Measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays at the LHCb experiment and particle identification developments for the LHCb Upgrade



Fidan Suljik

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2025

To my family
for the eternal love and support

Abstract

This thesis presents a comprehensive study of the $B^{\pm} \to DK^{*\pm}$ decay channel and a measurement of the CKM angle γ using D meson decays to two- and four-body final states $K^+\pi^-$, π^+K^-, K^+K^-, $\pi^+\pi^-$, $K^+\pi^-\pi^+\pi^-$, $\pi^+K^-\pi^+\pi^-$, $\pi^+\pi^-\pi^+\pi^-$ and three-body final states $K^0_S\pi^+\pi^-$, $K^0_S K^+K^-$. The analysis uses the full dataset collected in proton-proton collisions during Run 1 (2011–2012) and Run 2 (2015–2018) at the LHCb experiment, corresponding to an integrated luminosity of $9\,\text{fb}^{-1}$. Observables in relation to *CP* violation are measured for the different D meson decay modes and the results are combined, yielding $\gamma = (63 \pm 13)^\circ$. This model-independent result is compatible with the world average and is included in the latest LHCb combination. This work also constitutes the first observation of the suppressed $B^{\pm} \to [\pi^{\pm}K^{\mp}]_D K^{*\pm}$ and $B^{\pm} \to [\pi^{\pm}K^{\mp}\pi^{\pm}\pi^{\mp}]_D K^{*\pm}$ decays.

Particle identification developments for the LHCb Upgrade detector are also presented. The main calibration sample used for kaons and pions at LHCb is $D^{*+} \to D^0\pi^+$ decays, for which a new fit model is developed and implemented. The new model is used to obtain the PID efficiencies and PID performance with data collected in July 2023 with the LHCb Upgrade detector.

Acknowledgements

First of all, I would like to sincerely thank my supervisor, Sneha Malde, for her continuous support, guidance and availability during my DPhil. It was an absolute pleasure to work with her on our research, which I am proud of. Sneha encouraged me to regularly present my work and supported me in attending major conferences in our field. She always provided quick feedback and we had fruitful discussions.

I extend my thanks to all my colleagues from the Oxford LHCb group and CERN, where I also had the opportunity to spend time during my Long Term Attachment. A special thanks to Sue and Kim for all their support in the department, which was greatly appreciated.

I was very happy to join the Oxford University Association Football Club (OUAFC) and play for the Blues. It was an honour to be part of this historic club and one of the oldest football clubs in the world. I thank my teammates and coaches for the good atmosphere and team spirit. University Football played an important role in my Oxford experience. One of the highlights was the 138th Football Varsity match, when I scored an important goal to secure the win. I was really happy to be in the Starting XI and play against Cambridge, and therefore be awarded a Blue one more time.

I would like to express my gratitude also to Lincoln College for these amazing years. I took advantage of a wide range of events organised in College, which were enriching in both academic and social aspects.

I am happy and honoured to be a Berrow Foundation scholar, having the opportunity to complete my DPhil at the University of Oxford, in a very enriching multidisciplinary environment. I am grateful to the Berrow Foundation for this opportunity. I would also like to thank the Fondation Zdenek et Michaela Bakala for the additional support.

Last but not least, I would like to thank my family and in particular my parents for everything they have done for me.

Thank you Oxford for the enriching experiences and lifelong memories, you will always have a special place in my heart.

Preface

This Preface was also submitted as a standalone document under the University regulations, in particular as part of the statement of authorship. Particle physics being a collaborative discipline, I explain here my contributions as presented in the thesis and my contributions outside of the thesis as part of the LHCb collaboration.

This thesis presents the measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays, which is published in the Journal of High Energy Physics (JHEP 02 (2025) 113, arXiv:2410.21115) [1]. The results of this analysis are also included in the latest LHCb γ combination (LHCb-CONF-2024-004) [2]. Two families of D meson final states are considered in the analysis. The first type of D meson final states, called ADS/GLW modes, correspond to two- and four-body final states $K^{\pm}\pi^{\mp}$, $K^{+}K^{-}$, $\pi^{+}\pi^{-}$, $K^{\pm}\pi^{\mp}\pi^{\pm}\pi^{\mp}$, $\pi^{+}\pi^{-}\pi^{+}\pi^{-}$. The second family of D meson final states are the three-body final states $K^{0}_{S}\pi^{+}\pi^{-}$, $K^{0}_{S}K^{+}K^{-}$, called BPGGSZ modes. These three-body self-conjugate final states are analysed for the first time at LHCb.

Moreover, a significant project was carried in parallel to the main analysis work presented in this thesis. This work focuses on Particle Identification (PID), contributing to the Real-Time Alignment project for LHCb. In particular, a detailed study was performed on the main PID calibration samples for kaons and pions used in LHCb, and the integration of the code developed during this project into the new PID calibration framework developed for Run 3. This work resulted in a public document (LHCb-FIGURE-2023-023) which presents the hadron PID performance with data collected in July 2023 with the LHCb Upgrade detector [3].

Thesis outline

- Chapter 1: Introduction and global context of the thesis in the field of particle physics.
- Chapter 2: Presentation of the theoretical framework of the CKM angle γ within the Standard Model and the method to measure it. Overview of the analysis strategy.
- Chapter 3: Introduction of the LHCb experiment and description of the detector and experimental setup.

- Chapter 4: Part I of the measurement, presenting the dataset and candidate selection, as well as the selection efficiencies.
- Chapter 5: Sensitivity study in order to check the feasibility of the BPGGSZ analysis and that it is a valuable measurement to perform. Preparation of the fit setup and studies of the stability of the fitting procedure.
- Chapter 6: Part II of the measurement, describing the invariant-mass fits performed to the data and the results.
- Chapter 7: Presentation of the LHCb Upgrade, focusing on the changes made to the LHCb detector.
- Chapter 8: Project on particle identification developments for the LHCb Upgrade.
- Chapter 9: Summary of the results and outlooks.

Contributions of the author

The work presented in this thesis is my own work unless clearly stated or referenced.

First, I processed the BPGGSZ data and performed a sensitivity study. I adapted the code used in Oxford for similar BPGGSZ analyses to apply the selection to the raw data samples. For the sensitivity study, I wrote a standalone package to perform the *CP* fit to measure the *CP* observables, after applying the candidate selection.

I then developed the whole code for the ADS/GLW analysis from the beginning. This includes the code to perform the selection on the ADS/GLW modes as well as the invariant-mass fits. The model developed in the ADS/GLW modes is also used for the BPGGSZ part.

For the BPGGSZ analysis, I worked closely with a postdoctoral colleague, who performed background studies using the framework I had setup, very similarly to the ADS/GLW modes described in Chapter 4. They carried out the work described in Section 4.3.5 and reoptimised the specific Boosted Decision Tree used for the BPGGSZ modes. Introduced in Chapter 6, the fit and systematic uncertainties relating to the BPGGSZ part in Sections 6.4.4 and 6.5.10 were also carried out by the colleague.

Finally, I performed the combination of the results in order to extract the final value of γ. I wrote the combiner using the software used in LHCb to perform similar combinations. This last step allowed the interpretation of the results from all modes and the determination of the CKM angle γ from this decay channel.

Other contributions and experience

On top of the work presented in this thesis, I had the opportunity to work on other projects and take part in other events which contributed to my research experience.

During my first year and in parallel to the doctoral course work, I processed data and performed studies for another analysis. I studied backgrounds and misidentification shapes in data and simulation for $B^0 \to \overline{D}^0 K\pi$ and $B^0 \to \overline{D}^0 \pi\pi$ decays. These were particularly interesting for an ongoing γ measurement in the group on which a postdoctoral colleague was working.

I was a Run 3 performance liaison for the Beauty to Open Charm (B2OC) Working Group. As part of this two-year mandate, I was responsible to attend meetings and report relevant news between B2OC and the Early Measurements Task Force (EMTF) in preparation of the early stages of the Run 3 of the experiment, as well as the WP4/5, which is the working group dealing with alignment, calibration, and data quality. In this context, I worked on an additional project with a LHCb colleague who was based at CERN. This consisted in a project aiming to develop code in order to provide automatic feedback on merge requests on GitLab, which was implemented in order to inform users about the status of the nightly builds.

I also took shifts as Data Manager in the LHCb control room during my time at CERN as part of my Long Term Attachment (LTA). During this time, I also had the chance to take part in the event organised for the launch of Run 3 and be present in the LHCb control room. There was a live broadcast, reaching over 1M views on YouTube, during which the world-record energy of 13.6 TeV was reached and for which first stable-beam collisions were recorded.

Finally, I had the chance to attend conferences where I presented results on behalf of the LHCb collaboration. In particular, I attended three major international conferences on behalf of the LHCb collaboration:

- CKM 2021 (International Workshop on the CKM Unitarity Triangle, Melbourne, Australia (online))
  Proceedings: Suljik, Fidan. γ measurements in ADS and GLW(-like) decays at LHCb, 2023. DOI:10.22323/1.411.0115

- Lepton Photon 2023 (31st International Symposium on Lepton Photon Interactions at High Energies, Melbourne, Australia)
  Proceedings: Suljik, Fidan. Measurement of the CKM angle gamma at LHCb, 2023. DOI:10.5281/zenodo.8348793

- Rencontres de Blois 2024 (35th Rencontres de Blois on Particle Physics and Cosmology, Château de Blois, France)

For the Rencontres de Blois 2024 conference, I was honoured to be invited to present the results from my measurement presented in this thesis.

Contents

1 Introduction **1**

2 Theory **5**

2.1 The Standard Model of Particle Physics . . . 5
2.2 Discrete symmetries . . . 8
2.3 *CP* violation . . . 9
2.4 Weak interaction of quarks and CKM matrix . . . 10
2.5 Unitary Triangle and CKM angle γ . . . 12
2.6 Measuring γ in $B^{\pm}$ decays . . . 14
2.6.1 Coherence factor . . . 15
2.6.2 GLW modes . . . 16
2.6.3 ADS modes . . . 18
2.6.4 BPGGSZ modes . . . 22
2.7 Analysis strategy and overview . . . 25

3 The LHCb experiment **27**

3.1 The Large Hadron Collider at CERN . . . 27
3.2 Running conditions and integrated luminosity in Run 1–2 . . . 29
3.3 The LHCb detector . . . 30
3.3.1 Tracking system . . . 31
3.3.1.1 Vertex Locator (VELO) . . . 32
3.3.1.2 Dipole magnet . . . 33
3.3.1.3 Tracking stations . . . 34
3.3.1.4 Track reconstruction . . . 35
3.3.2 Particle identification system . . . 36
3.3.2.1 Ring Imaging Cherenkov detectors (RICH) . . . 37
3.3.2.2 Calorimeters . . . 39
3.3.2.3 Muon stations . . . 39
3.3.2.4 Particle identification . . . 40
3.3.3 Trigger system . . . 41
3.3.3.1 L0 hardware trigger . . . 42
3.3.3.2 High Level software trigger . . . 42
3.3.4 Offline processing and stripping . . . 44
3.3.5 Simulation . . . 44

4 Measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays: Part I 45
4.1 Data and simulation samples . . . 45
4.2 Candidate selection . . . 48
4.2.1 Event reconstruction . . . 48
4.2.2 Truth matching . . . 48
4.2.3 Trigger requirements . . . 49
4.2.4 Preselection rectangular cuts . . . 50
4.2.4.1 K^* selection: non-resonant $B^{\pm} \to DK^0_S\pi^{\pm}$ background 52
4.2.4.2 Background from $B \to D\pi\pi\pi$ decays . . . 53
4.2.4.3 Charmless $B \to hh(hh)K^*$ background . . . 55
4.2.4.4 Clones . . . 55
4.2.5 Multivariate analysis: Boosted Decision Tree . . . 56
4.2.6 PID requirements . . . 63
4.2.7 Candidate vetoes . . . 64
4.2.7.1 Double misidentified (cross-feed) background . . . 64
4.2.7.2 Background from $D \to K^0_S\pi\pi$ in $D \to \pi\pi\pi\pi$ . . . 65
4.2.8 BDT optimisation . . . 65
4.2.9 Multiple candidates . . . 71
4.2.10 Summary of final offline cuts . . . 71
4.3 Efficiencies . . . 73
4.3.1 Selection efficiencies . . . 73
4.3.2 BDT and vetoes efficiencies . . . 74
4.3.3 PID efficiencies . . . 75
4.3.4 Total combined efficiency . . . 77
4.3.5 BPGGSZ phase-space efficiency . . . 79
4.4 Signal and backgrounds . . . 80
4.4.1 Signal shape . . . 80
4.4.2 Combined datasets . . . 82
4.4.3 Irreducible background shapes . . . 83
4.4.3.1 Combinatorial background . . . 84
4.4.3.2 Partially reconstructed $B \to D^*K^*$ backgrounds . . . 84
4.4.3.3 Charmless background . . . 88
4.4.3.4 Background from $\Lambda^0_b \to \Lambda^+_c K^{*-}$ . . . 93
4.4.3.5 Background from $B^0_s \to \bar{D}^0\bar{K}^{*0}$ . . . 95
4.4.4 Additional background studies . . . 98
4.4.4.1 Misidentified $B^{\pm} \to DK^0_S K^{\pm}$ background . . . 98
4.4.4.2 Track-swap backgrounds . . . 99
4.5 Summary . . . 102

5 Sensitivity study of the γ measurement in $B^{\pm} \to DK^{*\pm}$ decays using the $D \to K^0_S h^+ h^-$ channels 103
5.1 Introduction . . . 103
5.2 Candidate selection . . . 105
5.3 Invariant-mass fit . . . 107
5.4 Sensitivity study . . . 108
5.4.1 Expected yields . . . 109
5.4.2 *CP* fit . . . 111
5.4.3 Physics parameters fit . . . 113

6 Measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays: Part II 115
6.1 Extended maximum likelihood method . . . 115
6.2 Fit strategy . . . 116
6.3 Global fit . . . 117
6.3.1 Fit components . . . 118
6.3.2 Fit results . . . 119
6.4 *CP* fit . . . 122
6.4.1 Fit components . . . 122
6.4.2 ADS/GLW *CP* fit . . . 123
6.4.2.1 Asymmetry corrections . . . 123
6.4.2.2 Fit parametrisation . . . 126
6.4.2.3 Fit results . . . 129
6.4.2.4 Fit validation with pseudoexperiments . . . 136
6.4.3 First observation of suppressed modes . . . 140
6.4.4 BPGGSZ *CP* fit . . . 141
6.4.4.1 Intermediate fit . . . 141
6.4.4.2 Fit in Dalitz bins . . . 142
6.4.4.3 Fit results . . . 145
6.5 Systematic uncertainties . . . 146
6.5.1 Asymmetry corrections . . . 146
6.5.2 Branching fractions . . . 146
6.5.3 Efficiencies . . . 147
6.5.4 Signal shape . . . 148
6.5.5 Combinatorial background . . . 148
6.5.6 Partially reconstructed backgrounds . . . 149
6.5.7 Charmless background . . . 150
6.5.8 Background from $\Lambda^0_b \to \Lambda^+_c K^{*-}$ . . . 151
6.5.9 Background from $B^0_s \to \overline{D}^0 K^{*0}$ . . . 151
6.5.10 BPGGSZ systematic uncertainties . . . 151

6.5.11 Summary of systematic uncertainties . . . 154
6.6 Summary of results . . . 157
6.7 Combination and interpretation of results . . . 158
6.7.1 ADS/GLW . . . 159
6.7.2 BPGGSZ . . . 161
6.7.3 Combination . . . 162

7 The LHCb Upgrade detector 165
7.1 Running conditions in Run 3 . . . 165
7.2 LHCb Upgrade detector . . . 166
7.2.1 Tracking system Upgrade . . . 166
7.2.2 Particle identification system Upgrade . . . 166
7.2.3 Trigger system Upgrade . . . 167
7.3 Physics motivations and trigger efficiency . . . 168

8 Particle identification developments for the LHCb Upgrade 169
8.1 PID calibration at LHCb . . . 169
8.2 Calibration method . . . 171
8.2.1 Requirements of the calibration . . . 171
8.2.2 $D^{*+} \to D^0\pi^+$ calibration sample . . . 172
8.2.3 Fit-and-count method . . . 174
8.2.4 *sWeights* method . . . 174
8.3 Kinematic variables and binning scheme . . . 176
8.4 Signal selection and shape . . . 179
8.5 Backgrounds . . . 182
8.6 Two-dimensional model . . . 189
8.7 Previous calibration in Run 2 . . . 189
8.8 New two-dimensional data fit . . . 189
8.9 PID performance with 2023 data . . . 193

9 Summary and outlooks 199

Appendices

A *CP* fit projections 205

References 209

I, a universe of atoms, an atom in the universe.

— Richard Feynman

1 Introduction

Discovering and understanding the world around us has always been a fundamental part of human nature, driving progress and development forward. The curiosity and desire to understand the phenomena taking place in the universe is at the heart of physics, which studies the fundamental constituents of matter and their complex interactions, forming everything in the universe, from the smallest to the largest scales. Particle physics focuses on the exploration of the fundamental particles and their interactions, for which the best knowledge to date is contained in the theory of the Standard Model (SM). While successfully describing and predicting elementary particles, such as the Higgs boson discovered in 2012, it is known to be incomplete and leaves unanswered questions.

One of the limitations of the Standard Model, which is discussed in more detail in Chapter 2, is the lack of explanation of the matter-antimatter asymmetry that is observed in the universe. Particles of matter and antimatter formed in the Big Bang, assuming both types were produced in the same amount, would annihilate and no matter would remain. All the matter remaining and seen in the universe finds no explanation in the SM, where the observed *CP* violation (CPV) is not sufficient to describe this phenomenon. In fact, the only known source of CPV in the SM comes from a single phase γ in the Cabibbo-Kobayashi-Maskawa (CKM) matrix [4, 5] (and similarly if a phase appears in the Pontecorvo-Maki-Nakagawa-Sakata (PMNS)

matrix describing neutrino mixing). This thesis aims to contribute in addressing this problem, which is one of the fundamental open questions in particle physics and in science in general, by performing a measurement of γ. Direct measurements of this weak *CP*-violating phase γ are a benchmark of the SM. Therefore, the weak force and interaction of the quarks is a key place to look for the presence of New Physics (NP), corresponding to physics beyond the Standard Model (BSM).

The best place to reproduce similar conditions as the ones present shortly after the Big Bang is the Large Hadron Collider (LHC) at CERN, which is the largest particle collider and machine in the world, where particles with high energy that existed at that time can be produced in order to look for particles and study phenomena that could reveal the extension of the SM. By accelerating beams of protons at a speed close to the speed of light, proton-proton (pp) collisions take place at given interaction points, where high energy particles and antiparticles are produced. The Large Hadron Collider beauty (LHCb) experiment, located at one of these interaction points, collected pp collision data corresponding to an integrated luminosity of $9\,\text{fb}^{-1}$ during the Run 1 (2011–2012) and Run 2 (2015–2018) of the LHC, which is used for the measurement presented in this thesis. The LHCb collaboration designed a detector, presented in Chapter 3, to study the physics of the b (beauty) and c (charm) quarks and the particles they form. In this experimental setup, B mesons are produced and analysed by exploring differences between B^+ and B^- particles.

Making use of $B^\pm \to DK^{*\pm}$ decays, where the D meson is a superposition of D^0 and $\bar{D}^0$ states, the phase γ is accessed from interference effects which take place between the different paths of the B meson decay to the final state. The $K^{*\pm}$ refers to the $K^*(892)^\pm$ resonance, which is reconstructed through the decay $K^{*\pm} \to K^0_S\pi^\pm$. Different families of D meson final states are considered in this analysis. The first type of D meson final state corresponds to two- and four-body final states $K^+\pi^-$, π^+K^-, K^+K^-, $\pi^+\pi^-$, $K^+\pi^-\pi^+\pi^-$, $\pi^+K^-\pi^+\pi^-$, $\pi^+\pi^-\pi^+\pi^-$. This work supersedes a previous measurement published by LHCb on these decay modes using partial Run 2 data (2011–2016) [6]. The measurement presented in this thesis benefits

from the increased dataset by about a factor two, as well as many improvements introduced in the analysis. These improvements include an improved selection of the candidates, a better and detailed understanding of the different backgrounds as well as new D meson decay modes which are added. These additional modes correspond to the second family of D meson final states analysed, which are the three-body final states $K^0_S\pi^+\pi^-$, $K^0_S K^+K^-$. These three-body self-conjugate final states are analysed for the first time at LHCb, while previous measurements from other B-factories have been published (but not in a model-independent way). The first step of the work is to perform a sensitivity study, presented in Chapter 5, in order to check that the analysis is valuable to perform in these three-body final states, as well as to prepare the fit setup and study the stability of the fitting procedure. The measurement itself is detailed over two chapters: Chapter 4 introduces the dataset, the selection of the candidates as well as the selection efficiencies, while Chapter 6 presents the invariant-mass fits performed to the data and the results.

During the LHC Long Shutdown 2 (LS2) from 2019–2022, the LHCb detector underwent a major upgrade in order to collect data at an increased instantaneous luminosity by a factor five. The LHCb Upgrade and the main changes implemented to the detector are summarised in Chapter 7.

One important aspect of the measurement presented in this thesis is particle identification (PID). In order to select the different datasets corresponding to the list of D meson final states analysed, the main difference in the selection applied to the candidates is the PID requirement placed on the final-state tracks. A significant project carried in parallel to the analysis is also presented in this thesis, which consists in PID developments for the LHCb Upgrade. A new fit model for the main calibration sample used by LHCb for kaons and pions, using the $D^{*+} \to D^0\pi^+$ decay channel, is developed and implemented into the new PID calibration framework to be used in Run 3. This project and the detailed study is presented in Chapter 8.

The results presented in this thesis are summarised in Chapter 9, and the outlook on future expected sensitivities is discussed.

The eternal mystery of the world
is its comprehensibility.

— Albert Einstein

2 Theory

This chapter lays out the theoretical framework of the Standard Model and how to measure the CKM angle γ. The measurement of γ is described for the different families of D meson decay modes considered in this work. Finally, the strategy and an overview of the analysis is given.

2.1 The Standard Model of Particle Physics

The current best knowledge of particle physics is included in the theoretical framework known as the Standard Model. This gives a description of all the known fundamental elementary particles and the interactions governing them, which are the building blocks of matter and the universe as a whole. In total, 17 particles are included in the Standard Model and divided into two groups, 12 fermions and 5 bosons, which are presented in Figure 2.1. This table lists the particles contained in the SM and their properties, in a similar way to the chemical elements in the periodic table. In the case of the SM, the properties of the particles include the mass, the charge and the spin.

The fermions, the matter particles, come in two types, 6 quarks and 6 leptons. These particles have a half-integer spin and are described by the Fermi-Dirac statistics and satisfy the Pauli exclusion principle. The quarks are the building

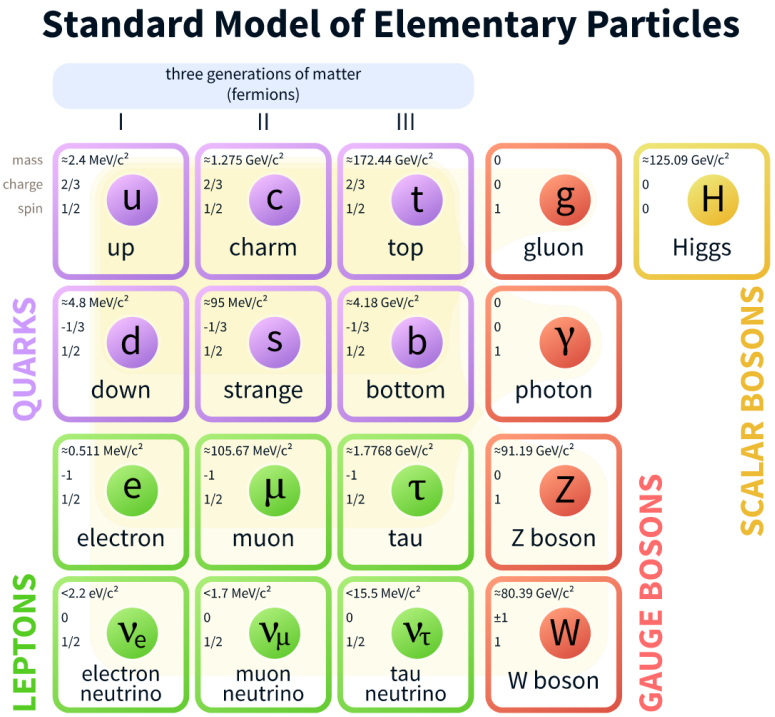


Figure 2.1: Elementary particles included in the Standard Model [7].

blocks of particles such as protons and neutrons, which form the nucleus of atoms through the strong interaction, while the leptons are not sensitive to the strong interaction. There are three generations of fermions. All visible matter in the universe is made from the first generation, consisting of the quarks up (u) and down (d), as well as the leptons of the first generation which are the electron (e) and the corresponding electron neutrino (ν_e). The second and third generations of quarks are the pairs: charm (c) and strange (s); top (t) and bottom/beauty (b). The muon (μ) and the tau (τ) are the leptons of the second and third generations, which are composed of heavier and unstable particles. Each of the charged leptons l has a corresponding neutrino ν_l, with $l \in e, \mu, \tau$, which is electrically neutral and interacts only weakly with matter. There are corresponding antimatter particles for the fermions, with additive quantum numbers being opposite such as the charge. The antimatter version of the quarks q are the antiquarks $\bar{q}$. Two or more quarks combine to form composite particles known as hadrons, which come in two groups, the baryons and mesons. The baryons are made of three quarks and belong to the fermions, while mesons are bosons formed of quark-antiquark pairs. Exotic

hadrons composed of more than three quarks, such as tetraquarks and pentaquarks, also exist and constitute recent discoveries by LHCb [8, 9]. In this thesis, $B^{\pm}$ mesons are studied, where the B^{+} meson is composed of an up quark u and a beauty antiquark $\bar{b}$. The $B^{\pm}$ mesons result from the subsequent hadronisation of b quarks formed in the collisions of protons.

The particles are governed by fundamental forces, which are mediated by force carrier particles, which belong to the second group of elementary particles in the SM, the bosons. These integer-spin particles obey the Bose-Einstein statistics. Four of the five bosons are force-mediating particles. The gluon (g) is responsible for the strong interaction, the photon (γ) mediates the electromagnetic interaction, while the weak interaction is carried by the $W^{\pm}$ and Z bosons. The last piece of the SM is the Higgs boson (H), which is a scalar boson in the Higgs mechanism through which the particles acquire mass.

Only three fundamental forces are incorporated in the SM. The fourth fundamental interaction, which is gravity, is not included in the SM despite being the most obvious in the large scale and day-to-day life. However, at the subatomic scale, gravity is much weaker and becomes negligible compared to the other three fundamental forces, being around 38 orders of magnitude weaker than the strong interaction. At the gigaelectronvolt (GeV) scale, the intensity ratios between the different forces are given in Table 2.1. The weak force interacts at short distances and can be stronger than the electromagnetic force.

Table 2.1: Intensity ratios of the fundamental interactions at the GeV scale [10].

Force	Relative strength
Strong	1
Electromagnetic	10^{-2}
Weak	10^{-5}
Gravity	10^{-38}

The limitations of the SM go beyond the non-incorporation of gravity, for which other theories propose the existence of a boson particle which would act as the

mediator of the gravity fundamental force. While the SM represents the best current knowledge and successfully describes and explains results to a very high precision, it still remains incomplete and fails to explain different phenomena.

The SM does not provide a dark matter particle candidate, while cosmological observations of the rotational speed of galaxies show that additional mass is needed to explain the higher speed than theory predictions which take into account all known visible matter [11]. Moreover, the SM does not provide an explanation for the fact that exactly three generations of fermions exist, nor for the large difference in mass scales between the particles which is known as the mass hierarchy problem. Concerning neutrinos and their nature, it is also unknown if their behaviour follows the one of Majorana or Dirac fermions. In the case of Majorana particle, this would imply that the neutrino is its own antiparticle.

One of the main open questions in particle physics and science is the matter-antimatter asymmetry problem, which is not solved in the SM. It is assumed that an equal amount of matter and antimatter was formed in the Big Bang, in which case all particles of both types would annihilate into pure energy, not leaving any remaining matter as observed today. Three conditions were proposed by Andrei Sakharov which are necessary for the baryon asymmetry in the universe to happen [12], one of which is CP violation, the combined violation of Charge-Conjugation (C) and Parity (P) symmetry. However, the current amount of observed CP violation in the SM is insufficient to describe and explain the proportion of matter seen in the universe. This is the area of interest of the work detailed in this thesis, which presents a CP violation measurement in charged B decays, which aims to contribute in answering this open question of particle physics.

2.2 Discrete symmetries

Symmetry is a fundamental concept and its properties play an important role in physics and in science in general. Physical systems presenting a continuous symmetry display quantities which are conserved. This is contained in Noether's theorem [13], for which an example is the conservation of momentum in systems

which are invariant under translation. Discrete symmetries are of particular interest in particle physics and the SM, such as the

- C: Charge-conjugation operation, transforming a particle into the corresponding antiparticle by conjugating the internal quantum numbers
- P: Parity operation, inverting the spatial coordinates $(t, \vec{x}) \rightarrow (t, -\vec{x})$
- T: Time reversal operation, inverting the time coordinate $(t, \vec{x}) \rightarrow (-t, \vec{x})$.

These symmetries have been found to be broken individually, by experiments, while the CPT theorem [14] states that physical systems are invariant under the simultaneous transformation of the three operators. The SM predicts CPT to be a fundamental and exact symmetry of the universe. In 1956, Chien-Shiung Wu's experiment with cobalt-60 nuclei showed P symmetry breaking in weak interactions [15], which constitutes the first observation of violation of such discrete symmetry. The experiment was previously proposed by T. D. Lee and C. N. Yang [16], for which they were awarded the Nobel Prize. While C violation was also observed in pion decays, the CP symmetry was still assumed to be a true symmetry.

2.3 CP violation

CP violation can appear in three different forms:

- Direct CP violation: CP violation in decay which takes place when a given decay and its CP conjugate process present different amplitudes $\Gamma(B \rightarrow f) \neq \Gamma(\bar{B} \rightarrow \bar{f})$.
- Indirect CP violation: CP violation in mixing occurs in neutral meson systems where the rates of the two CP-conjugate states are different $\Gamma(B^0 \rightarrow \overline{B}^0) \neq \Gamma(\overline{B}^0 \rightarrow B^0)$.
- CP violation in interference: CP violation in interference between decay and mixing, which can appear in the decay of a neutral meson to a final state which is also accessible to its CP-conjugate particle.

The *CP* symmetry was observed to be violated in 1964 in an experiment by Cronin and Fitch [17]. This *CP* violation was observed in neutral kaon decays, which were expected to decay to the *CP*-odd three-pion final states, while decays to the *CP*-even two-pion final states were also observed. In the late 1990s, results on direct *CP* violation in the neutral kaon system were presented by the NA48 and KTeV collaborations [18, 19]. In 2001, *CP* violation was also observed in the B^0 system by the B-factories BaBar and Belle [20, 21], while LHCb reported the observation of *CP* violation in the charged $B^{\pm}$ decays in the early 2010s [22]. In 2013, LHCb also observed *CP* violation in the B_s^0 system [23]. In 2019, the LHCb experiment measured the first *CP* violation in the D^0 meson decay [24], while recent evidence of *CP* violation in neutrino oscillation was published by T2K [25]. More recently, in 2025, the LHCb collaboration announced the first observation of *CP* violation in Λ_b^0 baryon decays [26].

The matter-dominated universe observed today is currently not explained by the SM, where all the sources of *CP* violation observed are not sufficient to solve the matter-antimatter asymmetry problem. However, the *CP* violation measurements observed within the quarks are described by the SM.

2.4 Weak interaction of quarks and CKM matrix

The mixing of quarks was initially explained in the SM by the matrix proposed by Nicola Cabibbo in 1963 [4], which contained one parameter and successfully described the weak interactions of the first two generations of quarks. At that time, the third generation of quarks was still unknown. Following the discovery of *CP* violation in 1964, which was not explained within the 2×2 Cabibbo matrix, effort was devoted to develop the theory in order to incorporate and explain the *CP* violation within the SM. In 1973, Kobayashi and Maskawa extended this matrix to the three generations of quarks [5], for which work they received the Nobel Prize in Physics in 2008. This resulted in a three-dimensional matrix V_{CKM}, called the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which describes the weak transition

between the three generations of quarks. This relates the mass eigenstates (d,s,b) to the weak eigenstates (d',s',b') in the following way:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.1}$$

where each element of the CKM matrix is proportional to the amplitude of the transition between the corresponding quarks. The CKM matrix is a 3×3 complex matrix, which is unitary and has 9 real and independent parameters. Five of the parameters can be reabsorbed as arbitrary non-physical phases, which leaves four parameters needed for the parametrisation. These are three Euler mixing angles and one single complex phase. In the standard parametrisation, or Chau-Keung parametrisation [27], the matrix can be written as

$$V_{\mathrm{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{2.2}$$

where θ_{12}, θ_{23} and θ_{13} are three mixing angles, with the notation $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The *CP*-violating complex phase δ is the only source of all the *CP* violation in the quark sector of the SM. This is due to the fact that this phase enters with the opposite sign in the amplitude terms of the antiparticles. While the SM requires the unitarity of V_{CKM}, the values of the individual elements of the CKM matrix are not predicted and can be measured independently. From a global fit, the values of the elements are determined to be [28]:

$$|V_{\mathrm{CKM}}| = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22501 \pm 0.00068 & 0.003732^{+0.000090}_{-0.000085} \\ 0.22487 \pm 0.00068 & 0.97349 \pm 0.00016 & 0.04183^{+0.00079}_{-0.00069} \\ 0.00858^{+0.00019}_{-0.00017} & 0.04111^{+0.00077}_{-0.00068} & 0.999118^{+0.000029}_{-0.000034} \end{pmatrix}. \tag{2.3}$$

The values of the diagonal elements are close to unity, which shows the fact that transitions of quarks within the same generation are favoured. These transitions appear in favoured decays in the SM and are called Cabibbo-favoured decays. The off-diagonal elements are smaller and correspond to Cabibbo-suppressed transitions. In terms of the standard parametrisation, this hierarchy of the size of the CKM elements can be written as $1 \gg s_{12} \gg s_{23} \gg s_{13}$. It is convenient to introduce another parametrisation, called the Wolfenstein parametrisation [29], which gives

an approximation of the standard parametrisation, incorporating the hierarchy of the elements. This is done by defining the parameter $\lambda \equiv s_{12}$, which is used to perform a series expansion. Using this parametrisation with an approximation to the order λ^3, the CKM matrix can be written as

$$V_{\mathrm{CKM}} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\bar{\rho}-i\bar{\eta}) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\bar{\rho}-i\bar{\eta}) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)\,, \tag{2.4}$$

where $A\lambda^2 \equiv s_{23}$ and $A\lambda^3(\bar{\rho}-i\bar{\eta}) \equiv s_{13}s_{13}e^{i\delta}$. It can be seen that *CP* violation is directly related to the parameters $\bar{\rho}$ and $\bar{\eta}$ through the complex term $\bar{\rho}-i\bar{\eta}$.

2.5 Unitary Triangle and CKM angle γ

From the unitarity condition of the CKM matrix, $V_{\mathrm{CKM}}V_{\mathrm{CKM}}^{\dagger} = V_{\mathrm{CKM}}^{\dagger}V_{\mathrm{CKM}} = 1$, nine independent equations are obtained, of which one can be written as:

$$V_{ud}V_{ub}^{*} + V_{cd}V_{cb}^{*} + V_{td}V_{tb}^{*} = 0. \tag{2.5}$$

This can be represented by a triangle in the complex plane, as illustrated in Figure 2.2, whose area is proportional to the amount of *CP* violation present in the SM.

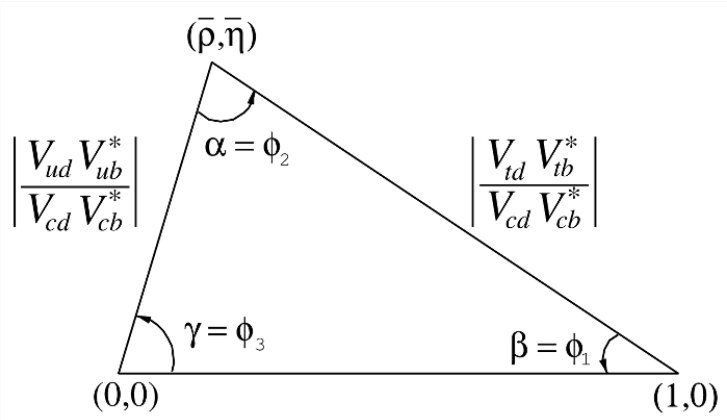


Figure 2.2: Illustration of the CKM unitary triangle in the complex plane [28].

One of the three internal angles of the triangle is the CKM angle γ, which is defined as

$$\gamma \equiv \arg\left(-\frac{V_{ud}V_{ub}^{*}}{V_{cd}V_{cb}^{*}}\right). \tag{2.6}$$

From this definition, it can be seen that γ is not dependent on the elements of the CKM matrix which involve the top quark. Therefore, the weak phase γ is the only angle which is easily accessible at tree level and measured in tree-level decays, where BSM physics effects are not expected to contribute [30, 31]. Direct measurements of the CKM angle γ at tree-level can thus be benchmarks of the SM. Flavour physics observables often can have large theoretical uncertainties due to the fact that quarks are not free but come in bound states and therefore the impact of quantum chromodynamics (QCD) has to be taken into account. As will be shown, in γ measurements it is possible to extract the QCD part directly from the data. Therefore, since the QCD effects can be determined directly from data, the remaining theoretical uncertainty on γ is negligible [32].

On the other hand, indirect measurements of γ can be obtained by performing a global fit to the unitary triangle, assuming a closed triangle. In this global fit, independent measurements of the other angles and the sides of the triangles, which depend on the elements of the CKM matrix, are used. This indirect determination of γ is dominated by the measurement of the angle β and neutral B-meson mixing. These are loop-level processes, where New Physics effects would be expected to contribute [33]. Therefore, a key place to look for BSM physics is within the weak interactions of the quarks. This is done by comparing direct and indirect measurements of γ and looking for any discrepancy between the direct and indirect measurements which would be a clear sign of New Physics. The status as of Summer 2023[1] is a value of $\gamma = (66.29^{+0.72}_{-1.86})^\circ$ obtained from the global fit, while direct measurements yield a result of $\gamma = (65.9^{+3.3}_{-3.5})^\circ$ [34]. The two corresponding CKM triangles are illustrated in Figure 2.3. A recent combination of direct measurements from LHCb yields a value of $\gamma = (63.8^{+3.5}_{-3.7})^\circ$ [35]. It is therefore important to further improve the precision of the direct measurement in order for it to be compared with the indirect determination of γ which is currently more precise.

[1]The status is presented here using results that do not include the work from this thesis.

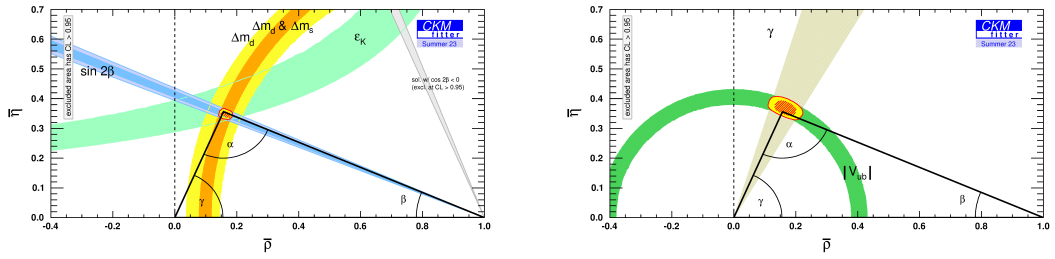


Figure 2.3: The unitary triangle illustrating the results from indirect (left) and direct (right) measurements [34].

2.6 Measuring γ in $B^{\pm}$ decays

This thesis presents a measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays. This section presents how γ can be measured in tree-level B meson decays and the corresponding method used for each family of D meson final states analysed. In $B^{\pm} \to DK^{*\pm}$ decays, the D meson is a superposition of D^0 and $\overline{D}^0$ states that decay to the same final state. For the B^- decay, the favoured decay involves a $b \to c$ quark transition in $B^- \to D^0 K^{*-}$, while the suppressed decay $B^- \to \overline{D}^0 K^{*-}$ involves a $b \to u$ quark transition. The Feynman diagrams of the two corresponding decays for B^- are illustrated in Figure 2.4.

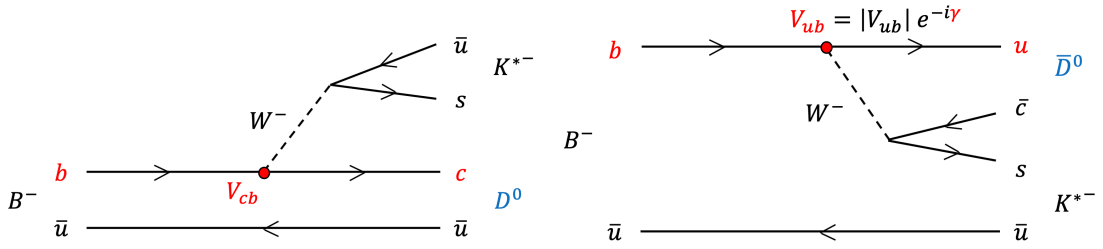


Figure 2.4: Feynman diagrams of $B^- \to DK^{*-}$ decays, for the favoured (left) and suppressed (right) decays.

The use of common D decays accessible to both the D^0 and $\overline{D}^0$ states gives an interference between the $b \to c\bar{u}s$ and $b \to u\bar{c}s$ transitions occurring in the $B^{\pm} \to DK^{*\pm}$ decays. The amplitudes of the $B^- \to D^0 K^{*-}$ and $B^- \to \bar{D}^0 K^{*-}$ decays are proportional to the CKM elements V_{cb} and V_{ub}, respectively. The interference between the two amplitudes gives access to the phase information, which can be obtained by studying final states which are accessible to both flavours

of the D meson, D^0 and $\overline{D}^0$. The amplitude of the $B^- \to D^0 K^{*-}$ is defined as

$$\mathcal{A}(B^- \to D^0 K^{*-}) \equiv A_B, \tag{2.7}$$

while the amplitude of $B^- \to \bar{D}^0 K^{*-}$ is given by

$$\mathcal{A}(B^- \to \bar{D}^0 K^{*-}) = A_B \cdot r_B \cdot e^{i(\delta_B - \gamma)}. \tag{2.8}$$

The parameter δ_B represents the strong-phase difference between the two amplitudes and r_B corresponds to the ratio between the two decay amplitudes:

$$r_B = \frac{|\mathcal{A}(B^- \to \bar{D}^0 K^{*-})|}{|\mathcal{A}(B^- \to D^0 K^{*-})|}. \tag{2.9}$$

The strong-phase δ_B comes from strong and electromagnetic interactions and is a *CP*-conserving phase, unlike the weak phase which is a *CP*-violating phase. The expressions of the amplitudes for the corresponding *CP*-conjugate B^+ decays can be obtained by changing the sign of γ ($\gamma \leftrightarrow -\gamma$). The two physics parameters r_B and δ_B are determined alongside γ in the measurement presented in this thesis.

2.6.1 Coherence factor

In the $B^\pm \to DK^{*\pm}$ decays used in this analysis, the $K^{*\pm}$ refers to the $K^*(892)^\pm$ resonance, which is reconstructed through the decay $K^{*\pm} \to K^0_S \pi^\pm$. Given the large natural width of the K^* resonance, the amplitudes and strong phases defined in Section 2.6 depend on the coordinate p of the phase space of the final state $DK^0_S\pi^\pm$ in the following way,

$$\mathcal{A}(B^- \to D^0 K^0_S \pi^-) = A_c(p) \cdot e^{i(\delta_c(p))}, \tag{2.10}$$

$$\mathcal{A}(B^- \to \bar{D}^0 K^0_S \pi^-) = A_u(p) \cdot e^{i(\delta_u(p) - \gamma)}. \tag{2.11}$$

In these expressions, the magnitudes of the amplitudes of $b \to c$ and $b \to u$ transitions are $A_c(p)$ and $A_u(p)$ respectively, with the corresponding strong phases being $\delta_c(p)$ and $\delta_u(p)$. For the *CP*-conjugate decays of B^+ mesons, the amplitudes are obtained by flipping the sign of γ. Both the resonant $K^{*\pm}$ and the non-resonant $K^0_S\pi^\pm$ contributions, or contributions from other resonances, are included in these

decays. However, the analysis focuses on the $K^{*\pm}$ resonance. The amplitude ratio and the difference in strong phase between the two B^- decays can be expressed as

$$r_B^2 = \frac{\Gamma(B^- \to \bar{D}^0 K^{*-})}{\Gamma(B^- \to D^0 K^{*-})} = \frac{\int_{K^*} |A_u(p)|^2 \mathrm{d}p}{\int_{K^*} |A_c(p)|^2 \mathrm{d}p}, \tag{2.12}$$

$$\kappa e^{i\delta_B} = \frac{\int_{K^*} A_c(p) A_u(p) e^{i(\delta_u(p) - \delta_c(p))} \mathrm{d}p}{\sqrt{\int_{K^*} |A_c(p)|^2 \mathrm{d}p} \sqrt{\int_{K^*} |A_u(p)|^2 \mathrm{d}p}}, \tag{2.13}$$

where κ is the coherence factor that takes into account the contribution not coming from the $K^*(892)^{\pm}$ resonance. The presence of these other interfering contributions would dilute the sensitivity to the weak phase γ. The coherence factor satisfies $0 \leq \kappa \leq 1$, where a value of 1 indicates a pure resonant data sample. The coherence factor has been estimated using an amplitude model developed by considering resonant and non-resonant $K^0_S \pi^{\pm}$ contributions in the region of the $K^*(892)^{\pm}$ resonance of interest, giving a value of (0.95 ± 0.06) [36]. This value is taken as an input in the analysis presented here.

2.6.2 GLW modes

Different D meson final states are analysed in the measurement presented in this thesis. The first class of D meson final states considered are the *CP* eigenstates $D \to K^- K^+$ and $D \to \pi^- \pi^+$. These are referred to as GLW modes, after the theorists Gronau, London and Wyler [37, 38]. Due to the fact that these are *CP*-even eigenstates, the amplitudes for the D^0 and $\bar{D}^0$ meson decaying into these final states are equal, as illustrated in Figure 2.5. This means that the overall amplitudes between the two decay paths are of different size. In turn, the amount of *CP* violation that can be observed in these decays is relatively small, driven by the value of the r_B ratio which is small.

The total amplitude for the corresponding B^- and B^+ decays can be written as

$$\begin{aligned} \mathcal{A}(B^- \to [f]_D K^{*-}) &= A_B A_D + A_B r_B e^{i(\delta_B - \gamma)} A_D \\ &= A_B A_D [1 + r_B e^{i(\delta_B - \gamma)}], \end{aligned} \tag{2.14}$$

$$\begin{aligned} \mathcal{A}(B^+ \to [\bar{f}]_D K^{*+}) &= A_B A_D + A_B r_B e^{i(\delta_B + \gamma)} A_D \\ &= A_B A_D [1 + r_B e^{i(\delta_B + \gamma)}], \end{aligned} \tag{2.15}$$

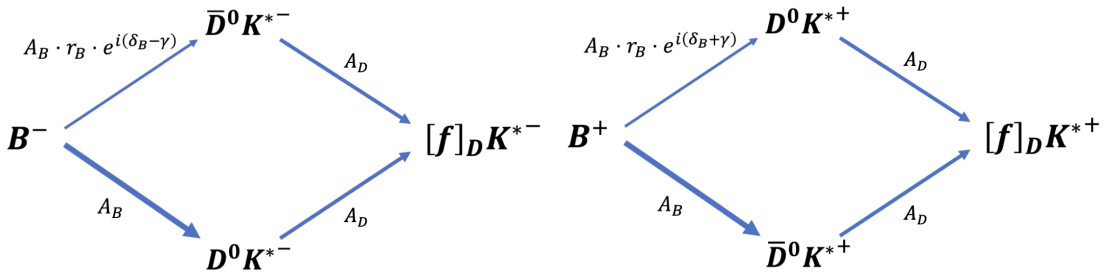


Figure 2.5: Interference diagram for $B^- \to DK^{*-}$ (left) and $B^+ \to DK^{*+}$ (right) with CP-eigenstates of the D decay.

where f is the D meson final state and A_D the magnitude of the amplitude of the D meson decay. The total decay rates, integrating over the K^* phase space, are given by

$$\Gamma(B^- \to [f]_D K^{*-}) \propto 1 + r_B^2 + 2\kappa r_B \cos(\delta_B - \gamma), \tag{2.16}$$

$$\Gamma(B^+ \to [\bar{f}]_D K^{*+}) \propto 1 + r_B^2 + 2\kappa r_B \cos(\delta_B + \gamma), \tag{2.17}$$

where it can be seen that γ enters each expression with a differing sign. In order to measure γ, CP observables are constructed which are rate ratios and rate asymmetries between B^- and B^+ decays. These are experimentally preferred compared to measuring rates directly. The CP observables which are directly related to the physics parameters of interest, including γ, are given by

$$A_{KK} = \frac{\Gamma(B^- \to [K^+K^-]_D K^{*-}) - \Gamma(B^+ \to [K^+K^-]_D K^{*+})}{\Gamma(B^- \to [K^+K^-]_D K^{*-}) + \Gamma(B^+ \to [K^+K^-]_D K^{*+})} = \frac{2\kappa r_B \sin\delta_B \sin\gamma}{R_{KK}}, \tag{2.18}$$

$$A_{\pi\pi} = \frac{\Gamma(B^- \to [\pi^+\pi^-]_D K^{*-}) - \Gamma(B^+ \to [\pi^+\pi^-]_D K^{*+})}{\Gamma(B^- \to [\pi^+\pi^-]_D K^{*-}) + \Gamma(B^+ \to [\pi^+\pi^-]_D K^{*+})} = \frac{2\kappa r_B \sin\delta_B \sin\gamma}{R_{\pi\pi}}, \tag{2.19}$$

$$\begin{aligned} R_{KK} &= \frac{\Gamma(B^- \to [K^+K^-]_D K^{*-}) + \Gamma(B^+ \to [K^+K^-]_D K^{*+})}{\Gamma(B^- \to [K^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [K^+\pi^-]_D K^{*+})} \times \frac{\mathcal{B}(D^0 \to K^-\pi^+)}{\mathcal{B}(D^0 \to K^+K^-)} \\ &= 1 + r_B^2 + 2\kappa r_B \cos\delta_B \cos\gamma, \end{aligned} \tag{2.20}$$

$$\begin{aligned} R_{\pi\pi} &= \frac{\Gamma(B^- \to [\pi^+\pi^-]_D K^{*-}) + \Gamma(B^+ \to [\pi^+\pi^-]_D K^{*+})}{\Gamma(B^- \to [K^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [K^+\pi^-]_D K^{*+})} \times \frac{\mathcal{B}(D^0 \to K^-\pi^+)}{\mathcal{B}(D^0 \to \pi^+\pi^-)} \\ &= 1 + r_B^2 + 2\kappa r_B \cos\delta_B \cos\gamma. \end{aligned} \tag{2.21}$$

This method can be extended to the quasi-GLW [39] final state $D \to \pi^-\pi^+\pi^-\pi^+$, which is predominantly a *CP*-even eigenstate. This is achieved by taking into account the fraction of *CP*-even states F_+ in the mixture of *CP*-even and *CP*-odd states composing this final state. This parameter has been measured to be $0.735 \pm 0.015 \pm 0.005$ and is taken from Reference [40]. The corresponding *CP* observables are

$$\begin{aligned} A_{\pi\pi\pi\pi} &= \frac{\Gamma(B^- \to [\pi\pi\pi\pi]_D K^{*-}) - \Gamma(B^+ \to [\pi\pi\pi\pi]_D K^{*+})}{\Gamma(B^- \to [\pi\pi\pi\pi]_D K^{*-}) + \Gamma(B^+ \to [\pi\pi\pi\pi]_D K^{*+})} \\ &= \frac{2\kappa(2F_+ - 1) r_B \sin\delta_B \sin\gamma}{1 + r_B^2 + 2\kappa(2F_+ - 1) r_B \cos\delta_B \cos\gamma}, \end{aligned} \tag{2.22}$$

$$\begin{aligned} R_{\pi\pi\pi\pi} &= \frac{\Gamma(B^- \to [\pi^+\pi^-\pi^+\pi^-]_D K^{*-}) + \Gamma(B^+ \to [\pi^+\pi^-\pi^+\pi^-]_D K^{*+})}{\Gamma(B^- \to [K^-\pi^+\pi^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [K^+\pi^-\pi^+\pi^-]_D K^{*+})} \\ &\qquad \times \frac{\mathcal{B}(D^0 \to K^-\pi^+\pi^-\pi^+)}{\mathcal{B}(D^0 \to \pi^+\pi^-\pi^+\pi^-)} \\ &= 1 + r_B^2 + 2\kappa(2F_+ - 1) r_B \cos\delta_B \cos\gamma. \end{aligned} \tag{2.23}$$

2.6.3 ADS modes

The second type of final states considered are the non-*CP* eigenstates $D \to K^{\pm}\pi^{\mp}$. These consist of the Cabibbo-favoured (CF) decays $D^0 \to K^-\pi^+$, $\overline{D}^0 \to K^+\pi^-$, and the doubly Cabibbo-suppressed (DCS) decays $D^0 \to \pi^-K^+$, $\overline{D}^0 \to \pi^+K^-$. These final states are called ADS modes, after the method proposed by the theorists Atwood, Dunietz and Soni [41, 42]. In this case, the amplitudes of the D^0 and $\overline{D}^0$ meson decaying into the same final state are therefore of different size.

Two categories need to be distinguished in this method, which are the favoured and suppressed modes. The favoured modes correspond to the decays where both the B meson and the D meson decay via favoured decays, such as $B^- \to D^0K^{*-}$, followed by $D^0 \to K^-\pi^+$. The favoured modes correspond to the decays where the pion charge from the D meson is opposite to the B charge, and are denoted $D \to K\pi$ throughout the document. As the amplitudes of the D^0 and $\overline{D}^0$ meson decaying to the same final state are of different size, it is necessary to describe the physics of the D meson decay. This is achieved through the parameters

$r_D^{K\pi}$, $\delta_D^{K\pi}$, which appear in the decay rates. These correspond to the amplitude ratio between the favoured and suppressed D meson decay, and the corresponding strong-phase difference, respectively.

$$r_D^{K\pi} e^{-i\delta_D^{K\pi}} = \frac{\mathcal{A}(\overline{D}^0 \to \pi^+ K^-)}{\mathcal{A}(D^0 \to K^- \pi^+)}. \tag{2.24}$$

The $r_D^{K\pi}$, $\delta_D^{K\pi}$ parameters, which appear in the corresponding CP observables, are needed and taken as external inputs [35]. The total amplitudes and decay rates for these modes are given by

$$\begin{aligned} \mathcal{A}(B^- \to [f]_D K^{*-}) &= A_B A_D + A_B r_B e^{-i(\delta_B - \gamma)} A_D r_D e^{-i\delta_D} \\ &= A_B A_D [1 + r_B r_D e^{i(\delta_B - \gamma)} e^{-i\delta_D}], \end{aligned} \tag{2.25}$$

$$\begin{aligned} \mathcal{A}(B^+ \to [\bar{f}]_D K^{*+}) &= A_B A_D + A_B r_B e^{i(\delta_B + \gamma)} A_D r_D e^{-i\delta_D} \\ &= A_B A_D [1 + r_B r_D e^{i(\delta_B + \gamma)} e^{-i\delta_D}], \end{aligned} \tag{2.26}$$

$$\Gamma(B^- \to [f]_D K^{*-}) \propto 1 + r_B^2 r_D^2 + 2\kappa r_B r_D \cos(\delta_B - \delta_D - \gamma), \tag{2.27}$$

$$\Gamma(B^+ \to [\bar{f}]_D K^{*+}) \propto 1 + r_B^2 r_D^2 + 2\kappa r_B r_D \cos(\delta_B - \delta_D + \gamma). \tag{2.28}$$

However, in these favoured decays, the amplitudes between the two decay paths are of significantly different size. The interference is therefore highly suppressed in these modes and the sensitivity to γ is limited. These modes are used as normalisation channels in the analysis, with respect to which the rates are measured, while the asymmetry is measured in the favoured mode through the CP observable

$$\begin{aligned} A_{K\pi} &= \frac{\Gamma(B^- \to [K^-\pi^+]_D K^{*-}) - \Gamma(B^+ \to [K^+\pi^-]_D K^{*+})}{\Gamma(B^- \to [K^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [K^+\pi^-]_D K^{*+})} \\ &= \frac{2\kappa r_B r_D^{K\pi} \sin(\delta_B - \delta_D^{K\pi}) \sin\gamma}{1 + r_B^2 (r_D^{K\pi})^2 + 2\kappa r_B r_D^{K\pi} \cos(\delta_B - \delta_D^{K\pi}) \cos\gamma}. \end{aligned} \tag{2.29}$$

The second category is made of the suppressed modes, corresponding to decays where the B meson decays via favoured decay, followed by a suppressed D meson decay, such as $B^- \to D^0 K^{*-}$, followed by $D^0 \to \pi^- K^+$. These decays correspond to the decays where the pion charge coming from the D meson and the B charge are the same, and are denoted $D \to \pi K$. The interference diagram is illustrated in Figure 2.6 for both B^- (left) and B^+ (right) decays.

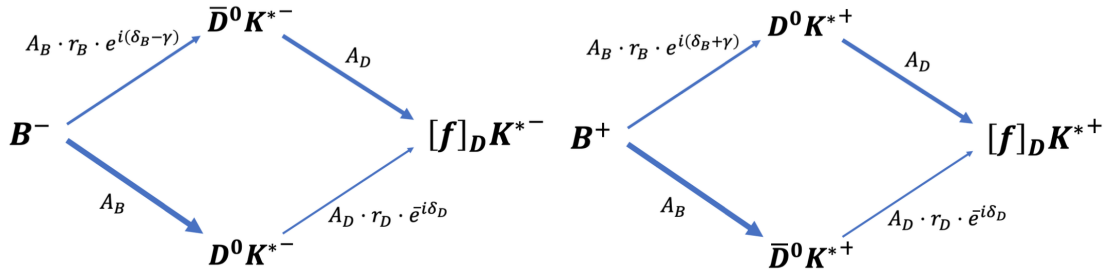


Figure 2.6: Interference diagram for $B^- \to DK^{*-}$ (left) and $B^+ \to DK^{*+}$ (right) with non-CP eigenstates of the D decay.

The favoured B decay which is followed by a suppressed D decay, interferes with the suppressed B decay which is followed by a favoured D decay. This is a key difference which makes the two paths of the decay to have similar sized amplitudes overall, where larger interference effects can occur. For these modes, the total amplitudes and decay rates are written as

$$\begin{aligned}\mathcal{A}(B^- \to [\bar{f}]_D K^{*-}) &= A_B A_D r_D e^{-i\delta_D} + A_B r_B e^{i(\delta_B-\gamma)} A_D \\ &= A_B A_D [r_B e^{i(\delta_B-\gamma)} + r_D e^{-i\delta_D}],\end{aligned} \tag{2.30}$$

$$\begin{aligned}\mathcal{A}(B^+ \to [f]_D K^{*+}) &= A_B A_D r_D e^{-i\delta_D} + A_B r_B e^{i(\delta_B+\gamma)} A_D \\ &= A_B A_D [r_B e^{i(\delta_B+\gamma)} + r_D e^{-i\delta_D}],\end{aligned} \tag{2.31}$$

$$\Gamma(B^- \to [\bar{f}]_D K^{*-}) \propto r_B^2 + r_D^2 + 2\kappa r_B r_D \cos(\delta_B + \delta_D - \gamma), \tag{2.32}$$

$$\Gamma(B^+ \to [f]_D K^{*+}) \propto r_B^2 + r_D^2 + 2\kappa r_B r_D \cos(\delta_B + \delta_D + \gamma). \tag{2.33}$$

The CP observables for the suppressed modes, constructed in a similar way by taking rate ratios and asymmetries, are given by

$$\begin{aligned}A_{\pi K} &= \frac{\Gamma(B^- \to [\pi^- K^+]_D K^{*-}) - \Gamma(B^+ \to [\pi^+ K^-]_D K^{*+})}{\Gamma(B^- \to [\pi^- K^+]_D K^{*-}) + \Gamma(B^+ \to [\pi^+ K^-]_D K^{*+})} \\ &= \frac{2\kappa r_B r_D^{K\pi} \sin(\delta_B + \delta_D^{K\pi}) \sin\gamma}{R_{\pi K}},\end{aligned} \tag{2.34}$$

$$\begin{aligned}R_{\pi K} &= \frac{\Gamma(B^- \to [\pi^- K^+]_D K^{*-}) + \Gamma(B^+ \to [\pi^+ K^-]_D K^{*+})}{\Gamma(B^- \to [K^- \pi^+]_D K^{*-}) + \Gamma(B^+ \to [K^+ \pi^-]_D K^{*+})} \\ &= r_B^2 + (r_D^{K\pi})^2 + 2\kappa r_B r_D^{K\pi} \cos(\delta_B + \delta_D^{K\pi}) \cos\gamma.\end{aligned} \tag{2.35}$$

The ADS method is also extended and used for the favoured and suppressed four-body final states $D \to K^-\pi^+\pi^-\pi^+$, $D \to \pi^-K^+\pi^-\pi^+$ [41, 42]. This is possible by using the corresponding four-body D decay parameters $r_D^{K3\pi}$, $\delta_D^{K3\pi}$. However, in these four-body modes, a single strong-phase parameter is not sufficient to parametrise the strong phase as in the two-body modes. This is due to the fact that the strong phase varies across the phase space of the D meson decay. The hadronic parameters are averaged by integrating over the phase space, where an additional parameter $\kappa^{K3\pi}$ is introduced to take into account the variation of the amplitudes which dilutes the interference effects.

$$(r_D^{K3\pi})^2 = \frac{\int |A_{\pi K\pi\pi}(p)|^2 \mathrm{d}p}{\int |A_{K\pi\pi\pi}(p)|^2 \mathrm{d}p}, \tag{2.36}$$

$$\kappa^{K3\pi} e^{i\delta_D^{K3\pi}} = \frac{\int A_{K\pi\pi\pi}(p) A_{\pi K\pi\pi}(p) \mathrm{d}p}{\sqrt{\int |A_{K\pi\pi\pi}(p)|^2 \mathrm{d}p}\sqrt{\int |A_{\pi K\pi\pi}(p)|^2 \mathrm{d}p}}, \tag{2.37}$$

where the amplitudes in a given phase-space coordinate $A_{K\pi\pi\pi}(p)$ and $A_{\pi K\pi\pi}(p)$ correspond to the amplitudes of the favoured and suppressed four-body D decays, respectively. The four-body D decay parameters $r_D^{K3\pi}$, $\delta_D^{K3\pi}$, $\kappa^{K3\pi}$ have been measured and are taken as inputs from a combination of results from BESIII, CLEO-c and LHCb collaborations [43]. Similarly to the two-body modes, the *CP* observables for the four-body modes are

$$\begin{aligned} A_{K\pi\pi\pi} &= \frac{\Gamma(B^- \to [K^-\pi^+\pi^-\pi^+]_D K^{*-}) - \Gamma(B^+ \to [K^+\pi^-\pi^+\pi^-]_D K^{*+})}{\Gamma(B^- \to [K^-\pi^+\pi^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [K^+\pi^-\pi^+\pi^-]_D K^{*+})} \\ &= \frac{2\kappa r_B \kappa^{K3\pi} r_D^{K3\pi} \sin(\delta_B - \delta_D^{K3\pi}) \sin\gamma}{1 + r_B^2 (r_D^{K3\pi})^2 + 2\kappa r_B \kappa^{K3\pi} r_D^{K3\pi} \cos(\delta_B - \delta_D^{K3\pi}) \cos\gamma}, \end{aligned} \tag{2.38}$$

$$\begin{aligned} A_{\pi K\pi\pi} &= \frac{\Gamma(B^- \to [\pi^-K^+\pi^-\pi^+]_D K^{*-}) - \Gamma(B^+ \to [\pi^+K^-\pi^+\pi^-]_D K^{*+})}{\Gamma(B^- \to [\pi^-K^+\pi^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [\pi^+K^-\pi^+\pi^-]_D K^{*+})} \\ &= \frac{2\kappa r_B \kappa^{K3\pi} r_D^{K3\pi} \sin(\delta_B + \delta_D^{K3\pi}) \sin\gamma}{R_{\pi K\pi\pi}}, \end{aligned} \tag{2.39}$$

$$\begin{aligned} R_{\pi K\pi\pi} &= \frac{\Gamma(B^- \to [\pi^-K^+\pi^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [\pi^+K^-\pi^+\pi^-]_D K^{*+})}{\Gamma(B^- \to [K^-\pi^+\pi^-\pi^+]_D K^{*-}) + \Gamma(B^+ \to [K^+\pi^-\pi^+\pi^-]_D K^{*+})} \\ &= r_B^2 + (r_D^{K3\pi})^2 + 2\kappa r_B \kappa^{K3\pi} r_D^{K3\pi} \cos(\delta_B + \delta_D^{K3\pi}) \cos\gamma. \end{aligned} \tag{2.40}$$

2.6.4 BPGGSZ modes

The final family of D meson decays considered are the three-body self-conjugate final states $D \to K^0_S\pi^+\pi^-$ and $D \to K^0_S K^+K^-$. These are called the BPGGSZ modes from the formalism developed in References [44–46]. These final states form complex systems of resonances, where large variations of the strong phase across the phase space occur. This means that the interference term is maximally diluted, leading to a limited sensitivity to γ if measured by averaging the strong phase over the phase space as for the four-body final states. The goal is instead to perform the analysis in these three-body modes by analysing regions of the phase space in order to exploit high-coherence regions.

At a given coordinate in the D meson phase space, the density of the B^- decay can be obtained by taking the square of the amplitude and is given by:

$$\begin{aligned} P_B &= |A|^2 + r_B^2|\overline{A}|^2 + r_B(A^*\overline{A}e^{i(\delta_B-\gamma)} + A\overline{A}^*e^{-i(\delta_B-\gamma)}) \\ &= |A|^2 + r_B^2|\overline{A}|^2 + 2\kappa r_B|\overline{A}||A|(\cos\Delta\delta_D\cos(\delta_B-\gamma) - \sin\Delta\delta_D\sin(\delta_B-\gamma)) \\ &= |A|^2 + r_B^2|\overline{A}|^2 + 2\kappa\sqrt{|A|^2|\overline{A}|^2}(C\,x_- - S\,y_-). \end{aligned} \tag{2.41}$$

The amplitudes A (with complex conjugate A^*) and $\overline{A}$ correspond to the amplitudes of the D^0 and $\overline{D}^0$ decays respectively, and $\Delta\delta_D$ is the corresponding strong-phase difference. In this expression, $C \equiv \cos\Delta\delta_D$, $S \equiv \sin\Delta\delta_D$, and $x_- = r_B\cos(\delta_B-\gamma)$, $y_- = r_B\sin(\delta_B-\gamma)$. For the B^+ charge-conjugate decay, a similar expression is obtained by replacing $\gamma \leftrightarrow -\gamma$, $A \leftrightarrow \overline{A}$ and $x_- \leftrightarrow x_+$, $y_- \leftrightarrow y_+$, where $x_+ = r_B\cos(\delta_B+\gamma)$, $y_+ = r_B\sin(\delta_B+\gamma)$.

