



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Studies of b Hadron Decays to Charmonium, the LHCb Upgrade and Operation

Konstantin Gizdov



Doctor of Philosophy
The University of Edinburgh
2022



Except where otherwise noted, this work is licensed under a
Creative Commons Attribution 4.0 International License
© Konstantin Gizdov, University of Edinburgh March 2022 CC BY

Abstract

Precise measurements of CP violation provide stringent tests of the Standard Model towards the search for signs of new physics. Using LHC proton-proton collision data, collected by the LHCb detector during 2015 and 2016 at the centre-of-mass energy of 13 TeV corresponding to an integrated luminosity of 1.9fb^{-1} , presented is the latest measurement of the CP -violating phase, ϕ_s , using $B_s^0 \rightarrow J/\psi \phi$ decays. The machine-learning-based data selection, data-driven corrections to simulated event samples, and the control of systematic effects using dedicated samples are discussed. The values $\phi_s = -0.083 \pm 0.041 \pm 0.006 \text{ rad}$, $\Delta\Gamma_s = 0.077 \pm 0.008 \pm 0.003 \text{ ps}^{-1}$ (*i.e.* the decay width difference between the light and the heavy mass eigenstates in the B_s^0 system) and $\Gamma_s - \Gamma_d = -0.0041 \pm 0.0024 \pm 0.0015 \text{ ps}^{-1}$ (*i.e.* the difference of the average B_s^0 and B_d^0 meson decay widths) are obtained, yielding the World's most precise determination of these quantities ¹.

Furthermore, shown are the efforts and contributions towards the LHCb Upgrade: the quality assurance and testing of the LHCb RICH Upgrade components, and the redesign and upgrade of the fully online software trigger – LHCb HLT Upgrade. Regarding the former, an original implementation of a parallelised, robust and highly available automation system is introduced. In connection to the latter, a novel neural network architecture and optimisation methods are laid out, enabling complex machine learning to be performed in a low latency high-throughput environment. Those directly influence the future deployment of the experiment and its data collecting and analysis capabilities. Thus, they are essential for future more precise and stringent research.

¹the first uncertainty is statistical, and the second is systematic

Lay Summary

According to our present understanding, the universe started as an extremely dense uniform object then expanded outwards into what is observed today. Moreover, it should contain equal parts matter and antimatter. Yet, appears to be dominated by matter only. One possible resolution to this discrepancy lies in a phenomenon called CP violation, which was discovered in the 1960s in particular decays of particles called kaons. It describes the subtle difference in the behaviour of matter and antimatter. Unfortunately, the amount of CP violation in the Standard Model, the state-of-the-art theory of particle physics, is too small to explain the vast matter-antimatter asymmetry observed in the universe. This unexplained discrepancy motivates a lot of research into additional sources of CP violation that do not fit into current theories. One of the ways to do this is to study the decays of composite particles, called b mesons, which constantly change between their matter and antimatter counterparts.

The LHCb Experiment, sitting at the Large Hadron Collider at CERN, is at the forefront of this kind of research and was built specifically to detect such particles and their decays. The bulk of the work presented in this thesis uses data collected by LHCb during the 2015 and 2016 data-taking periods to analyse the decays of one type of b meson, the B_s^0 . Specifically, it focuses on a particular type of decay involving a direct probe of one type of CP violation. Furthermore, it delves into the technologies used to collect data, process and analyse them efficiently and describes the development of two separate improvements planned for the LHCb Upgrade. One has to do with upgrading the Ring-imaging Čerenkov subdetector system and another with the algorithms used to filter and select collision events for further analysis.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own, except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification, except as specified.

Parts of this work have been published in [1–5].

(Konstantin Gizdov, March 2022)

I want to live, however briefly, knowing my life is finite. Mortality gives meaning to human life, captain. Piece, love, friendship. These are precious because we know they cannot endure. A butterfly that lives forever is really not a butterfly at all.

Data ([6])

Acknowledgements

I would like to thank my supervisor Greig Cowan for his excellent guidance, feedback and support throughout my PhD. The Edinburgh PPE Group has been vital to the success of my research. In particular, I am deeply grateful to Silvia Gambetta, Stephan Eisenhardt and last but not least Franz Muheim for sharing their expertise and showing genuine understanding when needed.

Thanks must also go to my dear colleagues. They provided much of the needed skills, time, effort, and thoughtful discussions, especially including the $B_s^0 \rightarrow J/\psi \phi$ proponents, the LHCb RICH Collaboration and the LHCb High Level Trigger group. Most notably, I give my thanks to Sean Benson, Veronika Chobanova, Francesca Dordei, Katya Govorkova, Jennifer Zonneveld, Stefano Gallorini, Antonis Papanestis, Conor Fitzpatrick, Christoph Hasse, Rosen Matev, Niklas Nolte, Alex Pearce, Sascha Stahl and many more.

I would like to thank CERN, the LHCb Collaboration, the University of Edinburgh, SUPA and the STFC for providing the infrastructure, training, resources and funding to make my doctoral research possible in the first place. I would also like to express my gratitude to the developers of ROOT, RooFit, RooStats, EvtGen, Pythia, Geant4, RapidFit, RapidSim, The Scikit-HEP Project, TensorFlow and the suite of LHCb packages for providing the necessary software used to analyse and simulate the data used in my research.

Finally, I thank my family, long-standing friends and the friends I made along the way in and outside of physics. Thanks to Denitsa Angelova, Tim Bristow, Emmy Gabriel, Paul Glaysheer, Ruta Liepina, Adam Morris, Nikola Petkov, Yavor Petruncyashchev, Monika Slavova, Iwan Smith, Georgi Tinchev for being supportive, encouraging me and tolerating my attempts to explain what I do.

Contents

Abstract	i
Lay Summary	iii
Declaration	v
Acknowledgements	ix
Contents	xi
List of Figures	xvii
List of Tables	xxiii
Listings	xxv
1 Introduction	1
2 Theory	7
2.1 Standard Model.....	7
2.1.1 Electroweak (EW) interaction and symmetry breaking (Higgs mechanism).....	8
2.1.2 Strong interaction (QCD)	10
2.1.3 Quarks and leptons	11
2.1.4 Beyond the Standard Model (BSM) Physics.....	12
2.2 <i>CP</i> Violation.....	13
2.2.1 <i>CP</i> violation in decay	14
2.2.2 <i>CP</i> violation in mixing.....	16
2.2.3 <i>CP</i> violation in mixing and decay	17
2.3 CKM Matrix.....	18
2.4 <i>CP</i> violation in the B_s^0 sector	21
2.4.1 B_s^0 - \bar{B}_s^0 mixing	23
2.4.2 $b \rightarrow c\bar{c}s$ transitions	25

2.4.3	B_s^0 - \bar{B}_s^0 mixing and $B_s^0 \rightarrow J/\psi \phi$ decay	27
2.4.4	Angular distributions of pseudoscalar-to-vector-vector (P2VV) meson decays.....	28
2.5	ϕ_s measurement and its implications.....	32
2.5.1	New physics in B_s^0 mixing	32
3	The LHCb Experiment	35
3.1	The Large Hadron Collider	35
3.2	LHCb Overview	39
3.2.1	Coordinate system	40
3.3	Tracking and Vertexing	40
3.3.1	Vertex Locator (VELO)	40
3.3.2	Vertexing Performance.....	43
3.3.3	Magnet	44
3.3.4	Silicon and Straw tube tracking stations (TT, IT and OT). ..	46
3.3.5	Tracking reconstruction.....	50
3.3.6	Tracking performance	52
3.4	Particle Identification.....	54
3.4.1	RICH system.....	55
3.4.2	Čherenkov effect.....	55
3.4.3	Ring-imaging Čherenkov (RICH) detectors	57
3.4.4	RICH system performance	60
3.4.5	Calorimeters.....	62
3.4.6	Muon chambers.....	65
3.5	Trigger, Reconstruction and Data Processing.....	69
3.5.1	Trigger (L0 and HLT).....	69
3.5.2	Reconstruction	74
3.5.3	Simulation	74
3.5.4	Analysis	75
3.5.5	Distributed computing (Dirac and the Grid)	76
4	The LHCb Upgrade of the High Level Trigger (HLT)	77
4.1	Introduction	77
4.2	LHCb HLT Upgrade	78
4.2.1	LHCb HLT Upgrade requirements	78
4.2.2	Core Software	79
4.2.3	HLT	80
4.2.4	Machine learning algorithms.....	81

4.2.5	Machine learning algorithms at LHCb	81
4.3	NNDrone	82
4.3.1	Drone neural network training	83
4.3.2	Drone mutation and morphing during learning	85
4.3.3	High energy physics applications	86
4.3.4	Transferability to low latency environments	95
4.3.5	NNDrone at LHCb.....	96
4.3.6	Results.....	97
5	The LHCb Upgrade of the Ring-imaging Čherenkov (RICH) system	99
5.1	Introduction	100
5.2	Photon Detector Quality Assurance.....	100
5.2.1	Custom test bench.....	101
5.2.2	Goals.....	103
5.2.3	System automation	106
5.2.4	Microservices	106
5.2.5	Finite state machine.....	108
5.2.6	Results.....	110
6	Simulation and event selection of $B_s^0 \rightarrow J/\psi \phi$ decays with LHC Run 2 data	113
6.1	Introduction	113
6.2	Data and simulation samples	113
6.2.1	Dataset	114
6.2.2	Simulation samples	114
6.2.3	Momenta corrections	116
6.3	Event selection	118
6.3.1	Trigger selection.....	119
6.3.2	STRIPPING selection.....	124
6.4	Corrections to simulated samples.....	126
6.4.1	Corrections of particle PID	128
6.4.2	Reweighting of simulation kinematic, event and selection variables.....	136
6.5	Multivariate selection.....	143
6.5.1	MVA training and optimisation.....	144

7	Measurement of the CP-violating phase ϕ_s using $B_s^0 \rightarrow J/\psi \phi$ decays with LHC Run 2 data	153
7.1	Introduction	153
7.2	Invariant mass fit	154
7.2.1	Primary vertex constraint in J/ψ mass estimation	154
7.2.2	Peaking backgrounds and their subtraction	156
7.2.3	Per-event-error mass model	158
7.3	Decay time resolution	160
7.3.1	Model and parametrisation	160
7.4	Decay time efficiency	162
7.5	Angular efficiency	165
7.5.1	Angular efficiency model and parametrisation	165
7.6	Flavour Tagging	167
7.7	Time-dependent angular fit to data	170
7.7.1	Statistical background subtraction	170
7.7.2	The $B_s^0 \rightarrow J/\psi \phi$ probability density function	170
7.7.3	K^+K^- S-wave interference	171
7.7.4	Inclusion of resolution, detector acceptance and flavour tagging	171
7.8	Systematics	174
7.8.1	Contribution from B_c^+ decays	174
7.9	Results	182
8	Conclusion	185
A	Further detail on some parts of Chapter 2	189
A.1	The Standard Model	189
A.1.1	The unconventional distinction used to portray chirality	189
A.2	CP Violation	190
A.2.1	Time-reversal operator	190
A.2.2	Weak decay Hamiltonian construction	191
A.2.3	Two-mode decay amplitude	191
A.2.4	CP in mixing	192
A.3	On-shell and off-shell particles	194
A.4	Expected ϕ_s statistical uncertainty against amount of collected data at LHCb	195
A.5	P2VV angular differential decay rate in the helicity formalism	196

B	Further detail on some parts of Chapter 3	199
B.1	Čherenkov effect	199
B.1.1	Čherenkov angle	199
B.1.2	Čherenkov ring radius	200
B.1.3	RICH Optics	200
B.2	Alignment and calibration	201
B.2.1	Data samples and tasks	201
B.3	Trigger	201
B.3.1	TIS and TOS	201
C	Further detail on some parts of Chapter 6 and Chapter 7	203
C.1	Maximum likelihood estimation (fit)	203
C.1.1	Fisher information	205
C.1.2	Minimising the negative logarithm of the likelihood	206
C.2	Fit to weighted data	208
C.3	Nuisance parameters	210
C.4	Detailed information about the so-called TISTOS bug	211
C.5	Impact of VELO track fit mismatch between data and simulation .	213
C.6	Flavour Tagging	214
C.7	Details for combined fit between Run 1, 2015 and 2016 results	215
C.8	Wrong PV component in decay time resolution studies	215
C.9	S-wave fractions parameters	216
	Bibliography	217

List of Figures

(1.1)	ϕ_s^{ccs} and $\Delta\Gamma_s$ 68 % confidence-level contours and theoretical predictions.	3
(2.1)	Elementary Particles in the Standard Model	8
(2.2)	CKM Unitary Triangles.....	21
	(2.2 (a)) d–b unitary triangle.....	21
	(2.2 (b)) s–b unitary triangle.....	21
(2.3)	Global fit to CKM unitary triangle in $(\bar{\rho}, \bar{\eta})$ plane	23
(2.4)	Global fit to CKM unitary triangle in $(\bar{\rho}_s, \bar{\eta}_s)$ plane	24
(2.5)	B_s^0 - \bar{B}_s^0 mixing diagrams	25
	(2.5 (a)) s -channel.....	25
	(2.5 (b)) t -channel.....	25
(2.6)	$B_s^0 \rightarrow J/\psi \phi$ decay diagrams.....	26
	(2.6 (a)) tree diagram.....	26
	(2.6 (b)) penguin diagram.....	26
(2.7)	Helicity angular basis.....	29
(2.8)	$B_s^0 \rightarrow J/\psi \phi$ final state polarisations	30
(2.9)	Constraints on Δ_s as calculated by CKMFitter group.....	33
(3.1)	Schematic diagram of CERN’s Accelerator complex	36
(3.2)	LHCb luminosity during a typical fill	37
(3.3)	The LHCb detector	38
(3.4)	Distributions of simulated b quark production in polar angles and pseudorapidity	39
(3.5)	LHCb VELO Layout and Acceptance	41
(3.6)	LHCb VELO RF Foil and r/ϕ Sensors.....	42

(3.7)	LHCb VELO primary vertex and impact parameter resolutions	43
(3.8)	LHCb Magnet	45
(3.9)	LHCb Magnetic Field	46
(3.10)	LHCb Tracker Turicensis	47
(3.11)	LHCb Inner Tracker	48
(3.12)	LHCb Outer Tracker	49
(3.13)	LHCb Outer Tracker drift time	50
(3.14)	LHCb track types	51
(3.15)	LHCb track efficiency	53
(3.16)	LHCb track resolution	54
(3.17)	Čerenkov Emission	55
	(3.17 (a)) $v < c/n$	55
	(3.17 (b)) $v > c/n$	55
(3.18)	Čerenkov Angle	56
	(3.18 (a)) $v < c/n$	56
	(3.18 (b)) $v > c/n$	56
(3.19)	LHCb RICH1 and RICH2	57
(3.20)	LHCb RICH HPD schematic and photograph	59
(3.21)	LHCb RICH PID efficiency	61
(3.22)	Particle shower signature of electrons, hadrons and photons	62
(3.23)	SPD, PS, ECAL and HCAL segmentation	63
	(3.23 (a)) SPD, PS and ECAL	63
	(3.23 (b)) HCAL	63
(3.24)	SPD/PS module schematic	63
(3.25)	Inner ECAL module	64
(3.26)	HCAL submodule schematic	64
(3.27)	Schematic diagram of the muon system and its stations	66
	(3.27 (a)) Muon system side view	66
	(3.27 (b)) Muon system segmentation	66
(3.28)	LHCb data flow	68
(3.29)	LHCb Trigger rates and algorithms	70
	(3.29 (a)) Trigger Algorithms	70

(3.29 (b))	Trigger Rates.....	70
(3.30)	SPD multiplicity for events containing $B^+ \rightarrow \bar{D}^0 \pi^+$	72
(4.1)	Input distributions for simulated signal and background events for $B_s^0 \rightarrow J/\psi \phi$ Keras classifier	88
(4.1 (a))	Mother p_T	88
(4.1 (b))	Mother η	88
(4.1 (c))	Min p_T	88
(4.1 (d))	Max p_T	88
(4.1 (e))	Min η	88
(4.1 (f))	Min η	88
(4.2)	Input distributions for simulated signal and background events for Keras jet-separation classifier.....	91
(4.2 (a))	Jet η	91
(4.2 (b))	$\sigma(\eta_{\text{charged}})$	91
(4.2 (c))	$\sigma(\eta_{\text{neutral}})$	91
(4.2 (d))	Jet ϕ	91
(4.2 (e))	$\sigma(\phi_{\text{charged}})$	91
(4.2 (f))	$\sigma(\phi_{\text{neutral}})$	91
(4.2 (g))	Avg. η_{charged}	91
(4.2 (h))	Avg. η_{neutral}	91
(4.3)	Input distributions for simulated signal and background events for Keras jet-separation classifier.....	92
(4.3 (a))	Avg. ϕ_{charged}	92
(4.3 (b))	Avg. ϕ_{neutral}	92
(4.3 (c))	Jet E_{charged}	92
(4.3 (d))	Jet E_{neutral}	92
(4.3 (e))	$\sum p_{T\text{charged}}$	92
(4.3 (f))	$\sum p_{T\text{neutral}}$	92
(4.4)	Input distributions for simulated signal and background events for Keras jet-separation classifier.....	93
(4.4 (a))	K_T^2	93
(4.4 (b))	Ket m	93
(4.4 (c))	Ket p	93
(4.5)	Drone loss function as a function of training epoch	93
(4.6)	Drone loss convergence as a function of training epoch	94
(4.7)	Signal efficiency vs background rejection comparison between the drone and original model	94
(5.1)	PDQA test box	101
(5.2)	Single-channel response of a R13742 PMT	105

