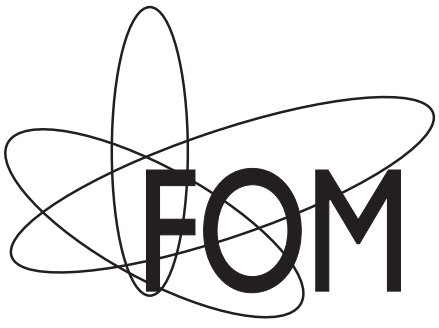
Search for a Strange Phase in Beautiful Oscillations

Cover: 'A Close Look at Interference' by Die zijn van '82.
Photo by Brian Valentine (www.flickr.com/photos/lordv).

Printed in 2010 by Ipskamp Drukkers, Enschede.



This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO).

VRIJE UNIVERSITEIT

Search for a Strange Phase in Beautiful Oscillations

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op vrijdag 22 oktober 2010 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Tristan Arnoldus du Pree

geboren te Rotterdam

promotor: prof.dr. M.H.M. Merk
copromotor: dr. H.G. Raven

het heelal

hoe verder men keek
hoe groter het leek

-J.A. Deelder

Contents

Chapter 1

Introduction

One of the contemporary big puzzles in physics is the observed asymmetry in the presence of matter and anti-matter in the universe. On Earth, most, if not all, matter consists of particles and not of anti-particles. In satellite experiments it is observed that all cosmic ray particles are matter particles. No antimatter dominated regions have been found in the visible universe. The standard model of cosmology assumes the whole universe to consist predominantly of matter. Therefore, during the evolution of the universe, a mechanism must have created the observed abundance of matter over anti-matter.

In physics symmetries are important. As first shown by Sakharov [1], one of the symmetries that must be broken to generate the current matter abundance is the CP symmetry. Violation of this combined symmetry of charge (C) conjugation and space inversion or parity (P) operation has first been experimentally observed in a laboratory in 1964, in the weak decay of neutral K mesons [2].

Building on the work of Cabibbo, the mechanism of violation of CP symmetry has been incorporated in the Standard Model of particle physics (SM) by Kobayashi and Maskawa in the CKM matrix [3]. They have done this by imposing the existence of at least three generations of quarks, which allows for a complex coupling constant in weak interaction amplitudes between quarks, necessary for CP violation to occur. The prediction of a third generation of quarks has been confirmed and the description of violation of the CP symmetry in both the kaon and B_d^0 is consistent with observations, which the Noble prize committee in 2008 rewarded with a share of the Noble prize.

However, the amount of CP violation predicted in the SM is too small to explain the observed matter abundance. Therefore, new sources of CP violation have to be sought for. CP violation in the B_s^0 system has not (yet) been observed, although a first evidence is suggested by the DØ collaboration [4]. It is one of the main goals of the LHCb experiment, an experiment at the Large Hadron Collider (LHC), a proton-proton accelerator at CERN (European Organisation for Nuclear Research), which has just begun taking data. At this accelerator B_s^0 mesons are expected to be copiously produced [5].

One of the key measurements for the discovery of CP violation in the B_s^0 system is the measurement of CP violation in the decay of a B_s^0 to a J/ψ and a ϕ meson, and the subsequent decays $J/\psi \to \mu^+\mu^-$ and $\phi \to K^+K^-$. On of the first $B_s^0 \to J/\psi\phi$ candidates as reconstructed by the LHCb experiment is shown in figure 1.1. This decay is the equivalent of the $B_d^0 \to J/\psi K_S$ decay for the B_s system. A theoretically interesting aspect of the $B_s^0 \to$

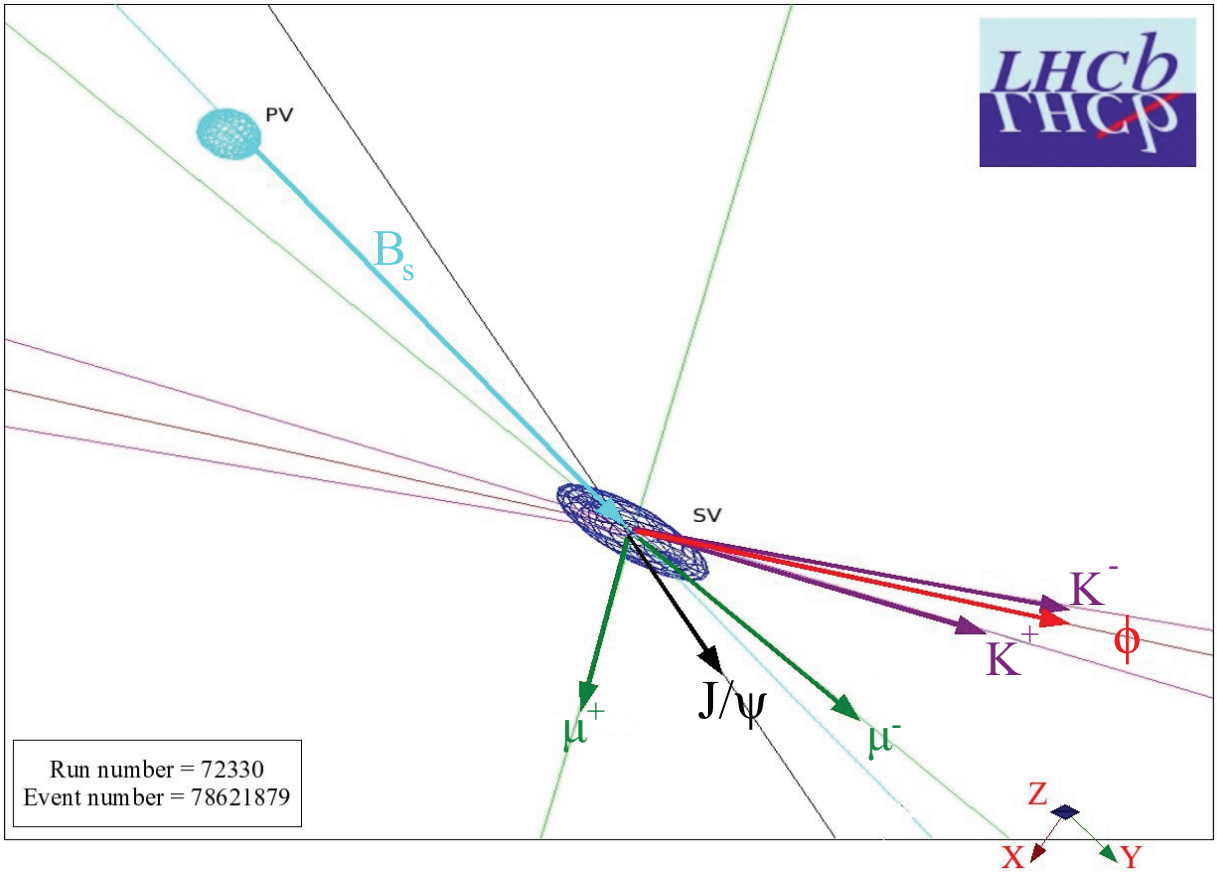


Figure 1.1: *One of the first $B_s^0 \to J/\psi\phi$ candidates as reconstructed by the LHCb detector.*

$J/\psi\phi$ channel is that the CP-violating phase in this decay is doubly Cabibbo suppressed, leading to an expectation of CP violation that is practically zero in the SM.

The $B_s^0 \to J/\psi\phi$ decay offers the best sensitivity to the CP violating parameter ϕ_s, which, in the Standard Model, is expected to be small (-0.04 rad). Non-SM new virtual processes might enhance observable CP-violation effects. In this way this measurement is both a search for CP violation, a test of the SM, and a probe to new physics, complementary to direct searches of new particles. Measurements by CDF and DØ have hinted in the past towards deviations from the value expected in the SM. *If* new physics contributes significantly to the amplitude of processes, the measurement allows to determine both the magnitude and phase of the couplings of these new particles.

Outline

In this dissertation the experimental method to analyze $B_s^0 \to J/\psi\phi$ signal channel is studied, using events from Monte Carlo simulations.

In chapter 2 the occurrence of CP violation and particle-antiparticle mixing, in particular in relation to B_s^0 and $\bar{B}_s^0$ mesons, is described. Possible deviations from the contributions expected in the Standard Model, due to new physics, are discussed. The theoretically ex-

pected differential decay rates, both for the signal channel, and for the control channel, $B_d^0 \to J/\psi K^*$, are given. Finally, the analysis strategy to measure the phase ϕ_s is outlined.

After this, in chapter 3, the LCHb detector is described. Most attention is given to the subdetector systems used in the measurement of CP violation in the $B_s^0 \to J/\psi\phi$ channel.

In order to limit the trigger rate of events selected by the LHCb detector and to obtain an optimal sensitivity to the CP-violating phase in the $B_s^0 \to J/\psi\phi$ channel, a selection method is developed. In chapter 4 this selection is described. An optimization procedure is presented, from which the optimal offline selection follows. A trigger selection procedure is proposed that keeps both the trigger rate acceptable and the sensitivity maximal.

Since both the reconstruction and the event selection affect the expected theoretical decay distribution, methods are developed to take these effects into account. In chapter 5 methods are described to take into account acceptance and resolution effects, and to include background in a multi-dimensional fit to recover the theoretically expected physics parameters.

Finally, in chapter 6 the efficiencies, resolutions and background distributions, caused by the described selection and detector effects, are presented. Using the proposed fit methods, these effects are included in the description of the event distributions. From the expected yields and purities, the sensitivity of the LHCb experiment to the physics parameters in the $B_s^0 \to J/\psi\phi$ decay, and to CP violation in particular, is estimated.

Chapter 2

Mixing and CP violation in the B_S system

The main subject of B-physics research is the study of flavor changing interactions of quarks $q_i \to W q_j$. In the Standard Model (SM) these interactions are governed by the Cabibbo Kobayashi Maskawa (CKM) mechanism, which contains a (3×3) matrix with complex coupling constants V_{ij}. Measurements of B decays determine both the sizes of CKM matrix elements and, interestingly, their complex phases. A non-zero irreducible phase of CKM couplings is required to generate CP-violating processes in the SM, such that the decay width of a B to a final state f ($\Gamma_{B\to f}$) and the width of the CP-conjugated decay ($\Gamma_{\bar{B}\to \bar{f}}$) are unequal, and the CP asymmetry,

$$A_{CP} = \frac{\Gamma_{B\to f} - \Gamma_{\bar{B}\to\bar{f}}}{\Gamma_{B\to f} + \Gamma_{\bar{B}\to\bar{f}}}, \tag{2.1}$$

is nonzero.

In this chapter the occurrence of CP violation and particle-antiparticle mixing, in particular in relation to B_s^0 and $\bar{B}_s^0$ mesons, is described. In particular, CP-violating phases originating from phase-convention independent combinations of CKM-matrix elements in the form of $V_{ij}V_{ik}^*/V_{lj}V_{lk}^*$, with a focus on the parameter $\beta_s \equiv \arg(-V_{cs}^*V_{cb}/V_{ts}^*V_{tb})$, are described. Subsequently, the origin of CP violation in the $B_s^0 \to J/\psi\phi$ decay, namely CP violation arising from interference between direct decay and decay after mixing, is described.

The final state of the $B_s^0 \to J/\psi\phi$ decay is a superposition of CP eigenstates with odd and even orbital angular momentum. Hence the final state is an admixture of states with positive and negative CP eigenvalue, which contribute with different signs to the CP asymmetry, hereby diluting the observed CP asymmetry. In order to determine the sizes of the different polarizations, and hence the true CP violating phase, an angular analysis needs to be performed.

In this chapter the proper time and angular distribution originating from the leading order SM processes contributing to the $B_s^0 \to J/\psi\phi$ decay are described. Possible deviations from this distribution due to both higher order Standard Model contributions and new physics contributions are given. Subsequently, the proper time and angular distribution of the decay $B_d^0 \to J/\psi K^*$, the most important control channel for this analysis, is described. Finally, the requirements for the LHCb experiment to measure the physics parameters are

outlined.

2.1 CP violation through interference

Since the SM Lagrangian is hermitian, every term in the Lagrangian appears with its adjoint. For example, the Yukawa part of the SM Lagrangian is

$$\begin{aligned} -\mathcal{L}_{\text{Yukawa}} &= Y_{ij}\overline{\psi_{L_i}}\phi\psi_{R_j} + h.c. \\ &= Y_{ij}\overline{\psi_{L_i}}\phi\psi_{R_j} + Y^*_{ij}(\overline{\psi_{L_i}}\phi\psi_{R_j})^\dagger. \end{aligned} \tag{2.2}$$

Under CP the field terms transform as

$$CP(\overline{\psi_{L_i}}\phi\psi_{R_j}) = (\overline{\psi_{L_i}}\phi\psi_{R_j})^\dagger. \tag{2.3}$$

Effectively, this means that under CP the couplings in (2.2) get complex conjugated. In other words, the imaginary phases in complex couplings flip sign in the CP-conjugated process. Furthermore, imaginary phases that do not flip sign can contribute to the total decay amplitude, e.g. due to final state interactions, or due to mixing, as shown below.

As such the decay probability of a decay dominated by a single amplitude is equal for the particle and the anti-particle case. Namely, physical observables are invariant under global phase transformations. Assume, however, that a particle decay occurs through two coherently contributing amplitudes, A_1 and A_2. For CP violation to occur in this decay process, a number of requirements must be fulfilled.

Assume that the phase difference between A_1 and A_2 contains two parts: one 'strong' phase, which does not change sign under CP conjugation, and one 'weak' phase, which *does* change sign under CP conjugation. Without loss of generality, the phases of A_1 can be chosen to equal zero, and the amplitudes can be written in terms of their strong phase difference δ, and weak phase difference ϕ, as

$$A_1 = |A_1|\,, \quad \frac{A_2}{A_1} = \frac{|A_2|}{|A_1|}e^{i\delta}e^{i\phi}. \tag{2.4}$$

Thus the decay rate of the unstable particle is found to be

$$\begin{aligned} |A|^2 &= |A_1 + A_2|^2 = ||A_1| + |A_2|e^{i\delta}e^{i\phi}|^2 \\ &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2|\cos(\delta+\phi), \end{aligned} \tag{2.5}$$

whereas the CP-conjugated decay rate for the anti-particle is

$$\begin{aligned} |\bar{A}|^2 &= ||A_1| + |A_2|e^{i\delta}e^{-i\phi}|^2 \\ &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2|\cos(\delta-\phi). \end{aligned} \tag{2.6}$$

Then the CP-asymmetry A_{CP} equals

$$A_{CP} \equiv \frac{|A|^2 - |\bar{A}|^2}{|A|^2 + |\bar{A}|^2} = \frac{-2\sin\phi\sin\delta}{r + 1/r + 2\cos\phi\sin\delta}, \tag{2.7}$$

with $r = |A_1|/|A_2|$. A graphical illustration of this example is shown in figure 2.1.

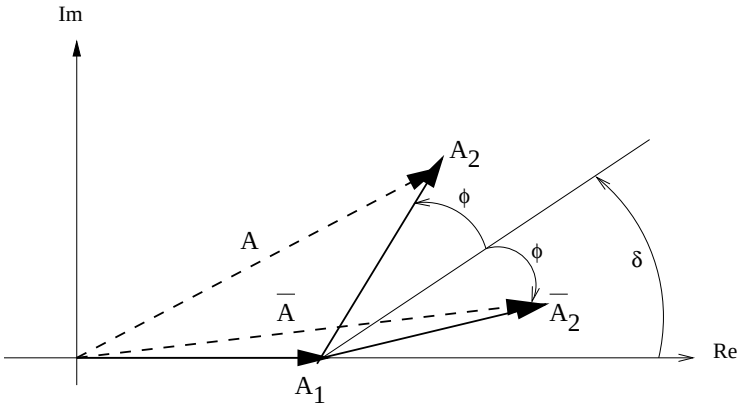


Figure 2.1: *A graphical illustration of CP violation, describing complex decay amplitudes as vectors in the complex plane. Without loss of generality A_1 can be chosen real, i.e. $A_1 = \bar{A}_1$. The difference in length of the two vector sums $A = A_1 + A_2$ and $\bar{A} = \bar{A}_1 + \bar{A}_2$ is a measure of the amount of CP violation.*

From this example, it is clear that, for an observable violation of CP to be present, at least two interfering amplitudes are required. A single amplitude is not sufficient to generate CP violation. Furthermore, both a nonzero relative strong phase and a nonzero relative weak phase are required: only if $\phi \neq 0$ *and* $\delta \neq 0$ are the difference between the decay rate and its CP conjugate nonzero. Finally, CP violation, for certain δ and ϕ, is largest if $|A_1| = |A_2|$ and vanishes as one of the two amplitudes goes to zero.

2.2 CP violation in weak interactions

CP violation is accommodated in the Standard Model (SM) in the charged weak current interactions between left handed quark fields. Writing the charged weak current as

$$J^{\mu-} = (\bar{u}_L, \bar{c}_L, \bar{t}_L)\gamma^\mu V_{CKM} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}, \tag{2.8}$$

with V_{CKM} a 3×3 matrix, the charged weak current interaction in the SM reads

$$\mathcal{L}_{\text{weak}} = \frac{g}{\sqrt{2}} J^{\mu-} W^-_\mu + h.c. \tag{2.9}$$

From this equation it can be seen that the weak current does not couple to the quark mass eigenstates, but to the weak interaction eigenstates, which are rotated from the mass eigenstates by the operation

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.10}$$

V_{CKM} is the Cabibbo-Kobayashi-Maskawa (CKM) matrix, a unitary matrix that describes the relative strength of the weak charge couplings between different quark flavors. Writing the couplings between flavor i and flavor j as V_{ij}, it has the following form

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2.11}$$

Following from unitarity, the CKM matrix, for three quark generations, contains four free parameters: three real angles and one irreducible phase [3]. Since there is only one irreducible phase in the 3×3 CKM matrix, every measurement of CP violation is a probe of the same quantity, assuming only SM processes contribute. The amount of CP violation in the weak decays in the SM is proportional to the Jarlskog determinant J [11], which for any choice of $i, j, k, l = 1, 2, 3$ [13] satisfies

$$\mathrm{Im}(V_{ij} V_{kl} V_{il}^* V_{kj}^*) = J \sum_{m,n=1}^{3} \epsilon_{ikm} \epsilon_{jln}, \tag{2.12}$$

and which is phase-convention independent, since the index of every quark field enters both with and without complex conjugate. Comparison of the measurements of both kaon, B_d and B_s decays is therefore a test of the consistency of the model of CP violation in the weak interactions.

2.3 Unitarity Triangles

The orthonormality of the CKM matrix

$$V_{CKM} V_{CKM}^\dagger = \mathbf{1}, \tag{2.13}$$

with $\mathbf{1}$ the identity matrix, leads to nine constraints. Two of these relate the couplings of the b-quarks, and as such play an important role in B-physics:

$$\begin{aligned} (B_d:)\ V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* &= 0, \\ (B_s:)\ V_{us} V_{ub}^* + V_{cs} V_{cb}^* + V_{ts} V_{tb}^* &= 0. \end{aligned} \tag{2.14}$$

Both these relations can be represented by a triangle in the complex plane. These triangles are shown in figure 2.2, using the measured values of the CKM matrix elements. The equations have been normalized such that one of the sides of the triangle is real, and has length 1. As a result, the only free corner is the apex. The apex of the Unitarity Triangle relevant for the B_d^0 system (UT) is $-V_{ud} V_{ub}^* / V_{cd} V_{cb}^*$. For the Unitarity Triangle relevant for the B_s^0 system (UT_s), the apex lies at $-V_{us} V_{ub}^* / V_{cs} V_{cb}^*$.

It can be seen that the interior angles of the UT are equal to

$$\alpha \equiv \arg[-\frac{V_{td} V_{tb}^*}{V_{ud} V_{ub}^*}];\ \beta \equiv \arg[-\frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*}];\ \gamma \equiv \arg[-\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*}]. \tag{2.15}$$

Since every (arbitrary) phase of a quark field appears either both in numerator and denominator, or both in V and V^*, these phases cancel. Hence the angles are invariant under

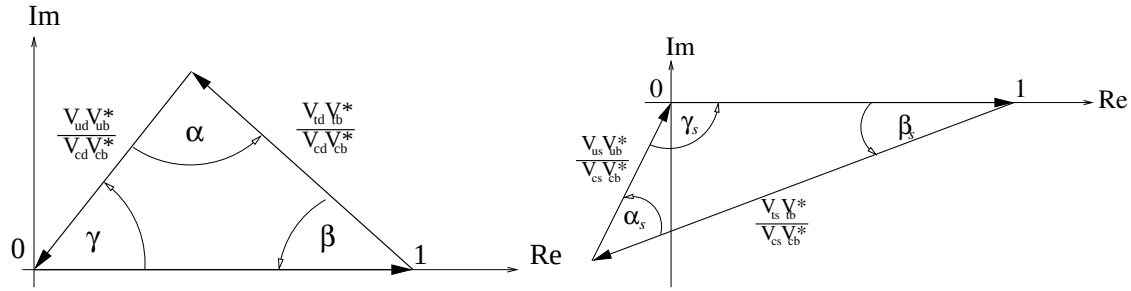


Figure 2.2: *The Unitarity Triangles relevant for the B_d^0 system (left) and the B_s^0 system (right). The sides are scaled such that one of the sides is equal to one. The scale of the imaginary axes are arbitrary and different; the directions of the arrows in the complex plane are governed by the measured values of the CKM elements.*

rephasing of the quark fields. The angles are related through the relation $\alpha + \beta + \gamma = \pi$. The angles in UT_s are given by

$$\alpha_s \equiv \arg[-\frac{V_{ts}^* V_{tb}}{V_{us}^* V_{ub}}];\ \beta_s \equiv \arg[-\frac{V_{cs}^* V_{cb}}{V_{ts}^* V_{tb}}];\ \gamma_s \equiv \arg[-\frac{V_{us}^* V_{ub}}{V_{cs}^* V_{cb}}], \tag{2.16}$$

and also satisfy $\alpha_s + \beta_s + \gamma_s = \pi$.

The angle β is related to the phase of $B_d - \bar{B}_d$ mixing process, whereas the angle β_s is related to the phase of $B_s - \bar{B}_s$ mixing, as will be shown later. Note that due to the values of the CKM matrix elements in the Standard Model and the fact that the observables are chosen to be the internal angles of the triangles, the angles in UT are related to the angles in UT_s by substituting $d \to s$ *and* taking the complex conjugate. This difference appears to be important for the sign convention of the mixing phase[1,2].

Finally, note that the surface area of all unitarity relations is equal to $|J|/2$, where the observable J is given in (2.12) and is at the time of writing measured to be

$$|J| = (2.93^{+0.15}_{-0.16}) \times 10^{-5} [9]. \tag{2.17}$$

A non-zero area of the unitarity triangles is a requirement for the presence of CP violation in the weak interactions[3].

2.3.1 Wolfenstein parameterization

The CKM matrix is often represented in the Wolfenstein parameterization, in which the hierarchy of the observed magnitude of the couplings is apparent. This parameterization is constructed by expanding the CKM matrix in terms of the parameter $\lambda \approx 0.23$ [6]. In this scheme the CKM matrix is parameterized in terms of the four real numbers λ, A, ρ, η as

[1]The angle β_s in UT_s related to $B_s - \bar{B}_s$ mixing is encountered in the literature in different ways, e.g. as χ or $\delta\gamma$.

[2]A more systematic nomenclature is proposed in [8].

[3]For CP-violation to occur in addition the quark masses are required to be not degenerate.

follows [12]:

$$V_{CKM} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{2.18}$$

The parameters A, ρ, η are real and of order unity. More details on the choice of parameterization can be found in [6] and references therein.

In this parameterization up to fourth order, imaginary couplings only show up in V_{ub} and V_{td}. As will be shown later, for measurements of β_s, higher order terms are in fact required:

$$\begin{pmatrix} -\frac{1}{8}\lambda^4 + \mathcal{O}(\lambda^6) & \mathcal{O}(\lambda^7) & 0 \\ \frac{1}{2}A^2\lambda^5[1-2(\rho+i\eta)] + \mathcal{O}(\lambda^7) & -\frac{1}{8}\lambda^4(1+4A^2) + \mathcal{O}(\lambda^6) & \mathcal{O}(\lambda^8) \\ \frac{1}{2}A\lambda^5(\rho+i\eta) + \mathcal{O}(\lambda^7) & \frac{1}{2}A\lambda^4(1-2(\rho+i\eta)) + \mathcal{O}(\lambda^6) & -\frac{1}{2}A^2\lambda^4 + \mathcal{O}(\lambda^6) \end{pmatrix}. \tag{2.19}$$

The apex of UT, up to fourth order in λ, is given by

$$-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} = (\bar{\rho}, \bar{\eta})\,; \;\; \text{with} \;\; \bar{\rho} \simeq (1-\lambda^2/2)\rho\,, \;\; \bar{\eta} \simeq (1-\lambda^2/2)\eta. \tag{2.20}$$

Similarly, the apex of UT_s is equal to

$$-\frac{V_{us}V_{ub}^*}{V_{cs}V_{cb}^*} = (\bar{\rho}_s, \bar{\eta}_s)\,; \;\; \text{with} \;\; \bar{\rho}_s \simeq \frac{-\lambda^2}{1-\lambda^2/2}\rho\,, \;\; \bar{\eta}_s \simeq \frac{-\lambda^2}{1-\lambda^2/2}\eta. \tag{2.21}$$

Using this parameterization it can be seen that the angles γ and γ_s are related[4] according to

$$\gamma_s = \pi - \gamma + \mathcal{O}(\lambda^3). \tag{2.22}$$

An instructive way to write down the CKM matrix is in terms of the angles β, γ and β_s of the Unitarity Triangles. Then, in the Wolfenstein phase convention, the CKM matrix reads

$$V_{CKM} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}|e^{-i\gamma} \\ -|V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}|e^{-i\beta} & -|V_{ts}|e^{+i\beta_s} & |V_{tb}| \end{pmatrix} + \mathcal{O}(\lambda^5). \tag{2.23}$$

In next chapter it is shown that, in this phase convention, the phase of V_{ts} is responsible for the CP violation in the decay process $B_s^0 \to J/\psi\phi$.

2.3.2 Constraints on the CKM matrix from measurements

The CKM model of weak interactions can be tested by verifying the consistency of different measurements of the CKM matrix elements. The current status of this combination of measurements is shown in figure 2.3. The constraints of the CP violating angles α, β and γ by direct measurements are dominated by the decays $B \to \rho^+\rho^-$, $B \to J/\psi K_S$ and $B^+ \to D(K_s^0\pi^+\pi^-)K^+$. The indirect constraints come from the measurements of the B_d

[4]The channel $B_s \to D_s^{\pm}K^{\mp}$ is sensitive to γ_s and β_s, and strictly speaking not to γ as is often claimed.

mixing frequency Δm_d and the B_s mixing frequency Δm_s, from observation of CP violation in kaon decays (ε_k), and from the measurement of the size of V_{ub}. The latter constraint is obtained from two distinct types of measurements: from inclusive and exclusive semi-leptonic decays on one hand, and from the branching ratio of $B^+ \to \tau^+\nu$ on the other hand.

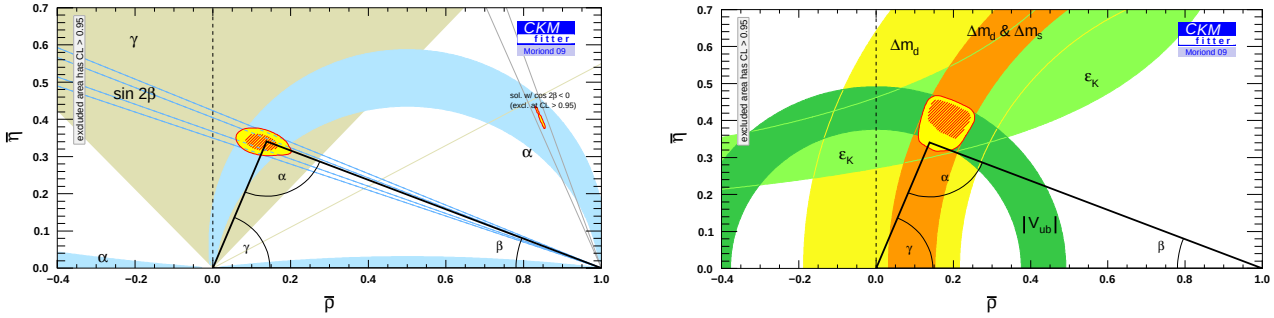


Figure 2.3: *The (left) direct and (right) indirect constraints, from measurements, on the Unitarity Triangle relevant for the B_d^0 system.*

In figure 2.4 the current combination of all constraints on the UT is shown. The available indirect constraints on UT_s are also shown. Combining (2.20) and (2.21) it can be seen that the constraints on UT_s are equal to the constraints on UT, multiplied by $V_{us}V_{cd}/V_{ud}V_{cs} \approx -\lambda^2$. Since this small number is negative, the constraints are point mirrored (and scaled) in the origin.

The current direct and indirect measurements of the $\mathrm{UT}_{(s)}$ angles are summarized in Table 2.1. The least well known angle of the UT angles is γ. Further constraining this angle or, the approximately equivalent angle γ_s, is one of the main goals of the LHCb experiment [20]. The constraint on β_s from indirect measurements has approximately the same relative precision as that of β, as can be understood from figure 2.4. The direct measurement of β_s, at the time of writing, does not deviate significantly from this indirect constraint.

CP angle	Indirect measurements (°)	Direct measurements (°)
α	$95.6^{+3.3}_{-8.8}$	$89.0^{+4.4}_{-4.2}$
β	$27.4^{+1.3}_{-1.9}$	$21.07^{+0.90}_{-0.88}$
γ	$67.8^{+4.2}_{-3.9}$	70^{+27}_{-30}
β_s	$1.032^{+0.049}_{-0.046}$	22 ± 10 or 68 ± 10

Table 2.1: *The current indirect and direct measurements of the UT angles α, β, and γ [9] and the indirect [9] and direct [7] measurements of the UT_s angle β_s. No direct measurements of γ_s and α_s exist.*

From table 2.1 follows that at the current level of precision, all measurements that constrain the $\mathrm{UT}_{(s)}$ are compatible with each other. It can therefore be concluded that the CKM mechanism provides the dominant source of CP violation in the B system.

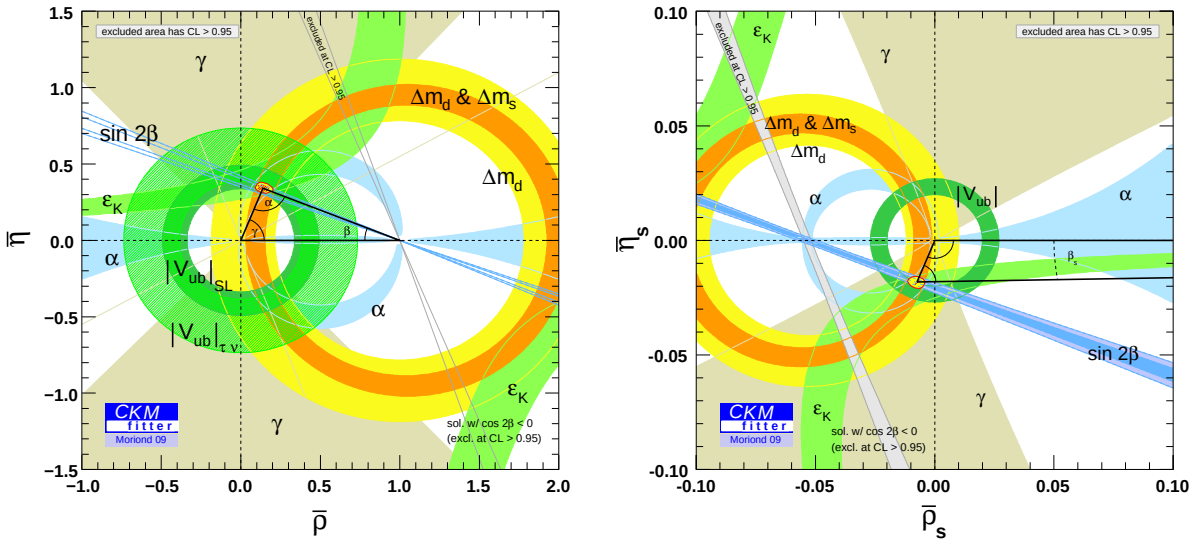


Figure 2.4: *The constraints from measurements on the Unitarity Triangles relevant for the B_d^0 system (left) and the B_s^0 system (right) in the Standard Model. The constraints on UT_s are the constraints on UT, multiplied by $V_{us}V_{cd}/V_{ud}V_{cs}$. Since this (small) number is negative, the constraints are point mirrored in (0,0). Since the factor has a small imaginary component there are small differences in the plots as the distance from the origin increases.*

2.4 B-physics

B mesons are two-quark states of which one quark is a b ('bottom' or 'beauty') quark. B mesons have a relatively long lifetime, which is experimentally useful for their identification. Furthermore, neutral B-mesons have a large oscillation probability to become anti-B mesons (and vice versa), allowing for sizeable interference effects in decay processes, which in turn can be used to probe the complex couplings of the CKM parameters.

The existing neutral B mesons are the B_d^0 (and $\bar{B}_d^0$) with quark content $\bar{b}d$ (and $b\bar{d}$), and the B_s^0 (and $\bar{B}_s^0$) with quark content $\bar{b}s$ (and $b\bar{s}$). Some general properties of the neutral B-mesons can be found in table 2.2.

	B_d^0	B_s^0
m_q(MeV)	5279.53 ± 0.33	5366.3 ± 0.6
τ_q(ps) $\equiv 1/\Gamma_q$	1.530 ± 0.009	$1.470^{+0.026}_{-0.027}$
Δm_q (ps^{-1})	0.507 ± 0.005	17.77 ± 0.12

Table 2.2: *Properties of the B_d^0 and B_s^0 mesons [6]. The subscript q distinguishes between the B_d and the B_s system.*

The oscillation or mixing of neutral B mesons occurs with a frequency Δm_q and was discovered in 1987 in the Argus experiment [14]. In this section the origin of B mixing is

explained. Furthermore, it is described how CP-violating processes can occur in B decays.

2.4.1 $B\bar{B}$ mixing

Transitions from the B^0 to $\bar{B}^0$ flavor states, and vice versa, are possible since a produced B-meson quantum state develops in time according to its projection on the mass eigenstates. To show this, the B^0 system is written as a superposition of the flavor eigenstates $|B^0\rangle$ and $|\bar{B}^0\rangle$

$$|\Psi(t)\rangle = a(t)|B^0\rangle + b(t)|\bar{B}^0\rangle. \tag{2.24}$$

The time evolution follows from the Schrödinger equation

$$i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = H\begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = (M - \frac{i}{2}\Gamma)\begin{pmatrix} a(t) \\ b(t) \end{pmatrix}, \tag{2.25}$$

where the Hamiltonian H has been written in terms of the two hermitian matrices M and Γ: the mass and lifetime matrices, respectively. From CPT invariance it follows that $\Gamma_{11} = \Gamma_{22} \equiv \Gamma$, and $M_{11} = M_{22} \equiv M$. And since, by construction, $M^\dagger = M$ and $\Gamma^\dagger = \Gamma$, it also follows that $\Gamma_{12} = \Gamma_{21}^*$ and $M_{12} = M_{21}^*$. Hence the Hamiltonian can be written as

$$H = \begin{pmatrix} H_0 & H_{12} \\ H_{21} & H_0 \end{pmatrix} = \begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M - \frac{i}{2}\Gamma \end{pmatrix}. \tag{2.26}$$

Contributions to Γ_{12} originate from intermediate real final states into which both B and $\bar{B}$ can decay [15]. M_{12} is dominated by short distance (virtual) processes.

For clarity, the conventions for the transformation of meson states and quark currents in this thesis are chosen[5,6] such that

$$CP|B^0\rangle = -|\bar{B}^0\rangle\,; \;\; (CP)\bar{q}_L\gamma_\mu b_L(CP)^{-1} = -\bar{b}_L\gamma^\mu q_L, \tag{2.27}$$

following [16]. This choice defines the CP eigenstates, which, in this convention, equal

$$|B^0_{even}\rangle = \frac{1}{\sqrt{2}}(|B^0\rangle - |\bar{B}^0\rangle)\,, \;\; |B^0_{odd}\rangle = \frac{1}{\sqrt{2}}(|B^0\rangle + |\bar{B}^0\rangle). \tag{2.28}$$

The Hamiltonian (2.26) has eigenvalues

$$\lambda_{1,2} = H_0 \pm \sqrt{H_{12}H_{21}}. \tag{2.29}$$

Now let us choose the H(eavy) and L(ight) mass eigenstates of the Hamiltonian as

$$\begin{aligned} |B_L\rangle &\equiv p|B^0\rangle + q|\bar{B}^0\rangle \\ |B_H\rangle &\equiv p|B^0\rangle - q|\bar{B}^0\rangle, \end{aligned} \tag{2.30}$$

with $|p|^2 + |q|^2 = 1$. Writing the eigenvalue of the eigenstate in terms of the mass and width of the eigenstate

$$\lambda_{H,L} = M_{H,L} - \frac{i}{2}\Gamma_{H,L}, \tag{2.31}$$

[5]For treatments in which this convention is not fixed, see [18] and [19].
[6]Metric $g_{\mu\nu} = (1, -1, -1, -1)$.

and choosing the convention that the heavy eigenstate has the largest mass, it follows that

$$\lambda_H = H_0 + \sqrt{H_{12}H_{21}}\,, \quad \lambda_L = H_0 - \sqrt{H_{12}H_{21}}. \tag{2.32}$$

Since $m_{H,L} = \mathrm{Re}\lambda_{H,L}$, the mass difference equals

$$\begin{aligned} \Delta m \equiv m_H - m_L &= 2\mathrm{Re}\sqrt{H_{12}H_{21}} \\ &= 2\mathrm{Re}\sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)}, \end{aligned} \tag{2.33}$$

which, by construction, is positive. Since $\Gamma_{H,L} = -2\mathrm{Im}\lambda_{H,L}$, the width difference equals

$$\begin{aligned} \Delta\Gamma \equiv \Gamma_L - \Gamma_H &= 4\mathrm{Im}\sqrt{H_{12}H_{21}} \\ &= 4\mathrm{Im}\sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)}. \end{aligned} \tag{2.34}$$

This is chosen such that it is positive in the Standard Model [16]. Since from experiment it is known that $\Delta m \gg \Delta\Gamma$, it follows model independently that $|\Gamma_{12}/M_{12}| \ll 1$. As a result, the mass and width difference of the mass eigenstates can be approximated by

$$\begin{aligned} \Delta m &= 2|M_{12}| \left[1 + O(|\frac{\Gamma_{12}}{M_{12}}|^2)\right], \\ \Delta\Gamma &= 2|\Gamma_{12}| \cos\phi_{M/\Gamma} \left[1 + O(|\frac{\Gamma_{12}}{M_{12}}|^2)\right]. \end{aligned} \tag{2.35}$$

Here the phase $\phi_{M/\Gamma}$ is convention independent, and defined by

$$\frac{M_{12}}{\Gamma_{12}} = -\left|\frac{M_{12}}{\Gamma_{12}}\right| e^{i\phi_{M/\Gamma}}. \tag{2.36}$$

Solving the Schrödinger equation, the time evolution of the eigenstates can be written as

$$|B_{H,L}(t)\rangle = e^{-i\lambda_{H,L}t}|B_{H,L}(0)\rangle = e^{-i(m_{H,L} - i\Gamma_{H,L}/2)t}|B_{H,L}(0)\rangle. \tag{2.37}$$

Casting these equations back into the basis of flavor eigenstates, the following time dependent solutions for the wave functions of physical B-mesons is found

$$\begin{aligned} |B^0_{\mathrm{phys}}(t)\rangle &= g_+(t)|B^0\rangle + \frac{q}{p}g_-(t)|\bar{B}^0\rangle \quad \text{(initially produced as } B^0\text{)}, \\ |\bar{B}^0_{\mathrm{phys}}(t)\rangle &= \frac{p}{q}g_-(t)|B^0\rangle + g_+(t)|\bar{B}^0\rangle \quad \text{(initially produced as } \bar{B}^0\text{)}, \end{aligned} \tag{2.38}$$

where $g_\pm(t)$ are the oscillation amplitudes and are given by

$$g_\pm(t) = \frac{1}{2}(e^{-(im_L + \Gamma_L/2)t} \pm e^{-(im_H + \Gamma_H/2)t}). \tag{2.39}$$

It should be noted that the i in the exponent delivers a relative phase between the two oscillation amplitudes that does not flip sign under CP conjugation, which is necessary for CP violation to occur. This can be easily seen if $\Delta\Gamma = 0$, in which case the oscillation

amplitudes are $g_+(t) = \cos\Delta mt$ and $g_-(t) = i\sin\Delta mt$. Since the size of the accompanying amplitudes is time-dependent, this relative phase which does not flip sign under CP conjugation allows for time-dependent CP violation.

With the mass eigenstates defined in (2.30), and the choice of the eigenvalues in (2.32), it then follows, in the convention defined by (2.27), unambiguously that q/p is negative and equal to

$$\frac{q}{p} = -\frac{\sqrt{H_{12}H_{21}}}{H_{12}} = -\sqrt{\frac{M_{12}^* - i\Gamma_{12}^*/2}{M_{12} - i\Gamma_{12}/2}}. \tag{2.40}$$

Defining the (convention dependent) phase

$$\phi_M \equiv \arg M_{12}, \tag{2.41}$$

(2.40) can be written as

$$\frac{q}{p} = -e^{-i\phi_M}\sqrt{\frac{|M_{12}| - i|\Gamma_{12}|e^{i\phi_{M/\Gamma}}}{|M_{12}| - i|\Gamma_{12}|e^{-i\phi_{M/\Gamma}}}}. \tag{2.42}$$

In the limit $|\Gamma_{12}/M_{12}| \ll 1$ the convention dependent mixing fraction q/p can be approximated by

$$\frac{q}{p} = -e^{-i\phi_M}[1 - a_{SL}/2], \text{ with } a_{SL} = |\frac{\Gamma_{12}}{M_{12}}|\sin(\phi_{M/\Gamma}). \tag{2.43}$$

Here a_{SL} is a convention *independent* measure for the CP asymmetry in mixing, which is an observable in semi-leptonic decays, as is illustrated below.

Finally, since for B mesons $\cos\phi_{M/\Gamma} \simeq 1$ in the SM, (2.28) and (2.30) are approximately equal, and the mass eigenstates can be related to the CP eigenstates in the SM as

$$\begin{aligned} |B_L\rangle &\simeq |B^0_{even}\rangle, \\ |B_H\rangle &\simeq |B^0_{odd}\rangle. \end{aligned} \tag{2.44}$$

Since the branching fraction of B_s to CP-even final states is larger than to CP-odd final states, due to the D_SD_S final states, the width difference between the CP eigenstates $\Delta\Gamma_{CP} \equiv \Gamma(B_s^{\text{even}}) - \Gamma(B_s^{\text{odd}}) = 2|\Gamma_{12}|[1 + O(|\frac{\Gamma_{12}}{M_{12}}|^2)]$, which is positive. In the SM the width difference of the mass eigenstates approximately equals that of the CP eigenstates: $\Delta\Gamma \simeq \Delta\Gamma_{CP}$. $\Delta\Gamma_{CP}$ can be measured from e.g. the branching ratios of B_s mesons to the odd/even final states $D_S^{(*)+}D_S^{(*)-}$ and is unaffected by new physics [16], unlike $\Delta\Gamma$, as is discussed below.

2.4.2 CP Violation in B^0 decays

The decay rates of B decay processes can differ from those of their CP conjugated processes. Three types of CP violation in B decays are distinguished.

CP violation in the mixing

CP violation in mixing transitions occurs when the magnitude ratio of the mixing amplitudes $|q/p| \neq 1$, since then

$$|\langle B^0_{\rm phys}(t)|\bar{B}^0\rangle|^2 = |q/p|^2 \neq |p/q|^2 = |\langle \bar{B}^0_{\rm phys}(t)|B^0\rangle|^2. \tag{2.45}$$

This implies that the decay rate of an initially produced B meson to a final state specific for a $\bar{B}$ meson differs from the CP conjugated process. In the neutral B system this can be observed through determinations of the asymmetry between initially tagged B^0 and $\bar{B}^0$ mesons in semileptonic decays:

$$a_{SL} = \frac{\Gamma(\bar{B}^0_{phys}(t) \to l^+\nu X) - \Gamma(B^0_{phys}(t) \to l^-\nu X)}{\Gamma(\bar{B}^0_{phys}(t) \to l^+\nu X) + \Gamma(B^0_{phys}(t) \to l^-\nu X)}. \tag{2.46}$$

CP is conserved if the mass eigenstates equal the CP eigenstates, since then the relative phase between M_{12} and Γ_{12} vanishes. The weak phase originates from the possible difference between the mass eigenstates and the CP eigenstates, as is present in the kaon system. The presence of the required strong phase originates from the mixing process, as discussed below.

This observable a_{SL} is a constraint in estimating new physics occurring in B^0_s mixing, as is discussed below. Current measurements are consistent with $a_{SL} = 0$, or $|q/p| = 1$, which is used below.

CP violation in decay

An obvious class of CP-violating processes occurs when the decay amplitudes $\bar{A}_{\bar{f}}$, the amplitude of $\bar{B} \to \bar{f}$, and A_f, the amplitude of $B \to f$, are not equal:

$$|A_f/\bar{A}_{\bar{f}}| \neq 1. \tag{2.47}$$

Since then, by definition, the decay rate of a B^0 into a final state differs from the case, where a $\bar{B}^0$ decays to the CP-conjugated final state. An example is the $B \to K\pi$ decay, which can either occur via a on-shell decay process or via an off-shell loop process. In this case the weak phase originates from the difference between the tree and the loop amplitudes, and the strong phase originates from strong interactions between the final state particles.

CP violation in the interference between decays with and without mixing

In the absence of both direct CP violation in decay *and* CP violation in mixing, there is still a possibility for CP violation to occur. It can happen when there is an amplitude for the B-meson to directly decay into the final state, as well as an amplitude to first oscillate into its anti-particle and *then* decay into the same final state. A relative weak phase difference then leads to time-dependent CP violation. The strong phase originates from the dynamics of the mixing process.

To parameterize this type of CP violation the λ_f variable is introduced:

$$\lambda_f \equiv \frac{q}{p}\frac{\bar{A}_f}{A_f}. \tag{2.48}$$

The parameter λ_f is the ratio of the amplitudes of the decay and the CP-conjugated decay. CP violation in the interference of mixing and decay occurs when $\arg(\lambda_f) + \arg(\lambda_{\bar{f}}) \neq 0$. As it is relevant for $B_s^0 \to J/\psi\phi$, this type of CP violation is described in more detail in the next section.

2.5 Interference between mixing and decay to a CP-eigenstate

A special case of CP violation due to the interference of mixing and decay is when the final state f is a CP eigenstate. Supposing f is a CP eigenstate with eigenvalue η_f,

$$CP|f_{CP}\rangle = \eta_f|f_{CP}\rangle, \tag{2.49}$$

it follows that

$$\bar{A}_f = \eta_f \bar{A}_{\bar{f}}\,, \text{ and therefore } \lambda_f = \eta_f \frac{q}{p}\frac{\bar{A}_{\bar{f}}}{A_f}. \tag{2.50}$$

In that case the overall decay amplitude of an initially produced B-meson reads:

$$\begin{aligned} A(t) &= \langle f_{CP}|H|B(t)\rangle \\ &= A[g_+(t) + \lambda_f g_-(t)], \end{aligned} \tag{2.51}$$

since the B can either decay directly via an amplitude $Ag_+(t)$, or mix and decay with an amplitude $A\lambda_f g_-(t)$, as is shown in figure 2.5. Using the expressions of the oscillation

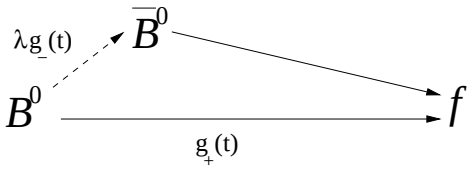


Figure 2.5: *The two amplitudes contributing to the decay of a B^0 meson to a final state to which both the meson and the anti-meson can decay. The lower, direct, decay amplitude equals $g_+(t)$; the upper amplitude, where the meson first mixes and then decays, equals $\lambda g_-(t)$.*

amplitudes given in 2.39, the decay rates of the B and the $\bar{B}$ to the final state f read

$$\begin{aligned} \Gamma_{B\to f}(t) &= |A_f|^2(1+|\lambda_f|^2)\frac{e^{-\Gamma t}}{2}\times \\ &\quad \left(\cosh\frac{\Delta\Gamma t}{2} - D_f \sinh\frac{\Delta\Gamma t}{2} + C_f\cos(\Delta m t) - S_f \sin\Delta m t\right), \\ \Gamma_{\bar{B}\to f}(t) &= |A_f|^2|\frac{p}{q}|^2(1+|\lambda_f|^2)\frac{e^{-\Gamma t}}{2}\times \\ &\quad \left(\cosh\frac{\Delta\Gamma t}{2} - D_f \sinh\frac{\Delta\Gamma t}{2} - C_f\cos(\Delta m t) + S_f \sin\Delta m t\right), \end{aligned} \tag{2.52}$$

where

$$D_f = \frac{2\mathrm{Re}\lambda_f}{1+|\lambda_f|^2},\ C_f = \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2},\ S_f = \frac{2\mathrm{Im}\lambda_f}{1+|\lambda_f|^2}. \tag{2.53}$$

Then the CP asymmetry between the B and the $\bar{B}$ decay is

$$\begin{aligned} A_{CP}(t) &= \frac{\Gamma_{B\to f}(t) - \Gamma_{\bar{B}\to f}(t)}{\Gamma_{B\to f}(t) + \Gamma_{\bar{B}\to f}(t)} \\ &= \frac{a_- \cosh\frac{\Delta\Gamma t}{2} - a_- D_f \sinh\frac{\Delta\Gamma t}{2} + a_+ C_f \cos(\Delta m t) - a_+ S_f \sin\Delta m t}{a_+ \cosh\frac{\Delta\Gamma t}{2} - a_+ D_f \sinh\frac{\Delta\Gamma t}{2} + a_- C_f \cos(\Delta m t) - a_- S_f \sin\Delta m t}, \end{aligned} \tag{2.54}$$

where the following variables have been defined: $a_+ \equiv (1+|\frac{p}{q}|^2)$, $a_- \equiv (1-|\frac{p}{q}|^2)$. For small a_{SL} these variables satisfy $a_+ \approx 2 + a_{SL}$, $a_- \approx -a_{SL}$.

To correctly determine the above CP asymmetry, the flavor of the B at production needs to be determined. The inclusion of this so-called experimental flavor tagging in the analysis, and the consequences of doing it occasionally incorrectly, are discussed in chapter 5.

2.6 CP violation in the interference of mixing and $b \to c\bar{c}s$ transitions.

The mixing phenomenon of a B_s^0 into a $\bar{B}_s^0$ is described by the off-diagonal elements of the mass and decay matrices of (2.26). The dominating non-zero off-diagonal element M_{12}

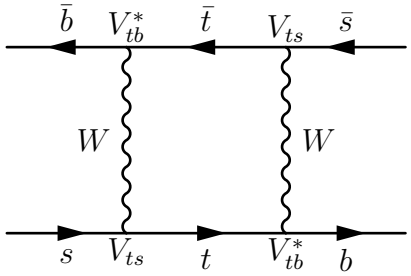


Figure 2.6: *The Feynman 'box' diagram describing $B_s - \bar{B}_s$ mixing.*

stems from the Feynman 'box' diagram in figure 2.6, whereas Γ_{12} originates from real (on-shell) final states [16] and is dominated by CKM-favoured, tree-level, decays. Since Γ_{12} is determined by long-distance physics, it is practically insensitive to new physics, which is assumed to involve higher mass scales. On the other hand, M_{12} originates from short-distance physics. Since virtual (off-shell) high-mass particles can contribute to the diagram, M_{12} can be sensitive to new physics.

The Feynman diagram of the $B_s^0 \to J/\psi\phi$ decay is shown in figure 2.7.

In the limit $|\Gamma_{12}| \ll |M_{12}|$, (2.40) becomes

$$\frac{q}{p} = -\sqrt{\frac{M_{12}^*}{M_{12}}}. \tag{2.55}$$

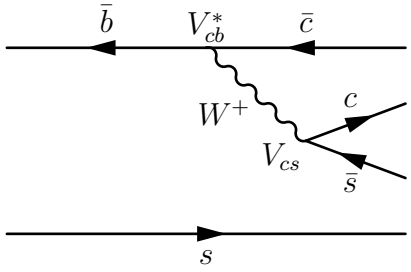


Figure 2.7: *The Feynman diagram describing the $B_s^0 \to J/\psi\phi$ decay.*

The convention dependent phases of the mixing and the decay can be combined, such that a convention independent parameter λ specific for $B_s^0 \to J/\psi\phi$ can be constructed. It can be seen from figure 2.6 and figure 2.7 that $\lambda_{J/\psi\phi}$ equals

$$\lambda_{J/\psi\phi} = \left(\frac{q}{p}\right)_{B_s} \frac{\bar{A}_{J/\psi\phi}}{A_{J/\psi\phi}} = \eta \left(\frac{V_{tb}^* V_{ts}}{V_{tb} V_{ts}^*}\right) \left(\frac{V_{cb} V_{cs}^*}{V_{cb}^* V_{cs}}\right). \tag{2.56}$$

The first term originates from the B-mixing diagram, the second term from the decay. The factor η is the CP eigenvalue of the final state. It can be seen from 2.16 that this fraction, in the SM, is related to the angle β_s of the Unitarity Triangle:

$$\lambda_{J/\psi\phi} = \eta e^{+2i\beta_s}. \tag{2.57}$$

Now assume, consistent with experimental data, that $|q/p| = 1$, then the decay rates read, in terms of η and β_s,

$$\begin{aligned}
\Gamma_{B\to f(t)} &= |A_f|^2 \frac{e^{-\Gamma t}}{2} \times \\
&\quad \left(\cosh\frac{\Delta\Gamma t}{2} - \eta\cos 2\beta_s \sinh\frac{\Delta\Gamma t}{2} - \eta \sin 2\beta_s \sin\Delta m t\right), \\
\Gamma_{\bar{B}\to f(t)} &= |A_f|^2 \frac{e^{-\Gamma t}}{2} \times \\
&\quad \left(\cosh\frac{\Delta\Gamma t}{2} - \eta\cos 2\beta_s \sinh\frac{\Delta\Gamma t}{2} + \eta \sin 2\beta_s \sin\Delta m t\right).
\end{aligned} \tag{2.58}$$

Then the CP-asymmetry reads

$$A_{CP} = \frac{-\eta \sin 2\beta_s \sin\Delta m_s t}{\cosh\frac{\Delta\Gamma t}{2} - \eta\cos 2\beta_s \sinh\frac{\Delta\Gamma t}{2}}. \tag{2.59}$$

This CP violation can be understood from figure 2.8. The B_s^0 meson can decay via two paths to the final state f: either directly, or by first oscillating to a $\bar{B}_s^0$. Similarly, in the CP-mirrored process the $\bar{B}_s^0$ meson can decay via two paths to the final state f: either directly, or by first oscillating to a B_s^0.

The relative weak phase β_s, which flips sign in the CP-conjugated decay, originates from the mixing and subsequent decay. The phase that does not flip sign in the CP-conjugated

decay originates from the dynamics of the B_s-$\bar{B}_s$ oscillation, since there is a relative phase of 90° between $g_+(t)$ and $g_-(t)$ in (2.39), as explained above. Since the size of the amplitude accompanied by the strong phase varies as a function of proper time, the CP asymmetry is proper-time dependent.

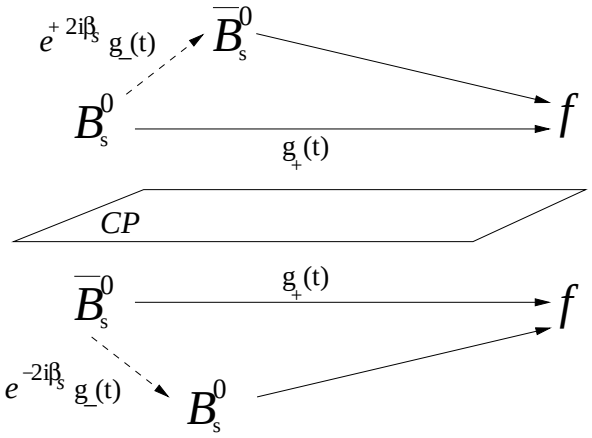


Figure 2.8: *CP violation from interference. The upper figure describes the decay of the B_s^0 meson to the CP-eigenstate f, the lower the CP-conjugated decay of the $\bar{B}_s^0$ to the same final state. The upper decay occurs with a weak phase $+2\beta_s$, the lower with $-2\beta_s$*

2.6.1 New Physics contributions

In the case that supersymmetry particles exist, these are expected to contribute to the box diagram, as shown by the Feynman diagram example shown in figure 2.9 [21]. As an alternative example, contributions from Z' bosons are discussed in [22].

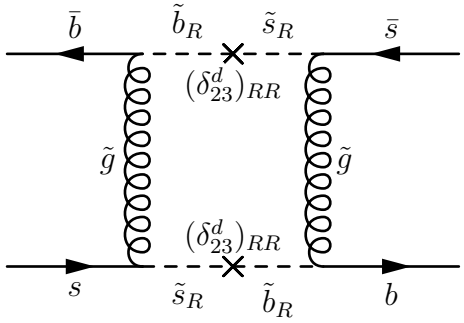


Figure 2.9: *An example of a diagram where supersymmetry particles contribute to B_s-$\bar{B}_s$ oscillation [21].*

In general, new particles are expected not to contribute to on-shell processes, due to their larger masses. Therefore, any non-SM contribution to $\Delta B = 2$ mixing transitions is expected to only affect M_{12}, and not Γ_{12}. Such a contribution can be parameterized by

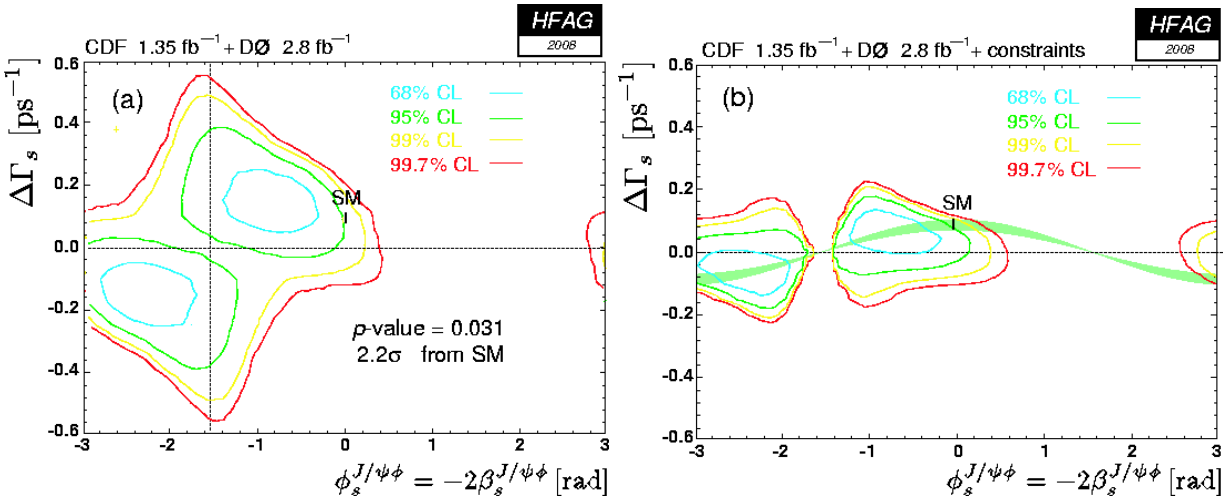


Figure 2.10: *Constraints on $\Delta\Gamma$ and ϕ_s from the combination of $B_s^0 \to J/\psi\phi$ measurements of CDF and DØ at the time of writing. In the right plot also the information from measurements of lifetimes and semi-leptonic asymmetries in B_s^0 decays is used.*

introducing a parameter Δ_s. With the inclusion of new physics, the off-diagonal term in the Hamiltonian can be written as

$$M_{12} \quad \to \quad M_{12}^{SM}\Delta_s \equiv M_{12}^{SM}|\Delta_s|e^{i\phi_s^{\Delta}}. \tag{2.60}$$

Since this changes the overall phase as

$$\arg M_{12} \to \arg M_{12}^{SM} + \arg \Delta_s, \tag{2.61}$$

(2.57) becomes

$$\lambda_{J/\psi\phi} \to \eta e^{+2i\beta_s - i\phi_s^{\Delta}}. \tag{2.62}$$

This implies that the NP phase augments the SM CP violating phase in the following way

$$-2\beta_s \to -2\beta_s + \phi_s^{\Delta}. \tag{2.63}$$

Now the CP-violating observable $\phi_s \equiv -2\beta_s + \phi_s^{\Delta}$ can be introduced[7], such that, finally

$$\lambda_{J/\psi\phi} = \eta e^{-i\phi_s}. \tag{2.64}$$

Then the observables given in (2.35) and (2.43) are influenced in the following manner:

$$\begin{aligned} \Delta m_s &\to 2|M_{12}^{SM}||\Delta_s| \\ \Delta\Gamma_s &\to 2\Gamma_{12}\cos(\phi_{M/\Gamma} + \phi_s^{\Delta}) \\ a_{SL} &\to |\frac{\Gamma_{12}}{M_{12}^{SM}}|\frac{\sin(\phi_{M/\Gamma} + \phi_s^{\Delta})}{|\Delta_s|} \end{aligned} \tag{2.65}$$

[7]This convention independent ϕ_s should not be confused with the same variable often used for the convention dependent mixing phase in the B_s system. Furthermore, β_s is the angle of UT_s and hence in presence of new physics contributions not an observable. Finally, $\Phi_{J/\psi\phi}$ in [20] is aesthetically less appealing.

A combination of the constraints on $\Delta\Gamma_s$ and ϕ_s, at the time of writing, is shown in figure 2.10 [7]. The 'p-value', the probability of obtaining the measurement in the figure or a larger deviation assuming that the SM values are the true values, is, at the time of writing, 3% [7].

The constraints on Δ_s, at the time of writing, are shown in figure 2.11, combining the measurements of the semi-leptonic asymmetries of different B_q decays a^q_{SL} as defined in 2.46, the lifetimes and widths differences $\Delta\Gamma_q$, the mass differences Δm_q and the measurements of ϕ_s. The SM value $(|\Delta_s|, \phi_s^{\Delta}) = (1, 0)$ is, at the time of writing, excluded at 2σ.

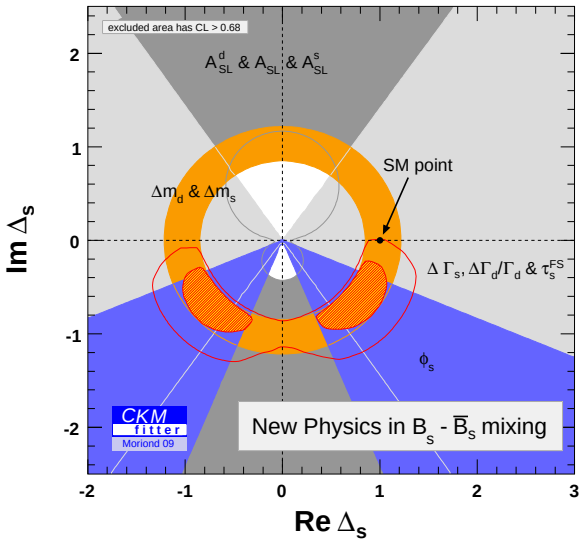


Figure 2.11: *Current constraints on New Physics from the measurements of ϕ_s, widths and width differences, mass differences and semi-leptonic asymmetries. Deviations from the SM values according to (2.63) and (2.65) are assumed. Due to the large uncertainties, only the* 68% *CL are shown for the individual constraints. Both* 68% *and* 95% *contours are shown for the global constraint. The SM point ($\Delta_s = (1, 0)$) is excluded at 2σ [10].*

2.6.2 Penguin contributions

The described NP contributions can only be measured if all SM contributions are understood. If not only tree decay amplitudes, but also loop amplitudes known as penguin amplitudes contribute, the total amplitude becomes a sum of the tree amplitude A_T and the penguin amplitude A_P, with strong phases $\delta_{T,P}$ and weak phases $\phi_{T,P}$

$$A = A_T e^{i\delta_T} e^{i\phi_T} + A_P e^{i\delta_P} e^{i\phi_P}. \tag{2.66}$$

Only if the penguin amplitude is negligible, the fraction of the two amplitude reads

$$\frac{A}{\bar{A}} \cong \frac{A_T e^{i\delta_T} e^{i\phi_T}}{A_T e^{i\delta_T} e^{-i\phi_T}} = e^{2i\phi_T}, \tag{2.67}$$

and the relative weak and strong phases of the tree and penguin amplitudes are not important in the measurement of CP-violation [23].

If, next to tree contributions T, also penguin contributions P to the final amplitude are considered, the total amplitude reads

$$A = V_{cs}V_{cb}^{*}T + V_{cs}V_{cb}^{*}P_c + V_{us}V_{ub}^{*}P_u + V_{ts}V_{tb}^{*}P_t, \tag{2.68}$$

where T and P_j are tree and penguin amplitudes, (with internal $j \in (u, c, t)$ quarks) respectively. Using the unitarity relations ($V_{us}V_{ub}^{*} + V_{cs}V_{cb}^{*} + V_{ts}V_{tb}^{*} = 0$), this can be written as

$$A = V_{cs}V_{cb}^{*}T + V_{cs}V_{cb}^{*}(P_c - P_t) + V_{us}V_{ub}^{*}(P_u - P_t). \tag{2.69}$$

The penguin contributions P_i can be expressed in terms of perturbatively calculable Wilson coefficient functions and non-perturbative hadronic matrix elements. Calculating the perturbative calculable Wilson coefficients, a correction of less than a permille of the observed CP violation in the decay $B^0 \to J/\psi K_s$ is found [24]. With similar arguments, the contributions from penguin amplitudes are expected to be negligible for $B_s^0 \to J/\psi\phi$ as well. However, the non-perturbative hadronic matrix elements which appear in P_i are associated with large uncertainties. From an analysis of $B^0 \to J/\psi\pi^0$ decays, penguin uncertainties in $B_s^0 \to J/\psi\phi$ as large as $\mathcal{O}(10\%)$ are not ruled out [25]. Therefore, such uncertainties should be calibrated experimentally.

This can be done using $B_s \to J/\psi\bar{K}^*$ as a control channel. Measuring its branching ratio and the direct CP asymmetry with $\bar{B}_s \to J/\psi K^*$, the penguin contributions to this channel can be measured. Assuming SU(3) flavor symmetry of the strong interactions, these contributions can then assumed to occur also in $B_s^0 \to J/\psi\phi$. Finally, also $B_d \to J/\psi\phi$ decay has been suggested to be used to check the 'penguin annihilation' and 'exchange' diagrams, which contribute to $B_s^0 \to J/\psi\phi$, but which are not present in $B_s \to J/\psi\bar{K}^*$ [25].

2.7 Time dependent angular distribution

CP-violation in the $B_s^0 \to J/\psi\phi$ decay originates from the interference between direct decay and decay after mixing, both to the same CP eigenstate. However, the $B_s^0 \to J/\psi\phi$ channel involves the decay of a spin-zero (pseudo-scalar) B_s into two spin one (vector) mesons, in which the total angular momentum $\vec{J}$, the sum of the orbital momentum $\vec{L}$ and spin $\vec{S}$, is conserved: $\vec{J} = \vec{L} + \vec{S} = 0$. Hence the final state can have angular momentum $|L| = (0, 1, 2)$, which leads to the following CP eigenvalue possibilities for the final state:

$$\begin{aligned} CP|J/\psi\phi\rangle &= \eta_f|J/\psi\phi\rangle \\ &= (-1)^L \times (-1) \times (-1) \times (+1)|J/\psi\phi\rangle \\ &= (-1)^L|J/\psi\phi\rangle, \end{aligned} \tag{2.70}$$

leading to either even ($L = 0, 2$) or odd ($L = 1$) eigenvalues.

At leading order, the CP-even and CP-odd components contribute with opposite signs to the CP asymmetry, thereby diluting the observable CP asymmetry. Therefore, to determine the CP-violating phase, these contributions must be disentangled. The relative sizes of these polarization amplitudes can be determined from the angular distribution of the decay

particles. To be able to determine the fractions of the polarizations, and hence the CP-violating phase, an angular analysis therefore needs to be performed.

In the final state there are three different linear polarizations: longitudinal (A_0), transverse parallel ($A_\parallel$), which both are CP-even; and the transverse perpendicular ($A_\perp$) polarization, which is CP-odd. The three different polarizations are shown in figure 2.12.

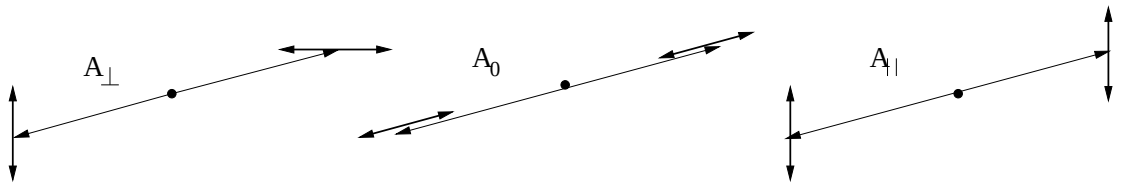


Figure 2.12: *Linear polarizations of the vector mesons for the different amplitudes. Since the initial B is spinless, the spin of the final state needs to be compensated by angular momentum of the system. From this plot it is clear that the 'transverse and perpendicular' polarized amplitude $A_\perp$ has angular momentum $L = 1$, and therefore is CP-odd. The longitudinal (A_0) and 'transverse and parallel' ($A_\parallel$) polarized amplitudes are superpositions of $L = (0, 2)$ states and are CP-even.*

Since there are four final state particles, three angles suffice to describe the decay. The CP properties are most prominent in the transversity frame. The transversity frame is shown in figure 2.13. In this frame the xy-plane is defined by the plane in which the daughters

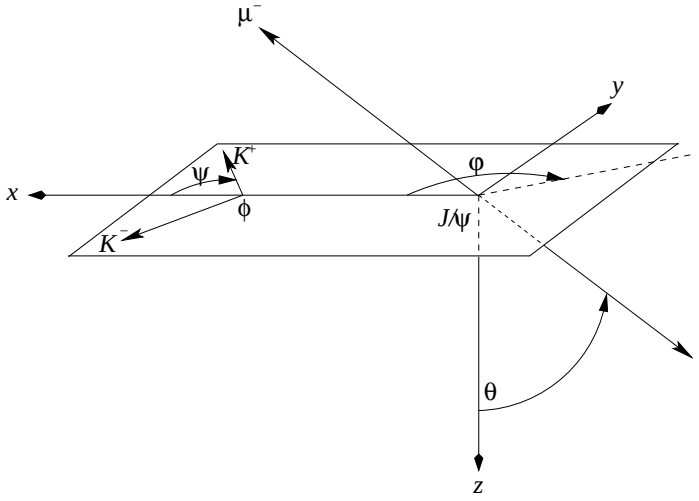


Figure 2.13: *Definition of the transversity frame. The angles ϕ, θ are defined in the frame of the J/ψ, whereas the angle ψ is defined in the frame of the ϕ.*

of the ϕ-meson decay. The $\hat{x}$-axis is defined by the direction of the ϕ-meson momentum vector in the B_s^0 rest frame; the $\hat{y}$-axis in this plane is defined perpendicular to the $\hat{x}$-axis and such that $p_y(K^+) > 0$. The $\hat{z}$-axis is chosen to complete the right-handed system. The

angles (ϕ, θ) describe the decay direction of the positive muon in the J/ψ transversity rest frame; ψ is the helicity angle of the positive kaon in the ϕ rest frame.