The kinematics of these three-body $D \to K^0_S h^+h^-$ decays can be fully represented in two dimensions in a Dalitz plot [47], whose coordinates are the square of the invariant masses defined as $s_+ \equiv m^2(K^0_S h^+)$ and $s_- \equiv m^2(K^0_S h^-)$. Figure 2.7 illustrates a Dalitz plot for $D \to K^0_S\pi^+\pi^-$ from Belle collaboration [48], where it can be seen that the Dalitz plot contains information on the different resonances involved in the final state. In particular, the $K^*(892)^- \to K^0_S\pi^-$ resonance can be observed as a vertical distribution at around $s_- \approx 0.8\,\mathrm{GeV}^2/c^4$.

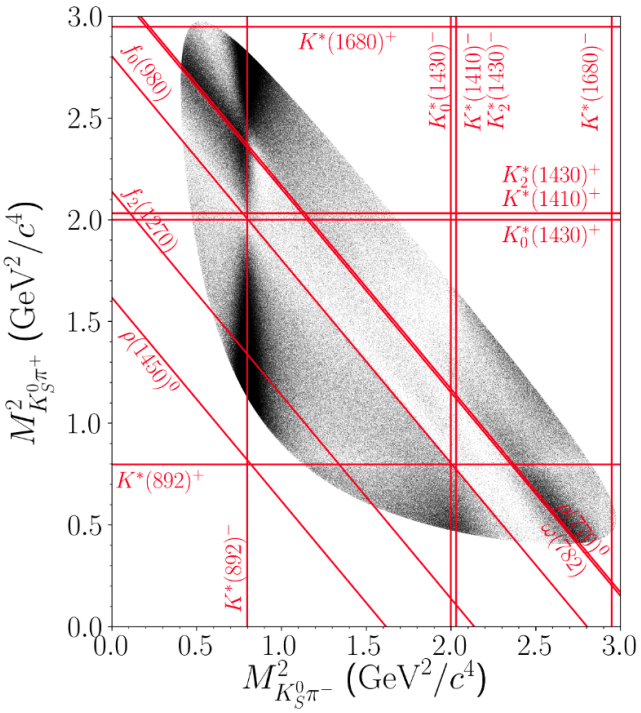


Figure 2.7: Example of a Dalitz plot for $D \to K^0_S\pi^+\pi^-$ from Belle collaboration [48].

The interference effects appear as different distributions of the D meson Dalitz plot for the B^- and B^+ decays. In order to observe these differences, the phase space is binned, essentially reducing the measurement to a counting experiment in each of these Dalitz bins. The binning of the Dalitz plots is chosen in a way that bin pairs are formed such that bins i and $-i$ are CP conjugate and symmetric with respect to the $s_- = s_+$ axis, with positive bins corresponding to bins where $s_- > s_+$. By integrating the density in Equation (2.41), the partial decay rate can be obtained for B^- and B^+ in each bin i:

$$\Gamma_i^- \propto [F_i + (x_-^2 + y_-^2)F_{-i} + 2\kappa\sqrt{F_i F_{-i}}(c_i x_- + s_i y_-)], \tag{2.42}$$

$$\Gamma_i^+ \propto [F_{-i} + (x_+^2 + y_+^2)F_i + 2\kappa\sqrt{F_i F_{-i}}(c_i x_+ - s_i y_+)], \tag{2.43}$$

where the c_i, s_i parameters correspond to the amplitude-weighted averages of the strong-phase differences over bin i, and the F_i coefficients represent the fractional yields of flavour-tagged D^0 decays in the corresponding bin. In other terms, this gives the number of observed D^0 meson decays in bin i including experimental effects such as reconstruction and selection efficiency. The F_i coefficients are taken

from the LHCb measurement in $B^{\pm} \to D(\to K^0_S h^{\pm} h^{\mp}) h^{\pm}$ decays [49], assuming that the relative efficiency between the two analyses is the same across the Dalitz plot. This assumption is verified using simulated samples of $B^{\pm} \to DK^{*\pm}$ and $B^{\pm} \to Dh^{\pm}$ decays.

The c_i, s_i parameters have been measured in bins by CLEO [50] and BESIII [51, 52]. The number of bins chosen was that which was found to be optimal by the data available at these charm factories. The number of bins are 16 and 4 for $D \to K^0_S \pi^+ \pi^-$ and $D \to K^0_S K^+ K^-$ respectively. The binning scheme was optimised such that the expected statistical sensitivity to the *CP* observables $x_{\pm}$, $y_{\pm}$ is maximised [50]. Figure 2.8 shows the optimal binning schemes used for $K^0_S \pi^+ \pi^-$ (left) and $K^0_S K^+ K^-$ (right).

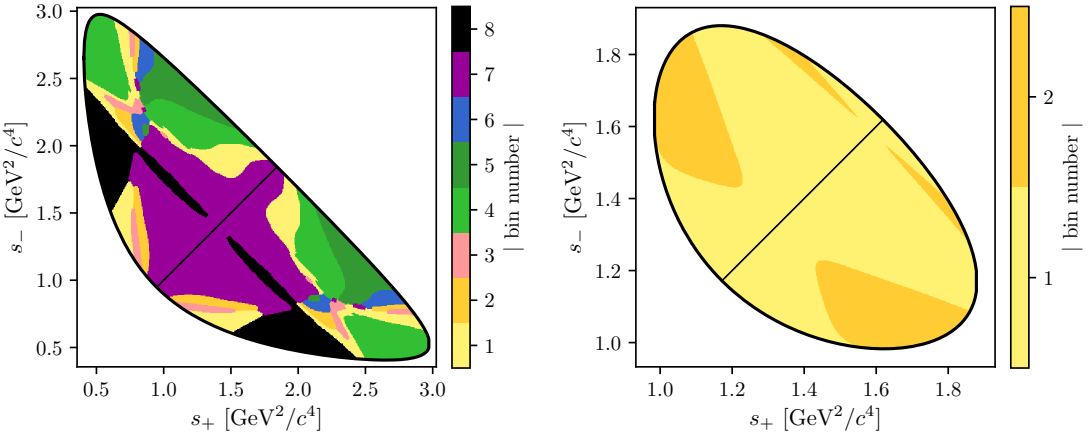


Figure 2.8: The optimal Dalitz plot binning schemes for $D \to K^0_S \pi^+ \pi^-$ (left) and $D \to K^0_S K^+ K^-$ (right) decays [50].

These binning schemes are therefore also used in this analysis in order to be able to use external measurements of the coefficients c_i, s_i. By using these external inputs in the analysis, there is no need to rely on a model of the D meson decay in the phase space [53]. Therefore, the analysis can be performed as a model-independent measurement with no associated systematic uncertainties based on the choice of the model, which are more difficult to estimate. The Equations (2.42)–(2.43) relating the rate in each bin to the *CP* parameters can be expressed in terms of yields in

each bin, introducing normalisation factors $h_B^{\pm}$ for each charge:

$$N_i^- = h_B^-[F_i + (x_-^2 + y_-^2)F_{-i} + 2\kappa\sqrt{F_i F_{-i}}(c_i x_- + s_i y_-)], \tag{2.44}$$

$$N_i^+ = h_B^+[F_{-i} + (x_+^2 + y_+^2)F_i + 2\kappa\sqrt{F_i F_{-i}}(c_i x_+ - s_i y_+)]. \tag{2.45}$$

In the system of Equations (2.44)–(2.45), there are a sufficient number of constraints in order to be able to extract the *CP* observables and the normalisation factors $h_B^{\pm}$ from the measured yields in each bin. The measured *CP* observables are

$$x_{\pm} = r_B \cos(\delta_B \pm \gamma), \tag{2.46}$$

$$y_{\pm} = r_B \sin(\delta_B \pm \gamma). \tag{2.47}$$

2.7 Analysis strategy and overview

The overall LHCb strategy is to cover and combine as many B and D meson decays to improve the sensitivity in the era of high precision measurement. The measurements in different channels add further constraints on γ, as well as a check of consistency due to the fact that different backgrounds and systematic uncertainties are involved in the different decays.

The goal of the analysis presented in this thesis is to determine the value of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays, using ADS/GLW and BPGGSZ modes, where the favoured mode is used as a normalisation and control channel. The *CP* observables, which are directly related to the physics parameters of interest, are measured for both the ADS/GLW and BPGGSZ modes. The measured *CP* observables are defined in Equations (2.18)–(2.23) (GLW, quasi-GLW), Equations (2.29), (2.34)–(2.35) (two-body ADS), Equations (2.38)–(2.40) (four-body ADS) and Equations (2.46)–(2.47) (BPGGSZ).

Measure what is measurable,
and make measurable what is not so.

— Galileo Galilei

3 The LHCb experiment

The LHCb experiment is presented in this chapter. The CERN accelerator complex is introduced and the experimental conditions are detailed. The LHCb detector is described in more details, especially focusing on parts and sub-detectors which are most relevant for the analysis presented in this thesis.

3.1 The Large Hadron Collider at CERN

Located between Switzerland and France at the European Organisation for Nuclear Research (CERN), the Large Hadron Collider (LHC) is the largest particle accelerator in the world [54]. Hosting four experiments and sitting 100 metres underground, the 27-kilometre circumference collider accelerates protons close to the speed of light. Figure 3.1 shows an illustration of the LHC.

The acceleration of the protons is performed in stages in different parts of the accelerator complex. By applying an electric field, protons are extracted from a bottle of hydrogen gas and their energy is increased up to 50 MeV by a linear accelerator (LINAC 2). In the next step, the protons enter the Proton Synchrotron (PS) and further increase their energy to about 25 GeV. Before being injected into the LHC, the protons reach 450 GeV in the Super Proton Synchrotron (SPS). These

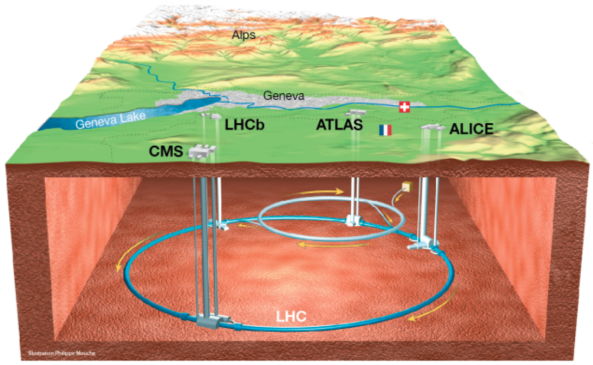


Figure 3.1: Illustration of the Large Hadron Collider, located 100 metres underground between Switzerland and France [55].

different accelerators can be seen in Figure 3.2, showing the different accelerators and detectors of the CERN accelerator complex.

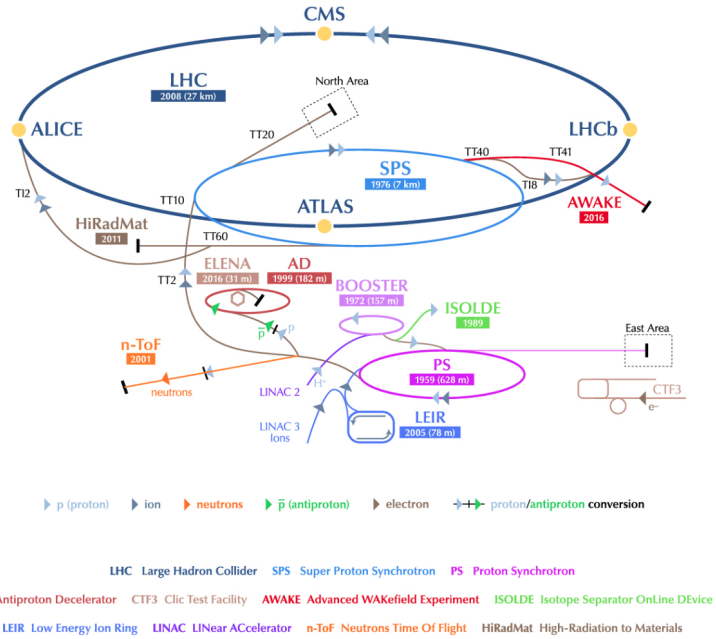


Figure 3.2: Illustration of the CERN accelerator complex [56].

In the LHC, two different beam pipes are used where protons are injected in opposite directions, before making the two proton beams collide at specific interaction points within the four LHC experiments, which are ALICE, ATLAS, CMS and LHCb experiments.

3.2 Running conditions and integrated luminosity in Run 1–2

The two proton beams are made of 2808 bunches, where each bunch contains 1.15×10^{11} protons. The bunch-crossing frequency is 40 MHz, corresponding to a separation of 25 ns between the bunches. During Runs 1–2, the proton-proton collisions took place at a total centre-of-mass energy of $\sqrt{s} = 7, 8$ TeV in 2011 and 2012 respectively, and $\sqrt{s} = 13$ TeV in 2015–2018.

To measure the rate of pp collisions recorded and quantify the performance of the accelerator, the luminosity ($\mathcal{L}$) is used, which relates the event rate dN/dt to the cross-section σ:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma \,. \tag{3.1}$$

Figure 3.3 shows the integrated luminosity recorded for each year by the LHCb detector during Runs 1–2, with a total corresponding to approximately 9 fb^{-1} [57].

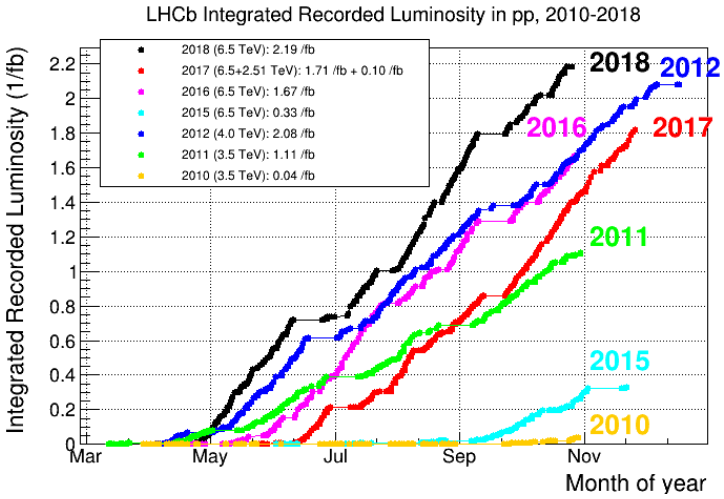


Figure 3.3: Integrated luminosity recorded at LHCb for each year of Run 1–2 [57].

One feature of the LHCb experiment [58, 59] is that the instantaneous luminosity is maintained at a constant value of $\mathcal{L} = 4 \cdot 10^{32}\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, which is lower than the other LHC experiments. This is achieved by luminosity levelling, which consists in adjusting the beam separation as the luminosity decreases during each fill of data taking. This allows data-taking conditions to be kept constant, while reducing the

number of pp collisions in each bunch crossing and therefore the event multiplicity. There are on average 1.1–1.6 pp collisions at the LHCb interaction point, which corresponds to the number of primary vertices per visible event. This is an advantage in terms of reduced radiation damage on the detector as well as simpler data analysis when reconstructing tracks from cleaner events. Indeed, levelling the luminosity helps in maintaining a good resolution when measuring displaced vertices, which is a key signature in beauty and charm decays studied at LHCb. This comes with the price of obtaining datasets of smaller integrated luminosity.

3.3 The LHCb detector

Unlike the other three LHC detectors which are shaped as cylinders and cover the full solid angle, the LHCb detector [58, 59] is a single-arm forward spectrometer which is designed and built to study the physics of b and c quarks. Figure 3.4 presents a layout of the detector.

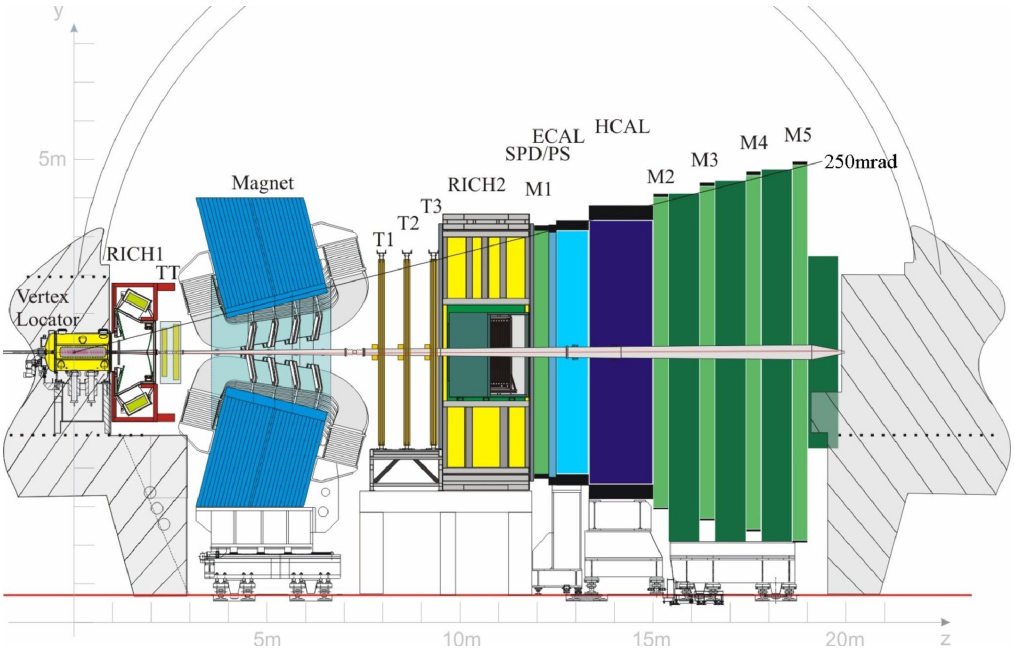


Figure 3.4: The LHCb detector layout [58], presenting a side-view of the detector and showing the different sub-detectors.

The pseudorapidity range covered is $2 < \eta < 5$, where the pseudorapidity is defined as $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$. The angle θ corresponds to the angle between the axis of the

proton beam and the direction of the particle momentum of interest. The acceptance of the detector in the x and y directions covers 15–300 and 15–250 mrad, respectively. With this specific configuration and geometry of the detector, the production cross-section of $b\bar{b}$ pairs peaks within the detector acceptance, as illustrated in Figure 3.5.

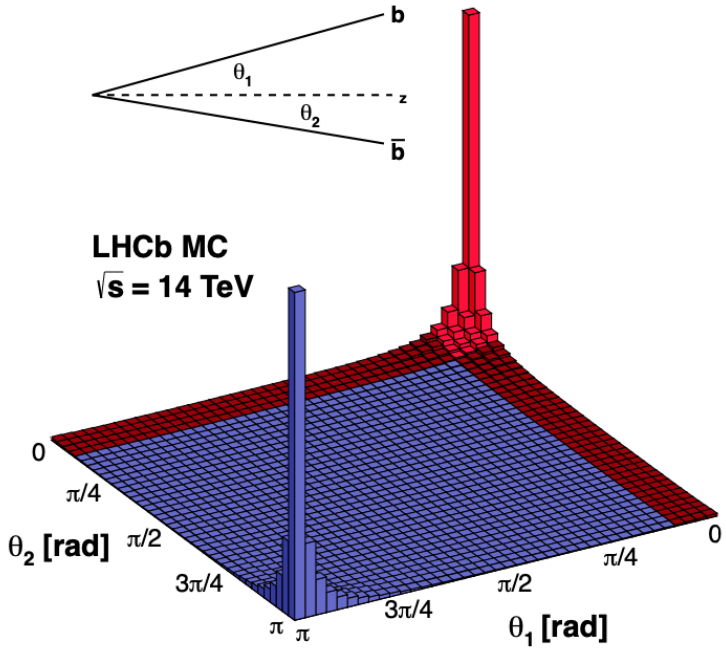


Figure 3.5: Production cross-section of $b\bar{b}$ pairs as a function of their angle with respect to the beam axis [60]. The acceptance of the LHCb detector is shown in red.

A series of sub-detectors, each designed for specific tasks, form the LHCb detector and are aligned in layers around the interaction point. The three main systems forming the LHCb detector are the tracking system, the particle identification system and the trigger system, which are described in the following sections.

3.3.1 Tracking system

The task of the tracking system is to measure the momentum of charged particles traversing the detector. This is done by measuring and reconstructing tracks through the different parts of the tracking system which are described below.

3.3.1.1 Vertex Locator (VELO)

The first sub-detector of the tracking system is the Vertex Locator (VELO) [61, 62], which is placed closest to the interaction point of the *pp* collision. It allows primary vertices to be measured with high precision, as well as secondary vertices which are displaced from the primary vertices. The VELO consists of a silicon detector, made of two halves of silicon strip modules, where each half contains 21 modules. Each module is made of two 300 µm half-disc sensors. A top view of the VELO modules is presented in Figure 3.6.

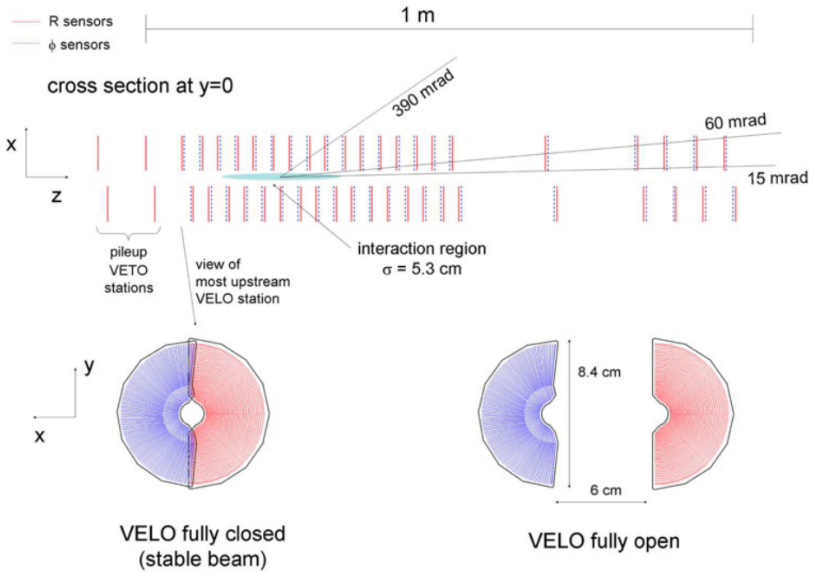


Figure 3.6: Top view of the VELO modules [58].

The two parts can be moved mechanically and are retracted by up to 29 mm during injection phase in order to avoid radiation damage. The two halves are then joined around the interaction point in stable-beam conditions, with a distance of about 8 mm from the beam to the innermost active region of the VELO. The distance r from the center and the azimuthal angle ϕ of the vertex are provided by the silicon sensors. The resolution of the impact parameter (IP) of tracks is $\sigma_{\text{IP}} = 15 + \frac{29}{p_T}$ µm. The primary vertices are measured with a typical resolution of 10 µm and 50 µm in the x,y directions and z direction respectively. Measuring secondary vertices with a good resolution is particularly important for the performed

analysis in order to reduce combinatorial background, given the fact that the B and D mesons are long-lived particles.

3.3.1.2 Dipole magnet

A dipole magnet [63], which provides a magnetic field B of about 4 Tm, is installed in between the tracking stations in order to bend the trajectory of the charged particles as they traverse the detector.

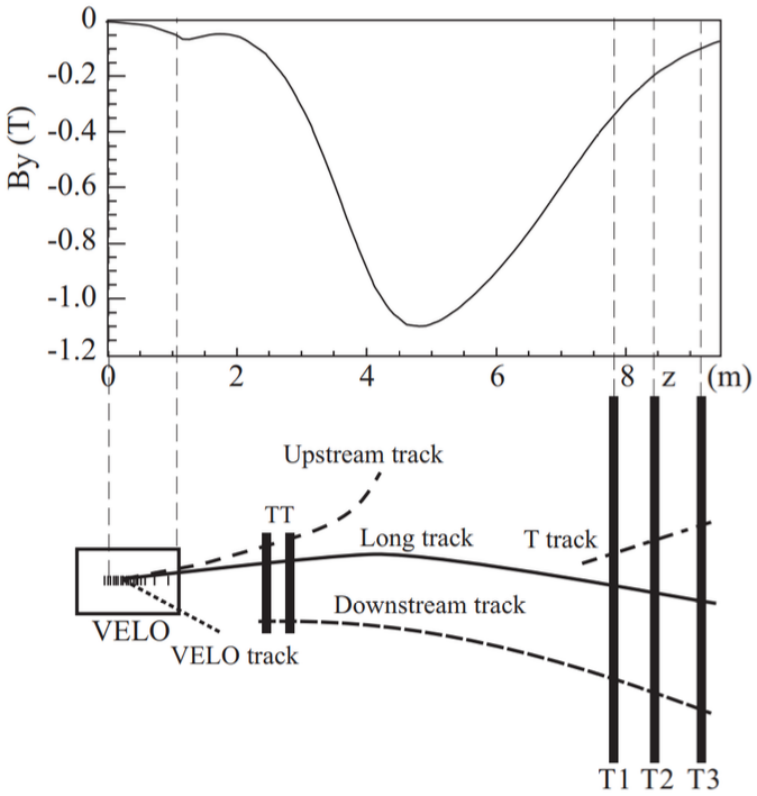


Figure 3.7: Magnetic field provided by the dipole magnet which is installed between tracking stations [59].

The curvature of a track R is used, together with its elementary electric charge q, to determine the momentum of the particle p through the equation $p = qRB$. An illustration of the magnetic field and bent tracks from the VELO to the tracking stations is shown in Figure 3.7. The magnetic field from the dipole magnet is vertical, therefore the tracks are bent in the plane defined by the x and z axes. Any differences between the two halves of the detector could lead to systematic effects

such as detection asymmetries between the two charges, which are relevant for the analysis presented in this thesis. To reduce and cancel this effect, the magnet polarity is changed between *Up* and *Down* in order to collect approximately equal amounts of data in both directions of the magnetic field.

3.3.1.3 Tracking stations

The last part of the tracking system consists of four tracking stations, placed before (upstream) and after (downstream) the dipole magnet. The four tracking stations are the Tracker Turicensis, or Trigger Tracker (TT), placed upstream of the magnet, as well as tracking stations T1-T3 downstream of the dipole magnet. Each of the four tracking stations is made of four layers, which are arranged in a X–U–V–X geometry. The two central layers are tilted with respect to the first and last vertical planes, with an angle of ±5°. Two layers of the TT are illustrated in Figure 3.8, where a vertical (left) and tilted (right) planes are shown. Arranging the planes in this geometry allows measurements of the y and z coordinates of the particles, which reduces ghost tracks coming from multiple-hit ambiguities.

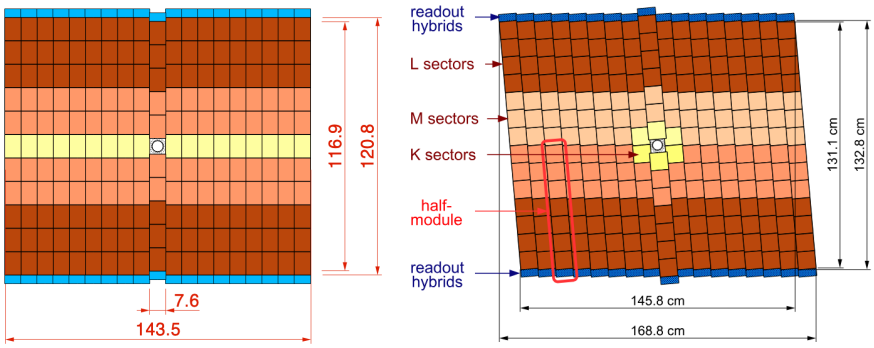


Figure 3.8: Layers of a vertical plane [64] (left) and a tilted plane [58] (right) of the TT.

The Silicon Tracker (ST) is made of the Trigger Tracker (TT) and the Inner Tracker (IT) [64], where the IT is the inner part of the stations T1-T3, where silicon microstrip detectors with pitches of 200 µm are used. The Outer Tracker (OT) [65, 66], which refers to the outer part of the three tracking stations T1-T3 which is not in the IT, is a drift time detector. It is composed of straw tubes which are filled with a gas mixture of $Ar/CO_2/O_2$ (70%/28.5%/1.5%). Figure 3.9 shows a

module of the OT (a) and the geometry arrangement of the OT which is the same as for the IT (b). There are 18 separate modules in each plane of the OT, while each module is composed of 128 straw tubes which are arranged in two staggered monolayers. In the straw tubes, the gas atoms are ionised by the particles going through them and the anode wire collects the electrons with a drift time of about 35 to 50 ns. The single-hit spatial resolution achieved in Run 2 is 171 µm, which was improved from 205 µm in Run 1 by performing the space-time alignment of the OT online, during data taking [66].

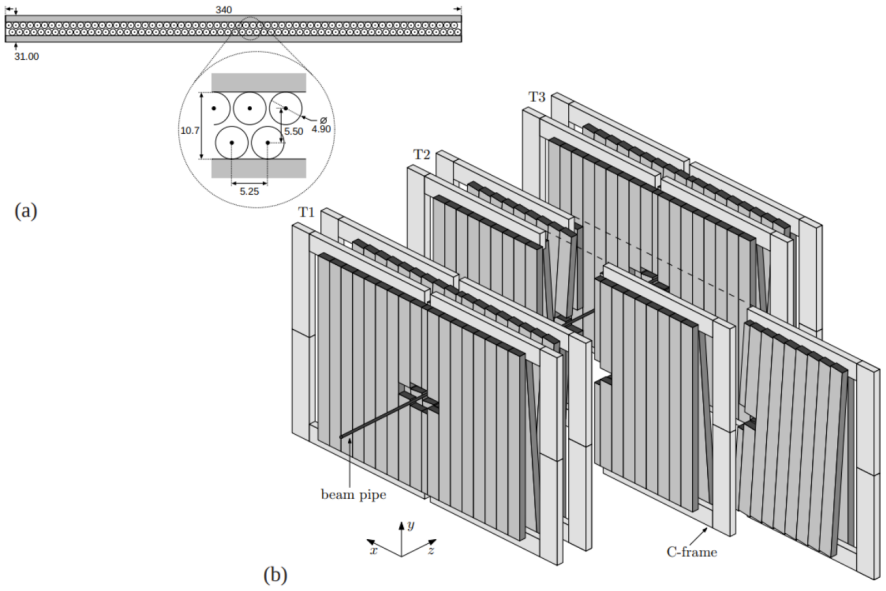


Figure 3.9: Illustration of a module of the OT (a) and the modules and their arrangement in the X–U–V–X geometry (b) [65].

3.3.1.4 Track reconstruction

Measuring the momentum of particles with good resolution plays an important role in the analysis as it leads to narrow signal peaks which are well separated from partially reconstructed backgrounds located at a lower invariant mass. As shown in Figure 3.10, the relative uncertainty obtained from the tracking system on the momentum of charged particles varies between 0.5% and 1.0% based on the momentum.

Different track types are defined based on the information used for their reconstruction, which are shown in Figure 3.7. The two types of interest for

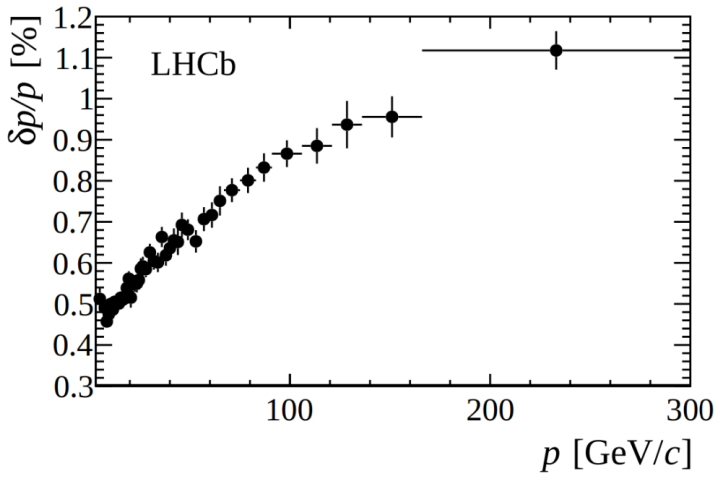


Figure 3.10: Relative uncertainty on the momentum measurement of charged long tracks in $J/\psi \to \mu^+\mu^-$ decays [59].

the analysis presented in this thesis are Long and Downstream tracks. The former refers to tracks which are reconstructed with hits recorded within the VELO, while the latter represents tracks which only make use of track hits within the tracking stations. In the analysis, this is relevant for K^0_S particles which are reconstructed in their decay to two charged pions $\pi^+\pi^-$.

The track reconstruction for Long tracks is done by extrapolating VELO tracks to match them with hits in the tracking stations, using the *forward tracking* [67, 68] and *track matching* [69, 70] algorithms. Concerning Downstream tracks, these are reconstructed by extrapolating T-tracks (tracks in the T1-T3 stations) to match them with compatible TT hits [71]. A Kalman filter [72] is then used to reprocess all tracks which are reconstructed. This step helps to take into account multiple scattering and apply energy loss corrections [73].

3.3.2 Particle identification system

Being able to distinguish pions from kaons is crucial in flavour physics and in particular in the measurement presented in this thesis, where different D decay modes are considered with final states involving kaons and pions. The particle identification (PID) system, which includes Ring Imaging Cherenkov (RICH) detectors providing information on the types of particles going through the detector, makes it possible

to separate kaons from pions and remove backgrounds coming from decays with misidentified particles. The rest of the particle identification system makes use of calorimeters and muon stations.

3.3.2.1 Ring Imaging Cherenkov detectors (RICH)

Ring Imaging Cherenkov (RICH) detectors [74–76] are designed to identify charged particles of known momentum, making use of the Cherenkov effect. This corresponds to electromagnetic radiation which is emitted by the medium when a charged particle goes through a dielectric medium faster than the speed of light in that specific medium, analogously to the sonic boom. The light from the Cherenkov radiation, which is emitted in a cone of angle θ_c, is directly related to the particle velocity $\beta = v/c$ and the refraction index n of the medium:

$$\cos\theta_c = \frac{1}{n \cdot \beta} \,. \tag{3.2}$$

Therefore, the velocity β can be determined by measuring the Cherenkov angle θ_c. In turn, the velocity is used to obtain the mass of the particle which identifies the particle of interest. This is achieved by using the relation $p = \gamma m v$, where $\gamma = (1 - \beta^2)^{-1/2}$ is the Lorentz factor, together with the measurement of the momentum p provided by the tracking system.

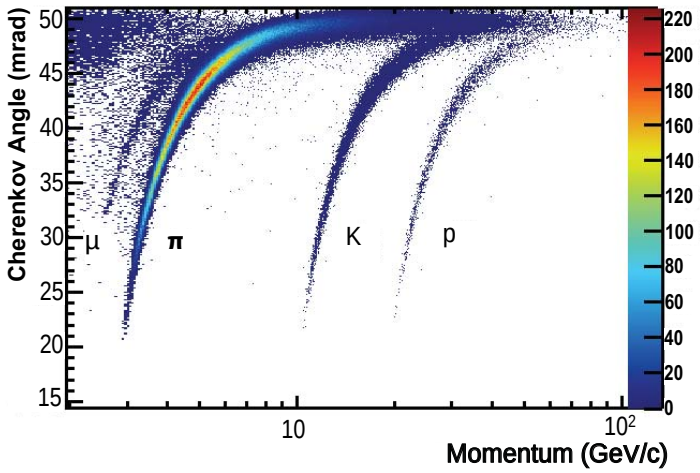


Figure 3.11: Cherenkov angle as a function of momentum for different types of particles [75].

The Cherenkov angle as a function of the momentum for different types of particles is shown in Figure 3.11, where it can be seen that RICH detectors provide good PID to distinguish kaons from pions as well as other particles such as protons or muons.

Two RICH detectors are used in order to cover and improve the performance over a wider momentum range. Upstream of the dipole magnet and using a C_4F_{10} radiator, RICH1 is more sensitive for lower momentum particles in the range 1–60 GeV/c. To cover higher momentum values in the range 15–100 GeV/c, RICH2 is placed downstream of the magnet and uses CF_4 as radiator. Figure 3.12 presents an overview of the RICH1 (left) and RICH2 (right) detectors.

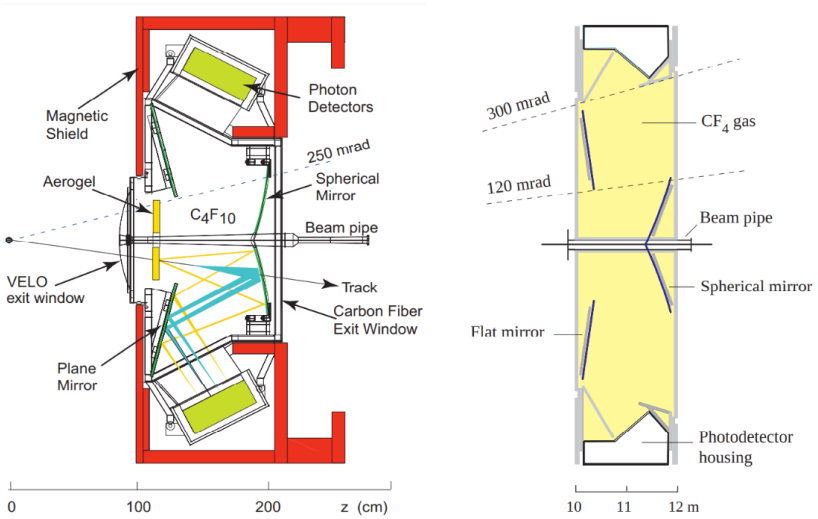


Figure 3.12: Schematic representation of the RICH1 (left) and RICH2 detectors [58, 74].

For RICH1, an aerogel radiator was also included in Run 1 in order to provide better sensitivity at even lower momentum not accessible with the C_4F_{10} radiator. Due to poor performance and not meeting requirements, this was removed in Run 2 [76]. Mirrors are used in both RICH detectors which reflect photons from the Cherenkov radiation to Hybrid Photon Detectors (HPD). Arrays of HPD detectors are used where rings are formed from the photons generated from the particle tracks, such that the radius of the ring is given by the Cherenkov angle.

3.3.2.2 Calorimeters

The calorimeter system [77, 78] provides PID for electrons, photons and hadrons, and is used to trigger events which helps in rejecting backgrounds. The calorimeter system is placed downstream of RICH2 and is composed of four sub-detectors: a Scintillating Pad Detector (SPD), a Pre-Shower Detector (PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). The calorimeter system relies on scintillators coupled to photomultiplier tubes (PMTs).

The energy and position of particles are measured using the ECAL and HCAL. The calorimeters stop the particles traversing the detector, where they deposit their energy. The calorimeters are made of metal layers and plastic scintillator plates. The metal used for the ECAL is lead, while iron is used for the HCAL. Secondary particles are produced in the metal when the incoming particle interacts with the material. In turn, photons are emitted by excitation when the secondary particles create interactions in the scintillator. This produced light is proportional to the energy which is lost by the initial particle and is collected by wavelength-shifting fibres and transported to the PMTs.

The SPD and PS, which are placed upstream of the two calorimeters, are used to trigger events and reject background. The SPD, which is made of rectangular scintillating pads, makes it possible to distinguish electrons from photons, as only charged particles interact in the SPD. A lead absorber, which is 15 mm thick, is placed in between the two layers of scintillating pads of the SPD and PS. This helps to separate electrons from pions using the PS, since electrons deposit energy in the interaction with the lead layer.

3.3.2.3 Muon stations

The muon system [79, 80] consists in five muon stations M1–M5. Since the muons traverse the whole detector, the muon system is placed downstream of all the other sub-detectors, except from the M1 station which is placed upstream of the calorimeter system. This station is used for an independent measurement of muon transverse momentum which is used in the trigger.

The innermost region of the first muon station M1 is a gas electron multiplier (GEM) detector, while stations M2–M5 are Multi Wire Proportional Chambers (MWPC) which are filled with three different gases $Ar/CF_4/CO_2$ (45%/40%/15%). The GEM is used in the first station due to the larger amount of tracks present in the central region before the calorimeter system. The M2–M5 stations are separated by iron absorbers which are 80 cm thick, which favours the selection of high-momentum muons.

3.3.2.4 Particle identification

In order to identify particles, the information from the different parts of the particle identification system is combined. Of particular importance in this thesis is the ability to separate hadrons and especially pions from kaons. This is mainly achieved from the information provided by the RICH detectors, which is compared with expected patterns for specific particles [81]. In this way, differences between the hypotheses that the hits correspond to those produced by two specific particles are compared, using likelihood functions $\mathcal{L}$. The pion hypothesis is chosen as the reference hypothesis, to which other hypotheses are compared by looking at the corresponding likelihood functions. For kaons, the variable of interest is given by

$$\Delta \log \mathcal{L}(K-\pi) = \log \mathcal{L}(K) - \log \mathcal{L}(\pi)\,, \tag{3.3}$$

which represents the likelihood of a given track being a kaon relative to the hypothesis of the track being a pion. This $\Delta \log \mathcal{L}(K-\pi)$ variable, which is referred to as PIDK in the analysis, is used to apply PID requirements and select kaons. Figure 3.13 presents the efficiency to correctly identify a kaon for two different PID requirements, as well as the corresponding misidentification of a pion as a kaon determined in 2017 data. The method used to obtain the PID efficiency is discussed in more detail in Chapter 8. It can be seen that a good separation for kaons and pions is achieved, where the rate of correctly identified kaons is large while maintaining a low rate of misidentified pions.

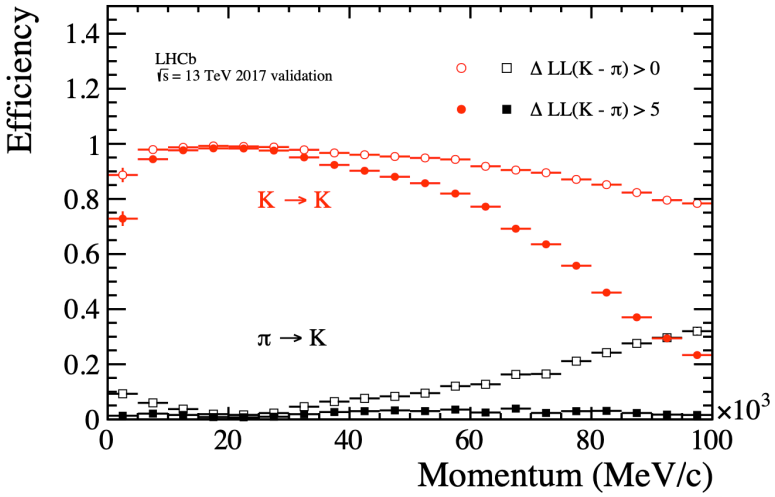


Figure 3.13: Efficiency of kaon identification and pion misidentification as a function of momentum determined in 2017 data [82].

3.3.3 Trigger system

As explained in Section 3.3, $b\bar{b}$ quark pairs are produced with a high cross section at LHCb, which is also the case for $c\bar{c}$ quark pairs. These are the events of interest at LHCb for which large input rates are obtained due to the large corresponding cross sections. This can be seen in Figure 3.14, which presents different event rates and production cross sections at LHC energies. This is the reason why processing data is more challenging at LHCb where the events of interest occur at the MHz level, in comparison to ATLAS or CMS for example. These general purpose LHC experiments deal with events of interest at or below the kHz level, which makes it easier to select the data and maintain an output rate of the order of the kHz without discarding the events of interest. Since storing all data is not feasible in terms of storage costs for this large amount of signal to be processed, the trigger system [83–85] is designed to reject background and select interesting events for further processing. The 40 MHz bunch crossing rate is reduced in stages corresponding to three different levels of the trigger, which are a first level hardware trigger and two high level software triggers. An overview of the trigger for Run 2 is presented in Figure 3.15.

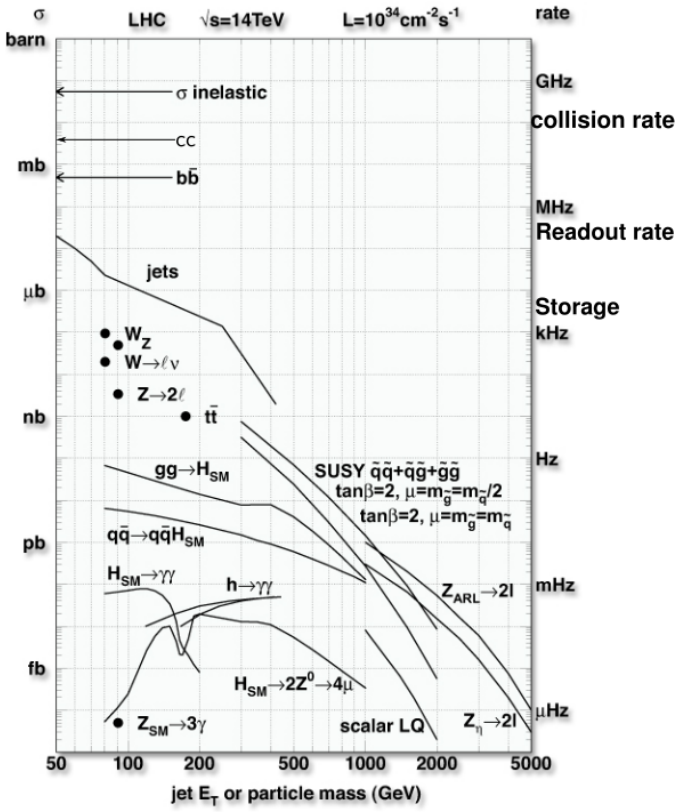


Figure 3.14: Event rates and production cross sections at LHC energies [86].

3.3.3.1 L0 hardware trigger

The first level of the trigger is a hardware trigger, called the Level-0 (L0) hardware trigger, where the rate is reduced to the level of 1 MHz by selecting high momentum and high transverse energy events by combining information from the calorimeter system and the muon system.

3.3.3.2 High Level software trigger

The High Level Trigger (HLT) is a software trigger composed of two levels, HLT1 and HLT2. A partial event reconstruction is performed at the HLT1 level by placing requirements on momentum and the quality of reconstructed tracks. This inclusive selection further reduces the rate to the order of 100 kHz. These selections are based on the position of the PVs for long tracks which are selected over a p_T threshold. The HLT1 therefore selects events with at least one track with high

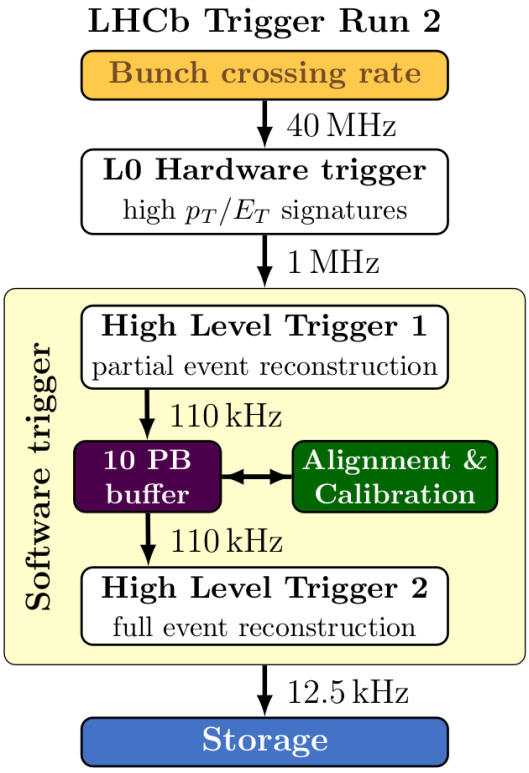


Figure 3.15: Overview of the trigger in Run 2 [85].

transverse momentum and which is significantly displaced from the PV, which is a signature of B and D decays in this analysis.

The full event reconstruction is performed in the final step which is the HLT2 by combining information from the full detector, where both inclusive and exclusive selections are applied, reducing the rate to about 12.5 kHz which is stored on disk. A disk buffer of around 10 PB is used between the two HLT stages for alignment and calibration. Using this disk buffer gives time for events to be calibrated based on detector conditions during data-taking, such as alignment of the tracking system and RICH detectors, before moving to HLT2. During Run 1, the events were directly moved to HLT2 before applying the calibration offline. This change allowed to improve the trigger efficiency in Run 2. In the analysis presented in this thesis, multivariate algorithms [87] are used in HLT2 to keep events for which secondary vertices are consistent with two-, three- or four-body b-hadron decays.

3.3.4 Offline processing and stripping

The output of the HLT2, which is saved on disk and contains the calibration and alignment of the full detector, is used for analysis. The data is further processed offline with a selection developed specifically for a given decay chain, called *stripping*. This step is executed on the Worldwide LHC Computing Grid (WLCG) [88] as it requires significant time and processing power. This makes use of computing centres in many countries around the world in order to run the code in parallel by splitting the data in smaller datasets. The selection for the decay of interest in the measurement presented in this thesis is described in Chapter 4.

3.3.5 Simulation

Simulation of the signal decays plays an important role in many LHCb analyses. In the measurement presented here, simulation samples are used to model the invariant-mass distribution of decays involved in the analysis as well as estimating selection efficiencies. The first step of the simulation consists in simulating pp collisions, which are generated using Pythia [89] for a given LHCb configuration [90]. For unstable particles in the decay chain, their evolution and decay are described by EvtGen [91], where final-state radiation is generated using Photos [92]. Finally, using the Geant4 toolkit [93, 94], the response of the detector to the interaction of particles in the detector is implemented as described in Reference [95].

The last step of the simulation, which is the full detector simulation, requires significant computing power, which limits the amount of events that can be generated. This issue can be mitigated by reusing the detector simulation multiple times for non-signal particle tracks using ReDecay [96]. This allows only the signal events to be generated while keeping the detector simulation of the rest of the event many times. Moreover, fast simulation is also used in the analysis for studying some of the background decays. Using the RapidSim package [97], events are generated using EvtGen with kinematics corresponding to b and c quark hadrons decaying in LHCb. The LHCb detector resolution effects are modelled by smearing the momentum of particles.

An experiment is a question which science poses to Nature and a measurement is the recording of Nature's answer.

— Max Planck

4

Measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays

Data, candidate selection and efficiencies

This chapter describes the first part of the main work of this thesis, which is a model-independent measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays at LHCb. The full Run 1 and Run 2 dataset, collected between 2011 and 2018, is used for the measurement, which supersedes the ADS/GLW results published with 2011–2016 data [6]. This comprehensive study of this decay channel uses ADS/GLW and BPGGSZ D decay final states, where the BPGGSZ modes are analysed for the first time at LHCb. This first part details the candidate selection and the selection efficiencies.

4.1 Data and simulation samples

The analysis makes use of the LHCb dataset collected during Run 1 (2011–2012) and Run 2 (2015–2018), corresponding to $9\,\text{fb}^{-1}$ of integrated luminosity. The details for each year concerning the integrated luminosity and the centre-of-mass energy are given in Table 4.1. As described in Section 3.3.4, the candidates are first selected from the stripping. Specific stripping lines are used for the different D decay modes. For the ADS/GLW analysis, the candidates are selected by two

different lines for the two-body and four-body D final states. Similarly, a dedicated line is used to select the BPGGSZ modes. For each of the three lines, the K^0_S candidates are reconstructed separately for the two track types. The track is said to be of type Long (LL) if hits from the tracks are recorded inside the VELO, and the track is defined as Downstream (DD) if the reconstruction is performed only using hits within the tracking stations.

Table 4.1: Integrated luminosity and centre-of-mass energy for each year of data taking used in the analysis.

Year	$\mathcal{L}$ [fb^{-1}]	$\sqrt{s}$ [TeV]
2011	1.1	7
2012	2.1	8
2015	0.3	13
2016	1.7	13
2017	1.7	13
2018	2.1	13

In order to model the signal distribution and determine the efficiencies, which are included in the fit to measure the *CP* observables in Section 6.4, simulation samples are used in the analysis. Full LHCb Monte Carlo (MC) samples are generated for the list of signal decays, shown in Table 4.2.

Table 4.2: List of signal MC samples generated in the analysis.

Decay
$B^\pm \to [K^\pm\pi^\mp]_D K^{*\pm}$
$B^\pm \to [K^+K^-]_D K^{*\pm}$
$B^\pm \to [\pi^+\pi^-]_D K^{*\pm}$
$B^\pm \to [K^\pm\pi^\mp\pi^\pm\pi^\mp]_D K^{*\pm}$
$B^\pm \to [\pi^+\pi^-\pi^+\pi^-]_D K^{*\pm}$
$B^\pm \to [K^0_S\pi^+\pi^-]_D K^{*\pm}$
$B^\pm \to [K^0_S K^+K^-]_D K^{*\pm}$

In these simulation samples, the $K^{*\pm}$ resonance is forced to decay as $K^{*\pm} \to K^0_S\pi^\pm$, with $K^0_S \to \pi^+\pi^-$, while the D meson decay is uniformly distributed over the phase space. As discussed in more detail in Section 4.4, other simulation samples are

also generated and used for background studies. These background samples are used to model the invariant-mass distribution or estimate the expected number of candidates of specific backgrounds. The partially reconstructed background is modelled using the corresponding simulated decays. Three different helicity states are generated for the partially reconstructed backgrounds, as explained in Section 4.4.3.2. The full list of background decays for which samples are generated in the analysis is given in Table 4.3.

Table 4.3: List of background MC samples generated in the analysis.

Decay
$B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ (Helicity 001)
$B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ (Helicity 010)
$B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ (Helicity 100)
$B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ (Helicity 001)
$B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ (Helicity 010)
$B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ (Helicity 100)
$B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$ (Helicity 001)
$B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$ (Helicity 010)
$B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$ (Helicity 100)
$B^{\pm} \to DK^0_S\pi^{\pm}$
$B^{\pm} \to DK^0_S K^{\pm}$
$B^{\pm} \to DK^{*\pm}\pi^0$
$B^{\pm} \to K^{*\pm}K^{\pm}\pi^{\mp}$
$B^{\pm} \to K^{*\pm}K^+K^-$
$B^{\pm} \to K^{*\pm}\pi^+\pi^-$
$B^{\pm} \to K^{*\pm}\pi^{\pm}K^{\mp}$
$\Lambda_b^0 \to \Lambda_c^+(\to pK^-\pi^+)K^{*-}$
$B_s^0 \to DK^0_S\pi^+\pi^-$
$B_s^0 \to D_s^-(\to K^+K^-\pi^-)K^{*+}$
$B^0 \to D_s^-(\to K^+K^-\pi^-)K^{*+}$
$B^0 \to D_s^-(\to \pi^+\pi^-\pi^-)K^{*+}$
$B_s^0 \to D_s^{*-}(\to D_s^-(K^+K^-\pi^-)\gamma/\pi^0)K^{*+}$
$B^{\pm} \to D(\to K^0_S\pi^+\pi^-)K^{*\pm}$
$B^{\pm} \to D(\to K^0_S K^+K^-)K^{*\pm}$

4.2 Candidate selection

An important step of the analysis is the selection of the candidates, in order to keep the events of interest and reduce the background level as much as possible. To this aim, the selection of the B meson signal candidates is performed in stages, where the data is filtered further at each stage by removing or reducing backgrounds of different types.

4.2.1 Event reconstruction

The first step of the selection, before applying the selection in stages, consists of the reconstruction of the decay chain by combining the final state particles to reconstruct the decay of interest. The $K^*(892)^\pm$ resonance is reconstructed through the $K^{*\pm} \to K^0_S\pi^\pm$ decay, where the charged pion $\pi^\pm$ is referred to as the bachelor pion. The K^0_S particles are very long-lived particles and can therefore decay in different parts of the detector. These are reconstructed through their decay to two charged pions $\pi^+\pi^-$. The fact that these decays are reconstructed in different regions of the detector leads to different momentum and vertex resolutions. Two categories are defined for the track types of K^0_S particles, LL and DD.

In order to improve the resolution, the B meson is constrained to come from the corresponding PV, and the invariant masses of the D meson and K^0_S particles are constrained to their PDG known values. This is achieved by performing a refit of the decay chain using the `DecayTreeFitter` (`DTF`) tool [98]. This also leads to a reduction in the difference in B mass resolution between the LL and DD K^0_S track types, which allows both invariant-mass distributions to be fit together in the fit, as shown in Section 4.4.2.

4.2.2 Truth matching

For the signal simulation samples, an additional step after the reconstruction is required in the selection, which is referred to as truth matching. This consists in making sure that the simulated events correspond to true signal events. This is ensured by matching each simulated particle with the truth information generated

for each particle in the decay. The `TRUE_ID` of each particle is required to match the corresponding correct PDG ID number or have `TRUE_ID` = 0. The latter condition, `TRUE_ID` = 0, corresponds to cases where the particles are not matched to any MC particles and do therefore not have any associated tracks. These are known as ghost tracks, which could be background events. These can also be real particles not associated correctly due to poorly reconstructed tracks or decaying in flight. This is the reason why these candidates are kept to correctly model the tails of the invariant-mass distribution of the B meson candidates. For decay products, it is also required to match the corresponding mother and grandmother particles to particles with the correct identity.

Moreover, different categories of events are selected for the truth matching of the simulation samples. One category is for signal events and two other categories are also used to select candidates based on the background categories in which they are classified. The first background category contains events where one or more particles are missed in the reconstruction but the identification of all MC particles is correct. For example, these can be real decays with final state radiation (FSR) photons not reconstructed in the decay. In the second background category, one or more particles are identified as ghost tracks. This can happen in decays in flight or due to a poor reconstruction of tracks in the event, as described previously for events with `TRUE_ID` = 0. The events contained in these two background categories are kept in the truth matching of the simulation samples in order to correctly model the tails of the invariant-mass distribution of the B meson candidates.

4.2.3 Trigger requirements

As introduced in Section 3.3.3, the selected candidates are required to satisfy trigger requirements at each trigger level. At the hardware trigger level L0, a high transverse energy deposit in the HCAL (`L0HadronDecision`) is required to select the mode of interest, with a trigger on signal (TOS). This means that the trigger requirements are met by a particle associated with the signal decay. Candidates are also formed for events with a trigger independent of signal (TIS), triggering on a

particle not related to the signal decay in the given event. In this case (`L0Global`), it is required that at least one candidate satisfies the transverse energy requirements to be selected by the L0 trigger. At the HLT level, loose preselection requirements, based on displaced vertices and vertex quality, are used to further reduce the rate. Inclusive TOS trigger lines are used, which contain the list of requirements to be satisfied for events to be selected. The trigger lines used in the analysis are listed in Table 4.4 for the different trigger levels.

Table 4.4: Summary of the trigger requirements applied to the data.

Run	Trigger level	Trigger lines
Run 1	L0	`L0HadronDecision_TOS OR L0Global_TIS`
	HLT1	`Hlt1TrackAllL0Decision_TOS`
	HLT2	`Hlt2Topo{2, 3 OR 4}BodyBBDTDecision_TOS`
Run 2	L0	`L0HadronDecision_TOS OR L0Global_TIS`
	HLT1	`Hlt1{Track OR TwoTrack}MVADecision_TOS`
	HLT2	`Hlt2Topo{2, 3 OR 4}BodyDecision_TOS`

4.2.4 Preselection rectangular cuts

In addition to the stripping and trigger requirements, a set of rectangular preselection cuts are applied in the next stage of the selection. It is required that the `DTF` fit converged and requirements are placed on invariant-mass ranges.

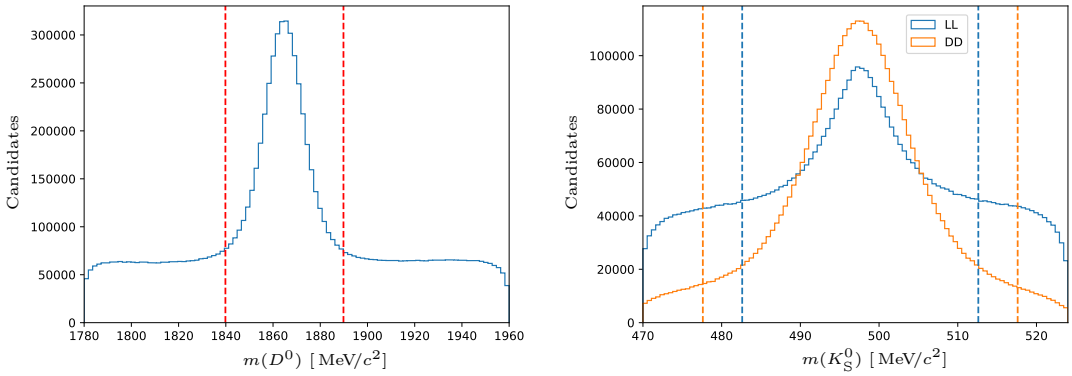


Figure 4.1: Invariant-mass distribution in the $K\pi$ mode of the D (left) and the K^0_S (right) before the preselection cuts. The lines show the selected mass window regions.

The signal events are selected by requiring the mass of the resonances to be around their known values from the PDG [28]. The D (left) and K^0_S (right) invariant-mass distributions are shown in Figure 4.1, as well as the chosen mass window cuts. These mass ranges correspond to approximately three standard deviations (three times the mass resolution). Different windows are chosen for DD and LL candidates in the ADS/GLW modes, due to the difference in mass resolution between the two track types. The bachelor pion and D meson daughters are required to have information from the RICH detector (`hasRich` = `True`) for subsequent PID requirements to be applied with high performance. In order to remove or reduce specific backgrounds, a list of preselection cuts are used and summarised in Table 4.5.

Table 4.5: List of preselection cuts applied in the analysis.

Variable	Cut
All modes	
`DecayTreeFitter` converged	`True`
Bachelor pion momentum (`Bach_P`)	< 150 GeV/c
Bachelor pion `hasRich`	`True`
D daughters `hasRich`	`True`
B invariant mass	[4750, 5800] MeV/c^2
D invariant mass	$m_D \in m_D^{PDG} \pm 25\,\text{MeV}/c^2$
K^0_S invariant mass (DD/LL tracks)	$m_{K^0_S} \in m_{K^0_S}^{PDG} \pm 20/15\,\text{MeV}/c^2$
K^* invariant mass	$m_{K^*} \in m_{K^*}^{PDG} \pm 75\,\text{MeV}/c^2$
Helicity angle $\lvert\cos(\theta^*)\rvert$	> 0.3
$\Delta z^{K_S-B}_{\text{significance}}$ (LL tracks)	> 5
ADS/GLW modes	
$\Delta z^{D-B}_{\text{significance}}$	> 2
Min opening angle θ_{ij}	$> 0.03°$
BPGGSZ modes	
D daughter K^0_S invariant mass	$m_{K^0_S} \in m_{K^0_S}^{PDG} \pm 20\,\text{MeV}/c^2$
$\Delta z^{D-B}_{\text{significance}}$	> 0.5
Flight-distance χ^2 of D daughter K^0_S	> 49

The requirements are either applied to all modes or specifically to the ADS/GLW or BPGGSZ modes, and the reasons behind them are described in this section. For the BPGGSZ modes, the D daughter K^0_S has a flight-distance χ^2 requirement

in order to remove backgrounds from $D \to \pi\pi\pi\pi$ and $D \to \pi\pi KK$ decays in the $D \to K^0_S\pi\pi$ and $D \to K^0_S KK$ modes respectively.

4.2.4.1 K^* selection: non-resonant $B^\pm \to DK^0_S\pi^\pm$ background

The K^* is a wide resonance which decays via strong interaction. In order to remove non-resonant background, that is $B^\pm \to DK^0_S\pi^\pm$ decays, a tight mass requirement on the K^* mass is applied. The invariant mass is required to be within 75 MeV/c^2 around the PDG nominal value [28]. To further suppress this background, a condition is applied on the K^0_S helicity angle θ^*, which is defined as the angle between the vectors of the companion K^0_S and the B momentum in the K^* rest frame. This is due to the fact that in $B^\pm \to DK^{*\pm}$ decays, which are Scalar $\to$ Scalar Vector decays, the K^* vector particle (spin J=1) is longitudinally polarised. The helicity angle is required to satisfy $|\cos\theta^*| > 0.3$. Figure 4.2 shows the $\cos\theta^*$ (normalised) distribution obtained from simulated signal events, as well as for background candidates in the high sideband of the B invariant mass. The candidates in the region between the red vertical lines are rejected in the selection.

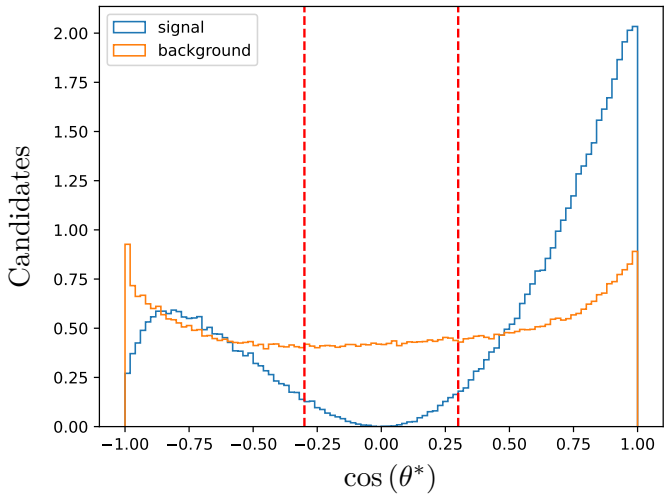


Figure 4.2: Cosine of the helicity angle for simulated signal events and for background candidates taken from the high sideband of the B invariant mass, for the favoured $K\pi$ mode. The two distributions are normalised and the rejected window is shown with the red vertical lines.

The combinatorial background, which is distributed more uniformly as a function of $\cos\theta^*$, is therefore also reduced by this helicity requirement.

4.2.4.2 Background from $B \to D\pi\pi\pi$ decays

The most important background in the analysis comes from the peaking background due to $B^{\pm} \to D\pi^{\pm}\pi^{\mp}\pi^{\pm}$ decays. This background peaks in the same place as the signal and would be larger than the signal without the selection criteria which reduce it, due to the larger branching fraction $(5.5 \pm 2.0) \times 10^{-3}$ of these $B \to D\pi\pi\pi$ decays compared to $(5.3 \pm 0.4) \times 10^{-4}$ for the signal [28]. In order to suppress this background, it is important to ensure that the companion K^0_S particle has a significant flight distance from the B vertex, which is illustrated in Figure 4.3. This is to make sure that the $\pi^+\pi^-$ final state is reached through the intermediate K^0_S particle.

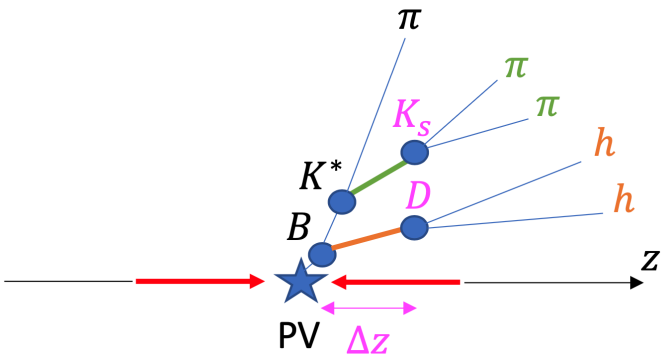


Figure 4.3: Topology of the signal decay, illustrating that the D and K^0_S particles must be well displaced from the B vertex in order to suppress peaking backgrounds coming from decays where the final states are reached without passing by the intermediate particles.

This is achieved by using the flight distance significance of the companion K^0_S, defined as

$$\Delta z^{K_S - B}_{significance} = \frac{z^{K_S}_{vtx} - z^{B}_{vtx}}{\sqrt{\sigma^2(z^{K_S}_{vtx}) + \sigma^2(z^{B}_{vtx})}}, \tag{4.1}$$

where $z^{K_S}_{vtx}$ and z^{B}_{vtx} represent the positions of the decay vertex in the z direction of the K^0_S particle and B meson respectively, and $\sigma(z^{K_S}_{vtx})$ and $\sigma(z^{B}_{vtx})$ are the corresponding errors on these quantities. In order to study this background, the B invariant-mass distribution obtained from the candidates in the K^0_S mass sidebands is studied, where the refit of the decay chain is removed in order to be able to

estimate this background. The invariant mass is shown in Figure 4.4 for LL (left) and DD (right) candidates. For the LL tracks, a tight cut is placed on the flight distance significance, which is required to be larger than 5. It can be seen that this requirement for the long-lived K^0_S particles is efficient in reducing the peaking background to a negligible level.

