

INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Pablo
Baladrón Rodríguez

PhD Thesis

Controlling the Penguin Pollution in the CKM Angle α with Axial B Decays at LHCb

Santiago de Compostela, 2024

Doctoral Programme in Nuclear and Particles Physics

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

DOCTORAL THESIS

CONTROLLING THE PENGUIN POLLUTION IN THE CKM ANGLE α WITH AXIAL B DECAYS AT LHCb

Author

Pablo Baladrón Rodríguez

Supervisor/s: Jeremy Peter Dalseno and Diego Martínez Santos

Tutor: Diego Martínez Santos

PHD PROGRAMME IN NUCLEAR AND PARTICLES PHYSICS

SANTIAGO DE COMPOSTELA / LUGO

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

To Celsa

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ABSTRACT

A branching fraction measurement of a four-body charged B meson decay within an inclusive charmless mass window is presented. The mass window focuses on axial decays, which decay into three final-state particles.

The full Run 1 and Run 2 LHCb dataset, corresponding to an integrated luminosity of $9fb^{-1}$, is analysed. A novel approach for estimating efficiencies is introduced, employing a flat phase space simulation in a data-driven method to extract the signal amplitudes.

This measurement serves as the preamble to an amplitude analysis that can be performed on the same data. Through $SU(3)$ symmetry, this analysis aims to constrain the penguin pollution in the CKM angle α extraction from the mixing-induced CP violation in the neutral $SU(3)$ related channel.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

LIMIAR

Nesta tese preséntase unha medida da frecuencia de decaemento do mesón B cargado a catro corpos dentro dunha rexión do espazo fásico sen quarks c. Ésta rexión defínese de tal xeito que permite estudar resonanciais axiais, as cales decaen en tres partículas do estado final.

Analízase todo o conxunto de datos de LHCb dispoñible, correspondente a unha luminosidade integrada de $9fb^{-1}$. Introdúcese un enfoque novo para estimar as eficiencias, empregando unha simulación de espazo fásico plano nun método baseado en datos para extraer as amplitudes do sinal.

Esta medida serve como preámbulo para unha análise de amplitudes que se pode realizar cos mesmos datos. A través da simetría $SU(3)$, ésta análise ten como obxectivo restrinxir a contaminación de procesos pingüíns na extracción do ángulo α de CKM na canle neutral relacionada a través da simetría $SU(3)$. Nesta canle, a medida de α obtense a partir da medida da violación de CP inducida pola mestura da oscilación do mesón B e da súa desintegración, é dicir, a medida da violación CP na interferencia destes dous procesos.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CONTENTS

Abstract iii

Limiar (abstract in Galician) v

Contents vi

1 Introduction 1
1.1 Goals and objectives . . . 3

2 Phenomenology 5
2.1 The Standard Model of Particle Physics . . . 5
2.1.1 CKM Matrix . . . 8
2.1.2 The α angle . . . 9
2.2 Controlling the penguin pollution in α from $B^0 \to a_1^{\pm}\pi^{\mp}$ . . . 10
2.2.1 α from $B^0 \to a_1^{\pm}\pi^{\mp}$ . . . 13
2.2.2 The $B^+ \to K_S^0\pi^+\pi^-\pi^+$ decay . . . 15

3 The LHCb experiment 19
3.1 Subdetectors . . . 19
3.1.1 Tracking . . . 20
3.1.1.1 VeLo . . . 21
3.1.1.2 Magnet . . . 21
3.1.1.3 T stations . . . 22
3.1.1.4 Reconstruction, Calibration and Types of tracks . . . 22
3.1.2 Particle identification . . . 22
3.1.3 Calorimetry and muon detection . . . 23
3.1.3.1 The Electromagnetic Calorimeter . . . 23
3.1.3.2 The Hadronic Calorimeter . . . 24
3.1.3.3 Muon stations . . . 24
3.2 Trigger . . . 24
3.2.1 Level-0 Trigger . . . 24
3.2.2 High Level Trigger . . . 25
3.2.3 TIS and TOS categories . . . 25
3.2.4 Triggering on Run 3 . . . 25
3.2.4.1 HLT2 selection study for Run3 . . . 26
3.3 LHCb data-flow . . . 27
3.3.1 Stripping . . . 27
3.3.2 Simulation . . . 30
3.3.2.1 MC corrections . . . 31

4 Methodology: Data-driven models on the Phase Space 33
4.1 The 5D Phase Space . . . 33
4.1.1 Phase Space density . . . 34
4.1.2 The quasi two-body approximation . . . 34
4.1.3 The axial resonances problem . . . 35
4.2 Data-driven models . . . 36

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.2.1 Kernel Density Estimation . . . 37
4.2.2 Gradient Boosted re-weighting . . . 37
4.3 Performance example: Efficiency parameterisation . . . 38
4.4 Discussion . . . 39
4.5 Complementary statistical and computing tools . . . 43

5 Charmless branching fraction measurement of the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay 45
5.1 Data sets . . . 46
5.2 Selection . . . 47
5.2.1 Trigger . . . 47
5.2.2 Stripping . . . 47
5.2.3 Final selection and background suppression . . . 48
5.2.3.1 Mass window . . . 48
5.2.3.2 Pre-selection and charm vetoes . . . 49
5.2.3.3 BDT selection . . . 49
5.2.3.4 Figure of merit study . . . 51
5.3 Yields . . . 54
5.3.1 Mass fit components . . . 56
5.3.1.1 Signal channel - Wide region . . . 57
5.3.1.2 Signal channel - Signal region . . . 57
5.3.1.3 Control channel - Wide region . . . 57
5.3.2 K^0_S selection . . . 57
5.3.3 Wide region . . . 58
5.3.3.1 Control channel . . . 59
5.3.4 Signal region . . . 60
5.3.4.1 sWeights . . . 60
5.4 Efficiencies . . . 62
5.4.1 Corrections on the PHSP . . . 64
5.4.2 PID correction . . . 65
5.4.3 Tracking correction . . . 66
5.4.4 Lo Hadron TIS/TOS correction . . . 66
5.4.5 Amplitudes correction . . . 67
5.4.6 Results . . . 74
5.4.6.1 Signal channel . . . 74
5.4.6.2 Control channel . . . 74
5.5 Systematic uncertainties . . . 80
5.5.1 Fit bias . . . 81
5.5.2 Signal model . . . 81
5.5.3 Background model . . . 83
5.5.4 BDT variables re-weighting . . . 83
5.5.5 Sample statistics . . . 85
5.5.6 Efficiencies correction . . . 85
5.5.6.1 TIS/TOS Lo Correction . . . 87
5.5.6.2 Track Correction . . . 87
5.5.6.3 PID Correction . . . 87
5.5.7 GBrw hyper-parameters . . . 87
5.6 Branching fraction result . . . 88
5.7 Discussion . . . 90

6 Momentum scale calibration at LHCb using the Armenteros-Podolanski plot 93
6.1 The Armenteros-Podolanski plot . . . 93
6.1.1 The modified Armenteros space . . . 95
6.2 Experimental effects on the Armenteros-Podolanski plot . . . 96
6.2.1 Momentum scale (S_p) . . . 96
6.2.2 Charge asymmetry (C_A) . . . 96
6.2.3 Parent mass scale (S_M) . . . 97
6.3 Theil index maximisation fitting method . . . 98

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

6.3.1 Theil index . . . 98
6.3.2 Fitting procedure . . . 98
6.3.3 Statistical uncertainties estimation . . . 100
6.4 Results . . . 100
6.4.1 $K^0_S \to \pi^+\pi^-$ . . . 100
6.4.2 $\Lambda^0 \to p\pi^-$ . . . 101
6.5 Discussion . . . 101

7 Conclusion **107**

8 Resumo **109**
8.1 Fenomenoloxía . . . 110
8.2 O Experimento LHCb . . . 111
8.3 Metodoloxía . . . 112
8.4 Medida da frecuencia relativa de decaemento de $B^+ \to K^0_S\pi^+\pi^-\pi^+$ . . . 112
8.5 Calibración da escala do momento en LHCb empregando o diagrama de Armenteros-Podolanski . . . 113
8.6 Conclusións . . . 114

Bibliography **115**

Glossary **121**

Appendix A Appendix **123**
A.1 MC fits . . . 123
A.1.1 Wide region . . . 123
A.1.1.1 Signal channel . . . 123
A.1.1.2 Control channel . . . 126
A.1.2 Signal region . . . 128
A.1.2.1 Signal channel . . . 128

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CHAPTER 1

INTRODUCTION

In this thesis, many different approaches have been considered to develop the objectives; some of them failed. Various goals arose, some of which were discarded, while others are still present in this text. Some past work was rescued from the ashes of oblivion and were further developed. In this thesis, as in many others, a final mixture of the work developed during these years of heavy learning is presented; it may not be a representation of the hard work, in the same way as not all the prospects of any research are fulfilled, but it will show my little grain of sand in the huge desert that is particle physics and data analysis.

The central work of the thesis is the branching fraction measurement of the $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ decay within a charmless mass window of the Phase Space (PHSP) at the Large Hadron Collider beauty (LHCb) experiment at CERN (*Centre Européen de Recherche Nucléaire*). In this analysis, a novel approach to estimate the efficiencies is investigated. Without prior knowledge of the amplitudes in data, they are parameterised using a data-driven approach to correct the efficiency obtained from a naive simulation of the signal of interest. This method is highly applicable to any research field where simulations need correction using data in high-dimensional spaces.

This measurement is part of a plan to determine the α angle of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [1, 2]. This plan represents the world's first attempt to measure α following this strategy, which involves the study of the axial mesons in the PHSP, decaying to three final-state particles. Additionally, the branching fraction is also a world-first measurement in the particular mass window of interest, laying the groundwork for an α measurement at the LHCb experiment. The CKM matrix quantifies the strength of different quark transitions within the Electroweak theory of Flavour Physics [3, 4, 5]. It contains a CP-violating phase that encapsulates the information about CP violation. In CP-violating processes, quark transitions do not occur with equal probabilities between the matter sector and its antimatter counterpart.

In physics, flavour refers to different families of quarks and leptons and differentiates the matter from the antimatter sector. Ordinary matter consists of quarks and leptons from the first family, while two additional families are headed by the strange and bottom quarks. Flavour Physics examines how these families relate and studies transitions between them. Specifically, the α angle is part of a system of equations ensuring the unitarity of the CKM matrix for quark transitions related to B-decays. The B meson contains a b quark and a quark from the first generation, depending on its charge. Mesons are composed of two quarks, whereas baryons, such as neutrons and protons (p), consist of three quarks and constitute ordinary matter together with the electrons (e^-), that are leptons from the first family.

The unitary relations can be represented geometrically as a triangle in the complex plane. The α angle is the least precisely known parameter of the triangle relations involving B-mesons; the other two angles are known with much greater precision and without ambiguity in their combination. Consequently, while the sum of the triangle's angles must equal π, the precision assessment of α serves as a consistency check within the CKM paradigm.

This measurement could shed light on one of the main problems of the Standard Model: the baryonic matter-antimatter asymmetry. According to Sakharov's conditions [6], CP violation is required to explain this asymmetry. However, the predicted CP violation within the Standard Model is insufficient to account for the observed asymmetry in our universe, as indicated by cosmological parameters [7]. Given that we inhabit a universe predominantly composed of matter, the absence of antimatter raises the question: where did it go? The discrepancy may arise from various sources, including cosmological models or the Standard

UB

Model itself, which lacks a key source of matter-antimatter asymmetry necessary to explain observations on the cosmological scale.
The Standard Model does not account for Dark Matter and Dark Energy, which together constitute more than 95% of the observed universe. Among other issues, it does not explain the non-zero masses of neutrinos, which could allow for a CKM-like matrix in the leptonic sector (the PMNS matrix), potentially revealing new sources of CP violation [8]. Despite these shortcomings, the Standard Model provides an exceptionally precise description of interactions between fundamental particles. Many observables are accurately predicted and match experimental results with remarkable precision [9]. However, discrepancies between predictions and experimental data persist in some cases [10], suggesting incompleteness in the theory and indicating that something essential may be missing.
Numerous theories Beyond the Standard Model (BSM) have been proposed [11], yet to date, none have replaced the Standard Model or fully resolved its puzzles.

On the experimental side, numerous particle physics experiments have been testing the Standard Model since the previous century. Key experiments that investigated b quark transitions, often referred to as B-factories, were BaBar and Belle. Currently, both the LHCb and Belle II experiments are actively taking data. In B-factories, B-mesons are produced from asymmetric $e^+ - e^-$ collisions, where the $\Upsilon(4S)$ meson emerges boosted in the forward direction and decays to a $B\bar{B}$ meson pair. However, in LHCb the colliding particles are protons where B-mesons emerge from a much more complex collision. The colliding particles are usually gluons present inside the protons, from where a $b\bar{b}$ quark pair emerges boosted in the beam-pipe direction. Then, the quarks hadronisates on different types of mesons and baryons as B, B_s or Λ_b particles.
It is crucial for science to have multiple independent experiments with similar physics programs to identify discrepancies between experiments, which may indicate misunderstanding of certain physics processes or underestimation of systematic uncertainties.
LHCb has achieved unprecedented results in studying CP violation in the b-quark sector [12], as well as in the charm sector [13]. However, LHCb does not currently plan to measure α due to the complexities involved in reconstructing neutral particles in the highly contaminated environment of $p - p$ collisions. In this thesis an alternative to measure α at LHCb will be presented.

On the theoretical side, I would like to mention that asking questions is one of the most fundamental characteristics of humankind. Since the birth of philosophy, a framework of thought has been present in our society, allowing us to express and compare different ideas. This led to critical thinking and the birth of science. Finally, physics was able to address the questions: Where do we come from? and What are we made of? Despite the answers in science being difficult to interpret, as the target of science is not the final truth but the path to it [14], some light has been shed on these profound questions.
Focusing on the particular matter of the Standard Model and Flavour Physics, it is important to remark that this is a very active field where numerous theoretical studies are still ongoing, and many aspects of the theory remain to be clarified and further understood. For example, in this thesis, all presented phenomenology assumes QCD factorisation, which means that the strong interaction and the weak interaction act on such different scales that there is typically no significant interplay between them. The strong interaction binds quarks into particles, while the weak interaction explains particle decay and quark transitions. However, this assumption is not universally correct for all energy scales, and currently, researchers are studying its impact of this assumption on many important observables for the weak theory [15]. With increasing data statistics, these effects can become comparable to the experimental uncertainties in measurements and may even rival new physics effects. Therefore, disentangling QCD effects at any energy scale from new physics effects in weak interactions is crucial for properly interpreting experimental results. In conclusion, it is crucial to understand that there is always an interplay between experiment and theory; they are not independent fields of study. This interplay becomes especially evident in global fits, where a model can be tested using multiple experimental inputs from various observables to constrain theory parameters. Such approaches are widely used to test BSM theories, as exemplified by recent work [16].

In addition to the branching fraction measurement, further work was developed in this thesis. A new method for momentum scale calibration was investigated, which differs fundamentally from current calibration methods. Currently, the momentum scale is calibrated using reference decays where the invariant mass is constructed from information about the daughter particles, allowing comparison with a reference value [17]. Any bias in the measurement is then attributed to the momentum scale bias.
In our method, however, no input from particles is needed. Instead, we use the Armenteros-Podolanski plot [18], where resonances appear as ellipses. In our proof-of-principle approach, we extract the momen-

tum scale by comparing data to simulation using a Theil index-based fitting method [19]. This method allows simultaneous extraction of both the momentum scale and the mass of the parent particle. This contrasts with current methods, which cannot separate the effects of inaccuracies in photon radiation characterisation from momentum scale effects. Here, both parameters are fitted simultaneously by fixing children particles to their known values.
This method has the potential to enhance detector calibration, leading to improved resolution and potentially advancing understanding of physical processes with a lower amount of data needed.

Furthermore, further efforts were made in preparation for Run 3 data-taking at LHCb. This research explored alternatives for storing candidates of our decays of interest. Optimising data storage in particle physics detectors is crucial for conserving disk space and efficiently managing computing resources.
A deep understanding of the ultimate physics goals is essential for determining the optimal approach to data selection and analysis as there is always an interplay between how data is selected and how it is analysed to achieve the final measurement. In our case, achieving a high acceptance in the threshold region of the PHSP for axial resonances is critical to reaching our physics goals. Therefore, it is important to investigate which selection criteria maintain optimal performance in this region of the PHSP. This region is crucial for our analysis, as the observables of interest for the α measurement will be extracted from here.

1.1 GOALS AND OBJECTIVES

An analysis on a particle physics experiment is a highly complex task, involving numerous contributions from detector planning, construction, and calibration, all the way to achieving the final physics goal. It is crucial to meticulously control every aspect that could potentially interfere with the measurement. Moreover, achieving an analysis goal typically requires a series of measurements that collectively fulfil various objectives, ultimately enabling the extraction of the final physics parameter.

The analysis presented in this thesis is part of an effort to measure the CKM α angle using LHCb data without requiring the reconstruction of neutral particles and involving the study of the axial mesons in the PHSP. Thus, the primary objective is the α measurement. This involves conducting a time-dependent flavour-tagged amplitude analysis of the $B^0 \to a_1^{\pm}\pi^{\mp}$ decay to extract CP-violation parameters related to interference between mixing and decay. Additionally, to mitigate penguin pollution in the α determination, a $SU(3)$ symmetry approximation is applied, which involves a branching fraction and amplitude analysis to extract branching ratios and phases of axial mesons in the $B^+ \to K_S^0\pi^+\pi^-\pi^+$ decay: $a_1^+ \to \pi^+\pi^-\pi^+$ and $K_1 \to K_S^0\pi^+\pi^-$. Note that charge conjugation is implied throughout this thesis, also considering the B^- decay and all the charge-conjugated processes.
The fundamental difference between a branching fraction analysis and an amplitude analysis lies in the type of information each method extracts. In an amplitude analysis, both the magnitude and phase of the different amplitude processes are assessed, with these processes parameterised as complex values. This allows for the disentanglement of specific processes leading to the final state measured in the branching fraction. In contrast, a branching fraction measurement only determines the squared magnitude of the amplitude, thus providing no information about the phases. Essentially, a branching fraction measurement is similar to a counting experiment, quantifying the absolute number of events for a specific process in the data. Conversely, an amplitude analysis measures only the ratios between different amplitudes, with every quantity obtained relative to a reference amplitude fixed to a standard value.
The specific contribution of this thesis is the measurement of the branching fraction of the $B^+ \to K_S^0\pi^+\pi^-\pi^+$ decay within a charmless mass window. In this analysis, Run 1 and Run 2 LHCb data are used, corresponding to an integrated luminosity of $9\,fb^{-1}$. An amplitude analysis can be performed on the same dataset, thereby achieving all the objectives related to the analysis of the $B^+ \to K_S^0\pi^+\pi^-\pi^+$ decay channel in the α measurement plan at LHCb. Section 2.2.2 provides a brief introduction to the amplitude analysis.

This initiative is challenging and could gain substantially from increased statistical precision for more accurate amplitude analyses of both neutral and charged B modes. Therefore, it is crucial to establish a robust data storage plan for future data-taking periods at LHCb. Thus, another objective is to develop a strategy for data storage to facilitate the α measurement during the Run 3 of the LHCb experiment.

Finally, in this thesis some past work is recovered to be further developed at the LHCb experiment. Initially studied as a theoretical feasibility, this work [20] has been further developed and applied in experimental conditions at LHCb. Chapter 6 explores the momentum scale calibration using the Armenteros-Podolanski plot as a proof-of-principle method at LHCb, aiming to establish a methodology suitable for the LHCb data-taking conditions.

In conclusion, this thesis is organised as follows: In Chapter 2 the theoretical introduction is presented, showing the phenomenology needed to understand the basics of the α plan at LHCb and the branching fraction measurement, that is presented in Chapter 5. Chapter 3 introduces the LHCb experiment, showing the studies developed to store the Run 3 data for the α plan at LHCb. In Chapter 4 the most relevant and novel methodology needed to develop the branching fraction measurement is presented, being an introduction to the methods that are utilised on the measurement, presented in Chapter 5 as already mentioned. Chapter 6 explores the momentum scale calibration method using the Armenteros plot at LHCb. Finally, the conclusions are drawn in Chapter 7.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 2

PHENOMENOLOGY

In this chapter, the main ideas for developing the analysis are presented, starting from the Standard Model of particle physics, going through the CKM matrix, and finishing with the analysis strategy to measure α at LHCb. In addition, in the last section, an introduction to the amplitude analysis of the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay is also given.

Physics has evolved enormously since the 20th century when Quantum Physics, Special Relativity, and Quantum Field Theories were developed. Thus, it is not possible to explain all the details of the physics involved in the analysis. Therefore, we are presenting here the basic concepts needed to understand the importance of the analysis's goal.

2.1 THE STANDARD MODEL OF PARTICLE PHYSICS

The Standard Model is the theory that studies the interactions between the fundamental particles that constitute matter and radiation on the smallest scale of our known universe. It is a renormalisable Quantum Field Theory (QFT) that unifies three of the four fundamental interactions in nature: the electromagnetic interaction, and the nuclear weak and strong interactions. Each interaction is mediated by a set of bosons. The exchange of these bosons between different fermionic particles produces the various observed interactions. The bosons are fundamental particles with integer spins, while the fermions, which are the matter constituents, have half-integer spins. Spin is a fundamental property of particles, like mass and charge, and it defines how particles interact with others of the same kind. Fermions cannot be in the same state, leading to Pauli's exclusion principle, whereas bosons can occupy the same state, leading, for example, to Bose-Einstein condensates.

This theory does not include the study of gravity, as it does not yet have a quantum formulation; this is one of the main open questions in physics. Additionally, the matter-antimatter asymmetry, dark matter, and dark energy remain among other significant unresolved issues.

QFT has some fundamental properties. They are renormalisable, which means that the energy scale of the theory is not fixed and can be set to a value defined, for example, by the fundamental constants of the theory. This is a key feature as the ultimate goal of these theories is to explain interactions at any energy scale so that any phenomena observed in nature can be explained by this theory. Additionally, another fundamental property of almost any QFT is CPT invariance. Where C, P and T are known as discrete symmetries, they are transformations that can be defined as:

- C is a charge transformation, it flips the sign of the electric charge.
- P is the parity transformation, which produces a mirror image of a given process. Quantum states which are eigenstates of the parity transformation, as mesons, can have eigenvalues +1 or −1. Together with the spin, it defines some properties of mesons. A spin-0 particle with negative parity is a pseudoscalar meson, as B mesons, pions or the K^0_S. A spin-1 meson is a vector if the parity is negative or an axial-vector if the parity is positive.
- T is the time reversal transformation.

Thus, CPT invariance states that the CPT transformation does not change a physical process as the physical

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

laws are invariant under this transformation. However, CP invariance can be violated, and this is a key property for the Standard Model as already mentioned.
There are some more symmetries that are always assumed in these theories, including Lorentz and gauge symmetries. Lorentz invariance states that the laws of physics are independent of the reference system, independent of rotations, translations, or any given boost or change in velocity.

Gauge symmetry is a local symmetry that describes the fields used in the formulation of the fundamental interactions. These fields are associated with a gauge boson that carries the interaction between particles that have a charge associated with the field; they are the mediating bosons. Each boson-interaction follows some rules defined by an operator in an algebraic symmetry group. Each operator is associated with the field, and the generators of the field are the gauge bosons. Thus, each particle with the associated charge can interact with the gauge bosons.
The electromagnetic interaction involves the electric charge, where photons mediate the interaction between charged particles. The gauge symmetry that describes this process is represented by the unitary group $U(1)$. The only degree of freedom is represented by the photon, and this group can be expressed as the circle in the complex plane.
The weak interaction involves the flavour charge, with $W^{\pm}$ and Z^0 bosons mediating the interactions between flavoured particles, allowing the transition or decay of some particles into others with different flavours. The gauge symmetry of the weak interaction is described by the special unitary group of the second order: $SU(2)$. There are 3 degrees of freedom represented by the 3 bosons. The $SU(N)$ group can be represented by $N \times N$ matrices with determinants equal to 1.
The strong interaction involves the color charge, with gluons mediating the interactions between colored particles, binding the nucleus and protons together. The gauge symmetry describing the strong interaction is given by the $SU(3)$ group, with 8 degrees of freedom represented by 8 gluons with different colors that bind the colored quarks together. The group representation is made by triplets representing the 3 possibilities for the color of each quark in analogy to the RGB color representation. In this way, each quark carries a color (or anti-color) charge; the gluons carry 2 color charges, binding the quarks together. There are 8 possible combinations of 2 colors among the 3 possible values for charge and anti-charge, these 8 gluons are the generators of the group.
However, it's important to note that the fact that gluons themselves carry color charge makes the strong force unique among the fundamental forces of nature. This self-interaction of gluons is one of the reasons why the strong force becomes stronger at longer distances, leading to the phenomenon known as confinement, where quarks and gluons are never observed in isolation but always bound together within composite particles such as protons, neutrons, and mesons.

The gauge symmetry that describes the Standard Model is:

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y \,,$$

where C refers to the color charge, L to the fact that the EW theory [3, 4, 5] couples to left-handed doublets, and Y is the hypercharge defined as $Y = 2(Q - T_3)$, where Q is the electric charge in units of the charge of the electron and T_3 is the weak isospin projection. Under this symmetry, we must express the different constituents of the Standard Model.

The Electroweak (EW) theory ($SU(2)_L \times U(1)_Y$) is a chiral theory where only the left-handed fermion doublets (featuring the lepton and the corresponding neutrino, and the two quarks) are coupled to the mediating bosons. Their right-handed counterparts are singlets for the EW group. Regarding the right-handed neutrinos, we can consider that as all their quantum numbers are 0 and they are massless in the Standard Model, they do not exist for us. Hence, in the Standard Model, there are 5 representations of the group, each having 3 different families. In Fig. 2.1, the particles, quarks and leptons, that constitute these representations are shown, together with the mediating bosons. The Standard Model representation can be written as:

- The left-handed quark doublets:

$$Q_L^i(3,2)_{+1/3} \;:\quad Q_L^1 = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad Q_L^2 = \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \quad Q_L^3 = \begin{pmatrix} t_L \\ b_L \end{pmatrix}$$

- The left-handed lepton doublets:

$$L_L^i(1,2)_{-1/2} \;:\quad L_L^1 = \begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}, \quad L_L^2 = \begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}, \quad L_L^3 = \begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$$

- The right-handed up-like quarks:

$$U_R^i(3,1)_{+4/3} \;:\quad U_R^1 = u_R \,, \quad U_R^2 = c_R \,, \quad U_R^3 = t_R$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

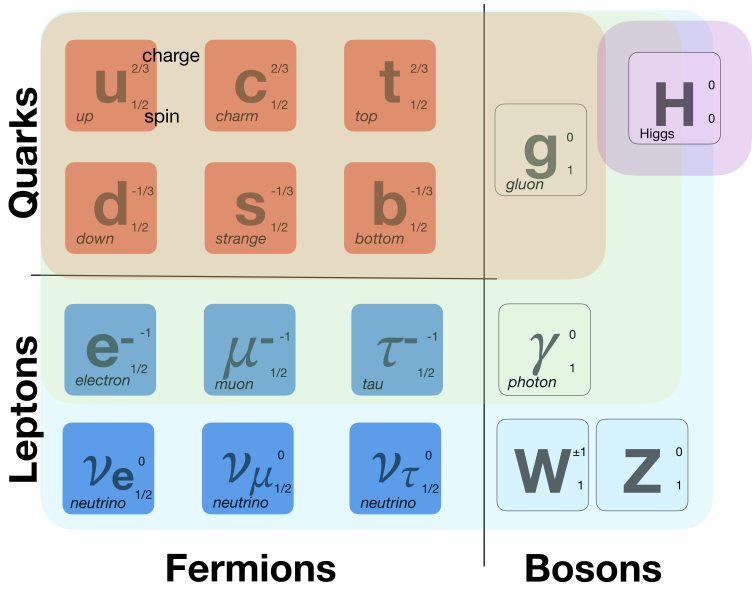


FIGURE 2.1: Chart summarising the particles within the Standard Model. Shaded regions denote the particles that are sensitive to each interaction mediated by the bosons.

- The right-handed down-like quarks:

$$D_R^i(3,1)_{-2/3}: \quad D_R^1 = d_R\,, \quad D_R^2 = s_R\,, \quad D_R^3 = b_R$$

- The right-handed fermions:

$$E_R^i(1,1)_{-1}: \quad E_R^1 = e_R\,, \quad E_R^2 = \mu_R\,, \quad E_R^3 = \tau_R$$

where the notation between parentheses and as a subscript of the parentheses denotes that, e.g., $Q_L^i(3,2)_{+1/3}$ are $SU(3)_C$ triplets and $SU(2)_L$ doublets and have hypercharge $Y = +1/3$.

Now it is time to talk about the Higgs mechanism and how the bosons and the fermions obtain mass through the spontaneous symmetry breaking mechanism. Without the Higgs, the $SU(2)_L \times U(1)_Y$ bosons would be massless.

We described the Electroweak interaction as a unified interaction for the weak and electromagnetic processes. So far, the mediating bosons are massless, hence, the distance at which these bosons can act is infinite. This is the case for the electromagnetic interaction, but for the weak interaction, we know that its effect has a short distance range; hence, in nature, these bosons have some mass, and given the experimental observations, this must be a heavy mass. To solve this problem, we need an extra doublet in the theory with a large vacuum expectation value (vev) which indicates that this Electroweak symmetry is spontaneously broken. This process is known in the literature as the Higgs mechanism [21]. The new scalar field can be represented as

$$\phi(1,2)_{+1} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}.$$

With a vacuum expectation value of

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}.$$

The Higgs potential has the shape of a Mexican hat and can be expressed as:

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2\,,$$

where μ^2 must be negative. The † symbol means the Hermitian conjugate of the matrix, i.e., the transposed and complex conjugated matrix. The vev is then:

$$v = \frac{-\mu^2}{2\lambda}\,.$$

USC UNIVERSIDADE DE SANTIAGO

Consequently, we have the EW symmetry broken with a large vev. Using a rotation, we can transform the gauge bosons into the physical ones. These new rotated fields then acquire mass, leaving the photon

massless. It can be seen that the EW bosons obtain the high mass values as expected, but to obtain masses for the quarks, we need a new coupling. We add to the Standard Model a Yukawa-like coupling.

Finally, the Standard Model Lagrangian can be written as:

$$\mathcal{L}_{SM} = \mathcal{L}_{Gauge} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}\,,$$

where

$$\mathcal{L}_{Gauge} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi\,,$$

which encapsulates the information from the gauge fields and their interactions with the fermions,

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi)\,,$$

and the Yukawa term for the different kinds of fermions,

$$\mathcal{L}_{Yukawa} = -\sum_{ij} y^d_{ij}\bar{Q}^i_L\phi d^j_R - \sum_{ij} y^u_{ij}\bar{Q}^i_L\tilde{\phi}u^j_R - \sum_{ij} y^e_{ij}\bar{L}^i_L\phi e^j_R + \text{h.c.}\,,$$

where y^d_{ij} and y^u_{ij} are 3×3 complex matrices for the d and u type quarks respectively. Note that both Yukawa couplings cannot be diagonalised at the same time in a given gauge, this being the first hint of the presence of the CKM matrix in the Standard Model that relates mass eigenstates with the weak eigenstates of the quark representations.

2.1.1 CKM MATRIX

The fact that the Yukawa coupling produces a set of fermions with mass, and the interacting fermions through the weak currents are massless in their interaction with the gauge bosons, produces a rotation matrix in the flavour space that links the propagating fermions with the weak-interacting fermions. This gauge rotation is fixed by the masses of the fermions of the different families, constraining the rotation matrix that defines the transition between the fermions of the different families through the charged weak currents. In the Standard Model, the neutrinos are massless, so this gauge is free in the leptonic sector and this rotation matrix can be trivially diagonalised, producing no effective difference between the mass and weak leptonic eigenstates. Some BSM theories propose alternatives where the non-zero mass of the neutrinos produces a rotation matrix in the leptonic sector [8]. For the goal of this thesis, we will only consider the rotation matrix in the quark sector, as our decays of study are purely hadronic and the leptons do not play a role.

Through Noether's theorem [22], the symmetries of the Lagrangian can be related to some conserved currents. In the Electroweak theory, the charged currents that are mediated by the $W^\pm$ bosons can be written as:

$$J^\mu_{W^\pm} = -\frac{g}{\sqrt{2}}\left(\bar{u}_L\gamma^\mu W^+_\mu V_{CKM}d_L + \bar{d}_L\gamma^\mu W^-_\mu V^\dagger_{CKM}u_L\right) \tag{2.1}$$

The u_L and d_L terms are the weak eigenstates for the quarks of type up and down respectively. These are related to the mass eigenstates (U_L, D_L) through a rotation in the flavour space that can be written as:

$$U_L = \mathcal{U}^U_L u_L \qquad D_L = \mathcal{U}^D_L d_L\,.$$

These matrices are unitary and their product is the $V_{CKM} = \mathcal{U}^{U\dagger}_L\mathcal{U}^D_L$ term that appears in Eq.2.1. If the weak and mass eigenstates were the same, this matrix would be the identity. The matrix was first proposed by Cabibbo [1] in the context of two quark families and presented in its current form using the three quark families by Kobayashi and Maskawa [2]. It is a 3×3 matrix that can be expressed as:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.042 \\ 0.009 & 0.041 & 0.999 \end{pmatrix} \tag{2.2}$$

USC

The approximate values are extracted from the latest report of the CKMFitter [23] group.

Despite there being 9 complex elements in the matrix, there are only 4 free parameters, due to the unitarity conditions and the fact that the 3 free phases for each quark family are absorbed into a single

one due to the arbitrariness of each phase. This is known as rephasing freedom. Thus, this matrix can be parameterised using 3 mixing angles and a single phase. This phase carries all the information about CP violation within the Standard Model, assuming a non-vanishing Jarlskog determinant. Many different parameterisations were developed for this matrix. The standard parameterisation uses 3 rotating angles (θ_{ij}) related to the transitions between the different families and includes the CP-violating phase (δ) in one of them. Defining $s_{ij} = \sin(\theta_{ij})$ and $c_{ij} = \cos(\theta_{ij})$, this parameterisation can be written as:

$$V_{\text{CKM}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \\ \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{2.3}$$

It is known experimentally that $s_{12} > s_{23} > s_{13}$, with $s_{12} \approx 0.225$. It is then convenient to use the Wolfenstein [24] parameterisation, an empirical representation that makes use of this hierarchy, defining:

$$\begin{aligned} s_{12} = \lambda = |V_{us}| = \sqrt{|V_{ud}|^2 + |V_{us}|^2}, \\ s_{23} = A\lambda^2 = \lambda \left| \frac{V_{cb}}{V_{us}} \right|, \\ s_{13}e^{i\delta} = V_{ub}^* = A\lambda^3(\rho + i\eta) = \\ A\lambda^3(\bar{\rho} + i\bar{\eta})\sqrt{1 - A^2\lambda^4}\sqrt{1 - \lambda^2}\left[1 - A^2\lambda^4(\bar{\rho} + i\bar{\eta})\right] \end{aligned} \tag{2.4}$$

These relations ensure that:

$$\bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \tag{2.5}$$

With this definition, we can write the Wolfenstein parameterisation to any order in λ without compromising the unitarity of the matrix. We can write this representation as:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4)\,. \tag{2.6}$$

Obtaining, in this way, the most widely used representation of the CKM matrix. The λ parameter is the most important one in this representation, as one can easily know the strength of some quark transition by its order in this parameter. The ρ and η parameters carry the information of the CP violation. Many interesting observables related to CP-violation that can be experimentally measured are usually represented in the complex plane, this is a way to represent the CP-violating parameters constraining the CKM mechanism.

Charge-Parity symmetry can be violated in the Standard Model in many different quark transitions. This means that the processes where some quark content is present do not happen with the same probability as those processes where the charge and parity conjugated particles are involved. We know that the weak interaction maximally violates the parity, as only the left-handed fermions are coupled to the weak bosons. However, the presence of the complex phase in the CKM matrix also allows the possibility of CP-violating processes, as the CP transformation changes the sign of this complex phase. These processes are very interesting because they can help us to explain the matter-antimatter asymmetry in our universe. In addition, it is very interesting to explore these processes as new sources of CP violation may arise outside the scope of the Standard Model, leading to new discoveries and physics beyond the Standard Model.

2.1.2 THE α ANGLE

The unitarity of the CKM matrix, $V_{\text{CKM}}V_{\text{CKM}}^\dagger = \mathbb{1}$, yields the following relation, among many others, for the matrix elements involved in the quark transitions related to B^0-meson decays, whose quark content is $\begin{pmatrix} b \\ \bar{d} \end{pmatrix}$:

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0\,. \tag{2.7}$$

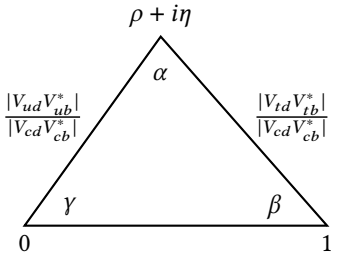


FIGURE 2.2: Triangle representation of the CKM matrix relations on the B^0 sector.

We can redefine this equation by dividing all the terms by $V_{cd}V_{cb}^*$ to obtain a matching triangle in the complex plane, as shown in Fig. 2.2. The vertices of the triangle in the complex plane are $(0,0)$, $(1,0)$, and $(\bar{\rho}, \bar{\eta})$ according to the definition of the Wolfenstein parameterisation (Eq. 2.4).
We can then define the 3 angles of this triangle as:

$$\begin{aligned} \alpha &= \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \\ \beta &= \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \\ \gamma &= \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \end{aligned} \tag{2.8}$$

By over-constraining this triangle through many measurements of different CP-violating observables, one can study the consistency of the CKM mechanism in the Standard Model and search for new sources of CP violation in precision studies.
In Fig. 2.3, the current status of many measurements constraining this triangle is shown.

The α angle is the least well-known angle of this triangle. Assessing its precision could lead us to discover sources of new physics. The α angle can be measured in $b \to d(u\bar{u})$ quark transitions, where $B^0 - \bar{B}^0$ mixing occurs, dominated by $b \to d(t\bar{t})$ transitions. Through the study of mixing-induced CP violation using time-dependent and flavour-tagged amplitude analysis, the α angle can be extracted.
In contrast with the study of the other two angles, the α sector faces an added complication: the contribution of penguin processes, which involve a different CKM phase, can be of the same order of magnitude as tree-level contributions. This results in an effective measurement of the angle rather than its direct determination, expressed as $\alpha + \Delta\alpha$, where $\Delta\alpha$ represents the penguin pollution.
These penguin processes are next-to-leading-order processes where the same final state is achieved but through a different process involving different CKM elements. These processes are typically represented as Feynman diagrams, with the tree-level being the first-order process. Higher orders involve internal loops that require additional interactions, which can be mediated by a weak boson (weak penguin) or a gluon (strong penguin).
These processes are generally highly suppressed compared to the tree level. However, due to the small V_{ub} term in the CKM matrix, the size of the penguin contamination can be significant in $b \to d(u\bar{u})$ decay processes. The Feynman diagrams illustrating these processes are shown in Fig. 2.4.

It is also important to remark that quarks, carrying color charge, are sensitive to the strong interaction. Typically, since the scale of the strong interaction is much smaller than that of the weak interaction, we assume QCD factorisation [25]. This assumption implies that the hadronisation process does not affect the CKM weak parameters. While generally a good approximation, it can be a limiting factor for very precise measurements, and corrections may depend on the energy scale at which the observables are measured and estimated.
In our study, we will assume that QCD effects factorise. However, this approximation becomes limiting when attempting to extract the impact of penguin processes on the CKM phase.

2.2 CONTROLLING THE PENGUIN POLLUTION IN α FROM $B^0 \to a_1^{\pm}\pi^{\mp}$

USC UNIVERSIDADE DE SANTIAGO

In this section, various approaches to measure α are presented. Subsequently, the strategy chosen to measure it at the LHCb experiment will be detailed. This approach does not require the reconstruction

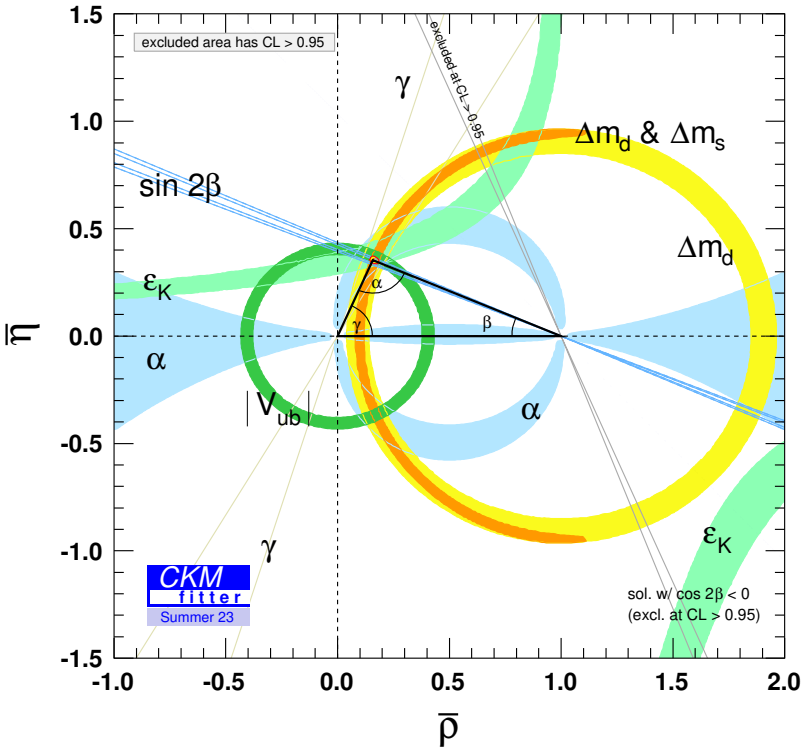
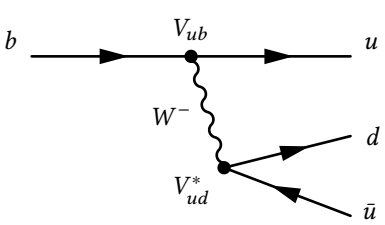
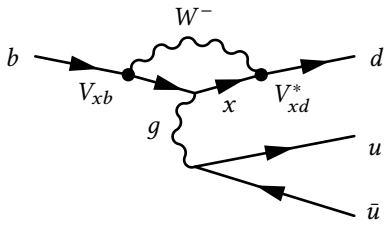


FIGURE 2.3: CKMFitter group [23] constraints in the $(\bar{\rho}, \bar{\eta})$ plane.



(A) Tree level diagram.



(B) Penguin diagram, $x = t, c, u$.

FIGURE 2.4: Feynman diagrams for the $b \longrightarrow d(u\bar{u})$ quark transition.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

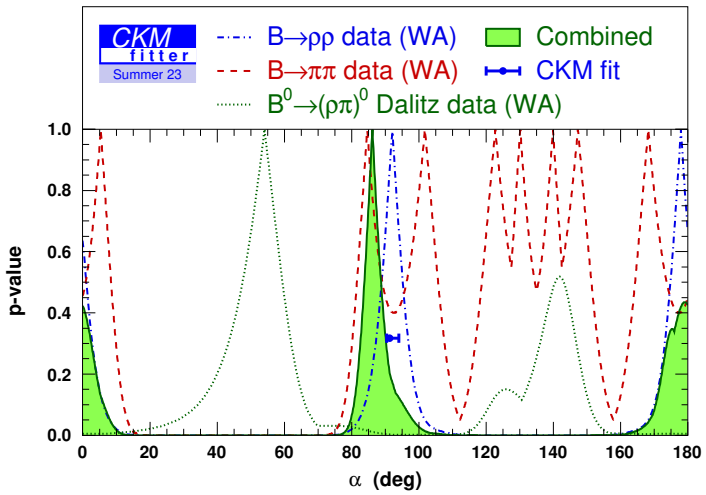


FIGURE 2.5: CKMFitter group [23] combination of the α measurement from $B \to \pi\pi$ and $B \to \rho\rho$ isospin analysis and $B^0 \to \rho\pi$ Dalitz analysis.

of any neutral particle in the final state, which, as discussed in Chapter 3, is particularly challenging at LHCb. Finally, the role of the charmless branching fraction measurement in the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay, presented in this thesis, will be clarified in the context of the α strategy at LHCb, concluding with an introduction to the next steps of analysis using the same data.

The decay channels from which an α measurement can be extracted are usually represented as $B \to hh$, where h represents a hadron containing only u and d quarks. These analyses typically involve complicated isospin relations [26], considering all possible charge configurations and invoking $SU(2)$ symmetry to control penguin pollution in the α measurement. However, these approaches often fail to provide a unique measurement due to ambiguities resulting in multiple solutions within the $[0, \pi]$ range.

On the $B \to \pi\pi$ system, there exists an 8-fold degeneracy in the α solution. The isospin relations can be represented as triangles in the complex plane, where ambiguity arises from the orientation of these triangles relative to each other. If the penguin contribution is negligible, these triangles would collapse into a line, reducing the ambiguity to a 2-fold degeneracy. Analysis of mixing-induced CP-violation, extracting $\sin(2\alpha)$, has been performed at the B-factories BaBar [27, 28], Belle [29, 30, 31], and LHCb [32].

Conversely, the $B \to \rho\rho$ system presents a more favorable situation with a smaller penguin contribution, leading to a 2-fold degeneracy in this system. This analysis has also been conducted at the three B-factories: BaBar [33, 34], Belle [35, 36], and LHCb [37].

Additionally, the $B \to \rho\pi$ system has been studied using the Dalitz plot of the three pions in the final state, employing a different approach. However, the solutions obtained from this analysis do not align with those from the isospin analysis. It is important to note that this analysis involves studying the three-body amplitude, which includes the additional challenge of neutral pion reconstruction. This analysis has been performed at BaBar [38] and Belle [39].

The α measurement from these systems is combined to reduce the number of multiple solutions that arise for α. The current status of combining these measurements to extract the α angle is shown in Fig. 2.5.

In this thesis, the strategy involves studying the axial resonances present in the $B^0 \to \pi\pi\pi\pi$ decay, extracting $B^0 \to a_1^{\pm}\pi^{\mp}$ using a time-dependent flavour-tagged amplitude analysis. It should be noted that this final state is the same as $B^0 \to \rho^0\rho^0$, from which α can also be extracted. Therefore, these axial contributions can also be used to constrain penguin pollution in the α measurement from the $\rho\rho$ system using $SU(2)$ isospin relations [40].

However, a different approach will be considered where a single α solution may be extracted by studying $SU(3)$-related channels, such as $B^+ \to K^0_S\pi^+\pi^-\pi^+$. These channels can be used to control the penguin pollution in the effective α measurement from $B^0 \to a_1^{\pm}\pi^{\mp}$. Some work on the $B^0 \to a_1\pi$ decay has already been performed by the BaBar [41] and Belle [42] experiments.

2.2.1 α FROM $B^0 \to a_1^\pm \pi^\mp$

As previously shown, the CKM angle α is the least known parameter among the three angles of the unitary triangle. Therefore, improving its precision can further test the CKM mechanism and provide potential hints of new physics in the flavour sector.
The previously presented strategies that invoke $SU(2)$ symmetry arguments fail to yield a single solution for α, despite being theoretically attractive due to the small mass difference between the u and d quarks. In our approach, we will employ $SU(3)$ symmetry, assuming equal masses for the u, d, and s quarks. With this premise, we aim to leverage the study of axial vectors to better constrain the solution for the α angle in charmless B four-body decays.
This idea has been explored in previous theoretical works [43] and experimentally by the BaBar collaboration [44], utilising the branching fractions measured at the B factories. With the increased luminosity at these facilities today, a more detailed analysis is feasible. An amplitude analysis of these axial decays could also utilise phase information from the amplitudes to enhance constraints on the α measurement through the $SU(3)$ approximation [45].
Although it may appear less appealing from a theoretical standpoint, we will leverage $SU(3)$ flavour symmetry to mitigate the penguin pollution in the effective α solution obtained from the $B^0 \to a_1^\pm \pi^\mp$ decay. Given the numerous strategies in the past that have failed to determine a unique solution for this angle, employing this symmetry in axial resonances could provide additional constraints to refine the α determination and reduce its ambiguities.

To extract α from $B^0 \to a_1^\pm \pi^\mp$, a time-dependent flavour-tagged amplitude analysis must be performed, as α can be extracted from the phase shift caused by mixing-induced CP-violation, i.e., the interference between B^0-meson mixing and decay.
Since the final state is not a CP-eigenstate, we need to distinguish the flavour of the decaying B^0 meson through a flavour tagging algorithm.
The amplitudes involved in the decay are defined as:

$$\begin{aligned} A(B^0 \to a_1^+ \pi^-) &= T^+ e^{i\gamma} + P^+ , \\ A(B^0 \to a_1^- \pi^+) &= T^- e^{i\gamma} + P^- , \end{aligned} \tag{2.9}$$

where T and P represent complex amplitudes involving only strong dynamics, with the weak phase explicitly represented by γ. The presence of γ in the definitions arise from using the C-convention [26] when constructing the amplitudes. Furthermore, due to the unitarity of the CKM matrix, the re-phasing freedom allows us to absorb phases such that γ multiplies the tree-level (T) amplitude. Consequently, the amplitudes from the diagrams are redefined such that $T = \mathcal{T}_u + \mathcal{P}_u - \mathcal{P}_c$ and $P = \mathcal{P}_t - \mathcal{P}_c$, where $\mathcal{T}$ and $\mathcal{P}$ represent the tree and penguin Feynman diagrams respectively, as shown in Fig. 2.4.
The CP-conjugate amplitudes are given by:

$$\begin{aligned} \bar{A}(\bar{B}^0 \to a_1^- \pi^+) &= T^+ e^{-i\gamma} + P^+ , \\ \bar{A}(\bar{B}^0 \to a_1^+ \pi^-) &= T^- e^{-i\gamma} + P^- . \end{aligned} \tag{2.10}$$

The CP-violating parameter λ_{CP}, whose phase is related to the α angle, is defined as:

$$\begin{aligned} \lambda_{CP}^+ &= \frac{\bar{A}(\bar{B}^0 \to a_1^- \pi^+)}{A(B^0 \to a_1^+ \pi^-)} e^{i(2\pi - 2\beta)}, \\ \lambda_{CP}^- &= \frac{\bar{A}(\bar{B}^0 \to a_1^+ \pi^-)}{A(B^0 \to a_1^- \pi^+)} e^{i(2\pi - 2\beta)}, \end{aligned} \tag{2.11}$$

where we have assumed negligible CP-violation due to mixing ($|q/p| = 1$). In our convention, β is the phase of $B^0 - \bar{B}^0$ mixing. We introduce a 2π redundant phase shift to align with the CKM relation: $\alpha = \pi - \beta - \gamma$. Finally, the effective α measurement from this time-dependent flavour-tagged amplitude analysis is given by:

$$\alpha_{\text{eff}}^\pm = \arg(\lambda_{CP}^\pm)/2 = \alpha + \Delta\alpha^\pm, \tag{2.12}$$

where the difference between these solutions for α_{eff} is a strong phase, δ_{strong}. This strong phase enters into the penguin contribution, which does not carry any weak phase. However, it is not trivially cleared from Eq. 2.12 due to the amplitude ratio. If there were no penguin contribution, the impact of the strong phase difference between the two solutions for λ_{CP} would be negligible.

USC

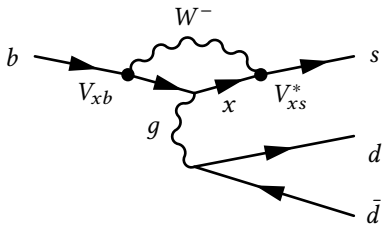


FIGURE 2.6: Penguin Feynman diagram for the $b \to s(d\bar{d})$ quark transition, $x = t, c, u$.

To control the penguin pollution, we invoke the $SU(3)$ argument. The $SU(3)$ partner of the a_1 particle is the $K_{1,A}$ meson, both being spin-1 axial resonances. The $K_{1,A}$ meson is a weak eigenstate and represents a mixture of the $K_1(1270)$ and $K_1(1400)$ mass eigenstates, which complicates the analysis but does not impede the extraction of the strong phase induced by the penguin pollution. A detailed explanation of how to address this challenge in the amplitude analysis is provided in the following section.

In the $SU(3)$ partner channels $B^+ \to K^0_{1,A}\pi^+$ and $B^+ \to K^0_S a_1^+$, the quark transition is $b \to s(d\bar{d})$. This constitutes a Flavour Changing Neutral Current (FCNC) process without a tree-level contribution, thus it is purely a penguin process. The penguin Feynman diagram for the $b \to s(d\bar{d})$ quark transition is illustrated in Fig. 2.6.

The amplitudes from the $SU(3)$ related channels can be expressed as:

$$\begin{aligned} A(B^+ \to K^0_{1,A}\pi^+) &= -\frac{1}{\bar{\lambda}}\frac{f_{K_1}}{f_{a_1}}P^+, \\ A(B^+ \to a_1^+ K^0_S) &= -\frac{1}{\bar{\lambda}}\frac{f_K}{f_\pi}P^-, \end{aligned} \tag{2.13}$$

where $\bar{\lambda} = \frac{|V_{us}|}{|V_{ud}|} = \frac{|V_{cd}|}{|V_{cs}|}$, and f represents decay constants estimated from lattice QCD calculations. These constants account for factorisable $SU(3)$ breaking effects, considering the different CKM elements involved in the amplitudes and the use of different resonances to extract amplitude parameters.

Additionally, more decay channels could be added to the analysis to further constrain the penguin pollution. For example, the $B^0 \to K^+\pi^-\pi^+\pi^-$ channel could be included using a similar strategy to control the amplitude-level contributions. This approach aims to address non-factorisable $SU(3)$ breaking effects if the experimental precision is sufficient. Such effects may arise from new Feynman topologies, theoretical uncertainties, or sources of unknown origin. An example includes the Annihilation topology present in the $b \to s(d\bar{d})$ quark transition but absent in the $b \to d(u\bar{u})$ transition [46].

Summarising, a new approach to measure α involving the axial resonances of charmless B four-body decays was presented. A solution for α can be extracted using a frequentist approach similar to the CKMFitter Group [23]. A χ^2 can be constructed based on the theoretical relations given by Eqs. 2.12, 2.11, 2.9, 2.10, and 2.13, where the experimental results can be incorporated. The value of $\Delta\chi^2$ across the α range can be converted into a p-value scan, from which confidence intervals can be derived.

On one hand, from the time-dependent flavour-tagged amplitude analysis of $B^0 \to a_1^\pm\pi^\mp$, the $\alpha^\pm_{eff}$ solutions can be obtained. However, this information alone is insufficient to determine the α solution due to uncontrolled penguin pollution.

On the other hand, the pure-penguin process $B^+ \to K^0_S\pi^+\pi^-\pi^+$ allows extraction of the strong phase difference between $K_{1,A}$ and a_1, which relates to the strong phase difference between the $\alpha^\pm_{eff}$ solutions (δ_{strong}). Furthermore, the branching fractions of $B^+ \to K^0_{1,A}\pi^+$ and $B^+ \to a_1^+ K^0_S$ can be utilised to constrain the size of penguin pollution in the $B^0 \to a_1^\pm\pi^\mp$ system, denoted as P^+ and P^-.

With this approach, a unique solution for α can potentially be derived depending on the precision achieved by various experimental observables and the yet unknown value of the strong phase difference to control penguin pollution [45].

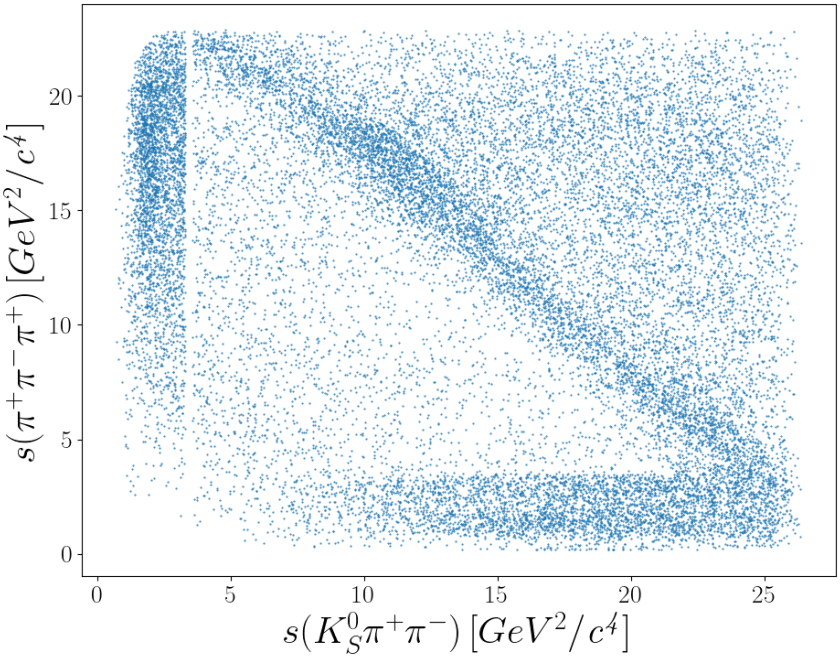


FIGURE 2.7: Axial Dalitz plot of a the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay.

2.2.2 THE $B^+ \to K^0_S\pi^+\pi^-\pi^+$ DECAY

In this section, we present the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay. Initially, we aim to perform a branching fraction measurement to obtain the absolute values before conducting an amplitude analysis. The amplitude analysis will help us extract the relative branching fractions and phases between the $B^+ \to K^0_{1,A}\pi^+$ and $B^+ \to a_1^+K^0_S$ amplitudes. Our ultimate goal in both the $B^0 \to \pi^+\pi^-\pi^+\pi^-$ and $B^+ \to K^0_S\pi^+\pi^-\pi^+$ channels is to determine the relative phases between the axial resonances and its branching fractions through this amplitude analysis.

However, as the axial resonances do not overlap in the PHSP, it is challenging to directly extract the relative phases between these amplitudes a_1^+ and a_1^- in the B^0 decay or K_1^0 and a_1^+ in the B^+ decay. To address this issue, we introduce the vector-vector (VV) contribution to the data. This contribution connects these resonances within the PHSP, serving as a reference for determining the relative phase between the axial resonances. In the B^0 decay, the $\rho^0\rho^0$ amplitude is used, while in the B^+ decay, the $K^{*+}(892)\rho^0$ amplitude is utilised.

Thus, the analysis employs an inclusive mass window covering both axial and VV contributions within the charmless region of the PHSP. Figure 2.7 illustrates the axial Dalitz plot for the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay within this inclusive charmless mass window. The plot's axes represent the two axial contributions or three-body resonances: $a_1^+ \to \pi^+\pi^-\pi^+$ and $K_1^0 \to K^0_S\pi^+\pi^-$. In the diagonal of the plot, the $K^{*+}\rho^0$ VV contribution can be appreciated, linking both axial resonances located in the corners of the PHSP.

A detailed explanation of the inclusive branching fraction measurement is provided in Chapter 5. The following section offers a brief introduction to the amplitude analysis of the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay.

One of the most critical aspects of performing a successful amplitude analysis is identifying the different amplitudes contributing to our final state within the studied mass window. This requires recognising the decay possibilities of the a_1 and $K_{1,A}$ mesons, as well as the VV contributions and other potential amplitudes within the mass window.

USC

THE a_1 RESONANCE This case is relatively simple. The dominant decay mode is $a_1^+ \to \rho^0\pi^+$, with the a_1^+ studied through its decay into three charged pions. Additional intermediate states may be considered, such as the σ, $f_0(980)$, $f_0(1370)$, and $f_2(1270)$, all of which decay to $\pi^+\pi^-$. Thus, the main process to be analysed is $B^+ \to a_1^+(\to \rho^0(\to (\pi^+\pi^-)\pi^+)K^0_S$.

Decay Channel	$K_1^0(1270)$	$K_1^0(1400)$
$K_S^0\rho^0$	(42 ± 6) %	(3.0 ± 3.0) %
$K_S^0\omega^0$	(11.0 ± 2.0) %	(1.0 ± 1.0) %
$K^{*+}(892)\pi^-$	(16 ± 5) %	(94 ± 6) %
$K_S^0 f_0^0(1370)$	(3.0 ± 2.0) %	(2.0 ± 2.0) %
$K_0^{*+}(1430)\pi^-$	(28 ± 4) %	-

TABLE 2.1: Decay modes and the partial widths of the $K_1(1270)$ and $K_1(1400)$ resonances. The values are extracted from the PDG [49].