(5.3)	Microservices	107
(5.4)	Finite state machine	109
(5.5)	R13742 (1 in) PMT gain	110
(5.6)	R13743 (2 in) PMT gain	111
(6.1)	Momentum and impact parameter resolutions for data and simulation	117
(6.2)	Muon and DiMion L0 trigger decisions	120
(6.3)	Global L0 trigger decisions	120
(6.4)	Comparison of the distributions of J/ψ trigger response between collision data and signal MC simulation sample	123
(6.5)	Boundary effect of trigger cut on decay time	127
(6.6)	Binning scheme used in PIDCALIB for ProbNN $\mu(\mu)$ resampling	132
	(6.6 (a)) p_T projection	132
	(6.6 (b)) p projection	132
	(6.6 (c)) N_{tracks} projection	132
(6.7)	Binning scheme used in PIDCALIB for ProbNN $k(K)$ resampling	133
	(6.7 (a)) p_T projection	133
	(6.7 (b)) p projection	133
	(6.7 (c)) N_{tracks} projection	133
(6.8)	Simulation PID response corrections	134
	(6.8 (a)) No PIDCALIB, No GB weighting	134
	(6.8 (b)) PIDCALIB without GB weighting	134
	(6.8 (c)) PIDCALIB with GB weighting	134
(6.9)	Distribution of $m(J/\psi K^+ K^-)$ after the STRIPPING and trigger selection in 2016 data	135
(6.10)	Simulation PID response corrections	136
	(6.10 (a)) No PIDCALIB, No GB weighting	136
	(6.10 (b)) PIDCALIB without GB weighting	136
	(6.10 (c)) PIDCALIB with GB weighting	136
(6.11)	Distributions of track multiplicity	137
(6.12)	Distributions before and after GB reweighting corrections (part. 1).	139
(6.13)	Distributions before and after GB reweighting corrections (part. 2).	140
(6.14)	Background rejection versus signal efficiency of the different MVAs trained to select $B_s^0 \rightarrow J/\psi \phi$ events	144

(6.15)	Distribution of the FOM used to optimise the threshold cut on the BDTG3	148
(6.16)	Distribution of $m(J/\psi K^+ K^-)$ in 2016 after applying all selections, including the MVA's optimised cut	148
(6.17)	Distributions of MVA training variables in the signal and background samples	149
(6.18)	BDTG3 consistency checks.....	150
(6.19)	Comparison of BDTG3 response distribution in simulation and data	150
(6.20)	Distributions of $m_{K^+ K^-}$ and $m_{\mu^+ \mu^-}$ in data after selection	151
(7.1)	Values of resolutions of $m(J/\psi K^+ K^-)$ mass shape in bins of B_s^0 decay time	155
(7.2)	Distribution of the invariant mass of B_s^0 candidates from simulation	156
(7.3)	Distribution of the invariant mass of B_s^0 candidates in data	159
(7.4)	Distribution of the decay time of $B_s^0 \rightarrow D_s^- \pi^+$ candidates.....	160
(7.5)	Standard deviation of the effective single-Gaussian decay time resolution, σ_{eff}	162
(7.6)	Distribution of $m(J/\psi K^+ \pi^-)$ in 2016 after applying all selections, including MVA's optimised cut	164
(7.7)	Overview of flavour tagging at LHCb	168
(7.8)	Gaussian fits to P distributions from Eq. 7.21 of parameters in $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ simulated experiments	177
(7.9)	Gaussian fits to P distributions from Eq. 7.21 of parameters in $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ simulated experiments	178
(7.10)	Gaussian fits to P distributions from Eq. 7.21 of parameters in $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ simulated experiments	179
(7.11)	Gaussian fits to P distributions from Eq. 7.21 of parameters in $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ simulated experiments	180
(7.12)	Gaussian fits to P distributions from Eq. 7.21 of parameters in $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ simulated experiments	181
(7.13)	Decay-time and angular fit 1-D projections	183
(8.1)	Regions of 68% confidence level in the ϕ_s - $\Delta\Gamma_s$ plane	187
(A.1)	Expected ϕ_s statistical uncertainty vs. integrated luminosity	195

(C.1)	Comparison of the distributions of K and μ VELO $\chi^2_{\text{track}}/\text{ndf}$	213
-------	--	-----

List of Tables

(3.1)	L0 thresholds for the different trigger lines.....	71
(4.1)	Requirements for full software trigger	79
(4.2)	Hyperparameter number comparison between original classifier and drone	95
(4.3)	Processing time comparison between original classifier and drone....	95
(4.4)	Statistics from Moore test of NNDrone	96
(6.1)	List of Monte-Carlo simulation samples.....	116
(6.2)	Decay model parameters for Sim09b Monte-Carlo simulation	117
(6.3)	The analysis-relevant triggers	122
(6.4)	Criteria used to select $B_s^0 \rightarrow J/\psi \phi$ candidates in the STRIPPING.....	125
(6.5)	PIDCALIB efficiency conditions per resampled track type	129
(6.6)	Decay modes used by PIDCALIB calibration samples.....	131
(6.7)	BDTG3 variable importance ranking.....	147
(7.1)	Angular and time-dependent functions used in the fit to the data ...	173
(7.2)	Summary of systematic uncertainties.....	176
(7.3)	ϕ_s fit correlation matrix.....	182
(C.1)	Calibration parameters for the OS and SSK taggers.....	214
(C.2)	Overall tagging performance for $B_s^0 \rightarrow J/\psi K^+ K^-$	214
(C.3)	Correlation matrix for ϕ_s combined fit of Run 1 and the 2015 and 2016 data.....	215
(C.4)	Correlation matrix for ϕ_s combined fit of Run 1 and the 2015 and 2016 data including systematics	215

(C.5)	Values of the S-wave parameters in each $m(K^+K^-)$ bin.....	216
-------	--	-----

Listings

(4.1)	$B_s^0 \rightarrow (J/\psi \rightarrow \mu^+\mu^-)(\phi \rightarrow K^+K^-)$ Keras classifier implementation..	87
(4.2)	Jet separation Keras classifier implementation	89
(C.1)	DAVINCI configuration needed to mitigate the TISTOS bug	212

Chapter 1

Introduction

One of the remaining puzzles of modern physics is the unexplained abundance of matter compared to antimatter in the universe. Current theory indicates that matter and antimatter should have been created in equal amounts at the Big Bang. However, modern observations of the interstellar medium put the fraction of antimatter to matter, f , at about $f < 10^{-15}$ [7]. Other observations of star clusters put the limit closer to $f < 10^{-9}$. In all of these cases, it is clear that very few antimatter particles roam the universe compared to the matter we observe in galaxies, stars, planets and life. To account for this imbalance, Sakharov came up with a set of conditions that need to be fulfilled by any yet undiscovered phenomena that aim to address the matter-antimatter imbalance [8]. These processes carry the catch-all name of Baryogenesis. Sakharov's conditions require the violation of baryon/lepton number ($n_{B/L}$) conservation and violation of C and CP symmetries while happening outside thermal equilibrium. Violation in C symmetry is necessary so that processes producing more baryons than antibaryons do not balance with processes producing more antibaryons than baryons. The violation of CP symmetry (CP violation) means that equivalent processes involving matter and antimatter must occur at a different rate so that interactions that do not conserve $n_{B/L}$ do not cancel each other out. Considering the constraints on baryon and lepton number violation, the Standard Model cannot generate the large enough matter-antimatter imbalance observed today [9–11]. Thus, looking for new sources of CP violation beyond the Standard Model is essential. One of the ways to perform such a search is by employing “B Physics”: an active area of particle physics research where scientists study hadrons containing b quarks.

The LHCb experiment is designed primarily for this type of research – to look for new physics in b decays. It continues the work of BaBar and Belle [12], which focused on studies of B^0 and B^+ decays. Still, being a dedicated “B Physics” experiment operating at a hadron collider, LHCb can study new hadrons, such as B_s^0 mesons, B_c^+ mesons and b baryons with unrivalled precision.

This thesis focuses on the study of B_s^0 meson decays to Charmonium (a flavourless meson whose constituents are a heavy c quark and its antiquark) and on two projects about improving the performance of the LHCb experiment. Specifically, the decay $B_s^0 \rightarrow J/\psi \phi$ is studied to measure the CP -violating ϕ_s phase, a sensitive precision test of the Standard Model. Also, a new algorithm, based on a new type of neural network, is developed to supplement the LHCb High Level Trigger Upgrade to be more versatile and handle more collision data. Last but not least, an automated control system, crucial to the success of the LHCb Ring-imaging Čerenkov (RICH) Detector Upgrade, is deployed to vet components for the upgraded RICH system.

The measurement detailed in Chapter 6 looks at the B_s^0 - \bar{B}_s^0 meson system and $b \rightarrow c\bar{c}s$ transition. The B_s^0 is a neutral meson with quark content $s\bar{b}$, while the \bar{B}_s^0 is made up from $\bar{s}b$. The mass eigenstates of the B_s^0 - \bar{B}_s^0 meson system are a linear combination of the flavour eigenstates. Thus, the meson oscillates between its states through the flavour changing weak current as time goes on. Namely, it is in constant flux between its particle and antiparticle state. So the decay to $J/\psi \phi$ ($b \rightarrow c\bar{c}s$ transition) can occur two ways: $B_s^0 \rightarrow J/\psi \phi$ or $B_s^0 \rightarrow \bar{B}_s^0 \rightarrow J/\psi \phi$. Both decays are allowed at tree level as well as through a QCD penguin diagram in the Standard Model. The ϕ_s phase is thus a measurement that can access interference between mixing and direct decay. The theoretical predictions for ϕ_s are more precise than any measurement performed so far (Figure 1.1). So any experimental improvement will serve as a stringent test of the Standard Model and potential clues for physics beyond.

The analysis reported in this thesis focuses on improving previous measurements done by LHCb (References [14, 15]). One of the main ways to do this was to add new data from LHCb Run 2 data-taking that would increase the statistical significance. Also, several key changes to signal selection strategy helped improve measurement precision, and further enhancements were made to the mass model, removal of peaking backgrounds, and other relevant analysis methods.

Further in the thesis, discussed are some future improvements to the LHCb

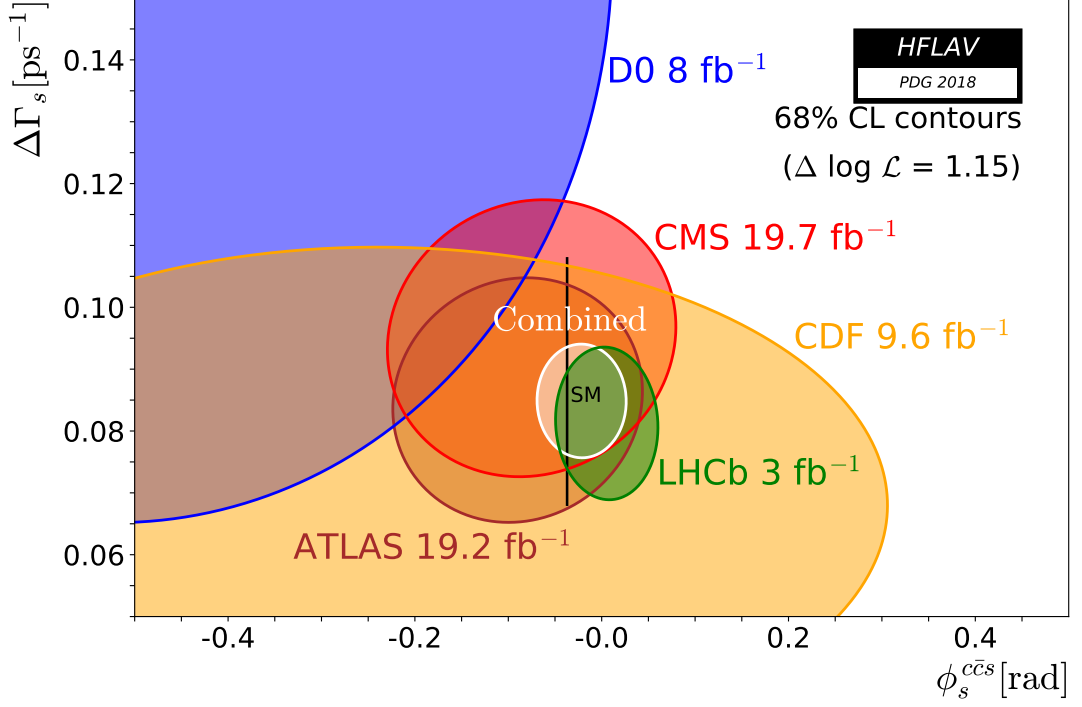


Figure 1.1 *Individual 68% confidence-level contours of ATLAS, CMS, CDF, D0 and LHCb in the $(\phi_s^{c\bar{c}s}, \Delta\Gamma_s)$, their combined contour (solid line and shaded area), as well as the Standard Model predictions (thick black rectangle) as performed by the HFLAV [13] averaging group.*

experiment. For Run 3 at the Large Hadron Collider, the LHCb detector will be upgraded to cope with the new conditions. The luminosity will be increased by a factor of 5 (from $4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ to $2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$), so the collision rate and pileup (overlapping collisions in the detector) will increase. Thus, all of the subdetectors as well as the data acquisition system will be upgraded.

The higher luminosity presents unique challenges for data acquisition. Suppose the current computing model and software framework were kept. In that case, the data storage capacity and computing power required to process data at this rate would exceed the current capacity by at least one order of magnitude [16]. The system currently consists of a Level-0 (L0) hardware trigger, software High Level Triggers (HLT1 and HLT2) and a software Offline Reconstruction. The hardware of L0 trigger has a fixed limit in readout frequency and dead-time that combined with the rest of the hardware limitations of different subdetectors and data acquisition systems inhibits the full detector readout to 1MHz.

During Run 3, the L0 trigger will be removed entirely, and the HLT and Offline Reconstruction will be combined into a common HLT/Turbo stream of data, while

the detector readout rate will increase to 40 MHz [17–19]. The surge in data flow will require novel solutions to cope with detector readout and triggering. So far, most of the fast decisions in the L0 trigger have been pre-programmed based on thresholds. This approach is a basic way to filter the influx of data to the HLT, which in Run 3 will not be available. A more innovative way to approach this problem is to use multivariate algorithms (MVAs). However, MVAs are typically expensive computationally. In Section 4.3, this thesis presents a new type of neural network that can run much faster and still preserve its performance, significantly increasing background rejection and improving signal retention.

Of course, challenges for Run 3 are also directly relevant to the subdetectors themselves. One of the most distinctive features of LHCb is its RICH system, which is instrumental in particle identification (PID) [20]. It provides signal separation for charged hadrons in the momentum range 1.5–100 GeV/ c . This is essential for studying hadronic final states and very important for precision CP violation measurements. One example is the $B_s^0 \rightarrow J/\psi K^+ K^-$ analysis presented in this thesis. The final state includes two charged kaons, and without PID the final state signature will suffer from severe combinatorial background for lacking K - π separation.

Moreover, PID is used to tag particles and antiparticles. For example, distinguishing a b from a \bar{b} by looking at its decay to charged kaons through the $b \rightarrow c \rightarrow s$ chain proves very useful for time-dependent measurements. To accomplish this, the current RICH system consists of two RICH detectors and covers an angular acceptance of 15–300 mrad [21]. The current system is equipped with hybrid photon detectors (HPDs), which have a fixed 1 MHz readout rate in the electronics encapsulated within their tube. Thus, for the upgrade of the system, the HPDs will be replaced with commercial multianode photomultipliers (MaPMTs) with external electronics. Chapter 5 discusses the programme used to select, vet and deliver the new MaPMTs for the LHCb RICH Upgrade. It consists of an entire two-year campaign called the Photon Detector Quality Assurance (PDQA). The PDQA entails designing a test bench, control software and conducting extensive testing and characterisation on 4000 MaPMT units. It required replicating the testing equipment in two separate labs and two different countries, automating the testing process to remove external factors and delivering all fully characterised MaPMTs in a state ready to be mounted in the new RICH detectors.

In this thesis, the decay $B_s^0 \rightarrow J/\psi \phi$ is considered equivalent to the $B_s^0 \rightarrow J/\psi K^+ K^-$ decay, where $m(K^+ K^-)$ is in the mass range of the $\phi(1020)$ resonance.

The inclusion of charge-conjugate processes is implied throughout, unless explicitly stated otherwise.

Chapter 2

Theory

2.1 Standard Model

The Standard Model (SM) of particle physics [22] is the theory describing three of the four known fundamental interactions in nature, all except gravity. Electromagnetic, weak and strong interactions are all incorporated into the Standard Model by elementary particles and their interactions 2.1. Electromagnetism (or quantum electrodynamics, QED) and the weak force are unified into the theory of electroweak interactions (EW) [23–28]. The strong force is described by Quantum Chromodynamics (QCD) [29]. SM is a quantum field theory where the constituents of matter are represented by 12 fundamental types of fermions (fermion fields) plus their antipartners. The fermions are grouped into three generations based on *flavour* and *mass* and further split into two types of quarks (up-type and down-type) and two types of leptons (charged and neutral). The force interactions are realised by twelve fundamental gauge bosons (boson fields). Eight gluons, g , for the strong interaction, the W^+ , W^- , and Z^0 for the weak interaction, and the photon, γ , for the electromagnetic interaction. Finally, a scalar boson (complex scalar doublet field), the Higgs boson (H^0 , discovered in 2012 [30, 31]), is responsible for the interactions giving quarks, fermions and the weak gauge bosons their masses and thus reproducing the observed interactions at all temperatures below a very high characteristic energy scale¹. Figure 2.1 shows the particle content of the Standard Model. The theory is built upon the internal symmetries of the $SU(3)_c \times SU(2)_L \times U(1)_Y$ ².

¹in the Standard Model, the Higgs field characteristic energy scale is about ~ 200 GeV [32].

