\,K^+K^-}$ with this dataset is expected to achieve a statistical sensitivity on the measurement of ϕ_{s} of $0.008\,\mathrm{rad}$. As the theoretical precision on ϕ_{s} is $0.003\,\mathrm{rad}$ [19], a measurement made using the $50\,\mathrm{fb}^{-1}$ of data from the upgrade will place a very strong constraint on the effects of new physics.

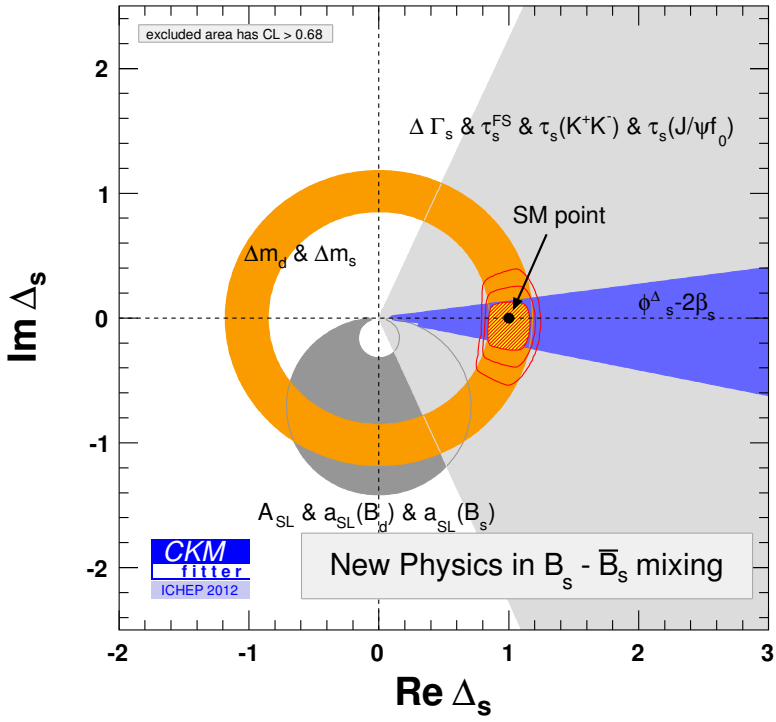


Figure 10.2: *Constraints placed on Δ_s as calculated by the CKMFitter group [87]. With the constraints from the latest set ϕ_s measurements given by the solid blue region. The 68% CL region is given by the red dashed region, and the 68%, 95% and 99% C.L. are given by the solid red lines.*

A

Time Evolution of States

This appendix describes the time evolution of a neutral B meson state composed to two eigenstates with different lifetimes.

A.1 Mass and lifetime splitting

The general equation describing the time evolution of two states is given in A.1. From this it is possible to relate the mass and lifetime splitting in terms of elements within the given matrices.

$$\mathbf{R} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} .$$

$$\mathbf{R} = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{21} & \mathcal{M}_{22} \end{bmatrix} - \frac{i}{2}\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} \mathrm{M} - \frac{i}{2}\Gamma & \mathcal{M}_{12} - \frac{i}{2}\Gamma_{12} \\ \mathcal{M}^*_{12} - \frac{i}{2}\Gamma^*_{12} & \mathrm{M} - \frac{i}{2}\Gamma \end{bmatrix} . \quad \text{(A.1)}$$

First the eigenstates of A.1 are calculated making use of the relation $\mathbf{R} - \lambda 1 = 0$. From this it is possible to construct equation A.2.

$$\lambda_\pm = \left(M - \frac{i}{2}\Gamma\right) \pm \sqrt{\left(\mathcal{M}_{12} - \frac{i}{2}\Gamma_{12}\right)\left(\mathcal{M}^*_{12} - \frac{i}{2}\Gamma^*_{12}\right)} . \quad \text{(A.2)}$$

Defining the eigenstates to take the form of A.3, makes it clear that collecting the real and imaginary terms gives the definitions given in A.4.

$$\lambda_{\mathrm{H}} = \left(M + \frac{\Delta M}{2}\right) + \frac{i}{2}\left(\Gamma + \frac{\Delta\Gamma}{2}\right) \ , \ \lambda_{\mathrm{L}} = \left(M - \frac{\Delta M}{2}\right) + \frac{i}{2}\left(\Gamma - \frac{\Delta\Gamma}{2}\right) \ . \tag{A.3}$$

$$\begin{aligned}
\Delta M &= 2\mathfrak{Re}\sqrt{\left(\mathcal{M}_{12} - \frac{i}{2}\Gamma_{12}\right)\left(\mathcal{M}^*_{12} - \frac{i}{2}\Gamma^*_{12}\right)} = \mathrm{M_H} - \mathrm{M_L} \ , \\
\Delta\Gamma &= 4\mathfrak{Im}\sqrt{\left(\mathcal{M}_{12} - \frac{i}{2}\Gamma_{12}\right)\left(\mathcal{M}^*_{12} - \frac{i}{2}\Gamma^*_{12}\right)} = \Gamma_{\mathrm{L}} - \Gamma_{\mathrm{H}} \\
\overline{\Gamma} &= \frac{\Gamma_{\mathrm{L}} + \Gamma_{\mathrm{H}}}{2} \ .
\end{aligned} \tag{A.4}$$

Using the eigenstates given in equation A.2 and solving the eigenstate problem in equation A.5 the fraction $\frac{p}{q}$ is defined as in equation A.6.

$$\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = \lambda_{\pm} \begin{bmatrix} p \\ q \end{bmatrix} \ . \tag{A.5}$$

$$\frac{q}{p} = \sqrt{\frac{\mathcal{M}^*_{12} - \frac{i}{2}\Gamma^*_{12}}{\mathcal{M}_{12} - \frac{i}{2}\Gamma_{12}}} \ . \tag{A.6}$$

A.2 Time evolution of neutral B mesons

By definition the $\mathcal{CP}$ eigenstates of neutral B mesons are composed of two different flavour mesons as defined in A.7.

$$|\mathrm{B_H}\rangle = p|\mathrm{B}\rangle + q|\overline{\mathrm{B}}\rangle \quad |\mathrm{B_L}\rangle = p|\mathrm{B}\rangle - q|\overline{\mathrm{B}}\rangle \ . \tag{A.7}$$

Here $\mathrm{B_H}$ is defines as the Heavy mass eigenstate and $\mathrm{B_L}$ is referred to as the Light mass eigenstate. The time evolution of these mass eigenstates is governed by equation A.8.

$$|\mathrm{B}_{\binom{\mathrm{H}}{\mathrm{L}}}(\mathrm{t})\rangle = e^{i\mathrm{M}\binom{\mathrm{H}}{\mathrm{L}}\mathrm{t}} e^{-\frac{1}{2}\Gamma\binom{\mathrm{H}}{\mathrm{L}}\mathrm{t}} |\mathrm{B}_{\binom{\mathrm{H}}{\mathrm{L}}}(\mathrm{t}=0)\rangle \ . \tag{A.8}$$

From equations A.7 and A.8 the time evolution of the two different flavour eigenstates is given in equation A.9.

$$
\begin{aligned}
|\mathrm{B(t)}\rangle &= \frac{1}{2p}\left[e^{i\left(\mathrm{M_H}+\frac{i}{2}\Gamma_\mathrm{H}\right)\mathrm{t}}|\mathrm{B_H\,(t=0)}\rangle + e^{i\left(\mathrm{M_L}+\frac{i}{2}\Gamma_\mathrm{L}\right)t}|\mathrm{B_L\,(t=0)}\rangle\right] \\
&= \frac{1}{2p}\left[e^{i\left(\mathrm{M_H}+\frac{i}{2}\Gamma_\mathrm{H}\right)t}\left(p|\mathrm{B\,(0)}\rangle - q|\overline{\mathrm{B}}\,(0)\rangle\right) + e^{i\left(\mathrm{M_L}+\frac{i}{2}\Gamma_\mathrm{L}\right)t}\left(p|\mathrm{B\,(0)}\rangle + q|\overline{\mathrm{B}}\,(0)\rangle\right)\right] \\
&= \begin{bmatrix} \left(\frac{e^{i\left(\mathrm{M_H}+\frac{i}{2}\Gamma_\mathrm{H}\right)t}+e^{i\left(\mathrm{M_L}+\frac{i}{2}\Gamma_\mathrm{L}i\right)t}}{2}\right) & |\mathrm{B\,(0)}\rangle \\ +\frac{q}{p} \quad \left(\frac{e^{\left(\mathrm{M_H}+\frac{i}{2}\Gamma_\mathrm{H}\right)t}-e^{i\left(\mathrm{M_L}+\frac{i}{2}\Gamma_\mathrm{L}\right)t}}{2}\right) & |\overline{\mathrm{B}}\,(0)\rangle \end{bmatrix} .
\end{aligned} \tag{A.9}
$$

Here the identities $g_+(t)$ and $g_-(t)$ A.10 are also introduced for convenience.

$$
g_+(t) = \frac{e^{-i\left(\mathrm{M_L}-\frac{i}{2}\Gamma_\mathrm{L}t\right)} + e^{-i\left(\mathrm{M_H}-\frac{i}{2}\Gamma_\mathrm{H}t\right)}}{2} \ , \ g_-(t) = \frac{e^{-i\left(\mathrm{M_L}-\frac{i}{2}\Gamma_\mathrm{L}t\right)} - e^{-i\left(\mathrm{M_H}-\frac{i}{2}\Gamma_\mathrm{H}t\right)}}{2} \ . \tag{A.10}
$$

Using these definitions equation A.9 to A.11 and the equivalent equation governing the evolution of an initial $\overline{\mathrm{B}}$ state is given.

$$
\begin{aligned}
|\mathrm{B(t)}\rangle &= g_+(t)\,|\mathrm{B\,(0)}\rangle + \frac{q}{p}g_-(t)\,|\overline{\mathrm{B}}\,(0)\rangle \ , \\
|\overline{\mathrm{B}}\mathrm{(t)}\rangle &= \frac{p}{q}g_+(t)\,|\mathrm{B\,(0)}\rangle + g_-(t)\,|\overline{\mathrm{B}}\,(0)\rangle \ .
\end{aligned} \tag{A.11}
$$

Using this notation the evolution of an initial B meson at t = 0 meson to a B meson at time t is given in by equation A.12.

$$
\begin{aligned}
\mathrm{P\,(B\,(0) \to B\,(t))} &= |\langle \mathrm{B\,(0)\,|B\,(t)}\rangle|^2 \\
&= \left|g_+(t)\,\langle \mathrm{B\,(0)\,|B\,(0)}\rangle + \frac{q}{p}g_-(t)\,\langle \mathrm{B\,(0)\,|}\overline{\mathrm{B}}\,(0)\rangle\right|^2 \\
&= |g_+(t)\,\langle \mathrm{B\,(0)\,|B\,(0)}\rangle|^2 = g_+(t)^2 \\
&= \frac{e^{-\overline{\Gamma}\mathrm{t}}}{2}\left[\cosh\left(\frac{1}{2}\Delta\Gamma t\right) + \cos\left(\Delta m t\right)\right] \ .
\end{aligned} \tag{A.12}
$$

The expression governing the evolution of an initial B meson into a $\overline{\mathrm{B}}$ meson at time t is given in equation A.13.

$$\begin{aligned}\mathrm{P}\left(\mathrm{B}\left(0\right)\rightarrow\overline{\mathrm{B}}\left(\mathrm{t}\right)\right) &= \frac{p^2}{q^2}g_{-}\left(t\right)^2 \\ &= \frac{p^2}{q^2}\frac{e^{-\overline{\Gamma}\mathrm{t}}}{2}\left[\cosh\left(\frac{1}{2}\Delta\Gamma t\right)-\cos\left(\Delta m t\right)\right] \ . \end{aligned} \tag{A.13}$$

Similarly the equivalent expressions governing the time evolution of an initial $\overline{\mathrm{B}}$ meson are given in equations A.14 and A.15.

$$\mathrm{P}\left(\overline{\mathrm{B}}\left(0\right)\rightarrow\overline{\mathrm{B}}\left(\mathrm{t}\right)\right) = g_{+}\left(t\right)^2 = \frac{e^{-\overline{\Gamma}\mathrm{t}}}{2}\left[\cosh\left(\frac{1}{2}\Delta\Gamma t\right)+\cos\left(\Delta m t\right)\right] \ . \tag{A.14}$$

$$\mathrm{P}\left(\overline{\mathrm{B}}\left(0\right)\rightarrow\mathrm{B}\left(\mathrm{t}\right)\right) = \frac{p^2}{q^2}g_{-}\left(t\right)^2 = \frac{p^2}{q^2}\frac{e^{-\overline{\Gamma}\mathrm{t}}}{2}\left[\cosh\left(\frac{1}{2}\Delta\Gamma t\right)-\cos\left(\Delta m t\right)\right] \ . \tag{A.15}$$

Using these definitions it is possible to calculate the probability of a given meson decaying into the final state f. This is calculated in equation A.16. Here the identities $A_f = \langle f|\mathcal{H}|\mathrm{B}\rangle$ and $\overline{A}_f = \langle f|\mathcal{H}|\overline{\mathrm{B}}\rangle$ are used.

$$\begin{aligned}&\mathrm{P}\left(\mathrm{B(t)}\rightarrow f\right) = \left|\langle f|\mathrm{B(t)}\rangle\right|^2 = \langle f|\mathrm{B(t)}\rangle\langle f|\mathrm{B(t)}\rangle^{\dagger} \\ &= \left[g_{+}\left(t\right)\langle f|\mathrm{B}\left(0\right)\rangle + \frac{q}{p}g_{-}\left(t\right)\langle f|\overline{\mathrm{B}}\left(0\right)\rangle\right]\times\left[g_{+}\left(t\right)\langle f|\mathrm{B}\left(0\right)\rangle + \frac{q}{p}g_{-}\left(t\right)\langle f|\overline{\mathrm{B}}\left(0\right)\rangle\right]^{\dagger} \\ &= \begin{bmatrix} |g_{+}\left(t\right)|^2|A_f\left(0\right)|^2 & +\left|\frac{q}{p}\right|^2|g_{-}\left(t\right)|^2\left|\overline{A}_f\left(0\right)\right|^2 & +\frac{q}{p}g_{-}\left(t\right)g_{+}\left(t\right)^{*}|A_f\left(0\right)|^{*}\left|\overline{A}_f\left(0\right)\right| \\ & & +\left(\frac{q}{p}\right)^{*}g_{+}\left(t\right)g_{-}\left(t\right)^{*}|A_f\left(0\right)|\left|\overline{A}_f\left(0\right)\right|^{*} \end{bmatrix} \ . \end{aligned} \tag{A.16}$$

