



**This electronic thesis or dissertation has been
downloaded from the University of Bristol Research
Portal, <http://research-information.bristol.ac.uk>**

Author:

Wang, Rui

Title:

**Unbinned amplitude analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay using
an amplitude ansatz method at the LHCb experiment**

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited on the University of Bristol Research Portal. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Unbinned amplitude analysis of the
 $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay
using an amplitude ansatz method
at the LHCb experiment

By

WANG RUI



Department of Physics
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Science.

OCTOBER 2024

Abstract

An amplitude analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay is presented in this thesis. The measurement makes use of data samples collected by the LHCb experiment between the years 2011 and 2018, which correspond to an integrated luminosity of 9 fb^{-1} pp collision data. This analysis performs an unbinned fit using a model that considers the differential branching fractions of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay, which can be written in the basis of the K^{*0} transversity amplitudes, with acceptance effects and background taken into account. The transversity amplitudes are measured as functions of dimuon invariant mass squared, q^2 , and are modelled with Legendre polynomial ansatzes aiming to remove model dependencies as much as possible from the description of experimental datasets. These amplitudes contain P-wave amplitudes, while the S-wave amplitudes of the $K\pi$ system are also accounted for.

This analysis is performed in two q^2 regions, $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ and $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$. In both regions, the likelihood surfaces of individual parameters are carefully examined. The results in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region present correct statistical coverage while the results from the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region still need further investigation. The best-fit points of the P-wave amplitudes with uncertainties and correlations will be provided as the final result of this analysis, such that synthetic datasets with the acceptance effects and background subtracted can be generated. Theorists can then use any model to fit the synthetic data to determine the vector current ($\mathcal{C}_9$) and axial-vector current ($\mathcal{C}_{10}$) Wilson coefficients. The theoretical models chosen can potentially provide an explanation for the anomalies seen in $b \rightarrow s \ell^+ \ell^-$ transitions.

Acknowledgements

Pursuing a PhD degree in a country 5000 miles away from home during a pandemic can never be easy, but I am lucky enough to stay in a wonderful place and meet some amazing people who made it possible. The Bristol Particle Physics group has been my home for the past four years and I will always remember the golden days I spent there.

I would first like to thank my supervisor Kostas Petridis, for being the best supervisor and a great friend. You have led me into a truly inspiring project that I wholeheartedly enjoyed and this thesis would have been impossible without you. You have taken your time to educate me on not only flavour physics, but all things related as well. Sitting down and coding together with you, online or offline, will always be something I cherish. I would also like to thank my second supervisor Jonas Rademacker for being a constant source of enthusiasm and positivity. I am grateful to the China Scholarship Council – University of Bristol joint-funded PhD Scholarship that allowed me to do this work.

Special thanks to Lakshan Ram Madhan Mohan for the constant support and selfless help whenever I need, especially in my first two years. I owe a huge debt to your wisdom, without which the construction of this analysis framework could have been much more painful.

I am lucky to have Matthew Birch, not only as a great collaborator but also as a kind friend. All the physics and non-physics conversations were enjoyable, as well as the delicious cuisine we had together. I am also enormously grateful to Mike McCann, for physics and the delicious gelato. My thanks extend to Paula Álvarez Cartelle for useful discussions. You guys have taken care of the wibbly wobbly q^2 thing really well.

I would also like to thank the rest members of the Bristol LHCb group, Paras, Alex, Zahra, Camille, Ben, Richard, Jake A, Jake R, Jozie and Zak, for making the group a big family. Shout out to my deskmate Florian for the constant offer of alcohol. Thank you to the class of 2019 squad: Mei-Li, Charlotte, Rob, Katie, Chris and Raul. I always miss the pub lunches on Fridays before the pandemic.

My final appreciation goes to my Mom BI Yonghua and Dad WANG Huipeng. I have had good times and made bad mistakes, but you were with me throughout the way. I want you to know that I am as proud of you as you are of me unconditionally. Thank you LI Yue for being the best partner. You have completed my soul.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE:

Author's Contribution

The measurement presented in this thesis was carried out by the author between September 2019 and November 2023, with the support of the Bristol LHCb group, the Bristol Particle Physics group and an analysis team of members of the LHCb collaboration. The author's contributions are detailed in the following list:

- Chapter 4: The data analysed in this analysis is collected by the LHCb experiment operated by the LHCb collaboration. The simulation samples are produced by the LHCb simulation group. Both the data and simulation samples have the LHCb software applied to them. The software was written by subgroups of the LHCb collaboration. The author developed the preselection detailed in Section 4.2 with other collaborators. The author recognises the development of the multivariate classifiers to reduce the peaking backgrounds, the $B^+ \rightarrow K^+ \mu^+ \mu^-$ veto and the combinatorial BDT illustrated in Section 4.4 - 4.6 by other collaborators. The Part-Reco BDTs outlined in Section 4.7 is the result of the author's own work inspired by the work of other members of the LHCb collaboration.
- Chapter 5: The author recognises the development of the corrections applied to simulation samples and of the acceptance effect models by other collaborators.
- Chapter 6: The fit model presented in this chapter is developed and constructed by the author. The study on the control mode and rare mode pseudodata is the author's own work. The results of the control mode fit are compared with the work by other members of the LHCb collaboration.
- Chapter 7: The results presented in this chapter are the results of the author's own work.
- Chapter 8: The evaluation of the two leading systematic uncertainties is the result of the author's own work.
- Appendix B: The author recognises studies documented in this appendix chapter by other members of the LHCb collaboration.
- Appendix C: The author recognises studies documented in this appendix chapter by other collaborators.

Table of Contents

	Page
List of Tables	xiii
List of Figures	xv
1 Introduction	1
2 Theory and Background	5
2.1 Standard Model	5
2.1.1 Lagrangian of the SM	7
2.1.2 Spontaneous symmetry breaking and the Higgs mechanism	8
2.1.3 Flavour structure in the SM	10
2.1.4 Interactions of fermions and gauge bosons in the SM	13
2.1.5 Qualitative description of Quantum Chromodynamics	14
2.2 Effective Field Theory of $b \rightarrow s\ell^+\ell^-$ transitions	15
2.2.1 Operator product expansions	16
2.2.2 The effective Hamiltonian of $b \rightarrow s\ell^+\ell^-$ transitions	17
2.3 Differential decay rate of $B^0 \rightarrow K^{*0}\mu^+\mu^-$	18
2.3.1 Definition of the angular basis	19
2.3.2 Construction of a pure P-wave differential decay rate	21
2.3.3 Treatment of the $m_{K\pi}$ dependence	24
2.4 Helicity amplitudes of the K^{*0} meson	26
2.4.1 Infinitesimal symmetries of the angular distribution	27
2.4.2 Amplitude ansatz method	27
2.5 Experimental status of anomalies in $b \rightarrow s\ell^+\ell^-$ transitions	28
2.5.1 Branching fraction	28
2.5.2 Angular observable	29
2.5.3 Lepton Flavour Universality ratio	30
2.6 Global fits of the Wilson coefficients	32

3	LHCb Experiment at the LHC	35
3.1	Layout of the LHCb detector	35
3.2	Tracking system and magnet	38
3.2.1	VELO	38
3.2.2	Tracking stations and magnets	41
3.3	Calorimeters	42
3.4	Particle identification with Cherenkov detectors	43
3.5	Muon stations	45
3.6	Trigger system	46
3.6.1	Low level trigger	47
3.6.2	High level trigger	48
3.6.3	Changes to the trigger system for the Run 2 period	49
3.6.4	Summary of the trigger lines used in this analysis	50
3.7	Generation of simulated data	50
4	Selection	53
4.1	Stripping and trigger requirements	54
4.2	Preselection	55
4.3	Multivariate classifier	56
4.3.1	Decision tree mechanism	56
4.3.2	Gradient boosting and XGBOOST algorithm	58
4.3.3	kFolding technique to reduce bias	59
4.4	Peaking backgrounds	59
4.5	$B^+ \rightarrow K^+ \mu^+ \mu^-$ veto	67
4.6	Combinatorial BDT	67
4.6.1	BDT training variables	67
4.6.2	Combinatorial BDT training results	68
4.7	Random particle backgrounds and Part-Reco BDTs	79
4.8	Conclusion	84
5	Detector Efficiency Effects	87
5.1	Correction for the differences between data and simulation	87
5.1.1	PID	88
5.1.2	Tracking efficiency	88
5.1.3	Trigger and B^0 kinematic corrections	89
5.2	Parametrisation of the acceptance function	90
5.3	The combined Run 1 and Run 2 acceptance function	95
6	Fitter Setup	97

6.1	The TensorFlow analysis framework	97
6.2	The signal model	98
6.2.1	Angular PDF	98
6.2.2	B mass lineshape	99
6.2.3	Treatment of the B_s^0	100
6.3	The background model	100
6.4	Extended log likelihood fit	101
6.5	Fit strategy	102
6.6	Validation with pseudodata	102
6.6.1	Toy study in the region $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$	103
6.6.2	Toy study in the region $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$	106
6.7	Validation with control mode fit	118
7	Results	123
7.1	Fit to Run 1 + 2 data in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region	123
7.2	Fit to Run 1 + 2 data in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region	129
7.3	Statistically sampling the amplitudes with bifurcated Gaussians	132
7.3.1	Amplitudes and observables in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region	133
7.3.2	Amplitudes and observables in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region . . .	137
7.4	Global fits of Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$	141
8	Systematic Uncertainties	143
8.1	L0 efficiency	144
8.2	Kinematic corrections	145
8.3	Tracking corrections	145
8.4	Acceptance simulation statistics	145
8.5	$m_{K\pi}$ model	147
8.6	B^0 mass model	148
9	Conclusions and Outlook	149
A	Definition of a Crystal Ball function	151
B	Goodness-of-fit	153
C	Fit Bias Study Using Toys Generated from SM Predictions	155
	Bibliography	159

List of Tables

Table	Page
2.1 Integrals of the $m_{K\pi}$ lineshapes used in this analysis.	26
3.1 Conditions for the <code>isMuon</code> variable.	46
3.2 Thresholds of the <code>L0Muon</code> and <code>L0DiMuon</code> variables.	47
3.3 Trigger requirements.	51
4.1 Requirementnets in the <code>B2XMuMu</code> stripping line.	55
4.2 Preselection requirements.	56
4.3 A summary of the peaking background quoted as rates with respect to the signal. .	60
4.4 Training variables used in the peaking background BDTs.	61
4.5 Signal efficiency and background rejection efficiency for each background.	66
4.6 Optimum BDT cut points and efficiencies of the combinatorial BDTs.	78
5.1 The applied <code>nTracks</code> scaling factors for each run period.	88
5.2 Combined acceptance run period scaling	95
6.1 Observables from control mode fits	120
6.2 Comparisons of observables from control mode fits between different periods	121
6.3 Comparisons of P-wave observables between this and the ongoing binned analysis. .	121
8.1 L0 trigger efficiency associated systematic uncertainties.	144
8.2 Kinematic correction associated systematic uncertainties.	146

List of Figures

Figure	Page
1.1 Feynman diagrams of the $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decay in the SM or NP models.	2
2.1 Visualisation of the scalar field potential.	9
2.2 Hierarchy structure of the CKM matrix.	13
2.3 Physical processes described by the operators $\mathcal{O}_{7,9,10}$	17
2.4 Definition of the angular basis for $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ and $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$ decays. . . .	20
2.5 Gluon effects in the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay.	29
2.6 Anomalies in the differential branching fractions of $b \rightarrow s \ell^+ \ell^-$ transitions.	30
2.7 Anomaly in the P'_5 observable.	31
2.8 Global fits of Wilson coefficients C_9 and C_{10}	33
3.1 The LHC complex.	36
3.2 Side view of the LHCb detector.	37
3.3 Simulated $b\bar{b}$ production with respect to the opening angles.	38
3.4 Schematic of a b-hadron decay topology within the VELO.	39
3.5 The schematic representation of a $R\phi$ sensor and the overview of the VELO left half.	39
3.6 IP resolution from the VELO.	40
3.7 PV resolution from the VELO.	40
3.8 Layout of the LHCb tracking system and reconstructed tracks.	41
3.9 Schematic of how the calorimeter system works.	42
3.10 Reconstructed Cherenkov angle as a function of track momentum in the $C_4 F_{10}$ radiator.	44
3.11 Performance of the DLL based PID variable.	44
3.12 Performance of the ProbNN based PID variable.	45
3.13 Schematic of the muon stations as seen from the side and along the beamline. . . .	46
3.14 Dataflow diagrams of the Run 1 and Run 2 trigger systems.	50
4.1 A schematic illustration of the BDT working mechanism of maximum depth 3. . .	57

LIST OF FIGURES

4.2	Result of the training and the ROC curve of the double mis-ID BDT on signal and background.	62
4.3	Result of the training and the ROC curve of the $B_s^0 \rightarrow \phi \mu^+ \mu^-$ BDT on signal and background.	62
4.4	Result of the training and the ROC curve of the muon-pion swap from the J/ψ BDT on signal and background.	63
4.5	Result of the training and the ROC curve of the muon-pion swap from the J/ψ BDT on signal and background.	63
4.6	Result of the training and the ROC curve of the muon-pion swap from the $\psi(2S)$ BDT on signal and background.	64
4.7	Result of the training and the ROC curve of the muon-pion swap from the $\psi(2S)$ BDT on signal and background.	64
4.8	Result of the training and the ROC curve of the $\Lambda_b^0 \rightarrow p K \mu^+ \mu^-, p \rightarrow \pi$ BDT on signal and background.	65
4.9	Result of the training and the ROC curve of the $\Lambda_b^0 \rightarrow p K \mu^+ \mu^-, p \rightarrow K, K \rightarrow \pi$ BDT on signal and background.	65
4.10	Training variables of the Run 1 combinatorial BDT.	69
4.11	Training variables of the Run 2 combinatorial BDT.	72
4.12	Comparison of combinatorial BDT training variables in simulation of sWeighted control mode data for year 2018.	75
4.13	Result of the training of the combinatorial BDT for Run 1 and Run 2.	78
4.14	ROC curves from the training of the combinatorial BDT for Run 1 and Run 2.	78
4.15	Schematics of the $\cos \theta_K$ asymmetry in the range of $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$	79
4.16	Distributions of the variables to train the $K^{*+} \rightarrow K^+ \pi^0$ BDT.	81
4.17	Distributions of the variables to train the $K^{*+} \rightarrow K_S^0 \pi^+$ BDT.	82
4.18	BDT output and ROC curve for the $K^{*+} \rightarrow K^+ \pi^0$ BDT.	83
4.19	BDT output and ROC curve for the $K^{*+} \rightarrow K_S^0 \pi^+$ BDT.	83
4.20	The expected yields of peaking backgrounds for data taken in 2012.	84
4.21	The expected yields of peaking backgrounds for data taken in 2016.	85
4.22	The expected yields of peaking backgrounds for data taken in 2017.	85
4.23	The expected yields of peaking backgrounds for data taken in 2018.	86
5.1	Distribution of nTracks in the 2016 data and simulation before and after the correction scaling.	88
5.2	L0 efficiency derived with the TISTOS method	90
5.3	Kinematics reweighting for Run 1	91
5.4	Kinematics reweighting for 2016	92
5.5	Kinematics reweighting for 2017	93
5.6	Kinematics reweighting for 2018	94

5.7	Distribution of q^2 at generator level	94
5.8	Acceptance projections	96
6.1	Example projections from fitting to a toy generated in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	103
6.2	Residuals of the P-wave parameters from ensembles of fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	104
6.3	Coverage of the P-wave parameters from the ensemble of toy fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	107
6.4	Blinded distribution of the P-wave amplitudes from the ensemble of toy fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	111
6.5	Example projections from fitting to a toy generated in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	112
6.6	Residuals of the P-wave parameters from ensembles of fits in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	113
6.7	Coverage of the P-wave parameters (left-handed parallel amplitude components) from the ensemble of toy fits in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	115
6.8	Blinded distribution of the P-wave amplitudes from the ensemble of toy fits in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	117
6.9	Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to Run 1 data.	118
6.10	Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to 2016 data.	119
6.11	Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to 2017 data.	119
6.12	Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to 2018 data.	120
7.1	Projections from fitting to Run 1 + Run 2 data in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	124
7.2	Likelihood profiles of the P-wave amplitude coefficients in the fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	125
7.3	Projections from fitting to Run 1 + Run 2 data in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	129
7.4	Likelihood profiles of the P-wave amplitude coefficients in the fit in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	130
7.5	Likelihood profiles of two amplitude coefficients in the fit to a pseudoexperiment with $10\times$ current statistics in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	132
7.6	The ensemble of amplitudes calculated based on the bifurcated Gaussian sampling from the data fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	134
7.7	The ensemble of observables calculated based on the bifurcated Gaussian sampling from the data fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	135
7.8	The ensemble of amplitudes calculated based on the bifurcated Gaussian sampling from the data fit in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	138

7.9	The ensemble of observables calculated based on the bifurcated Gaussian sampling from the data fit in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.	139
7.10	Global fits of the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ to pseudoexperiments generated from fits to LHCb Run 1 + 2 data samples using different physics models. The orange contour shows a global fit to a $10\times$ statistics pseudoexperiment generated from the dispersion model fit [149] setting $C_9^{\text{NP}} = C_{10}^{\text{NP}} = 0$. The green contour shows this for a $10\times$ statistics pseudoexperiment generated from the amplitude ansatz model which was fit to the same dispersion model pseudoexperiment mentioned above. The red contour shows an example $1\times$ statistics pseudoexperiment. The axes are blinded.	141
8.1	The ensemble of the S-wave amplitude $\mathcal{A}_{00}^L$ calculated based on the bifurcated Gaussian sampling from the data fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.	147
B.1	Goodness-of-fit test for the modelling of acceptance effects on the 2016 FLATQ2 simulation.	154
C.1	Distribution of fits of $Re(A_{\parallel}^L)$ to toys generated from the SM amplitude predictions.	155
C.2	S3 and S4 at $q^2 = 6$ where only $Re(A_{\parallel}^L)$ and the RH amplitudes are floated.	156
C.3	Toys generated close to the S3/S4 boundary and distribution of fits to $Re(A_{\parallel}^L)$	156
C.4	Toys generated away to the S3/S4 boundary distribution of fits to $Re(A_{\parallel}^L)$	156
C.5	Distribution of $Re(A_{\parallel}^L)$ in the ensemble of toy fits where no, upper and lower 5%, 10%, 15%, 20% of toys in the other amplitudes are cut out.	157

Chapter 1

Introduction

The Standard Model (SM) of particle physics has been the most successful theory to date describing the fundamental constituents of matter and how they interact with each other. However, the SM has its limitations and drawbacks. For example, it does not explain the content of dark matter [1] in the universe, nor the predominance of matter over antimatter [2]. Gravity is not incorporated either. Hence, it is widely agreed that the SM is an incomplete description of nature with new particles and interactions yet to be discovered, motivating the search for New Physics (NP).

Among all high energy physics experiments, there are typically two types of searches conducted to discover NP. Direct searches attempt to produce NP particles directly in high energy particle beam collisions which can be subsequently measured by, e.g. a peak in the invariant mass spectrum of the decay products. The ATLAS and CMS collaborations conduct such experiments. The discovery of the Higgs boson in 2012 at ATLAS [3] and CMS [4] follows the direct search principle, marking the discovery of the final building block of the SM. Direct searches are limited by the energy of the particle beam that creates the NP particles. Another type is indirect searches, accomplished by performing very precise measurements of known decays and comparing them with SM predictions. Any discrepancies, also referred to as anomalies, would be a hint of virtual NP contributions to the decay. Indirect searches are not limited by beam energy, but a good understanding of the SM is vital to making reliable predictions, which often rely on approximation techniques.

In recent years, the anomalies in the $b \rightarrow s\ell^+\ell^-$ transitions became one of the most promising candidate signals for new physics, including branching fraction measurements [5–12], Lepton Flavour Universality (LFU) checks [13–17] and angular observables [18–22]. A large contribution to these flavour anomalies comes from the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay. The analysis of this decay can be performed using an Effective Field Theory (EFT) approach. All short-distance physics, including NP, can be encoded in the so-called Wilson coefficients. The $b \rightarrow s\ell^+\ell^-$ process is forbidden at the tree level in the SM and can only occur through electroweak penguin diagrams, where a new quark-antiquark pair emerges from a particle loop mediated by a boson, and box

diagrams, where particles exchange bidirectionally that results in the quark types swapped [23]. These processes are shown at the top of Figure 1.1. In NP theories, the anomalies seen in the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay can be explained by introducing contributions from new vector or axial-vector particles to the decays at the loop level, such as a Z' boson or leptoquarks. The Feynman diagrams for these processes are also shown at the bottom of Figure 1.1. One can then construct observables which are less affected by hadronic uncertainties for the angular distribution of the $K^+ \pi^- \mu^+ \mu^-$ system in the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay.

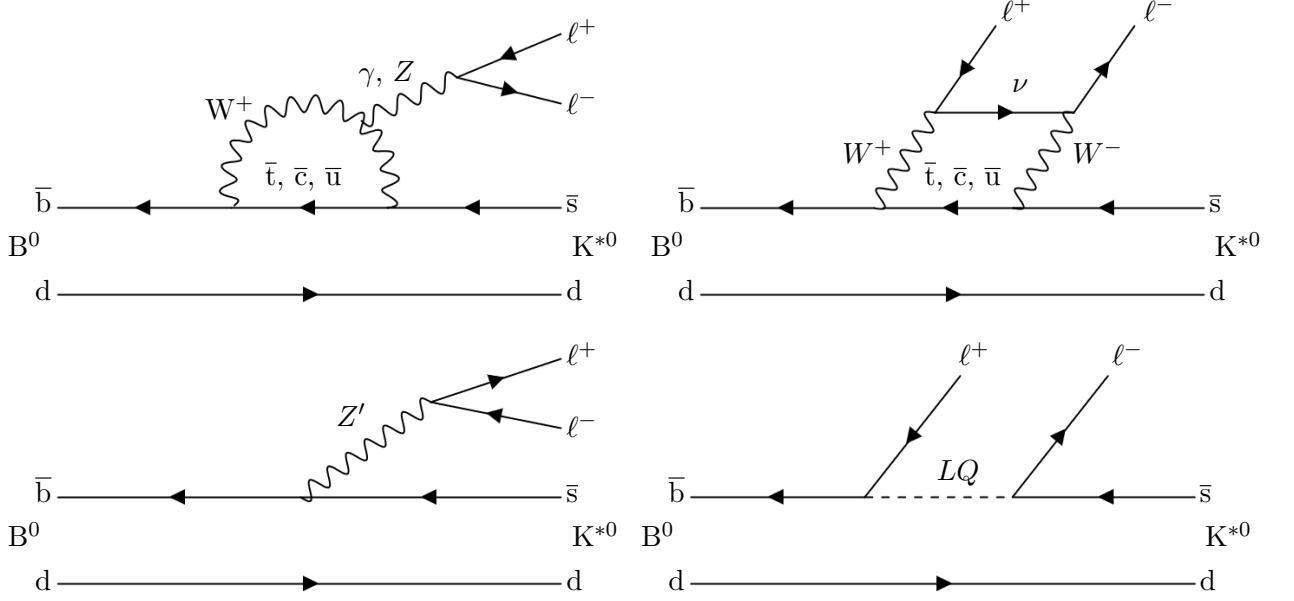


Figure 1.1: Feynman diagrams in the SM of the $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decay, mediated through the electroweak penguin (top left) and box (top right) diagrams. Possible NP contributions include a tree-level diagram mediated by a new gauge boson Z' (bottom left) or a Leptoquark LQ (bottom right).

In previous and one other ongoing analyses on the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay, the analysis is carried out by integrating over bins in q^2 . Such integrations in bins cannot fully exploit the information in q^2 and will result in a decrease in the precision of the final results. The probability density function (PDF) of the model may fall into unphysical regions as well. There are also some unbinned analyses ongoing in the LHCb collaboration, but all of them depended on a certain theory model to describe the decay. They are thus highly dependent on the choice of theoretical models and the associated uncertainties. Motivated by the method proposed in Ref. [24], this analysis determines all P-wave K^{*0} amplitudes as a Legendre parametric function of q^2 , such that the fit can be performed unbinned and the theoretical model dependencies are removed as much as possible.

This thesis is organised as follows:

- Chapter 2 provides the theory foundation of this analysis. It gives a brief introduction to different elements of the SM Lagrangian density before introducing the EFT of $b \rightarrow s \ell^+ \ell^-$ transitions. It then gives the differential decay rate of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, introduces the K^* helicity amplitudes and explains the Legendre polynomial ansatz to model the amplitudes. Finally, an overview of the experimental status of b anomalies is given.
- Chapter 3 introduces the LHCb experiment at CERN and its detector elements as the experimental setup.
- Chapter 4 outlines the development of the selection used in this analysis and the corresponding background studies.
- Chapter 5 illustrates the acceptance effects and the way to model it. Corrections applied to simulation samples to match data are also briefly mentioned.
- Chapter 6 combines all necessary parts mentioned above together to construct a PDF describing the experimental data, based on which an extended negative log likelihood (NLL) function to be minimised in the fit is defined. The fitter is validated through toy studies and fits in the $B^0 \rightarrow J/\psi K^{*0}$ control mode. The toy study also validates the method to construct the error intervals in this analysis.
- Chapter 7 presents the results from fit of data in two q^2 regions. The likelihood profiles of the amplitude parameters are shown and the amplitudes and observables built on top of these parameters are also examined. These results can be interpreted through global fits of Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$.
- Chapter 8 introduces two systematic uncertainties that are expected to be dominant. They are shown as fractions of the statistical uncertainties.
- Chapter 9 draws the conclusions and presents the outlook.

Chapter 2

Theory and Background

This chapter introduces the necessary theoretical and experimental concepts that are relevant to the amplitude analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay. Section 2.1 gives an overview of the Standard Model, with references to [25–30]. Section 2.2 and 2.3 introduces the concept of effective field theories and construct the theoretical description of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay, while Section 2.4 provides the ansatz method modelling the amplitudes of the decay. Finally, a brief overview of the current status of experimental measurements for the $b \rightarrow s \ell^+ \ell^-$ transitions is given in Section 2.5.

2.1 Standard Model

The Standard Model (SM) is a Quantum Field Theory (QFT) that successfully describes three out of four fundamental forces in nature, namely the electromagnetic, weak and strong interaction. It is based on the gauge group

$$(2.1) \quad SU(3)_C \times SU(2)_L \times U(1)_Y$$

The gauge symmetry $SU(3)_C$ is the symmetry of the quantum chromodynamics (QCD) that describes strong interactions. The quantum number associated with this gauge symmetry is called colour and takes three values denoted as red, green and blue. This symmetry remains exact in the SM Lagrangian.

The product groups $SU(2)_L \times U(1)_Y$ describe the unified electroweak interaction. The charges related to each of the two groups are called weak isospin and hypercharge respectively. There are two weak isospin states, $\pm \frac{1}{2}$, and the hypercharge whose value satisfies the Gell-Man-Nishima relation

$$(2.2) \quad Q = T^3 + Y/2 ,$$

where Q is the electric charge, T^3 is the weak isospin charge and Y denotes the hypercharge. The product group $SU(2)_L \times U(1)_Y$ is spontaneously broken down to the symmetry group of quantum electrodynamics (QED), $U(1)_{EM}$.

The fundamental SM particles can be divided into two categories: fermions and bosons. The fermion field ψ can be decomposed into fields with left- and right-handed chiralities

$$(2.3) \quad \begin{aligned} \psi &= \psi_L + \psi_R, \\ \psi_L &= P_L \psi, \quad \psi_R = P_R \psi \end{aligned}$$

where P_L and P_R are chiral projection operators given by

$$(2.4) \quad P_L \equiv \frac{1 - \gamma_5}{2}, \quad P_R \equiv \frac{1 + \gamma_5}{2}$$

with $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^4$ being the product of four Dirac matrices γ^μ .

The fermions observed so far can be further categorised into quarks and leptons. Both quarks and leptons consist of three generations. Fermions among the three generations share exactly the same quantum numbers but are different in mass. The six quarks grouped in three generations are the up (u) and down (d) quark, the charm (c) and strange (s) quark, and the top (t) and bottom (b) quark. Their fields can be expressed by three doublets

$$(2.5) \quad Q^i = \begin{bmatrix} u_L^i \\ d_L^i \end{bmatrix} \in \left\{ \begin{bmatrix} u_L \\ d_L \end{bmatrix}, \begin{bmatrix} c_L \\ s_L \end{bmatrix}, \begin{bmatrix} t_L \\ b_L \end{bmatrix} \right\}$$

and three singlets of both up- and down-types

$$(2.6) \quad u^i = (u_R^i) \in \{u_R, c_R, t_R\}$$

and

$$(2.7) \quad d^i = (d_R^i) \in \{d_R, s_R, b_R\}.$$

In order of increasing mass, the charged leptons consist of electrons (e), muons (μ) and taus (τ), accompanied by the electron neutrino (ν_e), the muon neutrino (ν_μ) and the tau neutrino (ν_τ). Their fields can be arranged in a similar way, where the three weak isospin doublets are expressed as

$$(2.8) \quad L^i = \begin{bmatrix} \nu_L^i \\ \ell_L^i \end{bmatrix} \equiv \left\{ \begin{bmatrix} \nu_{eL} \\ e_L \end{bmatrix}, \begin{bmatrix} \nu_{\mu L} \\ \mu_L \end{bmatrix}, \begin{bmatrix} \nu_{\tau L} \\ \tau_L \end{bmatrix} \right\}$$

and three singlets are expressed as

$$(2.9) \quad \ell^i = (\ell_R^i) \equiv \{e_R, \mu_R, \tau_R\}.$$

As right-handed neutrinos have never been observed, they are not included in the SM Lagrangian.

Quarks carry the colour charge and can feel the strong force, while leptons do not. All fermions feel the weak force since they all carry weak isospins. Particles carrying an electric charge include all quarks, electrons, muons and taus, and they feel the electromagnetic force. All

of these particles have their corresponding antiparticle states with electric and colour charges flipped.

The invariance of the local gauge predicts the existence of massless gauge bosons with spin-1 for each gauge group in 2.1. There are eight gluons G_μ^a for $SU(3)_C$, three charged current W_μ^a for $SU(2)_L$ and one neutral current B_μ for $U(1)_Y$. The electroweak symmetry breaking of $SU(2)_L \times U(1)_Y$ mix W_μ^a and B_μ into $W_\mu^\pm$, Z^0 and γ . γ belongs to the unbroken symmetry $U(1)_{EM}$ and is the only boson that stays massless.

2.1.1 Lagrangian of the SM

The Lagrangian density of the SM can be written as

$$(2.10) \quad \mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Fermion}} + \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} .$$

It can be seen that the SM Lagrangian is split into four terms. The first two terms will be briefly introduced below and the latter two will be discussed in detail in later sections.

The first term $\mathcal{L}_{\text{Fermion}}$ describes the kinematics of the SM fermions and their couplings to gauge bosons. It is given by

$$(2.11) \quad L_{\text{fermion}}^{\text{SM}} = \sum_{j=1,2,3} \bar{L}_j i D_\mu \gamma^\mu L_j + \bar{L}_{Rj} i D_\mu \gamma^\mu L_{Rj} + \bar{Q}_j i D_\mu \gamma^\mu Q_j + \bar{u}_{Rj} i D_\mu u_{Rj} + \bar{d}_{Rj} i D_\mu d_{Rj}$$

The covariant derivative D_μ is defined as

$$(2.12) \quad D_\mu = \partial_\mu - i g_1 Y B_\mu - i g_2 \frac{\sigma^a}{2} W_\mu^a \quad \text{for lepton doublets } L_j,$$

$$(2.13) \quad D_\mu = \partial_\mu - i g_1 Y B_\mu \quad \text{for lepton singlets } L_{Rj},$$

$$(2.14) \quad D_\mu = \partial_\mu - i g_1 Y B_\mu - i g_2 \frac{\sigma^a}{2} W_\mu^a - i g_s \frac{\lambda^a}{2} G_\mu^a \quad \text{for quark doublets } Q_j,$$

$$(2.15) \quad D_\mu = \partial_\mu - i g_1 Y B_\mu - i g_s \frac{\lambda^a}{2} G_\mu^a \quad \text{for quark singlets } u_{Rj}, d_{Rj},$$

and is the same for all three generations $j = 1, 2, 3$. Here σ^a are the Pauli matrices, λ^a are the Gell-mann matrices, and g_1, g_2, g_s stand for the couplings for the gauge groups $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$ correspondingly. It can be seen that gauge boson fields couple to fermion fields.

The second term describes the kinematics of the gauge bosons and is expressed as

$$(2.16) \quad L_{\text{gauge}}^{\text{SM}} = -\frac{1}{4} G_{\mu\nu}^a G^{\mu\nu,a} - \frac{1}{4} W_{\mu\nu}^a W^{\mu\nu,a} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu},$$

where

$$(2.17) \quad \begin{aligned} G_{\mu\nu}^a &= \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \\ W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_2 \epsilon^{abc} W_\mu^b W_\nu^c, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned}$$

with ϵ^{abc} and f^{abc} being the structure constants of the group $SU(2)_L$ and $SU(3)_C$ respectively. This expression suggests that gauge bosons couple to themselves.

2.1.2 Spontaneous symmetry breaking and the Higgs mechanism

The SM forbids mass terms. For the fermions, mass terms of the form $-m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$ involves couplings between left- and right-handed fields that transform differently under $SU(2)_L$ and does not leave the Lagrangian invariant. Analogously, mass terms for the gauge bosons $M^2 A_\mu^a A^{\mu,a}$ break gauge symmetry and are hence forbidden. However, both fermions and bosons are observed to be massive. The Higgs mechanism introduces a complex scalar field charged under $SU(2)_L$ with a non-zero vacuum expectation value that provides a solution to this problem. This results in the spontaneously symmetry breaking $SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM}$ while preserving the invariance of the Lagrangian. The kinematic term of the scalar field in the SM Lagrangian is given by

$$(2.18) \quad \mathcal{L}_{\text{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) ,$$

where the covariant derivative D_μ is as defined in Equation 2.12. It contains a kinetic term and a scalar term. The scalar term $V(\Phi)$ takes the form

$$(2.19) \quad V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 .$$

For $\mu^2 > 0$, this potential has a minimum to be

$$(2.20) \quad \Phi^\dagger \Phi = \frac{\mu^2}{2\lambda} ,$$

yielding the vacuum expectation value (vev) of the Higgs field Φ to be

$$(2.21) \quad \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{1}{\sqrt{2}} v .$$

A schematic diagram 2.1 illustrates the shape of the scalar field potential. Without loss of generality, the vacuum expectation value of this field can be chosen to be

$$(2.22) \quad \langle 0 | \Phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} ,$$

and can then be expanded with respect to this expectation value to get

$$(2.23) \quad \Phi = \begin{bmatrix} h_+ + i\chi_+ \\ \frac{v}{\sqrt{2}} + h_0 + i\chi_0 \end{bmatrix} .$$

with h_+ , $\chi_{+,0}$ being real scalar fields. In fact, they are the fields of the so-called Goldstone bosons and can be removed under local gauge transformations, simplifying 2.23 to

$$(2.24) \quad \Phi = \begin{bmatrix} 0 \\ \frac{v}{\sqrt{2}} + h_0 \end{bmatrix} .$$

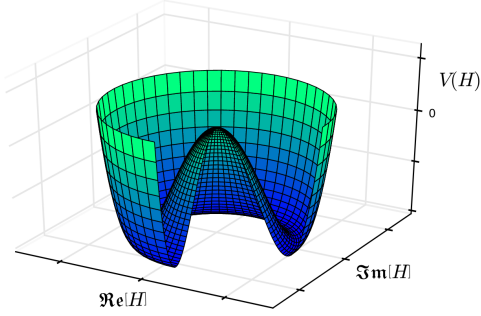


Figure 2.1: A visualisation of the potential in the form of Equation 2.19 with $\mu^2 > 0$. A sector is omitted to display the inside. The ground state of this potential has a non-zero vev , which breaks the symmetry spontaneously. This diagram is taken from [31].

If the Higgs field Φ of Equation 2.19 is rewritten in the form of 2.23, one can get only one mass term

$$(2.25) \quad 2\lambda\nu^2 h_0^2 \equiv m_H^2 h_0^2 \implies m_H = \sqrt{2\lambda}\nu = \sqrt{2}|\mu|.$$

Expanding the first term of the Higgs Lagrangian 2.18 with 2.12, the following expression can be derived:

$$(2.26) \quad \begin{aligned} & \frac{1}{2}(0, v) \left(-igW_\mu^a \tau^a - \frac{i}{2}g'B_\mu \right) \left(igW^{a\mu} \tau^a + \frac{i}{2}g'B^\mu \right) \begin{pmatrix} 0 \\ v \end{pmatrix} \\ &= \frac{1}{2} \frac{v^2}{4} \left[g^2 (W^1)^2 + g^2 (W^2)^2 + (-gW^3 + g'B)^2 \right] \\ &= \frac{1}{2} \left(m_W^2 (W_\mu^+)^{\dagger} W^{\mu+} + m_W^2 (W_\mu^-)^{\dagger} W^{\mu-} + m_Z^2 (Z_\mu^0)^{\dagger} Z^{\mu 0} \right). \end{aligned}$$

The last line of the above expression is obtained by introducing the mass terms of the gauge fields $W_\mu^\pm$ and Z_μ^0 :

$$(2.27) \quad \begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2), \\ Z_\mu^0 &= \frac{1}{\sqrt{g^2 + g'^2}} (gW_\mu^3 - g'B_\mu), \end{aligned}$$

They correspond to the mass terms of the $W^\pm$ and Z_0 boson $m_W = vg/2$ and $m_Z = v\sqrt{g^2 + g'^2}/2$. As can be seen in Equation 2.26, a mass term for the gauge field of the photon A_μ does not exist, which is in alignment with the fact that photons are massless. The gauge field of the photon A_μ is related to W_μ^3 and B_μ through

$$(2.28) \quad A_\mu = \frac{1}{\sqrt{g^2 + g'^2}} (g'W_\mu^3 + gB_\mu).$$

A weak mixing angle θ_W can be defined to mix the gauge fields W_μ^3 and B_μ , given by

$$(2.29) \quad \begin{pmatrix} Z_\mu^0 \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}$$

where

$$(2.30) \quad \cos \theta_W = \frac{g}{\sqrt{g^2 + g'^2}}, \quad \sin \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}.$$

The weak mixing angle θ_W also links the masses of the W and Z bosons via $m_W = m_Z \cos \theta_W$ and the electromagnetic coupling g_e to the couplings g and g' via $g_e = g \sin \theta_W = g' \cos \theta_W$.

2.1.3 Flavour structure in the SM

The mass of the fermions and the flavour structure in the SM originate from the Yukawa interaction between the quark and lepton Dirac fields with the Higgs scalar field. In the SM Lagrangian 2.10, this is described by the Yukawa term $\mathcal{L}_{\text{Yukawa}}$. For quarks and leptons, the Yukawa term can be written as

$$(2.31) \quad \mathcal{L}_{\text{Yukawa}}^q = -y_{ij}^d \overline{Q}_L^i \Phi d_R^j - y_{ij}^u \overline{Q}_L^i \Phi^* u_R^j + \text{h.c.},$$

for the quarks and

$$(2.32) \quad \mathcal{L}_{\text{Yukawa}}^l = -y_{ij}^l \overline{L}_L^i \Phi L_R^j + \text{h.c.},$$

for the leptons. Here Q_L , d_R and u_R are the quark doublets and quark singlets respectively as defined in Equations 2.5 - 2.7 and L_L and L_R are the lepton doublets and lepton singlets respectively as defined in Equations 2.8 - 2.9. y are the couplings between different generations represented by the indices i, j . Finally, h.c. stands for the hermitian conjugate of the corresponding terms, which does not exist for the neutrinos in the SM, and $H^* = i\sigma^2 H$.

Substituting 2.24 into Equation 2.31, the quark Yukawa term becomes

$$(2.33) \quad \begin{aligned} \mathcal{L}_{\text{Yukawa}}^q &= -y_{ij}^d \begin{bmatrix} \overline{u}_L^i & \overline{d}_L^i \end{bmatrix} \begin{bmatrix} 0 \\ \nu/\sqrt{2} + h_0 \end{bmatrix} d_R^j - y_{ij}^d d_R^j \begin{bmatrix} 0 & \nu/\sqrt{2} + h_0 \end{bmatrix} \begin{bmatrix} u_L^j \\ d_L^j \end{bmatrix} \\ &\quad - y_{ij}^u \begin{bmatrix} \overline{u}_L^i & \overline{d}_L^i \end{bmatrix} \begin{bmatrix} \frac{\nu}{\sqrt{2}} + h_0 \\ 0 \end{bmatrix} u_R^j - y_{ij}^u u_R^j \begin{bmatrix} \frac{\nu}{\sqrt{2}} + h_0 & 0 \end{bmatrix} \begin{bmatrix} u_L^j \\ d_L^j \end{bmatrix} \\ &= -\frac{y_{ij}^d \nu}{\sqrt{2}} \overline{d}_L^i d_R^j - \frac{y_{ij}^d \nu}{\sqrt{2}} \overline{d}_R^i d_L^j - \frac{y_{ij}^u \nu}{\sqrt{2}} \overline{u}_L^i u_R^j - \frac{y_{ij}^u \nu}{\sqrt{2}} \overline{u}_R^i u_L^j \\ &\quad + \text{terms with } h_0. \end{aligned}$$

The indices (i, j) in the above expression denote the summation over three generations. Therefore, by defining

$$(2.34) \quad \mathbf{d}_L \equiv \begin{bmatrix} d_L \\ s_L \\ b_L \end{bmatrix}; \quad \mathbf{d}_R \equiv \begin{bmatrix} d_R \\ s_R \\ b_R \end{bmatrix}; \quad \mathbf{Y}^d \equiv \begin{bmatrix} y_{11}^d & y_{12}^d & y_{13}^d \\ y_{21}^d & y_{22}^d & y_{23}^d \\ y_{31}^d & y_{32}^d & y_{33}^d \end{bmatrix}$$

and

$$(2.35) \quad \mathbf{u}_L \equiv \begin{bmatrix} u_L \\ c_L \\ t_L \end{bmatrix}; \quad \mathbf{u}_R \equiv \begin{bmatrix} u_R \\ c_R \\ t_R \end{bmatrix}; \quad \mathbf{Y}^u \equiv \begin{bmatrix} y_{11}^u & y_{12}^u & y_{13}^u \\ y_{21}^u & y_{22}^u & y_{23}^u \\ y_{31}^u & y_{32}^u & y_{33}^u \end{bmatrix};$$

one can write 2.33 in the matrix notation

$$(2.36) \quad \mathcal{L}_{\text{Yukawa}}^q = -\frac{\nu}{\sqrt{2}} \left(\overline{\mathbf{d}}_L \mathbf{Y}^d \mathbf{d}_R + \overline{\mathbf{d}}_R \mathbf{Y}^d \mathbf{d}_L \right) - \frac{\nu}{\sqrt{2}} \left(\overline{\mathbf{u}}_L \mathbf{Y}^u \mathbf{u}_R + \overline{\mathbf{u}}_R \mathbf{Y}^u \mathbf{u}_L \right) + \text{terms with } h_0.$$

Analogously, substituting 2.24 into Equation 2.32 the lepton Yukawa term becomes

$$(2.37) \quad \begin{aligned} \mathcal{L}_{\text{Yukawa}}^\ell &= -y_{ij}^\ell \begin{bmatrix} \overline{\nu}_L^i & \overline{\ell}_L^i \end{bmatrix} \begin{bmatrix} 0 \\ \nu/\sqrt{2} + h_0 \end{bmatrix} \ell_R^j - y_{ij}^\ell \ell_R^i \begin{bmatrix} 0 & \nu/\sqrt{2} + h_0 \end{bmatrix} \begin{bmatrix} \nu_L^j \\ \ell_L^j \end{bmatrix} \\ &= -\frac{y_{ij}^\ell \nu}{\sqrt{2}} \overline{\ell}_L^i \ell_R^j - \frac{y_{ij}^\ell \nu}{\sqrt{2}} \ell_R^i \overline{\ell}_L^j + \text{terms with } h_0. \end{aligned}$$

The first terms are Dirac mass terms for the lepton field. By defining

$$(2.38) \quad \ell_L \equiv \begin{bmatrix} e_L \\ \mu_L \\ \tau_L \end{bmatrix}; \quad \ell_R \equiv \begin{bmatrix} e_R \\ \mu_R \\ \tau_R \end{bmatrix}; \quad \mathbf{Y}^\ell \equiv \begin{bmatrix} y_{11}^\ell & y_{12}^\ell & y_{13}^\ell \\ y_{21}^\ell & y_{22}^\ell & y_{23}^\ell \\ y_{31}^\ell & y_{32}^\ell & y_{33}^\ell \end{bmatrix},$$

Equation 2.37 becomes

$$(2.39) \quad \mathcal{L}_{\text{Yukawa}}^\ell = -\frac{\nu}{\sqrt{2}} \left(\overline{\ell}_L \mathbf{Y}^\ell \ell_R + \overline{\ell}_R \mathbf{Y}^\ell \ell_L \right) + \text{terms with } h_0.$$

Any matrix can be diagonalized with a bi-unitary transformation. Such transformations can be performed on the matrix form of the fermion fields defined in Equation 2.5 to 2.9 with

$$(2.40) \quad \begin{aligned} \ell_L &\rightarrow U_L^\ell \ell'_L, & \ell_R &\rightarrow U_R^\ell \ell'_R, \\ \mathbf{u}_L &\rightarrow U_L^u \mathbf{u}'_L, & \mathbf{u}_R &\rightarrow U_R^u \mathbf{u}'_R, \\ \mathbf{d}_L &\rightarrow U_L^d \mathbf{d}'_L, & \mathbf{d}_R &\rightarrow U_R^d \mathbf{d}'_R, \end{aligned}$$

which leads to the diagonalisation of the Yukawa matrices

$$(2.41) \quad \begin{aligned} \mathbf{Y}_{\text{diag}}^d &= U_L^{d\dagger} \mathbf{Y}^d U_R^d = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{bmatrix}, \\ \mathbf{Y}_{\text{diag}}^u &= U_L^{u\dagger} \mathbf{Y}^u U_R^u = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{bmatrix}, \\ \mathbf{Y}_{\text{diag}}^\ell &= U_L^{\ell\dagger} \mathbf{Y}^\ell U_R^\ell = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{bmatrix}, \end{aligned}$$

where $m_\alpha^{d,u,l}$ are the mass matrices of the quarks and leptons. After this transformation, the Yukawa Lagrangian of quarks and leptons can be written as

$$(2.42) \quad \mathcal{L}_{\text{Yukawa}}^q = -\frac{v}{\sqrt{2}} \left(\overline{\mathbf{d}}'_L \mathbf{Y}_{\text{diag}}^d \mathbf{d}'_R + \overline{\mathbf{u}}'_L \mathbf{Y}_{\text{diag}}^u \mathbf{u}'_R \right) + \text{terms with } h_0 + \text{h.c.} ,$$

$$(2.43) \quad \mathcal{L}_{\text{Yukawa}}^l = -\frac{v}{\sqrt{2}} \left(\overline{\ell}'_L \mathbf{Y}_{\text{diag}}^\ell \ell'_R \right) + \text{terms with } h_0 + \text{h.c.} .$$

However, such transformations cannot diagonalise $\mathbf{Y}^u$ and $\mathbf{Y}^l$ simultaneously. This is because only left-handed fields exist in the quark and lepton doublets and transformations shown in 2.40 only involve U_L^u and U_L^l that diagonalise $\mathbf{Y}^u$ and $\mathbf{Y}^l$, but not U_L^d that diagonalises $\mathbf{Y}^d$. Alternatively, $\mathbf{Y}^u$ cannot be diagonalised with U_L^d . The term $\overline{\mathbf{d}}'_L \mathbf{Y}_{\text{diag}}^d \mathbf{d}'_R$ thus requires an additional transformation to be diagonalised:

$$(2.44) \quad \mathbf{d}'_L \equiv U_L^{u\dagger} U_L^d \mathbf{d}'_L .$$

It can be seen that up and down states are mixed in this transformation, so the transformation is responsible for flavour transitions in the SM. The matrix elements $U_L^{u\dagger} U_L^d$ are the terms of the Cabbibo-Kobayashi-Maskawa (CKM) matrix [32, 33]. Written in the matrix notation, the CKM matrix is given by

$$(2.45) \quad U_L^{u\dagger} U_L^d \equiv \mathbf{V}_{\text{CKM}} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}$$

It is a 3×3 unitary matrix, so there are 9 degrees of freedom. Among this, 5 of the phases can be absorbed by the quark fields. The remaining one-phase and three-angle parameters can be determined experimentally.

The structure of the CKM matrix is visualised in Figure 2.2. Wolfenstein gives a parameterisation of the CKM matrix [34] that better shows its hierarchy structure,

$$(2.46) \quad \mathbf{V}_{\text{CKM}} = \begin{bmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\varrho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \varrho - i\eta) & -A\lambda^2 & 1 \end{bmatrix} + O(\lambda^4) ,$$

with the expansion parameter

$$(2.47) \quad \lambda = |V_{us}| \approx 0.22$$

and parameters A , ϱ and η having values between 0 and 1. The complex elements are the only source of CP asymmetry in the SM, making the determination of the CKM matrix one of the main goals in flavour physics.

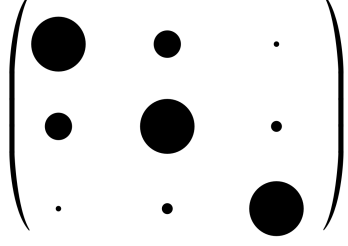


Figure 2.2: Visualisation of the CKM matrix elements. The radius of each circle is proportional to the structure of the corresponding CKM matrix parameter, i.e. $r \approx \sqrt{|V_{ij}|}$. The hierarchical structure of the near-diagonal matrix elements can be seen in this diagram.

2.1.4 Interactions of fermions and gauge bosons in the SM

Revisiting the SM Lagrangian of the fermions in Equation 2.11, the interactions between the quark and lepton fields with gauge boson fields arise from the kinetic terms $i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi$. After spontaneous symmetry breaking and introducing the gauge boson fields $W_\mu^\pm$, Z_μ and A_μ , the covariant derivatives for the quark doublets 2.14 can be rewritten in the mass basis as

$$(2.48) \quad \mathcal{D}_\mu Q^i = \left(\partial_\mu + igA_\mu^g T^g + i\frac{g_2}{\sqrt{2}} \left(W_\mu^+ \frac{\sigma^+}{2} + W_\mu^- \frac{\sigma^-}{2} \right) + i\sqrt{g_1^2 + g_2^2} \left(\frac{\sigma^3}{2} - \sin^2 \theta_W \mathcal{Q} \right) Z_\mu + ie\mathcal{Q}A_\mu \right) Q^i$$

Here the electric charge generator

$$(2.49) \quad \mathcal{Q} = \frac{\sigma^3}{2} + Y = \begin{bmatrix} 1+Y & 0 \\ 0 & Y-1 \end{bmatrix}$$

the electric coupling strength

$$(2.50) \quad e \equiv g_2 s_W$$

where the abbreviation $s_W \equiv \sin \theta_W$ is introduced and

$$(2.51) \quad \sigma^\pm = \sigma^1 \pm i\sigma^2 = \begin{bmatrix} 0 & 1 \pm 1 \\ 1 \mp 1 & 0 \end{bmatrix}$$

are all introduced. The hypercharge Y is assigned as $-1/2$ for L^i and $1/6$ for Q^i correspondingly. Expansions of the coupling term $i\bar{Q}^i\gamma^\mu\mathcal{D}_\mu Q^i$ can thus be obtained for quark states with the charged bosons,

$$(2.52) \quad i\bar{Q}^i\gamma^\mu \frac{ig_2}{\sqrt{2}} \left(W_\mu^+ \frac{\sigma^+}{2} + W_\mu^- \frac{\sigma^-}{2} \right) Q^i,$$

and with the neutral bosons

$$(2.53) \quad i\bar{Q}^i\gamma^\mu i\sqrt{g_1^2 + g_2^2} \left(\frac{\sigma^3}{2} - s_W^2 \mathcal{Q} \right) Z_\mu Q^i + i\bar{Q}^i\gamma^\mu ie\mathcal{Q}A_\mu Q^i.$$

In the mass basis, Equation 2.52 can be written in element notation as

$$\begin{aligned}
 & i\overline{Q}^i\gamma^\mu\frac{ig_2}{\sqrt{2}}\left(W_\mu^+\frac{\sigma^+}{2}\right)Q^i + \text{h.c.} \\
 (2.54) \quad & = \frac{-g_2}{\sqrt{2}}W_\mu^+\left[\overline{u'_{Lp}}V_{pq}d'_{Lq}\right]\left(U_{(L)}^{u\dagger}\right)_{pi}\left(U_{(L)}^u\right)_{ij}\begin{bmatrix} 0 & \gamma^\mu \\ 0 & 0 \end{bmatrix}\begin{bmatrix} u'_{Lj} \\ V_{jk}d'_{Lk} \end{bmatrix} + \text{h.c.} \\
 & = \frac{-g_2}{\sqrt{2}}W_\mu^+V_{jk}\overline{u'_{Lj}}\gamma^\mu d'_{Lk} + \text{h.c.}
 \end{aligned}$$

where V_{jk} stands for the jk^{th} element of the CKM matrix. This expression describes the flavour-changing transition of a down-type quark to an up-type quark mediated by the charged current. In the absence of $W_\mu^\pm$ fields in the covariance of right-handed quarks 2.15, the charged $W^\pm$ boson only couples to left-handed quarks. Parity is hence maximally violated in charged current interactions.

Coming back to the neutral bosons, 2.53 can also be expressed in the mass basis as

$$\begin{aligned}
 (2.55) \quad & i\overline{Q}^i\gamma^\mu i\sqrt{g_1^2 + g_2^2}\left(\frac{\sigma^3}{2} - s_W^2\mathcal{Q}\right)Z^\mu Q^i + i\overline{Q}^i\gamma^\mu ie\mathcal{Q}A_\mu Q^i \\
 & = -\sqrt{g_1^2 + g_2^2}Z_\mu\left[\begin{array}{cc} \overline{u'_{Lp}} & \overline{V_{pq}d'_{Lq}} \end{array}\right]\left(U_{(L)}^{u\dagger}\right)_{pi}\left(U_{(L)}^u\right)_{ij}\begin{bmatrix} \left(\frac{1}{2} - \frac{2}{3}s_W^2\right)\gamma^\mu & 0 \\ 0 & \left(-\frac{1}{2} + \frac{1}{3}s_W^2\right)\gamma^\mu \end{bmatrix}\begin{bmatrix} u'_{Lj} \\ V_{jk}d'_{Lk} \end{bmatrix} \\
 & \quad - eA_\mu\left[\begin{array}{cc} \overline{u'_{Lp}} & \overline{V_{pq}d'_{Lq}} \end{array}\right]\left(U_{(L)}^{u\dagger}\right)_{pi}\left(U_{(L)}^u\right)_{ij}\begin{bmatrix} \frac{2}{3}\gamma^\mu & 0 \\ 0 & -\frac{1}{3}\gamma^\mu \end{bmatrix}\begin{bmatrix} u'_{Lj} \\ V_{jk}d'_{Lk} \end{bmatrix} \\
 & = -\sqrt{g_1^2 + g_2^2}Z_\mu\left(\left(\frac{1}{2} - \frac{2}{3}s_W^2\right)\overline{u'_{Lj}}\gamma^\mu u'_{Lj} + \left(-\frac{1}{2} + \frac{1}{3}s_W^2\right)\overline{d'_{Lk}}\gamma^\mu d'_{Lk}\right) \\
 & \quad - eA_\mu\left(\frac{2}{3}\overline{u'_{Lj}}\gamma^\mu u'_{Lj} - \frac{1}{3}\overline{d'_{Lk}}\gamma^\mu d'_{Lk}\right).
 \end{aligned}$$

This expression describes the Flavour Changing Neutral Current (FCNC) process. It implies that FCNC processes do not mix different quark flavours and are hence forbidden at the tree level in the SM and can only occur at the loop level. FCNC decays can only be mediated by electroweak penguin and box diagrams. An example is the $B^0 \rightarrow K^{*0}\ell^+\ell^-$ transitions whose Feynmann diagrams are shown in the first two plots of Figure 1.1. The branching fractions of the FCNC decays are heavily suppressed in the SM, so if NP effects already enter at the tree level, the decay properties will be significantly changed. FCNC decays are thus excellent probes of NP.

2.1.5 Qualitative description of Quantum Chromodynamics

The strong interactions, described by the theory of QCD, play an important role in $b \rightarrow s\ell^+\ell^-$ transitions. The $b \rightarrow s\ell^+\ell^-$ transitions receive high order corrections from diagrams involving gluons at the quark level. In the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay that this analysis focuses on, there are

meson systems which are bound states of quark and anti-quarks and are described by QCD. Therefore, it is necessary to qualitatively study QCD effects.

The theory of QCD describes the interactions between quarks and gluons. A feature of QCD is called asymptotic freedom, which states that the QCD coupling becomes weaker at high-energy scales, or short distances. Reversely, the coupling becomes stronger at lower energies. Similar to the fine structure constant of electrodynamics, $\alpha_s \equiv g_s^2/4\pi$ is defined. At a certain energy scale ϱ , the strong coupling can be expressed as

$$(2.56) \quad \alpha_s(\varrho) = \frac{6\pi}{(33 - 2n_f) \log(\varrho/\Lambda_{\text{QCD}})} ,$$

where $n_f = 3$ is the number of quark flavours. The scale of QCD is denoted Λ_{QCD} , which is defined as the scale where α_s becomes large with ϱ^2 decreasing. Experimental measurements suggest that $\Lambda_{\text{QCD}} \approx 200 \text{ MeV}/c^2$. This behaviour suggests that perturbative calculations of QCD become difficult at low energies, which is the leading source of theoretical uncertainties in previous studies of $b \rightarrow s\ell^+\ell^-$ transitions.

2.2 Effective Field Theory of $b \rightarrow s\ell^+\ell^-$ transitions

Effective field theory (EFT) is commonly used in the study of flavour physics. The EFT only focuses on the physics relevant to the time or length scale in which an experiment is conducted. Different energy scales are separated, such that low-energy effects, which are difficult to predict due to the non-perturbative nature of the QCD, can be separated from high-energy effects that could potentially include NP contributions. In $b \rightarrow s\ell^+\ell^-$ transitions, the energy threshold of the system is chosen to be

$$(2.57) \quad \varrho = m_b \approx 5\text{GeV}/c^2 \gg \Lambda_{\text{QCD}} ,$$

which describes the dynamics of a single b quark.

A generic starting point of a phenomenology study for weak decays of hadrons is the effective Hamiltonian, which can be written as

$$(2.58) \quad \mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \sum_i V_{\text{CKM}}^i C_i(\varrho) \mathcal{O}_i ,$$

where G_F is the Fermi constant and $\mathcal{O}_i$ are the relevant local operators that describe the decay studied. The V_{CKM}^i is the corresponding CKM factor and $C_i(\varrho)$ is known as Wilson coefficients at the energy scale ϱ , both of which describe the coupling strength of the relevant operator within the effective Hamiltonian. This effective Hamiltonian is obtained by the Operator Product Expansion (OPE) approach. This approach was historically used in, e.g. Ref. [35], for the description of the nuclear beta decay although the CKM factor V_{ud} was not taken into account. A similar type of EFT approach is employed when probing $b \rightarrow s\ell^+\ell^-$ decays because EFT makes no assumptions about the particular structure of the physics, such that the NP effects are included in a model-independent way.

2.2.1 Operator product expansions

The effective Hamiltonian $\mathcal{H}_{\text{eff}}$ in Equation 2.58 can be regarded as a series of effective vertices multiplied by effective coupling constants under the OPE technique. With the energy scale set to ϱ , contributions from scales higher than ϱ are summarised in the Wilson coefficients $C_i(\varrho)$. These Wilson coefficients can be calculated using a perturbation theory if ϱ is not too small as a result of the asymptotic freedom of QCD. These contributions are also referred to as ‘heavy degrees of freedom’ in literature. The interactions from the SM $W^\pm$ - and Z^0 -bosons and t quark all fall into the definition of ‘heavy’ and are thus included in the Wilson coefficients. NP contributions beyond the SM are also included if short distance QCD effects are properly considered. On the other hand, all low energy long distance effects are contained within the operators $\mathcal{O}_i(\varrho)$.

For a transition from an initial state $|i\rangle$ to a final state $\langle f|$, the decay amplitudes can be conveniently computed by constructing the equivalent effective Hamiltonian evaluated at the energy scale ϱ

$$(2.59) \quad \langle f | \mathcal{H}_{\text{eff}} | i \rangle = \sum_i C_i(\varrho) \langle f | \mathcal{O}_i(\varrho) | i \rangle .$$

By using the approach of OPE, the determination of the decay amplitudes is essentially separated into two distinct parts: the short distance calculation of the couplings $C_i(\varrho)$ and the long distance calculation of the matrix elements $\langle \mathcal{O}_i(\varrho) \rangle$. It should be noted that the dependencies of ϱ in both parts cancel out since the amplitude does not depend on a certain energy scale. The Wilson coefficients can be calculated by comparing the effective theory with the full theory in its perturbative regime in the SM [36]. This process is called *matching*. Any heavy NP effects will modulate the observed decay process and they can be studied by comparing the SM predicted and experimentally measured values of the Wilson coefficients by treating them as generalised couplings. Deviations from the SM predicted value of the Wilson coefficients thus provide sensitivity of a broad class of NP particles independent of a specific NP model. Practically the Wilson coefficients can be separated into predictions from the SM and NP contributions as $C_i \equiv C_i^{\text{SM}} + C_i^{\text{NP}}$.

On the other hand, the calculation of matrix elements $\langle \mathcal{O}_i(\varrho) \rangle$ are theoretically challenging. The existence of low energy long distance contributions enforces using non-perturbative methods such as lattice calculations, QCD sum rules, hadronic sum rules, etc. The hadronic matrix elements can be parameterised as summations of Lorentz structures and functions of kinematic variables, which are also known as ‘form factors’. The form factors are the leading source of uncertainties in the prediction of the observables in some $b \rightarrow s \ell^+ \ell^-$ analyses, such as Refs. [9, 12, 37]. This motivates the definition of optimised observables using ratios of form factors such that the uncertainties largely cancel, e.g. the P'_5 observable in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ analyses (see Section 2.5.2).

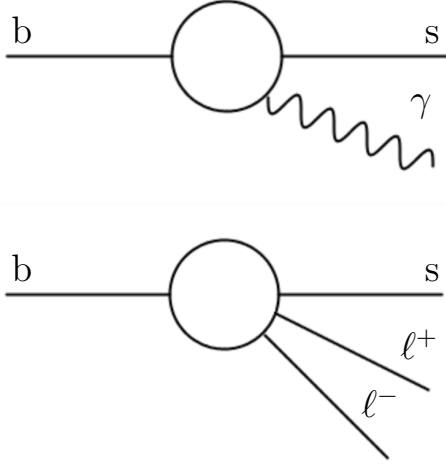


Figure 2.3: Schematic diagrams showing the physical processes described by the operators $\mathcal{O}_7$ (top) and $\mathcal{O}_{9,10}$ (bottom). $\mathcal{O}_7$ relates to diagrams mediated by a radiated photon, $\mathcal{O}_9$ relates to diagrams mediated by a vector current while $\mathcal{O}_{10}$ relates to diagrams mediated by an axial current.

2.2.2 The effective Hamiltonian of $b \rightarrow s \ell^+ \ell^-$ transitions

The effective Hamiltonian of $b \rightarrow s \ell^+ \ell^-$ transitions can be written as [38]

$$(2.60) \quad \mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} \left(\lambda_t \mathcal{H}_{\text{eff}}^{(t)} + \lambda_u \mathcal{H}_{\text{eff}}^{(u)} \right),$$

with λ_i representing the combination of CKM matrix elements $\lambda_i = V_{ib} V_{is}^*$ and

$$\begin{aligned} \mathcal{H}_{\text{eff}}^{(t)} &= C_1 \mathcal{O}_1^c + C_2 \mathcal{O}_2^c + \sum_{i=3}^6 C_i \mathcal{O}_i + \sum_{i=7,8,9,10,S,P,T} (C_i \mathcal{O}_i + C'_i \mathcal{O}'_i), \\ \mathcal{H}_{\text{eff}}^{(u)} &= C_1 (\mathcal{O}_1^c - \mathcal{O}_1^u) + C_2 (\mathcal{O}_2^c - \mathcal{O}_2^u). \end{aligned}$$

The contribution of $\mathcal{H}_{\text{eff}}^{(u)}$ is doubly Cabibbo-suppressed (i.e. $V_{ub} V_{us}$ is tiny) and hence dropped.

The operators $\mathcal{O}_i$ in $\mathcal{H}_{\text{eff}}^{(t)}$ stand for different interaction processes, among which $\mathcal{O}_{7,9,10}$ are expected to be dominant. $\mathcal{O}_7$ is the radiation photon operator that describes a photon decaying into a dilepton pair, whereas $\mathcal{O}_9$ is the vector and $\mathcal{O}_{10}$ the axial vector operator. These processes are illustrated in a schematic diagram 2.3.

