$\mathcal{CP}$ Violation in $B_s \to J/\psi\phi$ decays at LHCb and Sensitivity to the Weak Mixing Phase β_s

Author:
Colin David MCLEAN

THESIS SUBMITTED IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

THE SCHOOL OF PHYSICS AND ASTRONOMY
THE UNIVERSITY OF EDINBURGH
2009

Acknowledgements

There are many people I wish to thank during the time of my PhD. I wish to thank my parents and brother, for your patience, kindness and support.

I wish to thank my close friends: T. Reiter, F. Diaz and L. Edwards. The joyful memories and words of wisdom that have kept me sane over these past three years. A big thanks to all the PhD students in the physics department, for making the department a wonderful research environment. Thanks to everyone I have met in the Edinburgh university triathlon club and those associated with Scottish touch rugby, your good natured banter has kept my enthusiasm and love for sport alive during the PhD.

Finally, I thank the academic staff at Edinburgh for helping me carry out my research. In particular, Prof. Clarke for the knowledge of parameter fitting that I have learnt, and Dr. Andres Osorio-Oliveros, from whom I have learnt so much. Thank you Andres for your help, and for being a great friend and research companion: I look forward to that cold beer with you Andres in CERNs restaurant One.

Declaration

This work represents the efforts of many members of the LHCb collaboration at CERN. I have been an integral part of a small team of people investigating the physics potential of the $B_s \rightarrow J/\psi\phi$ decay channel at LHCb. I hereby declare that the writing of this thesis is my own work, and that the work has not been submitted for any other degree or professional qualification except as specified.

signature: _____

name: _____

date: _____

Abstract

The LHCb experiment (CERN) will offer a precise hadronic probe with which to study $\mathcal{CP}$ violation, rare decays and possible New Physics (NP) effects occurring in the b-system. Contained within the vast physics program is a dedicated and intense effort to measure the $\mathcal{CP}$ violating weak mixing phase, β_s. This phase, observable in the neutral B_s-system, presents a sensitive indicator to NP occurring at the TeV scale.

This phase appearing through the theoretically clean $\bar{b} \to \bar{c}c\bar{s}$ quark-level transition, can be measured precisely using $B_s \to J/\psi\phi$ decays. The analysis of $B_s \to J/\psi\phi$ decays requires using its decay angular information to separate out the $\mathcal{CP}$ eigenstates. The sensitivity to β_s can then be extracted from the angular differential rates for this mode. This thesis presents the selection and reconstruction of this decay channel, analysis of the background specific to this channel and the method for fitting the data to extract β_s.

Previous studies within the collaboration have shown the β_s sensitivity using a reduced angular expression for the decay rates. By studying the full angular expression, we find both a quantitative and qualitative improvement in the precision with which β_s is obtained: the precision improves by approximately 20%, from $\pm$ 0.027 to $\pm$ 0.022 radians. In addition, we find it is possible to extract the b-tagging parameter from data itself, which aids to minimise the overall systematic effect. This thesis, in preparation for data taking at the LHCb, presents this expected improvement in measuring β_s.

Contents

List of Figures

List of Tables

1 Introduction

Shortly after the Big Bang, a mechanism unsettling the matter/anti-matter symmetry came into effect which ultimately shaped the universe we see today. The question why the magnitude of this asymmetery remains so large — the baryonic asymmetry is calculated to be $\mathcal{O}(10^{-10})$ [1] — still remains unanswered, and finding an explanation continues to be an important research topic in modern particle physics. Within the Standard Model (SM) of particle physics, it is linked to the effect known as $\mathcal{CP}$ violation ($\not{\mathcal{CP}}$), that we observe in weak interactions. Indeed, the requirement of $\mathcal{C}$ and $\mathcal{CP}$[1] violating interactions was one of the three conditions postulated by the physicist Andrei Sakharov (1967) to explain the asymmetry; the other two conditions being:

2. that baryon number conservation is violated. This condition is necessary, otherwise the same number of baryons and anti-baryons in each reaction would be created and destroyed.

3. a deviation from the thermal equilibrium. In thermal equilibrium, the density depends only on the temperature and the mass, which is the same for particles and antiparticles under $\mathcal{CPT}$ theorem[2].

Together, these three conditions provide the possibility for a universe initially containing equal amounts of matter and antimatter to evolve into a matter dominated universe.

The discovery of $\mathcal{CP}$ violation through weak interactions dates back to the early 1960s with the first observation in the kaon system through $K_L \rightarrow \pi^+\pi^-$ decays [2]. Until recently the only evidence of $\mathcal{CP}$ violation was in the kaon system. However, in the last decade or so, there has been huge experimental effort to further explore $\mathcal{CP}$ violation in the quark flavour sector of the Standard Model through B mesons. The first observation of $\mathcal{CP}$ violation outside the kaon system was in 2001 by the B factories, BaBar and Belle, with the measurement of $\sin(2\beta_d)$ from $B_d \rightarrow J/\psi(\mu^+\mu^-)K_s$ decays [3, 4][3]. In addition to this, the Tevatron has also been investigating $\mathcal{CP}$ violation mainly with B mesons, with the CDF experiment recently showing the possibility for $\mathcal{CP}$ violation to occur in the B_s system [5, 6].

[1] $\mathcal{CP}$ here refers to the combination of the charge-conjugation ($\mathcal{C}$) and parity ($\mathcal{P}$) transformation respectively.

[2] $\mathcal{T}$ here is called the time reversal operator. When applied to a quantum state it reverses the time coordinates.

[3] Here β refers to an angle within the unitarity triangle, shown in Figure 2.1. Both BaBar and Belle operating at the $\Upsilon(4S)$ resonance have analysed of $\mathcal{O}(10^8)$ B^0_d meson pairs.

The interest of studying $\mathcal{CP}$ violation beyond the current experimental energies of the B factories and the Tevatron, is because of the expectation of finding new sources of physics: new sources of $\mathcal{CP}$ violation for instance, or unravelling the flavour and/or fermion mass structure of the SM. At the energy scale of the Large Hadron Collider (LHC), $\mathcal{O}$(14TeV), we anticipate these phenomena to reveal themselves.

One of the interesting places to search for new physics will be through $B_s \to J/\psi(\mu^+\mu^-)$–$\phi(K^+K^-)$ decays. Indeed, this channel is often regarded as one of the "golden" modes due to its experimentally clean signature and its immediate signal for new physics contributing through B_s mixing. This channel will provide the main focus for this thesis. One of the primary objectives of the LHCb experiment (CERN's dedicated b physics experiment) will be to measure the weak mixing phase $-2\beta_s$ within this decay. The LHCb experiment provides an excellent environment with which to make this measurement, including several key subdetectors important for reconstructing this channel: the VELO, the RICH and the MUON systems. A description of the LHCb experiment and each of its subdetectors is given in chapter 3.

Reconstruction of the $B_s \to J/\psi\phi$ channel requires knowledge of the background specific to it. The background which presents a danger includes generic $b\bar{b}$-inclusive events and prompt decays, including J/ψ. Understanding this background allows us to develop a set of discriminating variables to help cut away this background. One of the most powerful variables for doing this was found to be the transverse momentum $p_\perp$ of the muons and kaons, this and the other discriminating variables used are presented in chapter 4. Besides the traditional cut based approach, the use of multivariate classifiers to help discriminate signal from background was investigated. In this thesis, the Boosted Decision Tree classifier (BDT) was found to give optimal separation of our signal with respect to prompt J/ψ background.

A complication of the $B_s \to J/\psi\phi$ decay arises due to its final state being one of two different possible $\mathcal{CP}$ eigenstates. In order to extract any physics from this decay we first need to disentangle this admixture of states, which can be achieved on a statistical basis using the decay angles: θ_{tr}, ϕ_{tr} and θ_ϕ associated with the $B_s \to J/\psi\phi$ decay. With the decay angles (defined in section 2.6) and the propertime of the B_s, we can then perform a time dependent angular analysis to extract the physics parameters of interest, specifically $-2\beta_s$.

Previous studies within the LHCb collaboration have done this using only one of these angles. In this thesis, we used all three angles to extract the sensitivity to $-2\beta_s$ using the full angular differential decay rate expressions, given in chapter 2. The sensitivities were then extracted from fast (toy) Monte Carlo simulations, using as input the results from previous studies on fully simulated data: yields, background levels, resolution and flavour tagging. The results obtained are presented and compared to that of the one angle analysis in chapter 6.

2

The Standard Model and $\mathcal{CP}$ violation

The Standard Model (SM) of particle physics encapsulates our current knowledge of the fundamental building blocks of matter and the forces acting between them. As summarised in Table 2.1, the SM is experimentally well tested and established.

The bosons mediating the forces fall into two groups at the current level of understanding: the photon γ, the $W^{\pm}$ and the Z^0 mediating the electroweak interactions and the gluons g mediating the strong interaction. The strong force is described by a theory known as Quantum ChromoDynamics (QCD) and is responsible for binding quarks together into hadrons.

		Leptons			Quarks		Force Carriers		
	Particle	Mass ($\mathrm{MeV/c^2}$)	$\mathcal{Q}/e$	Particle	Mass († $\mathrm{MeV/c^2}$ $^{\dagger\dagger}\mathrm{GeV/c^2}$)	$\mathcal{Q}/e$	Gauge Boson	$\mathcal{Q}/e$	Force
Generations 1^{st}	ν_e	$< 2.2\times10^{-6*}$	0	u	$^{\dagger}2.55^{+0.75}_{-1.025}$	2/3			Strong
	e^-	$0.511\pm(1.3\times10^{-8})$	-1	d	$^{\dagger}5.04^{+0.96}_{-1.54}$	-1/3	g	-	
2^{nd}	ν_μ	$< 0.19^{**}$	0	c	$^{\dagger\dagger}1.27^{+0.07}_{-0.11}$	2/3			
	μ^-	$< 105.66\pm(4.0\times10^{-6})$	-1	s	$^{\dagger}104^{+26}_{-34}$	-1/3	γ	0	Electroweak
3^{rd}	ν_τ	$< 18.2^{**}$	0	t	$^{\dagger\dagger}171.2\pm2.1$	2/3	Z	0	
	τ^-	1777.0 ± 0.17	-1	b	$^{\dagger\dagger}4.20^{+0.17}_{-0.07}$	-1/3	W	±1	

Table 2.1: Properties of the elementary particles in the SM taken from [7]. Here, * indicates the Confidence Level (CL) at 90% and ** the CL at 95%.

Electroweak interactions fall naturally into two classes: charged and neutral currents. Interactions exchanging a $W^{\pm}$ boson are described by charged currents, while those interactions exchanging a neutral Z^0 boson yield neutral currents. Interactions which allow transformations of one species of quarks (known as flavour) into other species are called flavour changing processes. The main interest in this thesis will be on charged currents entering the Electroweak Lagrangian.

2.1 Electroweak Lagrangian

The Standard Model Lagrangian has two distinct parts: the QCD and the Electroweak Lagrangian. Accordingly, we can distinguish between two types of theoretically possible $\mathcal{CP}$ violation: $\mathcal{CP}$ violation originating from the strong interaction Lagrangian, which is usually referred to as the strong $\mathcal{CP}$ problem, and $\mathcal{CP}$ violation occurring in the Electroweak Lagrangian. The Lagrange density describing Electroweak interactions can be written symbolically as:

$$\mathcal{L}_{EW} = \mathcal{L}(f,g) + \mathcal{L}(f,H) + \mathcal{L}(g,H) + \mathcal{L}(g) - V(H) \tag{2.1}$$

Where in the notation above:

- f: corresponds to the fermions in the theory; quarks and leptons.
- g: corresponds to the gauge bosons, $\vec{W}$ and $\mathcal{B}$, before Spontaneous Symmetry Breaking (SSB).
- H: corresponds to the Higgs doublet, introduced in equation (2.3).

Of the terms in equation (2.1) only $\mathcal{L}(f,H)$ is found to violate $\mathcal{CP}$, the other terms are $\mathcal{CP}$ conserving.[1] All the known $\mathcal{CP}$ violation resides in the $\mathcal{L}(f,H)$ Lagrangian, which describes interactions of the fermions with the Higgs doublet. The Lagrangian takes the form:

$$\mathcal{L}(f,H) = \sum_{j,k=1}^{3} \left[Y^u_{j,k} (\bar{q}^u, \bar{q}^d)_{j,L} H^c q^u_{k,R} + Y^d_{j,k} (\bar{q}^u, \bar{q}^d)_{j,L} H q^d_{k,R} + h.c. \right]. \tag{2.2}$$

where q_L and $q^{u/d}_{j,R}$ describe the left handed[2] quark fields arranged into $\mathcal{SU}(2)_L$ doublet and singlet fields. The terms $Y^{u/d}_{j,k}$ appearing in equation 2.2 are the Yukawa coupling constants. Due to the Lorentz invariance of each term in equation (2.2), the Yukawa couplings are arbitrary complex numbers in the SM. The scalar fields, $H^{(c)}$, given in equation (2.2) are defined as:

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad H^c \equiv i\sigma_2 H^\star = \begin{pmatrix} (\phi^0)^\star \\ -\phi^- \end{pmatrix}, \tag{2.3}$$

and denote the Higgs doublet and its charge-conjugate respectively. The Higgs doublet represents four real scalar fields, which may be introduced as $\phi^+ = \frac{1}{\sqrt{2}}(\phi^1 + i\phi^2)$ and $\phi^0 = \frac{1}{\sqrt{2}}(\phi^3 + i\phi^4)$. Both H and H^c transform[3] as doublets under $\mathcal{SU}(2)_L$.

[1] As mentioned earlier $\mathcal{CP}$ violation may enter into the QCD Lagrangian, however it has yet to be experimentally verified [8].

[2] All particles in the SM acquire both right and left chirality (or handedness); apart from the neutrino which is found to exist only in left handed states.

[3] The action of $\mathcal{SU}(2)_L$ on H^c is: $H^c \rightarrow e^{i(\vec{\alpha}\cdot\vec{\sigma})} H^c = \mathcal{U}(\vec{\alpha}) H^c$.

Spontaneous symmetry breaking and mass

Under Spontaneous Symmetry Breaking (SSB), the mechanism[4] by which the gauge bosons and fermions in equation (2.2) acquire mass, the scalar field ϕ^0 becomes perturbed by an amount $\phi^0 \rightarrow \phi^0 + v$. The amount by which ϕ^0 is shifted is a real number and is called the vacuum expectation value of the field. The three scalar fields ϕ^j (with $j \in \{1\ ,2\ ,3\}$ are absorbed by the $W^{\pm}$ and the Z bosons which consequently become massive. The essential point after SSB is that in $\mathcal{L}(f,H)$ only terms involving the field ϕ^0 remain:

$$\mathcal{L}(f,H) \quad \underrightarrow{\text{SSB}} - \sum_{j,k=1}^{3} \left[M^u_{j,k} \overline{q}^u_{j,L} q^u_{k,R} + M^d_{j,k} \overline{q}^d_{j,L} q^d_{k,R} + h.c \right] (1 + \tfrac{1}{v}\phi^0) \tag{2.4}$$

where the terms $M^{u/d}_{j,k}$ appearing in equation (2.4) are called the quark mass matrices and are defined as:

$$M^u_{j,k} = -\frac{v}{\sqrt{2}} Y^u_{i,k}, \quad M^d_{j,k} = -\frac{v}{\sqrt{2}} Y^d_{j,k}.$$

Due to the three generations in the SM these are 3×3 matrices in flavour space.

In the discussion so far we have been dealing with non-physical, i.e massless fields. But when describing the dynamics of elementary particles, we need to express the fields given in equation (2.4), in terms of the physical fields. This is achieved by diagonalising the quark mass matrices M^u and M^d:

$$\begin{aligned} \mathcal{U}^u_L M^u \mathcal{U}^{u\dagger}_R = \mathcal{D}_u = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \\ \mathcal{U}^d_L M^d \mathcal{U}^{d\dagger}_R = \mathcal{D}_d = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} \end{aligned} \tag{2.5}$$

Here, $\mathcal{U}^{u/d}_{L,R}$ denote unitary matrices[5] which multiply the mass matrix $M^{u/d}$ from the left and right respectively. The matrices $\mathcal{D}_{u,d}$ are diagonal 3×3 matrix, with the quantities $m_u, m_d, \ldots$ in equation (2.5), denoting the eigenvalues of the mass matrices, i.e. the quark masses. Substituting

[4] It is the Higgs mechanism [9] that gives the quarks their mass in the SM; this mechanism assumes a scalar field called the Higgs field is present in the vacuum that can slow down some (otherwise massless) elementary particles. Such particles then become massive particles travelling at speeds less that the speed of light. Other particles - like the photon - are immune to this field: they don't slow down and they remain massless. Although the Higgs field is not detectable, accelerators with sufficient energy and luminosity, will hopefully discover the Higgs particle.

[5] Unitarity of $\mathcal{U}_{L,R}$ implies that $\mathcal{U}_{L,R}\mathcal{U}^{\dagger}_{L,R} = 1$.

equation (2.5) into equation (2.4) gives for the up-type quarks:

$$\begin{aligned}
\bar{q}^u_{j,L} M^u_{j,k} q^u_{k,R} &= \bar{q}^u_L M^u q^u_R \\
&= \bar{q}^u_L \mathcal{U}^\dagger_L \mathcal{U}^u_L M^u \mathcal{U}^\dagger_R \mathcal{U}^u_R q^u_R \\
&= \overline{\mathcal{U}^u_L q^u_L} \mathcal{D} \mathcal{U}^u_R q^u_R \\
&= \overline{\mathcal{U}^u_L q^u_L} \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \mathcal{U}^u_R q^u_R
\end{aligned} \tag{2.6}$$

A similar expression is obtained for the down-type quarks. Thus the physical fields, entering Lagrangian $\mathcal{L}(f,g)$ given in equation (2.1), can be expressed as:

$$\begin{aligned}
q_L^{u(d),phys} &= \mathcal{U}_L^{u(d)} q_L^{u(d)} = \mathcal{U}_L^{u(d)} \begin{pmatrix} u(d)_L \\ c(s)_L \\ t(b)_L \end{pmatrix}, \\
q_R^{u(d),phys} &= \mathcal{U}_R^{u(d)} q_R^{u(d)} = \mathcal{U}_R^{u(d)} \begin{pmatrix} u(d)_R \\ c(s)_R \\ t(b)_R \end{pmatrix}.
\end{aligned} \tag{2.7}$$

Expressed in terms of the physical fields, the neutral currents are found not to mix quark flavours in the SM. This important feature is describe by the Glashow-Illiopoulos-Maiani (GIM) mechanism [10], which is further discussed in Appendix A. Extensions of the SM often encounter difficulties in this respect by predicting FCNCs larger than the observational bounds. From equation (2.2), the charge currents can be expressed as:

$$\begin{aligned}
\mathcal{L}_{cc} &\equiv W^+ \bar{q}^u_L \gamma^\mu q^d_L + h.c. \\
&= W^+ \bar{q}^{u,phys}_L \gamma^\mu \mathcal{U}^u_L \mathcal{U}^{d\dagger}_L q^{d,phys}_L + h.c. \\
&= W^+ \bar{q}^{u,phys}_L \gamma^\mu V_{CKM} q^{d,phys}_L + h.c. \\
&= W^+ J^\mu_{cc} + h.c,
\end{aligned} \tag{2.8}$$

with $V_{CKM} \equiv \mathcal{U}^u_L \mathcal{U}^{d\dagger}_L$ and J^μ_{cc} denoting the charge current:

$$J^\mu_{cc} \equiv (u,c,t)_L \gamma^\mu \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}_L \tag{2.9}$$

The matrix V_{CKM}, often referred to as the Cabbibo-Kobayashi-Maskawa (CKM) matrix [11], is unitary following the requirements of unitarity for the $\mathcal{U}^{u/d}$ matrices in equation (2.5). Each complex element,V_{ij}, in V_{CKM} expresses the coupling of the i^{th} up-type quark to the j^{th} down-type quark and as such represents the mixing of the quark flavours within the SM.

Parametrisation of the CKM matrix

Since the CKM matrix is 3×3 and contains complex elements it has, in principle, eighteen free parameters. The condition of unitarity imposes nine additional constraints on V_{CKM} leaving only nine free parameters. In general given a $N \times N$ complex unitary matrix, $N(N-1)/2$ of its elements may be taken as Euler angles, which are introduced when dealing with rotations in an N dimensional Euclidean space. The remaining parameters of which there will be $N^2 - N(N-1)/2$ are called phases. Therefore, the nine free parameters of the CKM matrix will be expressed in terms of three Euler angles and six phases.

To understand how to deal with the phases appearing in V_{CKM}, it is necessary to realise that for $\mathcal{CP}$ conservation in equation (2.8), the CKM matrix must be real. This is because in quantum field theory, and indeed quantum mechanics in general, what matters is not the absolute phases but the relative phases of the different fields[6]. The phases entering the fields in equation (2.8) are arbitrary and can be redefined without changing the physics as follows:

$$\begin{aligned} u_L &\to e^{i\phi(u)} u_L, \quad c_L \to e^{i\phi(c)} c_L, \quad etc \\ d_L &\to e^{i\phi(d)} d_L, \quad s_L \to e^{i\phi(s)} s_L, \quad etc \end{aligned} \tag{2.10}$$

Where $\phi(x)$, $x = u, c, \cdots, d, s, \cdots$ are arbitrary real numbers. In general there will be $2N$ such quantities. Under the quark field rephases given in equation (2.10), the CKM matrix can be expressed as:

$$\begin{aligned} V_{CKM} \quad \to \quad & \begin{pmatrix} e^{-i\phi(u)} & 0 & 0 \\ 0 & e^{-i\phi(c)} & 0 \\ 0 & 0 & e^{-i\phi(t)} \end{pmatrix} \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} e^{i\phi(d)} & 0 & 0 \\ 0 & e^{i\phi(s)} & 0 \\ 0 & 0 & e^{i\phi(b)} \end{pmatrix} \\ = \quad & \begin{pmatrix} e^{i(\phi(d)-\phi(u))} V_{ud} & e^{i(\phi(s)-\phi(u))} V_{us} & e^{i(\phi(b)-\phi(u))} V_{ub} \\ e^{i(\phi(d)-\phi(c))} V_{cd} & e^{i(\phi(s)-\phi(c))} V_{cs} & e^{i(\phi(b)-\phi(c))} V_{cb} \\ e^{i(\phi(d)-\phi(t))} V_{td} & e^{i(\phi(s)-\phi(t))} V_{ts} & e^{i(\phi(b)-\phi(t))} V_{tb}. \end{pmatrix} \end{aligned} \tag{2.11}$$

Of the phase differences $(\phi(j) - \phi(k))$ appearing in equation (2.11), five are linearly independent and can be freely rotated away by choosing:

$$V_{ud} \geq 0 \quad V_{us} \geq 0 \quad V_{ub} \geq 0 \quad V_{cd} \leq 0 \quad V_{td} \leq 0 \tag{2.12}$$

However the last remaining phase cannot. This remaining phase, denoted by δ, becomes our source of $\mathcal{CP}$ violation in the Standard Model. Since V_{CKM} is unitary, $V_{CKM}^\dagger V_{CKM} = V_{CKM} V_{CKM}^\dagger =$

[6] From quantum mechanics the phase of a wavefunction is not a measurable quantity. A wavefunction ψ and $e^{i\eta}\psi$, where η is a real number, are physically equivalent.

1, a set of twelve equations given by $\sum_k V_{ki}V_{kj}^* = \delta_{ij}$ can be obtained:

$$\begin{aligned}
\text{rows } 1\times 1 \quad & uu : |V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 1, && (2.13)\\
\text{rows } 2\times 2 \quad & cc : |V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2 = 1, && (2.14)\\
\text{rows } 3\times 3 \quad & tt : |V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2 = 1, && (2.15)\\
\text{columns } 1\times 1 \quad & dd : |V_{ud}|^2 + |V_{cd}|^2 + |V_{td}|^2 = 1, && (2.16)\\
\text{columns } 2\times 2 \quad & ss : |V_{us}|^2 + |V_{cs}|^2 + |V_{ts}|^2 = 1, && (2.17)\\
\text{columns } 3\times 3 \quad & bb : |V_{ub}|^2 + |V_{cb}|2 + |V_{tb}|^2 = 1, && (2.18)\\
\text{rows } 1\times 2 \quad & cu : V_{ud}^*V_{cd} + V_{us}^*V_{cs} + V_{ub}^*V_{cb} = 0, && (2.19)\\
\text{rows } 1\times 3 \quad & tu : V_{ud}^*V_{td} + V_{us}^*V_{ts} + V_{ub}^*V_{tb} = 0, && (2.20)\\
\text{rows } 2\times 3 \quad & tc : V_{cd}^*V_{td} + V_{cs}^*V_{ts} + V_{cb}^*V_{tb} = 0, && (2.21)\\
\text{columns } 1\times 2 \quad & sd : V_{ud}^*V_{us} + V_{cd}^*V_{cs} + V_{td}^*V_{ts} = 0, && (2.22)\\
\text{columns } 1\times 3 \quad & bd : V_{ud}^*V_{ub} + V_{cd}^*V_{cb} + V_{td}^*V_{tb} = 0, && (2.23)\\
\text{columns } 2\times 3 \quad & bs : V_{us}^*V_{ub} + V_{cs}^*V_{cb} + V_{ts}^*V_{tb} = 0. && (2.24)
\end{aligned}$$

The twelve unitary relations and five inequality relations given above, lead to the canonical form for the V_{CKM} [12]. From its canonical form, we can obtain many other possible parametrisation of V_{CKM} [13]. We will look at two parametrisations: the standard parametrisation, which has been adopted by the Particle Data Group (PDG), and the Wolfenstein parametrisation. The standard parametrisation can be defined by introducing the terms $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$ $(i, j = 1, 2, 3)$ such that:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -s_{23}c_{12} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \qquad (2.25)$$

where δ again describes the $\mathcal{CP}$ violating phase. We obtain from phenomenological applications measurements of s_{13} and s_{23} of the $\mathcal{O}(10^{-3})$ and $\mathcal{O}(10^{-2})$ respectively [14]. Consequently c_{13}=c_{23}=1 and we can define the four independent parameters as:

$$s_{12} = |V_{us}|, \qquad s_{13} = |V_{ub}|, \qquad s_{23} = |V_{cb}|, \qquad \delta. \qquad (2.26)$$

The two main advantages of this parametrisation to the canonical one are: that the terms s_{12}, s_{13} and s_{23} can be measured independently and that the $\mathcal{CP}$ violating phase is always multiplied by something which is very small, namely s_{13}. This last point shows the suppression of the $\mathcal{CP}$ violating phase independent of its actual size.

A more transparent approximation and certainly more popular when discussing flavour physics

is the one introduced by Wolfenstein [15]. The Wolfenstein parametrisation is an approximate parametrisation of the CKM matrix emphasising the hierarchical nature of the CKM elements. In this parametrisation, the elements are expanded in a power series of the small parameter λ=$|V_{us}|$=0.22 and our four independent parameters are replaced by:

$$\lambda, \qquad A, \qquad \rho, \qquad \eta. \tag{2.27}$$

Where the imaginary part of η is responsible for the $\not{\mathcal{CP}}$ phase. The relationship between the two parameter sets, in the standard and Wolfenstein parametrisation then becomes:

$$s_{12} = \lambda, \qquad s_{23} = A\lambda^2, \qquad s_{13}e^{-i\delta} = A\lambda^3(\rho - i\eta). \tag{2.28}$$

Using this change of variables we can expand each CKM element to $\mathcal{O}(\lambda^5)$, which is high enough to match the level of sensitivity that will be obtained at the LHCb. We can express the CKM matrix through the standard parametrisation as [16]:

$$V_{CKM} = \begin{array}{c} \\ \mathrm{u} \\ \mathrm{c} \\ \mathrm{t} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{d} & \mathrm{s} & \mathrm{b} \end{array} \\ \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda + \frac{1}{2}A^5\lambda^5\left[1 - 2(\rho + i\eta)\right] & 1 - \frac{1}{2}\lambda^2 + \frac{1}{8}\lambda^4(1 - 4A^2) & A\lambda^2 \\ A\lambda^3(1 - \bar{\rho} - i\bar{\eta}) & -A\lambda^2 + A\lambda^4(\frac{1}{2} - \rho - i\eta) & 1 - \frac{1}{2}A^2\lambda^4 \end{pmatrix} \end{array} + \mathcal{O}(\lambda^6). \tag{2.29}$$

Where:

$$\bar{\rho} = \rho(1 - \frac{\lambda^2}{2}), \qquad \bar{\eta} = \eta(1 - \frac{\lambda^2}{2}). \tag{2.30}$$

Consequently we find that:

$$\begin{aligned} V_{us} &= \lambda, & V_{cb} &= A\lambda^2, \\ V_{ub} &= A\lambda^3(\rho - i\eta), & V_{td} &= A\lambda^3(1 - \bar{\rho} - i\bar{\eta}). \end{aligned} \tag{2.31}$$

The strength of $\not{\mathcal{CP}}$ in the SM can be expressed through the Jarskog invariant [17]:

$$J_{CP} = |\Im(V_{i\alpha}V_{j\beta}V_{i\beta}V_{j\alpha})|, \qquad (i \neq j, \alpha \neq \beta). \tag{2.32}$$

Where $\mathcal{CP}$ is violated when the imaginary part of the products of the CKM matrix elements is different from zero. The parameter J_{CP} can be used to measure the strength of $\mathcal{CP}$ violation in the SM. Using equation (2.29) we find that:

$$J_{CP} \approx \lambda^6 A^2 \eta(1 - \lambda^2/2) \sim 10^{-5}. \tag{2.33}$$

$\mathcal{CP}$ violation in the SM is therefore a small effect. However, this can be enhanced by new complex couplings, arising from NP, which would constitute additional sources of $\mathcal{CP}$ violation.

2.2 The unitarity triangles

The six orthogonality relations equations (2.19) to (2.24) can be represented as triangles in the complex plane each having the same area, $A_{\triangle} = \frac{1}{2}J_{CP}$. By measuring the angles and sides of these triangles it is possible to over constrain the CKM picture of the SM. The six unitarity triangles are shown in Figure 2.1, where each side has been expressed in terms of the Wolfenstein parametrisation. From Figure 2.1 it can be seen that only two of the triangles have all three sides of similar length, $\mathcal{O}(A\lambda^3)$, namely (bd) and (tu). The other triangles are quite flat and will require very high precision to prove experimentally that they are not degenerate to a line. The (bd) triangle is often referred to as the unitarity triangle and has been the central target for the B-factories studying $\mathcal{CP}$ violation in the SM. It is also important to constrain the angles and sides of the other triangles. The (tu) triangle differs only sightly from the unitarity triangle via $\mathcal{O}(\lambda^2)$ corrections. Differences, or even non-closure of these two triangles may signify deviations from or extensions to the SM. After the B factories, it will be left to the LHC experiments to overconstain this triangle.

The CKM matrix may also be parametrised using four independent phases, the four angles associated with the unitarity triangles given in equations (2.34). These angles are rephase invariant terms, meaning that they resemble physical quantities independent of the CKM parametrisation. The angles are given by:

$$\begin{aligned}
\gamma &\equiv \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right], \\
\beta_d &\equiv \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \\
\beta_s &\equiv \arg\left[-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right], \\
\beta_K &\equiv \arg\left[-\frac{V_{cs}V_{cd}^*}{V_{us}V_{ud}^*}\right],
\end{aligned} \tag{2.34}$$

where by construction $\alpha + \beta + \gamma = \pi$ and are related to the (bd) triangle. It is clear that triangle (bd) will be mainly determined by the B_d system, since it contains the term $V_{td}V_{tb}^*$ which controls B_d oscillations. Likewise, the angle β_s which is contained within the (sb) triangle is doubly Cabibbo-suppressed ($\mathcal{O}(\lambda^2)$) and is related to the B_s system since the $V_{tb}V_{ts}^*$ controls $B_s - \bar{B}_s$ oscillations. Finally, it can be shown using the Wolfenstein parametrisation how each of the angles are related by the $\mathcal{CP}$ violation parameter, η.

$$\begin{aligned}
\gamma &= \arg\left[-1(\rho - i\eta)\right] \approx \eta, \\
\beta_d &= \arg\left[-1 + \rho + i\eta\right] \approx -\eta, \\
\beta_s &= \arg\left[1 - \lambda^2(\frac{1}{2} - \rho - i\eta) + \mathcal{O}(\lambda^4)\right] \approx \lambda^2\eta, \\
\beta_K &= \arg\left[1 - A^2\lambda^4(\frac{1}{2} - \rho - i\eta) + \mathcal{O}(\lambda^6)\right] \approx A^2\lambda^4\eta.
\end{aligned} \tag{2.35}$$

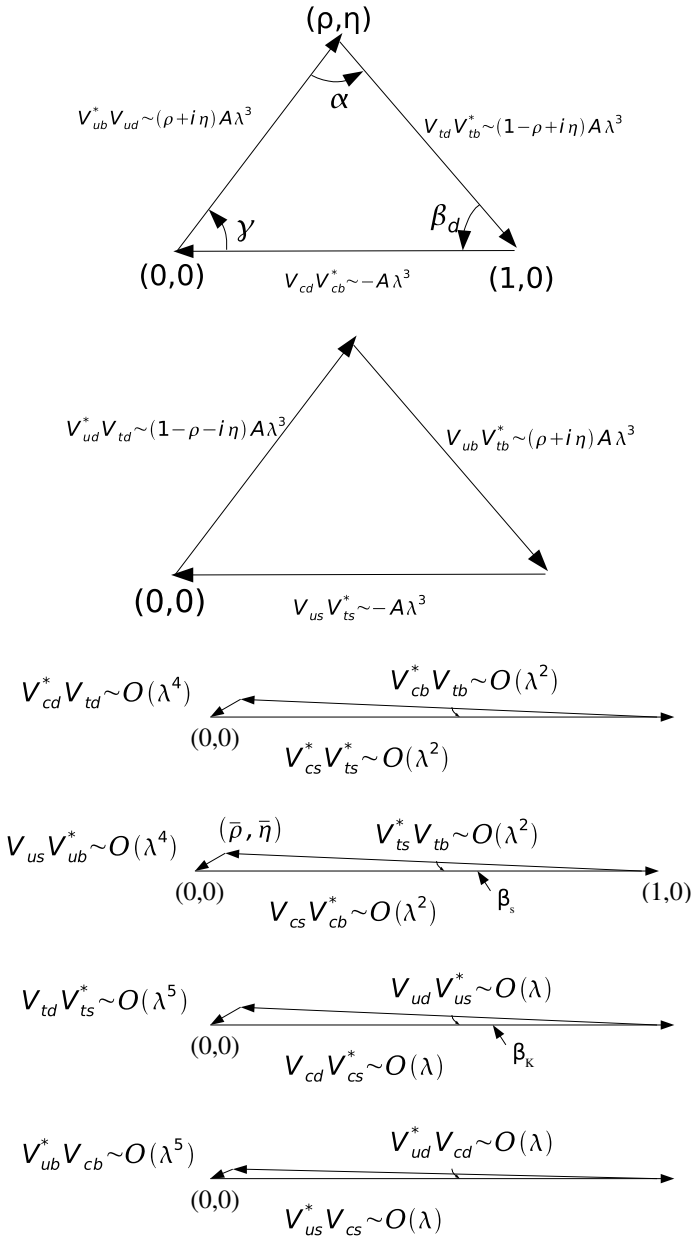


Figure 2.1: The three types of unitarity triangles. The top two triangles are the bd and tu triangles, corresponding to equations (2.23) and 2.20 respectively. The bd triangle is usually denoted as the unitarity triangle and defines the angles: α, β_d and γ. The next two triangles are tc and bs, corresponding to equations (2.21) and (2.24), respectively. Here the bs triangle defined the angle β_s. The final two triangles are the sd and cu triangles, corresponding to equations (2.22) and (2.19) respectively. The sd triangle leads to the angle β_K. All triangles have the same area.

2.3 The neutral $B_q^0 - \bar{B}_q^0$ system

Neutral meson systems have the peculiar and interesting property that its particle and anti-particle are distinguished only by the flavour quantum numbers, which is not conserved by the weak interactions responsible for their decay. Several families of this kind are known: neutral kaons, neutral D mesons and neutral B_d and B_s mesons. The purpose of this section is to describe the standard quantum mechanical formalism applicable for the time evolution of the B_q system, where $q \in \{d, s\}$ representing the flavour content of the B meson.

We begin by discussing the flavour eigenstates of the B_q^0 system. Denoting the neutral meson by B_q^0 and its antiparticle by $\overline{B_q^0}$, with non-zero flavour eigenvalues, the flavour eigenstates are defined as:

$$F|B_q^0\rangle = +|B_q^0\rangle, \quad F|\overline{B_q^0}\rangle = -|\overline{B_q^0}\rangle. \tag{2.36}$$

Where F is the Flavour quantum operator.[7] Focusing on the lowest mass states of the system (all higher mass states will rapidly decay into these via strong interactions), the neutral mesons are found to be pseudo-scalar particles. The effect of Parity($\mathcal{P}$) on these eigenstates is therefore,

$$\mathcal{P}|B_q^0\rangle = -|B_q^0\rangle, \quad \mathcal{P}|\overline{B_q^0}\rangle = -|\overline{B_q^0}\rangle. \tag{2.37}$$

The effect of charge conjugation ($\mathcal{C}$) will be to transform B_q^0 into $\overline{B_q^0}$ and vice versa, this introduces a phase factor η_C such that:

$$\mathcal{C}|B_q^0\rangle = \eta_C|\overline{B_q^0}\rangle, \quad \mathcal{C}|\overline{B_q^0}\rangle = \eta_C^*|B_q^0\rangle, \tag{2.38}$$

with $|\eta_C|^2$ = 1. The combined effect of $\mathcal{CP}$ is then given by,

$$\mathcal{CP}|B_q^0\rangle = \eta_{CP}|B_q^0\rangle, \quad \mathcal{CP}|\overline{B_q^0}\rangle = \eta_{CP}^*|\overline{B_q^0}\rangle, \tag{2.39}$$

introducing the $\mathcal{CP}$ phase η_{CP}=$e^{i\xi_{CP}}$ (such that $|\eta_{CP}|^2$=1).

Weak interactions do not conserve flavour and therefore F does not commute with the full Hamiltonian, $\mathcal{H}$, of the system. The flavour eigenstates are therefore not the physical eigenstates of the system. It is possible to use this fact to split the full Hamiltonian of the system into the flavour invariant (strong) Hamiltonian $\mathcal{H}_0$ and non-flavour invariant (weak) Hamiltonian $\mathcal{H}_W$. The states $|B_q^0\rangle$ and $|\overline{B_q^0}\rangle$ can be defined eigenstates of $\mathcal{H}_0$, while the $\mathcal{H}_W$ term leads to B_q-$\overline{B_q}$ mixing.

[7] When acting on a flavour eigenstate, F defines the mesons flavour. For example, $F|B_{\bar{b}s}\rangle = \frac{-\frac{1}{3}}{N_d}|B_{\bar{b}s}\rangle = +|B_{\bar{b}s}\rangle$. Where $N_d = \frac{-1}{3}$

Under the influence of the full Hamiltonian, the time evolution of a general state $|\psi(t)\rangle$ (a superposition of $|B_q^0\rangle$ and $|\overline{B_q^0}\rangle$) will be an admixture of decay and mixing described by:

$$|\psi(t)\rangle = a(t)|B_q^0\rangle + b(t)|\overline{B_q^0}\rangle + \sum_n c_n(t)|n\rangle \tag{2.40}$$

where $|n\rangle$ represents any state of any number of particles which are decay modes of the original mesons (i.e. $|n\rangle$=$|\pi\pi\rangle, |3\pi\rangle, |\pi l\nu_l\rangle \cdots$) and where t denotes the proper time measured in the B_q^0-$\overline{B_q^0}$ rest frame. Since the states B_q^0 and $\overline{B_q^0}$ are coupled together via mixing and can decay into other states, the exact solution for this general state is particularly complicated. The problem can be greatly simplified using the Weisskopf-Wigner approximation [12, 18] in which;

- initial states that are linear combinations of B_q^0 and $\overline{B_q^0}$ only are considered; the effect of the weak interaction Hamiltonian on the intermediate states $|n\rangle$ are considered zero, $\langle n|\mathcal{H}_W|n\rangle$=0.

- the time evolution of the coefficients describing these two components, namely $a(t)$ and $b(t)$, are studied for times much larger than the typical strong-interaction scale.

Using this approximation, the time evolution of the general state $|\psi(t)\rangle$ projected onto the B_q^0-$\overline{B_q^0}$ subspace is effectively governed by the Schrödinger like equation:

$$i\frac{\partial}{\partial t}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \mathcal{H}_{eff}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} \equiv \begin{pmatrix} \langle B_q^0|\mathcal{H}_{eff}|B_q^0\rangle & \langle B_q^0|\mathcal{H}_{eff}|\overline{B_q^0}\rangle \\ \langle \overline{B_q^0}|\mathcal{H}_{eff}|B_q^0\rangle & \langle \overline{B_q^0}|\mathcal{H}_{eff}|\overline{B_q^0}\rangle \end{pmatrix}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix}, \tag{2.41}$$

Where the effective Hamiltonian, $\mathcal{H}_{eff}$, has been introduced. It is common to split $\mathcal{H}_{eff}$ into a hermitian and an anti-hermitian part, such that:

$$\mathcal{H}_{eff} = M - \frac{i}{2}\Gamma,$$

where both M (mass matrix) and Γ (decay matrix) are hermitian,

$$M = \frac{1}{2}\left(\mathcal{H}_{eff} + \mathcal{H}_{eff}^{\dagger}\right) = M^{\dagger} \quad \Gamma = i\left(\mathcal{H}_{eff} - \mathcal{H}_{eff}^{\dagger}\right) = \Gamma^{\dagger}. \tag{2.42}$$

A consequence of the $\mathcal{CPT}$ theorem is that the masses and lifetimes of the particles are identical: $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$ (or H_{11} = H_{22}). This reduces the number of parameters in $\mathcal{H}_{eff}$ to six, which can then be expressed as:

$$\mathcal{H}_{eff} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} M_{11} - \frac{i}{2}\Gamma_{11} & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M_{22} - \frac{i}{2}\Gamma_{22} \end{pmatrix}. \tag{2.43}$$

By expanding the effective Hamiltonian in powers of the weak interaction Hamiltonian $\mathcal{H}_W$ to

second order we can read off the matrix elements for M and Γ:

$$\mathcal{H}_{eff} = \mathcal{H}_0 + \mathcal{H}_W + \sum_n \mathcal{H}_W|n\rangle\langle n|\mathcal{H}_W \left(\mathcal{P}\frac{1}{m_0 - E_n + i\epsilon} - i\pi\delta(m_0 - E_n)\right) + \dots \quad (2.44)$$

where m_0 is the unperturbed mass of the $B_q^0(\overline{B_q^0})$ such that $\mathcal{H}_0|B_q^0\rangle = m_0|B_q^0\rangle$, and E_n is the energy of the intermediate states, $|n\rangle$, such that $\mathcal{H}_0|n\rangle = E_0\mathcal{H}_0|n\rangle$. The explicit expressions for the matrix elements are then,

$$M_{11} = M_{11}^* = m_0 + \langle B_q^0|\mathcal{H}_W|B_q^0\rangle + \sum_n \mathcal{P}\frac{|\langle n|\mathcal{H}_W|B_q^0\rangle|^2}{m_0 - E_n} \quad (2.45)$$

$$M_{22} = M_{22}^* = m_0 + \langle \overline{B_q^0}|\mathcal{H}_W|\overline{B_q^0}\rangle + \sum_n \mathcal{P}\frac{|\langle n|\mathcal{H}_W|\overline{B_q^0}\rangle|^2}{m_0 - E_n} \quad (2.46)$$

$$M_{12} = M_{21}^* = \langle B_q^0|\mathcal{H}_W|\overline{B_q^0}\rangle + \sum_n \mathcal{P}\frac{\langle B_q^0|\mathcal{H}_W|n\rangle\langle n|\mathcal{H}_W|\overline{B_q^0}\rangle}{m_0 - E_n} \quad (2.47)$$

$$\Gamma_{11} = \Gamma_{11}^* = 2\pi\sum_n \delta(m_0 - E_n)|\langle n|\mathcal{H}_W|B_q^0\rangle|^2 \quad (2.48)$$

$$\Gamma_{22} = \Gamma_{22}^* = 2\pi\sum_n \delta(m_0 - E_n)|\langle n|\mathcal{H}_W|\overline{B_q^0}\rangle|^2 \quad (2.49)$$

$$\Gamma_{12} = \Gamma_{21}^* = 2\pi\sum_n \delta(m_0 - E_n)\langle B_q^0|\mathcal{H}_W|n\rangle\langle n|\mathcal{H}_W|\overline{B_q^0}\rangle. \quad (2.50)$$

Here $\mathcal{P}$ indicates the principle part description which is effectively a sum over all the intermediate states to which B_q^0 and $\overline{B_q^0}$ can decay. The intermediate states contributing to the off-diagonal term M_{12} are virtual, and within the B_s-system these states are dominated by top quark exchange as shown in Figure 2.2.[8] The intermediate states contributing to Γ_{12} however, are physical. This is due to the energy conserving delta function.[9].

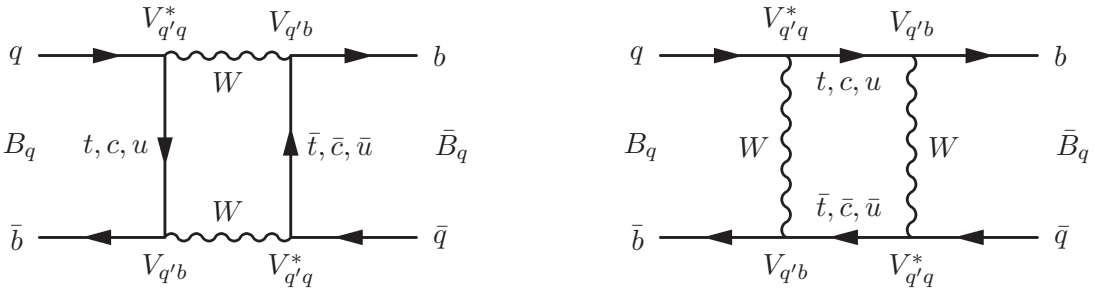


Figure 2.2: The B_q^0 box diagrams. Where $q \in \{d, s\}$ and $q' \in \{t, c, u\}$, which is dominated by the t-quark due to the GIM mechanism (see Appendix A).

[8] M_{12} is often called the dispersive part of $\mathcal{H}_{eff}$. An example of this is the transition amplitude between B_q^0 and $\overline{B_q^0}$, i.e. $B_q^0 \to X \to \overline{B_q^0}$, occurring via some virtual intermediate states (X).

[9] Γ_{12} is often called the absorptive part of $\mathcal{H}_{eff}$. An example of this, outside the B-system, is $K^0 \to n(\pi\pi) \to \bar{K^0}$, where the common state n is real, since both K^0 and $\bar{K^0}$ have enough energy to decay into it.

Since $\mathcal{H}_{eff}$ is not hermitian, its eigenvalues can be complex and can be written in the general form, $M_{a,b} - \frac{i}{2}\Gamma_{a,b}$. Where M_a and M_b are the masses of B_a and B_b respectively, while Γ_a and Γ_b are their decay widths. At this point we can also define the mass and decay width differences of the two states via $\Delta M = M_a - M_b$, $\Delta\Gamma = \Gamma_b - \Gamma_a$. The labels a and b carry no physical meaning and as such, the signs of ΔM and $\Delta\Gamma$ are arbitrary. However their relative sign has physical significance, it dictates whether it is the heaviest or the lightest state which lives longer. In the B_s-system, it has become customary to assign the mass of the eigenstates to these labels, such that: a = H and b = L for the heavy and light eigenstates respectively. Then ΔM = M_H - M_L is positive by definition and sign of $\Delta\Gamma_s$ has to be determined by experiment: combined measurements indicate $\Delta\Gamma_s$ to be positive, $0.138^{+0.068}_{-0.074}$ps [19]. Although this result is not conclusive, it is customary to set $\Delta\Gamma_s$ = Γ_L - Γ_H.

The eigenstates of $\mathcal{H}_{eff}$ are the $\mathcal{CP}$ eigenstates of the system which can be denoted by $|B^0_{L,H}\rangle$ with complex eigenvalues are:

$$\mathcal{H}_{eff}|B^0_{L,H}\rangle = \lambda_{L,H}|B^0_{L,H}\rangle, \quad \lambda_{L,H} \equiv M_{L,H} - i\frac{\Gamma_{L,H}}{2}, \tag{2.51}$$

with $M_{L,H}$ and $\Gamma_{L,H}$ the mass and decay widths. Their time dependence takes the form,

$$|B^0_{L,H}(t)\rangle = e^{-i(M_{L,H} \pm \frac{1}{2}\Gamma_{L,H})t}|B^0_{L,H}(0)\rangle. \tag{2.52}$$

The $\mathcal{CP}$ eigenstates can be written more generally as,

$$|B^0_L\rangle = p|B^0_q\rangle + q|\overline{B^0_q}\rangle, \quad |B^0_H\rangle = p|B^0_q\rangle - q|\overline{B^0_q}\rangle, \tag{2.53}$$

by introducing the complex numbers p and q (normalised by $|p|^2 + |q|^2$ = 1), which represent the amount of meson state mixing. The relationship between the mixing parameters (p and q) and the matrix elements of $\mathcal{H}_{eff}$ in the flavour eigenbase can be expressed through [20]:

$$\frac{q}{p} = e^{in\pi}\left(\frac{H_{21}}{H_{12}}\right)^{1/2} = \left(\frac{M^*_{12} - \frac{i}{2}\Gamma^*_{12}}{M_{12} - \frac{i}{2}\Gamma_{12}}\right)^{1/2}, \tag{2.54}$$

where n denotes the sign of the square root. The time evolution of the $|B^0_q\rangle$, $|\overline{B^0_q}\rangle$ meson states can be obtained by rearranging equation (2.43) and substituting in the time dependent expressions of the $|B^0_{L,H}\rangle$ states. For instance, the time evolution of the $|B^0_q\rangle$ state is given by:

$$\begin{aligned}
|B^0_q(t)\rangle &= \frac{1}{2p}\left[|B^0_L(t)\rangle + |B^0_H(t)\rangle\right] \\
&= \frac{1}{2p}\left[e^{-i(M_L - \frac{1}{2}\Gamma_L)t}|B^0_L(0)\rangle + e^{-i(M_H - \frac{1}{2}\Gamma_H)t}|B^0_H(0)\rangle\right] \\
&= \frac{1}{2}\left[e^{-i(M_L - \frac{i}{2}\Gamma_L)t} + e^{-i(M_H - \frac{i}{2}\Gamma_H)t}\right]|B^0_L\rangle \\
&+ \frac{q}{p}\frac{1}{2}\left[e^{-i(M_L - \frac{i}{2}\Gamma_L)t} - e^{-i(M_H - \frac{i}{2}\Gamma_H)t}\right]|B^0_H\rangle.
\end{aligned} \tag{2.55}$$

By letting:

$$\mathcal{F}_{\pm}(t) = \frac{1}{2}\left[e^{-i(M_L - \frac{i}{2}\Gamma_L)t} \pm e^{-i(M_H - \frac{i}{2}\Gamma_H)t}\right],$$

the expressions for the time evolution of our physical states entering the B_q^0-$\overline{B_q^0}$ system simplifies to:

$$|B_q^0(t)\rangle = \mathcal{F}_+(t)|B_q^0\rangle + \frac{q}{p}\mathcal{F}_-(t)|\overline{B_q^0}\rangle, \quad |\overline{B_q^0}(t)\rangle = \mathcal{F}_+(t)|\overline{B_q^0}\rangle + \frac{p}{q}\mathcal{F}_-(t)|B_q^0\rangle. \tag{2.56}$$

These expression will be used again in section 2.4 when describing the different types of $\mathcal{CP}$ entering the B_s system.

2.4 $\mathcal{CP}$ violation in the neutral meson system

The purpose of this section is to describe how $\mathcal{CP}$ violation enters into the neutral meson system. A complication arises when studying $\mathcal{CP}$ violation in the neutral meson system, due to the presence of complex phases. In quantum mechanics in general, complex phases are rather elusive since they are usually completely arbitrary and irrelevant for any single state. Thus pinning down which phases are physically relevant and which are not, in the neutral meson system, is usually not simple.

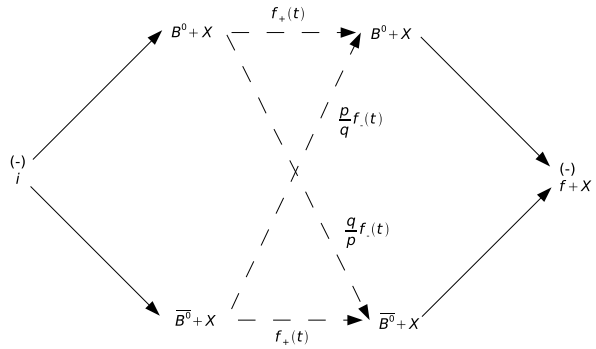


Figure 2.3: Schematic description of the decay chain $i \to B_q + X \to f + X$. The initial state, i, takes into account the production of the neutral meson system through strong and weak interactions.

The neutral mesons can exhibit a rich phenomenology of $\mathcal{CP}$ violation, indeed they are one of only a few systems in which this phenomenon has been observed so far[10]. In general we find three types of $\mathcal{CP}$ violation entering the evolution and decay of the neutral meson system. To quantify these $\mathcal{CP}$ violating effects, we need to find parameters which are independent of any

[10]Hints for new physics have been found in modes such as $B^+ \to J/\psi K^+$ and $B^+ \to K^+\pi^0$ [21]

phase convention. The different types of $\mathcal{CP}$ violation and the parameters used to measure their effect will now be discussed. We begin by discussing the general decay chain $i \to B_q + X \to f + X$ shown in Figure 2.3. Here a neutral B meson system originates from an initial state i, which accounts for production by both strong and weak interactions, and evolves in time before decaying into a final state f. X in this instance refers to the set of particles produced in association with the neutral B meson. The amplitude of this decay chain (and its $\mathcal{CP}$ conjugate) depends on the amplitudes of the initial processes, namely:

$$\begin{aligned} \mathcal{A}_{i\to B_q} &\equiv \langle B_q|\mathcal{H}^{eff}|i\rangle, \quad \mathcal{A}_{\bar{i}\to B_q} \equiv \langle B_q|\mathcal{H}^{eff}|\bar{i}\rangle, \\ \mathcal{A}_{i\to \overline{B_q}} &\equiv \langle \overline{B_q}|\mathcal{H}^{eff}|i\rangle, \quad \mathcal{A}_{\bar{i}\to \overline{B_q}} \equiv \langle \overline{B_q}|\mathcal{H}^{eff}|\bar{i}\rangle, \end{aligned} \tag{2.57}$$

and on the parameters describing the time evolution of the neutral B_q system. This includes $\frac{q}{p}$ and also on the amplitudes for the decay into the final states,

$$\begin{aligned} \mathcal{A}_f &\equiv \langle f|\mathcal{H}^{eff}|B_q\rangle, \quad \mathcal{A}_{\bar{f}} \equiv \langle \bar{f}|\mathcal{H}^{eff}|B_q\rangle, \\ \bar{\mathcal{A}}_f &\equiv \langle f|\mathcal{H}^{eff}|\overline{B_q}\rangle, \quad \bar{\mathcal{A}}_{\bar{f}} \equiv \langle \bar{f}|\mathcal{H}^{eff}|\overline{B_q}\rangle. \end{aligned} \tag{2.58}$$

All these states maybe redefined by an arbitrary phase transformation, which will change the mixing parameters and the transition amplitudes. However, the magnitude of the transition amplitudes and $\frac{q}{p}$ will be invariant under these arbitrary phase transformations. Using these magnitudes we can define phase invariant (thus physical) quantities, which describe the mixing and the transitions occurring in Figure 2.3:

$$\lambda_f \equiv (\pm)\frac{q}{p}\frac{\bar{\mathcal{A}}_f}{\mathcal{A}_f}, \quad \lambda_{\bar{f}} \equiv (\pm)\frac{q}{p}\frac{\bar{\mathcal{A}}_f}{\mathcal{A}_f}, \tag{2.59}$$

$$\eta_{i\to B_q} \equiv \frac{\mathcal{A}_{i\to\overline{B_q}}}{\mathcal{A}_{i\to B_q}}\frac{p}{q}, \quad \eta_{\bar{i}\to B_q} \equiv \frac{\mathcal{A}_{\bar{i}\to\overline{B_q}}}{\mathcal{A}_{\bar{i}\to B_q}}\frac{p}{q}. \tag{2.60}$$

Parameters given in equation (2.59) describe the interference between mixing in the $B_q - \overline{B_q}$ system and the subsequent decay from that system into the final states f and $\bar{f}$ respectively. In contrast, the parameters given in equation (2.60), describe the interference between the production and the mixing in the $B_q - \overline{B_q}$ system.

Using these phase invariant observables, we can classify the three types of $\mathcal{CP}$ violation occurring in these systems. Each type will be discussed in turn.

1. $\mathcal{CP}$ in the mixing

This occurs when the physical states do not coincide with $\mathcal{CP}$ eigenstates, such that,

$$\left|\frac{q}{p}\right| \neq 1. \tag{2.61}$$

This implies that virtual transitions, $B_q \rightarrow \overline{B_q}$, induced by the flavour-changing part of the Hamiltonian (M_{12} and Γ_{12}), have a different probability with respect to the $\mathcal{CP}$ conjugate transition $\overline{B_q} \rightarrow B_q$. This is shown for the neutral kaon system in Figure 2.4.

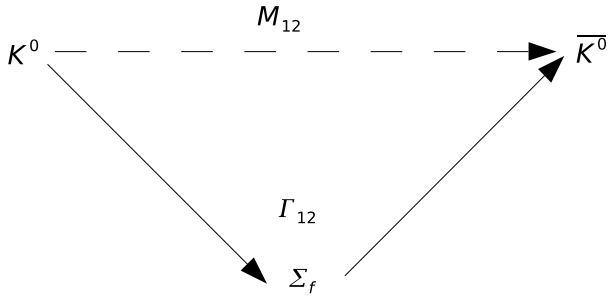


Figure 2.4: Schematic of $\mathcal{CP}$ violation through mixing in the neutral kaon system.

2. $\mathcal{CP}$ in the decays

This occurs when the physical decay amplitudes for $\mathcal{CP}$ conjugate processes into final states f and $\bar{f}$ are different in modulus,

$$\left|\frac{\overline{A}_{\bar{f}}}{A_f}\right| \neq 1. \tag{2.62}$$

This requires the presence of at least two interfering decay amplitudes, as shown in Figure 2.5 for the neutral kaon system, with different weak and strong phases. This type of $\mathcal{CP}$ violation is the only type possible for charged particles, which are forbidden to mix under charge conservation.

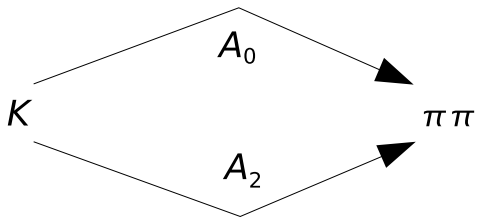


Figure 2.5: Schematic of direct $\mathcal{CP}$ violation in neutral kaon decay into two pions. Where A_0 is the amplitude for $K^0 \rightarrow \pi^0\pi^0$ and A_2 is the amplitude of $K^0 \rightarrow \pi^+\pi^-$.

3. $\mathcal{CP}$ in the interference of mixing and decay

This is $\mathcal{CP}$ which arises from the interference between mixing in the neutral meson system and its subsequent decay into the final states f and $\bar{f}$. This is true even if the amplitudes for both flavour oscillations and physical decays have the same magnitude for the $\mathcal{CP}$-conjugate states,

$$\left|\frac{q}{p}\right| = 1 \quad \text{and} \quad \left|\frac{\overline{A}_f}{A_f}\right| = 1. \tag{2.63}$$

This type of $\mathcal{CP}$ violation, shown in Figure 2.6, occurs through $B_s \to J/\psi\phi$ decays when measuring $\sin(2\beta_s)$.

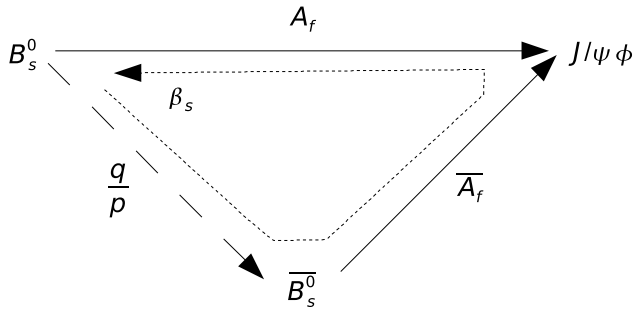


Figure 2.6: Schematic of $\mathcal{CP}$ violation through the interference of mixing and decay in $B_s \to J/\psi\phi$.

In Figure 2.6, the final state can be reached via the decay amplitude $\mathcal{A}_{B\to f} \equiv \langle f|\mathcal{H}^{eff}|B_q\rangle$ and also through the process in which B_q first oscillates to $\overline{B_q}$ before decaying into f, namely $\mathcal{A}_{B\to\bar{B}\to f} = \langle f|\mathcal{H}^{eff}|\overline{B_q}\rangle\langle\overline{B}_q|\mathcal{H}^{eff}|B_q\rangle$. In the second case the two processes act together so that any relative phase between the corresponding amplitudes is important. The processes in which the meson does or does not mix before decaying cannot be distinguished and can therefore interfere in the overall amplitude. Two interfering amplitudes are therefore always present, which can be identified as the meson and anti-meson decay amplitudes, for the evolving coherent mixture of B_q and $\overline{B_q}$. The overall decay amplitude is then,

$$\mathcal{A}_f = \langle f|\mathcal{H}^{eff}|B_q\rangle + \langle f|\mathcal{H}^{eff}|\overline{B_q}\rangle\langle\overline{B_q}|\mathcal{H}^{eff}|B_q\rangle. \tag{2.64}$$

Where the phase for the mixing term in equation (2.64) is given by θ_M and the phase for the decay terms by θ_f ($\overline{\theta_f}$) for B_q ($\overline{B_q}$) decay amplitudes into f. The relative phase in equation (2.64) can then be expressed as $\theta_f - \overline{\theta_f} - \theta_M$.

For the $\mathcal{CP}$-conjugate process in which an initial $\overline{B_q}$ is considered. The phase difference[11] is then given by $\overline{\theta_f} - \theta_f - \overline{\theta_M}$. For this $\mathcal{CP}$ conjugate process it is possible to equalise either the

[11] Where in the phase difference $\overline{\theta_M}$ is the phase of the $\overline{B_q} \to B_q$ amplitude.

mixing phases ($\theta_M = \overline{\theta_M}$) and/or the decay phases ($\theta_f = \overline{\theta_f}$) [12]. This means that the decay of an initial B_q and an initial $\overline{B_q}$ into the final state f can have a different phase difference between the direct decay amplitude and the oscillation amplitude.

$\mathcal{CP}$ through the interference of mixing and decay therefore arises from the phase mismatch between the mixing and decay amplitudes. It can be measured through the physically observable complex quantity λ_f. For final states which are also $\mathcal{CP}$ eigenstates,

$$\mathcal{CP}|f\rangle = |\bar{f}\rangle = \eta_{\mathcal{CP}}(f)|\bar{f}\rangle, \quad \text{where } \eta_{\mathcal{CP}}(f) = \pm 1, \tag{2.65}$$

λ_f can be written as[13],

$$\lambda_f = \eta_{\mathcal{CP}}(f)\frac{q}{p}\frac{\overline{\mathcal{A}}_f}{\mathcal{A}_f}. \tag{2.66}$$

All three types of $\mathcal{CP}$ violation can be expressed through λ_f.

2.5 Decay into a $\mathcal{CP}$ eigenstate $B_s \to J/\psi\phi$

The $B_s \to J/\psi\phi$ decay is topologically equivalent to that of $B_d \to J/\psi(\mu^+\mu^-)K_s$ decays, with s quark replacing the d quark. While $B_d \to J/\psi(\mu^+\mu^-)K_s$ decays measures the relevant phase in $B_d - \overline{B_d}$ mixing (as discussed in section A.3), the decay $B_s \to J/\psi\phi$ will measure the $B_s - \overline{B_s}$ mixing phase.

It is possible to find different and sometimes confusing definitions of the B_s mixing phase, measured through $B_s \to J/\psi\phi$ decays, depending on: author, experiment, sign convention(s) and possible new physics scenarios. The term ϕ_s for the observable $\mathcal{CP}$ violating phase has often been used [6], however this definition suffers some ambiguity in the literature. For instance, ϕ_s is also used for the observable $\mathcal{CP}$ phase as, $\phi_s = \arg(-\frac{M_{12}}{\Gamma_{12}})$ [22] [14]. It is also possible to see the expression, $\phi_s = \arg(M_{12})$, used for the $\mathcal{CP}$ violating mixing phase in the B_s-system [23, 24].

In this thesis we avoid using the term ϕ_s to denote the weak B_s mixing phase, instead we define it to be $-2\beta_s$, the convention adopted by CDF in [5]. In order to measure the $\mathcal{CP}$ violation occurring from this decay, we need to consider the convention dependent (unobservable) and convention independent (observable) phases occurring in this decay[15]. Where the convention dependent phases include: $\arg(M_{12})$ the phase of the off-diagonal mixing element M_{12} (see Appendix A) and ϕ_T the phase originating from the tree level decay in $B_s \to J/\psi\phi$, as illustrated

[12] With the 'and' condition holding only if $\mathcal{CP}$ symmetry holds.

[13] Choosing the convention that η_{CP}=1 in equation (2.39)

[14] Where The value of ϕ_s is calculated to be (4.2 $\pm$ 1.4)$\cdot 10^{-3}$, which implying the values of $\arg(M_{12})$ and $\arg(\Gamma_{12})$ are close to each other.

[15] Throughout this chapter the phase convention adopted by the PDG [7] has been used.

in Figure 2.7. Convention independent phases are invariant under the particular convention used to rotate away the phases differences found in quark fields. Phases of this type generally involve the product of four CKM elements; the unitary angle β_s is one such observable phase.

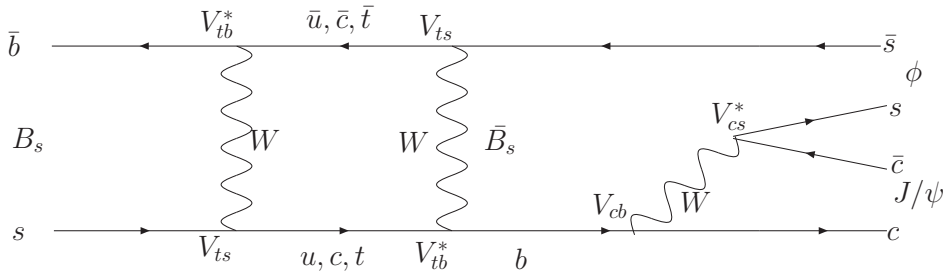


Figure 2.7: Shown on the left is mixing diagram for B_s going to $\overline{B_s}$, while on the right is one of the decay diagrams, in this instance giving $\mathcal{A}_{J/\psi\phi}$.

The angle β_s occurs in the squashed (sb) triangle as shown in Figure 2.1. In the SM, β_s is doubly Cabbibo suppressed [7, 25] and from equation (2.35) is ~ 0.02 radians or $\mathcal{O}(1^o)$. Within the normal phase convention introduced in equation (2.29), β_s is completely determined by the CKM element V_{ts} such that[16],

$$-2\beta_s^{SM} = \arg(M_{12}) + 2\phi_T. \tag{2.67}$$

In $B_s \to J/\psi\phi$ decays, the complex quantity arising in the interference between mixed and non mixed decays is then given by:

$$\lambda_{J/\psi\phi} \equiv \frac{q}{p}\frac{\bar{\mathcal{A}}_{\overline{\mathcal{F}_{\mathcal{CP}}}}}{\mathcal{A}_{\mathcal{F}_{\mathcal{CP}}}} = \eta_{\mathcal{CP}}\frac{q}{p}\frac{\bar{\mathcal{A}}_{\mathcal{F}_{\mathcal{CP}}}}{\mathcal{A}_{\mathcal{F}_{\mathcal{CP}}}}, \tag{2.68}$$

where $\eta_{\mathcal{CP}} = \pm 1$ denotes the $\mathcal{CP}$ eigenvalue of the final state[17] $\mathcal{F}_{\mathcal{CP}}$. $\mathcal{CP}$ is found to be violated in the B_s-system when there is a relative phase between $\frac{q}{p}$(mixing) and $\frac{\bar{\mathcal{A}}_{\overline{\mathcal{F}_{\mathcal{CP}}}}}{\mathcal{A}_{\mathcal{F}_{\mathcal{CP}}}}$ (decay). From Figure 2.7 and equation (2.68) it is possible to deduce that:

$$\begin{aligned}
\lambda_{J/\psi\phi} &= \left(\frac{V_{tb}^* V_{ts}}{V_{tb} V_{ts}^*}\right)\left(\frac{V_{cb} V_{cs}^*}{V_{cb}^* V_{cs}}\right) \\
&= e^{i\arg(\lambda_{J/\psi\phi})} \\
&= \left[\cos(\arg(\lambda_{J/\psi\phi}) + i\sin(\arg(\lambda_{J/\psi\phi}))\right].
\end{aligned} \tag{2.69}$$

[16] In many phenomenological B_s mixing papers [26, 22, 5], it is common to see V_{ts}^* used instead of V_{ts} in the definition of β_s. This occurs when $\arg\left(-\frac{V_{ts}^* V_{tb}}{V_{cs}^* V_{cb}}\right)^\star = \arg\left(-\frac{V_{cs} V_{cb}^*}{V_{ts} V_{tb}^*}\right) = -\beta_s$.

[17] Such that, $\mathcal{CP}|\mathcal{F}_{\mathcal{CP}}\rangle = \eta_f|\mathcal{F}_{\mathcal{CP}} = \pm|\mathcal{F}_{\mathcal{CP}}\rangle$ and where $\mathcal{A}_{\mathcal{F}_{\mathcal{CP}}}(\bar{\mathcal{A}}_{\overline{\mathcal{F}_{\mathcal{CP}}}})$ are the decay amplitudes of $B_s(\overline{B_s})$ into the final states $|\mathcal{F}_{\mathcal{CP}}\rangle(|\overline{\mathcal{F}_{\mathcal{CP}}}\rangle)$ respectfully.

In the standard phase convention this can be expressed as:

$$\arg(\lambda_{J/\psi\phi}) = \arg(M_{12}) + 2\phi_T, \tag{2.70}$$

In the presence of NP, the observable $\mathcal{CP}$ violating phase entering the $B_s \to J/\psi\phi$ decays can then be expressed as,

$$\begin{aligned} \arg(\lambda_{J/\psi\phi}) &= \arg(\lambda_{J/\psi\phi})^{SM} + \phi_s^{NP} \\ &= -2\beta_s^{SM} + \phi_s^{NP}. \end{aligned} \tag{2.71}$$

where in equation (2.71), the presence of NP entering the off-diagonal mixing element M_{12}, has been parametrised by the term ϕ_s^{NP}. In this thesis the measurable $\mathcal{CP}$ violating phase given in equation (2.71) is expressed as,

$$-2\beta_s = \arg(\lambda_{J/\psi\phi}). \tag{2.72}$$

2.6 Decay rates for B_q^0 into $\mathcal{CP}$ conjugate final states

We have already met the expression for the evolution of flavour eigenstates, B_q and $\overline{B_q}$ given in equation (2.56). The amplitudes of the B_q and $\overline{B_q}$ to decay into some final state f at time t is then,

$$\begin{aligned} \mathcal{A}(B_q(t) \to f) = \langle f|\mathcal{H}_{eff}^{\dagger}|B_q(t)\rangle &= \mathcal{F}_+(t)\mathcal{A}_f^{(q)} + \frac{q}{p}\mathcal{F}_-(t)\overline{\mathcal{A}_f}^q, \\ \mathcal{A}(\overline{B_q}(t) \to f) = \langle f|\mathcal{H}_{eff}|\overline{B_q}(t)\rangle &= \mathcal{F}_+(t)\overline{\mathcal{A}}_f^{(q)} + \frac{p}{q}\mathcal{F}_-(t)\mathcal{A}_f^q, \end{aligned} \tag{2.73}$$

taking the modulus squared of the above amplitudes, we find the corresponding theoretical decay rates (Γ_{the}) of the B_q and $\overline{B_q}$ mesons into the final state f (and the $\mathcal{CP}$ conjugated $\overline{f}$ state) as:

$$\begin{aligned} \Gamma_{the}(B_q(t) \to f) &= N_f|\mathcal{A}_f^q|^2\{|\mathcal{F}_+(t)|^2 + |\lambda_f|^2|\mathcal{F}_-|^2 + 2\Re\left[\lambda_f\mathcal{F}_+^*(t)\mathcal{F}_-(t)\right]\}, \\ \Gamma_{the}(\overline{B_q}(t) \to f) &= N_f|\mathcal{A}_f^q|^2\left|\frac{p}{q}\right|^2\{|\mathcal{F}_-(t)|^2 + |\lambda_f|^2|\mathcal{F}_+|^2 + 2\Re\left[\lambda_f\mathcal{F}_+(t)\mathcal{F}_-^*(t)\right]\}, \\ \Gamma_{the}(B_q(t) \to \bar{f}) &= N_{\bar{f}}|\overline{\mathcal{A}}_{\bar{f}}^q|^2\left|\frac{q}{p}\right|^2\{|\mathcal{F}_-(t)|^2 + \left|\frac{1}{\lambda_{\bar{f}}}\right|^2|\mathcal{F}_+|^2 + 2\Re\left[\frac{1}{\lambda_{\bar{f}}}\mathcal{F}_+(t)\mathcal{F}_-^*(t)\right]\}, \\ \Gamma_{the}(\overline{B_q}(t) \to \bar{f}) &= N_{\bar{f}}|\overline{\mathcal{A}}_{\bar{f}}^q|^2\{|\mathcal{F}_+(t)|^2 + \left|\frac{1}{\lambda_{\bar{f}}}\right|^2|\mathcal{F}_-|^2 + 2\Re\left[\frac{1}{\lambda_{\bar{f}}}\mathcal{F}_+^*(t)\mathcal{F}_-(t)\right]\}, \end{aligned} \tag{2.74}$$

where $\frac{1}{\lambda_{\bar{f}}} = \frac{q}{p}\frac{\bar{\mathcal{A}}_{\bar{f}}}{\mathcal{A}_{\bar{f}}}$ and $N_f = N_{\bar{f}}$ is a normalisation factor arising from kinematics. These are our master formulae [27] from which we can to derive the differential decay rates for $B_s \to J/\psi\phi$. The transition probabilities can be simplified using the approximation $\left|\frac{q}{p}\right|^2 = 1$, i.e. no $\mathcal{CP}$

violation in mixing and ignoring higher order corrections in $\left|\frac{\Gamma_{12}}{M_{12}}\right|$ to give,

$$\begin{aligned}
\Gamma_{the}(B_q(t)\to f) &= N_f|\mathcal{A}^q_f|^2 e^{-\Gamma_q t}\Big[\frac{1+|\lambda_f|^2}{2}\cosh\frac{\Delta\Gamma_q t}{2} \\
&\quad + \frac{1-|\lambda_f|^2}{2}\cos(\Delta M_q t) - \Re\lambda_f \sinh\frac{\Delta\Gamma_q t}{2} - \Im\lambda_f\sin(\Delta M_q t)\Big] \\
\Gamma_{the}(\overline{B_q}(t)\to f) &= N_f|\mathcal{A}^q_f|^2 e^{-\Gamma_q t}\Big[\frac{1+|\lambda_f|^2}{2}\cosh\frac{\Delta\Gamma_q t}{2} \\
&\quad - \frac{1-|\lambda_f|^2}{2}\cos(\Delta M_q t) - \Re\lambda_f \sinh\frac{\Delta\Gamma_q t}{2} + \Im\lambda_f\sin(\Delta M_q t)\},\Big] \\
\Gamma_{the}(B_q(t)\to \bar f) &= N_{\bar f}|\overline{\mathcal{A}}^q_{\bar f}|^2 e^{-\Gamma_q t}\Big[\frac{1+\left|\lambda_{\bar f}\right|^{-2}}{2}\cosh\frac{\Delta\Gamma_q t}{2} \\
&\quad - \frac{1-\left|\lambda_{\bar f}\right|^{-2}}{2}\cos(\Delta M_q t) - \Re\frac{1}{\lambda_{\bar f}} \sinh\frac{\Delta\Gamma_q t}{2} + \Im\frac{1}{\lambda_{\bar f}}\sin(\Delta M_q t)\Big], \\
\Gamma_{the}(\overline{B_q}(t)\to \bar f) &= N_{\bar f}|\overline{\mathcal{A}}^q_{\bar f}|^2 e^{-\Gamma_q t}\Big[\frac{1+\left|\lambda_{\bar f}\right|^{-2}}{2}\cosh\frac{\Delta\Gamma_q t}{2} \\
&\quad + \frac{1-\left|\lambda_{\bar f}\right|^{-2}}{2}\cos(\Delta M_q t) - \Re\frac{1}{\lambda_{\bar f}} \sinh\frac{\Delta\Gamma_q t}{2} - \Im\frac{1}{\lambda_{\bar f}}\sin(\Delta M_q t)\Big].
\end{aligned} \tag{2.75}$$

The corresponding differential decay rate expressions for the $B_s \to J/\psi\phi$ decay are obtained using equations (2.75) and (2.71) leading to,

$$\begin{aligned}
\Gamma_{the}(B_q(t)\to f) &= N_f|\mathcal{A}^q_f|^2 e^{-\bar\Gamma_s t}\Big[\cosh\frac{\Delta\Gamma_s t}{2} - \eta_f\cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} - \eta_f\sin(2\beta_s)\sin(\Delta M_s t)\Big], \\
\Gamma_{the}(\overline{B_q}(t)\to f) &= N_f|\mathcal{A}^q_f|^2 e^{-\bar\Gamma_s t}\Big[\cosh\frac{\Delta\Gamma_s t}{2} - \eta_f\cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} + \eta_f\sin(2\beta_s)\sin(\Delta M_s t)\Big]
\end{aligned} \tag{2.76}$$

where $\bar\Gamma_s = \frac{\Gamma_H+\Gamma_L}{2}$ and $\Delta\Gamma_s = \Gamma_L - \Gamma_H$. Assuming a Standard Model $B_s - \overline{B_s}$ mixing phase and neglecting terms quadratic in $-2\beta_s$ (i.e. $\cos(-2\beta_s)\approx 1$) the above time evolution equations will correspond to the following experimental decay rates:

$$\begin{aligned}
\Gamma_{the}(B_s(t)\to f_e) &= N_{f_e}|\mathcal{A}^q_{f_e}|^2\Big[e^{-\Gamma_L t} - e^{-\bar\Gamma_s t}\sin(2\beta_s)\sin(\Delta M_s t)\Big], \\
\Gamma_{the}(\overline{B_s}(t)\to f_e) &= N_{f_e}|\mathcal{A}^q_{f_e}|^2\Big[e^{-\Gamma_L t} + e^{-\bar\Gamma_s t}\sin(2\beta_s)\sin(\Delta M_s t)\Big], \\
\Gamma_{the}(\overline{B_s}(t)\to f_o) &= N_{f_o}|\mathcal{A}^q_{f_o}|^2\Big[e^{-\Gamma_H t} + e^{-\bar\Gamma_s t}\sin(2\beta_s)\sin(\Delta M_s t)\Big], \\
\Gamma_{the}(B_s(t)\to f_o) &= N_{f_o}|\mathcal{A}^q_{f_o}|^2\Big[e^{-\Gamma_H t} - e^{-\bar\Gamma_s t}\sin(2\beta_s)\sin(\Delta M_s t)\Big]
\end{aligned} \tag{2.77}$$

where f_e, f_o denote the $\mathcal{CP}$ even and $\mathcal{CP}$ odd final states respectively. The effects of flavour tagging, i.e. the identification (or tagging) of the original flavour of the b or $\bar b$ quark in the

detected B_s meson, can now be introduced. Tagging procedures are by no means 100% perfect, and this imperfection results in a dilution of the $\mathcal{CP}$ asymmetry through the $\sin(2\beta_s)$ term. The dilution factor is given by $\mathcal{D} = 1 - 2\omega_{\rm tag}$, where $\omega_{\rm tag}$ is the probability of having a wrong identification ($\omega_{\rm tag}$ = $\frac{1}{2}$ in the case there is no tag). With flavour tagging, the theoretical decay rate expressions (2.76) become more realistic (or measurable), and are defined as:

$$\begin{aligned}\Gamma_{phy}(B_q(t) \to f) &= (1-\omega_{\rm tag})\Gamma_{the}(B_q(t) \to f) + \omega_{\rm tag}\Gamma_{the}(\overline{B_q}(t) \to f),\\ \Gamma_{phy}(\overline{B_q}(t) \to f) &= \omega_{\rm tag}\Gamma_{the}(B_q(t) \to f) + (1-\omega_{\rm tag})\Gamma_{the}(\overline{B_q}(t) \to f).\end{aligned}$$

this leads to:

$$\begin{aligned}\Gamma_{phy}(B_q(t) \to f) &= N_f|\mathcal{A}_f^q|^2 e^{-\bar{\Gamma}_s t}\Big[\cosh\frac{\Delta\Gamma_s t}{2} - \eta_f\cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} - \eta_f\mathcal{D}\sin(2\beta_s)\sin(\Delta M_s t)\Big],\\ \Gamma_{phy}(\overline{B_q}(t) \to f) &= N_f|\mathcal{A}_f^q|^2 e^{-\bar{\Gamma}_s t}\Big[\cosh\frac{\Delta\Gamma_s t}{2} - \eta_f\cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} + \eta_f\mathcal{D}\sin(2\beta_s)\sin(\Delta M_s t)\Big].\end{aligned} \tag{2.78}$$

From equation (2.78), we see that even when $\mathcal{D}$=0 (i.e. when we have no tagging), we still have access to $-2\beta_s$ through the cosine term. This implies that information from untagged events can be used as well. For the SM value of $-2\beta_s$, untagged events are expected to yield a small sensitivity to $-2\beta_s$, this is studied in section 6.4.

The $B_s \to J/\psi\phi$ differential decay rates

The $B_s \to J/\psi\phi$ decay involves a spinless (J=0) pseudo-scalar ($J^{\mathcal{P}} = 0^-$) B_s^0 meson decaying into two vector ($J^{\mathcal{CP}} = 1^{--}$) mesons. In the rest frame of the B_s, it can therefore decay into states with relative orbital angular momentum values $\mathcal{L} = 0,\ 1,\ 2$, due to total spin conservation of this isolated system. The $\mathcal{CP}$ eigenvalues of the $J/\psi\phi$ final state are then given by $\mathcal{CP}(J/\psi\phi)$ = $\mathcal{CP}(J/\psi)\mathcal{CP}(\phi)(-1)^{\mathcal{L}}$ = +1, -1, +1. The final state is therefore an admixture of $\mathcal{CP}$-even and $\mathcal{CP}$-odd final eigenstates, which needs to be disentangled in order to make any $\mathcal{CP}$ violation measurement.

In general, the decay amplitude for the $B_s \to J/\psi\phi$ decay can be expressed in terms of the linear polarisation states of the J/ψ and ϕ vector mesons, $\mathcal{A}(B_s \to J/\psi\phi)$ = $A_0(t)$, $A_\parallel(t)$, $A_\perp(t)$. The amplitude $A_\perp(t)$ results in decays in which the two vector mesons are emitted with relative orbital angular momentum $\mathcal{L}$=1, thus is associated with a $\mathcal{CP}$-odd decay. The amplitudes $A_0(t)$ and $A_\parallel(t)$ are associated with (mixtures of) the $\mathcal{L}$=0 and $\mathcal{L}$=2 decays and are $\mathcal{CP}$-even, assuming the magnitude of the corresponding decay amplitudes is equal for particle and anti-particle.

The final state is then an admixture of $\mathcal{CP}$ eigenstates with three independent polarisation

states normalised such that the decay rate is given by:

$$\Gamma_{the}(t) \propto |A_0(t)|^2 + |A_{\|}(t)|^2 + |A_{\perp}(t)|^2, \tag{2.79}$$

The physical decay rates for a B_s into a $\mathcal{CP}$ even (η_{f_e}=+1) or $\mathcal{CP}$ odd (η_{f_o}=-1) states, following the expressions given in (2.76), are then:

$$\Gamma_{phy}(B_s(t) \to f_e) = N_{f_e}|\mathcal{A}^q_{f_e}|^2 e^{-\bar{\Gamma}_s t}\{\cosh\frac{\Delta\Gamma_s t}{2} - \cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} - \mathcal{D}\sin(2\beta_s)\sin(\Delta M_s t)\},$$
$$\Gamma_{phy}(B_s(t) \to f_o) = N_{f_o}|\mathcal{A}^q_{f_o}|^2 e^{-\bar{\Gamma}_s t}\{\cosh\frac{\Delta\Gamma_s t}{2} - \cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} + \mathcal{D}\sin(2\beta_s)\sin(\Delta M_s t)\}$$

where $f_e \in \{0, \|\}$ and $f_o \in \{\perp\}$. In a similar fashion we find the observed amplitudes for a $\overline{B_s} \to f_{e,o}$ transition are:

$$\Gamma_{phy}(\overline{B_q}(t) \to f_e) = N_{f_e}|\mathcal{A}^q_{f_e}|^2 e^{-\bar{\Gamma}_s t}\{\cosh\frac{\Delta\Gamma_s t}{2} - \cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} + \mathcal{D}\sin(2\beta_s)\sin(\Delta M_s t)\},$$
$$\Gamma_{phy}(\overline{B_q}(t) \to f_o) = N_{f_o}|\mathcal{A}^q_{f_o}|^2 e^{-\bar{\Gamma}_s t}\{\cosh\frac{\Delta\Gamma_s t}{2} + \cos(2\beta_s)\sinh\frac{\Delta\Gamma_s t}{2} - \mathcal{D}\sin(2\beta_s)\sin(\Delta M_s t)\},$$

The reason for wanting to express the decay rates in terms of the polarisation states given in equation (2.79), is to be left with amplitudes (i.e. moduli squares of the decay rates) which provide a handle on measuring our physical parameters, $\Delta\Gamma_s$, $-2\beta_s$ etc. In section 2.4 it was shown that the $\mathcal{CP}$ eigenstates in the B_s system are very close to the mass eigenstates (or physical eigenstates of well defined lifetime). The observed final states of the $B_s \to J/\psi\phi$ decay are thus to a very good approximation $\mathcal{CP}$ eigenstates, with the $\mathcal{CP}$-odd angular terms ($|A_{0,\|}|^2$) developing in time in accordance with $e^{-\Gamma_L t}$ and the $\mathcal{CP}$-even angular terms (involving $|A_{\perp}|^2$) developing according to $e^{-\Gamma_H t}$. The $B^0_{s,H}$ and $B^0_{s,L}$ states will therefore decay into distinct angular distributions. This will enhance the statistical precision with which the lifetime difference can be measured, since the B^0_s-system the lifetime difference $\Delta\Gamma_s$ ($\sim \mathcal{O}(10\%)$ in SM), is big enough to be measure.

It is now time to discuss the two bases in which the $B_s \to J/\psi\phi$ decay is commonly defined. These are called the helicity and transversity bases respectively. This thesis will exclusively use the tranversity bases. For further details on the helicity basis, the reader is directed to section A.4. Before defining each basis it is useful to introduce some common notation. The B_s meson is usually referred to as the parent, mother or initial state particle. The two vector mesons are referred to as the daughter particles and their decay products as the final state particle: this is because they are long lived and are stable enough to interact with the detector.

The full decay chain to be investigated is $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$. Each of the four final state particles has four degrees of freedom (four components of four-momentum per particle). There are seven particles of well defined mass in the decay, which provides seven constraints from the relation between four momentum and invariant mass. The physics of the B_s

decay in the laboratory frame is translationally invariant, since the spatial momentum of the B is irrelevant to the decay process. The decay is also rotationally invariant owing to the spinless nature of the mother. This means that the decay angular distribution is isotropic with respect to the B_s rest frame. Of the nine remaining degrees of freedom, three are accounted for by the spatial momentum, another three originate from the orientation (Euler angles) of the coordinate system. The three remaining degrees of freedom arise from the decay angles of the final state particles with respect to each other and the mother.

In the tranversity formalism, the $B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay is fixed in the rest frame of the J/ψ. The ϕ flight direction defines the positive x-axis, while the plane of the K^+K^- system defines the $X-Y$ plane with the y-axis orientated such that the momentum of the K^+ is positive: $p_y(K^+) > 0$. The two fold ambiguity in choosing the z-axis is solved by adopting a right-handed co-ordinated system, as shown in Figure 2.8.

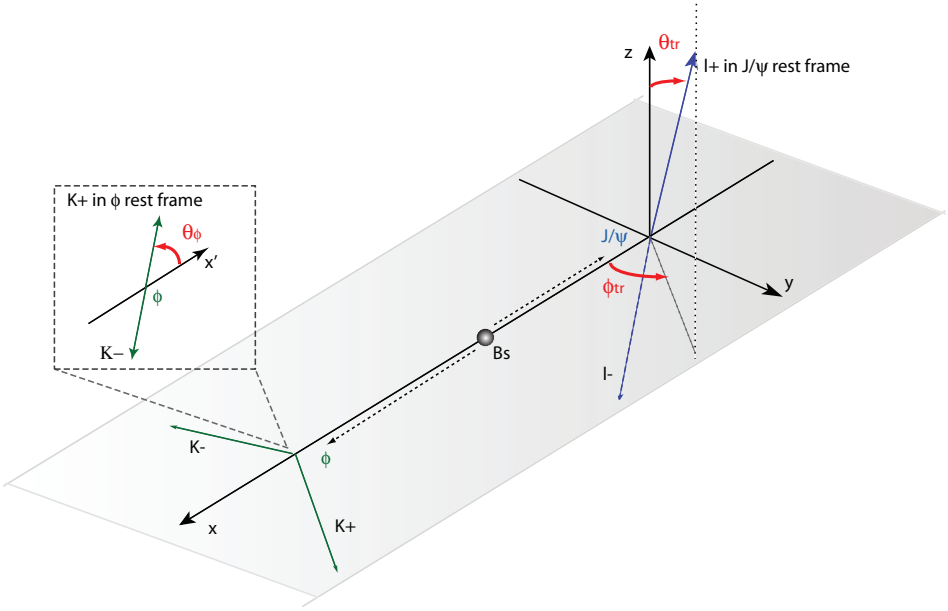


Figure 2.8: Angle definition (Transversity Basis): θ_{tr} is the angle formed by the positive lepton (ℓ^+) in the J/ψ rest frame and the z axis. The angle ϕ_{tr} of ℓ^+ is the azimuthal angle in the same frame. θ_ϕ is the polar angle between the x' axis and the decaying K^+ in the ϕ meson rest frame.

The angle θ_{tr} is defined as the angle between μ^+ and the positive direction of the z-axis. The angle ϕ_{tr} is defined as the positive angle between the x-axis and the projection of the μ^+ onto the $X-Y$ plane. The final decay angle, θ_ϕ is defined as the angle between the K^+ and the x-axis in the rest frame of the ϕ. The description of the decay angles in the tranversity formalism

can be summarised as [28]:

$$\hat{x} = \hat{p}_\phi, \quad \hat{y} = \frac{p_{K^+} - \hat{p}_\phi(\hat{p}_\phi \cdot \hat{p_{K^+}})}{|\hat{p}_{K^+} - \hat{p}_\phi(\hat{p}_\phi \cdot \hat{p}_{K^+})|}, \quad \hat{z} = \hat{x} \times \hat{y}$$

$$\sin\theta_{tr}\cos\phi_{tr} = \hat{p}_{l^+} \cdot \hat{x}, \quad \sin\theta_{tr}\sin\phi_{tr} = \hat{p}_{l^+} \cdot \hat{y}, \quad \cos\theta_{tr} = \hat{p}_{l^+} \cdot \hat{z} \tag{2.80}$$

$$\cos\theta_\phi = -\hat{p}'_{K^+} \cdot \hat{p}'_{J/\psi} \tag{2.81}$$

where the unprimed vectors are all unit three vectors measured in the rest frame of the J/ψ, and the primed quantities are unit vectors measured in the rest frame of the ϕ vector meson.

The physical three angle differential decay rate expressions of an initially produced B_s are obtained from equation (2.79) in terms of the transveristy angles defined in terms of equations (2.80) and (2.81):

$$\frac{d^3\Gamma(t)}{d\cos\theta_{tr}\, d\cos\theta_\phi\, d\phi_{tr}} = \sum_k^6 h^{(k)}(t)\Theta^{(k)}(\theta_{tr}, \theta_\phi, \phi_{tr}),$$

and for an initially produced $\overline{B_s^0}$:

$$\frac{d^3\bar{\Gamma}(t)}{d\cos\theta_{tr}\, d\cos\theta_\phi\, d\phi_{tr}} = \sum_k^6 \bar{h}^{(k)}(t)\Theta^{(k)}(\theta_{tr}, \theta_\phi, \phi_{tr}),$$

where each time h^k and angular $\Theta^k(\theta_{tr}, \theta_\phi, \phi_{tr})$ component is defined in Table 2.2 below and where the full expression for each of the time-dependant amplitudes can be found in section A.5. It is often useful to help reduce the formula given in section A.5, by introducing definitions $R_\perp$ and R_0:

k	$h(t)$	$\bar{h}(t)$	$\Theta(\theta_{tr}, \theta_\phi, \phi_{tr})$
1	$\lvert A_0(t)\rvert^2$	$\lvert\bar{A}_0(t)\rvert^2$	$2\cos^2\theta_\phi(1 - \sin^2\theta_{tr}\cos^2\phi_{tr})$
2	$\lvert A_{\lvert\lvert}(t)\rvert^2$	$\lvert\bar{A}_{\lvert\lvert}(t)\rvert^2$	$\sin^2\theta_\phi(1 - \sin^2\theta_{tr}\sin^2\phi_{tr})$
3	$\lvert A_\perp(t)\rvert^2$	$\lvert\bar{A}_\perp(t)\rvert^2$	$\sin^2\theta_\phi\sin^2\theta_{tr}$
4	$Re\{A_0^*(t)A_{\lvert\lvert}(t)\}$	$Re\{\bar{A}_0^*(t)\bar{A}_{\lvert\lvert}(t)\}$	- $\frac{1}{\sqrt{2}}\sin 2\theta_\phi\sin^2\theta_{tr}\sin 2\phi_{tr}$
5	$Im\{A_{\lvert\lvert}^*(t)A_\perp(t)\}$	$Im\{\bar{A}_{\lvert\lvert}^*(t)\bar{A}_\perp(t)\}$	$\sin^2\theta_\phi\sin 2\theta_{tr}\sin\phi_{tr}$
6	$Im\{A_0^*(t)A_\perp(t)\}$	$Im\{\bar{A}_0^*(t)\bar{A}_\perp(t)\}$	$\frac{1}{\sqrt{2}}\sin 2\theta_\phi\sin 2\theta_{tr}\cos\phi_{tr}$

Table 2.2: Time and angular components in the $B_s \to J/\psi\phi$ decays.

$$R_\perp = \frac{|A_\perp(0)|^2}{|A_\perp(0)|^2 + |A_{||}(0)|^2 + |A_0(0)|^2}, \quad R_0 = \frac{|A_0(0)|^2}{|A_\perp(0)|^2 + |A_{||}(0)|^2 + |A_0(0)|^2}, \quad (2.82)$$

where the magnitude of the amplitudes, as indicated, are taken at time t=0.

By integrating over two of the angles, ϕ_{tr} and θ_ϕ, the expressions for the decay rate simplify considerably, although this will reduce our sensitivity in measuring $-2\beta_s$ (see chapter 6). This reduced differential decay rate expression, in terms of the single angle θ_{tr}, also allows separation of the $\mathcal{CP}$-even and $\mathcal{CP}$-odd components and is given by[18]:

$$\begin{aligned}
\frac{d\Gamma(t)}{d\cos\theta_{tr}} \propto\ & (1-R_\perp)[(1+\cos(2\beta_s))\, e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\, e^{-\Gamma_{\rm H}t} \\
-\ & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\frac{1}{2}(1+\cos^2\theta_{tr}) \\
+\ & R_\perp[(1-\cos(2\beta_s))\, e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\, e^{-\Gamma_{\rm H}t} \\
+\ & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\sin^2\theta_{tr} \qquad (2.83)
\end{aligned}$$

with the corresponding expression for the $\overline{B_s^0}$ decay rate given by:

$$\begin{aligned}
\frac{d\bar{\Gamma}(t)}{d\cos\theta_{tr}} \propto\ & (1-R_\perp)[(1+\cos(2\beta_s))\, e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\, e^{-\Gamma_{\rm H}t} \\
+\ & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\frac{1}{2}(1+\cos^2\theta_{tr}) \\
+\ & R_\perp[(1-\cos(2\beta_s))\, e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\, e^{-\Gamma_{\rm H}t} \\
-\ & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\sin^2\theta_{tr} \qquad (2.84)
\end{aligned}$$

Both the reduced one angular and full three angular expression for the differential decay rates will be used in this thesis to extensively quantify the LHCb's sensitivity to the physics parameters $-2\beta_s$ and $\Delta\Gamma_s$, within the framework of the SM.

2.7 Constraining the CKM picture

The picture of the CKM matrix and the unitarity triangle taken from current experimental and theoretical findings [29, 19] will be given in this section. A summary of the current, as of Summer 2008, experimental status of measurements on the CKM matrix is shown in Figure 2.9. Specifically we will discuss the following set of observables:

$$V_{us}, \quad V_{ud}, \quad V_{ub}, \quad V_{cb}, \quad \Delta M_d, \quad \Delta M_s, \quad \beta_{(d,s)}, \quad \alpha, \quad \gamma.$$

[18] A fuller derivation for the reduced differential decay rate expressions is given in section A.6

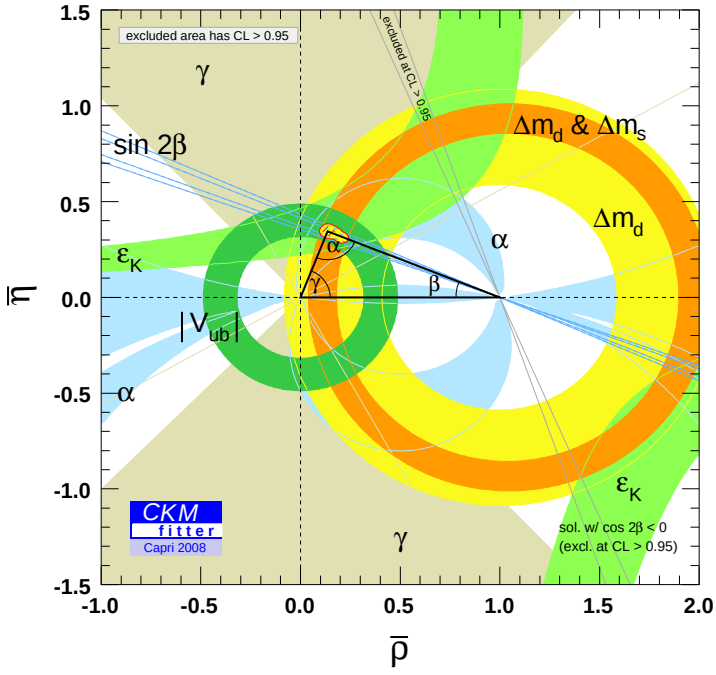


Figure 2.9: Experimental constraints on the ($\bar{\eta}$,$\bar{\rho}$) complex plane from the CKMFitter group [29] taken from the Summer 2008 conference in Capri.

$|V_{us}| \equiv \lambda(0.2228 \pm 0.039 \pm 0.0018)$

The determination of this term is usually taken from the decays $K^+ \to \pi^0 l^+ \nu_l$ and semileptonic K_L^0 decays, $K_L^0 \to \pi^- l^+ \nu_l$. Recent measurements have also used the BF($K^+ \to \pi e^{\pm} \nu$) taken from the Brookhaven National Laboratory (BNL) E865 experiment [29].

$|V_{ud}|(0.9740 \pm 0.0001 \pm 0.0008)$

This term is obtained from a combined likelihood fit of super-allowed nuclear β decays, neutron β-decays, pion β-decays, (which is taken from the branching ratio of BF($\pi^+ \to \pi^0 e^+ \nu_e$)) and the pion lifetime.

$|V_{ub}|(3.90 \pm 0.08 \pm 0.68) \times 10^{-3}$

This term can be extracted either from inclusive $B \to X_\mu l \nu_l$ decays or exclusive $b \to u l \nu_l$ transitions such as $B \to \pi l \nu_l$ or $B \to \rho l \nu_l$. In both the exclusive and inclusive processes, theoretical errors dominate. Particularly in exclusive processes, the errors are dominated by lattice and light-cone sum rules calculations [30, 31]. Given this fact a conservative treatment is adopted when deriving $|V_{ub}|$. The inclusive and exclusive $|V_{ub}|$ central values are averaged with the theoretical error taken as the larger of the two processes.

$|V_{cb}|(42.0 \pm 0.6 \pm 0.8) \times 10^{-3}$

In the Wolfenstein parametrisation $|V_{cb}|$ determines the parameter A, and helps constrain the apex of the unitarity triangle shown in Figure 2.9. Its most accurately obtained from semileptonic $B \to D^* l \bar{\nu}_l$ decays.

$R_t \approx (0.9)$

This parameter involves the ratio $|\frac{V_{td}}{V_{ts}}|$ and can be extracted by making use of the ratio of the B_d and B_s-system mass differences. This is done to minimise the theoretical uncertainties introduced by the hadronic mixing parameters ($\mathcal{F}_{B_q} B_q^{1/2}$). The hadronic parameters in the B_s-system can be re-expressed into B_d-system by introducing the $\mathcal{SU}(3)$-breaking parameter $\xi = \frac{\mathcal{F}_{B_s}\sqrt{B_s}}{\mathcal{F}_{B_d}\sqrt{B_d}}$ which can be calculated accurately in lattice QCD, ξ = 1.16 $\pm$ 0.05 [29]. This then leads to the ratio:

$$\frac{\Delta M_s}{\Delta M_d} \sim \frac{M_{B_s}}{M_{B_d}} \xi^2 \left| \frac{V_{ts}}{V_{td}} \right|^2, \tag{2.85}$$

with the ratio, $\left|\frac{V_{ts}}{V_{td}}\right|$, being determined from the ratio of B_s and B_d mixing and exclusive decay rates $B \to \rho\gamma$ to $B \to K^*\gamma$.

$R_b \approx (0.43 \pm 0.1)$

This parameter corresponds to a circle of radius R_b centred at (0,0) in Figure 2.9. It can be determined from the ratio $|\frac{V_{ub}}{V_{cb}}|$ and can be estimated using the above CKM matrix element results and using:

$$R_b \quad = \sqrt{1 + R_t^2 - 2R_t \cos\beta_d}. \tag{2.86}$$

$\alpha(87.8^{+5.8}_{-5.4})^o$

Although access to this phase cannot be achieved directly. It can be extracted from experimental measurements of $\mathcal{CP}$-violating asymmetries in $B_d - \overline{B_d}$, using $B_d \to \pi^+\pi^-$ decays. Neglecting any pollution coming from penguin diagrams and assuming no $\mathcal{CP}$ in mixing (i.e. $|\frac{q}{p}|$=0), we can extract $\sin(2\alpha)$ from the co-efficient coming from the sinusoidal term given in:

$$\begin{aligned} a_{\mathcal{CP}(t)} \quad &= \tfrac{\Gamma(\bar{B}^0(t)\to\pi^+\pi^- - \Gamma(B^0(t)\to\pi^+\pi^-)}{\Gamma(\bar{B}^0(t)\to\pi^+\pi^- + \Gamma(B^0(t)\to\pi^+\pi^-)}, \\ &= S^{+-}_{\pi\pi} \sin(\Delta_d t) - C^{+-}_{\pi\pi} \cos(\Delta_d t). \end{aligned} \tag{2.87}$$

$\sin(2\beta_d)$ (0.739$\pm$0.048)

This angle has been well measured from $\bar{b} \to \bar{c}c\bar{s}$ quark level transitions in B_d decays; where the phase of the dominate tree-level amplitude is approximately real. The time-dependent $\mathcal{CP}$-

violating parameter measured from the interference between decays with and without mixing, is approximately equal to $\sin(2\beta_d)$. The golden channel for this measurement is $B_d \rightarrow J/\psi(\mu^+\mu^-)K_s$, with the dominated measurement coming from BaBar and Belle [32].

$\beta_s \in$[0.28-1.29] rad

This is the unitarity angle previously shown in Figure 2.1 and defined in section 2.5. The current measurement of $-2\beta_s$ is shown in Figure 2.10 at the 68% confidence level [33]. β_s is the strange counterpart to the β_d angle and can be determined through $\bar{b} \rightarrow \bar{c}c\bar{s}$ quark level transitions in B_s decays. The gold-plated decay for this measurement is $B_s \rightarrow J/\psi\phi$, the focus of this thesis, but other decays can contribute including, $B_s \rightarrow J/\psi\eta^{'}$, $B_s \rightarrow \eta_c\phi$ and $B_s \rightarrow D_sD_s$. The $\mathcal{CP}$ asymmetry for these channels will probe the $\mathcal{CP}$ phase $-2\beta_s \equiv [V_{ts}V_{tb}^*]$ which in the SM corresponds to $\sim -0.0368 \pm 0.0017$ rad.