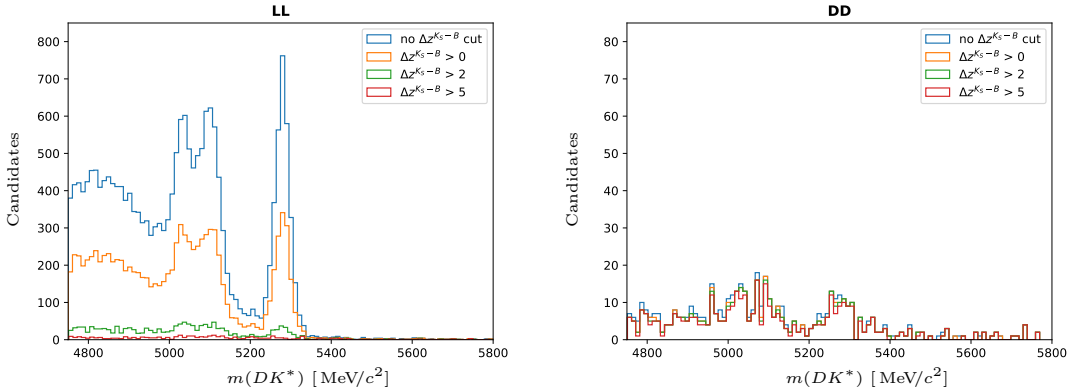


Figure 4.4: Invariant-mass distribution in the $K\pi$ mode for different values of the flight distance significance cut for LL (left) and DD (right) tracks, obtained by taking the candidates in the K^0_S mass sidebands.

As would be expected for DD candidates, there is no peak at the B mass as the condition is already satisfied for DD tracks. Indeed, these tracks do not have hits in the VELO but only hits in the tracking stations, meaning that the K^0_S particle flies significantly from the B vertex.

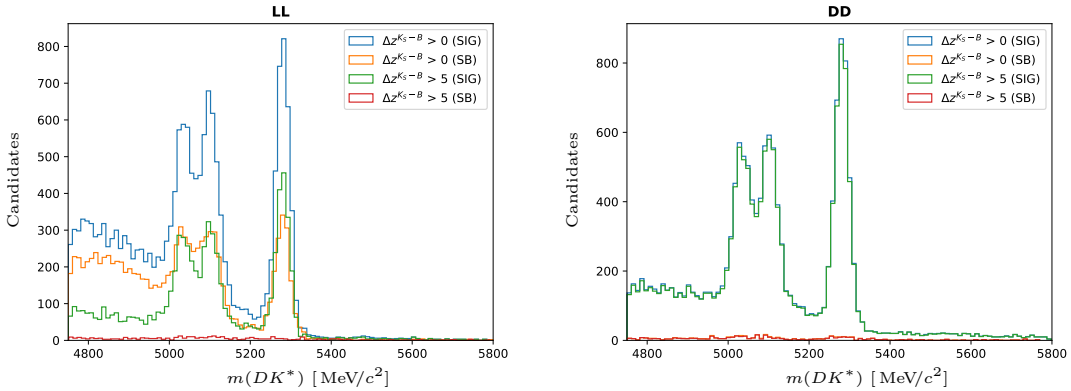


Figure 4.5: Invariant-mass distribution in the $K\pi$ mode for different values of the flight distance significance cut for LL (left) and DD (right) tracks, obtained by taking the candidates in the K^0_S mass sidebands (SB), compared to candidates within the signal region (SIG).

The invariant mass from the sidebands (SB) is then also compared to the B invariant-mass distribution for candidates within the signal region (SIG), which is shown in Figure 4.5 for LL (left) and DD (right) tracks. The cut has a negligible effect on DD candidates since there is no background coming from $B \to D\pi\pi\pi$ decays, while efficiently retaining signal candidates. Concerning the LL candidates, it is possible to see that a significant amount of the peaking background is removed to a negligible level, while keeping a large portion of the signal with an efficiency of about 96%.

4.2.4.3 Charmless $B \to hh(hh)K^*$ background

Another important peaking background in the analysis comes from $B \to hh(hh)K^*$ decays. This is called charmless background as the final state is reached without proceeding through the charmed D meson. In order to reduce this background, a similar quantity to the flight distance significance (Equation (4.1)) is defined for the D meson instead of the K^0_S particle. In this case, the D meson decay vertex is required to be well separated from the B vertex, by satisfying $\Delta z^{D-B}_{\text{significance}} > 2$ for the ADS/GLW modes and $\Delta z^{D-B}_{\text{significance}} > 0.5$ for the BPGGSZ modes. The charmless background is reduced efficiently while keeping a high signal efficiency. However, the charmless background is not entirely removed with these cuts. This is dealt with and taken into account in the final fit, as detailed in Section 4.4.3.3.

4.2.4.4 Clones

The last background which is addressed in the preselection comes from clones, which are tracks that are used more than one time to reconstruct a particle in a given event. This background is relevant in the four-body D final states, where more hits and particles are combined to form the D meson. In order to suppress clones, the opening angle between these tracks is studied. For background candidates coming from clones, it is expected that the opening angle between two tracks used to reconstruct a given candidate would tend to small values. Figure 4.6 presents the minimum opening angle between the pairs of tracks in the event, for the favoured four-body decay. The minimum of this opening angle θ_{ij}, which is defined as the angle between any pair of tracks i, j of the D decay products, is required to be

larger than 0.03° [99]. This selection cut maintains the signal efficiency at a high level of about 98% while removing the clones efficiently.

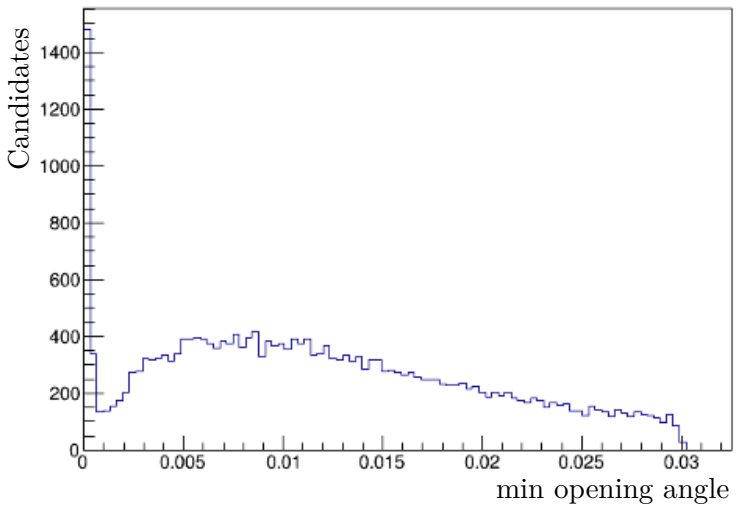


Figure 4.6: Minimum opening angle (in radians) between pairs of tracks in the 4-body D meson favoured decay, with a clear peak close to 0 made of clone tracks.

4.2.5 Multivariate analysis: Boosted Decision Tree

After reducing the background and dealing with specific backgrounds in the preselection, the next stage of the selection makes use of machine learning techniques to further reduce the background level. These methods are established and have many applications in high energy physics. In this analysis, gradient Boosted Decision Trees (BDTs) [100] are used to classify events as signal or background based on the value of a set of variables. This is illustrated in Figure 4.7, which shows an example tree where events are classified as signal (S) or background (B). This is implemented using ROOT and the included TMVA package [101]. This multivariate analysis is used as it performs better for a higher number of variables, which can be correlated, than a simple rectangular cut-based selection. The BDT is boosted by using a forest of trees, called weak learners. The ensemble of trees is used to form a single classifier output, based on the sequential training of the individual trees. In this procedure, events which are misclassified in a training step are assigned a higher weight for the next step, allowing the algorithm to increase the overall performance.

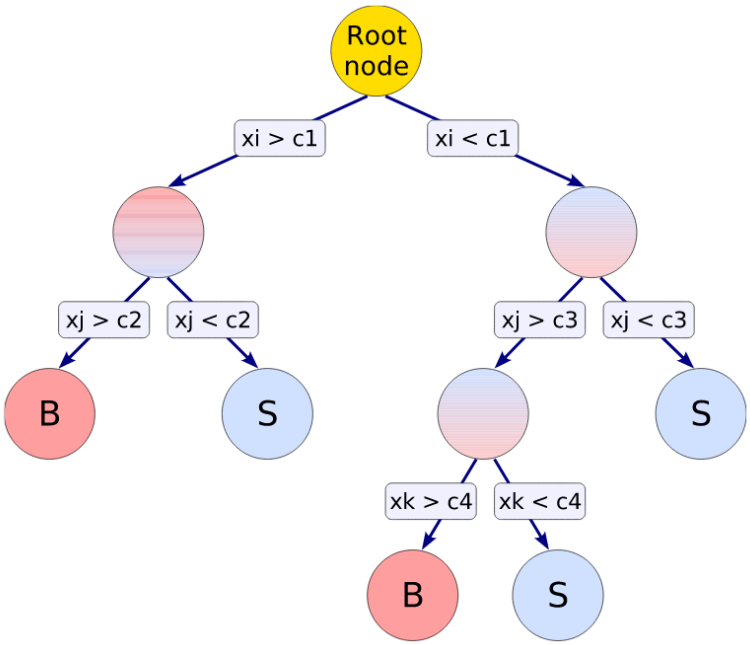


Figure 4.7: Example of the structure of a Boosted Decision Tree used in the analysis to classify events as signal or background based on the value of a set of input variables [101].

The aim of the BDT is to further reduce the amount of combinatorial background still present in the samples at this stage. The events are classified in a range between -1 and 1 based on their probability to be a background or signal event respectively. Separate BDTs are used for ADS/GLW and BPGGSZ modes, for each number of final-state particles.

Concerning the ADS/GLW modes, BDTs are trained separately for two- and four-body modes, as well as for DD and LL tracks. Two categories of events are used as proxy samples to train the BDT in order for it to identify signal and background events. These samples have the preselection cuts applied. As a signal proxy, MC simulation samples of signal events are used. The background is taken from the high-mass sideband in data above 5400 MeV/c^2, where combinatorial background is present. The MC samples and data sidebands are split into two sub-samples, which are the training and testing samples. This method is used in order to test that the BDT is not overtrained. Due to the limited available high-mass sideband in data above the fit range, more events are needed in order to make sure to avoid overtraining of the BDT. Indeed, the fit range extends up to 5600 MeV/c^2, while

the available sideband from stripping goes up to 5800 MeV/c^2. To increase the number of events and obtain a larger training sample, events which are closer to the signal and within the fit range are used. Therefore, candidates from an invariant mass of 5400 MeV/c^2 are selected in the training sample which is used as proxy for combinatorial background. To ensure that the data events within the fit range are not used for the training as well as subsequently classified by the BDT, two independent BDT classifiers A and B are formed for each BDT, which is known as k-folding method [102]. This is achieved by forming the A and B BDTs based on the event number being odd or even, respectively. In this way, the BDT A, trained with odd events, is applied independently to the dataset B corresponding to the data with an even event number. A schematic representation is shown in Figure 4.8, which illustrates the procedure of the training and the application of the BDT, where half of the events is used for training and half for testing.

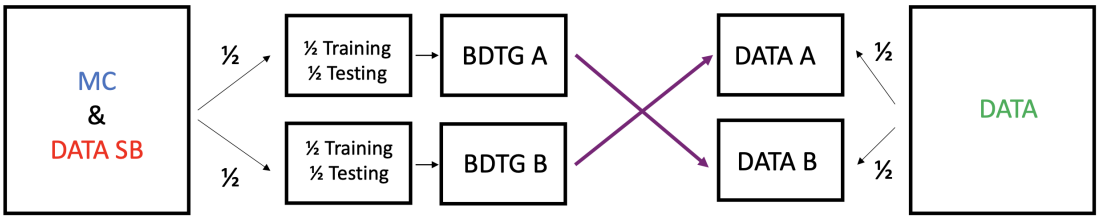


Figure 4.8: Schematic representation of the application of the BDT classifiers to data and MC samples.

In order to distinguish signal and background events, the BDT makes use of a set of variables. The list of variables used is shown in Table 4.6, where the importance and rank of the variables is given for the 4-body, LL tracks BDT B. This includes momenta of the particles and their vertex positions, geometry considerations or the fit quality of the reconstruction. An efficient variable in combinatorial background rejection is the transverse momentum (p_T) asymmetry between the B candidate and the other tracks contained in a cone around this candidate, defined as

$$A_{p_T} = \frac{p_T(B) - p_T(cone)}{p_T(B) + p_T(cone)}. \tag{4.2}$$

This variable, where $p_T(cone)$ refers to the sum of the p_T of all tracks in the cone whose circle has a radius of 1.5 units, quantifies the isolation of the B candidate and

efficiently reduces combinatorial background. Some of the variables in Table 4.6 are only used for a specific number of final states or track type. Variables 26–27 are only used in the four-body BDT and variables 23–25 are only used for LL tracks.

The distributions of the input variables for signal and background events are presented in Figures 4.9– 4.10. Figure 4.11 (left) presents the classifier output of the BDT for signal and background events, for both training and testing samples. It can be seen that the training and testing distributions are in agreement and no overtraining is seen for the classifier. Figure 4.11 (right) displays the ROC curve, which gives the background rejection as a function of the signal efficiency. The larger the area under curve is, the more efficient is the BDT in rejecting background while keeping a high signal efficiency. The BDT is then applied to data and the resulting distribution of the classifier output applied to the favoured mode is shown in Figure 4.12. A good separation between background and signal can be seen, with values close to -1 and 1 respectively. In order to determine the cut value to apply on this BDT response variable, an optimisation is performed and described in Section 4.2.8. For the BPGGSZ modes, similar strategy and variables are used [103].

Table 4.6: BDT input variables and corresponding importance as an example for the 4-body, LL tracks BDT B.

Number	Variable	Variable name	Importance [%]	Rank
1	χ^2_{IP} of the D child K	log(h1_IPCHI2_OWNPV)	4.32	6
2	χ^2_{IP} of the D child π (same sign as K for 4-body)	log(h2/3_IPCHI2_OWNPV)	4.93	1
3	χ^2_{IP} of the companion π	log(Bach_IPCHI2_OWNPV)	4.01	10
4	χ^2_{IP} of the B	log(Bu_IPCHI2_OWNPV)	4.01	11
5	χ^2_{IP} of the D	log(D0_IPCHI2_OWNPV)	3.11	24
6	B displacement χ^2	log(Bu_FDCHI2_OWNPV)	3.90	14
7	$p_{\mathrm{T}}(B)$	log(Bu_PT)	3.60	18
8	$p(D)$	log(D0_P)	4.37	5
9	$p_{\mathrm{T}}(D)$	log(D0_PT)	3.97	12
10	companion momentum $p(\pi)$	log(Bach_P)	3.51	20
11	companion transverse momentum $p_{\mathrm{T}}(\pi)$	log(Bach_PT)	3.69	17
12	$p(K^0_{\mathrm{S}})$	log(BachKs_P)	4.21	7
13	$p_{\mathrm{T}}(K^0_{\mathrm{S}})$	log(BachKs_PT)	3.43	22
14	Radial distance between best PV and B	log(Bu_RHO_BPV)	3.87	15
15	Radial distance between best PV and D	log(D0_RHO_BPV)	3.96	13
16	Distance of closest approach of B vertex	Bu_MAXDOCA	3.49	21
17	Asymmetry parameter of the B (A_{p_T})	Bu_PTASY_1.5	4.04	9
18	χ^2 of vertex fit of the B	log(Bu_VTXCHI2DOF)	4.08	8
19	χ^2 of vertex fit of the K^0_{S}	log(BachKs_VTXCHI2DOF)	2.57	25
20	Direction angle of the B	log10(1-Bu_DIRA_BPV)	3.59	19
21	Direction angle of the D	log10(1-D0_DIRA_BPV)	3.37	23
22	DecayTreeFitter χ^2/d.o.f	log(Bu_constD0KSPV_CHI2NDOF)	4.48	4
23	χ^2_{IP} of the K^0_{S}	log(BachKs_IPCHI2_OWNPV)	4.69	3
24	Min χ^2_{IP} of the K^0_{S} children	log(min_BachKsh_IPCHI2_OWNPV)	4.92	2
25	Max χ^2_{IP} of the K^0_{S} children	log(max_BachKsh_IPCHI2_OWNPV)	3.79	16
26	Min χ^2_{IP} of the D child pions	min_D0pi_IPCHI2_OWNPV	0.87	27
27	Max χ^2_{IP} of the D child pions	max_D0pi_IPCHI2_OWNPV	1.25	26

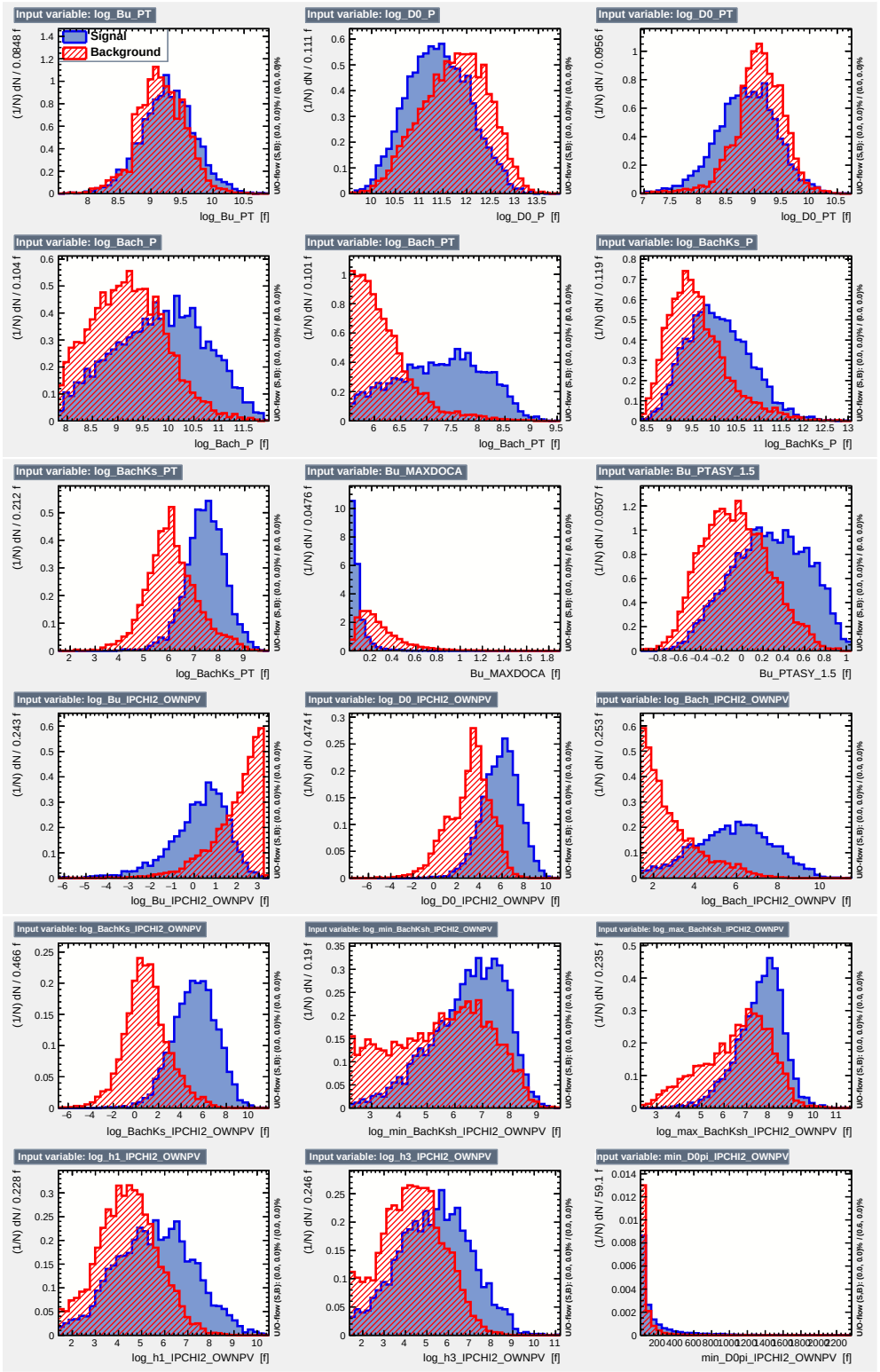


Figure 4.9: Distribution of the variables (part 1) used in the BDT for signal (blue) and background (red).

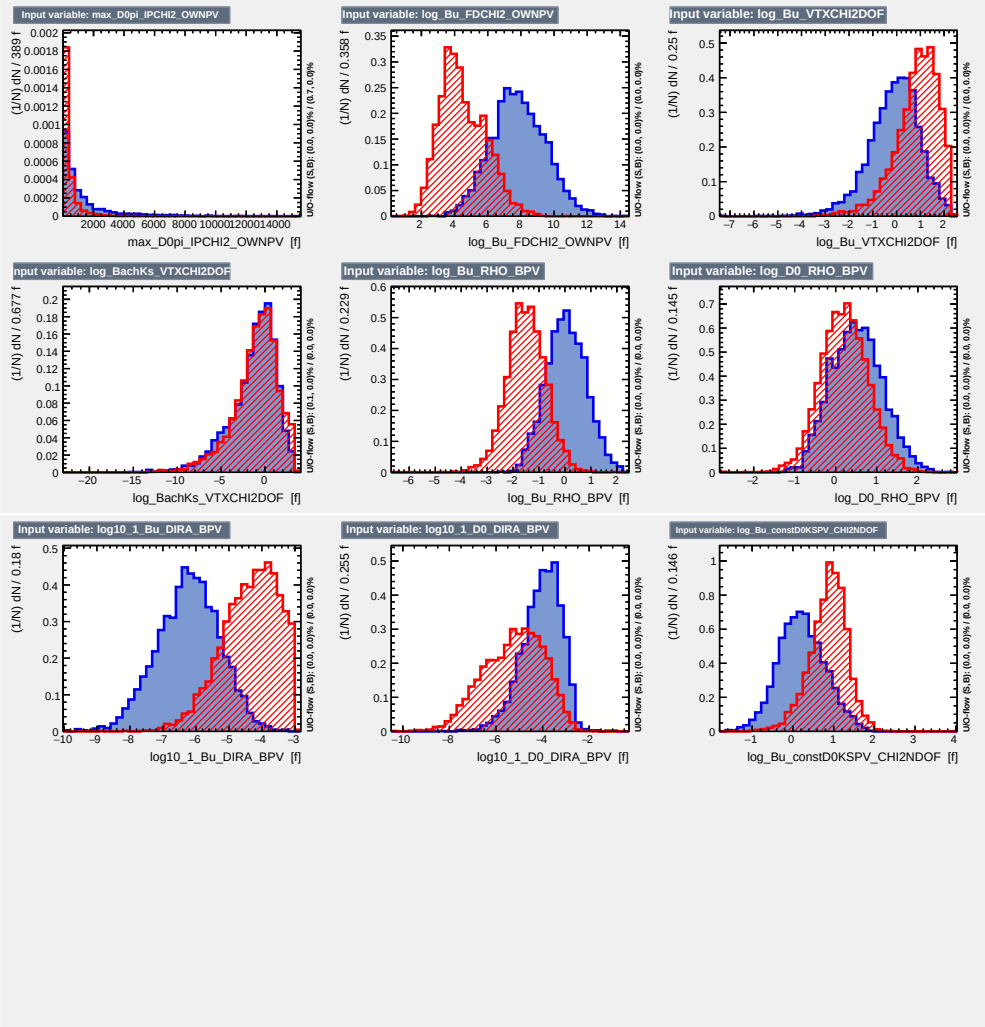


Figure 4.10: Distribution of the variables (part 2) used in the BDT for signal (blue) and background (red).

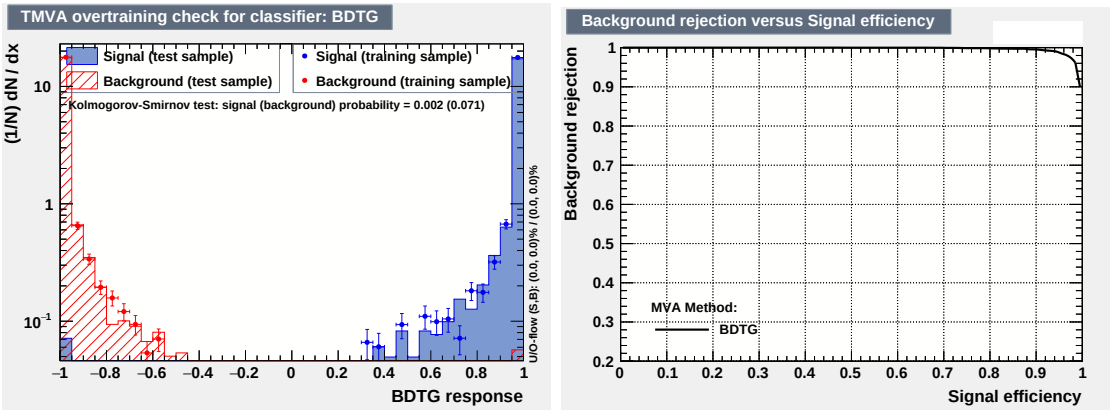


Figure 4.11: Classifier output (left) and background rejection versus signal efficiency (ROC curve) (right).

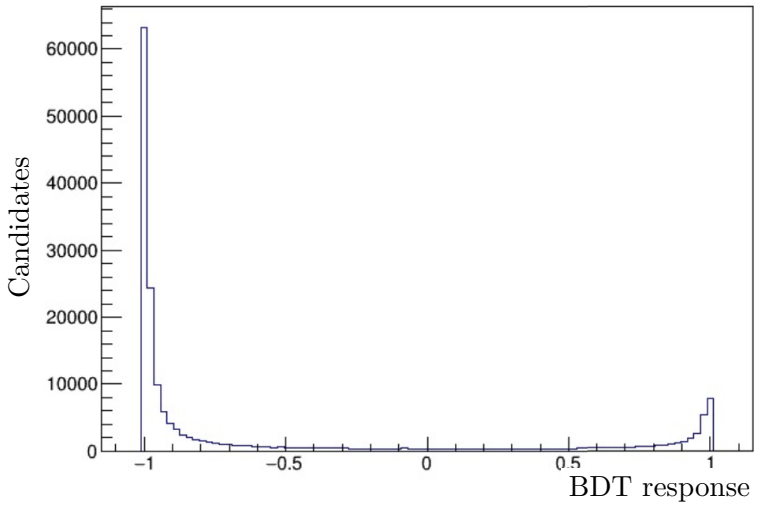


Figure 4.12: Output distribution of the BDT classifier applied to data.

4.2.6 PID requirements

As introduced in Section 3.3.2, particle identification plays a key role in many LHCb analyses, including the analysis presented in this thesis. In particular, the selection is very similar and common to all the different D decay modes. The candidates are placed in a unique D decay mode category by identifying the D daughters as being kaons or pions. This is achieved by applying PID requirements on the D daughter particles. The kaons are required to satisfy PIDK > 1 (2) for Run 2 (Run 1), while pions are required to satisfy PIDK < -1 (-2) for Run 2 (Run 1). These requirements imply that no candidate appears in two categories. In the four-body final states $K^{\pm}\pi^{\mp}\pi^{\pm}\pi^{\mp}$ and $\pi^{\pm}K^{\mp}\pi^{\pm}\pi^{\mp}$, only the two same-sign pions are required to satisfy the PID cuts. For the $\pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^{\mp}$ mode, the cut is applied on the two pions that have the same charge as the B meson (first and third pions). Concerning the three-body self-conjugate modes $K^0_S K^+ K^-$ and $K^0_S \pi^+ \pi^-$, the kaons are selected with PIDK > -5 and the pions with PIDK < 20. These loose PID cuts are applied to reduce misidentified D decays. In addition, it is required that the kaons are within the RICH momentum range ($p < 100$ GeV/c) and have information from the RICH detector (`hasRich = True`).

Background could arise from candidates being assigned to the wrong D decay mode category. This is called cross-feed background and could contribute to the

selected events as misidentified D daughter tracks. The selection on the D mass window removes cross-feed background events between the favoured mode and the CP-eigenstates (KK, $\pi\pi$). This is due to the fact that the D reconstructed mass for such an event would not be within the selected mass window for a $K \leftrightarrow \pi$ misidentification. However, the PID requirements do not suppress the cross-feed background between the favoured and suppressed modes. The residual cross-feed background between $K\pi$ and πK modes is dealt with in Section 4.2.7.1.

In order to reduce the small contribution from $B^{\pm} \to DK^0_S K^{\pm}$ decays, where the companion kaon is misidentified as a pion, a condition is added on the PID variable of the companion pion, coming from the K^* resonance. The requirement is PIDK< 4, which is studied and found to be sufficient in reducing this background to a negligible level. Section 4.4.4.1 presents the details of the calculation which shows this contamination is negligible.

Background from semileptonic decays $D \to K^{*-}(K^0_S\pi^-)l^+\nu_l$ are suppressed by adding additional muon and electron PID requirements on the D daughter kaons and pions. In both $D \to K^0_S K^+K^-$ and $D \to K^0_S\pi^+\pi^-$ decays, kaons and pions are required to have `isMuon` $= 0$. Pions in the $K^0_S\pi^+\pi^-$ mode have an additional PIDe < 0 cut [49].

4.2.7 Candidate vetoes

Additional specific backgrounds are removed by applying vetoes to reject candidates. These vetoes are based on the value of the reconstructed invariant mass of the candidates.

4.2.7.1 Double misidentified (cross-feed) background

The favoured and suppressed modes have the same particles in the final state but with opposite charges. Contamination could arise from the favoured mode to the suppressed mode when both tracks are misidentified ($K \leftrightarrow \pi$). The kaon would be misidentified as a pion and the pion would be misidentified as a kaon. A double mis-ID veto is added and applied to the suppressed mode to remove this background,

which would peak under the signal. Each candidate is reconstructed with both tracks swapped, and the candidates with a reconstructed mass within 15 MeV/c^2 of the D mass are removed. This reduces the double misidentified background to a negligible level. For the four-body final states, a similar background would contribute. In this case, there are two possible combinations of a pion to be misidentified as the kaon from the favoured $D \to K^{\pm}\pi^{\mp}\pi^{\pm}\pi^{\mp}$ mode into the suppressed $D \to \pi^{\pm}K^{\mp}\pi^{\pm}\pi^{\mp}$ mode. Therefore, two vetoes are applied in the suppressed mode for the two pions of the same sign (and opposite sign to the kaon).

4.2.7.2 Background from $D \to K^0_S\pi\pi$ in $D \to \pi\pi\pi\pi$

In the $D \to \pi\pi\pi\pi$ mode, background events could contribute from $D \to K^0_S\pi\pi$ decays. To remove this background, vetoes are applied on the K^0_S invariant mass. This requirement was introduced in Reference [104] and was also applied in the $B \to D(hh\pi\pi)h$ analysis [105]. For any $\pi^+\pi^-$ pair in the $D \to \pi^+\pi^-\pi^+\pi^-$ mode, the invariant mass of that pair is required to be away from the K^0_S known mass and therefore outside of the range [480, 505] MeV/c^2.

4.2.8 BDT optimisation

After applying all the selection criteria until this stage and detailed above, the cut on the BDT output is optimised. This is determined such that the sensitivity to the *CP* observables is maximised. In other words, the goal is to find the value of the BDT output above which the events are kept in the final sample for which the uncertainty on the measured observables is minimised. The ADS/GLW and the BPGGSZ BDTs are optimised separately.

ADS/GLW

In order to optimise the BDT selection, an ensemble of pseudoexperiments is generated. These are used to determine the expected uncertainties on the *CP* observables instead of performing the optimisation directly on the data in order to avoid any bias based on the statistical fluctuations in data. To this aim, the expected yields and efficiencies are needed, which are inputs to the *CP* fit used to

extract the *CP* observables, for different values of the BDT cut. The fit procedure is detailed in Section 6.4. The expected signal yields are estimated from the global mass fit to the favoured mode, described in Section 6.3. The MC simulation samples are used to obtain the signal efficiencies. The combinatorial background level is estimated by determining combinatorial background efficiencies from the high-mass sideband above 5600 MeV/c^2. The *CP* observables are computed for fixed values of the physics parameters obtained from a previous LHCb gamma combination [106]. The pseudodatasets are therefore generated with $r_B = 0.1$, $\delta_B = 35°$ and $\gamma = 65.4°$.

For the favoured and GLW modes, the optimisation scans are shown in Figures 4.13–4.14, which present the uncertainty on the *CP* observables as a function of the BDT cut value. For each observable, the uncertainty is defined as the mean of the uncertainty distribution obtained form the ensemble of pseudoexperiments, performed at each BDT point. The error on that uncertainty is taken to be the standard deviation of that uncertainty distribution. The combinatorial background is very efficiently removed by the BDT, which makes it possible to achieve a high signal purity even for relatively low values of the cut on the BDT response. For example, more than 50% of the combinatorial background level is reduced by moving from a cut value of 0 to a cut value of 0.6, while only dropping the signal efficiency from 95% to 90% for the $K\pi$ mode and from 96% to 92% for the $K\pi\pi\pi$ mode. Therefore, the expected sensitivity is not significantly increased between these working points in the favoured modes and the choice is made to reject the additional background and select a cut value of 0.6 on the BDT response, which corresponds to the optimal (or very close to the optimal) working point for the GLW modes. This gives a common BDT cut value for all favoured and GLW modes.

Since the suppressed modes have not been observed to date, the optimisation was initially done separately for the suppressed modes. The figure of merit to maximise was chosen to be the significance $S/\sqrt{S+B}$, where S and B are the estimated number of signal and background events in the signal region defined as [5230, 5330] MeV/c^2. Subsequently, the optimisation was also performed with the default metric, that is the uncertainty on the *CP* observables, and found to yield the

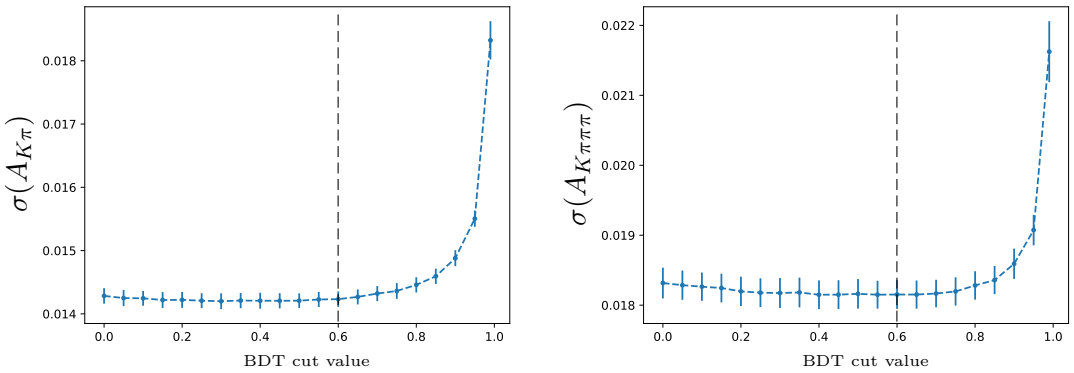


Figure 4.13: Expected uncertainty on the *CP* observables for the favoured modes for different values of the cut on the BDT response, obtained using an ensemble of pseudoexperiments.

same optimal BDT cut. The optimisation scans for the suppressed modes are given in Figure. 4.15. For both suppressed modes in the analysis, the optimal working point is found to be at a value of 0.95. This requirement on the BDT response minimises the uncertainty on the *CP* observables measured in these suppressed modes, which also corresponds to the point where the significance is maximised. For comparison, Figure 4.16 shows the scans obtained by using the significance as a figure of merit for the favoured and GLW modes. In summary, a BDT cut of 0.6 is applied for the favoured and GLW modes, and a value of 0.95 is chosen for the suppressed modes.

BPGGSZ

The BDT selection cut for the BPGGSZ modes is chosen in a similar way. Pseudo-experiments are generated with fixed values of the physics parameters, $r_B = 0.1$, $\delta_B = 127°$ and $\gamma = 64°$, using inputs from the global mass fit. Using the equations relating the *CP* observables to the physics parameters, γ and its corresponding uncertainty is determined. This uncertainty is obtained for a range of BDT cut values, giving a minimised uncertainty at a cut value of 0.9 [103].

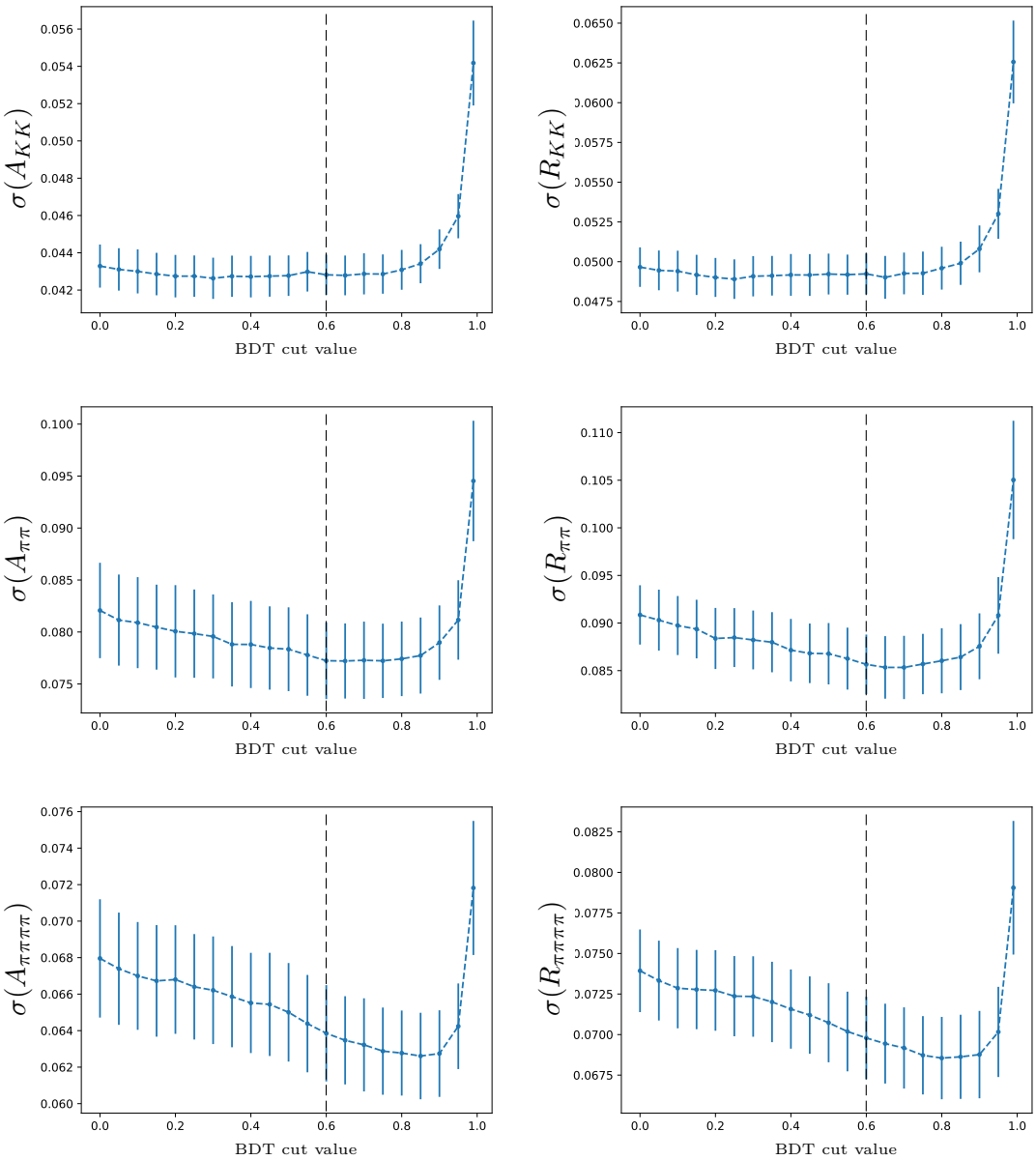


Figure 4.14: Expected uncertainty on the *CP* observables for the GLW modes for different values of the cut on the BDT response, obtained using an ensemble of pseudoexperiments.

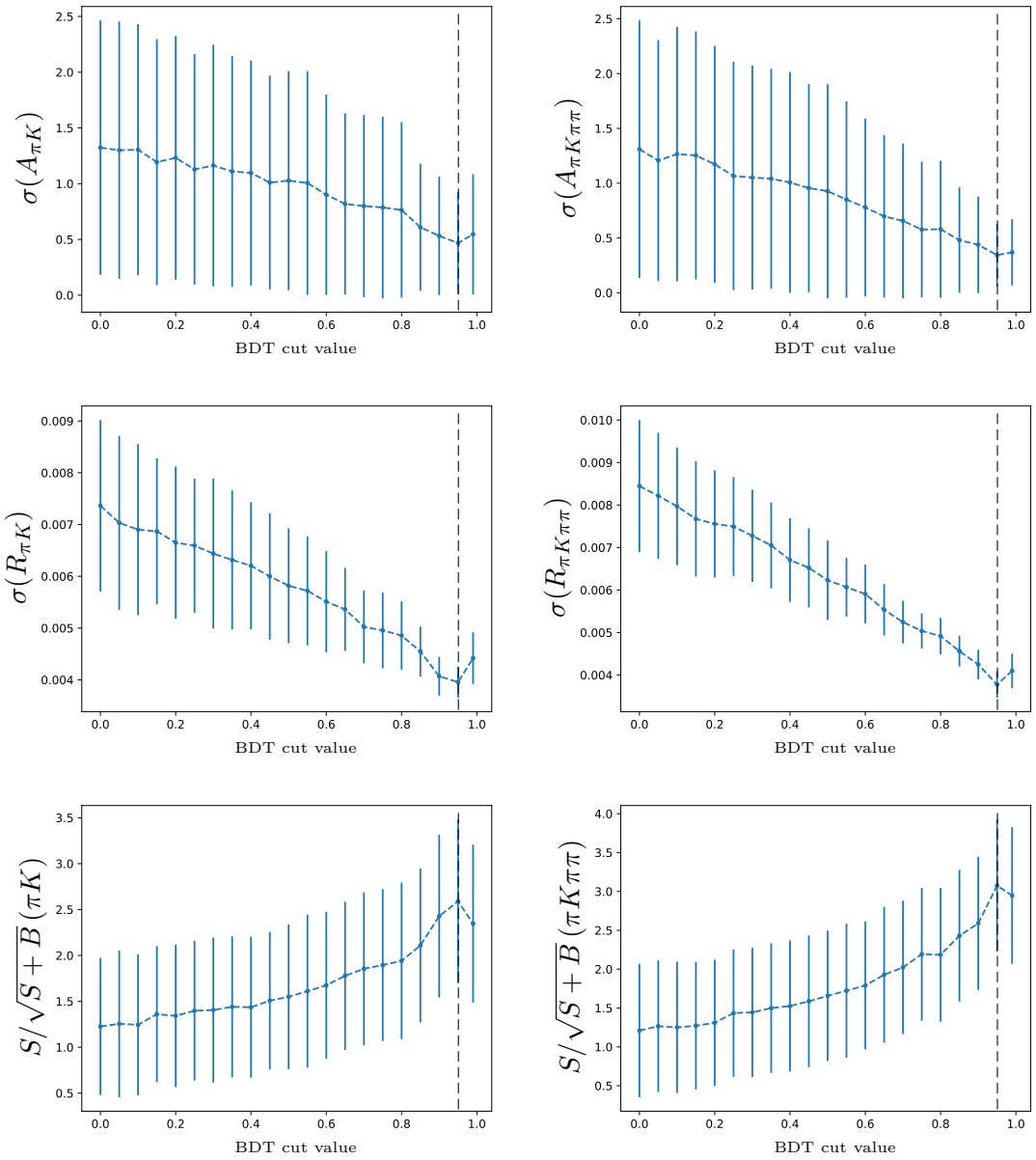


Figure 4.15: BDT optimisation for the suppressed modes: expected uncertainty on the *CP* observables (top and middle) as a function of the BDT cut value, obtained using an ensemble of pseudoexperiments, and significance as a function of the BDT working point (bottom), used to optimise the two suppressed modes.

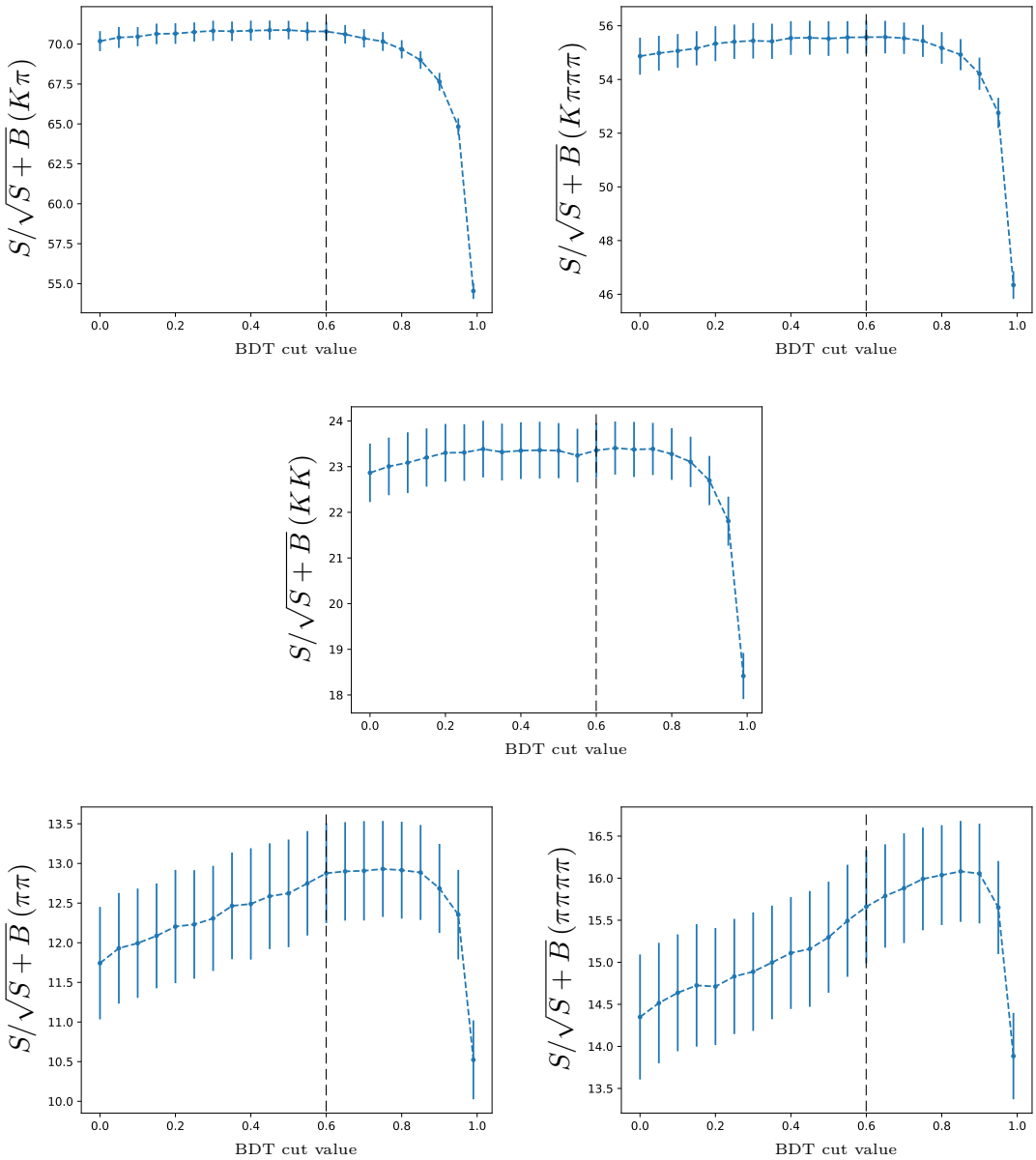


Figure 4.16: Significance as a function of the BDT working point for the favoured (top) and GLW (middle and bottom) modes, obtained using an ensemble of pseudoexperiments.

4.2.9 Multiple candidates

The final step of the selection consists in dealing with multiple candidates in the samples. These concern candidates where more than one candidate is selected from the same single event. This can for example occur when two candidates share part or all of the objects used in the reconstruction, such as tracks. For data samples, the rate of these multiple candidates after all selection steps is given in Table 4.7 for different D decay modes. It is found that the rate is $< 0.7\%$ for two-body final states, $< 2.6\%$ for four-body final states and $< 1\%$ for three-body final states. Given the low rate of multiple candidates present in the final data samples, one candidate is chosen randomly in datasets where multiple candidates are still present after the selection.

Table 4.7: Rate of multiple candidates in different D decay modes, where the numbers are given in % for the data after the full selection.

D decay mode	DD	LL
$K\pi$	0.4	0.7
KK	0.3	0.7
$\pi\pi$	0.1	0.2
πK	0.2	0.2
$K\pi\pi\pi$	1.7	2.3
$\pi\pi\pi\pi$	1.9	2.6
$\pi K\pi\pi$	1.6	1.6

4.2.10 Summary of final offline cuts

The final offline selection cuts are summarised in Table 4.8, which are applied after the preselection. These are the BDT cut, the PID requirements, the vetoes as well as the treatment of the multiple candidates. To show the effect of these cuts, the invariant-mass distribution for the two-body favoured mode is presented in Figure 4.17, before and after applying the final offline selection.

Table 4.8: Summary of the final offline cuts, including the requirements on the BDT, PID requirements, candidate vetoes and multiple candidates.

Variable (modes)	Cut
BDT	
BDT ($K\pi$, KK, $\pi\pi$, $K\pi\pi\pi$, $\pi\pi\pi\pi$)	> 0.6
BDT (πK, $\pi K\pi\pi$)	> 0.95
BDT ($K^0_S hh$)	> 0.9
PID	
Bachelor pion	PIDK< 4
ADS/GLW modes	
Kaons ($K\pi$,KK,πK,$K\pi\pi\pi$,$\pi K\pi\pi$)	PIDK > 2 (Run 1)
	PIDK > 1 (Run 2)
Pions ($K\pi$,$\pi\pi$,πK,$K\pi\pi\pi$,$\pi\pi\pi\pi$,$\pi K\pi\pi$)	PIDK < -2 (Run 1)
	PIDK < -1 (Run 2)
BPGGSZ modes	
Kaons ($K^0_S KK$)	PIDK > -5
Pions ($K^0_S \pi\pi$)	PIDK < 20
	PIDe < 0
Kaons and pions ($K^0_S hh$)	`isMuon` $= 0$
Vetoes	
Double mis-ID (πK, $\pi K\pi\pi$)	$\lvert m_D(K \leftrightarrow \pi) - m_D^{PDG}\rvert > 15\,\text{MeV}/c^2$
$\pi^+\pi^-$ invariant mass ($\pi\pi\pi\pi$)	$\notin [480, 505]\,\text{MeV}/c^2$
Multiple candidates	one candidate randomly

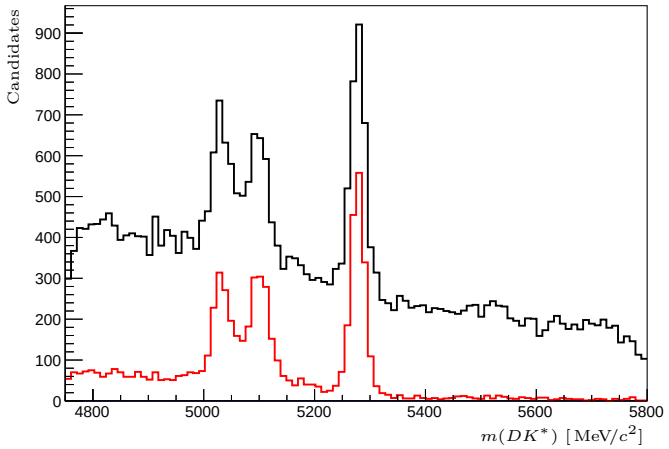


Figure 4.17: B invariant-mass distribution of the two-body favoured mode after the preselection (black) and after the final offline cuts (red), obtained with the full dataset.

4.3 Efficiencies

The candidate selection presented in this chapter can have a different efficiency in selecting signal for each D decay mode. These efficiency differences are taken into account in the *CP* fit parametrisation, described in Section 6.4, where the selection and PID efficiencies appear as inputs in the equations relating the *CP* observables to the yields. In order to determine the efficiencies, the signal MC samples which are generated for each year of data taking of Run 1 and Run 2 are used.

4.3.1 Selection efficiencies

As the selection is performed in stages, it is important to correctly determine the efficiency at each step in order to obtain the total efficiency of the selection, which is composed of the generator-level efficiency, the reconstruction efficiency and the offline efficiency:

$$\epsilon_{sel} = \epsilon_{gen} \times \epsilon_{reco} \times \epsilon_{off}. \tag{4.3}$$

As introduced in Section 3.3.5, simulating events going through the detector is computationally expensive, especially compared to the generation of the events. Therefore, cuts can be applied at the generator level to only keep interesting events in the detector simulation, saving disk space and CPU resources. This is done by removing events for which the decay products are not within the LHCb acceptance. This information is available in the generated MC statistics tables, where the generator-level efficiency ϵ_{gen} and the corresponding uncertainty are given. The stripping and reconstruction efficiency ϵ_{reco} corresponds to the ratio of the number of events which are reconstructed, and truth matched, to the total number of generated events. Lastly, the offline selection efficiency ϵ_{off} is determined by counting the number of events remaining after applying the full selection to the MC sample, except from the PID requirements, which are dealt with separately in Section 4.3.3. The uncertainties on the efficiencies are computed as binomial uncertainties:

$$\frac{1}{N_f}\sqrt{N_i(1-\frac{N_i}{N_f})}, \tag{4.4}$$

where N_i and N_f are the number of candidates before and after each selection stage respectively. The efficiencies and corresponding uncertainties are listed in Table 4.9 for the different D decay modes.

Table 4.9: Generator-level efficiency, stripping and reconstruction efficiency and offline selection efficiency for each D decay mode, determined using signal MC samples.

Mode	ϵ_{gen} [%]	ϵ_{reco} [%]	ϵ_{off} [%]	ϵ_{sel} [%]
$K\pi$	18.33 ± 0.04	0.6262 ± 0.0013	43.87 ± 0.11	0.0500 ± 0.0002
KK	19.16 ± 0.04	0.5891 ± 0.0013	43.39 ± 0.11	0.0486 ± 0.0002
$\pi\pi$	17.59 ± 0.04	0.6587 ± 0.0014	42.88 ± 0.10	0.0495 ± 0.0002
$K\pi\pi\pi$	17.30 ± 0.04	0.2153 ± 0.0008	44.10 ± 0.18	0.0164 ± 0.0001
$\pi\pi\pi\pi$	16.59 ± 0.04	0.2288 ± 0.0008	38.18 ± 0.17	0.0145 ± 0.0001

It can be seen that the generator-level efficiency is similar between the D decay modes, while the reconstruction efficiency is about a factor three lower for the four-body final states compared to the two-body final states. This is expected given that more tracks need to be reconstructed in the four-body modes. The offline selection efficiencies are close to each other for the different modes, except for the $\pi\pi\pi\pi$ mode where it is lower by about 5 percentage points. This is due to the K^0_S vetoes described in Section 4.2.7.2.

4.3.2 BDT and vetoes efficiencies

Concerning the suppressed modes, the only difference in the selection efficiency with respect to the favoured modes comes from the different BDT working points as well as the additional double mis-ID vetoes applied in the suppressed modes. Although ϵ_{BDT} is included in ϵ_{off}, it is computed separately as the ratio of selection efficiencies between the suppressed and favoured modes appears in the *CP* fit, in the following way:

$$\frac{\epsilon_{sel}(\pi K(\pi\pi))}{\epsilon_{sel}(K\pi(\pi\pi))} = \frac{\epsilon_{BDT}(\pi K(\pi\pi))}{\epsilon_{BDT}(K\pi(\pi\pi))} \times \epsilon_{veto}(\pi K(\pi\pi)). \tag{4.5}$$

The two efficiencies ϵ_{BDT} and ϵ_{veto} are determined using $K\pi(\pi\pi)$ MC samples and listed in Table 4.10.

Table 4.10: Offline selection efficiencies which are different between the favoured and suppressed modes.

Mode	ϵ_{BDT} [%]	ϵ_{veto} [%]
$K\pi$	89.59 ± 0.09	–
πK	71.41 ± 0.13	92.68 ± 0.09
$K\pi\pi\pi$	91.77 ± 0.15	–
$\pi K\pi\pi$	77.92 ± 0.22	89.31 ± 0.19

4.3.3 PID efficiencies

As mentioned in this section, the PID efficiencies are treated separately, due to the fact that PID variables are known to be poorly simulated. Correctly determining PID efficiencies is important as the PID requirements applied to the D daughters are the main difference in selection between the different modes. Therefore, instead of using the MC samples to estimate the PID efficiency for each cut, a data-driven method with calibration samples is used. Specific decays are used to form the calibration samples, where the particle type can be determined without making use of PID requirements. The main source of calibration tracks used for kaons and pions comes from $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decays. Due to the fact that the D^* decays via strong interaction, the charge of the bachelor pion is known, which tags the flavour of the D meson. In turn, this tags the type of the D daughter particles. This procedure is described in more detail in Chapter 8.

Using the tools from the PIDCalib2 package [107], the PID efficiencies are measured in bins of momentum p and pseudorapidity η. The calibration tracks obtained from the calibration samples are used to create histograms containing the efficiencies for different PID cuts and a given binning scheme for (p, η). The signal MC samples in the analysis, called reference samples, are then used to determine per-track efficiencies. Each event is assigned an efficiency based on the efficiency histograms created in the first step. The obtained PID efficiencies are presented in Table 4.11 for the different decay modes and separately for Run 1 and Run 2.
The first uncertainty on the PID efficiencies is statistical and due to the limited amount of simulation available. This statistical uncertainty is assessed by bootstrap-

Table 4.11: PID efficiencies for Run 1 and Run 2 obtained for each D decay mode using PIDCalib2.

D decay mode	Run 1 PID efficiency [%]	Run 2 PID efficiency [%]
$K\pi$	$66.86 \pm 0.14 \pm 0.29$	$77.48 \pm 0.07 \pm 0.08$
KK	$74.92 \pm 0.13 \pm 0.29$	$81.38 \pm 0.06 \pm 0.11$
$\pi\pi$	$60.07 \pm 0.13 \pm 0.31$	$73.89 \pm 0.07 \pm 0.17$
$K\pi\pi\pi$	$58.85 \pm 0.16 \pm 0.26$	$78.34 \pm 0.09 \pm 0.27$
$\pi\pi\pi\pi$	$64.34 \pm 0.17 \pm 0.22$	$81.76 \pm 0.08 \pm 0.15$

ping the reference samples, and assigned as the RMS of the distribution obtained from the bootstrapping [108]. The second uncertainty comes from systematic effects that must be considered and which are related to the method used to determine the efficiencies. The systematic uncertainty is composed of two parts.

The first component comes from the uncertainty due to the *sPlot* method [109] which is used to extract the signal candidates from the calibration tracks. For each D daughter track on which a PID cut is applied, a 0.1% uncertainty is assigned [107]. Concerning the bachelor pion, the assumption is made that the momentum of the D meson does not change significantly between the different modes to have an impact on the PID efficiency of the bachelor pion. To verify this assumption, the momentum of the bachelor pion for the different D decay modes is shown in Figure 4.18, where it can be seen that the assumption is met.

The second source of systematic uncertainty is related to the choice of the binning scheme in (p, η), with $p \in [2, 200]$ GeV/c and $\eta \in [1.5, 5.0]$. To evaluate this component of the systematic uncertainty, alternative binning schemes are used to compute the efficiency. The default binning is made of 4 uniform bins in η. For the p bins, the bin edges for the first bins are placed at [2, 9.3, 15.6, 19] GeV/c, and a larger last bin between [100, 200] GeV/c. In between, the region from 19 to 100 GeV/c contains 18 uniformly distributed bins. The default binning scheme is referred to as (18, 4), which corresponds to the number of uniform bins in (p, η). Four alternative binning schemes are used to estimate the systematic uncertainty related to the choice of the binning scheme, (14, 2), (16, 3), (20, 5)

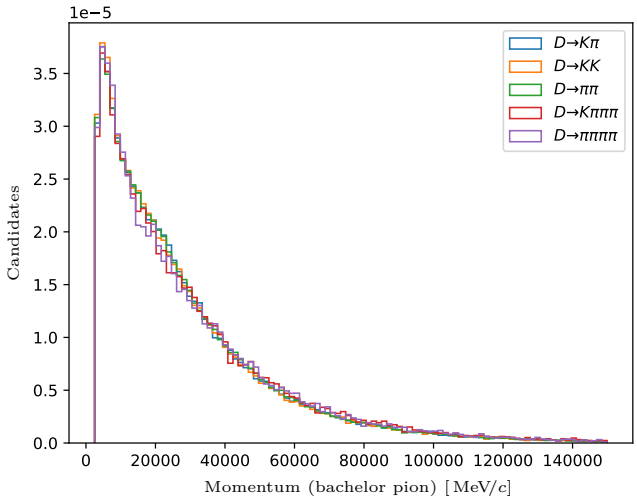


Figure 4.18: Momentum of the bachelor pion in simulation for the different D decay modes.

and (22, 6), numbered 1–4 respectively. Table 4.12 presents the PID efficiencies (and corresponding statistical uncertainties) obtained with the alternative binning schemes, as well as the associated systematic uncertainty. The latter is taken to be the standard deviation of the efficiencies obtained with the different binning schemes.

4.3.4 Total combined efficiency

The efficiency used as input in the *CP* fit is the total efficiency, which is the product of the selection and PID efficiencies:

$$\epsilon_{tot} = \epsilon_{sel} \times \epsilon_{PID}. \tag{4.6}$$

For each year of data taking and track type, the total efficiency is computed separately. However, since the analysis is performed by combining the track types (DD and LL) and using the full dataset (Run 1 and Run 2), a combined total efficiency needs to be determined. This combined efficiency is obtained from the individual efficiencies, as illustrated in Figure 4.19. A weighted average is first performed to combine the individual total efficiencies from each year, where the weights correspond to the respective integrated luminosity of data collected, listed in Table 4.1.
This procedure is performed separately for each track type, DD and LL tracks, giving the efficiencies ϵ_{DD} and ϵ_{LL} respectively. As the track types are combined

Table 4.12: Determination of the systematic uncertainty due to the binning scheme chosen to evaluate the PID efficiencies.

D mode	Binning	Run 1 PID efficiency [%]	Run 2 PID efficiency [%]
$K\pi$	Default	66.86 ± 0.14	77.48 ± 0.07
	Alternative 1	66.99 ± 0.13	77.36 ± 0.06
	Alternative 2	66.44 ± 0.13	77.28 ± 0.07
	Alternative 3	66.33 ± 0.13	77.32 ± 0.07
	Alternative 4	66.44 ± 0.14	77.44 ± 0.07
	Systematic uncertainty	0.294224	0.082299
KK	Default	74.92 ± 0.13	81.38 ± 0.06
	Alternative 1	75.05 ± 0.14	81.22 ± 0.07
	Alternative 2	74.49 ± 0.13	81.28 ± 0.07
	Alternative 3	74.48 ± 0.13	81.47 ± 0.06
	Alternative 4	74.42 ± 0.14	81.44 ± 0.07
	Systematic uncertainty	0.289884	0.106071
$\pi\pi$	Default	60.07 ± 0.13	73.89 ± 0.07
	Alternative 1	60.19 ± 0.12	73.76 ± 0.07
	Alternative 2	59.62 ± 0.14	73.50 ± 0.07
	Alternative 3	59.46 ± 0.12	73.50 ± 0.07
	Alternative 4	59.68 ± 0.13	73.71 ± 0.07
	Systematic uncertainty	0.312885	0.170497
$K\pi\pi\pi$	Default	58.85 ± 0.16	78.34 ± 0.09
	Alternative 1	58.13 ± 0.16	77.70 ± 0.08
	Alternative 2	58.65 ± 0.17	78.15 ± 0.09
	Alternative 3	58.50 ± 0.17	78.22 ± 0.09
	Alternative 4	58.58 ± 0.18	78.37 ± 0.09
	Systematic uncertainty	0.263965	0.269238
$\pi\pi\pi\pi$	Default	64.34 ± 0.17	81.76 ± 0.08
	Alternative 1	63.82 ± 0.18	81.36 ± 0.09
	Alternative 2	64.28 ± 0.17	81.57 ± 0.08
	Alternative 3	63.99 ± 0.17	81.53 ± 0.09
	Alternative 4	64.22 ± 0.17	81.68 ± 0.09
	Systematic uncertainty	0.220232	0.154338

in the final dataset used in the analysis, it is important to account for differences in the selection efficiency between DD and LL tracks. This is done by combining ϵ_{DD} and ϵ_{LL} by performing a weighted average where the weights $\omega(\text{DD})$ and $\omega(\text{LL})$ reflect the number of candidates of each type in data. These weights are determined by extracting the number of signal candidates of each type in the dataset with the full selection applied. The weights obtained are about 2/3 and 1/3 for DD and LL tracks respectively. The total combined efficiencies, as well as the selection

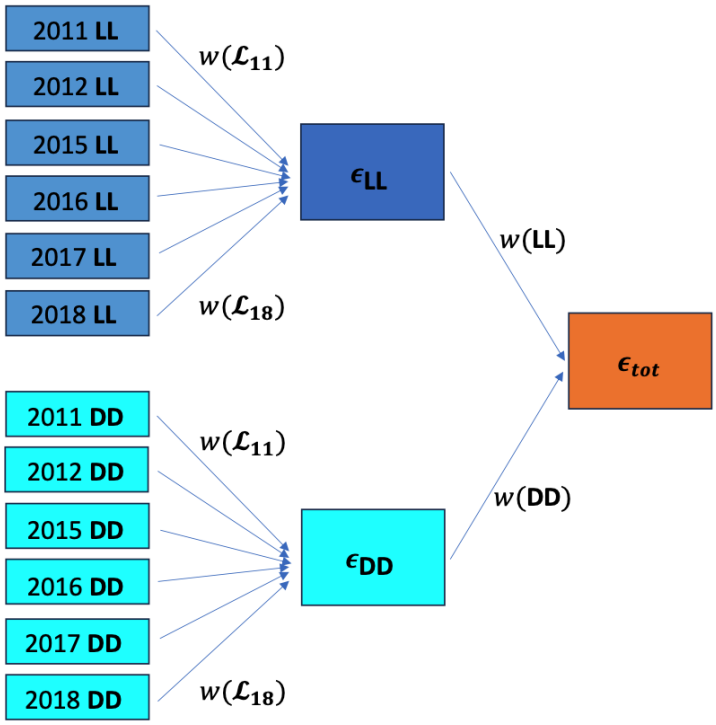


Figure 4.19: Illustration of the procedure used to combine the efficiencies.

and PID efficiencies, are listed in Table 4.13.

Table 4.13: Selection, PID and total combined efficiencies. Note that the total combined efficiency, shown in the last column in the table and used as input in the *CP* fit, does not correspond exactly to the product of the selection and PID efficiencies due to the procedure for combining the efficiencies.