THE $K_{1,A}$ PARTICLE This case is more complex because the $K_{1,A}$ particle is a weak eigenstate. It must be studied through the mixing of mass eigenstates, which are the observable states in the PHSP. The $K_{1,A}$ and $K_{1,B}$ weak eigenstates are related to the $K_1(1270)$ and $K_1(1400)$ mass eigenstates through a weak rotation parameterised by the θ_{K_1} angle:

$$\begin{pmatrix} K_1(1270) \\ K_1(1400) \end{pmatrix} = \begin{pmatrix} \sin\theta_{K_1} & \cos\theta_{K_1} \\ \cos\theta_{K_1} & -\sin\theta_{K_1} \end{pmatrix} \begin{pmatrix} K_{1,A} \\ K_{1,B} \end{pmatrix} . \tag{2.14}$$

In this way, the amplitude $A(B^+ \to K_{1,A}^0\pi^+)$ can be expressed in terms of the mass eigenstates:

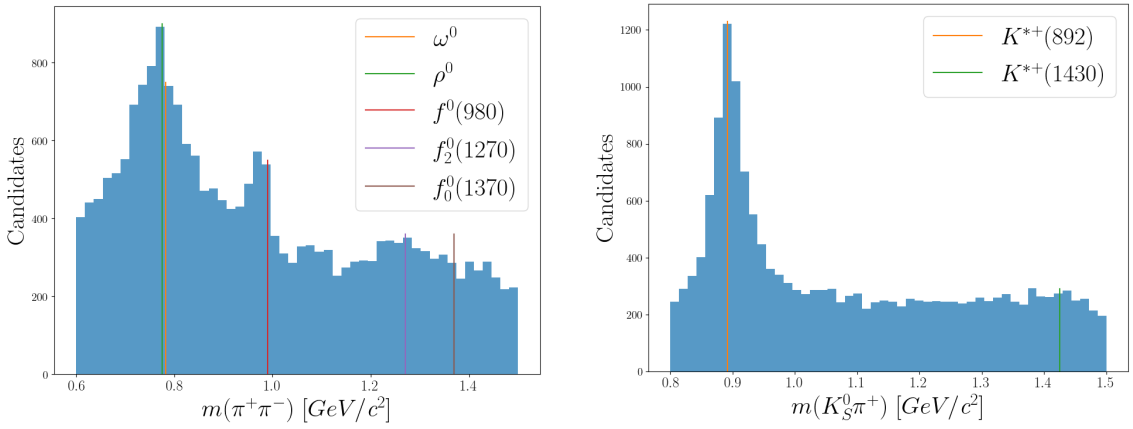
$$A(B^+ \to K_{1,A}^0\pi^+) = \sin\theta_{K_1} A(B^+ \to K_1^0(1270)\pi^+) + \cos\theta_{K_1} A(B^+ \to K_1^0(1400)\pi^+) . \tag{2.15}$$

This introduces additional complexity to the analysis because of the extra phase, for which there is limited information [47, 48]. Thus, the $B^+ \to K_{1,A}^0\pi^+$ amplitude necessitates studying the two resonances across the PHSP as independent amplitudes with their own decay processes. This can be seen as a nested 3-body amplitude analysis of the $K_1(1270)$ and $K_1(1400)$ resonances within the PHSP of a 4-body B-meson decay. The decay modes for the axial strange mesons are reported in Table 2.1, The decay amplitude can be represented as $B^+ \to K_1^0(1270)(\to K_S^0\pi^+\pi^-)\pi^+$ and $B^+ \to K_1^0(1400)(\to K_S^0\pi^+\pi^-)\pi^+$.

THE VV CONTRIBUTIONS In our mass window, in order to determine the phase between the axial contributions, we include the most dominant VV amplitude, that is $B^+ \to K^{*+}(892)(\to K_S^0\pi^+)\rho^0(\to \pi^+\pi^-)$. However, other amplitudes might also fall within our mass window, complicating the analysis. This could include higher spin resonances such as $K_0^{*+}(1430)$ or $f_2(1270)$, or spin o resonances like $f_0(980)$. Identifying and disentangling these contributions from the axial cascade decays is crucial. In other words, we must differentiate the axial cascade topologies from the VV topologies for resonances within our mass window. These additional amplitudes include $K^{*+}(1430)f_2(1270)$ and $K^{*+}(892)f_0(980)$, where the K^{*+} mesons decay to $K_S^0\pi^+$ and the f mesons decay to $\pi^+\pi^-$.

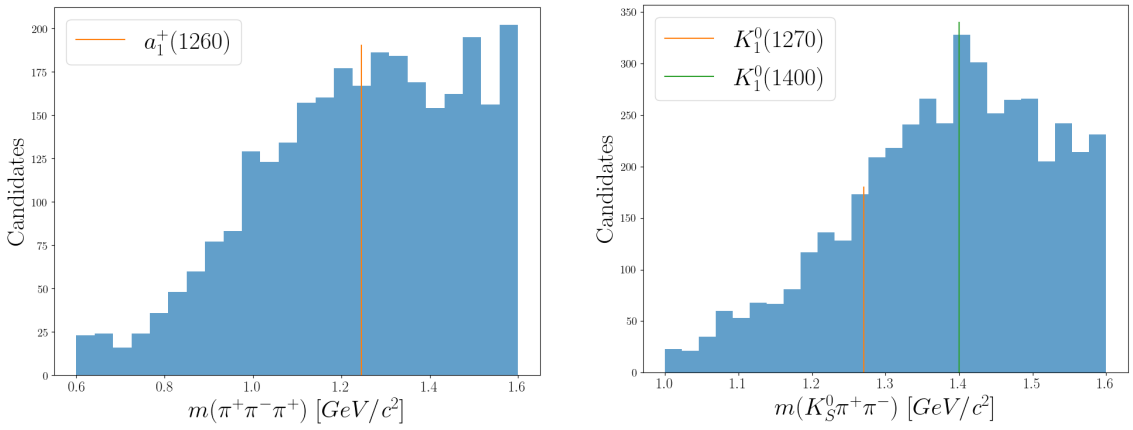
Finally, to conclude this chapter, we highlight the aforementioned resonances on the data PHSP projections. Fig. 2.8 shows the two-body resonances, while Fig. 2.9 displays the three-body ones. A significant contribution from the K^{*+} and ρ^0 resonances is evident, as shown in Fig. 2.7, due to the dominant VV contribution to the PHSP. The main resonances resulting from the K_1 and a_1 contributions are also visible, indicating the feasibility of the amplitude analysis on this data. The data used for these plots corresponds to one of the samples utilised for the branching fraction measurement presented in Chapter 5.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) $\pi^+\pi^-$ (B) $K_S^0\pi^+$

FIGURE 2.8: 2-body resonances in the PHSP of the $B^+ \longrightarrow K_S^0\pi^+\pi^-\pi^+$ decay.



(A) $\pi^+\pi^-\pi^+$ (B) $K_S^0\pi^+\pi^-$

FIGURE 2.9: 3-body resonances in the PHSP of the $B^+ \longrightarrow K_S^0\pi^+\pi^-\pi^+$ decay.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 3

THE LHCB EXPERIMENT

The LHCb experiment, is a single-arm spectrometer that covers the forward region of the $p-p$ collision, where most of the $b\bar{b}$ quark pairs are produced. Its primary focus of study is the CP violation in B decays [12], but thanks to the precision on the momentum and vertexing determination in the low p_T region, it can be used to study CP violation in charm decays [13], obtaining an outstanding precision in strange physics [50] as well. The geometrical acceptance of the detector is $2 < \eta < 5$ [51], where η is the pseudorapidity defined as

$$\eta = \frac{1}{2} \ln \frac{p + p_z}{p - p_z} = -\ln \tan \frac{\theta}{2}, \tag{3.1}$$

and the z-axis in the LHCb experiment is considered to be the forward direction, i.e. the direction of the proton beam-pipe.

It is located at CERN, on one of the $p-p$ interaction points at the LHC, which is the biggest particle accelerator ever created, with a main ring of 27 km long [52]. In the LHC, the protons are compacted in bunches and accelerated to collide in the interaction points, where a collision happens every 50 ns. To inject protons into the LHC, a large infrastructure is required. First, the protons must be linearly accelerated and grouped into bunches. Then, before being injected into the LHC, they are further accelerated using two different circular accelerators to provide the necessary boost.

In the LHCb detector between 1-2 $p-p$ collisions per bunch crossing are expected, which is a lower luminosity than the nominal LHC value. This is because the LHCb detector is built on the low p_T region, close to the beam-pipe, with a lot of radiation that can damage the detectors. The luminosity is decreased right before the detector, this is obtained by a magnet that distort the beam trajectory so that the beams do not collide front to front but slightly displaced. The energy collision varied through the different data taking periods at the LHC, from $7-8\,TeV$ at the centre-of-mass during Run 1 from 2011 to 2012, to $13\,TeV$ in Run 2 from 2015 to 2018. During the current Run 3 running period, the luminosity conditions will increase, the collision rate will be increased to 25 ns and the luminosity will be increased by a factor 5. In addition, the energy will be also increased to $14\,TeV$ [53], that is the maximum energy for the protons inside the LHC.

Data taking at LHCb is very challenging due to the high collision rate and the high number of tracks per event. There is a complex trigger system that allows the detector to store the desired data and remove the non-interesting events online. Afterwards, the data is further selected offline through stripping lines, from which one can apply certain requirements to data obtaining the signal of interest; in the case of this thesis, the $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ decay. Once the data is selected one must apply a tighter selection in order to obtain a pure sample of the signal of study to be finally analysed.

In this work, the full data released by the LHCb detector will be analysed, corresponding to the Run 1 and Run 2 of the LHC, yielding to an integrated luminosity of $9\,fb^{-1}$ in the LHCb experiment.

3.1 SUBDETECTORS

USC UNIVERSIDADE DE SANTIAGO

The LHCb detector is a multi-purpose instrument where various sources of information are combined to extract the signals of interest. The primary measures used are tracking, particle identification, calorimetry,

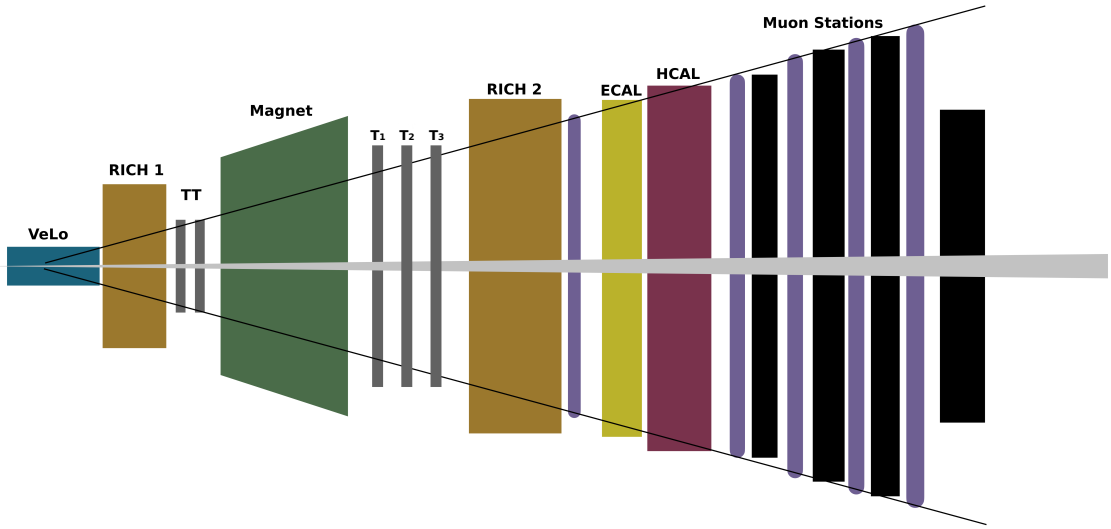


FIGURE 3.1: Schematic picture of the LHCb detector

and muon detection. Understanding how these different measurements are combined is crucial for extracting meaningful data from the detector. A schematic representation of the LHCb detector, showing the different subdetectors, is presented in Fig. 3.1.
Tracking measures the momentum and decay time of particles within the detector. This is achieved through the Vertex Locator (VeLo), tracking stations and the magnet. These components work together to determine the trajectory and momentum of charged particles as they move through the detector.
The particle identification system determines the velocity of charged particle tracks. By combining velocity information from PID detectors (such as RICH1 and RICH2) with momentum data from tracking, the mass of the particle associated with the track can be calculated, thereby identifying the particle.
Calorimeters measure the energy of particles. The electromagnetic calorimeter (ECAL) measures the energy of photons and electrons, while the hadronic calorimeter (HCAL) measures the energy of hadrons. Calorimetry also helps in particle identification, as electrons are stopped in the ECAL and hadrons are stopped in the HCAL, facilitating the separation of these types of particles.
The muon system identifies muons, which are the only particles that can travel through the entire detector without being stopped. If a track leaves hits in the muon stations, it is highly likely to be a muon.

Combining information from these subdetectors enables comprehensive particle identification and track reconstruction. However, due to the high track multiplicity in each event, achieving effective track separation and particle identification remains challenging. Statistical techniques are employed to probabilistically reconstruct tracks and classify particles, thereby obtaining the necessary variables for the particle physics analysis.
Additionally, researchers continuously develop new ideas to expand the capabilities of the LHCb detector beyond its original design. It remains competitive in measuring the mass of heavy leptons [54] and is actively involved in dark matter searches [55]. These advancements, utilising established hardware, underscore the evolving role of the LHCb detector in particle physics research and emphasise the importance of developing new tools to push the boundaries of scientific exploration.

3.1.1 TRACKING

Now that we understand the relevance of the different systems of the LHCb detector and how they can be combined, let us dig into each one, starting with the tracking system.
The tracking system measures the momentum of charged particles. It consists of silicon-based detectors and a magnet. By measuring how the magnet curves the trajectory of charged particles, one can infer their momentum by knowing the value of the magnetic field. The biggest limitation on the precision of the momentum value is the knowledge of the magnetic field. In addition, the acceptance in momentum of the detector also depends on the value of the magnetic field. For very low momentum tracks, the magnet will reject the particle from the LHCb geometrical acceptance, and at very high momentum values, the magnet will not be powerful enough to curve the trajectory of the particle in a meaningful way for the tracking

detectors, so no measure of its momentum can be inferred. The acceptance in momentum is between 5 GeV/c and 200 GeV/c.
The momentum precision ($\Delta p/p$) is about 0.4% for the lowest reconstructible momentum values of charged particles and about 1% for particles around 200 GeV/c [56].

3.1.1.1 VELO

The main goal of the LHCb detector is the study of CP violation in B-mesons. One of the most important features of the LHCb detector is its ability to locate the secondary vertices emerging from the p-p collisions, from which the timing information related to the decay lifetime is obtained. To fulfill this goal, the VeLo is used. Thanks to the incredible spatial resolution of this detector, the lifetime measurements of B-decays can be performed with high precision. This precision is achieved with a system of layers located around the collision point. These layers are separated by a few centimetres over a metre in length.
The layers consist of silicon strips in the radial and azimuthal directions [57], so that, with these coordinates, the position of a charged particle's hit in the layer can be located.

There are some important variables commonly known at the LHCb to study these features from B decays. These are:

- **Primary Vertex (PV):** The vertex of the colliding protons.
- **Secondary Vertex:** The decay vertex of a particle produced at the primary vertex.
- **Flight Distance (FD):** The distance between the PV and the secondary vertex where the particle decays.
- **Impact Parameter (IP):** IP is the shortest distance between the trajectory of a particle and the PV.
- **DIRA:** Defined as the cosine of the direction angle between the reconstructed momentum and the line that connects the primary vertex with the secondary one. For perfectly reconstructed particles, it is 1.
- $\mathbf{p_T}$: The component of the momentum perpendicular to the beam-pipe. In this way, the z-axis is defined as the pipe line direction.
- **DOCA:** Distance of Closest Approach, the minimum distance between two tracks inside the detector.

The calculation of some of these variables, such as the IP, DIRA or FD, requires the knowledge of the primary vertex associated with the candidate. Therefore, these variables have an associated uncertainty, as they depend on the precision obtained in the determination of the vertices or the tracks. Thus, in many algorithms used to reduce the background noise, information from the quality (or χ^2 value) of these variables is used, as B decays usually have these variables well-defined.

3.1.1.2 MAGNET

The LHCb detector uses a warm dipole magnet with 4 T · m of bending power [58]. To avoid systematic biases, the polarity of the magnet is changed periodically. As the dipole magnet surrounds the beam-pipe, it also bends the protons inside it. In order to correct for this effect in the colliding protons, two magnets are located on the sides of the LHCb detector, refocusing the beams.

The momentum of the charged particles is computed from the curvature caused by the magnetic field on the bending plane. The magnetic field in LHCb is primarily perpendicular to the beam-pipe and is parameterised by obtaining a field map along the detector. The bending plane is defined as the plane perpendicular to the direction of the magnetic field, in which the curvature of the particle's trajectory is observed. From the Lorentz formula for the magnetic force and the fact that this force acts as a centripetal force, the momentum of the particle can be measured from the radius of curvature on the bending plane and the value of the integrated magnetic field through the particle's trajectory.

$$\vec{\mathrm{F}} = q\vec{\mathrm{v}} \wedge \vec{\mathrm{B}} = m\frac{\mathrm{v}^2}{R} \quad \rightarrow \quad R = \frac{m\mathrm{v}}{q\mathrm{B}\sin(\alpha)} = \frac{p}{q\mathrm{B}\sin(\alpha)} \quad \rightarrow \quad p = q\,R\,\mathrm{B}\,\sin(\alpha)\,. \tag{3.2}$$

In this expression, α is the angle between the magnetic field and the velocity of the particle, and R is the curvature radius on the bending plane. From this expression, it is clear that the precision with which the radius of curvature and the magnetic field are measured are the limiting factors in determining the particle's momentum.

3.1.1.3 T STATIONS

Besides the VeLo, there are two additional modules to measure the momentum of charged particles; these are the T stations. One is located right before the magnet, or upstream, and the other module is downstream of the magnet. The TT (Tracker Turicensis), before the magnet, is utilised to improve the reconstruction algorithms. Thanks to the residual magnetic field, naive momentum estimations can be associated with each track, facilitating the momentum guess to match the tracks upstream and downstream of the magnet. Moreover, it is crucial to measure the momentum of particles that can decay after the VeLo, such as the K_S^0. Downstream of the magnet, the last tracking stations are located: $T1$, $T2$, and $T3$. These stations are constituted by two different structures: the inner tracker and the outer tracker. The inner tracker, being closer to the beam pipe, has a higher occupancy; therefore, it is constructed with pixel silicon sensors [59]. The outer tracker consists of drift tubes with a gas mixture that ensures a fast readout [60].

3.1.1.4 RECONSTRUCTION, CALIBRATION AND TYPES OF TRACKS

The track reconstruction is performed from the hits in the different tracking detectors; by combining all the information, the track and the associated momentum can be obtained. Once the track is located along the detector, more information from other subdetectors can be added. In addition, these tracks can be combined to build signal candidates and compute more complex physical variables such as invariant masses or timing information from the parent candidate.
As our final state is purely hadronic and no neutrals are involved in the reconstruction[1], we will not give further details on the reconstruction of the different types of particles. The reconstruction efficiency for B-tracks crossing the entire detector is 95% [56].
It is important to note that, due to the dirty environment of the $p-p$ collisions where lots of different tracks are produced, the reconstruction algorithms may find some tracks that are just a random combination of hits and are not related to any physical particle. We refer to these tracks as *Ghost* tracks.

Many features enter into the calibration of the detector, with the most important being the detector misalignment and the momentum scale calibration. Both calibrations are performed by fitting to well-known resonances that are widely used in the detector, such as $J/\Psi \rightarrow \mu^+\mu^-$ or $B^+ \rightarrow J/\Psi K^+$ [17]. While misalignment is usually spotted by the difference in the reconstruction of positive and negative charge particles, the momentum scale bias is typically corrected by the measured bias in the mass of the reconstructed particle.

Due to the three different modules used to measure the track of charged particles, we can find different types of tracks depending on which of these modules are used for reconstruction:

1. **VeLo track:** These are tracks that are only present inside the VeLo. These tracks can be used to estimate with better precision the position of the primary vertices.
2. **Long track:** When all the modules are used to reconstruct the track, i.e., the track has associated hits in the three modules. This type of track has the highest precision in momentum determination.
3. **Downstream track:** This type of track has no associated hits on the VeLo but has hits on the other modules, possibly because the parent particle has a significant lifetime, such as the K_S^0. The precision in momentum determination is lower.
4. **Upstream track:** In this type of track there are only hits before the magnet, no momentum information can be associated with this track. These tracks may be interesting to use in Flavour Tagging algorithms.
5. **T track:** These tracks only leave hits downstream of the magnet and are usually related to long-lived particles.

3.1.2 PARTICLE IDENTIFICATION

The particle identification of charged particles is mainly performed by the Cherenkov detectors, where the velocity of the tracks can be measured. By combining this information with the track's momentum, one can identify the particle by its mass,

$$m = \frac{p}{v}\,. \tag{3.3}$$

These Ring-Imaging Cherenkov Detectors (RICH) are located before and after the magnet. The second detector is designed to measure the velocity of particles with higher momentum [61].

[1] The K_S^0 particles in LHCb are always reconstructed by their children pions.

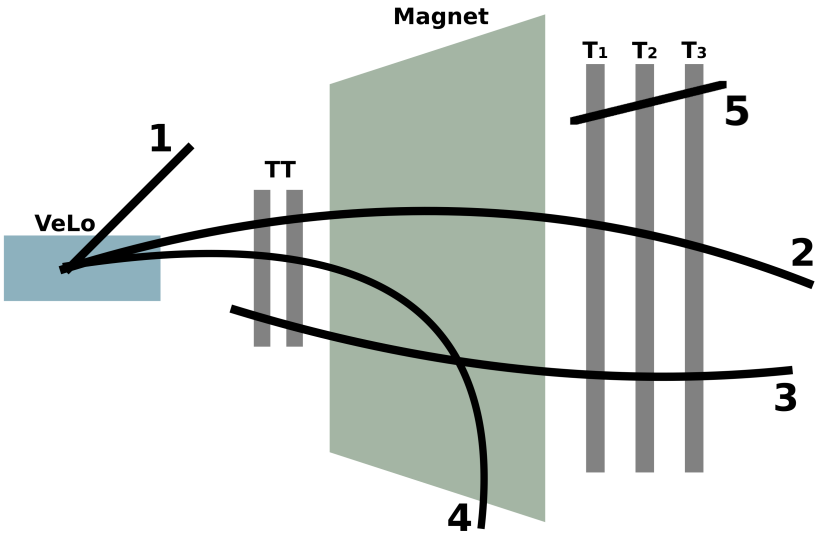


FIGURE 3.2: Type of tracks at LHCb. 1 represents a VeLo track, 2 is a Long track, 3 represents a Downstream track, 4 is an Upstream track and 5 a T track.

To measure the velocity, these detectors make use of Cherenkov radiation, which occurs when a particle travels faster than the speed of light in a medium. This radiation forms a cone with an aperture that depends on the particle's velocity. Projected into the photon detector plane, the signature of each track is a ring. From the diameter of these rings, the velocity of the associated track can be obtained. The likelihood of accurately identifying each charged hadron consistently remains above 90% for a misidentification probability of 5% [62].

These RICH detectors are crucial for separating pions from kaons and protons, as there is no other way to identify them along the detector. However, the information from these detectors is not the only source used to identify the different particles. The most common variable used in LHCb for particle identification is a Multivariate Analysis (MVA) based variable that utilises information from the RICH detectors as well as all the other detector components to obtain PID decisions. This variable is called *ProbNN*.

3.1.3 CALORIMETRY AND MUON DETECTION

The calorimetry and the muon stations at the LHCb detector are mainly used to separate different kinds of particles. These detectors generally have a very low resolution; they consist of multiple cells where the relevant information is whether the detector fired and where.

As particles travel through matter, they have a probability of being stopped in the medium. Photons and electrons are the easiest particles to stop; hence the electromagnetic calorimeter is located first in the detector. Then the hadrons, such as pions, kaons, protons, and neutrons, are stopped in the hadronic calorimeter. After these calorimeters, some iron material is placed in the detector to stop the most energetic hadrons that could pass through the hadronic calorimeter. Finally, the remaining particles in the detector are the muons; hence, the last detectors in the LHCb are the muon stations.

These subdetectors are key in the first stage of the trigger. In this stage, the information from these detectors is used to spot signatures of decay products coming from B-meson decays. In addition, the calorimeter system is crucial for reconstructing neutral particles such as photons, which leave no other footprint in the detector.

3.1.3.1 THE ELECTROMAGNETIC CALORIMETER

The electromagnetic calorimeter [63] stops the electrons and photons that travel through the detector. As this detector is the only calorimeter with a high energy resolution [64], it can measure the energy of the photons and electrons. Electrons are very difficult to reconstruct due to bremsstrahlung radiation; highly energetic electrons emit photons when traveling through matter. However, the electromagnetic calorimeter can also be used to reconstruct the energy loss of the electrons. Additionally, before the electromagnetic

calorimeter, there are two submodules: the Scintillating Pad Detector (SPD) and the Preshower (PS). These submodules help to identify whether the incoming particles are neutral or charged, distinguishing also between light and heavy particles.

3.1.3.2 THE HADRONIC CALORIMETER

The hadronic calorimeter [63] is a detector that measures the energy of hadrons by stopping them. It has a very low energy resolution [64]; therefore, it is not used to directly measure the energy of hadrons, but instead, it is employed at the low-level trigger. When a high-energy deposition is detected by this detector, the trigger system is activated, making it an important component of the hardware-based trigger system, as will be discussed later. These high-energy depositions may indicate the presence of high-momentum hadrons likely originating from the decay of a B-meson.

3.1.3.3 MUON STATIONS

Finally, the muon stations are used to identify the muons, as these are the only particles that reach these detectors. If a track has several hits in the muon stations, the flag *is_muon* is set to True for that track. They are Multi Wire Proportional Chambers detectors [65] that count hits and can be used for muon reconstruction. Instead of locating the position of a hit, these detectors fire when a muon passes through them. By arranging many modules on a grid, the approximate location of the muon on each plane can be determined to produce a final muon track that can be associated with a track in the tracking system.

3.2 TRIGGER

The triggering system of a particle physics detector is a key part for its correct functioning as it decides how the collected data is stored. In p-p machines, a large number of different particles are produced, making it fundamental to decide what to store in order to manage computing resources effectively and to select the candidates of interest that will be analysed. The strategy is to decrease the detector readout rate through different stages, starting from 40 MHz and reducing it to 12.5 kHz during Run 2 [66], although it was originally planned to be reduced to 400 Hz [67]. In LHCb there are three different trigger stages, in each step a different strategy to handle the detector information is used to further reduce the detector readout and to finally store the candidates of interest. The different trigger lines select the interesting signatures in the detector to perform physics analysis. There are some general-purpose lines as well as very specific lines for the most relevant decays for the experiment. Deciding what to store and how depends on the physics program of the collaboration.

Triggering must be understood as a continuous process. During data-taking, the detector information must be reconstructed to obtain different physical signatures used to apply various trigger requirements. The detector reconstruction in the first trigger steps is very naive. The data is then stored in buffers that are utilised to perform detector calibration and alignment. In these buffers, the data can be further processed, and more complex trigger requirements can be applied before finally storing the data.

3.2.1 LEVEL-0 TRIGGER

The level-0 (L0) trigger corresponds to a hardware level. This is a crucial step for the trigger where a very basic detector readout is performed. Generally, some layers of the VeLo help to locate the primary vertices, then some information from the calorimeters and the muon chambers is used to locate key signatures on the detector. There are three main L0 trigger categories: the electromagnetic, the hadronic, and the muon L0 trigger. Events with high energy depositions in the calorimeters or muons with high p_T are stored to be processed by the following trigger stages. As the maximum readout rate of the LHCb detector is 1 MHz, this is the maximum allowed rate after the L0 trigger selection.

This stage is crucial for the correct functioning of the detector with the available computing resources. Since we are limited by the detector readout, having this initial stage that relies on simple signals in the detector, which do not require full reconstruction and detector calibration, is essential to reduce the load on the software trigger.

3.2.2 HIGH LEVEL TRIGGER

This is the software-based trigger stage. In this stage, the reconstruction must be performed and the tracks must be combined to establish decision criteria. A full detector readout is first performed at this trigger level. There are two stages of the High-Level Trigger (HLT): a first selection (HLT1) where a naive event reconstruction is performed, selecting only candidates of a single track or 2-track combinations, and a final selection (HLT2) where a full event reconstruction is performed to decide which events are stored and how. Some physics analyses need full event information for offline reconstruction, while for others, the information from the signal tracks is sufficient.

In the HLT1 stage, tracks with high impact parameter (IP) and/or high p_T are selected. Thus, a primary vertex reconstruction is performed in the VELO, where the tracks used for the reconstruction are limited to a small momentum window. In addition to the information from the VeLo detector, information from the electromagnetic calorimeter is also used to seek leptonic or photonic signatures, and the muon chambers are used for the muonic lines. The purpose of this stage is to reduce the rate to approximately 80 kHz [56].

The last stage of the trigger is the HLT2, where the final selection is applied to the data that will be analysed. Each HLT2 line is more specific and depends on the physics program of LHCb. In this stage, full event reconstruction is performed. There are mainly four different types of lines: the generic beauty trigger, the charm trigger, the muon trigger, and exclusive lines. The HLT2 category of interest for our analysis is the beauty trigger. In this trigger, the topological information from the detector is used, such as displaced vertices, distance between the signal candidate tracks, etc. With this information, partially reconstructed B decays are targeted, allowing a wide spectrum of B decays to be stored for further selection in the stripping stage.

3.2.3 TIS AND TOS CATEGORIES

Before going further, there is an important concept that must be learned for every analysis at the LHCb experiment. When a particular decay is being studied, dedicated trigger requirements must be imposed on an event to be selected. However, a signal event can be stored due to a different signal candidate whose tracks fired those trigger requirements. When this happens, we say that the trigger fired independently of the signal under study, defining the TIS (Trigger Independent of Signal) category. When a certain trigger stage is fired by the signal tracks, we say that the trigger fired on signal, i.e., the TOS (Trigger On Signal) category.

In general, a certain trigger requirement can be both TIS and TOS because there can be several triggering tracks in the event, leading to an overlap between both categories.

In our analysis, the signal under study is a purely hadronic decay; thus, our signal can only fire the Hadronic L0 trigger. Consequently, in our analysis, only the L0 Hadronic TOS and the L0 Global TIS categories will be analysed. We include the L0 Global TIS because there can be many events where our signal candidate is present, but those tracks did not fire the L0 trigger, while a different track from the event did fire it.

3.2.4 TRIGGERING ON RUN 3

The basic differences between the old and the upgraded detector are the adaptation to the new data-taking conditions and the removal of the L0 trigger [68]. In Run 3, there is an increase in energy, luminosity, and collision rate; hence, the detectors must be fast enough and ready for the new higher radiation environment. The primary modification in the subdetectors is the transition to a pixel-based tracking system, which permits greater occupancy during the detector readout.

The primary benefit of eliminating the hardware trigger is the ability to remove the p_T cut from data selection, allowing for the reconstruction of tracks near the beam pipe. This significantly affects decays involving strange quarks, as they propagate close to the beam pipe, and also influences the efficiency in the threshold region of the PHSP, where the decay products have lower momentum. To accomplish this demanding task, a significant modification has been made to the HLT1 step, which now employs Graphical Processing Units (GPUs) instead of CPUs for computation [69]. By using gaming GPUs, substantial computing power can be achieved at a lower cost. Utilising multiple GPUs allows for the parallelisation of detector readout and reconstruction, thus achieving this remarkable objective.

3.2.4.1 HLT2 SELECTION STUDY FOR RUN3

The HLT2 strategy for Run 3 is a bit different from the previous data-taking periods. Currently, the strategy is to develop more exclusive lines with the purpose of reducing the amount of combinatorial background in our stored data. The combinatorial background comes from the combination of random tracks in the detector that pass the selection criteria but do not belong to any signal of interest. Usually, they are combinations of kaons and pions that can resonate into light mesons. Therefore, by being more strict in the selection criteria and focusing on very specific signals of study, one can reduce the amount of background present in the data that would be removed anyway, leading to a reduction in disk space usage.

The HLT2 for Run 3 is currently being developed by the analysts with the help of the experts on the HLT software. Part of the work in this thesis involved developing trigger lines for specific decay channels. These lines are part of the strategy to measure the CKM α angle using only LHCb data, without requiring the reconstruction of neutral particles. With the help of additional analysts and the experts on the matter, we started to develop the trigger lines associated to $B^+ \to K^0_S hhh$ and $B^0 \to hhhh$ decays, where $h = \pi, K$. When $h = \pi$, these lines would constitute the lines to measure α.

Many different approaches were studied to obtain the data for these lines during Run 3. When building the lines, it is important to balance the efficiency and the rate of the line. Since these lines are inclusive in the PHSP, meaning the three-body and the two-body combinations are included in the mass window, it is very difficult to obtain a high efficiency and an affordable rate at the same time. A tight cut could bias the efficiency distribution across the PHSP, resulting in a poor acceptance of, for example, the axial, or three-body, contributions. Below are the main ideas that were studied:

- Build inclusive lines on PID, so that there were no PID cuts on the lines. However, the rate needed to perform this approach with reasonable efficiency would be very high, exceeding the maximum acceptable rate.
- Remove the p_T cut that highly impacts the PHSP production threshold of these decays. This idea is still present in the lines, but to control the rate, another tight cut must be applied.
- Use the MVA trained for the topological[2] lines as a previous filter for these HLT2 lines, so that only events passing this filter would be considered. This idea was very interesting, as the topological lines have a tight cut set on the MVA; we could smooth that MVA cut and apply a set of cuts more adequate to our signal of study. In this way, we could reduce the rate of the line while keeping the desired efficiency. The main problem with this approach is that a different tuning of the MVA during data-taking would impact the efficiency distribution on the PHSP and would make the extraction of the efficiencies more difficult, but this is very unlikely to happen. More concerning is the fact that with this approach, there could be a significant overlap between these lines and the topological ones, resulting in a waste of computing resources and disk space.

Once all these possibilities were considered, it was found that, to obtain a reasonable value for the rate, a tight MVA cut should be applied or a tight cut on the p_T of the lines. The tight cut on p_T is something that we want to avoid as it impacts the threshold of the PHSP. Thus, we need to study the impact of having a tight MVA cut on the different regions of the PHSP. If a particular region is clearly removed by the tight MVA cut, a different approach must be considered. If a tight MVA cut is affordable in terms of efficiency, we could use the topological inclusive line as our HLT2 line and apply a tighter selection afterward to isolate our signal of study. This can be performed in the *Sprucing*, which is a central selection that runs offline on top of some HLT2 line.

In order to test the feasibility of this approach and decide which strategy to follow, the following test was performed: We generated a Monte Carlo (MC) sample in the PHSP mass window of interest. We obtained a selection with an affordable rate using the strategy with the topological MVA as a previous filter for the line, so a looser MVA cut could be applied. Then we applied to this line the tight cut for the MVA that the topological HLT2 line uses. Thus, we had two different versions of the HLT2 line, one matching the MVA cut from the topological HLT2 line (*tight MVA*) and the other with a looser MVA cut (*loose MVA*). Finally, we passed the trigger selection to our MC sample and compared the PHSP distribution obtained from the output of these lines. If the distributions from the different MVA cuts are similar to the original MC distribution, this means that the MVA is not biasing our efficiency across the PHSP, so we can use the topological HLT2 line for our physics purpose.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[2] In Run 3, the topological trigger is performed by a MVA algorithm that uses similar information as in Run 1 and Run 2 but employs a machine learning algorithm to distinguish the signal candidates from the background. The MVA is trained with a mixture of two- and three-body B decays.

We performed this study for the $B^+ \to K^0_S \pi^+\pi^-\pi^+$ HLT2 line, as shown in Fig. 3.3, and for the $B^0 \to \pi^+\pi^-\pi^+\pi^-$ line, shown in Fig. 3.4. In both cases, despite the low statistics in the selected sample due to the low efficiency of multi-body hadronic lines, we can appreciate that the PHSP distributions using the tight and loose MVA cuts are very similar, populating all the regions of the PHSP. Thus, we conclude that these lines can be studied by applying further selection criteria to the HLT2 topological line and no specific lines are needed. It is worth noting that the efficiency value obtained from this strategy is very close to the efficiency obtained in Run 1 and Run 2.

3.3 LHCB DATA-FLOW

We have presented the detector and the data-saving strategy so far. To understand how the data is treated [70, 71] and finally analysed, we still need the big picture of the LHCb data-flow, as shown in Fig. 3.5. This chart shows the different steps to produce a physics result at LHCb, along with the applications used to perform each step. So far, we have discussed the triggering and the reconstruction. To develop the trigger software (HLT), the *Moore* application is used, whereas *Brunel* runs the software for the reconstruction. All the applications used in the central LHCb software are embedded in the *Gaudi* framework [72].

The LHCb data-flow can be revisited starting from the point of view of the colliding protons. Once a colliding p-p pair is found inside the VeLo detector, the emerging particles start to travel through space. Those particles that fall within the acceptance of the detector are captured. Before the software system realises that an interesting signature has just occurred in the detector, the next proton collision is already happening. During this time, a naive detector readout is deciding whether to send the information from this event to the buffer or not. If a positive decision is given by any of the hardware triggers, the information is sent to be processed. In the buffer, many things are happening simultaneously; it is a CPU farm where all the computations are performed, from the tracking reconstruction, triggering, and storage to the detector calibration. The calibration is performed online as the data from the present run is taken.

Finally, if the trigger requirements are fulfilled, the information from the reconstruction is stored on disk or tape to be ultimately analysed. Depending on the needs of the analysis, the information from all the subdetectors can be stored to perform a more detailed reconstruction offline.

However, the LHCb data-flow not only covers what is happening online. Once the data is stored, a central Stripping campaign saves the information needed for the analysis. In addition, the detector simulation also happens centrally to avoid wasting computing resources; as each event costs a lot of computing power, the produced information from the simulation must be properly stored. In the following sections, these two central productions will be treated. The *DaVinci* application is utilised for the Stripping selection, whereas *Gauss* is the software that includes all the tools needed for the simulation.

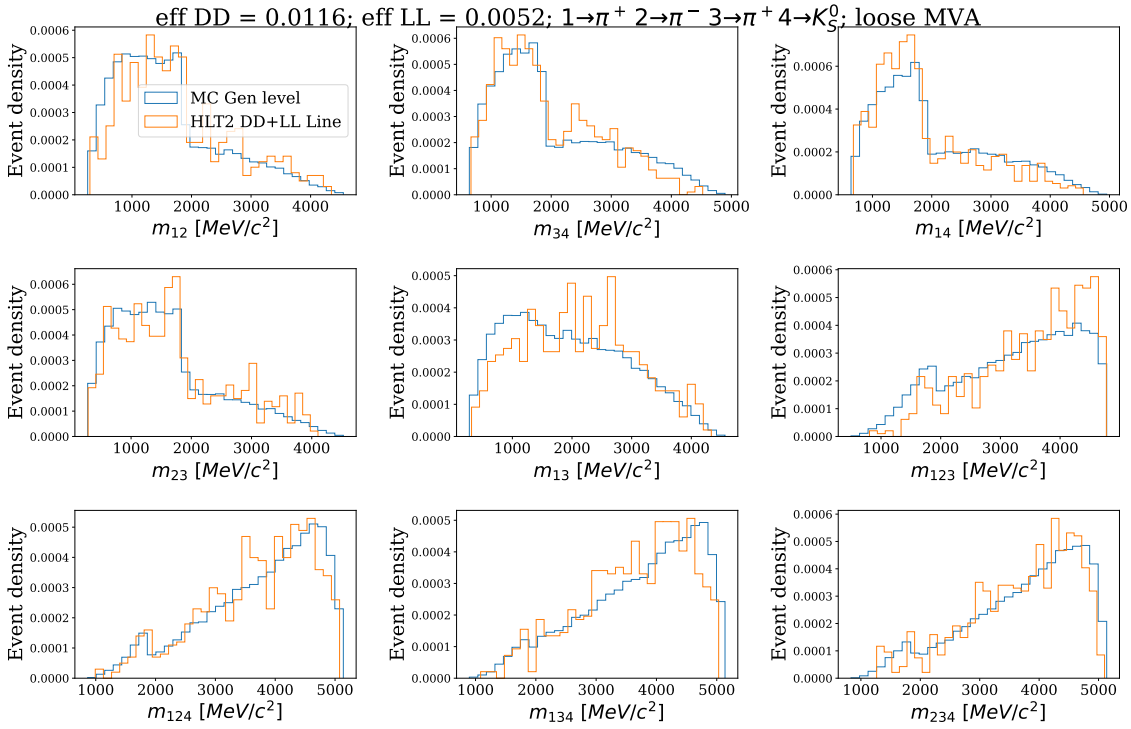
Moreover, to finally produce a result, the data must be analysed. Thus, one needs to obtain the desired physical observable of the signal candidates from the detector information or from the output of the reconstruction algorithms. The *DaVinci* software allows the analyst to obtain the physical variables from the detector information, acting as a bridge between the engineering and the physics. In Chapter 5, the analysis will be presented, where many tools such as *Python* and *ROOT* are utilised to produce the data tuples and to compute the different algorithms to statistically process the data, obtaining, in this way, the final result.

3.3.1 STRIPPING

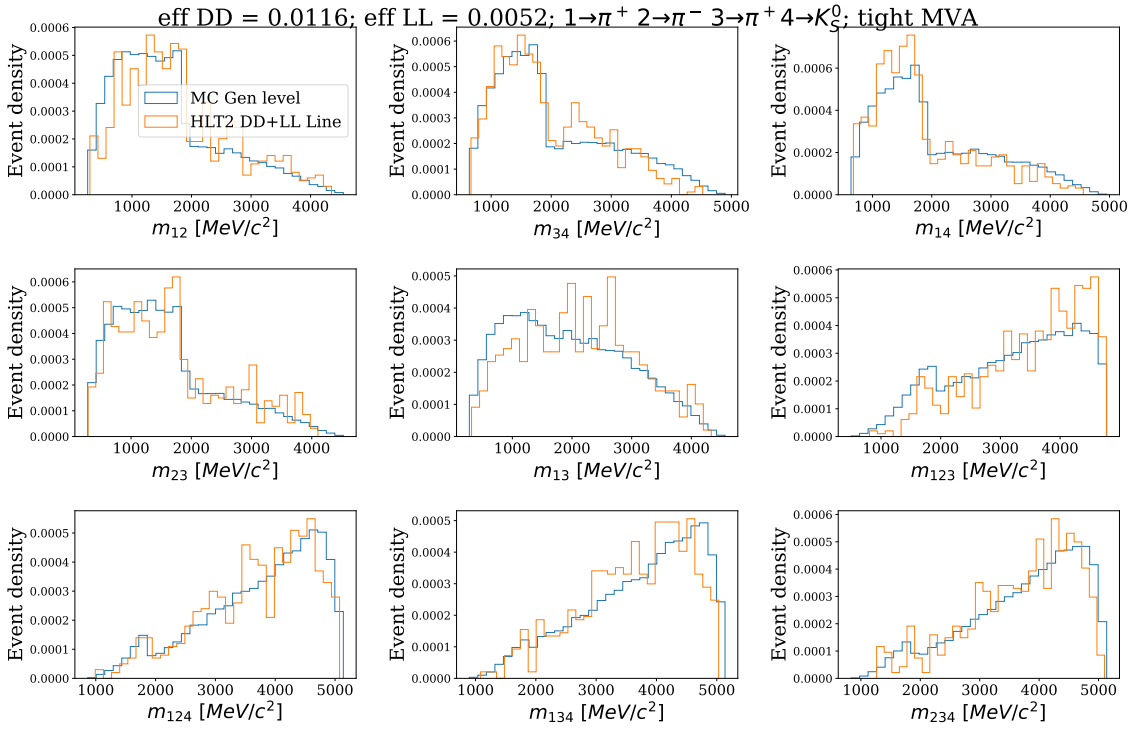
The stripping at LHCb is a central selection process that allows users to store on disk the physical candidates to be analysed. It can be stored in many formats and, depending on the analysis, some information from the event can be removed to save disk space. The stripping lines usually require some trigger criteria and then apply further selection on the data that fulfill these requirements. The data after the stripping can be stored as *DST* files, where the information from all the reconstructed particles and vertices in the event is stored, or in *MDST* format, where only the tracks and vertices related to the signal of study are stored.

At the user level, one can build the data tuples for an analysis from a stripping line, by adding some extra requirements and the desired information for the analysis, such as constraining the mass of some intermediate particles of the decay, re-running the reconstruction, or computing new variables of interest for the analysis. There are periodic re-stripping campaigns to add new stripping lines that may be interesting for new analyses with LHCb data.

USC



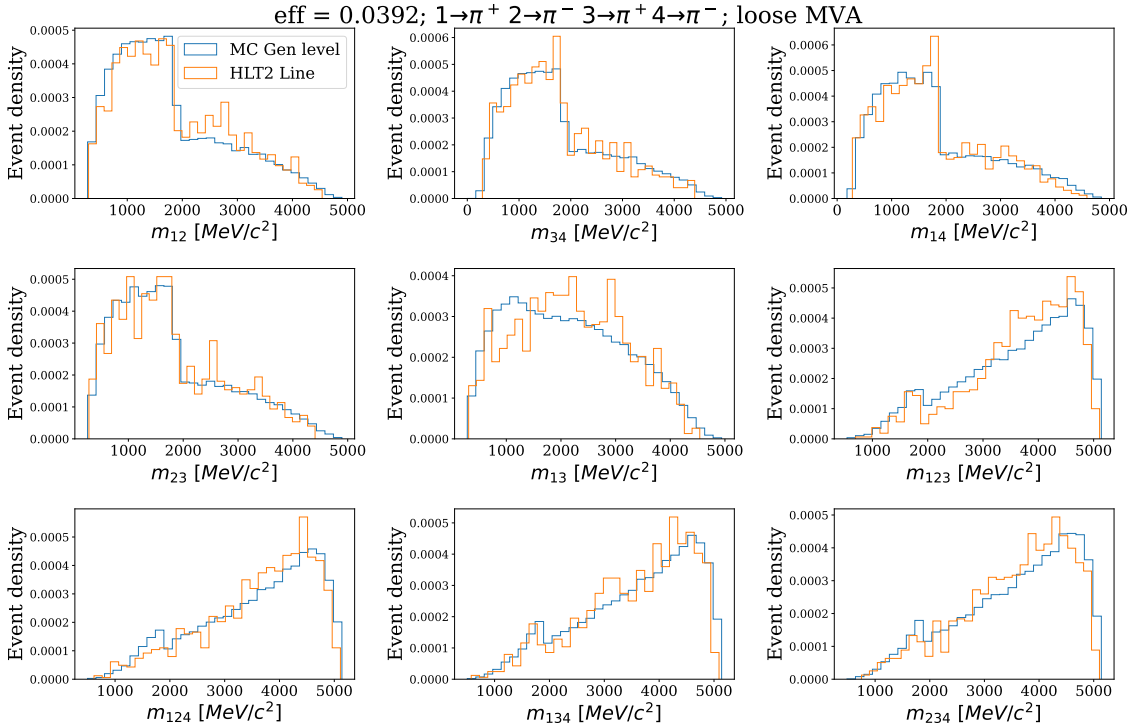
(A) Loose MVA cut.



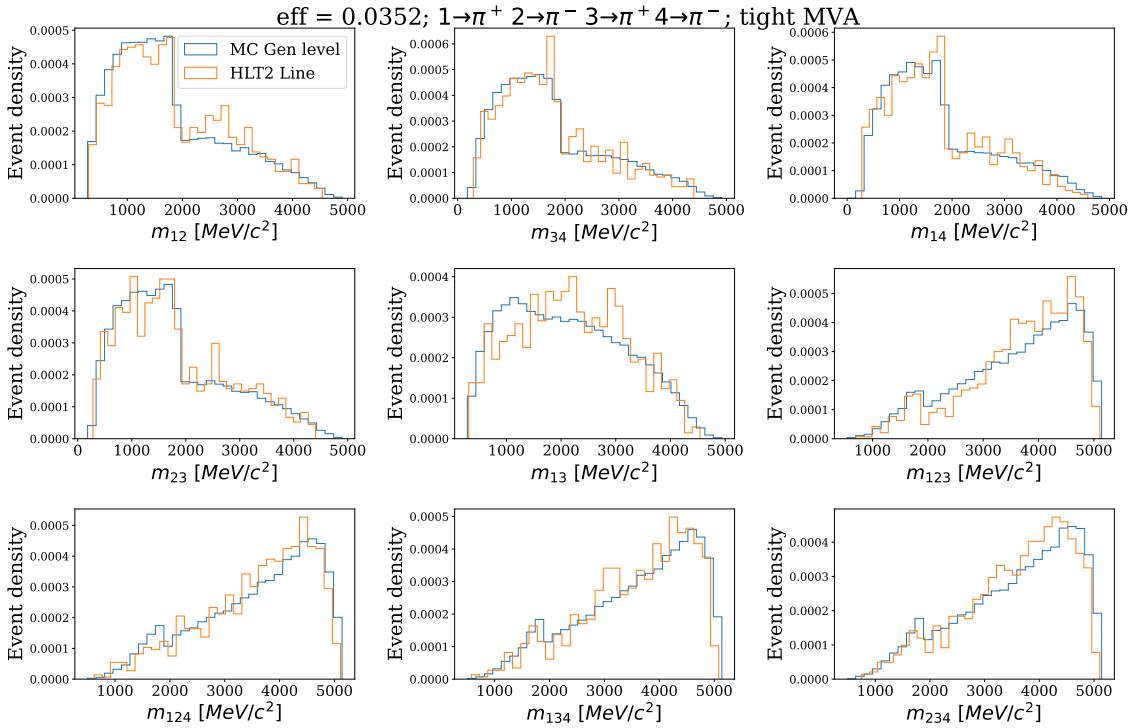
(B) Tight MVA cut.

FIGURE 3.3: PHSP projections for the $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ decay, the MC is generated with Run 3 conditions. The distributions before and after the trigger selection are shown, in blue the distribution of the MC at generator level and in orange the distribution of the selected events by the HLT2 lines.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) Loose MVA cut.



(B) Tight MVA cut.

FIGURE 3.4: PHSP projections for the $B^0 \to \pi^+\pi^-\pi^+\pi^-$ decay, the MC is generated with Run 3 conditions. The distributions before and after the trigger selection are shown, in blue the distribution of the MC at generator level and in orange the distribution of the selected events by the HLT2 line.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

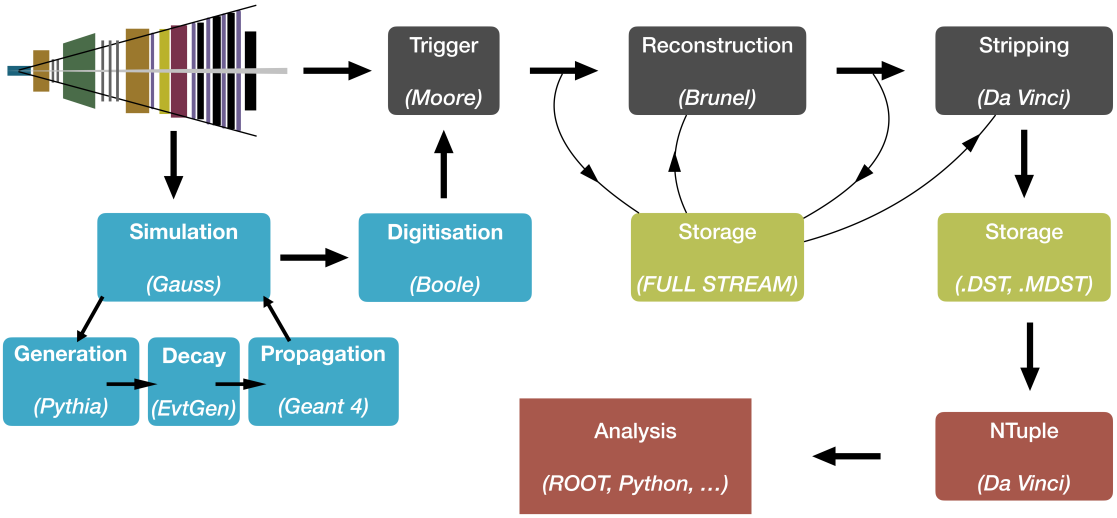


FIGURE 3.5: LHCb data-flow and its applications embedded in the *Gaudi* framework. In blue, MC simulation software is represented; in grey, central LHCb software; in yellow, data storage; and in red, user-level software. The thin lines in the chart indicate that some processes can be iterative, for example, the reconstruction can be run many times depending on specific demands.

3.3.2 SIMULATION

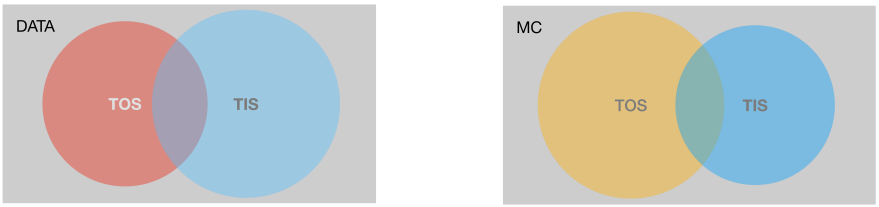
The simulation is one of the most important steps for any analysis at the LHCb experiment, as it is the only way to statistically treat all the physics processes that happen inside the detector, understanding, in an isolated way, the signal of interest. These processes range from the generation of the p-p collisions to the production of the particular physics process of interest and its interactions with the detector. In real data situations, it is very difficult to separate the signal of interest from other similar processes that pollute the signal region, known as background. These background sources can be similar physical processes or random combinations of tracks that survive the data-selection criteria. In this way, simulation plays a key role in analysis validation, as it is the only way to study the signal of interest without any contamination from different background sources.

The simulation is performed in many steps before entering the detector chain to mimic the data selection, as shown in Fig. 3.5. These steps are:

1. **Generation**: The generation step mimics the LHC conditions of energy and luminosity of the colliding protons to generate the collisions in a similar environment as in the LHCb detector. *Pythia* [73] is the tool used to generate the p-p collisions; it can be tuned to generate $b\bar{b}$ quark pairs from the collisions, producing the B-mesons.
2. **Decay**: Once the B-mesons (or the parent particle of study) emerge from the simulated p-p collision, they must be decayed through the model that follows the process of interest. *EvtGen* [74] allows us to generate a model and decay the parent particle to its children following that model. For example, some vector-vector decays can be generated with the polarisation given by some experiment or theoretical model.
3. **Propagation**: The generated particles must be propagated through the detector material, taking into account all the possible interactions with it. To simulate the behavior of the particles, a *Geant4* [75] model of the detector is built. With this software generating a reliable model of the detector, all the material interactions can be simulated. Additionally, the photon radiation through the medium is generated with *PHOTOS* [76], which describes more accurately the photon radiation of all the particles in the detector.
4. **Digitisation**: Now that we understand how the particles emerging from the p-p collisions in the LHCb environment travel through the detector material, we must simulate the response of the detector to these particles. As the detector produces digital signals, the digitisation process must be simulated to mimic this detector response. This means transforming the detector interactions with the traveling particles into digital signals that can be processed by the trigger system. *Boole* [77] is the software utilised for this step.

At this stage, the simulation can be incorporated into the same software that processes the data, beginning

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) Data. (B) MC.

FIGURE 3.6: A schematic representation of the TIS and TOS categories in data and MC samples. The colors used in each category highlight the varying efficiency distributions within each category and among the samples.

with the Lo trigger, passing through the reconstruction chain and subsequent trigger criteria, and ultimately being stripped. This ensures that the simulation complies with all the selection criteria applied to real data.

3.3.2.1 MC CORRECTIONS

Despite ensuring that the simulation passes through all the selection criteria as the data do, the simulation process is not perfect and there may be differences between data and simulation. There are specific features of the detector that are very difficult to simulate due to their complexity, with many different effects occurring simultaneously. Therefore, it is crucial to identify where these mismatches occur and find ways to address these differences. The primary consequence of discrepancies between real data and simulation is the inaccurate estimation of efficiencies, which are typically studied using simulation.
Some discrepancies are well-known effects, and there are established methods to handle them. The most commonly used method to address differences between data and MC is through data-driven approaches, where a well-understood process is simulated and concurrently extracted from real data. Reference variables are selected that represent the main dependencies of the studied effects. For example, the η and momentum of each track are often chosen, as these variables accurately describe geometric and acceptance dependencies resulting from inaccuracies in simulating detector occupancy in specific regions. Lastly, a correction histogram can be constructed based on these variables. The correction factor in each bin of the histogram is the ratio of events observed in data to those in MC.

The corrections typically applied at LHCb and relevant for our analysis include:

- **PID correction:** The particle identification system is not accurately simulated at LHCb due to the complexity of these subdetectors. To properly estimate the impact of the PID selection on the signal MC sample, a data-driven approach is employed. Correction histograms are built using the *PIDCalib* package [78]. A reference channel is selected where PID variables are constructed on real data and MC for that channel, ensuring that the kinematic cuts match the requirements applied to the MC sample where the PID efficiencies will be computed. The correction histogram is obtained from the reference channel, using the η and momentum of the track, as well as the global occupancy of the detectors, as variables to compute the correction. When applying the correction histograms to the MC sample with no cuts on these PID variables, a per-event efficiency value can be obtained for the MC sample. Obtaining the real impact of the PID selection on the MC sample.
- **Tracking correction:** Tracking efficiencies differ between data and MC [79], necessitating corrections to align MC tracking efficiencies with those observed in data. Kinematic information from each track is sufficient for this correction; the corrected efficiency for a track can be obtained from the value of the momentum and η for that track. By combining the corrections for each signal track, a per-event tracking correction can be obtained.
- **Lo Hadron correction:** The hadronic calorimeter plays a critical role in the Lo trigger for hadronic lines. As there are five pions in the final state of our decay of interest, we must understand the behaviour of the signal tracks in the hadronic calorimeter to properly estimate the Lo Hadron TOS efficiencies. The Lo Hadron correction [80] is used to estimate the TOS efficiencies of our signal. This tool computes energy depositions from signal tracks and possible overlaps in calorimeter cells, providing a per-event estimation of Lo Hadron efficiency. Additionally, the proportions of TIS and TOS categories differ between data and MC. In this correction, the MC events in each category are scaled to match the proportions in data. A graphical representation of TIS and TOS categories for signal and MC is shown in Fig. 3.6.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 4

METHODOLOGY: DATA-DRIVEN MODELS ON THE PHASE SPACE

The methodology developed to perform the branching fraction analysis includes a wide variety of tools and methods that allows to extract the final observable of interest. However, in this chapter, we are presenting the main novelty in the development of this kind of analysis. Usually, in multi-body decays, to extract the efficiencies, one assumes that the signal comes from the dominant process, which does not cover all the physic processes happening in real data. To include all these processes in the efficiency estimation without modeling them, we decided to use a data-driven approach on the 5D PHSP of a four-body decay, extracting hence the unbiased signal efficiency. Estimating the efficiency is crucial to determine the number of event candidates that emerged from the $p-p$ collisions based on the measured event yield.

These data-driven methods allow us to avoid binning the 5D PHSP. Binning would be a significant limitation, as a huge amount of data would be needed to populate all the bins of the PHSP. Instead, we extract the desired distributions from data in a new sample. By weighting each event of this sample, we can transform it to resemble the input data. We replace the binned scheme with a new weighted sample where we can compute numerical integration to extract our parameters of interest.
The methods we studied are the Kernel Density Estimation (KDE) and the Gradient Boosted re-weighting (GBrw). Both methods can transform the reference sample by weighting it. However, each method calculates the target weights in a unique way. They will be presented in Sec. 4.2. In the following section, the multidimensional space where the corrections will be computed is presented.

4.1 THE 5D PHASE SPACE

The four-body PHSP constitutes the set of four-vectors that respect the PHSP density, i.e., that respect the energy-momentum relations between the decaying particle and its children particles. Remember that the norm of the four-vector represents the invariant mass of the particle.

The number of degrees of freedom in the PHSP of a four-body decay can be obtained from the following argument:

- There are 4 four-momentum vectors in the final state: $4 \times 4 = 16$ variables.
- The energy-momentum conservation must be imposed; the children particles come from the four-vector described by the parent particle. Hence, there are 4 degrees of freedom constrained, leaving $16 - 4 = 12$ free variables.
- The mass of each child particle in the final state is fixed, implying 4 new constraints, one for each children particle: $12 - 4 = 8$ independent variables.
- Finally, our system is invariant under global rotations, i.e., the parent particle is unpolarised as it comes from the $p-p$ collisions. Hence, there are 3 extra degrees of freedom that can be integrated out. These correspond to the 3 dimensions of space or the three Euler angles: $8-3 = 5$, thus obtaining the 5D PHSP of a four-body decay.

As we have just seen, there are 5 independent variables in a four-body decay when the parent particle is

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

unpolarised. The follow-up question is, can we describe the PHSP density as a function of the position in a 5D space built from the PHSP variables?

4.1.1 PHASE SPACE DENSITY

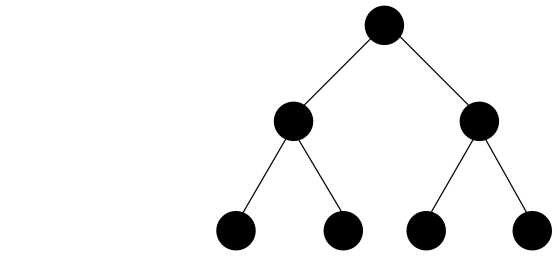
The PHSP density describes the set of four-vectors that respect energy-momentum conservation in a particle decay to n-bodies. Thus, it represents the potential states of these momenta and their associated probabilities, or the state density, for every PHSP point.

For a n number of decaying particles, the PHSP density is given by:

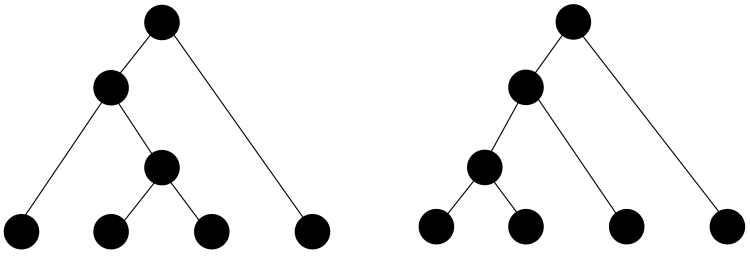
$$d\Phi_n = \delta(P - \sum_i^n p_i) \prod_i^n \frac{d^4 p_i}{(2\pi)^2} \delta(p_i^2 - m_i^2)\,. \tag{4.1}$$

where P refers to the parent's momentum and p_i to the children's momentum. The first delta-function sets the energy-momentum conservation between the parent particle and the children, and the second one in the sum reflects the fact that the energy of each child particle is described by its three-momentum and its fixed mass, adding an extra constraint to the PHSP density as mentioned earlier.

As we will see in the following section, in our modelling approach, some data-driven methods can benefit from the value of the PHSP density for each data event or PHSP position. Nevertheless, it is not easy to obtain the value of the integral of Eq. 4.1 in a general case. Although the integral can be solved numerically, we will address the computation of this integral in the quasi two-body approximation, as the analytical expression of the integral can improve the modelling using data-driven methods. In addition, we will study the viability of applying this approximation to study the axial resonances of the four-body decay.



(A) VV decay topology.



(B) Axial decay topologies.

FIGURE 4.1: Different topolgies for a four-body decay.