$\gamma(72^{+34}_{-30})^o$

This angle γ can be constrained most precisely from pure tree decays of charged B mesons of the type $B \rightarrow D^*K^*$ [34]. The angle can then be obtained from the phase carried by V_{ub} through direct $\mathcal{CP}$ effects. The measurement on γ can be further improved using neutral B decays, such as $B_d \rightarrow D^{*\pm}\pi^{\mp}$ and $B_d \rightarrow D^{\pm}\rho^{\mp}$, constraining $|\sin(2\beta_d + \gamma)|$ to 0.37 at the 95% CL.

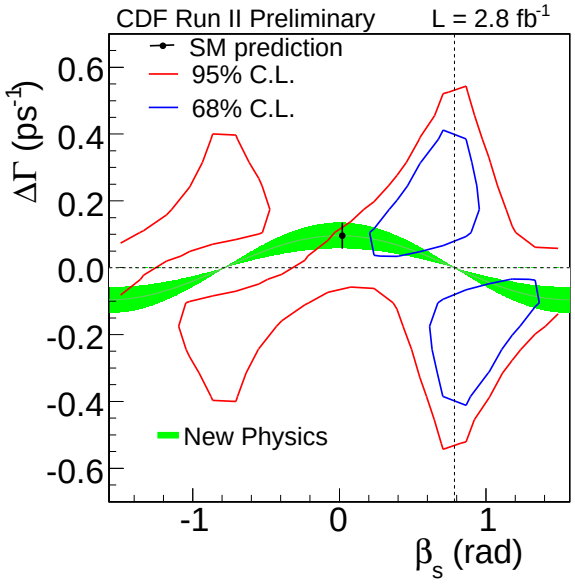


Figure 2.10: The Feldman-Cousins confidence regions for the two dimensional fit for $\Delta\Gamma_s$ and β_s. The plot is taken from CDF's tagged analysis [33] with ~3000 signal $B_s \rightarrow J/\psi\phi$ events. The deviation of $\sim 1.8\sigma$ from the SM prediction is also shown.

2.8 Experimental parameters entering the $B_s \to J/\psi\phi$ decay

We give a brief summary of the currently measured parameters entering the B_s system and $B_s \to J/\psi\phi$ decays. These parameter values listed in Table 2.8, and equations (2.89) and (2.89), will be required when performing the selection of, and fitting to, $B_s \to J/\psi\phi$ events.

	Mass [$\mathrm{MeV/c}$]	average lifetime [ps]	Δm_s [ps^{-1}]	$\frac{\Delta\Gamma_s}{\Gamma_s}$	$\frac{1}{\Gamma_s}$[ps]
B_s	$5367.5^{+0.026}_{-0.027}$	$1.47^{+0.026}_{-0.027}$	17.77±0.1 ±0.07	0.133±0.074	$1.514^{0.034}_{0.037}$

Table 2.3: The neutral B_s-meson parameters, with the values taken from [19].

The fraction of $\mathcal{CP}$-odd and $\mathcal{CP}$-even components entering the differential decay rate expressions given in equation A.42, can also be obtained from recent measurements [5, 35] of the decay polarisation states.

$$\begin{aligned} |A_0(0)|^2 &= 0.530 \pm 0.021(\text{stat}) \pm 0.007(\text{sys}), \\ \left|A_{||}(0)\right|^2 &= 0.23 \pm 0.27(\text{stat}) \pm 0.009(\text{sys}), \\ &\Rightarrow \quad R_\perp = 0.23, \quad R_0 = 0.53. \end{aligned} \tag{2.88}$$

The strong phases, δ_1 and δ_2 (equation (2.89)), have still to be measured accurately from $B_s \to J/\psi\phi$ events and from theory they are currently taken to be 0 and π respectively [27]. However, in current analyses of $B_s \to J/\psi\phi$ data [6] use is made of the $\mathcal{SU}(3)$ symmetry, between the s and d quarks, to extract the value of the strong phases coming from $B_d \to J/\psi(\mu^+\mu^-)K^{\star 0}$ decays [19].

$$\begin{aligned} \delta_1 &= \arg\left(A^*_{||}(0)A_\perp(0)\right) = -0.46 \pm 0.08(\text{stat}) \pm 0.03(\text{sys}), \\ \delta_2 &= \arg\left(A^*_0(0)A_\perp(0)\right) = 2.97 \pm 0.06(\text{stat}) \pm 0.01(\text{sys}). \end{aligned} \tag{2.89}$$

2.9 New Physics and the $B_s \to J/\psi\phi$ decay

The SM provides an excellent effective field theory description of almost all particle physics phenomenology performed thus far. We are capable, for example, of accommodating data from the B-meson and kaon systems within present experimental and theoretical uncertainties (illustration in Figure 2.9). However, this does not mean that NP contributions are absent.The purpose of this final section is to discuss the current status of NP found from $B_s \to J/\psi\phi$ data.

Any NP physics effects may appear at tree level via new interactions, or more likely at the loop level via the exchange of new particles through SM box diagrams as illustrated in Figure 2.7. These effects are especially applicable to those observables which are suppressed in the SM, such as the $B_s - \bar{B_s}$ mixing phase. Indeed, in a large variety of NP models found in the literature [36, 37, 38], a large contribution to NP effects mainly originate from the $B_s - \bar{B_s}$ mixing

amplitude.

Due to the relatively large number of NP models that exist, it is often desirable to parametrise the effects of NP in as model independent way as possible. This can be achieved with a simple Cartesian parametrisation by adding two additional parameters, $\mathbf{Re\Delta_s}$ and $\mathbf{Im\Delta_s}$, such that [29]:

$$\langle B_s|\mathcal{H}^{eff(SM+\mathbf{NP})}_{\Delta B=2}|\bar{B}_s\rangle \equiv \langle B_s|\mathcal{H}^{eff(SM)}_{\Delta B=2}|\bar{B}_s\rangle \times (\mathbf{Re}(\mathbf{\Delta_s}) + i\mathbf{Im}(\mathbf{\Delta_s})), \tag{2.90}$$

here $\mathcal{H}^{SM+\mathbf{NP}}_{\Delta B=2}$ denotes the effective Hamiltonian containing both NP and the SM, whilst $\mathcal{H}^{SM}_{\Delta B=2}$ denotes the contribution from the SM only. As NP effects through $B_s \rightarrow J/\psi\phi$ decays are likely to occur in the SM box diagrams (characterised by the frequency (Δm_s) and $\mathcal{CP}$ violating phase $-2\beta_s$), it is possible to parametrise the NP contribution with $|\Delta_s|\, e^{i\arg(\Delta_s)}$ [29] such that,

$$\begin{aligned} \Delta m_s &= \Delta M^{SM}_{12}\, |\Delta_s|\, e^{i\arg(\Delta_s)}, \\ -2\beta_s &= \arg(\lambda_{J/\psi\phi})^{SM} + \arg(\Delta_s). \end{aligned} \tag{2.91}$$

Here $\Delta_s \geq 0$ measures the relative strength of NP contribution with respect to the SM and $\arg(\Delta_s)$ measures the new $\mathcal{CP}$-violating phase. The SM is therefore located at $|\Delta_s|$=1 and $\arg\Delta_s$=0.

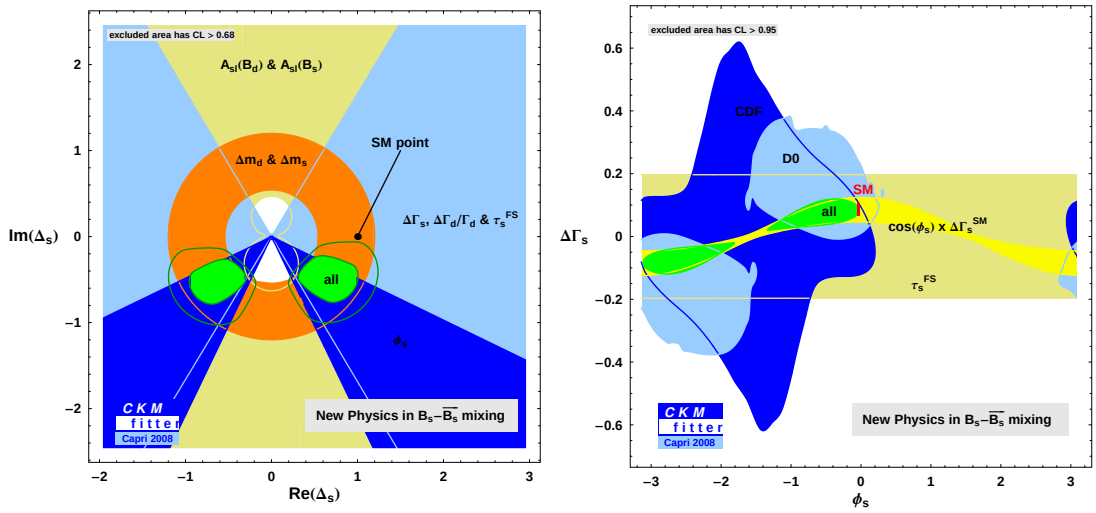


Figure 2.11: Experimental constraints on NP entering $B_s - \bar{B}_s$ mixing from the CKMFitter group [29] taken from the Summer 2008 conference in Capri. Only the 68% confidence interval has been defined. The left-hand plot shows global fit to the NP effects within the ($Re\Delta_s$, $Im\Delta_s$) plane, while the right-hand plot show the effect of NP within the ($\Delta\Gamma_s$, $-2\beta_s$) plane.

Assuming NP contributions to tree level mediated decays are negligible[19], it is then possible to perform a global fit[20] for ($Re\Delta_s$, $Im\Delta_s$), to constrain any NP in the ($\bar{\rho}$, $\bar{\eta}$) plane as shown in the left-hand plot in Figure 2.11, which is taken from the CKMFitter [29] using recent tagged $B_s \to J/\psi\phi$ Tevatron data [39, 40]. In this figure, the dominant[21] parameters constraining NP arise from Δm_s and $-2\beta_s$. A global fit for the parameters given in this plot led to a $\sim 2.5\sigma$ deviation of $-2\beta_s$ from its SM prediction. New physics effects have also been constrained in the ($\Delta\Gamma_s$, $-2\beta_s$) plane, as shown in the right-hand plot in Figure 2.11. Here, the deviation of $-2\beta_s$ from its Standard Model predication (the red line) is found to be 1.9σ. The conclusion is that, at the current level of statistics there is no strong evidence for NP occurring through $B_s \to J/\psi\phi$ decays.

2.10 Conclusion

Different generations of quarks can mix through the CKM matrix (equation (2.9)) and it is this mixing matrix which introduces the $\mathcal{CP}$ violation we see in the quark sector of the SM. The CKM matrix leads to six unitarity triangles (Figure 2.1), and through one of these triangles (the bs triangle) we can define the angle β_s, which is doubly Cabbibo suppressed within the Wolfenstein parametrisation. This angle, due to its definition given in equation 2.34, is related to the B_s system and in particular, the $\mathcal{CP}$ violating phase (equation 2.72) observed in $B_s \to J/\psi\phi$ decays in the SM. This phase, which we denote as $-2\beta_s$, is very small and is well determined in the Standard Model, -0.0368$\pm$0.0017 rad. We can extract our sensitivity to $-2\beta_s$, through $B_s \to J/\psi\phi$ decays using the physical three angular differential decay rate expression defined within the transversity basis (equation (2.82)).

Recent calculations using experimental data from the Tevatron experiments, suggest a deviation of $-2\beta_s$ from its SM prediction is calculated to be $\sim 2.5\sigma$ [29]. This means there is currently no solid experimental evidence for NP occurring in B_s-$\bar{B_s}$ mixing. Over the next few years, it will be the task of the dedicated B physics experiment, LHCb, with its higher statistics to further constrain this picture. In the first year of data taking at an integrated luminosity of $0.5fb^{-1}$, LHCb expects a sensitivity to $-2\beta_s$ of $\pm$0.06 rad [41]. In chapter 3 we introduce the LHCb experiment, describe its different subdetectors, and the experimental software required to reconstruct $B_s \to J/\psi\phi$ events.

[19] This assumption allows the unitarity of the CKM matrix to hold in the presences of NP, allowing the SM contribution to $B_s - \bar{B_s}$ mixing to keep its usual expression as a function of ($\bar{\rho}$, $\bar{\eta}$)

[20] The global fit parameters used by the CKMFitter group include: $(A)_q$, $\Delta\Gamma_q$, ΔM_q, τ_s^{SM} where $q \in \{d, s\}$.

[21] Additional information constraining the NP is brought in by measurements of the semileptonic asymmetries A_{SL}^s (circle) and the width difference $\Delta\Gamma_s = \cos(\phi_s)\Delta\Gamma^{SM}$ (straight line).

3

The LHC and LHCb

3.1 Introduction

The European Organisation for Nuclear Research (CERN) situated on the French-Swiss border is the world's largest particle physics laboratory. Founded in 1954, its primary objective was to unite the European countries (and scientists) after the second world war. Despite the success of the LEP (Large Electron-Positron) collider[1] in confirming the electroweak part of the Standard Model (SM), it could not answer some fundamental questions. For instance: "Where do the fundamental particles gain their mass from?"; "Where does the asymmetry of matter and anti-matter stem from?"; "Is there new physics beyond the SM to allow for the unification of the fundamental forces?". This prompted scientists to build the Large Hadron Collider (LHC), making use of infrastructure existing from the LEP project: primarily the 27 km-long accelerator ring situated 100 m underground, and old detector caverns to hold the new experiments. An aerial view of the LHC ring is shown in Figure 3.1. One LHC experiment to be housed here will be the Large Hadron Collider beauty (LHCb) experiment, designed to study $\mathcal{CP}$ violation in the B-meson system and to look for physics beyond the SM. From 2008, the LHC will offer physicists a microscope to new physics at the TeV energy scale.

3.2 The Large Hadron Collider

The LHC ring consists of two counter-rotating beams of protons at 7 $\mathrm{TeV/c}$ each, colliding at a rate of 40 MHz. The protons are first accelerated using the Super Proton Synchrotron (SPS) to 450 $\mathrm{GeV/c}$, before being injected into the LHC. The centre-of-mass energy of the colliding protons, which determines the available energy for producing new particles, will thus be $\sqrt{s} = 14\ \mathrm{TeV/c}$. This is the highest energy reached by any accelerator so far[2]. The physics motivation of this high operating energy, $\mathcal{O}(\mathrm{TeV})$, is the expectation of finding the unobserved Higgs particle and to open a large window into the unknown, increasing the chances of finding new phenomena. This high energy scale also motivates the construction of a circular pp collider, as opposed to a e^+e^- machine, which loses more energy due to synchrotron radiation loss[3].

[1] LEP ran from 1989 to 2000

[2] The $\mathrm{p\bar{p}}$ Tevatron collider at Fermilab (USA), operating at $\sqrt{\mathrm{s}} = 2\ \mathrm{TeV/c}$ is currently the highest energy collider.

[3] Energy loss through synchrotron radiation is inversely proportional to the mass of the accelerating particle.

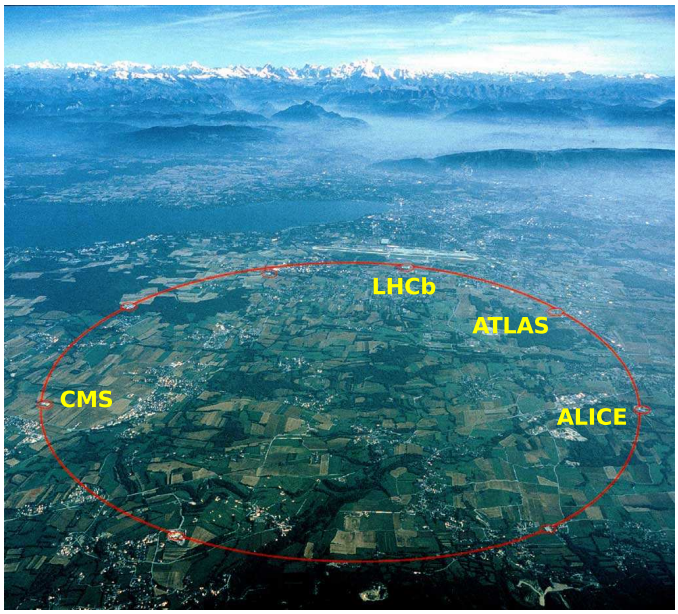


Figure 3.1: Aerial view mapping on the surface of the underground LHC ring, taken from [42]. The position of the four main experiments has been highlighted in yellow.

3.3 The LHC parameters

To keep the energetic protons in their circular orbit requires a strong magnetic field, with a maximal strength of about 8.33 T.[4] This field produced by 1232 superconducting dipole magnets placed into a cryostat containing superfluid helium, cooled to 1.9 K.

The protons travelling around the beam are contained in bunches. There are $\sim$ 2800 bunches (N_{bunches}) each consisting of $\sim 10^{11}$ protons colliding at a rate (ν_{bunches}) of 40 MHz, i.e. proton bunches colliding every 25 ns. The number of pp collisions at each of the four interaction point depends on two quantities:

- The total pp cross-section $\sigma_{pp}^{tot} = \sigma_{pp}^{inel} + \sigma_{pp}^{el}$. The part due to elastic scattering leaves the protons intact which in general are not observed in the detectors. The inelastic part gives rise to a large multiplicity of interactions passing through the detector setup. At $\sqrt{s} = 14TeV$ the expected inelastic cross-section is $\sigma_{pp}^{inel} = 80$ mb.
- The luminosity $\mathcal{L}$ measures the particle flux. It depends on: the compactness of the beam (the emittance), the ability of the magnets to focus the beam at the interaction point (betatron function), the crossing angle and the bunch crossing frequency. The design luminosity of LHC yields $\mathcal{L}$=10^{34} $\text{cm}^{-2}\text{s}^{-1}$.

The number of inelastic pp collisions (N_{pp}) over a given time interval and the mean number of interactions per bunch crossing ($\langle N_{pp} \rangle$) is given by:

[4] This strength is around 139,000 times the strength of the Earth's magnetic field at its magnetic poles.

$$N_{pp} = \sigma_{pp}^{inel} \int \mathcal{L} dt, \qquad \text{and} \qquad \langle N_{pp} \rangle = \frac{\sigma_{pp}^{inel} \mathcal{L}}{\nu_{bx}}. \tag{3.1}$$

Where, ν_{bx} is the nominal LHC bunch crossing frequency of 40 MHz. At a luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ this corresponds to $\sim$ 25 interactions per bunch crossing. This is true for the general purpose detectors, ATLAS and CMS. For LHCb operating at the lower luminosity of 10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, corresponding to a bunch crossing frequency of 40 MHz, the mean number of interactions is ~ 1.

3.4 The LHCb experiment

The Large Hadron Collider beauty experiment (LHCb) is dedicated to $\mathcal{CP}$ violation studies and the study of rare decays in the B-meson system [43].

The LHCb detector has been optimised to detect the decay products of B-mesons. It allows the identification of b-quarks from their decay products with high efficiency and purity. It precisely identifies the b-decay vertex and the corresponding primary interaction vertex. The daughter particles from the decaying B-meson, and the final state particles from possible subsequent decays can be determined by their momentum $\vec{p}$ vector, charge and mass. The charged particles can be precisely tracked from their momentum and charge as they curve in the magnetic field. Neutral and charged particles are stopped and measured in the electromagnetic and hadronic calorimeters.

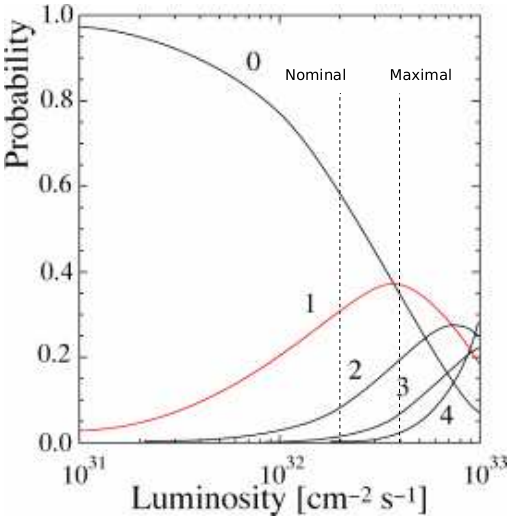


Figure 3.2: Probability of having n=0,1,2,3,4 inelastic interactions as a function of luminosity. The dashed lines represent LHCb's nominal and maximal luminosities.

The number of p-p interactions occurring in a given bunch crossing follows the Poisson distribution:

$$P(\langle N_{pp} \rangle, N_{pp}) = \frac{\langle N_{pp} \rangle^{N_{pp}}}{N_{pp}!} e^{-\langle N_{pp} \rangle} \tag{3.2}$$

The high-multiplicity of primary vertices at the LHC, is not appropriate for the studies performed at LHCb. Therefore, LHCb is operated at a significantly lower nominal luminosity of $\mathcal{L} = 2 \times 10^{32}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which is achieved by defocusing the beams at the interaction point of LHCb. The multiplicity distribution achieved by this measure is shown in Figure 3.2, where the dashed lines represent LHCb's nominal and maximal ($4 \times 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$) luminosities.

At the LHC energies, protons can interact via gluon↔gluon fusion [44], and the distribution of produced $b\bar{b}$ pairs will be strongly peaked in both the forward and backward direction as illustrated in Figure 3.3.

This figure illustrates why LHCb has been designed as a single-arm forward spectrometer. With LHCb's forward design and angular coverage, it manages to cover about 35% of the phase space of the produced b mesons. The expected phase space coverage depends on extrapolating hadronic parameters for b-quark production, to measurements performed by the UA1 experiment (operated on the SPS ring in the 1980's) and the current CDF and D0 experiments at the Tevatron. Expectation for the cross-section $\sigma_{b\bar{b}}$ range from 175 to 950 μb, depending on the extrapolation [44]. The mean value of these estimates, 500 μb, is currently used by all LHC experiments in their Monte Carlo simulations. The true value will need to be measured using LHC data.

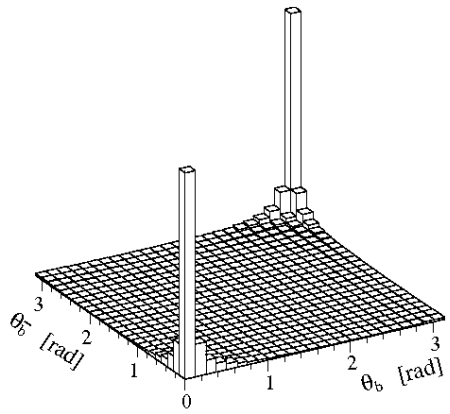


Figure 3.3: Polar angle for b- and $\bar{b}$-hadron directions at the LHC [45].

The LHCb detector is located at LHC interaction point 8, the site of the former DELPHI experiment. A schematic of the LHCb detector as it is fitted into the DELPHI cavern is shown in Figure 3.4. The right-handed coordinate system is determined such that the vertical axis y points upwards and the z-axis points from the interaction point downstream. The main magnetic field component is along the y-axis, thus defining the horizontal or bending plane $x - z$ and the vertical or non-bending plane $y - z$. The angular coverage is 10-300 mrad in the horizontal plane and 10-250 mrad in the vertical plane, where the acceptance is given in terms of the polar angle θ with respect to the z-axis. Equivalently, the acceptance corresponds to a given range in pseudo-rapidity[5] of 1.9 < η < 4.9. The dimensions of the detector are approximately (x=6 m)×(y=5 m)×(z=20 m).

[5] $\eta = -ln\tan(\frac{\theta}{2})$ where θ is the polar angle.

The geometry enables the vertex detectors to be situated much closer to the interaction point, allowing better vertex resolution than for the central detectors. Finally, the forward geometry is more open, allowing for a simpler mechanical design and maintenance.

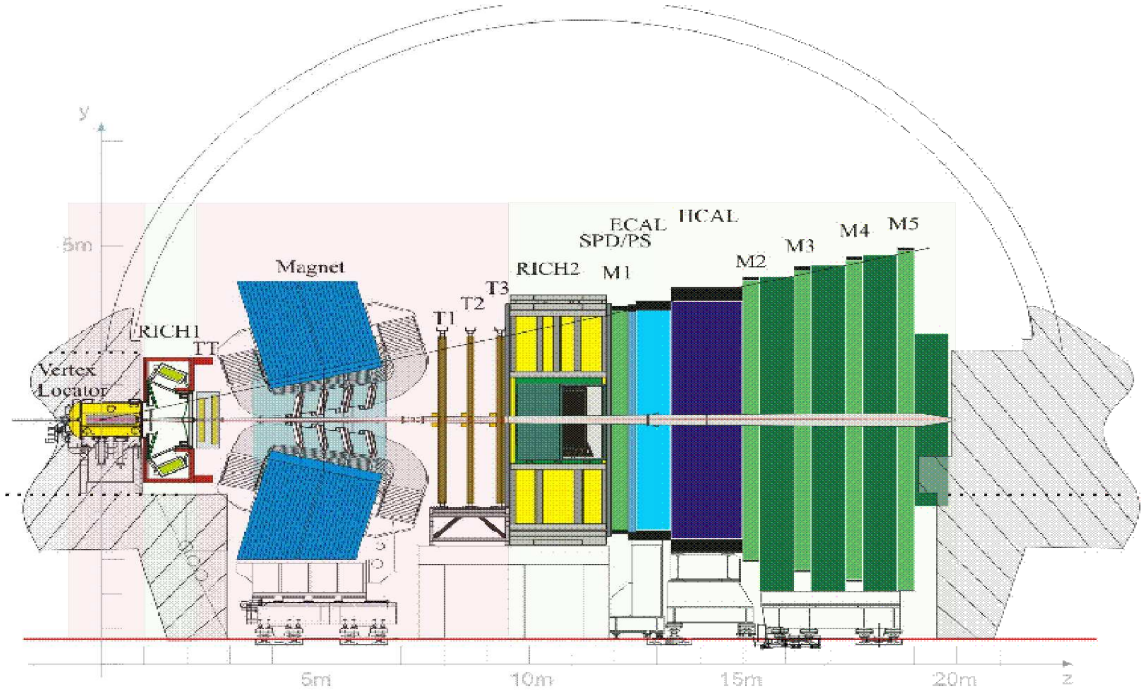


Figure 3.4: The LHCb detector in the vertical (non-bending) plane [43]. The subdetectors can be categorised into tracking devices (highlighted in red) and particle identification (PID) detectors (highlighted in green).

The subdetectors of the LHCb experiment are identified in Figure 3.4. Starting from the interaction point (IP), they are:

- The Vertex Locator (VELO): a high resolution silicon-strip detector to identify the vertices of the primary interaction as well as the decay vertices of short lived particles.
- The first Ring Imaging Cherenkov counter (RICH1): with a large potential acceptance, optimised to identify low momentum particles which may occur at large polar angles.
- The Trigger Tracker (TT): a silicon-strip detector for finding tracks and to provide fast information used in the trigger.
- The warm Magnet: with an integrated magetic field of 4 Tm.

- The main Tracking stations (T1,T2,T3). each consists of a high resolution Inner Tracker (IT), made of silicon-strip detectors, and an Outer Tracker (OT), consisting of straw tubes.
- The second Ring Imaging Cherenkov Detector (RICH2): with a acceptance limited by the opening of the magnet optimised for high momentum particles.
- The first Muon station (M1): Gas Electron Multiplier (GEM) detectors are used in the central part of this station due to the high particle flux, while Multi Wire Proportional Chambers (MWPC) are used for the lower occupancy outer region.
- The Scintillating Pad Detector and Pre-Shower (SPD/PS): using plastic scintillators and lead absorbers, the SPD counts the charged particles prior to electromagnetic showering, while the PS helps separate fast moving pions from electrons.
- The Electromagetic Calorimeter (ECAL): consists of 25 radiation lengths[6] of lead-scintillator readout by scintillator fibres in Shashlik layout.
- The Hadronic Calorimeter (HCAL): made from 5.6 interaction lengths[7] of iron-scintillator titles orientated in beam direction and read out by scintillating fibres.
- And finally the Muon Stations (M2,M3,M4,M5): consisting of 1368 Multi Wire Proportional Chambers (MWPC), providing good transverse momentum resolution of the muons.

The B-mesons produced at LHCb are real and highly boosted in the laboratory frame. With the average lifetime of B_d and the B_s mesons being $\sim$ 1.53 ps and $\sim$ 1.47 ps respectively, these mesons typically travel 10 mm in the laboratory frame before decaying. Due to the production in pairs of b and $\bar{b}$ quarks at the interaction point, the two B-mesons are linked to each other such that information from the second B-meson may help interpret the decay of the first B-meson.

In this thesis we focus on B_s mesons decaying into the $J/\psi\phi$ final state, with the J/ψ resonant state instantaneously decaying into $\mu^+\mu^-$ and the ϕ decaying into K^+K^-. We are interested in reconstructing the two muons and two kaons of the final state. All subdetectors of the LHCb experiment are used for this analysis, but the VELO, the RICH detectors (1 and 2) and the MUON systems are particularly important. Each subdetector will now be discussed in turn.

[6]Defined as the distance traversed in a material over which the electron energy is reduced by a factor of 1/e due to radiation loss only.

[7]Defined as the distance traversed in a material before a particle undergoes an interaction that is neither elastic nor quasi-elastic.

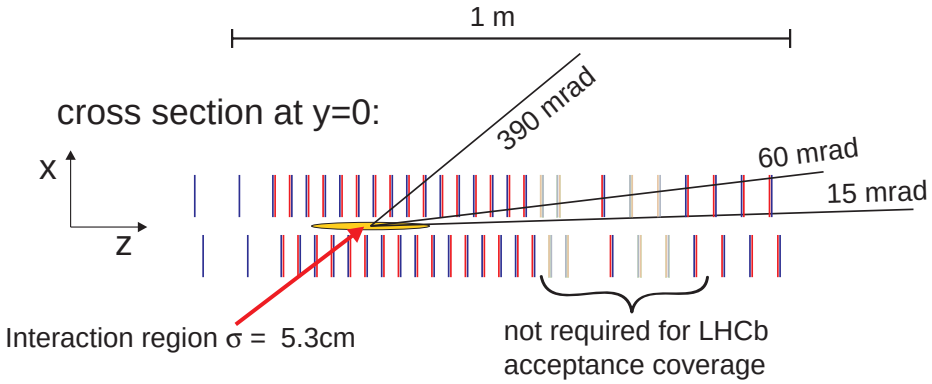
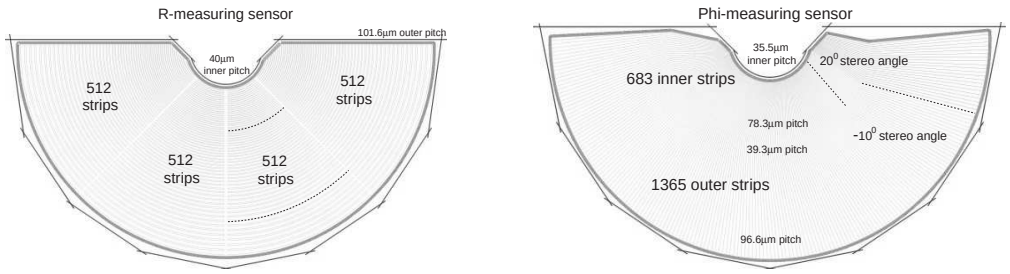


Figure 3.5: Schematic cross-section of the LHCb VELO detector layout [43], viewed from the top at y=0, i.e. in the (x,z) plane. The two radial pile-up veto sensors, left of the interaction point, are shown in blue.

3.4.1 Vertex Locator (VELO)

The VELO [46, 47] consists of 21 main tracking stations. Their arrangements within the LHC beam pipe is schematically shown in Figure 3.5. Each station is split into two semicircular silicon-strip sub-detectors which can be retracted by 3 cm from the proton beam during injection and acceleration. Each semi-circular sub-detector consists of two sensor layers, with circularly oriented strips measuring the radial (r) and the other with radial orientated strips measuring the azimuthal (Φ) coordinate of each track. Each r- and Φ-sensor comprises of 2048 silicon strips, of which the layout is shown in Figure 3.6.



(a) radial (b) azimuthal

Figure 3.6: Layout of the radial (r) and azimuthal (Φ) sensors, with some strips indicated with dotted lines for illustration. The images are taken from [43].

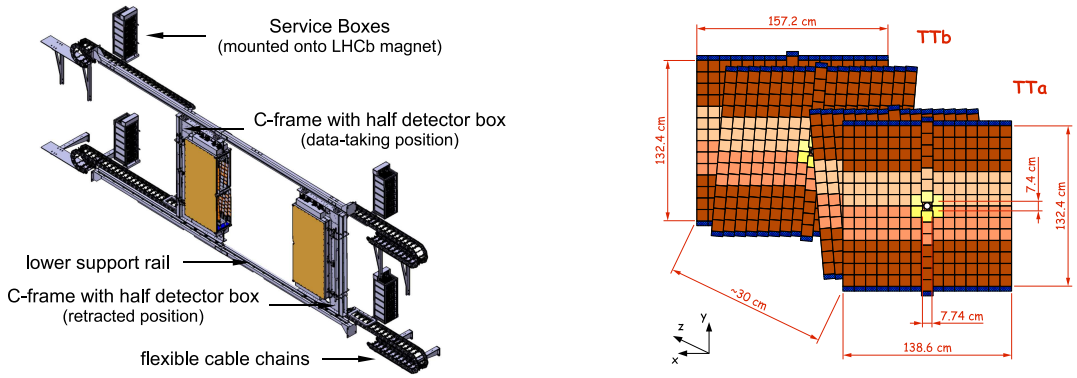
The VELO layout has been optimised to reduce the amount of material in the detector acceptance, whilst providing the required geometrical coverage of the charged tracks. Any charged particle produced within the acceptance region around the interaction point will pass through at least five stations. There a charged particle generally leaves one or two adjacent hits as it passes through a sensor. The resolution with which the tracks can be reconstructed depends on the sen-

sor pitch and projection angle. The best resolution[8] achieved with the VELO is 9 μm in both the r and Φ directions [47].

The VELO's role is of crucial importance for B-physics measurements. In $B_s \rightarrow J/\psi\phi$ decays for example, the B_s lives long enough to measure its lifetime, provided we can separate the secondary vertex, i.e. the B_s decay vertex, from the primary interaction point. This measurement is provided by the VELO with high precision. In addition it provides the information needed to veto pile-up[9] and to identify candidates from displaced vertices. The former is used by the L0 trigger to select clean events, the latter by the High Level Trigger to select B meson events.

3.4.2 The Tracking stations (TT) and T1,T2,T3

Further tracking of charged particles is provided by multiple stations (the red section shown in Figure 3.4) along the path of the particles. Tracking consisting of: the VELO, the Trigger Tracker (TT) and the tracking stations (T1, T2, T3) [48, 49].



(a) Trigger Tracker

(b) TTa and TTb stations

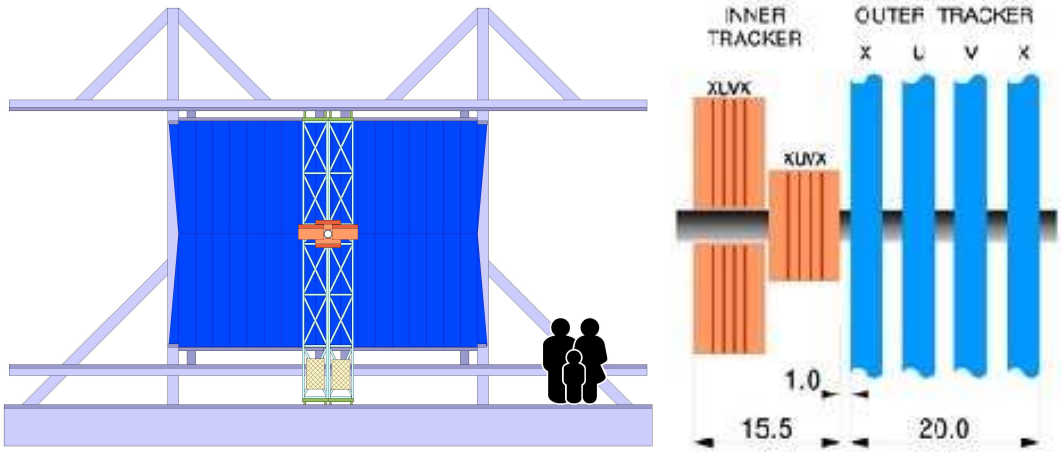
Figure 3.7: (a) Isometric view of the Trigger Tacker (TT) support mechanism. Two half stations, one on either side of the beam pipe, can be moved horizontally. (b) View of the x- and u-layers in TTa and of the v- and x-layers in TTb (the square elements represent the silicon sensors).

The Trigger Tracker consists of four silicon strip detector layers grouped into two stations of two layers each, called TTa and TTb. They are shown in Figure 3.7. The station TTa is centred at z=232 cm and TTb is centred at z=262 cm with a clearance of 27 cm between the two stations. In total both stations cover an active area of $\sim$8.4 m^2. The sensors of each layer are staggered by 1 cm in the z-direction and overlap by a few mm in the x-direction. TTa consists of a x- and u-layer, with the strips in the x-layer running in y-direction to measure the deflection of charged particles in x-direction. The u-layer is rotated with respect to the x-layer by a stereo angle[10]

[8] The worst resolution being 25 μm

[9] Pile-up is multiple primary vertices per bunch crossing.

[10] The stereo angle is defined as the angle between the strips direction and the y axis, in the $x-y$ plane. Positive



(a) tracking substations (b) IT and OT

Figure 3.8: (a) Schematic view, looking from the interaction point, on one of the tracking sub-stations in the $x-y$ plane showing: the beam pipe in the centre, surrounded by the Inner Tracker (IT orange) and the Outer Tracker (OT, straw tubes blue). (b) A cut-through of the centre part of the tracking sub-stations parallel to the beam axis displaying the staggering of the IT and OT layers along the beam pipe. Figures are taken from [48].

of -5° when looking from the interaction point. TTb consists of a x- and a v-layer, with the second rotated by $+5^{\circ}$. Each TTa(TTb) layer is split into 15(17) vertical ladders which in turn are subdivided into several readout sectors. Finer segmentation is given to the central sectors to help improve the trigger performance. These sectors are indicated by various colours in Figure 3.7.

The main purpose of the tracking stations (T1, T2 and T3) is to reconstruct the charged tracks which pass through the opening in the magnet in order to determine their momentum. They also provide an extrapolation of tracks used in the reconstruction of Cherenkov rings by the RICH detectors. Finally they provide seeds for the reconstruction of clusters in the calorimeter and muon chambers. Each of the tracking substations (T1,T2,T3) consists of an Inner Tracker (IT), which covers the central 2% of the area with high track occupancy and an Outer Tracker (OT) covering the remaining 98% of the area with lower occupancy. To cope with the high track multiplicity in the central region, the IT is constructed using the same silicon strip technology as employed for the TT. For each substation the IT covers a total active area of $\sim$1.4m^2 around the beam pipe. The four layers are arranged in (x,u,v,x) stereo orientation, as discussed for the TT.

With increasing polar angles, the track occupancy drops quickly. Below an occupancy of 15% strawtubes can be employed to achieve the designed track resolution [49]. For each substation the OT covers an active area of, $\sim$ 28.9 m^2. The four detection layers are arranged in the

angles occur when rotating from the positive x axis to the positive y axis in the transverse plane.

same (x,u,v,x) stereo geometry as the IT and TT stations. Each detection layer is made from straw drift tubes filled with a gas mixture of Argon (70%) and $\mathrm{C0_2}$ (30%). This guarantees fast drift times and a readout of the signal of below 50 ns [47].

The main purpose of the TT station is to assign momentum information to the charged particles located by the VELO, prior to the deflection by the magnetic field. The charged particles are slightly deflected by the low magnetic field fringe between the VELO and TT which has an integrated bending power of only 0.15 Tm [50]. This small deflection can be used to roughly estimate the momentum of particles with a limited resolution. Particles with a high momentum, with a momentum of a few GeV/c, experience in the bending plane a deviation of only a few millimeters compared to a straight trajectory. But this is sufficient to determine the momentum with 10-40% accuracy without the use of other stations. This information is used in the $\mathrm{L0}$ trigger to determine the tranverse momentum of the high impact parameter tracks. The TT is also used to reconstruct long-lived neutral particles, like the K_s^0, which usually decay outside the VELO acceptance.

3.4.3 Magnet

The LHCb magnet, Figure 3.9, is used to deflect the path of charged particles. It provides an overall bending power of $\int \mathrm{Bdl}$ =4 Tm. With this a momentum resolution of $\frac{\delta p}{p} \sim 0.4\%$ is achieved for high momentum tracks, with a momentum of $\mathrm{p} \sim 40\ \mathrm{GeV/c}$, which are detected in the VELO, the TT and in the tracking stations T1-T3. Its funnel-like geometry is determined by the acceptance requirements of LHCb, its opening allows for a large forward acceptance of $\pm$250 mrad vertically and $\pm$300 mrad horizontally. The magnet is built using the conventional warm technology [45], with saddle-shaped coils in a window-frame iron yoke. The poles are sloped to match the detector acceptance. The use of a warm magnet allows to reverse its polarity. This is regarded as an important systematic cross-check, allowing the study of systematic effects introduced by the residual left-right asymmetries in the detector.

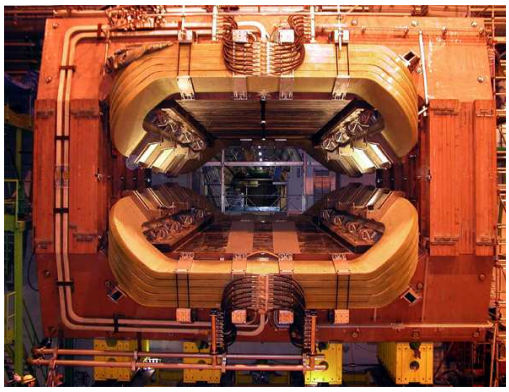
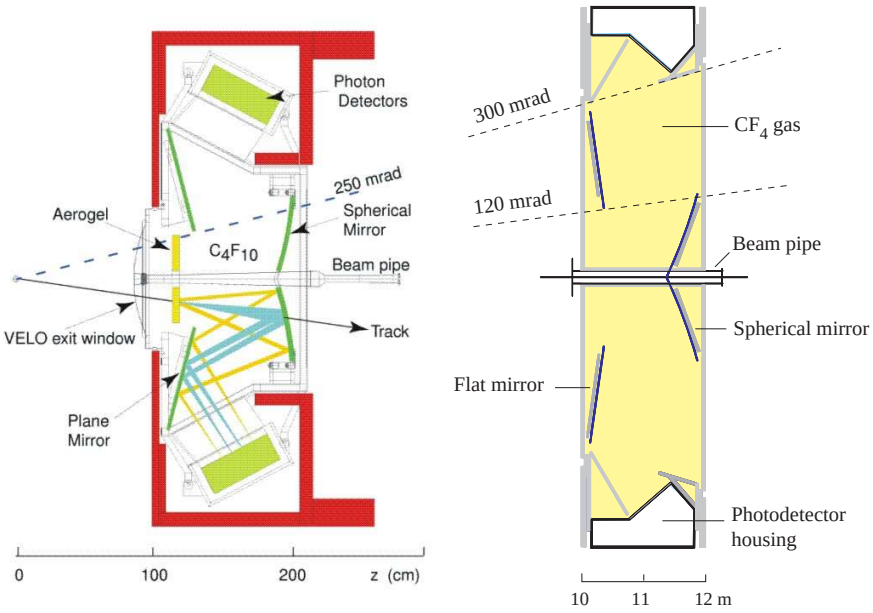


Figure 3.9: The LHCb dipole magnet and surrounding iron yoke, looking long the z-axis back to the interaction point. The photograph is taken from [42].

3.4.4 The RICH system

Particle identification is provided by two Ring Imaging Cherenkov counters (RICH) [43, 47]. They are designed to cover the broad angular and momentum spectrum of charged particles found within the detectors acceptances. RICH1 is located between the VELO and TT before the magnet, and RICH2 is situated between the three tracking stations and the calorimeters behind the bending magnet. RICH1 is designed to identify low to intermediate momentum tracks ranged from ~ 1 GeV/c to 60 GeV/c using an aerogel and a C_4F_{10} radiator, RICH2 provides particle identification from ~ 15 GeV/c up to ~ 150 GeV/c using a CF_4 radiator. RICH1 covers the full LHCb acceptance, while RICH2 covers a limited acceptance of $\sim \pm 10$ mrad to ± 120 mrad ($\pm$ 100 mrad) in the horizontal (vertical) direction.

Particle identification requires knowledge of the particle mass, charge and interaction properties with matter. In Ring Imaging Cherenkov counters, the velocity of charged particles ($\beta = \frac{v}{c}$) is measured. Together with the momentum measurement for the charged tracks the mass of the particles can be deduced. This technique is particularly well suited for high momentum tracks where other particle identification methods fail.



(a) RICH1 (b) RICH2

Figure 3.10: (a) Schematic view of the RICH1 detector from the side and (b) the RICH2 detector from the top. The figure were taken from [51].

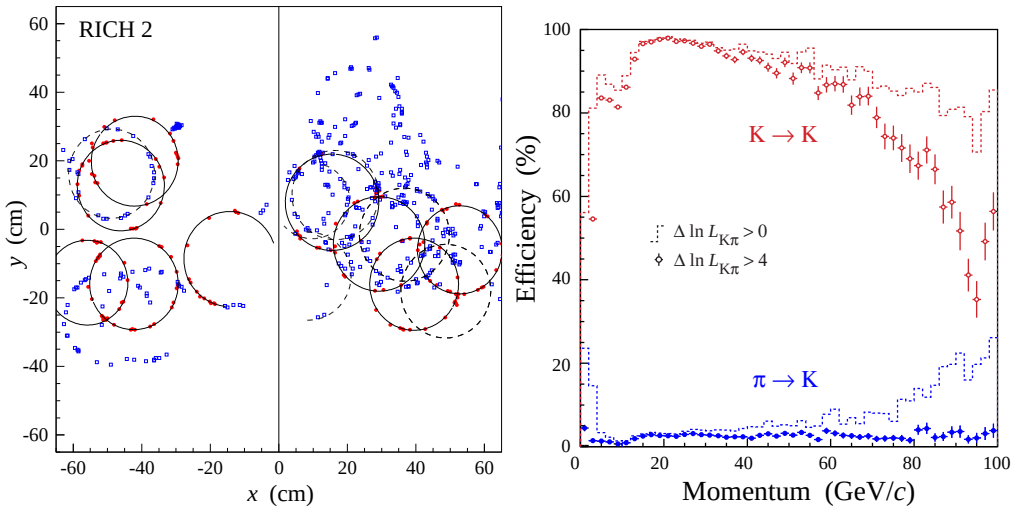
The principle of RICH detectors relies on the Cherenkov effect. Charged particles which travel through a medium with a refractive index $n > 1$, i.e. not in the vacuum, are forced to emit electromagnetic radiation if their velocity is larger than the velocity of light in that medium,

$v = \beta c > \frac{c}{n}$. This is a direct consequence of the Maxwell equations. The light emitted along the track interferes and forms a cone of opening angle $\cos\theta_c = \frac{1}{n\beta}$ with respect to the trajectory. As shown in Figure 3.10, spherical mirrors collect and focus the light into ring images. At LHCb the mirrors are tilted to allow the photodetectors to be placed outside the geometrical acceptance. This distorts the ring images to ellipses. A second set of plane mirrors are employed to increase the lever arm, allowing the photodetectors to be placed compactly in the z-direction. The light is focused on detector planes equipped with Hybrid Photon detectors (HPDs). The HPDs are sensitive to single photons in the range of 200 to 600 nm, with a high Quantum Efficiency[11] ($\langle \mathrm{QE} \rangle$ @ 270 nm = 27%) and covers an active area ratio of $\sim$ 65%. Three radiator media are employed in the RICH system to produce Cherenkov light over a wide range of particle momentum:

Arogel n=1.03, optimal for low momentum particles inside RICH1;

$\boldsymbol{C_4F_{10}}$ n=1.0015, for intermediate momentum particles inside RICH1;

$\boldsymbol{CF_4}$ n=1.00046, for high momentum particles inside RICH2.



(a) Cherenkov rings (b) Kaon identification

Figure 3.11: (a) Detected photon hits of a single event in the two detection planes of RICH2 and the reconstructed ring images associated with detected tracks. (b) The kaon identification efficiency (top curve) and the probability for a pion being misidentified as a kaon (bottom curve) versus the particle momenta. The plots are taken from [45].

[11] The Quantum Efficiency is defined as the probability that a measurable photoelectron is produced when a photon hits the photocathode.

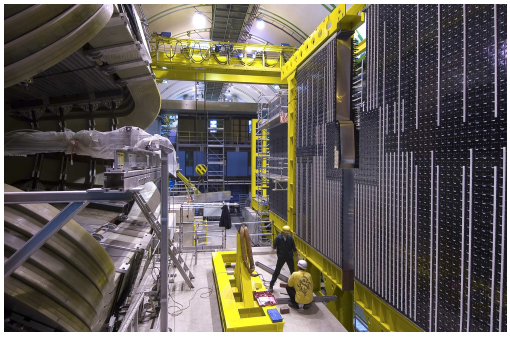
For each event a pattern recognition algorithm reconstructs the ring images as illustrated in Figure 3.11(a) and an association with the tracks found by the tracking system is sought. Rings without a track reconstruction, scattered hits and clusters of noise are treated as background. A maximum likelihood is calculated testing the various particle hypotheses against the measured Cherenkov angle (θ_C), forming the most likely candidate for each ring.

An example of the particle identification (PID) performance is shown in Figure 3.11(b). The average efficiency for kaon identification $\epsilon(K \rightarrow K)$ between 2 and 100 $\mathrm{GeV/c}$ is 95% for a likelihood[12] of $\Delta\mathcal{L}_{k-\pi} > 0$, with a corresponding average pion misidenification rate $\epsilon(\pi \rightarrow K)$ of $\sim$ 5%. Around 30 $\mathrm{GeV/c}$ the identification probability is $\sim$ 97% and the misidentification probability $\sim$ 5%. The efficiency of the two identifications can be traded off by the probability of the misidentification of pions by the variation of the likelihood cut. In the same way the separation power between all pairs of particles can be determined. The overall RICH PID performance using fully simulated Monte Carlo data, is found to be good over the entire momentum range, 2 to 100 $\mathrm{GeV/c}$ [45].

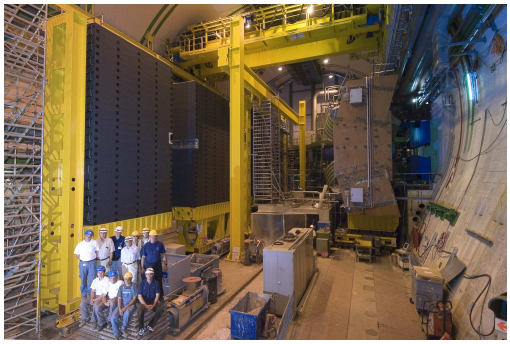
The RICH counters provide the essential particle identification for physics analysis at the LHCb. Especially the separation between charged pions and kaons, which help to distinguish between decay channels with the same topology, such as $B_d \rightarrow \pi^+\pi^-$ which requires rejection of the following two-body backgrounds: $B_d \rightarrow K^+\pi^-$, $B_s \rightarrow K^-\pi^+$ and $B_s \rightarrow K^+K^-$.

[12] A discussion and definition for the delta log likehood function is presented in section 4.3

3.4.5 Calorimetry



(a) ECAL



(b) HCAL

Figure 3.12: (a) The electromagnetic calorimeter (ECAL) during installation opposite the dipole magnet. (b) The hadronic calorimeter (HCAL), during installation. The images where taken from [42].

The main purpose of the calorimeter system is the identification of particles including: hadrons, electrons and photons, by measuring their energies, position and shower profile. The two main calorimeters, illustrated in Figure 3.12, are the Electromagnetic and Hadronic Calorimeters (ECAL and HCAL). Identification of electrons and photons, is primarily achieved by the ECAL. This is essential for flavour tagging semileptonic decays and to enable reconstruction of B decays containing prompt γ or π^0. The system is used to identify events with high transverse energy B-hadron decay products, which is the basis of the L0 trigger decision.

The calorimeter system is placed towards the end of the LHCb detector, between the first and second muon stations, as illustrated in Figure 3.4.

The Scintillating Pad Detector (SPD) and Pre-Shower (PS)

These subdetectors are located in front (SPD) and after (PS) a 12 mm thick lead converter, and just before the main calorimeters. Both consist of 15 mm thick scintillator tiles, with wavelength shifting fibres used to transmit scintillating light to conventional photomultiplier tubes for readout. The SPD counts the number of charged tracks prior to showering. This specifically helps to distinguish between photons and electrons. The lead converter initiates the electromagnetic shower, with the showers initial evolution being picked up by the PS. For 50 $\mathrm{GeV/c}$ moving neutral pion and electron beams, the retention efficiency of the PS is $\sim$ 99.7% and $\sim$ 97% respectively [47]. The PS therefore helps to separate fast moving pions (hadrons) and electrons.

Electromagnetic Calorimeter (ECAL)

The LHCb ECAL is built from individual Shashlik-type modules that are made from 66 lead absorbing plates (2 mm thick) interspaced with scintillator tiles (4 mm thick) as the active material[13]. Its total thickness corresponds to 25 radiation lengths and 1.1 interaction lengths. Wavelength shifting fibres penetrate the lead/scintillator stack through holes and are readout through photomultipliers. The ECAL structure, shown in Figure 3.13, is segmented into three sections (the SPD and PS calorimeters also have the same structure) with one type of module per section. All three module types have identical squares size of 121.2 mm, but differ by the number of readout cells. The ECAL section closest to the beam pipe consists of 167 modules containing 9 readout cells each, the middle section has 448 modules 4 cells each, and the outer section has 2688 modules made from a single cell.

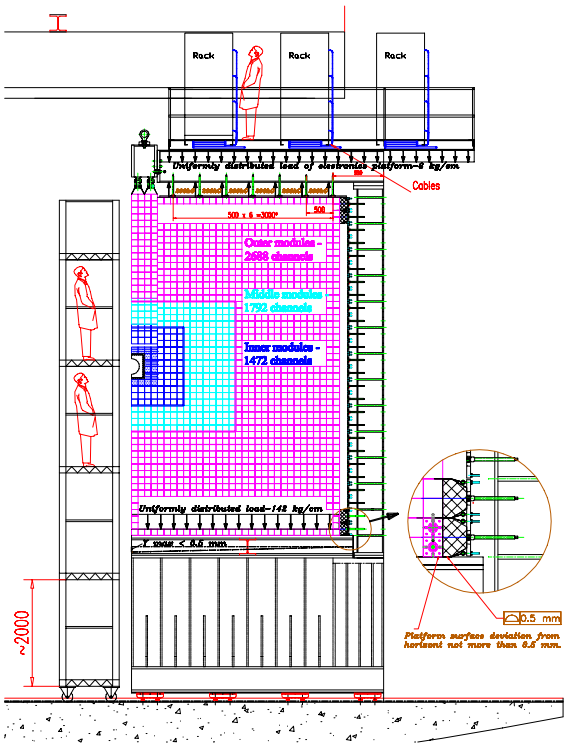


Figure 3.13: Front view of the electromagnetic calorimeter, looking from the interaction point. The electromagnetic calorimeter consists of three sections, the inner section (blue), the middle section (cyan) and the outer section (magenta), which decrease in the number of readout cells moving away from the beam pipe. The image is taken from [52].

The energy resolution of the ECAL can be described by:

$$\frac{dE}{E} = \frac{a}{\sqrt{E}} \oplus b \oplus \frac{c}{E}(\text{E in GeV/c}),$$

where a stands for the stochastic term, due to statistical fluctuations in the shower and b for the constant term, representing the systematics induced by the detector. The c term describes the noise induced by the electronics in the detector. Using one of the outer modules, with a test beam of 50 GeV/c electrons, the terms are measured to be 8.5% < a < 9.5%, $b \sim$ 0.8% and c $\sim$10.1% [53]. The energy resolution for electrons generally varies from $\sim$3% (at 20 GeV/c) to $\sim$1.1% (at 100 GeV/c).

[13]This type of calorimeter use first used for the PHENIX detector at the Brookhaven National Laboratory. The active material is made from polystyrene-based scintillators, with 2.5% PTP and 0.01% POPOP.

Hadronic Calorimeter (HCAL)

The HCAL is placed behind the ECAL and stops hadrons via their interactions with the detector's material. The HCAL is made out of layers of steel (16 mm thick) and scintillating tiles (4 mm thick), as the absorber and active material respectively. The scintillating tiles and absorbers run parallel to the beam, with the overall material thickness of 1.2 meters. The HCAL structure is segmented into two section each being built of 26 detector modules with each module being subdivided into 8 submodules, one half-section is shown in Figure 3.14. Wavelength shifting fibres run along the edges of the scintillator tiles to bring the scintillating light to photomultiplier tubes that are fixed at the back of the structure. The expected energy resolution of the HCAL is $\frac{\sigma(E)}{E} = \frac{80\%}{\sqrt{E}} \oplus 10\%$.

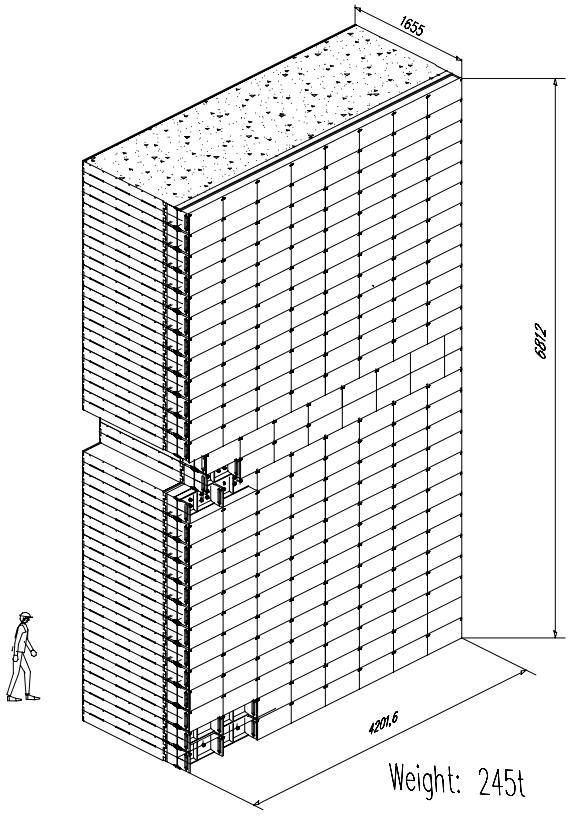


Figure 3.14: View of one half of the hadronic calorimeter. There are 36 horizontal modules stacked on top of each other. Two central modules are shorter to allow the accelerator beam to pass through. The images is taken from [52].

3.4.6 Muon Chambers

The muon detector is divided into 5 tracking stations (M1 to M5). Stations M2-M5 are placed behind the LHCb calorimetry system, which acts as a shield designed to attenuate all photons, electrons and hadrons before reaching the outer muon stations. M1 improves the transverse momentum measurements of the muon track and is in the L0 trigger. Due to the high particle flux received by the inner part of this station, up to $\sim 8.3 \times 10^{-3}/\mathrm{cm}^2$, triple GEM detectors[14] are used [54]. The four remaining stations are separated by muon shields made from 80 cm thick iron attenuators. Each station is subdivided into 4 regions, R1 to R4, with varying pad granularity. The lower occupancy level allows the use of Multi Wire Proportional Chambers (MWPC). The chambers are built with 2 mm wire spacing and a small gas chamber of 5 mm width. Each chamber contains a gas mixture of carbon dioxide, argon, and tetrafluoromethane (45:15:40). This mixture is used to achieve a fast time resolution of 3 ns [47].

[14] 24 triple-GEM detector are used, which consist of three gas electron multiplier (GEM) foils sandwiched between anode and cathode plates.

Figure 3.15: View between two of LHCb Muon stations. The image was taken from [42].

The main purpose of the muon stations, shown in Figure 3.15, is to identify muons from electrons and high energy hadrons. This information is used for the level-0 and high level triggers[15], and in off-line analysis. M1 is placed in front of the SPD/PS system and is important for transverse momentum measurements of the muon track used in the L0 muon trigger. M2-M3 are primarily used to obtain good (20%) transverse momentum resolution for the L0 trigger, while M4 and M5 are used to confirm the presence of penetrating muons. Muon identification indicates the presence of an interesting event in our detector. For example, they are present in the final states of many $\mathcal{CP}$-sensitive B-decays, including $B_s \rightarrow J/\psi(\mu^+\mu^-)\phi$ or rare decays such as $B_s \rightarrow \mu^+\mu^-$.

3.4.7 The Trigger system

Proton-proton bunches collide at the LHC every 25 ns leading to event rate of 40 MHz. At LHCb's designed luminosity of 2×10^{32} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ an average of 10 million proton events per second will be produced. Currently 2 kHz of selected events can be written to permanent storage. It is the task of the trigger system to select with high efficiency, and in real time those events which bear the signatures of the physical processes to be studied (signal), and with a high purity those processes to be discarded (background). In addition, the trigger system has to provide the means to deal with highly varying frequencies of different signal event types. For instance, the system should be able to pass all

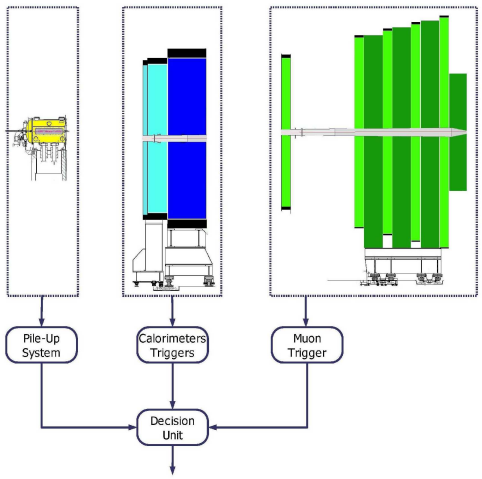


Figure 3.16: The Level-0 trigger system [47], consisting of: the pile-up system of the VELO, the Level-0 calorimeter trigger and the Level-0 muon trigger system.

[15]The muon detector provides information to the L0 on high transverse momentum muons, helping to enrich the triggered content of heavy-flavour events.

events of rare signal types, while limiting in a controlled way the amount of signal types to fit the available bandwidth for data transfer and storage.

The LHCb trigger system is based on a two-level system [55], and uses known properties of b-hadrons to select the decays of B mesons:

- their long lifetime, which leads to a displaced secondary vertex and therefore to tracks which exhibit large impact parameter with respect to the primary interaction point. The decay products of the B-mesons generally are measured to have a large transverse momentum $p_\perp$ or energy ($E_\perp$)[16], which is due to large mass difference between the B-hadrons and their decay products.
- Global event variables such as charged track multiplicity and number of interactions per events are also used.

The data initially collected from an event is first passed through the Level-0 (L0) trigger, reducing the event rate to 1 MHz. The event rate is further reduced using the High Level Trigger (HLT), which selects events at a rate of 2 kHz for mass storage and subsequent offline analysis.

The Level-0 trigger

This is a fast[17] hardware-based triggering system relying on four subsystems: the pile-up system, the calorimeter trigger, the muon trigger, and the decision unit that compiles the global decision. This system is schematically shown in Figure 3.16 where:

The Pile-up system detects multiple primary vertices. This is achieved by measuring the radial position of tracks using two pairs of overlapping r-sensors, see Figure 3.5, placed upstream of the VELO. Track hits coming from each sensor pair are then used to determine the position of primary vertices along the beam line. This is used to veto pile-up (multiple interaction events).

The Calorimeter trigger looks for high $E_\perp$ particles: electrons, γ, π^0 or hadrons. An event is accepted if clusters identified as:

- electrons, using information from M1, the SPD, the PS and the ECAL, have an $E_\perp$ > 2.6GeV.
- γ, using information from the SPD, PS and the ECAL, have a $E_\perp$ > 2.3GeV.
- hadrons, using information from M1, the PS and the HCAL, have a $E_\perp$ > 3.5GeV, or > 4.0 GeV for π^0.

[16] $p_\perp$ is measured with respect to the z-axis: $p_\perp \equiv \sqrt{p_x^2 + p_y^2}$. $E_\perp$ is defined to be $E_\perp \equiv \sin(\theta)E$, where E is the total energy of the particle and θ is the angle the particles trajectory makes with the interaction point.

[17] The latency of L0 is fixed to 4 μs; the maximum time between a pp interaction and the time for the L0 decision to reach the Front End (FE) electronics.

If the total energy of the calorimeter is below 5.0GeV the event is vetoed.

The Muon trigger will accept an event if a reconstructed track identified as a muon, with hits in all five muon chambers, has a $p_\perp > 1.3$ GeV/c. The event is also accepted if the sum of the $p_\perp$ of the two largest $p_\perp$ muons is above 1.5 GeV/c.

The Decision Unit receives information from the each Level-0 subtrigger and produces a single decision per event. Events are accepted when at least one test of the calorimeter trigger or muon trigger is passed and there is no veto from the pile up. If the $p_\perp$ of the two largest $p_\perp$ muons in an event is above threshold, the decision unit sends out a positive result irrespective of any vetos acquired.

The Higher-Level triggers

The Higher-level triggers (HLT), shown in Figure 3.17, are software based trigger systems. They are implemented as C++ applications on a PC farm which consists of about 2000 CPU nodes. An fraction of this cluster is used for the HLT1 which reduces the event rate from 1 MHz down to 30 kHz. The HLT1 first refines the L0 trigger decision. For this it uses the data available to the L0 but allows for more computing time and basic event reconstruction using multiple subdetectors. Depending on the characteristics of the events accepted at this stage they are passed to one or several independent trigger alleys to further refine the selection, including:

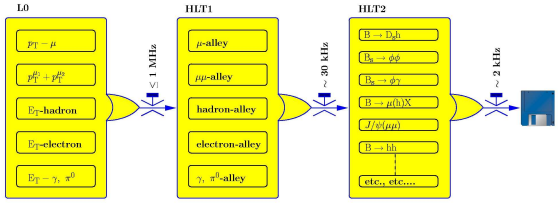


Figure 3.17: [47] Global view of the LHCb trigger scheme. Only a subset of the HLT1 lines is applied depending on the types of L0 candidates present in the event, while all HLT2 selections are run for every event.

the muon only alley: an event is selected when it contains two muon candidates with an invariant mass > 2.5 GeV/c^2, or when the invariant mass of the pair is > 0.5 GeV/c^2 and the impact parameter (IP) with respect to the primary vertex is > 100 μm.

the muon and hadron alley: this alley is run when an event which passes the L0-muon trigger, but fails the HLT1 muon alley. Events are selected with a muon and hadron track with high IP and $p_\perp$.

the hadron only alley: an event is selected if there is at least one hadron with a $p_\perp > 2.5$ GeV/c and an $IP > 150\mu$m, or if there are two hadrons one with a $p_\perp > 1.1$ GeV/c and the other with a $p_\perp > 0.9$ GeV/c.

the electromagnetic (e, γ and π^0) alley: events passing this alley are selected when one calorimeter object has $E_\perp > 3$ GeV/c and $p_\perp > 1.2$ GeV/c.

This way each alley selects more specifically on information provided by one of the L0 subsystems. The event characterisation becomes more complete by the addition of tracks from the tracking subdetectors, the VELO and/or T1-T3 stations. Tracks with large impact parameters are searched with additional criteria, the $p_\perp$ of the tracks, being used, i.e. events with at least one long track[18] and with a $p_\perp$ of at least 1.5 $\mathrm{GeV/c}$ will pass.

If the event is selected by at least one of the HLT1 allies, it proceeds to the second HLT trigger level, the HLT2. This trigger is executed for all events passing the HLT1 trigger and is independent of which alley the event passed. The aim of HLT2 is to reduce the combined output rate of the alleys from around 30 kHz down to 2 kHz (20 Mbits/s) the rate at which the data is written to storage. First a full reconstruction of particles in the event is preformed. The reconstructed event is then processed through exclusive and inclusive selection algorithms in parallel. The inclusive algorithms select partial B decays such as: ϕX, $J/\psi X$, $D^* X$, $\mu^\pm X$, $\mu^\pm h X$, $\mu^+\mu^- X$. The exclusive algorithms select individual final states with the highest possible efficiency. An event is accepted and potentially stored if it passes either the inclusive or exclusive selections.

3.5 LHCb simulation software

The LHCb software environment is built upon an Object Orientated (OO) C++ framework called GAUDI [56]. This framework provides a common environment for the production of Monte Carlo (MC) events, the simulation of the detector, the reconstruction of simulated and real data, and the analysis of either data. Figure 3.18 shows the structure of the software environment used and the data flow between the logical components. The first stage in the simulation software is the event generator. This is software used to simulate collisions of two protons, the underlying event, parton showering, hadronisation etc. A dedicated software package, EVTGEN [57], is then used to simulate b-hadron decays. The final states produced by the event generator are then propagated through a simulation of the LHCb detector provided by the GAUSS [58] software package. This software is based on the simulation package GEANT 4 [59]. These simulations emulate, as accurately as possible, the passage of the particles through the apparatus and the corresponding detector response.

In the third step a digitisation of the detector electronics is simulated, using the BOOLE [60] package. In parallel to this real pp collisions are produced by the LHC accelerator and their decay products produce a response in the real LHCb detector, which in turn gets digitised in the real readout electronics. The digitised data coming out of either the simulated or the real detector is then passed into the common reconstruction stage, BRUNEL [61], where the high-level particles are built from the available detector information to characterise the event. This data is

[18] These are tacks that traverse the whole LHCb tracking system, from the VELO down to the T stations.

then permanently stored in so called Data Storage Tapes (DST), with the raw data still attainable. Finally, the reconstructed events can then be processed at a later time through the analysis stage, provided by the DAVINCI package [62], where the physics parameters of interest can be extracted.

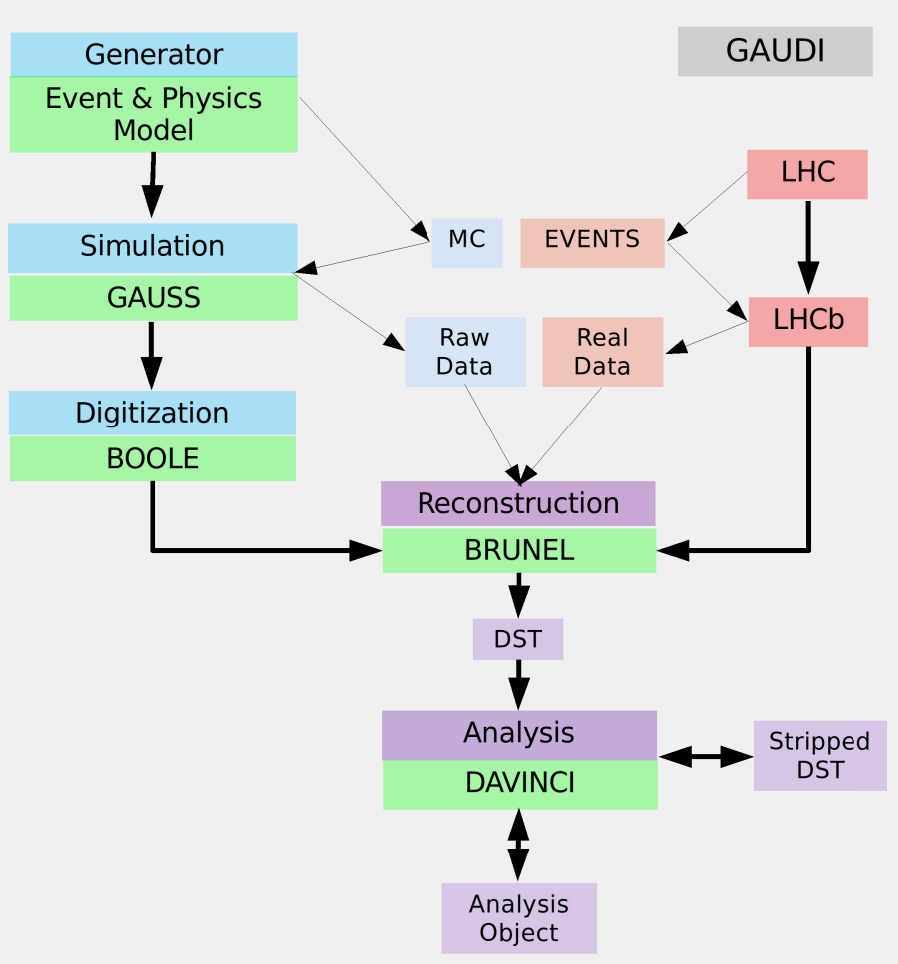


Figure 3.18: Flow diagram illustrating the different LHCb software components embedded in GAUDI. Marked in red are the stages which process real data. Boxes in blue represent software components processing simulated data. The components marked as Reconstructed and Analysis deal with both real and simulated data, and is given in purple.

The DAVINCI software package allows the off-line selection and analysis of specific final states, e.g. b-hadron decays. Final state particles are identified from all available detector information, from primary and secondary vertices are reconstructed with the highest available precision. A series of custom defined C++ algorithms are used to select final state particles and event topologies characterising the decay process under study. These algorithms contain selection criteria including: transverse momentum, particle mass, flight distance, etc to help filter out the signal events. DAVINCI versions v19r7 to v19r13 have been used in this thesis.

3.5.1 Generation of data samples

Before the LHC is fully operational, only generated (or fully simulated) Monte Carlo (MC) data can be used to test the above event generation framework, and estimate the detectors performance to physics decays of interest. The software chain starts with PYTHIA [63], which generates all the pp minimum-bias interactions with a centre of mass energy of $\sqrt{14}$ TeV and deals with showering and hadronisation. EVTGEN is then used to simulate the decay of these hadrons, specifically B-meson decays. This software originated from the BABAR B-physics program and is designed to handle complex $\mathcal{CP}$ violating decays, like $B_d \rightarrow J/\psi K_s$. Specific decay channels of interest including:

Minimum-bias events All events which are generated by PYTHIA and passing through EVTGEN.

$b\bar{b}$-inclusive events All events generated by PYTHIA which contain at least one b hadron within the acceptance cone of 400 mrad with respect to the beam axis are kept. The b-hadron decays through EVTGEN.

Signal events Events containing a b hadron (or any other signal) of interest within the 400 mrad geometrical acceptance cone. PYTHIA repeats its hadronization until the interaction contains the required hadron type. The signal candidate is then forced to decay in EVTGEN according to the decay chain specified by a decay file and all the other underlying particles from the event are decayed.

Monte Carlo data

There is a distinction between data which passes through the generator[19] level (MC truth data) and those events which also pass through detector reconstruction (MC data) stage.

Even after applying an off-line selection, the original MC truth information is still preserved in the reconstructed data. This makes it possible to determine efficiencies. After off-line selection, a candidate particle needs to be re-associated with its corresponding generated MC event. This match is provided by high level tools, called Linkers and Relations tables [64]. The match

[19]Generated MC events passing through the LHCb software model, end up stored in DST (Data Storage Tape) files. These files are used during the analysis stage, by applying selection cuts, to reconstruct the events.

follows the criteria that a reconstructed track or particle is said to be associated to a MC generated particle if the clusters used to form the underlying track or particle are matched to a certain fraction of hits coming from the same MC particle. If no association is found the track is called a ghost track. Two or more tracks associated to the same MC generated particle are then called clones.

Except for the minimum-bias sample, all other samples have a simulated generator level acceptance cut of < 400 mrad. The efficiency of this cut varies with the hadron type under generation [43]. This needs to be well understood since these specific efficiencies significantly effect the event yields and background-to-signal ratios. The specific geometrical efficiencies for the difference hadrons under investigation in chapter 4 are quoted in Table 3.1.

Generator-level cut [%]			
$\epsilon^{\theta}_{sigB_u}$	$\epsilon^{\theta}_{sigB_d}$	$\epsilon^{\theta}_{B_s \to J/\psi\phi}$	$\epsilon^{\theta}_{sigB_{b\bar{b}}}$
34.9 ± 0.3	34.9 ± 0.3	18 ± 0.006	43.4 ± 0.3

Table 3.1: Generator-level geometrical acceptance cut [65].

Data samples

Table 3.2 below gives the data samples used for the selection (chapter 4), background (chapter 5) and sensitivity (chapter 7) studies preformed in this thesis. The data samples and the software used are all part of the LHCb Data Challenge 06 (DC06). The sensitivity studies carried out in chapter 6, used values obtained from a previous LHCb Data Challenge DC04.

LHCb Event type	Description	type	# of Events generated events (10^6)
13144002	$B_s \to J/\psi\phi$	Sig	1.6
13264001	$B_s \to D_s\pi$	Sig	2.0
24142001	J/ψ-inclusive	Bkg	1.9
11144001	$B_d \to J/\psi K^{*0}$	Bkg	3.1
10000000	$b\bar{b}$-inclusive	Bkg	27

Table 3.2: The LHCb DC06 data samples used in this study and the corresponding number of generated events available for this study; these samples were generated using Gauss version v30r0. The equivalent luminosity for the $B_s \to J/\psi\phi$ signal sample is $\sim 2.5\ fb^{-1}$.

Figure 3.19 shows one of the $B_s \to J/\psi\phi$ events obtained from the LHCb event type 13144002 using Panoramix[20]. This event type is generated with the EVTGEN PVV_CPHL

[20]This is the LHCb model for data visualisation, version v16r0 has been used to view the $B_s \to J/\psi\phi$ events shown in Figure 3.19.

model [66], which allows for a $\mathcal{CP}$ violating time dependent asymmetry. Events generated from this event type also required that each of the $B_s \rightarrow J/\psi\phi$ decay products lie within the geometrical acceptance of the detector. In the next chapter, the characteristics of the $B_s \rightarrow J/\psi\phi$ event data will be investigated.

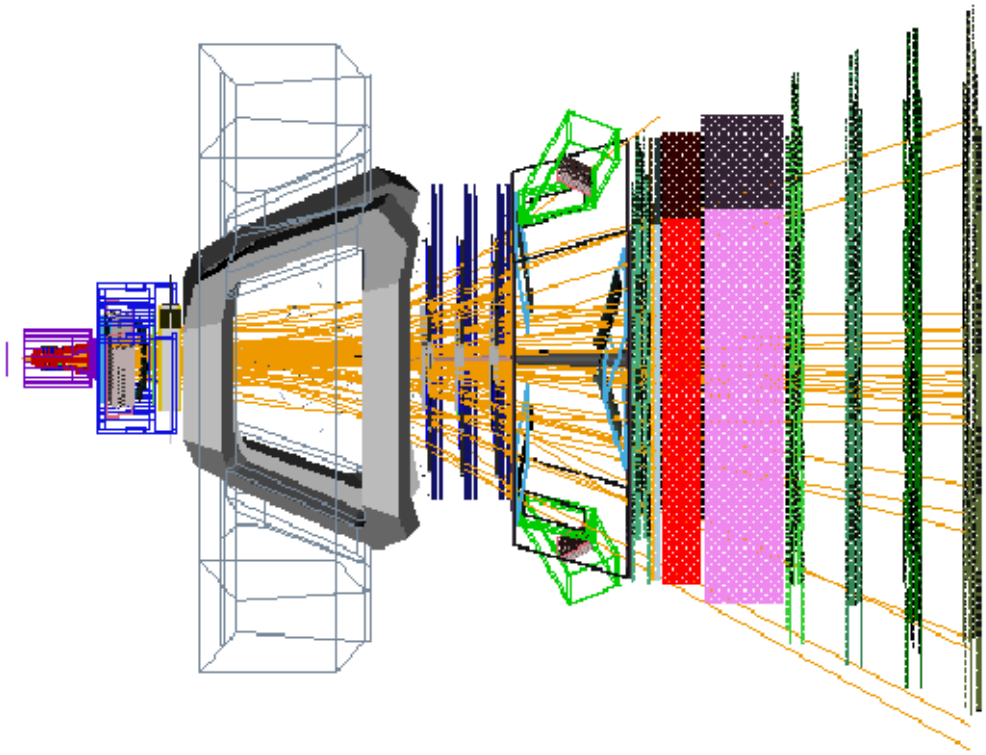


Figure 3.19: Panoramix (version v16r0) view looking down at the LHCb detector in the $x - z$ plane, with the interaction point on the left-hand side. The tracks in orange show a typically fully simulated $B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ event.

4

$B_s \rightarrow J/\psi\phi$ Event selection

We first need to reconstruct and select the $B_s \rightarrow J/\psi\phi$ decays. The first step in selecting events is the trigger system, which reduces the overall event rate from 40MHz down to 2kHz. This is then followed by an off-line selection algorithm to further filter the decays of interest. In this chapter we will present the off-line selection procedure for the $B_s \rightarrow J/\psi\phi$ channel, using the fully simulated Monte Carlo data.

4.1 Decay topology

From the proton-proton interaction point, it is possible to create a $b\bar{b}$ pair, with either one having the possibility of hadronising to form a B_s^0 meson. Given the natural decay length, $\tau \sim 418.5$ μ m, and average momentum, ~ 150 GeV/c, each B_s^0 produced at the LHCb will, on average, travel $c\tau p/m \approx 1.2$ cm before decaying. This leads to the B_s^0 most characteristic signature at LHCb which is that of a displaced secondary vertex. The B_s^0 can then decay into J/ψ and ϕ resonant states[1], which can instantaneously decay into a di-muon and di-kaon pair[2]. The J/ψ and ϕ are very short lived, meaning the di-muon and di-kaon pair effectively originate from a common vertex. In turn this means that there are only two vertices of interest, the B_s^0 production and decay vertex. A cartoon illustration of the $B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay along the boost z-direction of the detector is shown in Figure 4.1

4.2 Selection procedure

This procedure will be divided into five stages:

- **Determination of the selection variables**. The first step is to identify a set of reconstruction parameters that provide a good separation between signal and background events. In general this will require: making use of log likelihood difference $\Delta ln\mathcal{L}$ requirements for PID charged candidate tracks; transverse momentum p_t cuts on final state particles; the

[1] Use of J/ψ and ϕ to explicitly mean $J/\psi(1S)$ and $\phi(1020)$ respectively, will be used throughout this thesis.

[2] In this study use is made of the $J/\psi \rightarrow \mu^+\mu^-$ and $\phi \rightarrow K^+K^-$ channels. It is also possible to for the J/ψ to decay into e^+e^- which occurs 5.94% of the time.

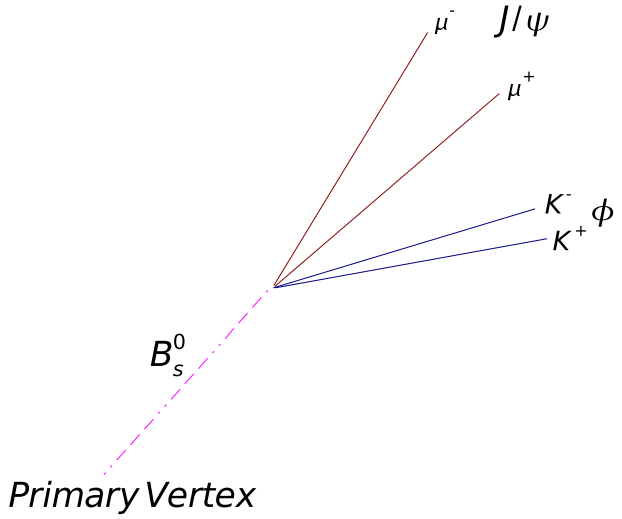


Figure 4.1: Decay topology of $B_s \rightarrow J/\psi\phi$. A $b\bar{b}$ pair is created at the primary vertex (PV) with either one hadronising into the B_s^0. The long lifetime of the B_s^0 means it decays a distance ~ 1.2 cm before decaying. This is shown by the dotted red line. The B_s^0 can then decay into a J/ψ and ϕ which, due to their short lifetimes, can decay almost instantaneously into a di-muon (solid magenta lines) and di-kaon pair (solid blue lines), respectively.

goodness of track fitting to a common vertex, given by the χ^2; an invariant mass window on candidate particles.

- **Pre-Selection**. The next step is to apply a loose set of cuts on the selection variables. This is to reduce the combinatorial background in the signal and especially in background samples. This will be achieved by applying $B_s \rightarrow J/\psi\phi$ pre-selection cuts to $B_s \rightarrow J/\psi\phi$ signal and background data samples [67]. These preselection cuts were not developed during these studies.

- **Selection**. After applying a pre-selection, a more specific set of selection cuts on p_t, reconstructed vertex and mass window on the J/ψ, ϕ and B_s candidates, in accordance with [68], are applied to further purify the signal to background ratio.

- **Analysis of the selection results**. Events passing through the final selection cuts are then studied. Calculation of the optimal signal yield, selection efficiencies and background-to-signal ratio are calculated. The quality of reconstruction in terms of mass, propertime, decay angles - associated with the final state particles - and tagging are studied.

- **Cut Optimisation**. After selection, optimisation of the selection cuts are studied. This stage, discussed in chapter 7, will employ Root's MultiVariate Analysis (MVA) classifiers, in particular the Boosted Decision Tree (BDT). The study will be performed to optimise the background-to-signal ratio using prompt J/ψ's as the source of background.

4.3 Selection variables

The selection variables can be divided into four groups namely: particle identification, $p_\perp$ of the decay products, the goodness of a vertex fit to charged tracks and invariant mass windows.

4.3.1 Particle identification

At the beginning of the selection the reconstructed tracks of each event are assigned a particle identification (PID) based upon information from the different sub-detector. The PID is provided by the RICH counters (π, K, p), the ECAL, the HCAL (e, γ, hadrons) and the muon system (μ). The PID is not unique and several PID's can be assigned to a signal track.

The PID information is used in the offline selection through the likelihood hypothesis combining the information of all the sub-detectors. For example using the RICH, calorimeter and MUON subdetectors, likelihood hypothesis that the track is a μ or hadron is given as:

$$\mathcal{L}(\mu) = \mathcal{L}^{RICH}(\mu)\mathcal{L}^{CALO}(\text{non } \mu)\mathcal{L}^{MUON}(\mu)$$
$$\mathcal{L}(h) = \mathcal{L}^{RICH}(h)\mathcal{L}^{CALO}(\text{non } e)\mathcal{L}^{MUON}(\text{non } \mu)$$

where h represents a hadron, e stands for an electron and μ for a muon. Each function computes the probability of being of the given type of particle. If there is more than one estimator from a given detector, they can be simply combined by taking the product of their individual likelihoods.

Particles are then selected by cutting on the ratio of likelihoods between different hypothesis, by forming a delta log likelihood (DLL) function comparing the two hypothesis. The ratio is defined as:

$$\Delta \ln \mathcal{L}_{AB} = \ln \mathcal{L}(A) - \ln \mathcal{L}(B) = \ln \left[\frac{\mathcal{L}(A)}{\mathcal{L}(B)}\right] \tag{4.1}$$

where A, B stand for the hypothesis to have A, B. The ratio (4.1) tends to be positive for correctly A-type identified A-type particles and negative for correctly identified B-type particles. This method can only be applied to charged tracks, γ and π^0 need another treatment[3].

A more specific example of log likelihoods is given by the performance of the RICH system to distinguish kaons and pions, in $B_s \to D_s^- K^+$ events, which is given by:

[3]Photon identification is based on whether there are hits on the SPD cell in the ECAL. Neutral pions are reconstructed from two photons. They are mostly reconstructed at low momentum (3 GeV/c) where the ECAL can separate the pair of separate photons, while at high P_T the pair cannot be properly resolved by the ECAL (the reconstructed π^0 is then referred to as being "merged")

$$\Delta \ln \mathcal{L}_{K\pi} = \ln \mathcal{L}(K) - \ln \mathcal{L}(\pi) = \ln \left[\frac{\mathcal{L}(K)}{\mathcal{L}(\pi)} \right] \tag{4.2}$$

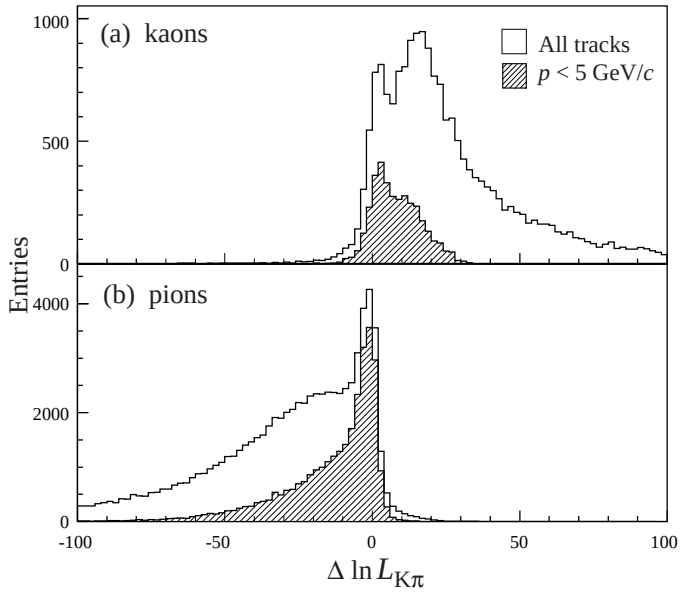
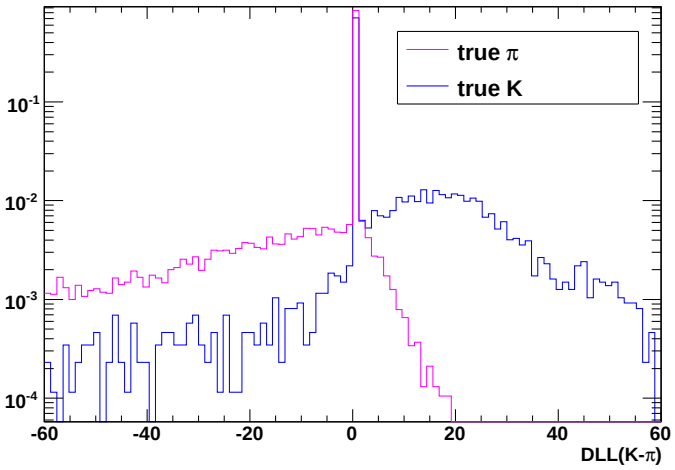


Figure 4.2: Difference in log-likelihood between the kaon and pion hypotheses from the RICH system [43]. The top plot describes the decision for kaons, while the distribution for pions is shown on the bottom, for a sample of $B_s \rightarrow D_s^- K^+$ events; shaded areas show low momentum tracks.

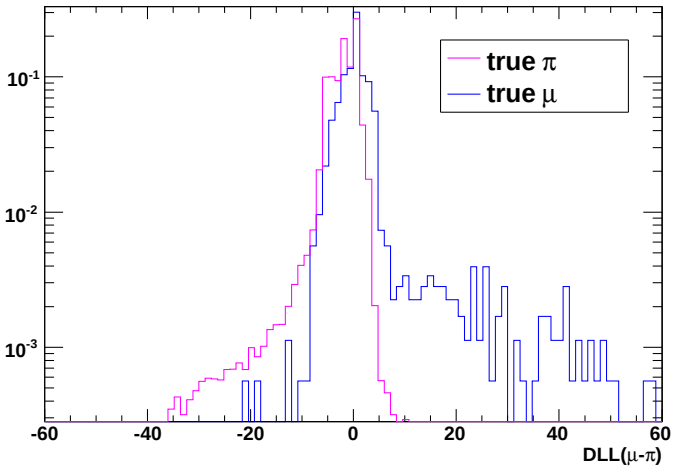
Figure 4.2, shows clearly how kaons and pions can be distinguished. Since $\Delta \ln \mathcal{L}_{K\pi}$ tends to have positive values for kaons, whilst pions tend to have negative values. The double-peaked structure in Figure 4.2(a) is due to the momentum-dependence of the $\pi - K$ separation of the RICH system.

In Figure 4.3, the distribution of true pions is shown in magenta with respect to true kaons (top) or true muons (bottom) shown in blue. Here the $\mu - \pi$ separation[4] is seen to be less distinct with respect to the $K - \pi$ separation. This is because the masses of the muons (μ_{mass}=105 $\mathrm{MeV/c^2}$) and pions (π_{mass}=139 $\mathrm{MeV/c^2}$) are nearly degenerate with respect to the kaon mass (K^+_{mass}=493 $\mathrm{MeV/c^2}$), meaning its more difficult to distinguish the muons and pions. As the particle identification information in this example is taken as the global decision (the decision used from all sub-detectors). The separation power coming from the muon system is washed out by the other sub-detectors where its more difficult to distinguish the pions and muons.

[4]Where a $\Delta \ln \mathcal{L}_{K\mu}$ > -5 is applied as part of the offline selection cuts, as discussed in Table 4.5.



(a) $\Delta ln_{K-\pi}$

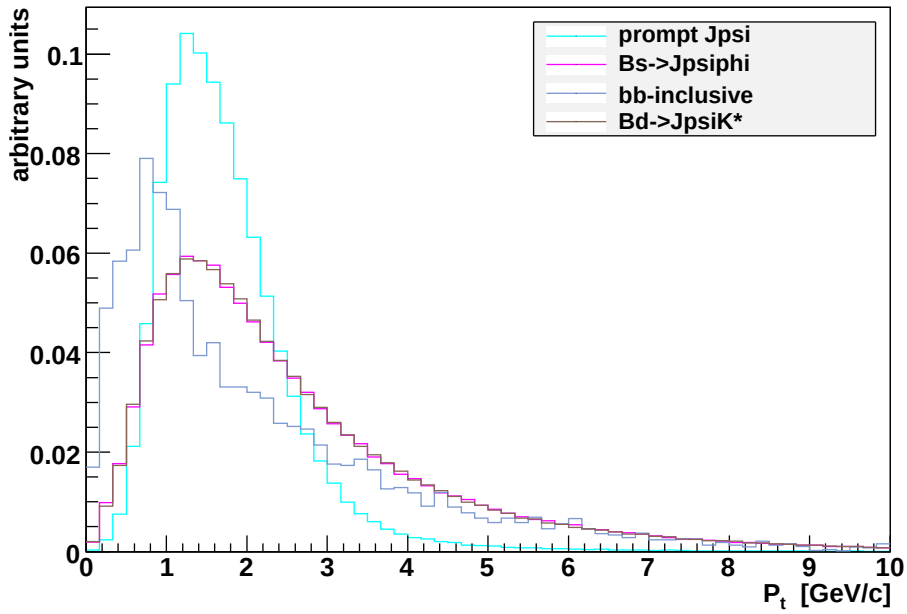


(b) $\Delta ln_{\mu-\pi}$

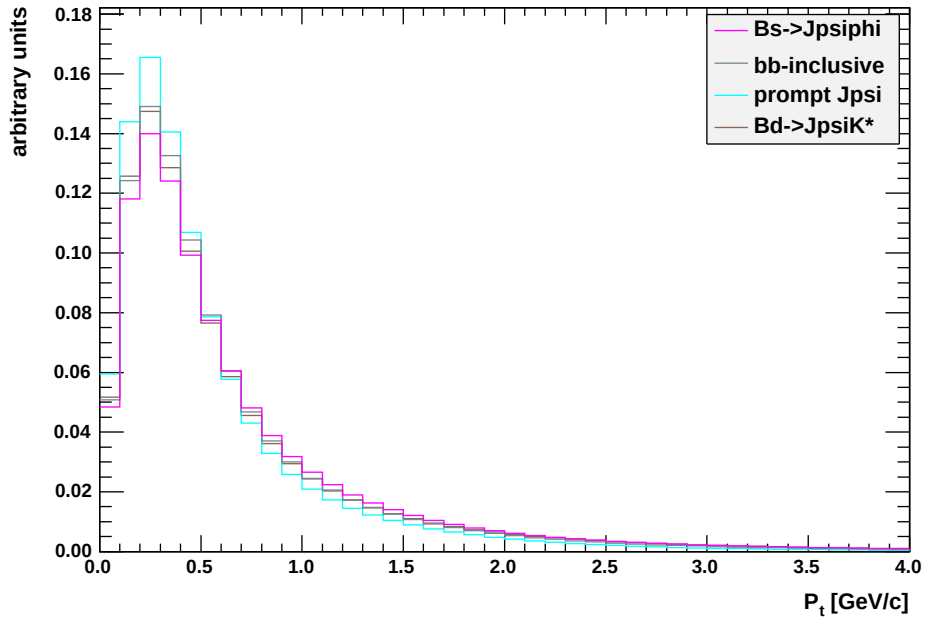
Figure 4.3: $\Delta ln\mathcal{L}_{X-\pi}$ separation for true kaons and muons $X \in \{K, \mu\}$ with respect to true pions obtained using the combined sub-detector decision with $B_s \rightarrow J/\psi\phi$ event data. (a) Show the separation for kaons (blue) verses pions (magenta), while (b) shows the separation for muons (blue) with respect to pions (magenta).

4.3.2 $P_\perp$ of B_s decay products

A property of b-decays is that the final state decay products typically have higher $p_\perp$ distributions than that from non-b-decays. It is therefore possible to reject some of the background to $B_s \rightarrow J/\psi\phi$ by cutting on the $p_\perp$ of the muons and kaons. A cut of greater than 750 $\mathrm{MeV/c}$ to both muons and kaons is applied as described in Table 4.5. The $p_\perp$ distribution for both μ^+ and K^+ is shown in Figure 4.4, where the $p_\perp$ is higher for muons coming from the $B_s \rightarrow J/\psi\phi$ signal compared to the non-prompt background.



(a) $p_\perp$ of μ^+ candidates

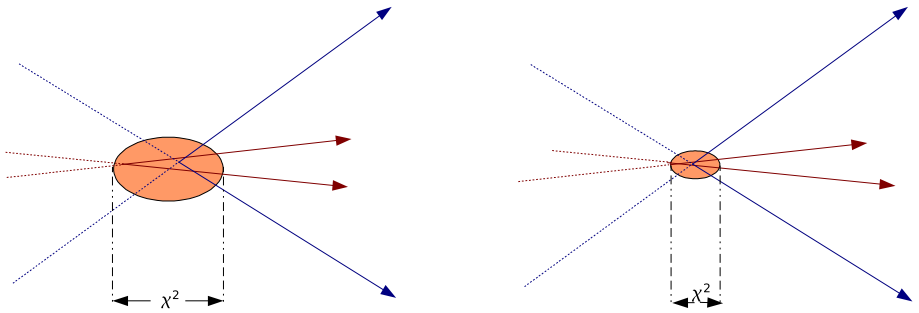


(b) $p_\perp$ of K^+ candidates

Figure 4.4: Distribution of $p_\perp$ and momentum for μ^+ (a) and K^+ (b) from $B_s \rightarrow J/\psi\phi$ (magenta), $b\bar{b}$-inclusive (grey), prompt J/ψ (cyan) and $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ (brown) data samples before any cuts. The distributions shown are normalised to unity.

4.3.3 The vertex χ^2 requirement

To effectively reject random combinations of tracks when forming reconstructed particles, a criteria widely used is that of the quality of the vertex fit, χ^2. This criteria determines how precisely tracks can be fitted to form a common vertex. The smaller the χ^2 the more probable the tracks are from originating from a common vertex. The χ^2 requirement is illustrated in Figure 4.5 for the reconstruction of four tracks to a common vertex.



(a) lose χ^2 (b) tight χ^2

Figure 4.5: Fit of four tracks to a common vertex. The shaded orange area illustration the quality of the vertex fit. (a) shows a lose vertex fit corresponding to a high vertex χ^2, (b) corresponds to a tight vertex fit corresponding to a low χ^2.

4.3.4 Mass window requirement

An invariant mass window requirement is simply to reject unwanted events when the mass of the reconstructed particle is well known. For example, we would apply a mass window[5] to the reconstructed B_s^0 to help get rid of decays such as $B_s^0 \to J/\psi(2S)\phi$ or $B_s^0 \to J/\psi\pi\pi$, which have mass distributions 5366.5 $\pm$ 1.9 MeV/c^2 and 5366.63 $\pm$ 3.2 [69]. The J/ψ, ϕ and B_s^0 nominal mass peaks are given in Table 4.1, with their respective distributions after applying the offline selection cuts shown in Figure 4.6. Further discussion on the mass resolutions, for each plot given in Figure 4.6 can be found in section B.2.

[5]The canonical mass window requirement is to selected candidates (X), such that:

$$\left|(M_x - M_{PDG(x)})\right| = |\Delta M_x| < \frac{M_{\text{window}}}{2}.$$

Where M_x, $M_{PDG(x)}$ are the reconstructed and PDG masses respectfully and M_{window} is the mass window.

Nominal Particle Masses ($\mathrm{MeV/c^2}$)	
$M_{J/\psi}$	3096.92 ± 0.01
M_{ϕ}	1019.46 ± 0.02
$M_{B_s^0}$	5369.6 ± 2.4

Table 4.1: Nominal J/ψ, ϕ and B_s^0 masses taken from [7].

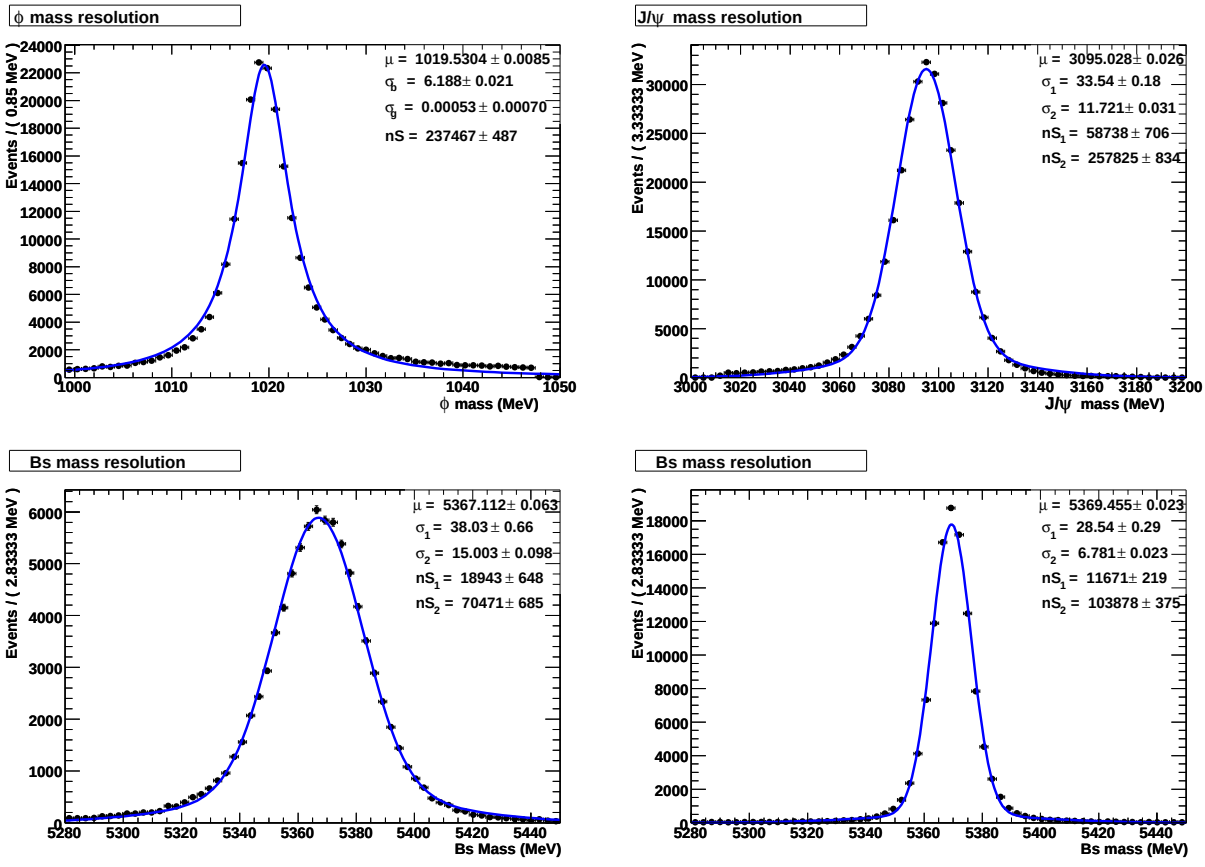


Figure 4.6: The reconstructed mass of the ϕ (top left), the J/ψ (Top right) and the B_s^0 (bottom left) mesons from the selected (see Table 4.5) signal events. The ϕ distribution is fitted with a Breit Wigner convoluted with a Gaussian (σ_b and σ_g are the widths of the Breit Wiger and Gaussians functions respectfully), whilst the J/ψ and the B_s^0 mass distributions are fitted with a double Gaussian. The bottom right plot shows the fitted B_s^0 mass after constraining the masses of the ϕ and J/ψ to their nominal values given in Table 4.1.

4.4 Trigger performance on signal

The LHCb trigger is of crucial importance in selecting amongst the collisions those events of interest for b-physics studies. The trigger is based on a two-level system [70] exploiting the fact that b-flavoured hadrons are heavy and long lived. The trigger decision is built on two components: the level-0 (L0) decision, which is mainly based on high $p_\perp$ energy particles: hadrons, electrons, neutral pions and muons, and the High Level Trigger (HLT) decision, which refines the L0 selected events by passing them through a series of dedicated trigger alleys (HLT1). The alleys perform a partial event reconstruction based on the high transverse momentum and large impact parameter (with respect to the primary vertex) characteristics of the B-decay products; events passing at least one of these alleys then pass through HLT2. This is a set of inclusive and exclusive algorithms.