In the plane in which the (spinless) kaons decay, the polarization of the J/ψ is directly related to the orbital angular momentum of the system. The axis perpendicular to that plane, $\hat{z}$, corresponds to the polarization axis for which $L = 1$, the CP-odd transversity amplitude $A_\perp$. This can also be understood from figure 2.12, which, in combination with figure 2.13, leads to the following relation between the axes and linear polarization amplitudes [17]

$$A_0 : \epsilon = \hat{x}\,;\;\; A_\parallel : \epsilon = \hat{y}\,;\;\; A_\perp : \epsilon = \hat{z}. \tag{2.71}$$

The CP-even amplitudes $A_{0,\parallel}$ are superpositions of the $L = (0, 2)$ states.

Using the helicity formalism, the differential time and angular decay rate is derived in appendix A. In the transversity convention the expression reads

$$\frac{\mathrm{d}^4\Gamma}{\mathrm{d}t\mathrm{d}\vec{\Omega}} = \frac{\sum_{i=1}^{6} A_i(t|\vec{\lambda}) f_i(\vec{\Omega})}{\int\int \sum_{j=1}^{6} A_j(t|\vec{\lambda}) f_j(\vec{\Omega}) \mathrm{d}t\mathrm{d}\vec{\Omega}}. \tag{2.72}$$

The sums run over products of time dependent amplitude functions and the corresponding angular dependent functions, as given below. $A_i(t|\vec{\lambda})$ represent the time (t) dependent functions, with $\vec{\lambda}$ the physics parameters. $f_i(\vec{\Omega})$ are the angular dependent functions in the transversity frame, with the angular observables $\vec{\Omega} \equiv (\cos\theta, \cos\psi, \phi)$, which are defined in figure 2.13. The infinitesimal $\mathrm{d}\vec{\Omega}$ is defined as $\mathrm{d}\vec{\Omega} \equiv \mathrm{d}\cos\theta \mathrm{d}\phi \mathrm{d}\cos\psi$. The implementation of the Monte Carlo (MC) generation of the proper time and angular dependent distribution of the decay is explained in appendix B.

Angular dependent functions

The angular transversity functions $f_i(\vec{\Omega}) = f_i(\cos\theta, \cos\psi, \phi)$ in (2.72) are given by the formulas in table 2.3. The angles are defined in figure 2.13. The conversion to a different convention, used by e.g. the Babar experiment, is given in appendix A.2.

i	$f_i(\cos\theta, \cos\psi, \phi)$	$A_i(t\|\lambda)$	$\bar{A}_i(t\|\lambda)$
1	$2\cos^2\psi[1 - \sin^2\theta\cos^2\phi]$	$\lvert A_0(t)\rvert^2$	$\lvert\bar{A}_0(t)\rvert^2$
2	$\sin^2\psi[1 - \sin^2\theta\sin^2\phi]$	$\lvert A_\parallel(t)\rvert^2$	$\lvert\bar{A}_\parallel(t)\rvert^2$
3	$\sin^2\psi\sin^2\theta$	$\lvert A_\perp(t)\rvert^2$	$\lvert\bar{A}_\perp(t)\rvert^2$
4	$-\sin^2\psi\sin 2\theta\sin\phi$	$\mathrm{Im}(A_\parallel^*(t)A_\perp(t))$	$\mathrm{Im}(\bar{A}_\parallel^*(t)\bar{A}_\perp(t))$
5	$\frac{1}{\sqrt{2}}\sin 2\psi\sin^2\theta\sin 2\phi$	$\mathrm{Re}(A_\parallel^*(t)A_0(t))$	$\mathrm{Re}(\bar{A}_\parallel^*(t)\bar{A}_0(t))$
6	$\frac{1}{\sqrt{2}}\sin 2\psi\sin 2\theta\cos\phi$	$\mathrm{Im}(A_0^*(t)A_\perp(t))$	$\mathrm{Im}(\bar{A}_0^*(t)\bar{A}_\perp(t))$

Table 2.3: *The angular functions f_i and amplitude terms A_i of the $B_s^0 \to J/\psi\phi$ decay in the transversity frame.*

Time dependent functions

Using (2.51), and writing the strong phases of the polarization amplitudes as δ_i, the time dependent amplitude terms in (2.72) read:

$$\begin{aligned}
A_1 &= |A_0(0)|^2 e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) - \cos\phi_s \sinh(\Delta\Gamma_s t/2) + \sin\phi_s \sin(\Delta m_s t)] \\
A_2 &= |A_\parallel(0)|^2 e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) - \cos\phi_s \sinh(\Delta\Gamma_s t/2) + \sin\phi_s \sin(\Delta m_s t)] \\
A_3 &= |A_\perp(0)|^2 e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) + \cos\phi_s \sinh(\Delta\Gamma_s t/2) - \sin\phi_s \sin(\Delta m_s t)] \\
A_4 &= |A_\parallel(0)||A_\perp(0)| e^{-\Gamma_s t}[\sin(\delta_\perp - \delta_\parallel)\cos(\Delta m_s t) - \cos\phi_s \cos(\delta_\perp - \delta_\parallel)\sin(\Delta m_s t) \\
&\qquad - \sin\phi_s \cos(\delta_\perp - \delta_\parallel)\sinh(\Delta\Gamma_s t/2)] \\
A_5 &= |A_\parallel(0)||A_0(0)|\cos(\delta_\parallel - \delta_0) \\
&\qquad \times e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) - \cos\phi_s \sinh(\Delta\Gamma_s t/2) + \sin\phi_s \sin(\Delta m_s t)] \\
A_6 &= |A_0(0)||A_\perp(0)| e^{-\Gamma_s t}[\sin(\delta_\perp - \delta_0)\cos(\Delta m_s t) - \cos\phi_s \cos(\delta_\perp - \delta_0)\sin(\Delta m_s t) \\
&\qquad - \sin\phi_s \cos(\delta_\perp - \delta_0)\sinh(\Delta\Gamma_s t/2)]. \qquad (2.73)
\end{aligned}$$

The decay rates of the $\bar{B}$ are retrieved by substituting $\phi_s \to -\phi_s$ and $A_\perp \to \bar{A}_\perp = -A_\perp$:

$$\begin{aligned}
\bar{A}_1 &= |A_0(0)|^2 e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) - \cos\phi_s \sinh(\Delta\Gamma_s t/2) - \sin\phi_s \sin(\Delta m_s t)] \\
\bar{A}_2 &= |A_\parallel(0)|^2 e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) - \cos\phi_s \sinh(\Delta\Gamma_s t/2) - \sin\phi_s \sin(\Delta m_s t)] \\
\bar{A}_3 &= |A_\perp(0)|^2 e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) + \cos\phi_s \sinh(\Delta\Gamma_s t/2) + \sin\phi_s \sin(\Delta m_s t)] \\
\bar{A}_4 &= -|A_\parallel(0)||A_\perp(0)| e^{-\Gamma_s t}[\sin(\delta_\perp - \delta_\parallel)\cos(\Delta m_s t) - \cos\phi_s \cos(\delta_\perp - \delta_\parallel)\sin(\Delta m_s t) \\
&\qquad + \sin\phi_s \cos(\delta_\perp - \delta_\parallel)\sinh(\Delta\Gamma_s t/2)] \\
\bar{A}_5 &= |A_\parallel(0)||A_0(0)|\cos(\delta_\parallel - \delta_0) \\
&\qquad \times e^{-\Gamma_s t}[\cosh(\Delta\Gamma_s t/2) - \cos\phi_s \sinh(\Delta\Gamma_s t/2) - \sin\phi_s \sin(\Delta m_s t)] \\
\bar{A}_6 &= -|A_0(0)||A_\perp(0)| e^{-\Gamma_s t}[\sin(\delta_\perp - \delta_0)\cos(\Delta m_s t) - \cos\phi_s \cos(\delta_\perp - \delta_0)\sin(\Delta m_s t) \\
&\qquad + \sin\phi_s \cos(\delta_\perp - \delta_0)\sinh(\Delta\Gamma_s t/2)]. \qquad (2.74)
\end{aligned}$$

It should be noted that there is an ambiguity in the solution of these equations, as they are invariant under the simultaneous substitution

$$\begin{aligned}
\phi_s &\leftrightarrow \pi - \phi_s, \\
\Delta\Gamma_s &\leftrightarrow -\Delta\Gamma_s, \\
\delta_\parallel &\leftrightarrow 2\pi - \delta_\parallel, \\
\delta_\perp &\leftrightarrow \pi - \delta_\perp. \qquad (2.75)
\end{aligned}$$

The possibility of resolving this ambiguity is discussed in section 2.7.2, where the S-wave contribution in the K^+K^- amplitude is discussed.

Finally, note that due to the interference terms, also an untagged sample of $B_s^0 \to J/\psi\phi$ events provides sensitivity to the CP asymmetry. This is due to the fact that some terms in which ϕ_s appears (e.g. the $\cos\phi_s \sinh(\Delta\Gamma_s t/2)$ terms) do not flip sign in the CP-conjugated decay. This is discussed in more detail in chapter 5. However, for optimal sensitivity to ϕ_s, it is needed to measure the flavor of the B_s^0 meson at production.

2.7.1 Single transversity angle analysis

To demonstrate the necessity of an angular analysis, it is instructive to discuss the single angle (transversity) analysis. In the single angle analysis only the transversity angle θ is

measured. Integrating the decay rate (2.73) over the angles $\cos\psi, \phi$, the decay rate reads

$$\begin{aligned}\frac{\mathrm{d}^2\Gamma}{\mathrm{d}t\mathrm{d}\cos\theta} &= (1-|A_\perp(t)|^2)\times(1+\cos^2\theta)/2 + |A_\perp(t)|^2\times(1-\cos^2\theta)\\ &= (1-R_{odd})\times(1+\cos^2\theta)/2\\ &\quad\times e^{-\Gamma_s t}[\cosh(\Delta\Gamma t/2) - \cos\phi_s\sinh(\Delta\Gamma t/2) + \sin\phi_s\sin(\Delta m_s t)]\\ &\quad + R_{odd}\times(1-\cos^2\theta)\\ &\quad\times e^{-\Gamma_s t}[\cosh(\Delta\Gamma t/2) + \cos\phi_s\sinh(\Delta\Gamma t/2) - \sin\phi_s\sin(\Delta m_s t)]. \quad (2.76)\end{aligned}$$

Here R_{odd} ($\equiv |A_\perp(0)|^2/(|A_\perp(0)|^2+|A_\parallel(0)|^2+|A_0(0)|^2)$) is the fraction of CP-odd polarized states.

Example: $B_d^0 \to J/\psi K^*(K_S\pi_0)$

As an example, consider the $B_d^0 \to J/\psi K^*(K_S\pi_0)$ decay: the B_d equivalent of $B_s^0 \to J/\psi\phi$. Here a B_d decays into a CP-eigenstate consisting of two decaying vector mesons. In the B_d system $\Delta\Gamma_d = 0$ and the substitution[8] $\phi_s \to \phi_d \equiv 2\beta$ is made. In this case (2.76) reads

$$\begin{aligned}\frac{\mathrm{d}^2\Gamma}{\mathrm{d}t\mathrm{d}\cos\theta} &= (1-R_{odd})e^{-\Gamma_d t}[1+\sin\phi_d\sin(\Delta m_d t)]\times(1+\cos^2\theta)/2\\ &\quad + R_{odd}e^{-\Gamma_d t}[1-\sin\phi_d\sin(\Delta m_d t)]\times(1-\cos^2\theta), \quad (2.77)\end{aligned}$$

and the observable oscillation amplitude a_{obs} of the oscillation equals

$$a_{obs} = (1-2R_{odd})\times\sin\phi_d. \quad (2.78)$$

This example illustrates why the observable $\cos\theta$ can be used for a single angle analysis: the transversity angle shows different behavior for CP-even and CP-odd amplitudes and makes it possible to estimate the dilution due to the admixture.

An example of transversity angle and proper time distributions with both CP-even and CP-odd components is shown in figure 2.14. As shown, from the angular distribution the fraction of CP-odd polarized decays R_{odd} can be estimated. This fraction gives the dilution in the observable proper time oscillation a_{obs}, due to the admixture of CP-even and CP-odd eigenstates in the final state. Only by performing an angular analysis one can then determine $\sin\phi_d$. The same reasoning holds for $B_s^0 \to J/\psi\phi$: since the observable oscillation amplitude is damped by a factor $1-2A_\perp$, the decrease in sensitivity is the same for an increased mistagging, as explained in chapter 6.

2.7.2 S-wave contribution

In the previous section only resonant final state K mesons in an $L=1$ state (P-wave), originating from an intermediate ϕ meson are considered. However, the K meson pair can also be in an $L=0$ state, either resonant (a^0, f^0) or non-resonant, possibly with an invariant mass in the ϕ mass region. These so-called S-wave contributions are described and studied

[8]Note that $\phi_d = 2\beta$, whereas $\phi_s = -2\beta_s$. This is due to the difference in the definition of β and β_s in the SM.

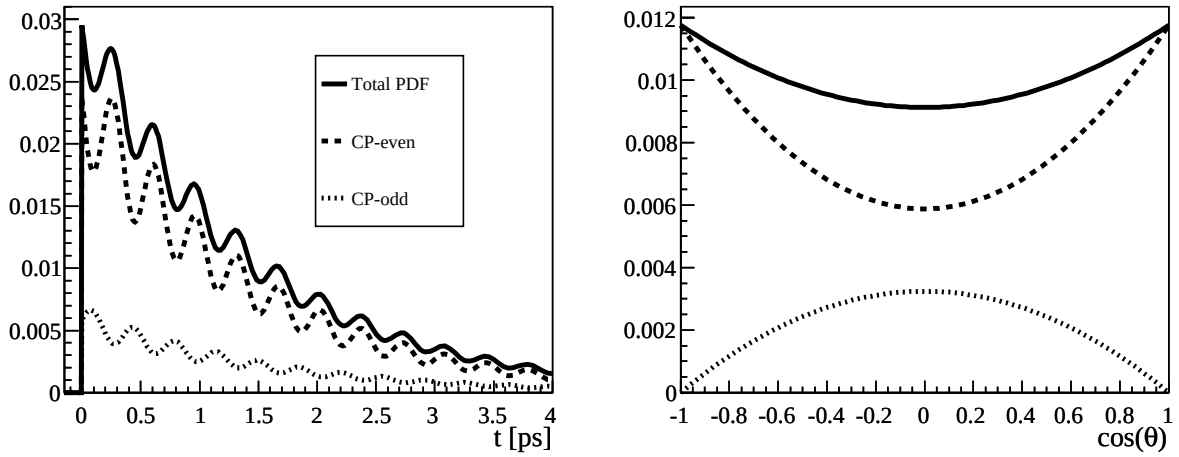


Figure 2.14: *Example of the theoretical proper time PDF (left) and the* $\cos\theta$ *PDF (right) of the* $B_s^0 \to J/\psi\phi$ *decay. The behavior of the CP-even component and the CP-odd component are different in both observables. The proper time-dependent CP-violating amplitude (*$\phi_s = -0.2$*) is diluted due to the admixture of both components, but the dilution (the fraction CP-odd* $|A_\perp(0)|^2 = 20\%$*) can be estimated from the angular distribution.*

in [31], for the helicity frame. This section summarizes the main results and describes the relevant angular formulas in the transversity frame.

The extra terms are found by taking states of the K mesons with angular momentum $L = 0$ into account. The contributions of these possible states lead to the extra angular functions of the S-wave fraction and the interference terms, given in table 2.4 for the transversity frame.

i	$f_i(\cos\theta, \cos\psi, \phi)$	$A_i(t\|\lambda)$
7	$2[1 - \sin^2\theta\cos^2\phi]$	$\|A_S(t)\|^2$
8	$\sqrt{6}\sin\psi\sin^2\theta\sin 2\phi$	$\mathrm{Re}(A_S^*(t)A_\parallel(t))$
9	$\sqrt{6}\sin\psi\sin 2\theta\cos\phi$	$\mathrm{Im}(A_S^*(t)A_\perp(t))$
10	$4\sqrt{3}\cos\psi[1 - \sin^2\theta\cos^2\phi]$	$\mathrm{Re}(A_S^*(t)A_0(t))$

Table 2.4: *The angular functions for the S-wave contribution of the* $B_s^0 \to J/\psi\phi$ *decay in the transversity frame. The time dependence of the amplitude terms can be found in [31].*

As can be seen from the table, the CP-odd S-wave component has a $\cos^2\theta$ angular distribution similar to the CP-even signal component, when compared with table 2.3. However, the time-dependent behavior is behaving like the P-wave CP-odd component. Therefore, when the S-wave contribution is not taken into account, two effects will occur. First of all, from the angular distribution, the fraction CP-odd P-wave, $|A_\perp|^2$, will be underestimated. Secondly, the dilution of the proper time dependent oscillation due to the admixture of different polarization states will become smaller. Therefore, when neglected, the CP-violating phase will be estimated erroneously, and will, in realistic conditions, be underestimated.

The bias in the estimate of the CP violating phase, for realistic conditions, is found to be 10% towards zero, for an S-wave component of 10% [31].

The different (resonant and non-resonant) components can be taken into account by modelling the different di-kaon invariant mass (m_{KK}) distributions. An alternative option is to fit the angular distribution according to a P-wave amplitude A_P and an S-wave amplitude A_S, hereby determining $|A_P|/|A_S|$ per bin of m_{KK} [31].

When taken into account, the S-wave contribution can be used to resolve the sign ambiguity in the estimate of ϕ_s, since the strong phase of the S-wave component is known to increase as the invariant mass of the di-muon pair increases. This means the strong phase has to be estimated as a function of m_{KK} [31].

2.8 Control channel: $B_d^0 \to J/\psi K^*(K^+\pi^-)$

To calibrate the analysis of the $B_s^0 \to J/\psi\phi$ decay, the analysis of the $B_d^0 \to J/\psi K^*(K^+\pi^-)$ decay is used. It serves as a control channel for the analysis of both the proper time, the tagging, and the angular distribution of the $B_s^0 \to J/\psi\phi$ decay.

The $B_d^0 \to J/\psi K^*(K^+\pi^-)$ decay is, similar to $B_s^0 \to J/\psi\phi$, a decay of a scalar particle into two vector mesons, decaying into two leptons and two scalars respectively. Therefore, the angular distribution of the $B_d^0 \to J/\psi K^*(K^+\pi^-)$ decay is similar to the $B_s^0 \to J/\psi\phi$ angular distribution. By performing an angular analysis the polarization amplitudes of the $B_d^0 \to J/\psi K^*(K^+\pi^-)$ can be measured.

Since the B_d^0 meson, as does the B_s^0 meson, oscillates, the analysis of the decay rate of $B_d^0 \to J/\psi K^*(K^+\pi^-)$ events can be used to estimate mixing in the B_d^0 system. To measure mixing the flavor of the B_d^0 meson both at production and at decay must be determined. Since the final state of this decay is flavor specific, this channel is self-tagging, which means that the flavor of the B_d^0 at decay can be determined from the charges of the final state particles. As a consequence, the mistag fraction of the B_d^0 meson at production can be determined from the amplitude of the mixing signal. This measurement of the mistag fraction in the $B_d^0 \to J/\psi K^*$ decay can subsequently be used to control the tagging performance of the $B_s^0 \to J/\psi\phi$ decay. Then by performing a tagged proper time analysis the mixing frequency can be determined.

Since the mixing parameter Δm_d, the polarization amplitudes, and the strong phases of $B_d^0 \to J/\psi K^*(K^+\pi^-)$ are already measured in experiments at the B-factories, a measurement of these parameters can be used to check the analysis of the proper time, tagging, and angular distribution of $B_s^0 \to J/\psi\phi$. Hence the $B_d^0 \to J/\psi K^*$ channel is the ideal control channel for the $B_s^0 \to J/\psi\phi$ decay.

Decay rates

The flavor of the signal B-meson can be determined from the decay products using the quark contents of the final state particles. For example, if the final state is $K^* \to K^+\pi^-$, a B_d^0 must have decayed; if the final state is $\bar{K}^* \to K^-\pi^+$, a $\bar{B}_d^0$ must have decayed. This also means that if the direct decay $B \to f$ is possible, the decay of a $\bar{B}$ into f can not be reached directly, but only via mixing: $\bar{B} \to B \to f$.

It follows that the decay amplitudes are:

$$\begin{aligned} A(B \to f) &= g_+, \\ A(\bar{B} \to \bar{f}) &= g_+, \\ A(B \to \bar{f}) &= \frac{q}{p} g_-, \\ A(\bar{B} \to f) &= \frac{p}{q} g_-. \end{aligned} \tag{2.79}$$

The first two are the 'unmixed amplitudes', the latter are 'mixed amplitudes'. The differential proper time and angular decay rate is

$$\frac{\mathrm{d}^4\Gamma}{\mathrm{d}t\mathrm{d}\Omega} = \frac{\sum_{i=1}^{6} A_i(t|\vec{\lambda}) f_i(\Omega)}{\sum_{j=1}^{6} A_j(t|\vec{\lambda}) f_j(\Omega) \mathrm{d}t\mathrm{d}\Omega}. \tag{2.80}$$

The angular functions f_i are the same as for $B_s^0 \to J/\psi\phi$, and given in table 2.3.

Assuming $|q/p| = 1$ and $\Delta\Gamma = 0$, the time dependent amplitudes are the following. For the events occurring via a direct decay (i.e. the unmixed events) the amplitude terms read

$$\begin{aligned} |A_i(t)|^2 &= |A_i(0)|^2 \frac{e^{-\Gamma_d t}}{2}(1 + \cos\Delta m_d t) \\ \mathrm{Im}(A_{\parallel,0}^*(t) A_\perp(t)) &= |A_{\parallel,0}(0)||A_\perp(0)| \sin(\delta_{\parallel,0} - \delta_\perp) e^{-\Gamma_d t}(1 + \cos\Delta m_d t) \\ \mathrm{Re}(A_\parallel^*(t) A_0(t)) &= |A_\parallel(0)||A_0(0)| \cos(\delta_\parallel - \delta_0) e^{-\Gamma_d t}(1 + \cos\Delta m_d t). \end{aligned} \tag{2.81}$$

For the events decaying after mixing (i.e. the mixed events) the amplitude terms read

$$\begin{aligned} |A_i(t)|^2 &= |A_i(0)|^2 \frac{e^{-\Gamma_d t}}{2}(1 - \cos\Delta m_d t) \\ \mathrm{Im}(A_{\parallel,0}^*(t) A_\perp(t)) &= |A_{\parallel,0}(0)||A_\perp(0)| \sin(\delta_{\parallel,0} - \delta_\perp) e^{-\Gamma_d t}(1 - \cos\Delta m_d t) \\ \mathrm{Re}(A_\parallel^*(t) A_0(t)) &= |A_\parallel(0)||A_0(0)| \cos(\delta_\parallel - \delta_0) e^{-\Gamma_d t}(1 - \cos\Delta m_d t). \end{aligned} \tag{2.82}$$

2.9 Analysis strategy

To perform a measurement of CP violation in the $B_s^0 \to J/\psi\phi$ decay, reconstructing signal and suppressing backgrounds, the experimental requirements follow below.

In order to reconstruct the decay accurately, three requirements have to be met. First of all, since $\Delta m_s = 17.8\ \mathrm{ps}^{-1}$, the proper time dependent oscillations of the CP asymmetry occur with high frequency, and the decay distance and time needs to be determined with high precision. Hence precise vertexing and track reconstruction is required. Precise track reconstruction is also needed for the angular analysis of the decay particles. Secondly, an untagged sample of $B_s^0 \to J/\psi\phi$ decays does provide some sensitivity to ϕ_s, however only at second order. Hence, flavor tagging (determining whether the B meson at production was a B or $\bar{B}$) enhances the sensitivity for ϕ_s. Finally, the current uncertainty on the amplitude of the CP asymmetry in $B_s^0 \to J/\psi\phi$ is $\sim 10°$, following from measurements at the Tevatron. In order for LHCb to contribute to the sensitivity significantly, assuming similar performance as the Tevatron experiments, at least ~ 10 k $B_s^0 \to J/\psi\phi$ signal decays are required.

In order to minimize misidentifications and suppress background, particle identification of muons and kaons needs to be performed, since the studied final state consists of two muons and two kaons. To improve the mass resolution, which increases the possibility to separate background and signal, good particle momentum resolution is required.

In next chapter the LHCb experiment is described, showing that in the design of this experiment all ingredients needed for an analysis of the $B_s^0 \to J/\psi\phi$ decay are present.

Chapter 3

The LHCb experiment

The LHCb experiment is a dedicated B-physics experiment situated at the LHC accelerator. Its main purpose is to study CP-violation and weak interactions in the B-meson system, and to measure the branching ratios of rare B-decays. The goal is to overconstrain the CKM matrix of weak interactions and search for new physics.

In the previous chapter the theoretical decay distributions of the $B_s^0 \to J/\psi\phi$ channel have been described, which allow to make a measurement of the CP asymmetry in this channel. To detect and reconstruct these decays, requirements need to be fulfilled, as described in section 2.9. This chapter describes the design of the LHCb subdetectors and their expected performances, and it is shown that all 'ingredients' are present to meet these requirements. The description is based on [5], in which more detailed information and references can be found.

After describing the properties of the LHC accelerator, a short overview of the LHCb detector is given. Subsequently, the subdetectors used for tracking and particle identification are described. Finally, the system that makes it possible to select the most interesting events while reducing the event rate, the trigger, is described.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is an approximately circular proton-proton collider at CERN. Before being accelerated by the LHC, protons pass through several pre-accelerators, as shown in figure 3.1. From an initial linear accelerator (LINAC2), they are subsequently fed through the Proton Synchroton Booster (BOOSTER), the Proton Synchrotron (PS), and the Super Proton Synchroton (SPS) from which they are finally injected into the LHC at an energy of 450 GeV.

At the four interaction points, shown in figure 3.1, the four main experiments at the LHC are situated: ALICE, ATLAS, CMS and LHCb. ALICE is an experiment dedicated to heavy ion physics; ATLAS and CMS are general purpose experiments, which primarily aim to discover on-shell production of new particles. Since these particles are expected to have large masses, and since the processes by which they are produced have small cross sections, the LHC is designed with both a center-of-mass energy and a luminosity as large as possible. At the time of writing, the accelerator is being commissioned, and a record center-of-mass energy of 7 TeV has been reached.

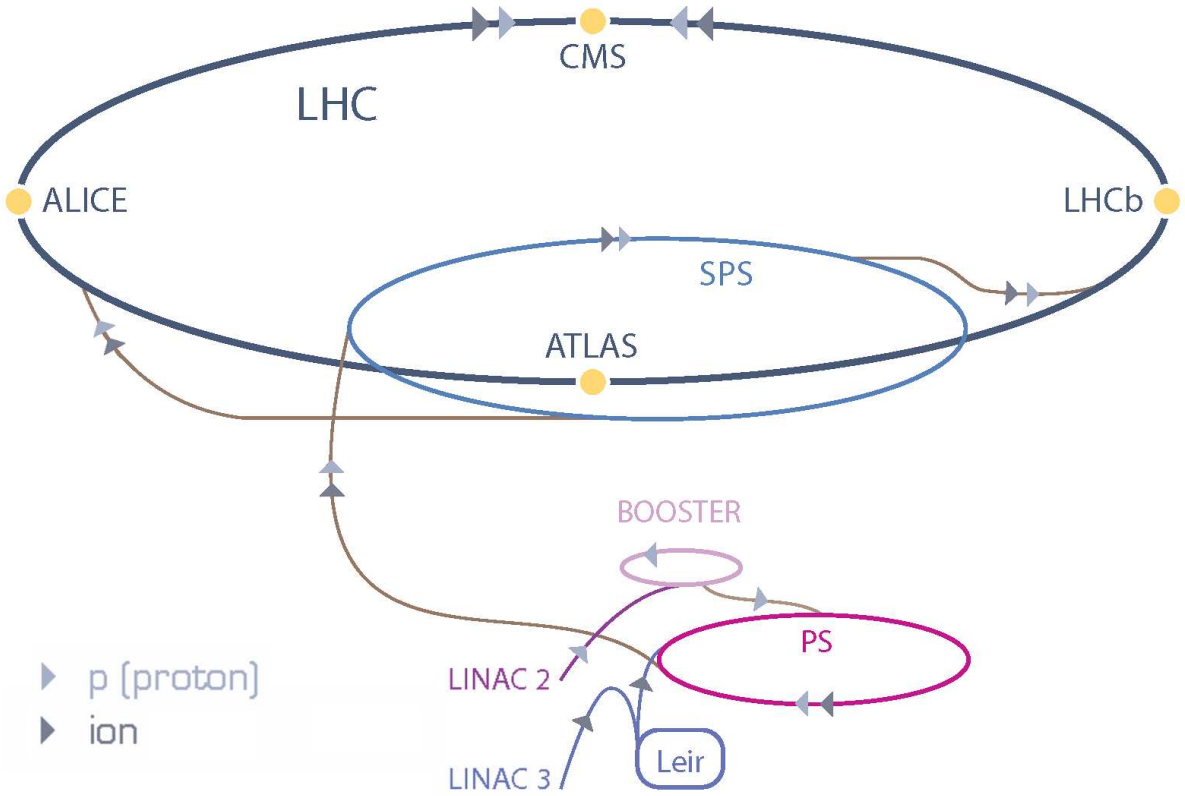


Figure 3.1: *Layout of part of the CERN accelerator complex. Before being accelerated in the LHC and collided at the interaction points of ALICE, ATLAS, CMS, ans LHCb, protons pass through the LINAC2, the BOOSTER, the PS and the SPS. Ions are accelerated by LINAC3 and Leir before being injected in the BOOSTER.*

The operation of the LHC can be summarized as follows. Two beams move in opposite direction and are kept in orbit around the 27 km circumference of the accelerator by the magnetic field of superconducting magnets. The superconducting coils are kept at a temperature of 1.9 K (by a cryostat), at which a maximum magnetic field of 8.33 T is generated. This field allows to reach the design center-of-mass energy of $\sqrt{s} = 14$ TeV. Finally, the bunches in which the particles are concentrated are designed to collide with a frequency of 40 MHz, and can be squeezed at the interaction points to achieve a design luminosity of 10^{34} cm^{-2}s^{-1}. The LHC design parameters are summarized in table 3.1.

Yield of $B_s^0 \to J/\psi\phi$ events

At the time of writing the LHC is not able to reach the design center-of-mass energy of 14 TeV. Since the on-shell production of new particles depends strongly on the center-of-mass energy, this is critical for the direct searches for new physics processes. But since the cross sections of the CP-violating B meson decays LHCb aims to perform measurements of depend less on the center of mass energy, the center of mass energy is less crucial to

Circumference	26659 m
Center of mass energy	14 TeV
Injection energy	450 GeV
Field at 2×450 GeV	0.535 T
Field at 2×7 TeV	8.33 T
Helium temperature	1.9 K
Luminosity	$10^{34}\text{cm}^{-2}\text{s}^{-1}$
Bunch spacing	25 ns
Luminosity lifetime	10 h
Time between 2 fills	7 h

Table 3.1: *The main LHC design parameters.*

the indirect searches for new physics. In this thesis the $b\bar{b}$ cross section is assumed to be $\sigma_{b\bar{b}} = 500$ μb.

Since the ratio of the cross sections of inelastic collisions and $b\bar{b}$ collisions is $\sigma_{inel}/\sigma_{b\bar{b}} \approx 160$, and since the branching fraction of interesting decays is small ($\leq \mathcal{O}(10^{-3})$), a selective trigger is crucial for the LHCb experiment to reduce the event rate to storage. To improve the performance of the trigger, as explained in section 3.5, the number of bunch-crossings with single interactions is maximized. Therefore, the instantaneous luminosity at LHCb, when compared to the general purpose experiments, is reduced to $2{\times}10^{32}$ cm^{-2}s^{-1}, leading to an expected integrated annual[1] luminosity at LHCb of $\mathcal{L} = 2.0$ fb^{-1}.

The annual yield S of $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decays at LHCb can be calculated as follows[2],

$$S = \mathcal{L} \times \sigma_{b\bar{b}} \times 2 \times f_{b\to B_s} \times \text{BR}_{B_s^0 \to J/\psi\phi} \times \text{BR}_{J/\psi\to\mu\mu} \times \text{BR}_{\phi\to KK}, \tag{3.1}$$

with $f_{b\to B_s}$ the $b \to B_s^0$ hadronization fraction, and BR_i the relevant branching ratios. Since the hadronization fraction $f_{b\to B_s} = 11\%$, and

$$\text{BR}_{B_s^0 \to J/\psi(\to\mu\mu)\phi(\to KK)} = 9.3 \times 10^{-4} \times 5.93\% \times 49.2\% = 2.7 \times 10^{-5}, \tag{3.2}$$

it finally follows that $S = 6.0$ M events in which $B_s^0 \to J/\psi(\to \mu\mu)\phi(\to KK)$ occur at the LHCb interaction point in a nominal year. Since approximately 20% of these decays have all final state tracks in the acceptance of the detector [5], the LHCb detector at the LHC accelerator can detect enough signal events to perform a competitive measurement of $B_s^0 \to J/\psi\phi$.

3.2 The LHCb detector

The choice of the LHCb detector geometry is justified by the fact that, at the LHC design energy, both the b- and $\bar{b}$-hadrons are predominantly produced in the same forward, or

[1] An LHCb year is defined as 10^7 s.

[2] The factor 2 originates from the fact that both the b and the $\bar{b}$ quark can hadronize into an (anti) B_s meson.

backward, cone, as shown in figure 3.2(b). Therefore, the LHCb detector is designed as a single-arm spectrometer with a forward angular coverage ranging from approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane. The restriction to the forward region allows for a tracking volume with a large $\int Bdl$ and hence a good momentum resolution. Furthermore, equipment needed for readout of the subdetectors can be placed outside of the detector acceptance, keeping the interactions in the detector to a minimum.

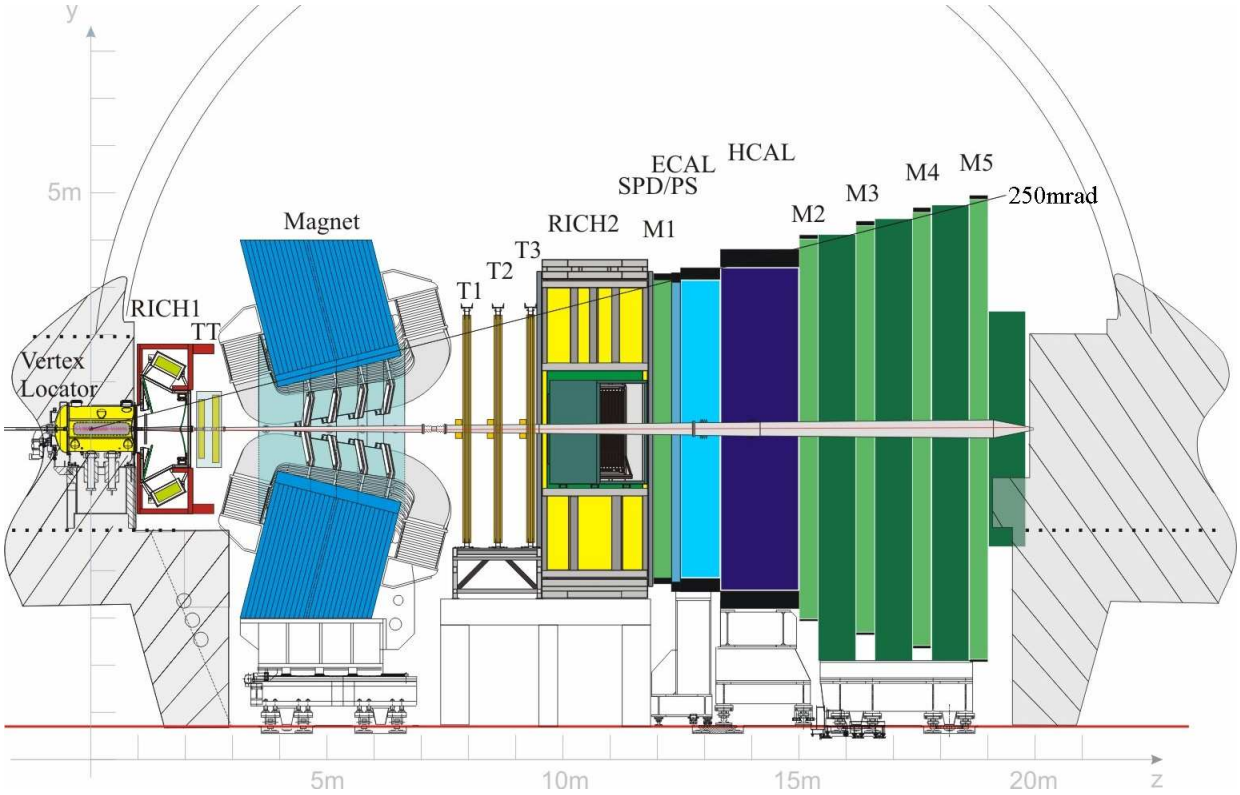


Figure 3.2: *The setup of the LHCb detector, with the different subdetectors shown in the vertical plane. The right-handed coordinate system adopted has the z axis along the beam and the y axis along the vertical. The magnetic field in this plane is vertically, hence this is referred to as the non-bending plane.*

The layout of the LHCb spectrometer is shown in figure 3.2(a). From the point of view of the analysis of the $B_s^0 \to J/\psi\phi$ decay channel, the detector can be divided in two systems: the tracking system (consisting of VELO, TT, T1-T3 and a dipole magnet) and the subdetectors for particle identification (RICH1,2, the calorimeters and M1-5).

The conical shape of the LHCb detector plays an important role in the $B_s^0 \to J/\psi\phi$ analysis. Due to this shape final state particles with relatively large transverse momentum can escape the detector without detection, which, via correlations with angular observables, can cause non-uniform efficiencies in the angular observables of the decay.

For the same reason does the region around the beam axis, in which no detector material is placed, play an important role in the analysis of $B_s^0 \to J/\psi\phi$. Due to this hole in the detector acceptance final state particles with relatively little transverse momentum

can escape the detector without detection. Because of this, and because of correlations between kinematical observables of final state signal particles in the lab frame and angular observables in the rest frame of the decay, the presence of the beam pipe causes non-uniform efficiencies in the angular observables of the $B_s^0 \to J/\psi\phi$ decay.

The LHC beam is accommodated inside the LHCb experiment by a beam pipe, which is designed to minimize the creation of secondary particles, while withstanding the external air pressure at a design beam vacuum of 10^{-9} mbar. The beam pipe consists of a thin exit window (located at $z = 858$ mm), sealed to the VELO vacuum tank, followed by two conical parts of 25 mrad and 10 mrad, respectively. Both parts are made of beryllium to minimize the secondary interactions of the particles with material. After $z = 13$ m, where the amount of material is less critical, the beam pipe is constructed of stainless steel.

3.3 Tracking system

Charged particle detection is important in the $B_s^0 \to J/\psi\phi$ analysis for two main reasons: vertexing (determining the distance between the production and the decay vertex of the B_s) and momentum reconstruction. These two together provide the mass, the angular and the proper time resolution, important for the offline signal reconstruction and background suppression. Furthermore, information about momentum and decay distance are used in the trigger.

The tracking system consists of a dipole magnet and the subdetectors used for tracking. These are the vertex locator system (VELO) and four planar tracking stations: the Tracker Turicensis (TT), consisting of four layers upstream of the dipole magnet, and T1-T3, consisting of at least 12 layers downstream of the magnet. For occupancy reasons, silicon microstrip detectors are used in VELO and TT, as well as in T1-T3 in the region close to the beam pipe (Inner Tracker, IT). In the outer region of T1-T3 (Outer Tracker, OT), where the occupancy is lower, straw tubes are employed.

3.3.1 Vertex Locator

To provide precise measurements of track coordinates close to the interaction region, the Vertex Locator (VELO), consisting of a series of silicon modules, is arranged along the beam direction. It is used to identify the detached secondary vertices typical for b-hadron decays and makes it possible to meet the requirement to reconstruct $B_s^0 \to J/\psi\phi$ decays with a proper time resolution good enough to resolve the fast time-dependent oscillations of the CP asymmetry.

To provide accurate measurements of the position of the vertices, the silicon modules of the VELO are placed as close as possible to the beam axis, namely at 8 mm. In order to detect the majority of the tracks originating from the beam spot ($\sigma = 5.3$ cm), the detector is designed such that tracks emerging up to $z = 10.6$ cm downstream from the nominal interaction point cross at least 3 VELO stations, for a polar angular window between 15 and 300 mrad, as shown in figure 3.3.

To enable fast reconstruction of tracks and vertices in the LHCb trigger, a cylindrical geometry with silicon strips measuring $r\phi$ coordinates is chosen for the modules. The

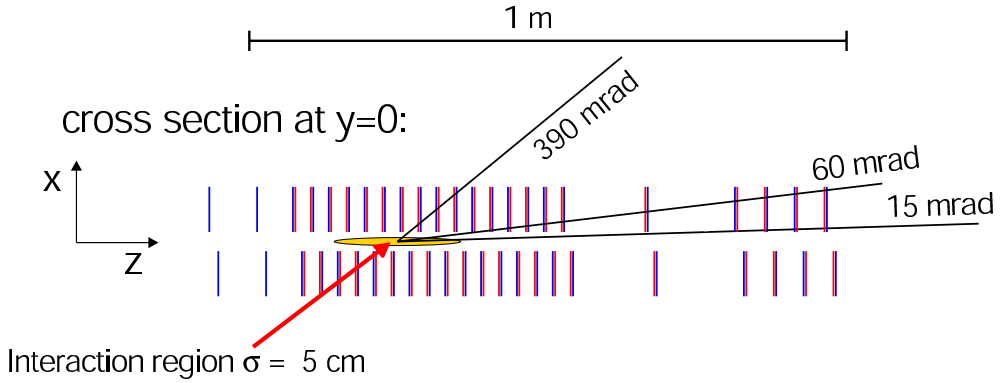


Figure 3.3: *The setup of the VELO silicon modules along the beam direction. The left two pairs form the pile-up system. Indicated are the average crossing angle for minimum bias events (60 mrad), and the minimial (15 mrad) and maximal (390 mrad) angle for which at least 3 VELO stations are crossed. 390 mrad is the opening angle of a circle that encloses a rectangular opening angle of* 250×300 *mrad [36].*

strips of the r sensor are concentric semi-circles, the ϕ sensors measure a coordinate almost orthogonal to the r-sensor. The geometry is shown in figure 3.4. A 2D reconstruction in the r-z plane alone allows to detect tracks originating from close to the beam line in the high-level trigger. These measurements are used to compute the impact parameter of tracks with respect to the production vertex, which is used in the trigger to discriminate between signal and background. The level-0 trigger uses information from the pile-up veto system, two stations located upstream, which make it possible to reject events with multiple pp interactions in one beam crossing. A detailed discussion on the usage of the VELO in the trigger can be found in section 3.5.

The setup of the VELO is as follows. The half disc sensors are arranged in pairs of r and ϕ sensors and are mounted back-to-back. The sensors are 300 μm thick, radiation tolerant, n-implants in n-bulk technology [35]. The minimal pitch of both the r and the ϕ sensors is 32 μm, linearly increasing towards the outer radius at 41.9 mm. To reduce the strip occupancy and pitch at the outer edge of the ϕ-sensors, the ϕ-sensor is divided in two parts. The outer region starts at a radius of 17.25 mm and has approximately twice the number of strips as the inner region. The strips in both regions make a 5° stereo angle with respect to the radial to improve pattern recognition, and adjacent stations are placed with opposite angles with respect to the radial. In order to fully cover the azimuthal angle with respect to the beam axis, the two detector halves overlap, as is shown in figure 3.4.

To minimize the amount of material between the particle vertices active detector layers the detector is placed inside vacuum. To separate the primary beam vacuum from the secondary vessel vacuum and shield the detector from RF pickup from the beam, the sensors are separated from the beam vacuum by a thin aluminium foil. Both the sensors and this so-called RF-foil are contained inside a vacuum vessel. During beam injection the two halves of the VELO are retracted 3 cm away from the nominal beam position. The RF-foil is designed to minimize interactions.

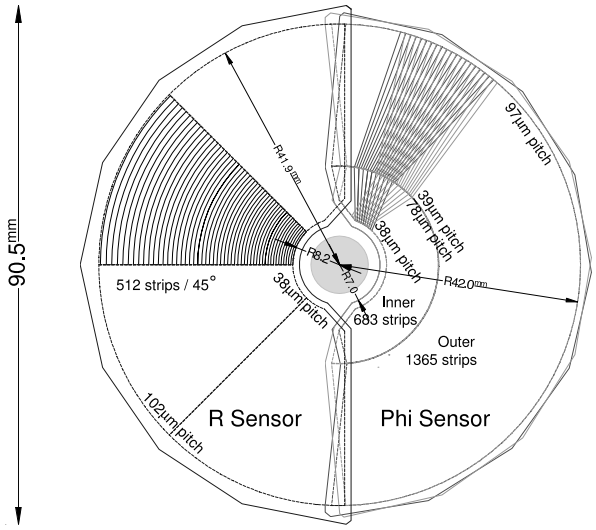


Figure 3.4: *A VELO module and its dimensions. Shown are the r and ϕ sensors.*

3.3.2 Magnet

Following from the requirement of a good momentum resolution, the LHCb experiment utilizes a (dipole) magnet, which bends the tracks of charged particles. The warm (i.e. not superconducting) magnet consists of two saddle-shaped coils. These are placed mirror symmetrically, such that the gap left open by the magnet is slightly larger than the LHCb acceptance, and the principal field component is vertical throughout the detector acceptance.

The quantity important for momentum resolutions, and hence for the analysis of $B_s^0 \to J/\psi\phi$ channel, is the integrated magnetic field the magnet delivers. For tracks passing through the entire tracking system this is [5]

$$\int B\mathrm{d}l = 4\,\mathrm{Tm}, \tag{3.3}$$

making it possible to measure charged particles' momenta up to 200 GeV within 0.5 % uncertainty.

3.3.3 Inner Tracker

To perform accurate momentum estimates, important for mass, angular and proper time resolutions in the reconstruction of the $B_s^0 \to J/\psi\phi$ channel, hit information downstream of the magnet is required, which is provided by three tracking stations. Since the magnet bends particles in the horizontal direction perpendicular to the beam pipe ($\hat{x}$), the track density is largest in an elliptically shaped region around the beam pipe. In order to have similar occupancies over the plane, a detector with finer detector granularity is required in

this region. Therefore, the Inner Tracker (IT), 120 cm wide and 40 cm high, as shown in figure 3.5, is located in the center of the three tracking stations.

Because of the high track density near the beam pipe, silicon strip detectors are used. The total active detector area covers 4.0 m^2, consisting of 129024 readout strips of either 11 cm or 22 cm in length. To improve track reconstruction, the four detector layers are arranged in an x-u-v-x geometry, in which the strips are vertical in the first and in the last layer, whereas the other two (u, v) layers are rotated by stereo angles of $\pm 5°$, providing the sensitivity in the vertical $(\hat{y})$ direction.

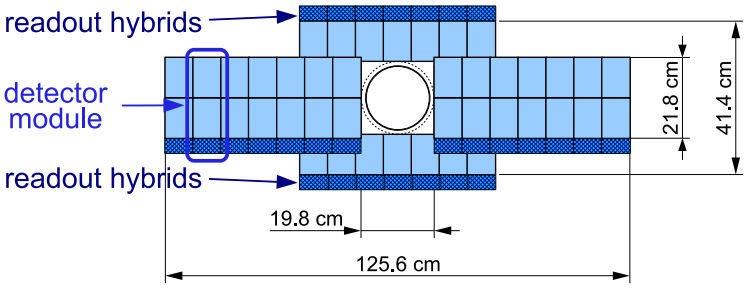


Figure 3.5: *Layout of the IT.*

The pitch of the single-sided p^+-on-n strips is 198 μm. In order to have similar performance in terms of signal-to-noise, the thickness of the sensors is 320 μm for the single-sensor ladders below and above the beam pipe, and 410 μm for the double sensors at the sides of the beam pipe. The four layers are housed in 4 boxes, which are placed such that they overlap. These overlaps avoid gaps in the detector and, more importantly, make it possible to perform alignment using reconstructed tracks.

3.3.4 Outer Tracker

Similar to the IT, the Outer Tracker (OT) performs track measurements downstream of the magnet, allowing to determine the momenta of charged particles. The OT covers the outer region of the three tracking stations T1-T3.

Since the track density further away from the beam pipe is lower, straw tubes are used. The total active area of one station is 5971×4850 mm^2, and the OT and the IT together cover the full acceptance of the experiment. As is the case for the IT, these layers are also arranged in an x-u-v-x geometry, as shown in figure 3.6(a).

The OT is designed as an array of individual, gas-tight straw-tube modules as shown in figure 3.6(b). Each module contains two layers of drift-tubes with an inner diameter of 4.9 mm. The front-end (FE) electronics measures the drift time of the ionization clusters produced by charged particles traversing the straw tubes, digitizing it with respect to every bunch crossing. Given the bunch crossing rate of 25 ns and the diameter of the tube, and in order to guarantee a fast drift time (below 50 ns) and a sufficient drift-coordinate resolution (200 μm), a mixture of Argon (70%) and CO_2 (30%) is used as counting gas.

Crucial quality assurance tests of the Outer Tracker are described in appendix C. The response of the individual channels are tested using radioactive sources, both after production and after installation. Furthermore, tests of the gas tightness of individual modules

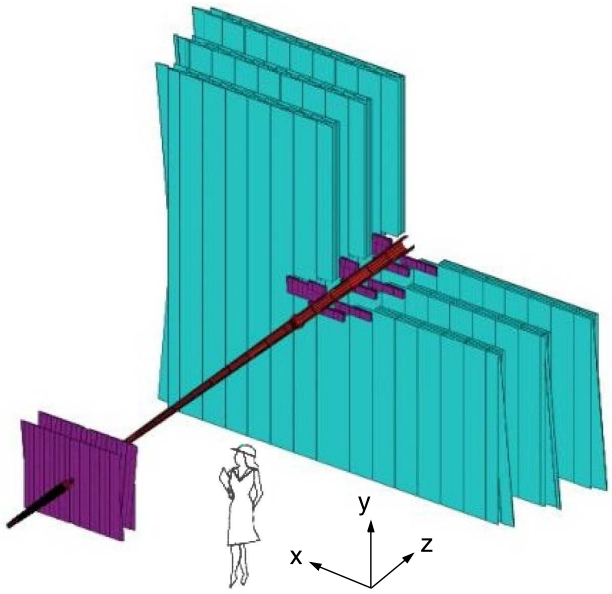


Figure 3.6: *The layout of the OT, shown together with the IT and the TT*

and of the gas propagation are performed. Also, the performance of the FE electronics after assembly is tested.

3.3.5 Tracker Turicensis

To improve the momentum estimate of charged particles, track measurements are performed before these enter the magnet. Therefore, the Tracker Turicensis[3] (TT), a planar tracking station, is located between the VELO and the LHCb dipole magnet. It is also used to perform the track measurements of long lived neutral particles which decay after the VELO, in particular K_S^0 and Λ. Furthermore, by using the (weak) magnetic field inside the tracker, track information from the TT is used by the High Level Trigger to confirm candidates between the VELO and the tracking stations.

In order to cover the full acceptance of the experiment, the TT is constructed 150 cm wide and 130 cm high. It consists of four detector layers, with a total active area of 8.4 m^2, with 143360 readout channels, up to 38 cm in length. To improve track reconstruction, the four detector layers are arranged in two pairs that are separated by approximately 27 cm along the LHCb beam axis. And again, like the IT and the OT, the TT detection layers are in an x-u-v-x arrangement.

The layout of one of the detector layers is illustrated in figure 3.7. Its basic building block is a half module that covers half the height of the LHCb acceptance. It consists of a row of seven silicon sensors, named a ladder. The silicon sensors for the TT are 500 μm thick, single sided p^+-on-n sensors, as for the IT. They are 9.64 cm $\times$ 9.44 cm long and carry 512 readout strips with a strip pitch of 183 μm.

[3]Turicensis is Latin for Zürich. The abbreviation used to stand for Trigger Tracker.

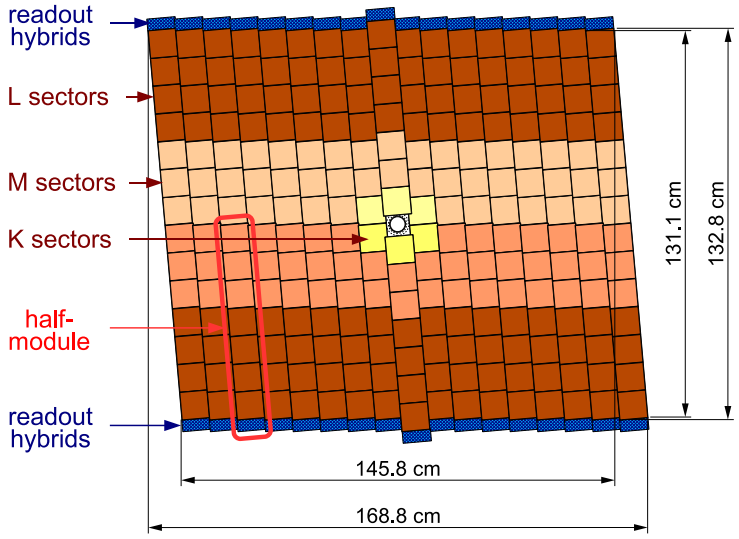


Figure 3.7: *Layout of one of the stereo plane detector layers of the TT.*

3.3.6 Tracking performance

The dimensions of the strips of all three silicon detectors, VELO, TT, and IT, are chosen such that the signal to noise ratio is greater than 10, and such that the high occupancies can be dealt with. Given the intrinsic resolutions of VELO, TT, IT, and OT and the magnetic field described in section 3.3.2, the combined performance, important for the analysis of the $B_s^0 \rightarrow J/\psi\phi$ channel, can be summarized as follows. The momentum resolution is $\frac{\delta p}{p} \sim (0.3 - 0.5)\%$ [5], depending on the momentum of the charged particle. Together with the vertexing performed by the VELO this leads to a B mass resolution of typically 15-20 MeV/c^2 and a proper time resolution, typical for B decays at LHCb, of $\sim$ 40 fs [**?**]. The angular resolutions of reconstructed $B_s \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decays are typically $\sim$ 25 mrad, as is shown below. Finally, the tracking efficiency from simulations is $> 95\%$ for tracks from B decays [5], crossing the entire detector, ensuring that most of the $B_s^0 \rightarrow J/\psi\phi$ decays inside the acceptance of the detector are reconstructed.

3.4 Particle identification

The final state of the $B_s^0 \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay consists of two kaons and two muons. As discussed in section 2.9, this leads to requirements on the particle identification in the LHCb experiment. First of all, particle identification allows to suppress backgrounds, in particular from pions, particle identification needs to be performed. Furthermore, information about the particle types in an event is used by the trigger. Finally, it is used in the determination of the flavor of B mesons at the time of their production.

In this section the RICH, calorimeter and muon system are described. To suppress backgrounds from pions in the $\phi \rightarrow K^+K^-$ reconstruction, information from the RICH system is used to determine the mass of the particles in which the ϕ meson decays. Furthermore,

to suppress backgrounds from pions in the $J/\psi \to \mu^+\mu^-$ reconstruction, information from the muon system is used. Using the muon system also an efficient and selective trigger is made, and muons from semi-leptonic b decays are used to provide a tag of the flavor of the B meson at the time of production. Finally, the calorimeter is of less importance in the analysis of $B_s^0 \to J/\psi\phi$.

3.4.1 RICH system

In order to be able to distinguish kaons from pions, the Ring Imaging Cherenkov (RICH) system measures the velocity of a charged particle. Since at large polar angles the momentum spectrum is softer than at small polar angles, the system is composed of two detectors which, when combined, cover the full momentum range. The upstream detector, RICH 1, covers the full detector acceptance and the low momentum particle range $\sim 1 - 60$ GeV/c, using aerogel and C_4F_{10} radiators. The downstream detector, RICH 2, covers the high momentum range from ~ 15 GeV/c up to and beyond 100 GeV/c, using a CF_4 radiator, and covers the region where high momentum particles are produced.

Both detectors use the same principle to determine the particle type: when particles traverse a medium with a speed larger than the local macroscopic speed of light, Cherenkov radiation in the shape of a cone is emitted. Knowledge of the angle of the Cherenkov cone θ_C and of the refractive index of the medium n suffices to determine the speed v of the particle, using the relation

$$\cos\theta_C = \frac{1}{n\beta}, \tag{3.4}$$

where $\beta = v/c$, with c the speed of light. Together with the momentum of the particle as determined using the tracking system, the mass of the particle can be calculated, and hence the type of the particle. The relationships between θ_C and the momentum, for different particle masses, are shown in figure 3.8(b).

To analyze the emitted Cherenkov radiation, and have as little material as possible in the detector acceptance, the cones of emitted light are reflected by spherical and flat mirrors. This can be seen in figure 3.8(a), where the layout of RICH 1 is shown. Outside of the detector acceptance, Hybrid Photon Detectors are placed, on which the cones appear as circles. The layout of RICH 2 is similar.

After analysis of the circles in the RICH detectors, the K-identification efficiency is 90%, whereas the probability to misidentify a pion as a kaon is $\sim 3\%$ [5]. The estimated likelihood that a track reconstructed as a kaon truly is a kaon is used in the selection process.

3.4.2 Muon system

Muon triggering and offline muon identification are fundamental requirements to select events of the $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay. The muon system provides fast information to construct a high-p_T muon trigger at the earliest level (L0) and clearly identify muons in the high-level trigger (HLT) and in the offline analysis.

The muon system, shown in figure 3.9(a), is composed of five stations (M1-M5) of rectangular shape, placed along the beam axis at the most downstream end of the spectrometer. The transverse dimensions scale with the distance from the interaction point, such that the

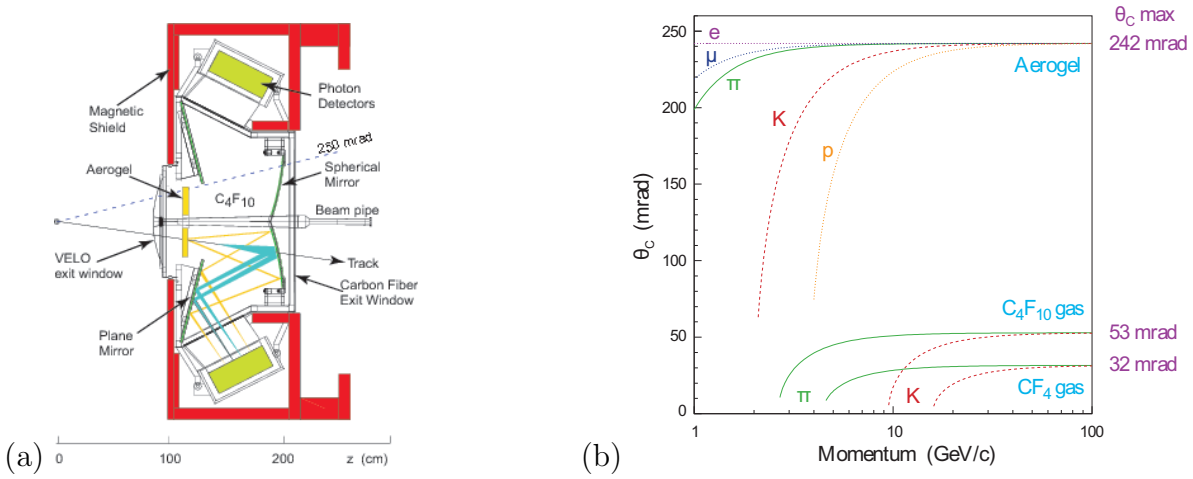


Figure 3.8: *Schematic layout of the RICH 1 detector (a). Graph of the Cherenkov angle θ_C versus the momentum. Combining the information of the velocity provided by the RICH system with the momentum information provided by the tracking system gives information about the mass (b).*

full detector acceptance is covered. Stations M2-M5 are placed downstream of the calorimeters and are interleaved with iron absorbers 80 cm thick to select penetrating muons. The minimum momentum of a muon to cross the five stations is approximately 6 GeV/c, since the total absorber thickness, including the calorimeters, is approximately 20 interaction lengths [5]. To improve the p_T measurement in the trigger, station M1 is placed in front of the calorimeters.

The muon trigger uses standalone muon track reconstruction and p_T measurement and requires aligned hits in all five stations. Therefore, the efficiency of each station must be high enough to obtain a trigger efficiency of at least 95%. Stations M1-M3 have a high spatial resolution along the x coordinate. They are used to define the track direction, and to calculate the p_T of the candidate muon with a resolution of 20% [5]. Stations M4 and M5 have a limited spatial resolution, as their main purpose is the identification of penetrating particles.

The layout of the muon stations is shown in figure 3.9(b). Multi-wire proportional chambers (MWPC) are used for all regions, except the innermost region of station M1, where the occupancy is highest, and where, because of ageing considerations, triple-GEM detectors are used. Each muon station is divided into four regions, R1 to R4, with increasing distance from the beam axis. The linear dimensions of the regions, and their segmentation, scale in the ratio 1:2:4:8. With this geometry, the particle flux and channel occupancy are expected to be roughly the same over the four regions of a given station. The layout also shows the partitioning into logical pads and the (x,y) granularity.

The likelihood a track in the muon stations originates from a true muon is determined by analyzing the hits in a window around this track. This likelihood can be used in the event selection algorithm. Furthermore, information from the muon system is used for tagging purposes.

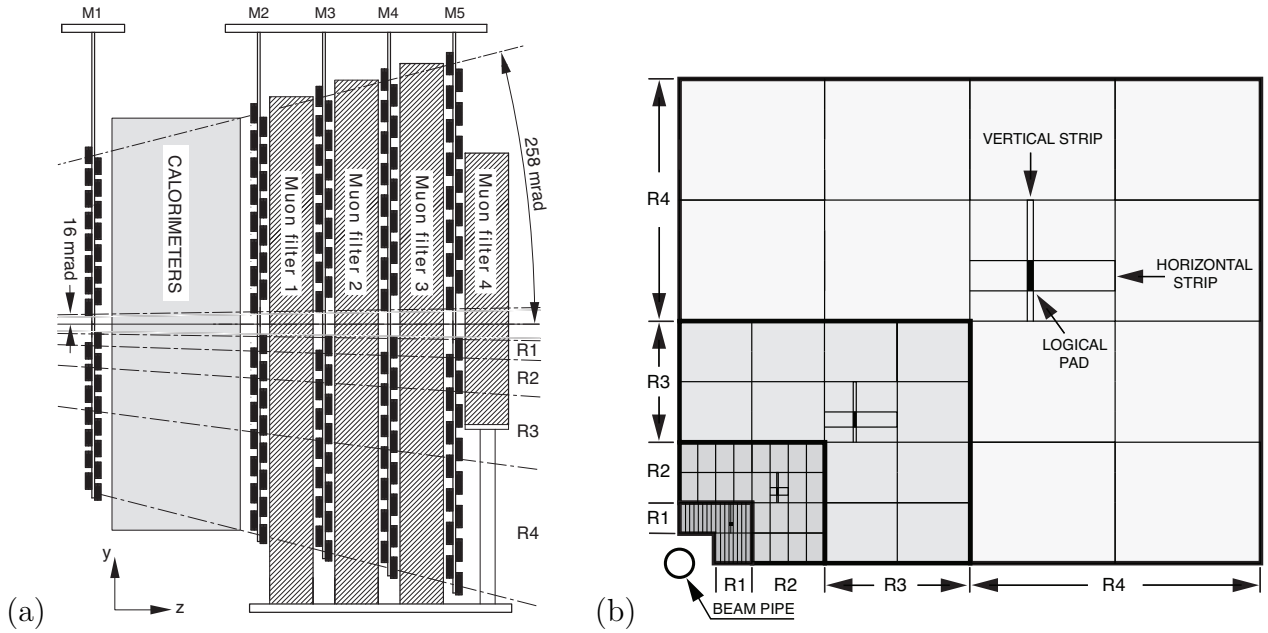


Figure 3.9: *The setup of the muon stations (a). Layout of one muon station, where the horizontal and vertical pads and logical pads are indicated (b).*

3.4.3 Calorimeter system

To select hadron, electron and photon candidates with large transverse energy, and thereby distinguish signal from background, information from the calorimeter system is used by the first trigger level (L0). Furthermore, the calorimeter system provides the separate identification of electrons, photons and hadrons, as well as the measurement of their energy and position. Since for the $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay the muon trigger is already efficient and selective, the calorimeter system does not play an important role in the analysis of this decay.

In downstream order, the calorimeter system consists of the following detectors: a scintillator pad detector (SPD) is followed by a preshower detector (PS), after which an electromagnetic calorimeter (ECAL) and a hadron calorimeter (HCAL) are installed. Information from the ECAL and HCAL is used to select electron, photon, and hadron candidates with large transverse energy. The SPD is used to reject π^0 mesons in the electron trigger, whereas the PS is used to reject charged pions.

Because the hit density varies by two orders of magnitude over the calorimeter surface, the PS/SPD, ECAL and HCAL adopt a variable lateral segmentation (shown in figure 3.10). The ECAL is divided into three different sections, whereas the HCAL is segmented into two zones and has larger cell sizes, because of the differences in shower sizes.

The most important role the calorimeter plays in the analysis of the $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay comes from the thickness of the calorimeters. The ECAL is 25 radiation lengths, the thickness of the HCAL is 5.6 interaction lengths, which prevents most hadrons from entering the muon chambers, helping muon identification.

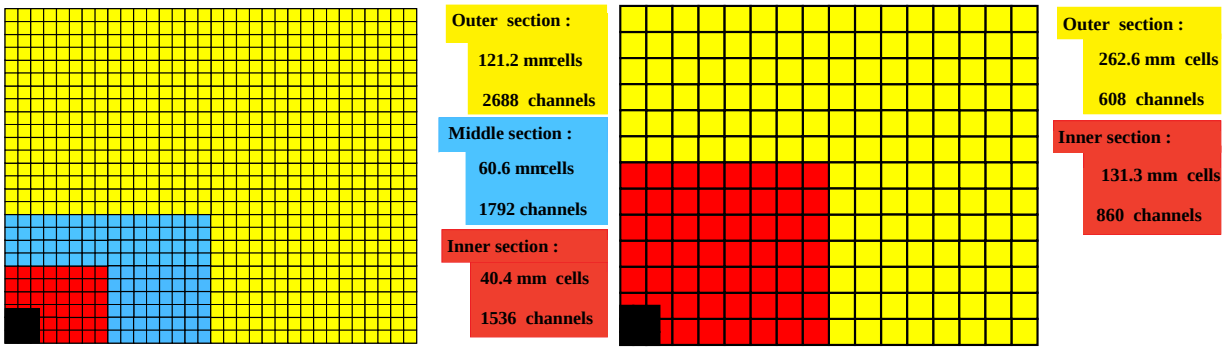


Figure 3.10: *Segmentation and dimensions of the ECAL (left) and the HCAL (right).*

3.5 Trigger

Reading out the channels of all described subdetectors of the LHCb detector, one event under normal conditions has a typical size of $\sim$ 30 kB. Because of this event size and the limited storage, events are written to tape, for offline analysis, with a rate of $\sim$ 2 kHz. To perform this reduction in a way which is efficient for the signal events, given the available resources, the trigger is divided into different levels.

The two levels into which the trigger is divided, are: the Level-0 (L0), implemented in hardware, and the High Level Trigger (HLT), a C++ application executed on the Event Filter Farm (EFF). The L0 reduces the visible interactions of about 10 MHz [5] to a rate of 1 MHz, at which rate the HLT can have access to the full event data. Since the HLT is implemented in software, it can be adjusted to developments in event reconstruction and selection.

In order to be able to perform a complete reconstruction in the HLT, it is divided in two parts. The HLT1 first reduces the rate to $\sim$ 30 kHz, aimed to confirm the candidates found in the L0 trigger by performing partial pattern recognition on selected track candidates. After this reduction the HLT2 can perform both inclusive and exclusive trigger selections using full pattern recognition. A diagram of the trigger flow is shown in figure 3.11.

Finally, it should be noted that the selection performed offline on the events on storage should be the optimal selection: the offline selection, using the offline reconstruction, should be made with the aim to have the best sensitivity to the signal parameters. The mere purpose of the trigger is to reduce the rate written to storage. Furthermore, the online reconstruction in general is worse than the offline reconstruction. Therefore, the online selection cuts are, in general, relaxed compared to the offline selection.

3.5.1 L0 trigger

The L0 electronics is implemented in custom-made electronics and operates synchronously with the 40 MHz brunch crossing frequency. The purpose of the L0 trigger is to reduce the LHC beam crossing rate of 40 MHz to the rate of 1 MHz at which the entire detector can be read out by the HLT.