4.1.2 THE QUASI TWO-BODY APPROXIMATION

In this approximation, we consider that the four-body decay happens through sequential or nested two-body decays, with the topology shown in Fig. 4.1A. With this consideration, we can compute the integral of the three two-body decays independently. Each integral is computed in the reference frame of the decaying particle [81]. This process corresponds to a Vector-Vector (VV) decay, as in the $B^+ \to K^{*+}\rho^0$ decay, where each vector, the intermediate resonances, decays into one pair of final-state particles. In this way, Eq. 4.1 for a four-body decay can be simplified to:

$$d\Phi_{4,VV} = \frac{ds_{12}}{2\pi}\frac{ds_{34}}{2\pi}\, d\Phi_2(q_{12}, q_{34})\, d\Phi_2(\hat{p}_1, \hat{p}_2)\, d\Phi_2(\hat{p}_3, \hat{p}_4)\,, \tag{4.2}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

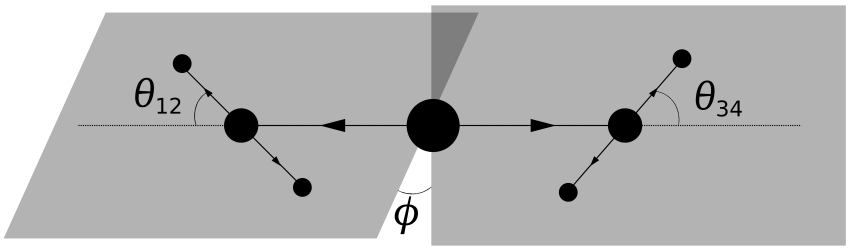


FIGURE 4.2: Helicity angles definition on the quasi two-body approximation of a four-body decay.

where $d\Phi_2$ are the two-body PHSP that factorise within this approximation, we labelled the final-state particles using numbers from 1 to 4. The hatted variables represent the momentum values in the reference frame where each vector parent is at rest. Note that the squared masses ($s = m^2$) of the intermediate states s_{12} and s_{34} are free, hence, this dependency is explicitly added to the expression. As the three two-body decays factorise, they can be computed independently from the two-body PHSP integral:

$$\int d\Phi_2 = \frac{q}{4\pi M} \int \frac{d\cos(\theta)}{2} \frac{d\phi}{2}\,, \tag{4.3}$$

where q is the children break-up momentum in the reference frame of the parent particle with mass M. Note that a two-body decay always happens within a plane. Thus, one of the angles can be integrated out, parameterising all the PHSP dependencies on a single angle; in the parent rest-frame, both children particles carry the same momentum value, q, and they fly in opposite directions.

From Eq. 4.2, it is clear that the masses of the intermediate vector particles are some of the PHSP density dependencies. To obtain the five variables that fully describe the PHSP density of a four-body decay, we need to find three more dependencies. These dependencies are the angles entering into each two-body PHSP in the quasi two-body decomposition, known as the helicity angles. Their schematic representation is shown in Fig. 4.2. The ϕ angle is the angle between the planes formed by the children particles of each vector decay in the parent particle rest frame, and the other angles are the angles between the children particles and the boost direction in each vector's rest frame.

Finally, to obtain an analytical expression for the PHSP density, we can introduce Eq. 4.3 into Eq. 4.2. Ignoring the constant terms as they add no information on the functional form of the distribution, we can write the analytical expression of the PHSP density within the quasi two-body approximation:

$$d\Phi_{4,VV} \sim \frac{q_{12} q_{34} q_B}{m_{12} m_{34}} ds_{12}\, ds_{34}\, d\cos\theta_{12}\, d\cos\theta_{34}\, d\phi\,, \tag{4.4}$$

In this expression, q_{12}, q_{34}, and q_B are the break-up momenta in the reference frame of each two-body decay. These three magnitudes depend on m_{12} and m_{34}. The 5D PHSP can be represented by the integration variables:

$$(s_{12}, s_{34}, \cos\theta_{12}, \cos\theta_{34}, \phi)\,, \tag{4.5}$$

These variables are widely used to study VV four-body decays.

4.1.3 THE AXIAL RESONANCES PROBLEM

In an axial (A) decay where the intermediate resonance decays to three bodies, there are two different possibilities for the decay of the intermediate resonance to the final state, as shown in Fig. 4.1B. The $B^+ \to K_1^0(1270)\pi^+$ decay is an example of axial topology. The $K_1^0(1270) \to K_S^0\pi^+\pi^-$ decay can happen through a $K^{*+} \to K_S^0\pi^+$ resonance or a $\rho^0 \to \pi^+\pi^-$ intermediate state. These topologies stop the two-body decomposition from factorising, making the derivation of a practical analytical solution impossible with a quasi two-body approximation.

For the full PHSP of our four-body decay, assuming all chains follow a quasi two-body approximation, we can still recover an analytical solution to the PHSP density and describe all the features using the independent variables outlined in Eq. 4.5.

The issue arises when the PHSP distribution is not considered as a whole, but just a specific region of the axial range. When cutting the PHSP in a certain region of any of the variables obtained from the quasi two-body parameterisation ($s_{12}, s_{34}, \cos\theta_{12}, \cos\theta_{34}, \phi$) or a combination of them, the impact of this

cut on the analytical expression in Eq. 4.4 can be easily recovered. However, When cutting the PHSP on a different projection, such as the three-body mass (m_{123}), the PHSP distribution on the other variables is modified due to the correlations between the PHSP variables and m_{123}. In this way, these modifications or distortions on the axial projections are more difficult to model by the analytical expression obtained from the quasi two-body approximation.

Moreover, to obtain an analytical expression, if possible, one would need significant dedication to derive it in a general way for the two possibilities of an axial cascade decay. The solution would need to develop a mass-dependent Dalitz plot approach where the parent mass in the Dalitz decay would be an integrating variable. The Dalitz plot is a widely used approach to parameterise the PHSP dependencies of a three-body decay in a 2D space.

To overcome this issue, we opted for a numerical strategy: a numerical integration technique is executed utilising a reference sample. This sample is simulated so that it follows the PHSP distribution within a desired mass window. In this way, this sample would carry all the information about the PHSP density distribution. Performing any integration in a specific region of the PHSP would be equivalent to performing a sum on the events of the specific region. If we then apply some weighting to that sample, the integral would be the sum of the weights of the sample. Finally, we will test the validity of the quasi two-body approximation in the charmless axial region of the PHSP to assess the size of the bias imposed by the approximation in our case.

4.2 DATA-DRIVEN MODELS

Nowadays, as the information era floods all aspects of our lives [82], the amount of stored data is increasing at an alarming rate. Unfolding data distributions, from which one can extract very useful information, is a key feature to understand certain patterns in data. Data-driven models are therefore crucial in this process, being the main tool to unfold these distributions with no prior knowledge of them.

In the context of our measurement, a branching fraction in a charmless mass window that includes vector-vector and axial resonances, these methods can be used to extract certain distributions that are key to controlling certain aspects of the measurement without assuming any model that could bias the result. As will be presented in Chapter 5, in this analysis we aim to parameterise the amplitudes in data that fall within our charmless mass window. As the final goal is to understand how data amplitudes would be distributed at the generator level, we need to account for the efficiency effects. These efficiency effects arise from the detection process, reconstruction, and data selection.

In this way, the data amplitudes and the efficiency distributions must be parameterised in a data-driven approach. The efficiency distribution will be parameterised using a flat PHSP simulation that undergoes all the generation, detection, and selection steps as real data does. Any deviation from the PHSP distribution in this sample can then be associated with the efficiency effect. Additionally, the data amplitudes will be extracted from real data. These distributions can be treated as Probability Density Functions (PDF or $\mathcal{P}(\Phi_4)$) within the PHSP. A more detailed explanation on PDFs and parameter estimation is found in Sec. 4.5.

Finally, by dividing the PDF extracted from real data by the PDF extracted from the flat PHSP simulation, the PDF describing the amplitudes at the generator level can be obtained.

Many different approaches can be considered to extract the PDF distribution from data. In this thesis, we use the following strategy: a flat PHSP sample is simulated within the desired region of the PHSP, i.e., the analysis mass window. We refer to this sample as the sample at generator level. The data-driven method can then transform this sample by adjusting each event's weight to match the reference or input distribution. This input distribution is derived from the sample that describes the desired feature to parameterise. In this way, the simulated sample that only describes the PHSP density (i.e., the sample at generator level or integration sample), acts as a template where we copy the PDFs of interest by modifying its weights. Note that these PDFs are non-parametric as there is no analytical expression to describe them. This approach is model-independent and only depends on the characteristics of the algorithm used to unfold the distributions.

In the following sections, we will present two different algorithms to perform this re-weighting process that allows us to obtain the desired PDFs from data.

4.2.1 KERNEL DENSITY ESTIMATION

The KDE algorithm extracts a PDF from data, it can be understood as a transformation of each data-point in a kernel function. By summing the all the kernel functions from each data-point, the data distribution can be unfolded. Thus, there is no need for a second sample; with this algorithm, one can obtain the PDF distribution directly from the data. Conceptually, however, two samples are involved despite one can be used twice by folding it. One sample is used to build the PDF and then the other is used to evaluate it, assigning a per-event weight based on the PDF values at each data point.
The computation involves constructing a matrix of distances between data points from both samples. These distances account for correlations in the multidimensional space, ensuring they are properly considered when computing the PDFs. The matrix is then summed row-wise after transforming each distance in a kernel value. This approach results in the final PDF, which aggregates contributions from each event in the sample.

The KDE algorithm can be formulated as:

$$w_{j,\mathrm{KDE}} = \sum_i K((\mathbf{x}_j - \mathbf{x}_i), H)\,, \tag{4.6}$$

where variables in bold denote vectors in an arbitrary number of dimensions. Here, H represents the bandwidth, determining the kernel (K) size used to build the PDF from each data point. H is linked to the covariance matrix of the space, preserving correlations within the data. Sec. 4.5 introduces the concept of space correlations in detail. Typically, a Gaussian kernel is chosen, where the bandwidth relates to the standard deviation σ:

$$K(\mathbf{x}, H) = \frac{\exp\left[-\frac{1}{2}\mathbf{x}^\top H^{-1}\mathbf{x}\right]}{(2\pi)^{-d/2}|H|^{1/2}}\,, \tag{4.7}$$

where H is the covariance matrix of the Gaussian kernel, representing its multidimensional standard deviation and d is the number of dimensions of the space.

In summary, the KDE algorithm has two main parameters: the kernel function and the bandwidth. The implementation used in this thesis is described in [83]. This implementation handles sharp edges, such as boundaries of a square distribution, by reflecting them across the space boundaries and then restricting the distribution to its original range. This feature is key to avoid having edge-effects on the parameterised distribution, extending its range beyond the physical limit or adding undesirable features to it.

Regarding our analysis strategy, it is crucial to note that this tool extracts the PDF directly from a data sample. Therefore, when using the flat PHSP sample at generator level to parameterise PDF distributions, we do not want to extract the PHSP density component as it is already considered in the sample. Thus, we must use the expression in Eq. 4.4 to suppress this contribution from data. For example, when obtaining the efficiency from the flat PHSP sample passing through the LHCb selection process, we must weight each event by

$$\frac{m_{12}m_{34}}{q_{12}q_{34}q_B}\,,$$

in order not to account twice for the PHSP density in the sample at generator level. As previously mentioned, this expression is not universally valid for all mass-window cuts in the PHSP and may present particular challenges when excluding specific regions in the projections of three-body combinations. Section 4.3 will explore this limitation further.

4.2.2 GRADIENT BOOSTED RE-WEIGHTING

This algorithm is widely used at the LHC, particularly at LHCb, to correct MC simulations to match real data distributions. Several variables can be corrected simultaneously by obtaining weights on the MC sample that transform it to match data distributions.

GBrw is a machine learning technique that minimises differences between two samples based on a χ^2 criterion across multiple projections. The algorithm employs a Boosted Decision Tree architecture. The key parameters of this algorithm are the tree depth, the number of estimators, and the learning rate, a factor that smooths the minimisation process to prevent over-training. The implementation used in this thesis is summarised in [84].
By controlling these parameters, one can tune the re-weighting process to meet specific data requirements. For instance, when both samples exhibit significant differences and complex distributions, such as those

seen in $\pi\pi$ projections, a larger number of estimators is needed to capture these distribution effectively. Conversely, for smoother distributions like efficiency variations across PHSP, fewer estimators suffice, mitigating over-training risks due to limited statistics.
Over-training can be identified by examining the distribution of output weights, as well as in the KDE algorithm. A prevalence of zero-weight events with a long tail towards high values suggests potential over-training, where the sample's statistical power concentrates heavily on a few outlier events.

Furthermore, as GBrw compares two different samples, any shared features between the samples, such as PHSP density distribution, are ignored. In this approach, there is no need of an analytical expression for the PHSP density, as it is already accounted for on the sample at generator level.

However, a notable drawback of GBrw is its failure to account for multidimensional space correlations during the minimisation process. The algorithm focuses on minimising differences across 1D projections, thereby potentially ignoring overall correlations within the space [84]. While this limitation could impact our approach, it is important to highlight that the PHSP correlations are inherently considered in the generator level PHSP sample. Nevertheless, the correlation features present in the samples to be copied may not be properly captured.
To address this, extending the PHSP to 6D within the GBrw algorithm could mitigate some correlation issues. However, potential redundancies in this expanded space would need further investigation.
Additionally, it is important to acknowledge that statistical limitations persist in these studies, where sample size may be a limitation to fully capture data features, whether in correlations or 1D projections.

4.3 PERFORMANCE EXAMPLE: EFFICIENCY PARAMETERISATION

We have presented two different algorithms for performing data-driven parameterisation in the 5D PHSP of a four-body decay. Now, we will compare their performance in the case of efficiency parameterisation. The efficiency distribution is obtained from simulation: the flat PHSP sample at generator level is propagated through the entire detector simulation process, as detailed in Sec. 3.3.2. Additionally, we simulate the data-taking process and the data selection steps to enhance the significance of the signal. Since the sample is generated flat in PHSP, any deviation from flatness indicates efficiency effects across PHSP.
In this analysis, the PHSP mass window includes both two-body (VV) and three-body (A) mass regions that do not contain c quarks, i.e., regions before the D-meson mass. Therefore, we do not analyse the entire PHSP but focus on specific regions of interest.
In this test, we will evaluate the efficiency parameterisation in two distinct scenarios: one for a VV-like region of the PHSP and the other for an A-like cut. Remembering that the decay under study is $B^+ \to K^0_S\pi^+\pi^-\pi^+$, these regions can be defined as follows:

- VV mass window: $m(K^0_S\pi^+) < m(D) \quad AND \quad m(\pi^+\pi^-) < m(D)$.
- A mass window: $m(K^0_S\pi^+\pi^-) < m(D)$.

Regarding the data-driven models, there are subtle differences between each case as already mentioned. Our ultimate goal is to extract the amplitudes at the generator level; thus, data amplitudes will be divided by the efficiency distribution. In order to prevent infinite values when dividing by the efficiency weights, they must be non-zero. Specific details of this strategy are provided in Sec. 5.4.5. Therefore, we select hyperparameters for each algorithm to achieve smooth distributions in PHSP and to ensure non-zero weights. The parameters used for testing each algorithm are as follows:

- KDE: A Gaussian kernel where the bandwidth was chosen using criteria for multivariate distributions derived by Wand and Jones [85]. We include a scaling factor for the kernel width adjustment (typically between 1 and 3, depending on data statistics) to prevent events having zero weight. Controlling the kernel size is crucial for smoothing the parameterisation process and avoiding over-fitting.
- GBrw: This algorithm does not have a straightforward method to select optimal hyperparameters, especially concerning how multidimensional distributions are handled. Extending the χ^2 criteria to a multidimensional space poses challenges. However, as we want the efficiency distribution to be smooth, we use a low value for the *max_depth* parameter, specifically $max_depth = 2$, while keeping default values for other parameters in our implementation. This test aims to evaluate algorithm performance in a basic scenario without further tuning. In practical applications, the number of estimators would also be fine-tuned.

It is important to note that finding optimal values for these parameters in multivariate distributions is not straightforward, as evaluations are limited to 1D projections. Combining criteria across different projections is non-trivial. Thus, a significant source of systematic uncertainty in our analysis strategy will

be associated to variations in these hyperparameters. These parameters can be analogous to the binning scheme in histogram-based strategies.

To conclude this section, the results of efficiency parameterisation for the $B^+ \to K^0_S\pi^+\pi^-\pi^+$ decay using both algorithms are presented. Each algorithm's performance is demonstrated across two mass windows. The set of variables used to capture the PHSP distributions, referred to as training variables, includes:

$$(m_{K^0_S\pi^+}, m_{\pi^-\pi^+}, \cos\theta_{K^0_S\pi^+}, \cos\theta_{\pi^-\pi^+}, \phi)\,,$$

where the masses of the two vectors are used instead of their squared masses. Thus, the PHSP density distribution needs adjustment according to the Jacobian of the transformation: $ds = 2mdm$. Removing constant terms, the PHSP density in this space can be described as:

$$d\Phi_4 \sim q_{K^0_S\pi^+} q_{\pi^-\pi^+} q_B \, dm_{K^0_S\pi^+} \, dm_{\pi^-\pi^+} \, d\cos\theta_{K^0_S\pi^+} \, d\cos\theta_{\pi^-\pi^+} \, d\phi\,.$$

To extract the efficiency distribution from the flat PHSP LHCb simulation using the KDE algorithm, this sample must be weighted by $1/q_{K^0_S\pi^+} q_{\pi^-\pi^+} q_B$ to isolate the efficiency distribution.

Performance comparison of both methods across different mass projections of the PHSP is shown in Fig. 4.3. In the A mass window, some mismatches in certain projections with the KDE algorithm are noticeable. Fig. 4.4 shows the performance of the variables used in the training space. Here, the wrong characterisation of the efficiency by the KDE algorithm in the axial mass window is also noticeable. This issue arises because efficiency effects cannot be separated from incorrect parameterisation of the PHSP distribution in this mass window. The plot labelled as *PHSP* shows the sample at generator level, demonstrating that in the VV mass window, the sample forms a square block. Any deviation from this square block in the 5D space can be attributed to efficiency effects.
Lastly, Fig. 4.5 displays some 2D projections of the training space on the sample at generator level. In the VV mass window, the space exhibits square boundaries not only in 1D projections but also in multiple dimensions. This square boundary feature is crucial for the KDE algorithm to function correctly, as it utilises boundary reflections with respect to the edge of each dimension. This feature is needed to reflect the multidimensional distribution accurately. In contrast, this characteristic is not observed in the A block. Adding an additional complication to the efficiency parameterisation by the KDE in the A block.

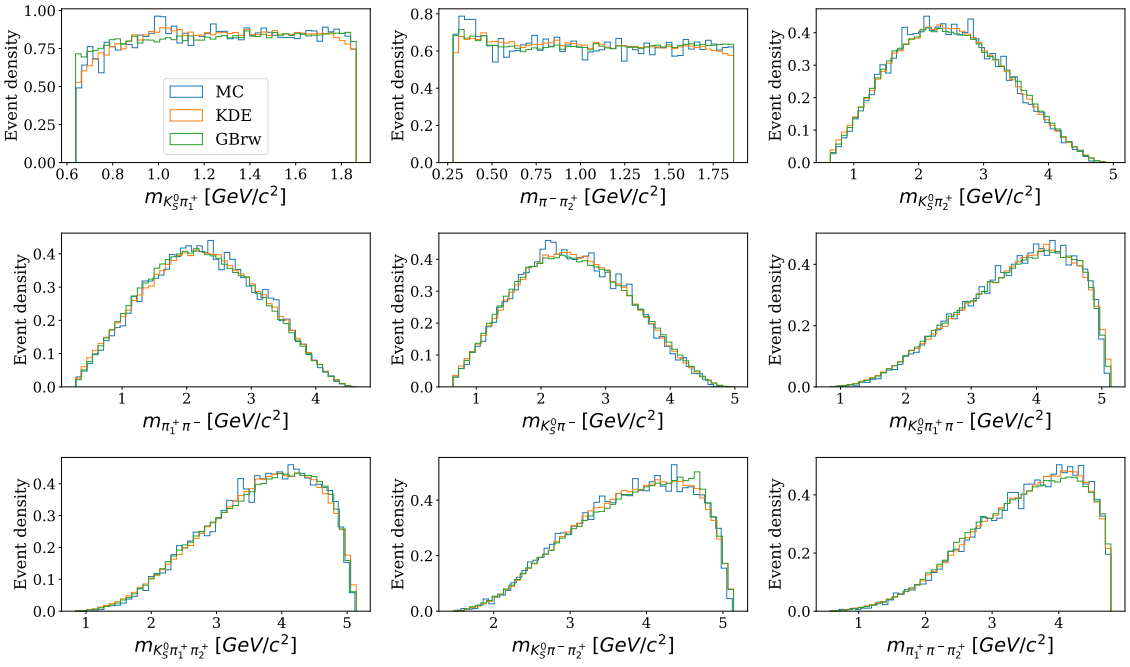
To sum up, using a very narrow kernel could yield plots that describe the efficiency distribution in the axial region as well. However, this would be an example of over-fitting, as most weights in the sample would tend towards zero, capturing noise related to limited statistics rather than the underlying distribution.

4.4 DISCUSSION

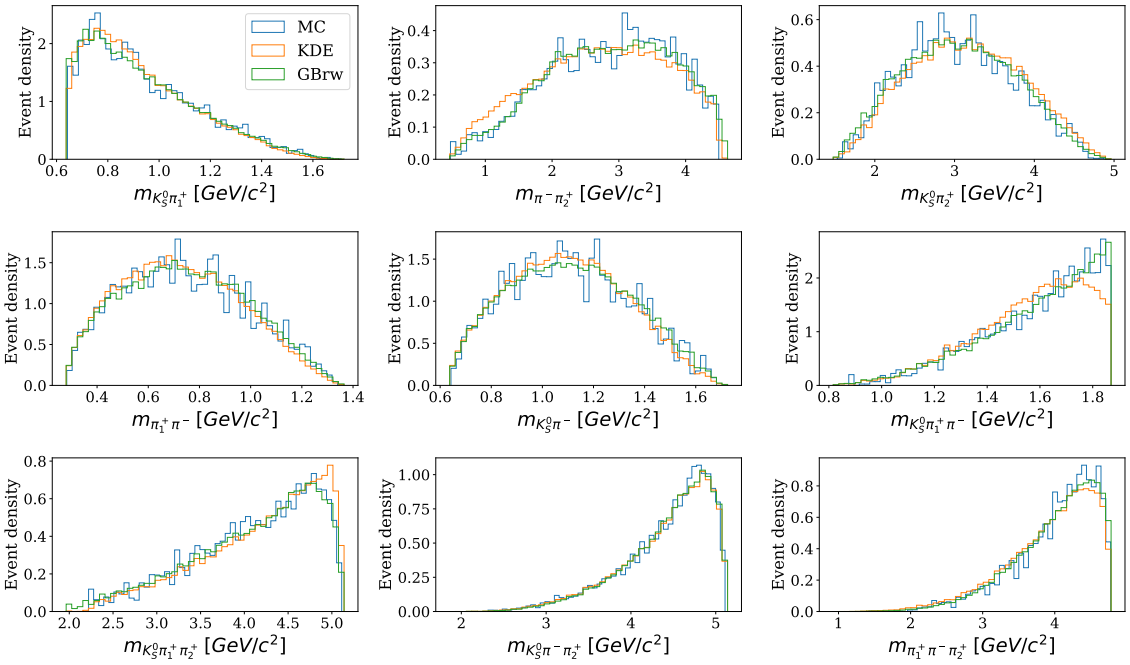
We have seen how these two algorithms work and how they perform on the VV and the axial mass windows. The main difference between these two algorithms is that, while the GBrw compares the two samples and transforms each event in the output sample to look like the input one, the KDE does not use the information from the two samples. Instead, it uses the input sample and one event in the output sample where the PDF ($\mathcal{P}(\Phi_4)$ is evaluated, transforming a flat sample within the space into a sample that follows the same PDF as the input one.
This is a crucial difference for our analysis, because with the KDE algorithm we need to undo the PHSP density distribution, as we are not interested in copying it, but only in the efficiency effect in this case. As we do not know how to integrate the PHSP density in an axial decay topology, the KDE does not perform well when some portion of the A projection of the PHSP is removed.

This issue with the KDE algorithm could be solved in two different ways: one is to use a flat 5D sample (a sample generating random events in five dimensions, or a 5D cube) as the integration sample that is transformed. In this way, there would be no need to un-weight the PHSP distribution from the input sample. Another option is to obtain the integration values with the GBrw algorithm, computing the PHSP integral by transforming the flat PHSP sample to match the flat 5D sample. These strategies may have different problems, such as not respecting the PHSP correlation in the integral performed with the GBrw, or having data points with no physical meaning in the square 5D block, which acts as the generator-level sample.

As data statistics in our case are indeed a limitation, we decided to implement our analysis strategy using the GBrw algorithm. Despite some correlations not being properly captured, the bias this could imply



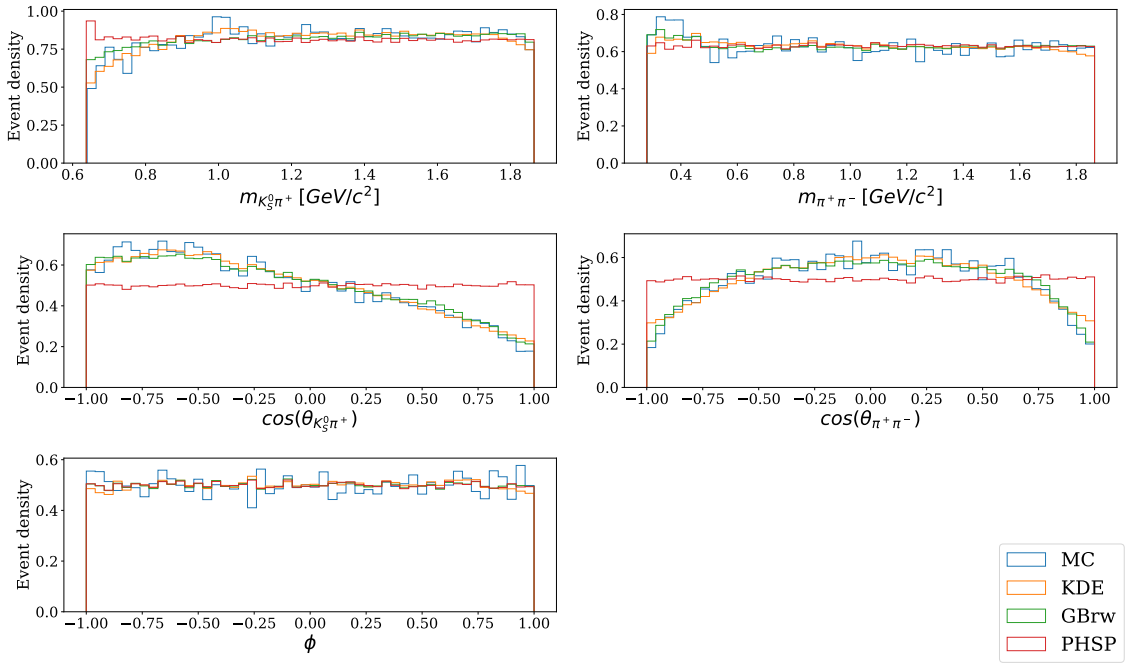
(A) VV mass window.



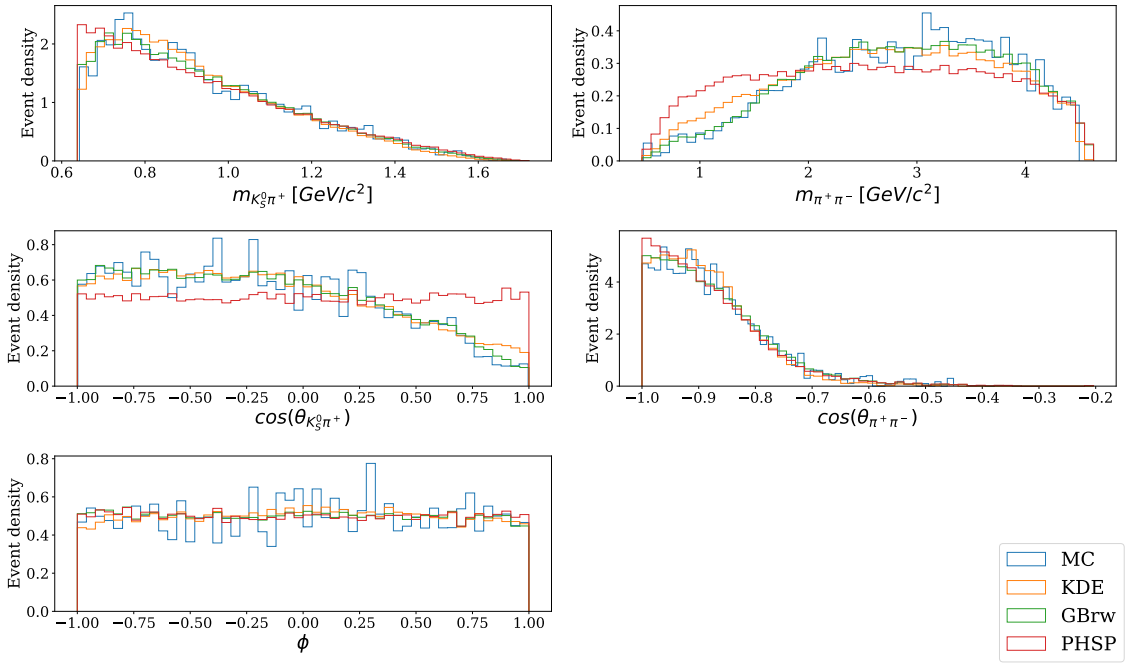
(B) A mass window.

FIGURE 4.3: PHSP two-body and three-body mass projections. The performance of the two data-driven algorithms on the efficiency parameterisation is shown. *MC* label refers to the LHCb simulated sample, *KDE* and *GBrw* labels represent the output of each algorithm.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



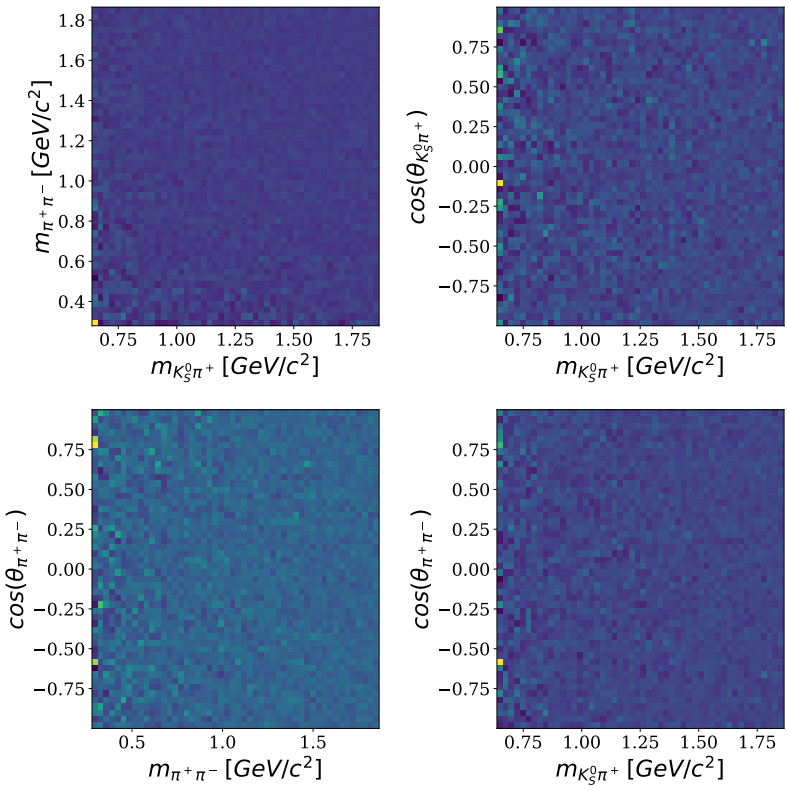
(A) VV mass window.



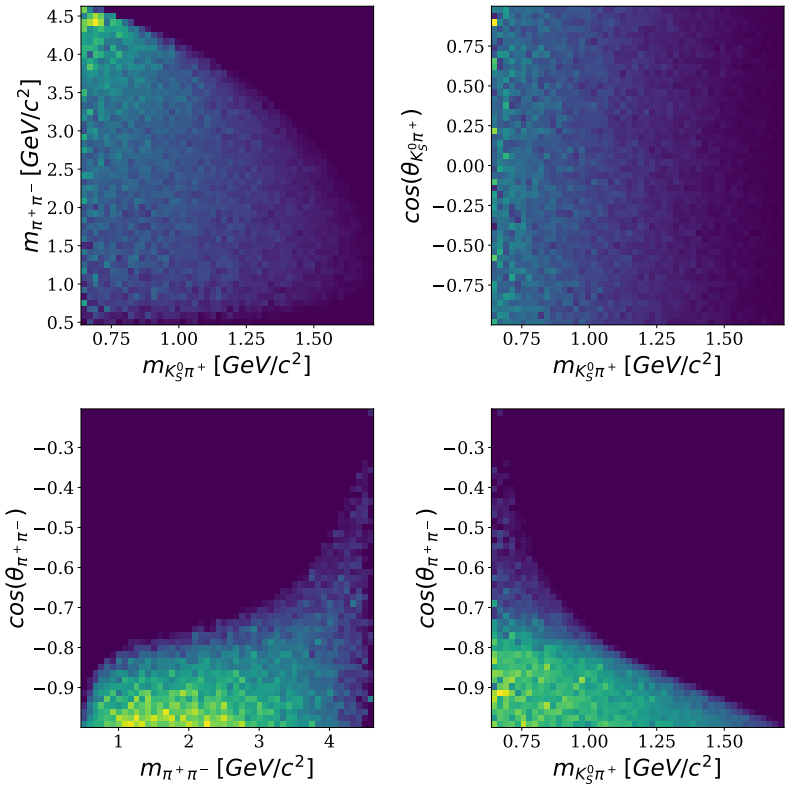
(B) A mass window.

FIGURE 4.4: The PHSP training projections of the data-driven algorithms are shown. *MC* label refers to the LHCb simulated sample, *KDE* and *GBrw* labels represent the output of each algorithm. Additionally, the flat PHSP sample at generator level weighted by the inverse of the PHSP density, is also shown; it is labelled as *PHSP*.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) VV mass window.



(B) A mass window.

FIGURE 4.5: Some 2D PHSP projections of the training variables on the sample at generator level weighted by the inverse of the PHSP density.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

would always be smaller than the bias imposed by assuming a naive model to represent our amplitudes in data.
Lastly, it is important to remark that, if the whole PHSP is analysed, the quasi two-body approximation would be valid for the whole space and the KDE algorithm could be used, being more consistent with the correlations of data in the PHSP with this procedure. The downside of this approach is that the amplitude analysis in the same data would be very impractical, as the number of amplitudes to model would be enormous, having thus a large number of parameters to minimise with all the computing overhead that it implies.

In conclusion, we decided that for our analysis case, using the GBrw algorithm to re-weight the PHSP sample at generator level is the best way to proceed. In addition, the KDE algorithm could be used to test the analysis strategy with a different algorithm, being a useful consistency check for the analysis. However, one would need to solve the difficulties presented.

4.5 COMPLEMENTARY STATISTICAL AND COMPUTING TOOLS

We have presented so far the methodology that allows us to model the desired PHSP distributions, which is key to developing the main novel contribution of our analysis. However, many different tools are used in a particle physics analysis. Nowadays, one of the main novelties is the inclusion of Graphical Processing Units (GPUs) in the computations. Thanks to their ability to process thousands of computations in parallel at the same time, they are a very useful tool for handling the millions of data events that a particle physics analysis needs to process. In this thesis, many of the computations are accelerated using GPUs or parallel computing in different CPU threads.
Conversely, there are also many tools used since the very beginning of this field, such as the statistical tools for fitting parameters to extract the PDF that data follows. We are presenting here the very basics of data statistics to define a Maximum Likelihood fit, which is used many times in the Branching Fraction analysis and is considered to be the optimal way to extract the PDF parameters from data. The aforementioned data-driven methods can also be considered fitting methods; however, they are non-parametric estimations of the underlying PDF in data. Nowadays, there are many different ways to extract parameters or distributions from data. As an example, in Sec. 6.3, a fitting method in a 2D space with a binned strategy and using an entropy measure estimator is presented. Below we present a set of key concepts to properly introduce the idea of Maximum Likelihood Estimation and some related concepts widely used in particle physics analysis.

PDF A Probability Density Function ($\mathcal{P}$) is a function that describes the probability distribution of a continuous random variable $\mathbf{X}$ in an arbitrary number of dimensions. It must be defined as positive in all its range and it must be normalised (integral equals 1). Its distribution can also depend on some parameters μ_i. It can be expressed as $\mathcal{P}(\mathbf{X}|\mu_i)$.

LIKELIHOOD FUNCTION A Likelihood Function represents, given specific observed data, how likely it is to observe the given data under different parameter values of the PDF. It is a function of the parameters of a statistical model. Given a set of n equally distributed data points $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, the Likelihood as a function of the μ_i parameters can be written as: $L(\mu_i) = \prod_j^n \mathcal{P}(\mathbf{X}_j|\mu_i)$.

MAXIMUM LIKELIHOOD ESTIMATION The purpose of the Maximum Likelihood Estimation is to find the PDF parameters that best describe the observed data. In this way, the maximum value for the Likelihood is the goal. Usually, in order to obtain a more stable value for the likelihood, we take the logarithm, multiplying it by -2 to obtain $-2\log L(\mu_i)$, this being the estimator to minimise. To find the minimum, the partial derivatives with respect to the μ_i parameters are taken. To perform the computation we use the *Minuit* [86] algorithm embedded in the *ipanema* [87] software.

COVARIANCE MATRIX AND CORRELATIONS When several fitting parameters are considered at the same time, how one parameter varies with respect to other parameters may have an impact on the fit result. The covariance matrix is a square and symmetric matrix $N \times N$, where N is the number of μ_i parameters, with the diagonal terms representing the uncertainty associated with the fitting parameters and the off-diagonal terms indicating how the estimates of different parameters vary together. Totally correlated parameters will vary together so that when one of them is varied, the other must vary by the same quantity to respect

the distribution of the whole multidimensional space. Thus, the covariance matrix informs about the uncertainties of the fitting parameters and their correlation.

SPLOT TECHNIQUE The *sPlot* [88] or the sWeights, are obtained from a procedure that transforms the fitting data so that it follows one of the PDF components. It makes use of the covariance matrix obtained in the fit. If our PDF is defined by a signal and a background component, this technique finds the set of weights (sWeights) that transforms the data to follow only the signal PDF component. This is a technique widely used in LHCb analysis to correct the MC distributions to match the data ones, being able to extract the signal PDF from the data, removing any remaining background contribution.
In our branching fraction analysis, we will use this technique to extract the signal component from the data so that we can extract the signal amplitudes using a data-driven approach on the PHSP.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CHAPTER 5

CHARMLESS BRANCHING FRACTION MEASUREMENT OF THE $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ DECAY

In this chapter, an inclusive charmless branching fraction measurement of the $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ decay is presented. The full LHCb available data is utilised, corresponding to both Run 1 and Run 2 data-taking periods. The analysed mass window includes any vector-vector (VV) or axial (A) resonance in the charmless region of the PHSP, i.e., before the D-meson mass. As there are two identical pions (π^+) in the final state, one must take into account the two possible Bose combinations when building the two and three-body invariant masses involving the π^+, explicitly handling the pion exchange. In order to reduce the systematic uncertainties, the branching fraction is measured relative to a control channel, which is $B^+ \to D^0(K^0_S \pi^+ \pi^-)\pi^+$. Both channels have the same final state, the difference between them is the different mass window. The control channel is chosen to be the main charm contribution to the PHSP. In this way, it is not necessary to estimate the number of parent particles produced from the $p-p$ collisions, as this number cancels out when computing the branching fraction ratio.

This branching fraction measurement is the preamble to an amplitude analysis (AmAn) to the same decay channel and in the same mass window, where one can constrain the penguin-induced $\Delta\alpha$ that pollutes the effective CKM angle $\alpha + \Delta\alpha$ measurement of $B^0 \to a_1^\pm \pi^\mp$ obtained from a $B^0 \to \pi^+\pi^-\pi^+\pi^-$ time-dependent flavour-tagged AmAn, through an $SU(3)$ symmetry [43, 45]. The BaBar Collaboration has already performed a similar analysis [44]. It is a LHCb-friendly approach that does not involve neutral particles with the added possibility to directly constrain $SU(3)$-breaking effects. The goal of this AmAn would be to measure the physics properties of the $SU(3)$ partner channels, $K^0_S a_1^+$ and $K^0_{1A}\pi^+$, including their branching fractions and strong phase difference, δ_{strong}. As the axial contributions do not overlap on the PHSP, the VV amplitudes are also included to act as a binding to perform the AmAn allowing to extract δ_{strong}. K_{1A} is a flavour eigenstate and can be written as a mixture of the $K_1(1270)$ and $K_1(1400)$ mass eigenstates, so the involved weak-phase must be taken into account and extracted from the AmAn or as an input from another analysis [89]. The available results on the CKM angle α from $B^0 \to \rho^0\rho^0$ and $B^0 \to a_1^\pm \pi^\mp$ decays comes from BaBar [33, 41], Belle [35, 42] and LHCb [37] experiments.

The analysis is performed in 3 main steps: selection, mass fits and efficiencies calculation. From the selection, the significance of the signal is enhanced, then in the mass fit, the yield of the signal component can be extracted through a maximum likelihood estimation fit. Finally, on the efficiencies, one can estimate the impact of the detection and selection processes on the signal yield. The selection is performed in 2 steps, first a soft PID selection and then a BDT on topological variables is utilised to reduce the combinatorial background. The mass fits for the signal channel are also performed in 2 steps, in this way the correlation between the PHSP variables and the B-mass are reduced by focusing on a narrow region when obtaining the sWeights with the sPlot technique [88]. Finally, in the efficiencies calculation a novel approach is considered: as there is no well-known amplitude model to simulate on the signal channel mass window, a data-driven method is used to extract the signal amplitudes of the decay. In this way, the MC sample for the signal channel is simulated flat in the PHSP in order to capture the efficiency variation across it as

the amplitudes are taken from the data sWeights. This novel approach helps in reducing the dominant systematic uncertainty related to analyses where the PHSP amplitudes are not known. Typically, a naive model that considers only the dominant contribution to the PHSP is used to extract efficiencies. With this approach, we can reduce the systematic uncertainty associated with the model used in efficiency calculations.
In the case of the control channel, as its amplitudes are well known, the standard procedure to obtain the efficiencies from MC is followed, and all the signal amplitudes presented in data are captured in the MC sample. This procedure implies ensuring that the distributions used to enhance the significance of the signal, i.e. the variables used to apply the selection requirements which impact the efficiencies, are equally distributed in data and in MC.
Furthermore, as there are many backgrounds that fall in the fitting region, most of the possible backgrounds are simulated in order to constraint their line-shapes in the mass fits, the samples are passed through the same selection as data and the line-shapes are fixed on the fits to data from the fit results to these MC samples.

Finally, the branching fraction is computed as follows:

$$BF_{signal} = BF_{\text{control}} \frac{f_{\text{signal}}}{\epsilon_{\text{signal}}} \frac{\epsilon_{\text{control}}}{f_{\text{control}}},$$

- BF is the Branching Fraction (from PDG [49] for the control channel),
- f is the signal yield for each channel obtained from the mass fits,
- ϵ is the efficiency. If taken from MC, the efficiency encompasses each simulation step such as the generation, which accounts for the detector acceptance, the efficiency of the reconstruction algorithm for those tracks that fall within the detector acceptance, and the efficiency of each trigger, stripping, and selection requirement: $\epsilon = \epsilon_{\text{gen.}} \cdot \epsilon_{\text{reco.}} \cdot \epsilon_{\text{trigger}} \cdot \epsilon_{\text{strip.}} \cdot \epsilon_{\text{selection}}$. Note that the efficiency of the signal channel is not simply taken from MC.

5.1 DATA SETS

The full Run 1 and Run 2 data are analysed on the branching fraction measurement. In addition, several MC samples are simulated to count for all the contributions that fall in the mass window. Then, these MC samples are produced with the statistic scaled based on the luminosity and efficiency for each year, as shown in Table 5.1. Moreover, the MC samples are stripping filtered due to the low efficiency in order not to waste the disk spaces. The different simulated decays are shown in Table 5.2. The efficiency of each sample is computed by the efficiency at generator level times the retention rate of the stripping filtering script. Accounting, in this way, for different data selection steps that impact the efficiency on the simulation, such as: the generator level efficiency, which considers the LHCb geometrical acceptance; the efficiency of the reconstruction algorithms on the channels of study; and, finally, also considering the efficiency on the trigger and stripping requirements: $\epsilon_{MC} = \epsilon_{gen.} \cdot$ Ret. rate.

The data sets are produced with the *AnalysisProductions* tool that runs the *DaVinci* application on the different samples. We use the same scripts for the real-data and MC samples production. The only differences between the real and simulated samples are the addition of some *TRUE* variables in the MC samples and the use of different algorithms for the momentum scale calibration. Furthermore, the input stream for the real-data sample is *BHADRON.MDST* whether the MC is stored on full *DST*. The PHSP variables are constrained using *DecayTreeFitter* constraining the value of the parent and child masses. Here, the mass of the K^0_S candidate, which decays to two pions, is constrained matching the PDG [49] value. This constraint ensures that the four-body PHSP density is respected. Then, the reconstruction is re-computed obtaining the constrained children momentum.

It is also important to remark that, due to the presence of a K^0_S in the final-state, we have 2 different reconstruction possibilities for it whether the children pions deploy some hits in the *VELO* or not. If so, the K^0_S candidate is reconstructed as *long* (LL), if not, it is reconstructed as *downstream* particle (DD). The main differences between these 2 categories are the resolution and the efficiency. Thus, they will be treated as different samples during the whole analysis. Finally, the samples are also split by groups of two-years categories as the efficiency correction may be different for each of these samples: 2011 and 2012 (1112) corresponding to Run 1, 2015 and 2016 (1516), and, 2017 and 2018 (1718). We split the Run 2 data-taking period due to the change on the threshold of the hadronic L0 trigger in 2017 [66]. Obtaining then 6 different samples for each channel. This division is implemented as such given the expected similar efficiency across

Year	CoM Energy $[TeV]$	Luminosity $[fb^{-1}]$	Stripping version
2011	7	1,11	21r1p1
2012	8	2,08	21r0p1
2015	13	0,33	24r1
2016	13	1,67	28r1
2017	13	1,71	29r2
2018	13	2,19	34

TABLE 5.1: Luminosity and Stripping versions used on the simulation of each data sample.

Decay	Eventtype	Model	Events
$B^+ \to K^0_S\pi^+\pi^-\pi^+$	12105151	PHSP mass window	2M
$B^+ \to \bar{D}^0(K^0_S\,\pi^+\pi^-)\pi^+$	12165121	Dalitz model for D^0	4M
$B^+ \to K^0_S K^\pm\pi^+\pi^\mp$	12105162(3)	PHSP	13M
$B^+ \to (D^0 \to K^0_S\pi^+\pi^-\pi^0)\pi^+$	12165529	PHSP	2M
$B^+ \to K^{*+}(K^0_S\pi^+)\eta'(\pi^-\pi^+\gamma)$	12105315	PHSP	9M
$B^0 \to (D^{*-} \to (D^0 \to K^0_S\pi^+\pi^-)\pi^-)\pi^+$	11166144	PHSP	5M
$B^0_s \to (D^-_s \to K^0_S\pi^+\pi^-\pi^-)\pi^+$	13166172	PHSP	3M

TABLE 5.2: MC samples used in the analysis. The number of events is reported at the stripping level.

these combined samples. Furthermore, they are not further split because we are aggregating the statistics to enhance the accuracy on the efficiency and amplitudes estimations with the data-driven techniques. Ultimately, these samples will be combined on a χ^2 to report the final branching fraction result.

5.2 SELECTION

There is an online selection by the trigger, a central selection by the stripping that selects the data to store on disk, and finally, an offline selection which purpose is to enhance the significance of the signal of study.

5.2.1 TRIGGER

There are 3 main steps in the trigger system from the point of view of fully hadronic decays: L0, HLT1 and HLT2. The L0 is the first step and it is hardware-based, as the signal decays are purely hadronic the only L0 relevant system for the signal is the hadronic one: it selects tracks with high energy deposited in the hadronic calorimeter. The next step is the high level trigger with two different steps which are software based. In the first level a single track or two-track combinations are selected; in the second one, more combinations of three or four tracks are considered.
In one $p-p$ collision event there can be several signals for different physics-cases, an event can be stored due to the signal of study (Triggered On Signal, TOS) or due to another track independent from the signal of study (Triggered Independent of Signal, TIS). In this analysis only Global L0 TIS or Hadron L0 TOS events are analysed.

The stripping line has some trigger requirements, these are:

Run 1:

- Hlt1Filter: *Hlt1TrackAllL0Decision*
- Hlt2Filter: *Hlt2Topo[234]Body.*Decision*

Run 2:

- Hlt1Filter: *Hlt1(Two)?TrackMVADecision*
- Hlt2Filter: *Hlt2Topo[234]BodyDecision*

These trigger requirements are related to the topological B lines and are specialised in selecting partially reconstructed B decays.

5.2.2 STRIPPING

In the stripping selection, the signal candidates are built and stored to be read in order to define the physical variables that will be used for the analysis.

USC UNIVERSIDADE DE SANTIAGO

Selection Category	Variable	Cut
K^0_S input tracks	FDχ^2	> 5
$\pi^\pm$ input tracks	IPχ^2	> 4
	P_T $[MeV/c]$	> 500
	P $[MeV/c]$	> 1500
	PIDπ	> 0
	PIDK	< 0
	Ghost. Prob.	< 0.4
$B^\pm$ candidate	AP_T $[MeV/c]$	> 1500
	$B^\pm P_T$ $[MeV/c]$	>1500
	ADOCA-χ^2(i,j)	< 10
	Vtx χ^2/n_{dof}	< 12
	$B^\pm$ DIRA	> 0.99995
	$B^\pm\tau[ps]$	> 0.1

TABLE 5.3: Set of stripping requirements applied in *B2KSHHH_PiPiPi_(LL,DD)Line*. The variables used in the stripping are defined in Sec. 3.1.1. When the χ^2 symbol is added, means that the cut is based on the performance of the fit to extract the tracks during the reconstruction algorithm. AP_T variable is the scalar sum of the p_T of each final-state track. Finally, τ represents the B-candidate decay time.

As there are 2 different possibilities to build the K_S candidate at LHCb, 2 different stripping lines are considered: *B2KSHHH_PiPiPi_DDLine* and *B2KSHHH_PiPiPi_LLLine*.
The K^0_S particles are selected from the *StdLooseKs(LL,DD)* particle container and the $\pi^\pm$ from *StdNoPIDsPions*. These containers are loose selections of the final-state tracks that will be used to build the signal candidates by applying the stripping requirements shown in Table 5.3. Note that both, signal and control channels use the same stripping lines. This is very useful in reducing the sources of systematic uncertainties.

5.2.3 FINAL SELECTION AND BACKGROUND SUPPRESSION

The main goal of this step is to suppress the combinatorial background maximising the significance of the signal. Also, as we are not interested in the full PHSP of the final-state, we perform a mass window selection to keep the regions of the PHSP that we want to include in the branching fraction measurement. Finally, the remaining charm contributions to the signal channel are removed.

5.2.3.1 MASS WINDOW

The mass window includes any Axial combination that fall before the D^0-meson mass and the $K^{*+}\rho^0$ VV decay with both Bose combinations on the exchange of the π^+ particles. Labelling the 2 π^+ as $B^+ \to K^0_S\,\pi^+_1\pi^-\pi^+_2$ the mass window can be written in the following way:

$$\begin{aligned} m(K^0_S\pi^+_1) < 1864.83 \quad AND \quad m(\pi^-\pi^+_2) < 1864.83 \quad OR \\ m(K^0_S\pi^+_2) < 1864.83 \quad AND \quad m(\pi^+_1\pi^-) < 1864.83 \quad OR \\ m(K^0_S\pi^+_1\pi^-) < 1869.65 \quad OR \quad m(K^0_S\pi^+_1\pi^+_2) < 1869.65 \quad OR \\ m(K^0_S\pi^-\pi^+_2) < 1869.65 \quad OR \quad m(\pi^+_1\pi^-\pi^+_2) < 1869.65\,, \end{aligned} \tag{5.1}$$

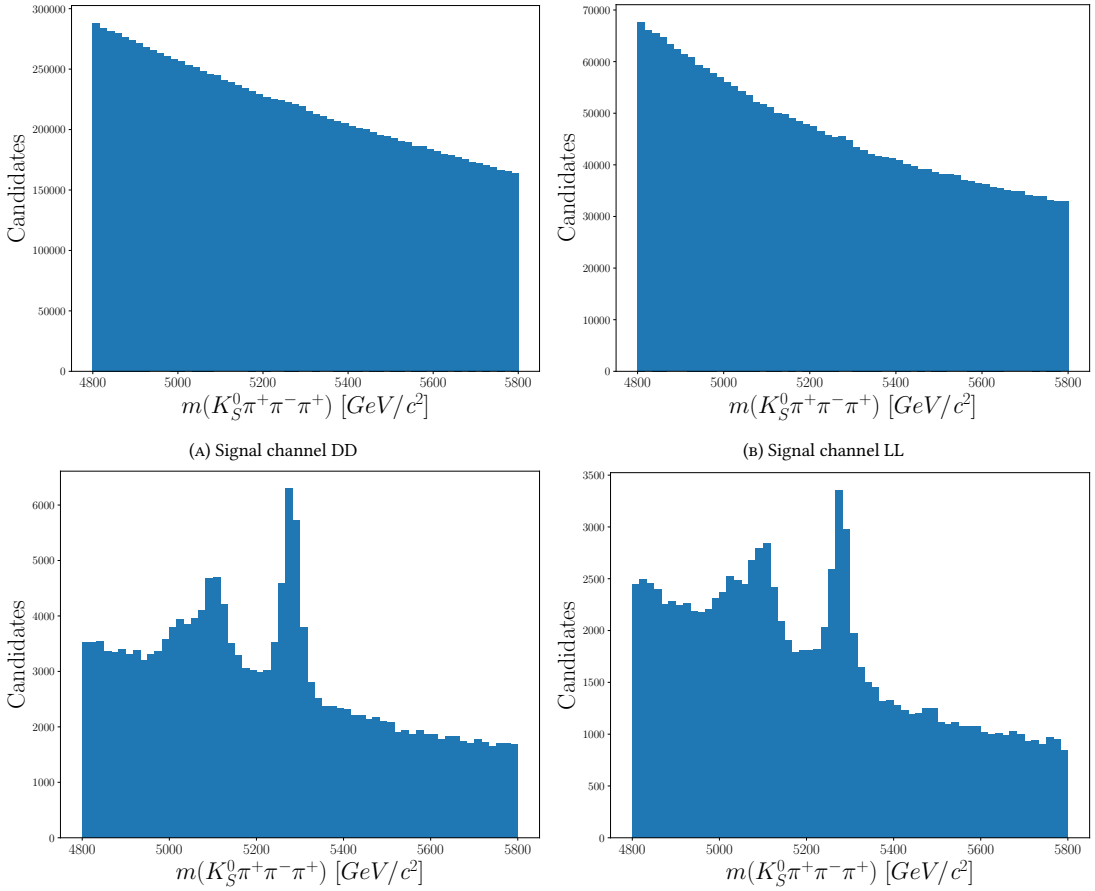
where masses are in MeV/c^2. This is a general mass window that allows us to perform the amplitude analysis to extract the phase difference between the $K^0_1 \to K^0_S\pi^+\pi^-$ and the $a^+_1 \to \pi^+\pi^-\pi^+$ decays. Fig. 2.7 shows how the VV contribution, $K^{*+}\rho^0$, connects both axial decays in the PHSP allowing to extract the phase difference between them.

As the control channel is the main charm contribution to the PHSP: $B^+ \to D^0(K^0_S\pi^+\pi^-)\pi^+$, we select the control channel mass window as:

$$(1820 < m(K^0_S\pi^+_1\pi^-) < 1900) \quad OR \quad (1820 < m(K^0_S\pi^-\pi^+_2) < 1900) \quad , \tag{5.2}$$

where masses are also in MeV/c^2.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) Signal channel DD

(B) Signal channel LL

(C) Control channel DD

(D) Control channel LL

FIGURE 5.1: Real-data B-mass projection after the pre-selection cuts.

5.2.3.2 PRE-SELECTION AND CHARM VETOES

The main contribution to the pre-selection is a soft PID requirement that can be summarised in the following way:

- $ProbNNpi > 0.001$, for the 5 pions on the final state.
- $\sum_i^5 is_muon < 2$, vetoing any di-muon on the PHSP.

The B-mass projection after these cuts is shown in Fig. 5.1. Furthermore, the *Decay Tree Fitter* status must be 0, which means that the fit with the constrained masses of the parent particle and the 4 particles of the final-state of study must have converged. Finally, all the events must be or *L0_Hadron_TOS* or *L0_Global_TIS*, which are the 2 cases analysed in the efficiency estimation:

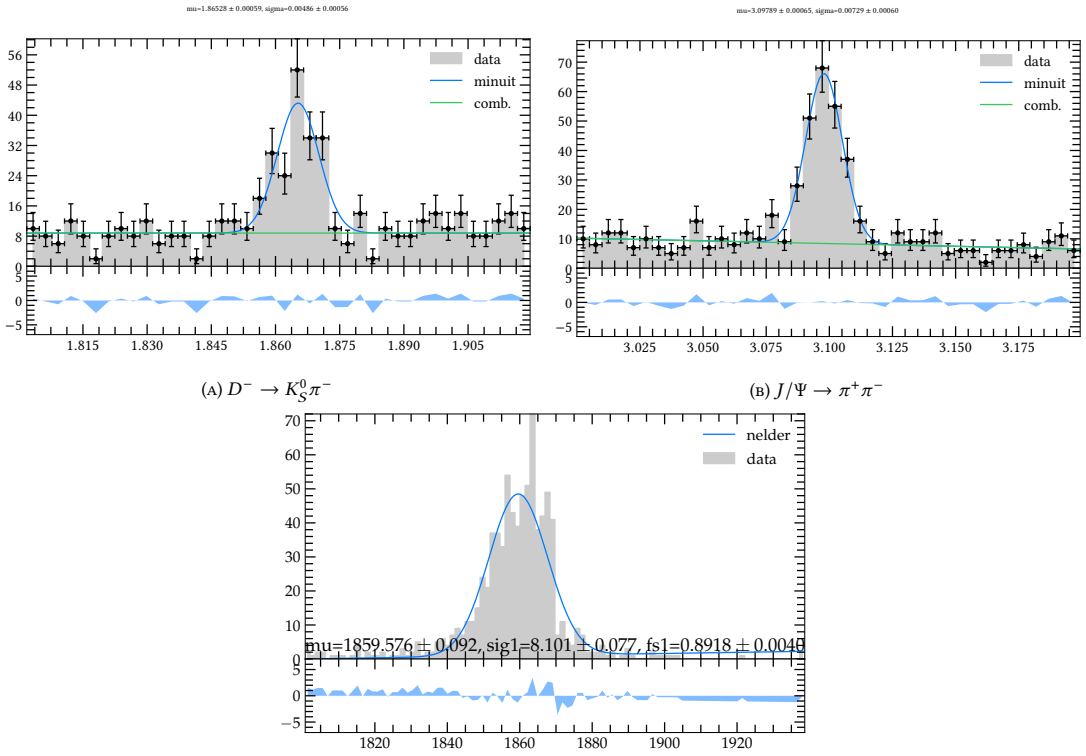
- $L0_Hadron_TOS = 1 \quad OR \quad L0_Global_TIS = 1$.

For each charm contribution that remains in the PHSP reflections after the signal mass window is applied, a fit is performed and a veto is set to $\mu \pm 3\Gamma$ from the fit result. The fitting model for these contributions is a double Gaussian with mean (μ) and 2 different parameters for the 2 widths, then the width of the line-shape is $\Gamma = f_s\Gamma_1 + (1 - f_s)\Gamma_2$. In addition, a 2^{nd}-order Chebyshev polynomial is added to account for the combinatorial background. The vetoed decays are $J/\Psi \rightarrow \pi^+\pi^-$, $D^- \rightarrow K^0_S\pi^-$ and $D^0 \rightarrow K^0_S\pi^+\pi^-$ which is the control channel. The fits are shown in Fig. 5.2. There are no vetoes needed for the control channel due to the tight mass window.

Finally, in Fig. 5.3 the PHSP mass projections of the signal channel are shown on a toy flat PHSP sample. In this plot, the signal channel mass window can be appreciated.

5.2.3.3 BDT SELECTION

After the soft PID cuts, a BDT selection on topological variables is performed. A MVA Boosted Decission Tree is trained with truth-matched signal MC (simulated flat on the PHSP) as the signal category and



(A) $D^- \to K^0_S\pi^-$

(B) $J/\Psi \to \pi^+\pi^-$

(C) $D^0 \to K^0_S\pi^+\pi^-$

FIGURE 5.2: Charm vetoes applied to the signal channel: The $\mu \pm 3\Gamma$ range from each fit is vetoed.