Mode	ϵ_{sel} [%]	ϵ_{PID} [%]	ϵ_{tot} [%]
$K\pi$	0.0500 ± 0.0002	73.66 ± 0.12	0.0376 ± 0.0002
KK	0.0486 ± 0.0002	79.12 ± 0.12	0.0390 ± 0.0002
$\pi\pi$	0.0495 ± 0.0002	68.90 ± 0.13	0.0350 ± 0.0002
$K\pi\pi\pi$	0.0164 ± 0.0001	71.49 ± 0.14	0.0122 ± 0.0001
$\pi\pi\pi\pi$	0.0145 ± 0.0001	75.67 ± 0.12	0.0113 ± 0.0001

4.3.5 BPGGSZ phase-space efficiency

For the BPGGSZ modes, the F_i coefficients are taken from the $B^{\pm} \to D(K^0_S h^+ h^-)h^{\pm}$ analysis [49]. Since absolute changes in efficiency and the overall efficiency are irrelevant, the efficiencies are not determined in this mode. However, it is important to check if the relative efficiency across the phase space of the D meson is different

in this analysis since the same F_i coefficients are used. In this aim, simulation samples from both decay channels are used, where the full selection is applied. The distributions of kinematic variables for the D meson and the D daughter particles are compared. Moreover, the squared invariant-mass distributions $m^2(K^0_S\pi^+)$, $m^2(K^0_S\pi^-)$ and $m^2(\pi^+\pi^-)$, as well as the Dalitz plot distributions, are compared. It is found that the distributions agree with each other between the two analyses, except from $m^2(\pi^+\pi^-)$ for which this effect is taken into account as a systematic uncertainty as described in Section 6.5.10 [103].

4.4 Signal and backgrounds

4.4.1 Signal shape

In order to model the shape of the signal, the truth-matched MC samples of $B^\pm \to DK^{*\pm}$ decays for each D mode are used. The chosen signal shape is the sum of a Gaussian G and a modified Cruijff function C [49]. The latter is a modified Gaussian function which is found to model the LHCb resolution well and account for radiative decays in the low-mass tail due to final-state radiation (FSR):

$$C(m;\,\mu,\sigma,\alpha_L,\alpha_R,\beta) \propto \begin{cases} \exp\left[\frac{-(m-\mu)^2\left(1+\beta(m-\mu)^2\right)}{2\sigma^2+\alpha_L(m-\mu)^2}\right] & \text{if } m-\mu<0 \\ \exp\left[\frac{-(m-\mu)^2\left(1+\beta(m-\mu)^2\right)}{2\sigma^2+\alpha_R(m-\mu)^2}\right] & \text{if } m-\mu>0\,. \end{cases} \tag{4.7}$$

The total probability density function (PDF) for the signal shape is written as

$$F(m;\,\mu,\sigma,\alpha_L,\alpha_R,\beta) = f_C \times C(m;\,\mu,\sigma,\alpha_L,\alpha_R,\beta) + (1-f_C)\times G(m;\,\mu,\sigma), \tag{4.8}$$

where the shape parameters α_L, α_R and β are determined from simulation, and the mean μ and width σ are shared between the two functions. The fraction f_C represents the fraction of the total PDF taken by the modified Cruijff function. For each D decay mode, fits to the signal simulation samples are performed, where a full description of the LHCb detector is included. The fits are shown in Figure 4.20 for $K\pi$ (top left), $K\pi\pi\pi$ (top right), KK (middle), $\pi\pi$ (bottom left) and $\pi\pi\pi\pi$ (bottom right).

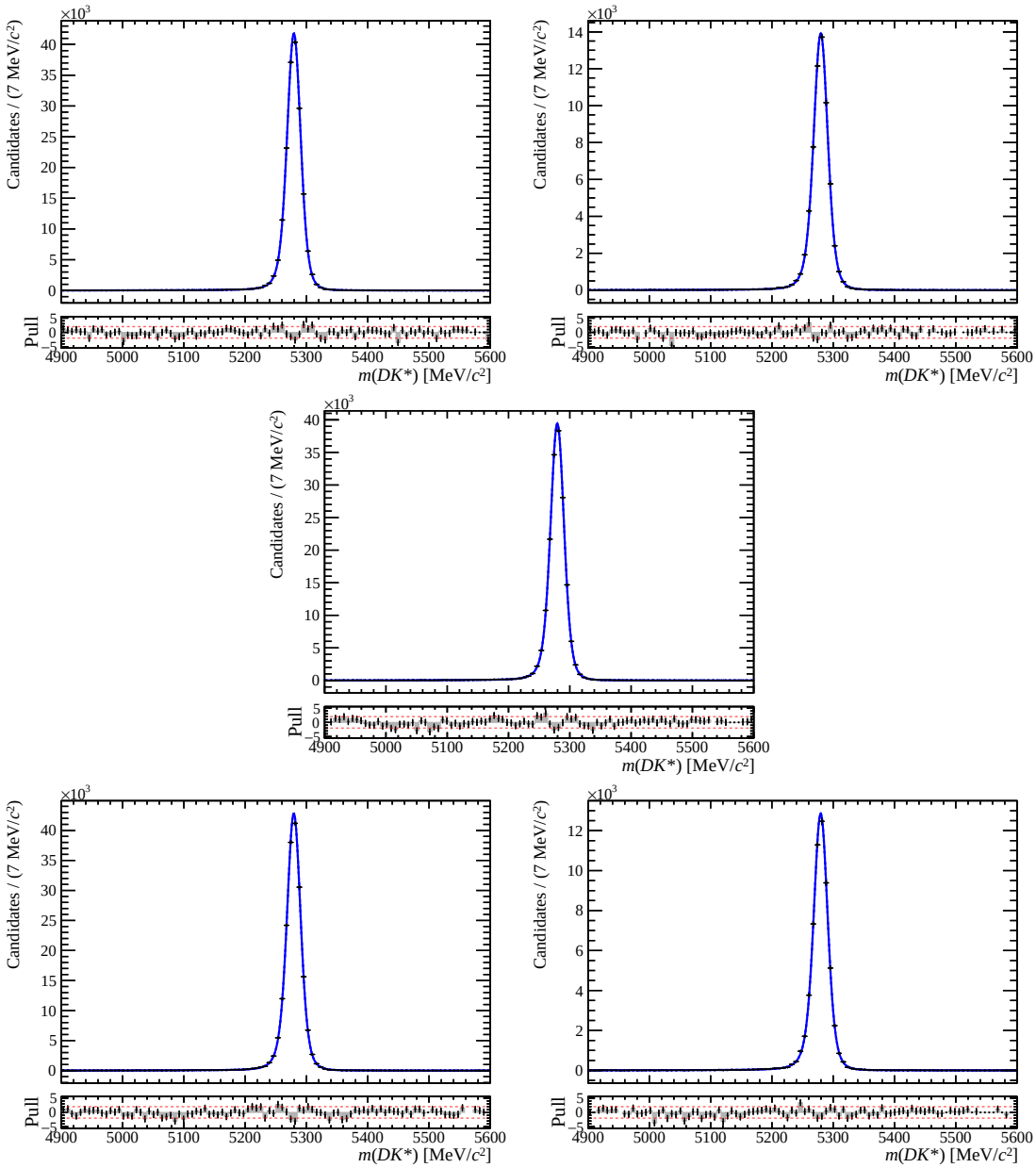


Figure 4.20: Fit to the simulated $B^{\pm} \to [K\pi]_D K^{*\pm}$ (top left), $B^{\pm} \to [K\pi\pi\pi]_D K^{*\pm}$ (top right), $B^{\pm} \to [KK]_D K^{*\pm}$ (middle), $B^{\pm} \to [\pi\pi]_D K^{*\pm}$ (bottom left) and $B^{\pm} \to [\pi\pi\pi\pi]_D K^{*\pm}$ (bottom right) signal events. The MC samples for all years are combined, each with a weight corresponding to the integrated luminosity of that year in data.

The parameters obtained as a result of these fits are listed in Table 4.14. For the BPGGSZ modes, the signal shape is determined by performing a fit to the $K^0_S\pi\pi$ signal MC sample. The shape is modelled by a modified Cruijff function, as the addition of the Gaussian function is not required.

Table 4.14: Fit parameter results obtained from the fits to simulated data for the different signal modes.

Mode	μ	σ	α_L [$\times 10^{-2}$]	α_R [$\times 10^{-2}$]	β [$\times 10^{-6}$]	f_C
$K\pi$	5279.53 ± 0.04	10.72 ± 0.08	16.1 ± 0.2	12.2 ± 0.2	2.8 ± 0.2	0.78 ± 0.04
KK	5279.57 ± 0.04	10.69 ± 0.08	16.2 ± 0.2	12.3 ± 0.2	2.7 ± 0.2	0.77 ± 0.04
$\pi\pi$	5279.40 ± 0.04	10.70 ± 0.08	15.6 ± 0.2	12.2 ± 0.2	2.3 ± 0.2	0.87 ± 0.04
$K\pi\pi\pi$	5279.63 ± 0.08	11.11 ± 0.15	16.3 ± 0.4	12.7 ± 0.3	2.4 ± 0.3	0.83 ± 0.07
$\pi\pi\pi\pi$	5279.56 ± 0.08	10.97 ± 0.17	16.5 ± 0.5	12.6 ± 0.3	2.9 ± 0.3	0.86 ± 0.08

4.4.2 Combined datasets

As already mentioned, the data from the different years and track types is combined to form the final dataset of the analysis. Similarly, the MC samples generated for each year are combined with weights corresponding to the integrated luminosity for each year in data (Table 4.1). The Run 1 and Run 2 datasets being combined in the analysis, the corresponding shapes are shown in Figure 4.21 (left). The shapes are obtained by performing individual fits to the invariant-mass distributions from the corresponding MC samples, using the $K\pi$ mode. Moreover, the DD and LL tracks are combined as it is found that the signal shapes between the two track types agree with each other, as shown in Figure 4.21 (right). The values of the shape parameters, as well as the mean and width, are checked and found to be compatible between the individual fits performed by splitting the samples into the categories which are compared. The fit parameter results for these fits are given in Table 4.15.

Table 4.15: Fit parameter results obtained from the fits to simulated data for different data samples. The comparisons between Run 1 and Run 2, and DD and LL, are shown for the $K\pi$ mode.

	μ	σ	α_L [$\times 10^{-2}$]	α_R [$\times 10^{-2}$]	β [$\times 10^{-6}$]	f_C
Run 1	5279.26 ± 0.08	10.87 ± 0.13	16.2 ± 0.5	12.3 ± 0.4	3.0 ± 0.4	0.67 ± 0.06
Run 2	5279.63 ± 0.05	10.63 ± 0.08	16.0 ± 0.2	12.2 ± 0.2	2.7 ± 0.2	0.84 ± 0.04
DD	5279.56 ± 0.05	10.67 ± 0.09	16.2 ± 0.3	12.2 ± 0.2	2.8 ± 0.3	0.76 ± 0.05
LL	5279.48 ± 0.08	10.82 ± 0.14	16.1 ± 0.4	12.2 ± 0.3	2.6 ± 0.3	0.83 ± 0.07
$K\pi$	5279.53 ± 0.04	10.72 ± 0.08	16.1 ± 0.2	12.2 ± 0.2	2.8 ± 0.2	0.78 ± 0.04
$K\pi\pi\pi$	5279.63 ± 0.08	11.11 ± 0.15	16.3 ± 0.4	12.7 ± 0.3	2.4 ± 0.3	0.83 ± 0.07

Finally, the shapes are also compared between the favoured $K\pi$ and $K\pi\pi\pi$ modes,

where larger differences can be seen, as shown in Figure 4.22. The corresponding results for the fit parameters are also listed in Table 4.15. From these results, it is found that the mean and width can be shared across modes with the same number of final states, while the datasets can be combined over years of data taking and track types.

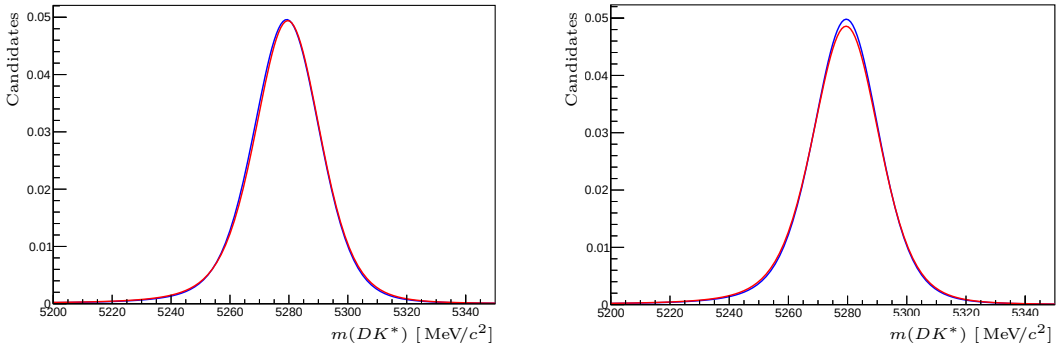


Figure 4.21: Shapes obtained from individual fits to simulation samples for unweighted Run 1 (blue) and Run 2 (red) simulation samples (left), and the shapes for DD (blue) and LL (red) tracks with the full MC weighted dataset (right) using the $K\pi$ mode.

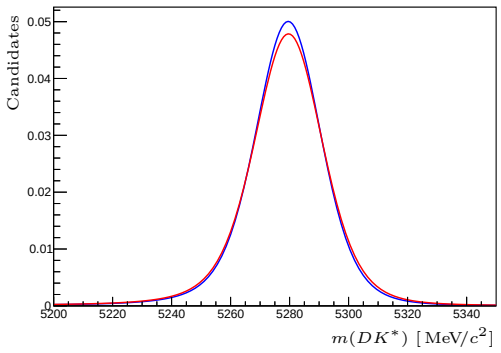


Figure 4.22: Shapes obtained from individual fits to simulation samples for the $K\pi$ (blue) and $K\pi\pi\pi$ (red) modes with the full MC weighted dataset.

4.4.3 Irreducible background shapes

After the full selection is applied, some backgrounds remain in the data samples and need to be taken into account when performing the measurement. These are irreducible backgrounds, which are listed in this subsection and for which the expected yields and invariant-mass shapes are studied.

4.4.3.1 Combinatorial background

The remaining combinatorial background in the sample is modelled by an exponential function. The slope parameter is left free to be determined directly by the data. In the *CP* fit, the slope parameter is shared across modes with the same number of final states.

4.4.3.2 Partially reconstructed $B \to D^*K^*$ backgrounds

Partially reconstructed backgrounds refer to events for which a particle is not reconstructed in a given decay and reach the same final state as the signal. In the analysis presented here, these are decays of the type $B \to D^*K^*$, with $D^* \to D[\pi^0/\gamma]$, where the pion/photon coming from the D^* is missed and not reconstructed. These backgrounds are not reducible due to their similar topology to signal candidates. However, due to the missing particle, these backgrounds sit at a lower reconstructed invariant mass where they form the main and largest contribution. Three specific partially reconstructed decays are considered in this analysis:

- $B^\pm \to D^{*0}(\to D^0[\gamma])K^{*\pm}$
- $B^\pm \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$
- $B^0 \to D^{*\pm}(\to D^0[\pi^\pm])K^{*\mp}$.

This type of background is common in analyses of similar B-hadron decays. These $B \to D^*K^*$ partially reconstructed decays are of the type Pseudoscalar $\to$ Vector Vector. Due to angular momentum conservation, three possible helicity states for the D^* resonance are possible (-1, 0, $+1$). In turn, the helicity of the missed particle is directly related to the missing momentum and thus to the B reconstructed mass. The ± 1 helicity states form two invariant-mass distributions which are physically indistinguishable [110].

These partially reconstructed decays are modelled using analytical PDFs, for which the parameters are fixed from simulation, which are of the form

$$F_{\mathrm{PR}}(m) = \int_a^b dx\ p_{\mathrm{H}}(x)\ \mathrm{G2}(x|m,\sigma,f_G,R_G)\ \left(\frac{1-\xi}{b-a}x + \frac{b\xi - a}{b-a}\right). \tag{4.9}$$

The integral is performed between the kinematic endpoints a and b, which are fully determined from the masses of the particles involved in these decays. The function $p_{\mathrm{H}}(x)$ is a parabola in the region between the kinematic endpoints, while its value is zero outside of the kinematic range. To take into account resolution effects in the LHCb detector, this parabolic function is convolved with a double Gaussian function $\mathrm{G2}(x|m, \sigma, f_G, R_G)$, where f_G is the fraction of the first Gaussian with respect to the second and R_G gives the relative width of the second Gaussian with respect to the width of the first Gaussian. Finally, the last part of the PDF, involving the parameter ξ, accounts for selection effects where events are preferentially selected based on their transverse momentum, or invariant mass. This would introduce a height difference between the two peaks, taken into account by the linear function with a slope $1 - \xi$, where $\xi = 1$ would correspond to the case where the two peaks are of equal height.

HORNS

In the case of $B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ and $B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$ decays with the D^* in the helicity state 0, the spin-0 π^0 which is missed in the reconstruction is predominantly produced along angles $\theta = 0°$ or $\theta = 180°$. This helicity angle θ is defined as the angle between the momentum of the π^0 in the D^* rest frame and the D^* boost vector in the B rest frame. This observation comes from the helicity state of the D^*, combined with angular momentum conservation. In the case where $\theta = 180°$, the momentum taken away by the π^0 is maximal, and therefore the B reconstructed mass is lower. Conversely, the B reconstructed mass is higher when the minimal momentum is carried by the π^0, corresponding to the case where $\theta = 0°$. This direct one-to-one relation between the B reconstructed mass and the momentum of the missing π^0 particle gives an invariant-mass distribution with two peaks, parametrised by the following corresponding parabola $p_{\mathrm{H}}(x)$:

$$p_{\mathrm{HORNS}}(x) = \left(x - \frac{a+b}{2}\right)^2. \tag{4.10}$$

HILL

For the partially reconstructed decays with a spin-1 missing particle, such as $B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ decays with the D^* in the helicity state 0, the missing neutral particle is predominantly produced at angles $\theta = 90°$ or $\theta = 270°$. In this case, the curvature of the parabola is negative and no double-peaked distribution is observed, since the momentum taken away by the photon is equal in both directions. The corresponding parabola can be written as

$$p_{\text{HILL}}(x) = -(x-a)(x-b). \tag{4.11}$$

LITTLEHORNS

Finally, the last case concerns $B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ decays with the D^* in the helicity state ± 1 for which the helicity angle of the photon can be described by adding a uniform term to the parabolic p_{HORNS} function:

$$p_{\text{LITTLEHORNS}}(x) = \left(x - \frac{a+b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2. \tag{4.12}$$

A summary of the partially reconstructed decays with their corresponding analytical PDF is given in Table 4.16, while the corresponding fits to the MC samples are shown in Figure 4.23.

Table 4.16: List of partially reconstructed decays considered in the analysis. For each decay and helicity state, the name of the analytical PDF used to model the corresponding distribution is given.

Decay	Helicity state	Analytical PDF
$B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$	± 1	LITTLEHORNS
	0	HILL
$B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$	± 1	HILL
	0	HORNS
$B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$	± 1	HILL
	0	HORNS

The yield ratios between the three different partially reconstructed decays, which are used in the global fit described in Section 6.3, are defined separately for the

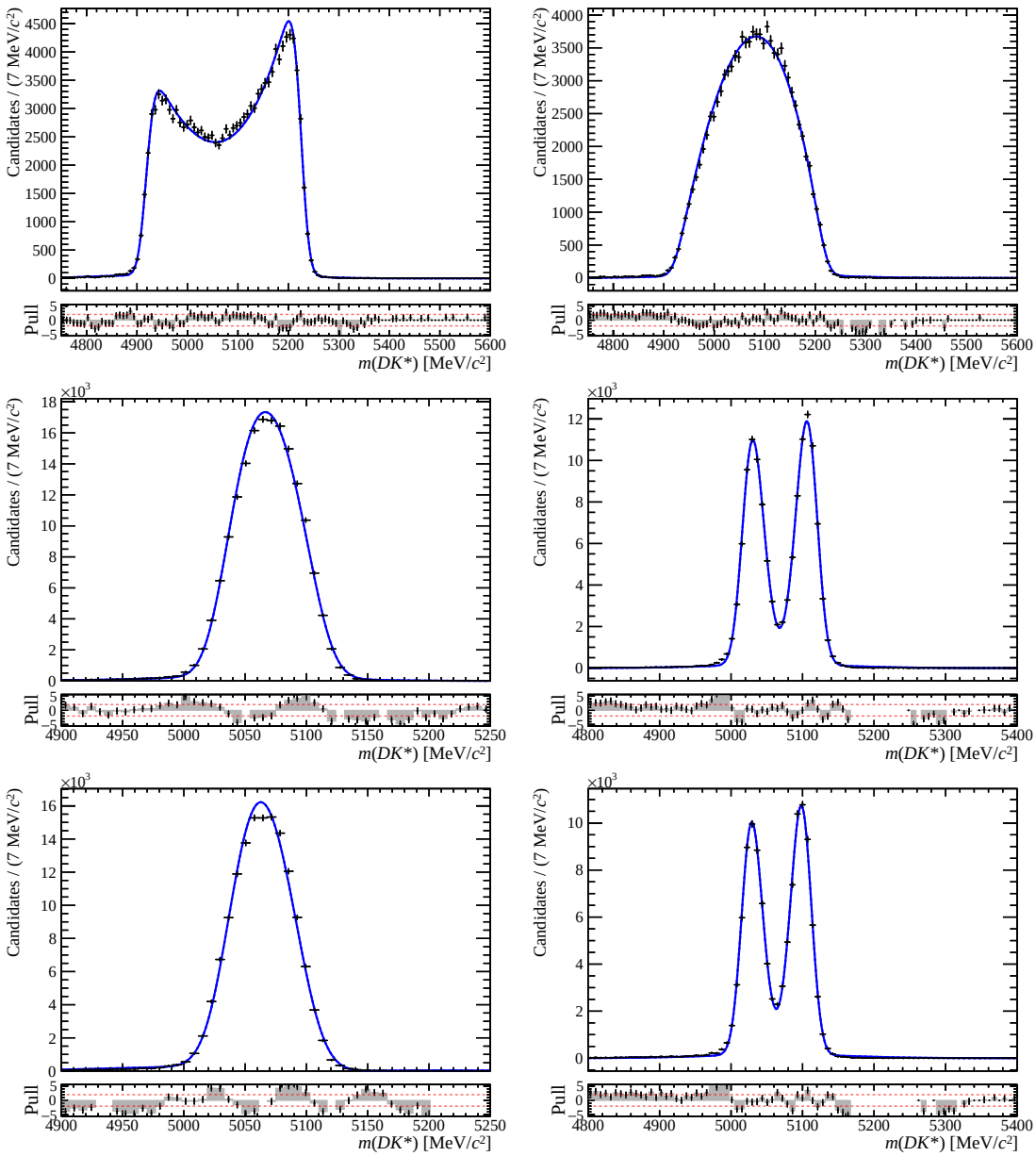


Figure 4.23: Fits to simulated data for the partially reconstructed $B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ (top), $B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ (middle) and $B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$ (bottom) backgrounds, for the two helicity states ± 1 (left) and 0 (right), using analytical PDFs. The MC samples for all years are combined, each with a weight corresponding to the integrated luminosity of that year in data.

0 and ± 1 helicity states in the following way:

$$\frac{N(B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm})}{N(B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} = \frac{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm})}{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} \times \frac{\epsilon_{sel}(B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm})}{\epsilon_{sel}(B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm})}, \quad (4.13)$$

$$\frac{N(B^0 \to D^{*\pm}(\to D^0[\pi^\pm])K^{*\mp})}{N(B^\pm \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} = \frac{\mathcal{B}(B^0 \to D^{*\pm}(\to D^0[\pi^\pm])K^{*\mp})}{\mathcal{B}(B^\pm \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} \times \frac{\epsilon_{sel}(B^0 \to D^{*\pm}(\to D^0[\pi^\pm])K^{*\mp})}{\epsilon_{sel}(B^\pm \to D^{*0}(\to D^0[\pi^0])K^{*\pm})}. \quad (4.14)$$

The selection efficiencies ϵ_{sel} are determined from simulation samples of these partially reconstructed decays and the branching fraction ratios are given by [111]:

$$R^\gamma = \frac{\mathcal{B}(B^\pm \to D^{*0}(\to D^0[\gamma])K^{*\pm})}{\mathcal{B}(B^\pm \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} = 0.55 \pm 0.02\,, \quad (4.15)$$

$$R^{\pi^\pm} = \frac{\mathcal{B}(B^0 \to D^{*\pm}(\to D^0[\pi^\pm])K^{*\mp})}{\mathcal{B}(B^\pm \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} = 0.43 \pm 0.11\,. \quad (4.16)$$

4.4.3.3 Charmless background

Another source of background comes from $B^\pm \to hh(hh)K^{*\pm}$ decays, where the B meson decays directly to the final-state particles without proceeding through a D meson, as introduced in Section 4.2.4.3. This charmless background peaks at the signal mass and is reduced by the flight distance cuts applied in the preselection. This ensures that the D meson flies significantly from the B vertex and therefore that the D meson daughters are not produced directly from the B decay, as illustrated in Figure 4.3. To study and estimate the yield of this background in each mode, the D mass sidebands are used. The idea is to estimate the charmless yield for each mode by performing a fit to the B invariant-mass distribution obtained by taking the events which are present in the D sidebands. These sidebands are defined to be [1770, 1820] and [1910, 1960] MeV/c^2, as shown in Figure 4.24 for the $K\pi$ favoured mode.

For the favoured and suppressed modes, events from both sidebands are used to estimate the charmless yield. Given that each sideband is of the size of the signal region window, the yield obtained from the fit is divided by two to extrapolate the charmless yield expected in the signal region. In the KK mode, only the low sideband is used due to the fact that the high sideband is dominated by favoured $K\pi$ decays where the pion is misidentified as a kaon. Similarly, only the high sideband is used for the $\pi\pi$ and $\pi\pi\pi\pi$ modes, the low sideband being dominated by misidentified favoured decays. In these cases where only one sideband is used, the yield obtained from the fit is directly ported to the signal region, which has the same

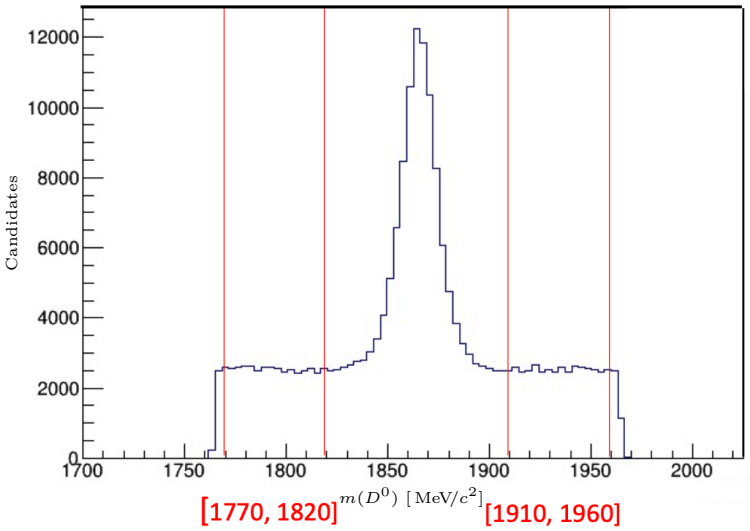


Figure 4.24: D invariant-mass distribution obtained without refitting the decay chain for the favoured $K\pi$ mode, showing the sidebands used to study the charmless background.

size as one sideband. The fits to the B invariant-mass distribution for each mode, performed separately for both charges, are shown in Figures 4.25–4.26. Table 4.17 presents the residual charmless yield in each mode obtained from the sidebands fits, separately for both charges. Negligible background is found in all modes, except from the pionic modes $\pi\pi$ and $\pi\pi\pi\pi$, where visible charmless background is seen, even after applying the flight distance cut $\Delta z^{D-B}_{\text{significance}} > 2$. This is due to the larger branching fraction of the $B^{\pm} \to \pi\pi K^{*\pm}$ decay compared to the $B^{\pm} \to K\pi K^{*\pm}$ decay. The chosen flight distance cut reduces the charmless background to a low level while maintaining a relatively high signal efficiency.

Table 4.17: Residual charmless yields obtained by fitting the D sidebands separately for each charge, after applying $\Delta z^{D-B}_{\text{significance}} > 2$.

D decay mode	B^-	B^+
$K\pi$	4.2 ± 3.5	3.6 ± 3.2
KK	0.4 ± 10.5	0.6 ± 10.9
$\pi\pi$	23.1 ± 6.4	21.5 ± 6.6
πK	0 ± 2	0 ± 2
$K\pi\pi\pi$	0 ± 28	7.6 ± 4.3
$\pi\pi\pi\pi$	12.8 ± 6.2	17.2 ± 6.8
$\pi K\pi\pi$	1.7 ± 1.5	1.0 ± 1.4

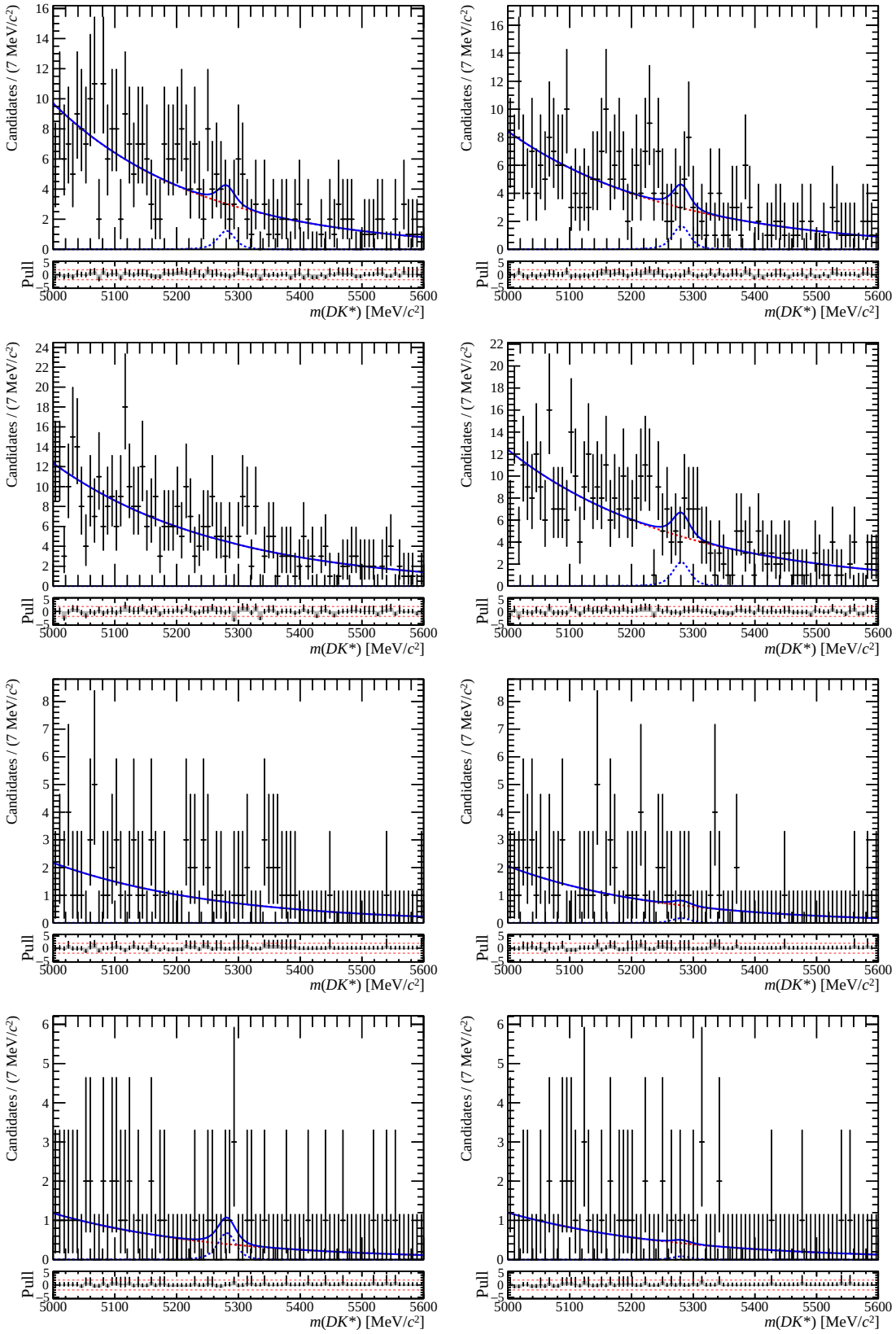


Figure 4.25: Fit to the B invariant-mass distributions, obtained by taking the candidates in the D mass sidebands for B^- (left) and B^+ (right) in the $K\pi$ (1st row), $K\pi\pi\pi$ (2nd row), πK (3rd row) and $\pi K\pi\pi$ (4th row) modes with $\Delta z^{D-B}_{\text{significance}} > 2$.

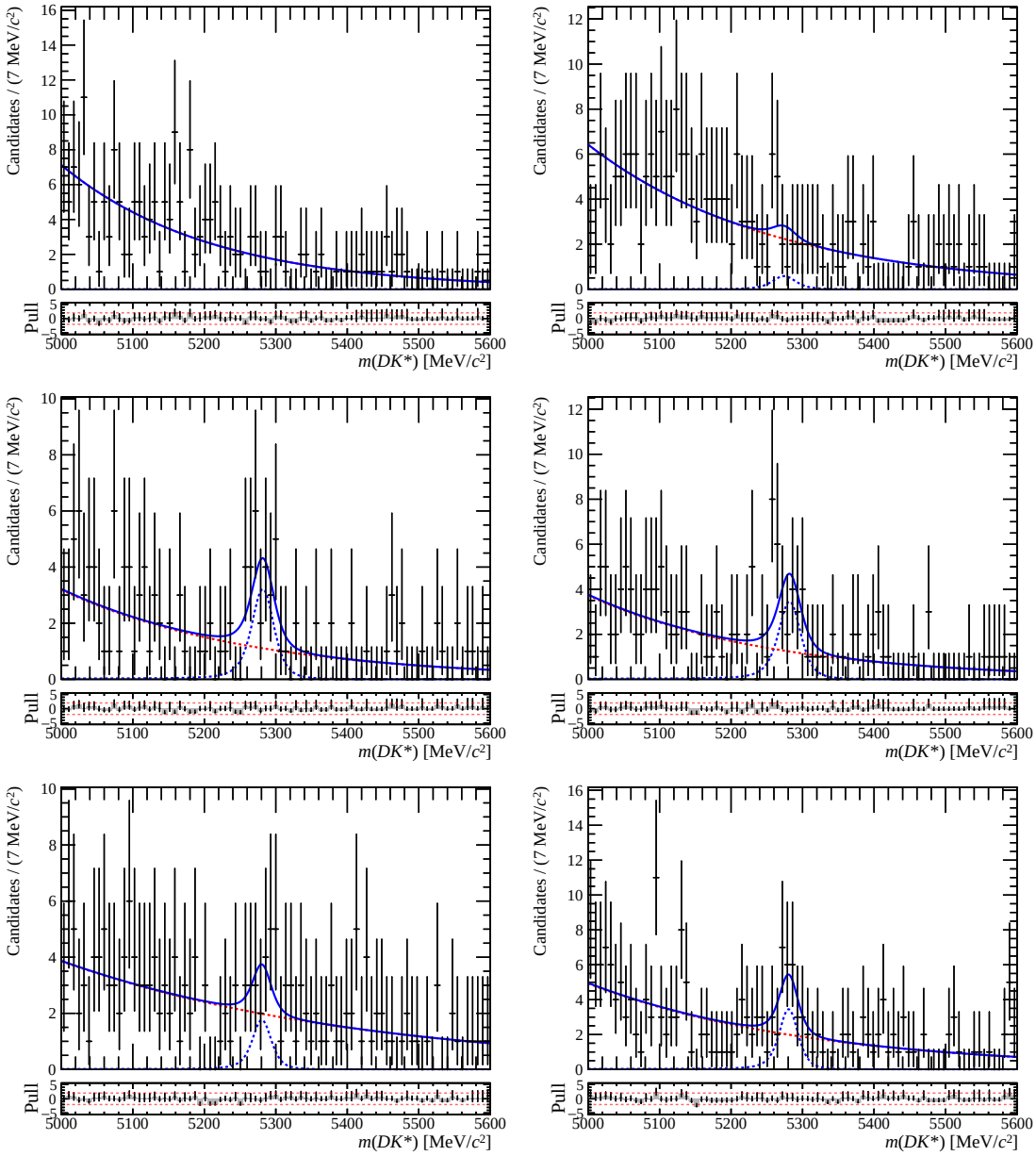


Figure 4.26: Fit to the B invariant-mass distributions, obtained by taking the candidates in the D mass sidebands for B^- (left) and B^+ (right) in the KK (top), $\pi\pi$ (middle) and $\pi\pi\pi\pi$ (bottom) modes, with $\Delta z^{D-B}_{\text{significance}} > 2$.

In order to further remove the residual charmless background, a tighter cut could be applied. However, this would strongly impact the selection efficiency. As an example, tightening the flight distance cut to a value of 4 would imply a loss of signal efficiency of about 16% and 18% for the $\pi\pi$ and $\pi\pi\pi\pi$ modes respectively. Therefore, it is decided to deal with the residual background in the *CP* fit.

This is done by adding a component to model this contribution in the $\pi\pi$ and $\pi\pi\pi\pi$ modes. The shape is obtained from simulation, by fitting the invariant-mass distribution of MC samples obtained from $B^{\pm} \to \pi\pi K^{*\pm}$ decays. The fit is shown in Figure 4.27, where the distribution is modelled using the same shape as for the signal.

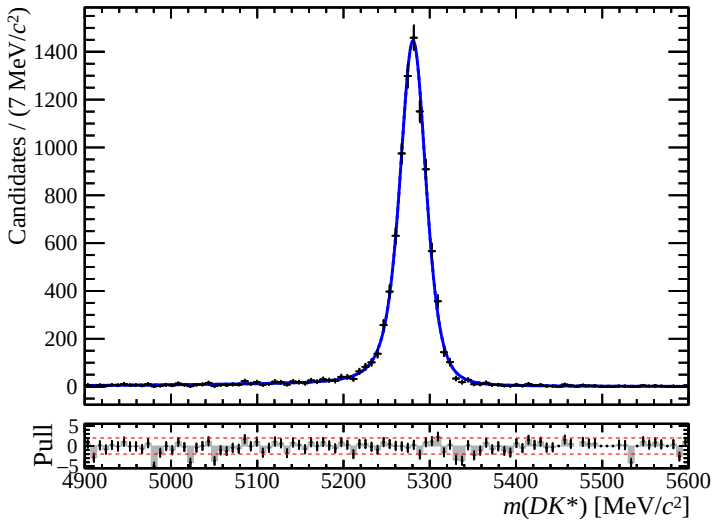


Figure 4.27: Fit to the invariant-mass distribution obtained by taking the candidates from the charmless MC sample $B^{\pm} \to \pi\pi K^{*\pm}$.

From this fit, the shape parameters are fixed and used in the *CP* fit. Concerning the mean and width of the function, these are fixed from an intermediate fit to the data where a looser $\Delta z^{D-B}_{\text{significance}} > 0.5$ cut is applied, to take into account any data-MC differences. Signal leakage coming from signal events which would sit in the D sidebands would result in an overestimation of the charmless yield. In this procedure used to estimate the charmless yield, this small leakage in the sidebands is taken into account by subtracting it from the charmless yield obtained from the baseline sidebands fit.

The charmless background is similarly studied for the BPGGSZ modes by looking at the D sidebands, defined by excluding the ± 25 MeV/c^2 signal region around the D^0 mass. Both sidebands are used for the $K^0_S\pi^+\pi^-$ and $K^0_S K^+K^-$ modes. The B invariant-mass distributions are fitted with a Gaussian for the signal and an exponential function for the background. No significant charmless background is found for both BPGGSZ modes [103].

4.4.3.4 Background from $\Lambda_b^0 \to \Lambda_c^+ K^{*-}$

A background coming from $\Lambda_b^0 \to \Lambda_c^+(\to pK[\pi])K^{*-}$ decays can contribute in the $D \to KK$ mode. This can happen when the pion coming from the Λ_c^+ is missed in the reconstruction, while the proton is misidentified as a kaon, since no proton suppression cut is used in this analysis. This background does not contribute in the other modes as the PID requirements suppress $p \to \pi$ misidentification but not $p \to K$ misidentification. The topology of this background is illustrated in Figure 4.28.

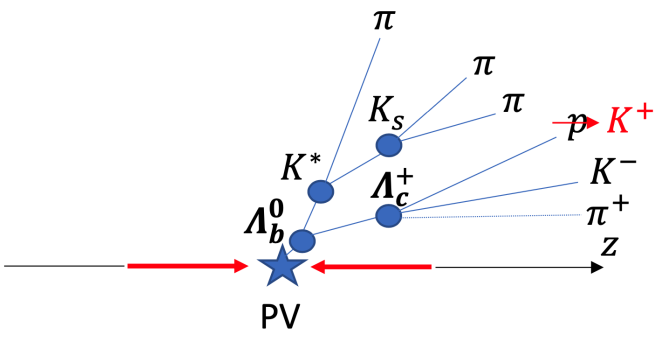


Figure 4.28: Topology of the $\Lambda_b^0 \to \Lambda_c^+(\to pK[\pi])K^{*-}$ decay. This is a background contributing in the KK mode, where the same final state is reached when the proton is misidentified as a kaon and the pion coming from the Λ_c^+ is missed in the reconstruction.

In order to study this background, MC samples of simulated $\Lambda_b^0 \to \Lambda_c^+ K^{*-}$ events are generated. These simulation samples are used to estimate the expected number of events from this background in the data, as well as to obtain the shape of this background. The latter is obtained from a fit to the invariant-mass distribution of these simulated events, shown in Figure 4.29. To model this background, a

Cruijff function is used:

$$\text{Cruijff}(m;\,\mu,\sigma_L,\sigma_R,\alpha_L,\alpha_R,) \propto \begin{cases} \exp\left[\frac{-(m-\mu)^2}{2\sigma_L^2+\alpha_L(m-\mu)^2}\right] & \text{if } m-\mu<0 \\ \exp\left[\frac{-(m-\mu)^2}{2\sigma_R^2+\alpha_R(m-\mu)^2}\right] & \text{if } m-\mu>0\,, \end{cases} \tag{4.17}$$

and the results for the parameters obtained from the fit are listed in Table 4.18.

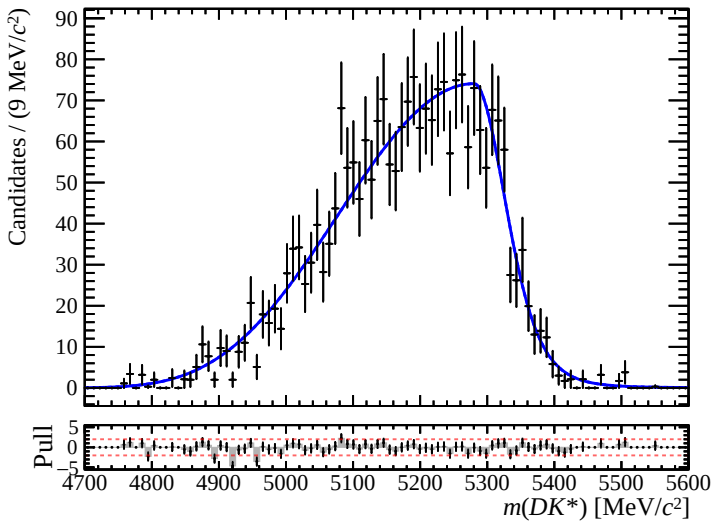


Figure 4.29: Invariant-mass distribution of simulated events for the $\Lambda_b^0 \to \Lambda_c^+ K^{*-}$ background with the fit result superimposed using a Cruijff function.

Table 4.18: Fit parameter results obtained from the fits to simulated data for the $\Lambda_b^0 \to \Lambda_c^+ K^{*-}$ background.

μ	σ_L	σ_R	α_L	α_R
5280.27 ± 9.30	195.77 ± 12.83	45.91 ± 6.63	-0.094 ± 0.042	0.108 ± 0.029

Concerning the yield of this background $N_{\Lambda_b^0}$, this is estimated using the following equation:

$$N_{\Lambda_b^0} = N_{B^+} \times \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ K^{*-})}{\mathcal{B}(B^+ \to DK^{*+})} \times \frac{\mathcal{B}(\Lambda_c^+ \to pK\pi)}{\mathcal{B}(D \to KK)} \times f_{\Lambda_b^0/B^+} \times \epsilon_{\Lambda_b^0/B^+}, \tag{4.18}$$

where N_{B^+} is the signal yield in the KK mode, $\epsilon_{\Lambda_b^0/B^+}$ represents the efficiency ratio between this background and the signal decay and $f_{\Lambda_b^0/B^+}$ is the ratio of production fraction taken from Reference [112]. Taking the branching fractions of

the decays from the PDG [28], the number of expected events obtained is presented in Table 4.19. The corresponding ratio of this background with respect to the signal $D \to KK$ events is also shown, which is found to be around 2%. To take into account this background, a component is added to the model for the KK mode in the *CP* fit.

Table 4.19: Expected yield from $\Lambda_b^0 \to \Lambda_c^+ K^{*-}$ in the $D \to KK$ mode.

D decay mode	$N_{\Lambda_b^0}/N_{B^+}$	$N_{\Lambda_b^0}$
KK	0.021 ± 0.011	12.88 ± 7.14

4.4.3.5 Background from $B_s^0 \to \bar{D}^0 \bar{K}^{*0}$

Backgrounds from B_s decays become dominant in the suppressed modes of this measurement, given the low signal yield in these modes. These backgrounds must be studied and considered since these could potentially leak and contribute into the signal region. These B_s backgrounds are of the type $B_s^0 \to \bar{D}^0 \bar{K}^{**0}$, with $\bar{K}^{**0} \to K^{*-}[\pi^+]$ and $K^{*-} \to K_S^0 \pi^-$, where the pion from the $\bar{K}^{**0}$ decay is missed and not reconstructed. These decays appear in the low mass region where partially reconstructed backgrounds are present. To study this background in detail, fast simulation [97] is used to generate samples of these backgrounds for five different $\bar{K}^{**0}$ resonances:

- $B_s^0 \to \bar{D}^0 \bar{K}^*(1410)^0$
- $B_s^0 \to \bar{D}^0 \bar{K}_1(1400)^0$
- $B_s^0 \to \bar{D}^0 \bar{K}_1(1270)$
- $B_s^0 \to \bar{D}^0 \bar{K}_0^*(1430)^0$
- $B_s^0 \to \bar{D}^0 \bar{K}_2^*(1430)^0$.

These samples are used to model the distribution of this B_s background, for which the tail could leak into the signal region of the suppressed modes. The shapes of the five individual samples for the different resonances are presented in Figure 4.30 (left).

These are obtained from individual fits to the simulated samples. Figure 4.30 (right) shows the average invariant-mass distribution of these B_s decays, with the fit result superimposed. In this fit, the distribution is modelled using a kernel estimation PDF, making use of a superposition of Gaussian functions at each point contributing to the total integral with the same weight. A systematic uncertainty is considered in the *CP* fit for the shape used to model the B_s background, discussed in Section 6.5.

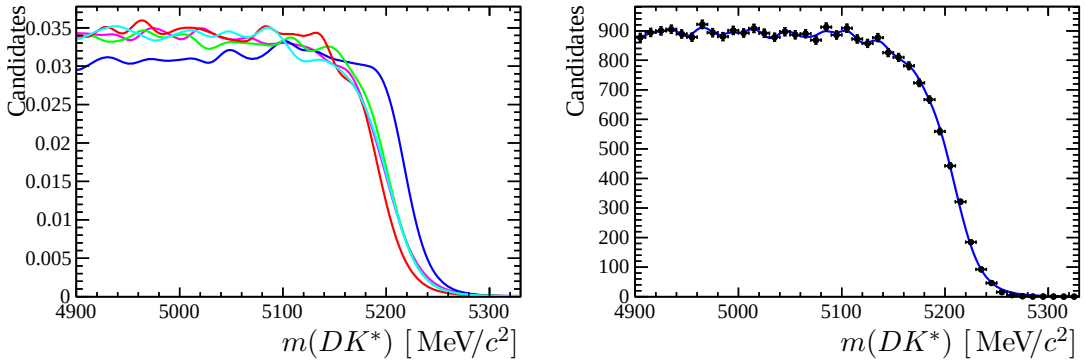


Figure 4.30: Shapes of the five resonances obtained by fitting individually each generated sample (left), and invariant-mass distribution of the average of the five simulated samples for the B_s background with the fit result superimposed (right).

The goal is then to determine the yield of this background in the signal region. This is done by studying the invariant-mass distribution of the suppressed πK mode, using the full dataset as shown in Figure 4.31 (left). In this plot, the signal region is blinded. The idea is to fit the B_s background distribution in the low-mass region, to the left of the signal region, in order to be able to extrapolate into the signal region. To this aim, the expected number of combinatorial and partially reconstructed backgrounds is subtracted from the data in the low-mass region, on a bin by bin basis.

To determine the number of combinatorial events, a fit is performed to the data in the high-mass sideband, to the right of the signal region of the suppressed mode. By extrapolating the fit to the low-mass region, the number of combinatorial events in each bin can be determined. Concerning the partially reconstructed background, the interference in the partially reconstructed region is determined by using the same values of the physics parameters in the following equation:

$$R^{\pm}_{\pi K} = \frac{N(B^{\pm} \to [\pi^{\pm}K^{\mp}]_D K^{*\pm})}{N(B^{\pm} \to [K^{\pm}\pi^{\mp}]_D K^{*\pm})} \times \frac{\epsilon_{sel}(K\pi)}{\epsilon_{sel}(\pi K)} \times \frac{1}{\epsilon_{veto}(\pi K)}. \tag{4.19}$$

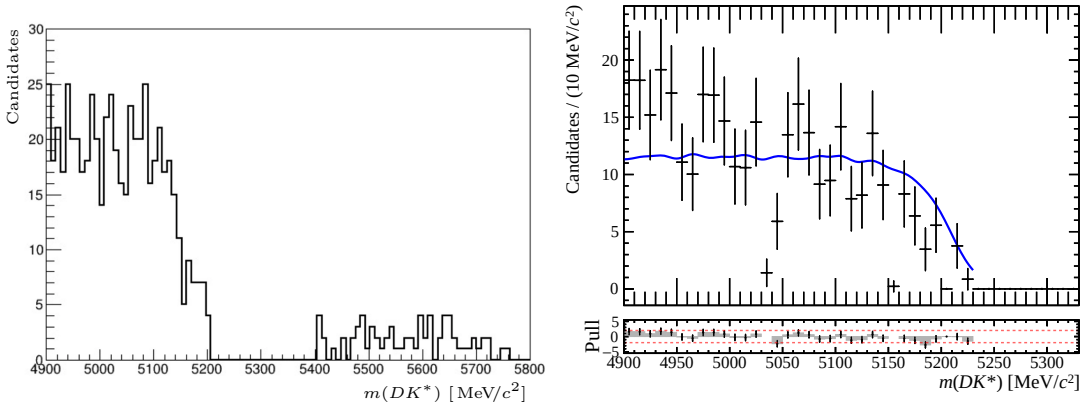


Figure 4.31: Invariant-mass distribution for the suppressed πK mode, obtained with the full dataset and where the signal peak is blinded (left), and the distribution obtained in the low-mass region after subtracting the expected combinatorial and partially reconstructed background events, with the fit result superimposed (right).

The number of partially reconstructed events in bin i in the suppressed mode, $N^i_{pr,\pi K}$, can then be computed as

$$N^i_{pr,\pi K} = N^i_{pr,K\pi} \times R_{\pi K} \times \frac{\epsilon_{sel}(\pi K)}{\epsilon_{sel}(K\pi)} \times \epsilon_{veto}(\pi K), \tag{4.20}$$

where $N^i_{pr,K\pi}$ is the number of partially reconstructed events in bin i for the $K\pi$ mode. The assumption is made that the resulting distribution after subtracting the combinatorial and partially reconstructed backgrounds from the low-mass region is attributed to the $B^0_s \to \bar{D}^0 \bar{K}^{*0}$ background. Using the PDF obtained from simulation, a fit is performed in the low-mass region, as shown in Figure 4.31 (right). By extrapolating into the signal region, the expected yield N_{B_s} is found to be less than 3 events in total, for both charges, as presented in Table 4.20.

Table 4.20: Expected yield from B_s backgrounds in the suppressed mode.

D decay mode	N_{B_s}
πK	2.51 ± 1.59

A component is added to the mass model for the suppressed modes in the *CP* fit in order to control this background. The yield of this background for each charge is fixed to be half of the determined total expected yield, and the shape is also fixed

in the final fit. The yield of this B_s background in the four-body suppressed $\pi K\pi\pi$ mode is fixed through the ratio of signal events $N_{K\pi\pi\pi}/N_{K\pi}$ in the favoured modes.

4.4.4 Additional background studies

This section presents additional background studies performed to consider other sources of background that could contribute in the analysis. These are not included in the *CP* fit as they are found to not be present in the analysis or reduced to a negligible level.

4.4.4.1 Misidentified $B^\pm \to DK^0_S K^\pm$ background

Background coming from $B^\pm \to DK^0_S K^\pm$ decays, where the kaon is misidentified as a pion, could contribute in the analysis. To study this background, MC samples of this decay are generated. The corresponding invariant-mass distribution, obtained after applying the full signal selection except from the PID requirements, is shown in Figure 4.32 (left). The selection, especially the K^* selection requirements, efficiently suppresses this background. Retaining only the events within the *CP* fit range, above 5230 MeV/c^2, it is found that there is about 7% of this background which remains in the range. Figure 4.32 (right) is obtained after applying the PID requirement on the companion (misidentified) kaon, which further reduces this background.

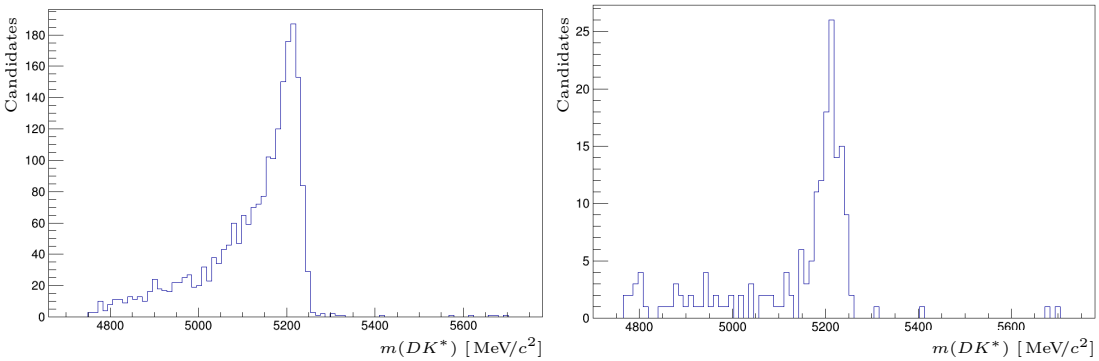


Figure 4.32: Distribution of $B^\pm \to DK^0_S K^\pm$ events reconstructed after the full selection without PID requirements (left) and after the PID requirement on the bachelor pion (right).

An estimation of the number of expected events from this background is given by

$$N_{K\pi}^{DK^0_S K^\pm} = N_{K\pi}^{sig} \times \frac{\mathcal{B}(B^\pm \to DK^0_S K^\pm)}{\mathcal{B}(B^\pm \to DK^{*\pm})} \times \frac{\epsilon_{DK^0_S K^\pm}}{\epsilon_{DK^{*\pm}}}, \tag{4.21}$$

where $N_{K\pi}^{sig}$ is the signal yield in the favoured $K\pi$ mode and $\epsilon_{DK_S^0K^{\pm}}/\epsilon_{DK^{*\pm}}$ is the total efficiency ratio of this background with respect to the signal. The branching ratio of this background decay $\mathcal{B}(B^{\pm} \to DK_S^0K)$ is comparable to the branching ratio of the signal decay $\mathcal{B}(B^{\pm} \to DK^{*\pm})$ [28]. This gives an expected number of events of 1.8 ± 0.6 coming from this misidentified $B^{\pm} \to DK_S^0K^{\pm}$ background, corresponding to about 0.03% of the signal yield. This background is therefore considered to be negligible and not included in the *CP* fit.

4.4.4.2 Track-swap backgrounds

Finally, the last source of background investigated is track-swap backgrounds. These correspond to decays where the final-state tracks are swapped in the reconstruction. To study these backgrounds, the invariant-mass distributions are reconstructed by swapping tracks from the final-state particles in order to check for any peaks that would be present. The following cases are considered:

- The K_S^0 particle is swapped with a $\pi^+\pi^-$ pair from the D meson.
- The bachelor pion is swapped with a pion of the same charge from the D meson.

The invariant-mass distributions are reconstructed in each case by performing the combinations of tracks in the four-body modes. For $B^{\pm} \to [K^{\pm}\pi^{\mp}\pi^{\pm}\pi^{\mp}]_D K_S^0\pi^{\pm}$ and $B^{\pm} \to [\pi^{\pm}K^{\mp}\pi^{\pm}\pi^{\mp}]_D K^{*\pm}$ modes, these are presented in Figure 4.33, while the invariant-mass distributions for the $B^{\pm} \to [\pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^{\mp}]_D K_S^0\pi^{\pm}$ mode are shown in Figure 4.34. These backgrounds are studied in a similar way for the BPGGSZ modes. In any of these distributions, no evident peak is found and therefore no track-swap background is found to be present in the selected data.

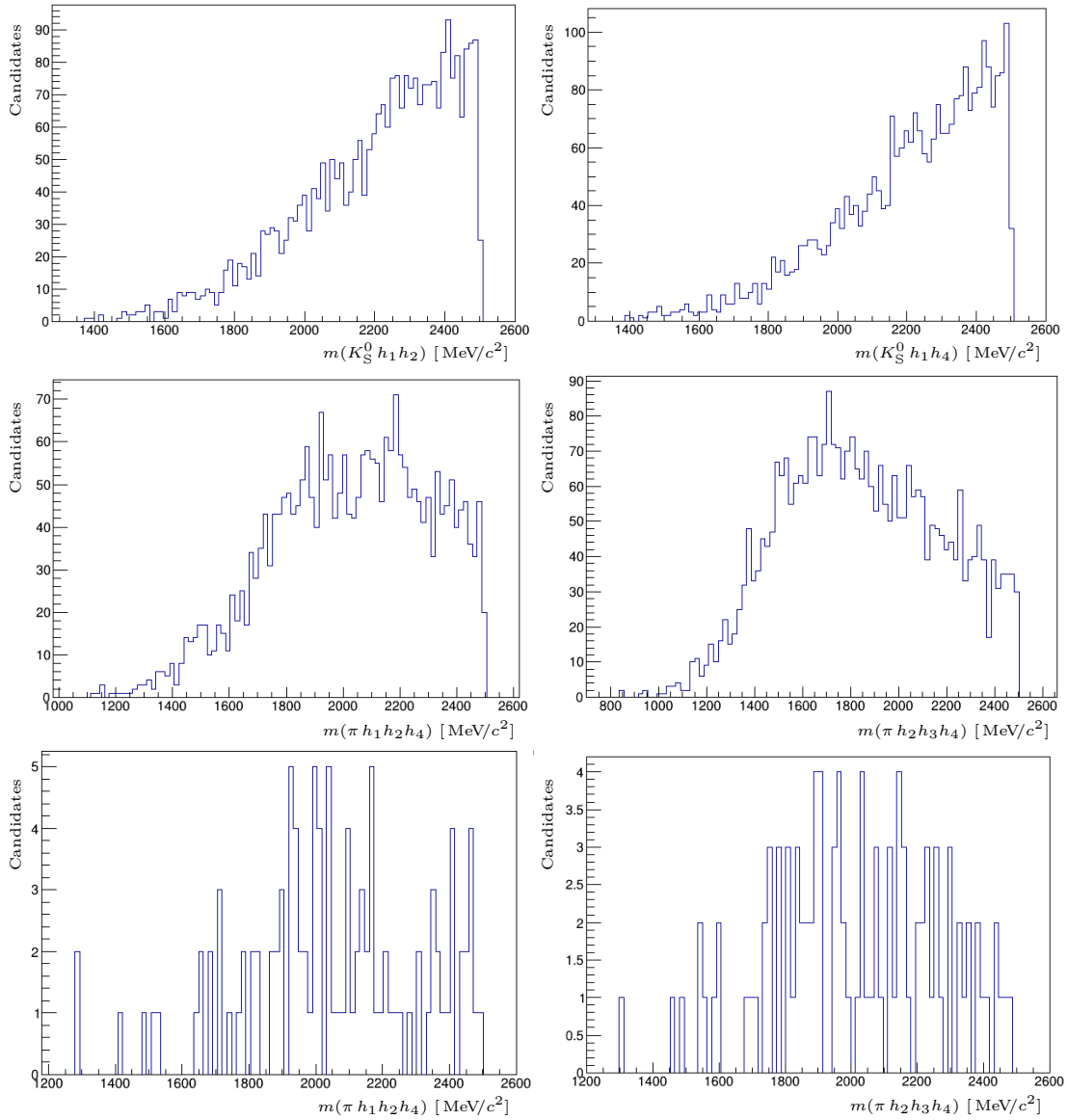


Figure 4.33: Invariant-mass distributions in data obtained by swapping tracks. Top: in $B^\pm \to [K^\pm\pi^\mp\pi^\pm\pi^\mp]_D K^0_S\pi^\pm$ where the K^0_S is swapped with a $\pi^+\pi^-$ pair from the D meson. Middle: in $B^\pm \to [K^\pm\pi^\mp\pi^\pm\pi^\mp]_D K^0_S\pi^\pm$ where the bachelor pion is swapped with a D daughter particle of the same charge. Bottom: in $B^\pm \to [\pi^\pm K^\mp\pi^\pm\pi^\mp]_D K^{*\pm}$ where the bachelor pion is swapped with a D daughter pion of the same charge.

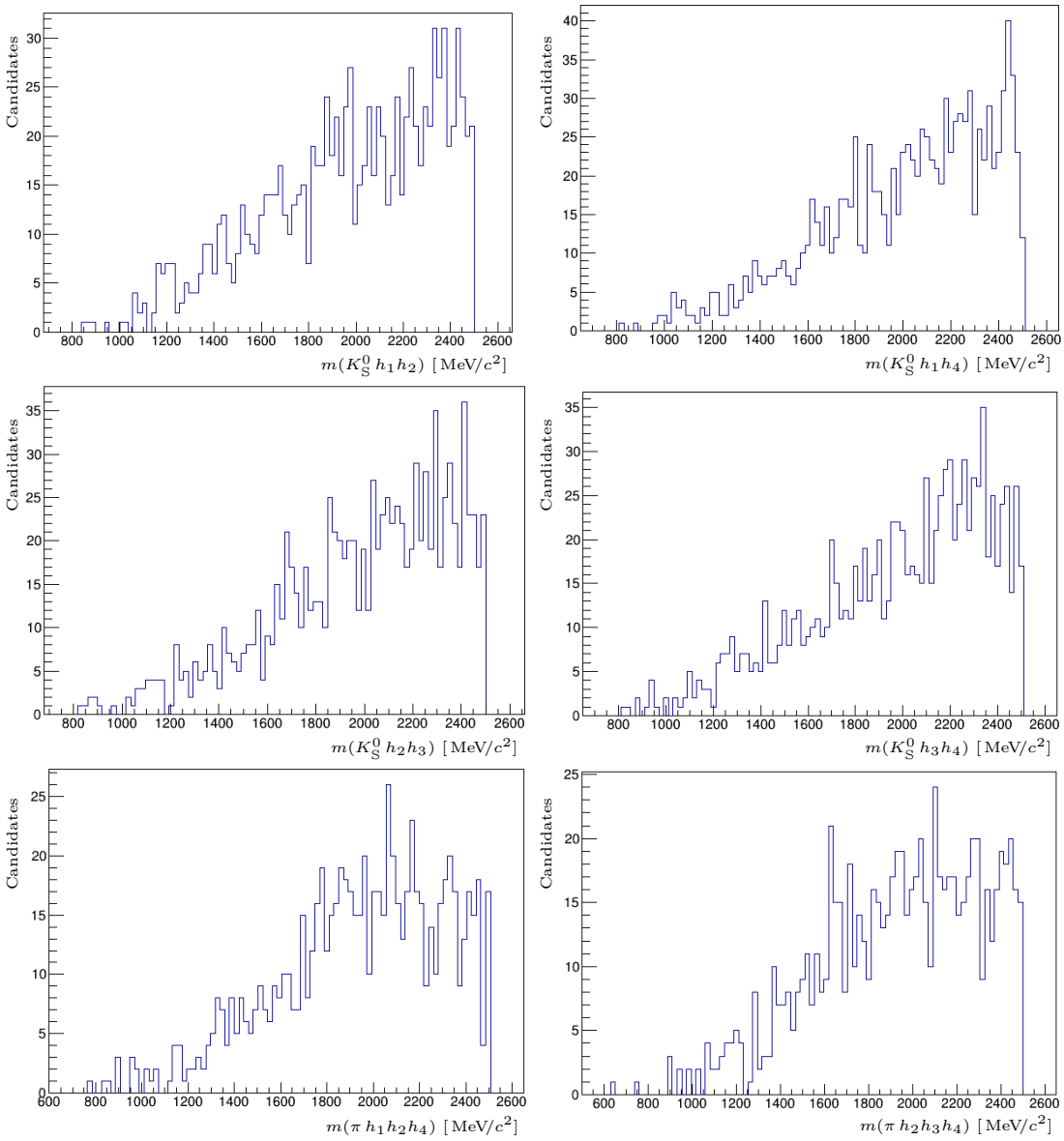


Figure 4.34: Invariant-mass distributions in data obtained by swapping tracks in $B^{\pm} \to [\pi^{\pm}\pi^{\mp}\pi^{\pm}\pi^{\mp}]_D K^0_S \pi^{\pm}$. Top and middle: the K^0_S is swapped with a $\pi^+\pi^-$ pair from the D meson. Bottom: the bachelor pion is swapped with a D daughter pion of the same charge.

4.5 Summary

This chapter presented the candidate selection, which efficiently removed and suppressed important backgrounds in the analysis. The selection efficiencies were computed, which are taken into account in the *CP* fit. The irreducible backgrounds which remain in the data sample in all D decay modes are:

- Combinatorial background
- Partially reconstructed $B \to D^* K^*$ backgrounds.

Additional backgrounds are found in specific modes:

- $B^\pm \to \pi\pi(\pi\pi) K^{*\pm}$ charmless background in $D \to \pi\pi(\pi\pi)$ modes
- $\Lambda_b^0 \to \Lambda_c^+ (\to pK[\pi]) K^{*-}$ background in the $D \to KK$ mode
- $B_s^0 \to \bar{D}^0 \bar{K}^{*0}$ background in the suppressed $D \to \pi K(\pi\pi)$ modes.

The invariant-mass shape used to model the signal was presented in this chapter, as well as the shapes and expected yields for the irreducible backgrounds, which are included in the *CP* fit.

I think that the discovery of antimatter was perhaps the biggest jump of all the big jumps in physics in our century.

— Werner Heisenberg

5

Sensitivity study of the γ measurement in $B^{\pm} \to DK^{*\pm}$ decays using the $D \to K^0_S h^+ h^-$ channels

This chapter summarises the sensitivity study that was performed at the beginning of the analysis. The goal is to determine the expected uncertainty on the measurement of γ for the BPGGSZ modes in the $B^{\pm} \to DK^{*\pm}$ channel, which is measured for the first time at LHCb in this decay channel. This allows the check that the analysis is valuable to perform, as well as to explore the fit setups and stability of the fitting procedure. This was performed using a preliminary version of the selection and invariant-mass fit. This sensitivity study predates the final selection presented in Chapter 4, and the final fit to data used in the measurement which is described in Chapter 6.

5.1 Introduction

As the BPGGSZ modes have not been studied at LHCb in the $B^{\pm} \to DK^{*\pm}$ channel, the first step was to perform sensitivity projections for the measurement of γ using these decays. This study focuses on the $D \to K^0_S \pi^+ \pi^-$ final state, which contributes

about 90% of the total number of candidates obtained from both BPGGSZ modes of interest $D \to K^0_S\pi^+\pi^-$ and $D \to K^0_S K^+ K^-$. The objective is to check that this is a valuable measurement to perform. Previous results for this mode have been published by the Belle and BaBar collaborations [113, 114].