To efficiently and quickly reduce the rate, the L0 uses simple observables to discriminate

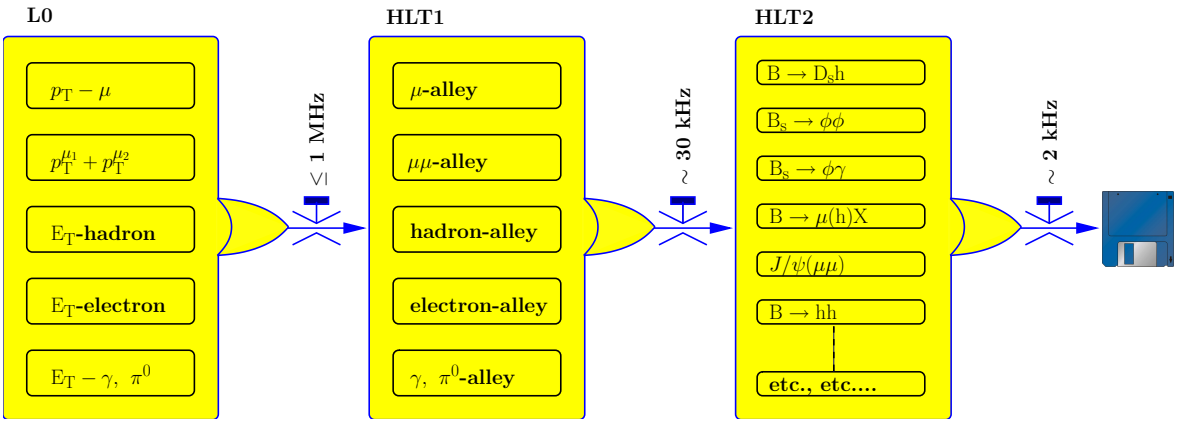


Figure 3.11: *Diagram of the flow of events in the trigger. The output of the L0 is* ≤ 1 *MHz, the output of the HLT1 is* ~ 30 *kHz, and the output of the HLT2 is* ~ 2 *kHz.*

between signal and background events. Since B mesons, due to their relatively large mass, often decay into particles with relatively large transverse momentum p_T and energy E_T, the L0 trigger attempts to reconstruct the hadron, electron and photon clusters with highest E_T in the calorimeter, as well as the two muons in the muon chambers with highest p_T. A nominal set of thresholds in the L0 are

1. At least one cluster in the HCAL with $E_T^{\text{hadron}} > 3.5$ GeV.
2. At least one cluster in the ECAL with $E_T^{e,\gamma,\pi^0} > 2.5$ GeV.
3. A muon candidate in the muon chambers with $p_T^{\mu} > 1.2$ GeV or two muons with $p_T^{\mu_1} + p_T^{\mu_2} > 1$ GeV.

These criteria correspond to the L0 decisions shown in the flow diagram in figure 3.11.

Furthermore, since events with a large number of tracks and primary vertices would occupy a too large fraction of the computing power of the HLT, these events may be rejected. In order to avoid events with high multiplicities, information of the calorimeters is used to calculate the total observed energy and estimate the number of tracks, using the number of hits in the SPD. In order to reject events with a large number of primary vertices, the pile-up system in the VELO allows to reconstruct the number of primary pp interactions in each bunch crossing.

By combining the measurements of the clusters with highest p_T and E_T, and the estimates of the number of tracks and primary vertices, the decision to accept an event is derived.

3.5.2 High level trigger

HLT1

To further reduce the L0 output event rate from 1 MHz to a rate at which both inclusive and exclusive selections can be performed in the HLT2, the HLT1 demands a confirmation

of the previous trigger stage. Therefore, the HLT1 divides L0 candidates into independent *lines*, shown in figure 3.12. The specific HLT1 lines which are executed depend on the L0 decision output.

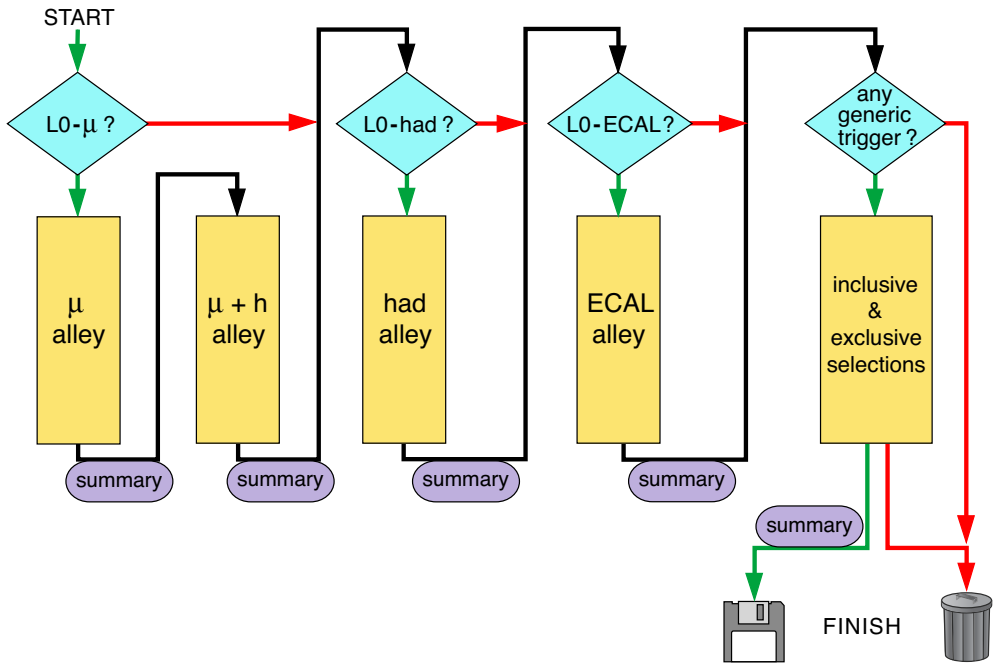


Figure 3.12: *Flow of events in the HLT1. Depending on the L0 decision an HLT1 line follows. After that, a confirmation of the previous trigger stage by an HLT1 selection is required.*

The confirmation can happen in two, similar, ways.

1. In the first case L0→T matching is performed: L0 calorimeter or muon particle candidates are assumed to originate from the interaction region, and seeds in the T-stations, matching in both space and momentum, are sought for. Then T→VELO matching is performed: given the matched T-seeds, matching VELO tracks are sought for.

2. In the other case L0→VELO matching is performed first: r-z tracks in the VELO, matching an L0 object, are demanded to match also using the VELO ϕ-sensor information, resulting in 3D tracks. Then VELO→T matching is performed: given a 3D VELO track, T-station tracks matching in both space and momentum are sought for.

Finally, the expected HLT1 performance, is as follows. The most important lines for the $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ analysis are the HLT1 μ lines. The input rate into the HLT1 μ lines is expected to be $\sim$ 230 kHz. After confirmation and possible impact parameter cuts, this rate is reduced to 10 kHz, which is about 30% of the total bandwidth. HLT1 reduces the total rate to about 30 kHz, at which rate both inclusive and exclusive selections can be performed.

HLT2

The output rate of events accepted by the HLT1 lines is sufficiently low to perform selections using the full event information. Composite particles are selected by HLT2, using requirements on e.g. the invariant mass or the lifetime of the particle. The inclusive and exclusive selections aim to reduce the rate to about 2 kHz, a rate at which data is written to storage for further analysis. If the total rate is too large, lines can be prescaled (or postscaled), meaning that only a fraction of the triggered events as input (or output) of a certain line is randomly[4] selected.

In order to write a limited amount of events to tape, while rejecting as little signal events as possible, an efficient HLT2 selection is needed for the $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay. The optimized offline selection, and the resulting HLT2 selection is described in the following chapter.

[4]The selection is performed 'quasi' random, meaning that the output is predictable.

Chapter 4

Selection

All ingredients required to reconstruct the $B_s^0 \to J/\psi\phi$ decay are present in the LHCb experiment. However, to keep the number of events selected both by the trigger and by the offline selection at an acceptable rate, discriminating event selection criteria are required. Furthermore, the trigger and offline selection should be performed in such a way that the sensitivity to ϕ_s is maximized.

In this chapter a selection algorithm is presented which meets both design goals by dividing the selected sample in two parts. After critically reviewing the existing selections, the concept of dividing the sample of events in 'prescaled' and 'detached' samples is explained. Following the description of the optimization procedure to maximize the sensitivity to ϕ_s, an offline selection is estimated. Subsequently, the way to divide the sample in two parts is determined. After estimating the selection performance in terms of efficiencies, purities, and trigger rates, guidelines to operate the trigger are given. Finally, the improvement compared to the selection in [20] is presented and discussed.

4.1 Introduction to $B_s^0 \to J/\psi\phi$ selections

Previous MC studies of the $B_s^0 \to J/\psi\phi$ channel in LHCb show that the major background component is the so-called 'prompt background', which has a lifetime distribution peaking around zero. These background events are due to combinations of tracks, of which at least one originates from the primary proton-proton interaction vertex. A significant fraction of this background consists of events where a prompt J/ψ was produced [20].

This background increases the rate significantly and decreases the contribution to the sensitivity of the physics parameters of signal events with small t. To keep the trigger rate limited, two methods have been used in the LHCb experiment for the $B_s^0 \to J/\psi\phi$ analysis. However, neither have been optimized for the extraction of ϕ_s, and both solutions introduce disadvantages.

One solution adopted in previous MC studies is to introduce proper time significance cuts [45]. A disadvantage of this method is that an efficiency is introduced which is not uniform as a function of proper time and which goes to zero for small proper time. Usually, the events at small negative proper time are used to determine the proper time resolution. Since the events in this region are removed, the proper time resolution can not be determined

from this sample alone. Furthermore, the shape of the proper time efficiency can not be determined from this sample alone.

An alternative solution is to implement a set of multiple cuts [44] in order to keep the trigger rate at an acceptable level, while ensuring a constant proper time efficiency. The disadvantage of such a method is that many cuts are required, each introducing possible systematic effects. In particular, multiple p_T cuts are introduced, which, in general, introduce angular acceptance effects, on top of the effects introduced by the detector and trigger. Besides this, the reconstruction mass windows of the daughter mesons must be chosen narrow in order to suppress the rate. In particular, the mass window of the ϕ meson is only 12 MeV/c^2, whereas for a study of the S-wave component, as described in section 2.7.2, a mass window of at least 20 MeV/c^2 would be preferred, as explained in [31]. Finally, to achieve an efficiency which is constant as a function of the proper time, only a limited set of trigger criteria can be used, resulting in a a sub-optimal efficiency and a reduced event yield.

4.2 Introduction to detached and prescaled selections

Here a third selection method is introduced and described, avoiding the above mentioned problems by dividing the sample into sub-samples. Furthermore, one of the sub-samples is used to optimize the sensitivity to ϕ_s.

The most important motivation of this selection method is that the events in the region around $t = 0$ form a significant contribution to the trigger rate, whereas the signal events in this region hardly contribute to the sensitivity to ϕ_s, as will be shown later. Therefore, the main sample is selected with an explicit lifetime cut $t > \tau_0$ applied, thereby avoiding the dominant prompt background around $t = 0$. And since this sample is the main contributor to the sensitivity to ϕ_s, the selection is optimized on only this sample. This sample, with mainly signal events, is called the *detached sample*.

The second sample is selected without any cut that induces a non-uniform proper time efficiency. The main purpose of this sample is to allow determination of the proper time resolution from the data, which is not possible with the detached sample. Furthermore, it is used to determine the efficiency as a function of the proper time as induced by the selection of the detached sample. Since the determination of the proper time resolution from the prompt background does not require a high selection efficiency, this sample is prescaled, as explained in section 3.5.2. This prescaling keeps the event rate at an acceptable level. This sample is named the *prescaled sample* or *control sample*. Since this sample hardly contributes to the sensitivity to ϕ_s, it is not used in the selection optimization.

In figure 4.1 the efficiency as a function of proper time for the two samples is schematically shown. The proper time efficiency of the detached sample is a step function: equal to 0 for $t < \tau_0$ and equal to 1 for $t > \tau_0$. The efficiency of the prescaled sample is constant as a function of the proper time, but low and equal to the known prescale fraction.

In case the lifetime cutoff is placed online, the efficiency after offline reconstruction is smeared, due to the difference between online and offline reconstruction. This complicates the shape of the proper time efficiency, but this shape can be determined from data by comparing the detached sample and the prescaled control sample, as explained in more

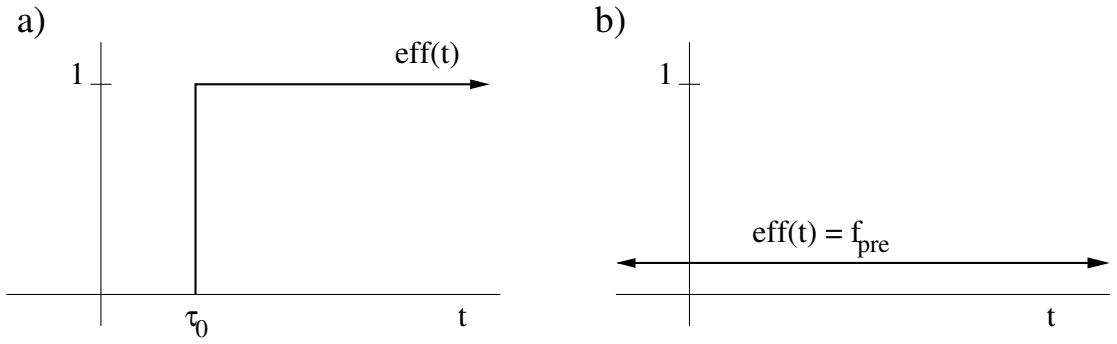


Figure 4.1: *The relative efficiency of the detached sample (a) and the efficiency of the prescaled sample (b), as induced by the selection. The (online) lifetime cut in the selection of the detached sample is equal to τ_0, the efficiency of the selection of the prescaled sample is equal to the prescale factor f_{pre}. In order to guarantee a uniform efficiency, events in the prescaled sample are required to be triggered by triggers which do not induce non-uniform proper time efficiencies.*

detail in chapter 5.

It should be noted that the prescaled sample requires events to be selected by triggers which do not induce non-uniform proper time efficiencies. The detached sample can include events which are selected by any trigger, also by those which induce a non-uniform proper time efficiency. This further complicates the shape of the relative proper time efficiency, but, as mentioned, this can be determined from data. The fact that more trigger lines can be used, adds sensitivity to ϕ_s as well as robustness.

Furthermore, both efficiencies get multiplied by the possibly non-trivial overall proper time efficiency introduced by the offline selection, the detector acceptance, and L0. This is a separate issue, which is discussed below, where it is shown that ignoring this efficiency is a valid assumption for statistics corresponding to the first years of data taking.

After the optimization procedure, using the detached sample, the effects of the lifetime cut (τ_0) and the prescale fraction ($f_{\rm pre}$) are studied. The optimal values of these depend on the maximal allowed (trigger) rate and are studied at the end of the chapter.

4.3 Optimization procedure

To determine the optimal selection cuts an optimization procedure is used. Using a figure of merit, the sensitivity to the measurement of ϕ_s is optimized, iteratively adjusting a set of rectangular selection cuts. A small number of selection criteria is used, each of which does not affect the proper time distribution and which are expected to be uncorrelated. Furthermore, usage of a minimal number of kinematic cuts is proposed, minimizing the number of possible systematic effects in the distortion of the angular distribution of the final state particles.

The goal of the optimization procedure is to minimize $\sigma(\phi_s)$, the uncertainty of ϕ_s.

Therefore a figure of merit, named FOM,

$$\text{FOM} \propto 1/\sigma(\phi_s), \tag{4.1}$$

is introduced. In the optimization procedure FOM is maximized, which is equivalent to minimizing the uncertainty of ϕ_s.

The correct figure of merit in terms of signal and background event yields depends on the specific analysis. In a counting experiment, in which a signal box is defined, the uncertainty on the number of signal events in the signal box is

$$\sigma(S) = \sqrt{S+B}, \tag{4.2}$$

where B represents the number of background events in the box. This means that if the mere purpose of an optimization procedure is to minimize the relative uncertainty $\sigma(S)/S$ in a counting experiment, then FOM $= S/\sqrt{S+B}$ is the correct variable to use in the optimization procedure.

If, instead of an analysis based solely on counting, an analysis is performed in which the shape of the distribution of both signal and background events are described after the event selection, then signal and background can be distinguished better. An example is the simultaneous description of both an invariant mass distribution and a proper time distribution by a single PDF. Not only the description of the peaking invariant mass distribution makes it possible to distinguish signal from background, also the different behavior of signal and background in the proper time distribution contributes to the discriminating power. Therefore, the dilution of the sensitivity to signal parameters due to background events is smaller in the case an observable is used in the description of the data than in the case that merely the number of events is estimated.

In other words, when performing a fit, describing the event distributions as function of the observables not used in the event selection, the background dilutes the sensitivity to the signal parameters less than in the counting experiment, in the presence of the same number of background events in the signal box. Therefore, (4.2) is not correct anymore, and the sensitivity to S becomes

$$\sigma(S) = \sqrt{S+B^*}, \tag{4.3}$$

where the 'effective background' B^* is introduced.

To finally determine the correct variable to be used in the optimization of the $B_s^0 \to J/\psi\phi$ selection, it should be noted that the sensitivity $\sigma(\phi_s)$ is proportional to the error of the number of signal events (S), depending on the effective number of background events (B^*), as

$$\begin{aligned} \sigma(\phi_s) &\propto \frac{\sigma(S)}{S} \\ &= \frac{\sqrt{S+B^*}}{S}. \end{aligned} \tag{4.4}$$

Therefore, the figure of merit,

$$\text{FOM} = \frac{S}{\sqrt{S+B^*}} \tag{4.5}$$

is the correct variable to maximize when the $B_s^0 \to J/\psi\phi$ selection is optimized. B^*, with in general $B^* \leq B$, provides the effective dilution due to background, for certain signal and background distributions, and depends on the shape of the signal and background distributions.

Before the optimization procedure is described, first the data samples and selection variables are discussed.

4.3.1 Data samples

The events used in the selection optimization result from time-consuming MC simulations. This means the simulated minimum bias sample is too small to be used for the description of the background events in the selection optimization.

Instead, two samples are used which together are expected to be representative for the background. An 'inclusive J/ψ' sample, consisting of the pp interactions in which a J/ψ is produced, is used for the description of the prompt component of the background; a '$b\bar{b}$ di-μ' sample, consisting of the pp interactions in which a $b\bar{b}$ quark pair is produced and two muons are contained in the final state, is used for the description of the long-lived component of the background.

In table 4.1, the MC data samples used for the selection optimization can be found. To optimize the selection and calculate the annual yields and trigger rates, all samples used have been weighted to represent an equivalent of 2.0 fb^{-1}, which is a nominal year of data taking. In general, the samples are produced with requirements on the decay products to be in the acceptance of the detector. The combination of this generation efficiency, the sample size and the cross section lead to the 'effective luminosity' of each sample.

Purpose	Monte Carlo sample	Cross section	Luminosity	V
Optimization				
Signal	$B_s^0 \to J/\psi\phi$	$2.713 \times 10^{-5} \times 500$ μb	17 fb^{-1}	1
Background	$b\bar{b}$ di-μ (no signal)	$2.21\% \times 500$ μb	5.4 10^{-3} fb^{-1}	2
Background	Inclusive J/ψ (no $b\bar{b}$)	286 μb	7.7 10^{-5} fb^{-1}	2
Performance				
Background	Minimum bias L0 accepted	1 MHz	1.1 10^{-6} fb^{-1}	2
Background	Inclusive J/ψ	286 μb	7.7 10^{-5} fb^{-1}	3
Signal	$B_s^0 \to J/\psi\phi$	$2.713 \times 10^{-5} \times 500$ μb	20 fb^{-1}	3
Signal	$B_s^0 \to J/\psi\phi$ ($\Delta\Gamma = \phi_s = 0$)	$2.713 \times 10^{-5} \times 500$ μb	0.4 fb^{-1}	3
Signal	$B_s^0 \to J/\psi\phi$ ($\phi_s = -0.7$)	$2.713 \times 10^{-5} \times 500$ μb	19 fb^{-1}	3
Calibration	$B_d^0 \to J/\psi K^*$	5.1 10^{-5} $\times 500$ μb	0.3 fb^{-1}	3

Table 4.1: *The data samples used for the selection optimization, with their sizes and versions specified. Two types of reconstruction have been used for the samples used in the optimization: signal and background are reconstructed with different definitions of the variables used for particle identification. The versions V are the following: 1=(DaVinci v21r0p1, Brunel v31, DC06), 2=(DaVinci v21r0p1, Brunel v30, DC06), 3=(DaVinci v24r4, MC09).*

It should be noted that various event types might be present in more than one sample. For example, $B_d^0 \to J/\psi K^*$ events appear in both the '$b\bar{b}$ di-μ' and in the 'inclusive J/ψ' sample. To avoid double counting caused by overlapping samples, signal events have been removed from all background samples. First, signal events are removed from the inclusive '$b\bar{b}$ di-μ' sample. Subsequently, all events originating from a $b\bar{b}$ decay have been removed from the inclusive J/ψ sample, such that this samples is also complementary. Finally, it should be noted that the background and signal samples have been reconstructed with different software versions in which the information used for particle identification has been defined differently.

Finally, also the samples that are used for studies of the performance of the selection in terms of the sensitivity to ϕ_s and the trigger rate can be found. These are the signal samples with SM values of $\Delta\Gamma$ and ϕ_s, a signal sample with an exaggerated value of ϕ_s, and a signal sample with settings $\Delta\Gamma = \phi_s = 0$. The ~ 9 M events in the minimum bias sample, used to estimate the expected trigger rate as realistically as possible, have been accepted by the L0 trigger and correspond to a realistic L0 output rate of 1 MHz.

4.3.2 Selection variables

The number of selection criteria is kept small in order to minimize the number of possible systematic effects. Furthermore, the variables are chosen to be minimally correlated, such that (efficiency) effects due to selection cuts can be studied separately, and calibration channels can be applied easier. The following selection variables, described below, are used: $p_T, \chi^2, \Delta M, \Delta \log \mathcal{L}_{\mathrm{K},\mu/\pi}$.

p_T: the transverse momenta of final state and of composite particles. Particles originating from decays of heavy particles in general have high transverse momentum. Hence, this is an effective observable to distinguish signal (originating from a B_s) from background originating from lighter quarks, in particular to remove the combinatoric prompt background originating from the primary vertex. A disadvantage of the use of p_T cuts (especially when p_T cuts are applied on individual final state particles, e.g. muons and/or kaons) is that these distort the observed angular distributions[1].

χ^2: the goodness of fit of composite particle reconstruction. The first one that is used is the χ^2 of the B_s vertex, which is a measure for the probability that the tracks assigned to the B_s originate from the same vertex. The second one that is used is the χ^2 of the B_s lifetime fit, which is a measure of consistency to the hypothesis that the given reconstructed B_s candidate originates from the assigned primary vertex. Both variables allow to suppress 'long-lived' backgrounds, since long-lived combinatoric background events in general do not originate from a vertex which points back to the interaction point. However, tracks in prompt background events do originate from a vertex which coincides with the interaction point, and hence prompt background events are not suppressed by cutting on these variables.

ΔM: the absolute mass difference of the reconstructed particle compared to its nominal mass in [6]. The absolute difference between the measured mass and the nominal mass is not used in the final event selection to discriminate between signal and background. Actually,

[1] For example: on average the relationship between helicity angle ψ and $p_T(\mathrm{K})$ is $p_T(K) = \text{offset} + \text{constant} \times \cos\psi$, as shown in chapter 6.

$|\Delta M(V)|$ of the vector mesons will be chosen wide enough to study the background and such that the distributions can be described. However, $|\Delta M(V)|$ criteria *are* used to define the signal and background samples in the optimization procedure . The mass range of the reconstructed B candidate is made very wide during the optimization (500 MeV/c^2) in order to increase the statistics.

$\Delta \log \mathcal{L}_{\mathrm{K},\mu/\pi}$, with $\Delta \log \mathcal{L}_{S/B} \equiv \log \frac{\mathrm{Prob}(S)}{\mathrm{Prob}(B)}$: the probability that a reconstructed particle is a signal particle or a mis-reconstructed background particle, for a given particle identity hypothesis. The delta-log likelihood is acquired by combining information from the trackers, calorimeters and the Cherenkov detectors. Since the definition used in the different versions of reconstruction software is different, the $\Delta \log \mathcal{L}$ cuts are not optimized, but chosen to be $\Delta \log \mathcal{L}_{\mathrm{K},\mu/\pi} > -5$. This value is chosen since it provides a high signal efficiency.

4.3.3 Optimization steps

Having described the set of selection variables, the process aimed to find the set of selection cuts leading to the best sensitivity to ϕ_s can be outlined.

According to (4.4), the precision to measure ϕ_s depends on the number of signal events, S, and the effective number of background events, B^*. In practice, B^* can be rewritten as

$$B^* \equiv \alpha B, \tag{4.6}$$

where B represents the number of background events in the 3σ mass window around the B_s, and α a weighting factor. It then follows that the precision of the ϕ_s measurement behaves as

$$\sigma(\phi_s) \propto \sqrt{S + \alpha B}/S. \tag{4.7}$$

Since a fit, e.g. a simultaneous fit to the invariant B mass distribution and the proper time distribution, is sensitive to differences between signal and background distributions in the observables used in this fit, the weight α depends on the distinguishing power of the fit between signal and background, e.g. the difference between the proper time distribution of signal and background after selection. Therefore, α is a measure of the power of the fit to distinguish signal and background using their individual distributions in all observables, and of the dilution of the sensitivity due to the presence of background events.

Since FOM $\propto 1/\sigma(\phi_s)$, the selection is optimized by maximizing the figure of merit (FOM)

$$\mathrm{FOM} \equiv S/\sqrt{S + \alpha B}. \tag{4.8}$$

As shown below, the weight α can be determined from fast parameterized (or 'toy') experiments[2], using the signal and background distributions as estimated from fully simulated events after selection. Optimizing the selection and determining the figure of merit, the two main ingredients in the procedure, is repeated iteratively, until the figure of merit converges.

Schematically, the procedure is as follows

[2]Alternatively, α can be determined from an extended fit to both signal and background after selection, fitting for the number of (signal) events. Using the relative uncertainty of the number of signal events the resulting FOM can be determined.

1. First the selection is optimized by maximizing FOM $\equiv S/\sqrt{(S+\alpha B)}$ with an arbitrary value of α. Adjusting the set of rectangular cuts in the optimization variables $(p_T, \chi^2, \Delta M)$, FOM is maximized for different values of τ_0.

2. From the ranges of maximal FOM, which are found by eye, an optimal selection is chosen. Ranges are looked for, since these guarantee stability of the performance of the selection. An example of signal and background distributions, and the resulting FOM distributions (for certain τ_0) is shown in figure 4.2.

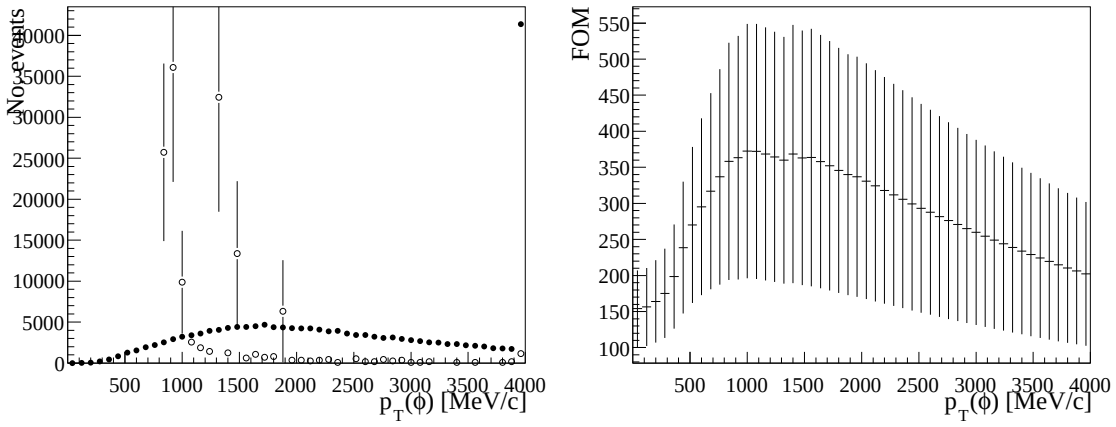


Figure 4.2: *Example of signal (closed dots) and background (open dots) (left) and FOM (right) distributions for the $p_T(\phi)$ observable. Here $\alpha = 0.4$, $\tau_0 = 0.2$ ps. It can be seen that two samples are used for the background. The prompt J/ψ event sample includes ϕ-mesons with lower p_T as compared to those in the $b\bar{b}$-dimuon sample. Also, due to the available statistics, the ϕ mesons in the dimuon sample have smaller weight. The reason that the FOM peaks as a function of this observable is due to the fact that the background behaves different from the signal. The maximum of FOM is reached at $p_T = 1500$ MeV/c, and the plateau of maximal FOM is estimated by eye to be 1000-1700 MeV/c. Since every FOM value is calculated using a single sample of signal and background events, the bin-to-bin errors of the FOM are correlated.*

3. The selection providing maximal FOM is used to determine the resulting background proper time distribution.

4. Subsequently, the background distribution determined after the selection in previous step is used to perform toy experiments. Varying S/B, the behavior of $\sigma(\phi_s) \propto \sqrt{S+\alpha B}/S$ is studied. In this way α can be determined, and hence a new definition of FOM.

5. The found value of α is used to repeat the procedure of steps 2-3-4: successively FOM is optimized, the optimal stable cuts are determined, and α is estimated. This procedure is repeated until the outcome stabilizes and an optimal selection is determined. The selection is chosen from the plateaus of the FOM, of which examples for different observables are shown in figure 4.3 and 4.4.

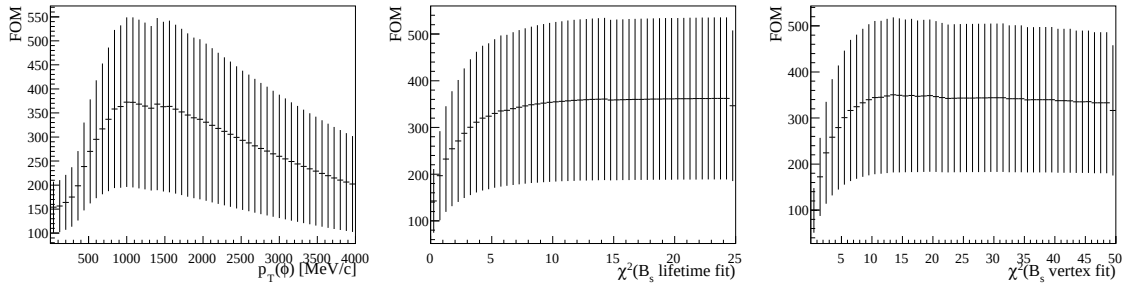


Figure 4.3: *Example of FOM distributions of the observables used on in the selection. Here* $\alpha = 0.4$, $\tau_0 = 0.2$ *ps. The plateaus of maximal FOM are* $p_T(\phi) \in (1000 - 1700)$ *MeV/c;* χ^2*(*B_s *lifetime fit)* ≥ 8 *;* χ^2*(*B_s *vertex fit)* ≥ 10*. Since every FOM value is calculated using shared signal and background events, the bin-to-bin errors of the FOM are correlated.*

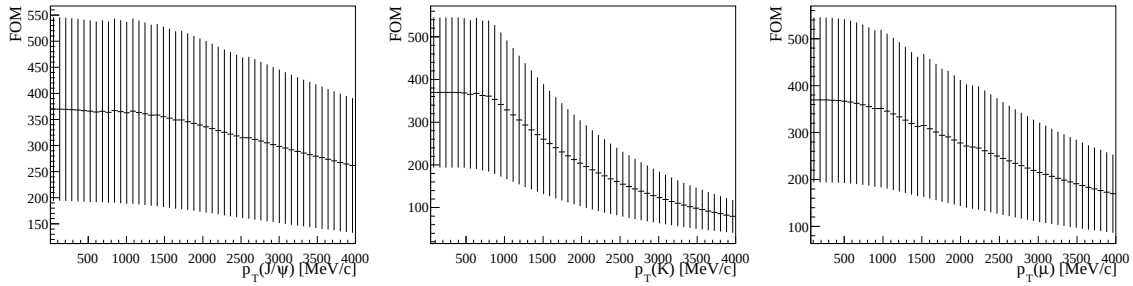


Figure 4.4: *Example of FOM distributions of the observables which are not used in the selection. Here* $\alpha = 0.4$, $\tau_0 = 0.2$ *ps. Since these observables in general behave similar to the signal, these variables are not appropriate to distinguish signal from background. The FOM are within fluctuations monotonically decreasing, which means the optima are consistent with applying no cut. Since every FOM value is calculated using shared signal and background events, the bin-to-bin errors of the FOM are correlated.*

6. Finally, when the selection as a function of the variables $(p_T, \chi^2, \Delta M, \Delta \log \mathcal{L}_{\mathrm{K},\mu/\pi})$ is estimated, the optimal τ_0 and prescale fraction for a given maximal selection rate are determined, by studying their combined effect on $\sigma(\phi_s)$.

This procedure provides a selection with maximum FOM, given the set of selection variables. The figure of merit used to optimize the selection, is found by studying $\sigma(\phi_s)$ and therefore the selection provides a maximal sensitivity to ϕ_s. Furthermore, since the selection is chosen from the plateau maxima of FOM, it is stable with respect to variations of the cuts and (small) variations in the signal and background distributions, as can occur between data and MC, or between different versions of MC. Summarizing, the procedure provides a stable selection with maximal sensitivity to ϕ_s, for the given set of cut observables.

4.4 Selection optimization

In this section the optimization is performed using the procedure described in the previous section. Using an arbitrary initial value for α, the selection is optimized, and the procedure of maximizing and determining FOM is iterated, until it converges. Finally, a set of optimal selection cuts is determined.

4.4.1 First iteration

Choosing as a starting value[3] $\alpha = 2/3$, the figure of merit (4.8) becomes

$$\text{FOM} = S/\sqrt{S + \frac{2}{3}B}. \tag{4.9}$$

Maximizing this to determine the selection criteria, the behavior of FOM as a function of τ_0 as shown in figure 4.5 is found. Since all measurements in the plot are retrieved by optimizing the same signal and background samples, the measurements are statistically correlated.

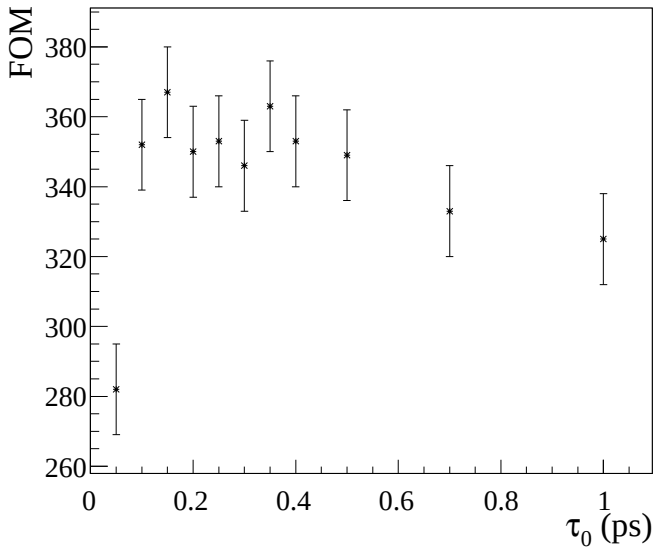


Figure 4.5: *The optimal values of FOM for different values of τ_0. The set of cuts differ per τ_0. The errors of the different measurements are correlated. A plateau in the range $(0.15, 0.50)$ ps can be seen.*

Due to the large amount of prompt background, FOM decreases rapidly[4] for $\tau_0 < 0.15$ ps. The maximal value of FOM is found for $\tau_0 = 0.15$ ps. In the region $\tau_0 \in (0.15, 0.5)$ ps a

[3] $\alpha = 2/3$ corresponds to the number of background events in the 2σ mass window.

[4] The fact that FOM decreases does not mean that the sensitivity decreases if τ_0 is lowered. It only means that the prompt background that appears in this region influences the sensitivity differently than long lived background. Since removing a long-lived background event is more useful than removing a prompt background event, the FOM can not be used to optimize the selection with in this region.

plateau can be seen, where FOM is constant within statistical fluctuations. The fluctuations in that region originate from the fact that the background sample has limited statistics. For values $\tau_0 > 0.5$ ps FOM decreases, since the lifetime cutoff removes predominantly signal events.

Hence the selection is chosen from the common maxima of the selection observables in the region of $\tau_0 \in (0.15, 0.5)$ ps. This is done by determining the plateau of maximal FOM for all different selection observables, as shown in figure 4.3 and 4.4. Then the selection criteria of the variables are chosen such that FOM is maximal for $\tau_0 \in (0.15, 0.50)$. The selection after this first iteration, with maximal FOM for the definition in (4.9), are shown in table 4.2.

$\chi^2(B_s$ vertex fit)	$\chi^2(B_s$ lifetime fit)	$p_T(\phi)$	$p_T(J/\psi, K, \mu)$
< 16	< 9	> 1100 MeV/c	$-$

Table 4.2: *Selection following from maximizing FOM, after the first optimization iteration ($\alpha = 2/3$). Since the values of $p_T(J/\psi,K,\mu)$ for maximal FOM are consistent with applying no cut, these cuts are avoided (or equal to zero). The mass windows of the mesons are set equal to $\Delta m(\phi) < 9$ MeV/c^2 and $\Delta m(J/\psi) < 34$ MeV/c^2 after the optimization.*

4.4.2 FOM after first iteration

As explained before, not only the number of background events, but also the (proper time) distribution of the background influences the sensitivity to ϕ_s. For example, the effect on $\sigma(\phi_s)$ from extra prompt background is in general small, whereas long living background decreases the sensitivity to ϕ_s more. To determine the influence of the background on the sensitivity, parameterized by α in (4.7), and thereby the correct figure of merit (4.8) for the optimization, the distribution of the background, after the selection found in section 4.4.1, needs to be determined.

The proper-time dependent background distribution is described by the PDF $B(t)$, as

$$B(t) = G(t, t'|\mu, \sigma_1, \sigma_2, f_{\sigma_1}) \otimes \Big(f_{\text{prompt}}\delta(t') + (1 - f_{\text{prompt}})B_{LL}(t') \Big). \tag{4.10}$$

The variable t' represents the actual true decay time of the events, while the variable t represents the experimentally observed decay time. The prompt background PDF, $\delta(t')$, and the long-lived background PDF $B_{LL}(t')$ are summed according to the prompt event fraction f_{prompt}. This sum is convolved with the resolution model $G(t, t'|\mu, \sigma_1, \sigma_2, f_{\sigma_1})$, representing two Gaussians with the same mean μ and different widths σ_1, σ_2, summed with a fraction f_{σ_1}. The long-lived background is described as

$$B_{LL}(t') = f_{\tau_1} B^1_{LL}(t'|\tau^1_{LL}) + (1 - f_{\tau_1}) B^2_{LL}(t'|\tau^2_{LL}), \tag{4.11}$$

with B^1_{LL} and B^2_{LL} two exponential distributions with lifetimes τ^1_{LL} and τ^2_{LL}, summed with a fraction f_{τ_1}. A fit to the selected background events in the whole t-region, using this description, can be found in figure 4.6.

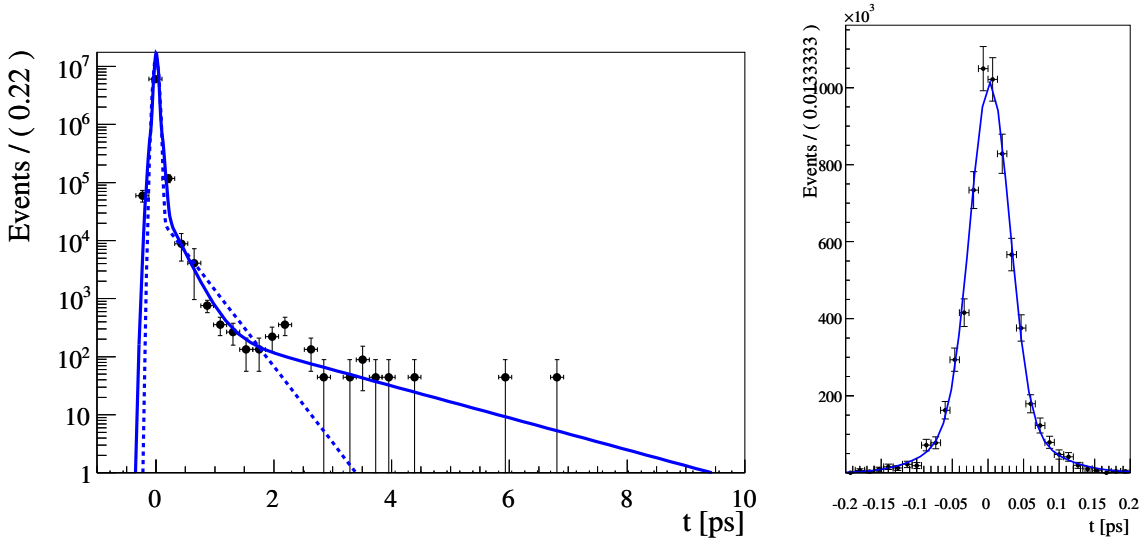


Figure 4.6: *Fit to the proper time distribution of the background after the optimal selection for $\alpha = 2/3$, the first optimization step. Events and errors are weighted to an equivalent of 2.0 fb^{-1} of data taking. The dashed curve is the PDF as used for the determination of α. The right plot is the same as the left plot, zoomed in around $t = 0$, to show the prompt background.*

To determine the weight α used in the optimization, toy experiments are performed, based on the expected proper-time and invariant-mass distributions of signal and background after the selection. Compared to (4.10), a simplified model is used, describing the long-lived background by an exponential distribution with a single lifetime τ_{LL}, and describing the resolution by a single gaussian with a width of σ_t. The settings used in the toy experiments are summarized in table 4.3. As is shown below, these simplifications do not change the estimate of the optimal selection.

In order to determine α, the statistical error of ϕ_s is determined as a function of B/S, using the background distribution after the selection. Setting $\tau_0 = 0.2$ ps and choosing the prescale factor equal to $f_{\rm pre} = 0.1\%$, the relationship between $\sigma(\phi_s)$ and B/S in figure 4.7 is found.

A χ^2 fit of $\sigma(\phi_s) = c\sqrt{S + \alpha B}/S$, with c a constant, returns $\alpha = 0.4 \pm 0.1$. Therefore, in the next iteration of optimization the number of background events in a 1.2σ B_s mass window, equivalent to $\alpha = 0.4$, is used for αB. Hence

$$\text{FOM} = \frac{S}{\sqrt{S + 0.4B}} \tag{4.12}$$

is the correct variable to use in the next step of the optimization procedure.

variable	toy value	fit result
τ_{LL} [ps]	0.33	0.33 ± 0.05
σ_t [fs]	39	-
μ [fs]	0	2.9 ± 7
B/S	47	47 ± 1
f_{prompt} [%]	99.3	99.3 ± 0.2
ϕ_s [rad]	-0.04	-

Table 4.3: *The parameters used as input for the toy experiments, and the results of the fit to the fully simulated background distribution. B/S is retrieved from a 3σ B_s mass window and the meson mass windows given in table 4.2. It should be noted that B/S, which consists mainly of prompt background, is very large, since the selection is optimized to minimize $\sigma(\phi_s)$, and not to suppress the large fraction of prompt background. Using a resolution consisting of two Gaussians to describe the background, the following parameters are found: $\sigma_1 = (28 \pm 1)$ fs, $\sigma_2 = (67 \pm 4)$ fs, $f_{\sigma_1} = (75 \pm 4)$ %. In the toy experiments ϕ_s is set to the SM value.*

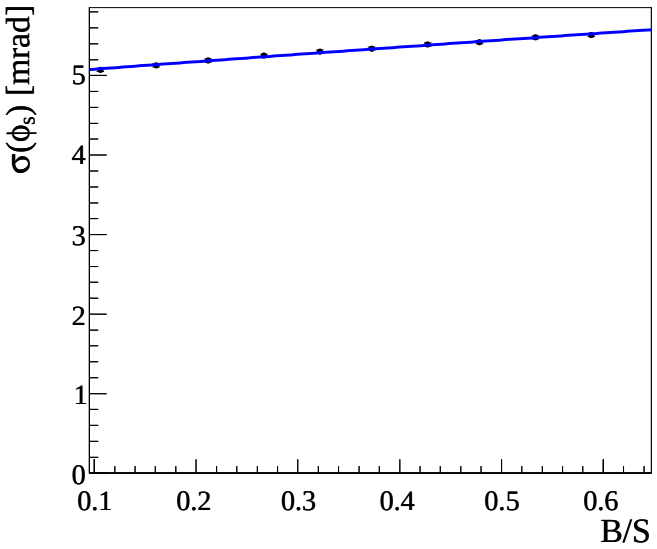


Figure 4.7: *A χ^2 fit to $\sigma(\phi_s) \propto \sqrt{S + \alpha B}/S$, assuming constant errors for $\sigma(\phi_s)$. The fraction background over signal B/S varies, the number of signal events S is fixed. B is the number of background events in a 3σ B mass window and for $t > \tau_0 = 0.2$ ps. The fit returns $\alpha = 0.4 \pm 0.1$.*

4.4.3 Optimal selection

Having determined that, after the first selection optimization, the FOM used in the following optimization should have $\alpha = 0.4$, the optimization of the selection by maximizing FOM is repeated. Again, for different τ_0 the maximal FOM are determined, which appears to be in the region where $\tau_0 \in (0.15 - 0.4)$ ps. The selections in this region, leading to a maximal FOM, can be found in table 4.4. The values of the variables are given for which FOM is maximal, the plateau ranges (given between brackets) are found by judging the histograms by eye.

τ_0 [ps]	χ^2_{vertex}	χ^2_{lifetime}	$p_T(\phi)$ [MeV/c]	$p_T(J/\psi)$ [MeV/c]	$p_T(K)$ [MeV/c]	$p_T(\mu)$ [MeV/c]
0.15	19 (> 10)	14.5 (> 8)	1440 ($1000 - 1600$)	85 ($-$)	120 (< 900)	150 (< 500)
0.20	20 (> 10)	14.5 (> 8)	1440 ($900 - 1600$)	85 ($-$)	120 (< 900)	150 (< 500)
0.25	20 (> 10)	14.5 (> 8)	1120 ($1000 - 1500$)	85 ($-$)	120 (< 600)	225 (< 500)
0.30	20 (> 10)	10.5 (> 8)	1040 ($800 - 1400$)	85 ($-$)	440 (< 500)	150 (< 500)
0.35	20 (> 10)	16 (> 8)	1040 ($800 - 1300$)	85 ($-$)	120 (< 600)	225 (< 500)
0.40	20 (> 10)	16 (> 8)	1040 ($800 - 1300$)	85 ($-$)	120 (< 800)	225 (< 500)

Table 4.4: *Optimal cuts for the second iteration ($\alpha = 0.4$) of optimizations around the maxima. Between brackets the plateau if any. The plateaus for the absolute mass difference of the J/ψ and the ϕ are > 30 MeV/c^2 and > 6.5 MeV/c^2 respectively. The plateaus of FOM of all obervables overlap, which means a stable selection can be chosen.*

From the table two things can be noted. Firstly, the FOM of all observables have plateaus which overlap, implying that an optimal selection that is stable under changing τ_0 can be determined. Secondly, three out of four p_T cuts can be avoided when optimizing the sensitivity for ϕ_s, since in the whole relevant τ_0 range the plateaus contain $p_T(J/\psi, K, \mu) = 0$. Since angular distributions in general are disturbed by p_T cuts, especially when applied on the individual daughter particles, avoiding these p_T cuts means that the number of possible systematic effects in the distortion of the angular distribution are minimized.

Combining the plateaus of the different selections, the estimated selection in table 4.5 is chosen. Since the optimal selection criteria for $\alpha = 0.4$ are the same as found in the first iteration where $\alpha = 2/3$, the background distribution after this selection is the same for both $\alpha = 0.4$ and $\alpha = 2/3$, and the iteration procedure has converged. This means the FOM in (4.12), with $\alpha = 0.4$, is the correct figure of merit for the optimization. Furthermore, the optimal selection is not only stable for uncertainties of the background distribution in the selection observables, but also in uncertainties of the background proper time distribution.

Finally, the mass windows for the composite particles in the final selection are chosen as in table 4.5. These are wide enough in order to study the background distribution in the sidebands. For the ϕ particle there is no lower mass bound, and the upper limit is high enough to study the non-resonant (or S-wave) background.

variable	cut
$\chi^2(B_s$ vertex fit)	< 16
$\chi^2(B_s$ lifetime fit)	< 9
$p_T(\phi)$	> 1100 MeV/c
$\min(\Delta \log \mathcal{L}_{K/\pi})$	> -5
$\min(\Delta \log \mathcal{L}_{\mu/\pi})$	> -5
$\Delta m(B_s)$	140 MeV/c^2 (7σ)
$\Delta m(J/\psi)$	< 84 MeV/c^2 (7σ)
$m(\phi)$	< 1050 MeV/c^2

Table 4.5: *The final selection after optimization. The first three cuts follow from the optimization procedure, the min($\Delta \log \mathcal{L}$) cuts are chosen conservatively. The mass windows are chosen wide enough to be able to study the background.*

4.5 Optimization of the lifetime cut and the prescale fraction

In the previous section the selection has been optimized. According to table 4.3 it follows that for this selection $B/S = 47$. As explained above, this large fraction of background is due to the fact that the selection has been optimized to have a maximal the sensitivity to ϕ_s, and not to suppress the prompt background (contrary to the selection in [44]). For an expected annual signal yield of ~ 200 k signal events, it follows that a total yield of ~ 10 M is expected to be selected by the selection without lifetime cut, equivalent to an offline rate of $\sim$1 Hz, in narrow mass windows. In conclusion, the selection without any lifetime cut criterium has a too large output rate.

Since prompt background is the dominating contribution to the rate, a suppression can be achieved by dividing the selected sample in two, as explained in the introduction of this chapter. The sample with no lifetime cut, the 'control' sample, dominated by prompt background, is therefore prescaled by a factor, which means that only a fraction of the events passing the selection are accepted. The events in the detached sample, the sample with lifetime cut, are all selected.

In this section the optimal lifetime cut criterium and prescale fraction are determined. For this purpose toy MC experiments are performed, using a number of events corresponding to approximately a year of nominal data taking. An example of such a toy experiment is shown in figure 4.8. In these toy experiments the proper time resolution is determined from the fit to the proper time distribution of the prompt background events, a crucial aspect of the method.

First, the effect of the lifetime cut (removing events at low t) on the sensitivity to ϕ_s is determined. Secondly, the effect of the prescale fraction (resulting in less prompt background events to determine the proper time resolution parameters) on the sensitivity to ϕ_s is determined. Finally, the optimal combination of the lifetime criterium and the prescale factor, given a maximally allowed rate, is determined.

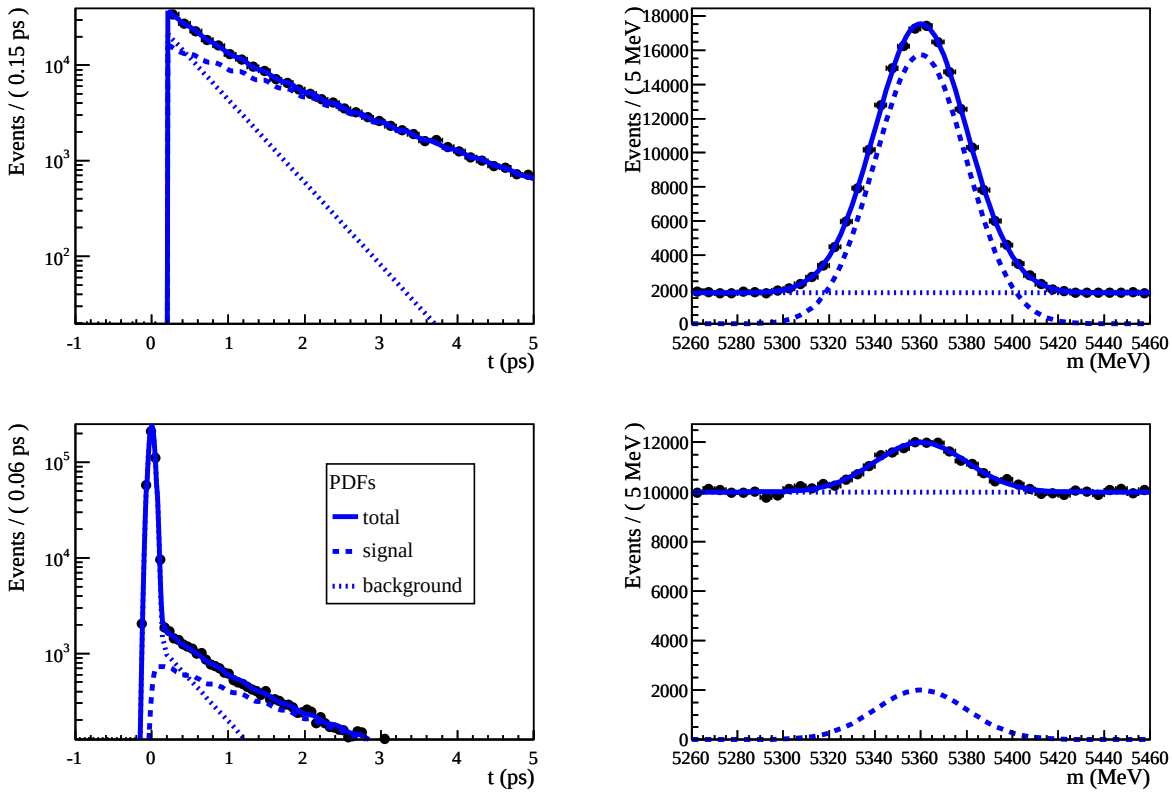


Figure 4.8: *Example of one toy experiment corresponding to approximately one year of data taking at* $\mathcal{L} = 2 \times 10^{32}$. *On the left the reconstructed proper time distributions are shown, on the right the invariant mass distributions. At the bottom the prescaled sample is shown, at the top the detached sample. The total PDF and the signal and background components after a fit are shown. The lifetime cut is* $\tau_0 = 0.2$ *ps. Clearly the cutoff caused by the lifetime cut* $\tau_0 = 0.2$ *ps can be seen. The prescale fraction is* 10%. *The effect of the prescaling is clear from the relatively small number of events per bin of* t *in the prescaled sample, at proper times larger than the cutoff value, and from the smaller number of signal events in the invariant mass peak. From the invariant mass distributions it can be seen that the prescaled sample is dominated by background, whereas the detached sample consists almost purely of signal events. The purpose of the two samples is clear from this plot: the detached selection contributes the biggest part to the sensitivity of the physics parameters, whereas the prescaled sample is used to calibrate the proper time resolution and acceptance.*

4.5.1 Effect of the lifetime cut on $\sigma(\phi_s)$

The effect of varying the lifetime cut on the sensitivity $\sigma(\phi_s)$ to ϕ_s is studied, assuming infinite bandwidth. The lifetime cut applied in the selection is motivated by the fact that the events at low t are predominantly prompt background events. Therefore, events in this region are expected to contribute little to the sensitivity to ϕ_s. Removing these events would therefore decrease the sensitivity little, while the event rate would be reduced significantly.

In figure 4.9 the sensitivity for ϕ_s is plotted for different values of the cut-off τ_0, while the prescale fraction of the prescaled sample, from which the resolution is determined, is kept at a constant value of 1%. It can be seen that the uncertainty of ϕ_s decreases as the lifetime cut decreases. However, for small lifetime cuts the increase in sensitivity comes to a halt. This is due to the large amount of prompt background around $t = 0$, due to which the signal events in this region contribute hardly to the sensitivity to ϕ_s.

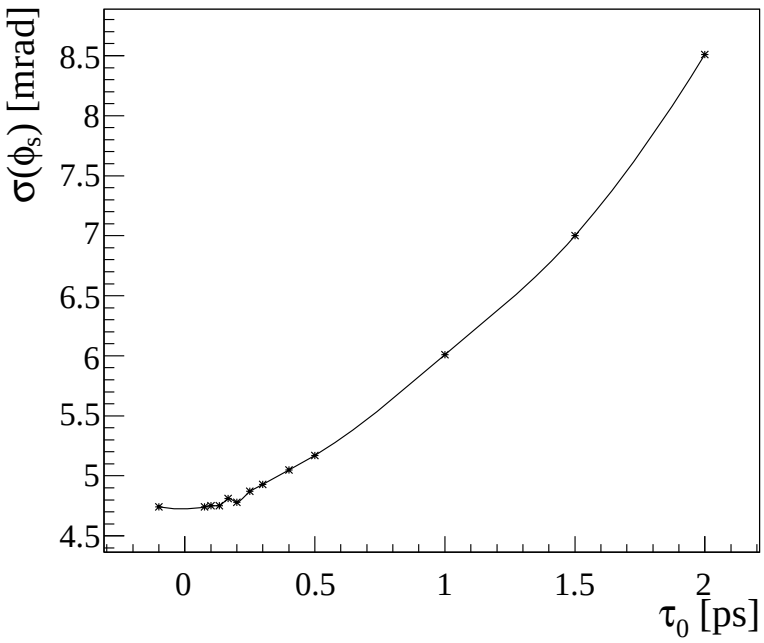


Figure 4.9: *The uncertainty of ϕ_s for different values of the cutoff lifetime τ_0. The prescale rate is constant and equal to 1%. Since no constraint is applied on the total bandwidth, the curve is monotomically increasing. It should be noted that the axis of $\sigma(\phi_s)$ is zero-suppressed.*

It can be concluded that a lifetime cut in the detached sample $\tau_0 < 0.20$ ps hardly decreases the sensitivity to ϕ_s. This explains why a detached selection can still lead to a maximal sensitivity to ϕ_s: the signal events around $t = 0$ hardly contribute to the sensitivity, since the largest fraction of the events is prompt background.

4.5.2 Effect of the prescale fraction on $\sigma(\phi_s)$

Having determined the effect of varying the lifetime cut, the effect of varying the prescale fraction of the control sample is studied, assuming again infinite bandwidth. Since not all prompt background events are required to determine the proper-time resolution with a sensitivity sufficient to determine ϕ_s, a fraction of the sample can be ignored. The clear advantage of this prescaling is that this suppresses the rate.

In figure 4.10 the precision $\sigma(\phi_s)$ is shown for different prescale rates. No clear dependency is observed. The distribution supports the hypothesis that ϕ_s is insensitive to a prescale factor larger than 0.1%.

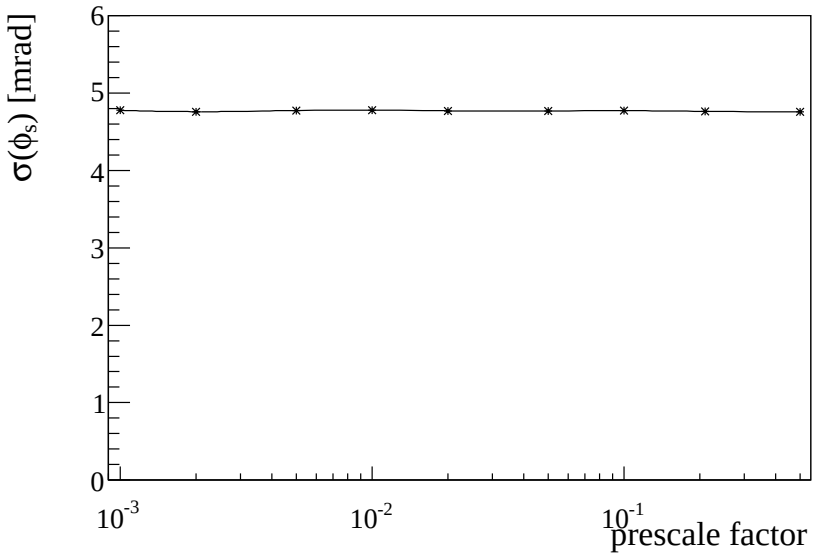


Figure 4.10: *The precision of ϕ_s for constant $\tau_0 = 0.2$ ps as a function of different prescale factors. The distribution supports the hypothesis that the sensitivity for ϕ_s is constant for prescale rates larger than 0.1%. It should be noted that in the study with a prescale factor of* 0.1%*, still ∼10 k events are in the prompt, prescaled background.*

From this study it can be concluded that for a lifetime cut of $\tau_0 = 0.2$ ps, the sensitivity to ϕ_s is independent of a prescale factor larger than 0.1%. This is due to the fact that for 2.0 fb^{-1} of data, there are always sufficient events in the prompt background peak to determine the resolution precisely enough not to dilute the sensitivity to ϕ_s. Therefore, a prescaled selection can be used in combination with a detached sample, without losing sensitivity to ϕ_s; the events at low proper time t that are not included in the detached sample do not contribute significantly to the signal sensitivity, whereas the events that are not included in the prescaled sample only influence the accuracy of the calibration but do not alter the sensitivity to ϕ_s.

It should be noted that for smaller periods of data taking, prescale factors this small might not suffice. A solution to this problem is proposed later, when the operational aspects are discussed.

4.5.3 Effect of τ_0 and $f_{\mathbf{pre}}$ on the event rate and $\sigma(\phi_s)$

Given the constraint of a maximal event selection rate, the optimal lifetime cutoff for the detached sample and prescale rate for the control sample can be determined.

From the results in sections 4.5.1 and 4.5.2 it can be concluded that no sensitivity is lost for a prescale factor as small as 0.1% and a lifetime criterium $\tau_0 < 0.2$ ps. To suppress the rate, while keeping optimal sensitivity to ϕ_s, it is most advantageous to prescale the events at small t by a big factor, while limiting the lifetime cut to 0.20 ps.

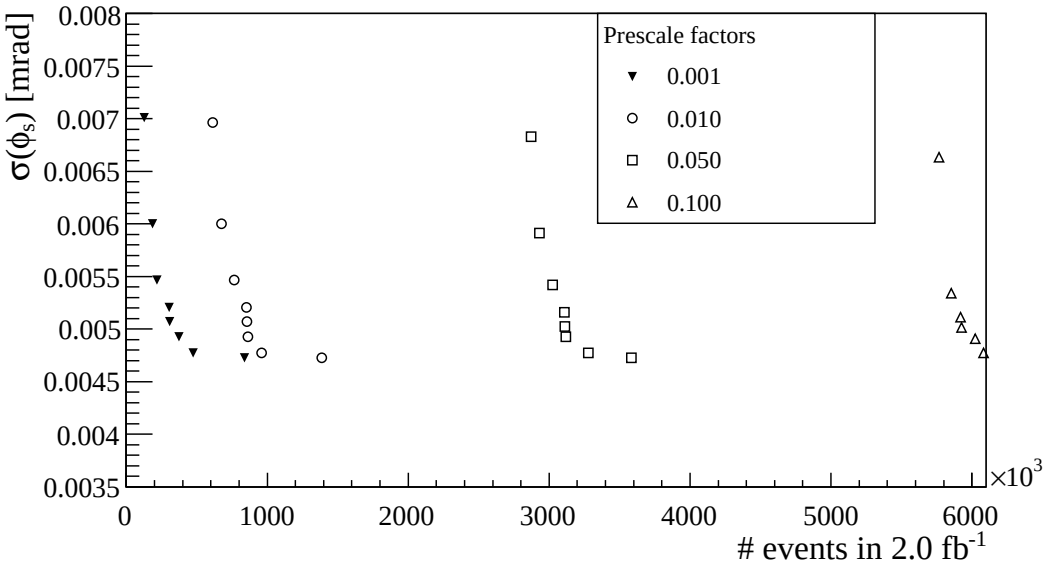


Figure 4.11: *The sensitivity for ϕ_s vs the total yield for different values of the prescale factors, for 2.0 fb^{-1}. As the cutoff τ_0 decreases in the range (1.5, 1.0, 0.7, 0.5, 0.4, 0.3, 0.2, 0.1) ps, the rate increases, just like the sensitivity to $\sigma(\phi_s)$ improves.*

This can be seen in figure 4.11, in which the uncertainty of ϕ_s has been plotted against the annual yield for different combinations of τ_0 and f_{pre}. The different markers in the graph denote prescale factors ranging from 0.1% to 10%. The lifetime cuts range from 0.10 to 0.50 ps. An asymptote can be clearly seen: lowering the cutoff below a certain value ($\tau_0 < 0.2$ ps) only increases the rate, but does not improve the sensitivity. Adjusting the prescale factor does not alter the sensitivity, but lowering the prescale factor below 1% hardly decreases the rate, since for that region the rate is dominated by detached events.

It can therefore be concluded that a combination of a prescaled selection and a detached selection leads to an optimal selection, while the event rate is kept limited. For the presented selection, it appears that $\tau_0 < 0.2$ ps suffices, whereas any value of $f_{\text{pre}} > 0.1\%$ can be used for a nominal year of data.

4.6 Efficiencies, yields, purities, and rates of the selections

Having optimized both the selection criteria as well as the prescale factor of the unbiased event sample, quantitative statements about the performance of the selection algorithm can be made. To compare the effects of the selection criteria, the efficiency for each cut is studied. In order to estimate the annual yield for the signal sample and a statistical error of the measurement of ϕ_s, the efficiency and purity of the event sample is estimated. To estimate the performance of the trigger, the trigger rate is determined. Furthermore, to allow optimization of the trigger algorithm the trigger efficiency is factorized, normalized to the offline selection efficiency. Finally, for the performance of the selection also the efficiencies as function of the angular observables and the proper time, e.g. due to the τ_0 cut in the selection, need to be determined.

4.6.1 Signal efficiency per cut

The efficiency of each selection criterium, as obtained in the optimization procedure, is shown in table 4.6. Of the first three cuts, the $p_T(\phi)$ cut appears to be least efficient, consistent with the observation in the optimization procedure that it discriminates between signal and background best. The $\Delta \log \mathcal{L}$ cuts have not been optimized. The $\Delta \log \mathcal{L}_{K/\pi}$ cut appears to be very efficient for signal events, whereas 13.6% of the events do not survive the $\Delta \log \mathcal{L}_{\mu/\pi}$cut. The selection cuts on the PID information requires further study, after which they can be included in the optimization procedure. As an illustration, also the efficiency of the lifetime cut is indicated. The value of this cut can be adjusted to adapt to the allowed event rate.

variable	cut	efficiency
$p_T(\phi)$	> 1100 MeV/c	87.3%
$\chi^2(B_s$ vertex fit)	< 16	96.0%
$\chi^2(B_s$ lifetime fit)	< 9	90.5%
$\min(\Delta \log \mathcal{L}_{K/\pi})$	> -5	99.5%
$\min(\Delta \log \mathcal{L}_{\mu/\pi})$	> -5	86.4%
t	> 0.2 ps	86.4%

Table 4.6: *Efficiency for each cut in the final selection. The numbers are retrieved from cuts on associated signal events. The efficiency of every cut is defined with respect to the previous cut. The shown lifetime cut is adjustable.*

4.6.2 Annual yield, purity and rate

To determine the annual signal yield, the selection is performed on ∼10 k simulated signal events produced at the interaction point of LHCb. A fit is performed to the reconstructed invariant mass distribution, which can be found in figure 4.12. The signal distribution is

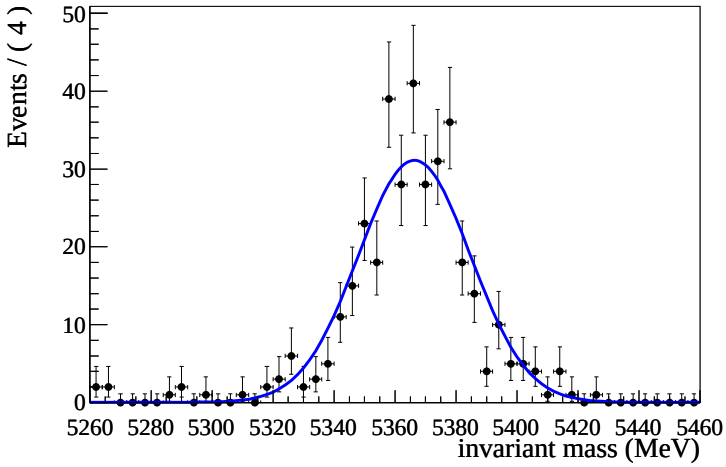


Figure 4.12: *Fit to reconstructed and selected events from a* ~ 10 *k signal sample. The signal is described by a single Gaussian; the background is described by an exponential.*

described by a single Gaussian, background by an exponential curve. Scaling the found number of signal events to a nominal year, the result corresponds to an expected annual signal yield of (207 ± 10) k events, when no lifetime cut is applied.

The expected annual yields and purities are summarized in table 4.7. In the control sample the background over signal ratio for long lived background equals 0.33 ± 0.09. In the detached sample, where a lifetime cut $t > \tau_0$ is applied, this number is smaller. Finally, it should be noted that the dominating background is prompt when no lifetime cut is applied, but that this background vanishes in the detached sample.

The expected event rate of the selection is determined by running the offline selection over L0 accepted minimum bias events. This sample is chosen, since it is the most complete simulation of an expected LHCb data sample. Since the corresponding running time is small, the uncertainty on the expected rate is relatively large. After performing selection cuts on L0 accepted minimum bias events, the event rates shown in table 4.7 are found. If no lifetime cut is applied, the rate is clearly too large. However, the rate of the detached sample is acceptable, depending on the chosen lifetime cut. The rate of the sample without lifetime cut should be prescaled, as shown in previous section, such that it also becomes acceptable.

As an example, including a lifetime cut of $\tau_0 = 0.2$ ps in the selection of the detached sample, the rate is[5] (0.05 ± 0.05) Hz for the detached selection. The rate of the prescaled selection is $f_{\text{pre}}(5.4 \pm 0.5)$ Hz, where f_{pre} is the prescale fraction. The total rate in that case is hence

$$(0.05 \pm 0.05)\text{Hz} + f_{\text{pre}}(5.4 \pm 0.5)\text{Hz}. \tag{4.13}$$

For a prescale factor of $f_{\text{pre}} = 1\%$ this would mean a total, acceptable, rate of (0.10 ± 0.05) Hz, or about 1M events per year.

Finally, the three long-lived surviving background events from the sample of L0-accepted

[5]The uncertainty on the rate is estimated using a $\sqrt{N}$ error for the number of selected events N.

Sample	τ_0 cutoff	S	B_{LL}/S	B_{prompt}/S	rate [Hz]
Control	no τ_0	$f_{\text{pre}} \times (207 \pm 10)$ k	0.33 ± 0.09	47	$f_{\text{pre}} \times (5.4 \pm 0.5)$
Detached	$\tau_0 = 0.20$ ps	(179 ± 10) k	0.21 ± 0.07	0	(0.05 ± 0.05)
	$\tau_0 = 0.15$ ps	-	-	-	(0.10 ± 0.07)
	$\tau_0 = 0.10$ ps	-	-	-	(0.15 ± 0.09)

Table 4.7: *Final yields and purities after the selection. The background over signal fractions B/S are defined in a 3σ B mass window and in the 'optimal' mass ranges of the J/ψ and the ϕ, as given in table 4.2. Furthermore, the final rate after L0 and offline selection are given. For the control sample the yield and the rate depend on the prescale factor f_{pre}*

minimum-bias events all show the same origin: the selected kaons are misreconstructed pions, whereas the muons originate from pions which decay in flight, in the magnet region.

4.6.3 Trigger efficiencies

So far only an offline selection algorithm has been considered. However, before the offline selection is performed, events should have been selected by the trigger first. To determine whether the trigger would have kept these offline selected events, the trigger efficiency for this offline event sample is determined. The HLT selections shown are tuned such that their efficiency for these events is as large as possible.