² c stands for colour, L stands for left-handed, and Y stands for hypercharge

Standard Model of Elementary Particles

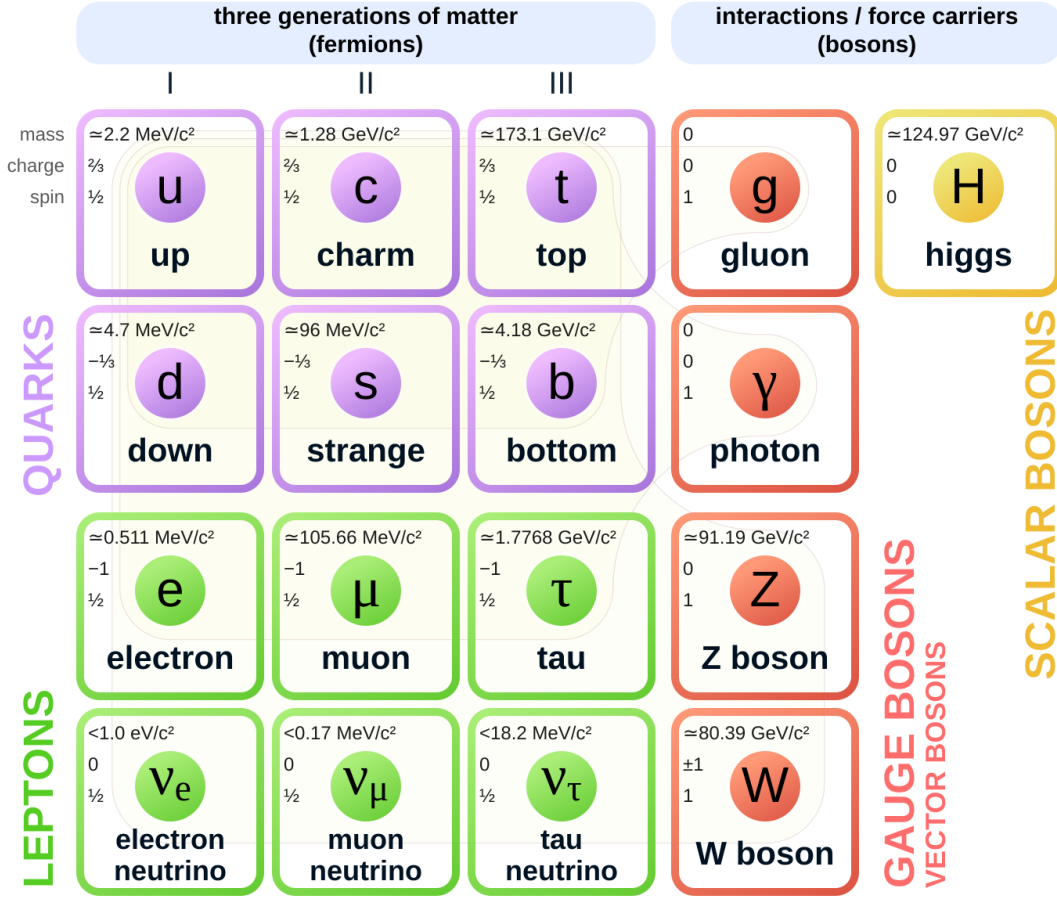


Figure 2.1 *Elementary Particles in the Standard Model: 5 bosons and 12 fermions [33].*

2.1.1 Electroweak (EW) interaction and symmetry breaking (Higgs mechanism)

Electroweak theory describes the electroweak interactions. It is a combination of electromagnetism and the weak interaction.

Electromagnetism (EM) is the everyday interaction that stands behind familiar phenomena such as electricity and magnetism. Its quantum-mechanical extension Quantum Electrodynamics (QED) completes that description to include all fundamental processes between charged particles via the exchange of a photon³. QED completely describes the interaction between light and matter as well as

³the photon (γ) is a massless vector boson.

processes such as chemical interactions and the structure of matter.

On the other hand, the weak interaction, is a short-range phenomenon responsible for processes like radioactive beta decay, nuclear fusion inside stars and neutrino scattering. It is mediated by three massive vector bosons (W^+ , W^- , and Z^0) and couples to all types of fermions, but its charged bosons only couple to left-handed fermions and right-handed antifermions. The “handedness” term refers to the chirality of a particle. Chirality is a fundamental quantum property of each particle type and should not be confused with helicity⁴, a frame dependent quantity for massive particles. The weak interaction is thus a chiral theory, and the interaction is charged under weak isospin. For example, the left-handed electron is charged with weak isospin of $-\frac{1}{2}$ and, therefore, can couple to a neutrino through an exchange of a W boson, while the positron is right-handed, carries 0 weak isospin and cannot. Here, an unconventional distinction is made: the pure eigenstates of the (anti)electron and (anti)positron are not each other’s antiparticle. Instead, they are considered separate particles altogether and thus possess different chirality (Section A.1.1). In reality, the physical particles are linear combinations of the pure weak eigenstates. Contrary to the strong and electromagnetic forces, weak interactions can change quark flavour⁵.

The electroweak interaction is the generalised description and unified theory of electromagnetism and weak interaction. The particles that mediate it are four massless vector bosons: W^1 , W^2 , W^3 with symmetry $SU(2)_L$ and coupling g , and B with symmetry $U(1)_Y$ and coupling g' .

The observed EW interactions mediated by the massive bosons are due to the Higgs mechanism [26–28]. It introduces a complex scalar doublet field with a non-zero vacuum expectation value. The introduction of this field results in a spontaneous breaking of the electroweak symmetry. Thus, electric charge arises as a linear combination⁶ of weak hypercharge (Y_W) and the third component of weak isospin (T_3):

$$Q = T_3 + \frac{1}{2}Y_W \quad (2.1)$$

while the physical bosons acquire mass and appear as a linear combination of the

⁴helicity is the projection of a particle’s spin onto its momentum.

⁵flavour-changing neutral currents (FCNC) needed to change the flavour of all types of fermions are still unobserved.

⁶the linear combination is non-trivial, and any other will result in the Higgs field interacting with the photon, *e.g.* the Higgs boson acquiring an electric charge.

four massless vector bosons (W^1 , W^2 , W^3 and B):

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \quad (2.2)$$

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$$

where the θ_W is a free parameter called the Weinberg angle [24]. Since the axes representing the particles have been rotated by θ_W , a mismatch is introduced between the mass of the Z^0 and W^\pm bosons: $M_W = M_Z \cos \theta_W$. In the low energy limit, $\mathcal{O}(E) < 200 \text{ GeV}$, since the weak bosons are no longer massless, weak influence decreases exponentially with distance and $SU(2)_L \times U(1)_Y$ reduces to $U(1)_{\text{EM}}$ [32].

2.1.2 Strong interaction (QCD)

The strong interaction is responsible for binding quarks together into hadrons, *e.g.* mesons, baryons and nucleons (protons and neutrons) into nuclei. It is charged under the “colour” charge. In contrast to electric charge, colour charge can have three values for quarks: red, green and blue, and three for antiquarks: antired, antigreen and antiblue. This does not mean the particles are indeed coloured in the colloquial sense, a purely visual effect confined to the human experience of electromagnetic waves in the visible spectrum. Instead, the colour charge is a fundamental property of quarks, and the name is only chosen for descriptive purposes – those will become apparent in the following. To form hadrons quarks need to combine such that the overall colour charge is neutral (0 colour charge). Since there are three colour charge options plus their antistates, a neutral colour is when a colour is combined with its anticolour or when all three colours are combined. The latter option is similar to how human vision perceives the combination of red, green and blue light as neutral white light. This statement can be more succinctly expressed as:

$$R\bar{R} = G\bar{G} = B\bar{B} = RGB = \overline{RGB} = 0 \quad (2.3)$$

where R , G , B stand for red, green and blue colour charge, \bar{R} , \bar{G} , \bar{B} stand for antired, antigreen and antiblue colour charge, respectively, while 0 stands for neutral, zero or white colour charge.

Thus, the quantum field theory describing the strong force is named Quantum Chromodynamics, or QCD, (from the Greek word $\chi\rho\tilde{\omega}\mu\alpha$ [khrôma] meaning colour). The mediator particles are eight massless vector boson fields, called gluons. The symmetry of QCD is $SU(3)_c$ with coupling g_s . Gluons have no weak isospin and no electric charge; they carry the colour charge. Each gluon has one of the allowed colour configurations based on the Gell-Mann matrices. Explicitly if:

$$R = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (2.4)$$

are the definitions of colour for quarks, while the equivalent row vectors are the definitions of anticolour for antiquarks, then gluons can carry the following combinations:

$$g^i = \begin{pmatrix} R & G & B \end{pmatrix} \lambda^i \begin{pmatrix} \overline{R} \\ \overline{G} \\ \overline{B} \end{pmatrix} \quad (2.5)$$

where λ^i are the Gell-Mann matrices. Unlike the electroweak interaction, the gauge boson interaction is not suppressed by charge, chirality, mass or flavour, meaning that the coupling between the bosons and quarks is the same. In fact, given the size of g_s at low energy scales and the fact gluons are charged under the interaction they mediate means they can self-interact “strongly”. Consequently, an interesting phenomenon is observed, where the field potential of the strong interaction does not diminish with distance (r). Instead, it remains proportional to $\sim r$ regardless of separation. Consequently, it turns out that when quarks are separated by a significant distance, the energy put into their separation is enough to produce a quark-antiquark pair anew. The latter effect is known as quark or colour confinement [34]. This also means that at low energy scales, perturbation theory cannot be used to extract any meaningful dynamical quantities about strong interactions.

2.1.3 Quarks and leptons

The force interactions in the SM are between the fundamental fermions. Fermions are split into two main kinds: quarks and leptons. Leptons do not interact with the strong force and thus are not confined like quarks to hadrons.

Quarks are the fundamental fermions that interact with all fundamental forces.

The six flavours of quark are split into two types. Up-type are the up (u), charm (c) and top (t) quarks, which have electric charge $+2/3$ and weak isospin $+1/2$. Down-type are the down (d), strange (s) and bottom, also known as beauty, (b) quarks with electric charge $-1/3$ and weak isospin $-1/2$. Quarks cannot be observed in free unbound colour-doublet states because of colour confinement and are only found in colour-singlet states – hadrons. Hadrons can be either mesons, quark-antiquark ($q\bar{q}$) states, or baryon states with three or more quarks. Most of the time, baryons are represented by qqq or $\bar{q}\bar{q}\bar{q}$ states. However, some exotic hadron states can be $q\bar{q}qq$, $qqq\bar{q}$ or similar. The meson colour-singlet state is a linear combination of the colour states:

$$\frac{1}{\sqrt{3}}(R\bar{R} + G\bar{G} + B\bar{B}) \quad (2.6)$$

and similarly for the baryon (qqq):

$$\frac{1}{\sqrt{6}}(RGB - RBG + GBR - GRB + BRG - BGR) \quad (2.7)$$

On the other hand, leptons are fermions not interacting with the strong force and are split into electrically charged and electrically neutral. The charged leptons are the electron (e), muon (μ) and tau (τ) with electric charge -1 and weak isospin $-1/2$. The neutral leptons are the corresponding neutrinos: the electron (ν_e), muon (ν_μ) and tau (ν_τ) neutrino with 0 electric charge and weak isospin $+1/2$.

2.1.4 Beyond the Standard Model (BSM) Physics

While the Standard Model has been very successful and accurate at describing a wide range of observable phenomena in our universe, it is far from a complete theory of everything and does not encompass all observations. For example, the Standard Model is currently incompatible with General Relativity, the most accurate theory describing the gravitational phenomena. Furthermore, the current framework of SM does not account for the observed amount of matter-antimatter (baryon) asymmetry in the universe. However, while the SM does not account for the above-mentioned, there is little evidence to contradict it. At the time of writing of this thesis, the only observational evidence directly contradicting the model is that neutrinos oscillate between flavour states,

indicating that they possess non-zero mass, meaning they flavour eigenstates do not align with their observable mass eigenstates [35–37].

Consequently, the goal of experiments in particle physics is to subject the Standard Model to more and more stringent tests to identify other effects that do not align with it and eventually arrive at a better and more complete theory.

2.2 CP Violation

Modern theoretical models have relied on the principle of symmetries to describe physical observations. Emmy Noether showed at the beginning of the 20th century that continuous symmetries lead to conserved quantities and that conserved quantities indicate symmetries in the system [38, 39]. However, the Standard Model includes several discrete symmetries, while Noether’s theorem only applies to continuous symmetries of closed systems. For example, the symmetry concerning continuous translation in space implies momentum conservation. However, there is no continuous(-ly differentiable) translation that would equate to a coordinate reversal (a parity transformation). Hence, Noether’s theorem does not directly apply to parity symmetry⁷. Charge conjugation (C), parity (P) and time-reversal (T) symmetries are all discrete symmetries, and they all have a special place in the Standard Model but are not under the umbrella of Noether’s theorem.

Explicitly, the C operator converts a particle (ψ) to its antiparticle or vice versa, $\psi \rightarrow \bar{\psi}$, the P operator inverts the spatial coordinate, $r \rightarrow -r$, and the T operator reverses the direction of time, or in other words, inverts the sign of the time coordinate, $t \rightarrow -t$ (Section A.2.1). Thus, the C , P and T symmetries are represented by an observable being invariant under the corresponding transformation.

CP symmetry thus implies invariance under both transformations combined:

$$CP |\psi(r, t)\rangle = |\bar{\psi}(-r, t)\rangle \quad (2.8)$$

Electromagnetic and strong interactions are known to respect both C , P and also their combination CP . In the case of the strong interactions CP symmetry

⁷Extensions of Noether’s theorem to discrete systems exist (*e.g.* [40, 41]), but these are generally not directly applicable to the topic of this thesis.

is associated with the *strong CP problem*. Essentially, the formulation of QCD theory does not forbid CP violation but actually contains two competing terms, one for the fermion mass with a chiral phase and another for gluon pseudoscalar density, both contributing to CP violation [42]. These two competing terms cannot be cancelled simultaneously since their contributions are phase dependent inversely, meaning their phase difference is constant:

$$\theta' + \theta \rightarrow (\theta' - \alpha) + (\theta + \alpha) = \theta' + \theta = \bar{\theta} \quad (2.9)$$

where α is the chiral transformation angle needed to cancel the quark mass term, θ' and θ are the phases of the quark mass term and the gluon density, respectively, while $\bar{\theta}$ is the resultant CP -violating phase. However, the strong interaction has never been observed to violate CP symmetry.

Conversely, the weak interaction and so electroweak interactions do not preserve it. Charged electroweak interactions do not couple to right-handed particles, and so they violate P symmetry. They also do not couple to left-handed antiparticles and thus violate C symmetry. Finally, these interactions seemed to follow CP symmetry because they involve both left-handed particles and right-handed antiparticles. However, they were observed to violate CP symmetry in the $K_L^0 \rightarrow \pi^+\pi^-$ decay, which won James Cronin and Val Fitch their Nobel prize in 1980 [43, 44].

Coincidentally, C and CP violations are needed to satisfy the second of Sakharov's conditions [8]. This has implications in finding an explanation for the observed matter-antimatter asymmetry in the universe. It would allow for certain physical processes⁸ to happen with different rates depending on their CP state and thus provide a mechanism for matter to be treated preferentially.

In the context of this thesis, three types of CP violation will be considered: CP violation in decay (direct), CP violation in mixing (indirect), CP violation in interference between mixing and decay.

2.2.1 CP violation in decay

CP violation in decay is a difference in the rates between CP conjugate decays.

⁸specifically processes violating baryon number conservation.

For a state, $|\psi\rangle$, decaying to a final state, $|f\rangle$, the respective amplitudes will be:

$$A_f = \langle f | H | \psi \rangle \text{ and } \bar{A}_{\bar{f}} = \langle \bar{f} | H | \bar{\psi} \rangle \quad (2.10)$$

where H is the Hamiltonian (Section A.2.2). CP violation occurs when:

$$\frac{\bar{A}_{\bar{f}}}{A_f} \neq 1 \quad (2.11)$$

since the decay rate (probability) is the square of the amplitude.

A total decay amplitude is taken to be the sum of the individual amplitudes corresponding to all modes (processes) involved. The individual amplitudes are complex-valued and carry two phases: a strong phase, δ , and a weak phase, φ . The strong phase is invariant under a CP transformation as it follows from Section 2.1.2. However, the weak phase is conjugated. A general form of an amplitude and its CP counterpart can be written as:

$$\begin{aligned} A_f &= \sum_n |A_n| e^{i\delta_n} e^{i\varphi_n} \\ \bar{A}_{\bar{f}} &= \sum_n |A_n| e^{i\delta_n} e^{-i\varphi_n} \end{aligned} \quad (2.12)$$

This puts an interesting constraint on the possible CP -violating processes. For example, constructing a CP -violating amplitude with just one mode is impossible:

$$\begin{aligned} |A_f|^2 &= A_f A_f^* = |A_1|^2 e^{i(\delta_1 - \delta_1)} e^{i(\varphi_1 - \varphi_1)} = |A_1|^2 \\ |\bar{A}_{\bar{f}}|^2 &= \bar{A}_{\bar{f}} \bar{A}_{\bar{f}}^* = |A_1|^2 e^{i(\delta_1 - \delta_1)} e^{-i(\varphi_1 - \varphi_1)} = |A_1|^2 \end{aligned} \quad (2.13)$$

Thus, direct CP violation can only occur for decays that have multiple contributing modes. For example with two modes:

$$\begin{aligned} |A_f|^2 &= A_f A_f^* = |A_1|^2 + |A_2|^2 + |A_1||A_2| e^{i(\delta_1 - \delta_2)} e^{i(\varphi_1 - \varphi_2)} \\ &\quad + |A_1||A_2| e^{-i(\delta_1 - \delta_2)} e^{-i(\varphi_1 - \varphi_2)} \\ &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2|(\cos(\Delta\delta) \cos(\Delta\varphi) - \sin(\Delta\delta) \sin(\Delta\varphi)) \\ |\bar{A}_{\bar{f}}|^2 &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2|(\cos(\Delta\delta) \cos(\Delta\varphi) + \sin(\Delta\delta) \sin(\Delta\varphi)) \end{aligned} \quad (2.14)$$

where $\Delta\delta = \delta_1 - \delta_2$ and $\Delta\varphi = \varphi_1 - \varphi_2$ with further clarifications in Section A.2.3.