Using: $\overline{\Gamma} = \dfrac{\Gamma_{\mathrm{H}} + \Gamma_{\mathrm{L}}}{2}$ and $\Delta\Gamma = \Gamma_{\mathrm{L}} - \Gamma_{\mathrm{H}}$, the time dependent functions are defined as A.17.

$$
\begin{aligned}
g_{+}(t)^2 &= \frac{e^{-\overline{\Gamma}t}}{2}\left[\cosh\left(\frac{1}{2}\Delta\Gamma t\right) + \cos\left(\Delta m t\right)\right] \ , \\
g_{-}(t)^2 &= \frac{e^{-\overline{\Gamma}t}}{2}\left[\cosh\left(\frac{1}{2}\Delta\Gamma t\right) - \cos\left(\Delta m t\right)\right] \ , \\
g_{+}(t)\, g_{-}(t)^{\dagger} &= \frac{e^{-\overline{\Gamma}t}}{2}\left[-\sinh\left(\frac{1}{2}\Delta\Gamma t\right) + i\sin\left(\Delta m t\right)\right] \ , \\
g_{-}(t)\, g_{+}(t)^{\dagger} &= \frac{e^{-\overline{\Gamma}t}}{2}\left[-\sinh\left(\frac{1}{2}\Delta\Gamma t\right) - i\sin\left(\Delta m t\right)\right] \ . \qquad \text{(A.17)}
\end{aligned}
$$

$$
\mathrm{P}\left(\mathrm{B(t)} \to f\right) = \frac{e^{-\overline{\Gamma}t}}{2}\begin{bmatrix}
 & |A_f(0)|^2 & \cosh\left(\frac{1}{2}\Delta\Gamma t\right) + \cos\left(\Delta m t\right) \\
+ & \left|\overline{A}_f(0)\right|^2\left|\frac{q}{p}\right|^2 & \left\{\cosh\left(\frac{1}{2}\Delta\Gamma t\right) - \cos\left(\Delta m t\right)\right\} \\
+ & |A_f(0)|^*\left|\overline{A}_f(0)\right|\frac{q}{p} & \left\{-\sinh\left(\frac{1}{2}\Delta\Gamma t\right) - i\sin\left(\Delta m t\right)\right\} \\
+ & |A_f(0)|\left|\overline{A}_f(0)\right|^*\left(\frac{q}{p}\right)^* & \left\{-\sinh\left(\frac{1}{2}\Delta\Gamma t\right) + i\sin\left(\Delta m t\right)\right\}
\end{bmatrix} ,
$$

$$
\mathrm{P}\left(\mathrm{B(t)} \to f\right) = \frac{e^{-\overline{\Gamma}t}}{2}|A_f(0)|^2\begin{bmatrix}
\left(1 + \left|\frac{q}{p}\right|^2\left|\frac{\overline{A}_f(0)}{A_f(0)}\right|^2\right) & \cosh\left(\frac{1}{2}\Delta\Gamma t\right) \\
+\left(1 - \left|\frac{q}{p}\right|^2\left|\frac{\overline{A}_f(0)}{A_f(0)}\right|^2\right) & \cos\left(\Delta m t\right) \\
-2\mathfrak{Re}\left(\frac{q}{p}\left|\frac{\overline{A}_f(0)}{A_f(0)}\right|\right) & \sinh\left(\frac{1}{2}\Delta\Gamma t\right) \\
-2\mathfrak{Im}\left(\frac{q}{p}\left|\frac{\overline{A}_f(0)}{A_f(0)}\right|\right) & \sin\left(\Delta m t\right)
\end{bmatrix} . \qquad \text{(A.18)}
$$

From the identities in A.17 the time dependent probability A.16 can be expanded to take the form in A.18. Although this equation is the full PDF expressing the time dependent decay rate of $\mathrm{B} \to f$ it is normally written in a more convenient form using the parameter $\lambda_f^0 = \dfrac{q}{p}\dfrac{\overline{A}_f}{A_f}$ as in equation A.19.

$$\mathrm{P}\left(\mathrm{B(t)} \to f\right) = \frac{e^{-\overline{\Gamma}t}}{2}\left|A_f\left(0\right)\right|^2 \begin{bmatrix} \left(1+\left|\lambda_f^0\right|^2\right) & \cosh\left(\frac{1}{2}\Delta\Gamma t\right) \\ +\left(1-\left|\lambda_f^0\right|^2\right) & \cos\left(\Delta m t\right) \\ -2\Re\mathfrak{e}\left(\lambda_f^0\right) & \sinh\left(\frac{1}{2}\Delta\Gamma t\right) \\ -2\Im\mathfrak{m}\left(\lambda_f^0\right) & \sin\left(\Delta m t\right) \end{bmatrix} . \tag{A.19}$$

Here the parameter λ_f^0 contains all of the $\mathcal{CP}$-dependent terms and as such it is important to consider how it behaves under the $\mathcal{CP}$ operator. As such because the behaviour of the possible final states differs under application of the $\mathcal{CP}$ operator the λ_f^0 parameter must be changed for different final state configurations.

The final state is known to be either $\mathcal{CP}$-odd or $\mathcal{CP}$-even. In order to describe the state in a way that is invariant of this transform the parameter λ_f is used. This parameter is defined to include the eigenvalue of the $\mathcal{CP}$ eigenstate as in A.20.

$$\mathcal{CP}\ \lambda_f^0 = \eta_f\ \lambda \quad \text{where:} \begin{cases} \eta_f = +1 & (\mathcal{CP}\text{-even}) \\ \eta_f = -1 & (\mathcal{CP}\text{-odd}) \end{cases} . \tag{A.20}$$

This allows for the different $\mathcal{CP}$ states be constructed using the same definition of λ. This results in the 'master equation' A.21.

$$\mathrm{P}\left(\mathrm{B(t)} \to f\right) = \frac{e^{-\overline{\Gamma}t}}{2}\left|A_f\left(0\right)\right|^2\left(1+\lambda^2\right)\left[\begin{array}{r} a\cosh\left(\frac{1}{2}\Delta\Gamma t\right)+b\sinh\left(\frac{1}{2}\Delta\Gamma t\right)+c\cos\left(\Delta m t\right) \\ +d\sin\left(\Delta m t\right) \end{array}\right] . \tag{A.21}$$

$$a = 1\ ,\ b = \frac{-2\Re\mathfrak{e}\left(\lambda\right)}{\left(1+\lambda^2\right)} = D\ ,\ c = \frac{1-\lambda^2}{\left(1+\lambda^2\right)} = C\ ,\ d = \frac{-2\Im\mathfrak{m}(\lambda)}{\left(1+\lambda^2\right)} = S\ . \tag{A.22}$$

This master equation makes use of the factors a, b, c, d (equation A.22) which vary depending on the physics present.

The equivalent function for the decay of $\overline{\mathrm{B_s^0}} \rightarrow \mathrm{J/\psi\, K^+K^-}$ is shown in equation A.23.

$$\mathrm{P}\left(\overline{\mathrm{B}}(\mathrm{t}) \rightarrow f\right) = \left|\frac{p^2}{q^2}\right| \frac{e^{-\overline{\Gamma}t}}{2} \left|A_f\left(0\right)\right|^2 \left(1+\lambda^2\right) \left[\begin{array}{r} a\cosh\left(\frac{1}{2}\Delta\Gamma t\right) + b\sinh\left(\frac{1}{2}\Delta\Gamma t\right) - c\cos\left(\Delta m t\right) \\ - d\sin\left(\Delta m t\right) \end{array}\right] . \tag{A.23}$$

Under the assumption that $\left|\frac{p^2}{q^2}\right| = 1$ adding both equations A.21 and A.23 gives equation A.24.

$$\begin{pmatrix} \mathrm{P}\left(\mathrm{B}(\mathrm{t}) \rightarrow f\right) \\ +\mathrm{P}\left(\overline{\mathrm{B}}(\mathrm{t}) \rightarrow f\right) \end{pmatrix} \propto e^{\frac{-(\Gamma_\mathrm{H}+\Gamma_\mathrm{L})t}{2}} \begin{pmatrix} & \frac{\left(e^{\frac{(\Gamma_\mathrm{L}-\Gamma_\mathrm{H})t}{2}} + e^{\frac{(\Gamma_\mathrm{H}-\Gamma_\mathrm{L})t}{2}}\right)}{2} \\ + \left[\frac{2\mathfrak{Re}\,\lambda}{1+|\lambda|^2}\right] & \frac{\left(e^{\frac{(\Gamma_\mathrm{L}-\Gamma_\mathrm{H})t}{2}} - e^{\frac{(\Gamma_\mathrm{H}-\Gamma_\mathrm{L})t}{2}}\right)}{2} \end{pmatrix} . \tag{A.24}$$

Assuming no $\mathcal{CP}$ violation, $|\lambda| = 1$, $A_f = \overline{A}_f$. Equation A.24 can be simplified to equation A.25.

$$\begin{pmatrix} \mathrm{P}\left(\mathrm{B}(\mathrm{t}) \rightarrow f\right) \\ +\mathrm{P}\left(\overline{\mathrm{B}}(\mathrm{t}) \rightarrow f\right) \end{pmatrix} \propto \begin{cases} A_1 e^{-\Gamma_\mathrm{L} t} & \mathcal{CP}\text{-odd} \\ + \quad A_2 e^{-\Gamma_\mathrm{H} t} & \mathcal{CP}\text{-even} \end{cases} . \tag{A.25}$$

From this the final untagged time dependence can be seen to be the sum of the two different lifetimes A.25. Here A_1 and A_2 are the amplitudes of the $\mathcal{CP}$-odd and $\mathcal{CP}$-even final states.

B

Angular Weights

This appendix describes the calculation of angular acceptance weights used to incorporate the effect of angular detector acceptance into a PDF denominator.

The angular weights (ξ_i) used to describe the behaviour of the angular acceptance in the denominator of a fit function are defined as in B.1.

$$\begin{aligned}\int \varepsilon^a \mathrm{PDF}^a\left(\vec{\Omega}\right) d\vec{\Omega} &= \sum_i \int \varepsilon_i^a \mathrm{PDF}_i^a\left(\vec{\Omega}\right) d\vec{\Omega} \\ \xi_i &= \int \varepsilon^a\left(\vec{\Omega}\right) \mathrm{PDF}_i^a\left(\vec{\Omega}\right) d\vec{\Omega}\end{aligned} \tag{B.1}$$

From this definition of the angular acceptance weights it is possible to express the angular weights in a more generic form as described in B.2.

$$\begin{aligned}\xi_k &= \left(\int \frac{\mathcal{S}_{\mathrm{obs}}\left(t\right)}{\langle\varepsilon^a\rangle\left(t\right)} dt\right) \times \int \varepsilon^a\left(\vec{\Omega}\right) f_k\left(\vec{\Omega}\right) d\vec{\Omega} \Big/ \int \frac{\mathcal{S}_{\mathrm{obs}}\left(t\right)}{\langle\varepsilon^a\rangle\left(t\right)} dt \\ \text{defining: } \mathrm{E}^a &= \left[\int \frac{\mathcal{S}_{\mathrm{obs}}\left(t\right)}{\langle\varepsilon^a\rangle\left(t\right)\, dt}\right]^{-1} \\ \frac{1}{\mathrm{E}^a}\xi_k &= \left(\int \frac{\mathcal{S}_{\mathrm{obs}}\left(t\right)}{\langle\varepsilon^a\rangle\left(t\right)} dt\right) \times \int \varepsilon^a\left(\vec{\Omega}\right) f_k\left(\vec{\Omega}\right) d\vec{\Omega} \\ &= \int \mathcal{S}_{\mathrm{obs}}\left(t\right) \frac{\varepsilon^a\left(\vec{\Omega}\right)}{\langle\varepsilon^a\rangle\left(t\right)} f_k\left(\vec{\Omega}\right) dt d\vec{\Omega}\end{aligned} \tag{B.2}$$

Here the parameter E^a reduces to the time averaged angular acceptance when the time and angular acceptance effects are uncorrelated. The parameter $\mathcal{S}_{\mathrm{obs}}\left(t\right)$ is the observed distribution of events in decay time from MC, $f_k\left(\vec{\Omega}\right)$ is the k-th angular term of the PDF describing the underlying physics, and $\langle\varepsilon^a\rangle\left(t\right)$ is the time-averaged angular acceptance. From this equation it is required that the term $\frac{\varepsilon^a\left(\vec{\Omega}\right)}{\langle\varepsilon^a\rangle\left(t\right)}$ is to be expressed using using information from MC. Equation B.3 relates this term to the ratio of the observed and underlying angular distributions of the particles from MC.

$$
\begin{aligned}
\mathcal{S}_{\mathrm{obs}}\left(\vec{\Omega}\,|\,t\right) &= \frac{\varepsilon\left(t,\vec{\Omega}\right)\mathrm{PDF}\left(t,\vec{\Omega}\right)}{\int\varepsilon\left(t,\vec{\Omega}\right)\mathrm{PDF}\left(t,\vec{\Omega}\right)\,d\vec{\Omega}} \\
\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}\,|\,t\right) &= \frac{\mathrm{PDF}\left(t,\vec{\Omega}\right)}{\int\mathrm{PDF}\left(t,\vec{\Omega}\right)\,d\vec{\Omega}} \\
\left[\frac{\mathcal{S}_{\mathrm{obs}}\left(\vec{\Omega}\,|\,t\right)}{\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}\,|\,t\right)}\right] &= \frac{\varepsilon\left(t,\vec{\Omega}\right)\mathrm{PDF}\left(t,\vec{\Omega}\right)}{\int\varepsilon\left(t,\vec{\Omega}\right)\mathrm{PDF}\left(t,\vec{\Omega}\right)\,d\vec{\Omega}}\Bigg/\frac{\mathrm{PDF}\left(t,\vec{\Omega}\right)}{\int\mathrm{PDF}\left(t,\vec{\Omega}\right)\,d\vec{\Omega}} \\
&= \frac{\varepsilon\left(t,\vec{\Omega}\right)}{\int\varepsilon\left(t,\vec{\Omega}\right)\left\{\dfrac{\mathrm{PDF}\left(t,\vec{\Omega}\right)}{\int\mathrm{PDF}\left(t,\vec{\Omega}\right)d\vec{\Omega}}\right\}d\vec{\Omega}} \\
&= \frac{\varepsilon\left(t,\vec{\Omega}\right)}{\int\varepsilon\left(t,\vec{\Omega}\right)\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}\,|\,t\right)d\vec{\Omega}} = \frac{\varepsilon\left(t,\vec{\Omega}\right)}{\varepsilon^t\left(t\right)\langle\varepsilon^a\rangle\left(t\right)} = \frac{\varepsilon^a\left(\vec{\Omega}\right)}{\langle\varepsilon^a\rangle\left(t\right)}
\end{aligned}
\tag{B.3}
$$