The sensitivity study follows the BPGGSZ method explained in Section 2.6.4. The kinematics of the three-body final state $D \to K^0_S\pi^+\pi^-$ can be fully represented in two dimensions in the Dalitz plot. The common decay $D \to K^0_S\pi^+\pi^-$ gives rise to interference, which appears as different distributions of the Dalitz plots of the D meson for B^- and B^+ decays. The CP-conjugate of the $D^0 \to K^0_S\pi^+\pi^-$ decay corresponds to the $\overline{D}^0 \to K^0_S\pi^-\pi^+$ decay. Since the two charged pions have swapped momenta, the CP-conjugate of a given state in the phase space for B^- decays corresponds to a state which is symmetric with respect to the $m^2_- = m^2_+$ axis in the phase space for B^+ decays. This is the reason why the axes are exchanged for the Dalitz plots of B^- and B^+ decays, as illustrated in Figure 5.1 which shows the Dalitz plots obtained from data using the full Run 1 and Run 2 dataset for B^- (left) and B^+ (right), and where the quantity of data is not sufficient to see the differences due to CP violation.

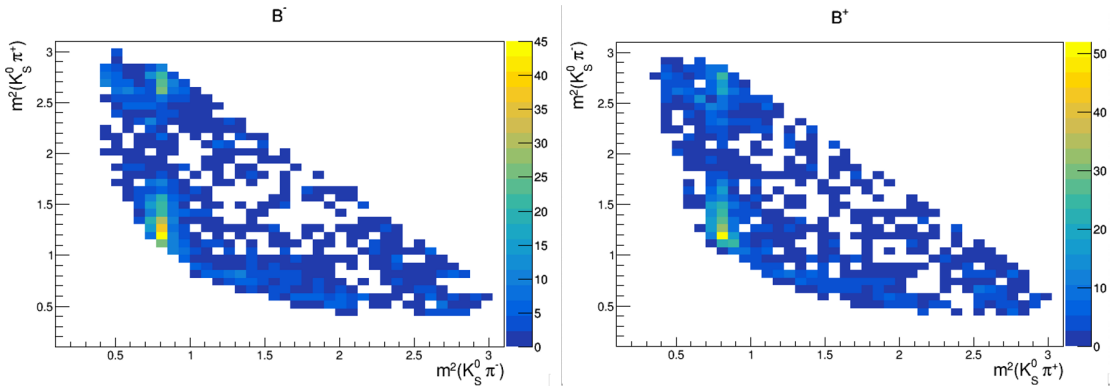


Figure 5.1: Dalitz plot of the D meson in data for B^- (left) and B^+ (right) obtained from the full dataset.

In order to observe these differences the Dalitz plots are binned. The optimal binning scheme with 16 bins is used, for which the CP observables $x_\pm$, $y_\pm$ have a maximised expected statistical sensitivity. In this way, the measurement is reduced to a counting experiment in each Dalitz bin and the differences can be

interpreted in terms of SM considerations by expressing the yields in each bin as in Equations (2.44)–(2.45) and repeated here:

$$\begin{aligned} N_i^- &= h_B^-[F_i + (x_-^2 + y_-^2)F_{-i} + 2\kappa\sqrt{F_i F_{-i}}(c_i x_- + s_i y_-)], \\ N_i^+ &= h_B^+[F_{-i} + (x_+^2 + y_+^2)F_i + 2\kappa\sqrt{F_i F_{-i}}(c_i x_+ - s_i y_+)]. \end{aligned} \tag{5.1}$$

The *CP* observables $x_\pm$, $y_\pm$ are then directly related to the physics parameters of interest (r_B, δ_B and γ):

$$x_\pm + iy_\pm = r_B \cdot e^{i(\delta_B \pm \gamma)} \Rightarrow \begin{cases} x_\pm = r_B \cos(\delta_B \pm \gamma) \\ y_\pm = r_B \sin(\delta_B \pm \gamma). \end{cases} \tag{5.2}$$

In Equations (5.1), the c_i, s_i and F_i coefficients are fixed from external measured constraints, while the coherence factor κ is assumed to be equal to 1 for these studies. The c_i, s_i coefficients are directly measured at CLEO [50] and BESIII [51, 52], while the F_i values are taken from the $B^\pm \to D(\to K^0_S h^\pm h^\mp)h^\pm$ channel at LHCb [49]. This leaves 6 parameters of interest to be determined by the fit, which are the *CP* observables $x_\pm$, $y_\pm$, as well as the normalisation factors $h_B^\pm$.

5.2 Candidate selection

The main goal of the selection of the candidates is to have a good estimate of the signal and background in order to obtain the number of signal candidates in the dataset. The candidate selection for the sensitivity study predates and is similar to the final selection described in detail in Chapter 4. Since the F_i values are shared from the $B^\pm \to Dh^\pm$ analysis, the selection of the candidates is based on that analysis in order to maintain a similar efficiency across the Dalitz plot. In particular, the BDT trained in that analysis was used in the channel analysed in this work. Figure 5.2 summarises the selection by showing the invariant-mass distribution of the candidates before and after the selection is applied to the raw data. The combinatorial background is efficiently reduced by the BDT, while partially reconstructed backgrounds remain at the low-mass region, to the left of the signal peak.

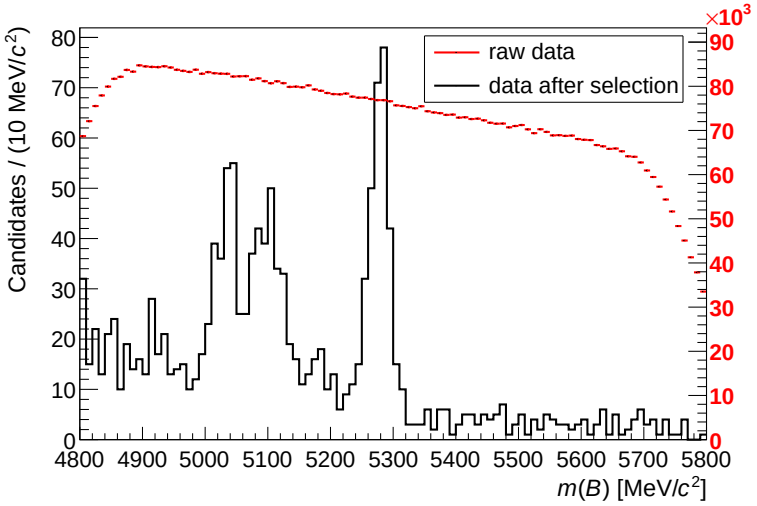


Figure 5.2: Invariant-mass distribution before and after the selection shown in two adapted scales.

While the BDT is taken from the $B^{\pm} \to Dh^{\pm}$ analysis, the cut used for this study is optimised in order to minimise the expected statistical uncertainty on the γ measurement. In this case, the significance is optimised, since the optimal point is expected to be close to the maximum significance. This figure of merit corresponds to $S/\sqrt{S+B}$, where S and B are the number of signal and background events in the signal region., which are obtained from a fit to the B invariant-mass distribution presented in Section 5.3.

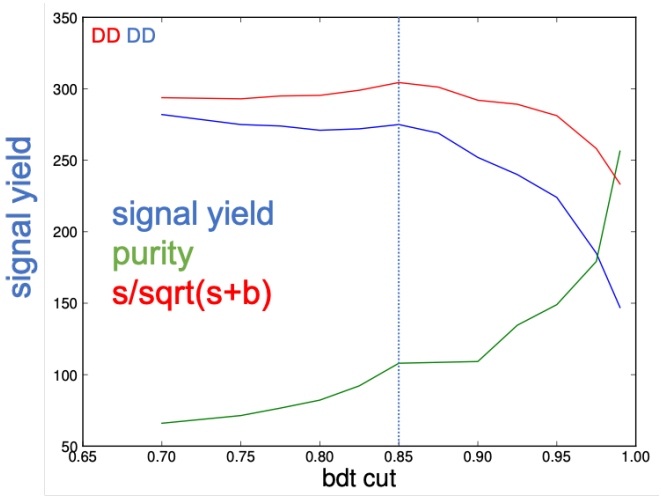


Figure 5.3: Optimisation of the BDT cut by performing a scan of BDT values.

This optimisation is performed by scanning different values of the BDT cut, as

shown in Figure 5.3. The chosen working point is highlighted with the vertical dotted line, corresponding to a BDT cut value of 0.85. It can be seen that the signal yield at this working point is of the same order as looser BDT cut values, while the signal yield drops quickly for tighter BDT cuts. The purity S/B increases for higher BDT cuts as more background is removed.

5.3 Invariant-mass fit

The signal is well separated from the complex structure of the partially reconstructed background and therefore the invariant-mass fit is performed in a reduced range from 5230 MeV/c^2. Two categories are considered for the candidates, signal and combinatorial background, and each component is modelled by a probability density function (PDF). The combinatorial background is simply modelled by a first-order polynomial with a slope which is left to be determined by the data. For the signal, the shape is studied using simulation samples of $B^{\pm} \to D(\to K^0_S \pi^+ \pi^-) K^{*\pm}$ decays. The signal PDF is made by the sum of two Crystal Ball (CB) functions [115], which consists of a modified Gaussian function where the tails are power law distributions:

$$CB(m;\, \mu, \sigma, \alpha, n) \propto \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2\right] & \text{if } \frac{m-\mu}{\sigma} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \cdot \left(\frac{n}{|\alpha|} - |\alpha| - \left(\frac{m-\mu}{\sigma}\right)\right)^{-n} & \text{otherwise.} \end{cases} \tag{5.3}$$

The two CB functions share the same tail parameters α and n, which are fixed from a fit to MC samples which include a full description of the LHCb detector. The mean μ and width σ are floating in the fit and determined by the data, which allows for differences between data and simulation. The mean μ is shared between the two CB shapes. The invariant-mass distribution with the fit result superimposed is shown in Figure 5.4 for one of the four categories defined by the track types of the pions used to reconstruct the two K^0_S particles in the decay chain.

By summing the yields obtained in each category, the total signal yield is found to be 547 events for the $D \to K^0_S \pi^+ \pi^-$ mode. This number of candidates is compared to the number of candidates obtained in the previous results from

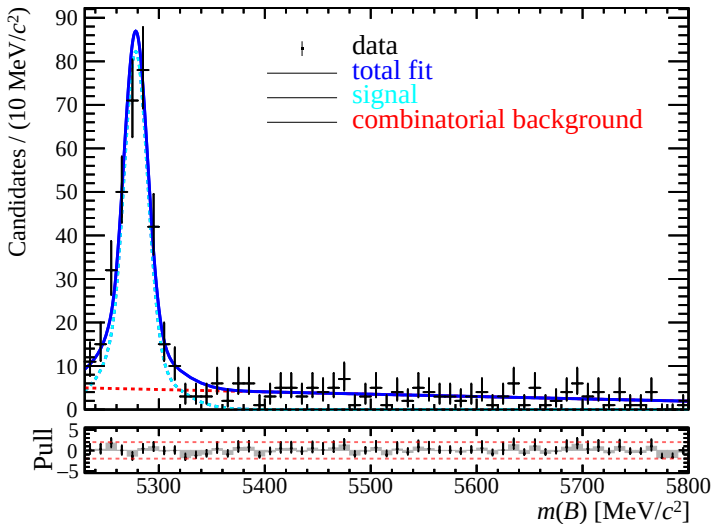


Figure 5.4: Invariant-mass distribution and fit to data in the reduced range. The distribution is shown here for one category of the K^0_S mesons, and the fit is composed of the signal and the combinatorial background components.

B-factories, corresponding to 54 candidates obtained by Belle with 357 fb^{-1} of data and 163 events obtained by BaBar. In this work with the full Run 1 and Run 2 dataset from LHCb, there are around 10 times more candidates than the Belle measurement and more than 3 times the number of events obtained by BaBar.

5.4 Sensitivity study

A sensitivity study is then performed to explore the feasibility and sensitivity of the measurement of γ using $B^\pm \to D(\to K^0_S\pi^+\pi^-)K^{*\pm}$ decays. This is achieved by determining the expected statistical uncertainties on the CP observables, and in turn on γ. Equations (5.1), repeated here with $\kappa = 1$, can be rewritten in a simplified way as:

$$\begin{aligned} N_i^- &= h_B^-[F_i + (x_-^2 + y_-^2)F_{-i} + 2\sqrt{F_i F_{-i}}(c_i x_- + s_i y_-)] \\ N_i^+ &= h_B^+[F_{-i} + (x_+^2 + y_+^2)F_i + 2\sqrt{F_i F_{-i}}(c_i x_+ - s_i y_+)], \end{aligned} \tag{5.4}$$

$$\begin{aligned} N_i^- &= h_B^- Y_i^- \\ N_i^+ &= h_B^+ Y_i^+. \end{aligned} \tag{5.5}$$

The normalisation coefficients h_B^-, h_B^+ can be written in terms of the total number of events for each charge N_{total}^-, N_{total}^+ as $h_B^{\pm} = N_{total}^{\pm} / \sum_i Y_i^{\pm}$:

$$\begin{aligned} N_i^- &= \frac{N_{total}^-}{\sum_i Y_i^-} Y_i^-, \\ N_i^+ &= \frac{N_{total}^+}{\sum_i Y_i^+} Y_i^+. \end{aligned} \tag{5.6}$$

The terms $Y_i^{\pm}$ contain the *CP* observables $x_{\pm}$, $y_{\pm}$ to be measured, as well as the fixed inputs c_i, s_i and F_i.

5.4.1 Expected yields

The expected yield in each Dalitz bin can therefore be computed from the total number of events for each charge and the *CP* observables. The total yield of about 550 events obtained from the invariant-mass fit to the data is used and equally divided between the two charges, such that $N_{total}^{\pm} = 275$. The values of the *CP* observables $x_{\pm}$, $y_{\pm}$ are determined using Equations (5.2) from fixed values of the physics parameters, $r_B = 0.1$, $\delta_B = 40°$ and $\gamma = 72°$ [6]. The expected yields in each bin N_i^- and N_i^+ are computed in this way for both B^- and B^+ decays, respectively.

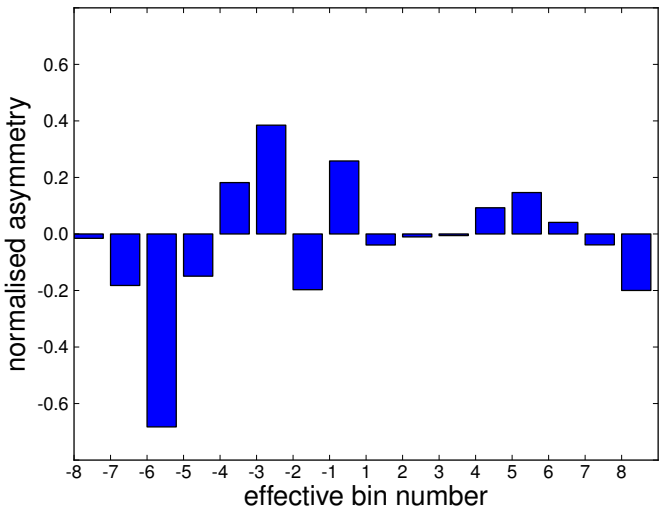


Figure 5.5: Normalised expected asymmetry in each Dalitz bin.

In order to visualise the expected distribution of *CP* violation, a normalised asymmetry in each bin is computed, where an effective bin number is defined as i for B^+ and $-i$ for B^-. This normalised asymmetry is defined for this study

as the difference between the number of candidates of each charge, divided by the total number of candidates in the considered pair of bins:

$$\text{normalised asymmetry} = \frac{N_{B^+} - N_{B^-}}{N_{B^+} + N_{B^-}}. \tag{5.7}$$

In the absence of *CP* violation, the asymmetry is expected to be zero. Figure 5.5 presents the normalised expected asymmetry for each effective bin number, where large visible asymmetries are present in some of the bins. This shows how *CP* violation manifests itself in these decays.

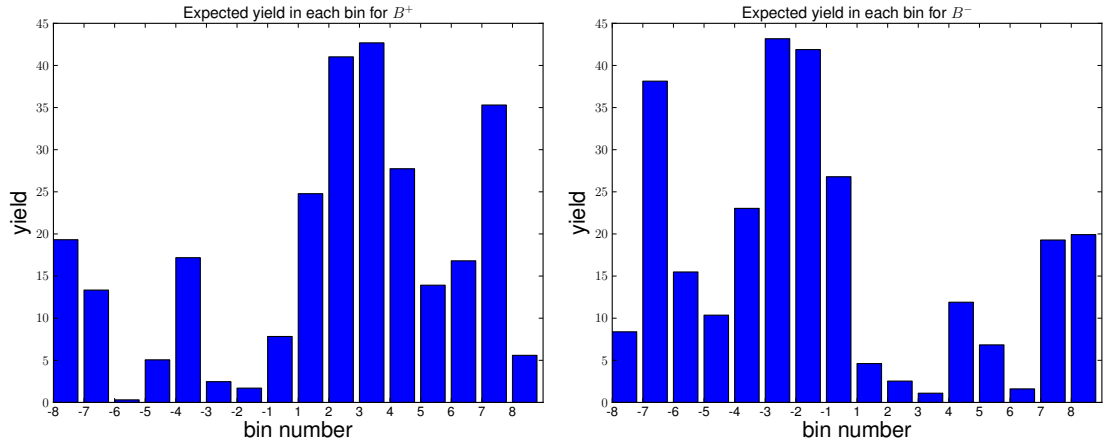


Figure 5.6: Expected yields in each bin of the Dalitz plot for the B^+ (left) and B^- (right) decays.

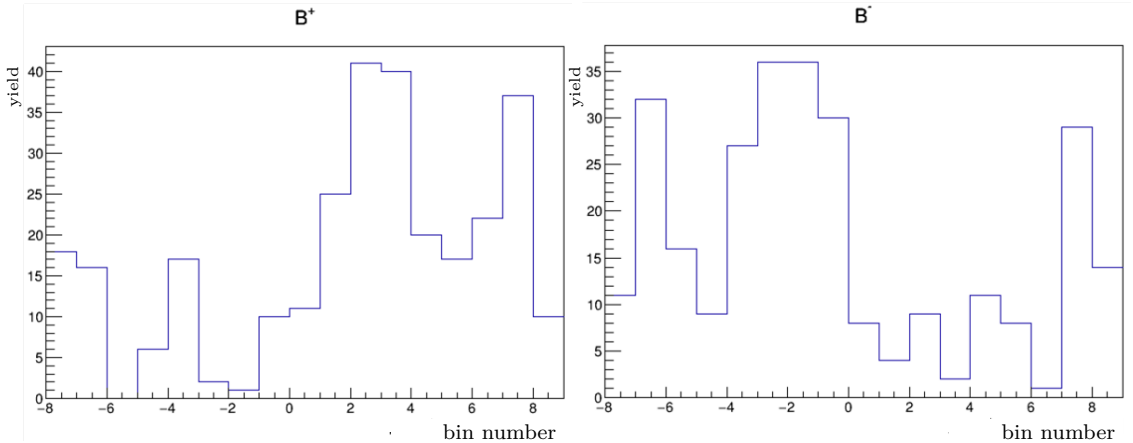


Figure 5.7: Yields in each bin of the Dalitz plot for the B^+ (left) and B^- (right) decays in data, for a region of the invariant-mass around the B mass.

The expected yields in each bin are presented in Figure 5.6 for B^+ (left) and B^- (right). In order to compare these expected yields with the data, the number of

events in each Dalitz bin in the signal region around the B mass in data is shown in Figure 5.7. It can be seen that the distribution of the expected yields follows a similar distribution to that of the number of events observed in data, which was studied as a cross-check for the computed expected yields in each bin.

5.4.2 *CP* fit

In order to perform the sensitivity study and determine the expected statistical uncertainties on the *CP* observables $x_{\pm}$, $y_{\pm}$, pseudoexperiments are used where signal-only pseudoexperiments are generated. The uncertainties are not expected to degrade significantly when performing the measurement in data, which includes background, given the high purity and low level of background. The idea is to generate a random dataset for each pseudoexperiment based on the computed expected yields, where the yield in each bin follows a Poisson distribution. For such distribution, given the expectation λ_i, the probability of observing x_i events in bin i is given by

$$f(x_i|\lambda_i) = \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!}. \tag{5.8}$$

A fit can then be performed to each pseudodataset in order to extract the *CP* observables $x_{\pm}$, $y_{\pm}$, using Equations (5.4) with a Poisson Maximum Likelihood (ML) fit. The principle of the ML method is to maximise the likelihood $\mathcal{L}$, defined as

$$\mathcal{L} = \prod_i f(x_i|\lambda_i). \tag{5.9}$$

This corresponds to the product over the bins i of the Poisson distributions. In the ML method, it is stated that the best estimators for the λ_i are obtained at the working point for which $\mathcal{L}$ is maximised. Given that $\mathcal{L}$ and $\ln(\mathcal{L})$ are maximised for the same λ_i, it is convenient to maximise $\ln(\mathcal{L})$ given the properties of the logarithm, allowing the product to be replaced by a sum:

$$\begin{aligned} \ln(\mathcal{L}) &= \sum_i \ln\left[f(x_i|\lambda_i)\right] \\ &= \sum_i \left[x_i \cdot \ln(\lambda_i) - \lambda_i\right] - \sum_i \ln(x_i!). \end{aligned} \tag{5.10}$$

This fit is repeated for each pseudoexperiment, from which the values of the *CP* observables are extracted. In this way, distributions of the values of the *CP* observables are formed and the mean value is taken to be the observed value of the *CP* observables $x_{\pm}^{obs}$, $y_{\pm}^{obs}$. Similarly, the expected uncertainty on each *CP* observable $\sigma_{x_{\pm}}^{obs}$, $\sigma_{y_{\pm}}^{obs}$ is defined as the mean of the uncertainty distribution obtained from the pseudoexperiments. Moreover, a pull distribution is constructed for each observable. The pull for observable x is defined as

$$\text{pull}(x) = \frac{x^{fit} - x^{gen}}{\sigma^{fit}}, \tag{5.11}$$

where x^{fit} and σ^{fit} are the fitted value and its uncertainty obtained from the fit, while x^{gen} corresponds to the generated value of the *CP* observables obtained from the chosen fixed values of the physics parameters when generating the pseudodatasets. Figure 5.8 shows the pull distribution obtained for the x_+ observable, with a Gaussian fit superimposed.

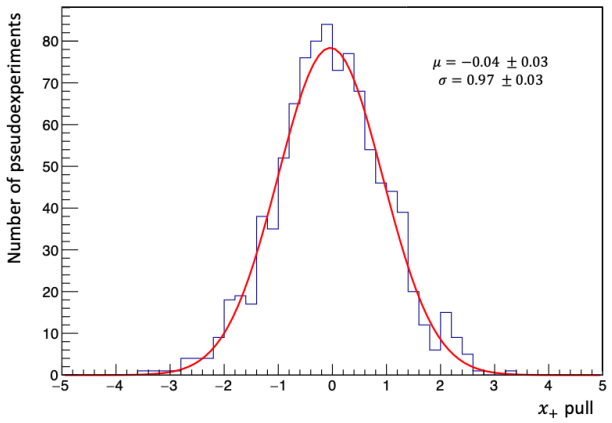


Figure 5.8: Pull distribution of the *CP* observable x_+.

The pull distributions present a good behaviour in general, reflected by the mean of the Gaussian function being consistent with 0 and the width consistent with 1. The uncertainties are well estimated from the fit, shown by the fact that the widths of all but one of the pull distributions are consistent with 1. Some small biases b of the mean value are observed for some observables with respect to the generated value. The bias observed is below 7% of the statistical uncertainty. To correct for

this and account for any bias, a systematic uncertainty is added to the uncertainties $\sigma^{obs}_{x_\pm}$, $\sigma^{obs}_{y_\pm}$. This is done by adding the bias in quadrature to the mean uncertainties:

$$\sigma'_x = \sqrt{(\sigma^{obs}_x)^2 + (b_x)^2}. \tag{5.12}$$

It is found that these biases are purely statistical effects, given the very low yields in some bins. This is done by repeating the pseudoexperiments study at higher signal yields. Since the bin yields are larger and therefore statistically further away from 0, the biases are reduced.

5.4.3 Physics parameters fit

The final step of the sensitivity study is to determine the expected uncertainty on the determination of γ. The physics parameters, including γ, are directly related to the *CP* observables through Equations (5.2). To obtain the physics parameters from the *CP* observables, a fit is performed by finding the global minimum of the χ^2 defined as

$$\chi^2 = \vec{x}^T M^{-1}_{cov} \vec{x}, \tag{5.13}$$

where $\vec{x}$ is the vector containing the *CP* observables:

$$\vec{x} = \vec{x}^{obs} - \vec{x}^{exp}(r_B, \delta_B, \gamma) = \begin{pmatrix} x^{obs}_- \\ y^{obs}_- \\ x^{obs}_+ \\ y^{obs}_+ \end{pmatrix} - \begin{pmatrix} x^{exp}_-(r_B, \delta_B, \gamma) \\ y^{exp}_-(r_B, \delta_B, \gamma) \\ x^{exp}_+(r_B, \delta_B, \gamma) \\ y^{exp}_+(r_B, \delta_B, \gamma) \end{pmatrix}. \tag{5.14}$$

The covariance matrix M_{cov} is computed from the correlation matrix M_{corr}, determined from the ensemble of pseudoexperiments, for each pair of *CP* observables i, j:

$$M_{cov}[i,j] = \sigma'_i \sigma'_j M_{corr}[i,j]. \tag{5.15}$$

In the χ^2 fit, the uncertainties σ'_x with the bias correction are used. Figure 5.9 presents the two-dimensional scans of the physics parameters which can be used to display the sensitivity on the physics parameters. These present the χ^2 difference at each point of the phase space with respect to the global minimum. Contours at $\Delta\chi^2 = 2.30, 6.18$ corresponding to the $1\sigma, 2\sigma$ bands (68.3% and 95.5% confidence levels) are shown for γ vs r_B (left) and γ vs δ_B (right).

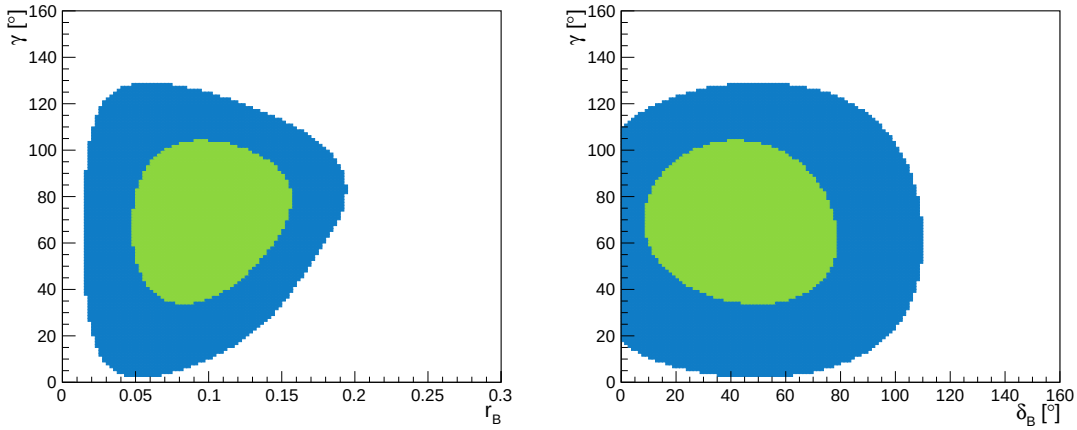


Figure 5.9: Two-dimensional physics parameters scan showing the contours at $\Delta\chi^2 = 2.30, 6.18$ corresponding to the $1\sigma, 2\sigma$ bands (68.3% and 95.5% confidence levels).

While the uncertainty is broader at a 2σ level, it is found that the uncertainty on γ obtained from the fit is $\sigma(\gamma) \simeq 16^\circ$ at the global minimum. Despite the fact that separate measurements using $B^\pm \to DK^{*\pm}$ alone are not available from the previous Belle and BaBar measurements, an improvement of about a factor 2 is achievable from this LHCb measurement. This is estimated by inspecting the confidence level plots from the published results by Belle and BaBar [113, 114]. This shows that this measurement is a valuable addition to the family of γ measurements and that this analysis is valuable to perform. This study therefore motivated the addition of these BPGGSZ modes to the ADS/GLW modes previously studied [6].

For every one billion particles of antimatter there were one billion and one particles of matter. And when the mutual annihilation was complete, one billionth remained – and that's our present universe.

— Albert Einstein

6

Measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays

Invariant-mass fits and results

This chapter presents the second part of the analysis, where the CKM angle γ is determined from measurements of *CP* observables in $B^{\pm} \to DK^{*\pm}$ decays at LHCb. The invariant-mass fits used to measure the *CP* observables are detailed. The results and corresponding systematic uncertainties are given, before performing a combination of the results from the different D decay modes to obtain the physics parameters of interest.

6.1 Extended maximum likelihood method

The fits to data are performed using the extended maximum likelihood method [116], from which the values of the fit model parameters are determined. For a given set of parameters $\boldsymbol{\theta}$ and n observed quantities $\boldsymbol{x} = (x_1, ..., x_n)$ which follow the PDF $f(x_i; \boldsymbol{\theta})$, the likelihood $\mathcal{L}$ is defined as

$$\mathcal{L} = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) \,. \tag{6.1}$$

The likelihood function represents the probability to observe the data for different values of the parameters used to describe the fit model. The statement of the

maximum likelihood method is that the best estimators for the parameters $\boldsymbol{\theta}$ correspond to the working point where the likelihood function is maximised. In practice, it is more convenient to work with $\ln(\mathcal{L})$ given that $\ln(\mathcal{L})$ is maximised at the same parameters $\boldsymbol{\theta}$ as $\mathcal{L}$. This is due to the properties of the logarithm and in particular the fact that the product can be replaced by a sum in the definition of the likelihood:

$$\ln(\mathcal{L}) = \sum_{i=1}^{n} \ln\left[f(x_i;\boldsymbol{\theta})\right]. \tag{6.2}$$

This method can be extended for the case where n is not fixed but taken as an additional parameter and corresponds to the number of candidates contained in the PDF. This is represented by a Poisson random variable, with mean μ, which is added as an additional term to the likelihood. The extended likelihood can therefore be written as

$$\mathcal{L}(\mu,\boldsymbol{\theta}) = \frac{\mu^n}{n!} e^{-\mu} \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}), \tag{6.3}$$

$$\ln\left[\mathcal{L}(\mu,\boldsymbol{\theta})\right] = \sum_{i=1}^{n} \ln\left[\mu(\boldsymbol{\theta}) f(x_i;\boldsymbol{\theta})\right] - \mu(\boldsymbol{\theta}) + C, \tag{6.4}$$

where C is an additional constant. In the case where μ is independent of $\boldsymbol{\theta}$, the best values obtained with the extended maximum likelihood and the default likelihood are the same. In this instance, the estimator for the mean μ corresponds to n.

6.2 Fit strategy

After selecting the data samples and studying the different backgrounds present in the data, the fitting procedure can take place. The fits are performed to the $m(DK^{*\pm})$ invariant-mass distributions. As described in Section 4.4, there are three components which contribute in all D decay modes. The signal, combinatorial background and partially reconstructed backgrounds coming from $B \to D^*K^*$ decays. There is a small overlap in the invariant-mass distributions of the $B \to D^*K^*$ and signal $B \to DK^*$ decays. It is possible to reliably parametrise the $B \to D^*K^*$ background for the $B \to DK^*$ favoured D decays. However, this is not the case for

the other D decay modes. The reason for this is is due to the CP violation present amongst the other modes. As given in Equations (4.13)–(4.14), the yield ratios between the three different partially reconstructed decays are fixed through the ratios of branching fractions and selection efficiencies, which assumes no CP violation in the favoured modes. The assumptions on the relative branching fractions no longer hold if there is CP violation. The partially reconstructed $B \to D^*K^*$ decays are of the type Scalar $\to$ Vector Vector. There are three different amplitudes, resulting from the three possible helicity states due to angular momentum conservation. The three amplitudes cannot interfere, hence each has its own strong phase. There is not sufficient data to constrain the low-mass region simultaneously with the rest of the degrees of freedom in the fit.

Therefore, the strategy is to fit the full spectrum for the favoured modes, called global fit, and then place a cut on the data to perform the CP fit in a reduced range. Given the fact that the partially reconstructed background decays are well separated from the signal peak, most of the partially reconstructed background is removed. The reduced invariant-mass range is chosen to start from $5230\,\mathrm{MeV}/c^2$. The global fit is then used to understand the extrapolation of the partially reconstructed background into the signal region of the other channels.

6.3 Global fit

The first step in the fitting procedure consists of a global fit, performed in the high yield favoured $K\pi$ and $K\pi\pi\pi$ modes using an unbinned extended maximum likelihood fit. The goal is to perform a fit to the invariant-mass distributions of the selected candidates in these modes in order to determine the amount of partially reconstructed background leaking into the signal region. Moreover, this global fit helps to fix parameters related to these backgrounds in the subsequent CP fit, where the CP observables are measured.

6.3.1 Fit components

Three components contribute in the favoured modes. These contributions consist of the signal, the combinatorial background, and the partially reconstructed backgrounds. Each component is modelled by a PDF, f. In fact, within the partially reconstructed backgrounds there are six subcomponents, which correspond to the three considered decays and the two distinguishable possible helicity states for each decay. The total PDF, G, used as a parametrisation for the global fit in the favoured $K\pi(\pi\pi)$ mode can then be written as

$$\begin{aligned} G_{K\pi(\pi\pi)}(m) = & N^{sig}_{K\pi(\pi\pi)} \cdot f^{sig}_{K\pi(\pi\pi)} + N^{comb}_{K\pi(\pi\pi)} \cdot f^{comb}_{K\pi(\pi\pi)} \\ & + N^{pr,\pi^{\pm}(0)}_{K\pi(\pi\pi)} \cdot f^{pr,\pi^{\pm}(0)} + N^{pr,\pi^{\pm}(\pm1)}_{K\pi(\pi\pi)} \cdot f^{pr,\pi^{\pm}(\pm1)} \\ & + N^{pr,\gamma(0)}_{K\pi(\pi\pi)} \cdot f^{pr,\gamma(0)} + N^{pr,\gamma(\pm1)}_{K\pi(\pi\pi)} \cdot f^{pr,\gamma(\pm1)} \\ & + N^{pr,\pi^0(0)}_{K\pi(\pi\pi)} \cdot f^{pr,\pi^0(0)} + N^{pr,\pi^0(\pm1)}_{K\pi(\pi\pi)} \cdot f^{pr,\pi^0(\pm1)}. \end{aligned} \tag{6.5}$$

The yields $N^{pr,\pi^0(0)}$, $N^{pr,\gamma(0)}$ and $N^{pr,\pi^{\pm}(0)}$ are the yields of the corresponding partially reconstructed decays $B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$, $B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ and $B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$, for the 0 helicity state. Similarly, $N^{pr,\pi^{\pm}(\pm1)}$, $N^{pr,\gamma(\pm1)}$ and $N^{pr,\pi^0(\pm1)}$ refer to the corresponding yields for the ±1 helicity state. Moreover, two of the yields of the partially reconstructed decays are fixed with respect to the third one, as already mentioned in Equations (4.13)–(4.14). This is done separately for each helicity state. For the 0 helicity state (and similarly for ±1), the yields $N^{pr,\gamma(0)}$ and $N^{pr,\pi^{\pm}(0)}$ are related to $N^{pr,\pi^0(0)}$ in the following way:

$$\begin{aligned} N^{pr,\gamma(0)} = & \frac{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm})}{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} \\ & \times \frac{\epsilon_{sel}(B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm})}{\epsilon_{sel}(B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} \times N^{pr,\pi^0(0)}, \end{aligned} \tag{6.6}$$

$$\begin{aligned} N^{pr,\pi^{\pm}(0)} = & \frac{\mathcal{B}(B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp})}{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} \\ & \times \frac{\epsilon_{sel}(B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp})}{\epsilon_{sel}(B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm})} \times N^{pr,\pi^0(0)}. \end{aligned} \tag{6.7}$$

This essentially means that the yields of $B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ for both helicity states, $N^{pr,\pi^0(0)}$ and $N^{pr,\pi^0(\pm1)}$, are free in the fit, while the other two

are fixed through ratios of branching fractions and selection efficiencies. The branching fraction ratios,

$$R^{\gamma} = \frac{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^{0}[\gamma])K^{*\pm})}{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^{0}[\pi^{0}])K^{*\pm})}\,, \tag{6.8}$$

$$R^{\pi^{\pm}} = \frac{\mathcal{B}(B^{0} \to D^{*\pm}(\to D^{0}[\pi^{\pm}])K^{*\mp})}{\mathcal{B}(B^{\pm} \to D^{*0}(\to D^{0}[\pi^{0}])K^{*\pm})}\,, \tag{6.9}$$

are Gaussian constrained in the fit. The mean and the width of the corresponding Gaussian functions are taken to be the values given in Equations (4.15)–(4.16).

Concerning the signal component, the shape parameters of the PDF are fixed to the MC fit results presented in Section 4.4.1, while the mean and width of the signal peak is determined by the data, which allows for differences between data and simulation. Finally, the combinatorial background is modelled by a simple exponential function, with parameter τ, which is free in the fit and determined by the data.

6.3.2 Fit results

The invariant-mass distributions, with the superimposed global fit results, are shown in Figures 6.1 and 6.2 for the $K\pi$ and $K\pi\pi\pi$ modes respectively. The corresponding fit parameter results are presented in Table 6.1. It can be seen that 5500 and 3445 signal candidates are present in the data for the favoured $K\pi$ and $K\pi\pi\pi$ modes respectively. The partially reconstructed backgrounds, which are well separated from the signal peak, have most of the events contained within the double-peak structure in the low-mass region. The number of partially reconstructed events leaking into the signal region is determined to be 25 ± 1 for $K\pi$ and 17 ± 1 for $K\pi\pi\pi$, where the signal region is $m(DK^{*\pm}) > 5230\,\text{MeV}/c^2$. In the other D decay modes, the number of partially reconstructed background events is fixed by scaling the number obtained in the favoured mode using the corresponding branching fractions and selection efficiencies. The two-body final-state modes are scaled with respect to the yield in $K\pi$, while the four-body modes are scaled with respect to the $K\pi\pi\pi$ global fit results. Table 6.2 presents the number of events of partially reconstructed backgrounds leaking into the signal region for each D decay mode.

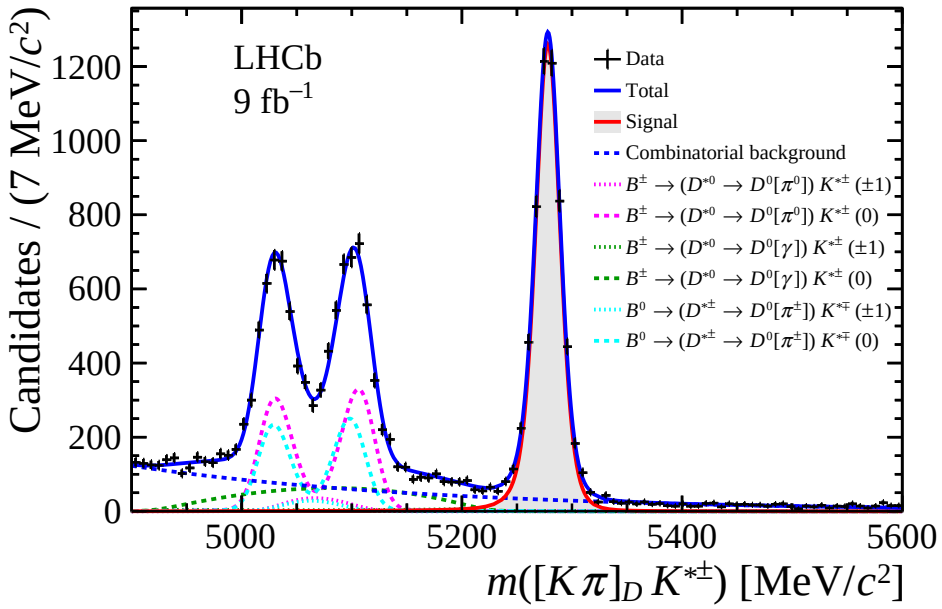


Figure 6.1: Invariant-mass distribution of the two-body favoured mode $K\pi$ with the fit result superimposed, using the full Run 1 and Run 2 dataset corresponding to an integrated luminosity of 9 fb^{-1}.

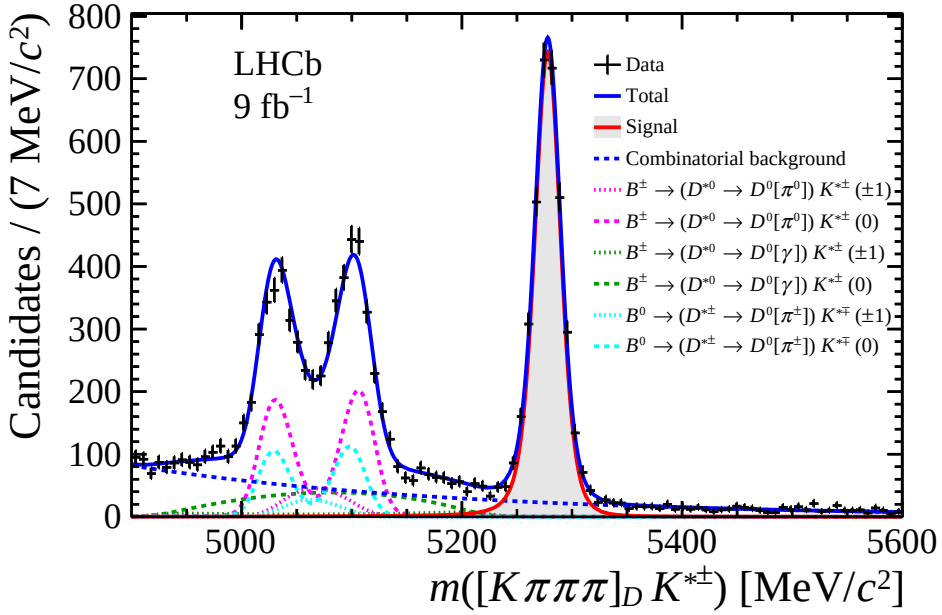


Figure 6.2: Invariant-mass distribution of the four-body favoured mode $K\pi\pi\pi$ with the fit result superimposed, using the full Run 1 and Run 2 dataset corresponding to an integrated luminosity of 9 fb^{-1}.

Table 6.1: Fit parameter results obtained from the global fit to data for the favoured modes $K\pi$ and $K\pi\pi\pi$.

Parameter	$K\pi$	$K\pi\pi\pi$
N^{sig}	5500 ± 79	3445 ± 63
N^{comb}	4451 ± 108	3107 ± 91
$N^{pr,\pi^0(0)}$	3565 ± 137	2187 ± 104
$N^{pr,\pi^0(\pm1)}$	349 ± 59	446 ± 58
μ	5278.1 ± 0.2	5278.0 ± 0.3
σ	10.8 ± 0.2	11.3 ± 0.2
τ	-0.0036 ± 0.0001	-0.0034 ± 0.0001
R^{γ}	0.54 ± 0.02	0.54 ± 0.02
$R^{\pi^{\pm}}$	0.76 ± 0.07	0.56 ± 0.08

Table 6.2: Number of partially reconstructed events leaking into the signal region for each D decay mode.

D decay mode	Part. Reco. events
$K\pi$	25
KK	2.7
$\pi\pi$	0.9
πK	0.06
$K\pi\pi\pi$	17
$\pi\pi\pi\pi$	1.5
$\pi K\pi\pi$	0.04

6.4 *CP* fit

The second step in the fitting procedure consists of the *CP* fit, for which the goal is to measure the *CP* observables of interest. This is done separately for the ADS/GLW and BPGGSZ modes. For the favoured modes, the number of background events are determined from the global fits. To take into account any small leaking partially reconstructed background in the other modes, corrections are consequently made by scaling the leaking component obtained from the global fit through corresponding branching ratios and efficiencies.

6.4.1 Fit components

- Signal:

  The signal shape parameters are fixed from the MC fits for all D decay modes (Section 4.4.1). The mean and width are free in the fit and shared across modes with the same number of final-state particles.

- Combinatorial background:

  The yield of the combinatorial background is free for each decay mode, as well as the parameter of the exponential function. The slope parameter defining the shape is shared across modes with the same number of final-state particles.

- Partially reconstructed background:

  The shapes are completely fixed from the MC fits (Section 4.4.3.2). The yield of $B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ in both helicity states 0 and ± 1, as well as the branching fraction ratios, are fixed to the values obtained from the global fit results. The corresponding number of partially reconstructed events for each D decay mode is given in Table 6.2.

- Charmless background in $\pi\pi$ and $\pi\pi\pi\pi$ modes:

  Shape and yield entirely fixed from the sideband studies (details in Section 4.4.3.3).

- Λ_b^0 background in KK mode:
  Fixed shape and yield (for details see Section 4.4.3.4) .

- B_s background in πK and $\pi K \pi\pi$ modes:
  Fixed shape and yield (for detail see Section 4.4.3.5).

6.4.2 ADS/GLW *CP* fit

6.4.2.1 Asymmetry corrections

When measuring physics *CP* asymmetries of interest A^{phys}, any other asymmetries A^{corr} present in data coming from other sources must be taken into account. The raw asymmetry A^{raw} can thus be written as

$$A^{raw} = A^{phys} + A^{corr}. \tag{6.10}$$

Such corrections are included in the fit when measuring the *CP* observables of interest. The two main sources of asymmetries taken into account and corrected for in this analysis are the $B^{\pm}$ production asymmetries A^{prod} and the detection asymmetries A^{det} for kaons and pions:

$$A^{corr} = A^{prod} + A^{det}. \tag{6.11}$$

Firstly, production asymmetries come from the fact that particles and antiparticles have different heavy flavour production rates [117–119]. For this analysis, the quantity of interest is the $B^{\pm}$ production asymmetry, which reflects the difference in production rates of B^+ compared to B^-. This quantity was directly measured in Reference [120], obtaining a value of $A_{B^+} = (+0.028 \pm 0.068)\%$, which is used in this work.

Secondly, it is important to correctly account for detection asymmetries for $K^{\pm}$ and $\pi^{\pm}$ in this analysis where kaons and pions form all final states considered. The kaon detection asymmetries arise from differences in the nuclear interaction lengths between K^+ and K^-, which interact differently with the material as they traverse the LHCb detector. The detection asymmetries are studied and measured in Reference [121] using Run 1 data, where calibration samples of $D^- \to K^+\pi^-\pi^-$

and $D^- \to K^0_S\pi^-$ are used. The kinematics are weighted to match signal kaons from $B^- \to D\pi^-$ favoured decays. Using these two D decays, the detection asymmetry for $K^+\pi^-$ pairs is determined. The calibration samples are compared by determining the charge asymmetries in bins of transverse momentum p_T and rapidity y, and taking the average across all kinematic bins. The asymmetries are further studied in Reference [120] to understand the uncertainty that should be applied due to the fact that the asymmetry is applied to the combined Run 1 and Run 2 dataset, resulting in a value of $A_{K\pi} = (-0.960 \pm 0.134)\%$.

Due to the fact that more K^- particles are likely to be absorbed in the detector compared to K^+, more K^+ could reach the calorimeter, triggering the L0 hadron trigger. Therefore, any L0 hadron trigger asymmetry must be considered, which was also studied in Reference [120]. This is done by studying favoured $B^\pm \to D(K^\pm\pi^\mp)\pi^\pm$ decays, where only a kaon asymmetry is involved since two opposite-charge pions are present in the final state. This is achieved by selecting two independent samples using two different trigger lines. The first sample selects events which are triggered on signal (`L0HadronDecision_TOS`), while the second sample is formed by only selecting events with a trigger independent of signal (`L0Global_TIS`). The latter does not contain any L0 hadron effects. This makes it possible to study the kaon asymmetry by performing independent fits to the two samples and measuring A_{B^+} in both cases, and looking at the difference in the values of A_{B^+} to estimate the effect coming from the L0 hadron trigger in the kaon asymmetry. An additional kaon asymmetry of $A^{\mathrm{L0,trig}}_{K\pi} = (-0.17 \pm 0.08)\%$ is obtained in this way, which is taken into account in this work by assigning a systematic uncertainty on the kaon detection asymmetry $A_{K\pi}$.

As mentioned previously, the $A_{K\pi}$ asymmetry correction only allows the correction for $K^+\pi^-$ pairs. In order to be able to correct for kaon or pion final-state tracks in isolation, the pion detection asymmetry A_π is needed, which is also measured in Reference [121] in kinematic bins using Run 1 data. Similarly to $A_{K\pi}$, the pion detection asymmetry is assigned a systematic uncertainty in Reference [120] to account for any differences in the kinematics between the Run 1 and Run 2 data,

yielding a value of $A_\pi = (-0.064 \pm 0.019)\%$. The values used in the *CP* fit for the asymmetry corrections are summarised here:

- $A_{B^+}^{corr} = (+0.028 \pm 0.068)\%$
- $A_{K\pi}^{corr} = (-0.960 \pm 0.134)\%$
- $A_{\pi}^{corr} = (-0.064 \pm 0.019)\%$

While the production asymmetry correction is applied for all modes, to apply the detection asymmetry corrections to the *CP* observables for the different D decay modes, the number of kaons and pions present in the final state must be determined. For kaon or pion pairs h^+h^-, the detection asymmetries cancel as the net number of charged particles of a given type is zero. This is for example the case for the pions coming from the K^0_S particle following the $K^{*\pm}$ decay. Therefore, the remaining particles to consider are the bachelor pion and the D daughters. For $K^+\pi^-$ pairs, the $A_{K\pi}$ asymmetry correction is used. For pions in isolation, the A_π is applied. For example, in the favoured mode $B^+ \to [K^+\pi^-]_D K^0_S\pi^+$, there is a $K^+\pi^-$ pair coming from the D meson decay and the additional bachelor pion. Therefore, the correction applied is $A_{K\pi} + A_\pi$.

Table 6.3: Summary of the asymmetry corrections applied in each D decay mode.

D decay mode	decay	asymmetry correction
$K\pi$	$B^+ \to [K^+\pi^-]_D K^0_S\pi^+$	$A_{B^+} + A_{K\pi} + A_\pi$
KK	$B^+ \to [K^+K^-]_D K^0_S\pi^+$	$A_{B^+} + A_\pi$
$\pi\pi$	$B^+ \to [\pi^+\pi^-]_D K^0_S\pi^+$	$A_{B^+} + A_\pi$
πK	$B^+ \to [\pi^+K^-]_D K^0_S\pi^+$	$A_{B^+} - A_{K\pi} + A_\pi$
$K\pi\pi\pi$	$B^+ \to [K^+\pi^-\pi^+\pi^-]_D K^0_S\pi^+$	$A_{B^+} + A_{K\pi} + A_\pi$
$\pi\pi\pi\pi$	$B^+ \to [\pi^+\pi^-\pi^+\pi^-]_D K^0_S\pi^+$	$A_{B^+} + A_\pi$
$\pi K\pi\pi$	$B^+ \to [\pi^+K^-\pi^+\pi^-]_D K^0_S\pi^+$	$A_{B^+} - A_{K\pi} + A_\pi$

Table 6.3 summarises the correction applied for each mode. It should be noted that these asymmetry corrections are smaller than the uncertainties measured in the fit used in the analysis presented here.

6.4.2.2 Fit parametrisation

This subsection details the *CP* fit parametrisation used to measure the *CP* observables, defined in Equations (2.18)–(2.23), (2.29), (2.34)–(2.35), (2.38)–(2.40). This consists of a simultaneous extended maximum likelihood fit, performed to the B invariant-mass distributions in different categories of events. These categories are defined by the charge of the B meson and the D decay mode considered, giving $2 \times 7 = 14$ categories. The floating parameters in the *CP* fit are summarised below:

- *CP* observables:
  $A_{K\pi}$, A_{KK}, $A_{\pi\pi}$, $A_{\pi K}$, R_{KK}, $R_{\pi\pi}$, $R_{\pi K}$, $A_{K\pi\pi\pi}$, $A_{\pi\pi\pi\pi}$, $A_{\pi K\pi\pi}$ $R_{\pi\pi\pi\pi}$, $R_{\pi K\pi\pi}$
- Total signal yields of the favoured $K\pi$ and $K\pi\pi\pi$ modes:
  $N^{sig}_{K\pi}$, $N^{sig}_{K\pi\pi\pi}$
- Signal mean and width for the 2-body and 4-body final states:
  μ_{2body}, σ_{2body}, μ_{4body}, σ_{4body}
- Combinatorial background yield in each category:
  $N^{-,comb}_{K\pi}$, $N^{+,comb}_{K\pi}$, $N^{-,comb}_{KK}$, $N^{+,comb}_{KK}$, $N^{-,comb}_{\pi\pi}$, $N^{+,comb}_{\pi\pi}$, $N^{-,comb}_{\pi K}$, $N^{+,comb}_{\pi K}$,
  $N^{-,comb}_{K\pi\pi\pi}$, $N^{+,comb}_{K\pi\pi\pi}$, $N^{-,comb}_{\pi\pi\pi\pi}$, $N^{+,comb}_{\pi\pi\pi\pi}$, $N^{-,comb}_{\pi K\pi\pi}$, $N^{+,comb}_{\pi K\pi\pi}$
- Combinatorial slope parameter for the 2-body and 4-body final states:
  τ_{2body}, τ_{4body}.

The *CP* observables can be rewritten in terms of the measured yields in the fit, taking into account branching ratios and selection efficiencies as

$$A_f^{raw} = \frac{N(B^- \to [f]_D K^{*-}) - N(B^+ \to [f]_D K^{*+})}{N(B^- \to [f]_D K^{*-}) + N(B^+ \to [f]_D K^{*+})}, \tag{6.12}$$

$$R_{hh(hh)} = \frac{N(B^- \to [hh(hh)]_D K^{*-}) + N(B^+ \to [hh(hh)]_D K^{*+})}{N(B^- \to [K\pi(\pi\pi)]_D K^{*-}) + N(B^+ \to [K\pi(\pi\pi)]_D K^{*+})} \times \frac{\mathcal{B}(D \to K\pi(\pi\pi))}{\mathcal{B}(D \to hh(hh))} \times \frac{\epsilon_{tot}(K\pi(\pi\pi))}{\epsilon_{tot}(hh(hh))}, \tag{6.13}$$

$$R_{\pi K(\pi\pi)} = \frac{N(B^- \to [\pi K(\pi\pi)]_D K^{*-}) + N(B^+ \to [\pi K(\pi\pi)]_D K^{*+})}{N(B^- \to [K\pi(\pi\pi)]_D K^{*-}) + N(B^+ \to [K\pi(\pi\pi)]_D K^{*+})} \times \frac{\epsilon_{tot}(K\pi(\pi\pi))}{\epsilon_{tot}(\pi K(\pi\pi))}, \quad (6.14)$$

where $hh(hh)$ refers to the KK, $\pi\pi$ or $\pi\pi\pi\pi$ modes. The D meson final state is defined as f. For simplicity and clarity, the total yield, as well as the yields in each charge, for each final state f are defined as follow:

$$\begin{aligned} N_f^- &= N(B^- \to [f]_D K^{*-}) \\ N_f^+ &= N(B^+ \to [f]_D K^{*+}) \\ N_f &= N_f^- + N_f^+. \end{aligned} \quad (6.15)$$

This allows Equation (6.12) to be written as

$$A_f^{raw} = \frac{N_f^- - N_f^+}{N_f^- + N_f^+}. \quad (6.16)$$

In order to take into account and apply the asymmetry corrections (detailed in Section 6.4.2.1), it is more convenient to define a parameter $a_f^{raw} = N_f^-/N_f^+$. This allows the asymmetry corrections to be applied as products, as detailed below. Using Equation (6.16), the parameter a_f^{raw} can be expressed in terms of the asymmetry A_f^{raw}:

$$a_f^{raw} = \frac{1 + A_f^{raw}}{1 - A_f^{raw}}. \quad (6.17)$$

The yields in each charge can therefore be expressed as a function of the total yield and the parameter a_f^{raw}:

$$\begin{aligned} N_f^+ &= \frac{1}{1 + a_f^{raw}} N_f, \\ N_f^- &= \frac{1}{1 + \frac{1}{a_f^{raw}}} N_f. \end{aligned} \quad (6.18)$$

Since the raw asymmetry defined in Equation (6.16) is made of the CP asymmetry of interest and the asymmetry corrections:

$$\begin{aligned} A_f^{raw} &= A_f + A_f^{corr} \\ a_f^{raw} &= a_f \cdot a_f^{corr}, \end{aligned} \quad (6.19)$$

where the asymmetry correction parameter a_f^{corr} can be written as the product of the different sources which are the production asymmetry a^{prod} and detection asymmetry corrections a^{det}. The detection asymmetry correction depends on each final state:

$$a^{corr}_{K\pi(\pi\pi)} = a^{corr}_{B^+} \cdot a^{corr}_{K\pi} \cdot a^{corr}_{\pi} = \frac{1+A^{corr}_{B^+}}{1-A^{corr}_{B^+}} \cdot \frac{1+A^{corr}_{K\pi}}{1-A^{corr}_{K\pi}} \cdot \frac{1+A^{corr}_{\pi}}{1-A^{corr}_{\pi}} \tag{6.20}$$

$$a^{corr}_{hh(hh)} = a^{corr}_{B^+} \cdot a^{corr}_{\pi} = \frac{1+A^{corr}_{B^+}}{1-A^{corr}_{B^+}} \cdot \frac{1+A^{corr}_{\pi}}{1-A^{corr}_{\pi}} \tag{6.21}$$

$$a^{corr}_{\pi K(\pi\pi)} = a^{corr}_{B^+} \cdot a^{corr}_{-K\pi} \cdot a^{corr}_{\pi} = \frac{1+A^{corr}_{B^+}}{1-A^{corr}_{B^+}} \cdot \frac{1-A^{corr}_{K\pi}}{1+A^{corr}_{K\pi}} \cdot \frac{1+A^{corr}_{\pi}}{1-A^{corr}_{\pi}} \tag{6.22}$$

Compiling Equations (6.12)–(6.22) and the corresponding definitions, the parametrisation of the *CP* fit can be summarised in the following way:

$$a_f = \frac{1+A_f}{1-A_f} \tag{6.23}$$

$$N^+_{K\pi(\pi\pi)} = \frac{1}{1+a_{K\pi(\pi\pi)} \cdot a^{corr}_{K\pi(\pi\pi)}} N_{K\pi(\pi\pi)} \tag{6.24}$$

$$N^-_{K\pi(\pi\pi)} = \frac{1}{1+\frac{1}{a_{K\pi(\pi\pi)} \cdot a^{corr}_{K\pi(\pi\pi)}}} N_{K\pi(\pi\pi)} \tag{6.25}$$

$$N^+_{hh(hh)} = \frac{1}{1+a_{hh(hh)} \cdot a^{corr}_{hh(hh)}} N_{hh(hh)} \tag{6.26}$$

$$N^-_{hh(hh)} = \frac{1}{1+\frac{1}{a_{hh(hh)} \cdot a^{corr}_{hh(hh)}}} N_{hh(hh)} \tag{6.27}$$

$$N^+_{\pi K(\pi\pi)} = \frac{1}{1+a_{\pi K(\pi\pi)} \cdot a^{corr}_{\pi K(\pi\pi)}} N_{\pi K(\pi\pi)} \tag{6.28}$$

$$N^-_{\pi K(\pi\pi)} = \frac{1}{1+\frac{1}{a_{\pi K(\pi\pi)} \cdot a^{corr}_{\pi K(\pi\pi)}}} N_{\pi K(\pi\pi)} \tag{6.29}$$

$$N_{hh(hh)} = R_{hh(hh)} \times N_{K\pi(\pi\pi)} \times \frac{\mathcal{B}(D \to hh(hh))}{\mathcal{B}(D \to K\pi(\pi\pi))} \times \frac{\epsilon_{tot}(hh(hh))}{\epsilon_{tot}(K\pi(\pi\pi))} \tag{6.30}$$

$$\begin{aligned} N_{\pi K(\pi\pi)} &= R_{\pi K(\pi\pi)} \times N_{K\pi(\pi\pi)} \times \frac{\epsilon_{tot}(\pi K(\pi\pi))}{\epsilon_{tot}(K\pi(\pi\pi))} \\ &= R_{\pi K(\pi\pi)} \times N_{K\pi(\pi\pi)} \times \frac{\epsilon_{BDT}(\pi K(\pi\pi))}{\epsilon_{BDT}(K\pi(\pi\pi))} \cdot \epsilon_{veto}(\pi K(\pi\pi)) \,. \end{aligned} \tag{6.31}$$

For the total yield of the suppressed modes, corresponding to Equation (6.31), it can be seen that the ratio of efficiencies between the suppressed and favoured decays simplifies such that only the BDT efficiencies and the additional veto efficiency in the suppressed mode are involved. This is due to the fact that the rest of the selection is identical between the two decays and cancels in the efficiency ratio.

In summary, the signal yields in the different modes are parametrised in terms of the *CP* observables and the signal yield in the favoured modes, as well as the asymmetry corrections and efficiency ratios. Therefore, the *CP* observables are embedded in Equations (6.24)–(6.31) and are freely varying components in the fit and hence directly extracted.

6.4.2.3 Fit results

The B invariant-mass distributions for each category, that is for each B charge and D mode, are presented in Figures 6.3–6.5, where the *CP* fit results are superimposed. The corresponding fit projections in logarithmic scale are given in Appendix A, showing the small background contributions. The value of the *CP* observables obtained from the fit are given in Table 6.4. For the favoured mode, the asymmetries between B^+ and B^- are small and consistent with zero as expected given the current precision. Larger asymmetries can be observed for some of the other modes, where hints of *CP* violation can be seen in the KK mode for example. The largest asymmetries are seen in the suppressed modes, where $A_{\pi K}$ exhibits large *CP* violation which can be seen by eye in Figure 6.5. The correlation matrix of the statistical uncertainties on the *CP* observables for the ADS/GLW modes is given in Table 6.5. The largest correlations occur between the *CP* observables of the same mode, which is the case in particular for the πK mode. This is expected from the definition of the asymmetry and ratio *CP* observables which are directly

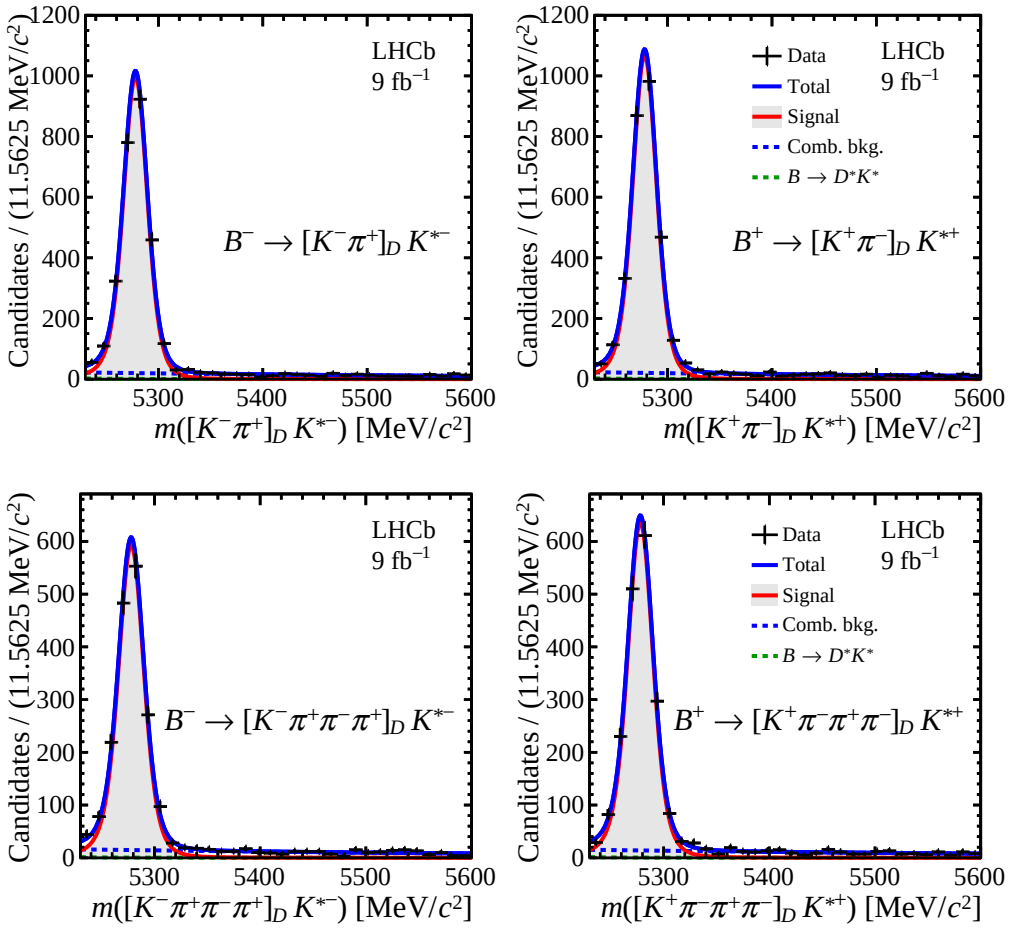


Figure 6.3: Invariant-mass distributions for the favoured $K\pi$ (top row) and $K\pi\pi\pi$ (bottom row) modes for B^- (left) and B^+ (right) decays, obtained with data corresponding to the full Run 1 and Run 2 dataset. The fit result is superimposed, showing the different components.

related, as can be seen for example in Equation (2.34). This correlation matrix is used in the combination of results to determine γ. The results for the rest of the fit parameters are listed in Table 6.6. These consist of the signal yield in the $K\pi$ and $K\pi\pi\pi$ modes and the combinatorial background yield in each category, as well as the corresponding shape parameters.