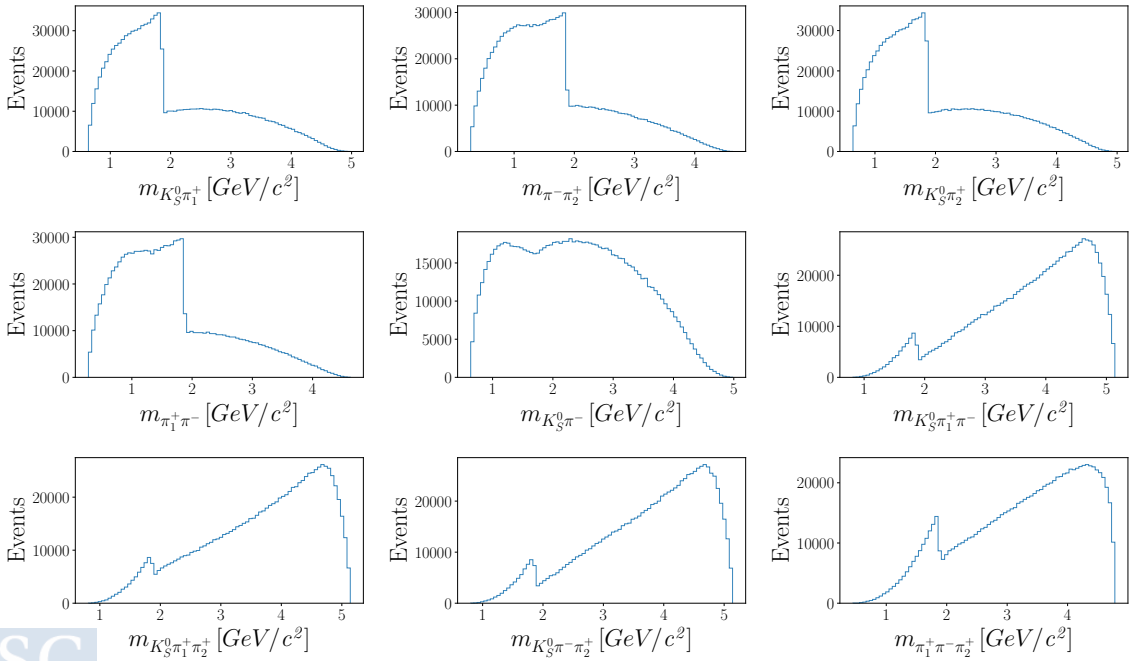
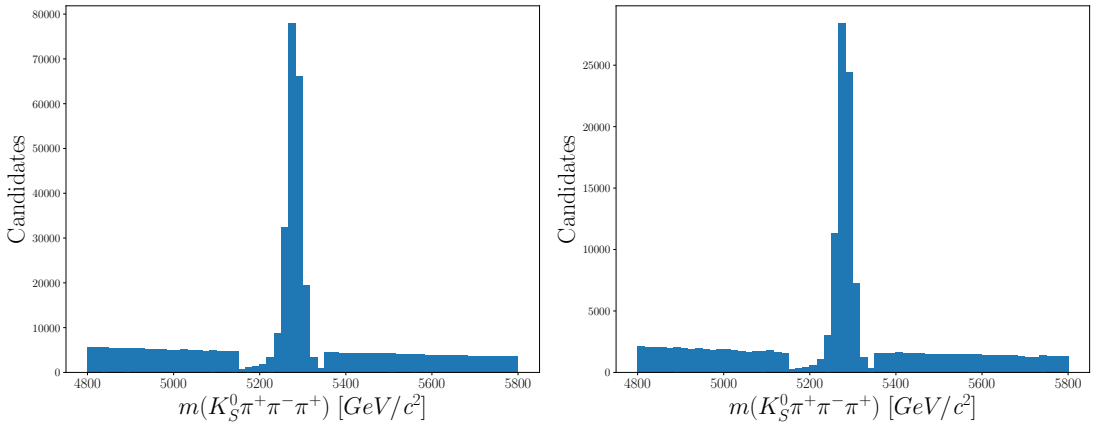


FIGURE 5.3: Flat PHSP toy signal sample showing the mass window on many mass projections.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Variable	Description	Rank (DD)	Rank (LL)
B_ETA	B pseudorapidity	0.10	0.10
B_PT	B transverse momentum	0.095	0.10
B_ENDVERTEX_CHI2	B vertex fit χ^2	0.10	0.12
log_Ks_IPCHI2_OWNPV	Ks IP χ^2 w.r.t. PV	0.088	0.090
logf_Ks_DIRA_OWNPV	Ks pointing angle	0.099	0.11
log_B_FDCHI2_OWNPV	B flight distance χ^2 w.r.t. PV	0.096	0.075
log_Ks_FDCHI2_OWNPV	Ks flight distance χ^2 w.r.t. PV	0.079	**0.12**
log_B_IPCHI2_OWNPV	B IP χ^2 w.r.t. PV	**0.13**	0.096
logf_B_DIRA_OWNPV	B pointing angle	0.11	0.099
Sumhilog_IPCHI2_OWNPV	Sum of 3 charged tracks IP χ^2	0.095	✗
SumhiKslog_IPCHI2_OWNPV	Sum of 5 charged tracks IP χ^2	✗	0.088

TABLE 5.4: Variables used in the BDT and the rank of importance for both K^0_S-reconstruction cases. Where log represents the logarithm and $f(x) = \frac{x}{1-x}$. The sum of the IP χ^2 of the children excludes the K^0_S pions for the DD case.



(A) DD (B) LL

FIGURE 5.4: B-mass projection showing the BDT training data for DD and LL samples. MC is used for the signal and side-bands from real data as backgrounds. Showing only the 1516 sample for clarity.

side-band data as the background category; the side-band data are taken from the B-mass candidate invariant mass range: $[4800, 5150] \cup [5350, 5800] MeV/c^2$. The B-mass projections are shown in Fig. 5.4. All the data used for the training pass through the same selection. Note that there is no specific training for the control channel, the same BDT training for both channels is used.
The training is split in different samples as previously mentioned: 1112, 1516 and 1718. In addition, the K_S reconstruction categories are also split, with slightly different variables used in the BDT as reported in Table 5.4. Thus, 6 different samples are used for the BDT training, its performance is shown in Fig. 5.5 and in Fig. 5.6. In the plots, only one year category is shown to avoid overcrowding the text with plots. Similar performances are obtained for all the samples. As shown in the χ^2 combination to extract the final branching fraction result in Fig. 5.41, all the measurements from the different samples are in agreement.

5.2.3.4 FIGURE OF MERIT STUDY

We then evaluate the trained BDT in real-data and MC, so we can study the impact of the BDT on the efficiency. It is important to remark that for the control channel we do not train a different BDT, we use the same BDT training as for the signal channel. This should have little impact on the final control channel yield as the training variables are similar on signal and control channels. We show in Fig. 5.7 the sWeighted real-data from the control channel against the signal channel MC to show that we can use the same BDT training on the control channel. The impact on the efficiency of the BDT on the control channel will be evaluated.

After the BDT is evaluated on data, we can perform the figure of merit (FoM) study. We fit the data for different BDT cuts, we use the same fitting model as in the mass fit from which we extract the signal

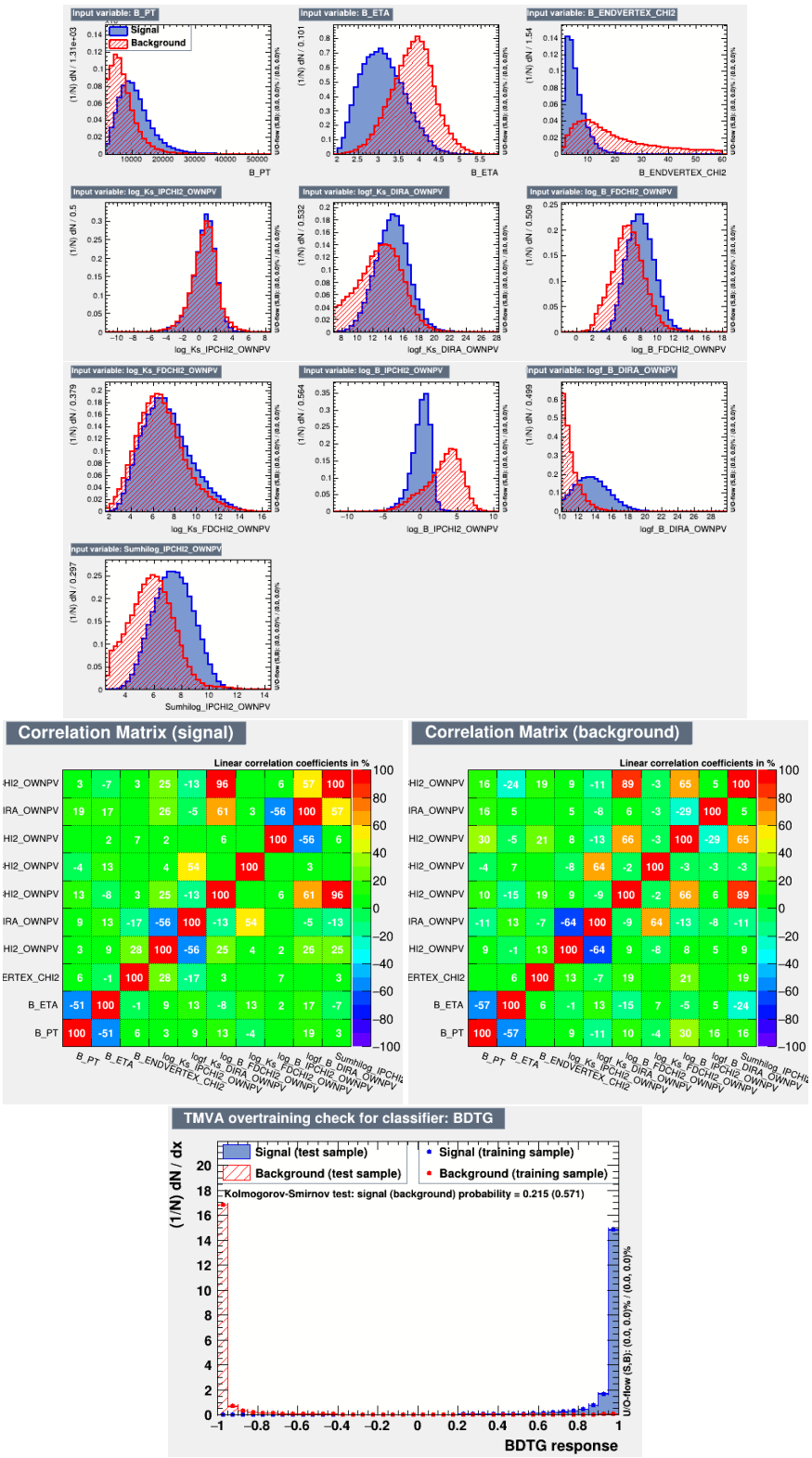


FIGURE 5.5: BDT training performance for 1516 DD sample.

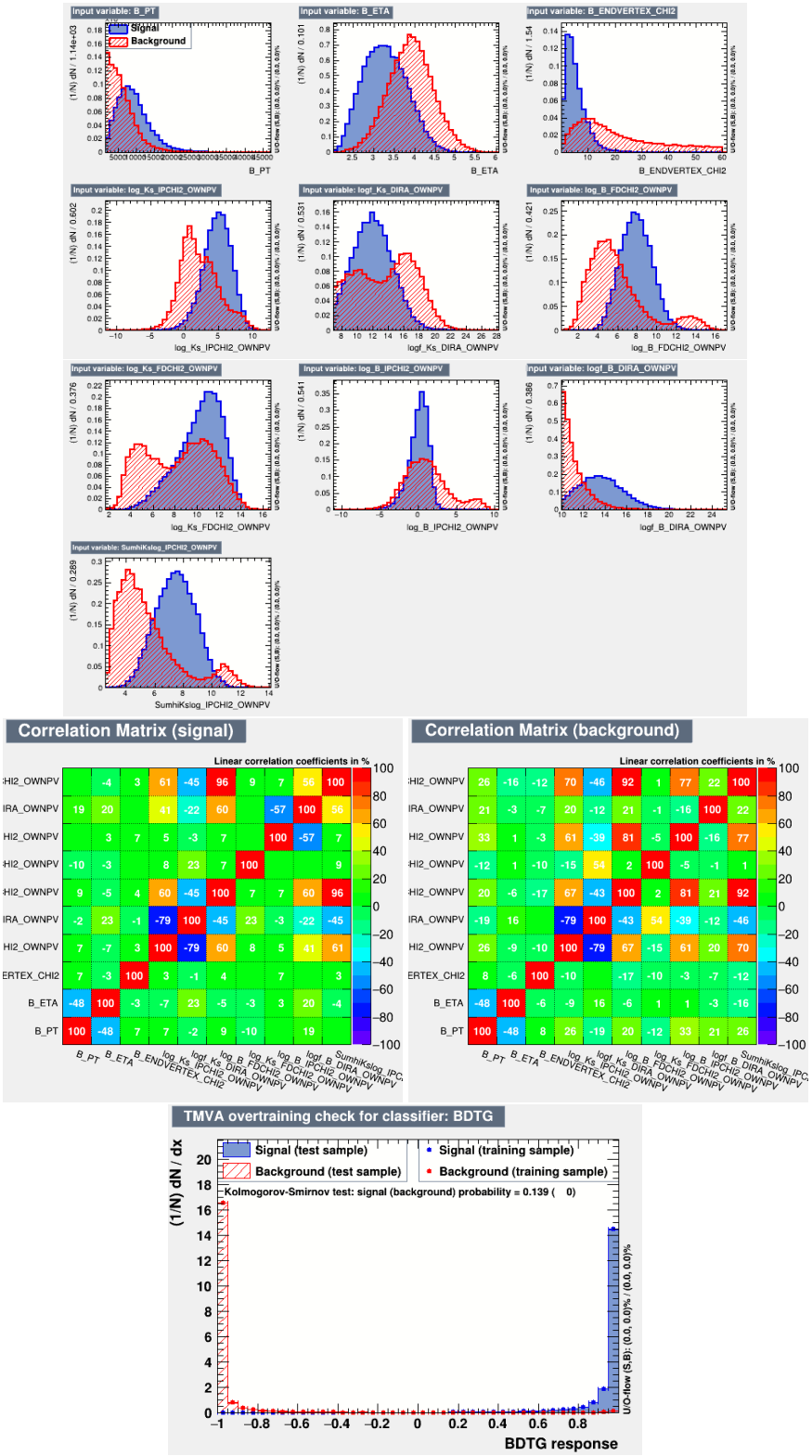
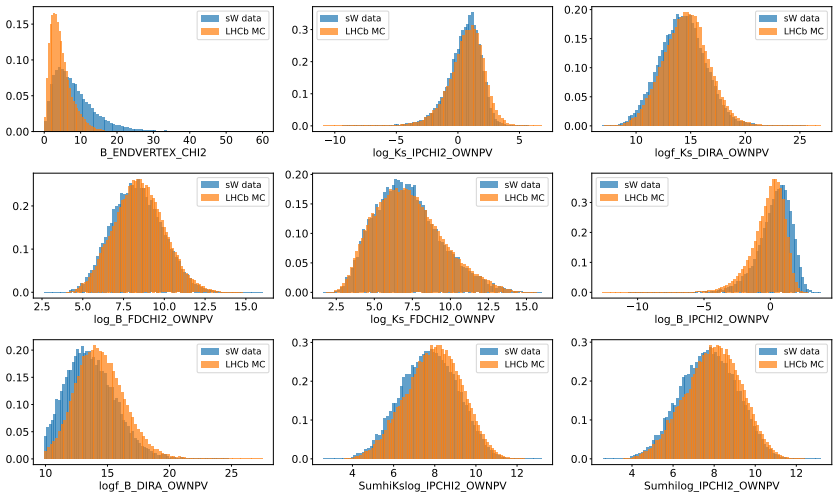
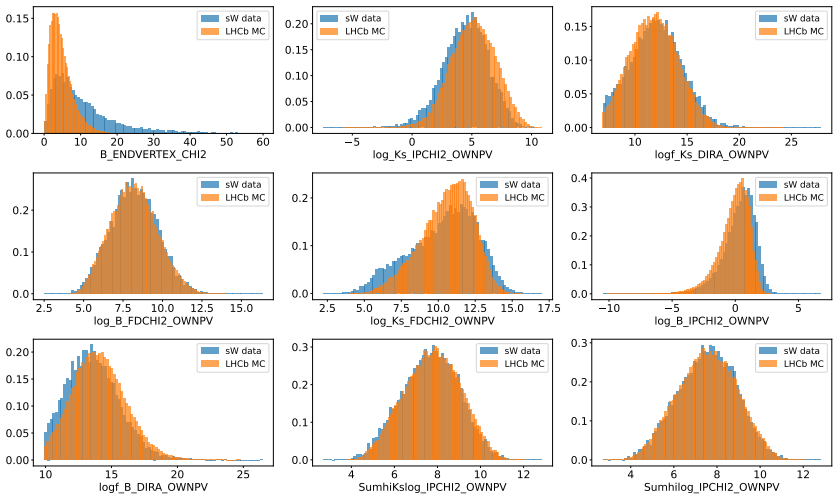


FIGURE 5.6: BDT training performance for 1516 LL sample.



(A) DD



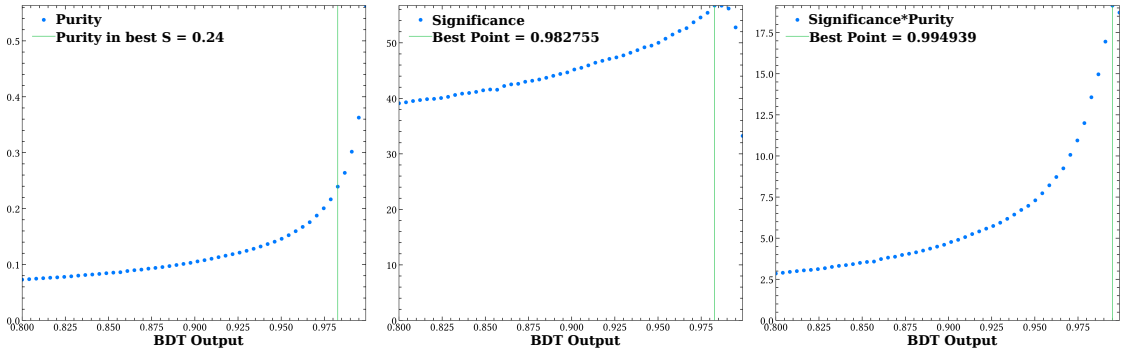
(B) LL

FIGURE 5.7: sWeigted control channel data against signal channel MC for the BDT training variables. The only significant mismatch is due to the flight-distance of the D-meson on the top-left plot. We demonstrate with this plot that specific BDT training for the control channel would not have a big impact on the extracted yield from real-data. Showing only the 1718 sample for clarity.

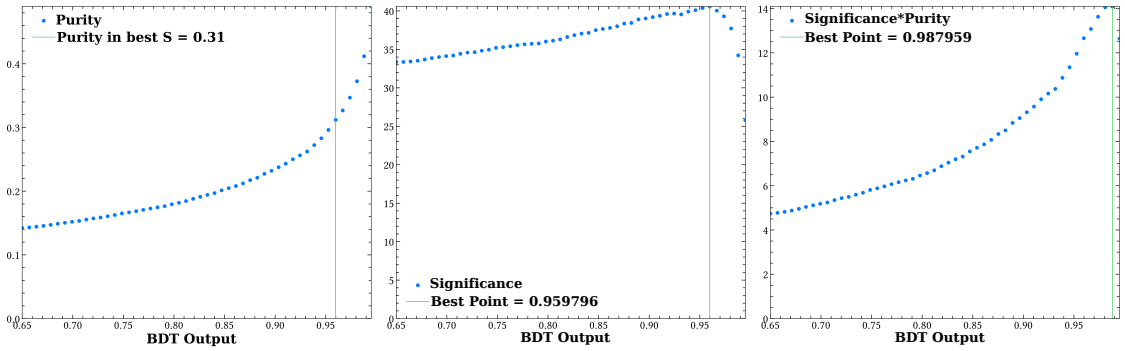
yields, as is shown in the following section. We select the fit with the highest significance of the signal. We define the significance as $S = \frac{N_S}{\sqrt{N_S+N_B}}$, we also report the purity of the sample, that is, $P = \frac{N_S}{N_S+N_B}$ and their multiplication. N_S is the number of signal events and N_B the background yield. The results are shown in Fig. 5.8 for the signal channel and in Fig. 5.9 for the control channel.

5.3 YIELDS

This is a crucial stage for the analysis, where one can disentangle the different contributions to the data by performing the mass fits. Mass fits are executed in 2 steps, a first fit in a wide region $[5.0, 5.8]\ GeV/c^2$ and then a second mass fit is performed in $[5.2, 5.5]\ GeV/c^2$ but fixing all the fitting components except for the relative yield between signal and background. The purpose of the second mass fit is to obtain the sWeights, as we are interested in the PHSP projections and they might be correlated with the B mass, we focus on the signal region in order to reduce these correlations. In the control channel, as the sWeights will not be used to extract the signal amplitudes from data, the second step is skipped.

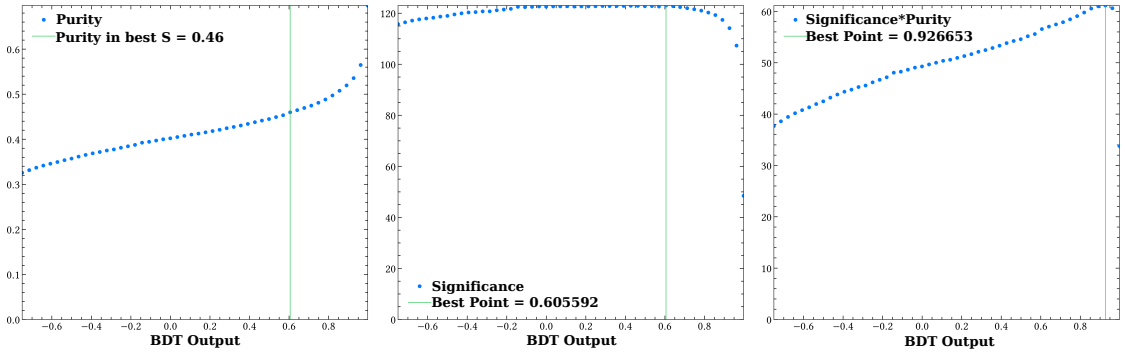


(A) DD

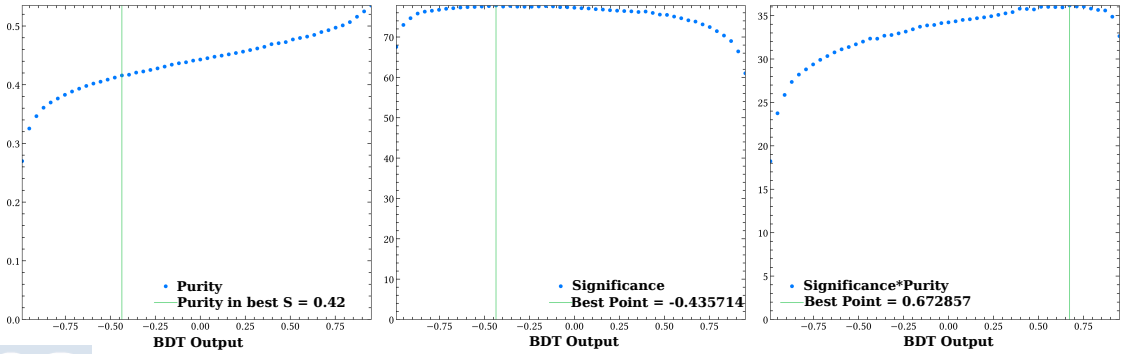


(B) LL

FIGURE 5.8: FoM study for signal channel. On the left, the purity is shown, at the middle the significance, which is the FoM used, and, on the right, the purity times the significance. Showing only the 1718 sample for clarity.



(A) DD



(B) LL

FIGURE 5.9: FoM study for control channel. On the left, the purity is shown, at the middle the significance, which is the FoM used, and, on the right, the purity times the significance. Showing only the 1718 sample for clarity.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.3.1 MASS FIT COMPONENTS

The MC samples are used to constrain the line-shapes of the different fitting components in the mass fits. An important remark is that there is no isolation variables used in the selection, so there are several partially reconstructed backgrounds that appear in the fitting region that must be identified. We generate MC samples in order to constraint these contributions. Table 5.2 shows the generated MC samples. If some isolation variables were used in the data selection, the partially reconstructed backgrounds would be suppressed and the mass fit would be easier to perform. Therefore, integrating isolation variables into the data selection process could improve the analysis.

The signal and misidentified samples are truth-matched. In the case of the signal component, as the reconstructed and simulated decay chains are identical, we can use the information from the *BKG_CAT* variable, using the 0, 20 and 50 categories. These variable values represent reconstructed decay chains that match the final state of the simulated ones, including radiative decays. For the misidentified component, we cannot use this variable. Instead, we isolate the component to be parameterised by matching the decay vertex of the parent particle with the origin vertex of the children particles using the *TRUE* MC information. This ensures that only the component of interest is included in the sample. Consequently, when fitting to constrain the line-shapes, there is no need to account for any additional contributions in the MC fit.
However, in other background MC samples, such as the partially reconstructed ones, truth-matching is not possible because the decay chains in the reconstructed data differ from those in the simulated data. Thus, in addition to the component under study, there is additional background that must be considered. To avoid including this background in the parameterised fitting component, we extract it from the MC data by modeling it. We then add an extra 2^{nd}-order Chebyshev polynomial to the fit.
In the fit to real data we only use the component of interest, removing in this way the combinatorial background that is present in the MC samples due to the combination of random tracks that pass the selection criteria. Finally, we fix all the line-shape parameters from these fits to MC. In this way, we constrain the fitting functions that will be used in the fit to real data. Appendix A.1 shows the MC fits used to extract the various components for the mass fit.
Moreover, it is crucial to highlight that, as stated earlier, the fits for the signal channel are conducted in two distinct regions. Since we are using several functions with a relative yield to model each contribution from the MC sample, when changing the fitting range, we must scale the line-shape to the new region. We do the scaling by simply repeating the fits to MC in the new region.

The free parameters on the mass fits will be the yields of each component and also some parameters on the signal component that account for data/MC differences, these are $\delta\mu$ and $\delta\sigma$ that accounts for momentum scale and resolution differences between data and MC. In addition, the combinatorial background model is also freed on the mass fit.

Note that, before performing each fit, a K^0_S selection is applied, thus, the K^0_S projection of the signal MC sample is also fitted to parameterise the K^0_S line-shape.

Before specifying each fitting component that contributes to each fit, we are showing how we generally parameterise them:

- Signal: Double Crystal Ball line-shape, with 2 Gaussian cores and 2 power-law tails. Note that the mean and sigma of the 2 Gaussian cores are let to differ. The parameters that count for data/MC differences impacts the whole distribution: as in the Crystal Ball function the tails are parameterised with respect to the Gaussian core, the data/MC corrections are propagated to the tails also. The signal component on the K^0_S fit also uses this model.

$$\mathcal{P}_{sig} = f_s CB(x, \mu + \delta\mu, \sigma \cdot \delta\sigma, \alpha_a, n_1, \alpha_2, n_2) + (1 - f_s)G(x, \mu + \delta\mu + \delta\mu_G, \sigma \cdot \delta\sigma \cdot \delta\sigma_G) \tag{5.3}$$

$$CB(x, \mu, \sigma, \alpha_1, n_1, \alpha_2, n_2) \begin{cases} (\frac{n1}{|\alpha_1|})^{n_1} (\frac{n_1}{|\alpha_1|} - |\alpha_1| - \frac{x-\mu}{\sigma})^{-n_1} e^{-\frac{|\alpha_1|^2}{2}}, & \text{if } \frac{x-\mu}{\sigma} \leq -|\alpha_1| \\ e^{\frac{-(\frac{x-\mu}{\sigma})^2}{2}}, & \text{if } -|\alpha_1| < \frac{x-\mu}{\sigma} < |\alpha_2| \\ (\frac{n2}{|\alpha_2|})^{n_2} (\frac{n_2}{|\alpha_2|} - |\alpha_2| - \frac{x-\mu}{\sigma})^{-n_2} e^{-\frac{|\alpha_2|^2}{2}}, & \text{if } \frac{x-\mu}{\sigma} \geq |\alpha_2| \end{cases} \tag{5.4}$$

- Mis-ID: same as signal but with no parameters to take into account for data/MC differences. Note that the mis-ID channel is Cabibbo suppressed with respect to the signal channel. As this contribution is small, there is no sensitivity to extract its data/MC corrections.
- Partially reconstructed: parameterised with double Argus line-shape. In the case of the part. reco. η' decay, another Argus contribution is added: As we have produced more events to model this contribution, because it overlaps the signal, adding an extra Argus ensures a good precision determining

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the model.

$$\mathcal{P}_{partially\ reconstructed} = f_s A(x, m_{01}, c_1, p_1) + (1 - f_s) A(x, m_{02}, c_2, p_2) \tag{5.5}$$

or for the η' model:

$$\begin{aligned}\mathcal{P}_{partially\ reconstructed} = f_{s1} A(x, m_{01}, c_1, p_1) + f_{s2} A(x, m_{02}, c_2, p_2) + \\ (1 - f_{s1} - f_{s2}) A(x, m_{03}, c_3, p_3)\end{aligned} \tag{5.6}$$

where

$$A(x, m_0, c, p) = \frac{2^{-p} c^{2(p+1)}}{(\Gamma(p+1) - \Gamma(p+1, c/2)} \frac{x}{m_0} \left[1 - (\frac{x}{m_0})^2\right]^p e^{c\left(1 - (\frac{x}{m_0})^2\right)} \tag{5.7}$$

- Combinatorial background: The contribution is parameterised using a 2^{nd}-order Chebyshev polynomial, the line-shape parameters are freed on the fit to real-data.

$$\mathcal{P}_{combinatorial} = Chebyshev(x, C_1, C_2) = C_1 + C_2 x \tag{5.8}$$

5.3.1.1 SIGNAL CHANNEL - WIDE REGION

The selected range for the mass fit in the wide region is $[5.0, 5.8]\, GeV/c^2$. The shared fitting components for all the mass fits are the signal, the mis-ID and the combinatorial background. The MC fits to model the different contributions are shown in Appendix A.1. In Fig. A.1 the signal component is modelled and in Fig. A.2 the mis-ID characterisation is shown. In the signal channel mass fit in the wide region there are some extra partially reconstructed contributions, these are:

- $B^+ \to K^*(K^0_S\pi)\eta'(\pi\pi\gamma)$ missing the photon. MC fit is shown in Fig. A.3.
- $B^0_s \to D^-_s(K^0_S\,\pi^+\pi^-\pi^-)\pi^+$ missing a charged pion from the D^-_s. MC fit is shown in Fig. A.4.
- $B^+ \to D^0(K^0_S\,\pi^+\pi^-\pi^0)\pi^+$ missing the neutral pion from the D^0. MC fit is shown in Fig. A.5.
- $B^0 \to D^{*-}(D^0(K^0_S\,\pi^+\pi^-)\pi^-)\pi^+$ missing a charged pion from the D^0. MC fit is shown in Fig. A.6.

5.3.1.2 SIGNAL CHANNEL - SIGNAL REGION

The narrower selected range for the second signal channel mass fit is $[5.2, 5.5]\, GeV/c^2$. Within this region, in addition to the signal component, mis-ID, and combinatorial background, only some partially reconstructed decays remain, as the others are excluded from the fitting region. The fit to signal MC is shown in Fig. A.8 and the fit to the mis-ID in Fig. A.9. Finally, the surviving partially reconstructed components are:

- $B^+ \to K^*(K^0_S\pi)\eta'(\pi\pi\gamma)$ missing the photon. MC fit is shown in Fig. A.14.
- $B^0_s \to D^-_s(K^0_S\,\pi^+\pi^-\pi^-)\pi^+$ missing a charged pion from the D^-_s. The line-shape characterisation is taken from the only Argus distribution surviving to this range in the fit on Fig. A.4.

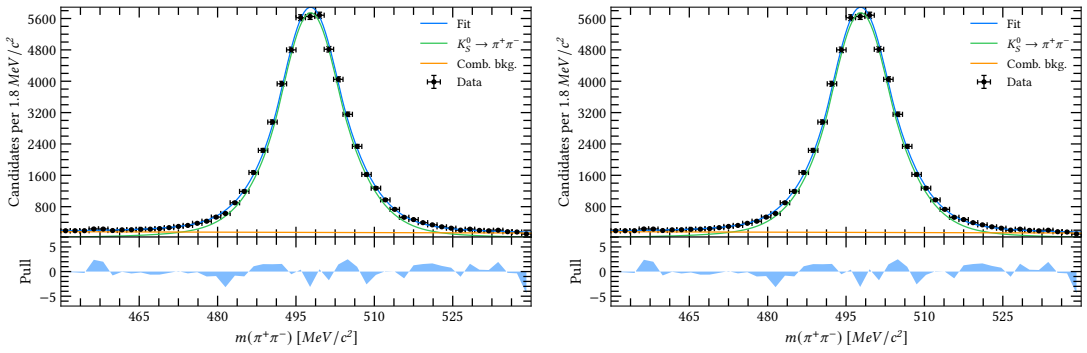
5.3.1.3 CONTROL CHANNEL - WIDE REGION

In the control channel there is a different PHSP mass window, thus, the partially reconstructed fitting components are different, the only sharing component with the signal channel is the $B^0_s \to D^-_s(K^0_S\,\pi^+\pi^-\pi^-)\pi^+$ decay. Then, besides the signal (which fit to MC is shown in Fig. A.8), mis-ID (fit to MC in Fig. A.9) and combinatorial background, there are only 2 more fitting components:

- $B^0_s \to D^-_s(K^0_S\,\pi^+\pi^-\pi^-)\pi^+$ missing a charged pion from the D^-_s. MC fit is shown in Fig. A.10.
- In the fitting range appears a partially reconstructed decay from which there is no MC sample. We parameterise this component as a double Argus distribution, as defined in Eq. 5.5. This line-shape does not overlap with the bulk of the signal component, thus, we let it free in the fit constraining the range of its parameters to match the desired distribution. This component can be appreciated in Fig.5.15 peaking at $5100\, MeV/c^2$. This background is likely coming from a D^* decay.

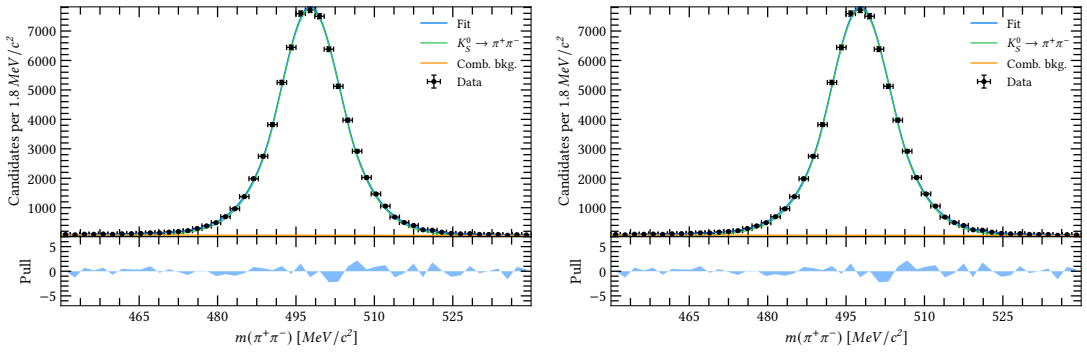
5.3.2 K^0_S SELECTION

Before performing the fit to the B-mass, we fit the K^0_S-mass spectra for the selected BDT cut and fit range. The line-shape is defined from a fit to MC data using a double Crystal Ball. This fit is shown in Fig. A.7 for the signal channel (note that for the K^0_S we do not change the range of the fit, so the same line-shape is used for the selection on both regions) and in Fig. A.7 for the control channel. The fit to real data is performed



(A) DD (B) LL

FIGURE 5.10: K^0_S selection for signal channel mass fit on the wide region. Showing only the 1718 sample for clarity.



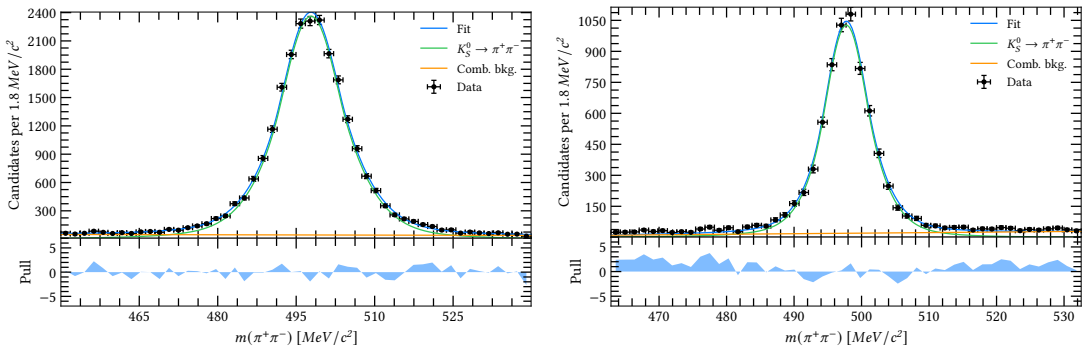
(A) DD (B) LL

FIGURE 5.11: K^0_S selection for control channel mass fit. Showing only the 1718 sample for clarity.

by adding a 2^{nd}-order Chebyshev polynomial function to count for the combinatorial background and, as for the B-mass signal component fit, we add the momentum scale and resolution parameters to count for possible data/MC differences. Once the fit is performed, we select $\mu \pm 3\Gamma$ of the K^0_S signal parameters. This selection is applied before each mass fit, so there are 3 different K^0_S-selection fits. The fit to the signal channel wide region is shown in Fig. 5.10, the one to the control channel appears in Fig. 5.11 and the K^0_S selection applied in the signal region to obtain the signal channel sWeights is shown in Fig. 5.12.
Note that the performance of this last fit in the LL category, where the K^0_S candidates decay earlier in the detector, is worst than in the other K^0_S selection fits (Fig. 5.12B). This effect is caused by the tight selection on the signal K^0_S. The candidates come mainly from VV decays on the charmless region of the PHSP, i.e. the region where these K^0_S have less energy and thus a lower p_T. In addition, the VV decays are highly longitudinally polarised, thus the K^0_S are boosted in the forward direction, close to the beam-pipe, where the material interactions are more likely to happen. In conclusion, these K^0_S candidates are more difficult to model and they are hihgly polluted by the material interactions. Anyhow, the model extracted from this fit is not propagated through the analysis, the only purpose of these fits is to obtain the mass window to select the K^0_S candidates for the analysis.

5.3.3 WIDE REGION

In the initial fit, the mass region $[5.0, 5.8]\,GeV/c^2$ is selected, and a fit is performed with the yields, resolution and momentum scale correction factors, and the combinatorial background line-shape as free parameters. As there are many line-shapes in the low-mass tail of the signal and there are few events to fit, the $B^+ \to K^{*+}\eta'$ contribution is Gaussian constrained in order to retrieve an accurate number of events for each contribution. We choose this contribution as there is available information of the expected yield with respect to the signal. Moreover, the PHSP efficiencies of this contribution should be very close to those of the signal component, as the only difference between these decays is the addition of the radiating photon, it is a soft photon due to the range that is selected on the fit. In the BaBar paper [90] we can find a measure of the branching fraction for this channel, where the K^{*+} can be: $K^{*+}(892)$, $K^{*+}(1400)$ or



(A) DD (B) LL

FIGURE 5.12: K^0_S selection for signal channel mass fit on the signal region. Showing only the 1718 sample for clarity.

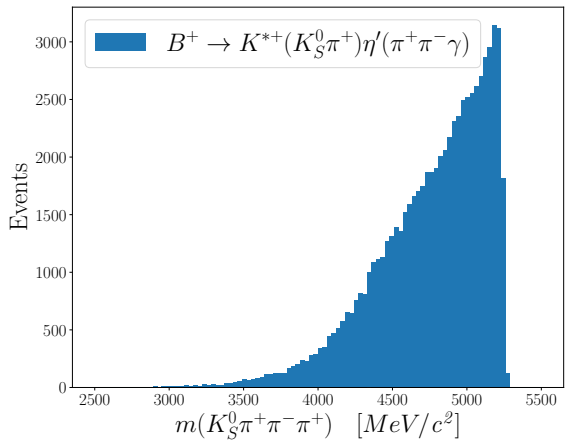
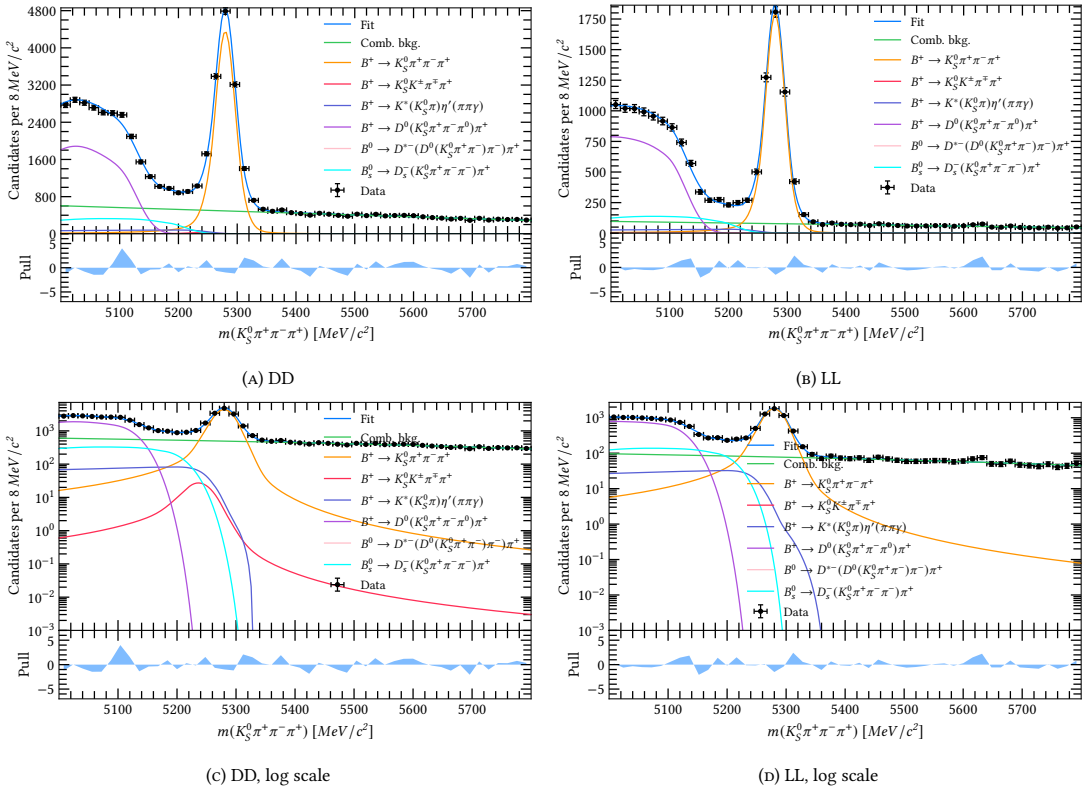


FIGURE 5.13: Toy simulation to extract the efficiency of the $B^+ \to K^*(K^0_S\pi)\eta'(\pi\pi\gamma)$ partially reconstructed decay to fall in the fitting region.

$K^0_S\pi^+$ S-wave. As we perform the constraint with respect to the signal component, we need the branching fraction of the signal as an input for this constraint. It is assumed that the signal being fitted is composed of a combination of $B^+ \to K^*(K^0_S\pi)\pi\pi$ and $B^+ \to K^0_S a_1(\pi\pi\pi)$ decays, which branching fractions are taken from BaBar analysis [91, 92]. Finally, in order to perform the constrain from these results, one must take into account the efficiency of the $B^+ \to K^{*+}\eta'$ contribution to fall in the fitting region. To do so, we generate a toy simulation for this background (without passing it through the detector) and we compute the ratio of the events that fall in the fitting region by the number of generated events, the efficiency of this fiducial cut is 0.3333 ± 0.0025; in Fig. 5.13 the distribution of this decay on the B-mass projection, for the toy sample, is shown. In conclusion, one can compute the constraint multiplying this efficiency by the ratio of branching fractions $\frac{BF_{B^+\to K^*(K^0_S\pi)\eta'(\pi\pi\gamma)}}{BF_{B^+\to(K^*(K^0_S\pi)\pi\pi)+(K^0_S a_1(\pi\pi\pi))}}$, the value for the final constraint is 0.091 ± 0.033. To perform the Gaussian constraint we sum to the likelihood fitting function a Gaussian cost function. The value of the $B^+ \to K^{*+}\eta'$ yields that does not match the constraint are penalised; the uncertainty of the constraint is associated to the standard deviation of the Gaussian cost function. The fit result is shown in Fig. 5.14.

5.3.3.1 CONTROL CHANNEL

In contrast, in the control channel there are some differences with respect to the signal channel. As there are not many overlapping components on a particular region of the fit, there is no need to use any Gaussian constraint to get a meaningful yield for all the fitting components. In addition, the partially reconstructed decays are different because of the different PHSP acceptance on both channels. Moreover, as there is no MC to model the extra partially reconstructed decay, we include an extra free Argus distributions to this fit, as defined in Eq. 5.5, limiting the range of their parameters to match this component. This free Argus



(A) DD

(B) LL

(C) DD, log scale

(D) LL, log scale

FIGURE 5.14: Mass fit on the signal channel wide region. The $K^0_S\,\pi^+\pi^-\pi^+$ invariant mass projection is shown together with all the fitting components. Showing only the 1718 sample for clarity.

does not impact the signal yield as it is clearly separated from the signal bulk, then it can be freed in the fit. The fit result is shown in Fig. 5.15. The yields for the control channel to be used on the branching fraction calculation are extracted from this fit and reported in Table 5.6. A set of sWeights is also generated for the signal of the control channel, it will be used to re-weight the BDT variables as is shown in Sec. 5.4.6.2.

5.3.4 SIGNAL REGION

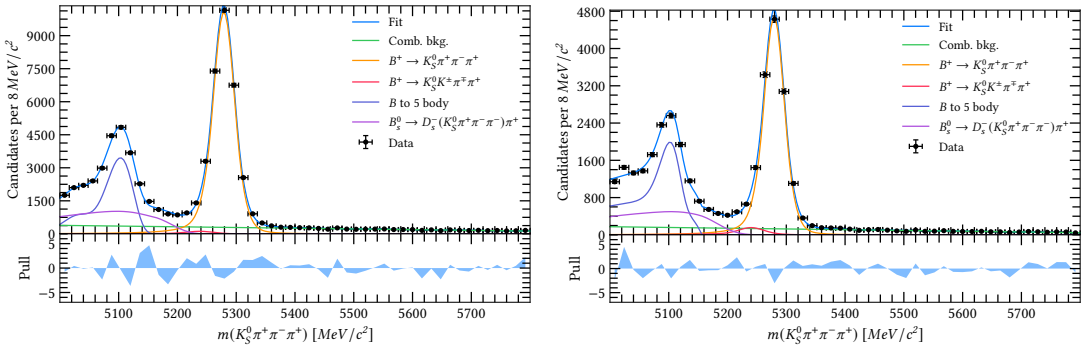
After performing the fit in the wider region, we focus on the signal region, defined as $[5.2, 5.5]\,GeV/c^2$. This is interesting in order to produce a set of sWeights less correlated with the PHSP projections. There are several parameters that are fixed from the previous fit, such as the combinatorial background model parameters and the data/MC differences parameters $\delta\mu$ and $\delta\sigma$. The background line-shape is considered as a whole, summing all the contributions together. To do so, the integral ratio of each line-shape in the new region relative to the wider region must be considered. This approach results in a single model that parameterise all the background contributions in the signal region.

In this way, the only free parameter is the signal yield. As there are only two fitting distributions, the background yield can be defined relative to the signal yield. The fit result is shown in Fig. 5.16. The yields for the signal channel to be used on the branching fraction calculation are extracted from this fit and reported in Table 5.6.

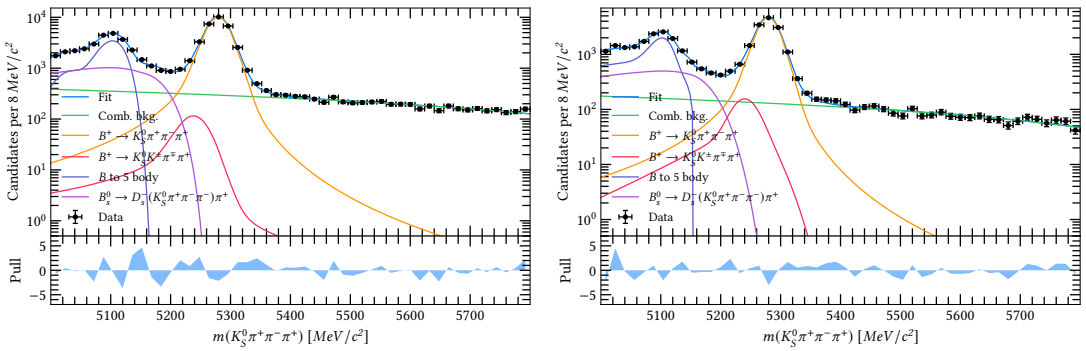
5.3.4.1 SWEIGHTS

The sPlot technique [88], allow to statistically suppress a certain PDF contribution from data, remaining the desired one. If we have 2 PDF contributions in a fit, signal and background, we can apply the technique to get a per-event weight that transform the data to follow the signal distribution. These weights are the sWeights for the signal fitting component.

The main issue when using the sWeights is that the technique assumes that there is no correlation between the fitting variable and the variables where the sWeights are used. As we already mentioned, we

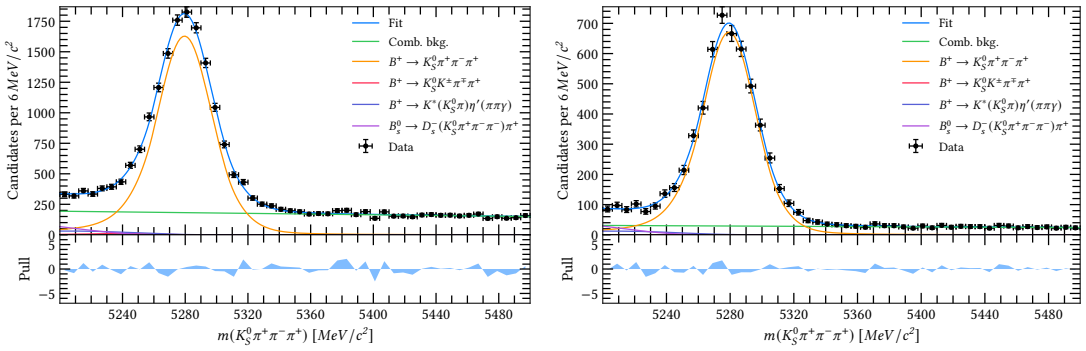


(A) DD (B) LL

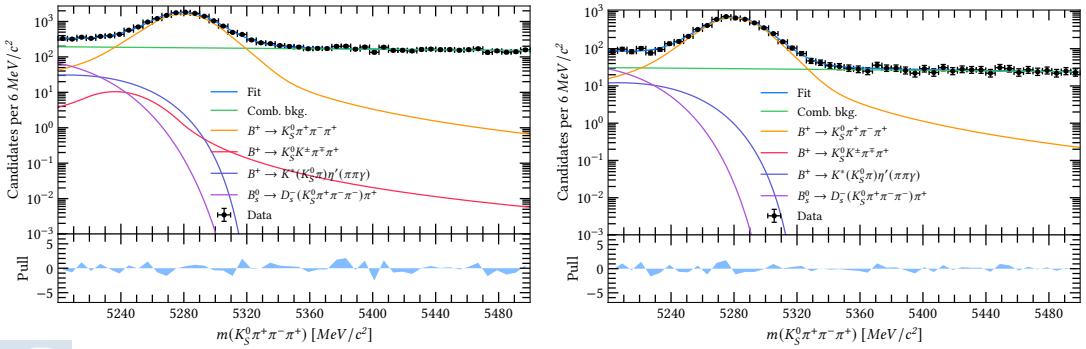


(C) DD, log scale (D) LL, log scale

FIGURE 5.15: Mass fit on the control channel wide region. The $K^0_S\,\pi^+\pi^-\pi^+$ invariant mass projection is shown together with all the fitting components. Showing only the 1718 sample for clarity.



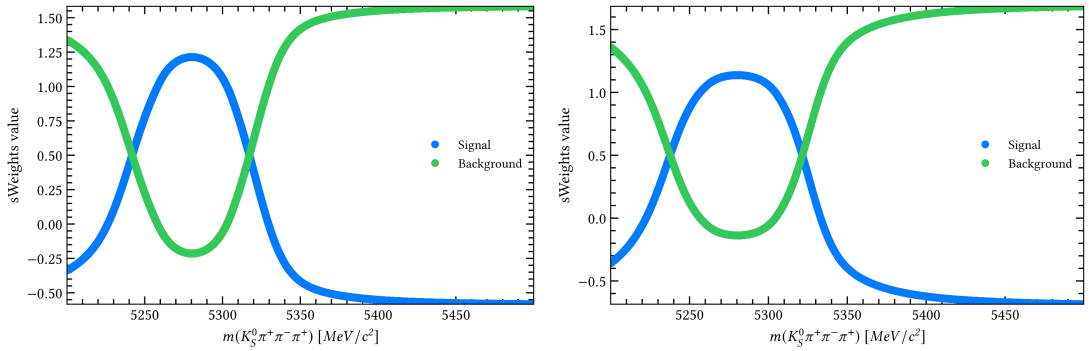
(A) DD (B) LL



(C) DD, log scale (D) LL, log scale

FIGURE 5.16: Mass fit to the signal channel on signal region. The $K^0_S\,\pi^+\pi^-\pi^+$ invariant mass projection is shown together with all the fitting components. Showing only the 1718 sample for clarity.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) DD (B) LL

FIGURE 5.17: sWeight values obtained from each event in the signal region mass fit plotted against the $K^0_S\,\pi^+\pi^-\pi^+$ invariant mass projection. Showing only the 1718 sample for clarity.

use the two-step mass fit in order to reduce these correlations, as the PHSP variables and the B-mass might be correlated. In Sec. 5.4.1 these correlations are further studied.

The signal sWeights are crucial for the analysis; they will be used to obtain how the signal amplitudes are distributed in the PHSP in a data-driven approach using a Gradient Boosted re-weighting algorithm (GBrw). In Fig. 5.17 a scatter plot of the sWeights against the B-mass is shown.

5.4 EFFICIENCIES

Controlling the efficiencies is key to accurately measuring the branching fraction ratio. It is one of the main sources of bias in this kind of analysis since they are obtained from simulation, which implies that a good understanding of the simulation is needed in order to correct for differences with real data. These differences are the primary source of bias in the measurement.
To estimate the efficiencies, a signal MC sample is generated mimicking the amplitudes in data, and, after some small corrections that make the MC distributions closely resemble the real data, the efficiency can be estimated by dividing the number of MC events surviving the data selection by the number of generated events. This method is used for the control channel because their amplitudes are well-known.
However, in the current analysis this is not the procedure for the signal mode. We generate a signal channel MC that is distributed flat in the PHSP, in this way, any deviation from flatness is due to the efficiency effects in the detector simulation that mimics data acquisition. In addition, one can apply the offline data selection to the MC and estimate the overall efficiency for the flat PHSP sample. Finally, to determine the efficiency in the data, it is necessary to consider the influence of data amplitudes absent in the MC. For this purpose, a GBrw is employed within the 5-dimensional PHSP of a four-body decay. This allows us to model any data amplitude for any PHSP projection. By applying the same approach, we estimate the efficiency distribution from the flat MC sample. After adjusting for known discrepancies that are not accurately simulated, we ensure that our MC distributions closely match the actual efficiencies in data.
To maintain consistency with the yield extraction, we calculate the efficiencies by dividing them into groups of two years and K^0_S reconstruction categories, similar to the mass fits. Since the MC has been generated and scaled by luminosity, there is no need to re-scale the MC by year or magnet polarity categories. It is worth noting that there is a noticeable dependency of the efficiency distribution on the K^0_S reconstruction category.

In this analysis, our simulated signal sample is distributed flat across the PHSP, thus, in this sample there is no information about the signal amplitudes. The novel approach for this analysis is to obtain these amplitudes from the data. There are some methods in the literature to extract a PDF distribution from data, as GBrw, or Kernel Density Estimation. The GBrw appeared to be the most accurate one for our use case, but both algorithms were investigated.
The ultimate objective of this procedure is to obtain a generator level MC that follows the data amplitudes, thus, we need 2 PDF components: the efficiency ($\epsilon(\Phi_4)$) and the signal amplitudes ($|A(\Phi_4)|^2$). Here Φ_4 refers to the four-body PHSP where these distributions are evaluated. We can extract from data $\epsilon(\Phi_4)\cdot|A(\Phi_4)|^2$, where the contributions from the detector and data selection efficiencies and the data amplitudes are present. As we have generated a flat MC sample, any deviation from flatness is due to the efficiency, thus,

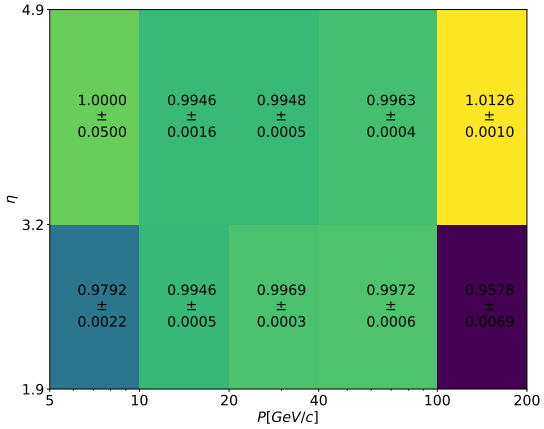


FIGURE 5.18: Tracking efficiency correction from *TrackCalib2* package, the plot shows the correction for 2018 data as a function of the track η and momentum values.

the efficiency distributions in the PHSP can be extracted from MC. In this way, we already have all the information that we need to obtain the efficiency of the data.

The main downside of this procedure is that, as the MC has limitations in reproducing the real data distributions, the MC is usually re-weighted to match the real data distributions used to enhance the signal significance. In the current strategy, this re-weighting cannot be performed, as we would be copying the amplitudes in data and we want our signal MC sample to only reproduce the efficiency effects.
However, what can be done is to use the available tools, developed by the collaboration, to correct our MC for the following features:

- PID efficiencies can be corrected using the *PIDCalib2* package [78].
- Tracking efficiencies [79] can be corrected using *TrackCalib2* package. An example of the tracking corrections as a function of the track momentum and η is shown in Fig. 5.18.
- Lo Hadron trigger efficiency can be corrected using the *Lo Calo* tool [80], it properly reproduces the Lo trigger efficiency. In this way we can parameterise the PHSP efficiency distribution of the TIS and TOS samples. By correcting these distribution and scaling each trigger category by its amount in data, we can correct the efficiency due to the Lo Hadron trigger. We follow the same procedure as in [93].

On the other hand, for the control channel our MC is generated following the amplitudes in data, so we can re-weight our MC to be sure that the efficiencies that we are estimating are the ones in data. The correction procedure for the control channel is the following:

- First, we must re-weight the variables that are used in the packages for the efficiency corrections, these are the η and p of the child particles. We want to ensure that the corrections have an accurate impact.
- Then we apply the three efficiency corrections previously mentioned: PID, tracking and trigger efficiencies.
- Finally, we must take into account the impact of the BDT to suppress the combinatorial background on the efficiency. To evaluate accurately the BDT efficiency from the MC sample, the distributions of the variables used in the BDT training must match the real data distributions. We use a set of sWeights for the signal component of the control channel to re-weight the MC on these variables. The BDT training variables are shown in Table 5.4. Now that these distributions match, we can ensure an adequate efficiency extraction from MC when cutting on the same BDT value as the one selected on the Figure of Merit study for real data.

As naturally follows, for the signal channel we will consider this re-weighting procedure as a systematic uncertainty source, because we cannot perform this re-weighting procedure in the baseline as already explained. However, a more exhaustive analysis could be performed if we find a set of variables to apply the selection criteria that are not correlated with the PHSP variables. In this scenario, since the selection variables would be uncorrelated with the PHSP variables, we would not be copying the signal amplitudes when re-weighting them.

The aforementioned re-weightings are performed using the GBrw algorithm. These include re-weightings to match the real data distributions on the BDT variables of the MC sample as well as

re-weightings used to capture the PHSP distributions. In the three-body analogy, these distributions are captured by performing a binning of the 2D PHSP and comparing data and MC yields in each histogram bin. However, we will use a different strategy in which the GBrw algorithm is key to capture the PHSP distributions, as shown in Sec. 5.4.5. Consequently, we perform the per-event corrections using the GBrw by comparing the target distributions with the current or output distributions that we want to modify.
For the corrections applied to the LHCb MC sample, we will end up with a set of weights for each correction. These weights from each correction will be multiplied to obtain a global correction factor for each event in the MC sample. Finally, we will compute the efficiency by dividing the sum of these weights by the number of events at the generator level.
We can express the re-weighting algorithm as:

$$GBrw\ (\text{input/target}, w_{\text{input}};\ \text{output}) \to w_{\text{output}}\,, \tag{5.9}$$

obtaining the set of weights w_{output} that transform the output sample to match the input (or target) distribution. The variables utilised in the re-weighting process depend on each case; an arbitrary number of variables can be used simultaneously.
It is also important to remark that, just as the results may depend on the binning scheme for a three-body analysis, in this four-body analogy, the results may depend on the hyper-parameters of the GBrw. Systematic uncertainty estimations will be required due to the variation in these hyper-parameters.

5.4.1 CORRECTIONS ON THE PHSP

On the corrections where we want to parameterise the PHSP distributions, such as the TIS/TOS correction or the amplitudes correction, the usual PHSP variables of a VV decay will be used, the two vector masses and the three helicity angles:

$$\left(m_{K^0_S\pi^+_1},\ m_{\pi^-\pi^+_2},\ cos(\theta_{K^0_S\pi^+_1}),\ cos(\theta_{\pi^-\pi^+_2}),\ \phi\right)\,. \tag{5.10}$$

Note that, to ensure a good performance of the method, when plotting the results we do not simply show the re-weighted projections, we show many different projections of the PHSP to be sure that there is no mismatch on any mass projection.