The resulting difference:

$$|A_f|^2 - |\bar{A}_{\bar{f}}|^2 = -4|A_1||A_2| \sin(\Delta\delta) \sin(\Delta\varphi) \quad (2.15)$$

contains both strong and weak phases, but is purely existent because of the weak phase transforming (changing sign) under CP . Otherwise, if the weak phase did not acquire an extra minus sign (Equation 2.12 and Section A.2.3), the phases will cancel and the result would lack any phase contributions. In experiments, this difference can be quantified by the \mathcal{A}_{CP} observable, defined as:

$$\mathcal{A}_{CP} = \frac{|A_f|^2 - |\bar{A}_{\bar{f}}|^2}{|A_f|^2 + |\bar{A}_{\bar{f}}|^2} = \frac{\Gamma(\psi \rightarrow f) - \Gamma(\bar{\psi} \rightarrow \bar{f})}{\Gamma(\psi \rightarrow f) + \Gamma(\bar{\psi} \rightarrow \bar{f})} \quad (2.16)$$

where $\Gamma(\psi \rightarrow f)$ is the partial decay width of $\psi \rightarrow f$ decay.

2.2.2 CP violation in mixing

As for CP violation in mixing, one can look at the process of neutral particle oscillations. Neutral particle oscillations can occur when a neutral particle has distinct internal quantum eigenstates which do not coincide with its external energy (mass) eigenstate. For example, some neutral mesons have this property, allowing them to transmute or oscillate between their particle and antiparticle state without breaking any conservation laws.

Consider a meson with two distinct flavour eigenstates, $|\psi\rangle$ and $|\bar{\psi}\rangle$. They cannot also be CP eigenstates, as the CP operator transforms one into the other. Thus, the CP eigenstates must be a mixture of the pure flavour states:

$$|\psi_+\rangle = \frac{1}{\sqrt{2}}|\psi\rangle - \frac{1}{\sqrt{2}}|\bar{\psi}\rangle \quad \text{and} \quad |\psi_-\rangle = \frac{1}{\sqrt{2}}|\psi\rangle + \frac{1}{\sqrt{2}}|\bar{\psi}\rangle \quad (2.17)$$

where the subscript indicates the CP eigenvalue (+ is even, - is odd):

$$CP|\psi_+\rangle = +|\psi_+\rangle \quad \text{and} \quad CP|\psi_-\rangle = -|\psi_-\rangle \quad (2.18)$$

By the same argument, the flavour eigenstates are transformed by the weak interactions, and they cannot be weak eigenstates:

$$|\psi_1\rangle = p|\psi\rangle - q|\bar{\psi}\rangle \quad \text{and} \quad |\psi_2\rangle = p|\psi\rangle + q|\bar{\psi}\rangle \quad (2.19)$$

where $|\psi_1\rangle$ and $|\psi_2\rangle$ are orthogonal superpositions of $|\psi\rangle$ and $|\bar{\psi}\rangle$. Here, p and q are complex mixing amplitudes, and CP violation in mixing occurs when the rate of mixing is different in one direction from the other, in other words:

$$\frac{q}{p} \neq 1 \quad (2.20)$$

where the amplitudes do have the same magnitude.

2.2.3 CP violation in mixing and decay

Finally, the third way CP can manifest itself is in combining the above two processes, *i.e.*, in the interference of mixing and decay. This can happen when a neutral particle's distinct flavour eigenstates, $|\psi\rangle$ and $|\bar{\psi}\rangle$, can both decay to the same final state f . In this scenario, there are two ways for each flavour state to transition into the final state: a state can oscillate into its anti-state and then the anti-state can decay to the final state, or it can directly decay into the final state. CP violation, in this case, is due to the difference in weak phases of the amplitudes of both possible paths. Let the weak phase of the direct decay be φ_D , while the mixing amplitude have a weak phase φ_M . The difference between the two paths will be:

$$\phi = [\varphi_M + (-\varphi_D)]_{\text{path 1}} - [\varphi_D]_{\text{path 2}} = \varphi_M - 2\varphi_D \quad (2.21)$$

Since mixing and decay are independent processes, one can define a total measure of the CP asymmetry, λ by the product of the respective conditions defined in Equations 2.11 and 2.20:

$$\lambda \equiv \frac{q}{p} \frac{\bar{A}_f}{A_f} \quad \text{and} \quad \bar{\lambda} \equiv \frac{p}{q} \frac{A_{\bar{f}}}{\bar{A}_{\bar{f}}} \quad (2.22)$$

As per previous arguments, f can only be a CP eigenstate if $\lambda = \bar{\lambda}$, meaning the CP operator does not transform f into some other state. Similarly, if $\lambda \neq 1$, then CP is violated. Experimentally, this can be observed in the time-dependent decay rate asymmetry, defined as:

$$\mathcal{A}_{CP}(t) = \frac{d\Gamma(\psi \rightarrow f)/dt - d\Gamma(\bar{\psi} \rightarrow f)/dt}{d\Gamma(\psi \rightarrow f)/dt + d\Gamma(\bar{\psi} \rightarrow f)/dt} \quad (2.23)$$

Curiously, the definition of λ allows CP to be violated here without being broken individually in decay or mixing. This can happen by λ acquiring phase, $\arg(\lambda) \neq 0$, while $|\lambda| = 1$.

2.3 CKM Matrix

The weak interaction has only been observed to couple between the same generation of leptons (*e.g.* e and ν_e , but not μ and ν_τ). However, it can couple between different generations of quarks. This process is known as quark mixing. To date, the only observed CP violation [43, 44] is in quark mixing⁹, and the mixing coupling strengths between quarks are described in the Standard Model by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [45, 46].

By convention, only the down-type quarks take part in the mixing process. This is because the up-type quark weak and mass eigenstates are set to overlap, while the down-type weak eigenstates are disparate from their mass eigenstates. For example, if d' is the weak eigenstate of the down quark, it will be the one interacting with the weak eigenstate, u , of the up quark at the flavour-changing W^\pm vertex. More generally, the down-type quark weak states are constructed by a rotation of the down-type mass eigenstates:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} \equiv V_{\text{CKM}} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} \quad (2.24)$$

where d' , s' , b' are the weak eigenstates, the elements V_{ij} are numbers representing corresponding $i - j$ coupling strengths and d , s , b are the corresponding mass eigenstates. The CKM matrix must be a complex matrix that must also be unitary ($V_{\text{CKM}} V_{\text{CKM}}^* = V_{\text{CKM}}^* V_{\text{CKM}} = I$) to preserve probability amplitudes.

In more detail, when transforming from quarks to antiquarks, the CKM matrix elements in the weak states are replaced with their complex conjugates. If there were no complex phase, the CKM matrix would be real-valued, and the weak interactions will be invariant under a CP transformation. However, as per observations, CP -asymmetry does exist and introducing the complex phase is the means to break CP symmetry and give rise to CP violation. Thus, the V_{ij}

⁹other potential sources are the strong interaction and neutrino oscillations but continue to remain unobserved.

elements are complex, and the by-element unitarity conditions are as follows:

$$\sum_{n \in \{d,s,b\}} V_{in}^* V_{jn} = \delta_{ij} \quad \text{and} \quad \sum_{n \in \{u,c,t\}} V_{ni}^* V_{nj} = \delta_{ij} \quad (2.25)$$

These conditions and the fact that some of the complex phases can be absorbed in the arbitrary phases of individual quark fields limit the number of independent physical parameters needed to specify the values of V_{CKM} fully. In this case, the number of free parameters is 4. In the Standard Model, their expression is convention dependent and frequently being three mixing angles and one complex CP -violating phase. Another commonly chosen parametrisation is the so-called Wolfenstein parametrisation which has the benefit of real-valued V_{ud} , V_{us} , V_{cb} and V_{tb} and a complex phase. The Wolfenstein parametrisation still only uses 4 parameters: λ , A , ρ , η . They are all real and defined in terms of the following:

$$\lambda \equiv \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}} \quad A\lambda^2 \equiv \frac{|V_{cb}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}} \quad A\lambda^3(\rho + i\eta) \equiv V_{ub}^* \quad (2.26)$$

where the complex phase is introduced with the combination of ρ and η parameters. While it looks strange initially, this parametrisation is motivated by its explanatory power. Namely, the diagonal elements showing the coupling strengths between quarks of the same generation are approximately equal to 1 in magnitude. While the magnitudes of off-diagonal elements responsible for mixing between first and second generation are 4 to 5 times smaller. Second to third generation and first to third generation mixing elements are further suppressed by higher powers of λ^2 and λ^3 respectively. Neglecting higher-order terms (up to $\mathcal{O}(\lambda^3)$), the V_{CKM} matrix can be expanded in terms of the small λ parameter:

$$V_{\text{CKM}} \approx \begin{pmatrix} 1 - \frac{1}{2}\lambda & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} \quad (2.27)$$

A common suppression factor is not evident, which makes elements farther from the diagonal smaller. This is called Cabibbo suppression after Nicola Cabibbo, who introduced the first of such factors¹⁰ for the then-only two generations of quarks known.

Still, these are all convention dependent parametrisations, while a CP violation measure must not be. To construct such an invariant measure, the unitarity

¹⁰Cabibbo angle, θ_c

conditions are used. Let us examine some of the conditions from Equation 2.25. For example:

$$\begin{aligned}
V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* &= 0 \quad | : V_{cd}V_{cb}^* \\
\implies \\
1 + \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} &= 0
\end{aligned} \tag{2.28}$$

Here, the “b–d” condition has been normalised as per convention so that one of its elements is unit length. This condition is associated with the B_d^0 meson, while the B_s^0 meson is associated with the following:

$$\begin{aligned}
V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* &= 0 \quad | : V_{cs}V_{cb}^* \\
\implies \\
1 + \frac{V_{us}V_{ub}^*}{V_{cs}V_{cb}^*} + \frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*} &= 0
\end{aligned} \tag{2.29}$$

An observation can be made about the complex phases of these combinations of CKM elements. It is interesting to note that the phase is of the form:

$$\arg\left(\frac{V_{ij}V_{i\beta}^*}{V_{\alpha j}V_{\alpha\beta}^*}\right) = \arg(V_{ij}V_{\alpha\beta}V_{i\beta}^*V_{\alpha j}^*) \tag{2.30}$$

The quark indices appear twice for each quark in the expression, once in the complex element and once for the complex conjugate element. As previously pointed out, each quark field has absorbed a phase, which cancels out phases for the combination of four elements in this way. These combinations become convention independent.

Using these properties, the so-called unitarity triangles can be constructed in the complex plane. The unit size elements serve as the unit length side oriented in the direction of the real number line and starting at zero. The complex-valued elements complete the rest of the sides, summing back to zero and completing the triangle. The d–b and s–b triangles are shown in Figure 2.2. The coordinates of

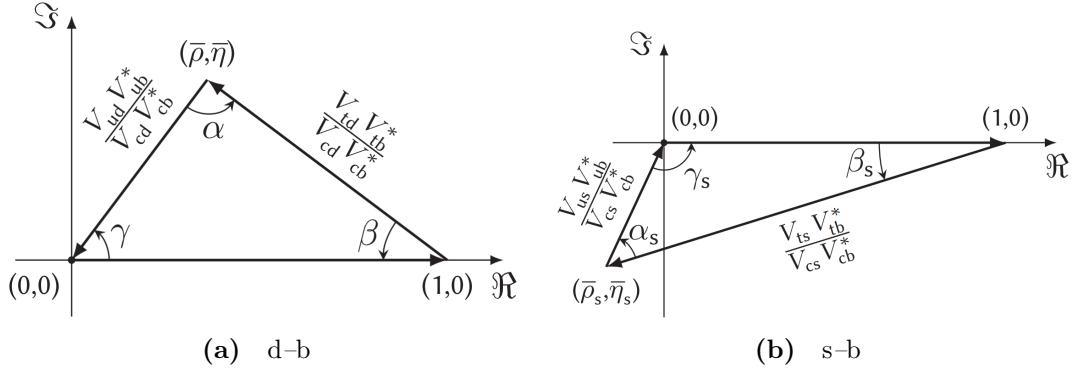


Figure 2.2 CKM Unitary triangles: (a) d-b unitary triangle constructed from Equation 2.28, (b) s-b unitary triangle constructed from Equation 2.29

the vertex solely in the complex plane are defined to be $(\bar{\rho}, \bar{\eta})$ and $(\bar{\rho}_s, \bar{\eta}_s)$ for the d-b and s-b triangles, respectively. The angles of the triangles also have names per convention, which for the d-b triangle are:

$$\begin{aligned} \alpha \equiv \phi_2 &\equiv \arg \left(-\frac{V_{td} V_{tb}^*}{V_{ud} V_{ub}^*} \right) & \beta \equiv \phi_1 &\equiv \arg \left(-\frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*} \right) \\ \gamma \equiv \phi_3 &\equiv \arg \left(-\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*} \right) \end{aligned} \quad (2.31)$$

while for the s-b triangle are:

$$\begin{aligned} \alpha_s \equiv \phi_{2_s} &\equiv \arg \left(-\frac{V_{us} V_{ub}^*}{V_{ts} V_{tb}^*} \right) & \beta_s \equiv \phi_{1_s} &\equiv \arg \left(-\frac{V_{ts} V_{tb}^*}{V_{cs} V_{cb}^*} \right) \\ \gamma_s \equiv \phi_{3_s} &\equiv \arg \left(-\frac{V_{cs} V_{cb}^*}{V_{us} V_{ub}^*} \right) \end{aligned} \quad (2.32)$$

The β_s phase is by this construction quite small. That is important for the CP violation measurement of the observable phase ϕ_s , since it is approximately equal to $-2\beta_s$ in the Standard Model. Thus, even if new unknown physics contributions are small, their effect could still be significant compared to β_s . That makes precise measurements of ϕ_s quite a sensitive probe for BSM physics.

2.4 CP violation in the B_s^0 sector

One of the preferred ways to discover new physics effects is to challenge the current theoretical models. Specifically, when it comes to CP violation, this is related to confirming or refuting the CKM unitarity triangles. The currently-

accounted-for CP violation due to phenomena related to the CKM matrix is too small to explain the observed matter-antimatter asymmetry.

For example, the observed matter-antimatter asymmetry discussed in Chapter 1, puts the fraction of antimatter per matter at about $f < 10^{-9}$. Furthermore, nucleosynthesis constraints require a baryon number to entropy ratio in the observed part of the universe $\eta_B/s \sim (4,6)10^{-11}$ at a temperature $T > 1$ GeV [47]. Since there is no known mechanism for matter-antimatter separation on cosmological scales, it is assumed that the laws of physics governing the microscopic scales are responsible for providing a process creating the baryonic excess [8, 47]. This process is called baryogenesis.

The Big Bang theory states that the early Universe was hot and dense with temperatures ranging from a few eV up to the Planck scale $M_{Pl} \sim 10^{19}$ GeV [48]. As the Universe cooled down it went through several phase transitions. The highest temperatures phase transitions are important for grand unified theories (GUTs) and may have an impact on cosmic inflation. At lower temperatures, about $M_W \sim 100$ GeV, the electroweak phase transition occurs late in the cosmic evolution compared to GUT transitions and is mainly important for baryogenesis [47, 48]. Sakharov's condition [8] of processes occurring outside thermal equilibrium can be satisfied by such a phase transition. That is because it would occur on the surfaces of expanding bubbles where the boundary sits around the propagation of the newly-reached non-zero vacuum expectation value of the Higgs field when $M_W < M_H^{\text{crit}}$. These bubbles are growing and filling the space where the Higgs field is still in its pre-existing state.

The contributions of CP violation in baryogenesis has been studied extensively and is believed to be much smaller than expected. Specifically, to explain the observed η_B/s as above, the effective asymmetry from CP needs to be $\Delta_{CP} \geq 10^{-8}$ [47, 49, 50]. However, best estimates due to the known sources of CP violation in the Standard Model put this value at $\Delta_{CP} \leq 10^{-21}$ [47, 49, 50]. Thus, CKM processes cannot account for at many orders of magnitude difference with the observed baryon asymmetry.

Figure 2.3 shows the global fit to the (d-b) unitary triangle in the $(\bar{\rho}, \bar{\eta})$ plane performed by *The CKMFitter Group* [51], which combines measurements from multiple sources (LHCb, BaBar, Belle, *etc.*). Similarly, Figure 2.4 shows the global fit of the (s-b) unitary triangle in the $(\bar{\rho}_s, \bar{\eta}_s)$ plane. It can be clearly seen that the size of the β_s angle is much smaller than that of the β angle.