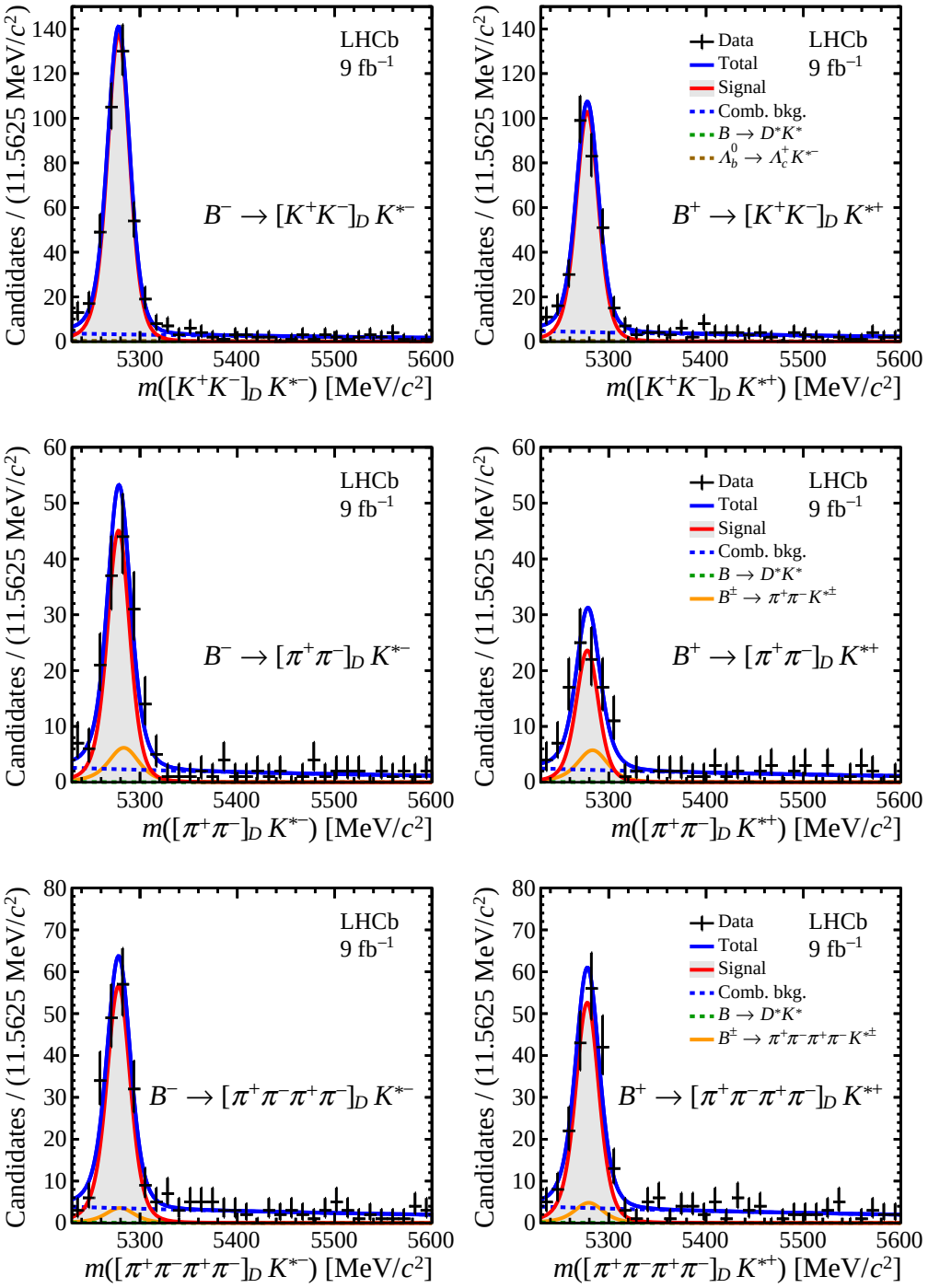


Figure 6.4: Invariant-mass distributions for the KK (top row), $\pi\pi$ (middle row) and $\pi\pi\pi\pi$ (bottom row) modes for B^- (left) and B^+ (right) decays, obtained with data corresponding to the full Run 1 and Run 2 dataset. The fit result is superimposed, showing the different components.

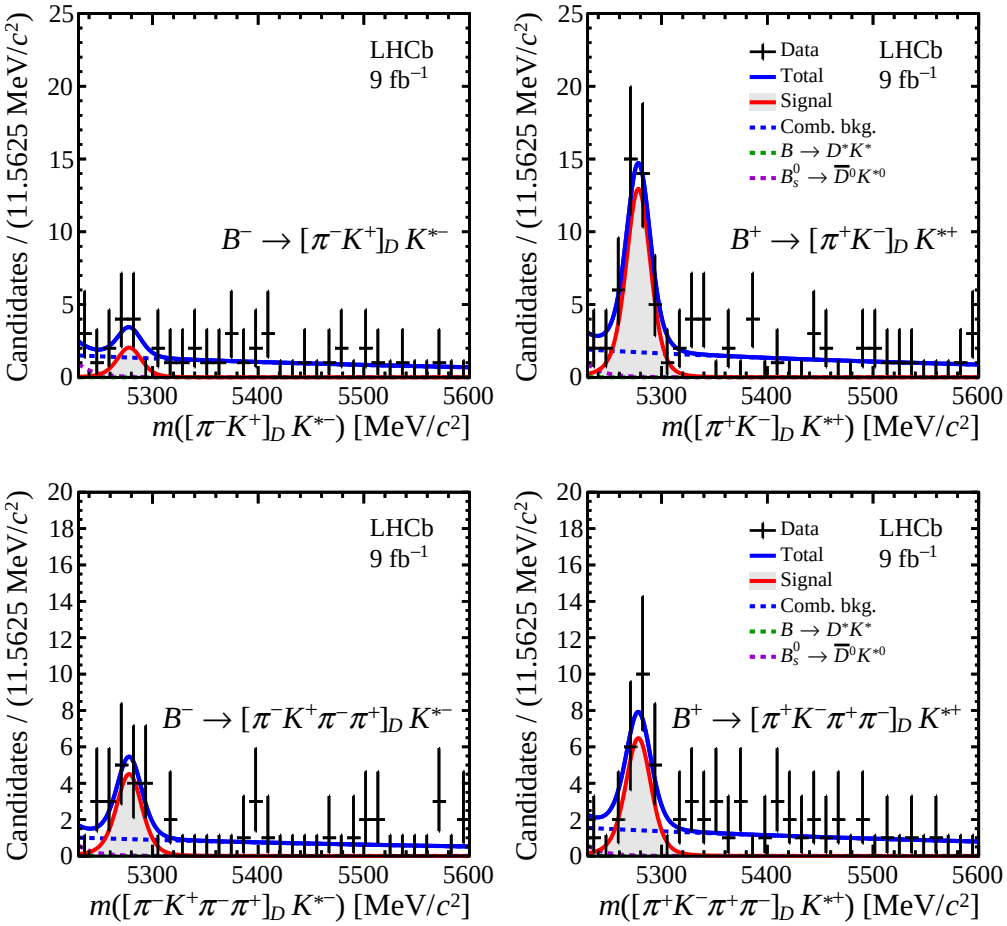


Figure 6.5: Invariant-mass distributions for the suppressed πK (top row) and $\pi K\pi\pi$ (bottom row) modes for B^- (left) and B^+ (right) decays, obtained with data corresponding to the full Run 1 and Run 2 dataset. The fit result is superimposed, showing the different components.

Table 6.4: Fit results obtained for the *CP* observables.

CP observable	Fit value
$A_{K\pi}$	-0.024 ± 0.014
A_{KK}	0.14 ± 0.04
$A_{\pi\pi}$	0.31 ± 0.09
$A_{\pi K}$	-0.73 ± 0.16
R_{KK}	1.10 ± 0.05
$R_{\pi\pi}$	0.96 ± 0.09
$R_{\pi K}$	0.0098 ± 0.0019
$A_{K\pi\pi\pi}$	-0.024 ± 0.018
$A_{\pi\pi\pi\pi}$	0.04 ± 0.06
$A_{\pi K\pi\pi}$	-0.19 ± 0.22
$R_{\pi\pi\pi\pi}$	1.05 ± 0.07
$R_{\pi K\pi\pi}$	0.0118 ± 0.0026

Table 6.5: Correlation matrix of the statistical uncertainties on the *CP* observables for two- and four-body modes. Values smaller than 0.1% are not shown and indicated with "–".

	$A_{K\pi}$	A_{KK}	$A_{\pi\pi}$	$A_{\pi K}$	R_{KK}	$R_{\pi\pi}$	$R_{\pi K}$	$A_{K\pi\pi\pi}$	$A_{\pi\pi\pi\pi}$	$A_{\pi K\pi\pi}$	$R_{\pi\pi\pi\pi}$	$R_{\pi K\pi\pi}$
$A_{K\pi}$	1	–	–	–	–	–	–	–	–	–	–	–
A_{KK}		1	–	–	−0.019	–	–	–	–	–	–	–
$A_{\pi\pi}$			1	–	–	−0.085	–	–	–	–	–	–
$A_{\pi K}$				1	–	–	0.183	–	–	–	–	–
R_{KK}					1	0.055	0.028	–	–	–	–	–
$R_{\pi\pi}$						1	0.020	–	–	–	–	–
$R_{\pi K}$							1	–	–	–	–	–
$A_{K\pi\pi\pi}$								1	–	–	–	–
$A_{\pi\pi\pi\pi}$									1	–	−0.022	–
$A_{\pi K\pi\pi}$										1	–	0.040
$R_{\pi\pi\pi\pi}$											1	0.029
$R_{\pi K\pi\pi}$												1

Table 6.6: Fit parameter results obtained from the *CP* fit to data, except from the *CP* observables which are presented separately.

Parameter	Fit value
$N^{sig}_{K\pi}$	5500 ± 81
$N^{sig}_{K\pi\pi\pi}$	3448 ± 65
$N^{-,comb}_{K\pi}$	482 ± 29
$N^{+,comb}_{K\pi}$	498 ± 30
$N^{-,comb}_{KK}$	79 ± 12
$N^{+,comb}_{KK}$	103 ± 13
$N^{-,comb}_{\pi\pi}$	57 ± 9
$N^{+,comb}_{\pi\pi}$	54 ± 9
$N^{-,comb}_{\pi K}$	33 ± 6
$N^{+,comb}_{\pi K}$	42 ± 7
$N^{-,comb}_{K\pi\pi\pi}$	372 ± 26
$N^{+,comb}_{K\pi\pi\pi}$	354 ± 25
$N^{-,comb}_{\pi\pi\pi\pi}$	90 ± 11
$N^{+,comb}_{\pi\pi\pi\pi}$	91 ± 11
$N^{-,comb}_{\pi K\pi\pi}$	24 ± 6
$N^{+,comb}_{\pi K\pi\pi}$	36 ± 7
μ_{2body}	5278.0 ± 0.2
μ_{4body}	5278.0 ± 0.2
σ_{2body}	11.1 ± 0.2
σ_{4body}	11.6 ± 0.2
τ_{2body}	-0.0021 ± 0.0004
τ_{4body}	-0.0018 ± 0.0005

Table 6.7: A summary of the observed signal yields for each D decay channel and for each B meson charge. The uncertainties are statistical only. The purity within the region defined as $[5230, 5330]\,\mathrm{MeV}/c^2$ is given for both charges combined.

D decay mode	B^-	B^+	Purity (%)
$K\pi$	2656 ± 55	2844 ± 57	94
KK	366 ± 20	274 ± 18	90
$\pi\pi$	121 ± 13	63 ± 10	69
πK	5 ± 4	35 ± 7	58
$K\pi\pi\pi$	1665 ± 44	1783 ± 45	93
$\pi\pi\pi\pi$	160 ± 14	149 ± 14	77
$\pi K\pi\pi$	13 ± 5	18 ± 5	58

From the *CP* fit results, the yields for each charge and each D decay mode are derived and given in Table 6.7, together with the purity for each D decay mode, for both charges combined. The purity is also derived from the fit results by taking the number of signal and background events within the region defined as $[5230, 5330]\,\text{MeV}/c^2$. As expected with the addition of data collected in 2017–2018, the signal yields are about 2.5 times larger than the ones obtained in the partial Run 2 analysis [6].

Table 6.8: Comparison of the fit results obtained for the *CP* observables with respect to the partial Run 2 results [6].

CP observable	New result	Previous result	$\mathcal{P}_0$
$A_{K\pi}$	-0.024 ± 0.014	-0.004 ± 0.023	1.1
A_{KK}	0.14 ± 0.04	0.06 ± 0.07	1.4
$A_{\pi\pi}$	0.31 ± 0.09	0.15 ± 0.13	1.7
$A_{\pi K}$	-0.73 ± 0.16	-0.81 ± 0.33	0.3
R_{KK}	1.10 ± 0.05	1.22 ± 0.09	1.6
$R_{\pi\pi}$	0.96 ± 0.09	1.08 ± 0.14	1.1
$R_{\pi K}$	0.0098 ± 0.0019	0.011 ± 0.004	0.3
$A_{K\pi\pi\pi}$	-0.024 ± 0.018	-0.013 ± 0.031	0.4
$A_{\pi\pi\pi\pi}$	0.04 ± 0.06	0.02 ± 0.11	0.2
$A_{\pi K\pi\pi}$	-0.19 ± 0.22	-0.45 ± 0.43	0.7
$R_{\pi\pi\pi\pi}$	1.05 ± 0.07	1.08 ± 0.13	0.3
$R_{\pi K\pi\pi}$	0.0118 ± 0.0026	0.011 ± 0.005	0.2

A clear improvement in the precision of the *CP* observables can also be seen, compared to the previous results published with partial Run 2 data. Not only is the dataset around two times larger, but different parts of the analysis are improved and the backgrounds are better understood and studied in detail. To check the agreement between the results obtained from the two analyses, the fact that datasets are overlapping must be taken into account. Therefore, a pull is determined as a quantity used to compare the results and defined as

$$\mathcal{P}_0 = \frac{|x - x_0|}{\sqrt{|\sigma^2 - \sigma_0^2|}}, \tag{6.32}$$

where x and x_0 represent the new result from this work and the previous result respectively. The difference $\sigma^2 - \sigma_0^2$ of the uncertainties squared between the two

sets of results is assumed to provide a good estimate of the uncorrelated uncertainty. The pull is computed for each *CP* observable and given in Table 6.8, where it can be seen that the results agree at the level between 0.2 and 1.7 standard deviations.

One change in this analysis compared to the previous one is a change in the choice of *CP* observables. Since the other analysis measured different *CP* observables, the quantities of interest are derived from the results in order to perform the comparison. The reason for this change is explained in Section 6.4.2.4, where the *CP* fit is validated using pseudoexperiments.

6.4.2.4 Fit validation with pseudoexperiments

In order to validate the *CP* fit setup and the results, pseudoexperiments are used to check for the stability and any potential bias of the fit. A set of 1000 pseudodatasets are generated using fixed values of the physics parameters, $r_B = 0.108$, $\delta_B = 34^\circ$ and $\gamma = 63.8^\circ$ [35]. Each dataset is then fitted back with the *CP* fit. The convergence rate of the pseudoexperiments is found to be larger than 95%. The pull for *CP* observable x is defined as

$$\mathcal{P} = \frac{x_{fit} - x_{gen}}{\sigma_{fit}}, \tag{6.33}$$

where x_{fit} and σ_{fit} are the fitted value and the corresponding uncertainty, and x_{gen} the generated value of the *CP* observable. The pull distribution for each observable is shown in Figures 6.6 and 6.7. The pulls are fitted with a Gaussian function, which is superimposed in the plots. A width consistent with 1 and the mean with 0 reflects that no significant bias is observed.

The choice of *CP* observables is different compared to the partial Run 2 analysis. This is the case for the suppressed modes, for which the $A_{\pi K(\pi\pi)}$, $R_{\pi K(\pi\pi)}$ parametrisation is used in the analysis presented here. It has been found that these observables behave better in this context compared to the $R^+_{\pi K(\pi\pi)}$, $R^-_{\pi K(\pi\pi)}$ parametrisation which was used in the partial Run 2 analysis and defined as

$$R^+_{\pi K(\pi\pi)} = \frac{\Gamma(B^+ \to [\pi^+K^-(\pi^+\pi^-)]_D K^{*+})}{\Gamma(B^+ \to [K^+\pi^-(\pi^+\pi^-)]_D K^{*+})}, \tag{6.34}$$

$$R^{-}_{\pi K(\pi\pi)} = \frac{\Gamma(B^- \to [\pi^- K^+(\pi^-\pi^+)]_D K^{*-})}{\Gamma(B^- \to [K^-\pi^+(\pi^-\pi^+)]_D K^{*-})}\,. \tag{6.35}$$

Indeed, the ratios in the suppressed modes were previously measured separately for each charge. This is usually a good parametrisation for low yields as the two observables $R^{+}_{\pi K(\pi\pi)}$ and $R^{-}_{\pi K(\pi\pi)}$ are statistically independent. However, performing the pseudoexperiments in this work, it was observed that some observables present a larger bias using the $R^{\pm}_{\pi K(\pi\pi)}$ parametrisation. It was also observed that this bias depends on the physics input parameters (r_B, δ_B, γ) used to generate the pseudoexperiments. The biases could be explained by the fact that the yields in the suppressed modes are very small, with the presence of a large combinatorial background compared to the signal. Using the $A_{\pi K(\pi\pi)}$, $R_{\pi K(\pi\pi)}$ parametrisation, which measures the yield in both charges, resulted in more symmetrical statistical uncertainties compared to the $R^{\pm}_{\pi K(\pi\pi)}$ parametrisation. In summary, the $A_{\pi K(\pi\pi)}$, $R_{\pi K(\pi\pi)}$ parametrisation seems to present a better statistical behaviour in this case.

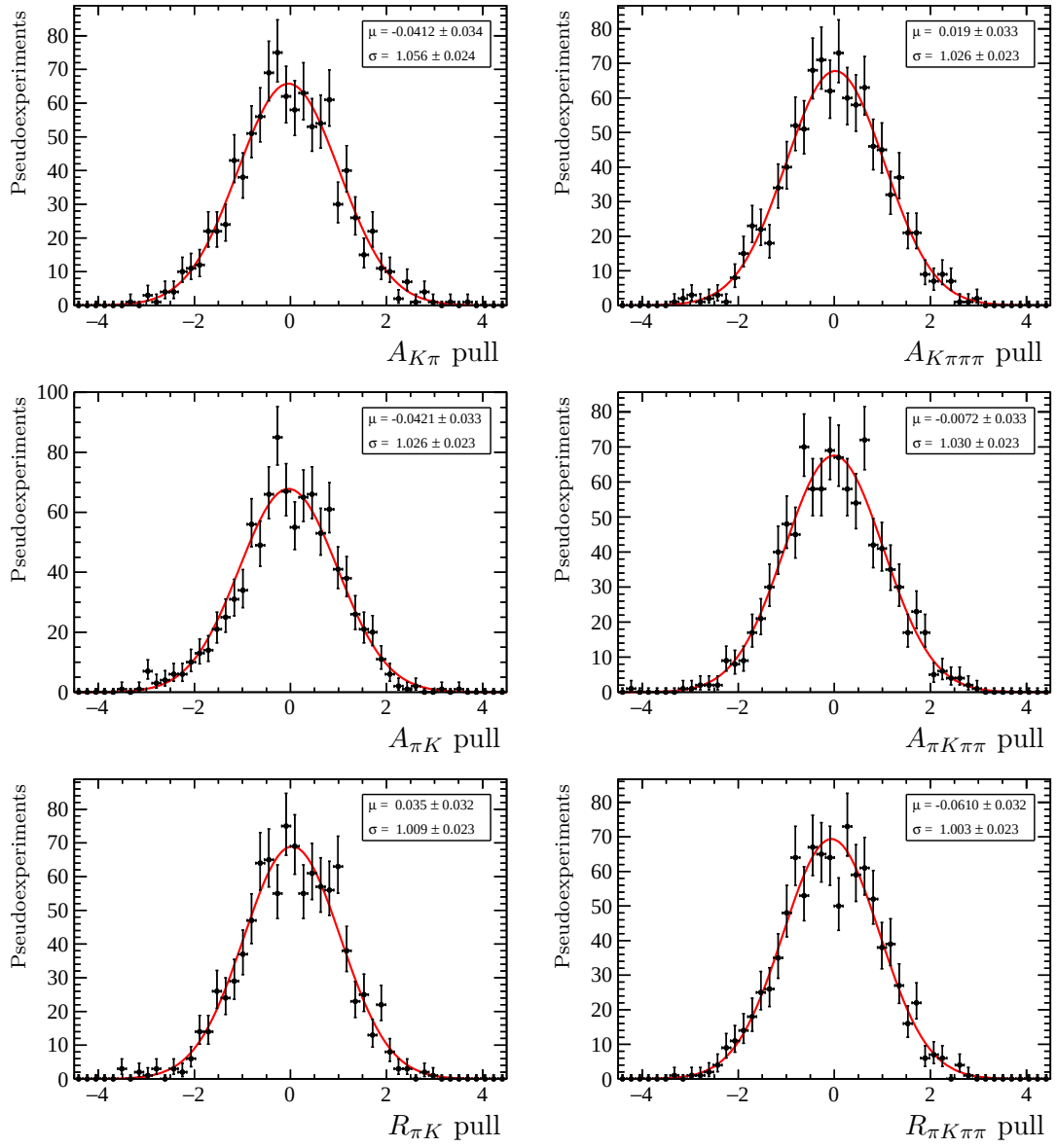


Figure 6.6: Pulls for the *CP* observables of the favoured and suppressed modes.

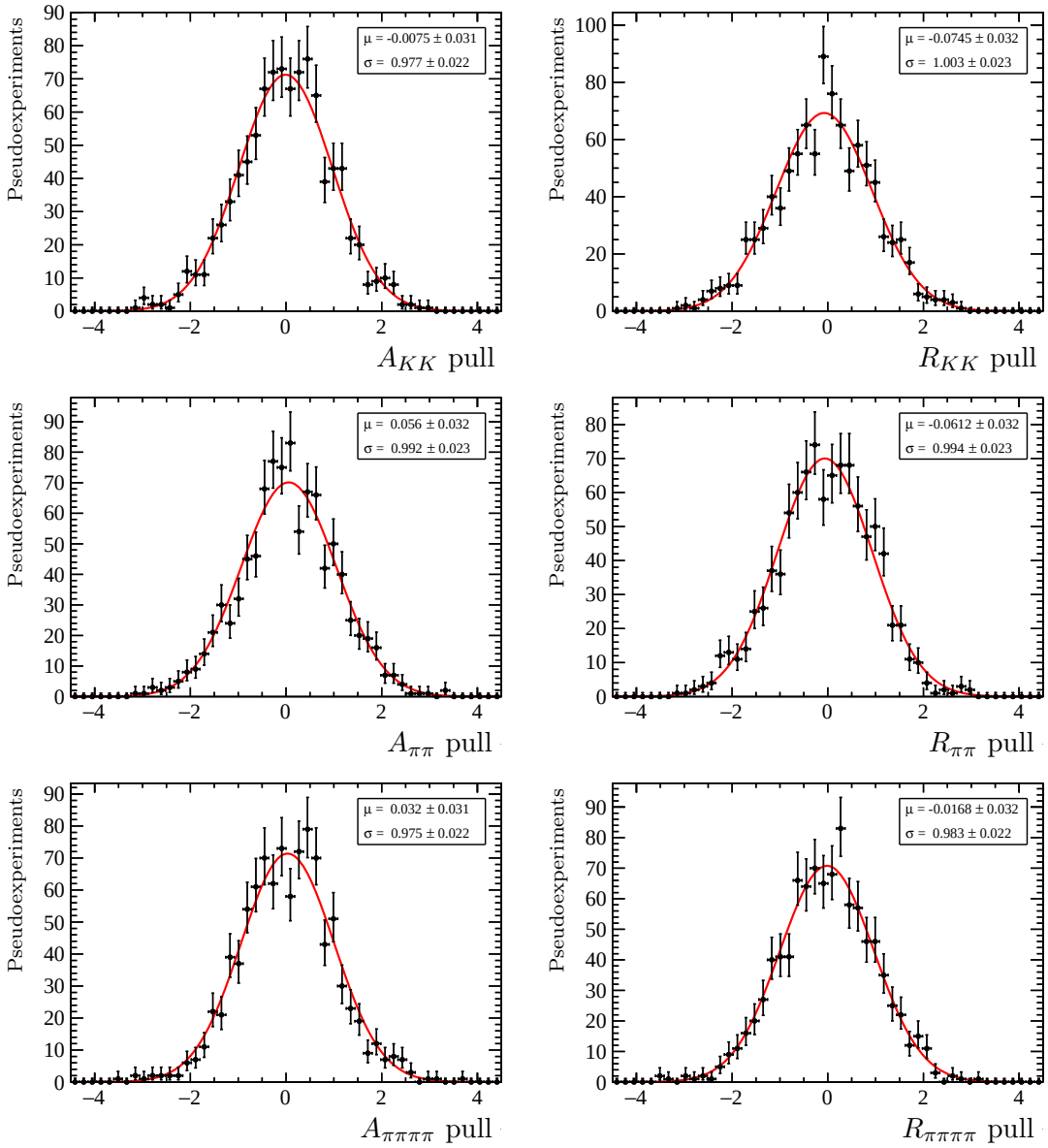


Figure 6.7: Pulls for the *CP* observables of the GLW modes.

6.4.3 First observation of suppressed modes

For the suppressed modes, an additional invariant-mass fit is performed. This consists in a charge-integrated invariant-mass fit in order to determine the significance of an observation of these modes, using Wilks' theorem [122]. The statement is that the significance is given by

$$\mathcal{Z} = \sqrt{-2\ln\left(\frac{\mathcal{L}_0}{\mathcal{L}_1}\right)}, \tag{6.36}$$

where $\mathcal{L}_0$ represents the likelihood of the null model, which consists of the default model but does not include the signal. The likelihood $\mathcal{L}_1$ represents the likelihood of the alternative model for which the signal component is included. The invariant-mass fit is performed twice, as shown in Figure 6.8, once without the signal component (bottom) and once with the signal (top), for both the πK (left) and $\pi K\pi\pi$ (right) modes. The statistical significance of an observation for the πK and $\pi K\pi\pi$ modes is found to be 5.8 and 5.6 standard deviations respectively. Therefore, this analysis provides the first observation of the suppressed $B^{\pm} \to [\pi^{\pm}K^{\mp}]_D K^{*\pm}$ and $B^{\pm} \to [\pi^{\pm}K^{\mp}\pi^{\pm}\pi^{\mp}]_D K^{*\pm}$ decays.

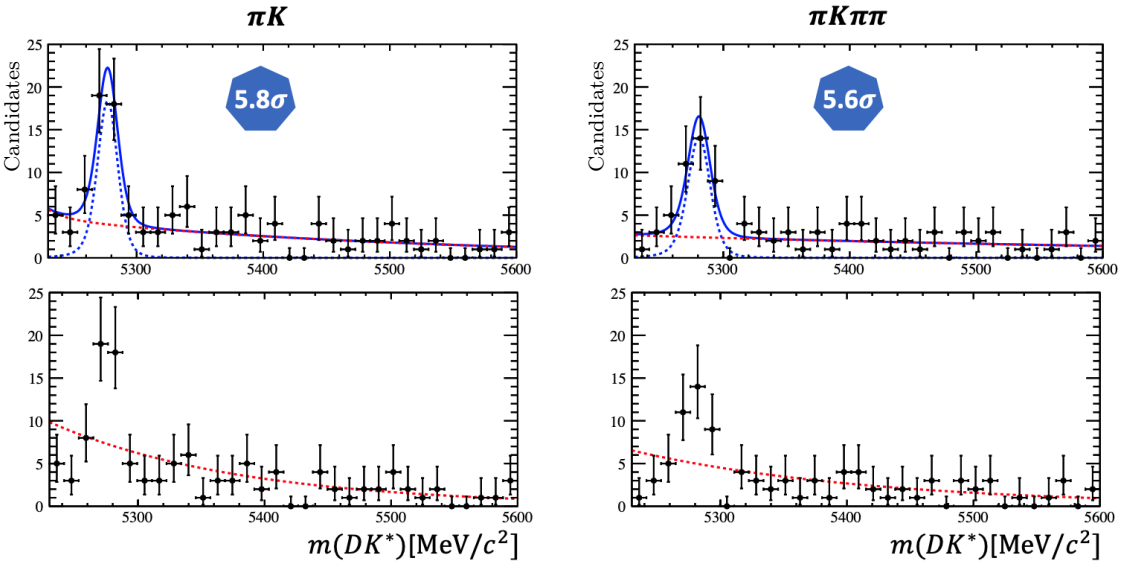


Figure 6.8: Charge-integrated invariant-mass distributions of the suppressed modes and the fits performed to determine the significance of an observation for both modes, using Wilks' theorem.

6.4.4 BPGGSZ *CP* fit

6.4.4.1 Intermediate fit

For the BPGGSZ modes, the *CP* fit is performed in two stages. First, an intermediate fit is done in the full integrated phase space, simultaneously in both $K^0_S\pi\pi$ and $K^0_S KK$ modes.

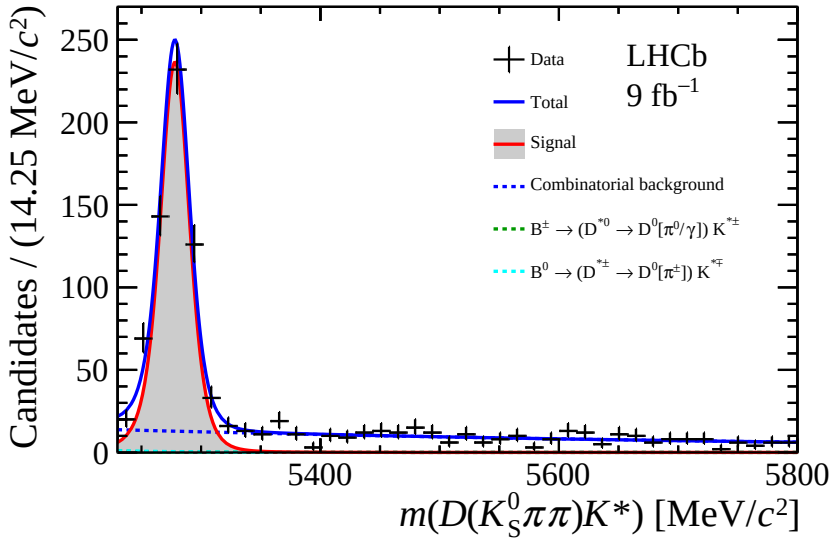


Figure 6.9: Invariant-mass distribution of the three-body $K^0_S\pi\pi$ mode in the integrated phase space with the fit result superimposed, using the full Run 1 and Run 2 dataset.

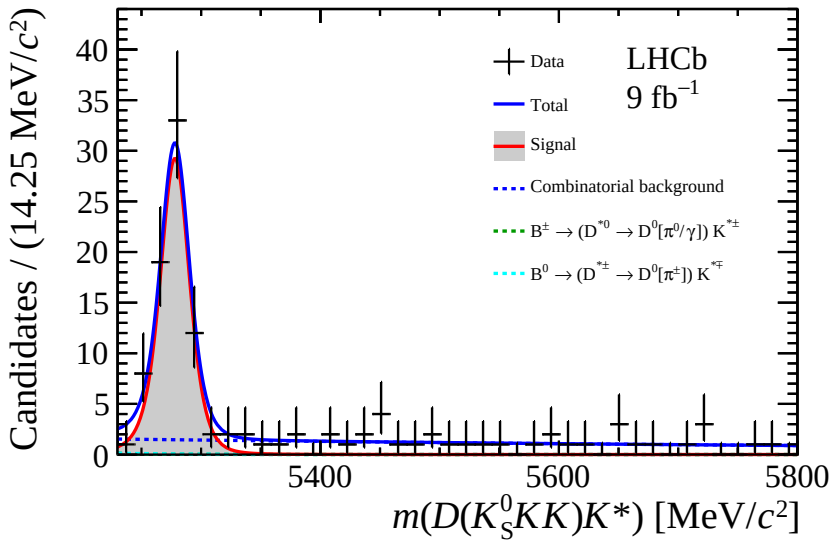


Figure 6.10: Invariant-mass distribution of the three-body $K^0_S KK$ mode in the integrated phase space with the fit result superimposed, using the full Run 1 and Run 2 dataset.

The fit projections are shown in Figures 6.9 and 6.10 for $K^0_S\pi\pi$ and $K^0_S KK$ respectively. This intermediate fit is used to obtain the total signal yield for both modes, as well as the combinatorial background shape to be used in the subsequent fit to determine the *CP* observables of interest. The floating parameters in this intermediate fit, for which the fit results are given in Table 6.9, are the following:

- Total signal yield of the $K^0_S\pi\pi$ and $K^0_S KK$ modes:
  $N^{sig}_{K^0_S\pi\pi}$, $N^{sig}_{K^0_S KK}$
- Signal mean and width for the 3-body final states:
  μ_{3body}, σ_{3body}
- Combinatorial background yield in the $K^0_S\pi\pi$ and $K^0_S KK$ modes:
  $N^{comb}_{K^0_S\pi\pi}$, $N^{comb}_{K^0_S KK}$
- Combinatorial slope parameter for the $K^0_S\pi\pi$ and $K^0_S KK$ modes:
  $\tau_{K^0_S\pi\pi}$, $\tau_{K^0_S KK}$.

Table 6.9: Fit parameter results obtained from the intermediate fit for the BPGGSZ modes.

Parameter	Fit value
$N^{sig}_{K^0_S\pi\pi}$	550 ± 27
$N^{sig}_{K^0_S KK}$	68 ± 9
$N^{comb}_{K^0_S\pi\pi}$	379 ± 24
$N^{comb}_{K^0_S KK}$	47 ± 8
μ_{3body}	5278.1 ± 0.7
σ_{3body}	11.6 ± 0.6
$\tau_{K^0_S\pi\pi}$	-0.0014 ± 0.0004
$\tau_{K^0_S KK}$	-0.0010 ± 0.0010

6.4.4.2 Fit in Dalitz bins

The second step of the fit procedure for the BPGGSZ modes is the fit in Dalitz bins, from which the *CP* observables $x_\pm$ and $y_\pm$ are measured. This is a simultaneous extended maximum likelihood fit performed in different categories defined by the

B charge, the D decay mode and the Dalitz bin. Given the 16 and 4 bins in the binning scheme used in this analysis for $K^0_S\pi\pi$ and $K^0_S KK$ respectively, there are $2 \times 16 = 32$ categories for $K^0_S\pi\pi$ and $2 \times 4 = 8$ categories for $K^0_S KK$. The free fit parameters are given here:

- *CP* observables:

  $x_{\pm}$, $y_{\pm}$

- Total signal yield of B^+ and B^- for each decay mode:

  $N^{+,sig}_{K^0_S\pi\pi}$, $N^{-,sig}_{K^0_S\pi\pi}$, $N^{+,sig}_{K^0_S KK}$, $N^{-,sig}_{K^0_S KK}$

- Signal mean and width for the 3-body final states:

  μ_{3body}, σ_{3body}

- Combinatorial background yield in each category.

The values of the external inputs c_i, s_i are listed in Table 6.10, while the F_i values are given in Table 6.11.

Table 6.10: Values of the c_i, s_i coefficients taken as external inputs [50–52].

bin i	c_i	s_i
	$D \to K^0_S\pi\pi$	
1	-0.037 ± 0.049	0.829 ± 0.097
2	0.837 ± 0.067	0.286 ± 0.152
3	0.147 ± 0.067	0.786 ± 0.154
4	-0.905 ± 0.021	0.079 ± 0.059
5	-0.291 ± 0.041	-1.022 ± 0.064
6	0.272 ± 0.082	-0.977 ± 0.176
7	0.918 ± 0.017	-0.184 ± 0.065
8	0.773 ± 0.033	0.277 ± 0.118
	$D \to K^0_S KK$	
1	0.713 ± 0.032	-0.107 ± 0.132
2	-0.758 ± 0.037	-0.394 ± 0.173

The fact that the yields of the two B charges are separately normalised implies that there is no dependence on production and detection asymmetries for the BPGGSZ

Table 6.11: Values of the F_i coefficients taken as external inputs [49].

bin i	LL	DD
	F_i for $D \to K^0_S\pi\pi$	
-8	0.024 ± 0.001	0.024 ± 0.000
-7	0.127 ± 0.001	0.133 ± 0.001
-6	0.062 ± 0.001	0.056 ± 0.001
-5	0.046 ± 0.001	0.042 ± 0.001
-4	0.095 ± 0.001	0.095 ± 0.001
-3	0.160 ± 0.001	0.160 ± 0.001
-2	0.153 ± 0.001	0.153 ± 0.001
-1	0.095 ± 0.001	0.097 ± 0.001
1	0.022 ± 0.001	0.020 ± 0.000
2	0.005 ± 0.000	0.005 ± 0.000
3	0.004 ± 0.000	0.004 ± 0.000
4	0.055 ± 0.001	0.056 ± 0.001
5	0.027 ± 0.001	0.022 ± 0.000
6	0.004 ± 0.000	0.003 ± 0.000
7	0.055 ± 0.001	0.057 ± 0.001
8	0.067 ± 0.001	0.072 ± 0.001
	F_i for $D \to K^0_S KK$	
-2	0.207 ± 0.004	0.202 ± 0.003
-1	0.222 ± 0.004	0.230 ± 0.003
1	0.290 ± 0.005	0.296 ± 0.003
2	0.281 ± 0.005	0.271 ± 0.003

modes. The yield in each Dalitz bin is parametrised using Equations (2.44)–(2.45) and the yields are related to the *CP* observables $x_\pm$ and $y_\pm$, which are measured through these equations. For the partially reconstructed backgrounds, the total yield determined by scaling appropriately the yield obtained from the global fit in the favoured mode is distributed across the Dalitz bins. The number of partially reconstructed background events is about 6 for $D \to K^0_S\pi\pi$ and 1 for $D \to K^0_S KK$. Assuming that B^- only decays to D^0, which is the favoured B meson decay, the background yield is multiplied by F_i, which corresponds to the fractional yield of flavour-tagged D^0 meson decays in bin i. Similarly, the background yield for B^+, assumed to decay to $\overline{D}^0$, is multiplied by F_{-i}. Due to the low yield of this background leaking into the signal region, the impact of this assumption would be negligible.

6.4.4.3 Fit results

The results obtained for the *CP* observables, where the uncertainties are statistical only, are given by

$$\begin{aligned} x_+ &= 0.003^{+0.051}_{-0.052} \\ y_+ &= 0.056^{+0.059}_{-0.064} \\ x_- &= 0.139^{+0.051}_{-0.060} \\ y_- &= -0.175^{+0.114}_{-0.103}\,. \end{aligned}$$

The corresponding correlation matrix of statistical uncertainties for the BPGGSZ modes is given in Table 6.12, which is obtained from the fit.

Table 6.12: Correlation matrix of statistical uncertainties on *CP* observables for three-body modes. Values below 0.1% are not shown and indicated with "–".

	x_+	y_+	x_-	y_-
x_+	1	−0.144	–	–
y_+		1	–	–
x_-			1	0.448
y_-				1

The yields obtained from the fit for each charge for the BPGGSZ modes are given in Table 6.13. The purity for both charged combined derived from the fit results within the region defined as $[5230, 5330]\,\mathrm{MeV}/c^2$ is also given. After validating the fit setup using pseudoexperiments, similarly to the ADS/GLW modes, very small biases were corrected for by looking at the difference between the mean value measured by the fit and the generated value for each *CP* observables.

Table 6.13: Observed signal yields for the BPGGSZ modes. The uncertainties are statistical only and the purity within the region defined as $[5230, 5330]\,\mathrm{MeV}/c^2$ is given.

D decay mode	B^-	B^+	Purity (%)
$K^0_S\pi\pi$	279 ± 18	268 ± 18	85
$K^0_S KK$	29 ± 6	40 ± 7	86

6.5 Systematic uncertainties

The final step in getting the values of the *CP* observables consists of the evaluation of the systematic uncertainties. Different sources of systematic uncertainties contribute to the results presented in this analysis. This section describes how these systematic uncertainties are computed, where two main methods are used.

To evaluate the systematic uncertainties due to fixed inputs in the *CP* fit, the *CP* fit to data is repeated thousands of times while varying the value of the given fixed parameter. The parameter is varied according to a Gaussian function whose width corresponds to the uncertainty on the parameter. For each *CP* observable measured in the fit, the impact of the fixed input is quantified by taking the standard deviation of the fit parameter distribution obtained from the ensemble of fits. This is assigned as the systematic uncertainty on the *CP* observable.

Systematic uncertainties related to the fit shapes are computed by generating thousands of pseudoexperiments with an alternate model and fitting back with the baseline fit. The difference between the generated value and the mean of the fit parameter distribution is taken as the systematic uncertainty to be assigned.

6.5.1 Asymmetry corrections

As described in Section 6.4.2.1, the asymmetry corrections are fixed inputs in the *CP* fit. The systematic uncertainties associated to their values are obtained by repeating the *CP* fit many times, while varying each time the value of the correction within its uncertainty. This is done separately for each of the three asymmetry corrections. Moreover, the additional source of kaon asymmetry due to the L0 hadron trigger is also taken into account. This is done by adding the value of $A_{K\pi}^{\mathrm{L0,trig}}$ to $A_{K\pi}$ while varying its value within its uncertainty.

6.5.2 Branching fractions

The branching fractions of the different D decay modes appear as fixed inputs in the definition of the *CP* observables, as detailed in Section 6.4.2.2. For the two-body

final states, these branching fractions are taken from the HFLAV average [123]:

$$\begin{aligned} \mathcal{B}(D^0 \to K^-\pi^+) &= (3.999 \pm 0.045)\% \\ \mathcal{B}(D^0 \to \pi^+\pi^-) &= (0.1490 \pm 0.0027)\% \\ \mathcal{B}(D^0 \to K^+K^-) &= (0.4113 \pm 0.0051)\% \\ \mathcal{B}(D^0 \to \pi^-K^+) &= (1.372 \pm 0.0017) \times 10^{-4}\,. \end{aligned} \tag{6.37}$$

These branching fractions are obtained from an averaging procedure, including final state radiation (FSR) corrections. From the corresponding fit, the correlation matrix for the branching fractions of the $K\pi$, $\pi\pi$ and KK modes is also obtained and given in Reference [123], and reproduced in Table 6.14. In this case, the systematic uncertainties for these two-body modes are computed by repeating the *CP* fit many times, varying the values of the branching fractions according to the correlation matrix.

Table 6.14: Correlation matrix for the two-body D decay modes obtained from the fit in the averaging procedure in Reference [123].

	$K^-\pi^+$	$\pi^+\pi^-$	K^+K^-
$K^-\pi^+$	1	0.77	0.76
$\pi^+\pi^-$		1	0.58
K^+K^-			1

For each of the other D decay modes, the systematic uncertainty associated to the branching fraction is computed by repeating the fit while varying the value according to a Gaussian with a width corresponding to the uncertainty on the branching fraction fixed value. For the four-body modes, the branching fraction values are taken from the PDG [28]:

$$\begin{aligned} \mathcal{B}(D^0 \to K^-\pi^+\pi^-\pi^+) &= (8.22 \pm 0.14)\% \\ \mathcal{B}(D^0 \to \pi^+\pi^-\pi^+\pi^-) &= (0.756 \pm 0.020)\% \\ \mathcal{B}(D^0 \to \pi^-K^+\pi^-\pi^+) &= (2.49 \pm 0.07) \times 10^{-4}\,. \end{aligned} \tag{6.38}$$

6.5.3 Efficiencies

Other fixed inputs which are included in the *CP* fit are the selection and PID efficiencies given in Section 4.3. The uncertainty on the selection efficiencies, which

are determined from simulation, is mainly due to the limited size of the signal MC samples used to compute the efficiencies. The systematic uncertainty related to the selection efficiencies is again computed by repeating the *CP* fit while varying the efficiencies within their uncertainty.

The PID efficiencies, which are computed using the PIDCalib2 package [107] and described in Section 4.3.3, have additional sources of systematic uncertainty on top of the statistical uncertainty due to the size of the samples. These are systematic effects due to the *sWeights* procedure used in the PIDCalib method, as well as the corresponding choice of the $(p,\ \eta)$ binning.

6.5.4 Signal shape

For the signal shape, only the shape parameters are fixed in the fit. The values and corresponding uncertainties are obtained from the MC fits described in Section 4.4.1. To evaluate a systematic effect on this component of the fit, the fit to data is repeated while varying the shape parameters within their uncertainty, taking correlations between the parameters into account by using the correlation matrix obtained from the MC fits. The standard deviation of the ensemble of fits is then assigned as the systematic uncertainty.

6.5.5 Combinatorial background

While the yield of the combinatorial background in each category is free in the fit, the slope parameter is shared across modes with the same number of final-state particles. To consider a systematic uncertainty for this assumption, the slope parameter is obtained by fitting the high-mass sideband of the B invariant-mass distribution separately for each D decay mode. These fits are shown in Figure 6.11. Pseudoexperiments are then generated using the combinatorial shapes obtained from these sideband fits, while fitting the pseudodatasets back with the default *CP* fit model. The systematic uncertainty is then taken to be the difference between the mean of the fit parameter distribution from the ensemble of pseudoexperiments and the generated value. While the combinatorial background in data is driven and

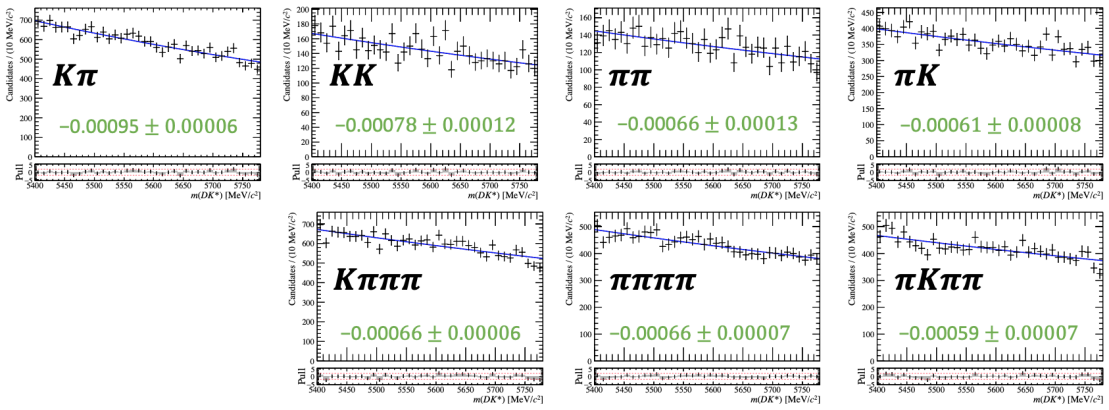


Figure 6.11: Fit to the high-mass sideband in the B invariant-mass distribution for each D decay mode in order to obtain the slope parameter separately for each decay.

determined by the low-mass region to a large extent, this study gives an estimate of the impact of the choice of the combinatorial background shape on the *CP* observables and the corresponding systematic uncertainty.

6.5.6 Partially reconstructed backgrounds

The small component coming from partially reconstructed backgrounds leaking into the signal region is fixed in the *CP* fit from the value obtained in the global fit. Different sources of systematic uncertainty are considered for this background, coming from the associated shape and yield used in the fit. The partially reconstructed background shape is entirely fixed from the MC fits, described in Section 4.4.3.2. The corresponding systematic uncertainty is computed in a similar way to the signal shape, taking the correlations between the different parameters into account using the correlation matrix obtained from the fit.

The yield of partially reconstructed background leaking into the signal region is fixed from the global fit (Section 6.3). A systematic effect related to the yield of partially reconstructed decays comes from the efficiency ratios which are fixed in the global fit, as can be seen in Equations (6.6)–(6.7). The spread in the yield obtained when repeating the fit while varying the efficiencies within their uncertainty is used to estimate the associated systematic effect.

Finally, another source of systematic uncertainty related to the yield comes from the fact that the yield is equally divided between B^+ and B^- in the *CP* fit,

assuming no *CP* violation in the partially reconstructed background. This choice is assigned a systematic uncertainty by introducing an asymmetry in the generation model of the pseudoexperiments. The *CP* violation in the partially reconstructed backgrounds is introduced by using the equations which define the *CP* observables in terms of the physics parameters and using the same parametrisation defined for the *CP* fit. The same equations as for the signal are used. The physics parameters are taken to be $r_B = 0.108$ and $\gamma = 63.8°$ [35]. However, the strong phases for the partially reconstructed backgrounds are unknown. A random number generator is thus used to generate three strong phases between [0, 360]° for the $B^{\pm} \to D^{*0}(\to D^0[\pi^0])K^{*\pm}$ decays, one for each helicity state. The strong phases for the $B^{\pm} \to D^{*0}(\to D^0[\gamma])K^{*\pm}$ decays are obtained by adding 180° (modulo 360°) to the generated strong phases, due to the 180° strong-phase difference between the $D^{*0} \to D^0\pi^0$ and $D^{*0} \to D^0\gamma$ decays. For the $B^0 \to D^{*\pm}(\to D^0[\pi^{\pm}])K^{*\mp}$ decays, r_B is fixed to zero and therefore the strong phases for these decays do not play any role. The default model is then used to fit back the generated pseudodatasets in order to obtain the systematic uncertainty.

The total systematic uncertainty for the partially reconstructed backgrounds is taken as the sum in quadrature between the different sources. The largest effect comes from the partially reconstructed background yield in the suppressed modes, in particular for the πK mode. This would be expected given the low yield and large asymmetry observed in the suppressed πK mode.

6.5.7 Charmless background

The shape and yield of the residual charmless background (Section 4.4.3.3) are completely fixed in the *CP* fit. The shape is obtained from the fits to simulation samples, while the yield is fixed from the data sidebands fits. These two sources of systematic uncertainties are computed in a similar way by varying the fixed values within their uncertainties.

6.5.8 Background from $\Lambda_b^0 \to \Lambda_c^+ K^{*-}$

The uncertainty on the Λ_b^0 background (Section 4.4.3.4) is computed in a similar way, by assessing the uncertainty coming from the shape and yield fixed in the KK mode for this background.

6.5.9 Background from $B_s^0 \to \overline{D}^0 K^{*0}$

The B_s^0 background present in the suppressed modes (Section 4.4.3.5) is also entirely fixed in the *CP* fit, that is both in terms of yield and shape. The systematic uncertainty associated to the yield is computed by varying the yield within its uncertainty. The shape is obtained by taking the average of the different resonances obtained from fast simulation. The systematic uncertainty is estimated by taking the shape of the resonance with the largest difference to the average shape.

6.5.10 BPGGSZ systematic uncertainties

Similarly, the systematic uncertainties are computed for the BPGGSZ modes, which are summarised in this subsection.

Fit model

The signal shape is assigned a systematic uncertainty by using an alternate PDF, which is a Crystal Ball [115] function which has different widths on both tails. Concerning the partially reconstructed backgrounds, the corresponding uncertainty is assessed by performing the fit without this component. These have a negligible impact and a small associated systematic uncertainty.

External inputs c_i and s_i

The c_i and s_i values are taken as external inputs. The associated systematic uncertainty is estimated by varying their values within their uncertainty, taking correlations into account.

Efficiency profile of strong phase inputs

The c_i and s_i inputs do not account for the efficiency profile in this analysis. Therefore, a systematic uncertainty is considered for the effect of the non-flat efficiency across the Dalitz plot. This is assessed by looking at the difference in results obtained when using c_i and s_i values from an amplitude model [48, 124] which assumes a flat efficiency and the values from the efficiency profile in this analysis obtained from simulation.

F_i from $B^\pm \to Dh^\pm$ analysis

This analysis used the F_i values measured from $B^\pm \to D\pi^\pm$ decays. Therefore, any differences in the efficiency across the Dalitz plot with respect to $B^\pm \to DK^{*\pm}$ decays are considered as a systematic uncertainty. This is done by reweighting the $F_{\pm i}$ values based on simulation and generating pseudoexperiments based on these values. These pseudodatasets are then fitted back with the nominal $F_{\pm i}$ values to assign a systematic uncertainty to account for this effect.

Coherence factor κ

A systematic uncertainty is assigned for the coherence factor κ, which is also fixed in the fit, by generating pseudoexperiments with $\kappa \pm \sigma_\kappa$ and fitting back with the original model.

Bin migration

Due to momentum resolution effects in the detector, candidates could be reconstructed in a wrong bin. This bin migration effect is, to first order, accounted for in the $F_{\pm i}$ values which are measured in data. Second-order effects would be due to *CP* violation differences between $B^\pm \to DK^{*\pm}$ and $B^\pm \to D\pi^\pm$ decays. This is considered as a systematic uncertainty by comparing the momentum resolution between the two decays in simulation.

Bias correction

The small bias corrections applied to the *CP* observables and mentioned in Section 6.4.4.3 are also assigned a systematic uncertainty. This is achieved by generating pseudoexperiments with varied values of the physics parameters and looking at the bias distribution to assign the systematic uncertainty.

Table 6.15: Summary of systematic uncertainties for the two- and four-body modes ($\times 10^{-2}$). Contributions which are less than 0.1% of the statistical uncertainty are not shown.

	$A_{K\pi}$	A_{KK}	$A_{\pi\pi}$	$A_{\pi K}$	R_{KK}	$R_{\pi\pi}$	$R_{\pi K}$	$A_{K\pi\pi\pi}$	$A_{\pi\pi\pi\pi}$	$A_{\pi K\pi\pi}$	$R_{\pi\pi\pi\pi}$	$R_{\pi K\pi\pi}$
Asymmetry corrections	0.17	0.072	0.067	0.078	–	–	–	0.17	0.073	0.16	–	–
Branching fractions	–	–	–	–	0.88	1.2	–	–	–	–	3.5	–
Selection efficiencies	–	–	–	–	0.87	0.76	0.0024	–	–	–	1.2	0.0047
PID efficiencies	–	–	–	–	0.22	0.23	–	–	–	–	0.36	–
Signal shape	–	–	0.046	0.067	0.20	0.26	0.0011	–	0.020	0.069	0.31	0.0021
Combinatorial shape	0.034	0.053	0.14	2.6	0.30	0.29	0.021	0.014	0.22	0.097	0.14	0.0071
Part. reco. background	–	–	–	0.16	0.072	0.12	0.0043	–	–	–	–	–
Charmless background	–	–	4.9	0.034	–	4.5	–	–	2.9	–	3.0	–
Λ_b^0 background	–	–	0.016	0.044	0.030	0.039	–	–	–	–	–	–
B_s^0 background	0.046	0.011	0.38	1.1	0.020	0.032	0.0093	0.038	0.12	0.54	0.27	0.0054
Total systematic	0.18	0.09	4.9	2.8	1.31	4.7	0.02	0.17	2.9	0.5	4.8	0.01
Statistical	1.4	4.0	9.0	16.4	5.0	9.0	0.19	1.8	6.0	21.8	7.0	0.26

6.5.11 Summary of systematic uncertainties

A summary of the systematic uncertainties for the ADS/GLW modes is given in Table 6.15, where contributions which are less than 0.1% of the statistical uncertainty are not shown. The largest systematic uncertainty is related to the charmless background. Another systematic effect which is considerable is the one associated to the branching ratios for the four-body modes. The total systematic uncertainty is computed by taking the sum in quadrature of the different sources of systematic uncertainty. It can be seen that the statistical uncertainty (which is also included in the table for reference) is larger than the systematic uncertainty for all *CP* observables.

The systematic uncertainties for the BPGGSZ modes are listed in Table 6.16. In this case, the leading contribution on the $y_\pm$ observables comes from the external c_i and s_i inputs. The $x_\pm$ are less impacted due to the better precision on the c_i coefficients compared to s_i. The systematic uncertainty coming from the F_i values is the largest for the $x_\pm$ observables. Overall, the analysis is statistically limited for both sets of D decay modes with the current available dataset.

Table 6.16: Summary of systematic uncertainties for the three-body modes ($\times 10^{-2}$).

Source	$\sigma(x_+)$	$\sigma(y_+)$	$\sigma(x_-)$	$\sigma(y_-)$
c_i, s_i uncertainty	0.4	1.9	0.9	3.9
F_i inputs	1.5	0.4	1.7	0.4
Value of κ	0.8	0.4	0.6	0.8
Efficiency correction to c_i, s_i	0.0	0.0	0.2	0.6
Bin migration	0.4	0.2	0.3	0.4
Mass model	0.1	0.1	0.1	0.3
Bias correction	0.4	0.6	0.3	0.6
LHCb syst. uncertainty	1.8	0.9	1.9	1.3
Total systematic.	1.8	2.1	2.1	4.1
Statistical	5.2	6.4	6.0	11.4

To determine the correlation matrix of the systematic uncertainties, the correlation coefficient for each pair of *CP* observables is computed from the ensemble of fits/pseudoexperiments. In this way, the correlation matrix for each source of

systematic uncertainty can be built and converted into the corresponding covariance matrix. The total correlation matrix of systematic uncertainties is then obtained by summing the different sources. The correlation matrices of systematic uncertainties for the ADS/GLW and BPGGSZ modes are given in Tables 6.17 and 6.18 respectively. The largest correlations for the ADS/GLW modes are between the *CP* observables of the same mode, similarly to the statistical correlations, which is expected from the fact that the asymmetry and ratio *CP* observables are directly related in their definitions. Moreover, the favoured modes are connected to all the other modes through the *CP* fit parametrisation and thus present larger correlations. The largest correlation occurs between the two asymmetries in the favoured modes, $A_{K\pi}$ and $A_{K\pi\pi\pi}$. These two modes are the only modes where the yield is free in the fit and the charge-combined yield in both determines the overall normalisation. In the BPGGSZ modes, the correlations are mainly due to the correlations in the uncertainty on the c_i, s_i measurements.

Table 6.17: Correlation matrix of systematic uncertainties on *CP* observables for two- and four-body modes. Values smaller than 0.1% are not shown and indicated with "–".

	$A_{K\pi}$	A_{KK}	$A_{\pi\pi}$	$A_{\pi K}$	R_{KK}	$R_{\pi\pi}$	$R_{\pi K}$	$A_{K\pi\pi\pi}$	$A_{\pi\pi\pi\pi}$	$A_{\pi K\pi\pi}$	$R_{\pi\pi\pi\pi}$	$R_{\pi K\pi\pi}$
$A_{K\pi}$	1	0.314	–	−0.014	–	–	−0.016	0.918	0.010	−0.169	–	–
A_{KK}		1	–	−0.017	−0.039	–	−0.011	0.319	0.032	0.155	−0.019	0.048
$A_{\pi\pi}$			1	–	–	−0.292	–	–	–	–	–	–
$A_{\pi K}$				1	–	–	0.220	−0.017	–	0.086	–	0.023
R_{KK}					1	0.051	–	–	–	–	–	–
$R_{\pi\pi}$						1	–	–	–	–	–	–
$R_{\pi K}$							1	–	–	−0.013	–	0.063
$A_{K\pi\pi\pi}$								1	0.013	−0.173	–	0.013
$A_{\pi\pi\pi\pi}$									1	–	−0.605	0.025
$A_{\pi K\pi\pi}$										1	–	0.345
$R_{\pi\pi\pi\pi}$											1	–
$R_{\pi K\pi\pi}$												1

Table 6.18: Correlation matrix of systematic uncertainties on *CP* observables for three-body modes.

	x_+	y_+	x_-	y_-
x_+	1	-0.055	0.030	0.022
y_+		1	-0.158	-0.478
x_-			1	0.319
y_-				1

6.6 Summary of results

This section summarises the results obtained from the *CP* fit. The results for the *CP* observables, including the statistical and systematic uncertainties, are given by:

$$
\begin{aligned}
A_{K\pi} &= -0.024 \pm 0.014 \pm 0.002 \\
A_{KK} &= 0.14 \pm 0.04 \pm 0.001 \\
A_{\pi\pi} &= 0.31 \pm 0.09 \pm 0.05 \\
A_{\pi K} &= -0.73 \pm 0.16 \pm 0.03 \\
R_{KK} &= 1.10 \pm 0.05 \pm 0.01 \\
R_{\pi\pi} &= 0.96 \pm 0.09 \pm 0.05 \\
R_{\pi K} &= 0.0098 \pm 0.0019 \pm 0.0002 \\
A_{K\pi\pi\pi} &= -0.024 \pm 0.018 \pm 0.002 \\
A_{\pi\pi\pi\pi} &= 0.04 \pm 0.06 \pm 0.03 \\
A_{\pi K\pi\pi} &= -0.19 \pm 0.22 \pm 0.01 \\
R_{\pi\pi\pi\pi} &= 1.05 \pm 0.07 \pm 0.05 \\
R_{\pi K\pi\pi} &= 0.0118 \pm 0.0026 \pm 0.0001
\end{aligned}
$$

$$
\begin{aligned}
x_+ &= 0.003^{+0.051}_{-0.052} \pm 0.018 \pm 0.004 \\
y_+ &= 0.056^{+0.059}_{-0.064} \pm 0.009 \pm 0.019 \\
x_- &= 0.139^{+0.051}_{-0.060} \pm 0.019 \pm 0.009 \\
y_- &= -0.175^{+0.114}_{-0.103} \pm 0.013 \pm 0.039
\end{aligned}
$$

6.7 Combination and interpretation of results

The final step of the analysis consists in the interpretation of the results in terms of the physics parameters of interest r_B, δ_B and γ. These are determined from the values of the *CP* observables determined in the *CP* fit, as well as the uncertainties and the corresponding correlation matrices. The *CP* observables are related to the physics parameters through Equations (2.18)–(2.23), (2.29), (2.34)–(2.35), (2.38)–(2.40) and Equations (2.46)–(2.47) for the ADS/GLW and BPGGSZ modes respectively. A summary of the parameters involved in this system of equations is given here:

- Physics parameters to be measured:
  r_B, δ_B, γ
- *CP* observables obtained from the *CP* fit
  $A_{K\pi}$, A_{KK}, $A_{\pi\pi}$, $A_{\pi K}$, R_{KK}, $R_{\pi\pi}$, $R_{\pi K}$, $A_{K\pi\pi\pi}$, $A_{\pi\pi\pi\pi}$, $A_{\pi K\pi\pi}$ $R_{\pi\pi\pi\pi}$, $R_{\pi K\pi\pi}$,
  x_+, x_-, y_+, y_-
- Required inputs, for which the values are given in Table 6.19:
  κ, $r_D^{K\pi}$, $\delta_D^{K\pi}$, $r_D^{K3\pi}$, $\delta_D^{K3\pi}$, $\kappa^{K3\pi}$, F_+

Table 6.19: Values of the external inputs used in the combination, which are the B decay coherence factor [36], the two-body D decay parameters [35], the four-body D decay parameters taken as inputs from a combination of results from BESIII, CLEO-c and LHCb collaborations [43], and the fraction of *CP*-even states for the $\pi\pi\pi\pi$ mode [40].

Parameter	Value
κ	0.95 ± 0.06
$r_D^{K\pi}$	$0.05865^{+0.00014}_{-0.00015}$
$\delta_D^{K\pi}$	$(190.2 \pm 2.8)^\circ$
$r_D^{K3\pi}$	0.0550 ± 0.0007
$\delta_D^{K3\pi}$	$(161^{+28}_{-18})^\circ$
$\kappa^{K3\pi}$	$0.44^{+0.10}_{-0.09}$
F_+	$0.735 \pm 0.015 \pm 0.005$

The external D decay parameters which are required as external inputs are taken from published measurements. In particular, the two-body D decay parameters $r_D^{K\pi}$,

$\delta_D^{K\pi}$ are taken from Reference [35], while the four-body parameters $r_D^{K3\pi}$, $\delta_D^{K3\pi}$, $\kappa^{K3\pi}$ are taken from a combination of measurements performed by the BESIII, CLEO-c and LHCb collaborations [43]. For the $\pi\pi\pi\pi$ mode, the required CP-even fraction F_+ is taken from Reference [40], measured by BESIII. Moreover, the correlation matrices for the D decay parameters are also required and used in the combination.

The physics parameters are determined using the GammaCombo package [125, 126], which is used to compute confidence intervals of parameters when combining measurements. This framework implements two statistical methods to determine the parameters of interest, which are the Prob and Plugin methods. The former is a simpler profile-likelihood method, while the latter makes use of pseudoexperiments to obtain frequentist intervals for the parameters [127], based on the Feldman-Cousins approach [128]. The Plugin method is considered to provide a better coverage compared to the Prob method. In this section, the results are first determined separately for the ADS/GLW and BPGGSZ modes, before performing a combination of all results.

6.7.1 ADS/GLW

In the ADS/GLW modes, there are multiple solutions due to the more complicated trigonometric relations between the *CP* observables and the physics parameters of interest. In particular, two solutions are found in the range $[0, 180]^\circ$ for γ. The one-dimensional scan of γ is presented in Figure 6.12, where both the Prob (shaded area) and Plugin (data points) methods are shown. Moreover, the two-dimensional scans of the physics parameters are presented in Figure 6.13. These plots show the contours corresponding to the 1σ, 2σ bands. While the preferred solution from the fit sits at around 120°, the solution which is compatible with the world average is found to be

$$\gamma = (58^{+7}_{-6})^\circ$$

$$r_B^{DK^{*\pm}} = 0.102^{+0.010}_{-0.012}$$

$$\delta_B^{DK^{*\pm}} = (57^{+6}_{-7})^\circ.$$

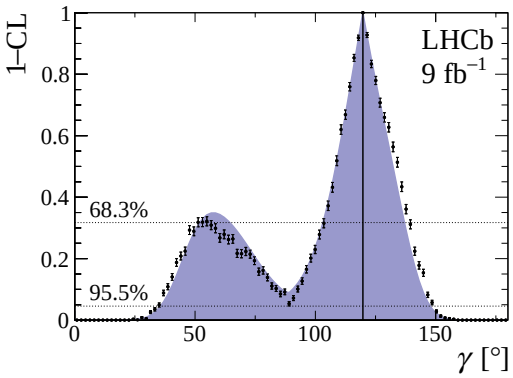


Figure 6.12: One-dimensional scan on the confidence limits for γ for ADS/GLW modes, where both the PROB (shaded area) and PLUGIN (data points) methods are shown.

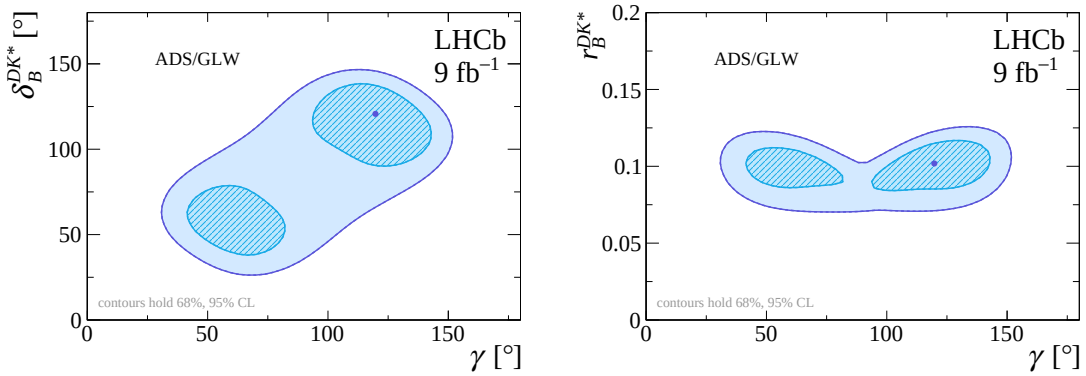


Figure 6.13: Two-dimensional physics parameter scans showing the contours corresponding to the 1σ, 2σ bands for ADS/GLW modes.

In order to show the improvement in precision of this new analysis with respect the the partial Run 2 results, corresponding to about 5 fb^{-1}, the two-dimensional scans are performed for both sets of results and shown in Figure 6.14.

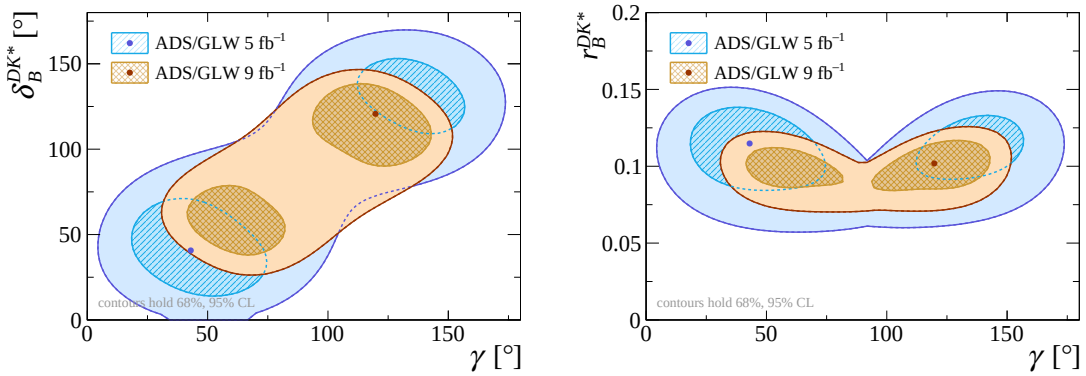


Figure 6.14: Two-dimensional physics parameter scans showing the 1σ, 2σ contours for ADS/GLW modes, compared to results from the Run 2 partial dataset.

6.7.2 BPGGSZ

For the BPGGSZ modes, the one-dimensional scan for γ and the two-dimensional contours are presented in Figures 6.15 and 6.16 respectively.