Combining these two equations (B.3 and B.2) leads to the relation which is used to calculate the angular weights B.4.

$$
\begin{aligned}
\frac{1}{\mathrm{E}^a}\xi_k &= \int\mathcal{S}_{\mathrm{obs}}\left(t\right)\frac{\mathcal{S}_{\mathrm{obs}}\left(\vec{\Omega}\,|\,t\right)}{\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}\,|\,t\right)}f_k(\vec{\Omega})\,d\vec{\Omega}dt \\
&= \int\mathcal{S}_{\mathrm{obs}}\left(\vec{\Omega}\,,\,t\right)\frac{f_k\left(\vec{\Omega}\right)}{\mathcal{S}_{\mathrm{phys}}\left(\vec{\Omega}\,|\,t\right)}\,d\vec{\Omega}dt
\end{aligned}
\tag{B.4}
$$

This particular relation is actually approximated in the equation that is used to calculate the acceptance weights from data. The final approximation takes the

k	ξ_k			
	Data Subset:			
	Untagged Events	Tagged Events	*almost unbiased*	*exclusively biased*
1	$+0.9779 \pm 0.0011$	$+0.9835 \pm 0.0015$	$+0.9792 \pm 0.0010$	$+0.9851 \pm 0.0024$
2	$+1.0308 \pm 0.0018$	$+1.0252 \pm 0.0024$	$+1.0295 \pm 0.0016$	$+1.0239 \pm 0.0039$
3	$+1.0312 \pm 0.0020$	$+1.0235 \pm 0.0026$	$+1.0291 \pm 0.0017$	$+1.0231 \pm 0.0042$
4	-0.0012 ± 0.0019	-0.0007 ± 0.0025	-0.0012 ± 0.0016	$+0.0003 \pm 0.0040$
5	-0.0011 ± 0.0013	$+0.0009 \pm 0.0016$	-0.0009 ± 0.0011	$+0.0028 \pm 0.0027$
6	-0.0005 ± 0.0012	-0.0004 ± 0.0015	-0.0003 ± 0.0010	-0.0014 ± 0.0025
7	$+0.9903 \pm 0.0012$	$+0.9937 \pm 0.0015$	$+0.9919 \pm 0.0010$	$+0.9899 \pm 0.0025$
8	$+0.0009 \pm 0.0018$	$+0.0011 \pm 0.0023$	$+0.0008 \pm 0.0015$	$+0.0020 \pm 0.0037$
9	$+0.0026 \pm 0.0017$	-0.0023 ± 0.0021	$+0.0009 \pm 0.0014$	$+0.0004 \pm 0.0035$
10	-0.0038 ± 0.0035	-0.0129 ± 0.0045	-0.0067 ± 0.0030	$+0.0103 \pm 0.0075$

Table B.1: *The values of the angular acceptance weights calculated using different subsets from the fully simulated MC. These weights are for illustration purposes only and aren't used in fitting.*

limit of using a large MC dataset to integrate over the observed MC distribution. This is expressed in equation B.5.

$$\mathrm{E}\left(\frac{1}{\mathrm{E}^a}\xi_\mathrm{k}\right) = \frac{1}{\mathrm{N_{obs}}}\sum_{\mathrm{e}=1}^{\mathrm{N_{obs}}}\frac{f_\mathrm{k}\left(\Omega_\mathrm{e}\right)}{\mathcal{S}_\mathrm{phys}\left(\vec{\Omega}_e \,|\, t_e\right)} \tag{B.5}$$

As the parameter E^a is a constant it is often ignored in fitting to data.

For the analysis of $\mathrm{B_s^0} \to \mathrm{J/\psi\, K^+K^-}$ the angular weights calculated using the whole dataset are given in Table 4.8. The angular weights are also recalculated using various subsets of the data as shown in Table B.1. It is seen that the results calculated from the various subsets of the whole dataset are compatible with the numbers used in the main analysis.

C

Mass Fits in Six Bins

This appendix describes the results from fitting to the $m(\mathrm{J/\psi\ K^+K^-})$ mass distribution fitted in six bins of $m(\mathrm{K^+K^-})$

When performing the full $\mathrm{B_s^0} \to \mathrm{J/\psi\ K^+K^-}$ analysis the $m_{\mathrm{B_s^0}}$ fit is performed in six bins in $m(\mathrm{K^+K^-})$ as described in Section 6.5. The PDF used in fitting to the mass distribution is described in this section. In order to construct the fit this way each bin has an independent signal fraction and background model. A separate signal fraction is used for the *almost unbiased* (f^{UB}) and the exclusively biased (f^{B}) datasets. The numerical fit results for this fit are summarised in Table C.1. The projections of the fits to each of the six bins are shown in Figure C.2. The number of signal events from these fits is $\mathrm{N_{sig}} = 27617 \pm 117$.

	Parameter	Fit result and error
	$f_{\sigma_m{}^1}$	0.75977 $\pm$0.034441
	$\sigma_m{}^1$	6.0801 $\pm$0.13159
	r_{21}	2.0655 $\pm$0.088238
	m_{B_s}	5368.2 $\pm$0.048375
	α_M	0.0017313 $\pm$0.00014683
	f^{UB}	0.10793 $\pm$0.0063747
bin 1	f^B	0.15143 $\pm$0.01756
	N^{UB}	4515
	N^B	716
	α_M	0.0017486 $\pm$0.00017719
	f^{UB}	0.4729 $\pm$0.007945
bin 2	f^B	0.51274 $\pm$0.01945
	N^{UB}	5276
	N^B	845
	α_M	0.0017361 $\pm$0.00019785
	f^{UB}	0.79351 $\pm$0.0044383
bin 3	f^B	0.81683 $\pm$0.010221
	N^{UB}	10858
	N^B	1776
	α_M	0.0012316 $\pm$0.00019773
	f^{UB}	0.76111 $\pm$0.0050647
bin 4	f^B	0.80411 $\pm$0.011173
	N^{UB}	9351
	N^B	1585
	α_M	0.0015366 $\pm$0.00015527
	f^{UB}	0.4537 $\pm$0.0071086
bin 5	f^B	0.48973 $\pm$0.017756
	N^{UB}	6607
	N^B	1039
	α_M	0.0015859 $\pm$9.9538e-05
	f^{UB}	0.1689 $\pm$0.0046759
bin 6	f^B	0.21286 $\pm$0.012794
	N^{UB}	10594
	N^B	1593

Table C.1: *Fit Results corresponding to $m_{\mathrm{B_s^0}}$ fit in 6 bins of $m(\mathrm{K^+K^-})$ as described in Section 6.5. The signal $m_{\mathrm{B_s^0}}$ mass distribution is common to all six bins and each bin has a separate signal/background fraction and background model for the almost unbiased(UB) and the exclusively biased(B) datasets.*

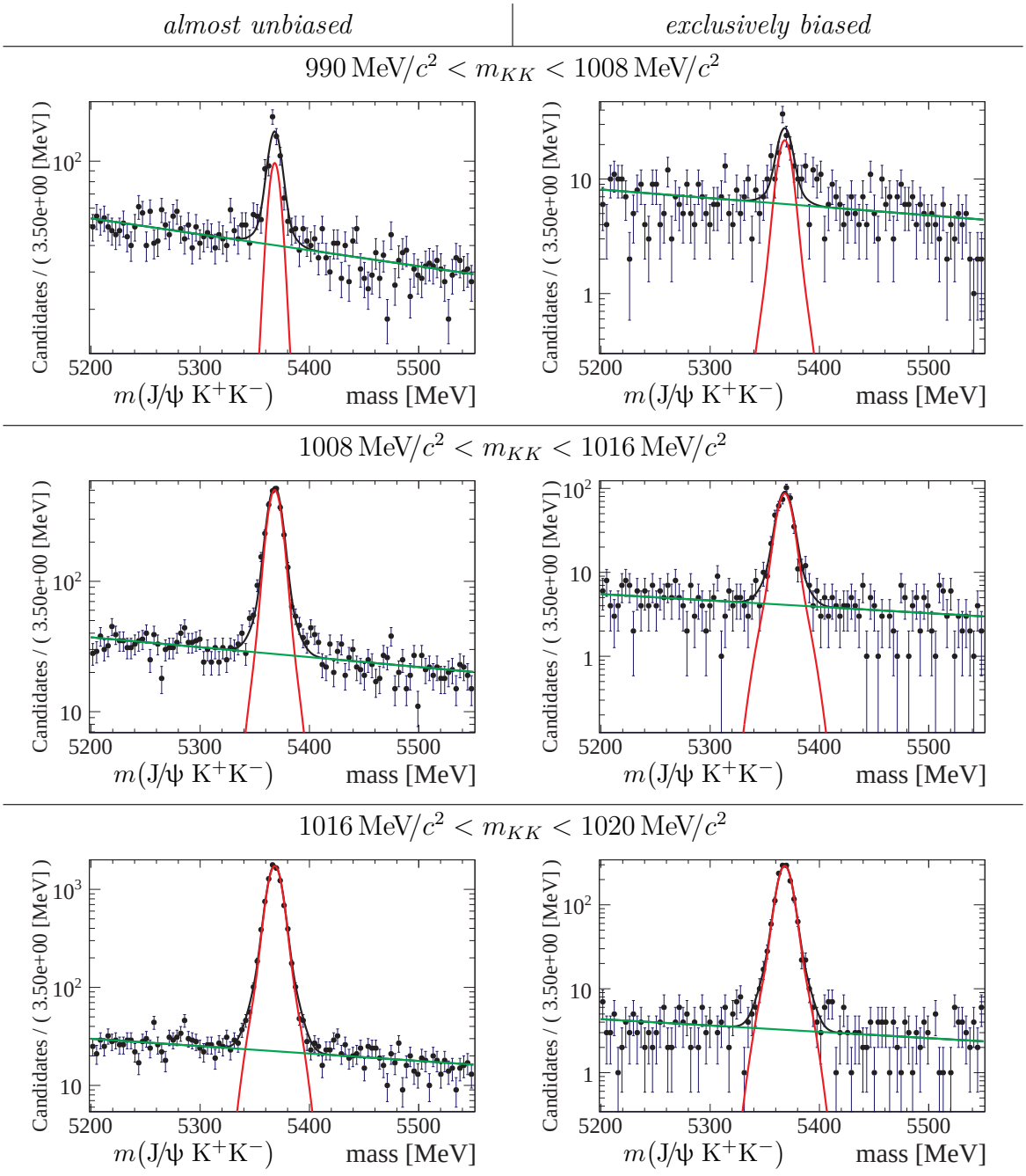


Figure C.1: *Mass Distributions for the lower 3 bins in $m(\mathrm{K^+K^-})$, bins 1, 2 and 3.*

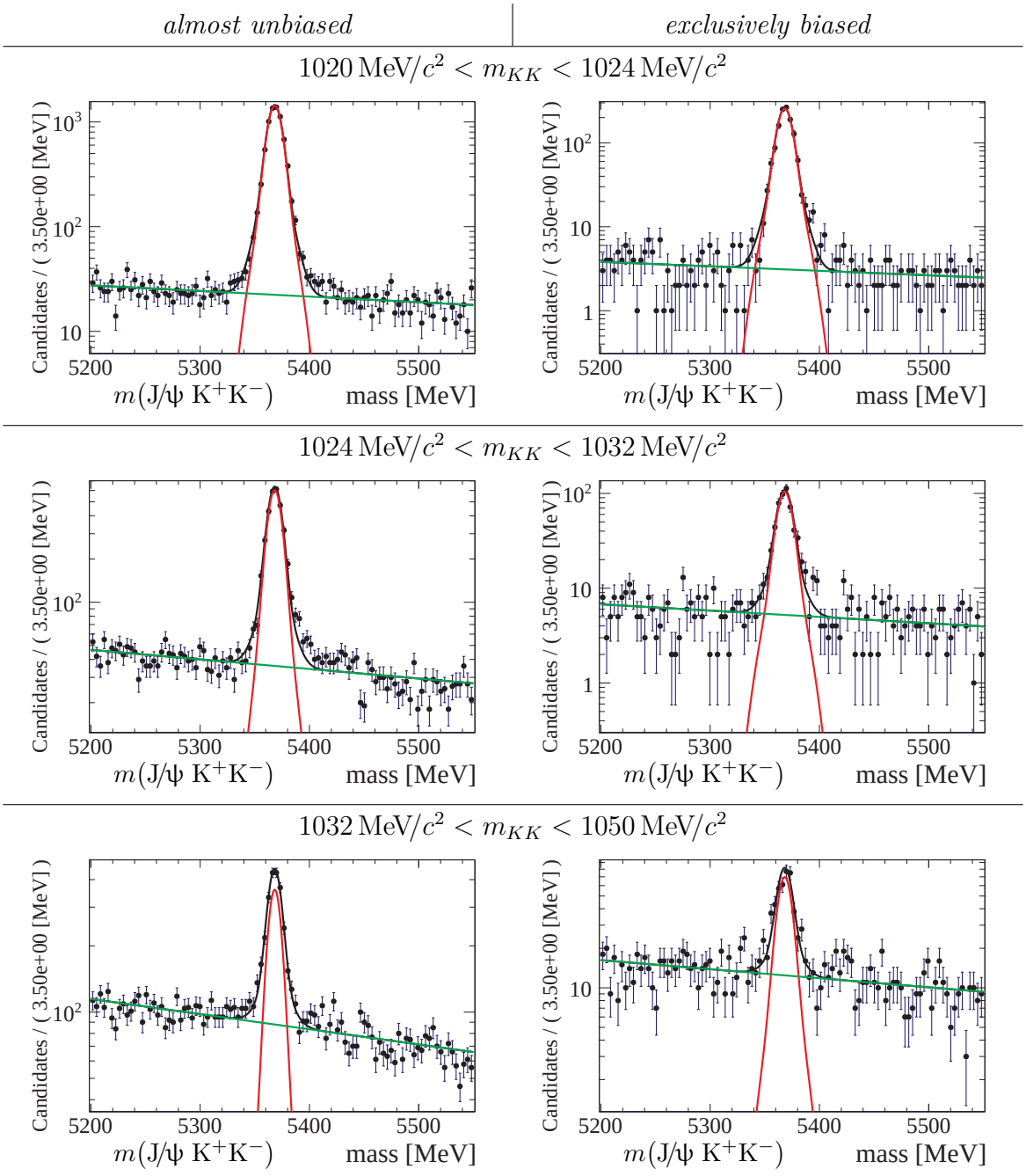


Figure C.2: *Mass Distributions for the upper 3 bins in $m(\mathrm{K^+K^-})$, bins 4, 5 and 6.*

D

Measured Likelihood Profiles

This appendix summarises the likelihood profiles which have been extracted for the physics parameters in the full $B_s^0 \to J/\psi\, K^+K^-$ analysis.