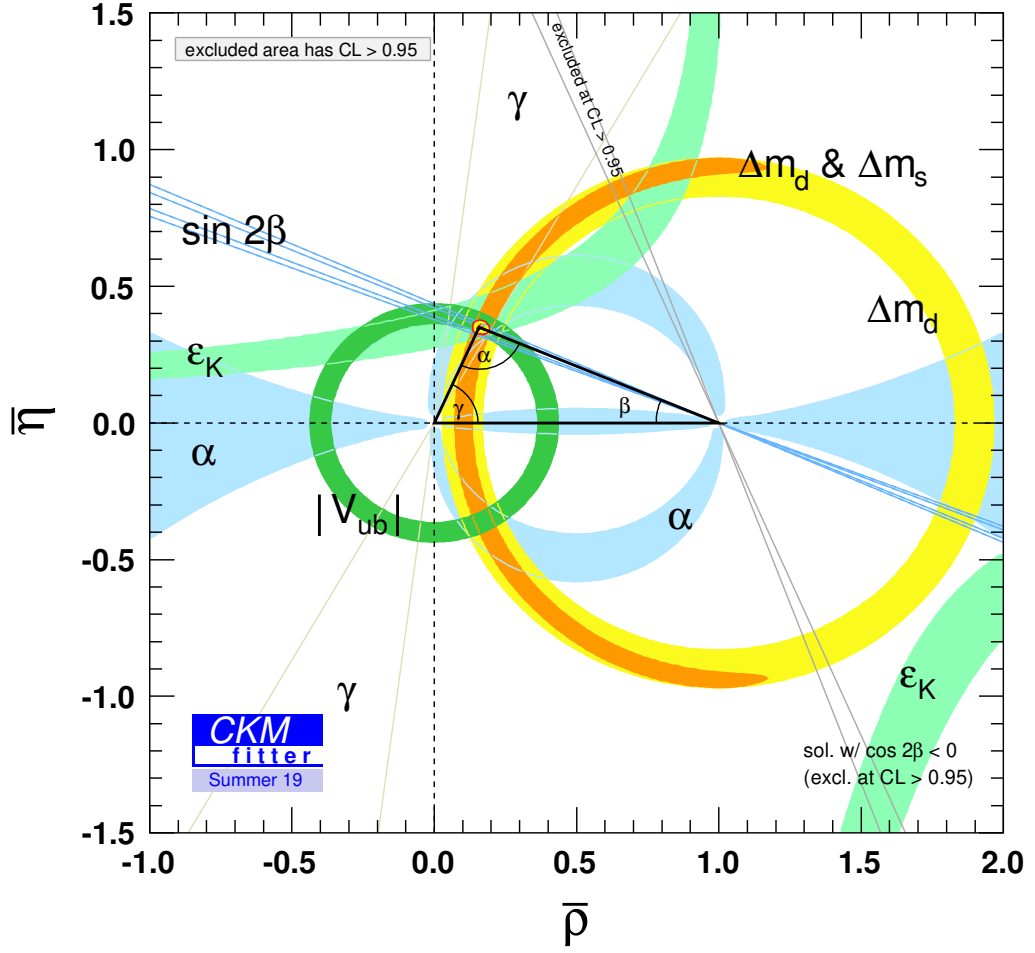


Figure 2.3 Global fit to (d-b) CKM unitary triangle in $(\bar{\rho}, \bar{\eta})$ plane [51]

Thus, the neutral B_s^0 meson and its $B_s^0 \rightarrow J/\psi \phi$ decay offer a candidate to improve upon the currently known sources of CP violation. The decay is associated with an observable phase, ϕ_s , due to interference between mixing in the $B_s^0 - \bar{B}_s^0$ system and the $B_s^0 \rightarrow J/\psi \phi$ decay. And the phase, ϕ_s , is yet to be measured to a high enough precision to unambiguously confirm or refute the current Standard Model theoretical predictions.

2.4.1 $B_s^0 - \bar{B}_s^0$ mixing

The B_s^0 meson is a neutral meson, a bound state of an \bar{b} and an s , capable of the neutral particle oscillations discussed in Section 2.2. The lowest order diagrams of this oscillation are shown in Figure 2.5. Thus, the $B_s^0 - \bar{B}_s^0$ oscillations

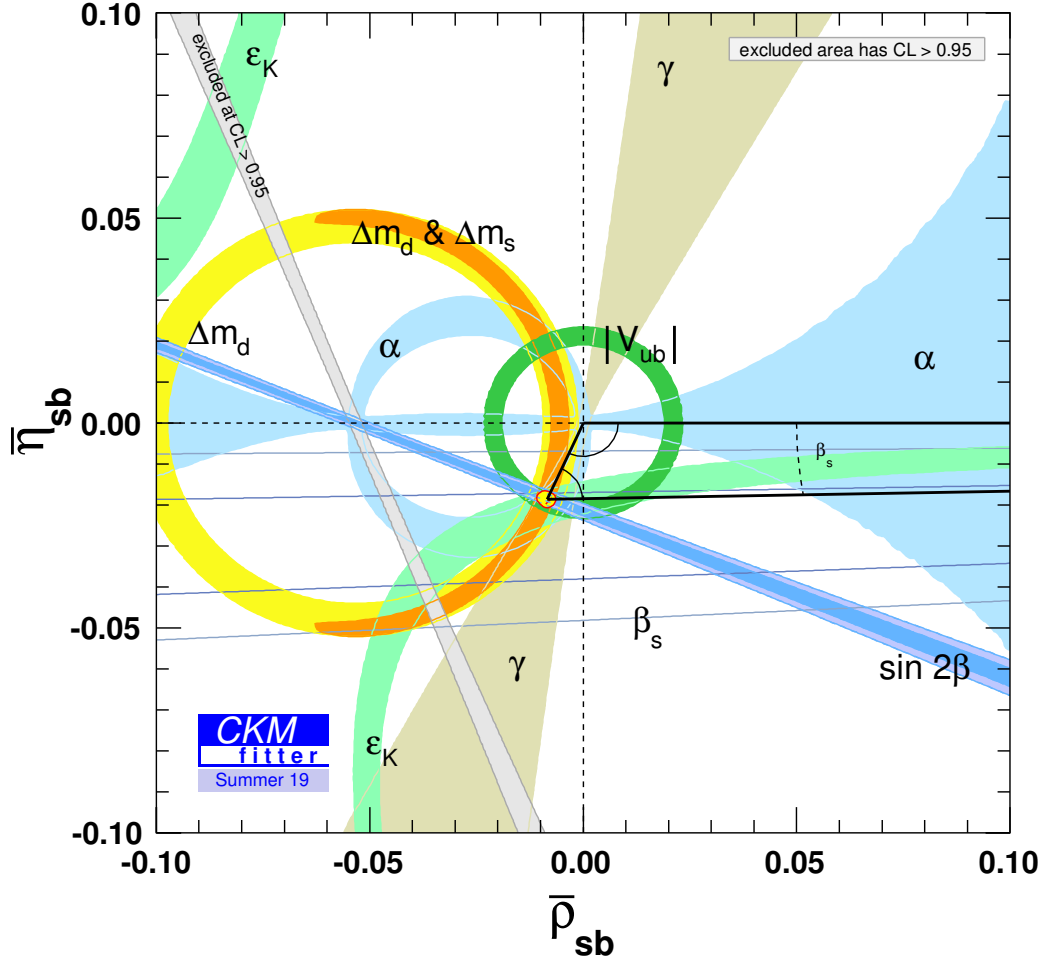


Figure 2.4 Global fit to (s-b) CKM unitary triangle in $(\bar{\rho}_s, \bar{\eta}_s)$ plane [51]

can permit CP violation if over time one of the transitions (either $B_s^0 \rightarrow \bar{B}_s^0$ or $\bar{B}_s^0 \rightarrow B_s^0$) happens with a different frequency, resulting in one of the transitions being preferred.

The equation governing the time-evolution of the B_s^0 - \bar{B}_s^0 systems is as follows:

$$i\hbar \frac{d}{dt}(|B(t)\rangle) = \left(\mathbf{M} - \frac{i}{2} \mathbf{\Gamma} \right) |B(t)\rangle \quad (2.33)$$

where \mathbf{M} and $\mathbf{\Gamma}$ are the mass and decay matrix respectively (Section A.2.4). The equations is subject to two solutions of the form:

$$|B_k(t)\rangle = e^{(-im_k - \Gamma/2)t} |B_k(0)\rangle \quad (2.34)$$

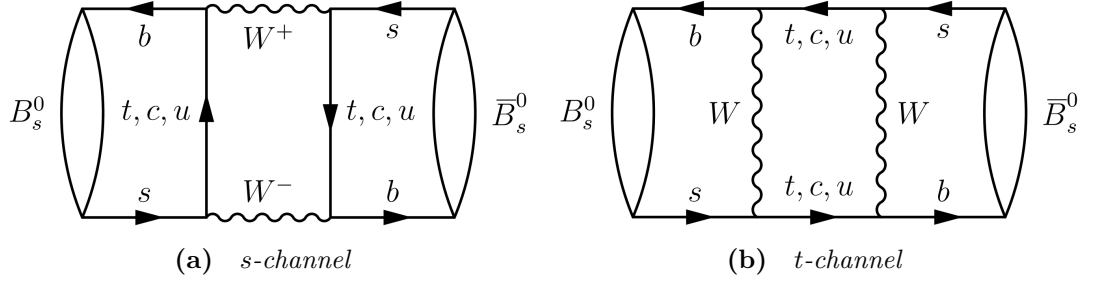


Figure 2.5 *Lowest order B_s^0 - \bar{B}_s^0 mixing diagrams: (a) s-channel, (b) t-channel [15]*

where k denotes the two possible mass eigenstates, normally referred to as the heavy (‘H’) and the light (‘L’) states. They are mixtures of the flavour eigenstates:

$$|B_L\rangle = p |B_s^0\rangle - q |\bar{B}_s^0\rangle \quad (2.35)$$

$$|B_H\rangle = p |B_s^0\rangle + q |\bar{B}_s^0\rangle \quad (2.36)$$

where p and q are complex numbers equivalent from the parametrisation in Equation 2.19. The time evolution of the B_s^0 - \bar{B}_s^0 system is governed by:

$$\begin{aligned} |B_s^0(t)\rangle &= g_+(t) |B_s^0\rangle + \frac{q}{p} g_-(t) |\bar{B}_s^0\rangle \\ |\bar{B}_s^0(t)\rangle &= g_+(t) |\bar{B}_s^0\rangle + \frac{p}{q} g_-(t) |B_s^0\rangle \end{aligned} \quad (2.37)$$

where $|B_s^0\rangle = |B_s^0(t=0)\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|\bar{B}_s^0\rangle = |\bar{B}_s^0(t=0)\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and the functions g_{\pm} are specified in Equation A.20 in Section A.2.4.

As the B_s^0 - \bar{B}_s^0 mixing diagrams are dominated by internal t transitions, the mixing phase, ϕ_M , in the system is thus proportional to:

$$\phi_M = \arg(M_{12}) \approx \arg([V_{ts}V_{tb}^*]^2) \quad (2.38)$$

which is not directly observable but contributes to the ratio q/p as detailed in Section A.2.4.

2.4.2 $b \rightarrow c\bar{c}s$ transitions

The $B_s^0 \rightarrow J/\psi \phi$ decay is due to a $b \rightarrow c\bar{c}s$ transition, which happens mainly through two decay diagrams, a tree diagram (Figure 2.6a) and a penguin diagram

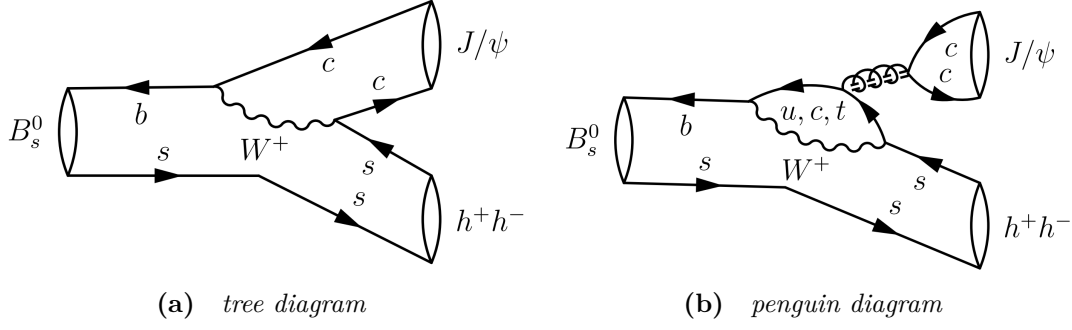


Figure 2.6 $B_s^0 \rightarrow J/\psi \phi$ decay diagrams: (a) lowest order tree diagram, (b) penguin diagram [15].

(Figure 2.6b). The total amplitude is the sum of the individual tree and penguin amplitudes:

$$A_{\bar{c}cs} = V_{cs}V_{cb}^*(A^{tree} + A_c^{peng}) + V_{us}V_{ub}^*A_u^{peng} + V_{ts}V_{tb}^*A_t^{peng} \quad (2.39)$$

Using the unitarity relation from Equation 2.29, this can be reduced to:

$$A_{\bar{c}cs} = V_{cs}V_{cb}^*(A^{tree} + A_c^{peng} - A_t^{peng}) + V_{us}V_{ub}^*(A_u^{peng} - A_t^{peng}) \quad (2.40)$$

As per the discussed Wolfenstein parametrisation in Section 2.3, $V_{us}V_{ub}^*$ is suppressed by a quadratic factor of λ^2 compared to $V_{cs}V_{cb}^*$. Thus, the amplitude may be approximately considered to be:

$$A_{\bar{c}cs} = V_{cs}V_{cb}^*(A^{tree} + A_c^{peng} - A_t^{peng}) + \mathcal{O}(\lambda^2) \quad (2.41)$$

Therefore the weak direct decay phase, φ_D , will approximately be $\varphi_D \approx \arg(V_{cs}V_{cb}^*)$. Since the $J/\psi \phi$ final state is its own charge conjugate, *i.e.* $|f\rangle = |\bar{f}\rangle$, the relative phase difference between a B_s^0 decaying into $|f\rangle$ and \bar{B}_s^0 decaying into the same $|f\rangle$ is:

$$\frac{\bar{A}_f}{A_f} = \eta_f \left| \frac{\bar{A}_f}{A_f} \right| e^{i(2\varphi_D - \delta_n)} \sim e^{i(2\varphi_D - \delta_n)} \quad (2.42)$$

with $\eta_f = \pm 1$ indicating the CP eigenvalue in the case the final state is a CP eigenstate and δ_n is the unobservable phase from Equation 2.12. The total mixing and decay process results in:

$$\lambda_f = \frac{q}{p} \frac{\bar{A}_f}{A_f} \sim e^{i(\varphi_M - \delta_n)} e^{i(2\varphi_D - \delta_n)} = e^{-i(\varphi_M - 2\varphi_D)} \quad (2.43)$$

similarly to Equation 2.21. The phase of λ_f is thus called ϕ_s :

$$\phi_s = -\arg(\lambda_f) = \varphi_M - 2\varphi_D \quad (2.44)$$

In the Standard Model this can be expressed with the elements of the CKM matrix to give:

$$\phi_s = \arg([V_{ts}V_{tb}^*]^2) - 2\arg(V_{cs}V_{cb}^*) = 2\arg\left[\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right] = -2\beta_s \quad (2.45)$$

Thus, in the Standard Model ϕ_s is predicted to be very small: $\phi_s^{\text{SM}} = -0.03696_{-0.0084}^{+0.0072}$ rad [52, 53]. Given the size of this prediction, any experimental deviations from the ϕ_s^{SM} value are effectively deviations from zero. This makes measuring ϕ_s very sensitive to any new effects. Similarly, the phase difference, ϕ between mixing and decay matrix elements, M_{12} and Γ_{12} , can also be affected¹¹. Specifically, M_{12} is more sensitive as the mixing occurs through a box diagram, while the allowed tree-level decay will suppress the possible contributions to Γ_{12} . The latter can manifest itself in the measured value of $\Delta\Gamma_s = 2|\Gamma_{12}|\cos\phi$.

2.4.3 B_s^0 - \bar{B}_s^0 mixing and $B_s^0 \rightarrow J/\psi \phi$ decay

As per the discussion in Sections 2.2.2 and A.2.4, the flavour states of B_s^0 - \bar{B}_s^0 system are not CP eigenstates. The B_s^0 - \bar{B}_s^0 mass eigenstates are a mixture of the flavour eigenstates and the system oscillates between its observable states B_L and B_H in time. Also, the final state of the $B_s^0 \rightarrow J/\psi \phi$ decay cannot be CP eigenstates, since the decaying B_s^0 meson has spin zero, while both the J/ψ and the ϕ are spin-one mesons. Thus, the final state is a combination of three possible orbital angular momentum configurations. Consequently, any CP related measurements require observations of the behaviour of the system in time to ascertain the time-evolution of the CP eigenstate mixture.