L0 and HLT1

First, the efficiencies of the L0 and HLT1 trigger lines are determined. The following HLT1 lines appear to have efficiencies above 10%. All other HLT1 lines have efficiencies $< 10\%$ and are neglected.

1. Hlt1DiMuonNoIPL0Di: events triggered by the L0 di-muon trigger and confirmed as di-muon pair.

2. Hlt1DiMuonNoIP2L0: events triggered by the L0 single muon trigger, with a second muon candidate in the event, together confirmed as di-muon pair.

3. Hlt1DiMuonNoIPL0Seg: events triggered by the L0 single muon trigger, with another track segment in the event, together confirmed as di-muon pair.

4. Hlt1MuTrack: triggered by the L0 single muon trigger, with another track in the event. A non-trivial proper time efficiency inducing impact parameter (IP) criterium is applied on the muon, and the muon and the track are required to form a vertex .

5. Hlt1SingleMuonIPCL0: events triggered by the L0 single muon trigger, with the muon candidate confirmed as a muon. An IP cut is applied in the selection of the muon candidate.

HLT1 line	ε(line)	ε(uniform and non-uniform)	ε
Hlt1DiMuonNoIPL0Di	40%	75%	85%
Hlt1DiMuonNoIP2L0	49%		
Hlt1DiMuonNoIPL0Seg	55%		
Hlt1MuTrack	61%	72%	
Hlt1SingleMuonIPCL0	64%		

Table 4.8: *The efficiencies ε of the most efficient HLT1 lines for events selected by the $B_s^0 \to J/\psi\phi$ offline selection. Shown are the efficiencies per line, the efficiency of the (non-)uniform lines and the efficiency of all lines. The events in the Hlt1DiMuonNoIPL0DiDecision line originate from the DiMuon L0 trigger. All other lines get their input from the L0 Muon trigger. All given efficiencies are including L0 efficiency.*

It is not only important whether these lines are efficient, but also whether cuts are included which induce non-uniform proper time efficiencies, as this needs to be taken into account during the analysis. In the lines 1-3 no cuts are applied that induce non-uniform proper time efficiencies. To suppress the trigger rate in these lines a dimuon invariant mass cut (>2.5 GeV) is applied, which does not affect the events selected by the offline selection. In the lines 4-5 cuts are applied that induce non-uniform proper time efficiencies, in particular IP cuts.

In table 4.8 the efficiencies of these HLT1 lines, with respect to the offline selected events, can be found. As can be seen, the inclusion of lines with cuts that induce non-uniform proper time efficiencies leads to a relative increase in statistics of 13%.

It should be noted that the increase of events in the 'sensitive region' is larger, hence the sensitivity to ϕ_s will be increased more than a factor $\sqrt{1.13}$. This can be understood from figure 4.13. The plateau of the proper time efficiency approaches 92%. Since the selected event sample at low t consists of mainly prompt background events, the signal events at low t, as shown before, hardly contribute to the sensitivity to ϕ_s.

It can be concluded that the efficiency relevant for the sensitivity to ϕ_s for the inclusion of all muon HLT1 lines is 92%. If only the HLT1 lines with a uniform proper time efficiency are included, this efficiency drops to 75%. This profit in the trigger of 23% in effective signal statistics is due to the fact that the selected sample is divided in two, and samples from HLT1 lines that induce non-constant efficiencies can be included. The possible inclusion of extra HLT1 lines also improves the robustness of the selection.

HLT2

After being triggered by L0 and confirmed by HLT1, the event rate needs to be suppressed further in HLT2. Since in HLT2 exclusive event reconstruction can be performed, this selection can be performed in a similar way as the offline selection.

The selection cuts in the exclusive HLT2 lines are shown in table 4.9. These selection cuts are the same for both the prescaled and the detached line. In the prescaled line the prescale factor is applied, whereas in the detached line instead the lifetime cut is applied. Both these variables can be changed to adjust to the rate, as explained in previous section.

variable	HLT2 cut
$p_T(\phi)$	> 500 MeV/c
$\chi^2(B_s$ vertex fit)	< 100
$\chi^2(B_s$ lifetime fit)	< 36
$\lvert\Delta M(B)\rvert$	< 300 MeV/c^2
$\lvert\Delta M(J/\psi)\rvert$	< 100 MeV/c^2
$M(\phi)$	< 1100 MeV/c^2

Table 4.9: *Selection cuts in the exclusive HLT2 lines for the prescaled and detached selections.*

It should be noted that the HLT2 cuts are loosened with respect to the offline selection, since the offline and the online reconstruction differ. The cuts are chosen such that the HLT2 selection is maximally efficient with respect to the offline selected sample.

As can be seen from figure 4.14, the plateau efficiency of the exclusive, detached $B_s^0 \to J/\psi\phi$ HLT2 selection is $> 90\%$. Again, the plateau efficiency is the important quantity, since the events at low t hardly contribute to the sensitivity to ϕ_s. The main inefficiency is due to the fact that for an exclusive selection all four signal tracks are needed, which, due to online tracking inefficiencies, are not always reconstructed.

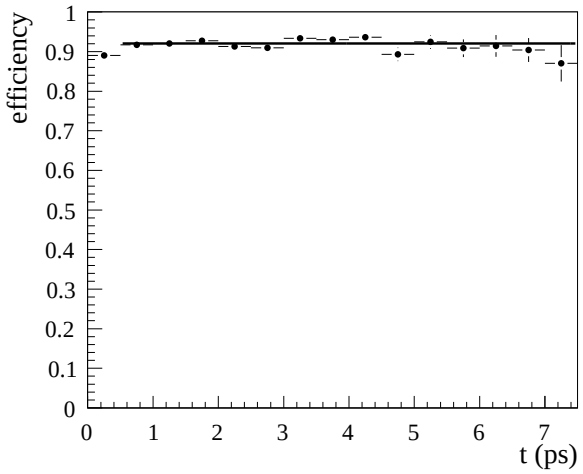


Figure 4.13: *HLT1 selection efficiency as a function of the proper time, with respect to offline selected events, for all muon lines. The plateau efficiency is* 92%*, and is fit for in the region* $t > 0.5$ *ps. The selection efficiency of the HLT1 di-muon lines is* 75% *and constant as a function of proper time.*

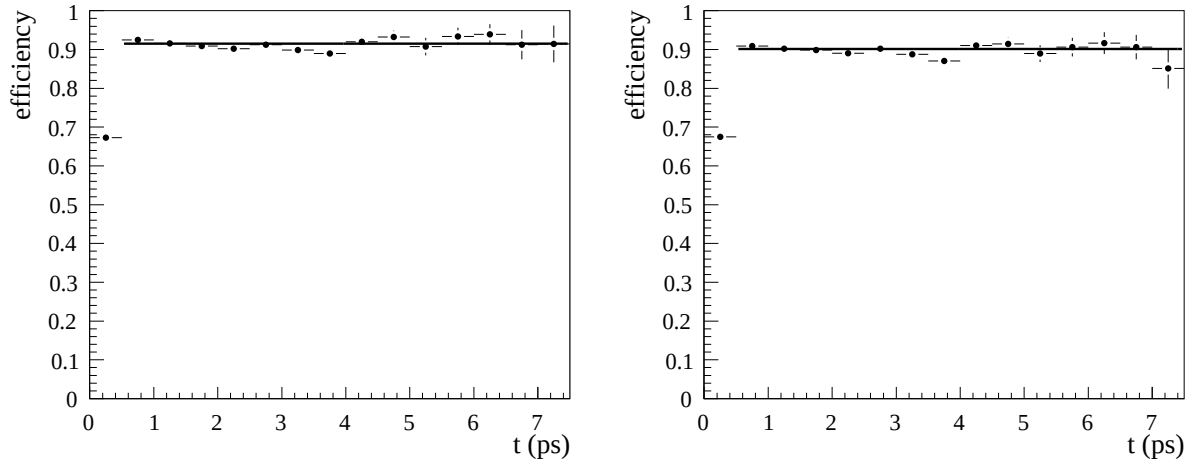


Figure 4.14: *Efficiency as a function of the proper time for the detached* $B_s^0 \to J/\psi\phi$ *HLT2 selection. The efficiency of the exclusive detached events with respect to events coming from the Hlt1DiMu lines as a function of proper time (left), the efficiency of the events from the Hlt1Mu lines (right). The former plateau is* 92%*, the latter plateau is* 90%*. The fit is performed in the region* $t > 0.5$ *ps. The inefficiency at low* t *is due to the lifetime cutoff in the HLT2 selction.*

4.6.4 Efficiencies as function of the angular and proper time observables

Finally, the efficiencies as a function of the angular and proper time observables, induced by both the offline and online selections, are presented. For this study the $\Delta\Gamma = 0$ sample has been used, for which the difference of the B-heavy decay width (Γ_H) and the B-light decay width (Γ_L) is equal to zero. This sample is most suitable to study angular and proper time acceptances, since for $\Delta\Gamma = 0$, the proper time and angular distributions factorize, such that proper-time and angular efficiencies do not influence each other.

The efficiencies are determined by dividing the distributions of the true MC observables of reconstructed events by the theoretical distribution. The overall scale of the efficiencies is governed by the number of events produced at the LHCb interaction point and the number of reconstructed events, and is a product of the reconstruction efficiency and the selection efficiency.

The efficiencies as projected on the angular observables $\cos\theta$ and $\cos\psi$, as defined in chapter 2, after the offline selection, are shown in figure 4.15. Both efficiencies are lowest for the angular observables $|\cos\theta|, |\cos\psi| \to 1$. In figure 4.16 the efficiency of decay angle ϕ and proper time t are shown. The behavior of the angular efficiencies is qualitatively in agreement with the efficiencies as given in [20]. The effects are due to (indirect) p_T cuts, e.g. from the detector acceptance and tracking efficiencies. It should be noted that the angular efficiencies are partly due to 'apparent' efficiencies, caused by acceptance effects in other,d correlated, angular observables.

Table 4.10 shows the relative deviations of the different angular efficiencies and the goodness of fit of the hypothesis of a constant efficiency. The angular efficiencies are non-uniform and vary within $\sim 15\%$, whereas the proper time efficiency is consistent with being constant, which can also be seen from figure 4.16.

variable	Δ_{rel}	χ^2/N_{DoF}
$\cos\theta$	13%	46/19
$\cos\psi$	15%	52/19
ϕ	15%	48/24
t	-	17/19

Table 4.10: *Efficiencies of the different observables. The relative difference between the peaks is defined as* $\Delta_{rel} \equiv \frac{max-min}{max}$. *Only the* χ^2/N_{DoF}, *the* χ^2 *per number of degrees of freedom, of the constant line fit to the proper time efficiency supports the hypothesis that the proper time efficiency is constant. All other observables have non-trivial efficiencies.*

Finally, the efficiency as a function of the proper time as induced by the lifetime cut in the exclusive detached HLT2 selection is shown in figure 4.17. This efficiency is defined with respect to the events in the (prescaled) control sample: in each bin the number of events selected by both the prescaled and the detached trigger are divided by the number of events selected by the prescaled trigger. Since a lifetime cutoff is used in the detached trigger, the jump around $t = 0.15$ ps is caused by the explicit lifetime cut. The fact that the jump is not exactly a step function, but more smoothly, is due to differences between

offline and online reconstruction.

When the effects of above efficiencies appear to be significant, the efficiencies need to be included in the analysis. Methods to do so are discussed in next chapter.

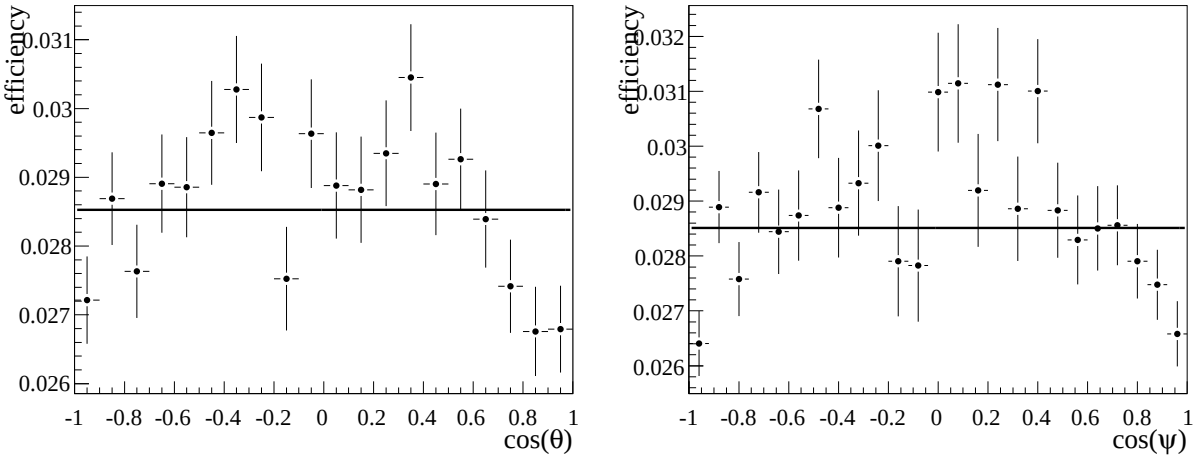


Figure 4.15: *Efficiency as a function of the angular observables* $\cos\theta$ *(left) and* $\cos\psi$ *(right) due to the offline selection. The true MC values of the reconstructed events are normalized to the theoretical distribution and the number of events produced at the interaction point of LHCb. Superimposed is the result of a fit of a constant efficiency.*

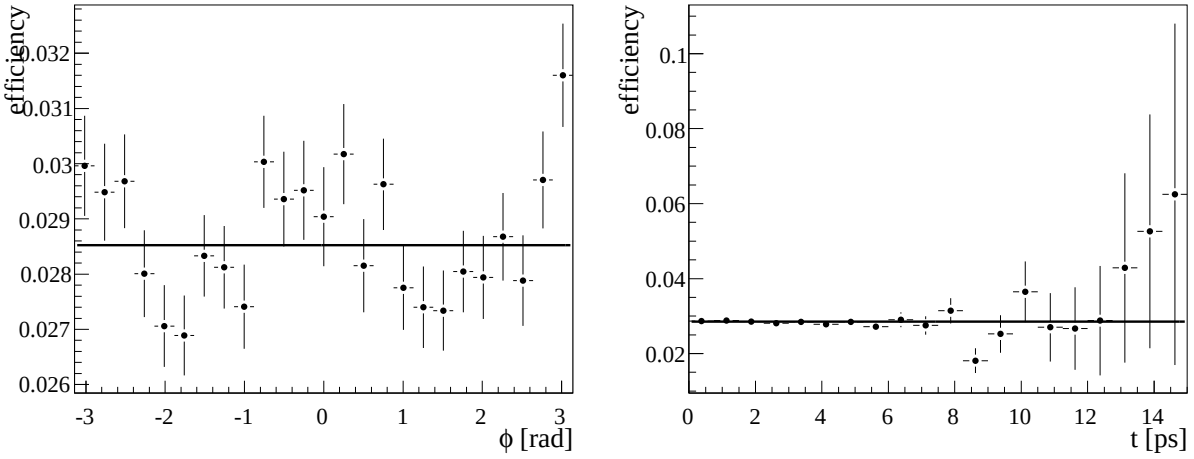


Figure 4.16: *Efficiency as a function of the angle* ϕ *(left) and the proper time* t *(right) due to the offline selection. The true MC values of the reconstructed events are normalized to the theoretical distribution and the number of events produced at the interaction point of LHCb. Superimposed is the result of a fit of a constant efficiency.*

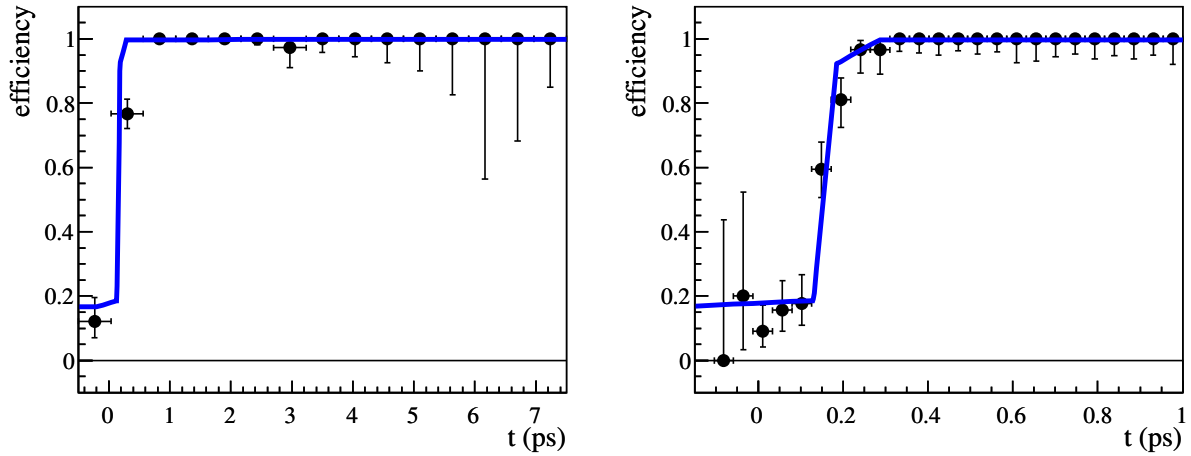


Figure 4.17: *Non-constant efficiency as a function of the proper time, due to the lifetime cut in the exclusive HLT2 selection. For each bin the number of events in both the prescaled and detached sample are divided by the number of events in the prescaled sample. The prescaled sample is required to be constructed by events selected by the di-muon HLT1 lines and the prescaled HLT2 trigger, which do not induce a non-constant efficiency as a function of the proper time. In the detached sample used in this figure, events selected by the detached HLT2 trigger are used, and only events selected by HLT1 di-muon lines are considered. The fit method, simultaneously fitting for the physics parameters and the efficiency on data, is explained in next chapter.*

4.7 Operational guidelines

In previous section the efficiencies of different trigger lines have been presented. These provide a procedure that can be followed during operation, and that is described in this section. Furthermore, possible procedures for the case the experiment appears to behave differently than expected from the simulation, are proposed.

First, the prescaling procedure is discussed. Prescaling, which in general happens in HLT2, is done to satisfy operational constraints on the trigger, i.e. bandwidth limits for storing data to disk and CPU usage by the HLT. In the previous section the HLT2 line that selects a prescaled sample of $B_s^0 \to J/\psi\phi$ decays is described. Since this trigger line is required to select a sample with a uniform proper time efficiency, events in this prescaled sample must also be selected by the HLT1 lines in which no non-uniform proper time efficiency inducing criteria are applied. The reduction of events, either due to prescaling, or due to restricting to specific HLT1 lines, or due to possible unexpected inefficiencies in the exclusive selection, as has been shown before, do not affect the sensitivity to ϕ_s. Therefore, performing the prescaling in an exclusive HLT2 selection is a robust procedure and does not need to be adjusted in any expected situation because a small fraction of the events suffices.

On the contrary, events in the detached sample are selected to have a maximal sensitivity on the determination of ϕ_s. Therefore, events in the detached sample *are* allowed to be

selected by HLT1 lines that use cuts which induce non-uniform proper time efficiencies. This leads to multiple possible procedures that can be followed.

1. To select events in the detached sample with the same criteria as in the prescaled sample, an exclusive HLT2 selection is performed in which a lifetime cutoff is used, and only HLT1 lines are included in which no cuts are applied that induce non-uniform proper time efficiencies. As the events in the prescaled sample are selected by the same HLT1 lines, by construction the prescaled and the detached samples are, except for the lifetime cutoff, selected with identical criteria. Since this HLT2 selection relies on exclusive reconstruction, a decay time cut can be applied online. This is implemented in the exclusive Hlt2Bs2JpsiphiDetached line.

2. To select a larger number of events, also other HLT1 lines, e.g. HLT1 muon lines implementing impact parameter criteria, can be included in the detached selection. As shown, this increases the number of selected signal events. Furthermore, the shape of the proper time efficiency is affected, and the inclusion of HLT1 lines not present in the prescaled sample, could introduce extra systematic effect. Therefore, these lines should only be included if they provide a significant increase in statistics.

3. To increase the rate even further, an inclusive selection can be considered for the detached sample. This in general is more robust against unexpected low (online) reconstruction efficiencies, since not all of the four final state tracks need to be reconstructed. This means that, *if* the HLT single track reconstruction efficiency would be *significantly* lower, many events would be lost. In the case of the described exclusive HLT2 algorithm, the expected efficiency is however larger than 90%, as shown before. In case an inclusive HLT2 selection is employed, all HLT1 lines can be included, which increases the signal efficiency, but increases also the number of possible systematic effects that can occur.

In case an additional, tighter lifetime criterium is applied offline, the overall proper time efficiency can be a perfect step function. In case the lifetime cut is only applied online, as possible in the first two procedures described before, the proper time efficiency curve is less trivial. Either way, to determine the proper time efficiency of the detached sample, as a check or to correct the description of the proper time distribution, the detached and the prescaled sample must be compared.

To determine the proper time efficiency induced by the lifetime cut, the relative and absolute sizes of the sample matter. As shown, for a nominal year of data taking the event samples are large and the prescale fraction can be very small in order to determine the proper time resolution with sufficient precision. In the start-up phase, this might not suffice. Since at the start-up the rate is expected not to be a critical issue, prescaling is not yet required, and the prescale fraction can be set up to 100%. As soon as the rate becomes critical, the prescale fraction can be lowered, making a smooth transition to a prescaled and a detached sample.

It can be concluded that, using the prescaled and the detached selections together, there are sufficient methods to adjust to bandwidth requirements, and adjust to unexpected inefficiencies, without losing sensitivity. To adapt to the rate, only the prescale fraction and lifetime cut need to be adjusted, whereas the 'real' selection criteria can remain unaffected.

4.8 Performance improvement

To demonstrate the advantage of the method of splitting a sample into a prescaled and a detached sample, the performance of this selection is compared below with the selection described in [44]. That selection consists of a single sample and has been developed with the restriction of an efficiency which is constant as a function of the proper time. To demonstrate the advantages of the division of the samples, the expected increase in sensitivity to ϕ_s is given, and causes of the increase explained.

To compare the performance, first the sizes and shapes of the backgrounds are compared. The lifetimes of the long-lived backgrounds after both selections are the same within uncertainties. Comparing the purities of the different selections given in table 4.11, it can be concluded that the dilutions of the sensitivity due to long lived background are approximately the same for both selections.

Selection	$\lvert\Delta m(B_s)\rvert$	$\lvert\Delta m(J/\psi)\rvert$	$\lvert\Delta m(\phi)\rvert$	B_{LL}/S
Detached & Prescaled selections	60 MeV/c^2	34 MeV/c^2	9 MeV/c^2	0.3
Selection in [44]	50 MeV/c^2	42 MeV/c^2	12 MeV/c^2	0.5

Table 4.11: *Purities of the selection described in this thesis and the selection described in [44], given the mass windows. All numbers are given in the whole t range.*

Next, the trigger efficiencies are compared. Whereas the selection in [44] can include only events selected by the HLT1 di-muon lines, in the detached selection events selected by all HLT1 muon lines can be included. Whereas the first has a (plateau) efficiency of 74%, the latter has a plateau efficiency of 92%, which clearly leads to an increase in yield of signal events.

In table 4.12 the yields for different trigger scenarios are given, including the lifetime cut $t > \tau_0 = 0.2$ ps in the detached selection. The increase in signal yield only, for the detached selection with $t > \tau_0 = 0.2$ ps, including all HLT1 muon lines, compared with the selection in [44], is 26%. This increase is due to an increase of the HLT1 efficiency as well as a loosening of the offline selection criteria.

	HLT1 muon × HLT2 detached	HLT1 di-muon × HLT2 detached	Selection in [44]
S	148 k ($t > \tau_0$)	122 k ($t > \tau_0$)	117 k (no τ_0)

Table 4.12: *Number of signal events S for different trigger scenarios used for the detached and prescaled selection, and for the selection in [44]. The number of events of the selection in [44] are given for the whole decay time region. The number of events given for the different trigger scenarios, are for the detached selection, assuming $t > \tau_0 = 0.2$ ps.*

Since the shapes of the offline efficiency as a function of the proper time and the angular observables are approximately the same, the relative sensitivity between the selection described in this thesis and the selection described in [44] (labelled with $'$), is calculated

using:

$$\begin{aligned}\frac{\sigma(\phi_s)'}{\sigma(\phi_s)} &= \frac{S/\sqrt{S+\alpha B}}{(S/\sqrt{S+\alpha B})'} \\ &= \frac{\sqrt{S}}{\sqrt{S'}}\frac{\sqrt{1+\alpha' B'/S'}}{\sqrt{1+\alpha B/S}}.\end{aligned} \tag{4.14}$$

Since both the purity and the behavior of the long-lived background in the sensitive region is the same, it follows that $\alpha \approx \alpha'$ and $S/B \approx S'/B'$. Therefore, (4.14) simplifies to

$$\frac{\sigma(\phi_s)'}{\sigma(\phi_s)} \approx \sqrt{\frac{S}{S'}}. \tag{4.15}$$

Since not only for the detached selection, but also for the selection in [44], events at small t hardly contribute to the sensitivity to ϕ_s, the number of signal events in the $t > \tau_0$ region should be compared. If only events from HLT1 di-muon lines are included, the increase is

$$\frac{\sigma(\phi_s)'}{\sigma(\phi_s)} \approx \sqrt{\frac{S(t>\tau_0)}{S'(t>\tau_0)}} = \sqrt{\frac{207\text{k} \times 74\% \times 92\% \times 86\%}{117\text{k} \times 86\%}} = \sqrt{120\%}. \tag{4.16}$$

If in the detached sample also the events triggered by the other HLT1 lines are included, the improvement increases to

$$\frac{\sigma(\phi_s)'}{\sigma(\phi_s)} \approx \sqrt{\frac{S(t>\tau_0)}{S'(t>\tau_0)}} = \sqrt{\frac{207\text{k} \times 92\% \times 90\% \times 86\%}{117\text{k} \times 86\%}} = \sqrt{145\%}. \tag{4.17}$$

It can be concluded that for a realistic scenario, including all HLT1 muon lines, and using the exclusive detached HLT2 selection, an increase of the sensitivity of $\sim 20\%$ can be expected for the described selection. This improvement is due to two facts: events from more HLT1 lines *with* cuts that induce non-uniform proper time efficiencies can be included in the detached sample, and cuts previously needed to suppress prompt background are not needed anymore, avoiding the unnecessary removal of signal events.

In addition to the above explained advantage of an increase of signal sensitivity, there is more flexibility in the described selection. The total rate of the two selections is adjustable, without influence on $\sigma(\phi_s)$; the mass windows are wider, in order to be able to study the background of the composite particles; and additional HLT1 lines can be included, thereby improving the robustness of the selection.

4.9 Conclusions

A selection procedure for $B_s^0 \to J/\psi\phi$ with a small number of cuts has been described and optimized on MC events. Iteratively optimizing the figure of merit FOM $= S/\sqrt{S+\alpha B}$, and determining with toy experiments the background weight $\alpha = 0.4 \pm 0.1$, the optimal selection criteria $p_T(\phi) > 1100$ MeV/c, $\chi^2(B_s$ vertex fit$) < 16$ and $\chi^2(B_s$ lifetime fit$) < 9$ have been determined. Additional p_T cuts are shown to be not necessary. The selection

is shown to be stable under changes of both the background distribution of the selection observables as well as the proper time distribution. Since the applied set of cuts is small and independent, the selection yields a small number of possible systematic effects.

Two selection criteria on the particle identity have been added: $\Delta \log \mathcal{L}_{K,\mu/\pi} > -5$ is required for all final state particle candidates. Since these have not been included in the optimization procedure, they could be studied and optimized, together with the other criteria, to improve the selection and possibly loosen the other selection criteria.

It has been demonstrated that prescaling the events with a lifetime smaller than a certain lifetime cut $\tau_0 < 0.20$ ps with a prescale fraction $f_{\text{pre}} > 0.1\%$, does not lead to a significant decrease of the sensitivity to ϕ_s. Furthermore, it has been shown which combination of lifetime cutoff and prescale rate gives the optimal sensitivity for ϕ_s for a given maximum yield.

The selection yields (207 ± 10) k signal events in the whole t-region for 2.0 fb^{-1}. The S/B ratio is equal to 0.33 ± 0.9 in the control sample for the long lived background only, while it is $S/B = 1/47$ including the prompt background. In the detached sample the prompt background disappears, and the fraction long-lived background decreases. Of the total background sample 99.3% are prompt background events and the long-lived background is observed to have a lifetime of 0.33 ± 0.05 ps.

To determine the expected performance of the trigger, the most efficient HLT1 lines have been presented and their efficiencies shown. Also, the exclusive HLT2 selection has been presented. With this knowledge the operational guidelines have been outlined. The proposed prescaling can be performed both online (in HLT2) or offline (in the stripping phase). In case the event rate is not critical, the selection can be configured without prescaling.

The performance of the selection has been compared to the selection in [44], which consists of a single sample and does not place requirements that induce an efficiency which is non-constant as a function of the proper time. Compared to this selection, selection criteria can be loosened or avoided, leading to an increase in signal yield and a decrease of the number of possible systematic effects. Furthermore, events selected by more HLT1 lines can be selected. When an exclusive HLT2 selection is used, and all HLT1 muon lines are included, an improvement in the sensitivity to $\sigma(\phi_s)$ of $\sim 20\%$ can be expected, due to an increase of the signal yield of 26%, of which relatively more events are in the sensitive high-t region.

Prescaling the control sample with a factor $f_{\text{pre}} = 1\%$, the total trigger rate is maximally (0.20 ± 0.09) Hz for the offline selection, for a lifetime cutoff $\tau_0 > 0.10$ ps. It can be concluded that using the lifetime cut and the prescale rate, one has 2 parameters to suppress the trigger rate if necessary, while maintaining an optimal sensitivity to ϕ_s, and being robust to a changing performance of the trigger.

Chapter 5

Fit methods

Event distributions as function of decay observables following from theory in practice get altered by detector acceptance, reconstruction and selection effects. Both the LHCb experiment, as described in chapter 3, as well as the event selection procedure, described in the previous chapter, cause non-uniform efficiencies, imply resolution effects, and add backgrounds to the theoretical distribution of the $B_s^0 \to J/\psi\phi$ decay described in chapter 2. In a fit to data to extract the physics parameters of interest, these effects must be accounted for in the description. In this chapter methods are described to implement these effects for all observables of the $B_s^0 \to J/\psi\phi$ decay, within situations applicable to the LHCb experiment.

Methods are described to take into account the efficiencies, resolutions, and backgrounds distorting the theoretically expected event distributions in the angular observables. A non-parametric inclusion of non-uniform angular efficiencies is given. The method is extended to also include angular resolutions. To include background in the description, the non-parametric description of the angular efficiency must be extended, and therefore both the efficiencies and the background are parameterized in terms of polynomials, hereby using the advantages of the non-parametric description of the angular efficiency.

When discussing the angular efficiencies, the problem of 'apparent' efficiencies is discussed, and it is shown that the correction for this effect is automatically implemented in the described methods. This is of particular importance to the prescaled and detached event selection procedures, aimed to reduce background as described in previous chapter. To perform a fit to events selected by these two triggers, a method is described to determine the induced proper-time acceptance from data.

Subsequently, to determine which fraction of the mesons have been labelled erroneously as an (anti-)B meson, a method is described to estimate the flavor mistag fraction from a control sample.

Finally, to improve the robustness of the fit of the total description to the data, an alternative parameterization of the strong phases is given.

A sample resulting from MC simulations is only employed for the determination of efficiencies. For all other described methods control samples are used. For the description of the background distribution the distributions in the mass sidebands are used; for the description of the proper time resolution and of the proper time efficiency as introduced in the trigger, a prescaled sample is used as described in previous chapter; for the tagging properties a data sample of a control channel is used.

To validate all fitting methods described in this chapter, a proper time dependent, angular analysis on the data sample of flavor tagged $B_d^0 \to J/\psi K^*$ events will be performed. As a cross check the physics parameters found for this decay channel will be compared to earlier results obtained by the B factories, before applying the same fitting methods to events of the $B_s^0 \to J/\psi\phi$ channel.

5.1 Introduction to angular efficiency corrections

In the MC an efficiency in three dimensions as function of the angular observables can be defined as the probability to reconstruct an event, given that it was generated with a certain set of decay angles. The efficiency can then be determined by dividing the number of reconstructed events in a certain bin by the number of generated events in this 3D bin. If a 3D efficiency is determined in this way several problems can arise.

First of all, an unpractically large data sample is required in order to get sufficiently small statistical uncertainties per bin. Secondly, a multi-dimensional function fitted to model such angular efficiency is in general not factorizable into different one-dimensional efficiencies; many terms can be required for a correct parameterization and the fit procedure can become unstable. Furthermore, when multiplying the resulting multi-dimensional function with the probability density function (PDF) of the signal decay as described in chapter 2, fitting and plotting can become slow in practice because of the numerical integrations which have to be performed, leading to time-consuming studies. Finally, systematic effects due to any parameterization must be considered.

The method of 'normalization weights' avoids the above problems. This method has been described and used by the BaBar experiment for the analysis of $B_d^0 \to J/\psi K^*$ events [26], [46], [47]. Contrary to the decay of the B_d, the B_s decays are described by $\Delta\Gamma_s \neq 0$, which implies that the angular distributions depend on the decay time. For this reason the method is extended in this chapter to describe $B_s^0 \to J/\psi\phi$.

Likelihood

Since the method described here uses a likelihood maximization, first a brief review of the maximum log-likelihood fitting procedure [6] is given.

The method of maximum likelihood consists of finding, for a set of independently measured quantities, say $\vec{x}$, the set of values for the parameters $\vec{\lambda}$ that maximizes the joint probability density for all the data, given by

$$\mathcal{L}(\vec{\lambda}) = \prod_e S(\vec{x}_e|\vec{\lambda}), \tag{5.1}$$

where the index e denotes the event. S is the (signal) PDF, which is larger than, or equal to, zero everywhere and is normalized over the observables. In the convention used in this thesis a PDF is written with a capital letter and is implicitly normalized over the variables written to the left of the $|$. For convenience the *unnormalized* signal function, s, is introduced, for which the following equation applies:

$$S(\vec{x}_e|\vec{\lambda}) = \frac{s(\vec{x}_e|\vec{\lambda})}{\int s(\vec{x}|\vec{\lambda})\mathrm{d}\vec{x}}. \tag{5.2}$$

Maximizing $\mathcal{L}$ is equivalent to maximizing $\ln \mathcal{L}$, so it suffices to solve the log-likelihood equation

$$\frac{\mathrm{d} \ln \mathcal{L}}{\mathrm{d}\lambda_n} = \frac{\mathrm{d}}{\mathrm{d}\lambda_n} \sum_e \ln \frac{s(\vec{x}_e|\vec{\lambda})}{\int s(\vec{x}|\vec{\lambda})\mathrm{d}\vec{x}} = 0. \tag{5.3}$$

In other words, having measured a set of observables $\vec{x}_e$, and assuming a signal PDF $S(\vec{x}|\vec{\lambda})$, this equation makes it possible to find the set of parameters $\vec{\lambda}$ for which the total probability of all the measurements (which is the product of the individual probabilities of each measurement) is maximal.

Including an efficiency, $\varepsilon(\vec{x})$, the maximum likelihood equation becomes

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_n} \sum_e \ln \frac{s(\vec{x}_e|\vec{\lambda})\varepsilon(\vec{x})}{\int s(\vec{x}|\vec{\lambda})\varepsilon(\vec{x})\mathrm{d}\vec{x}} = 0. \tag{5.4}$$

Since $\ln(AB) = \ln(A)+\ln(B)$, and the efficiency does not depend on the physics parameters $\vec{\lambda}$, the solution of (5.4) is independent of the efficiency in the numerator and therefore can be ignored. As a result, the efficiency only enters through the normalization in the denominator.

5.2 Formalism of the angular efficiency weights

In this section a method is described which corrects for an efficiency in the fit by determining the integral that appears in the denominator of (5.4). This is done for the case where the efficiency, as well as certain components of the PDF, are independent of the parameters $\vec{\lambda}$, and the efficiency factorizes in efficiencies of subsets of the observables. It is shown how to determine this integral for a general angular efficiency by summing over the selected MC events. The method is independent of the shape of the efficiency, and the efficiency does not need to be parameterized.

5.2.1 Likelihood maximization for $B \to J/\psi V$

In the signal channel $B_s^0 \to J/\psi\phi$ and the control channel of interest $B_d^0 \to J/\psi K^*$ four of the observables are the proper time t and the decay angles $\vec{\Omega} = (\cos\theta, \cos\psi, \phi)$, as defined in chapter 2. Since $\Delta\Gamma_s \neq 0$, the PDF of the $B_s^0 \to J/\psi\phi$ channel can not be written as a product of an angular dependent PDF $P(t)$ and a time dependent PDF $P(\vec{\Omega})$:

$$S(t, \vec{\Omega}) \neq P(t) \times P(\vec{\Omega}), \tag{5.5}$$

which means that the angular distribution and the proper time distribution need to be analyzed simultaneously[1].

To show the intertwined nature of the angular and proper time dependence, the signal PDF can be written as a normalized sum of products of angular and time dependent

[1]For the $B_d^0 \to J/\psi K^*$ channel $\Delta\Gamma = 0$, and the angular dependent and proper time dependent PDFs do factorize and can be studied seperately.

functions, as described in chapter 2:

$$S(t_e, \vec{\Omega}_e|\vec{\lambda}) = \frac{\sum_{i=1}^{6} A_i(t_e|\vec{\lambda}) f_i(\vec{\Omega}_e)}{\int\int \sum_{j=1}^{6} A_j(t|\vec{\lambda}) f_j(\vec{\Omega}) \mathrm{d}t \mathrm{d}\vec{\Omega}}. \tag{5.6}$$

Here i denotes the index of the angular functions $f_i(\vec{\Omega})$ and the time dependent amplitude terms $A_i(t|\vec{\lambda})$. The fact that angular dependent functions $f_i(\vec{\Omega})$ do not depend on the parameters $\vec{\lambda}$ is one of the most important ingredients of the method.

Since the summation over the indices of the functions speaks for itself, the explicit summation sign will be left out, using the summation convention $A_i f_i \equiv \sum_i A_i f_i$. In the case of no (or an ideal or constant) angular efficiency, the likelihood equation which has to be solved then reads

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k} \ln \frac{\prod_e s(t_e, \vec{\Omega}_e|\vec{\lambda})}{\int\int s(t, \vec{\Omega}|\vec{\lambda}) \mathrm{d}\vec{\Omega} \mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{A_i(t_e|\vec{\lambda}) f_i(\vec{\Omega}_e)}{\int\int A_j(t|\vec{\lambda}) f_j(\vec{\Omega}) \mathrm{d}\vec{\Omega} \mathrm{d}t} = 0, \tag{5.7}$$

where the PDF is normalized over the observables $t, \vec{\Omega}$ and differentiated with respect to the parameters $\vec{\lambda}$, which have to be determined.

For the decay rates at hand the integrals in the denominator can be calculated analytically. It should be noted that since the individual angular dependent functions do not depend on any of the physics parameters $\vec{\lambda}$, the expression factorizes

$$\int\int A_j(t|\vec{\lambda}) f_j(\vec{\Omega}) \mathrm{d}\vec{\Omega} \mathrm{d}t = \int A_j(t|\vec{\lambda}) \mathrm{d}t \int f_j(\vec{\Omega}) \mathrm{d}\vec{\Omega}, \tag{5.8}$$

and hence the second integral has to be determined only once in the likelihood optimization procedure.

5.2.2 Inclusion of efficiencies

Introducing a general angular and proper time dependent efficiency,

$$\varepsilon(t, \vec{\Omega}), \tag{5.9}$$

the observed shape of the distribution is

$$\begin{aligned} s_{\mathrm{obs}}(t, \vec{\Omega}|\vec{\lambda}) &\equiv s(t, \vec{\Omega}|\vec{\lambda}) \varepsilon(t, \vec{\Omega}) \\ &= A_i(t|\vec{\lambda}) f_i(\vec{\Omega}) \varepsilon(t, \vec{\Omega}). \end{aligned} \tag{5.10}$$

Since the efficiency in the numerator of (5.4) does not influence the estimation of the maximum of the likelihood, and it does not depend on the fitted parameters, it is irrelevant for the derivative, and (5.4) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{A_i(t_e|\vec{\lambda}) f_i(\vec{\Omega}_e)}{\int A_j(t|\vec{\lambda}) \int f_j(\vec{\Omega}) \varepsilon(t, \vec{\Omega}) \mathrm{d}\vec{\Omega} \mathrm{d}t} = 0. \tag{5.11}$$

From this it can be seen that the integral over the angles is independent of the physics parameters. Furthermore, no description of the efficiency function is needed: only the normalization integral over the angles has to be determined to include the angular efficiency, and *thus* it only has to be done once. While the fitted parameters change value in the likelihood maximization, the value of the angular integral remains constant. How this integral can be estimated from an unbinned data set is shown in the next subsection.

5.2.3 Normalization weights

It is shown here how the integrals over the angles $\vec{\Omega}$

$$\xi_j(t) \equiv \int f_j(\vec{\Omega})\varepsilon(t,\vec{\Omega})\mathrm{d}\vec{\Omega} \tag{5.12}$$

are determined. Whether an event gets accepted or rejected, not only depends on $(t,\vec{\Omega})$, but potentially also depends on all other observables in the event such as momentum, impact parameter, number of tracks, etc. Naming all other observables $\vec{z}$, the efficiency equals

$$\varepsilon(\vec{\Omega},t) = \frac{\int \varepsilon(t,\vec{\Omega},\vec{z})S(t,\vec{\Omega},\vec{z}|\vec{\lambda})\mathrm{d}z}{S(t,\vec{\Omega}|\vec{\lambda})}. \tag{5.13}$$

Since the probability to *generate* an event with certain $(\vec{\Omega},t)$ is independent of $\vec{z}$, it follows that $S(\vec{\Omega},t|\vec{\lambda}) = \int S(\vec{\Omega},t,\vec{z},\vec{\lambda})\mathrm{d}z$. It should be noted that the probability to generate an event with certain angles, however, *does* depend on the parameters $\vec{\lambda}$ and, since in general $\Delta\Gamma \neq 0$ (in which case the proper time and angular dependents PDF do not factorize), also on the proper time t.

Using $S(\vec{\Omega},t) = S(\vec{\Omega}|t)S(t)$, the integrals (5.12) can be written as

$$\xi_j(t) = \int \frac{f_j(\vec{\Omega})}{S(\vec{\Omega}|t,\vec{\lambda})}\varepsilon(t,\vec{\Omega},\vec{z})S(\vec{\Omega},\vec{z}|t,\vec{\lambda})\mathrm{d}z\mathrm{d}\vec{\Omega}. \tag{5.14}$$

To proceed further it should be noted that $S(\vec{\Omega},\vec{z}|t,\vec{\lambda})\mathrm{d}z\mathrm{d}\vec{\Omega}$ is the probability to generate an event for given $(t,\vec{\lambda})$ with observables between $[\vec{\Omega},\vec{\Omega}+\mathrm{d}\vec{\Omega}]$ and $[\vec{z},\vec{z}+\mathrm{d}\vec{z}]$. In other words, $S(\vec{\Omega},\vec{z}|t,\vec{\lambda})\mathrm{d}z\mathrm{d}\vec{\Omega}$ is the weight with which events are generated, and the integrals can be written

$$\begin{aligned}\xi_j(t) &\simeq \frac{1}{N_{\mathrm{gen}}}\sum_{e\in\{\mathrm{generated}\}}\frac{f_j(\vec{\Omega}_e)}{S(\vec{\Omega}_e|t_e,\vec{\lambda})}\varepsilon(t_e,\vec{\Omega}_e,z_e)\\ &= \frac{1}{N_{\mathrm{gen}}}\sum_{e\in\{\mathrm{accepted}\}}\frac{f_j(\vec{\Omega}_e)}{S(\vec{\Omega}_e|t_e,\vec{\lambda})},\end{aligned} \tag{5.15}$$

where, in the last step, it has been used that for each *individual* event the efficiency is a boolean (a zero or a one), since an event either gets accepted or not. Note that the overall factor $1/N_{\mathrm{gen}}$ is irrelevant for the likelihood maximization.

As a result, the equation which has to be solved, is

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k}\ln\mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k}\sum_e \ln\frac{A_i(t_e|\vec{\lambda})f_i(\vec{\Omega}_e)}{\int A_j(t|\vec{\lambda})\xi_j(t)\mathrm{d}t}, \tag{5.16}$$

where the normalization integrals can be determined as in (5.15) as a function of t.

It can be concluded that there is an expression for the normalization which can be determined from MC. The integral in which the angular efficiency appeared has become a sum over the accepted MC events. Furthermore, the ξ_j can be determined in the presence of a non-trivial proper time efficiency. The relationship between angular and proper time efficiencies is discussed in more detail in section 5.5.1, where it is shown that the weights can be determined correctly in the presence of so-called apparent efficiencies.

5.2.4 Inclusion of tagging flavour

In previous section it is assumed that the efficiency depends on the proper time and angular observables, and on observables that are not present in the theoretical signal PDF (5.6). This PDF, however, also depends on the flavor of the generated B_s meson at production. Therefore, as the most general case, the possibility of an efficiency which also depends on the flavor of the meson is considered, and the flavor of the generated meson (as an element of $\vec{z}$) must be taken into account explicitly. This changes the definition of the angular efficiency weights, as given in (5.15), since these are only valid for an efficiency and a PDF which do not depend on the flavor of the generated meson.

Defining a production flavor of the B_s, q, with different labels (B_s or $\bar{B}_s$) which are equally probable, the most general case in which the efficiency depends on this category is studied. Then

$$\varepsilon(t,\vec{\Omega}) \quad \rightarrow \quad \varepsilon(t,\vec{\Omega},q), \tag{5.17}$$

for which the most general efficiency reads

$$\varepsilon(t,\vec{\Omega},q) = \frac{\int \varepsilon(t,\vec{\Omega},q,\vec{z}) S(t,\vec{\Omega},q,\vec{z}|\vec{\lambda})\mathrm{d}z}{S(t,\vec{\Omega},q|\vec{\lambda})}. \tag{5.18}$$

Following the same reasoning as in the previous subsection, the following likelihood maximization must be solved:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k}\ln\mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k}\sum_e \ln\frac{A_i(t_e,q_e|\vec{\lambda})f_i(\vec{\Omega}_e)}{\sum_q\int A_j(t,q|\vec{\lambda})\xi_j(t,q)\mathrm{d}t}, \tag{5.19}$$

where the normalization integrals as a function of t, q are

$$\xi_j(t,q) = \frac{1}{N_{\mathrm{gen}}}\sum_e^{\mathrm{accepted}}\frac{f_i(\vec{\Omega}_e)}{S(\vec{\Omega}_e|t_e,q_e,\vec{\lambda})}. \tag{5.20}$$

Equation (5.20) implies that in general the integrals ξ_j do depend on time and the flavor. This is a consequence of the fact that the (angular) efficiency can depend on time or on tagging flavor. In case the angular efficiency factorizes, the expression of the normalization weights simplifies.

5.2.5 Special case: factorization of efficiencies

Assuming the angular efficiency factorizes, the total efficiency can be written as

$$\varepsilon(\vec{\Omega},t,q) = \varepsilon(\vec{\Omega})\varepsilon(t,q). \tag{5.21}$$

The weights can then be written as $\xi_j(t,q) = \varepsilon(t,q)\times\xi_j$, and the likelihood equation which has to be maximized is

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k}\ln\mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k}\sum_e \ln\frac{A_i(t_e,q_e|\vec{\lambda})f_i(\vec{\Omega}_e)}{\sum_q\int A_j(t,q|\vec{\lambda})\varepsilon(t,q)\mathrm{d}t\times\xi_j}. \tag{5.22}$$

In the case the efficiency is constant as a function of the tagging flavor and proper time, the likelihood maximization becomes

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k} \ln \mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{A_i(t_e, q_e|\vec{\lambda}) f_i(\vec{\Omega}_e)}{\sum_q \int A_j(t, q|\vec{\lambda}) \mathrm{d}t \times \xi_j}, \tag{5.23}$$

where the normalization weights have become (t, q) independent and equal to

$$\xi_j = \frac{1}{N_{\mathrm{gen}}} \sum_e^{\mathrm{accepted}} \frac{f_i(\vec{\Omega}_e)}{S(\vec{\Omega}_e|t_e, q_e, \vec{\lambda})}. \tag{5.24}$$

5.2.6 Summary of the method

The method to perform the log likelihood maximization can be summarized as follows.

1. First start by estimating the maximum likelihood equation (5.4).
2. Take $\vec{\lambda}$-independent terms ($\varepsilon(\Omega)$, N_{gen}) out of the differentiation with respect to λ_k.
3. Take $\vec{\Omega}$-independent terms A_i out of the normalization integral with respect to $\vec{\Omega}$.
4. Rewrite the normalization integral over the angles $\vec{\Omega}$ as a sum over selected MC events.

Although the formalism has been described for a general efficiency, in the rest of this chapter a special case is studied: a proper time and tagging independent angular efficiency. In the next section, where the method is validated, only an angular efficiency is considered. The log-likelihood maximization then becomes

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k} \ln \mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{A_i(t_e|\vec{\lambda}) f_i(\vec{\Omega}_e)}{\int A_j(t|\vec{\lambda}) \mathrm{d}t \times \xi_j}, \tag{5.25}$$

where the normalization integrals are determined as

$$\xi_j = \frac{1}{N_{\mathrm{gen}}} \sum_e^{\mathrm{accepted}} \frac{f_j(\vec{\Omega}_e)}{S(\vec{\Omega}_e|t_e, q_e, \vec{\lambda})}. \tag{5.26}$$

Since these integrals ξ_j do not depend on the physics parameters, they need to be determined only once, and can then be cached in the calculation of the PDF, as described in appendix D.

It should be noted that each event must be weighted by the probability to generate a set of angular observables for a given flavor and proper time. Since the per-event probability to generate a set of angular observables for the $B_s^0 \to J/\psi\phi$ decay depends on the values of the other observables, it is not correct to integrate over the other observables, as is possible for the $B_d^0 \to J/\psi K^*$ decay. Therefore, every event should be weighted by the probability *conditional* on t and q.

Since the integrals weight the efficiency for each separate angle-dependent function, the physical meaning of the normalization numbers ξ_j can be seen: these are the numbers which weight the efficiency of each $f_j(\vec{\Omega})$. Hence these are referred to as 'efficiency weights'.

5.3 Validation

The validation of using angular efficiency weights is presented in [48], and is summarized here. Two event samples are used: a 'reference sample' to determine the angular efficiency correction, and a 'fit sample' to use the efficiency correction in a fit of the physics parameters. Toy MC is studied, which implies that a simplified MC experiment is repeated many times, in order to test if the central values and the errors of the parameters are estimated correctly.

The first validation in [48] of the method of angular efficiency weights is performed by comparing the integrals for a constant efficiency

$$\xi_j = \int f_j(\vec{\Omega}) \mathrm{d}\vec{\Omega}, \tag{5.27}$$

with the numbers determined numerically from a corresponding toy MC using (5.26). The two results are identical within statistical uncertainties, as is summarized in table 5.1. The errors of the weights for each experiment are estimated from the expectation value E of the set of per-event measurements $\xi_j^e \equiv f_j(\vec{\Omega}_e)/S(\vec{\Omega}_e|t_e, q_e, \vec{\lambda})$, as

$$\sigma_j(\xi_j) = \sqrt{V_{jj}} = \sqrt{E(\xi_j^e \times \xi_j^e) - E(\xi_j^e) \times E(\xi_j^e)}. \tag{5.28}$$

Comparing the difference of the expected weights ξ_j^{ana} and the determined weights ξ_j^{num}, given the calculated errors $\sigma(\xi_j)$, a pull distribution is constructed, of which the results are summarized in table 5.1. The calculated uncertainties are identical to the width of the measured distribution, within statistical fluctuations, which means that the uncertainties of the efficiency corrections are estimated correctly.

	ξ_j^{ana}	$\langle \xi_j^{\mathrm{num}} \rangle$	$\sigma_p \equiv (\xi_j^{\mathrm{num}} - \xi_j^{\mathrm{ana}})/\sigma(\xi_j)$
ξ_1	11.170	11.171 ± 0.001	0.97 ± 0.03
ξ_2	11.170	11.169 ± 0.002	0.96 ± 0.03
ξ_3	11.170	11.170 ± 0.002	1.06 ± 0.03
ξ_4	0	-0.001 ± 0.002	0.96 ± 0.03
ξ_5	0	$+0.001 \pm 0.001$	1.01 ± 0.03
ξ_6	0	-0.001 ± 0.001	1.02 ± 0.03

Table 5.1: *Average $\langle \xi_j^{num} \rangle$ of the numerical estimates of ξ_j when no angular efficiency is applied. The central values are consistent with the expected analytical integral ξ_j^{ana}. Since the widths of the pull distributions σ_p are consistent with 1, the uncertainty $\sigma(\xi_j)$ is estimated correctly. More details can be found in [48].*

Subsequently, it is shown that the method is self-consistent. Performing a fit on an event sample, using the weights as retrieved from the sample *itself*, all fits return exactly the input values of the polarization amplitudes, without any statistical fluctuations. This means that all the information about the efficiency is indeed contained in the efficiency weights.

Furthermore, it is demonstrated that the method gives unbiased estimates for different parameter settings. Both for different efficiencies, and for different distributions (i.e. different polarization amplitudes $|A_i(0)|^2$) of the samples, the fit returns the correct physics parameters using the angular efficiency correction, as summarized in table 5.2.

Setting	$\mu_p(\|A_0\|^2)$	$\sigma_p(\|A_0\|^2)$	$\mu_p(\|A_\perp\|^2)$	$\sigma_p(\|A_\perp\|^2)$
Constant efficiency	-0.07 ± 0.06	0.97 ± 0.04	-0.07 ± 0.07	1.03 ± 0.05
'$B_d^0 \to J/\psi K^*$-like efficiency'	-0.05 ± 0.06	0.95 ± 0.04	-0.05 ± 0.06	1.00 ± 0.04
'Step efficiency'	-0.02 ± 0.06	0.97 ± 0.04	-0.07 ± 0.06	1.02 ± 0.05
$\|A_0\|^2 = 0.50$; $\|A_\perp\|^2 = 0.25$	-0.05 ± 0.06	0.96 ± 0.04	-0.08 ± 0.06	0.99 ± 0.04
$\|A_0\|^2 = 0.25$; $\|A_\perp\|^2 = 0.50$	-0.05 ± 0.06	0.99 ± 0.04	-0.04 ± 0.06	0.99 ± 0.04
$\|A_0\|^2 = 0.25$; $\|A_\perp\|^2 = 0.25$	-0.07 ± 0.06	0.97 ± 0.05	-0.03 ± 0.06	1.02 ± 0.05

Table 5.2: *The mean (μ_p) and the width (σ_p) of the pull distributions of the polarization amplitudes, for different settings. The amplitudes are normalized such that $|A_\parallel(0)|^2 = 1 - |A_0(0)|^2 - |A_\perp(0)|^2$. The $B_d^0 \to J/\psi K^*$-like efficiency and the step efficiency are shown in figure 5.2 and in figure 5.3, respectively. The $B_d^0 \to J/\psi K^*$-like efficiency resembles the expected angular efficiency for the $B_d^0 \to J/\psi K^*$ channel and is used since it is more extreme than the efficiency expected for $B_s^0 \to J/\psi\phi$. In the latter three experiments, in which the polarization amplitudes are varied, the '$B_d^0 \to J/\psi K^*$-like' efficiency has been applied. More details can be found in [48].*

Also, the systematical error due to the limited statistics in the reference sample is studied. Introducing $z = A \times \xi$, the product of $A \equiv |A_j(0)|^2$ and $\xi \equiv \xi_j$, and assuming no correlations between the two parameters, the relative uncertainty of z is

$$(\frac{\sigma_z}{z})^2 = (\frac{\sigma_A}{A})^2 + (\frac{\sigma_\xi}{\xi})^2. \tag{5.29}$$

Since the relative uncertainty of $|A_j|^2$ is inversely proportional to the number of events in the fit sample, N_{fit}, and the relative uncertainty of ξ_j is inversely proportional to the number of events N_{ref} used in the reference sample to determine the efficiency correction, the relative uncertainty of z is

$$\frac{\sigma_z}{z} \propto \sqrt{1 + N_{\text{fit}}^2/N_{\text{ref}}^2}. \tag{5.30}$$

Defining $n \equiv N_{\text{ref}}/N_{\text{fit}}$, the relative size of the reference sample compared to the fit sample, and fixing the efficiency weights, the statistical uncertainty on the amplitudes grows as

$$\sigma(|A_i|^2) \propto \sqrt{1 + (1/n)^2}. \tag{5.31}$$

This makes it possible to estimate the size of the reference sample required as not to significantly degrade the intrinsic precision of the measurement. The growth of the uncertainty as the reference sample shrinks, is tested and shown in table 5.3 to be consistent with (5.31).

Finally, it is shown that the method can also be used to describe a factorizing 3D angular efficiency for the realistic case for B_s mesons when $\Delta\Gamma \neq 0$. In particular, the method is

n	$\sqrt{1+(1/n)^2}$	pull($\|A_0\|^2$)	pull($\|A_\perp\|^2$)
5	1.02	1.04 ± 0.05	1.07 ± 0.05
3	1.05	1.11 ± 0.05	1.07 ± 0.05
2	1.12	1.15 ± 0.05	1.14 ± 0.05
1	1.41	1.38 ± 0.06	1.40 ± 0.06

Table 5.3: *Pulls for different n, the relative size of the reference samples compared to the fit samples.*

validated in the presence of a non-trivial proper time efficiency. This relationship between angular and proper time studies and the occurrence of 'apparent' efficiencies is described in more detail in section 5.5.

5.4 Angular resolutions

In the discussion hitherto a limited experimental precision on the decay angle measurement, the so-called angular resolution $R(\Omega, \Omega')$, has not been taken into account. To make a realistic model this should be included. Convoluting the PDF with an angular resolution modifies the likelihood equation (5.4) to

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k} \ln \mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{\int s_{\mathrm{obs}}(\vec{\Omega}) R(\Omega, \Omega'_e) \mathrm{d}\vec{\Omega}}{\int\int s_{\mathrm{obs}}(\vec{\Omega}) R(\Omega, \Omega') \mathrm{d}\vec{\Omega}' \mathrm{d}\vec{\Omega}}. \tag{5.32}$$

In this equation s_{obs} is the true shape of the PDF including an angular efficiency, as defined in (5.10). The true angular observables are denoted $\vec{\Omega}$, those after the inclusion of resolution are denoted $\vec{\Omega}'$.

Applying an efficiency *before* smearing with the resolution the equivalent of (5.25) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_k} \ln \mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{A_i(t_e|\vec{\lambda}) \int f_i(\Omega) \varepsilon(\Omega) R(\Omega, \Omega'_e) \mathrm{d}\vec{\Omega}}{\int A_j(t|\vec{\lambda}) \mathrm{d}t \times \xi'_j}, \tag{5.33}$$

where the normalization weights, including an angular resolution, have become

$$\xi'_j = \int\int f_j(\Omega) \varepsilon(\Omega) R(\Omega, \Omega') \mathrm{d}\vec{\Omega}' \mathrm{d}\vec{\Omega}. \tag{5.34}$$

In the limit of high angular resolution in comparison to the variation of $f_j(\vec{\Omega})\varepsilon(\vec{\Omega})$, $R(\vec{\Omega}, \vec{\Omega}')$ can be treated as a delta distribution, and it follows that $\xi'_j \to \xi_j$ and that the efficiency can be factorized out of the integral in the numerator, such that it again is irrelevant in the likelihood maximization.

In order to include a finite angular resolution in the fit such that the angular efficiency can be ignored, the efficiency should be defined differently, namely as a function of the *reconstructed* angles. Then the likelihood maximization becomes

$$0 = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \ln \mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{\int s(\vec{\Omega}) R(\vec{\Omega}, \vec{\Omega}'_e) \mathrm{d}\vec{\Omega} \times \varepsilon'(\vec{\Omega}')}{\int\int s(\vec{\Omega}) R(\vec{\Omega}, \vec{\Omega}') \mathrm{d}\vec{\Omega} \varepsilon'(\vec{\Omega}') \mathrm{d}\vec{\Omega}'}, \tag{5.35}$$

where $\varepsilon'(\Omega')$ is the efficiency with respect to the *convoluted* PDF. Then the efficiency drops out of the numerator, and it suffices to solve

$$0 = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \ln \mathcal{L}(\vec{\lambda}) = \frac{\mathrm{d}}{\mathrm{d}\lambda_k} \sum_e \ln \frac{A_i(t_e|\vec{\lambda}) \int f_i(\Omega) R(\Omega, \Omega'_e) \mathrm{d}\vec{\Omega}}{\int A_j(t|\vec{\lambda}) \mathrm{d}t \times \xi'_j}. \tag{5.36}$$

In that case the normalization integrals including angular resolutions are equal to

$$\begin{aligned} \xi'_j &\equiv \int f'_j(\vec{\Omega}') \varepsilon'(\vec{\Omega}') \mathrm{d}\vec{\Omega}' \\ &= \int\int f_j(\Omega) R(\Omega, \Omega') \mathrm{d}\vec{\Omega} \varepsilon'(\Omega') \mathrm{d}\vec{\Omega}'. \end{aligned} \tag{5.37}$$

This means that the integrals can be determined as

$$\begin{aligned} \xi'_j &= \frac{1}{N_{\mathrm{gen}}} \sum_e \frac{f'_j(\vec{\Omega}'_e)}{S'(\vec{\Omega}'_e|t_e, q_e, \vec{\lambda})} \\ &= \frac{1}{N_{\mathrm{gen}}} \sum_e \frac{\int f_j(\vec{\Omega}) R(\vec{\Omega}, \vec{\Omega}'_e) \mathrm{d}\vec{\Omega}}{\int S(\vec{\Omega}|t_e, q_e, \vec{\lambda}) R(\vec{\Omega}, \vec{\Omega}'_e) \mathrm{d}\vec{\Omega}}, \end{aligned} \tag{5.38}$$

in which the convoluted PDF and angular basis functions, respectively, equal

$$S'(\vec{\Omega}') \equiv \int S(\vec{\Omega}) R(\vec{\Omega}, \vec{\Omega}') \mathrm{d}\vec{\Omega}\,; \quad f'_j(\vec{\Omega}') \equiv \int f_j(\vec{\Omega}) R(\vec{\Omega}, \vec{\Omega}') \mathrm{d}\vec{\Omega}. \tag{5.39}$$

Summarizing, the angular resolution can be taken into account in the weights in a similar way as the angular efficiency.

Alternatively, the angular resolutions can be chosen to be ignored. In order to do so safely, they have to be small in comparison with the scale on which the distribution and the efficiency vary. In that case the same fit as in the previous section can be performed. To determine when this approximation is sufficient and to determine possible systematic effects induced by this approximation, a toy study is performed in next chapter.

5.5 Apparent efficiencies

It has been shown in section 5.2 that the efficiency weights correct for a non-uniform angular efficiency. The formalism has been based on [26], where it has been applied to the $B^0_d \to J/\psi K^*$ analysis. In that decay $\Delta\Gamma = 0$, and as such the angular and the proper time distributions factorize. As a consequence, the angular efficiency corrections have been determined by integrating over the proper time distribution.

However, in case that $\Delta\Gamma_s \neq 0$, which is valid for the $B^0_s \to J/\psi\phi$ decay, this is no longer correct, since a non-uniform proper time efficiency induces extra 'apparent' angular efficiencies[2]. In order to include proper time and angular efficiencies consistently, 'true' and 'apparent' efficiencies should be separated. This is of particular importance to the detached sample of $B^0_s \to J/\psi\phi$ events, which is obtained by applying an explicit decay time criterium. The method described in this chapter is therefore extended to be used for $\Delta\Gamma \neq 0$, and it is shown that the weights handle the apparent efficiencies correctly.

[2]Apparent efficiencies are also known as 'pseudo' efficiencies.

5.5.1 Apparent angular efficiencies

In the $B_s^0 \rightarrow J/\psi\phi$ decay, due to the nonzero value of $\Delta\Gamma_s$, the angular distributions are different for different slices of proper time. This is due to the fact that the angular distributions of the polarization amplitudes differ, and the relative contributions of the polarization amplitudes change as a function of proper time. An example of the angular distributions for different time slices is shown in appendix B.

Expressed in terms of the notation used in the previous sections, the probability of a set of $\vec{\Omega}$ for given t does not equal the time integrated probability for that particular $\vec{\Omega}$ to occur. Consequently, 'apparent' efficiencies can be introduced: integrated efficiencies which can be introduced through the physics correlations between the proper time and angular distributions. Inefficiencies of amplitudes in a certain observable, automatically lead to apparent efficiencies of the same amplitude in other observables.

The cause of such apparent efficiencies can be seen as follows. Consider an event with a certain set of decay angles for which a certain proper time is highly probable, and suppose that detection of this proper time is very inefficient. It is probable that this event will not end up in the final sample. Since these kind of events are unlikely to be accepted, also this set of angles becomes very unlikely to be accepted. On the other hand, events with a set of angles for which these particular inefficient proper times are highly *im*probable are relatively more likely to be observed. Hence, integrated over the proper time, this correlation leads to an 'apparent' efficiency in the angular acceptance.

In figure 5.1 an example of a toy simulation of this effect is shown: the apparent angular efficiency as induced by applying a proper time efficiency. This result is obtained in the following way. First, a proper time and angular dependent MC is generated from the theoretical PDF of the $B_s^0 \rightarrow J/\psi\phi$ decay. Then an efficiency is applied, on the proper time only. The proper time efficiency can be seen in figure 5.1(a) and is low for small t and constant for large t, as for the detached selection. The apparent efficiencies, as induced on the angular observables, are shown in figure 5.1(b,c,d).

The deviation from constant efficiencies as a function of $\cos\theta$, $\cos\psi$ and ϕ, integrated over the other observables, arises solely from the shape of the proper time efficiency. For example, the central region of $\cos\theta$, around $\cos\theta \sim 0$, has a larger contribution from events at large t, compared to the regions around $|\cos\theta| \sim 1$. Accepting relatively more events at large t automatically leads to accepting relatively more events with angles which are probable for these t. To summarize: the 'true' angular efficiency is constant, however the (time) *integrated* 'apparent' angular efficiencies are not.

This effect has consequences for the determination of angular efficiencies. It means that, in general, efficiencies can not be determined correctly while integrating over other observables, even when they factorize. When including not only the true, but also the apparent efficiencies, the efficiency is overcompensated. This overcompensation can be avoided by making four dimensional efficiencies, which lead to PDFs with CPU intensive performance.

It is shown in table 5.4 that one of the most striking features of the method of efficiency weights is the immunity to this problem: in the presence of a proper time efficiency, which induces an apparent angular efficiency, the angular efficiency correction, as determined with the weights, is found to be consistent with a constant true angular efficiency. This means that the method of normalization weights can be used to determine true angular

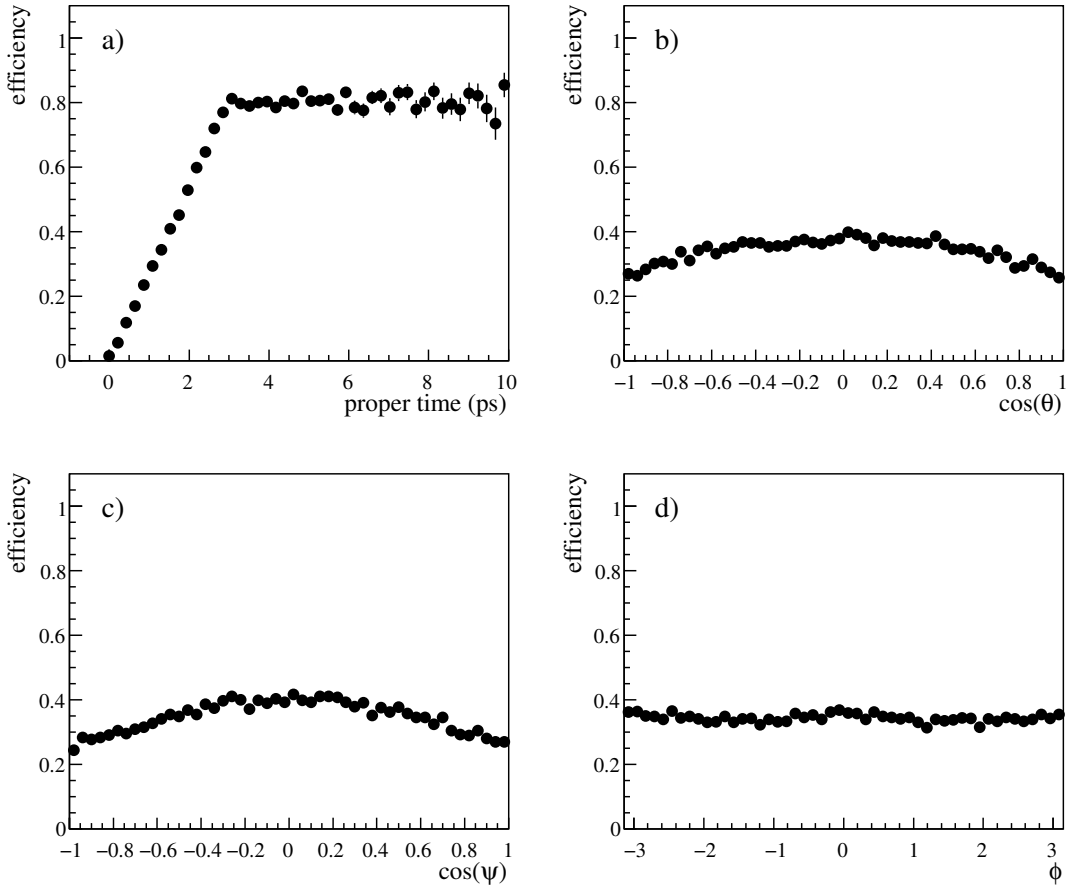


Figure 5.1: *The apparent efficiencies in the angular observables are shown in (b), (c), (d). These are induced by a true efficiency function, which is generated in the proper time observable as shown in (a), where the height of the plateau is arbitrary. For illustration purposes, the value of* $\Delta\Gamma_s/\Gamma_s$ *is taken to be* 0.72*, approximately ten times the SM value in the* B_s^0 *system. Since the lifetime induces a difference in the efficiency of the polarization amplitudes, are the shapes of the efficiencies in the angular observables superpositions of the angular functions* $f_i(\vec{\Omega})$*.*

efficiencies in the presence of apparent efficiencies and hence to describe a factorizing 3D angular efficiency in the presence of a non-trivial proper time efficiency. This is important for the 'detached' $B_s^0 \to J/\psi\phi$ selections, in which, on purpose, a lifetime cut is imposed. The weights are determined correctly, since the value of the PDF is evaluated *conditional* on t.