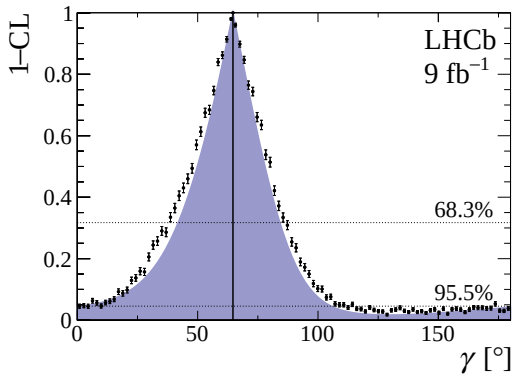


Figure 6.15: One-dimensional scan on the confidence limits for γ for BPGGSZ modes, where both the PROB (shaded area) and PLUGIN (data points) methods are shown.

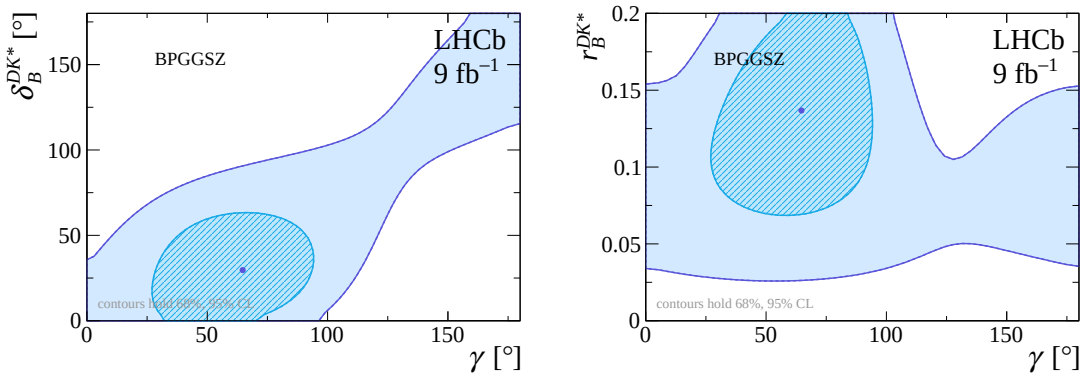


Figure 6.16: Two-dimensional physics parameter scans showing the contours corresponding to the 1σ, 2σ bands for BPGGSZ modes.

While no values of γ are excluded at the 95% confidence level, there is a single solution for γ, which is less precise than the one obtained from the ADS/GLW modes and found to be

$$\gamma = (65^{+21}_{-26})^{\circ}$$

$$r_B^{DK^{*\pm}} = 0.137^{+0.047}_{-0.050}$$

$$\delta_B^{DK^{*\pm}} = (30 \pm 24)^{\circ}.$$

6.7.3 Combination

Finally, all results are combined in order to determine the physics parameters of interest. By combining the results from ADS/GLW and BPGGSZ modes, the overall sensitivity can be improved. In particular, the BPGGSZ result helps to break the multiple solutions from the ADS/GLW result, which is more precise. Indeed, a single precise solution is found from the combination, for which the one-dimensional γ scan is shown in Figure 6.17 (left). It can be seen that the same result is found with both the PROB and PLUGIN methods. Figure 6.17 (right) summarises the one-dimensional scans obtained from both ADS/GLW and BPGGSZ results, as well as the combination.

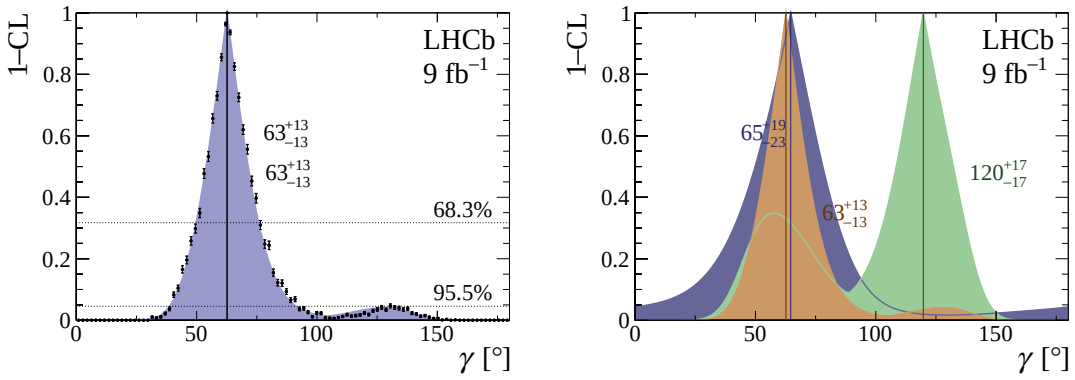


Figure 6.17: One-dimensional γ scan with the PROB (shaded area) and PLUGIN (data points) methods for the combination (left), and ADS/GLW modes (green), BPGGSZ modes (blue) and the combination (orange) one-dimensional scans (right).

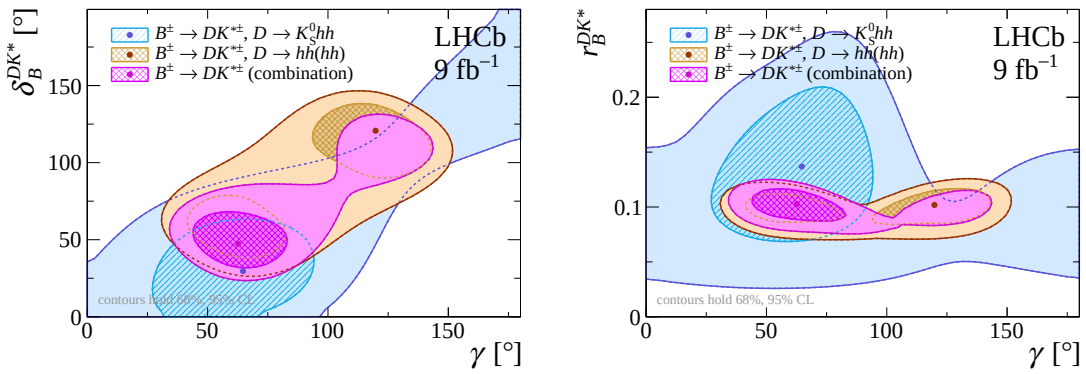


Figure 6.18: Two-dimensional physics parameter scans showing the 1σ, 2σ contours from the combination. Yellow, blue and magenta contours show the results for two- and four-body modes, three-body modes and all D decay modes, respectively.

All the two-dimensional physics parameter scans showing the 1σ, 2σ contours are summarised and presented in Figure 6.18. It can be seen that the combination is powerful in improving the overall precision on γ. The values obtained for the physics parameters from the fit are

$$\gamma = (63 \pm 13)^{\circ}$$

$$r_B^{DK^{*\pm}} = 0.103 \pm 0.010$$

$$\delta_B^{DK^{*\pm}} = (47^{+14}_{-12})^{\circ}.$$

This model-independent result of γ is consistent with the LHCb γ combination and the world average. This is a valuable addition to the family of γ measurements and is included in the latest 2024 LHCb combination [2].

Not everything that can be counted counts,
and not everything that counts can be counted.

— Albert Einstein

7 The LHCb Upgrade detector

A summary of the changes made for the LHCb Upgrade are presented in this chapter. This highlights the main changes with respect to the LHCb experiment detailed in Chapter 3. On top of the analysis presented in this thesis, a significant project was performed for the LHCb Upgrade and is the subject of Chapter 8.

7.1 Running conditions in Run 3

At the end of Run 2 in December 2018, the LHCb experiment underwent an upgrade of its detector. During this period called Long Shutdown 2 (LS2) until 2022, different sub-detectors were removed or upgraded in order to be able to operate at a higher instantaneous luminosity compared to Run 2. This represents a factor five increase, corresponding to a luminosity $\mathcal{L} = 2 \cdot 10^{33}\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ and an average number of pp collisions of 5 per visible event. The upgraded LHCb detector [129, 130] is designed to be able to run with a 40 MHz read-out, for which one of the main changes consists in a full-software trigger system upgraded such that a full reconstruction of the events at the LHC non-empty bunch crossing rate is possible.

7.2 LHCb Upgrade detector

In order to cope with the Run 3 conditions which offer a more challenging environment and be able to read out the full detector at the full collision event rate, parts of the LHCb detector needed to be removed or replaced. Moreover, new trigger-less electronics replaced the electronics of the sub-detectors to be adapted to the read-out bunch crossing rate. Figure 7.1 presents the layout of the LHCb detector in Run 2 (left) and the Upgrade (right), where the main changes are highlighted.

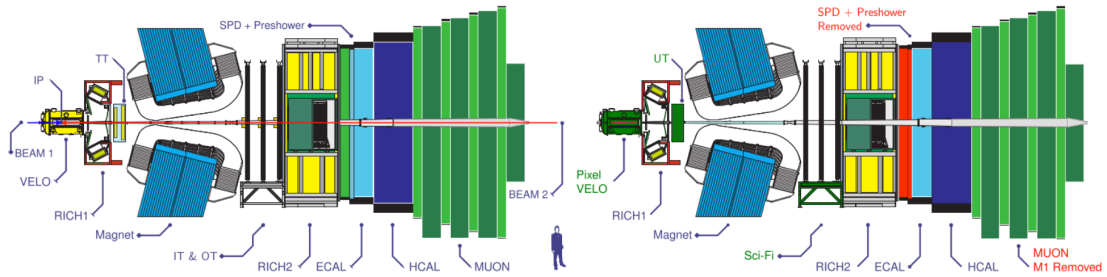


Figure 7.1: The layout of the LHCb detector (left) and the Upgrade (right) [58, 130].

7.2.1 Tracking system Upgrade

The VELO silicon strip detector was replaced by pixel sensors placed closer to the beam [131]. The pixel silicon tracker is made of 52 modules covering the two halves, which can come to a distance of about 5 mm from the beam when closed. Instead of the TT, an Upstream Tracker (UT) is used in the tracking system [132]. The UT consists in four layers of silicon strip detectors, which can collect track hits closer to the beam axis due to a thinner pitch size around the beam pipe. The IT and OT are replaced by a Scintillating Fibre Tracker (SciFi). The SciFi is made of plastic fibre mats arranged in 12 layers in total, 4 layers of three stations each. The read-out is performed by using silicon photomultipliers (SiPMs). The same X–U–V–X geometry is kept for the UT and SciFi.

7.2.2 Particle identification system Upgrade

For the particle identification system in the Upgrade [133], the main change is the upgrade of the electronics to accommodate the increased luminosity and radiation.

To accommodate the full read-out, multi-anode PMTs are used in the RICH detectors for the photon detection to obtain a higher granularity. While the layout of the RICH detectors is kept unchanged, the occupancy in the inner region is reduced in RICH1 by increasing the focal length. The SPD and PS from the calorimeter system were removed as the L0 hardware trigger is removed in the Upgrade. Concerning the muon system, the first station M1 was also removed due to the increased occupancy in the Upgrade which makes the association of hits to muon tracks in M1 more challenging. Moreover, to reduce the hit occupancy in M2, shielding is added between the HCAL and the muon chambers.

7.2.3 Trigger system Upgrade

At higher luminosities, the L0 hardware trigger displays a saturation of the efficiency for hadronic channels. It is therefore removed and replaced by the upgraded trigger [134, 135] which consists of a software only trigger. The trigger scheme for Run 2 (left) and the Upgrade (right) is shown in Figure 7.2.

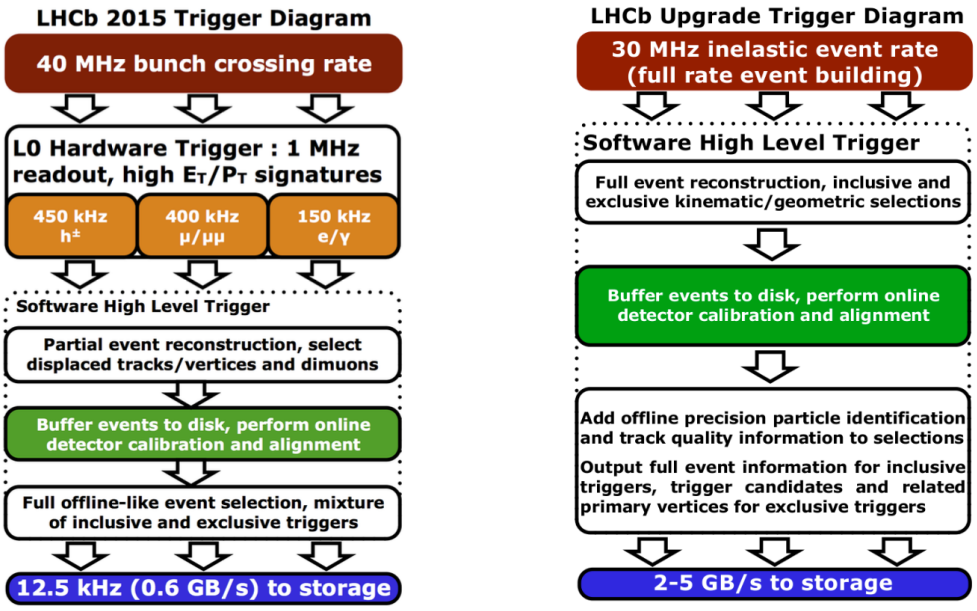


Figure 7.2: The trigger system layout in Run 1–2 (left) and the Upgrade (right) [136].

From the nominal LHC bunch crossing rate, 10 MHz correspond to empty crossings, leaving 30 MHz of inelastic event rate. The goal is to perform a read-out of

this full collision rate using the full software trigger, where requirements can be placed in order to reduce the high input rates at LHCb (see Section 3.3.3). The upgraded trigger system consists in a two-stage trigger which is similar to the Run 2 software trigger.

7.3 Physics motivations and trigger efficiency

With the upgraded detector and the ability to read out the detector at the full collision rate, it will be possible to advance the physics goals in Run 3. In the era of high precision measurement, many analyses which are currently statistically dominated will benefit from the increased dataset, while others would become possible for the first time. For example, analyses of rare decays will take advantage of the larger datasets and new channels with low branching fraction would become accessible.

The removal of the hardware trigger in Run 3 is especially advantageous for hadronic final states such as those used for γ measurements. Figure 7.3 shows the trigger efficiencies for $B^+ \to \overline{D}^0(K^+\pi^-)\pi^+$ decays in Run 3 (2024) data and in Run 2 [137]. It can be seen that there is a large increase in yields at low transverse momentum, which is an important point to improve the sensitivity to *CP* violation and rare processes.

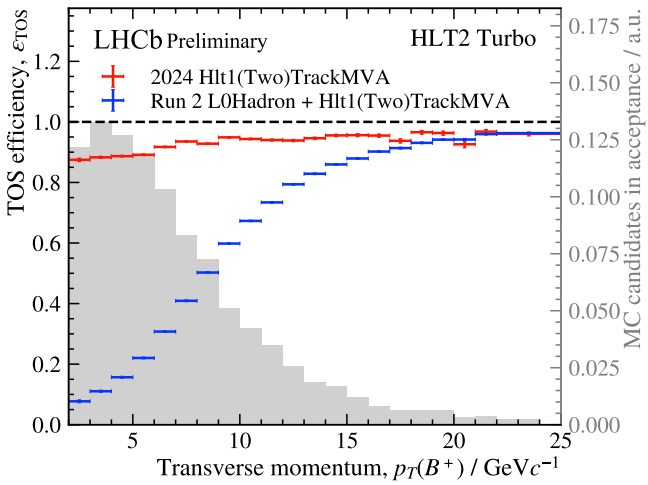


Figure 7.3: Trigger efficiency for $B^+ \to \overline{D}^0(K^+\pi^-)\pi^+$ decays in Run 3 data (red) and in Run 2 (blue) as a function of transverse momentum [137].

What we observe is not nature itself,
but nature exposed to our method of questioning.

— Werner Heisenberg

8 Particle identification developments for the LHCb Upgrade

This chapter presents a significant project that was carried in parallel to the analysis, contributing to the Real-Time Alignment project for LHCb. The main PID calibration sample for kaons and pions used in LHCb was studied in detail and the code developed was integrated into the new PID calibration framework to be used in Run 3. This was development work in preparation for Run 3 and hence the testing and improvement are performed on Run 2 data which was the only data available at the time. The chapter ends by showing the PID performance on the small amount of Run 3 data collected in 2023. With the increase of instantaneous luminosity by a factor 5 in Run 3, the LHCb Upgrade detector was designed to accommodate this higher data rate, as detailed in Chapter 7.

8.1 PID calibration at LHCb

Correctly identifying particles is a key aspect of particle physics experiments and PID plays an important role in many LHCb analyses. For example, rejecting background coming from misidentified B-meson decays is crucial in searches for rare decays. For instance, in the $B^0_{(s)} \to \mu^+\mu^-$ rare decays, it is important to efficiently remove

backgrounds coming from misidentified $B^0_{(s)} \to h^+h'^-$ decays, with $h^{(\prime)} = \pi, K$ [138]. The invariant-mass distribution for this example is shown in Figure 8.1.

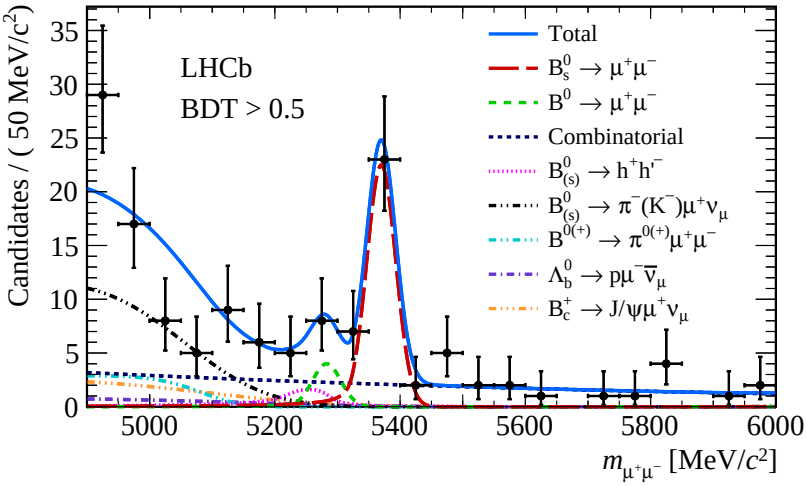


Figure 8.1: Invariant-mass distribution for $B^0_{(s)} \to \mu^+\mu^-$ decays. The different components of the fit are also shown, in particular backgrounds coming from misidentified decays such as from $B^0_{(s)} \to h^+h'^-$ decays [138].

Another example concerns final states with the same topology for which the events need to be classified correctly in the different categories, which are mutually exclusive. This is the case in $B^0 \to h^+h^-$ analyses where cross-feed backgrounds need to be correctly determined from a calibration of the PID variables.

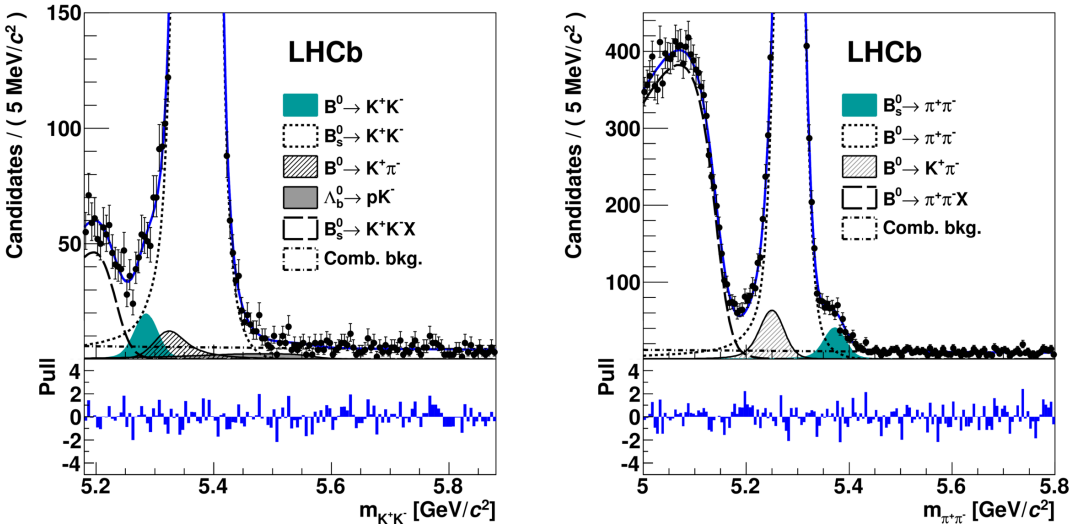


Figure 8.2: Invariant-mass distribution for the $B^0 \to K^+K^-$ (left) and $B^0_s \to \pi^+\pi^-$ (right) decays [139].

Figure 8.2 presents the invariant-mass distribution for the $B^0 \to K^+K^-$ (left) and $B_s^0 \to \pi^+\pi^-$ (right) decays [139]. In this case, the PID efficiencies are used to accurately determine the yield of the cross-feed backgrounds in the data samples. The usage and importance of PID requirements and efficiencies can also be seen in the $B^\pm \to DK^{*\pm}$ analysis presented in this thesis, where the main difference in selection between the different D decay modes considered is the PID requirement applied on the D daughter particles. It is therefore crucial to control and evaluate without bias the efficiencies of these requirements applied in the selection.

By combining information from the sub-detectors, in particular from the two RICH detectors, PID quantities are formed, which give information about the type of a particle going through the detector. Simulating the response of the PID detectors is complex due to its dependence on the kinematics of the particle going through the detector and the detector occupancy, which has a sensitivity to the beam conditions and can therefore vary between different events. The response of the PID detectors also depends on the conditions of the experiment such as the temperature, gas pressure and detector alignments, which can also change between different runs of data-taking [107]. Therefore, using MC samples of the signal decay of interest to directly determine PID efficiencies would only give an approximation.

8.2 Calibration method

Due to the fact that the PID variables are known not to be perfectly reproduced in simulation samples, a data-driven method is used to determine the PID performance [107] and presented in this section.

8.2.1 Requirements of the calibration

The response of the PID detectors, and therefore the PID performance, is non-uniform and depends on various variables, even for particles of the same type. In particular, the PID response depends on the momentum of the particle, the pseudo-rapidity and the track multiplicity, which corresponds to the number of tracks present in the event. Therefore, different analyses performed at LHCb use

different values of the PID requirements applied in the selection. This is the reason why a flexible calibration system is needed in order for analysts to be able to obtain PID efficiencies for the given cuts used in their analyses. The analysts need to have the ability to choose any cut to calibrate. Therefore, the calibration needs to provide efficiency tables and not a number for a fixed PID requirement. Moreover, the data-driven calibration must be analyst friendly and easy to use, provide accurate PID efficiencies and be computationally fast. It is also important to preserve the reproducibility of the results given by the calibration.

The data-driven approach consists of using specific decays where no PID information is needed in order to infer the particle species. Different calibration samples are used depending on the type of particle on which PID requirements are placed, as well as on the kinematic region of interest. In addition, these calibration samples need to have large yields in order to reduce the statistical uncertainties, as well as high purity to reduce the associated systematic uncertainties. Any residual background must be removed or accurately modelled in order to provide pure, correctly identified, signal tracks. This is achieved by performing invariant-mass fits to the calibration data, modelling both signal and background components. The calibration samples are selected without PID requirements in order to make sure that no bias is introduced.

8.2.2 $D^{*+} \to D^0\pi^+$ calibration sample

The main calibration sample for kaon and pion tracks used in LHCb is $D^{*+} \to D^0\pi^+$, where $D^0 \to K^-\pi^+$. This decay is chosen such that the hadrons can be identified without PID requirements. In this case, the flavour of the D^0 meson is tagged by the charge of the bachelor slow pion coming from the strong decay of the D^*, where $D^{*+} \to \overline{D}^0\pi^+$ is forbidden. The topology of this decay is shown in Figure 8.3, where both the D^{*+} (left) and D^{*-} (right) cases are presented. In turn, the kaon and pion coming from the D decay can be distinguished, based on the determined flavour of the D^0, since $D^0 \to K^+\pi^-$ is doubly Cabibbo-suppressed. This is shown in Figure 8.4, where the Feynman diagrams are given for both the favoured (left) and suppressed

(right) D^0 decays. In almost all cases, the D^0 daughter pion is therefore the track with the same charge to the slow pion, while the kaon has the opposite charge.

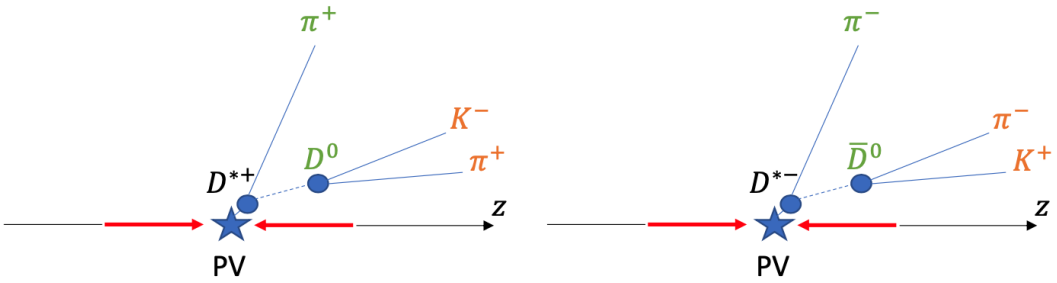


Figure 8.3: Topologies of the D^{*+} (left) and D^{*-} (right) decays. The pion produced in the strong decay of the D^* tags the flavour of the D meson, distinguishing in turn the charge of the kaon and pion from the D decay.

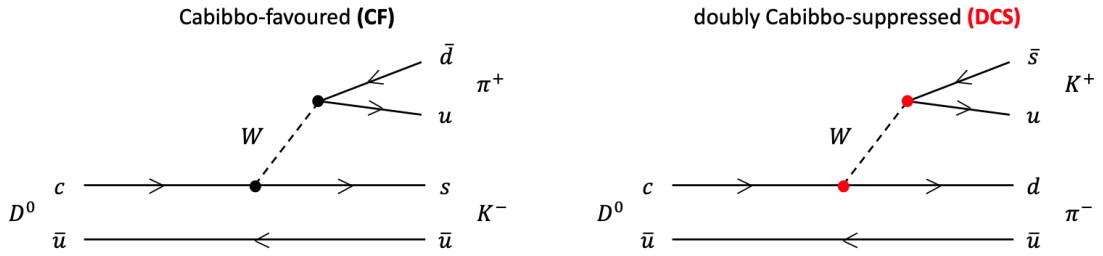


Figure 8.4: Feynman diagrams for the Cabibbo-favoured $D^0 \to K^-\pi^+$ (left) and doubly Cabibbo-suppressed $D^0 \to K^+\pi^-$ (right) decays.

Thanks to the large charm production in the LHC pp collisions, an abundance of these decays is collected within the LHCb detector. In this high-yield sample, the residual background sources must be identified and dealt with.

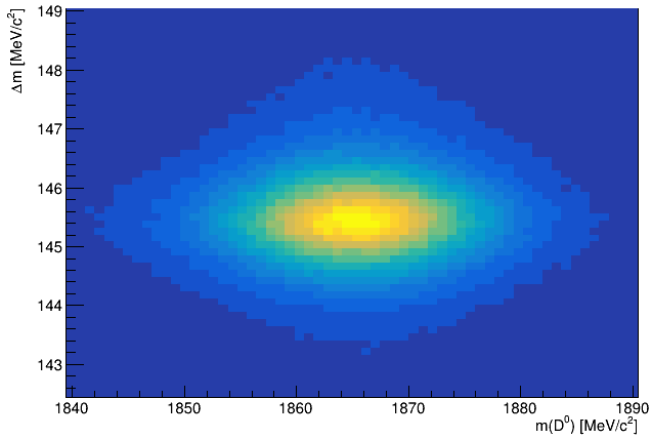


Figure 8.5: Two-dimensional distribution of $m(D^0)$ and Δm in data.

This is achieved by performing a 2D invariant-mass fit to the data, presented in Section 8.8, using the variables $m(D^0)$ and $\Delta m = m(D^*) - m(D^0)$, shown in Figure 8.5.

8.2.3 Fit-and-count method

The efficiency could be computed by using the fit-and-count technique. This consists in performing the fit to the invariant-mass distribution before and after the PID cut, in order to obtain the amount of signal passing the PID requirement. This implies that the data, signal and background, must be well modelled for the specific cut applied. It is known that distributions, especially the background, vary significantly depending on the PID cut, as illustrated in Figure 8.6 for a specific PIDK requirement on the probe pion.

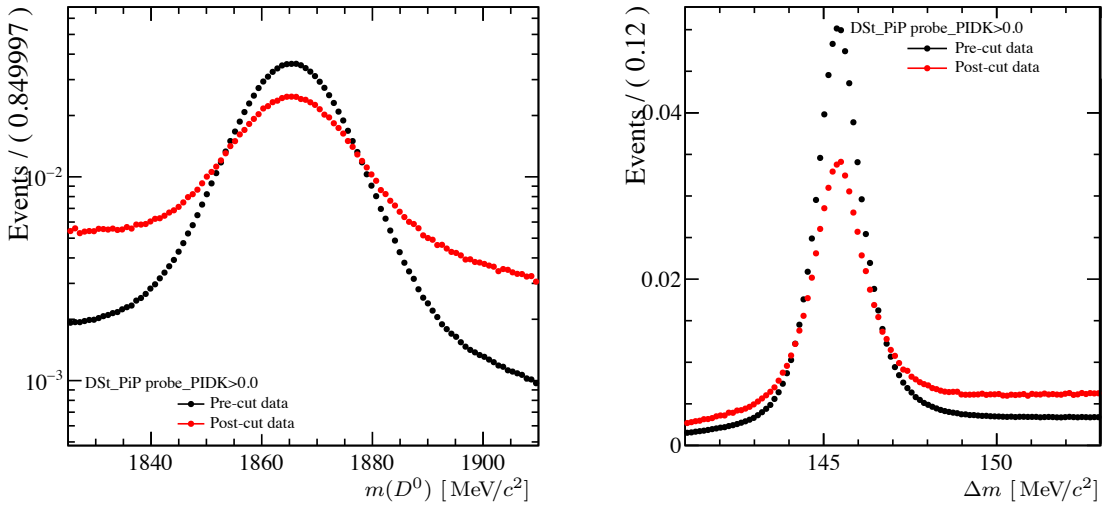


Figure 8.6: Invariant-mass distributions for $m(D^0)$ (left) and Δm (right) for a specific PID requirement.

The main issue with the fit-and-count method is that the misidentified background shapes change significantly for different PID requirements. Therefore, the fit-and-count method does not satisfy the calibration requirements as a detailed study for each PID requirement would need to be performed.

8.2.4 *sWeights* method

A flexible calibration is obtained by performing the calibration using the *sPlot* technique [109], with which the residual backgrounds are modelled and removed

statistically. Given the two sources of events in the data, signal and background, the goal is to unfold the two contributions coming from the two different sources. To this aim, a discriminating variable is used for which the PDFs used to model the different sources are known, in order to disentangle the two components in another reconstructed variable. This method assumes that the reconstructed variable used is uncorrelated with the discriminating variable for each source.

In this case, the two-dimensional invariant-mass distribution is used as a discriminating variable to model the signal and background candidates. A first fit is then performed to determine the yield N_i for each source s. For each source, the PDF can be written as f_i, while $f_i(m_e)$ corresponds to the value of the PDF f_i for a given event e associated with values m_e for the discriminating variables m. The second step consists in giving signal weights, called *sWeights*, depending on the probability for the candidate to be a signal or background event. For each source, the *sWeights* for a given event are defined as

$$W_e^s = \frac{\sum_{j=1}^{N_s} \boldsymbol{V}_{sj}\, f_j(m_e)}{\sum_{k=1}^{N_s} N_k f_k(m_e)}. \tag{8.1}$$

The number of sources is given by $N_s = 2$ and $\boldsymbol{V}$ represents the covariance matrix defined as

$$\boldsymbol{V}_{sj}^{-1} = \sum_{e=1}^{N} \frac{f_s(m_e) f_j(m_e)}{(\sum_{k=1}^{N_s} N_k f_k(m_e))^2}, \tag{8.2}$$

where the total number of events is given by N. The sum of the *sWeights* corresponds to the number of signal candidates in the sample. In order to determine the efficiency of a given PID requirement, the strategy is to sum all the *sWeights* W_i before and after the PID cut. This gives the ratio of the amount of signal before and after applying the PID selection, which corresponds to the efficiency:

$$\epsilon_{\text{PID}}\,(\text{cut}) = \frac{\sum_i W_i\,(\text{cut})}{\sum_i W_i}. \tag{8.3}$$

The use of the *sWeights* to compute the efficiency allows analysts to use this technique for different PID cuts, without performing a detailed analysis for each PID requirement considered. This comes with a systematic uncertainty, which is currently

estimated to be of the order of 0.1% per track as an absolute uncertainty [107]. This is negligible and appropriate for most analyses, while the few analyses requiring a better precision in the PID efficiency determination need to perform additional studies using the fit-and-count method.

The calibration method using the *sWeights* satisfies the requirements and is implemented in the PIDCalib package [107], which provides the LHCb analysts with the tools to compute the PID efficiencies from the calibration data samples, or perform a reweighting of MC samples to be used in the analysis. Indeed, the provided *sWeights* are also made available as an input into the reweighting package, which allows analysts to correct the PID response of MC samples [140]. Therefore, the obtained *sWeights* feed into almost every single analysis for LHCb from the proton-proton data.

For Run 3, PIDCalib2 was developed and superseded the original PIDCalib package. The new package was written in Python, with no dependence on the Root software. This transition allowed a gain of speed compared to the original. With the increase of yields in Run 3 and the required level of precision, it is therefore important to ensure that the calibration data is well described by the PDF used in the *sPlot* and ensure a smooth integration of the new PID calibration framework to be used in Run 3. The aim of this project is to provide the new model for the main source of calibration tracks used for kaons and pions, and implement it in the new framework. Two main goals for the new model are to find a more accurate and robust fit model, as well as reduce any potential biases from the *sWeights* procedure.

8.3 Kinematic variables and binning scheme

The PID response of a particle depends on the kinematics since the momentum p is related to the RICH Cherenkov angle and the deposited energy in the calorimeter. The PID response also depends on the occupancy and the conditions of the detector. For the calibration procedure, it is assumed that this quantity can be completely parametrised by a given number of variables, such as the momentum of the track p, the pseudo-rapidity η and the track multiplicity. The value of the PID variable

can then be assumed to be constant within small portions of the sample, if the sample is divided using sufficiently small intervals for the binning. In this way, the efficiency can be determined in each subsample and each track whose kinematics belong to this subsample can be assigned the same efficiency as the PID response is assumed to be fully parametrised in these binning variables. In the PIDCalib package, this is provided as a histogram containing the per-event efficiencies.

The strategy in Run 2 was to perform the invariant-mass fit in the full (p, η) phase space for the extraction of the *sWeights*. This approach requires that the distribution is uniform across the phase space in these kinematic variables. However, it was found that the invariant-mass distributions vary as a function of the kinematics. Not only are the shape and the resolution of the signal different in different regions of the phase space, but the background shapes vary as well. This can be seen in Figure 8.7, which shows the invariant-mass distributions of $m(D^0)$ (left) and Δm (right) for the probe pion for two regions of the phase space in data.

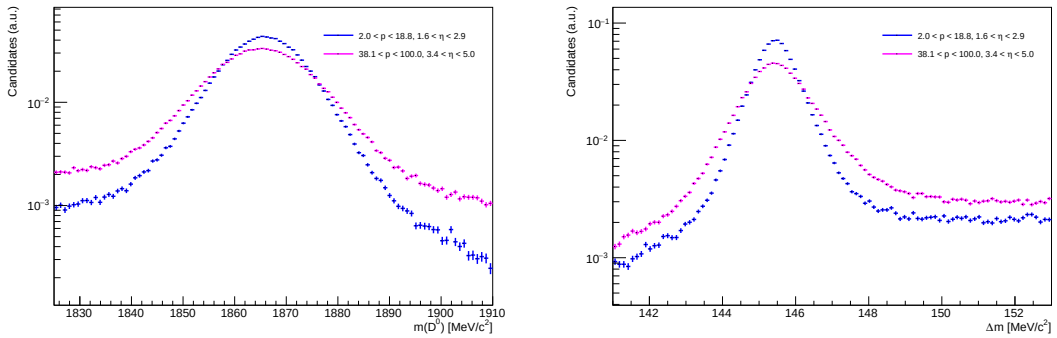


Figure 8.7: Comparison of the invariant-mass distributions in data for $m(D^0)$ (left) and Δm (right) for two regions of the phase space, where p is expressed in GeV/c.

This motivates a change of the fit procedure and the new strategy proposed for Run 3 is to perform the fit in different kinematic bins in order to reduce any uncertainty coming from the *sWeights* procedure. Figure 8.8 presents the kinematic phase space for the probe kaon (left) and the probe pion (right) in simulation. This phase space was studied in order to choose the binning scheme for the fit procedure. The goal was to design a binning scheme such that the number of signal events in each bin is approximately the same, where no bin contains a much higher or much

lower number of the signal candidates. Different binning schemes were studied, such as 3×3, 4×2, 4×3 and 4×4 bins in p and η respectively. The phase space is first binned in the first variable and then optimised in the second. The selected binning scheme is shown in Table 8.1 and superimposed in Figure 8.8. A regular binning scheme was not chosen to avoid regions with a small number of events, which were merged with adjacent bins so that the signal fraction in each bin is at least around 10%, with some bins containing up to 25% of the signal.

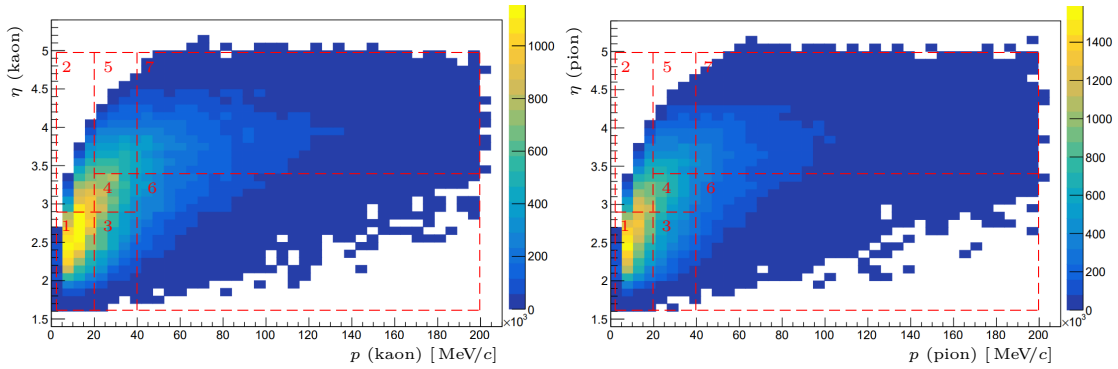


Figure 8.8: Two-dimensional distribution of p and η in simulation for the probe kaon (left) and the probe pion (right), with the binning scheme superimposed.

Table 8.1: Binning scheme in the (p, η) phase space.

		η bins		
		1.6 - 2.9	2.9 - 3.4	3.4 - 5.0
p bins [GeV/c]	2 - 19	Bin 1	Bin 2	
	19 - 39	Bin 3	Bin 4	Bin 5
	39 - 200	Bin 6		Bin 7

This means that the fit is studied in these different kinematic regions, which is expected to give a better precision with more accurate *sWeights* and therefore a smaller associated systematic error. The systematic uncertainties on the fit-and-count method are of the same order as the difference that one might expect from these two *sWeights* type pictures and therefore there is no reliable mechanism in which it is possible to know that the binned fit is a better description. However, due to the high yield in this channel, performing the fit in kinematic bins helps for the stability of the fit and is beneficial and worth doing for the calibration procedure. Moreover, with the other improvements in the fit model, as described and shown in

the next sections, the fits are also better in each kinematic region and it is therefore expected to yield better fit results overall compared to the Run 2 procedure.

8.4 Signal selection and shape

The first step is to select a sample of $D^{*+} \to D^0\pi^+$ decays, without using PID variables. Dedicated HLT2 trigger lines are used to obtain the raw data, where basic preliminary requirements are applied, such as on the momentum, or the track and vertex quality. For example, the signal candidates are selected by applying momentum and transverse momentum requirements on the particles, as well as on the quality of the impact parameter of tracks. Flight distance requirements are also applied such that the D^0 meson is well separated from the decay vertex. The HLT2 line is `Hlt2PIDD02KPiTagTurboCalib` in Run 2 [141], while this is renamed as `Hlt2PID_DstToD0Pi_D0ToKPi` for Run 3, where some adjustments are made to accommodate the new rate.

Additional offline cuts are added before proceeding with the fit, in order to remove background from misidentified or swapped D^0 daughters. Vetoes are added to remove events that can be reconstructed as $D^0 \to \pi^+\pi^-$, $D^0 \to K^+K^-$ or $D^0 \to \pi^-K^+$ within 25 MeV/c^2 of the D^0 mass:

$$\begin{aligned} &|m(K^- \to \pi^-, \pi^+ \to \pi^+) - m_{D^0}| > 25\,\text{MeV}/c^2, \\ &|m(K^- \to K^-, \pi^+ \to K^+) - m_{D^0}| > 25\,\text{MeV}/c^2, \\ &|m(K^- \to \pi^-, \pi^+ \to K^+) - m_{D^0}| > 25\,\text{MeV}/c^2. \end{aligned} \tag{8.4}$$

These vetoes are illustrated in Figure 8.9, showing the invariant-mass distribution of the signal as well as the wrong mass distributions on which the vetoes are placed. It can be seen that the D^0 mass is restricted relatively tightly in order to remove the remaining background from misidentified decays, which can be seen in red and green in the low and high mass range for $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ respectively. The double misidentified background, or swapped daughters, from the DCS $D^0 \to \pi^-K^+$ decay peaks in the same place as the signal and is easily removed by the veto. The fit range, [1825, 1910] MeV/c^2, is chosen to minimise these components on

the sides of the mass range to avoid modelling any remaining contribution, while keeping enough of the tails of the D^0 mass distribution.

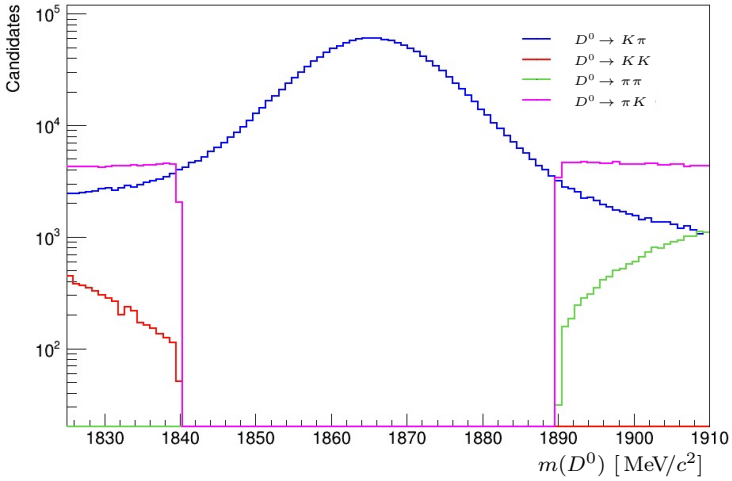


Figure 8.9: Vetoes used to remove background from misidentified or swapped D^0 daughters.

The vetoes applied do not sculpt the signal shape. This can be seen in Figure 8.10, which shows the D^0 invariant-mass distribution before and after applying the vetoes. However, applying the different vetoes to the samples can distort the background distributions. This needs to be taken into account when studying the backgrounds to be modelled in the fit.

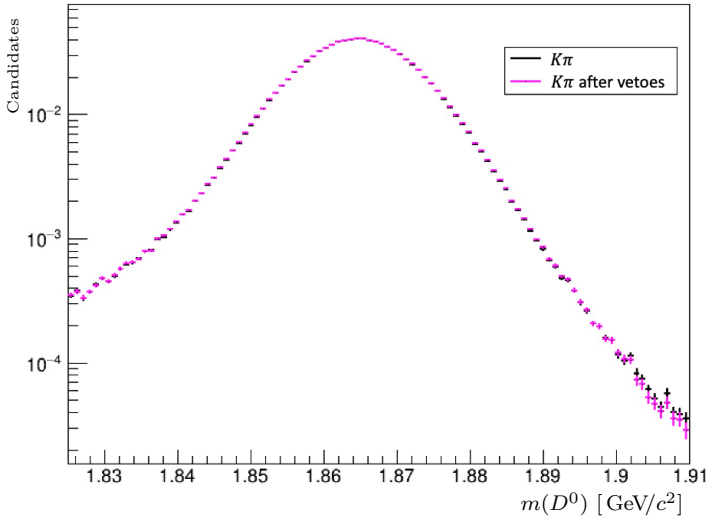


Figure 8.10: Invariant-mass distribution of the $D^0 \to K\pi$ signal before and after applying the vetoes.

The signal shape was studied with Run 2 MC samples. The D^0 invariant-mass

distribution was modelled with two Gaussians. Concerning the Δm distribution, the old parametrisation was a sum of a Crystal Ball function and two Gaussians. The result using this old model is shown in Figure 8.11.

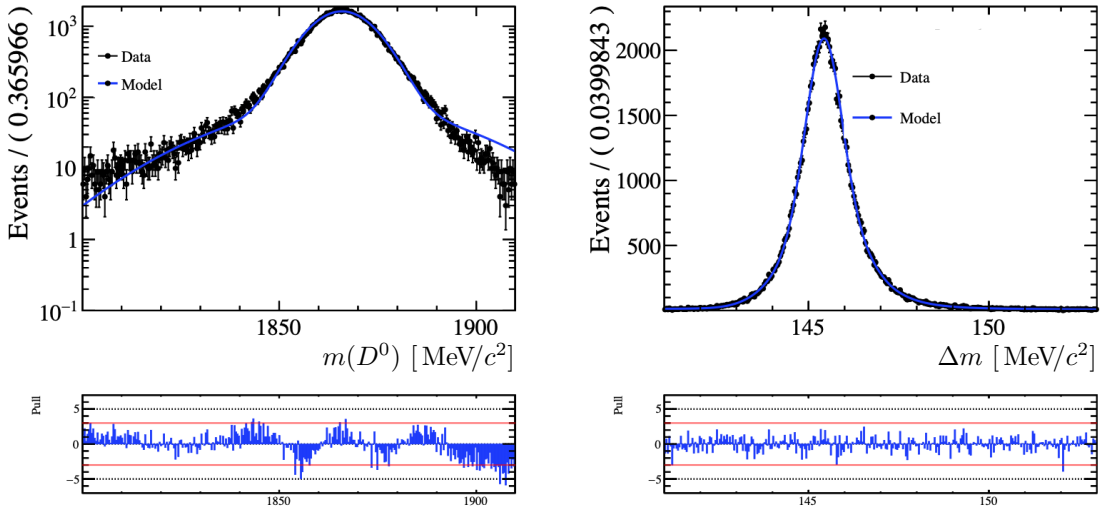


Figure 8.11: Projections of the 2D fit to the truthmatched MC distribution for D^0 (left) and Δm (right) using the old model.

The parametrisation used in Run 2 was changed, motivated by 2D fits to the MC samples with different PDFs. Many different shapes were studied in order to find the PDF providing the best fit quality and stability. The revisited model is a sum of the modified Cruijff function and a Gaussian. The modified Cruijff function, which is defined as

$$C(m;\,\mu,\sigma,\alpha_L,\alpha_R,\beta) \propto \begin{cases} \exp\left[\frac{-(m-\mu)^2\left(1+\beta(m-\mu)^2\right)}{2\sigma^2+\alpha_L(m-\mu)^2}\right] & \text{if } m-\mu<0 \\ \exp\left[\frac{-(m-\mu)^2\left(1+\beta(m-\mu)^2\right)}{2\sigma^2+\alpha_R(m-\mu)^2}\right] & \text{if } m-\mu>0\,, \end{cases} \tag{8.5}$$

is used to account for the radiative tail, which models well the LHCb resolution. This function allows an asymmetry in the tails of the signal shape, which improves the fit quality, evaluated by looking at the pulls of the fit. Similarly, the Δm model was changed to a simple modified Cruijff function. The result of the projections of the 2D fit to the truthmatched MC distribution is shown in Figure 8.12 for D^0 (left) and Δm (right).

The improvement of the quality of the fit for the D^0 projection with the new model compared to the old one can also be seen in Figure 8.13, which presents MC

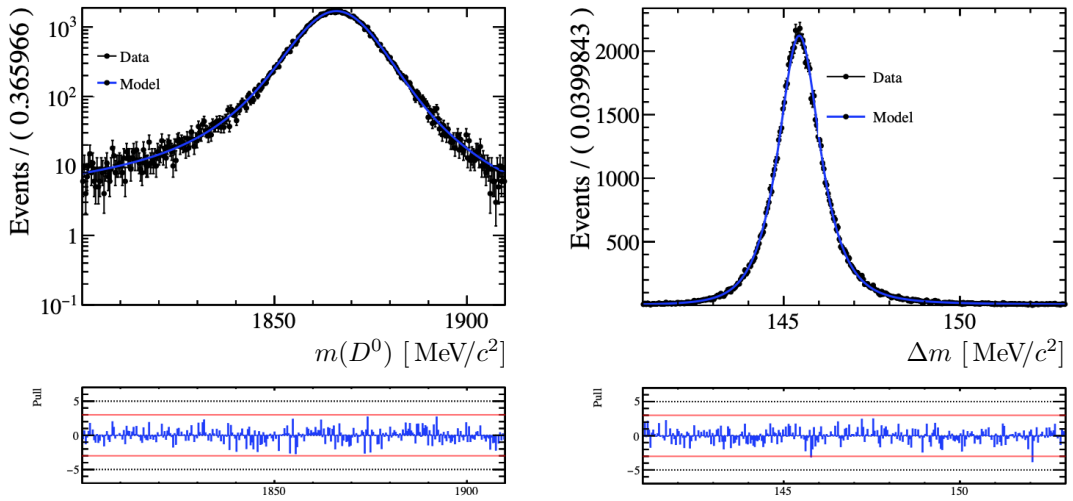


Figure 8.12: Projections of the 2D fit to the truthmatched MC distribution for D^0 (left) and Δm (right) using the new model.

fits to the D^0 invariant-mass distributions. This shows how the signal shape, with asymmetric tails, can impact the other components of the 2D fit model which is described later. In particular, the new signal shape prevents the background from compensating for any missing or mismodelled signal in the tails.

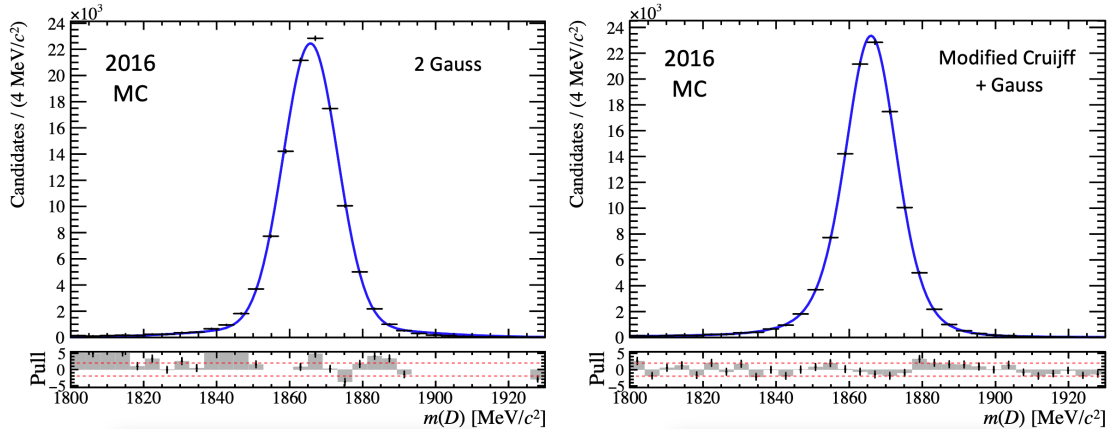


Figure 8.13: Distribution of the D^0 mass in simulation, with the fit result superimposed for the old model (left) and new model (right).

8.5 Backgrounds

After the selection of the calibration samples, three sources of background remain in the data:

- **Random** π: Real D meson combined with random soft pion
- **Combinatorial** background
- **Multibody** D^*: Real D^* with other (multibody) D decay

The most important source of background comes from decays where the flavour of the D meson is not known, which provides a dilution of the signal distribution. These are decays where a real D meson is paired and combined with a random, unrelated pion from the event. This background peaks in the $m(D^0)$ distribution and has the same shape as the signal. This is the main reason why a 2D fit is needed in this channel. To resolve this background, the Δm distribution is used, where the background does not peak. An empirical function with kinematic threshold is used to model the background in Δm:

$$f(m;\,\alpha,\beta,m_0) \propto (m-m_0)^{1/2} + \alpha\,(m-m_0)^{3/2} + \beta\,(m-m_0)^{5/2}\,, \qquad (8.6)$$

where the threshold m_0 is fixed and given by the pion mass, $m_0 = 139.57\,\mathrm{MeV}/c^2$. The parameters α and β are left free in the fit to data. This threshold function was updated and is different from the Run 2 model, which used another empirical threshold function.

The combinatorial background is non-peaking in both variables. A first order polynomial is used to model the combinatorial background in $m(D^0)$, as in Run 2. The empirical function with kinematic threshold in Equation (8.6) is used for Δm as well, where the threshold m_0 is left free in this case.

The final source of background is the multibody D^* background. These are real D^* decays but where the D^0 meson decays to other final states. This background is expected to peak in Δm and the shape to be the same as the signal, given that a real D^* is selected and decaying to a D^0 and a slow pion. However, the distribution of this background needs to be studied in detail in the $m(D^0)$ variable. In Run 2, this background was simply modelled by a first order polynomial in $m(D^0)$. This background is low but the precision required for the calibration means that these have to be well understood and modelled in order to be able to control

them in the fit. Therefore, a detailed study of this background is performed in order to obtain a good description.

In order to study these backgrounds, fast simulation was used to generate several D^0 decays, listed in Table 8.2 together with their branching fractions.

Table 8.2: List of D^0 decay modes studied for the multibody D^* background with the corresponding branching fractions [142].

D^0 decay mode	Branching fraction [%]
$D^0 \to K^-\pi^+$	3.950 ± 0.031
$D^0 \to K^-K^+$	0.408 ± 0.006
$D^0 \to \pi^-\pi^+$	0.1455 ± 0.0024
$D^0 \to \pi^-K^+$	0.01364 ± 0.00026
$D^0 \to K^-\mu^+\nu_\mu$	3.41 ± 0.04
$D^0 \to K^-e^+\nu_e$	3.542 ± 0.035
$D^0 \to \pi^-\pi^+\pi^0$	1.49 ± 0.06
$D^0 \to K^-\pi^+\pi^0$	14.4 ± 0.5

These decays contribute as background when particles are misidentified or missed in the reconstruction, and therefore the same final state as signal is reached. Figure 8.14 shows the multibody background distributions obtained from fast simulation. For each decay, the misidentified particles are shown in red in the legend, while missed particles are written in squared brackets. For example, the $D^0 \to K^-\mu^+[\nu_\mu]$ decay contributes as background when the muon is misidentified as a pion, while the neutrino is missed in the reconstruction. From the corresponding plot in Figure 8.14, it can be seen that this background contributes to the signal region. However, the $D^0 \to K^-\pi^+\pi^0$ decay, which has the largest branching fraction, does not contribute to the signal region as it sits at a lower reconstructed mass due to the neutral pion which is not reconstructed. The mass range in which they appear is relevant in the final distribution considered for the fit. Moreover, the decays are weighted by their corresponding branching fraction to obtain the total multibody D^* background contribution.

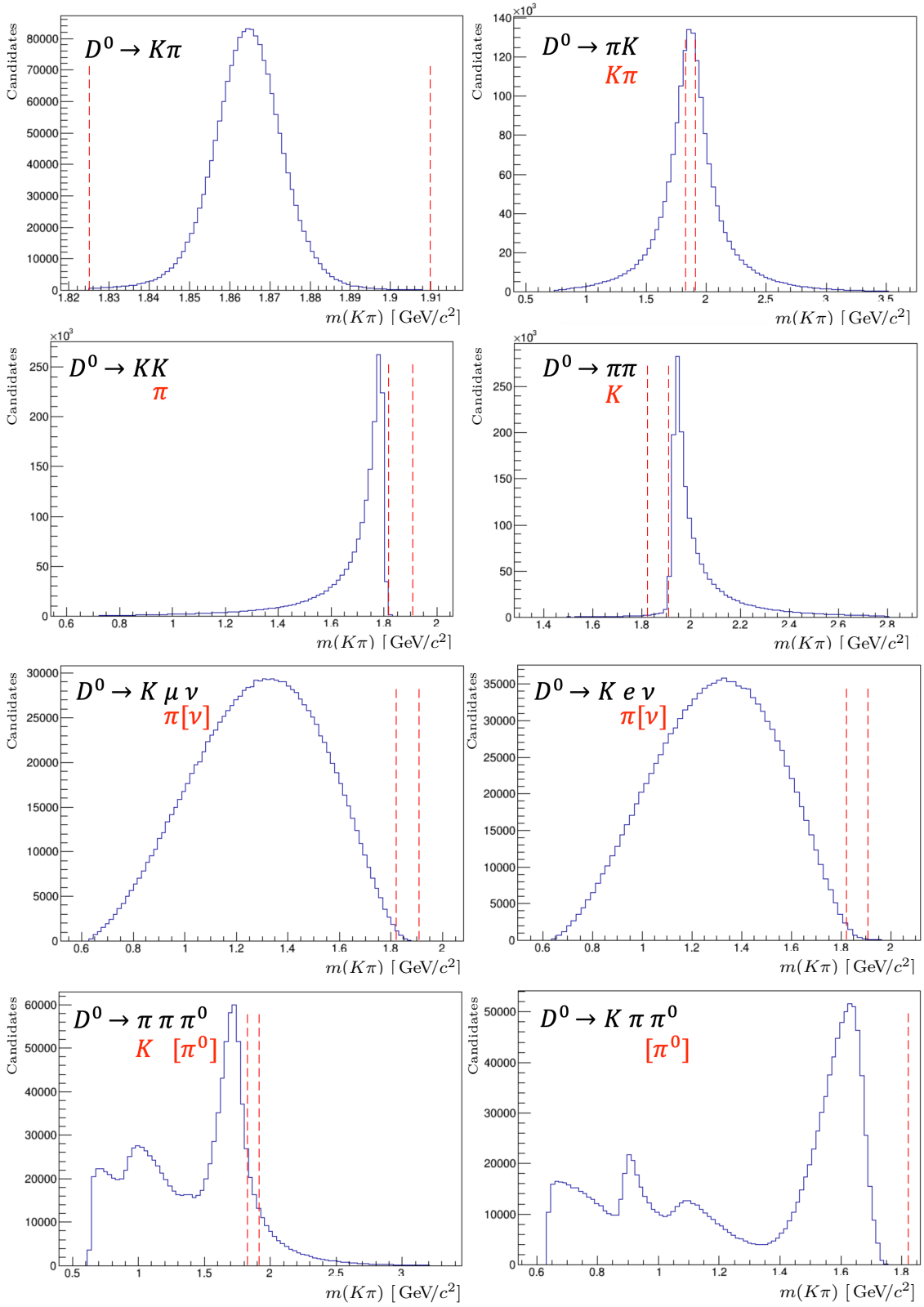


Figure 8.14: Invariant-mass distributions of the D^0 decay modes studied for the multibody D^* backgrounds. For each decay, the misidentified particles are shown in red in the legend, while missed particles are written in squared brackets. The fit range is shown by the red vertical lines.

Figure 8.15 presents the invariant-mass distributions for all D^0 decays considered, together with the signal, before (left) and after (right) applying the vetoes. This illustrates the fact that the background shapes are sculpted by these vetoes. From the plots, it can already be seen that the largest background comes from $D^0 \to \pi^-\pi^+\pi^0$ decays, which peaks around the D^0 mass. It can also be seen that these backgrounds are very small compared to the signal. Therefore, the idea is to fix the yield and the shape of this background in the final fit, which would help with the stability of the fit and prevent this component from going to zero. Moreover, this also prevents this component from compensating for signal in the tails.

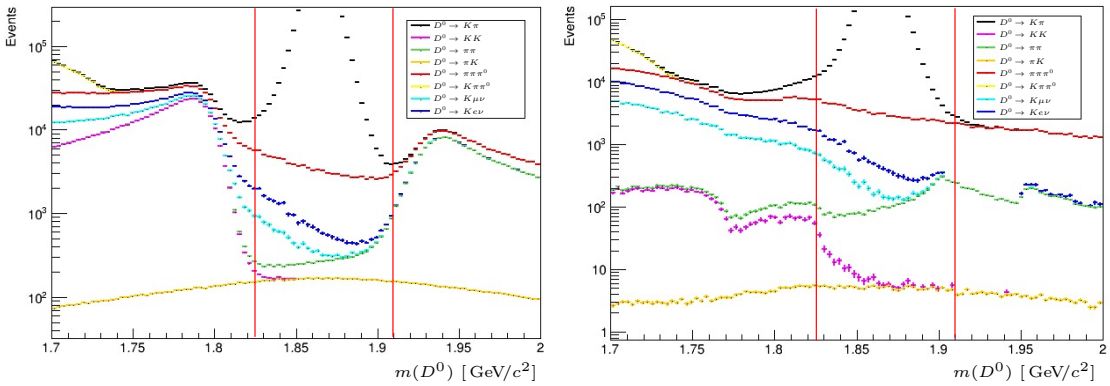


Figure 8.15: Superimposed invariant-mass distributions of the D^0 decay modes studied for the multibody D^* backgrounds, before (left) and after (right) applying the vetoes.

Due to the fact that the background shape varies across the phase space, a fit to the total multibody D^* background distribution is performed in each of the kinematic bins. The background is modelled using a second order polynomial. These fits are shown in Figure 8.16 for the probe pion and in Figure 8.17 for the probe kaon. It is found that a peak at around 1.9 GeV/c^2 is present in some kinematic bins. Therefore, an additional Gaussian is added in the model to account for this, while only a second order polynomial is used in the other bins. The shape parameters are completely fixed from these fits in the final invariant-mass fit to extract the *sWeights*.

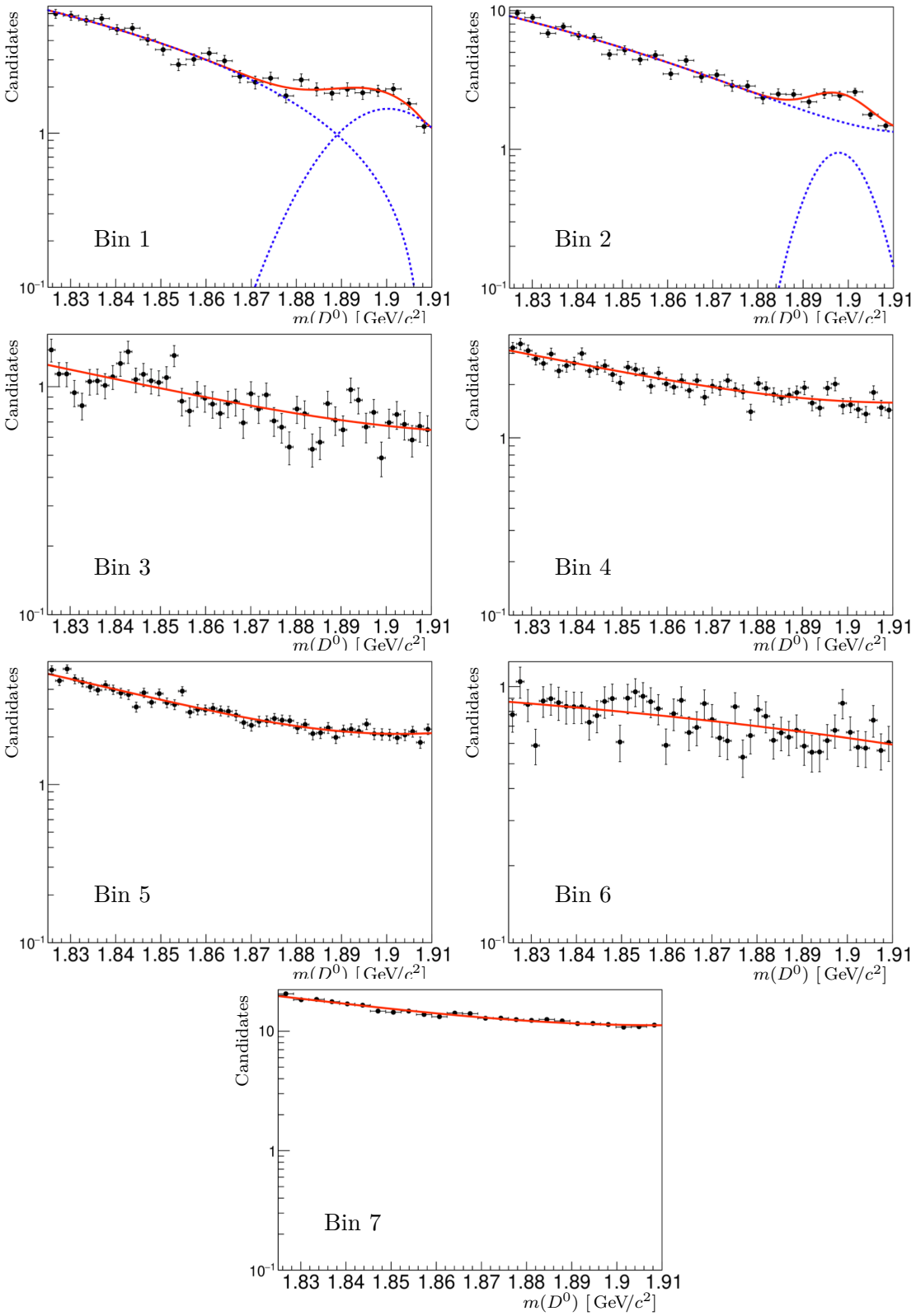


Figure 8.16: Invariant-mass distribution of the total multibody D^* background in the different bins for the probe pion, with the fit result superimposed.

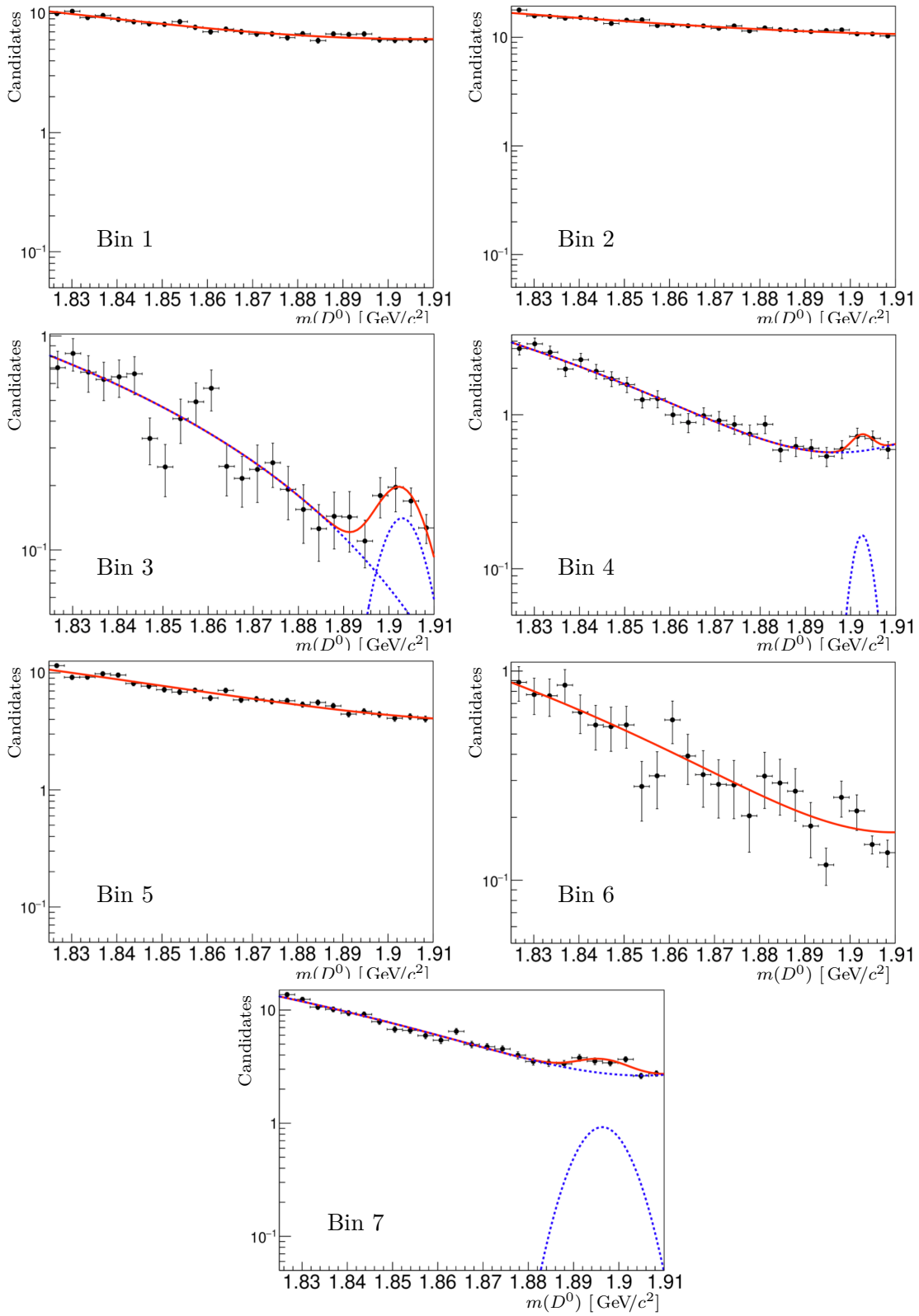


Figure 8.17: Invariant-mass distribution of the total multibody D^* background in the different bins for the probe kaon, with the fit result superimposed.