The time-dependent amplitude for mixing and decay can be obtained by combining the mixing amplitudes from Equation 2.37 with the decay amplitudes of each possible decay. For a system produced as B_s^0 (\bar{B}_s^0) and decaying into the final state $|f\rangle$ ($|\bar{f}\rangle$), the notation of labelling the decay amplitude as A_f ($\bar{A}_{\bar{f}}$) is adopted. In the case of the $B_s^0 \rightarrow J/\psi \phi$ decay, both B_s^0 and \bar{B}_s^0 can decay into either $|f\rangle$ or $|\bar{f}\rangle$ as the particle content is $K^\pm K^\mp \mu^\pm \mu^\mp$. Thus, in terms of the

¹¹refer to Section A.2.4

initial conditions, there are four possibilities to consider:

$$\begin{aligned} A(B_s^0 \rightarrow f) &\sim g_+ A_f + \frac{q}{p} g_- \bar{A}_f & A(B_s^0 \rightarrow \bar{f}) &\sim \frac{q}{p} (g_- \bar{A}_f + \frac{p}{q} g_+ A_f) \\ A(\bar{B}_s^0 \rightarrow f) &\sim g_+ A_f + \frac{p}{q} g_- \bar{A}_f & A(\bar{B}_s^0 \rightarrow \bar{f}) &\sim g_+ \bar{A}_f + \frac{p}{q} g_- A_f \end{aligned} \quad (2.46)$$

where the conditional amplitude of a system produced as B_s^0 , then decaying at time t as B_s^0 with an amplitude A_f into state $|f\rangle$ is taken to be $\langle B_s^0 | B(t) \rangle A_f \sim g_+(t) A_f$. It can be seen that the structure of the individual amplitudes is very similar and that, for example, the amplitude for $\bar{B}_s^0 \rightarrow f$ can be obtained from the expression for the amplitude of $B_s^0 \rightarrow f$ by swapping g_+ for g_- and multiplying by a factor of $\frac{p}{q}$. Furthermore, the differential decay rate, $\frac{d\Gamma(B_s^0 \rightarrow f)}{dt}$, can be obtained by squaring the corresponding decay amplitude. Using the following relations derived from Equation A.20:

$$\begin{aligned} |g_\pm|^2 &= \frac{1}{2} e^{-\Gamma_s t} \left[\cosh\left(\frac{1}{2} \Delta\Gamma_s t\right) \pm \cos(\Delta m_s t) \right] \\ g_+^* g_+ &= \frac{1}{2} e^{-\Gamma_s t} \left[-\sinh\left(\frac{1}{2} \Delta\Gamma_s t\right) + i \sin(\Delta m_s t) \right] \end{aligned} \quad (2.47)$$

the differential decay rate takes the form:

$$\begin{aligned} \frac{d\Gamma(B_s^0 \rightarrow f)}{dt} &\sim \frac{1}{2} |A_f|^2 (1 + |\lambda_f|^2) e^{-\Gamma_s t} \left[\cosh\left(\frac{1}{2} \Delta\Gamma_s t\right) + C_f \cos(\Delta m_s t) \right. \\ &\quad \left. + D_f \sinh\left(\frac{1}{2} \Delta\Gamma_s t\right) + S_f \sin(\Delta m_s t) \right] \end{aligned} \quad (2.48)$$

where λ_f is defined as in Equation 2.43, and the parameters C_f , D_f and S_f are:

$$C_f = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \quad D_f = -\frac{2 \operatorname{Re}(\lambda_f)}{1 + |\lambda_f|^2} \quad S_f = \frac{2 \operatorname{Im}(\lambda_f)}{1 + |\lambda_f|^2} \quad (2.49)$$

2.4.4 Angular distributions of pseudoscalar-to-vector-vector (P2VV) meson decays

The final state of the $B_s^0 \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)(\phi \rightarrow K^+ K^-)$ decay comprises four particles, all of which are assumed to be on-shell¹². In particular, the kaons and muons are selected with momenta such that in the frame of reference of the LHCb detector they are observed to traverse it fully without decaying further. Consequently, due to energy-momentum conservation and all final state particles

¹²Section A.3

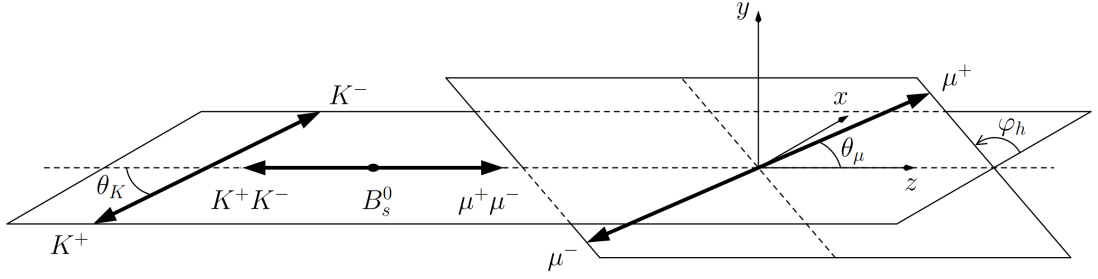


Figure 2.7 *definitions of the $B_s^0 \rightarrow J/\psi \phi$ decay angles in the helicity basis: φ_h – the angle between the $\mu^+\mu^-$ and K^+K^- decay planes; θ_K – the angle between the momentum of the K^+ and the opposite direction of the B_s^0 momentum in the K^+K^- centre-of-mass frame; and θ_μ – the angle between the momentum of the μ^+ and the opposite direction of the B_s^0 momentum in the $\mu^+\mu^-$ centre-of-mass frame decay angles in the helicity basis [15].*

being on-shell, the kinematics are constrained from a maximum of sixteen degrees of freedom down to only twelve, which can fully describe them.

Additionally, B_s^0 has zero spin and negative parity making it a pseudoscalar meson. While the J/ψ and the ϕ have spin one, *i.e.*, vector mesons. The decay $B_s^0 \rightarrow J/\psi \phi$ is thus referred to as a pseudoscalar-to-vector-vector (P2VV) decay. Here the important point is that this means the decay amplitude of the B_s^0 cannot depend on its momentum or the orientations of the J/ψ and ϕ momenta (Section A.5). Thus, only five degrees of freedom remain. Finally, treating the J/ψ as an on-shell particle removes another degree of freedom by assuming its mass is a constant. Thus, the differential decay rate is a function of the four remaining variables:

$$d\Gamma \sim \sum_i |A_i|^2 dm_{J/\psi K^+}^2 dm_{K^+K^-}^2 d\cos\theta_\mu d\varphi_h \quad (2.50)$$

where $m_{J/\psi K^+}$ and $m_{K^+K^-}$ are the invariant masses of the $J/\psi K^+$ and K^+K^- systems, respectively. The angles θ_μ and φ_h are the helicity angles defined in Figure 2.7 (refer to Section A.5). To fully conform to the helicity formalism to simplify the expressions needed to accommodate for the angular dependence, a further change of variables is applied from $m_{J/\psi K^+}^2$ and $m_{K^+K^-}^2$ to $m_{K^+K^-}$ and $\cos\theta_K$. This introduces a Jacobian as expected, but given the fact all particles involved are on-shell, kinematic relations between invariant mass and momenta can be exploited to arrive at an expression of the form:

$$d\Gamma \sim \sum_i |A_i|^2 |\mathbf{p}_{J/\psi}| |\mathbf{p}_{K^+}| dm_{K^+K^-} d\cos\theta_K d\cos\theta_\mu d\varphi_h \quad (2.51)$$

where $\mathbf{p}_{J/\psi}$ and \mathbf{p}_{K^+} are the three momenta of the J/ψ and K^+ respectively, while $\cos\theta_K$ is the third helicity angle defined in Figure 2.7. However, these momenta are indeed dependent on the K^+K^- system invariant mass through said kinematic relations but do not depend on any of the helicity angles. As such, these can be absorbed into $|A_i(m_{K^+K^-})|^2$. Namely, the $m_{K^+K^-}$ dependent part of the amplitude. As already mentioned, the $B_s^0 \rightarrow J/\psi \phi$ decay is a P2VV transition,

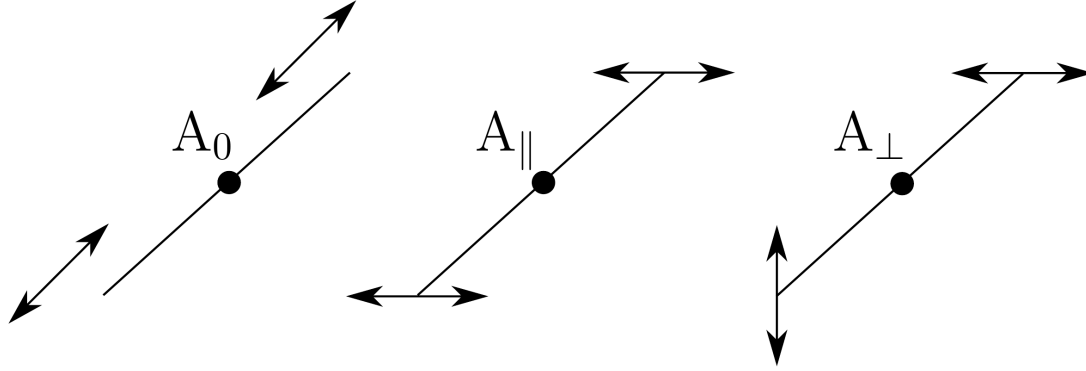


Figure 2.8 *Relative spin polarisations of the $B_s^0 \rightarrow J/\psi \phi$ intermediate states and their corresponding amplitudes: 0 or longitudinal polarisation – A_0 ; \parallel or parallel polarisation – A_{\parallel} ; \perp or perpendicular polarisation – A_{\perp} [54].*

which means there are three possible intermediate angular-momentum states. In terms of spin-polarisation, one state is when the polarisation vectors of both the J/ψ and ϕ are parallel to the helicity axis (called the longitudinal polarisation). Another state is when the polarisation vectors are perpendicular to the helicity axis but parallel to each other (called parallel polarisation). One more state exists where the polarisation vectors are perpendicular to the helicity axis and to each other (called perpendicular polarisation). These are shown in Figure 2.8 with their shorthand labels: 0, \parallel and \perp . The individual amplitudes associated with each state are then A_0 , A_{\parallel} and A_{\perp} .

Consequently, the $J/\psi \phi$ system is not a pure CP state but a mixture of CP eigenstates, and its behaviour under a CP transformation depends on the overall orbital angular momentum configuration (Section 2.4.3). Both the J/ψ and $\phi(1020)$ are CP -odd. That means that the overall CP state is determined by the parity of the orbital angular momentum state. There are two CP -even states: the longitudinal S-wave ($L = 0$) state and the parallel D-wave ($L = 2$) state. The perpendicular state is a P-wave ($L = 1$) and is CP -odd.

Furthermore, there is one more state: an non-resonant S-wave state of the K^+K^-

system. It shall be labelled by the letter “S” with amplitude A_S [55]. Since the K^+K^- system has $L = 0$ in that state, the $J/\psi K^+K^-$ system needs to compensate by being in a P-wave arrangement. The underlying K^+K^- system is also odd under CP in this case.

Looking back at Equation 2.46, it is possible to derive expressions for the B_s^0 decaying into each final state through each of the possible intermediate states described. For example, the $B_s^0 \rightarrow f$ amplitude may look like so:

$$A(B_s^0 \rightarrow f) \sim g_+ A_f + \frac{q}{p} g_- \bar{A}_f = g_+ \sum_i A_i + \frac{q}{p} g_- \sum_j \bar{A}_j \quad (2.52)$$

where A_i and \bar{A}_j are the amplitudes for the B_s^0 and \bar{B}_s^0 decaying through any intermediate state $i, j \in (0, \parallel, \perp, S)$. Now let:

$$A_i = a_i \times r_i(\theta_K, \theta_\mu, \varphi_h) \times q_i(m_{K^+K^-}) \quad (2.53)$$

the amplitude of each intermediate state will consist of a complex number, a_i , the angular dependence, r_i , and the dependence on the K^+K^- system invariant mass, q_i . Given that in the case of $B_s^0 \rightarrow J/\psi \phi$, the B_s^0 and \bar{B}_s^0 both go through the same intermediate state and the fact that each intermediate state has the equivalent dependences, respectively, it is easy to see that the total amplitudes will have the form:

$$A(B_s^0 \rightarrow f) = \sum_i (g_+ + \lambda_f^i g_-) A_i \quad A(\bar{B}_s^0 \rightarrow f) = \sum_i \frac{p}{q} (g_- + \lambda_f^i g_+) A_i \quad (2.54)$$

where $\lambda_f^i = \frac{q}{p} \frac{\bar{A}_i}{A_i}$ as per previous familiar expression, but individually here for every CP eigenstate. In fact, in this case, λ_f^i is also equivalently specified in terms of $\lambda_f^i = \eta_i |\lambda_f^i| e^{-i\phi_s^i}$ as the intermediate states are eigenstates of CP . The individual ϕ_s^i phase is given by $\phi_s^i = -\arg(\frac{1}{\eta_i} \lambda_f^i)$. Therefore, there may exist CP violation dependence according to each intermediate state. However, if CP violation occurs in the same amount in all such states, then $\phi_s^i \rightarrow \phi_s$.

Finally, the total angular time-dependent differential decay rate can be constructed by the same exercise from Section 2.4.3, except this time it contains each intermediate state and their interference [56]:

$$\frac{d^4\Gamma(t)}{dm_{K^+K^-}^2 d\cos\theta_K d\cos\theta_\mu d\varphi_h} = \sum_k^{10} N_k C_{SP}^k h_k(t) f_k(\theta_K, \theta_\mu, \varphi_h) \quad (2.55)$$

where $N_k = |A_i \times A_j|$ for all possible combinations of $i, j \in (0, \parallel, \perp, S)$, $h_k(t)$ are time-dependent mixing terms of the form found in Equation 2.48, and f_k are angular dependence terms similar to those in Equation 2.53, except here the dependence on $m(K^+K^-)$ is factorised in the C_{SP}^k terms. The exact definition of all the terms is discussed in detail in Chapter 7. The terms of Equation 2.55 above are fully defined in Table 7.1, while the C_{SP}^k are defined in Section 7.7.3.

2.5 ϕ_s measurement and its implications

New physics (NP) beyond the current Standard Model theory can potentially contribute to significant deviations on the measurements of ϕ_s from its SM predicted value.

2.5.1 New physics in B_s^0 mixing

Due to the nature of the $B_s^0 \rightarrow J/\psi \phi$ decay and the already discussed properties in this chapter, one promising place where new physics effects might have an impact is towards modifying the $B_s^0 - \bar{B}_s^0$ mixing phase discussed in Equation 2.44, φ_M [57]. These effects can manifest themselves through new particles contributing to the $B_s^0 - \bar{B}_s^0$ mixing diagram in Figure 2.5. Of course, there is still the possibility of NP presenting itself in the penguin decay diagram from Figure 2.6. However, such contributions are limited by the fact that the strong force has not yet been observed to violate CP symmetry as discussed in Section 2.2.

That leaves the search for NP to be concentrated at the mixing phase, φ_M . The $B_s^0 - \bar{B}_s^0$ mixing can be described by the matrix:

$$\mathbf{H}_{\text{mix}}^s = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} \equiv \begin{pmatrix} M_{11}^s & M_{12}^s \\ M_{12}^{s,*} & M_{22}^s \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma_{11}^s & \Gamma_{12}^s \\ \Gamma_{12}^{s,*} & \Gamma_{22}^s \end{pmatrix} \quad (2.56)$$

as defined in Equation A.9, where the mixing phase is defined as $\varphi_M \equiv \arg(M_{12}^s)$. To include possible modifications in a model independent way, one can include them as a complex scale factor into M_{12}^s :

$$M_{12}^s \rightarrow M_{12}^{SM,s} \cdot \Delta_s = |\Delta_s| M_{12}^{SM,s} e^{-i\varphi_{\Delta_s}} \quad (2.57)$$

where Δ_s is the complex scale factor related to the NP effects. The definitions

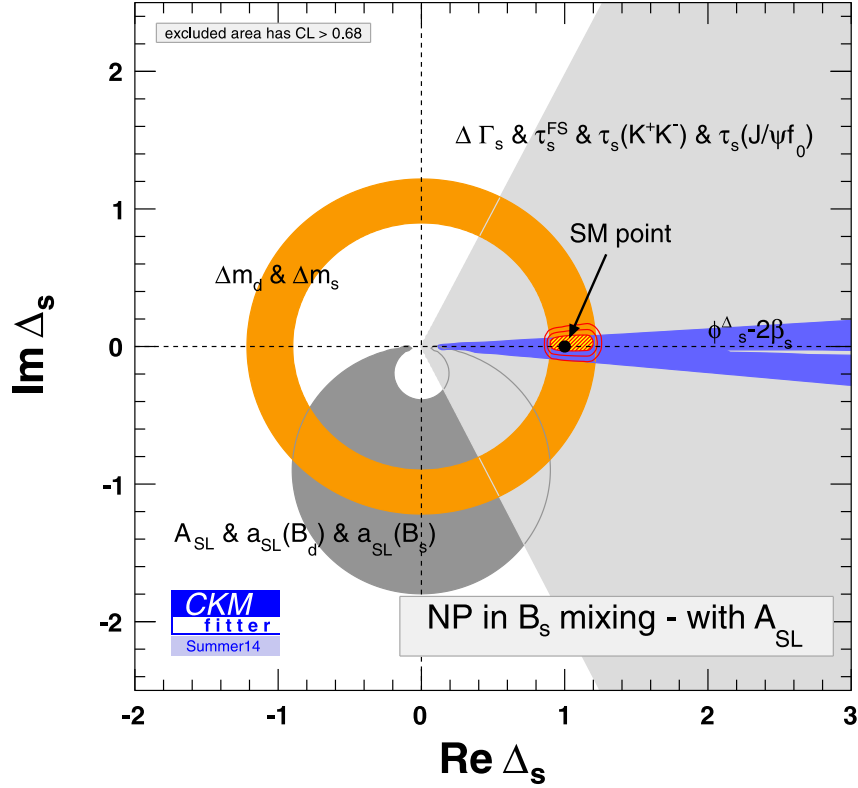


Figure 2.9 Constraints on Δ_s as calculated by CKMFitter group [58]. The coloured areas represent regions with $1 - p < 68.3\%$ for the constraints as labelled. The blue region is from the ϕ_s constraint and the orange is from the combined Δm_d and Δm_s constraint. The solid red lines around the SM point represent the boundaries for the 68%, 95% and 99% confidence intervals, respectively.

are equivalent for the $B_d^0 - \bar{B}_d^0$ mixing as well.

The physical observables $|\Gamma_{12}^q|$, $|M_{12}^q|$ and φ_M^q can be determined from the corresponding mass difference $\Delta M_q \approx 2|M_{12}^q|$ of the $B_{L,H}$ states, their decay width difference $\Delta\Gamma_q \approx 2|\Gamma_{12}^q| \cos \varphi_q$ and the semileptonic CP asymmetry, defined as:

$$a_{\text{SL}}^q \equiv \text{Im} \frac{\Gamma_{12}^q}{M_{12}^q} = \frac{|\Gamma_{12}^q|}{|M_{12}^q|} \sin \varphi_q = \frac{\Delta\Gamma_q}{\Delta M_q} \tan \varphi_q \quad (2.58)$$

where $q \in (d, s)$ labels the respective quark flavour [58].