In this analysis, the mass window includes multiple separated regions of the PHSP, to ensure a good parameterisation of each of these regions, we split them on different mass window blocks when computing the GBrw. These blocks are defined so that one of this disjoint regions is added whereas the other regions are excluded, ensuring that there is no event counted many times. In addition, as some of these blocks are the Bose symmetric refection of another block, we can join these blocks together to double the statistics on the re-weighting process. The mass window blocks are built from the mass window defined in Eq. 5.1. On the mass window a set of OR logical cuts are performed, these define the disjoint regions of the PHSP. We then associate each OR cut with a mass window block ensuring no overlap of events. For example, the $V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}$ block is defined as $\left(m(K^0_S\pi^+_1) < 1864.83\ AND\ m(\pi^-\pi^+_2) < 1864.83\right)$. We define the mass window blocks in the following way (! represents the *NOT* logic operator):

$$\begin{aligned}
&V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}\,,\\
&A_{K^0_S\pi^+_1\pi^-}\ AND\ !\left(V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}\right),\\
&V_{K^0_S\pi^+_2}V_{\pi^+_1\pi^-}\ AND\ !\left(V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}\ OR\ A_{K^0_S\pi^+_1\pi^-}\right),\\
&A_{K^0_S\pi^-\pi^+_2}\ AND\ !\left(V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}\ OR\ A_{K^0_S\pi^+_1\pi^-}\ OR\ V_{K^0_S\pi^+_2}V_{\pi^+_1\pi^-}\right),\\
&A_{K^0_S\pi^+_1\pi^+_2}\ AND\ !\left(V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}\ OR\ A_{K^0_S\pi^+_1\pi^-}\ OR\ V_{K^0_S\pi^+_2}V_{\pi^+_1\pi^-}\ OR\ A_{K^0_S\pi^-\pi^+_2}\right),\\
&A_{\pi^+_1\pi^-\pi^+_2}\ AND\ !\left(V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}\ OR\ A_{K^0_S\pi^+_1\pi^-}\ OR\ V_{K^0_S\pi^+_2}V_{\pi^+_1\pi^-}\ OR\ A_{K^0_S\pi^-\pi^+_2}\ OR\ A_{K^0_S\pi^+_1\pi^+_2}\right).
\end{aligned} \tag{5.11}$$

Using $m_{K^0_S\pi^+_1}$ and $m_{\pi^-\pi^+_2}$ on the $V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2}$ block, would be equivalent to use $m_{K^0_S\pi^+_2}$ and $m_{\pi^+_1\pi^-}$ on the $V_{K^0_S\pi^+_2}V_{\pi^+_1\pi^-}$ block. Thus, explicitly exchanging the π^+ child particles on one of the blocks, these blocks can be added. Note that this Bose-exchange also affects the definition of the angles. The same happens to the $A_{K^0_S\pi^+_1\pi^-}$ and $A_{K^0_S\pi^-\pi^+_2}$ blocks where the 2 π^+ can be exchanged on one of the blocks adding then the blocks together. Fig. 5.19 shows the different mass window blocks on some mass projections for the flat PHSP sample at generator level.
Finally, to summarise how we add the different blocks we write them explicitly here:

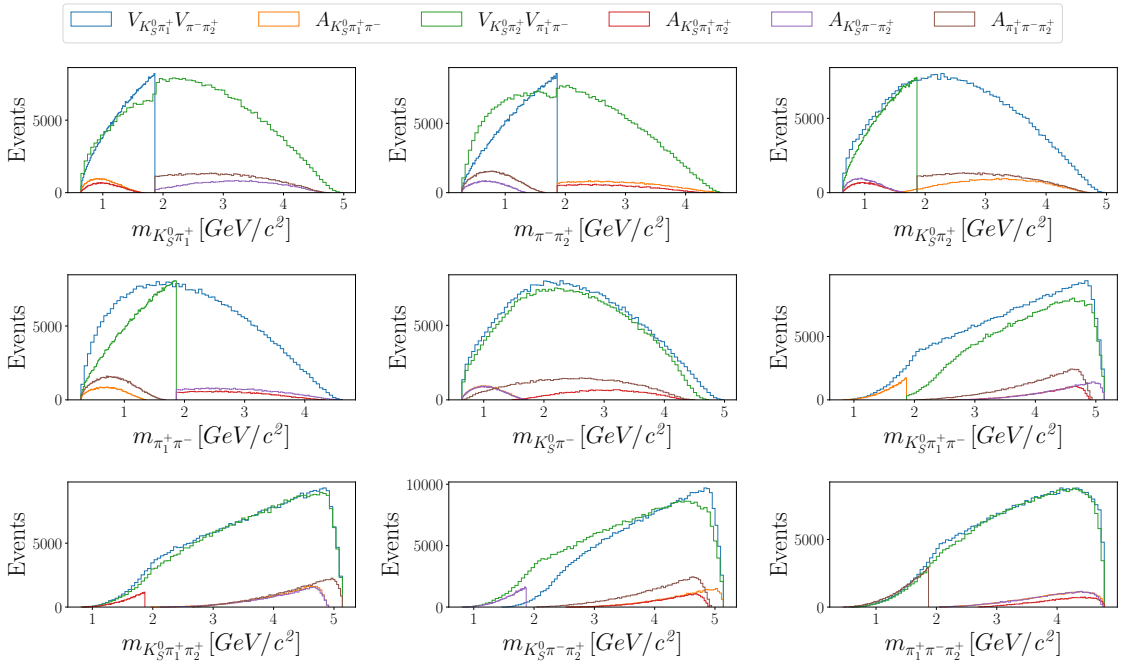


FIGURE 5.19: A flat PHSP sample at generator level on some mass projections of the PHSP. The sample is split by blocks as defined in Eq. 5.11. The number of events in each block can be observed. In the whole sample there are $1M$ events.

$$\begin{aligned} V_{K^0_S\pi^+_1}V_{\pi^-\pi^+_2} \quad &\oplus \quad V_{K^0_S\pi^+_2}V_{\pi^+_1\pi^-}(\pi^+_1 \leftrightarrow \pi^+_2)\,, \\ A_{K^0_S\pi^+_1\pi^-} \quad &\oplus \quad A_{K^0_S\pi^-\pi^+_2}(\pi^+_1 \leftrightarrow \pi^+_2)\,, \\ & \qquad A_{K^0_S\pi^+_1\pi^+_2}\,, \\ & \qquad A_{\pi^+_1\pi^-\pi^+_2}\,. \end{aligned} \tag{5.12}$$

Having finally four different training blocks.

As we are using the GBrw separately for each block, we then must join the samples back again, to join the output weights we must scale them by the number of events in each input block:

$$w_{\text{output}} = w^{block_1}_{\text{output}} \frac{N^{block_1}_{input}}{\sum w^{block_1}_{\text{output}}} \oplus w^{block_2}_{\text{output}} \frac{N^{block_2}_{input}}{\sum w^{block_2}_{\text{output}}} \oplus \ldots \tag{5.13}$$

Before going further, as we use the sWeights to extract the signal component on the PHSP and the sWeights are obtained from the B mass fit, it is important to check the correlation between the PHSP variables and the B-mass. High correlation between these variables can lead to suboptimal performance of the sWeights. To assess the correlation, we compute the Pearson correlation coefficient. The highest correlations are obtained for the $\cos(\theta_{K^0_S\pi^+_1(\pi^-\pi^+_2)})$ with a correlation of approximately 5%. We consider that this level of correlation is harmless for our analysis. The correlation plot is shown in Fig. 5.20.

5.4.2 PID CORRECTION

In this correction we count for differences between data and MC on the particle identification system. With the *PIDCalib2* package one can build the data-driven correction histograms for samples that match the kinematic selection of our data, i.e. the stripping-level kinematic, topological and PID cuts. The histograms are built on the η and p of the particle's track and, N_{tracks} in Run 1 and $N_{\text{SPD Hits}}$ in Run 2 as global event variables. We compute a correction for the five final-state signal tracks, combining them to obtain a per-event correction (κ_{PID}) with the associated uncertainty. The uncertainty of each track is obtained when building the correction histograms and depends, basically, on the sample-size used to build them. From

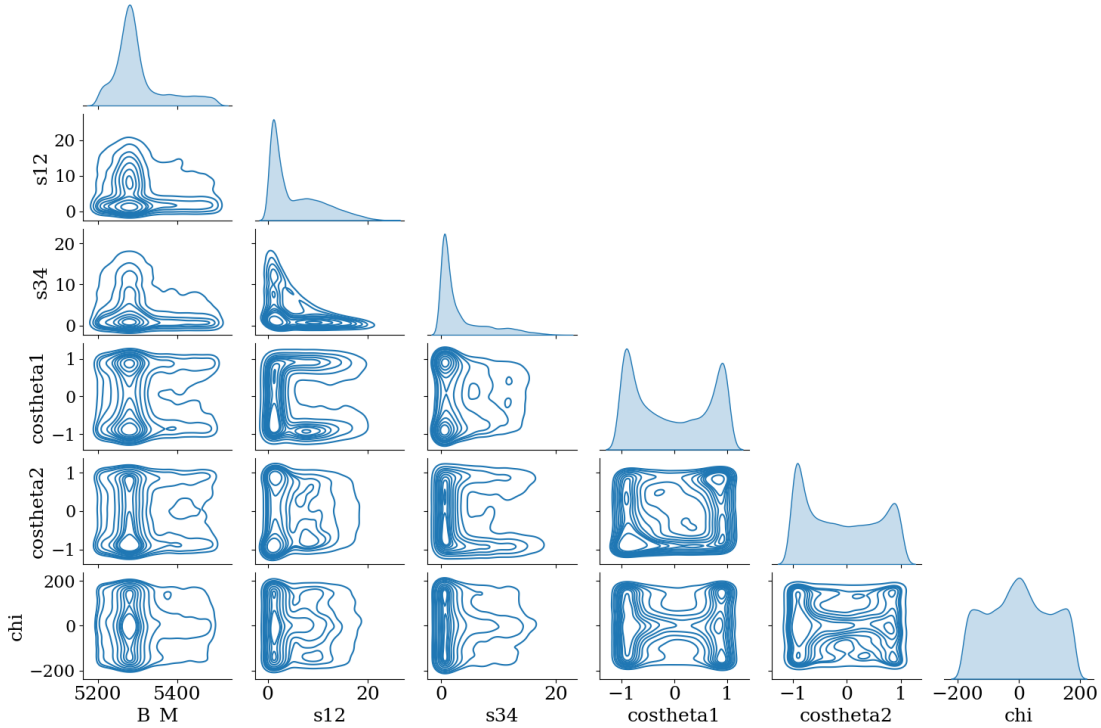


FIGURE 5.20: Correlation plot of the PHSP variables and the B-mass. The B-mass appears in the first row, its contour plot with the other variables appears in the first column. The other plots correspond to the correlation plots between the PHSP variables.

this correction, the accurate PID efficiencies of a given PID cut are obtained; therefore, the input sample MC does not have PID cuts on the variables that we want to estimate their true impact on the efficiency.

5.4.3 TRACKING CORRECTION

The tracking efficiencies are also not well simulated, we correct for any deviation with *TrackCalib2*. The package works in a similar fashion as *PIDCalib2*, but the correction histograms are 2-dimensional on η and p of each track. In this case we use the available calibration tables. We compute a per-event correction (κ_{Track}) by combining the correction and associated error for each of the five signal tracks. The uncertainties of each correction also depend on the size of the samples used to obtain the tables.

5.4.4 L0 HADRON TIS/TOS CORRECTION

The Lo system is not properly simulated for hadronic decays due to the poor knowledge on the hadronic calorimeter. Thus, in order to simulate the behaviour of the Lo hadron trigger, we use the *Lo Calo* tool. As the TOS MC sample is biased due to this effect, we use only the TIS MC sample and this tool to correct for the efficiencies. The output of the *Lo Calo* tool is the probability that the signal tracks fired the Lo Hadron trigger system (ϵ_{TOS}), it also takes into account possible overlapping in the calorimeter cells.

Moreover, it is important to remark that the TIS trigger category refers to the whole Lo system and not only the hadronic one, so we use the TIS Global trigger category in our correction. This is simply because the TIS category can be fired by any track present in the $p-p$ event that does not belong to the purely hadronic signal. Therefore, in this correction we are taking into account two different effects. We must correct the impact on the efficiency of the Lo hadronic trigger. But we also need to scale the amount of each trigger category to match the amount of TIS and TOS data events. The Lo trigger correction impacts both categories due to the overlap between them, but the PHSP efficiency distribution of the TIS sample is not biased by this effect. Thus, we use the TIS category to calibrate the TOS one.

One can define the following samples using the *Lo Calo* tool on the TIS sample:

- TIS weighted by ϵ_{TOS}: representing the TIS and TOS sample, i.e. the overlap between both categories. Fig. 5.21 and Fig. 5.22 show the performance of the tool.

- TIS weighted by $1-\epsilon_{TOS}$: TIS not TOS sample. That represents the TIS sample excluding the overlap with the TOS sample.

We will use the TIS and TOS sample to correct for the whole TOS. So there are two different corrections, one for the TOS sample and other for the TIS not TOS, joining the two samples we obtain the full data sample.

We use the GBrw as explained in Sec. 5.4.1 to correct the PHSP distribution of the MC with the output of this tool. For the TOS sample (the superscript refers to the sample that is being copied):

1. Copy the TIS and TOS distribution to the TIS sample:

$$GBrw(\text{TIS and TOS};\text{TIS}) \rightarrow w_{\text{TIS}}^{\text{TIS and TOS}} .$$

2. Apply the correction $\epsilon_{\text{TOS}}/w_{\text{TIS}}^{\text{TIS and TOS}}$ to the TOS sample:

$$GBrw(\text{TIS}, w=\epsilon_{\text{TOS}}/w_{\text{TIS}}^{\text{TIS and TOS}};\text{TOS}) \rightarrow \kappa_{\text{TOS}} .$$

In this way we have created in the TIS sample, the correction factor for the TOS sample using the TIS and TOS sample and the *Lo Calo* tool. Finally we evaluate this correction on the TOS sample obtaining the desired correction (κ_{TOS}).

What remains is the TIS not TOS sample, so we proceed in a similar way for this sample obtaining a correction also here:

1. Copy the TIS not TOS distribution to the TIS sample:

$$GBrw(\text{TIS not TOS};\text{TIS}) \rightarrow w_{\text{TIS}}^{\text{TIS not TOS}}$$

2. Apply the correction $(1-\epsilon_{\text{TOS}})/w_{\text{TIS}}^{\text{TIS not TOS}}$ to the TIS not TOS sample:

$$GBrw(TIS, w=(1-\epsilon_{\text{TOS}})/w_{\text{TIS}}^{\text{TIS not TOS}};\text{TIS not TOS}) \rightarrow \kappa_{\text{TIS not TOS}}$$

Finally, we need to scale each correction to join the TOS, and, TIS not TOS samples back together, obtaining thus the TIS/TOS correction ($\kappa_{\text{TIS/TOS}}$). We scale each correction with respect to the amount in data of each category:

$$\kappa_{\text{TIS/TOS}} = \frac{f_{\text{TOS}}^{\text{data}}}{f^{\text{data}}} \times \frac{\kappa_{\text{TOS}}}{\sum(\kappa_{\text{TOS}})} \oplus \frac{f_{\text{TIS not TOS}}^{\text{data}}}{f^{\text{data}}} \times \frac{\kappa_{\text{TIS not TOS}}}{\sum(\kappa_{\text{TIS not TOS}})} . \tag{5.14}$$

It is important to remark that we cannot appreciate any change on the distributions after applying this correction, as shown in Figs. 5.23 and 5.24. Also, when using the *Lo Calo* tool we see very small differences. We would need a bigger MC sample to appreciate the effect of this correction.

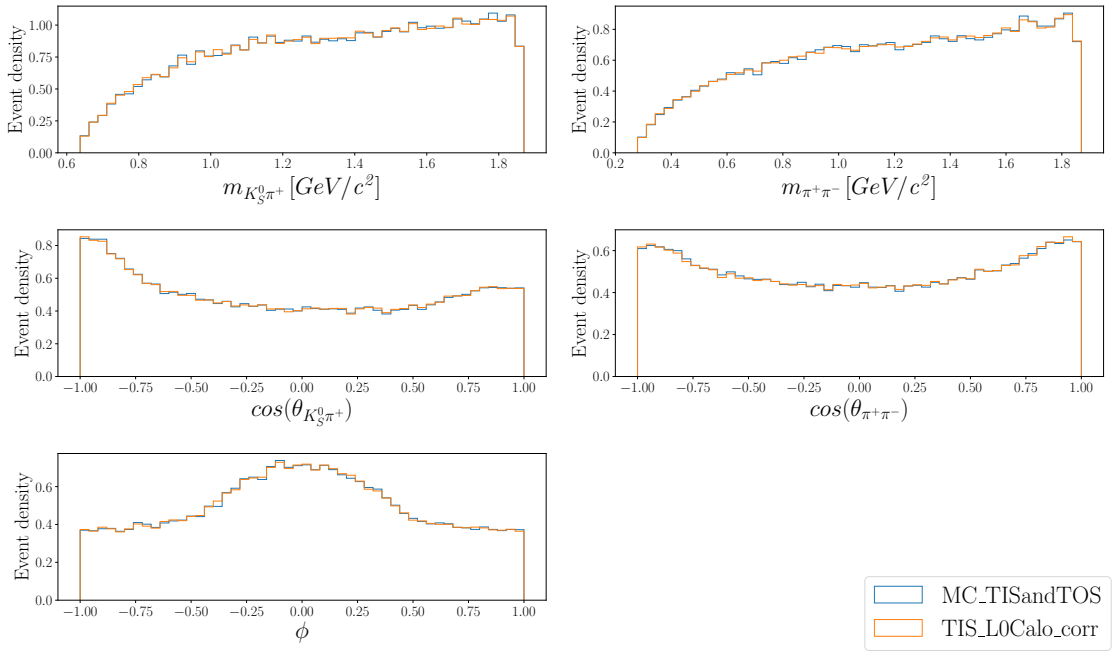
5.4.5 AMPLITUDES CORRECTION

Now that we have corrected our MC by the main effects that have an impact on the efficiency, we can use it to extract the efficiency distribution of data across the PHSP ($\epsilon_{data}(\Phi_4) \approx \epsilon_{MC}(\Phi_4) \equiv \epsilon(\Phi_4)$). From data sWeights we can extract the amplitudes times the efficiency ($\epsilon(\Phi_4)\,|A(\Phi_4)|^2$). Having thus all the ingredients to extract the efficiency of data. This correction will only be applied to the signal channel, for the control channel the amplitudes at generator level are already known.

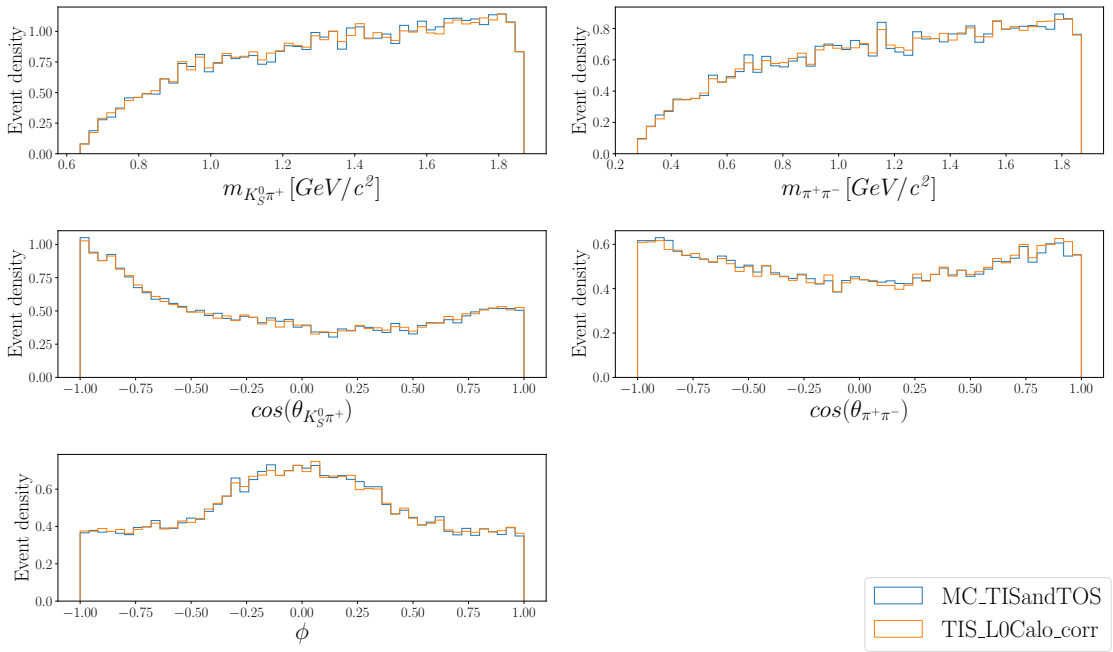
The goal of this correction is to obtain how data-amplitudes would be distributed at generator level, i.e. $|A(\Phi_4)|^2$. In order to obtain the amplitudes at generator level we simulate the signal MC (flat PHSP in the analysis mass window) at generator level, we use a sample with $1M$ events. This sample represents the identity ($\mathbb{1}(\Phi_4)$) and will be used as a blackboard to copy the features that we need for the correction. We consider the flat PHSP sample as the identity because it is our reference distribution, but this sample follows the four-body PHSP distribution defined by:

$$d\Phi_4 = \delta(P-\sum_i^4 p_i)\prod_i^4 \frac{d^4p_i}{(2\pi)^2}\delta(p_i^2-m_i^2) . \tag{5.15}$$

Where P refers to parent's momentum, p_i to the children's momentum and δ is the delta function. Summarising, to obtain the efficiency value of the real data we need the following ingredients:



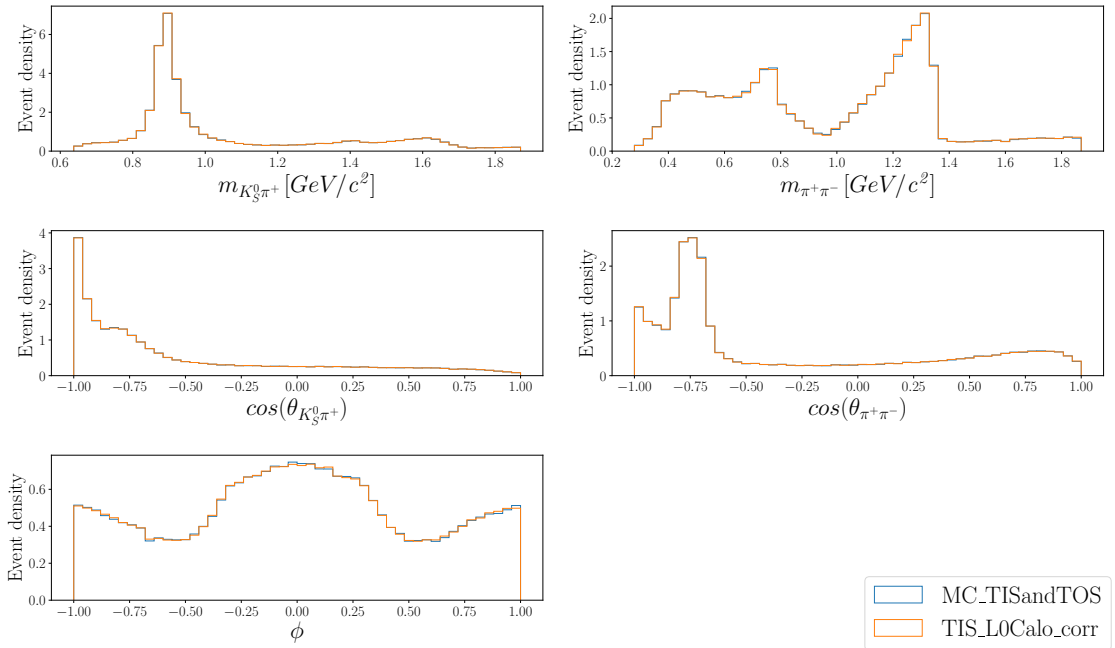
(A) DD



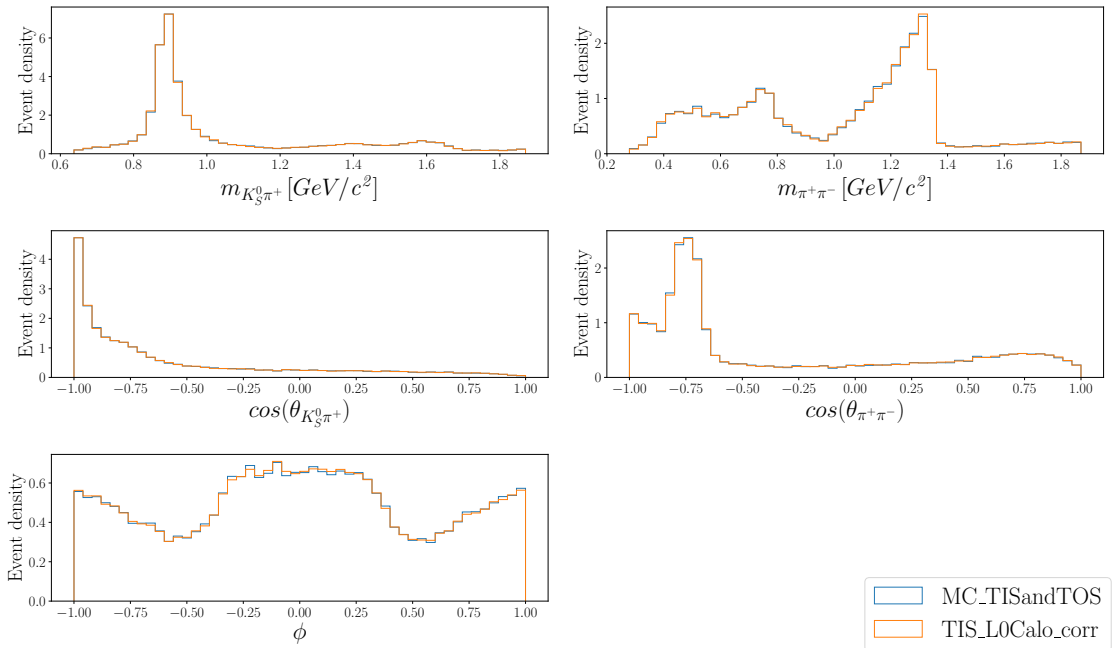
(B) LL

FIGURE 5.21: Performance of *Lo Calo* tool on the GBrw space: MC TIS weighted by ϵ_{TOS} in orange on top of the TIS and TOS MC sample in blue for the signal channel. Showing only the 1718 sample for clarity.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



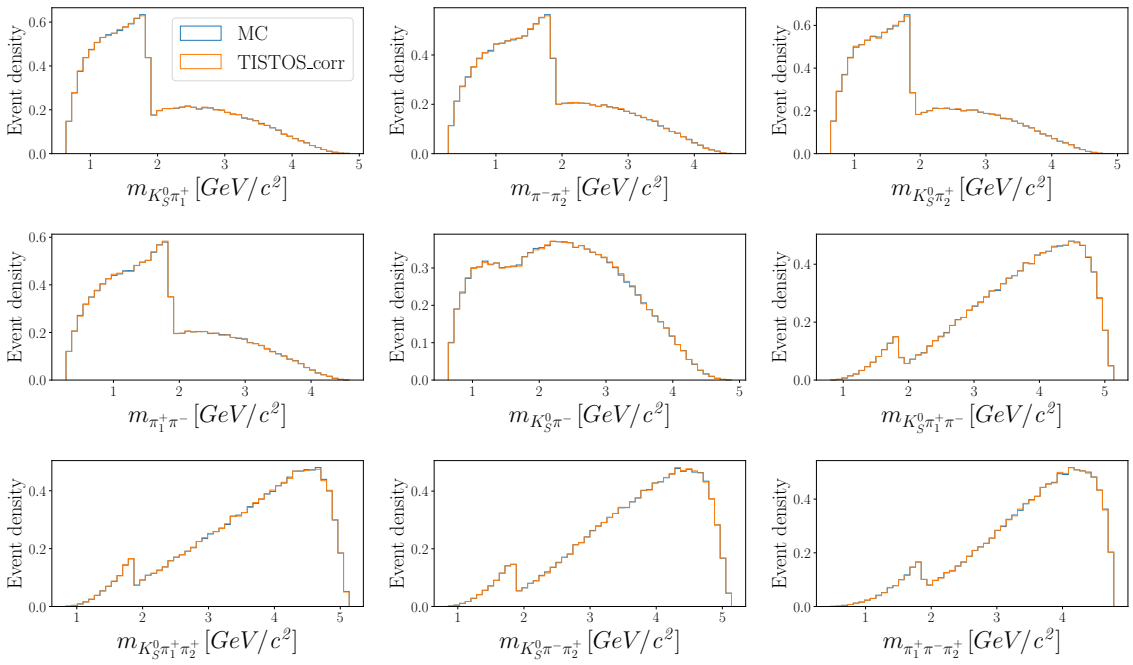
(A) DD



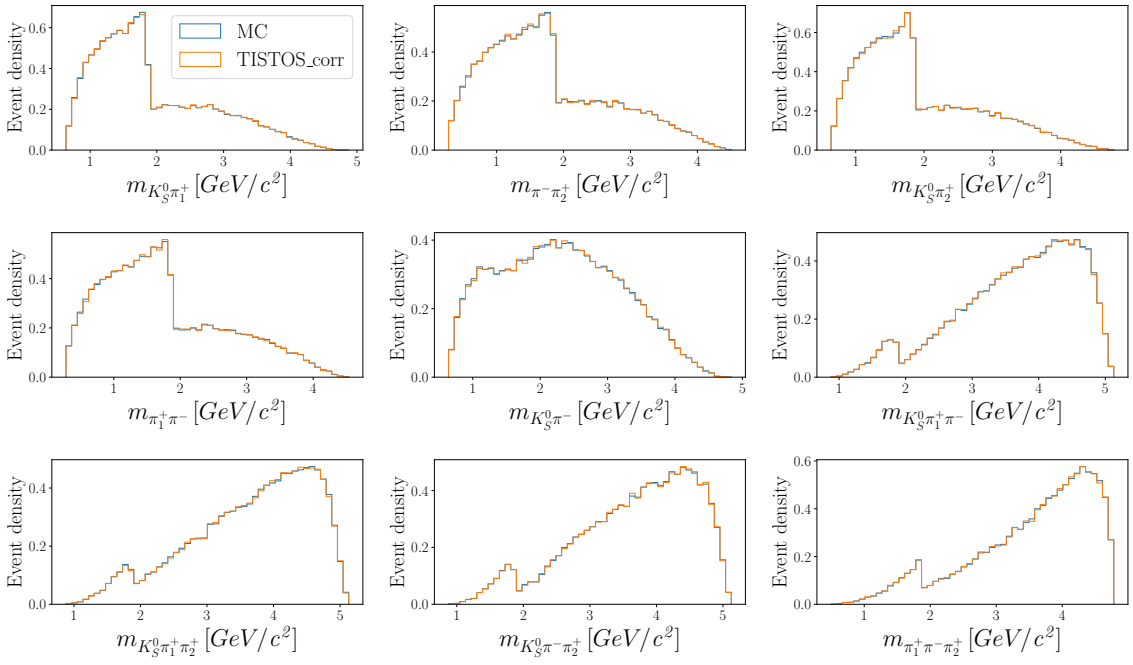
(B) LL

FIGURE 5.22: Performance of *Lo Calo* tool on the GBrw space: MC TIS weighted by ϵ_{TOS} in orange on top of the TIS and TOS MC sample in blue for the control channel. Showing only the 1718 sample for clarity.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



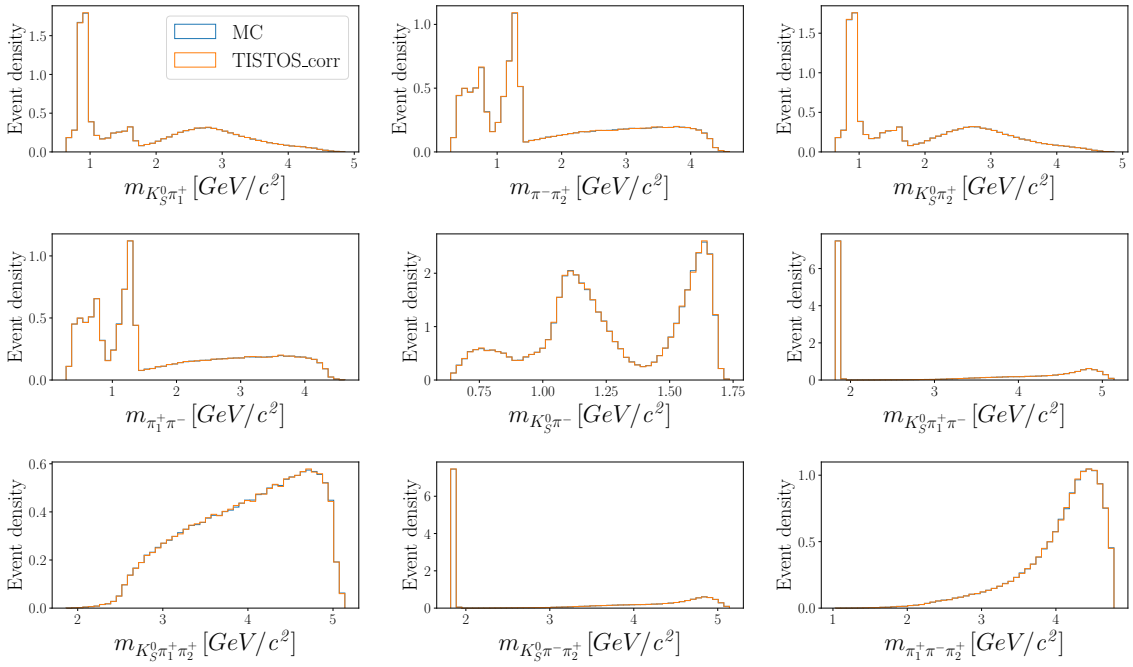
(A) DD



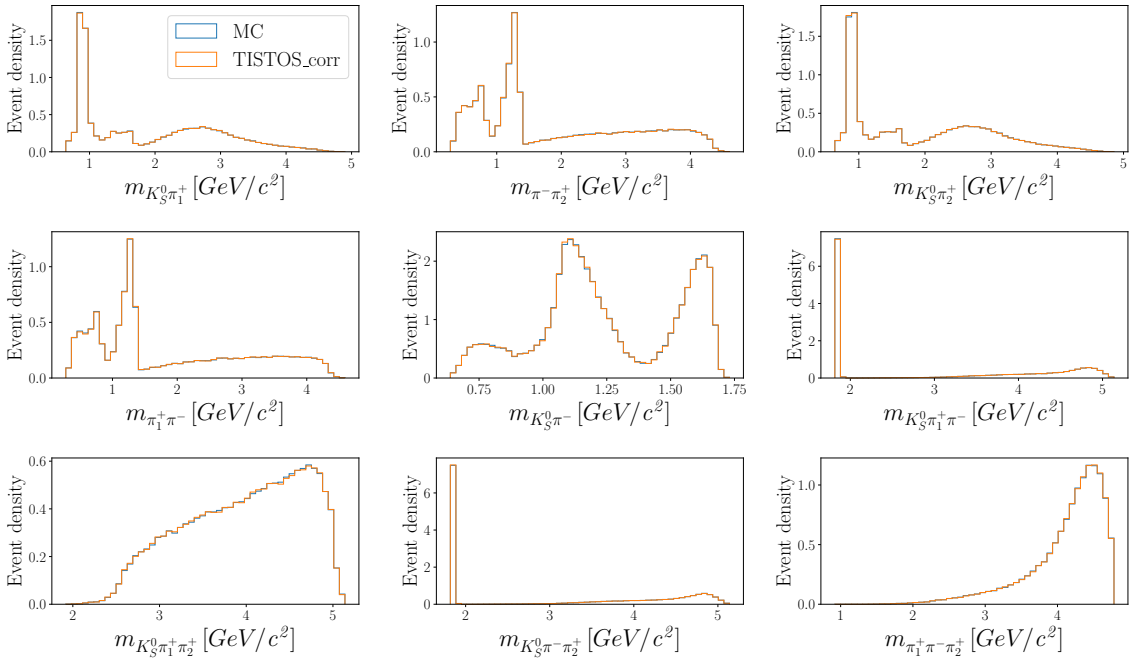
(B) LL

FIGURE 5.23: Performance of TIS/TOS correction for many mass projections of the PHSP: MC before (in blue) and after the TIS/TOS correction (in orange) for the signal channel are shown. No sizable effect can be appreciated. Showing only the 1718 sample for clarity.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) DD



(B) LL

FIGURE 5.24: Performance of TIS/TOS correction for many mass projections of the PHSP: MC before (in blue) and after the TIS/TOS correction (in orange) for the control channel are shown. No sizable effect can be appreciated. Showing only the 1718 sample for clarity.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

- The efficiency of each event on the sample at generator level properly scaled by the efficiency value from the LHCb MC.
- The weights on the sample at generator level that describe the signal amplitudes in data without the efficiency effect, i.e. the amplitudes at generator level.

These features will be obtained using the aforementioned samples which PDFs describe the following distributions:

- LHCb MC: $\mathcal{P}_{\text{MC}}(\Phi_4) = \epsilon(\Phi_4)$.
- Data sWeights: $\mathcal{P}_{\text{data}}(\Phi_4) = \epsilon(\Phi_4)\,|A(\Phi_4)|^2$.
- Generator level MC: $\mathcal{P}_{\text{gen}}(\Phi_4) = \mathbb{1}(\Phi_4)$.

It is important to highlight that we have presumed the detector resolution does not affect the extraction of amplitudes at the generator level. This presumption is typically accurate since the resonances under investigation possess a natural width that exceeds the detector's resolution. However, this assumption does not hold for the $\omega - \rho$ interference characterisation, as the interference occurs within a narrower region of the PHSP than the detector resolution. To accurately parameterise this line-shape, we should consider the resolution effects, given that the PDF extracted from the data sWeights is: $\epsilon(\Phi_4)|A(\Phi_4)|^2 \otimes R(\Phi'_4 - \Phi_4)$, where R represents the resolution. Consequently, we should also parameterise the resolution effects and remove them from the real data sample to obtain the generator-level amplitudes. Nevertheless, in our modeling procedure, data statistics is a limiting factor. Therefore, to decide whether to account for the resolution effects, we must assess if the data statistics are sufficient to observe the $\omega - \rho$ interference. Based on various plots of the $\pi^+\pi^-$ spectrum, we consider this a borderline case, as the interference is noticeable in some samples but not in others. Lastly, since the amplitude correction is a second-order effect that multiplies the efficiency weights, we will not include the resolution effects in our amplitude characterisation.

We use the GBrw as described in Sec. 5.4.1 to extract each feature and copy it to the sample at generator level, obtaining the weights that transform the sample at generator level to look like the input data. In Eq. 5.12 the different mass blocks utilised in the re-weighting process are shown.

In order to obtain the efficiency and the amplitudes separately on the sample at generator level we do the following:

1. Evaluate the GBrw on the LHCb MC as input, applying the efficiency correction weights ($\kappa_{\text{MC}} = \kappa_{\text{TIS/TOS}} \times \kappa_{\text{Track}} \times \kappa_{\text{PID}}$) obtaining thus the efficiency weights (w_ϵ) on the sample at generator level, the performance of this re-weighting is shown in Fig. 5.28:

$$GBrw(\text{MC}, w = \kappa_{\text{MC}};\, \text{gen}) \to w'_\epsilon\,.$$

   It is important to note that, to obtain the efficiency weights properly scaled, we must multiply them by the efficiency of the LHCb MC, that is $\epsilon_{\text{LHCb MC}} = \frac{\sum \kappa_{\text{MC}}}{N_{\text{gen}}}$, where N_{gen} is the number of LHCb MC events at generator level, this efficiency depends on the sample and is around the $10^{-4} - 10^{-5}$ level. The values for the different samples are shown in Table 5.6.

$$w_\epsilon = w'_\epsilon \times \frac{\epsilon_{\text{LHCb MC}}}{\frac{\sum w'_\epsilon}{N}}\,,$$

   where N is the number of events in the generator-level MC sample.
2. Evaluate the GBrw on the data weighted by the sWeights, obtaining the amplitudes times the efficiency ($w_{\epsilon|A|^2}$) on the sample at generator level, the performance of this re-weighting is shown in Fig. 5.29:

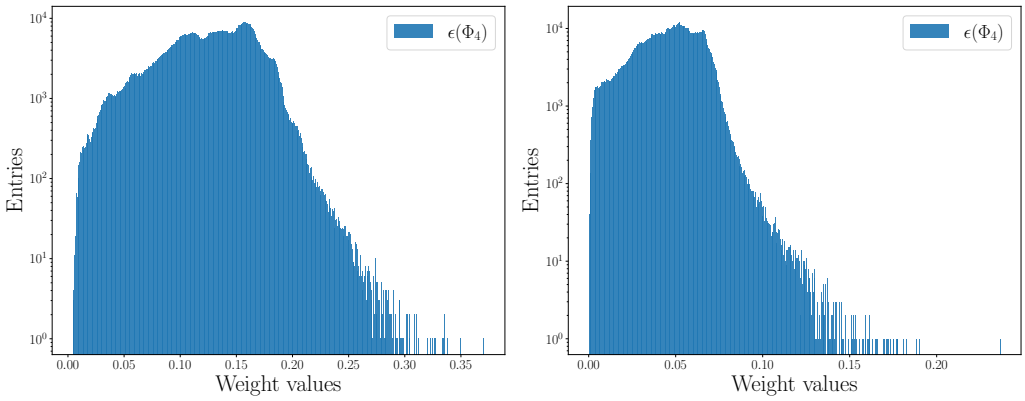
$$GBrw(\text{data}, w = \text{sW};\, \text{gen}) \to w_{\epsilon|A|^2}\,.$$

3. Divide both sets of weights to obtain the amplitudes at generator level ($w_{|A|^2}$), by subtracting the impact of the efficiency on the sWeighted data. Fig. 5.30 shows how the amplitudes are distributed at generator level:

$$w_{|A|^2} = \frac{w_{\epsilon|A|^2}}{w_\epsilon}\,.$$

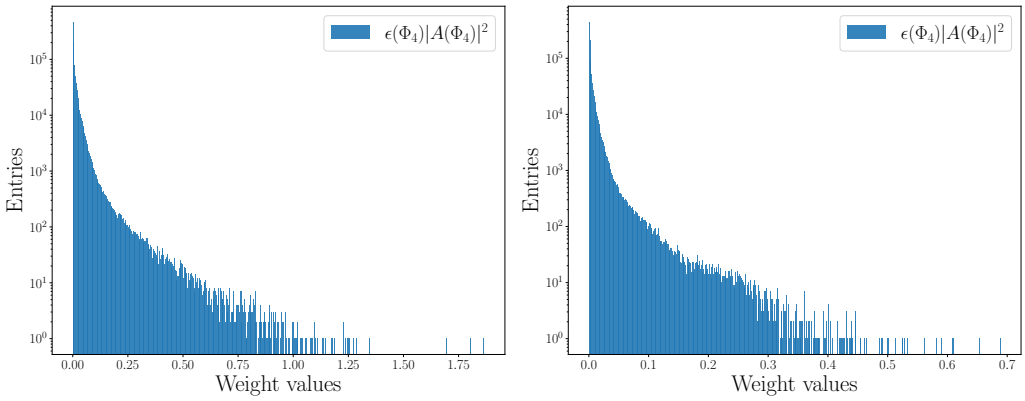
   Note that we are dividing by the efficiency weights that we got from MC, we must be sure that there is no value $w_\epsilon = 0$ so that we have infinite values on $w_{|A|^2}$. We use a set of hyper-parameters on the GBrw that ensure a smooth behaviour of these weights. However, for the amplitudes extraction, as there are more sharp structures, we need a tighter set of hyper-parameters that ensures that we are properly capturing the resonances. Fig. 5.25 shows the efficiency weights, Fig. 5.26 shows

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) DD (B) LL

FIGURE 5.25: Distribution of the efficiency weights w_ϵ on the sample at generator level that match the efficiency on the LHCb MC sample. There are values close to 0 but there are no values $w_\epsilon = 0$ that could retrieve infinite values in the amplitudes at generator level. Showing only the 1718 sample for clarity.



(A) DD (B) LL

FIGURE 5.26: Distribution of the amplitudes times efficiency weights $w_{\epsilon|A|^2}$ on the sample at generator level that match the sWeighted data. Showing only the 1718 sample for clarity.

the weight distribution extracted from the sWeights, i.e. $w_{\epsilon|A|^2}$ and Fig. 5.27 shows the weights for the amplitudes at generator level $w_{|A|^2}$. In the amplitudes at generator level in Fig. 5.30, some fluctuations and differences between the DD and LL samples can be appreciated. These differences arise as a consequence of the limited sample sizes, the high-dimensional training space, and the sharp distributions that are being copied. In Sec. 5.5, systematic uncertainties due to the size of these samples will be computed.

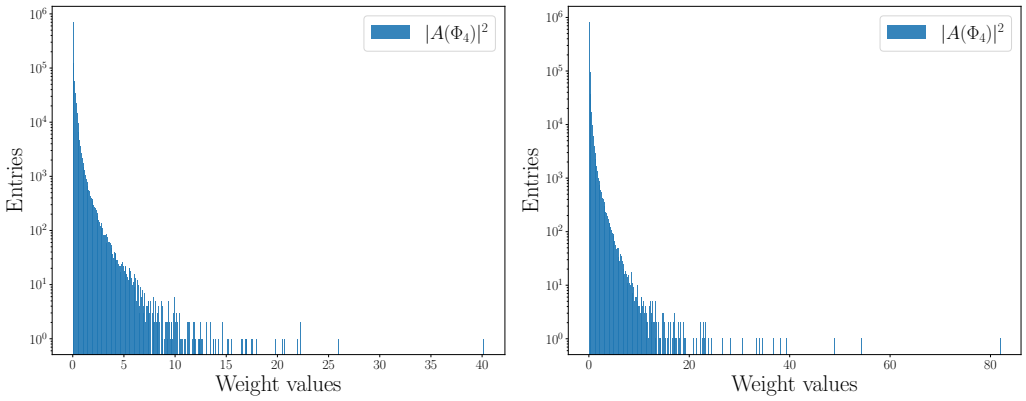
To determine the data efficiency, we need to integrate the product of efficiency and amplitude distributions over the PHSP range. These distributions are parameterised as weights in the sample at generator level.

To obtain a meaningful value from this integration, which reflects the true efficiency, the efficiency distribution must be scaled to represent the global efficiency value, not only the PHSP dependency. Remember that we scaled the efficiency weights in the generator-level sample so that its efficiency equals the efficiency of the LHCb MC sample.

Next, the generator-level signal amplitude distribution must be normalised by dividing by the sum of the weights during the integration. Since we have discrete values, this integral can be performed as a weighted average of the efficiency $(\epsilon(\Phi_4), w_\epsilon)$ by the generator-level amplitudes $(|A(\Phi_4)|^2, w_{|A|^2})$:

$$\epsilon_{\text{signal}} = \int_{\text{Mass Window}} |A(\Phi_4)|^2 \cdot \epsilon(\Phi_4)\ d\Phi_4\,, \tag{5.16}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) DD (B) LL

FIGURE 5.27: Distribution of the amplitudes weights $w_{|A|^2}$ on the sample at generator level that represent the amplitudes at generator level. Showing only the 1718 sample for clarity.

or in terms of weights in the sample at generator level:

$$\epsilon_{\text{signal}} = \frac{1}{\sum w_{|A|^2}} \sum_{i=1}^{N} \left(w_{|A|^2,\, i} \cdot w_{\epsilon,\, i} \right) , \tag{5.17}$$

being N the number of events in the generator-level MC sample.

In conclusion, to show the performance of the method, in Figs. 5.31 and 5.32 we compare the PHSP distributions of the different PDFs. On one side, the amplitudes at generator level and the sWeighted data are shown. On the other side, the flat PHSP sample at generator level and the LHCb MC sample are shown. The distributions on each side differ due to the efficiency effect. Therefore, by examining these plots, we can evaluate the method's performance, demonstrating that we can remove the efficiency effect from the real data distribution. This allows us to obtain the signal amplitudes at generator level and accurately compute the real data efficiency.

5.4.6 RESULTS

We show now the results of the efficiency calculation for the signal and control channels.

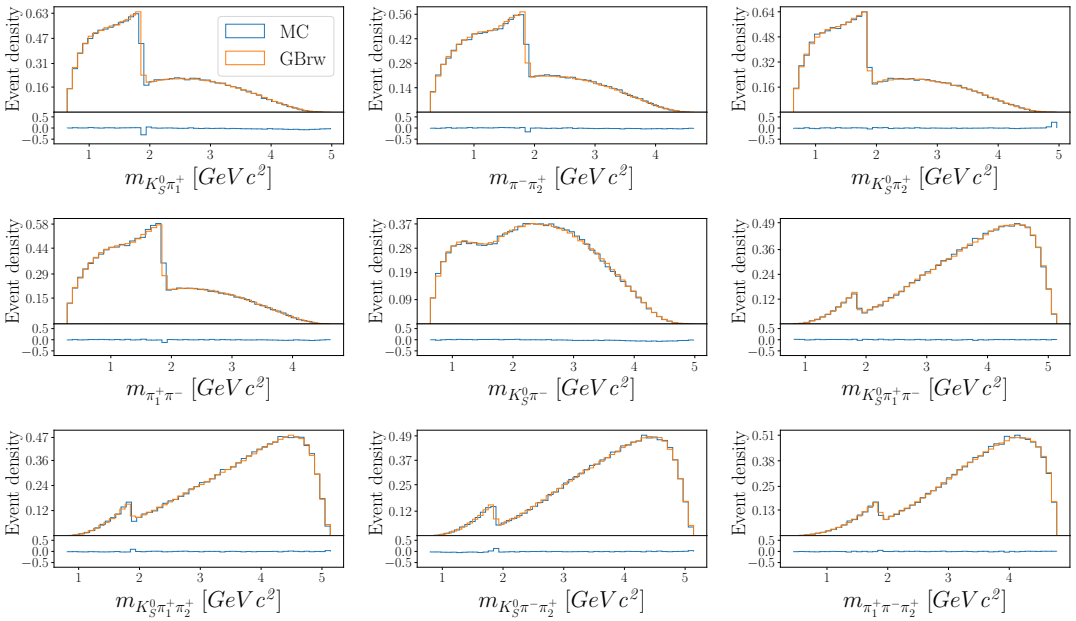
5.4.6.1 SIGNAL CHANNEL

We use the output from the mass fit to the narrow signal region where we obtain the sWeights to compute the amplitudes-corrected efficiency. As we are using the data to obtain the signal amplitudes, we do not implement further corrections to MC besides the PID, tracking and Lo correction. In conclusion, we obtain the efficiency as described in Sec. 5.4.5.

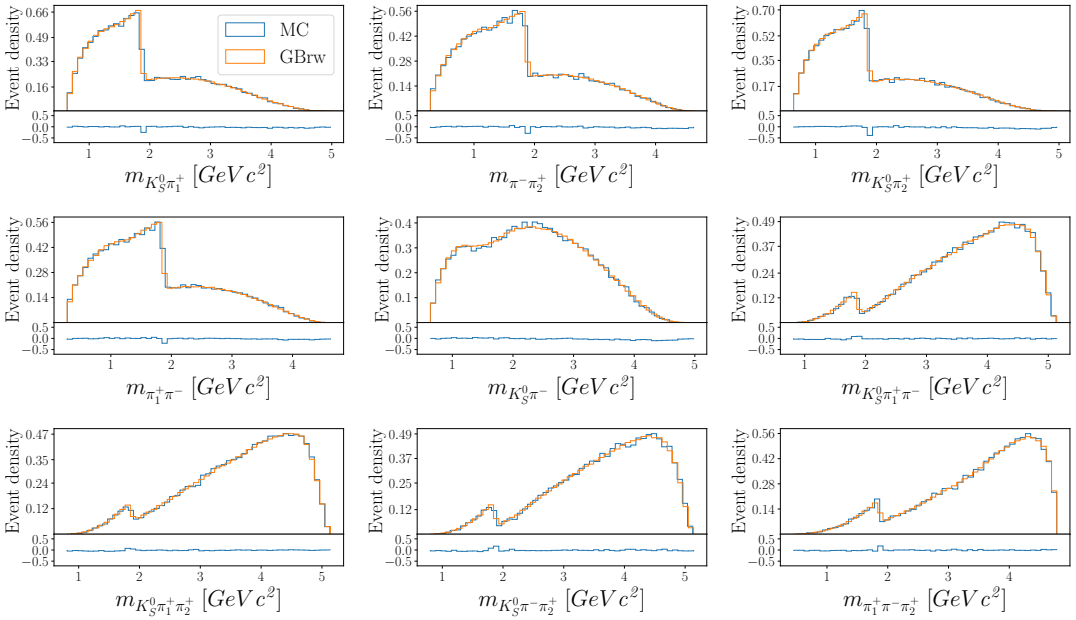
5.4.6.2 CONTROL CHANNEL

We use the output from the mass fit to the wide region, we also compute the sWeights from this mass fit. We do not repeat the mass fit on the signal region because we do not use the sWeights to re-weight the PHSP projections, we use them to re-weight the BDT variables and the p and η of the child particles which are less correlated with the B-mass. In Fig. 5.33 and Fig. 5.34 we can see some of the BDT variables before and after re-weighting on top of the sWeighted variables.

The efficiency is obtained from the simulated MC samples; after passing these samples though the same selection as data, several MC distributions are re-weighted against sWeighted data to ensure an accurate extraction of the efficiencies from the simulation.
Firstly, we re-weight the variables to be used in the three efficiency corrections, ensuring an accurate impact of these corrections. These variables are the η and p of the five children particles. Then, we apply the three efficiency corrections.



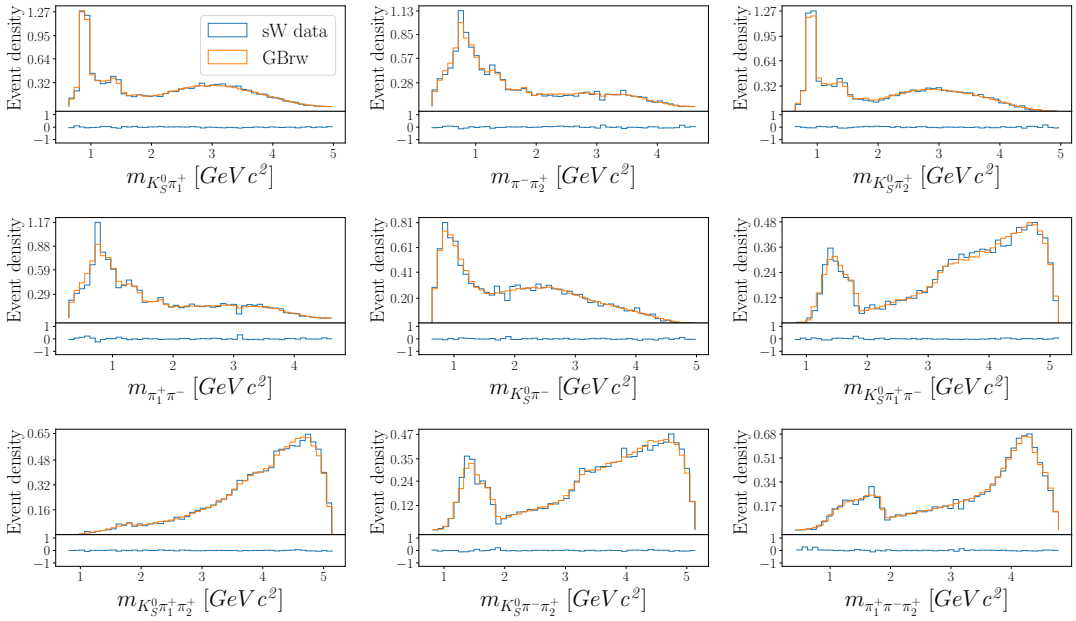
(A) DD



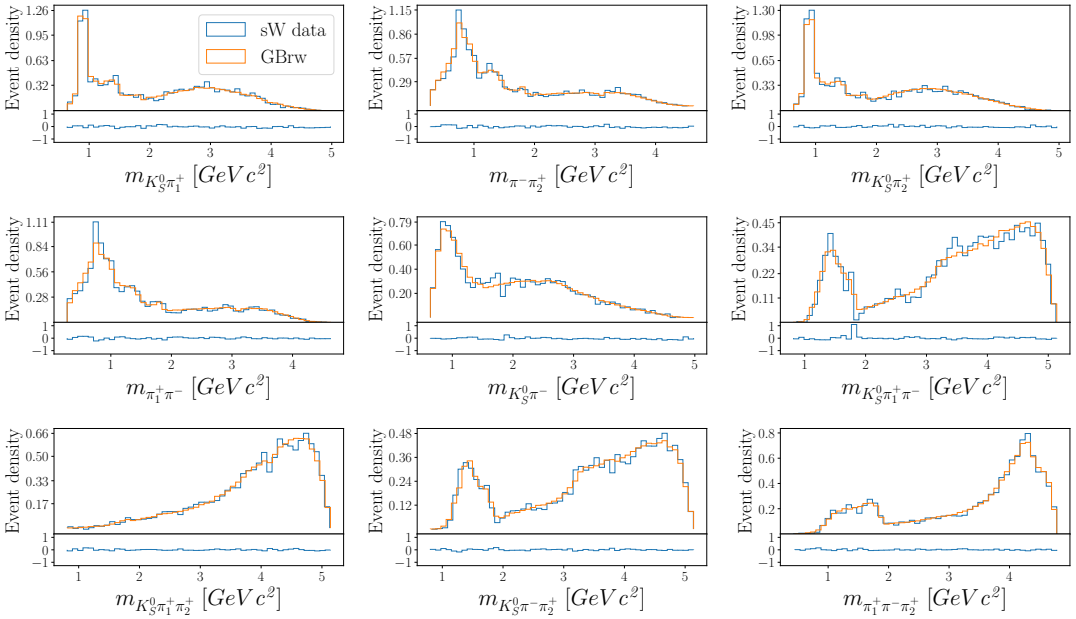
(B) LL

FIGURE 5.28: Performance of the GBrw to extract the efficiency ($\epsilon(\Phi_4)$) on the sample at generator level on some mass projections of the PHSP. The plots show the efficiency corrected LHCb MC in blue on top of the sample at generator level weighted by w_ϵ in orange. Showing only the 1718 sample for clarity.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



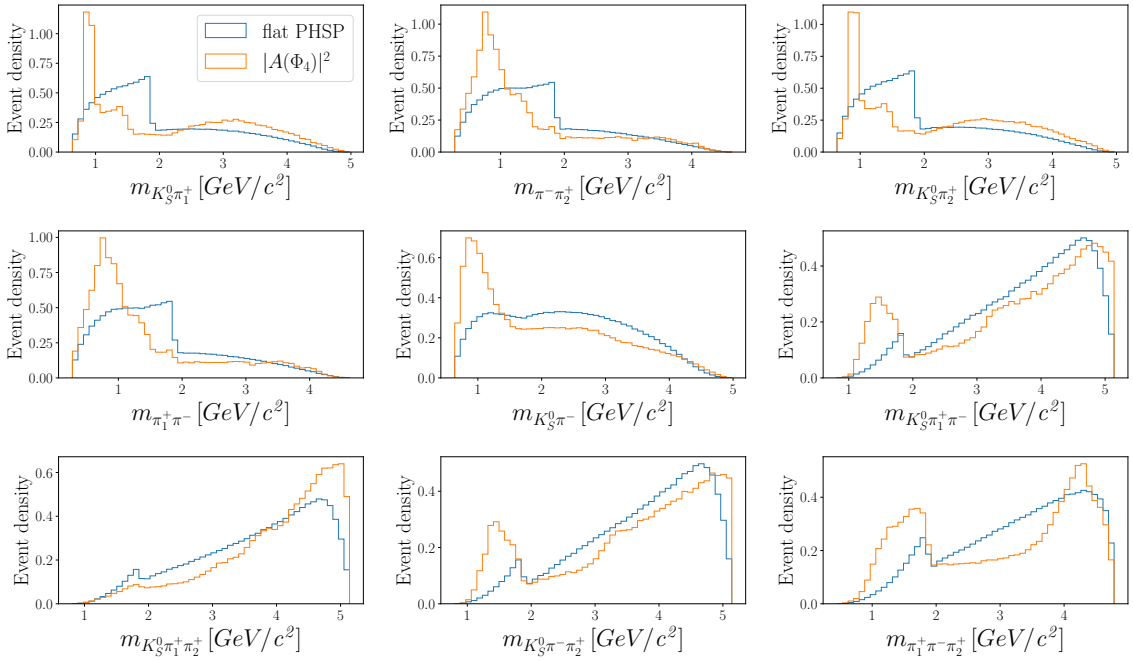
(A) DD



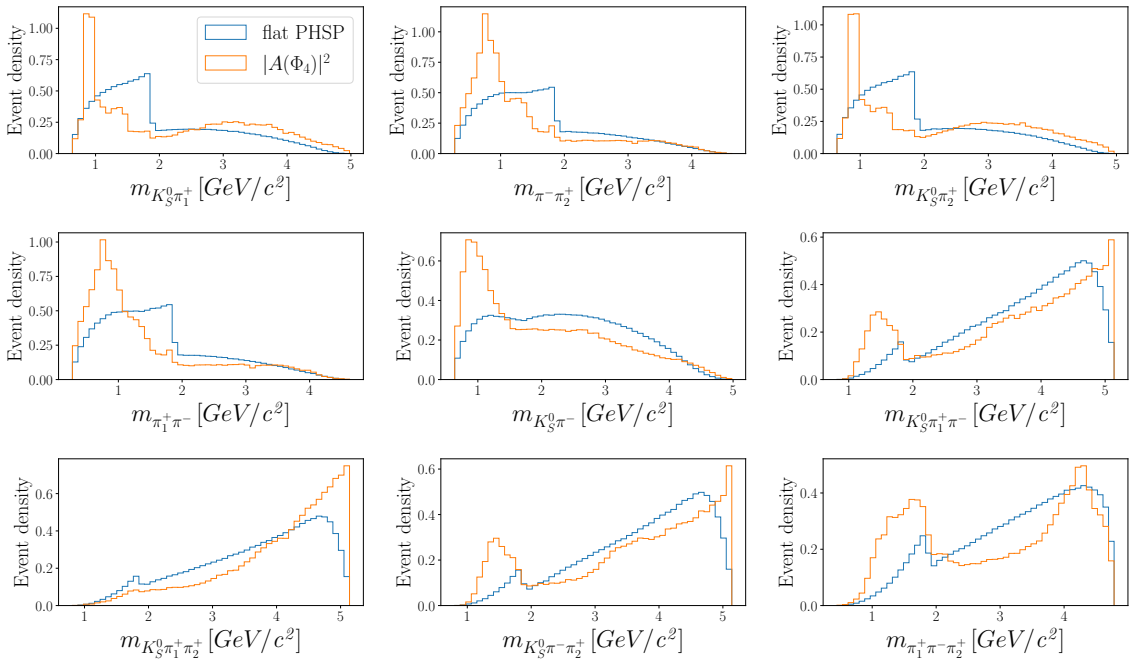
(B) LL

FIGURE 5.29: Performance of the GBrw to extract the efficiency times the amplitudes ($\epsilon(\Phi_4)|A(\Phi_4)|^2$) on the sample at generator level on some mass projections of the PHSP. The plots show the sWeighted data in blue on top of the sample at generator level weighted by $w_{\epsilon|A|^2}$ in orange. Showing only the 1718 sample for clarity.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



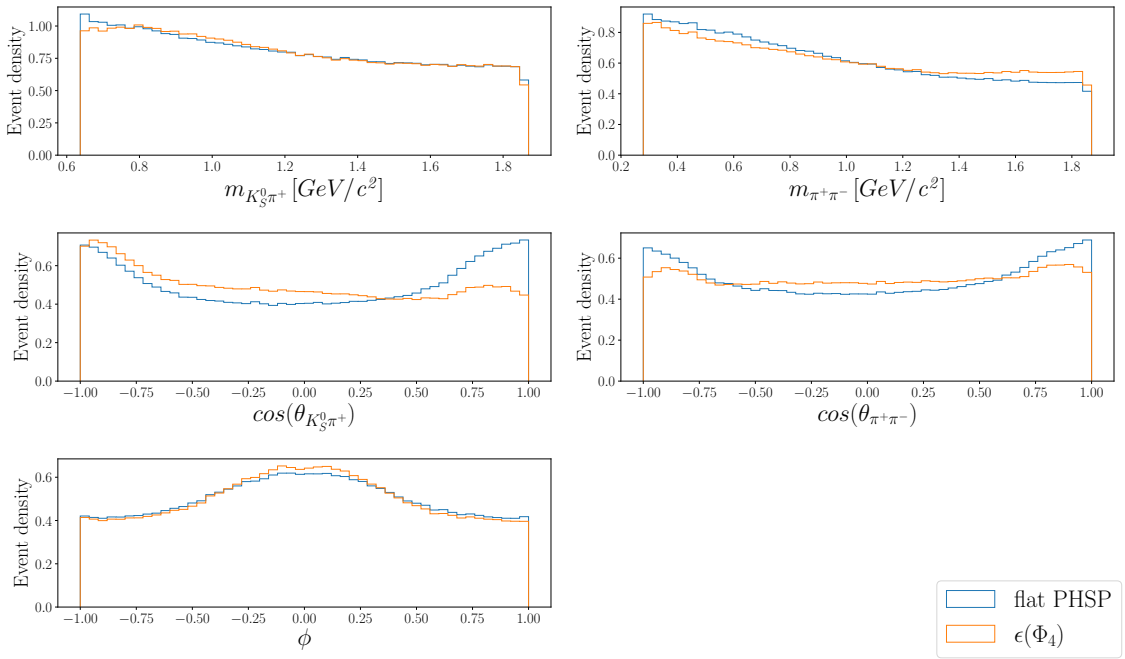
(A) DD



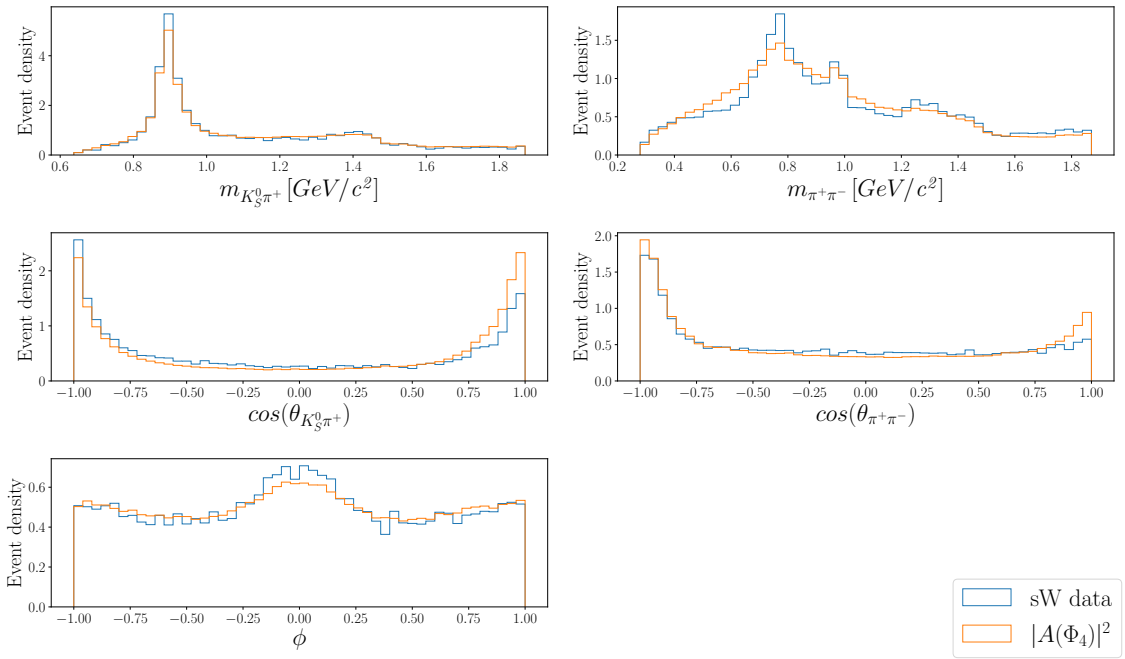
(B) LL

FIGURE 5.30: Amplitudes distribution at generator level ($|A(\Phi_4)|^2$) on some mass projections of the PHSP. The plots show the sample at generator level weighted by $w_{|A|^2}$ in orange on top of the same sample with no weights in blue. This plot highlights the performance of the method, showing that we can undo the efficiency effects in data obtaining the amplitudes at generator level. Showing only the 1718 sample for clarity.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



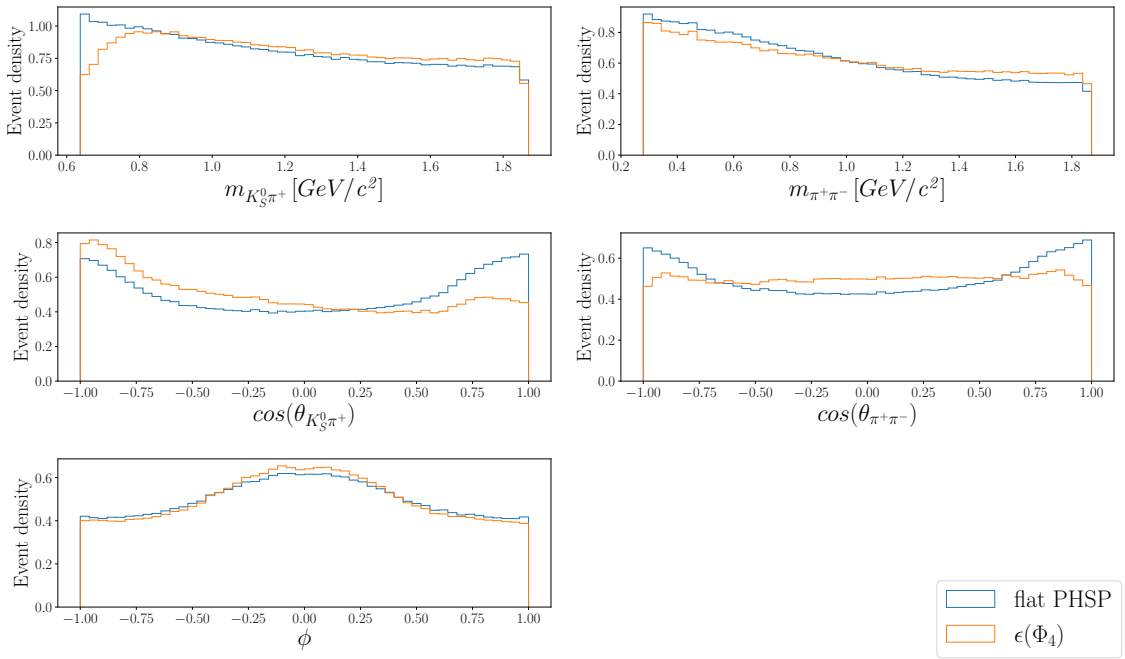
(A) Flat PHSP sample at generator level, in blue with no weights applied and, in orange, with the efficiency weights applied matching the LHCb MC efficiency distribution.