These operators are written as

$$(2.61) \quad \begin{aligned} \mathcal{O}_7 &= \frac{e}{g^2} m_b (\bar{s} \sigma_{\mu\nu} P_R b) F^{\mu\nu}; & \mathcal{O}'_7 &= \frac{e}{g^2} m_b (\bar{s} \sigma_{\mu\nu} P_L b) F^{\mu\nu}, \\ \mathcal{O}_9 &= \frac{e^2}{g^2} (\bar{s} \gamma_\mu P_L b) (\bar{\mu} \gamma^\mu \mu); & \mathcal{O}'_9 &= \frac{e^2}{g^2} (\bar{s} \gamma_\mu P_R b) (\bar{\mu} \gamma^\mu \mu), \\ \mathcal{O}_{10} &= \frac{e^2}{g^2} (\bar{s} \gamma_\mu P_L b) (\bar{\mu} \gamma^\mu \gamma_5 \mu); & \mathcal{O}'_{10} &= \frac{e^2}{g^2} (\bar{s} \gamma_\mu P_R b) (\bar{\mu} \gamma^\mu \gamma_5 \mu). \end{aligned}$$

where e is the electromagnetic coupling strength, g is the strong coupling constant and m_b is the running mass of the b quark. $P_{L,R}$ are the chiral projection operators and $F^{\mu\nu}$ is the electromagnetic field strength tensor. The operators with primes have different chirality from the unprimed ones and are vanished or highly suppressed in the SM. For example, the Wilson

coefficient C'_7 is helicity suppressed as $C'^{SM}_7 = (m_s/m_b) C^{SM}_7$, while $C'^{SM}_{9,10} = 0$ due to the absence of right-handed (axial-)vector interactions.

The remainder of the operators is either highly suppressed or the corresponding Wilson coefficients are already well constrained by experiments. Therefore, they are only briefly mentioned. For example, $\mathcal{O}_{1i}$ describe the processes whose diagram contains an internal quark of type i and $\mathcal{O}_{3-6}$ are QCD penguin operators. These operators are generated in some NP models, but their impacts are studied to be small or heavily constrained, so they can be safely neglected. $\mathcal{O}_8$ is the radiative gluon operator, and $\mathcal{O}_{S,P,T}$ describe (pseudo-)scalar or tensor couplings. Their effects are extremely suppressed.

The Wilson coefficients C_i are evaluated by matching them to the full SM theory at the scale $\varrho = m_W$, whose value is big enough for a solid calculation with perturbation theory, before evolving down to the scale of the b-quark, which introduces a mixing between different Wilson coefficients:

$$\begin{aligned}
 C_7^{\text{eff}} &= \frac{4\pi}{\alpha_s} C_7 - \frac{1}{3} C_3 - \frac{4}{9} C_4 - \frac{20}{3} C_5 - \frac{80}{9} C_6, \\
 C_9^{\text{eff}} &= \frac{4\pi}{\alpha_s} C_9 + Y(q^2), \\
 C_{10}^{\text{eff}} &= \frac{4\pi}{\alpha_s} C_{10}, \\
 C'^{\text{eff}}_{7,9,10} &= \frac{4\pi}{\alpha_s} C'_{7,9,10}.
 \end{aligned}
 \tag{2.62}$$

The function $Y(q^2)$ accounts for all non-local long distance effects coupling to a lepton pair via a virtual photon. It is in fact a suppressed combination of further Wilson coefficients and is presented in detail in Ref. [36, 38]. It should be noted that C_9^{eff} receives corrections from charm-loop diagrams $\mathcal{O}_{1c,2c}$ through the inclusion of the coefficients $C_{1c,2c}$ in the function $Y(q^2)$. The coefficients shown in the equation above are called effective Wilson coefficients and appear commonly in the amplitudes. The values of C_7^{eff} , C_9 and C_{10} within the SM are calculated up to the next-to-next-to-leading-logarithm accuracy at the energy scale of $\mu = m_b = 4.8 \text{ GeV}/c^2$ [38]. The values are $C_7^{\text{eff}} = -0.304$, $C_9 = +4.211$ and $C_{10} = -4.103$.

2.3 Differential decay rate of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$

The $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay is an important channel to study the $b \rightarrow s \ell \ell$ transitions. As this decay includes two muons in the final state, the LHCb detector has a better Particle Identification (PID) efficiency and cleaner detector significance compared to that with electrons, so it is experimentally favoured. Since K^{*0} is a short-lived vector particle, the actual decay of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ seen in the experiment is $B^0 \rightarrow K^{*0} (\rightarrow K^+ \pi^-) \mu^+ \mu^-$, where the polarisation states are accessible. Different kinds of observables can be constructed from the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay, including the differential branching fraction [9, 12], CP-asymmetries [39], isospin asymmetries [9]

and angular observables. The angular observables have dependencies on the angular states of the daughter particles and are of particular interest in this analysis.

This section is devoted to constructing the angular distributions of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays. Before describing the differential decay rate, it is necessary to define the angular basis where the differential decay rate is constructed in Section 2.3.1. The $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay contains contributions from both P- and S-wave states, where the $K^+ \pi^-$ system is in a vector state with angular momentum of 1 and a scalar state with angular momentum of 0 respectively. The differential decay rate is first constructed for pure P-wave decays in Section 2.3.2. The treatment of S-wave amplitudes is then discussed in Section 2.3.2.1.

2.3.1 Definition of the angular basis

This decay has four degrees of freedom which are chosen to be the dimuon mass squared q^2 and three angles, θ_K , θ_l and ϕ conventionally. Meanwhile, information about the resonance spectrum is encoded in the invariant mass of the $K^+ \pi^-$ system $m_{K^+ \pi^-}$.

There is conflict in the conventions of the angles between theorists and experimentalists. This conflict has resulted in a lot of confusion on the signs of the terms in the differential decay rate of the B^0 and its CP-conjugated partner as can be seen in the next section. This analysis follows the experimental angular convention adopted by previous measurements [37, 40–42], while the conventions chosen by a majority of theorists are stated in [38, 43].

For the B^0 decays, the angle θ_l and θ_K are defined as

$$(2.63) \quad \begin{aligned} \cos \theta_l &= \left(\hat{p}_{\mu^+}^{(\mu^+ \mu^-)} \right) \cdot \left(\hat{p}_{\mu^+ \mu^-}^{(B^0)} \right) = \left(\hat{p}_{\mu^+}^{(\mu^+ \mu^-)} \right) \cdot \left(-\hat{p}_{B^0}^{(\mu^+ \mu^-)} \right), \\ \cos \theta_K &= \left(\hat{p}_{K^+}^{(K^{*0})} \right) \cdot \left(\hat{p}_{K^{*0}}^{(B^0)} \right) = \left(\hat{p}_{K^+}^{(K^{*0})} \right) \cdot \left(-\hat{p}_{B^0}^{(K^{*0})} \right), \end{aligned}$$

and the angle ϕ is defined as

$$(2.64) \quad \begin{aligned} \cos \phi &= \left(\hat{p}_{\mu^+}^{(B^0)} \times \hat{p}_{\mu^-}^{(B^0)} \right) \cdot \left(\hat{p}_{K^+}^{(B^0)} \times \hat{p}_{\pi^-}^{(B^0)} \right), \\ \sin \phi &= \left[\left(\hat{p}_{\mu^+}^{(B^0)} \times \hat{p}_{\mu^-}^{(B^0)} \right) \times \left(\hat{p}_{K^+}^{(B^0)} \times \hat{p}_{\pi^-}^{(B^0)} \right) \right] \cdot \hat{p}_{K^{*0}}^{(B^0)}. \end{aligned}$$

For the $\bar{B}^0$ decays, the angle θ_l and θ_K are defined as

$$(2.65) \quad \begin{aligned} \cos \theta_l &= \left(\hat{p}_{\mu^-}^{(\mu^+ \mu^-)} \right) \cdot \left(\hat{p}_{\mu^+ \mu^-}^{(\bar{B}^0)} \right) = \left(\hat{p}_{\mu^-}^{(\mu^+ \mu^-)} \right) \cdot \left(-\hat{p}_{\bar{B}^0}^{(\mu^+ \mu^-)} \right), \\ \cos \theta_K &= \left(\hat{p}_{K^-}^{(K^{*0})} \right) \cdot \left(\hat{p}_{K^{*0}}^{(\bar{B}^0)} \right) = \left(\hat{p}_{K^-}^{(K^{*0})} \right) \cdot \left(-\hat{p}_{\bar{B}^0}^{(K^{*0})} \right), \end{aligned}$$

and the angle ϕ is defined as

$$(2.66) \quad \begin{aligned} \cos \phi &= \left(\hat{p}_{\mu^-}^{(\bar{B}^0)} \times \hat{p}_{\mu^+}^{(\bar{B}^0)} \right) \cdot \left(\hat{p}_{K^-}^{(\bar{B}^0)} \times \hat{p}_{\pi^+}^{(\bar{B}^0)} \right), \\ \sin \phi &= - \left[\left(\hat{p}_{\mu^-}^{(\bar{B}^0)} \times \hat{p}_{\mu^+}^{(\bar{B}^0)} \right) \times \left(\hat{p}_{K^-}^{(\bar{B}^0)} \times \hat{p}_{\pi^+}^{(\bar{B}^0)} \right) \right] \cdot \hat{p}_{\bar{K}^{*0}}^{(\bar{B}^0)}. \end{aligned}$$

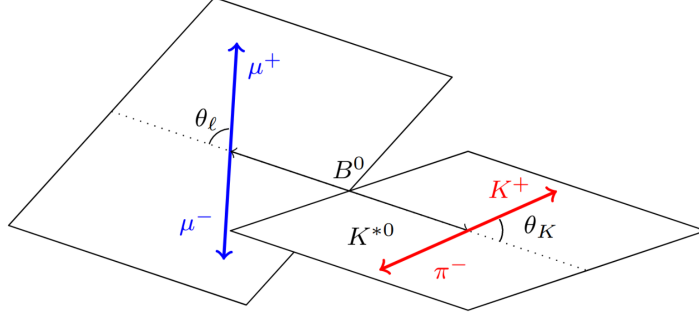
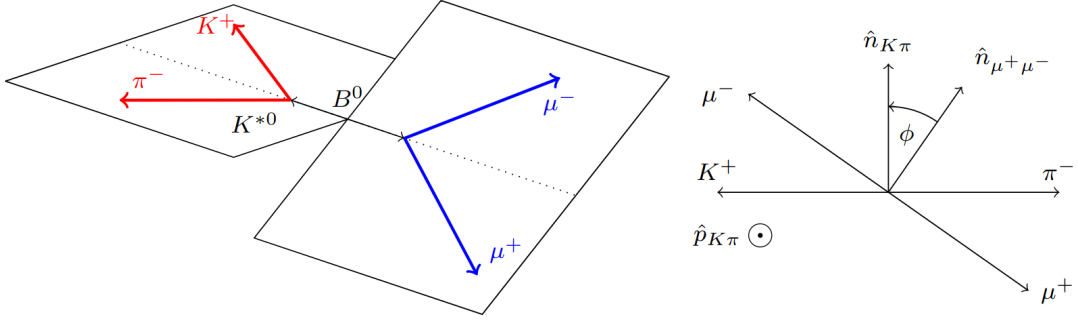
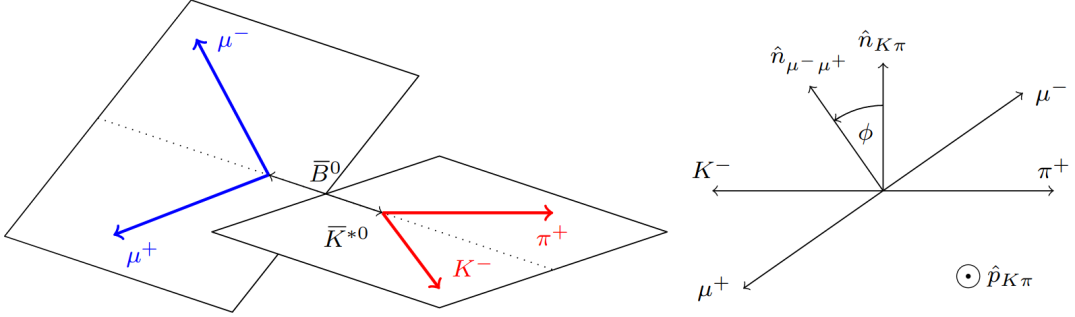

 (a) The definition of θ_l and θ_K for the B^0 decays.

 (1b) The definition of ϕ for the B^0 decays.

 (2b) The definition of ϕ for the $\bar{B}^0$ decays.

Figure 2.4: The schematic diagram showing the definition of the angular basis used for the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ and $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$ decays in this analysis. $\hat{n}_{ab}$ denotes the normal to the plane containing particles a and b in the B^0 or $\bar{B}^0$ rest frame. This diagram is taken from [44].

In all of the equations above, $\hat{p}_A^{(B)}$ stand for unit vectors on the direction of particle A in the rest frame of the B system. The schematic diagram of the angular basis for both decays is illustrated in Figure 2.4. All particle momenta are boosted to the B^0 or $\bar{B}^0$ rest frame in the first place. The angular convention of the $\bar{B}^0$ decay is thus a CP transformation of the B^0 decay with the definition of the angular basis.

2.3.2 Construction of a pure P-wave differential decay rate

The differential decay rate of the $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$ is calculated in detail in Refs. [36, 38, 45]. For a $K\pi$ system in a P-wave only configuration and ignoring scalar contributions, the differential decay rate can be written as [24]

$$\begin{aligned}
 \frac{d^4 \bar{\Gamma} [\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-]}{d \cos \theta_\ell d \cos \theta_K d\phi dq^2} = & \frac{9}{32\pi} [\mathcal{J}_{1s} \sin^2 \theta_K + \mathcal{J}_{1c} \cos^2 \theta_K + \mathcal{J}_{2s} \sin^2 \theta_K \cos 2\theta_\ell \\
 & + \mathcal{J}_{2c} \cos^2 \theta_K \cos 2\theta_\ell + \mathcal{J}_3 \sin^2 \theta_K \sin^2 \theta_\ell \cos 2\phi \\
 & + \mathcal{J}_4 \sin 2\theta_K \sin 2\theta_\ell \cos \phi + \mathcal{J}_5 \sin 2\theta_K \sin \theta_\ell \cos \phi \\
 & + \mathcal{J}_{6s} \sin^2 \theta_K \cos \theta_\ell + \mathcal{J}_7 \sin 2\theta_K \sin \theta_\ell \sin \phi \\
 & + \mathcal{J}_8 \sin 2\theta_K \sin 2\theta_\ell \sin \phi + \mathcal{J}_9 \sin^2 \theta_K \sin^2 \theta_\ell \sin 2\phi]
 \end{aligned}
 \tag{2.67}$$

Here the $\mathcal{J}_i(q^2)$ functions are q^2 -dependent angular observables composed of bilinear combinations of the six complex amplitudes $\mathcal{A}_\parallel^{L,R}$, $\mathcal{A}_\perp^{L,R}$ and $\mathcal{A}_0^{L,R}$, where the superscripts L and R denote different chiralities of the dimuon system and the subscripts $\parallel, \perp, 0$ denote the three polarisation states of the $\bar{K}^{*0}$ meson. The $\mathcal{J}_i(q^2)$ observables are written as

$$\begin{aligned}
 \mathcal{J}_{1s} = & \frac{(2 + \beta_\mu^2)}{4} \left[|\mathcal{A}_\perp^L|^2 + |\mathcal{A}_\parallel^L|^2 + (L \rightarrow R) \right] + \frac{4m_\mu^2}{q^2} \text{Re} \left(\mathcal{A}_\perp^L \mathcal{A}_\perp^{R*} + \mathcal{A}_\parallel^L \mathcal{A}_\parallel^{R*} \right) \\
 \mathcal{J}_{1c} = & |\mathcal{A}_0^L|^2 + |\mathcal{A}_0^R|^2 + \frac{4m_\mu^2}{q^2} [2 \text{Re} (\mathcal{A}_0^L \mathcal{A}_0^{R*})] \\
 \mathcal{J}_{2s} = & \frac{\beta_\mu^2}{4} \left[|\mathcal{A}_\perp^L|^2 + |\mathcal{A}_\parallel^L|^2 + (L \rightarrow R) \right] \\
 \mathcal{J}_{2c} = & -\beta_\mu^2 \left[|\mathcal{A}_0^L|^2 + (L \rightarrow R) \right] \\
 \mathcal{J}_3 = & \frac{\beta_\mu^2}{2} \left[|\mathcal{A}_\perp^L|^2 - |\mathcal{A}_\parallel^L|^2 + (L \rightarrow R) \right] \\
 \mathcal{J}_4 = & \frac{\beta_\mu^2}{\sqrt{2}} \left[\text{Re} (\mathcal{A}_0^L \mathcal{A}_\parallel^{L*}) + (L \rightarrow R) \right] \\
 \mathcal{J}_5 = & \sqrt{2} \beta_\mu \left[\text{Re} (\mathcal{A}_0^L \mathcal{A}_\perp^{L*}) - (L \rightarrow R) \right] \\
 \mathcal{J}_{6s} = & 2\beta_\mu \left[\text{Re} (\mathcal{A}_\parallel^L \mathcal{A}_\perp^{L*}) - (L \rightarrow R) \right] \\
 \mathcal{J}_7 = & \sqrt{2} \beta_\mu \left[\text{Im} (\mathcal{A}_0^L \mathcal{A}_\parallel^{L*}) - (L \rightarrow R) \right] \\
 \mathcal{J}_8 = & \frac{\beta_\mu^2}{\sqrt{2}} \left[\text{Im} (\mathcal{A}_0^L \mathcal{A}_\perp^{L*}) + (L \rightarrow R) \right] \\
 \mathcal{J}_9 = & \beta_\mu^2 \left[\text{Im} (\mathcal{A}_\parallel^{L*} \mathcal{A}_\perp^L) + (L \rightarrow R) \right]
 \end{aligned}
 \tag{2.68}$$

with $\beta_\mu^2 = (1 - 4m_\mu^2/q^2)$. The notation $(L \rightarrow R)$ refers to the same term correspondingly but with flipped chirality. In the q^2 range where this analysis is performed, the effect of massive

leptons is found not negligible, especially when q^2 is close to $1.25 \text{ GeV}^2/c^4$. Therefore, the β_μ terms in Equation 2.68 are included, but the terms with a factor m_μ^2/q^2 in the front, e.g. the last term in $\mathcal{J}_{1s}$, are still set to zero. This yields the relationship $\mathcal{J}_{1s} = 3\mathcal{J}_{2s}$ and $\mathcal{J}_{1c} = -2\mathcal{J}_{2c}$. A more complicated relationship between $\mathcal{J}_{2c}$ and the other $\mathcal{J}$ terms is explicitly written out in Ref. [46]:

$$\begin{aligned} \mathcal{J}_{2c} = & -2 \frac{(2\mathcal{J}_{2s} + \mathcal{J}_3)(4\mathcal{J}_4^2 + \beta_\mu^2 \mathcal{J}_7^2) + (2\mathcal{J}_{2s} - \mathcal{J}_3)(\beta_\mu^2 \mathcal{J}_5^2 + 4\mathcal{J}_8^2)}{16\mathcal{J}_{2s}^2 - (4\mathcal{J}_3^2 + \beta_\mu^2 \mathcal{J}_{6s}^2 + 4\mathcal{J}_9^2)} \\ & + 4 \frac{\beta_\mu^2 \mathcal{J}_{6s}(\mathcal{J}_4 \mathcal{J}_5 + \mathcal{J}_7 \mathcal{J}_8) + \mathcal{J}_9(\beta_\mu^2 \mathcal{J}_5 \mathcal{J}_7 - 4\mathcal{J}_4 \mathcal{J}_8)}{16\mathcal{J}_{2s}^2 - (4\mathcal{J}_3^2 + \beta_\mu^2 \mathcal{J}_{6s}^2 + 4\mathcal{J}_9^2)}. \end{aligned}$$

Therefore, the number of independent angular observables is reduced from eleven to nine.

The angular convention is differently defined between the theory and the LHCb collaboration, so caution needs to be taken to determine which angular convention the decay rate is defined with. The angular conventions defined in the previous section are in alignment with the LHCb experimental convention: when particles are mapped into anti-particles, corresponding to a CP-transformation, the momenta of all particles are reversed, yielding the angle change $\phi \rightarrow 2\pi - \phi$. In other words, when going from the angular distribution of $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$ to that of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$, sign flips need to be applied to the terms containing $\mathcal{J}_{7,8,9}$, as illustrated in Ref. [40].

The sum of the differential decay rates of the B^0 and $\bar{B}^0$ mesons can be written as the sum of angular observables respectively, in other words making the substitution $\mathcal{J}_i \rightarrow \mathcal{J}_i + \bar{\mathcal{J}}_i$ in Equation 2.67. The q^2 -dependent CP-average observables S_i can be defined as

$$(2.69) \quad S_i \equiv (\mathcal{J}_i + \bar{\mathcal{J}}_i) / \left(\frac{d\Gamma}{dq^2} + \frac{d\bar{\Gamma}}{dq^2} \right),$$

where $d\Gamma/dq^2$ denotes the differential decay rate of P-wave contributions given by

$$(2.70) \quad \frac{d\Gamma}{dq^2} + \frac{d\bar{\Gamma}}{dq^2} = |\mathcal{A}_\parallel^L|^2 + |\mathcal{A}_\perp^L|^2 + |\mathcal{A}_0^L|^2 + (L \rightarrow R).$$

With the relationship $\frac{3}{4}(2S_{1s} + S_{1c}) - \frac{1}{4}(\frac{2}{3}S_{1s} + \frac{1}{2}S_{1c}) = 1$ (see Ref. [47]), the number of independent S_i observables is reduced from nine to eight. Furthermore, the longitudinal polarisation of the K^{*0} meson, F_L , and the forward-backward asymmetry of the dimuon system A_{FB} are related to CP-averaged observables with the following relationship:

$$(2.71) \quad \begin{aligned} F_L &= 1 - \frac{4}{3}S_{1s}, \\ A_{FB} &= \frac{3}{4}S_{6s}. \end{aligned}$$

Therefore, one can write the differential decay rate of the $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$ decay in the basis of the CP-averaged observables S_i as [44]

$$(2.72) \quad \frac{1}{d(\Gamma + \bar{\Gamma})/dq^2} \frac{d^4(\Gamma + \bar{\Gamma})}{d \cos \theta_\ell d \cos \theta_K d\phi dq^2} = \frac{9}{32\pi} \left[\frac{3}{4} (1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \right. \\
 + \frac{1}{4} (1 - F_L) \sin^2 \theta_K \cos 2\theta_l \\
 - F_L \cos^2 \theta_K \cos 2\theta_l + S_3 \sin^2 \theta_K \sin^2 \theta_l \cos 2\phi \\
 + S_4 \sin 2\theta_K \sin 2\theta_l \cos \phi + S_5 \sin 2\theta_K \sin \theta_l \cos \phi \\
 + \frac{4}{3} A_{FB} \sin^2 \theta_K \cos \theta_l + S_7 \sin 2\theta_K \sin \theta_l \sin \phi \\
 \left. + S_8 \sin 2\theta_K \sin 2\theta_l \sin \phi + S_9 \sin^2 \theta_K \sin^2 \theta_l \sin 2\phi \right].$$

Similarly, the CP-asymmetric observables A_i can be defined as

$$A_i \equiv (\mathcal{J}_i - \bar{\mathcal{J}}_i) / \left(\frac{d\Gamma}{dq^2} + \frac{d\bar{\Gamma}}{dq^2} \right)$$

Since this analysis is performed in the CP-averaged configuration, the measurement of these CP-asymmetric observables is beyond the scope of this thesis.

Furthermore, 'optimised' observables that have a smaller theoretical uncertainty can be built. They are defined as

$$(2.73) \quad \begin{aligned} P_1 &= \frac{2S_3}{(1 - F_L)}, \\ P_2 &= \frac{2}{3} \frac{A_{FB}}{(1 - F_L)}, \\ P_3 &= \frac{-S_9}{(1 - F_L)}, \\ P'_{4,5,8} &= \frac{S_{4,5,8}}{\sqrt{F_L (1 - F_L)}}, \\ P'_6 &= \frac{S_7}{\sqrt{F_L (1 - F_L)}}. \end{aligned}$$

2.3.2.1 S-wave interference

The expression 2.67 assumes the $K^+ \pi^-$ system is in a P-wave configuration, which is the case for the K^{*0} (892) meson. However, the $K^+ \pi^-$ system can also be in an S-wave configuration where K^{*0} has a broad scalar resonance structure. The differential decay rate 2.67 thus needs

to be modified as follows, introducing $\mathcal{J}_i$ of S-wave contribution and interference terms:

$$\begin{aligned}
 (2.74) \quad & \frac{d^4\Gamma}{d\cos\theta_\ell d\cos\theta_K d\phi dq^2} \rightarrow \frac{d^4\Gamma}{d\cos\theta_\ell d\cos\theta_K d\phi dq^2} \\
 & + \frac{9}{32\pi} [\mathcal{J}'_{1c} (1 - \cos 2\theta_\ell) + \mathcal{J}''_{1c} \cos\theta_K (1 - \cos 2\theta_\ell) \\
 & + \mathcal{J}'_4 \sin 2\theta_\ell \sin\theta_K \cos\phi + \mathcal{J}'_5 \sin\theta_\ell \sin\theta_K \cos\phi \\
 & + \mathcal{J}'_7 \sin\theta_\ell \sin\theta_K \sin\phi + \mathcal{J}'_8 \sin 2\theta_\ell \sin\theta_K \sin\phi]
 \end{aligned}$$

where

$$\begin{aligned}
 (2.75) \quad & \mathcal{J}'_{1c} = \frac{1}{3} |\mathcal{A}_{00}^L|^2 + \frac{1}{3} |\mathcal{A}_{00}^R|^2 \\
 & \mathcal{J}''_{1c} = \frac{2}{\sqrt{3}} [\text{Re}(\mathcal{A}_{00}^L \mathcal{A}_0^{L*}) + (L \rightarrow R)] \\
 & \mathcal{J}'_4 = \sqrt{\frac{2}{3}} [\text{Re}(\mathcal{A}_{00}^L \mathcal{A}_{\parallel}^{L*}) + (L \rightarrow R)] \\
 & \mathcal{J}'_5 = 2\sqrt{\frac{2}{3}} [\text{Re}(\mathcal{A}_{00}^L \mathcal{A}_{\perp}^{L*}) - (L \rightarrow R)] \\
 & \mathcal{J}'_7 = 2\sqrt{\frac{2}{3}} [\text{Im}(\mathcal{A}_{00}^L \mathcal{A}_{\parallel}^{L*}) - (L \rightarrow R)] \\
 & \mathcal{J}'_8 = \sqrt{\frac{2}{3}} [\text{Im}(\mathcal{A}_{00}^L \mathcal{A}_{\perp}^{L*}) + (L \rightarrow R)],
 \end{aligned}$$

where $\mathcal{A}_{00}^{L,R}$ are S-wave amplitudes. Note that the CP-conjugated mode leads to a sign flip to terms $\mathcal{J}'_{7,8}$. An extra observable representing the fraction of the S-wave contribution can be thus defined as

$$(2.76) \quad F_S = \frac{|\mathcal{A}_{00}^L|^2 + |\mathcal{A}_{00}^R|^2}{d\Gamma/dq^2}$$

with $d\Gamma/dq^2$ denoting the total differential decay rate with both P- and S-wave contributions given by

$$\begin{aligned}
 (2.77) \quad & \frac{d\Gamma}{dq^2} = \frac{d\Gamma_P}{dq^2} + \frac{d\Gamma_S}{dq^2} \\
 & = |\mathcal{A}_{\parallel}^L|^2 + |\mathcal{A}_{\perp}^L|^2 + |\mathcal{A}_0^L|^2 + |\mathcal{A}_{00}^L|^2 + (L \rightarrow R).
 \end{aligned}$$

Note that F_S is always positive by definition.

2.3.3 Treatment of the $m_{K\pi}$ dependence

Although the angular distribution is written in terms of q^2 , $\cos\theta_\ell$, $\cos\theta_K$ and ϕ , there is actually a $m_{K\pi}$ dependence as well due to the finite width of the K^{*0} and the inclusion of P- and S-wave amplitudes and the interference terms between them. Therefore the correct four-dimensional

angular distribution should be obtained by writing out the five-dimensional angular distribution and integrating over $m_{K\pi}$.

It is assumed that $m_{K\pi}$ is factorisable with q^2 , i.e. $\mathcal{A}(m_{K\pi}, q^2) = \mathcal{A}(m_{K\pi}) \mathcal{A}(q^2)$. The $m_{K\pi}$ contribution is thus accounted for in the angular distribution by performing the transformation

$$(2.78) \quad \mathcal{A}_P \rightarrow \mathcal{A}_P \mathcal{A}_{BW}$$

to the P-wave amplitudes $\mathcal{A}_P$ and

$$(2.79) \quad \mathcal{A}_S \rightarrow \mathcal{A}_S \mathcal{A}_{LASS}$$

to the S-wave amplitudes $\mathcal{A}_S$. The $m_{K\pi}$ lineshape in the P-wave is modelled by relativistic Breit-Wigner distributions

$$(2.80) \quad \mathcal{A}_{BW}(m_{K\pi}) = \frac{\sqrt{pp'} \left(\frac{p}{p_0}\right)^{L_{B^0}} \left(\frac{p'}{p'_0}\right)^{L_{K\pi}} B'_{L_B}(p, p_0) B'_{L_{K\pi}}(p', p'_0)}{m_{K^{*0}}^2 - m_{K\pi}^2 - im_{K^{*0}} \Gamma(m_{K\pi})},$$

with p being the momentum of the $K^+ \pi^-$ system in the B^0 rest frame, p' being the momentum of the kaon in the $K^+ \pi^-$ rest frame and p_0 and p'_0 being them evaluated at the resonance peak. The orbital angular momentum between the B^0 and the dimuon system is given by $L_B = 0$ while that between the $K^+ \pi^-$ system and the dimuon system is given by $L_{K\pi} = 1$, and $m_{K^{*0}}$ denotes the pole mass of the K^{*0} . B'_L are Blatt-Weisskopf barrier modelling the production and decay of the resonance and are given by

$$(2.81) \quad \begin{aligned} B'_0(p', p'_0) &= 1 \\ B'_1(p', p'_0) &= \sqrt{\frac{1 + (p'_0 d)^2}{1 + (p' d)^2}}, \end{aligned}$$

where $d = 1.6 \text{ GeV}^{-1}$ according to Ref. [48]. The decay width $\Gamma(m_{K\pi})$ is given by

$$\Gamma(m_{K\pi}) = \Gamma(m_{K^{*0}}) \left(\frac{p'}{p'_0}\right)^{(2L_{K\pi}+1)} \frac{m_{K^{*0}}}{m_{K\pi}} (B'_{L_{K\pi}}(p', p'_0))^2.$$

The $m_{K\pi}$ lineshape in the S-wave is parameterised by the LASS model [49] written as

$$(2.82) \quad \mathcal{A}_{LASS}(m_{K\pi}) = \left[\sqrt{pp'} \left(\frac{p}{p_0}\right)^{L_{B^0}} \left(\frac{p'}{p'_0}\right)^{L_{K\pi}} B'_{L_B}(p, p_0) B'_{L_{K\pi}}(p', p'_0) \right] \times \left(\frac{m_{K\pi}}{\cot \delta_B - i} + \frac{e^{2i\delta_B}}{\cot \delta_R - i} \right),$$

where $L_{B^0} = 1$ and $L_{K\pi} = 0$. $\cot \delta_B$ describes the non-resonant component of the LASS model and is defined as

$$\cot \delta_B = \frac{1}{ap'} + \frac{rp'}{2}$$

with $r = 1.7 \text{ GeV}^{-1}$ and $a = 2.48 \text{ GeV}^{-1}$ according to Ref. [50].

To satisfy the relationship 2.77, the $m_{K\pi}$ lineshapes are normalised to the $m_{K\pi}$ window by redefining them to be

$$(2.83) \quad \begin{aligned} \mathcal{A}_{\text{BW}}(m_{K\pi}) &\rightarrow \frac{\mathcal{A}_{\text{BW}}(m_{K\pi})}{\sqrt{\int_{0.796}^{0.996} |\mathcal{A}_{\text{BW}}(m_{K\pi}, q^2)|^2 dm_{K\pi}}} \\ \mathcal{A}_{\text{LASS}}(m_{K\pi}) &\rightarrow \frac{\mathcal{A}_{\text{LASS}}(m_{K\pi})}{\sqrt{\int_{0.796}^{0.996} |\mathcal{A}_{\text{LASS}}(m_{K\pi}, q^2)|^2 dm_{K\pi}}}. \end{aligned}$$

These normalised shapes hardly vary as a function of q^2 and the integrals are approximated to be flat in q^2 . This effect, along with the choice of the parameterisation of the $m_{K\pi}$ lineshape, is studied further as a systematic effect in Section 8.5.

Therefore, when integrating over $m_{K\pi}$ in the five-dimensional angular distribution, the pure P- and S-wave $m_{K\pi}$ lineshapes integrate to 1 and only the interference integral remains, namely

$$(2.84) \quad \int_{0.796}^{0.996} \mathcal{A}_{\text{LASS}}(m_{K\pi}) \mathcal{A}_{\text{BW}}^*(m_{K\pi}) dm_{K\pi}.$$

The value of the integrals is summarised in Table 2.1.

Table 2.1: Summary of the integrals of the $m_{K\pi}$ lineshapes performed in $m_{K\pi} \in [796, 996]$ GeV/ c^2

Term	Value
$\int_{0.796}^{0.996} \mathcal{A}_{\text{BW}}(m_{K\pi}) ^2 dm_{K\pi}$	1.0
$\int_{0.796}^{0.996} \mathcal{A}_{\text{LASS}}(m_{K\pi}) ^2 dm_{K\pi}$	1.0
$\int_{0.796}^{0.996} \mathcal{A}_{\text{LASS}}(m_{K\pi}) \mathcal{A}_{\text{BW}}^*(m_{K\pi}) dm_{K\pi}$	$0.385 - 0.467i$

Therefore, the four-dimensional angular distribution in 2.74 is modified as

$$(2.85) \quad \frac{d^4\Gamma(B^0 \rightarrow K^{*0}\mu^+\mu^-)}{d\cos\theta_\ell d\cos\theta_K d\phi dq^2} = \frac{d^4\Gamma_P(B^0 \rightarrow K^{*0}\mu^+\mu^-)}{d\cos\theta_\ell d\cos\theta_K d\phi dq^2} + \frac{9}{32\pi} \left[\mathcal{J}'_{1c} f'_{1c}(\vec{\Omega}) + \mathcal{J}_i f_i(\vec{\Omega}) \int_{0.796}^{0.996} \mathcal{A}_{\text{LASS}}(m_{K\pi}) \mathcal{A}_{\text{BW}}^*(m_{K\pi}) dm_{K\pi} \right]$$

where $\mathcal{J}_i$ stand for the remaining S-wave $\mathcal{J}$ -terms shown in 2.75 apart from $\mathcal{J}'_{1c}$ and $f_i(\vec{\Omega})$ stand for the corresponding angular terms.

2.4 Helicity amplitudes of the K^{*0} meson

The amplitude ansatz method is proposed in Ref. [24]. All of the amplitudes can be determined as a parametric function of q^2 . The dependencies of form factors in the modelling of the amplitudes can be removed by this method, making the analysis theoretically model independent. This

method enables a single unbinned fit to determine all K^{*0} helicity amplitudes. The fit is unbinned without the need to average observables over q^2 bins, enabling the full exploitation of the information in the q^2 space.

2.4.1 Infinitesimal symmetries of the angular distribution

Ignoring the S-wave components, the angular distribution of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ (or $\bar{B}^0 \rightarrow \bar{K}^{*0}\mu^+\mu^-$) decay can be written in the basis of 11 $\mathcal{J}_i$ terms as shown in 2.67, or in the basis of the CP-averaged observables S_i as shown in 2.72. These expressions represent the *experimental* degrees of freedom. Since there are 3 relationships between the 11 $\mathcal{J}_i$ terms, the number of *experimental* degrees of freedom is $11 - 3 = 8$, which can be further supported by the fact that there are 8 independent S_i observables. When writing the angular distribution in the basis of the amplitudes, the *theoretical* degrees of freedom it represents must match that of the *experimental* to eliminate the degeneracy. Since there are 6 complex P-wave amplitudes in total, there are hence $6 \times 2 - 8 = 4$ continuous symmetry transformations of the amplitudes that leave each of the $\mathcal{J}_i$ term invariant, and therefore the angular distribution.

As shown in [47], the eight amplitudes can be arranged into four complex vectors

$$(2.86) \quad n_{\parallel} = \begin{pmatrix} \mathcal{A}_{\parallel}^L \\ \mathcal{A}_{\parallel}^{R*} \end{pmatrix}, \quad n_{\perp} = \begin{pmatrix} \mathcal{A}_{\perp}^L \\ -\mathcal{A}_{\perp}^{R*} \end{pmatrix}, \quad n_0 = \begin{pmatrix} \mathcal{A}_0^L \\ \mathcal{A}_0^{R*} \end{pmatrix}, \quad n_{00} = \begin{pmatrix} \mathcal{A}_{00}^L \\ \mathcal{A}_{00}^{R*} \end{pmatrix}.$$

Certain transformations can be applied to these amplitude vectors leaving the angular distribution invariant, which can be written as

$$(2.87) \quad n'_i = U n_i = \begin{pmatrix} e^{i\phi_L} & 0 \\ 0 & e^{-i\phi_R} \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cosh i\omega & -\sinh i\omega \\ -\sinh i\omega & \cosh i\omega \end{pmatrix} n_i.$$

The amplitude basis is thus not unique and needs to be defined to break the degeneracies in the angular distribution. This choice of which amplitudes to fix is model-independent and several bases have been proposed by [51] and [24]. To guarantee a simple q^2 parametrisation of the amplitude ansatz introduced in the next section, the amplitude basis chosen in this analysis is

$$(2.88) \quad \text{Im}(\mathcal{A}_{\perp}^R) = \text{Im}(\mathcal{A}_0^L) = \text{Re}(\mathcal{A}_0^R) = \text{Im}(\mathcal{A}_{00}^R) = 0.$$

As four amplitudes are fixed, the number of free amplitudes matches the degrees of freedom in the angular distribution.

2.4.2 Amplitude ansatz method

The initial ansatz to parameterise the amplitudes is

$$\mathcal{A} = \alpha + \beta q^2 + \gamma/q^2$$

where α , β and γ are coefficients to be fit for. However, it was later found that high correlations persist among the three coefficients. The ansatz thus needs to be modified. For the P-wave amplitudes, the modified ansatz is

$$(2.89) \quad \mathcal{A}_{\parallel,\perp,0}^{L,R} = \sum_i \alpha_{i,(\parallel,\perp,0)}^{L,R} L_i(q^2)$$

where L_i are Legendre polynomials of order i and α_i are corresponding Legendre coefficients that are complex. Taking the natural orthogonality of Legendre polynomials, the correlations between these Legendre coefficients can be largely reduced. These Legendre coefficients are not to be interpreted in terms of short- or long-distance effects.

On the other hand, the S-wave amplitudes are treated as nuisance parameters in the fit due to the lack of precise theory predictions and are described by

$$(2.90) \quad \mathcal{A}_{00}^{L,R} = \alpha_{00}^{L,R},$$

and are assumed to be flat in q^2 . This approximation is acceptable because the fit is performed in a four-dimensional configuration and has no $m_{K\pi}$ dependence, resulting in small statistical precision uncertainties of the S-wave amplitudes that are much larger than their variations in q^2 .

2.5 Experimental status of anomalies in $b \rightarrow s\ell^+\ell^-$ transitions

Over the past few years, $b \rightarrow s\ell^+\ell^-$ transitions have been one of the most heated topics in flavour physics. This is largely due to the anomalies of the experimental measurements seen in this channel. Common measurements can be categorised into three types with different levels of theoretical uncertainty.

2.5.1 Branching fraction

The prediction of the branching fractions is theoretically challenging. The branching fractions are related to the decay amplitudes, whose determination is dependent on the q^2 region as different assumptions are valid when q^2 is low or high. Additionally, intermediate resonances with the same final state as the primary one may start to enter with the q^2 increasing. In the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay for example, in the low q^2 region the method of QCD factorisation is utilised to obtain theoretical predictions [52]. Form factors are used to describe the transition of the B^0 meson to a K^{*0} meson, which are approximated using lattice QCD calculations [53] or computed using the principle of Light-Cone Sum Rules (LCSR) [54], but have large uncertainties due to the non-perturbative nature of QCD. In the higher q^2 region, the heavy quark effective field theory is employed [55]. In the region $8.0 < q^2 < 15\text{GeV}^2/c^4$, two peaking resonances $B^0 \rightarrow J/\psi K^{*0}$ and $B^0 \rightarrow \psi(2S)K^{*0}$ become significant, which mediates via tree-level $b \rightarrow c\bar{c}s$

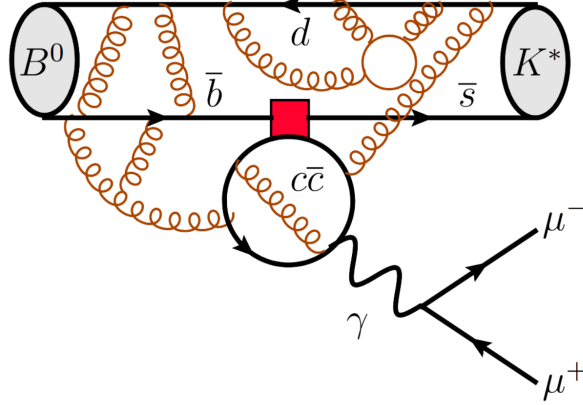


Figure 2.5: Diagram showing the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay where the interaction of a $\bar{b}$ quark and a $\bar{s}$ quark and a spectator d quark is involved. The effect of gluons results in form factor uncertainties while the charm loop exchanging gluons with the three quarks causes further non-factorisable hadronic uncertainties. This diagram is taken from [44].

transitions. For the high q^2 region, the interference between the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay and decays including charmonium resonances dominate. The charm loop exchanges gluons with the $\bar{b}$ and $\bar{s}$ quarks, bringing additional hadronic uncertainties [56, 57]. A Feynman diagram illustrating the interference of charm loop is illustrated in Figure 2.5. As can be seen in the diagram, the presence of a virtual photon suggests that charm loops may point to a vector-like coupling. This has been extensively discussed among the theory community.

Measurements of differential branching fractions of several $b \rightarrow s\ell^+\ell^-$ decays have been made by the LHCb and other collaborations and compared with SM predictions, shown in Figure 2.6. For most of the decays, the measurements in the lower q^2 bins lie below SM predictions.

2.5.2 Angular observable

Angular observables are constructed to reduce uncertainties related to form factors, as mentioned in Section 2.3. For example, in the P'_5 observable shown in 2.73, the form factors cancel to first order. However, the prediction of the observables still suffers from non-factorisable hadronic effects.

Although all angular observables are of equal importance, efforts were dedicated to measuring P'_5 by various experiments in recent years due to its smaller theory uncertainty related to form factors. The measurements are shown in Figure 2.7 and compared with theoretical predictions based on [18–20, 37]. The LHCb collaboration measurements are the most precise and present large discrepancies with SM predictions, especially in the range $4.0 < q^2 < 8.0 \text{ GeV}^2/c^4$. This hints at contributions from NP models or a lack of understanding of hadronic effects. The data analysed were recorded during the years 2011 and 2012 (Run 1) of the LHCb experiment, corresponding to integrated luminosities of 1.0 and 2.0 fb^{-1} at center-of-mass energies of 7 and 8 TeV respectively. In this analysis, the angular observables are determined with the different

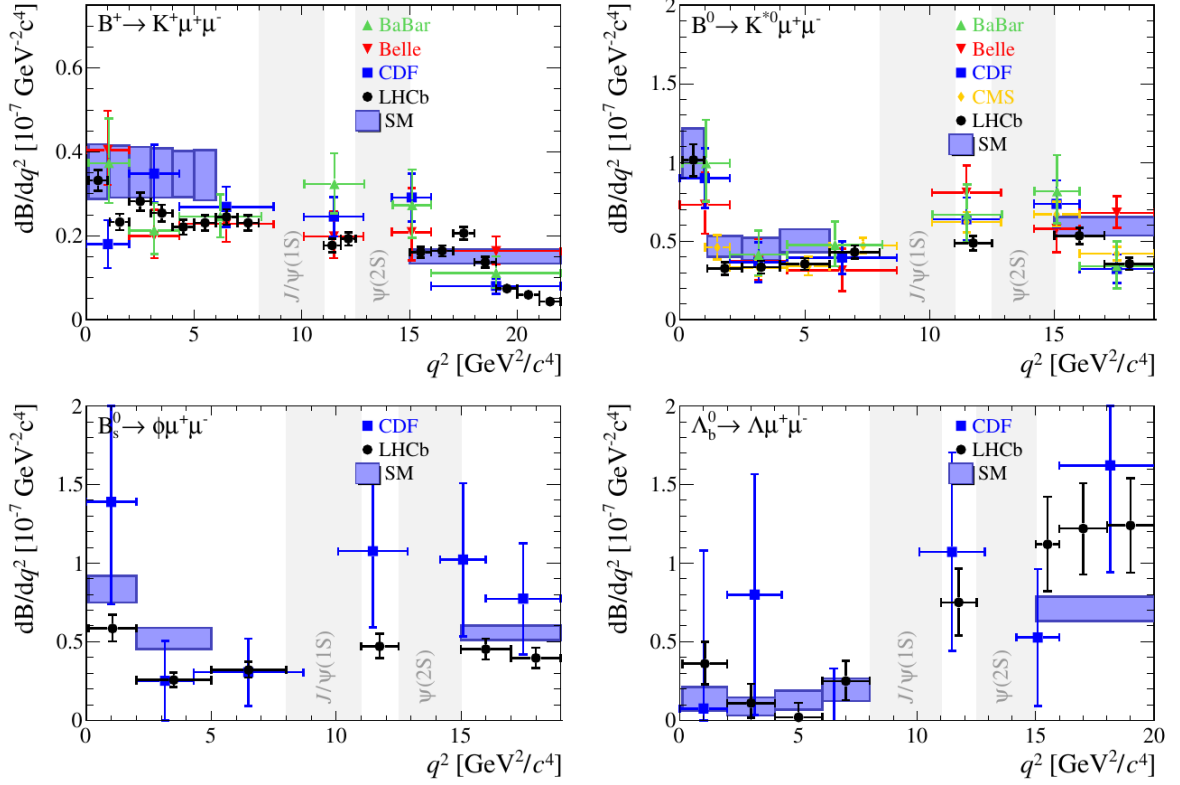


Figure 2.6: Measurements and SM predictions of the differential branching fractions of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ (top left), $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ (top right), $B_s^0 \rightarrow \phi(1020) \mu^+ \mu^-$ (bottom left) and $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ (bottom right) decays. The experimental measurements are from Babar [5], Belle [6], CDF [7], CMS [8] and LHCb [9–12] collaborations respectively. The SM predictions are from Refs. [58–61]. Figure from [62].

amplitude ansatz method, and with data taken during the Run 1 and 2 periods, with extra data corresponding to integrated luminosities of 6.0 fb^{-1} at center-of-mass energy of 13 TeV.

2.5.3 Lepton Flavour Universality ratio

In the SM, the coupling of gauge bosons to leptons is independent of lepton flavour. Therefore, the predictions of the branching fractions of e.g. $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ where the lepton pair is $e^+ e^-$ or $\mu^+ \mu^-$ should be identical apart from a factor that accounts for the difference in mass and the difference in radiative corrections between electrons and muons. This yields the concept of lepton flavour universality (LFU).

Experimentally, the LFU observables R_K and R_{K^*} are defined as [15, 16]

$$(2.91) \quad R_{(K,K^*)} \equiv \frac{\frac{\mathcal{N}}{\epsilon} (B^{(+,0)} \rightarrow K^{(+,*)} \mu^+ \mu^-)}{\frac{\mathcal{N}}{\epsilon} (B^{(+,0)} \rightarrow K^{(+,*)} J/\psi (\rightarrow \mu^+ \mu^-))} \bigg/ \frac{\frac{\mathcal{N}}{\epsilon} (B^{(+,0)} \rightarrow K^{(+,*)} e^+ e^-)}{\frac{\mathcal{N}}{\epsilon} (B^{(+,0)} \rightarrow K^{(+,*)} J/\psi (\rightarrow e^+ e^-))} ,$$

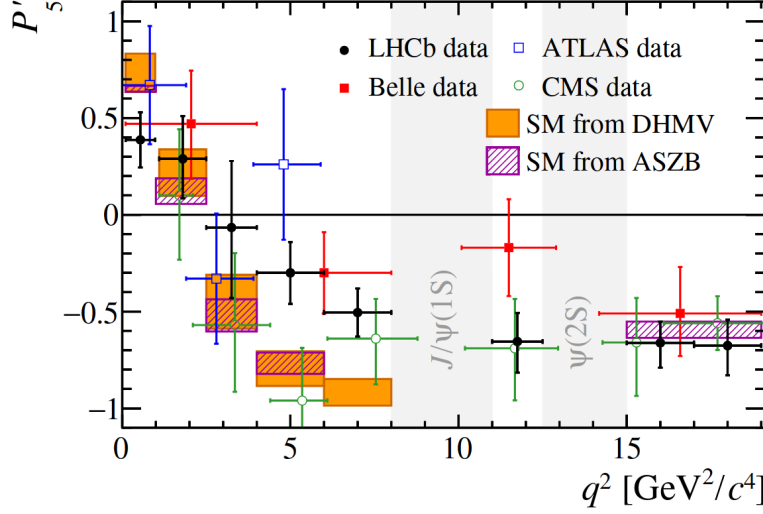


Figure 2.7: Measurements of the P'_5 observable presented by the LHCb, ATLAS, CMS and Belle collaborations. The theory predictions were made in [58] (DHMV) and [63] (ASZB). This diagram is taken from [44] .

where $\frac{N}{\epsilon}(X)$ stands for the efficiency corrected yield for process X . The theoretical predictions of these ratios are extremely clean, with an uncertainty of $\mathcal{O}(1\%)$ [64]. The LHCb collaboration determined R_K using the Run 1 data in the range $1.1 < q^2 < 6.0 \text{ GeV}^2/c^4$ [14]:

$$(2.92) \quad R_K^{[1.1,6]} = 0.846^{+0.060}_{-0.054} (\text{stat.}) \quad {}^{+0.016}_{-0.014} (\text{syst.}),$$

and also for R_{K^*} in two q^2 bins $0.045 < q^2 < 1.1 \text{ GeV}^2/c^4$ and $1.1 < q^2 < 6.0 \text{ GeV}^2/c^4$ [13]:

$$(2.93) \quad \begin{aligned} R_{K^*}^{[0.045,1.1]} &= 0.66^{+0.11}_{-0.07} (\text{stat.}) \pm 0.03 (\text{syst.}) , \\ R_{K^*}^{[1.1,6]} &= 0.69^{+0.11}_{-0.07} (\text{stat.}) \pm 0.05 (\text{syst.}) , \end{aligned}$$

where the first uncertainty is statistical and the second systematic. These results suggest tensions with the SM predictions at a level of 2.5 standard deviations (σ).

As of the time this thesis was written, the LHCb collaboration made an update using data taken during the Run 1 and 2 periods [15, 16]. The updated values are

$$(2.94) \quad \begin{aligned} R_K^{[0.1,1.1]} &= 0.994^{+0.090}_{-0.082} (\text{stat.}) \quad {}^{+0.029}_{-0.027} (\text{syst.}) , \\ R_K^{[1.1,6]} &= 0.949^{+0.042}_{-0.041} (\text{stat.}) \quad {}^{+0.022}_{-0.022} (\text{syst.}) , \\ R_{K^*}^{[0.1,1.1]} &= 0.927^{+0.093}_{-0.087} (\text{stat.}) \quad {}^{+0.036}_{-0.035} (\text{syst.}) , \\ R_{K^*}^{[1.1,6]} &= 1.027^{+0.072}_{-0.068} (\text{stat.}) \quad {}^{+0.027}_{-0.026} (\text{syst.}) . \end{aligned}$$

These measurements are in agreement with the predictions of the SM. The differences from the measurements above are understood to be partially due to tighter electron identification criteria and partially due to the modelling of the residual misidentified hadronic backgrounds [15, 16].

Another LFU ratio $R_D^{(*)}$ is constructed for the $b \rightarrow c\ell\nu$ transitions. It is defined as

$$(2.95) \quad R_{(D^{(*)})} \equiv \frac{\mathcal{B}(B \rightarrow D^{(*)}\tau^+\nu_\tau)}{\mathcal{B}(B \rightarrow D^{(*)}\mu^+\nu_\mu)}$$

Combining measurements performed by the Belle [65–68], LHCb [69, 70] and BaBar [71, 72] collaboration, the results of $R_D^{(*)}$ present a discrepancy with the SM prediction at the level of 3σ . This anomaly hints at a hierarchical NP effect with the largest contribution to the tauonic modes. In the context of $b \rightarrow s\ell^+\ell^-$ transitions, $b \rightarrow s\tau^+\tau^-$ decays such as $B^+ \rightarrow K^+\tau^+\tau^-$ could be significantly enhanced by 3 orders with respect to the SM value [73–75]. Such a value is achievable within current experimental limits and attempts of observation of this decay are ongoing in the LHCb collaboration.

There are various measurements ongoing on the LFU ratio in different decay channels in the LHCb collaboration, as well as searches for lepton flavour-violating decays. A lot of updates and new results can be foreseen in the near future.

2.6 Global fits of the Wilson coefficients

To interpret the b-anomalies, all available experimental results of $b \rightarrow s\ell^+\ell^-$ transitions are collected by multiple phenomenology groups and analysed within a model-independent Weak Effective Theory framework [76, 77]. Within such a framework, numerical Wilson coefficients can be derived by integrating the short-distance effects out, while long-distance effects are involved in local operators. Possible NP effects with a general hypothesis of being short-distance can thus be expressed as contributions to Wilson coefficients. Global fits are performed to assess scenarios that can describe all experimental measurements.

The scenario where a universal contribution to $b \rightarrow se^-e^+$ and $b \rightarrow s\mu^+\mu^-$ exists is first proposed in Ref. [78] and updated in Ref. [79] with the first measurement on LFV published [13]. Other independent works on these global fits by different groups ended up with very similar conclusions at that time, albeit choosing different theoretical inputs and statistical treatments [80–85]. Models involving a heavier neutral current Z' or a LeptoQuark were proposed, and their capability to explain these anomalies was demonstrated in Ref. [86].

With the latest results on R_K and R_{K^*} mentioned in the section above [15, 16] and an update on the measurement of the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-)$ by the CMS collaboration [87], the global fits are updated in Ref. [88]. This update also benefits from theoretical improvements in the prediction of form factors from B-meson LCSR [89, 90] and lattice-QCD computations [91, 92]. Figure 2.8 displays the hypothesis that NP contributes to $C_{9,\mu}$, $C_{10,\mu}$ and $C_{9,e}$, $C_{9,\mu}$. They present deviations with the SM predictions with a significance of 4.4σ and 5.6σ respectively. It should be noted that since the latest results on R_K and R_{K^*} support the SM, the anomalies do not only concern the muon modes. It is a golden time to study heavy flavour physics now, and obtaining more information from $b \rightarrow s\ell^+\ell^-$ transitions is more important than ever.

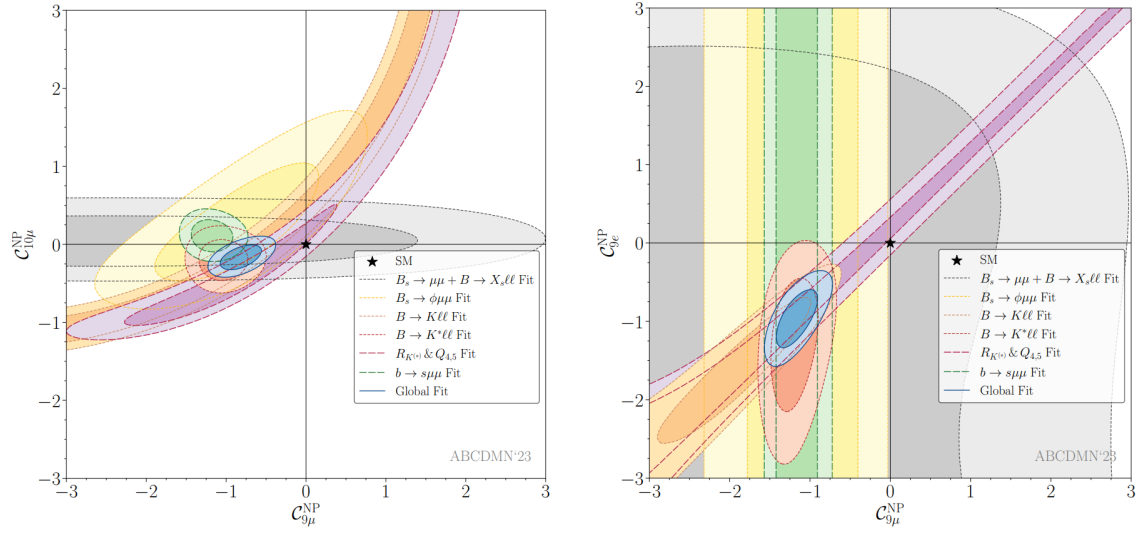


Figure 2.8: The global fit of NP contributions to C_9 , C_{10} in the muon mode (left) and to C_9 in the electron/muon mode (right). The 1 and 2 σ confidence regions are shown dark-shaded and light-shaded respectively. The experimental measurements are shown with the same principle with the colour scheme outlined by the legend. Figure from [88].

Chapter 3

LHCb Experiment at the LHC

The Large Hadron Collider (LHC) is built by the European Organisation of Nuclear Research (CERN) in Geneva, Switzerland. The LHC is a circular collider that uses the tunnel previously made for the Large Electron Positron collider (LEP) along the French-Swiss border. Measuring 26.7km in circumference, the LHC is the largest particle collider in the world. It is also the world's most powerful collider that is capable of accelerating proton beams to energies of up to 6.8 TeV. The protons are first accelerated to an energy of up to 450 GeV using a set of boosters and synchrotrons before being injected into the LHC. Further acceleration of the proton beams is achieved by 16 radio-frequency chambers along the beam line, and the beams are bent along the ring by 1232 superconducting dipole magnets that provide a magnetic field of up to 7.7 T. When the target beam energies are reached, the two proton beams collide at four different points along the accelerator ring. Four main particle detectors at the LHC are located at the four collision points: ATLAS [93] and CMS [94] that are two general purpose experiments, ALICE [95] that studies heavy ion collisions and LHCb that is dedicated to studying heavy flavour physics. An illustration of the LHC experiments is shown in Figure 3.1.

The data used in this analysis are collected in proton-proton (pp) collisions with a collision frequency of 40MHz. They are taken in two run periods of the LHC, commonly known as LHC Run 1 (2010-2012) and LHC Run 2 (2015-2018). Run 1 data collected during 2011 and 2012 are used when the LHC was operated at reduced collision centre-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV respectively, corresponding to an integrated luminosity of 1 fb^{-1} and 2 fb^{-1} respectively. Run 2 data are collected at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of about 2 fb^{-1} for each year.

3.1 Layout of the LHCb detector

The LHCb detector is primarily targeted at looking for indirect evidence of NP in CP violation or rare decays of hadrons containing b and c quarks. This physics goal largely instructs the design of the LHCb detector. Unlike the general-purpose detectors ATLAS and CMS, the LHCb

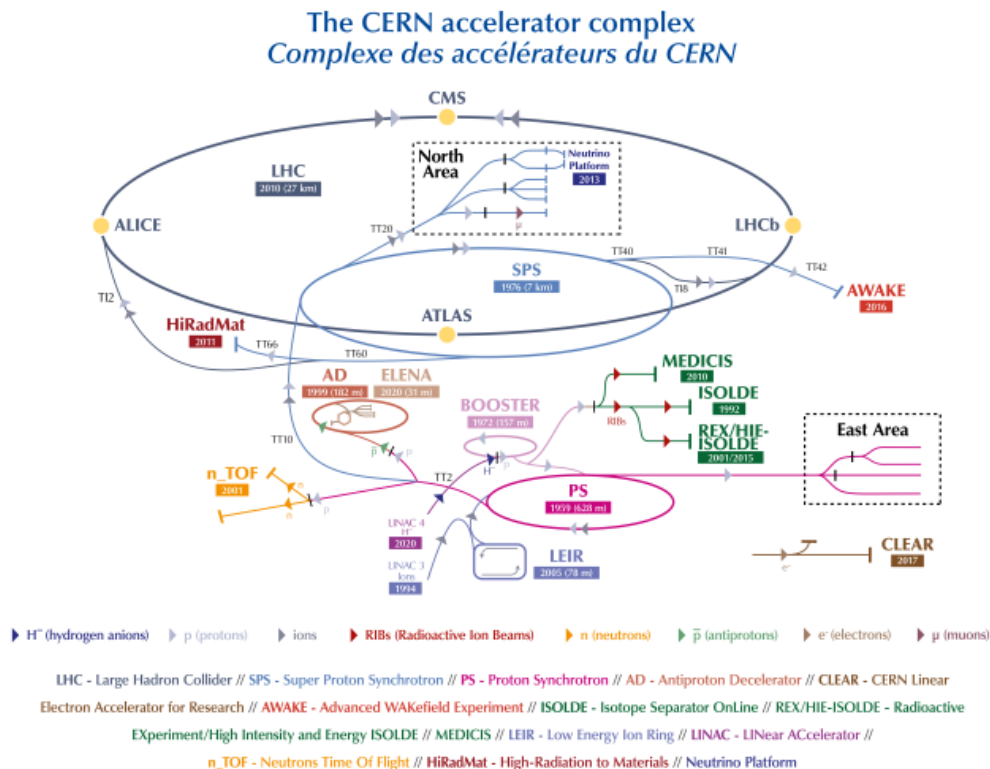


Figure 3.1: A schematic of the main experiments at the LHC. Figure from [96].

detector is a single-arm forward spectrometer that covers the azimuthal beam angle range $10 < \theta < 250$ mrad or the pseudorapidity¹ range $2 < \eta < 5$ equivalently. The LHCb detector has a forward single-arm configuration because it is built along the beam line covering positive θ . The choice of this angular acceptance is motivated by the fact that the $b\bar{b}$ production is concentrated around $\theta = 0$, as shown in Figure 3.3. The production cross section of $pp \rightarrow b\bar{b}X$ within the angular acceptance of the LHCb detector is measured to be $\sigma(pp \rightarrow b\bar{b}X) = 72.0 \pm 0.3 \pm 6.8 \mu\text{b}$ at $\sqrt{s} = 7 \text{ TeV}$ [97] and $\sigma(pp \rightarrow b\bar{b}X) = 144.3 \pm 1 \pm 21 \text{ mb}$ at $\sqrt{s} = 13 \text{ TeV}$ [98], where the first uncertainty is statistical and the second systematic. Specifically for the LHCb detector, the two colliding proton beams are slightly offset from each other. This technique is called luminosity levelling. Such a technique is employed to provide a uniform instantaneous luminosity to the LHCb detector that can be changed independently of other experiments. This is crucial for the experiment as the frequency of overlapping interactions involving multiple vertices is significantly diminished, such that tracks and vertices of short-lived particles can be well reconstructed.

The following sections give a brief overview of the sub-detectors. A cross-sectional schematic of the LHCb detector is illustrated in Figure 3.2. Starting from the left of the Cartesian origin

¹Pseudorapidity is defined as $\eta = -\ln \tan \theta/2$

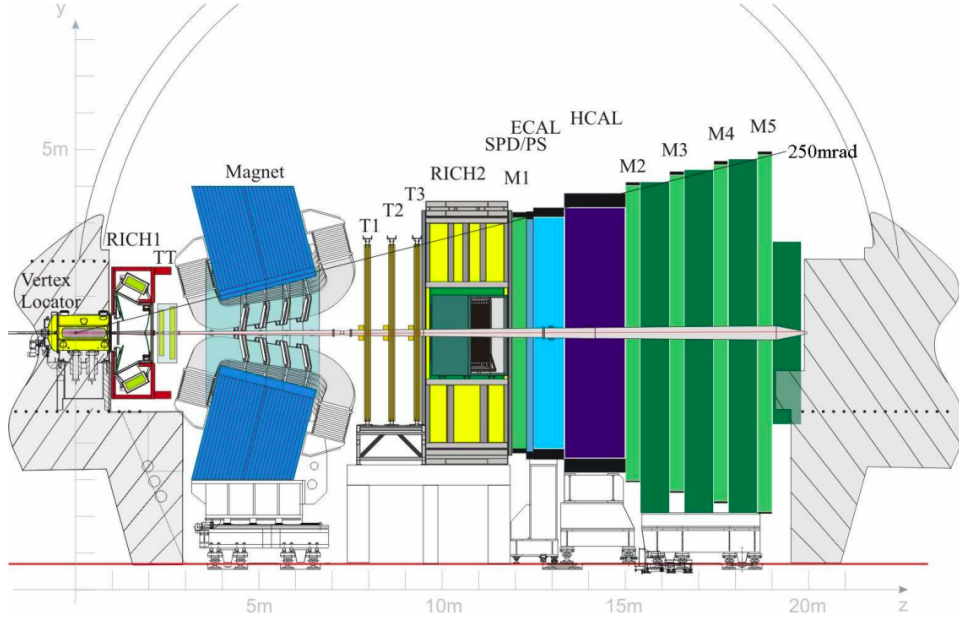


Figure 3.2: The side view of the LHCb detector. The z axis is defined along the direction of the beamline and the x - y plane is perpendicular to the z axis. The collision roughly happens at the origin point of this coordinate system. Figure taken from [99].

where the pp collision happens, the particle beams go forward to the right of the diagram. The parts close to the interaction point are also referred to as *upstream* of the dipole magnet in literature and those away from it as *downstream* of the dipole magnet. The tracking and vertexing information is provided by the VERtEX LOcator (VELO) and tracking stations as described in Section 3.2.1 and 3.2 respectively. The identification of the particle species is performed using Ring Imaging Cherenkov (RICH) detectors, calorimeters and muon systems as outlined in Section 3.4, 3.3 and 3.5 correspondingly.