The process for extracting the likelihood profiles was to fix the value of a chosen parameter and to re-minimise the likelihood function to determine the best minima. The fit was always started close to the global minima to reduce the computational requirements and to help keep the nuisance parameters in each scan well behaved.

D.1 Likelihood profiles in nominal fit

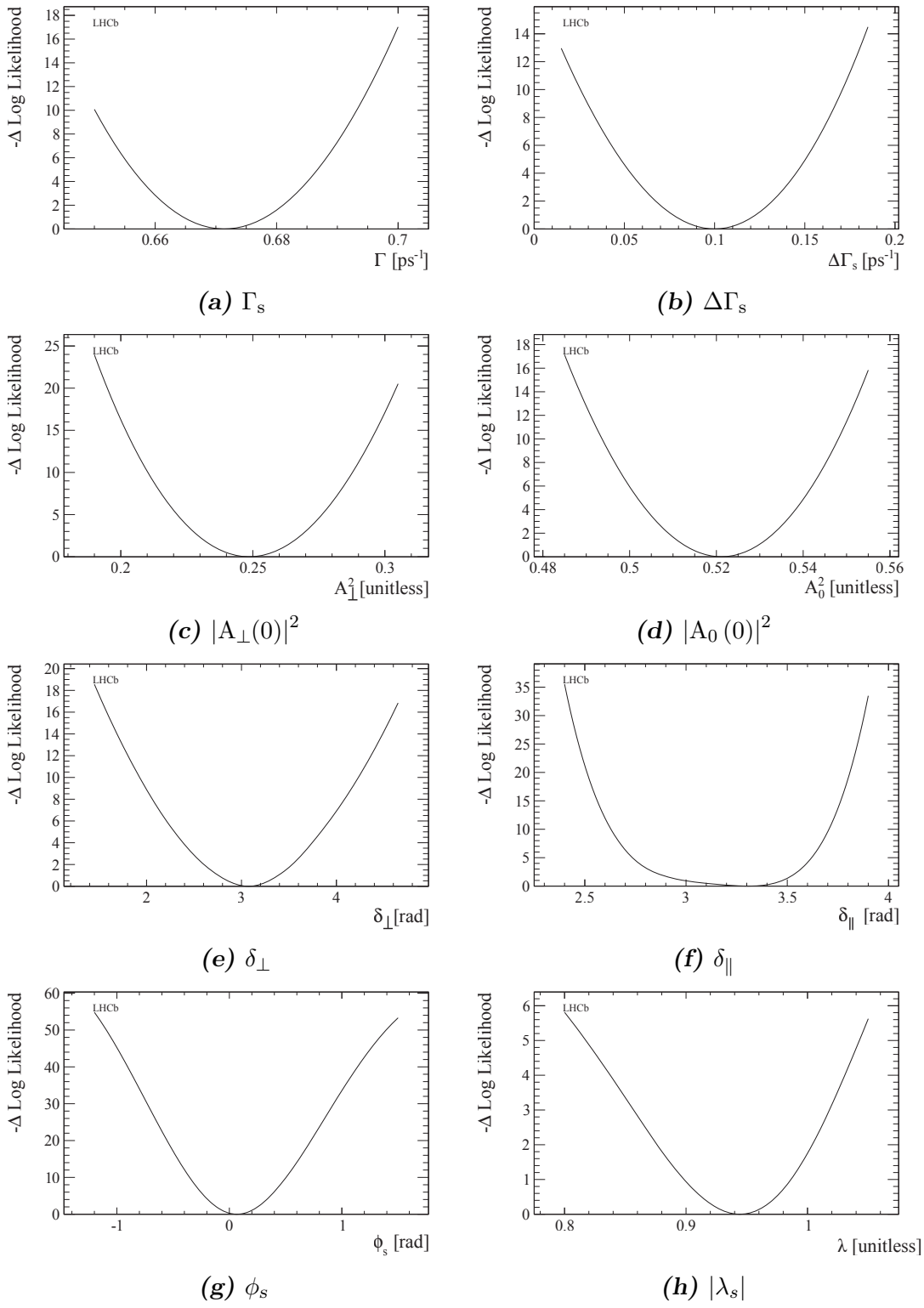
The likelihood profiles for the key physics parameters are given in Figure D.1. The likelihood profiles for all of the key physics parameters apart from $\delta_\parallel$ are well behaved. In the case of $\delta_\parallel$ this is due to the parameter's minima being close to its second degenerate minima and so the function isn't parabolic. This allows for all parameters bar $\delta_\parallel$ to have symmetric errors quoted about their global minima.

The likelihood profiles for the S-wave parameters in the six bins of $m(K^+K^-)$ are shown in Figure D.2. Unlike the key physics parameters none of the S-wave parameters have parabolic likelihood profiles. This is related to the fact that the S-wave fractions F_S are close to the physical limit of 0.

This discontinuity in the likelihood function is also responsible for the structure of the $\delta_{S-\perp}$ likelihood function in each bin. Because none of these parameters have parabolic likelihood profiles the errors quoted are asymmetric and correspond to the limits where the likelihood function rises by 0.5.

D.2 Likelihood profiles for Δm_s

In the nominal fit to $B_s^0 \to J/\psi\, K^+K^-$ the parameter Δm_s is constrained using an external measurement of this parameter Ref [65], an additional fit is performed with this parameter free. The likelihood function in Figure D.3 (a) shows that the fit is able to fit to a single well defined value for Δm_s over a wide range. The function shown Figure D.3 (b).



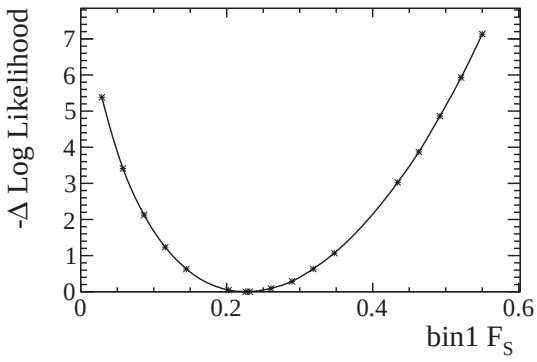
(a) Γ_s ***(b)*** $\Delta\Gamma_s$

(c) $|A_\perp(0)|^2$ ***(d)*** $|A_0(0)|^2$

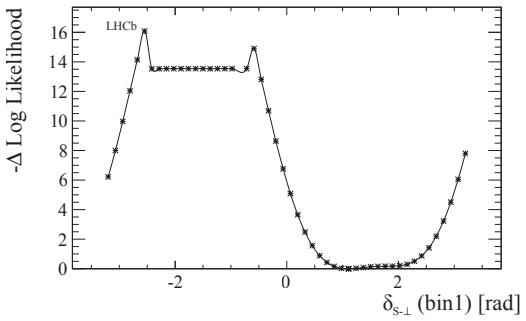
(e) $\delta_\perp$ ***(f)*** $\delta_\parallel$

(g) ϕ_s ***(h)*** $|\lambda_s|$

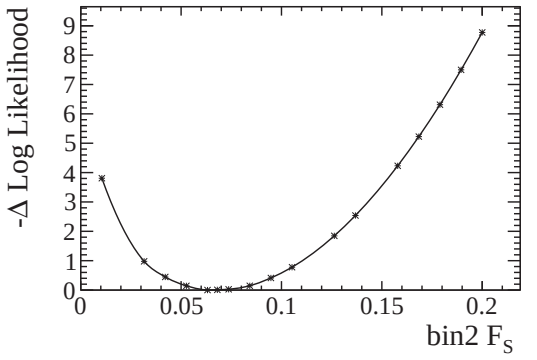
Figure D.1: *Likelihood profiles for key physics parameters*



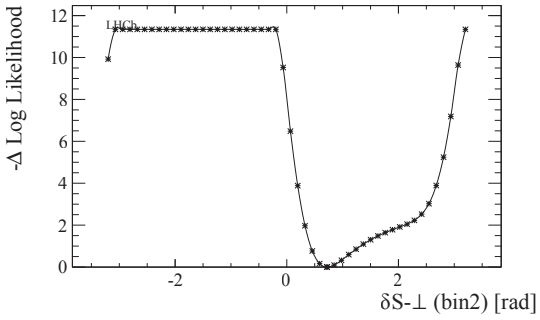
(a) *bin 1* F_S



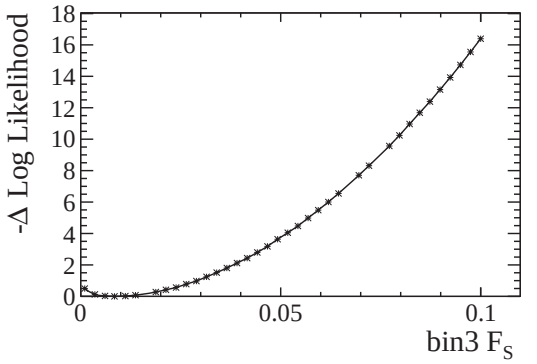
(b) *bin 1* $\delta_{S-\perp}$



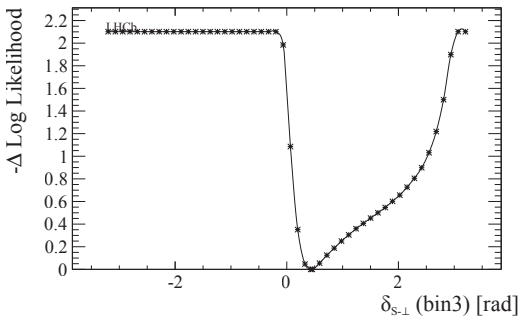
(c) *bin 2* F_S



(d) *bin 2* $\delta_{S-\perp}$

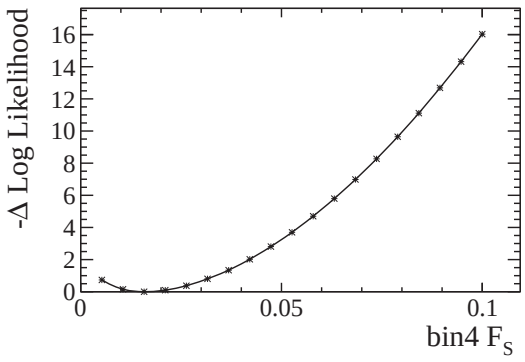


(e) *bin 3* F_S

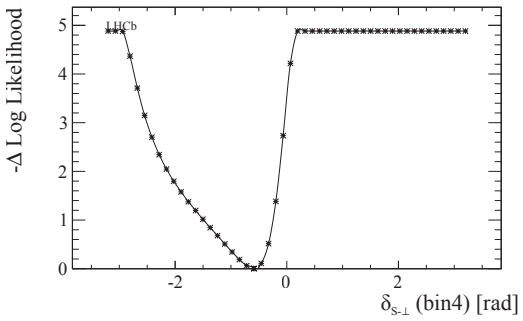


(f) *bin 3* $\delta_{S-\perp}$

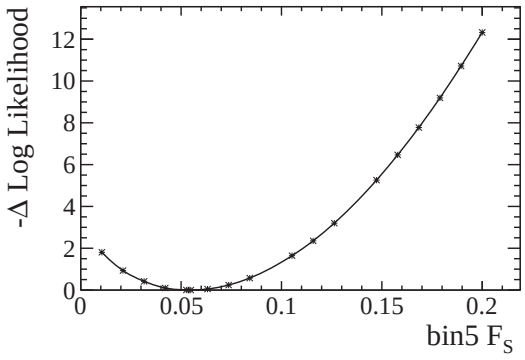
Figure D.2: *Likelihood profiles S-wave parameters*