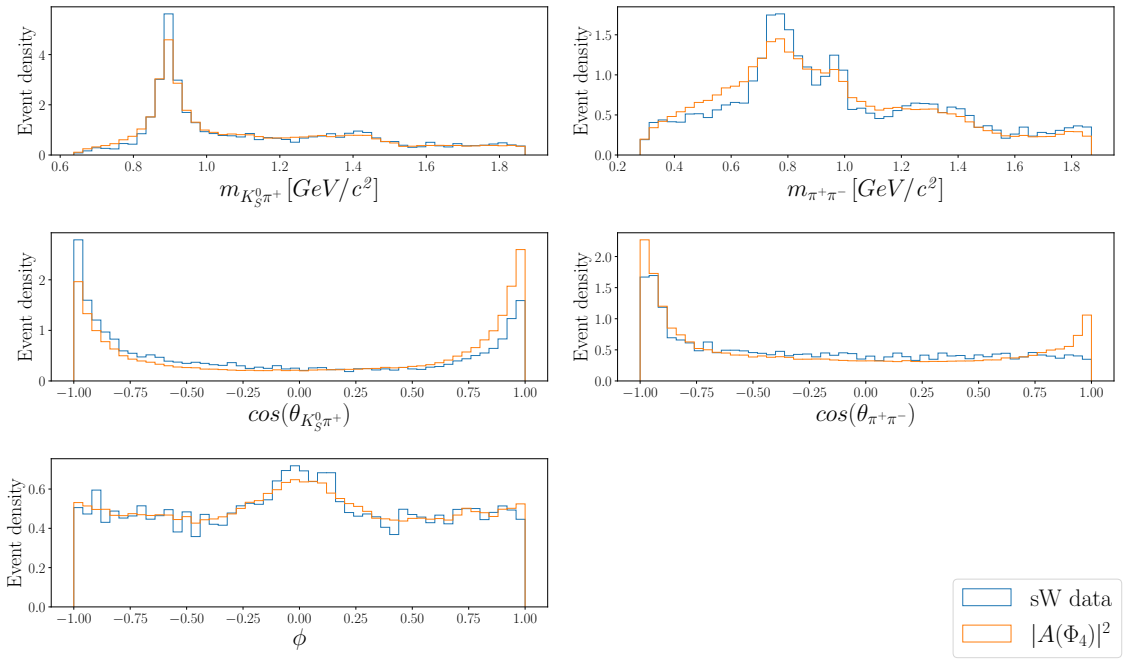
(B) In blue the sWeighted data are plotted, in orange the sample at generator level weighted by the amplitudes with no efficiency effects, the difference between these samples is the efficiency distribution.

FIGURE 5.31: A comparison between the samples with and without efficiency effects, showing that we can recover the amplitudes at generator level. The corrected value of the efficiency is the weighted average of the efficiency by the amplitudes at generator level, both in orange in these plots. The plot shows the 1718 DD sample on the GBrw training space.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



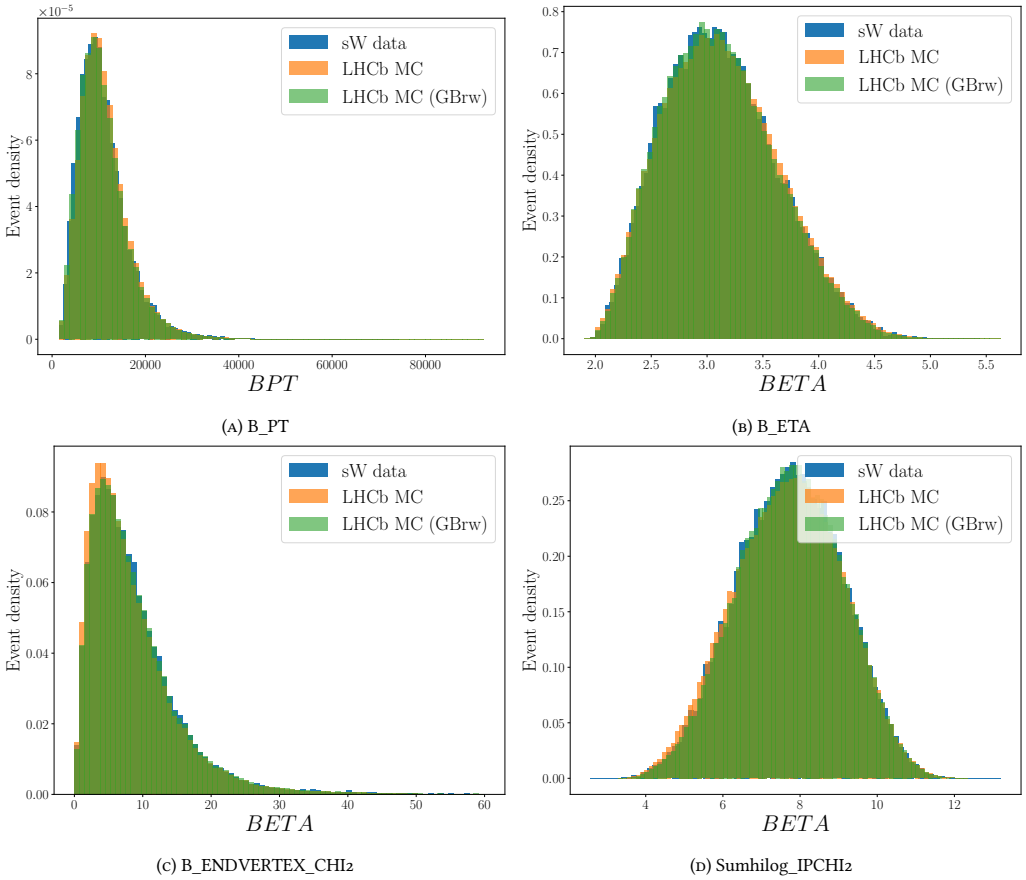
(A) Flat PHSP sample at generator level, in blue with no weights applied and, in orange, with the efficiency weights applied matching the LHCb MC efficiency distribution.



(B) In blue the sWeighted data are plotted, in orange the sample at generator level weighted by the amplitudes with no efficiency effects, the difference between these samples is the efficiency distribution.

FIGURE 5.32: A comparison between the samples with and without efficiency effects, showing that we can recover the amplitudes at generator level. The corrected value of the efficiency is the weighted average of the efficiency by the amplitudes at generator level, both in orange in these plots. The plot shows the 1718 LL sample on the GBrw training space.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) B_PT

(B) B_ETA

(C) B_ENDVERTEX_CHI2

(D) Sumhilog_IPCHI2

FIGURE 5.33: Some BDT variables on the control channel for sWeighted data and MC before and after BDT re-weighting. The plot shows the DD K^0_S-reco sample. In blue the sWeighted data are shown, in orange the MC with no re-weighting applied and in green the MC after the BDT variables re-weighting. Showing only the 1718 sample for clarity.

Finally, we re-weight the BDT variables against the sWeighted data, ensuring that the impact of the BDT selection is accurately described by the MC sample. The two re-weightings are performed with the GBrw algorithm but using the respective variables to be re-weighted:

$$GBrw(\text{data}, w = \text{sW};\ \text{MC}) \to w_{\text{MC}}\,.$$

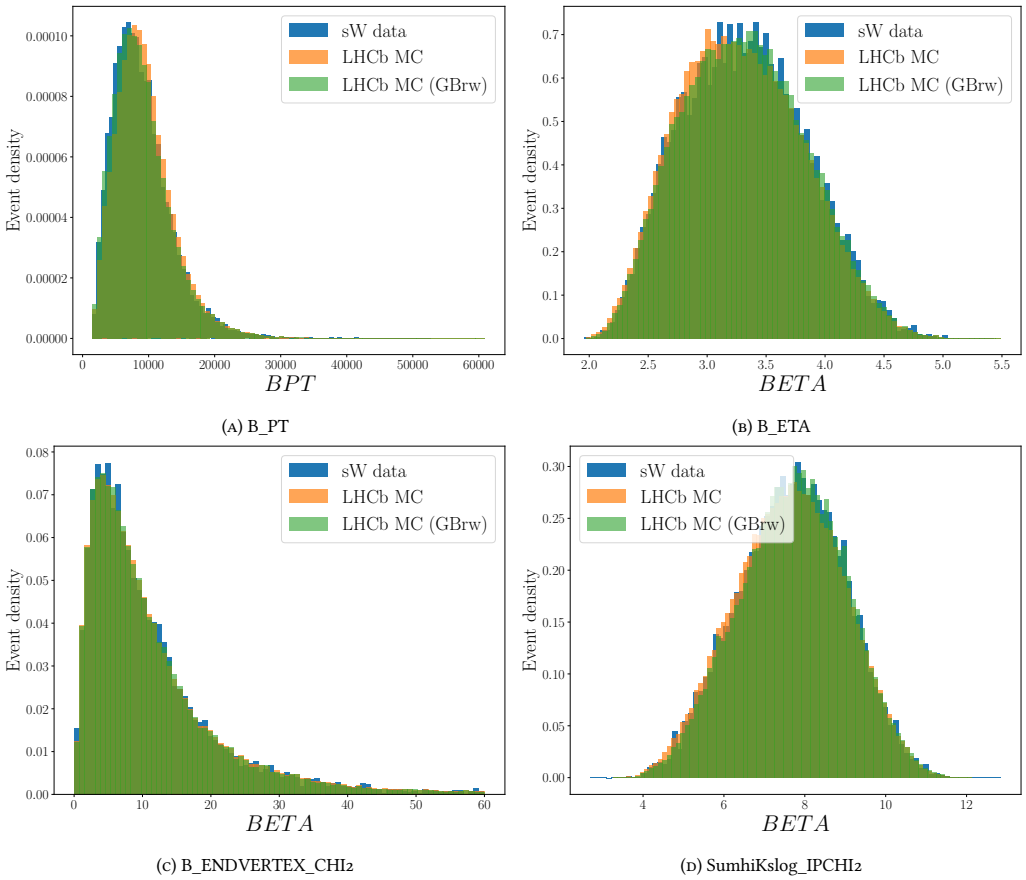
In this way we obtain $w_{(\eta,p)}$ from the GBrw algorithm evaluated on the η and p of the five child particles; in a similar fashion, and after applying the three efficiency corrections as mentioned in Sec. 5.4.2, 5.4.3 and 5.4.4, we obtain $w_{\text{BDT vars.}}$ from the GBrw algorithm evaluated on the BDT variables space.
The final efficiency is then:

$$\epsilon_{\text{control}} = \frac{\sum(w_{(\eta,p)} \times \kappa_{\text{TIS/TOS}} \times \kappa_{\text{Track}} \times \kappa_{\text{PID}} \times w_{\text{BDT vars.}})}{N_{\text{gen}}}\,.$$

To obtain the N_{gen} we simply multiply the number of MC events just after the generation step, before passing trough the detector chain, by the generator level efficiency. We ultimately verify through plotting that the data and re-weighted MC projections of the PHSP are equally distributed. Fig. 5.35 shows these distributions on several mass projections of the PHSP.

5.5 SYSTEMATIC UNCERTAINTIES

In this analysis, we employ a novel approach for estimating the efficiency, which requires an adaptation in the estimation of systematic uncertainties. The key difference from similar analyses is the use of sWeights to correct the efficiency obtained from MC for the decay amplitudes present in the data. In this procedure, the MC sample must capture only the effect of the efficiencies, without capturing any amplitude, thus, the MC must be generated flat in the PHSP. To preserve this flatness, we avoid re-weighting the MC on the



(A) B_PT

(B) B_ETA

(C) B_ENDVERTEX_CHI2

(D) SumhiKslog_IPCHI2

FIGURE 5.34: Some BDT variables on the control channel for sWeighted data and MC before and after BDT re-weighting. The plot shows the LL K^0_S-reco sample. In blue the sWeighted data are shown, in orange the MC with no re-weighting applied and in green the MC after the BDT variables re-weighting. Showing only the 1718 sample for clarity.

variables utilised for the selection, as they may have a correlation with the PHSP variables, compromising the flatness of the MC sample and the effectiveness of the method.
This approach is especially valuable when the signal's amplitude model is not known. Although it may be less optimal if a well-defined model is available, it guarantees accurate parameterisation of signal efficiencies in the absence of a known model. By not depending on a basic model for the efficiency extraction, this method mitigates potential biases in the measurement.

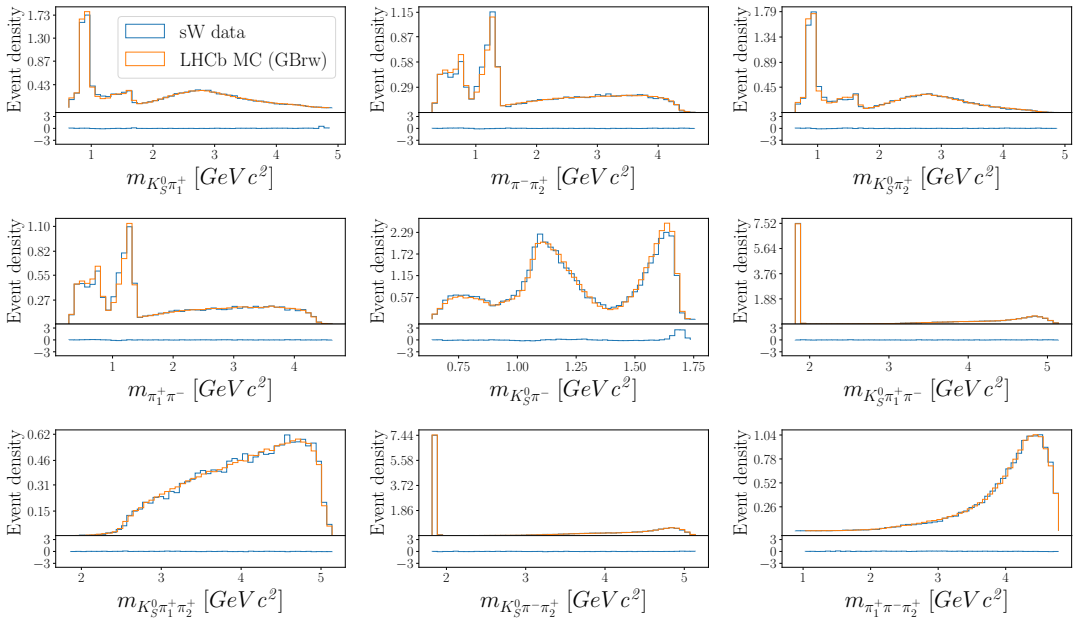
In the following the different systematic sources of uncertainty that has been evaluated will be presented. Note that the systematic uncertainties are computed in the ratio of branching fractions for each sample independently: $\frac{BF_{\text{signal}}}{BF_{\text{control}}}$.
The results of this section are shown in Table 5.5.

5.5.1 FIT BIAS

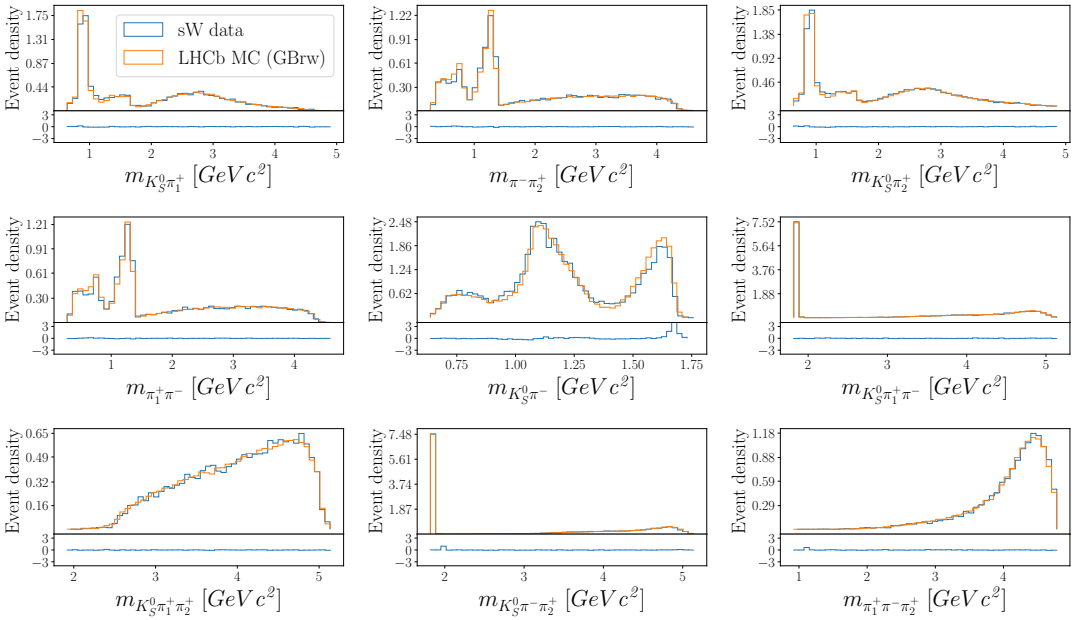
We compute this systematic uncertainty source by sampling data from the mass fit PDF until the sampled statistics match the number of events in the data sample. Then, we perform the mass fit again to these samples. The fit bias is evaluated in the signal region mass fit for the signal channel and in the control channel mass fit. 100 different samples for each fit were created.
The associated uncertainty is the distance between the baseline result and the mean of the distribution of the branching fraction ratio obtained from the 100 calculations on each sample. The result of the different computations is shown in Fig. 5.40E.

5.5.2 SIGNAL MODEL

This systematic uncertainty source impacts the yield extraction and also the efficiency calculation. We change the signal model on the mass fit to the signal channel, from double Crystal-Ball to Hypathia [94],



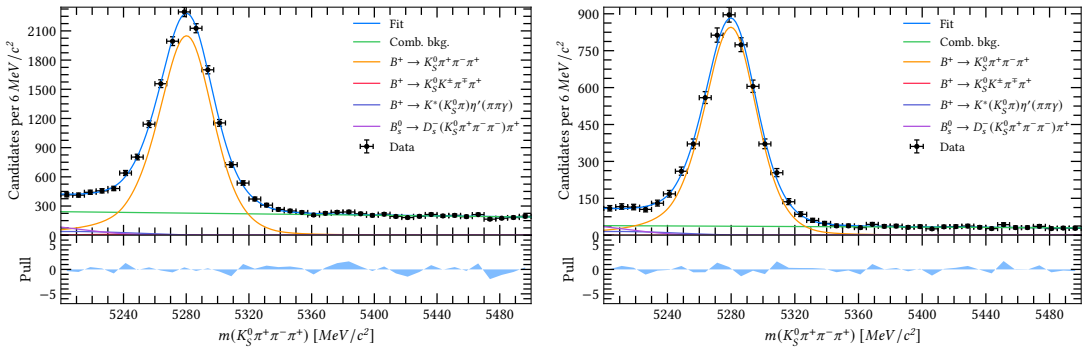
(A) DD



(B) LL

FIGURE 5.35: Control channel PHSP projections for sWeighted data in blue and re-weighted MC in orange from where we extract the efficiencies. We can appreciate some small differences between the samples. Showing only the 1718 sample for clarity.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) DD (B) LL

FIGURE 5.36: Mass fit to the signal channel on signal region using the Hypathia distribution. Showing only the 1718 sample for clarity.

performing the mass fit on the signal region to obtain the sWeights and the new yield. From the point of view of the efficiencies, as the sWeights are used to obtain the efficiency correction, one would check the impact on the efficiency calculation by performing it with the new set of sWeights from the Hypathia distribution.

We compute the systematic uncertainty as the difference between the baseline value and the alternative model. Note that the difference is computed on the branching fraction ratio and both the yield and efficiency play a role. The mass fit result with the Hypathia distribution is shown in Fig. 5.36 and the plot showing the amplitudes extraction with the alternative set of sWeights in Fig. 5.37.

5.5.3 BACKGROUND MODEL

As shown in Table 5.2, the model used to generate the background MC, from which the line shapes for the mass fits are extracted, is flat on the PHSP. These channels involve partially reconstructed decays, and thus their line shapes can be sensitive to the amplitude model. However, these channels are multi-body decays, and there is very little information on the amplitude model. Additionally, the yield observed from the mass fits is very low.

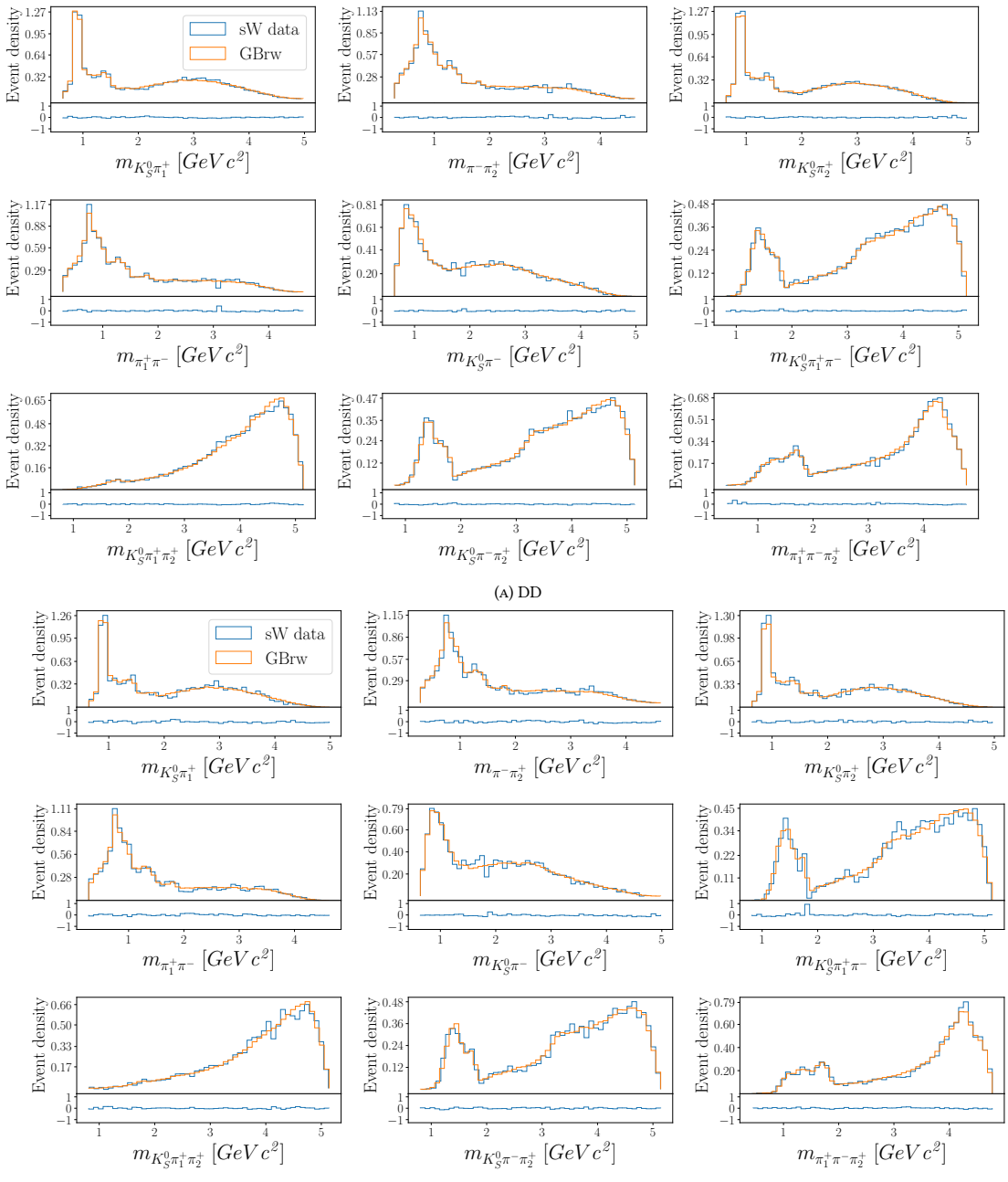
To control this source of systematic uncertainty, we decided to adopt a conservative approach which not involves modelling the amplitudes from these partially reconstructed decays or from the suppressed misID component.

There are three competing contributions to the low-mass tail of the signal candidate in the signal channel mass fit: misidentified $B^+ \to K^0_S K^\pm \pi^\mp \pi^+$, and partially reconstructed $B^+ \to K^{*+}(K^0_S \pi^+)\eta'(\pi^-\pi^+\gamma)$ and $B^0_s \to D^-_s(K^0_S \pi^+\pi^-\pi^-)\pi^+$ decays. We propose excluding these three background sources one by one in the B candidate mass fit to estimate the systematic uncertainty arising from the unknown background models. The associated uncertainty is computed as the difference between the baseline value and the value obtained from these fits, the maximum deviation obtained from these three mass fits will be used to report the uncertainty value.

5.5.4 BDT VARIABLES RE-WEIGHTING

To accurately account for the BDT's impact on data efficiency, it is essential to ensure that the distributions of BDT variables in the MC (from which the efficiency is calculated) match those in the data. As we are not re-weighting the MC in the baseline to preserve the flatness of the PHSP ensuring that none of the amplitudes in data are present in the MC, we assign this computation as a systematic uncertainty source. To minimise the uncertainty from this source, we have used only topological variables in the BDT, as they contain very little information about the PHSP variables. In addition, as we are using the η and p variables of the child particles in the efficiencies correction, we will also take their effect into account in this calculation by computing the MC correction in the same way as in the control channel, as shown in Sec. 5.4.6.2. In Fig. 5.38 we can see the BDT variables distributions on sWeighted data and MC with no re-weighting, we can appreciate that both distributions are very similar. Therefore, we expect the impact of this systematic uncertainty source to be on a similar level as other systematic uncertainty estimations.

To compute this systematic source we follow the exact same procedure as in the control channel:

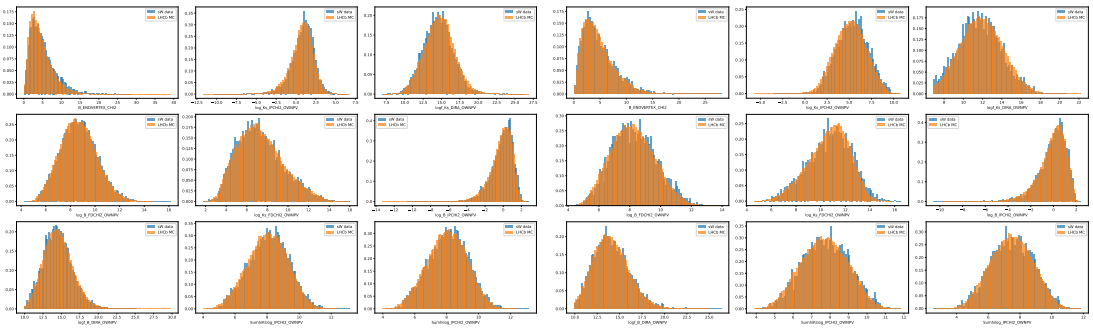


(A) DD

(B) LL

FIGURE 5.37: Performance of the GBrw to extract the efficiency times the amplitudes ($\epsilon(\Phi_4)A(\Phi_4)$) on the sample at generator level with an alternative set of sWeights from the Hypathia distribution. The plots show the sWeighted data in blue on top of the sample at generator level weighted by $w_{\epsilon|A|^2}$ in orange. Showing only the 1718 sample for clarity.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) DD (B) LL

FIGURE 5.38: sWeigted signal channel data against signal channel MC for the BDT training variables. Very little differences between the samples on these variables can be appreciated. Showing only the 1718 sample for clarity.

- Re-weight the variables that are used as input for the efficiency corrections, theη and p of the child particles ($w_{(\eta,p)}$).
- Apply the three efficiency corrections: PID, Tracking and TIS/TOS.
- Finally, re-weight the MC once again to match the BDT variables of the real data sample, taking into account the impact of the BDT on the efficiency ($w_{\text{BDT vars.}}$).

Then one would compute the amplitudes correction with this set of weights for the MC when obtaining the efficiency: $\kappa_{\text{MC}} = w_{(\eta,p)} \times \kappa_{\text{TIS/TOS}} \times \kappa_{\text{Track}} \times \kappa_{\text{PID}} \times w_{\text{BDT vars.}}$. The result of this computation is shown in Fig. 5.39 where we see the efficiency extraction with the new set of weights. The systematic uncertainty source is computed as the difference of the branching fraction ratio with respect to the baseline value. This correction only impacts the efficiency of the signal. Note that this is a very conservative approach to estimate this uncertainty source.

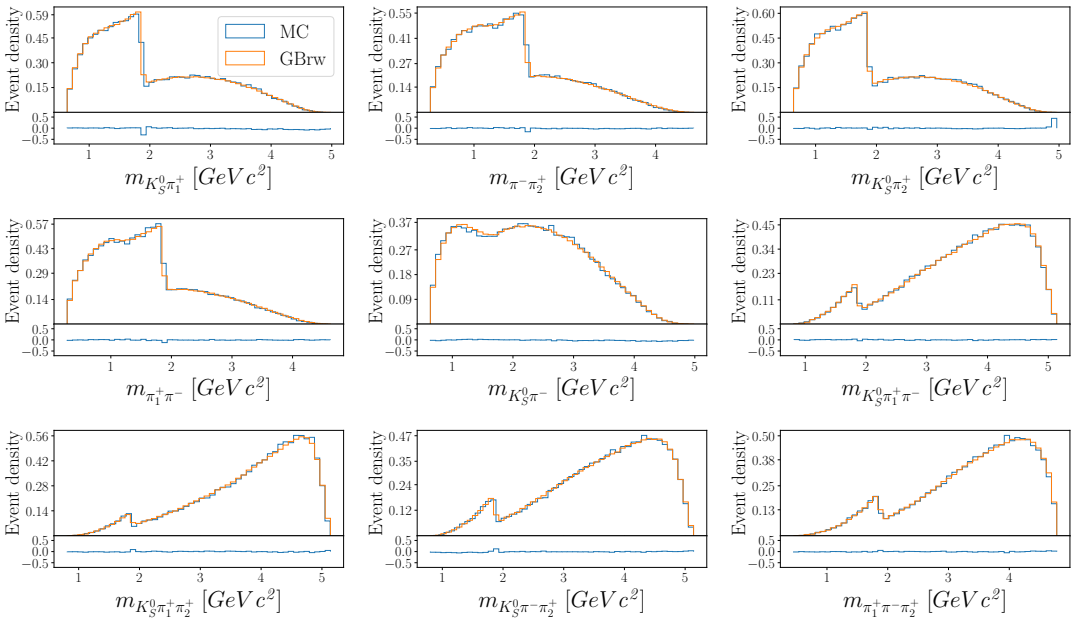
5.5.5 SAMPLE STATISTICS

The size of the samples used to extract and transfer amplitudes and efficiencies to the third sample at the generator level can affect the final correction. Since the third sample contains 1 million events, it is not statistically limited compared to the other two samples. However, the limited statistics in these smaller samples, which are both used for efficiency correction, can interact and impact the results. For example, if there is a specific region of the PHSP greatly populated by one of the samples, but on the other sample there are few events in that region, the uncertainty on the amplitudes correction due to the sample size would be high. This could be the case for a resonance near the threshold of the PHSP, where the flat MC sample has lower density of events. To address this possible effect or interplay, the systematic uncertainty due to limited statistics is computed simultaneously for both samples.

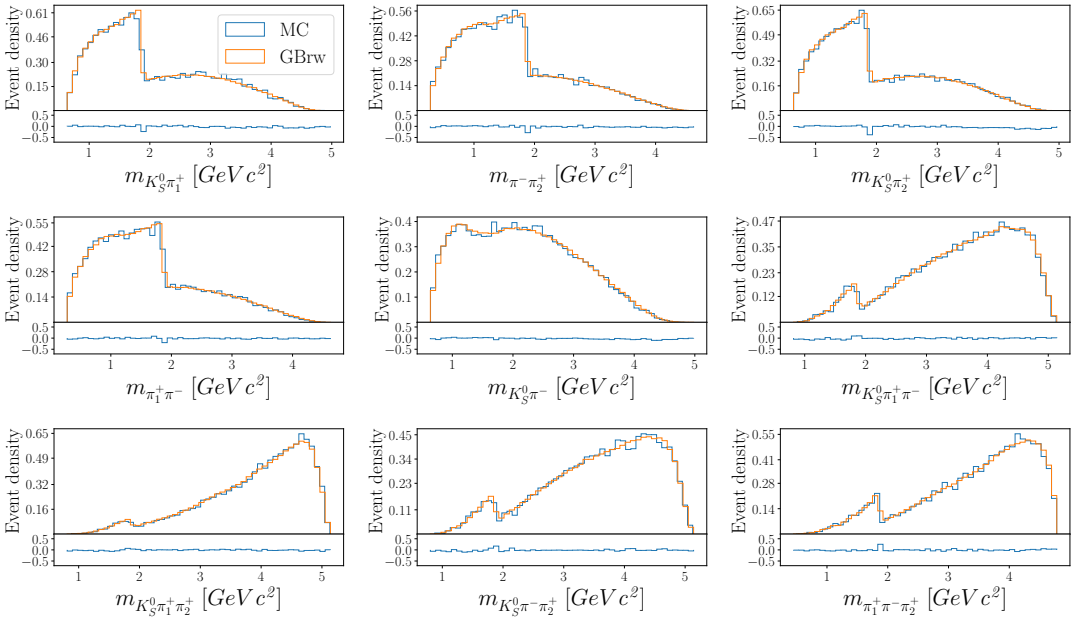
We perform a bootstrapping on these samples so that a random event is taken from each sample until the statistics of the new sample match the original one. We repeat this bootstrapping 100 times for each sample, and in each iteration, we compute the final corrected efficiency. The systematic error is then assign as the standard deviation of the final ratio of branching fraction histogram. The histograms for the different samples are shown in Fig. 5.40F.

5.5.6 EFFICIENCIES CORRECTION

All the efficiency corrections are derived from data-driven histograms, created by taking the ratio of a real data sample to the MC simulation for a specific decay. The primary source of uncertainty in these histograms is the statistical uncertainty, which arises from the sample sizes in each bin used to compute the corrections. Although the chosen binning scheme can also affect the results, it is considered a secondary factor and will not be investigated in this analysis. As these are not the main source of systematic uncertainty, this secondary factor would not have a significant impact on the systematic uncertainty value. Consequently, each event has a correction with an associated uncertainty, which we can use to estimate the overall uncertainty on the correction.



(A) DD



(B) LL

FIGURE 5.39: Performance of the GBrw to extract the efficiency ($\epsilon(\Phi_4)$) on the sample at generator level, the BDT variables are re-weighted to match the sWeighted data. The plots show the efficiency corrected LHCb MC in blue on top of the sample at generator level weighted by w_ϵ in orange. Showing only the 1718 sample for clarity.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.5.6.1 TIS/TOS LO CORRECTION

In this systematic uncertainty computation, we vary the correction factor of each event by randomising it 100 times within its uncertainty. For each iteration, a random number is drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 1. This random number is used as a factor to multiply the uncertainty of the correction for each event. The resulting value is then added to the correction factor, scaling the correction for each event. This process ensures that the uncertainty is varied in a correlated manner across all events in the sample and for both the signal and control channels. Finally, we calculate the ratio of branching fractions for the 100 iterations, with the associated systematic uncertainty being the standard deviation of the histogram of these 100 values. The result of the computations is shown in Fig. 5.40D.

5.5.6.2 TRACK CORRECTION

Same procedure as the TIS/TOS Lo correction. The result of the computations is shown in Fig. 5.40C.

5.5.6.3 PID CORRECTION

Same procedure as the TIS/TOS Lo correction. The result of the computations is shown in Fig. 5.40B.

5.5.7 GBRW HYPER-PARAMETERS

It was observed that the GBrw hyper-parameter *max_depth* significantly impacts the results. There are three main re-weightings to compute the efficiencies: the BDT re-weighting on the control channel, and the efficiency and amplitude extraction in the signal channel. Although certain values of *max_depth* $(3, 4, 5, 6, 7)$ do not affect how the plotted result looks, they do produce substantial differences in value of the efficiency results. For the baseline, *max_depth* was set to 5 for all three re-weightings. To assess this systematic uncertainty, we compute the efficiencies using all five possible values of *max_depth* for the three re-weightings, resulting in $5^3 = 125$ different configurations. We then obtain the efficiency values for all these configurations. The final uncertainty is determined by the standard deviation of the histogram generated from the 125 branching fraction ratio values for each sample. Fig. 5.40A shows the results of these computations.

sample	GBrw	Bootstrap.	Fit Bias	BDT rw	sW	Bkg Model
1112 LL	$2.5{\cdot}10^{-2}$	$\mathbf{6.8 \cdot 10^{-2}}$	$1.7{\cdot}10^{-3}$	$\mathbf{1.0 \cdot 10^{-1}}$	$6.6{\cdot}10^{-3}$	$3.2{\cdot}10^{-3}$
1516 LL	$2.6{\cdot}10^{-2}$	$\mathbf{4.6 \cdot 10^{-2}}$	$1.3{\cdot}10^{-5}$	$2.6{\cdot}10^{-2}$	$5.6{\cdot}10^{-3}$	$5.2{\cdot}10^{-3}$
1718 LL	$\mathbf{2.4 \cdot 10^{-2}}$	$1.1{\cdot}10^{-2}$	$4.1{\cdot}10^{-5}$	$1.6{\cdot}10^{-2}$	$1.2{\cdot}10^{-2}$	$1.5{\cdot}10^{-3}$
1112 DD	$1.4{\cdot}10^{-2}$	$\mathbf{6.3 \cdot 10^{-2}}$	$1.2{\cdot}10^{-3}$	$5.3{\cdot}10^{-2}$	$3.6{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$
1516 DD	$\mathbf{1.2 \cdot 10^{-2}}$	$8.7{\cdot}10^{-3}$	$1.3{\cdot}10^{-5}$	$1.1{\cdot}10^{-2}$	$7.4{\cdot}10^{-3}$	$1.0{\cdot}10^{-3}$
1718 DD	$\mathbf{1.1 \cdot 10^{-2}}$	$2.8{\cdot}10^{-3}$	$7.2{\cdot}10^{-4}$	$9.3{\cdot}10^{-3}$	$9.8{\cdot}10^{-3}$	$5.3{\cdot}10^{-6}$

TABLE 5.5: Results from the systematic uncertainties estimations. The distributions that retrieve these results are shown in Fig.5.40. The details on each computation are given in Sec. 5.5. The dominant systematic for each sample is highlighted in bold. GBrw column refers to the systematic uncertainty source presented in Sec. 5.5.7, Bootstrap. to the systematic uncertainty source explained in Sec. 5.5.5, Fit Bias column is explained in Sec. 5.5.1, BDT rw in Sec. 5.5.4, sW is the systematic error source related to the signal model as explained in Sec. 5.5.2, Bkg Model is related to the signal channel backgrounds amplitude model explained in Sec. 5.5.3 and, lastly, the systematic error sources due to the efficiency corrections as shown in Secs. 5.5.6.2, 5.5.6.3 and 5.5.6.1.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

sample	Tracking	PID	Lo
1112 LL	$5.9{\cdot}10^{-3}$	$2.1{\cdot}10^{-3}$	$3.7{\cdot}10^{-3}$
1516 LL	$1.1{\cdot}10^{-2}$	$2.0{\cdot}10^{-3}$	$3.4{\cdot}10^{-3}$
1718 LL	$7.0{\cdot}10^{-3}$	$2.0{\cdot}10^{-3}$	$2.0{\cdot}10^{-3}$
1112 DD	$4.1{\cdot}10^{-3}$	$3.5{\cdot}10^{-4}$	$1.1{\cdot}10^{-3}$
1516 DD	$8.3{\cdot}10^{-3}$	$5.8{\cdot}10^{-4}$	$1.2{\cdot}10^{-3}$
1718 DD	$6.0{\cdot}10^{-3}$	$2.6{\cdot}10^{-4}$	$7.2{\cdot}10^{-4}$

TABLE 5.5: Results from the systematic uncertainties estimations. The distributions that retrieve these results are shown in Fig.5.40. The details on each computation are given in Sec. 5.5. The dominant systematic for each sample is highlighted in bold. GBrw column refers to the systematic uncertainty source presented in Sec. 5.5.7, Bootstrap. to the systematic uncertainty source explained in Sec. 5.5.5, Fit Bias column is explained in Sec. 5.5.1, BDT rw in Sec. 5.5.4, sW is the systematic error source related to the signal model as explained in Sec. 5.5.2, Bkg Model is related to the signal channel backgrounds amplitude model explained in Sec. 5.5.3 and, lastly, the systematic error sources due to the efficiency corrections as shown in Secs. 5.5.6.2, 5.5.6.3 and 5.5.6.1.

5.6 BRANCHING FRACTION RESULT

After obtaining the yields and efficiencies for both the signal and control channels across all samples, we can merge these measurements to determine the signal branching fraction.
We combine the ratio of branching fractions from the different samples in a χ^2 fit. Once we obtain the combined branching fraction ratio value, we use the branching fraction from the control channel as input to obtain the final result. Following Eq. 5.1, the ratio of branching fractions can be written as:

$$\frac{BF_{\text{signal}}}{BF_{\text{control}}} = \frac{f_{\text{signal}}}{\epsilon_{\text{signal}}} \frac{\epsilon_{\text{control}}}{f_{\text{control}}} .$$

The different quantities needed for the branching fraction ratio are obtained from the following computations:

- The signal channel yield (f_{signal}) is obtained from the mass fit to the signal narrow region, as shown in Sec. 5.3.4.
- The control channel yield (f_{control}) is obtained from the mass fit to the wide region, as shown in Sec. 5.3.3.1.
- Signal channel efficiency (ϵ_{signal}) is taken from the output of the amplitudes correction, as explained in Sec. 5.4.6.1.
- Control channel efficiency ($\epsilon_{\text{control}}$) is taken from the output of the BDT variables re-weighting, as explained in Sec. 5.4.6.2.

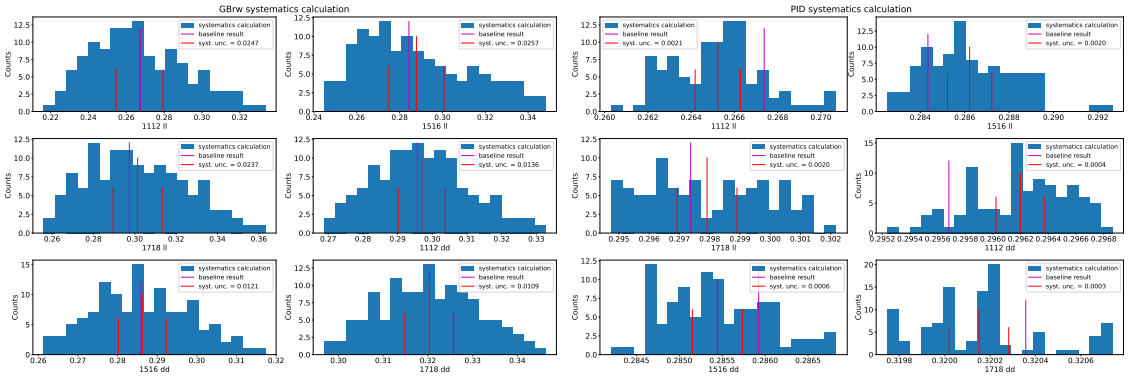
The values used to compute the branching fraction ratio for each sample are shown in Table 5.6 and the ratio values along with their uncertainties used in the χ^2 combination fit are presented in Table 5.7. By redefining $\frac{BF_{\text{signal}}}{BF_{\text{control}}} \equiv r$, we can formulate our estimator to be minimised as a function of the fitting parameter μ, as follows:

$$\chi^2(\mu) = \sum_i \left(\frac{(r_i - \mu)}{\sqrt{(u_{stat.}(r_i))^2 + (u_{syst.}(r_i))^2}} \right)^2 , \tag{5.18}$$

where the index i runs over the 6 different samples, and u represents the statistical and systematic uncertainties, respectively. The parameter μ that minimises the χ^2 estimator is considered our fit result. The result from the χ^2 combination fit is:

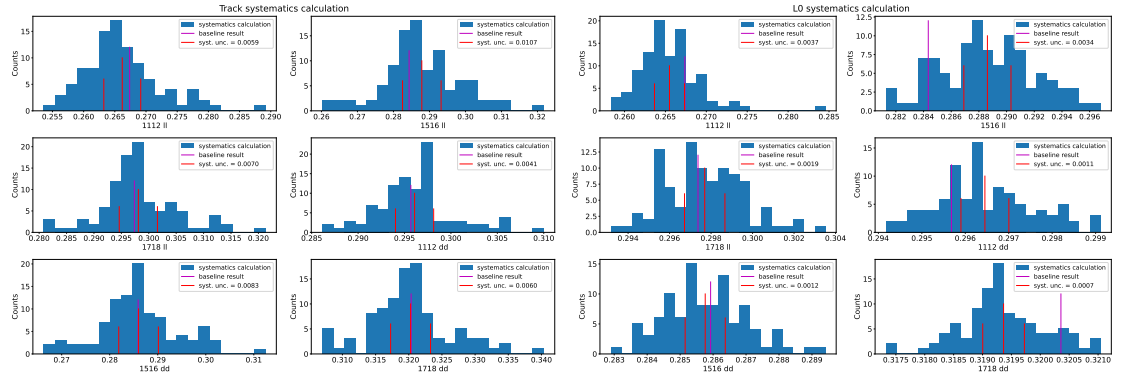
$$\frac{BF_{\text{signal}}}{BF_{\text{control}}} = 0.296 \pm 0.013 , \tag{5.19}$$

with a relative uncertainty of 3%, remember that the statistical and systematic uncertainties are combined in quadrature for each sample before the fit. In Fig. 5.41, the result is shown alongside the input values used to obtain the combined value. The plot demonstrates good agreement between the measurements from the different samples, highlighting the accurate performance of the method within the estimated sources of uncertainty. The minimum value of the estimator found in the minimisation is $\chi^2 = 1.5$, which further certifies the good agreement between the measurements.



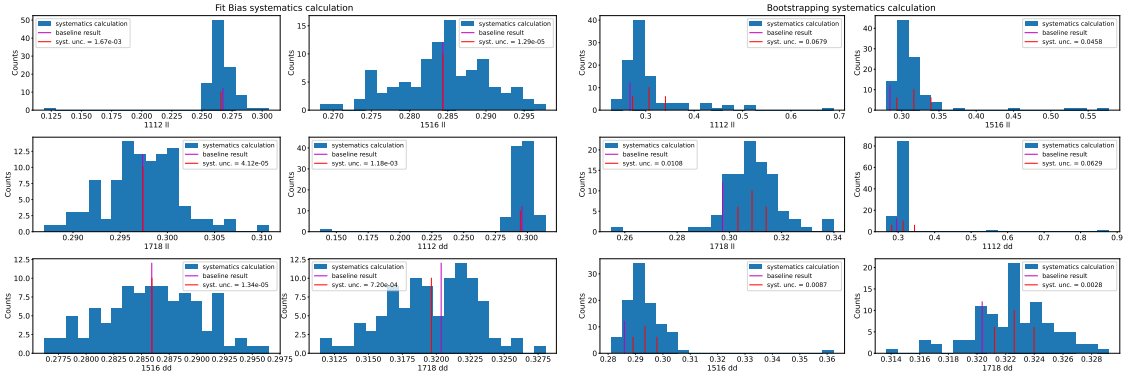
(A) GBrw parameters

(B) PID correction



(C) Tracking correction

(D) Lo TIS/TOS correction



(E) Fit bias

(F) Bootstrapping

FIGURE 5.40: Results of the multiple calculations in the systematic uncertainty sources estimation. Exclusively the systematic uncertainty sources evaluated by sampling distributions are shown in this figure.

sample	$\epsilon^{\text{MC}}_{\text{signal}}$	ϵ_{signal}	$\epsilon_{\text{control}}$	f_{signal}	$u(f_{\text{signal}})$	f_{control}	$u(f_{\text{control}})$
1112 LL	$4.42 \cdot 10^{-5}$	$4.02 \cdot 10^{-5}$	$2.45 \cdot 10^{-5}$	$1.45 \cdot 10^{3}$	$2.78 \cdot 10^{1}$	$3.46 \cdot 10^{3}$	$5.85 \cdot 10^{1}$
1516 LL	$1.03 \cdot 10^{-4}$	$8.46 \cdot 10^{-5}$	$6.43 \cdot 10^{-5}$	$2.97 \cdot 10^{3}$	$4.49 \cdot 10^{1}$	$8.13 \cdot 10^{3}$	$1.05 \cdot 10^{2}$
1718 LL	$8.48 \cdot 10^{-5}$	$7.01 \cdot 10^{-5}$	$5.59 \cdot 10^{-5}$	$5.09 \cdot 10^{3}$	$4.83 \cdot 10^{1}$	$1.41 \cdot 10^{4}$	$1.42 \cdot 10^{2}$
1112 DD	$1.30 \cdot 10^{-4}$	$1.17 \cdot 10^{-4}$	$9.65 \cdot 10^{-5}$	$3.00 \cdot 10^{3}$	$4.89 \cdot 10^{1}$	$8.38 \cdot 10^{3}$	$9.45 \cdot 10^{1}$
1516 DD	$2.48 \cdot 10^{-4}$	$2.19 \cdot 10^{-4}$	$1.85 \cdot 10^{-4}$	$5.96 \cdot 10^{3}$	$7.61 \cdot 10^{1}$	$1.81 \cdot 10^{4}$	$1.59 \cdot 10^{2}$
1718 DD	$2.33 \cdot 10^{-4}$	$2.09 \cdot 10^{-4}$	$1.59 \cdot 10^{-4}$	$1.30 \cdot 10^{4}$	$1.04 \cdot 10^{2}$	$3.17 \cdot 10^{4}$	$2.02 \cdot 10^{2}$

TABLE 5.6: Efficiencies and yields used in the χ^2 fit to combine the measurements. $\epsilon^{\text{MC}}_{\text{signal}} \equiv \epsilon_{\text{LHCb MC}}$ shows the efficiency before the amplitudes correction, for the branching fraction ratio calculation we use the amplitudes corrected one: ϵ_{signal}.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

sample	$\frac{BF_{\text{signal}}}{BF_{\text{control}}}$	$u_{stat.}(\frac{BF_{\text{signal}}}{BF_{\text{control}}})$	$u_{syst.}(\frac{BF_{\text{signal}}}{BF_{\text{control}}})$
1112 LL	$2.56 \cdot 10^{-1}$	$6.56 \cdot 10^{-3}$	$1.24 \cdot 10^{-1}$
1516 LL	$2.76 \cdot 10^{-1}$	$5.53 \cdot 10^{-3}$	$6.02 \cdot 10^{-2}$
1718 LL	$2.87 \cdot 10^{-1}$	$3.97 \cdot 10^{-3}$	$3.37 \cdot 10^{-2}$
1112 DD	$2.94 \cdot 10^{-1}$	$5.82 \cdot 10^{-3}$	$8.37 \cdot 10^{-2}$
1516 DD	$2.80 \cdot 10^{-1}$	$4.34 \cdot 10^{-3}$	$2.18 \cdot 10^{-2}$
1718 DD	$3.12 \cdot 10^{-1}$	$3.20 \cdot 10^{-3}$	$1.86 \cdot 10^{-2}$

TABLE 5.7: Input ratios used in the χ^2 fit to combine the measurements. $u_{stat.}$ is the statistical uncertainty and $u_{syst.}$ is the systematic one, the systematic uncertainty is combined in quadrature from the different estimated sources.

To compute the final branching fraction we use the value from the control channel as input to obtain our final result. The input branching fraction is extracted from the PDG [49]. Two different branching fraction values are used, firstly, the branching fraction of the $B^+ \to \bar{D}^0\pi^+$ decay, and it is then multiplied by the $D^0 \to K^0_S\pi^+\pi^-$ branching fraction, obtaining then the control channel branching fraction:

- $BF(B^+ \to \bar{D}^0\pi^+) = (4.68 \pm 0.13) \cdot 10^{-3}$.
- $BF(D^0 \to K^0_S\pi^+\pi^-) = (2.80 \pm 0.18)\%$.

The control channel branching fraction is then:

$$BF_{\text{control}} = BF(B^+ \to \bar{D}^0\pi^+) \times BF(D^0 \to K^0_S\pi^+\pi^-) = (1.310 \pm 0.092) \cdot 10^{-4}, \tag{5.20}$$

note that the input BF_{control} has a 7% of relative uncertainty.

Finally, we multiply both values to obtain our branching fraction measurement:

$$BF_{\text{signal}} = BF_{\text{control}} \times \frac{BF_{\text{signal}}}{BF_{\text{control}}} = (3.87 \pm 0.32) \cdot 10^{-5}\,. \tag{5.21}$$

We can report the result separating the different uncertainty sources, first the uncertainty on the ratio and then, the uncertainty on the PDG input.

$$\boxed{BF_{\text{signal}} = (3.87 \pm 0.17 \pm 0.27) \cdot 10^{-5}} \tag{5.22}$$

5.7 DISCUSSION

To contextualise our findings, it is helpful to reference previous studies and results, which can provide an indication of the expected magnitude of our result. There are some previous results in the literature that give us a hint of what the order of magnitude of our result should be. However, the mass window including VV and Axial decays on the charmless region of the PHSP used in this analysis has never been explored before. In addition, there is no theoretical input on these purely penguin process that fall in our mass window. Then, we are going to present these previous results remarking the similarities with our analysis in each case but taking into account that no direct comparison can be done with any of them:

- $BF(B^+ \to K^*(K^0_S\pi)\pi\pi) = (7.5 \pm 1.0) \cdot 10^{-5}$, from BaBar paper [91]. In this analysis the full PHSP is analysed with vetoes on the main charm contributions. Despite our mass window includes more $K\pi$ resonances, such as the $K^*(1430)$, it is very clear that the $K^*(980)$ is the dominant one, as can be seen in Fig. 5.29. In addition, the covered PHSP on the $\pi\pi$ projection of our mass window is clearly more limited than on a full PHSP analysis. In conclusion, this result could be considered as an upper limit on our branching fraction.
- $BF(B^+ \to K^0_S a_1^+(\pi\pi\pi)) = (17.4 \pm 2.5 \pm 2.2) \cdot 10^{-6}$, from BaBar paper [92]. In this analysis a similar mass window to our a_1^+ mass block ($A_{\pi_1^+\pi^-\pi_2^+}$) is set. This contribution can be considered to fall inside our mass window.
- $BF(B^+ \to (K_1^0(1400)\pi^+ + K_1^0(1270)\pi^+)) = (2.9^{+2.9}_{-1.7}) \cdot 10^{-5}$, from BaBar paper [44], the result is obtained on these exclusive decays that indeed are present in our mass window. Note that the uncertainty on this result is the biggest one between the 3 results presented here.