Some definitions need to be made before being referred to in the following sections. For the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays with $K^{*0} \rightarrow K^+ \pi^-$ that this thesis studies, an *event* occurs when the two proton beams collide in the LHCb detector. A B hadron is produced in such an *event* and travels a few millimetres before decaying within the VELO. The charged kaons and pions produced in this decay traverse the VELO and tracking stations while leaving energy deposits in the detector pads that record the *hits*. A *track* is reconstructed by associating these *hits* with particle trajectories in each *event*. The *tracks* are further associated with vertices and a Particle Identification (PID) or mass hypothesis is assigned to them using information from the RICH detectors, along with the information from muon detectors on the charged daughter muons of the decay.

A $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ *candidate* with $K^{*0} \rightarrow K^+ \pi^-$ is constructed by combining a K^+ , a π^- and two oppositely charged muon tracks that originate from the same vertex. This could be from

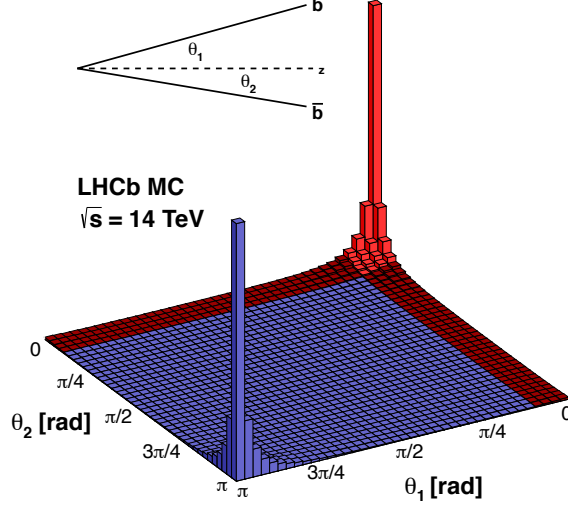


Figure 3.3: The simulated $b\bar{b}$ production as a function of the opening angle, $\theta_1(\theta_2)$ of the b ($\bar{b}$) quark, defined as shown in the inset. The red-coloured region shows the acceptance of the LHCb detector [100].

a genuine $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay with $K^{*0} \rightarrow K^+ \pi^-$ or a background process. An interesting *event* is initially selected based on some simple tracking and vertex reconstruction algorithm. This process is also known as triggering and is further described in Section 3.6. Finally, the generation of simulated events is discussed in Section 3.7.

3.2 Tracking system and magnet

The tracking system of charged particles is constituted by the VELO, the Tracker Turicensis (TT) upstream of the magnet and tracking stations T1-T3 downstream of the magnet (see Figure 3.2). The VELO is crucial for giving detailed information of the b hadron decays and is therefore described in more detail in Section 3.2.1, whereas Section 3.2.2 discusses the other tracking stations and their combined performance.

3.2.1 VELO

Hadrons containing b and c quarks have a long lifetime, so they fly for a few millimetres before decaying at LHC collision energies. In the case of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays, the B^0 meson originates from pp collisions at the Primary Vertex (PV), flies a short distance and eventually decays into its daughter particles at the Secondary Vertex (SV). Figure 3.4 illustrates this process schematically. The main purpose of the VELO detector is to reconstruct VELO tracks and separate PV from SV to build long tracks.

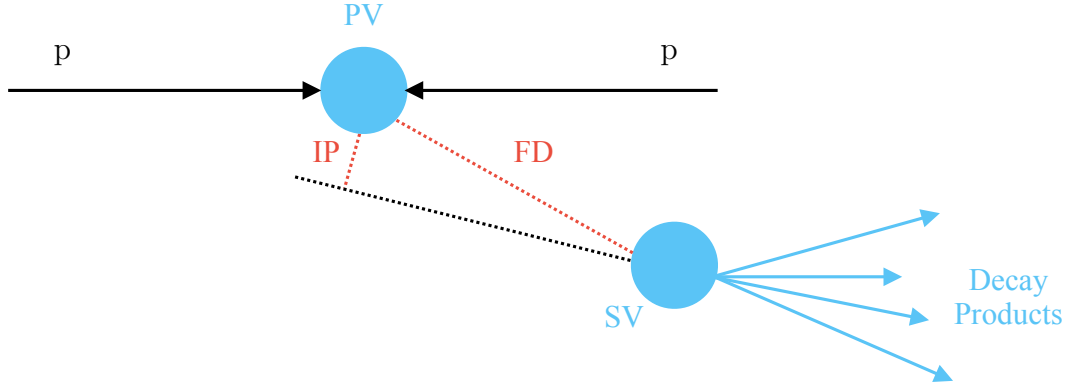


Figure 3.4: Schematic of a b-hadron decay topology within the VELO. A b-hadron is produced in the pp collision at the Primary Vertex (PV) and travels for a length of Flight Distance (FD) before decaying into the daughter particles at the Secondary Vertex (SV). The Impact Parameter (IP) is defined as the line perpendicular to the direction vector of the decaying particle and the PV.

The VELO comprises two half-doughnut-shaped parts on either side of the beam, each containing 21 silicon modules. One of the parts has the silicon strips lined circularly around the module to measure the axial distance of the hits R , and the other has strips lined in the radial direction to provide information on the polar angle of hits ϕ . The structure of the $R\phi$ sensor and the VELO is illustrated in Figure 3.5. When the proton beams are stable, the two VELO modules are positioned 8 mm from the beam. Whenever the beams are unstable, such as during the ramp-up of the beam energy, both VELO parts are retracted by 3 cm to protect them from damage by stray beam particles.

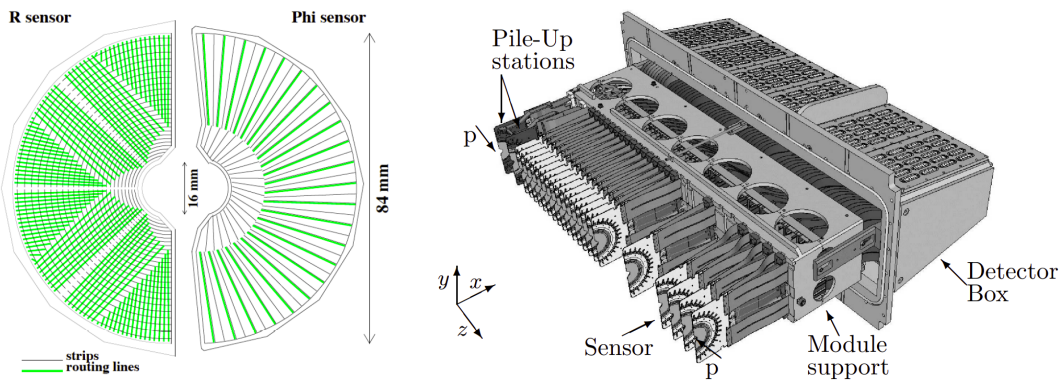


Figure 3.5: (Left) The schematic representation of a $R\phi$ sensor [101]. (Right) Overview of the VELO left half [102].

The cylindrical geometry of the $R\phi$ sensor is motivated by the need for fast reconstruction of the Impact Parameter (IP). The IP is defined as the closest distance from a track to the PV (see Figure 3.4). Tracks from decays containing b- and c-hadrons are generally displaced

from the PV and are hence characterised by tracks with large IP with respect to the PV. The resolution of the IP is dependent on the transverse momentum p_T , the component of momentum transverse to the beam line, as illustrated in Figure 3.6. For tracks with high p_T , a better IP resolution is obtained at $15\ \mu\text{m}$.

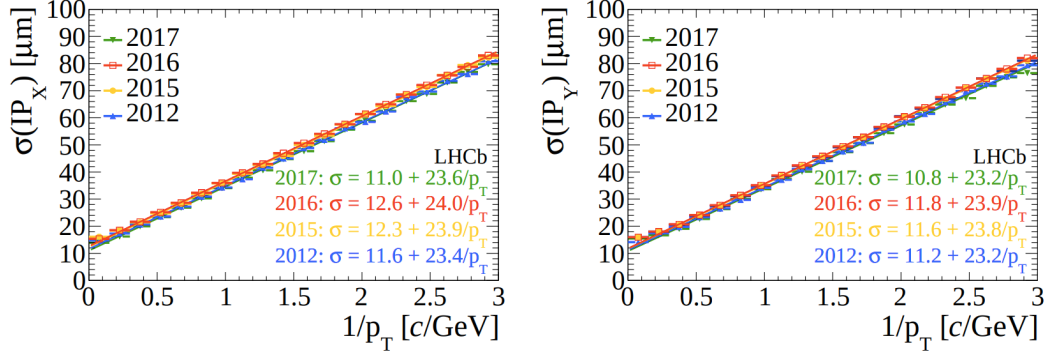


Figure 3.6: Resolution of the x (left) and y (right) components of the IP comparing four different data-taking periods. The resolution is given in the bottom right corner as a function of p_T . Figure from [103].

The hit resolution depends on the width of the strips (pitches as referred to) and the charge shared between strips, which is related to the position of strips in the VELO. The pitch varies from around $40\ \mu\text{m}$ close to the beam to around $100\ \mu\text{m}$ away from the beam, obtaining a hit resolution between 4 to $30\ \mu\text{m}$. The PV resolution is determined by N , the number of tracks used to construct the PV. The PV resolution follows the relation of $\sigma_{xV} = -1 + 97.7/N^{0.61}$, $\sigma_{yV} = -1.5 + 92.7/N^{0.59}$ and $\sigma_{zV} = -16 + 923/N^{0.69}$ along the x , y and z coordinates of the vertex respectively, as demonstrated in Figure 3.7. With the minimum number of required tracks to be 5, the resolution on the x and y directions of a vertex is around $35\ \mu\text{m}$ and the resolution on the z direction is around $300\ \mu\text{m}$.

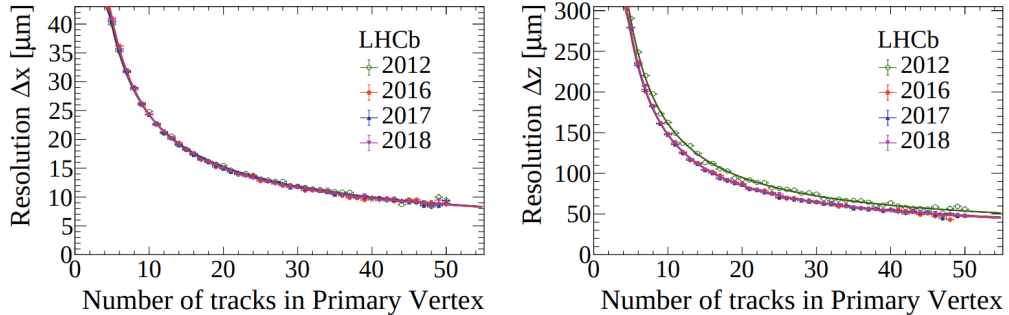


Figure 3.7: The PV x (left) and z (right) resolution as a function of the number of tracks in the PV for the Run 1 and Run 2 PV reconstruction algorithms. Figure from [103].

3.2.2 Tracking stations and magnets

Several more tracking detectors are installed in the LHCb detector further downstream the beamline from the VELO, TT upstream of the magnet and the T1-T3 downstream of the magnet (see Figure 3.2). Their primary goal is to measure the momentum of charged particles from the curvature of the tracks in a known magnetic field provided by the magnet. The magnet has a bending power of 4 Tm. A schematic of the different tracking stations is shown in Figure 3.8.

The TT is made of four layers of silicon strip detectors. The single hit resolution of these strips ranges between 50 to 55 μm with an excellent hit efficiency of above 99% [104]. The T1-T3 track detectors also have four layers each. Each layer is split into two detector technologies, the Inner Tracker (IT) and the Outer Tracker (OT). The IT is made of high-resolution silicon microstrip due to the higher occupancy of particles transversing through it. The OT is away from the beam and thus consists of straw tube detectors filled with gas, which are cheaper but have a lower resolution. The relative momentum resolution, $\delta p/p$, ranges from $\delta p/p = 0.4\%$ at $p = 5 \text{ GeV}/c$ to $\delta p/p = 0.6\%$ at $p = 100 \text{ GeV}/c$.

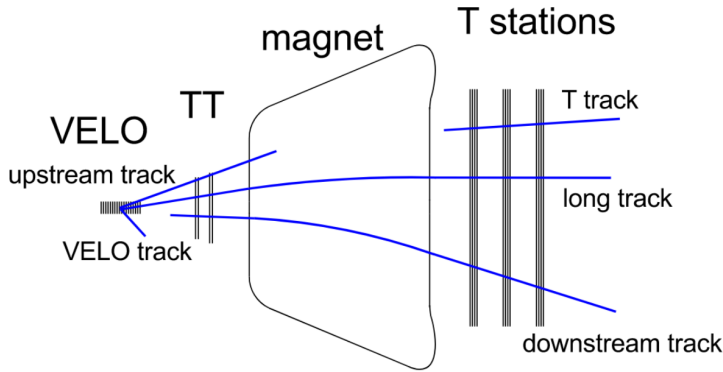


Figure 3.8: Schematic illustrating the modules of the tracking system and the different track types reconstructed by different algorithms. Figure taken from [105].

A Kalman filter algorithm [106, 107] is used to fit tracks to hits in the trackers including the VELO. As the input of the Kalman filter algorithm naturally contains measurement inaccuracies and statistical noise, the algorithm may find nonphysical *ghost tracks*, which mainly consist of fake hits or hits that were produced by other particles. Such *ghost tracks* have larger χ^2 values in general and leave fewer hits, so they can be strongly suppressed. One can also estimate the ghost probability, $\text{prob}_{\text{ghost}}$, by combining information from the track fit [108, 109].

3.3 Calorimeters

The calorimeter system at the LHCb experiment allows photons, electrons and hadrons to be identified and their energy measured. The system includes four sub-detectors: the Scintillating Pad Detector (SPD), Preshower Detector (PD), Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL). The energy of the particles passing the sub-detectors is absorbed and re-emitted as light, and the light is led to the Photomultiplier Tubes (PMT) through wavelength shifter fibres (WLS). A schematic illustrates this process in Figure 3.9. The SPD, PD and ECAL contribute to the trigger and offline identification of electrons, photons and neutral pions. The HCAL is primarily used in the trigger systems by measuring the transverse energy of the hadronic shower and to improve the distinction between electrons and hadrons.

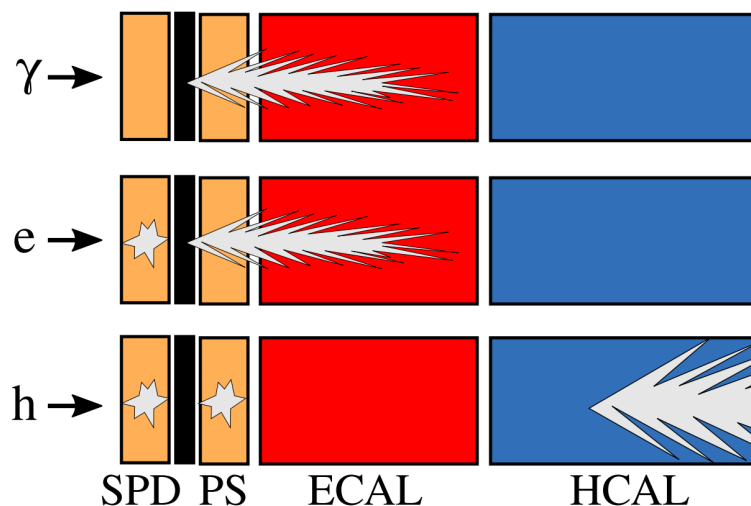


Figure 3.9: Illustration of the energy deposit in the calorimeter system of a photon (γ), an electron (e) and a charged hadron (h). Taken from [110].

The SPD and PS are made out of two plastic scintillator layers with a 15 mm thick lead plate separating them. Electrons and photons are distinguished by the SPD which detects charged particles only. These particles traverse the SPD and radiate in the lead plate. When photons meet the lead plate, an electromagnetic shower is initiated before converting to electron-positron pairs. PS detects these pairs and utilises this information to differentiate between photons and hadrons. Both the ECAL and HCAL have modules made of several layers of scintillating tiles separated by layers of absorbing material. In ECAL the absorber is made out of lead while in HCAL it is made out of iron. WLS fibres are used again to transport the scintillating light to the PMT [102].

The calorimeter system plays a minor role in the angular analysis of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays. However, the information it provides is used in PID which is discussed in the following two sections.

3.4 Particle identification with Cherenkov detectors

The correct identification of particle species plays a vital role in the reconstruction of specific decays. For the reconstruction of $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)\mu^+\mu^-$ candidates, the correct identification of the kaons and pions is crucial. The PID is primarily carried out by the Ring Imaging Cherenkov (RICH) system including RICH1 upstream of the magnet and RICH2 downstream of the magnet just after the tracking stations T1-T3 (see Figure 3.2).

The RICH detectors are based on the Cherenkov radiation phenomenon [111]. When a charged particle traverses through a medium faster than the speed of light in that medium, it emits a light cone through Cherenkov radiation. This light cone is collected by the RICH detectors using mirrors and photo-detectors, and an angle of the light cone with respect to the particle's trajectory, θ_c , can be determined. One can thus determine the mass m of the particle through the relationship between θ_c , m , the momentum of the particle p and the refractive index n of the medium:

$$(3.1) \quad \cos(\theta_c) = \frac{\sqrt{m^2c^2 + p^2}}{pn} .$$

Different gases with different refractive indices are filled in RICH1 and RICH2 as the medium, providing different momentum ranges in the two sub-detectors. RICH1 is filled with C_4F_{10} with $n = 1.0018$, providing PID of kaons and pions with momenta in the range of $2 < p < 40$ GeV within the full LHCb angular acceptance of $25 - 300$ mrad. RICH2 contains CF_4 with $n = 1.0005$ and provides PID for kaons and pions with momenta in the range of $15 < p < 100$ GeV, focusing on the region closer to the beamline of $15 - 120$ mrad. The Cherenkov angle with respect to the track momentum is shown in Figure 3.10. The choice of two RICH detectors with a different angular coverage is driven by the strong correlation between momentum and polar angle as higher momentum particles are produced at smaller polar angles.

Combining the momentum information from the tracking system and the Cherenkov angles reconstructed in the RICH, each track is associated with a mass hypothesis by fitting Equation 3.1 with different mass hypotheses to each track. This allows for the DLL variable, defined as the difference in Log-Likelihoods, to quantify the difference between two mass hypotheses of a certain track. For example, the DLL between the kaon and pion hypothesis is defined as:

$$(3.2) \quad \text{DLL}_{K\pi} = \log \mathcal{L}(\theta_c, p \mid K) - \log \mathcal{L}(\theta_c, p \mid \pi) .$$

The DLL variables play a crucial role in the selection of signal candidates as outlined in Chapter 4.

The kaon identification efficiencies, along with the rate of a pion misidentified as a kaon, are plotted versus the particle momentum in Figure 3.11, in which two cuts on the $\text{DLL}_{K\pi}$ are compared. In general, a stricter cut of $\text{DLL}_{K\pi} > 5$ keeps the pion misidentification rate below 5% in the full range of momenta but causes the kaon identification to drop significantly as the

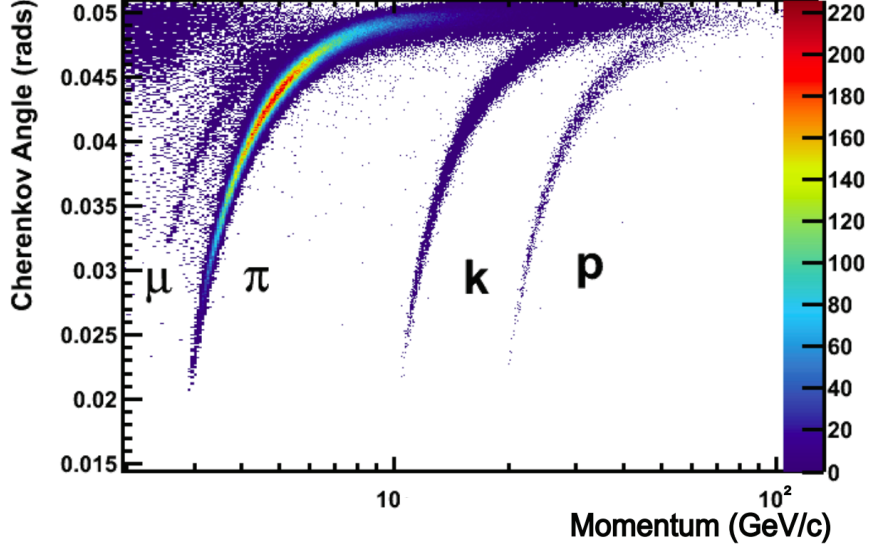


Figure 3.10: Reconstructed Cherenkov angle as a function of track momentum in the C_4F_{10} radiator. Figure taken from [112].

momenta increase. On the other hand, a looser cut of $DLL_{K\pi}$ keeps the kaon identification efficiency above 80% for $p < 80$ MeV/c but results in the pion misidentification rate reaching above 20% at high momenta ($p > 80$ MeV/c).

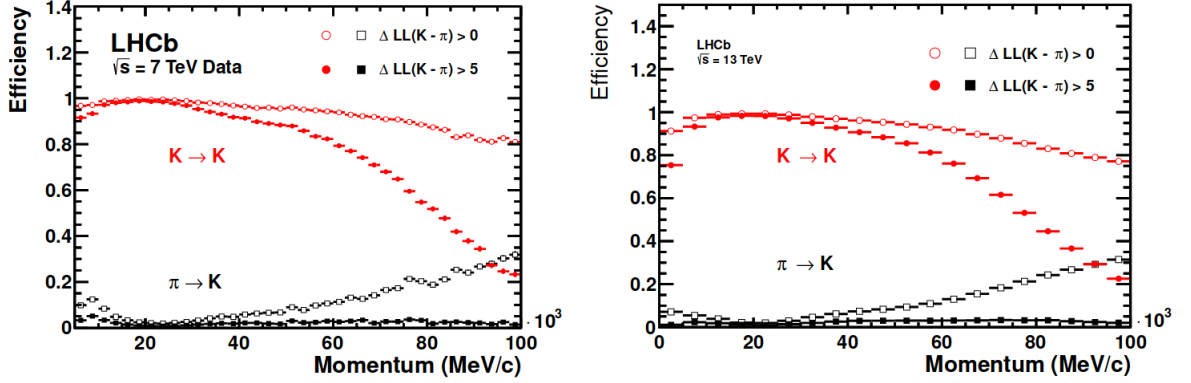


Figure 3.11: The kaon identification efficiency and pion misidentification rate evaluated with calibration samples for Run 1 (left) and Run 2 (right) under different cuts of the $DLL_{K\pi}$ variable. The $\Delta LL(K - \pi)$ is the synonym of the $DLL_{K\pi}$ introduced in this thesis. Figure taken from [112, 113].

In addition to the likelihood method, a Neural Network (NN) based classifier is trained and the classification probabilities are stored in the variables ProbNNx (where $x = K, \pi, \mu$). A full list of these ProbNNx variables is presented in Reference [114]. The kinematic variables from the tracking system and the likelihood ratios are used to train the neural networks, along with the information from the PID system that is relevant to the calculation of the likelihood ratio, such as the geometrical acceptance of the RICH radiators. The performance of the ProbNNK

variable evaluated on the 2018 data is shown as an example in Figure 3.12, and the efficiency is similar for other data-taking periods.

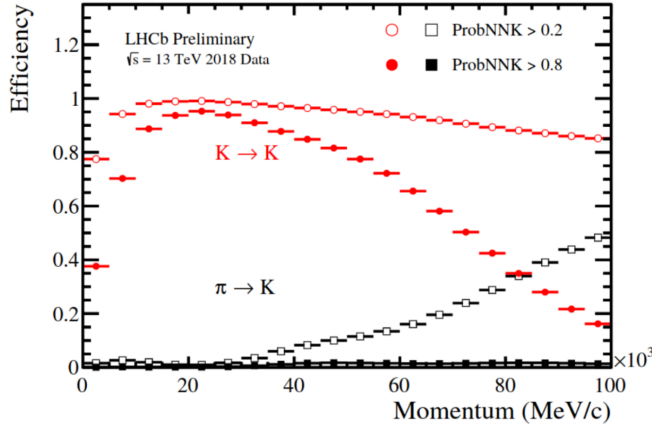


Figure 3.12: The kaon identification efficiency and pion misidentification rate evaluated with calibration samples for Run 2 under different cuts of the **ProbNNK** variable. Figure from [115].

As can be seen from the left bands of Figure 3.10, the separation power between muons and pions of the RICH detectors is fairly poor. Therefore, the muon system is needed to further identify muons and is outlined in the next section.

3.5 Muon stations

The correct identification of muons is of fundamental importance for the $B^0 \rightarrow K^{*0}(\rightarrow K^+\pi^-)\mu^+\mu^-$ decay. Experimentally, the muons penetrate material much better than most other particles and have a lifetime long enough to traverse the full length of the detector. Taking advantage of this property, the muon stations are placed at the end of the LHCb detector.

The muon system consists of five rectangular stations, with M1 placed downstream of RICH2 in front of the calorimeters while M2-M5 is located downstream of the calorimeters, as shown in the left plot of Figure 3.13. Each of the stations is segmented into four regions, marked as R1-R4 on the right hand side of Figure 3.13. For the R1 region of the M1 station, 12 triple gas electron multipliers (triple-GEM) detectors are installed whereas the R2-R4 modules are made up of 264 multi-wire proportional chambers (MWPC) [116]. The M2-M5 stations are separated by 80 cm thick iron shields and each station consists of 276 MWPCs in total. The size of the MWPCs is increased as a function of the distance to ensure consistent particle flux and channel occupancy across the four regions, and the triple-GEM replaces the MWPCs in the R1 region of the M1 station to account for the high risk of radiation damage.

The main purpose of M1 is to give an improved measurement of the transverse momentum p_T of the muons for triggering. The rest of the muon stations, along with M1, determine the PID information for muons. There are two important PID variables that make use of the muon

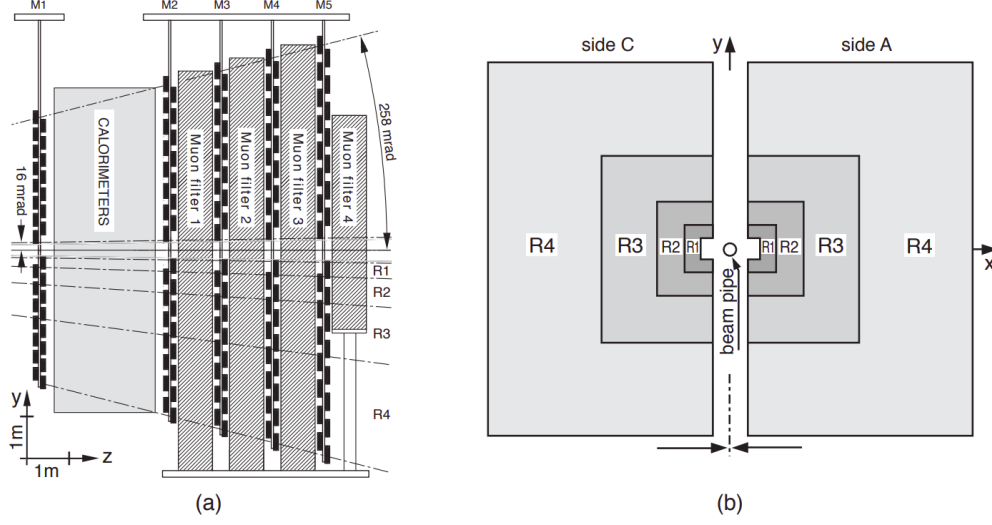


Figure 3.13: Schematic of the muon stations as seen from the side (left) and along the beamline (right). Figure taken from [117].

stations. The first is a binary value called `isMuon`, which can be simply determined by requiring hits in the muon stations taking advantage of the stopping power against all other particles by the calorimeters and iron shields. The conditions for `ismuon` is summarised in Table 3.1. The `ismuon` variable has an excellent identification efficiency for muons of about 95% for tracks with $0.8 \text{ GeV}/c < p_T < 1.7 \text{ GeV}/c$ and 100% for tracks with $p_T > 5 \text{ GeV}/c$, while keeping the identification probability for protons, kaons and pions less than 2%.

Table 3.1: Conditions for the `isMuon` variable. Taken from [118].

momentum range	Required hits in muon stations
$3 \text{ GeV}/c < p < 6 \text{ GeV}/c$	M2 and M3
$6 \text{ GeV}/c < p < 10 \text{ GeV}/c$	M2, M3 and either M4 or M5
$p > 10 \text{ GeV}/c$	M2, M3, M4 and M5

The second variable is a DLL variable defined similarly to $\text{DLL}_{K\pi}$ in Equation 3.2:

$$(3.3) \quad \text{DLL}_{\mu\pi} = \log \left(\mathcal{L}(\theta_c, p, D^2 | \mu) \right) - \log \left(\mathcal{L}(\theta_c, p, D^2 | \pi) \right) ,$$

where θ_c denotes the Cherenkov angle obtained from the RICH, p is the particle momentum and D^2 represents the average squared distance significance of the hits in the muon stations. The detailed definition of D^2 can be found in Ref. [118].

3.6 Trigger system

At the LHC the pp-collision happens at a rate of 40 MHz during normal operations. This results in a huge amount of visible events that cannot be processed or stored at the same time. Only

1% of the visible events are expected to be $b\bar{b}$ events, and only a fraction of these events is expected to produce b (c)-hadrons. A trigger system is thus designed to determine when to read out the detector and efficiently select the interesting events to be written to long-term storage.

The trigger system is split into two stages: the low level trigger Level-0 (L0) trigger implemented in specialised hardware, and the software-based high level trigger referred to as the High Level Trigger (HLT). Both stages are explained in the following sections.

3.6.1 Low level trigger

The L0 trigger is implemented using Field Programmable Gate Arrays (FPGA). It makes use of information obtained from the calorimeters and muon stations to decide when to read out the detector, reducing the data rate from 40 MHz to 1 MHz within 4 μ s. The L0 calorimeter trigger and the L0 muon trigger work independently. Since events triggered by the L0 calorimeter trigger are not used in this analysis, this section focuses more on the L0 muon trigger.

The L0 muon trigger accepts events with two thresholds, denoted as the L0Muon trigger line for a single muon with high transverse momentum p_T and the L0DiMuon trigger line for a pair of high- p_T muons. This is driven by the fact that heavy-flavour hadrons can be characterized by particles with high transverse momenta. The p_T of a track is estimated from the spatial coordinates detected by the five muon stations, and two tracks with the largest p_T are taken to make the decision. The L0Muon trigger requires the larger p_T of the two to be above a certain threshold while the L0DiMuon trigger has requirements on the product of p_T of the two tracks. These requirements for the two trigger lines are different for individual data-taking periods and are summarised in Table 3.2.

Table 3.2: Thresholds of the L0Muon and L0DiMuon in each year [44].

	L0Muon threshold	L0DiMuon threshold
	p_T^{largest}	$p_T^{\text{largest}} \times p_T^{2^{\text{nd}}\text{largest}}$
2011 [119]	$> 1.48 \text{ GeV}/c$	$> (1.296 \text{ GeV}/c)^2$
2012 [119]	$> 1.76 \text{ GeV}/c$	$> (1.6 \text{ GeV}/c)^2$
2016 [103]	$> 1.8 \text{ GeV}/c$	$> 2.25 \text{ GeV}^2/c^2$
2017 [103]	$> 1.35 \text{ GeV}/c$	$> 1.69 \text{ GeV}^2/c^2$

The L0 trigger further imposes requirements on the maximum number of hits in the SPD, denoted as SPD multiplicity, to reduce the complexity of the events and thus speed up the decision-making. The maximum SPD multiplicity allowed for L0Muon is 450. For L0DiMuon, as it accepts events at a much lower rate, the upper threshold of the SPD multiplicity is chosen to be 900. These SPD requirements do not cause a significant signal efficiency loss [103].

3.6.2 High level trigger

The events passing the L0 trigger stage are further processed by the software-based HLT, which reduces the output rate from 1 MHz to 5 kHz in Run 1 and 12 kHz in Run 2. The HLT can be split into two successive algorithms, High Level Trigger 1 (HLT1) and High Level Trigger 2 (HLT2). A particular sequence of triggering algorithms and selections is summarised in the so-called trigger lines. A trigger line provides a boolean value for accepting or rejecting the event. Events passing any of the HLT1 or HLT2 trigger lines are stored for offline analysis.

High Level Trigger 1 (HLT1)

In HLT1, a fast reconstruction of tracks in events is performed making use of the information provided by the full tracking system. The first step involves a reconstruction of the PV using the tracks in the VELO, followed by two algorithms to determine whether an event should be examined by the final requirements of the HLT1 trigger.

The first algorithm focuses on selecting muon candidates. For each track in the VELO, it is examined whether it leaves a hit in a defined search window of the muon station M3. This search window is based on the bending angle of a track assuming it corresponds to a muon and has a momentum greater than 6 GeV/c, which is the minimum required momentum of a muon candidate. The hits within this search window are then recalculated combining the information from the VELO to from candidate tracks. Additional search windows in the M2, M4 and M5 stations are then opened and whether these candidate tracks leave hits in the search windows are examined. If at least one of the stations contains a hit, the tracks are provisionally tagged as muon candidates. The quality of the muon candidate tracks is assessed by performing linear χ^2 fits until the χ^2 value from the fits divided by the number of degrees of freedom $\chi^2/\text{ndf} < 25$ is fulfilled, at which point the algorithm stops. This track is passed onto the next stage of reconstruction. The second algorithm selects tracks with a sufficiently large IP with respect to any PV. Tracks selected by the second algorithm are not necessarily muon tracks and are passed onto the next stage as well.

The next stage of reconstruction in HLT1 is called forward tracking [120]. The search windows in the T1-T3 stations are determined by tracks passing either or both of the algorithms above. Imposing thresholds on minimum momentum in forward tracking helps reduce the size of these search windows and hence the processing time required for the reconstruction. A Kalman filter based algorithm [106, 107] is then utilised to fit the reconstructed tracks to obtain their χ^2 and covariance matrix.

In this analysis, the HLT1 trigger lines used to select Run 1 $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ candidates are called `Hlt1TrackAllL0` and `Hlt1TrackMuon` triggers. The first trigger selects all events with a track candidate with a $p_T > 1.6$ GeV/c. The second trigger line filters out all events with a track candidate with both a true `isMuon` value and a $p_T > 1$ GeV/c. For the Run 2 period,

the `Hlt1TrackAllL0` trigger is replaced by the multivariate-analysis (MVA) based trigger lines `Hlt1TrackMVA` and `Hlt1TwoTrackMVA` to increase the selection efficiencies.

High Level Trigger 2 (HLT2)

The full event reconstruction is performed in HLT2 for all tracks with $p_T > 500 \text{ MeV}/c$ ($p_T > 300 \text{ MeV}/c$ for the Run 1 period) using the information from all sub-detectors. There are two types of trigger lines implemented in HLT2, the inclusive line and the exclusive line. Inclusive lines, also referred to as topological lines, select decays with a certain topology, whereas exclusive lines are dedicated to selected specific decays. For this analysis, six inclusive trigger lines and two exclusive muon trigger lines are used to select potential $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ candidates.

The inclusive lines select candidates using MVA techniques to exploit special tracks called Topo-Track. A Topo-Track is an HLT2 track with extra requirements on track fit quality (χ^2/ndf), IP and muon and electron identification. Topo-Tracks are combined to form a 2-, 3- or 4-body decay object with the upper limit of their Distance Of Closest Approach (DOCA) being 0.2 mm. Such a decay object containing at least one muon is triggered by `Hlt2TopoMu[2,3,4]Body` while that not necessarily containing a muon is triggered by `Hlt2Topo[2,3,4]Body`. Additionally, all six trigger lines ensure that the tracks come from a displaced B-hadron decay.

For the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay in the presented analysis, two exclusive muon algorithms are employed, which are the `Hlt2SingleMuon` and the `Hlt2DiMuonDetached` trigger lines. Events meeting either or both of the following requirements pass the `Hlt2SingleMuon` trigger line: an event has a muon candidate with a $p_T > 10 \text{ GeV}/c$ and the muon track in an event has a $p_T > 1.3 \text{ GeV}/c$, a very low χ^2/ndf and a significant IP. On the other hand, the `Hlt2DiMuonDetached` trigger line is based on two identified muons in the final state. The two muon candidate tracks should originate from the same PV and have a significant IP and a low χ^2/ndf . There are also lower limits on the dimuon invariant mass of the tracks being $1 \text{ GeV}/c$ and on the dimuon p_T being $1.5 \text{ GeV}/c$.

3.6.3 Changes to the trigger system for the Run 2 period

The HLT2 trigger decisions are further optimised for the Run 2 period. This is achieved by performing quality reconstruction offline at the trigger level in real time. With the improvements in computing power and algorithms, full calibration Run 2 data can be used in HLT2, which makes it possible to achieve optimised PID variables with the highest possible resolution.

The differences between the trigger system in Run 1 and 2 are shown in the flow diagrams of Figure 3.14. The hardware-based L0 trigger largely remains unchanged apart from the thresholds that are optimised each year to different pp collision rates. For the software-based trigger, a 10PB buffer is introduced between HLT1 and HLT2 in Run 2, which allows for data

taken in two consecutive weeks to be stored before HLT2 needs to be executed. The computing resources in Run 2 are also increased. The full alignment and calibration can thus be performed with the stored data before HLT2, such that the complete offline quality reconstruction at the trigger level is enabled.

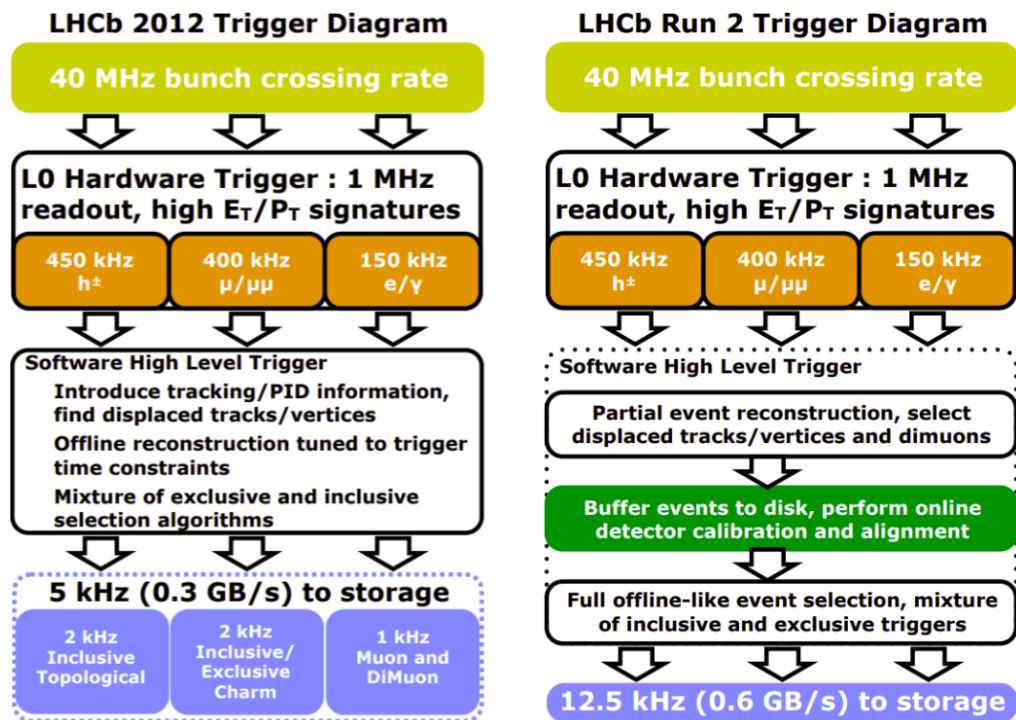


Figure 3.14: Dataflow diagrams of the Run 1 (left) and Run 2 (right) trigger systems. Figure from [121].

3.6.4 Summary of the trigger lines used in this analysis

In summary, the full list of trigger lines used for the analysis is kept in alignment with the current binned analysis [122] and summarised in Table 3.3. Data passing these trigger lines are further processed as discussed in detail in the next Chapter.

3.7 Generation of simulated data

Full Monte Carlo (MC) simulated events are used in multiple places in the angular analysis of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$. For example, they are used to discriminate signal candidates against peaking and combinatorial background with the application of MVA classifiers as described in Chapter 4. They are also vital to understand the detector efficiency as introduced in Chapter 5. These simulated events are generated centrally by the LHCb collaboration. The pp-collisions are generated using PYTHIA8 [123, 124] with a specific LHCb configuration [125] and the

Table 3.3: Trigger lines used in the presented analysis for different data-taking periods.

	2011+2012	2016	2017+2018
L0	Muon DiMuon	Muon DiMuon	Muon DiMuon
HLT1	TrackAllL0 TrackMuon	TrackMVA TrackMuon TwoTrackMVA	TrackMVA TrackMuon TwoTrackMVA
HLT2	Topo [2,3,4]BodyBBDT TopoMu [2,3,4]BodyBBDT DiMuonDetached SingleMuon	Topo [2,3,4]Body TopoMu [2,3,4]Body	Topo [2,3,4]Body TopoMu [2,3,4]Body DiMuonDetached

subsequent hadronic decay is described by EVTGEN [126]. The final state radiation is described by PHOTOS [127]. The GEANT4 toolkit [128] is used to describe the detector geometry and the detector response of the generated particles' interaction with the detector. The same reconstruction and selection procedure as for the collected data can be applied to simulated events.

There are differences between simulated events and collected data. They are corrected by resampling the PID or reweighting in some of their variables. These corrections will be further discussed in the following parts of this thesis where the corresponding simulated events are used.

Chapter 4

Selection

The trigger requirements described in Section 3.6 offer a first-level filter on the raw data to reduce it to a manageable size. The events are fully reconstructed offline and saved to disk. However, further steps are needed to select genuine $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ candidates while suppressing background candidates as much as possible. These steps are referred to as selection and can be achieved using the reconstructed particle momenta, vertex and track qualities, and particle identification criteria.

Background events may be separated into two categories: *peaking* and *combinatorial*. As the name implies, peaking backgrounds come from particular fully reconstructed B-hadron decays that pass the trigger selection and form a peaking structure in the reconstructed B^0 mass. Their topology may be mis-reconstructed or the tracks of these decays may be misidentified. In contrast, combinatorial background events are composed of four-body candidate events where the tracks passing the trigger selection are randomly combined from all tracks in the event.

The selection is applied in five steps. The first step is to satisfy stripping requirements which are summarised in Section 4.1. The next step is preselection, which removes a large portion of combinatorial background candidates. The requirements are summarised in Section 4.2. The following steps are essentially applications of multivariate classifiers. A brief introduction to multivariate classification algorithms is given by Section 4.3. Peaking backgrounds are then dealt with in the third step, where individual multivariate classifiers are designed and implemented for different processes. They are described in Section 4.4. Another multivariate classifier is designed and trained to further suppress combinatorial backgrounds and is described in Section 4.6. Finally, two specific backgrounds are studied in Section 4.7 and are treated by training two more multivariate classifiers. The performance of the full selection is shown in Section 4.8.

Before discussing the detail of the selection stages, the variables involved are firstly defined:

η The pseudorapidity of a track, as defined in Section 3.1.

ϕ The azimuthal angle of a track as shown in Section 3.2.1.

Muon and hadron Isolation variables The Isolation variable of a muon (hadron) can be described as the quantification of the capability for extra tracks to form a vertex with a given muon (hadron) track. Muon Isolation variables also have an additional multivariate approach employed on top of that. Further details can be found in Ref. [129].

DIRA DIRA of a particle is the angle between the momentum vector of the certain particle and the directional vector between the PV and the FD vector of the particle.

IP χ^2 variables IP is short for Impact Parameter as defined in Figure 3.4. The IP χ^2 variable of a particle α stands for the change in the χ^2 of a vertex fit when a track of α is included.

Flight distance χ^2 The FD χ^2 represents the change in the vertex fit χ^2 when including the meson child candidates (see Figure 3.4). The meson of interest should have a large flight distance. This means that the decay vertex formed by the child particles of the meson should have a large FD χ^2 when compared with the PV.

End vertex χ^2 of the K^{*0} End vertex refers to the position of the reconstructed decay vertex of the particle, and End vertex χ^2 is the χ^2 value of the vertex fit. This variable shows how well the tracks form a vertex.

End vertex χ^2 of the K^{*0} is also included. It can help identify the K^{*0} meson, which eventually decays into a kaon and a pion.

isMuon The `isMuon` variable refers to the boolean information to identify muons provided by the muon stations, as discussed in Section 3.5.

DLL $DLL_{\alpha\beta}$ denotes the difference in the log-likelihood between the α and β mass hypotheses. It is outlined in Section 3.4 and 3.5.

ProbNN The `ProbNNx` variable makes use of NN techniques for the PID of particle x ($x = K, \pi, \mu$). Further details may be referred to in Section 3.4.

4.1 Stripping and trigger requirements

All of the data samples used in this analysis come from the B2XMuMu stripping line. This stripping line is designed to select displaced B-meson candidates which decay into final states containing a pair of oppositely charged muons and at least one hadron.

The stripping versions for this analysis benefit from the most recent re-stripping campaign, which applies identical selections across all data-taking periods. This campaign imposes two new hadron isolation variables to Run 1 data samples which used to be only available on Run 2 data

samples. The explicit requirements imposed on $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ candidates for this stripping line are given in Table 4.1.

Table 4.1: Requirements in the B2XMuMu stripping line.

Candidate	Requirement
B^0	Measured mass $\in [4600, 7100]$ MeV/ c^2
	Mass $\in [4700, 7000]$ MeV/ c^2
	End vertex χ^2 per DOF < 8
	DIRA > 0.9999
	Flight distance $\chi^2 > 64$
	Maximum daughter $\chi_{IP}^2 > 9$
K^{*0}	Measured mass $\in [0, 6200]$ MeV/ c^2
	Mass $\in [0, 6200]$ MeV/ c^2
	End vertex χ^2 per DOF < 12
	DIRA > -0.9
	Flight distance $\chi^2 > 16$
	Minimum $\chi_{IP}^2 > 0$
$\mu^+ \mu^-$	Maximum daughter $\chi_{IP}^2 > 6$
	Measured mass < 7100 MeV/ c^2
	End vertex χ^2 per DOF < 12
	DIRA > -0.9
	Flight distance $\chi^2 > 9$
K^+, π^-	Maximum χ_{IP}^2 of either μ^+ or $\mu^- > 6$
	Minimum $\chi_{IP}^2 > 6$
	Hits in the RICH subsystem
μ^+, μ^-	Minimum $\chi_{IP}^2 > 6$
	$DLL_{\mu\pi} > -3$
	Positive isMuon detected, based on muon chamber hits
Tracks	Ghost probability < 0.5
Event	Maximum hits in SPD sub-detector < 600

4.2 Preselection

The preselection requirements are listed in Table 4.2. They are largely in alignment with that used in the current binned analysis, which was adopted from a previous LHCb measurement [39]. It was proven to have a good signal efficiency of more than 99%. All candidates must satisfy the requirements on the angular separation between track pairs to reduce the number of duplicated tracks. The vertex fit χ^2 of the B candidates is also constrained to ensure a good-quality vertex fit. Furthermore, conditions on the momentum of the kaon in the $x-z$ -plane need to be satisfied. This is because for a track originating at the edge of the detector acceptance, tracks of one charge may be discarded due to the limitation of the detector acceptance in the bending plane, whereas tracks of the opposite charge are deflected back into the detector. This behaviour

leads to large raw CP-asymmetry, which can be removed from the samples by fulfilling such conditions.

Additionally, three general PID requirements are applied, as shown in the last three lines in Table 4.2. Such requirements help remove backgrounds involving KK, $\pi\pi$ and $K\pi$ with $K \leftrightarrow \pi$ swap, as well as mis-identified backgrounds such as $B^0 \rightarrow K^{*0} J/\psi$ where the mass hypotheses of the muon and the pion are swapped with each other. The PID requirements for the muons result in a signal efficiency of 98% and reduces background by a further 20% after the full selection is applied.

Table 4.2: Preselection requirements.

Candidate	Requirement
Track pairs	Angular separation $\theta > 1$ mrad
B^0	End-vertex $\chi^2/\text{DOF} < 6$
K^+	$p_Z(K^+) > 2000 \text{ MeV}/c$ $ p_X(K^+) < 0.33(p_Z(K^+) - 2000)$
K	ProbNN _k > 0.2
π	ProbNN π > 0.2
μ	ProbNN μ > 0.3

4.3 Multivariate classifier

Constructing a selection based on individual variables is generally good at separating signal candidates from background. However, such methods do not fully exploit the information available from the variables, including the correlations between variables. The background can thus be further suppressed, making use of the better separating power obtained from an MVA method.

Boosted decision tree (BDT) is the MVA method adopted by this analysis and is discussed in Section 4.3.1. The BDTs in this analysis are trained using the `scikit-learn` python package [130] with the implementation of the XGBOOST algorithm [131]. This algorithm will be explained in Section 4.3.2. The `kFolding` technique described in Section 4.3.3 is utilised to reduce the bias during the training. This strategy is adopted in all the following sections. Multiple BDTs are trained to suppress the peaking backgrounds. A combinatorial BDT is then trained to separate signal candidates from the background, after which two BDTs are used to reduce specific partially reconstructed backgrounds. The detailed training procedure and performance of the BDTs are described in Sections 4.4 - 4.7 respectively.

4.3.1 Decision tree mechanism

The BDT algorithm is an extension of the decision tree algorithm. A single decision tree works as a sequence of splitters that maximise the classification power that separates the signal

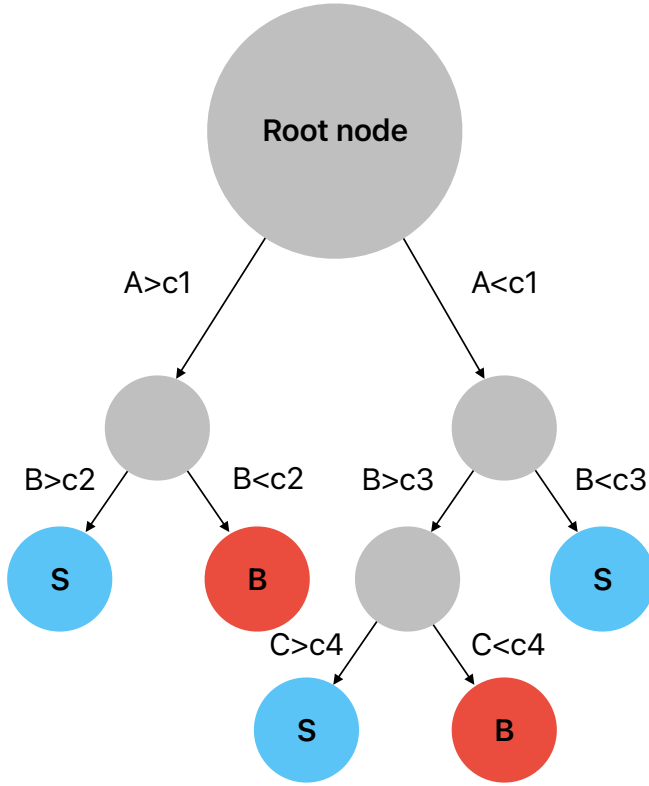


Figure 4.1: A schematic illustration of the BDT working mechanism of maximum depth 3. The split starts from the root node and continues until the maximum depth is reached or certain criteria are fulfilled, ending up in a signal (S) or background (B) node. A, B and C are exemplary MVA variables classifying candidates.

sample from the background for a given set of variables. The algorithm requires a dataset with candidates labelled as signal or background as the training sample. Each training sample is then split into two subsamples, or leaf nodes. The initial split is also known as the root node. A tree grows from the root node and the nodes are continuously split. Each of these splits is based on one variable at a time. The growth of the tree can be terminated by constraining the number of nodes, which is often referred to as the maximum depth of the tree in literature. The terminated leaf nodes are then classified as signal or background nodes, depending on the dominant category in the nodes. An illustration of the decision tree algorithm is shown in Figure 4.1.

To achieve the best splitting power, the condition on the variable for each node is determined based on the Gini impurity. The Gini impurity scheme measures the likelihood of a random event of the full data set being misclassified based on the number of signal and background. It is defined as,

$$(4.1) \quad Gini_{node} \equiv P_S(1 - P_S) + P_B(1 - P_B) = 2P_S(1 - P_S) ,$$

where $P_{S(B)}$ stands for the purity of the signal or background. It can be written as

$$(4.2) \quad P_{S(B)} = \frac{N_{S(B)}}{N_{tot}} ,$$

where $N_{S(B)}$ is the signal (background) yield in the node and N_{tot} is the total number of candidates. The quantity $\mathcal{L}(\theta)$ is minimised in a certain node that is split into two daughter nodes according to the threshold θ . The quantity $\mathcal{L}(\theta)$ is defined as

$$(4.3) \quad \mathcal{L}(\theta) = \frac{N_{daughter1}}{N_{parent}} Gini_{daughter1} + \frac{N_{daughter2}}{N_{parent}} Gini_{daughter2} .$$

Here, $N_{daughter1(2)}$ represents the number of candidates in either daughter node, and N_{parent} represents the number of candidates in the parent node. This process is repeated until the daughter nodes reach the layer whose depth equals the maximum depth constraint. The classifier obtained as stated above is a weak classifier on its own and is only slightly better than random guessing. The ensemble method, such as gradient boosting, is thus introduced to create a better classifier.

4.3.2 Gradient boosting and xgboost algorithm

The idea of boosting decision trees is to train a sequence of decision trees, where the latter ones compensate for the shortcomings of the former ones. This forms a forest of decision trees. For example, in the classical boosting scheme the events incorrectly classified as signal and background are given higher weights whereas those that ended up in the correct nodes are given lower weights. Such weights can be understood as an evaluation of the shortcomings that the boosting scheme is trying to compensate for.

Gradient boosting, however, uses an error function to model the shortcomings. The process of generating decision trees iteratively is formalized as a gradient descent algorithm over a target error function. In this analysis, the error function is chosen to be the logistic regression for binary classification [132]. It is defined as

$$(4.4) \quad \sum_{i=1}^n \mathcal{L}(\hat{y}_i, y_i) = -[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]$$

Here, i stands for an individual event and n stands for the total number of the training sample. $\hat{y}$ represents the predicted value of event i and in our case, the probability of the event i being signal. y represents the actual value of the event i , 1 if it is signal and 0 if it is background. The gradient descent algorithm is then performed, minimising the error function to achieve the optimal split criterion. The first classifier starts with all parameters randomised, and the latter classifiers inherit the predicted errors of the previous model as the starting point of fitting. The final BDT classifier can be obtained by taking a weighted sum of the BDT ensemble.

The term XGBOOST is short for Extreme Gradient Boosting [131]. Unlike the normal gradient boosted decision trees (GBDT) where trees are built sequentially, XGBOOST builds trees in

parallel. This algorithm is thus compatible with GPU usage and the time to train a classifier is a lot shorter. In XGBOOST, the trees are pruned by setting the maximum depth of the trees, instead of a greedy strategy where criteria are first. The algorithm is also optimised to make full use of the hardware resources. These optimisations ensure the speedy performance of XGBOOST. To prevent overfitting, XGBOOST penalise complex models by adding LASSO (L1) regularisation term

$$(4.5) \quad \alpha \propto \sum_{j=1}^m |\beta_j|$$

and Ridge (L2) regularisation term

$$(4.6) \quad \lambda \propto \sum_{j=1}^m \beta_j^2$$

to the loss function. In both terms, β_j stands for any coefficient of the model. Moreover, by employing the distributed weighted Quantile Sketch algorithm and the built-in cross-validation feature, XGBOOST can find the optimal split criteria sufficiently from a weighted training sample.

4.3.3 kFolding technique to reduce bias

An important factor limiting the performance of the MVA technique is the small sample size available for the training sample, and this problem will worsen when splitting the full data sample. The same BDT cannot be evaluated on the same training sample. There might be statistical fluctuations embedded in the training sample, and the BDT might be biased if it is trained and evaluated on the same data sample. This effect is also known as overfitting. Furthermore, the BDT needs to be optimised on an unbiased dataset for the continuous BDT classifier obtained as described above. The k -folding cross-validation technique is therefore employed to deal with this problem.

At the start of training, the available data is split into $k = 10$ subsets. Each subset contains the same proportions of the different classes as the full dataset, the classes being signal and background in this case. In every iteration, a BDT is trained on $k - 1 = 9$ subsets, namely 90% of the data, after which the resulting BDT classifier is evaluated on the remaining subset that contains 10% of the data. It is therefore ensured that each of the BDTs is evaluated on the part of data that was not used for training. This process is repeated 10 times and 10 statistically independent BDT classifiers are obtained. The final BDT score is set to be the average score of the 10 BDTs evaluated on the corresponding subset that was not used for training.

4.4 Peaking backgrounds

Several kinds of fully reconstructed backgrounds have peaking structures in the B^0 mass and may cause bias in the fit. They are listed in Table 4.3. The estimation of the level of

different peaking backgrounds makes use of the fragmentation fractions $f_s/f_d = 0.26$ [133] and $f_\Lambda/f_d = 0.46$ [134] to measure the relevant branching fractions relative to the branching fraction of the signal mode.

Table 4.3: A summary of the peaking backgrounds. The quoted rate with respect to the signal is before any selection (product of the relevant hadronisation and branching fractions)

Mode	Type	MisID	Rate wrt signal
$B^0 \rightarrow K^{*0}\mu^+\mu^-$	Double hadron mis-ID	$K \leftrightarrow \pi$	1.
$B_s^0 \rightarrow \phi\mu^+\mu^-$	Single hadron mis-ID	$K \rightarrow \pi$	0.89
$B^0 \rightarrow K^{*0}J/\psi$	Hadron-muon swap	$\mu \leftrightarrow K$	1351
$B^0 \rightarrow K^{*0}J/\psi$	Hadron-muon swap	$\mu \leftrightarrow \pi$	1351
$B^0 \rightarrow K^{*0}\psi(2S)$	Hadron-muon swap	$\mu \leftrightarrow K$	632
$B^0 \rightarrow K^{*0}\psi(2S)$	Hadron-muon swap	$\mu \leftrightarrow \pi$	632
$\Lambda_b^0 \rightarrow pK\mu^+\mu^-$	Single hadron mis-ID	$p \rightarrow \pi$	0.28
$\Lambda_b^0 \rightarrow pK\mu^+\mu^-$	Double hadron mis-ID	$p \rightarrow K, K \rightarrow \pi$	0.28

For each of these background modes, an XGBOOST BDT is trained. The BDT cuts are then combined to act as a veto and all events that failed the cut are excluded. The BDTs are trained on simulated samples calibrated using control regions in the data as described in Section 5.1. For all the peaking background BDTs, the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ simulation is used as the signal. For the background simulations, the relevant misidentification hypotheses are enforced. The trigger, stripping, preselection and invariant mass cuts are applied to both the signal and background. As the sample sizes are limited after applying these initial selection requirements, the k -folding technique is employed with 10 folds. One fold out of 10 is chosen randomly for use in this analysis. Most of the BDT cut points are determined the same as the ongoing binned analysis [122], and harsher cuts are sometimes needed to reduce the overall estimated peaking backgrounds to a reasonable level, e.g. below 1% of the signal after the full selection is applied.

The BDTs use invariant masses under the swapped hypotheses, such as $m_{swap}(K\pi)$, and PID variables for training. Since there is the possibility that **ProbNN** is correlated with other kinematic variables, the **DLL** variables are chosen instead. The training variables of the BDTs are summarised in Table 4.4. Since the preselection, including the relatively harsh general PID requirements on the hadrons (see Table 4.2), are already applied to the training samples, the BDTs are optimised to remove peaking backgrounds which are similar to the signals.

The responses of the BDTs are presented in Figure 4.2 - 4.9. The signal and background rejection efficiencies are outlined in Table 4.5. For BDTs involving events from the J/ψ , only events with the swapped dimuon mass hypothesis in the J/ψ region ($3050 < m(\mu^+\mu^-) < 3150$ MeV/ c^2) are considered in the training. Likewise for BDTs involving events from the $\psi(2S)$, only events with the swapped dimuon mass hypothesis in the $\psi(2S)$ region ($3626 < m(\mu^+\mu^-) < 3746$

Table 4.4: Training variables used in the peaking background BDTs.

Background	mis-ID	Training variables
$B^0 \rightarrow K^{*0} \mu^+ \mu^-$	$K \leftrightarrow \pi$ mis-ID	dihadron mass with the background hypothesis, $DLL_{K\pi}$ for the hadrons
$B_s^0 \rightarrow \phi \mu^+ \mu^-$	$K \rightarrow \pi$ mis-ID	dihadron mass with the background hypothesis, $DLL_{K\pi}$ for the pion
$B^0 \rightarrow K^{*0} J/\psi$	$\mu \leftrightarrow K$ swap	dimuon mass with the background hypothesis, $DLL_{K\pi}$ and $DLL_{\mu\pi}$ for the kaon and muon
$B^0 \rightarrow K^{*0} J/\psi$	$\mu \leftrightarrow \pi$ swap	dimuon mass with the background hypothesis, $DLL_{\mu\pi}$ for the pion and muon
$B^0 \rightarrow K^{*0} \psi(2S)$	$\mu \leftrightarrow K$ swap	dimuon mass with the background hypothesis, $DLL_{K\pi}$ and $DLL_{\mu\pi}$ for the kaon and muon
$B^0 \rightarrow K^{*0} \psi(2S)$	$\mu \leftrightarrow \pi$ swap	dimuon mass with the background hypothesis, $DLL_{\mu\pi}$ for the pion and muon
$\Lambda_b^0 \rightarrow p K \mu^+ \mu^-$	$p \rightarrow \pi$ mis-ID	dihadron mass with the background hypothesis, B mass with the background hypothesis, $DLL_{p\pi}$ for the pion
$\Lambda_b^0 \rightarrow p K \mu^+ \mu^-$	$p \rightarrow K$, $K \rightarrow \pi$ mis-ID	dihadron mass with the background hypothesis, B mass with the background hypothesis, $DLL_{K\pi}$ for the pion, $DLL_{p\pi}$ for the kaon

MeV/ c^2) are considered. The efficiencies shown are still for the full samples.