This section discusses the performance of each trigger decision on the $B_s \to J/\psi\phi$ signal events. The presence of the muons, from the J/ψ mesons, means the $B_s \to J/\psi\phi$ channel will be relatively clean to trigger on: the L0 trigger, running at the nominal LHC bunch cross rate of 40 MHz, is found to be approximately 90% efficient [71] at triggering on muonic channels. The other advantage this channel has comes from the RICH system, which provides good kaon identification from the large hadronic background, this has been discussed in section 3.4.4. Running[6] the L0 decision over the $B_s \to J/\psi\phi$ event data leads to an overall selection efficiency of 93%. A breakdown of each L0 alley is detailed in Table 4.2, where the L0 single muon alley is seen to have the highest efficiency, $\sim 97\%$. After the L0 trigger the event then passes through the HLT.

The HLT1 alleys detailed in Table 4.3 then refine the L0 decision. The nomenclature and a description for each alley is given in [72, 73]. At the time of these studies, the HLT alleys used still require tuning. From Table 4.3 we see that the four most significant alleys are: the single muon trigger (MuonTriggerSingle) with an efficiency of 52.7%, the dimuon alley (MuonTriggerDiMuon) with an efficiency of 26.3%, the dimuon alley with an impact parameter cut on the primary vertex (MuonTriggerIPDiMuon), and the hadron muon alley (MuHadTrigger) with an efficiency of 39.9%. Events which pass at least one of the HLT alleys are then processed by the inclusive and exclusive trigger algorithms, shown in Table 4.4, with most events passing the shared dimuon algorithm. The overall HLT efficiency being 81%, as given in Table 4.4.

[6]The results of the trigger studies given in Tables 4.2 and 4.3 have been obtained using the HLT trigger software, version v1r10 and DaVinci version v19r14. At the time these studies were carried out, using the analysis of software DaVinci v19r14, the trigger system was in a state of transition. The HLT trigger structure changing from a one-tier, to a two-tier arrangement. The restructuring of the HLT means that the HLT1 and HLT2 triggers have still to be tuned using fully generated DC06 event data.

L0 Alleys	Selection efficiency ϵ[%]			
	Signal	$b\bar{b}$-inclusive	prompt-J/ψ	$B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$
LO Muon	53.96	43.0	74.6	87.8
L0 Single Muon	97.84	96.2	93.0	96.8
L0 DiMuon	53.07	98.7	98.3	98.9
L0 All Muon	80.24	100.0	100.0	100.0
L0 Total	93.0	56.0	81.0	77.0

Table 4.2: Efficiencies for signal and background passing the L0 trigger.

Hlt Alleys	Selection efficiency ϵ[%]			
	Signal	$b\bar{b}$-inclusive	prompt-J/ψ	$B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$
MuonTriggerSingle	52.8	62.4	29.9	77.9
MuonTriggerSingleNoIP	8.62	11.7	1.6	12.7
MuonTriggerDiMuon	26.3	53.9	93.4	90.4
MuonTriggerIPDiMuon	23.9	71.5	39.0	82.3
MuHadTrigger	39.9	79.8	47.8	84.3
HadTriggerSingle	1.04	19.5	11.5	18.9
HadTriggerDiHadron	2.32	4.88	54.4	80.5
EleTriggerSingle	0.37	73.3	100.0	65.3
PhoTriggerEleTrk	0.93	0.0	0.0	0.0

Table 4.3: HLT trigger efficiencies for signal and background events. A description of each alley can be found in [72, 73]

Hlt Excusive & Inclusive Alg	Selction efficiency ϵ[%]
HltSharedDiMuon	74.8
HltSelBiasDiMuon	57.1
HltSelB2JpsiX_MuMu	65.3
HltSelUnbiasedDiMuon	39.5
HTL Total	81.0

Table 4.4: Selected HLT inclusive and exclusive and trigger decision running over $B_s \rightarrow J/\psi\phi$ signal events. The overall HLT decision is 81%.

4.5 Selection cuts

J/ψ and ϕ selection

All final state particles, $K^{\pm}$ and $\mu^{\pm}$ are required to have a minimum transverse momentum of more than 750 $\mathrm{MeV/c}$ [67]. From Table 4.5, the J/ψ reconstruction is required to have an unconstrained vertex fit with a χ^2 convergence of less than 6. The final requirement for the J/ψ reconstruction is for each candidates mass to lie within a mass window of $\pm$ 85 $\mathrm{MeV/c^2}$. The effect of applying these cuts to the mass and $p_\perp$ of the J/ψ candidates, is shown in Figure 4.7 and 4.8 respectively.

The ϕ is reconstructed via its decay to K^+K^- using the cuts given in Table 4.5. The ϕ reconstruction requires an unconstrained vertex fit with a χ^2 of less than 40. This selection is less stringent than that for the J/ψ selection since the kaons coming from the ϕ are nearly co-linear. The invariant ϕ mass is then required to lie within $\pm$ 28 $\mathrm{MeV/c^2}$ of the nominal ϕ mass. The effect of applying these cuts to the mass and $p_\perp$ of the J/ψ candidates, is shown in Figure 4.9 and 4.10 respectively.

Selection variables	Requirement
$\Delta \ln \mathcal{L}_{\mu\pi}(\mu^+,\mu^-)$	> 0
$p_\perp(K^+,K^-,\mu^+,\mu^-)$	> 750 $\mathrm{MeV/c}$
$\chi^2(\mu^+\mu-)$	< 6
$\Delta M_{J/\psi}(\mu^+\mu-)$	$\pm$ 85 $\mathrm{MeV/c^2}$
$\chi^2(K^+K-)$	< 40
$\Delta M_{\phi}(K^+K-)$	$\pm$ 28 $\mathrm{MeV/c^2}$
$\chi^2(J/\psi\phi)$	< 22.5
$\Delta M^{sig}_{B_s}(J/\psi\phi)$	$\pm$ 50 $\mathrm{MeV/c^2}$
$\Delta M^{bkg}_{B_s}(J/\psi\phi)$	$\pm$ 500 $\mathrm{MeV/c^2}$

Table 4.5: Selection cuts applied to the $B_s \rightarrow J/\psi\phi$ channel [68].

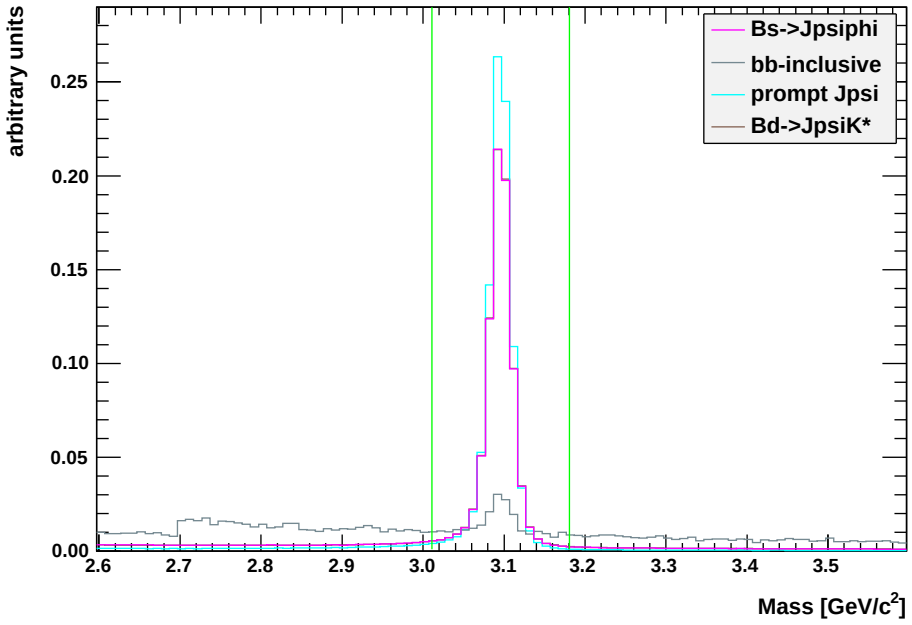
B_s^0 selection

The final part of the selection is to combine the reconstructed J/ψ and ϕ particles into a B_s^0 candidate. The four charged tracks, two muons and kaons, must fit to a single vertex of χ^2 less than 22.5. Each candidates mass is then required to lie within a tight mass window of $\pm$ 50 $\mathrm{MeV/c^2}$. The results of applying the cuts, given in Table 4.5, on the mass and $p_\perp$ of each B_s^0 candidate is shown in Figures 4.11 and 4.12 respectively. A summary of the J/ψ, ϕ and B_s candidates passing the $B_s \rightarrow J/\psi\phi$ selection cuts, is given in Table 4.6.

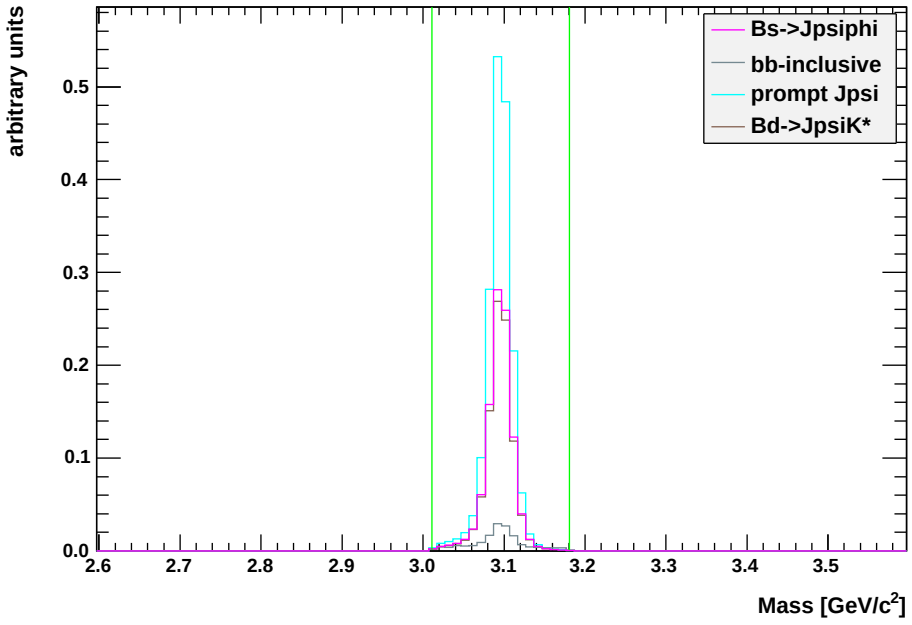
The current $B_s \rightarrow J/\psi\phi$ selection requires no other cuts to be applied. This is in contrast to previous selection studies [67] where an additional propertime significance cut, of greater than 5, was applied to the B_s^0 candidates. The cut was applied to reject events originating from the primary interaction point, specifically prompt $J/\psi \rightarrow \mu^+\mu^-$ background events. The effect of this cut will be raised again when discussing the $\frac{B}{S}$ ratio for prompt-J/ψ in chapter 5.

selection and efficiencies summary for J/ψ, ϕ and B_s					
Candidate	Description	$B_s \rightarrow J/\psi\phi$ Signal	$b\bar{b}$-inclusive	J/ψ-inclusive	$B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$
J/ψ	events before cuts	2.2M	925k	889k	7M
	after $p_\perp$ cut [%]	67.0	93.0	53.0	67.8
	after χ^2 cut [%]	27.0	55.0	13.0	26.6
	after $\lvert\Delta M\rvert$ cut [%]	19.0	7.0	9.0	18.8
ϕ	events before cuts	51.8M	54.3M	12.6M	152M
	after $p_\perp$ cut [%]	93.0	95.0	94.0	94.0
	after χ^2 cut [%]	31.0	39.0	32.0	34.5
	after $\lvert\Delta M\rvert$ cut [%]	87.0	97.0	97.0	97.8
B_s	events before cuts	186k	225	4k	58k
	after χ^2 cut [%]	17.2	68.8	50.0	82.7
	after $\lvert\Delta M\rvert$ cut [%]	7.0	91.0	92.0	88.6
	events after cuts	142k (149k)	6 (33)	162 (764)	1137 (4123)

Table 4.6: selection cut flow for J/ψ, ϕ and B_s selection from the $B_s \rightarrow J/\psi\phi$ decay, where the efficiencies are calculated with respect to the previous cuts. The quantities in the parenthesis indicate the number of B_s candidates passing the wide mass window requirement. Where the generated luminosity for each data sample is arbitrary.

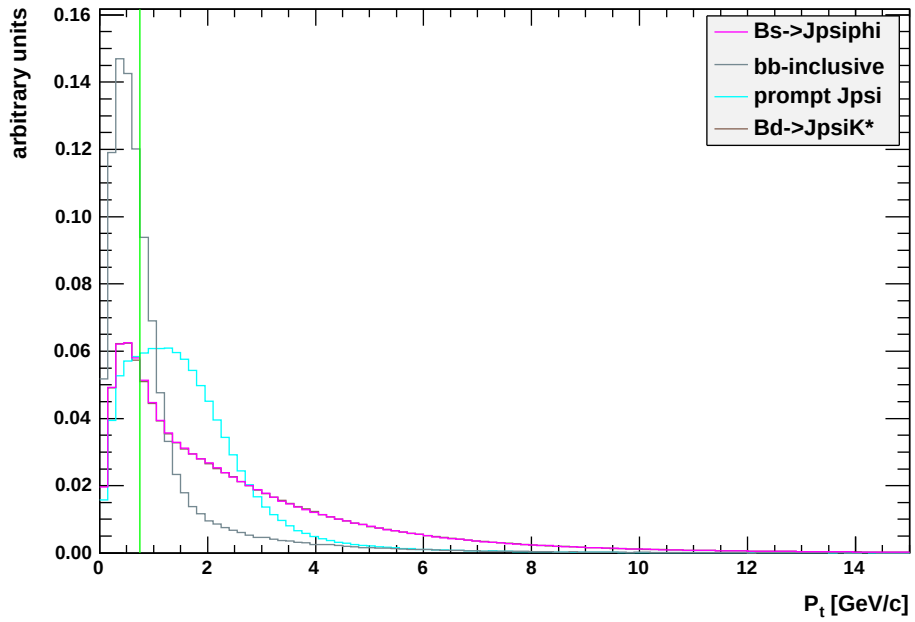


(a) J/ψ mass before cuts

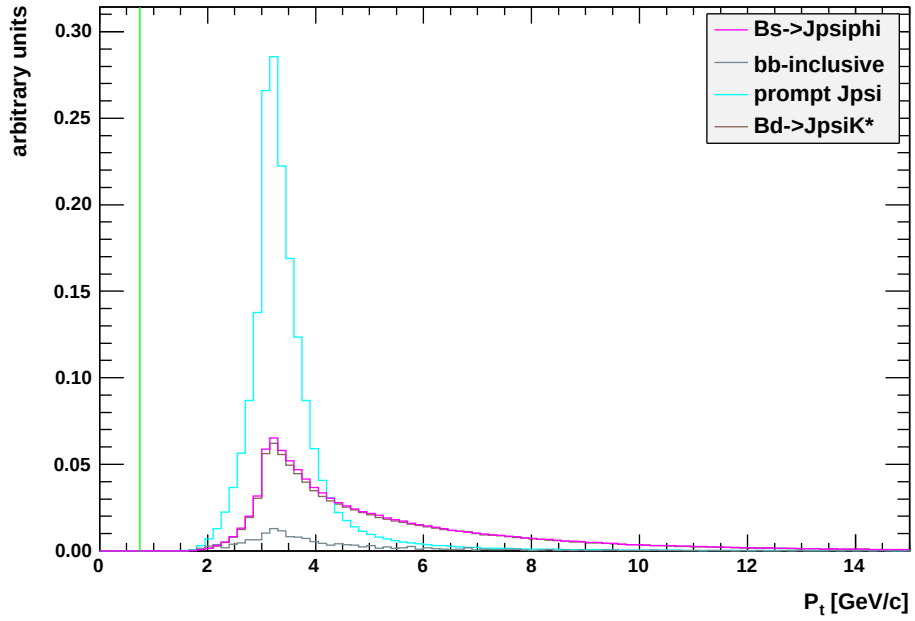


(b) J/ψ mass after all cuts

Figure 4.7: (a) Distribution of mass for J/ψ from $B_s \rightarrow J/\psi\phi$ (magenta), $b\bar{b}$ inclusive (grey), prompt J/ψ (cyan) and $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ (blue) background specific data samples before selection. (b) J/ψ mass distributions after all J/ψ cuts given in Table 4.5. The distributions shown are normalised to an arbitrary scale.

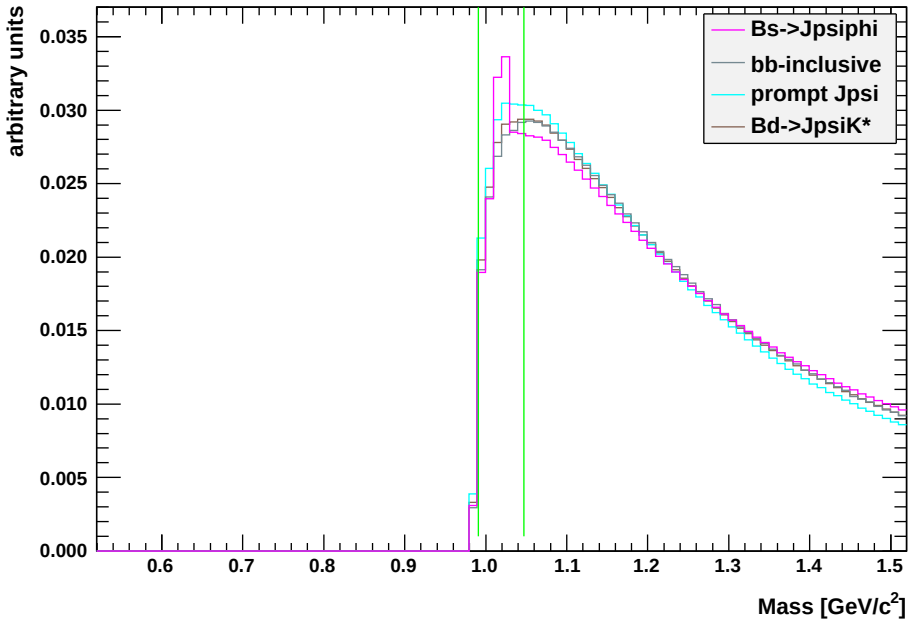


(a) $J/\psi\ p_\perp$ before cuts

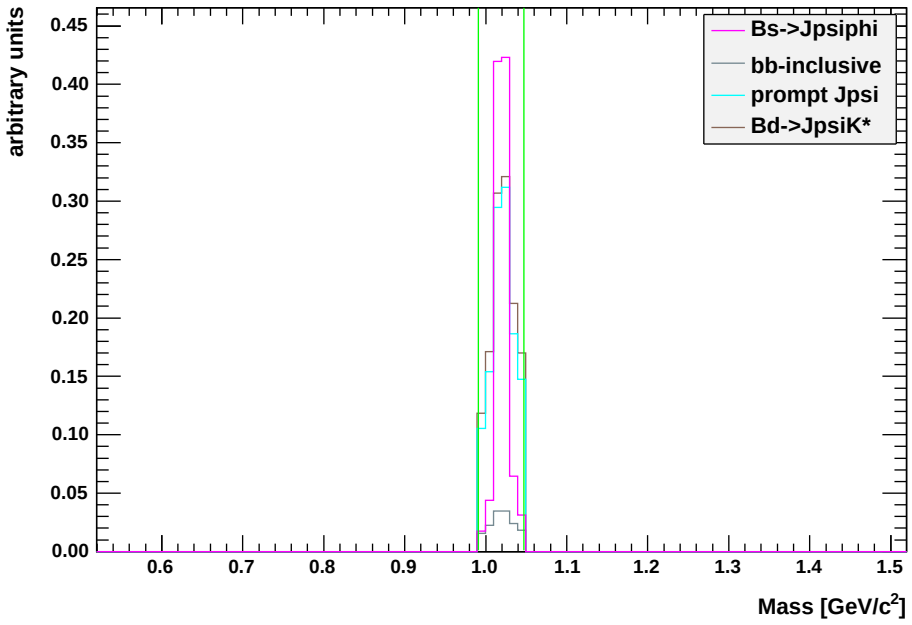


(b) $J/\psi\ p_\perp$ after all cuts

Figure 4.8: (a) Distribution of $p_\perp$ for J/ψ from $B_s \rightarrow J/\psi\phi$ (magenta), $b\bar{b}$ inclusive (grey), prompt J/ψ (cyan) and $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ (brown) background specific data samples before selection. (b) $J/\psi\ p_\perp$ distributions after all J/ψ cuts given in Table 4.5. The distributions shown are normalised to an arbitrary scale.

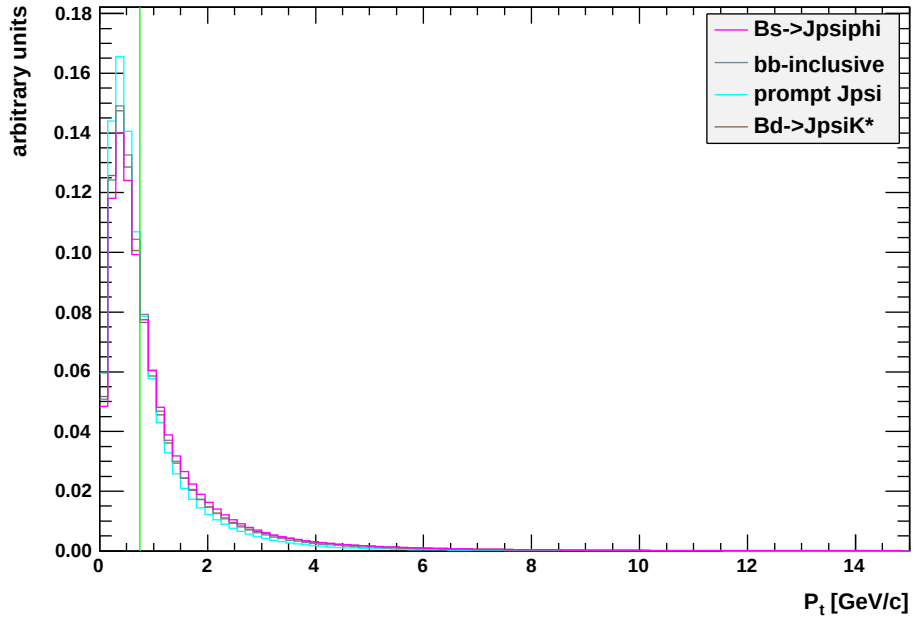


(a) ϕ mass before cuts

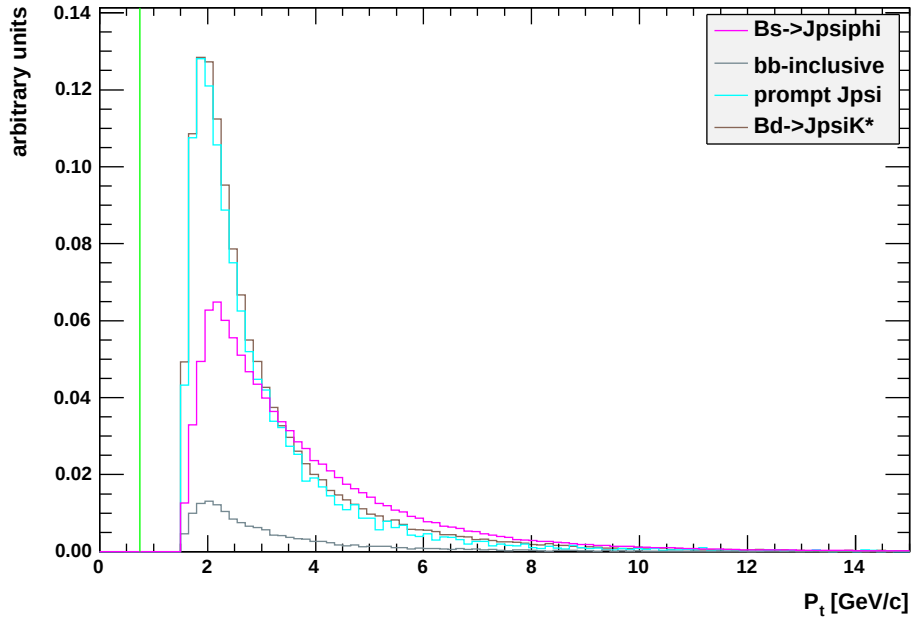


(b) ϕ mass after all cuts

Figure 4.9: (a) Distribution of mass for $\phi(K^+K^-)$ from $B_s \rightarrow J/\psi\phi$ (magenta), $b\bar{b}$ inclusive (grey), prompt ϕ (cyan) and $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ (brown) background specific data samples before selection. (b) ϕ mass distributions after all ϕ cuts given in Table 4.5. after all cut. The distributions shown are normalised to an arbitrary scale.

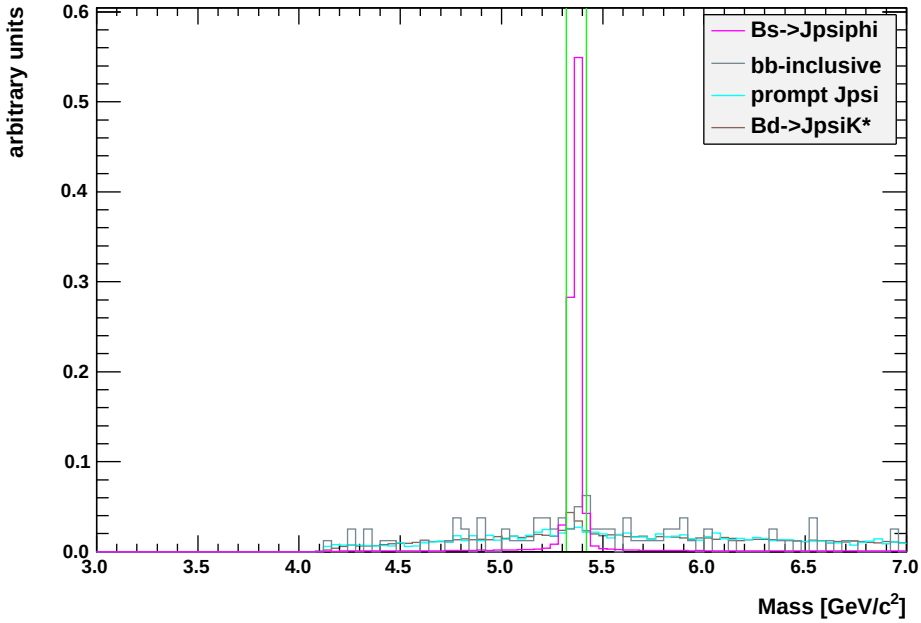


(a) $\phi\, p_\perp$ before cuts

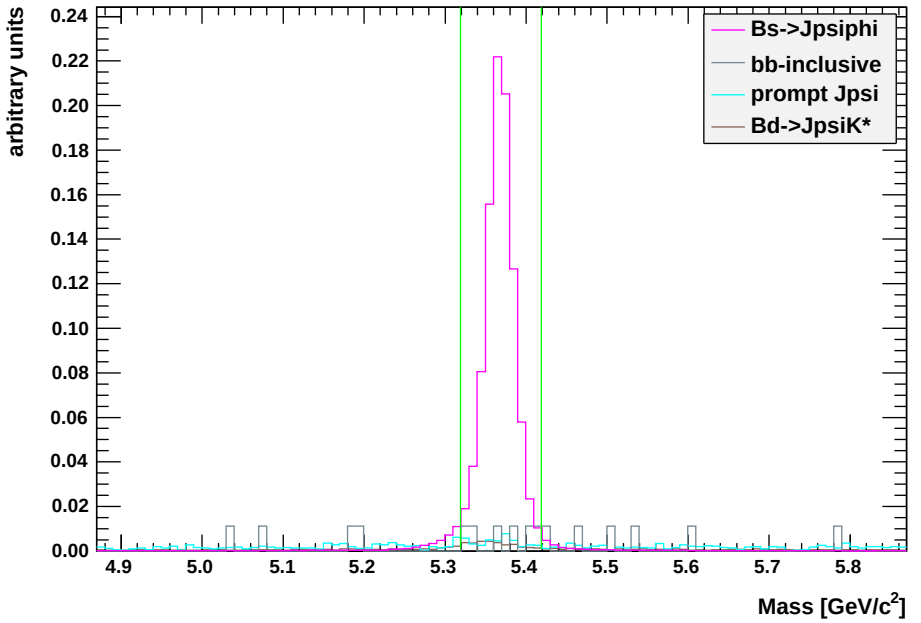


(b) $\phi\, p_\perp$ after all cuts

Figure 4.10: (a) Distribution of $p_\perp$ for $\phi(K^+K^-)$ from $B_s \rightarrow J/\psi\phi$ (magenta), $b\bar{b}$ inclusive (grey), prompt ϕ (cyan) and $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ (brown) background specific data samples before selection. (b) $\phi\, p_\perp$ distributions after all ϕ cuts given in Table 4.5. The distributions shown are normalised to an arbitrary scale.

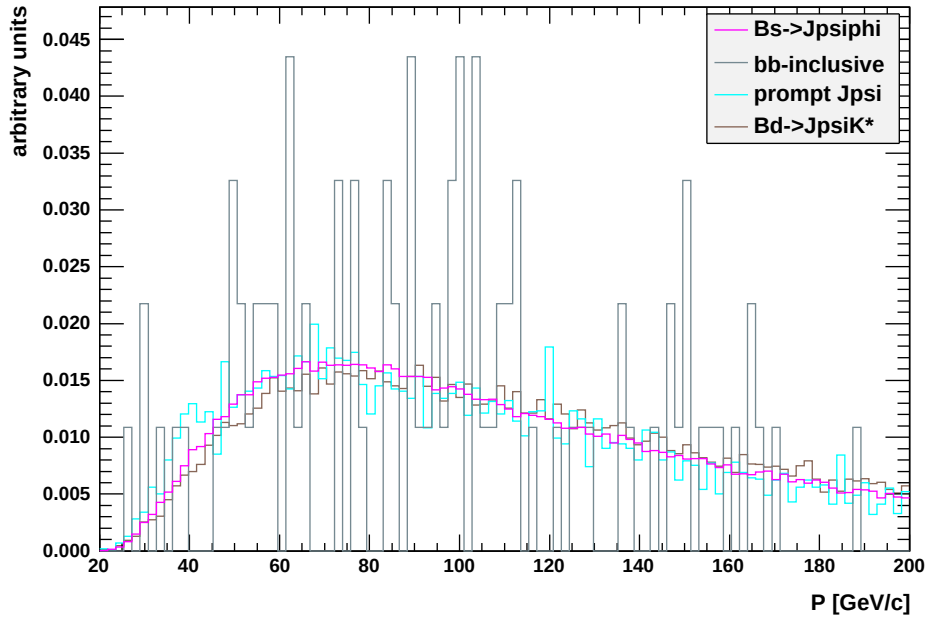


(a) B_s mass before cuts

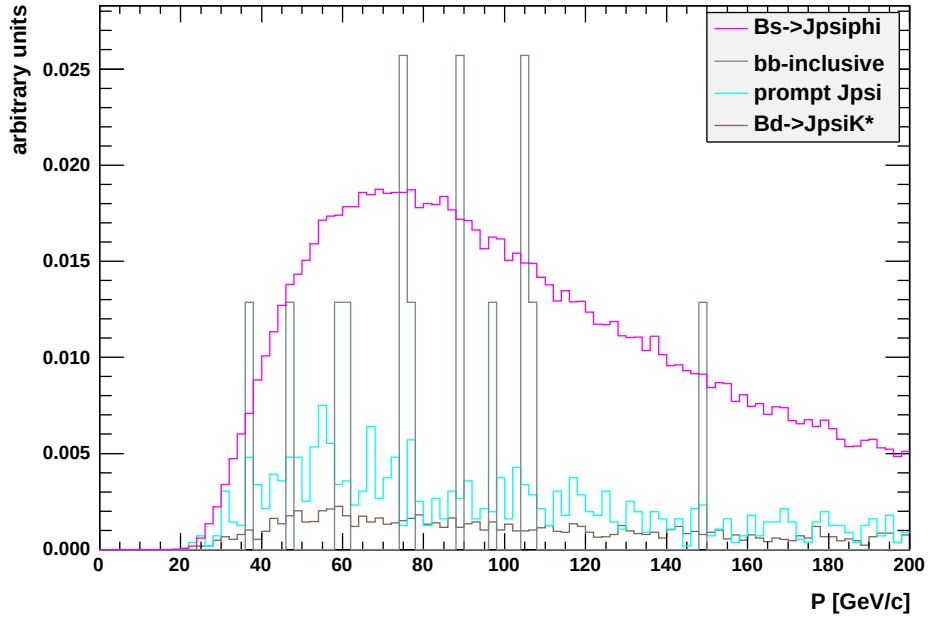


(b) B_s mass after all cuts

Figure 4.11: (a) Distribution of mass for B_s^0 from $B_s \to J/\psi\phi$ (magenta), $b\bar{b}$ inclusive (grey), prompt J/ψ (cyan) and $B_d \to J/\psi(\mu^+\mu^-)K^{\star 0}$ (brown) data samples before applying the selection cuts. (b) The B_s mass distributions after all cuts (using a wide $\pm$ 500 $\mathrm{MeV/c^2}$ background mass window). The tight signal mass window cut ($\pm$ 50 $\mathrm{MeV/c^2}$) is shown in green. The distributions shown are normalised to an arbitrary scale.



(a) B_s $p_\perp$ before cuts



(b) B_s $p_\perp$ after all cuts

Figure 4.12: (a) Distribution of $p_\perp$ for B_s^0 from $B_s \rightarrow J/\psi\phi$ (magenta), $b\bar{b}$ inclusive (grey), prompt J/ψ (cyan) and $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ (brown) data samples before selection cuts. (b) The B_s $p_\perp$ distributions after all cuts (using a wide $\pm$ 500 $\mathrm{MeV/c^2}$ mass window). The distributions shown are normalised to an arbitrary scale.

4.6 Branching fraction and production yield

The LHC will be a source of a copious numbers of b-hadrons. At the machine's design luminosity of $\mathcal{L} \approx 2 \times 10^{32}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ and with an expected $b\bar{b}$ cross-section ($\sigma_{b\bar{b}}$) of 500 μb the annual yield of $b\bar{b}$ pairs at the LHCb will reach approximately 10^{12} $b\bar{b}$ pairs:

$$N_{b\bar{b}} \approx \mathcal{L} \cdot \sigma_{b\bar{b}} \cdot 10^{7} s \approx 10^{12}. \tag{4.3}$$

The annual production yield of a certain decay channel at LHCb is obtained using the following equation:

$$N^{exp}_{signal} = 2 \times \sigma_{b\bar{b}} \times \mathcal{L}^{int}_{year} \times BF(\bar{b} \to B^0_s) \times BF_{\mathrm{VIS}} \tag{4.4}$$

Where the factor of 2 deals with possible production of either B^0_s or $\bar{B^0_s}$ mesons from a $b\bar{b}$ pair. The hadronisation probability is BF($\bar{b} \to B^0_s$) = BF($b \to \bar{B^0_s}$) = (10.3 $\pm$ 1.4) $\times 10^{-2}$ [7]. The designed integrated luminosity for one nonimal year - $\mathcal{L}^{int}_{year}$ = $\int \mathcal{L} dt$ - is equal to 2 fb^{-1}. The visible branching fraction BF_{VIS} is the product of the individual branching fractions associated with the decay. For $B_s \to J/\psi\phi$ this is $(3.1 \pm 1.2) \times 10^{-5}$ and is calculated from:

$$\begin{aligned} BF_{VIS} &= \mathrm{BF}(\mathrm{B^0_s} \to \mathrm{J}/\psi\phi) \times \mathrm{BF}(\mathrm{J}/\psi \to \mu^+\mu^-(\gamma)) \times \mathrm{BF}(\phi \to \mathrm{K^+K^-}) \\ &= (3.1 \pm 1.2) \times 10^{-5} \end{aligned} \tag{4.5}$$

Where,

$$\begin{aligned} BF(B^0_s \to J/\psi\phi) &= (9.3 \pm 3.3) \times 10^{-4} \\ BF(\phi \to K^+K^-) &= (49.0 \pm 0.6) \times 10^{-2} \\ BF(J/\psi \to \mu^+\mu^-) &= (5.93 \pm 0.1) \times 10^{-2} \\ BF(J/\psi \to \mu^+\mu^-\gamma) &= (0.88 \pm 0.14) \times 10^{-2} \\ BF(J/\psi \to \mu^+\mu^-(\gamma)) &= BF(J/\psi \to \mu^+\mu^-) + BF(J/\psi \to \mu^+\mu^-\gamma) \\ &= (6.81 \pm 1.10) \times 10^{-2} \end{aligned} \tag{4.6}$$

The expected signal production for $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ is therefore $(6.2 \pm 2.0) \times 10^6$ decays per year: in other words, about 1 in 1.7×10^5 $b\bar{b}$ pairs per year will decay as $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$. Taking into account that only a third of the $b\bar{b}$ pairs are within the LHCb acceptance, the number of $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ events available for reconstruction is $\sim 2.0 \times 10^6$.

4.7 Signal yield and efficiencies

This section describes the performance of the $B_s \rightarrow J/\psi\phi$ selection on signal data. For this discussion, the following definitions for a given event are introduced:

- $N^{\text{signal}}_{\text{gen } 4\pi}$: total number of generated Monte Carlo signal events within 4π.
- $N^{\text{signal}}_{\text{gen}}$: total number of generated Monte Carlo signal events within the 400 mrad geometric acceptance.
- $N_{\text{,ted}}$: number of reconstructed events, with all final state particles reconstructed as long tracks.
- $N^{\text{signal}}_{\text{sel}}$: number of off-line signal events after the final selection cuts, but before trigger.

Using the above definitions we can define the total selection efficiency before trigger by:

$$\epsilon_{\text{tot}} = \epsilon^{\theta}_{J/\psi\phi} \times \epsilon_{\text{sig}} = \frac{N^{\text{signal}}_{\text{gen}}}{N^{\text{signal}}_{\text{gen } 4\pi}} \times \frac{N^{\text{signal}}_{\text{sel}}}{N^{\text{signal}}_{\text{gen}}}, \tag{4.7}$$

where $\epsilon^{\theta}_{J/\psi\phi}$ (= 18$\pm$0.006 %) is the fraction of generated $B_s \rightarrow J/\psi\phi$ signal events within the 400 mrad generator-level acceptance of the detector. ϵ_{sig} is the number of reconstructed signal events over the number of generated events.

The above efficiencies and numbers for the reconstructed events are given in Table 4.7, where the calculated uncertainties are statistical[7] and take into account the fraction of independent events.

Channel	$N^{\text{signal}}_{\text{gen}}$	$N_{\text{,ted}}$	$N^{\text{signal}}_{\text{sel}}$	ϵ_{sig}[%]	ϵ_{tot}[%]
$B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$	1.6M	157k	142k	10.9 $\pm$ 0.003	1.98 $\pm$ 0.00065

Table 4.7: Reconstruction, off-line selection numbers and selection efficiencies for $B_s \rightarrow J/\psi\phi$, before the trigger. The uncertainties are statistical.

The annual signal yield before trigger can then be calculated using,

$$\begin{aligned} N_{signal} &= 2 \times \sigma_{b\bar{b}} \times \mathcal{L}^{int}_{year} \times BF(\bar{b} \rightarrow B^0_s) \times BF_{VIS} \times \epsilon_{tot} \\ &= 132680. \end{aligned} \tag{4.8}$$

We therefore expect at LHCb running at the nominal luminosity of $2fb^{-1}$, appoximately 133k $B_s \rightarrow J/\psi\phi$ signal events before applying the trigger. This number is lower than $N^{\text{signal}}_{\text{sel}}$

[7]The error on an efficiency ϵ=$\frac{n}{N}$ within a sample of independent events is given by σ_ϵ=$\sqrt{\frac{\epsilon(1-\epsilon)}{N}}$, which is often called the binomial error. The statistical error on ϵ_{tot} is computed using the first equality in (4.7).

given in Table 4.7, which is due to $N_{\text{gen}}^{\text{signal}}$ (in Table 4.7) being generated for 2.5 fb^{-1} of data. In this case, only the efficiency numbers given in Table 4.7 are meaningful and will be used to calculate the background-to-signal ratios in chapter 5.

4.8 Background contribution

In order to achieve the optimal background-to-signal ratio – $\frac{B}{S}$ – for this channel, an appreciation of the background types which can contaminate, or fake, our selection cuts is required. This is important since we have seen above that one 1 in 1.7×10^5 $b\bar{b}$ pairs are expected to decay as $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$. The background can be distinguished into three categories:

i. Combinatorial background: random combinations of tracks.

ii. Prompt background: particles produced at the primary vertex, in particular prompt J/ψ.

iii. b-inclusive background: b-decays that have the same topology and/or final state particles as the decay of interest. For the $B_s \to J/\psi\phi$ channel, decays of the type $B_d \to J/\psi(\mu^+\mu^-)K^{\star 0}$ have the possibility of passing through the pre-selection and faking $B_s \to J/\psi\phi$ events. For early physics measurements with $B_s \to J/\psi\phi$ this source of background may well be important, for instance if the $\pi - K$ miss-identification rate of the RICH system is not operating as expected from DC06 data. Other $K_{S,L}$ channels including, $B_s \to J/\psi(\mu^+\mu^-)K_{S/L}(\pi^+\pi^-)$, will hardly contaminate the $B_s \to J/\psi\phi$ channel since the $K_{S,L}$ fly a long distance. The decay lengths being $c\tau_S \sim 2.68$ cm and $c\tau_L \sim 15.51$ cm [7] respectively, which means that these events will lack the characteristic four track vertex of the $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ channel.

Fortunately, its possible to distinguish the above background types based on the topology and kinematics of the $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay. The first two backgrounds categorised can be reduced by exploiting the following characteristic of B_s^0-meson decays:

- A displaced secondary vertex due to the long B_s^0 lifetime. It has already been noted that the average decay length of the B_s^0 is about 1.2 cm. This is in contrast to prompt decays which by definition decay at the interaction point.

- Final state particles of the B_s-meson decay have typically higher $p_\perp$ than prompt particles or final state particles from non-b-decays.

Unfortunately the third category of background, b-inclusive events, carry the same properties as the B_s-meson decays. They are more difficult to remove. A full discussion on the different background samples can be found in chapter 5.

4.9 Selection summary

This section summarises the most important results of the $B_s \rightarrow J/\psi\phi$ off-line selection to fully simulated Monte Carlo data.

B_s mass resolution: Using the core Gaussian, $\sigma_m^{res}(B_s^0)$ = (15.003$\pm$ 0.1) $\mathrm{MeV/c^2}$, without applying a mass constrained vertex fit to the daughters; with a mass constrained fit this becomes, $\sigma_m^{res}(B_s^0)$ = (6.78$\pm$ 0.023) $\mathrm{MeV/c^2}$.

Annual event yield: The expected annual yield (for $2\mathrm{fb}^{-1}$) before L0 and HLT triggers is 133k events. After applying the L0 trigger this is expected to be 122k events and 99.4k after applying the global L0 and HLT decision.

The remaining issue we need to discuss, with regards the off-line selection, is the different background contributions to the $B_s \rightarrow J/\psi\phi$ signal channel. There are two main sources of background for this channel: $b\bar{b}$-inclusive and prompt J/ψ. Both will investigated in the following chapter.

5

$B_s \rightarrow J/\psi\phi$ Background studies

The off-line selection cuts, detailed in section 4.5, reduce the original data samples to a more manageable number of events to be used by the MultiVariate Analysis (MVA) classifiers in chapter 7. Before moving on to optimise the selection cuts with these classifiers, we need to identify what the most relevant sources of background are for the $B_s \rightarrow J/\psi\phi$ channel. The strategy in this chapter will be to focus on the $b\bar{b}$-inclusive data sample, which has previously been discussed as the main source of background for this channel [74, 75] and indeed most $\mathcal{CP}$ violating channels at LHCb. We will also investigate specific sources of background to our off-line selection, including prompt-J/ψ and $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ decays.

The impact of the prompt-J/ψ background sample on our selection cuts will certainly be of interest, and during these studies was found to be a major contributor to the total $\frac{B}{S}$ ratio. The $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ data sample has been included in our studies since this decay is topologically similar to the signal channel. It may well contribute to a source of low mass background for this analysis, especially if the π-K separation efficiencies of the RICH is not optimal.

Before calculating the $\frac{B}{S}$ ratio, the individually selected background events must be inspected in more detail. To do this we make use of an enlarged mass window, $\mathcal{F}_{mass}$. The approach of using a large mass window to study the background is commonly seen in LHCb analysis studies [76]. An example of such a reconstructed event is $B_s \rightarrow J/\psi(\mu^+\mu^-,\gamma)\phi(K^+K^-)$ where the radiative γ is used in the invariant mass calculation. The mass of this reconstructed B-meson will always be on the lower side of the mass window. To classify the reconstructed background events, use will be made of a custom designed background classification tool [1] [77].

The background classification tool (BCT) categorises events into one of twelve categories, these are summarised below[2]:

[1] The background classification tool is effectively a "labour" saving device useful to gain a rough idea of the background composition before selecting specific background events for closer investigation. The tool preforms classification of the reconstructed background event by assigning it one of twelve classifications, which include combinatoric, ghost, mis-identification, etc.

[2] Given in the parenthesis, is the numerical code for each background category returned by the BCT tool.

Signal (cat: 0): An event reconstructed as signal is required to fulfil four conditions:

- **(a)** Each final state particles is associated to a MC particle, where both the final state and MC particle have the same PIDs.
- **(b)** All such MC associated final state particles have a common mother, which itself is MC associated.
- **(c)** Daughters of the MC mother particle must be MC associated and must correspond to the final state particles in (a).
- **(d)** The MC associated mother in (b) must have the same PID as the mother in (b).

Fully reconstructed decay (cat: 20): This is where the decay has been topologically and correctly reconstructed, but the mother particle has been incorrectly identified.

Reflections (cat: 30): This is background caused by one of the final state particles being misidentified, for example a pion instead of a kaon.

Partially Reconstructed decay (cat: 40): These are potentially dangerous sources of background as they can appear under the signal peak. It applies to the case when a fragment of the decay is incorrectly identified as a signal decay. This can involve missing and miss identified particles.

Low mass background (cat: 50): The is a subset of Partially reconstructed decays above, where there are no misidentified particles, only missing particles. Hence, these reconstructed events systematically fall below the signal peak.

Ghost background (cat: 60): This is any reconstructed event in which one or more of the final state particles id found to have no MC association.

Primary vertex background (cat: 70): Any reconstructed decay in which one or more final state particles are found to come from the primary vertex is classified in this category. Impact parameter cuts protect with respect to the PV should help reduce this background. This category will be a dangerous source of background whilst whilst running over prompt-J/ψ data.

Badly reconstructed primary vertex (cat: 80): These are decays which have originated from an incorrectly reconstructed primary vertex. The particles which originate for these vertices will have high momenta and appear as displaced vertices. Since these are two characteristics of our B_s candidate, event with this classification need to be scrutinised.

Pile up background (cat: 100): These are decay in which the final state particles are found to come from more than one primary vertex.

b $\bar{\text{b}}$ background (cat: 110): These are decay which do not satisfy the above criteria, but have at least one final state particle coming from mother containing a bottom quark.

c $\bar{c}$ background (cat: 120): These are decay which do satisfy the above criteria, but have at least one final state particle coming from a mother containing a charm quark.

uds background (cat: 130): This the "default", any event which does not fit into the above classifications.

To calculate the $\frac{B}{S}$ fraction for each samples under investigation, the procedure will be to consider all candidates which lie within a loose $\pm 500\ \mathrm{MeV/c^2}$ B_s mass window. Use is then made of the BCT, the decay chain structure of the reconstructed candidate and whether-or-not the event is unique. The BCT will allow us to separate signal events, which will not be considered, from physical (codes 20-50 above) and technical (codes 60-130) backgrounds. The technical backgrounds which encompass the definition of combinatorial events, will be considered. Finally, to ensure each considered candidate is unique, each reconstructed event is subjected to a test indicating whether it is considered a clone[3] on not. A cloned event occurs when an event contains more than one reconstructed candidate which have essentially identical momenta[4]. In this case the proceedure is to take one of the candidates, the one with a mass lying closest to the nominal B_s mass. In order to investigate the effectiveness of the clone proceedure for the $b\bar{b}$-inclusive sample, use will be made of the best χ^2 criteria. In this criteria if an event contains more than one reconstructed candidate, the one with the smallest vertex χ^2 is used.

5.1 $b\bar{b}$-inclusive background

We write the expression for $\frac{B}{S}$ as [7]:

$$\left(\frac{B}{S}\right) = \frac{1}{2 \cdot \mathrm{BF}(\bar{\mathrm{b}} \to \mathrm{B_s^0}) \cdot \mathcal{BF}_{\mathrm{total}}^{\mathrm{signal}}} \times \frac{\epsilon_\theta^{b\bar{b}}}{\epsilon_\theta^{\mathrm{signal}}} \times \frac{\epsilon_{\mathrm{sel}}^{\mathrm{bkg}}}{\epsilon_{\mathrm{sel}}^{\mathrm{signal}}} \times \mathcal{F}_{mass}, \tag{5.1}$$

Where we have:

-. $\epsilon_\theta^{b\bar{b}}$ and $\epsilon_\theta^{\mathrm{signal}}$ are the efficiencies of the $b\bar{b}$-inclusive and signal channels respectively being found within the 400 mrad acceptance requirement of the detector;

-. $\mathrm{BF}(\bar{\mathrm{b}} \to \mathrm{B_s^0})$ = 10.7% $\pm$ 1.4 % is the $\bar{b} \to B_s^0$ production fraction. The factor 2 takes into account the possible production of both B_s^0 and $\overline{B_s^0}$ mesons from the $b\bar{b}$ pair.

-. $\mathcal{BF}_{total}^{signal}$ is the total branching fraction of the B_s^0 decay into the final state.

[3] An event containing a clone couple is one where more than 70% of the Inner Tracker (IT) and Outer Tracker (OT) hits of at least one track are used in both tracks [78]; essentially, we have more than one candidate reconstructed per event, where two (or more) of these candidates are reconstructed with almost identical momentum. What happens is that the tracking procedure has built two tracks from almost the same number of hits (i.e. the same MC particle) and these two tracks are nearly identical. But since their are two tracks, assume for illustration two positive kaons (K_1^+ and K_2^+), the tracking software will build two ϕ, $\phi(K^-,K_1^+)$ and $\phi(K^-,K_2^+)$ and then two B_s, so essentially the same event will be count twice.

[4] The test for clones events will be as follows: If an event contains two (or more) reconstructed candidates, and these candidates momenta differ by less than 2%, i.e. if $\frac{B_1^{\mathrm{reco}} - B_2^{\mathrm{reco}}}{B_1^{\mathrm{reco}} + B_2^{\mathrm{reco}}} < 0.02$, the event is considered to be a clone.

-. $N_{\text{sel}}^{\text{bkg}}$ is the number of selected background events in the tight mass window of $\pm$ 50 MeV/c^2. In the case no background events are selected, the number of events within the wider mass window $\pm$ 500 MeV/c^2 will be used. In this case, equation 5.1 requires a mass window enlargement factor $\mathcal{F}_{mass}$;

-. $\epsilon_{\text{sel}}^{\text{bkg}} = \frac{N_{\text{sel}}^{\text{bkg}}}{N_{\text{gen}}^{\text{bkg}}}$ is the efficiency for selecting background events, where $N_{\text{gen}}^{\text{bkg}}$ is the number of generated background events within the fiducial cut.

-. $\epsilon_{\text{sel}}^{\text{signal}} = \frac{N_{\text{sel}}^{\text{signal}}}{N_{\text{gen}}^{\text{signal}}}$ is the efficiency for selecting signal events, where $N_{\text{sel}}^{\text{signal}}$ is the number of signal events selected and $N_{\text{gen}}^{\text{signal}}$ is the number of generated signal events within the fiducial cut.

-. $\mathcal{F}_{mass} = \frac{\Delta M^{\text{tight}}}{\Delta M^{\text{loose}}} = \frac{100}{1000} = 0.1$ is the mass window enlargement factor for background events. The large mass window is used to allow enough background events to pass the selection cuts in order have enough statistics to study them.

Running over the $b\bar{b}$-inclusive data sample (Table 3.2) with an enlarged mass window of $\pm$500 MeV/c^2 around the nominal B_s mass, 33 candidates out of 6M pass the $B_s \rightarrow J/\psi\phi$ selection. Of these candidates, 6 lie within the tight $\pm$50 MeV/c^2 signal window, 16 lie above and 11 lie below the tight mass window. After taking into account clones and signal candidates, a total of 14 will be considered. A detailed breakdown of this selection process is given in section C.1.

$b\bar{b}$-inclusive selection summary

As expected and discussed above, the majority of events reconstructed from the bb-inclusive data sample are combinatorial or ghost like in nature. We now summarise the results of running the signal selection cuts on the $b\bar{b}$-inclusive data sample.

- From the 6M stripped events[5]

  33 candidates were reconstructed:

  - Of the reconstructed candidates:
  - $\sim$ 3% of the events are signal.
  - $\sim$ 39% are ghost.
  - $\sim$ 12% of the events contained clones.
  - The other major contribution, $\sim$ 42%, coming from badly reconstructed primaries.
  - $J/\psi(1S)$ and $\phi(1210)$ reconstructed candidates:

[5]The $b\bar{b}$-inclusive data sample analysed has been stripped. This is where the data is first run over the HLT pre-selection algorithms [79] which for instance, require the event to contain a heavy di-muon pair. This process reduces the $b\bar{b}$ inclusive sample down from an initial 21M to 800,000 events. This final number of stripped events corresponds to $\sim$ 500 seconds of data taking at LHCb: $t_{b\bar{b}} = \frac{N_{\text{stripped}}}{N_{\text{events/second}}} \times \frac{1}{\epsilon_{\text{stripped}}\epsilon_{\text{gen}}\epsilon_{\text{LHCb}}}$.

- $\sim$ 9% of the events contained true $\phi(1210) \rightarrow K^+K^-$ decays, $\sim$ 3% of the events contained $\phi(1020)$ from B_s, 15% from prompt $\phi(1210)$ and $\sim$ 72% of the events did not contain a $\phi(1210)$

- $\sim$ 3% of the events contained true $J/\psi(1S) \rightarrow \mu^+\mu^-$ decays, $\sim$ 6% of the events contained $J/\psi(1S)$ from B_s, 9% from B_u, $\sim$ 6% from B_d and $\sim$ 76% of the events did not contain a $J/\psi(1S)$

- Final states:

- $\sim$ 51% final state K^- are identified as K^-, $\sim$ 6% were mis-identified as π^- and $\bar{P}$, $\sim$ 3% as μ^- and 33% had no association.

- $\sim$ 36% final state K^+ are identified as K^+, $\sim$ 24% as π^+, $\sim$ 9% as a proton (P) and $\sim$ 30% have no association.

- $\sim$ 45% final state μ^- are identified as μ^-, $\sim$ 12% as K^-, $\sim$ 6% as π^- and $\sim$ 36% are not identified.

- $\sim$ 39% final state μ^+ are identified as μ^+, $\sim$ 12% as K^+, $\sim$ 15% as π^+ and $\sim$ 33% are not identified.

From the background event yield in each mass region, the background selection efficiencies can be computed. From the 4 background events found within the tight mass window, we compute a background selection efficiency, $\epsilon_{\text{sel}}^{\text{bkg}}$, of $\sim$(4.53 $\pm$ 0.56 $\times$)10^{-7} is found. With the expected event yields for signal and background and using equation (5.1) an estimate of the background-to-signal ratio in the signal mass region can be made:

$$\left(\frac{B}{S}\right)^{b\bar{b}-\text{inclusive}}_{B_s\rightarrow J/\psi\phi} = \underbrace{0.14 \pm 0.056}_{\text{sig. win}}. \tag{5.2}$$

Repeating the above exercise, but replacing the clone criteria with the best vertex χ^2 criteria, it can be seen that out of the 6M events analysed, 29 candidates now pass the selection criteria, such that:

- Three events lie within the $\pm 50\ \mathrm{MeV/c^2}$ signal window:

- One candidate is classified as signal, and is neglected.

- 12 candidates are found below the $\pm 50\ \mathrm{MeV/c^2}$ signal window:

- One candidate is classified as low mass background, and is neglected.

- 14 candidates lie above the $\pm 50\ \mathrm{MeV/c^2}$ signal window:

Table 5.1 summarises the $\frac{B}{S}$ ratios using both filter criteria with $b\bar{b}$-inclusive data.

Filter criteria applied	tight signal region
	Number of candidates
clone	4
vertex χ^2	2
	$(\frac{B}{S})$
clone	0.07 ± 0.028
vertex χ^2	0.14 ± 0.056

Table 5.1: The $\frac{B}{S}$ ratios with $b\bar{b}$ background data using the clone and χ^2 filter criteria.

If we consider all 33 candidates lying within the enlarged mass window, and neglecting only events identified as signal (i.e 1 candidate) the ratio becomes:

$$\left(\frac{B}{S}\right)^{b\bar{b}-\text{inclusive}}_{B_s\rightarrow J/\psi\phi} = 1.17 \pm 0.445 \tag{5.3}$$

5.2 Specific background

The selection code was also run over specific background samples. Prompt-J/ψ decays poses a difficult problem for selection of $B_s \rightarrow J/\psi\phi$ signal events. Their contribution can be limited by introducing a propertime significance cut on the reconstructed B_s, as investigated in [74]. Another potential source specific background comes from $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ which has a similar event topology to our signal. Calculation of the $\frac{B}{S}$ ratio with respect to $B_d \rightarrow J/\psi(\mu^+\mu^-)K_s$ can be found in section C.3.

prompt-J/ψ background

Prompt J/ψ, that is J/ψ's that do not originate from the b hadron, pose a specific source of background for the $B_s \rightarrow J/\psi\phi$ reconstruction. These decays have the potential of producing detached vertices that could be reconstructed as fake $B_s \rightarrow J/\psi\phi$ combinations.

The EVTGEN data sample used, see Table 3.2, is generated with an admixture of prompt J/ψ and J/ψ coming from b hadrons. The contribution from J/ψ originating from b hadrons is already included in the $\frac{B}{S}$ ratio obtained from the $b\bar{b}$-inclusive data sample. Therefore, assuming that the fraction of J/ψ from b decays in the $b\bar{b}$-inclusive sample is the same in the present sample, we can estimate the fraction of prompt J/ψ using the following calculation:

$$\begin{aligned}
\sigma^{incl}_{b\bar{b}} &= 0.627mb, \quad \sigma^{pr}_{J/\psi} = 0.313mb, \\
\mathrm{BF(b \to J/\psi X)} &= 1\%, \\
\sigma(b\bar{b} \to J/\psi X) &= 2 \times \sigma^{incl.}_{b\bar{b}} \times \mathrm{BF(b\bar{b} \to J/\psi X)} = 0.0125\mathrm{mb}, \\
\Rightarrow \sigma(c\bar{c} \to J/\psi X) &= \sigma^{pr}_{J/\psi} - \sigma(b\bar{b} \to J/\psi X) = 0.301mb, \\
\Rightarrow \mathcal{F}^{J/\psi_{pr}} &= \frac{\sigma(c\bar{c} \to J/\psi X)}{\sigma^{pr}_{J/\psi}} = 96\% \qquad (5.4)
\end{aligned}$$

Where in the above calculation used was made of the measured PYTHIA cross sections from generation [65] and the branching ratios [7]. The $\frac{B}{S}$ ratio can be calculated using:

$$\left(\frac{B}{S}\right) = \frac{\sigma^{pr}_{J/\psi}}{\sigma_{bb}} \times \frac{\mathcal{F}^{J/\psi_{pr}} \times \mathcal{BF}^{J/\psi_{pr}}_{total}}{2 \times \mathrm{BF(\bar{b} \to B^0_s)} \times \mathcal{BF}^{\mathrm{signal}}_{\mathrm{total}}} \times \frac{\epsilon^{J/\psi}_{\theta}}{\epsilon^{\mathrm{signal}}_{\theta}} \times \frac{\epsilon^{spec}_{sel}}{\epsilon^{\mathrm{signal}}_{\mathrm{sel}}} \times \mathcal{F}_{mass}. \qquad (5.5)$$

Where:

-. $\epsilon^{J/\psi}_{\theta}$ = 19.6% is the fraction of J/ψ-inclusive events lying in the 400 mrad acceptance.

-. $\sigma^{pr}_{J/\psi}$ = 0.313 mb, is the prompt J/ψ cross-section.

-. σ_{bb} = 0.627 mb, is the $b\bar{b}$-inclusive cross-section.

-. $\mathcal{F}^{J/\psi_{pr}}$ = 96% is the fraction of prompt J/ψ occurring in the J/ψ-inclusive data sample.

-. $\mathcal{BF}^{J/\psi_{pr}}_{total} = \mathrm{BF(J/\psi \to \mu^+\mu^-, \gamma)} = \mathrm{BF(J/\psi \to \mu^+\mu^-)} + \mathrm{BF(J/\psi \to \mu^+\mu^-\gamma)} = (6.81 \pm 1.10) \times 10^{-2}$

Running over the J/ψ-inclusive data sample with an enlarged mass window of $\pm$500 $\mathrm{MeV/c^2}$ around the nominal B_s mass, 939 candidates out of 1.8M analysed events were selected. Of these candidates, a total of 99 were found below the mass window, 278 were found within the $\pm$ 50 $\mathrm{MeV/c}$ window and 343 candidates lying above the mass window. A full description for the events lying in each region can be obtained from section C.2.

J/ψ-inclusive selection summary

- From the 1.8M analysed J/ψ-inclusive events 939 candidates where reconstructed:
  - Of the reconstructed candidates:
  - $\sim$ 7% of the reconstructed events were considered signal.
  - $\sim$ 14% reconstructed events where considered as ghosts.
  - $\sim$ 63% events where considered to have one or more final state coming from the same primary as the reconstructed candidate.

- $\sim$ 20% of the events contain cloned candidates.

• $J/\psi(1S)$ and $\phi(1210)$ reconstructed candidates:

- $\sim$ 0.3% of the events contained true $\phi(1210) \rightarrow K^+K^-$ decays, $\sim$ 21% contained a $\phi(1020)$ coming from B_s, 53% from prompt $\phi(1210)$ and $\sim$ 67% of the events did not contain a $\phi(1210)$

- $\sim$ 1% of the events contained true $J/\psi(1S) \rightarrow \mu^+\mu^-$ decays, $\sim$ 10% of the $J/\psi(1S)$ came from B_s, 5% from B_u, $\sim$ 4% from B_d, $\sim$75% from prompt J/ψ and $\sim$ 5% of the events did not contain a $J/\psi(1S)$

• Final states:

- $\sim$ 49% final state K^- are identified as K^-, $\sim$ 12% were mis-identified as π^-, 7% as a positron, $\sim$ 6% as μ^- and 26% had no association.

- $\sim$ 51% final state K^+ are identified as K^+, $\sim$ 8% as π^+ and as a proton and $\sim$ 25% have no association.

- $\sim$ 97% final state μ^- are identified as μ^-, $\sim$ 0.2% as K^-, $\sim$ 0.7% as π^- and $\sim$ 1.6% are not identified.

- $\sim$ 98% final state μ^+ are identified as μ^+, $\sim$ 0.1% as K^+ and as a proton, $\sim$ 0.3% as π^+ and $\sim$ 1.9% are not identified.

The $\frac{B}{S}$ ratio for each mass region can now be calculated. Using the 99 background events found within the tight mass window, we calculate the background selection efficiency to be $\sim$(5.5 $\pm$ 0.55 $\times$)10^{-5}. With the expected event yields for signal and background and using equation (5.5) an estimate of the background-to-signal ratio in the tight mass region can be made:

$$\left(\frac{B}{S}\right)^{J/\psi-\text{inclusive}}_{B_s\rightarrow J/\psi\phi} = \underbrace{0.31 \pm 0.12}_{\text{sig. win}}. \tag{5.6}$$

If we now repeat the above exercise, but replace the clone criteria with the best vertex χ^2 the following is observed. Out of the 1.8M events analysed, 764 candidates pass the selection criteria:

• 149 events within the $\pm$50 $\mathrm{MeV/c^2}$ window:

- 23 events are considered as signal $B_s \rightarrow J/\psi\phi$ events, which will not be included as background.

- 4 candidates occur from partially reconstructed primary vertices.

- 11 candidates are considered as ghosts.

- 9 candidates are considered to have final states coming from a primary.

- 95 are considered event where a final state has originated from the same primary as the candidate.

- 5 candidates where the a final state and candidate come from different primaries.
- 2 are considered as $b\bar{b}$ background.

- 297 events are found to lie below the $\pm 50\ \mathrm{MeV/c^2}$ mass window:
  - 19 signal events below the mass window, which will not be considered.
  - 2 events are considered as reflections.
  - 2 events are considered as originating for partially reconstructed primary vertices.
  - 7 low mass: not considered as combinatorial background.
  - 26 events are considered as ghost.
  - 14 events where at least one final state originate from a primary vertex.
  - 220 events are considered to have final states coming from the same primary vertex.
  - 4 events are considered to have final states coming from different primary vertices.
  - 3 events are consider as $b\bar{b}$ background events.

- 378 events are above the $\pm 50\ \mathrm{MeV/c^2}$ mass window:
  - 21 signal events below the mass window, which will not be considered:
  - 8 events are considered as partially reconstructed background events.
  - 8 events are regarded as low mass background events.
  - 32 events are considered as ghost.
  - 23 events are regarded as having at least one final state originating from a primary vertex.
  - 211 events are considered where the final state particles originate from the same primary vertex.
  - 11 events with final states coming from different primary vertices.
  - 5 events are considered as $b\bar{b}$ background events.

Table 5.2 summarises the $\frac{B}{S}$ ratios using the various filter criteria with $b\bar{b}$-inclusive data.

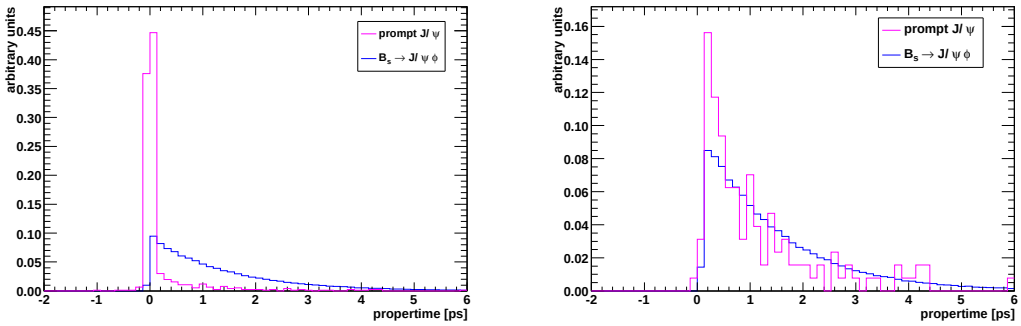
Applied Filter Criteria	tight signal region
	Number of candidates
clone	99
vertex χ^2	149
	$(\frac{B}{S})$
clone	0.13 ± 0.12
vertex χ^2	0.40 ± 0.20

Table 5.2: The $\frac{B}{S}$ ratios with J/ψ-prompt background data using the clone and 2χ filter criteria.

If we consider candidates selected via the χ^2 criteria lying within the the enlarged mass window and neglect only signal events (62) the $\frac{B}{S}$ ratio becomes:

$$\left(\frac{B}{S}\right)^{J/\psi-\text{inclusive}}_{B_s\rightarrow J/\psi\phi} = 1.9 \pm 1.14 \tag{5.7}$$

It is clear this ratio is rather large. The problem is the events concentrated at very short lifetimes as shown in Figure 5.1, which can be removed by applying a propertime significance cut as done in DC04 studies [67]. The advantage of not applying this IP cut is that no properime acceptance function[6] will be needed when determining the sensitivity to $\sin(2\beta_s)$, helping reduce systematic effects.



(a) no IP cut (b) 5σ IP cut

Figure 5.1: (a) shows the propertime distribution of events passing the selection criteria (Table 4.5). Blue is for $B_s \rightarrow J/\psi\phi$ and magenta for prompt $J/\psi \rightarrow \mu^+\mu^-$ events. (b) shows the propertime distribution of events passing the selection cuts at a 5σ IP cut. The signal events are normalised to the number of events in the prompt $J/\psi \rightarrow \mu^+\mu^-$ sample.

[6] In chapter 6, studies using a simple top hat propertime acceptance function showed no deterioration to our ϕ sensitivity.

5.3 Background summary

In this section we summarise the effect of applying the offline unbiased selection cuts to full reconstructed sources of background for the $B_s \to J/\psi\phi$ signal data.

$b\bar{b}$-inclusive The nominal background level is:

$$\left(\frac{B}{S}\right)^{b\bar{b}-inclusive}_{B_s\to J/\psi\phi} = 1.17 \pm 0.445.$$

Of the events passing the selection: 39% are ghosts with the other major contribution about 42% coming from badly reconstructed primary vertices. Of the reconstructed events: 51% and 36% of the K^+ and K^- were identified correctly, while 45% and 49% of the μ^+ and μ^- were correctly identified.

prompt-J/ψ The nominal background level is:

$$\left(\frac{B}{S}\right)^{\text{prompt}-J/\psi}_{B_s\to J/\psi\phi} = 1.9 \pm 1.14$$

Of the events passing the selection: 14% were considered as ghosts, while 63% were consider to have one or more states coming from the same primary vertex. Of the reconstructed events: 49% and 51% are correctly reconstructed as K^+ and K^-, while 97% and 98% of the μ^+ and μ^- are correctly reconstructed.

Background Level: Using a lifetime unbiased selection therefore, we find the total nominal background level is: $\frac{B}{S}=\frac{B}{S}|_{b\bar{b}}+\frac{B}{S}|_{\text{prompt-}J/\psi}+\frac{B}{S}|_{B_d\to J/\psi(\mu^+\mu^-)K_s}$ =1.2 + 1.9 + 9.6$\times 10^{-3}$=3.11, using the central values from the $b\bar{b}$, the prompt-J/ψ and the $B_d \to J/\psi(\mu^+\mu^-)K_s$ contributions.

We have now discussed the performance of our selection cuts on fully reconstructed signal and background data. The next topic to discuss is our expected sensitivity to the physics parameters of interest namely: $-2\beta_s$ and $\Delta\Gamma_s$, using as input the knowledge we have gained over the previous chapters.

Upon completing the sensitivity studies, the final topic of discussion will be to use the prominent background contribution from prompt-J/ψ's to help optimise the selection cuts using multivariate analysis classifiers. One particular classifier, the Boosted Decision Tree (BDT), will be extensively studied for this purpose: the variation in sensitivity with respect to classifier output will also be studied.

6

Sensitivity studies with $B_s \to J/\psi\phi$

6.1 Introduction

The $B_s \to J/\psi\phi$ channel provides us the opportunity to determine the B_s mixing phase $-2\beta_s$ and the decay width difference $\Delta\Gamma_s$. This chapter presents the sensitivities with which these parameters can be measured at the LHCb. In comparison to previous LHCb studies [75, 80, 81] the novel work of this thesis is to extend to use the full three decay angular analysis. We find using the full differential decay rate expressions (equation (2.82)) a quantitative and qualitative improvement in the precision on all parameters. In addition, we found the mistag fraction $\omega_{\rm tag}$ may be obtained from the data themselves in the full angular analysis. The parameters under investigation in this chapter include:

- $-2\beta_s$: the measurable phase arising as a result of the interference between mixing and decay in the analysis of $B_s \to J/\psi\phi$ events (defined in equation (2.72)).
- $\Delta m_s \equiv M_H - M_L$: this is mass difference between physical heavy (B_H)and light (B_L) states, which was discussed in section 2.3.
- $\Delta\Gamma_s \equiv \Gamma_L - \Gamma_H$: the decay width difference between B_L and B_H, discussed in section 2.3.
- $\bar{\Gamma}_s = \frac{1}{2}(\Gamma_L + \Gamma_H)$: the average decay width.
- $R_\perp$ and R_0: the fraction of $\mathcal{CP}$-odd and $\mathcal{CP}$-even components at time t=0 defined in equation (2.82).
- δ_1 and δ_2 : the $\mathcal{CP}$ conserving strong phases (equation (2.89)).
- $\omega_{\rm tag}$: the B_s and $\bar{B}_s$ mistag fraction (see section 6.2.4) defined as (number of wrong tags)/(total number of tags).

The following studies, use the full angular information available from the three angles θ_{tr}, ϕ_{tr} and θ_ϕ. These angles are formed in the so called transversity basis (hence the subscript tr). A schematic representation of the decay and description for the decay angles can be found in

section 2.6.

The structure of this chapter is as follows: in section 6.2 we present the expected precision the physical observables associated with the $B_s \to J/\psi\phi$ events can be measured. These observables include the propertime and angular resolution and acceptances, and the tagging efficiencies associated with the $B_s \to J/\psi\phi$ decays. Section 6.3 describes the procedures adopted for the simulation of data and the fitting technique. In this section we also describe how to approximate the effects of detector resolution and background contributions. An in depth study is given in section 6.4 into parameter sensitivity between the reduced (equation (A.41)) and full (equation (2.82)) angular analysis. In this section we observe an improvement on the $-2\beta_s$ sensitivity. Section 6.5 describes our sensitivity to $-2\beta_s$ using the full three angular analysis and procedure adopted in section 6.3. Section 6.8 describes additional studies into the sensitivity of $-2\beta_s$ varying the central value of: $\omega_{\rm tag}$, the strong phases, and the $p_\perp$ cut applied to the muons and/or kaons.

6.2 Physics observables associated with $B_s \to J/\psi\phi$ events

6.2.1 Propertime studies

To perform an accurate measurement of the time-dependent $\mathcal{CP}$-asymmetry, it is important to determine the B_s^0 decay time with high precision, in order to resolve the fast oscillating B_s^0 mesons. In general, the reconstructed B_s^0 decay time[1] can be found via [74]:

$$\tau_{rec} = M \times \frac{\vec{p} \cdot \vec{L}}{|\vec{p}^2|}, \tag{6.1}$$

where M is the nominal B_s^0 mass, $\vec{p}$ is the reconstructed momentum vector and $\vec{L}$ the decay length vector. The decay length vector is the difference between the primary ($\vec{x}_{PV}$) and secondary ($\vec{x}_{SV}$) vertices, $\vec{L}$ = $\vec{x}_{SV}$ - $\vec{x}_{PV}$. This section presents the reconstructed B_s propertime, resolution and acceptance effects to be used when fitting for $-2\beta_s$.

Illustrated in Figure 6.1 is the B_s^0 decay time distribution τ_{rec} (top left plot), together with its estimated per-event error τ_{rec}^{error}(top right plot) for events passing the selection stage. As the selection requirements outlined above impose only lifetime unbiased cuts, events with a negative propertime can be observed. This is a consequence of the smearing, or resolution, associated with primary[2] and secondary vertices of the event. In section B.3, we summarised the residuals

[1]The B_s^0 decay time, τ, is related to the B_s^0 meson lifetime in the laboratory system τ_{Lab} via $\tau = \frac{\tau_{Lab}}{\gamma}$, where γ is the relativistic Lorentz factor.

[2]The primary vertex assigned to the candidate B_s^0 is given as the average position of all primary vertices associated to the B_s^0 candidate; where on average, ~ 1.7 primary vertices were found associated to each candidate B_s^0. In previous studies of the B_s^0 propertime from $B_s \to J/\psi\phi$ decays [75], a single primary vertex has been assigned to each candidate using as a selection criteria the primary vertex with the smallest impact parameter significance associated to the B_s^0.

and pulls for each component associated with the B_s momentum, B_s decay vertex and primary vertex.

The good resolution on the primary vertex, $\sigma_{P_z^{res}} \sim 49.9$ μm, is the result of the large multiplicity of tracks. There are ~ 115 tracks per event as shown in Figure 6.2, which are used to determine the primary vertex particularly at large angles [82]. The B_s decay vertex resolution ($\sigma_{S_z^{res}} \sim 260.1$ μm) on the other hand, is significantly poorer since the signal tracks are mainly produced at low angles.

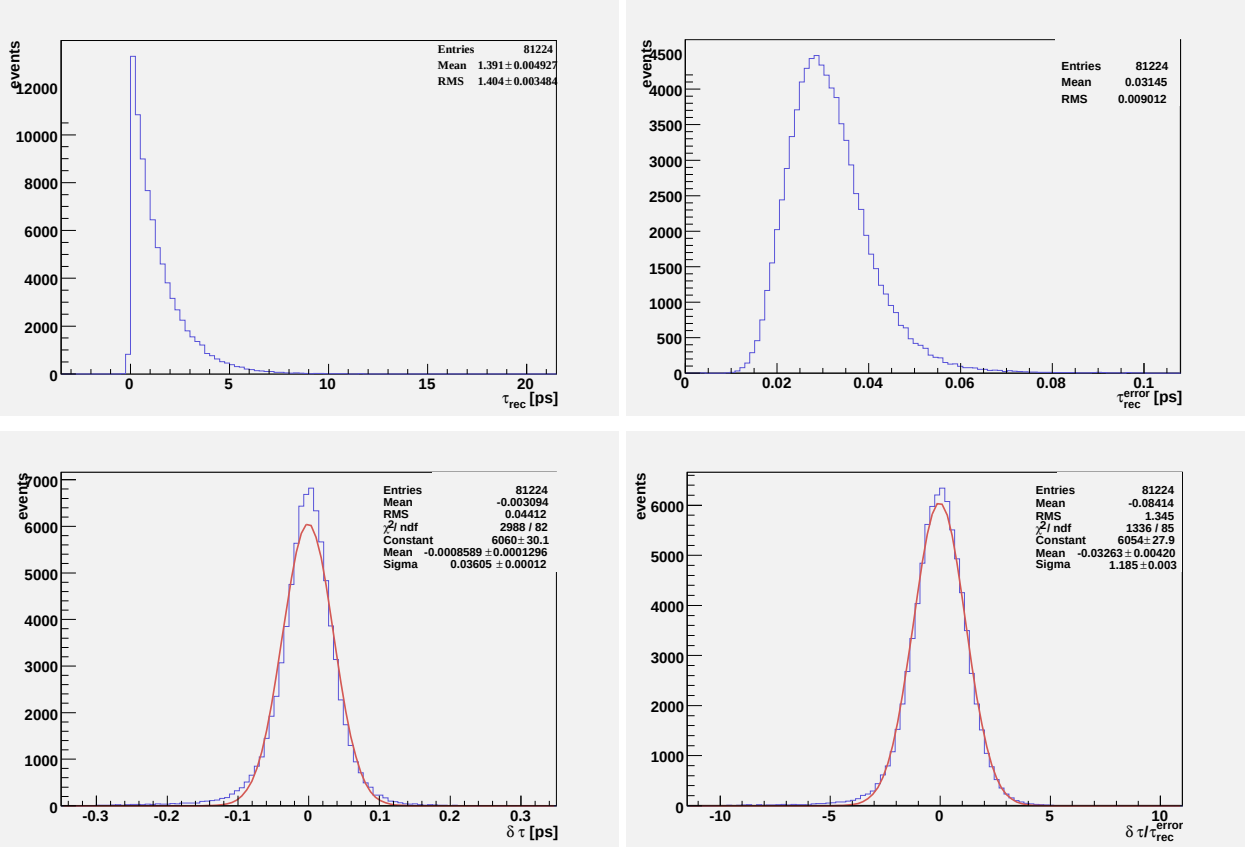


Figure 6.1: The B_s^0 decay time (top left) and its error (top right) obtained from the lifetime fit for pre-selected events. The B_s^0 decay time residual (bottom left) and pull distribution (bottom right) have been fitted with a single Gaussian.

The propertime errors shown in Figure 6.1 vary in a wide range, from 0.01 to 0.1 ps, which is due to the spread in the distribution of the decay distance and the B_s^0 momentum. If the errors were correctly estimated, the B_s^0 propertime resolution would equal the mean value of its error distribution, which is 31.5 fs. However, as shown in the bottom left hand plot in Figure 6.3, the resolution[3] is found to be 36.1 fs; meaning there is $\sim$ 13 % underestimate of the errors.

[3]The propertime residual has also been fitted using a double Gaussian and is found to be 34.0 fs for the core Gaussian, as shown in Figure B.4.

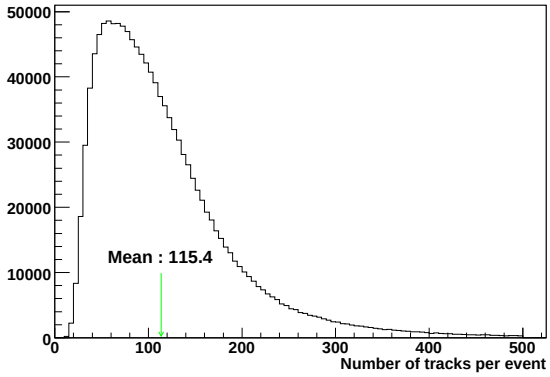


Figure 6.2: Track multiplicity distribution for DC06 generator-level $B_s \rightarrow J/\psi\phi$ signal data. Where the green arrow, at 115.4, indicates the average number of tracks per event.

This effect is confirmed by the pull distribution shown in the bottom right hand plot in Figure 6.3, which shows a postive bias of $\sim$ 19%. The propertime resoltution can then be corrected by introducing a scaling factor of, 1.19. The performance for the $B_s \rightarrow J/\psi\phi$ propertime is therefore:

$$\begin{aligned}
\text{Mean Error} : \tau_{fit}^{err} &= 31.5 fs; \\
\text{Resolution} : \sigma_\tau &= (36.5 \pm 0.12)\ \text{fs}, \quad \text{mean} = (-0.86 \pm 0.13)\ \text{fs}; \\
\text{Pull} : \sum_\tau &= 1.185 \pm 0.003, \quad \text{mean} = (-0.0362 \pm 0.0042); \\
\text{Scaled Resolution} \sum_\tau^{cor} : &= 1.185 \times 31.5 = (37.3 \pm 0.09)\ \text{fs}
\end{aligned}$$

For the B_s propertime, the acceptance effect with the lifetime unbiased (Table 4.5) and biased [74] selection cuts (and trigger) has also been investigated. Results of the lifetime unbiased selection can be found in section B.5. These studies use fully simulated Monte Carlo data to determine the acceptance effects. These effects are parametrised by fitting to the bin-to-bin ratio of the histogram of the propertime after applying selection cuts to the true propertime. The true propertime for these studies was calculated at the generator level, before applying any selection cuts.

Using the lifetime unbiased selection cuts and no trigger, we would therefore expect to see a flat distribution for the B_s propertime acceptance. This is indeed shown in the top left-hand plot in Figure 6.3, which has been fitted using a zeroth order polynomial and shows a relatively flat distribution. However, the fit to the B_s propertime acceptance distribution can be improved by using a 1^{st} order polynomial, as shown in the top right hand plot of Figure 6.3, this fit emphasises a slight negative slope in the propertime acceptance distribution. This slight bias in longer

lived B_s candidates is due to the associated track starting at a high z position in the VELO. The reconstructiblity of these candidates will be worse since the track originating from the B_s will travel through fewer VELO stations, leaving fewer hits in the VELO to reconstruct the track.

The effect of then applying the global L0 trigger decision is shown in the middle left-hand Figure 6.3, where the propertime acceptance has been parametrised using:

$$\mathcal{A}(\tau) = b\left(\frac{(a\tau)^c}{1+(a\tau)^c}\right). \tag{6.2}$$

The effect of applying the HLT and L0 decision is also shown in Figure 6.3, in the middle right-hand plot. This propertime acceptance has been fitted using [83]:

$$\mathcal{A}(\tau') = \left(\frac{\alpha}{1+e^{\beta(\tau')}}\right), \tag{6.3}$$

which was found to give an more accurate parametrisation of the acceptance function than the parametrisation given in equation (6.2). Where in equation (6.3), $\tau' = \tau - \tau^0$. All the parametrisations discussed have been for propertimes in the range [0,6] ps.

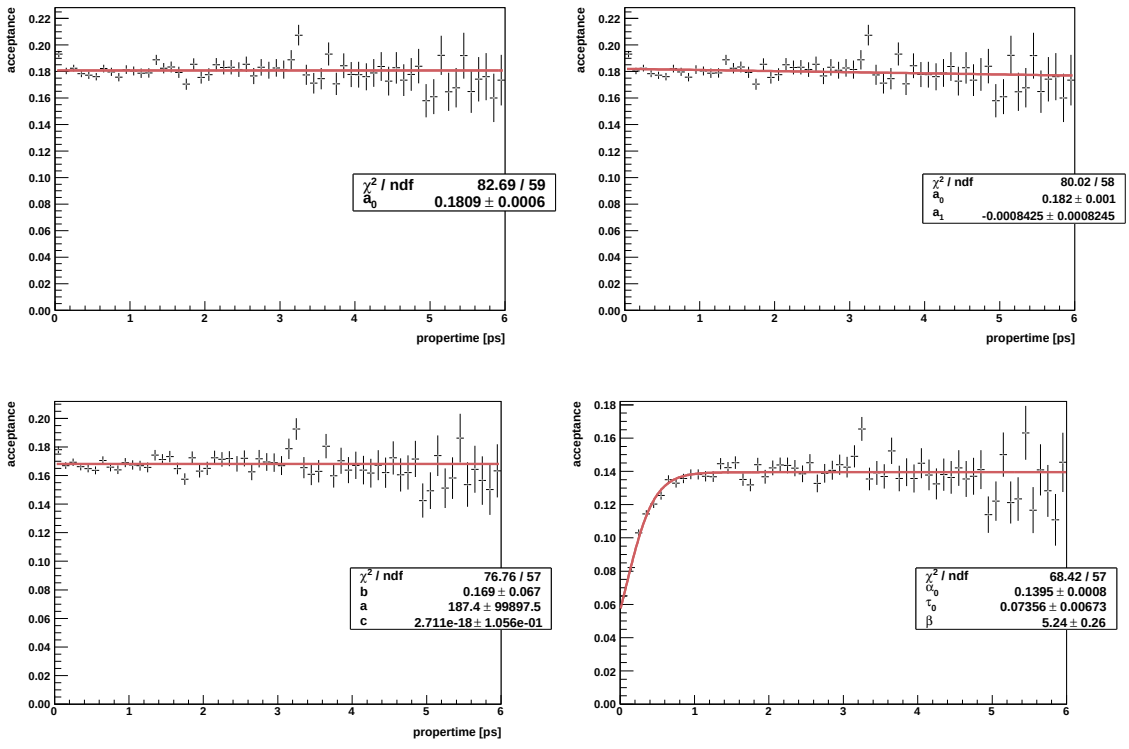


Figure 6.3: Shown in the top left- and right-hand plots is the acceptance of lifetime unbiased selected events as a function of propertime before L0 and HLT triggers. The middle left-hand plot shows the acceptance effect due to the global L0 decision, while the middle right-hand plot shows the effect of L0 and HLT trigger decision using the parametrisation found using equation 6.3.

6.2.2 Decay angle studies

Resolution and acceptance effects for the $B_s \to J/\psi\phi$ decay angles using fully simulated signal data have also been investigated. Figure 6.4 shows the true distribution for each of the decay angles: θ_{tr}, ϕ_{tr} and θ_ϕ, within the 400 mrad geometrical detector acceptances[4]. The effect of then applying the off-line selection cuts on each decay angle, is shown in Figure 6.6. The effect of applying the L0, and L0 with the HLT triggers are described and fitted for in Figures 6.7 and 6.8 respectively. In this figure we observe noticeable structure for each of the angular acceptance distributions, where each angular accetance distrubition has been fitted using an n^{th} degree polynomial[5]. The effect due to the 400 mrad generator-level cut has already been investigated [75] to have negligible effect on the distribution of the decay angles. This angular acceptance structure is further investigated in chapter 6, where a simple generator-level study revealed that the $p_\perp$ cut applied to the muons and kaons is a major contributor. The angular acceptance effects caused by cutting on the $p_\perp$ is important to understand. In section 6.3.3, we find that this leads to a bias in fit parameters R_t and R_p, and will thus ultimately bias any measurement of $-2\beta_s$. In a real analysis using this channel, the effects due to angular acceptance will need to be accounted for.

The residuals for each of the decay angles have also been investigated and are presented in Figure 6.5. A double Gaussian (with the core given in red) has been used to fit for each residual, such that:

$$\begin{aligned}
\theta_{tr} : \sigma^{res}_{\theta_{tr}} &= (29.6 \pm 0.324) \text{ mrad} \quad (\text{core } 0.71\%, \sigma = 11.86 \text{ mrad}), \\
\phi_{tr} : \sigma^{res}_{\phi_{tr}} &= (28.8 \pm 0.268) \text{ mrad} \quad (\text{core } 0.74\%, \sigma = 8.31 \text{ mrad}), \\
\theta_{\phi} : \sigma^{res}_{\theta_{\phi}} &= (20.1 \pm 0.332) \text{ mrad} \quad (\text{core } 0.87\%, \sigma = 14.77 \text{ mrad}).
\end{aligned}$$

Since these resolutions are negligible small, they have little effect on the distribution of the decay angles (Figure 6.4) and thus on our sensitivity studies to $-2\beta_s$; this effect will therefore be neglected in our sensitivity studies.

6.2.3 Tagging studies

Identifying the flavour of the reconstructed B-meson at time of production is a crucial task for LHCb. Two different tagging strategies are used [84]:

[4] These distribution are obtained at the generator level.

[5] The order of the polynomial being chosen to ensure the χ^2 per degree of freedom is close to 1.

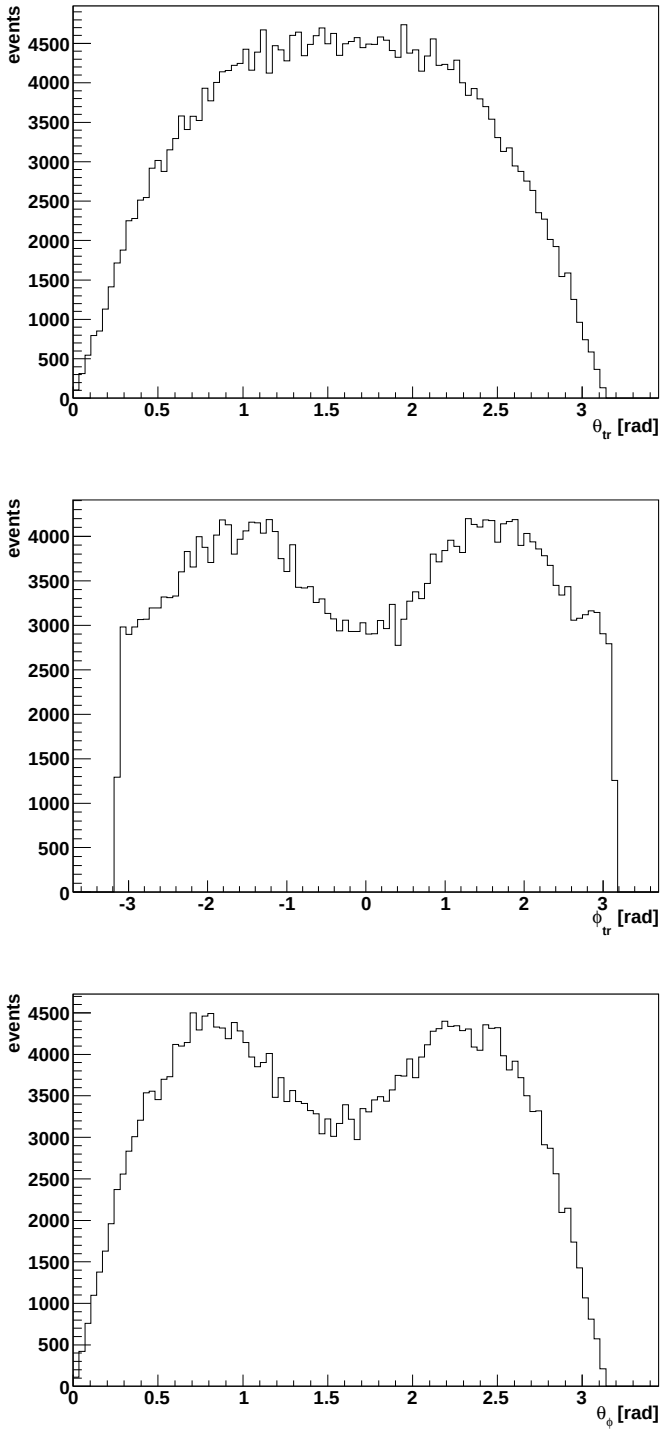


Figure 6.4: The tranversity angular distribution, θ_{tr}, ϕ_{tr} and θ_ϕ for $B_s \rightarrow J/\psi\phi$ events at the generator level (EVTGEN), where the polarisation fractions are $\mathrm{R}_\perp$ = 0.24 and R_0 = 60.

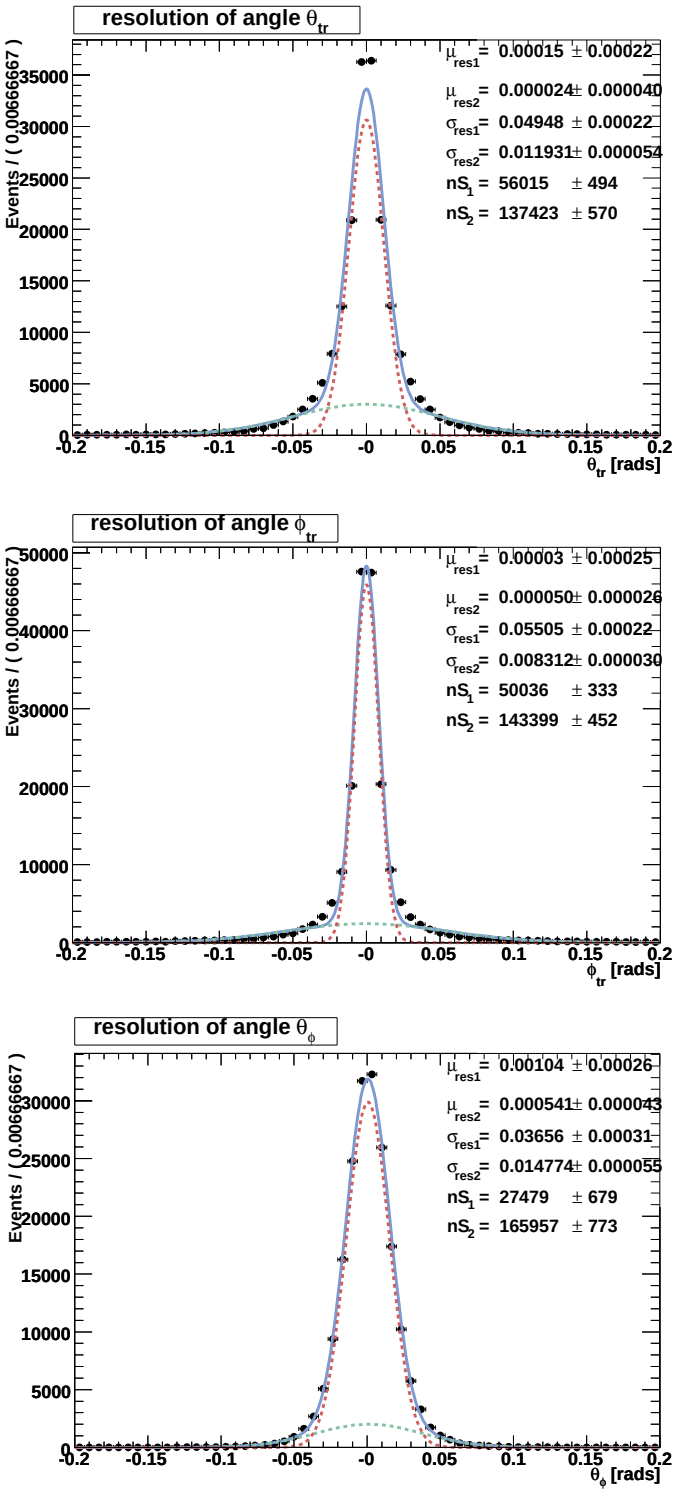


Figure 6.5: Resolution distributions for the decay angle with fully simulated data after applying the selection cuts. The left plots describes the θ_{tr} resolution distribution, middle plot the ϕ_{tr} distribution and the θ_ϕ resolution profile is shown on the right.

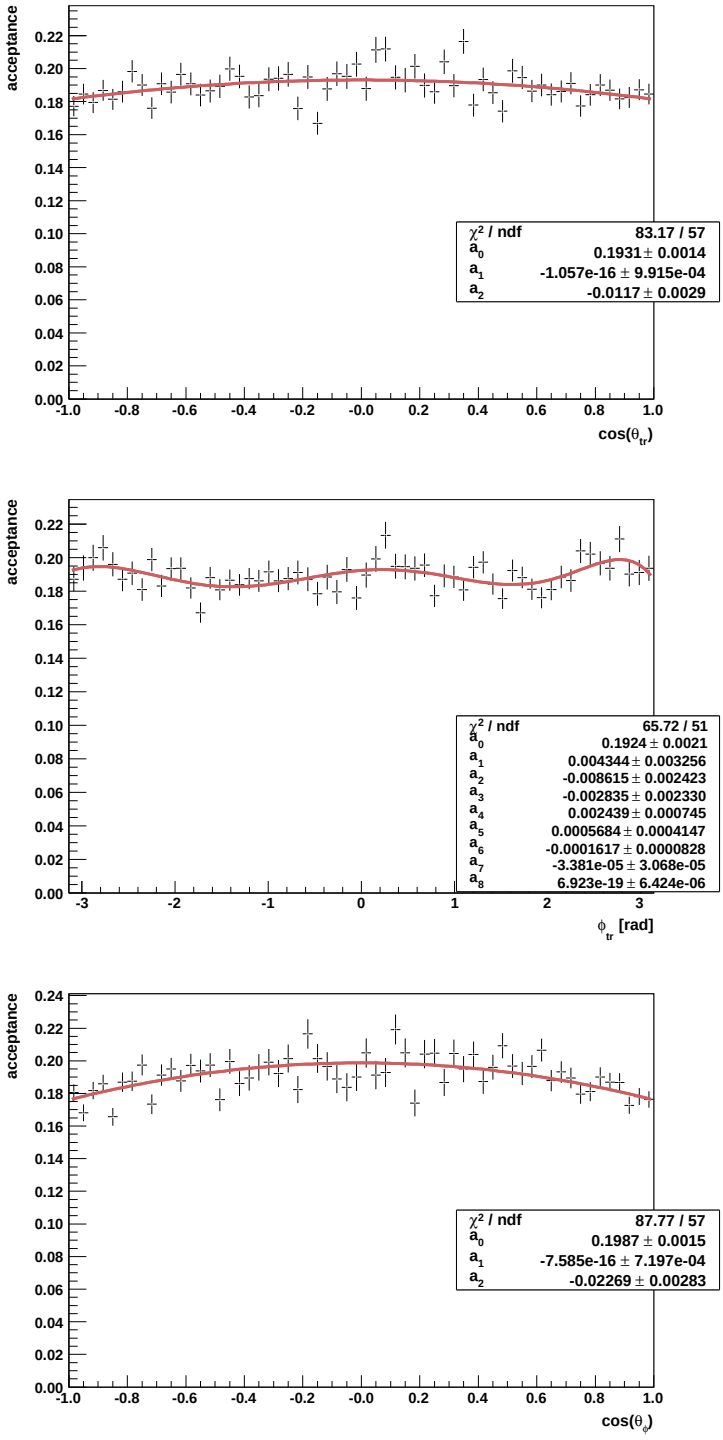


Figure 6.6: The acceptance effect for each decay angle in $B_s \rightarrow J/\psi\phi$ events: (top) $\cos(\theta_{tr})$, (middle) ϕ_{tr}, (bottom) $\cos(\theta_\phi)$, before trigger.

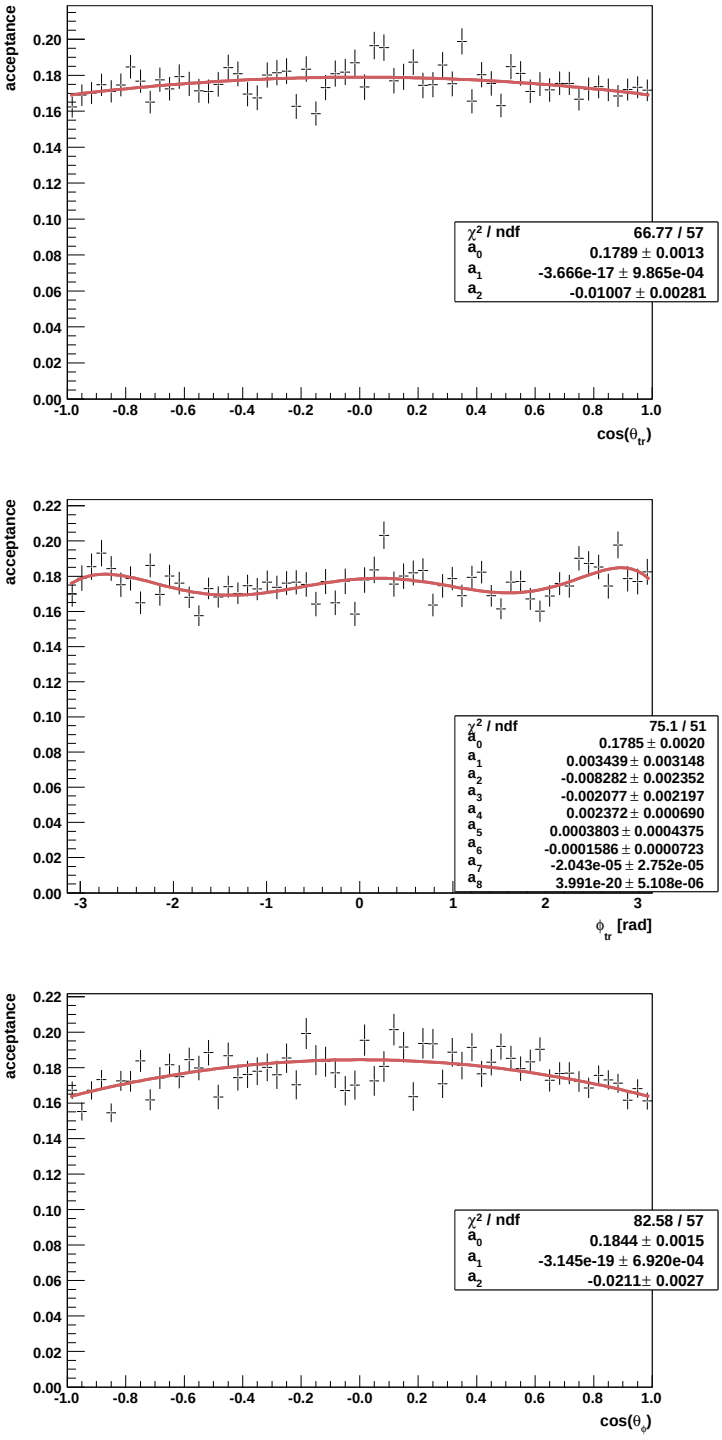


Figure 6.7: The acceptance effect for each decay angle in $B_s \to J/\psi\phi$ events: (top) $\cos(\theta_{tr})$, (middle) ϕ_{tr}, (bottom) $\cos(\theta_\phi)$, after the L0 decision.

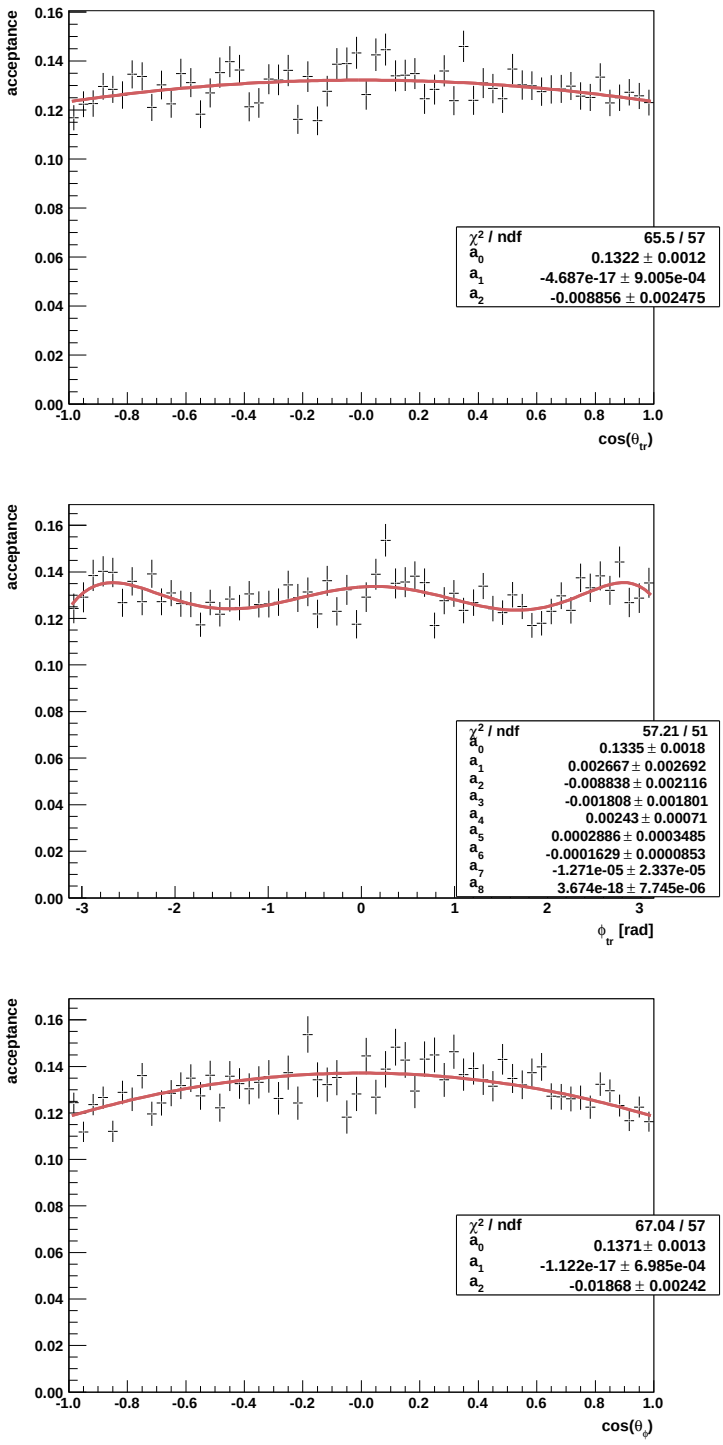


Figure 6.8: The acceptance effect for each decay angle in $B_s \rightarrow J/\psi\phi$ events: (top) $\cos(\theta_{tr})$, (middle) ϕ_{tr}, (bottom) $\cos(\theta_\phi)$, after both the L0 and HLT decision.