These results align well with our final findings. While a direct comparison suggests that our branching fraction might be slightly underestimated, the large uncertainties associated with these results and the fact that no direct comparison is possible must be considered. Overall, we find that our results are in fair agreement with the state-of-the-art measurements.

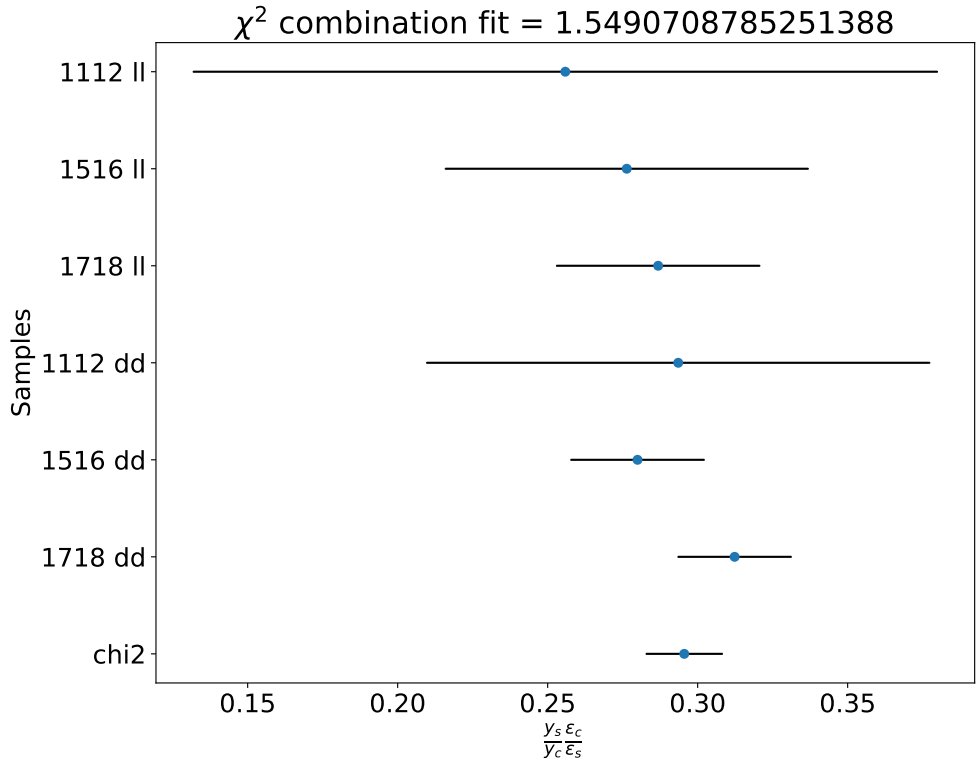


FIGURE 5.41: χ^2 combination fit to obtain $\frac{BF_{\text{signal}}}{BF_{\text{control}}}$. In the y-axis, the input from the different samples is shown, and the last entry labelled as 'chi2' represents the fit result of the combination.

In addition to this comparison with other results, it is also interesting to report the impact in the measurement of the novel strategies presented in the analysis. The novel strategy used for efficiency estimation in this branching fraction analysis is motivated by the necessity for high accuracy in determining efficiencies for novel decay channels where the amplitude model is not known. Usually, when the model is not known, a naive model of the signal is utilised. This assumption may impact the accuracy of the measurement, as the PHSP efficiencies may differ from the simulated ones, leading to a bias in the measurement. In the present work, we push our effort to reduce this bias by employing a model-independent approach using data-driven methods on the PHSP to estimate the signal efficiency.
This method uses information from the data amplitudes to obtain a correction to the efficiency derived from the simulation. Hence, we can compute the size of this correction by comparing the branching fraction ratio reported in the analysis with the case where no correction is applied. The size of the correction is 10% for the DD K^0_S reconstruction case and 17% for the LL case, highlighting that this is indeed a significant correction for our analysis. This size is computed as $(r_{\text{no correction}} - r)/r$, where r represents the branching fraction ratio.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 6

MOMENTUM SCALE CALIBRATION AT LHCB USING THE ARMENTEROS-PODOLANSKI PLOT

The momentum scale calibration is crucial for the correct functioning of a particle physics detector, as it ensures a good accuracy on the momentum determination of the charged tracks. In this chapter, a new approach to estimate the momentum scale bias is presented. The momentum scale bias can occur due to several different sources, such as magnetic field map mis-modelling, sub-detectors mis-alignment [95], or other different sources that lead to poor accuracy in momentum determination.

In a previous work [20], we showed how the momentum scale calibration can be benefited by the fact that in the Armenteros-Podolanski plot [18] no prior knowledge of any particle's mass involved in the decay is needed to fit them. Now, this idea is going to be extended to use a modified version of the Armenteros plot to fit, at the same time, the mass of the parent particle (which is generally poorly known) and the momentum scale. In this proof of principle, we use real LHCb data, and, to obtain the fit parameters, we use a full LHCb simulation that we change in each fit iteration to match the desired mass and momentum scale parameters.

The last report on the momentum scale calibration at the LHCb detector was published on 2024 [17]. The current method relies on fitting the parent masses of different decays such as $J/\Psi \to \mu^+\mu^-$ and $B^+ \to J/\Psi K^+$. By comparing the reconstructed mass of the parent particles with the reference values and combining the measurements, the bias of the global momentum scale can be extracted, leading to a relative accuracy of 3×10^{-4}.

6.1 THE ARMENTEROS-PODOLANSKI PLOT

In the Armenteros-Podolanski plot, two-body decays ($M \to m_1 m_2$) are distributed as ellipses. From the parameters that describe the ellipse, one can get the masses of the three particles involved in the decay. To build the Armenteros-Podolanski plot, one must measure the momentum values of the child particles. The space is defined as:

$$(\alpha, p_T), \tag{6.1}$$

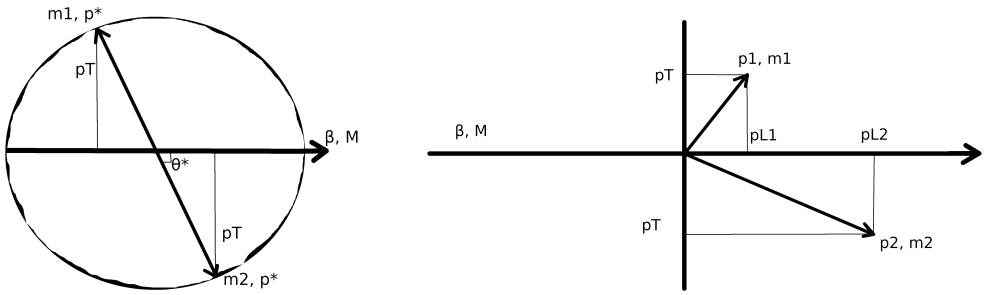
where p_T is the transverse momentum of the children particles with respect to the parent momentum and α is the asymmetry of the longitudinal momentum:

$$\alpha = \frac{p_{L1} - p_{L2}}{p_{L1} + p_{L2}}, \tag{6.2}$$

where subscript 1 refers to the positively charged particle and 2 to the negative one. In Fig. 6.1b, one can see a sketch with the definition of the variables used to build the Armenteros plot. Note that, due to momentum conservation the p_T of both particles is the same and, as the transverse component of a Lorentz boost remains invariant, is also the same as the p_T in the centre-of-mass reference frame. Thus, the p_T of a two-body decay is an invariant quantity when the reference frame is changed. A representation of the

USC

momentum values at the centre-of-mass is shown in Fig. 6.1A, where one can see that the momentum of the child particles is opposite and have the same magnitude due to momentum conservation. Later on, we will realise why these properties are key to modify the mass of the parent particle in the full LHCb simulation in order to perform a fit to real data.



(A) CM (B) LAB

FIGURE 6.1: Two-body decay PHSP variables on the centre-of-mass (CM) and laboratory (LAB) reference frames.

The next question that naturally arise is how to build the ellipse from the Armenteros variables so that we can get the mass parameters of the decay from the ellipse. The first thing to do is to build the four-momentum (k^μ) conservation at the CM. Note that in the following we are using natural units, i.e. $c = 1$. In addition, the variables with an asterisk (*) refers to the variables in the centre-of-mass system:

$$k^\mu_{CM,before} = (M, \vec{0}), \tag{6.3}$$

$$k^\mu_{CM,after} = (E_1^* + E_2^*, \mathbf{p}^* + (-\mathbf{p}^*)). \tag{6.4}$$

$$(k^\mu_{CM,before})^2 = (k^\mu_{CM,after})^2 \Leftrightarrow M^2 = (E_1^* + E_2^*)^2 - (\mathbf{p}^* + (-\mathbf{p}^*))^2 = (E_1^* + E_2^*)^2, \tag{6.5}$$

and, getting rid of the energy-momentum relativistic relation: $E = \sqrt{p^2 + m^2}$, one can get the momentum of the child particles in the CM:

$$p^{*2} = \frac{1}{4M^2}(M^4 + m_1^4 + m_2^4 - 2M^2(m_1^2 + m_2^2) - 2m_1^2 m_2^2). \tag{6.6}$$

Note that p^* is the children momentum in the CM frame, also known as the break-up momentum and expressed as q. Finally, applying the energy-momentum relation to Eq. 6.6 one can get the energy of the child particles in the CM:

$$E_1^* = \frac{1}{2M}(M^2 + m_1^2 - m_2^2), \qquad E_2^* = \frac{1}{2M}(M^2 - m_1^2 + m_2^2). \tag{6.7}$$

From Fig. 6.1A we can easily get that:

$$\sin\theta^* = \frac{p_T}{p^*}, \tag{6.8}$$

being θ^* the angle between the child momentum and the boost direction (or parent momentum) in the centre-of-mass system.

In order to get the children's momentum in the LAB system we must apply the Lorentz boost to the CM momentum. Remember that we just need to apply the boost to the longitudinal components as the p_T is a Lorentz invariant, thus:

$$p_L = \gamma\beta E^* + \gamma p_L^*, \tag{6.9}$$

where $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ is the Lorentz factor and $\beta \equiv \frac{v}{c}$ is the parent velocity in c units.

Including this expression on the Armenteros α variable taking into account that $p_L^* = p^* \cos\theta^*$ and using Eq. 6.7:

$$\alpha = \frac{2p^*}{\beta M}\cos\theta^* + \frac{m_1^2 - m_2^2}{M^2}, \tag{6.10}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

in this equation, all the information from the decay is included: the masses that participate in the decay, the value of the boost, i.e. parent velocity, and the angle between the boost and the children's momentum in the CM.

If we now define:

$$\alpha_0 = \frac{m_1^2 - m_2^2}{M^2}; \qquad r_\alpha = \frac{2p^*}{M}, \tag{6.11}$$

and then impose the ultra-relativistic limit with $\beta \longrightarrow 1$, we obtain:

$$\frac{\alpha - \alpha_0}{r_\alpha} = \cos\theta^* . \tag{6.12}$$

Finally, in order to get the ellipse equation, we just need to consider this last equation, Eq. 6.8 and the trigonometric relation $\cos^2 a + \sin^2 a = 1$:

$$\frac{(\alpha - \alpha_0)^2}{r_\alpha^2} + \frac{p_T^2}{p^{*2}} = 1 . \tag{6.13}$$

Consequently, we derive the equation of the ellipse that represents a two-body decay in the Armenteros-Podolanski plot.

6.1.1 THE MODIFIED ARMENTEROS SPACE

The ultimate goal of this study is to obtain the best accuracy we can in the parameters of interest in determining the momentum scale. The data used for calibration is very naively selected, thus, these samples are very polluted due to different background sources. In order to fit the polluted data on a 2D space, we use a binned fitting method; hence, the accuracy is then strongly dependent on the binning scheme. To avoid having empty bins during the fit calculation, which results in a waste of computing resources, we transform the space to elliptical coordinates (ϕ, r), where a reference ellipse appears as a straight line ($r = 1$) and we can do a very fine rectangular binning around it.
The variable change is written as follows:

$$\begin{aligned} \alpha(\phi, r) &= \tilde{\alpha}_0 + \tilde{r}_\alpha\, r\, \cos\phi , \\ p_T(\phi, r) &= \tilde{p}^*\, r\, \sin\phi . \end{aligned} \tag{6.14}$$

Where the parameters with a tilde, e.g.: $\tilde{p}^*$, are the parameters of the ellipse we want to focus on. These parameters are usually fix from the known masses of the two-body decay we want to fit.
We then can write the (ϕ, r) variables as a function of the Armenteros ones:

$$\begin{aligned} r(\alpha, p_T) &= \sqrt{\frac{(\alpha - \tilde{\alpha}_0)^2}{\tilde{r}_\alpha^2} + \frac{p_T^2}{\tilde{p}^{*2}}} , \\ \phi(\alpha, p_T) &= \arccos\left(\frac{\tilde{p}^*\,(\alpha - \tilde{\alpha}_0)}{\sqrt{(\tilde{p}^*\,(\alpha - \tilde{\alpha}_0))^2 + (p_T\,\tilde{r}_\alpha)^2}} \right) . \end{aligned} \tag{6.15}$$

Changing the x-range of the space from $(-1, 1)$ to $(0, \pi)$. Finally, one can plug the variable change into the ellipse equation to obtain the new parametric formula of the two-body decays in the modified Armenteros space:

$$\frac{(\tilde{\alpha}_0 + \tilde{r}_\alpha\, r\, \cos\phi - \alpha_0)^2}{r_\alpha^2} + \frac{(\tilde{p}^*\, r\, \sin\phi)^2}{p^{*2}} = 1 . \tag{6.16}$$

Solving the quadratic equation one can obtain the general solution:

$$\begin{aligned} r &= \frac{-B + \sqrt{B^2 - 4\,A\,C}}{2\,A}, \\ A &= \tilde{r}_\alpha^2\, p^{*2}\, \cos^2\phi + r_\alpha^2 \tilde{p}^{*2} \sin^2\phi , \\ B &= 2\,\tilde{r}_\alpha\, p^{*2}\, \cos\phi(\tilde{\alpha}_0 - \alpha_0) , \\ C &= p^{*2}\left((\tilde{\alpha}_0 - \alpha_0)^2 - r_\alpha^2\right) . \end{aligned} \tag{6.17}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Finally, in order to simplify the equation, we can take the case of a symmetric decay where $m \equiv m_1 = m_2$:

$$
\begin{aligned}
r &= \sqrt{\frac{1}{\left(\frac{\tilde{r}_\alpha}{r_\alpha}\cos\phi\right)^2 + \left(\frac{\tilde{p}^*}{p^*}\sin\phi\right)^2}}, \\
r_\alpha &= \frac{\sqrt{M^2 - 4\,m^2}}{M}, \quad \tilde{r}_\alpha = \frac{\sqrt{\tilde{M}^2 - 4\,\tilde{m}^2}}{\tilde{M}}, \\
p^* &= \sqrt{M^2/4 - m^2}, \quad \tilde{p}^* = \sqrt{\tilde{M}^2/4 - \tilde{m}^2}\,.
\end{aligned}
\tag{6.18}
$$

In this equation is easy to spot that, when the ellipse mass parameters (M, m) match the variable change parameters (the ones with the tilde), the analytic expression for the decay is a straight line ($r = 1$). However, in our computations we will use the generic formula to cover both cases, the symmetric one as the $K^0_S \to \pi^+\pi^-$ decay and the general case as $\Lambda^0 \to p\pi^-$ decay.

6.2 EXPERIMENTAL EFFECTS ON THE ARMENTEROS-PODOLANSKI PLOT

Up to now we have seen that we can parameterise the decay ellipse as a function of the three masses that participate in a two-body decay. Nevertheless, to match the observed real data on the Armenteros plot, we must take into account many experimental effects. The easiest effect to realise is the finite momentum of the parent particle, as to build the ellipse we assumed the ultra-relativistic limit ($\beta \to 1$). This effect makes that the decays where the parent particle has the lowest momentum, the data points does not follow the ellipse equation. When comparing real data to simulation is then important that both have a similar distribution on the parent momentum magnitude. Another experimental effect that can be easily spotted is that the Armenteros plot is not uniformly populated on the x-axis. The distribution of α (or ϕ) depends on the acceptance of highly asymmetric decays. For example, if one of the particles is too slow, it may be ejected by the magnet.
These experimental effects will be accounted for in the simulation. Therefore, the momentum acceptance must be similar in real data and in the MC simulation of the decay that will be used to obtain the momentum and mass scale parameters. In addition, the resolution in the momentum measurement must be also properly simulated.

In the following, we are going to present three different experimental effects that we can add to our simulated samples to measure their impact on data. Thus, later on, these may be our fitting parameters. These are the momentum scale, the mass scale, and the charge asymmetry. We will not include the charge asymmetry in this proof of principle study, but we will demonstrate its impact on the Armenteros-Podolanski plot, indicating that it could serve as the third fitting parameter. In Fig. 6.2, the MC simulation on the modified Armenteros space is shown, where the impact of different experimental effects can be appreciated. In this figure, both the MC true momentum variables and the reconstructed ones are visualised.

6.2.1 MOMENTUM SCALE (S_p)

The momentum scale, as said, is the bias on the measurement of each track's momentum. It can be affected by different sources such as the path of the track within the detector, the value of the momentum itself, etc. Thus, it is important to measure it on different channels with different decay topology and different momentum coverage for the reconstructed particles.

In order to measure its effect, as the bias affects in the same way the momentum of all the tracks, we can just parameterise it as a factor that multiplies to the p_T Armenteros variable, as in the momentum asymmetry its effect is cancelled.

$$
p_T \to p'_T = S_p \cdot pT\,. \tag{6.19}
$$

6.2.2 CHARGE ASYMMETRY (C_A)

The charge asymmetry usually arise as a side-effect of tracking sub-detectors mis-alignment. As in the case of the momentum scale, its effect may vary as a function of the track momentum, so using different decay channels may be interesting to over-constraint the result, obtaining an unbiased global correction parameter.

Contrary to the previous case, the effect of the charge asymmetry only impacts the α Armenteros variable and can be written as:

$$\alpha \rightarrow \alpha' = \frac{p_{L1} - C_A \cdot p_{L2}}{p_{L1} + C_A \cdot p_{L2}}, \tag{6.20}$$

remember that subscript 1 refers to the positive charge and 2 to the negative one. Defining in this way, a $C_A > 1$ would enhance the negative momentum and $C_A < 1$ would enhance the positive one.

6.2.3 PARENT MASS SCALE (S_M)

The primary objective of fitting the parent mass when obtaining the momentum scale is to address the commonly inadequate knowledge of it, which typically imposes limitations on the accuracy of the momentum scale determination. In addition, it can also take into account the final-state radiation that directly affects the reconstructed parent mass. Usually, the final-state radiation is modelled by the MC simulation; however, the simulation accuracy is on the level of 5×10^{-5} [17]. Thus, for very precise measurements, this may be a limitation. By adding this fitting parameter, we could avoid this limitation. With this in mind, we are going to present how to scale the parent mass on the Armenteros variables. The obvious definition of the scaling is:

$$M \rightarrow M' = S_M \cdot M\,. \tag{6.21}$$

Before going further, one must identify what to do, as, in order to properly scale the mass, one must undo the boost and go back to the CM system. Then, we want to identify the impact of the scaling in the CM frame. Nevertheless, it is even more interesting to identify the invariant quantities in the mass scaling process; the θ^* angle between the boost and the child momentum in the CM remains invariant. From Eq. 6.8:

$$\sin\theta^* = \frac{p_T}{p^*} = \frac{p'_T}{p^{*\prime}}\,. \tag{6.22}$$

Thus, the first quantity we want to obtain the impact of the scaling on is p^*; from Eq. 6.6:

$$p^{*2} \rightarrow p^{*\prime 2} = \frac{1}{4M'^2}(M'^4 + m_1^4 + m_2^4 - 2M'^2(m_1^2 + m_2^2) - 2m_1^2 m_2^2). \tag{6.23}$$

Consequently,

$$p_T \rightarrow p'_T = \sin\theta^* \cdot p^{*\prime 2}\,. \tag{6.24}$$

In order to prevent having a nonphysical value on p_T due to the resolution effect, we compute the mass scaling on the TRUE variables from the MC simulation. Then, we smear the data for each p_T and p_{Li} taking the information of the resolution as the difference between the reconstructed and TRUE momentum variables:

$$\Delta p_{(T,Li)} = p^{reco}_{(T,Li)} - p^{TRUE}_{(T,Li)}\,. \tag{6.25}$$

We already obtained the mass-scaled p_T Armenteros variable, as it remains invariant on the Lorentz boost, we now need to boost the longitudinal components of the decay. We then need to obtain the new parent velocity to use in the boost.
We fix the parent momentum value before the scaling (p_M), this is because the parent momentum depends on the hadronisation process and the quark momentum from the $p-p$ collision, so this value should not depend on the mass scaling. Thus, we recompute the velocity using the scaled mass as:

$$\beta \rightarrow \beta' = \frac{p_M}{E'} = \frac{p_M}{\sqrt{(p_M)^2 + (M')^2}}\,. \tag{6.26}$$

Then we can recompute γ to be used in the boost:

$$\gamma \rightarrow \gamma' = \frac{1}{\sqrt{1-\beta'^2}}\,. \tag{6.27}$$

Finally, from Eq. 6.9, using $P^*_L = p^*\cos\theta^*$, and, applying the relativistic energy relation, we can recompute the longitudinal components of the child particles on the TRUE MC information:

$$p_{L,i} \rightarrow p'_{L,i} = \gamma'\left(\beta'\sqrt{p^{*\prime 2} + m_i^2} \pm p^{*\prime}\cos\theta^*\right)\,. \tag{6.28}$$

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The last thing to do is to apply the resolution to the children momentum variables and build the mass-scaled Armenteros-Podolanski plot with the resolution effects from MC.

$$\begin{aligned} p_T^{\prime,reco} &= p_T^{\prime} + \Delta p_T \,, \\ p_{Li}^{\prime,reco} &= p_{Li}^{\prime} + \Delta p_{Li} \,. \end{aligned} \tag{6.29}$$

Obtaining the mass-scaled Armenteros variable α':

$$\alpha \rightarrow \alpha^{\prime,reco} = \frac{p_{L1}^{\prime,reco} - p_{L2}^{\prime,reco}}{p_{L1}^{\prime,reco} + p_{L2}^{\prime,reco}} \,. \tag{6.30}$$

In this proof of principle, we use the dispersion information in Eq. 6.25 from each event. As we smear the parent mass, which affects the momentum value of the child particles, the resulting dispersion might not match the one given by the event. This is because the momentum resolution depends on the momentum value. As we are smearing this factor a small quantity during the fit, we will not correct the dispersion value. The appropriate dispersion value could be obtained from the dispersion distribution as a function of the momentum on the MC data. This can be done, for example, using a KDE algorithm.

6.3 THEIL INDEX MAXIMISATION FITTING METHOD

Once we know how to generate the experimental effects we want to fit in our MC, we need a fitting method that allows us to use these procedure to obtain our parameters of interest, in this case S_p and S_M. As during the real calibration of a particle-physics detector, very naive selections are applied to data; we try to mimic this situation by using minimum bias samples. In our real data samples a mixture of decays $K_S^0 \rightarrow \pi^+\pi^-$ and $\Lambda^0 \rightarrow p\pi^-$ is present. We use two different dedicated stripping lines to fit each decay in order to enhance the contribution we are interested on. The MC simulation samples mimics the signal of interest and the data selection process.

To achieve a great performance on this polluted data, we need a robust method when some outliers are present. Great performance was obtained when using the Theil index as the fit estimator.

6.3.1 THEIL INDEX

The Theil index [19] is an estimator that measures, in terms of entropy, the inequality of a system. In other terms, it measures how far is a system to the situation where all the individuals have the same value of the variable of interest. It is widely used in economics to measure economic inequality within a population.

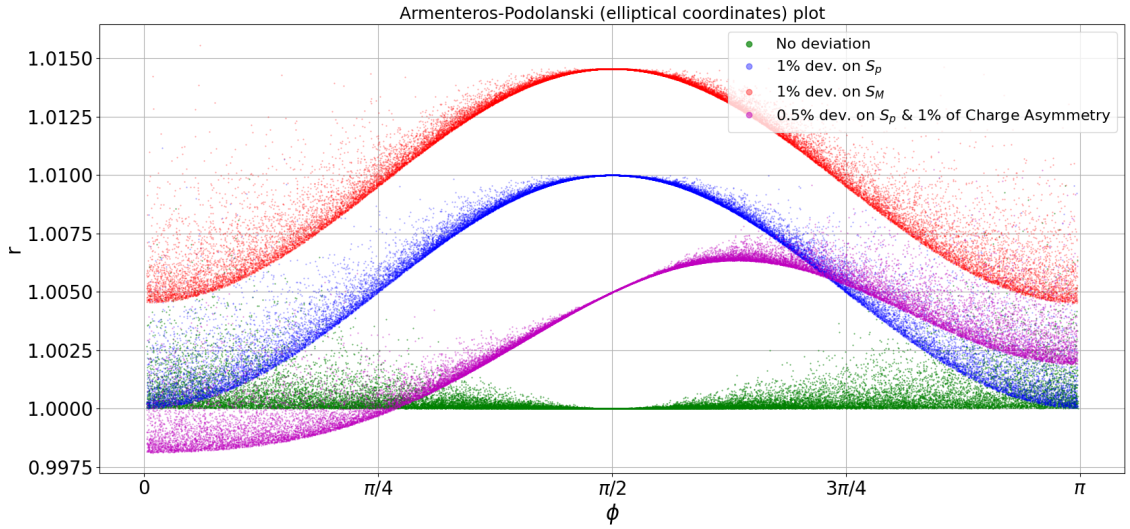
Our system of study is the binned 2D Armenteros modified space, where the different parts, or individuals, of the system are the different bins, and the variable of interest is the number of events in each bin. The Theil index estimator can be written as:

$$\begin{aligned} T &= \frac{1}{N \ln(N)} \sum_{i=1}^{N} \frac{x_i}{\mu} \ln(\frac{x_i}{\mu}) \,, \\ \mu &= \frac{1}{N} \sum_{i=1}^{N} x_i \,. \end{aligned} \tag{6.31}$$

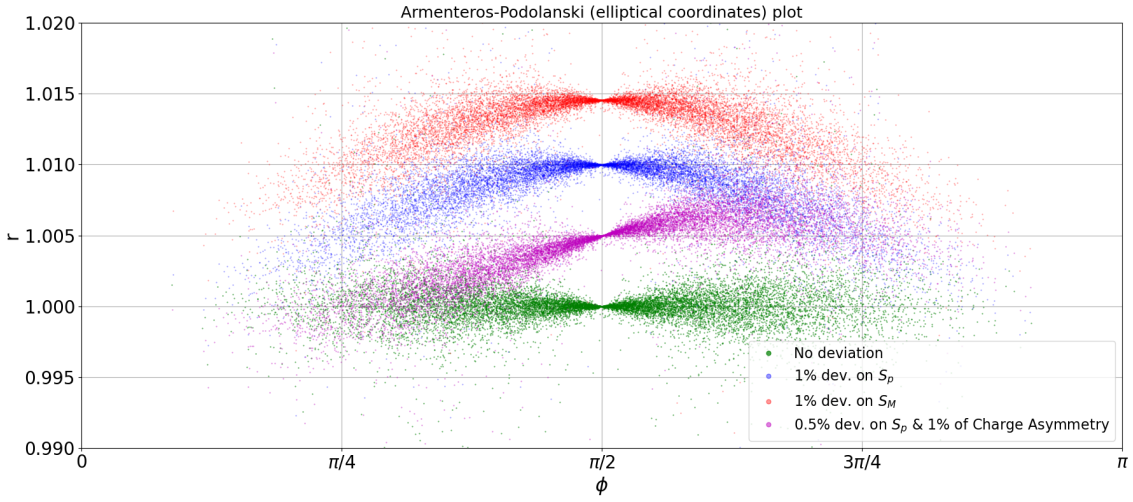
Where N is the number of bins, x_i is the number of events in each bin, and, μ is the mean of x_i. Described as such, $T \in [0, 1]$, where $T = 0$ means that the bins are equally populated and if $T = 1$ all the counts fall in a single bin.

6.3.2 FITTING PROCEDURE

The fits are performed by applying different variations to MC data, this modified MC data is added to the real data, we build the 2D histogram of the Armenteros plot and compute its Theil index for those fitting parameters. We generate the fitting parameters randomly, scanning the full parameter space during each fit. We do not use a minimisation algorithm because the Theil index retrieves discrete values. Consequently, using a minimisation algorithm on the Theil index would experience convergence issues. To scan the parameter space with a nice resolution, a lot of different parameters must be computed. As we want to



(A) The MC data using true momentum values is shown. In this plot the effect of the dispersion due to the ultra-relativistic approximation can be appreciated. An uniform acceptance in θ^* is used.



(B) LHCb MC data using the reconstructed momentum values is shown. In this plot the effect of the reconstruction smears the data points making more difficult to differentiate which experimental effect has been applied. The data-points follow the LHCb acceptance, thus, the distribution on ϕ is not uniform.

FIGURE 6.2: Armenteros-Podolanski space with elliptical coordinates. The coordinate change parameters match the ellipse parameters with no deviations applied.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

obtain the fit result in a reasonable time, the *CUDA* programming language that runs on a *NVidia* GPU is used; computing each fit iteration on a kernel of the GPU.

The binned Armenteros modified space is used, the variable change parameters match the mass parameters used on the MC simulated sample. The binning range is selected so that it covers the signal in data. The number of bins is limited by the GPU memory, we use a grid of 140×140 bins. The fitting parameters S_p and S_M are defined so that all the modified MC data falls within the binning range, we want to have the same amount of data in each fit iteration to rely on the Theil index estimation. These fitting parameters are generated randomly in the parameter space for each fit iteration. The number of fit iterations is $2000 \times BLOCKSIZE$, and $BLOCKSIZE = 256$. The $BLOCKSIZE$ defines how the information is sent to the GPU. Thus, there are 512000 iterations in each fit.

For each fit iteration the Theil index of the 2D histogram is computed and, finally, the parameters that retrieve the highest Theil index value are selected as the fit result. In this case, the MC data best overlap the real data points, populating the same histogram bins and thus achieving the highest inequality in the histogram system.

6.3.3 STATISTICAL UNCERTAINTIES ESTIMATION

The only uncertainty we will take into account is the statistical one. Other sources of uncertainty could be related to binning scheme, number of iterations in each fit, number of MC events, etc.

In order to compute the statistical uncertainties, we split the sample into 25 different subsamples of the same size. For each subsample, we repeat the fit with the exact same configuration and compute the value of our parameters S_p and S_M. Finally, we obtain a Gaussian-like distribution for our parameters of interest, where the statistical uncertainty is the uncertainty on the mean of the distribution for each parameter. In addition, we also check that the mean value of the distribution is compatible with the value obtained on the fit to the whole sample.

6.4 RESULTS

Now we are going to present the results of this method on the $K^0_S \to \pi^+\pi^-$ and $\Lambda^0 \to p\pi^-$ decay channels. Only long tracks were used on this study, as they retrieve the best accuracy on the momentum determination and all the tracking sub-detectors are involved on the reconstruction. Only the momentum and mass scale parameters are fitted. In addition, the samples are form the 2018 data-taking period with the down magnet polarity.
The masses for the decays used in the simulated MC samples and on the variable change parameters of the modified Armenteros space are listed in Table 6.1. The Armenteros-Podolanski plot of the real data is shown in Fig. 6.3

Particle	**Mass** [MeV/c^2]
K^0_S	497.611
Λ^0	1115.683
p	938.27208
$\pi^\pm$	139.57061

TABLE 6.1: Values of the input masses used for the momentum scale determination.

6.4.1 $K^0_S \to \pi^+\pi^-$

The fit result is shown in Fig. 6.4, the modified Armenteros space is used so that the variable change parameters match the mass parameters used on the MC simulation. In the plot of Fig. 6.4A it can be seen that the number of events in the MC sample is lower than in real data. Nevertheless, a nice performance of the fit is achieved. In Fig. 6.4B the scan of the parameters space is shown, as well as the fit result. In Fig. 6.5 the uncertainties estimation is show for both fitting parameters. In Table 6.2 the values obtained in the fit result are shown, where we can see that the relative uncertainty obtained from the fit is of the same order as the accuracy reported by the LHCb collaboration in 2024 [17].

On the fit scan in Fig. 6.4B, a high correlation between both parameters can be appreciated. The main reason is because, when the detector acceptance is applied, as shown in Fig.6.2B, both experimental effects

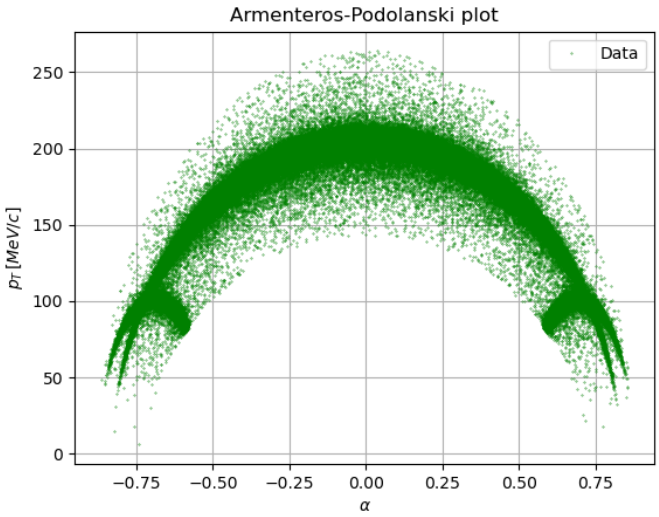


FIGURE 6.3: Armenteros-Podolanski plot of a real data sample. It shows the $K^0_S \to \pi^+\pi^-$ ellipse, symmetric with respect to $\alpha = 0$, and the two $\Lambda^0 \to p\pi^-$ ellipses.

have a similar signature on the Armenteros plane. The mass scale factor separates the scaled ellipse from the original one whereas the momentum scale applies a similar distortion but does not affect to the $\phi = 0$ and $\phi = \pi$ regions of the plot, as shown in Fig. 6.2A. As these regions are out of the LHCb acceptance, because these are cases where one of the particles have very low momentum, i.e. high asymmetry; the two effects are more difficult to differentiate with few MC events.

S_p	$u(S_p)$	**Rel. Unc.**	S_M	$u(S_M)$	**Rel. Unc.**
0.98888	0.00062	6.3×10^{-4}	1.00697	0.00041	4.1×10^{-4}

TABLE 6.2: Fit result for the $K^0_S \to \pi^+\pi^-$ decay. The value and uncertainty, along with the relative uncertainty of both fitted parameters, are shown.

6.4.2 $\Lambda^0 \to p\pi^-$

In Fig. 6.6 the fit result is shown, the Armenteros modified space appears in Fig. 6.6A and the scan of the parameter space with the Theil index in Fig. 6.6B. The uncertainties calculation is shown in Fig. 6.7 and the measured values are reported in Table 6.3. Here we can see that the accuracy on the momentum scale is slightly improved from the previous fit but the mass accuracy is greatly improved. This improvement in the mass scale is not only due to the higher MC statistics but also due to the kinematics of the decay, as the lower the p_T of the decay, the more sensitive are the ellipse parameters to the parent mass.
In this fit, the statistics of the MC are closer to the data size, achieving a similar performance to the previous fit, although the parameters obtained are less correlated in this case. To further investigate the impact of the MC statistics on the result, some systematic study should be performed. Furthermore, the two Λ^0 and $\bar{\Lambda}^0$ ellipses are summed to double the statistics, thus, no charge asymmetry could be spotted from this fit.

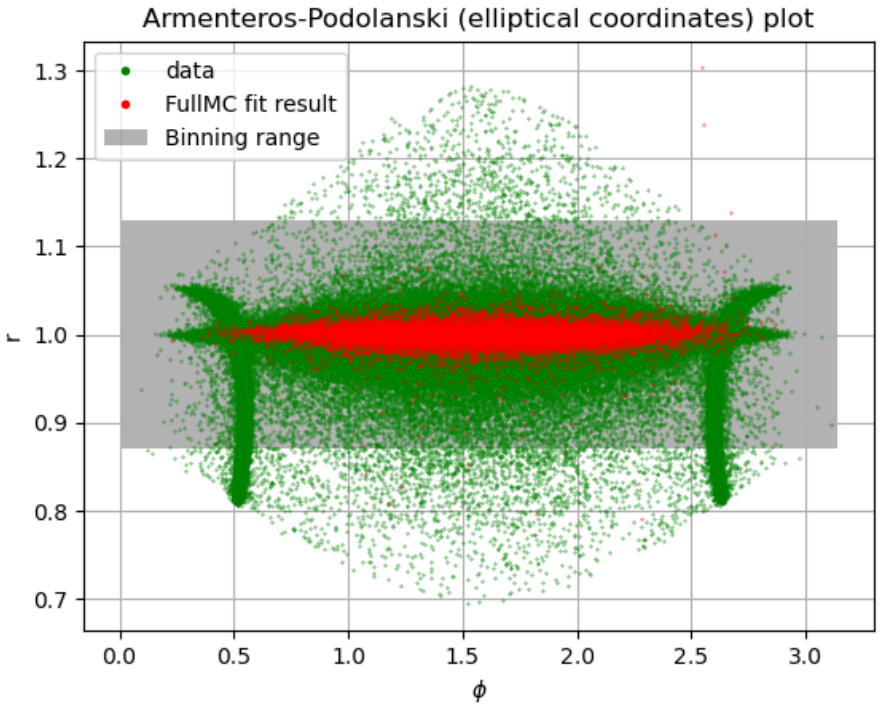
S_p	$u(S_p)$	**Rel. Unc.**	S_M	$u(S_M)$	**Rel. Unc.**
0.99723	0.00057	5.7×10^{-4}	1.000126	2.6×10^{-5}	2.6×10^{-5}

TABLE 6.3: Fit result for the $\Lambda^0 \to p\pi^-$ decay. The value and uncertainty, along with the relative uncertainty of both fitted parameters, are shown.

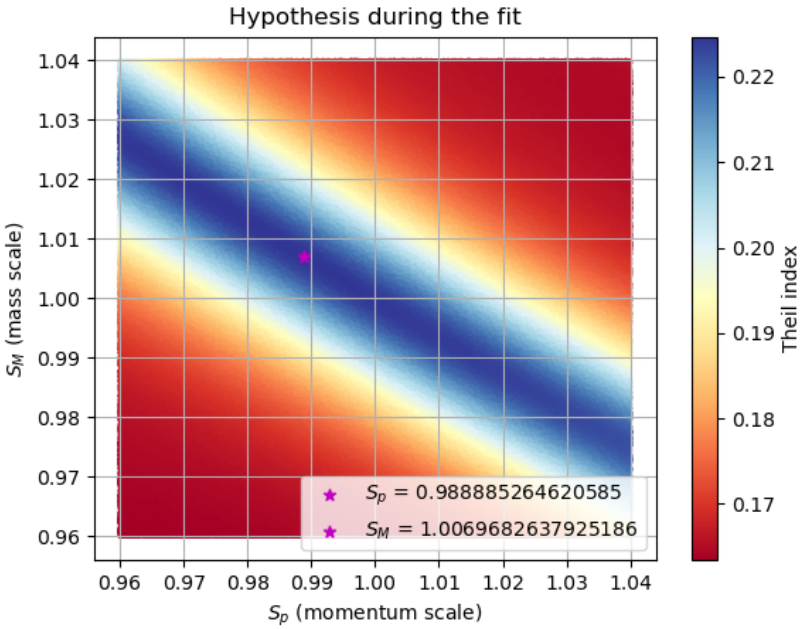
6.5 DISCUSSION

USC UNIVERSIDADE DE SANTIAGO

In this proof of principle we have shown a completely new method to extract the momentum scale bias of a particle physics detector. We have obtained very promising results with very limited data and computing



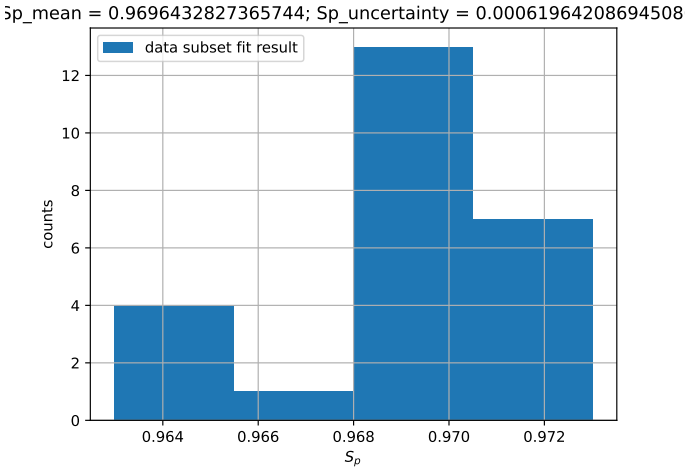
(A) Armenteros modified space showing the fit configuration, the real data and the MC sample scaled by the fit result parameters.



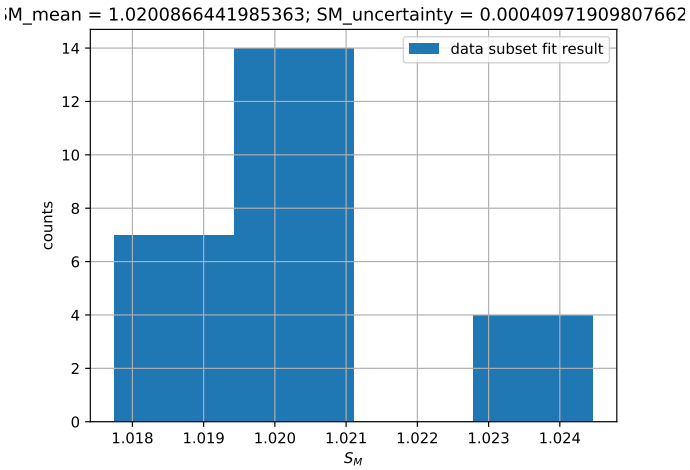
(B) Parameter space obtained during the fit showing the result.

FIGURE 6.4: $K^0_S \to \pi^+\pi^-$ fit result.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) S_p parameter.

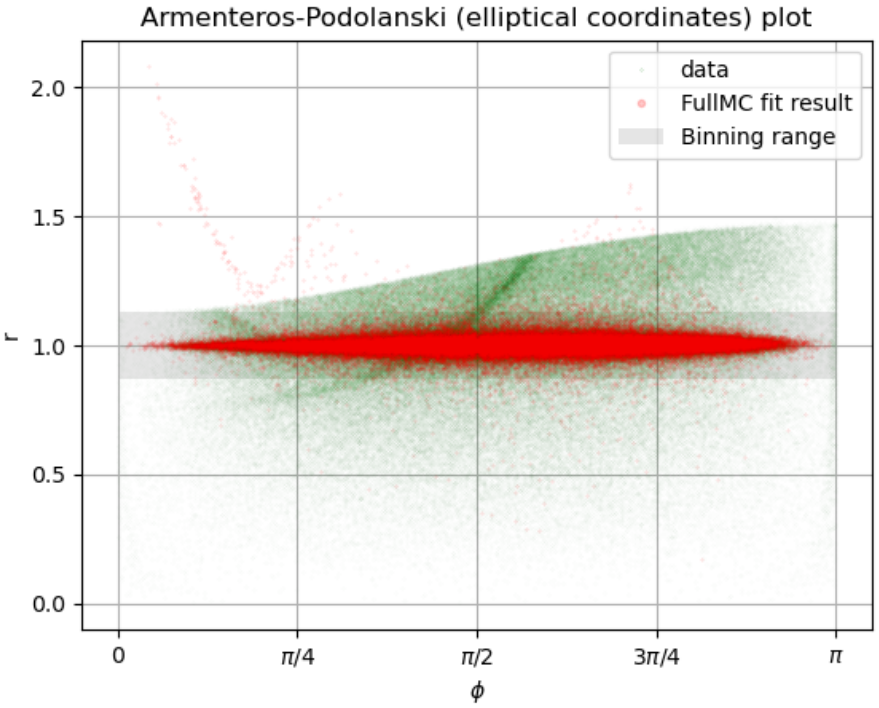


(B) S_M parameter.

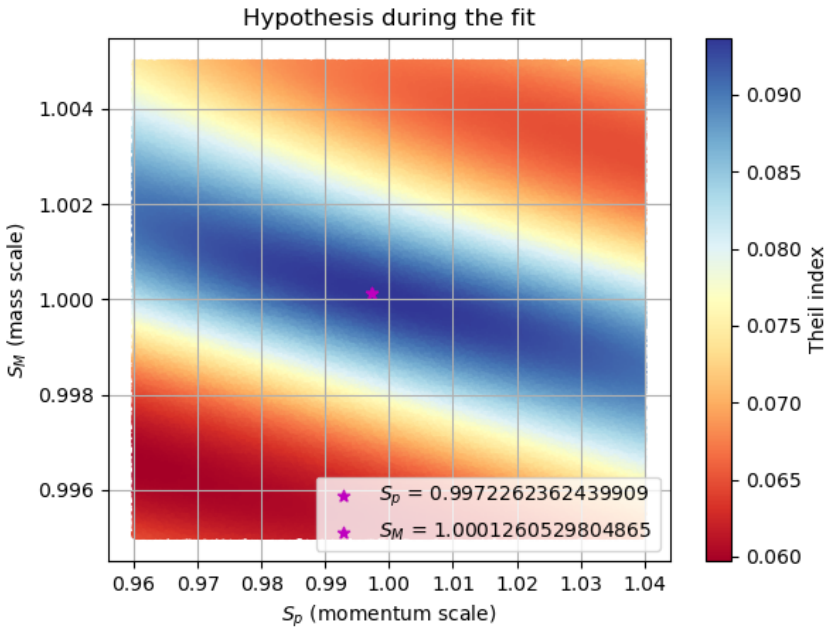
FIGURE 6.5: $K^0_S \to \pi^+\pi^-$ statistical uncertainties calculation.

resoures; achieving a similar performance as the LHCb in 2024 [17]. Given that in the Armenteros plot there is no input parameters needed to build the decay signature, we can fit at the same time all the masses that participate in the decay. With this is mind, we can parameterise the detector effects on the Armenteros variables and fit the limiting parameters; obtaining the highest possible accuracy on the momentum scale determination. The biggest novelty of this method is the fact that we can fit, at the same time, the parent mass and the momentum scale. In this way, we can avoid the limiting factor due to the parent mass precision and, also, take into account possible mis-modellings due to the final-state radiation on the MC. The best performance in determining the momentum scale bias was achieved using the $\Lambda^0 \to p\pi^-$ decay, where an accuracy of 5.7×10^{-4} was obtained. When compared with the current method used at LHCb, which has an accuracy of 3×10^{-4}, that is twice the performance we achieved; we can then conclude that our result is very competitive with state-of-the-art approaches. This suggests that it would be worthwhile to investigate this method further, potentially exploiting other statistical tools such as the maximum likelihood approach, which could yield optimal statistical performance. However, it is also necessary to investigate the sources of systematic uncertainty in more detail, understanding where the limitations of the method may arise and how computing power and the amount of statistics can limit the method's performance.

In the Run 3 data-taking period at the LHC, the LHCb detector, with the removal of the hardware trigger, can reconstruct particles at very low p_T, this means that we can asses the region close to the beam pipe where a big number of strange particles are produced [50]. Thus, we expect to have large samples of



(A) Armenteros modified space showing the fit configuration, the real data and the MC sample scaled by the fit result parameters.

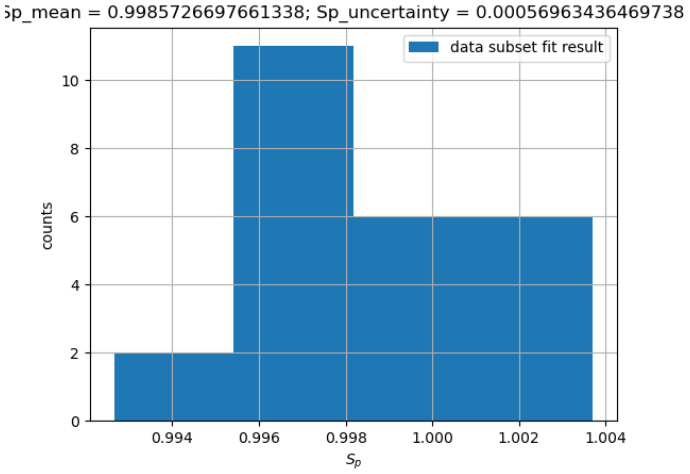


(B) Parameter space obtained during the fit showing the result.

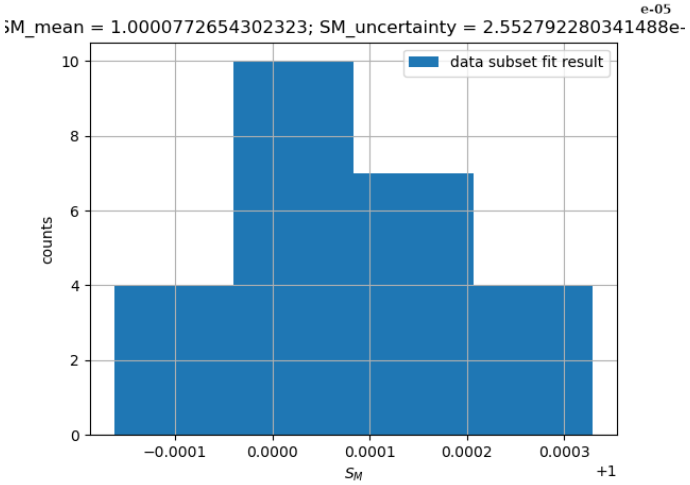
FIGURE 6.6: $\Lambda^0 \to p\pi^-$ fit result.

K^0_S and Λ^0 decays. In addition, as these decays have a low p_T at the center of mass, and the value of the boost is very high due to the energy of the colliding protons, the child particles are produced in the forward direction. As a result, we can select specific regions of the detector and compute some "local" momentum scale, being able to spot effects of the momentum scale bias on particular regions of the detector.

In addition, decay signatures at different mass scales can be exploited to obtain the momentum scale in the modified Armenteros space on different decays. For example, the very precise $J/\Psi \to \mu^+\mu^-$ decay



(A) S_p parameter.



(B) S_M parameter.

FIGURE 6.7: $\Lambda^0 \rightarrow p\pi^-$ uncertainties calculation.

or the $\Upsilon \rightarrow B^0\bar{B}^0$ decay can be used. In the case of the Υ decay, the p_T in the center of mass frame is very low. Despite the challenges in reconstruction, the masses can be obtained with high precision in the Armenteros-Podolanski plot, resulting in an accurate determination of the momentum scale.

Finally, this method could be used on different physic studies where a high precision on the masses of the decaying particles is needed. This method may be interesting on measurements systematically limited, as this new method could improve the precision on the mass of the particles involved in the decay. An use-example could be the measurement of the weak boson $W^{\pm}$ mass. Moreover, the method could be also exploited on random searches for two-body decays.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 7

CONCLUSION

To conclude this thesis, the different achievements during these years will be highlighted. The research work was mainly developed on the $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ decay channel at the LHCb experiment. However, it was not the only time-consuming task performed during the thesis. This task involved many different aspects of the learning process. A profound understanding of the detector and how it works was crucial to develop new strategies for the preparation of future data takings with more challenging conditions. In addition, to perform the analysis, a deep background in computation, physics, and statistics is needed. Without proper knowledge, it is impossible to understand where the many problems arise and how to face them, despite sometimes looking for the perfect solution being the biggest enemy of the solution itself.

A different task was developed on the more technical side, where a new method to measure the momentum scale bias was also investigated. Despite this research not only including the work on this thesis, as a preliminary version of the method on toy data with no consideration of experimental effects was investigated previously [20]. In this thesis, this method was adapted to the LHCb experiment, developing a fitting method where instead of a functional form, MC data are used to fit the real data. This eliminates the process of parameterising the MC data and allows for per-event determination of the parameters, which may be interesting to study features of the simulation that do not match the performance of the real data, such as photon radiation. Moreover, this method showed very promising performance in a sub-optimal case, which indicates that further exploiting this method might be very worthwhile.

Finally, in this thesis, the central work was included in the narrative of the CKM angle α at LHCb. Despite the central work presented on this strategy, that is the time-dependent flavor-tagged amplitude analysis of $B^0 \to a_1^{\pm} \pi^{\mp}$ was not performed yet, it did not stop us from presenting the strategy and clarifying the context of the measurement presented in this thesis. Laying the foundations for the α measurement at the LHCb experiment.

Nevertheless, some work was introduced on the $B^0 \to \pi^+ \pi^- \pi^+ \pi^-$ channel, performing the time-integrated analysis studying both the axial and vector contributions but focusing on the observables in the VV mass window [96], as an starting point for the more complex analysis.

In addition to the branching fraction, the preliminary work on the amplitude analysis of the $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ decay was also performed, despite not being deeply presented here, as the complexity of the amplitudes in the mass window is very high and makes the amplitude analysis fit very difficult to parameterise and to perform.

What was indeed performed is the branching fraction analysis.

In this analysis, we have presented a branching fraction measurement in an inclusive charmless mass window of the $B^+ \to K^0_S \pi^+ \pi^- \pi^+$ decay. This is the first measurement of this decay in this particular mass window performed by any experiment.

We employ an innovative method to compute the efficiencies by using a flat PHSP MC simulation combined with a data-driven model to determine the signal amplitudes. This approach allows us to manage the influence of data amplitudes on the efficiency, both in mass and angle projections of the PHSP, without relying on any simplistic model that could bias the efficiency calculation.

The major drawback of this method is the inability to disentangle differences between data and MC due to the signal amplitudes from those caused by incorrect simulation of signal acceptance, i.e., efficiency.

Additionally, a significant source of systematic uncertainty in the data-driven method arises from the various hyper-parameters that can be used to extract the signal amplitudes and the efficiency distributions in the PHSP, due to the lack of best-fit criteria in multidimensional spaces. Arbitrary choices of these parameters can bias the results, so we consider multiple potential values for these parameters when computing systematic uncertainty sources. However, this systematic uncertainty is still less significant than the uncertainty introduced by the simplistic signal models assumed in previous similar analyses. Furthermore, different approaches, such as KDE, can be explored to reduce this systematic uncertainty. It is also important to note that the analysis mass window complicates the extraction of signal amplitudes because it involves different disjoint regions of the PHSP. If the entire PHSP were analysed, the data-driven models could be trained on the entire space without needing this splitting, resulting in smoother distributions of the PDFs of interest. However, this approach would complicate the extraction of the primary goal of this analysis, which is determining the δ_{strong} difference between the K_1^0 and a_1^+ resonances and their branching fractions through an amplitude analysis.

In conclusion, the measurement presented in this thesis is highly intriguing and incorporates numerous innovative developments. Additionally, it is the starting point of the amplitude analysis that can be performed in the same data with the same efficiency parameterisation; using the generator level MC as the normalisation sample for the amplitude analysis. Moreover, the method presented in this analysis for extracting signal amplitudes can be similarly applied to control the backgrounds affecting the amplitude analysis.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 8

RESUMO

Nesta tese explóranse novos enfoques para a análise de datos na física de partículas, tanto no ámbito de análises para obter parámetros físicos interesantes na teoría, coma métodos para mellorar o funcionamento do detector.

Estes enfoques céntranse principalmente no uso de datos para transformar as simulacións de tal xeito que poidan ser corrixidas por diferentes efectos, xa sexa a través de parametrizacións ou a través de métodos estatísticos non paramétricos que permiten obter información directamente dos datos. Estes métodos son cruciais para obter a medida sen sesgo da eficiencia de procesos físicos non coñecidos, xa que non se asumen modelos simplificados para a simulación destes procesos.

O traballo principal da tese céntrase na análise da frecuencia de decaemento do mesón B^+ en $K^0_S\pi^+\pi^-\pi^+$, é dicir, cantas veces sucede ese decaemento en relación a todos os decaementos do mesón B^+. O proceso de decaemento pode expresarse como $B^+ \to K^0_S\pi^+\pi^-\pi^+$, onde as partículas do estado final son mesóns, constituídos por dous quarks.

Os quarks son as partículas fundamentais que constitúen a materia segundo o modelo estándar da física de partículas. Os quarks xúntanse de dous en dous para formar mesóns (coma os pións e kaóns antes presentados) ou de tres en tres para formar barións. Os barións, como os protóns e neutróns, constitúen a materia. Ademais dos quarks, os leptóns son a outra peza fundamental que constitúe a materia, como o electrón ou o neutrino. Por último, para conter a todo o modelo estándar, precisamos os bosóns. Os bosóns son as partículas que median as interaccións, son os responsables das diferentes forzas observadas na natureza. Existen seis bosóns no modelo estándar: tres bosóns que median a interacción feble, un para a interacción electromagnética, outro para a interacción nuclear forte e finalmente, o bosón de Higgs, que lle dá masa ás partículas.

O modelo estándar, a pesar de predicir con moita precisión e explicar con moito detalle certos procesos físicos, é incapaz de predicir correctamente certos patróns na natureza, como a asimetría observada entre barións e anti-barións no universo. As antipartículas teñen as mesmas propiedades que as partículas pero teñen carga e paridade oposta. Ademais, o modelo estándar non inclúe a gravidade, tampouco é capaz de explicar a materia e a enerxía escuras, así coma a oscilación dos neutrinos.

Polo tanto, estudar en profundidade o modelo estándar é crucial para entender cales son os seus problemas e conseguir arranxalos.

Para estudar o modelo estándar da física de partículas empréganse os colisores de partículas. Nestes colisores, as partículas aceléranse a velocidades moi cercanas ás da luz no baleiro e fanse colidir. Destas colisións emerxen novas partículas que decaen rapidamente. Os produtos do decaemento destas partículas son capturados polos detectores situados ao redor do punto de colisión. Deste xeito, reconstruíndo a información do que sucedeu durante a colisión grazas aos detectores, pódense analizar moitos decaementos similares para, a través do uso da estatística, sacar conclusións sobre os procesos físicos que acontecen. Os resultados obtidos pódense comparar a posteriori co predito polo modelo estándar ou outras teorías para rexeitalas ou aceptalas.

Dende o s.XX construíronse moitos aceleradores e detectores para analizar os decaementos de mesóns B, o instrumento empregado para medir as colisións que foron analizadas nesta tese é o experimento LHCb situado no CERN. Os mesóns B prodúcense a partir de colisións de protóns cunha gran enerxía.

USC

A análise plantexada nesta tese atópase dentro dun plan para medir o ángulo α da matriz de Cabibbo-Kobayashi-Maskawa (CKM). Este ángulo é unha medida da violación CP no modelo estándar, que é un dos procesos que explican a asimetría materia-antimateria no universo. A simetría CP refírese ao comportamento dun determinado proceso cando se cambia a carga e paridade das partículas intervintes. É dicir, cando un proceso e o proceso formado polas antipartículas non ocorren coa mesma probabilidade, dícese que a paridade CP é violada.
O ángulo α é o ángulo menos coñecido dos pertencentes a un triángulo que describe a violación CP no modelo estándar en decaementos de mesóns B. Como os outros dous ángulos son coñecidos con maior precisión, os parámetros de violación CP dentro da teoría están completamente constreñidos. O interese de medir o ángulo α radica en que é un xeito de constatar a consistencia do mecanismo da matriz CKM para estudar a violación CP. Estudar con máis detalle este ángulo pode desentrañar algún dos principais problemas do modelo estándar.

A maiores da análise mencionada, nesta tese tamén se inclúe o traballo relacionado co desenvolvemento do sistema de selección e almacenado dos datos producidos polo detector. Realizouse un estudo para decidir cal é o mellor xeito para detectar e almacenar os decaementos que son interesantes para a medida de α. Este estudo decidiu cal ía ser a estratexia empregada durante a toma de datos que está a acontecer durante a escritura desta tese no LHC.
Ademais, tamén se desenvolveu un método para a calibración da escala do momento do detector co emprego do diagrama de Armenteros-Podolanski. Este método permite obter ao mesmo tempo o sesgo na medida da masa e na medida do momento. Isto é fundamentalmente diferente a métodos actuais que empregan o sesgo no valor de masas coñecidas para corrixir a escala de momento. A principal vantaxe deste método é que a limitación coa que se coñece a masa da partícula empregada como referencia pode ser eliminada.

8.1 FENOMENOLOXÍA

A análise plantexada pretende constreñir o ángulo α da matriz CKM definida no marco do modelo estándar. Esta matriz infórmanos sobre a probabilidade de transición dunha familia de quarks a outra a través dos procesos de correntes cargadas que cambian o sabor. O sabor é unha propiedade que nos informa a que familia pertence unha partícula, habendo tres principais no sector dos quarks: a primeira familia, formada polos quarks u e d, que constitúen os protóns; a segunda familia, formada por quarks c e s; e a terceira familia formada por t e b.
Ademais, os quarks que se propagan como estados físicos de masa e os quarks que interaccionan a través de correntes cargadas para cambiar o seu sabor non son exactamente iguais. Relaciónanse a través dunha matriz de rotación no espazo de sabor. Esta matriz é a que dá lugar á existencia da matriz CKM na cal existe unha fase que cambia de signo ao facer a transformación CP, explicando así a violación CP dentro do modelo estándard. A matriz CKM pode ser escrita como:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.042 \\ 0.009 & 0.041 & 0.999 \end{pmatrix} \tag{8.1}$$

Se impoñemos a condición de unitariedade a esta matriz, chegamos a unha serie de relacións, entre as cales nos interesa a seguinte:

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

Onde os diferentes termos V representan os diferentes elementos da matriz CKM que informan sobre a probabilidade de diferentes transicións de quarks. Desta expresión, podemos obter un triángulo no plano complexo, coñecido como o triángulo unitario do sector do mesón B. Neste triángulo, o ángulo α pódese medir a partir do estudo da violación CP na interferencia entre a oscilación do mesón B^0 e o seu decaemento a mesóns formados por quarks da primeira xeración. Uns exemplos de canles onde estudar o ángulo α son: $B^0 \to \pi\pi$, $B^0 \to \rho\rho$ e $B^0 \to \rho\pi$. Onde ρ é outro mesón que decae moi rapidamente en dous pións (π).
Actualmente, o valor de α obtense da análise destes procesos onde aparecen moitas posibles solucións. Escaneando en función de α as diferentes solucións pódese obter un valor de α combinado por todas as medidas.