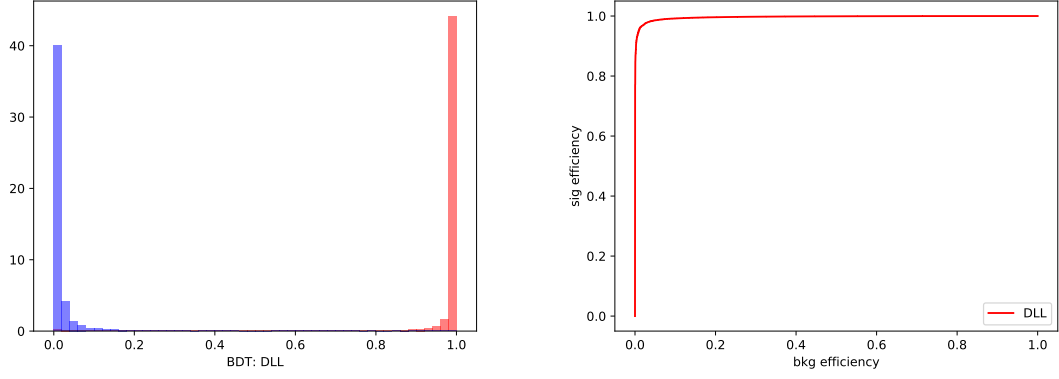


Figure 4.2: Result of the training (left) and the ROC curve (right) of the double mis-ID BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

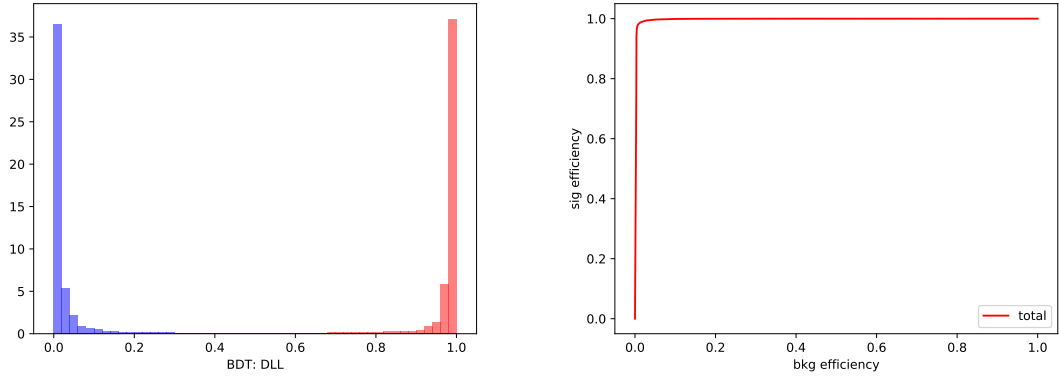


Figure 4.3: Result of the training (left) and the ROC curve (right) of the $B_s^0 \rightarrow \phi \mu^+ \mu^-$ BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

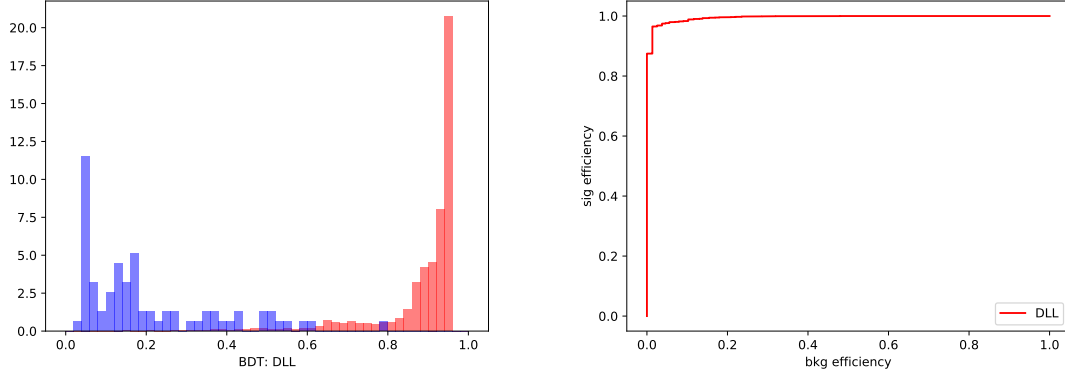


Figure 4.4: Result of the training (left) and the ROC curve (right) of the muon-pion swap from the J/ψ BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

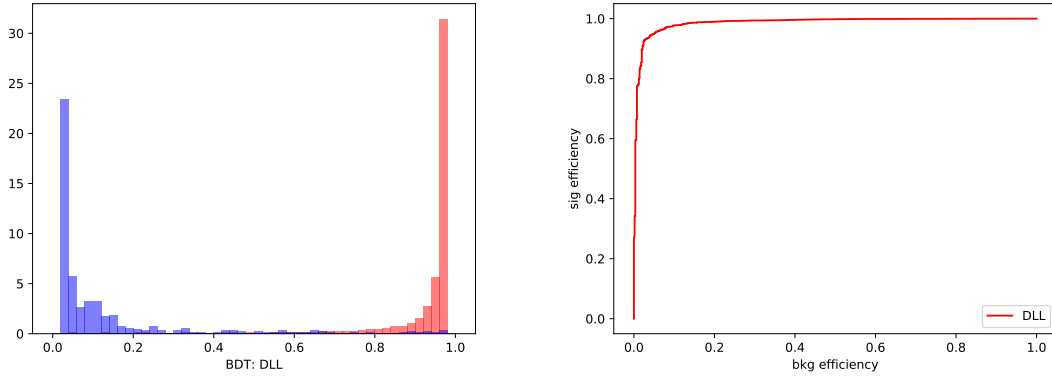


Figure 4.5: Result of the training (left) and the ROC curve (right) of the muon-pion swap from the J/ψ BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

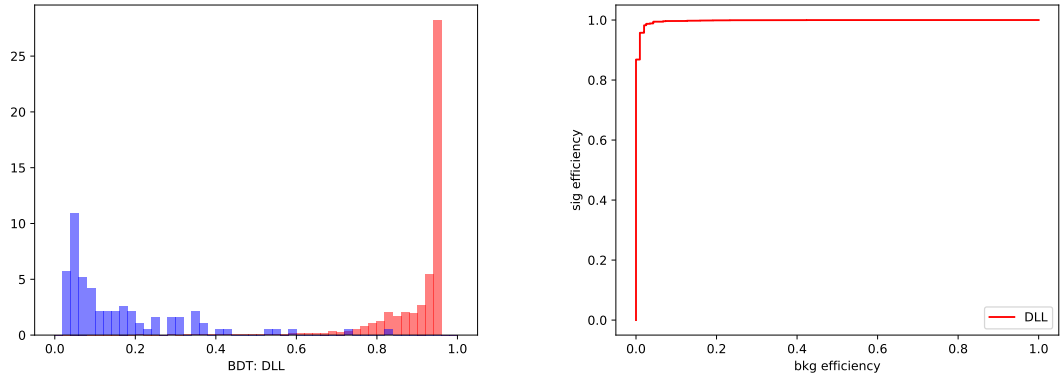


Figure 4.6: Result of the training (left) and the ROC curve (right) of the muon-pion swap from the $\psi(2S)$ BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

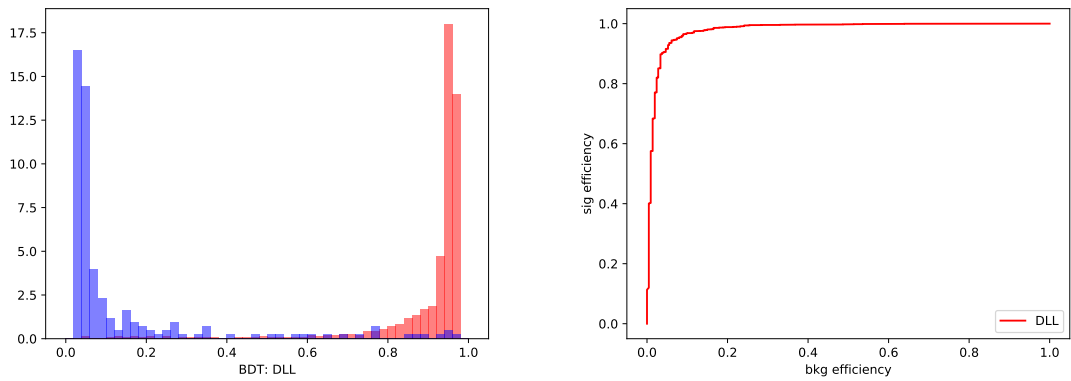


Figure 4.7: Result of the training (left) and the ROC curve (right) of the muon-pion swap from the $\psi(2S)$ BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

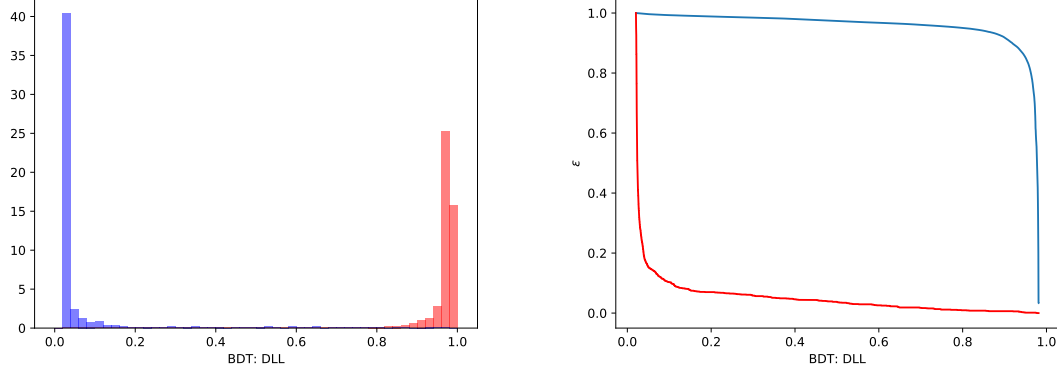


Figure 4.8: Result of the training (left) and the ROC curve (right) of the $\Lambda_b^0 \rightarrow p K \mu^+ \mu^-$, $p \rightarrow \pi$ BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

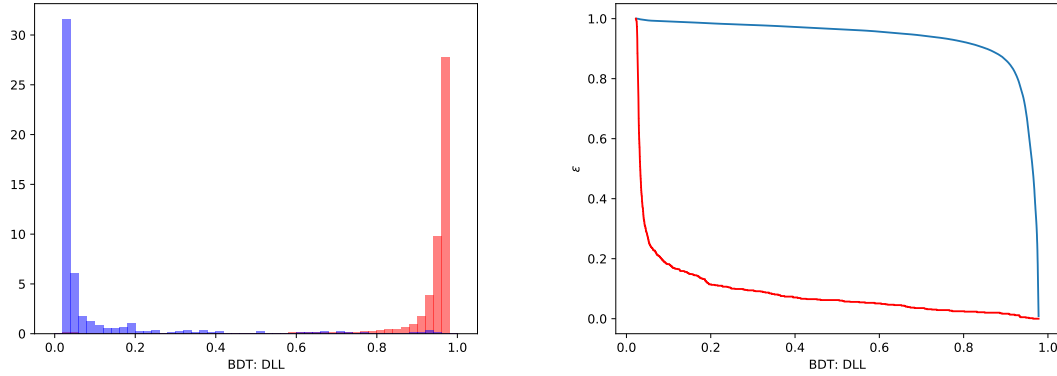


Figure 4.9: Result of the training (left) and the ROC curve (right) of the $\Lambda_b^0 \rightarrow p K \mu^+ \mu^-$, $p \rightarrow K$, $K \rightarrow \pi$ BDT on signal and background. The data is kFolded and the response shown here is on the testing sample.

Table 4.5: Signal efficiency and background rejection efficiency for each background, with figures showing the BDT response for each background indicated.

Background	mis-ID	Figure	Signal efficiency	Background rejection efficiency
$B^0 \rightarrow K^{*0} \mu^+ \mu^-$	$K \leftrightarrow \pi$ mis-ID	4.2	$99.24 \pm 0.01\%$	$89.3 \pm 0.4\%$
$B_s^0 \rightarrow \phi \mu^+ \mu^-$	$K \rightarrow \pi$ mis-ID	4.3	$99.896 \pm 0.004 \%$	$90.2 \pm 0.2 \%$
$B^0 \rightarrow K^{*0} J/\psi$	$\mu \leftrightarrow K$ swap	4.4	$99.23 \pm 0.01\%$	$96 \pm 2\%$
$B^0 \rightarrow K^{*0} J/\psi$	$\mu \leftrightarrow \pi$ swap	4.5	$99.634 \pm 0.007\%$	$97.5 \pm 0.7\%$
$B^0 \rightarrow K^{*0} \psi(2S)$	$\mu \leftrightarrow K$ swap	4.6	$99.557 \pm 0.008\%$	$96 \pm 2\%$
$B^0 \rightarrow K^{*0} \psi(2S)$	$\mu \leftrightarrow \pi$ swap	4.7	$99.915 \pm 0.008\%$	$98 \pm 1\%$
$\Lambda_b^0 \rightarrow p K \mu^+ \mu^-$	$p \rightarrow \pi$ mis-ID	4.8	$99.304 \pm 0.008\%$	$89.6 \pm 0.5\%$
$\Lambda_b^0 \rightarrow p K \mu^+ \mu^-$	$p \rightarrow K,$ $K \rightarrow \pi$ mis-ID	4.9	$98.948 \pm 0.008\%$	$83.1 \pm 0.8\%$

4.5 $B^+ \rightarrow K^+\mu^+\mu^-$ veto

Another type of background is expected because the decay products of $B^+ \rightarrow K^+\mu^+\mu^-$ could be paired to a random pion from the rest of the event. These events are present in the upper mass sideband $m_{K\pi\mu\mu} > 5380 \text{ MeV}/c^2$, the mass of a B meson plus a pion. They can be excluded by applying a cut on the $K^+\mu^+\mu^-$ mass, namely $5220 < m(K^+\mu^+\mu^-) < 5340 \text{ MeV}/c^2$, which corresponds to a $60 \text{ MeV}/c^2$ window around the B^+ mass. Since the B^0 mass is not a variable to be fit for in this analysis, a histogram is created to describe this cut-out region by mapping out toys removed by this veto in regions of $m_{K\pi\mu\mu}$, $m_{K\pi}$, $\cos\theta_K$ and q^2 . This histogram is then integrated over $m_{K\pi}$ to remove its dependence. Data events that fall into this region are set to zero. Such a cut has a sculpting effect in the plane of $m_{K\pi\mu\mu}$, $\cos\theta_K$ and q^2 , which can be seen in the right plot of Figure 4.15, and the background PDF is normalised accordingly. This effect is considered to be trivial and thus neglected as it only takes effect in the upper mass sideband. More details of this veto can be found in Ref. [135].

4.6 Combinatorial BDT

The combinatorial background is reduced by training a BDT using the XGBOOST algorithm. The BDT is trained using $B^0 \rightarrow K^{*0}\mu^+\mu^-$ simulation that is calibrated to the data. Using simulation ensures that the topologies of the events are similar to what is expected in data. For the background, the BDT is trained on data in the upper B^0 -mass sideband, $5350 < m_{K\pi\mu\mu} < 5700 \text{ MeV}/c^2$. Both signal and background are constrained in the $K\pi$ mass region $0.796 < m_{K\pi} < 0.996 \text{ GeV}/c^2$ and the dimuon mass squared regions of interest for this analysis, namely $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ and $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$. The preselection mentioned in Section 4.2 and peaking background vetos described in Section 4.4 are applied to both candidates as well. Two separate BDT selections are trained, one for Run 1 and one for Run 2. When combining events taken in different years in Run 1 and Run 2, a mixture weight is assigned to individual events by taking the ratio of the MC yields in the J/ψ window over the J/ψ yields in data. The k -folding technique is used with 10 folds.

4.6.1 BDT training variables

The variables and training samples are chosen to enable the combinatorial BDT to effectively eliminate a wider spectrum of topologies, such as combinatorial events where both hadrons and muons are not displaced from the primary vertex [136]. The full BDT training variables are listed and discussed below.

- B^0 lifetime
- B^0 p_T

- B^0 end vertex χ^2
- B^0 DIRA
- $DLL_{\mu\pi}(\mu^+)$
- $DLL_{\mu\pi}(\mu^-)$
- $DLL_{K\pi}(K)$
- $DLL_{K\pi}(\pi)$
- Muon and hadron isolation variables
- IP χ^2 of the pion
- IP χ^2 of the kaon
- IP χ^2 of the μ^+
- IP χ^2 of the μ^-
- IP χ^2 of the B^0
- Flight distance χ^2 of the K^{*0}
- Flight distance χ^2 of the B^0
- End vertex χ^2 of the K^{*0}

The training variables for the Run 1 and Run 2 BDTs are shown in Figure 4.10 and Figure 4.11 respectively.

Cautions need to be taken to ensure that the simulation has the same behaviour as the data. The background contribution in $B^0 \rightarrow J/\psi K^{*0}$ data is first removed using the *sPlot* technique by assigning each event an *sWeight*. The combinatorial BDT is applied to $B^0 \rightarrow J/\psi K^{*0}$ simulation and *sWeighted* $B^0 \rightarrow J/\psi K^{*0}$ data in different years and different $m_{K\pi}$ regions. The distributions of the input variables on both sets are examined. Such an examination for 2018 is shown as an example in Figure 4.12. The agreements are good in all years.

4.6.2 Combinatorial BDT training results

The results of the training for Run 1 and Run 2 combinatorial BDTs can be seen in Figure 4.13. The ROC curves for Run 1 and Run 2 are shown in Figure 4.14.

The significance figure of merit $S/\sqrt{S+B}$, where S and B represent the expected signal and background yields inside the signal mass region of $5170 < m_{K\pi\mu\mu} < 5700$ respectively, is utilised to determine the optimum cut points. To estimate the signal yields, the selection

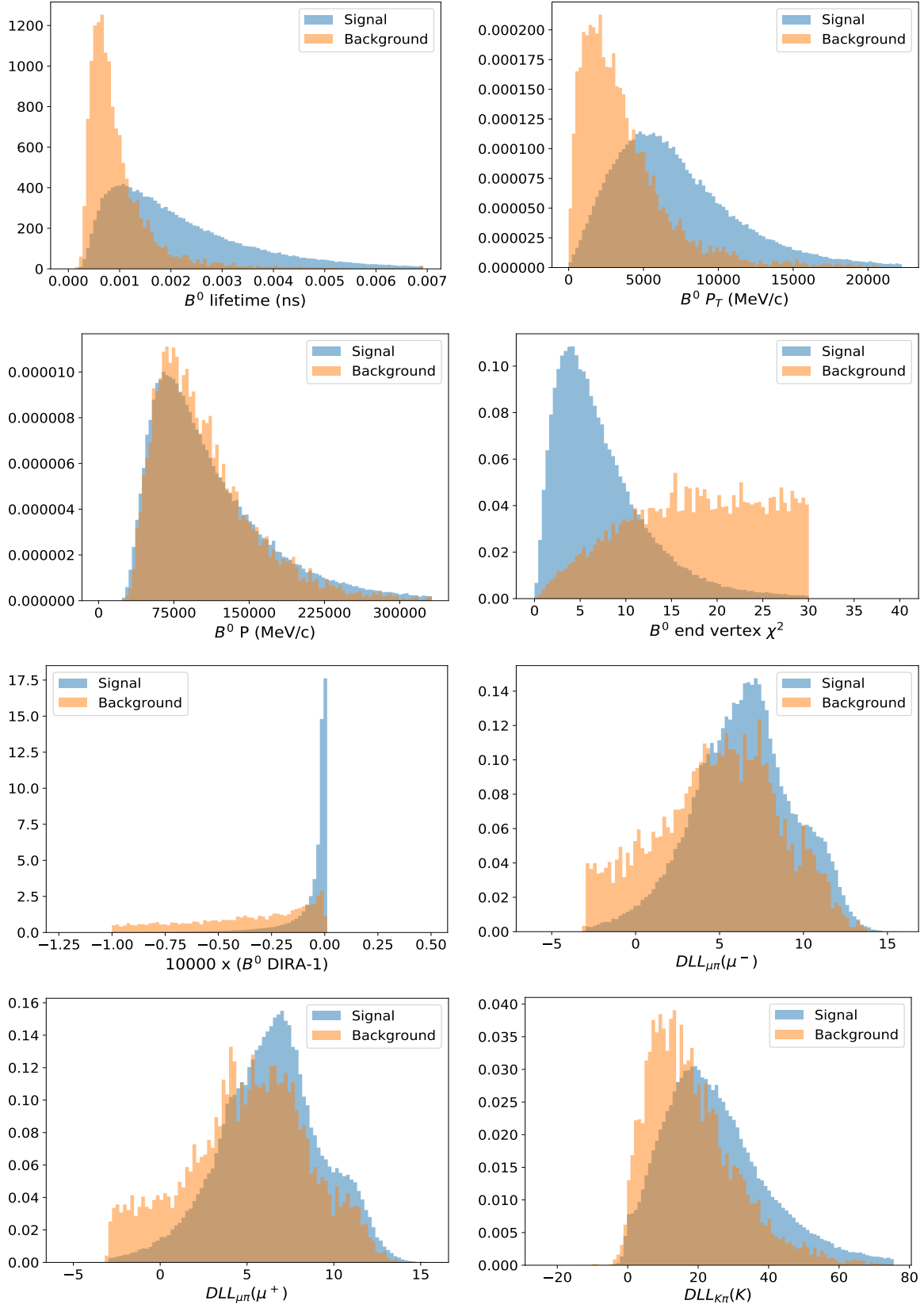


Figure 4.10: Training variables of the Run 1 combinatorial BDT.

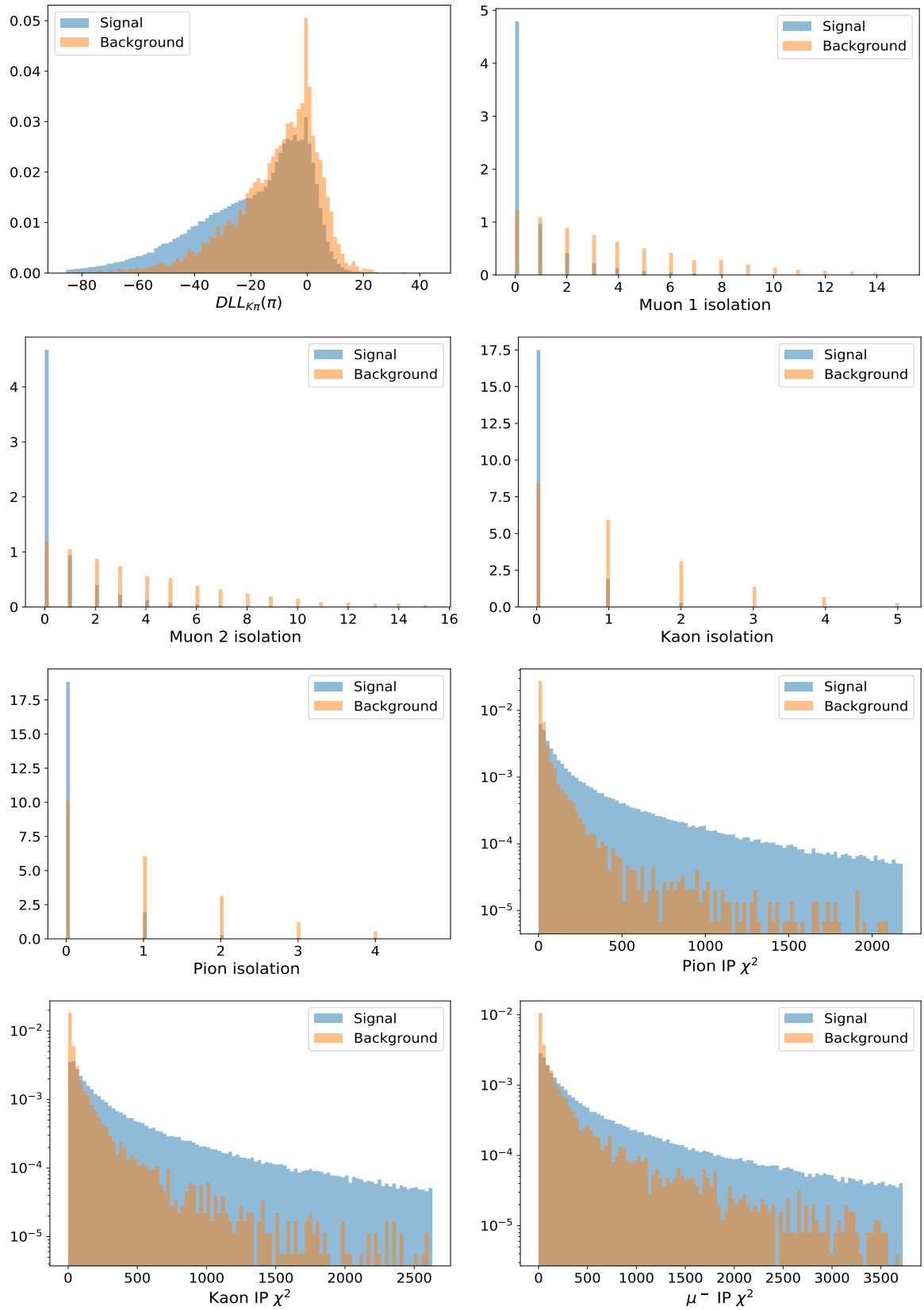


Figure 4.10 continued: Training variables of the Run 1 combinatorial BDT.

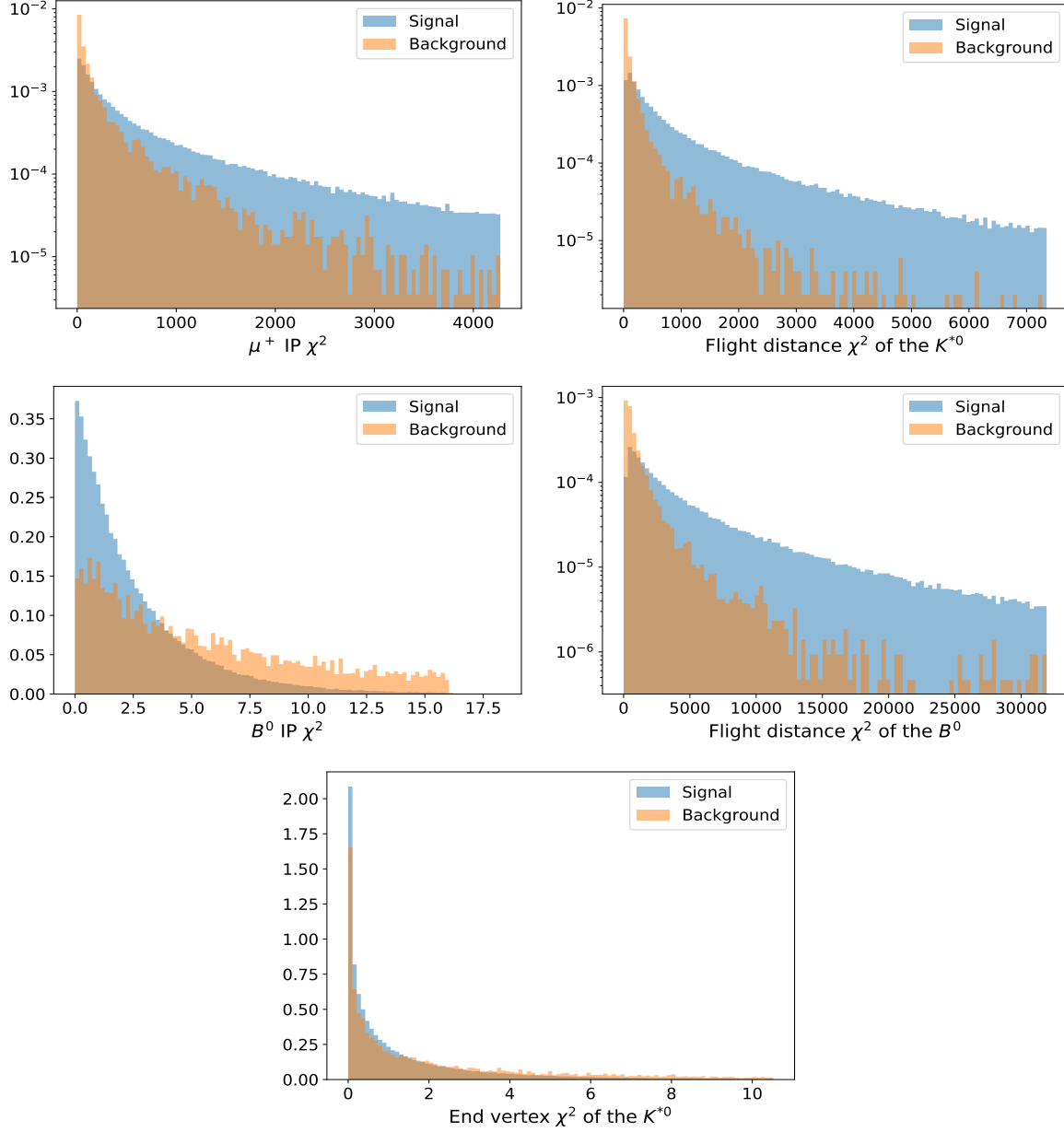


Figure 4.10 continued: Training variables of the Run 1 combinatorial BDT.

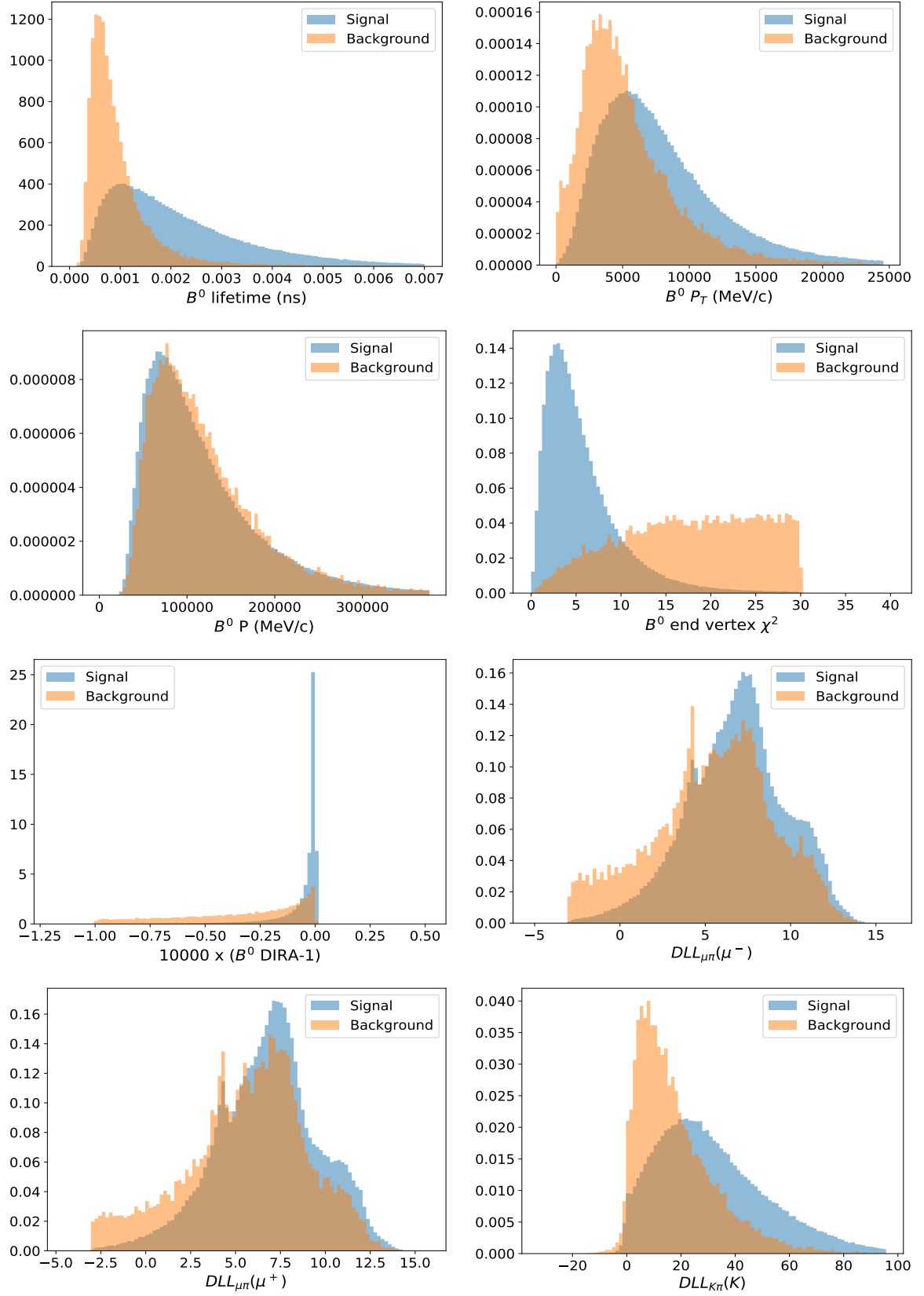


Figure 4.11: Training variables of the Run 2 combinatorial BDT.

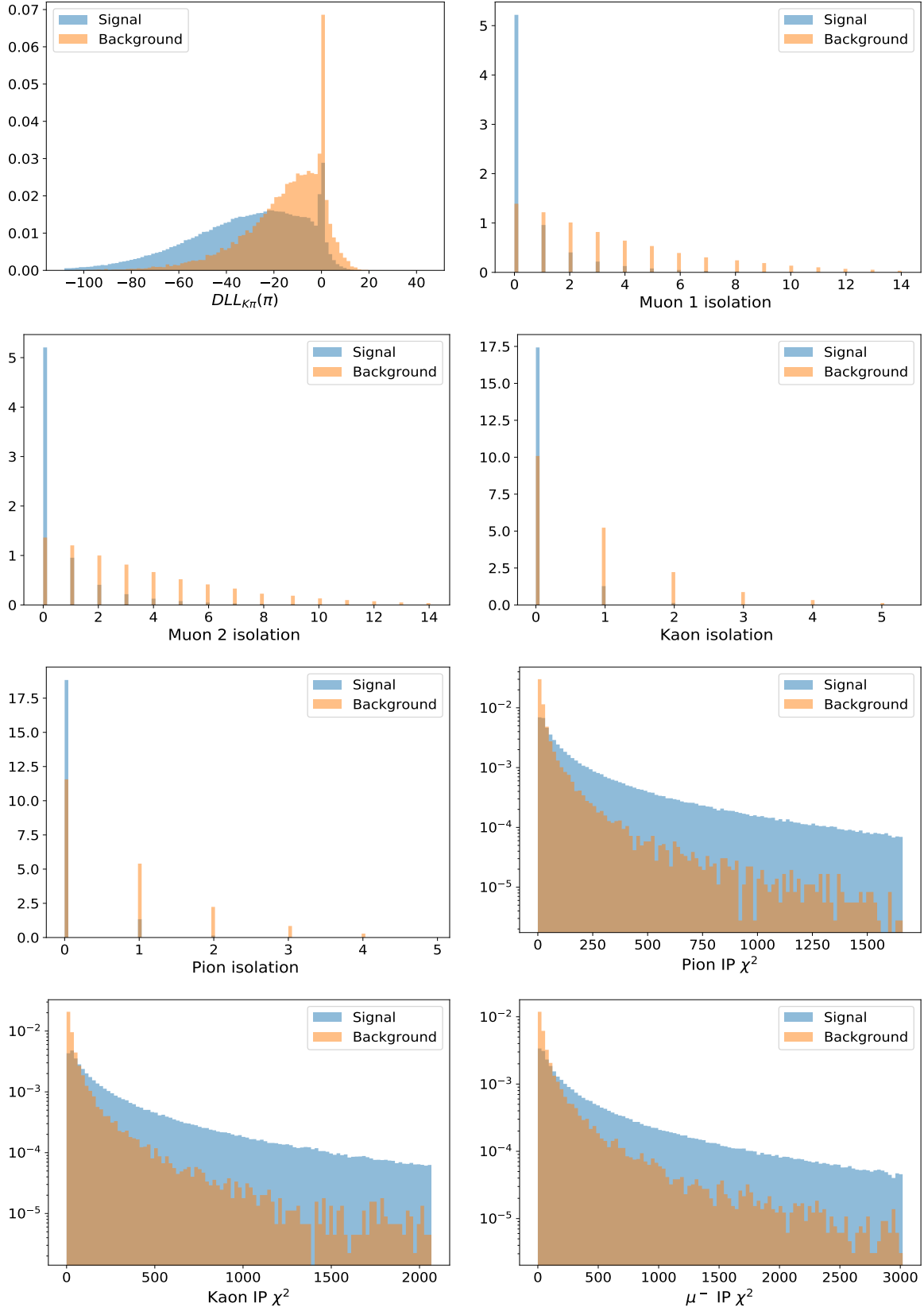


Figure 4.11 continued: Training variables of the Run 2 combinatorial BDT.

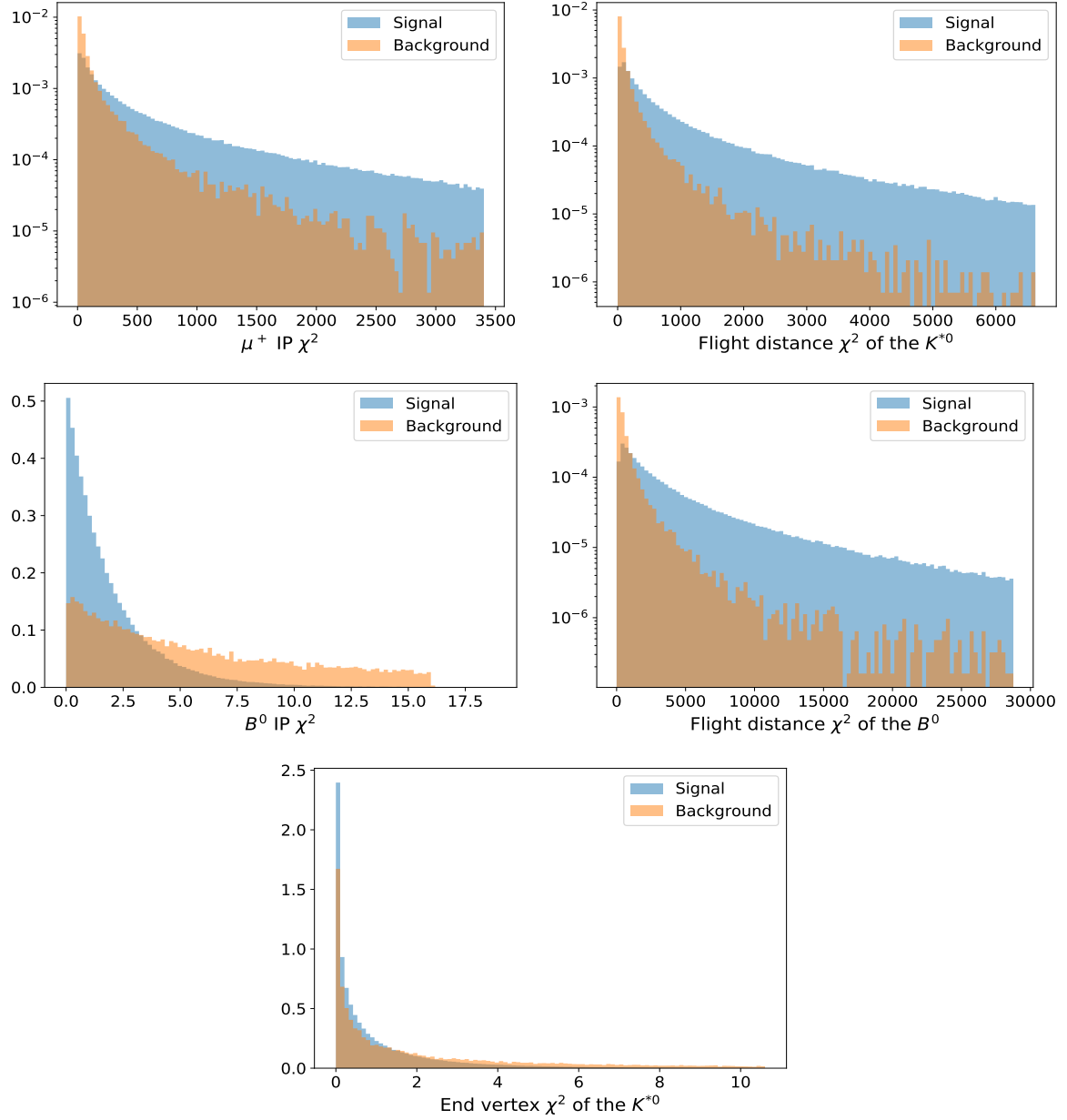


Figure 4.11 continued: Training variables of the Run 2 combinatorial BDT.

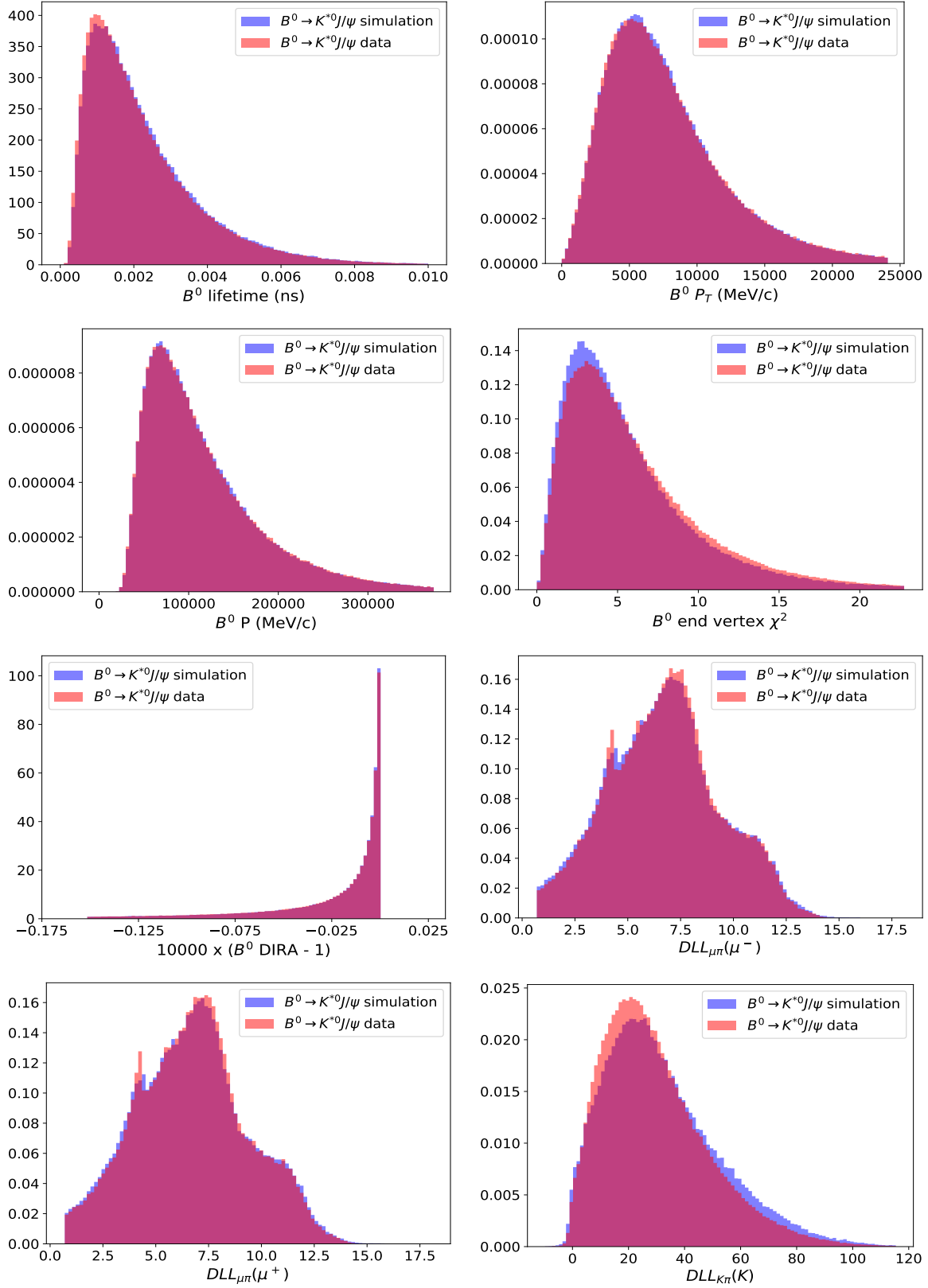
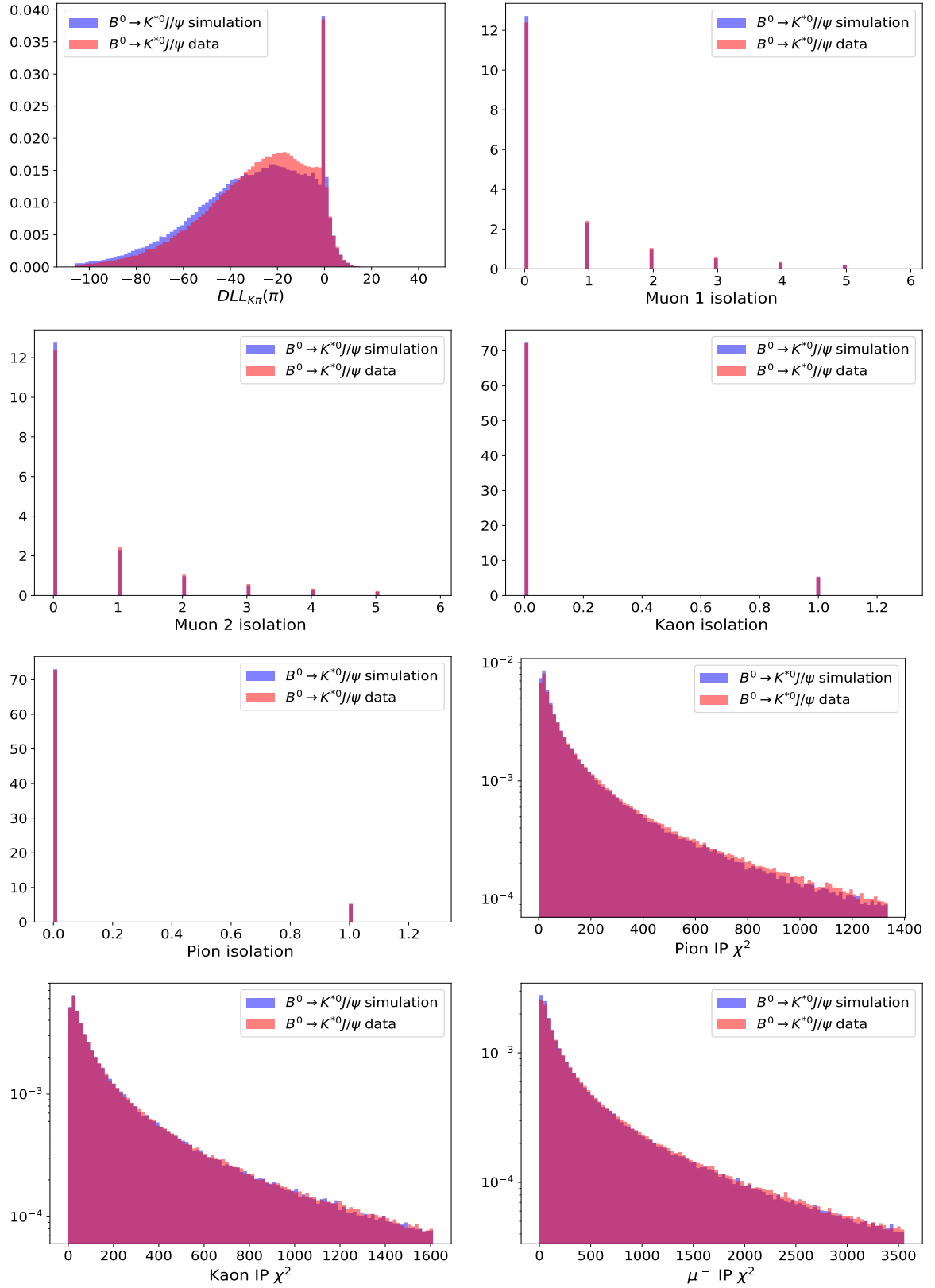


Figure 4.12: Comparison between $B^0 \rightarrow K^{*0} J/\psi$ simulation (blue) and sWeighted $B^0 \rightarrow K^{*0} J/\psi$ data (red) for the BDT training variables in year 2018.



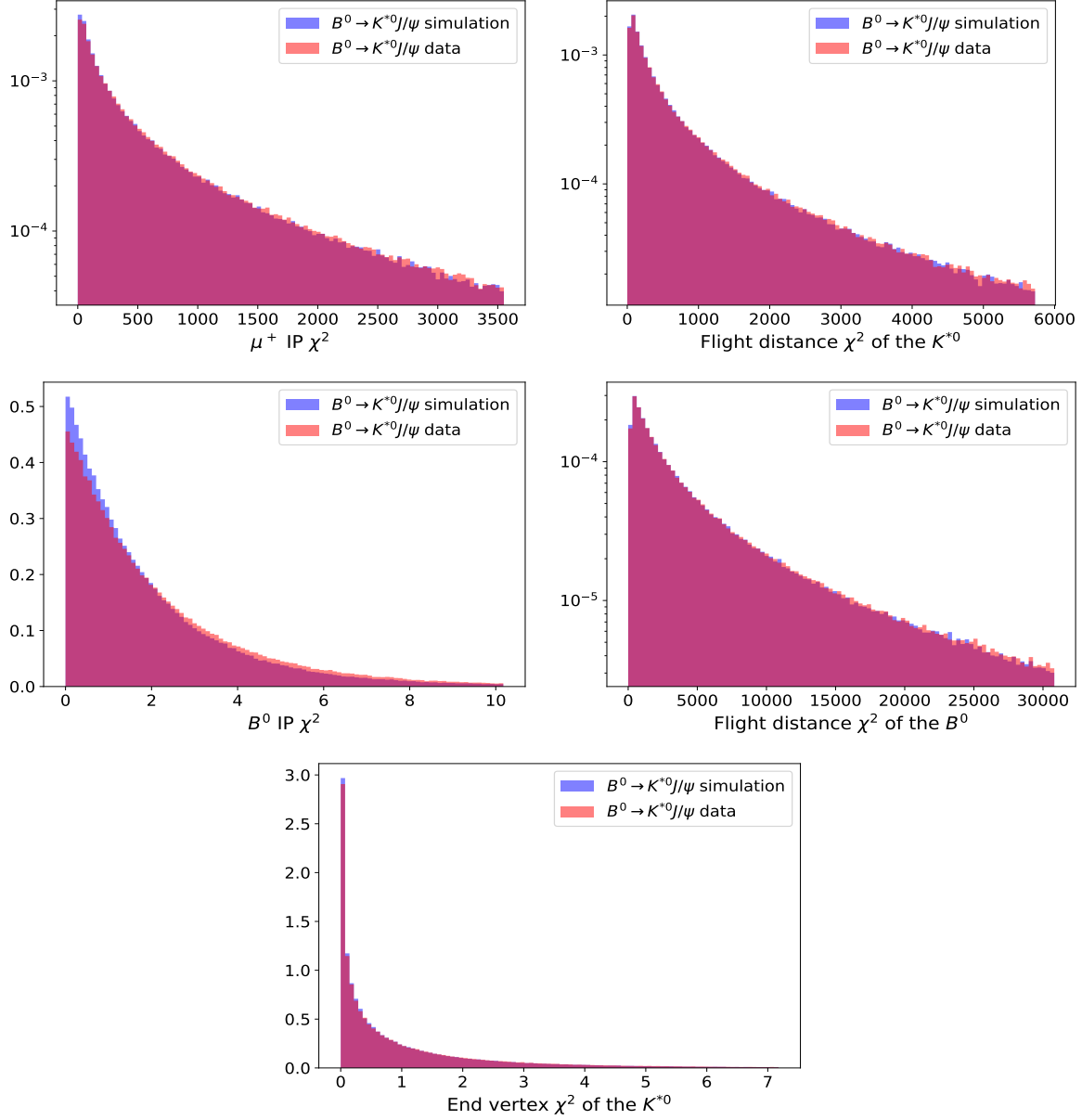


Figure 4.12 continued: Comparison between $B^0 \rightarrow K^{*0} J/\psi$ simulation (blue) and sWeighted $B^0 \rightarrow K^{*0} J/\psi$ data (red) for the BDT training variables in year 2018.

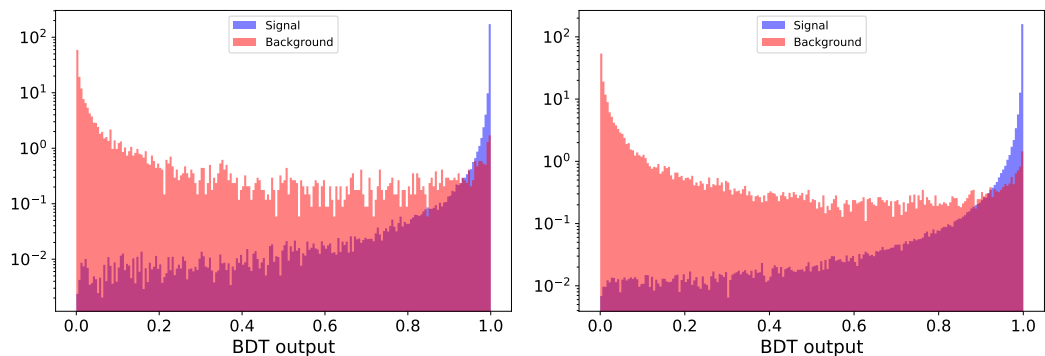


Figure 4.13: Result of the training of the combinatorial BDT for (left) Run 1 and (right) Run 2.

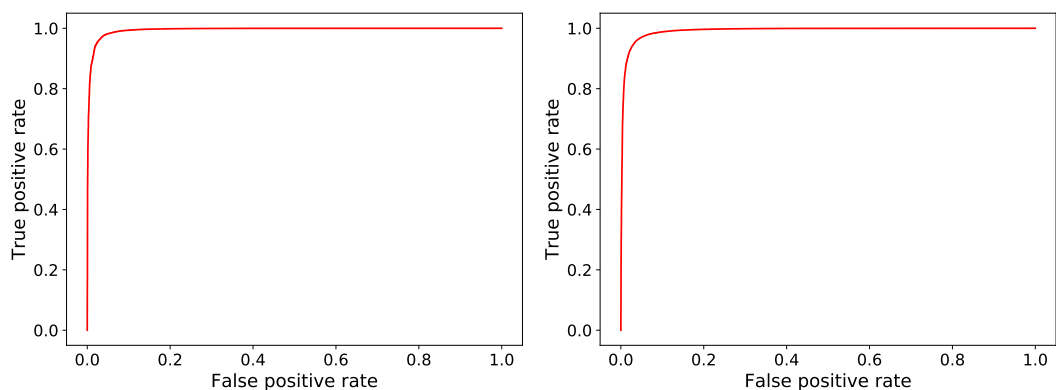


Figure 4.14: ROC curves from the training of the combinatorial BDT for (top) Run 1 and (bottom) Run 2.

efficiency is taken from signal and $B^0 \rightarrow J/\psi K^*$ simulation, and the $B^0 \rightarrow J/\psi K^*$ data is scaled to the ratio of rare mode and control mode branching fractions. As for the background yields, the data events in the sideband are fitted with an exponential function and the yields are extrapolated into the signal B^0 mass range. The significance $S/\sqrt{S+B}$ can then be maximised for the Run 1 and Run 2 BDTs. The optimal BDT cut points and efficiencies are summarised in Table 4.6.

Table 4.6: Optimum BDT cut points and efficiencies

Run Period	BDT cut	Significance	ε_{sig}	ε_{bkgrej}
1	0.97672	30.4	0.924	0.984
2	0.97983	60.8	0.917	0.987

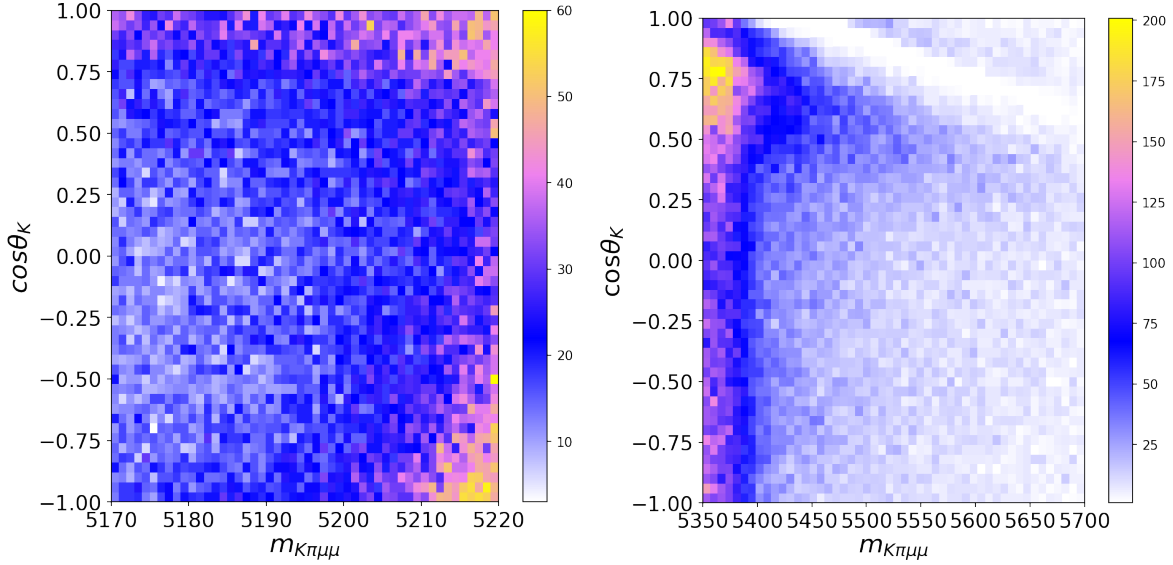


Figure 4.15: Plots showing the $\cos\theta_K$ asymmetry in the range of $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$. The left plot is plotted in the lower mass sideband and the right in the upper mass sideband. A diagonal gap can be spotted on the top of the right plot, which is due to the effect of the $B^+ \rightarrow K^+ \mu^+ \mu^-$ veto mentioned in Section 4.5.

4.7 Random particle backgrounds and Part-Reco BDTs

Further studies revealed that the data suffers from the presence of an extra mis-ID background which involves pions and kaons from the PV. A significant contribution comes from the process of $B^+ \rightarrow K^{*+} \mu^+ \mu^-$, with $K^{*+} \rightarrow K^+ \pi^0$ or $K^{*+} \rightarrow K_S^0 \pi^+$. The neutral particle from the K^{*+} meson may be missed and a random pion or kaon (misidentified as a pion) from the PV is picked up instead. These backgrounds originate from a rare decay with an equivalent branching fraction to the signal process but can have a very similar topology to the signal, and thus a large reconstruction efficiency, due to the presence of multiple hadrons from the PV. Such backgrounds are particularly dangerous as they are present across the whole B^0 -mass range (owing to the random hadron momentum from the PV). In addition, their angular distribution may differ from that of the combinatorial background. Both of these points mean that such backgrounds introduce a dependence of the angular distribution on the B^0 -mass distribution that can bias the angular distribution of the combinatorial background and therefore of the signal amplitudes. Such structures can be seen in Figure 4.15. The triangular-like shape in the lower mass sideband at low and high $\cos\theta_K$ and the clustering behaviour at high $\cos\theta_K$ in the upper mass sideband clearly suggest that correlations persist between the $\cos\theta_K$ and $m_{K\pi\mu\mu}$ background. The existence of such peaking structures in $\cos\theta_K$ makes it difficult to model the background with the current combinatorial background shape described in Section 6.3.

The selection steps described above help suppress such backgrounds, especially the background from $B^+ \rightarrow K^{*+} (\rightarrow K_S^0 \pi^+) \mu^+ \mu^-$. However, a large amount persists. Two dedicated

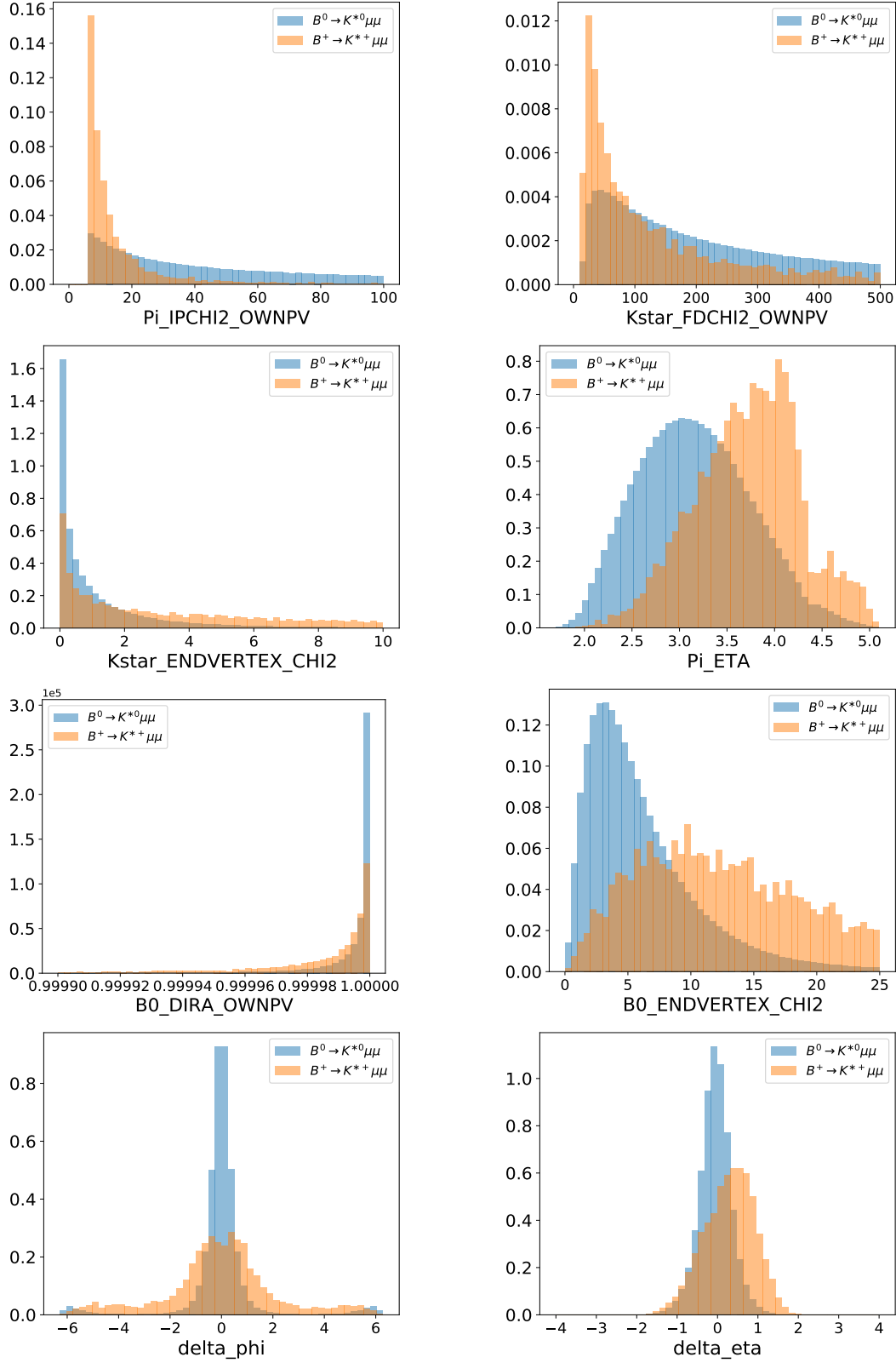
BDTs are hence trained to further suppress such backgrounds, one each for the random pion or kaon.

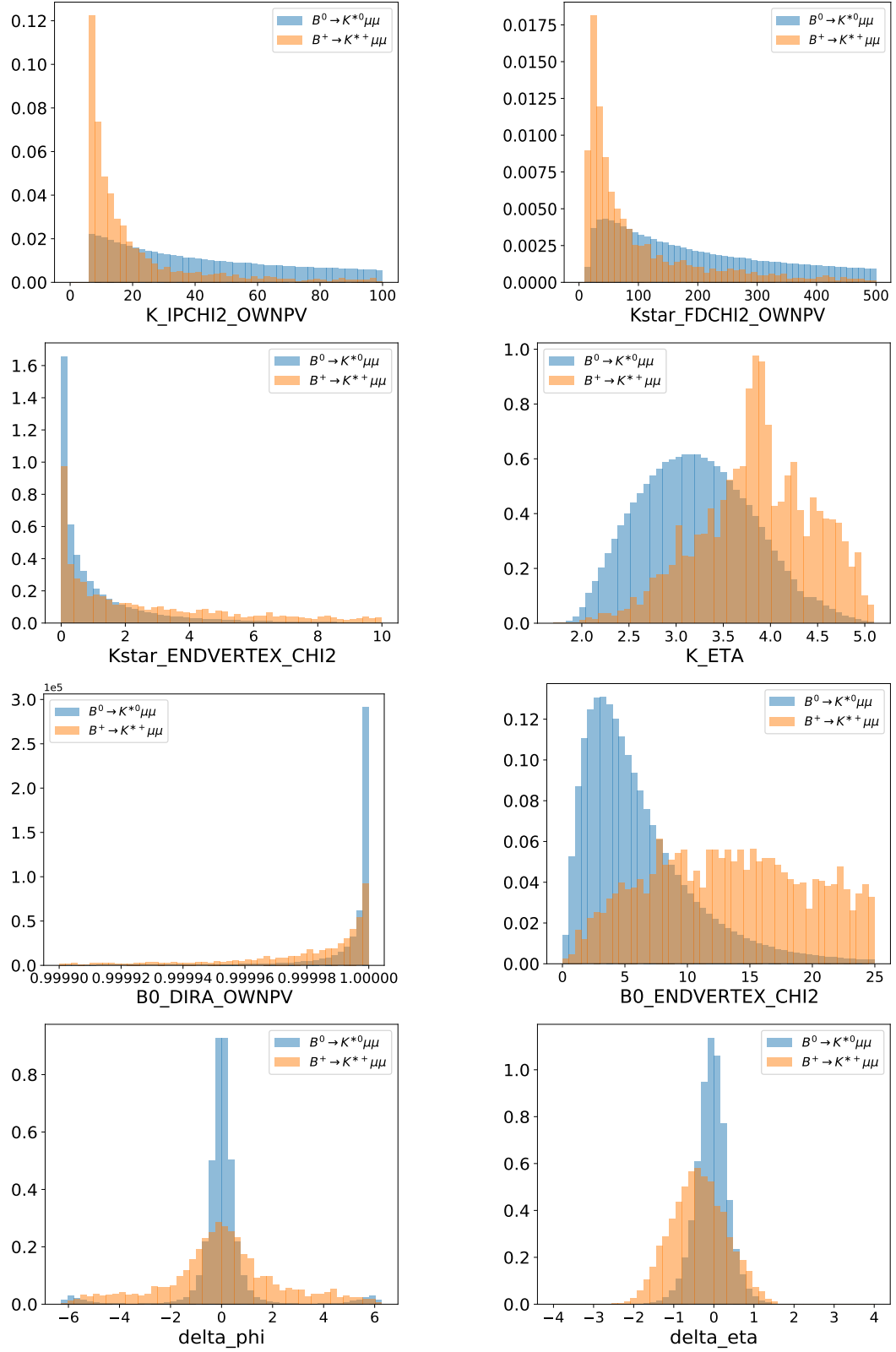
The proxy of the signal combines simulated $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ candidates from all data-taking periods. Again a mixture weight is assigned to individual events by taking the ratio of the MC yields in the J/ψ window over the J/ψ yields in data. This weight along with the MC weights are multiplied to obtain the final weights, which are fed to the BDTs afterwards. For the training of the background, simulated $B^+ \rightarrow K^+ \mu^+ \mu^-$ for all data-taking periods are combined as well. This provides enough background samples, such that no auxiliary simulations are needed, such as the charmonia $B^+ \rightarrow K^+ J/\psi$ simulation. The algorithm used is XGBOOST [131], and the k -folding technique is employed with 10 folds. The training variables are:

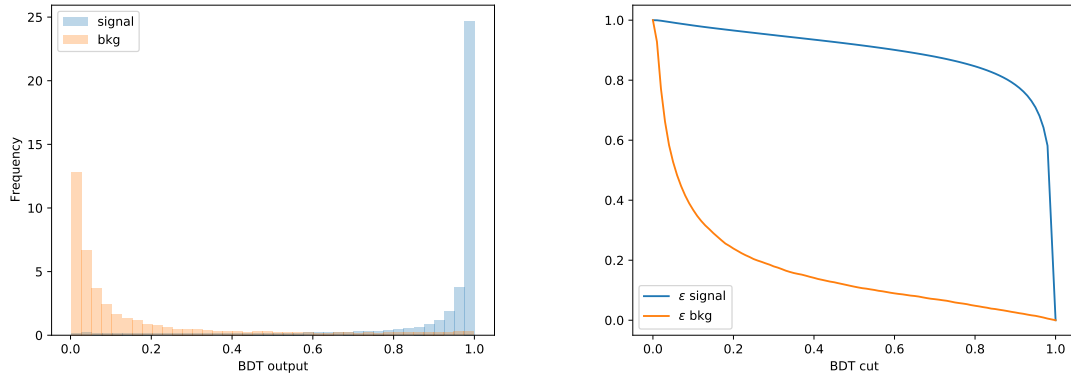
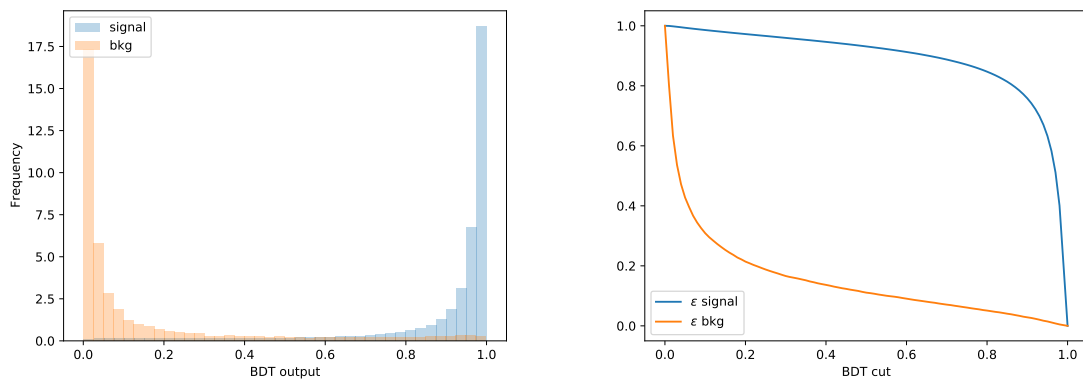
- Random track candidate (kaon or pion) χ_{IP}^2 with respect to the PV
- Track η
- Difference in η between the hadrons
- Difference in ϕ between the hadrons
- K^{*0} candidate χ^2 separation from the PV
- K^{*0} vertex χ^2
- B^0 candidate vertex χ^2
- B^0 DIRA wrt PV

It was found that the χ_{IP}^2 with respect to the PV of the random track candidate provides the biggest discrimination power and the vertex χ^2 of the B^0 candidate comes as the second. All the rest variables are of equal importance.