6.2.4 Tagging studies

Identifying the flavour of the reconstructed B-meson at time of production is a crucial task for LHCb. Two different tagging strategies are used [84]:

Opposite-side tagging: where the reconstructed B's flavour is inferred from the other B-hadron in the event. It can be determined by detecting the charge of the lepton from semileptonic B decays, or the charge of the kaon from $b \to c \to s$ transitions. The strategy is to search for leptons and kaons with a high transverse momentum and a large impact parameter.

Same-side tagging: where the reconstructed B's flavour is determined directly. This method is applied to tag B_s mesons. When a $B_s^0(\bar{b}s)$ meson is produced in a pp collision, it is accompanied by an extra $\bar{s}$ quark. Similarly, a s quark will accompany the production of a $\bar{B}_s^0$ meson. As the s or $\bar{s}$ quarks hadronise into a charged kaon $\sim 50\%$ of the time, we can use the charge of the kaon to infer back to the flavour of the reconstructed B.

By wrongly tagging a B_s^0 meson decay, a systematic effect is incurred diluting our sensitivity to B_s^0 mixing parameters. The performance of the tagging procedure is characterised by three quantities: the tagging efficiency $\epsilon_{\rm tag}$, which gives the fraction of events in which the tagging procedure gives an answer; the mistag fraction $\omega_{\rm tag}$, which gives the probability for the answer (tagging decision) to be incorrect when a tag is present and the effective tagging efficiency ϵ_{eff}, which is directly related to the tagging power and thus the statistical uncertainly on $\mathcal{CP}$ asymmetry measurements. The probabilities $\omega_{\rm tag}$, $\epsilon_{\rm tag}$ and ϵ_{eff} are calculated as:

$$\omega_{\rm tag} = \frac{N_{\rm wrong}}{N_{\rm wrong} + N_{\rm right} + N_{\rm no}}, \quad \epsilon_{\rm tag} = \frac{N_{\rm right} + N_{\rm wrong}}{N_{\rm right} + N_{\rm wrong} + N_{\rm no}}, \quad \epsilon_{eff} = \epsilon_{\rm tag}(1 - 2\omega_{\rm tag})^2. \tag{6.4}$$

Where $N_{\rm right}$, $N_{\rm wrong}$ and $N_{\rm no}$ are the number of correctly tagged, incorrectly tagged and untagged events, respectively.

The beauty of the $B_s \to J/\psi\phi$ data is that it is self tagging, that is we can extract $\omega_{\rm tag}$ from the data themselves as we discuss in section 6.4.2. However, the canonical approach adopted by LHCb is to evaluate these quantities from self-tagging channels called control channels; the control channel used for $B_s \to J/\psi\phi$ events will be $B_s^0 \to D_s^-\pi^+$ due to its high signal yield, $\sim$ 80,000 events in 2 $\rm fb^{-1}$ and decay topology, which is similar to our signal topology [76].

The purpose of this section is quantify the tagging performance[6], and briefly discuss any difference between the $B_s \to J/\psi\phi$ signal and the $B_s \to D_s\pi$ control channel. For the studies carried out in this section use is made of the $B_s \to J/\psi\phi$ selection cuts as described in section

[6]Where the additional LHCb analysis software has been used, including; FlavourTagging (version v7r3) and TriggerTisTosTool (version v2r1).

4.5, whilst for the $B_s \rightarrow D_s\pi$ selection, use is made of the selection cuts studied in [76].

The tagging performance for each data sample (signal and control channel) under investigation, is evaluated in terms of the mistag fraction $\omega_{\rm tag}$, the tagging efficiency $\epsilon_{\rm tag}$ and the effective tagging efficiency ϵ_{eff} using MC truth information, as shown in Table 6.1. It is clear from this table that there is difference between the tagging performance, especially the $\omega_{\rm tag}$, obtained from the control channel and our signal channel.

Channel	$B_s \rightarrow J/\psi\phi$	$B_s \rightarrow D_s\pi$
reconstructed	142000	73711
selected (w. MC truth)	132060	61 931
$\omega_{\rm tag}$ (%)	37.43 ± 0.15	35.41 ± 0.24
$\epsilon_{\rm tag}$ (%)	62.1 ± 0.12	66.23 ± 0.19
ϵ_{eff} (%)	3.924 ± 0.095	5.637 ± 0.18

Table 6.1: Tagging performance based on truth MC association.

In previous studies [85, 86], it has been shown that it is not straight forward to apply the mistag fraction from one decay channel to the other. Both decays have distinct signatures (one has a pair of muons while the other is purely hadronic) and therefore the events are triggered differently. This difference affects systematically the tagging performance in signal and control channels.

6.3 General procedures

6.3.1 Toy Monte Carlo

A Toy Monte Carlo model written in C++, and using a standard acceptance/rejection method, was developed to simulate four observables: the lifetime of the decay and the three decay product angles.

6.3.1.1 Experimental parameters

The expected annual yield for this decay, taken from studies carried out in section 4.7, is 133k events for nominal luminosity of 2 fb^{-1}. We take into consideration the signal tagging efficiency ε_{tag} and the mistag fraction ω_{tag} from section 6.2.4, with values 37.4% and 62.1% respectively, and choose (see section 6.2.1) an average proper time resolution of $\sigma_\tau = 37 \pm 0.5$ fs. From chapter 5, the background contribution for the $b\bar{b}$-inclusive and prompt-J/ψ samples are taken to be 1.17 and 1.9 respectively. The tagging efficiencies for the $b\bar{b}$-inclusive and the prompt J/ψ samples are taken from [41]. These parameters are summarised in Table 6.2.

yield [k]	$\left(\frac{B}{S}\right)^{bb}$	$\left(\frac{B}{S}\right)^{J/\psi}$	σ_τ [fs]	ω_{tag} [%]	ε_{tag} [%]	$\epsilon_{tag}^{b\bar{b}}$ [%]	$\epsilon_{tag}^{J/\psi}$ [%]
133	1.17	1.9	37 $\pm$0.5	37.4 $\pm$1.0	62.1 $\pm$1.0	30.0 $\pm$1.0	60.0 $\pm$1.0

Table 6.2: Expected experimental parameters for LHCb obtained in this thesis.

6.3.1.2 The B_s ($\bar{B}_s$) model

The ingredients to make the B_s ($\bar{B}_s$) p.d.f. are: the differential cross sections for both decay channels described in section 2.6; the tagging efficiency ε_{tag} and the mistag-fraction ω_{tag}. The tagging procedure is modelled by introducing a random variable q, which can take three possible values: +1 if the event is tagged at production time as a B_s; -1 if the event is tagged at production time as a $\bar{B}_s$; 0 if the event has no tag, which becomes equivalent of having $\omega_{tag} = 0.5$. We do not consider B_s-$\bar{B}_s$ production asymmetry effects in this study.

Defining the functions $W^+(t,\Omega)$ and $W^-(t,\Omega)$ to be the differential decay rates for $B_s \to J/\psi\phi$ and $\bar{B}_s \to J/\psi\phi$ given by equation (6.5):

$$W^+(t,\Omega) = \frac{d^3\Gamma(t)}{d\cos\theta_{tr}\, d\cos\theta_\phi\, d\phi_{tr}}, \quad W^-(t,\Omega) = \frac{d^3\bar{\Gamma}(t)}{d\cos\theta_{tr}\, d\cos\theta_\phi\, d\phi_{tr}} \tag{6.5}$$

where the differential decay rates have been defined in section 2.6. Here Ω stands for the set $[\cos\theta_{tr}, \cos\theta_\phi, \phi_{tr}]$. The underlying p.d.f. used for all of our event generation and fitting $\mathcal{F}(t,\Omega)$ is then:

$$\mathcal{F}(t,\Omega) = \frac{\epsilon_1 W^+(t,\Omega) + \epsilon_2 W^-(t,\Omega)}{\int \left(\epsilon_1 W^+(t,\Omega) + \epsilon_2 W^-(t,\Omega)\right) dt d\Omega} \tag{6.6}$$

The factors ϵ_1 and ϵ_2 take three different values according to the given q parameter value. Using equation (6.5), the generated distributions for the B_s propertime and decay angles are shown in Figures 6.9 and 6.10 respectively.

	q = +1	q = -1	q = 0
ϵ_1	$(1-\omega_{tag})$	ω_{tag}	0.5
ϵ_2	ω_{tag}	$(1-\omega_{tag})$	0.5

Table 6.3: Definition of tagging parameters

6.3.2 Including resolution

The proper time resolution was approximated by a single fixed Gaussian of width σ_τ for all events (given in Table 6.2). We convolve $\mathcal{F}(t,\Omega)$ with this Gaussian function:

$$\mathcal{F}'(t',\Omega) = \mathcal{F}(t,\Omega) \otimes \mathcal{G}(t-t';\sigma_\tau) \tag{6.7}$$

where the $\prime$ denotes "with resolution incorporated". The effect of this resolution on the propertime is shown in Figure 6.11.

Resolution on the decay angles was studied in chapter 4, where its effect was concluded to be negligible. Previous work [81] has also shown that angular resolution makes little impact on the results. We did not therefore included any angular resolution effects, but this will be studied in more detail by the collaboration as part of future work.

6.3.3 Including acceptance

For the purposes of this study, we have made use of the $B_s \rightarrow J/\psi\phi$ event selection given in Table 4.5, before applying the LHCb trigger[7]. The result of these cuts and of not applying the trigger is a flat acceptance distribution for the B_s propertime as highlighted in Figure 6.3. For the studies presented in section 6.5, we do not include an acceptance effect for the B_s propertime.[8] In this study, we also neglect acceptance effects on the decay angles.

However, from studies carried out on fully simulated data in section 6.2.2, we see an non negligible acceptance effect for the decay angles. In section 6.7 we examine the angular acceptance effect more closely. This is achieved by investigating what effect the $p_\perp$ cut, applied to both muons and/or kaons (see Table 4.5), has on the decay angular distributions and ultimately our physics parameters of interest. This acceptance effect should be addressed in detail in any future extensions to the studies presented in this thesis.

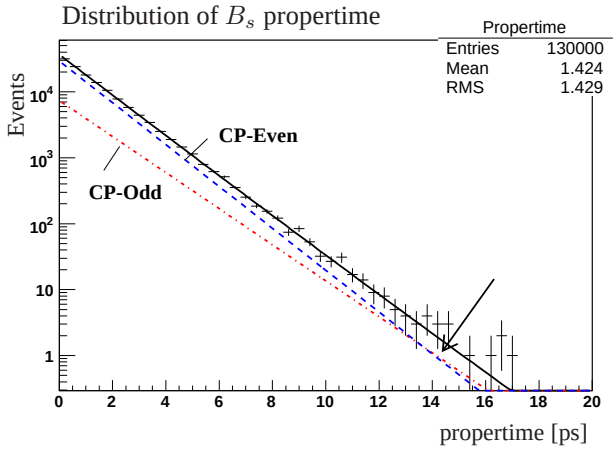


Figure 6.9: The Toy MC propertime distribution for $B_s \rightarrow J/\psi\phi$ events. The plot illustrates, for MC generated data (show in black), the $\mathcal{CP}$-even contribution (shown in blue) and the $\mathcal{CP}$-odd contribution (shown in red). The plot also illustrates (denoted by the arrow) the propertime at which the $\mathcal{CP}$-odd contribution begins to dominate over the $\mathcal{CP}$-even contribution.

[7] Post-trigger experimental parameter numbers have not been used, due to the restructuring of the trigger system, as discussed in section 4.4.

[8] The effect of modelling the propertime acceptance with a simple step function, can be found in [87].

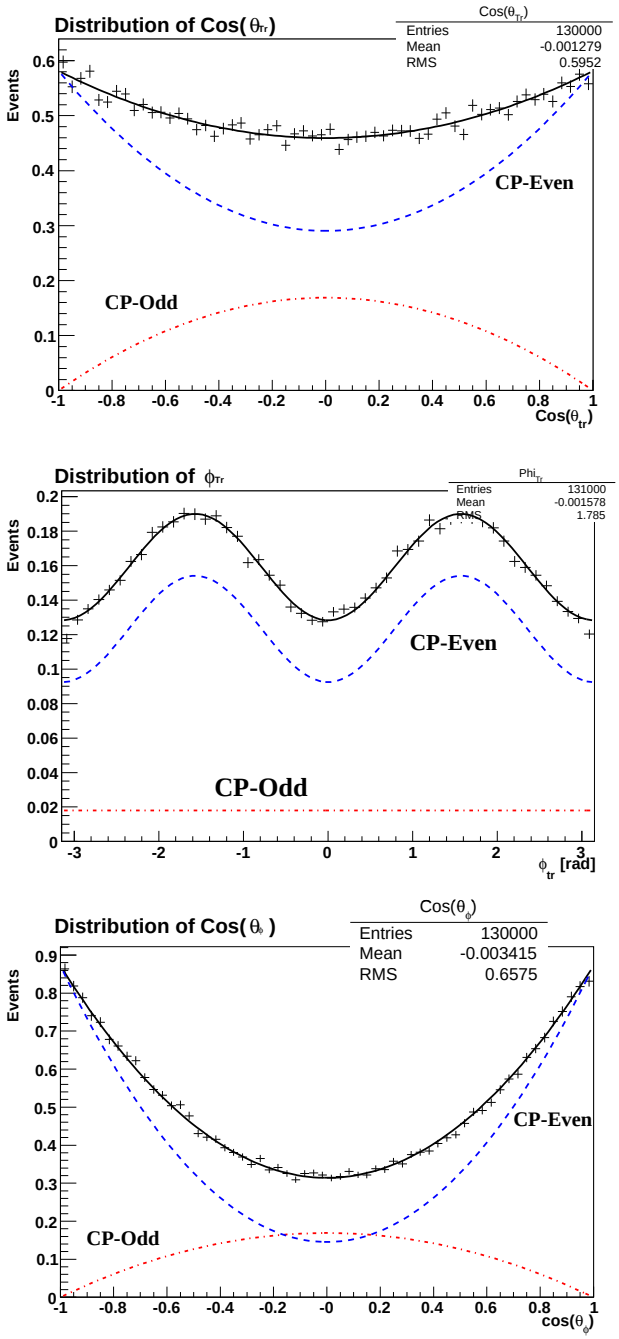


Figure 6.10: The Toy MC generated decay angular distributions, $\cos(\theta_{tr})$, ϕ_{tr} and $\cos(\theta_{\phi})$ for $B_s \rightarrow J/\psi\phi$ events. The data presented in this plot is only for illustration purposes, showing the $\mathcal{CP}$-even and $\mathcal{CP}$-odd component in each of the decay angles. The total distribution is represented by the solid-black line while the $\mathcal{CP}$-even and $\mathcal{CP}$-odd contributions have are indicated by the dashed blue and red lines respectively.

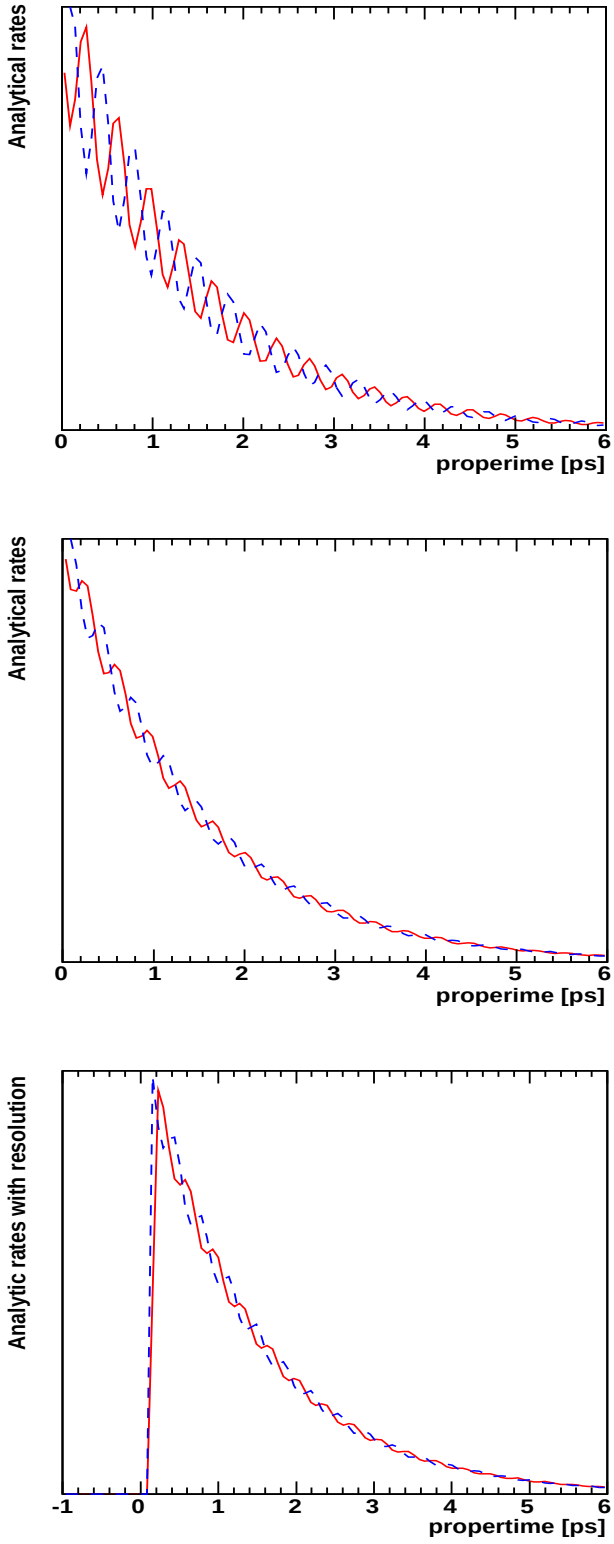


Figure 6.11: Signal decay rates [ps] for the $B_s \to J/\psi\phi$ transition to pure $\mathcal{CP}$-even eigenstates for initially B_s (red) and $\bar{B}_s$ (dashed blue) tagged mesons. The amplitude has been magnified by a factor of 10. The top plot shows the analytical decay rates, the effect of wrong-tag is shown in the middle plot, while the bottom plot show the effect of including constant propertime resolution.

6.3.4 Including background

The background is modelled as a simple exponential in proper time for the $b\bar{b}$ and a delta function for the prompt-J/ψ component respectively. For simplicity in the likelihood function, we have assumed that the angular distributions of the backgrounds are flat. However, we see from Figures 6.12 and 6.13 that the angular distributions of the backgrounds have a definite non flat structure. In any extensions to these studies, these structures will need to be taken into account. The background p.d.f. $\mathcal{B}$ is written as:

$$\mathcal{B}(t,\Omega) = f_{\rm exp}\exp^{-t/\tau_0}/\tau_0 + f_{\rm prompt}\delta(t) \tag{6.8}$$

This was subject to the same procedures for inclusion of tagging and resolution as described above to yield $\mathcal{B}''(t',\Omega)$ in a similar way as for $\mathcal{F}''$. In equation 6.8, the value of τ_0 is taken as 1 ps, which is based on DC04 full reconstruction studies [75]. $f_{\rm exp(prompt)}$ are the fraction of $b\bar{b}$-inclusive and J/ψ-prompt events respectively, such that:

$$f_{\rm exp(prompt)} = \frac{N \times \left(\frac{B}{S}\right)^{b\bar{b}(J/\psi)}}{N + N \times \left(\left(\frac{B}{S}\right)^{b\bar{b}} + \left(\frac{B}{S}\right)^{J/\psi}\right)}, \tag{6.9}$$

where N is the total number of signal events. The p.d.f. including both signal and background is then given by

$$\mathcal{P}''(t',\Omega) = f_{\rm sig}\mathcal{F}''(t',\Omega) + (1 - f_{\rm sig})\mathcal{B}''(t',\Omega) \tag{6.10}$$

where $f_{\rm sig} = 1 - f_{\rm exp} - f_{\rm prompt}$ is the signal fraction expected in the sample.

6.3.5 Physics input parameter central values

The central values of measurements of $\bar{\Gamma}_s$ [88] and Δm_s [89] were taken. $\Delta\Gamma_s$, $R_\perp$, R_0 and $-2\beta_s$ were assumed to have their predicted value [23, 90], while the values for δ_1 and δ_2 are motivated from [19, 91]. Table 6.4 lists the 8 nominal values needed for the three-angle studies:

$\bar{\Gamma}_s$ ps^{-1}	$\Delta\Gamma_s$ ps^{-1}	$R_\perp$	Δm_s ps^{-1}	$-2\beta_s$ [rad]	R_0	δ_1 [rad]	δ_2 [rad]
0.68	0.10	0.20	17.77	-0.04	0.60	-0.46	2.97

Table 6.4: Assigned values to the physical parameters

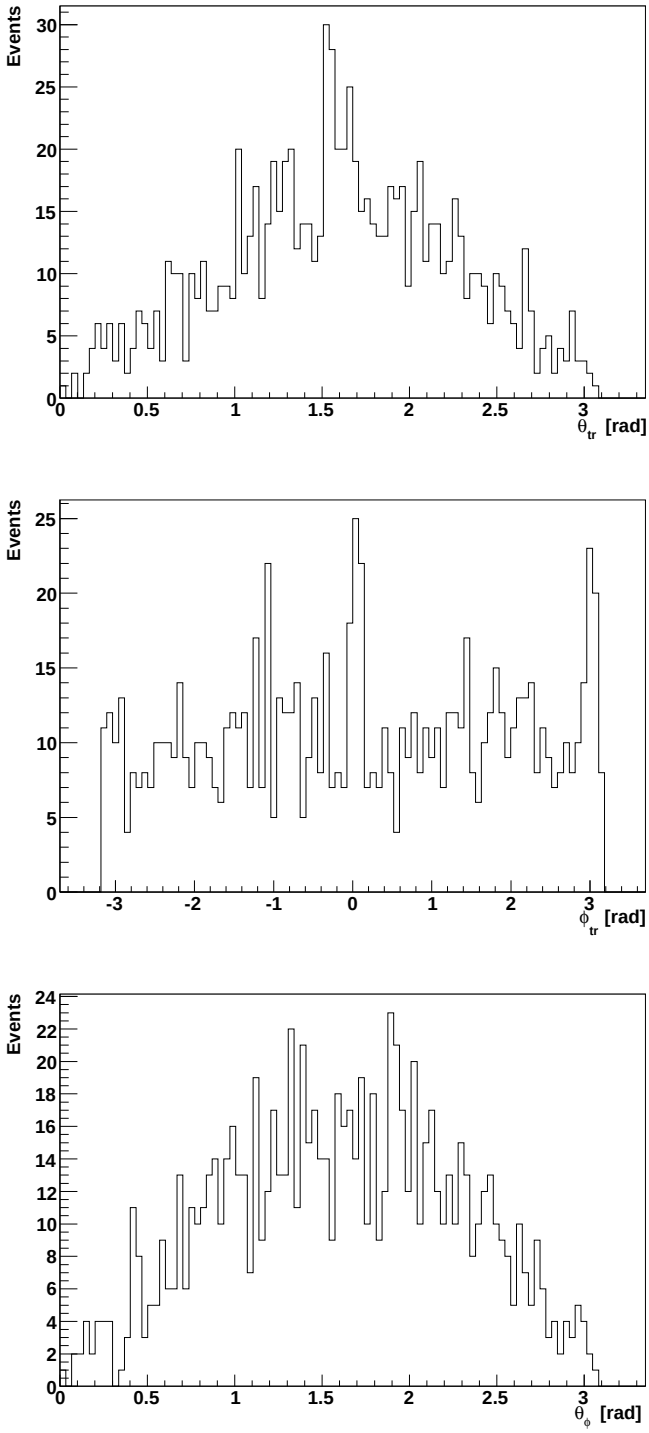


Figure 6.12: The transversity angular distribution, θ_{tr}, ϕ_{tr} and θ_{ϕ} for J/ψ-inclusive events after applying the selection cuts in Table 4.5.

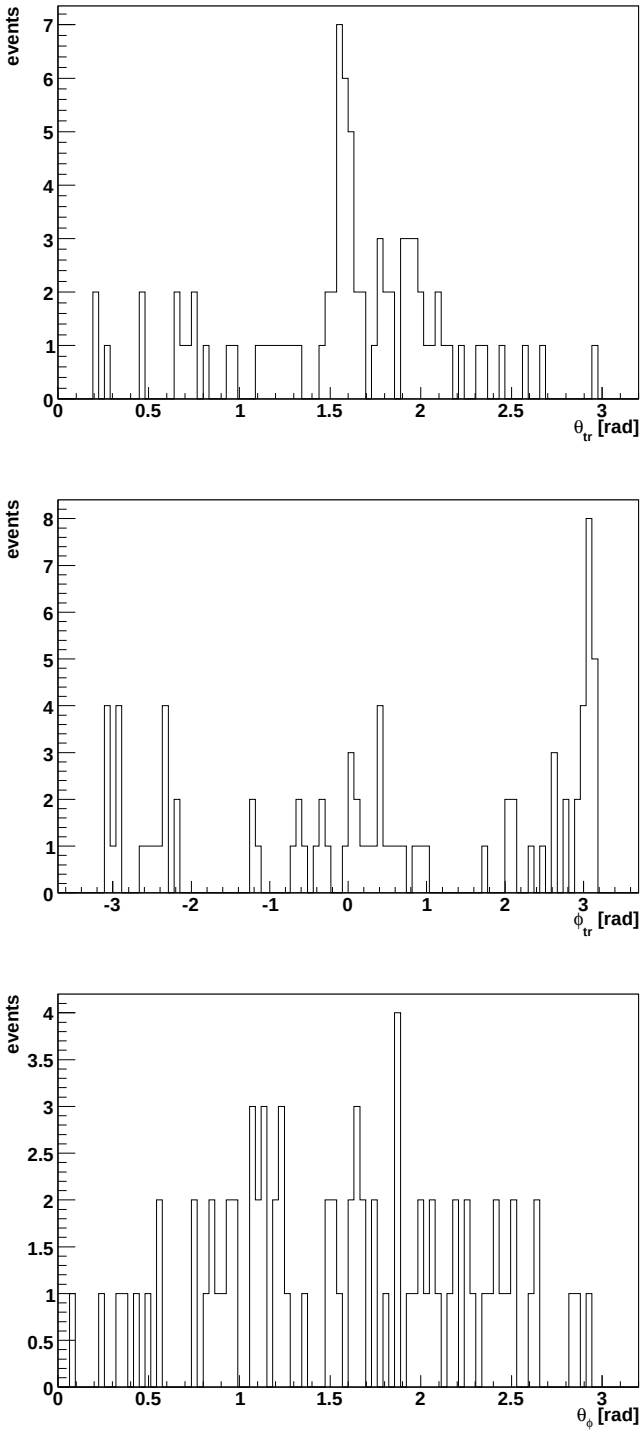


Figure 6.13: The transversity angular distribution, θ_{tr}, ϕ_{tr} and θ_{ϕ} for $b\bar{b}$-inclusive events after applying the selection cuts in Table 4.5.

6.3.6 Unbinned maximum likelihood fit

An unbinned maximum likelihood fit to the Toy MC data was preformed. The likelihood function being constructed as:

$$\mathcal{L} = \prod_{i}^{N} P(t_i, \Omega_i) \tag{6.11}$$

where $P(t, \Omega)$ is either (i) $\mathcal{F}(t, \Omega)$ (ii) $\mathcal{F}'(t', \Omega)$ or (iii) $\mathcal{P}''(t', \Omega)$ according to whether we are presenting (i) ideal (ii) with resolution or (iii) with resolution and background respectively.

In all cases we ran the toy MC for at least 500 repeat "experiments", and in each case fitted for the physics parameters using MINUIT[9][92].

For each physics quantity being studied we plot the returned best fit value, the returned error, and the pull distribution. These are all included in Appendix D. We checked that the pull distribution was not significantly biased, and that its width was ≈ 1. This was true in all cases to within 10%. We also checked the width of the returned fit value distribution was consistent with the average returned error. Again no inconsistencies were found. The precisions quoted in the tables are then taken as the standard deviation of the returned fit value distribution. For a complete description of the results obtained, see Appendix D.1.

6.4 Comparing the one and three angular $B_s \to J/\psi\phi$ analysis

In this section, we compare the parameter sensitivities obtained using the one and the three angular analysis. In this section we also investigate the effects of propertime resolution and long lived background. The studies in this section pre-date those studies given in section 6.5, using experimental and input parameter values obtained from earlier DC04 studies [75, 67].

For studies in this section, the expected annual yield is taken as 131k events for the nominal luminosity of 2 fb^{-1}[75]. The tagging efficiency and the mistag fraction are taken to be 33% and 57% respectively [76], and an average proper time resolution of $\sigma_\tau = 30 \pm 0.5$ fs [75] is considered. Only $b\bar{b}$-inclusive background is considered, with a $\frac{B}{S}$ of 0.12, and is modeled by an exponential in propertime (see equation 6.8). The input parameters retain their central values as given in Table 6.2, with the strong phases, δ_1 and δ_2, set to 0 and π respectively.

The studies in this section also include external constraints on Δm_s and $\omega_{\rm tag}$. The data in the $B_s \to J/\psi\phi$ channel alone constrains Δm_s to a precision of ≈ 0.2 in the single-angle fit and ≈ 0.02 in the three-angle fit, as discussed in section 6.8.1. However, Δm_s will also be constrained by data external to this channel (e.g. $B_s \to D_s\pi$) to a precision of ≈ 0.007 [75]. We therefore included in our fits a Gaussian constraint on Δm_s to its central value ± 0.007.

[9] We use MINUIT strategy=2 and we do an explicit call to HESSE after the minimisation execution.

The one-angle fit to determine $-2\beta_s$ cannot be performed without an external constraint on $\omega_{\rm tag}$ as it multiplies $\sin(2\beta_s)$ (and is hence 100% correlated). In the three-angle fit, $\omega_{\rm tag}$ is determined independently of $-2\beta_s$ to a precision of ≈ 0.01 and is almost uncorrelated to all other parameters. This arises due to the richness of different terms appearing in the full three-angle differential cross section. In our fits we also include a Gaussian constraint on $\omega_{\rm tag}$ to its central value of ± 0.0036 [75]. We have studied the effect of the $\omega_{\rm tag}$ constraint in section 6.8.1. The salient features of these studies is discussed below. A full description can be obtained in [87].

6.4.1 Sensitivity studies using the one decay angle analysis

In this section presents results where only the single transversity angle θ_{tr} has been used in the analysis. This will be referred to as the *one-angle fit*. We use the full 131k event sample, with each event being classified as either a b, $\bar{b}$ or untagged. The one-angle differential cross section for tagged events is (if tagging for a B_s):

$$\begin{aligned}(1-\omega_{\rm tag})\frac{d\Gamma(t)}{d\cos\theta_{tr}} \quad &+ \quad \omega_{\rm tag}\frac{d\bar{\Gamma}(t)}{d\cos\theta_{tr}} \propto (1-R_\perp)[(1+\cos(2\beta_s))e^{-\Gamma_L t}+(1-\cos(2\beta_s))e^{-\Gamma_H t} \\ &- \quad 2(\mathbf{1}-\mathbf{2}\omega_{\rm tag})e^{-\bar{\Gamma_s}t}\sin(\Delta m_s t)\sin(2\beta_s)]\frac{1}{2}(1+\cos^2\theta_{tr}) \\ &+ \quad R_\perp[(1-\cos(2\beta_s))e^{-\Gamma_L t}+(1+\cos(2\beta_s))e^{-\Gamma_H t} \\ &+ \quad 2(\mathbf{1}-\mathbf{2}\omega_{\rm tag})e^{-\bar{\Gamma_s}t}\sin(\Delta m_s t)\sin(2\beta_s)]\sin^2\theta_{tr}\end{aligned}$$

The different lifetime components are statistically separated through the angular distributions and hence $\Delta\Gamma_s$, $\bar{\Gamma}_s$ and $R_\perp$ may be measured. This is illustrated in the top plot in Figure 6.10, where we see the two different shapes of the θ_{tr} distribution for $\mathcal{CP}$-even and $\mathcal{CP}$-odd contributions.

In Table 6.5 the baseline results are presented where $\Delta\Gamma_s$, $\bar{\Gamma}_s$, $R_\perp$, Δm_s and $-2\beta_s$ are determined simultaneously, with the external Δm_s and $\omega_{\rm tag}$ constraints included. The correlation matrix is shown in the following table (Table 6.6) . The first three parameters remain uncorrelated with Δm_s or $-2\beta_s$ and hence this sector of the fit factorises and the results are identical to the untagged case. The error on Δm_s is completely determined by the external constraint, and is uncorrelated with anything else. The error on $-2\beta_s$ is uncorrelated with anything else and hence is the same as per the single parameter fits. Fitted values, errors and pulls distributions are shown in Appendix D.1.

The precision on $-2\beta_s$ degrades from 0.023 to 0.026 as resolution is incorporated. This may be expected as the proper time resolution (30 fs) is now 10% of the period of the sinusoidal oscillation (≈ 300 fs). Background was found to have only a small additional effect.

Parameter errors	Ideal	Resolution	Background
$\bar{\Gamma}_s$	0.0036	0.0035	0.0034
$\Delta\Gamma_s$	0.013	0.013	0.014
$R_\perp$	0.0052	0.0052	0.0059
Δm_s	0.007*	0.007*	0.007*
$-2\beta_s$	0.023	0.026	0.027
$\omega_{\rm tag}$	0.0036*	0.0036*	0.0036*

Table 6.5: Baseline results: Simultaneous one-angle fit to $\Delta\Gamma_s$, $\bar{\Gamma}_s$, $R_\perp$, Δm_s and $-2\beta_s$ using tagged events and the external Δm_s and $\omega_{\rm tag}$ constraints (indicated by *)

	$\bar{\Gamma}_s$	$\Delta\Gamma$	$R_\perp$	Δm_s	$-2\beta_s$	$\omega_{\rm tag}$
$\bar{\Gamma}_s$	1.0	-0.82	0.65	-0.00	-0.04	0.00
$\Delta\Gamma_s$		1.0	-0.68	0.00	0.03	0.00
$R_\perp$			1.0	-0.00	-0.04	-0.00
Δm_s				1.0	0.01	0.00
$-2\beta_s$					1.0	-0.02
$\omega_{\rm tag}$						1.0

Table 6.6: Baseline results: Correlation coefficients for simultaneous one-angle fit to $\Delta\Gamma_s$, $\bar{\Gamma}_s$, $R_\perp$, Δm_s and $-2\beta_s$ using tagged events and the external Δm_s and $\omega_{\rm tag}$ constraints

6.4.2 Sensitivity studies using the three decay angle analysis

In this section we describe the results obtained using the full three decay angle analysis. Unlike in the *one-angle fit*, we can fit for $-2\beta_s$ in the three-angular fit using un-tagged data. This section then begins with a discussion of the features, and parameter sensitivity to the un-tagged three angular analysis, before discussing the three angular tagged analysis.

Un-tagged events

From sections 2.6 and A.5, the three-angle differential cross section components for untagged B_s and $\bar{B_s}$ decays (i.e a 50% mix of each) are:

$$|A_0(t)|^2 + |\bar{A}_0(t)|^2 = \frac{|A_0(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t}\Big] \quad (6.12)$$

$$|A_\parallel(t)|^2 + |\bar{A}_\parallel(t)|^2 = \frac{|A_\parallel(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t}\Big] \quad (6.13)$$

$$|A_\perp(t)|^2 + |\bar{A}_\perp(t)|^2 = \frac{|A_\perp(0)|^2}{2}\Big[(1-\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t}\Big] \quad (6.14)$$

$$\mathrm{Re}\{A_0^*(t)A_\parallel(t)\} + \mathrm{Re}\{\bar{A}_0^*(t)\bar{A}_\parallel(t)\} = \frac{1}{2}|A_0(0)||A_\parallel(0)|\cos(\delta_2-\delta_1)\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t}\Big] \tag{6.15}$$

$$\mathrm{Im}\{A_\parallel^*(t)A_\perp(t)\} + \mathrm{Im}\{\bar{A}_\parallel^*(t)\bar{A}_\perp(t)\} = \frac{1}{2}|A_\parallel(0)||A_\perp(0)|\Big[\left(e^{-\Gamma_{\rm H}t}-e^{-\Gamma_{\rm L}t}\right)\cos\delta_1\sin(2\beta_s)\Big] \tag{6.16}$$

$$\mathrm{Im}\{A_0^*(t)A_\perp(t)\} + \mathrm{Im}\{\bar{A}_0^*(t)\bar{A}_\perp(t)\} = \frac{1}{2}|A_0(0)||A_\perp(0)|\Big[\left(e^{-\Gamma_{\rm H}t}-e^{-\Gamma_{\rm L}t}\right)\cos\delta_2\sin(2\beta_s)\Big] \tag{6.17}$$

Several features are observed:

-. Unlike the one-angle case, a $\sin(2\beta_s)$ term remains. However, it would only be possible to consider measuring $-2\beta_s$ with any decent accuracy if the strong phases were known from some external source or could be simultaneously fit.

-. In principle it is possible to fit for everything simultaneously. However for small $-2\beta_s$ (as in the case of the SM expectation) the imaginary cross terms are strongly suppressed (vanishing in the limit $\sin(2\beta_s)=0$). In this case it would be impossible to fit for both δ_1 and δ_2; it would only be possible to fit to the combination $\delta_1-\delta_2$ arising in the real cross term.

-. It is only if $-2\beta_s$ were different from zero (and preferably large) and the δ_i had values different from $0(\mathrm{mod}\pi)$ that it would be possible to reliably fit simultaneously the full set $\bar{\Gamma}_s$, $\Delta\Gamma_s$, $R_\perp$, R_0, δ_1, δ_2 and $-2\beta_s$. Otherwise this fit becomes problematic.

Table 6.7 present the baseline results where all are determined simultaneously. The correlation matrix is shown in Table 6.8. The correlations between the first four parameters are large and this is reflected in the much increased errors compared to single parameter fits.

Parameter errors	Ideal	Resolution	Background
$\bar{\Gamma}_s$	0.0026	0.0027	0.0028
$\Delta\Gamma_s$	0.0074	0.0075	0.0081
$R_\perp$	0.0030	0.0030	0.0035
R_0	0.0024	0.0023	0.0026

Table 6.7: Baseline results: Simultaneous three-angle fit to $\Delta\Gamma_s$, $\bar{\Gamma}_s$, $R_\perp$ and R_0 using untagged events. Values of $-2\beta_s=-0.04$, $\delta_1=0$ and $\delta_2=\pi$ were fixed.

Tagged events

We use the full 131k event sample, with the events classified as either a b, $\bar{b}$ or untagged. All three tagging categories are then used in the fit. The three-angle differential cross section components

	$\bar{\Gamma}_s$	$\Delta\Gamma_s$	$R_\perp$	R_0
$\bar{\Gamma}_s$	1.0	-0.67	0.51	-0.22
$\Delta\Gamma_s$		1.0	-0.65	0.23
$R_\perp$			1.0	-0.65
R_0				1.0

Table 6.8: Baseline results: Correlation coefficients for simultaneous fit to $\Delta\Gamma_s$, $\bar{\Gamma}_s$, $R_\perp$ and R_0 using untagged events

for tagged B_s decays are:

$$(1-\omega_{\rm tag})|A_0(t)|^2 + \omega_{\rm tag}|\bar{A}_0(t)|^2 = \frac{|A_0(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} - 2\,(\mathbf{1}-\mathbf{2}\omega_{\rm tag})\,e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad (6.18)$$

$$(1-\omega_{\rm tag})|A_\parallel(t)|^2 + \omega_{\rm tag}|\bar{A}_\parallel(t)|^2 = \frac{|A_\parallel(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} - 2\,(\mathbf{1}-\mathbf{2}\omega_{\rm tag})\,e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad (6.19)$$

$$(1-\omega_{\rm tag})|A_\perp(t)|^2 + \omega_{\rm tag}|\bar{A}_\perp(t)|^2 = \frac{|A_\perp(0)|^2}{2}\Big[(1-\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} + 2\,(\mathbf{1}-\mathbf{2}\omega_{\rm tag})\,e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad (6.20)$$

$$(1-\omega_{\rm tag})\mathrm{Re}\{A_0^*(t)A_\parallel(t)\} + \omega_{\rm tag}\mathrm{Re}\{\bar{A}_0^*(t)\bar{A}_\parallel(t)\} = \frac{1}{2}|A_0(0)||A_\parallel(0)|\cos(\delta_2-\delta_1)\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} - 2\,(\mathbf{1}-\mathbf{2}\omega_{\rm tag})\,e^{-\Gamma_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad (6.21)$$

$$(1-\omega_{\rm tag})\mathrm{Im}\{A_\parallel^*(t)A_\perp(t)\} + \omega_{\rm tag}\mathrm{Im}\{\bar{A}_\parallel^*(t)\bar{A}_\perp(t)\} = |A_\parallel(0)||A_\perp(0)|\Big[(\mathbf{1}-\mathbf{2}\omega_{\rm tag})\,e^{-\Gamma_s t}\left\{\sin\delta_1\cos(\Delta m_s t) - \cos\delta_1\sin(\Delta m_s t)\cos(2\beta_s)\right\} + \frac{1}{2}\left(e^{-\Gamma_{\rm H}t} - e^{-\Gamma_{\rm L}t}\right)\cos\delta_1\sin(2\beta_s)\Big] \quad (6.22)$$

$$(1-\omega_{\rm tag})\mathrm{Im}\{A_0^*(t)A_\perp(t)\} + \omega_{\rm tag}\mathrm{Im}\{\bar{A}_0^*(t)\bar{A}_\perp(t)\} = |A_0(0)||A_\perp(0)|\Big[(\mathbf{1}-\mathbf{2}\omega_{\rm tag})\,e^{-\Gamma_s t}\left\{\sin\delta_2\cos(\Delta m_s t) - \cos\delta_2\sin(\Delta m_s t)\cos(2\beta_s)\right\} + \frac{1}{2}\left(e^{-\Gamma_{\rm H}t} - e^{-\Gamma_{\rm L}t}\right)\cos\delta_2\sin(2\beta_s)\Big] \quad (6.23)$$

Several features are observed:

-. The sinusoidal Δm_s terms appear multiplied by the $1 - 2\omega_{\rm tag}$ term. Unlike the one-angle case, however, they are not suppressed by $\sin(2\beta_s)$ in the imaginary cross terms. Therefore we expect a substantially improved measurement of both Δm_s and $\omega_{\rm tag}$.

-. The $\sin\delta_1$ and $\sin\delta_2$ terms are present and unlike in the untagged case, they are separated from the $-2\beta_s$ term. They appear in a way which changes the phase of the Δm_s oscillation. We therefore expect a clean fit for δ_1 and δ_2 to be possible with some correlation with Δm_s.

Table 6.9 presents the baseline results where all are determined simultaneously with the external Δm_s and $\omega_{\rm tag}$ constraints included. The correlation matrix is shown in the following Table 6.10. Fitted values, errors and pulls distributions are shown in Appendix D.1. Observations:

-. The first four parameters remain largely uncorrelated with any of the others and hence this sector of the fit factorises and the results are identical to the untagged case.

-. The precisions obtained for $\bar{\Gamma}_s$ and $\Delta\Gamma_s$ are improved over the one-angle case.

-. The error on Δm_s is mainly determined by the external constraint, and is somewhat correlated with δ_1 and δ_2 as expected. These three appear to form a factorised set.

-. The error on $-2\beta_s$ is uncorrelated with anything else and shows a useful improvement over the one-angle fits.

-. In addition we show in Appendix D that we can in fact remove the $\omega_{\rm tag}$ constraint completely and it does not affect any of the results. As expected the data themselves make very precise measurement of $\omega_{\rm tag}$ to ± 0.01 which is uncorrelated with anything else.

parameter errors	Ideal	Resolution	Background
$\bar{\Gamma}_s$	0.0025	0.0025	0.0027
$\Delta\Gamma_s$	0.0075	0.0074	0.0079
$R_\perp$	0.0032	0.0031	0.0035
R_0	0.0025	0.0024	0.0027
δ_1	0.071	0.080	0.083
δ_2	0.064	0.071	0.075
Δm_s	0.007*	0.007*	0.007*
$-2\beta_s$	0.019	0.022	0.022
$\omega_{\rm tag}$	0.0036*	0.0036*	0.0036*

Table 6.9: Baseline results: Simultaneous three-angle fit to all parameters using tagged events. The external constraints are applied.

	$\bar{\Gamma}_s$	$\Delta\Gamma$	$R_\perp$	R_0	δ_1	δ_2	Δm_s	$-2\beta_s$	$\omega_{\rm tag}$
$\bar{\Gamma}_s$	1.0	-0.64	0.50	-0.22	-0.02	0.01	0.00	-0.02	0.01
$\Delta\Gamma_s$		1.0	-0.66	0.23	0.01	-0.01	0.00	-0.02	-0.01
$R_\perp$			1.0	-0.65	-0.	0.05	-0.01	0.03	0.04
R_0				1.0	0.12	-0.07	0.01	-0.01	-0.03
δ_1					1.0	0.10	0.25	0.00	-0.03
δ_2						1.0	0.27	0.00	0.00
Δm_s							1.0	0.00	-0.03
$-2\beta_s$								1.0	-0.03
$\omega_{\rm tag}$									1.0

Table 6.10: Baseline results: Correlation coefficients for simultaneous fit to all parameters using tagged events

6.5 Sensitivity results using a three decay angle analysis

This section describes the results obtained using the full three decay angle analysis and the fitting procedure described in section 6.3. The study presented here includes both background contributions and neglects acceptance effects on the propertime and decay angles. We use the full 133k event sample, with the events classified as either a b, $\bar{b}$ or untagged. All three tagging categories are then used in the fit[10].

In Table 6.12 the baseline results are presented where $\Delta\Gamma_s$, $\bar{\Gamma}_s$, $R_\perp$, R_0, δ_1, δ_2, Δm_s and $-2\beta_s$ are determined simultaneously. The sensitivity on $-2\beta_s$ is found to be $\pm$ 0.037 rad. The correlation matrix is shown in Table 6.13, and closely follows the correlation structure shown in Table 6.10. We also note that the strong phases are now heavy correlated between themselves and the unconstrained Δm_s. We also presented in Table 6.12 the simultaneous fit results using the latest LHCb experimental parameter values [41], which are given in Table 6.11. The sensitivity for $-2\beta_s$ coming from column (b) in Table 6.12 is found to be in good agreement with that quoted in [41], but is $\sim$ 20% different from the corresponding value quoted in column (a). The reason can attributed to the different $\omega_{\rm tag}$ values used. The fitted values, errors and pulls distributions for each study in Table 6.12 are shown in section D.1.

yield [k]	$\left(\frac{B}{S}\right)^{b\bar{b}-\text{inclusive}}_{B_s\to J/\psi\phi}$	$\left(\frac{B}{S}\right)^{\text{prompt}-J/\psi}_{B_s\to J/\psi\phi}$	σ_τ [fs]	ω_{tag} [%]	ε_{tag} [%]
130	0.12	1.9	40 $\pm$0.5	33 $\pm$1.0	57.1 $\pm$1.0

Table 6.11: Latest LHCb experimental parameter numbers for $B_s \to J/\psi\phi$ decay, taken from the CKM 2008 conference [41].

[10]The three-angle differential cross section components for tagged B_s decays are given in section 6.4.2.

Parameter	Input	Sensitivity	
		(a) Using Table 6.2	(b) Using Table 6.11
$\bar{\Gamma}_s$	0.68	0.0033	0.0036
$\Delta\Gamma_s$	0.1	0.012	0.012
$R_\perp$	0.20	0.0045	0.0047
R_0	0.60	0.0041	0.0044
δ_1	-0.46	0.11	0.091
δ_2	2.97	0.12	0.12
Δm_s	17.77	0.051	0.041
$-2\beta_s$	-0.04	0.037	0.030

Table 6.12: Simultaneous three-angle fit to all parameters using tagged events, using: (a) experimental parameters values given in Table 6.2 and (b) in Table 6.11.

	$\bar{\Gamma}_s$	$\Delta\Gamma$	$R_\perp$	R_0	δ_1	δ_2	Δm_s	$-2\beta_s$
$\bar{\Gamma}_s$	1.0	-0.75	0.60	-0.5	-0.015	-0.055	-0.016	-0.025
$\Delta\Gamma_s$		1.0	-0.67	0.70	0.007	0.037	0.007	0.033
$R_\perp$			1.0	-0.89	-0.07	-0.23	-0.08	-0.028
R_0				1.0	0.06	0.18	0.065	0.023
δ_1					1.0	0.89	0.76	0.021
δ_2						1.0	0.71	0.02
Δm_s							1.0	0.027
$-2\beta_s$								1.0

Table 6.13: Baseline results: Correlation coefficients for simultaneous fit to all parameters (given in Table 6.12) using tagged events.

6.6 Extracting $\omega_{\rm tag}$ and knowledge of the strong phases.

As indicated in the above studies, the data appears rich enough to be able to extract $\omega_{\rm tag}$, at least in the situation where $-2\beta_s$ is close to its expected SM value.

This is understood as follows: consideration of the form of the differential cross section shows that the relevant (imaginary interference) terms may be recast to read (when $-2\beta_s = 0$)

$$(1 - 2\omega_{\rm tag}) \times \sin(\delta - \Delta m_s t) \tag{6.24}$$

where we see that the fit can determine the sinusoid period (Δm_s) as well as its phase offset from the time distribution alone. This would leave the amplitude information to determine $(1-2\omega_{\rm tag})$.

This expectation is true for all values of the two strong phases, and was tested by performing series of fits for the strong phases, set to combinations of: 0, $\pi/4$, $\pi/2$, π and $3\pi/2$. In addition to this, we used the world average values of $\delta_1 = -0.46$ and $\delta_2 = 2.92$ from $J/\psi K^*$ decays [19, 91]. In all cases the fits were found to behave well and the errors parabolic for all parameters, except for the δ's.

When $-2\beta_s \neq 0$ the situation is more complicated. A similar recasting can be done to give

$$\frac{(1 - 2\omega_{\rm tag})}{2} \times \{(1 + \cos(2\beta_s))\sin(\delta - \Delta m_s t) + (1 - \cos(2\beta_s))\sin(\delta + \Delta m_s t)\} \tag{6.25}$$

It is somewhat more difficult to make the simplistic arguments of factorisation, but it is certainly true that:

- there is phase offset information related to the δ's,
- there is amplitude information related to $(1 - 2\omega_{\rm tag})(1 \pm \cos(2\beta_s))$, and
- there is amplitude information related to $(1 - 2\omega_{\rm tag})\sin(2\beta_s)$ from the other diagonal and real interference cross section components.

It is thus plausible that the fit has enough information to separately determine all quantities. To test this we have performed a series of fits with $-2\beta_s$ ranging from $0 \rightarrow \pi$. We have also performed a set of fits with $-2\beta_s = -0.8$ and the strong phases set to combinations of 0, $\pi/4$, $\pi/2$, π and $3\pi/2$ and also the set $\delta_1 = -0.46$, $\delta_2 = 2.92$. In all cases we find the fits well behaved and the errors parabolic for all parameters, except for the δ's. To clarify this, some of the distributions are shown in Figures D.21 to D.24 in Appendix D.

However, the situation in reality is not as simple as these rather idealised studies would suggest. The Log Likelihood (LL) scan for δ_1 nicely illustrates this. The studies undertaken have used the luxury of starting all fit parameters close to their generated values. In this case the fits

tend to always converge to the correct central values. For well behaved parabolic parameters this is safe and for those for which there will be external constraints available (such as $\bar{\Gamma}_s$, Δm_s and $\omega_{\rm tag}$) it is even more safe. However, the situation for the strong phases is evidently rather different. The LL scans are not parabolic and in the worst scenario a second false minimum can be found, which would in turn affect other parameters.

In practise this will be controlled by performing a series of analyses, with different assumptions about strong phases. LL scans should be performed to understand the structure of the LL surface. The $\omega_{\rm tag}$ parameters will be compared to, and probably at least loosely constrained by, the value obtained from control channels. However, it will always remain true that care will be needed in this sector.

6.7 Angular acceptance studies

In our toys we choose not to model the effect of angular acceptance. However, studies carried out using fully simulated data as shown in section 6.2.2, show a non negligible acceptance effect for the decay angles. In this final section, the effect of applying varying $p_\perp$ cuts on both muons and kaons, with respect to the decay angular distributions will be investigated.

To isolate the effect of varying the $p_\perp$ cut, an event generator-level (EVTGEN) study was performed, using a set of 75 $B_s \rightarrow J/\psi\phi$ data samples (see Table 3.2) each containing 300k events. The $p_\perp$ cut was varied from 100 $\mathrm{MeV/c}$ to 1.2 $\mathrm{GeV/c}$ in steps of 50 $\mathrm{MeV/c}$ for the following three situations:

(1) applying the $p_\perp$ to both muons and kaons,

(2) applying the $p_\perp$ to muons only and,

(3) applying the $p_\perp$ to kaons only.

The effect these cuts have on the individual angular acceptance are shown in Figures: D.8, D.9 and D.10 for the angle θ_{tr}, D.11, D.12 and D.13 for the angle ϕ_{tr} and D.14, D.15 and D.16 for the angle θ_ϕ. In these plots, the $p_\perp$ cut is increased going from the top to bottom. The effect of $p_\perp$ cuts on both muons and kaons is shown in the left-hand plots, the effect on just the muons in the middle plots and the effect only on kaons in the right-hand plots.

$p_\perp$ studies with three decay angles

Simultaneously fitting for the parameters $\{\bar{\Gamma}_s,\ \Delta\Gamma_s,\ R_\perp,\ R_0,\ -2\beta_s\}$, whilst increasing the $p_\perp$ cut on both muons and kaons, we observe a bias[11] on the fitted values for $R_\perp$ and R_0 as shown in Figures D.19 and D.20. The bias on these parameters can be seen to depend on the $p_\perp$ cut applied to the kaons only. This can be understood from the $p_\perp$ distribution of kaons and muons shown in Figure 6.14. We see from this figure that kaons favor lower values of $p_\perp$ than the muons. From the differential decay rate expressions, we see that it is the angular information which greatly helps in obtaining our sensitivity to the physics parameter $-2\beta_s$. When we alter the $p_\perp$ cut and incur a perceived bias on the fraction of $\mathcal{CP}$-even and -odd components, we inevitably cause a bias on $-2\beta_s$. This bias is shown in the top plot in Figure 6.15, which increases from essentially nothing for a 100 $\mathrm{MeV/c}$ cut to a 1σ effect after applying a 1.2 $\mathrm{GeV/c}$ cut. Here we have used the average statistical error.

In the three angular case applying $p_\perp$ cuts to muons and/or kaons generates a 3-D angular acceptance effect, which has been shown to slightly bias $-2\beta_s$. To correct for this, the acceptance curve ($\eta(\Omega)$) describing the angular sculpting due to the detector, needs to be introduced. This will lead to normalisation angular dependent integrates of the form:

$$\int \Theta_k(\Omega)\eta(\Omega)d\Omega, \qquad \text{k=1,2...6} \tag{6.26}$$

in the Likelihood function (equation (6.11)). It is possible for LHCb collaboration to correct for this, by use a normalised 3-D histogram of the angular acceptance distribution, modelled from Monte Carlo truth data (i.e. EVTGEN data). The integral in equation (6.26), will then be in the form of a mathematical function multiplied by an histogram, and will need to be computed numerically. One simple approach is to approximate this integral to a discrete sum over the histogram bins. The integral can then be expressed as:

$$\int \Theta_k(\Omega)\eta(\Omega)d\Omega \approx \sum_{j=1}^{N_{Bins}} \frac{[F_k(\Omega_j^{\max}) - F_k(\Omega_j^{\min})]\cdot\eta(\Omega_j^{\text{med}})}{\text{BinVol}}, \qquad \text{k=1,2...6} \tag{6.27}$$

[11] The error bars drawn here are the average statistical error, defined for a parameter X by:

$$\begin{aligned}
\underbrace{\langle Pull_X\rangle\langle\sigma(X)\rangle}_{\langle f\rangle} &= \underbrace{\langle X_{Fit}\rangle}_{\bar{X}} - \underbrace{X_{Gen}}_{\text{constant}},\\
\langle f\rangle &= \langle(\bar{X}+X_{Gen})^2\rangle - \langle\bar{X}+X_{Gen}\rangle^2,\\
&= \langle\bar{X}^2\rangle + 2X_{Gen}\langle\bar{X}\rangle + X_{Gen}^2 - \langle\bar{X}^2\rangle - 2X_{Gen}\langle\bar{X}\rangle - X_{Gen}^2,\\
&= \langle\bar{X}^2\rangle - \langle\bar{X}\rangle^2,\\
&= V(\bar{X}) = \frac{\sigma_X}{\sqrt{N}}.
\end{aligned}$$

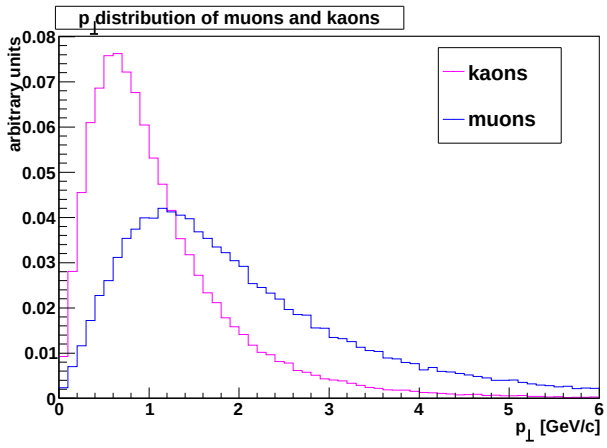


Figure 6.14: The $p_\perp$ distribution of K^+ (magenta) and μ^+ (blue) from $B_s \rightarrow J/\psi\phi$ EVTGEN data. The distributions shown are normalised to an arbitrary scale.

Where,

- $\sum_{j=1}^{N_{\rm Bins}}$, is the sum is over the bins along $\cos(\theta_{tr})$,ϕ_{tr} and $\cos(\theta_\phi)$.
- BinVol, is the bin volume, $\Delta\cos(\theta_{tr}) \times \Delta\phi_{tr} \times \Delta\cos(\theta_\phi)$.
- $F_k(\Omega)$. We have assumed each function $\Theta_k(\Omega)$ can compute the three dimensional angular integral analytically $\int \Theta_k(\Omega)d\Omega = F_k(\Omega), \qquad k = 1, 2, \ldots 6.$
- $\Omega_j^{max/min}$, is the bin value of Ω_k at the upper/lower edge of the 3-D space.
- Ω_j^{med}, is given by $\frac{\Omega_j^{min}+\Omega_j^{max}}{2}$.

Full details of this procedure can be obtained in [93]. Although this method is simple, it suffers from needing a very large amount of MC data to parametrise the 3-dimensional acceptance surface, at the required level of precision and resolution. The approach adopted by LHCb is follow that employed in [94], which does not rely on the exact shape of the acceptance function. The normalisation angular integral given in 6.26 is then:

$$\int \Theta_k(\Omega)\eta(\Omega)d\Omega \approx \sum^{acc\ events} \frac{\Theta_k(\Omega)}{\int h_k(t)\cdot\Theta_k(\Omega)dt} \qquad \text{k=1,2...6,} \tag{6.28}$$

where 'acc events' is the number of events which can be accepted in $\eta(\Omega)$, and where the approximation now provides a weighing factor for each angular dependent term $\Theta_k(\Omega)$. In this thesis we have not corrected for the systematic effect caused by angular acceptance. However, any extensions to these studies will need to account for these effects.

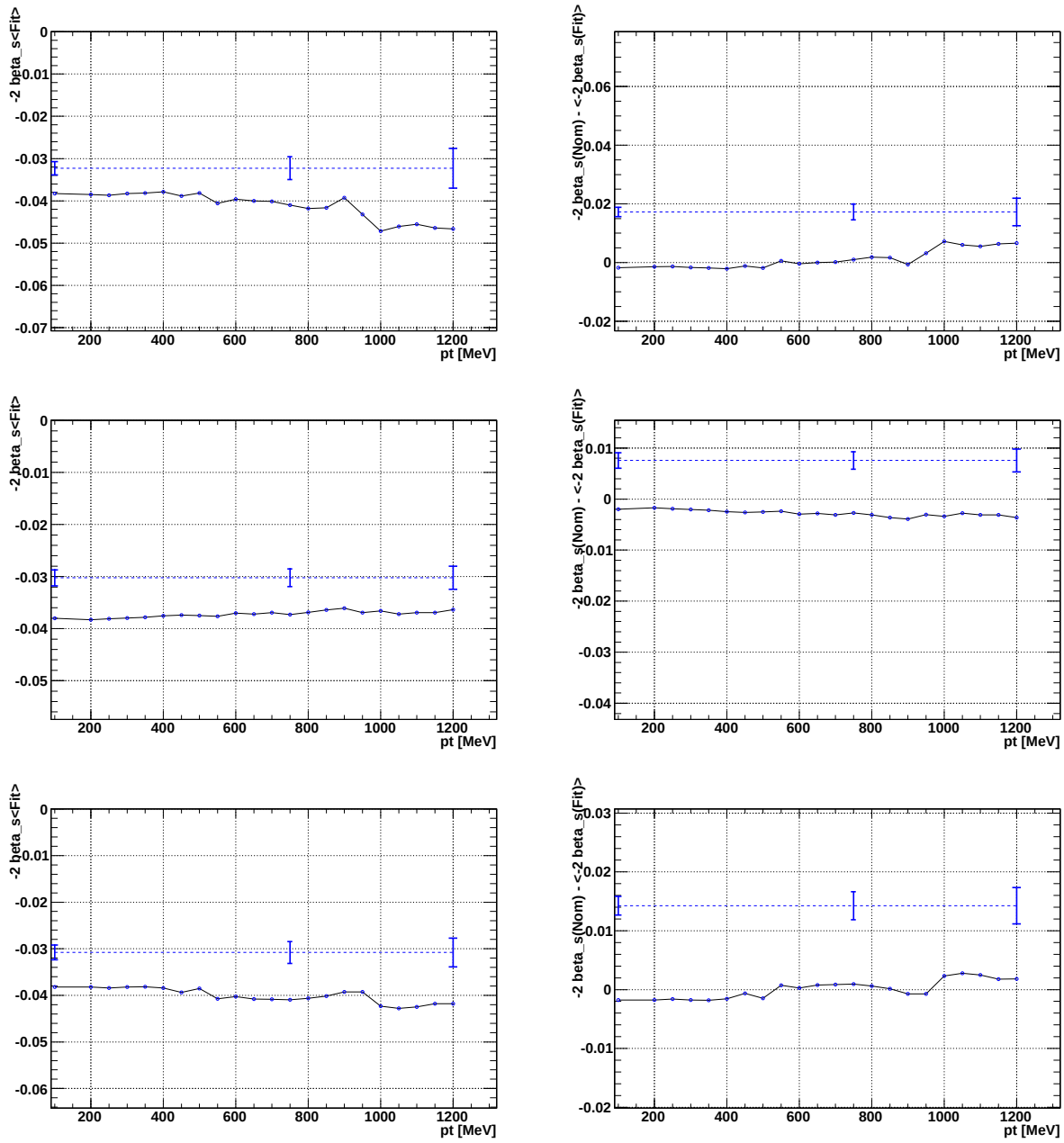


Figure 6.15: Shown in the left-hand plots is the mean fit value for $-2\beta_s$ using 75 EVTGEN data samples, obtain from a full angular fully tagged the simultaneous fit. The corresponding plots on the right-hand side shows the bias in the fitted value. The top plots illustrate the effect of applying a $p_\perp$ cut on both muons and kaons. The middle plot show the effect for the $p_\perp$ cut on muons only, while the bottom set of plots shows the $p_\perp$ effect on kaons only. The error bars shown at 100, 750, 1200 $\mathrm{GeV/c}$ show the standard error.

6.8 Systematic error studies

The statistical precision on $-2\beta_s$ will not only depend on the number of reconstructed events and the angular analysis strategy used, but also on the level of background present, the level of the mistag and propertime resolution, and other detector or model effects. In Table 6.14, based on previous analyses [6, 33], we give a non exhaustive list of the expected systematic effects on the $-2\beta_s$ measurement. Although extensive systematics studies were not performed on the measurement of $-2\beta_s$ during this work, we performed additional studies to investigate the effect of varying the central values of $\bar{\Gamma}_s$, $\Delta\Gamma_s$, $\omega_{\rm tag}$ $R_\perp$, R_0 and $-2\beta_s$ on the $-2\beta_s$ sensitivity.

6.8.1 Variations

The effect of the central value of $\bar{\Gamma}_s$, $\Delta\Gamma_s$, $\omega_{\rm tag}$, $R_\perp$, R_0 and $-2\beta_s$ on the $-2\beta_s$ sensitivity was also studied. Figure 6.16 shows the results obtained in the one angular analysis and Figure 6.18 shows the results obtained in the full angular analysis. Figure 6.17 shows the effect of the propertime resolution on both the one angular and three angular analysis. In all cases the quoted sensitivity is the RMS of the fit distribution for more than 300 experiments in the ideal case. We observe the following:

-. the variation on $\bar{\Gamma}_s$ appears to have on the average no effect on the $-2\beta_s$ sensitivity

-. an increment in the $\Delta\Gamma_s$ central value shows a tendency to improve the precision $-2\beta_s$. This is due to a better separation of the lifetime exponents Γ_L and Γ_H

-. as expected, the $-2\beta_s$ sensitivity has a correlation to the mistag fraction $\omega_{\rm tag}$. As the $\omega_{\rm tag}$ values increases, the precision on $-2\beta_s$ reduces significantly

-. the $-2\beta_s$ sensitivity is slightly affected by the central value of $|-2\beta_s|$ as it increases. This can be explained by the fact that $\sin(-2\beta_s)$ terms are not anymore small and therefore $-2\beta_s$ picks up correlations with other parameters usually suppressed in the limit of a small $|-2\beta_s|$ value.

Systematics	
Source	Effect
Tracking & alignment	Misalignment of the r and Θ sensors, and of the VELO modules in each half of the VELO detector, can cause a significant deterioration of the B_s^0 propertime resolution [95]. In some misalignment scenarios this can be as much as 60 fs [96]. As shown in Figure 6.17, deterioration of B_s^0 propertime resolution is shown to reduce precision on the $-2\beta_s$ measurement.
Production asymmetry	In our studies we have assumed that there is no production asymmetry. However, the LHC has an inherit non zero production asymmetry which can mimic and pollute the $\mathcal{CP}$ we are measuring. This asymmetry is expected to be of the order of a percent, with an uncertainty at the same level as our $-2\beta_s$ sensitivity: $-(1.9 \pm 1.3)\%$ for B_s^0 mesons [97]. This asymmetry therefore needs to be measured in order to disentangle it from our $-2\beta_s$ measurement. A measure of the production asymmetry in $B_s \to J/\psi\phi$ decays, can be obtained using a control channel such as $B_s \to D_s^- \pi^+$ or $B_s \to D_s^- \mu^+ \nu$.
Flavour tagging	From Figures 6.16 and 6.18 the effect of tagging is shown to cause a major bias on the $-2\beta_s$ sensitivity for both the one and three angular analysis. In section 6.6, we show that is possible to measure the mistag fraction $\omega_{\rm tag}$ from the data itself, with a precision of $\pm$ 0.010. We can therefore minimise this systematic effect by allowing $\omega_{\rm tag}$ be free in the fully tagged three angular analysis.
Background fractions	The fraction of prompt background is found to have a very limited effect on the $-2\beta_s$ sensitivity [41]. The long lived background fraction however, is found to cause a linear bias on the $-2\beta_s$ sensitivity. This is because the propertime distribution of the prompt background is centred at 0 ps and does not contaminate the signal where the oscillations occur.
Propertime resolution	The propertime resolution is found to cause a significant bias on the $-2\beta_s$ sensitivity. As illustrated in Figure 6.17 for the one and three angular angular analysis, we see that the value of propertime resolution has a significant effect on the $-2\beta_s$ sensitivity. A similar bias has been observed for the three angular analysis at ATLAS and CMS [41].
Decay angles	The distribution of the $B_s \to J/\psi\phi$ decay angles are directly effected by the contribution to $R_\perp$ and R_0. In Figure 6.18, we see the $-2\beta_s$ sensitivity remains unaffected whilst scanning over the given $R_\perp$ and R_0 range. However, the effect on $-2\beta_s$ due to non flat distribution of the decay angles with respect to background (Figures 6.12 and 6.13), has still to be investigated.

Table 6.14: Non exhaustive list of systematic effects which can bias the $-2\beta_s$ measurement in $B_s \to J/\psi\phi$ decays.

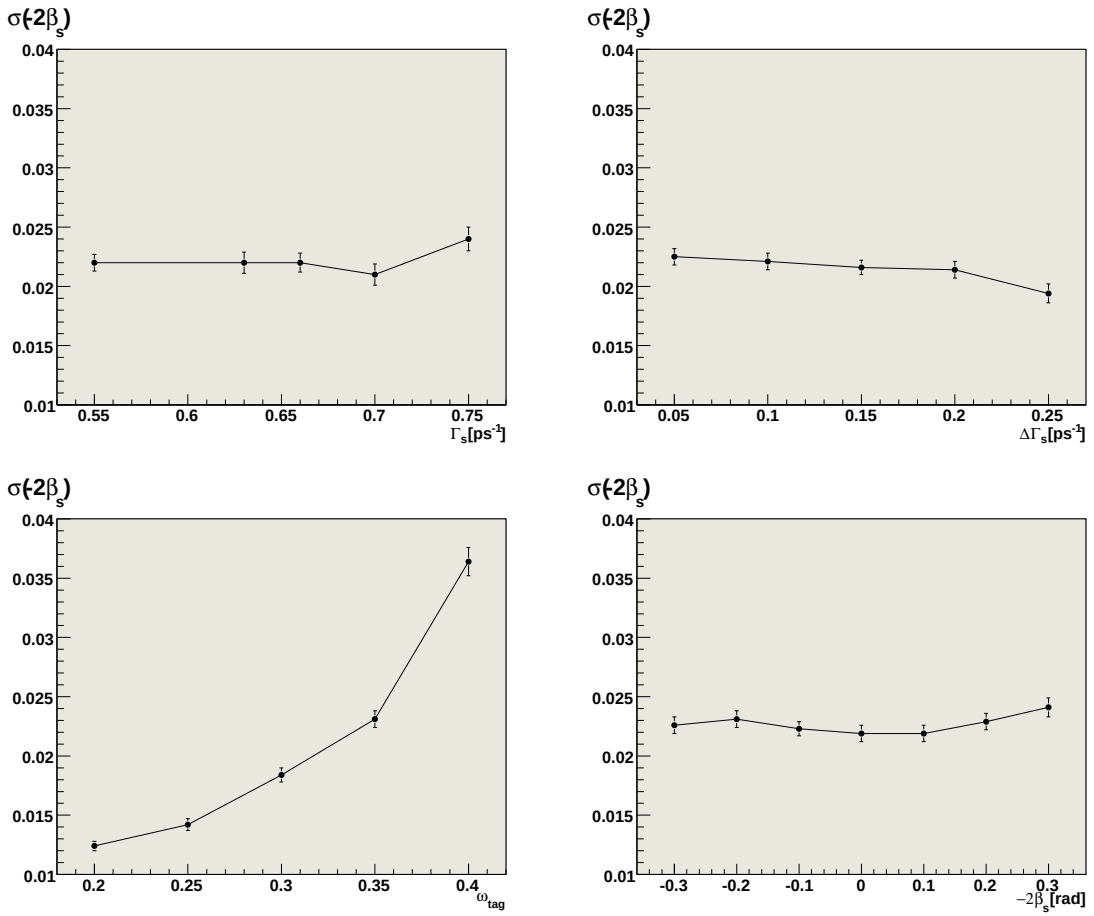


Figure 6.16: One-angle variation studies: From top-left we show here the effect on the $-2\beta_s$ sensitivity when the central value of $\bar{\Gamma}_s$, $\Delta\Gamma_s$, $\omega_{\rm tag}$ and $-2\beta_s$ is varied

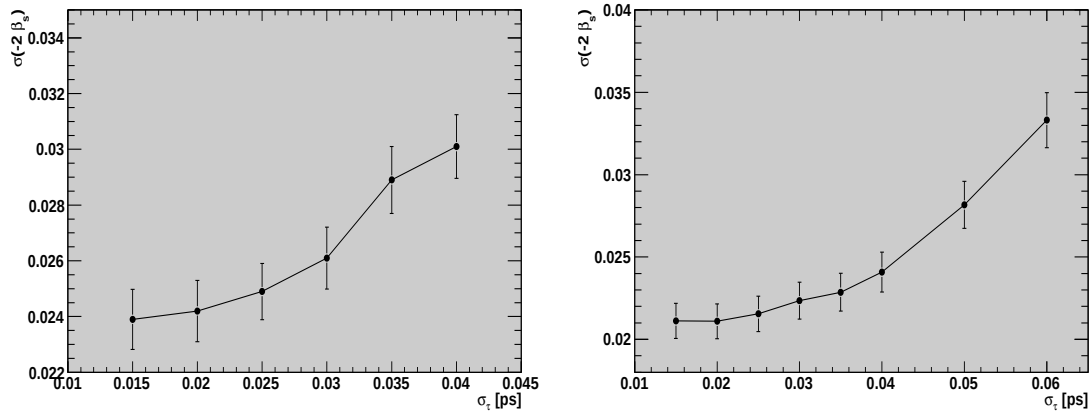


Figure 6.17: Variation studies on $-2\beta_s$ when the central value of σ_τ is varied. (Left) For the one angular analysis and (right) for the three angular analysis.

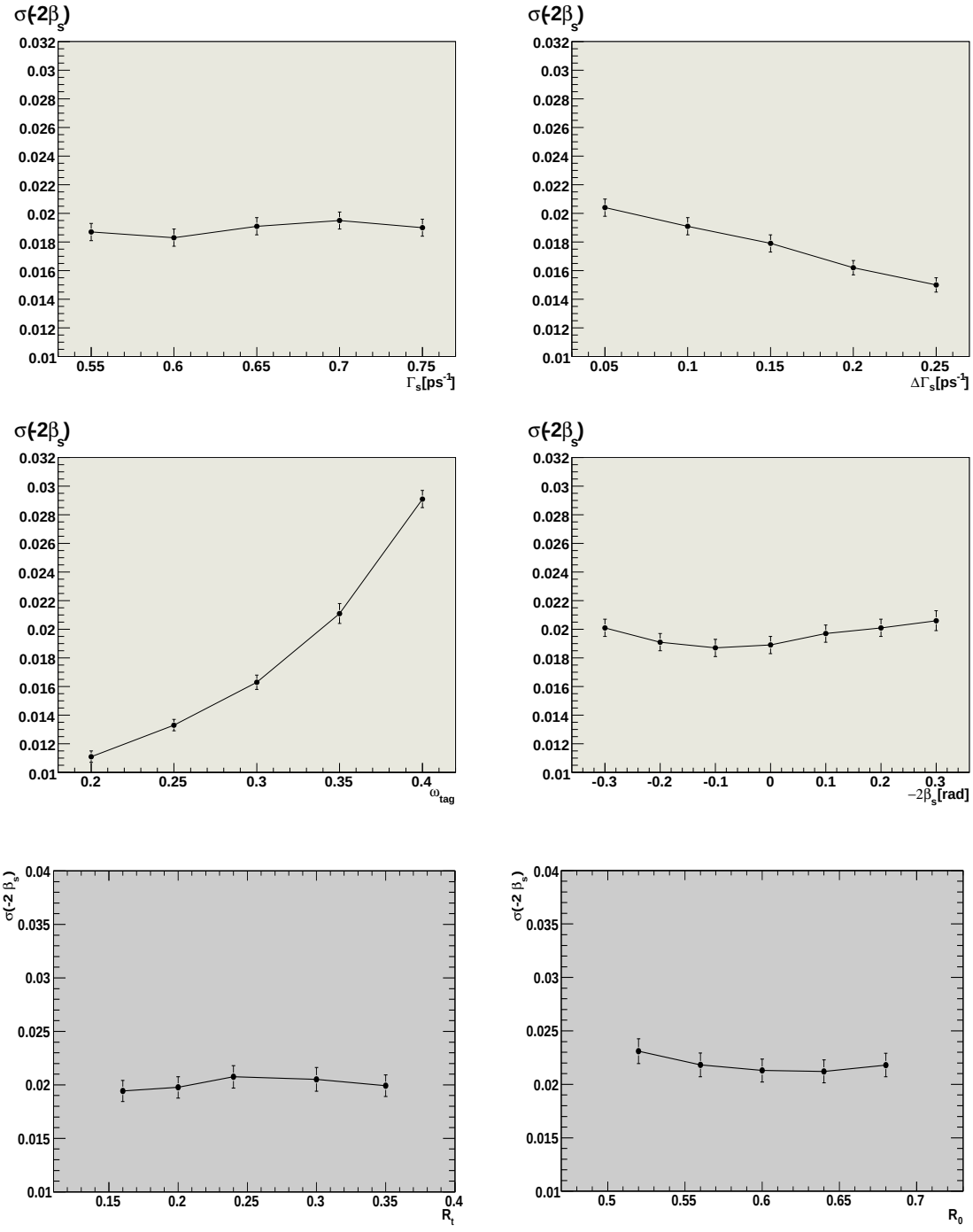


Figure 6.18: Three-angle variation studies: From top-left we show here the effect on the $-2\beta_s$ sensitivity when the central value of $\bar{\Gamma}_s$, $\Delta\Gamma_s$, $\omega_{\rm tag}$, $-2\beta_s$, $R_\perp$ and R_0 is varied

6.9 Conclusion

In these studies we have presented work which extends previous studies of the sensitivity of LHCb to the B_s mixing phase $-2\beta_s$ obtained in the $B_s \to J/\psi\phi$ channel, by including a full analysis using the three measurable decay product angles. We find the following:

The precision upon $-2\beta_s$ including both background contribution as described in section 6.3, is $\pm$ 0.037 rad, for one nominal year of data taking.

From a one to three angular analysis the precision upon $-2\beta_s$ improves by approximately 20%. The precision upon the measurements of $\bar{\Gamma}_s$ and $\Delta\Gamma_s$ are improved and the correlation between them reduced. The precision on $\bar{\Gamma}_s$ is improved from ± 0.0034 to ± 0.0027 and the precision on $\Delta\Gamma_s$ shows a large improvement from ± 0.014 to ± 0.0079. This is due to the improved statistical power in separating the $\mathcal{CP}$-odd and $\mathcal{CP}$-even components.

The mistag fraction, $\omega_{\rm tag}$, can be obtained from the data themselves with a precision of ± 0.010. This provides qualitative improvement through independence from, or at least the ability to cross check with a control channel.

In the final chapter, the sensitivity to $-2\beta_s$ will be investigated by making using of Multivariate analysis techniques to first optimise the selection cuts.

7

Optimising the selection cuts

7.1 Introduction

Particle physics events are in essence multivariate. By combining all possible pieces of information associated with these events, we would expect to gain at least the same level of discriminating power over regular cut-based techniques (as discussed in chapter 4), where each event has to satisfy a set of selection criteria. The hope is that by using all available information associated with these events, we can use multivariate techniques to increase the separation of signal and background, and thus the statistical power of a measurement.

The purpose of this chapter is to investigate the usefulness of multivariate classifiers in selection of the $B_s \to J/\psi\phi$ data.

A selection of linear, non-linear and shape (or parametrised) based classifiers will be used. The chapter begins by introducing each classifier, before discussing the procedure adopted to determine which classifier gives the optimal signal-to-background separation. We also give a brief discussion on the multivariate visualisation technique, Parallel Coordinates in section E.4, for visualising our event data. The optimal classifier was found to be the Boosted Decision Tree (BDT), which was then used to optimise the $\frac{B}{S}$ ratio and to subsequently investigate the sensitivity of $-2\beta_s$.

7.2 Multivariate Analysers (MVA)

7.2.1 The Fisher Linear discriminant

Suppose the result of measuring an individual event is a vector of attributes, $\vec{x} = x_1, \ldots, x_n$ where:

$$\begin{aligned} x_1 &= \text{number of muons,} \\ x_2 &= \text{mean } p_t \text{ of muons,} \\ &\vdots \\ x_n &= J/\psi \text{ mass.} \end{aligned}$$

The collection of attributes will follow some n-dimensional probability distribution function (p.d.f) which depends on the type of events we have. The events may be signal-like ($B_s \rightarrow J/\psi\phi$) or background-like (prompt-J/ψ) for instance. We assume the p.d.f, labelled by f, depends on two categories which determine whether the event is signal-like or not such that, $f(\vec{x}|H_0)$ and $f(\vec{x}|H_1)$ define the signal-like and background-like p.d.fs. Here H_0 is the signal category and H_1 is the background category. When we consider selecting events of type H_0, the question arises how to efficiently find the multidimensional boundary in order to accept events belonging to this type. The Fisher linear discriminant seeks this discriminant searching for a hyperplane, which allows the separation of data projected on a line perpendicular to this plane. A cartoon of this idea is shown in Figure 7.1.

The separation between signal and background depends on the mean of the signal and background distributions and on the spread of the data points about their respective means along the line orthogonal to the hyperplane. The plane of best separation requires maximising the separation between each sample's mean, whilst at the same time minimising each samples variance, so that we maximise the following criteria:

$$\text{separation} = \frac{|(\mu_{H_0} - \mu_{H_1})|^2}{\sigma^2_{H_0} + \sigma^2_{H_1}}. \tag{7.1}$$

Where the mean of each sample is defined as:

$$\mu_D = \frac{1}{N_D} \sum_{\vec{x}^j=1}^{N_D} \sum_{i=1}^{n} F_i x_i^j, \tag{7.2}$$

with $D \in \{H_0, H_1\}$ the number of signal and background events, n the number of variables in the n-dimensional space and F_i the coefficients defining the orientation of the hyperplane. Figure 7.2 shows how the data given in Figure 7.1 is projected onto the line of best separation

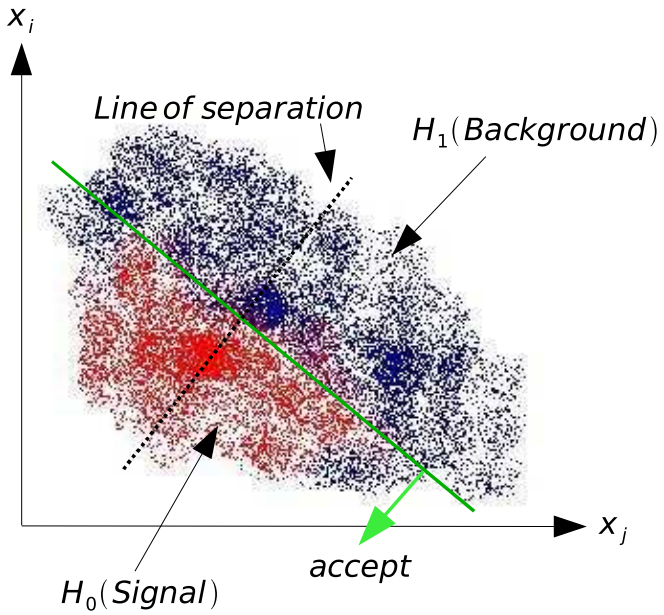


Figure 7.1: The distribution of signal-like events is coloured in red and background-like events coloured in blue, in the n-dimensional hyperspace. The hyperplane giving the best separation is coloured in green, with the line of projection orthogonal to this coloured black.

(the black line). A more rigorous description of the Fisher linear discriminant [98] is given in section E.1, however the specific definition which is used to classify the i^{th} event into either signal or background is given by:

$$y_{Fi}(i) = F_0 + \sum_{k=0}^{n} F_k x_k(i), \tag{7.3}$$

where F_k are called the Fisher coefficients and F_0 the offset, specifying where the hyperplane is positioned with respect to the data sample. The response of the Fisher discriminant is either $y_{Fi}(i) > 0$, which specifies signal-like events or $y_{Fi}(i) < 0$, which specifies background-like events.

7.2.2 Parametrised approach (Likelihood)

Likelihood discriminators are one of the most straightforward multivariate analysers. The likelihood classifier requires knowledge of the p.d.fs for both signal (P_S) and background (P_B), obtained from a vector of discriminating variables $\vec{x}$ after each event selection. The likelihood expressions for signal ($\mathcal{L}_S$) and background ($\mathcal{L}_B$) are determined from the product of the signal and background p.d.fs for each input variable (therefore potential correlations between variables

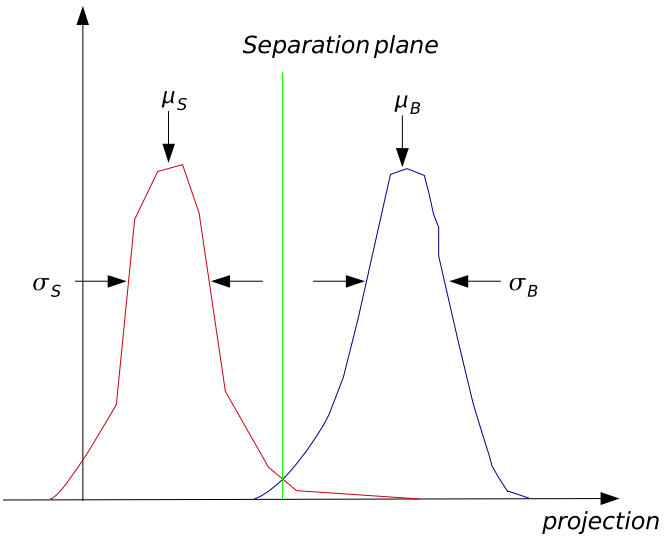


Figure 7.2: Projection of signal and background data given in Figure 7.1 along the direction of best separation.

are not taken into account), such that:

$$\mathcal{L}_S = \prod_{i=1}^{N_S} P_S(\vec{x}_i), \quad \mathcal{L}_B = \prod_{i=1}^{N_B} P_B(\vec{x}_i), \tag{7.4}$$

where $\int P_{S(B)}(\vec{x}_i)d\vec{x}$=1 and $N_{S/B}$ is the total number of signal or background events. A final discriminating variable can be constructed using this set to characterise each event as being either signal or background. The likelihood of the total sample is then given by:

$$y_{Li} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}. \tag{7.5}$$

Defined in this way, y_{Li} tends to be peaked at 1 for signal-like events and 0 for background-like events. The likelihood classifiers performance is known to suffer if there are intrinsic correlation amongst the input variables. If however, the input variables are Gaussian distributed and linearly correlated, an improved performance is possible by first decorrelating each of the input variables.

7.2.3 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN)s are widely used in particle physics providing a powerful method to separate signal data from its background. The D0 collaboration for instance, have been using ANNs extensively in single top quark searches [99, 100].

The goal of ANNs is to simulate on a computer, a parallel distributive system of many interactive elements (neurons) to correctly recall a response (i.e. if the data is likely to be signal or background) given some initial data. This effectively defines what we consider a ANN, which is any simulated collection of connected neurons with each neuron producing a response to a given set of input variables. Supplying the variables to the input neurons puts the network into a defined state that can be measured from the response of one or more of the output neurons. The idea of the MultiLayer Perceptrons (MLP) ANN is shown in Figure 7.3, where one hidden layer of neurons has been used. The reader can obtain a detailed description of ANNs in section E.2.

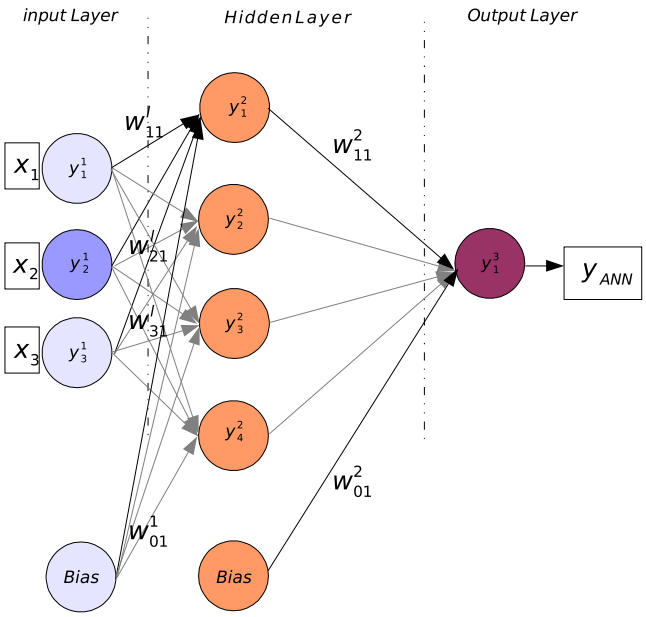


Figure 7.3: MultiPerceptron Layer with one hidden layer. The input variables to the input neurons ($1st$ layer) are denoted x_i, the output of each input neuron is y_i^1 and the weight associated to each input neuron is w_{ij}^1, where j denotes the j^{th} next neuron layer (in this example the hidden layer). The response of each neuron in the final hidden layer is then fed into the output neuron y_1^3, which given the ANNs response.

The specific response of an ANN, with one hidden layer[1] is given as:

$$y_{ANN}(i) = \sum_{j=1}^{n_h} y_j^{(2)} w_{jl}^{(2)} = \sum_{j=1}^{n_h} \tanh\left(\sum_{k=1}^{n} x_k(i) w_{kj}^{(1)}\right) \cdot w_{j1}^{(2)}. \tag{7.6}$$

[1] A $\tanh$ activation function has been used in the hidden layer.

7.2.4 Boosted Decision Trees (BDT)

The boosted decision tree is a machine learning technique widely used outwith particle physics [101]. The goal of decision trees is to extend the simple cut-based analysis into a multivariate technique by continuing to analyse events that fail a particular criterion. Mathematically, a decision tree is a sequence of binary operations (AND, OR) used to split the data into signal and background (Figure 7.4). Trees are trained using a set of known signal and background events (the training sample) and tested using a separate set of data, the test sample. The algorithm begins by considering all events to start on one node, the root node. All the events are then separated in turn into two children nodes - one most likely to be signal, the other mostly to be background - using the discriminating variable, taken from the variables set $\vec{x}$, which gives the best calculated separation value, or the splitting decision, for the given set of events. Events which pass the splitting decision are labelled as signal (S), while those which fail the decision are labelled as background (B). The algorithm is then applied recursively to these new nodes. The splitting stops when a given number of final nodes (called leaf nodes) are obtained, or until a node has too few events. The leaf nodes are classified as signal or background depending on the majority of events belonging to it, or its associated purity (signal fraction over the total training sample in this node). The most important part of the decision tree building process is determining the goodness of separation of signal and background events (the splitting criteria) and the optimal value at which to do this. If we assume each event is given a weight W_i, then the purity of a sample of events on the t^{th} node is:

$$P(t) = \frac{\sum_S W_S}{\sum_S W_S + \sum_B W_B}, \tag{7.7}$$

where $\sum_S W_S$ is the sum of signal event weights and $\sum_B W_B$ is the sum of background event weights. From equation (7.7), it is clear that $P_t(1 - P_t)$ will be 0 if the sample is pure signal or pure background. This allows us to define an index (or figure of merit) as to the purity of a given node. The Gini index $G(t)$, which has been used by both MiniBooNe for detecting electron neutrinos [102] and D0 for single top quark production [103], is defined as:

$$G(t) = \left(\sum_{i=1}^{n} W_i \right) P_t(1 - P_t), \tag{7.8}$$

where n is the number of events on the t^{th} node. The training algorithm uses the Ginix index to determines whether-or-not a node should be split into children nodes or not. Node splitting occurs if the separation between the Gini index of the parent node, and the sum of the Gini indies of the two children nodes is maximal[2]:

$$\text{Criterion}(t) = G(t) - \left(G(t_{\text{left child}}) + G(t_{\text{right child}}) \right). \tag{7.9}$$

[2]The children nodes can the labelled as left (or pass), and right (or fail)

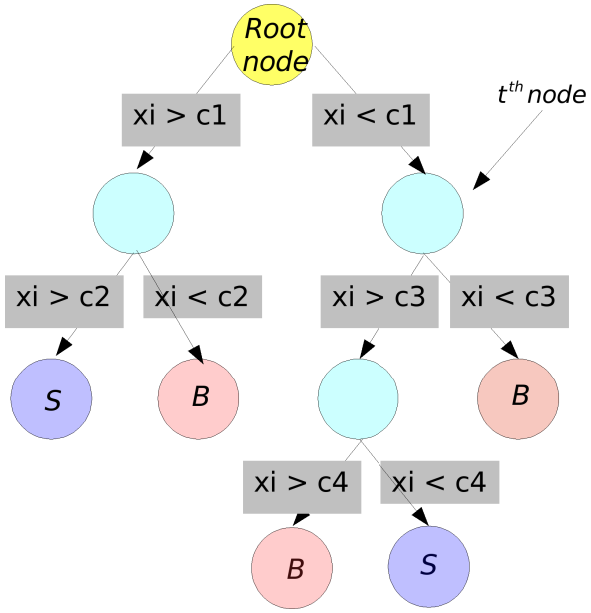


Figure 7.4: A Decision Tree (DT). The t^{th} node, where the splitting criteria (given in equation (7.9)) applied, has been highlighted.

Decision trees are powerful in discriminating signal and background, but unstable. A small statistical fluctuation in the training sample can produce a large change in the tree structure. This problem can be remedied by constructing a collection of decision trees and classifying the events using a procedure known as boosting [104]. With boosting, the training events which where misclassified, i.e. a signal event falling on a background leaf or vice versa, have their weights increased (or boosted) and a new tree is formed. In this way many trees are build up, a forest of say N_{Trees} trees. The classifier result for the i^{th} tree $T_i(\vec{x})$ is taken as +1 if the event falls on a signal leaf and -1 if the event fall on a background leaf. The final classifier result is then taken as a weighted sum of the individual trees: the idea of boosting a decision tree is shown in Figure 7.5.

There are several algorithms commonly used for boosting the weights of misclassified events [105]. The most popular algorithm is called AdaBoost (ADAptive Boost), the details of this algorithm can be found in section E.3. The resulting event classification $\text{y}_{\text{BDT}}(\vec{x})$ for the boosted classifier is then given by,

$$\text{y}_{\text{BDT}}(\vec{x}) = \sum_{i=1}^{N_{Trees}} \alpha_i \cdot T_i(\vec{x}). \tag{7.10}$$

Where α_i is the common re-weighting factor for each events in the i^{th} tree and $T(\vec{x})$ is the decision for that tree (see section E.3). Small values for $\text{y}_{\text{BDT}}(\vec{x})$ (values tending towards -1) indicate a background-like events, while large values (values tending towards 1) indicate signal-like events.

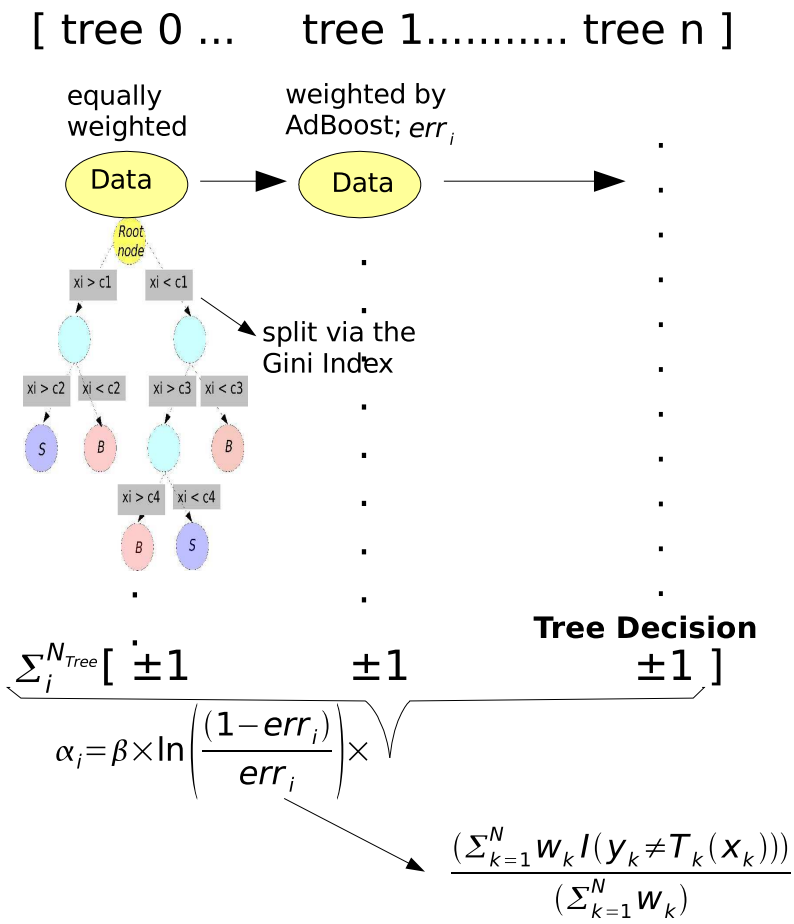


Figure 7.5: The idea of boosting. The classifier result for each decision tree T_i is taken as +1 if the event falls on a signal leaf and -1 if the event fall on a background leaf. The result from each decision tree is then boosted using the weighting algorithm α_i. Where the weighting algorithm makes use of the misclassification rate, err_{i-1}, of the previous tree.

7.3 Multivariate analysis software

The multivariate software package integrated within Root, called TMVA (Toolkit for Multivariate Data Analysis), has been developed to optimise event selection. TMVA is a toolkit which contains a large variety of multivariate classification algorithms,[3] ranging from cut optimisation algorithms, likelihood estimators, linear and non-linear neural networks to boosted decision trees [106]. TMVA manages simultaneously the training, testing and performance evaluation of all these classifiers. The toolkit consists of two independent phases: the training phase, where the multivariate classifiers are trained, tested and evaluated, and an application phase, where selected classifiers are applied to the concrete classification problem they have been trained for.

[3]For all optimisation studies discussed in this chapter the TMVA version 3.8.14 has been used.

The result of the training phase is the creation of weight files for each classifier which contains configuration options, controls and training weights. The trained classifiers are then applied to the test data, providing a scalar output upon which an event can be classified as either signal or background. The optimisation studies in this chapter have be carried out using the following set of TMVA classifiers:

- The Fisher.
- Likelihood and Decorrelated Likelihood.
- MultiLayer Perceptron (MLP) and Clermont-Ferrand (CF) Artifical Neural network.
- The Boosted Decision Tree.

A detailed description of the TMVA option used to train, test and run these classifiers is given in section E.5. The multivariate classifiers described above, were used to optimise signal purity after applying the selection cuts introduced in chapter 4.

7.4 Multivariate analysis studies

Fully reconstructed $B_s \rightarrow J/\psi\phi$ data is used for our signal sample, whilst the prompt-J/ψ sample is used as the only source of background. The prompt-J/ψ sample was chosen for the following reasons: the first is that for a lifetime unbiased selection as given in chapter 4, prompt-J/ψs were found to be the major background contributor. As discussed in chapter 5, the $\frac{B}{S}$ for prompt-J/ψs was found to be approximately 1.9. The second reason for using the prompt-J/ψ sample is statistics. In order for each classifier to have a responsable performance[4], they must be supplied with sufficient events for training and testing; few events will degrade the performance of each classifiers, apart from the Fisher. This was the reason for not using the $b\bar{b}$-inclusive data sample[5]. For each classifier and each variable set under investigation (see Figure 7.6), 10,000 signal and 400 background events were used during the training stage. The remaining number of events, $\sim$ 120,000 signal and $\sim$ 400 background events, were used for testing.

As the number of events contained in the generated signal and background data samples does not correspond to the number of events obtained in the same integrated time[6], both samples are required to be weighted. The integrated time $T_{S,B}$ (in seconds) for both signal and background is given by:

$$\begin{aligned} T_S &= \frac{N^S_{gen}}{2 \times \mathcal{L}^{int}_{\text{lumi}} \sigma(b\bar{b}) \text{BF}(\text{b} \rightarrow \text{B}_\text{s}) \text{BF}(\text{B}_\text{s} \rightarrow \text{J}/\psi\phi) \epsilon_\theta^{\text{signal}}} \times 10^7 \text{ s}, \\ T_B &= \frac{N^B_{gen}}{2 \times \mathcal{L}^{int}_{\text{lumi}} \sigma(b\bar{b}) \text{BF}(\text{J}/\psi \rightarrow \mu\mu) \epsilon_\theta^{\text{J}/\psi}} \times 10^7 \text{ s}. \end{aligned} \tag{7.11}$$

[4] This is the case for the ANN and Likelihood classifiers.
[5] The 32 background events lying within the wide B_s mass window, was consider too few statistics.
[6] This is an arbitrary time, which for these studies was taken to be a year.

The weighting factor is therefore given by:

$$\mathrm{W}_{S,B} \;=\; \frac{1}{T_{S,B}} \tag{7.12}$$

such that the weighting for signal is, $\mathrm{W}_S = 9.3 \times 10^{-3}\ s^{-1}$ and $\mathrm{W}_S = 2.23\ s^{-1}$ for the J/ψ background source.

The approach we adopted to optimise the signal purity after the canonical selection cuts with respect to the prompt J/ψ background is then:

Step 1 to select a full set of variables (Set 1) for each selected signal and background event, to enter the multivariate analysis stage. These are mostly kinematic variables, and are listed in section 7.4.1.

Step 2 to train and test each classifier on this full variable set. The list of classifiers used is: $\{\mathrm{Fisher},\ \mathrm{Likelihood},\ \mathrm{LikelihoodD},\ \mathrm{MLP},\ \mathrm{CF},\ \mathrm{BDT}\}$.

Step 3 to use the correlation matrices for signal and background to reduce Set 1 down to a set of variables which have correlation of $< 60\%$: this reduced set being called Set 2. This is to aid the performance of the Likelihood classifiers, which struggles with non-linear variable correlations. During this stage we will also make use of the BDT classifier; the zeroth decision tree[7] will also be used here to extract a BDT "inspired" variable set, denoted Set 3.

Step 4 to use sets 2 and 3 to find the optimal cut on each MVA classifier. The figure of merits calculated here include the $\frac{B}{S}$ ratio and the $\frac{S}{\sqrt{S+B}}$ ratio.[8]

Step 5 to use the optimal performing classifier to calculate the $\frac{B}{S}$ ratio, whilst retaining: $\{99\%,\ 90\%,\ 80\%,\ 70\%,\ 60\%,\ 50\%,\ 40\%,\ 30\%,\ 20\% \text{ and } 10\%\}$ of the signal.

The final step is to investigate the sensitivity to $-2\beta_s$ with respect to each signal rate given in point 6 above. The above selection procedure is summarised in Figure 7.6.

7.4.1 Discriminating variables

From the $B_s \to J/\psi\phi$ event information, we consider the full set of attributes defined within the laboratory frame. These are mostly kinematic parameters of the decay but also include detector effects such as the χ^2/dof of each final state track. A description of the variable later used in set 3 is defined below. A detailed description of each variable used in set 1 is given in section E.6, together with the signal and background distribution of each variable.

- **bs_e**: The energy associated with each B_s candidate.

[7] The zeroth decision tree is used since all events will initially have equal weightings

[8] This is often the criteria used when looking to optimise the branching ratio of a signal.

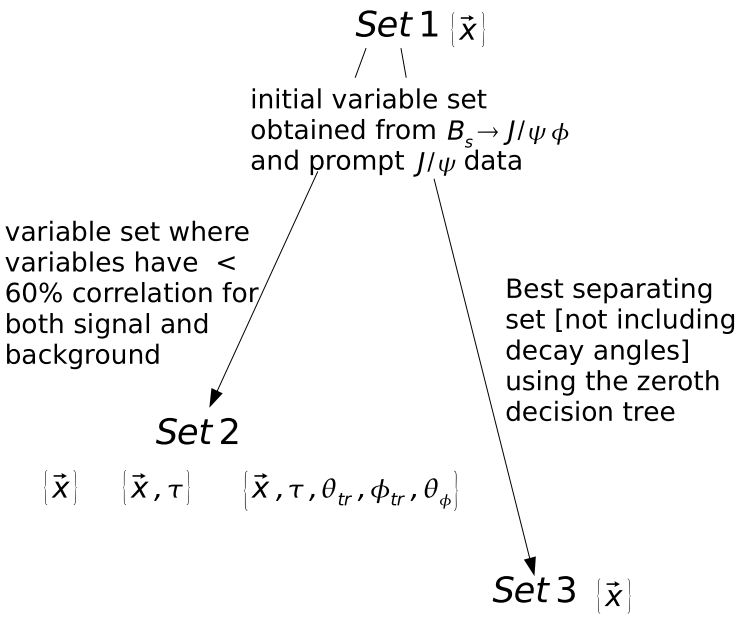


Figure 7.6: Schematic of the Multivariate analysis strategy. An initial set (Set 1) of $B_s \to J/\psi\phi$ and J/ψ kinematic variables is split into two further reduced sets. Variable Set 2 is obtained from Set 1 by selecting event variables which are less than 60% correlated for signal and background. Variable Set 3 is derived from Set 1 from the discriminating variables used by the zeroth decision tree. Variable set 3 is then used to determine the sensitivity of β_s using the prompt-J/ψ background sample.