Finally, also if $\Delta\Gamma = 0$ apparent angular efficiencies can be induced. For example, apparent angular efficiencies can be induced by a true efficiency in a different decay angle. This is shown in figure 5.2, simulating events of the $B_d^0 \to J/\psi K^*$ decay[3] in which angles

[3] $B_d^0 \to J/\psi K^*$ has $\Delta\Gamma = 0$, since it is a B_d decay. Since this is also a $P \to VV$ decay, the angular distribution of this decay is similar to that of $B_s^0 \to J/\psi\phi$.

variable	pull mean	pull width
ξ_1	0.04 ± 0.07	0.97 ± 0.05
ξ_2	0.06 ± 0.07	0.99 ± 0.05
ξ_3	-0.03 ± 0.07	1.00 ± 0.05
ξ_4	0.03 ± 0.07	0.99 ± 0.05
ξ_5	0.03 ± 0.07	1.08 ± 0.05
ξ_6	0.08 ± 0.07	1.05 ± 0.05

Table 5.4: *Fit results of gaussian PDFs to pull distributions of the ξ_j, for a sample with $\Delta\Gamma_s/\Gamma_s = 0.72$ on which a lifetime cut is imposed. The applied angular efficiency is constant, and the estimated ξ_j are compared with the ξ_j expected for a constant angular efficiencies. Even though the angular efficiencies appear to be not flat due to apparent efficiencies, the normalization weights ξ_j are determined as in agreement with a constant angular efficiency.*

$\cos\psi \to 1$ are accepted less likely. This induces low efficiencies for $|\cos\theta| \to 0$ and $|\phi| \to \pi/2$, but these efficiencies are apparent. This leads to a problem if factorizing angular efficiencies are determined one-dimensionally, but since the weights are determined three-dimensionally, this problem is avoided.

5.5.2 Apparent proper time efficiencies

Apparent efficiencies can be induced also in other than angular observables. For example, if $\Delta\Gamma \neq 0$, angular efficiencies can induce pseudo proper time efficiencies, following the same reasoning as for the apparent angular efficiencies. As before, events with proper times that preferably decay to an angle which is less efficient to detect, are themselves more inefficient than the events with proper times for which this inefficient decay angle is less probable.

Again, as an illustration, in figure 5.3 the results of a toy MC are shown, applying an efficiency to $\cos(\psi)$ when $\Delta\Gamma_s/\Gamma_s = 0.72$, ten times the SM value. Similar to the case $\Delta\Gamma = 0$, this results in non-trivial apparent efficiencies in the other angles. But, more importantly, it now also results in a significant non-uniform efficiency in the proper time.

Description of apparent proper time efficiencies

Apparent efficiencies, and in particular the apparent proper time efficiency, can be understood in terms of 'inefficient amplitudes' [49]. Describing the total PDF, as a simplification, in terms of two components, $A_{1,2}$, with lifetimes $\tau_{1,2}$, the PDF becomes

$$P_{\text{theory}}(t) = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}. \tag{5.40}$$

Suppose the components both have different, proper time *independent*, angular distributions. An angular efficiency leads to amplitude efficiencies $\varepsilon_{1,2}$, such that the PDF after this efficiency is

$$P_{\text{eff}}(t) = \varepsilon_1 A_1 e^{-t/\tau_1} + \varepsilon_2 A_2 e^{-t/\tau_2}, \tag{5.41}$$

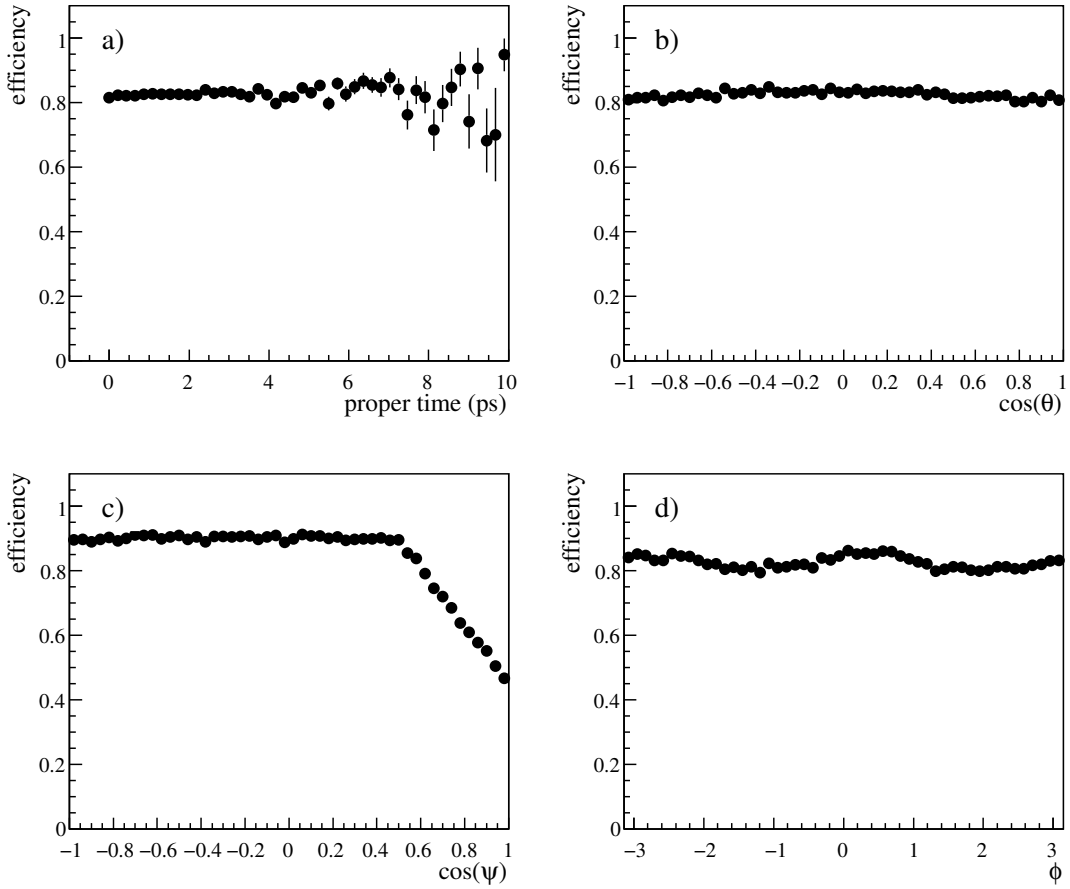


Figure 5.2: *Apparent efficiencies for $B_d^0 \to J/\psi K^*$ toy data ($\Delta\Gamma$=0) with realistic parameters, after applying the '$B_d^0 \to J/\psi K^*$-like' efficiency (with arbitrary scale) to* $\cos(\psi)$ *in (c) only. The proper time efficiency is constant, whereas the non-trivial efficiencies in* $(\cos\theta, \phi)$ *are apparent efficiencies. This effect is observed in $B_d^0 \to J/\psi K^*$ [20].*

which together with equation (5.40) lead to the following time dependent efficiency

$$\varepsilon(t) = \frac{P_{\text{eff}}(t)}{P_{\text{theory}}(t)} = \frac{\varepsilon_1 A_1 e^{-t/\tau_1} + \varepsilon_2 A_2 e^{-t/\tau_2}}{A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}}. \tag{5.42}$$

This equation is plotted in figure 5.4, which is in qualitative agreement with the apparent proper time efficiency seen in figure 5.3(a).

Using this simplified picture, an estimate of the magnitude of the effect can be determined. For $B_s^0 \to J/\psi\phi$ the maximal relative effect, substituting the efficiency determined in (5.42), is

$$\frac{\Delta\varepsilon}{\langle\varepsilon\rangle} \equiv \frac{\varepsilon(t\to\infty) - \varepsilon(t=0)}{[\varepsilon(t\to\infty) + \varepsilon(t=0)]/2} = 2 \times \frac{(1+\frac{A_s}{A_l}) - (1+\frac{\varepsilon_s A_s}{\varepsilon_l A_l})}{(1+\frac{A_s}{A_l}) + (1+\frac{\varepsilon_s A_s}{\varepsilon_l A_l})}, \tag{5.43}$$

where the index l denotes the long living component, and s the short living component. In the SM the long living component is the CP-odd component $|A_\perp|^2$, which is expected to be

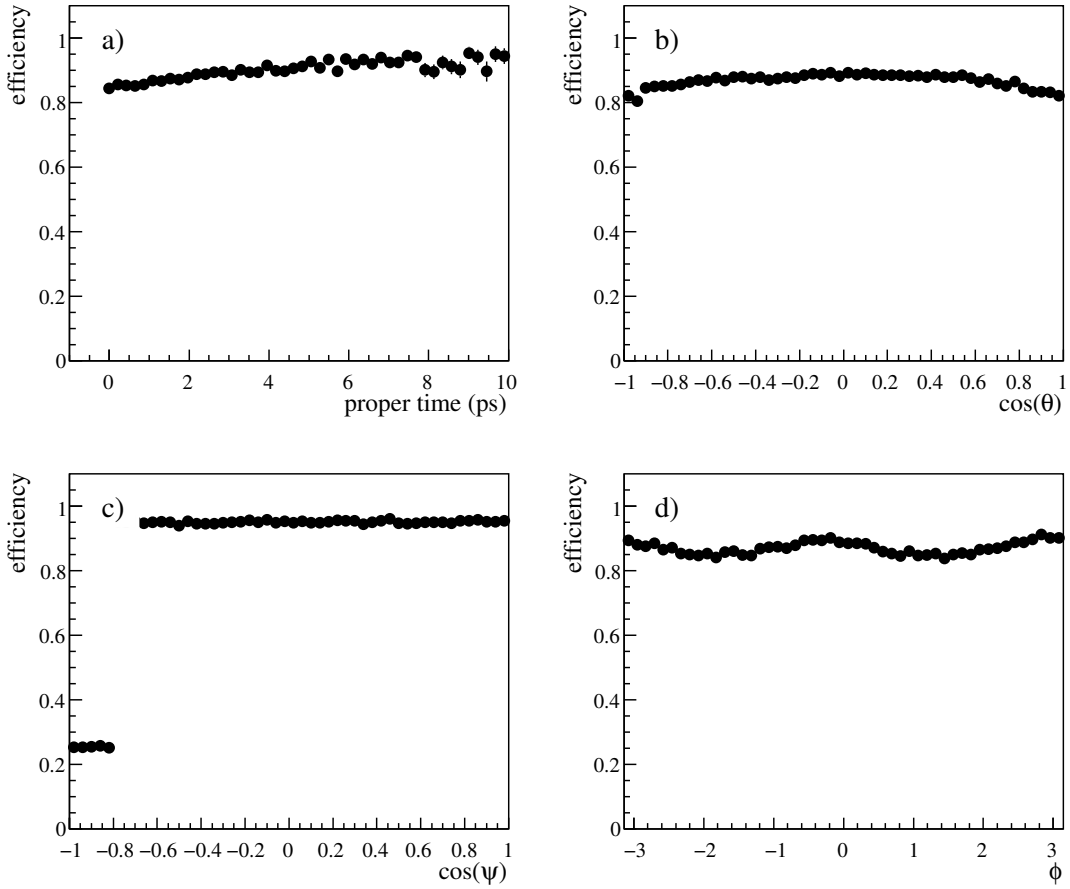


Figure 5.3: *The efficiencies for $B_s^0 \to J/\psi\phi$ toy data for the proper time (a) and angular observables (b), (c), (d) after applying an angular efficiency (with arbitrary scale) to the observable* $\cos(\psi)$*. The value of $\Delta\Gamma_s$ is ten times the default value, leading to a clear apparent proper time efficiency.*

20% of the total decay [60]. In that case

$$\frac{\Delta\varepsilon}{\varepsilon_l} = 2 \times \frac{5 - (1 + 4 \times \frac{\varepsilon_s}{\varepsilon_l})}{5 + (1 + 4 \times \frac{\varepsilon_s}{\varepsilon_l})}. \tag{5.44}$$

This means that if, due to a realistic angular acceptance, 10% more long living events are accepted than short living, a relative proper time acceptance of at most 8% is induced. The speed at which this asymptote is approached depends on the size of $\Delta\Gamma_s$.

5.5.3 Solutions to apparent efficiencies

Since apparent efficiencies can lead to incorrect compensation of the efficiencies, correction of apparent efficiencies should be avoided. Three solutions to avoid or control apparent efficiencies have been used in the analysis in this thesis.

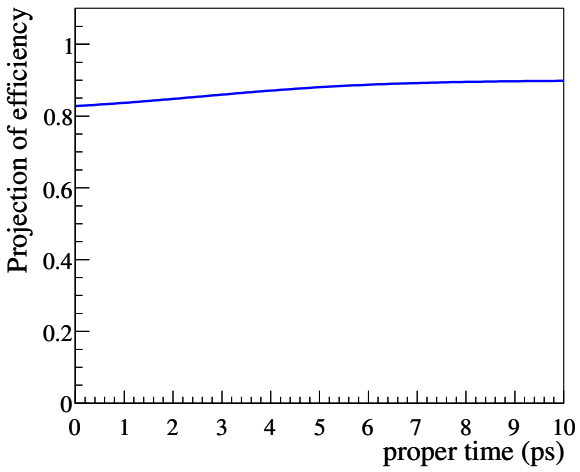


Figure 5.4: *The theoretical proper time efficiency according to equation (5.42) with the settings:* $\tau_1 = 2.50$ *ps,* $\tau_2 = 1.07$ *ps,* $A_1 = 0.2$*, and* $A_2 = 0.8$*. For the efficiencies* $\varepsilon_2/\varepsilon_1 = 0.9$ *is chosen. The behavior is qualitatively in agreement with figure 5.3(a).*

The first solution is to use a MC sample with $\Delta\Gamma_s = 0$. Then the angular and proper time distributions do factorize, and hence the angular efficiency does not influence the proper time efficiency, and vice versa, and these apparent efficiencies are avoided. This means two factorizing efficiencies can be determined, a one-dimensional efficiency for the proper time and a three-dimensional efficiency for the angular distribution.

A second solution is specific for the described method. The method of efficiency weights, contrary to more standard methods, is able to distinguish between true and apparent efficiencies. This is due to the fact that the efficiency is defined as the probability an event is accepted, given that it was generated. The efficiency is determined by weighting events with the per-event probability for a certain measurement of Ω. So the probability is determined conditional on t, exactly as it is defined. The proper time efficiency does not need to be known, and hence the probability to find a certain t does not need to be known, since for every event only the probability for a *given* t to find a certain $\vec{\Omega}$ is required. Hence, the method of normalization weights is immune to the problems of apparent efficiencies induced by proper time efficiencies.

As a third solution, control samples can be used which only differ by one selection criterium. For example, the prescaled and detached selection are exactly the same, up to a lifetime cut. Therefore, the apparent proper time efficiencies in both samples, due to e.g. p_T cuts, are exactly the same and cancel in the comparison. The proper time efficiency of the detached sample compared to the prescaled sample therefore is a true efficiency[4].

[4]Equivalently, the difference in angular efficiency is a pure apparent efficiency.

5.6 Fit to prescaled and detached samples

In the previous section it is explained how to deal with apparent efficiencies. Using the weights, apparent angular efficiencies due to proper time efficiencies are avoided. This is of importance to the detached sample, which is selected with a non-uniform proper time efficiency. Having solved the problem of apparent efficiencies, in this section a simultaneous fit to both the detached sample and the prescaled sample is described.

In order to suppress the rate originating from prompt background events, while having a maximal sensitivity to ϕ_s, two selections are used, as described in chapter 4. One selection selects events with a lifetime larger than a certain value ('detached' events) and provides sensitivity to ϕ_s. Another selection is prescaled, and provides a control sample without lifetime cut, necessary to determine the proper time resolution. A subset of the events is present in both samples.

Assume that the two samples are obtained from two lifetime dependent triggers, say $T_{\rm det}$ and $T_{\rm pre}$, which are running simultaneously. One trigger induces a lifetime cut with efficiency $\varepsilon_d(t)$, the other one prescales the events with an efficiency ε_p. Examples of such efficiencies are shown in figure 5.5.

A simultaneous fit to such two samples works in the following way. Using the overlap of the two samples, and the relative number of events in each sample, their relative efficiencies can be determined from the data. Furthermore, the uncertainty of the efficiency function $\varepsilon_d(t)$ due to the finiteness of both samples is propagated to the physics parameters, such that the uncertainties of the physics parameters are estimated correctly.

In case the decisions of both triggers are independent, there are four different possible combinations of trigger decisions. These decisions, and the corresponding efficiencies, are listed in table 5.5. The combination in which neither trigger fired is not used, since in general these events are not available.

$T_{\rm det}$ decision	$T_{\rm pre}$ decision	Efficiency $\varepsilon_i(t)$	PDF
1	1	$\varepsilon_{dp}(t) \equiv \varepsilon_d(t) \times \varepsilon_p$	P_{dp}
1	0	$\varepsilon_{d\bar{p}}(t) \equiv \varepsilon_d(t) \times (1-\varepsilon_p)$	$P_{d\bar{p}}$
0	1	$\varepsilon_{\bar{d}p}(t) \equiv (1-\varepsilon_d(t)) \times \varepsilon_p$	$P_{\bar{d}p}$
0	0	$\varepsilon_{\bar{d}\bar{p}}(t) \equiv (1-\varepsilon_d(t)) \times (1-\varepsilon_p)$	-

Table 5.5: *The possible trigger decisions in the case of two independent triggers T_{det} and T_{pre}, and the corresponding efficiencies and PDFs. The latter combination, in which neither trigger fires, is in general not observable.*

If the underlying PDF, prior to a trigger decision, is named $P(t)$, it follows that the PDFs observed after selection are

$$P_{ij}(t) \equiv \frac{P(t)\varepsilon_{ij}(t)}{\int P(t)\varepsilon_{ij}(t)\mathrm{d}t}. \tag{5.45}$$

Each event in the sample, triggered by one or both of these triggers, is described by the corresponding PDF. Given a combination of trigger decisions, the underlying PDF, is multiplied by the corresponding efficiency for that given combination of trigger decisions. Three

PDFs with shared parameters are then fitted simultaneously to a data set that is a sum of three non-overlapping subsets. Then, from data, the (physics) parameters of $P(t)$ *and* the shape of the separate efficiencies $\varepsilon_{d,p}(t)$ can be determined.

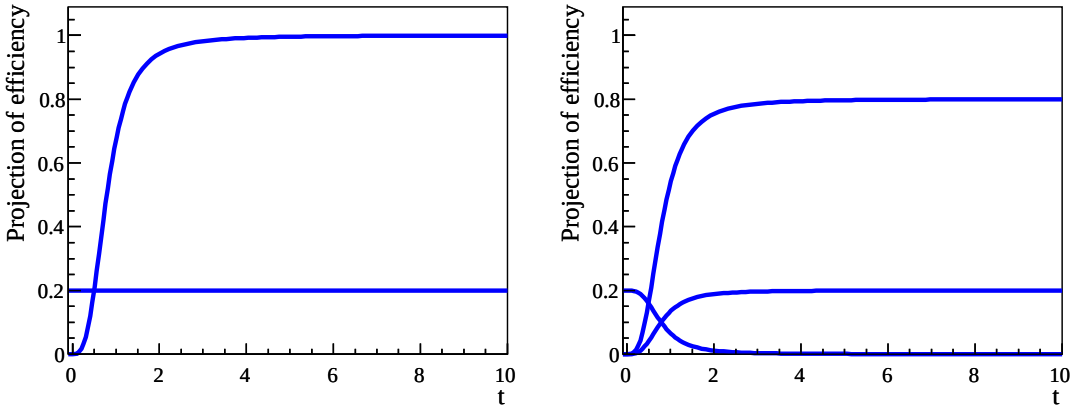


Figure 5.5: *Possible examples of efficiencies in the presence of a detached and a prescaled trigger. On the left the efficiency of the prescaled sample, $\varepsilon_p = 20\%$, and the efficiency of the detached sample, $\varepsilon_d(t)$, a smooth step function as a function of proper time. On the right the efficiencies, $\varepsilon_{ij}(t)$, of three non-overlapping samples are shown.*

Since most signal events are present in the detached sample and given that signal events around $t = 0$, due to the large prompt background, hardly contribute to the signal sensitivity, the sensitivity to the signal parameters originates from the events in the detached sample. Since the prompt background is only present in the prescaled control sample, the sensitivity to the proper time resolution originates from the events in the prescaled sample. The prescale factor is determined from the number of prescaled and detached events compared to the number of detached events. The sensitivity to the efficiency of the detached sample, $\varepsilon_d(t)$, originates from the comparison of the events which are both prescaled and detached compared to all prescaled events. This means that the prescaled and detached selection, as described in chapter 4, can be used in an analysis to determine ϕ_s.

Finally, 250 toy experiments are performed, fitting simultaneously to a prescaled and a detached sample. For every experiment 50 k events of both signal and background are generated, in a ratio of 1 : 4, corresponding to about 0.2 fb^{-1} luminosity at the LHC. The signal events have a lifetime of 1.47 ps, the main parameter of this test. The background is a purely prompt background sample, centered around a mean of zero with a width of 35 fs, as is the case for the signal proper time resolution. The prescale fraction is taken to be 10% and the detached efficiency is a 'smooth step' function, as shown in figure 5.5 leading to the distributions as already shown in previous chapter. Fitting for the decay time resolution, the prescale fraction and the efficiency shape, the results in figure 5.6 are found: the distribution of the fits to the lifetime are centered around the input value as also follows from the pull distribution, which furthermore shows that the uncertainty of the lifetime is correctly estimated.

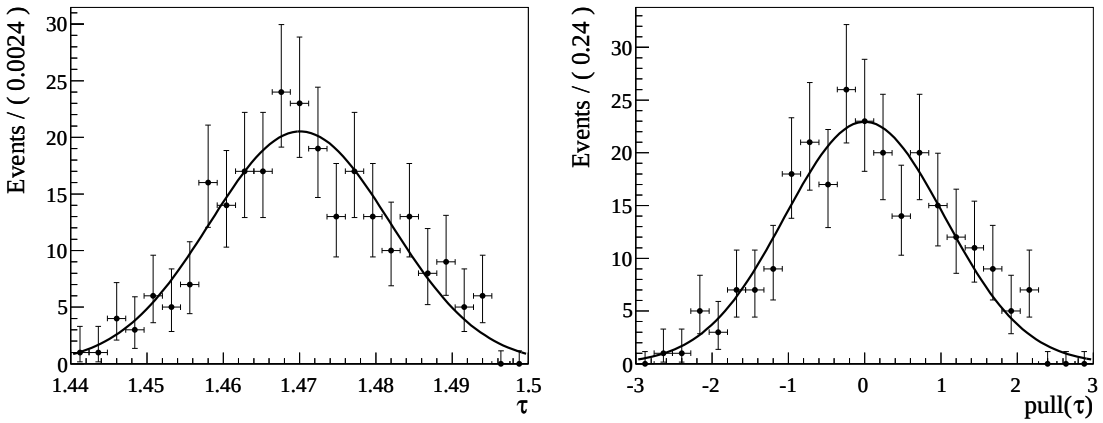


Figure 5.6: *Fit distribution and pull distribution of lifetime with prescaled and detached toy. The mean (μ) and the width (σ) of the lifetime distribution are $\mu = 1.4701 \pm 0.001$, $\sigma = 0.01$; the mean and width of the pull distribution are $\mu = 0.00 \pm 0.08$, $\sigma = 1.05 \pm 0.06$.*

5.7 Inclusion of background in the angular distribution

In previous section a description is given to perform a physics sensitivity fit in the presence of a background suppressing event selection. However, the remaining selected background events still need to be taken into account in the analysis. To understand how background affects the fit and the usage of the efficiency weights, the likelihood maximization procedure needs to be discussed. As is shown below, the method of weights must be extended in order to be able to make use of the advantages of the method also in the presence of background.

In case signal only is considered, the likelihood maximization requires no knowledge of the shape of the 3D angular efficiency as input. Only the numbers that act as weights in the normalization are needed. However, including a background PDF B, with a signal over signal-plus-background ratio f, the log likelihood maximization becomes

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\lambda_n} \ln(\prod_e \mathcal{L}(\vec{x}_e|\vec{\lambda})) &= \frac{\mathrm{d}}{\mathrm{d}\lambda_n} \sum_e \ln[f S_{\mathrm{obs}}(\vec{x}_e|\vec{\lambda}) + (1-f) B(\vec{x}_e|\vec{\lambda})] \\ &= \frac{\mathrm{d}}{\mathrm{d}\lambda_n} \sum_e \ln[f \frac{s(\vec{x}_e|\vec{\lambda})\varepsilon(\vec{x})}{\int s(\vec{x}_e|\vec{\lambda})\varepsilon(\vec{x})\mathrm{d}x} + (1-f) B(\vec{x}_e)] = 0. \quad (5.46) \end{aligned}$$

Thus, the signal efficiency in the numerator can no longer be neglected. In order to be able to include background by describing its distribution, the shape of the efficiency needs to be determined[5]. Furthermore, the knowledge of this shape of the efficiency is necessary for visualization purposes: to show the agreement between data distributions and projections of the physics PDF, the efficiency needs to be known.

[5]A different solution is to perform a background subtraction. Since in the fit to both the prescaled and detached sample the background is used to determine the resolution, this method adds extra complications for the used selection. Furthermore, this leads to an underestimation of the errors of the physics parameters, and therefore need to be recalculated, as shown in [47].

In this section a method is described to obtain a 3D description of the angular efficiency, such that background can be included in the fit and such that a plot of the angular efficiency and the accepted distribution can be made. This is done, as in [50], by decomposing the angular efficiency into a set of orthonormal basis functions. To still make maximal use of the advantages of the weights, the method is a continuation of the likelihood maximization method using normalization weights: the decomposition is determined by summing over a sample of accepted MC events. Furthermore, this method is extended to describe the angular distribution of the background.

5.7.1 Three-dimensional efficiency decomposed in polynomials

Any function of three observables $x, y, z \in [-1, 1]$ relevant for the description of a 3D efficiency can be decomposed into the monomials $x^a y^b z^c$, with (a, b, c) non-negative integers. Introducing coefficients c_{abc}, the 3D function $\varepsilon(x, y, z)$ can be written as

$$\varepsilon(x, y, z) = c_{abc} x^a y^b z^c. \tag{5.47}$$

Since any polynomial is a linear combination of these monomials, this remains valid when changing basis [50], in particular when choosing the basis functions in which the efficiency function is decomposed to be the Legendre polynomials. The orthonormal basis polynomials[6] $P_l(\omega)$, for an observable $\omega \in [-1, 1]$, equal

$$P_l(\omega) \equiv L_l(\omega)\sqrt{\frac{2l+1}{2}}, \tag{5.48}$$

with L_l the Legendre polynomials and l a non-negative integer. The basis obeys the orthonormality relations

$$\int P_i(\omega) P_k(\omega) \mathrm{d}\omega = \delta_{ik}. \tag{5.49}$$

Decomposing the angular efficiency in this basis, it can be written as

$$\varepsilon(\vec{\Omega}) = \sum_{abc} e_{abc} P_a(\cos\theta) P_b(\cos\psi) P_c(\phi). \tag{5.50}$$

The efficiency coefficients e_{abc} can now be determined in the same way as the normalization weights [51]. Using the orthonormality constraints (5.49), it follows that the coefficients can be determined as

$$\begin{aligned} e_{abc} &= \int [\varepsilon(\vec{\Omega}) P_a(\cos\theta) P_b(\cos\psi) P_c(\phi)] \mathrm{d}\cos\theta \mathrm{d}\cos\psi \mathrm{d}\phi \\ &\simeq \frac{1}{N_{\text{gen}}} \sum_{e}^{\text{generated}} \varepsilon(\vec{\Omega}_e) \frac{P_a(\cos\theta_e) P_b(\cos\psi_e) P_c(\phi_e)}{S(\vec{\Omega}_e | t_e, q_e, \vec{\lambda})}. \end{aligned} \tag{5.51}$$

So the efficiency coefficients can be determined as

$$e_{abc} = \frac{1}{N_{\text{gen}}} \sum_{e}^{\text{accepted}} \frac{P_a(\cos\theta_e) P_b(\cos\psi_e) P_c(\phi_e)}{S(\vec{\Omega}_e | t_e, q_e, \vec{\lambda})}. \tag{5.52}$$

[6] These should not be confused with the orthogonal polynomials $P_l(\cos\theta)$ used to define spherical harmonics, which differ by a normalization factor.

It should be noted that when plotting the (normalized) PDF $S_{\text{obs}} = \frac{S(\vec{\Omega})\times\varepsilon(\vec{\Omega})}{\int S(\vec{\Omega})\times\varepsilon(\vec{\Omega})\mathrm{d}\vec{\Omega}}$, the overall scale of the efficiency is irrelevant, as is the case for fitting. Only when plotting the efficiency separately the absolute values of the normalization functions matter, since they determine the scale of the efficiency function.

The 3D efficiency can now be described by using the determined efficiency coefficients (5.52) in (5.50) and can be multiplied with the PDF. Note that the Legendre polynomials are defined in the range $[-1, 1]$. In case the range of, say, observable ϕ is $[-\pi, \pi]$, the coefficients are determined as

$$e_{abc} \equiv \frac{1}{N_{\text{gen}}} \sum_{e}^{\text{accepted}} e^e_{abc} = \frac{1}{N_{\text{gen}}} \sum_{e}^{\text{accepted}} \frac{P_a(\cos\theta_e)P_b(\cos\psi_e)P_c(\phi_e/\pi)/\pi}{S(\vec{\Omega}_e|t_e, q_e, \vec{\lambda})}. \tag{5.53}$$

The sample of MC events should be large enough and the summation in (5.50) sufficiently long such that it is a good approximation.

To propagate the uncertainty due to the finite size of the MC sample on which the coefficients are determined, the width of the distribution of the sample of per-event coefficient e_{abc} can be used[7]. This width can be estimated, similar to (5.28), as

$$\sigma_{abc}(e_{abc}) = \sqrt{E(e^e_{abc} \times e^e_{abc}) - E(e^e_{abc}) \times E(e^e_{abc})}. \tag{5.54}$$

Using these estimated uncertainties, insignificant terms in the efficiency function can be removed, such that only the significant terms are left.

Furthermore, it should be noted that the implicit series of summation in (5.50) is infinitely long. Depending on the shape of the efficiency, a different order of the parametrization to describe the efficiency well enough, is required. The required order of parameterization of the decomposition in (5.50) can be determined numerically by comparing the normalization integrals ξ_j as determined by a sum over MC events with the integral when using this description of the efficiency:

$$\xi_j \equiv \int f_j(\vec{\Omega})\varepsilon(\vec{\Omega})\mathrm{d}\vec{\Omega} = \int f_j(\vec{\Omega})e_{abc}P_a(\cos\theta)P_b(\cos\psi)P_c(\phi)\mathrm{d}\vec{\Omega}. \tag{5.55}$$

Compared to the method of normalization weights, this method has similar advantages. It also allows to numerically check factorization of the efficiency $\varepsilon(t, \vec{\Omega}) = \varepsilon(t)\varepsilon(\vec{\Omega})$, by checking the proper time dependent behavior of the e_{abc} coefficients. Furthermore, the angular resolution can be corrected for in a similar way. And, just like the angular efficiency weights, apparent efficiencies are avoided, since the efficiency coefficients are determined conditionally on t, as shown in section 5.5. A disadvantage of the parameterization method is that due to the finiteness of the sum, systematic effects are possibly induced. However, as mentioned above, the weights can be used to see if significant terms in the description of the efficiency are missing, by comparing the integrals ξ_j as estimated by a sum over MC events with the integrals $\int f_j e_{abc} P_a P_b P_c \mathrm{d}\vec{\Omega}$. Also, similar as in the case of the weights,

[7]The errors can be propagated by introducing floating efficiency coefficients e'_{abc} and adding penalty terms to the likelihood. Such a penalty term then is a Gaussian as function of the floating e'_{abc}, with width $\sigma(e_{abc})$ and fixed central value e_{abc}. The coefficient e_{000} is fixed, since it only determines the scale of the efficiency, which is irrelevant when PDF and efficiency are multiplied.

the integral in the denominator can be cached. The construction and inclusion of the 3D efficiency function in the software is described in appendix D.

Finally, a result of the method is shown. Consider the case of a $B_s^0 \to J/\psi\phi$ data sample with a '$B_d^0 \to J/\psi K^*$-like' angular efficiency applied. The plotted PDF in figure 5.7 is a product of the theoretical PDF and the efficiency as determined with the polynomial method, with polynomials up to second order. It can clearly be seen that the curve follows

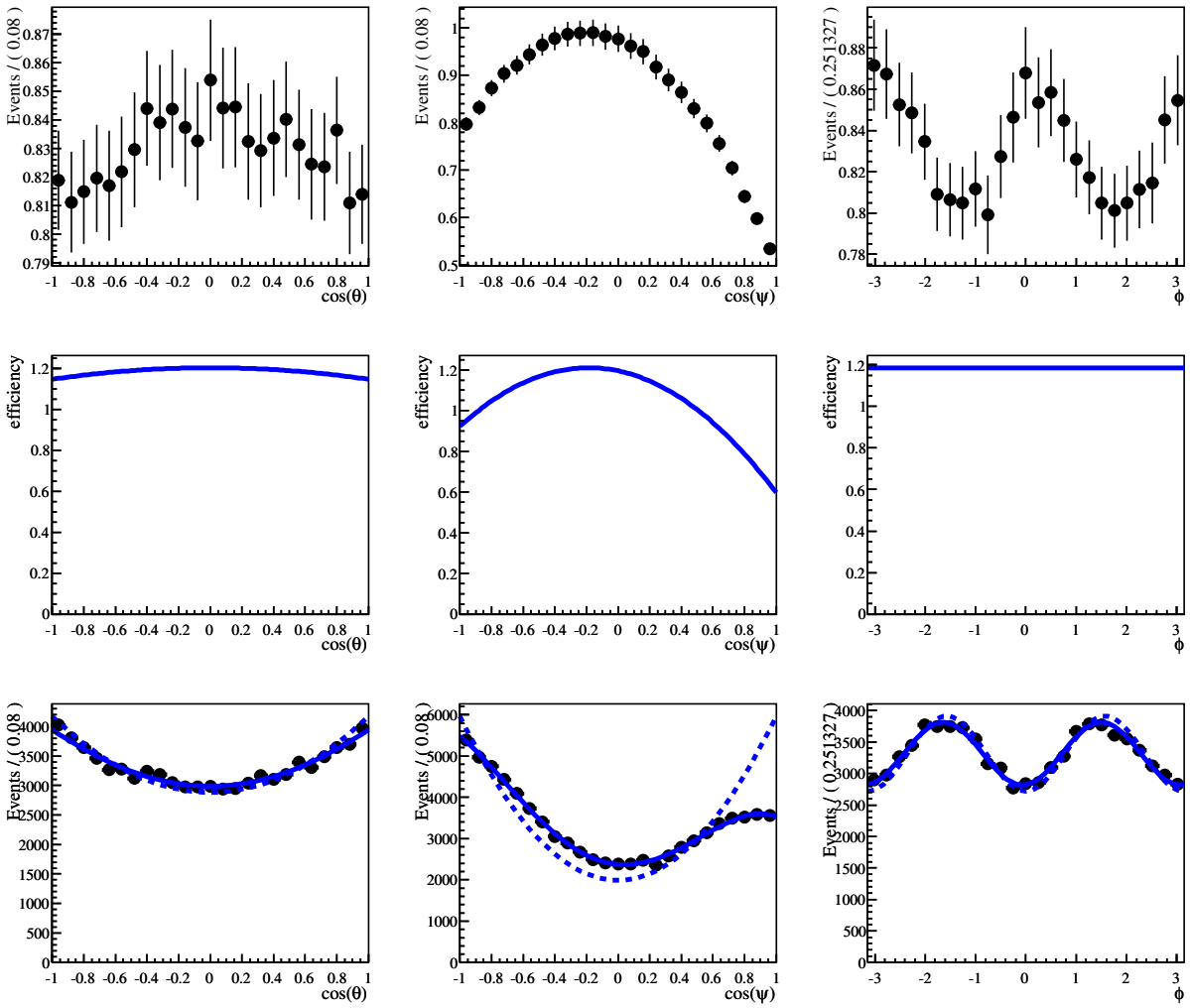


Figure 5.7: *From top to bottom: the toy efficiency data, the determined efficiency functions, and the observed distributions for $B_s^0 \to J/\psi\phi$ toy MC with an efficiency applied. A '$B_d^0 \to J/\psi K^*$-like' efficiency is chosen, since it varies more than a '$B_s^0 \to J/\psi\phi$-like' efficiency. The true efficiency is applied in the $\cos\psi$ observable only; the efficiencies in all other observables (top) are therefore apparent efficiencies. It can be seen that the curves (middle) determine the correct true efficiency. At the bottom the theoretical distribution is multiplied with the angular efficiency function as determined with the polynomials method. The dashed line denotes the uncorrected PDF. The PDF which takes into account the angular efficiency correctly describes the data, in particular in the inefficient region $\cos\psi \to 1$.*

5.7.2 Background PDF decomposed in polynomials

In addition to the signal efficiency, also the 3D angular distribution of the background can be decomposed in polynomials. The background PDF then reads

$$B(\vec{\Omega}) = \frac{\sum_{abc} b_{abc} P_a(\cos\theta) P_b(\cos\psi) P_c(\phi/\pi)}{\int\int\int \sum_{abc} b_{abc} P_a(\cos\theta) P_b(\cos\psi) P_c(\phi/\pi) \mathrm{d}\cos\theta \mathrm{d}\cos\psi \mathrm{d}\phi} \tag{5.56}$$

Using the same arguments as above, the coefficients can be determined, from the angular distribution of the events selected by the sideband mass regions, as

$$b_{abc} \simeq \frac{1}{N} \sum_e P_a(\cos\theta_e) P_b(\cos\psi_e) P_c(\phi_e/\pi). \tag{5.57}$$

In the same way as shown for the efficiency coefficients in (5.54), the widths of the background coefficients can be determined. Using these uncertainties, the irrelevant terms can be removed, and the significant terms can be used to describe the background. Both the events in the sidebands and the events of the signal region are used in a fit[8], in which the uncertainties due to the finite size of the sidebands are propagated to the other parameters. The construction of the 3D background PDF in the software is described in appendix D.

Finally, in figure 5.8 the result of this method to a 3D toy MC distribution is shown. The significant terms of polynomials up to fourth order are used in the fit. As can be seen in the plot, the generated MC and the determined PDF agree.

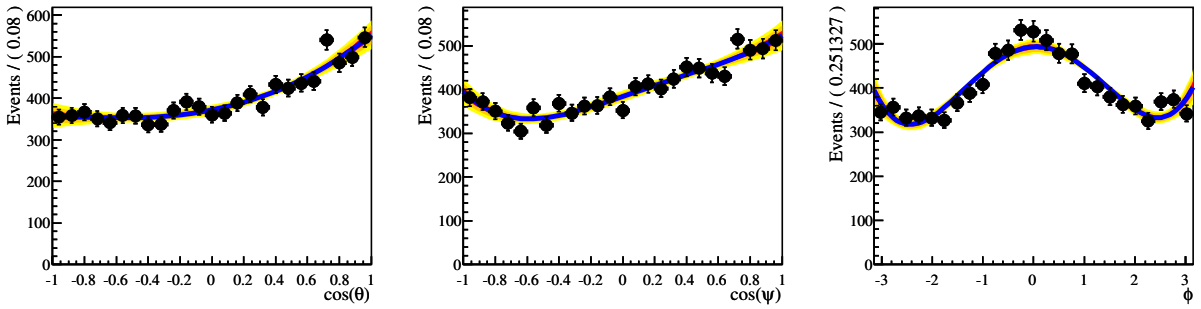


Figure 5.8: *Projections of a 3D background distribution and its description on the angular observables. The PDF is parameterized in terms of polynomials up to fourth order. The (1,2,3) sigma error bands are shown in the plot.*

5.8 Determination of the mistag fraction

Besides the resolutions and efficiencies in the angular and proper time observables, also the dilution of the tagging observables should be considered. As explained in chapter 2, the flavor of the produced B_s meson must be determined to measure ϕ_s with maximal precision.

[8]Fixing b_{000}, which cancels in the normalization and is irrelevant.

In case a fraction ε of the B_s and $\bar{B}_s$ mesons is tagged, and a fraction w of the tagged B_s mesons are wrongly tagged (or 'mistagged') as $\bar{B}_s$ mesons, and vice versa[9], then (2.73) and (2.74) change such that the tag dependent terms describing the tagged sample change as

$$\begin{aligned} \sin\phi_s \sin\Delta m_s t &\rightarrow D \sin\phi_s \sin\Delta m_s t, \\ \sin(\delta_\perp - \delta_i)\cos\Delta m_s t &\rightarrow D \sin(\delta_\perp - \delta_i)\cos\Delta m_s t, \end{aligned} \tag{5.58}$$

where D is given by $D = 1 - 2w$. From these expressions, and from (2.73) and (2.74), three observations can be made [20].

1. The terms proportional to $\sin\phi_s \sin\Delta m_s t$, are multiplied by D. Hence, B_s mesons wrongly tagged to be $\bar{B}_s$ mesons, and vice versa, in leading order, degrade the sensitivity to ϕ_s linearly with D as the mistag fraction increases. The mistag fraction dilutes the CP-asymmetry in the same way as $|A_\perp|^2$.

2. An untagged sample ($\varepsilon = 0$ or $w = 50\%$) still has sensitivity to ϕ_s. Although the terms in $A_{1,2,3}$ are proportional to $D\sin\phi_s$ and hence are in that case insensitive to ϕ_s, the interference terms keep sensitivity to ϕ_s. This sensitivity is, however, small: either second order in ϕ_s for small values of ϕ_s (due to the $\cos\phi_s$ terms), or proportional to $\sinh\frac{\Delta\Gamma_s t}{2}$.

3. Equivalently, the interference terms also introduce a sensitivity to the mistag parameter w. As explained in [30], it is possible to extract both the physics parameters ϕ_s, the strong phases *and* the mistag fraction w from data only, without external input of tagging information, using $B_s^0 \rightarrow J/\psi\phi$ only.

Therefore, for maximal sensitivity to ϕ_s determining the mistag fraction is essential, and it is most robust to calibrate the mistag using an other channel.

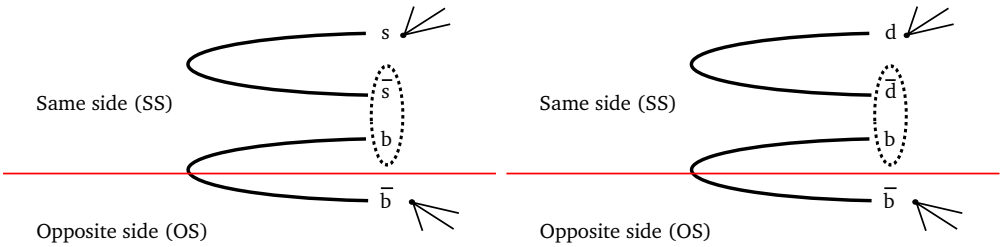


Figure 5.9: *Same side and opposite side tagging for B_s mesons (left) and B_d mesons (right).*

Calibration of flavor tagging is performed in two ways: via opposite side (OS) and same side (SS) tagging. Both methods use the anti-correlation of the charges of the signal quarks and the spectator quarks produced in the pp collision, as is illustrated in figure 5.9. In the

[9] In case a fraction ε of the B_s mesons (and a fraction $\bar{\varepsilon}$ of the $\bar{B}_s$) is tagged, a fraction w ($\bar{w}$) of the tagged B_s ($\bar{B}_s$) mesons are wrongly tagged as $\bar{B}_s$ (B_s) mesons, and a fraction a of the B_s mesons is produced more than $\bar{B}_s$ mesons, then the dilution of the sample of tagged B_s mesons becomes $D = \frac{(1+a)\varepsilon(1-w)-(1-a)\bar{\varepsilon}\bar{w}}{(1+a)\varepsilon(1-w)+(1-a)\bar{\varepsilon}\bar{w}}$, and the dilution of the sample of tagged $\bar{B}_s$ mesons $\bar{D} = \frac{(1-a)\bar{\varepsilon}(1-\bar{w})-(1+a)\varepsilon w}{(1-a)\bar{\varepsilon}(1-\bar{w})+(1-a)\varepsilon w}$ [52].

first method the charge of the decay products of the meson formed by the spectator b quark are used to determine the flavor of the signal B_s^0 meson; in the second method the charge of the meson formed by the spectator s quark is used. Furthermore, the correlation in phase space is used to determine which mesons originate from the same quark pair. Since the B_d^0, as opposed to the B_s^0, contains a d quark instead of an s quark, which forms a lighter meson, a sample of $B_d^0 \to J/\psi K^*$ events is only used to control the OS tagging of $B_s^0 \to J/\psi\phi$.

Since $B_d^0 \to J/\psi K^*(K^* \to K^+\pi^-)$ has a flavor specific final state, this channel is self-tagging, and therefore the OS mistag fraction can be determined from its mixing amplitude. Using this measurement of $w_{\rm OS}$, the dilution of the CP asymmetry in $B_s^0 \to J/\psi\phi$ can be estimated. The calibration of the OS mistag fraction is illustrated in figure 5.10, where the tagged PDFs describing the $B_s^0 \to J/\psi\phi$ decay and the $B_d^0 \to J/\psi K^*$ decay are shown, for different values of $w_{\rm OS}$. .

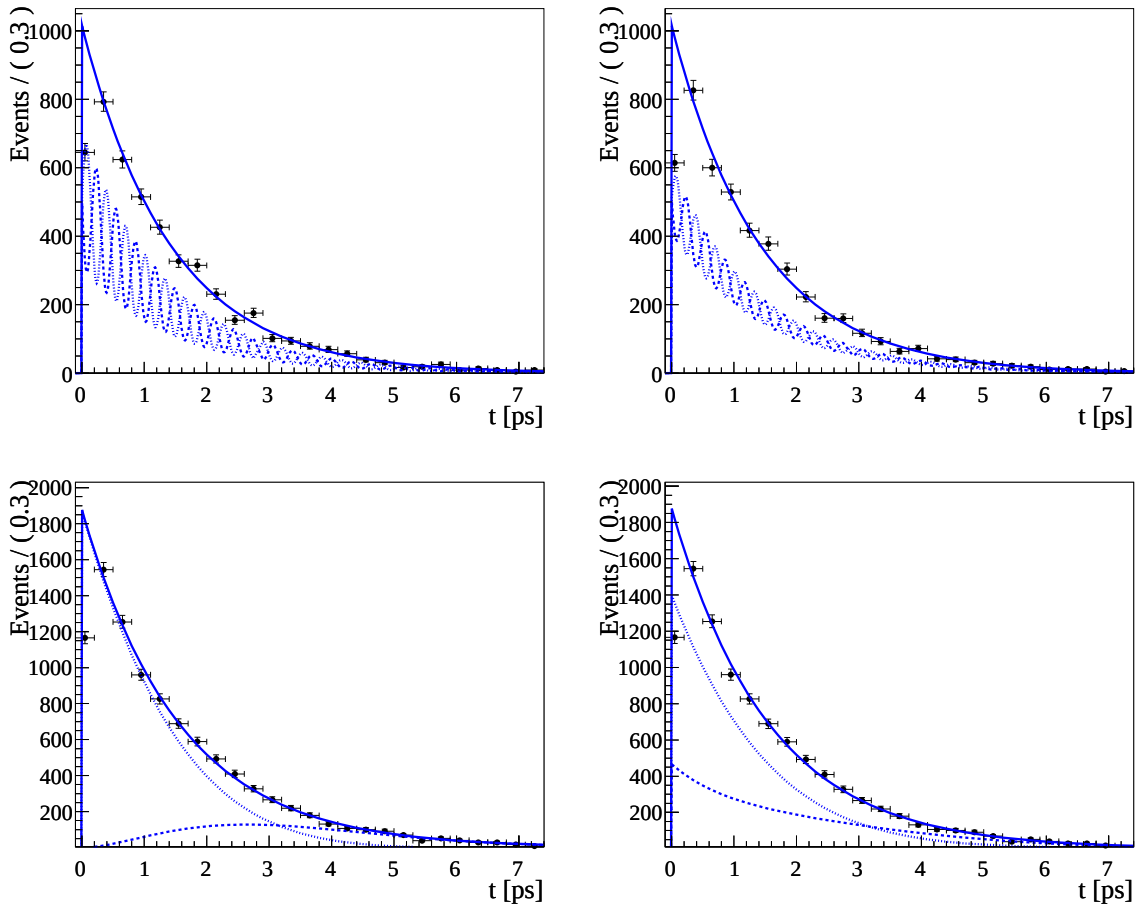


Figure 5.10: *The PDFs for samples of B_s and $\bar{B}_s$ mesons decaying to a $J/\psi\phi$ final state (up), and the PDFs for samples of mixed and unmixed events in the $B_d^0 \to J/\psi K^*$ channel (down). Left the mistag fraction $w_{OS} = 0$; right the mistag fraction $w_{OS} = 0.25$. From the dilution of the mixing oscillation in the $B_d^0 \to J/\psi K^*$ sample the OS mistag fraction of the $B_s^0 \to J/\psi\phi$ sample can be calibrated.*

A possible way to implement this is by performing a simultaneous fit. In this case the $B_s^0 \to J/\psi\phi$ and the $B_d^0 \to J/\psi K^*$ data sample are merged, and each event is labelled according to the corresponding decay. Then the combined PDF for $B_s^0 \to J/\psi\phi$ and $B_d^0 \to J/\psi K^*$ is fitted to this data sample, while having the opposite side mistag fraction $w_{\rm OS}$ as fit parameter in common. The PDFs P_i for each channel are given by

$$\begin{aligned} P_{B_d^0 \to J/\psi K^*} &= P_d(O_d|\lambda_d, w_{\rm OS}), \\ P_{B_s^0 \to J/\psi\phi} &= P_s(O_s|\lambda_s, w_{\rm OS}). \end{aligned} \tag{5.59}$$

Here O_i are the observables for each channel: for example the proper time, the angular observables and the flavour tag category. λ_i are the set of physics parameters for each channel: for example $\tau_i, \Delta m_i, \Delta\Gamma_i/\Gamma_i$ and the polarization amplitudes. Sharing the mistag fraction as fit parameter, w can be calibrated on the $B_d^0 \to J/\psi K^*$ decay [53], and its statistical uncertainty is propagated by construction to the parameters of interest in the $B_s^0 \to J/\psi\phi$ PDF.

The same method can be used to extract the 'same side' mistag fraction from the control channel $B_s \to D_s\pi$ [53] using the SS tag and propagating the statistical uncertainties due to the finite size of this control sample.

Division of the mistag fraction per tagging category

The tagging performance can be improved, as shown in next chapter, by splitting the event sample into sub-samples with different tagging performance[10]. In that case the mistag fractions w are compared per tagging category, which is provided by an analysis software tool, based on properties of the event. This means that for every category i a mistag fraction w_i is introduced, shared between the PDFs for the $B_d^0 \to J/\psi K^*$ and the $B_s^0 \to J/\psi\phi$ decay. This decreases the size of possible systematic effects, since the number of events per category are allowed to differ, and improves the sensitivity to ϕ_s.

As before, the $B_s^0 \to J/\psi\phi$ and $B_d^0 \to J/\psi K^*$ data samples are merged. In addition, each event carries information about the tagging category. The data set has then the contents as shown in table 5.6. The PDFs P_d and P_s are fitted to the one combined sample, with the appropriate PDF for $B_s^0 \to J/\psi\phi$ and $B_d^0 \to J/\psi K^*$ events, using the corresponding value of w_i, which are shared between the two channels. By performing this simultaneous fit, the statistical uncertainties in w_i, as retrieved from the control channel $B_d^0 \to J/\psi K^*$, are propagated to the parameters of $P_{B_s^0 \to J/\psi\phi}$, in particular to ϕ_s.

5.9 Parameterization of the strong phases

There are three issues that lead to potential difficulties related to the fit of the (strong) phases in the angular analyses of the pseudo scalar to vector vector decays (e.g. $B_d^0 \to J/\psi K^*$, $B_s^0 \to J/\psi\phi$), which can lead to unstable fits or bad estimates of the physics parameters.

[10] It should be noted that the number of events per mistag category can become small, which means point 3 above can become no longer valid, and a calibration of the mistag fraction can become not only preferable, but necessary.

	t	$\vec{\Omega}$	tag. flavor	rec. mode	w cat.
$B_s^0 \to J/\psi\phi$	t	$\vec{\Omega}$	$B_s, \bar{B}_s$	$J/\psi\phi$	i
$B_d^0 \to J/\psi K^*$	t	$\vec{\Omega}$	$B_d, \bar{B}_d$	$J/\psi K^*, J/\psi \bar{K}^*$	i

Table 5.6: *The observables of the merged data set of $B_s^0 \to J/\psi\phi$ and $B_d^0 \to J/\psi K^*$ events. Every event is labelled to which decay it belongs. The PDFs P_d and P_s are fitted simultaneously, sharing the mistag fractions w_i per mistag category i.*

1. When the absolute value of a complex number goes to zero the value of the phase is undefined.

2. When the PDF as a function of a parameter is not invertible (i.e. the first derivative of the PDF as a function of the parameter is zero for at least one value), multiple solutions can appear.

3. When background is included, the observable amplitude of an oscillation can become larger than one, as a consequence of correlations with other parameters[11].

Since for $B \to J/\psi V$ the sizes of the polarization amplitudes are significantly larger than zero, only the latter two problems are likely to appear. When factorization holds (i.e. strong interactions between the quark constituents of the vector mesons are negligible) the strong phases δ_i are expected to be zero. This means that the values of $\cos(\delta_i)$ are close to the physical limit of one. However, the experimental data may indicate an observation of a larger value, causing instabilities in the fit. How problematic these instabilities will be depends on the true values of the strong phases of $B_s^0 \to J/\psi\phi$, which are not known[12].

5.9.1 Examples of the problem

In figure 5.11 and 5.12 a manifestation of the above mentioned problems is shown. The first problem, in figure 5.11, is that the pull distribution of $\delta_\parallel$ is significantly wider than 1, implying that the error is underestimated. The second problem, in figure 5.12, is that the behavior of the fit becomes unreliable if the parameter approaches the boundary values. A necessary constraint on the range of the parameter causes boundary effects. This is due to the fact that the 'observable parameter' $\cos(\delta_i)$ is not cyclic, whereas the phase δ_i is.

In figure 5.13 the cause of the underestimated uncertainty of $\delta_\parallel$ is shown. The fit distribution of $\delta_\parallel$ is not gaussian. The fit values of both phases are close to a value of $+\pi$, the region where the fit can become unstable. The shown fit results appear when the fitting program is given the correct input value as start value of the fit. In reality a stable fit can become even more problematic to achieve, as the true values of the strong phases are not known.

[11] An example is the measurement of Δm_s [55], in which the mixing amplitude has been estimated to be larger than 100%.

[12] Although from theoretical considerations these can be expected to be similar to those of $B_d^0 \to J/\psi K^*$.

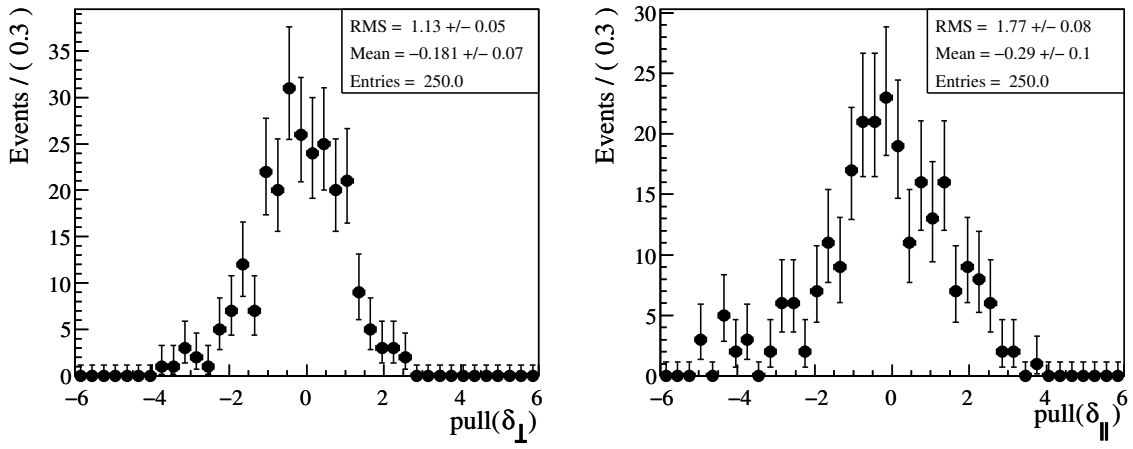


Figure 5.11: *Typical pulls of the fit results of the strong phases for $B_s^0 \to J/\psi\phi$ toy data. In (a) one can see the pull distribution for $\delta_\perp$, in (b) the pull distribution for $\delta_\parallel$. The error of $\delta_\parallel$ is clearly underestimated. The input values of the strong phases are $\delta_\perp = 2.91, \delta_\parallel = 3.35$. The differences between the two distributions are due to the differences between the input values of the strong phases and the differences between the terms in the PDF in which these parameters appear.*

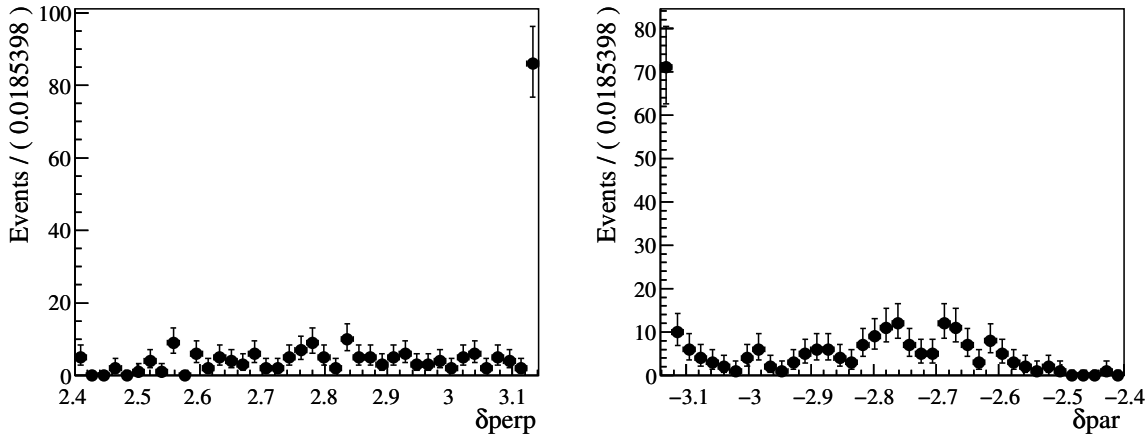


Figure 5.12: *Typical fit results of the strong phases for $B_s^0 \to J/\psi\phi$ toy data. In (a) one can see the distribution for $\delta_\perp$, in (b) the distribution for $\delta_\parallel$. The input values of the strong phases are $\delta_\perp = 2.91, \delta_\parallel = -2.93$.*

5.9.2 An alternative parameterization of the strong phases

To deal with this problem the following simple method is proposed, which amounts to *linearizing* the fit equation. Instead of fitting the sizes and phases of the amplitudes in (2.73), each term is fitted as a separate parameter. This means that the two phase differences

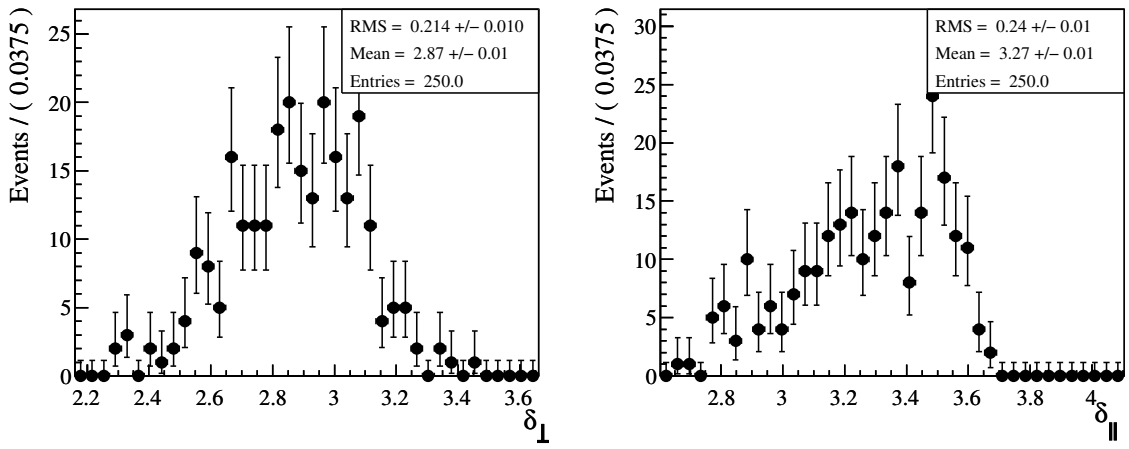


Figure 5.13: *Typical fit results of the strong phases for $B_s^0 \to J/\psi\phi$ toy data. In (a) the distribution for $\delta_\perp$ is shown, in (b) the distribution for $\delta_\parallel$. The distribution for $\delta_\parallel$ is not particularly gaussian. The input values of the strong phases are $\delta_\perp = 2.91, \delta_\parallel = 3.35$.*

between $\delta_\perp$, δ_0, and $\delta_\parallel$ are replaced by five parameters, the 'strong parameters':

$$\begin{aligned} C_{\perp,0} &= \cos(\delta_\perp - \delta_0), \\ S_{\perp,0} &= \sin(\delta_\perp - \delta_0), \\ C_{\parallel,0} &= \cos(\delta_\parallel - \delta_0), \\ C_{\perp,\parallel} &= \cos(\delta_\perp - \delta_\parallel), \\ S_{\perp,\parallel} &= \sin(\delta_\perp - \delta_\parallel). \end{aligned} \tag{5.60}$$

Fitting these strong parameters the fit becomes linear and hence more stable, and the estimate reliable. The values of the strong phases can be derived from the results for $(C_{\perp,0}, S_{\perp,0}, C_{\parallel,0}, C_{\perp,\parallel}, S_{\perp,\parallel})$, provided that correlations are correctly accounted for. For the main purpose of the current analysis, namely stabilizing the fit and determining the value of the weak phase ϕ_s, this parameterization suffices. To validate this method a toy experiment is performed, fitting for the polarization amplitudes $|A_i|^2$ and for the strong parameters $(C_{\perp,0}, S_{\perp,0}, C_{\parallel,0}, C_{\perp,\parallel}, S_{\perp,\parallel})$.

In figure 5.14 it is checked that indeed the parameters in the new basis have more Gaussian errors. It should be noted that a significant fraction have unphysical values: although the absolute value of a (co)sine can never become larger than 1, the absolute values of the fit parameters *do* become larger than 1. This fit issue is exactly the origin of the problem of the fit when determining the strong phases directly.

Furthermore, in [48] it is shown that with this parameterization:

1. The pulls of the strong parameters behave correctly.
2. The estimates of the other physics parameters ($|A_i|^2$, ϕ_s, $\Delta\Gamma_s$, etc.) are not biased.
3. The proposed parameterization does not alter the sensitivity to any of the physics parameters.

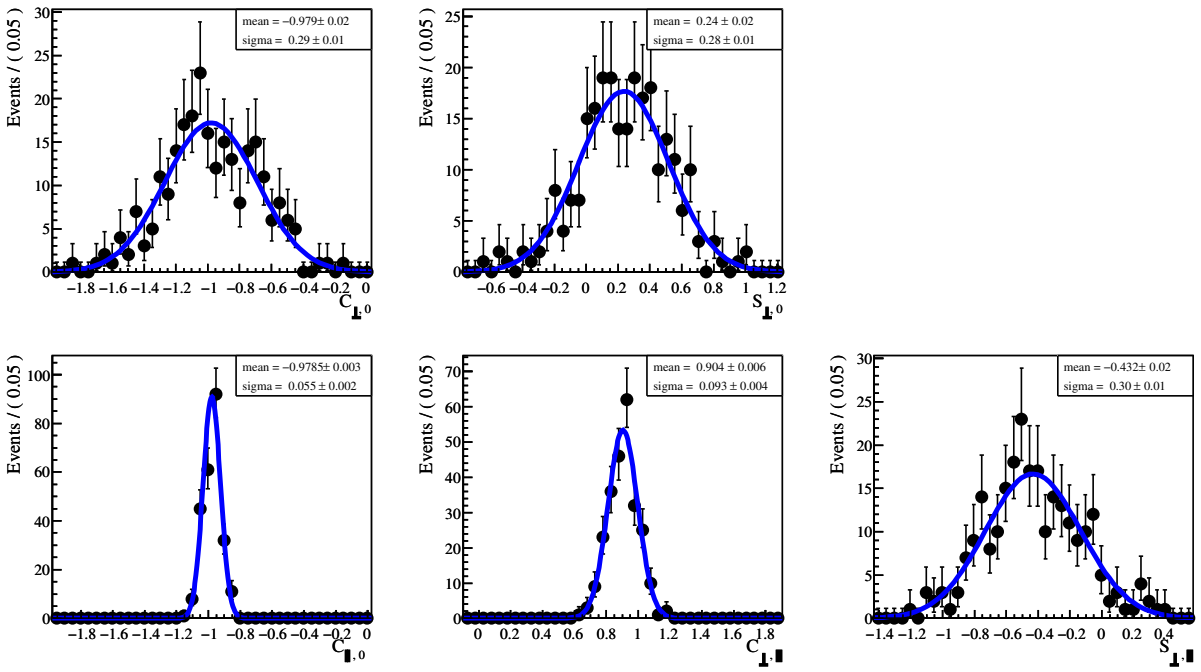


Figure 5.14: *Distributions of the fit results of the strong parameters, for $B_s^0 \to J/\psi\phi$ signal only. Contrary to the situation in which the phases are fitted directly, these distributions* are *Gaussian distributed. The input values are the central values of the x-axes: those corresponding to values of the strong phases $\delta_\perp = 2.91, \delta_\parallel = 3.35, \delta_0 = 0$. To demonstrate the difference in width of the individual parameters the plotted scale is the same for all parameters.*

5.9.3 Retrieval of the strong phases from the alternative parameterization

The strong phases themselves are extracted from the fit results $(C_{i,j}^{\rm fit}, S_{i,j}^{\rm fit})$ shown in figure 5.14. This is done by writing the strong parameters in terms of the original phase δ_i and subsequently minimizing the $\chi^2(\delta_i)$ with respect to δ_i, for the results as found from the parameter fits of the toy studies.

Introducing the residuals $\mathbf{R}$, with

$$\begin{aligned}
\mathbf{R}_1 &\equiv C_{\perp,0}^{\rm fit} - \cos(\delta_\perp - \delta_0), \\
\mathbf{R}_2 &\equiv S_{\perp,0}^{\rm fit} - \sin(\delta_\perp - \delta_0), \\
\mathbf{R}_3 &\equiv C_{\parallel,0}^{\rm fit} - \cos(\delta_\parallel - \delta_0), \\
\mathbf{R}_4 &\equiv C_{\perp,\parallel}^{\rm fit} - \cos(\delta_\perp - \delta_\parallel), \\
\mathbf{R}_5 &\equiv S_{\perp,\parallel}^{\rm fit} - \sin(\delta_\perp - \delta_\parallel).
\end{aligned} \tag{5.61}$$

Then, fixing δ_0 to zero, $\chi^2(\delta_\perp, \delta_\parallel)$ may be written as

$$\chi^2(\delta_\perp, \delta_\parallel) = \mathbf{R}^T V^{-1} \mathbf{R}, \tag{5.62}$$

with V the covariance matrix of the strong parameters obtained in the fitting procedure. The $\chi^2(\delta_i)$ contain all the information needed to extract the strong phases.

The maxima and likelihood curves of the strong phases, using the fitted mean and width of the average fit results of the strong parameters, can be found in figure 5.15 and 5.16. In figure 5.15 the shape of the χ^2 is shown for the central values and errors returned by the fits of the toy studies. In figure 5.16 the shape of the $\chi^2(\delta_i)$ is shown at a scale of $(0, 10)$. The non-Gaussian fit distribution of $\delta_\parallel$ translates in an asymmetric $\chi^2(\delta_\parallel)$. It has been shown in [48] that this method does not lead to a decrease in sensitivity to any of the physics parameters.

Finally, in case similar difficulties appear for ϕ_s, these can be solved by the similar parameterization $S = \sin\phi_s$, $C = \cos\phi_s$.

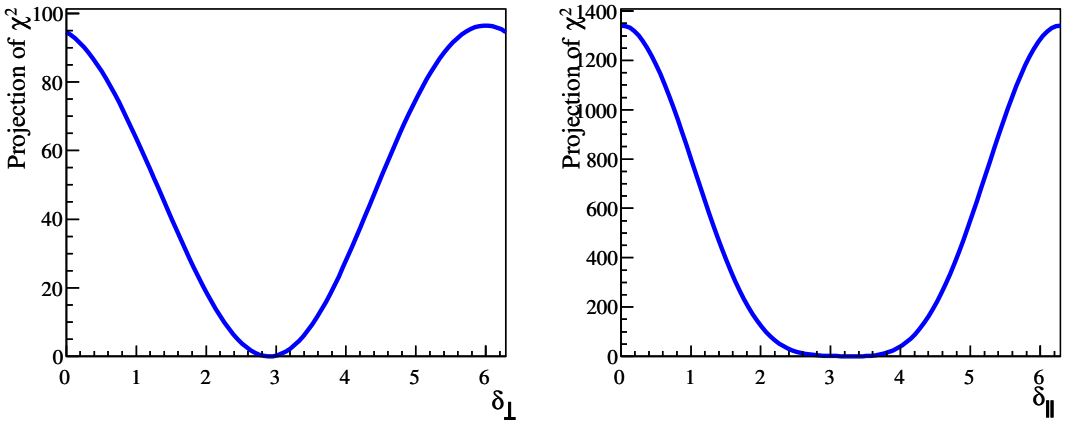


Figure 5.15: *Example of $\chi^2(\delta_i)$. It should be noted that the scale on which $\chi^2(\delta_1)$ are plotted is large.*

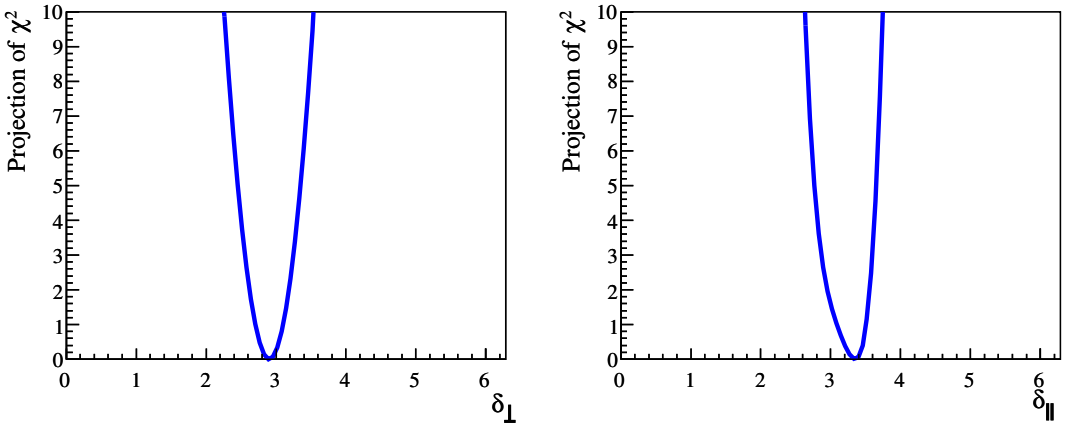


Figure 5.16: *The $\chi^2(\delta_i)$ on a scale of $(0, 10)$. The maximization returns $\delta_\perp = 2.90^{+0.22}_{-0.27}$ and $\delta_\parallel = 3.35^{+0.17}_{-0.40}$. The parabolic errors are 0.23 and 0.22 respectively.*

5.10 Conclusions

The LHCb detector acceptance, the reconstruction procedure, and the selection criteria induce inefficiencies, resolutions, mistagging, and backgrounds to the theoretical distribution of the $B_s^0 \to J/\psi\phi$ decay. To correctly determine the physics parameters, these effects in all observables need to be included in a fit.