The distributions of these training variables for the random pions training are shown in Figure 4.16, for the random kaons in Figure 4.17. The results of the training for random pions BDT are shown in Figure 4.18 and random kaons in Figure 4.19.


 Figure 4.16: Distributions of the variables to train the $K^{*+} \rightarrow K^+ \pi^0$ BDT.


 Figure 4.17: Distributions of the variables to train the $K^{*+} \rightarrow K_S^0 \pi^+$ BDT.


 Figure 4.18: BDT output and ROC curve for the $K^{*+} \rightarrow K^+ \pi^0$ BDT.

 Figure 4.19: BDT output and ROC curve for the $K^{*+} \rightarrow K_S^0 \pi^+$ BDT.

4.8 Conclusion

To conclude, the expected peaking background yields with respect to the expected signal yields are shown in Figure 4.20 - 4.23 for different years. The two bins in the plot represents two q^2 of interest in this analysis. It can be seen that the selection performs well. There is no bin that has more than 1.4% peaking backgrounds, with individual peaking backgrounds expected to be less than 0.4%.

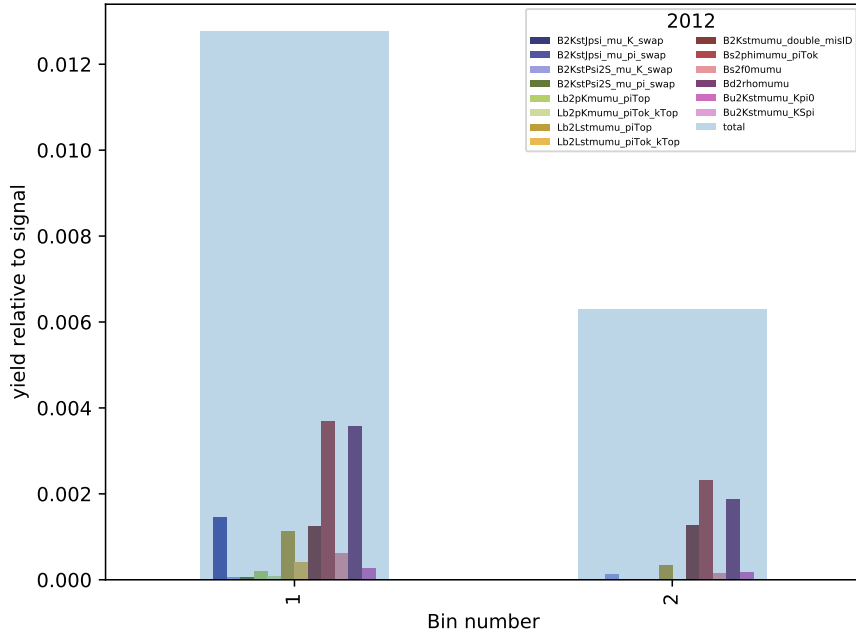


Figure 4.20: The expected yields of different peaking backgrounds with respect to the signal, shown in two q^2 regions of interest for data taken in 2012.

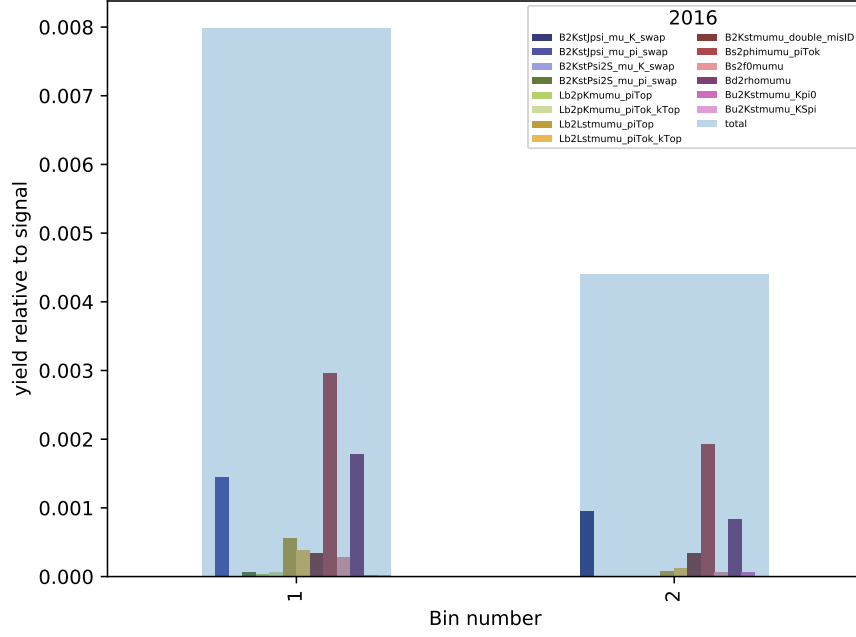


Figure 4.21: The expected yields of different peaking backgrounds with respect to the signal, shown in two q^2 regions of interest for data taken in 2016.

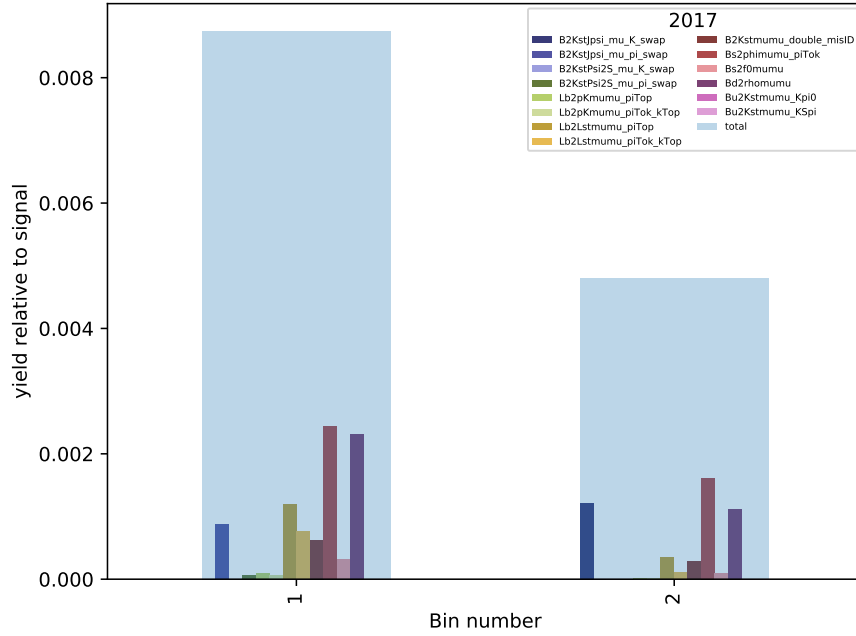


Figure 4.22: The expected yields of different peaking backgrounds with respect to the signal, shown in two q^2 regions of interest for data taken in 2017.

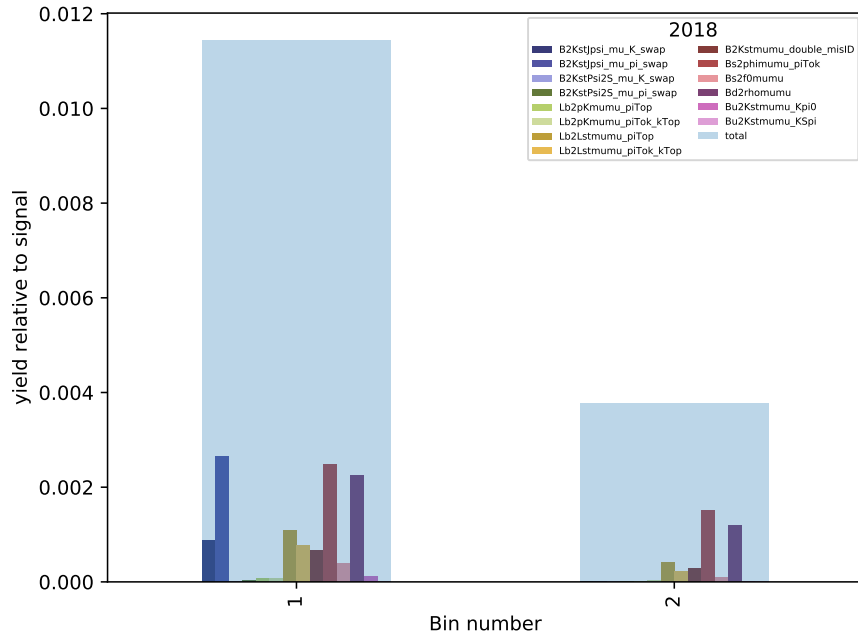


Figure 4.23: The expected yields of different peaking backgrounds with respect to the signal, shown in two q^2 regions of interest for data taken in 2018.

Chapter 5

Detector Efficiency Effects

The reconstruction and selection of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ signal candidates are not uniform across the angular and q^2 distributions, resulting in the so-called acceptance effect. This effect is accounted for using simulated samples. The identical selection procedure as in data (see Chapter 4) can thus be applied to the simulation.

The simulation is imperfect and has to be corrected to match the data. This procedure is introduced in Section 5.1. An acceptance function is then parameterised using the method of moments, where the angular and q^2 distribution are described using Legendre polynomials. The detailed form of the acceptance function is introduced in Section 5.2. Finally, the combined Run 1 and Run 2 acceptance function used in the fit of this analysis is shown in Section 5.3, largely following the method in the ongoing binned analysis of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays at the LHCb collaboration [122].

5.1 Correction for the differences between data and simulation

There are four run periods taken into consideration in this analysis: Run 1, 2016, 2017 and 2018. Large amounts of simulated events are generated for each run period. For Run 1, the 2011 and 2012 data samples are combined and compared with the 2012 simulation, such that the 2012 simulation is corrected to represent the whole of Run 1 alone.

Simulated events are known to have a lower particle multiplicity compared to data, which will affect the extraction of the PID performance that is known to depend on the event multiplicity. Therefore, the event multiplicities are corrected with a scaling factor based on the `nTracks` variable, the number of tracks associated with a certain event, to better approximate the multiplicity of data samples. The correction is achieved by multiplying the `nTracks` variable of each individual event with the scaling factor. An example that presents the distribution of `nTracks` before and after the application of the scaling factor for the 2016 data is shown in Figure 5.1. The applied scale factors for each run period are summarised in Table 5.1.

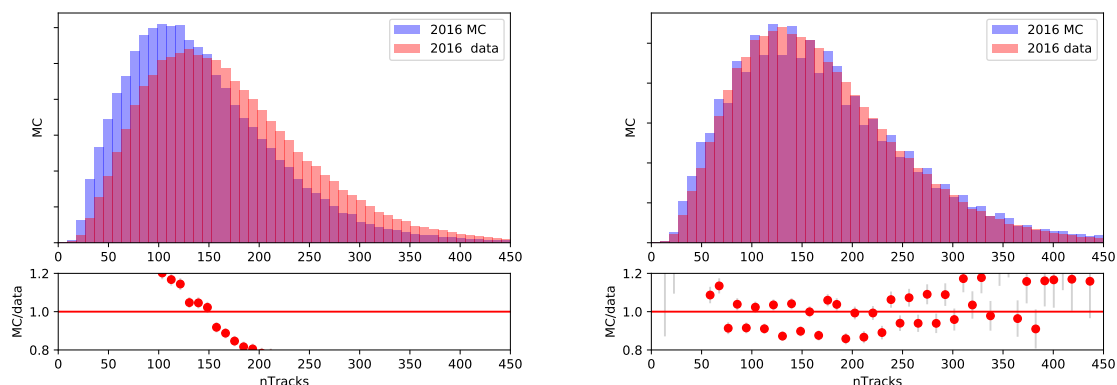


Figure 5.1: Distribution of `nTracks` in the 2016 data and simulation before (left) and after (right) the correction scaling. Figure from [122].

Table 5.1: The applied `nTracks` scaling factors for each run period.

Run period	Scale factor
Run 1	1.008
2016	1.196
2017	1.323
2018	1.366

5.1.1 PID

The PID variables in the simulation are corrected for differences in data. Among all of the PID variables, `isMuon`, which gives the PID requirements on the muon, is taken directly from the simulation, whereas other PID variables are ascertained using a PID sampling tool `PIDGEN`. There is the possibility that a simulated track falls beyond the phase space covered by the calibration sample. Under such circumstances, `PIDGEN` applies a correction such that the PID variables are resampled from a point close to the edge of the calibration sample instead of sampling from a uniform distribution. Further detail of the `PIDGEN` algorithm can be found in Ref. [137].

5.1.2 Tracking efficiency

The tracking efficiency for the simulation can be conveniently corrected using the standard tables provided by the LHCb tracking group [138–140]. The 2012 table is used for the Run 1 simulation, the 2016 table for the 2016 simulation and the 2018 table for both the 2017 and 2018 simulations due to the absence of the 2017 table. These tables are parameterised by the variables p_T and η , but they only cover a limited range which does not cover the full simulated samples. Therefore, if a track falls outside of the range of the table the correction weight is assigned to be that of the bin at the closest edge.

5.1.3 Trigger and B^0 kinematic corrections

The development of L0 efficiency and B^0 kinematics corrections require control mode data samples. An obvious choice of the control mode is $B^0 \rightarrow J/\psi K^{*0}$. However, this mode suffers from the S-wave components $B^0 \rightarrow J/\psi K^- \pi^+$ along with numerous exotic states, such as $B^0 \rightarrow Z (\rightarrow J/\psi \pi) K$. Though attempts have been made to model these contributions, large uncertainties persist in the physics model which may bias the derived corrections. Moreover, several partially reconstructed backgrounds, including those involving B_s^0 decays, can contribute to this mode.

Hence the primary control mode of this analysis is chosen to be $B^+ \rightarrow J/\psi K^+$. This mode has no S-wave contribution and no partially reconstructed backgrounds. It is also independent from the data samples from which physical results are derived. The channel $B^0 \rightarrow J/\psi K^- \pi^+$ is then kept as an alternative for the assessment of a systematic uncertainty on the corrections to the simulation.

The $B^+ \rightarrow J/\psi K^+$ data is selected from the B2XMuMu stripping line, the same as the signal samples. The dimuon invariant mass is required to be in the range $3050 < m(\mu^+ \mu^-) < 3150 \text{ MeV}$ to select the J/ψ and the B^+ mass is selected to be within $5170 < (B^+) < 5450 \text{ MeV}$. The same trigger requirements on HLT1 and HLT2 as the signal samples are used. For each of the periods (Run 1, 2016, 2017, 2018), the sample is split into two subsamples by odd or even event number. One subsample has the L0 requirements (L0Muon or L0DiMuon) applied to derive the kinematic corrections on the B^0 , while the other does not have such requirements to ascertain the L0 efficiency. In both subsamples, the backgrounds are removed using the *sPlot* technique [141], where the *sWeights* are determined by fits in B^+ invariant mass.

5.1.3.1 Trigger treatment and L0 efficiency corrections

The HLT1 and HLT2 selections are taken directly from the simulation, with the exception of the L0 efficiency for the muon triggers. For a given trigger configuration key (TCK), the L0 efficiency is well-modelled in the simulated samples. However, it is different in data since different TCKs with varying thresholds were used for different data-taking periods. Therefore, the primary source to derive the L0 efficiency for the muon triggers for a given TCK is data. The half of the *sWeighted* $B^+ \rightarrow J/\psi K^+$ data without the L0 selection is used to estimate the efficiency of the L0 selection using the TISTOS method [142]. The efficiency is parameterised by the transverse momentum p_T of the two decay muons, which is shown in Figure 5.2. These efficiencies are applied directly to the simulated samples.

5.1.3.2 B^0 kinematics

Finally, the simulation is corrected for the mismodelling of the B^0 meson production kinematics. The p and p_T spectra of the B^0 meson are corrected using the second half of the $B^+ \rightarrow J/\psi K^+$

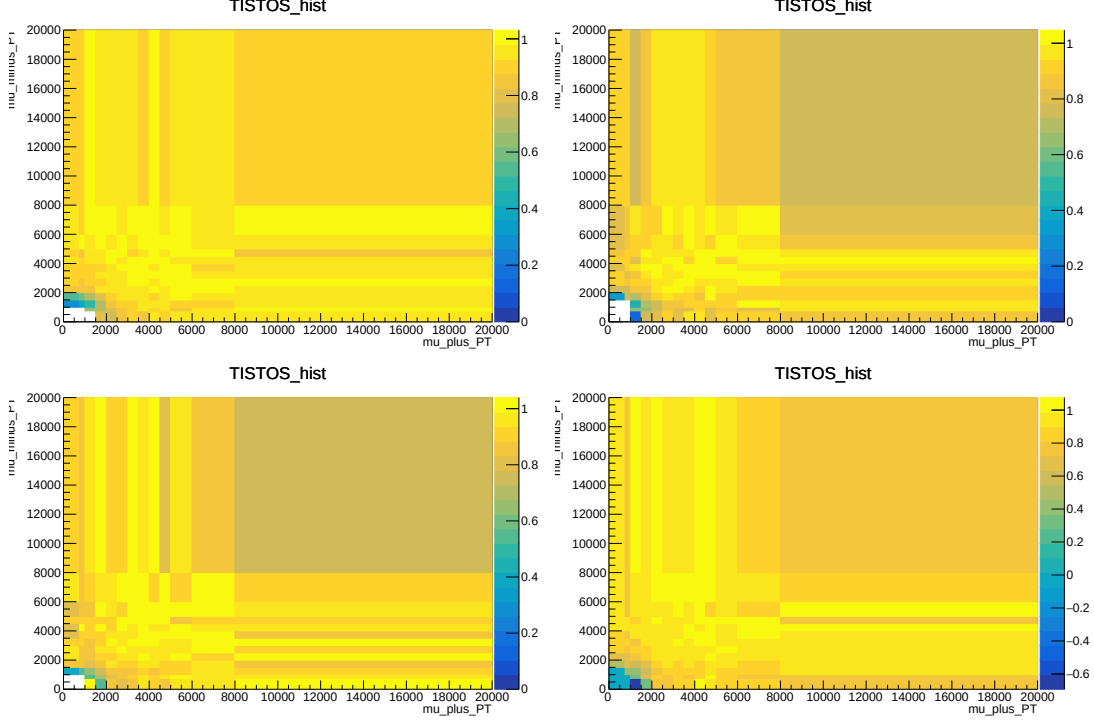


Figure 5.2: The L0 efficiency estimated using the TISTOS method for the p_T of the two muons, where the x axis denotes $p_T^{\mu^+}$ and y axis denotes $p_T^{\mu^-}$. Four data-taking periods are presented: Run 1 (top left), 2016 (top right), 2017 (bottom left) and 2018 (bottom right). The Run 1 histogram is a combination of the 2011 and 2012 samples.

data samples that have TISTOS derived L0 efficiency applied. The kinematic correction is derived using the *gbreweighter* BDT algorithm [143]. Four BDT reweighters are trained separately for each data-taking period. Each training sample is randomly split into half, with one training set to train the reweighters and one test set to check their performance. The p and p_T distributions of the B^0 in simulated and data samples in compared and shown in Figures 5.3 - 5.6 for each period. It can be seen that the agreement is good between the two samples.

5.2 Parametrisation of the acceptance function

The acceptance for the angular and q^2 distributions is modelled as a four-dimensional polynomial function given by

$$(5.1) \quad \varepsilon(\cos \theta_\ell, \cos \theta_K, \phi, q^2) = \sum_{hijk} c_{hijk} L_h(q^2) L_i(\cos \theta_\ell) L_j(\cos \theta_K) L_k(\phi),$$

where $L_m(x)$ is the Legendre polynomial of order m and q^2 , $\cos \theta_\ell$, $\cos \theta_K$ and ϕ are rescaled to the range $-1 < x < 1$ when evaluating the polynomial. The coefficients c_{hijk} are calculated with simulated signal events. The Legendre polynomials follow the orthogonality condition

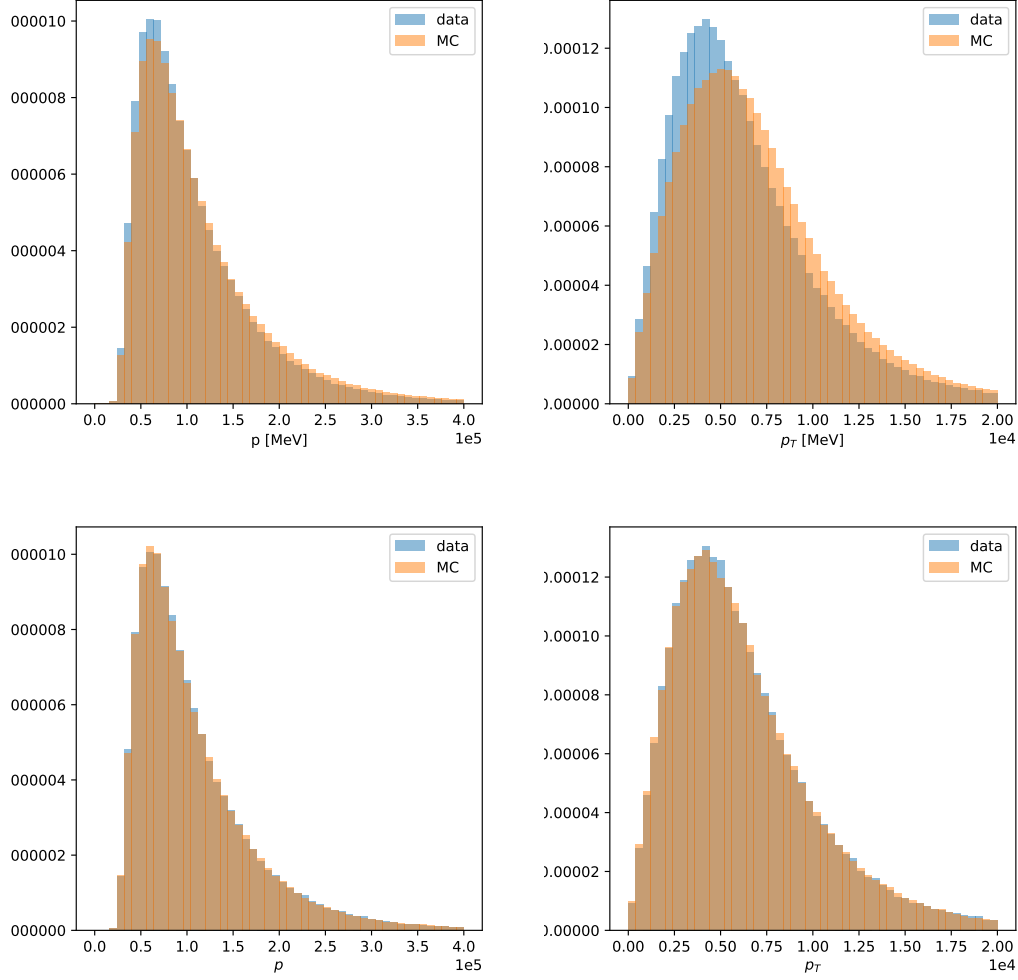


Figure 5.3: The training of the BDT reweighter using $B^+ \rightarrow J/\psi K^+$ as the control mode for the Run 1 data set. The top distributions are before weighting and the lower ones are after weighting.

given by

$$(5.2) \quad \int_{-1}^{+1} L_{i'}(x) L_i(x) dx = \frac{2}{2i+1} \delta_{ii'}.$$

where $\delta_{ii'}$ is the Kronecker delta function. Therefore, the coefficients c_{hijk} can be obtained using Monte Carlo integration. For the sake of clarity, only the derivation of the coefficient c_i for the $\cos \theta_\ell$ variable is illustrated below, but the same method applies to $\cos \theta_K$, ϕ and q^2 as well. If only terms related to $\cos \theta_\ell$ are kept, then Eqn. 5.1 becomes

$$(5.3) \quad \varepsilon(\cos \theta_\ell) = \sum_i c_i L_i(\cos \theta_\ell) .$$

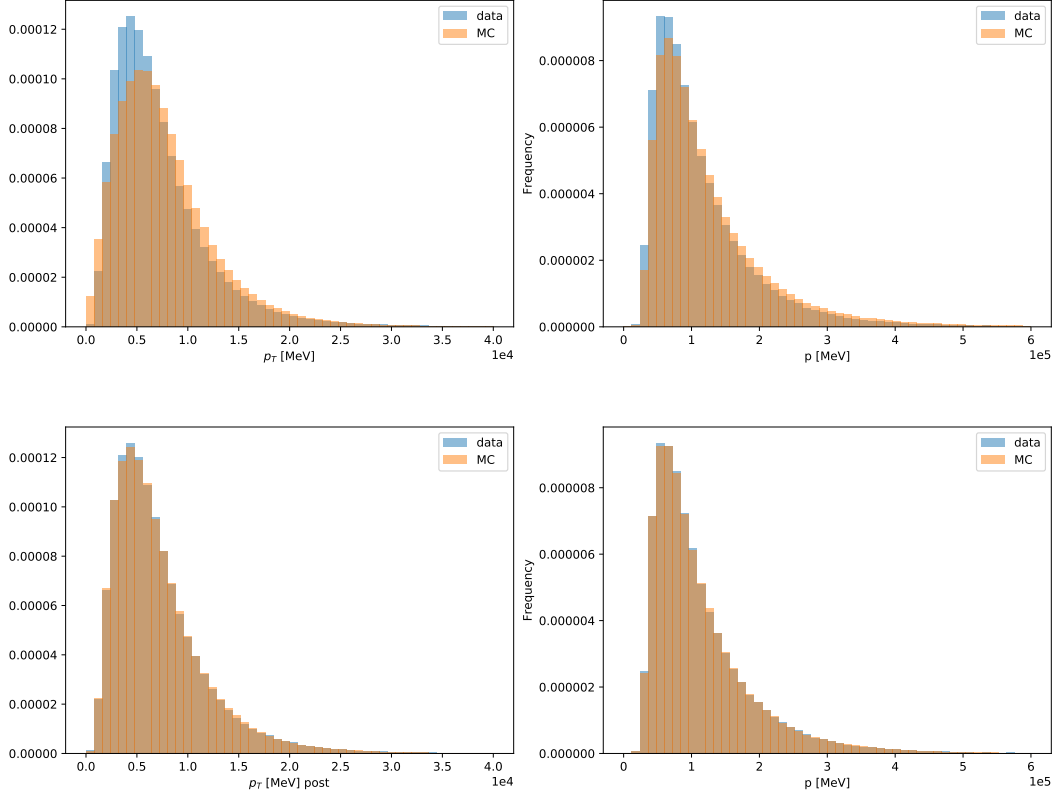


Figure 5.4: The training of the BDT reweighter using $B^+ \rightarrow J/\psi K^+$ as the control mode for the 2016 data set. The top distributions are before weighting and the lower ones are after weighting.

Multiplying Eqn. 5.3 by $L_{i'}(\cos \theta_\ell)$ and integrating over $\cos \theta_\ell$, the above expression becomes

$$(5.4) \quad \int_{-1}^{+1} L_{i'}(\cos \theta_l) \varepsilon(\cos \theta_l) d \cos \theta_l = \sum_i \int_{-1}^{+1} L_{i'}(\cos \theta_l) c_i L_i(\cos \theta_l) d \cos \theta_l .$$

Using the orthogonality condition of Legendre polynomials, one can further get

$$(5.5) \quad \int_{-1}^{+1} L_{i'}(\cos \theta_l) \varepsilon(\cos \theta_l) d \cos \theta_l = \sum_i c_i \frac{2}{2i+1} \delta_{ii'} .$$

Dropping the prime, the coefficients can be derived as

$$(5.6) \quad c_i = \int_{-1}^{+1} \left(\frac{2i+1}{2} \right) L_i(\cos \theta_l) \varepsilon(\cos \theta_l) d \cos \theta_l$$

The Monte Carlo integration approximation is given by

$$(5.7) \quad \varepsilon(\cos \theta_l) \approx \frac{1}{N} \sum_{e=1}^N \delta(\cos \theta_l - \cos \theta_{l,e})$$

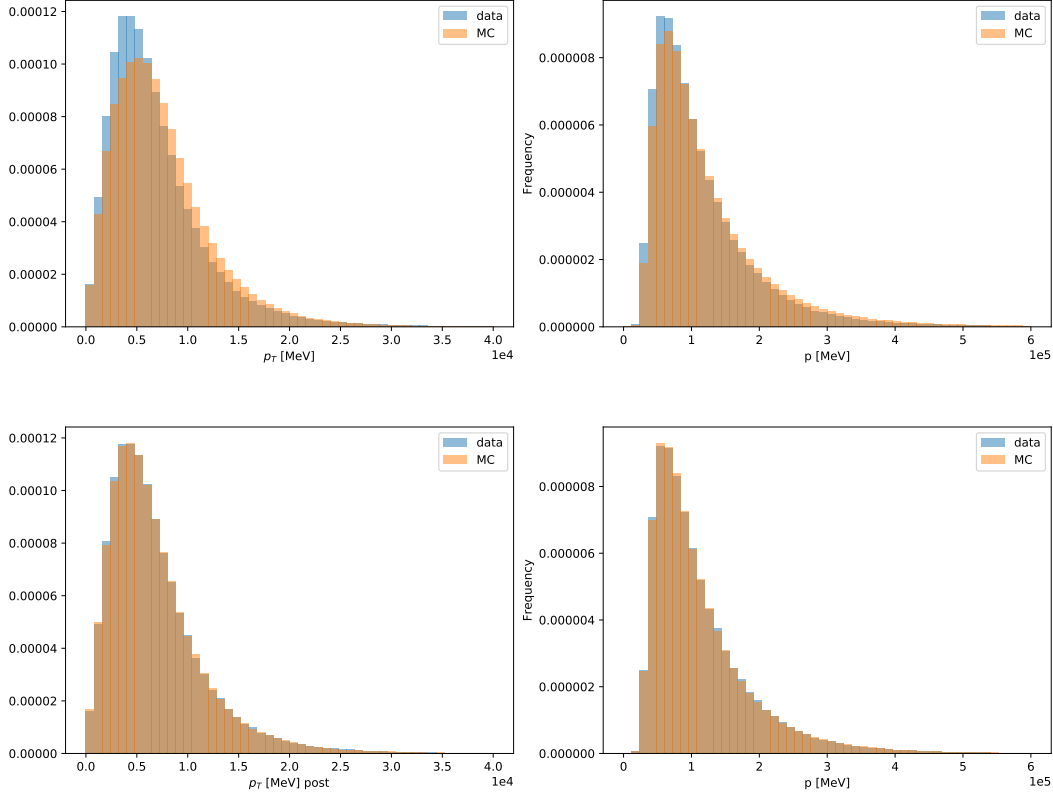


Figure 5.5: The training of the BDT reweighter using $B^+ \rightarrow J/\psi K^+$ as the control mode for the 2017 data set. The top distributions are before weighting and the lower ones are after weighting.

Here N denotes the number of all signal events in the simulation and the index e labels the simulated signal event. Inserting Eqn. 5.7 into Eqn. 5.6 and evaluating the Kronecker delta function, the expression of the coefficients becomes

$$(5.8) \quad c_i = \frac{1}{N} \sum_{e=1}^N \left(\frac{2i+1}{2} \right) L_i(\cos \theta_{l,e})$$

Generalising Eqn. 5.8 to include the $\cos \theta_K$, ϕ and q^2 terms gives the following expression

$$(5.9) \quad c_{hijk} = \frac{1}{N} \sum_{e=1}^N \left(\frac{2h+1}{2} \right) \left(\frac{2i+1}{2} \right) \left(\frac{2j+1}{2} \right) \left(\frac{2k+1}{2} \right) L_h(q_e^2) L_i(\cos \theta_{l,e}) L_j(\cos \theta_{K,e}) L_k(\phi_e) .$$

The events that are used to determine the coefficients c_{hijk} are generated with a FLATQ2 model. Based on the commonly adopted pure PHSP model which considers decays in phase space only, the FLATQ2 model ensures that the distribution of simulated events is flat in q^2 , $\cos \theta_\ell$, $\cos \theta_K$ and ϕ . Therefore, any subsequent sculpting effects in the shape can be conveniently interpreted as the distribution of the efficiency effect. The q^2 distribution of the FLATQ2 model is shown in Figure 5.7.

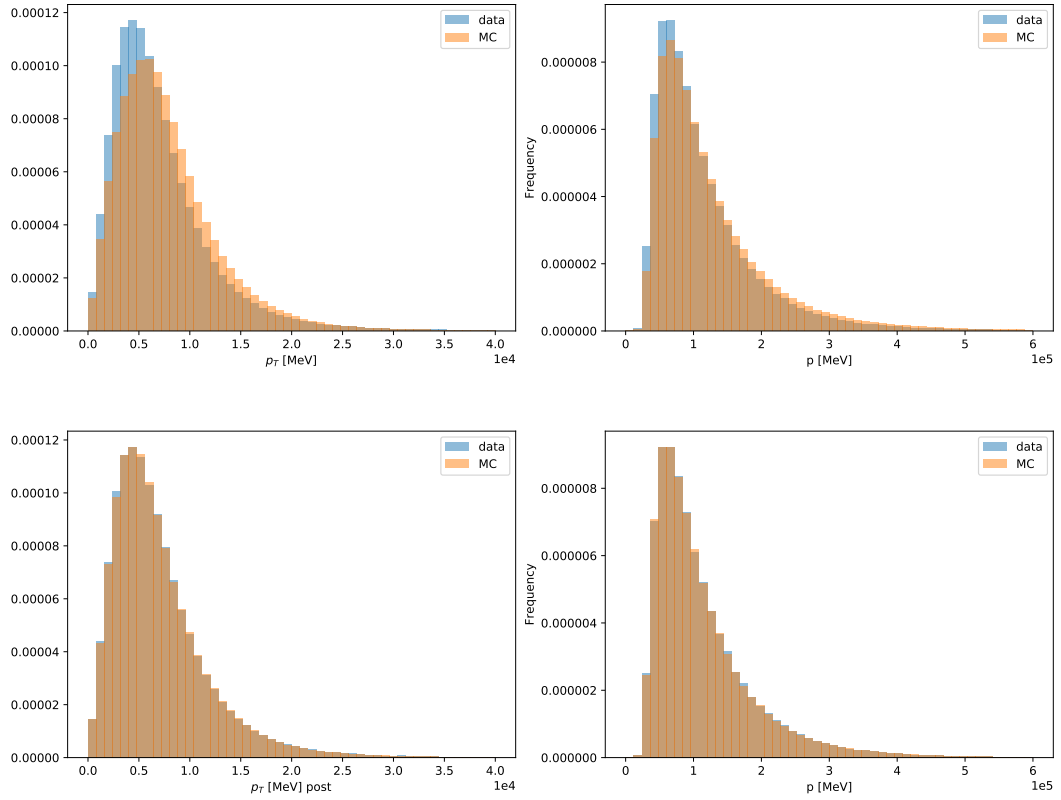


Figure 5.6: The training of the BDT reweighter using $B^+ \rightarrow J/\psi K^+$ as the control mode for the 2018 data set. The top distributions are before weighting and the lower ones are after weighting.

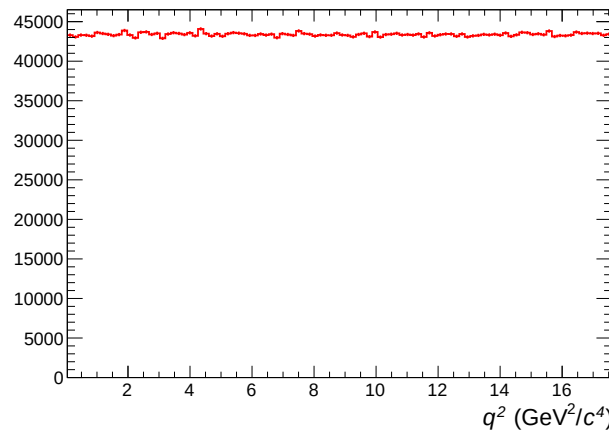


Figure 5.7: The q^2 simulation at generator level using the FLATQ2 model.

5.3 The combined Run 1 and Run 2 acceptance function

The orders of the Legendre polynomials in the acceptance function 5.1 are determined by fitting the acceptance function to simulated samples with the efficiency effects. They are chosen to be the lowest orders that make the acceptance function sufficient to describe the simulation accurately. The choice of the orders is further supported by a goodness-of-fit test, which is outlined in Appendix B.

The acceptance function is initially evaluated for each data-taking period (Run1, 2016, 2017, 2018) separately. Because this analysis combines the Run 1 and Run 2 data samples, the combined Run 1 + Run 2 acceptance needs to be computed. For each data period, the simulation used to compute the acceptance functions is normalised to the amount of simulation in the q^2 window corresponding to $B^0 \rightarrow J/\psi K^{*0}$ decays and then scaled to the $B^0 \rightarrow J/\psi K^{*0}$ yield obtained in the data J/ψ window and scaled according to the J/ψ data fits. The scaling factors are shown in Table 5.2.

Table 5.2: The number of acceptance events in simulation that passed the selection in the J/ψ window, J/ψ signal data yields, and weights used for scaling for each run period when computing the combined acceptance.

Year	Number of acceptance events	J/ψ data signal yield	Weight
Run 1	295385	303920	1.029
2016	269830	313483	1.162
2017	259796	336601	1.296
2018	249992	414027	1.656

For the combined Run 1 and Run 2 acceptance, Legendre polynomials of the seventh order are chosen for q^2 , $\cos\theta_\ell$ and ϕ , while the tenth order is chosen for $\cos\theta_K$, leading to 3430 coefficients in total. The one-dimensional projections of the combined acceptance function on q^2 and the three angles are shown in Figure 5.8. It can be seen that the acceptance function is in good agreement with the simulated samples in all four dimensions.

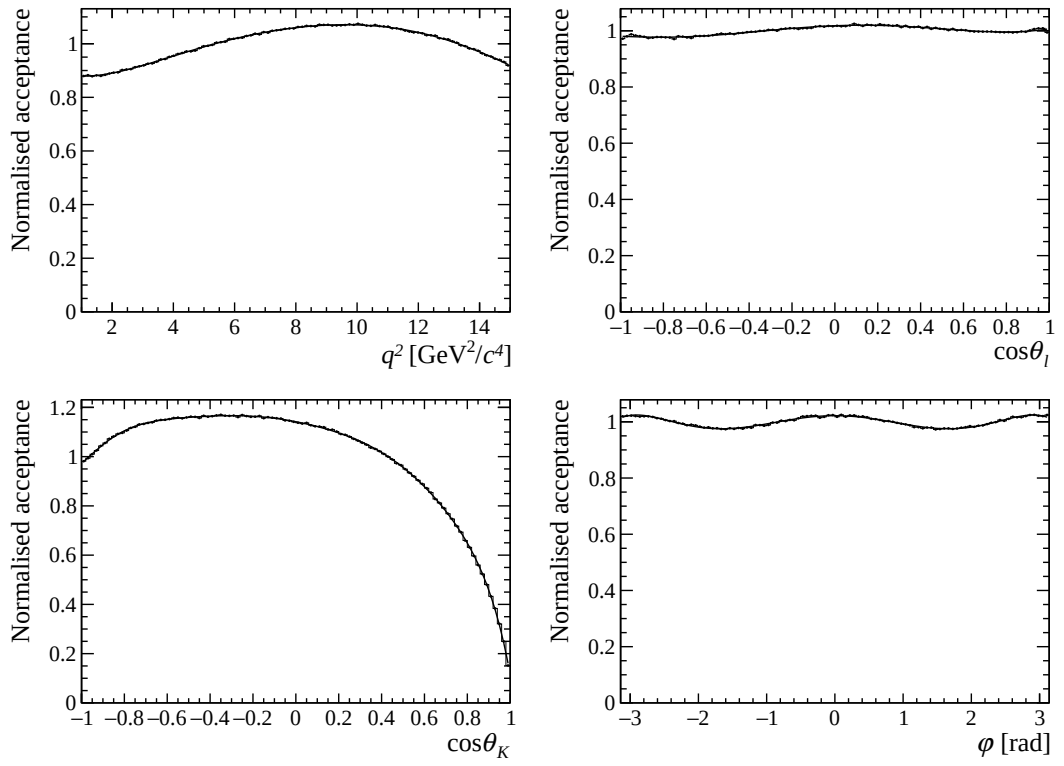


Figure 5.8: Projections of the combined B^0 and $\bar{B}^0$ acceptance in q^2 , $\cos\theta_\ell$, $\cos\theta_K$, and ϕ .

Chapter 6

Fitter Setup

Fitting for the amplitudes allows the fit to be performed in an unbinned setup. In binned analyses [21, 37, 122], the PDF is integrated over bins in q^2 which results in loss of information. Moreover, these analyses fit for angular observables which do not guarantee a physical PDF. For example, since there are no constraints on the angular observables, one can have one of the $\mathcal{J}$ terms, e.g. $\mathcal{J}_{1s}$, set to -1 and the rest $\mathcal{J}$ terms set to 0, which results in a negative PDF. Fitting for amplitudes, on the contrary, prevents the PDF from entering the nonphysical regions. An unbinned fit also ensures that the information in the q^2 space can be fully exploited and the precision of the measurement can be increased. Moreover, with the amplitudes described by the polynomial ansatz method, the model dependency of the theory is removed as much as possible¹. By describing the full LHCb Run 1 + 2 data with the amplitude ansatz method, synthetic model-independent data samples that are background subtracted and efficiency corrected can be generated. Theorists can use these synthetic data samples to fit any model they choose.

The analysis framework was constructed in TensorFlow motivated by the convenient utilisation of GPUs and is further discussed in Section 6.1. The PDF of the model is explicitly written out in Section 6.2 and Section 6.3, and the extended log likelihood function used in the fit is constructed in Section 6.4. The fit strategy is then summarised in Section 6.5. This validation of the fitter is done through studies with toy studies, as outlined in Section 6.6. The data from the control mode $B^0 \rightarrow J/\psi K^{*0}$ is also fit for validation purposes and is described in detail in Section 6.7.

6.1 The TensorFlow analysis framework

This analysis is implemented in a TensorFlow (TF) fit framework. TF is an open-source library developed by Google [144]. Primarily designed for machine learning purposes, TF has been

¹The model of the S-wave components as described in 2.3.2.1 is the only model dependency that has to be kept in this analysis. The amplitude ansatz method is not able to model the S-wave components precisely with the current statistics of the experimental data samples available.

widely adopted by both academia and industry supported by an active community. Essential features of TF include:

- When one defines operations, a computation graph is created. TF performs optimisation of the graph before it is executed, such as identifying operations that could be redone or results that could be taken from the cache, etc.
- TF calculates the derivatives of the calculated functions analytically, such that the minimisation is sped up.
- TF can run on various architectures, including multi-thread CPU and NVIDIA GPU.
- TF optimises the code on hardware backstage at the machine level. No prior knowledge is needed to run the code parallel on the CPU or GPU. One can thus focus on the model construction rather than the technical detail.

These features are important for maximum likelihood fits. However, TF alone is not the ideal optimiser. Although it has its own minimisation functions, they are optimised for machine learning but not for physics. For example, errors or likelihood scans are not provided when the minimisation is done. This is why Minuit [145], a longstanding and well-proven algorithm widely adopted in particle physics, has to be run as the minimiser. They are combined in the TensorFlow Analysis package [146]. In this analysis, the model is first defined in TF and is optimised for the hardware by TF when compiled. The likelihood function can then be constructed and given to Minuit to be minimised while TF is responsible for the analytical gradient calculation.

6.2 The signal model

6.2.1 Angular PDF

The signal PDF is the differential decay rate as defined in 2.3 with the detector efficiency effects applied, namely

$$(6.1) \quad \mathcal{P}_{\text{Sig}}(q^2, \cos \theta_\ell, \cos \theta_K, \phi) = \frac{d^4\Gamma + d^4\bar{\Gamma}}{dq^2 d\cos \theta_\ell d\cos \theta_K d\phi} \times \varepsilon(\cos \theta_\ell, \cos \theta_K, \phi, q^2) .$$

Here, the differential decay rate takes the form written in Equation 2.85:

$$\begin{aligned}
& \frac{d^4\Gamma + d^4\bar{\Gamma}}{dq^2 d\cos\theta_\ell d\cos\theta_K d\phi} \\
&= \frac{9}{32\pi} \{ [\mathcal{J}_{1s} \sin^2\theta_K + \mathcal{J}_{1c} \cos^2\theta_K + \mathcal{J}_{2s} \sin^2\theta_K \cos 2\theta_\ell + \mathcal{J}_{2c} \cos^2\theta_K \cos 2\theta_\ell \\
&\quad + \mathcal{J}_3 \sin^2\theta_K \sin^2\theta_\ell \cos 2\phi + \mathcal{J}_4 \sin 2\theta_K \sin 2\theta_\ell \cos \phi + \mathcal{J}_5 \sin 2\theta_K \sin \theta_\ell \cos \phi \\
&\quad + \mathcal{J}_{6s} \sin^2\theta_K \cos \theta_\ell + \mathcal{J}_7 \sin 2\theta_K \sin \theta_\ell \sin \phi + \mathcal{J}_8 \sin 2\theta_K \sin 2\theta_\ell \sin \phi \\
&\quad + \mathcal{J}_9 \sin^2\theta_K \sin^2\theta_\ell \sin 2\phi + \mathcal{J}'_{1c} (1 - \cos 2\theta_\ell)] \\
&\quad + [\mathcal{J}''_{1c} \cos \theta_K (1 - \cos 2\theta_\ell) + \mathcal{J}'_4 \sin 2\theta_\ell \sin \theta_K \cos \phi + \mathcal{J}'_5 \sin \theta_\ell \sin \theta_K \cos \phi \\
&\quad + \mathcal{J}'_7 \sin \theta_\ell \sin \theta_K \sin \phi + \mathcal{J}'_8 \sin 2\theta_\ell \sin \theta_K \sin \phi] \int_{0.796}^{0.996} \mathcal{A}_{\text{LASS}}(m_{K\pi}, q^2) \mathcal{A}_{\text{BW}}^*(m_{K\pi}, q^2) dm_{K\pi} \} ,
\end{aligned}$$

where the $\mathcal{J}$ -terms are bilinear combination of the helicity amplitudes as explicitly defined in Equations 2.68 and 2.75. Since this analysis is performed in the CP-averaged configuration, the $\mathcal{J}$ -terms in this PDF denote the averaged B^0 and $\bar{B}^0$ observables. The ε in Equation 6.1 represents the acceptance function as defined in Equation 5.1:

$$\varepsilon(q^2, \cos\theta_\ell, \cos\theta_K, \phi) = \sum_{hijk} c_{hijk} L_h(q^2) L_i(\cos\theta_\ell) L_j(\cos\theta_K) L_k(\phi) .$$

A discrete symmetry exists because the angular PDF consists of bilinear combinations of the helicity amplitudes. All the left-handed amplitudes and/or the right-handed amplitudes can be rotated by an overall phase of π , resulting in four exact global solutions of the PDF, namely $\pm\mathcal{A}_L$ and $\pm\mathcal{A}_R$. For example, in the term $\mathcal{J}_4$ written as

$$\mathcal{J}_4 = \frac{\beta_\mu^2}{\sqrt{2}} \left[\text{Re} \left(\mathcal{A}_0^L \mathcal{A}_\parallel^{L*} \right) + (L \rightarrow R) \right]$$

as shown in Equation 2.68, the transformation $\mathcal{A}_0^L \mapsto -\mathcal{A}_0^L$ and $\mathcal{A}_\parallel^L \mapsto -\mathcal{A}_\parallel^L$ leaves $\mathcal{J}_4$ invariant, hence the full PDF. This choice of global solution is arbitrary, and the chosen solution is extracted by performing many fits and mapping out the solution of interest in this analysis.

6.2.2 B mass lineshape

The $m_{K\pi\mu\mu}$ variable is factorised out separately. The lineshape of the B^0 mass is completely dependent on the detector resolution and is described by the combination of two Crystal Ball (CB) functions, whose mathematical formula is explicitly given in Appendix A. The tails of the two CB functions are on opposite sides. The B mass signal parameterisation is therefore

$$(6.2) \quad \mathcal{P}_{\text{sig}}(m_{K\pi\mu\mu}) = f_{\text{core}} P_{CB}(\mu, \sigma_1, \alpha_1, n_1) + (1 - f_{\text{core}}) P_{CB}(\mu, \sigma_2, \alpha_2, n_2) ,$$

where P_{CB} is the CB function and f_{core} is the fraction of one CB function. The tail parameters of the CB, namely α and n , are determined by means of simulation. Both parameters are fixed in rare mode fits. For the control mode fit, the α parameters are floated for better fit quality

while the n parameters are fixed due to high correlations with α . The mean, width and core fraction parameters are floated in the fit.

Therefore, the full signal PDF can be written as

$$(6.3) \quad \mathcal{P}_{\text{Sig}, B^0} = \mathcal{P}_{\text{Sig}}(q^2, \cos \theta_\ell, \cos \theta_K, \phi) \times \mathcal{P}_{\text{Sig}}(m_{K\pi\mu\mu}) .$$

6.2.3 Treatment of the B_s^0

As a result of the selection described in 4, the contributions from the $B_s^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$ channel become significant in data and can no longer be absorbed by the combinatorial background. The signal PDF hence needs to be modified to account for effects from the B_s^0 channel.

The angular and q^2 PDF of B_s^0 can be approximated by that of $\bar{B}^0$ under the assumption of minimal flavour violation [147]. As mentioned in Section 2.3, flipping the signs of $\mathcal{J}_{7,8,9}$ and $\mathcal{J}'_{7,8}$ in the angular and q^2 PDF of $\bar{B}^0$ can well approximate that of B_s^0 [40]. The mass model of B_s^0 is the same as that of B^0 , apart from the mean being shifted by the mass difference $\Delta m \equiv m_{B_s^0} - m_{B^0}$. The value of the mass shift is taken from the PDG [148], with the full B_s^0 PDF denoted by $\mathcal{P}_{\text{Sig}, B_s^0}$, the full signal PDF is modified to be

$$(6.4) \quad \mathcal{P}_{\text{Sig}} = (1 - f_{B_s^0}) \mathcal{P}_{\text{Sig}, \bar{B}^0} + f_{B_s^0} \mathcal{P}_{\text{Sig}, B_s^0}$$

where

$$f_{B_s^0} = \frac{N_{\text{Sig}, B_s^0}}{N_{\text{Sig}, B^0}} = \frac{\mathcal{B}(B_s^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-) f_s}{\mathcal{B}(B^0 \rightarrow K^{*0} \mu^+ \mu^-) f_d}$$

with f_s and f_d as the B_s^0 and B^0 fragmentation fractions. This ratio is floated in the control mode fit for the sake of better fit quality. In the rare mode fit, this ratio is fixed to previous measurements recorded in the PDG [148]. It should be noted that all amplitude parameters and mass parameters are shared between both signal PDFs, $\mathcal{P}_{\text{Sig}, \bar{B}^0}$ and $\mathcal{P}_{\text{Sig}, B_s^0}$, in the fit. This approximation is considered to be acceptable since the B_s^0 contributions are small compared to those from the main $\bar{B}^0$ process.

6.3 The background model

The background PDF is written as

$$(6.5) \quad \mathcal{P}_{\text{Bkg}}(q^2, \cos \theta_\ell, \cos \theta_K, \phi, m_{B^0}) = \mathcal{P}(q^2) \cdot \mathcal{P}(\cos \theta_\ell) \cdot \mathcal{P}(\cos \theta_K) \cdot \mathcal{P}(\phi) \cdot \mathcal{P}(m_{B^0})$$

The angular and q^2 background PDF is modelled as

$$(6.6) \quad \mathcal{P}(k) = P_0(k) + \alpha_1^k P_1(k) + \alpha_2^k P_2(k), \quad k \in \{q^2, \cos \theta_\ell, \cos \theta_K, \phi\}$$

where $P_i(k)$ represents Chebyshev Polynomials of order i and α_i^k is the background parameter fitted for. The m_{B^0} background, on the other hand, is described by an exponential function

$$(6.7) \quad \mathcal{P}(m_{B^0}) = e^{-\tau_m m_{B^0}}$$

There are therefore nine parameters in total describing the background PDF. As can be seen from Equation 6.5, the background is assumed to factorise among individual dimensions. This is a reasonable assumption since most of the non-factorisable background events can be removed by the selection developed in Chapter 4.

6.4 Extended log likelihood fit

The PDF of the signal model is not sensitive to global normalisation by construction. If all polynomial amplitudes were multiplied with a constant, this constant would cancel out when the signal PDF is normalised. Therefore, it is possible that the amplitude coefficients vary arbitrarily among different scaling during the fit, making the fit rarely converge. An overall constraint on the normalisation is applied to solve this problem by associating the normalisation of the angular and q^2 four-dimensional signal PDF with the signal yield, introducing an extended term into the likelihood written as

$$N_{\text{Sig}} = \bar{N}_{\text{Sig}} * \frac{\int \mathcal{P}_{\text{Sig}}|_{\text{Fit}}}{\int \mathcal{P}_{\text{Sig}}|_{\text{Init}}},$$

where $\bar{N}_{\text{Sig}}$ is the expected signal yield, $\int \mathcal{P}_{\text{Sig}}|_{\text{Init}}$ stands for the integral of the signal PDF calculated when all fit parameters are set to be the starting value and $\int \mathcal{P}_{\text{Sig}}|_{\text{Fit}}$ stands for the integral of the signal PDF during the fit. However, since this analysis does not aim to measure the branching fraction, the factor $\frac{\bar{N}_{\text{Sig}}}{\int \mathcal{P}_{\text{Sig}}|_{\text{Init}}}$ can be regarded as an overall scaling constant set arbitrarily, denoted as α . The extended term can thus be simplified as

$$(6.8) \quad N_{\text{Sig}} = \alpha \int \mathcal{P}_{\text{Sig}} \Big|_{\text{Fit}}.$$

This modified N_{Sig} can still represent the observed number of signal events since all of the amplitudes in $\mathcal{P}_{\text{Sig}}$ will scale according to α . Connecting the overall scaling of the signal PDF and the signal yield also has the advantage of allowing signal yield and background yield to vary independently. In this analysis, α is set to be 2,000,000 for the control mode fit and 100,000 for the rare mode fits.

Therefore, the full negative log-likelihood function to be minimised in the fit can be written as

$$(6.9) \quad -\log \mathcal{L} = - \sum_{N_{\text{Dat}}} \log (N_{\text{Sig}, B^0} P_{\text{Sig}, B^0} + N_{\text{Sig}, B_s^0} P_{\text{Sig}, B_s^0} + N_{\text{Bkg}} P_{\text{Bkg}}) + (N_{\text{Sig}, B^0} + N_{\text{Sig}, B_s^0} + N_{\text{Bkg}})$$

In this expression, N_{Sig} denotes the signal yield as defined in Equation 6.8 and N_{Bkg} denotes the background yield. P_{Sig} and P_{Bkg} are the signal and background PDFs defined as above respectively.

6.5 Fit strategy

The fitter is set to fit all coefficients in the amplitudes, the m_{B^0} lineshape parameters, the background parameters, and the background yields. The fit returns the fit values of the coefficients and their covariance matrix. The fit strategy is as follows:

1. Fit to rare mode data samples in the lower q^2 region $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$. Here the order of the P-wave amplitude ansatz in Equation 2.89 is set to be 4, denoted as

$$\mathcal{A}_{\parallel,\perp,0}^{L,R} = \alpha_{\parallel,\perp,0}^{L,R} L_0(q^2) + \beta_{\parallel,\perp,0}^{L,R} L_1(q^2) + \gamma_{\parallel,\perp,0}^{L,R} L_2(q^2) + \delta_{\parallel,\perp,0}^{L,R} L_3(q^2) .$$

The choice of the order of the ansatz is further supported by the Goodness-of-Fit test, using the same method described in Appendix B. Here the lower limit is chosen to maximally avoid the influence of the $\phi(2S)$ mode, while the upper limit is where the ansatz can both maintain the validity and refrain from the influence of the J/ψ peak.

2. Fit to data samples in the inter-resonance region $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$. Here the order of the amplitude ansatz 2.89 is set to be 2, written as

$$\mathcal{A}_{\parallel,\perp,0}^{L,R} = \alpha_{\parallel,\perp,0}^{L,R} L_0(q^2) + \beta_{\parallel,\perp,0}^{L,R} L_1(q^2) .$$

It is essentially a linear ansatz as the shapes of the data in this region are reasonably flat.

The B^0 mass window is selected to be $5170 < m_{K\pi\mu\mu} < 5700 \text{ MeV}/c$ in both q^2 regions. Individual fits in each q^2 region are carried out in the full data samples combining different data-taking periods, namely Run 1, 2016, 2017 and 2018. No information is shared between fits in the two regions. The fit results of the amplitude parameters, along with the covariance matrices, are taken to generate pseudoevents based on a multivariate Gaussian distribution. The amplitude functions and the observables are then calculated from these pseudoevents.

6.6 Validation with pseudodata

Pseudodata, or toy data as commonly referred to, are generated and fit back to validate the fitter setup. This procedure is called the toy study in the remaining part of this thesis.

Before generating the toy data samples, a blinded fit to the real Run 1 + 2 data in the corresponding q^2 region is performed. The results of the fit are taken as the *true* value of the toy sample generated, and at the start of each fit the initial values of the parameters are randomised based on a uniform distribution between the lower and upper limit of that parameter. The toy sample is a full representative of the data fit, including the acceptance, the background and the $B^+ \rightarrow K^+ \mu^+ \mu^-$ veto effect. They are therefore emulating real experiments with the *true* values of the parameters and their statistical properties already well-known. The toy fits are performed 500 times in each q^2 region and the ensemble of the toy fit results is studied.

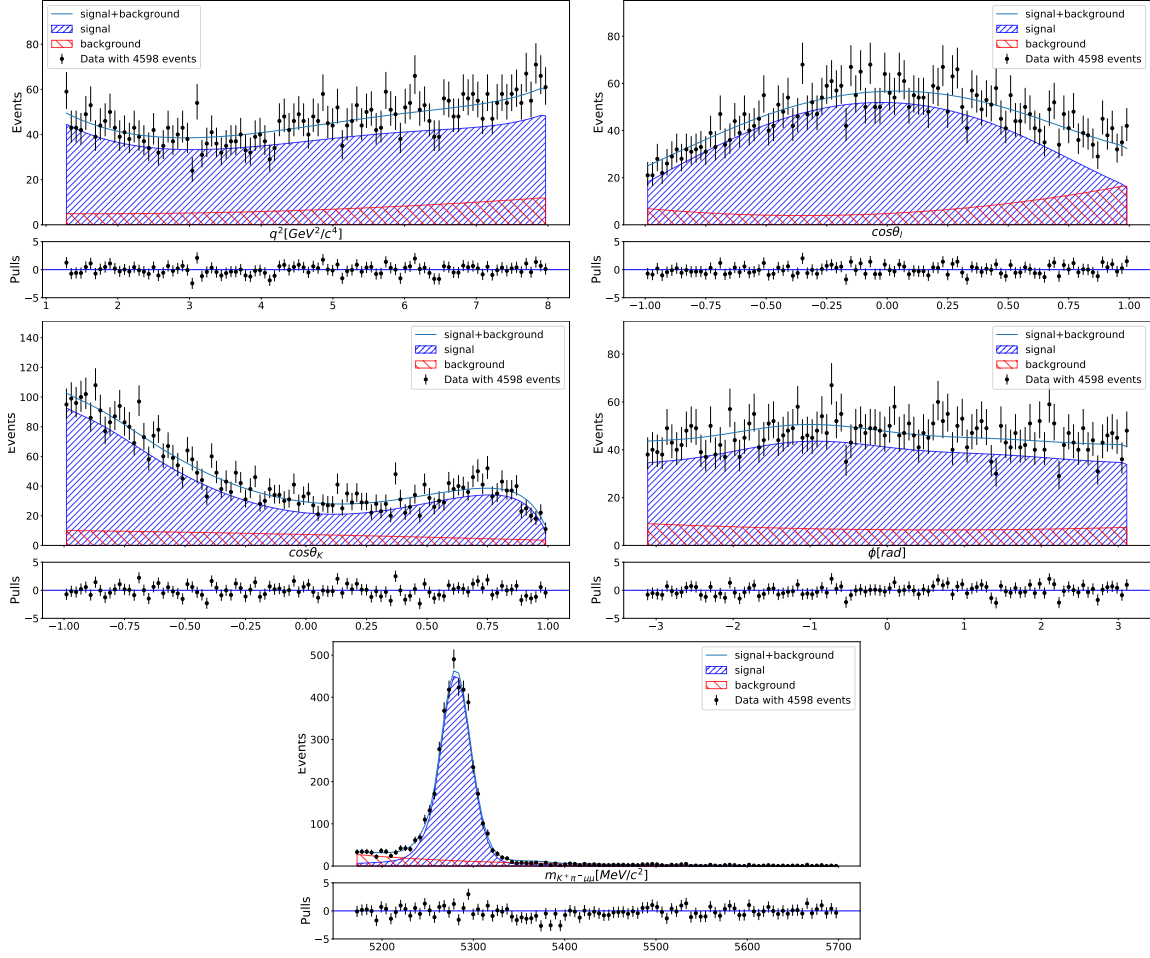


Figure 6.1: Example projections from fitting to a toy generated in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.

6.6.1 Toy study in the region $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$

This section describes the toy studies in the region $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$. The projections of an example toy fit is shown in Figure 6.1. In each plot, the subplot below shows the pull statistic of the histogram. For each bin X , the pull statistic is defined as

$$X_{\text{pull}} = \frac{X_{\text{true}} - X_{\text{fit}}}{X_{\text{error}}}.$$

The pull statistic is constructed to represent the quality of the fit. The smaller the pull statistic is, the closer the value of the PDF predicted by the fit is to the data. Here, it can be seen that the toys are fit back with good quality.

The ensemble of 500 toy fits is studied. The residuals of P-wave amplitude parameters are shown in Figure 6.2. Some biases are spotted in the parameters for the $\mathcal{A}_{\parallel}^R$ amplitude. They are due to physical boundary effects, which is explained in detail in Appendix C.