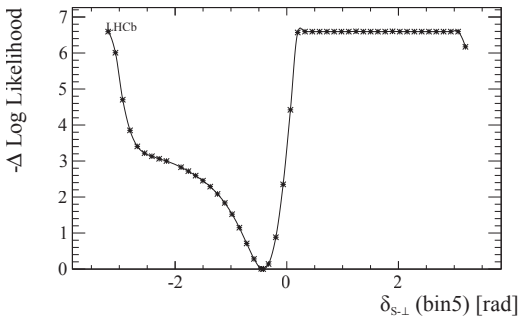
(g) *bin 4* F_S



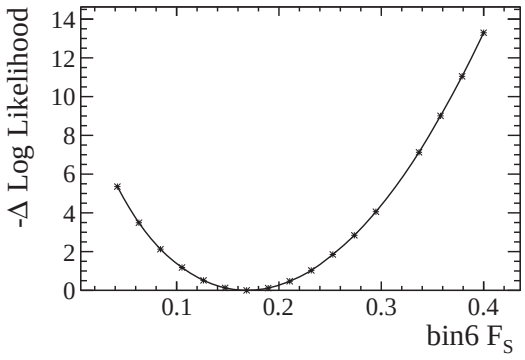
(h) *bin 4* $\delta_{S-\perp}$



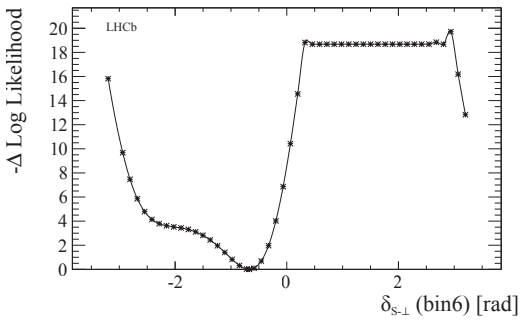
(i) *bin 5* F_S



(j) *bin 5* $\delta_{S-\perp}$

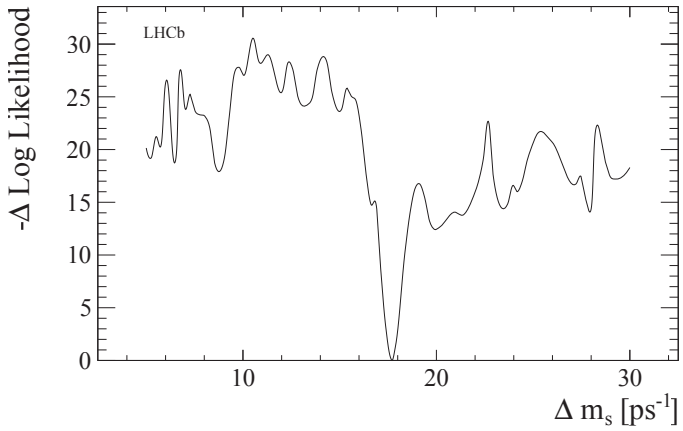


(k) *bin 6* F_S

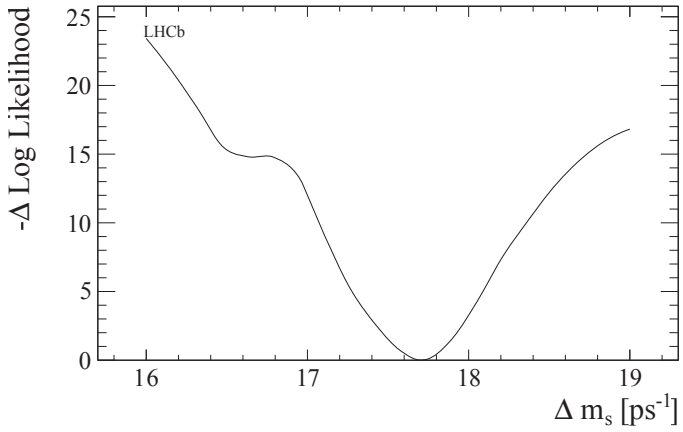


(l) *bin 6* $\delta_{S-\perp}$

Figure D.2: *Likelihood profiles S-wave parameters*



(a) Δm_s



(b) Δm_s

Figure D.3: *Δm_s likelihood profiles*

E

RapidFit Example Analysis

This appendix describes a minimal analysis within the RapidFit framework making use of a simple Gaussian function width a fixed central value and a variable width, which is analytically integrable.

Figure E.1 shows the source code for the user provided PDF along with the simplest XML required to actually perform a fit to a toy dataset using this PDF.

(a)

```
<RapidFit>

  <FitFunction>NegativeLogLikelihood</FitFunction>

  <Minimiser>Minuit</Minimiser>

  <ToFit>
    <PDF>
      <Name>Gaussian</Name>
    </PDF>
    <DataSet>
      <Source>Foam</Source>
      <NumberEvents>1000</NumberEvents>
      <PhaseSpaceBoundary>
        <Obsevable>
          <Name>x</Name>
          <Minimum>-5.</Minimum>
          <Maximum>5.</Maximum>
          <Unit>someUnit</Unit>
        </Observable>
      </PhaseSpaceBoundary>
    </DataSet>
  </ToFit>

  <ParameterSet>
    <PhysicsParameter>
      <Name>Sigma</Name>
      <Value>1.</Value>
      <Minimum>0.</Minimum>
      <Maximum>2.</Maximum>
      <Unit>unitless</Unit>
    </PhysicsParameter>
  </ParameterSet>

</RapidFit>
```

(b)

Gaussian.h:

```
#include "BasePDF.h"

class Gaussian : public BasePDF
{
  public:
    Gaussian();
    double Evaluate( DataPoint* );
    double Normalise( DataPoint*, PhaseSpaceBoundary* );
}
```

Gaussian.cpp:

```
#include "Gaussian.h"
#include <cmath>
#include <vector>

PDF_CREATOR( Gaussian );

Gaussian::Gaussian( PDFConfigurator* thisConfig ) : BasePDF()
{
  allObservables = std::vector<string>(1, "x");
  allParameters = ParameterSet( std::vector<string>(1, "Sigma");
}


double Gaussian::Evaluate( DataPoint* thisPoint )
{
  double thisX = thisPoint->GetObservable( "x" )->GetValue();
  double SigmaValue = allParameters.GetPhysicsParameter( "Sigma" )->GetValue();
  double exponent = - (thisX*thisX)/(2.*SigmaValue*SigmaValue);
  return std::exp( exponent );
}


double Gaussian::Normalise( DataPoint* thisPoint, PhaseSpaceBoundary* thisPhaseSpace )
{
  double pi=3.14159;
  double SigmaValue = allParameters.GetPhysicsParameter( "Sigma" )->GetValue();
  return SigmaValue * std::sqrt(2.*pi);
}
```

Figure E.1: *For demonstration purposes, (a) A user written XML file which fits an Amplitude of a Gaussian function to a toy Dataset. (b) A user written minimal Gaussian PDF in C++ which will be built by RapidFit.*

F

Pull Bias Measurements

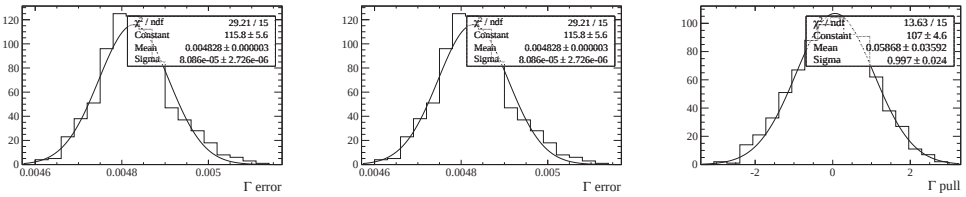
This appendix summarises the work done to extract the sensitivity on the various parameters in the $B_s^0 \to J/\psi\, K^+K^-$ analysis.

The expected sensitivity to each parameter of interest was determined using toy Monte Carlo events, generated with central values equal to the fit results measured with the data. The results are given in Table F.1. 24500 events were generated for each toy dataset.

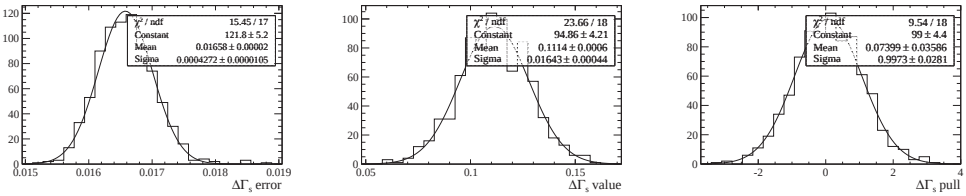
The distributions of central value, error and pull from each toy fit are shown in the subsequent histograms.

	Paramater	Sensitivity	Pull Bias
	Γ_s $[\mathrm{ps}^{-1}]$	0.00483	0.0587 $\pm$0.0359
	$\Delta\Gamma_s$ $[\mathrm{ps}^{-1}]$	0.0166	0.074 $\pm$0.0359
	$\lvert A_\perp(0)\rvert^2$	0.00888	-0.121 $\pm$0.035
	$\lvert A_0(0)\rvert^2$	0.0062	0.000116 $\pm$0.0365
	$\delta_\parallel$ [rad]	0.156	-0.155 $\pm$0.0471
	$\delta_\perp$ [rad]	0.226	-0.0858 $\pm$0.035
	Δm_s $[\mathrm{ps}^{-1}]$	0.0765	0.0431 $\pm$0.0294
	ϕ_s [rad]	0.0885	0.0313 $\pm$0.0342
	$\lvert\lambda_s\rvert^2$	0.0301	-0.0333 $\pm$0.0398
bin 1	F_S	0.0666	0.112 $\pm$0.0367
	$\delta_{S-\perp}$ [rad]	0.389	0.0718 $\pm$0.0459
bin 2	F_S	0.0279	0.00435 $\pm$0.0343
	$\delta_{S-\perp}$ [rad]	0.257	-0.152 $\pm$0.0444
bin 3	F_S	0.0119	0.205 $\pm$0.0326
	$\delta_{S-\perp}$ [rad]	0	-0.229 $\pm$0.0457
bin 4	F_S	0.0133	0.142 $\pm$0.0362
	$\delta_{S-\perp}$ [rad]	0.316	0.197 $\pm$0.0419
bin 5	F_S	0.025	-0.0262 $\pm$0.0357
	$\delta_{S-\perp}$ [rad]	0.24	0.107 $\pm$0.0421
bin 6	F_S	0.0352	0.0167 $\pm$0.038
	$\delta_{S-\perp}$ [rad]	0.168	0.0781 $\pm$0.0369

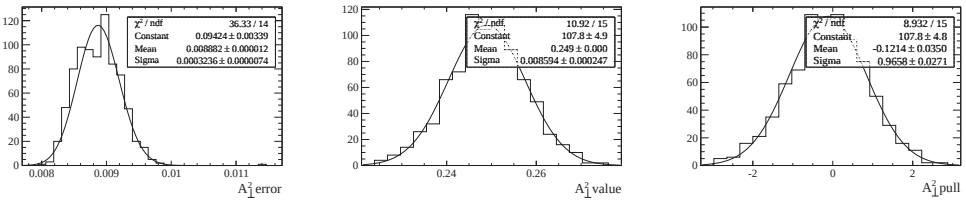
Table F.1: *Table showing the results of toy studies using 1000 toys of equivalent size to the unbiased+biased data sets. The first column shows the sensitivity to each parameter. The second column shows the pull bias (bias divided buy statistical error) derived from the set of histograms below*



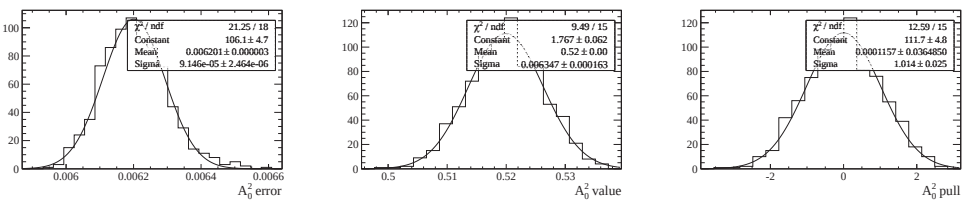
(a) Γ_s $[\mathrm{ps}^{-1}]$



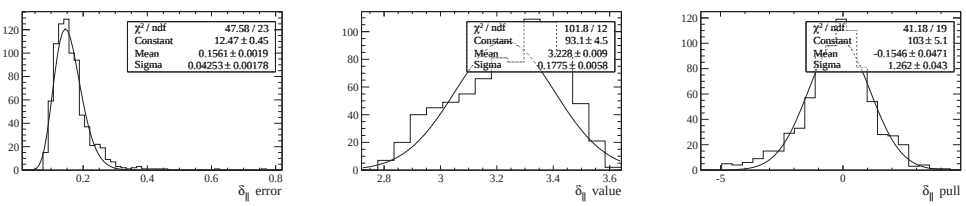
(b) $\Delta\Gamma_s$ $[\mathrm{ps}^{-1}]$



(c) $|\mathrm{A}_\perp(0)|^2$



(d) $|\mathrm{A}_0\,(0)|^2$



(e) $\delta_\parallel$ $[\mathrm{rad}]$

Figure F.1: *Toy study distributions. Left=central fit value, Centre=error, Right=pull.*

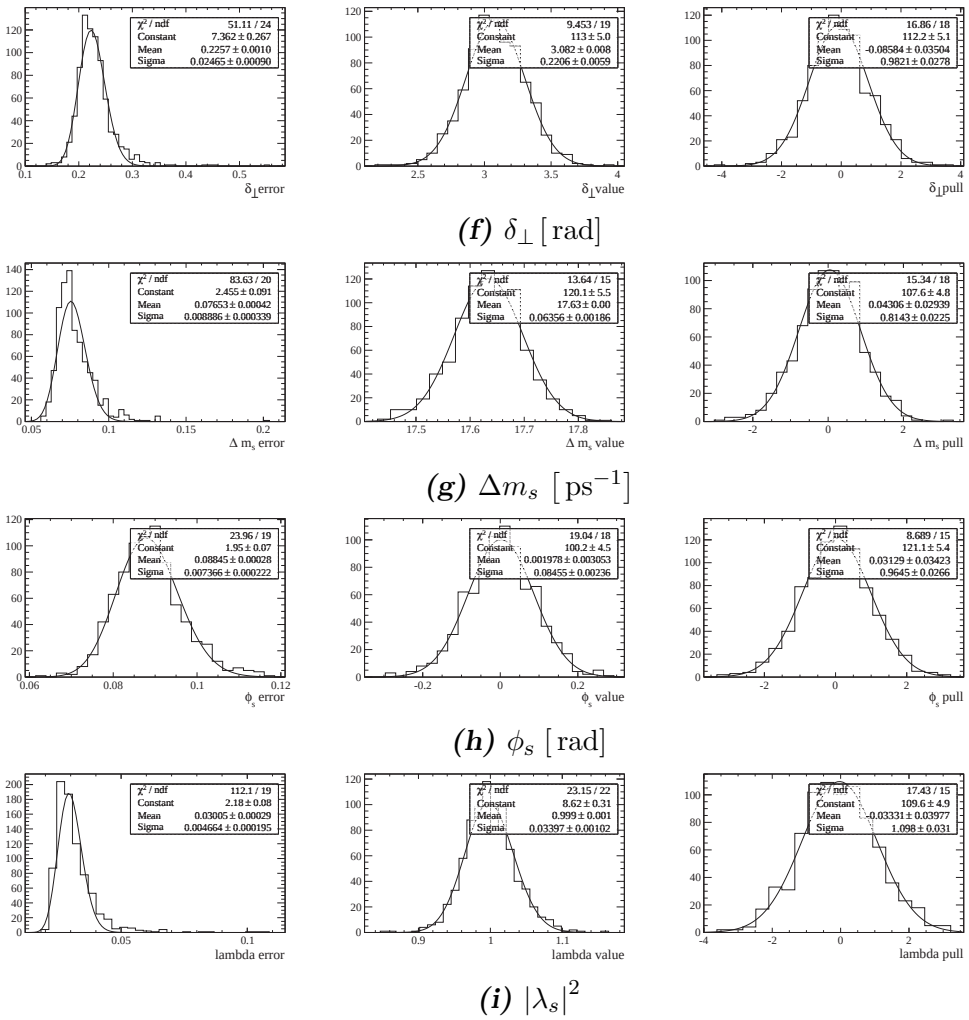
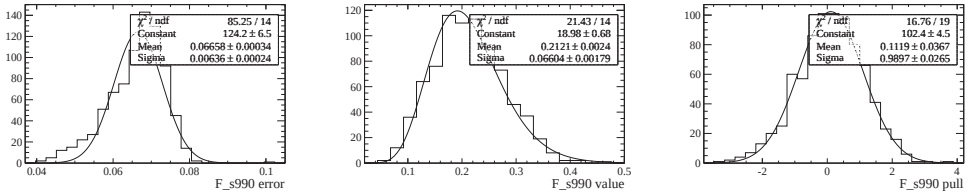
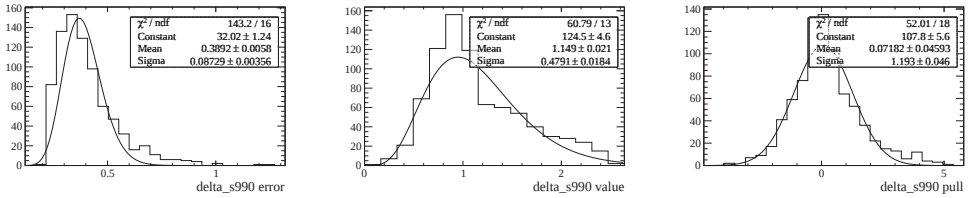