- **bs_p**: The momentum associated with each B_s candidate.
- **bs_pt**: The transverse momentum associated with each B_s candidate.
- **bs_tau**: The propertime of each B_s candidate.
- **bs_tau_err**: The propertime error associated with each B_s.
- **bs_tau_sig**: The propertime significance of each B_s candidate.
- **bs_dist**: The distance for the B_s decay vertex to the primary vertex position [9]
- **bs_dist_sig**: The distance significance of the B_s decay vertex to the primary vertex position. given as: $\frac{bs\_dist}{\sigma(bs\_dist)}$
- **bs_ip**: The impact parameter of each B_s candidate with respect to the primary vertex [10].
- **bs_ip_sig**: This impact parameter significance of each candidate with respect to the primary vertex, given as: $\frac{bs\_ip}{\sigma(bs\_ip)}$

[9] This has been calculated using the standard DaVinci method, calcVertexDis, found in the Davinci GeomDispCalculator class. This method calculates the absolute distance and errors between the two vertices.

[10] This is again calculated using the DaVinci GeomDispCalculator, using the calcImpactPar method. This returns the distance of closest approach and errors between the given particle and vertex.

7.4.2 Multivariate study: step one

From the full set of attributes, each classifier is trained and tested over the signal and background samples. The performance of each classifier is recorded and the correlation matrices, shown in Figures 7.2 and 7.3, used to further reduce the input parameter set.

Classifier Performance: The performance of each classifier is specified by three figures of merit, which include:

$\mathrm{Y_{rej}}$**:** the signal efficiencies at 90% background rejection,

$\mathrm{Y_{sep}}$**:** the separation of a classifier, and

$\mathrm{Y_{sig}}$**:** the significance of a classifier,

$\mathrm{Y_{rej}}$ effectively measuring the area under the background rejection versus signal purity plot. The larger the area the better the performance of the classifier is. The separation of a classifier measures the classifiers ability to separate signal and background, and is defined as:

$$\mathrm{Y_{sep}} = \frac{1}{2}\int_{-\infty}^{\infty} \frac{(\hat{\mathrm{y}}_{\mathrm{S}}(\mathrm{y}) - \hat{\mathrm{y}}_{\mathrm{B}}(\mathrm{y}))^2}{\hat{\mathrm{y}}_{\mathrm{S}}(\mathrm{y}) + \hat{\mathrm{y}}_{\mathrm{B}}(\mathrm{y})}\mathrm{dy}, \tag{7.13}$$

where $\hat{y}_S$ and $\hat{y}_B$ are the signal and background p.d.fs of y respectively. From equation (7.13), it is clear that $\mathrm{Y_{sep}}$ is bounded between 0 and 1, with 0 indicating identical signal and background shapes (no separation) and 1 indicating signal and background shapes with no overlap (perfect separation). The significance of a classifier however is unbounded and is defined by [107]:

$$\mathrm{Y_{sig}} = \frac{|\langle \mathrm{y_S}\rangle - \langle \mathrm{y_B}\rangle|}{\sqrt{\sigma^2(\mathrm{y_S}) + \sigma^2(\mathrm{y_B})}}, \tag{7.14}$$

where $\langle y_{S/B}\rangle$ is the classifiers mean and $\sigma(y_{S/B})$ is the root-mean-squared (rms) for signal (S) and background (B) respectively. From the classifiers performance profile summarised in Table 7.1 and Figure 7.7, it is evident that the BDT gives the optimal signal to background separation trained over the variables in Set 1. This is also evident in Figure 7.8, where the normalised classifier distribution on the left shows clear separation between signal (blue) and background (black).

Correlations: For the attributes defined in Set 1 the corresponding correlation matrices for signal and background is shown in Figure 7.2 and 7.3 respectively.

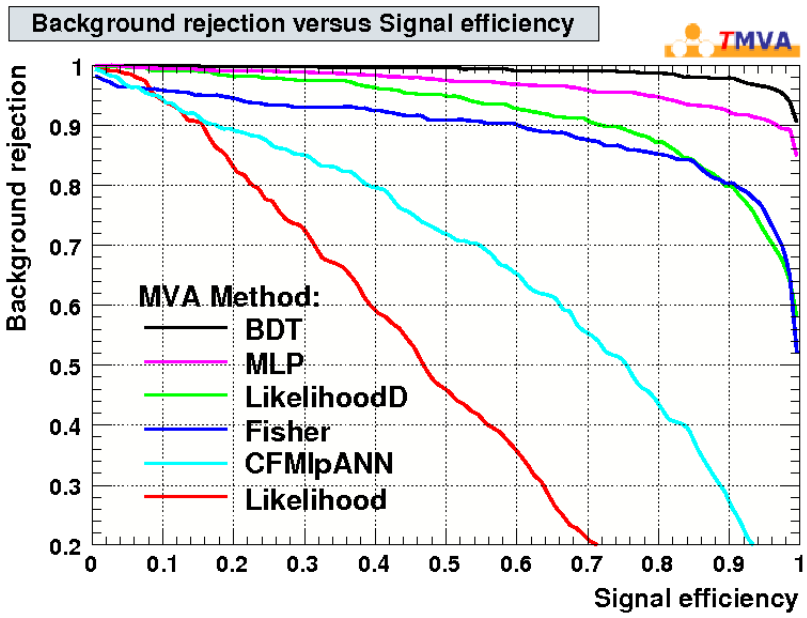


Figure 7.7: Background rejection versus signal efficiencies plots using the full variable set.

Classifier performance using Set 1	
Classifier	$(\mathrm{Y_{sig}}, \mathrm{Y_{sep}}, \mathrm{Y_{rej}})$
Likelihood	(0.64, 0.25, 0.16)
LikelihoodD	(0.29, 0.16, 0.71)
Fisher	(0.76, 0.32, 0.60)
MLP	(0.2, 0.98, 0.98)
CF	(0.43, 0.1, 0.16)
BDT	(2.45, 0.88, 1.00)

Table 7.1: The performance of each classifier to the attributes in Set 1 $\{\vec{x}, \tau_{B_s}, \theta_{tr}, \theta_{\phi}, \phi_{tr}\}$. The performance is characterised the significance ($\mathrm{Y_{sig}}$), separation ($\mathrm{Y_{sep}}$) and the signal efficiency at 90% background rejection ($\mathrm{Y_{rej}}$). Each classifiers performance is summerised in the parethises as: $(\mathrm{Y_{sig}}, \mathrm{Y_{sep}}, \mathrm{Y_{rej}})$.

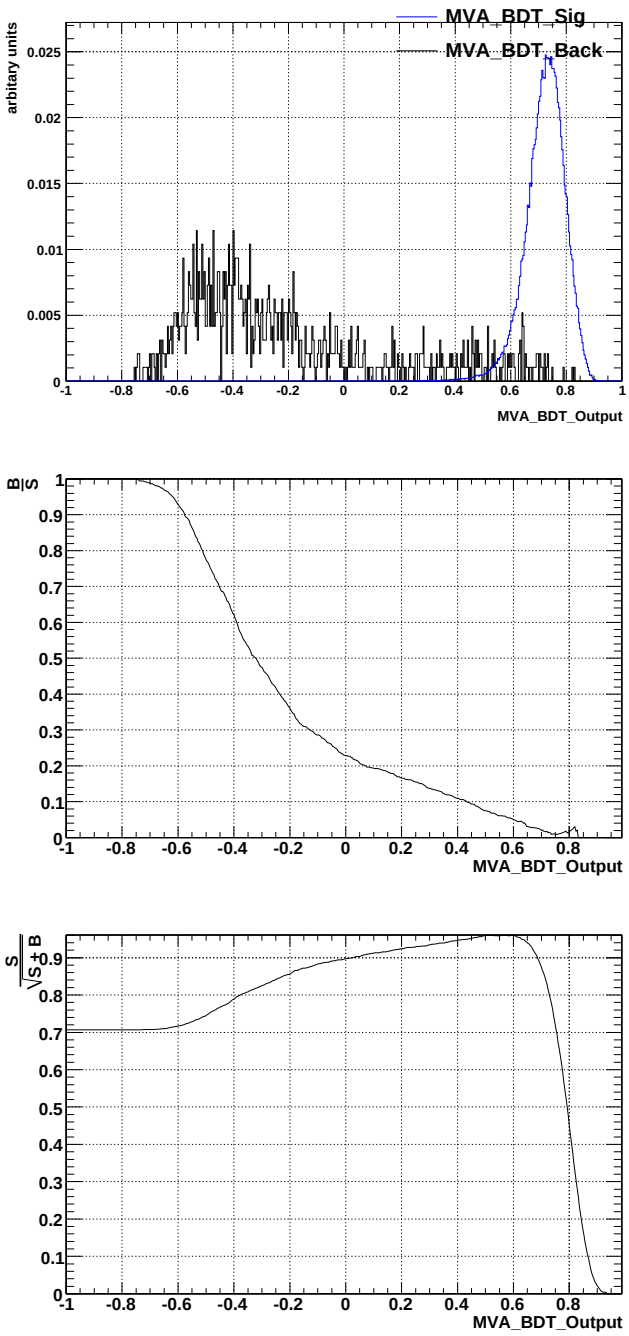


Figure 7.8: Normalised response of signal (blue) and background (black) to the BDT on the full data set: (Top) the BDT classifier response for signal and background (Middle), $\frac{B}{S}$ response of the BDT and (Bottom) $\frac{S}{\sqrt{S+B}}$.

	bs_e	bs_p	bs_pt	bs_tau	bs_tau_err	bs_ip	bs_dist	bs_tau_sig	bs_ip_sig	bs_dist_sig	theta	psi	phi	phi_mass	phi_e	phi_pt	jpsi_mass	jpsi_e	jpsi_pt	kp_p	kp_dll	kp_Tchi2	km_p	km_dll	km_Tchi2	mup_p	mup_dll	mup_Tchi2	mum_p	mum_dll	mum_Tchi2
mum_Tchi2	6	6	-3	1	1	-1	3	0	-1	0	0	0	0	0	4	-1	0	5	-3	4	-3	5	4	-3	3	2	-1	3	6	-3	100
mum_dll	18	18	-8	0	4	3	-9	-1	1	-1	1	-1	0	0	-14	4	-1	18	-8	13	4	-3	13	4	-4	-5	3	-2	21	100	-3
mum_p	71	71	37	-2	-19	-6	32	4	-2	5	-1	1	0	0	41	13	3	77	42	39	-9	8	40	-9	6	20	-9	2	100	-22	6
mup_Tchi2	5	5	-1	0	2	0	2	0	0	0	0	1	-1	0	4	0	0	5	-1	4	-4	2	4	-4	2	5	-3	100	2	-2	3
mup_dll	14	14	-9	-1	1	4	-8	-2	2	-2	0	-5	0	0	-16	-8	0	-10	-8	-16	30	-10	-15	30	-7	-7	100	-3	-9	3	-1
mup_p	71	71	37	-1	-18	-5	33	4	-1	5	0	0	0	0	40	12	-17	7	42	39	-8	53	8	-8	7	100	-7	5	20	-5	2
km_Tchi2	12	12	1	0	-1	0	6	0	0	0	0	-2	0	0	14	0	1	9	2	14	-10	7	13	-1	100	7	-7	2	6	-4	3
km_dll	-15	-15	10	0	3	3	-7	0	1	-1	0	4	-1	0	-16	-8	0	-11	-9	-15	100	-7	-16	100	-10	-8	30	-4	-9	4	-3
km_p	79	79	29	-1	-28	-3	37	7	1	8	0	19	-2	1	96	47	1	51	12	87	-16	11	100	-16	13	38	-15	4	40	-13	4
kp_Tchi2	11	11	0	0	-1	0	5	1	0	0	-1	1	-1	-1	11	-1	1	8	1	10	-7	100	11	-7	7	5	-10	2	8	-3	5
kp_dll	-15	-15	-10	0	3	3	-7	0	1	-1	0	4	-1	0	-16	-8	0	-11	-9	-15	100	-7	-16	100	-10	-8	30	-4	-9	4	-3
kp_p	79	79	30	-1	-28	-3	37	7	1	9	0	-15	0	1	96	48	1	51	12	100	-15	10	87	-15	14	39	-16	4	39	-13	4
jpsi_pt	43	43	91	-3	-47	-13	17	11	0	13	-2	0	0	0	12	39	0	54	100	12	-9	1	12	-9	2	42	-8	-1	42	-8	-3
jpsi_e	92	92	48	-2	-24	-7	43	5	-2	7	-1	0	0	0	52	16	0	100	54	51	-11	18	51	-11	9	77	-10	5	77	-18	5
jpsi_mass	1	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	100	1	0	1	0	1	1	0	1	-1	0	0	3	-1	0
phi_pt	33	33	72	-2	-39	-8	13	10	5	12	-1	0	0	2	49	100	0	16	39	48	-8	-1	47	-8	0	12	-8	0	13	-4	-1
phi_e	81	81	30	-1	-29	-3	38	7	1	9	0	2	-1	0	100	49	1	52	12	96	-16	11	96	-16	14	40	-16	4	41	-14	4
phi_mass	0	0	-1	0	-1	0	0	0	-1	0	0	0	0	100	1	2	0	0	0	1	0	-1	1	0	0	0	0	0	0	0	0
phi	0	0	0	-2	0	1	-1	-1	1	-1	0	-6	100	0	-1	0	1	0	0	0	-1	-1	-2	-1	0	0	0	-1	0	0	0
psi	1	1	0	0	-1	0	1	0	0	0	0	100	-6	0	2	0	1	0	0	-15	4	1	19	4	-2	0	-5	1	1	-1	0
theta	0	0	-2	0	1	1	0	-1	1	-1	100	0	0	0	0	-1	0	-1	-2	0	0	-1	0	0	0	0	0	0	-1	1	0
bs_dist_sig	9	9	15	89	-12	-1	76	98	-1	100	-1	0	-1	0	9	12	0	7	13	9	-1	0	8	-1	0	5	-2	0	5	-1	0
bs_ip_sig	-1	-1	1	-1	4	90	-1	-1	100	-1	1	0	1	-1	1	5	0	-2	0	1	1	0	1	1	0	-1	2	0	-2	1	-1
bs_tau_sig	7	7	13	92	-11	-1	75	100	-1	98	-1	0	-1	0	7	10	0	5	11	7	0	1	7	0	0	4	-2	0	4	-1	0
bs_dist	46	46	19	71	-3	1	100	75	-1	76	0	1	-1	0	38	13	0	43	17	37	-7	5	37	-7	6	33	-8	2	32	-9	3
bs_ip	-6	-6	-13	4	24	100	0	-1	90	-1	1	0	1	0	-3	-8	0	-7	-13	-3	3	0	-3	3	0	-5	4	0	-6	3	-1
bs_tau_err	-30	-30	-52	15	100	24	-3	-11	4	-12	1	-1	0	-1	-29	-39	0	-24	-47	-28	3	-1	-28	3	-1	-18	1	2	-19	4	1
bs_tau	-2	-2	-3	100	15	4	71	92	-1	89	0	0	-2	0	-1	-2	0	-2	-3	-1	0	0	-1	0	0	-1	-1	0	-2	0	1
bs_pt	46	46	100	-3	-52	-13	19	13	1	15	-2	0	0	1	30	72	0	48	91	30	-10	0	29	-10	1	37	-9	-1	37	-8	-3
bs_p	99	100	46	-2	-30	-6	46	7	-1	9	0	1	0	0	81	33	1	92	43	79	-15	11	79	-15	12	71	-14	5	71	-18	6
bs_e	100	99	46	-2	-30	-6	46	7	-1	9	0	1	0	0	81	33	1	92	43	79	-15	11	79	-15	12	71	-14	5	71	-18	6

Table 7.2: Correlation matrices for the signal variables found in Set 1.

	bs_e	bs_p	bs_pt	bs_tau	bs_tau_err	bs_ip	bs_dist	bs_tau_sig	bs_ip_sig	bs_dist_sig	theta	psi	phi	phi_mass	phi_e	phi_pt	jpsi_mass	jpsi_e	jpsi_pt	kp_p	kp_dll	kp_Tchi2	km_p	km_dll	km_Tchi2	mup_p	mup_dll	mup_Tchi2	mum_p	mum_dll	mum_Tchi2
mum_Tchi2	0	14	5	-3	-1	2	-2	-1	0	-1	0	0	-6	6	2	7	-6	-1	3	4	-1	10	4	-1	-3	11	-10	0	19	-10	100
mum_dll	1	-19	0	3	3	1	0	-2	-7	-1	-4	0	8	-11	-1	-1	2	3	-1	-13	7	-4	-13	7	-2	-2	4	0	-28	100	-10
mum_p	0	66	39	23	-5	12	37	0	2	-1	-2	2	-7	7	1	21	3	0	42	39	-9	7	39	-9	6	16	-9	-2	100	-28	19
mup_Tchi2	0	3	-3	0	1	0	2	1	1	1	-2	-1	-3	0	-1	-2	4	0	-3	1	0	-2	1	0	4	12	-5	100	-2	0	0
mup_dll	0	-12	-1	11	4	-4	0	2	8	0	4	-7	8	-5	3	-4	1	-1	0	-11	29	-7	-12	29	-17	-6	100	-5	-9	4	-10
mup_p	-2	67	31	10	-8	9	25	4	0	4	-4	-7	0	1	1	19	0	-5	30	45	-1	2	44	-1	14	100	-6	12	16	-2	11
km_Tchi2	0	17	9	-2	-2	1	3	3	-2	3	0	-3	-15	-3	0	13	1	-1	9	17	-21	8	15	-21	100	14	-17	4	6	-2	3
km_dll	0	-13	-2	5	-1	-12	-2	-6	9	-5	-10	0	4	-1	0	-4	1	0	-1	-14	100	-10	-16	100	-21	-1	29	0	-9	7	-1
km_p	6	86	29	16	-10	27	38	9	-2	8	-3	13	-12	-1	8	48	-7	3	8	91	-16	8	100	-16	15	44	-12	1	39	-13	4
kp_Tchi2	1	8	3	0	-4	15	4	0	-2	0	-9	10	-1	1	4	2	-2	0	2	6	-10	100	8	-10	8	2	-7	-2	7	-4	10
kp_dll	0	-13	-2	5	-1	-12	-2	-6	9	-5	-10	0	4	-1	0	-4	1	0	-1	-14	100	-10	-16	100	-21	-1	29	0	-9	7	-1
kp_p	5	87	37	23	-10	28	48	11	-2	11	-2	-14	-15	0	7	56	-1	2	15	100	-14	6	91	-14	17	45	-11	1	39	-13	4
jpsi_pt	-2	34	89	36	-6	-1	37	3	-1	4	0	-14	-4	0	-2	41	14	-2	100	15	-1	2	8	-1	9	30	0	-3	42	-1	3
jpsi_e	91	0	0	0	-31	2	1	12	-5	14	3	2	0	-3	53	2	-15	100	-2	2	0	0	3	0	-1	-5	-1	0	0	3	-1
jpsi_mass	-13	-1	12	7	3	0	10	-6	4	-5	1	-3	-5	3	-6	4	100	-15	14	-1	1	-2	-7	1	1	0	1	4	3	2	-6
phi_pt	4	45	74	26	-10	10	38	11	-2	11	1	-8	-6	-4	5	100	4	2	41	56	-4	2	48	-4	13	19	-4	-2	21	-1	7
phi_e	82	5	2	-1	-32	-4	0	8	-3	9	1	2	-1	-4	100	5	-6	53	-2	7	0	4	8	0	0	1	3	-1	1	-1	2
phi_mass	-4	3	0	1	1	0	5	2	7	3	-4	-8	-9	100	-4	-4	3	-3	0	0	-1	1	-1	-1	-3	1	-5	0	7	-11	6
phi	0	-11	-5	1	-1	-1	0	-5	-4	-5	0	1	100	-9	-1	-6	-5	0	-4	-15	4	-1	-12	4	-15	0	8	-3	-7	8	-6
psi	2	-2	-14	-4	-3	0	-4	-5	-3	-6	-3	100	1	-8	2	-8	-3	2	-14	-14	0	10	13	0	-3	-7	-7	-1	2	0	0
theta	2	-4	0	1	5	1	0	8	0	7	100	-3	0	-4	1	1	1	3	0	-2	-10	-9	-3	-10	0	-4	4	-2	-2	-4	0
bs_dist_sig	14	6	6	9	-14	8	12	98	-6	100	7	-6	-5	3	9	11	-5	14	4	11	-5	0	8	-5	3	4	0	1	-1	-1	-1
bs_tau_sig	-5	-1	-3	-5	7	0	-5	-6	100	-6	0	-3	-4	7	-3	-2	4	-5	-1	-2	9	-2	-2	9	-2	0	8	1	2	-7	0
bs_ip_sig	12	7	6	8	-13	6	11	100	-6	98	8	-5	-5	2	8	11	-6	12	3	11	-6	0	9	-6	3	4	2	1	0	-2	-1
bs_dist	0	48	44	82	-2	43	100	11	-5	12	0	-4	0	5	0	38	10	1	37	48	-2	4	38	-2	3	25	0	2	37	0	-2
bs_ip	0	24	2	35	-1	100	43	6	0	8	1	0	-1	0	-4	10	0	2	-1	28	-12	15	27	-12	1	9	-4	0	12	1	2
bs_tau_err	-35	-10	-9	2	100	-1	-2	-13	7	-14	5	-3	-1	1	-32	-10	3	-31	-6	-10	-1	-4	-10	-1	-2	-8	4	1	-5	3	-1
bs_tau	0	24	39	100	2	35	82	8	-5	9	1	-4	1	1	-1	26	7	0	36	23	5	0	16	5	-2	10	11	0	23	3	-3
bs_pt	1	45	100	39	-9	2	44	6	-3	6	0	-14	-5	0	2	74	12	0	89	37	-2	3	29	-2	9	31	-1	-3	39	0	5
bs_p	2	100	45	24	-10	24	48	7	-1	6	-4	-2	-11	3	5	45	-1	0	34	87	-13	8	86	-13	17	67	-12	3	66	-19	14
bs_e	100	2	1	0	-35	0	0	12	-5	14	2	2	0	-4	82	4	-13	91	-2	5	0	1	6	0	0	-2	0	0	0	1	0

Table 7.3: Correlation matrices for the background variables found in Set 1

7.4.3 Multivariate study: step two

After training each classifier over Set 1, we next remove using the signal and background correlation matrices (Figures 7.2 and 7.3), those variables with greater than 60% correlation. This reduces the set of variables from 31 to 11, as shown in Table 7.4. Although the **bs_tau** variable is highly correlated in both signal (**bs_tau_sig** and **bs_dist_sig**) and background (**bs_dist**) distributions, it has also been included into this set. The contents of this reduced set, now denoted as Set 2, is given in Table 7.4.

bs_tau	**bs_tau_err**	**km_Tchi2**	**phi**
phi_mass	**mup_dll**	**kp_Tchi2**	**psi**
jpsi_mass	**bs_ip**	**theta**	

Table 7.4: Variable Set 2. Each variable pertains from Set 1 with the requirement that its correlation, with other variables in this set, is less than 60%, for both signal and background correlation matrices. Also included in Set 2 is **bs_tau**. The list of all variable names can be obtained from section E.6.

Each classifier's performance was then tested under Set 2 and also two variants of this set: a set excluding the decay angles and a set excluding both the decay angles and propertime. Both the propertime (using an IP significance cut) and decay angles were found to have an acceptance effect in chapter 4. By including these variables in the multivariate fit, it is likely that we will retain these acceptance effects. It is necessary therefore to investigate each classifier's performance over the three sets: the propertime and decay angular acceptance effects using Set 2 for the case of the BDT is shown in section 7.5. The performance of each classifier is given in Table 7.5 and Figure 7.9. It is clear from this that the BDT classifier again performs optimally, under all three sets to distinguish signal from background data. The separation given by the BDT is visible in Figure E.21, for the normalised classifier response using Set 2.

Multivariate classifier can be over trained if they are given too many adjustable parameters for too few events. This is one of the reasons why events are separated into statistically independent training and testing samples; a classifier with many parameters, for example a decision tree, can "over-adapt" to the training sample, and show a very high efficiency. On an independent dataset, the efficiency would however be very different. A check for over training was carried out on each classifier and is summarised in Table 7.6. The check for over-training is performed using the Kolmogorov-Smirnov (KS) test [108], which searches for the maximum vertical deviation between test and training Cumulative Distribution Functions (CDFs) for signal and background respectively. The values quoted in Table 7.6 are the Kolmogorov-Smirnov Confidence Levels (KSCL). In the case of no over training the KSCL is uniformly distributed between 0 and 1, and on average is 0.5. If there is any significant over training, the training and test CDFs differ significantly, and the KSCL values tend to be very small, e.g. 1.0×10^{-5} or worse [106]. From

Table 7.6 it is clear that of all the classifiers, the CF classifier suffers the most from over training, whilst the Fisher Linear discriminate remains relatively unaffected. The BDT does not show any significant signs of over-training.

Classifier performance with Set 2			
Classifier	$(\mathrm{Y_{sig}}, \mathrm{Y_{sep}}, \mathrm{Y_{rej}})$		
	$\{\vec{x}, \tau_{B_s}, \theta_{tr}, \theta_\phi, \phi_{tr}\}$	$\{\vec{x}, \tau_{B_s}\}$	$\{\vec{x}\}$
Likelihood	(1.05, 0.45, 0.65)	(0.68, 0.26, 0.30)	(0.54, 0.23, 0.20)
LikelihoodD	(0.56, 0.27, 0.59)	(0.45, 0.17, 0.24)	(0.05, 0.19, 0.30)
Fisher	(0.76, 0.32, 0.45)	(0.22, 0.12, 0.09)	(0.21, 0.14, 0.05)
MLP	(1.25, 0.02, 0.70)	(0.68, 0.09, 0.29)	(0.08, 0.26, 0.28)
CF	(0.25, 0.012,0.29)	(0.68, 0.086, 0.02)	(0.08, 0.006,0.04)
BDT	(1.64, 0.71, 0.98)	(0.75, 0.29, 0.39)	(0.67, 0.24, 0.20)

Table 7.5: The performance $(\mathrm{Y_{sig}}, \mathrm{Y_{sep}}, \mathrm{Y_{rej}})$ of each classifier to the attributes in Set 2. Where the first column includes propertime τ_{B_s} and decay angles $\theta_{tr}, \theta_\phi, \phi_{tr}$. The second column excludes the decay angles and the third column excludes both propertime and decay angles.

Classifier over-training check with Set 2			
Classifier	signal (background) KS CL [0,1]		
	$\{\vec{x}, \tau_{B_s}, \theta_{tr}, \theta_\phi, \phi_{tr}\}$	$\{\vec{x}, \tau_{B_s}\}$	$\{\vec{x}\}$
Likelihood	0.997 (0.701)	0.858 (0.531)	0.830 (0.474)
LikelihoodD	0.962 (0.585)	0.987 (0.664)	0.874 (0.945)
Fisher	0.284 (0.934)	0.458 (0.328)	0.366 (0.349)
MLP	0.974 (0.585)	0.652 (0.992)	0.859 (0.833)
CF	0.159 (6.8×10^{-7})	0.449 (1.6×10^{-4})	0.410 (4.6×10^{-19})
BDT	0.620 (0.116)	0.901 (0.202)	0.369 (0.367)

Table 7.6: Over training check on signal (background) events for each classifier to the attributes in Set 2. Where the first column includes propertime τ_{B_s} and decay angles $\theta_{tr}, \theta_\phi, \phi_{tr}$. The second column excludes the decay angles and the third column excludes both propertime and decay angles. The values quoted are the Kolmogorov-Smirnov Confidence Levels (KSCL), which lie in the range [0,1].

7.5 Acceptance effects

The $B_s \to J/\psi\phi$ selection cuts (Table 4.5) and the B_s propertime significance cut have been shown, in Figures 6.6 and B.5 respectively, to cause an acceptance effect on the decay angles and B_s propertime. These effects were investigated using MV techniques by running the trained BDT over the signal and background data using Set 2. The acceptance distribution for retaining 90% and 98% of the signal is described in Figures 7.11 and 7.12, where the plots in Figure 7.11 show the propertime acceptance, while the θ_{tr}, ϕ_{tr} and θ_ϕ acceptance effects respectively, are shown in Figure 7.12. It is evident from these plots if we were to retain 98% of the signal, we

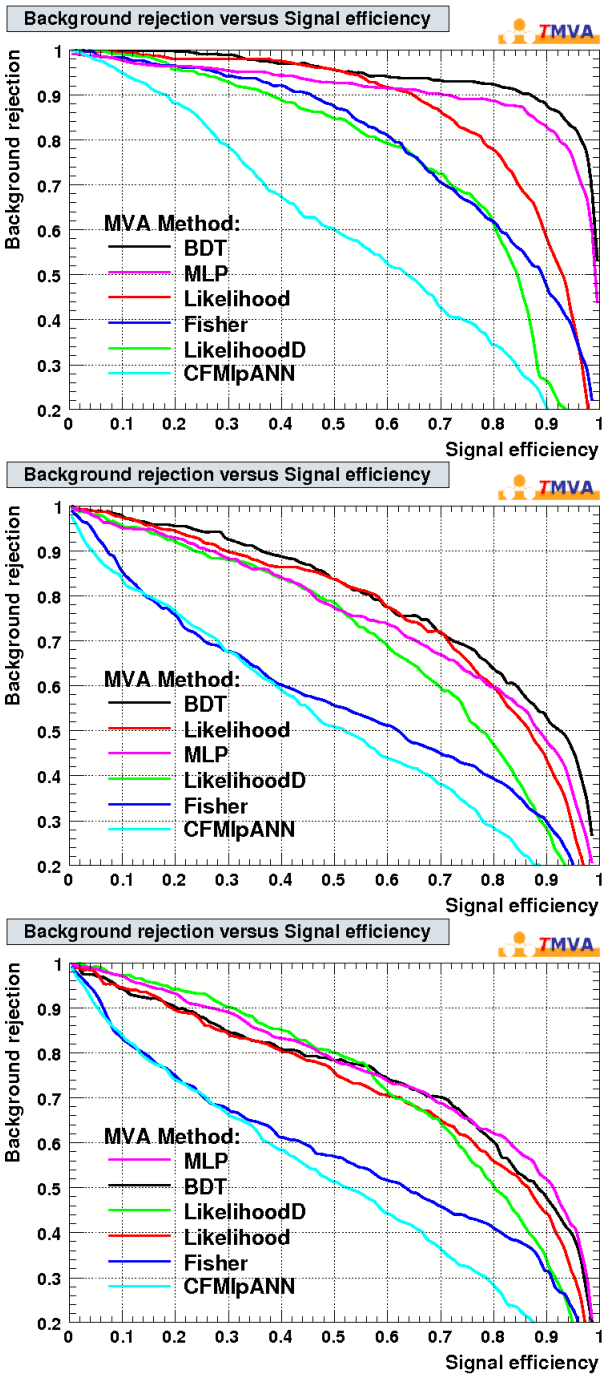


Figure 7.9: Background rejection verse signal efficiencies plots for Set 2, including: (top) propertime and the three decay angles; (Middle) neglects the propertime and (bottom) neglects propertime and decay angles.

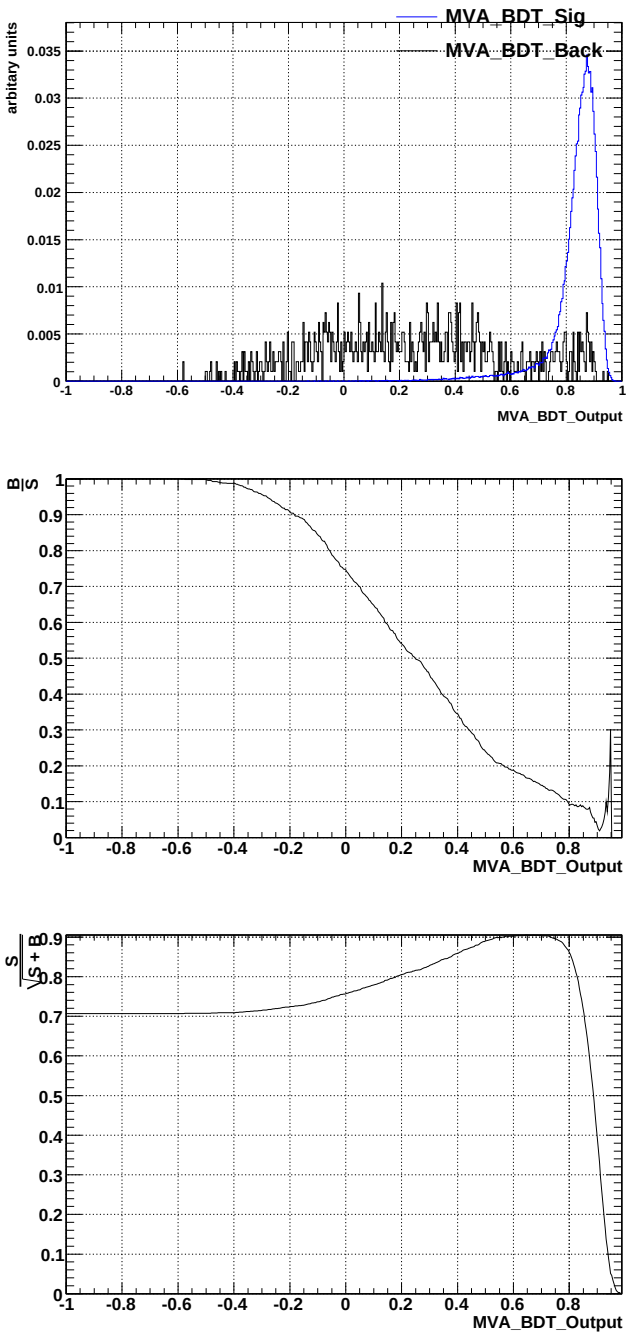


Figure 7.10: Normalised response of signal (blue) and background (black) to the classifiers on the reduced data set, including propertime and decay angles: (Left) the classifiers response for signal and background (Middle), the $\frac{B}{S}$ response of the classifier and (Right) the $\frac{S}{\sqrt{S+B}}$.

would not incure acceptance effect on either the propertime or angles. The implications of this is that we can use the highly separated B_s propertime distribution (see Figure E.16) within a final multivariate selection cut on the J/ψ background sample. The fact that a simple selection based on the BDT classifier may include the B_s propertime distribution, without incurring an acceptance effect, is potentially useful for the HLT. The HLT2 contains specific C++ algorithms for inclusive and exclusive event selection, that can include any combination of cuts. The only requirement being that these algorithms accept events below a specified bandwidth. It might be possible to build a BDT inspired HLT selection here, to retain $\sim$99% of the signal and reject a large fraction of the background.

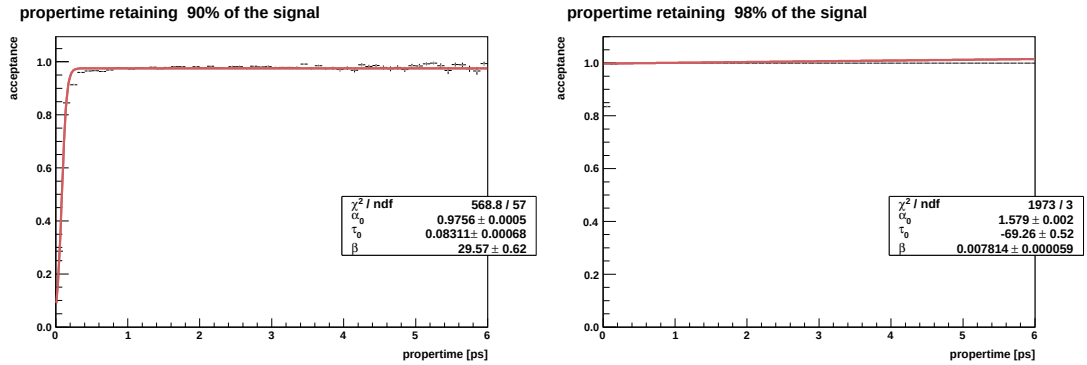


Figure 7.11: Acceptance effect on propertime using variable Set 2 and the BDT classifier. Retaining (Left) 90% and (Right) 98% of the the signal.

7.6 BDT inspired variable set

The BDT has been shown to perform optimally over variable Sets 1 and 2. We therefore make use of this classifier for the sensitivity studies in section 7.7. To obtain a variable set for these studies (Set 3), we start from Set 1 and make use of the zeroth decision tree (where all events are weighted equally) as a guide. This is shown in Figure 7.13. It is evident from this figure that the most discriminating variables in the event are those related to decay vertex of the B_s, including the variables: **bs_tau**, **bs_tau_sig**, **bs_dist_sig**. This is understandable given the data samples under investigation. The prompt J/ψ background occurs primarily around the interaction point, while the signal events have a significant displacement from it. Inspired by the zeroth decision tree, a third set of variable was generated as given in Table 7.7.

The performance of the BDT to the attributes in Set 3 is: 3.04, 0.97 and 1.00, for $\mathrm{Y_{sig}}$, $\mathrm{Y_{sep}}$, and Y_{rej} respectively, with the background to signal purity given in Figure 7.14. It is also evident from the top plot in Figure 7.15, that the BDT running over Set 3 gives clear separation between signal and background. It is also clear from the bottom plot in Figure 7.15, that the maximum $\frac{S}{\sqrt{S+B}}$ occurs at a classifier cut of 0.6. An over training test was also preformed on the classifiers

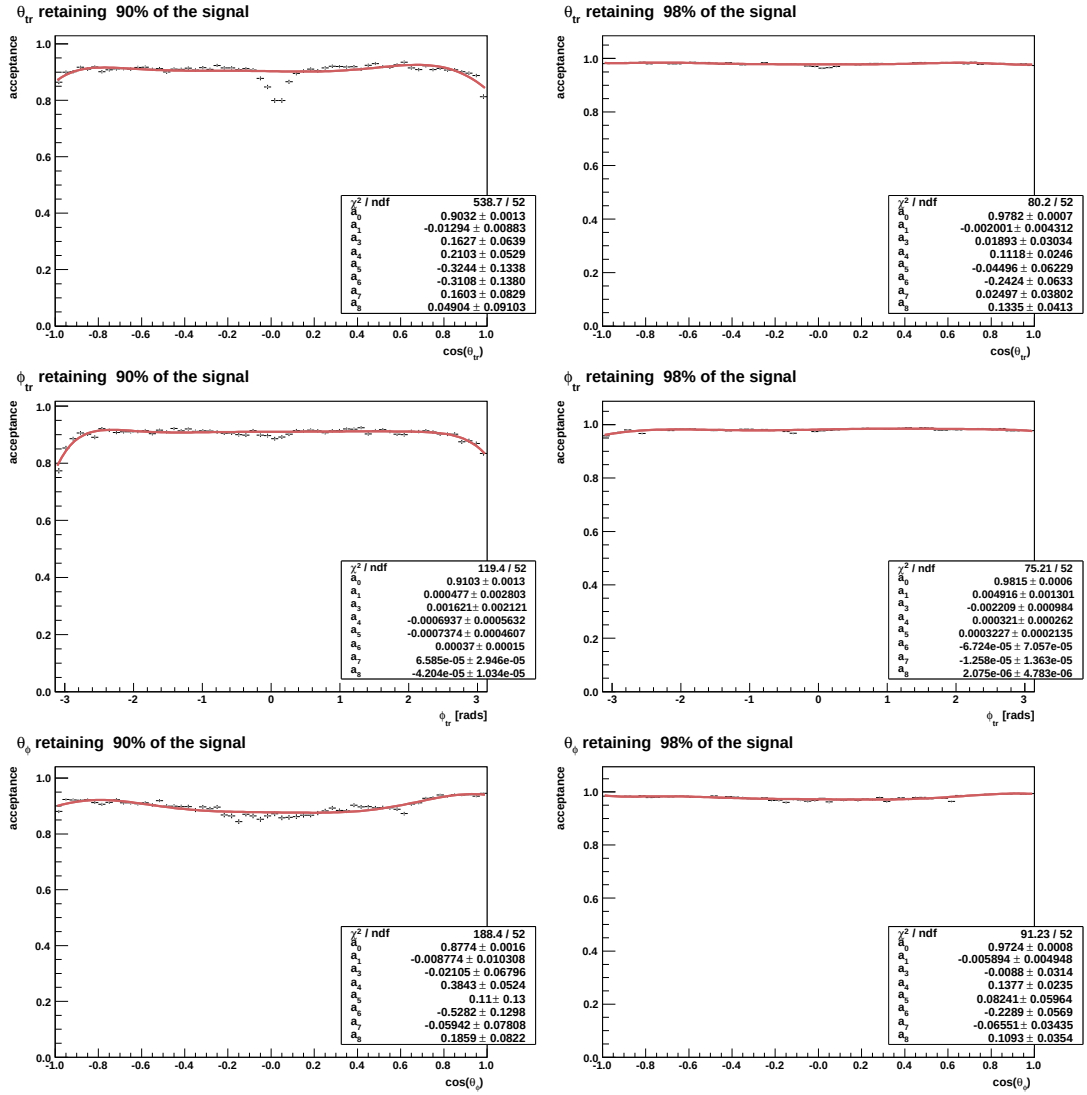


Figure 7.12: Acceptance effect on the decay angles using variable Set 2 and the BDT classifier. (Top) the θ_{tr} acceptance effect, retaining (Left) 90% and (Right) 98% of the the signal.

over variable set 3, and are illustrated in Figure E.22. Running over Set 3, the BDT does show signs of over training, with the Kolmogorov-Smirnov Confidence Level of 0.049 for signal and 0.00417 for background.

bs_e	**bs_p**	**bs_pt**	**bs_tau**
bs_tau_err	**bs_ip**	**bs_dist**	**bs_tau_sig**
bs_ip_sig	**bs_dist_sig**		

Table 7.7: Variable Set 3. Each variable from Set 1 is selected by studying the TMVA's output for the zeroth decision tree.

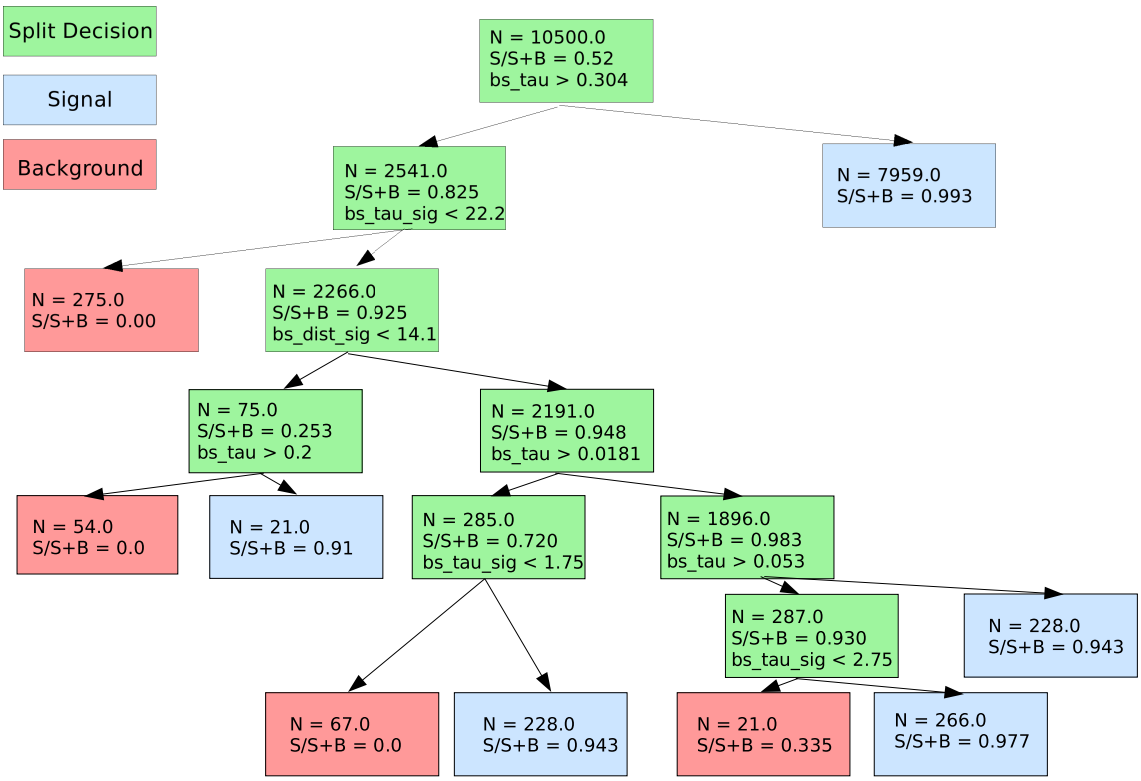


Figure 7.13: The response, taken from TMVA, after the training stage of the zeroth decision tree using as input the variable Set 1.

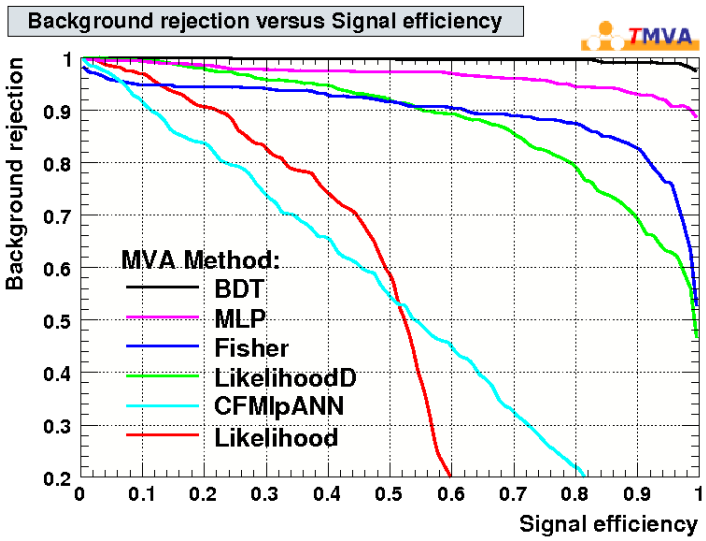


Figure 7.14: Background rejection verse signal efficiencies plot for each classifier running on Set 3.

7.7 Sensitivity to $-2\beta_s$ using the BDT

In this final section, we summarise the $\frac{B}{S}$ ratio using the specific J/ψ background sample obtained with the BDT and the sensitivity to $-2\beta_s$ with respect to the BDT output response to data Set 3. For sets 2 and 3, the $\frac{B}{S}$ fraction has been calculated[11] for different signal retentions, going from 5% up to 100%.

% of signal events retained	$\left(\frac{B}{S}\right)^{J/\psi}$			
	Set 2 $\{\vec{x}\}$	Set 2 $\{\vec{x},\tau\}$	Set 2 $\{\vec{x},\tau,\theta_{tr},\phi_{tr},\theta_{\phi}\}$	Set 3
5	1.54 ± 0.29	0.25 ± 0.12	0.19 ± 0.10	0.0
10	1.17 ± 0.18	0.17 ± 0.07	0.12 ± 0.06	0.0
20	1.08 ± 0.12	0.25 ± 0.06	0.22 ± 0.06	0.0
30	0.88 ± 0.09	0.23 ± 0.05	0.19 ± 0.04	0.0
40	0.99 ± 0.08	0.23 ± 0.04	0.19 ± 0.04	0.0
50	0.98 ± 0.07	0.19 ± 0.03	0.20 ± 0.03	0.0
60	0.98 ± 0.07	0.22 ± 0.03	0.22 ± 0.03	$4.2 \pm 4.4 \times 10^{-3}$
70	1.02 ± 0.06	0.24 ± 0.03	0.21 ± 0.03	$2.5 \pm 9.9 \times 10^{-3}$
80	1.09 ± 0.06	0.25 ± 0.03	0.24 ± 0.03	0.04 ± 0.012
90	1.18 ± 0.06	0.31 ± 0.03	0.30 ± 0.03	0.24 ± 0.027
95	1.29 ± 0.06	0.39 ± 0.04	0.41 ± 0.03	0.67 ± 0.044
98	1.51 ± 0.06	0.61 ± 0.03	0.64 ± 0.04	1.33 ± 0.06
99	1.64 ± 0.07	0.82 ± 0.05	0.84 ± 0.05	1.60 ± 0.07
100	1.90 ± 0.07	1.90 ± 0.07	1.90 ± 0.07	1.90 ± 0.07

Table 7.8: The $\frac{B}{S}$ numbers for parameter Sets 2 and 3, using the BDT response, retaining 5% up to 100% of the signal. For Set 2, the ratios excluding the propertime (τ) and the propertime and decay angles ($\tau,\ \theta_{tr},\ \phi_{tr},\ \theta_{\phi}$) has also been calculated.

From Table 7.8 and Figures E.21 and 7.15 it is clear as we increase the percentage of signal retained (going from right-to-left in the figures), we also increase the background contribution retained and thus increase the $\frac{B}{S}$ ratio. The maximum $\frac{B}{S}$ ratio is obtained, as expected, when we keep 100% of the signal. It is also clear from Table 7.8 and Figure 7.15 that Set 3 gives the best signal to background separation. It is also worth commenting that the negligible $\frac{B}{S}$ ratio below 60% signal remaining with Set 3, is due to a large B_s propertime cut ($\sim$0.3 ps), which wipes out huge portion of the prompt sample. We now use the $\frac{B}{S}$ numbers for this third set to investigate sensitivity to $-2\beta_s$ with respect to the BDT classifier output.

The experimental and input parameter set and values used for this study are outlined in Table 6.2, with the strong phases δ_1 and δ_2, set to 0 and π respectively. The background in these studies is modelled as a simple δ-function in propertime, whilst the angular distribution is initially taken as flat. Following the same nomenclature as in section 6.3, the background p.d.f including tagging and resolution is:

[11] The calculation has been preformed using equation (5.5).

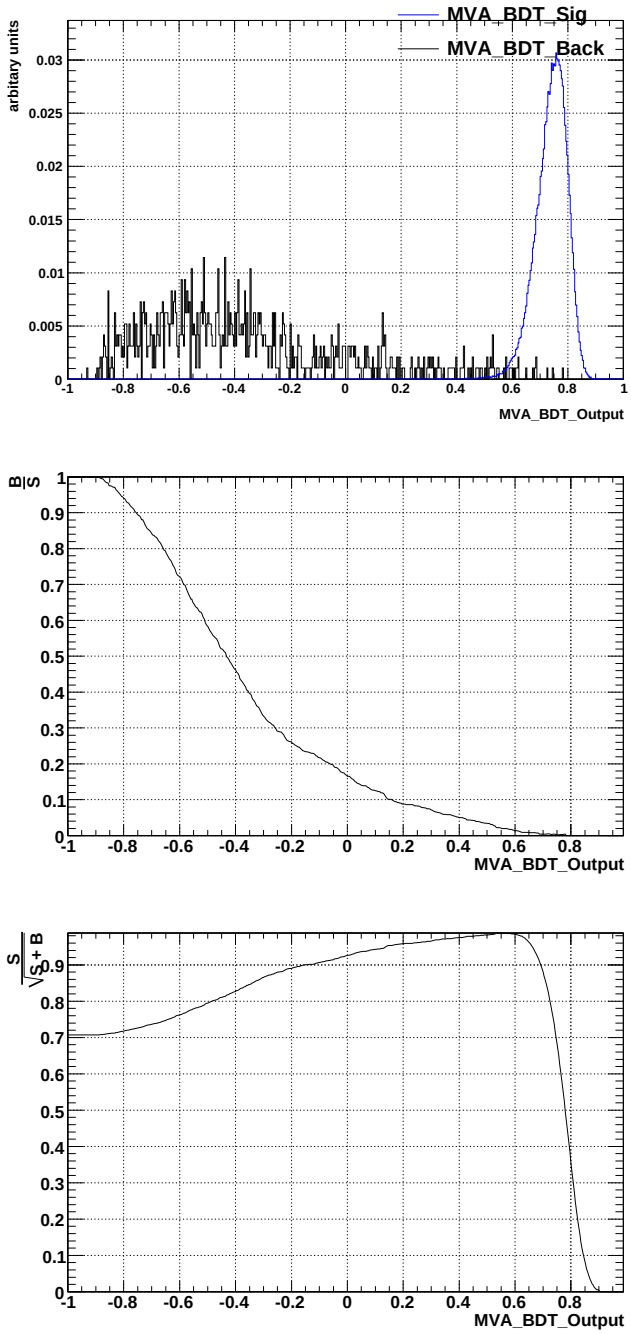


Figure 7.15: Normalised response of signal (blue) and background (black) to the BDT on variable Set 3: (Top) the BDT classifier response for signal and background (Middle), $\frac{B}{S}$ response of the BDT and (Bottom) $\frac{S}{\sqrt{S+B}}$.

$$\mathcal{B}'_{tag}(t',\Omega) = \delta(t,\Omega) \otimes \mathcal{G}_{tag}(t-t',\Omega;\sigma_t) = \mathcal{G}_{tag}(0,\Omega;\sigma_t). \tag{7.15}$$

That is a Gaussian of width σ_t centred at t=0. The p.d.f including both signal and background is then given by:

$$\mathcal{P}'(t',\Omega) = f_{sig}\mathcal{F}'(t',\Omega) + (1-f_{sig})\mathcal{B}'_{tag}(t',\Omega), \tag{7.16}$$

where $\mathcal{F}'$ following the same definition as in section 6.3.2 and f_{sig}=$\frac{N}{\left(N+N\times\left(\frac{B}{S}\right)^{J/\psi}\right)}$.

An unbinned maximum likelihood to the toy generated MC data can then be preformed, with the likelihood function being constructed following equation (6.11). In each study, for each $\frac{B}{S}$ value for Set 3 given in Table 7.8, we run at least 200 repeated experiments, fitting for the physics parameters using a ROOFIT based data model [109]. A simultaneous fit to the following parameters set $\{\Delta\Gamma_{\rm s},\ \bar{\Gamma}_{\rm s},\ {\rm R}_{\perp},\ -2\beta_{\rm s},\ {\rm R}_0\}$, was then performed. For each physics parameter, and each study under investigation, we plotted the returned fit value, the return error, and the pull distribution. These plots are included in section E.10 of Appendix E.

parameter sensitivity	$\Delta\Gamma_s$	$\bar{\Gamma}_s$	$R_\perp$	$-2\beta_s$	R_0
signal retention (%)					
100	0.0110	0.0033	0.0044	0.0270	0.0043
90	0.0118	0.0035	0.0046	0.0280	0.0044
80	0.0125	0.0037	0.0048	0.0290	0.0046
70	0.0133	0.0039	0.0051	0.0320	0.0050
60	0.0140	0.0043	0.0550	0.0340	0.0050
50	0.0160	0.0046	0.0059	0.0380	0.0058
40	0.0175	0.0052	0.0067	0.0420	0.0067
30	0.0200	0.0060	0.0077	0.0490	0.0075
20	0.0964	0.0073	0.0095	0.0600	0.0092
10	0.0350	0.0120	0.0130	0.0860	0.0129

Table 7.9: Errors obtained from the simultaneous fit, using parameter Set 3, for the varying signal retentions.

Table 7.9 shows the baseline results for the simultaneous fit to the above parameter set. The results quoted in Table 7.9 are the means of the parameter error distribution. Comparing the first row of this table with the two columns in Table 6.12, we see that the prompt background has negligible effect on the returned sensitivity. The $-2\beta_s$ precision increases from $\pm$ 0.027 rad retaining 100% of the signal, to $\pm$ 0.086 rad retaining only 10% of the signal. The error on the $-2\beta_s$ precision decreases as the signal retention rate goes down by $\frac{1}{\sqrt{N}}$. When comparing the first row of this table with the two columns in Table 6.12, we see the effect of the prompt

background has little effect on the returned sensitivity.

The extracted sensitivity for $-2\beta_s$ for each $\frac{B}{S}$ fraction, and thus the classifier output, is shown in Figure 7.16. Where it can be seen that the sensitivity remains virtually constant up to a classifier cut of 0.6. Above a cut of 0.6, although the $\frac{B}{S}$ ratio is virtually non-existant, the number of signal events decreases leading to an increased error. The bottom plot in Figure E.35, shows the enlarged region around the optimal $\frac{S}{\sqrt{S+B}}$ cut of namely 0.6; the plot show that a cut around 0.61, reduces the sensitivity of $-2\beta_s$ by less than 1% with respect to 100% signal retention.

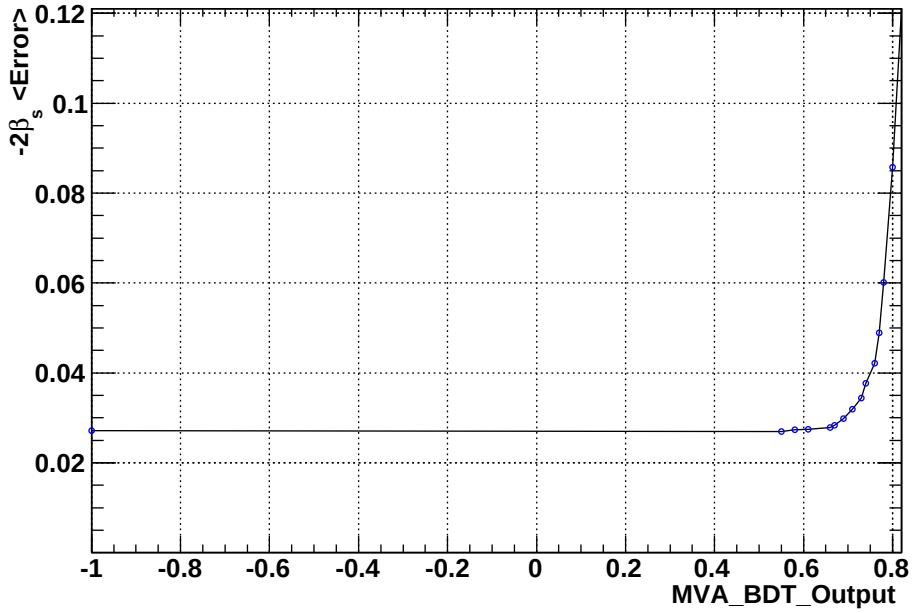


Figure 7.16: The $-2\beta_s$ sensitivity with respect to the BDT's response using parameter Set 3 and the $\frac{B}{S}$ ratios given in Table 7.8.

7.8 Conclusion

Multivariate analysis techniques including the Fisher, Log Likelihoods, ANN's and BDT's, have been studied in this chapter on $B_s \rightarrow J/\psi\phi$ signal and J/ψ-prompt background data. It was found that:

- using a reduced parameter set (Set 3), the BDT classifier gives optimal separation between signal and background data.
- with respect to the long-lived background, the sensitivity to $-2\beta_s$ increases by $\sim$20% from $\pm$0.022 to $\pm$0.027 radians using solely the J/ψ-prompt sample as background.
- using $\frac{B}{S}$ values extracted from the BDT classifiers, the $-2\beta_s$ sensitivity increases from $\pm$0.027 radians at a cut of -1 to $\pm$0.0860 radians at a cut of 0.80.
- Retaining $\sim$98% of the signal events, the propertime and decay angles can be included in a final MV cut using the BDT, without generating an acceptance effect.

8

Conclusions

The studies presented in this thesis have demonstrated LHCbs ability to extract the weak mixing phase β_s from $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decays. The novel aspect of this work has been to extend sensitivity studies of $-2\beta_s$, from the reduced angular analysis to the full (three) angular analysis for this channel. In doing so as detailed in section 6.4, we find there is a 20% improvement in our $-2\beta_s$ precision from $\pm$0.027 to $\pm$0.022 radians.

The knowledge needed to extract this sensitivity requires an understanding of the expected detector effects. From fully simulated MC data, we find a signal yield of 132k events before trigger. The L0 and HLT have a trigger efficiency of 93% and 81% respectively. A tagging efficiency and mistag rate of 62.1% and 37.4% respectively, and a background-to-signal ratio of 1.17 and 1.9 with respect to $b\bar{b}$-inclusive and J/ψ-inclusive data samples respectively. Using the above experimental parameters and the procedure outlined in section 6.3, we find that the sensitivity to $-2\beta_s$ using the full angular analysis is $\pm$0.037 radians.

Measurements of the weak B_s mixing phase through the $B_s \to J/\psi\phi$ channel are currently dominated by the Tevatron experiments, with the first tagged three angular analysis by both experiments being reported in 2008 [33, 6]. It will be a primary objective of the LHCb collaboration to improve these measurements already with the first year of data taking, with 0.5fb^{-1} of integrated luminosity available, i.e. starting from early running days. LHCb, due to its statistically large data sample and its optimal design characteristics, will dominate the determination of this phase to precision level. The status of the $-2\beta_s$ measurement is summarised in Table 8.1.

In addition, we also find that the mistag fraction (ω_{tag}) can be obtained from the $B_s \to J/\psi\phi$ data themselves, with a precision of ± 0.010. This provides qualitative improvement in measuring ω_{tag} as it either can be measured independently together with $-2\beta_s$ to a reasonable precision. This will be an important cross-check for systematical differences to external measurements of ω_{tag} which are more precise but may bear a systematic offset. This will allow to better determine the systematic uncertainty of using the most precise measurement of ω_{tag} when fitting for $-2\beta_s$.

Experiment	Luminosity [fb^{-1}]	Number of tagged events	$-2\beta_s$ [rad]
D0*	2.8	2k	$-0.57^{+0.24,+0.07}_{-0.30,-0.02}$
CDF	2.8	3k	[-0.56,-2.58] 68% CL
			$\sigma(-2\beta_s)$ [rad]
LHCb	0.5	28k **	$\pm$0.06
LHCb	2	130k	$\pm$0.037

Table 8.1: Comparison of the current measurements of $-2\beta_s$ from flavour tagged $B_s \to J/\psi\phi$ from the Tevatron experiments, and the label * indicates the simultaneous fit using the strong phases, constrained to the value obtained in $B_d \to J/\psi K^*$. The expected sensitivity of LHCb on the SM $-2\beta_s$ for 0.5 (where the label ** indicates untagged data) and 2 fb^{-1} respectively.

A full investigation of the systematic effects in the $B_s \to J/\psi\phi$ decays at LHCb has yet to be preformed. In this thesis, we have performed preliminary studies of the $-2\beta_s$ sensitivity with respect to the ω_{tag} and B_s propertime resolution. However, effects such as production asymmetry, mis-alignment of the VELO, angular acceptance effects and the angular distribution of the background have still to be investigated. Indeed, future work for this analysis will require an in-depth understanding of all the systematics effects originating from real data.

The work in this thesis also evaluated the use of several multivariate analysis classifiers, to determine the $-2\beta_s$ sensitivity. Using the J/ψ-inclusive data sample as the source of background, the optimal performing classifier was found to be the Boosted Decision Tree. The BDT classifier was then used as a final selection cut, to find the optimal background-to-signal ratio. The optimal sensitivity to $-2\beta_s$ was found to be $\pm$0.027 radians, which remained constant for BDT cut values between -1 to 0.6. The $-2\beta_s$ sensitivity was found to increasing to $\pm$0.086 radians at a cut value of 0.80. Important future work for these studies, will be to compare the results using an ensemble of independent data samples: this was not possible, due to the limited background statistics available. In addition, we found that whilst retaining 98% of the signal, lifetime depended variables could be cut on using the BDT classifier without inducing a propertime or angular acceptance effect. Work is currently underway to use this knowledge for a BDT inspired high level trigger selection for $B_s \to D_s\pi$ and $B_s \to J/\psi\phi$.

9

Literature List

- I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics Volume 1: From Relativistic Quantum Mechanics to QED*. 3^{rd} edition, Institute of Physics, 2003: [ISBN 0 7503 0864 8].
- I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics Volume 2: QCD and the Electroweak Theory*. 3^{rd} edition, Institute of Physics, 2003: [ISBN 0 7503 0950 5].
- L. H. Ryder, *Quantum Field Theory*, 2^{nd} edition, Cambridge University Press, 1996: [ISBN 0 521 47814 6].
- U. Mosel, *Fields, Symmetries and Quarks*. 2^{nd} edition, Springer, 1999: [ISBN 3 540 65235 3].
- T. Mannel, *Effective Field Theories in Flavour Physics*. Springer, 2004: [ISBN 3 540 21931 5].
- M. Beyer, *CP Violation in Particle, Nuclear and Astrophysics*. Springer, 2002: [IBSN 3 540 43705 3].
- J. J. Sakurai, *Modern Quantum Mechanics*. Addison-Wesley Publishing Company Inc, 1994. [IBSN 0 201 53929 2].
- M. L. Boas, *Mathematical Methods in the Physics Sciences*. 3^{rd} edition, Wiley, 2006. [IBSN 0 471 36580 7].
- O. Nachtmann, *Elementary Particle Physics Concepts and Phenomena*. Springer, 1989. [IBSN 0 387 51647 6].
- G. C. Branco, L. Lavoura, J. P. Silva, *International Series in Monographs on physics - 103. CP Violation*. Oxford Science Publications, 1999 [IBSN 0 19 850399 7].
- C. Jarskog, *Advance Series on Directions in High Energy Physics - Vol 3. CP Violation*. World Scientific, 1989 [IBSN 9971505614].
- D. Green,*World Scientific Lecture Notes in Physics - Vol. 55. Lectures in Particle Physics*. World Scientific, 1994 [IBSN 981 02 1683 1].
- M. S. Sozzi *Discrete Symmetries and CP violation*. Oxford Graduate Texts, 2008 [IBSN 978-0-19-92666-8].
- G. Gowan, *Statistical Data Analysis*. Oxford University Press, 1998. [IBSN 0 19 850155 2].
- R. J. Barlow *A Guide to the use of Statistical Methods in the Physical Science*. Wiley, 1989 [IBSN 0 471 92295 1]
- R. O. Duda *Pattern Classification*. Wiley, 2001 [IBSN 0 471 05669 3]

References

[1] A. D. Dolgov. *Baryogenesis, 30 years after*. arXiv:hep-ph/9707419v1, 1997.

[2] Christenson et al. *CP Violation in K^0 decays*. *Physics Review Letters*, 13, 1964.

[3] The BaBar Collaboration. *Observation of CP violation in the B^0 meson system*. *Phys. Rev. Lett*, hep-ex/0107013, 2001.

[4] The Belle Collaboration. *Obersvation of Large CP Violation in the Neutral B Meson System*. arXiv:hep-ex/0107061, 2001.

[5] The CDF Collaboration. *First Flavour-Tagged Determination of Bounds on Mixing Induced $\mathcal{CP}$ Violation in $B_s \to J/\psi\phi$ Decays*. *hep-ex*, arXiv:0712.2397v1, 2007.

[6] The D0 collaboration. *Measurement of B_s mixing parameters from the flavor-tagged decay $B_s \to J/\psi\phi$*. *hep-ex*, arXiv:0802.2255v1, 2008.

[7] W.-M Yao et al. *Review of Particle Physics*. *Journal of Physics G 33*, 2006.

[8] M. Dine. *TASI Lectures on the Strong CP Problem*. hep/ph/0011376, 2000.

[9] P. W. Higgs. *Broken symmetries, massless particles and gauge fields*. *Physics Letters*, 12(2), 1964.

[10] L. Maiani S. L. Glashow, J. Illiopoulos. *Weak Interactions with Lepton - Hadron Symmetry*. *Physics Review*, D2:1285–1292, 1970.

[11] M. Kobayashi T. Maskawa. *CP violation in the renormalizable theory of weak interactions*. *Progr. Theor. Phys.*, 49:625–657, 1973.

[12] O. Nachtmann. *Elementary Particle Physics Concepts and Phenomena*. Springer, 1989.

[13] J. Silva G. Branco, L. Lavoura. *CP Violation*. International Series of Monographs on Physics. Oxford Science Publications, Oxford, Great Britain, 1999.

[14] A. J. Buras. *Weak Hamiltonian, $\mathcal{CP}$ Violation and Rare Decays*. hep-ph/9806471, 1998.

[15] L. Wolfenstien. *Parametrization of the Kobayashi-Maskawa Matrix*. *Physical Review Letters*, 51, 1983.

[16] R. Fleischer. *Flavour Physics and CP Violation*. hep-ph/0405091, 2004.

[17] C. Jarsklog. *Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal $\mathcal{CP}$ nonconservation*. *Phys. Rev. Lett.*, 55.

[18] V. Weisskopf and E. Wigner. *Zeitschrift fur Physik.*, 63(50), 1930.

[19] E. Barberio et al. *Averages of b hadron properties at the end of 2006*. arXiv:hep-ex/0704.3575v1, 2007.

[20] J. P. Silva. *Phenomenological aspects of CP violation*. hep-ph/0410351, 2004.

[21] A. Soddu W. Hou, M. Nagashima. *Direct CP violation in $B^+ \to J/\psi K^+$ Decays as Probe for New Physics*. hep-ph/060508v1, 2006.

[22] U. Nierste A. Lenz. *Theoretical update of B_s-$\bar{B}_s$ mixing*. hep-ph/0612167, 2006.

[23] Partricia Ball and Robert Fleischer. *Probing New Physics through B mixing: Status, Benchmarcks and Prospects*. *Hep-ph*, hep-ph/0604249, 2006.

[24] Fleischer R. *CP Violation in the B System and Relations to $K \to \pi\nu\bar{\nu}$*. PhD thesis.

[25] The Tevatron collaboration. *B Physics at the Tevatron: Run 2 and beyond*. page 583, arXiv:hep-ph/0201071, 2002.

[26] U. Nierste Dunietz I., Fleisher R. *In persuit of new physics in B_s mixing*. *Phys. Rev. D.*, 63, 2001.

[27] I. Dunietz. *$B_s - \overline{B_s}$ mixing, $\mathcal{CP}$ violation, and extraction of CKM phases from untagged B_s data samples*. *Physical Review D*, 52(5), 1995.

[28] Dighe A.S. Dunietz I. and Fleisher R. *Extracting CKM phases and $B_s - \bar{B}_s$ mixing parameters from angular distributions of non-leptonic B decays*. *Eur. Phys. J.*, C6:647–662, 1999.

[29] J. Charles et al. *CP violation and the CKM matrix: assessing the impact of the asymetric B factories*. page 131, 2007.

[30] M. Okamoto et al. *Nucl. Phys. Proc. Suppl.*, 140.

[31] A. Khodjamirian et al. *Predictions on $B \to \pi l \nu$, $D \to \pi l \nu_l$, and $D \to K l \nu_l$ from QCD light-cone sum rules*. *Phys. Rev. D*, 62.

[32] The Belle Collaboration. *Measurement of Time-Dependent CP-Violating Asymmeries in $B^0 \to \phi K^0_s$, $K^+K^-K^0_s$, and $\eta' K^0_s$ Decays*. hep-ex/0308036.

[33] CDF Collaboration. *An Updated Measurement of the CP Violating Phase $\beta_s^{J/\psi}$*, 2008.

[34] on behalf of the BaBar Collaboration E. Latour. *Measurement of the CKM angle γ at Babar*, 2007.

[35] The CDF Collaboration. *Measurement of $\Delta\Gamma_s$ and ϕ_s in $B_s \to J/\psi\phi$ decays*. *PHYS*, CDF-8950, 2007.

[36] D. Guadagnoli M. Blanke, A. J. Buras and C. Tarantino. *Minimal Flavour Violation Waiting for Precise Measurements of Δm_s, $S_{\psi\phi}$, A^S_{SL},$|V_{ub}|$,γ and $B^0_{s,d} \to \mu^+\mu^-$*. *hep-ph*, arXiv:0604057v5, 2006.

[37] A. J. Buras M. Blanke. *A Guide to Flavour Changing Neutral Currents in the Littest Higgs Model with T-Parity*. *hep-ph*, arXiv:0703117v1, 2007.

[38] R. Fleischer. *CP Violation and B Physics at the LHC*. *hep-ph*, arXiv:0703127v2, 2007.

[39] CDF Collaboration. *First Flavor-Tagged Determination of Bounds on CP Violation in $B_s \to J/\psi\phi$*. `http://www-cdf.fnal.gov/physics/new/bottom/071214.blessed-tagged_BsJPsiPhi/`, year = 2008.

[40] D0 Collaboration. *Measurement of the Bs mixing parameters from the flavor-tagged decay $B_s \to J/\psi\phi$ Likelihood Scans*. `http://www-d0.fnal.gov/Run2Physics/WWW/results/final/B/B08A/likelihoods/`, year = 2008.

[41] G. Lanfranchi. *Search for new physics in $B_s \to J/\psi\phi$ at LHC*, 2008.

[42] CERN. *The CERN Document Server (CDS)*.

[43] LHCb collaboration. *LHCb Technical Design Report, Reoptimized Detector Design and Performance*. Technical Report 030, CERN, Scientific and Technical Services Division, CERN, Geneva, 2003.

[44] J. Baines et al. *Bottom Production*. *hep-ph*, arXiv:0003142v2, 2001.

[45] LHCb collaboration. *LHCb Technical Design Report*. Technical Report 004, CERN, Scientific and Technical Services Division, CERN, Geneva, 1998.

[46] D. Eckstein. *The LHCb VErtex LOcator*. CERN-LHCb/2003-099, 2003.

[47] LHCb Collaboration. *The LHCb Detector at LHC. Journal of Instrumentation (JINST)*, 2008.

[48] LHCb collaboration. *LHCb Technical Design Report, Inner Tracker*. Technical Report 029, CERN, Scientific and Technical Services Division, CERN, Geneva, 2002.

[49] LHCb collaboration. *LHCb Technical Design Report, Outer Tracker*. Technical Report 024, CERN, Scientific and Technical Services Division, CERN, Geneva, 2001.

[50] J. Nardulli. *The LHCb Tracking System and its performance*. CERN-LHCb/2006-031, 2006.

[51] LHCb collaboration. *LHCb Technical Design Report, RICH*. Technical Report 050, CERN, Scientific and Technical Services Division, CERN, Geneva, 2000.

[52] LHCb collaboration. *LHCb Technical Design Report, Calorimeters*. Technical Report 049, CERN, Scientific and Technical Services Division, CERN, Geneva, 2000.

[53] A. Arefiev et al. *Test Results of the LHCb Electromagnetic Calorimeter*. CERN-LHCb/2007-149.

[54] LHCb collaboration. *LHCb Technical Design Report, Muon*. Technical Report 029, CERN, Scientific and Technical Services Division, CERN, Geneva, 2001.

[55] J. A. Hernando. *The LHCb Trigger. Acta Physica Polonica B*, 38(3), 2006.

[56] G. Barrand et al. *A software architecure and framework for building LHCb data processing applications. Proc. of CHEP*, 2000.

[57] D. J. Lange. *The EvtGen particle decay simulation package. Nuclear Int. and Methods in Physics Research A*, 462, 2001.

[58] LHCb Collaboration. *Gauss: The LHCb simulation package*. `http://lhcb-comp.web.cern.ch/lhcb-comp/Simulation/Gauss.pdf`.

[59] GEANT 4 Collaboration. *GEANT 4 a simulation toolkit*. 2003.

[60] LHCb Collaboration. *Boole: The LHCb digitisation program: Users guide and reference manual*. `http://lhcb-release-area.web.cern.ch/LHCb-release-area/BOOLE/BOOLE_v6r5/doc/Boole.pdf`.

[61] LHCb Collaboration. *Brunel: The LHCb reconstruction program: Users guide*. `http://lhcb-release-area.web.cern.ch/lhcb-comp/Reconstruction/BUG.pdf`.

[62] LHCb Collaboration. *DaVinci: The LHCb analysis project*. `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/`.

[63] T. Sjöstrand et al. *Pythia 6.4 physics and manual*. hep-ph/0603175, 2006.

[64] LHCb Collaboration. *DaVinci Associators*.

[65] LHCb collaboration. *LHCb - Parameters of DC06 Monte Carlo productions*. 2007.

[66] T. du Pree. *PVV_CPLH*. `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/generator/PVV\_CPLH.pdf`.

[67] Raven G. *Selection of $B_s^0 \to J/\psi\phi$ and $B^+ \to J/\psi K^+$. PHYS*, LHCb/2003-118, 2003.

[68] J. Amoraal. *Panning for the Golden Decay $B_s \to J/\psi\phi$ at LHCb*. 2005.

[69] The CDF Collaboration. *Observation of $B_s \to \psi(2S)\phi$ and measurement of* BF*($B_s \to \psi(2S)\phi$)*/BF*($B_s \to J/\psi\phi$)*. CDF-7525, 2005.

[70] E. Rodrigues. *The LHCb Trigger System*. CERN-LHCb/2006-065, 2006.

[71] Fabio C. D. Metlica. *The LHCb Detector and Trigger*. *Nuclear Physics B (Proc. Suppl.)*, 167, 2007.

[72] H. Ruiz. *HLT Description*. `http://ruizh.home.cern.ch/ruizh/HltResults/5000v2r0Old/v2r0/`.

[73] H. Ruiz. *HLT Alley performance*. `http://ruizh.home.cern.ch/ruizh/HltResults/5000v2r0Old/v2r0/`.

[74] R. Raven. *Selection of $B_s \to J/\psi\phi$ and $B^+ \to J/\psi K^+$*.

[75] Fernandez L. *Exclusive Trigger Selections and Sensitivity to the $B_s - \bar{B}_s$ Mixing Phase at LHCb*. PhD thesis, EPFL, 2006.

[76] J. van Hunen M. Prokudin R. White A. Golutvin, R. Hierck. *$B_s^0 \to D_s^{\pm} K^{\mp}$ and $B_s^0 \to D_s^- \pi^+$ event selection*. CERN-LHCb/2003-127, 2003.

[77] V. V. Gligorov. *Reconstruction of the Channel $B_d \to D^+\pi^-$ and the background classification at LHCb*. CERN-LHCb/2007-044, 2007.

[78] O. Callot. *The Forward Tracking: Algorithm and Performance Studies*. CERN-LHCb/2007-015, 2007.

[79] H. Ruiz. *HLT Description*, 2008. `http://lhcb-trig.web.cern.ch/lhcb-trig/HLT/HltDescription.htm#inclusiveandexclusiveselections`.

[80] Fernandez L. *Sensitivity to the $B_s - \bar{B}_s$ Mixing Phase at LHCb*. *PHYS*, LHCb/2006-047, 2006.

[81] Raven G. *Sensitivity Studies of χ and $\Delta\Gamma$ with $B_s \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$*. *PHYS*, LHCb/2003-119, 2003.

[82] T. W. Versloot et al. *Position Reconstruction and Charge Distribution in LHCb VELO Silicon Sensors*. CERN-LHCb/2007-119, 2007.

[83] C. Gö bel S. Amato, M. Gandelman and L. de Paula. *Update to the LHCb sensitivity to $\sin(2\beta)$ from the CP-asymmerty in $B^0 \to J/\psi(\mu\mu)K_s^0$ decays*. CERN-LHCb/2007-045, 2007.

[84] M. Musy M. Calvi, O. Leroy. *Flavour Tagging Algorithms and Preformance in LHCb*.

[85] B. Souza de Paula. *Studies on Systematic Effects of the Trigger on Flavour Tagging at the Generator Level*. CERN-LHCb/2006-046, 2006.

[86] Somerville L. *Performance of the LHCb RICH Photon Detectors and Tagging Systematics for CP Violation Studies*. PhD thesis, University of Oxford, 2006.

[87] A. Osorio-Oliveros C. Clarke, C. Mclean. *Sensitivity studies to $-2\beta_s$ and $\Delta\Gamma_s$ using the full $B_s \to J/\psi\phi$ angular analysis at the LHCb*. CERN-LHCb/2007-101, 2007.

[88] Yao W.-M. et al. *Review of Particle Physics*. *Journal of Physics G*, 33:1+, 2006.

[89] A. Abulencia et al. *Observation of B_s^0 - $\bar{B}_s^0$ Oscillations*. *Phys. Rev. Lett.*, 97:242003, hep-ex/0609040, 2006.

[90] P. Ball et al. *B decays at the LHC*. hep-ph/0003238, 2000.

[91] The Babar collaboration. *Measurement of Decay Amplitudes of $B \to (c\bar{c})K^*$ with an Angular Analysis for $(c\bar{c}) = J/\psi$,$\psi(2S)$ and χ_{cl}*. hep-ex/0704.0522, 2007.

[92] James F. and Roos M. *'MINUIT' a System for Function Minimisation and Analysis of the Parameter Errors and Correlations*. *Comput. Phys. Commun.*, 10:343–367, 1975.

[93] G. Gomez-Ceballos K. Makhoul C. Paus A. Savoy-Navarro (CDF Collaboration) G. P. Di Giovanni, I. Furic. *Analysis of Untagged and Tagged $B_s \rightarrow J/\psi\phi$ decays with 1.7/1.3fb^{-1}. CDF/DOC/BOTTOM/CDFFR/8893*, 2007.

[94] The Babar Collaboration. *Measurement of decay amplitudes of $B \rightarrow J/\psi K^\star$, $\phi(2S)K^\star$, and $\chi_{C1}K^\star$ with an angular analysis. Phys. Rev. D*, 76, 2007.

[95] C. Parkes D. Petrie and S. Viret. *Study of the impact of VELO misalignments on the LHCb tracking and L1 trigger performance*. CERN-LHCb/2005-056, 2005.

[96] E. Rodrigues M. Gersabeck. *Impact of misalignments on the analysis of B decays*. CERN-LHCb/2008-012, 2008.

[97] S. Jonsell P. Ferola, T. Holopainen. *Asymmetries in B decays and their experimental control*. ATLAS PHYS No: 054, 1994.

[98] R. A. Fisher. *The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics*, 7:179–188, 1936.

[99] The D0 collaboration. *Search for single top quark production at D0 using neural netorks. Physics Letters B*, 2001.

[100] The D0 Collaboration. *Evidence for production of single top quarks. arXiv:0803.0739v1*, 2008.

[101] B. Joao. *Credit scoring with boosted decision trees. Munich Peronal RePEc Archive*, 2007.

[102] A. Ji Zhu B. P. Roe, H. Yang. *Boosted Decision trees, a powerful event classifier. Phystat*, 2005.

[103] The D0 Collaboration. *Evidence for Production of Single Top quarks and First Direct Measurement of $|V_{tb}|$. Physical Review Letters*, 2007.

[104] Y. Freund and R. E. Schapire. *Experiments with a new boosting algorthim, in Machine Learning. Proceedings of the Thirteenth International Conference*, 1996.

[105] J. Zhu H. Yang, B. Roe. *Studies of Boosted Decision Trees for MiniBooNE Identification. physics/0508045*, 2005.

[106] F. Tegenfelt H. Voss. K. Voss A. Höcker, J. Stelzer. *TMVA 3.7 - Toolkit for Multivariate Data Analysis with Root. arXviphysics/0703039*, 2007.

[107] et al. R. Barlow. *Recommended Statistical Procedures for BaBar*. Barbar/318, 2002.

[108] *Handbook of Methods of Applied Statistics*. John Wiley & Sons, year = 1967, author = I. Chakravarti, R. Laha and J. Roy,.

[109] Wouter Verkerke and David Kirkby. *The Roofit Toolkit for Data Modelling.* `http://roofit.sourceforge.net/`.

[110] T. Mannel. *Review of Heavy quark effective theory*. hep-ph/9611411v1, 1996.

[111] M. Neubert. *Hevy-Quark Effective Theory*. hep-ph/9610266v1, 1996.

[112] T. Inami and C.S. Lim. *Effects of superheavy and leptons in low-energy weak processes $K_L \rightarrow \mu\mu$,$K^+ \rightarrow \pi^+\nu\nu$ and K^0 to or from K^0. Journal of Physics G*, 65, 1981.

[113] A. I. Sanda A. B. Cater. *CP Nonconservation in Cascade Decays of B Measons. Physical Review Letters*, 45(12), 1981.

[114] A. I. Sanda I. I. Bigi. *Nuclear Physics*, B281.

[115] eprint = BaBar-515 year = 2000 Stephane T'Jampens), title = $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ Angular Distribution and Time Dependences.

[116] R. Fleischer A. Dighe, I. Dunietz. *Extracting CKM phases and $B_s - \overline{B_s}$ mixing parameters from angular distrubutions of non-leptonic B decays*. *The European Physical Journal C*, 6, 1999.

[117] M. Needham. *Performance of the LHCb Track Reconstruction Software*. CERN-LHCb/2008-001, 2008.

[118] A. Inselberg. *Parallel Coordinates: A tool for visulizing Multi-Dimensional Geometry*.

[119] V. Shelkov A. Hocker, S. Laplace. *Search for $B^0 \rightarrow a_0^{\pm}(980)\pi^{\mp}$*. BABAR/2001-141, 2001.

[120] J. Zhu Y. Liu I. Stancu G. McGregor B. P. Roe, H-J. Yang. *Boosted decision trees as an alternative to artificial neutral networks for particle identification*. *Nuclear Instruments and Methods in Physics Research A*, 2005.

[121] E. J. Wegman. *Hyperdimensional Data Analysis Parallel Coordinates*. *Journal of the American Statistical Association*, 85, 1990.

A

Theory of $B_s \to J/\psi\phi$ decays

A.1 Phenomenology in the B_s-system

The purpose of this section is to derive expressions for Δm_s and $\Delta\Gamma_s$, within the framework of an Effective Field Theory EFT. Effective field theories are nothing more than an approximation to a given theory, containing the appropriate degrees of freedom to describe the physical phenomena at a chosen length scale, while ignoring those degrees at shorter distances (or higher energies).

An example of a effective theory in QCD is the Heavy quark Effective Theory, which is applicable to B_s mesons. This theory uses the infinite mass limit of QCD, to approximate heavy quarks as static sources of colour. This approximation manifests into additional sources of flavour and spin symmetry in QCD, providing a model independent starting point for the description of weak interactions involving heavy quarks [110]. These approximate symmetries also work to suppress the hadronic uncertainties entering from QCD. Thus providing a clean theoretical framework for determining CKM matrix elements, for example V_{cb} from semi-leptonic $b \to c$ transitions [111]. The EFT applicable to the weak Hamiltonian takes the general structure:

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V^i_{CKM} C_i(\mu) Q_i \tag{A.1}$$

Here G_F is the usual Fermi constant and Q_i are the relevant local operators which govern the decay in question. V^i_{CKM} describes our usual CKM elements, while $C_i(\mu)$ are called the Wilson co-efficients. The couplings $C_i(\mu)$ summarise the physics contribution from scales higher than μ. Provided the scale μ is not too low[1] these couplings can be calculated pertubabtively. In general the couplings will depend on the top quark contribution, contributions from other heavy particles such as W, Z and new particles associated with extensions of the SM.

[1]The scale can be chosen arbitrary, but it is customary to choose μ to be of the order of the mass of the decaying hadron. Thus $\mathcal{O}(m_b) \sim 1$ GeV/c for B-decays; much greater than the typical QCD scale, λ_{QCD}, of a few hundred MeV/c.

A.1.1 Expressing B_s-$\bar{B}_s$ oscillation diagrams as effective box vertices

B-meson decays, like the oscillation diagrams as shown in Figure 2.2, are described by Feynman diagrams with W-, Z- and t quark exchange. Such diagrams are really describing what happens at very short distances $\mathcal{O}(M_w,M_z,M_t)$, whereas the more realistic picture of the decay is that of a process containing effective point-like vertices, which we can represent as local operators Q_i. The Wilson co-efficients can then be regarded as effective coupling constants associated to these effective vertices. The EFT picture of such a decays then looks like a series of effective vertices multiplied by effective coupling constants. This series is called the Operator Product Expansion (OPE).

These effective vertices generally involve both quarks and leptons, as depicted in Figure A.1. Where i,j,m,n denotes the external quarks or leptons and k,l the internal quarks and leptons.

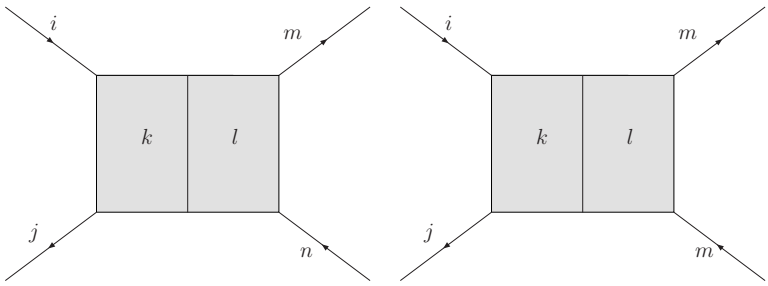


Figure A.1: Effective Box vertices

It is possible to calculate these effective vertices using elementary Feynman vertices and propagators found in the electroweak theory. For instance, the effective box vertex for $B_s - \overline{B_s}$ can be resolved in terms of its elementary vertices, shown in Figure A.2.

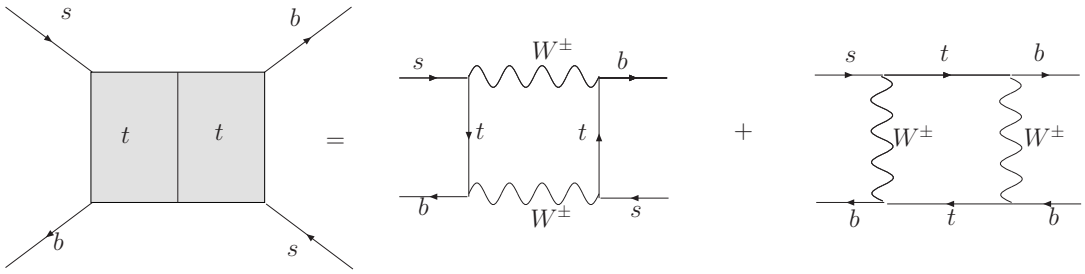


Figure A.2: Box vertices for the B_s box diagram resolved in terms of elementary vertices.

With help of elementary vertices and propagators, shown on the right-hand diagram in Figure A.2, it is possible to derive "Feynman rules" for the effective box vertex. It is usual to calculated

these elementary vertices and propagators in the t'Hooft-Feynman gauge and using the approximation that external quarks have zero momenta. This leads to the following box vertex formula for $B_q - \bar{B}_q$ mixing:

$$\text{Box } (\Delta s\text{=2}) = \lambda_i^2 \frac{G_F^2}{\sqrt{2}} M_W^2 S_0(x_i)(\bar{q}b)_{V-A}(\bar{q}b)_{V-A}, \tag{A.2}$$

where $\lambda_i = V_{is}^* V_{id}$. Equation (A.2) allows the calculation of the effective Hamiltonian for FCNC processes, without the inclusion of QCD corrections. There are some notable properties of equation (A.2) which are worth mentioning. The first is that the formula depends on the mass of the internal quarks and are calculable functions of:

$$x_i = \frac{M_i^2}{M_W^2}, \quad \text{i = u,c,t.} \tag{A.3}$$

The second point is that the function, $S_0(x_i)$, is one from a set of basic function[2] which govern the physics of all FCNC processes. The function $S_0(x_i)$ governs the physics of the $B_q - \bar{B}_q$ effective mixing processes[3] and takes the form:

$$\begin{aligned} S_0(x_t) = & \quad \frac{4x_t - 11x_t^2 + x_t^3}{4(1-x_t)^2} - \frac{3x_t^2 \ln x_t}{2(1-x_t)^2} \\ \sim & \quad 2.46 \left(\frac{m_t}{170\text{ GeV}}\right)^{1.52}, \quad \text{with } m_t \text{ in the range 150 GeV} \leq m_t \leq \text{200 GeV} \end{aligned} \tag{A.4}$$

It can actually be written more formally as:

$$S_0(x_t) \equiv F(x_t, x_t) + F_{x_u, x_u} - 2F(x_t, x_u), \tag{A.5}$$

where $F(x_i, x_{i(j)})$ are called the "base" functions corresponding to a given box diagram with i and j quark exchanges. These functions can be found by drawing all possible box diagrams and using the unitarity of the CKM matrix, in particular the relation:

$$\lambda_u + \lambda_c + \lambda_t = 0, \tag{A.6}$$

which allows the effective Hamiltonian to be obtained by summing over t and c quarks only. It is possible to formulate a general description of the effective Hamiltonian (A.1), using these base functions, for any decay process as:

$$\mathcal{H}_{eff}^{FCNC} = \sum_k C_k Q_k, \tag{A.7}$$

The co-efficients of these operators (Q_k) are simply linear combinations of the base functions - the Inami-Lim functions - multiplied by the corresponding CKM factors - λ_i. Consequently it is

[2]These base functions are called the Inami-Lim functions, the set of functions is defined in [112]

[3]The subscript "0" on $S_0(x_i)$, and in general all Lnami-lim functions, indicates that the functions do not include QCD corrections.

possible to write the amplitude for all FCNC processes and transitions as linear combinations of the basic (m_t-dependent) functions $F_r(x_t)$ with co-efficients P_r which characteristic the decay under investigation. This is often called the "Penguin-Box" expansion and takes the general form:

$$A(\text{decay}) = P_0(\text{decay}) + \sum_r P_r(\text{decay}) F_r(x_t). \tag{A.8}$$

Where the sum runs over all possible functions contributing to a given amplitude. P_0 summarises contributions stemming from internal quarks other than the top. Equation (A.8) is incredible useful, for example this description can be used to explain the GIM mechanism in FCNC processes.

A.1.2 The GIM mechanism in FCNC processes

It is the GIM mechanism which denies tree level FCNC process in the SM. Its effect is also felt upon the effective Hamiltonian given in equation (A.1) (or A.7), where its maximum effect occurs when the masses of the given internal quarks in the loop diagrams are equal: m_t = m_c = m_u. Focusing on the box diagrams for the B_s-system, we find that the CKM factors entering (A.8) via:

$$C_k \propto \sum_{i=u,c,t} \lambda_i F(x_i), \quad x_i = \frac{m_i^2}{m_W^2} \tag{A.9}$$

and satisfying (A.6) are given by,

$$\lambda_i = V_{ib}^* V_{is} \quad \text{B-decays} \quad B_s^0 - B_s^0, \tag{A.10}$$

which implies vanishing co-efficients (C_k) in the limit x_u=x_c=x_t. For complete CKM cancellation of FCNC beyond tree level we therefore need:

- unitarity of the CKM matrix.
- Exact equality of quark masses of a given charge, that is to say horizontal flavour symmetry.

In nature however, we do not see such a horizontal flavour symmetry and this is the reason why the GIM mechanism breaks down at the one-loop level. The size of this breakdown, and consequently the size of FCNC processes, depending on the mass differences of the quarks and the behaviour of the basic functions.

For small $x_i \ll 1$ relevant for $i \neq t$, we observe that the function $S_0(x_i)$ behaves as $S_0(x_i) \propto x_i$. This implies "hard" (quadratic) GIM suppression of FCNC processes governed by the function S_0, provided the top quark contribution can be neglected. For large x_t we have $S_0(x_t) \propto x_t$. Thus our $B_q - \bar{B}_q$ mixing diagrams are governed by the top quark

contributions, and the GIM suppression is not effective. This then signifies the special property that FCNC B transitions have. The appearance of the top quark, in the internal loop with $m_t \gg M_W \gg m_c, m_u$, removes the GIM suppression making B decays useful places for studying FCNC processes. In contrast, the FCNC decays of D-mesons are strongly suppressed since only d,s and b quarks, with m_d,m_s and $m_b \ll M_W$, enter the internal loops. The GIM mechanism for these decays is much more effective.

B_s mixing therefore proceeds to an excellent approximation through box diagrams with internal top quark exchanges, and with contributions from internal u and c quarks GIM suppressed[4]. The effective Hamiltonian for $B_s - \overline{B_s}$ mixing[5], at the scale μ=μ_b=$\mathcal{O}(m_b)$ [6], can be expressed as:

$$\mathcal{H}_{eff}^{\Delta B=2} = \frac{G_F^2}{16\pi^2} M_W^2 (V_{tb}^* V_{ts})^2 S_0(x_t) \{ \sum_{k=1}^{2} C_k(\mu) Q_k + \sum_{k=3}^{10} C_k Q_k \}. \tag{A.11}$$

With the summation $\sum_{k=3}^{10}$ taken over penguin operators being neglected. The only operators to consider are the tree operators $Q_{1,2}$:

$$\begin{aligned} Q_1(\Delta B = 2) &= (\bar{b}_\alpha s_\beta)_{V-A} (\bar{b}_\beta s_\alpha)_{V-A}, \\ Q_2(\Delta B = 2) &= (\bar{b}_\alpha s_\alpha)_{V-A} (\bar{b}_\beta s_\beta)_{V-A}. \end{aligned} \tag{A.12}$$

Where these matrix elements have different colour structure, α,β, but the same flavour structure. The effective Hamiltonian for B_s mixing, with leading order QCD corrections, is then given by:

$$\mathcal{H}_{eff}^{\Delta B=2} = \frac{G_F^2}{6\pi^2} M_W^2 (V_{tb}^* V_{V_{ts}})^2 B_{B_s} F_{B_s}^2 m_{B_s}^2 \eta_{B_s} S_0(x_t). \tag{A.13}$$

Where F_{B_s} is the B-meson decay constant, and B_{B_s} is bag parameter. These parametrises the non-perturbative effects in the hadronic matrix elements and are expected to be of $\mathcal{O}$(1). η_{B_s} is the QCD correction also of magnitude unity.

The dispersive (mass) part of the B_s mixing diagram is dominated by the t quark contribution, as is given by:

$$M_{12} = -\frac{G_F^2 M_W^2}{12\pi^2} \eta_{B_s} M_{B_s} F_{B_s^2} B_{B_s} S_0(x_t) (V_{ts}^* V_{tb})^2, \tag{A.14}$$

[4]Indeed these quark masses can be set to zero.

[5]The effective Hamiltonian for $B_s - \overline{B_s}$ describes a flavour changing process of 2 units, i.e. ΔF=2=ΔB. This processes can be described by the produce of two ΔF=1=ΔB processes. The effective Hamiltonian for this type of transition, again neglecting QCD effect, generally involves the inclusion of penguin operations

[6]Again noting that QCD effects are to be neglected.

whereas the absorptive (lifetime) part is determined by real intermediate states, which correspond to common decay products of the B_s and $\bar{B_s}$, such that:

$$\Gamma_{12} = \frac{G_F^2 M_W^2}{8\pi}\eta_{B_s} M_{B_s} F_{B_s}^2 B_{B_s} \left[(V_{ts}^* V_{tb})^2 + (V_{ts}V_{tb}V_{cs}V_{cb})\mathcal{O}(\frac{m_c^2}{m_b^2})\right] + (V_{cs}V_{cb})^2\mathcal{O}(\frac{m_c^4}{m_b^4}). \tag{A.15}$$

The above two equations cannot be reliably calculated due to the hadronic factors, $F_{B_s}\ B_{B_s}$. However in the ratio $\frac{\Gamma_{12}}{M_{12}}$ the largest uncertainties cancel leading to:

$$\left|\frac{\Gamma_{12}}{M_{12}}\right| = \frac{3\pi}{2}\frac{m_b^2}{m_W^2}\frac{1}{S_0(x_t)} \sim \mathcal{O}(\frac{m_b^2}{m_t^2}), \tag{A.16}$$

$$\frac{\Gamma_{12}}{M_{12}} = \left|\frac{\Gamma_{12}}{M_{12}}\right| e^{i\theta_{12}}, \quad \theta_{12} \equiv arg\left(-\frac{M_{12}}{\Gamma_{12}}\right) \sim \mathcal{O}\left(\frac{m_c^2}{m_b^2}\right). \tag{A.17}$$

Equations (A.16) and (A.17) can then be used to expand $\frac{q}{p}$ to obtain:

$$\begin{aligned}
\frac{q}{p} = &\quad e^{-i\theta_{12}}\left[1 - \left|\tfrac{\Gamma_{12}}{M_{12}}\right|\sin\theta_{12}\right] + \mathcal{O}\left(\left|\tfrac{\Gamma_{12}}{M_{12}}\right|^2\right), \\
\Delta M_s = &\quad 2\left|M_{12}\right|\left[1 + \mathcal{O}\left(\left|\tfrac{\Gamma_{12}}{M_{12}}\right|^2\right)\right], \\
\Delta\Gamma_s = &\quad 2\left|\Gamma_{12}\right|\cos\theta_{12}\left[1 + \mathcal{O}\left(\left|\tfrac{\Gamma_{12}}{M_{12}}\right|^2\right)\right].
\end{aligned} \tag{A.18}$$

Where $\theta_{12} \equiv arg\,(M_{12})$. To an excellent precision:

$$\Delta M = 2\left|M_{12}\right|, \quad \Delta\Gamma = 2\left|\Gamma_{12}\right|\cos\theta_{12} \simeq 2\left|\Gamma_{12}\right|, \tag{A.19}$$

with the right-hand sides given by (A.14) and (A.15). In the expression for $\Delta\Gamma_s$ above, the $\cos(\theta_{12})$ has been taken as unity since $\frac{m_c^4}{m_b^4} \approx 0.007$ within the SM. The above expressions for ΔM and $\Delta\Gamma$ are non-negative quantities in the Standard Model.

A.2 Decays from a neutral meson system

We now concentrate on the time-dependent decay rates from a neutral B meson system into a final state f, accessible to both B_q and $\overline{B_q}$. These rates depend on the decay amplitudes of a neutral B meson, of known initial flavor, as it evolves in time

$$\langle f|\mathcal{H}_{eff}|B_q^0(t)\rangle = \mathcal{A}(\mathcal{F}_+(t) + \lambda_f\mathcal{F}_-(t)), \quad \langle f|\mathcal{H}_{eff}|\overline{B_q^0}(t)\rangle = \mathcal{A}\frac{p}{q}(\mathcal{F}_-(t) + \lambda_f\mathcal{F}_+(t)).$$

Using these amplitudes the general time-dependent decay rates into a final state, f, can be expressed as

$$\begin{aligned}
\Gamma[B_q^0(t) \to f] &\to |\langle f|H_{eff}|B_q\rangle|^2 \\
&= \langle f|H_{eff}|B_q^0(t)\rangle\langle f|H_{eff}|B_q(t)\rangle^\dagger \\
&= \mathcal{A}_f(\mathcal{F}_+(t) + \lambda_f\mathcal{F}_-(t))\mathcal{A}_f^*(\mathcal{F}_+ + \lambda_f^*\mathcal{F}_-^*(t)) \\
&= |\mathcal{A}_f|^2\left[|\mathcal{F}_+(t)|^2 + |\lambda_f|^2|\mathcal{F}_-(t)|^2 + 2\Re\left[\lambda_f^*\mathcal{F}_-^*(t)\right]\right]
\end{aligned}$$

where the time evolution in the decay rates expression is governed by the functions $\mathcal{F}\pm(t)$ explicitly given by

$$\begin{aligned}
|\mathcal{F}_\pm(t)|^2 &= \frac{1}{4}\left(e^{-\Gamma_H t} + e^{-\Gamma_L t} \pm 2e^{-\Gamma t}\cos(\Delta M_q t)\right) + \frac{e^{-\Gamma t}}{2}\left(\cosh(\frac{\Delta\Gamma_q t}{2}) \pm \cos(\Delta M_q t)\right) \\
\mathcal{F}(t)_+^*\mathcal{F}_-(t) &= \frac{1}{4}\left(-e^{-\Gamma_H t} + e^{-\Gamma_L t} + 2ie^{-\Gamma t}\sin(\Delta M_q t)\right) + \frac{e^{-\Gamma t}}{2}\left(\sinh(\frac{\Delta\Gamma_q}{2}) + i\sin(\Delta M_q t)\right) \\
\mathcal{F}(t)_-^*\mathcal{F}_+(t) &= \frac{1}{4}\left(-e^{-\Gamma_H t} + e^{-\Gamma_L t} - 2ie^{-\Gamma t}\sin(\Delta M_q t)\right) + \frac{e^{-\Gamma t}}{2}\left(\sinh(\frac{\Delta\Gamma_q}{2}) - i\sin(\Delta M_q t)\right)
\end{aligned}$$

using (A.20) the decay rates can be expressed generally as

$$\begin{aligned}
\Gamma[B_q^0(t) \to f] \propto\ & \frac{1}{4}\left|\frac{q}{p}\right|^2|\overline{A_f}|^2\left[\left[\left|1 + \frac{1}{\lambda_f}\right|^2\right] + \left[\left|1 - \frac{1}{\lambda_f}\right|^2\right]\right. \\
& \left. - 2\left(1 - \frac{1}{|\lambda_f|^2}\right)e^{\Delta\bar{\bar{\Gamma}}_q t}\cos(\Delta M_q t) - 4\Im\left(\frac{1}{\lambda_f}\right)e^{-\Delta\bar{\Gamma}_q t}\sin(\Delta M_q t)\right]
\end{aligned}$$

From equation (A.20) is possible to see that $\mathcal{CP}$ in mixing or decay induces a $\cos(\Delta M_q t)$ term, while $\mathcal{CP}$ in the interference of mixing and decay will induce a $\sin(\Delta M_q t)$ term. Equation (A.20) can be rewritten more conveniently for neutral B meson systems as

$$\begin{aligned}
\Gamma[B_q^0(t) \to f] \propto\ & |\mathcal{A}_f|^2(1 + |\lambda_f|^2)\frac{e^{-\Gamma t}}{2} \\
& \left[\cosh(\frac{\Delta\Gamma_q t}{2}) - \mathcal{A}_{\Delta\Gamma}\sinh(\frac{\Delta\Gamma_q t}{2}) + \mathcal{A}_{\mathcal{CP}}^{dir}\cos(\Delta M_q t) - \mathcal{A}_{\mathcal{CP}}^{mix-ind}\sin(\Delta M_q t)\right]
\end{aligned}$$

where

$$\mathcal{A}_{\Delta\Gamma} = \frac{2\Re\lambda_f}{1+|\mathcal{A}_f|^2}, \quad \mathcal{A}_{\mathcal{CP}}^{dir} = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}, \quad \mathcal{A}_{\mathcal{CP}}^{mix-ind} = \frac{2\Im\lambda_f}{1+|\lambda_f|^2} \tag{A.20}$$

$$\begin{aligned}\bar{\Gamma}_{\overline{B_q^0}\to f(t)} \quad \propto \quad & |\mathcal{A}_f|^2|\frac{p}{q}|^2(1+|\lambda|^2)\frac{e^{-\Gamma t}}{2}\Big(\cosh(\frac{\Delta\Gamma_q t}{2}+\mathcal{A}_{\Delta\Gamma}\sinh(\frac{\Delta\Gamma_q t}{2}) \\ & -\mathcal{A}_{\mathcal{CP}}^{dir}\cos(\Delta M_q t)+\mathcal{A}_{\mathcal{CP}}^{mix-ind}\sin(\Delta M_q t))\Big)\end{aligned} \tag{A.21}$$

The other two decay rates, into the charge conjugate final states, are obtained by substituting f by $\bar{f}$ in equation (A.21). The factors $\mathcal{A}_{\Delta\Gamma}$ and $\mathcal{A}_{\mathcal{CP}}^{dir}$ are sensitive to the phase of $\lambda_f(\lambda_{\bar{f}})$ and thus $\mathcal{CP}$ violation.