Both $\Delta_{s,d}$ can vary significantly from their Standard Model prediction of $\Delta_{s,d} = 1$. Importantly, the NP phases $\varphi_{\Delta_{s,d}}$ do not only affect $a_{\text{SL}}^{s,d}$, but also shift the CP phases extracted from the mixing-induced CP asymmetries in $B_d^0 \rightarrow J/\psi K^{*0}$ and

$B_s^0 \rightarrow J/\psi \phi$ to $2\beta + \varphi_{\Delta_d}$ and $2\beta - \varphi_{\Delta_s}$, respectively [58]. Figure 2.9 shows the constraints on the parameter Δ_s as calculated by the CKMFitter group [51, 58] including a parameter A_{SL} measuring a linear combination of a_{SL}^d and a_{SL}^s . It can be seen that the contribution from NP to $B_s^0 - \bar{B}_s^0$ mixing must be quite small to remain compatible with the previous measurements of CP violation in this process.

Chapter 3

The LHCb Experiment

The LHCb Experiment is dedicated to flavour physics and focuses on precision measurements of hadrons, heavy quarks and CP violation. It is capable of cutting edge research such as the recent single-experiment $B_s^0 \rightarrow \mu^+ \mu^-$ [59] observation. Other prominent analyses include lepton universality tests such as $R(D^*)$ [60] or $R(K^*)$ [61]. Also, one of its design goals is the precise measurement of the CP -violating phase ϕ_s using $B_s^0 \rightarrow J/\psi \phi$ decays ([5, 14, 15] and the one discussed in Chapter 6). The latter was performed with data collected by the LHCb experiment at the Large Hadron Collider during the 2015 and 2016 data-taking periods. Moreover, the two projects in Chapters 4 and 5 directly deal with components of the experiment itself.

Thus, this chapter will overview the relevant details of the Large Hadron Collider and the LHCb experiment.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) Machine [62] is currently the world's largest and most powerful particle collider. It is a 27 km long accelerator ring located 100 m underground near the French-Swiss border around Geneva. The LHC is a proton-proton (pp) and heavy ion collider with a design luminosity of up to $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a centre-of-mass energy $\sqrt{s} = 14 \text{ TeV}$. During 2011, 2012, 2015-2018 it operated at $\sqrt{s} = 7 \text{ TeV}$, $\sqrt{s} = 8 \text{ TeV}$, $\sqrt{s} =$ and 13 TeV respectively. It is part of the European Organisation for Nuclear Research

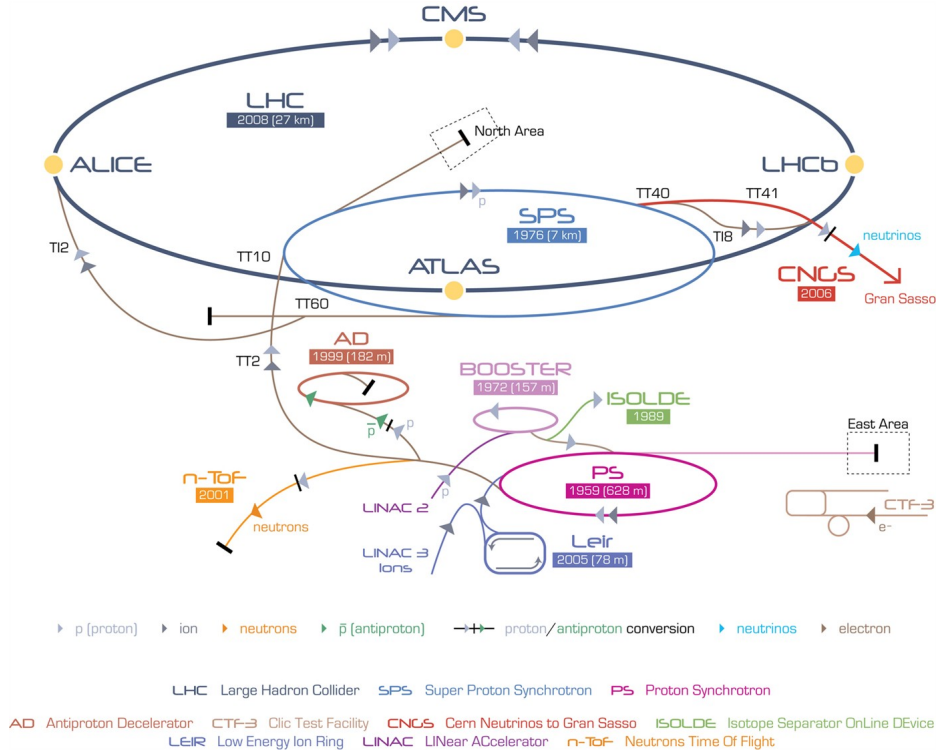


Figure 3.1 *Schematic diagram of CERN's Accelerator complex. Shown are the injector chain of the LHC and the experiments' locations at the corresponding interaction points [67].*

(CERN) and provides four beam crossing (interaction) points, occupied by the four large particle physics experiments – ALICE [63], ATLAS [64], CMS [65] and LHCb [66].

The ATLAS and CMS experiments are so-called $4\pi^1$ general-purpose particle physics detectors. Each carries design choices that grant them slight advantages or disadvantages in specific physics scenarios; however their fields of interest and operation overlap. They are notably responsible for the Higgs boson discovery [68]. Their programmes include top quark, t , physics, electroweak and searches of supersymmetric and other exotic particles.

The ALICE experiment focuses on the physics of quark-gluon plasma and thus takes most of its data during LHC's heavy-ion runs – typically only a few weeks per year. LHCb physics programme aims to search for CP violation sources beyond the Standard Model (BSM) and evidence for other new physics phenomena in rare decays.

¹ 4π or hermetic detector is a particle detector designed to cover as large a solid angle as possible around the interaction point.

The LHC machine relies on a large and complex injector chain, as illustrated in Figure 3.1. It starts with the proton source – hydrogen gas which is ionised by a breakdown electric field. Then the protons are sent to the LINAC2 linear accelerator, where they are accelerated to 50 MeV. This beam of protons then enters the chain of successively larger and more energetic accelerators – 1.4 GeV Proton Synchrotron Booster, 28 GeV Proton Synchrotron (PS) and the 450 GeV Super Proton Synchrotron (SPS). The beam that is finally injected into the LHC is comprised of bunches (1.2×10^{11} protons each) spaced at least 25 ns apart. The PS structures the protons in bunches with the maximum number of bunches possible of 2808, accounting for the gaps due to kicker magnet rise times (across the injector chain) and the beam dump system. The LHC has eight bending sections [69]. Each is equipped with superconducting dipole magnets capable of fields up to 8.33 T. Quadrupole, sextupole and octupole magnets are used to collimate and focus the beam. There are four crossing points where the collisions happen and thus where the large experiments are placed. The crossing point

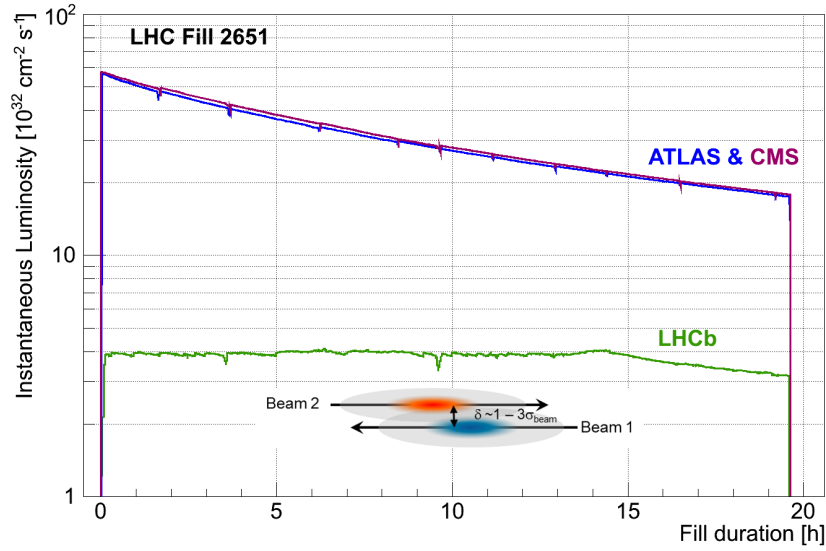


Figure 3.2 *Instantaneous luminosity over time at ATLAS (blue), CMS (red) and LHCb (green) during a typical fill in Run 1 [70].*

for LHCb is unique as the delivered instantaneous luminosity is kept constant through a process called “luminosity levelling” [71]. The proximity of the beams is adjusted opposite to the beam’s intensity to achieve a target luminosity at all times (Figure 3.2). This is done to keep a small number of collisions per event – simplifying reconstruction and thus improving precision.

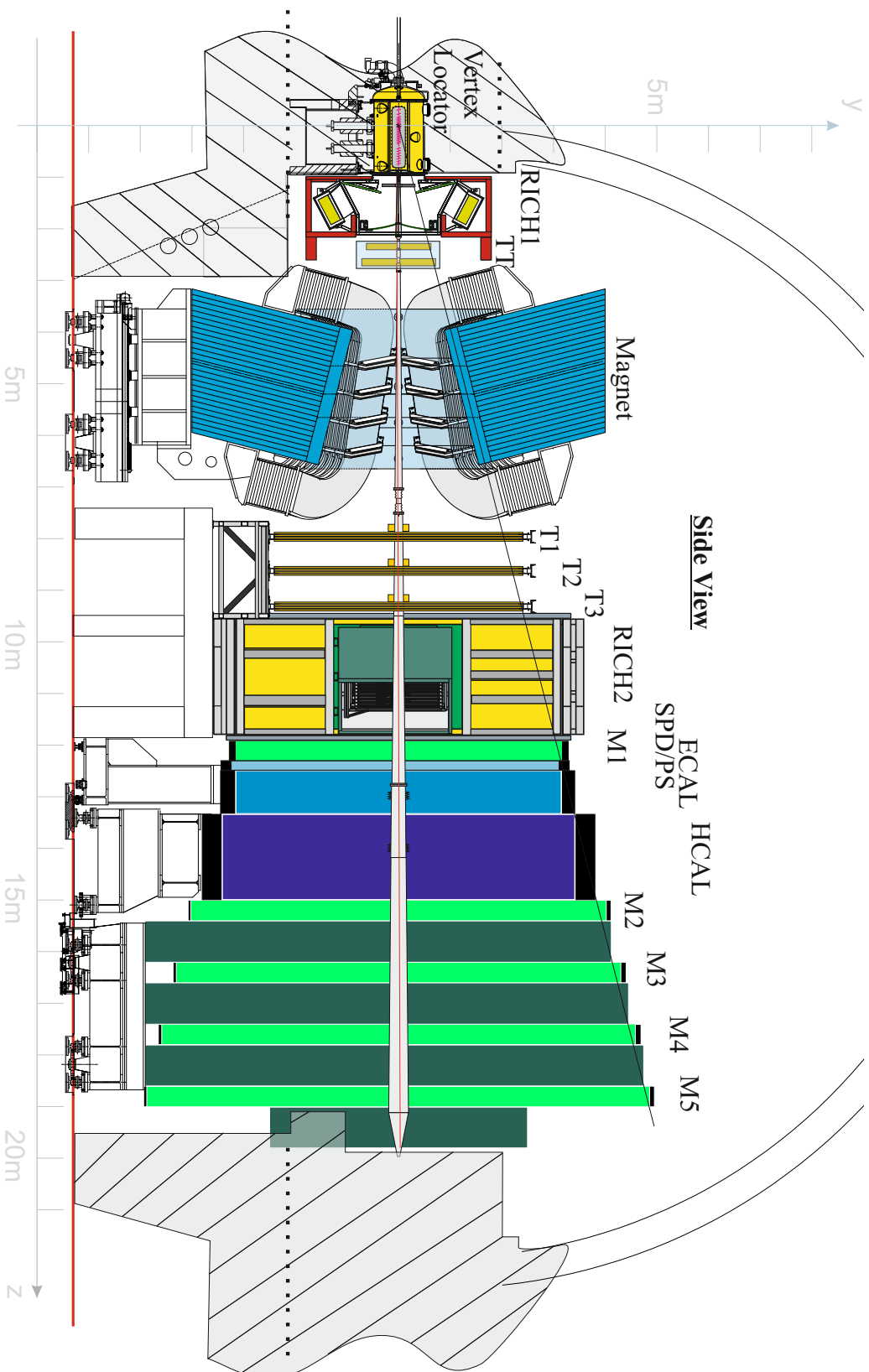


Figure 3.3 *The LHCb detector – side view in the y - z (non-bending) plane. [72].*

3.2 LHCb Overview

The Large Hadron Collider Beauty (LHCb) experiment is aimed at precision measurements of hadrons, specifically those containing the bottom or charm quarks. The name comes from the bottom (beauty) quark. Moreover, it is of unique design among the four big experiments at the LHC. It is not a full cylinder around the beam pipe but instead extends in the forward direction only. LHCb’s geometry is designed to take advantage of the angular distributions of $b\bar{b}$ production at the LHC, as shown in Figure 3.4. Since the $b\bar{b}$ pairs are mostly produced in the forward/backward regions of the collision (high pseudorapidity), the detector is constructed to be a single-arm forward spectrometer and concentrates its apparatus where it matters most. Furthermore, the LHC provides the largest $b\bar{b}$ production cross-section of any previous collider: $\sigma(p\bar{p} \rightarrow b\bar{b}) = 144 \pm 1 \pm 21 \mu\text{b}$ at $\sqrt{s} = 13 \text{ TeV}$ [73], six orders of magnitude higher than that of the B -factories, Belle [74] and BaBar [75]. Just during 2016 data-taking LHCb managed to accept $\mathcal{O}(10^9)$ b hadron events, and around $\mathcal{O}(10^8)$ of those were decays containing $\mu^+\mu^-$ pairs (dimuon) similar to $B_s^0 \rightarrow J/\psi \phi$. However, as mentioned in Section 3.1, events at a hadronic collider are much “messier” than those observed at previous electron-positron experiments – particle multiplicity is significantly higher and large pileup is a concern [76, 77].

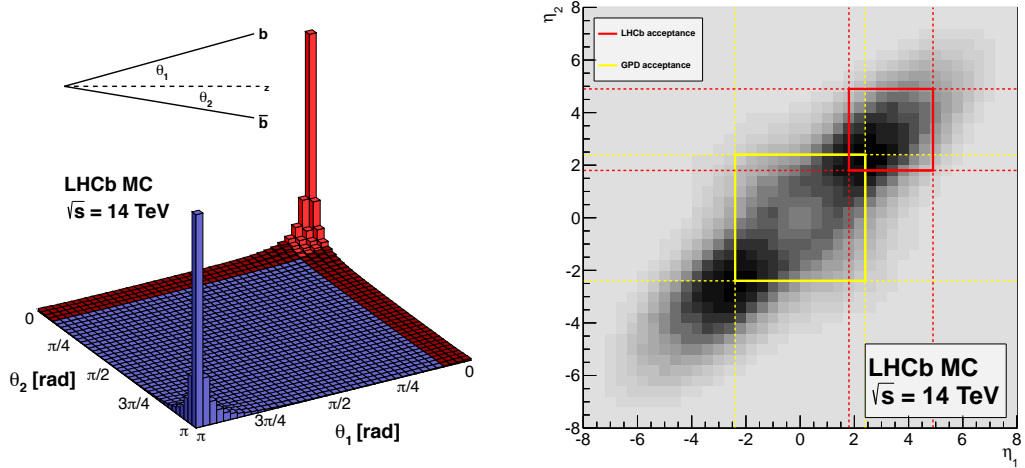


Figure 3.4 *Left: Distribution of simulated b quark production in polar angles. LHCb acceptance is marked by the red highlighted region. Right: Distribution of simulated b quark production in pseudorapidity. LHCb acceptance is marked by the red square and compared to general-purpose detectors (ATLAS, CMS) in the yellow square [78].*

To keep the detector occupancy from exceeding what the experiment can handle, the number of collisions per bunch crossing is limited to approximately 1.5 or equivalently to instantaneous luminosity of $\mathcal{L} = 4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. Figure 3.2 shows the time progression of the luminosity in ATLAS, CMS and LHCb during a typical LHC fill. LHCb luminosity stays constant for about 14 hours from the start of a fill. After this point, the beam intensity is low enough that zero separation cannot overshoot the desired luminosity threshold. Then it falls together with the other experiments [79]. This is due to the general LHC luminosity levelling performed by adjusting the beam offset to keep the luminosity constant for as long as possible and thus ensure maximal integrated luminosity. LHCb luminosity levelling makes it possible to reconstruct collision events more efficiently and allows for improved precision.

3.2.1 Coordinate system

LHCb uses a right-handed Cartesian coordinate system. The interaction point is chosen as the origin. The z -axis is set along the beam pipe (beam 1) and points towards the end of the detector. The y -axis is extended vertically upwards, while the x -axis lays horizontally aimed towards the centre of the LHC.

3.3 Tracking and Vertexing

The purpose of the tracking system is to reconstruct the paths charged particles took after a collision (tracks) by looking at the locations these particles interacted with the detectors (hits). The system is comprised of a vertex locator (VELO), a dipole bending magnet and four tracking stations (one before the magnet and three after it).