In this chapter, a method is described that implements a likelihood maximization, correcting for an angular efficiency function, without knowing or parameterizing its exact shape. This is done by the normalization numbers ξ_j, which are determined from the accepted events in an MC simulation and which weight the efficiency for each separate angle-dependent function. For signal events only (i.e. if no background is present), the full information of the angular efficiency is contained in these numbers.

The advantages of the method of normalization weights are the following. Since the ξ_j depend on t, q, the factorization of the angular efficiency can be studied numerically. Furthermore, the ξ_j can be used to control systematic effects due to parameterization of the efficiency. And to improve its performance, the possibility to cache multi-dimensional integrals is used in the construction of the fit PDF.

In a similar way as the weights are determined, a method to include the angular resolutions are determined. All advantages of the weights are in that case fully used.

Also, a correlation between proper time and angular efficiencies have been discussed. It is shown that proper time efficiencies can induce apparent angular efficiencies, and vice versa, due to the correlations between the observables in the physics distribution. A description of these apparent proper time efficiencies has been given. Solutions to avoid double corrections are given and it has been shown that using the method of weights true efficiencies are automatically determined correctly in the presence of 'apparent' efficiencies.

A situation in which the occurrence of apparent efficiencies is important, is the detached selection event sample. A method has been presented such that data can be estimated in the presence of both a prescaled proper time unbiased and a proper time 'detached' trigger. This fit method allows to determine both the proper time resolution and the relative proper time efficiency, using the prescaled sample, and to propagate the uncertainties to the other parameters. This means the background suppressing detached selection can be used for the analysis.

To include the remaining background events into the analysis, the weights can not be used without extra complications, and in general the shape of the angular efficiency can no longer be neglected. Therefore, a method has been described to determine 3D angular efficiencies in terms of polynomials. To benefit from the advantages of the weights method, the coefficients are determined in the same way as the weights, namely as a sum over MC events.

Using this parameterization, factorization of the angular efficiency can still be studied numerically; the performance of calculations when using the PDF for fitting and plotting can still be optimal; angular resolutions can be included; and apparent efficiencies are avoided. Furthermore, systematic effects due to a finite parameterization of the angular efficiency are controlled by a comparison with the weights. Using the method to determine the coefficients, insignificant terms to describe the angular efficiency can be neglected and the statistical fluctuations from the finite MC sample propagated.

Extending this method to describe the 3D angular efficiency, also the 3D angular background is described in terms of polynomials. As for the angular efficiency, insignificant terms can be estimated and ignored.

Subsequently, to correct for dilutions in the tagging observable, a method to use the mistag rates from a control channel are presented. The data samples of the signal channel and a control channel are merged, and the PDFs of both channels are fitted simultaneously. Sharing the mistag fraction as a fit parameter, the mistag fraction is commissioned on the control channel, while the statistical uncertainties from the limited size of this sample is propagated to the physics parameters of the signal channel.

Furthermore, an alternative parameterization of the strong phases, to improve the fit behavior, is described. In case a fit to a finite data set favors unphysical values of the strong phases, which the PDF does not have the freedom to describe, parameters can be biased and errors wrongly estimated. There are advantages in linearizing the problem, by fitting the different terms as separate parameters instead of fitting the physical parameters. The errors of the reparameterized strong phase parameters are gaussian and hence the estimate of other parameters in the fit become more reliable, without loosing sensitivity. Outliers in the fit disappear and the fit becomes more stable, whereas the pulls of the amplitudes remain correct. To extract the strong phases, even in the case data suggests unphysical values, statistical methods can be used. It has been demonstrated how to retrieve the strong phases from the linearized parameters, retrieving the correct minima and the positive and negative errors of the asymmetric χ^2 distribution of the strong phases. In case similar difficulties appear for ϕ_s, these can be solved by the similar parameterization $S = \sin\phi_s$, $C = \cos\phi_s$.

Finally, the usage of control samples is as follows. MC is used to use the described methods to correct for the angular efficiency. To control systematic effects, the method can be checked by applying the same method on the control channel $B_d^0 \to J/\psi K^*$. The angular distribution of the background is determined from the pure background in the B_s mass sidebands. The mistag fraction is determined from a control channel: $B_d^0 \to J/\psi K^*$ is suggested for OS tags, whereas $B_s \to D_s\pi$ is proposed for the SS tags. The proper time resolution of the signal and the proper time efficiency of the detached sample are determined using the prescaled control sample.

Chapter 6

Sensitivity study

In this chapter the expected sensitivity of the LHCb experiment to the physics parameters accessible in the decay of $B_s^0 \to J/\psi\phi$, in particular to the CP-violating phase ϕ_s, is determined. This is done by performing fast parameterized (or 'toy') MC experiments, simulating the expected distortions of the theoretical decay distribution introduced by the reconstruction and selection procedures. To determine the relevant resolution parameters, detection inefficiencies and backgrounds distorting the theoretical signal distributions in the proper time, angular, mass and tagging observables, fully simulated MC events are used.

The LHCb detector is described in chapter 3, and the simulation of its response to traversing particles is performed by GEANT [56]. The signal events in the simulation are generated according to the theoretical distribution described in chapter 2, which is implemented in EvtGen as described in appendix B.2. The B_s mesons used by EvtGen are generated by pythia [57], which is used to simulate the pp collisions. Because the events are retrieved from time-consuming simulation, the statistics of the background events is smaller, and hence the description of the background less accurate, than if real data would have been available.

After the event simulation step, the procedure is the same as for real data. Particles are reconstructed, and only events passing the selection, as described in chapter 4, are used to determine the distortions of the theoretical distribution due to selection and reconstruction. To describe these distortions, the methods described in previous chapter are used. Finally, to determine the expected measurement precision of the physics parameters (ϕ_s, $|A_\perp|$, $|A_0|$, $\delta_\perp$, $\delta_\parallel$, τ_s, $\Delta\Gamma_s$), toy experiments are performed, taking into account these distortions, using the software described in appendix D.

The outline of this chapter is the following. First, the model used to describe the observed distribution after selection and reconstruction is given. Subsequently, the inefficiencies, resolutions and backgrounds of the theoretical signal distribution in all observables, needed to extract the physics parameters, are shown and described. To determine the expected sensitivity of LHCb to the physics parameters (and their correlations), toy MC experiments are performed, including the distortions as determined on fully simulated MC events. Finally, an estimate is made of the required integrated luminosity needed to perform a measurement which is more accurate than the results available from the Tevatron at the time of writing.

6.1 The total PDF

The total observed PDF T, to describe the distribution obtained after selection and reconstruction, can be written in terms of the signal PDF S, and the background PDF B, as

$$T(t,\vec{\Omega},q,m|\vec{\lambda},\vec{\kappa},f_S) = f_S \times S(t,\vec{\Omega},q,m|\vec{\lambda}) + (1-f_S)\times B(t,\vec{\Omega},q,m|\vec{\kappa}). \quad (6.1)$$

The observables are the proper time t, the angular observables $\vec{\Omega}$, the tagging flavor q, and the reconstructed invariant masses m of the mesons. The parameters used to describe the signal PDF are denoted by $\vec{\lambda}$; the parameters used to describe the background PDF are denoted by $\vec{\kappa}$. Finally, f_S is the fraction of signal events. The fraction f_S is defined in the full selection mass windows and is directly related to the purities $S/(S+B)$, as given in chapter 4, via the relative sizes of the mass windows of the signal box.

6.1.1 The signal PDF

The signal PDF is written as the following product:

$$S(t,\vec{\Omega},q,m|\vec{\lambda}) = S(t,\vec{\Omega},q|\vec{\lambda}_p,\vec{\lambda}_t,\vec{\lambda}_\Omega,\vec{\lambda}_q)\times S_m(m|\vec{\lambda}_m). \quad (6.2)$$

Here S_m is the factorizing PDF of the signal mass distribution. Assuming an efficiency as function of the true proper time t' and true angular observables $\vec{\Omega}'$, $\varepsilon(t',\vec{\Omega}')$, independent of the tagging flavor, the rest of the signal PDF, as explained in previous chapter, reads

$$S(t,\vec{\Omega},q|\vec{\lambda}_p,\vec{\lambda}_t,\vec{\lambda}_\Omega,\vec{\lambda}_q) = \frac{A_i(t',q|\vec{\lambda}_{p,q})f_i(\vec{\Omega}')\varepsilon(t',\vec{\Omega}'|\vec{\lambda}_{t,\vec{\Omega}})}{\sum_q \int A_j(t,q|\vec{\lambda}_{p,q})f_j(\vec{\Omega})\varepsilon(t,\vec{\Omega}|\vec{\lambda}_{t,\vec{\Omega}})\mathrm{d}t\mathrm{d}\vec{\Omega}} \otimes R(t,t',\vec{\Omega},\vec{\Omega}'|\vec{\lambda}_{t,\vec{\Omega}}), \quad (6.3)$$

where R is the description of the proper time and angular resolution.

The physics parameters $\vec{\lambda}_p$ are: the CP-violating weak phase ϕ_s, the B_s oscillation frequency Δm_s, the lifetime τ_s, the width difference $\frac{\Delta\Gamma_s}{\Gamma_s}$, and the polarization amplitudes $|A_i|^2$. Depending on the parameterization, either the strong parameters $(C_{\perp,0}, S_{\perp,0}, C_{\parallel,0}, C_{\perp,\parallel}, S_{\perp,\parallel})$, as introduced in chapter 5, or the strong phase differences $(\delta_\perp - \delta_0, \delta_\parallel - \delta_0)$ are used.

The experimental parameters $\vec{\lambda}_t, \vec{\lambda}_\Omega, \vec{\lambda}_q, \vec{\lambda}_m$ describe the distortion of the theoretical signal distribution. They are divided in the parameters for the different observables:

1. λ_t: the parameters used for the description of the proper time resolution and efficiency, e.g. the width of the proper time resolution;

2. λ_Ω: the parameters used for the description of the angular time resolutions and efficiencies, e.g. the coefficients of the decomposition in polynomials;

3. λ_q: the parameters used for the description of the tagging distribution, e.g. the mistag fraction;

4. λ_m: the parameters used for the description of the signal invariant mass distributions of the mesons, e.g. the width of the reconstructed invariant B_s mass distribution.

Before being able to perform sensitivity studies of the physics parameters $\vec{\lambda}_p$, first the parameters $\vec{\lambda}_t, \vec{\lambda}_\Omega, \vec{\lambda}_q, \vec{\lambda}_m$, describing the distortions of the theoretical signal distributions, need to be determined.

6.1.2 The background PDF

The background is described as the product of four factorizing PDFs:

$$B(t, \vec{\Omega}, q, m|\vec{\kappa}) = B_t(t|\vec{\kappa}_t) \times B_\Omega(\vec{\Omega}|\vec{\kappa}_\Omega) \times B_q(q|\vec{\kappa}_q) \times B_m(m|\vec{\kappa}_m). \tag{6.4}$$

Here B_t is the description of the proper time distribution of the background, B_Ω the angular distribution of the background, B_q the tagging behavior of the background, and B_m the mass distribution of the background. $\vec{\kappa}_i$ are the parameters used to describe the background distribution in the given observable and need to be determined to perform sensitivity studies to determine the sensitivity to $\vec{\lambda}_p$.

6.2 Resolutions

Resolution effects, i.e. uncertainties in the reconstruction of observables, decrease the sensitivity to physics parameters due to smearing of the theoretical signal distribution and decrease the ability to distinguish signal and background. Furthermore, resolution effects not taken into account lead, in general, to biases in the estimates of the physics parameters. Therefore, resolution effects should be as small as possible, and the effects must be known and included in the description of the data.

The resolution effects in the proper time, mass, and angular observables are due to the finite momentum and vertex resolutions of the tracking system, described in chapter 3. In this section the resolutions in the observables of the $B_s^0 \to J/\psi\phi$ decay are interpreted and described. Furthermore, after either including them in the description of the data or neglecting them in case the effects are small, the resulting systematic effects on the physics parameters are estimated.

6.2.1 Proper time resolution

The proper time of the B_s meson is reconstructed in the laboratory frame as

$$t = m\frac{d}{p}, \tag{6.5}$$

with m the true PDG mass of the B meson, p the momentum of the reconstructed B, and d the distance between primary and secondary vertex. The decay time is reconstructed using a kinematic fit which constrains the B to originate from the primary vertex. The difference between this reconstructed value t and the (unobservable) true value t' is the proper time resolution. A proper time resolution in the reconstruction of the $B_s^0 \to J/\psi\phi$ decay dilutes the oscillation of the proper time dependent CP asymmetry, and hence decreases the sensitivity to ϕ_s.

In order to not bias the reconstructed proper time, the primary vertices should be reconstructed with the signal tracks excluded. As illustrated in figure 6.1, signal tracks contributing to the primary vertex bias the reconstructed proper time towards smaller values if they are included in the reconstruction of the primary vertex. Therefore, in this thesis, the primary vertex position is recomputed by removing any signal track. After this procedure, the shape of the resolution is symmetric, as shown and discussed in [42].

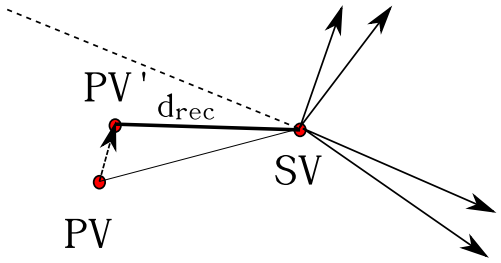


Figure 6.1: *Bias of the reconstructed proper time due to the inclusion of a signal track in the reconstruction of the true primary vertex (PV). If the dotted, extrapolated track originating from the secondary vertex (SV) is included in the reconstruction of the primary vertex, an incorrect primary vertex (PV') is estimated. This leads to an underestimate of the reconstructed decay distance d_{rec} and hence an underestimate of the lifetime.*

To include the uncertainty on the measurement of the proper time on a per-event basis, the per-event calculated error σ_t is used. This error is calculated based on the covariance matrix of the fitted tracks corresponding to the final state of the B-meson decay. The distribution of the per-event calculated proper time errors is shown in figure 6.2(a). Since the width of this distribution is finite, including the resolution on a per-event basis in general increases the sensitivity to physics parameters, as is shown below in a similar manner for the mistag fraction.

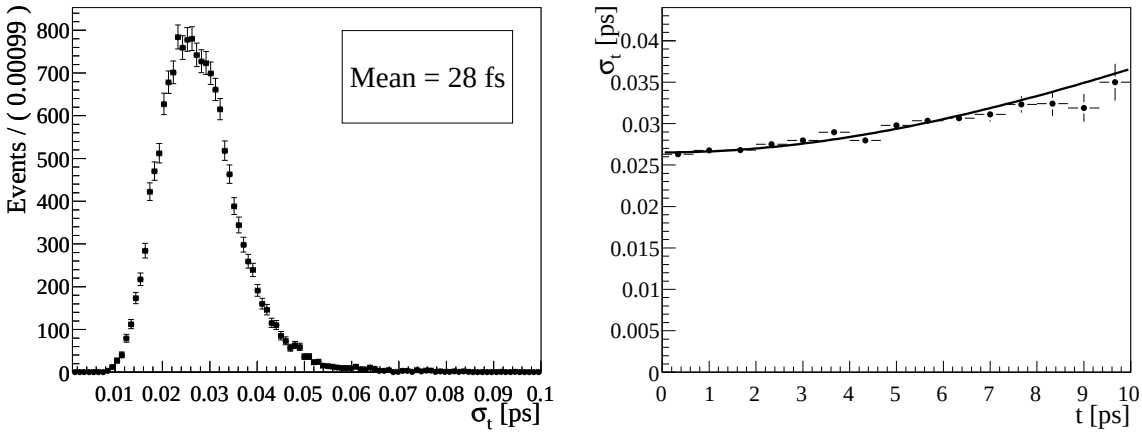


Figure 6.2: *The estimated proper time errors after refitting (left), and the average proper time resolution versus the true time (right). The proper time resolution depends on the decay time; the non-zero resolution at $t = 0$ in the right plot is due to the uncertainty in the estimate of the decay distance.*

The average σ_t as a function of t is shown in figure 6.2(b). To understand the dependence of the resolution as a function of the proper time, the proper time resolution is decomposed in the uncertainty from the estimate of the decay distance d and the estimate of the mo-

mentum p. From the definition of the proper time as a function of the measured d, p (6.5) it follows that the first contribution, which originates from the relationship between t and measured d, is proportional to the uncertainty in the distance d between the two vertices:

$$\sigma^d(t) = \frac{\partial t}{\partial d}\sigma(d) = \frac{m}{p}\sigma(d), \tag{6.6}$$

which is constant as a function of t for given p. The second contribution that constitutes the proper time resolution, which originates from the relationship between t and measured p, is proportional to the lifetime:

$$\sigma^p(t) = \frac{\partial t}{\partial p}\sigma(p) = -\frac{md}{p^2}\sigma(p) = -t\frac{\sigma(p)}{p}. \tag{6.7}$$

Neglecting correlations between the two parameters p and d, the relationship between t and σ_t is the following:

$$\sigma_t(t) = \sigma^d(t) \oplus \sigma^p(t) = \sqrt{(\frac{m\sigma(d)}{p})^2 + (\frac{\sigma(p)}{p})^2 t^2}. \tag{6.8}$$

The observed distributions of the free parameters in this description are shown in figure 6.3: the average momentum of the B_s meson is $\langle p_B \rangle \approx 100$ GeV and the width of the decay distance resolution is $\sigma(d) \approx 150\ \mu m$. Using these values and the PDG B_s mass, the constant contribution to the proper time resolution is expected to be $\frac{m\sigma(d)}{p} \approx 27$ fs. The width of the relative momentum resolution is $\sigma(p)/p \approx 0.3\%$. To test the description in (6.8), these parameters are fitted to the average σ_t as a function of t in figure 6.2(b), and compared to the expected values. Fitting the parameters of the model to the distribution, the following values are found: $\sigma(p_B)/p_B = (0.252 \pm 0.008)\%$ and $\frac{m\sigma(d)}{p} = (26.5 \pm 0.6)$ fs, which is in close agreement with the expectation.

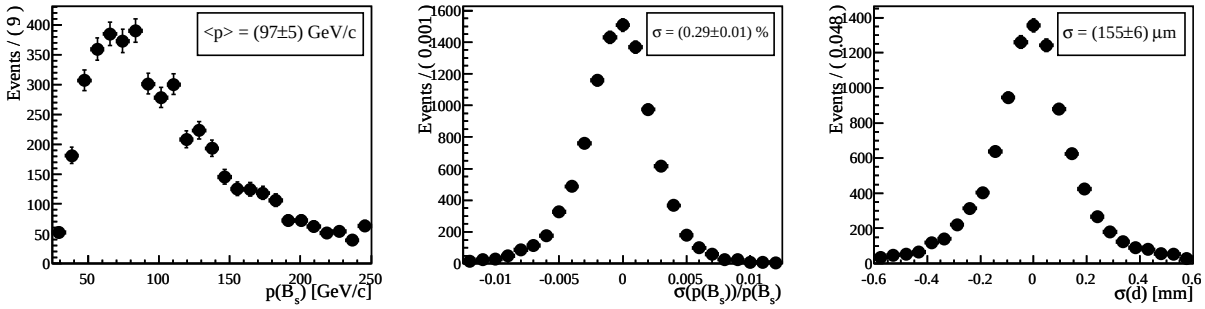


Figure 6.3: *The distributions of the B momentum p_B (left), the relative momentum resolution σ_p/p (center), and the decay distance resolution σ_d (right).*

The behavior of different parameters in (6.8) as a function of the B momentum p_B are shown in figure 6.4. The relative proper time resolution $\sigma(p_B)/p_B$ varies only $\pm 20\%$ in the region $50-250$ GeV. Remarkably, the ratio σ_d/p_B, the contribution in (6.8) which is constant

as a function of t, is also approximately constant as a function of p_B. As a consequence, the proper time resolution is constant as a function of p_B, for the range $p_B > 50$ GeV. For small p_B the proper time resolution increases, due to the nonzero decay distance resolution for zero B momentum.

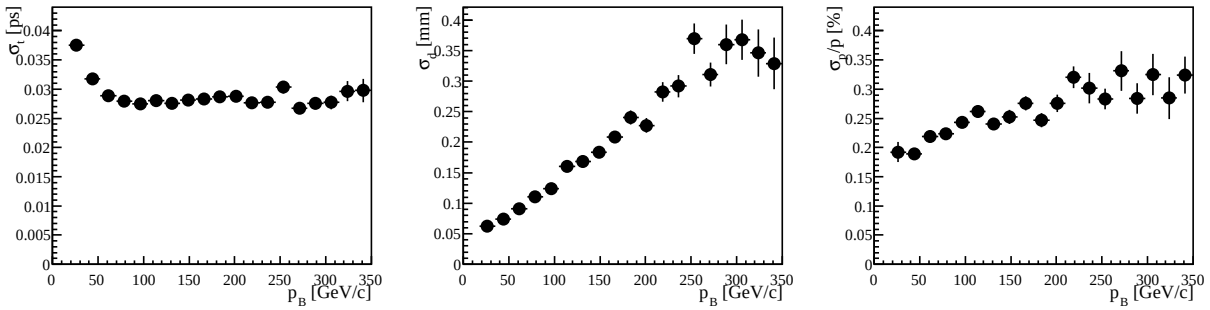


Figure 6.4: *Average resolutions as a function of the B momentum p_B: the proper time resolution σ_t (left), the decay distance resolution σ_d (center), and the relative momentum resolution $\sigma(p_B)/(p_B)$ (right).*

The calculated per-event error is included by writing the proper time resolution in (6.3) as

$$R(t, t'|\lambda_t) = G(t - t'|\sigma_t, C)E(\sigma_t) \tag{6.9}$$

Here $E(\sigma_t)$ is the PDF of the observable σ_t. G is the resolution model[1], which is conditional on σ_t. Per event, the PDF for the observed value of σ_t is used, which is multiplied by a scale factor C. This parameter corrects for an overall underestimate or overestimate of the calculated errors due to e.g. not perfectly aligned tracking detectors.

Using this model, the PDF in figure 6.5 is found with its parameters fitted to the data set. The per-event error distribution is determined from the data set itself, and described by a keys PDF [58]. Fitting for both the scale factor and the lifetime, the lifetime determined ($\tau = 1.47 \pm 0.02$ ps) is consistent with the input value ($\tau = 1.47$ ps). The fit of the scale factor ($C = 1.02 \pm 0.08$) is consistent with no scale factor, and the fitted mean of the resolution model is consistent with zero.

Finally, the fully simulated events and PDF in the region around $t = 0$ for different ranges of σ_t are shown in figure 6.6. It is shown that for increasing values of σ_t the resolution worsens, as expected. The PDF describes the events correctly for ranges of small, mediocre, and large observed values of σ_t.

It is concluded that the description of the proper time resolution can be improved by using the per-event calculated error σ_t. The effects of the usage of per-event errors to describe the data on the precision of the measurement of ϕ_s is discussed in more detail in [59]. The situation in the presence of background is discussed below.

[1]The resolution model can e.g. be a sum of multiple gaussians. Of these one can be chosen fixed with a large width, as a 'garbage term'.

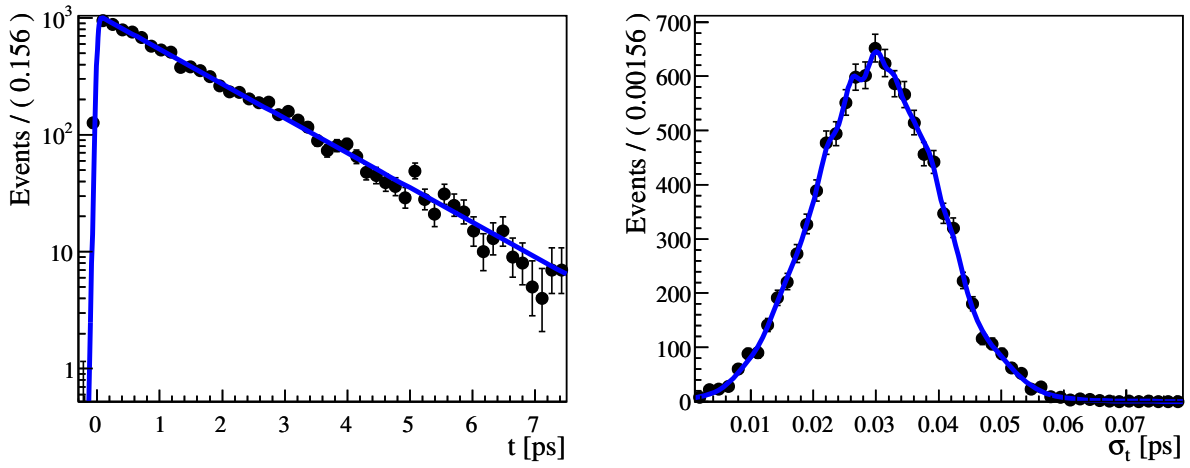


Figure 6.5: *Distribution as function of the proper time t (left) and the per-event error σ_t (right), with overlaid the best fit of a PDF to the data set. The per-event error distribution is described by a keys PDF.*

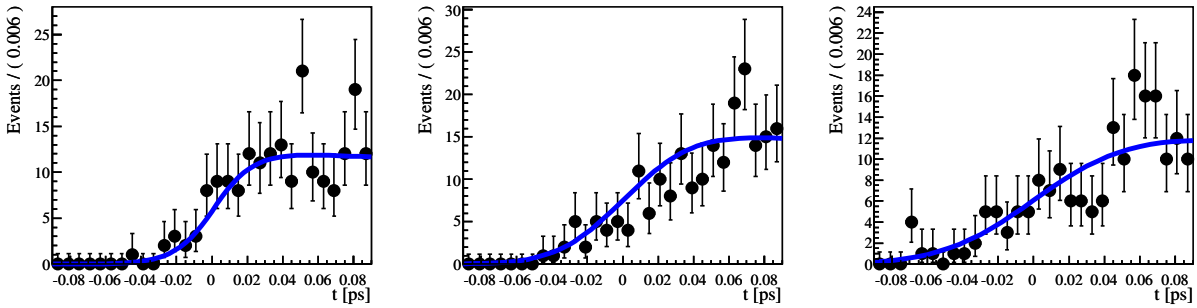


Figure 6.6: *The events and fitted PDF as a function of t for different ranges of the per-event error σ_t. From left to right: $\sigma_t \in [0, 0.025]$ ps, $\sigma_t \in [0.025, 0.035]$ ps, $\sigma_t \in [0.035, 0.100]$ ps. As the per-event error increases, it can be seen that the resolution worsens, as expected.*

6.2.2 Angular resolutions

The definitions of the angular observables $(\cos\psi, \cos\theta, \phi)$ in the rest frame of the mesons are shown in figure 2.13. Differences between the reconstructed and the true values of the observables are caused by imperfections in the tracking[2]. Since the polarization amplitudes A_i are determined from the angular distribution, non-zero angular resolutions lead to a decrease of the sensitivity to these physics parameters.

The angular resolutions for the $B_s^0 \to J/\psi\phi$ decay, as retrieved from fully simulated MC events, are shown in figure 6.7. The systematic effects due to neglecting the angular resolutions are studied in [48]. There, the angular resolutions are approximated by single

[2]Due to boosts and rotations, it is not intuitive to translate tracking resolutions in the lab frame to angular resolutions in the rest frame of the decay particles.

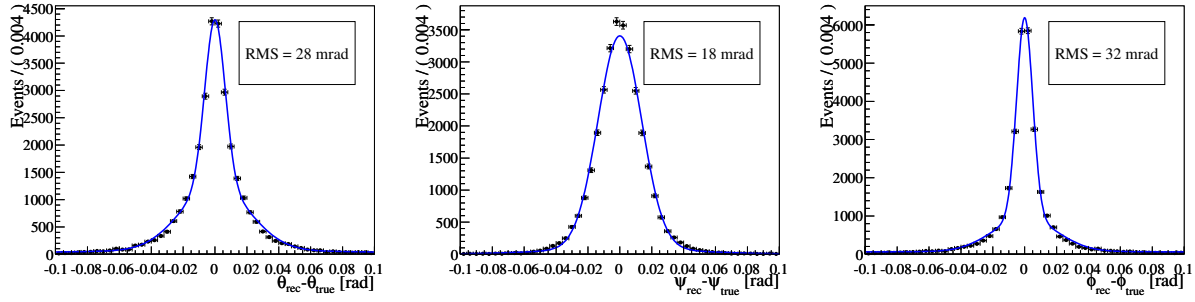


Figure 6.7: *The angular resolutions for $B_s^0 \to J/\psi\phi$ as retrieved from fully simulated MC events.*

Gaussian distributions with widths given by the values in table 6.1, based on fully simulated DC06 events.

θ	ψ	ϕ
29 mrad	19 mrad	28 mrad

Table 6.1: *The width of the Gaussian angular resolutions of the $B_s^0 \to J/\psi\phi$ decay used in [48] to study the systematic effects introduced by neglecting the angular resolutions.*

The angular resolutions in the $B_s^0 \to J/\psi\phi$ channel are worse than those of the $B_d^0 \to J/\psi K^*$ channel. This is due to the fact that the invariant masses of the mesons and the daughter particles are different for the two channels: the sum of the invariant masses of the kaons is closer to the ϕ invariant mass than the sum of the invariant masses of the kaon and the pion is to the mass of the K^*. This causes the opening angle in the lab frame to be smaller for the $B_s^0 \to J/\psi\phi$ channel, and thereby the angular resolutions to be worse.

Using 10 k $B_d^0 \to J/\psi K^*$ signal events, corresponding to ~ 0.04 fb^{-1} of integrated luminosity, no biases have been observed in the estimates of the physics parameters, even if the angular resolutions are assumed up to a factor of three worse than the expected angular resolutions.

Using 10 k $B_s^0 \to J/\psi\phi$ signal events, corresponding to ~ 0.2 fb^{-1} of integrated luminosity, no biases have been observed for the expected resolutions. However, a bias of the estimate of the polarization amplitudes has been observed, when assuming angular resolutions a factor two worse than the expected angular resolutions.

On a large statistics sample of 1 M signal events, assuming the expected angular resolutions, no significant biases are observed for 1M (toy) $B_d^0 \to J/\psi K^*$ events. However, for the $B_s^0 \to J/\psi\phi$ channel a significant bias is observed in the polarization amplitude $|A_\perp|^2$, equal to $(0.23 \pm 0.07) \times 10^{-3}$.

Since the statistical uncertainty of the estimate of $|A_\perp|^2$ for 2.0 fb^{-1} is expected to be 3×10^{-3}, the systematic effect introduced by neglecting the angular resolutions is small for this integrated luminosity, for the angular resolutions as expected from the fully simulated events used in this study.

6.2.3 Signal mass distributions

The width of the reconstructed mass distributions of unstable particles is due to both the detector mass resolution and (in particular for the ϕ meson) due to the intrinsic decay widths. The mass resolution, i.e. the difference between the reconstructed and the true mass, is due to momentum and track angle reconstruction resolutions of the tracking system. Since the observed mass distributions are used to distinguish signal and backgrounds, a finite mass resolution decreases the sensitivity to the physics parameters.

The signal mass distributions after selection and reconstruction is given in figure 6.8. The B_s^0 mass distribution can be described by a gaussian with central value 5369 MeV/c^2 and a width of 16 MeV/c^2; the J/ψ mass distribution has a width of 12 MeV/c^2; the peak of the ϕ mass distribution has a width of 5 MeV/c^2, dominated by its natural width.

It should be noted that no S-wave component has been used in the simulation of the K^+K^- invariant mass distribution, and that the B_s masses have been reconstructed without constraints on the invariant mass of the J/ψ.

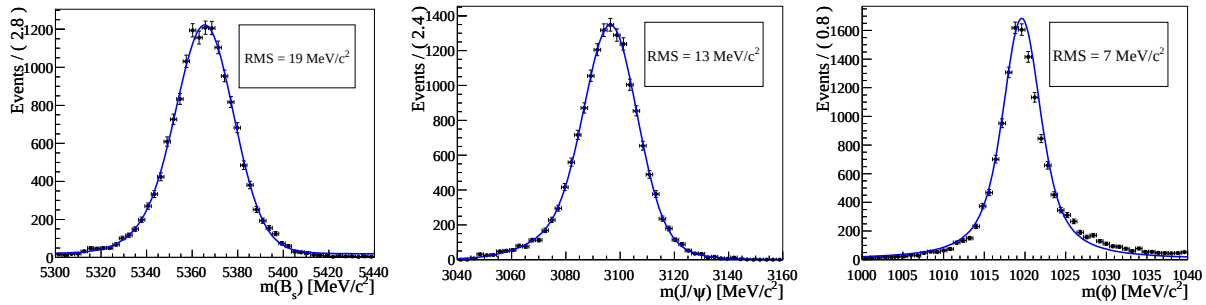


Figure 6.8: *Invariant mass distributions for the mesons of the signal decay, reconstructed without mass constraints.*

6.3 Efficiency functions

The theoretical decay distributions in practice get distorted by non-uniform efficiencies, i.e. a non-constant probability to reconstruct (and select) an event as a function of one of the observables used to analyze the $B_s^0 \to J/\psi\phi$ decay. This leads in general to systematic biases in the estimate of physics parameters when not taken into account. It is therefore important to keep acceptance effects as small as possible, and to include them in the description of the data.

The non-uniform efficiencies in the proper time and angular observables in the $B_s^0 \to J/\psi\phi$ decay are on one hand due to the geometrical shape of the detector and on the other hand due to the selection method. In this section the effects from non-uniform efficiencies are interpreted and, if necessary, included in the fit.

In order to avoid the appearance of apparent efficiencies, described in previous chapter, a special MC sample, in which $\Delta\Gamma = 0$, is used to determine the angular and proper time

efficiencies. Furthermore, since for $\Delta\Gamma = 0$ the angular distribution and the proper time distribution factorize, both can be analyzed separately[3].

6.3.1 Efficiency as function of the proper time

The largest non-uniformity in the efficiency as a function of the proper time originates from the online selection. As explained in chapter 4, events can be selected by two trigger lines: a detached line and a prescaled line. In the prescaled selection there is no explicit cut applied that induces a non-uniform proper efficiency as a function of proper time; in the detached selection, to reduce background, such a cut *is* explicitly applied.

Only events with an online reconstructed proper time larger than a certain value τ_0 are accepted by the detached line. Furthermore, in the sample of detached events, events selected by HLT1 lines in which cuts are applied that induce efficiencies that are non-constant as a function of proper time, are possibly included. Since non-uniform proper time efficiencies bias the estimates of the physics parameters (e.g. the lifetime and the width difference), they need to be taken into account.

The proper time efficiency of the detached sample, including only HLT1 lines with an efficiency which is constant as a function of the proper time, is shown in figure 4.17. The efficiency step in this plot is caused by the cut on the lifetime, $t > \tau_0 = 0.15$ ps, in HLT2. Due to differences between offline and online reconstruction it is not exactly a step function.

The proper time efficiency including HLT1 lines which *do* induce non-trivial proper time efficiencies is shown in figure 6.9. This efficiency is a product of the efficiency introduced by the cut on the lifetime in the exclusive HLT2 selection and the efficiency induced in the HLT1 lines.

The offline selection itself induces an efficiency which is constant as a function of the proper time, as shown in chapter 4. It is constant as a function of t within statistical fluctuations. Furthermore, it has already been shown in the previous section that the lifetime τ has been determined without a systematic bias in a fit to an offline selected event sample. This means the prescaled sample can be used as a control sample, in which no proper time efficiency correction needs to be taken into account. For a large statistics sample this assumption may need to be reconsidered, as shown for a different selection in [20].

Using no correction for the prescaled sample, and determining the relative efficiencies from the data itself, the plots in figure 6.10 and 6.11 are obtained. The plot shows a superimposed fit, on a sample with $\Delta\Gamma = 0$ and ignoring tagging information, using the method to fit for the proper time efficiency in the presence of a prescaled and a detached sample as described in section 5.6. The simultaneous fit to the detached and the prescaled sample are performed for two cases: namely including and excluding muon HLT1 lines in which non-uniform proper time inducing cuts are applied. The results are shown in figure 6.10 and 6.11, respectively. The prescaled control sample is the same for both studies; the number of events in the detached sample is larger for the first study, as more HLT1 lines are included. The results demonstrate that a fit for both the physics parameters and the proper time efficiency is possible.

[3]Because the study of CP violation in this case would serve no useful purpose, CP violation has not been included in the generation of this sample and flavor tagging information is ignored in the study.

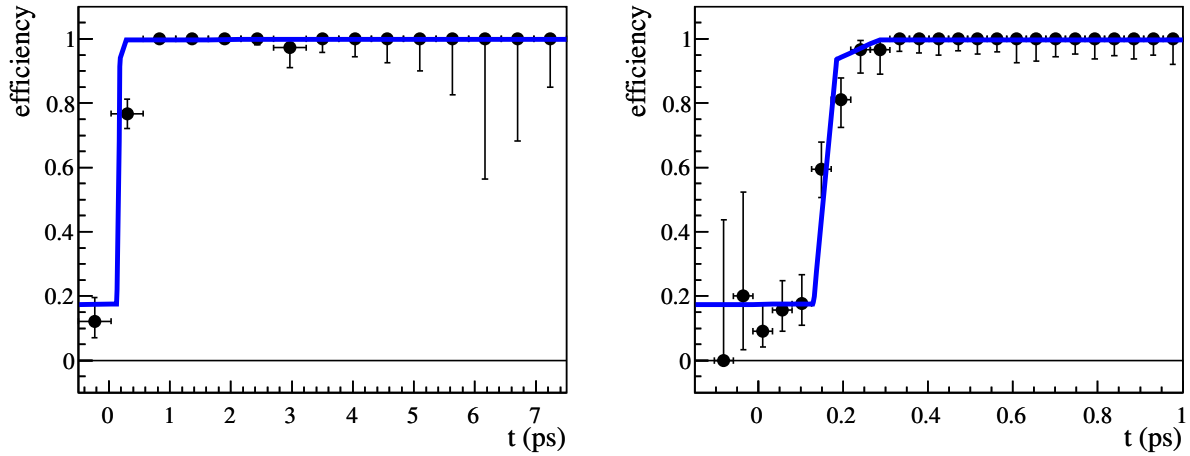


Figure 6.9: *Non-uniform efficiency as a function of the proper time due to the lifetime cut in the HLT2. All muon lines are considered, also those in which cuts are applied that cause a non-uniform proper time efficiency. The data points shown are the number of events selected by both the prescaled and the detached trigger, divided by the total number of events selected by the prescaled trigger. All events are offline selected. The right plot shows a zoom in of the decay time region around the lifetime cut* $t > 0.15$ *ps.*

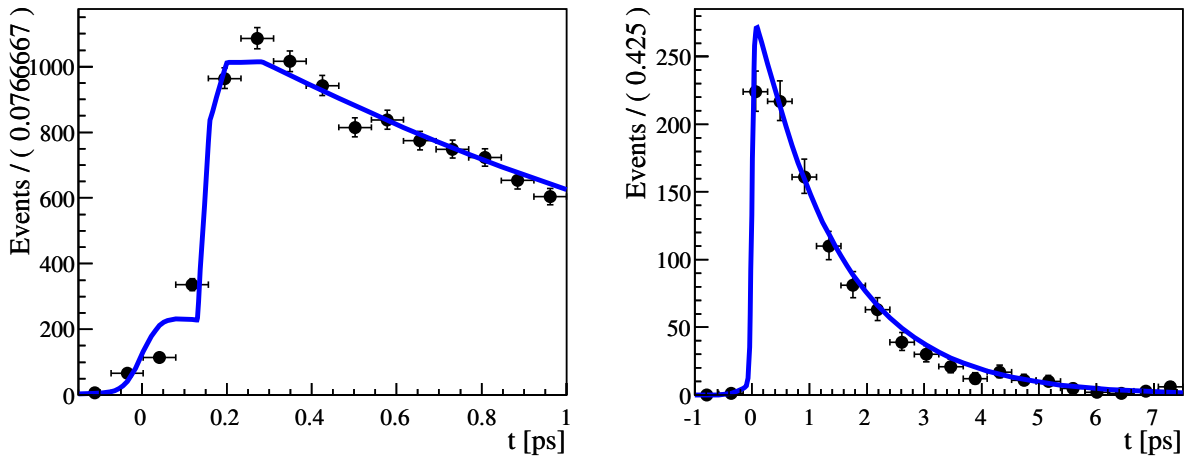


Figure 6.10: *Fit to detached (left) and prescaled (right) samples, with offline selected signal events only. Events passing any HLT1 muon line are considered, also the lines in which cuts are applied which cause a non-uniform proper time efficiency. The non-uniform proper time efficiency in the detached sample is due to the lifetime cut in the HLT2, and possible non-uniform proper time efficiencies inducing cuts in the HLT1.*

The events at low t have passed the lifetime selection criterium $t > \tau_0$ in the HLT2.

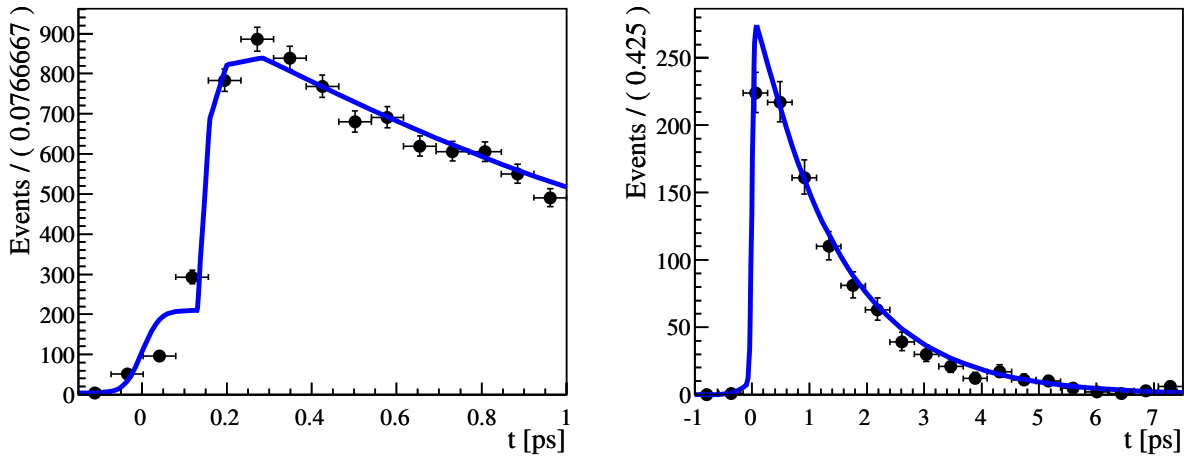


Figure 6.11: *Fit to detached (left) and prescaled (right) samples, with offline selected signal events only. Only events passing the HLT1 di-muon lines, in which no non-trivial proper time inducing cuts are applied, are considered. The non-trivial proper time efficiency of the detached sample is due to the lifetime cut in the HLT2.*

This is not caused by differences between online and offline tracking[4]. Also, the refitting of the primary vertex is not the reason, since refitting actually removes a bias towards values with small t.

It has been found that the events at low t that survive the lifetime cut have a relatively large average number of B candidates per event. These events possibly survive the selection criterium due to the fact that the event has been triggered independent of the signal decay. In other words: other tracks and/or vertices than those belonging to the signal decay have been used to reconstruct the lifetime. This means that these events are most likely triggered by a different candidate than the one used to reconstruct the lifetime.

6.3.2 Efficiency as function of the angular observables

Since the polarization amplitudes are determined from the angular distributions, non-uniform angular efficiencies, i.e. efficiencies which are non-constant as a function of the angular observables, induce biases in the estimates of the polarization amplitudes and correlated parameters. Non-uniform angular efficiencies in the $B_s^0 \to J/\psi\phi$ decay are induced by direct and indirect p_T cuts on the final state particles, e.g. by the detector acceptance or by selection effects.

The relationship between p_T cuts and non-uniform angular efficiencies can be understood from figure 6.12. It is shown that the p_T of the final state kaons is, on average, linearly correlated with $\cos\psi$, where ψ is the helicity angle of the kaons. This can be interpreted from the definition of the angles in figure 2.13: for $\cos\psi = +1$, the K^+ is emitted in the

[4]The efficiency for $t > \tau_0$ approaches the limit $\varepsilon(t) \to 1$ much quicker than the efficiency for $t < \tau_0$ approaches $\varepsilon(t) \to 0$. If the shape of the efficiency would be totally due to the online tracking resolution, this shape is expected be symmetric.

direction of the B, whereas for $\cos\psi = -1$ the K^+ is emitted in the opposite direction. Due to the sum of the momenta of the B and the K^+ meson, the p_T of the K^+ is largest for $\cos\psi \to +1$ and smallest for $\cos\psi \to -1$. For the K^- the relationship is the opposite. For the p_T of the ϕ meson no correlation with $\cos\psi$ is seen.

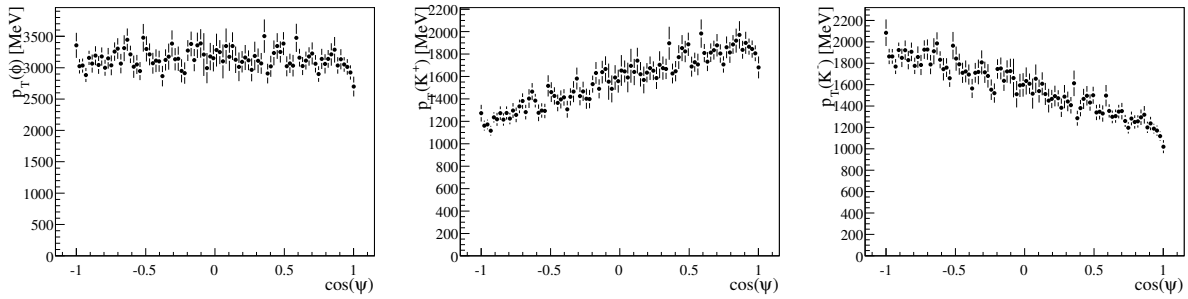


Figure 6.12: *Reconstructed average p_T as a function of $\cos\psi$ for the ϕ mesons (left), positively charged kaons (center) and negatively charged kaons (right) accepted by the $B_s^0 \to J/\psi\phi$ selection.*

In this way p_T cuts (due to trigger, detector acceptance and selection effects) can induce non-uniform angular efficiencies: since the p_T of the kaons, on average, is related to the observed value of $\cos\psi$, and the selection efficiency is non-constant as a function of p_T. This correlation results in the fact that values of $\cos\psi$ which are, on average, related to inefficient values of $p_T(K)$, are themselves less efficient too[5]. For this reason the number of p_T criteria and their values are kept minimal in the event selection procedure presented in this thesis. The selection in chapter 4 has been chosen such that no cuts on the p_T of the final state particles are required: only a cut on the p_T of the composite ϕ meson appears necessary. This means a minimal number of possible systematic effects in the angular distribution are induced.

The angular acceptance after trigger, selection and reconstruction is shown in section 4.6.4, where it has been shown that the efficiency in the region $\cos\psi \to \pm 1$ is smallest. The efficiencies of all angular observables, partly due to apparent efficiencies, fluctuate maximally $\sim 15\%$. Parts of all efficiencies are apparent efficiencies, due to the correlations between the angular observables.

The angular efficiencies in terms of the weights (5.24) can be found in table 6.2. Whereas the values found for $\xi_{4,5,6}$ are compatible with an efficiency which does not bias the amplitudes $A_{4,5,6}$, ξ_1 is significantly smaller than $\xi_{2,3}$, with a difference of 6%. This means that the amplitude accompanying $f_1(\vec{\Omega})$, namely $|A_0|^2$, will be biased to a smaller value in a fit in which an angular efficiency has not been corrected for.

As a check to test the validity of the assumption of factorization of the angular efficiency as a function of the proper time and the angular observables:

$$\varepsilon(t, \vec{\Omega}) = \varepsilon(t) \times \varepsilon(\vec{\Omega}), \tag{6.10}$$

[5]This effect is even more clear in $B_d \to J/\psi K^*$. In this channel the relation between $\cos\psi$ and $p_T(\pi)$ clearly causes the angles associated with low momentum pions ($\cos\psi \to +1$) to be less efficient.

weight	estimated value
ξ_1	3.46 ± 0.01
ξ_2	3.68 ± 0.02
ξ_3	3.73 ± 0.02
ξ_4	0.04 ± 0.02
ξ_5	0.01 ± 0.01
ξ_6	0.01 ± 0.01

Table 6.2: *The angular efficiency weights as determined on the* $\Delta\Gamma = 0$ *sample.*

the ξ_j are determined as a function of the proper time. In figure 6.13 the distributions of $\xi_j(t)$ can be found.

Describing ξ_j, with a constant function, the goodness of fit is given in table 6.3. All $\xi_j(t)$ are consistent with a constant value as a function of t. Since behavior of all weights as a function of proper time is the same, it can be concluded all angular efficiencies can be multiplied by the same efficiency as a function of proper time, and that the factorization in (6.10) is valid.

$\xi_j(t)$	χ^2	N_{DoF}	Prob(χ^2,N_{DoF})
$\xi_1(t)$	19	24	75%
$\xi_2(t)$	24	24	46%
$\xi_3(t)$	28	24	26%
$\xi_4(t)$	23	24	52%
$\xi_5(t)$	28	24	26%
$\xi_6(t)$	28	24	26%

Table 6.3: *The goodness of fit of the hypothesis of a constant function to a description of* $\xi_j(t)$*. From the probabilities, given the* χ^2 *for the number of degrees of freedom (*N_{DoF}*), it can be concluded that all* $\xi_j(t)$ *have a constant value as a function of proper time.*

To be able to include a correction for the efficiency, such that it can be plotted and background can be included in the description of the event sample, the efficiency as function of the three angular observables is decomposed in orthogonal polynomials:

$$\varepsilon(\vec{\Omega}) = e_{abc} P_a(\cos\theta) P_b(\cos\psi) P_c(\phi). \tag{6.11}$$

The result of a decomposition of the efficiency in terms of the 3D polynomials, as described in the previous chapter, can be found in table 6.4.

The systematic error due to the finite order in which the efficiency is decomposed can be determined by comparing the fit results using the weights ξ_j, as estimated by performing a sum over MC events, with the integral of the product of f_j and the efficiency. In the limit of an infinite order parameterization the following equation holds:

$$\xi_j \equiv \int f_j(\vec{\Omega})\varepsilon(\vec{\Omega})\mathrm{d}\vec{\Omega} = \int f_j(\vec{\Omega}) e_{abc} P_a(\cos\theta) P_b(\cos\psi) P_c(\phi)\mathrm{d}\vec{\Omega}. \tag{6.12}$$

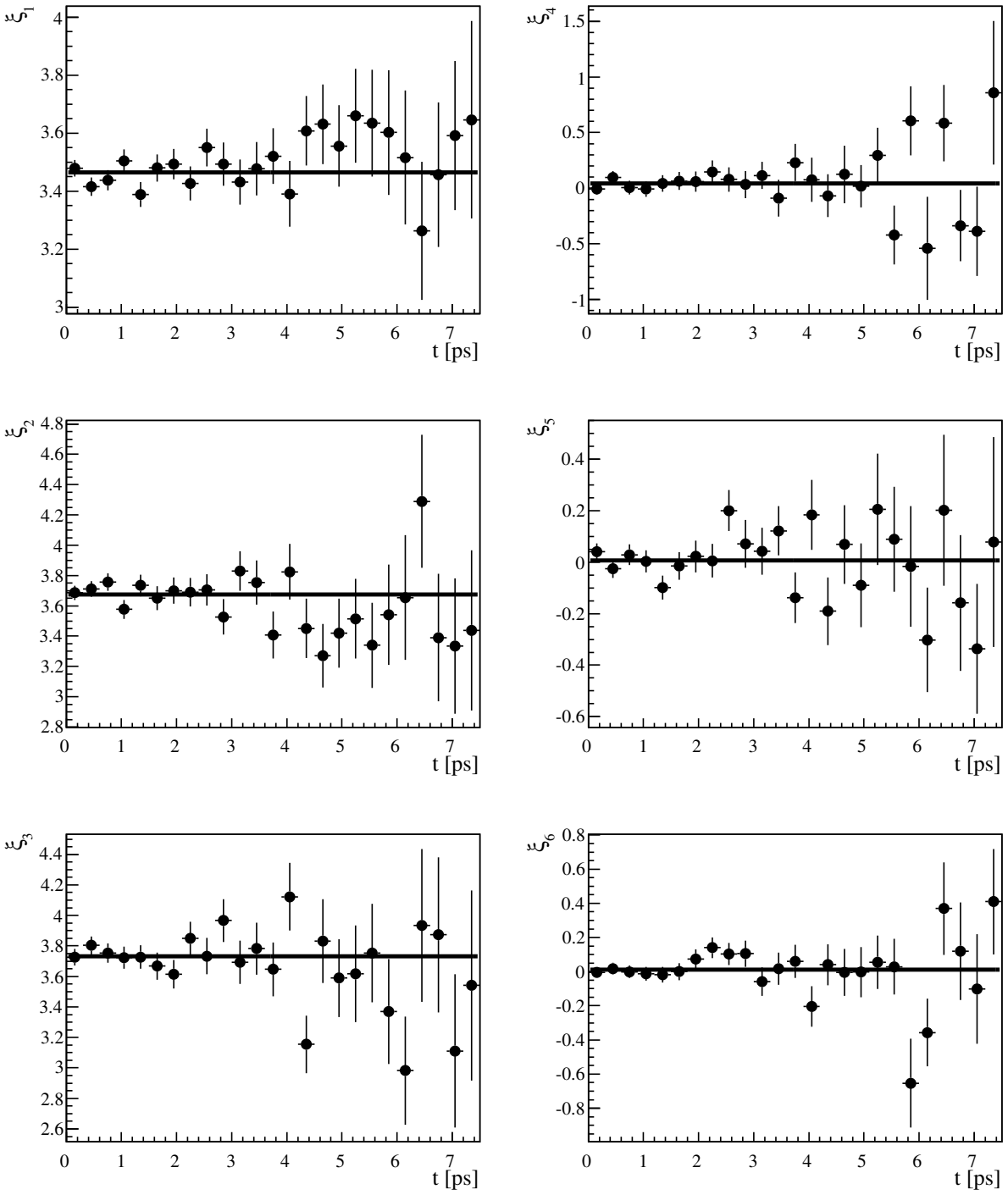


Figure 6.13: *The efficiency weights ξ_j determined as a function of the proper time t. The superimposed line is the fit of the hypothesis of a constant value.*

For a finite order parameterization the two corrections are no longer the same, and a systematic error needs to be estimated.

coefficient	estimated value
e_{000}	$+2.89$
e_{004}	$+0.07 \pm 0.02$
e_{020}	-0.07 ± 0.02
e_{111}	$+0.03 \pm 0.02$
e_{114}	-0.03 ± 0.01
e_{200}	$+0.11 \pm 0.02$
e_{203}	$+0.05 \pm 0.02$
e_{204}	-0.08 ± 0.02

Table 6.4: *The significant angular efficiency coefficients, for a decomposition of the efficiency in 3D polynomials. The estimated uncertainties are used to propagate the systematic effect due to the finite sample on which the angular efficiency has been determined.*

input parameter	without correction	with weights	polynomials
$\|A_0\|^2 = 0.60$	0.568 ± 0.004	0.600 ± 0.004	0.596 ± 0.004
$\|A_\perp\|^2 = 0.16$	0.187 ± 0.005	0.160 ± 0.005	0.164 ± 0.005

Table 6.5: *Fit results with and without corrections for an angular efficiency. Without corrections the estimates of the polarization amplitudes are biased, with correction the values are estimated correctly.*

The result of fits both with and without corrections for the angular efficiency are shown in table 6.5. If no correction is used to take into account the angular efficiency, the polarization amplitudes are significantly biased. If the efficiency is corrected for by the weights determined on the same sample, the input parameters are exactly returned. It is shown that in case the efficiency is decomposed in polynomials, the physics parameters are estimated correctly within statistical uncertainties.

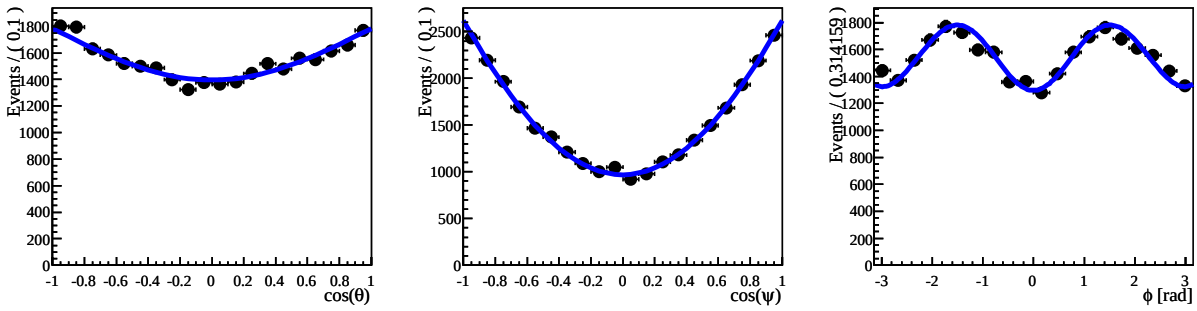


Figure 6.14: *Fit with a correction for the angular efficiency, describing it in terms of polynomials.*

The result of the fit with correction, using an efficiency decomposed in polynomials, is shown in figure 6.14. To show the deviations of both a fit with and a fit without correction,

in figure 6.15 the differences between the PDFs and the distribution of the event samples are shown. Using no correction the PDFs and the event distributions are inconsistent; including the efficiency correction the discrepancy largely disappears.

It can be concluded that a 3D efficiency decomposed in polynomials can be used to correct for the angular efficiency, and that the normalization weights can be used to control the systematic effects due to the parameterization. It should be noted that this does not include a separate systematic effect due to a potential difference between data and MC.

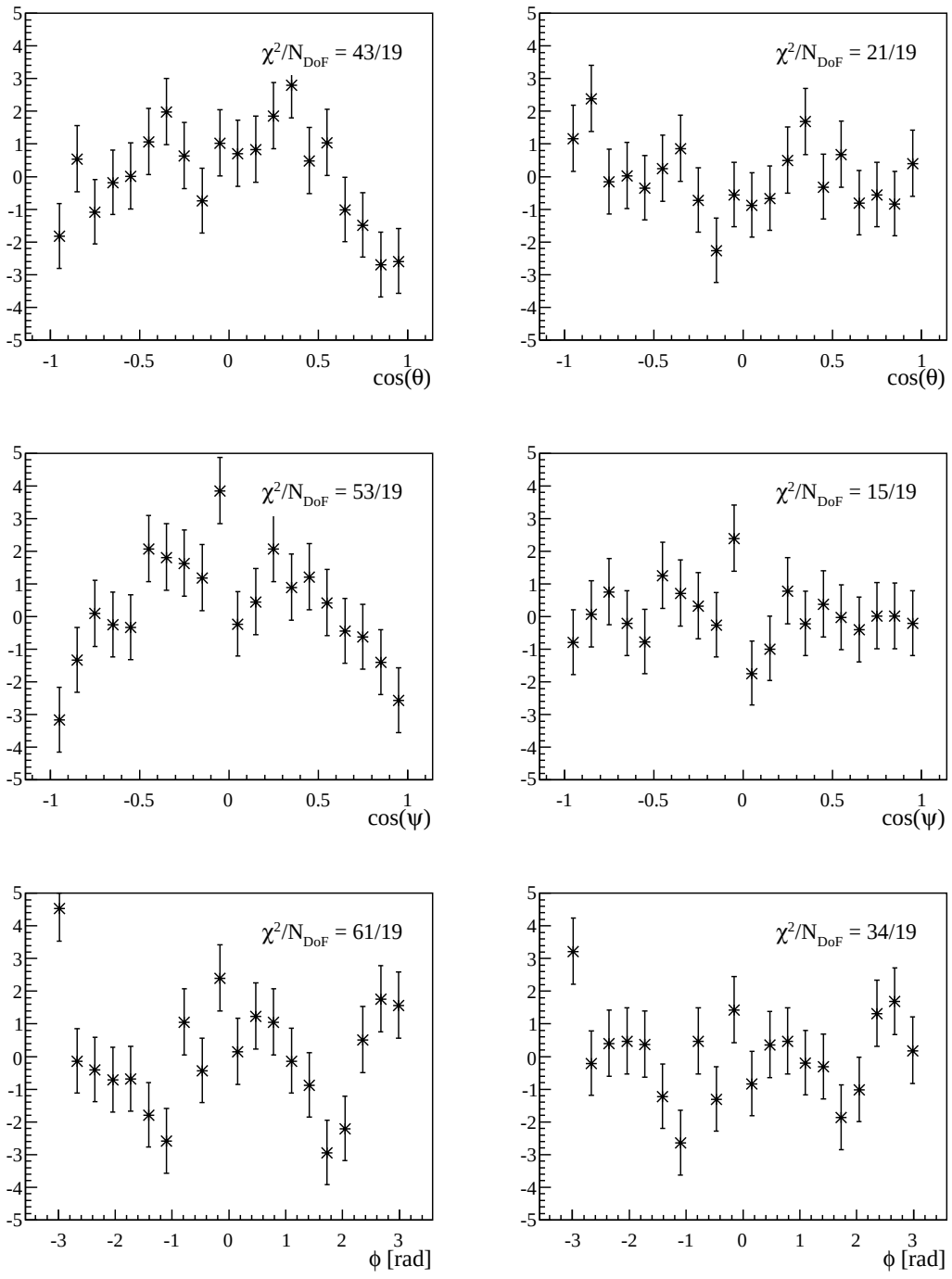


Figure 6.15: *The differences between the theoretical PDF used to describe the events and the number of events, normalized to the error. The comparison of the theoretical distribution and the observed events (left), are inconsistent, as can be concluded from the incompatibility of the points with zero and the values of χ^2/N_{DoF} of a fit of a constant line. The comparison of the corrected PDF and the data (right) are consistent.*

6.4 Tagging performance

Flavor tagging is required to measure a CP asymmetry and determine the phase ϕ_s. Due to various reasons, described in detail in [53], this cannot be done perfectly in LHCb, which leads to a decrease in the sensitivity to ϕ_s. Similar to the case of experimental resolutions and non-uniform efficiencies, the mistagging probability should preferably be as small as possible and must be taken into account, in order to avoid biases in the estimates of the physics parameters.

Since mistagging is important for the estimate of CP violation, a special MC signal sample with large CP violation is used in the study of the tagging performance in this section. Such a sample ensures that relatively large biases in the estimate of ϕ_s also become large on an absolute scale and thus easy to detect.

A calibration can be performed using the OS mistag fraction as estimated from the $B_d^0 \to J/\psi K^*$ channel. To improve the sensitivity to for ϕ_s, events can be grouped into categories with similar tagging properties. The fraction of events f_i per tagging category in the $B_s^0 \to J/\psi\phi$ sample is shown in table 6.6. A category is linked to a corresponding method to determine the flavor tag of the decay, using variables like the p_T of the tagging muon.

The improvement of the precision of the CP asymmetry measurement, and hence of the determination of ϕ_s, by splitting the sample according to categories can be determined by estimating the tagging power. The tagging power can be expressed in terms of the mistag fraction w_i and the fraction f_i of events tagged in that category. Since the sensitivity to the CP asymmetry is linearly related to the number of available events, the sensitivity to the asymmetry is related to the fraction f_i as

$$\sigma(\phi_s) \propto \frac{\sigma(A)}{A} \propto \frac{\sigma(N)}{N} = \frac{1}{\sqrt{f_i}}. \tag{6.13}$$

Since the asymmetry is linearly related to $1 - 2w_i$, the sensitivity to the asymmetry is related to the mistag fraction w_i as:

$$\sigma(\phi_s) \propto \frac{\sigma(A)}{A} \propto \frac{1}{1 - 2w_i}. \tag{6.14}$$

Since $\sigma(\phi_s) \propto 1/[\sqrt{f_i}(1 - 2w_i)]$, the tagging power is defined as

$$Q_i \equiv f_i(1 - 2w_i)^2. \tag{6.15}$$

Using the relationship $\sigma(\phi_s) \propto 1/\sqrt{NQ_i}$, it can be seen that a sample with N events and tagging power Q_i is equivalent to a sample of NQ_i events with perfect tagging.

The values of the tagging power per category, shown in table 6.6, are calculated from the mistag fractions w_i as determined from the self-tagging calibration channel $B_d^0 \to J/\psi K^*$. These values lead to the following three different tagging powers.

1. In case the sample is grouped according to a single category, the average mistag fraction is $\langle w \rangle = 44\%$, and the total tagging power $\langle \bar{Q}_1 \rangle$ equals

$$\bar{Q}_1 \equiv (1 - 2\langle w \rangle)^2 = 1.4\%. \tag{6.16}$$

   This case leads to the worst possible tagging power, but serves as a useful example.

Tagging category	$f_i(\%)$	$w_i(\%)$	$Q_i(\%)$
Cat. 0 (untagged)	55	50	0
Cat. 1	26	43	0.49
Cat. 2	7	34	0.73
Cat. 3	5	31	0.78
Cat. 4	4	25	0.98
Cat. 5	2	19	0.90

Table 6.6: *The fraction of events per tagging category (f_i), the mistag fraction per tagging category (w_i), and the tagging power per tagging category (Q_i) for the $B_s^0 \to J/\psi\phi$ sample, using OS tags only.*

2. If the sample is split according to two categories, namely into one sample of a fraction of $f_0 = 55\%$ of untagged events (with $w_0 = 0.5$), and another sample of a fraction of $f' = 45\%$ of tagged events with an average mistag rate $\langle w \rangle' = 37\%$, the total tagging power for splitting according to two tagging categories, $\bar{Q}_2$, equals

$$\bar{Q}_2 \equiv f'(1 - 2\langle w \rangle')^2 = 3.0\%. \tag{6.17}$$

3. Finally, in case the sample is divided according to six tagging categories, the total tagging power, $\bar{Q}_6$, with w_i as shown in table 6.6, is equal to

$$\bar{Q}_6 = \sum_i Q_i = \sum_i f_i(1 - 2w_i)^2 = 3.9\%. \tag{6.18}$$

The expected sensitivity improvement of ϕ_s due to splitting the sample in two is expected to be

$$\sqrt{\frac{Q_2}{Q_1}} = \sqrt{3.0/1.4} = 1.46. \tag{6.19}$$

If the sample is divided in six instead of two, the expected sensitivity improvement is

$$\sqrt{\frac{Q_6}{Q_2}} = \sqrt{3.9/3.0} = 1.14. \tag{6.20}$$

To check the expected improvements, a fit is performed simultaneously to the signal and the control sample, as explained in chapter 5. The result of such a fit, using six tagging categories, is shown in figure 6.16. Performing a fit with ~ 10 k $B_s^0 \to J/\psi\phi$ events in the signal sample and ~ 40 k $B_d^0 \to J/\psi K^*$ events in the control sample, and comparing the cases of one and six categories, the sensitivity improvement is found to be

$$\frac{\sigma_1(\phi_s)}{\sigma_6(\phi_s)} = 1.9 \pm 0.1, \tag{6.21}$$

which is compatible with the expected improvement from the increase in tagging power, $\sqrt{\frac{\bar{Q}_6}{\bar{Q}_1}} = 1.7$.

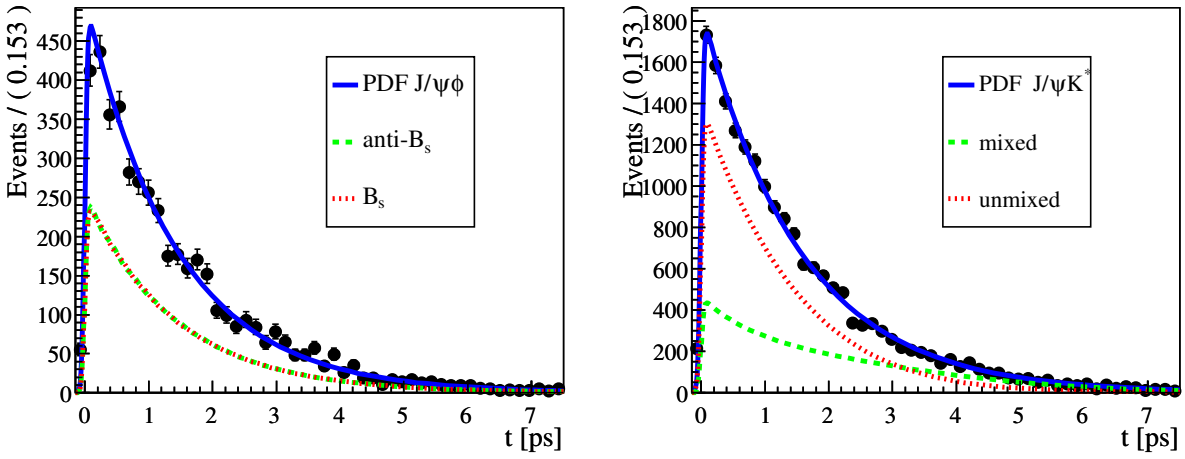


Figure 6.16: *The result of a simultaneous fit of the PDFs used to describe the 10 k $B_s^0 \to J/\psi\phi$ decays (left) and the 40 k $B_d^0 \to J/\psi K^*$ decays (right), projected on the proper time observable. The mistag fractions w_i per tagging category are shared parameters between the PDFs. Besides the total PDFs, the components are shown: the events tagged as $B_s/\bar{B}_s$ (left) and the events tagged as mixed/unmixed (right). The untagged events are not shown in the plot, which makes it possible to clearly distinguish the mixed and unmixed PDFs.*

Finally, the effect of error propagation is determined. By performing a simultaneous fit, the uncertainty of the mistag fraction due to the limited statistics in the control sample is propagated to the signal sample. Performing fits with and without propagating these uncertainties, the uncertainty in ϕ_s changes as

$$\frac{\sigma(\phi_s)}{\sigma(\phi_s)'} = 110\%. \tag{6.22}$$

This implies that a systematical error of 10%, due to a statistical uncertainty in the control sample, is avoided.