The general expression for the time-dependent $\mathcal{CP}$ asymmetry is then:

$$\begin{aligned}\mathcal{A}_{\mathcal{CP}}(t) \quad &= \quad \frac{\Gamma(B_q^0(t)\to f_{\mathcal{CP}})-\Gamma(\overline{B_q^0}(t)\to f_{\mathcal{CP}})}{\Gamma(B_q^0(t)\to f_{\mathcal{CP}})+\Gamma(\overline{B_q^0}(t)\to f_{\mathcal{CP}})}, \\ &= \quad \frac{|\mathcal{A}|^2(1+|\lambda_f|^2)\frac{e^{-\Gamma t}}{2}(\mathcal{A}_{\mathcal{CP}}^{dir}\cos(\Delta M_q t)-\mathcal{A}_{\mathcal{CP}}^{mix-ind}\sin(\Delta M_q t))}{|\mathcal{A}|^2(1+|\lambda_f|^2)^2\frac{e^{-\Gamma t}}{2}(\cosh(\frac{\Delta\Gamma_q t}{2})-\mathcal{A}_{\Delta\Gamma}\sinh(\frac{\Delta\Gamma_q t}{2}))}.\end{aligned}$$

In the absence of direct and $\mathcal{CP}$ violation in mixing, $|\lambda_f|$ = 1, this expression reduces to

$$\mathcal{A}_{CP}(t) = \frac{-\Im(\lambda_f)\sin(\Delta M_q t)}{\cosh(\frac{\Delta\Gamma_q t}{2})+\Re\lambda_f\sinh(\frac{\Delta\Gamma_q t}{2})}. \tag{A.22}$$

In the limit $\Delta\Gamma_q = 0$ which is a good approximation for the B_d system equation (A.22) reduces further to

$$\mathcal{A}_{CP}(t) = -\Im(\lambda_f)\sin(\Delta M_q t). \tag{A.23}$$

A.3 The $B_d \to J/\psi(\mu^+\mu^-)K_s$ decay

Interest in searching for $\mathcal{CP}$ violation in the B system, was sparked by two seminal articles by Carter and Sanda [113] and Bigi and Sanda [114] in the 1980's that discussed the prospect of observing $\mathcal{CP}$ violation outwith the kaon system using $B_d \to J/\psi(\mu^+\mu^-)K_s$ decays.

Although this is the most famous B decay for measuring $\mathcal{CP}$ violation, its calculation is quite complex. The calculation can be simplified without loss of generality, using the following approximations:

- Although the final state, $J/\psi K_s$, is not a $\mathcal{CP}$ eigenstate owing to the K_s the assumption that they are $\mathcal{CP}$ eigenstate is approximately valid, if we assume $\mathcal{CP}$ violation through kaon mixing, $\delta_K \sim 10^{-3}$, is negligible with respect to $\mathcal{CP}$ violation present in $B_d \to J/\psi(\mu^+\mu^-)K_s$ decay, which is of $\mathcal{O}(1)$.

- Using the spectator quark model, allowing only the decays $B_d \to J/\psi(\mu^+\mu^-)K_s$ and $\overline{B_d} \to J/\psi\overline{K^0}$. The $B_d \to J/\psi(\mu^+\mu^-)K_s$ is then only possible due to the $K-\overline{K}$ mixing which is accounted for through,

$$\frac{q_K}{p_K} = -\eta_K e^{i(\epsilon_K+\epsilon_d+\epsilon_s)}\frac{V_{us}^* V_{ud}}{V_{us}V_{ud}^*} \tag{A.24}$$

- Since the J/ψ vector meson and the kaon arise from a B decay, they must be in a relative l=1 state, which applying a $\mathcal{CP}$ ($\mathcal{P}$) transformation will yield an extra minus sign.

- The decay is dominated by the tree level diagram show in the bottom diagram in Figure A.3. But the decay also receives contributions from penguin diagrams as shown in the top plot in Figure A.3. The amplitudes for the two diagrams are proportional to,

$$\begin{aligned} \mathcal{A}_{tree} &\propto V_{cb}^* V_{cs} e^{i(\theta_t+\delta_t)}, \\ \mathcal{A}_{pen} &\propto V_{tb}^* V_{ts} e^{i(\theta_p+\delta_p)}. \end{aligned}$$

  The penguin diagrams are usually suppressed due to higher order weak interactions, and as a result the decay is overwhelmingly dominated by one weak phase, θ_t. Further more, since the final state is assumed a $\mathcal{CP}$ eigenstate, no direct $\mathcal{CP}$ violation is expected and we expect the same strong phases, $\delta_t = \delta_p = \delta$.

Using the above approximations and the CKM elements given in Figure A.3 the expression

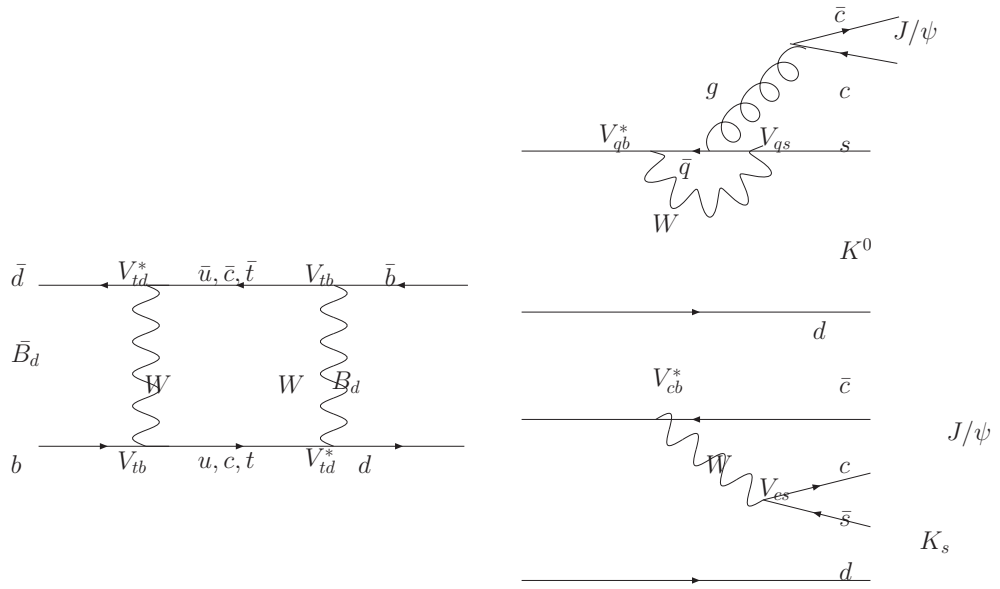


Figure A.3: Diagrams contributing to the $B_d \to J/\psi(\mu^+\mu^-)K_s$ decay. (Left) Box diagram contributing to $\overline{B_d} - B_d$ mixing, before decaying via an electroweak penguin (Top), or via the tree diagram (Bottom).

for $\mathcal{CP}$ violation in $B_d \to J/\psi(\mu^+\mu^-)K_s$ decays is given by

$$\begin{aligned}
\lambda_{J/\psi K_s} &= \left(\frac{q}{p}\right)\left(\frac{\bar{\mathcal{A}}_{J/\psi K_s}}{\mathcal{A}_{J/\psi K_s}}\right)\left(\frac{p}{q}\right)_K \\
&= (-\eta_B)\left(\frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}\right)(-\eta_K)\left(\frac{V_{us} V_{ud}^*}{V_{us}^* V_{ud}}\right)(-)\left(\frac{V_{cb} V_{cs}^*}{V_{cb}^* V_{cs}}\right) \\
&= -\eta_B \eta_K |z| e^{-i(\beta_d - \beta_K)} \frac{1}{|z|} e^{-i(\beta_d - \beta_K)} \\
&= -e^{-2i(\beta_d - \beta_K)}. \qquad \text{(A.25)}
\end{aligned}$$

Where $\eta_B = \eta_K = -1$ and the expressions for β_d and β_K come from equation 2.34 has been used. Neglecting the small angle, β_K, originating from the squashed (sd) triangle and using the assumption that there is only one weak phase, we find

$$\begin{aligned}
\mathcal{A}_{\mathcal{CP}}^{dir} &= 0 \\
\mathcal{A}_{\mathcal{CP}}^{mix-ind} &= \Im(\lambda_{J/\psi K_s}) = \eta_B \eta_K \sin(2\tilde{\beta}_d) \qquad \text{(A.26)}
\end{aligned}$$

where the phase $\tilde{\beta}_d$ [7] includes the possibility that there might be new contributions to the relevant phase in $B_d - \overline{B_d}$ mixing.

[7] In the SM $\tilde{\beta}_d$ coincides with the CKM phase β_d.

A.4 The helicity bases

There are many choices of co-ordinate system which can be used to define the decay angles. One originates from the standard helicity formalism [115]: using the helicity angles of the final state particles with respect to the daughter particles. The decay of each daughter via a two-body process can be used to define a decay plane in each daughter's rest frame. The polar angles between these decay planes and the parents decay plane provides two of the three decay angles. The third, comes from the azimuthal angle between the two daughter decay planes. This basis is illustrated in Figure A.4.

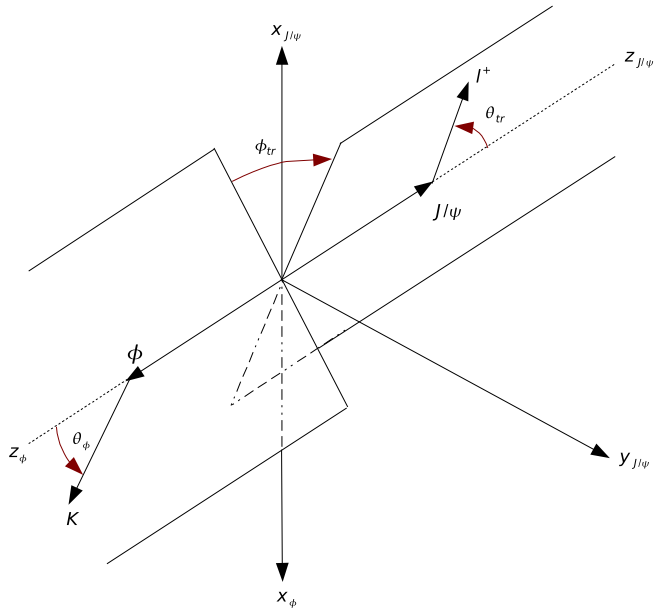


Figure A.4: Definition of the decay angles in the Helicity Basis [115].

A.5 Time dependent amplitudes of the $B_s \to J/\psi\phi$ decay rate expressions

$$|A_0(t)|^2 = \frac{|A_0(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} - 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.27)}$$

$$|A_\parallel(t)|^2 = \frac{|A_\parallel(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} - 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.28)}$$

$$|A_\perp(t)|^2 = \frac{|A_\perp(0)|^2}{2}\Big[(1-\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} + 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.29)}$$

$$\mathrm{Re}\{A_0^*(t)A_\parallel(t)\} = \frac{1}{2}|A_0(0)||A_\parallel(0)|\cos(\delta_2-\delta_1)\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} - 2e^{-\Gamma_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.30)}$$

$$\mathrm{Im}\{A_\parallel^*(t)A_\perp(t)\} = |A_\parallel(0)||A_\perp(0)|\Big[e^{-\Gamma_s t}\left\{\sin\delta_1\cos(\Delta m_s t) - \cos\delta_1\sin(\Delta m_s t)\cos(2\beta_s)\right\} + \frac{1}{2}\left(e^{-\Gamma_{\rm H}t} - e^{-\Gamma_{\rm L}t}\right)\cos\delta_1\sin(2\beta_s)\Big] \quad \text{(A.31)}$$

$$\mathrm{Im}\{A_0^*(t)A_\perp(t)\} = |A_0(0)||A_\perp(0)|\Big[e^{-\Gamma_s t}\left\{\sin\delta_2\cos(\Delta m_s t) - \cos\delta_2\sin(\Delta m_s t)\cos(2\beta_s)\right\} + \frac{1}{2}\left(e^{-\Gamma_{\rm H}t} - e^{-\Gamma_{\rm L}t}\right)\cos\delta_2\sin(2\beta_s)\Big] \quad \text{(A.32)}$$

In the case of the $\overline{B_s^0}$ the time evolution is given by the conjugate of the functions $\bar{h}^k(t)$ and the angular terms remain the same:

$$|\bar{A}_0(t)|^2 = \frac{|A_0(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_L t} + (1-\cos(2\beta_s))\,e^{-\Gamma_H t} + 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.33)}$$

$$|\bar{A}_\parallel(t)|^2 = \frac{|A_\parallel(0)|^2}{2}\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_L t} + (1-\cos(2\beta_s))\,e^{-\Gamma_H t} + 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.34)}$$

$$|\bar{A}_\perp(t)|^2 = \frac{|A_\perp(0)|^2}{2}\Big[(1-\cos(2\beta_s))\,e^{-\Gamma_L t} + (1+\cos(2\beta_s))\,e^{-\Gamma_H t} - 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.35)}$$

$$\mathrm{Re}\{\bar{A}_0^*(t)\bar{A}_\parallel(t)\} = \frac{1}{2}|A_0(0)||A_\parallel(0)|\cos(\delta_2-\delta_1)\Big[(1+\cos(2\beta_s))\,e^{-\Gamma_L t} + (1-\cos(2\beta_s))\,e^{-\Gamma_H t} + 2e^{-\Gamma_s t}\sin(\Delta m_s t)\sin(2\beta_s)\Big] \quad \text{(A.36)}$$

$$\mathrm{Im}\{\bar{A}_\parallel^*(t)\bar{A}_\perp(t)\} = -|A_\parallel(0)||A_\perp(0)|\Big[e^{-\Gamma_s t}\{\sin\delta_1\cos(\Delta m_s t) - \cos\delta_1\sin(\Delta m_s t)\cos(2\beta_s)\} - \frac{1}{2}\left(e^{-\Gamma_H t} - e^{-\Gamma_L t}\right)\cos\delta_1\sin(2\beta_s)\Big] \quad \text{(A.37)}$$

$$\mathrm{Im}\{\bar{A}_0^*(t)\bar{A}_\perp(t)\} = -|A_0(0)||A_\perp(0)|\Big[e^{-\Gamma_s t}\{\sin\delta_2\cos(\Delta m_s t) - \cos\delta_2\sin(\Delta m_s t)\cos(2\beta_s)\} - \frac{1}{2}\left(e^{-\Gamma_H t} - e^{-\Gamma_L t}\right)\cos\delta_2\sin(2\beta_s)\Big] \quad \text{(A.38)}$$

Where at time t=0, the angular distribution for $B_s \to J/\psi\phi$ depends on the observables $A_\parallel(0)$, $A_\perp(0)$, $A_0(0)$ and on the two phases $\delta_1 \equiv \{A_\parallel^*(0)A_\perp(0)\}$ and $\delta_2 \equiv \{A_0^*(0)A_\parallel(0)\}$. δ_1 and δ_2 are the $\mathcal{CP}$ conserving strong phases that are expected to be close to 0 and π respectively in the absence of significance final-state interactions [116].

A.6 The reduced $B_s \to J/\psi\phi$ decay rate expressions

$$\frac{d\Gamma(t)}{d\cos\theta_{tr}} \propto (|A_{\parallel}(t)|^2 + |A_0(t)|^2)(1+\cos^2\theta_{tr}) + \frac{1}{2}(|A_{\perp}(t)|^2)\sin^2\theta_{tr} \tag{A.39}$$

If we expand the time expressions, we obtain:

$$\begin{aligned}\frac{d\Gamma(t)}{d\cos\theta_{tr}} \quad \propto \quad & (|A_{||}(0)|^2 + |A_0(0)|^2)[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} \\ - \quad & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\frac{1}{2}(1+\cos^2\theta_{tr}) \\ + \quad & (|A_{\perp}(0)|^2)[(1-\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} \\ + \quad & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\sin^2\theta_{tr}\end{aligned} \tag{A.40}$$

which can be further simplified by introducing the $\mathcal{CP}$-odd fraction $R_{\perp}$:

$$\begin{aligned}\frac{d\Gamma(t)}{d\cos\theta_{tr}} \quad \propto \quad & (1-R_{\perp})[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} \\ - \quad & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\frac{1}{2}(1+\cos^2\theta_{tr}) \\ + \quad & R_{\perp}[(1-\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} \\ + \quad & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\sin^2\theta_{tr}\end{aligned} \tag{A.41}$$

with the corresponding expression for the $\overline{B_s^0}$ decay rate given by:

$$\begin{aligned}\frac{d\bar{\Gamma}(t)}{d\cos\theta_{tr}} \quad \propto \quad & (1-R_{\perp})[(1+\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1-\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} \\ + \quad & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\frac{1}{2}(1+\cos^2\theta_{tr}) \\ + \quad & R_{\perp}[(1-\cos(2\beta_s))\,e^{-\Gamma_{\rm L}t} + (1+\cos(2\beta_s))\,e^{-\Gamma_{\rm H}t} \\ - \quad & 2e^{-\bar{\Gamma}_s t}\sin(\Delta m_s t)\sin(2\beta_s)]\sin^2\theta_{tr}\end{aligned} \tag{A.42}$$

From equations (A.41) and (A.42) it is evident that the $\mathcal{CP}$ components are separated by their lifetime and by their distinct angular separation.

B

Selection of $B_s \rightarrow J/\psi\phi$ events

B.1 Resolution studies

The aim of this section is to outline the different physical resolutions associated with $B_s \rightarrow J/\psi\phi$ events. In the following plots in this section, the entries are for the number of candidates associated with the $B_s \rightarrow J/\psi\phi$ signal and passing the off-line selection described above, before applying the trigger selection. In order to describe the reconstructed quantities we will introduce the following definitions:

- **Residual δx:** Given an observable x, the residual is defined as the difference between the direct measurement x_{rec} (reconstructed or fitted quantity) and the Monte Carlo truth value x_{MC}, such that: $\delta x = x_{rec} - x_{MC}$. The residual therefore represents the error on x.

- **Resolution σ_x^{res}:** is defined as the standard deviation of a Gaussian fit to the residual δx.

- **Pull:** The pull of a variable x is the residual δx normalised to the estimated error σ_x^{err} of this difference. The error is generally determined from the covariance matrix of x_{rec}. Assuming Gaussian errors, the pull $(x_{rec} - x_{MC})/\sigma_x^{err}$ should exhibit a Normal Gaussian distribution with zero mean, i.e. no bias and unit width (denoted by $\sum_x$). Studying the pull distribution helps determine wrong error assignments and/or incorrect assumptions.

- **scale factor $\sum_x^{cor}$:** this is a correction factor which can be applied to obtain a more accurate measure of the mean of the error distribution, if the width of the pull distribution is biased: $\sum_x \times \bar{x}^{err}$

B.2 Mass resolution studies

In the EVTGEN package [57], the B_s^0, J/ψ, ϕ mesons are generated with the following central masses (M_{MC}) and lifetimes (τ_{MC}):

$$
\begin{aligned}
B_s^0\text{:}\ M_{MC} &= 5369.60\ \mathrm{MeV/c^2}, \quad \tau_{MC} = 1.461{\times}10^{-12}s;\\
J/\psi\text{:}\ M_{MC} &= 3096.87\ \mathrm{MeV/c^2}, \quad \tau_{MC} = 7.56567{\times}10^{-21}s;\\
\phi\text{:}\ M_{MC} &= 1019.42\ \mathrm{MeV/c^2}, \quad \tau_{MC} = 1.15045{\times}10^{-21}s;
\end{aligned}
$$

The related intrinsic decay widths of the particles is obtained from $\Gamma_{MC} \sim \frac{\hbar}{\tau_{MC}}$. In case of the B_s^0 meson, the width is of the order 10^{-10} $\mathrm{MeV/c^2}$. Therefore all B_s^0 mesons are generated with the same mass in the simulation. For the J/ψ and ϕ, the widths are equal to 0.087 $\mathrm{MeV/c^2}$ 4.458 $\mathrm{MeV/c^2}$ respectively, which is small compared the resolution of the detector. It is therefore possible to consider the J/ψ and ϕ mesons as particles with fixed generated mass. Figure 4.6 shows the mass resolutions of the B_s^0, J/ψ and ϕ mesons obtained after applying the off-line selection cuts.

The J/ψ mass distribution is fitted with a Breit-Wigner[1] giving a mean value of μ = (3095.1 $\pm$ 0.03) $\mathrm{MeV/c^2}$ and a resolution of $\sigma_m^{res}(J/\psi)$ = (7.84 $\pm$ 0.07) $\mathrm{MeV/c^2}$. This means that the J/ψ mass window cut of $\pm$ 85 $\mathrm{MeV/c^2}$ corresponds roughly to a 11 σ requirement on both sides of the nominal mass. A Breit Wigner is also used to fit the ϕ invariant mass distribution, which exhibits radiative tails either side of its nominal value. The mean value is μ = (1019.5 $\pm$ 0.009) $\mathrm{MeV/c^2}$ with a resolution of $\sigma_m^{res}(\phi)$ = (6.18 $\pm$ 0.02) $\mathrm{MeV/c^2}$. For the B_s^0, the invariant mass distribution is calculated using a double Gaussian giving a mean value of μ = (5.367 $\pm$ 0.063) $\mathrm{MeV/c^2}$. With the resolution of the core Gaussian, containing 70.2% of the events, begin $\sigma_m^{res}(B_s^0)$ = (15.003$\pm$ 0.1) $\mathrm{MeV/c^2}$ and that of the outer gaussian being $\sigma_m^{res}(B_s^0)$ = (38.033$\pm$ 0.7) $\mathrm{MeV/c^2}$.

The invariant mass of the B_s^0 was also calculated after applying a mass-constrained vertex fit to its daughters; J/ψ and ϕ mesons. After applying this criteria, the B_s^0 mass was again fitted using a double Gaussian, which in this instance leads an improved mean value of μ = (5369.46 $\pm$ 0.02) $\mathrm{MeV/c^2}$, approaching its nominal value. The resolution of the core Gaussian also narrows to $\sigma_m^{res}(B_s^0)$ = (6.78$\pm$ 0.023) $\mathrm{MeV/c^2}$ for the core Gaussian (0.89%) and $\sigma_m^{res}(B_s^0)$ = (28.54$\pm$0.29) $\mathrm{MeV/c^2}$ for the outer gaussian; this is shown in the bottom left-hand plot in Figure 4.6. By applying a mass constrained vertex fit to the daughters of the B_s^0, we find the resolution on the B_s^0 mass improves roughly by a factor of two, which is in agreement with the studies detailed in [71]. The B_s^0 mass window of $\pm$ 50 $\mathrm{MeV/c^2}$ used in this study represents a $\approx$ 7 σ requirement on both sides of its mean value.

[1] where the Breit-Wigner is given by: $BW(x, m, \Gamma) \approx \frac{1}{(x-m)^2+\frac{1}{4}\Gamma^2}$.

B.3 Propertime resolution studies

Summarised in equations (B.1) and (B.2), and shown in Figures B.3 and B.5 below, are the resolutions[2] and pulls for each primary and secondary vertex components used in calculating the B_s propertime given in equation (6.1):

$$\begin{aligned} P_x: \quad \sigma_{P_x}^{res} &= 13.6 \pm 3.25 \mu m, \quad \sum_{P_x} = 1.63 \pm 0.0045, \\ P_y: \quad \sigma_{P_y}^{res} &= 13.0 \pm 1.62 \mu m, \quad \sum_{P_y} = 1.61 \pm 0.0044, \\ P_z: \quad \sigma_{P_z}^{res} &= 49.9 \pm 2.10 \mu m, \quad \sum_{P_z} = 1.15 \pm 0.0032. \end{aligned} \tag{B.1}$$

$$\begin{aligned} S_x(B_s): \quad \sigma_{S_x}^{res} &= 16.6 \pm 2.77 \mu m, \quad \sum_{S_x} = 1.16 \pm 0.0032, \\ S_y(B_s): \quad \sigma_{S_y}^{res} &= 16.2 \pm 2.58 \mu m, \quad \sum_{S_y} = 1.14 \pm 0.0032, \\ S_z(B_s): \quad \sigma_{S_z}^{res} &= 260.1 \pm 3.14 \mu m, \quad \sum_{S_z} = 1.12 \pm 0.0031. \end{aligned} \tag{B.2}$$

From (B.1) and (B.2) we see that the transverse directions (x,y) have much better resolutions than in the z direction, this is due to the fact that the majority of the momentum is travelling down the beam pipe, and not perpendicular to it. For completeness, we also give the resolution and pulls for each of the components related to the B_s momentum in equation (B.3) and Figure B.4 below:

$$\begin{aligned} p_x(B_s): \quad \sigma_{p_x}^{res} &= 21.8 \pm 1.89 MeV/c, \quad \sum_{p_x} = 1.486 \pm 0.0041, \\ p_y(B_s): \quad \sigma_{p_y}^{res} &= 22.0 \pm 3.05 MeV/c, \quad \sum_{p_y} = 1.487 \pm 0.0041, \\ p_z(B_s): \quad \sigma_{p_z}^{res} &= 466 \pm 0.27 MeV/c, \quad \sum_{p_y} = 1.537 \pm 0.0042. \end{aligned} \tag{B.3}$$

Where the slightly worse momentum resolution as compared to [75] can be attributed to the increase in material simulated between the DC04 and DC06 generated Monte Carlo data [117].

[2] where the resolutions for each component has been computed using $\sigma = (f_1 \times \sigma_1 + f_2 \times \sigma_2)^{1/2}$, using the results of the double Gaussian fits to the residuals. Further details can be found in Appendix B.3.

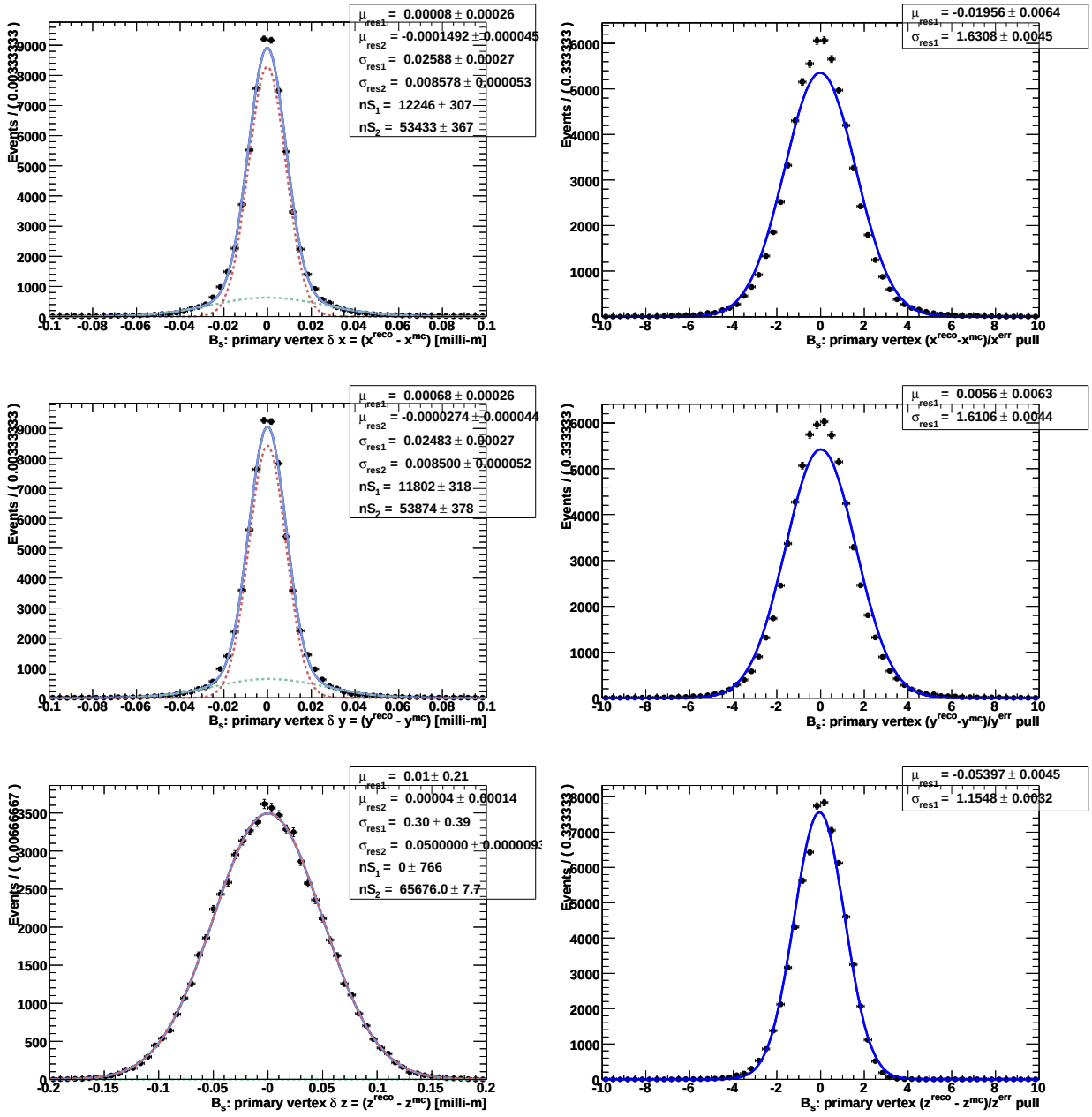


Figure B.1: Left plots: primary vertex residuals in the x direction (top), y direction and the z direction (bottom) [mm]. Right plots: the associated pull distributions for the primary vertex components in the x direction (top), y direction (middle) and z direction (bottom).

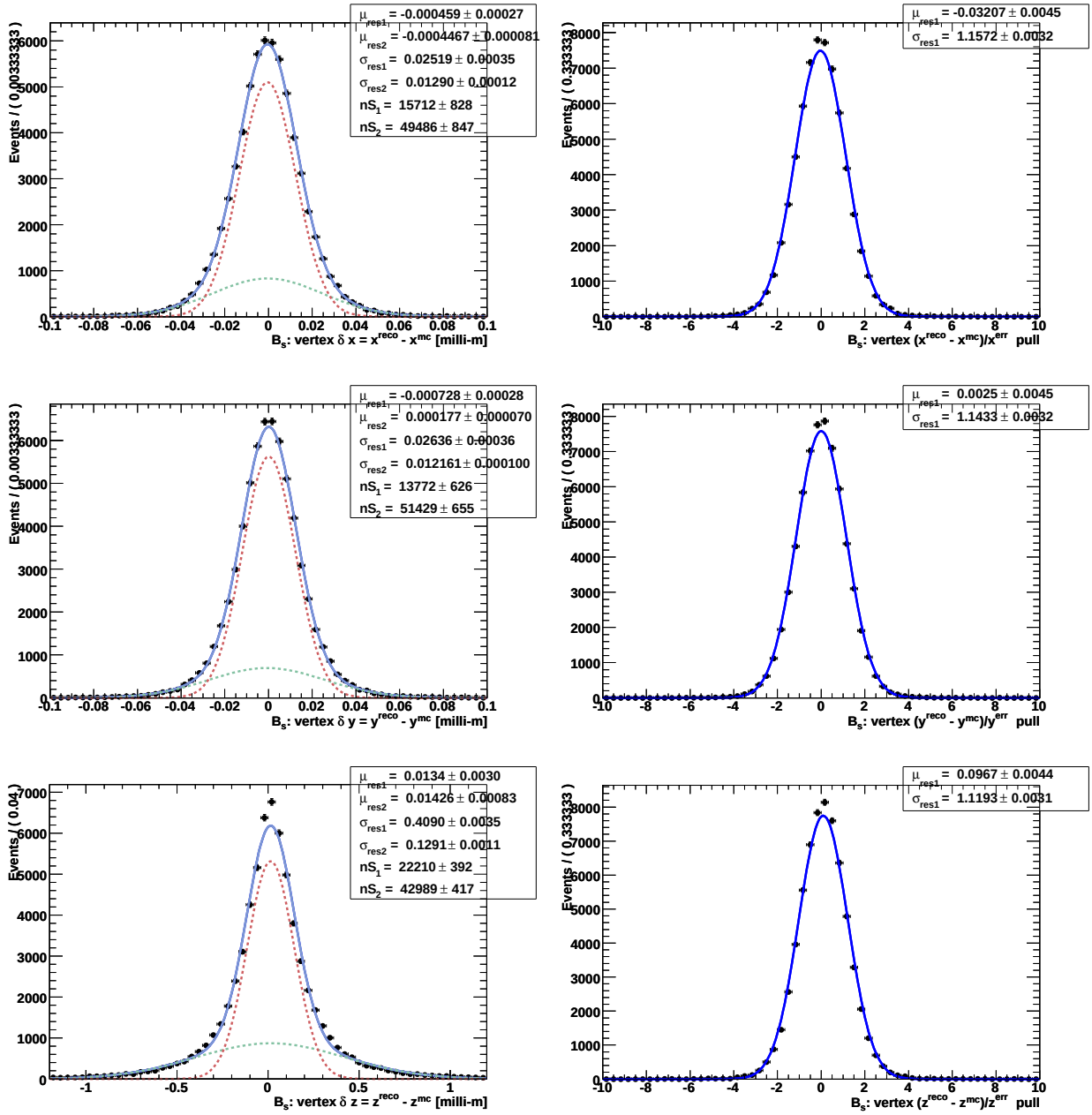


Figure B.2: Left plots: secondary decay vertex residuals in the x direction (top), y direction and the z direction (bottom) [mm]. Right plots: the associated pull distributions for the secondary decay vertex components in the x direction (top), y direction (middle) and z direction (bottom).

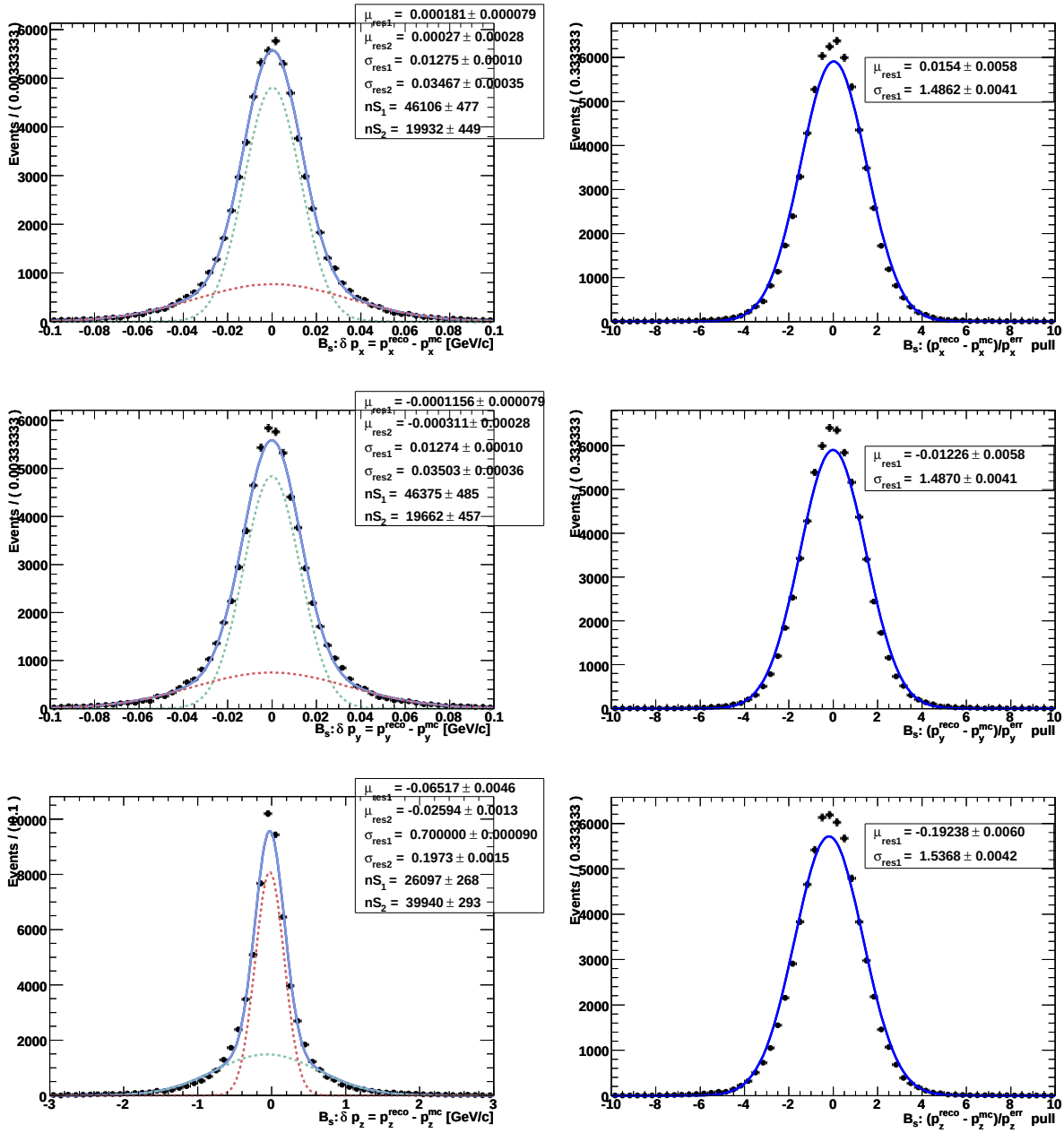


Figure B.3: Left plots: B_s momentum vector residuals in the x direction (top), y direction and the z direction (bottom) [$\mathrm{GeV/c}$]. Right plots: the associated pull distributions for the B_s momentum vector components in the x direction (top), y direction (middle) and z direction (bottom).

B.4 Propertime resolution double Gaussian fit

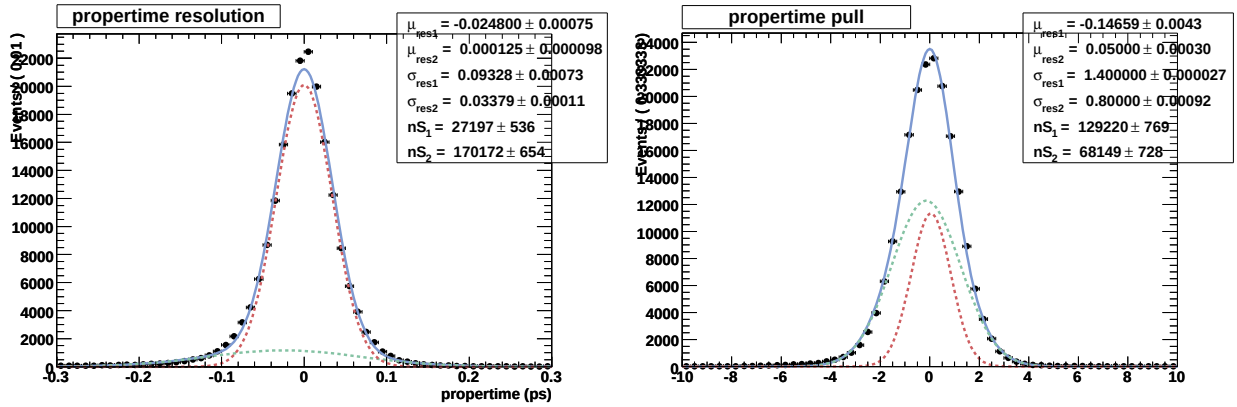


Figure B.4: The B_s^0 decay time residual (left) and pull distribution (right) have been fitted with a double Gaussian.

B.5 B_s propertime acceptance plots

By applying a cut on the propertime significance of > 5 [67], we would expect to bias the propertime acceptance distribution. This is illustrated in Figure B.5, where the top left-hand distribution shows the acceptance effect without using the trigger. The effect of applying the L0 and L0 plus HLT triggers being shown in the left-hand plot and bottom plot of Figure B.5, respectively. The effect of the lifetime bias selection is therefore to remove events at low propertime and thus the negative tail of these distribution, making it impossible to extract the propertime resolution.

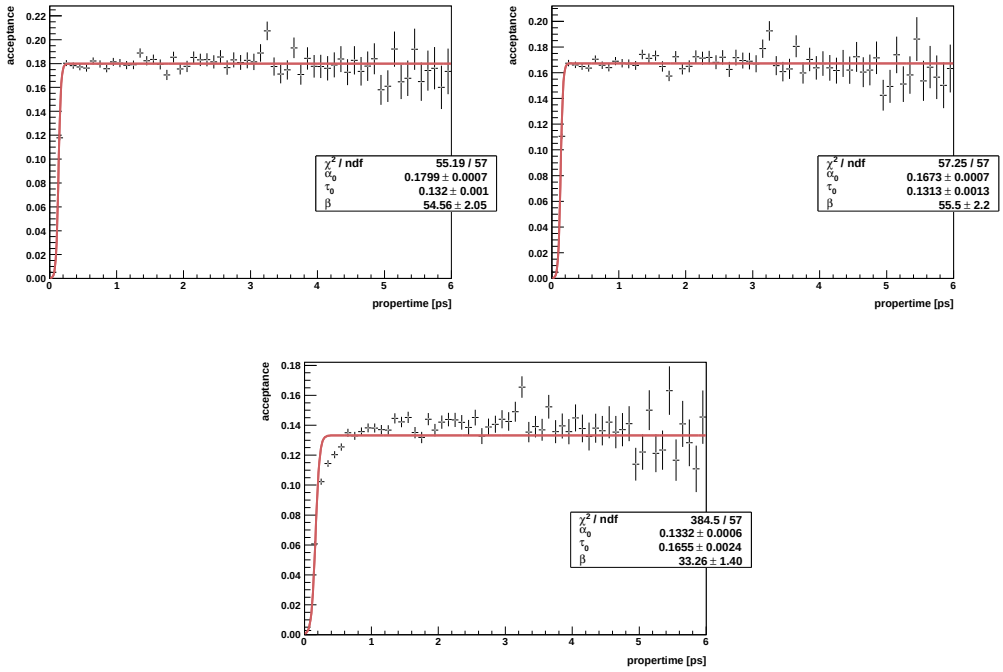


Figure B.5: Shown in the top left-hand plot is the acceptance of lifetime bias selected events as a function of propertime before the L0 and HLT triggers. Where the bias selection in this case implies the application of a 5σ propertime significance cut, in addition to the off-line selection cuts. The top right-hand plot shows the acceptance effect due to the global L0 decision, while the bottom plot shows the effect of L0 and HLT trigger decision.

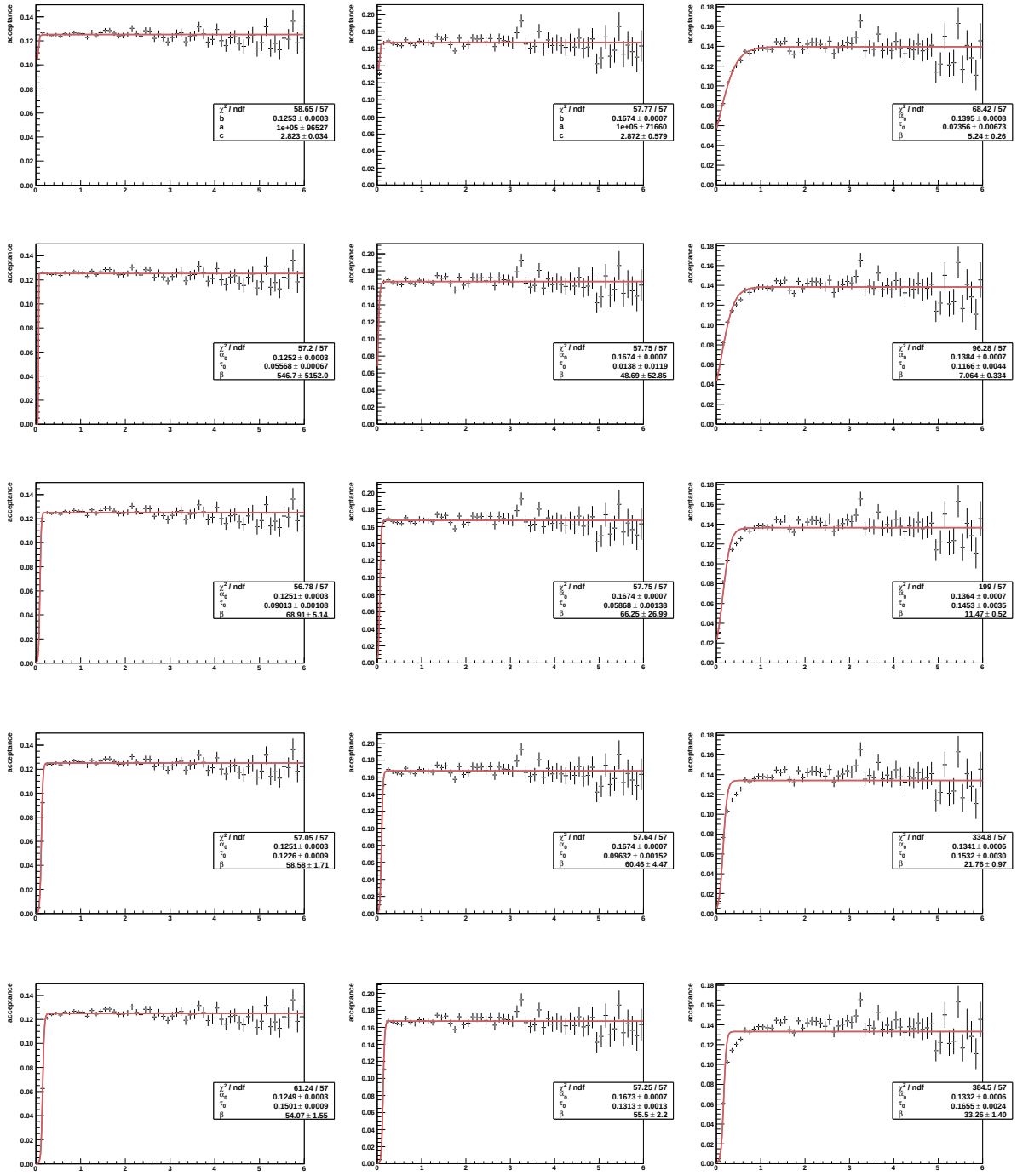


Figure B.6: Effect of the propertime significance cut on the selected B_s propertime distribution (measured in ps). The propertime significance cut is increased from top to bottom in set: $\{1\sigma,\ 2\sigma,\ 3\sigma,\ 4\sigma,\ 5\sigma\}$, and is shown before trigger (left), after the L0 decision (middle) and after both the L0 and HLT decision (right).

C

Background studies to $B_s \rightarrow J/\psi\phi$ events

C.1 Summary of $b\bar{b}$-inclusive data

- From the 6 candidates within the $\pm 50\ \mathrm{MeV/c^2}$ window:

  - Of the six candidates lying inside the tight mass window, one is considered as signal: $B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$, This event will not be counted as background.
  - Three candidates are considered as ghosts:
    $\bar{B}^0 \rightarrow (D^0 \rightarrow \omega(782)\pi^+K^-)\pi^+\mu^-\bar{\nu}_\mu$, with the misidentification of a μ^- with a K^-, the K^+ begin mis-identified as a P^+ and reconstructed via the partial $\Delta^+ \rightarrow \pi^0P^+$ decay, with additional ghosts coming from K^- and μ^+. It is evident this event can be classified as technical background: caused by ghost tracks and combinatorics. The event is also cloned with one other candidate, which falls above the signal region. This candidate is considered as background.
    The other two candidate under this category are also clones of each other. Both have the partial decay structure $b_1(1235)^- \rightarrow (\omega(782) \rightarrow \pi^0\pi^+\pi^-)\pi^-$, with the K^+ mis-identified as a π^+, and the K^- coming from the $\bar{K}^{*0}(892) \rightarrow \pi^+K^-$ partial decay. Both the μ^+ and μ^- are missing. Only one of these candidates will be considered as background.
  - One candidate is consider to have at least one final state originating from a primary vertex:
    $\bar{\lambda}_b^0 \rightarrow (D_s^+ \rightarrow \nu_\mu\mu^+)\bar{\lambda}_c^-$, with the K^+ being mis-identified as a π^+ coming from a partial reconstructed $b_1^+(1235) \rightarrow \pi^+(\omega(782) \rightarrow \pi^0\pi^+\pi^-)$ decay. Both the μ^- and K^- are missing. This candidate is also cloned with one other, lying above the signal region. The candidate will therefore be counted as background.
  - One candidate is considered to have originated from a badly reconstructed primary vertex, which will be counted as combinatorial background.

  From these 6 candidates, a total of 4 will be considered.

- Eleven of the reconstructed candidates lie below the $\pm 50\ \mathrm{MeV/c^2}$ mass window:

- One event is considered as low lowing background.
  This has been reconstructed as $B_s^0 \rightarrow J/\psi(\mu^+\mu^-,\gamma)\phi(K^+K^-)$, with it mass underestimated due to the radiative γ. This candidate will not be considered.
- Five candidates are considered as ghosts. With one or two of the final states missing, an example of this is:
  $B_u^+ \rightarrow (K_2^*(1430)^+ \rightarrow \pi^0 K^+)(J/\psi(1S) \rightarrow \mu^-\mu^+,\gamma,\gamma,\gamma)$, with an missing K^-.
- Four events are considered to have final state emanating from a primary vertex:
  A partially reconstructed $B_u^+ \rightarrow \bar{D}^{*0}(2010)\mu^+\nu_\mu$ with the μ^+ coming from the B_u^+, a partially reconstructed $\bar{B}_d^0 \rightarrow D_s^{*+}\mu^-\nu_{\bar{\mu}}$ with the μ^- coming from the $\bar{B}_s^0$, a partially reconstructed $K_1^-(1270) \rightarrow (\bar{K}^{*0}(892) \rightarrow \pi^+K^-)\pi^-$ with the K^- coming from the $\bar{K}^{*0}(892)$, and where the $K^+$is missing.
- One of the reconstructed decay is categorised as pile-up:
  Made up of the partially reconstructed $\rho(770)^+ \rightarrow \pi^0\pi^+$, with the μ^+ begin mis-identified as an π^+ and a $\rho^0(777) \rightarrow \pi^-\pi^+$, with the K^+ being mis-identified as an π^+.

From these 11 candidates a total of 10 will be considered.

- 16 reconstructed candidates lie above the ± 50 MeV/c^2 mass window:
  - Four events are consider as ghost:
    $\bar{\lambda}_b^0 \rightarrow K_1(1270)^+(J/\psi(S1) \rightarrow \mu^+\mu^-)\bar{P}$ and a partially reconstructed $\phi(1020) \rightarrow K^+$ and K^-, with the μ^+ missing.
    Two candidates under this category, made up of the partial decays $B^- \rightarrow D^*\mu^-\bar{\nu}_{mu}$ and $K^{*0}(1430)^- \rightarrow \pi^0K^-$ are cloned with each other. Only one of these clones will be considered.
  - Ten are consider to have more than one final state coming from a primary vertex, for instance:
    $\bar{B}_d^0 \rightarrow (D^0 \rightarrow \omega(782)\pi^+K^-)\pi^+\mu^-\nu_{\bar{\mu}}$, with K^+ mis-identified for a P originating from $\Delta^{++} \rightarrow \pi^+P$ decay and the K^- mis-identified for the $\bar{P}$ coming from a $\bar{\Delta}^0 \rightarrow \pi^+\bar{P}$ decay. With one of these candidates cloned with one other lying under the tight signal window.
  - One event is categorised as coming from pile-up:
    $\bar{B}_d^0 \rightarrow D^*(2010)^+\mu^-\bar{\nu}_\mu$, with the μ^- originating from the $\bar{B}_d^0$ together with the partial decay $\phi(1020) \rightarrow K^+K^-$, with the μ^+ being mis-identified as a K^+.
  - One candidate is considered to originate from a badly reconstructed primary: $a_2(1320) \rightarrow (\rho(770)^- \rightarrow \mu^-\pi^-)\pi^0$, with the μ^- mis identified with the K^+ and the μ^- with the π^-. With the K^+,K^- coming from a prompt $\phi(1020)$. This candidate is consider as combinatorial.

  In all a total of 14 candidates will be considered.

C.2 Summary of J/ψ-inclusive data

- 199 events within the $\pm 50\ \mathrm{MeV/c^2}$ window:
  - 66 candidates are considered as signal $B_s \rightarrow J/\psi\phi$ events. These will not be included as background.
  - 2 events are considered as reflections of the form: $\bar{B^0} \rightarrow J/\psi(1S)(\rightarrow \mu^+\mu^-\gamma,\gamma)\bar{K^{*0}}(892) \rightarrow \pi^+K^-$, with the K^+ mis-identified as the π^+. These physical background events caused by the mis-identification of one final state will not be considered.
  - 8 candidates occur from partially reconstructed primaries: $B^0 \rightarrow K_2^{*0}(1430)(\rightarrow \pi^-K^+\gamma)J/\psi(1S)(\rightarrow \mu^+\mu^-)$, with the K^+ mis-identified as a μ^+ and a K^- mis-identified as a π^-. These events will systematically have low mass and will not be considered as combinatorial background. Of the 8 candidate 6 are found to be clones of each other.
  - 1 candidate is characterised as low mass background: $\bar{B_s^0} \rightarrow J/\psi(1S)(\rightarrow \mu^+,\mu^-,\gamma,\gamma)\phi(1210)(\rightarrow K^+K^-)$. Events of this kind will not be considered as combinatorial background.
  - 28 candidates are considered as ghosts, for instance: $J/\psi(1S) \rightarrow \mu^+,\mu^-,\gamma,\gamma$ and $K_1(1270)^+ \rightarrow K^+,\omega(782)$ with the K^- missing. A total of 8 candidates in this type where found clones of each other.
  - 13 candidates are considered to have final states coming from a primary, for instance: $B^+ \rightarrow J/\psi(1S)(\rightarrow \mu^+\mu^-)K^*(892)^+$, with the K^+ and K^- missing. Of these, 9 where found clones of each other.
  - 73 are considered event where a final state has originated from the same primary as the candidate: $J/\psi(1S) \rightarrow \mu^+,\mu^-$ and $\phi(1210) \rightarrow K^+,K^-$. Of these, 17 are found clones of each other.
  - 6 candidates where the a final state and candidate come from different primaries: $J/\psi(1S) \rightarrow \mu^+\mu^-$ and $K^{*0}(892) \rightarrow \pi^-K^+$ and $K^*(892)^- \rightarrow \pi^-\bar{K^0}$, with the K^- mis-identified as an π^-. Two of these candidates are found to be clones.
  - 2 are considered as $b\bar{b}$ background: $B^+ \rightarrow J/\psi(1S)(\rightarrow \mu^+,\mu^-)K^*(892)^+\bar{K^0}K^0$ and $\bar{B^0} \rightarrow D^*(2010)^-(\rightarrow \pi^-\bar{D^0}(\rightarrow \pi^0\pi^-K^+))\eta K(892)^+D^0(\rightarrow \eta\bar{K^{*0}}(892)(\rightarrow \pi^+K^-))$

  A total of 99 candidates will be considered.

- 362 events are below the $\pm 50\ \mathrm{MeV/c^2}$ mass window:
  - 16 candidates are considered as originating from partially reconstructed primaries, for example:

$\bar{B^0} \to \bar{K^0} J/\psi(1S)(\to \mu^+\mu^-) f_2'(1525)(\to K^+K^-)$, with the K^+ miss identified as a μ^+. Of these candidates, 7 are considered clones of each other.

- 16 candidates are considered as low mass background, these events will not be considered for combinatorial background.
  Events of this type include, $B_s \to J/\psi(1S)(\to \mu^+\mu^-)\phi(1020)(\to K^+K^-)\gamma$.

- 58 candidates are considered as ghosts, for instance:
  $B^+ \to K^{*+}(892)(\to \pi^0 K^+) J/\psi(1S)(\to \mu^+\mu^-)$ were the K^- is missing. From these candidates, 24 are considered as clones.

- 24 candidates have at least one final state originating from a primary vertex, for example:
  $\lambda_{B_0} \to \nu^0 J/\psi(1S)(\to \mu^+\mu^-)\bar{K^{*0}}(892)$ and $\bar{\Delta} \to \pi^0 \bar{P}$, with the K^+ miss identified as a μ^+ and the K^- as a $\overline{P}$. Of these candidates, 7 are clones of each other.

- 235 candidates have originate from same primary vertex, for example:
  $B^+ \to K^{*+}(892)(\to \pi^+ K^-\gamma) J/\psi(1S)(\to \mu^+\mu^-)\phi(1020)(\to K^+K^-)$, with the K^+ miss identified as a π^+. Out of which 25 candidates are clones.

- 9 from different primary vertices:
  $J/\psi(1S)(\to \mu^+\mu^-)$, $K_1^-(1270)(\to K^-\omega(782))$ and $K^{*0}(892)(\to \pi^- K^+)$. In this example, each of the final state particles are correctly identified, but originate from different decay fragments. out of the 9 candidates 2 are clones.

- 4 candidates are consider as $b\bar{b}$ background, this includes for example:
  $\bar{B_s}(\to \pi^+\rho^-(770)\rho^+(770)D_s^-(\to \pi\phi(1020)(\to K^+K^-)))$, and $\lambda_{\bar{B_0}} \to \bar{\rho^-} J/\psi(1S)(\to \mu^+\mu^-)K^{*+}(892)$.

A total of 278 candidates will be considered.

- 378 events are above the $\pm 50\ \mathrm{MeV/c^2}$ mass window:

  - 3 candidates are considered as partially reconstructed background. This includes for example:
    $B^- \to K_2^*(1430)(\to \pi^+K^-) J/\psi(1S)(\to \mu^+\mu^-)$, with the K^+ miss identified as a μ^+. None of the candidates there found to be clones.

  - 49 of the candidates where considered as ghost, including:
    $J/\psi \to \mu^+\mu^-$, with both the K^+ and K^- missing. Out of these candidates, 11 where found to be clones of each other.

  - 27 candidates are considered to have at least one final state originating from a primary vertex. Of these candidates 7 are considered clones.

  - 284 candidates have final states originating from the same primary vetices. Of these candidates 27 are clones.

- 11 candidates from different primary vertices, including:
  $B_s \rightarrow \pi^+K^-K^0J/\psi(1S)(\rightarrow \mu^+\mu^-\gamma)$ and $f_2'(1525) \rightarrow K^+K^-$, with the K^- miss identified as a μ^-. Of these candidates, no clones were found.
- 4 candidates are considered as $b\bar{b}$ background, for instance:
  $B^- \rightarrow \pi^-\pi^+K^-\eta J/\psi(1S)(\rightarrow \mu^+\mu^-)$ and $B^0 \rightarrow D^*(2010)(\rightarrow \pi^-\bar{D}^0(\rightarrow \omega(782)K^{*0}(892)(\rightarrow \pi^-K^+)))\pi^+\pi^0\eta'$, with the K^- miss identified as a π^-. No clones where found here.

A total of 343 candidates will be considered.

C.3 $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ background

In the $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ decay, it is physically possible for the K^{*0} to decay via:

$$K^{*0} \rightarrow K^+\pi^-, \quad \text{or} \quad K^{*0} \rightarrow K^0\pi^0.$$

Following the Isospin(I), and its z-component(I_z), through each decay. It is possible to deduce that the K^{*0} will decay $\frac{2}{3}$ of the time into $K^+\pi^-$ and $\frac{1}{3}$ of the time into $K^{*0} \rightarrow K^0\pi^0$. [1] However, in the GVTGEN $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ generated data sample3.2, the K^{*0} has been forced to decay 100% of the time into $K^+\pi^-$.

Use will be made of the following decay to calculate the $\frac{B}{S}$ ratio for this specific decay:

$$\left(\frac{B}{S}\right) = \frac{\mathrm{BF}(\bar{\mathrm{b}} \rightarrow \mathrm{B_{u,d,s}})}{\mathrm{BF}(\bar{\mathrm{b}} \rightarrow \mathrm{B_s^0})} \frac{\mathcal{BF}_{total}^{spec}}{\mathcal{BF}_{total}^{signal}} \times \frac{\epsilon_\theta^{\mathrm{spec}}}{\epsilon_\theta^{\mathrm{signal}}} \times \frac{\epsilon_{\mathrm{sel}}^{\mathrm{bkg}}}{\epsilon_{\mathrm{sel}}^{\mathrm{signal}}} \times \mathcal{F}_{mass}. \tag{C.1}$$

where:

-. $\epsilon_\theta^{\mathrm{spec}}$ is the specific background fiducial acceptance requirement of the detector.

-. $\mathcal{BF}(\bar{b} \rightarrow B_{u,d,s})$ is the hadronization fraction of $\bar{b}$ into the specific hadrons $B_{u,d,s}$); where, $\bar{b} \rightarrow B_u$ = 38.9 $\pm$ 1.3 % and $\bar{b} \rightarrow B_d$ = 38.9 $\pm$ 1.3 %.

-. $\mathcal{BF}_{total}^{spec}$ is the total branching fraction of the specific decay channel into its final state.

-. $\epsilon_{\mathrm{sel}}^{\mathrm{bkg}} = \frac{N_{sel}^{spec}}{N_{gen}^{spec}}$ is the efficiency for selecting background events, where N_{sel}^{spec} is the specific number of background events selected and N_{gen}^{spec} is the number of generated specific background events analysed.

Following equation C.1 for the $B_d \rightarrow J/\psi(\mu^+\mu^-)K^{\star 0}$ decay channel. Requiring the specific branching fraction to be:

[1] The K^0 will decay 50% of the time into a K_L and the other 50% into a K_S. However, due to the large boost of the LHCb detector and the long decay length of the K_L (c$\tau \sim$ 7m in the lab frame [7]) only decay into the K_s will be seen.

$$\begin{aligned}
\mathcal{BF}^{spec}_{B_d\to J/\psi(\mu^+\mu^-)K^{*0}} &= \mathcal{BF}(B_d \to J/\psi K^{*0}) \times \mathrm{BF(J/\psi \to \mu^+\mu^-, \gamma)} \times \mathrm{BF(K^{*0} \to K^+\pi^-)}, \\
&= (6.04 \pm 1.00) \times 10^{-5}, \\
\mathcal{BF}(B_d \to J/\psi K^{*0}) &= (1.33 \pm 0.06) \times 10^{-3}, \\
\mathrm{BF(K^{*0} \to K^+\pi^-)} &= \frac{2}{3}, \\
\mathcal{BF}(J/\psi \to \mu^+\mu^-, \gamma) &= \mathcal{BF}(J/\psi \to \mu^+\mu^-) + \mathcal{BF}(J/\psi \to \mu^+\mu^-\gamma) = (6.81 \pm 1.10) \times 10^{-2}
\end{aligned} \tag{C.2}$$

leading to the specific branching fraction of $(5.99 \pm 1.0)\times 10^{-5}$ [2]

- 1163 events are found within the $\pm 50\ \mathrm{MeV/c^2}$ window:
  - 236 events are considered as ghost events: $B^0 \to K^*(892)(\to \pi^+K^-)J/\psi(1S)(\to \mu^+\mu^-)$ with the K^+ missing in this instance. Or $B^0 \to a_2^0(1320)(\to \pi^+\rho^-(770)(\to \pi^0\pi^-))$ with the K^- mis-identified as a π^- and the K^+ missing.
  - 57 of the events are consider as being partially reconstructed, for instance $B^0 \to K^{*0}(892)(\to \pi^+K^-)J/\psi(1S)(\to \mu^+\mu^-\gamma)$ with the K^+ mis-identified as a π^+.
  - 1 events is consider as a low mass background:
  - 125 events are considered as reflections: $\tilde{B} \to K^{*0}(892)(\to \pi^+K^-\gamma)J/\psi(1S)(\to \mu^-\mu^+)$ with the K^+ identified as a π^+.
  - 690 events originate from a Primary Vertex. An instance of this is the partially reconstructed decay: $B^0 \to K^*(892)J/\psi(1S)(\to \mu^+\mu^-)$ and $\phi \to K^-K^+$.
  - 2 event have final state originating from the same Primary Vertex, this includes for example: $K_1^-(1400) \to \pi^-\tilde{K}^*(820)(\to \pi^+K^-)$ with the K^+ identified as a π^+ and both the muons are missing.
  - 17 events come from different Primary Vertices: $B^0 \to K^*(892)(\to \pi^-K^+)J/\psi(1S) \to \mu^+\mu^-$ and $\omega(782) \to \pi^0\pi^+\pi^-\gamma$ with the K^- identified as a π^-.
  - 35 events are regarded as coming from $b\bar{b}$ background, including $B^0 \to K^*(892)J/\psi(1S)(\to \mu^+\mu^-)$ and $\tilde{B^0} \to \pi^-\pi^+D^+(\to \pi^+\phi(1210)(\to K^+K^-))D_s^{*-}(\to \gamma D_s^-(\to \bar{\nu}_\mu\mu^-\phi(1020)(\to K^+K^-)))$
- 1575 events are found below the $\pm 50\ \mathrm{MeV/c^2}$ mass window:
  - 301 events are considered as ghosts, for example: $B^0 \to K^{*0}(892)(\to \pi^-K^+)J/\psi(1S)(\to \mu^+\mu^-)$, where the K^- is missing.
  - 73 of the events are consider as being partially reconstructed, for instance:

[2] without radiative corrections this becomes 5.1 ± 10^{-5} which is the value quoted in the EVTGEN decay file.

- 165 events are considered as reflections, including:
  $\bar{B^0} \to K^{*0}(892)(\to \pi^- K^+)J/\psi(1S)(\to \mu^+\mu^-\gamma)$, with the K^- miss identified as a π^-.
- 952 events are considered with final states originating from a primary vertex, for instance:
  $B^0 \to K^{*0}(892)J/\psi(1S)(\to \mu^+\mu^-)$, where the K^+ and K^- are missing.
- 1 event is considered as have final state originating from the same primary vertex, this includes:
- 24 events are regarded as having final states originating from different primary vertices, for example:
  $\bar{K}^{*0}(892) \to \gamma\pi^+K^-$, $K_1^+(1270) \to \pi^-\pi^+K^+$ and $\rho^+(770) \to \pi^0\pi^+$, where the μ^+, μ^- and K^+ are miss identified as a π^+, K^+ and π^+ respectively.
- 59 events are regarded as $b\bar{b}$ background, including:
  $\bar{B}^0 \to K^{*0}(892)J/\psi(1S)(\to \mu^+\mu^-\gamma)$ and $B_s \to \eta\rho^+(770)\pi^0\pi^0 D_s^-(\to \phi(1020)(\to K^+K^-)e^-\mu_{e^-}\gamma\gamma)$.

- 1385 events are found above the $\pm 50\ \mathrm{MeV/c^2}$ mass window:
  - 258 events are considered as ghost, including:
    $B^0 \to J/\psi(1S)(\to \mu^+\mu^-\gamma)K^{*0}(892)(\pi^- K^+)$, with the K^- missing.
  - 72 of the events are considered as partially reconstructed, for instance:
    $B^0 \to J/\psi(1S)(\to \mu^+\mu^-\gamma\gamma)K^{*0}(892) \to \pi^- K^+$, where the K^- miss identified as a μ^-.
  - 127 events are considered as reflections, including:
  - 845 events where at lest one of the final states originate from a primary vertex. An instance of this is the partially reconstructed decay:
    $barB^0 \to J/\psi(1S)(\to \mu^+\mu^-)K^{*0}(892)$ and $a_2^0(1320) \to \pi^-\pi^+\omega(782)$, with the K^+ miss identified with π^+ and the K^- is missing.
  - 2 event are considered as having final state originating from the same primary vertex, this includes for example:
  - 24 events are considered as having final states coming from different primary vertices, including:
    $K^{*0}(892) \to \pi^- K^+$ and $\phi(1020) \to K^+K^-$, with the μ^- missing and the μ^+ miss identified as a K^+.
  - 57 events are regarded as $b\bar{b}$ background, including:
    $B^0 \to J/\psi(1S)(\mu^+\mu^-)K^{*0}(892)$ and $B^- \to \omega(782)(\to \pi^+\pi^-)\rho^-(770)\rho^-(770)\bar{K}^0 D^+$, with the K^+ miss identified as a μ^+ and the K^- as a π^-.

$B_d \to J/\psi(\mu^+\mu^-)K^{\star 0}$ Selection Summary

- from 3M events 4654 candidates where reconstructed within the wide mass window.
  - Reconstructed candidates:
  - $\sim$ 19% of these events were classified as ghosts.
  - $\sim$ 60% of the events have final states originating from a primary vertex.
  - $\sim$ 10% of these events originate from reflections.
  - $\sim$ 5% of these events occur as reflections.
  - $J/\psi(1S)$ and $\phi(1210)$ reconstructed candidates:
  - $\sim$ 94% of the events contain a $J/\psi(1S)$ coming from a B_d
  - $\sim$ 6% of the events contained no $J/\psi(1S)$'s.
  - $\sim$ 1% of the events contain a true $\phi(1210)$.
  - $\sim$ 6% of the events contain a prompt $\phi(1210)$.
  - $\sim$ 92% of the events did not contain a $\phi(1210)$.
  - Final states:
  - $\sim$ 36% of final state K^- were identified as K^-, $\sim$ 25% are mis-identified as a π^-, $\sim$ 4% as a μ^-, $\sim$ 5% as a $\bar{P}$ while $\sim$ 28% were not assocciated.
  - $\sim$ 36% of the final state K^+ were identified as K^+, $\sim$ 24% are mis-identified as a π^+, $\sim$ 4% as a μ^p, $\sim$ 5% as a P and 28% were not assocciated.
  - $\sim$ 97% of the final state μ^- were identified as μ^-, 1% as a π^- and $\sim$ 2% were not assocciated.
  - $\sim$ 97% of the final state μ^+ were identified as μ^+, $\sim$ 1% as a π^+ and $\sim$ 2% have no assocciation.

The specific background selection efficiency, assuming only the reconstructed events lying within the broad mass window, is then $\epsilon_{\text{sel}}^{\text{bkg}} \sim (38.2 \pm 1.13) \times 10^{-5}$; were the errors are statistical. This leads to a specific background to signal ratio in the mass window of:

$$\left(\frac{B}{S}\right)^{B_d \to J/\psi(\mu^+\mu^-)K^{\star 0}}_{B_s \to J/\psi\phi} = \underbrace{(3.5 \pm 1.4) \times 10^{-3}}_{\text{sig. win}}. \tag{C.3}$$

Including all events lying within the wide mass window this ratio becomes:

$$\left(\frac{B}{S}\right)^{B_d \to J/\psi(\mu^+\mu^-)K^{\star 0}}_{B_s \to J/\psi\phi} = (9.6 \pm 3.8) \times 10^{-3} \tag{C.4}$$

This ratio is small enough that we can neglect the specific background contribution coming from $B_d \to J/\psi(\mu^+\mu^-)K^{\star 0}$ decays.

D

Sensitivity studies with $B_s \to J/\psi\phi$

D.1 Results

The sensitivities are quoted as the RMS values (or the standard deviation) read from the distributions of the fitted values. For each study we generated > 500 Toy Monte Carlo experiments. For those parameters let free during the fitting stage, we plotted the fitted value distribution, the error distribution and the pull distribution. Figure D.1 summarises the results obtained from column (a) in Table 6.12. Figures D.2 to D.7 summarise the results obtained for each study done for the one-angle and three-angle analysis and for the two physics parameters of interest $\Delta\Gamma_s$ and $-2\beta_s$. We remind the reader that where the symbol ϕ_s appears in Figures D.2 to D.7 and following Figures this is identically equal to $-2\beta_s$

In addition, we summarise in Table D.1 the results obtained from column (a) and (b) given in Table 6.12. We also summarise in Tables D.2 to D.5 the results obtained in all cases and for all the parameters we fit for. First and second columns show the study performed and the parameters we fit for. On the third column we have the RMS of the fitted value distribution and next to it its corresponding error. The fifth column contains the mean value of the fitted error distribution (as returned by MINUIT), followed by the scaled mean, i.e. the mean error multiplied by the width of the pull distribution, and the corresponding error on the mean error. Columns eight and nine show the mean and width of the pull distribution. We also indicate the number of experiments used for each study.

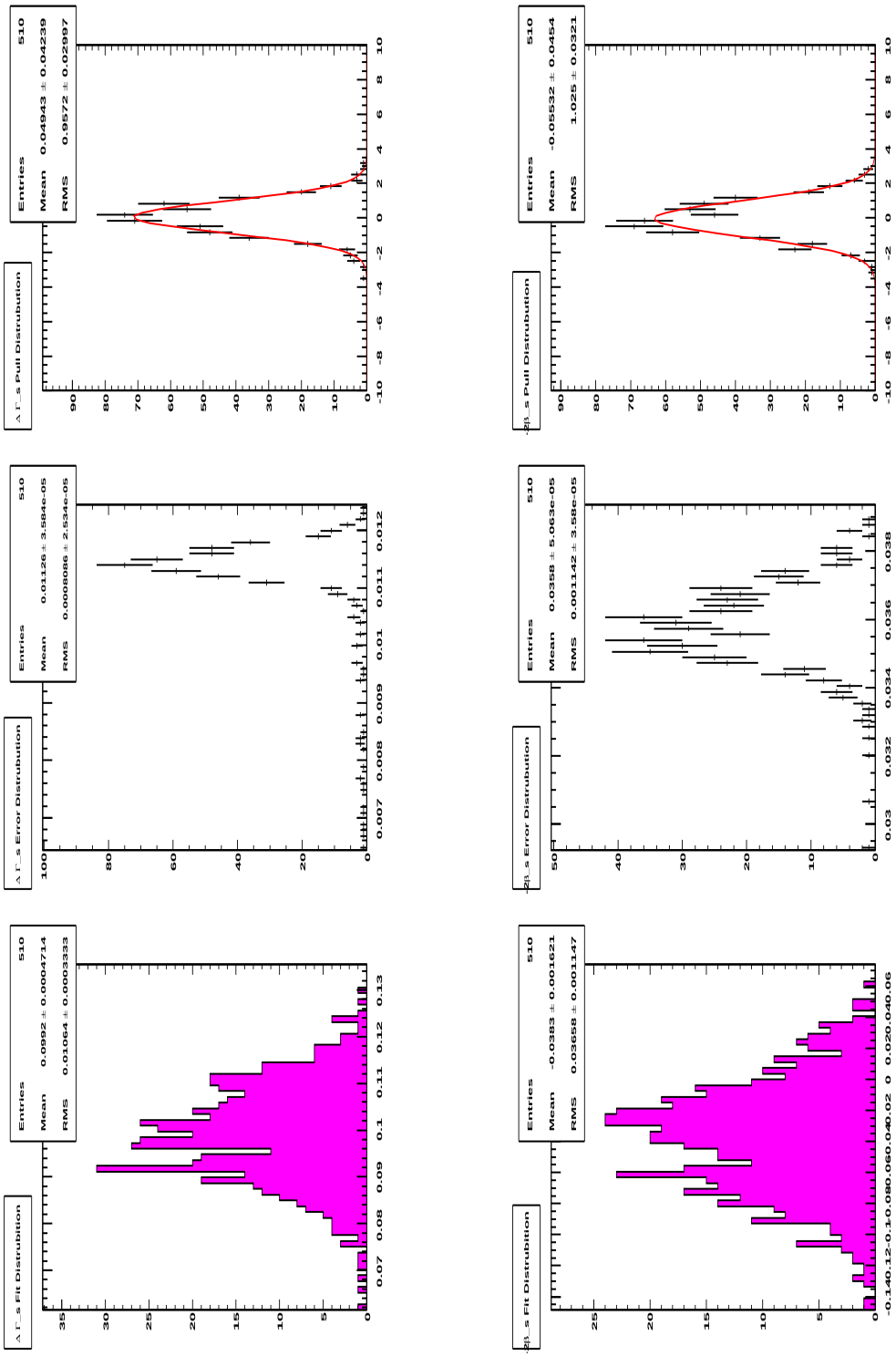


Figure D.1: Three-angle simultaneous studies (tagged data). $\Delta\Gamma_s$ and $-2\beta_s$ distributions

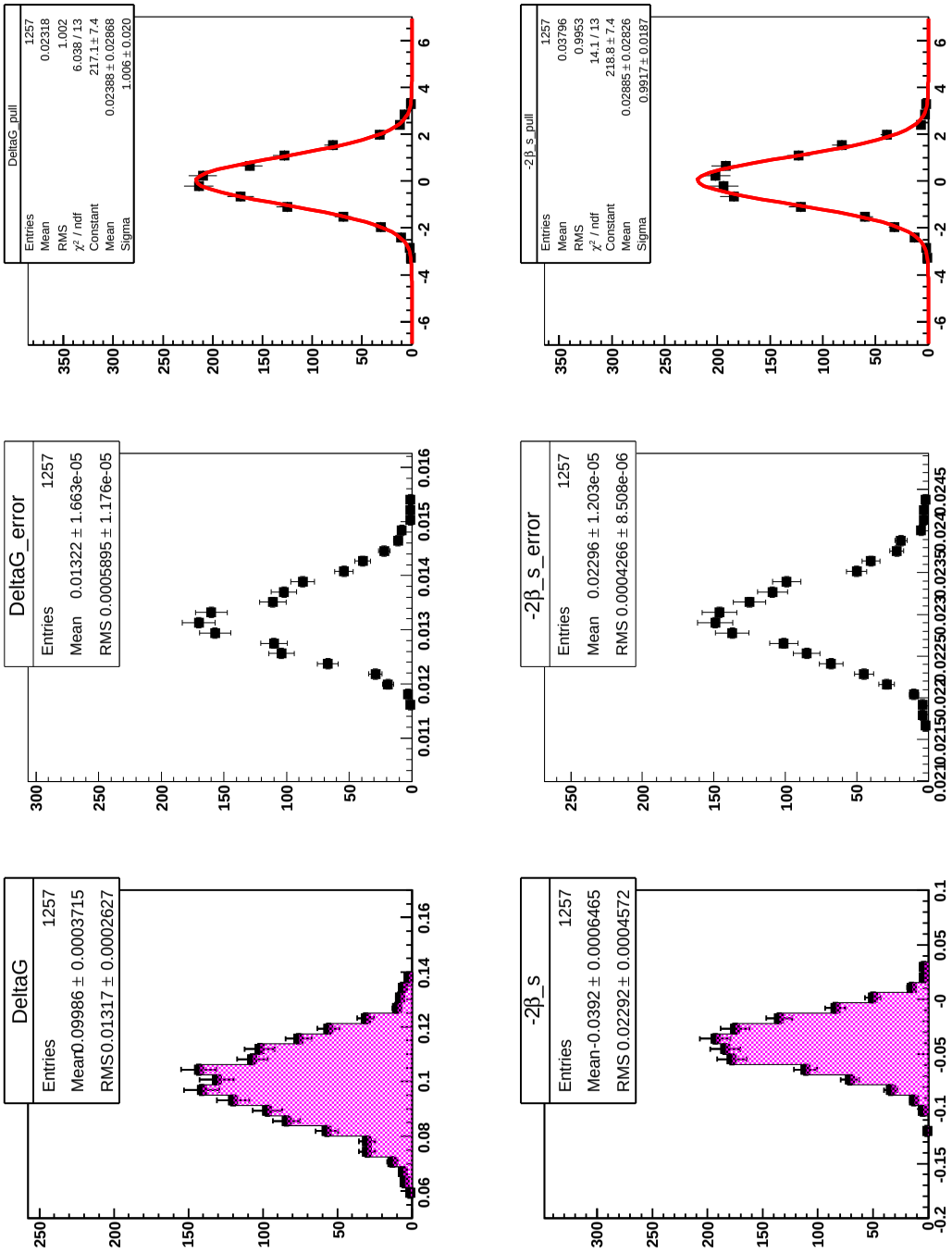


Figure D.2: Ideal: One-angle simultaneous studies (tagged data). $\Delta\Gamma_s$ and $-2\beta_s$ distributions

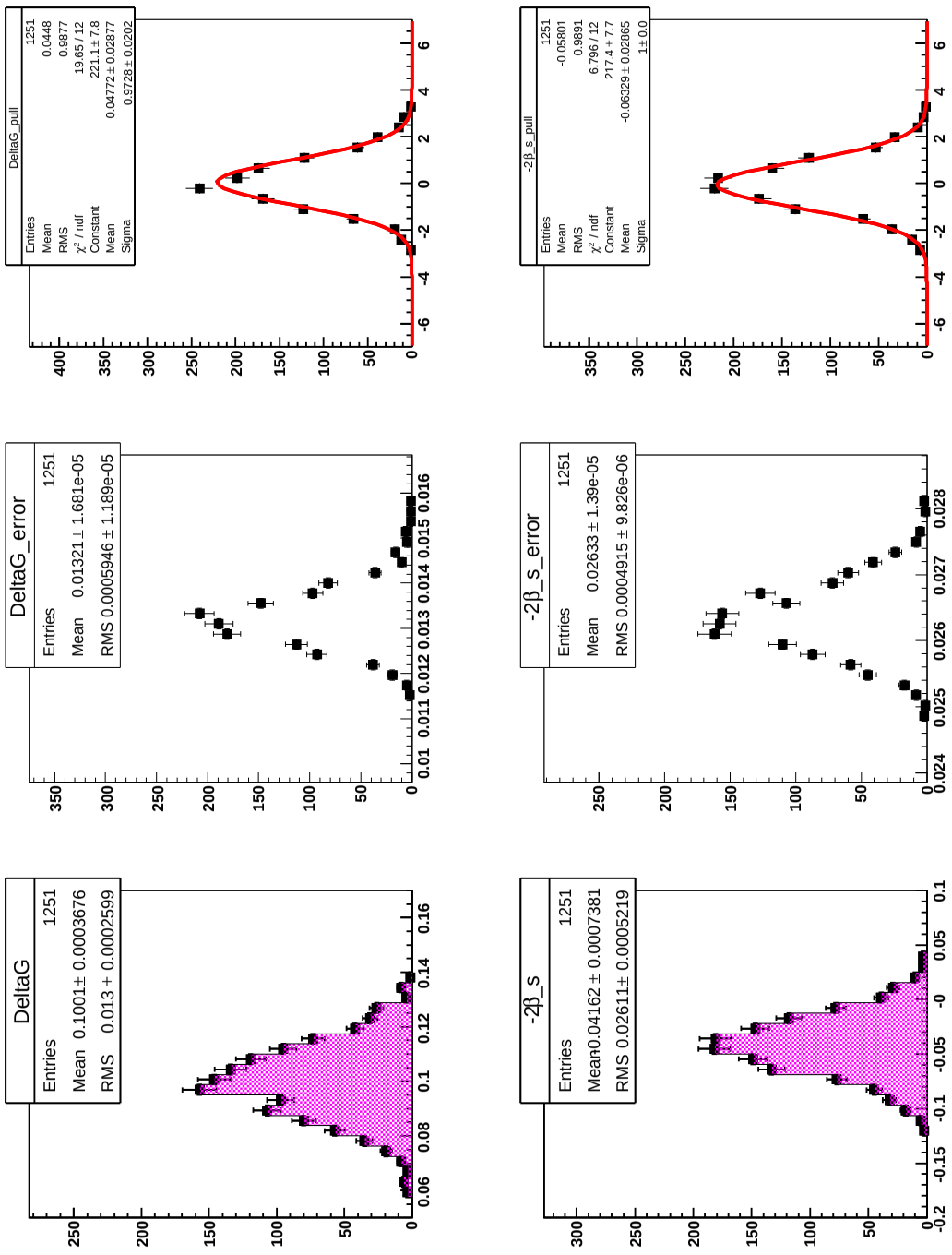


Figure D.3: Resolution: One-angle simultaneous studies (tagged data). $\Delta\Gamma_s$ and $-2\beta_s$ distributions

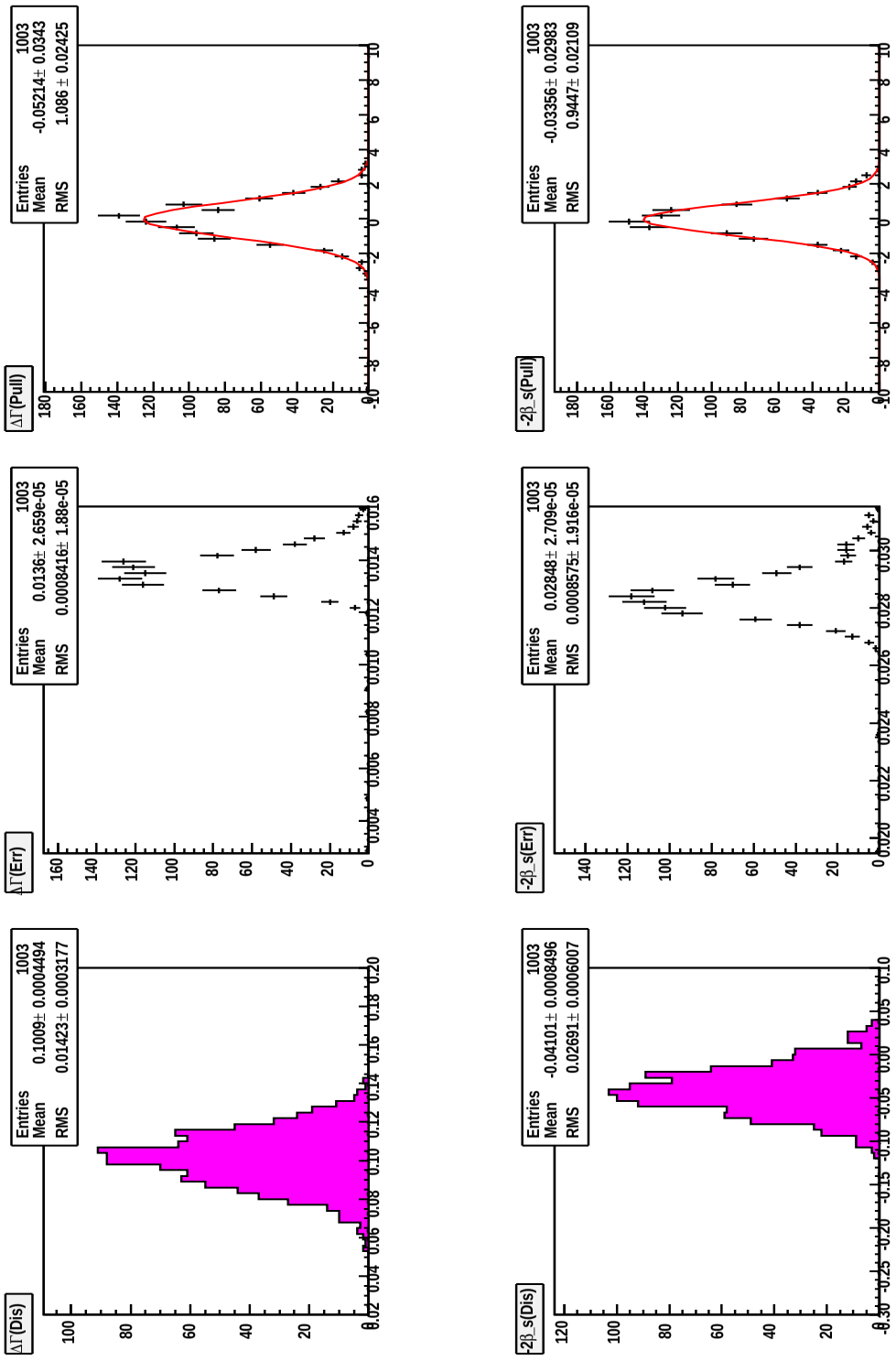


Figure D.4: Background: One-angle simultaneous studies (tagged data). $\Delta\Gamma_s$ and $-2\beta_s$ distributions

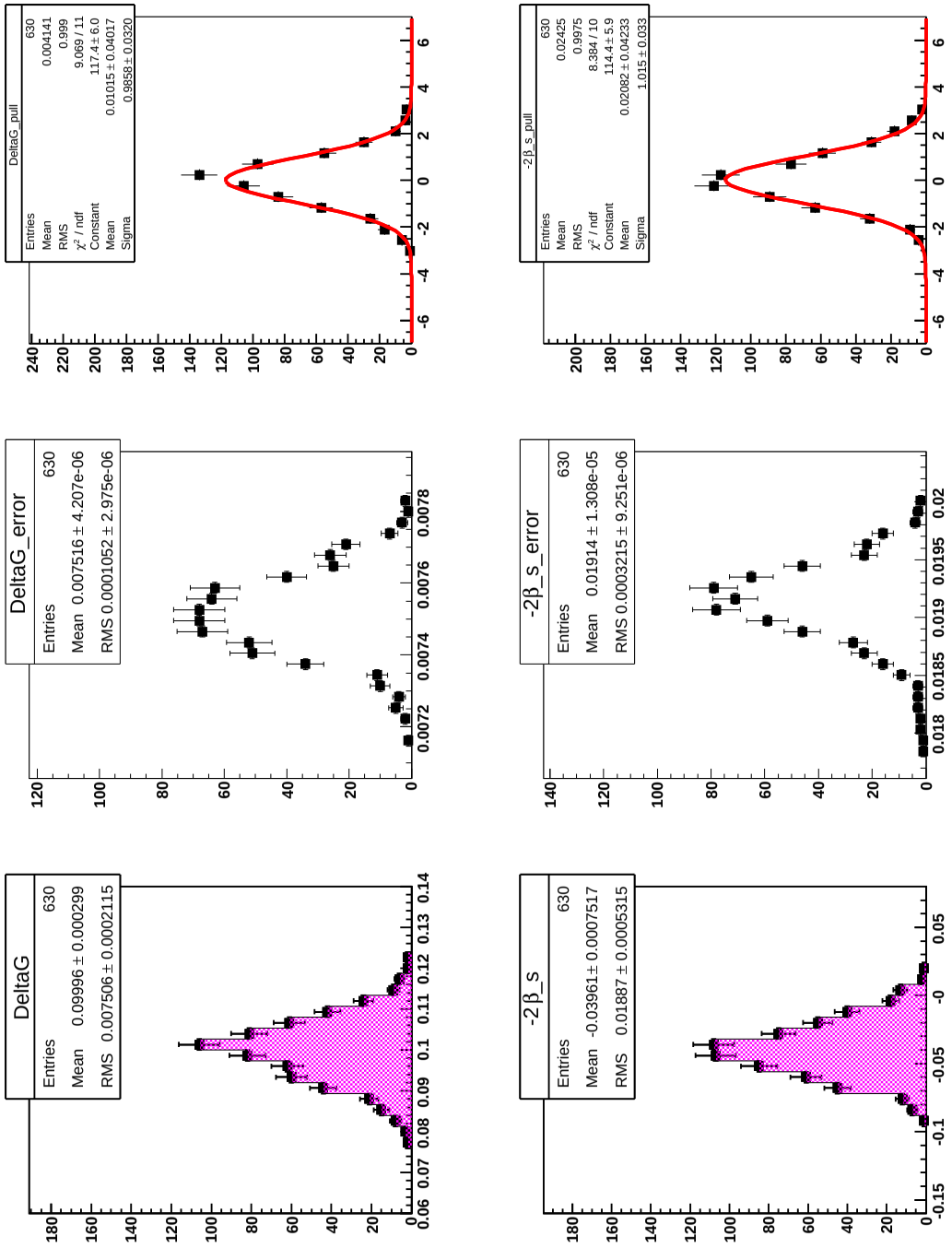


Figure D.5: Ideal: Three-angle simultaneous studies (tagged data). $\Delta\Gamma_s$ and $-2\beta_s$ distributions

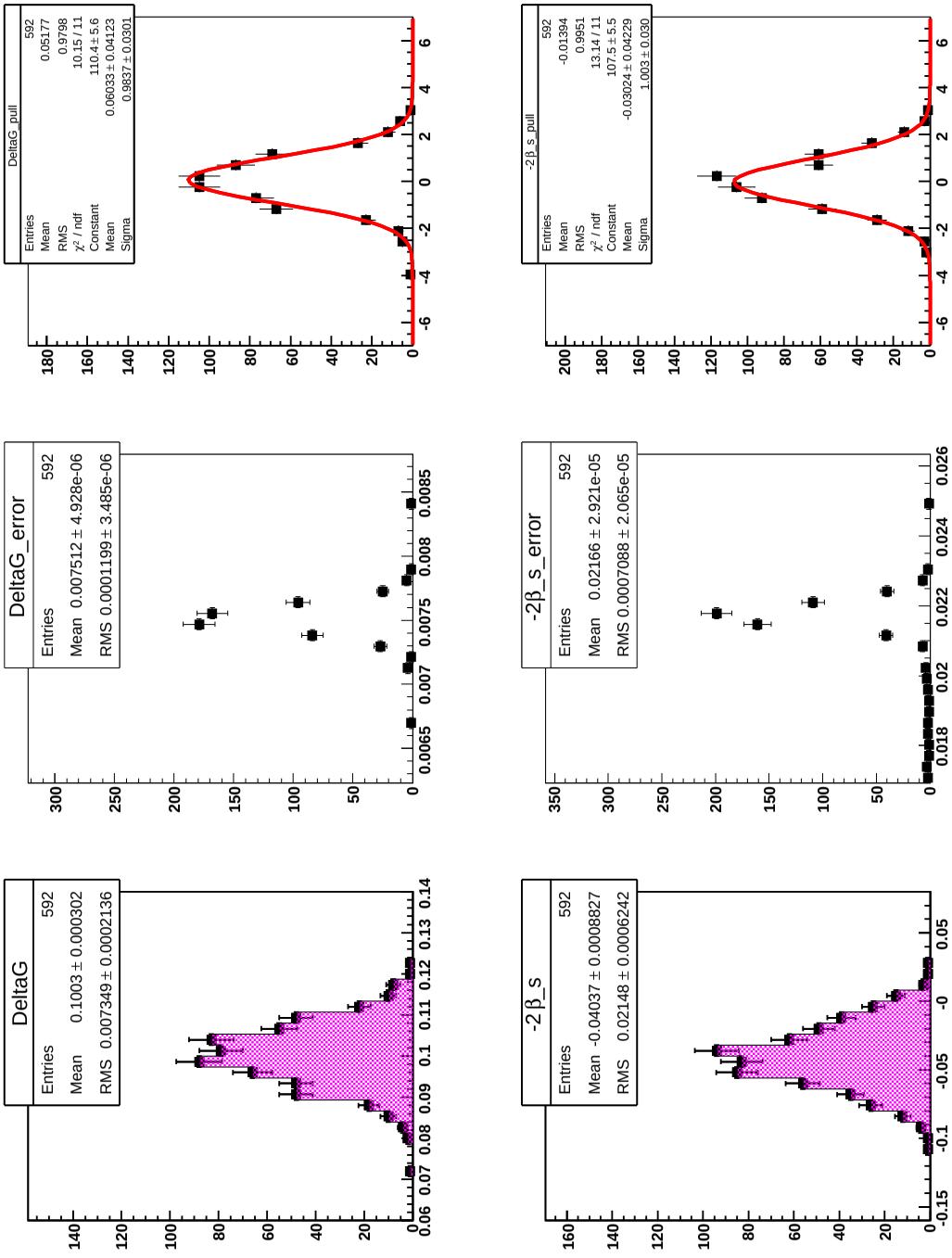


Figure D.6: Resolution: Three-angle simultaneous studies (tagged data). $\Delta\Gamma_s$ and $-2\beta_s$ distributions

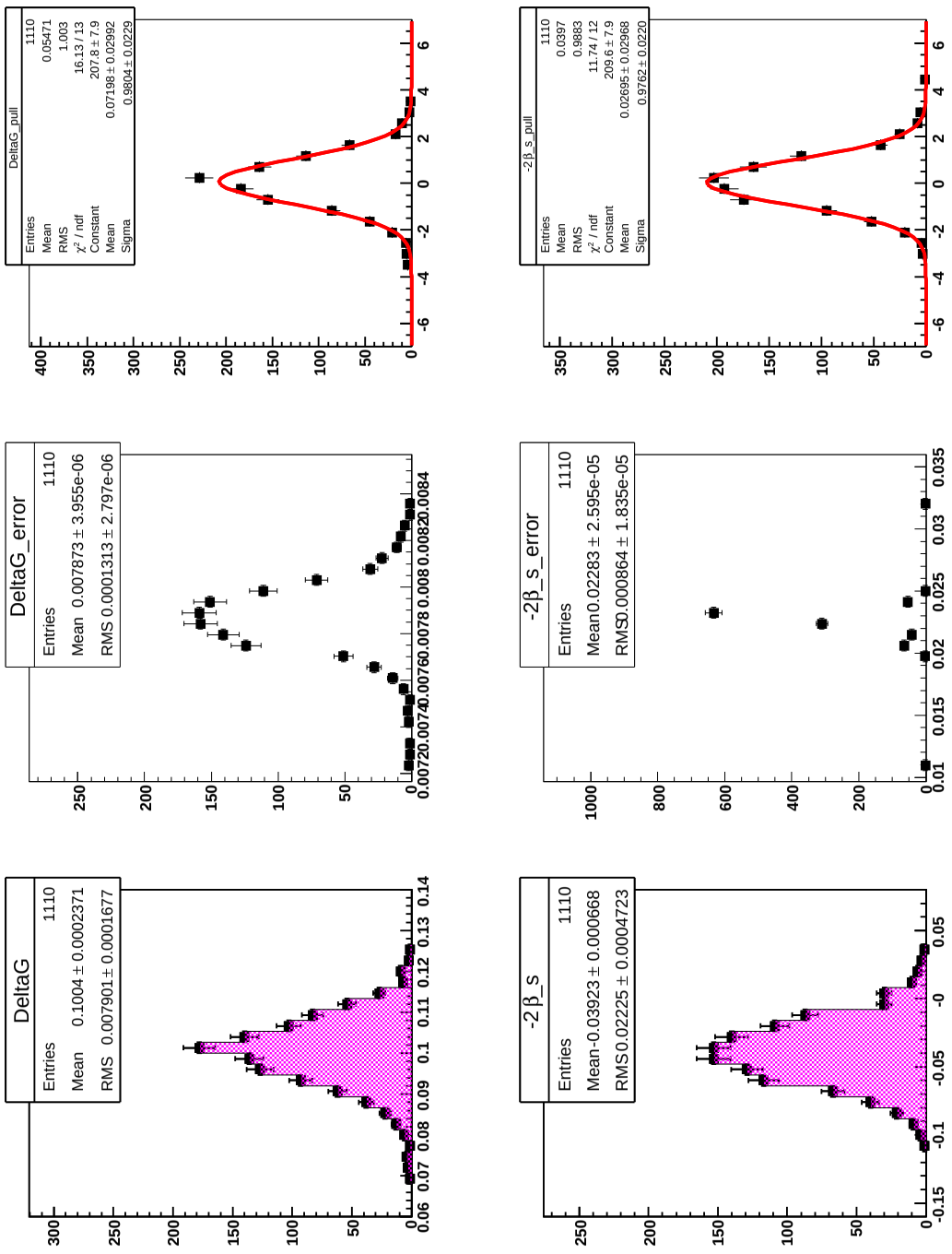


Figure D.7: Background: Three-angle simultaneous studies (tagged data). $\Delta\Gamma_s$ and $-2\beta_s$ distributions

Simultaneous fits							Pull	
Using Table 6.2	510 exps	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
	$\bar{\Gamma}_s$	0.00332	0.000104	0.00336	0.00333	0.000008	-0.096	0.991
	$\Delta\Gamma_s$	0.0106	0.00033	0.01126	0.0108	0.000036	0.0494	0.957
	$R_\perp$	0.00452	0.00014	0.00465	0.0045	0.00003	0.052	0.981
	R_0	0.0042	0.00013	0.0044	0.0042	0.00001	-0.049	0.96
	Δm_s	0.0502	0.00157	0.0553	0.051	0.00043	0.0173	0.926
	δ_1	0.01064	0.00033	0.0113	0.011	0.000034	0.049	0.957
	δ_2	0.123	0.00386	0.1368	0.145	0.0014	-0.066	1.06
	$-2\beta_s$	0.0366	0.0011	0.0358	0.0365	0.0000051	-0.0553	1.025
Using Table 6.11	556 exp							
	$\bar{\Gamma}_s$	0.00362	0.00011	0.00343	0.0035	0.0000065	0.024	1.025
	$\Delta\Gamma_s$	0.0117	0.00035	0.0116	0.0117	0.000025	-0.0306	1.011
	$R_\perp$	0.00471	0.00014	0.00476	0.00478	0.0000153	0.0411	1.004
	R_0	0.00439	0.00013	0.0045	0.0044	0.0000126	-0.0054	0.984
	Δm_s	0.0408	0.00122	0.0451	0.041	0.00025	0.081	0.916
	δ_1	0.0912	0.00274	0.1029	0.094	0.000544	0.0729	0.9128
	δ_2	0.107	0.0032	0.115	0.117	0.00087	0.0563	1.016
	$-2\beta_s$	0.0301	0.0009	0.0308	0.0304	0.0000369	-0.0019	0.99

Table D.1: Three-angle analysis: results of simulataneous fits with tagged data from Table 6.12.