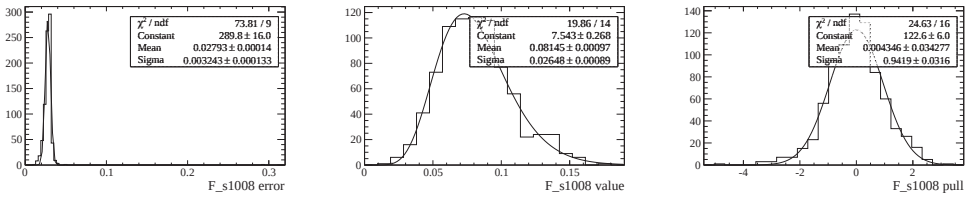
Figure F.1: *Toy study distributions. Left=central fit value, Centre=error, Right=pull.*



(j) *bin 1* F_{S}

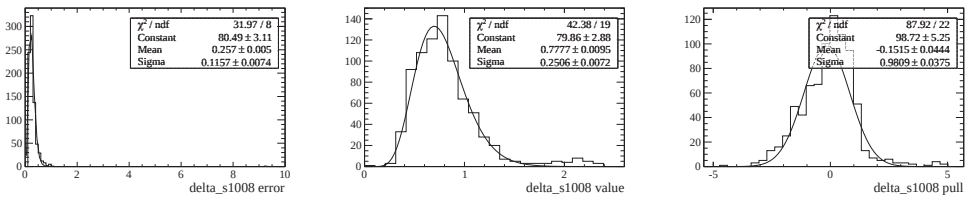


(k) *bin 1* $\delta_{\mathrm{S}-\perp}\,[\,\mathrm{rad}]$

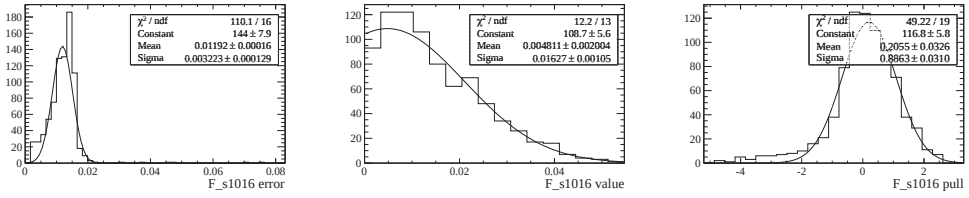


(l) *bin 2* F_{S}

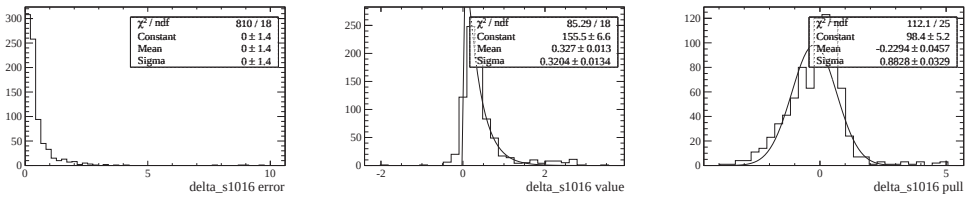
Figure F.1: *Toy study distributions. Left=central fit value, Centre=error, Right=pull.*



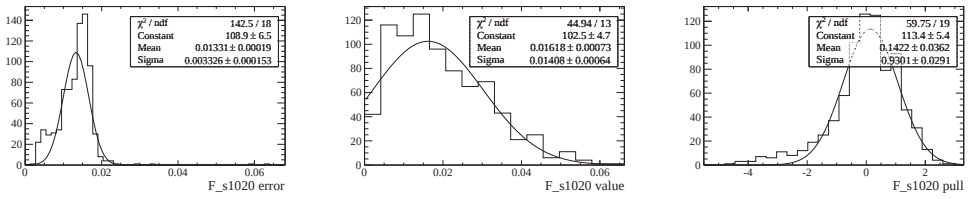
(m) *bin 2* $\delta_{S-\perp}$ [rad]



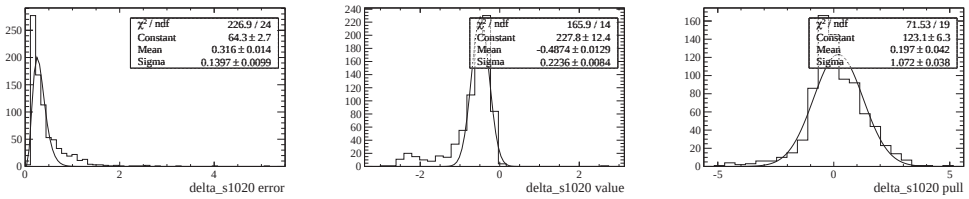
(n) *bin 3* F_S



(o) *bin 3* $\delta_{S-\perp}$ [rad]

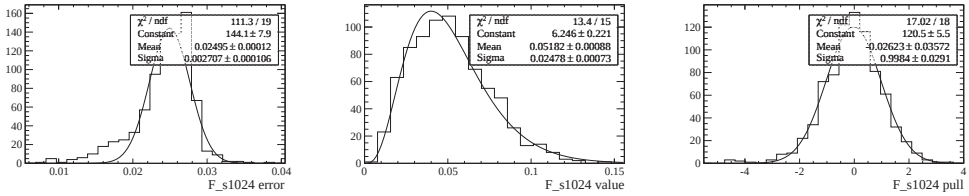


(p) *bin 4* F_S

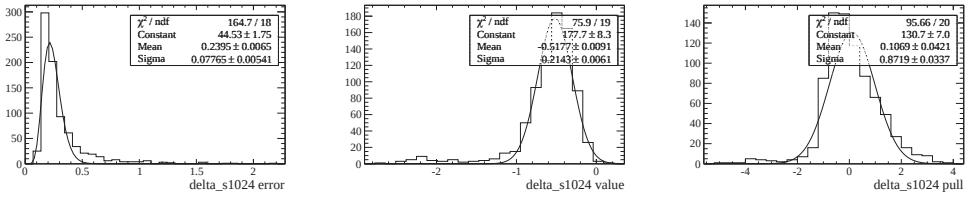


(q) *bin 4* $\delta_{S-\perp}$ [rad]

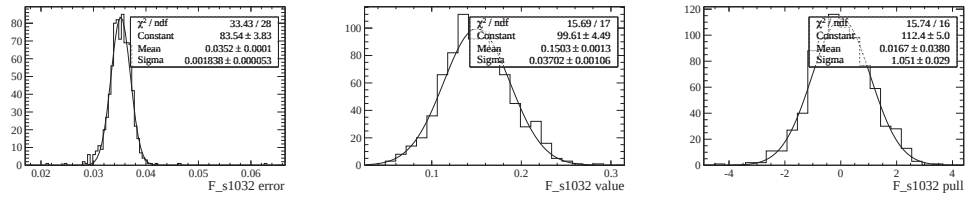
Figure F.1: *Toy study distributions. Left=central fit value, Centre=error, Right=pull.*



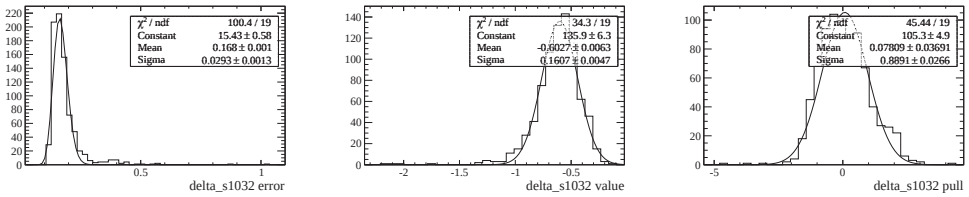
(r) *bin 5* F_{S}



(s) *bin 5* $\delta_{\mathrm{S}-\perp}\,[\mathrm{rad}]$



(t) *bin 6* F_{S}



(u) *bin 6* $\delta_{\mathrm{S}-\perp}\,[\mathrm{rad}]$

Figure F.1: *Toy study distributions. Left=central fit value, Centre=error, Right=pull.*

G

Supporting Decay Time Studies Material

This appendix describes supporting material for the measurements of Γ_s and $\Delta\Gamma_s$ from the angular moment analysis. In particular the generation of MC datasets to determine the behaviour of the fits to Γ_H and Γ_L are described in addition to the extraction of the correlation coefficient $\rho(\Gamma_H, \Gamma_L)$.

G.1 Estimated sensitivities

In order to replicate the full angular moment analysis, signal MC datasets are generated using the full $B_s^0 \to J/\psi\, K^+K^-$ PDF with the generation values in Table G.1. These MC datasets were each constructed with 27,600 signal events. A series of MC fits were performed to each dataset with angular weights calculated for each mass eigenstate. These used a single exponential function to extract the lifetime of the mass eigenstate for each dataset. Figure G.1 shows the distribution of measured values and errors for $\tau_H = 1/\Gamma_H$ and $\tau_L = 1/\Gamma_L$ from fits to 10,000 MC datasets. The distributions are well defined and centred around the input values.

Parameter	Generation Value
Γ_s *	0.6715 $[\mathrm{ps}^{-1}]$
$\Delta\Gamma_s$	0.100 $[\mathrm{ps}^{-1}]$
$\lvert A_0(0)\rvert^2$	0.521
$\lvert A_\perp(0)\rvert^2$	0.2485
F_S	0.0
δ_0	0.0 [rad]
$\delta_\perp$	3.31 [rad]
$\delta_\parallel$	3.08 [rad]
δ_S	0.0 [rad]
ϕ_s	0.0 [rad]
$\lvert\lambda\rvert$	1.0

Table G.1: *Central values used for the generation of the effective decay time MC datasets using the* $B_s^0 \to J/\psi\, K^+K^-$ *PDF. *The effect of upper decay time acceptance has been ignored here.*

Generated Values		Measurements from MC	
Γ_H $[\mathrm{ps}^{-1}]$	Γ_L $[\mathrm{ps}^{-1}]$	Γ_H $[\mathrm{ps}^{-1}]$	Γ_L $[\mathrm{ps}^{-1}]$
0.622	0.722	0.628 ± 0.021	0.722 ± 0.011
τ_H [ps]	τ_L [ps]	τ_H [ps]	τ_L [ps]
1.609	1.386	1.609 ± 0.055	1.386 ± 0.021

Table G.2: *A comparison between the generated and observed decay time values of the mass eigenstates for 10,000 MC datasets.*

Table G.2 shows a numerical comparison between the central values extracted from the MC distributions and the expected central values. The result of this shows that the extracted decay times are centred around the generated value with errors comparable to the measured times from data in Table 9.4.

The pull distribution for τ_L , Figure G.1 (c), shows this parameter is measured in an unbiased way and that the per-dataset uncertainties are correctly estimated. The pull distribution for τ_H , Figure G.1 (f), shows that although this parameter is unbiased the per-dataset uncertainty is systematically overestimated leading to the width of the pull distribution being 0.9.

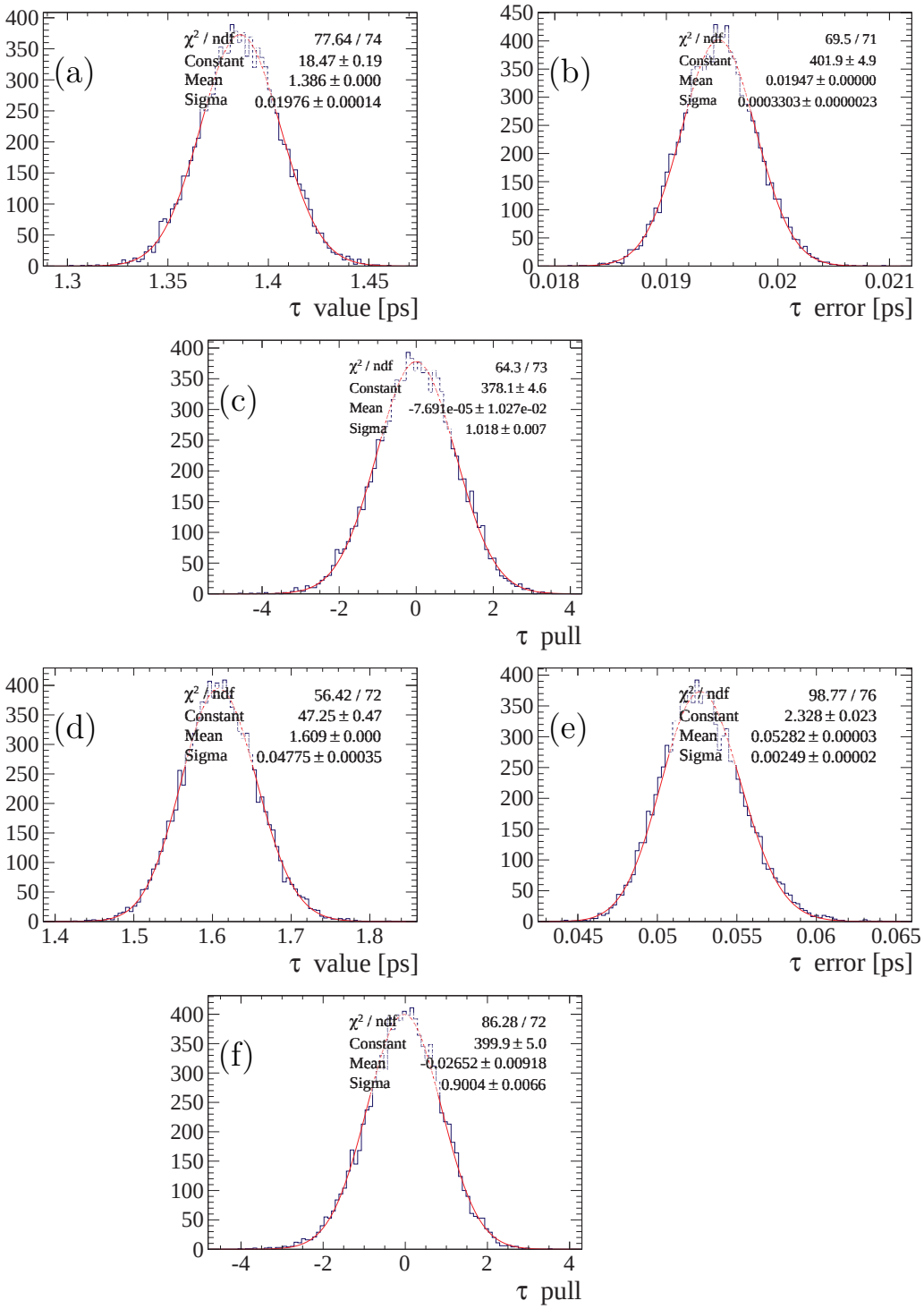


Figure G.1: *Distribution of: (a) τ_L value, (b) τ_L error, (c) τ_L pull, (d) τ_H value, (e) τ_H error and (f) τ_H pull. These results correspond to values extracted from 10,000 fast Monte Carlo datasets. The MC datasets were generated with 27,600 signal events using the values in Table G.1.*