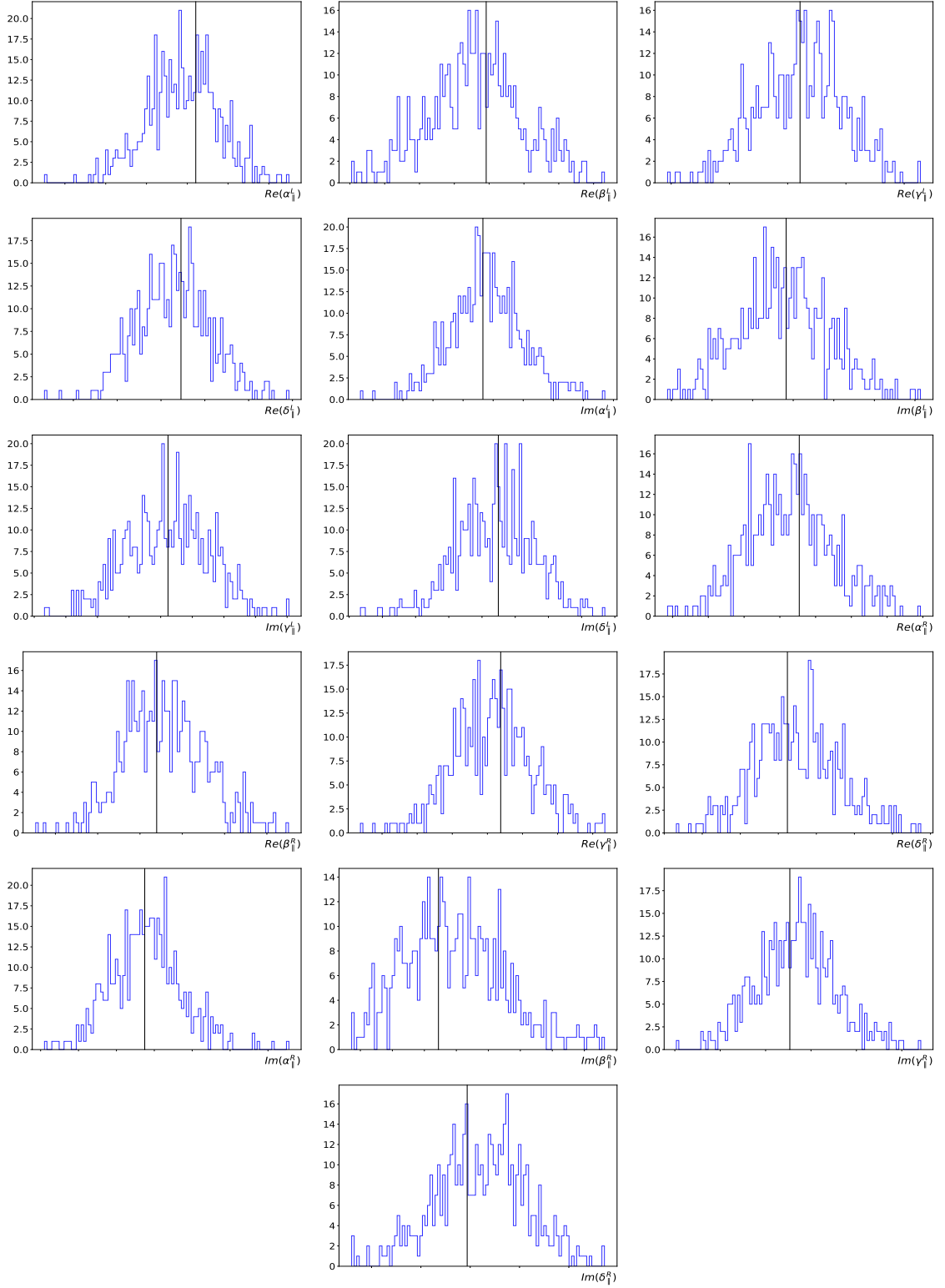


Figure 6.2: Residuals of the P-wave parameters (parallel amplitude components) from ensembles of fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. The true values of the parameters are shown by the black lines. The x axis is currently blinded.

6.6. VALIDATION WITH PSEUDODATA

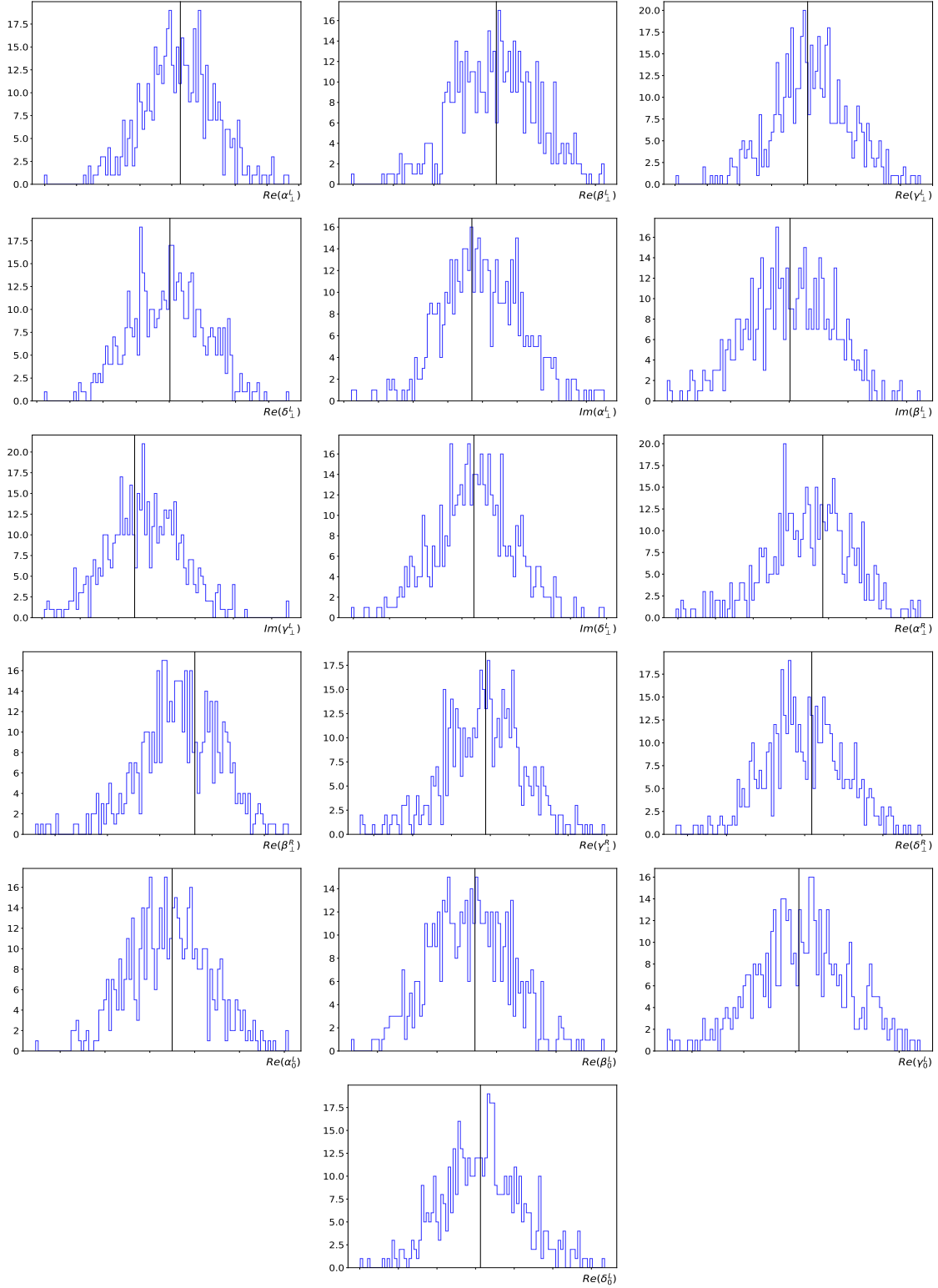


Figure 6.2 continued: Residuals of the P-wave parameters from ensembles of fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. The true values of the parameters are shown by the black lines. The x axis is currently blinded.

Since there are still correlations between the amplitude parameters, the residual plots should not be studied independently. The coverage of the toys is hence studied. For each amplitude parameter, the coverage indicated by Hessian uncertainties σ_{HESSE} of the fits are examined by calculating the fraction of fits where the *true* values are between the best-fit point μ minus a shift x of σ_{HESSE} and μ plus a shift x of σ_{HESSE} . This is compared to the absolute value of a cumulative distribution of a Normal distribution (i.e. a Gaussian with mean = 0 and $\sigma = 1$). These coverage are shown in Figure 6.3.

Although good coverage can be seen for most of the parameters, there are some plots where incorrect coverage is spotted. This motivates the estimation of the errors of each amplitude parameter using likelihood profiles. In a single toy fit, each amplitude parameter is profiled and a bifurcated parabola, i.e.

$$(6.10) \quad f = \begin{cases} a_l(x - \mu)^2, & x \leq \mu \\ a_r(x - \mu)^2, & x > \mu \end{cases},$$

is used to fit the likelihood profiles, where μ is the best-fit point of the corresponding amplitude parameter from the global fit and the subscript $l(r)$ represents the left-(right-)hand side of the best-fit point. The 1σ errors of an amplitude parameter are thus

$$(6.11) \quad \begin{cases} \sigma_l = f^{-1}(0.5), & x \leq \mu \\ \sigma_r = f^{-1}(0.5), & x > \mu \end{cases},$$

where f^{-1} denotes the inverse function of the bifurcated parabola as written in Equation 6.10.

Therefore, asymmetric uncertainties σ_{new} are calculated by fitting bifurcated parabolas to the likelihood profiles from the toys, following the formulas suggested in Equation 6.11. The coverage of the asymmetric uncertainties can thus be examined by comparing the fraction of fits where the *true* value is between $\mu - x\sigma_{\text{new}}$ and $\mu + x\sigma_{\text{new}}$ where σ_{new} with the cumulative distribution of the Normal distribution. As can be seen in Figure 6.3, the asymmetric uncertainties from fitting the likelihood profiles from the toys with bifurcated parabolas overall give coverage closer to the correct coverage than the Hessian error matrix does.

The distribution of toy fits of P-wave amplitudes is also studied and shown in Figure 6.4. In the figures, all amplitudes are shifted by the *true* value for blinding purposes, such that by definition the *true* value is at 0 (shown by the black line). Again, there are no large biases seen here.

6.6.2 Toy study in the region $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$

This section describes the toy studies in the region $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$. The projections of an example toy fit are shown in Figure 6.5 and no large pulls are spotted in the projections.

The ensemble of 500 toy fits is studied as well. The residuals of P-wave amplitude parameters are shown in Figure 6.6. It can be seen that some parameters are distributed very

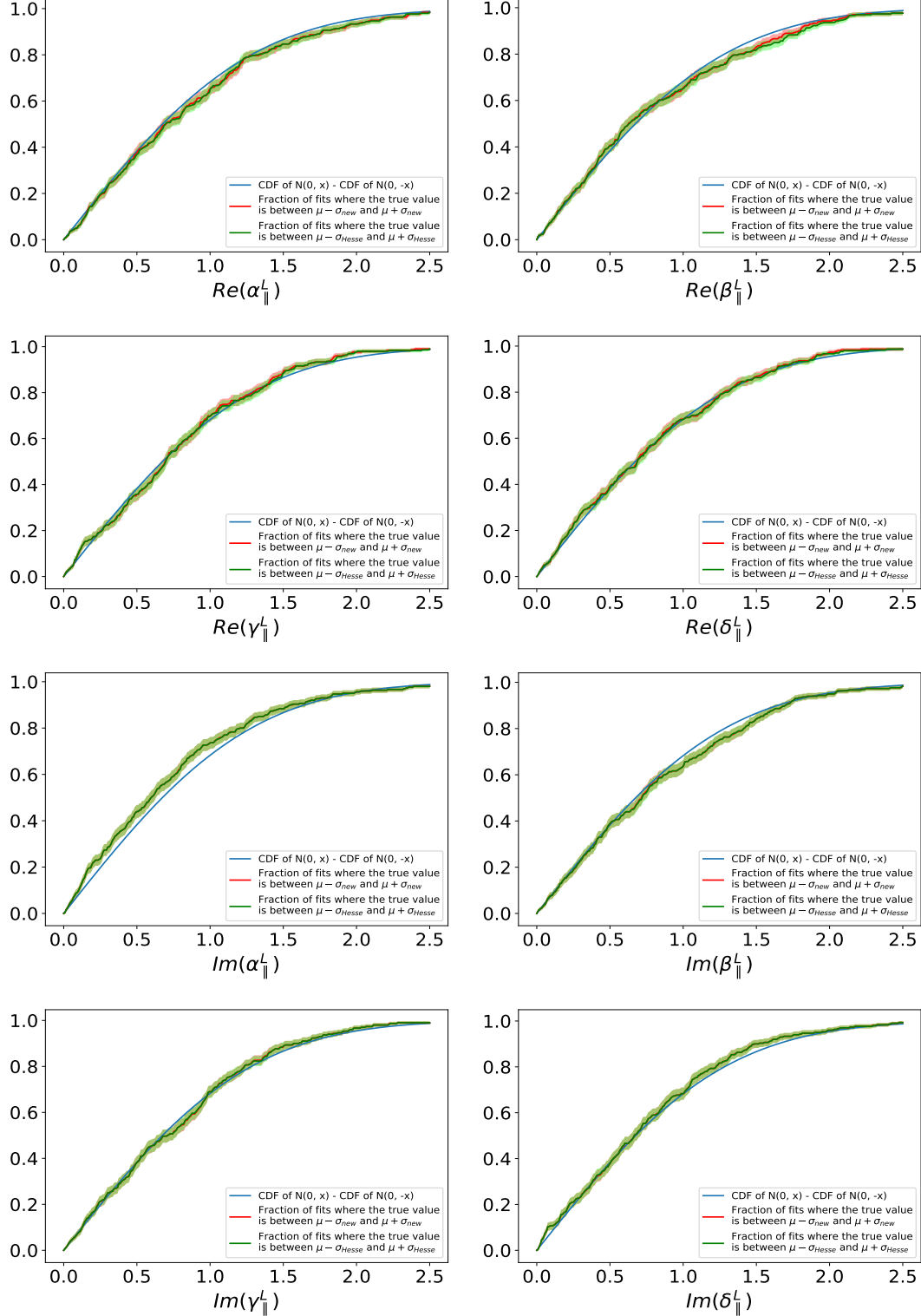


Figure 6.3: Coverage of the left-handed parallel amplitude components from the ensemble of toy fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. Those shown in green are the fraction of fits where the true value is between $\mu - x\sigma_{\text{HESSE}}$ and $\mu + x\sigma_{\text{HESSE}}$, while those shown in red is the same but for the uncertainties extracted from fitting bifurcated parabola to the likelihood profiles for each toy, denoted by σ_{new} . Both are compared to the absolute value of the cumulative distribution of a Normal distribution.

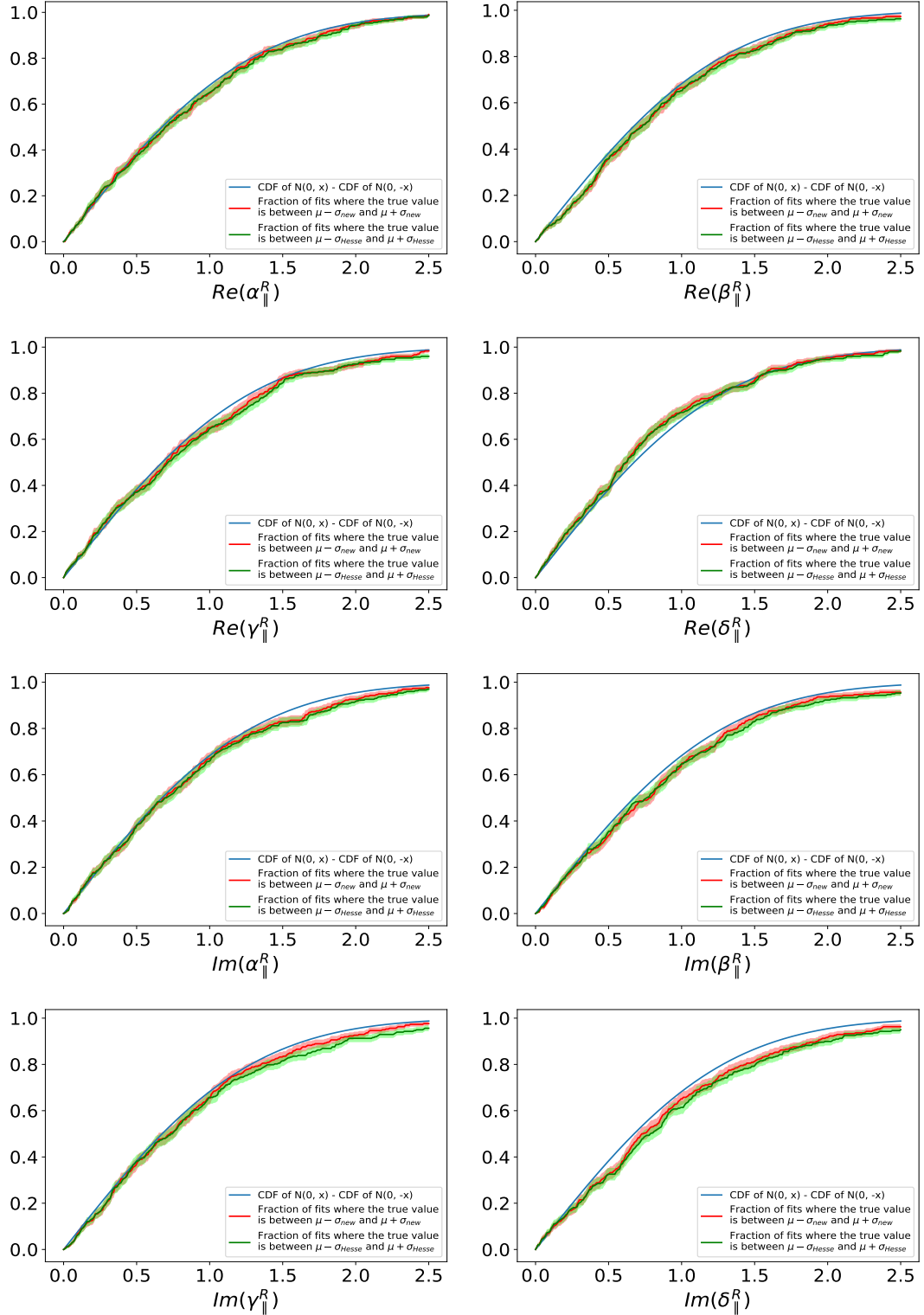


Figure 6.3 continued: Coverage of the left-handed parallel amplitude components from the ensemble of toy fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. Those shown in green are the fraction of fits where the true value is between $\mu - x\sigma_{\text{HESSE}}$ and $\mu + x\sigma_{\text{HESSE}}$, while those shown in red is the same but for the uncertainties extracted from fitting bifurcated parabola to the likelihood profiles for each toy, denoted by σ_{new} . Both are compared to the absolute value of the cumulative distribution of a Normal distribution.

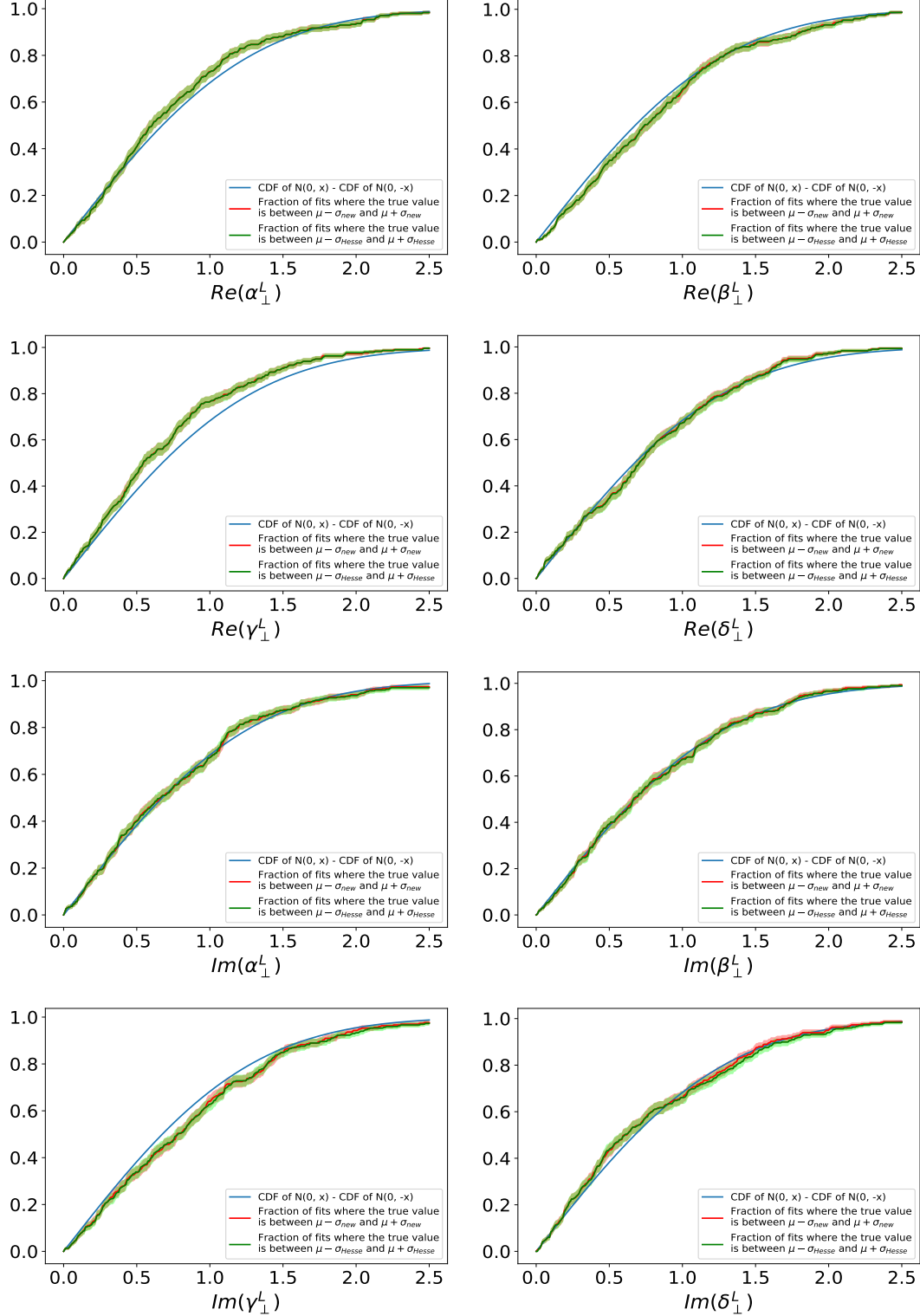


Figure 6.3 continued: Coverage of the left-handed parallel amplitude components from the ensemble of toy fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. Those shown in green are the fraction of fits where the true value is between $\mu - x\sigma_{\text{HESSE}}$ and $\mu + x\sigma_{\text{HESSE}}$, while those shown in red is the same but for the uncertainties extracted from fitting bifurcated parabola to the likelihood profiles for each toy, denoted by σ_{new} . Both are compared to the absolute value of the cumulative distribution of a Normal distribution.

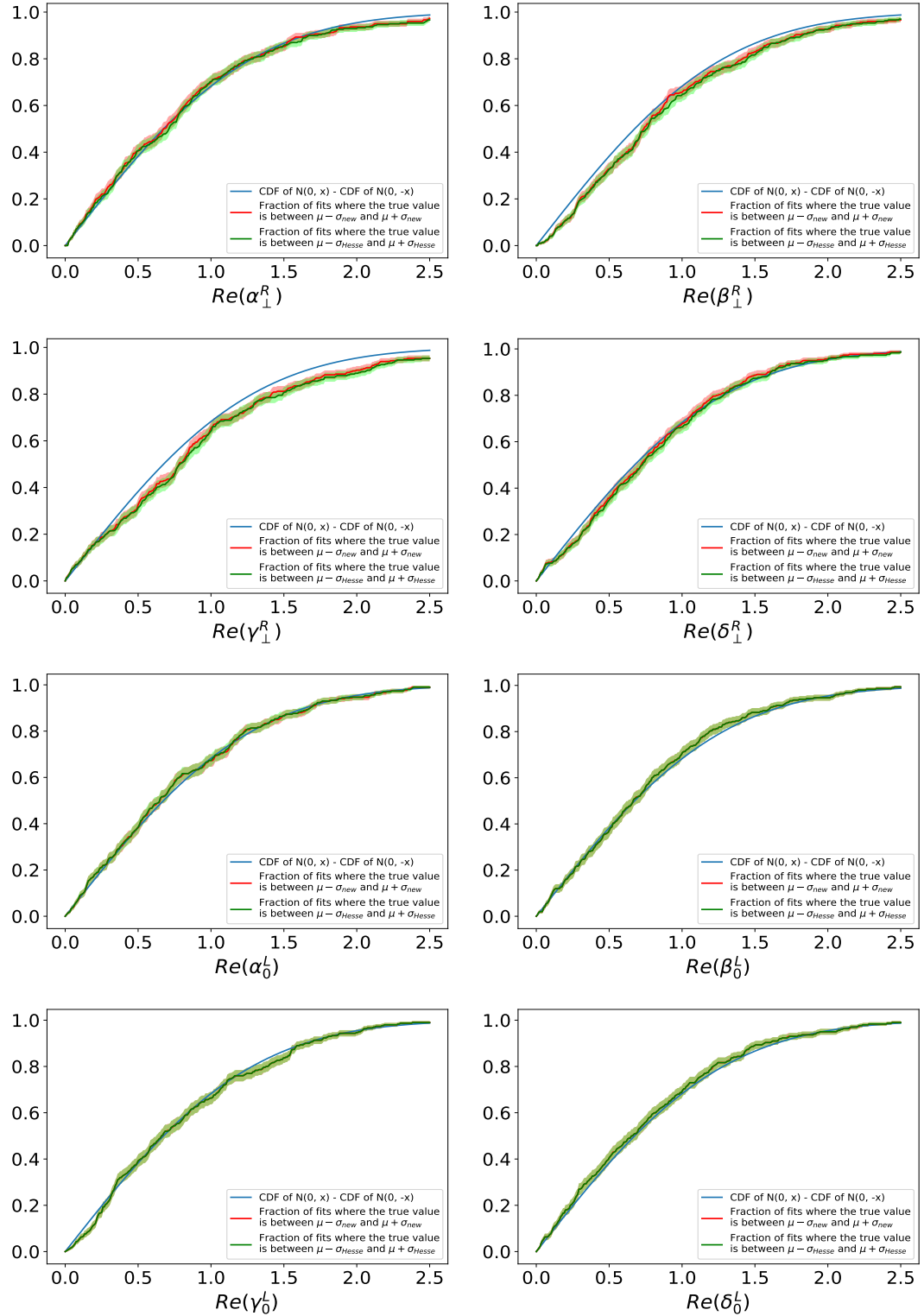


Figure 6.3 continued: Coverage of the left-handed parallel amplitude components from the ensemble of toy fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. Those shown in green are the fraction of fits where the true value is between $\mu - x\sigma_{\text{HESSE}}$ and $\mu + x\sigma_{\text{HESSE}}$, while those shown in red is the same but for the uncertainties extracted from fitting bifurcated parabola to the likelihood profiles for each toy, denoted by σ_{new} . Both are compared to the absolute value of the cumulative distribution of a Normal distribution.

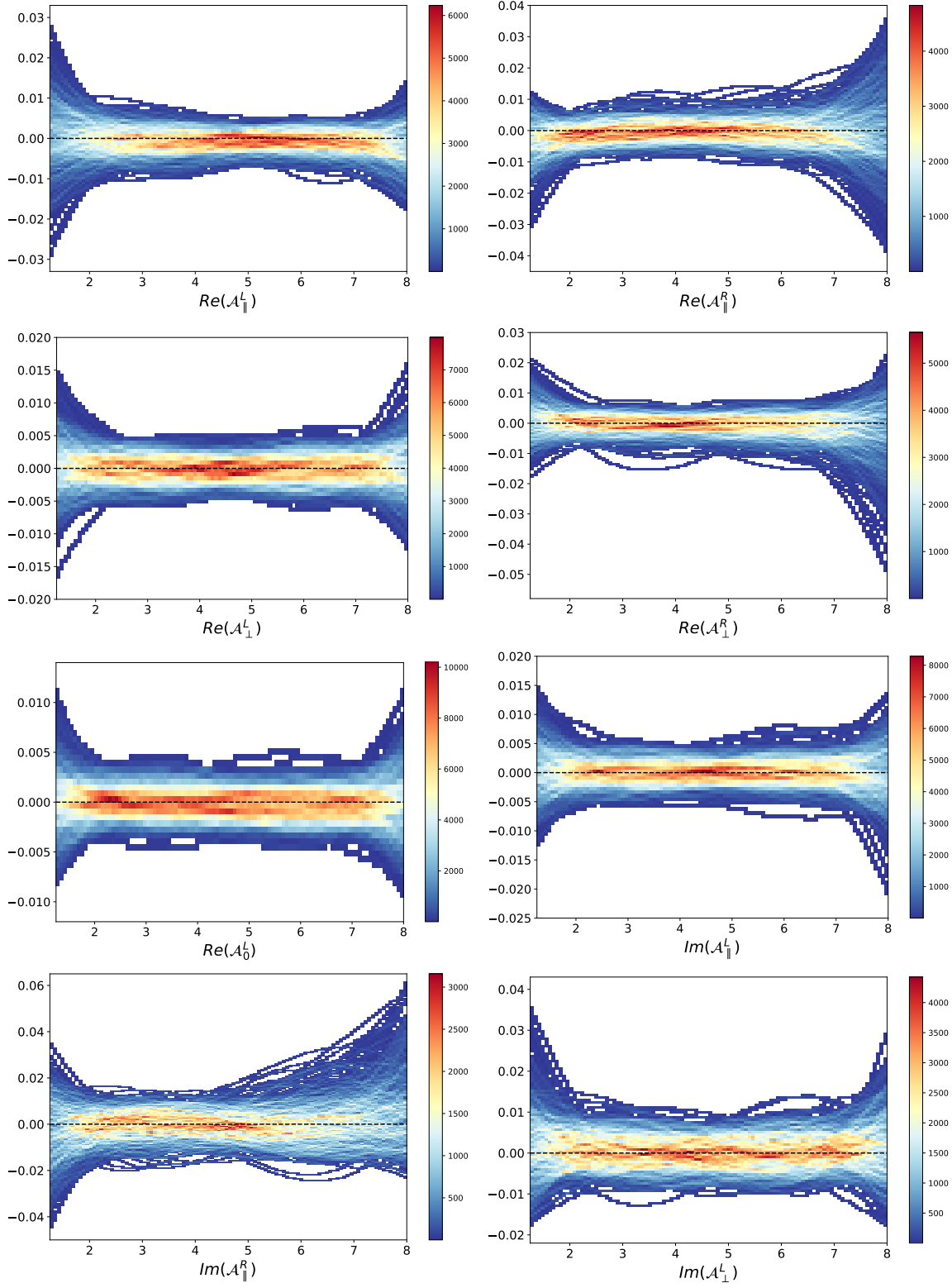


Figure 6.4: Distribution of fits of the P-wave amplitudes from the ensemble of toy fits in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. All fits are shifted by the true value for blinding.

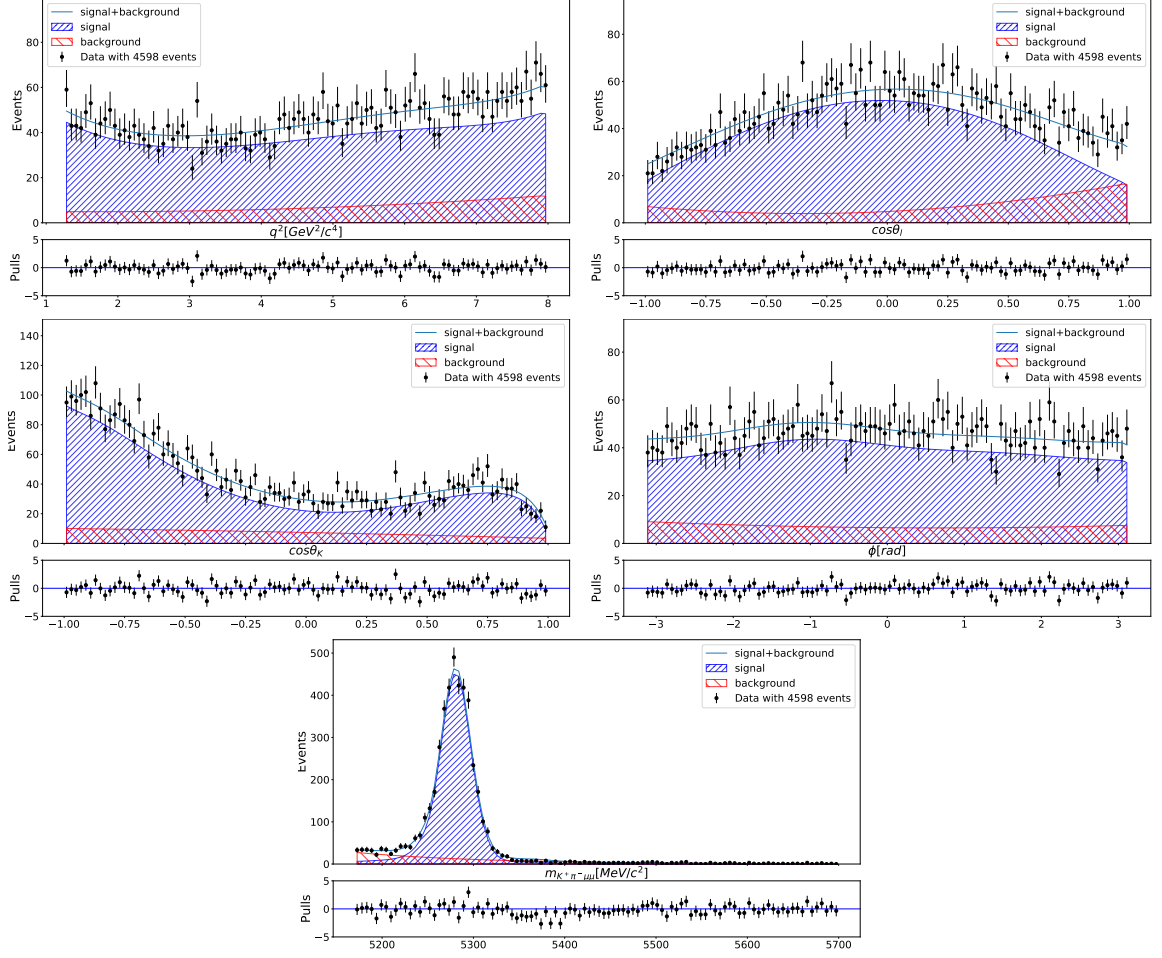


Figure 6.5: Example projections from fitting to a toy generated in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.

asymmetrically, which further motivates fitting bifurcated parabolas to the likelihood profiles to extract asymmetric errors. The biases due to physical boundary effects and correlation between parameters also exist. A second peak can be spotted for some of the parameters. This is due to the discrete symmetry of the PDF as mentioned in Section 6.2.1 and the value where the second peak is located has the same magnitude with the opposite sign as where the nominal peak is located.

The coverage of the toys was also studied. For each amplitude parameter, the fraction of fits where the *true* values are between the best-fit point μ minus a shift x of Hessian uncertainties $\sigma_{\text{HESS}}E$ and μ plus a shift x of $\sigma_{\text{HESS}}E$ is calculated. The fraction of fits where the *true* value is between $\mu - x\sigma_{\text{new}}$ and $\mu + x\sigma_{\text{new}}$ is also calculated, where the asymmetric uncertainties σ_{new} are obtained by fitting the likelihood profiles from the toys with bifurcated parabolas as suggested in Equation 6.11. Both fractions are compared to the absolute value of a cumulative distribution of a Normal distribution. These coverage are shown in Figure 6.7. Overall, the

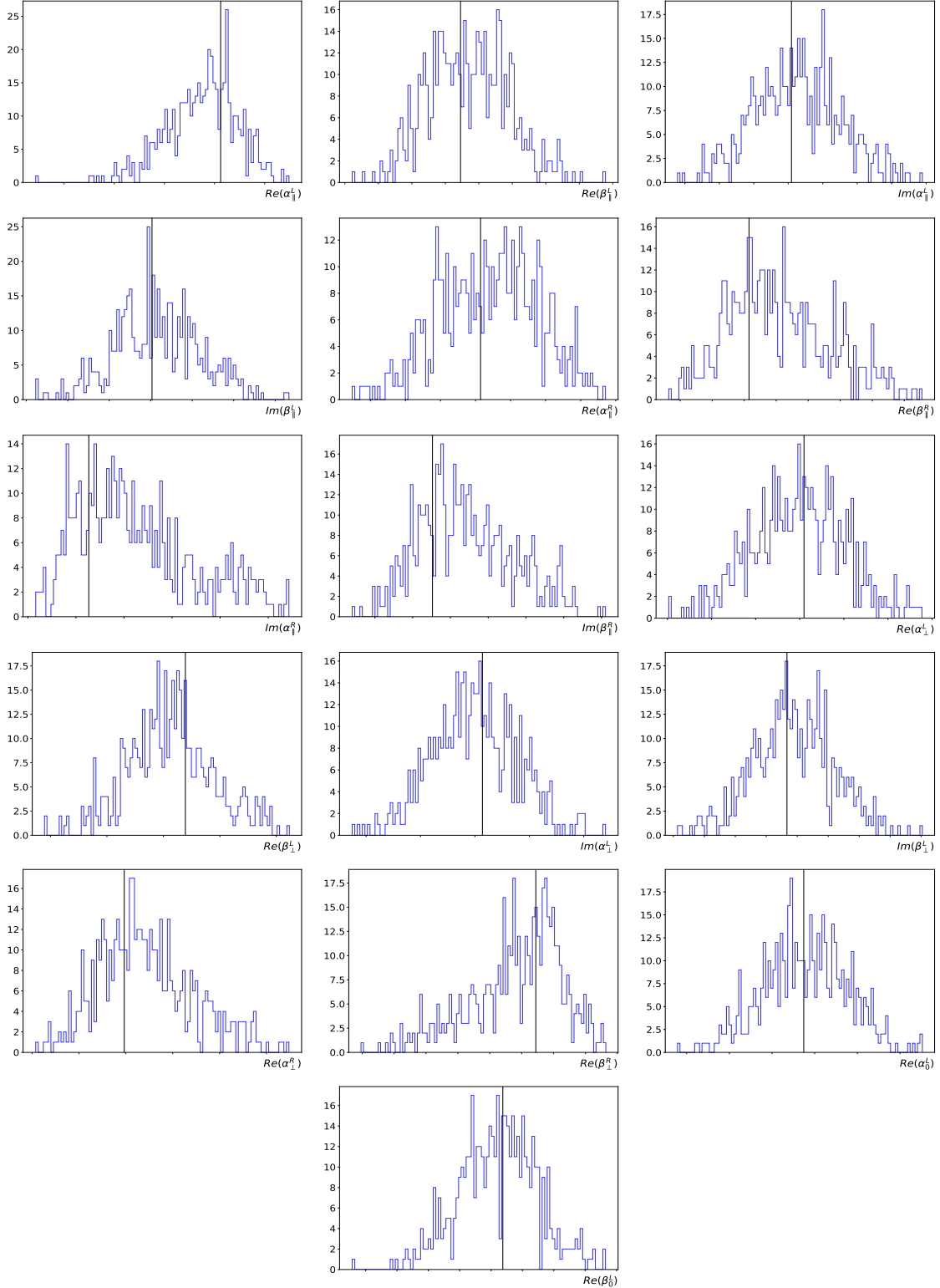


Figure 6.6: Residuals of the P-wave parameters (parallel amplitude components) from ensembles of fits in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. The value of the parameters used for the generation is from the data fit. The true values of the parameters are shown by the black lines. Note the x axis is currently blinded.

asymmetric uncertainties from fitting the likelihood profiles give coverage closer to the correct coverage than the Hessian error matrix.

The distribution of toy fits of P-wave amplitudes is also studied and shown in Figure 6.8. In the figures, all amplitudes are shifted by the *true* value for blinding purposes, such that by definition the *true* value is at 0 (shown by the black line). Again, there are no large biases seen here.

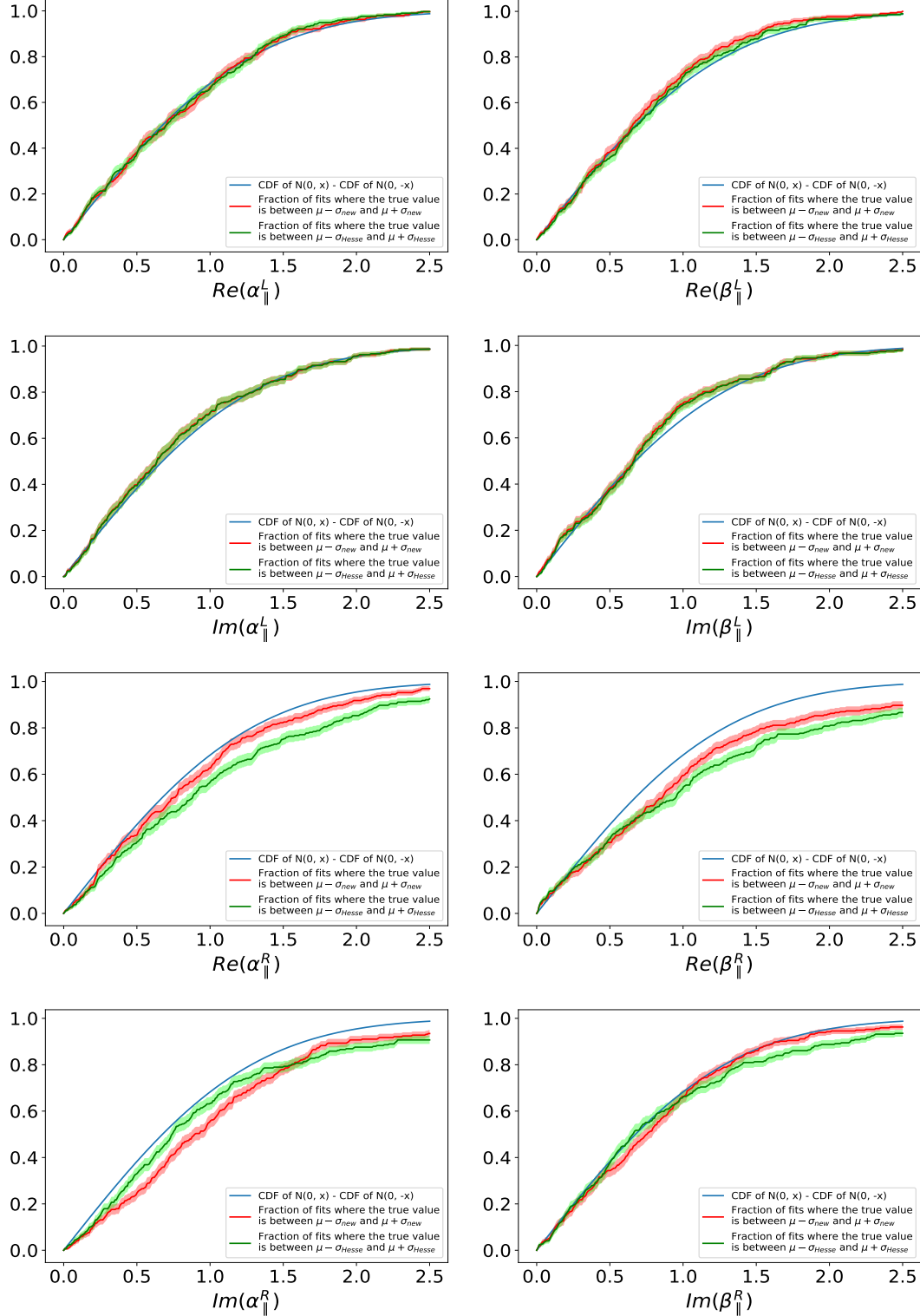


Figure 6.7: Coverage of the left-handed parallel amplitude components from the ensemble of toy fits in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. Those shown in green are the fraction of fits where the true value is between $\mu - x\sigma_{\text{HESSE}}$ and $\mu + x\sigma_{\text{HESSE}}$, while those shown in red is the same but for the uncertainties extracted from fitting bifurcated parabola to the likelihood profiles for each toy, denoted by σ_{new} . Both are compared to the absolute value of the cumulative distribution of a Normal distribution.

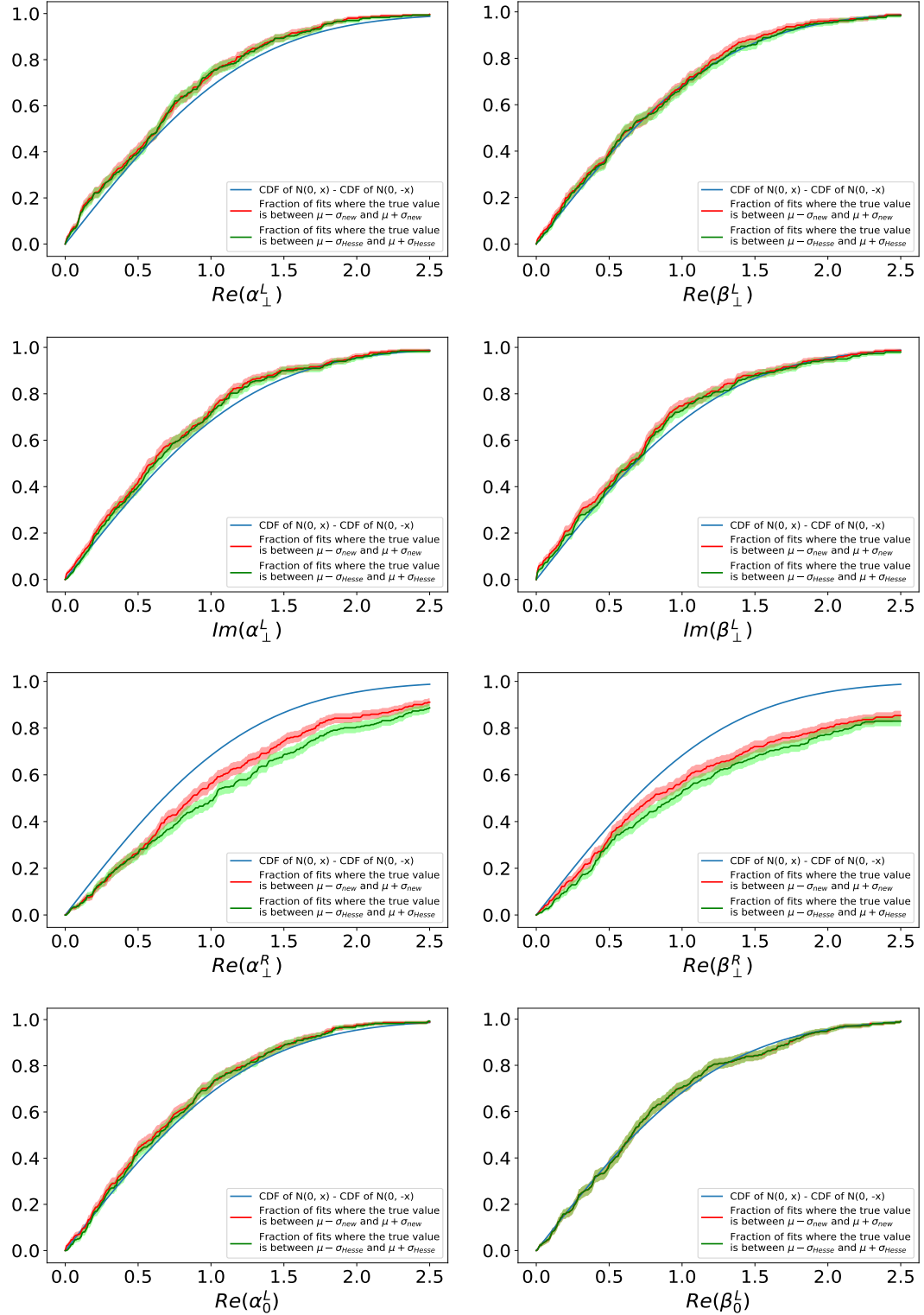


Figure 6.7 continued: Coverage of the left-handed parallel amplitude components from the ensemble of toy fits in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. Those shown in green are the fraction of fits where the true value is between $\mu - x\sigma_{\text{HESSE}}$ and $\mu + x\sigma_{\text{HESSE}}$, while those shown in red is the same but for the uncertainties extracted from fitting bifurcated parabola to the likelihood profiles for each toy, denoted by σ_{new} . Both are compared to the absolute value of the cumulative distribution of a Normal distribution.

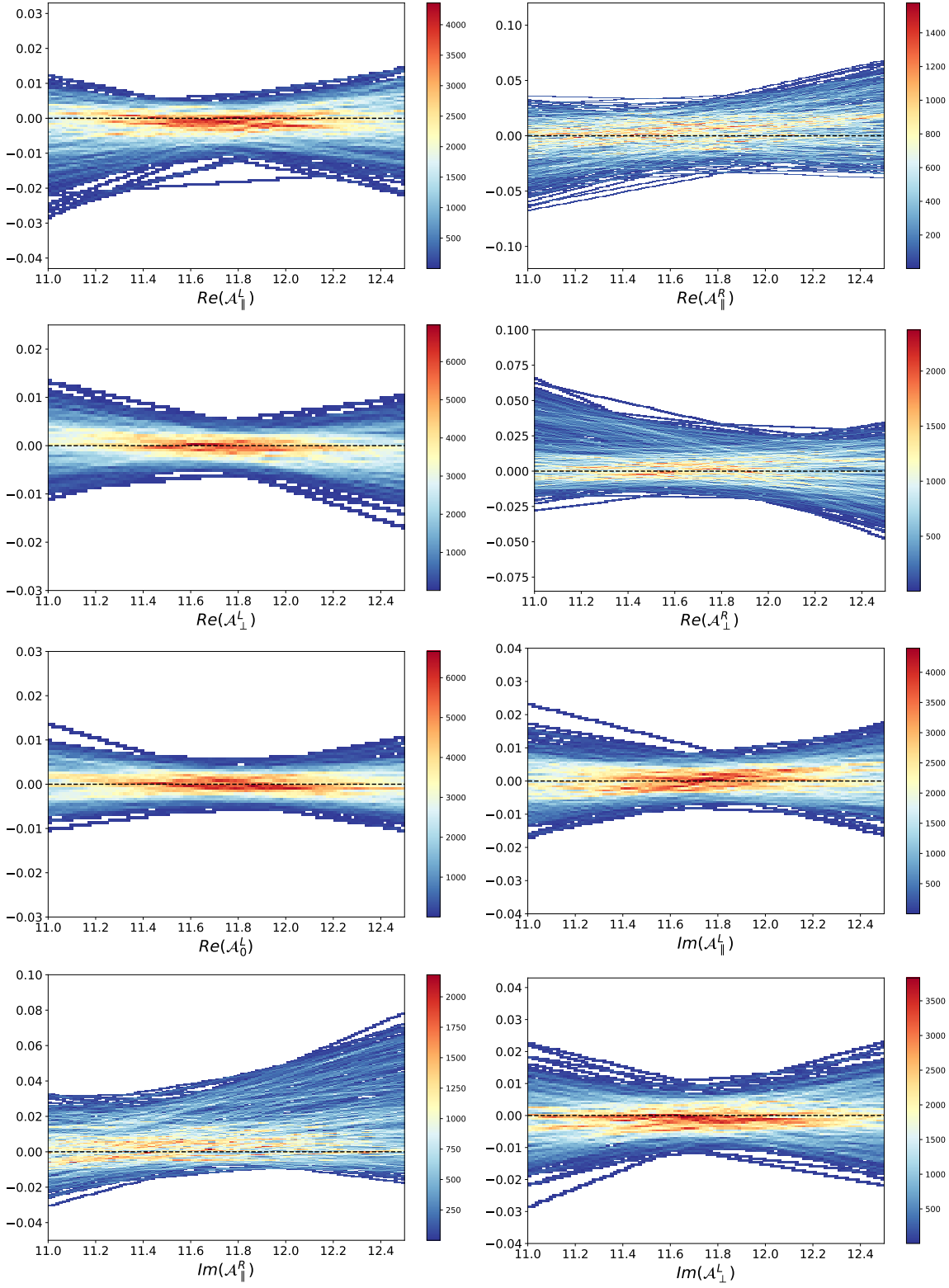


Figure 6.8: Distribution of fits of the P-wave amplitudes from the ensemble of toy fits in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. All fits are shifted by the true value for blinding.

6.7 Validation with control mode fit

In this analysis, the $B^0 \rightarrow J/\psi K^{*0}$ resonance is chosen to be the control mode taking advantage of its high statistics data samples. This control mode is fitted in order to validate the fitter. The fit is performed for each data-taking period, namely Run 1, 2016, 2017 and 2018 and the consistency of the observables between periods is examined.

To select the desired control mode data, the q^2 region here is chosen to be 60 MeV around the J/ψ mass peak, namely $9.223 < q^2 < 9.966 \text{ GeV}^2/c^4$. As the J/ψ resonance is peaking at a point in q^2 , the q^2 dependencies in the amplitudes should be removed. The P-wave amplitudes are hence set to be flat in q^2 . In other words, the maximum order of the ansatz 2.89 for P-wave amplitudes is set to be 0, i.e.

$$\mathcal{A}_{\parallel,\perp,0}^{L,R} = \alpha_{\parallel,\perp,0}^{L,R},$$

where α is the complex constant amplitude parameter to fit for.

For this fit configuration, the B_s^0 relative yields and mass difference, along with one set of CB tail parameters α , are additionally floated compared to the rare mode fit. The plots of the fits are shown in Figure 6.9 - 6.12.

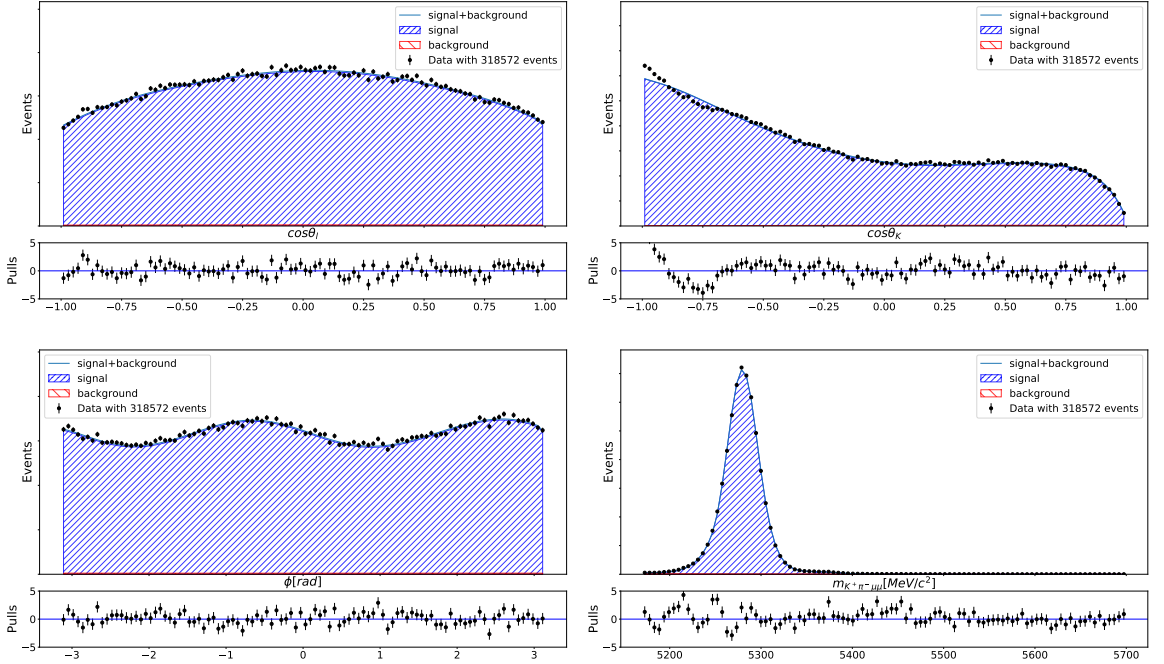
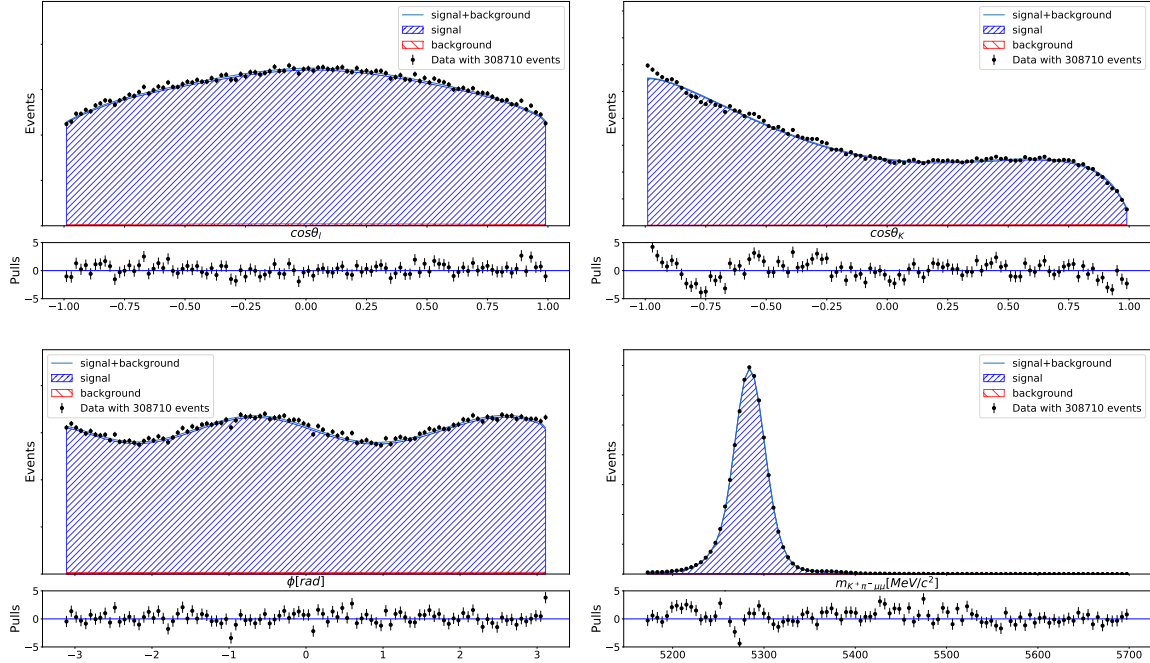
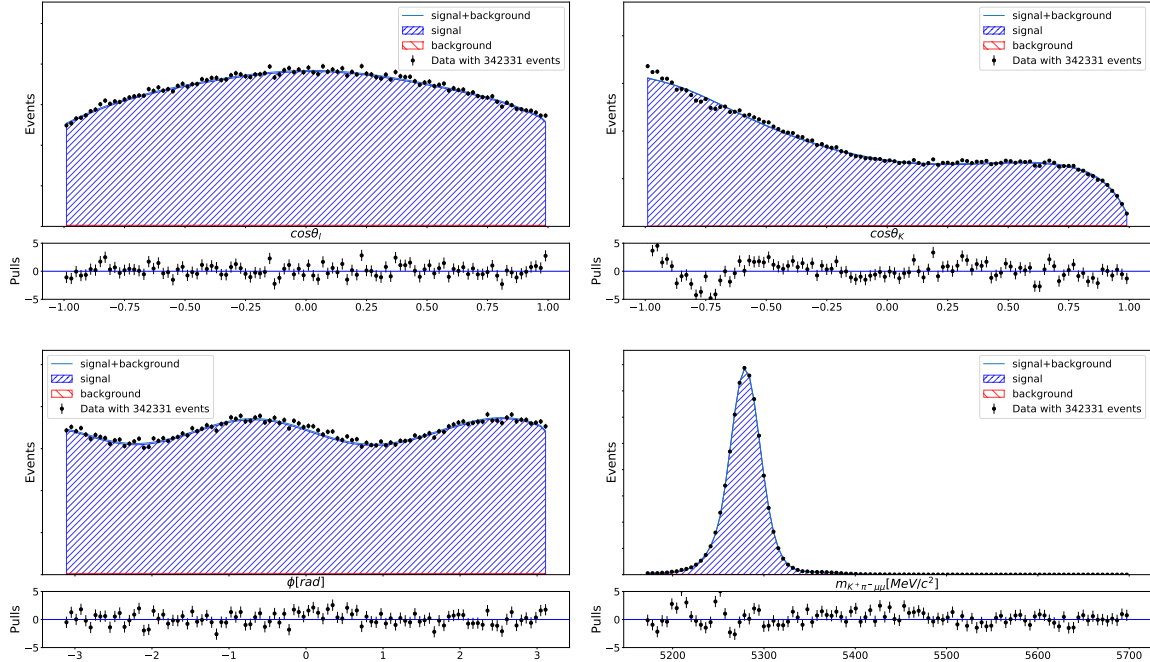
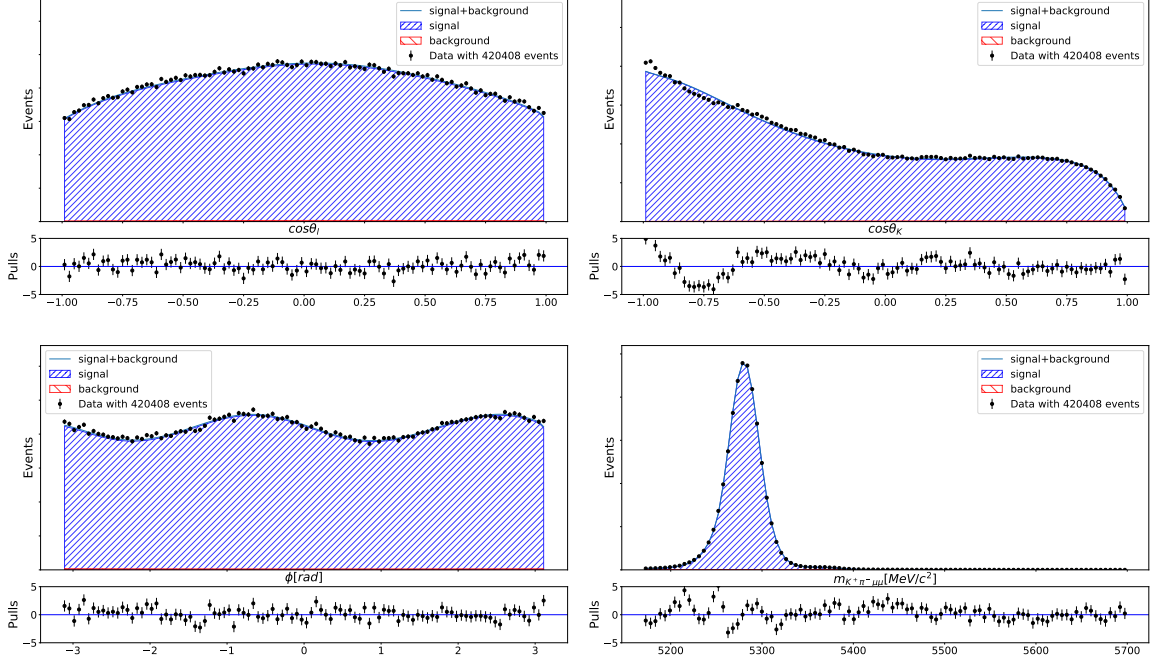


Figure 6.9: Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to Run 1 data.

Some large pulls can be spotted in the plots. The large pulls in low $\cos \theta_K$ are understood to be the pollution from $J/\psi \pi$ exotics, especially $B^0 \rightarrow Z (\rightarrow J/\psi \pi) K$. The pull in m_{B^0} can be resolved by using a more complex model for the B^0 mass lineshape. However, because the main purpose of control mode fits is to validate the fitter configuration and this effect is not present in rare mode fits, the B^0 mass model is kept the same in all regions.

Figure 6.10: Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to 2016 data.Figure 6.11: Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to 2017 data.


 Figure 6.12: Blinded plots of the $B^0 \rightarrow J/\psi K^{*0}$ fit to 2018 data.

The fit results and covariance matrix are then taken to sample an ensemble of amplitude coefficients, from which the observables can be computed. The observables in each data period are shown in Table 6.1. The comparisons between each data period are made by computing the pulls between the observables of two data periods and are shown in Table 6.2. It can be seen that there are no deviations more than 3σ in any of the observables and the fitter thus shows good stability.

 Table 6.1: Results for the observables from the $B^0 \rightarrow J/\psi K^{*0}$ fits.

	R1	2016	2017	2018
F_L	0.55813 ± 0.00156	0.56351 ± 0.00189	0.55965 ± 0.00150	0.56114 ± 0.00131
S_3	-0.00639 ± 0.00184	-0.0038 ± 0.00181	-0.00595 ± 0.00176	-0.00433 ± 0.00171
S_4	-0.24490 ± 0.00107	-0.24358 ± 0.00108	-0.24526 ± 0.00108	-0.24370 ± 0.00100
S_5	-0.00223 ± 0.00227	0.00171 ± 0.00210	-0.00030 ± 0.00220	0.00452 ± 0.00205
A_{FB}	-0.00039 ± 0.00141	-0.000180 ± 0.00141	-0.00143 ± 0.00131	-0.00263 ± 0.00120
S_7	0.00253 ± 0.00234	0.00274 ± 0.00234	-0.00238 ± 0.00220	0.00202 ± 0.00203
S_8	-0.05108 ± 0.00212	-0.05637 ± 0.00226	-0.05515 ± 0.00207	-0.05340 ± 0.00183
S_9	-0.09103 ± 0.00184	-0.09268 ± 0.00177	-0.09152 ± 0.00176	-0.09190 ± 0.00152
F_S	0.08099 ± 0.00266	0.08108 ± 0.00250	0.08814 ± 0.00265	0.08150 ± 0.00239

As a final check, the P-wave observables from control mode fits in this analysis are compared with those from the ongoing binned analysis [122] in Run 1 and 2016, as shown in Table 6.3. It can be seen that most of the P-wave observables present good compatibility between these two analyses. Although a few observables showed relatively large differences greater than 3σ , it should be noted that the systematic uncertainties of the observables are not included in this

Table 6.2: Comparisons of the observables from the $B^0 \rightarrow J/\psi K^{*0}$ fits between each data-taking period.