Simultaneous fits							
Ideal	1336 exps					Pull	
	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
$\bar{\Gamma}_s$	0.00358	0.00007	0.00354	0.003555	0.000005	0.0345	1.0030
$\Delta\Gamma_s$	0.01356	0.00026	0.01324	0.013531	0.000017	-0.0033	1.0220
$R_\perp$	0.00514	0.00010	0.00524	0.005131	0.000003	0.0310	0.9790
Resolution	1808 exps					Pull	
	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
$\bar{\Gamma}_s$	0.00359	0.00006	0.00354	0.003599	0.000004	-0.0329	1.0160
$\Delta\Gamma_s$	0.01353	0.00023	0.01322	0.013577	0.000014	0.0488	1.0270
$R_\perp$	0.00523	0.00009	0.00524	0.005236	0.000003	-0.0435	1.0000
Background	715 exps					Pull	
	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
$\bar{\Gamma}_s$	0.00354	0.00009	0.00359	0.003548	0.00001	-0.0445	0.9877
$\Delta\Gamma_s$	0.01348	0.00036	0.01356	0.013529	0.00002	0.0287	0.9977
$R_\perp$	0.00570	0.00015	0.00578	0.005697	0.00001	-0.0337	0.9852

Table D.2: One-angle analysis: simultaneous fits with untagged data in all four studies

Simultaneous fits							
Ideal	1257 exps					Pull	
	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
$\bar{\Gamma}_s$	0.00355	0.00007	0.00354	0.003541	0.000005	0.0054	1.0010
$\Delta\Gamma_s$	0.01317	0.00026	0.01322	0.013246	0.000017	0.0232	1.0020
$R_\perp$	0.00515	0.00010	0.00524	0.005111	0.000002	0.0360	0.9759
$-2\beta_s$	0.02292	0.00046	0.02296	0.022852	0.000012	0.0379	0.9953
Resolution	1251 exps					Pull	
	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
$\bar{\Gamma}_s$	0.00349	0.00007	0.00354	0.003479	0.000005	-0.0118	0.9841
$\Delta\Gamma_s$	0.01300	0.00026	0.01321	0.013048	0.000012	0.0448	0.9877
$R_\perp$	0.00518	0.00011	0.00523	0.005173	0.000003	-0.0118	0.9886
$-2\beta_s$	0.02611	0.00052	0.02633	0.026043	0.000014	-0.0580	0.9891
Background	1003 exps					Pull	
	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
$\bar{\Gamma}_s$	0.00371	0.00008	0.00360	0.003789	0.00001	0.0368	1.0520
$\Delta\Gamma_s$	0.01423	0.00032	0.01360	0.014770	0.00003	-0.0521	1.0860
$R_\perp$	0.00595	0.00013	0.00581	0.005985	0.00001	0.0796	1.0300
$-2\beta_s$	0.02691	0.00060	0.02848	0.026905	0.00003	-0.0336	0.9447

Table D.3: One-angle analysis: simultaneous fits with tagged data in all four studies

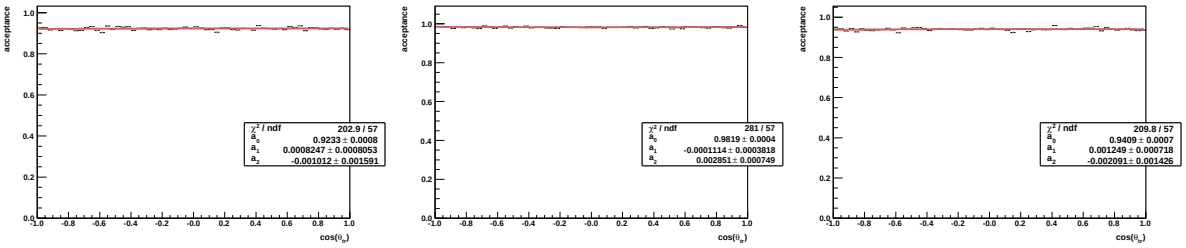
Simultaneous fits								
Ideal	866 exp						Pull	
		RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
	$\bar{\Gamma}_s$	0.002555	0.000061	0.002555	0.002552	0.000001	0.017	0.999
	$\Delta\Gamma_s$	0.007439	0.000179	0.007497	0.007454	0.000003	-0.010	0.994
	$R_\perp$	0.003044	0.000073	0.003007	0.003043	0.000001	-0.019	1.012
	R_0	0.002394	0.000058	0.002358	0.002393	0.000000	0.021	1.015
Resolution	589 exp						Pull	
		RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
	$\bar{\Gamma}_s$	0.002714	0.000079	0.002552	0.002710	0.000002	-0.082	1.062
	$\Delta\Gamma_s$	0.007491	0.000218	0.007492	0.007507	0.000004	0.043	1.002
	$R_\perp$	0.002978	0.000087	0.003008	0.002978	0.000001	0.000	0.990
	R_0	0.002316	0.000076	0.002358	0.002315	0.000054	0.040	0.982
Background	1653 exp						Pull	
		RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
	$\bar{\Gamma}_s$	0.002778	0.000048	0.002693	0.002779	0.000001	0.020	1.032
	$\Delta\Gamma_s$	0.008137	0.000142	0.007875	0.008135	0.000003	0.003	1.033
	$R_\perp$	0.003498	0.000061	0.003468	0.003499	0.000000	0.010	1.009
	R_0	0.002636	0.000046	0.002649	0.002635	0.000000	-0.031	0.995

Table D.4: Three-angle analysis: simulataneous fits with untagged data in all four studies

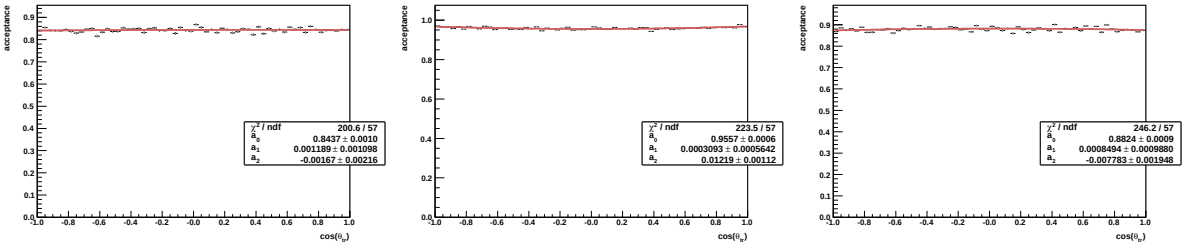
Simultaneous fits							Pull	
Ideal	630 exps	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
	$\bar{\Gamma}_s$	0.002484	0.000070	0.002567	0.002483	0.000002	-0.030	0.967
	$\Delta\Gamma_s$	0.007506	0.000212	0.007516	0.007508	0.000004	0.004	0.999
	$R_\perp$	0.003167	0.000089	0.003083	0.003172	0.000003	-0.140	1.029
	R_0	0.002479	0.000070	0.002353	0.002482	0.000001	0.014	1.055
	δ_1	0.071050	0.002002	0.065100	0.074149	0.000348	-0.104	1.139
	δ_2	0.064230	0.001810	0.061640	0.066633	0.000288	0.006	1.081
	$-2\beta_s$	0.018870	0.000530	0.019140	0.019092	0.000013	0.024	0.998
Resolution	592 exps	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
	$\bar{\Gamma}_s$	0.002464	0.000072	0.002560	0.002461	0.000002	-0.085	0.961
	$\Delta\Gamma_s$	0.007349	0.000214	0.007512	0.007360	0.000005	0.052	0.980
	$R_\perp$	0.003077	0.000089	0.003094	0.003086	0.000004	-0.161	0.997
	R_0	0.002429	0.000071	0.002351	0.002431	0.000001	0.016	1.034
	δ_1	0.079550	0.002312	0.070530	0.083084	0.000308	0.039	1.178
	δ_2	0.071110	0.002067	0.067390	0.072983	0.000268	0.014	1.083
	$-2\beta_s$	0.021480	0.000624	0.021660	0.021554	0.000029	-0.014	0.995
Background	1110 exps	RMS	RMS error	Mean	Scaled mean	Mean error	Mean	Width
	$\bar{\Gamma}_s$	0.002701	0.000057	0.002697	0.002705	0.000001	-0.021	1.003
	$\Delta\Gamma_s$	0.007901	0.000168	0.007873	0.007897	0.000004	0.055	1.003
	$R_\perp$	0.003507	0.000074	0.003551	0.003508	0.000003	-0.143	0.988
	R_0	0.002666	0.000057	0.002642	0.002668	0.000001	0.012	1.010
	δ_1	0.082620	0.001753	0.077000	0.086548	0.000249	-0.030	1.124
	δ_2	0.074720	0.001586	0.072810	0.076742	0.000197	0.012	1.054
	$-2\beta_s$	0.022250	0.000472	0.002283	0.002256	0.000026	0.040	0.988

Table D.5: Three-angle analysis: simulataneous fits with tagged data in all four studies

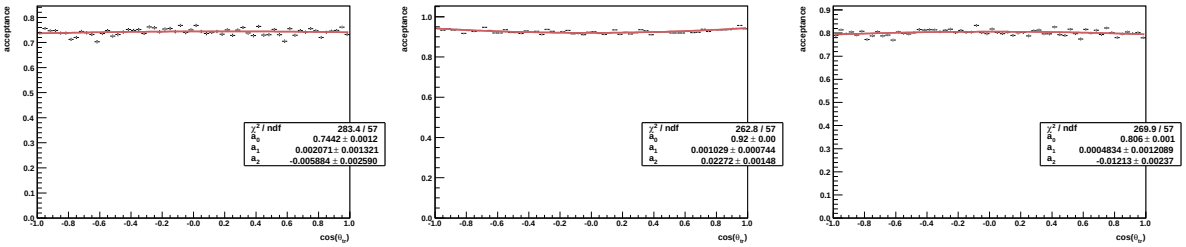
D.2 Angular acceptance studies



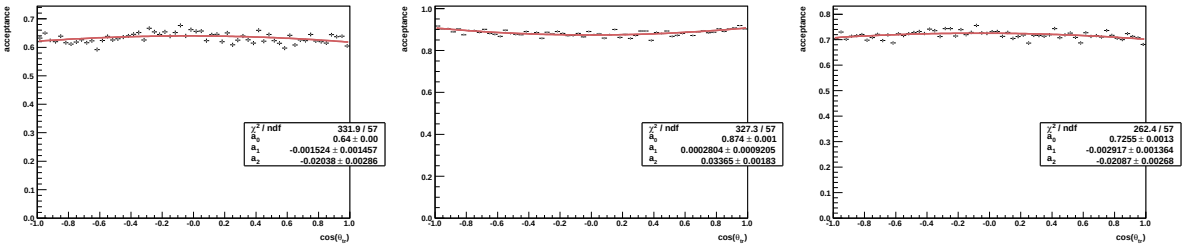
(a) 200 $\mathrm{MeV/c}$ $p_\perp$ cut



(b) 300 $\mathrm{MeV/c}$ $p_\perp$ cut

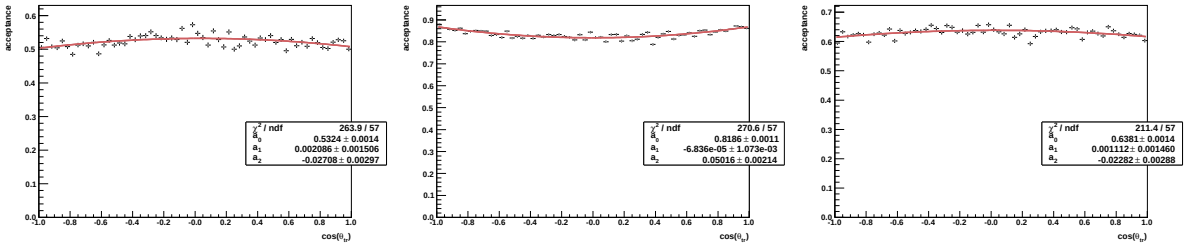


(c) 400 $\mathrm{MeV/c}$ $p_\perp$ cut

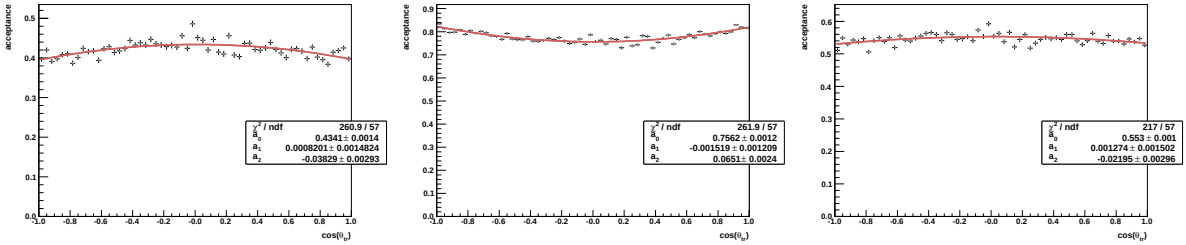


(d) 500 $\mathrm{MeV/c}$ $p_\perp$ cut

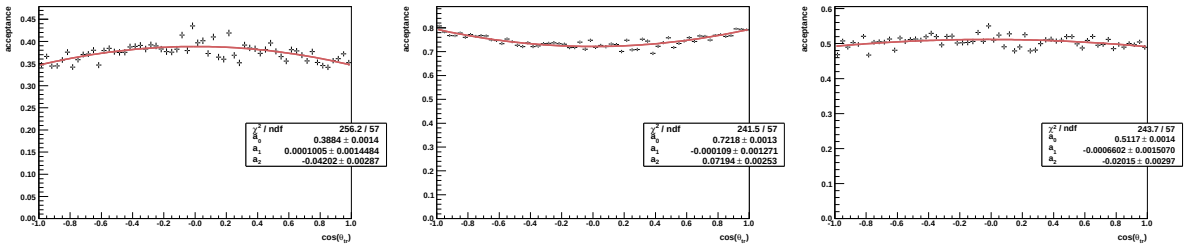
Figure D.8: EVTGEN study of the angular acceptance distribution for θ_{tr}, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



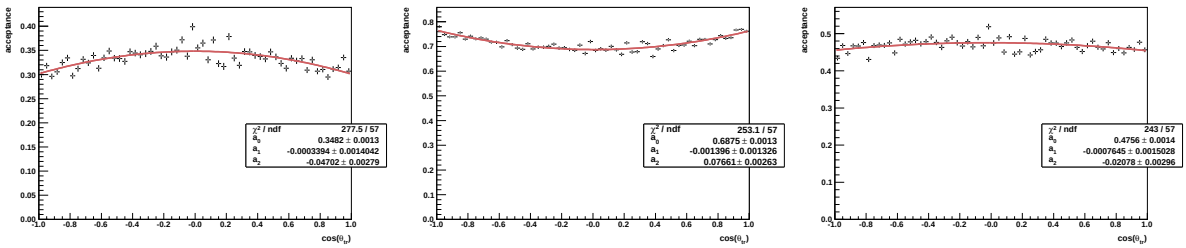
(a) 600 $\mathrm{MeV/c}$ $p_\perp$ cut



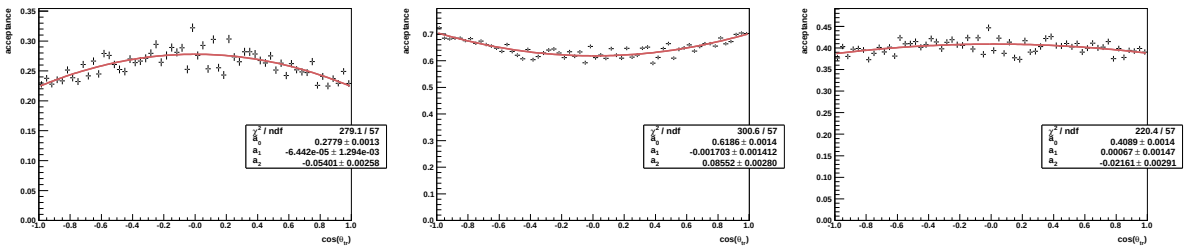
(b) 700 $\mathrm{MeV/c}$ $p_\perp$ cut



(c) 750 $\mathrm{MeV/c}$ $p_\perp$ cut

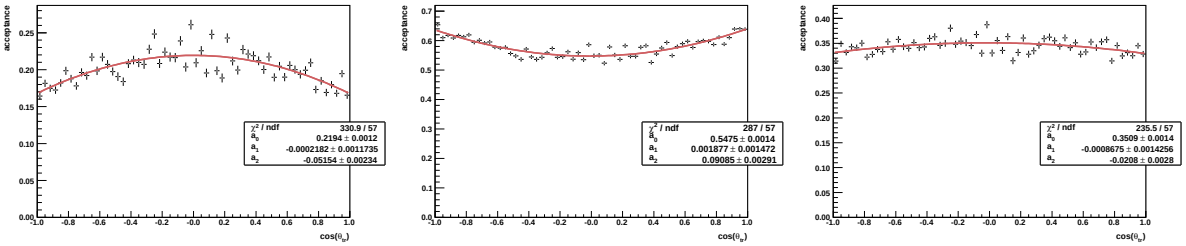


(d) 800 $\mathrm{MeV/c}$ $p_\perp$ cut

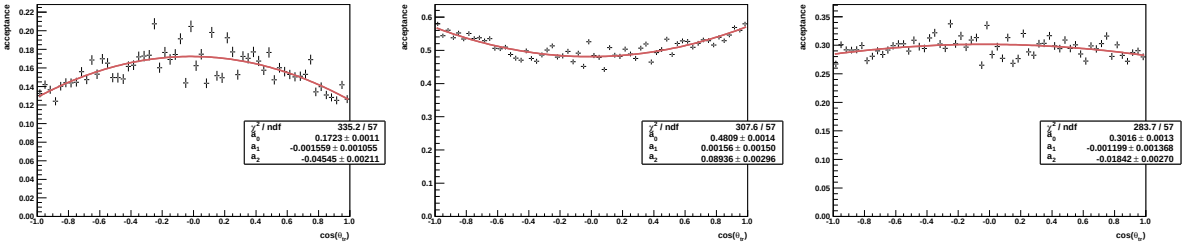


(e) 900 $\mathrm{MeV/c}$ $p_\perp$ cut

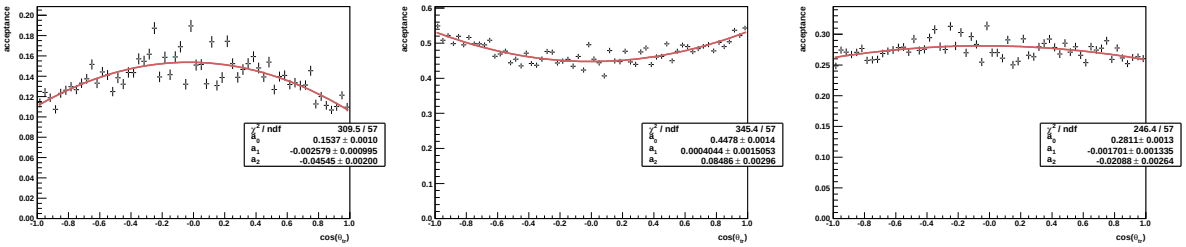
Figure D.9: EVTGEN study of the angular acceptance distribution for θ_{tr}, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



(a) 1000 $\mathrm{MeV/c}$ $p_\perp$ cut

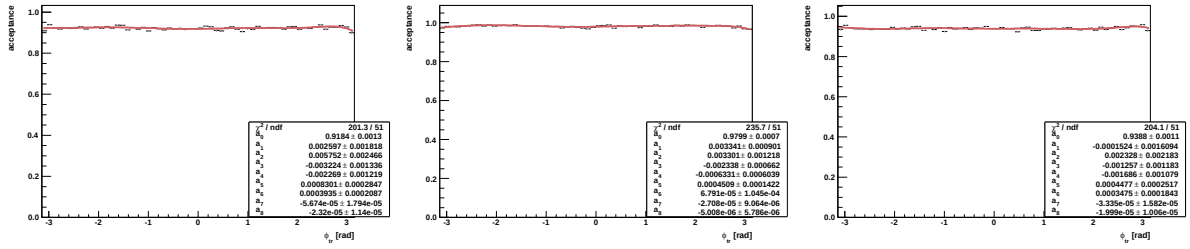


(b) 1100 $\mathrm{MeV/c}$ $p_\perp$ cut

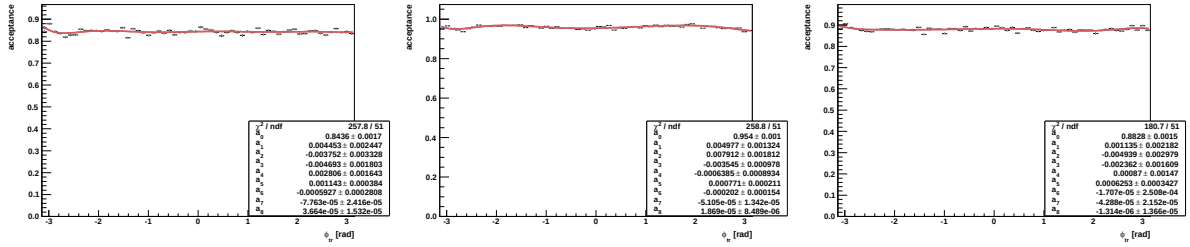


(c) 1150 $\mathrm{MeV/c}$ $p_\perp$ cut

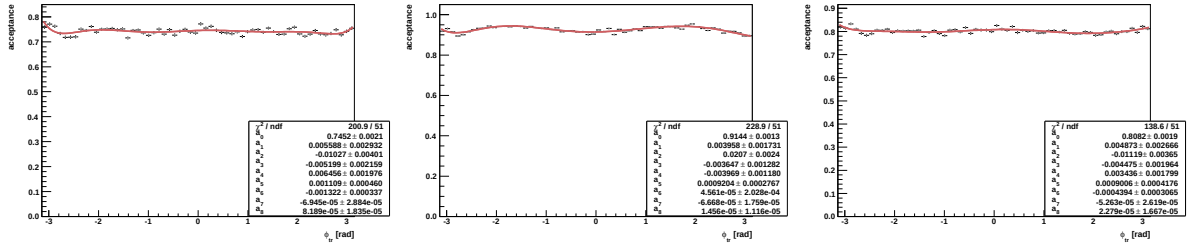
Figure D.10: EVTGEN study of the angular acceptance distribution for θ_{tr}, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



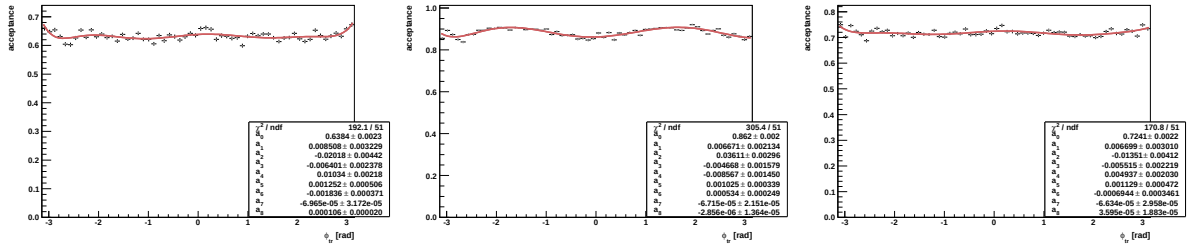
(a) 200 $\mathrm{MeV/c}$ $p_\perp$ cut



(b) 300 $\mathrm{MeV/c}$ $p_\perp$ cut

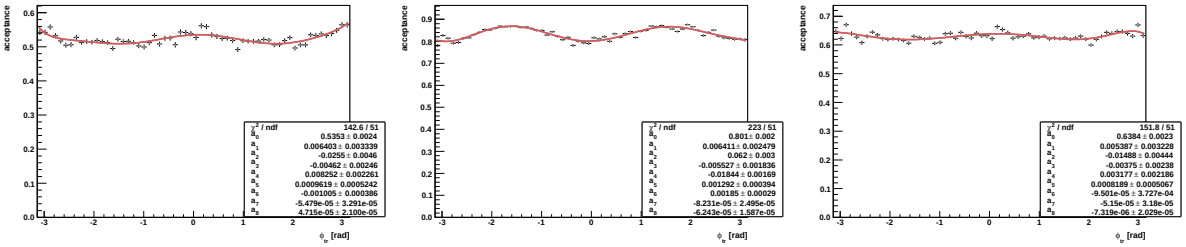


(c) 400 $\mathrm{MeV/c}$ $p_\perp$ cut

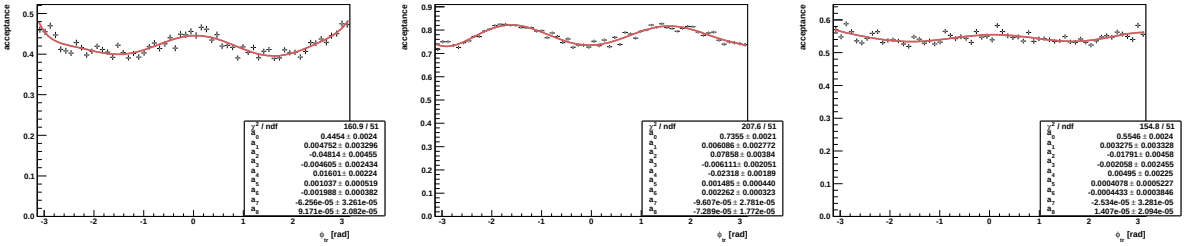


(d) 500 $\mathrm{MeV/c}$ $p_\perp$ cut

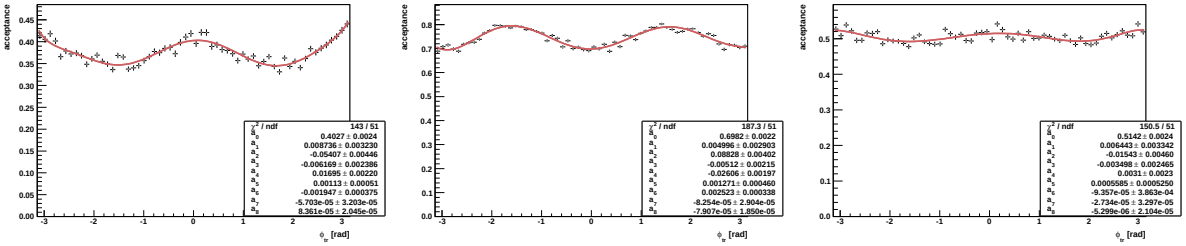
Figure D.11: EVTGEN study of the angular acceptance distribution for ϕ_{tr}, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



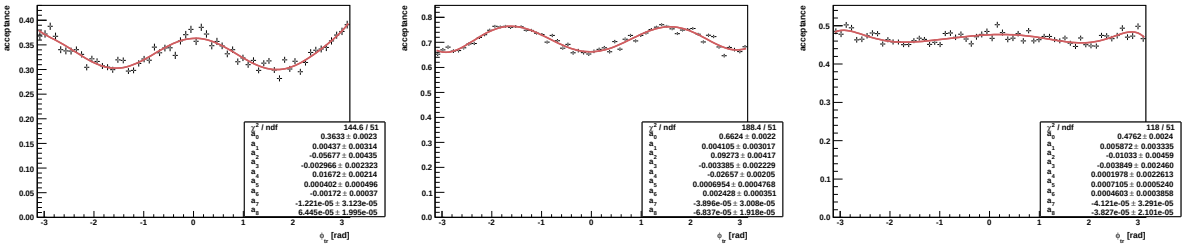
(a) 600 $\mathrm{MeV/c}$ $p_\perp$ cut



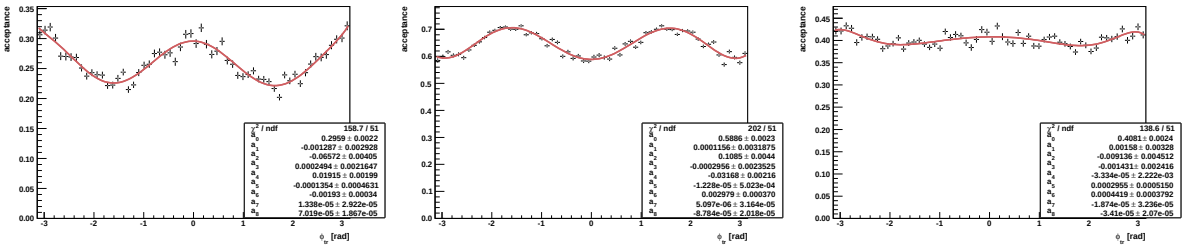
(b) 700 $\mathrm{MeV/c}$ $p_\perp$ cut



(c) 750 $\mathrm{MeV/c}$ $p_\perp$ cut

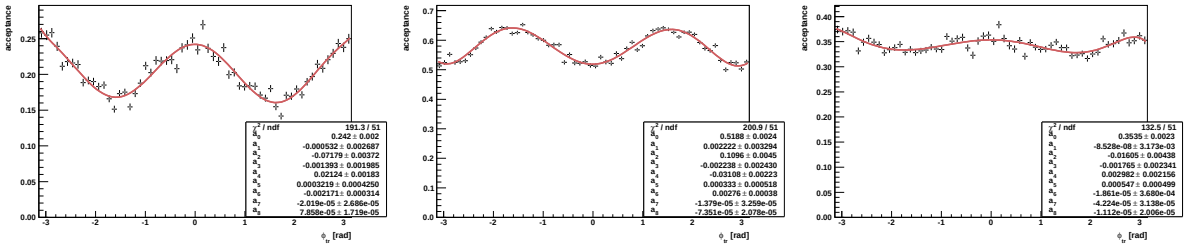


(d) 800 $\mathrm{MeV/c}$ $p_\perp$ cut

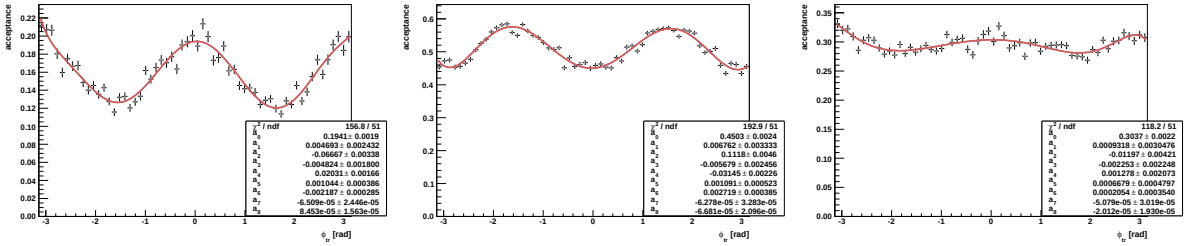


(e) 900 $\mathrm{MeV/c}$ $p_\perp$ cut

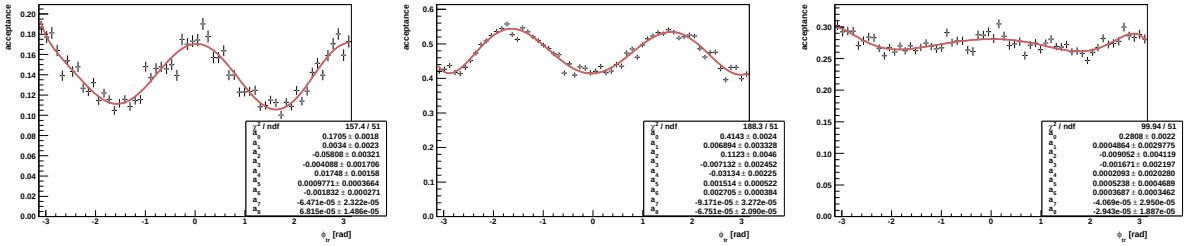
Figure D.12: EVTGEN study of the angular acceptance distribution for ϕ_{tr}, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



(a) 1000 $\mathrm{MeV/c}$ $p_\perp$ cut

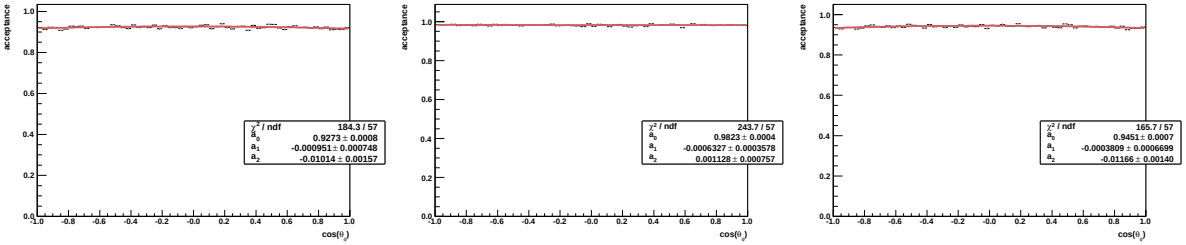


(b) 1100 $\mathrm{MeV/c}$ $p_\perp$ cut

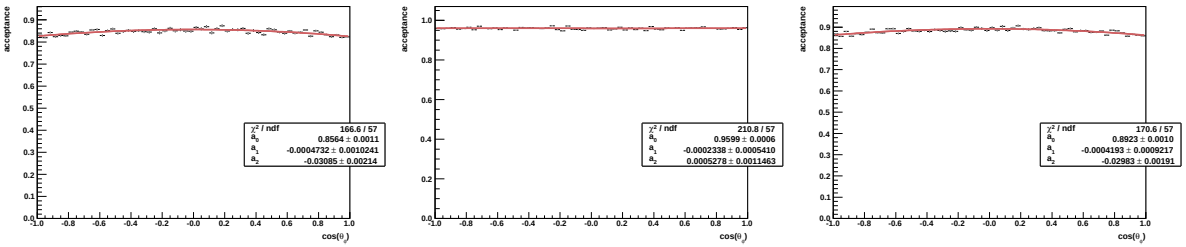


(c) 1150 $\mathrm{MeV/c}$ $p_\perp$ cut

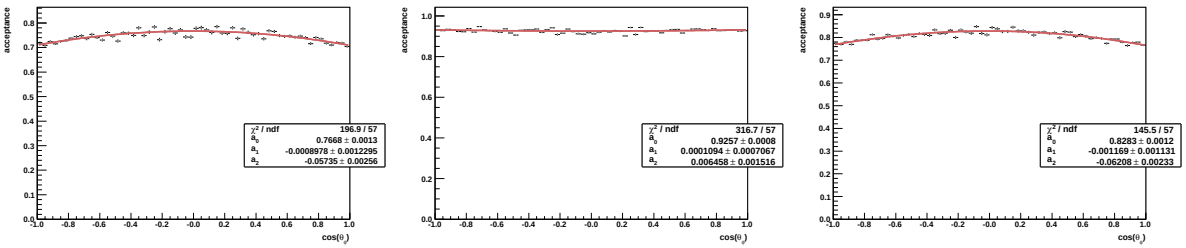
Figure D.13: EVTGEN study of the angular acceptance distribution for ϕ_{tr}, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



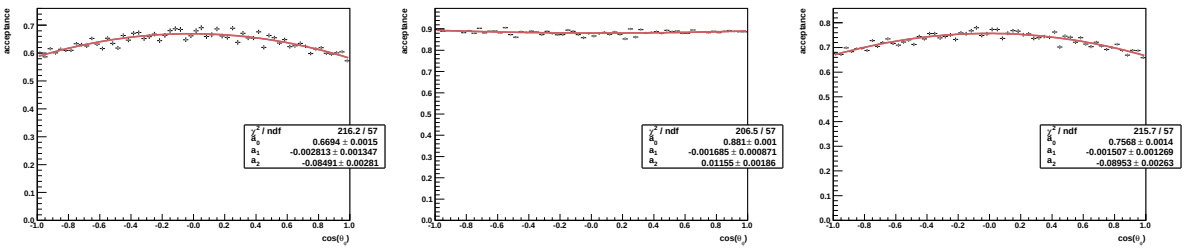
(a) 200 $\mathrm{MeV/c}$ $p_\perp$ cut



(b) 300 $\mathrm{MeV/c}$ $p_\perp$ cut

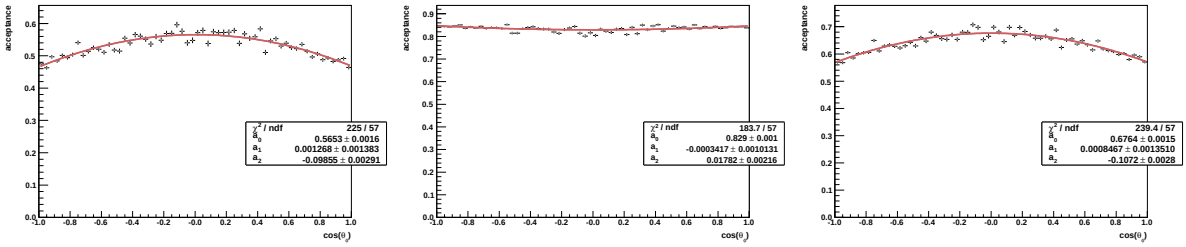


(c) 400 $\mathrm{MeV/c}$ $p_\perp$ cut

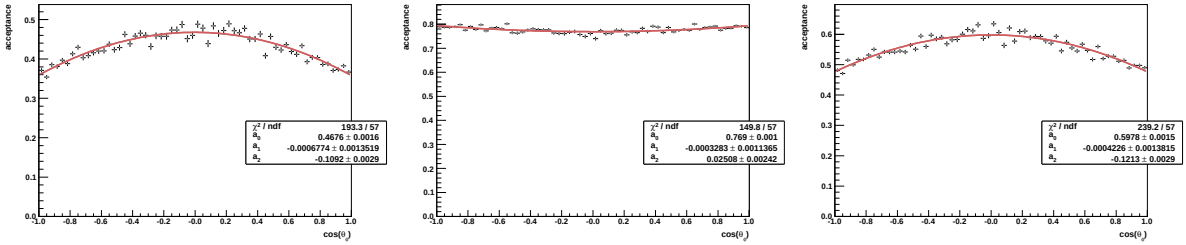


(d) 500 $\mathrm{MeV/c}$ $p_\perp$ cut

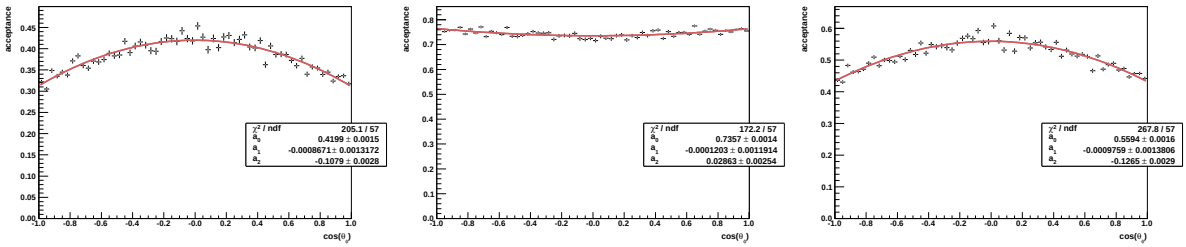
Figure D.14: EVTGEN study of the angular acceptance distribution for θ_{phi}, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



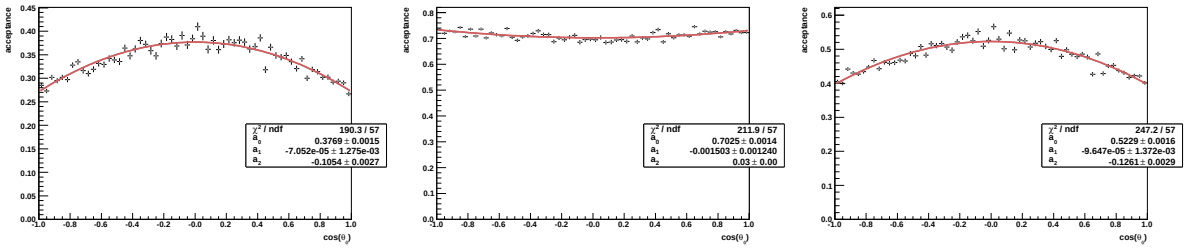
(a) 600 $\mathrm{MeV/c}$ $p_\perp$ cut



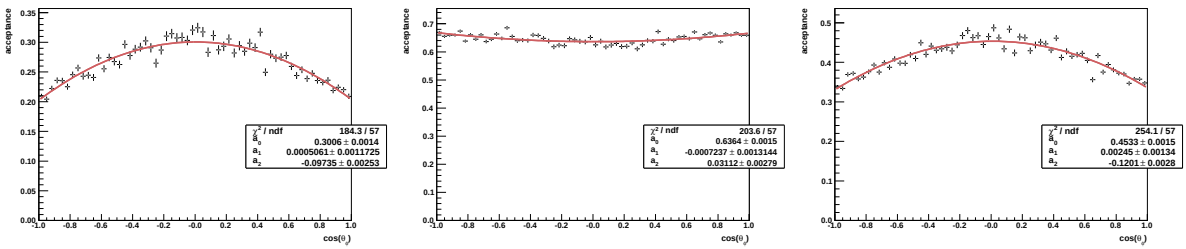
(b) 700 $\mathrm{MeV/c}$ $p_\perp$ cut



(c) 750 $\mathrm{MeV/c}$ $p_\perp$ cut

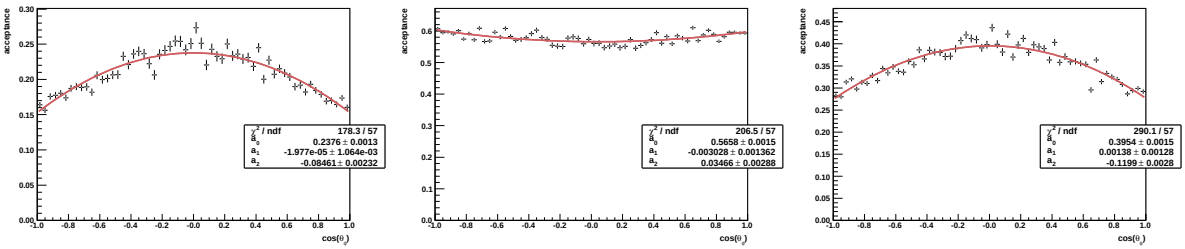


(d) 800 $\mathrm{MeV/c}$ $p_\perp$ cut

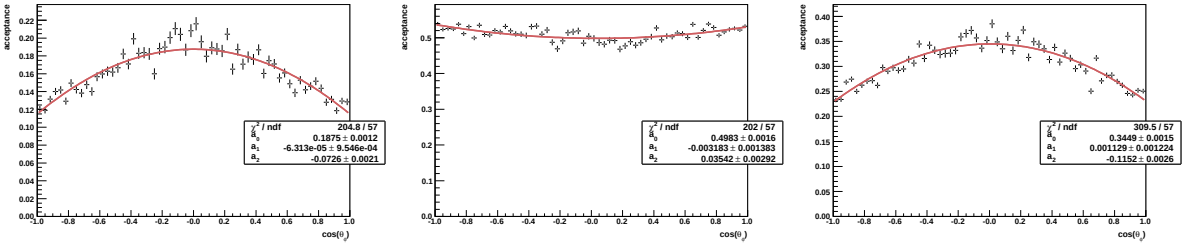


(e) 900 $\mathrm{MeV/c}$ $p_\perp$ cut

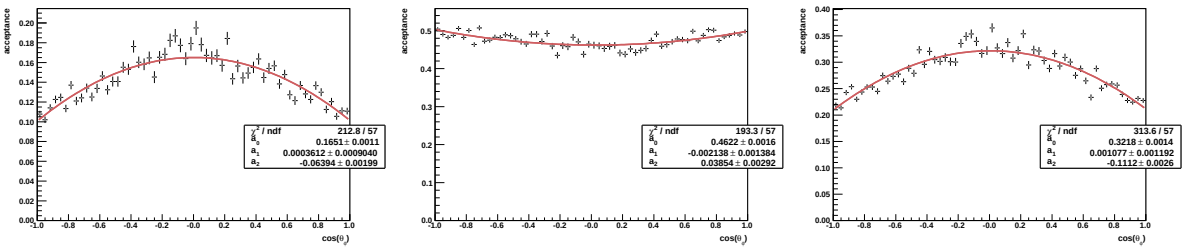
Figure D.15: EVTGEN study of the angular acceptance distribution for θ_ϕ, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.



(a) 1000 $\mathrm{MeV/c}$ $p_\perp$ cut



(b) 1100 $\mathrm{MeV/c}$ $p_\perp$ cut



(c) 1150 $\mathrm{MeV/c}$ $p_\perp$ cut

Figure D.16: EVTGEN study of the angular acceptance distribution for θ_ϕ, as a function of a $p_\perp$ cut on: (left) both muons and kaons, (middle) muons only and (right) kaons only.

D.3 $p_\perp$ studies with three decay angles

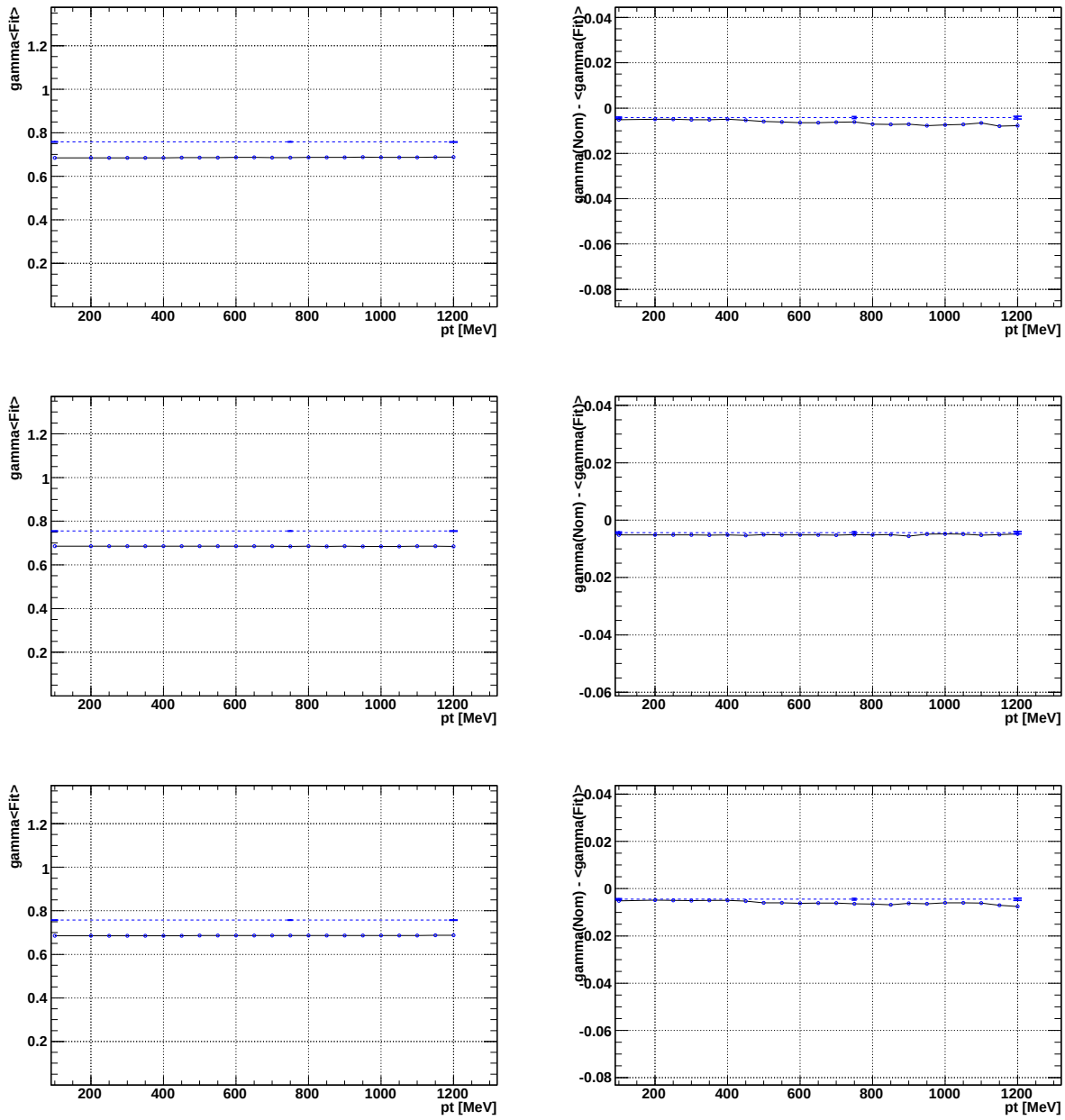


Figure D.17: Shown in the left-hand plots is the mean fit value for $\bar{\Gamma}_s$ using 75 EVTGEN data samples, obtain from a full angular fully tagged the simultaneous fit. The corresponding plots on the right-hand side shows the bias in the fitted value. The top plots illustrate the effect of applying a $p_\perp$ cut on both muons and kaons. The middle plot show the effect for the $p_\perp$ cut on muons only, while the bottom set of plots shows the $p_\perp$ effect on kaons only. The error bars shown at 100, 750, 1200 $\mathrm{GeV/c}$ show the standard error.

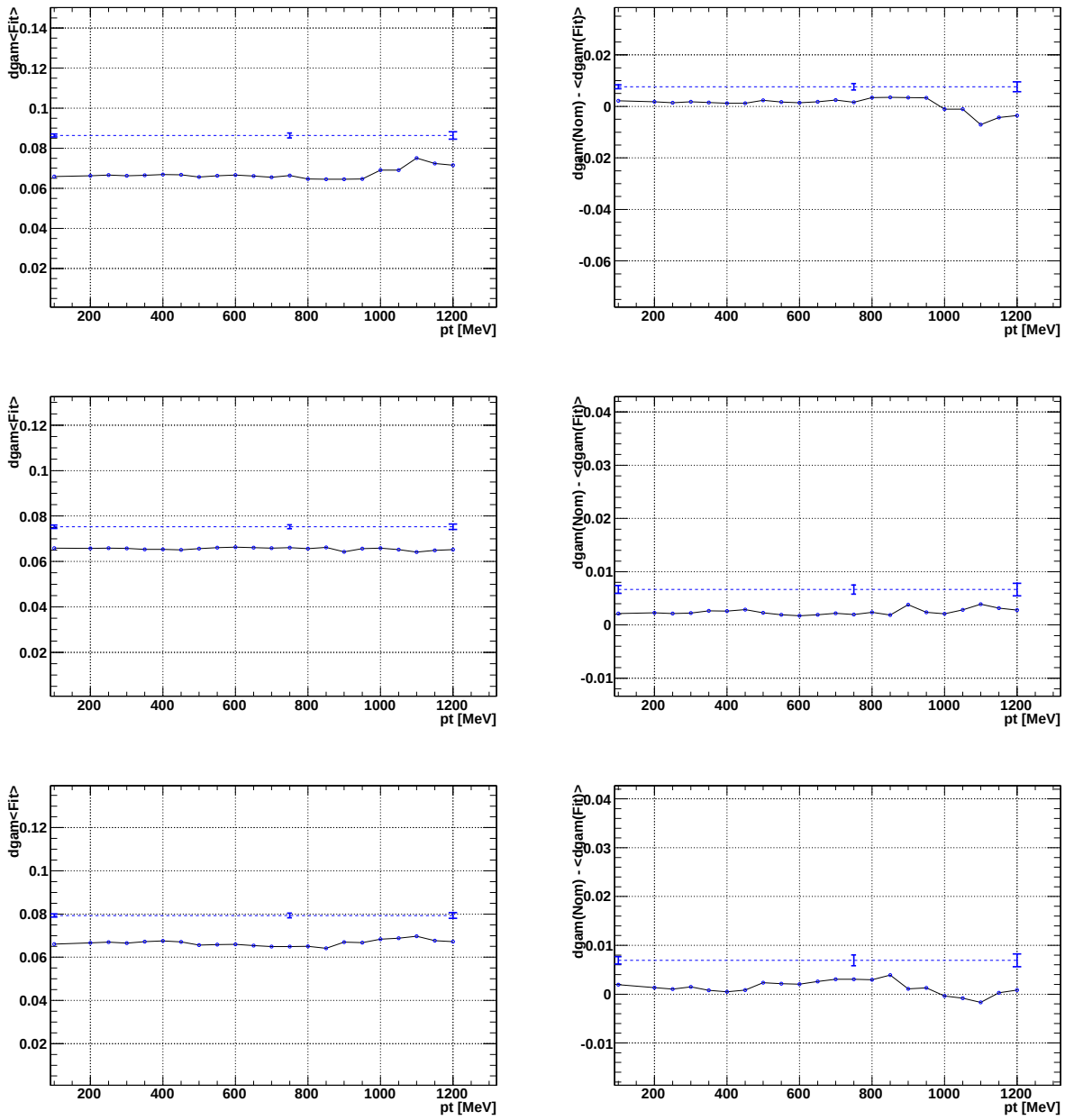


Figure D.18: Shown in the left-hand plots is the mean fit value for $\Delta\Gamma_s$ using 75 EVTGEN data samples, obtain from a full angular fully tagged the simultaneous fit. The corresponding plots on the right-hand side shows the bias in the fitted value. The top plots illustrate the effect of applying a $p_\perp$ cut on both muons and kaons. The middle plot show the effect for the $p_\perp$ cut on muons only, while the bottom set of plots shows the $p_\perp$ effect on kaons only. The error bars shown at 100, 750, 1200 $\mathrm{GeV/c}$ show the standard error.

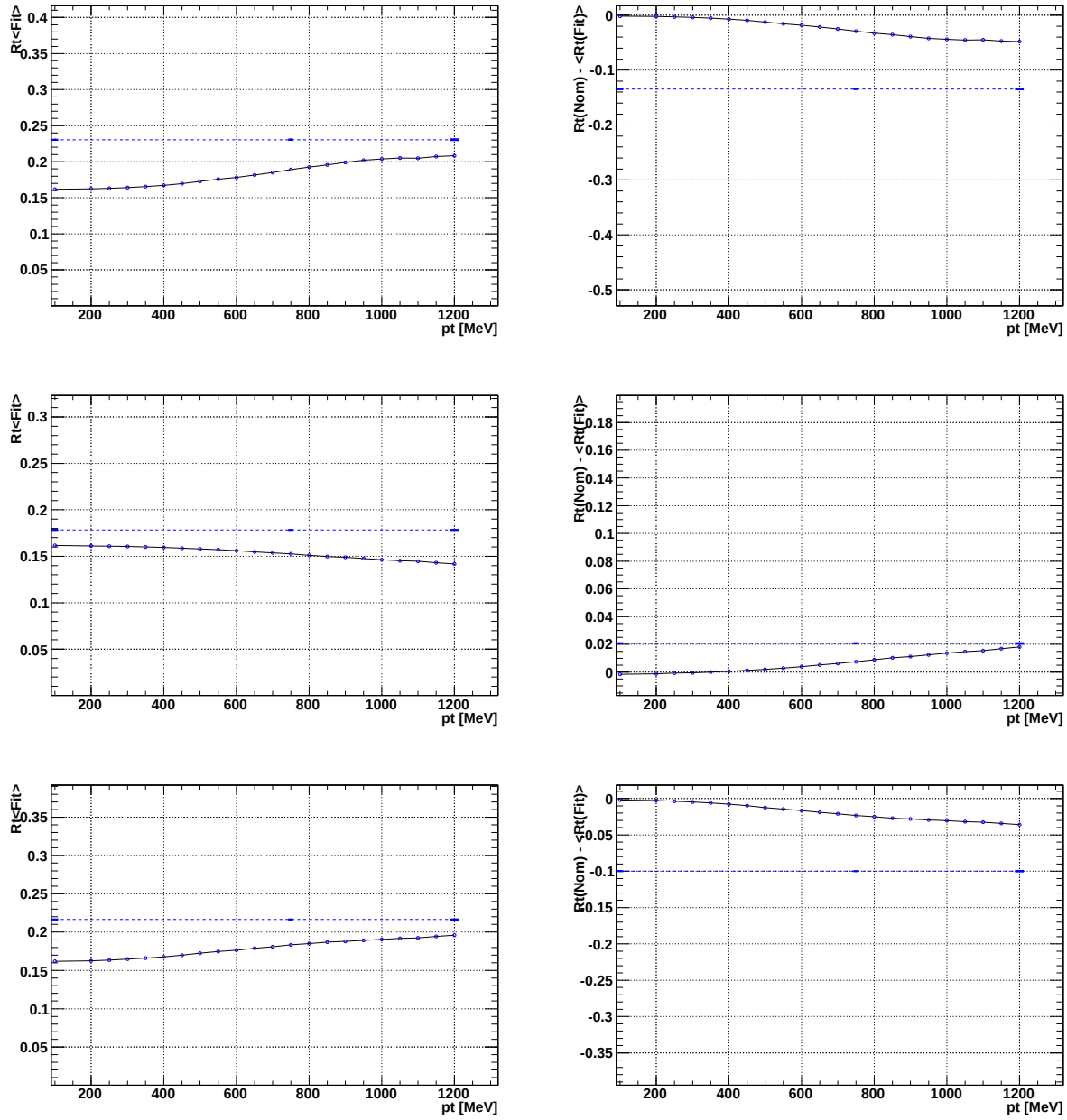


Figure D.19: Shown in the left-hand plots is the mean fit value for $R_\perp$ using 75 EVTGEN data samples, obtain from a full angular fully tagged the simultaneous fit. The corresponding plots on the right-hand side shows the bias in the fitted value. The top plots illustrate the effect of applying a $p_\perp$ cut on both muons and kaons. The middle plot show the effect for the $p_\perp$ cut on muons only, while the bottom set of plots shows the $p_\perp$ effect on kaons only. The error bars shown at 100, 750, 1200 $\mathrm{GeV/c}$ show the standard error.

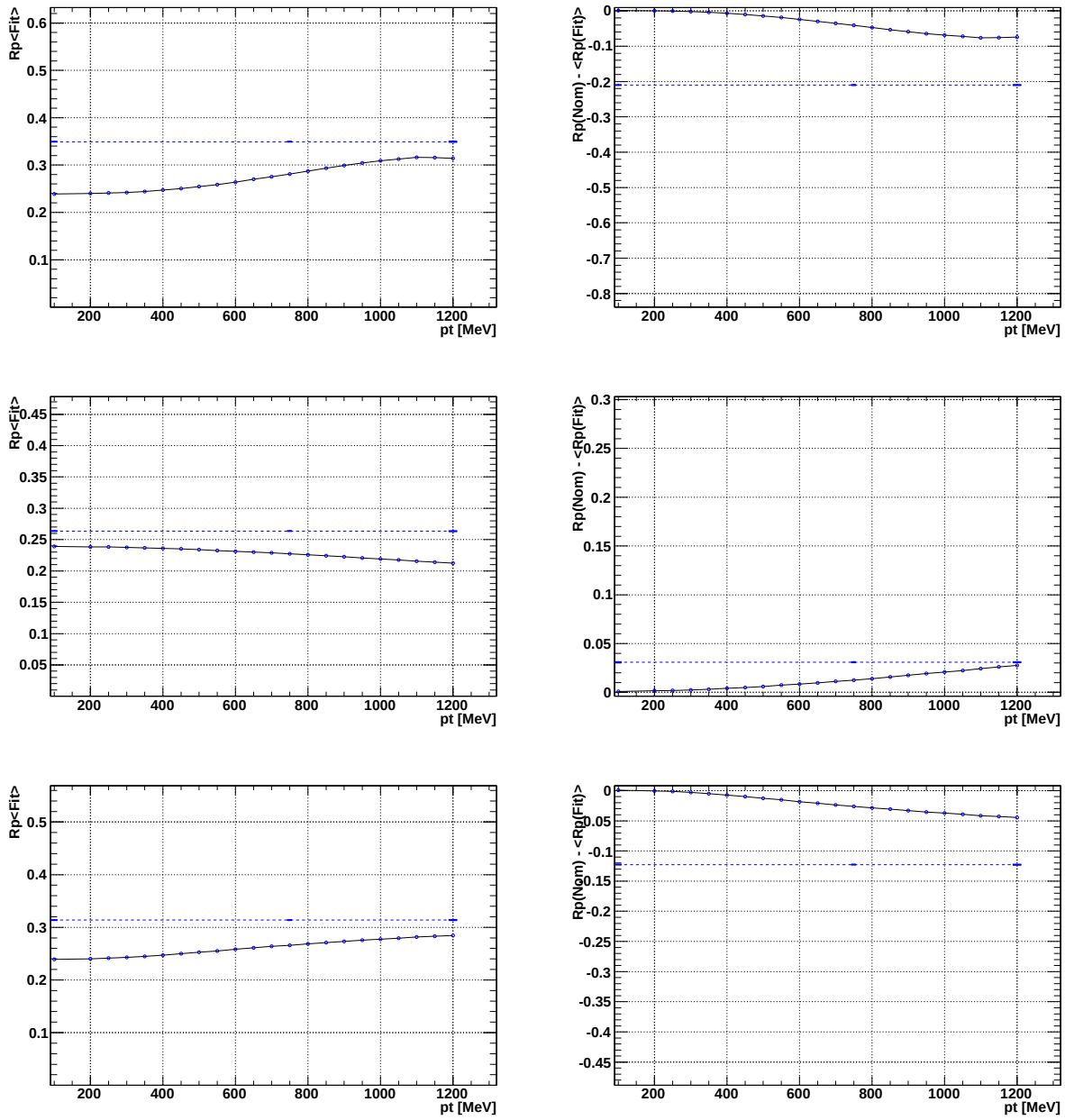


Figure D.20: Shown in the left-hand plots is the mean fit value for R_0 using 75 EVTGEN data samples, obtain from a full angular fully tagged the simultaneous fit. The corresponding plots on the right-hand side shows the bias in the fitted value. The top plots illustrate the effect of applying a $p_\perp$ cut on both muons and kaons. The middle plot show the effect for the $p_\perp$ cut on muons only, while the bottom set of plots shows the $p_\perp$ effect on kaons only. The error bars shown at 100, 750, 1200 $\mathrm{GeV/c}$ show the standard error.

D.4 $\omega_{\rm tag}$ and the Strong phases

Figures D.21 to D.22 show the results of 100 toy MC studies each containing 100,000 tagged events. In these studies, $-2\beta_s$ was set to -0.8 and the strong phases to two different sets: $[\pi/4, 3\pi/2]$ and [-0.46, 2.92] as motivated by $J/\Psi K^*$ decays [19, 91]. In these studies the following set of parameters were let free: $\{\bar{\Gamma}_s$,$\Delta\Gamma_s$,$R_\perp$,R_0,δ_1,δ_2,$-2\beta_s$,$\omega_{\rm tag}\}$, while Δm_s was fixed. All parameters, unless indicated above, were initially set to their nominal value as described in Section 3. In the studies we found the fits well behaved, meaning the fitted value for almost all parameters was close to its generated value. We also found the Log Likelihood (LL) scans parabolic, apart for the deltas.

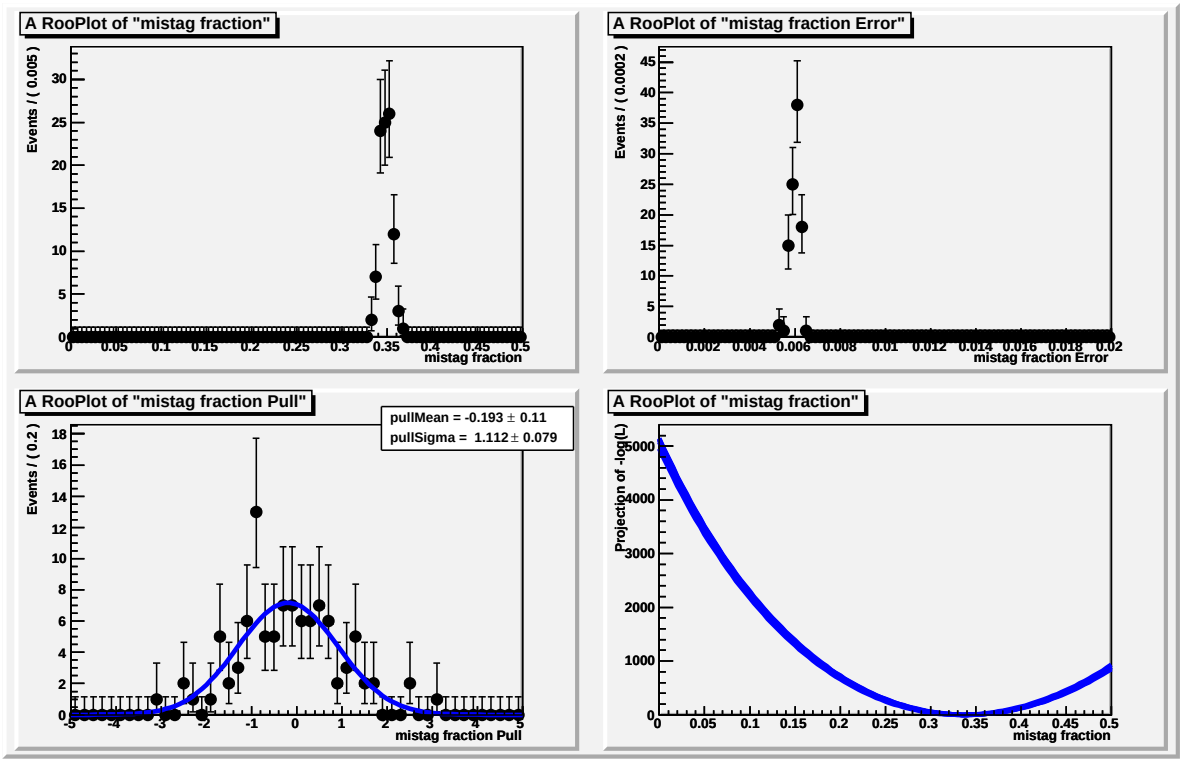


Figure D.21: Fit distributions for $\omega_{\rm tag}$. In this case $\delta_1 = \pi/4$ and $\delta_2 = 3\pi/2$. Shown are (i) central value (ii) Minuit error (iii) pull distribution (iv) 5 different Log Likelihood (LL) scans.

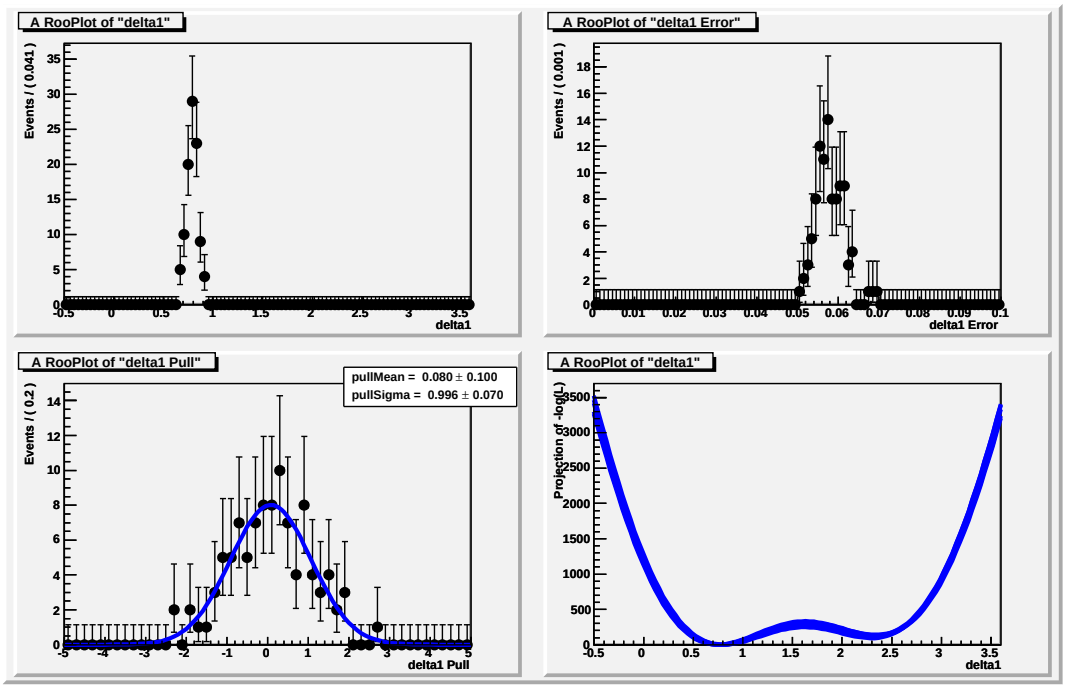


Figure D.22: Fit distributions for δ_1. In this case $\delta_1 = \pi/4$ and $\delta_2 = 3\pi/2$.Shown are (i) central value (ii) Minuit error (iii) pull distribution (iv) 5 different LL scans.

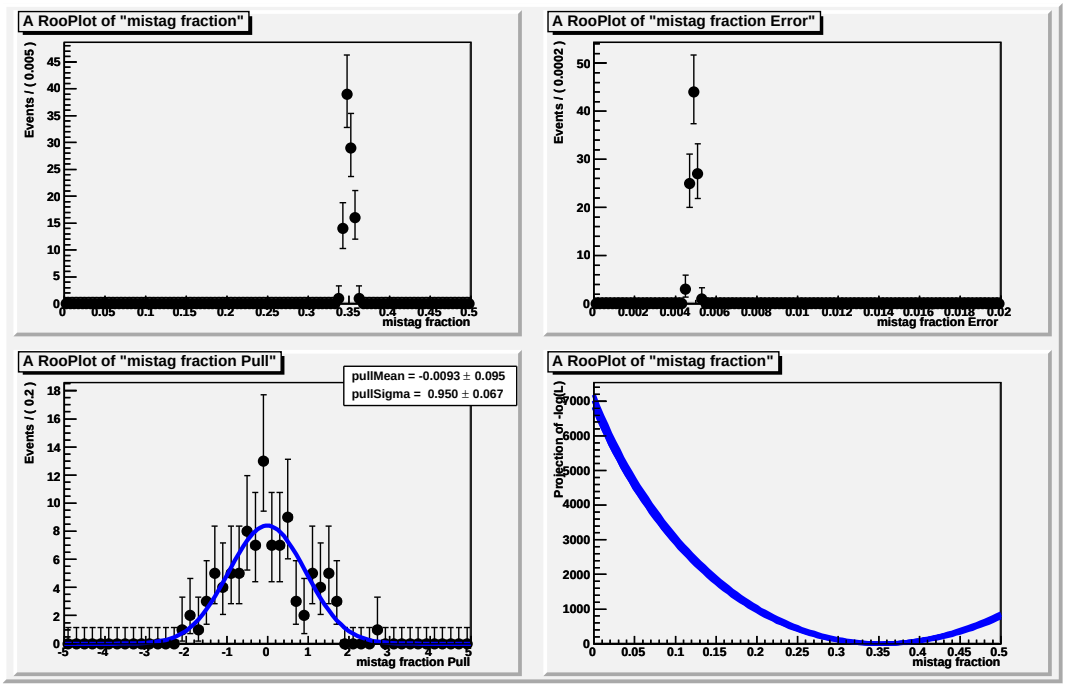


Figure D.23: Fit distributions for $\omega_{\rm tag}$. In this case $\delta_1 = -0.46$ and $\delta_2 = 2.92$. Shown are (i) central value (ii) Minuit error (iii) pull distribution (iv) 5 different LL scans.

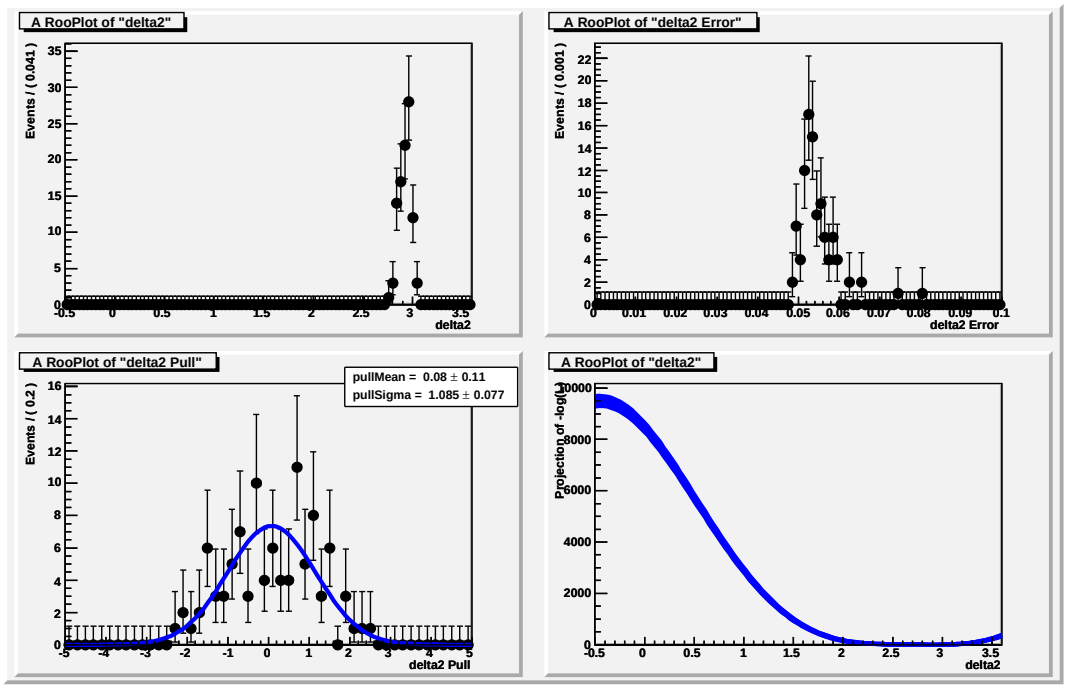


Figure D.24: Fit distributions for δ_2. In this case $\delta_1 = -0.46$ and $\delta_2 = 2.92$. Shown are (i) central value (ii) Minuit error (iii) pull distribution (iv) 5 different LL scans.

E

Optimisation of $B_s \rightarrow J/\psi\phi$ events

E.1 The Fisher

Suppose the result of measuring an individual event is a vector of attributes, $\vec{x} = x_1, \ldots, x_n$ where:

$$\begin{aligned} x_1 &= \text{number of muons,} \\ x_2 &= \text{mean } P_t \text{ of muons,} \\ &\vdots \\ x_n &= J/\psi \text{ mass.} \end{aligned}$$

The collection of attributes will follow some n-dimensional probability distribution function (p.d.f) which depends on the type of events. If the events are signal like ($B_s \rightarrow J/\psi\phi$) or background like (prompt-J/ψ) for instance. Assuming the p.d.f, labelled by f, depends on two categories which determines whether it is signal-or-not such that, $f(\vec{x}|H_0)$ and $f(\vec{x}|H_1)$ can be defined. Here H_0 is the signal category and H_1 is the background category. If we now consider selecting events of type H_0, that is to say signal events, the question arises how to efficiently find the multidimensional boundary in order to accept events belonging to this type. The Fisher linear discriminate seeks this discriminate by projecting the multidimensional space onto a line. A cartoon of this idea is shown in Figure E.1.

With a discriminate such as the Fisher a test statistic, y_{Fi}, can be constructed which compactifies, or projects, the data down into lower dimensions[1] without losing the ability to discriminate between categories in the multidimensional space. P.D.Fs for each categories can then be constructed, which depend on the test statistic, namely $g(t|H_0)$ and $g(t|H_1)$. The boundary decision between the hypothesis then depends on a single cut on y_{Fi}. This effectively divides the sample space into two regions as shown in the left-hand plot in Figure E.2, which either accepts or

[1] in the case of the Fisher linear discriminate, this will be the projection onto a line.

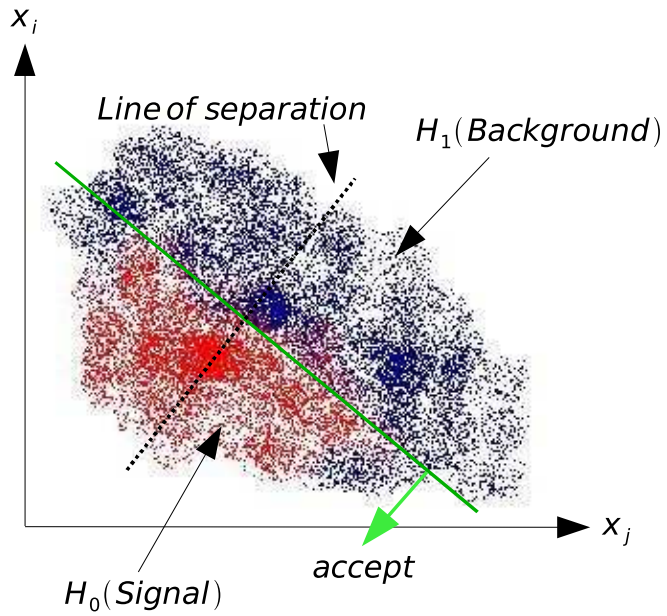


Figure E.1: Linear boundary decision, shown in green. The distribution of signal-like events is coloured in red, whilst that of background like events is coloured in blue.

rejects H_0. In general, the Fisher discriminate makes use of the following linear test statistic,

$$y_{Fi}(i) = \sum_{k=1}^{n} F_k x_k(i) \tag{E.1}$$

Where $y_{Fi}(i)$ is the projection for the i^{th} event in the n-dimensional data set onto the line in the direction of $\vec{F}$. The job now is to calculate the coefficients, $F_0 \ldots F_n$ and thus the direction of the line that maximises the separation between the p.d.fs. This can be achieved by maximising following,

$$J(\vec{F}) = \frac{|(\mu_{H_0} - \mu_{H_1})|^2}{\sigma^2_{H_0} + \sigma^2_{H_1}}, \tag{E.2}$$

Where the sample means are defined as:

$$\mu_D = \vec{F} \cdot \vec{m} = \frac{1}{N_D} \sum_{\vec{x}^{j=1}}^{N_D} \sum_{k=1}^{n} F_k x_k^j, \tag{E.3}$$

with $D \in \{H_0, H_1\}$. As illustrated in the right-hand plot in Figure E.2, $\mu_{H_{0,1}}$ and $\sigma^2_{H_{0,1}}$ are the mean and standard deviation of the H_0 and H_1 category respectively. From equation (E.2), classification of the events into signal and background relies on the mean of each projected class being far apart from each other, whilst at the same time minimising the standard deviation of

each class.

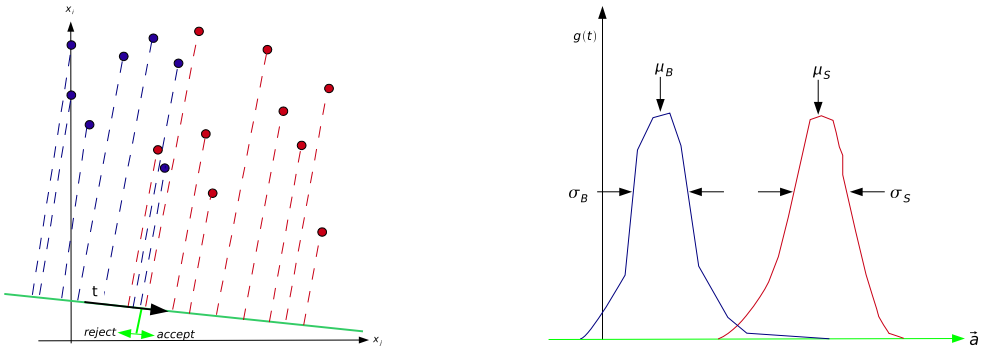


Figure E.2: (Left) Illustration of how the signal (H_0) and background (H_1) sample spaces could look projected onto the test statistic t. (Right) The test statistic can then be used to discriminate signal from background, where the quality of the separation is given by equation (E.2).

To obtain $J(\cdot)$ as an explicit function of $\vec{F}$, we introduce the scatter matrices $\boldsymbol{S}_W$ and $\boldsymbol{S}_B$ by:

$$\boldsymbol{S}_i = \sum_{x=1}^{n} (x - m_i)(x - m_i)^t,$$

where $i \in \{H_0, H_1\}$. The sum of the two sample variances, or scatter projection of the two samples $\sigma^2_{H_0}$ and $\sigma^2_{H_1}$ can be written as,

$$\sigma^2_{H_0} + \sigma^2_{H_1} = \vec{F}^t \boldsymbol{S}_W \vec{F} = \vec{F}^t (\boldsymbol{S}_{H_0} + \boldsymbol{S}_{H_1}) \vec{F}.$$

Similarly, the separation of the projected means obeys:

$$(\mu_{H_0} - \mu_{H_1})^2 = \vec{F}^t \boldsymbol{S}_B \vec{F},$$

with $\boldsymbol{S}_B$=$(m_{H_0} - m_{H_1})(m_{H_0} - m_{H_1})^t$. The matrix $\boldsymbol{S}_W$ is called the within-class scatter matrix. It is proportional to the sample covariance matrix ($\boldsymbol{C}$) for the pooled n-dimensional data. $\boldsymbol{S}_B$ is called the between-class scatter matrix. In terms of $\boldsymbol{S}_W$ and $\boldsymbol{S}_B$, the criterion function $J(\cdot)$ can be written as:

$$J(\vec{F}) = \frac{\vec{F}^t \boldsymbol{S}_B \vec{F}}{\vec{F}^t \boldsymbol{S}_W \vec{F}}$$

The direction of $\vec{F}$ that maximises $J(\cdot)$ is then given by:

$$\vec{F} = \boldsymbol{S}_W^{-1} (m_{H_0} - m_{H_1}).$$

The specific definition of the Fisher discriminant for classifying the i^{th} even is given by:

$$y_{Fi}(i) = F_0 + \sum_{i=0}^{n} F_k x_k(i). \tag{E.4}$$

Where F_k are the Fisher coefficients[2] and F_0 the offset[3].

[2]are given by:

$$F_k = \frac{\sqrt{N_{H_0} N_{H_1}}}{N_{H_0} + N_{H_1}} \sum_{p=0}^{n} \boldsymbol{S}^{-1}_{W,pk} (\mu_{m^{H_0},k} - m^{H_1,k}),$$

[3]designed to set the sample mean $\bar{y}_{Fi}$ of all $(N_S + N_B)$ the events to zero.

E.2 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN)s are widely used in particle physics providing a powerful method to separate signal data from its background. The D0 collaboration for instance, have been using ANNs extensively in single top quark searches [99, 100].

ANNs are artificial learning algorithms, and as such rely heavily on a computers memory. Within a digital computer: a number, name of a person or their face say, is stored in its memory as a string of bits associated to a particular address. To retrieve a memory location we need to know this address, which tells us nothing of the contents of the memory itself. This is the problem with an address-based memory system, one cannot read from memory cell without knowing its address. For example, given half a memory say someones face, we cannot recover the other half, say their name, unless we have that memory address. The other problem with address-based memory is that it is not robust. If one-bit is missing in the address then a completely different memory location will be retrieved. Biological memory systems however are just the opposite. They are associative: given a persons name, we can often recall their face or vice versa. Memories can be recalled spontaneously and not just upon request. We can often recover the correct memory given partial clues. For instance, if asked to recall the name connected with the following prompt: "A current head of state and diligent spokesperson who has represented a Northern hemisphere superpower for the past 10 years". Many people will mention president George W. Bush, even though this statement contains errors. Memories persist in our brain even if there in a continuous state of change; cells and proteins within the cells are continuously changing, becoming damaged, or destroyed, by natural processes.
Biological memory systems are therefore parallel, distributive (throughout our brain) and robust. It is these properties that have motivated the study of ANNs. The goal is to simulate on a computer a parallel distributive system of many interactive elements to correctly recall a response given some initial stimulant. This effectively defines what we consider as a ANN, which is any simulated collection of connected neurons, with each neuron producing a response to a given set of input variables. Supplying the variables to the input neurons puts the network into a defined state that can be measured from the response of one or more of the output neurons.

An ANN behaviour is determined by the output layer of neutrons, the weights of the inter-neuron connections, and by the response of the neurons to the input given by the neuron response function ρ. ANNs are commonly divided into two classes:

- **Supervised neural networks:** Where the ANNs are trained on a set of input-output pairs such that the network learns to model the dependency between them. In the case of MultiLayer Perceptrons (MLP), the training step is designed to optimise all the weights and biases, given in equation (E.5), to their optimal value for a given input-output pair ($s(t)$,$x(t)$); with the criterion optimised typically $\sum_t |f(s(t)| - x(t))^2$.

- **Unsupervised neural networks:** These are ANNs that are trained using a set of example datasets. The ANN's are then simply required to memorise the data in such a way that the examples can be recalled later. These networks are intended to discover 'patterns' in the data or underlying features from the examples.

One of the most common ANNs used is the supervised multilayer network, such as the MultiLayer Perceptron (MLP). This is a feedforward network, which implies all connections 'flow' in a single direction such that the network forms an acyclic graph as shown in Figure E.4. This network realise of the functionality of a basic perceptron, shown in Figure E.3. The basic idea of a simple perceptron is to map the neuron inputs, $y_1^l, \ldots, y_n^l$, onto a neuron output. This is achieved by using the neuron response function ρ. This function can be separated into a $\mathcal{R}^n \mapsto \mathcal{R}$ function[4] κ and a $\mathcal{R} \mapsto \mathcal{R}$ neuron activation function[5] α, such that $\rho = \alpha\dot{\kappa}$. This is mathematically written as:

$$y_j^l = \rho = \alpha\dot{\kappa} = \alpha\left(w_{0j}^{l-1}\sum_{i=1}^{n} y_i^{l-1}w_i^{l-1}\right) \tag{E.5}$$

where w_{ij}^{l-1} and y_i^{l-1} denote the input weights and variables of the $l-1$ layer into the j^{th} neuron, while w_{0j}^{l-1} denotes the bias on the $l-1$ layer.

Single perceptrons are used as building blocks to larger structure including the MLP. This network has input neurons, hidden neurons and output neurons arranged into layers: the first layer of a MLP being called the input layer, the last one being called the output layer, and all others being called hidden layers. The most common multilayer perceptrons have a single hidden layer, and are known as 'two-layer' networks; the number of layers being counted as the number of neuron layers not including the input layer. For a classification problem (such as discriminating signal from background) with n input variables and 2 output classes. The input layer consists of n neurons that hold the input values $x_1 \ldots x_n$, and one neuron in the output layer that holds the output variable, the neural net estimator[6] y_{ANN}. Each directional connection between the output of one neuron and the input of another has an associated weight. The value of the output neuron is multiplied with the weight to be used as input value for the next neuron. Optimisation of the output neuron is obtained during the training cycle. Here weights initially randomised during the first training cycle (or epoch), are adjusting using a supervised learning algorithm, until optimisation of the output neuron is achieved. The output from the

[4]this is effectively a linear combination of the input weights such that:

$\kappa : (y_1^l, \ldots y_n^l | w_{0j}^l, \ldots w_{nj}^l) \rightarrow w_{0j}^l + \sum_{i=1}^n y_i^l w_{ij}^l$. As this function connects the weights and variables in a given layer together and transmits the result to an activation function, it is often called the synapses function, in analogy with neurobiology.

[5]Commonly used activation functions include the Heaviside step, Sigmoid or the Tanh functions. These functions are convenient since they are close to linear near the origin, while saturating quickly away from the origin. This allows the MLP network to model well both strongly and mildly nonlinear mappings.

[6]If two neurons were used in the output layer one for signal and the other for background, their output values would be y_{ANN} and $1 - y_{ANN}$, respectively.

MLP; with a single hidden layer, Tanh activation function, and a linear output layer can be written mathematically as:

$$y_{ANN}(i) = \sum_{j=1}^{n_h} y_j^{(2)} w_{jl}^{(2)} = \sum_{j=1}^{n_h} \tanh\left(\sum_{k=1}^{n} x_k(i) w_{kj}^{(1)}\right) \cdot w_{j1}^{(2)}, \tag{E.6}$$

where n and n_h the number of neurons in the input layer and hidden layer respectively, $w_{kj}^{(l)}$ is the weight between input layer neuron i and hidden-layer neuron j, and $w_{j1}^{(2)}$ is the weight between the hidden layer neuron j and the output neuron.

The training cycle then enters the learning stage, where the network is supplied with n training events $x_a = (x_1, \ldots, x_n)_a$, $a = 1, \ldots n$. For each training event the neural network output y_{ANN}, a is calculated and compared to the desired output[7] $\hat{y}_a \in \{1, 0\}$ (1 for signal events, 0 for background events). This then feeds into an error function,

$$\text{Error}(x_1, \ldots\ldots, x_n|\mathbf{w}) = \sum_{a=1}^{n} E_a(x_a|\mathbf{w}) = \sum_{a=1}^{N} \frac{1}{2}(y_{ANN,a} - \hat{y}_a)^2, \tag{E.7}$$

which measures the level of agreement between the networks response and the desired one. In equation (E.7), $\mathbf{w}$ denotes the set of adjustable weights in the network. Network weights are adjusted in order to reduce the value of the error function, and the next epoch begins. After each epoch, the network is run on a set of independent signal and background events and the error function is recalculated. The procedure is repeated until the error function reaches a threshold[8].

Some drawbacks are that they are relatively slow to train and that the set of weights are sensitive to the training events (different training samples may lead to different sets of weights, although they may have similar performance).

[7] For an output of 0 or 1, use of a heavyside synapses function should be used.

[8] This early stopping technique avoids over-training, which happens when the training error improves while the testing error starts to go up again. At this point the network loses part of its generalisation power as it has learnt too much about specific events in the training sample.

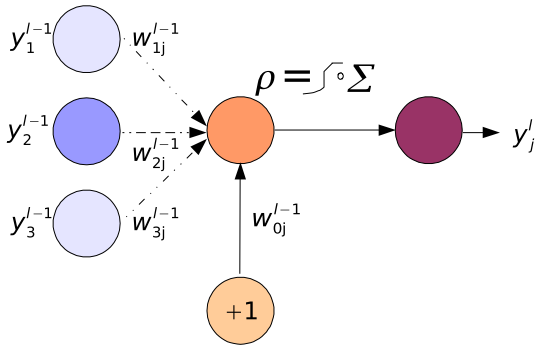


Figure E.3: Single perceptron, j, in the l^{th} layer with 3 input connects, carrying a weight w_{ij}^{l-1}, and a bias w_{0j}^{l-1}, which fixed to a constant,+1 and feeds into each neuron other than the input neurons.

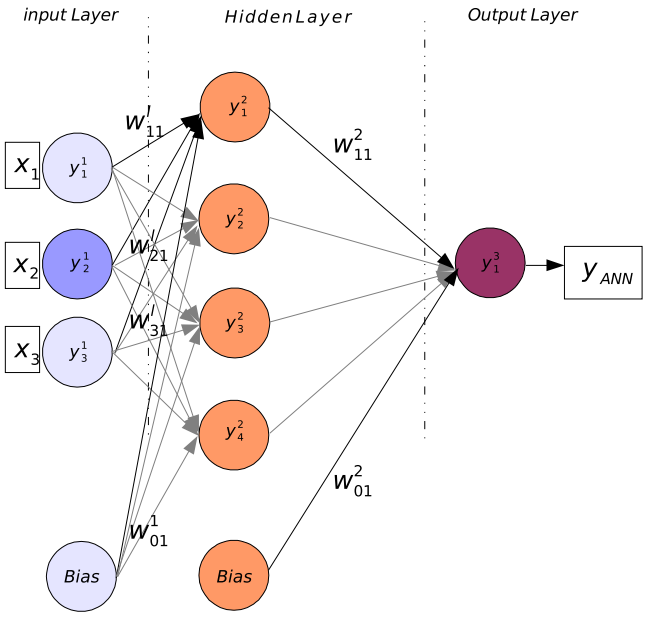


Figure E.4: MultiPerceptron Layer with one hidden layer. The input variables to the input neurons ($1st$ layer) are denoted x_i, the output of each input neuron is y_i^1 and the weight associated to each input neuron is w_{ij}^1, where j denotes the j^{th} next neuron layer (in this example the hidden layer). The response of each neuron in the final hidden layer is then fed into the output neuron y_1^3, which given the ANNs respose.

E.3 Boosted Decision Tree

In order to discuss the AdaBoost algorithm, let us assume the following:

- For the first decision tree there are a total of N events in the sample with the weight of each set to $\frac{1}{N}$.
- $\vec{x}$ is the set of variables for the i^{th} event.
- y_i equals 1 if the i^{th} event is a signal event and -1 if the event is a background event.
- w_i is the weight of the i^{th} event.
- The m^{th} trees decision: $T_{m,i}(\vec{x})$ equals 1 if the set of variables for the i^{th} event lands on a signal leaf and $T_{m,i}(\vec{x})$ equals -1 if the set of variables for that event lands it on a background leaf.
- The condition $I(y_i \neq T_{m,i}(\vec{x}))$ equals 1 if $y_i \neq T_{m,i}(\vec{x})$ and 0 if $y_i = T_{m,i}(\vec{x})$.

With the AdaBoost algorithm we then define for the $(m-1)^{th}$ tree the misclassified rate:

$$\mathrm{err}_{m-1} = \frac{\sum_{i=1}^{N} W_i I(y_i = T_{m-1,i}(\vec{x}))}{\sum_{i=1}^{N} W_i}. \tag{E.8}$$

The misclassified rate, err, for this tree is then used to modify the event sample in the next tree. This happens since the event weights in this new sample are modified using the previously misclassified events. This is achieved by multiplying the events by a common boost weighting factor α, which is derived from the misclassified rate err of the previous tree such that:

$$\alpha_m = \ln\left(\frac{1 - err_{m-1}}{err_{m-1}}\right), \tag{E.9}$$

which changes the weight of each event in the new tree,

$$w_i \rightarrow w_i \times e^{\alpha_m I(y_i \neq T_{m,i}(\vec{x}))} \tag{E.10}$$

The entire event sample is then renormalised to keep the total number of events in the tree, that is the sum of weights, constant:

$$w_i \rightarrow \frac{w_i}{\sum_{i=1}^{N} W_i}. \tag{E.11}$$

The resulting event classification $\mathrm{y_{BDT}}(\vec{x})$ for the boosted classifier is then given by,

$$\mathrm{y_{BDT}}(\vec{x}) = \sum_{i=1}^{N_{Trees}} \alpha_i \cdot T_i(\vec{x}). \tag{E.12}$$

Small values for $y_{\mathrm{BDT}}(\vec{x})$ (values tending towards -1) indicate a background-like events, while large values (values tending towards 1) indicate signal-like events.

E.4 MultiVariate Visualisation (MVV)

E.4.1 Parallel Coordinates (PC)

The scatter diagram is one of the most fundamental techniques for visualising data. It allows the eye to detect such structures in data as linearity, clustering, outliers and so on. However, scatter plots do not generalise readily beyond three dimensions and new approaches to visually representing multivariate data must be found.

Rather than adopting the approach of a scatter diagram, which preserves the orthogonality of the n-dimensional co-ordinate axes, the approach taken by the visualisation technique Parallel Coordinates, is to draw the axes parallel to each other at equidistances apart. A data vector of n attributes (x_1,x_2,...,x_n) is created by plotting x_1 on axis 1, x_2 on axis 2, and so on through to x_n on axis n, as illustrated in Figure E.5. These points are joint by a line such that each line on the Parallel Coordinate diagram represents and point in the n-dimensional space: we thus obtain a point-line duality between the two techniques [118].

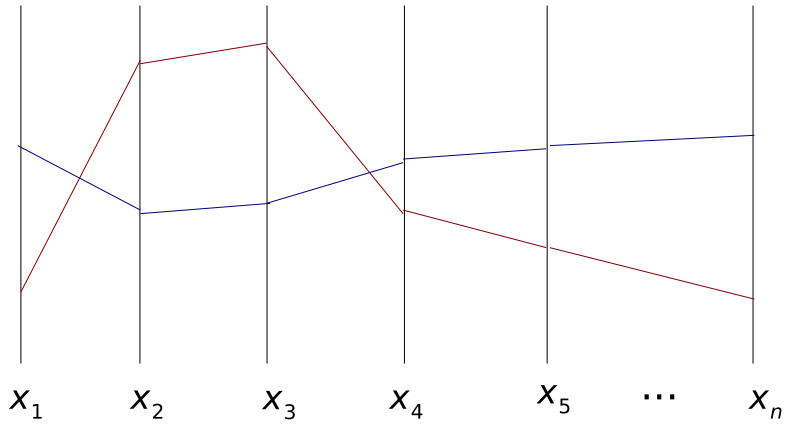


Figure E.5: Parallel coordinate representation of two n-dimensional points.

It is also possible to transform other objects for the usual Cartesian coordinates onto Parallel Coordinates. Consider a line $\mathcal{L}$ in the Cartesian coordinate plane given by: $y = mx + b$, and consider two points lying on that line, say $(a, ma+b)$ and $(c, mc+b)$ and illustrated in E.6. Next consider the xy Cartesian axes mapped into the xy parallel axes and superimpose the Cartesian coordinate axes tu on the xy parallel axes so that the y parallel axis has the equation $u = 1$. The point $(a, ma + b)$ in the xy Cartesian system maps into the line joining $(a, 0)$ to $(ma + b, 1)$ in

the tu coordinate axes. Similarly, $(c, mc + b)$ maps into the line joining $(c, 0)$ to $(mc + b, 1)$. These two lines intersect at a point (in the tu plane) given by $\bar{\mathcal{L}}$: $[b(1-m)^{-1}, (1-m)^{-1}]$. Thus $\mathcal{L}$ is the dual of $\bar{\mathcal{L}}$, giving the duality that points in Cartesian coordinates map into lines in Parallel Coordinates and lines in Cartesian coordinates map into points in Parallel Coordinates. For $0 < (1-m)^{-1} < 1$, m is negative and the intersection occurs between the Parallel Coordinate axes, so for negatively correlated data lines tend to intersect between Parallel Coordinate axes. For $m = -1$, the intersection is exactly midway. In the highly positively correlated data, the case where $(1-m)^{-1} < 0$ or $(1-m)^{-1} > 1$, lines tend to be parallel between the Parallel Coordinate axes, as shown in Figure E.7.

There are two further dualities. Rotation in Cartesian coordinates becomes translations in Parallel Coordinates and vice versa and points of inflection in Cartesian space become cusps[9] in Parallel Coordinate space and vice versa.

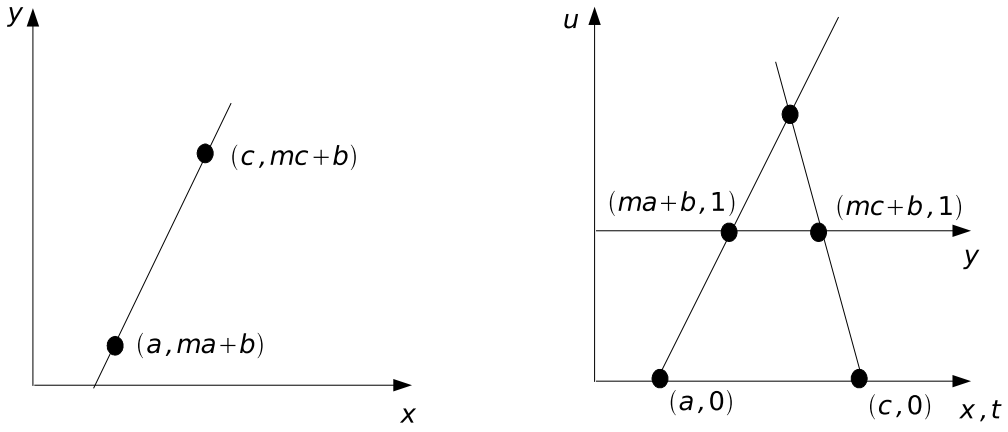


Figure E.6: Cartesian and Parallel Coordinate plots of two points. The tu Cartesian coordinates system is superimposed on the xy parallel coordinate system.

Uncorrelated data is also clearly visible seen in Parallel coordinates. If our data set was completely uncorrelated, we would expect the two-dimensional scatter diagram to fill substantially a circumscribing circle, the outline being in general an (point) ellipse. The corresponding Parallel coordinate transformation would in general be a (line) hyperbola with a (point) hyperbola as envelope, as shown in Figure E.8.

Clustering is also easily diagnosed in Parallel Coordinates. Figure E.9 (a) illustrate the separation of data in both x and y, while (b) shows the separation only in the x coordinate. Since Parallel Coordinate axes represent one-dimensional projections of the data, separation on any axis represents a view of the data that allows the detection of clustering. Due to the connectedness of the multidimensional Parallel Coordinate diagram, it is usually easy to see whether

[9] a point at which a curve crosses itself and at which the tangent of two lines coincide.

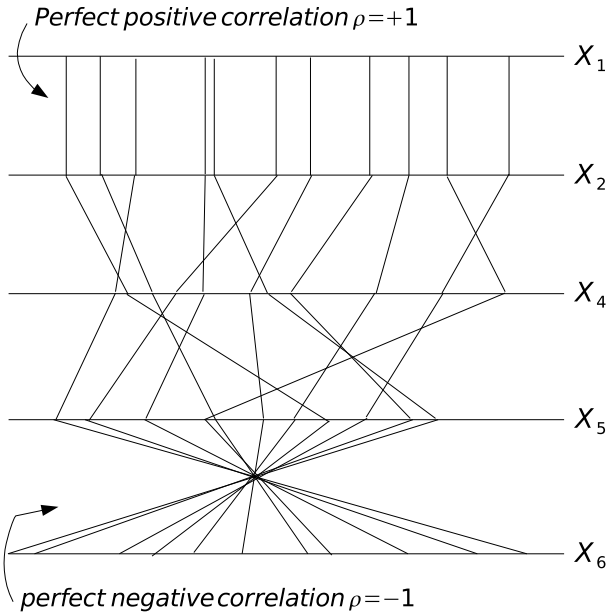


Figure E.7: Parallel Coordinates of six dimensional data illustrating the correlation of data: perfect positive correlation between axis, or dimensions, x_1 and x_2, while perfect negative correlation is shown between axis x_5 and x_6.

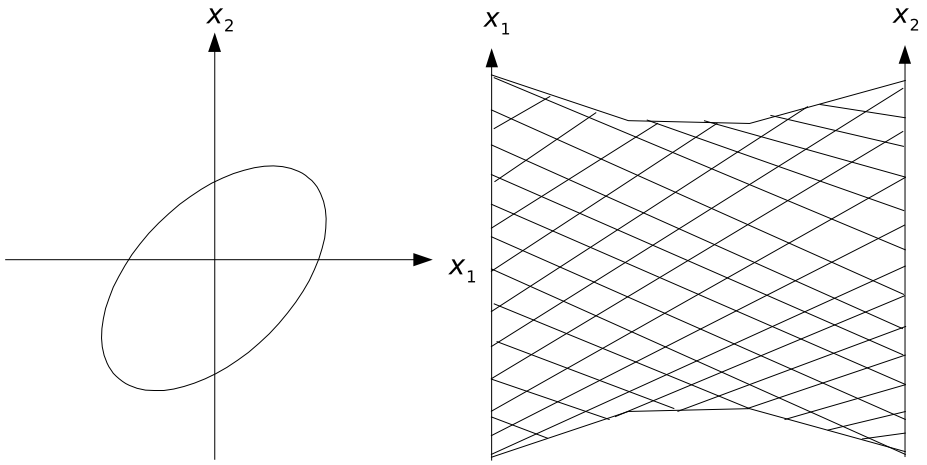


Figure E.8: Cartesian plot of points forming a two-dimensional ellipse and its transformation into Parallel coordinate; a line hyperbola with a point hyperbola as an envelope.

the clustering propagates through other dimensions. Figure E.9 (c) indicates the appearance of three clusters in both Cartesian plots and Parallel Coordinate plots. In neither projection do these clusters separate.

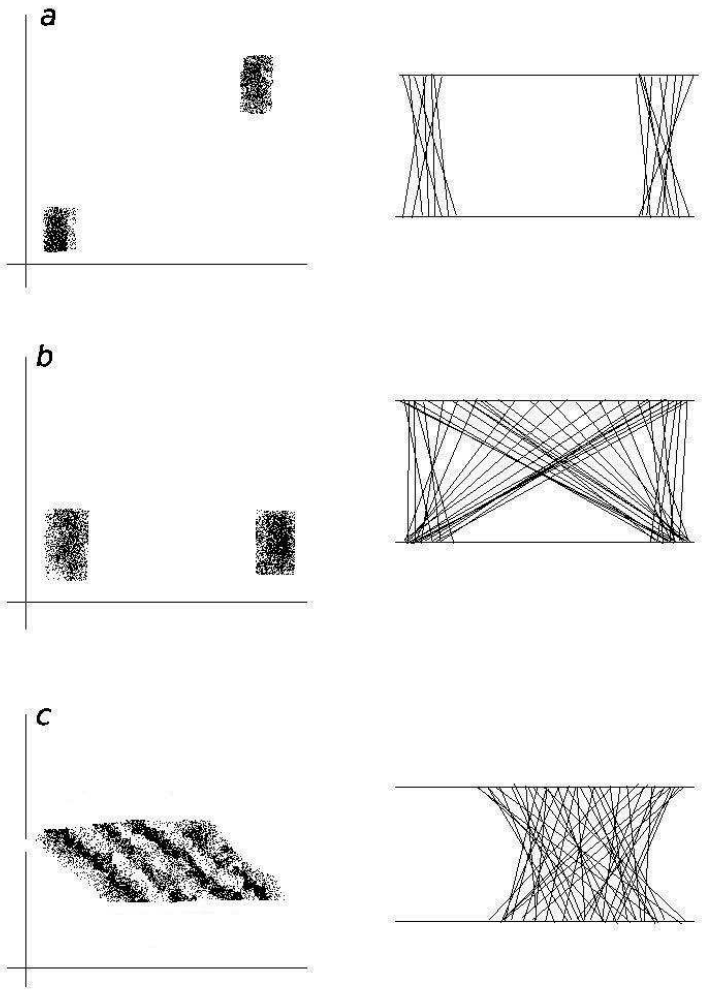


Figure E.9: (a) Clustering of data in Parallel Coordinates (right hand plots) that is separated in both x and y axis (conventional Cartesian plots shown on the left). (b) Clustering of data that is separated in the x but not in the y coordinate. (c) Clustering of data that is separated in neither projection.

E.5 MultiVariate analysis software

The Fisher

The following set of options given in Table E.1 where set when using the linear discriminant.

Option	value	Description
NbinsMVAPdf	10	Number of bins used for the p.d.fs of the classifier output
NsmoothMVAPdf	1	Number of smoothing iterations used per p.d.f

Table E.1: TMVA options for the Fisher classifier.

Performance

The Fisher discriminant depends (in-part) on the mean of each variable distribution for signal and background. This means in the situation where the sample means are the same, the Fisher gives very little discriminating power; even if the distributions are very different. If this situation is found, an improvement to the Fisher can be achieved by performing a transformation on the variables. For example, consider the idealistic situation for the Fisher discriminant, where the variables for signal and background have the following distribution shown in Figure E.10. On the contrary, the performance of the Fisher is seriously degraded if the variable distribution are similar to that shown in Situation 2; the middle plot of Figure E.10, where the variable means for signal and background are now identical. Situation 2 can be improved by first performing the transformation:

$$x'_S = |\mu_S - x_S|\,, \qquad x'_B = |\mu_B - x_B|$$

Performing this transformation leads to improved discrimination, as shown in the bottom plot of Figure E.10; the means for signal and background have now been shifted with respect to Situation 2. In the current study under investigation, many variables suffer the problem of their signal and background distribution means overlapping. From Figure E.16 for instance, it can be seen that the decays angles θ_{tr} and ϕ_{tr}, and the J/ψ mass have overlapping signal and background means. For these variables, the Fishers performance could be enhanced by taking the absolute value of these variables.

In certain cases. In particular when the discriminating variables are Gaussian distributed with linear correlation, the Fisher discriminate can compete with the Likelihood and the non-linear classifiers.

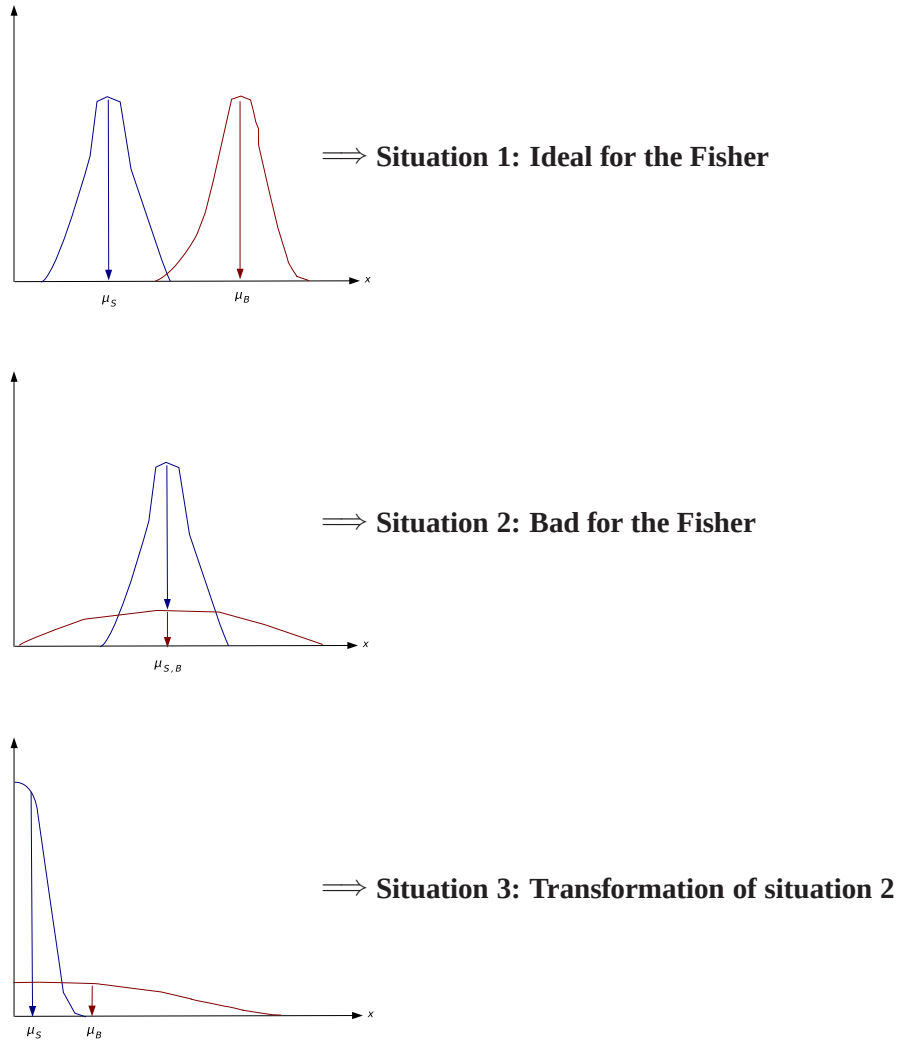


Figure E.10: Three situations for discriminating signal (red) and background (blue) using the Fisher discriminant. Situation 1 is ideal for the Fisher discriminant, since the signal and background distributions are well separated. The variable distribution in Situation 2, where the signal and background distributions overlap, is problematic for the Fisher discriminant. The variable distribution in Situation 2 is improved in Situation 3 by performing the variable transformation given in equation (E.13).