$B_s^0 \to J/\psi\, K^+K^-$ Analysis:

Γ_s $[\,ps^{-1}]$	$\Delta\Gamma_s$ $[\,ps^{-1}]$
$\Gamma_s{}^* = 0.6715 \pm 0.0048$	$\Delta\Gamma_s = 0.100 \pm 0.016$

Angular Moment Analysis:

Γ_s $[\,ps^{-1}]$	$\Delta\Gamma_s$ $[\,ps^{-1}]$
$\Gamma_s{}^* = 0.6744 \pm 0.0050$	$\Delta\Gamma_s = 0.094 \pm 0.025$

Table G.3: *This is a comparison between the results from the full analysis and the measurements of Γ_s and $\Delta\Gamma_s$ extracted from angular moment fits. The errors quoted here are statistical only. (* Here the measured decay width has not been corrected for the effect of upper lifetime acceptance.)*

G.2 Γ_{L} and Γ_{H} correlation

Both $(\Gamma_{\mathrm{L}}, \Gamma_{\mathrm{H}})$ and $(\Gamma_{\mathrm{s}}, \Delta\Gamma_{\mathrm{s}})$ are pairs of parameters which are strongly correlated. In order to propagate the errors from these measurements correctly this correlation between the values must be included. In order to correctly propagate the errors on Γ_{H} and Γ_{L} the value of $\rho(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}})$ is determined from MC simulations.

Using the distribution of fit results from the previous section allows the correlation between Γ_{H} and Γ_{L} $(\rho(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}}))$ to be extracted. This can be calculated though fitting a bivariate normal function to the distribution of Γ_{H} vs Γ_{L}. This function is defined as,

$$f(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}}) = \frac{1}{2\pi\sigma(\Gamma_{\mathrm{H}})\sigma(\Gamma_{\mathrm{L}})\sqrt{1-\rho(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}})^2}} \times \exp\left\{ -\frac{1}{2\left(1-\rho(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}})^2\right)} \left[\begin{array}{c} \dfrac{(\Gamma_{\mathrm{H}} - \mu(\Gamma_{\mathrm{H}}))^2}{\sigma(\Gamma_{\mathrm{H}})^2} + \dfrac{(\Gamma_{\mathrm{L}} - \mu(\Gamma_{\mathrm{L}}))^2}{\sigma(\Gamma_{\mathrm{L}})^2} \\ \\ -\dfrac{2\rho(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}})(\Gamma_{\mathrm{H}} - \mu(\Gamma_{\mathrm{H}}))(\Gamma_{\mathrm{L}} - \mu(\Gamma_{\mathrm{L}}))}{\sigma(\Gamma_{\mathrm{H}})\sigma(\Gamma_{\mathrm{L}})} \end{array} \right] \right\}. \tag{G.1}$$

An alternate approach to calculating the Γ_{H}, Γ_{L} correlation is to calculate the Pierson product-moment coefficient $\rho^{\mathrm{Pierson}}(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}})$,

$$\rho^{\mathrm{Pierson}}(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}}) = \frac{\sum_i^n (\Gamma_{\mathrm{H}i} - \mu(\Gamma_{\mathrm{H}}))(\Gamma_{\mathrm{L}} - \mu(\Gamma_{\mathrm{L}}))}{\sqrt{\sum_i^n (\Gamma_{\mathrm{H}i} - \mu(\Gamma_{\mathrm{H}}))^2}\sqrt{\sum_i^n (\Gamma_{\mathrm{L}} - \mu(\Gamma_{\mathrm{L}}))^2}}. \tag{G.2}$$

The resulting coefficient as calculated by the two methods is summarised in Table G.4. These values are in good agreement and have a small associated error allowing the correlation to be used to correctly propagate the errors on Γ_{H} and Γ_{L}.

In order to propagate the correlation between $(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}})$ to make a measurement of $(\Gamma_{\mathrm{s}}, \Delta\Gamma_{\mathrm{s}})$ equation G.3 is used to calculate the uncertainty associated on each event. The small error associated with the correlation allows this to be taken as a fixed constant.

Method	$\rho(\Gamma_{\mathrm{L}}, \Gamma_{\mathrm{H}})$ Results
Bivariate Normal Distribution	-0.7888 ± 0.0046
Pierson Product Moment	-0.792 ± 0.069

Table G.4: *Calculated $\rho(\Gamma_{\mathrm{L}}, \Gamma_{\mathrm{H}})$ values using different methods*

$$\sigma\left(\Gamma_{\mathrm{s}}\right)^2 = \left(\frac{1}{2}\right)^2 \sigma\left(\Gamma_{\mathrm{H}}\right)^2 + \left(\frac{1}{2}\right)^2 \sigma\left(\Gamma_{\mathrm{L}}\right)^2 + 2\rho\left(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}}\right)\left(\frac{1}{2}\right)^2 \sigma\left(\Gamma_{\mathrm{H}}\right)\sigma\left(\Gamma_{\mathrm{H}}\right) .$$
$$\sigma\left(\Delta\Gamma_{\mathrm{s}}\right)^2 = \sigma\left(\Gamma_{\mathrm{H}}\right)^2 + \sigma\left(\Gamma_{\mathrm{L}}\right)^2 - 2\rho\left(\Gamma_{\mathrm{H}}, \Gamma_{\mathrm{L}}\right)\sigma\left(\Gamma_{\mathrm{H}}\right)\sigma\left(\Gamma_{\mathrm{H}}\right) . \tag{G.3}$$

Bibliography

[1] A.D. Sakharov, "Violation of CP Invariance, c Asymmetry, and Baryon Asymmetry of the Universe," *Pisma Zh.Eksp.Teor.Fiz.*, vol. 5, pp. 32–35, 1967. DOI: {10.1070/PU1991v034n05ABEH002497} (cit. on p. 1).

[2] G. Hinshaw et al., "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," 2012. arXiv:1212.5226 (cit. on p. 2).

[3] N. Cabibbo, "Unitary Symmetry and Leptonic Decays," *Phys. Rev. Lett.*, vol. 10, 531–533, 12 1963. DOI: {10.1103/PhysRevLett.10.531}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.10.531 (cit. on pp. 2, 9).

[4] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics*, vol. 49, 652–657, 1973. DOI: {10.1143/PTP.49.652}. [Online]. Available: http://adsabs.harvard.edu/abs/1973PThPh..49..652K (cit. on pp. 2, 9).

[5] C. Jarlskog, "A basis independent formulation of the connection between quark mass matrices, CP violation and experiment," *Z Phys C Part Fields*, vol. 29, 491–497, 1985. DOI: {10.1007/BF01565198} (cit. on p. 2).

[6] G.W.S Hou, "Source of CP Violation for the Baryon Asymmetry of the Universe," 2008. arXiv:0810.3396 (cit. on p. 2).

[7] J. Beringer et al., "Review of Particle Physics," *Phys. Rev. D*, vol. 86, 010 001, 1 2012. DOI: {10.1103/PhysRevD.86.010001}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.86.010001 (cit. on pp. 2, 3, 5, 6, 48, 74, 83, 84, 95–97).

[8] M. E. Peskin, "Particle physics: Song of the electroweak penguin," *Nature*, vol. 452, 293–294, 2008. DOI: {10.1038/452293a} (cit. on p. 2).

[9] T. Aaltonen et al., "Measurement of the CP-violating phase $\beta_s^{J/\psi\phi}$ in $B_s^0 \to J/\psi\phi$ decays with the CDF II detector," *Phys. Rev. D*, vol. 85, 072 002, 7 2012. DOI: {10.1103/PhysRevD.85.072002}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.85.072002 (cit. on p. 3).

[10] V. M. Abazov et al., "Measurement of the CP-violating phase $\phi_s^{J/\psi\phi}$ using the flavor-tagged decay $B_s^0 \to J/\psi\phi$ in 8fb^{-1} of $p\overline{p}$ collisions," *Phys. Rev. D*, vol. 85, 032 006, 3 2012. DOI: {10.1103/PhysRevD.85.032006}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.85.032006 (cit. on pp. 3, 154).

[11] R. Aaij et al., "Measurement of the CP-violating phase ϕ_s in the decay $B_s^0 \to J/\psi\phi$," *Phys.Rev.Lett.*, vol. 108, p. 101 803, 2012. DOI: {10.1103/PhysRevLett.108.101803}. arXiv:1112.3183 (cit. on pp. 3, 120).

[12] R. Aaij at al., "Measurement of CP violation and the B_s^0 meson decay width difference with $B_s^0 \to J/\psi K^+K^-$ and $B_s^0 \to J/\psi\pi^+\pi^-$ decays," *Phys. Rev. D*, vol. D87, p. 112 010, 2013. DOI: {10.1103/PhysRevD.87.112010}. arXiv:1304.2600 (cit. on pp. 3, 21, 47, 48, 56, 59, 62, 64, 71, 72, 96, 123, 137, 139, 140, 154).

[13] E. Noether, "Invariant variation problems," *Transp. Theory Stat. Phys.*, vol. 1, 186–207, 1971. DOI: {10.1080/00411457108231446}. arXiv:0503066 (cit. on p. 7).

[14] J. Bell, *Birmingham University thesis*, 1954 (cit. on p. 7).

[15] W. Pauli, "Niels Bohr and the Development of Physics: Essays Dedicated to Niels Bohr on the Occasion of His Seventieth Birthday," 1955 (cit. on p. 7).

[16] C. S. Wu et al., "Experimental Test of Parity Conservation in Beta Decay," *Phys. Rev.*, vol. 105, 1413–1415, 4 1957. DOI: {10.1103/PhysRev.105.1413}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRev.105.1413 (cit. on p. 8).

[17] J. H. Christenson et al., "Evidence for the 2π Decay of the K_2^0 Meson," *Phys. Rev. Lett.*, vol. 13, 138–140, 4 1964. DOI: {10.1103/PhysRevLett.13.138}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.13.138 (cit. on p. 8).

[18] L. Wolfenstein, "Parametrization of the Kobayashi-Maskawa Matrix," *Phys. Rev. Lett.*, vol. 51, 1945–1947, 21 1983. DOI: {10.1103/PhysRevLett.51.1945}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.51.1945 (cit. on p. 10).

[19] J. Charles, et al., "Predictions of selected flavour observables within the Standard Model," *Phys. Rev. D*, vol. D84, p. 033 005, 2011. DOI: {10.1103/PhysRevD.84.033005}. arXiv:1106.4041 (cit. on pp. 23, 136, 137, 155).

[20] A. Badin et al., "Lifetime difference in B_s mixing: Standard model and beyond," *Phys. Lett. B*, vol. B653, pp. 230–240, 2007. DOI: {10.1016/j.physletb.2007.07.049}. arXiv:0707.0294 (cit. on p. 24).

[21] A. Lenz and U. Nierste, "Theoretical update of $B_s - \bar{B}_s$ mixing," *JHEP*, vol. 0706, p. 072, 2007. DOI: {10.1088/1126-6708/2007/06/072}. arXiv:0612167 (cit. on pp. 24, 136, 137).

[22] A. Badin et al., "Lifetime difference in B_s mixing: Standard model and beyond," *Phys. Lett. B*, vol. B653, pp. 230–240, 2007. DOI: {10.1016/j.physletb.2007.07.049}. arXiv:0707.0294 (cit. on pp. 24, 136, 137).

[23] A. Lenz, and U. Nierste, "Numerical Updates of Lifetimes and Mixing Parameters of B Mesons," 2011. arXiv:1102.4274 (cit. on pp. 24, 136, 137).

[24] The LHC*b* Collaboration, "The LHC*b* experiment at the CERN LHC," *JINST 3 S08005*, 2008. DOI: {doi:10.1088/1748-0221/3/08/S08005} (cit. on pp. 25, 26, 30, 31, 33–35, 37, 39, 41).

[25] The ALICE Collaboration, "The ALICE experiment at the CERN LHC," *JINST 3 S08002*, 2008. DOI: {10.1088/1748-0221/3/08/S08002} (cit. on p. 26).

[26] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST 3 S08003*, 2008. DOI: {doi:10.1088/1748-0221/3/08/S08003} (cit. on p. 26).

[27] The CMS Collaboration, "The CMS experiment at the CERN LHC," *JINST 3 S08004*, 2008. DOI: {doi:10.1088/1748-0221/3/08/S08004} (cit. on p. 26).

[28] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN," 2008. [Online]. Available: {http://cds.cern.ch/record/1260465?ln=en} (cit. on p. 27).

[29] T. Sjostrand et al., "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 0605, p. 026, 2006. DOI: {10.1088/1126-6708/2006/05/026}. arXiv:0603175 (cit. on pp. 27, 28, 43, 68).

[30] E. Christian, "$b\bar{b}$ production angle plots," [Online]. Available: http://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html (cit. on p. 27).