It can be concluded that using $B_d^0 \to J/\psi K^*$ as a control channel, the OS tags can be calibrated for $B_s^0 \to J/\psi\phi$. Furthermore, a simultaneous fit to both the signal and the control sample avoids a systematic uncertainty, of statistical origin, of $\sim 10\%$. Finally, a division of the signal sample according to six categories leads to an improvement of the sensitivity to ϕ_s of 14%, compared to a division according to two categories.

It should be noted that the method applied to $B_d^0 \to J/\psi K^*$ can only be used for OS tags, but a similar method can be used for SS tags. The SS mistag fraction can be determined for example from the mixing amplitude of the $B_s^0 \to D_s^{\pm}\pi^{\mp}$, in which case the total tagging power Q becomes 6.2% [20].

6.5 Background

Background events decrease the sensitivity to the physics parameters of the signal distribution, as explained in chapter 4. Furthermore, backgrounds not taken into account correctly

can lead to biases of the estimates of physics parameters. In this section the behavior of the background in terms of the proper time, angular, tagging, and mass observables is described.

6.5.1 Proper time distribution of background events

The dominant part of the background for the $B_s^0 \to J/\psi\phi$ channel is prompt background, i.e. background around $t = 0$. For this reason the selection consists of a detached and a prescaled sample. Whereas the detached selection is most sensitive to the signal parameters, the prescaled sample is necessary for calibration purposes, in particular to determine the proper time resolution.

The distribution of the background events after selection, excluding the cut on the lifetime, is shown in chapter 4 and is described as:

$$B_t(t) = G_1(t, t'|\mu, \sigma_1, \sigma_2, f_{\sigma_1}) \otimes (f_{\text{prompt}}\delta(t') + (1 - f_{\text{prompt}})B_{LL}(t'|\tau_{LL})). \tag{6.23}$$

In table 6.7 the values of the parameters are given, as determined in the fit. To describe the signal resolution in the presence of background, three possible scenarios are foreseen.

σ_1	(28 ± 1) fs
σ_2	(67 ± 4) fs
f_1	(75 ± 4) %
μ	0 ps
τ_{LL}	(0.33 ± 0.05) ps
f_{prompt}	99.3%

Table 6.7: *Values of the parameters used to describe the proper time distribution of the background events.*

1. If the amount of prompt background is small enough such that there is sensitivity to the proper time resolution of the signal distribution, the background can be modelled by this distribution from the mass sidebands, whereas the signal resolution, as described in section 6.2.1, can be fitted directly on the signal events, using the per-event calculated error.

2. In case the amount of prompt background is too large, a second option is to use the per-event error description for both the proper time distribution of the prompt background and for the the signal. A validation of this possibility has not been performed.

3. The last option would be to describe the signal proper time resolution directly by the proper time distribution of the prompt background.

The width of the proper time distribution of the prompt background described by a single gaussian, 38 ± 1 fs, is compatible with the average signal proper time resolution of 39 fs, as determined in [20]. This means the prompt background can be used as a description of the resolution of the signal. Systematic uncertainties due to assuming the wrong proper time resolution model, for example by overestimating/underestimating the width, are described in [59].

6.5.2 Angular distribution of the background events

As the angular distribution of the background events is not expected to be the same as for the signal, the estimate of the signal polarization amplitudes will be biased, when not taken into account correctly. The PDF used to describe this distribution, is

$$B_\Omega(\vec{\Omega}) = \frac{b_{abc} P_a(\cos\theta) P_b(\cos\psi) P_c(\phi)}{\int b_{abc} P_a(\cos\theta) P_b(\cos\psi) P_c(\phi) \mathrm{d}\vec{\Omega}}. \tag{6.24}$$

In previous chapter the method is described to determine the coefficients of this decomposition, and their uncertainties. After doing so, this PDF is fitted to a pure background distribution. Floating the significant terms only in the fit to the 3D angular distribution of the 'inclusive J/ψ' sample, the values in table 6.8 are found. The angular distribution of the events in the 'inclusive J/ψ' sample and the corresponding PDF obtained in the fit are shown in figure 6.17.

coefficient	estimated value
b_{000}	$+2.83$
b_{012}	-0.09 ± 0.05
b_{212}	$+0.10 \pm 0.04$
b_{222}	-0.22 ± 0.05
b_{310}	-0.12 ± 0.04
b_{400}	$+0.10 \pm 0.05$
b_{600}	-0.13 ± 0.05
b_{611}	$+0.08 \pm 0.05$
b_{612}	$+0.07 \pm 0.05$
b_{620}	-0.14 ± 0.05
b_{622}	-0.12 ± 0.05
b_{700}	$+0.09 \pm 0.05$

Table 6.8: *Parameters of the angular distribution of the background events in the 'inclusive J/ψ sample. The parameters in the table are the significant terms, after a fit.*

6.5.3 Tagging properties of the background

Since background events do not originate from a true (anti-)B_s^0 meson, a mistag fraction is an undefined variable for background events. However, the fraction of events per flavor and per tagging category can be estimated on a background sample, and as such can be used to distinguish signal from background. Since the tagging properties can also differ for different types of background, the fractions of events per tagging category, f_i^{bg}, are estimated for different types of background.

Using N_{cats} different tagging categories c_i, and assuming the same number of events for B and $\bar{B}$ tags per category, the background PDF reads

$$B_q \to \frac{f_i^{bg} b(q|c_i)}{\sum_q \sum_i^{N_{\mathrm{cats}}} f_i^{bg} b(q|c_i)}. \tag{6.25}$$

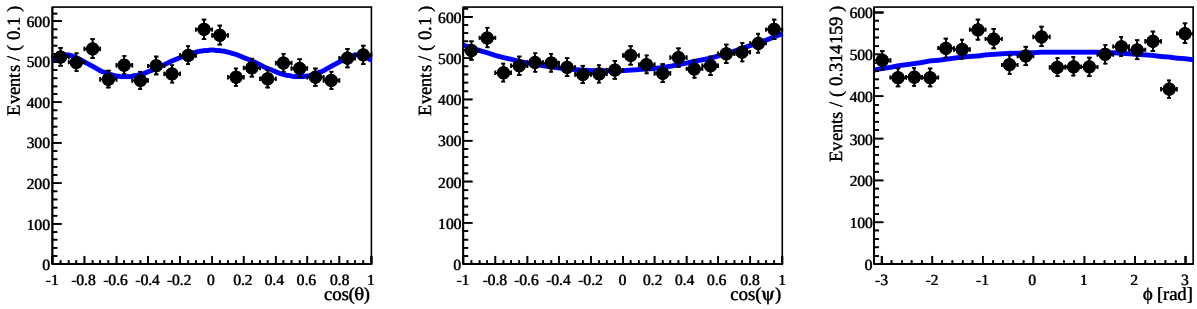


Figure 6.17: *Angular distribution of the 'inclusive J/ψ' sample. Superimposed is the PDF, constructed of the significant terms of a polynomial decomposition, after a fit of the coefficients to the event sample.*

In case the background PDF does not depend on the tagging flavor, this simplifies to

$$B_q \to \frac{f_i^{bg} b(c_i)}{\sum_i^{N_{\text{cats}}} f_i^{bg} b(c_i)}. \tag{6.26}$$

The fraction of events per tagging category for background events are given in table 6.9. Clearly, the fraction of mistag events differ for the signal and background, and for the different background samples.

Tagging category	f_i^{bg}(prompt)	f_i^{bg}(long-lived)
Cat. 0	65%	35%
Cat. 1	22%	37%
Cat. 2	7%	11%
Cat. 3	4%	8%
Cat. 4	1%	5%
Cat. 5	1%	4%

Table 6.9: *Fraction of events per tagging category for prompt and long-lived background [59].*

6.5.4 Invariant mass distribution of the background

The long-lived background distribution of the invariant B mass can be described by an exponential decay with a decay constant of 0.001 MeV/c^2, the prompt background by an exponential decay with a decay constant of 0.0006 MeV/c^2 [20]. The background properties in the signal region in the other observables $(t, \vec{\Omega}, q)$ are estimated from the sidebands of the invariant mass distributions.

6.6 Toy experiments for sensitivity study

In chapter 4 a rough estimate of the sensitivity to ϕ_s has been made of $\sigma(\phi_s) \sim 25$ mrad for 2.0 fb^{-1}, from simple considerations of the signal yield and purity. Here, the sensitivity is determined more accurately, using toy MC experiments. As PDFs, the signal and background PDFs in section 6.1 are used. As input for efficiencies, resolutions, and backgrounds the estimates obtained in previous sections are used.

First the parameter settings of the PDFs, following from the full MC studies, are given. When simplified, the simplifications, valid for at least 0.2 fb^{-1} of data taking, are motivated. Subsequently, toy MC experiments are performed to determine the sensitivity to the different physics parameters, in particular to ϕ_s. Also, the correlations between the different parameters are estimated. Finally, the amount of integrated luminosity needed to acquire a sensitivity better than the combined sensitivity of the Tevatron experiments, at the time of writing, is given.

6.6.1 Settings of the toy experiments

In a dedicated study [59] it has been observed that the proper time per-event error is ignored, since it does not influence the sensitivity, for a realistic value of Δm_s and realistic settings of the proper time resolution. Since the description of the proper time resolution for signal and prompt background events have been shown to be consistent, these are described with the same model: a single Gaussian with a width equal to $\sigma = 39$ fs.

The offline proper time efficiency is taken constant as a function of t. Furthermore, the variation of the online proper time efficiency due to the detached selection is chosen to be ignored, since it does not influence the sensitivity to ϕ_s, as has been shown in chapter 4. Therefore, only the number of events at larger t needs to be modelled correctly and the HLT plateau efficiencies $\varepsilon_{\text{HLT1,2}}(t > 0.2\,\text{ps})$ can be used. The effective number of signal events S_{eff} (for 2.0 fb^{-1}) is then defined as

$$S_{\text{eff}} = S_{\text{offline}} \times \varepsilon_{\text{HLT1}}(t > 0.2\,\text{ps}) \times \varepsilon_{\text{HLT2}}(t > 0.2\,\text{ps}). \tag{6.27}$$

As explained in chapter 4, this is a valid assumption to determine the sensitivities, since events at low t, where the efficiency decreases, do not contribute to the sensitivities to the signal parameters. Using the plateau efficiencies and the number of offline selected events found in chapter 4, the effective number of signal events equals

$$S_{\text{eff}} = 207\,\text{k} \times 92\% \times 90\% = 171\,\text{k}. \tag{6.28}$$

The purity of this number of signal events, i.e. the fraction of signal events, is estimated in section 4.6.2 and taken as input for the fractions of prompt and long-lived background: the fraction long-lived background events is $B_{LL}/S = 0.3$ with a lifetime $\tau_{LL} = 0.33$ ps, the fraction prompt background events is $B_{\text{prompt}}/S = 47$. The distribution of the invariant B_s mass[6] has a width of 16 MeV/c^2, and the invariant mass spectrum of the backgrounds is described as an exponential decay with the decay constant which describes the long-lived background: 0.001 MeV/c^2. The distributions of the daughter masses are ignored in

[6]Ignoring an improved resolution due to a mass-constrained fit.

the description, which means the S-wave contribution in the ϕ-mass distribution is not included. The effects of the S-wave contribution is evaluated in [31].

Furthermore, the efficiency is assumed to be constant as a function of the angular observables, as it does not influence the measurement precision of ϕ_s [48]. Also the angular resolution is ignored in the study; a possible systematic effect has been estimated earlier in this chapter. Finally, the angular distributions of the background events are taken to be constant, as the deviations from this assumption have been shown to constitute at most 10%.

As shown above, the tagging power is equal to $Q = 6.2\%$ in case the sample is divided according to six tagging categories. Implementing this tagging power by an equivalent single mistag fraction, the mistag fraction is set equal to an effective mistag fraction $w_{\text{eff}} = 37.5\%$ for all events in the sample. This value is fixed in the fit, and systematic uncertainties due to the limited size of the $B_d^0 \to J/\psi K^*$ sample have been determined earlier this chapter. Furthermore, the number of events tagged as B and as $\bar{B}$ in the background sample are assumed to be the same, and all background events are grouped into one tagging category, as the distributions in the tagging observables hardly differ between signal and background. Finally, it should be noted that in case same-side tagging information is ignored, the tagging power is 3.9% and the measurement precision ϕ_s increases by a factor 6.2/3.9.

Finally, the settings are summarized in table 6.10.

	Signal	Prompt Background	Long lived background
B_i/S	-	$B_{pr}/S = 47$	$B_{LL}/S = 0.3$
t	$\sigma = 39$ ps, $\mu = 0$	$\sigma = 39$ ps, $\mu = 0$	$\tau_{LL} = 0.33$ ps
$\vec{\Omega}$	Constant efficiency	Constant distribution	Constant distribution
q	$\varepsilon = 0, w_{\text{eff}} = 37.5\%$	Constant distribution	Constant distribution
m_B	$\sigma = 16$ MeV	$\tau = 0.001$ MeV	$\tau = 0.001$ MeV

Table 6.10: *The settings used in the toy experiments for the background fractions B_i/S and for the distributions in the different observables: proper time t, angular observables $\vec{\Omega}$, tagging flavor q, and B_s invariant mass m_B.*

6.6.2 Example of a toy experiment

To estimate the sensitivity to the physics parameters, 1000 toy experiments simulating signal and background events corresponding to an integrated luminosity of 2.0 fb^{-1}. In figure 6.18 an example of one such experiment is shown. The events shown are the sum distribution of signal and background toy events, where in addition to a fit of the total PDF also the signal and background components are shown.

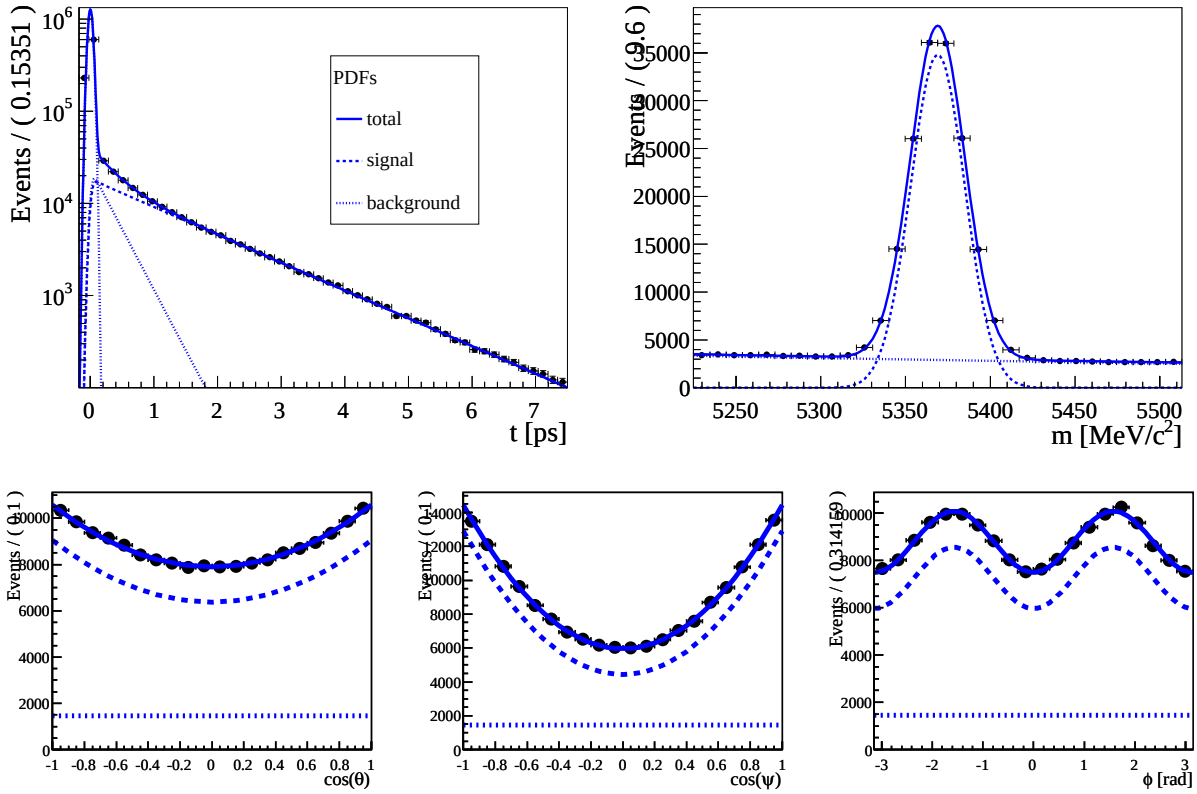


Figure 6.18: *Distribution of one toy experiment of* 2.0 *fb*$^{-1}$ *of data taking, with the fitted PDF overlaid. Furthermore, the signal (dashed) and background (dotted) components are shown. The distribution in the proper time observable t is shown in the three* σ B_s^0 *mass window; the distribution in the angular observables are shown in the three* σ B_s^0 *mass window and in the range* $t > 0.20$ *ps.*

In the proper time distribution the prompt background, peaking around $t = 0$, as well as the long living background contribution can be distinguished. In the angular observables the signal distribution behaves as described in chapter 2, whereas the distribution of the background is constant as a function of the angular observables. The invariant mass distribution of the signal PDF peaks at the B_s^0 mass, whereas the background decreases practically linearly as a function of the invariant mass. The events in the sidebands of the invariant mass distribution are used to derive a description for the proper time and angular distribution of the background in the region of the mass peak.

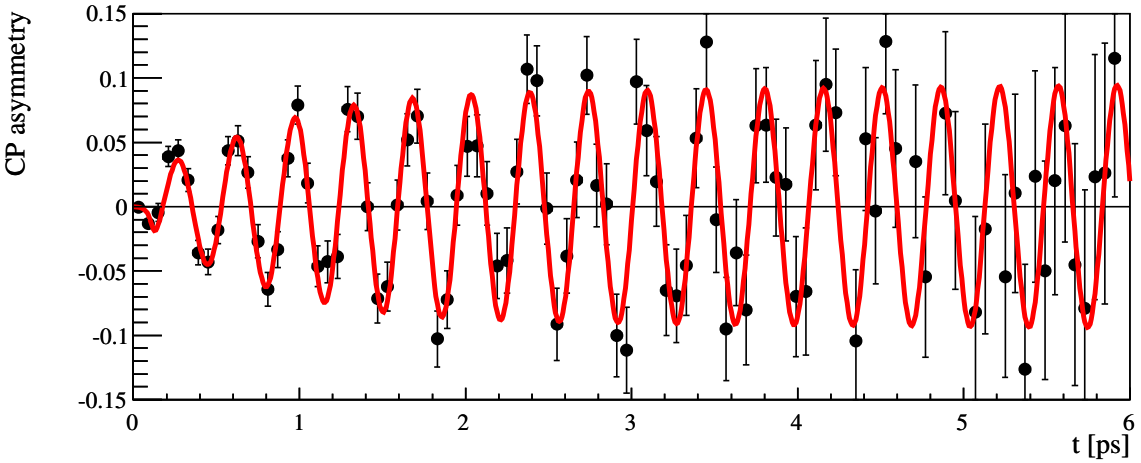


Figure 6.19: *The CP asymmetry for* 2.0 fb^{-1} *of data taking for* $\phi_s = -0.70$ *rad, as a function of proper time. The observable CP asymmetry approaches zero as* $t \to 0$*, due to the prompt background.*

The CP asymmetry of one such experiment is shown in figure 6.19. To demonstrate the asymmetry, the value of ϕ_s is chosen to be -0.70 rad. The proper time resolution is taken to be a single Gaussian with a width of 39 fs.

6.6.3 Correlations between signal parameters

Before the sensitivities to the physics parameters are presented, their correlations are discussed, for both parameterization methods of the strong phases. Correlations are important, since via correlations biases in one parameter can lead to biases in other parameters.

For the the first case, when including the strong phases as fit parameters, strong correlations are expected between the polarization amplitudes, since the three polarization fractions are required to add up to one. Furthermore, strong correlations between the polarization amplitudes and $\Delta\Gamma_s$ are expected, since individual polarization amplitudes have different lifetimes. Naturally, also a strong correlation between the decay width and the width difference is expected. And because the polarization phases always appear in a subtraction of two of these, strong correlations are expected between them.

In the correlation matrix presented in table 6.11, as expected, strong correlations, with absolute values of the correlations > 0.4, between $|A_\perp|^2$, $|A_0|^2$, Γ_s, $\Delta\Gamma_s$ are indeed found, for the method in which the strong phases $(\delta_\parallel, \delta_\perp)$ are directly used as fit parameters. On the other hand, ϕ_s appears to be very uncorrelated with all parameters. The correlation between the strong phases $(\delta_\parallel, \delta_\perp)$ is large $(+0.68)$, which is different from the results presented in [30], where low correlations between the strong phases have been found.

For the second case, including the strong parameters $(C_{\perp,0}, S_{\perp,0}, C_{\parallel,0}, C_{\perp,\parallel}, S_{\perp,\parallel})$ as fit parameters, again strong correlations between $|A_\perp|^2$, $|A_0|^2$, Γ_s, $\Delta\Gamma_s$ are found in the correlation matrix in table 6.12. Furthermore, ϕ_s appears to remain largely uncorrelated, as in the first fit. The strong parameters have small correlations amongst each other. A strong

(anti-)correlation of -0.47 is observed between $|A_\perp|^2$ and $C_{\parallel,0}$. This can be understood from the fact that $C_{\parallel,0}$ appears in (2.73) in the product $|A_\parallel||A_0|\cos(\delta_\parallel - \delta_0)$.

	$\|A_\perp\|^2$	$\|A_0\|^2$	Γ_s	$\Delta\Gamma_s$	ϕ_s	$\delta_\parallel$	$\delta_\perp$
$\|A_\perp\|^2$	+1	−0.64	+0.46	−0.60	−0.02	−0.08	−0.07
$\|A_0\|^2$		+1	−0.40	+0.61	−0.01	−0.01	+0.01
Γ_s			+1	−0.70	+0.00	+0.13	+0.04
$\Delta\Gamma_s$				+1	−0.08	−0.02	+0.05
ϕ_s					+1	−0.02	−0.09
$\delta_\parallel$						+1	+0.68
$\delta_\perp$							+1

Table 6.11: *Average correlation matrix for 125 toy experiments of* 2.0 fb^{-1} *of data taking, implementing the strong phases as fit parameters.*

	$\|A_\perp\|^2$	$\|A_0\|^2$	Γ_s	$\Delta\Gamma_s$	ϕ_s	$C_{\perp,0}$	$S_{\perp,0}$	$C_{\parallel,0}$	$C_{\perp,\parallel}$	$S_{\perp,\parallel}$
$\|A_\perp\|^2$	+1	−0.60	+0.44	−0.58	+0.06	+0.02	−0.01	−0.47	+0.03	−0.03
$\|A_0\|^2$		+1	−0.40	+0.61	−0.01	+0.05	+0.04	+0.06	+0.04	+0.04
Γ_s			+1	−0.70	+0.01	−0.13	−0.10	−0.09	−0.12	−0.08
$\Delta\Gamma_s$				+1	−0.09	+0.11	−0.01	+0.13	+0.10	+0.03
ϕ_s					+1	+0.01	+0.09	−0.04	−0.12	+0.01
$C_{\perp,0}$						+1	+0.04	−0.16	+0.10	−0.08
$S_{\perp,0}$							+1	+0.01	+0.08	+0.13
$C_{\parallel,0}$								+1	+0.06	+0.05
$C_{\perp,\parallel}$									+1	−0.05
$S_{\perp,\parallel}$										+1

Table 6.12: *Average correlation matrix for 125 toy experiments of* 2.0 fb^{-1} *of data taking, including the strong parameters as fit parameters.*

It is concluded that for both parameterizations the correlations between the polarization amplitudes and the widths is strong. This means that biases of the polarization amplitudes due to angular efficiencies, via correlations, can lead to biases of the decay widths. Vice versa, biases of the widths to proper time dependent efficiencies can, via correlations, lead to biases of the polarization amplitudes. The correlation of the CP-violating phase ϕ_s with the other physics parameters is small.

6.6.4 Expectations of the sensitivity

In table 6.13 the expected sensitivities to the physics parameters for 2.0 fb^{-1} of integrated luminosity at the LHCb experiment are given. These expectations follow from the yields and purities after the selection described in chapter 4, and using the fit methods to take into account the distortions of the signal distributions as described in previous chapter. In

particular, the expected sensitivity to the CP-violating phase ϕ_s is (24.0 ± 0.7) mrad for 2.0 fb^{-1} of data taking.

parameter	input value	sensitivity
ϕ_s [rad]	-0.04	$(2.40 \pm 0.07) \times 10^{-2}$
$\Delta\Gamma_s/\Gamma_s$	0.10	$(1.56 \pm 0.05) \times 10^{-2}$
τ_s [ps]	1.47	$(8.1 \pm 0.2) \times 10^{-3}$
$\lvert A_\perp \rvert^2$	0.16	$(3.5 \pm 0.1) \times 10^{-3}$
$\lvert A_0 \rvert^2$	0.60	$(2.6 \pm 0.1) \times 10^{-3}$
$\delta_\perp$ [rad]	2.91	$(7.6 \pm 0.2) \times 10^{-2}$
$\delta_\parallel$ [rad]	-2.93	$(7.4 \pm 0.2) \times 10^{-2}$

Table 6.13: *Expected sensitivities for 2.0 fb^{-1} of data taking.*

Finally, the expected two-dimensional likelihood profile in the $(\phi_s, \Delta\Gamma_s/\Gamma_s)$ plane for different luminosities, for an input value of $\phi_s = -0.70$ rad, is shown in figure 6.20. This input value is chosen, since it corresponds approximately to the central values of the combined results of the Tevatron experiments at the time of writing, as shown in chapter 2.

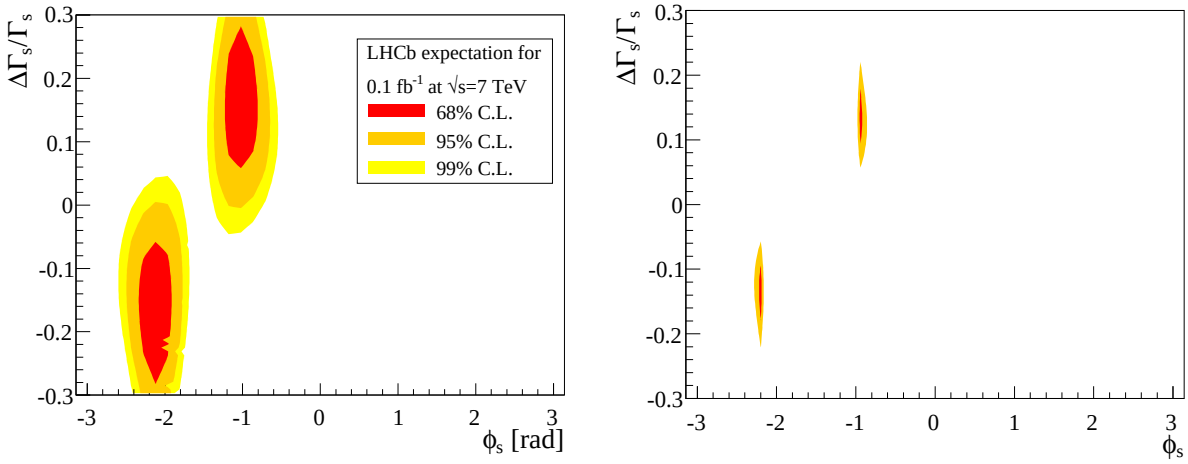


Figure 6.20: *Expectation of the likelihood profile in the (ϕ_s,$\Delta\Gamma_s/\Gamma_s$) plane for 0.1 fb^{-1} of data taking at a center of mass energy of 7 TeV (left) and 2.0 fb^{-1} at 14 TeV (right) at the LHCb experiment. For comparison, the input values are chosen (ϕ_s,$\Delta\Gamma_s/\Gamma_s$) = (-0.70 rad, 0.10), similar to the combined measurement of CDF and DØ shown in figure 2.10.*

Using the selection and the fit methods described in this thesis, the expected performance of the LHCb experiment is as follows. After 0.1 fb^{-1} of luminosity (at a center of mass energy of 7 TeV), the amount of integrated luminosity expected at the end of 2010, the precision of LHCb should surpass that of the Tevatron for a total luminosity of 18 fb^{-1} [20], as expected at the end of run 2 [60]. After 1.0 fb^{-1} of luminosity, the amount of integrated luminosity expected before the shutdown of 2012, the sensitivity is 48 mrad.

After 10 fb^{-1} (at a center of mass energy of 14 TeV) the statistical precision is expected to be 8 mrad, which means that for the SM value of the CP-violating phase, zero CP violation can be excluded at 5σ.

Appendix A

The angular distribution

The derivation of the angular distribution of a decay of a scalar B-meson into two vector mesons using the helicity formalism is explained extensively in [26], albeit for the $B_d^0 \to J/\psi K^*$ decay. The derivation using the helicity formalism is summarized here. Another derivation can be found in [17]. Finally, also the transformations between different conventions of the transversity frame are described.

A.1 Derivation of the angular distribution

The helicity formalism is described in [27] [28] [29]. It is particularly suited for relativistic problems because the helicity operator $\lambda = \vec{S} \cdot \hat{p}$ is invariant[1] under both rotations and boosts along $\hat{p}$ [27]. The matrix element of the decay of a particle $A \to B + C$, with the subsequent decays of unstable particles $B \to B_1 + B_2$ and $C \to C_1 + C_2$ is written as

$$\mathcal{M}(\lambda_{B_1}, \lambda_{B_2}, \lambda_{C_1}, \lambda_{C_2}) \propto \sum_{\lambda_B, \lambda_C} A^{A\to B+C}_{\lambda_B,\lambda_C} A^{B\to B_1+B_2}_{\lambda_{B_1},\lambda_{B_2}} A^{C\to C_1+C_2}_{\lambda_{C_1},\lambda_{C_2}}. \tag{A.1}$$

The following conventions discussed in [26] are used in the derivation:

- Rotations are defined actively, in which the rotations of the physical system with respect to the fixed coordinates with Euler angles are defined counterclockwise. $R(\alpha, \beta, \gamma) = R_z(\alpha)R_y(\beta)R_z(\gamma)$, where the indices denote rotations around that axis.
- The Jackson convention is used: z-axes in the frames of the vector mesons are defined in the direction of the meson and anti-parallel, the x-axes of the different frames are defined parallel. Therefore the Euler angles are $(\alpha, \beta, \gamma) \equiv (\varphi, \theta, 0)$, where θ is the polar angle and φ is the azimuthal angle.

The amplitudes can then be written as

$$A^{A\to B+C}_{\lambda_B,\lambda_C} = \sqrt{\frac{2J_A+1}{4\pi}} D^{J_A*}_{m_A,\lambda_B-\lambda_C}(\varphi_A, \theta_A, 0) H^A_{\lambda_B,\lambda_C}$$

[1] As long as the direction of the momentum is not reversed.

$$
\begin{aligned}
A^{B\to B_1+B_2}_{\lambda_{B_1},\lambda_{B_2}} &= \sqrt{\frac{2J_B+1}{4\pi}} D^{J_B*}_{m_B,\lambda_{B_1}-\lambda_{B_2}}(\varphi_B,\theta_B,0) H^B_{\lambda_{B_1},\lambda_{B_2}} \\
A^{C\to C_1+C_2}_{\lambda_{C_1},\lambda_{C_2}} &= \sqrt{\frac{2J_C+1}{4\pi}} D^{J_C*}_{m_C,\lambda_{C_1}-\lambda_{C_2}}(\varphi_C,\theta_C,0) H^C_{\lambda_{C_1},\lambda_{C_2}}.
\end{aligned} \tag{A.2}
$$

with the rotation matrices

$$
D^J_{m,m'}(\alpha,\beta,\gamma) = e^{-im\alpha} d^J_{m,m'}(\beta) e^{-im'\gamma}. \tag{A.3}
$$

Here the d-matrices equal

$$
d^J_{m,m'}(\beta) = \langle jm'|e^{-i\beta J_y}|jm\rangle. \tag{A.4}
$$

$\vec{J} = \vec{L}+\vec{S}$ is the total angular momentum, m_i is the spin projection and H the Hamiltonian. In general the unpolarized angular distribution is given by averaging over initial spins and summing over final spins:

$$
\frac{1}{\Gamma}\frac{\mathrm{d}^3\Gamma}{\mathrm{d}\Omega_A \mathrm{d}\Omega_B \mathrm{d}\Omega_C} \propto \frac{1}{2s_A+1} \sum_{\lambda_{B_1},\lambda_{B_2},\lambda_{C_1},\lambda_{C_2}} |\mathcal{M}(\lambda_{B_1},\lambda_{B_2},\lambda_{C_1},\lambda_{C_2})|^2. \tag{A.5}
$$

Since the initial particle is a spinless B^0, it follows that $J_A = 0$. In that case the secondary particles in the restframe of the B^0 must have opposite spins, so the secondary particles in the initial arbitrary frame have equal helicities. Now the arbitrary particle B, C are associated to the vector mesons of the $B^0_s \to J/\psi\phi$ decay: $B = J/\psi$ and $C = \phi$. Since the mesons move along their respective z-axes and since the (arbitrary) quantization axis is chosen to be in the J/ψ-direction, it follows that

$$
m_B \equiv m_{J/\psi} = \lambda, m_C \equiv m_\phi = -\lambda. \tag{A.6}
$$

Hence the matrix element reads

$$
\begin{aligned}
\mathcal{M} \;\propto\; & \sqrt{\frac{1}{4\pi}} D^{0*}_{0,0}(\varphi_{B^0},\theta_{B^0},0) H^{B^0\to J\psi\phi}_\lambda \\
& \sqrt{\frac{3}{4\pi}} D^{1*}_{\lambda,\lambda_{l^+}-\lambda_{l^-}}(\varphi_l,\theta_l,0) \\
& \sqrt{\frac{3}{4\pi}} D^{1*}_{-\lambda,\lambda_{K^+}-\lambda_{K^-}}(\varphi_K,\theta_K,0),
\end{aligned} \tag{A.7}
$$

where H^ϕ and $H^{J/\psi}$ are neglected, since they do not depend on λ.

In general the amplitude is the sum over all possible decay paths to the final state, therefore the angular distribution can be recovered by summing over all decay possibilities. Since the $K^\pm$ spin is equal to zero, it follows that $\lambda_{K^+} - \lambda_{K^-} = 0$. In addition, in the approximation of massless leptons only opposite helicities couple, hence $\alpha \equiv \lambda_{l^+} - \lambda_{l^-} \neq 0$. Defining $H_\lambda \equiv \langle J/\psi(\lambda)\phi(\lambda)|H|B^0\rangle$, (A.7) becomes

$$
\begin{aligned}
& \frac{1}{\Gamma}\frac{\mathrm{d}^4\Gamma(B\to J/\psi(\to l^+l^-)\phi(\to K^+K^-))}{\mathrm{d}\cos\theta_l \mathrm{d}\varphi_l \mathrm{d}\cos\theta_K \mathrm{d}\varphi_K} \\
\propto\; & \sum_{\alpha=\pm1} \left| \sum_{\lambda=\pm1,0} D^{1*}_{\lambda,\alpha}(\varphi_l,\theta_l,0) D^{1*}_{-\lambda,0}(\varphi_K,\theta_K,0) H_\lambda \right|^2.
\end{aligned} \tag{A.8}
$$

Substituting the rotation matrices gives

$$\begin{aligned}
&\frac{1}{\Gamma}\frac{\mathrm{d}^4\Gamma(B\to J/\psi(\to l^+l^-)\phi(\to K^+K^-))}{\mathrm{d}\cos\theta_l\mathrm{d}\varphi_l\mathrm{d}\cos\theta_K\mathrm{d}\varphi_K}\\
\propto\quad &\sum_{\alpha=\pm1}\left|\sum_{\lambda=\pm1,0} e^{+i\lambda(\varphi_l-\varphi_K)}d^1_{\lambda,\alpha}(\theta_l)d^1_{-\lambda,0}(\theta_K)H_\lambda\right|^2\\
=\quad &\sum_{\alpha=\pm1}\left|H_0 d^1_{0,\alpha}(\theta_l)d^1_{0,0}(\theta_K)+H_+e^{+i(\varphi_l-\varphi_K)}d^1_{1,\alpha}(\theta_l)d^1_{-1,0}(\theta_K)+H_-e^{-i(\varphi_l-\varphi_K)}d^1_{-1,\alpha}(\theta_l)d^1_{1,0}(\theta_K)\right|^2.
\end{aligned}\tag{A.9}$$

The H_0 terms originate from configurations in which the spin is directed perpendicular to the direction of motion of the vector mesons, the $H_{+,-}$ terms originate from configuration in which the spin is directed parallel, anti-parallel respectively.

In general a decay into four particles can be described by three angles. The angles $\varphi_{l,K}$ are unphysical, and are replaced by the physical angle φ which is defined as the angle between two decay planes

$$\varphi \equiv \varphi_l - \varphi_K. \tag{A.10}$$

Using the relation

$$d^j_{m',m} = (-1)^{m-m'}d^j_{m,m'} = d^j_{-m,-m'}, \tag{A.11}$$

and noticing that

$$ze^{i\varphi} + cc = 2(\mathrm{Re}z\cos\varphi - \mathrm{Im}z\sin\varphi), \tag{A.12}$$

it follows that the decay rate finally becomes

$$\begin{aligned}
&\frac{1}{\Gamma}\frac{\mathrm{d}^3\Gamma(B\to J/\psi(\to l^+l^-)\phi(\to K^+K^-))}{\mathrm{d}\cos\theta_l\mathrm{d}\cos\theta_K\mathrm{d}\varphi}\\
\propto\quad &|H_0|^2\sin^2\theta_l\cos^2\theta_K\\
&+\frac{1}{4}|H_+|^2\sin^2\theta_K(1+\cos^2\theta_l)\\
&+\frac{1}{4}|H_-|^2\sin^2\theta_K(1+\cos^2\theta_l)\\
&-\frac{1}{2}[\mathrm{Re}(H_+H_-^*)\cos2\varphi - \mathrm{Im}(H_+H_-^*)\sin2\varphi]\sin^2\theta_l\sin^2\theta_K\\
&+\frac{1}{4}[\mathrm{Re}(H_0(H_+^*+H_-^*))]\cos\varphi\sin2\theta_l\sin2\theta_K\\
&+\frac{1}{4}[\mathrm{Im}(H_0(H_+^*-H_-^*))]\sin\varphi\sin2\theta_l\sin2\theta_K.
\end{aligned}\tag{A.13}$$

The angles $\theta_{l,K}$ are helicity angles (the angles of the charged daughter particles with respect to the direction of the meson in the meson rest frame) and φ is the angle between the two decay planes.

Transversity conventions

Since the decay channel is CP-violating it is convenient to use the transversity conventions: choose the polarization amplitudes to be the CP-eigenstates and choose a frame in which the CP-even and the CP-odd amplitudes behave differently.

Since the helicity amplitudes transform under CP as

$$H_+ \to H_-,\ H_- \to H_+, \tag{A.14}$$

it follows that the polarization amplitudes

$$A_0 \equiv H_0,\ A_\parallel \equiv \frac{H_+ + H_-}{\sqrt{2}}, \tag{A.15}$$

are CP-even (L=0,2) eigenstates, whereas

$$A_\perp \equiv \frac{H_+ - H_-}{\sqrt{2}} \tag{A.16}$$

is a CP-odd (L=1) eigenstate. The proper time behavior of these amplitudes is derived in chapter 2.

As explained in chapter 2 the CP properties become more apparent in the transversity frame. This frame can be reached by transforming as

$$\begin{aligned} \sin\theta_l \sin\varphi &= \cos\theta_{tr} \\ \sin\theta_l \cos\varphi &= -\sin\theta_{tr}\sin\phi_{tr} \\ \cos\theta_l &= -\sin\theta_{tr}\cos\phi_{tr} \\ \theta_K &= \psi_{tr}. \end{aligned} \tag{A.17}$$

Since in this thesis only the transversity frame is used, the subscript 'tr' is omitted. With the transformations of both the amplitudes and the frame, the formulas in chapter 2 are finally obtained.

A.2 Transformations between different transversity conventions

Different conventions are used for the transversity frame. The definitions of the decay frames can be separated in two different groups. The conventions used by CLEO, Dighe, Dunietz, Fleischer, and LHCb on the one hand and the convention used by BaBar and DESY on the other hand. The convention used in this this thesis is the first one. The frame used in the convention used by Babar and DESY can be found in figure A.1. The conventions are

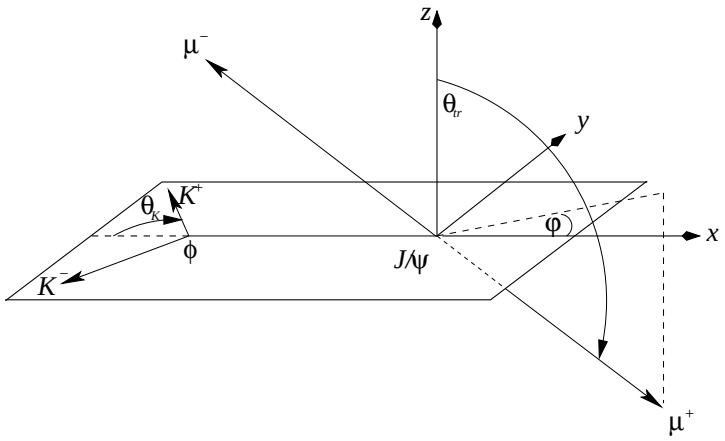


Figure A.1: *The transversity frame in the conventions used by Babar and DESY.*

related by the following transformations

$$\begin{aligned} x \to -x \quad &, \quad z \to -z; \\ \phi \to \pi - \phi \quad &, \quad \theta \to \pi - \theta_{\mathrm{tr}}; \\ f_4 \to -f_4 \quad &, \quad f_5 \to -f_5. \end{aligned} \tag{A.18}$$

It should be noted that in the two conventions the sign of the angular functions f_4 and f_5 are different.

Appendix B

Specific aspects of event generation for $B_s \to V_1 V_2$ decays

The decay of a $B_{(s)}$ meson into a final state is simulated by the EvtGen package [34]. Given a set of input parameters it calculates the probability to observe a given set of observables, and generates events accordingly. For $B_s^0 \to J/\psi\phi$ the calculation of the angular dependence of the distribution is performed using the scalar-to-vector-vector helicity amplitude (EvtSVVHelAmp) algorithm of EvtGen, the same as is used for $B_d^0 \to J/\psi K^*$. For the $B_d^0 \to J/\psi K^*$ decay the proper time and angular distributions factorize, therefore the simulation of the distributions is done independently.

However, for $B_s^0 \to J/\psi\phi$, in general $\Delta\Gamma \neq 0$. In that case the angular distribution evolves with proper time. This is caused by the fact that the different polarization amplitudes have different lifetimes and angular distributions. This can be understood in the following way. Assuming that $\phi_s = 0$, the one-angle distribution is

$$\begin{aligned} \frac{\mathrm{d}^4\Gamma}{\mathrm{d}t\mathrm{d}\theta} &\propto e^{-(\Gamma+\Delta\Gamma/2)t}(|A_0(0)|^2 + |A_\parallel(0)|^2)(1+\cos^2\theta)/2 \\ &+ e^{-(\Gamma-\Delta\Gamma/2)t}|A_\perp(0)|^2(1-\cos^2\theta). \end{aligned} \tag{B.1}$$

This means that the amplitudes distributed as $1+\cos^2\theta$ have a lifetime $\tau_L = 1/(\Gamma+\Delta\Gamma/2)$, whereas the amplitude distributed as $1-\cos^2\theta$ has a lifetime $\tau_H = 1/(\Gamma-\Delta\Gamma/2)$. As time evolves, the relative fraction of the different amplitudes therefore changes, and hence the observed angular distribution.

The angular distribution for different slices of time is shown in Fig. B.1. Due to a non-zero width difference the angular distribution is different for individual time bins and as a consequence the simulation can no longer factorize. Hence for the calculation of the proper time-dependent amplitudes a new model is developed.

The EvtPVVCPLH model takes care of this calculation, distributing for every generated time the correct fraction of polarization amplitudes to EvtSVVHelAmp. This model then generates the appropriate angular distribution.

The model allows for CP violating time asymmetries, as well as for different lifetimes for the different mass eigenstates: the Light and Heavy state, and both decay widths can be chosen to be the largest. The model is particularly intended to describe $P \to V_1 V_2$ decays

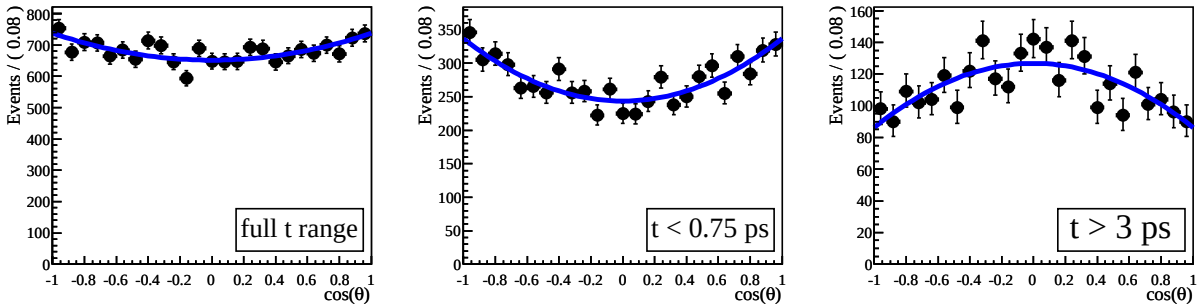


Figure B.1: *The angular distribution of* $\cos\theta$ *for different time windows. The angular distribution integrated over time (a) is different from the distribution at small times (b) and the distribution for large times (c).*

where the pseudo scalar has a nonzero width difference between the heavy and the light eigenstate, for example $B_s^0 \to J/\psi\phi$.

B.1 Functionality

The decay model works as follows. First of all the flavor of the 'other' or 'tagging' B at production is determined. This is used to determine the relevant flavor at production of the B or $\bar{B}$ meson decaying to $J/\psi\phi$. Since the meson decays to a CP-eigenstate, this flavor mixing of the produced B is irrelevant[1]. Therefore the (time dependent) mixing behavior beforehand generated by EvtGen by default, is ignored.

To generate the correct proper time dependent behavior correctly and efficiently, a proper time distribution with an envelope lifetime equal to $\tau_{\text{envelope}} \equiv 1/(\Gamma - |\Delta\Gamma|/2)$ is generated. Here $\Gamma - |\Delta\Gamma|/2$ is the smallest decay width of one of the mass eigenstates, leading to the longest possible lifetime. This is chosen such that the generation of the proper time distribution can be performed most efficiently. This is schematically shown in Fig. B.2.

Then the transversity amplitudes are calculated *with respect to* this envelope lifetime probability. Defining

$$g_{\pm} = \frac{1}{2}\left(\mathrm{e}^{-(im_{\mathrm{L}}+\Gamma_L/2)t} \pm \mathrm{e}^{-(im_{\mathrm{H}}+\Gamma_H/2)t}\right) , \tag{B.2}$$

with $\Gamma_H = \Gamma - \Delta\Gamma/2, \Gamma_L = \Gamma + \Delta\Gamma/2$, the proper time-dependent transversity amplitudes read

$$A_i(t) = A_i(0)[g_+(t) + \lambda_f g_-(t)] , \tag{B.3}$$

where $i \in (0, \|, \perp)$ and

$$\lambda_f = \eta \mathrm{e}^{+2i\beta_s} . \tag{B.4}$$

For the CP-even amplitudes $A_{0,\|}$ the CP eigenvalue equals $\eta = 1$ and for the CP-odd amplitude $A_\perp$ the CP eigenvalue equals $\eta = -1$.

[1] If a B meson decays to a CP-eigenstate, both an amplitude with and without mixing contribute. Like in quantum mechanics it is impossible to say if one decay happened via mixing or not.

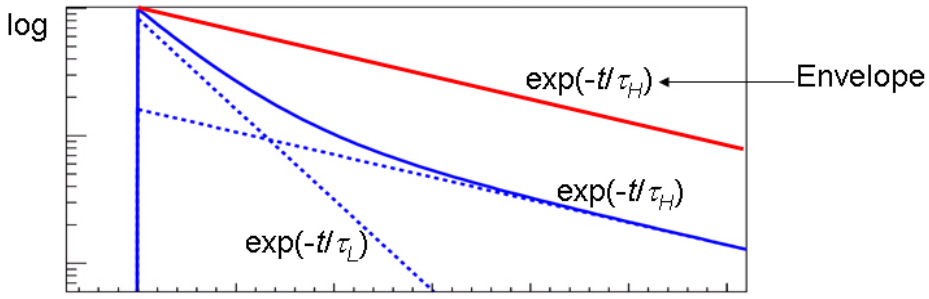


Figure B.2: *The envelope lifetime distribution, for the case* $\Delta\Gamma > 0$*, for which* $\tau_H > \tau_L$*. Dotted are shown the CP-even and CP-odd fractions, which in the case of no CP violation have lifetimes* τ_L*,* τ_H *respectively. For efficiency reasons firstly this envelope distribution is generated in EvtPVVCPLH. The amplitudes are then calculated with respect to the amplitude of this envelope rate.*

Since the envelope lifetime distribution has already been generated, it now suffices to calculate the amplitudes $A_i'(t)$, which are defined with respect to the amplitude of the envelope distribution as

$$A_i'(t) \equiv A_i(t)/\sqrt{\mathrm{e}^{-t/\tau_{\mathrm{envelope}}}} = \mathrm{e}^{(\Gamma/2-|\Delta\Gamma|/4)t}A_i(t). \tag{B.5}$$

These then equal

$$\begin{aligned} A_i(t) &= A_i(0)e^{-iMt}e^{-\Delta\Gamma t/4} \\ &\quad \times[(e^{-(-i\Delta m/2+\Delta\Gamma/4)t} + e^{-(i\Delta m/2-\Delta\Gamma/4)t}) + \eta e^{2i\beta_s}(e^{-(-i\Delta m/2+\Delta\Gamma/4)t} - e^{-(i\Delta m/2-\Delta\Gamma/4)t})], \end{aligned}$$

or

$$\begin{aligned} A_i(t) &= A_i(0)e^{-iMt} \\ &\quad \times[(e^{(i\Delta m/2)t} + e^{-(i\Delta m/2-\Delta\Gamma/2)t}) + \eta e^{2i\beta_s}(e^{(i\Delta m/2)t} - e^{-(i\Delta m/2-\Delta\Gamma/2)t})]. \end{aligned} \tag{B.6}$$

The expressions in (B.6) are the formulas implemented in EvtPVVCPLH. These amplitudes are distributed to EvtSVVHelAmp *for each event*, in order to estimate the correct angular distribution for the decay time of *that* event. Determining this angular distribution for every proper time, and accepting/rejecting accordingly, finally leads to the correct proper time and angular dependent distribution.

The method can be summarized as follows:

1. First events are generated with an envelope lifetime distribution.
2. For every event the polarization amplitudes are calculated for the generated time, and with respect to the envelope amplitude. The polarization amplitudes for that time are distributed to EvtSVVHelAmp.
3. Events are finally accepted/rejected with the proper time and angular rate, calculated with respect to the already generated envelope distribution.

B.2 Usage of the EvtGen decay file

The usage in the decay file is the following:

BrFr V1 V2 PVV_CPLH beta_s eta $|A_\parallel(0)|$ $\delta_\parallel$ $|A_0(0)|$ δ_0 $|A_\perp(0)|$ $\delta_\perp$

Here the different arguments represent the following parameters.

The first term is the branching fraction; the second and third term are the names of the two decay daughter vectors; the fourth argument is the name of the decay mode; the fifth argument is beta_s $\equiv \beta_s$, the relevant CKM angle in radians.

The sixth argument eta $\equiv \eta$ is the CP-eigenvalue of the final state. Since the final state is a mixture of CP-eigenstates this argument is not used.

The last six arguments (7-12) are the absolute values and phases of the polarization amplitudes in the transversity basis $A_\parallel$, A_0, and $A_\perp$, and their respective strong phases δ_i, at proper time equal to zero. Only the differences between the strong phases are physical.

The width difference $\Delta\Gamma_s$, possibly negative, and the mass difference, Δm_s, are not input parameters to the model. They are determined via the definition of deltaGamma and deltaMs in e.g. DECAY.DEC, in which the general settings of all simulations are defined.

Example

The example below generates B_s meson decays to the two vector mesons J/ψ and ϕ. Input values are the SM value of $\beta_s = 0.02$ rad, equivalent to $\phi_s = -0.04$ rad, and the approximate world averages of the polarization amplitudes: $|A_\parallel(0)|^2 = 0.24$, $|A_0(0)|^2 = 0.60$ and $|A_\perp(0)|^2 = 0.16$, with the phases $\delta_\parallel = 2.50$ rad, $\delta_\perp = -0.17$ rad and $\delta_0 \equiv 0$, which are the values used in the LHCb event generation.

```
Decay B_s0
   1.000  J/psi  phi  PVV_CPLH  0.02  1  0.49  2.50  0.775  0.0  0.4  -0.17
Enddecay
```

Versions

The above decay is used in the decay file 'Bs_Jpsiphi,mm=CPV.dec', starting from DecFiles/dkfiles version v13r9. The Gen/EvtGen code of EvtPVVCPLH is used since version v8r13 (also in DC06). Higher order terms have been corrected in November 2007 and are used in versions > v8r16.

Notes

The code is based on SVV_CPLH by Anders Ryd.

Appendix C

Outer Tracker Quality Assurance

The quality of the Outer Tracker modules and the their front end (FE) electronics is by applying several tests. In this section the following quality tests are described: final validation of modules with ^{90}Sr directly after their production; signal response of individual channels with ^{55}Fe after installation; module gas tightness and gas propagation after installation; performance test of FE electronics after assembly.

C.1 Validating the Modules with Radioactive Sources

The response of each straw channel of the OT modules is checked with a radioactive ^{90}Sr 2.3 MeV β source, as described in [37]. The full module is irradiated by a line source in steps of 1 cm along its length and the corresponding wire current is measured and recorded. This procedure allows to check the uniformity of the response of the wires.

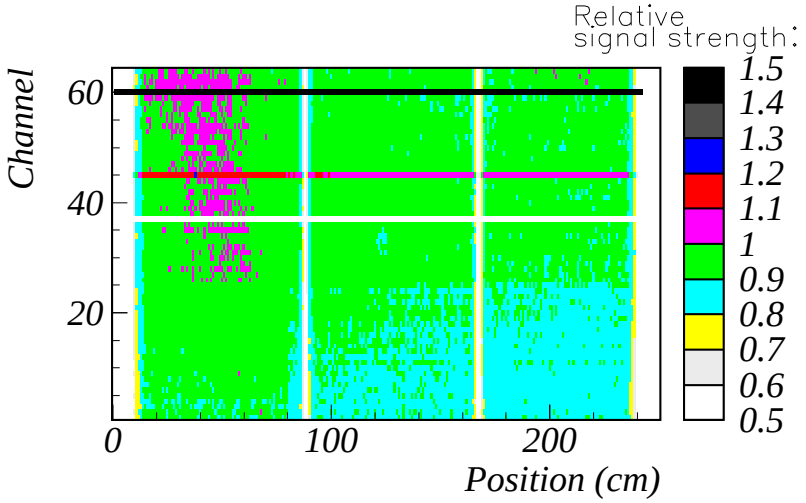


Figure C.1: *Example of a scan with a ^{90}Sr source of module S1L-065, side A. Wire locators and wires with low and high currents are clearly visible.*

After correcting for the source profile, module scans, like the one in figure C.1, are found. The presence of wire locators in the detectors is clearly visible as vertical lines. Light horizontal lines indicate channels with low currents, possibly caused by bad gas flow; dark lines correspond to channels with either shorts or high currents.

Flushing a module over a period of the order of a few weeks improves the response and makes the response of the module uniform to the level of ±5%. Disregarding the 10 worst modules out of 132, a total of approximately 27 individual bad channels are identified [40], corresponding to 0.1% of the total.

C.2 ^{55}Fe pulse

To verify proper operation of the detector after transport and installation, the individual straws of the OT are tested in situ using a ^{55}Fe 5.9 keV photon source [41]. The average pulse height of 128 pulses is recorded; only the A-side of the detector is (almost fully) tested. Since the gain in the straw depends on many parameters, which have not been recorded during the measurements, the test should be considered as a functionality test of the straws,

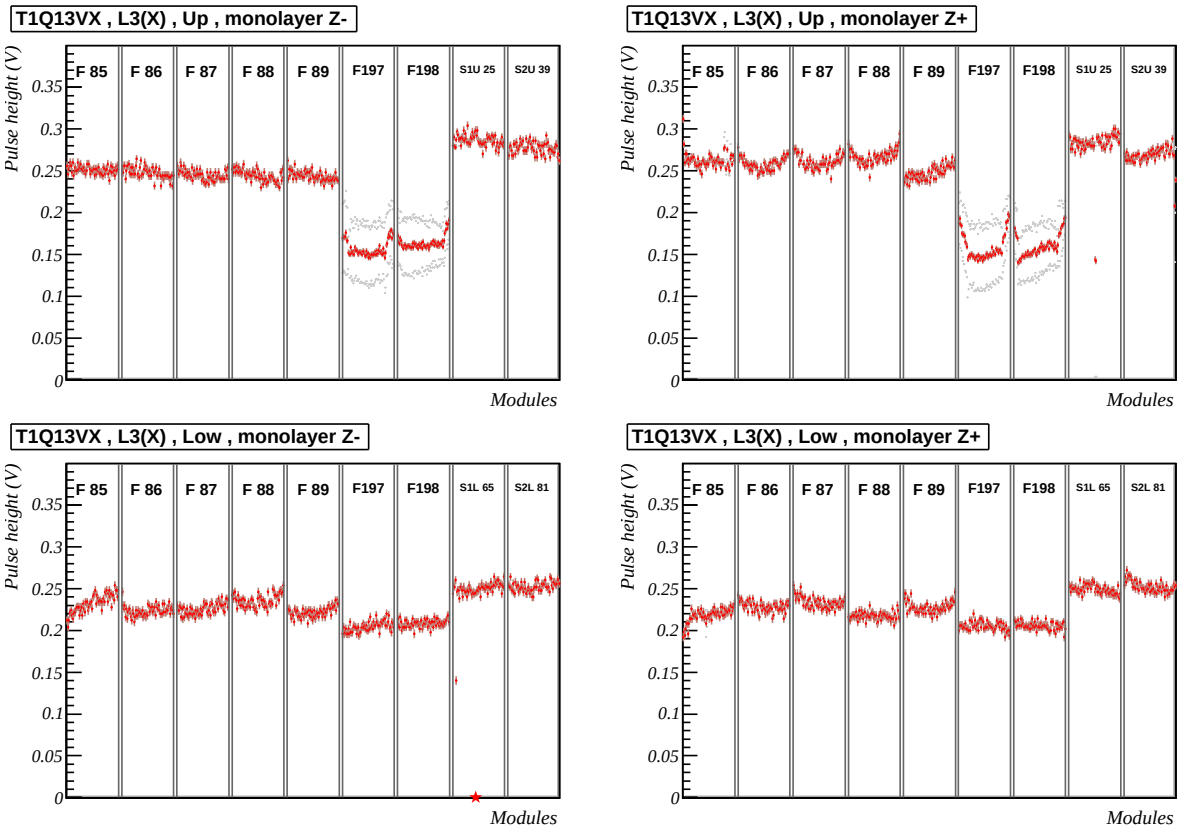


Figure C.2: *The individual (light) and average (dark) ^{5}5Fe pulse heights of 2304 straws of the upper and lower modules of the X layer of station T1-Q13-VX. The pulse height in modules F197 and F198 increased at the second measurement, presumably due to reduced gas flow.*

rather than a quantitative measurement of the straw performance.

The results, partly shown in figure C.2, demonstrate that 32 out of 27648 channels are below expectation, i.e. approximately 0.1%. Most of the problematic channels were already identified at the production sites. For example, the known noisy channel of module S1L-065, shown in figure C.1, appears as a channel with half the pulse height. In addition the known dead channel in S1L-065, shown in C.1, is identified.

The low pulse height in modules F197 and F198 is presumably caused by a wrong gas mixture, before reaching its equilibrium of Ar/CO_2 70/30, which is further investigated in the next section.

C.3 Gas tightness and gas propagation

To guarantee a proper behavior of the OT, the gas tightness and the gas propagation have individually been tested for each module.

Gas tightness

The gas tightness of each detector module is confirmed directly after installation. The pressure drop after applying an overpressure is measured for the total of installed modules in a C frame after each installation a module. The gas tightness is then calculated per module. Because the gas tightness of the full gas line connected to one module layer was measured, the gas-tightness of an individual module was corrected for the previously installed modules. This explains the occasional negative value shown in figure C.3.

Overall, an excellent agreement is observed between the measurements after production and at installation, showing that no additional leaks are introduced during the installation. The discrepancy of the gas-tightness results in T1-Q02-L1 is attributed to a wrong assumption of the installation sequence of the modules. For details on the gas tightness requirements, see [42].

Gas propagation

The OT modules are mounted vertically on a C-frame, with the gas input at the bottom, and the output at the top. In order to monitor gravitational effects on the gas flow, and to understand low pulse heights in the ^{55}Fe measurements, the vertical gas flow is measured during installation. This is done by measuring the pulse height using a ^{55}Fe source. The signal is monitored at various times, after flushing air filled modules at $t = 0$ with Ar/CO_2 70/30.

Measuring the pulse height in four modules at 15 cm distance from the gas input gives the results in figure C.4. All four modules show proper pulse heights after approximately 40 minutes. The total gas flow through the 7 F-modules is 0.5 l/min. From this it is concluded that the C-frame behaves as expected with the gas being distributed equally through the four modules.

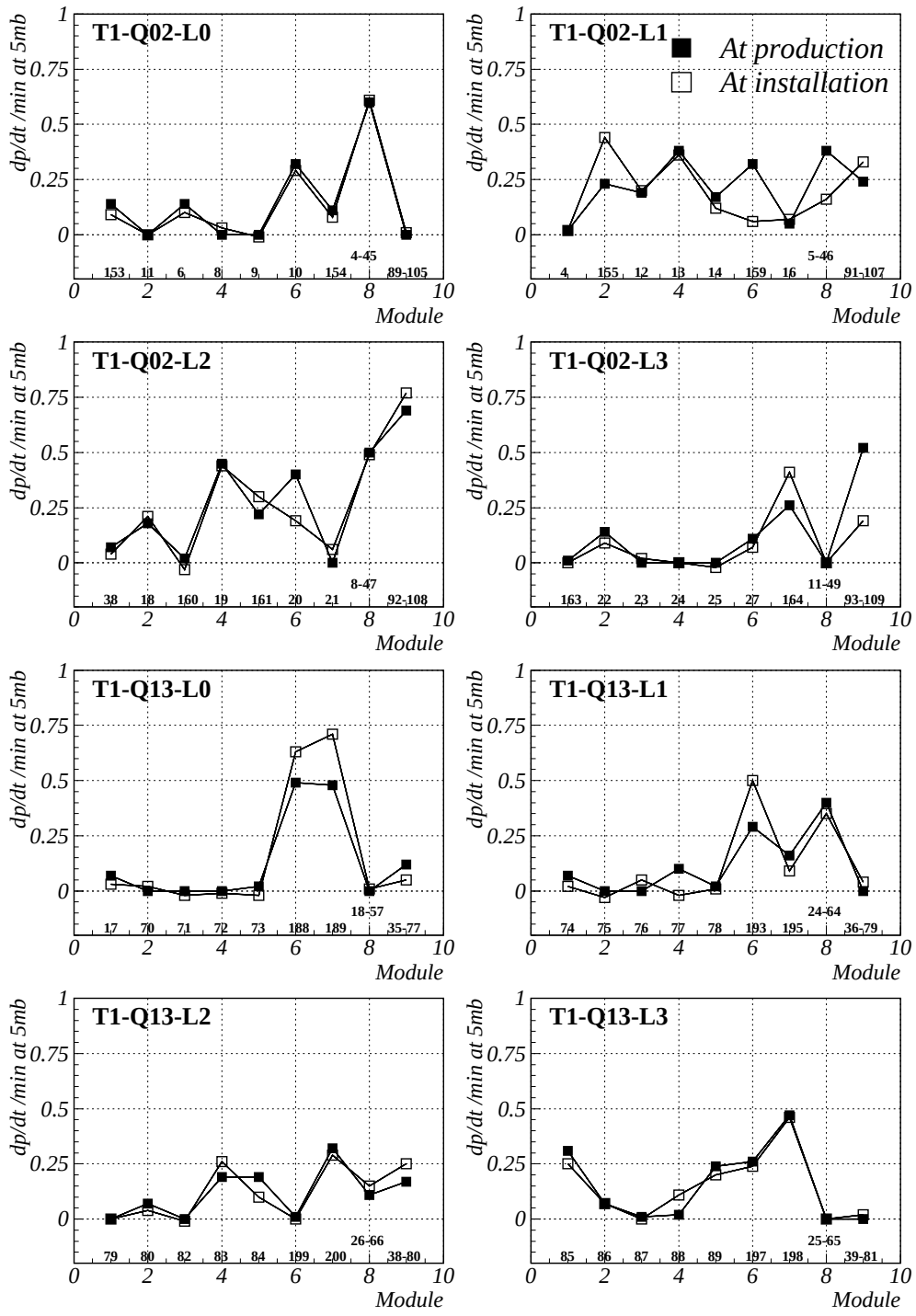


Figure C.3: *The pressure decay for the modules of station T1 are given, both the value after production, and at installation. Also the modules' serial number is shown.*

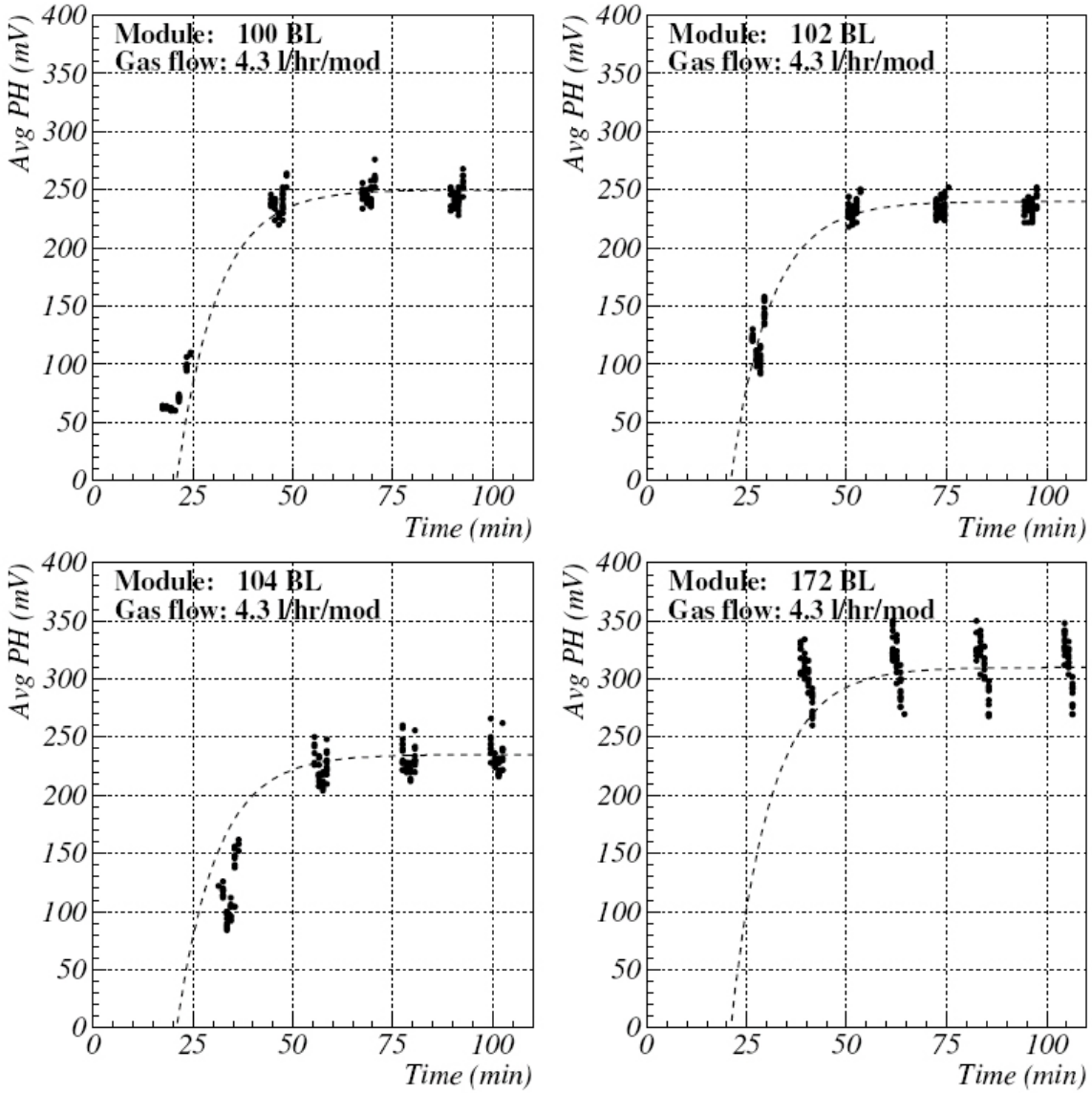


Figure C.4: *The four modules show proper pulse heights about 40 minutes after filling the C-frame with gas. The dotted line* $PH(t) = PH(\infty)(1 - e^{-0.1(t-21min)})$ *guides the eye.*

C.4 Tests of FrontEnd-boxes

The long OT modules are read out from both ends, where the so-called FrontEnd-box or FE-box is placed, in which several interconnected boards are housed. The main components of the OT readout electronics are the High Voltage board, the ASDBLR board, the OTIS board and the GOL board, as shown in figure C.5.