8.6 Two-dimensional model

As a summary, the four components of the fit model are listed in Table 8.3 in a schematic way. For each component, it is shown whether it is peaking or non-peaking in the two variables. Note that for the multibody D^* background, an additional Gaussian function is used in some of the bins to model the small peaking structure in the distribution.

Table 8.3: Summary of the components in the model. For each component, it is shown whether it is peaking or non-peaking in the $m(D^0)$ and Δm variables.

2D model		
Component	$m(D^0)$	Δm
Signal		
Random π		
Combinatorial		
Multibody D^*		

8.7 Previous calibration in Run 2

Due to the high yields of these D^* decays, the 2D fit is challenging to perform. In Run 2, the fit was performed in the full phase space. The fit result for data from 2016, taken from internal webpages, is shown in Figure 8.18. The fit presents relatively large pulls, with the pulls reaching values larger than 10σ and 25σ for the D^0 and Δm projections respectively.

8.8 New two-dimensional data fit

One of the key points is to make sure that the fit is stable and the parameters are correctly determined in each kinematic bin. The multi phase-space fit helps to improve the fit quality by performing the 2D fit in each bin, providing more accurate *sWeights*. With the new model developed in this work, the 2D fit is performed using

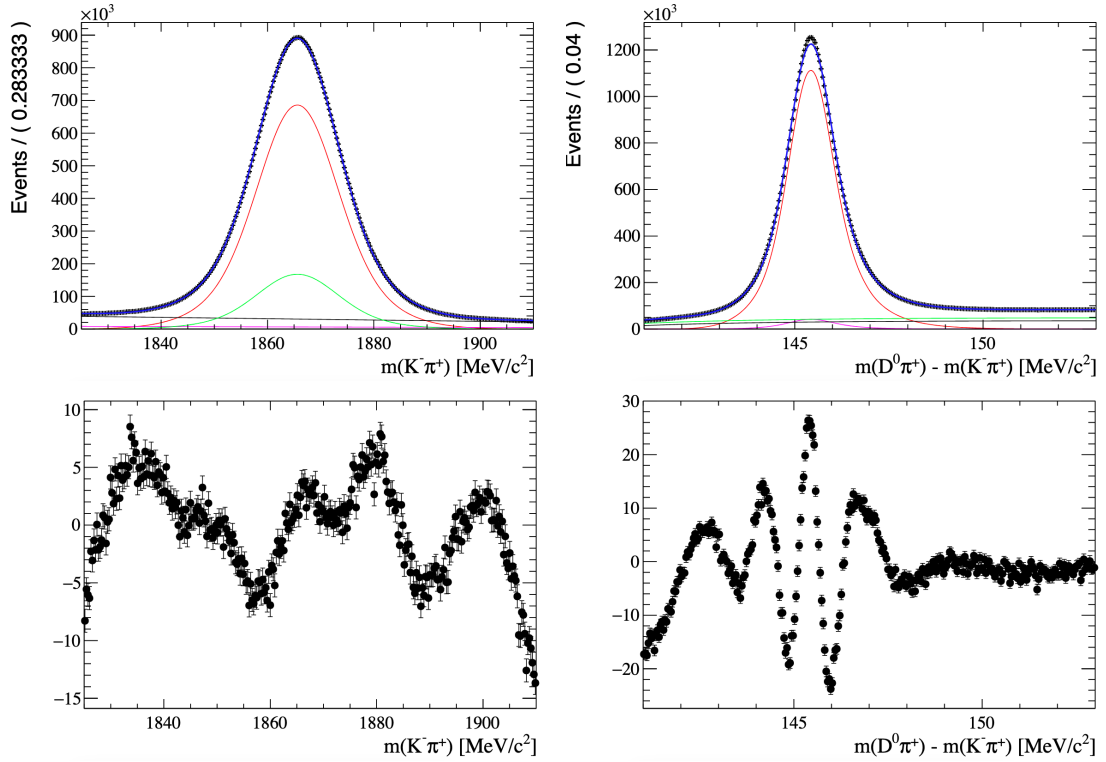


Figure 8.18: Two-dimensional fit to data from the Run 2 calibration procedure, with the residuals shown for each the $m(D^0)$ (left) and Δm (right) projections, taken from internal webpages.

the Run 2 calibration samples. Figure 8.19 presents the $m(D^0)$ projections from the 2D fit in the different kinematic bins, using calibration data from 2016. It can be seen that the pulls are better than the old fit, with pulls contained within around 3σ.

This new fit model was implemented in the new calibration framework, where the fit was also performed. Figure 8.20 presents the fit projections obtained using 2016 data, performing the fit in the full phase-space. This framework will be used in Run 3 to obtain the *sWeights*.

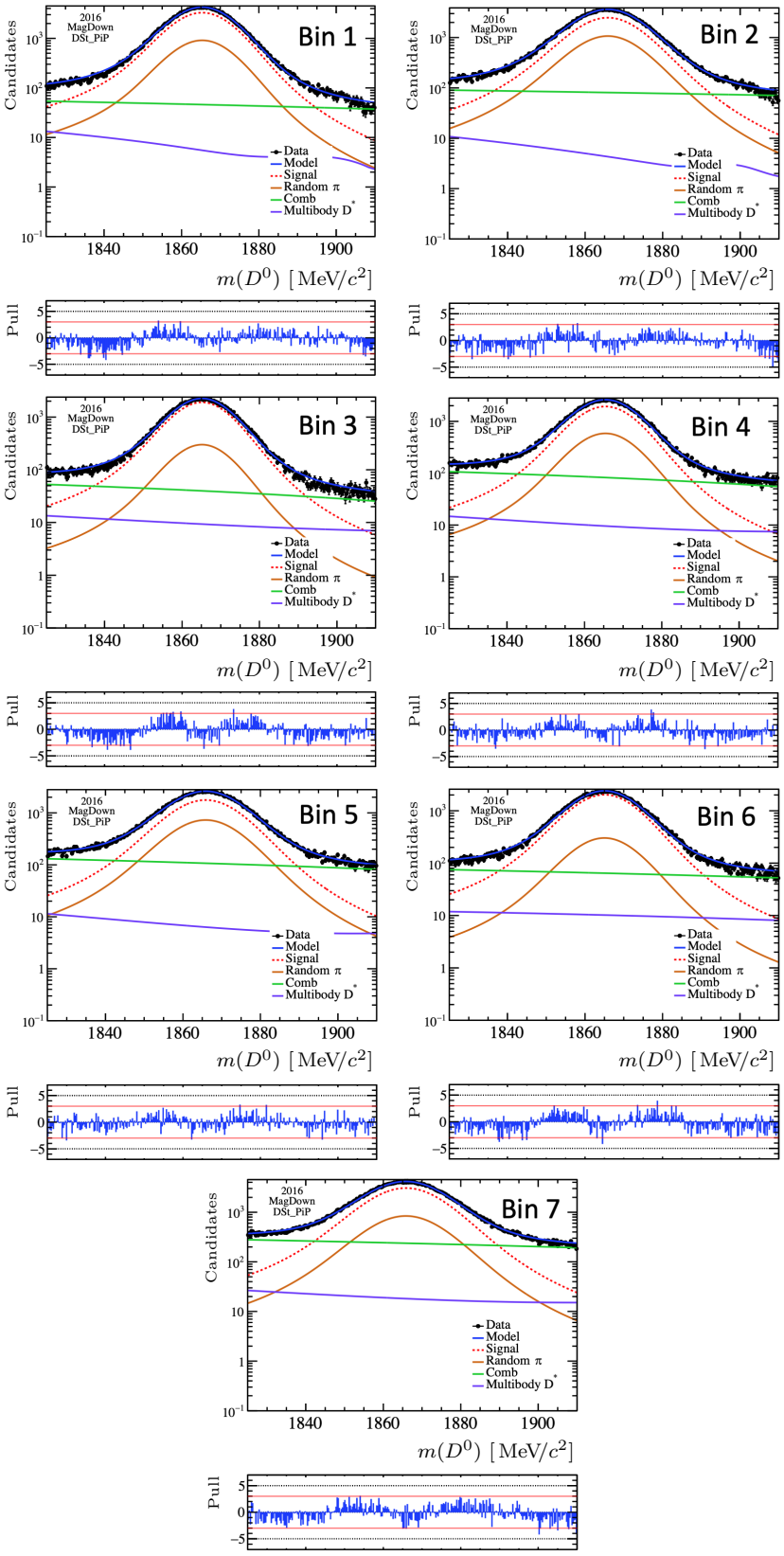


Figure 8.19: Projections of the 2D fit result for $m(D^0)$ in the different kinematic bins.

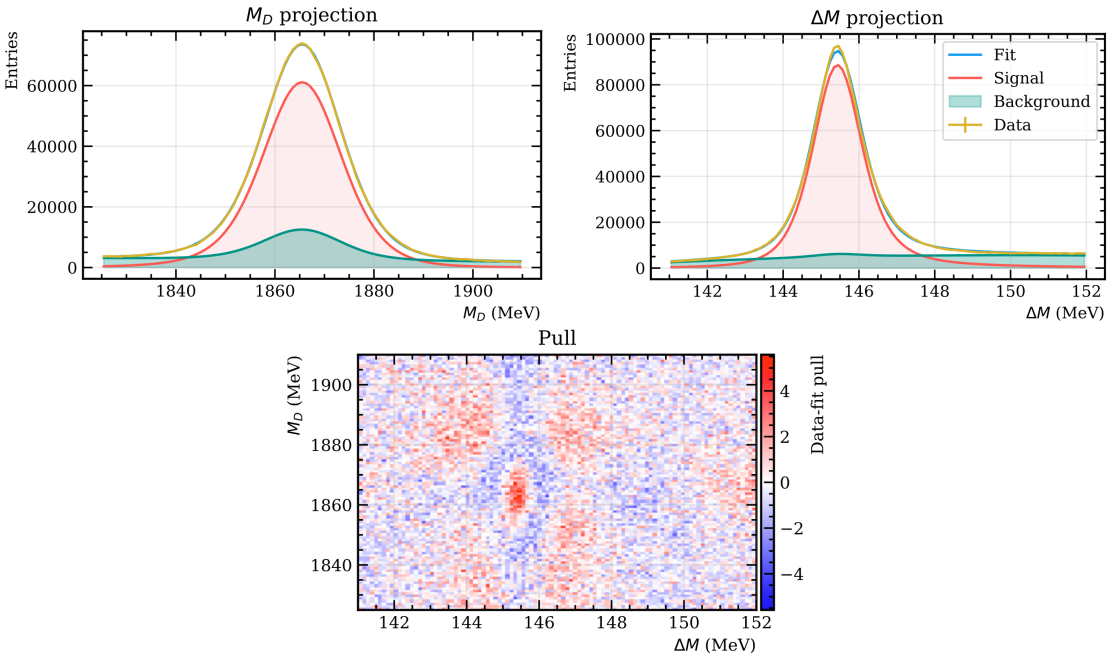


Figure 8.20: Projections of the 2D fit result in the new framework, using the integrated (full phase space) 2016 data. The two-dimensional residuals are also shown.

8.9 PID performance with 2023 data

This section presents the PID performance with data collected in July 2023. The dataset corresponds to run numbers between 269370 and 269542, with a down polarity. In this data taking period, the UT was not included, while the VELO was semi-closed (with a 49 mm gap). The sample was selected using the dedicated HLT2 line, where the cuts were adjusted to account for the higher rate and the level of background. Due to these cuts differences and the fact that the VELO was not closed, the high-end tail of the η coverage is reduced, as well as the lower end of the momentum spectrum. This can be seen in Figure 8.21, where the p (left) and η (right) distributions, with the *sWeights* applied, are compared between 2016 and 2023 data. Similarly, the p versus p_T distributions in Figure 8.22 show the missing lower end of the momentum spectrum. This can be recovered if necessary in Run 3 nominal conditions by adjusting the HLT2 line. In 2023, data was taken for commissioning purposes and the value of the mean number of reconstructed PVs was about $\mu = 1.8$. This is similar to that of Run 2, but well below the design intention of Run 3. A comparison is made between the performance in the 2016 and 2023 datasets.

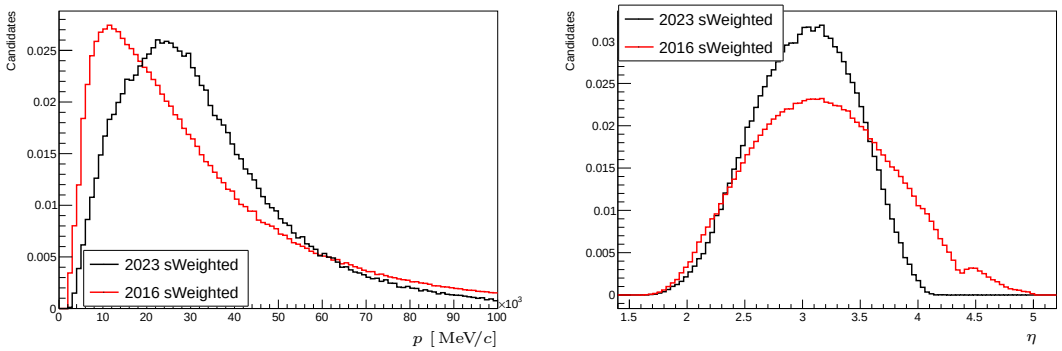


Figure 8.21: Distributions of p (left) and η (right) showing the coverage in 2016 and 2023.

Given the limited statistics in the 2023 data analysed here and the higher level of background, the 2D fit is performed to the integrated phase space, shown in Figure 8.23. Comparing the fits between 2016 and 2023, it can be seen that the level of background is higher in $m(D^0)$ in 2023. An improvement might be required

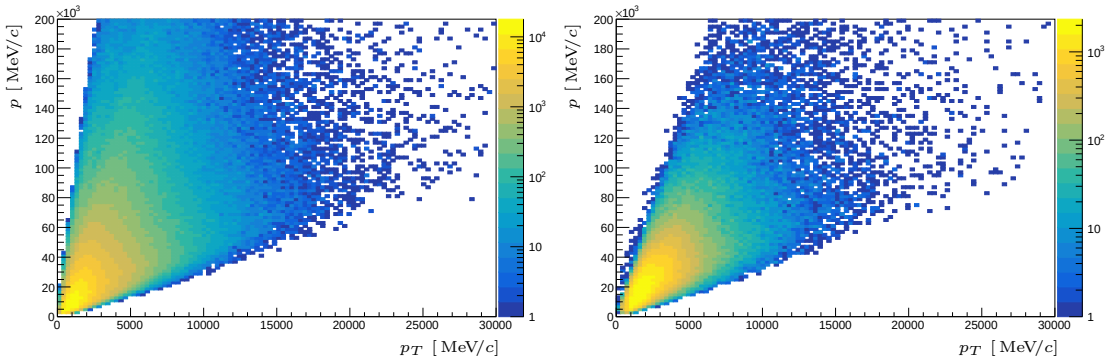


Figure 8.22: Coverage in the p-p_T phase space for 2016 (left) and 2023 (right) data.

for better calibration in the nominal Run 3 conditions. The resolution in the D^0 mass is slightly worse in 2023 compared to 2016, which is likely to be resolved by a better alignment of the detector with the start of Run 3. With the increased yields in Run 3, the default fit will be performed in the kinematic bins, improving the accuracy of the fit and reducing any bias from the procedure.

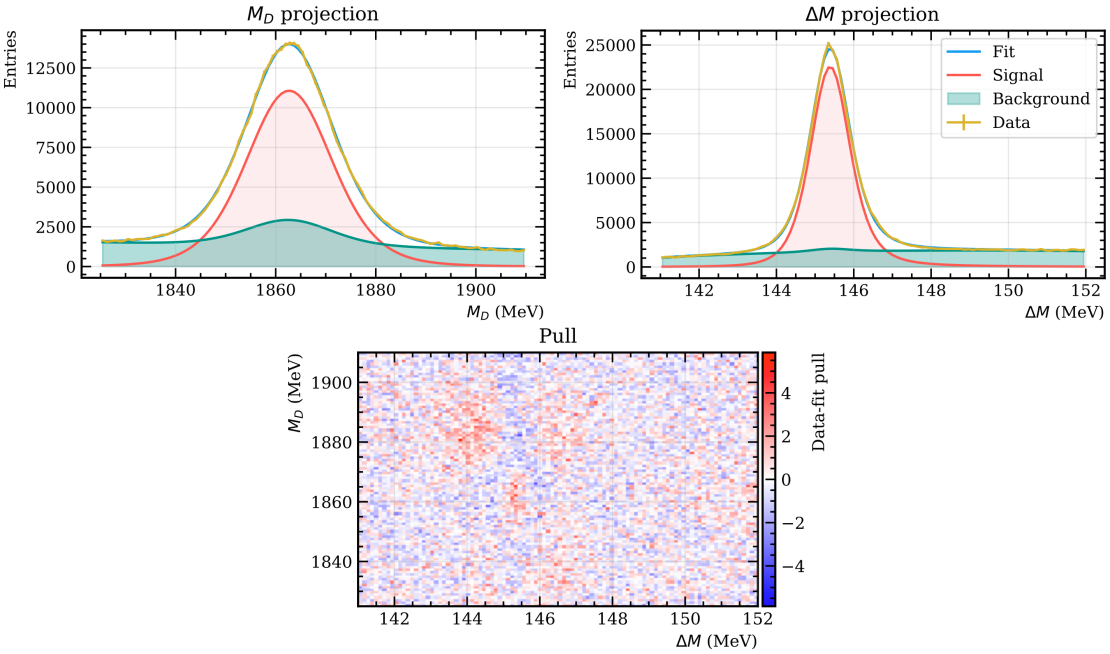


Figure 8.23: Projections of the 2D fit result in the new framework, using the integrated (full phase space) 2023 data. The two-dimensional residuals are also shown.

Using the *sWeights* obtained from this 2D fit to the full phase space, where $p \in [2,\ 200]$ GeV/c and $\eta \in [1.6,\ 5.0]$, the *sWeights* are extracted in order to evaluate the PID performance. This can be examined qualitatively by comparing

the distributions of the discriminating PID variables of different hadrons after subtracting the background. To discriminate between two particles, the difference in the log likelihood, ΔLL, between the hypotheses that the hits are produced by these two particular particles in the RICH detector is used. For example, $\Delta\mathrm{LL}(K-\pi)$ represents the difference in the log likelihood between the hypotheses of a kaon and pion. This distribution is shown in Figure 8.24, where excellent separation and little overlap between the two distributions can be observed.

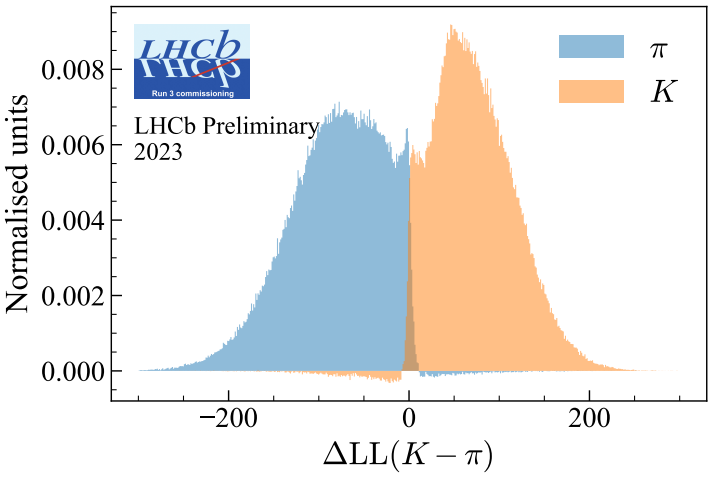


Figure 8.24: Distributions of the difference in the log likelihood between the hypotheses of a kaon and pion, where the distributions are background subtracted using the *sWeights* obtained from the 2D fit to the data.

The performance is also examined by computing the efficiencies from the *sWeights*, for two different PID cuts. This performance plot is shown in Figure 8.25, which shows the efficiency of the two considered PID requirements. The different curves represent the efficiency to correctly identify a kaon, as well as the corresponding misidentification of a pion as a kaon. Due to the limited yields, statistical variations in the *sWeights* can be seen, which cause the log likelihood distributions to be negative in some regions, as well as the efficiencies to be unphysical in a few kinematic bins. The efficiencies are restricted to be within [0, 1] for the purpose of these plots. This has been seen in the past, in the commissioning phase of the LHCb detector in Run 1 for example, and is expected to be small with a larger calibration sample. Moreover, the systematic uncertainty due to the *sWeights* procedure also

reduces with a lower level of background, which is expected to be achieved in Run 3. The systematic uncertainty will be evaluated in more detail in the future, with the default fit in kinematic bins and in the final setup when the selection is fully defined.

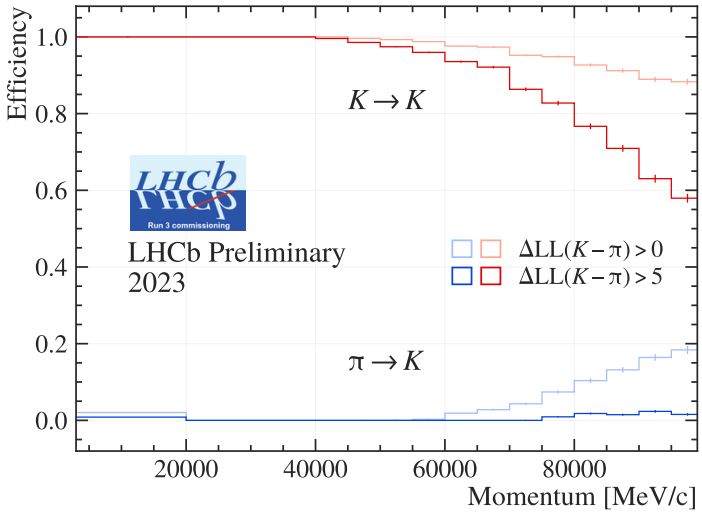


Figure 8.25: Efficiency of PID selections in momentum bins. Two selections on $\Delta\mathrm{LL}(K-\pi)$ are presented. The red curves represent the efficiency of correctly identifying a kaon, while the blue curves display the misidentification efficiency of a pion as a kaon.

Given the similar occupancy of the detector as in Run 2, a significant improvement of the PID performance is observed in this work with Run 3 data, which is expected due to improvements in the RICH detector, especially the gains obtained in the single-photon resolution [143] and yield. Similar plots to Figure 8.25 can be found in Reference [82] for Run 2. Even though the increase in occupancy expected in Run 3 will degrade the PID performance, this is still expected to be as good as in Run 2 or better. Figure 8.26 shows the pion misidentification versus the kaon identification efficiency for 2016 and 2023 data, for two different momentum ranges. This comparison shows again the improvement in the PID performance in 2023.

This improvement in the plots is largely due to the very low occupancy of the 2023 run. The design goal for the RICH was that it would have the same performance in Run 2 and Run 3. In studies done by LHCb colleagues on 2024 data, it is confirmed that the RICH performance in Run 3 is comparable to that in Run 2 when running at nominal design luminosity.

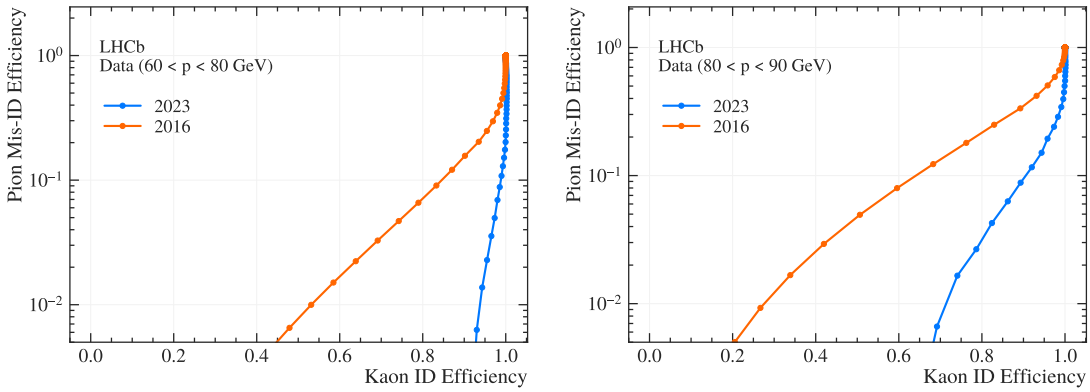


Figure 8.26: Pion misidentification efficiency versus kaon identification efficiency in 2016 and 2023, for two different momentum ranges.

Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning.

— Albert Einstein

9 Summary and outlooks

This thesis presents the measurement of the CKM angle γ in $B^{\pm} \to DK^{*\pm}$ decays at the LHCb experiment, using the full dataset collected in proton-proton collisions during Run 1 (2011–2012) and Run 2 (2015–2018), corresponding to an integrated luminosity of $9\,\text{fb}^{-1}$. This consists of a comprehensive study of this decay channel, where the D meson decays to two- and four-body final states $K^+\pi^-$, π^+K^-, K^+K^-, $\pi^+\pi^-$, $K^+\pi^-\pi^+\pi^-$, $\pi^+K^-\pi^+\pi^-$, $\pi^+\pi^-\pi^+\pi^-$ and three-body final states $K^0_S\pi^+\pi^-$, $K^0_S K^+K^-$.

This work supersedes a previous measurement in the two- and four-body modes using partial Run 2 data (2011–2016) published by LHCb [6], benefitting from the increased dataset by about a factor two and other improvements introduced in the analysis. Some of these improvements are an improved selection of the candidates and a detailed understanding of the backgrounds involved in the analysis. Moreover, this work includes the three-body modes which are added in the measurement. This analysis of the $B^{\pm} \to DK^{*\pm}$, $D \to K^0_S h^+h^-$ decays is performed for the first time at LHCb, while previous measurements from other B-factories have been published [113, 114].

The *CP* observables measured for the two- and four-body final states are found to

be

$$
\begin{aligned}
A_{K\pi} &= -0.024 \pm 0.014 \pm 0.002 \\
A_{KK} &= 0.14 \pm 0.04 \pm 0.001 \\
A_{\pi\pi} &= 0.31 \pm 0.09 \pm 0.05 \\
A_{\pi K} &= -0.73 \pm 0.16 \pm 0.03 \\
R_{KK} &= 1.10 \pm 0.05 \pm 0.01 \\
R_{\pi\pi} &= 0.96 \pm 0.09 \pm 0.05 \\
R_{\pi K} &= 0.0098 \pm 0.0019 \pm 0.0002 \\
A_{K\pi\pi\pi} &= -0.024 \pm 0.018 \pm 0.002 \\
A_{\pi\pi\pi\pi} &= 0.04 \pm 0.06 \pm 0.03 \\
A_{\pi K\pi\pi} &= -0.19 \pm 0.22 \pm 0.01 \\
R_{\pi\pi\pi\pi} &= 1.05 \pm 0.07 \pm 0.05 \\
R_{\pi K\pi\pi} &= 0.0118 \pm 0.0026 \pm 0.0001,
\end{aligned}
$$

where the first uncertainty is statistical and the second is systematic. The results for the three-body modes are

$$
\begin{aligned}
x_+ &= 0.003^{+0.051}_{-0.052} \pm 0.018 \pm 0.004 \\
y_+ &= 0.056^{+0.059}_{-0.064} \pm 0.009 \pm 0.019 \\
x_- &= 0.139^{+0.051}_{-0.060} \pm 0.019 \pm 0.009 \\
y_- &= -0.175^{+0.114}_{-0.103} \pm 0.013 \pm 0.039,
\end{aligned}
$$

where the first uncertainty is statistical and the second is systematic. The third uncertainty represents the propagated uncertainty from the c_i, s_i external strong-phase inputs.

The physics parameters of interest are determined through the equations which directly relate them to the *CP* observables. The *CP* observables results for the two- and four-body modes are consistent and supersede the results obtained in the previous analysis published with partial Run 2 data [6]. Given the nature of the

trigonometric equations for these two- and four-body modes, multiple solutions for γ are obtained. The three-body modes give a broader but single solution and help pick one of the two solutions. Therefore, combining the results obtained from the different modes is powerful and makes it possible to achieve a better overall precision on γ in this decay channel. The *CP* observables are combined, yielding

$$\gamma = (63 \pm 13)^\circ$$

$$r_B^{DK^{*\pm}} = 0.103 \pm 0.010$$

$$\delta_B^{DK^{*\pm}} = (47^{+14}_{-12})^\circ.$$

The result obtained in this analysis is model independent and compatible with the world average. This is included in the latest LHCb combination [2], providing a valuable addition to the family of γ measurements and the knowledge of γ. This work also constitutes the first observation of the suppressed $B^\pm \to [\pi^\pm K^\mp]_D K^{*\pm}$ and $B^\pm \to [\pi^\pm K^\mp \pi^\pm \pi^\mp]_D K^{*\pm}$ decays. With the Run 1 and Run 2 dataset used for the measurement, the analysis is statistically dominated compared to the different sources of systematic uncertainties involved in the measurement.

The precision obtained on the *CP*-violating observables in this decay channel, which contains an additional K^0_S particle coming from the K^* decay, is lower than the one achieved in the golden channel $B^\pm \to DK^\pm$. However, the decay channel studied in this measurement presents many advantages. These include a higher purity and a signal peak which is well separated from partially reconstructed backgrounds appearing at a lower reconstructed mass. Another key difference is that there is no corresponding misidentified background from $B^\pm \to D\pi^\pm$. Therefore, different sources of systematic uncertainties contribute in the analysis, which will become more important in the future with larger datasets. This also provides an important cross-check on the value of γ obtained from different decays.

The LHCb combination, which drives the world average, combines the results obtained for γ from different tree-level decays. The latest LHCb combination [2], which includes the measurement presented in this thesis, achieves a precision of 2.8°. This is better than the expected sensitivity of 4° from Run 1–2 data at LHCb.

This was obtained by studying a wide range of decay modes and improving and optimising the analyses performed in the different channels. Part of this is due to the measurement presented in this thesis.

With Run 3 and beyond, and the larger dataset to be collected, a precision better than 1° is expected to be reached [144, 145], which will allow a better comparison with the indirect determination of γ. It is shown in this thesis that the decay channel analysed lends itself well to analysis and presents advantages such as a clean signal peak and large purity which provide good sensitivity in this channel, which could be one of the first target goals for γ measurements in Run 3. With more precise tree-level determinations in the future, improving the knowledge of γ, which is a standard candle measurement of *CP* violation in the Standard Model, will continue to play an important role in the LHCb physics programme and particle physics in general.

Appendices

A

CP fit projections

The *CP* fit projections in logarithmic scale are presented in Figures A.1–A.3 for each category, where the small background contributions can be seen.

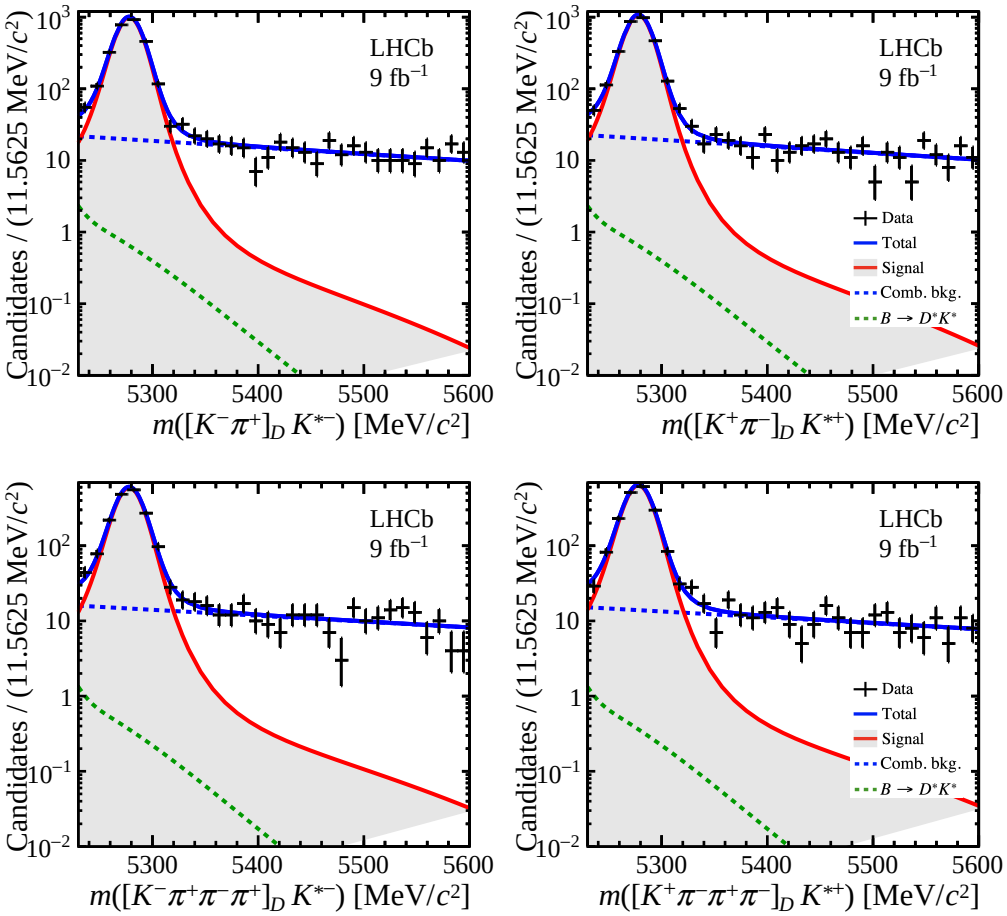


Figure A.1: Invariant-mass distributions for the favoured $K\pi$ (top row) and $K\pi\pi\pi$ (bottom row) modes for B^- (left) and B^+ (right) decays, obtained with data corresponding to the full Run 1 and Run 2 dataset. The fit result is superimposed, showing the different components.

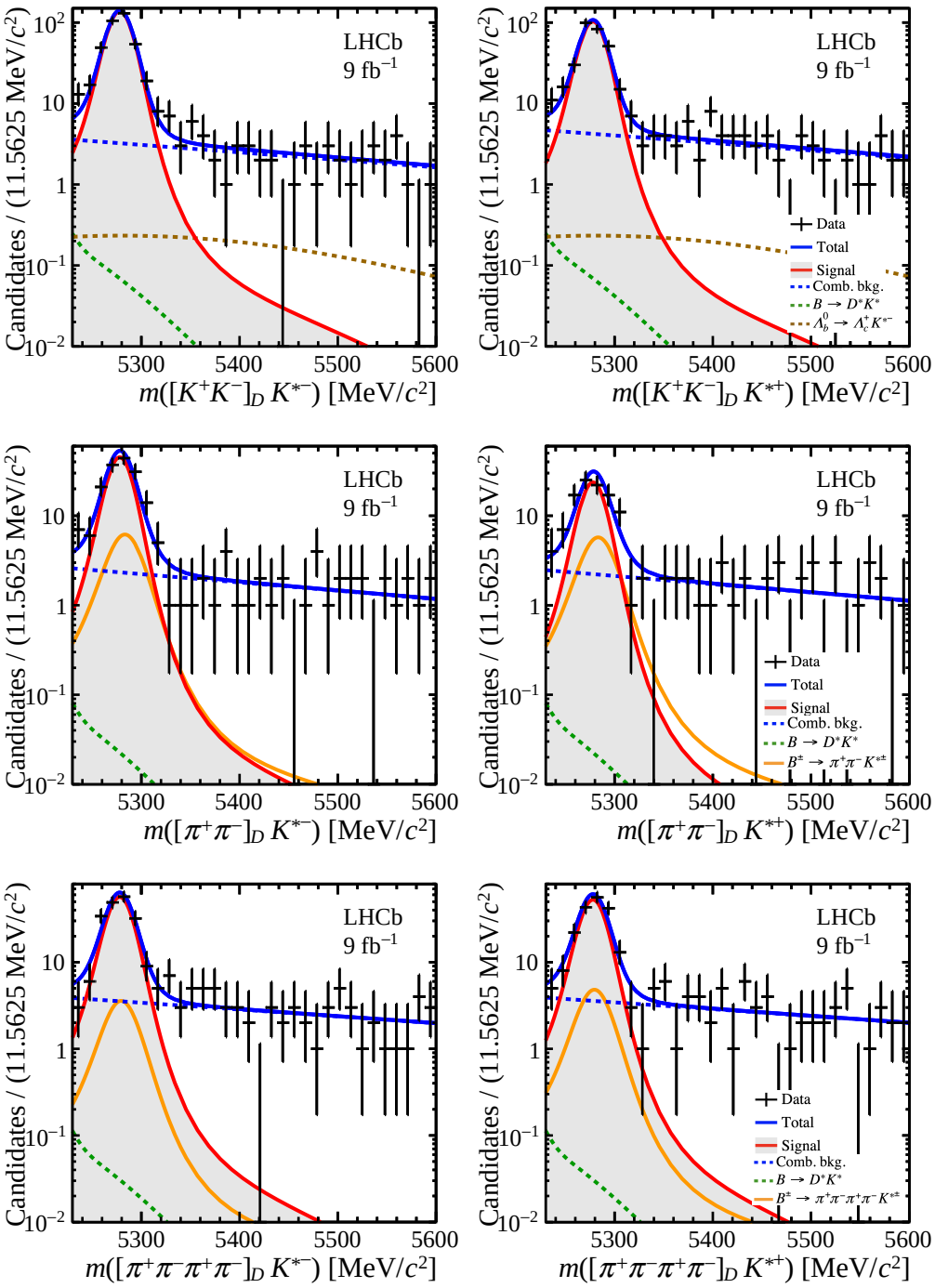


Figure A.2: Invariant-mass distributions for the KK (top row), $\pi\pi$ (middle row) and $\pi\pi\pi\pi$ (bottom row) modes for B^- (left) and B^+ (right) decays, obtained with data corresponding to the full Run 1 and Run 2 dataset. The fit result is superimposed, showing the different components.

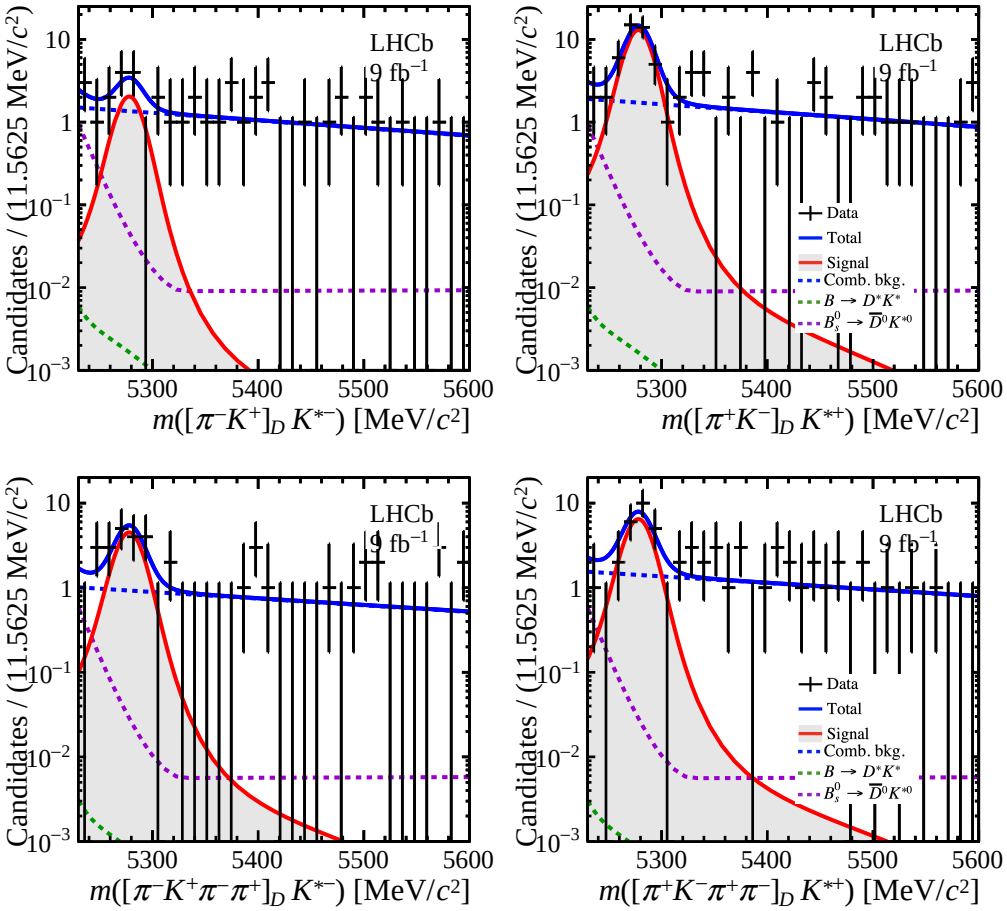


Figure A.3: Invariant-mass distributions for the suppressed πK (top row) and $\pi K\pi\pi$ (bottom row) modes for B^- (left) and B^+ (right) decays, obtained with data corresponding to the full Run 1 and Run 2 dataset. The fit result is superimposed, showing the different components.

References

[1] LHCb collaboration, *Measurement of the CKM angle γ in $B^{\pm} \to DK^{*}(892)^{\pm}$ decays*, JHEP **02** (2025) 113, arXiv:2410.21115, DOI: 10.1007/JHEP02(2025)113.

[2] LHCb collaboration, *Simultaneous determination of the CKM angle γ and parameters related to mixing and CP violation in the charm sector*, LHCb-CONF-2024-004, 2024, https://cds.cern.ch/record/2905625.

[3] LHCb collaboration, *Hadron PID performance with 2023 data*, LHCb-FIGURE-2023-023, 2023, https://lbfence.cern.ch/alcm/public/figure/details/720, https://cds.cern.ch/record/2871842.

[4] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531, DOI: 10.1103/PhysRevLett.10.531.

[5] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652, DOI: 10.1143/PTP.49.652.

[6] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP observables in $B^{\pm} \to DK^{*\pm}$ decays using two- and four-body D-meson final states*, JHEP **11** (2017) 156, Erratum ibid. **05** (2018) 067, arXiv:1709.05855, DOI: 10.1007/JHEP11(2017)156.

[7] Wikipedia, *The Standard Model of Elementary Particles*, https://en.wikipedia.org/wiki/Standard_Model.

[8] LHCb collaboration, R. Aaij *et al.*, *First observation of a doubly charged tetraquark candidate and its neutral partner*, Phys. Rev. Lett. **131** (2023) 041902, arXiv:2212.02716, DOI: 10.1103/PhysRevLett.131.041902.

[9] LHCb collaboration, R. Aaij *et al.*, *Observation of a $J/\psi\Lambda$ resonance consistent with a strange pentaquark candidate in $B^{-} \to J/\psi\Lambda\overline{p}$ decays*,

Phys. Rev. Lett. **131** (2023) 031901, arXiv:2210.10346, DOI: 10.1103/PhysRevLett.131.031901.

[10] A. Bay, *PARTICULES ELEMENTAIRES*, Laboratoire de Physique des Hautes Energies, EPFL.

[11] E. Corbelli and P. Salucci, *The extended rotation curve and the dark matter halo of M33*, Monthly Notices of the Royal Astronomical Society **311** (2000) 441, DOI: 10.1046/j.1365-8711.2000.03075.x.

[12] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991) 392, DOI: 10.1070/PU1991v034n05ABEH002497.

[13] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235.

[14] G. Luders, *On the Equivalence of Invariance under Time Reversal and under Particle-Antiparticle Conjugation for Relativistic Field Theories*, Kong. Dan. Vid. Sel. Mat. Fys. Med. **28N5** (1954) 1.

[15] C. S. Wu *et al.*, *Experimental Test of Parity Conservation in β Decay*, Phys. Rev. **105** (1957) 1413, DOI: 10.1103/PhysRev.105.1413.

[16] T. D. Lee and C. N. Yang, *Question of Parity Conservation in Weak Interactions*, Phys. Rev. **104** (1956) 254, DOI: 10.1103/PhysRev.104.254.

[17] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13** (1964) 138, DOI: 10.1103/PhysRevLett.13.138.

[18] NA48 collaboration, V. Fanti *et al.*, *A new measurement of direct CP violation in two pion decays of the neutral kaon*, Phys. Lett. B **465** (1999) 335, DOI: 10.1016/s0370-2693(99)01030-8.

[19] KTeV collaboration, A. Alavi-Harati *et al.*, *Measurements of direct CP violation, CPT symmetry, and other parameters in the neutral kaon system*, Phys. Rev. **D67** (2003) 012005, DOI: 10.1103/PhysRevD.67.012005.

[20] BaBar collaboration, B. Aubert *et al.*, *Observation of CP Violation in the B^0 Meson System*, Phys. Rev. Lett. **87** (2001) 091801, DOI: 10.1103/PhysRevLett.87.091801.

[21] Belle collaboration, K. Abe *et al.*, *Observation of Large CP Violation in the Neutral B Meson System*, Phys. Rev. Lett. **87** (2001) 091802, DOI: 10.1103/PhysRevLett.87.091802.

[22] LHCb collaboration, R. Aaij *et al.*, *Observation of CP violation in $B^{\pm} \to DK^{\pm}$ decays*, Phys. Lett. **B712** (2012) 203, Erratum ibid. **B713** (2012) 351, arXiv:1203.3662, DOI: 10.1016/j.physletb.2012.04.060.

[23] LHCb collaboration, R. Aaij *et al.*, *First observation of CP violation in the decays of B_s^0 mesons*, Phys. Rev. Lett. **110** (2013) 221601, arXiv:1304.6173, DOI: 10.1103/PhysRevLett.110.221601.

[24] LHCb collaboration, R. Aaij *et al.*, *Observation of CP violation in charm decays*, Phys. Rev. Lett. **122** (2019) 211803, arXiv:1903.08726, DOI: 10.1103/PhysRevLett.122.211803.

[25] T2K collaboration, K. Abe *et al.*, *Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations*, Nature **580** (2020) 339, DOI: 10.1038/s41586-020-2177-0.

[26] LHCb collaboration, R. Aaij *et al.*, *Observation of charge-parity symmetry breaking in baryon decays*, (2025), arXiv:2503.16954, submitted to Nature.

[27] L.-L. Chau and W.-Y. Keung, *Comments on the Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **53** (1984) 1802, DOI: 10.1103/PhysRevLett.53.1802.

[28] Particle Data Group, S. Navas *et al.*, *Review of particle physics*, Phys. Rev. D **110** (2024) 030001, DOI: 10.1103/PhysRevD.110.030001.

[29] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa matrix*, Phys. Rev. Lett. **51** (1983) 1945, DOI: 10.1103/PhysRevLett.51.1945.

[30] J. Brod, A. Lenz, G. Tetlalmatzi-Xolocotzi, and M. Wiebusch, *New physics effects in tree-level decays and the precision in the determination of the quark mixing angle γ*, Phys. Rev. D **92** (2015) 033002, DOI: 10.1103/PhysRevD.92.033002.

[31] A. Lenz and G. Tetlalmatzi-Xolocotzi, *Model-independent bounds on new physics effects in non-leptonic tree-level decays of B-mesons*, JHEP **07** (2020) 177, DOI: 10.1007/JHEP07(2020)177.

[32] J. Brod and J. Zupan, *The ultimate theoretical error on γ from $B \to DK$ decays*, JHEP **01** (2014) 51, arXiv:1308.5663, DOI: 10.1007/JHEP01(2014)051.

[33] M. Blanke and A. J. Buras, *Emerging ΔM_d-anomaly from tree-level determinations of $|V_{cb}|$ and the angle γ*, Eur. Phys. J. C **79** (2019) 159, DOI: 10.1140/epjc/s10052-019-6667-x.

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

[35] LHCb collaboration, *Simultaneous determination of the CKM angle γ and parameters related to mixing and CP violation in the charm sector*, LHCb-CONF-2022-003, 2022, https://cds.cern.ch/record/2838029.

[36] A. Nandi, *CP violation measurements using $B^{\pm} \to DK^{*\pm}$ decays at LHCb*, PhD thesis, University of Oxford, 2018.

[37] M. Gronau and D. London, *How to determine all the angles of the unitarity triangle from $B^0_d \to DK_s$ and $B^0_s \to D\phi$*, Phys. Lett. B **253** (1991) 483, DOI: 10.1016/0370-2693(91)91756-L.

[38] M. Gronau and D. Wyler, *On determining a weak phase from charged B decay asymmetries*, Phys. Lett. B **265** (1991) 172, DOI: 10.1016/0370-2693(91)90034-N.

[39] M. Nayak *et al.*, *First determination of the CP content $D \to \pi^+\pi^-\pi^0$ and $D \to K^+K^-\pi^0$*, Phys. Lett. B **740** (2015) 1, arXiv:1410.3964, DOI: 10.1016/j.physletb.2014.11.022.

[40] BESIII collaboration, M. Ablikim *et al.*, *Measurement of the CP-even fraction of $D^0 \to \pi^+\pi^-\pi^+\pi^-$*, Phys. Rev. D **106** (2022) 092004, arXiv:2208.10098, DOI: 10.1103/PhysRevD.106.092004.

[41] D. Atwood, I. Dunietz, and A. Soni, *Enhanced CP Violation with $B \to KD^0(\bar{D}^0)$ Modes and Extraction of the Cabibbo-Kobayashi-Maskawa Angle γ*, Phys. Rev. Lett. **78** (1997) 3257–3260, DOI: 10.1103/PhysRevLett.78.3257.

[42] D. Atwood, I. Dunietz, and A. Soni, *Improved methods for observing CP violation in $B^{\pm} \to KD$ and measuring the CKM phase γ*, Phys. Rev. D **63** (2001) 036005, DOI: 10.1103/PhysRevD.63.036005.

[43] BESIII collaboration, M. Ablikim *et al.*, *Measurement of the $D \to K^-\pi^+\pi^-\pi^+$ and $D \to K^-\pi^+\pi^0$ coherence factors and average strong-phase differences in quantum-correlated $D\overline{D}$ decays*, JHEP **05** (2021) 164, arXiv:2103.05988, DOI: 10.1007/JHEP05(2021)164.

[44] A. Bondar, *Proceedings of BINP special analysis meeting on Dalitz analysis*, 24-26 Sep. 2002, unpublished.

[45] A. Giri, Y. Grossman, A. Soffer, and J. Zupan, *Determining γ using $B^{\pm} \to DK^{\pm}$ with multibody D decays*, Phys. Rev. D **68** (2003) 054018, DOI: 10.1103/PhysRevD.68.054018.

[46] Belle collaboration, A. Poluektov *et al.*, *Measurement of ϕ_3 with Dalitz Plot Analysis of $B^{\pm} \to D^{(*)}K^{\pm}$ Decay*, Phys. Rev. D **70** (2004) 072003, DOI: 10.1103/PhysRevD.70.072003.

[47] R. H. Dalitz, *On the analysis of τ-meson data and the nature of the τ-meson*, Phil. Mag. Ser.7 **44** (1953) 1068, DOI: 10.1080/14786441008520365.

[48] BaBar and Belle collaborations, I. Adachi *et al.*, *Measurement of* $\cos 2\beta$ *in $B^0 \to D^{(*)}h^0$ with $D \to K^0_S\pi^+\pi^-$ decays by a combined time-dependent Dalitz plot analysis of BaBar and Belle data*, Phys. Rev. D **98** (2018) 112012, arXiv:1804.06153, DOI: 10.1103/PhysRevD.98.112012.

[49] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle γ in $B^{\pm} \to DK^{\pm}$ and $B^{\pm} \to D\pi^{\pm}$ decays with $D \to K^0_S h^+h^-$*, JHEP **02** (2021) 169, arXiv:2010.08483, DOI: 10.1007/JHEP02(2021)169.

[50] CLEO collaboration, J. Libby *et al.*, *Model-independent determination of the strong-phase difference between D^0 and $\overline{D}^0 \to K^0_{S,L}h^+h^-$ ($h = \pi$, K) and its impact on the measurement of the CKM angle γ/ϕ_3*, Phys. Rev. D **82** (2010) 112006, arXiv:1010.2817, DOI: 10.1103/PhysRevD.82.112006.

[51] BESIII collaboration, M. Ablikim *et al.*, *Model-independent determination of the relative strong-phase difference between D^0 and $\overline{D}^0 \to K^0_{S,L}\pi^+\pi^-$ and its impact on the measurement of the CKM angle γ/ϕ_3*, Phys. Rev. D **101** (2020) 112002, arXiv:2003.00091, DOI: 10.1103/PhysRevD.101.112002.

[52] BESIII collaboration, M. Ablikim *et al.*, *Improved model-independent determination of the strong-phase difference between D^0 and $\bar{D}^0 \to K^0_{S,L}K^+K^-$ decays*, Phys. Rev. D **102** (2020) 052008, arXiv:2007.07959, DOI: 10.1103/PhysRevD.102.052008.

[53] S. Chung *et al.*, *Partial wave analysis in K-matrix formalism*, Annalen der Physik **507** (1995) 404, DOI: 10.1002/andp.19955070504.

[54] L. R. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001, DOI: 10.1088/1748-0221/3/08/S08001.

[55] P. Mouche, *Overall view of the LHC. Vue d'ensemble du LHC*, 2014, General Photo, https://cds.cern.ch/record/1708847.

[56] C. De Melis, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, 2016, General Photo, https://cds.cern.ch/record/2119882.

[57] LHCb collaboration, https://lhcb-outreach.web.cern.ch/2018/12/03/end-of-the-first-phase-and-beginning-of-a-new-lhcb/.

[58] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005, DOI: 10.1088/1748-0221/3/08/S08005.

[59] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352, DOI: 10.1142/S0217751X15300227.

[60] LHCb collaboration, C. Elsasser, *$b\bar{b}$ production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html.

[61] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, 2001.

[62] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808, DOI: 10.1088/1748-0221/9/09/P09007.

[63] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007, 2000.

[64] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029, 2002.

[65] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893, DOI: 10.1088/1748-0221/9/01/P01002.

[66] Ph. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12** (2017) P11016, arXiv:1708.00819, DOI: 10.1088/1748-0221/12/11/P11016.

[67] D. Hutchcroft, *VELO Pattern Recognition*, CERN-LHCb-2007-013, 2007, https://cds.cern.ch/record/1023540.

[68] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, CERN-LHCb-2007-015, 2007, https://cds.cern.ch/record/1033584.

[69] M. Needham and J. Van Tilburg, *Performance of the track matching*, CERN-LHCb-2007-020, 2007, https://cds.cern.ch/record/1020304.

[70] M. Needham, *Performance of the Track Matching*, CERN-LHCb-2007-129, 2007, https://cds.cern.ch/record/1060807.

[71] O. Callot, *Downstream Pattern Recognition*, CERN-LHCb-2007-026, 2007, https://cds.cern.ch/record/1025827.

[72] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444, DOI: 10.1016/0168-9002(87)90887-4.

[73] J. Van Tilburg, *Track simulation and reconstruction in LHCb*, CERN-THESIS-2005-040, 2005, https://cds.cern.ch/record/885750.

[74] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037, 2000.

[75] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759, DOI: 10.1140/epjc/s10052-013-2431-9.

[76] LHCb RICH collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, Nucl. Instrum. Meth. **A876** (2017) 221, arXiv:1703.08152, DOI: 10.1016/j.nima.2017.03.009.

[77] LHCb collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036, 2000.

[78] C. Abellan Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, (2020), arXiv:2008.11556.

[79] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010, 2001.

[80] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346, DOI: 10.1088/1748-0221/8/02/P02022.

[81] R. W. Forty and O. Schneider, *RICH pattern recognition*, CERN-LHCb-98-040, 1998, https://cds.cern.ch/record/684714.

[82] R. Calabrese *et al.*, *Performance of the LHCb RICH detectors during LHC Run 2*, JINST **17** (2022) P07013, arXiv:2205.13400, DOI: 10.1088/1748-0221/17/07/P07013.

[83] LHCb collaboration, *LHCb trigger system: Technical Design Report*, CERN-LHCC-2003-031, 2003.

[84] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055, DOI: 10.1088/1748-0221/8/04/P04022.

[85] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790, DOI: 10.1088/1748-0221/14/04/P04013.

[86] Martin-Stirling-Thorne-Watt, *Parton Distribution Functions*, 2019, https://mstwpdf.hepforge.org/pdf4lhc/.

[87] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, DOI: 10.1088/1748-0221/8/02/P02013.

[88] I. Bird, *Computing for the Large Hadron Collider*, Annu. Rev. Nucl. Part. Sci. **61** (2011) 99, DOI: 10.1146/annurev-nucl-102010-130059.

[89] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820, DOI: 10.1016/j.cpc.2008.01.036.

[90] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047, DOI: 10.1088/1742-6596/331/3/032047.

[91] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152, DOI: 10.1016/S0168-9002(01)00089-4.

[92] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comp. Phys. Comm. **199** (2016) 86, arXiv:1011.0937, DOI: 10.1016/j.cpc.2015.09.013.

[93] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250, DOI: 10.1016/S0168-9002(03)01368-8.

[94] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270, DOI: 10.1109/TNS.2006.869826.

[95] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023, DOI: 10.1088/1742-6596/331/3/032023.

[96] D. Müller, M. Clemencic, G. Corti, and M. Gersabeck, *ReDecay: A novel approach to speed up the simulation at LHCb*, Eur. Phys. J. **C78** (2018) 1009, arXiv:1810.10362, DOI: 10.1140/epjc/s10052-018-6469-6.

[97] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, arXiv:1612.07489, DOI: 10.1016/j.cpc.2017.01.029.

[98] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191, DOI: 10.1016/j.nima.2005.06.078.

[99] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle γ with $B^{\mp} \to D[K^{\pm}\pi^{\mp}\pi^{\mp}\pi^{\pm}]h^{\mp}$ decays using a binned phase-space approach*, JHEP **07** (2023) 138, arXiv:2209.03692, DOI: 10.1007/JHEP07(2023)138.

[100] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, 1984, Wadsworth international group, Belmont, California, USA.

[101] H. Voss, A. Hoecker, J. Stelzer, and F. Tegenfeldt, *TMVA - Toolkit for Multivariate Data Analysis with ROOT*, PoS **ACAT** (2007) 040, DOI: 10.22323/1.050.0040.

[102] A. Blum, A. Kalai, and J. Langford, *Beating the hold-out: bounds for k-fold and progressive cross-validation*, Proceedings of the Twelfth Annual Conference on Computational Learning Theory, COLT '99, (New York, NY, USA) (1999) 203, DOI: 10.1145/307400.307439.

[103] F. Suljik *et al.*, *Measurement of the CKM angle γ in $B^\pm \to DK^{*\pm}$ decays*, Internal analysis note.

[104] S. Harnew *et al.*, *Model-independent determination of the strong phase difference between D^0 and $\overline{D}^0 \to \pi^+\pi^-\pi^+\pi^-$ amplitudes*, JHEP **01** (2018) 144, arXiv:1709.03467, DOI: 10.1007/JHEP01(2018)144.

[105] LHCb collaboration, R. Aaij *et al.*, *A study of CP violation in the decays $B^\pm \to [K^+K^-\pi^+\pi^-]_D h^\pm$ ($h = K, \pi$) and $B^\pm \to [\pi^+\pi^-\pi^+\pi^-]_D h^\pm$*, Eur. Phys. J. **C84** (2023) 547, arXiv:2301.10328, DOI: 10.1140/epjc/s10052-023-11560-5.

[106] LHCb collaboration, R. Aaij *et al.*, *Simultaneous determination of CKM angle γ and charm mixing parameters*, JHEP **12** (2021) 141, arXiv:2110.02350, DOI: 10.1007/JHEP12(2021)141.

[107] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021, 2016, https://cds.cern.ch/record/2202412.

[108] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1, DOI: 10.1214/aos/1176344552.

[109] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083, DOI: 10.1016/j.nima.2005.08.106.

[110] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP observables in $B^\pm \to D^{(*)}K^\pm$ and $B^\pm \to D^{(*)}\pi^\pm$ decays*, Phys. Lett. **B777** (2018) 16, arXiv:1708.06370, DOI: 10.1016/j.physletb.2017.11.070.

[111] Particle Data Group, R. L. Workman *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2022** (2022) 083C01, DOI: 10.1093/ptep/ptac097.

[112] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron production fractions in 7 TeV pp collisions*, Phys. Rev. **D85** (2012) 032008, arXiv:1111.2357, DOI: 10.1103/PhysRevD.85.032008.

[113] Belle collaboration, A. Poluektov *et al.*, *Measurement of ϕ_3 with Dalitz plot analysis of $B^+ \to D^{(*)}K^{(*)+}$ decay*, Phys. Rev. D **73** (2006) 112009, DOI: 10.1103/PhysRevD.73.112009.

[114] BaBar collaboration, P. del Amo Sanchez *et al.*, *Evidence for Direct CP Violation in the Measurement of the Cabbibo-Kobayashi-Maskawa Angle γ with $B^\mp \to D^{(*)}K^{(*)\mp}$ Decays*, Phys. Rev. Lett. **105** (2010) 121801, DOI: 10.1103/PhysRevLett.105.121801.

[115] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, DESY-F31-86-02, 1986, http://inspirehep.net/record/230779/.

[116] R. Barlow, *Extended maximum likelihood*, Nucl. Instrum. Meth. **A297** (1990) 496, DOI: 10.1016/0168-9002(90)91334-8.

[117] E. Norrbin, *Heavy Quark Production Asymmetries*, 1999, arXiv:hep-ph/9909437, DOI: 10.48550/arXiv.hep-ph/9909437.

[118] E. Norrbin and R. Vogt, *Bottom Production Asymmetries at the LHC*, 2000, arXiv:hep-ph/0003056, DOI: 10.48550/arXiv.hep-ph/0003056.

[119] E. Norrbin and T. Sjöstrand, *Production and hadronization of heavy quarks*, Eur. Phys. J. C **17** (2000) 137, DOI: 10.1007/s100520000460.

[120] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP observables in $B^\pm \to D^{(*)}K^\pm$ and $B^\pm \to D^{(*)}\pi^\pm$ decays using two-body D final states*, JHEP **04** (2021) 081, arXiv:2012.09903, DOI: 10.1007/JHEP04(2021)081.

[121] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B^\pm$ production asymmetry and the CP asymmetry in $B^\pm \to J/\psi K^\pm$ decays*, Phys. Rev. **D95** (2017) 052005, arXiv:1701.05501, DOI: 10.1103/PhysRevD.95.052005.

[122] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60, DOI: 10.1214/aoms/1177732360.

[123] Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021*, Phys. Rev. **D107** (2023) 052008, arXiv:2206.07501, updated results and plots available at https://hflav.web.cern.ch, DOI: 10.1103/PhysRevD.107.052008.

[124] BaBar collaboration, P. del Amo Sanchez *et al.*, *Measurement of D^0–$\bar{D}^0$ mixing parameters using $D^0 \to K^0_S\pi^+\pi^-$ and $D^0 \to K^0_S K^+ K^-$ decays*, Phys. Rev. Lett. **105** (2010) 081803, DOI: 10.1103/PhysRevLett.105.081803.

[125] M. Kenzie *et al.*, *GammaCombo: A statistical analysis framework for combining measurements, fitting datasets and producing confidence intervals*, DOI: 10.5281/zenodo.3371421.

[126] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle γ from a combination of LHCb results*, JHEP **12** (2016) 087, arXiv:1611.03076, DOI: 10.1007/JHEP12(2016)087.

[127] B. Sen, M. Walker, and M. Woodroofe, *On the unified method with nuisance parameters*, Statistica Sinica **19** (2009) 301.

[128] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. D **57** (1998) 3873, DOI: 10.1103/PhysRevD.57.3873.

[129] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, 2012.

[130] LHCb collaboration, R. Aaij *et al.*, *The LHCb Upgrade I*, JINST **19** (2024) P05065, arXiv:2305.10515, DOI: 10.1088/1748-0221/19/05/P05065.

[131] LHCb collaboration, *LHCb VELO Upgrade Technical Design Report*, CERN-LHCC-2013-021, 2013.

[132] LHCb collaboration, *LHCb Tracker Upgrade Technical Design Report*, CERN-LHCC-2014-001, 2014.

[133] LHCb collaboration, *LHCb PID Upgrade Technical Design Report*, CERN-LHCC-2013-022, 2013.

[134] LHCb collaboration, *LHCb Trigger and Online Upgrade Technical Design Report*, CERN-LHCC-2014-016, 2014.

[135] LHCb collaboration, *LHCb Upgrade GPU High Level Trigger Technical Design Report*, CERN-LHCC-2020-006, 2020.

[136] LHCb collaboration, *Trigger Schemes*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html.

[137] LHCb collaboration, *HLT1 trigger efficiencies in 2024 data*, LHCB-FIGURE-2024-030, 2024, https://lbfence.cern.ch/alcm/public/figure/details/3837, https://cds.cern.ch/record/2912743.

[138] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801, arXiv:1703.05747, DOI: 10.1103/PhysRevLett.118.191801.

[139] LHCb collaboration, R. Aaij *et al.*, *Observation of the annihilation decay mode $B^0 \to K^+K^-$*, Phys. Rev. Lett. **118** (2017) 081801, arXiv:1610.08288, DOI: 10.1103/PhysRevLett.118.081801.

[140] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2019) 1, arXiv:1803.00824, DOI: 10.1140/epjti/s40485-019-0050-z.

[141] L. Anderlini, V. Gligorov, O. Lupton, and B. Sciascia, *Calibration samples for particle identification at LHCb in Run 2*, LHCb-PUB-2016-005, 2016, https://cds.cern.ch/record/2134057.

[142] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01, DOI: 10.1093/ptep/ptaa104.

[143] LHCb collaboration, *RICH Performance Plots - Cherenkov Angle Resolutions for RICH 1 and RICH 2*, LHCB-FIGURE-2023-007, 2023, https://lbfence.cern.ch/alcm/public/figure/details/242, https://cds.cern.ch/record/2859118.

[144] LHCb collaboration, R. Aaij *et al.*, and A. Bharucha *et al.*, *Implications of LHCb measurements and future prospects*, Eur. Phys. J. **C73** (2013) 2373, arXiv:1208.3355, DOI: 10.1140/epjc/s10052-013-2373-2.

[145] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, CERN-LHCC-2018-027 LHCb-PUB-2018-009, 2018, arXiv:1808.08865.