[31] R. Aaij et al., "Measurement of at in the forward region," *Phys. Lett. B*, vol. 694, no. 3, pp. 209 –216, 2010, ISSN: 0370-2693. DOI: {10.1016/j.physletb.2010.10.010}. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0370269310012074 (cit. on p. 28).

[32] R. Lindner, "LHCb layout 2. LHCb schema 2," 2008, LHCb Collection. [Online]. Available: http://cds.cern.ch/record/5701 (cit. on p. 29).

[33] E. P. Olloqui, "LHCb preshower(PS) and scintillating pad detector (SPD): Commissioning, calibration, and monitoring," *J. Phys. Conf. Ser.*, vol. 160, p. 012 046, 2009. DOI: {10.1088/1742-6596/160/1/012046} (cit. on p. 37).

[34] C. Fitzpatrick, "A measurement of the CP-violating phase ϕ_s in the decay $B_s^0 \to J/\psi\,\phi$," *Edinburgh, University of Edinburgh thesis*, 2012 (cit. on pp. 38, 66).

[35] S. J. Alvsvaag et al., "The small angle tile calorimeter in the DELPHI experiment," *Nucl. Instrum. Methods Phys. Res., A*, vol. 425, 106–39. 24 p, 1998 (cit. on p. 39).

[36] A. A. Alves Jr et al., *J. Instrum.*, vol. 8, no. LHCb-DP-2012-002, P02022. 29 p, 2012. arXiv:1211.1346 (cit. on p. 41).

[37] R. Aaij et al., "The LHCb Trigger and its Performance," 2012. arXiv:1211.3055 (cit. on pp. 41, 42, 47).

[38] LHCb Speakers Bureau, "LHCb Trigger Scheme," [Online]. Available: http://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html (cit. on p. 43).

[39] M. Clemencic et al., "Recent developments in the LHCb software framework Gaudi," *J.Phys.Conf.Ser.*, vol. 219, p. 042 006, 2010. DOI: 10.1088/1742-6596/219/4/042006 (cit. on p. 42).

[40] R. Brun and F. Rademakers, "ROOT — An object oriented data analysis framework," *Nucl. Instrum. Meth. A*, vol. 389, no. 1–2, pp. 81 –86, 1997, ISSN: 0168-9002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016890029700048X (cit. on pp. 42, 44, 92, 104).

[41] I. Antcheva et al., "ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization," *Comput. Phys. Commun.*, vol. 180, no. 12, pp. 2499 –2512, 2009, 40 YEARS OF CPC: A celebratory issue focused on quality software for high performance, grid and novel computing architectures, ISSN: 0010-4655. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010465509002550 (cit. on pp. 42, 44, 92, 104).

[42] G. Corti et al., "Software for the LHCb experiment," *IEEE Trans. Nucl. Sci.*, vol. 53, pp. 1323–1328, 2006 (cit. on pp. 42–44).

[43] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth. A*, vol. 462, no. 1–2, pp. 152 –155, 2001, ISSN: 0168-9002. DOI: {10.1016/S0168-9002(01)00089-4}. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900201000894 (cit. on p. 43).

[44] P. Golonka and Z. Was, "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays," *Eur.Phys.J.*, vol. C45, pp. 97–107, 2006. DOI: {10.1140/epjc/s2005-02396-4}. arXiv:0506026 (cit. on p. 43).

[45] S. Agostinelli et al., "Geant4—a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, no. 3, pp. 250 –303, 2003, ISSN: 0168-9002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900203013688 (cit. on p. 43).

[46] B. Wynne, R. Currie, P. Clarke and G. Cowan, "RapidFit Fitting Framework," Developed internally within Edinburgh PPE group. [Online]. Available: {svn.cern.ch/reps/lhcbedinburgh/Code/Fitting/RapidFit/trunk} (cit. on p. 44).

[47] P. Charpentier and the LHCb Collaboration, "The LHCb Computing Model and Real Data," *J. Phys.: Conf. Ser.*, vol. 331, no. 7, 072 008, 2011. [Online]. Available: http://stacks.iop.org/1742-6596/331/i=7/a=072008 (cit. on pp. 44, 116).

[48] L. Arrabito et al., "Major changes to the LHCb Grid computing model in year 2 of LHC data," *J. Phys.: Conf. Ser.*, vol. 396, 032092. 8 p, 2012 (cit. on pp. 44, 116).

[49] R. Aaij et al., "The LHCb Trigger and its Performance in 2011," *JINST*, vol. 8, P04022, 2013. DOI: {10.1088/1748-0221/8/04/P04022}. arXiv:1211.3055 (cit. on p. 49).

[50] M. Needham, "Clone Track Identification using the Kullback-Liebler Distance," no. LHCb-2008-002. LPHE-2008-002, 2008 (cit. on p. 50).

[51] W. D. Hulsbergen, "Decay chain fitting with a Kalman filter," *Nuclear Instruments and Methods in Physics Research A*, vol. 552, 566–575, 2005. DOI: {10.1016/j.nima.2005.06.078}. arXiv:0503191 (cit. on p. 52).

[52] M. Pivk and F. R. Le Diberder, "Plots: A statistical tool to unfold data distributions," *Nuclear Instruments and Methods in Physics Research A*, vol. 555, 356–369, 2005. DOI: {10.1016/j.nima.2005.08.106}. arXiv:0402083 (cit. on pp. 55, 70, 71, 73, 76, 86).

[53] T. Du Pree, "Search for a Strange Phase in Beautiful Oscillations," *Amsterdam, Vrije Universiteit*, 2010 (cit. on p. 60).

[54] Y. Amhis et al., "Understanding decay time acceptances in $B_s^0 \to J/\psi\,\phi$ Decays," 2013, LHCb-ANA-2013-008 (cit. on p. 64).

[55] C. Linn, "Measurement of the CP-violating phase ϕ_s using $B_s^0 \to J/\psi\phi$ and $B_s^0 \to J/\psi\pi^+\pi^-$ decays with the LHCb experiment," *Universität Heidelberg*, 2013, Thesis Presented on 15 May 2013 (cit. on pp. 65, 95, 97).

[56] R. Aaij et al., "Opposite-side flavour tagging of B mesons at the LHCb experiment," *Eur.Phys.J.*, vol. C72, p. 2022, 2012. DOI: {10.1140/epjc/s10052-012-2022-1}. arXiv:1202.4979 (cit. on pp. 65–67).

[57] ——, "Optimization and calibration of the same-side kaon tagging algorithm using hadronic B_s^0 decays in 2011 data," 2012, LHCb-CONF-2012-033 Linked to LHCb-ANA-2011-103 (cit. on pp. 65, 68).

[58] R. Aaij at al., "Observation of the decay $B_c^+ \to B_s^0\pi^+$," *Phys. Rev. Lett.*, 2013. DOI: {10.1103/PhysRevLett.111.181801}. arXiv:1308.4544 (cit. on pp. 83, 131).

[59] R. Aaij et. al, "Measurements of B_c^+ production and mass with the $B_c^+ \to J/\psi\pi^+$ decay," *Phys. Rev. Lett.*, vol. 109, p. 232 001, 2012. DOI: {10.1103/PhysRevLett.109.232001}. arXiv:1209.5634 (cit. on p. 84).

[60] C. Qiao et al., "B_c Exclusive Decays to Charmonia and Light Mesons in QCD Factorization at Next-to-Leading Order Accuracy," *Phys. Rev. D*, vol. D89, p. 034 008, 2014. DOI: {10.1103/PhysRevD.89.034008}. arXiv:1209.5859 (cit. on p. 84).

[61] R. Aaji et al., "Measurement of b hadron production fractions in 7 TeV pp collisions," *Phys. Rev. D*, vol. 85, p. 032 008, 3 2012. DOI: 10.1103/PhysRevD.85.032008. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.85.032008 (cit. on p. 84).

[62] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations," *Comput.Phys.Commun.*, vol. 10, pp. 343–367, 1975. DOI: {10.1016/0010-4655(75)90039-9} (cit. on pp. 89–91, 110).

[63] Y. Xie, "sFit: a method for background subtraction in maximum likelihood fit," *ArXiv e-prints*, 2009. arXiv:0905.0724 (cit. on pp. 92, 101).

[64] M. Williams et al., "Multivariate side-band subtraction using probabilistic event weights," *JINST*, vol. 4, P10003, 2009. DOI: {10.1088/1748-0221/4/10/P10003}. arXiv:0809.2548 (cit. on p. 92).

[65] R. Aaij et al., "Measurement of the $B_s^0 - \bar{B}_s^0$ oscillation frequency Δm_s in $B_s^0 \to D_s^-(3)\pi$ decays," *Phys. Rev. Lett.*, vol. B709, pp. 177–184, 2012. DOI: {10.1016/j.physletb.2012.02.031}. arXiv:1112.4311 (cit. on pp. 94, 124, 126, 172).

[66] B. Aubert et al, "Time-integrated and time-dependent angular analyses of $B \to \mathrm{J/\psi\,K}\pi$: A measurement of $\cos 2\beta$ with no sign ambiguity from strong phases," *Phys. Rev. D*, vol. 71, p. 032 005, 3 2005. DOI: 10.1103/PhysRevD.71.032005. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevD.71.032005 (cit. on p. 95).

[67] R. Aaij et al., "Determination of the Sign of the Decay Width Difference in the B_s^0 System," *Phys. Rev. Lett.*, vol. 108, 241 801, 24 2012. DOI: {10.1103/PhysRevLett.108.241801}. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.108.241801 (cit. on pp. 95, 138).

[68] E. G. et al., "Design patterns: elements of reusable object-oriented software," 1994 (cit. on p. 104).

[69] W. Verkerke and D. Kirkby, "The roofit toolkit for data modelling, physstat 05, 2005," (cit. on p. 104).

[70] L. Moneta et al., "Minuit2, is project to re-implement Minuit in C++," [Online]. Available: `http://lcgapp.cern.ch/project/cls/work-packages/mathlibs/minuit/index.html` (cit. on p. 110).

[71] S. Jadach, "Foam: A general-purpose cellular Monte Carlo event generator," *Computer Physics Communications*, vol. 152, 55–100, 2003. DOI: `{10.1016/S0010-4655(02)00755-5}`. arXiv:`{0203033}` (cit. on p. 112).

[72] M. Galassi et al., "GNU Scientific Library Reference Manual (3rd Ed.)," [Online]. Available: `http://www.gnu.org/software/gsl/` (cit. on p. 115).

[73] I.A. Antonov and V.M. Saleev, "An Economic Method of Computing LP Tau-Sequences," *USSR Computational Mathematics and Mathematical Physics*, vol. 19, pp. 252–256, 1980 (cit. on p. 115).

[74] Harrison, K and Tan, C I and Liko, D and Maier, A and Moscicki, J T and Egede, U and Jones, R W I and Soroko, A and Patrick, G N, "GANGA : a Grid User Interface," 2006 (cit. on p. 116).

[75] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, 605–610, 1979. DOI: `{10.1093/biomet/66.3.605}`. eprint: `http://biomet.oxfordjournals.org/content/66/3/605.full.pdf+html`. [Online]. Available: `http://biomet.oxfordjournals.org/content/66/3/605.abstract` (cit. on p. 117).

[76] "Sur la sphère vide. A la mémoire de Georges Vorono," *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, 1934, issue 6, pages 793–800 (cit. on p. 119).

[77] R. Aaij et al., "Tagged time-dependent angular analysis of $B_s^0 \to J/\psi\,\phi$ decays with the 2010 LHCb data," 2011, LHCb-CONF-2011-006 (cit. on p. 120).

[78] R. Aaij at al., "Measurement of the CP-violating phase ϕ_s in $\overline{B}_s^0 \to J/\psi\pi^+\pi^-$ decays," *Phys. Rev. Lett.*, vol. B713, pp. 378–386, 2012. DOI: `{10.1016/j.physletb.2012.06.032}`. arXiv:`1204.5675` (cit. on pp. 122, 140).

[79] S.M. Flatté, "On the nature of 0+ mesons," *Phys. Lett. B*, vol. 63, no. 2, pp. 228 –230, 1976, ISSN: 0370-2693. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0370269376906559 (cit. on p. 132).

[80] A. S. Dighe et al., "Extracting CKM phases and $B_s - \bar{B}_s$ mixing parameters from angular distributions of nonleptonic B decays," *Eur.Phys.J.*, vol. C6, pp. 647–662, 1999. DOI: {10.1007/s100520050372}. arXiv:9804253 (cit. on pp. 143, 146, 147).

[81] R. Fleischer and R. Knegjens, "Effective Lifetimes of B_s Decays and their Constraints on the B_s^0-$\bar{B}_s^0$ Mixing Parameters," *Eur.Phys.J.*, vol. C71, p. 1789, 2011. DOI: {10.1140/epjc/s10052-011-1789-9}. arXiv:1109.5115 (cit. on p. 144).

[82] A. S. Dighe et al., "Angular distributions and lifetime differences in $B_s \to J/\psi\phi$ decays," *Phys. Rev. Lett.*, vol. B369, pp. 144–150, 1996. DOI: {10.1016/0370-2693(95)01523-X}. arXiv:9511363 (cit. on pp. 146, 147).

[83] HFAG, "Heavy flavor averaging group," [Online]. Available: http://www.slac.stanford.edu/xorg/hfag (cit. on pp. 153, 154).

[84] HFAG, "Updated ϕ_s coverage plot from HFAG working group to include latest ATLAS result (ATLAS-CONF-2013-039)," updated result from http://www.slac.stanford.edu/xorg/hfag available online at http://marwww.in2p3.fr/ oleroy/pro/hfag/may2013 (cit. on p. 154).

[85] T. Aaltonen et al., "Measurement of the Bottom-Strange Meson Mixing Phase in the Full CDF Data Set," *Phys. Rev. Lett.*, vol. 109, p. 171 802, 2012. DOI: {10.1103/PhysRevLett.109.171802}. arXiv:1208.2967 (cit. on p. 154).

[86] The ATLAS collaboration, "Flavour tagged time dependent angular analysis of the $B_s \to J/\psi\phi$ decay and extraction of $\Delta\Gamma$ and the weak phase ϕ_s in ATLAS," no. ATLAS-CONF-2013-039, 2013 (cit. on p. 154).

[87] A. Lenz et al., "Constraints on new physics in $B - \bar{B}$ mixing in the light of recent LHCb data," *Phys. Rev. D*, vol. D86, p. 033 008, 2012. DOI: {10.1103/PhysRevD.86.033008}. arXiv:1203.0238 (cit. on pp. 155, 156).