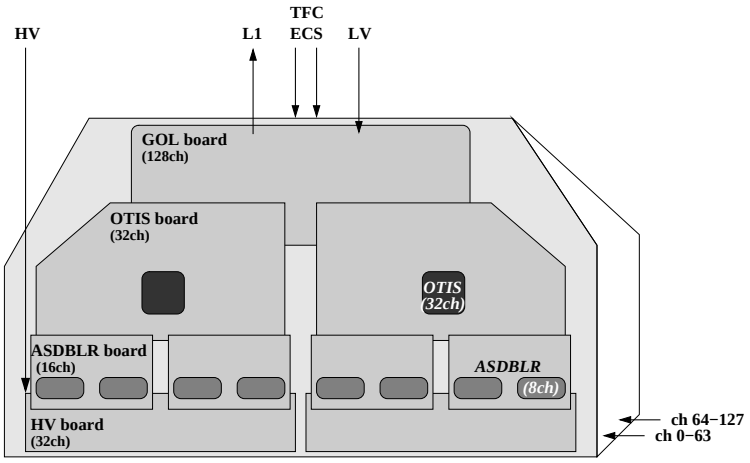


Figure C.5: *The layout of the OT readout electronics inside the FE-box.*

After assembly the FE-box is tested for threshold uniformity, cross talk, linearity and noise. The test procedure of the FE-box electronics is described in [38]. During assembly and testing a particular feature in the noise measurements was observed. In figure C.6 the number of detected hits is measured, when applying no test pulse. the results are shown. It appears that for both a threshold of 72 DAC and a threshold of 76 DAC the same patterns appears: the odd channels in the region of channel number 33-48 show noisy behavior.

This can be explained in the following way [43]. The OTIS uses the LHC clock to digitize the hit times in the OT, inducing a varying magnetic field, and thereby a current, in conductors. A test pulse line runs from the input of the GOL board at the OTIS board, over the OTIS board along the OTIS chip, to the ASDBLR channels. The test pulse line running closest to the chip, unshielded from the field of the clock signal from the OTIS, is the line feeding the test pulse to the odd channels of the left ASDBLR at the right OTIS board. This explains why the odd channels in the region 33-48 appear noisy.

This problem is solved by shielding the OTIS chip with a grounded plate. As a result the clock signal from the OTIS is no longer picked up by the test pulse line. The tests after this shielding are shown in figure C.7, where it is seen that the noise has disappeared.

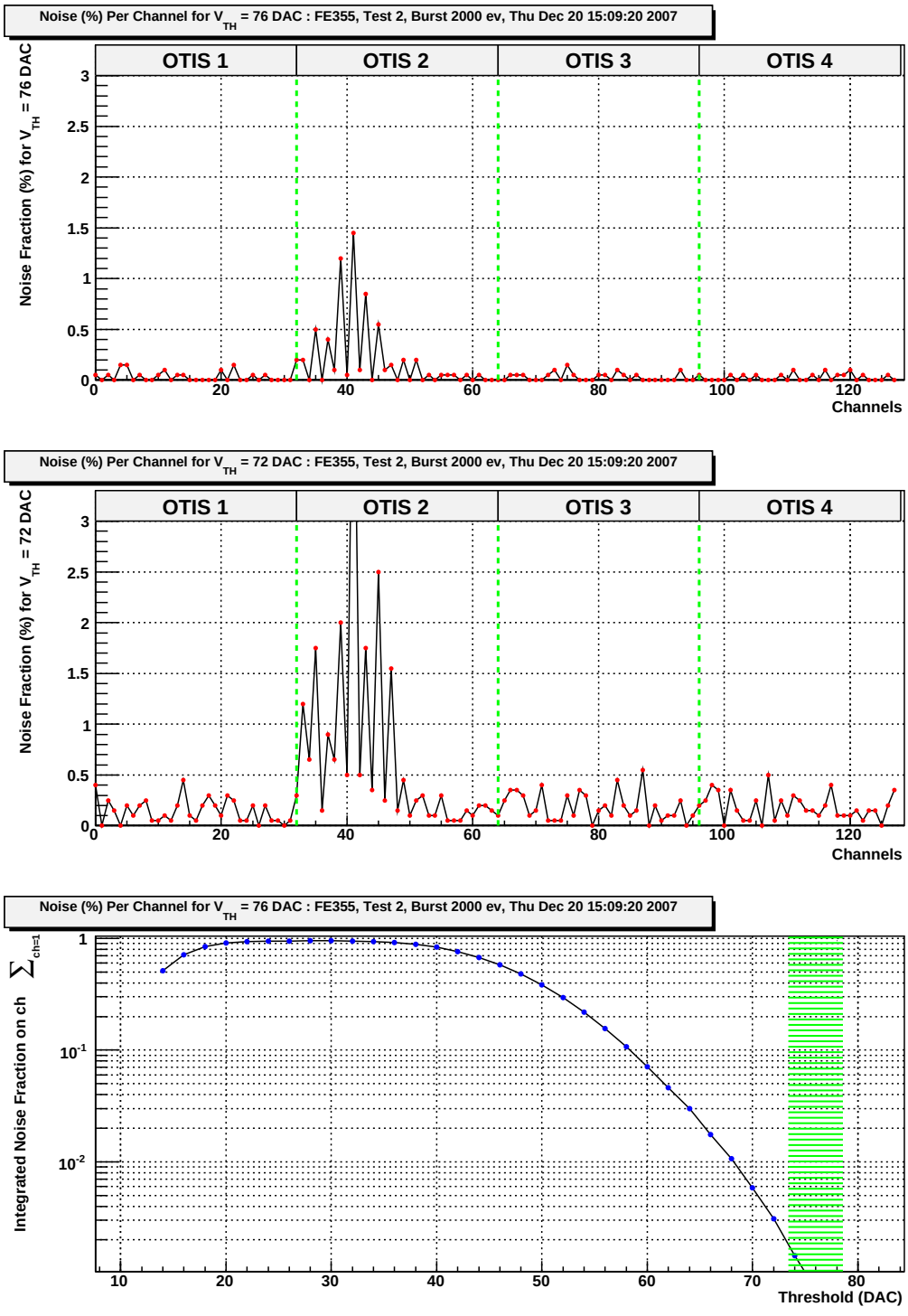


Figure C.6: *Noise scan without shielding. From top to bottom are shown: noise scan with high threshold (76 DAC), low threshold (72 DAC) and the integrated noise. Without shielding of the OTIS chip the odd channels in the region 33-48 appear to be noisy.*

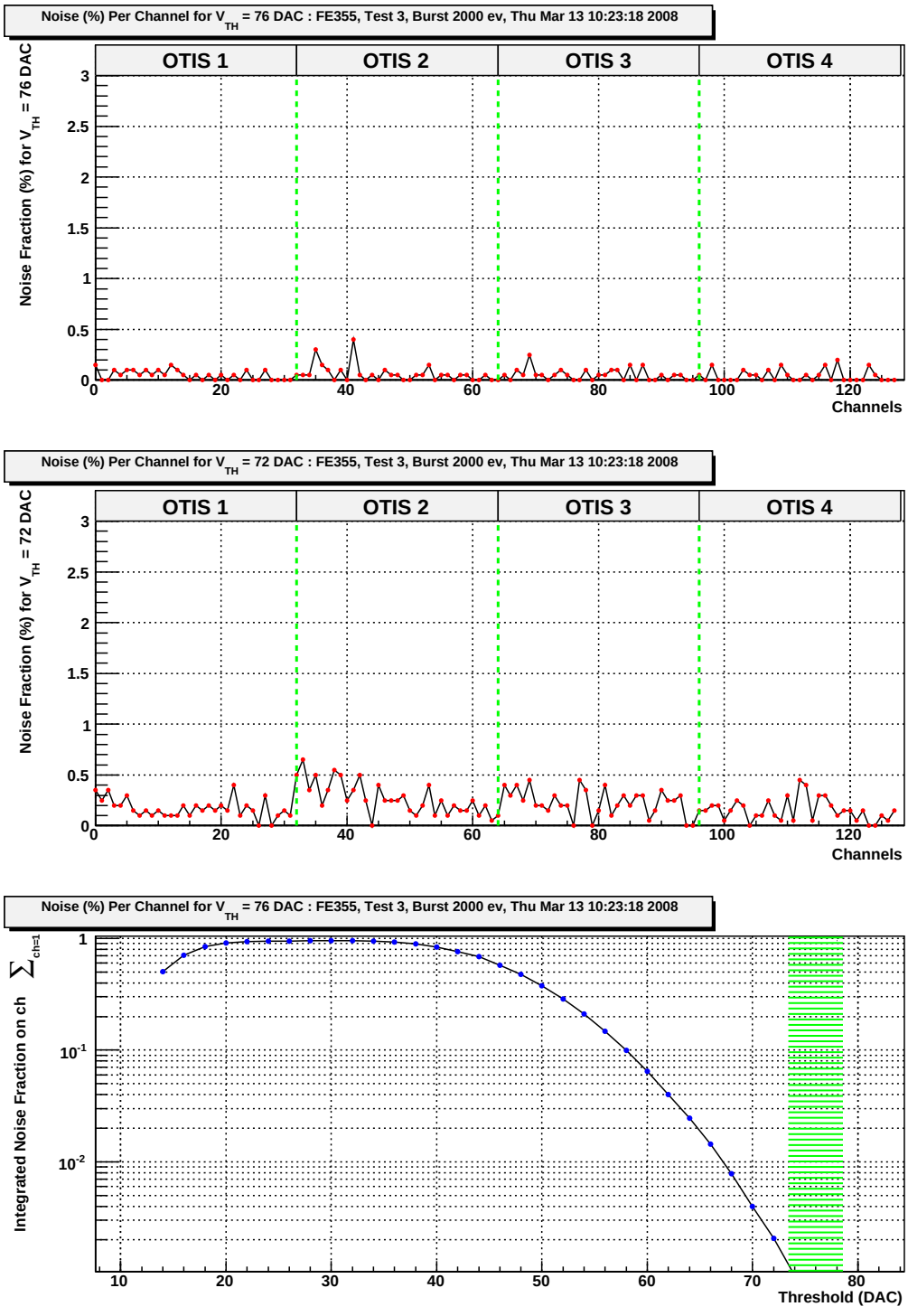


Figure C.7: *Noise scan with shielding. From top to bottom are shown: noise scan with high threshold (76 DAC), low threshold (72 DAC) and the integrated noise. After shielding the OTIS chip the noise in the odd channels in the region 33-48 has disappeared.*

Appendix D

Implementation of the fit methods

As shown in the chapter 2, the total decay description of the channels $B_s^0 \to J/\psi\phi$ and $B_d^0 \to J/\psi K^*$ can be written as a normalized sum of products of time and angular dependent functions. To correctly describe the distribution of signal events, experimental effects should be incorporated in the description of the PDF. Effects which should be included in the fit are inefficiencies and resolutions in the angular, proper time and tagging observables. Furthermore, backgrounds should be included in the description of data[1].

These requirements are general for pseudo-scalar to vector-vector (P2VV) decays. In addition are the angular and time dependence of other P2VV decays often similar. This means it should be possible to share analysis methods between different decays, using the functionality already provided by the RooFit fitting program [32] and to share the description of angular functions and their integrals between different decay channels. In order to share as much as possible these tools to construct PDFs for different channels in similar ways, the P2VVfit framework has been developed.

In this appendix the general scheme of the construction of the PDF is described: P2VVfit. This method can not only be applied to $B_s^0 \to J/\psi\phi$, but to general multi-dimensional PDFs, in particular to other $P \to V_1V_2$ decays of which proper time dependent angular distributions are analyzed. One advantage of the method is that the calculation of integrals is factorized as much as possible, making the fits and plots of the PDFs as fast as possible within RooFit. The intrinsic functionalities of RooFit to include efficiencies and resolutions are used as much as possible.

RooFit

RooFit [32] is a package of classes inside Root [33] which makes it possible to build PDFs, fit these to data, display them, and generate toy MC data sets according to the PDFs. Basic classes are RooRealVar (a real variable), and RooAbsPdf (an abstract pdf class), which both inherit from RooAbsReal (an abstract class for real functions and variables).

[1]The analysis is performed on reduced Data Storage Tapes (DSTs), so called microDSTs. These files are in the canonical LHCb event model format. To reduce storage space by a factor $O(100)$ only the information most relevant for the final analysis is copied from the DST to the microDST. This contains reconstructed particles, trigger decision reports or reconstructed vertices. Information which is not used in the final analysis is not stored. Examples are detailed track information or the particles that are not included in the decay chain.

RooFormulaVar is the RooFit equivalent of a formula and also inherits from RooAbsReal. A RooCategory is a class to provide integer valued variables, such as the variable with different values for different (tagging) categories. Finally, a RooResolutionModel is a general class for resolution models, possibly a sum of individual resolution models.

D.1 The signal PDF

The classes that are developed in addition to the RooFit classes are the base class modelling the angular dependence, and derived classes modelling specific cases. The latter are, for example, the angular distribution functions f_i in the transversity frame, called TransAngles, or those in the helicity frame, called HelAngles.

These angular basis functions are subsequently multiplied with time-dependent basis functions (provided by RooFit), the appropriate coefficients (RooFormulaVars) and a possible efficiency function (a RooAbsReal). The multiplication is performed using a RooProduct, which returns the product of the functions that are put in and recognizes the factorization of the angular and time dependent functions.

The PDF is constructed as schematically shown in figure D.1. The PDF is a RooRealSumPdf, a PDF class which inherits from RooAbsPdf. A RooRealSumPdf takes as input a list of functions and returns the normalized sum. The separate input functions need not be normalized or positive[2]. One element of the sum is a product of an angular and a proper time dependent functions, multiplied with the relevant coefficients.

The components of the PDF

The input of the RooRealSumPdf is a list of $n_t \times n_f$ RooProducts. Here n_f is the number of angular functions and n_t is the number of proper time dependent basis functions. Both numbers depend on the number of required basis functions, which differ for different decay channels and different frame conventions. Writing

$$\frac{1}{\Gamma}\frac{\mathrm{d}^4\Gamma}{\mathrm{d}t\mathrm{d}\Omega} = \frac{A_i(t,q|\vec{\lambda})f_i(\vec{\Omega})}{\int A_j(t,q|\vec{\lambda})f_j(\vec{\Omega})\mathrm{d}t\mathrm{d}\Omega}, \tag{D.1}$$

where the product can be written in terms of the components as

$$A_i(t,q|\vec{\lambda})f_i(\vec{\Omega}) = c_{ij}(q|\vec{\lambda})t_i(t)f_j(\vec{\Omega}). \tag{D.2}$$

Here $t_i(t)$ are the proper time basis functions, $f_j(\vec{\Omega})$ the angular functions and $c_{ij}(q|\vec{\lambda})$ are the relevant coefficients in front of every product. The coefficients in general depend on the physics parameters $\vec{\lambda}$ and a tagging category q. Every $(ij)^{th}$ product is an element of the list. RooProduct keeps track of the factorization of the different components.

[2]Actually a RooRealSumPdf takes a list of functions (dependent of observables) and a list of parameters (independent of observables). Since for $B_s^0 \to J/\psi\phi$ the tagging observables appear in the coefficients and hence in all components of the PDF, all components are grouped in the list of functions, and the list of parameters is a dummy list of ones.

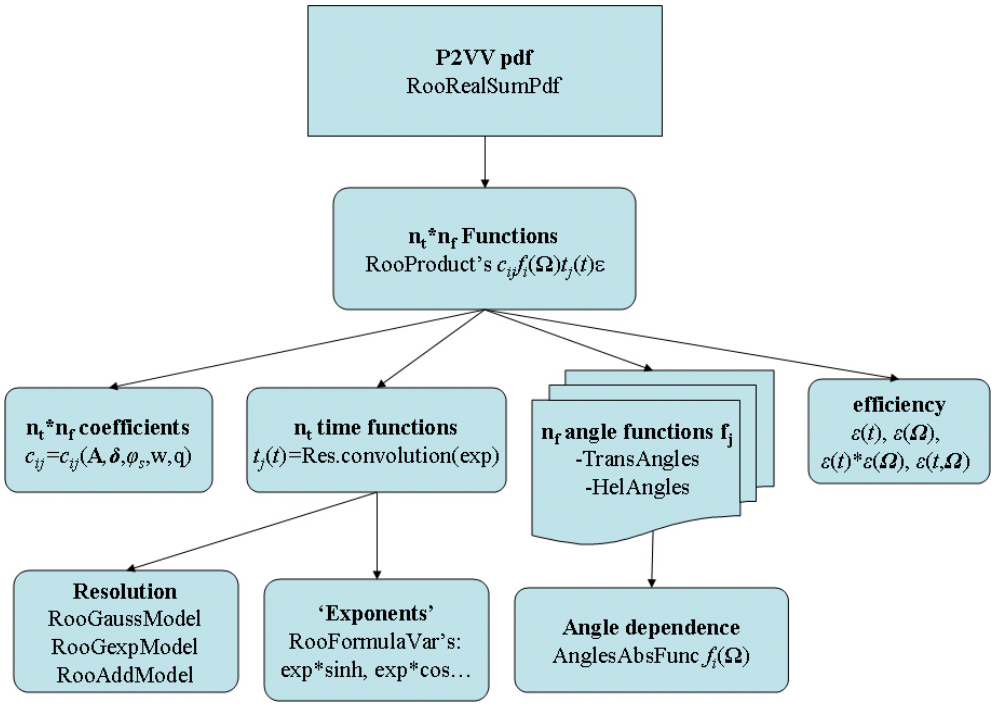


Figure D.1: *Schematic construction of the P2VV PDF. The PDF consists of a sum of $n_t \times n_f$ functions, the multiplication of the number of time and angular base functions. The functions depend on the observables and need not be normalized or positive. Every function appears with the appropriate coefficient. The PDF can be multiplied with an efficiency.*

The angular dependence

The only part of the PDF not incorporated in RooFit classes are the n_f angular functions $f_i(\vec{\Omega})$. The base class is AnglesAbsFunc, an abstract class for angular functions which inherits from RooAbsReal. Like any RooAbsReal it has a value and it can be analytically integrated, if the analytical integral is implemented. As input a set of angles is given and possibly a ξ_i. This ξ_i corrects for the angular acceptance by adjusting the normalization, as discussed in chapter 5.

An example of a specific implementation of the angular functions are the 6 angular functions in the $B_s^0 \to J/\psi\phi$ decay in the transversity basis as described in chapter **??**. These are implemented in the TransFi classes, which inherit from AnglesAbsFunc. In addition to the function values, also the analytical integral over the different angles is implemented. The HelFi classes are the equivalent of the TransFi classes, but in the helicity frame.

The proper time dependence

The used proper time functions $t_i(t)$ are base functions common in B physics and available in RooFit, namely $e^{-\Gamma t}\sin(\Delta m t), e^{-\Gamma t}\cos(\Delta m t), e^{-\Gamma t}\sinh(\Delta\Gamma t/2), e^{-\Gamma t}\cosh(\Delta\Gamma t/2)$. These functions can be convoluted with any RooResolutionModel. Examples are common resolutions like delta distributions, gaussians and combinations of these, and are implemented

analytically in RooFit.

The coefficients

The coefficients $c_{ij}(q|\vec{\lambda})$, which appear in products with the proper time and angular dependent functions, are RooFormulaVars. For $B_s^0 \to J/\psi\phi$ they depend on the tagging flavor observable and parameters like the polarization amplitudes, the strong phases, the mistag fraction, and the tagging flavor, and are given in chapter [**?**].

Inclusion of efficiencies

Finally, an efficiency function can be included. This can be any RooAbsReal: e.g. a function or a histogram. The following cases of increasing complexity are implemented:

1. the efficiency can be only a 1D proper time efficiency;
2. the efficiency can be a 3D angular efficiency;
3. the efficiency can be a product of an angular and a proper time efficiency (if the two efficiencies factorize);
4. the efficiency can be a 4D proper time and angular efficiency.

Introducing an efficiency $\varepsilon(t, \vec{\Omega})$, the PDF reads

$$\frac{\mathrm{d}^4\Gamma}{\mathrm{d}t\mathrm{d}\Omega} = \frac{A_i(t|\vec{\lambda}) f_i(\vec{\Omega}) \varepsilon(t, \vec{\Omega})}{\int A_j(t|\vec{\lambda}) f_j(\vec{\Omega}) \varepsilon(t, \vec{\Omega}) \mathrm{d}t\mathrm{d}\Omega}, \tag{D.3}$$

where the product is written in terms of the components as

$$A_i(t, q|\vec{\lambda}) f_i(\vec{\Omega}) \varepsilon(t, \vec{\Omega}) = c_{ij}(q|\vec{\lambda}) t_i(t) f_j(\vec{\Omega}) \varepsilon(t, \vec{\Omega}). \tag{D.4}$$

Finally, the product of the angular functions, the proper time functions, the coefficients, and possibly one or more efficiencies is calculated and normalized. RooProduct keeps track of the factorization of the different components.

1. In case of a 1D proper time efficiency, $\varepsilon(t, \vec{\Omega}) = \varepsilon(t)$, this integral is calculated numerically in one dimension.
2. In case of a 3D angular efficiency, $\varepsilon(t, \vec{\Omega}) = \varepsilon(\vec{\Omega})$, the angular normalization is numerically calculated *and* cached, as explained in chapter 5. The integral $\xi_i \equiv \int f_j(\vec{\Omega}) \varepsilon(\vec{\Omega}) \mathrm{d}\Omega$ is calculated only once and does not change during a likelihood maximization. This keeps the CPU performance of the program optimal. The same holds for any product of different factorizing angular efficiencies.
3. In case of a factorizing (3D) angular efficiency *and* a 1D proper time efficiency, $\varepsilon(t, \vec{\Omega}) = \varepsilon(t) \times \varepsilon(\vec{\Omega})$, the behavior is a combination of the two mentioned above.
4. If, finally, a general 4D proper time and angular efficiency $\varepsilon(t, \vec{\Omega})$ is given as input, 4D numerical integrations are performed when both fitting and plotting. This implies, however, a more CPU intensive, and therefore slower, program execution.

D.2 PDF for angular background

The only part of the background PDF chosen to be described as not-factorizable, is the angular dependence. This is written as

$$\frac{b_{abc}P_a(\cos\theta)P_b(\cos\psi)P_c(\phi)}{\int b_{abc}P_a(\cos\theta)P_b(\cos\psi)P_c(\phi)\mathrm{d}\Omega}. \tag{D.5}$$

This is implemented in a similar way as the signal PDF, namely:

$$\mathrm{RooRealSumPdf}(\mathrm{RooArgList}(\mathrm{RooProduct}(P_aP_bP_c)), \mathrm{RooArgList}(b_{abc})). \tag{D.6}$$

This means at most multiple 1D numerical integrals will be performed.

D.3 Three-dimensional angular efficiency

Finally, as described in chapter 5, the 3D angular efficiency can be decomposed as

$$e_{abc}P_a(\cos\theta)P_b(\cos\psi)P_c(\phi). \tag{D.7}$$

This is implemented in a similar way as the background PDF, namely:

$$\mathrm{RooAddition}(\mathrm{RooArgList}(\mathrm{RooProduct}(P_aP_bP_c)), \mathrm{RooArgList}(e_{abc})). \tag{D.8}$$

Bibliography

[1] A.D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asymmetry in the universe*, Phisma Zh. Eksp. Teor. Fiz. **5** (1967) 32.

[2] J.H. Christenson, J.W. Cronin, V.L. Fitch, and R. Turlay, *Evidence for the 2 π decay of the K_2^0 meson*, Phys. Rev. Lett. **13** (1964) 138.

[3] M. Kobayashi and T. Maskawa, *CP violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652.

[4] The DØ Collaboration, V. Abazov *et al.*, *Evidence for an anomalous like-sign dimuon charge asymmetry*, Submitted to Phys. Rev. **D**.

[5] The LHCb Collaboration, A. Alves *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[6] C. Amsler *et al.* (Particle Data Group), Physics Letters **B667** (2008) 1.

[7] Heavy Flavor Averaging Group, http://www.slac.stanford.edu/xorg/hfag/.

[8] P.F. Harrison, S. Dallison, and W.G. Scott, *The Matrix of Unitarity Triangle Angles for Quarks*, Phys. Lett. **B680** (2009) 328.

[9] CKMfitter group, J. Charles *et al.*, Eur. Phys. J. **C41** (2005) 1, updated results and plots available at http://ckmfitter.in2p3.fr/.

[10] CKMFitter group, O. Deschamps *et al.*, *CKM global fit and constraints on New Physics in the B meson mixing*, arXiv:0810.3139v2.

[11] C. Jarlskog, *Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal CP violation*, Phys. Rev. Lett. **55** (1985) 1039.

[12] L. Wolfenstein, *Parameterization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[13] P.F. Harrison and H.R. Quinn (editors), *The BaBar physics book*, SLAC-R-504 (1998).

[14] The ARGUS Collaboration, H. Albrecht *et al.*, *Measurement of the lifetime of bottom hadrons*, Phys. Rev. Lett. **51** (1983) 1316.

[15] V. Weisskopf, E. Wigner, *Over the natural line width in the radiation of the harmonius oscillator*, Z. Phys. **65** (1930), 18.

[16] I. Dunietz, R. Fleischer, and U. Nierste, *In Pursuit of New Physics with B_s decays*, Phys. Rev. **D63** (2001) 114015.

[17] A. Dighe, I. Dunietz, R. Fleischer, *Extracting CKM Phases and $B_s - \bar{B}_s$ Mixing Parameters from Angular Distributions of Non-Leptonic B Decays*, Eur. Phys. J. **C6** (1999) 647; I. Dunietz, H. Quinn, A. Snyder, W. Toki, and H. Lipkin, *How to extract CP-violating asymmetries from angular correlations*, Phys. Rev. **D43** (1991) 2193.

[18] J. Richman, *Les Houches summer school lectures* (1997).

[19] L. Fernandez, *Exclusive Trigger Selections and Sensitivity to the $B_s - \bar{B}_s$ Mizing Phase at LHCb*, Ph.D. thesis, Lausanne (2006).

[20] The LHCb Collaboration, B. Adeva *et al.*, *Road map for selected key measurements of LHCb*, LHCb-PUB-2009-029.

[21] D. Chang, A. Masiero, H. Murayama, *Neutrino mixing and large CP violation in B physics*, Phys. Rev. **D67** (2003) 075013; R. Harnik, D. Larson, H. Murayama, A. Pierce, *Atmospheric Neutrinos Can Make Beauty Strange*, Phys. Rev. **D69** (2004) 094024.

[22] V. Barger, L. Everett, J. Jiang, P. Langacker, T. Liu, and C. Wagner, *b to s Transitions in Family-dependent $U(1)'$ Models*, JHEP **0912** (2009) 048.

[23] R. Fleischer, *private communication.*

[24] H. Boos, T. Mannel, J. Reuter, *The Gold-plated mode revisited: $sin(2\beta)$ and $B^0 \to J/\psi K_S$ in the Standard Model*, Phys. Rev. **D70** (2004) 036006.

[25] S. Faller, R. Fleischer, T. Mannel, *Precision Physics with $B_s^0 \to J/\psi\phi$ at the LHC: The Quest for New Physics*, Phys. Rev. **D79** (2009) 014005.

[26] Stéphane T'Jampens, BaBar Thesis-03/016, Ph.D. thesis, Université Paris XI (18 Dec. 2002).

[27] Jeffrey D. Richman, *An Experimenter's Guide to the Helicity Formalism*, CALT-68-1148 (1984).

[28] Suh Urk Chung, *Spin Formalisms*, CERN **71-8** (1971).

[29] G. Ybeles-Smit, *Asymmetries in the decay of beauty*, Ph.D. thesis, Nikhef (2010).

[30] P. Clarke *et al.*, *Sensitivity studies to β_s and $\Delta\Gamma_s$ using the full $B_s^0 \to J/\psi\phi$ angular analysis at the LHCb*, CERN-LHCb-2007-101 (2007).

[31] Y. Xie, P. Clarke, G. Cowan, and F. Muheim, *Determination of $2\beta_s$ in $B_s \to J/\psi K^+K^-$ Decays in the Presence of a K^+K^- S-wave contribution*, JHEP **0909** (2009) 074.

[32] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, CHEP (2003), arXiv:0306116.

[33] http://root.cern.ch.

[34] D. Lange, *The EvtGen particle decay simulation package*, Nucl. Instr. and Methods **A462** (2001) 152.

[35] The LHCb Collaboration, P. Barbosa-Marinho *et al.*, *LHCb VELO Technical Design Report*, CERN-LHCC-2001-011 (2001).

[36] T. Ruf, *private communication.*

[37] G. van Apeldoorn *et al.*, *Outer Tracker Module Production at NIKHEF - Quality Assurance*, CERN-LHCb-2004-078.

[38] A. Pellegrino *et al.*, *OT FE-Box Test Procedures*, CERN-LHCb-2007-122.

[39] A. Berkien *et al.*, *The Front End Electronics of the LHCb Straw Tube Tracker*, CERN-LHCb-CONF-2009-021.

[40] N. Tuning, *private communication (LHCb week, 30 November 2005).*

[41] S. Bachmann *et al.*, *Test and Assembly of the Outer Tracker C-frames and Detector Modules at Installation*, CERN-LHCb-2008-033.

[42] P. Vankov, *Study of the B-Meson Lifetime and the Performance of the Outer Tracker at LHCb*, Ph.D. thesis, Nikhef (2008); L. Hooft van Huysduynen, master's thesis, *Propertime Resolution Model* Nikhef (2008).

[43] A. Zwart, *private communication.*

[44] M. Calvi, B. Khanji, G. Lanfranchi, O. Leroy, and S. Poss, *Lifetime unbiased selection of $B_s^0 \to J/\psi\phi$ and related control channels: $B^0 \to J/\psi K^*$ and $B^+ \to J/\psi K^+$*, CERN-LHCb-2009-025 (2009).

[45] G. Raven, *Selection of $B_s^0 \to J/\psi\phi$ and $B^+ \to J/\psi K^+$*, LHCb-2003-118 (2003).

[46] The BaBar Collaboration, B. Aubert *et al.*, *Measurement of the $B \to J/\psi K^*(892)$ decay amplitudes*, Phys. Rev. Lett. **93** (2001) 241801.

[47] The BaBar Collaboration, B. Aubert *et al.*, *Ambiguity-Free Measurement of $\cos 2\beta$: Time-Integrated and Time-Dependent Angular Analyses of $B \to J/\psi K\pi$*, Phys. Rev. **D71** (2004/2005) 032005.

[48] T. du Pree and G. Raven, *Methods for Angular Analyses of $B \to J/\psi X$*, CERN-LHCb-2009-024 (2009).

[49] I. Belyaev, *private communication.*

[50] A. Hicheur and G. Conti, *Parameterization of the LHCb Magnetic Field Map*, CERN-LHCb-2007-093 (2009).

[51] M. Baak, *private communication.*

[52] S. Rahatlou and S. Prell, BAD 467; R. Cahn, BAD 456.

[53] M. Calvi, G. Lanfranchi, O. Leroy, M. Musy, S. Poss, S. Vecchi, *Calibration of flavor tagging with $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^*$ control channels at LHCb*, CERN-LHCb-2009-020.

[54] The BaBar Collaboration, B. Aubert *et al.*, *Dalitz Plot Analysis of the Decay $B^0(\bar{B}^0) \to K^{\pm}\pi \mp \pi^0$*, Phys. Rev. **D78** (2008) 052005.

[55] The DØ Collaboration, V. Abazov *et al.*, *First Direct Two-Sided Bound on the B_s^0 Oscillation Frequency*, Phys. Rev. Lett. **97** (2006) 021802.

[56] GEANT4 Collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[57] T. Sjöstrand *et al.*, *High-Energy-Physics Event Generation with PYTHIA 6.1*, Comput. Phys. Commun. **135** (2001) 238.

[58] K. Cranmer, *Kernel Estimation in High-Energy Physics*, Comput. Phys. Commun. **136** (2001) 198.

[59] A. Carbone and S. Vecchi, *Systematic studies on the B_s proper time measurement for the determination of the mixing phase ϕ_s in the channel $B_s^0 \to J/\psi\phi$*, CERN-LHCb-PUB-2009-021 (2009).

[60] The CDF Collaboration, T. Aaltonen *et al.*, *First Flavor-Tagged Determination of Bounds on Mixing-Induced CP Violation in $B_s^0 \to J/\psi\phi$ Decays*, Phys. Rev. Lett. 100 (2008) 161802; The DØ Collaboration, V. Abazov *et al.*, *Measurement of B_s mixing parameters from the flavor-tagged decay $B_s^0 \to J/\psi\phi$*, Phys. Rev. Lett. **101** (2008) 241801.

[61] S. Stone, *First Physics Results from LHCb*, presented at 35th Int. Conf. on High Energy Physics, Paris, France (July 2010).

[62] The LHCb Collaboration, R. Aaij *et al.*, *Prompt K_S^0 production in pp collisions at $\sqrt{s} = 0.9$ TeV*, CERN-PH-EP-2010-027 (2010).

Summary, conclusions, and outlook

To explain the observed matter abundance in the universe, violation of the CP symmetry is required. Violation of this symmetry is incorporated in the standard model of particle physics in the weak interactions, but the amount of CP violation predicted in the standard model is by far too small to explain the size of the matter dominance. Therefore, to explain the observed matter abundance, sources of CP violation beyond the Standard Model are required. One of the goals of B physics is to search for new sources of CP violation, thereby probing models of new physics.

The source of CP-violating processes in the standard model is the irreducible complex coupling in the CKM matrix. Since this phase is opposite for particle and anti-particle decays, their decay rates can differ when this phase contributes to the sum of interfering decay amplitudes. CP asymmetries can be measured by comparing the time dependent decay rates for particle decays and anti-particle decays. This makes it possible to test the complex phases present in the CKM matrix.

The consistency of the different measurements of CP-violating processes can be checked by using the unitarity triangles, of which one angle is $\beta_s = \arg(-V_{cs}^* V_{cb} / V_{ts}^* V_{tb})$. This CP-violating phase is accessible in B_s decays for which both $B_s - \bar{B}_s$ mixing transitions and $b \to \bar{c}cs$ transitions contribute to the total decay amplitude. The small value of the amplitude of the time-dependent CP violation predicted in the standard model, related to $\phi_s = -2\beta_s = -0.04$ rad, can be changed due to contributions of off-shell particles, opening the possibility of finding contributions of heavy particles beyond the standard model.

The decay which offers the best sensitivity to this phase is the $B_s^0 \to J/\psi\phi$ decay. Since this is a decay of a pseudo scalar particle into two vector mesons, the final state is a superposition of states with different angular momentum, and hence of different CP eigenstates. Hence, an angular analysis is required to disentangle the different polarizations. To determine ϕ_s also the flavor of the B_s meson at production needs to be estimated. To suppress background the invariant mass of the final state particles needs to be determined.

An experiment designed to perform a measurement of this CP asymmetry is the LHCb experiment. The LHCb detector is located at the LHC accelerator, where B_s mesons will be produced copiously. The design of the tracking system leads to a proper time resolution of 40 fs, an angular resolution of $20 - 30$ mrad, and a B_s mass resolution of 16 MeV. Using the capabilities to perform particle identification, backgrounds can be suppressed, and a tagging power of 6.2% can be reached.

To reduce the rate of selected events to an acceptable level, the sample of selected events is divided in two samples: a sample of detached events and a sample of prescaled events. The sample of detached events is selected using an online lifetime cut, hereby suppressing background; the sample of prescaled events is selected by applying a prescale factor to

the trigger selection and is used as a control sample. The small set of additional selection criteria is optimized by maximizing a figure of merit $FOM = S/\sqrt{S+\alpha B}$, iteratively adjusting a set of rectangular cuts and estimating $\sigma(\phi_s) \propto \sqrt{S+\alpha B}/S$. When the trigger rate becomes critical, it can be lowered by adjusting the prescale fraction and the lifetime cut in the trigger selection, leaving the other selection criteria untouched.

To perform a measurement of the phase ϕ_s, a multi-dimensional simultaneous fit needs to be performed to the event distributions in the different observables: the proper time, the angular, the tagging and the invariant mass observables. In this thesis methods have been introduced to take into account the inefficiencies, resolutions, and backgrounds introduced by the selection method, the reconstruction algorithms, and the detector acceptance to the distributions in all observables. For the correction of the angular efficiency the usage of an MC sample is proposed, whereas for the other methods control samples are used. For all methods the $B^0_d \to J/\psi K^*$ decay is foreseen to serve as a control channel on which the methods can be validated first.

All methods have been shown to correct for biases in the estimates of the physics parameters on samples of fully simulated MC events. Using the expected experimental effects, resulting from full MC simulations, an estimate has been made of the expected precision of the measurement of the CP-violating phase ϕ_s. For 2.0 fb^{-1} the expected measurement precision is $\sigma(\phi_s) = 24 \pm 1$ mrad.

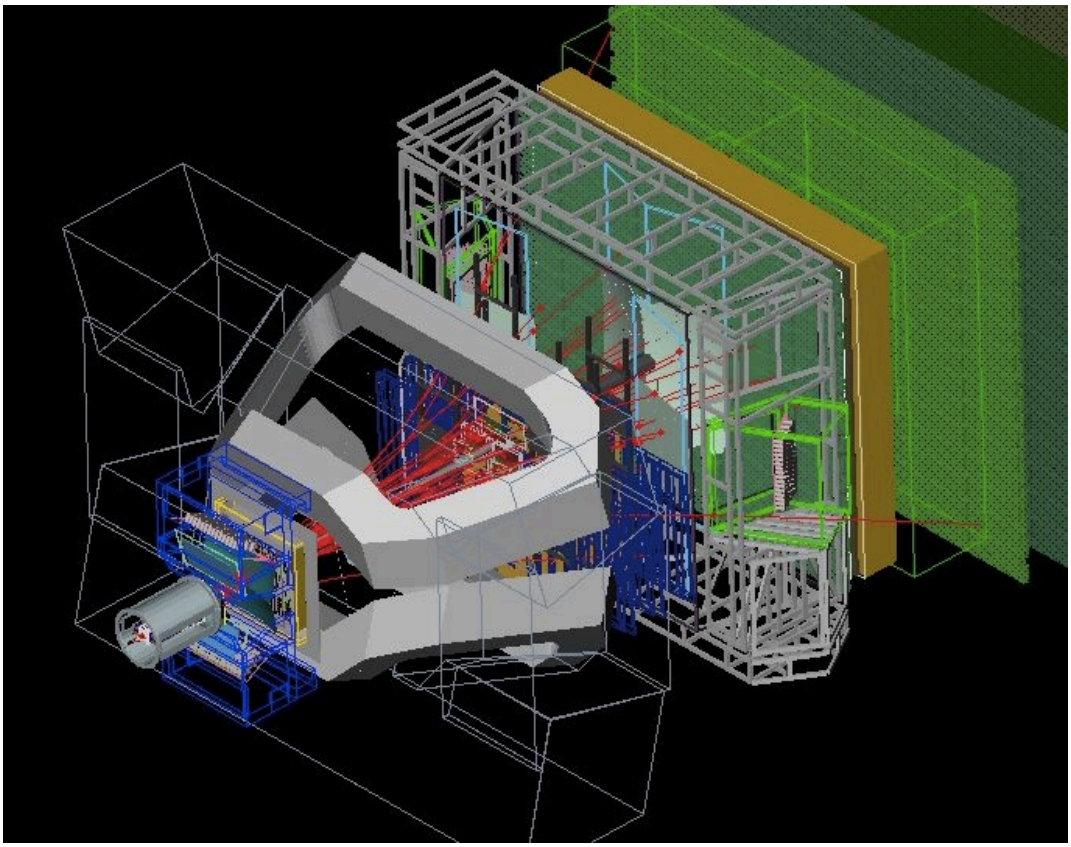


Figure 1: *Display of an event containing a $B^0_s \to J/\psi\phi$ candidate, as reconstructed by the LHCb detector.*

Current status

At the time of writing the first 600 nb^{-1} of data has been recorded at the LHCb experiment. For the decay channels $B^+ \to J/\psi K^+$, $B_d^0 \to J/\psi K^*$, and $B_s^0 \to J/\psi\phi$ candidates have been reconstructed, making it possible to test the predictions based on MC. An example of a $B_s^0 \to J/\psi\phi$ candidate is shown in figure 1.

In figure 2 the proper time and invariant mass of distributions of $B^+ \to J/\psi K^+$ candidates are shown. An abundance of long-living candidates can be seen in the signal region, hinting to reconstructed signal events. A significant invariant mass peak, with a fitted width of $\sigma_m = 23 \pm 2$ MeV/c^2, appears for long-living candidates around the B^+ invariant mass, containing an estimated number of signal events of $N_{\text{sig}} = 198 \pm 16$. As expected, the lifetime of the candidates in the signal region is larger than the lifetime of the candidates in the sidebands. The average proper time resolution is $\langle\sigma_t\rangle = 0.075$ ps.

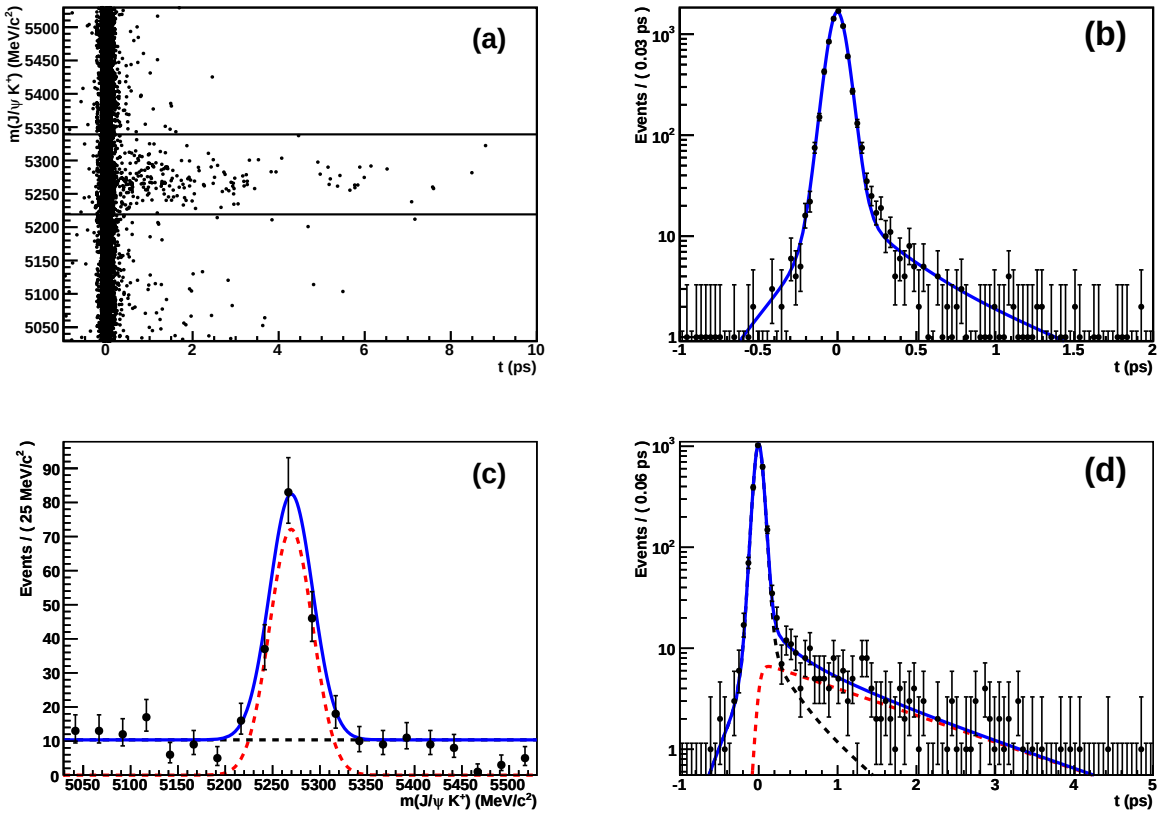


Figure 2: *Invariant mass and proper time distributions of $B^\pm \to J/\psi K^\pm$ candidates, indicating the signal region $|m(J/\psi K^+) - m(B^+)| < 60 MeV/c$ (a). Different projections of the data are shown, with a fit overlaid: the proper time distribution of the candidates outside the signal region (b), the proper time distribution of the candidates inside the signal region (d), and the invariant mass distribution for $t > 0.25$ ps (c).*

In figure 3 the proper time and invariant mass of distributions of $B_s^0 \to J/\psi\phi$ candidates

are shown, using the selection as described in this thesis. An abundance of long-living candidates can be seen in the signal region, hinting to reconstructed signal events. A significant invariant mass peak, with a fitted width of $\sigma_m = 37 \pm 9$ MeV/c^2 appears for long-living candidates around the B_s invariant mass, containing an estimated number of signal events of $N_{\text{sig}} = 23 \pm 5$. As expected, the lifetime of the candidates in the signal region is larger than the lifetime of the candidates in the sidebands. The average proper time resolution is $\langle \sigma_t \rangle = 0.074$ ps.

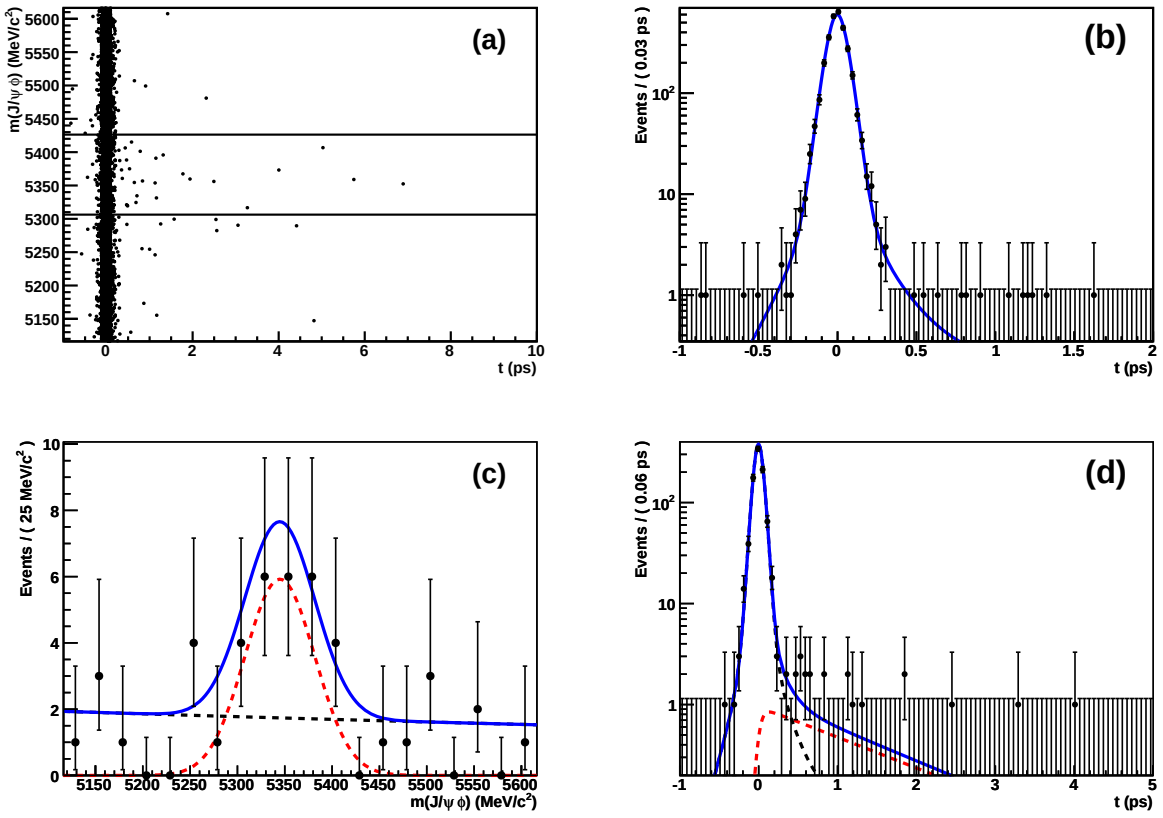


Figure 3: *Distributions of $B_s^0 \to J/\psi\phi$ candidates in the invariant mass and proper time observables, indicating the signal region $|m(J/\psi\phi) - m(B_s)| < 60 MeV/c$ (a). Different projections of the data are shown, with a fit overlaid: the proper time distribution of the candidates outside the signal region (b), the proper time distribution of the candidates inside the signal region (d), and the invariant mass distribution for $t > 0.25$ ps (c).*

From these distributions the following observations can be made. The proper time resolution, which is expected to be $\langle \sigma_t \rangle = 39$ fs based on MC simulations, is found to be a factor 1.9 larger. Since the sensitivity to the CP asymmetry scales as $e^{(\Delta m_s \times \sigma_t)^2/2}$, a maximal worsening of the sensitivity to ϕ_s of a factor two is expected. The mass resolution is, within statistical uncertainties, as large as expected from MC.

To calculate the expected signal yield, a $b\bar{b}$ cross section $\sigma_{b\bar{b}} = (292 \pm 15 \pm 43)\mu b$ at 7 TeV [61] is used and a relative uncertainty of the luminosity of 16% [62] is assumed. In that

case the number of signal events, using the selection described in this thesis, is expected to be 30 ± 7, which is in agreement with the found number of events. Furthermore, the number of events is too small to estimate the tagging performance or determine the polarization amplitudes from the angular distributions.

Hence, given the current situation it can be concluded that there are no limiting factors for the LHCb experiment to proceed to perform a measurement of ϕ_s.

Outlook

At the time of writing, the amount of statistics is too small to apply all analysis methods described in this thesis. When more data is taken, the methods are foreseen to be applied as follows.

The optimization procedure can be repeated using real data. Using pre-selected candidates, the figure of merit can be optimized, possibly adjusting the optimal selection criteria described in this thesis, in order to acquire maximal sensitivity to ϕ_s. As soon as the trigger rate becomes critical, it can be lowered by adjusting the prescale fraction and the lifetime cut in the trigger selection, leaving the other selection criteria untouched.

To test the fit methods described in this thesis, the $B_d^0 \to J/\psi K^*$ decay is foreseen as the control channel on which the methods can be validated. The methods to include inefficiencies, resolutions, and backgrounds in the angular and proper time resolutions can be checked on such a sample first, performing a measurement of mixing and the polarization amplitudes in this channel. When the physics parameters found for this channel are in agreement with the values estimated by the B factories, the same methods can be applied on the $B_s^0 \to J/\psi\phi$ channel, and the mistag fraction can be calibrated using the $B_d^0 \to J/\psi K^*$ channel.

Using the online and offline selection and the fit methods described in this thesis, the performance is expected to be as follows. After 0.1 fb^{-1} of luminosity (at a center of mass energy of 7 TeV), the amount of integrated luminosity expected at the end of 2010, the precision of LHCb should surpass that of the Tevatron for a total luminosity of 18 fb^{-1}, as expected at the end of run 2. After 1.0 fb^{-1} of luminosity, the amount of integrated luminosity expected before the shutdown of 2012, the sensitivity is 48 mrad. After 10 fb^{-1} (at a center of mass energy of 14 TeV) the statistical precision is expected to be 8 mrad, which means that for the SM value of the CP-violating phase, zero CP violation can be excluded at 5σ.

In case the estimated value appears to be different from the SM expected value, the measurement allows not only to discover new particles contributing to the amplitude of the decay processes, but also to determine both the magnitude and phase of the couplings of these new particles.

Populaire samenvatting

Zoektocht naar Vreemde Fases in Mooie Oscillaties

Een van de grote puzzels in de hedendaagse natuurkunde is het overschot van materie in het heelal. Materie en anti-materie zijn identiek in eigenschappen als massa en levensduur, maar verschillen in het feit dat ze een tegenovergestelde lading hebben. Zo kan uit hoogenergetisch licht een paar van deeltjes en anti-deeltjes ontstaan. Omgekeerd, als een deeltje en een anti-deeltje bij elkaar komen heffen ze elkaar op tot licht.

Daarom zou je verwachten dat er bij de oerknal evenveel materie als anti-materie is onstaan en dat beide elkaar vervolgens in de evolutie van het universum opgeheven zouden hebben. Toch komen we zowel op aarde als in de rest van het heelal alleen materie tegen, en is alle anti-materie verdwenen. Om dit overschot in materie te kunnen verklaren is er een verschil in het gedrag tussen de twee nodig.

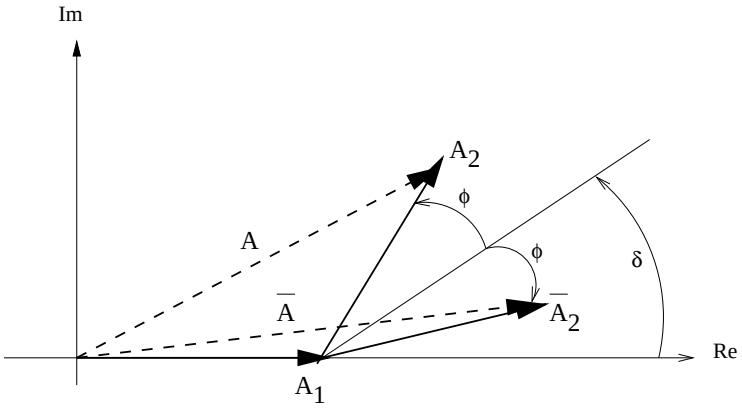
Echter, de huidige natuurkunde-modellen kunnen de huidige geobserveerde hoeveelheid materie niet verklaren, omdat het verschil in het gedrag van materie en anti-materie geproduceerd in versnellers veel te klein is. Daarom ben ik de laatste vier jaar beziggeweest met het voorbereiden van een onderzoek aan het LHCb experiment van een verschil tussen 'B_s mesonen' en 'anti-B_s mesonen'.

Een observatie van een groter verschil tussen deze twee deeltjes dan verwacht zou namelijk kunnen helpen in het verklaren waardoor er in het universum zo veel materie is, en dus eigenlijk, waardoor wij bestaan. In de komende pagina's zal ik proberen uit te leggen hoe ik die zoektocht heb voorbereid.

Complexe getallen

Om het verschil tussen deeltjes en anti-deeltjes te kunnen begrijpen, moet eerst het concept van een 'complex getal' bekend zijn. Iedereen leert op school dat $\sqrt{9} = 3$, immers $3 \times 3 = 9$. Echter, er wordt op school niet geleerd welk getal met zichzelf vermenigvuldigd moet worden om een negatief getal te krijgen. Natuurkundigen gebruiken daarvoor het 'imaginaire' getal i. Daarvoor geldt $i^2 = -1$ en, omgekeerd, $\sqrt{-1} = i$.

Zo kan een getal twee componenten hebben: een 'reële' component, bijvoorbeeld 3, en een 'imaginaire' component, bijvoorbeeld $2i$. Het getal $3 + 2i$ wordt dan een 'complex' getal genoemd. We zeggen dan ook wel dat het getal een 'complexe fase' heeft. Als complexe getallen bij elkaar opgeteld worden dan kan de grootte van de som, afhankelijk van de fases, verschillend zijn. Dit is getoond in figuur 4.



Figuur 4: *Een weergave van de som van complexe getallen. De lengte van een pijl geeft de kracht weer, de hoek van een pijl de fase. Doordat de fases verschillen, verschilt voor een som van de amplitudes de kracht van het verval voor deeltjes (A) en anti-deeltjes ($\bar{A}$).*

Quarks en mesonen

Verder moeten we ook weten dat de fundamentele bouwstenen van atoomkernen 'quarks' worden genoemd. Zoomen we in op de protonen en neutronen in een atoomkern, dan zijn, voor zover we weten, de kleinste deeltjes die we zien quarks. Deze quarks komen nooit alleen voor, maar altijd samen. Als quarks met z'n tweeën zijn, dan noemen we het samengestelde object een 'meson'.

Complexe koppelingen

Nu het begrip van een complex getal uiteengezet is, en we kennis hebben gemaakt met quarks en mesonen, kan het verschil tussen deeltjes en anti-deeltjes begrepen worden. Quarks en anti-quarks verschillen namelijk in het feit dat bij hun 'overgangen' de fase tegenovergesteld is. Dit betekent dat als een meson in een anti-meson overgaat, de fase tegenovergesteld is van een overgang van een anti-meson in een meson. Doordat de fase van een overgang van een B_s-meson naar een anti-B_s meson anders is dan het tegenovergestelde verval, kan een verschil ontstaan in het gedrag van de vervalstijd van deeltjes en anti-deeltjes.

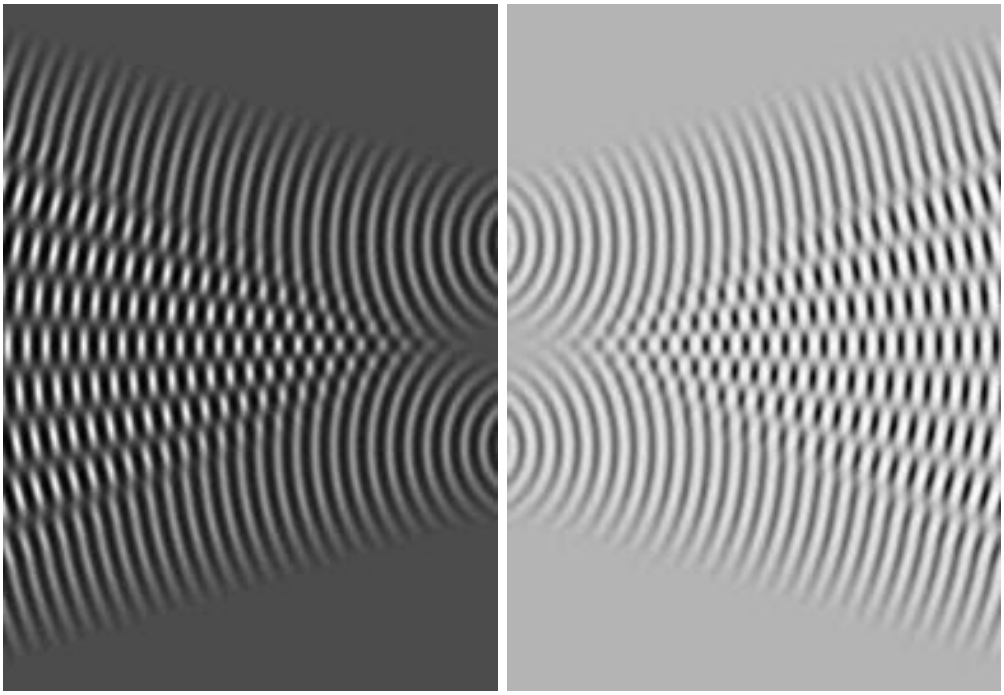
Het verschil in het gedrag van de vervalstijd komt door de wonderlijke eigenschappen van de kwantum-mechanica. Zoals watergolven met elkaar kunnen interfereren en aldus elkaar kunnen versterken danwel uitdoven, zo kunnen deeltjes in de kwantum-mechanica met zichzelf interfereren. Doordat de fase van deeltjes en anti-deeltjes verschilt, vervallen deeltjes en anti-deeltjes, naarmate de tijd vordert, verschillend. Door het verschil in vervalsintensiteit te meten, kan aldus de complexe fase van de koppelingen bepaald worden.

De kracht en de fase van het verval samen wordt ook wel de 'koppeling' genoemd. De koppelingen van deeltjes en anti-deeltjes zijn beschreven in de CKM matrix. Deze beschrijft de kracht van de koppelingen van overgangen tussen verschillende quark-smaken. Het doel van ons onderzoeksgebied, de 'B-fysica', is om zowel de groottes als de fases van de koppelingen te bepalen.

CP-schending

Voor een overschot aan deeltjes is een materie/anti-materie asymmetrie niet voldoende. Om van een oorspronkelijk evenwicht van deeltjes en anti-deeltjes toch een overschot aan materie te creëren moeten symmetrieën in de natuur gebroken zijn. Een van die symmetrieën die gebroken moet zijn is de symmetrie van de CP transformatie. Deze transformatie draait zowel de lading om (C) als de ruimte (P). C maakt van deeltjes anti-deeltjes, P laat iets wat rechtsom z'n as draait (zoals een ballerina voor een spiegel) linksom z'n as draaien.

Simpel gezegd: zou CP symmetrie een bestaande symmetrie van de natuur zijn, dan zouden er net zoveel deeltjes zijn die rechtsom hun as draaien, als anti-deeltjes die linksom hun as draaien, en zou er geen overschot van materie zijn in het heelal. Een grafische weergave van CP is getoond in figuur 5. Om het overschot aan materie te verklaren moeten we dus op zoek naar CP-schending.



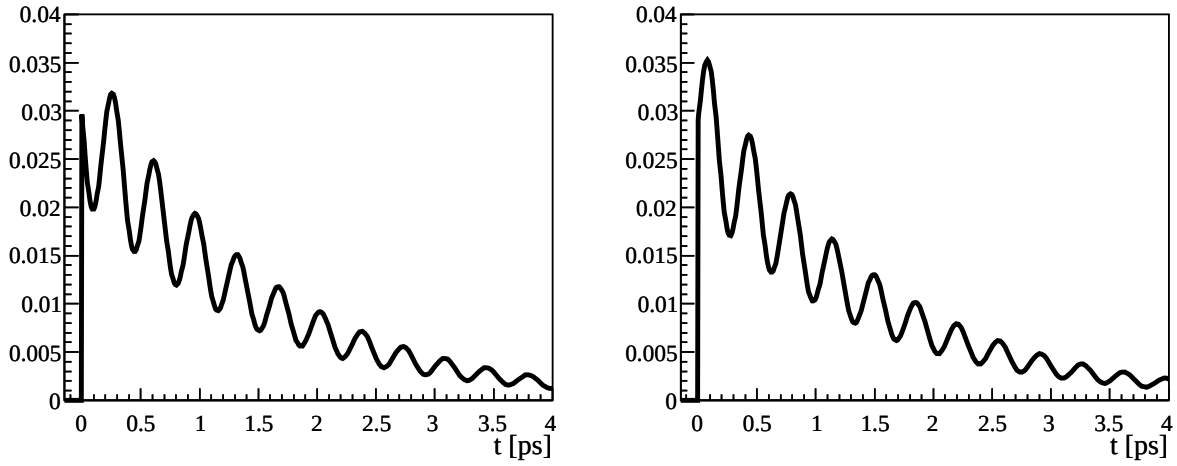
Figuur 5: *Schematische weergave van CP. P is weergegeven door de twee plaatjes te spiegelen. C, het omkeren van de lading, is gevisualiseerd door zwart en wit om te wisselen. Aangezien het gedrag van de watergolven hetzelfde is, is in dit geval CP niet geschonden.*

Vervallen van B_s-mesonen

Alle metingen van CP-schending in deeltjesvervallen bij versnellers zijn tot nog toe consistent te verklaren met het zojuist beschreven CKM-model. Echter, er is te weinig CP-schending geobserveerd om het overschot van materie in het heelal te kunnen verklaren. Er moeten dus onbekende bronnen van CP-schending zijn, veroorzaakt door nieuwe, nog niet ontdekte, deeltjes.

CP-schending is nog niet gemeten in het verval van B_s mesonen. Dit zijn mesonen bestaande uit een b ('beauty') quark en een s ('strange') quark. Doordat deze mesonen lang leven zijn ze uitermate geschikt om metingen van CP-schending aan te verrichten. Maar omdat deze mesonen relatief zwaar zijn, zijn B_s mesonen nog niet in grote getalen geproduceerd en aldus nog niet in detail bestudeerd. Immers, om zware deeltjes te maken, is er volgens $E = mc^2$ veel energie nodig bij een botsing.

Een theoretisch voorbeeld van CP-schending in B_s vervallen is getoond in figuur 6. De hoeveelheid CP-schending is gerelateerd aand het verschil in de hoogte van de oscillaties. De frequentie van de oscillaties is gerelateerd aan het massa-verschil Δm_s.



Figuur 6: *Een voorbeeld van CP-schending: het tijdsafhankelijke gedrag voor $B_s \to J/\psi\phi$ vervallen (links) verschilt van het gedrag voor $\bar{B}_s \to J/\psi\phi$ vervallen (rechts). De amplitude van de oscillaties is gerelateerd aan de CP-schendende fase, die in dit voorbeeld 29° is.*

De waarde van dit massa-verschil is $\Delta m_s = 2.2 \times 10^{-38}$ kg, oftewel een veertig-miljoenste van de massa van een elektron. De frequentie van de asymmetrie is 3×10^{12} Hz oftewel drie-duizend-miljard maal per seconde. Om nieuwe bronnen van CP-schending te vinden moeten dus eerst B_s mesonen gemaakt worden, en deze moeten vervolgens nauwkeurig bestudeerd worden.

LHCb detector

Voor de bestudering van de B_s mesonen is de LHC, de Large Hadron Collider uitermate geschikt. Dit is een deeltjesversneller bij Genève, 27 kilometer in omtrek en 100 meter onder de grond. Hierin worden protonen, de geladen kernen van waterstofatomen, tot bijna de lichtsnelheid in beide richtingen versneld. Elke seconde botsen protonen 40 miljoen keer op elkaar op bepaalde punten aan de ring. Op een van de snijpunten staat de LHCb detector om de botsingen, waarbij B_s mesonen worden geproduceerd, te analyseren.

De LHCb detector is uitermate geschikt om het verschil tussen B_s en anti-B_s mesonen te bepalen. De LHCb detector is namelijk gebouwd om zeer nauwkeurig de vervalsafstand van B_s mesonen te bepalen en daarmee heel nauwkeurig de levensduur van B_s vervallen. De

LHCb detector kan dit namelijk met een nauwkeurigheid van 40 fs, oftwel een vijfentwintigduizend-miljardste van een seconde.

Het meten van de CP-asymmetrie gaat als volgt. Van elk verval wordt de levensduur bepaald. Alle metingen worden verzameld, en een verdeling als in figuur 6 wordt gemaakt. Om de complexe fases van de koppelingen te bepalen moet de asymmetrie tussen de vervallen van de B_s mesonen en de anti-B_s mesonen bepaald worden.

Een analyse van de tijdsverdeling alleen is echter niet voldoende. Daarom is behalve een analyse van de tijdsverdeling in dit proefschrift ook een analyse van de verdeling in de drie vervalshoeken, en die van de invariante massa beschreven. Er is gekeken hoe verstoringen van de theoretisch verwachte verdelingen meegenomen kunnen worden door resoluties, inefficiënties en achtergronden mee te nemen in de beschrijven van de data. En, aangezien er een CP-asymmetrie gemeten moet worden, is er ook bestudeerd hoe de onnauwkeurigheid van de bepaling van het feit of er een B_s danwel een anti-B_s meson geproduceerd is bij de botsing, meegenomen kan worden in de analyse. Voor de methodes is ook aangegeven welke data-samples als controle kunnen fungeren.

Verder is er in dit proefschrift ook bestudeerd hoe de 'events' (voor ons interessante botsingen) geselecteerd kunnen worden. Hierbij is het van belang signaal events efficiënt te selecteren, en achtergrond events te onderdrukken. Verder moet er ook rekening gehouden worden met de hoeveelheid computerkracht die er aanwezig is om alle events te bestuderen. Daarvoor is in dit proefschrift een methode beschreven waardoor events niet alleen efficiënt, en met onderdrukking van achtergrond geselecteerd zijn, maar waarbij ook flexibel rekening gehouden kan worden met de hoeveelheid aanwezige computers.

Zoeken naar nieuwe natuurkunde

Voor het beschrijven van de natuurkunde van de aller kleinste deeltjes is het 'Standaard Model' de laatste decennia bijzonder succesvol gebleken. Echter, het kan niet verklaren waardoor er zo veel materie is in het heelal. Tegelijk voorspelt dit Standaard Model een CP-asymmetrie in het $B_s^0 \to J/\psi\phi$ verval die heel klein is. We gaan spoedig kijken of het Standaard Model er in dit geval met zijn voorspelling naast zit.

Op het moment van schrijven is de eerste data al genomen door LHCb, en zijn er al ongeveer 23 $B_s^0 \to J/\psi\phi$ events gereconstrueerd. Voordat er een statistisch significante meting kan worden gedaan van de CP-schending in dit kanaal zijn er meer data nodig. Gegeven de in dit proefschrift beschreven studie van de simulatie, de beschreven selectie van de vervallen, de analyse van de data en de huidige status van de detector, doet LHCb eind 2010 waarschijnlijk al de beste meting ter wereld van deze CP-asymmetrie. Eind 2011 zal de complexe fase, die het verschil tussen deeltjes en anti-deeltjes weergeeft, gemeten worden met een nauwkeurigheid van zo'n 3°.

Als deze fase significant verschilt met de door het Standaard Model voorspelde waarde, betekent dit dat er een onverwachte bron van CP-schending is ontdekt. Tegelijkertijd betekent dit dat er nieuwe deeltjes zijn ontdekt en dat zowel de grootte als de fase van de koppelingen van deze nieuwe deeltjes ontdekt zijn.

Acknowledgements

One of the things I like about experimental high energy physics is the fact that one works together with many people. Unfortunately, this means that the number of people that in some way have contributed to this thesis is too large to all personally thank here.

Certainly I would like to thank my supervisor, Gerhard Raven. Your enthusiasm has been enormous from day one and you were always full of suggestions. Thank you for showing me what a joy doing experimental physics research can be. I also would like to thank Marcel Merk for being my promotor and for taking the risk of hiring me when I was still a 'theory boy'. As a group leader you helped me with many things and you always did it with humor.

In the Nikhef LHCb group I have had the fortune to be surrounded by many other smart people. Juan Palacios, thank you for the help and advice, in particular concerning software. Vanya Belyaev, working with you on the development of the selection I learned a lot. Wouter Hulsbergen, the discussions and suggestions on multiple subjects concerning the analysis were always useful. Niels Tuning, thank you for the advice on basically all topics which I encountered during my promotion research, not in the last place concerning cars.

Outside of this group I would like to thank Max Baak for helping me to start on this subject. Also Stéphane T'Jampens has been a great help, in particular on the theoretical aspects of this analysis. Wouter Verkerke, thank you for answering questions about the usage of your program, I think it Rooles.

Concerning the time at CERN I would like to thank Yasmine Amhis for putting back the fun in fundamental physics. And Greig Cowan, I could thank you for many things, but in particular your help in tackling the Grid was unmissable. From the LHCb β_s working group I would especially like to thank Olivier Leroy, Basem Khanji, Gaia Lanfranchi, Yuehong Xie, and Peter Clarke for discussions and cooperation. I also would like to thank Géraldine Conti for her feedback on the P2VV fit package.

Concerning the work on the Outer Tracker I have to thank Antonio Pellegrino. Your humorous explanations of detector construction made sure testing even thousands of wires never became dull. Of the OT technicians I would in particular like to thank Loek Ceelie and Ruud Arink for the pleasant cooperation during the installation in the pit.

Both inside and outside the Nikhef LHCb group there are still too many people to mention. I must mention my 'favorite office mate' Gabriel Ybeles Smit. Furthermore, I very much enjoyed the discussions with Leo Wiggers during lunch. Roel Aaij and Daan van Eijk, it was nice to work together with you, especially on the processing and analysis of the first data. All the other people at Nikhef and within LHCb who helped, worked, drank coffee or spent a summer school with me: thank you for the good atmosphere at CERN and Nikhef.

Outside of nerdland I would like to thank my family and friends. And I would like to thank Isis and Mogwai for the accompanying four year long soundtrack.

Finally, I would like to thank Loes. Thank you for coming with me to Genève and after seven years still being my 'missie'.