Likelihood and Decorrelated Likelihood

Both TMVA's Likelihood classifiers where used with the following options given in Table E.2:

Performance

Parameter based discriminators, including the Likelihood, can achieve good separation power when model inaccuracies can be removed. Inaccuracies includes correlations between input

Option	value	Description
PDFInterpol	2^{nd}	2^{nd} order spline.
		Smoothing function used for each input variable distribution
Nsmooth	5	Number of smoothing iterations used for each variable distribution
NAvEvtPerBin	10	Average number of events per bin
VarTransform		Decorrelation of the input variables: Compute the square-root of the covariance matrix. True for the Decorrelated Likelihood.

Table E.2: TMVA options for the Likelihood and Decorrelated Likelihood classifiers.

variables, which can be removed by variable decorrelation procedures; which is just a rotation of the correlated variable parameter space.

Since the parametric form for each variable is generally unknown, the shapes can be approximated to from the training data by fitting polynomial splines to these histograms.

This approach can lead to model inaccuracies, if there are too few statistics in the data sample on the number of events per bin is too high. For instance, if one of the samples (background or signal) had $\sim$ 600 events, with $\sim$ 300 of these being used for training and the other 300 for testing, and if NAvEvtPerBin was set to 50, then there would be $\sim$ 5 bins with which to fit the distribution, which is inefficiently small number, leading to model inaccuracies. For the study discussed in this chapter, the background sample (of prompt J/ψ's) with 702 events, split evenly for training and testing leads to $\sim$ 350 events, and therefore 6 bins if NAvEvtPerBin equals 50. For the current study therefore, NAvEvtPerBin has been set to 10, which leads to 35 bins to model each background variable distribution. Another problem parametrising the variable shapes using splines is the amount of smoothing applied to the variable distribution, via the Nsmooth option. This option is used to remove the statistical fluctuations in the variable distribution. If Nsmooth is set too high, say 100, it can destroy non-statistical information removing much of the distributions shape. A high Nsmooth number is only beneficial to distributions with very little structure; almost flat distributions. On the contrary, variable distributions which have narrow structures, like peaks, require very little smoothing and a small Nsmooth number. As the majority of the variable distributions in this study are strongly peaked (Figure E.16), NSmooth has been set to a low value 5.An alternative to splines is to use an unbinned kernel estimator (KDE). This is also an non-parametric approximation, but now the shapes of the p.d.fs are obtained from unbinned data[10]. KDEs also include smoothing functionality.

The performance of the Likelihood classifiers are heavily dependent on accurate determination of each variable shape. This requires sufficient training statistics to populate each variable distribution and reducing the input variable correlations in the model. The problem with like-

[10] In the end, TMVA uses binned histograms in order to increase computational speed.

lihood classifier is that is ignores correlation amongst the discriminating input variables. In realistic situations this can be problematic and degrades the classifier performance. If the input variables in the training sample are found to be linearly correlated and Gaussian distributed, use can be made of a decorrelated Likelihood to improve the classifiers performance. Here the correlations between variables can be removed by computing the square-root C' of the correlation matrix C, namely $C = C'C'$ each variable can the be transformed from the original (x_i) into the decorrelated variable space (x'_i) by: $x' = C'^{-1}x$. This procedure cannot handle circular correlations however, and if applied to variables which are not linearly correlated with Gaussian distributions can even degrade the classifiers performance. In many studies use a large number of discriminating variables, the variable correlations can often be irreducible and difficult to account for. One approach [119] before using each classifier is to reduce further the list of input parameters, by removing those which are highly correlated. This procedure of reducing variable correlations has been adopted in this chapter to help improve the Likelihood's performances.

MultiLayer Perceptron & Clermont-Ferrand Artifical Neural network (CFANN)

The MLP and CF[11] neural network where used with the following architecture, given in Table E.4:

Option	value	Description
N_{var}	31, 12 (reduced set)	Number of input variables
Number of classes	2	signal & backgrounds
HiddenLayer	4	input/output layers plus 2 hidden layers
Number of neurons per layer	31,32,31,2(full set) 12,13,12,2(reduced set)	This comprises of N_{var} for the input layer, one for variable. $N_{var}+1$ for the second layer, each N_var for the third and 2 for the output layer.
NCycles	200	Number of trying cycles.
NeutronType	sigmoid	Hidden layer activation function

Table E.3: TMVA options for the MLP and the CF classifiers.

Performance

Neural networks are designed to deal with correlation amongst the input variables, as opposed to Likelihood classifiers. Of the two neural networks used, the CF is found to preform least well and computationally more expensive than the MLP. In this present study, no optimisation for the number of neurons in each hidden layer was performed: often, the number of neurons per layer is chosen to minimise the network's error function [99].

[11] In the TMVA nomenclature this classifier is called CFMlpANN

E.5.1 BDT

The Boosted Decision Tree was used with the options set given in Table E.4:

Option	value	Description
NTrees	400	Number of trees in the forest
BoostType	AdaBoost	Boosting Algorithm used for each tree
SeparationType	GiniIndex	Separation criterion or node splitting
nCuts	20	Number of steps during each node cut optimisation
PruneStrength	4.5	Pruning strength
PruneMethod	CostComplexity	Method used to remove (or prune) statistically insignificant branches

Table E.4: TMVA options for the BDT classifier.

Performance

The major advantage of boosted decision trees over other classifiers includes their stability and ability to handle large numbers of input variables. Often BDT require little tuning to produce good results, whilst being reasonably insensitive to poorly discriminating variables. This is because the BDT algorithm picks the best discriminating variable at each node splitting: variables which provide inefficient separation are not used. This is different to ANNs, their performance can often suffer if the additional information (variables) provides no separation power. On the contrary, using a small number of variables, ANNs are found to be competitive with BDTs [120]. Also, given sufficient training events and specifying sufficient leaves, the Boosting algorithms almost always converges to its optimal value [105].

E.5.2 Parallel Coordinates

For the multivariate visualisation studies use was made of two packages. The standard Parallel Coordinates package available in Root (version 5.17) and the dedicated visualisation package CrystalVision [121], created by E. Wegman. The Root based Parallel Coordinate package has the advantage over CrystalVision since it is already integrated within analysis framework. This enables a direct comparison between the multivariate classifiers embedded in Root, from the TMVA toolkit, with the Parallel Coordinate visualisation technique. The drawback however, which can be seen from Figures E.5.2, E.5.2 and E.5.2, is that the Root based implementation has no effective solution for tackling over-plotting[12], CrystalVison uses a technique called α blending. This is where the transparency of each pixel, as well as other forms of image composition, are dealt with in the hardware by a fourth component, the α-channel, in the RGB vector, (R,G,B,α).

[12] the situation where lines plotted on Parallel coordinate diagrams overlap each other.

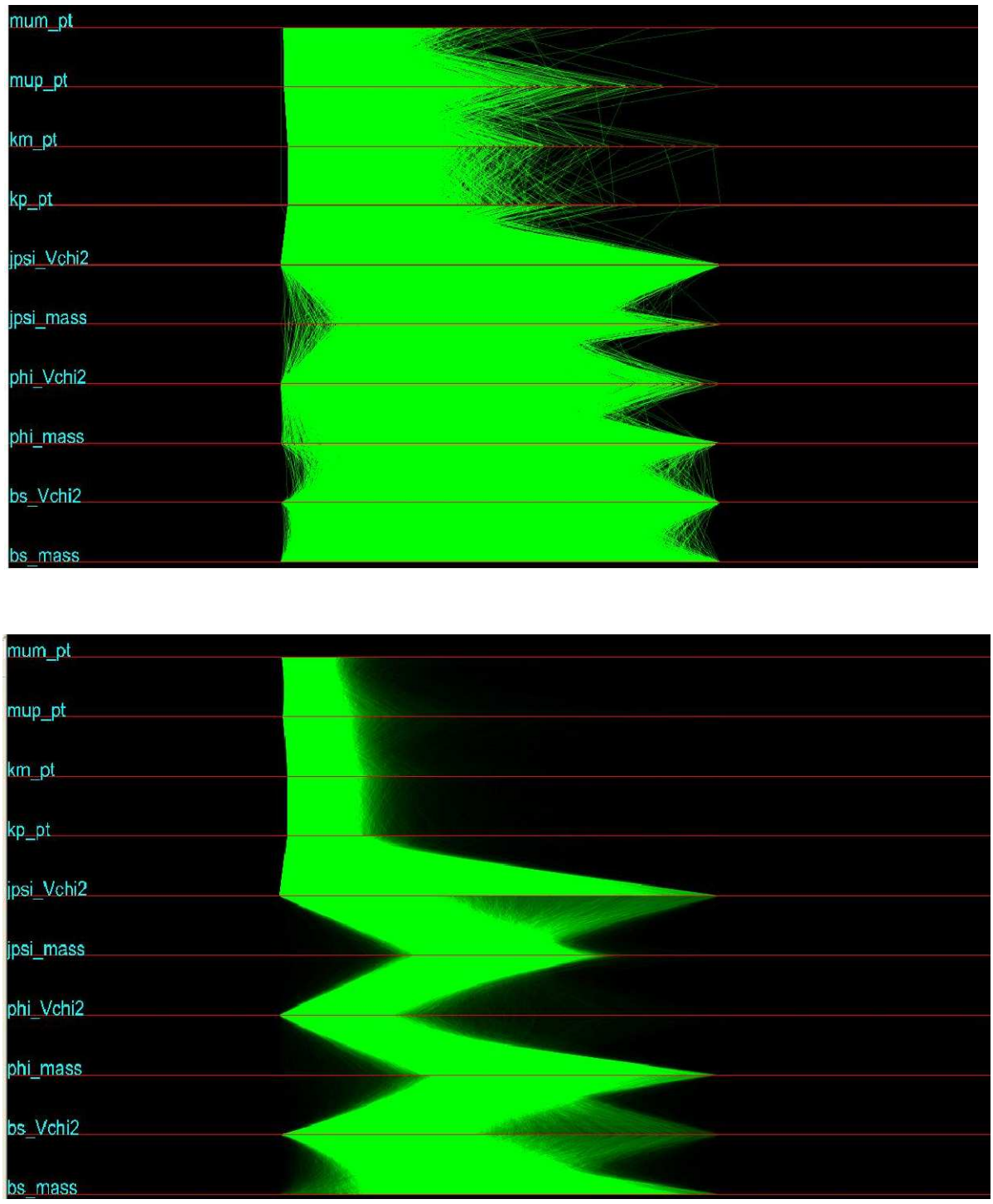


Figure E.11: Parallel Coordinate plots produced by the CystalVision software package. The Top plot has no α blending applied, whilst the bottom plot shows the effect of α blending. The Example shows the $B_s \to J/\psi\phi$ event data for each variable used in the selection, which includes: $p_t(\mu^{\pm}, K^{\pm})$, $\chi^2(\mu^+\mu^-)$, $\Delta M_{J/\psi}(\mu^+\mu^-)$,$\chi^2(K^+K^-)$,$\Delta M_{\phi}(K^+K^-)$, $\chi^2(B_s)$,ΔM_{B_s}.

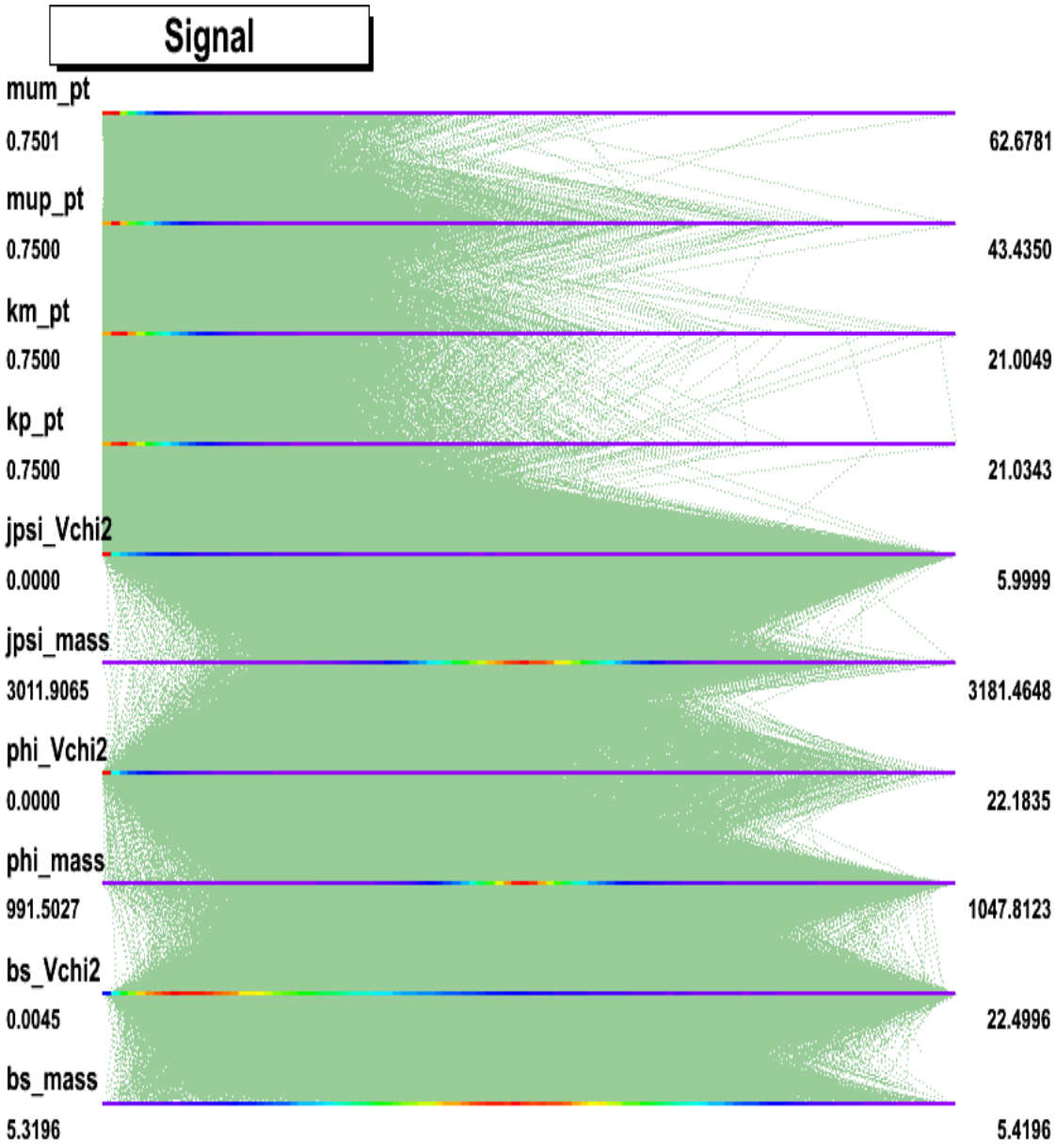


Figure E.12: Parallel Coordinate plots produced by the Root (version 5.17) software package. The plot shows the effect of overplotting. The Example shows the $B_s \rightarrow J/\psi\phi$ event data for each variable used in the selection, which includes: $p_t(\mu^{\pm}, K^{\pm})$, $\chi^2(\mu^+\mu^-)$, $\Delta M_{J/\psi}(\mu^+\mu^-)$,$\chi^2(K^+K^-)$,$\Delta M_{\phi}(K^+K^-)$, $\chi^2(B_s)$,ΔM_{B_s}.

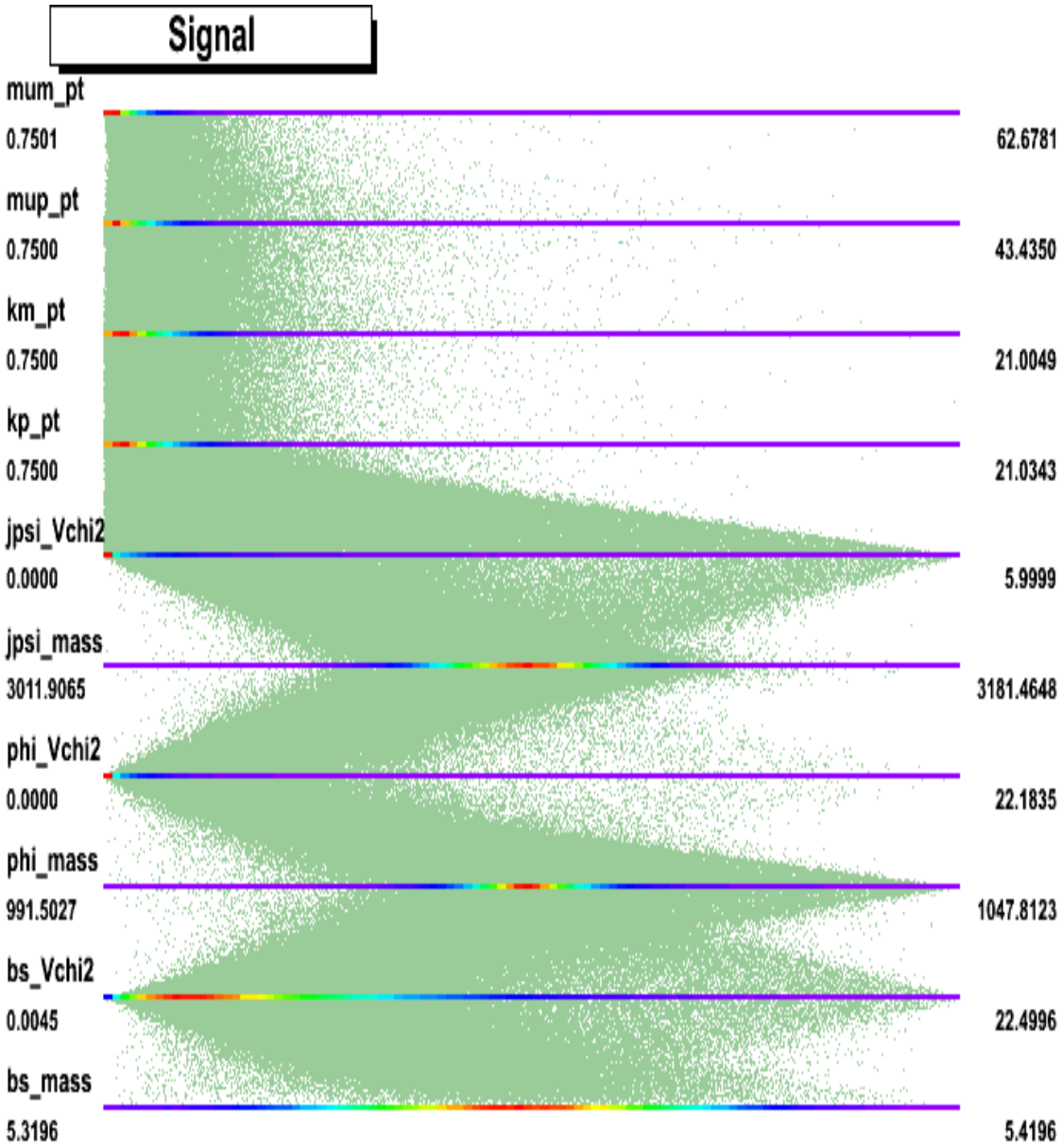


Figure E.13: Parallel Coordinate plots produced by the Root (version 5.17) software package. The plot shows the reduction in overplotting by increasing the pixelation of each line. The Example shows the $B_s \to J/\psi\phi$ event data for each variable used in the selection, which includes: $p_t(\mu^{\pm}, K^{\pm})$, $\chi^2(\mu^+\mu^-)$, $\Delta M_{J/\psi}(\mu^+\mu^-)$,$\chi^2(K^+K^-)$,$\Delta M_{\phi}(K^+K^-)$, $\chi^2(B_s)$,ΔM_{B_s}.

E.6 Discriminating variables

- **bs_e**: The energy associated with each B_s candidate.
- **bs_p**: The momentum associated with each B_s candidate.
- **bs_pt**: The transverse momentum associated with each B_s candidate.
- **bs_tau**: The propertime of each B_s candidate.
- **bs_tau_err**: The propertime error associated with each B_s.
- **bs_tau_sig**: The propertime significance of each B_s candidate.
- **bs_dist**: The distance for the B_s decay vertex to the primary vertex position[13]
- **bs_dist_sig**: The distance significance of the B_s decay vertex to the primary vertex position. given as: $\frac{bs\_dist}{\sigma(bs\_dist)}$
- **bs_ip**: The impact parameter of each B_s candidate with respect to the primary vertex[14].
- **bs_ip_sig**: This impact parameter significance of each candidate with respect to the primary vertex, give as: $\frac{bs\_ip}{\sigma(bs\_ip)}$
- **theta**: The polar decay angle, θ_{tr}, defined in the J/ψ rest frame between the μ^+ and the z-axis.
- **psi**: The azimuthal decay angle, ϕ_{tr}, defined in the x-y plane in the J/ψ rest frame.
- **phi**: The polar decay angle, θ_ϕ, defined between the K^+ and the x-axis in the ϕ rest frame.
- **phi_mass**: The mass of each the ϕ candidate.
- **phi_e**: The energy of each ϕ candidate.

[13] this has been calculated using the standard DaVinci method, calcVertexDis, found in the Davinci GeomDispCalculator class. This method calculates the absolute distance and errors between the two vertices.

[14] This is again calculated using the DaVinci GeomDispCalculator, using the calcImpactPar method. This returns the distance of closest approach between and errors between the given particle and vertex.

- **phi_pt**: The transverse momentum of each ϕ candidate.
- **jpsi_mass**: The mass of each J/ψ candidate.
- **jpsi_e**: The energy of each J/ψ candidate.
- **jpsi_pt**: The transverse momentum of each J/ψ candidate.
- **kp_p**: The momentum of each K^+ candidate.
- **kp_dll**: The global delta log likelihood hypothesis of each K^+ candidate with respect to pion: $\Delta ln\mathcal{L}_{K^+-\pi}$.
- **kp_Tchi2**: The track χ^2/dof associated to each K^+ candidate.
- **km_p**: The momentum of each K^- candidate.
- **km_dll**: The global delta log likelihood hypothesis of each K^- candidate with respect to pion: $\Delta ln\mathcal{L}_{K^--\pi}$.
- **km_Tchi2**: The track χ^2/dof associated to each K^- candidate.
- **mup_p**: The momentum of each μ^+ candidate.
- **mup_dll**: The global delta log likelihood hypothesis of each μ^+ candidate with respect to pion: $\Delta ln\mathcal{L}_{\mu^+-\pi}$.
- **mup_Tchi2**: The track χ^2/dof associated to each μ^+ candidate.
- **mum_p**: The momentum of each μ^- candidate.
- **mum_dll**: The global delta log likelihood hypothesis of each μ^- candidate with respect to pion: $\Delta ln\mathcal{L}_{\mu^--\pi}$.
- **mum_Tchi2**: The track χ^2/dof of associated to each μ^- candidate.

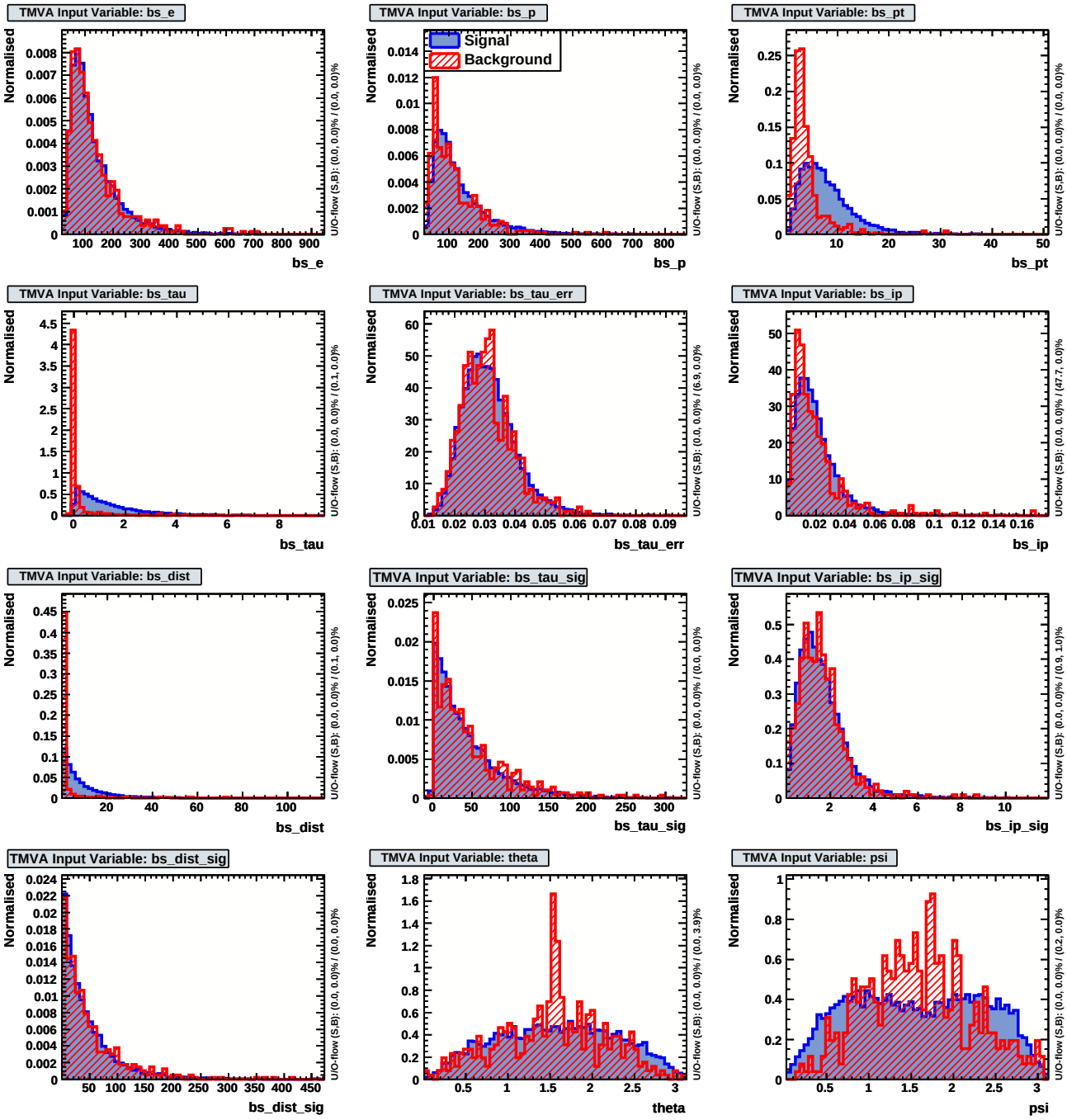


Figure E.14: Variable distributions for signal (solid blue) and background (hatched red) for Set 1, of input attributes initially used; the total number of attributes in this set is 31.

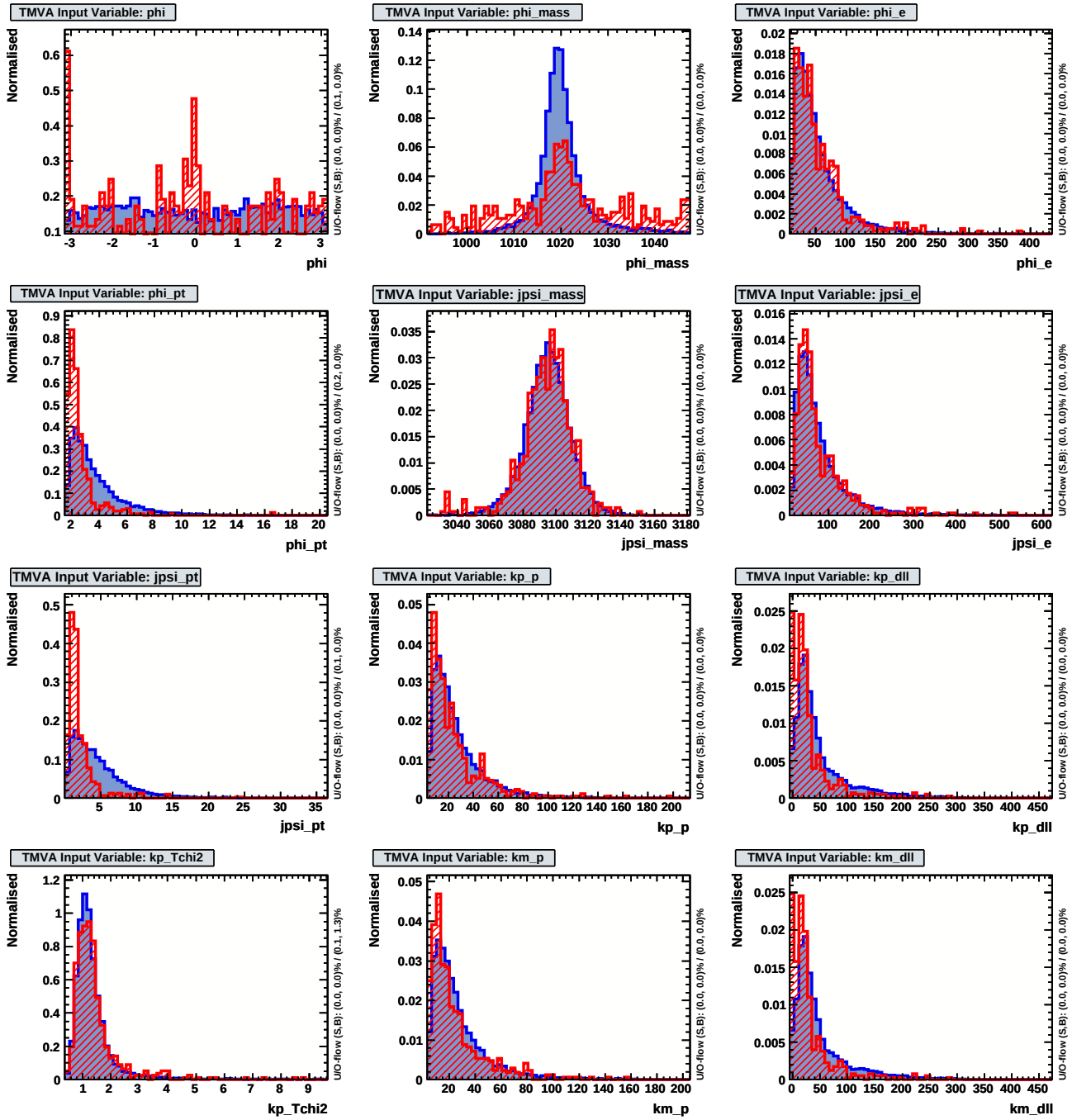


Figure E.15: Variable distributions for signal (solid blue) and background (hatched red) for Set 1, of input attributes initially used; the total number of attributes in this set is 31.

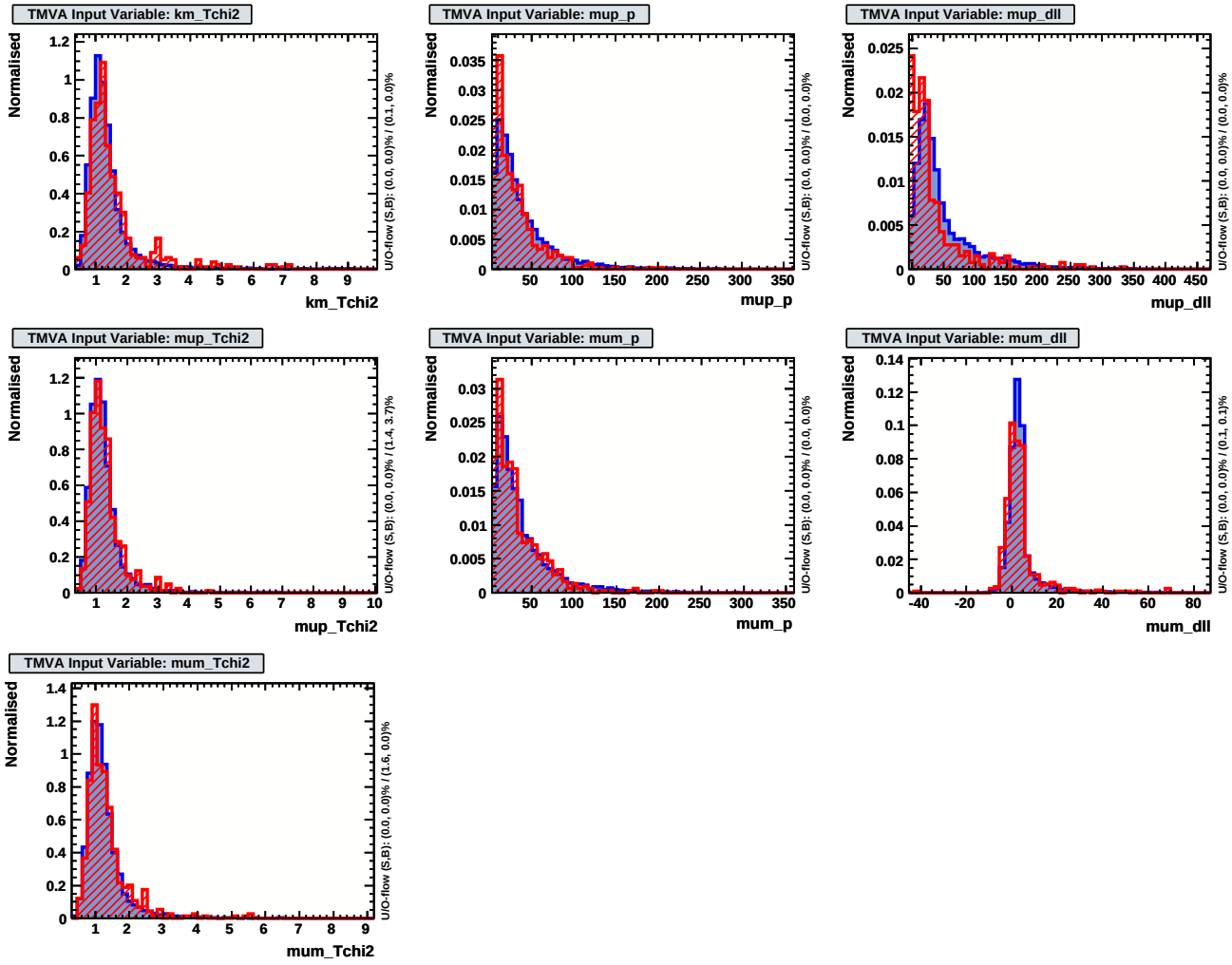


Figure E.16: Variable distributions for signal (solid blue) and background (hatched red) for Set 1, of input attributes initially used; the total number of attributes in this set is 31.

E.7 Correlation matrices for sets 2

	bs_tau	bs_tau_err	bs_ip	theta	psi	phi	phi_mass	jpsi_mass	kp_Tchi2	km_Tchi2	mup_dll	mup_Tchi2
mup_Tchi2	0	2	0	0	1	-1	0	0	2	2	-3	100
mup_dll	-1	1	4	0	-5	0	0	0	-10	-7	100	-3
km_Tchi2	0	-1	0	0	-2	0	0	1	7	100	-7	2
kp_Tchi2	0	-1	0	-1	1	-1	-1	1	100	7	-10	2
jpsi_mass	0	0	0	0	1	1	0	100	1	1	0	0
phi_mass	0	-1	0	0	0	0	100	0	-1	0	0	0
phi	-2	0	1	0	-6	100	0	1	-1	0	0	-1
psi	0	-1	0	0	100	-6	0	1	1	-2	-5	1
theta	0	1	1	100	0	0	0	0	-1	0	0	0
bs_ip	4	24	100	1	0	1	0	0	0	0	4	0
bs_tau_err	15	100	24	1	-1	0	-1	0	-1	-1	1	2
bs_tau	100	15	4	0	0	-2	0	0	0	0	-1	0

Table E.5: Correlation matrices for the signal variables found in Set 2.

	bs_tau	bs_tau_err	bs_ip	theta	psi	phi	phi_mass	jpsi_mass	kp_Tchi2	km_Tchi2	mup_dll	mup_Tchi2
mup_Tchi2	0	1	0	-2	-1	-3	0	4	-2	4	-5	100
mup_dll	11	4	4	-4	-7	8	-5	1	-7	-17	100	-5
km_Tchi2	-2	-2	1	0	-3	-15	-3	1	8	100	-17	4
kp_Tchi2	0	-4	15	-9	10	-1	1	-2	100	8	-7	-2
jpsi_mass	7	3	0	1	-3	-5	3	100	-2	1	1	4
phi_mass	1	1	0	-4	-8	-9	100	3	1	-3	-5	0
phi	1	-1	-1	0	1	100	-9	-5	-1	-15	8	-3
psi	-4	-3	0	-3	100	1	-8	-3	10	-3	-7	1
theta	1	5	1	100	-3	0	-4	1	-9	0	4	-2
bs_ip	35	-1	100	1	0	-1	0	0	15	1	-4	0
bs_tau_err	2	100	-1	5	-3	-1	1	3	-4	-2	4	1
bs_tau	100	2	35	1	-4	1	1	7	0	-2	11	0

Table E.6: Correlation matrices for the background variables found in Set 2.

E.8 Classifiers response to variable set 2

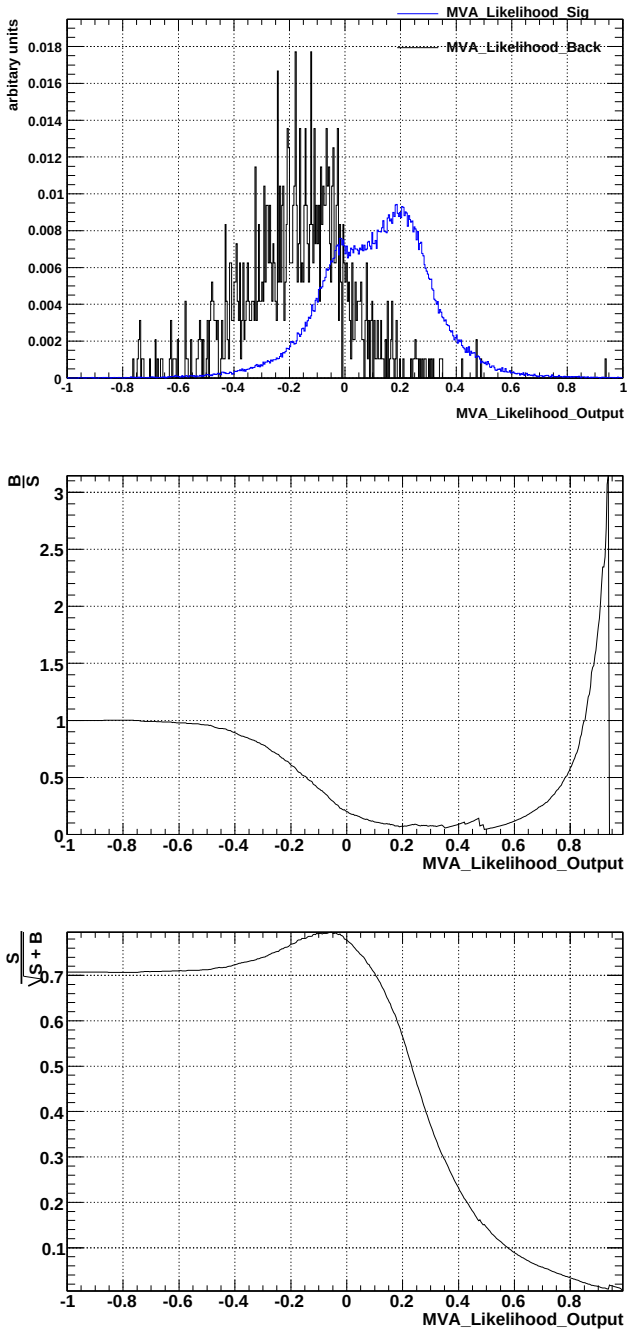


Figure E.17: Normalised response of signal (blue) and background (black) to the classifiers on the reduced data set, including propertime and decay angles: (Left) the classifiers response for signal and background (Middle), the $\frac{B}{S}$ response of the classifier and (Right) the $\frac{S}{\sqrt{S+B}}$.

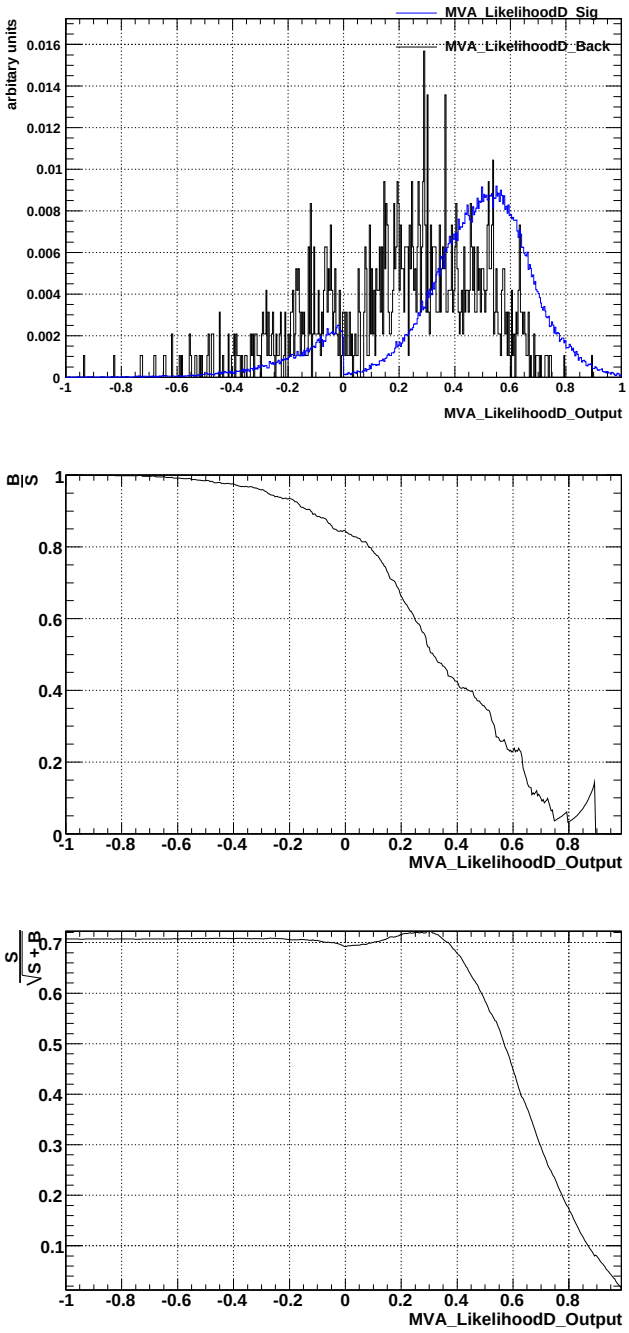


Figure E.18: Normalised response of signal (blue) and background (black) to the classifiers on the reduced data set, including propertime and decay angles: (Left) the classifiers response for signal and background (Middle), the $\frac{B}{S}$ response of the classifier and (Right) the $\frac{S}{\sqrt{S+B}}$.

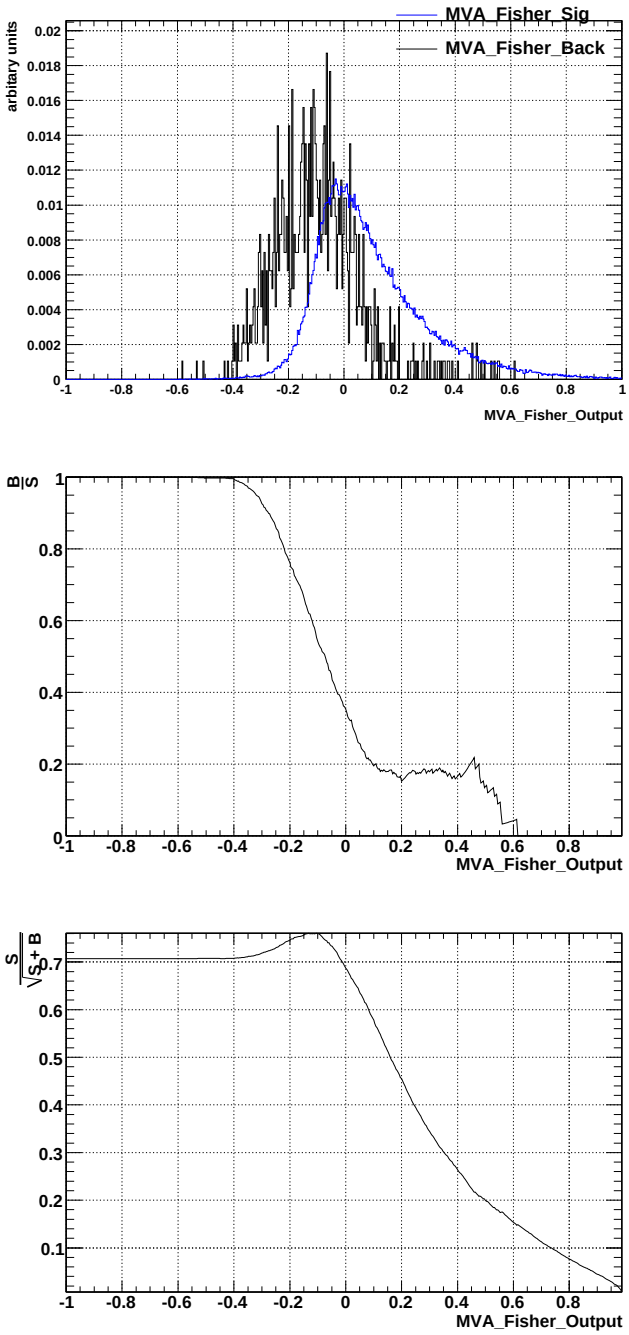


Figure E.19: Normalised response of signal (blue) and background (black) to the classifiers on the reduced data set, including propertime and decay angles: (Left) the classifiers response for signal and background (Middle), the $\frac{B}{S}$ response of the classifier and (Right) the $\frac{S}{\sqrt{S+B}}$.

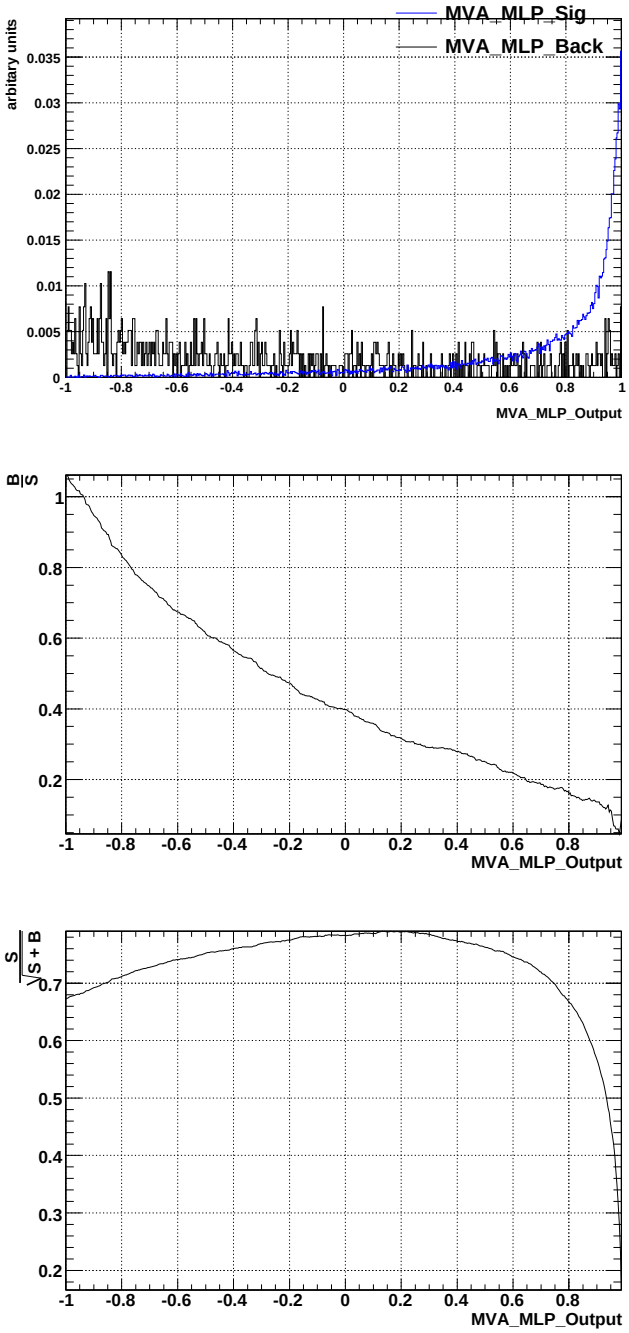


Figure E.20: Normalised response of signal (blue) and background (black) to the classifiers on the reduced data set, including propertime and decay angles: (Left) the classifiers response for signal and background (Middle), the $\frac{B}{S}$ response of the classifier and (Right) the $\frac{S}{\sqrt{S+B}}$.

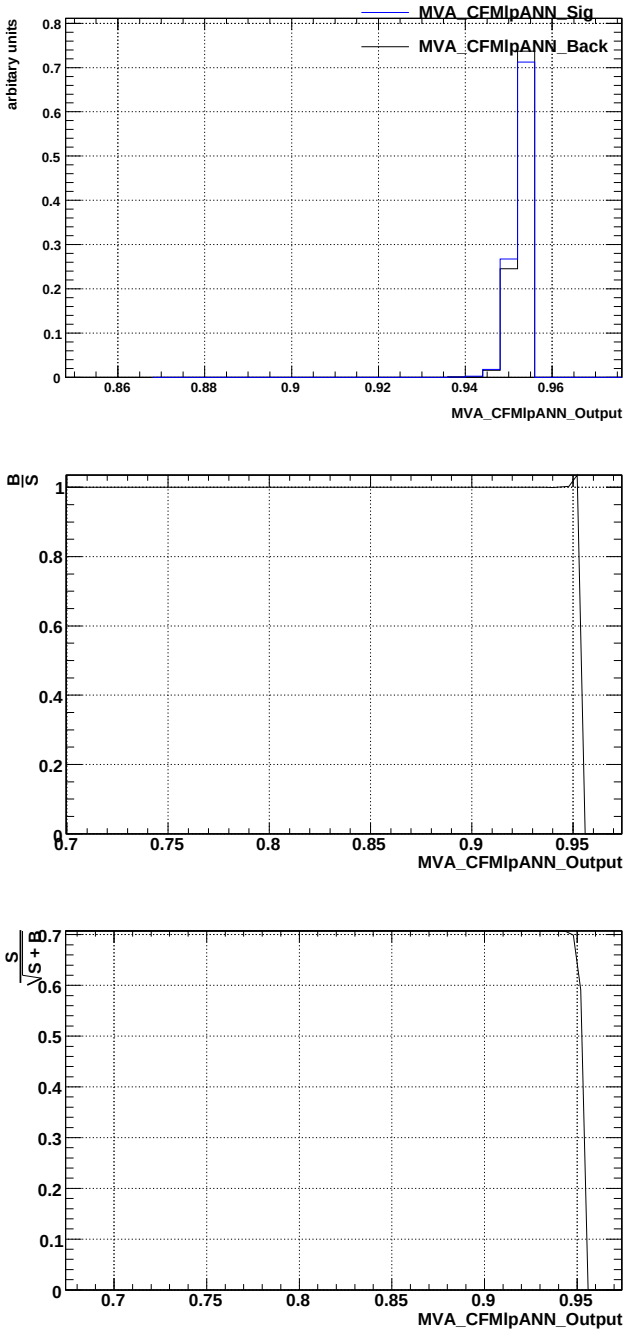


Figure E.21: Normalised response of signal (blue) and background (black) to the classifiers on the reduced data set, including propertime and decay angles: (Left) the classifiers response for signal and background (Middle), the $\frac{B}{S}$ response of the classifier and (Right) the $\frac{S}{\sqrt{S+B}}$.

E.9 Over training check for classifiers over variable Set 3

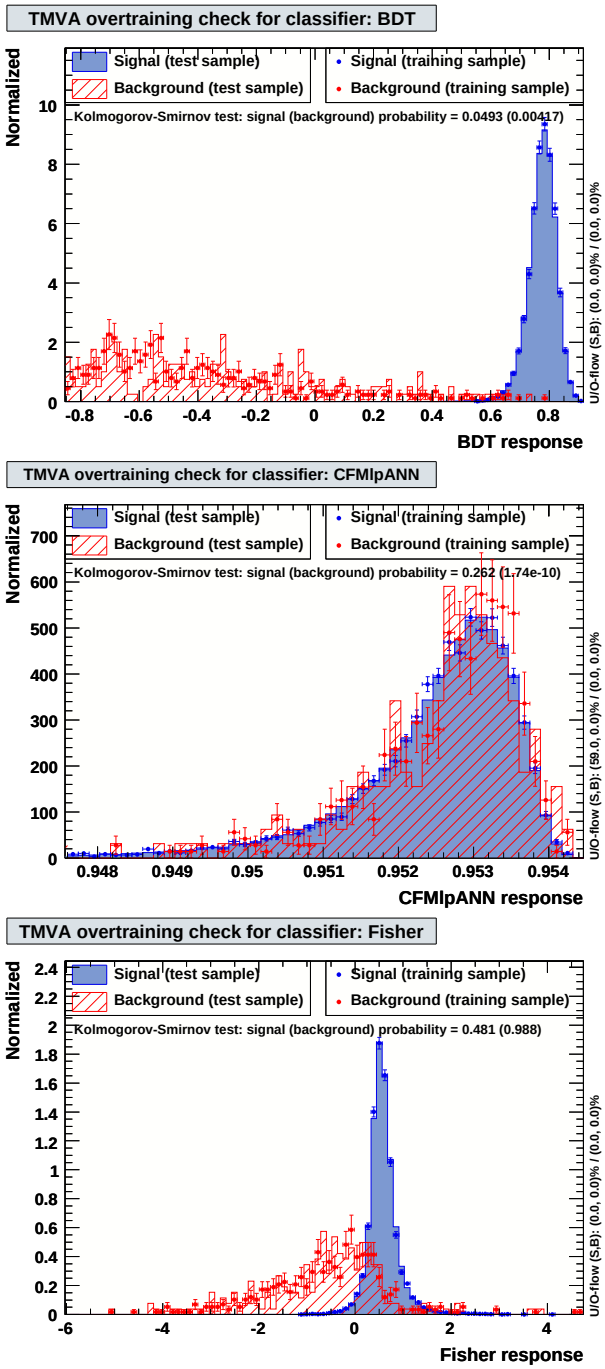


Figure E.22: Over training checks of the BDT, CF and Fisher distriminate classifiers over the variable set 3. The check uses the Kolmogorov-Smirnov Confidence Level, which varies from 0 to 1.

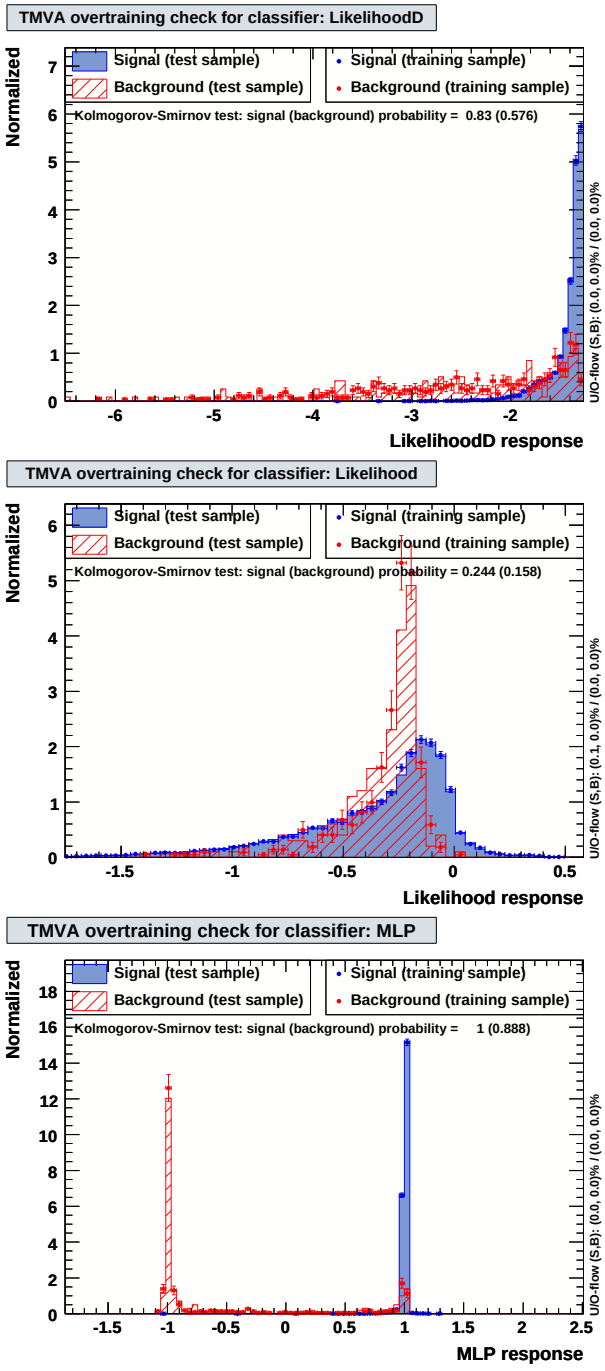


Figure E.23: Over training checks of the LikelihoodD, Likelihood and MLP classifiers over the variable set 3. The check uses the Kolmogorov-Smirnov Confidence Level, which varies from 0 to 1.

E.10 Pull distributions of $-2\beta_s\ \Delta\Gamma_s\ \bar{\Gamma}_s\ R_\perp$ and R_0

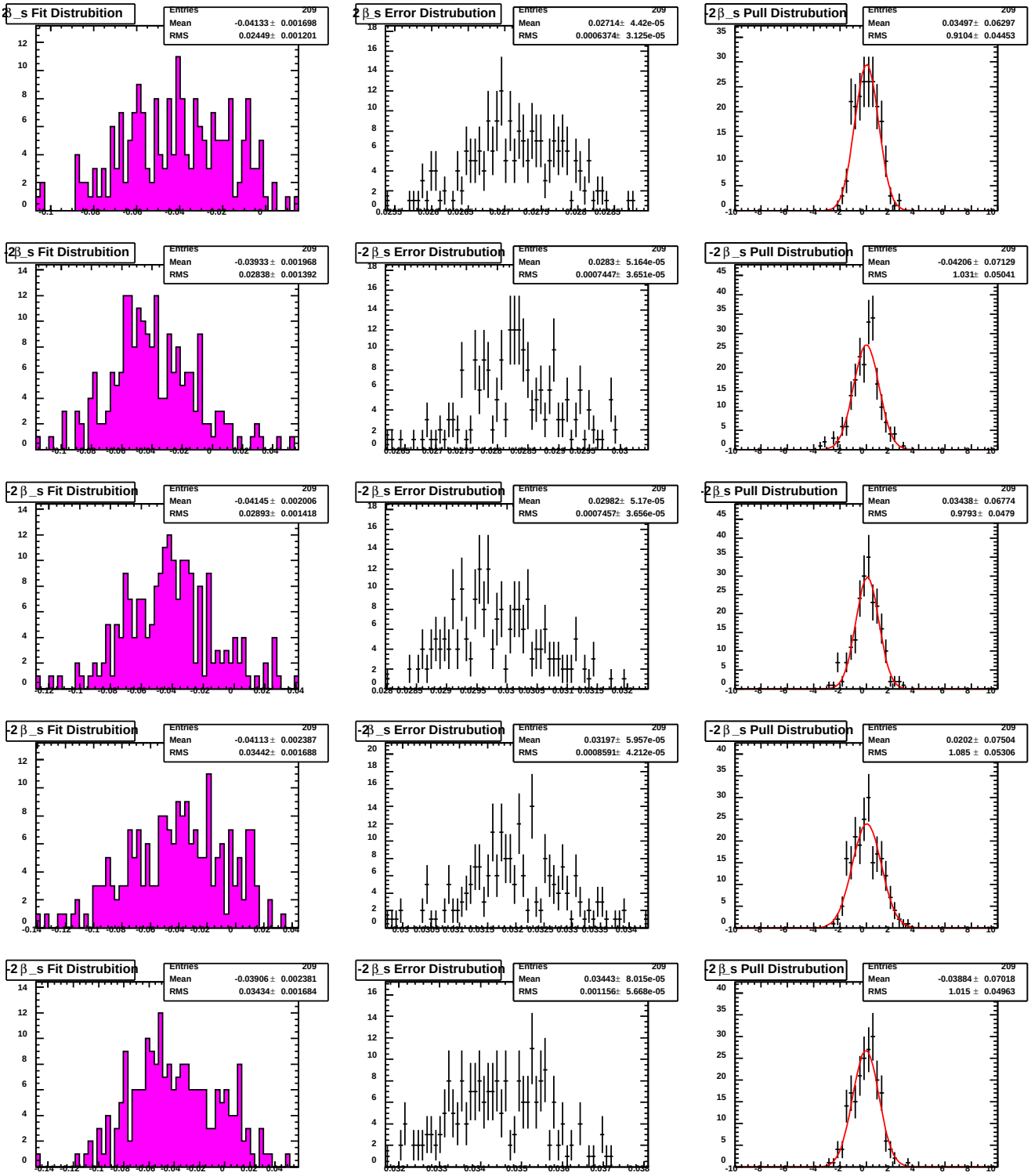


Figure E.24: $-2\beta_s$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{100\%,\ 90\%,\ 80\%,\ 70\%,\ 60\%\}$ of the signal.

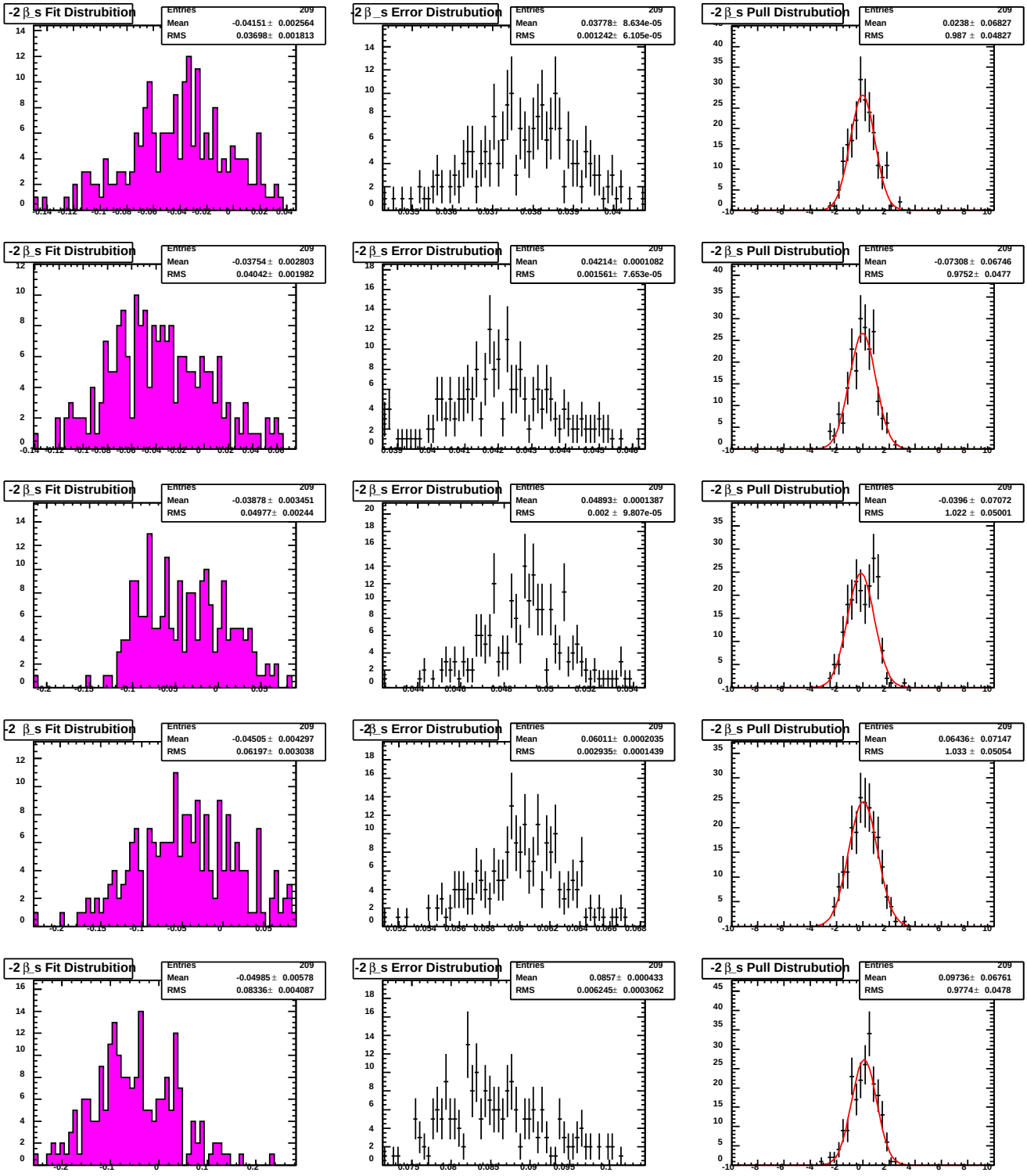


Figure E.25: $-2\beta_s$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{50\%,\ 40\%,\ 30\%,\ 20\%,\ 10\%\}$ of the signal.

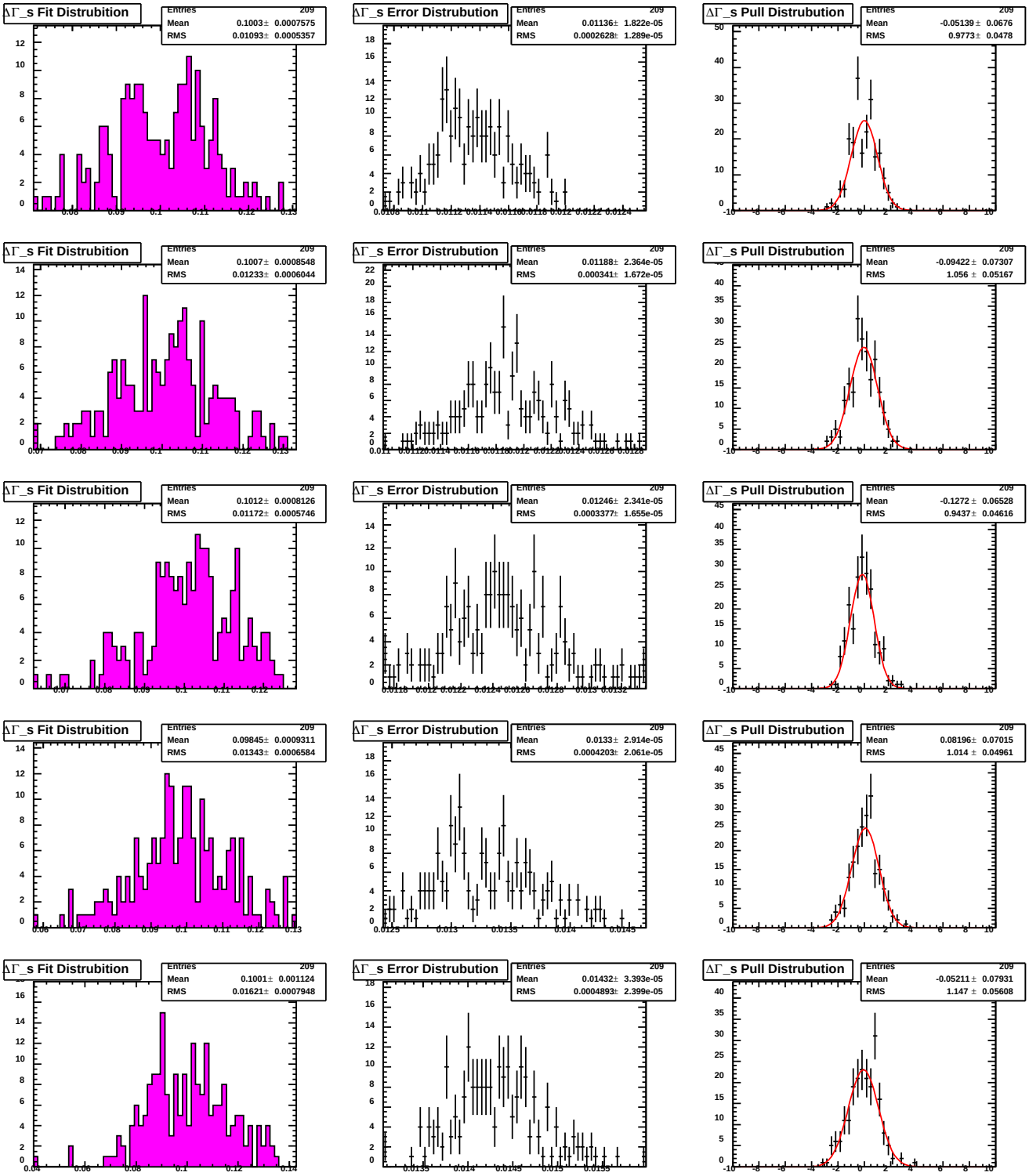


Figure E.26: $\Delta\Gamma_s$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{100\%,\ 90\%,\ 80\%,\ 70\%,\ 60\%\}$ of the signal.

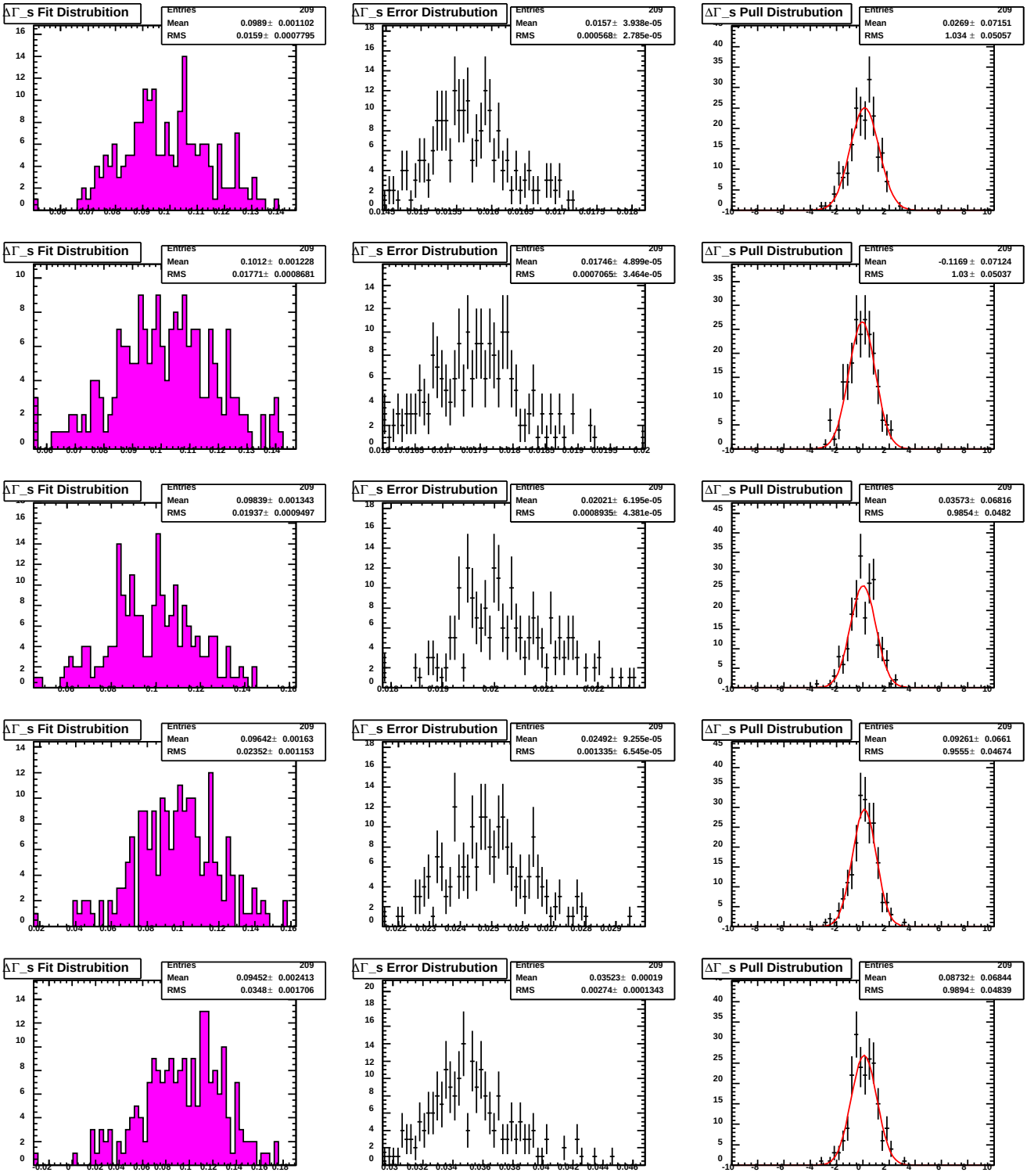


Figure E.27: $\Delta\Gamma_s$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{50\%,\ 40\%,\ 30\%,\ 20\%,\ 10\%\}$ of the signal.

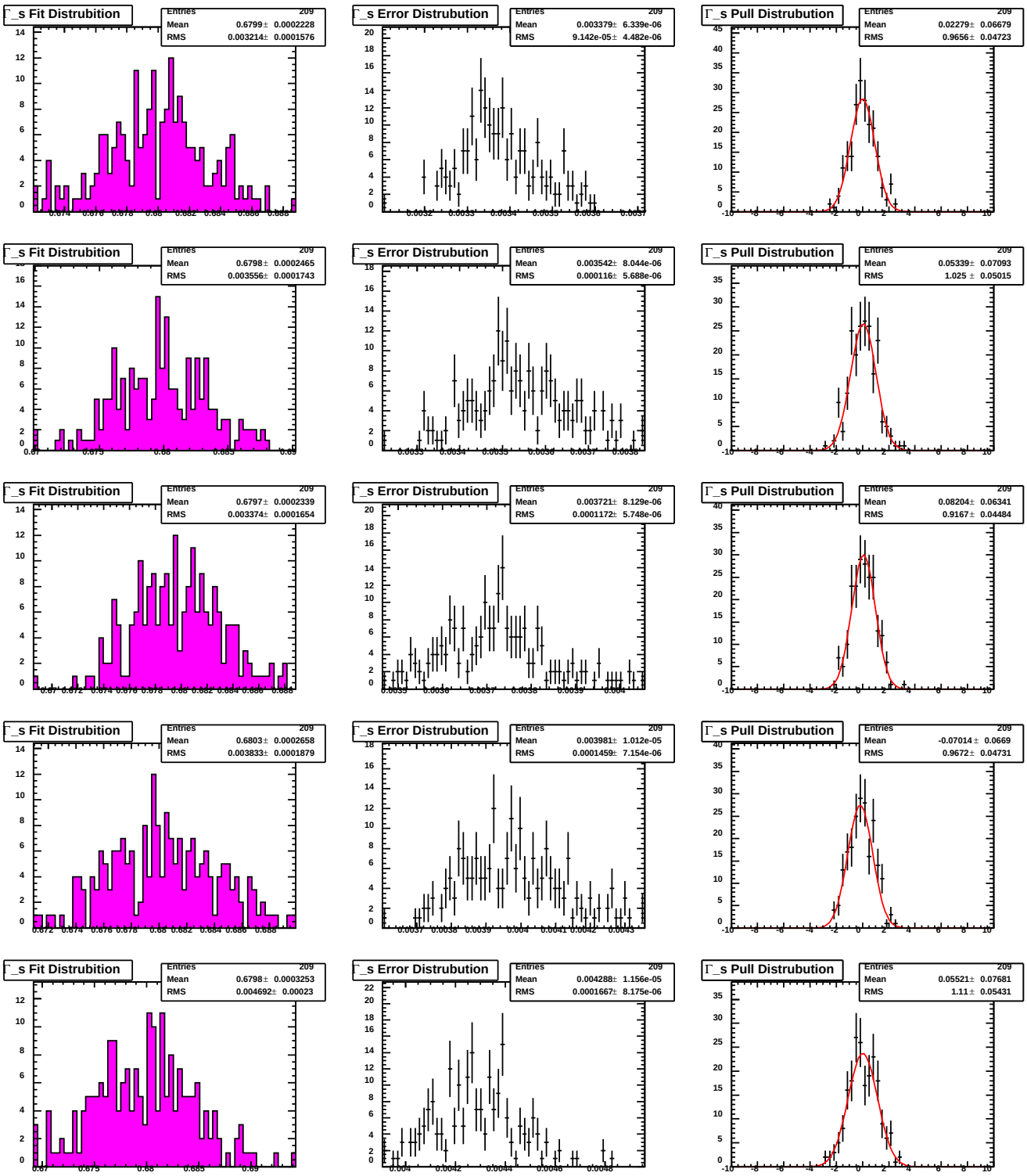


Figure E.28: $\bar{\Gamma}_s$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{100\%,\ 90\%,\ 80\%,\ 70\%,\ 60\%\}$ of the signal.

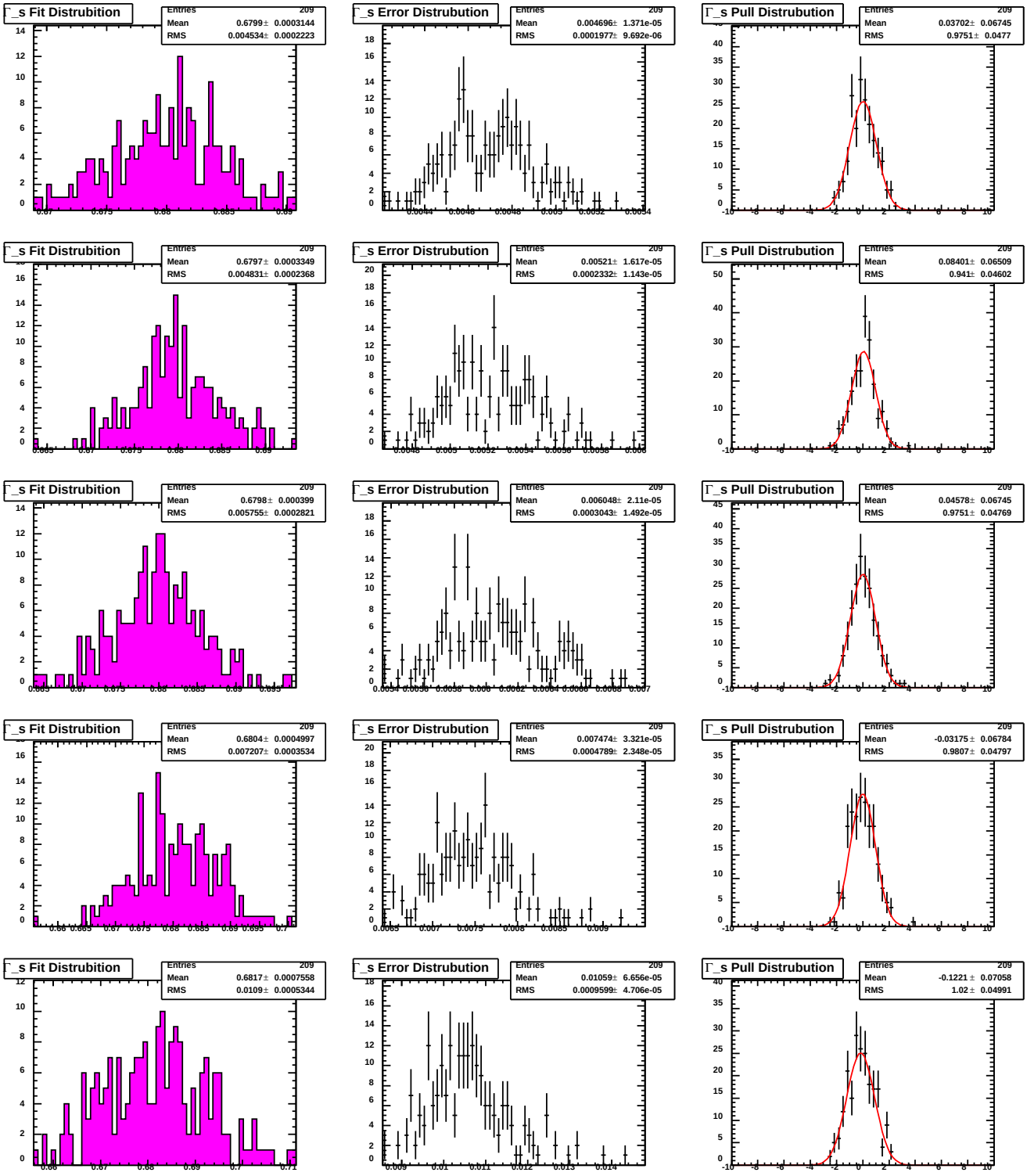


Figure E.29: $\bar{\Gamma}_s$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{50\%,\ 40\%,\ 30\%,\ 20\%,\ 10\%\}$ of the signal.

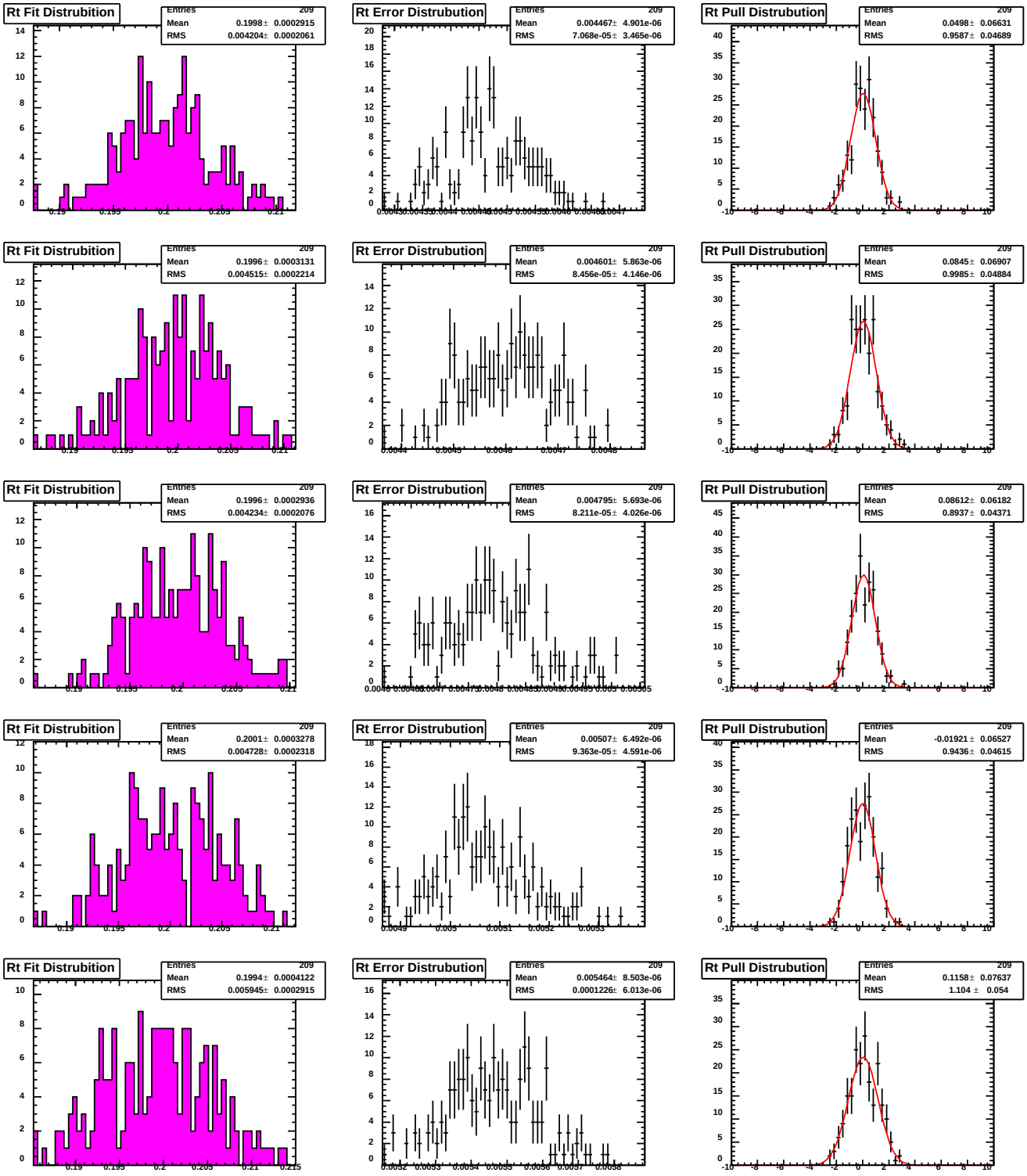


Figure E.30: $R_\perp$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{100\%,\ 90\%,\ 80\%,\ 70\%,\ 60\%\}$ of the signal.

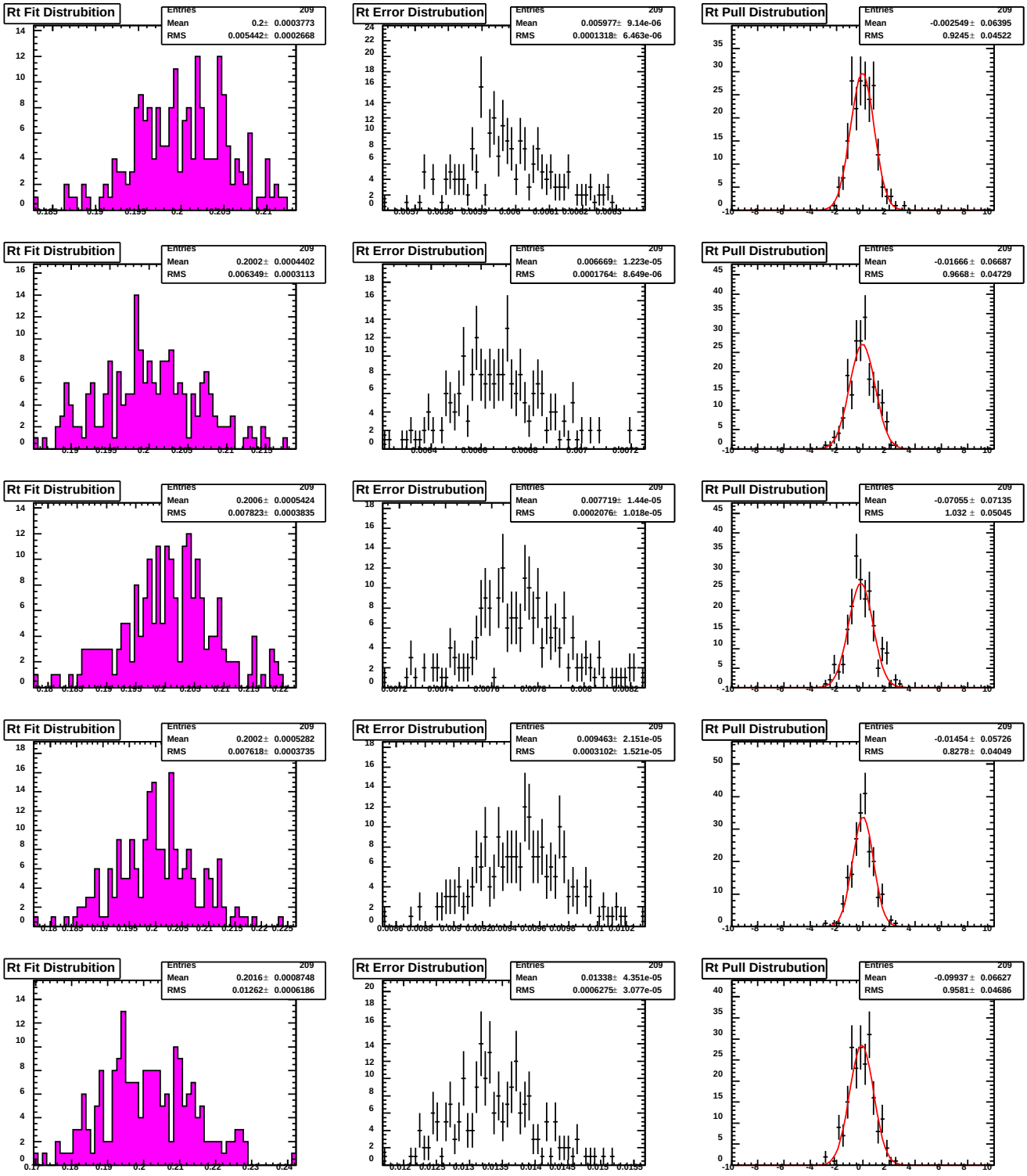


Figure E.31: $R_\perp$ sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{50\%,\ 40\%,\ 30\%,\ 20\%,\ 10\%\}$ of the signal.

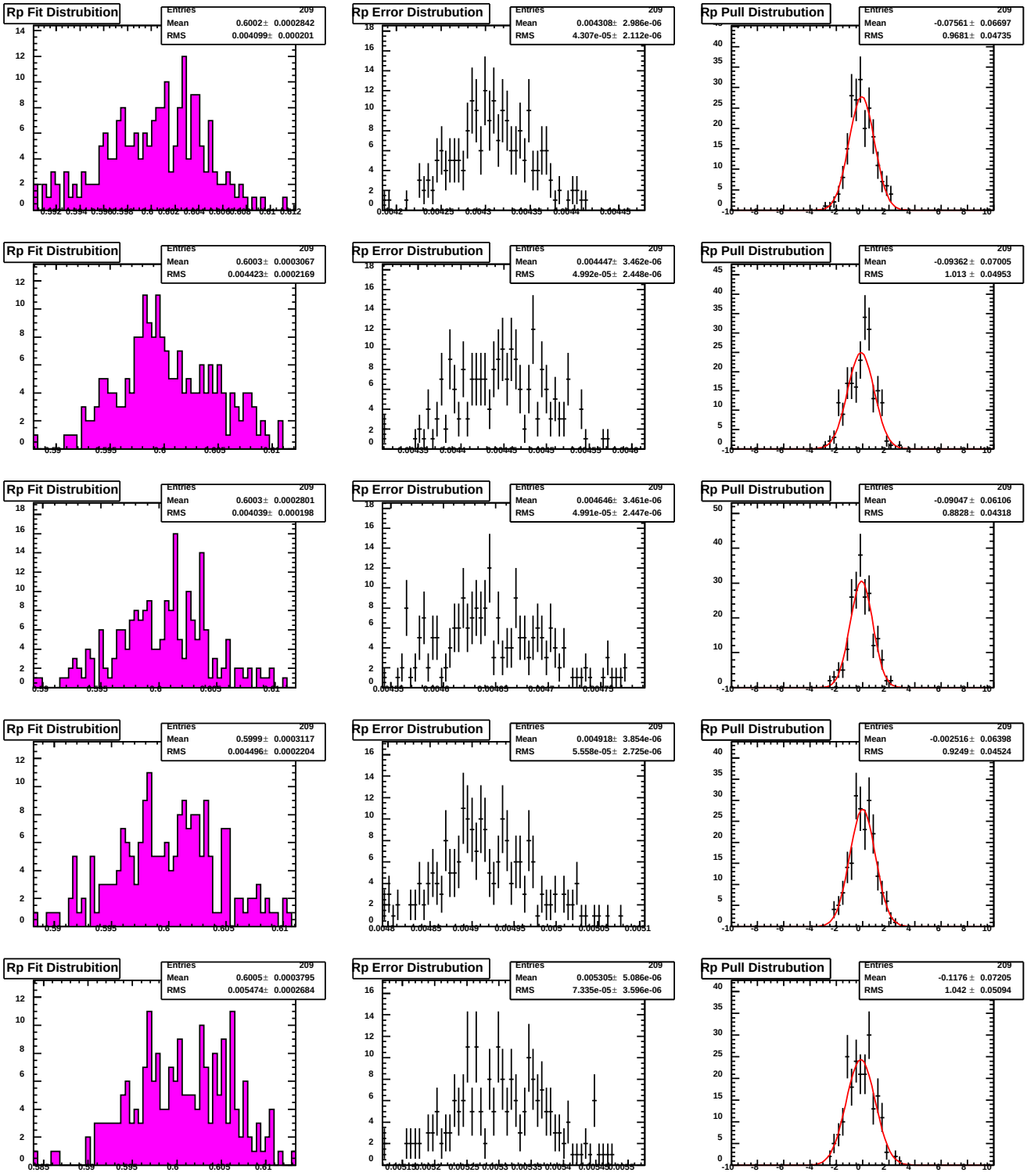


Figure E.32: R_0 sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{100\%,\ 90\%,\ 80\%,\ 70\%,\ 60\%\}$ of the signal.

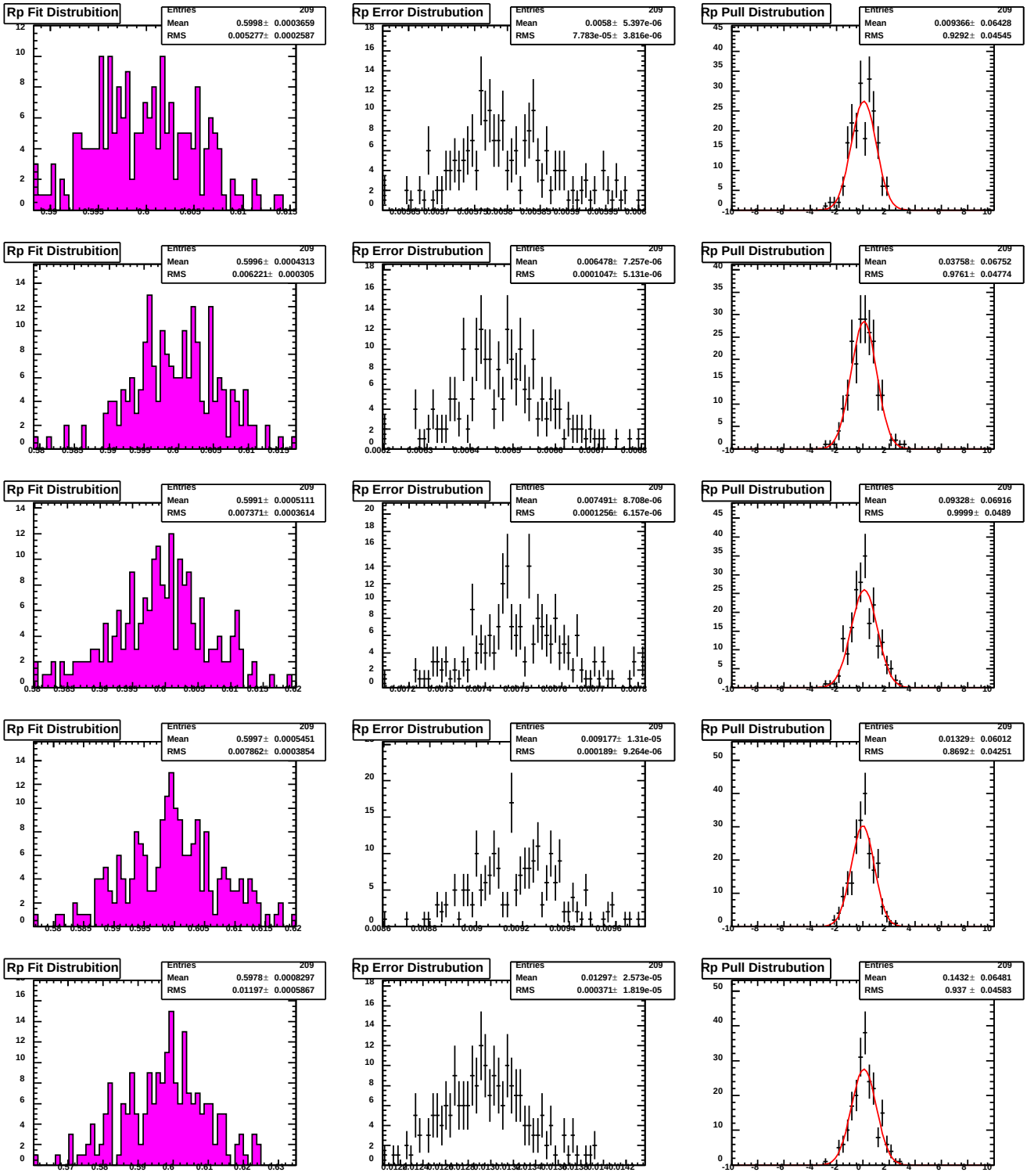


Figure E.33: R_0 sensitivity. Simultaneous studies using prompt-J/ψ data, retaining (from top to bottom) $\{50\%,\ 40\%,\ 30\%,\ 20\%,\ 10\%\}$ of the signal.

E.11 Sensitivity studies

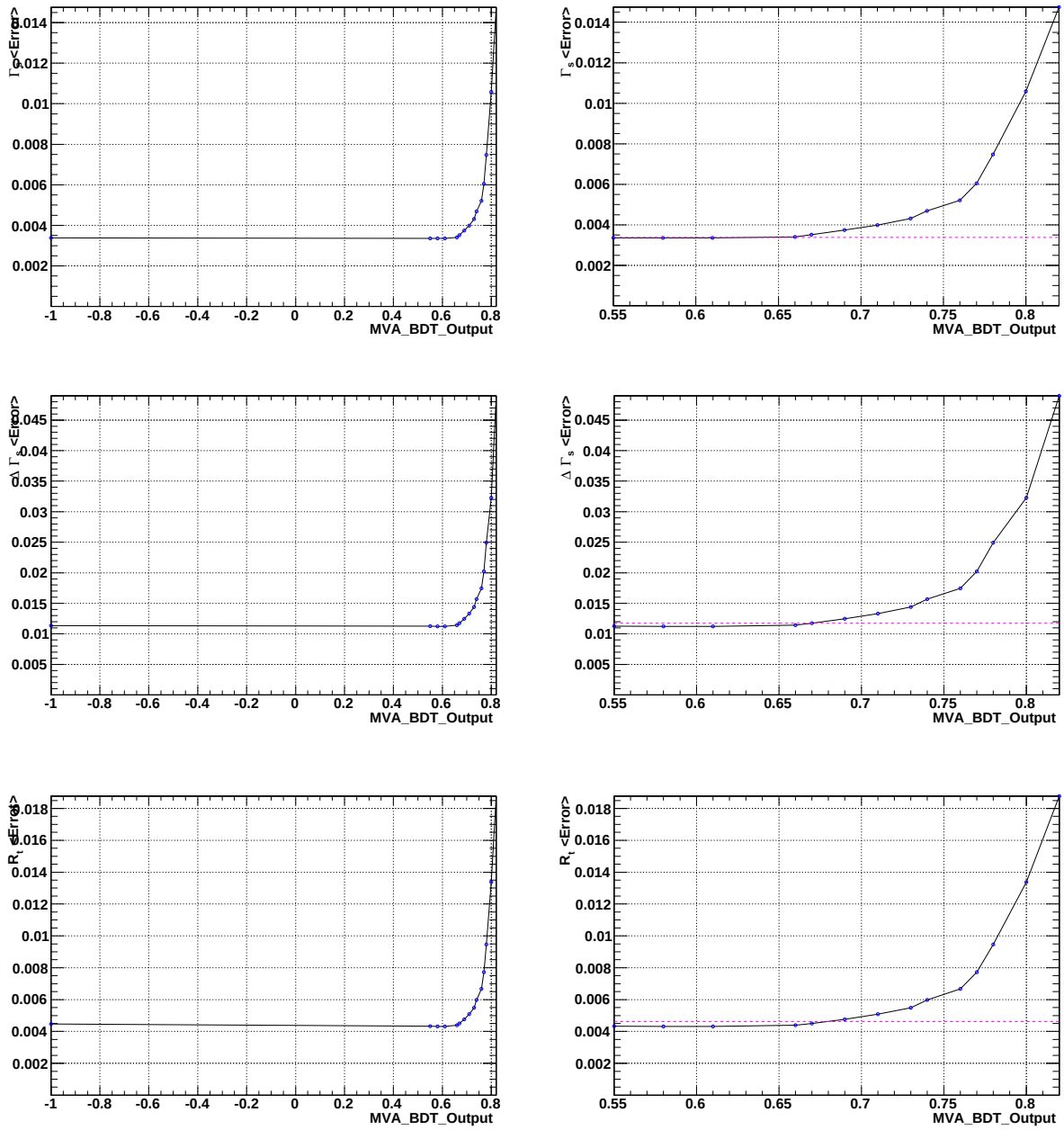


Figure E.34: (Left) Sensitivity of fit parameters with respect to the BDT's response using parameter Set 3 and the $\frac{B}{S}$ ratios given in Table 7.8 for: (Top) $\bar{\Gamma}_s$, (middle) $\Delta\Gamma_s$ and $R_\perp$ (Bottom) respectively. (Right) The corresponding parameter sensitivities enlarged around the maximal $\frac{S}{\sqrt{S+B}}$ value. The dotted magenta line indicating the sensitivity retaining 100% of the signal.

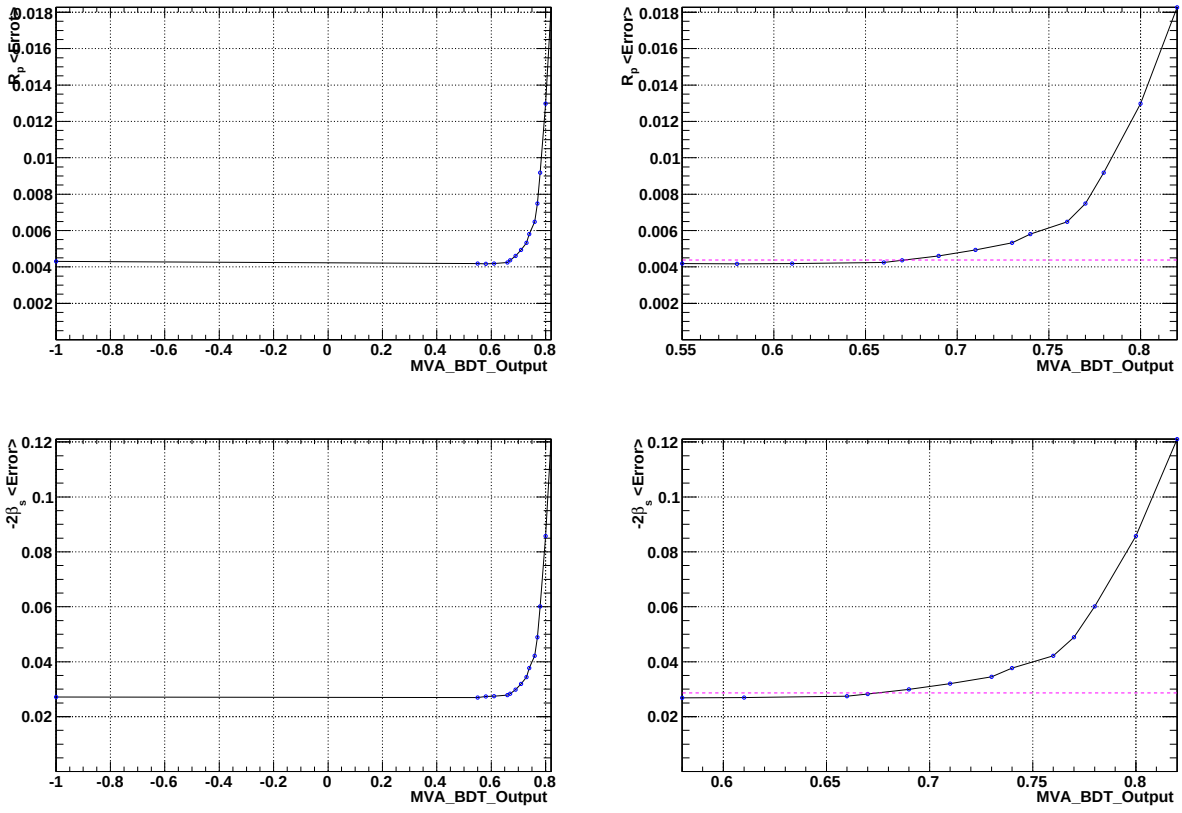


Figure E.35: (Left) Sensitivity of fit parameters with respect to the BDT's response using parameter Set 3 and the $\frac{B}{S}$ ratios given in Table 7.8 for: (Top) $R_\perp$, and (Bottom) $-2\beta_s$.(Right) The corresponding parameter sensitivities enlarged around the maximal $\frac{S}{\sqrt{S+B}}$ value. The dotted magenta line indicating the sensitivity retaining 100% of the signal.