	R1 – 2016	R1 – 2017	R1 – 2018	2016 – 2017	2016 – 2018	2017 – 2018
F_L	-2.18	-0.68	-1.47	1.57	1.04	-0.73
S_3	-1.01	-0.17	-0.84	0.88	0.22	-0.70
S_4	-0.86	0.24	-0.80	1.17	0.09	-1.13
S_5	-1.25	-0.61	-2.30	0.65	-0.97	-1.66
A_{FB}	-0.11	0.57	1.27	0.66	1.35	0.70
S_7	-0.06	1.52	0.17	1.58	0.23	-1.46
S_8	1.66	1.30	0.80	-0.39	-1.02	-0.62
S_9	0.67	0.19	0.37	-0.46	-0.34	0.16
F_S	-0.03	-1.87	-0.14	-1.92	-0.12	1.83

comparison. Furthermore, the rare modes where this analysis is performed have much smaller data samples than the control modes and therefore have a far lower statistical precision. These occasional discrepancies in P-wave observables between the two analyses are hence considered acceptable.

Table 6.3: Comparisons of the observables from the $B^0 \rightarrow J/\psi K^{*0}$ fits between each data-taking period.

	Run 1	2016
F_L	0.65	-1.27
S_3	-1.56	-0.95
S_4	3.35	0.63
S_5	0.33	-0.27
A_{FB}	-0.46	-0.74
S_7	0.50	-0.06
S_8	0.14	0.30
S_9	-2.69	-3.26

Chapter 7

Results

The results from rare mode fits to Run 1 + Run 2 data are presented in this Chapter. At this stage of analysis review, the signal parameters could potentially introduce observation biases to other measurements ongoing. They are therefore hidden and this procedure is commonly referred to as blinding. To summarise, the parameters that are varied in the rare mode fit include the parameters in the Legendre ansatzes describing the amplitudes, the mean and two variances of the dual Crystal Ball function describing the $m_{K\pi\mu\mu}$ lineshape, the parameters of the second-order Chebyshev polynomials modelling the angular background, the coefficient τ_m describing the $m_{K\pi\mu\mu}$ background, and finally the background yield. In the rare mode fits, the B_s^0 relative yields and mass difference are fixed to values from the PDG [148]. The tail and core fraction parameters of the Crystal Ball function are fixed to the result from the fit to simulation. The Run 1 and Run 2 data samples are combined and fit in two different q^2 regions, as described in Section 6.5. The four-dimensional acceptance function is included when constructing the PDF and an extended negative log-likelihood function is constructed to artificially put a constraint on the scale of the amplitude parameters.

In each q^2 region, the projections of the fitted PDF are first examined to ensure that the fitted PDF can correctly describe the data samples. The profile likelihoods of individual P-wave amplitude parameters are then examined to ensure the description of the likelihood surface is valid. The results from fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region are presented in Section 7.1 while those in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region are shown in Section 7.2.

7.1 Fit to Run 1 + 2 data in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region

The plots of the fit are shown in Figure 7.1. There are no large pulls seen in these fits.

The profile likelihoods are checked as shown in Figure 7.2. It can be spotted that Hessian errors cannot always model the likelihood surface precisely. Therefore, bifurcated parabolas are fit to the likelihood profiles to construct a new set of error intervals. This method is validated

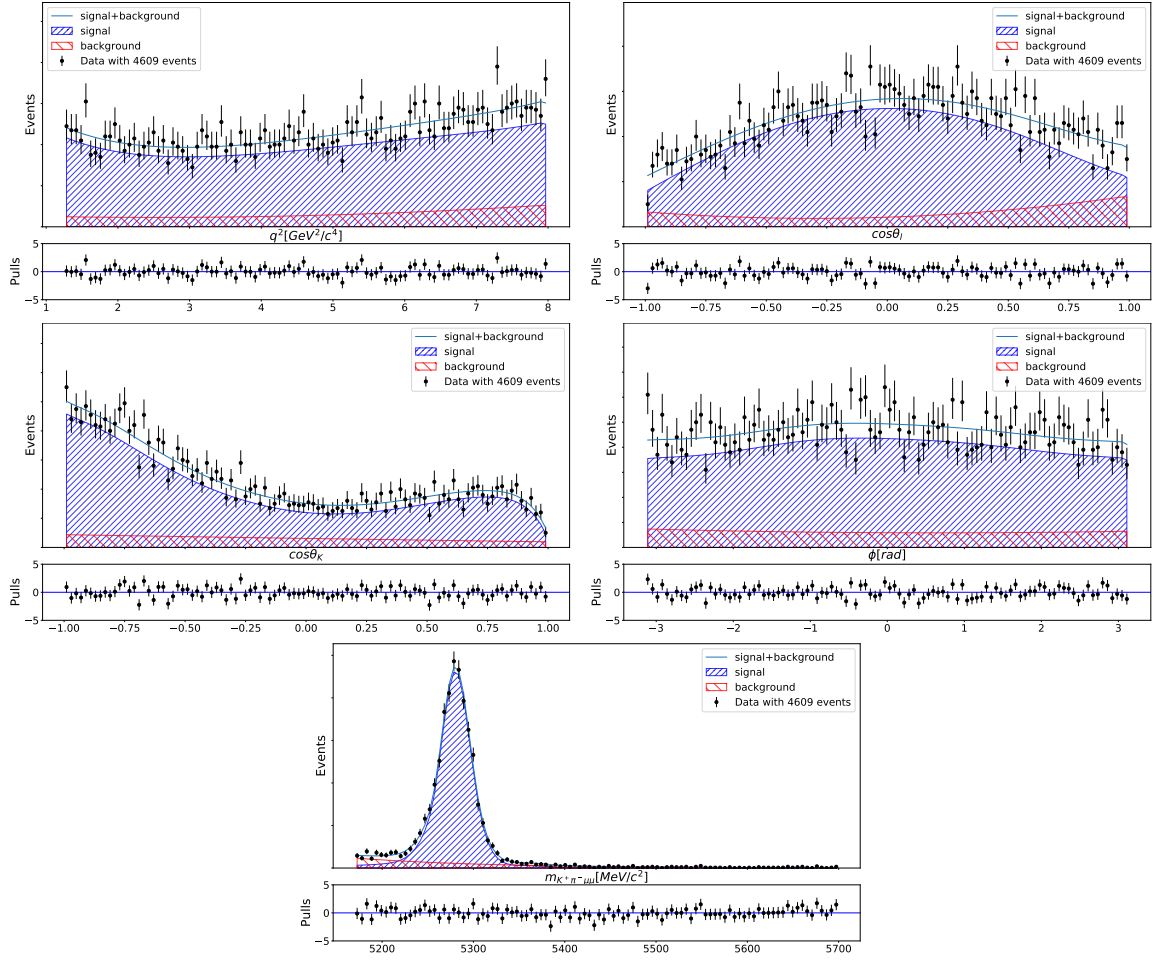


Figure 7.1: Projections from fitting to Run 1 + Run 2 data in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region.

in Section 6.6.1. As can be seen in the plots, the bifurcated parabolas can reasonably describe the likelihood surface within the 3σ region.

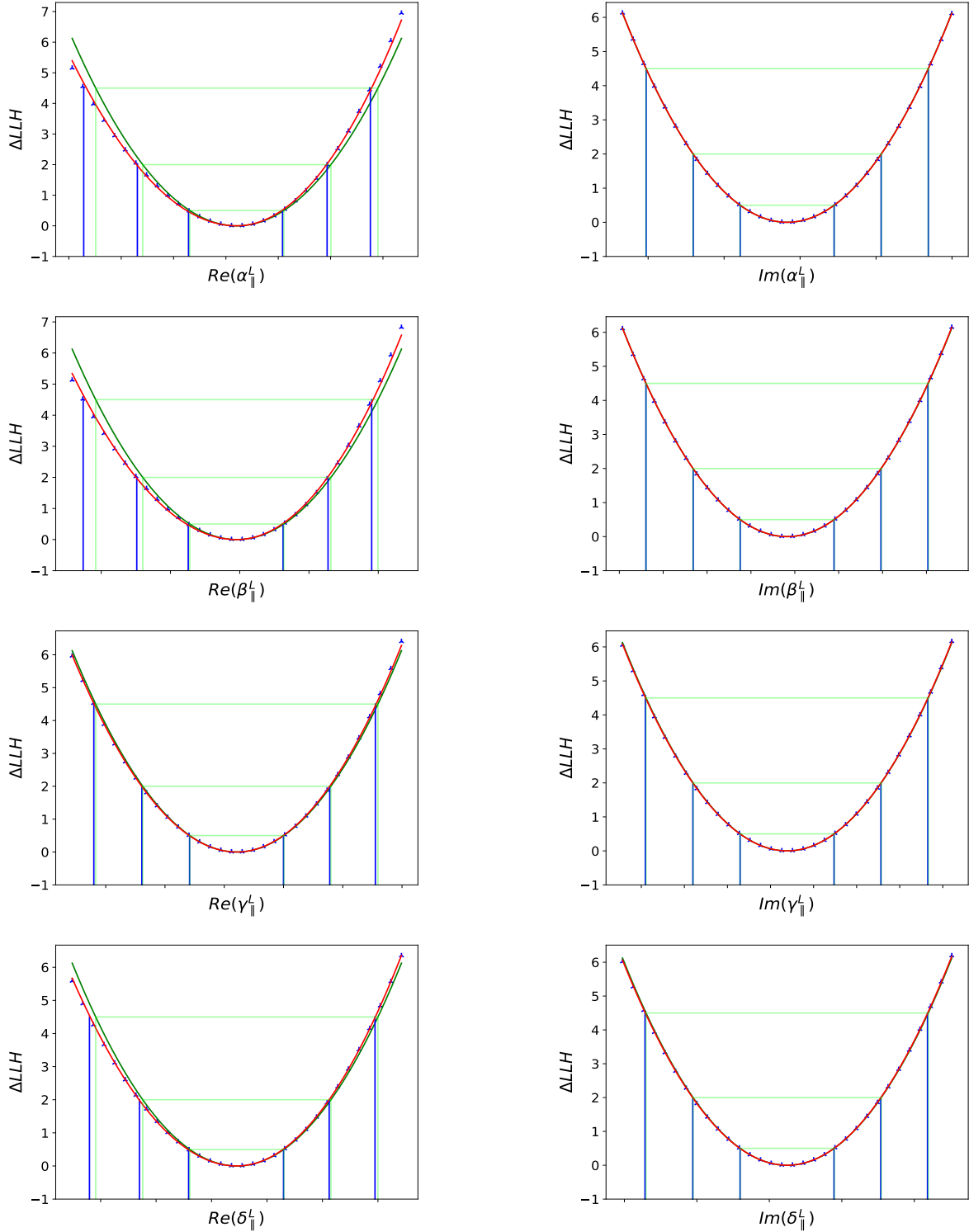


Figure 7.2: Likelihood profiles of the P-wave amplitude coefficients in the fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. The black points represent the likelihood profiles. The green and blue lines indicate the 1, 2 and 3 σ errors predicted by Hesse and the profiles respectively. The green parabola corresponds to the Hessian error matrix, and the red parabolas are fits of these profiles to bifurcated parabolas. Note the x axis is currently blinded.

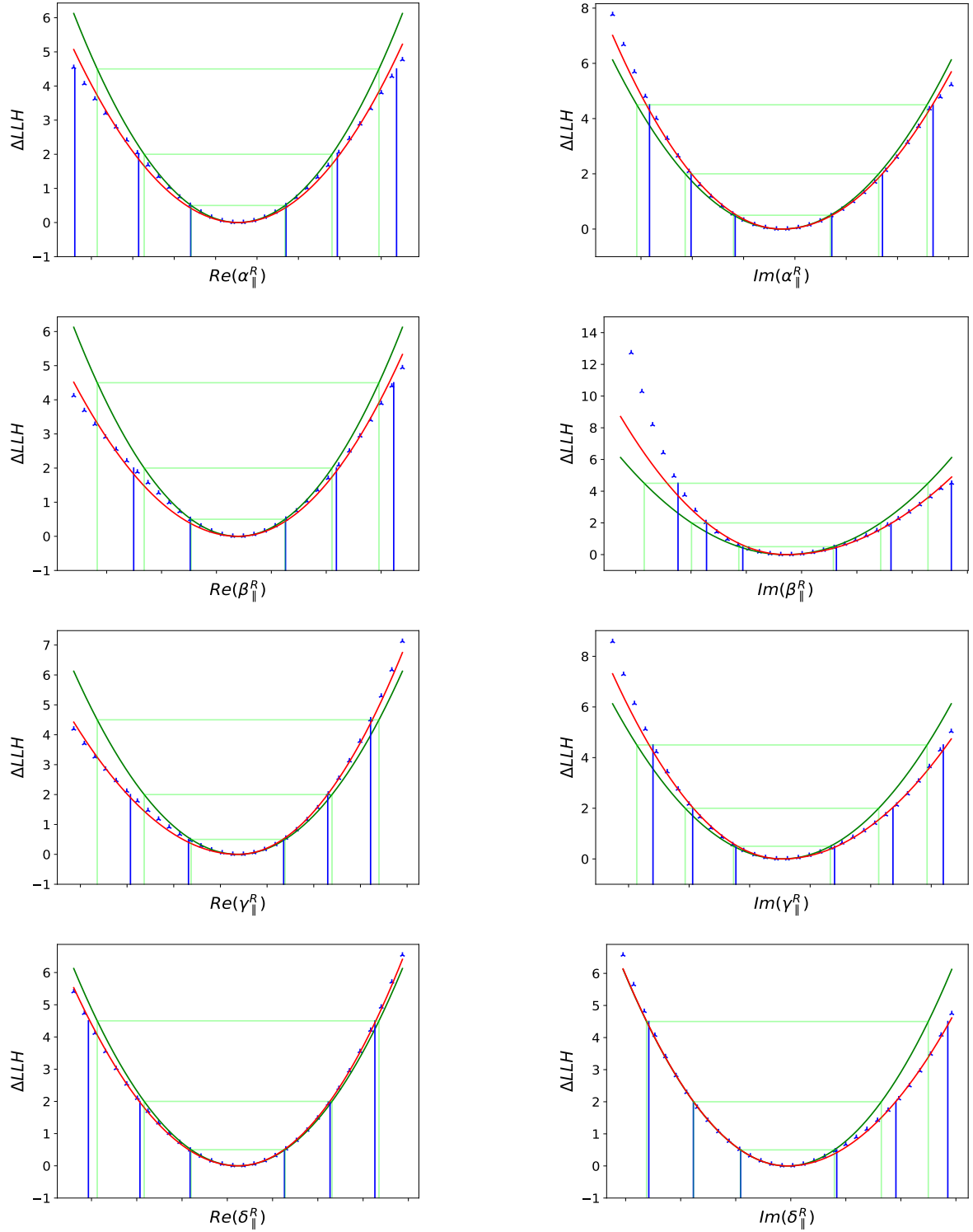


Figure 7.2 continued: Likelihood profiles of the P-wave amplitude coefficients in the fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. The black points represent the likelihood profiles. The green and blue lines indicate the 1, 2 and 3σ errors predicted by Hesse and the profiles respectively. The green parabola corresponds to the Hessian error matrix, and the red parabolas are fits of these profiles to bifurcated parabolas. Note the x axis is currently blinded.

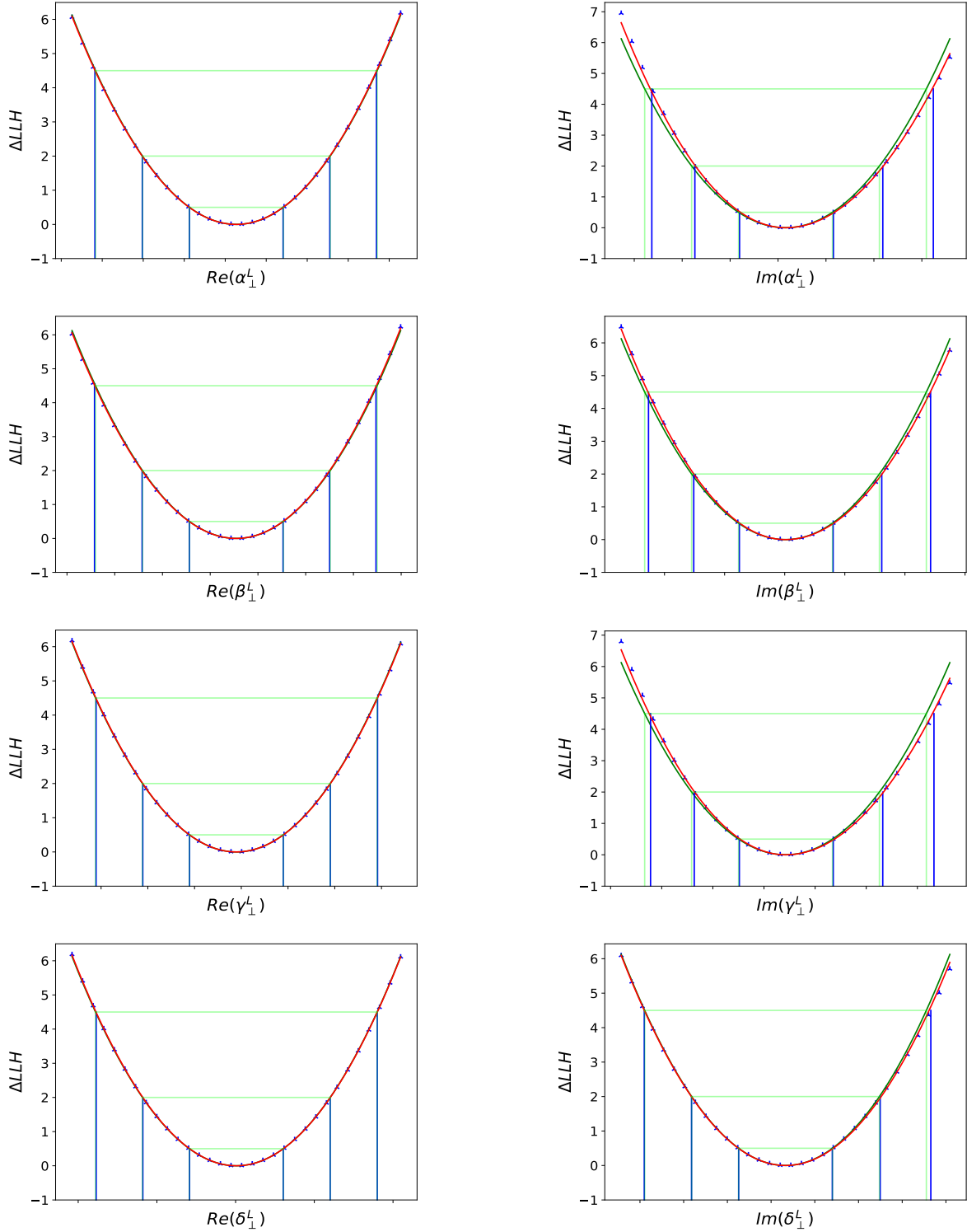


Figure 7.2 continued: Likelihood profiles of the P-wave amplitude coefficients in the fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. The black points represent the likelihood profiles. The green and blue lines indicate the 1, 2 and 3 σ errors predicted by Hesse and the profiles respectively. The green parabola corresponds to the Hessian error matrix, and the red parabolas are fits of these profiles to bifurcated parabolas. Note the x axis is currently blinded.

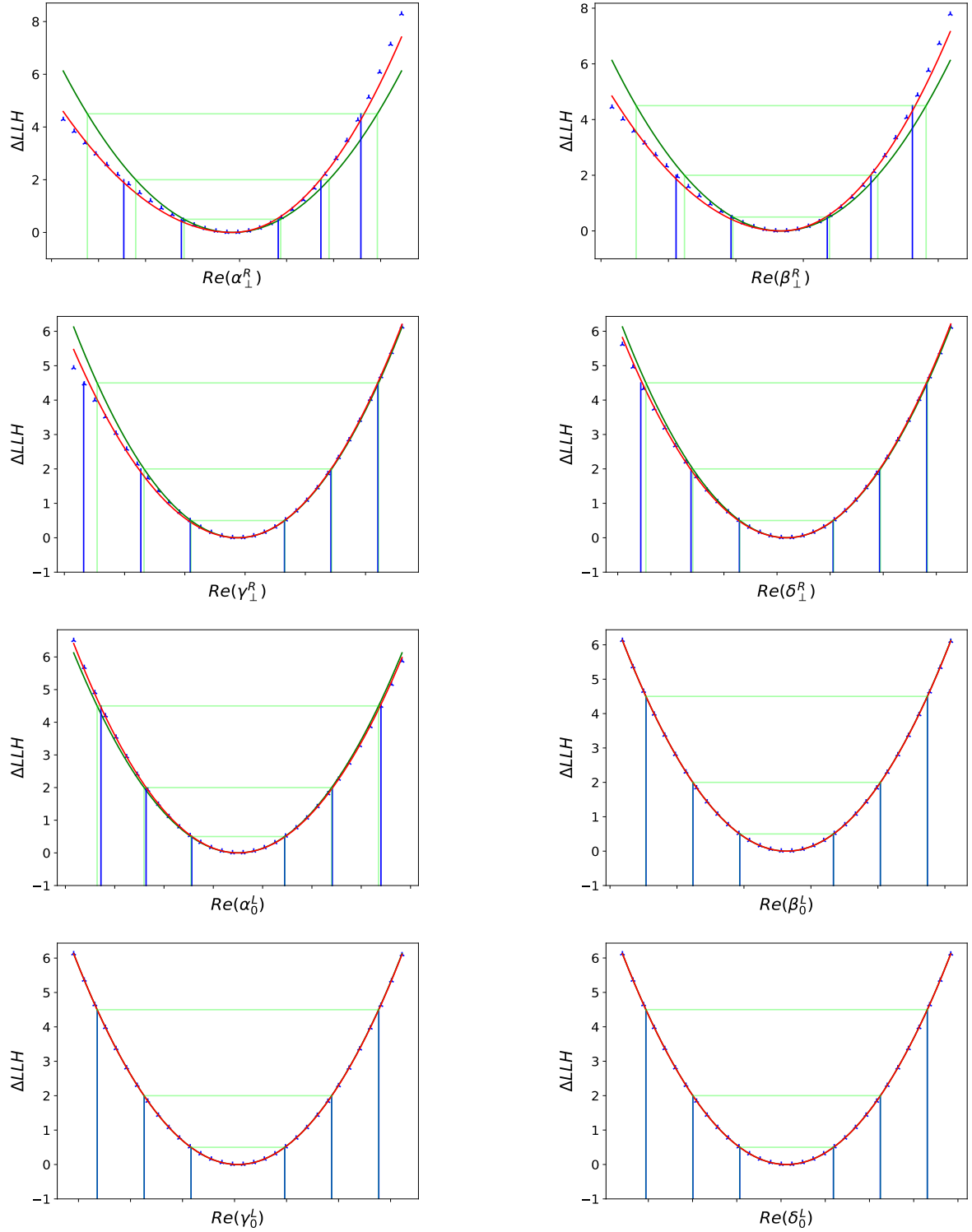


Figure 7.2 continued: Likelihood profiles of the P-wave amplitude coefficients in the fit in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. The blue points represent the profiles. The green and blue lines indicate the 1, 2 and 3 σ errors predicted by Hesse and the profiles respectively. The green parabola corresponds to the Hessian error matrix, and the red parabolas are fits of these profiles to bifurcated parabolas. Note the x axis is currently blinded.

7.2 Fit to Run 1 + 2 data in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region

The plots of the fit are shown in Figure 7.3. The fit presents good quality with no large pulls seen in these fits.

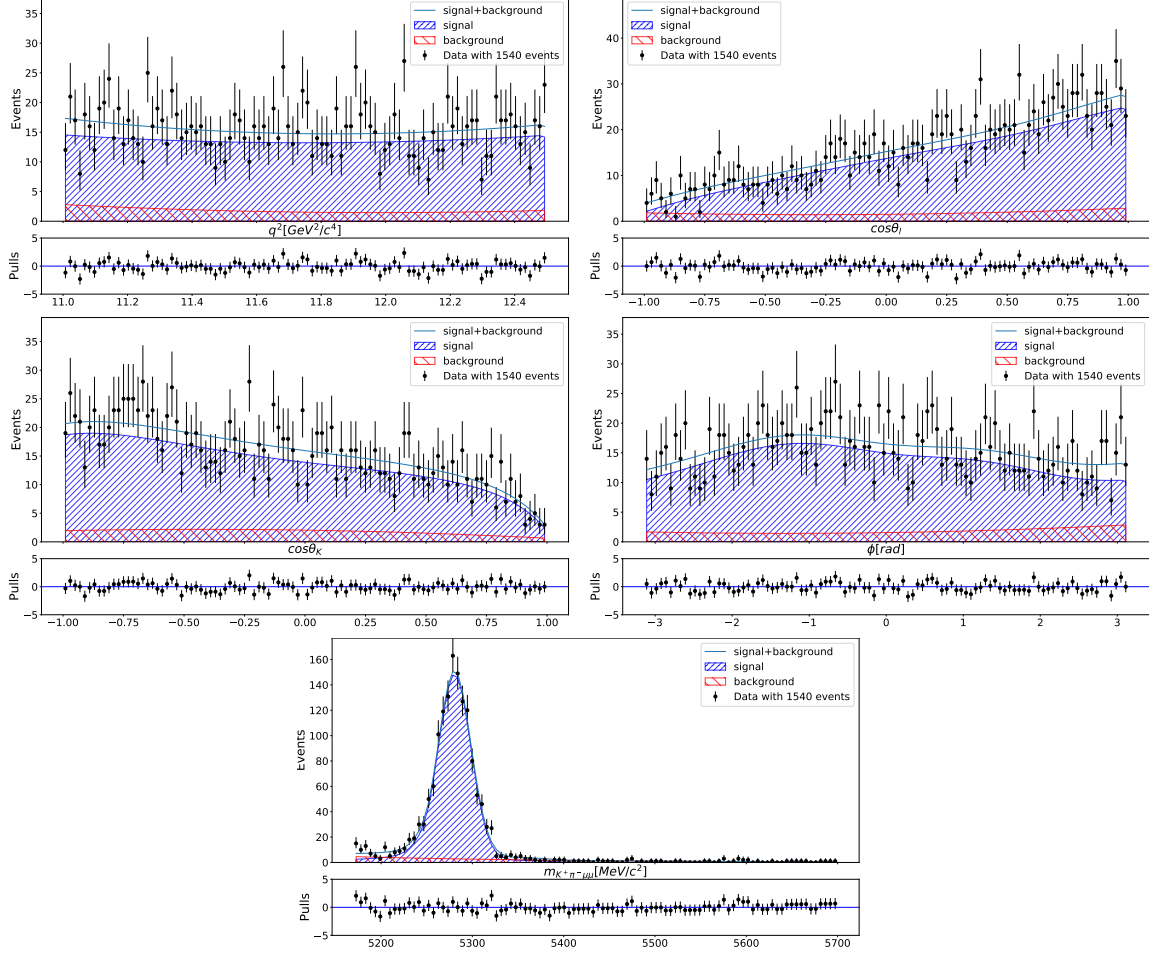


Figure 7.3: Projections from fitting to Run 1 + Run 2 data in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region.

The profile likelihoods are presented in Figure 7.4. Again bifurcated parabolas are fit to the likelihood profiles to construct a new set of error intervals. Some of the parameters are very close to zero and a second solution due to the discrete symmetry of the PDF as mentioned in Section 6.2.1 may be spotted in the likelihood profiles. They can be excluded in the bifurcated parabola fit. As seen from the plots, the likelihood surface for some parameters is very asymmetric and the bifurcated parabola fit struggles to describe it precisely.

In order to study the effect that the statistics of data samples have on the likelihood surface, a pseudoexperiment with 10 times the current statistics is generated and fit back. This pseudoexperiment has the same yield as what is expected after the full LHCb Run 3 data-taking period. The likelihood profiles of two right-handed amplitude parameters, $\text{Im}(\alpha_{\parallel}^R)$

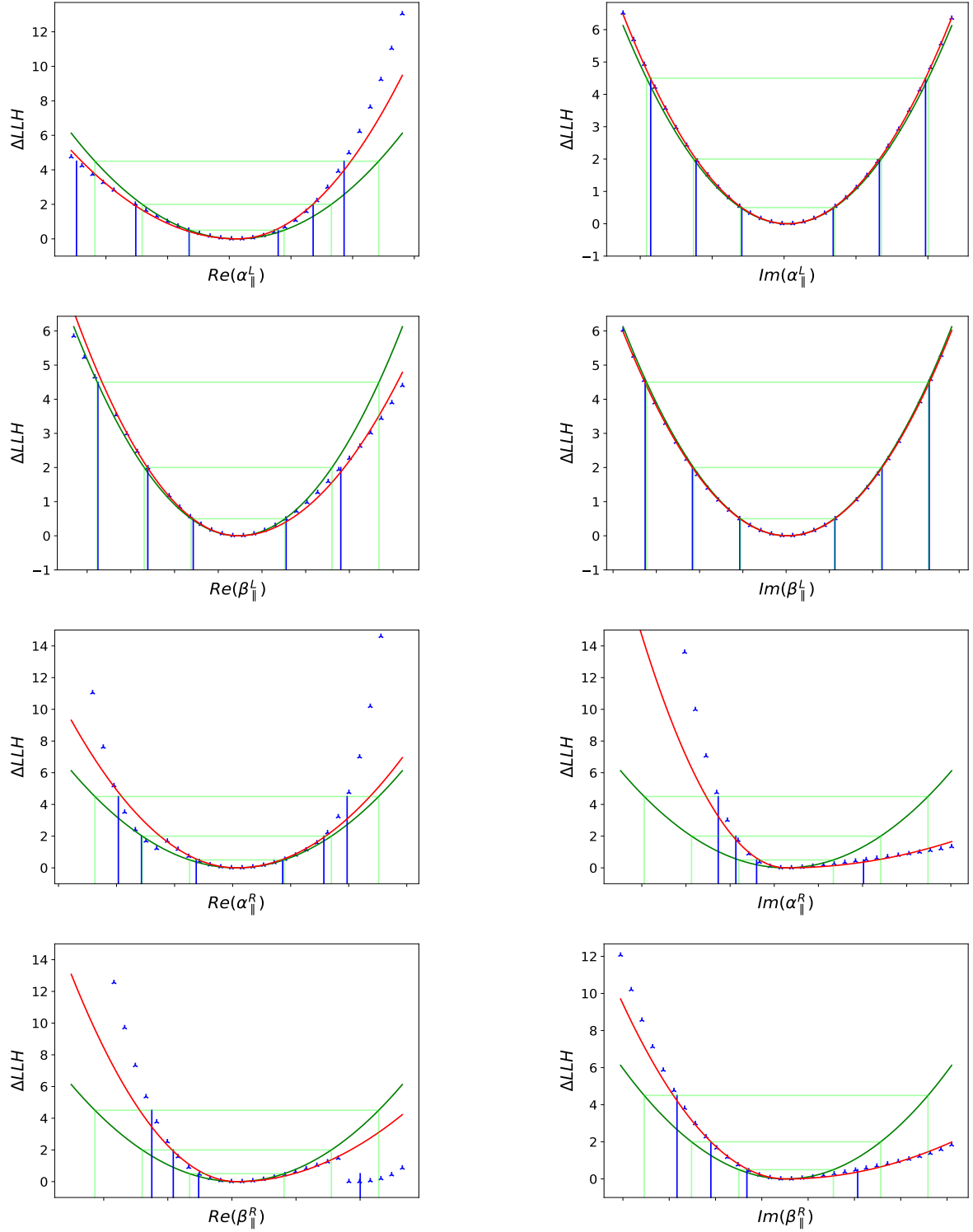


Figure 7.4: Likelihood profiles of the P-wave amplitude coefficients in the fit in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. The blue points represent the profiles. The green and blue lines indicate the 1, 2 and 3σ errors predicted by Hesse and the profiles respectively. The green parabola corresponds to the Hessian error matrix, and the red parabolas are fits of these profiles to bifurcated parabolas. Note the x axis is currently blinded.

7.2. FIT TO RUN 1 + 2 DATA IN THE $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ REGION

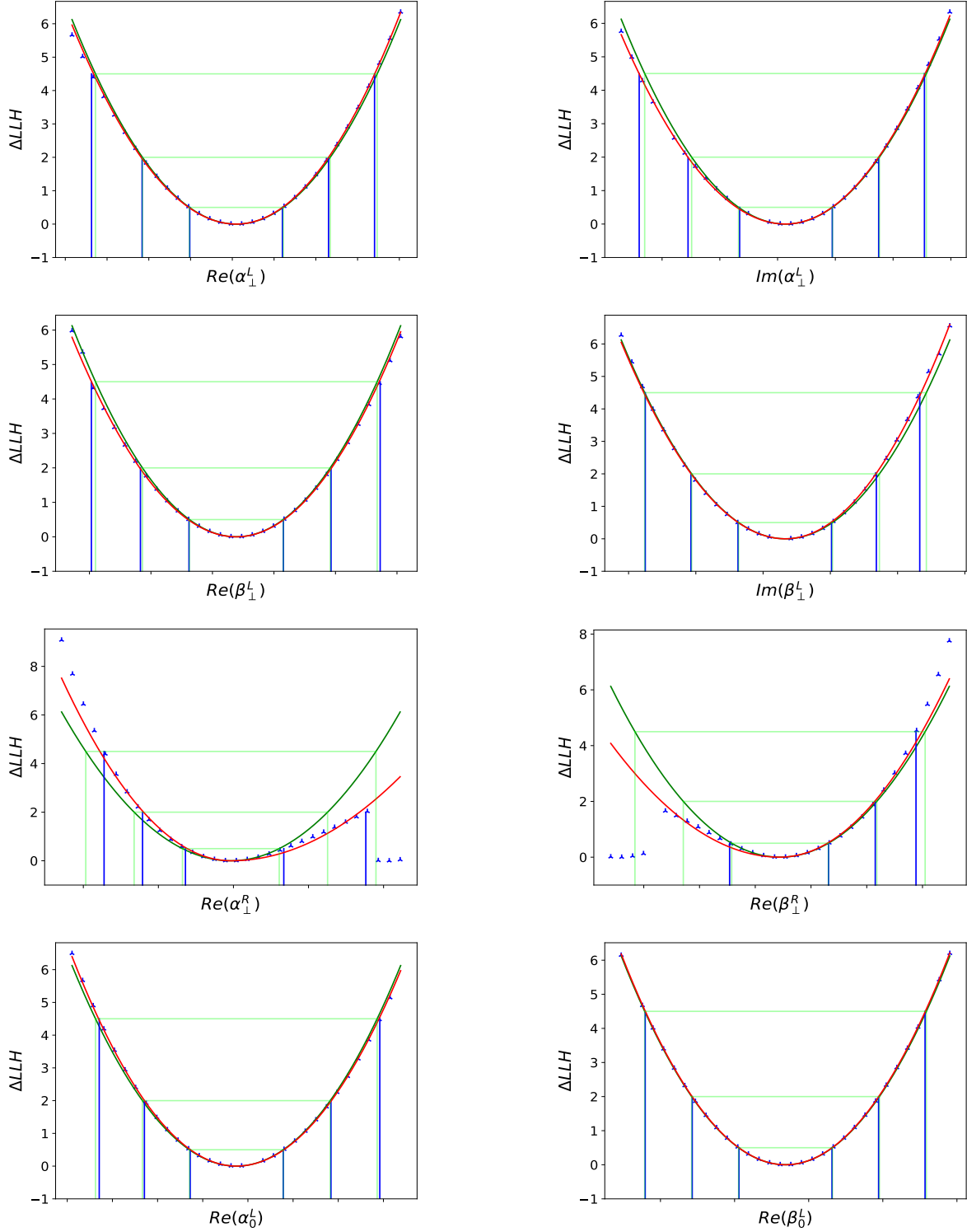


Figure 7.4 continued: Likelihood profiles of the P-wave amplitude coefficients in the fit in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. The blue points represent the profiles. The green and blue lines indicate the 1, 2 and 3σ errors predicted by Hesse and the profiles respectively. The green parabola corresponds to the Hessian error matrix, and the red parabolas are fits of these profiles to bifurcated parabolas. Note the x axis is currently blinded.

and $\text{Re}(\beta_{\parallel}^R)$, are examined as an example and are shown in Figure 7.5. The two parameters are chosen inspired by the very asymmetric shapes of their likelihood profiles and a second solution due to the discrete symmetry of the PDF, as was shown in Figure 7.4. It can be seen that for a pseudoexperiment with 10 times the current statistics, the shape of likelihood profiles becomes less asymmetric and can be better described by a parabola. The second solution can also be well separated from the solution of interest.

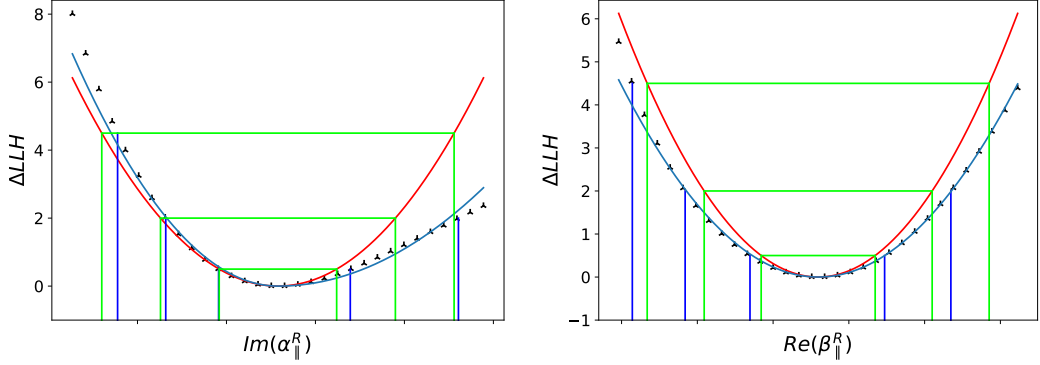


Figure 7.5: Likelihood profiles of two amplitude coefficients in the fit to a pseudoexperiment with $10\times$ current statistics in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. The blue points represent the profiles. The green and blue lines indicate the 1, 2 and 3σ errors predicted by Hesse and the profiles respectively. The green parabola corresponds to the Hessian error matrix, and the red parabolas are fits of these profiles to bifurcated parabolas. Note the x axis is currently blinded.

7.3 Statistically sampling the amplitudes with bifurcated Gaussians

With all the information provided by the fitter, pseudodata can be sampled for all of the amplitude parameters, and the ensemble of amplitudes and observables can be calculated. However as suggested by the likelihood profiles (see Figure 7.4) in both q^2 regions, the likelihood surface cannot be correctly described by Hesse. The errors are hence calculated by fitting bifurcated parabolas to likelihood profiles of individual amplitude parameters. Therefore, the likelihood surface of an individual amplitude parameter is described by a bifurcated Gaussian.

The best-fit values of each amplitude parameter will be taken as the central values of the bifurcated Gaussian, denoted as μ_i , while the bifurcated uncertainties are denoted as σ_i^{low} and σ_i^{high} . The correlation matrix obtained from Hesse is used and written as corr_{ij} between the i th and j th elements. A set of fluctuated amplitude parameters, α_i , may be sampled from the following PDF with the definitions above:

$$(7.1) \quad PDF(\mathbf{x}) = \mathcal{N} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T C^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

where $\mathcal{N}$ stands for the normalisation factor of the PDF, and C is the covariance matrix defined by

$$(7.2) \quad C_{ij} = \text{corr}_{ij} \sigma_i \sigma_j ,$$

in which σ represents the bifurcated errors

$$(7.3) \quad \sigma_i = \begin{cases} \sigma_i^{\text{low}} & x_i - \mu_i < 0 \\ \sigma_i^{\text{high}} & x_i - \mu_i \geq 0 \end{cases} .$$

These fluctuated amplitude parameters may then be used to construct the observables at any point in q^2 .

7.3.1 Amplitudes and observables in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region

Pseudodata in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region are sampled and amplitudes are constructed for the ensemble of the pseudodata, as shown in Figure 7.6 and 7.7 respectively. There are 10,000 pseudodata generated in total, and the intervals containing 68% and 95% of the events, corresponding to 1σ and 2σ intervals, are computed and presented in the figures below. The median of the ensemble is compared with the best-fit result computed directly from the fit itself. Overall, no large biases are spotted.

The observables are also constructed and are presented in the same way in Figure 7.7. No large biases are spotted either when comparing the median with best-fit values. Although there seem to be relatively large discrepancies in the plot of S_4 , it is not a genuine bias. This is studied with pseudoexperiments as outlined in Appendix C.

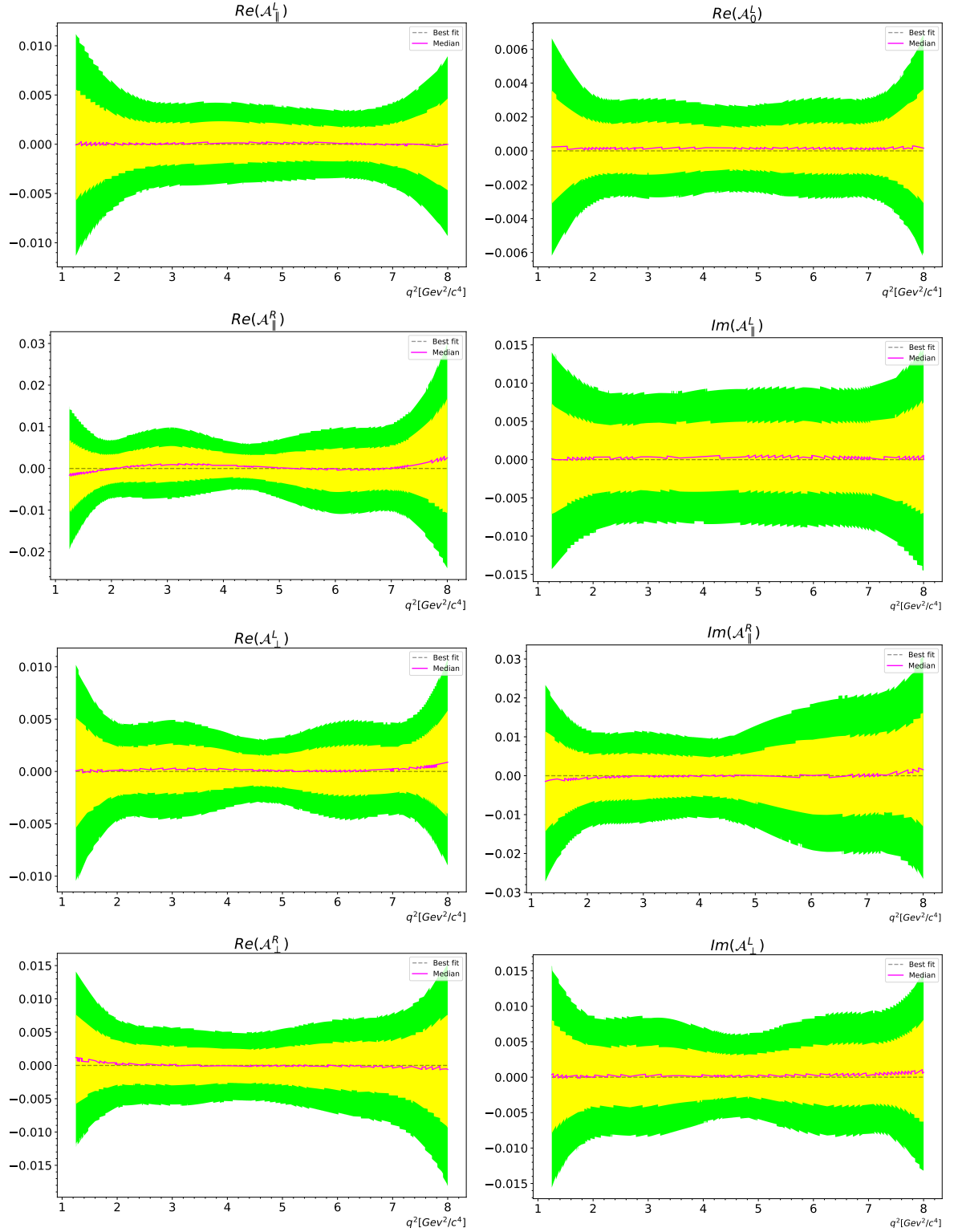


Figure 7.6: The ensemble of amplitudes calculated based on the bifurcated Gaussian sampling from the data fit. The yellow bands represent the 68% intervals, the green bands represent the 95% intervals and the magenta lines are the medians of the ensemble. All values are shifted by the best-fit points such that the best-fit points are located at 0 for blinding purposes.

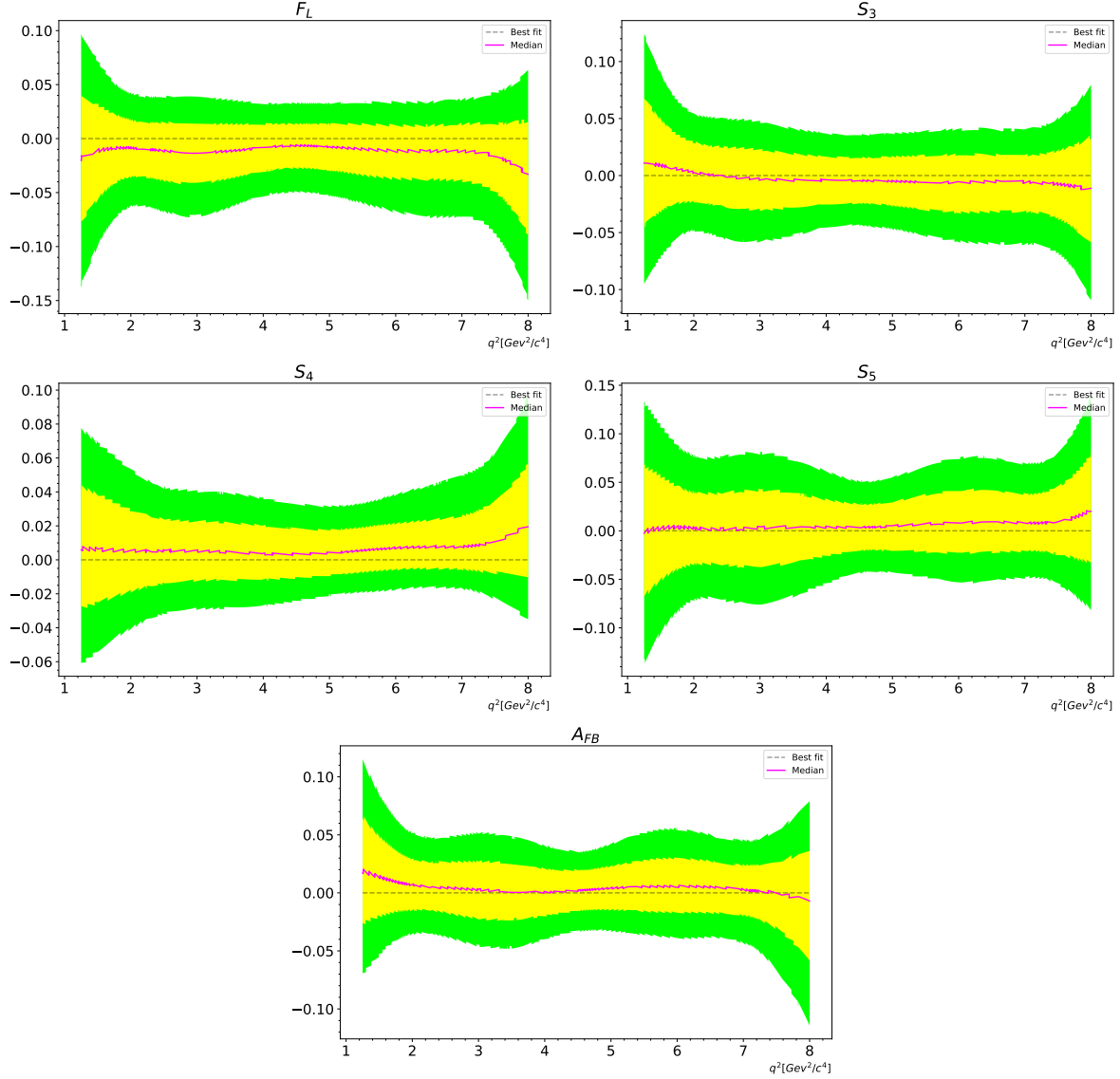


Figure 7.7: The ensemble of observables calculated based on the bifurcated Gaussian sampling from the data fit. The yellow bands represent the 68% intervals, the green bands represent the 95% intervals and the magenta lines are the medians of the ensemble. All values are shifted by the best-fit points such that the best-fit points are located at 0 for blinding purposes.

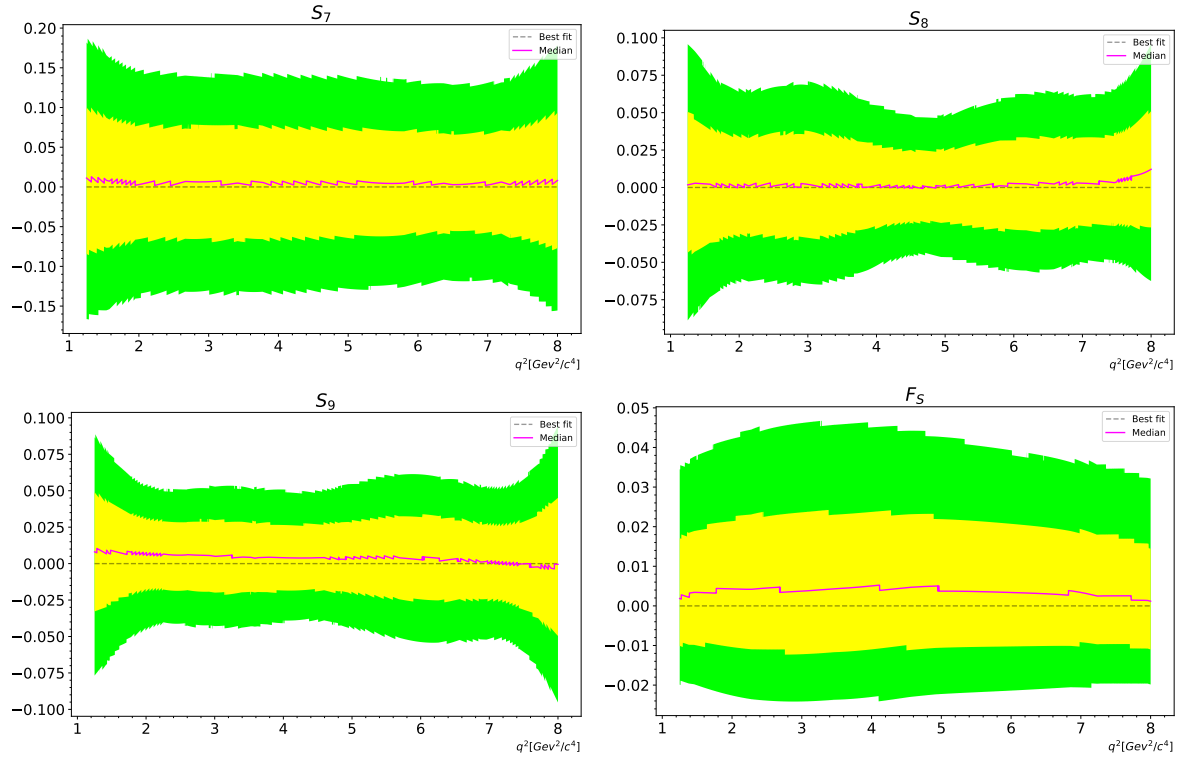


Figure 7.7 continued: The ensemble of observables calculated based on the bifurcated Gaussian sampling from the data fit. The yellow bands represent the 68% intervals, the green bands represent the 95% intervals and the magenta lines are the medians of the ensemble. All values are shifted by the best-fit points such that the best-fit points are located at 0 for blinding purposes.

7.3.2 Amplitudes and observables in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region

Pseudodata in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region are sampled as well to construct the amplitudes and observables, as illustrated in Figure 7.8 and 7.9 respectively. There are also 10,000 pseudodata generated in total with 1σ and 2σ intervals calculated. The median of the ensemble is compared with the best-fit result. It can be seen that biases exist to different extents in most of the observables and the largest one lies in F_S .

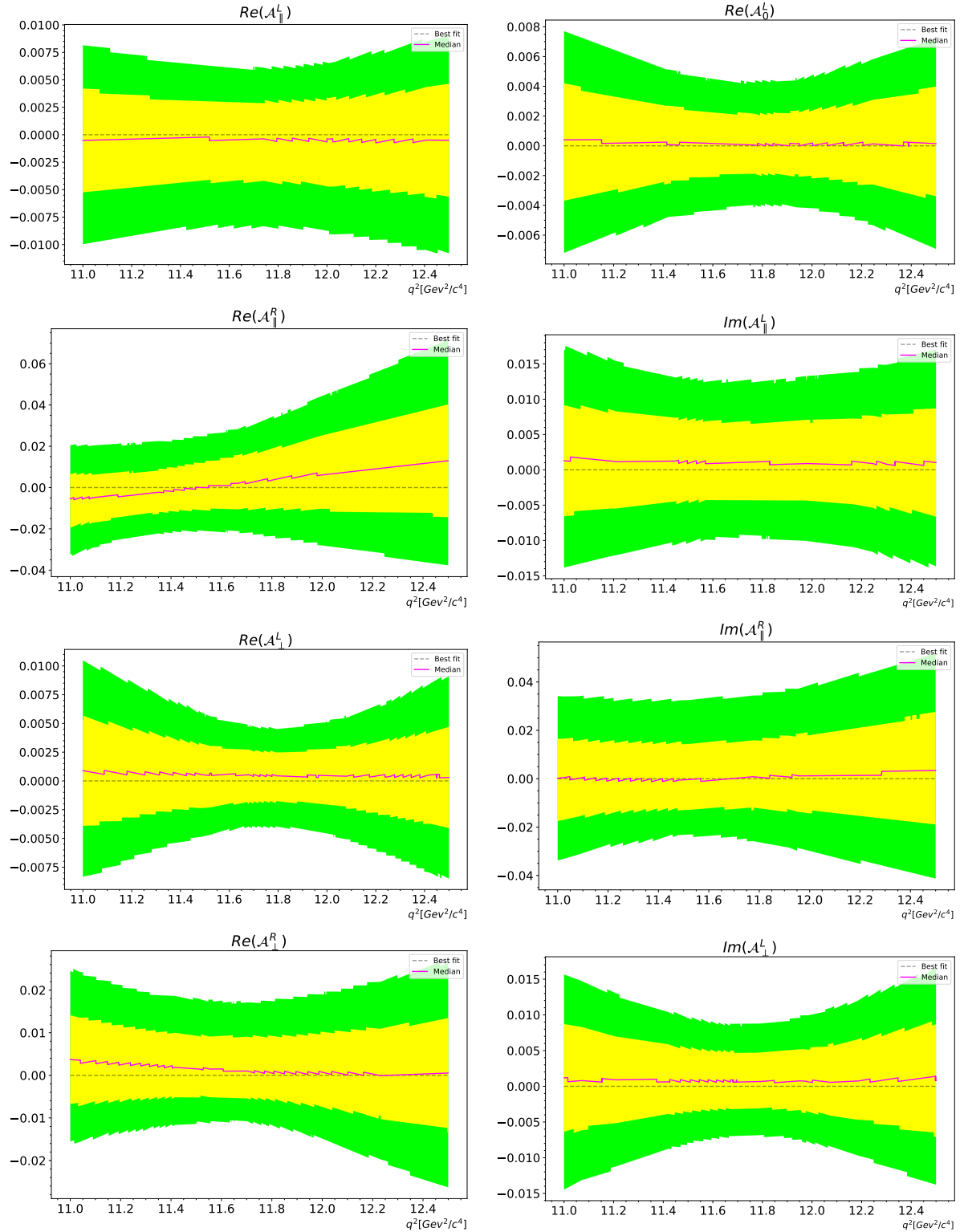


Figure 7.8: The ensemble of amplitudes calculated based on the bifurcated Gaussian sampling from the data fit. The yellow bands represent the 68% intervals, the green bands represent the 95% intervals and the magenta lines are the medians of the ensemble. All values are shifted by the best-fit points for blinding purposes.

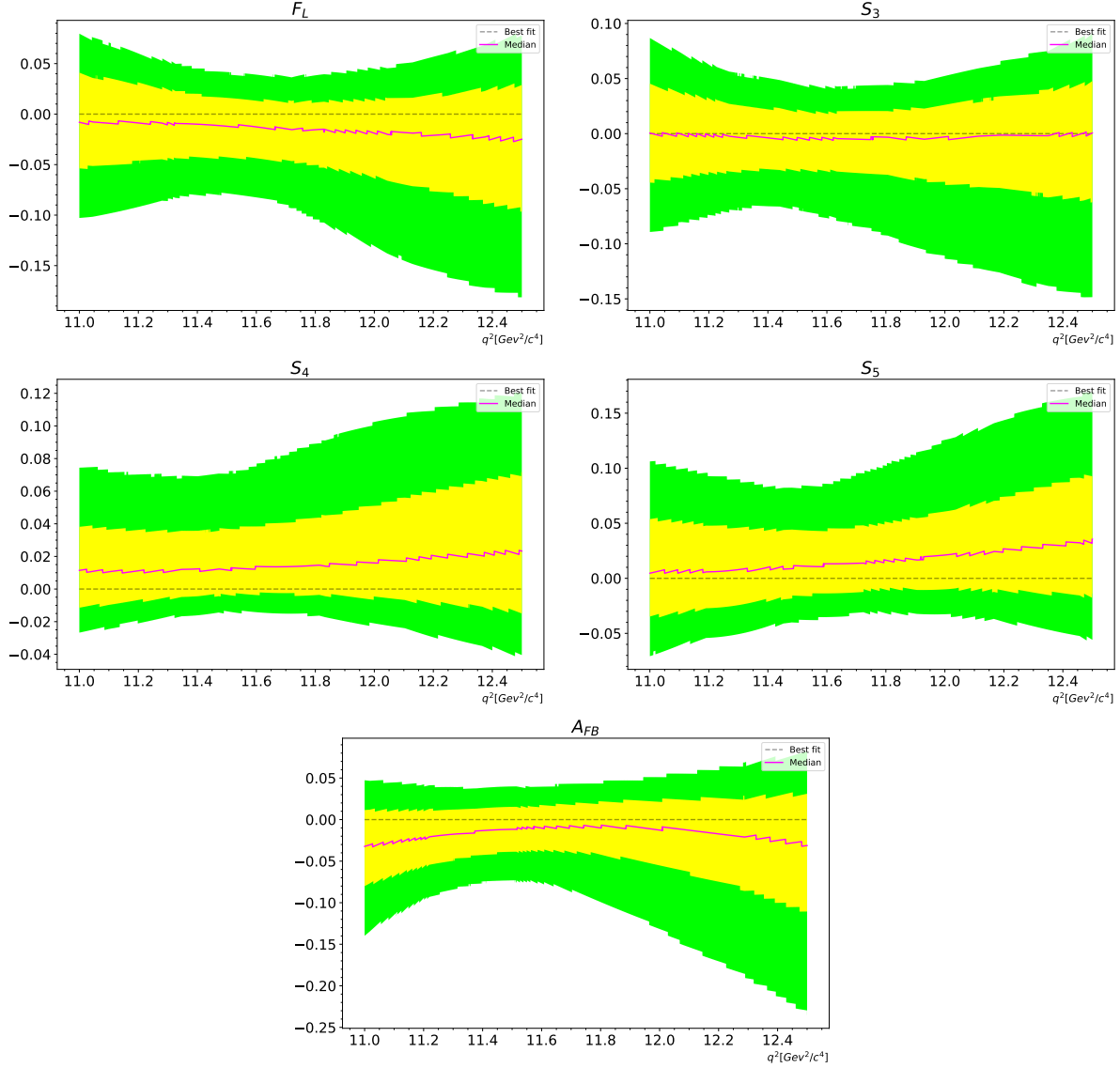


Figure 7.9: The ensemble of observables calculated based on the bifurcated Gaussian sampling from the data fit. The yellow bands represent the 68% intervals, the green bands represent the 95% intervals and the magenta lines are the medians of the ensemble. All values are shifted by the best-fit points such that the best-fit points are located at 0 for blinding purposes.

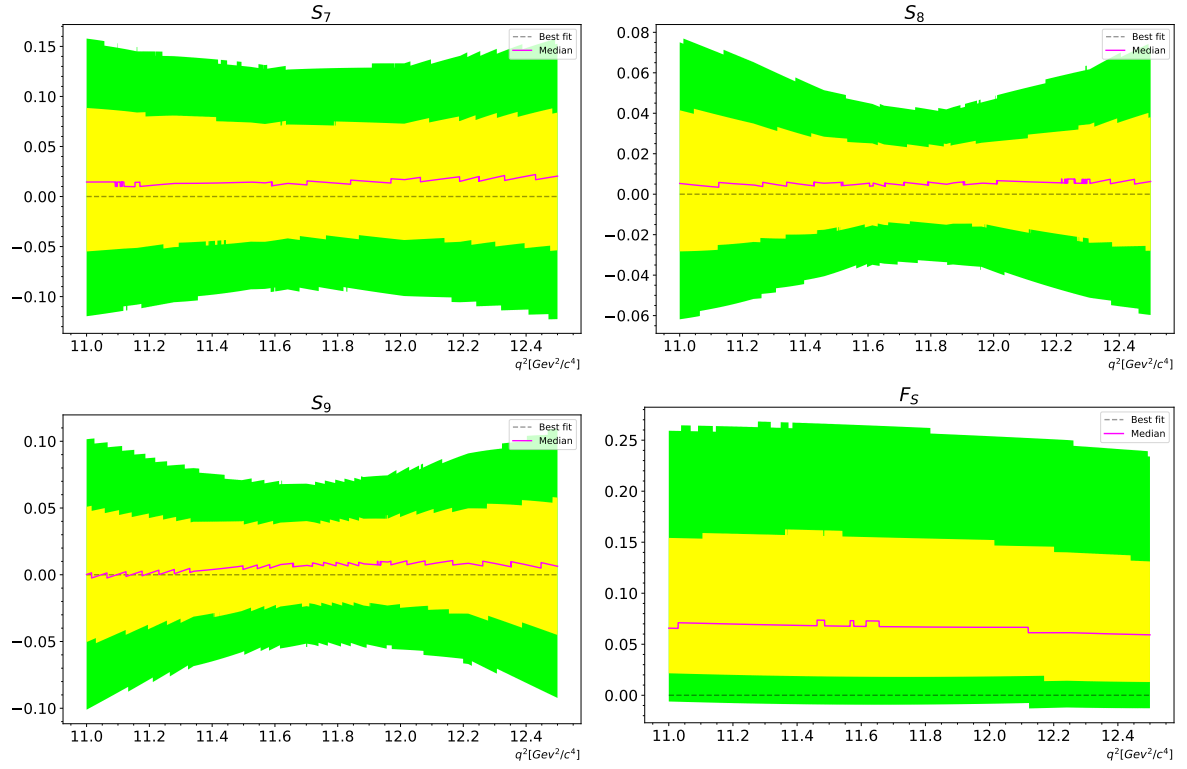


Figure 7.9 continued: The ensemble of observables calculated based on the bifurcated Gaussian sampling from the data fit. The yellow bands represent the 68% intervals, the green bands represent the 95% intervals and the magenta lines are the medians of the ensemble. All values are shifted by the best-fit points such that the best-fit points are located at 0 for blinding purposes.

7.4 Global fits of Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$

The previous section provides individual angular observables predicted by the fit. However, they alone are insufficient to judge the compatibility with theory model predictions as the observables have correlations with each other. An efficient way is to translate the information from observables to Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$. Synthetic data samples with high statistics, i.e. 1000 times the yield of current data samples available, can be generated from the fit result of this analysis following the method mentioned in the previous section. These synthetic data samples account for acceptance effects with background subtracted, so they are a good representation of the signal mode experimentally. Software packages such as FLAVIO [60] can then be used to carry out fits to the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ with these high statistics synthetic datasets by varying the value of theory parameters. The ensemble of such fits is utilised to determine a contour in the two-dimensional space of $\mathcal{C}_9$ and $\mathcal{C}_{10}$, which can be further compared with SM predictions of the values of $\mathcal{C}_9$ and $\mathcal{C}_{10}$. Such global fits of $\mathcal{C}_9$ and $\mathcal{C}_{10}$ using FLAVIO [60] in different physics scenarios are shown in Figure 7.10.