No noso caso, o valor de α queremos obtelo da análise da canle $B^0 \to a_1\pi$, onde a_1 é un mesón axial que decae a tres pións. A vantaxe de estudar esta canle é que non é necesario reconstruír ningunha partícula

neutra, como o pión neutro. Grazas a isto, pódese obter unha mellor precisión na medida de α en LHCb en decaementos a catro corpos.

A principal complicación engadida que ten o estudo do ángulo α é que o proceso do cal se obtén o ángulo non é o único que sucede na natureza. Existe un proceso chamado 'pingüín' que contamina a extracción do ángulo xa que contribúe cunha fase complexa diferente ao proceso de estudo. Ademais, este proceso pingüín ten unha probabilidade de suceder similar ao proceso principal de estudo no ángulo α, polo tanto, extraer a súa contribución é fundamental nesta medida.

Neste punto é onde entra a canle de decaemento presentada nesta tese, xa que se relaciona a través da simetría $SU(3)$ coa canle do mesón B neutro. Esta simetría asume que os quarks u, d e s son iguais. Polo tanto, a través do estudo da canle $B^+ \to K^0_S\pi^+\pi^-\pi^+$, que é un proceso puramente de tipo pingüín, pódese extraer a información para controlar esta contaminación na canle $B^0 \to a_1^\pm\pi^\mp$. Os procesos análogos que se queren estudar son $B^+ \to K^0_{1,A}\pi^+$ e $B^+ \to K^0_S a_1^+$, sendo o estado final de ambos procesos o sinalado previamente.

Esta análise das resonancias axiais xa foi plantexada previamente, pero nesta tese imos un pouco máis aló, xa que no noso estudo non só empregamos a información sobre a frecuencia de decaemento, o obxectivo é realizar tamén unha análise de amplitudes de onde se pode obter a información da fase coa que contribúen os diagramas de pingüíns neste proceso. Aportando así unha metodoloxía moito máis exhaustiva a pesar da necesidade de controlar os efectos da rotura da simetría $SU(3)$.

Así, o estudo da frecuencia de decaemento na canle $B^+ \to K^0_S\pi^+\pi^-\pi^+$ nunha rexión do espazo fásico que permita obter unha medida da diferenza de fase entre as resonancias axiais, é fundamental para poder extraer, posteriormente, a contaminación presente no ángulo α medido na canle $B^0 \to a_1^\pm\pi^\mp$ a través dunha análise de amplitudes dependente do tempo e na que o sabor do mesón decaendo é inferido.

8.2 O EXPERIMENTO LHCB

Os datos empregados para realizar o traballo presentado na tese foron recolectados polo experimento LHCb durante o período dende 2011 ata 2012 e máis dende 2015 ata 2018. Nese período de tempo mediuse unha luminosidade correspondente a $9\,\text{fb}^{-1}$ procedente de colisións entre protóns a moi alta enerxía, acelerados dentro do LHC, o anel principal que forma parte do conxunto de aceleradores de partículas situados no CERN.

O LHCb é un espectrómetro de masas constituído por un conxunto de detectores que permiten medir o momento e a masa de partículas cargadas, así como a enerxía das partículas xeradas dentro da aceptancia do detector. A aceptancia cobre a rexión frontal da colisión dos protóns, onde se producen a maioría dos quarks b emerxentes da colisión.

Dende o punto de vista da medida da frecuencia relativa de decaemento, o estado final de estudo está formado por hadróns cargados. Polo tanto, un bo funcionamento do sistema de detección de partículas cargadas é fundamental, tanto para a identificación de partículas, como para a reconstrución das trazas a partir das cales a medida do momento pode ser inferida.

Para a medida das trazas dentro do detector, empréganse unha serie de módulos e un imán que curva a súa traxectoria. O primeiro destes módulos, chamado VeLo, sitúase ao redor do punto de colisión dos protóns, de tal xeito que se poden facer medidas moi precisas do tempo de decaemento dos mesóns B, que teñen unha vida media curta. Para partículas de vida media máis longa, como o K^0_S, ás veces estas partículas poden decaer fóra da aceptancia do VeLo. Deste xeito, na reconstrución destas trazas, non se inclúe información deste primeiro módulo. O resultado son trazas reconstruídas cunha menor precisión. Polo tanto, durante a análise vanse estudar dúas categorías por separado dependendo de se os K^0_S decaeron dentro ou fóra do VeLo.

A través dun complexo proceso de selección no que interveñen diferentes aspectos do detector, e no cal unha alta capacidade de computación é necesaria, os datos interesantes para ser analizados son almacenados. Durante este proceso é necesario entender o comportamento dos diferentes detectores no seu conxunto, calibralos para obter o mellor funcionamento posible dos mesmos, e combinar toda a información para poder acadar os criterios necesarios de selección para unha almacenaxe óptima dos datos. Relacionado con este proceso de almacenado dos datos, fíxose un estudo para decidir cales son os criterios óptimos para almacenar os datos relacionados coa medida de α; estes son os relacionados cos decaementos $B^0 \to \pi^+\pi^-\pi^+\pi^-$ e $B^+ \to K^0_S\pi^+\pi^-\pi^+$. Fíxose un estudo para atopar os criterios que, respectando a capacidade de almacenaxe dedicada á medida de α, se obtivese unha boa aceptancia para as resonancias

axiais presentes nestes decaementos. Decidindo así, cal vai ser a estratexia seguida na toma de datos que está acontecendo durante estes meses de toma de datos.

Finalmente, ademais da detección dos decaementos de interese, é moi importante simulalos do xeito máis semellante a como obtén os datos reais o detector para, deste xeito, poder describir o impacto do proceso de detección e selección no sinal de interese. Así, na simulación téñense en conta tanto os procesos físicos que se queren estudar, como os procesos físicos asociados coa detección e selección destas partículas. O obxectivo final da simulación na nosa análise é a extracción das eficiencias. A pesar de simular estes procesos do mellor xeito posible, sempre existen lixeiras diferenzas entre os datos e as simulacións. Na nosa análise empregamos diferentes ferramentas para arranxar as diferenzas na identificación de partículas, no seu trazado dentro do detector e tamén, diferenzas relacionadas co proceso de almacenado e selección dos datos.

8.3 METODOLOXÍA

Na metodoloxía da tese inclúense moitas estratexias e ferramentas diferentes para obter os obxectivos plantexados. Moitas destas ferramentas son empregadas en calquera tipo de análise de datos nun experimento de física de partículas, como son os axustes a datos a través do método de máxima verosimilitude (*Maximum Likelihood fit*) ou o uso de redes neuronais para suprimir algunhas das fontes principais de ruído que contaminan os datos.
Non obstante, nesta tese destaca o novidoso xeito de empregar ferramentas ben establecidas para reducir as fontes de incertezas sistemáticas asociadas cos procedementos empregados neste tipo de análises. A principal fonte de incertezas sistemáticas que se conseguiu minimizar é a incerteza asociada ao descoñecemento do modelo de amplitudes en análises nas que non se coñecen con precisión. Coma a análise presentada na tese emprega unha rexión do espazo fásico moi complexa, incluíndo tanto decaementos a dous corpos coma a tres corpos, as amplitudes que entran dentro desta rexión son moi difíciles de parametrizar, e ademais, nunca foron medidas con anterioridade. Neste tipo de situacións, para obter a eficiencia emprégase unha simulación da principal compoñente no espazo fásico e estímase a incerteza sistemática asociada na eficiencia. No noso caso, empregaremos un modelo dirixido por datos no cal extraemos directamente o modelo dos datos estimando unha eficiencia sen este sesgo. O maior sistemático deste procedemento será asociado ao método empregado para extraer o modelo dos datos; xeralmente, evalúase a través da variación dos parámetros do método que se escollen para obter o modelo.

Para presentar a metodoloxía e as problemáticas que se tiveron que resolver, fíxose un estudo comparativo de dous métodos dirixidos por datos no espazo fásico dunha desintegración a catro corpos. O primeiro que debemos presentar é o espazo fásico de catro corpos, como se define e como pode calcularse a integral para obter unha forma funcional analítica. Neste punto xa aparece o primeiro problema, xa que a integral para o caso de resonancias a tres corpos dentro deste espazo fásico non ten unha solución coñecida.
Deste xeito, e despois dunha discusión detallada, escolleuse o método no que non é necesario coñecer esta integral para obter un bo resultado da parametrización das diferentes distribucións no espazo fásico necesarias para obter o valor real da eficiencia.

8.4 MEDIDA DA FRECUENCIA RELATIVA DE DECAEMENTO DE $B^+ \longrightarrow K^0_S\pi^+\pi^-\pi^+$

Ésta medida trátase do traballo central da tese, o principal obxectivo é obter cantas veces sucede o decaemento do mesón B^+ ao estado final $K^0_S\pi^+\pi^-\pi^+$ na rexión do espazo fásico a analizar. Como xa se mencionou, esta rexión correspondese cos decaementos a dous ou tres corpos de mesóns que conteñen quarks máis lixeiros que o quark c, é dicir u, d ou s. Trátase dunha rexión moi específica que nunca foi analizada, polo tanto non existen resultados previos para poder facer unha comparación directa. A maiores, o modelo de amplitudes non é coñecido. Este modelo explicaría cales son os procesos posibles, e a súa dinámica, a través dos cales obtense o estado final observado. Coñecer as amplitudes no espazo fásico é chave para poder obter un valor non sesgado da eficiencia da medida destes decaementos.
Xa que queremos obter cantas veces acontece un determinado proceso, é fundamental comprender cantas veces, de todas as que se producen dentro ou fóra do detector, o proceso é detectado. Este valor é coñecido como eficiencia.

Como nas colisións entre protóns é moi difícil saber cantos mesóns B^+ se produciron, para así estimar o número de veces que decaeu ao noso estado final de estudo, o que se fai neste tipo de medidas é considerar unha canle de decaemento semellante coa que comparar o resultado medido. Deste xeito, empregando a

información experimental sobre esta outra canle, á cal chamamos canle de normalización, podemos estimar o valor para a nosa canle de sinal ou de estudo.
Para este caso, para a canle de normalización escolleuse a canle de decaemento que maior contribúe ao espazo fásico do noso estado final, obviamente considerando todo o espazo fásico. Este proceso correspóndese co decaemento $B^+ \to D^0(K^0_S\pi^+\pi^-)\pi^+$, onde D^0 é un mesón que contén un quark de tipo c.

Deste xeito, para realizar a análise é necesario facer un paso previo de selección dos datos, despois eliminar a maior parte do ruído a través do uso de ferramentas de análise multivariable para finalmente obter o número de eventos medido en cada unha das canles.
Para obter o número de eventos é preciso realizar un axuste aos datos, non no espazo fásico, senón no espazo no cal se pode identificar a partícula nai: a masa invariante das catro partículas no estado final. Deste xeito, todas as compoñentes que aparecen nos datos deben ser parametrizadas. Para isto identifícanse e simulan, obtendo despois unha forma funcional destas simulacións. Unha vez obtemos cada representación para cada un destes procesos, tanto de sinal como ruído, podemos facer o axuste obtendo como resultado do axuste, o número de eventos de sinal para cada unha das compoñentes identificadas.
Como o modelo de amplitudes da canle de normalización é coñecido, a eficiencia de detección deste proceso pode obterse da simulación (dividindo o número de eventos despois do proceso de selección polo número de eventos de simulación xerados), despois de corrixirla por pequenas diferenzas cos datos. Pola contra, na nosa canle de sinal debemos facer un paso a maiores.

O proceso para obter o valor da eficiencia cando o modelo de amplitudes non é coñecido consta dos seguintes pasos:

- Primeiro é necesario simular cal é o impacto da eficiencia en cada posición do espazo fásico a analizar. Para iso simúlase a distribución do espazo fásico e faise pasar esta simulación a través do detector, seguindo así o mesmo proceso de selección que os datos reais (coma todas as simulacións empregadas durante a análise).
- Seguidamente, emprégase un método para extraer a compoñente de sinal a partir do axuste aos datos, asegurando que só a compoñente de sinal, e non o ruído, é representada polos datos.
- Finalmente, emprégase un método para facer un modelo dirixido por datos. O que fará este método é transformar un novo conxunto de datos, que é a simulación do espazo fásico a nivel de xerador, é dicir, sen pasar polo detector. O método transformará este conxunto de datos para que se asemelle ou aos datos reais ou á simulación. Obtendo así dous conxuntos de pesos nesta nova mostra que teñen a información que necesitamos para obter o valor da eficiencia real dos datos.

Deste xeito podemos obter a distribución da eficiencia no espazo fásico polos pesos que fan que se asemelle o espazo fásico a nivel de xerador ao espazo fásico que sufriu o proceso de selección dos datos. En cambio, a distribución das amplitudes virá dada polos pesos que fan que o xerador se asemelle aos datos. Agora ben, os datos teñen efectos de eficiencia e o que desexamos obter é a distribución das amplitudes a nivel de xerador. Deste xeito, dividimos os pesos das amplitudes obtidas dos datos polos pesos das eficiencias, obtendo así o terceiro conxunto de pesos que describen as amplitudes a nivel de xerador.

Agora que temos tanto a información das amplitudes a nivel de xerador así como a descrición da dependencia da eficiencia coa posición do espazo fásico, podemos obter a eficiencia corrixida. Este valor da eficiencia será a integral no espazo fásico da distribución das amplitudes a nivel de xerador multiplicada pola distribución da eficiencia. Como estas distribucións están representadas a través de pesos no conxunto de datos a nivel de xerador, a integral no espazo fásico pode verse como a media dos pesos que describen a eficiencia pesada polas amplitudes. Cómpre decatarse que os valores dos pesos de eficiencia deben sumar a eficiencia total obtida para a mostra da simulación que sufriu o proceso de selección.

Unha vez obtido o número de eventos medidos e o valor da eficiencia no canle de normalización e no canle de sinal, pódese extraer o valor da frecuencia de decaemento do canle de sinal introducindo o valor para o canle de normalización. Este valor é xeralmente obtido por outra medida independente, ben do mesmo experimento, doutros experimentos ou da combinación de todos os resultados presentes na literatura.

8.5 CALIBRACIÓN DA ESCALA DO MOMENTO EN LHCB EMPREGANDO O DIAGRAMA DE ARMENTEROS-PODOLANSKI

USC

Como xa se mencionou, además da medida central da tese, realizouse outro estudo diferente. Este estudo consiste no desenvolvemento dun novo método para a calibración da escala do momento do detector. A escala do momento refírese á exactitude coa que se pode medir o momento das partículas cargadas dentro

do detector. Isto determina a resolución do detector á hora de combinar estes momentos e construír as masas invariantes, que son empregadas para extraer os parámetros físicos de interese. Polo tanto, unha mellor exactitude na medida da escala do momento, resulta nunha mellor precisión á hora de determinar a masa invariante ou ángulos obtidos a través da medida do momento. Deste xeito, pode obterse unha maior precisión nos parámetros físicos coa mesma cantidade de datos.

O método actual que emprega a colaboración do LHCb consiste en construír a masa invariante de resonancias coñecidas, comparando o resultado obtido co resultado de referencia na literatura. O método que se presenta nesta tese é moito máis complexo e emprega o diagrama de Armenteros-Podolanski. Neste espazo os decaementos a dous corpos aparecen como elipses no plano, de tal xeito que dos parámetros desta elipse pódese extraer o valor das tres masas que participan no decaemento.
Neste estudo, conseguiuse extraer, ao mesmo tempo e de xeito independente, o valor da escala do momento e o valor da masa invariante da partícula nai. Esta diferenza con respecto ao método empregado actualmente no detector é fundamental, xa que estas dúas cantidades poden tratarse de xeito independente.
O principal beneficio do método sería entón a posibilidade de extraer o valor da escala do momento e, ao mesmo tempo, controlar os diferentes efectos experimentais que impactan a medida. Como por exemplo, controlar a limitación que impón a precisión coa que se coñece a masa da partícula nai ou corrixir a simulación debido ao mal coñecemento da radiación dos fotóns dentro do detector. Este último é un problema ben coñecido e sería un factor limitante á hora de obter a escala do momento cando a exactitude sexa mellorada no método actual.

8.6 CONCLUSIÓNS

Nesta tese desenvolveuse un traballo central, relacionado coa medida da frecuencia de decaemento que é crucial para constreñir a contaminación producida polos procesos pingüíns no ángulo α da matriz CKM. Este ángulo pode ser medido co detector LHCb a través do estudo das resonancias axiais no proceso $B^0 \to \pi^+\pi^-\pi^+\pi^-$. Sentando así as bases para unha futura medida.
Neste análise destaca o emprego dun método moi novidoso para obter o valor da eficiencia, resultando nunha redución das incertezas sistemáticas asociadas a este tipo de medidas.

Ademais, desenvolveuse un novo método para a calibración da escala do momento empregando o diagrama de Armentero-Podolanski. Con este método conséguese obter, empregando datos reais, unha exactitude semellante á obtida polo detector co método actual, pero empregando menos recursos. É dicir, empregando menos cantidade de datos e menos recursos computacionais. Demonstrando un gran potencial do método e sendo moi atractivo un posible estudo máis exhaustivo do mesmo.

Para concluír, convén resaltar o novidoso dos métodos empregados na tese e o gran potencial de mellora que teñen no marco da física de partículas, así como a súa posible aplicación a outros campos de estudo nos que o estudo de simulacións é chave para comprender os procesos que se queren estudar.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

BIBLIOGRAPHY

[1] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[2] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf. URL: https://doi.org/10.1143/PTP.49.652.

[3] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[4] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[5] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[6] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[7] N. Aghanim et al. "Planck2018 results: VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641 (Sept. 2020), A6. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833910. URL: http://dx.doi.org/10.1051/0004-6361/201833910.

[8] Jun Iizuka et al. "Parametrization of Pontecorvo–Maki–Nakagawa–Sakata mixing matrix based on CP-violating bipair neutrino mixing". In: *Modern Physics Letters A* 30.05 (Feb. 2015), p. 1550019. ISSN: 1793-6632. DOI: 10.1142/s0217732315500194. URL: http://dx.doi.org/10.1142/S0217732315500194.

[9] X. Fan et al. "Measurement of the Electron Magnetic Moment". In: *Phys. Rev. Lett.* 130 (7 2023), p. 071801. DOI: 10.1103/PhysRevLett.130.071801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.071801.

[10] T. Aoyama et al. "The anomalous magnetic moment of the muon in the Standard Model". In: *Physics Reports* 887 (2020). The anomalous magnetic moment of the muon in the Standard Model, pp. 1–166. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2020.07.006. URL: https://www.sciencedirect.com/science/article/pii/S0370157320302556.

[11] Eung Jin Chun, Jongkuk Kim, and Tanmoy Mondal. "Electron EDM and muon anomalous magnetic moment in two-Higgs-doublet models". In: *Journal of High Energy Physics* 2019.12 (2019), p. 68. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2019)068. URL: https://doi.org/10.1007/JHEP12(2019)068.

[12] R. Aaij et al. "Observation of Several Sources of $\mathcal{CP}$ Violation in $B^+ \to \pi^+\pi^+\pi^-$ Decays". In: *Physical Review Letters* 124.3 (Jan. 2020). ISSN: 1079-7114. DOI: 10.1103/physrevlett.124.031801. URL: http://dx.doi.org/10.1103/PhysRevLett.124.031801.

[13] R. Aaij et al. "Observation of $\mathcal{CP}$ Violation in Charm Decays". In: *Physical Review Letters* 122.21 (May 2019). ISSN: 1079-7114. DOI: 10.1103/physrevlett.122.211803. URL: http://dx.doi.org/10.1103/PhysRevLett.122.211803.

[14] Karl R. Popper. *The Logic of Scientific Discovery*. London, England: Routledge, 1935.

[15] Sébastien Descotes-Genon et al. "Light-Cone Sum Rules for S-wave $B \to K\pi$ Form Factors". In: *JHEP* 06 (2023), p. 034. DOI: 10.1007/JHEP06(2023)034. arXiv: 2304.02973 [hep-ph].

[16] The GAMBIT Collaboration et al. *Collider constraints on electroweakinos in the presence of a light gravitino*. 2023. arXiv: 2303.09082.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[17] R. Aaij et al. "Momentum scale calibration of the LHCb spectrometer". In: *Journal of Instrumentation* 19.02 (2024), P02008. DOI: 10.1088/1748-0221/19/02/P02008. URL: https://dx.doi.org/10.1088/1748-0221/19/02/P02008.

[18] J. Podolanski and R. Armenteros. "III. Analysis of V-events". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 45.360 (1954), pp. 13–30. DOI: 10.1080/14786440108520416.

[19] Henri Theil. "The Information Approach to Demand Analysis". In: *Henri Theil's Contributions to Economics and Econometrics: Volume II: Consumer Demand Analysis and Information Theory*. Ed. by Baldev Raj and Johan Koerts. Dordrecht: Springer Netherlands, 1992, pp. 627–651. ISBN: 978-94-011-2408-9. DOI: 10.1007/978-94-011-2408-9_3. URL: https://doi.org/10.1007/978-94-011-2408-9_3.

[20] P. Baladrón Rodríguez et al. "Calibration of the momentum scale of a particle physics detector using the Armenteros-Podolanski plot". In: *Journal of Instrumentation* 16.06 (2021), P06036. DOI: 10.1088/1748-0221/16/06/P06036. URL: https://dx.doi.org/10.1088/1748-0221/16/06/P06036.

[21] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[22] Emmy Noether. "Invariant variation problems". In: *Transport Theory and Statistical Physics* 1.3 (1971), pp. 186–207. DOI: 10.1080/00411457108231446. eprint: https://doi.org/10.1080/00411457108231446. URL: https://doi.org/10.1080/00411457108231446.

[23] J. Charles et al. "Current status of the standard model CKM fit and constraints on $\Delta F = 2$ new physics". In: *Phys. Rev.* D91 (2015). updated results and plots available at `http://ckmfitter.in2p3.fr/`, p. 073007. DOI: 10.1103/PhysRevD.91.073007. arXiv: 1501.05013 [hep-ph].

[24] Lincoln Wolfenstein. "Parametrization of the Kobayashi-Maskawa Matrix". In: *Phys. Rev. Lett.* 51 (21 1983), pp. 1945–1947. DOI: 10.1103/PhysRevLett.51.1945. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.1945.

[25] Hai-Yang Cheng and Kwei-Chou Yang. "Hadronic charmless B decays B —> AP". In: *Phys. Rev. D* 76 (2007), p. 114020. DOI: 10.1103/PhysRevD.76.114020. arXiv: 0709.0137 [hep-ph].

[26] J. Charles et al. "Isospin analysis of charmless B-meson decays". In: *Eur. Phys. J. C* 77.8 (2017), p. 574. DOI: 10.1140/epjc/s10052-017-5126-9. arXiv: 1705.02981 [hep-ph].

[27] J. P. Lees et al. "Measurement of *CP* asymmetries and branching fractions in charmless two-body *B*-meson decays to pions and kaons". In: *Phys. Rev. D* 87 (5 2013), p. 052009. DOI: 10.1103/PhysRevD.87.052009. URL: https://link.aps.org/doi/10.1103/PhysRevD.87.052009.

[28] B. Aubert et al. "Study of $B^0 \to \pi^0\pi^0$, $B^\pm \to \pi^\pm\pi^0$, and $B^\pm \to K^\pm\pi^0$ decays, and isospin analysis of $B \to \pi\pi$ decays". In: *Phys. Rev. D* 76 (9 2007), p. 091102. DOI: 10.1103/PhysRevD.76.091102. URL: https://link.aps.org/doi/10.1103/PhysRevD.76.091102.

[29] J. Dalseno et al. "Measurement of the *CP* violation parameters in $B^0 \to \pi^+\pi^-$ decays". In: *Phys. Rev. D* 88 (9 2013), p. 092003. DOI: 10.1103/PhysRevD.88.092003. URL: https://link.aps.org/doi/10.1103/PhysRevD.88.092003.

[30] Y.-T. Duh et al. "Measurements of branching fractions and direct *CP* asymmetries for $B \to K\pi$, $B \to \pi\pi$ and $B \to KK$ decays". In: *Phys. Rev. D* 87 (3 2013), p. 031103. DOI: 10.1103/PhysRevD.87.031103. URL: https://link.aps.org/doi/10.1103/PhysRevD.87.031103.

[31] T. Julius et al. "Measurement of the branching fraction and *CP* asymmetry in $B^0 \to \pi^0\pi^0$ decays, and an improved constraint on ϕ_2". In: *Phys. Rev. D* 96 (3 2017), p. 032007. DOI: 10.1103/PhysRevD.96.032007. URL: https://link.aps.org/doi/10.1103/PhysRevD.96.032007.

[32] R. Aaij et al. "Measurement of *CP* asymmetries in two-body $B^0_{(s)}$-meson decays to charged pions and kaons". In: *Phys. Rev. D* 98 (3 2018), p. 032004. DOI: 10.1103/PhysRevD.98.032004. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.032004.

[33] Bernard Aubert et al. "Measurement of the Branching Fraction, Polarization, and CP Asymmetries in $B^0 \to \rho^0\rho^0$ Decay, and Implications for the CKM Angle α". In: *Phys. Rev. D* 78 (2008), p. 071104. DOI: 10.1103/PhysRevD.78.071104. arXiv: 0807.4977 [hep-ex].

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[34] B. Aubert et al. "Improved Measurement of $B^+ \to \rho^+\rho^0$ and Determination of the Quark-Mixing Phase Angle α". In: *Phys. Rev. Lett.* 102 (14 2009), p. 141802. DOI: 10.1103/PhysRevLett.102.141802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.102.141802.

[35] I. Adachi et al. "Study of $B^0 \to \rho^0\rho^0$ decays, implications for the CKM angle ϕ_2 and search for other B^0 decay modes with a four-pion final state". In: *Phys. Rev. D* 89.7 (2014). [Addendum: Phys.Rev.D 89, 119903 (2014)], p. 072008. DOI: 10.1103/PhysRevD.89.072008. arXiv: 1212.4015 [hep-ex].

[36] J. Zhang et al. "Observation of $B^\mp \to \rho^\mp\rho^0$ Decays". In: *Phys. Rev. Lett.* 91 (22 2003), p. 221801. DOI: 10.1103/PhysRevLett.91.221801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.91.221801.

[37] R. Aaij et al. "Observation of the $B^0 \to \rho^0\rho^0$ decay from an amplitude analysis of $B^0 \to (\pi^+\pi^-)(\pi^+\pi^-)$ decays". In: *Physics Letters B* 747 (2015), pp. 468–478. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2015.06.027. URL: https://www.sciencedirect.com/science/article/pii/S0370269315004499.

[38] J. P. Lees et al. "Measurement of *CP*-violating asymmetries in $B^0 \to (\rho\pi)^0$ decays using a time-dependent Dalitz plot analysis". In: *Phys. Rev. D* 88 (1 2013), p. 012003. DOI: 10.1103/PhysRevD.88.012003. URL: https://link.aps.org/doi/10.1103/PhysRevD.88.012003.

[39] A. Kusaka et al. "Measurement of *CP* Asymmetry in a Time-Dependent Dalitz Analysis of $B^0 \to (\rho\pi)^0$ and a Constraint on the Quark Mixing Matrix Angle ϕ_2". In: *Phys. Rev. Lett.* 98 (22 2007), p. 221602. DOI: 10.1103/PhysRevLett.98.221602. URL: https://link.aps.org/doi/10.1103/PhysRevLett.98.221602.

[40] J. Dalseno. "Resolving the ϕ_2 (α) ambiguity in $B \to \rho\rho$". In: *JHEP* 11 (2018), p. 193. DOI: 10.1007/JHEP11(2018)193. arXiv: 1808.09391 [hep-ph].

[41] Bernard Aubert et al. "Measurements of CP-Violating Asymmetries in $B^0 \to a_1^\pm(1260)\pi^\mp$ decays". In: *Phys. Rev. Lett.* 98 (2007), p. 181803. DOI: 10.1103/PhysRevLett.98.181803. arXiv: hep-ex/0612050.

[42] J. Dalseno et al. "Measurement of Branching Fraction and First Evidence of CP Violation in $B^0 \to a_1^\pm(1260)\pi^\mp$ Decays". In: *Phys. Rev. D* 86 (2012), p. 092012. DOI: 10.1103/PhysRevD.86.092012. arXiv: 1205.5957 [hep-ex].

[43] M. Gronau and J. Zupan. "Weak phase α from $B^0 \to a_1^\pm(1260)\pi^\mp$". In: *Physical Review D* 73.5 (Mar. 2006). ISSN: 1550-2368. DOI: 10.1103/physrevd.73.057502. URL: http://dx.doi.org/10.1103/PhysRevD.73.057502.

[44] Bernard Aubert et al. "Measurement of branching fractions of B decays to $K_1(1270)\pi$ and $K_1(1400)\pi$ and determination of the CKM angle alpha from $B^0 \to a_1^\pm\pi^\mp$". In: *Phys. Rev. D* 81 (2010), p. 052009. DOI: 10.1103/PhysRevD.81.052009. arXiv: 0909.2171 [hep-ex].

[45] J. Dalseno. "Resolving the $\phi_2(\alpha)$ ambiguity in $B^0 \to a_1^\pm\pi^\mp$". In: *Journal of High Energy Physics* 2019.10 (Oct. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep10(2019)191. URL: http://dx.doi.org/10.1007/JHEP10(2019)191.

[46] Michael Gronau et al. "Decays of B mesons to two light pseudoscalars". In: *Physical Review D* 50.7 (Oct. 1994), 4529–4543. ISSN: 0556-2821. DOI: 10.1103/physrevd.50.4529. URL: http://dx.doi.org/10.1103/PhysRevD.50.4529.

[47] Mahiko Suzuki. "Strange axial-vector mesons". In: *Phys. Rev. D* 47 (3 1993), pp. 1252–1255. DOI: 10.1103/PhysRevD.47.1252. URL: https://link.aps.org/doi/10.1103/PhysRevD.47.1252.

[48] Hai-Yang Cheng. "Revisiting axial-vector meson mixing". In: *Physics Letters B* 707.1 (Jan. 2012), 116–120. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2011.12.013. URL: http://dx.doi.org/10.1016/j.physletb.2011.12.013.

[49] R. L. Workman and Others. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[50] A. A. Alves Junior et al. "Prospects for measurements with strange hadrons at LHCb". In: *Journal of High Energy Physics* 2019.5 (May 2019). ISSN: 1029-8479. DOI: 10.1007/jhep05(2019)048. URL: http://dx.doi.org/10.1007/JHEP05(2019)048.

[51] The LHCb Collaboration et al. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08005.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[52] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08001.

[53] A Leflat. "Upgrade of the LHCb Vertex Locator". In: *Journal of Instrumentation* 9.08 (2014), p. C08015. DOI: 10.1088/1748-0221/9/08/C08015. URL: https://dx.doi.org/10.1088/1748-0221/9/08/C08015.

[54] R. Aaij et al. "Measurement of the W boson mass". In: *Journal of High Energy Physics* 2022.1 (Jan. 2022). ISSN: 1029-8479. DOI: 10.1007/jhep01(2022)036. URL: http://dx.doi.org/10.1007/JHEP01(2022)036.

[55] Alexandre Brea Rodríguez et al. "Prospects on searches for baryonic Dark Matter produced in *b*-hadron decays at LHCb". In: *The European Physical Journal C* 81.11 (Nov. 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09762-w. URL: http://dx.doi.org/10.1140/epjc/s10052-021-09762-w.

[56] "LHCb detector performance". In: *International Journal of Modern Physics A* 30.07 (Mar. 2015), p. 1530022. ISSN: 1793-656X. DOI: 10.1142/s0217751x15300227. URL: http://dx.doi.org/10.1142/S0217751X15300227.

[57] P R Barbosa-Marinho et al. *LHCb VELO (VErtex LOcator): Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/504321.

[58] S Amato et al. *LHCb magnet: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/424338.

[59] P R Barbosa-Marinho et al. *LHCb inner tracker: Technical Design Report*. Technical design report. LHCb. revised version number 1 submitted on 2002-11-13 14:14:34. Geneva: CERN, 2002. URL: https://cds.cern.ch/record/582793.

[60] P R Barbosa-Marinho et al. *LHCb outer tracker: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/519146.

[61] "LHCb: RICH technical design report". In: (Sept. 2000).

[62] R. Calabrese et al. "Performance of the LHCb RICH detectors during LHC Run 2". In: *Journal of Instrumentation* 17.07 (2022), P07013. DOI: 10.1088/1748-0221/17/07/P07013. URL: https://dx.doi.org/10.1088/1748-0221/17/07/P07013.

[63] S Amato et al. *LHCb calorimeters: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/494264.

[64] C. Abellán Beteta et al. "Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC". In: *arXiv e-prints*, arXiv:2008.11556 (Aug. 2020), arXiv:2008.11556. DOI: 10.48550/arXiv.2008.11556. arXiv: 2008.11556 [physics.ins-det].

[65] P R Barbosa-Marinho et al. *LHCb muon system: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/504326.

[66] R. Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *Journal of Instrumentation* 14.04 (Apr. 2019), P04013–P04013. ISSN: 1748-0221. DOI: 10.1088/1748-0221/14/04/p04013. URL: http://dx.doi.org/10.1088/1748-0221/14/04/P04013.

[67] R Antunes-Nobrega et al. *LHCb trigger system: Technical Design Report*. Technical design report. LHCb. revised version number 1 submitted on 2003-09-24 12:12:22. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/630828.

[68] Roel Aaij et al. "The LHCb Upgrade I". In: *JINST* 19.05 (2024), P05065. DOI: 10.1088/1748-0221/19/05/P05065. arXiv: 2305.10515 [hep-ex].

[69] R. Aaij et al. "Allen: A High-Level Trigger on GPUs for LHCb". In: *Computing and Software for Big Science* 4.1 (2020). ISSN: 2510-2044. DOI: 10.1007/s41781-020-00039-7. URL: http://dx.doi.org/10.1007/s41781-020-00039-7.

[70] Gloria Corti. "LHCb computing". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 265–269. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00120-6. URL: https://www.sciencedirect.com/science/article/pii/S0168900201001206.

[71] Niko Neufeld. "The LHCb online system". In: *Nuclear Physics B - Proceedings Supplements* 120 (2003). Proceedings of the 8th International Conference on B-Physics at Hadron Machines, pp. 105–108. ISSN: 0920-5632. DOI: https://doi.org/10.1016/S0920-5632(03)01888-7. URL: https://www.sciencedirect.com/science/article/pii/S0920563203018887.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

[72] G Barrand et al. "GAUDI: The software architecture and framework for building LHCb data processing applications". In: (2000). DOI: 10.1016/S0010-4655(01)00254-5. URL: https://cds.cern.ch/record/467678.

[73] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: *SciPost Phys. Codeb.* 2022 (2022), p. 8. DOI: 10.21468/SciPostPhysCodeb.8. arXiv: 2203.11601 [hep-ph].

[74] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[75] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[76] P. Golonka and Z. Was. "PHOTOS Monte Carlo: a precision tool for QED correctionsin Z and W decays". In: *The European Physical Journal C* 45.1 (Jan. 2006), 97–107. ISSN: 1434-6052. DOI: 10.1140/epjc/s2005-02396-4. URL: http://dx.doi.org/10.1140/epjc/s2005-02396-4.

[77] Lukas Tlustos et al. "Imaging properties of the Medipix2 system exploiting single and dual energy thresholds". In: *IEEE Trans. Nucl. Sci.* 53 (2006), pp. 1323–1328. DOI: 10.1109/TNS.2006.872627. URL: http://cds.cern.ch/record/913827.

[78] Lucio Anderlini et al. *The PIDCalib package*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2202412.

[79] The LHCb collaboration. "Measurement of the track reconstruction efficiency at LHCb". In: *Journal of Instrumentation* 10.02 (Feb. 2015), P02007–P02007. ISSN: 1748-0221. DOI: 10.1088/1748-0221/10/02/p02007. URL: http://dx.doi.org/10.1088/1748-0221/10/02/P02007.

[80] Alexandra Martin Sanchez, Patrick Robbe, and Marie-Hélène Schune. "Performances of the LHCb Lo Calorimeter Trigger". In: (June 2012).

[81] Hitoshi Murayama. *Notes on Phase Space*. 2007. URL: http://hitoshi.berkeley.edu/233B/phasespace.pdf.

[82] Byung-Chul Han. *Infocracy*. Cambridge: Polity Press, 2022.

[83] Luke Zoltan Kelley. "kalepy: a Python package for kernel density estimation, sampling and plotting". In: *Journal of Open Source Software* 6.57 (2021), p. 2784. DOI: 10.21105/joss.02784. URL: https://doi.org/10.21105/joss.02784.

[84] Alex Rogozhnikov. "Reweighting with Boosted Decision Trees". In: *Journal of Physics: Conference Series* 762.1 (2016), p. 012036. DOI: 10.1088/1742-6596/762/1/012036. URL: https://dx.doi.org/10.1088/1742-6596/762/1/012036.

[85] M.C. Jones M.P. Wand. *Kernel Smoothing*. Chapman and Hall/CRC, 1994. DOI: https://doi.org/10.1201/b14876.

[86] F. James and M. Roos. "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations". In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.

[87] D. Martínez Santos et al. "Ipanema-\beta : tools and examples for HEP analysis on GPU". In: (June 2017). arXiv: 1706.01420 [hep-ex].

[88] M. Pivk and F.R. Le Diberder. "sPlot: A statistical tool to unfold data distributions". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1–2 (Dec. 2005), 356–369. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.08.106. URL: http://dx.doi.org/10.1016/j.nima.2005.08.106.

[89] K. Hayasaka, Z. Huang, and E. Kou. "The hadronic τ decay $\tau^- \to K_1^- \nu_\tau \to (K^-\omega)\nu_\tau \to (K^-\pi^+\pi^-\pi^0)\nu_\tau$ and the axial vector mixing angle". In: *Eur. Phys. J. C* 81.6 (2021), p. 502. DOI: 10.1140/epjc/s10052-021-09303-5. arXiv: 2102.00752 [hep-ph].

[90] P. del Amo Sanchez et al. "B-meson decays to $\eta'\rho$, $\eta' f_0$, and $\eta' K^{*}$". In: *Phys. Rev. D* 82 (2010), p. 011502. DOI: 10.1103/PhysRevD.82.011502. arXiv: 1004.0240 [hep-ex].

[91] B. Aubert et al. "Branching fraction measurements of charged B decays to $K^{*+}K^+K^-$, $K^{*+}\pi^+K^-$, $K^{*+}K^+\pi^-$ and $K^{*+}\pi^+\pi^-$ final states". In: *Phys. Rev. D* 74 (5 2006), p. 051104. DOI: 10.1103/PhysRevD.74.051104. URL: https://link.aps.org/doi/10.1103/PhysRevD.74.051104.

USC

[92] Bernard Aubert et al. "Observation of $B^+ \to a_1^+(1260)K^0$ and $B^0 \to a_1^-(1260)K^{+}$". In: *Phys. Rev. Lett.* 100 (2008), p. 051803. DOI: 10.1103/PhysRevLett.100.051803. arXiv: 0709.4165 [hep-ex].

[93] R. Aaij et al. "Dalitz plot analysis of $B_s^0 \to \bar{D}^0 K^- \pi^+$ decays". In: *Phys. Rev.* D90 (2014), p. 072003. DOI: 10.1103/PhysRevD.90.072003. arXiv: 1407.7712 [hep-ex].

[94] D. Martínez Santos and F. Dupertuis. "Mass distributions marginalized over per-event errors". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 764 (2014), pp. 150–155. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2014.06.081. URL: https://www.sciencedirect.com/science/article/pii/S0168900214008353.

[95] R. Aaij et al. "Curvature-bias corrections using a pseudomass method". In: *Journal of Instrumentation* 19.03 (2024), P03010. DOI: 10.1088/1748-0221/19/03/P03010. URL: https://dx.doi.org/10.1088/1748-0221/19/03/P03010.

[96] David Friday. "A time-integrated analysis of Bd $\to \pi+\pi-\pi+\pi-$ at LHCb and simulations for Low Gain Avalanche Detectors (LGAD)". PhD thesis. Glasgow U., Glasgow U., 2023. DOI: 10.5525/gla.thesis.83485.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GLOSSARY

GBrw Gradient Boosted re-weighting 33, 37–39, 43, 62–65, 67–69, 72, 75, 76, 78–80, 86–88

KDE Kernel Density Estimation 33, 37–39, 43, 98, 108

LHCb Large Hadron Collider beauty 1–5, 10, 12, 19–27, 30, 31, 37, 39–41, 44–46, 64, 72–75, 78, 79, 86, 93, 94, 99–101, 103, 107, 109, 111, 114

MC Monte Carlo 26, 28–31, 37, 44–46, 62–70, 72–75, 78–83, 85, 86, 96–104, 107

PHSP Phase Space 1, 3, 15–17, 25–29, 33–46, 48–50, 54, 57–60, 62–67, 70–83, 85, 90, 91, 94, 107, 108

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APPENDIX A

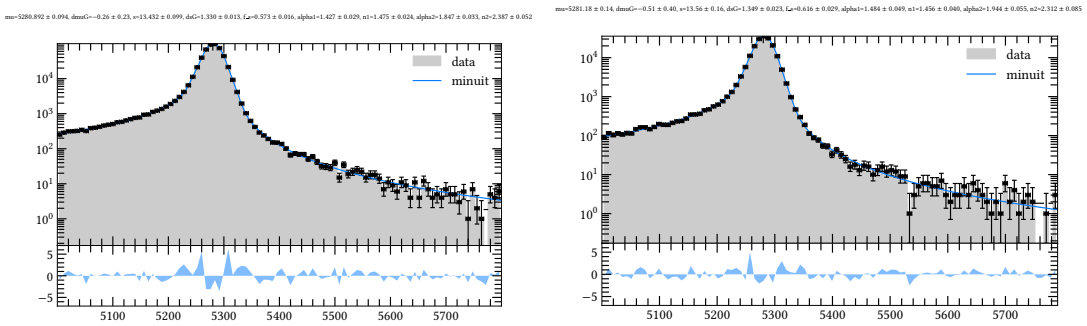
APPENDIX

A.1 MC FITS

The fits to extract the different line-shapes from MC on the branching fraction analysis are show.

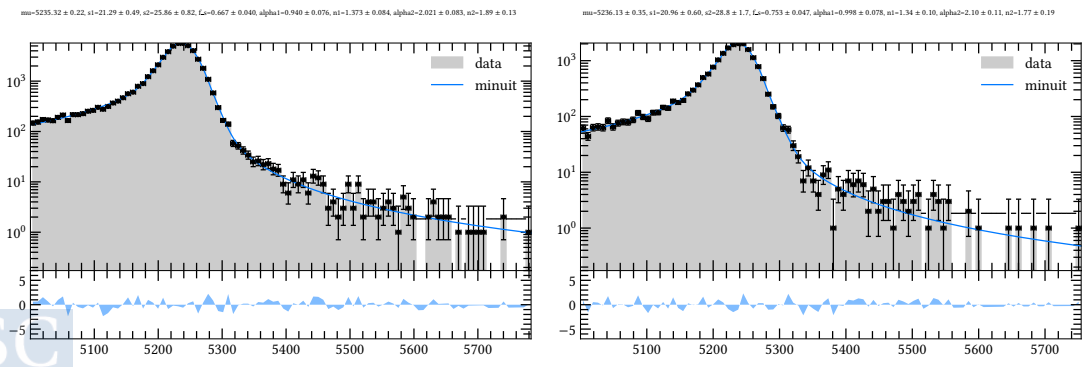
A.1.1 WIDE REGION

A.1.1.1 SIGNAL CHANNEL



(A) 1718 DD (B) 1718 LL

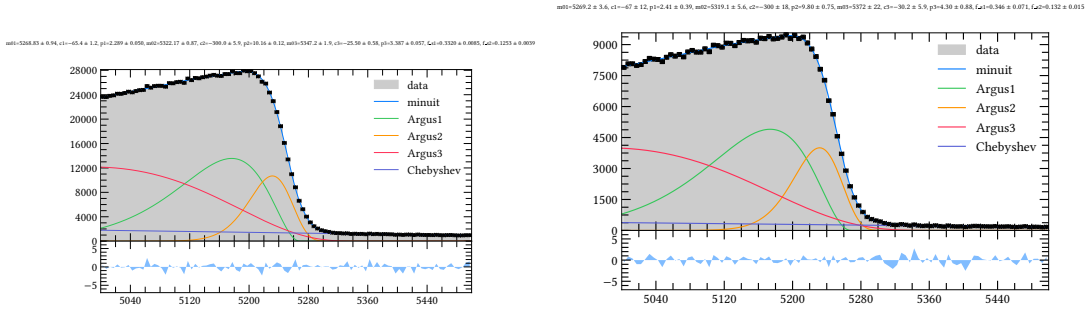
FIGURE A.1: Signal component: Double Crystal-Ball. Event-type: 12105151.



(A) 1718 DD (B) 1718 LL

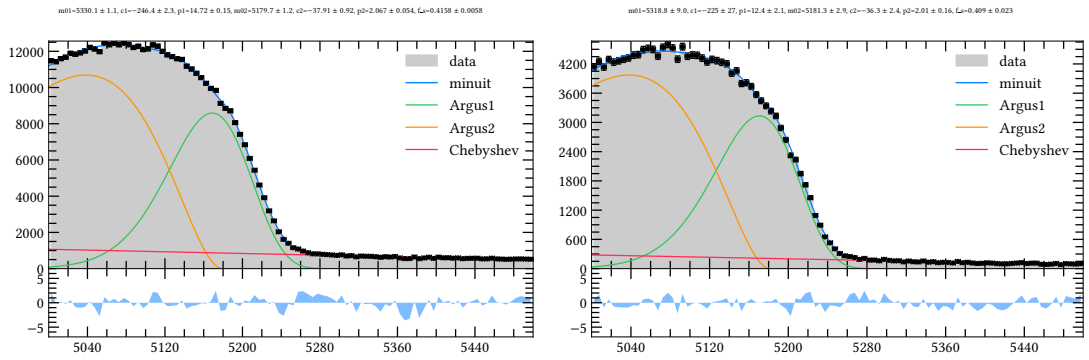
FIGURE A.2: Mis-ID component: Double Crystal-Ball. Event-type: 12105162(3).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



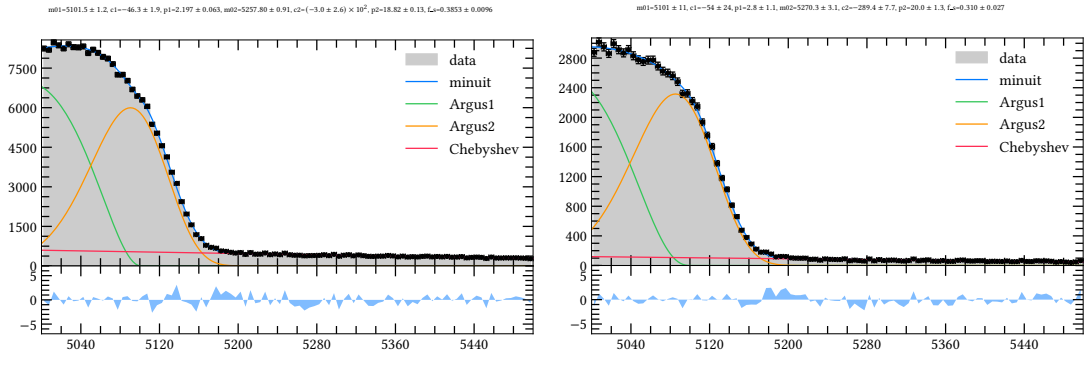
(A) 1718 DD (B) 1718 LL

FIGURE A.3: $B^+ \to K^*(K^0_S\pi)\eta'(\pi\pi\gamma)$ missing the photon: Triple Argus. Event-type: 12105315.



(A) 1718 DD (B) 1718 LL

FIGURE A.4: $B^0_s \to D^-_s(K^0_S\pi^+\pi^-\pi^-)\pi^+$ missing a charged pion from the D^-_s: Double Argus. Event-type: 131661725.



(A) 1718 DD (B) 1718 LL

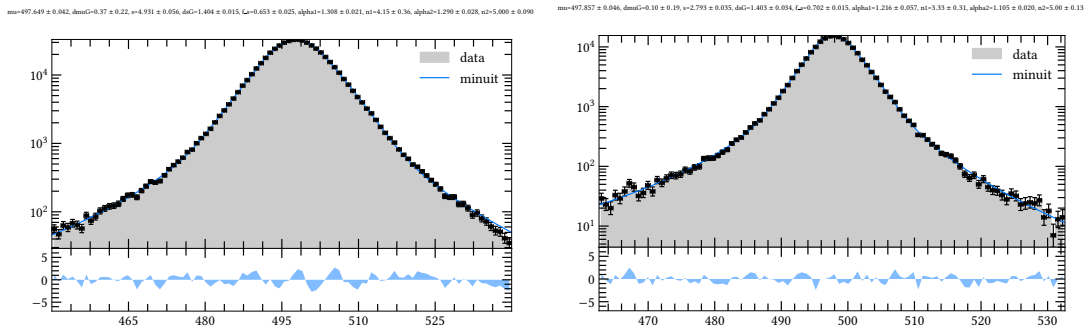
FIGURE A.5: $B^+ \to D^0(K^0_S\pi^+\pi^-\pi^0)\pi^+$ missing the neutral pion: Double Argus. Event-type: 12165529.



(A) 1718 DD (B) 1718 LL

FIGURE A.6: $B^0 \to D^{*-}(D^0(K^0_S\pi^+\pi^-)\pi^-)\pi^+$ missing a charged pion from the D^0: Double Argus. Event-type: 11166144.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



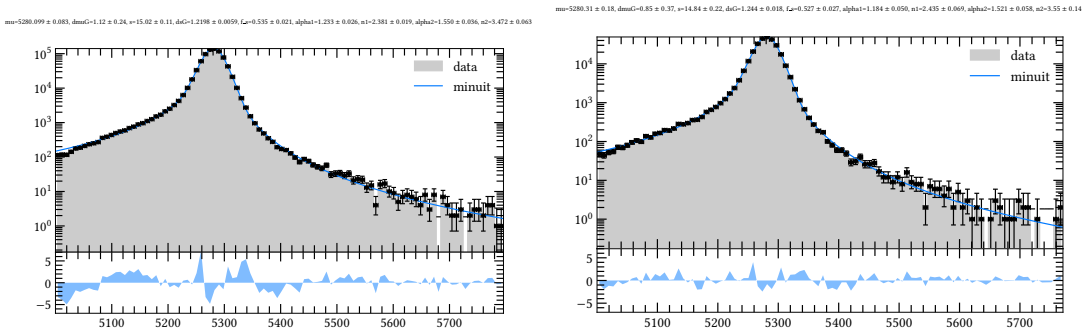
(A) 1718 DD

(B) 1718 LL

FIGURE A.7: Signal component K^0_S: Double Crystal-Ball. Event-type: 12105151.

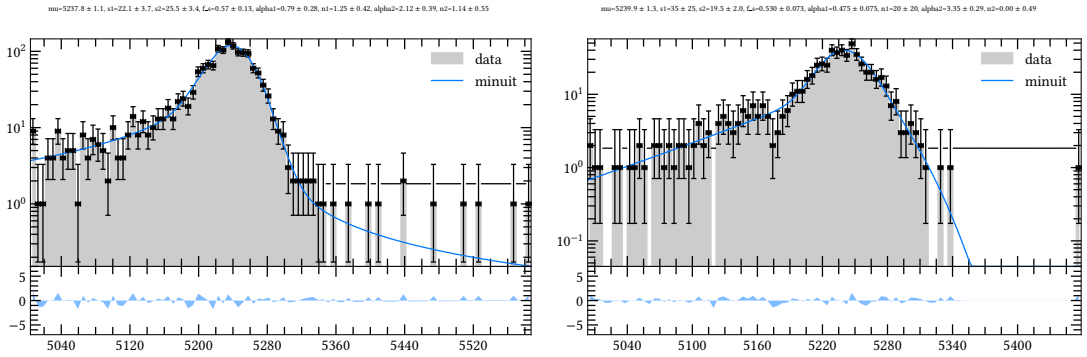
USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A.1.1.2 CONTROL CHANNEL



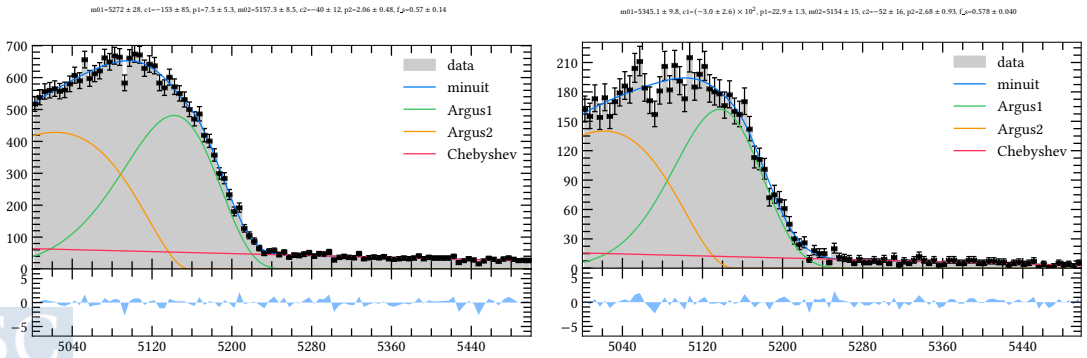
(A) 1718 DD (B) 1718 LL

FIGURE A.8: Signal component: Double Crystal-Ball. Event-type: 12165121.



(A) 1718 DD (B) 1718 LL

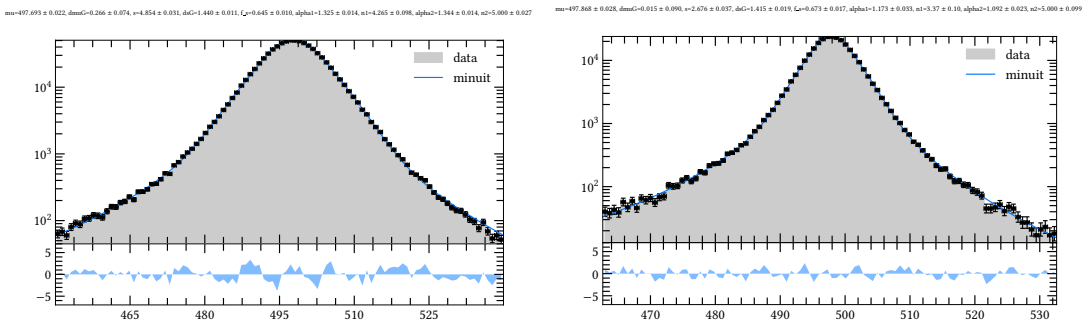
FIGURE A.9: Mis-ID component: Double Crystal-Ball. Event-type: 12105162(3).



(A) 1718 DD (B) 1718 LL

FIGURE A.10: $B_s^0 \to D_s^- (K_S^0 \pi^+\pi^-\pi^-)\pi^+$ missing a charged pion from the D_s^-: Double Argus. Event-type: 1316617**25**.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(A) 1718 DD

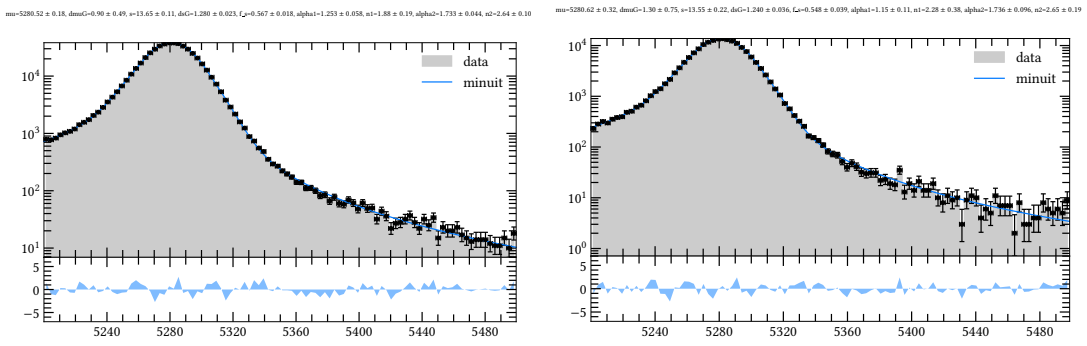
(B) 1718 LL

FIGURE A.11: Signal channel K^0_S: Double Crystal-Ball. Event-type: 12165121.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

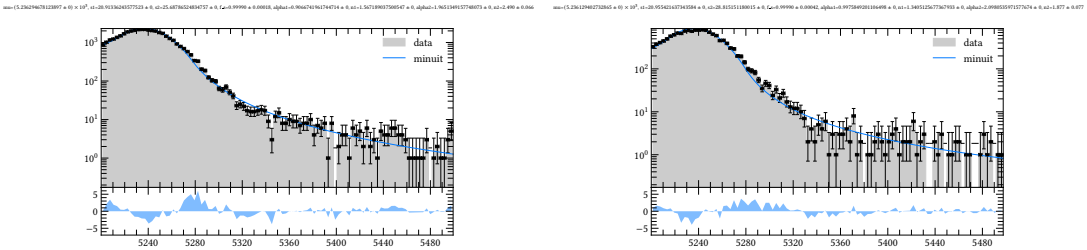
A.1.2 SIGNAL REGION

A.1.2.1 SIGNAL CHANNEL



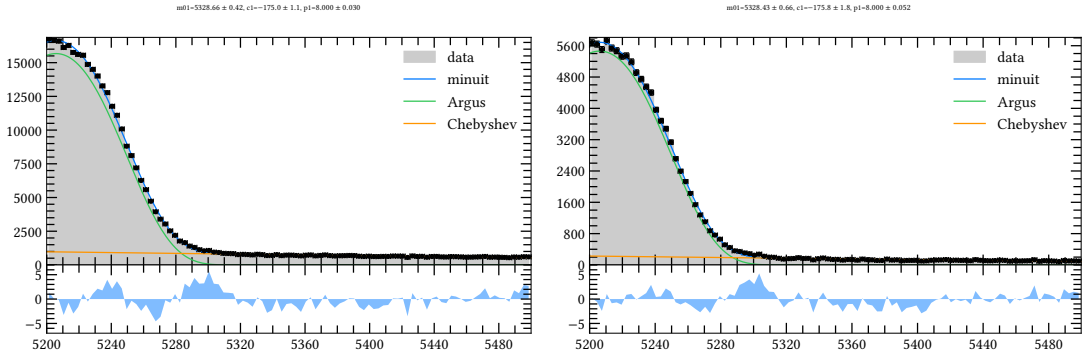
(A) 1718 DD (B) 1718 LL

FIGURE A.12: Signal component: Double Crystal-Ball. Event-type: 12105151.



(A) 1718 DD (B) 1718 LL

FIGURE A.13: Mis-ID component: Double Crystal-Ball. Event-type: 12105162(3).



(A) 1718 DD (B) 1718 LL

FIGURE A.14: $B^+ \to K^*(K^0_S\pi)\eta'(\pi\pi\gamma)$ missing the photon: Argus. Event-type: 12105315.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ACKNOWLEDGMENTS

Sometimes, it is not easy to find a moment to stop, look back, and be grateful to the people who helped you get where you are. However, I want to change this and express my gratitude to the many people who, without seeking any benefit, helped me during the development of my thesis.

I want to express my appreciation for the time and support of the senior staff, who I believe are crucial for students during the initial steps of their research careers and at crucial moments when important decisions must be made. Special thanks to Máximo Plo for helping me write this thesis, and to Bernardo Adeva for discussing my research program with me and accepting me as his teaching assistant. Finally, I want to extend my gratitude to Paul Soler for making my experience in Glasgow very enjoyable.

On the other side of senior support are the fellow students. Without the proper environment created by my office mates and colleagues at the university, this thesis would not have been possible. They helped me through moments of low motivation when nothing seemed to work and no one seemed to care. Even more importantly, they also made the good times enjoyable, spending countless hours with me drinking beer. Finally, I want to express my gratitude for the many hours spent discussing various interesting topics in physics, statistics, and computing. These discussions were invaluable for understanding some key concepts that are much more subtle than they initially appear. I am especially grateful to Moncho, Asier, Saúl, Alex, Adrián, and Marcos from the office, as well as Bastián, Alberto, Andrés, and many other university mates with whom I shared many laughs.

Despite not being from Santiago, I must mention David, the only person with whom I truly shared my work issues, and with whom I navigated the challenges of the α measurement. Finally, I want to mention Marcos again, who provided me with the tools to perform my mass fits on a GPU. We learned a lot from his computing skills, and he deserves this special mention. However, learning is a shared experience, and I also learned a lot from Moncho and Asier, as well as everyone else.

To conclude, I want to express my gratitude to all the people who have shared my life during these years and throughout my whole life: my family, and all my friends from everywhere around. Every individual has a different way of facing and experiencing life. Without any of you, I would not be the person I am. I carry a bit of each of you inside myself, and I am very proud of it.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA





A branching fraction measurement of a four-body charged B meson decay within an inclusive charmless mass window is presented. The mass window focuses on axial decays, which decay into three final-state particles.

The full Run 1 and Run 2 LHCb dataset, corresponding to an integrated luminosity of 9 fb^{-1}, is analysed. A novel approach for estimating efficiencies is introduced, employing a flat phase space simulation in a data-driven method to extract the signal amplitudes.

This measurement serves as the preamble to an amplitude analysis that can be performed on the same data. Through SU(3) symmetry, this analysis aims to constrain the penguin pollution in the CKM angle α extraction from the mixing-induced CP violation in the neutral SU(3) related channel.