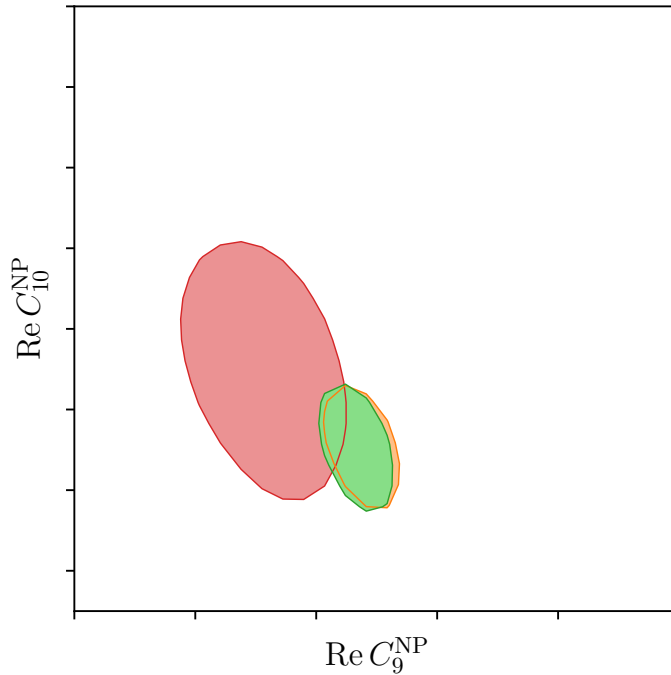


Figure 7.10: Global fits of the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$ to pseudoexperiments generated from fits to LHCb Run 1 + 2 data samples using different physics models. The orange contour shows a global fit to a $10\times$ statistics pseudoexperiment generated from the dispersion model fit [149] setting $C_9^{\text{NP}} = C_{10}^{\text{NP}} = 0$. The green contour shows this for a $10\times$ statistics pseudoexperiment generated from the amplitude ansatz model which was fit to the same dispersion model pseudoexperiment mentioned above. The red contour shows an example $1\times$ statistics pseudoexperiment. The axes are blinded.

Chapter 8

Systematic Uncertainties

The systematic effects of this analysis originate either from the evaluation of the acceptance function or the assumptions of the fit model. At the time this thesis is written, two of the systematic uncertainties predicted to be dominant are evaluated and presented in Section 8.1 and 8.2. They are due to L0 efficiency and kinematics corrections respectively. The estimations are performed in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region. Brief introductions of the remaining systematic uncertainties and how they can be assessed are given in the rest of the chapter. The systematic uncertainties are evaluated using large statistics pseudoexperiments. These large simulation samples are generated with the parameters systematically altered and fit back both with the nominal and systematically altered model. The fit results of the amplitude parameters using the nominal model and systematically altered model are recorded to compute the systematic uncertainties.

The systematic uncertainties are calculated using the root-mean-square deviation (RMSD) in this analysis, which is defined as

$$(8.1) \quad \text{RMSD}(\alpha) = \sqrt{\frac{\sum_N (\hat{\alpha} - \alpha)^2}{N}},$$

where α and $\hat{\alpha}$ denote the fit result of an amplitude parameter α using the nominal model and systematically altered model respectively and N is the total number of pseudoexperiments. Consider a model with a nominal acceptance ϵ_n and an alternative acceptance ϵ_a as an example. The systematically altered acceptance is then defined as

$$(8.2) \quad \epsilon(q^2, \cos \theta_\ell, \cos \theta_K, \phi) = (1 - x) \epsilon_n(q^2, \cos \theta_\ell, \cos \theta_K, \phi) + x \epsilon_a(q^2, \cos \theta_\ell, \cos \theta_K, \phi)$$

where x is the percentage by which the alternative acceptance is mixed with the nominal acceptance. With x evenly distributed within 100%, a set of fits is performed with these systematically altered acceptances that result in a set of $\hat{\alpha}$. This set of $\hat{\alpha}$ is used to estimate the systematic uncertainty of the amplitude parameter α associated with this alternative acceptance along with the result from a fit using the nominal acceptance.

8.1 L0 efficiency

The weights used to model the L0 trigger efficiency may introduce a systematic uncertainty. The nominal L0 trigger efficiency is derived from control mode data. An alternative way to derive the L0 efficiency is to emulate it from the simulated samples directly. The emulated L0 output is directly taken from DAVINCI. The L0 muon p_T thresholds from the TCKs are applied to sections of simulated samples which can represent the proportion of each in the data.

Table 8.1: L0 trigger efficiency associated systematic uncertainties, shown as a fraction of the corresponding statistical uncertainties in percentage (%).

Parameter	syst./ stat. uncertainty(%)	Parameter	syst./ stat. uncertainty (%)
$\text{Re}(\alpha_{\parallel}^L)$	1.524	$\text{Re}(\alpha_{\perp}^L)$	0.788
$\text{Im}(\alpha_{\parallel}^L)$	0.230	$\text{Im}(\alpha_{\perp}^L)$	0.396
$\text{Re}(\beta_{\parallel}^L)$	1.914	$\text{Re}(\beta_{\perp}^L)$	1.113
$\text{Im}(\beta_{\parallel}^L)$	0.473	$\text{Im}(\beta_{\perp}^L)$	0.433
$\text{Re}(\gamma_{\parallel}^L)$	0.532	$\text{Re}(\gamma_{\perp}^L)$	0.901
$\text{Im}(\gamma_{\parallel}^L)$	0.228	$\text{Im}(\gamma_{\perp}^L)$	0.829
$\text{Re}(\delta_{\parallel}^L)$	0.255	$\text{Re}(\delta_{\perp}^L)$	0.730
$\text{Im}(\delta_{\parallel}^L)$	0.122	$\text{Im}(\delta_{\perp}^L)$	0.335
$\text{Re}(\alpha_{\parallel}^R)$	0.732	$\text{Re}(\alpha_{\perp}^R)$	1.654
$\text{Im}(\alpha_{\parallel}^R)$	1.221	$\text{Re}(\beta_{\perp}^R)$	0.723
$\text{Re}(\beta_{\parallel}^R)$	1.979	$\text{Re}(\gamma_{\perp}^R)$	1.270
$\text{Im}(\beta_{\parallel}^R)$	2.035	$\text{Re}(\delta_{\perp}^R)$	0.369
$\text{Re}(\gamma_{\parallel}^R)$	1.254	$\text{Re}(\alpha_0^L)$	3.246
$\text{Im}(\gamma_{\parallel}^R)$	2.153	$\text{Re}(\beta_0^L)$	5.028
$\text{Re}(\delta_{\parallel}^R)$	0.225	$\text{Re}(\gamma_0^L)$	1.419
$\text{Im}(\delta_{\parallel}^R)$	0.385	$\text{Re}(\delta_0^L)$	2.539

An alternative acceptance function is derived with the L0 efficiency derived from simulation with varying mixture fractions to produce a covariance matrix for what is a discrete systematic uncertainty. The nominal acceptance function is mixed with this alternative acceptance function to generate 100 pseudoexperiments, which are fit with two models to compute the systematic uncertainty for kinematic corrections using the RMSD formula. The systematic uncertainties for individual amplitude parameters are outlined in Table 8.1, shown as a fraction of the corresponding statistical uncertainties. It can be seen that this systematic uncertainty is less than 5% of the statistical uncertainty for all of the parameters and is thus negligibly small.

8.2 Kinematic corrections

As mentioned in Section 5.1.3.2, $B^+ \rightarrow J/\psi K^+$ is used as the control mode to derive corrections for the differences between data and simulation. However, there are uncertainties associated with the method used to derive the corrections, the statistics of the control mode samples and the fact that $B^+ \rightarrow J/\psi K^+$ is chosen to be the control mode. A systematic uncertainty is therefore assessed due to these corrections. $B^0 \rightarrow J/\psi K^{*0}$, which is widely used in the literature of similar analyses, is used as the alternative control mode to derive another set of weights for corrections to simulation. An alternative acceptance function is derived with these new weights. Again the nominal acceptance function is mixed with this alternative acceptance function in varying mixture fractions to generate pseudoexperiments. 100 pseudoexperiments are generated in total and are fit with two models to compute the systematic uncertainty for kinematic corrections using the RMSD formula.

The systematic uncertainties for individual amplitude parameters are summarised in Table 8.2 as a fraction of the corresponding statistical uncertainties.

8.3 Tracking corrections

The tracking efficiency corrections are applied to simulation samples and matched to data. However, the tracking correction table has its own statistical and systematic uncertainties which introduce a systematic uncertainty to the fit result of this analysis. The tracking table can be varied within its uncertainties with which new weights can be computed. An alternative acceptance is then computed with the new weights and used to evaluate this systematic uncertainty following the same method illustrated in the previous sections. This systematic uncertainty is expected to be tiny with respect to the statistical uncertainty.

8.4 Acceptance simulation statistics

The simulation samples used to derive the acceptance function have limited statistics and thus result in a systematic uncertainty. This systematic uncertainty can be addressed by

Table 8.2: Kinematic correction associated systematic uncertainties, shown as a fraction of the corresponding statistical uncertainties in percentage (%).

Parameter	syst./ stat. uncertainty (%)	Parameter	syst./ stat. uncertainty (%)
$\text{Re}(\alpha_{\parallel}^L)$	1.079	$\text{Re}(\alpha_{\perp}^L)$	0.494
$\text{Im}(\alpha_{\parallel}^L)$	0.165	$\text{Im}(\alpha_{\perp}^L)$	0.456
$\text{Re}(\beta_{\parallel}^L)$	1.178	$\text{Re}(\beta_{\perp}^L)$	0.792
$\text{Im}(\beta_{\parallel}^L)$	0.380	$\text{Im}(\beta_{\perp}^L)$	0.550
$\text{Re}(\gamma_{\parallel}^L)$	0.539	$\text{Re}(\gamma_{\perp}^L)$	0.919
$\text{Im}(\gamma_{\parallel}^L)$	0.263	$\text{Im}(\gamma_{\perp}^L)$	0.631
$\text{Re}(\delta_{\parallel}^L)$	0.224	$\text{Re}(\delta_{\perp}^L)$	0.725
$\text{Im}(\delta_{\parallel}^L)$	0.132	$\text{Im}(\delta_{\perp}^L)$	0.391
$\text{Re}(\alpha_{\parallel}^R)$	0.741	$\text{Re}(\alpha_{\perp}^R)$	1.509
$\text{Im}(\alpha_{\parallel}^R)$	1.178	$\text{Re}(\beta_{\perp}^R)$	0.576
$\text{Re}(\beta_{\parallel}^R)$	2.090	$\text{Re}(\gamma_{\perp}^R)$	0.855
$\text{Im}(\beta_{\parallel}^R)$	1.808	$\text{Re}(\delta_{\perp}^R)$	0.312
$\text{Re}(\gamma_{\parallel}^R)$	0.807	$\text{Re}(\alpha_0^L)$	2.852
$\text{Im}(\gamma_{\parallel}^R)$	2.319	$\text{Re}(\beta_0^L)$	4.347
$\text{Re}(\delta_{\parallel}^R)$	0.244	$\text{Re}(\gamma_0^L)$	1.290
$\text{Im}(\delta_{\parallel}^R)$	0.318	$\text{Re}(\delta_0^L)$	2.468

bootstrapping the simulation samples, with which an alternative acceptance function is derived. Similar methods following that described in Section 8.1 and 8.2 are used to assign a number. This systematic uncertainty is expected to be negligible compared to the statistical uncertainty. The independence of the alternative acceptance in each toy generated is guaranteed by independently and randomly sampling from the original simulation sample.

8.5 $m_{K\pi}$ model

The lineshape of the S-wave is not fully understood. As the S-wave amplitudes in this analysis are parameterised by constant ansatzes, their variances are not taken into account by the flat lineshape. As the S-wave amplitudes are essentially treated as nuisance parameters, the systematic uncertainties of the S-wave amplitude parameters are beyond the scope of this analysis. Parameterising individual S-wave amplitudes with constants and ignoring their variances may impact the P-wave amplitudes and will result in a systematic uncertainty, which is studied quantitatively below.

Flat S-wave amplitudes affect P-wave amplitudes at the second-order level. The q^2 dependence of the S-wave amplitudes looks very similar to the q^2 dependence of the longitudinal P-wave amplitudes $\mathcal{A}_0^{L,R}$ [38, 47], and the $\mathcal{A}_0^L$ and $\mathcal{A}_{00}^L$ amplitudes are discussed below as an example. As can be seen in Figure 7.6 the biggest 1σ statistical variation over the full q^2 range is approximately 5% of its best-fit value for $\mathcal{A}_0^L$ (at around -0.065), so its lineshape is reasonably flat and so is the lineshape of $\mathcal{A}_{00}^L$. For $\mathcal{A}_{00}^L$, the biggest 1σ statistical variation over the full q^2 range is approximately 30% as shown in Figure 8.1. Hence the statistical power of the current data samples is not big enough to have the precision to tell the q^2 dependence in the shape of S-wave amplitudes. Since the effect is roughly at the level of 30% of statistical uncertainty of the S-wave amplitude, it is far less below 30% of the statistical uncertainty for P-wave amplitudes at the second order. Therefore, this systematic uncertainty is still negligible compared to the statistical uncertainty.

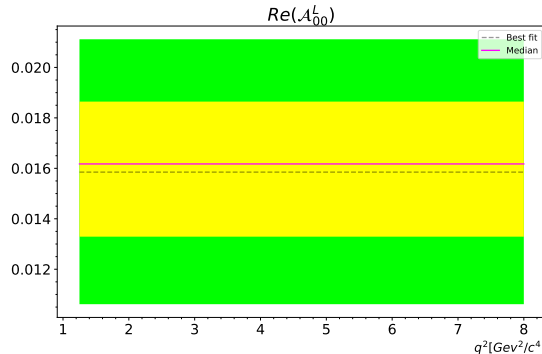


Figure 8.1: The ensemble of the S-wave amplitude $\mathcal{A}_{00}^L$ calculated based on the bifurcated Gaussian sampling from the data fit. The yellow bands represent the 68% intervals, the green bands represent the 95% intervals and the magenta lines are the medians of the ensemble.

8.6 B^0 mass model

As mentioned in Section 6.7, large pulls can be spotted in the mass fits to $B^0 \rightarrow J/\psi K^{*0}$ control mode data, as shown in Figures 6.9 - 6.12. An alternative model, such as two double Crystal Ball functions, can be used to describe the B^0 mass lineshape and evaluate a systematic uncertainty.

Chapter 9

Conclusions and Outlook

This thesis presents an amplitude analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay with a q^2 -dependent Legendre polynomial ansatz modelling the amplitudes. The data used are collected by the LHCb experiment during its Run 1 and Run 2 data-taking periods. The analysis presents the best-fit values of all amplitude parameters, the correlation matrix from Hesse and the bifurcated statistical errors of individual amplitude parameters derived by fitting bifurcated parabolas to the likelihood profiles. Two of the systematic uncertainties that were predicted to be dominant are evaluated to be negligible at the time this thesis is written, and the full set of systematic uncertainties will be estimated eventually and combined to construct a covariance matrix for all systematic effects. Unlike the other analyses of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay where angular observables and Wilson coefficients are determined using different theory models [149–151], this analysis provides a model-independent description of experimental data with acceptance effects accounted for and background subtracted. Synthetic data samples can be generated and theorists can use any theory model to fit the synthetic data samples to construct observables and determine the Wilson coefficients $\mathcal{C}_9$ and $\mathcal{C}_{10}$.

The fit is performed in two rare mode regions in q^2 , one in the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region and one in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region. The robustness of the fitter is tested on $B^0 \rightarrow J/\psi K^{*0}$ control mode samples in the $9.223 < q^2 < 9.966 \text{ GeV}^2/c^4$ region and with toy study in the rare modes. The observables constructed from the control mode fit demonstrate good consistency between data taken in different years, and the toy study shows good stability of the fitter. In the $1.25 < q^2 < 8.0 \text{ GeV}^2/c^4$ region, the method to derive the bifurcated errors is validated by coverage tests with toy studies, and the error intervals constructed from the best-fit values and bifurcated errors suggest no large bias of the fit. It is thus a good description of the experimental data. In the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region, although the bifurcated errors give a coverage closer to the correct one, it is not sufficient to describe the true likelihood surface. This conclusion is further supported by the biases when examining the error intervals constructed from the best-fit values and bifurcated errors. One possible reason is that the shape of the likelihood profiles cannot be well described by the bifurcated parabolas and

the likelihood surface is thus not Gaussian. Another problem is the modelling of the S-wave components, as suggested by the error interval plot of the F_S observable. Therefore, the fit in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region needs further investigation.

Looking into the future, there are exciting prospects that could be accomplished within the framework of this analysis. Albeit with the delay due to COVID-19, the LHC has already restarted to take data in 2023 after the Long Shutdown 2, marking a start of the scheduled four-year long Run 3 data-taking period. Thanks to the enhanced integrated luminosity and upgraded detectors, there could be 10 times the current data yield expected at the end of Run 3. This analysis clearly benefits from more data samples. Since the global fits for Wilson coefficients are scaled by $1/\sqrt{N}$ where N is the yield of data samples, more data allows for a more precise determination of the $\mathcal{C}_9$ and $\mathcal{C}_{10}$ contour, as illustrated in Figure 7.10. Moreover, the S-wave amplitudes are treated as nuisance parameters in the analysis currently and the $m_{K\pi}$ dependence in the differential decay rate is integrated. With more data samples available, the S-wave amplitudes can also be modelled using the same polynomial ansatz method and the fit can be performed in a five-dimensional configuration by adding the $m_{K\pi}$ dependence, lifting the remaining model dependency of the S-wave. More data samples in the $11.0 < q^2 < 12.5 \text{ GeV}^2/c^4$ region will also result in a likelihood surface of the amplitude parameters whose shape is expected to be less asymmetric and can be better described with the method described in this analysis. Finally, with the Run 3 dataset, there would be enough data samples to be split into two subsamples corresponding to B^0 and $\bar{B}^0$, with which a fit in the CP-asymmetric configuration with good statistical coverage can be performed. This analysis will then have the power to determine CP-asymmetric observables from the fit results.

In a wider picture of the $b \rightarrow s\ell^+\ell^-$ transitions, the anomalies still persist and need further measurement. The latest measurement on the magnetic moment of the muon by Fermilab [152] suggests that there are still problems yet understood in the muon channels and updates of the measurements mentioned in Section 2.5 are highly anticipated. More data may also clear the route of new analyses that were considered impracticable before. The Run 3 of the LHCb experiment will hopefully shed light on whether NP is observed in the $b \rightarrow s\ell^+\ell^-$ transitions.

Appendix A

Definition of a Crystal Ball function

A Crystal Ball function is a Gaussian function with a power-law tail on one side. The mathematical form of a Crystal Ball function is given by

$$(A.1) \quad P_{CB}(m; \mu, \sigma, \alpha, n) = N \begin{cases} A(B - \delta)^{-n} & \text{for } \delta < -|\alpha| \\ \exp(-\delta^2/2) & \text{for } \delta \geq -|\alpha| \end{cases},$$

where

$$(A.2) \quad \delta = \frac{m - \mu}{\sigma},$$

$$(A.3) \quad A = \left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{|\alpha|^2}{2}\right),$$

$$(A.4) \quad B = \frac{n}{|\alpha|} - |\alpha|.$$

Here N is a constant that normalises the Crystal Ball function and μ denotes the central mass.

The function describing the B^0 mass lineshape in this analysis is the summation of two Crystal Ball functions with tails on opposite sides, written as

$$(A.5) \quad \mathcal{P}_{\text{sig}}(m_{B^0}) = f_{\text{core}} P_{CB}(m_{B^0}; \mu, \sigma_1, \alpha_1, n_1) + (1 - f_{\text{core}}) P_{CB}(m_{B^0}; \mu, \sigma_2, \alpha_2, n_2).$$

In the rare mode fit, the fraction parameter f_{core} , along with the tail parameters α_1 , α_2 , n_1 and n_2 , are fixed to the fits to statistically large simulation samples, whereas μ , σ_1 and σ_2 parameters are floated.

Appendix B

Goodness-of-fit

This section largely follows Ref. [122] and all figures are taken from it.

In the evaluation of the acceptance function, there are two sources that lead to uncertainties pertaining to the acceptance function. The first is the limited size of the simulation samples and the second is the choice of orders of the polynomials to fit those simulation samples. If the orders are too low, then the polynomials cannot accurately model the shape of the samples; if the orders are too high, the polynomials would end up overfitting the samples and capturing the statistical fluctuations, which are correlated with the size of the simulation samples. Inspecting the one- or two-dimensional projections of the resulting acceptance function only is not sufficient to study the possible bias, because there might be mismodelling in the multi-dimensional space (four-dimensional in this case) that is not captured when projecting onto a lower dimension. A χ^2 -test statistic is insufficient either due to the relatively high order in dimensions. Therefore, a method to test the goodness-of-fit using boosted decision trees (BDTs), which are primarily designed for multivariate analysis, is developed.

As the first step, a four-dimensional fit is carried out on the simulation samples with polynomials of the chosen orders and it will produce an analytical polynomial function describing the simulation samples. This function is regarded as the nominal acceptance function with which a toy dataset with a large number of events (e.g. $1000\times$) is generated and treated as the nominal toy. The simulation samples are then resampled by reweighting with a continuous Poisson distribution and an alternative acceptance function is evaluated with the resampled samples. A toy is generated from this alternative acceptance function but with the same number of events ($1\times$) as that of the simulation samples used to derive the acceptance function. This resampling-evaluation-generation process is repeated. BDTs are then trained with the collection of $1\times$ toys and the $1000\times$ toy and optimised to discriminate between them (Toy-vs-Toy). The tuning of the hyperparameters of the BDTs makes use of the python package HYPEROPT [153].

The metric of the performance of the BDTs is the ROC AUC score, which is defined as the area under the Receiver Operating Characteristic curve (ROC AUC). If the ROC AUC score equals to 0.5, the BDT is equivalent to a random guess; if the ROC AUC score equals

to 1, the BDT can perfectly discriminate between the two toys. The ROC AUC scores of all discriminating BDTs are collected. Finally, another BDT with the same hyperparameters is trained on the $1000\times$ toy dataset and the original simulation samples (Toy-vs-Simulation). The ROC AUC score is obtained and compared with the distribution of the collection of those of all discriminating BDTs and a p -value is calculated. A good fit means that the BDT cannot differentiate between toy and simulation. If the fit is bad, the ROC AUC score of Toy-vs-simulation will be larger than the ROC AUC score distribution of Toy-vs-Toy.

An example of such a test is shown in Figure B.1. In this analysis, after reaching the high enough orders that make the acceptance function accurately describe the original simulation samples, the orders of the polynomials in the acceptance function are decreased and examined until the acceptance function can no longer result in a good fit. If the p -value is greater than 5%, the choice of orders is considered to result in an acceptance function that well describes the original simulation samples.

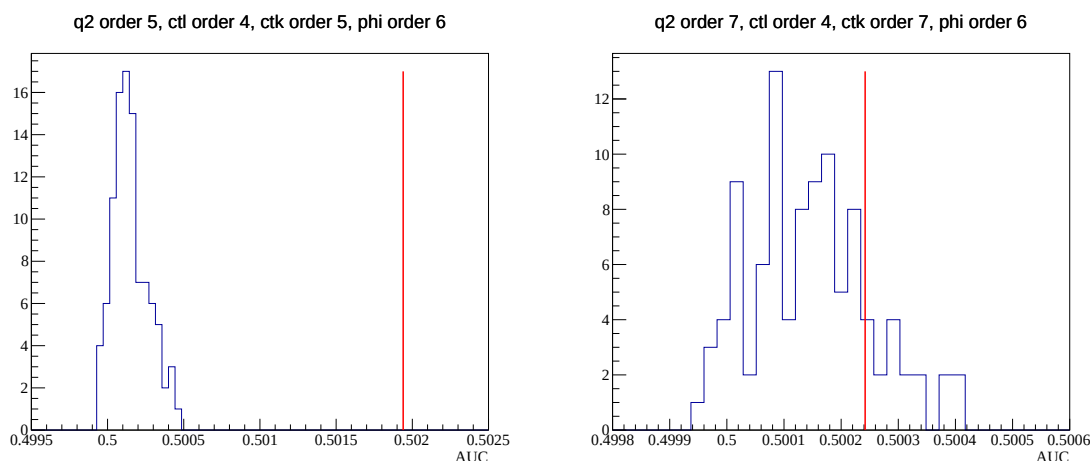


Figure B.1: Example goodness-of-fit plots on the 2016 FLATQ2 simulation. The left plot shows a bad description of the simulation with certain acceptance orders while the right shows a good description with alternative orders. The blue distribution is the ROC AUC scores for Toy-vs-Toy BDTs and the red line is the performance of the Toy-vs-Simulation BDT.

Appendix C

Fit Bias Study Using Toys Generated from SM Predictions

There seems to be an apparent fit bias in the angular observable S_4 . It is studied using pseudoexperiments generated from the SM predictions of the amplitudes in the $1.25 < q^2 < 8.0$ GeV^2/c^4 region. They are fit back based on the primary ansatz used in this analysis (Legendre polynomial four-parameter ansatz). The distribution of the fits for the amplitude $Re(A_{\parallel}^L)$ is shown in Figure C.1, where an apparent bias is also spotted.

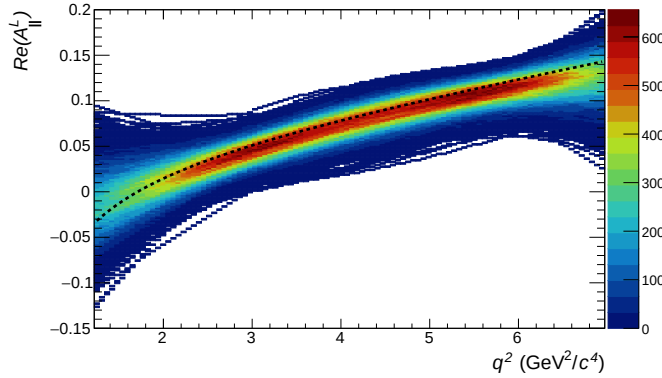


Figure C.1: Distribution of fits of $Re(A_{\parallel}^L)$ to toys generated from the SM amplitude predictions. The black line shows the true value.

A further set of fits is carried out with only $Re(A_{\parallel}^L)$ and the right-handed amplitudes are floated. A two-dimensional plot of angular observables S_3 and S_4 at $q^2 = 6$ is shown in Figure C.2. In this plot, the best-fit point is clearly unbiased. It can then be deduced that these seemingly apparent biases are in fact caused by projecting the physical boundary onto a single observable.

Toys are also generated close to/away from the S_3 - S_4 boundary and fit back. The parameter $Re(A_{\parallel}^L)$ appeared to biased and unbiased respectively, which are shown in Figures C.3 and C.4.

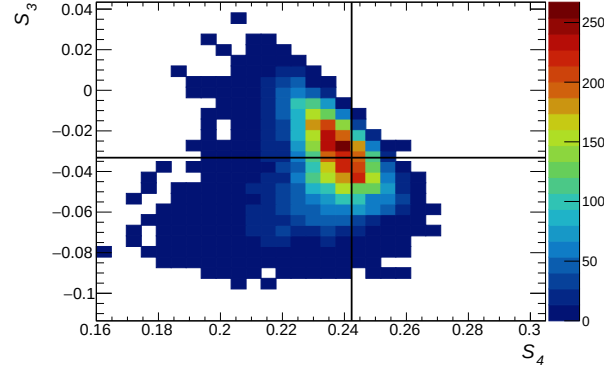


Figure C.2: S_3 and S_4 at $q^2 = 6$ where only $Re(A_{\parallel}^L)$ and the RH amplitudes are floated.

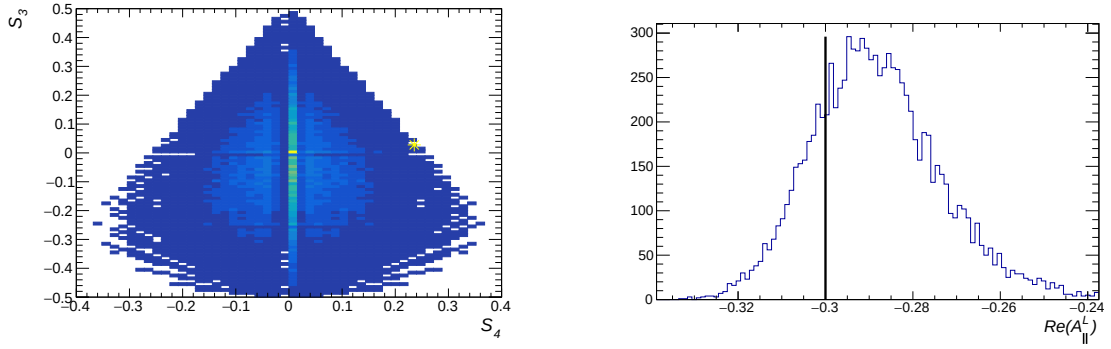


Figure C.3: Toys generated close to the S_3/S_4 boundary from where the yellow point is located (left) and distribution of fits to $Re(A_{\parallel}^L)$ (right) with the true value shown by the black line. There is an apparent bias in this amplitude.

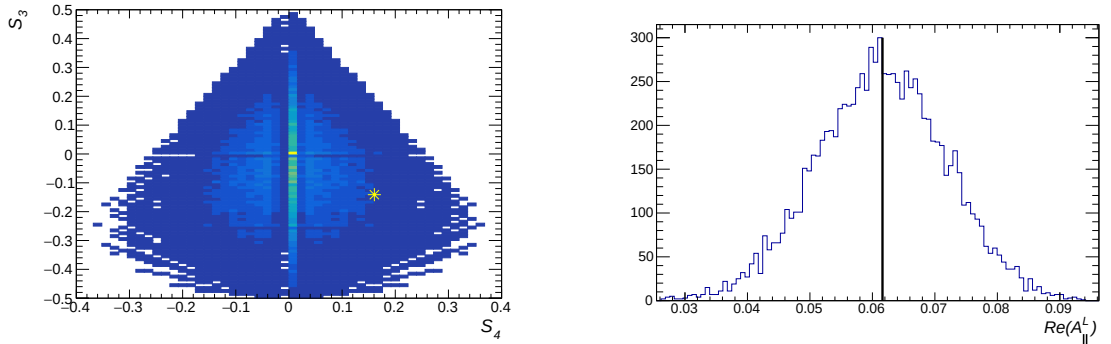


Figure C.4: Toys generated away to the S_3/S_4 boundary from where the yellow point is located (left) and distribution of fits to $Re(A_{\parallel}^L)$ (right) with the true value shown by the black line. This amplitude is not biased.

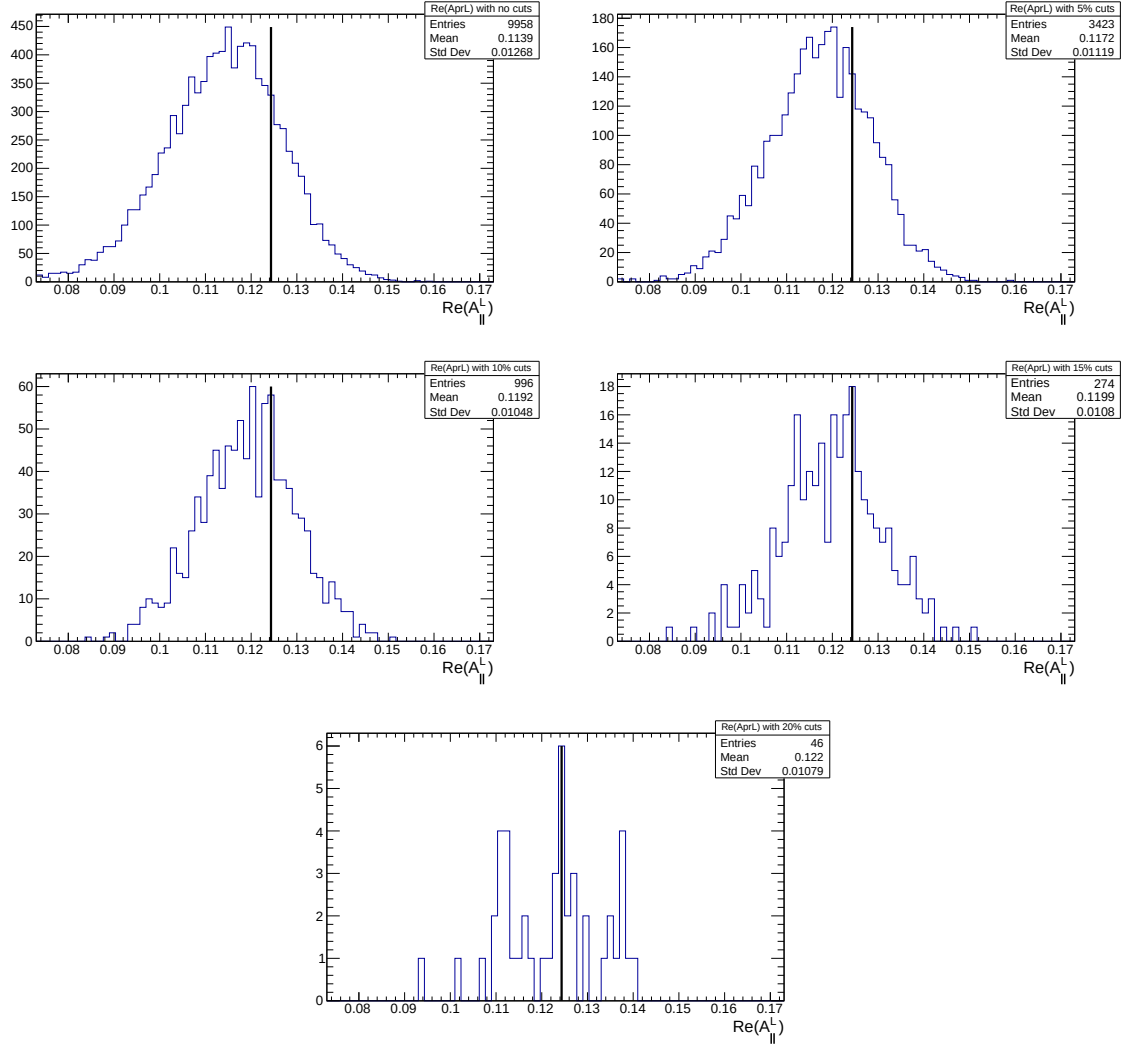


Figure C.5: Distribution of $Re(A_{\parallel}^L)$ in the ensemble of toy fits where no, upper and lower 5%, 10%, 15%, 20% of toys in the other amplitudes are cut out.

Finally, whether there is a genuine bias in the multidimensional space containing all parameters is checked. This is accomplished by selecting the region around the most likely value in the ensemble of toy fits and checking whether the amplitude of interest is unbiased. The lower and upper X% of fits of all other amplitudes are excluded while the amplitude $Re(A_{\parallel}^L)$ is examined. The distribution of the remaining fits is checked to see whether the bias decreases as X increases. Figure C.5 shows the distribution of $Re(A_{\parallel}^L)$ where no, upper and lower 5%, 10%, 15%, 20% of toys are cut out. As seen in this Figure, as the region containing the most likely value shrinks, the apparent bias seen in the one-dimensional distribution of $Re(A_{\parallel}^L)$ reduces. Thus it can be concluded that this apparent bias is not a genuine bias.

Bibliography

- ¹M. Schumann,
“Dark Matter 2014”,
EPJ Web Conf. **96**, edited by G. Giardina, S. Eidelman, G. Venanzoni, M. Battaglieri, and
G. Mandaglio, 01027
(2015).
- ²A. D. Sakharov,
“Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe”,
Soviet Physics Uspekhi **34**, 392
(1991).
- ³G. Aad et al. (ATLAS),
“Observation of a new particle in the search for the Standard Model Higgs boson with the
ATLAS detector at the LHC”,
Phys. Lett. B **716**, 1–29
(2012).
- ⁴S. Chatrchyan et al. (CMS),
“Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”,
Phys. Lett. B **716**, 30–61
(2012).
- ⁵J. P. Lees et al. (BaBar),
“Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \rightarrow$
 $K^{(*)}l^+l^-$ ”,
Phys. Rev. D **86**, 032012
(2012).
- ⁶J.-T. Wei et al. (Belle),
“Measurement of the Differential Branching Fraction and Forward-Backward Asymmetry for
 $B \rightarrow K^{(*)}l^+l^-$ ”,
Phys. Rev. Lett. **103**, 171801
(2009).

- ⁷T. Aaltonen et al. (CDF),
“Observation of the Baryonic Flavor-Changing Neutral Current Decay $\Lambda_b \rightarrow \Lambda \mu^+ \mu^-$ ”,
Phys. Rev. Lett. **107**, 201802
(2011).
- ⁸V. Khachatryan et al. (CMS),
“Angular analysis of the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV”,
Phys. Lett. B **753**, 424–448
(2016).
- ⁹R. Aaij et al. (LHCb collaboration),
“Differential branching fractions and isospin asymmetries of $\rightarrow K^{(*)} \mu^+ \mu^-$ decays”,
JHEP **06**, 133
(2014).
- ¹⁰R. Aaij et al. (LHCb collaboration),
“Angular analysis and differential branching fraction of the decay $B_s^0 \rightarrow \phi \mu^+ \mu^-$ ”,
JHEP **09**, 179
(2015).
- ¹¹R. Aaij et al. (LHCb collaboration),
“Differential branching fraction and angular analysis of $\Lambda_b^0 \rightarrow \Lambda \mu^+ \mu^-$ decays”,
JHEP **06**, 115
(2015).
- ¹²R. Aaij et al. (LHCb collaboration),
“Measurements of the S-wave fraction in $B^0 \rightarrow K^+ \pi^- \mu^+ \mu^-$ decays and the
 $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ differential branching fraction”,
JHEP **11**, 047
(2016).
- ¹³R. Aaij et al. (LHCb collaboration),
“Test of lepton universality with $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays”,
JHEP **08**, 055
(2017).
- ¹⁴R. Aaij et al. (LHCb collaboration),
“Search for lepton-universality violation in $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays”,
Phys. Rev. Lett. **122**, 191801
(2019).
- ¹⁵“Test of lepton universality in $b \rightarrow s \ell^+ \ell^-$ decays”,
(2022).

-
- ¹⁶“Measurement of lepton universality parameters in $B^+ \rightarrow K^+ \ell^+ \ell^-$ and $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays”,
(2022).
- ¹⁷“Measurement of the ratios of branching fractions $\mathcal{R}(D^*)$ and $\mathcal{R}(D^0)$ ”,
(2023).
- ¹⁸A. M. Sirunyan et al. (CMS),
“Measurement of angular parameters from the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV”,
Phys. Lett. B **781**, 517–541
(2018).
- ¹⁹M. Aaboud et al. (ATLAS),
“Angular analysis of $B_d^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”,
JHEP **10**, 047
(2018).
- ²⁰S. Wehle et al. (Belle),
“Lepton-Flavor-Dependent Angular Analysis of $B \rightarrow K^{*0} \ell^+ \ell^-$ ”,
Phys. Rev. Lett. **118**, 111801
(2017).
- ²¹R. Aaij et al. (LHCb collaboration),
“Measurement of CP -averaged observables in the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay”,
Phys. Rev. Lett. **125**, 011802
(2020).
- ²²R. Aaij et al. (LHCb collaboration),
“Angular analysis of the $B^+ \rightarrow K^{*+} \mu^+ \mu^-$ decay”,
to appear in Phys. Rev. Lett.
(2020).
- ²³M. E. Peskin,
“Particle physics: Song of the electroweak penguin”,
Nature **452**, 293–294
(2008).
- ²⁴U. Egede, M. Patel, and K. A. Petridis,
“Method for an unbinned measurement of the q^2 dependent decay amplitudes of $\bar{B}^0 \rightarrow \bar{K}^{*0} \mu^+ \mu^-$ decays”,
JHEP **06**, 084
(2015).

- ²⁵A. Buras,
Gauge Theory of Weak Decays
(Cambridge University Press, June 2020).
- ²⁶M. D. Schwartz,
Quantum Field Theory and the Standard Model
(Cambridge University Press, Mar. 2014).
- ²⁷M. Thomson,
Modern particle physics
(Cambridge University Press, New York, 2013).
- ²⁸M. E. Peskin and D. V. Schroeder,
An Introduction to quantum field theory
(Addison-Wesley, Reading, USA, 1995).
- ²⁹A. V. Manohar and M. B. Wise,
Heavy quark physics,
Vol. 10
(2000).
- ³⁰T. Mannel,
“Effective Field Theories in Flavor Physics”,
Springer Tracts Mod. Phys. **203**, 1–175
(2004).
- ³¹S. T. Cunliffe,
“The scalar component of $B^0 \rightarrow K^+ \pi^- \mu^+ \mu^-$ decays”,
PhD thesis (Imperial Coll., London, 2015).
- ³²N. Cabibbo,
“Unitary symmetry and leptonic decays”,
Phys. Rev. Lett. **10**, 531–533
(1963).
- ³³M. Kobayashi and T. Maskawa,
“CP Violation in the Renormalizable Theory of Weak Interaction”,
Prog. Theor. Phys. **49**, 652–657
(1973).
- ³⁴L. Wolfenstein,
“Parametrization of the kobayashi-maskawa matrix”,
Phys. Rev. Lett. **51**, 1945–1947
(1983).
- ³⁵R. P. Feynman and M. Gell-Mann,

- “Theory of Fermi interaction”,
Phys. Rev. **109**, edited by L. M. Brown, 193–198
(1958).
- ³⁶D. van Dyk,
“The Decays $\bar{B} \rightarrow \bar{K}^{(*)}\ell^+\ell^-$ at Low Recoil and their Constraints on New Physics”,
PhD thesis (Dortmund U., 2012).
- ³⁷R. Aaij et al. (LHCb collaboration),
“Angular analysis of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay using 3fb^{-1} of integrated luminosity”,
JHEP **02**, 104
(2016).
- ³⁸W. Altmannshofer et al.,
“Symmetries and asymmetries of $B \rightarrow K^{*}\mu^+\mu^-$ -decays in the Standard Model and beyond”,
JHEP **01**, 019
(2009).
- ³⁹R. Aaij et al. (LHCb collaboration),
“Measurement of CP asymmetries in the decays $B^0 \rightarrow K^{*0}\mu^+\mu^-$ and $B^+ \rightarrow K^+\mu^+\mu^-$ ”,
JHEP **09**, 177
(2014).
- ⁴⁰J. Gratex, M. Hopfer, and R. Zwicky,
“Generalised helicity formalism, higher moments and the $B \rightarrow K_{JK}(\rightarrow K\pi)\bar{\ell}_1\ell_2$ angular distributions”,
Phys. Rev. D **93**, 054008
(2016).
- ⁴¹R. Aaij et al. (LHCb collaboration),
“Differential branching fraction and angular analysis of the decay $B^0 \rightarrow K^{*0}\mu^+\mu^-$ ”,
JHEP **08**, 131
(2013).
- ⁴²R. Aaij et al. (LHCb collaboration),
“Measurement of form-factor-independent observables in the decay $B^0 \rightarrow K^{*0}\mu^+\mu^-$ ”,
Phys. Rev. Lett. **111**, 191801
(2013).
- ⁴³C. Bobeth, G. Hiller, and G. Piranishvili,
“CP Asymmetries in $\bar{B} \rightarrow \bar{K}^{*}(\rightarrow \bar{K}\pi)\bar{\ell}\ell$ and Untagged $\bar{B}_s, B_s \rightarrow \phi(\rightarrow K^+K^-)\bar{\ell}\ell$ Decays at NLO”,
JHEP **07**, 106
(2008).

- ⁴⁴F. J. Kress,
“Angular analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay at the LHCb experiment”,
PhD thesis (Imperial Coll., London, 2020).
- ⁴⁵F. Kruger, L. M. Sehgal, N. Sinha, and R. Sinha,
“Angular distribution and CP asymmetries in the decays $\bar{B} \rightarrow K^- \pi^+ e^- e^+$ and $\bar{B} \rightarrow \pi^- \pi^+ e^- e^+$ ”,
Phys. Rev. D **61**, [Erratum: Phys.Rev.D 63, 019901 (2001)], 114028
(2000).
- ⁴⁶J. Matias, F. Mescia, M. Ramon, and J. Virto,
“Complete Anatomy of $\bar{B}_d \rightarrow \bar{K}^{*0}(\rightarrow K\pi)\ell^+\ell^-$ and its angular distribution”,
JHEP **04**, 104
(2012).
- ⁴⁷M. Algueró, P. A. Cartelle, A. M. Marshall, P. Masjuan, J. Matias, M. A. McCann, M. Patel, K. A. Petridis, and M. Smith,
“A complete description of P- and S-wave contributions to the $B^0 \rightarrow K^+ \pi \ell^+ \ell^-$ decay”,
JHEP **12**, 085
(2021).
- ⁴⁸K. Chilikin et al. (Belle),
“Experimental constraints on the spin and parity of the $Z(4430)^+$ ”,
Phys. Rev. D **88**, 074026
(2013).
- ⁴⁹D. Aston et al.,
“A study of $K^- \pi^+$ scattering in the reaction $K^- p \rightarrow K^- \pi^+ n$ at 11 GeV/c”,
Nuclear Physics B **296**, 493–526
(1988).
- ⁵⁰J. Bressieux, G. Cowan, M. Kreps, O. Schneider, T. Skwarnicki, and Q. Wenbin,
“Evidence for the resonant character of the $Z(4430)^- \rightarrow \psi(2S)\pi^-$ mass peak observed in $B^0 \rightarrow \psi(2S)K^+\pi^-$ decays, and determination of the $Z(4430)^-$ spin-parity”,
(2013).
- ⁵¹U. Egede, T. Hurth, J. Matias, M. Ramon, and W. Reece,
“New physics reach of the decay mode $\bar{B} \rightarrow \bar{K}^{*0} \ell^+ \ell^-$ ”,
JHEP **10**, 056
(2010).
- ⁵²G. Buchalla,
“Heavy quark theory”,
In 55th Scottish Universities Summer School in Physics: Heavy Flavor Physics (SUSSP 2001)
(Feb. 2002),

- Pp. 57–104.
- ⁵³M. Della Morte, J. Heitger, H. Simma, and R. Sommer,
 “Non-perturbative Heavy Quark Effective Theory: An application to semi-leptonic B-decays”,
 Nucl. Part. Phys. Proc. **261-262**, edited by J. Blümlein, K. Jansen, M. Krämer, and J. H. Kühn, 368–377
 (2015).
- ⁵⁴S. Jäger and J. Martin Camalich,
 “Reassessing the discovery potential of the $B \rightarrow K^* \ell^+ \ell^-$ decays in the large-recoil region: SM challenges and BSM opportunities”,
 Phys. Rev. D **93**, 014028
 (2016).
- ⁵⁵M. Beylich, G. Buchalla, and T. Feldmann,
 “Theory of $B \rightarrow K^{(*)} \ell^+ \ell^-$ decays at high q^2 : OPE and quark-hadron duality”,
 Eur. Phys. J. C **71**, 1635
 (2011).
- ⁵⁶D. Melikhov,
 “Charming loops in exclusive rare FCNC B -decays”,
 EPJ Web Conf. **222**, edited by D. Melikhov and I. Volobuev, 01007
 (2019).
- ⁵⁷A. Kozachuk and D. Melikhov,
 “Revisiting nonfactorizable charm-loop effects in exclusive FCNC B -decays”,
 Phys. Lett. B **786**, 378–381
 (2018).
- ⁵⁸A. Bharucha, D. M. Straub, and R. Zwicky,
 “ $B \rightarrow V \ell^+ \ell^-$ in the Standard Model from light-cone sum rules”,
 JHEP **08**, 098
 (2016).
- ⁵⁹R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate,
 “Lattice QCD calculation of form factors describing the rare decays $B \rightarrow K^* \ell^+ \ell^-$ and $B_s \rightarrow \phi \ell^+ \ell^-$ ”,
 Phys. Rev. D **89**, 094501
 (2014).
- ⁶⁰D. M. Straub,
 “flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond”,
 (2018).

- ⁶¹R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate,
“Rare B decays using lattice QCD form factors”,
PoS **LATTICE2014**, 372
(2015).
- ⁶²J. Albrecht, D. van Dyk, and C. Langenbruch,
“Flavour anomalies in heavy quark decays”,
Prog. Part. Nucl. Phys. **120**, 103885
(2021).
- ⁶³S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto,
“On the impact of power corrections in the prediction of $B \rightarrow K^* \mu^+ \mu^-$ observables”,
JHEP **12**, 125
(2014).
- ⁶⁴M. Bordone, G. Isidori, and A. Pattori,
“On the Standard Model predictions for R_K and R_{K^*} ”,
Eur. Phys. J. C **76**, 440
(2016).
- ⁶⁵A. Abdesselam et al. (Belle),
“Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method”,
(2019).
- ⁶⁶Y. Sato et al. (Belle),
“Measurement of the branching ratio of $\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau$ relative to $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decays with a semileptonic tagging method”,
Phys. Rev. D **94**, 072007
(2016).
- ⁶⁷M. Huschle et al. (Belle),
“Measurement of the branching ratio of $\bar{B} \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ relative to $\bar{B} \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle”,
Phys. Rev. D **92**, 072014
(2015).
- ⁶⁸S. Hirose et al. (Belle),
“Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \rightarrow D^* \tau^- \bar{\nu}_\tau$ ”,
Phys. Rev. Lett. **118**, 211801
(2017).
- ⁶⁹R. Aaij et al. (LHCb collaboration),
“Measurement of the ratio of the $\mathcal{B}(B^0 \rightarrow D^{*-} \tau^+ \nu_\tau)$ and $\mathcal{B}(B^0 \rightarrow D^{*-} \mu^+ \nu_\mu)$ branching fractions using three-prong τ -lepton decays”,
Phys. Rev. Lett. **120**, 171802

- (2018).
- ⁷⁰R. Aaij et al. (LHCb collaboration),
 “Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$ ”,
 Phys. Rev. Lett. **115**, 111803
 (2015).
- ⁷¹J. P. Lees et al. (BaBar),
 “Measurement of an Excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons”,
 Phys. Rev. D **88**, 072012
 (2013).
- ⁷²J. P. Lees et al. (BaBar),
 “Evidence for an excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ decays”,
 Phys. Rev. Lett. **109**, 101802
 (2012).
- ⁷³B. Capdevila, A. Crivellin, S. Descotes-Genon, L. Hofer, and J. Matias,
 “Searching for New Physics with $b \rightarrow s\tau^+\tau^-$ processes”,
 Phys. Rev. Lett. **120**, 181802
 (2018).
- ⁷⁴C. Cornella, J. Fuentes-Martin, and G. Isidori,
 “Revisiting the vector leptoquark explanation of the B-physics anomalies”,
 JHEP **07**, 168
 (2019).
- ⁷⁵M. Algueró, J. Matias, B. Capdevila, and A. Crivellin,
 “Disentangling lepton flavor universal and lepton flavor universality violating effects in $b \rightarrow s\ell + \ell^-$ transitions”,
 Phys. Rev. D **105**, 113007
 (2022).
- ⁷⁶B. Grinstein, R. P. Springer, and M. B. Wise,
 “Effective Hamiltonian for Weak Radiative B Meson Decay”,
 Phys. Lett. B **202**, 138–144
 (1988).
- ⁷⁷G. Buchalla, A. J. Buras, and M. E. Lautenbacher,
 “Weak decays beyond leading logarithms”,
 Rev. Mod. Phys. **68**, 1125–1144
 (1996).
- ⁷⁸M. Algueró, B. Capdevila, S. Descotes-Genon, P. Masjuan, and J. Matias,

- “Are we overlooking lepton flavour universal new physics in $b \rightarrow s\ell\ell$?”,
Phys. Rev. D **99**, 075017
(2019).
- ⁷⁹M. Algueró, B. Capdevila, S. Descotes-Genon, J. Matias, and M. Novoa-Brunet,
“ $b \rightarrow s\ell^+\ell^-$ global fits after R_{K_S} and $R_{K^{*+}}$ ”,
Eur. Phys. J. C **82**, 326
(2022).
- ⁸⁰W. Altmannshofer and P. Stangl,
“New physics in rare B decays after Moriond 2021”,
Eur. Phys. J. C **81**, 952
(2021).
- ⁸¹L.-S. Geng, B. Grinstein, S. Jäger, S.-Y. Li, J. Martin Camalich, and R.-X. Shi,
“Implications of new evidence for lepton-universality violation in $b \rightarrow s\ell + \ell^-$ decays”,
Phys. Rev. D **104**, 035029
(2021).
- ⁸²T. Hurth, F. Mahmoudi, D. M. Santos, and S. Neshatpour,
“More Indications for Lepton Nonuniversality in $b \rightarrow s\ell^+\ell^-$ ”,
Phys. Lett. B **824**, 136838
(2022).
- ⁸³A. K. Alok, A. Dighe, S. Gangal, and D. Kumar,
“Continuing search for new physics in $b \rightarrow s\mu\mu$ decays: two operators at a time”,
JHEP **06**, 089
(2019).
- ⁸⁴M. Ciuchini, M. Fedele, E. Franco, A. Paul, L. Silvestrini, and M. Valli,
“Lessons from the $B^{0,+} \rightarrow K^{*0,+}\mu^+\mu^-$ angular analyses”,
Phys. Rev. D **103**, 015030
(2021).
- ⁸⁵A. Datta, J. Kumar, and D. London,
“The B anomalies and new physics in $b \rightarrow se^+e^-$ ”,
Phys. Lett. B **797**, 134858
(2019).
- ⁸⁶L. Alasfar, A. Azatov, J. de Blas, A. Paul, and M. Valli,
“ B anomalies under the lens of electroweak precision”,
JHEP **12**, 016
(2020).
- ⁸⁷A. Tumasyan et al. (CMS),

- “Measurement of the $B_S^0 \rightarrow \mu^+ \mu^-$ decay properties and search for the $B^0 \rightarrow \mu^+ \mu^-$ decay in proton-proton collisions at $\sqrt{s} = 13$ TeV”,
Phys. Lett. B **842**, 137955
(2023).
- ⁸⁸M. Algueró, A. Biswas, B. Capdevila, S. Descotes-Genon, J. Matias, and M. Novoa-Brunet,
“To (b)e or not to (b)e: no electrons at LHCb”,
Eur. Phys. J. C **83**, 648
(2023).
- ⁸⁹N. Gubernari, A. Kokulu, and D. van Dyk,
“ $B \rightarrow P$ and $B \rightarrow V$ Form Factors from B -Meson Light-Cone Sum Rules beyond Leading Twist”,
JHEP **01**, 150
(2019).
- ⁹⁰N. Gubernari, D. van Dyk, and J. Virto,
“Non-local matrix elements in $B_{(s)} \rightarrow \{K^{(*)}, \phi\} \ell^+ \ell^-$ ”,
JHEP **02**, 088
(2021).
- ⁹¹W. G. Parrott, C. Bouchard, and C. T. H. Davies ((HPQCD collaboration)§, HPQCD),
“ $B \rightarrow K$ and $D \rightarrow K$ form factors from fully relativistic lattice QCD”,
Phys. Rev. D **107**, 014510
(2023).
- ⁹²W. G. Parrott, C. Bouchard, and C. T. H. Davies (HPQCD),
“Standard Model predictions for $B \rightarrow K \ell^+ \ell^-$, $B \rightarrow K \ell^1 \ell^2$ and $B \rightarrow K \nu \nu^-$ using form factors from $N_f=2+1+1$ lattice QCD”,
Phys. Rev. D **107**, [Erratum: Phys.Rev.D 107, 119903 (2023)], 014511
(2023).
- ⁹³G. Aad et al. (ATLAS),
“The ATLAS Experiment at the CERN Large Hadron Collider”,
JINST **3**, S08003
(2008).
- ⁹⁴S. Chatrchyan et al. (CMS),
“The CMS Experiment at the CERN LHC”,
JINST **3**, S08004
(2008).
- ⁹⁵K. Aamodt et al. (ALICE),
“The ALICE experiment at the CERN LHC”,
JINST **3**, S08002

- (2008).
- ⁹⁶E. Lopienska,
“The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022”,
General Photo
(2022).
- ⁹⁷R. Aaij et al. (LHCb collaboration),
“Measurement of $\sigma(pp \rightarrow b\bar{b}X)$ at $\sqrt{s}=7$ TeV in the forward region”,
Phys. Lett. **B694**, 209
(2010).
- ⁹⁸R. Aaij et al. (LHCb collaboration),
“Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions”,
Phys. Rev. Lett. **118**, 052002
(2017).
- ⁹⁹A. A. Alves Jr. et al. (LHCb collaboration),
“The LHCb detector at the LHC”,
JINST **3**, S08005
(2008).
- ¹⁰⁰C. Elsasser (LHCb collaboration),
 $b\bar{b}$ Production angle plots,
https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html.
- ¹⁰¹R. Aaij et al.,
“Performance of the LHCb Vertex Locator”,
JINST **9**, P09007
(2014).
- ¹⁰²S. Borghi et al.,
“First spatial alignment of the LHCb VELO and analysis of beam absorber collision data”,
Nucl. Instrum. Meth. **A618**, 108–120
(2010).
- ¹⁰³R. Aaij et al.,
“Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC”,
JINST **14**, P04013
(2019).
- ¹⁰⁴R. Aaij et al. (LHCb collaboration),
“LHCb detector performance”,
Int. J. Mod. Phys. **A30**, 1530022

- (2015).
- ¹⁰⁵R. Aaij et al. (LHCb collaboration),
“Measurement of the track reconstruction efficiency at LHCb”,
JINST **10**, P02007
(2015).
- ¹⁰⁶R. E. Kalman,
“A New Approach to Linear Filtering and Prediction Problems”,
Journal of Basic Engineering **82**, 35–45
(1960).
- ¹⁰⁷R. Fruhwirth,
“Application of Kalman filtering to track and vertex fitting”,
Nucl. Instrum. Meth. A **262**, 444–450
(1987).
- ¹⁰⁸M. Needham,
Identification of Ghost Tracks using a Likelihood Method,
Tech. rep.
(CERN, Geneva, 2008).
- ¹⁰⁹A. Perieanu,
Identification of Ghost Tracks using Neural Networks,
Tech. rep.
(CERN, Geneva, 2008).
- ¹¹⁰T. Humair,
“Testing lepton universality in penguin decays of beauty mesons using the LHCb detector”,
Presented 02 May 2019
(2019).
- ¹¹¹R. Forty and O. Ullaland,
“Particle Identification: Time-of-Flight, Cherenkov and Transition Radiation Detectors”,
In *Particle Physics Reference Library: Volume 2: Detectors for Particles and Radiation*,
Edited by C. W. Fabjan and H. Schopper
(Springer, Cham, 2020),
Pp. 281–335.
- ¹¹²M. Adinolfi et al.,
“Performance of the LHCb RICH detector at the LHC”,
Eur. Phys. J. **C73**, 2431
(2013).
- ¹¹³R. Calabrese et al.,

- “Performance of the LHCb RICH detectors during LHC Run 2”,
JINST **17**, P07013
(2022).
- ¹¹⁴R. 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**, 1
(2018).
- ¹¹⁵“Global PID performance for charged particles”,
(2020).
- ¹¹⁶A. Kashchuk, C. Lippmann, W. Riegler, B. Schmidt, T. Schneider, and V. Suvorov,
“Asymmetric and double-cathode-pad wire chambers for the LHCb muon system”,
Nucl. Instrum. Meth. A **555**, 48–54
(2005).
- ¹¹⁷A. A. Alves Jr. et al.,
“Performance of the LHCb muon system”,
JINST **8**, P02022
(2013).
- ¹¹⁸F. Archilli et al.,
“Performance of the muon identification at LHCb”,
JINST **8**, P10020
(2013).
- ¹¹⁹J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk (LHCb HLT project),
“Performance of the LHCb High Level Trigger in 2012”,
J. Phys. Conf. Ser. **513**, edited by D. L. Groep and D. Bonacorsi, 012001
(2014).
- ¹²⁰O. Callot and S. Hansmann-Menzemer,
The Forward Tracking: Algorithm and Performance Studies,
Tech. rep.
(CERN, Geneva, 2007).
- ¹²¹“RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector”,
(2020).
- ¹²²K. Petridis, A. Marshall, P. A. Cartelle, M. Birch, M. Patel, M. McCann, and M. Smith,
ICB $K^0 \mu^+ \mu^-$ Run 1 + Run 2 binned angular analysis,
(2023)
<https://twiki.cern.ch/twiki/bin/view/Main/ImperialKstMuMu>.

- ¹²³T. Sjostrand, S. Mrenna, and P. Z. Skands,
“A Brief Introduction to PYTHIA 8.1”,
Comput. Phys. Commun. **178**, 852–867
(2008).
- ¹²⁴T. Sjostrand, S. Mrenna, and P. Z. Skands,
“PYTHIA 6.4 Physics and Manual”,
JHEP **05**, 026
(2006).
- ¹²⁵I. Belyaev et al. (LHCb),
“Handling of the generation of primary events in Gauss, the LHCb simulation framework”,
J. Phys. Conf. Ser. **331**, edited by S. C. Lin, 032047
(2011).
- ¹²⁶D. J. Lange,
“The EvtGen particle decay simulation package”,
Nucl. Instrum. Meth. A **462**, edited by S. Erhan, P. Schlein, and Y. Rozen, 152–155
(2001).
- ¹²⁷P. Golonka and Z. Was,
“PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays”,
Eur. Phys. J. C **45**, 97–107
(2006).
- ¹²⁸J. Allison et al.,
“Geant4 developments and applications”,
IEEE Trans. Nucl. Sci. **53**, 270
(2006).
- ¹²⁹C. Nägeli,
“Analysis of the rare decay $B_s^0 \rightarrow \mu^+ \mu^-$ using the compact muon solenoid experiment at
CERN’s Large Hadron Collider”,
PhD thesis (Zurich, ETH, 2013).
- ¹³⁰F. Pedregosa et al.,
“Scikit-learn: Machine learning in Python”,
J. Machine Learning Res. **12**, and online at <http://scikit-learn.org/stable/>, 2825
(2011).
- ¹³¹T. Chen and C. Guestrin,
“XGBoost: a scalable tree boosting system”,
In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and
data mining,
KDD ’16

- (2016),
Pp. 785–794.
- ¹³²J. Cramer,
The Origins of Logistic Regression,
Tinbergen Institute Discussion Papers 02-119/4
(Tinbergen Institute, Dec. 2002).
- ¹³³R. Aaij et al. (LHCb collaboration),
“Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics”,
JHEP **04**, f_s/f_d value updated in LHCb-CONF-2013-011, 001
(2013).
- ¹³⁴Y. Amhis et al. (Heavy Flavor Averaging Group),
“Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016”,
Eur. Phys. J. **C77**, updated results and plots available at <https://hflav.web.cern.ch>,
895
(2017).
- ¹³⁵M. Hecker,
“Searching for new physics in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays”,
PhD thesis (Imperial Coll., London, 2020).
- ¹³⁶S. Coquereau, M. Chrzaszcz, S. Cunliffe, C. Parkinson, K. Petridis, M. Tresch, M. De Cian, and C. M. Langenbruch,
*The $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ selection for 3fb^{-1} of LHCb data*,
Tech. rep. LHCb-INT-2013-058. CERN-LHCb-INT-2013-058
(CERN, Geneva, Nov. 2013).
- ¹³⁷L. Anderlini et al.,
“The PIDCalib package”,
(2016).
- ¹³⁸R. Aaij et al. (LHCb),
“Measurement of the track reconstruction efficiency at LHCb”,
JINST **10**, P02007
(2015).
- ¹³⁹M. De Cian, S. Hansmann-Menzemer, A. Jaeger, P. Seyfert, J. van Tilburg, and S. Wandernoth,
Measurement of the track finding efficiency,
Tech. rep.
(CERN, Geneva, 2012).

- ¹⁴⁰M. De Cian,
“Track Reconstruction Efficiency and Analysis of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ at the LHCb Experiment”,
PhD thesis (Zurich U., 2013).
- ¹⁴¹M. Pivk and F. R. Le Diberder,
“sPlot: A statistical tool to unfold data distributions”,
Nucl. Instrum. Meth. **A555**, 356–369
(2005).
- ¹⁴²S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino,
Data driven trigger efficiency determination at LHCb,
Tech. rep.
(CERN, Geneva, 2014).
- ¹⁴³A. Rogozhnikov,
“Reweightings with Boosted Decision Trees”,
J. Phys. Conf. Ser. **762**, edited by L. Salinas and C. Torres, 012036
(2016).
- ¹⁴⁴M. Abadi et al.,
“TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”,
(2016).
- ¹⁴⁵F. James,
“MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1”,
(1994).
- ¹⁴⁶A. Poluektov,
“Performing amplitude fits with TensorFlow: LHCb experience. HEP analysis ecosystem workshop”,
(2017).
- ¹⁴⁷G. Isidori and D. M. Straub,
“Minimal Flavour Violation and Beyond”,
Eur. Phys. J. C **72**, 2103
(2012).
- ¹⁴⁸R. L. Workman et al. (Particle Data Group),
“Review of Particle Physics”,
PTEP **2022**, 083C01
(2022).
- ¹⁴⁹M. Andersson, U. Egede, Y. Fuji, T. Hadavizadeh, M. Hecker, R. Henderson, K. Petridis,
G. Pomery, E. Smith, R. Song, and Z. Ghorbanimoghaddam,
*Unbinned analysis of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays using a dispersion relation mode*,

- (2024)
`https : / / twiki . cern . ch / twiki / bin / view / LHCbPhysics / BdToKstarMuMuUnbinnedDispersion.`
- ¹⁵⁰R. Aaij et al. (LHCb),
“Amplitude Analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ Decay”,
Phys. Rev. Lett. **132**, 131801
(2024).
- ¹⁵¹R. Aaij et al. (LHCb),
“Determination of short- and long-distance contributions in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decays”,
Phys. Rev. D **109**, 052009
(2024).
- ¹⁵²D. P. Aguillard et al. (Muon g-2),
“Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm”,
(2023).
- ¹⁵³J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox,
“Hyperopt: a python library for model selection and hyperparameter optimization”,
Computational Science Discovery **8**, 014008
(2015).