3.3.1 Vertex Locator (VELO)

The first subdetector that particles go through is the VERtex LOcator (VELO). It is responsible for determining the interaction vertex² with fine spatial granularity [80]. The required granularity comes from a particular feature of

²location of proton-proton collisions, also-called the primary vertex

b hadron decays where the decay vertex³ is displaced from the production vertex on the order of $\mathcal{O}(1\text{ cm})$. Therefore, it is crucial to have good vertex resolution to distinguish secondary vertices and reduce background. Also, given that B_s^0 - \bar{B}_s^0 mixing frequency is fast, any time-dependent measurements require time resolution on the order of $\mathcal{O}(50\text{ fs})$.

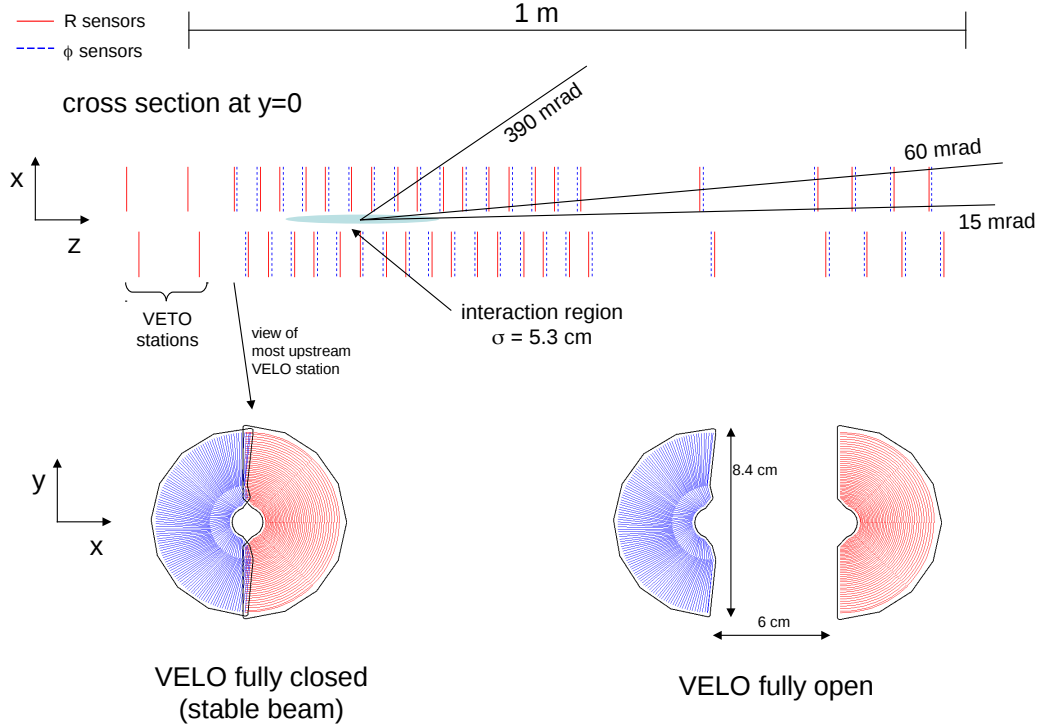


Figure 3.5 Schematic diagram of the LHCb VELO station layout, showing the open and close position, r and ϕ modules and acceptance. [80].

At LHCb, the VELO is equipped with several defining features to provide the needed performance. It is a movable silicon-strip detector that operates in a secondary vacuum vessel closely around the beam pipe. On the side facing the beam, it is separated from the LHC primary vacuum by a thin layer (0.3 mm in thickness) of corrugated aluminium alloy, called RF-foil (Figure 3.6). The VELO is designed to be in two parts (halves), each on either side of the beam pipe. LHCb is quite close to one of the injection sites for the LHC. Thus, the LHC beam is defocused and spatially spread out during beam injection. At those times, the VELO halves separate (open) and retract away to protect the modules from the increased radiation levels. When the beam is refocused again and no

³secondary vertex

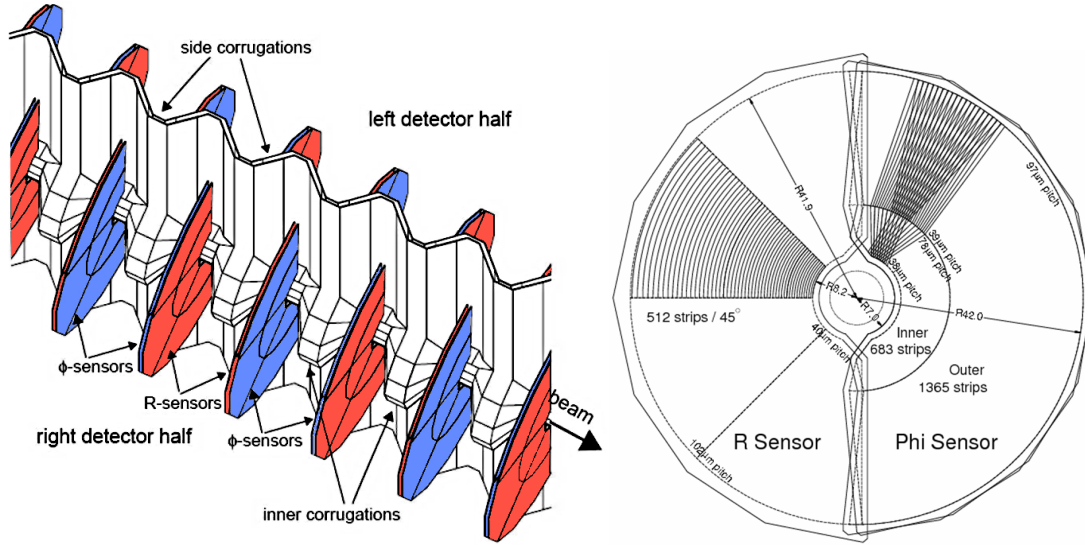


Figure 3.6 *Schematic diagram of the LHCb VELO. Left: VELO sensors and RF foil in the closed position. Right: Detail of r and ϕ sensors, showing the orientation of silicon microstrips. The diagram has two successive ϕ sensors overlaid to showcase the stereo angle design. [80].*

longer occupies such a wide area, the VELO closes again to achieve a minimum distance to the beam of 8 mm.

The detector consists of 42 detector modules split into 21 stations, 6 of which are located on the opposite side of the interaction vertex in relation to the main body of LHCb. The modules lie in the perpendicular plane to the z -axis. The modules perform three functions: hold the sensors in position, connect the electrical readouts to the sensors and enable cooling while in a vacuum. Each of the detector modules consists of two semi-circular silicon-strip sensors (Figure 3.6). Sensors come in two varieties: an r sensor, measuring the radial displacement from the beam of detector hits, and a ϕ sensor, measuring the azimuthal angle of hits. The r sensor modules are made up of concentric silicon-strip rings with variable pitch, decreasing from the outer part at $102\text{ }\mu\text{m}$ towards the centre at $38\text{ }\mu\text{m}$. This is done to follow the decreasing occupancy due to increasing track angle for particles leaving the interaction region at a wider aperture. Due to the bonding requirement of the r sensor pads, the sensor is slightly larger than the ϕ sensor. However, the total sensitive area is still the same between the two types. The ϕ sensor modules are wedge-shaped and divided further into two regions according to the silicon strip distribution. The inner region extends from the centre of the ϕ module up to a radius $r = 17.25\text{ mm}$ with a pitch starting at $38\text{ }\mu\text{m}$ up to $78\text{ }\mu\text{m}$. The outside region pitch begins at $39\text{ }\mu\text{m}$ up to $97\text{ }\mu\text{m}$ at the edge. This is done

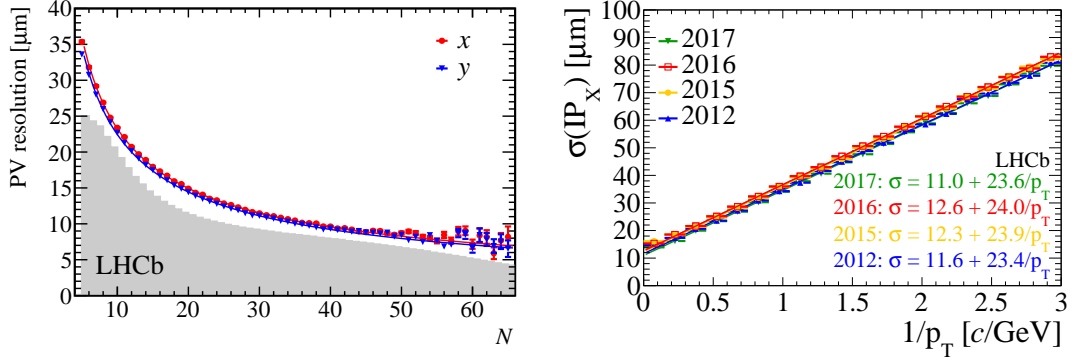


Figure 3.7 *Left: The primary vertex resolution, for events with one reconstructed primary vertex, as a function of track multiplicity. The superimposed histogram shows the distribution of number of tracks per reconstructed primary vertex for all events that pass the high level trigger. Right: The resolution of impact parameter in the x -direction as a function of the inverse transverse momentum. [4, 70].*

to distribute the occupancy and eliminate the need for huge pitch at the edges. Furthermore, to reduce reconstruction ambiguities arising from a fully symmetric coordinate system, the inner region strips are skewed from the outward radial direction by 20° in the inner region and by 10° in the outer region.

3.3.2 Vertexing Performance

The VELO performance can be expressed as its ability to accurately and precisely determine primary vertices and correlate tracks to them. To quantify this abstract concept, LHCb uses performance metrics such as the spatial resolution of vertices and the impact parameter of tracks.

The primary vertex resolution in data is determined by performing two independent measurements of its position and comparing the difference. More precisely, recorded tracks from the same events associated with the same vertex are randomly split into two sets, and each set is used to reconstruct its position. Then the difference in position of the two resulting vertices is taken. Finally, the width of the distribution of vertex position differences is $\sqrt{2}$ times the resolution⁴. The resolution strongly depends on the number of tracks recorded in the event as shown in Figure 3.7 (left).

The impact parameter (IP) is defined as the shortest perpendicular distance

⁴apparatus resolution, $\varepsilon_{\text{resolution}}$, and standard deviation, σ , relationship through degrees of freedom, N : $\varepsilon_{\text{resolution}} = \sigma_{\text{vtx}}/\sqrt{N}$

between the path of a particle and the primary vertex. The IP resolution is due to three main effects: multiple scattering of particles by the detector material, the spatial resolution of hits in the detector for reconstructed tracks and the distance required to extrapolate a track from the PV to its first hit in the detector. Usually, when quoting IP resolution, it is easier to talk about the resolution of the x and y projections of the IP, namely IP_x IP_y , which are the projections of IP onto the x - z and y - z planes. Their resolution in terms of inverse transverse momentum ($1/p_T$) is shown in Figure 3.7 (right). This dependence owes itself to multiple scattering in the detector material and the VELO's spatial resolution. At high p_T , typical of B meson decays, the resolutions are asymptotic, tending towards $\sim 12\text{ }\mu\text{m}$ and generally in the 13–20 μm range.

3.3.3 Magnet

Knowing the momentum of particles is an essential piece of the puzzle when talking about collision event reconstruction. More so, on LHCb, the individual momentum of charged particles helps uniquely identify them (Section 3.4). All of this, of course, requires a high precision measurement.

The method to achieve it is based on the Lorentz force:

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (3.1)$$

where \mathbf{p} is the particle's momentum, q is the charge, \mathbf{E} and \mathbf{B} are the electric and magnetic fields, and \mathbf{v} is the particle's velocity. Using only a magnetic field and no electric field, one can induce a change of momentum perpendicular to the velocity and magnetic field. Moreover, the particle's charge sign will determine the exact direction the deflection vector will be pointing: $+|q|(\mathbf{v} \times \mathbf{B})$ or $-|q|(\mathbf{v} \times \mathbf{B})$. Generally, given a spatially unbounded uniform magnetic field, any charged particle traversing it will execute circular motion around a point:

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{v} \times \mathbf{B}) = \frac{m_0\gamma|\mathbf{v}|^2}{\rho} \left(-\frac{\mathbf{r}}{|\mathbf{r}|} \right) \quad (3.2)$$

where m_0 is the particle's rest mass, γ is the Lorentz factor, ρ is the radius of the motion, and \mathbf{r} is a radius vector pointing outwards from the centre of rotation. For real non-uniform magnetic fields, it is then helpful to define the magnetic rigidity as the required magnetic bending for a given radius relative to

the particle momentum:

$$|\mathbf{B}|\rho = \frac{|\mathbf{p}|}{q} \quad (3.3)$$

Thus, utilising a well-known magnetic field and measuring the deflection particles experience traversing it, one can determine the momentum and the charge of these particles. For this purpose, LHCb employs a warm⁵ dipole bending magnet [81], illustrated in Figure 3.8. It provides the bending magnetic field needed by the tracking system to measure charged particle momentum. Given the length

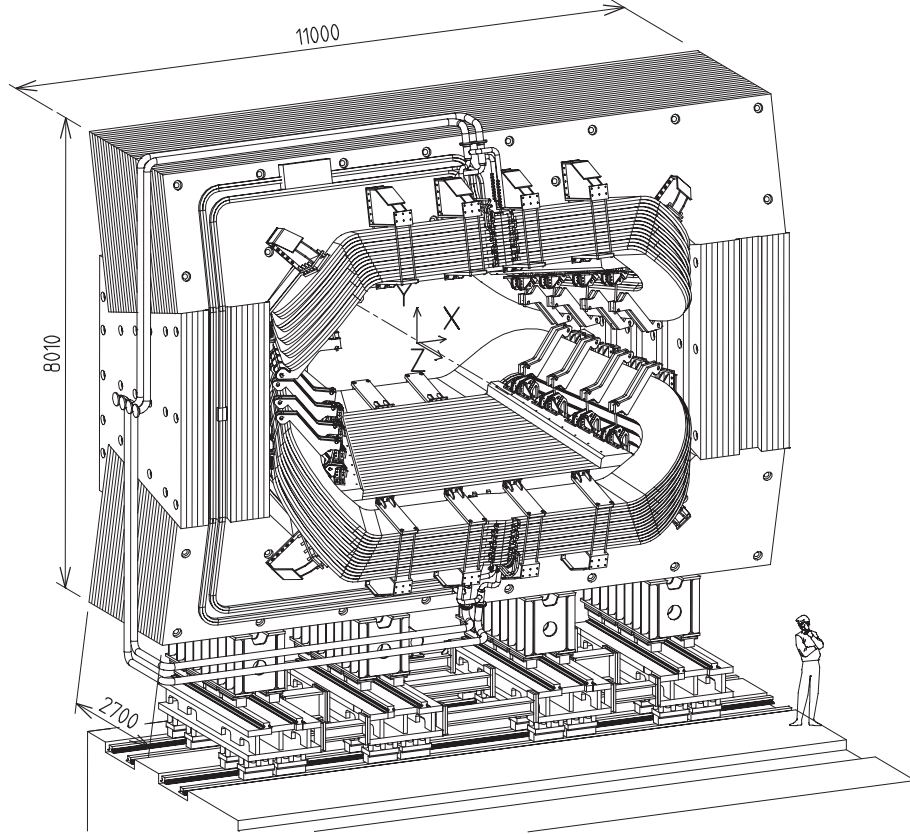


Figure 3.8 *Schematic diagram of the LHCb magnet. [81]. The magnetic field principal component lies along the y-axis.*

restrictions⁶ of the field in the z -direction, the magnetic field strength must be well suited to deflect particles sufficiently and with good spatial granularity for the momentum to be resolved with a satisfactory resolution.

The magnet at LHCb provides an integrated field of $\int \mathbf{B}dl = 4 \text{ Tm}$. The field strength as a function of position z is shown in Figure 3.9. The field spreads between the VELO (Section 3.3.1) and the tracking stations (Section 3.3.4).

⁵not superconducting

⁶Dependent on real-world factors such as physical magnet size, detector geometry and space restrictions

However, since the Hybrid Photon Detectors in the RICH detectors (Section 3.4.3) are affected by magnetic fields, it is tuned to be negligibly small in the region of RICH1. The field was measured to a relative precision of $\mathcal{O}(10^{-4})$ prior to data-taking to ensure it satisfies all the requirements. The magnet itself is comprised of two saddle-shaped coils attached to a two-part rectangular yoke with slanted poles

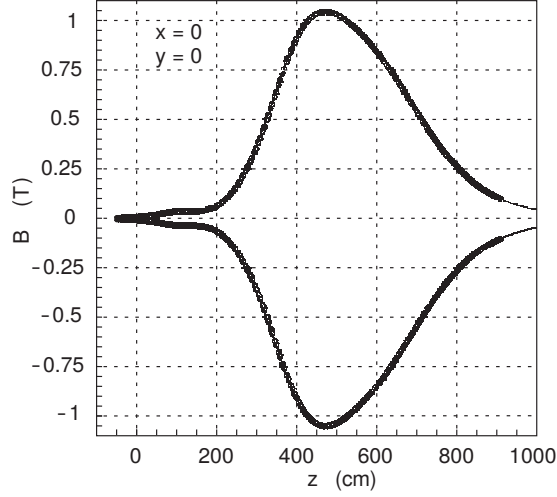


Figure 3.9 *Magnetic field of LHCb’s magnet (Figure 3.8) as a function of z in the plane of $(x, y) = (0, 0)$ for both polarities [81].*

forming a wedge-shaped window through the centre matching the acceptance of the detector. The yoke is composed of twenty-seven layers of laminated low-carbon steel. Each layer is 100 mm thick, giving the yoke a maximum mass of 25 tonnes. The coils are made of fifteen layers of aluminium conductor each. The conductors are hollow and have a central channel used for water-cooling. The current through the coils is $I = 5.85$ kA.

Finally, due to left-right asymmetries of the detector itself, the magnet’s polarity is reversed regularly during the operational year to cancel any of the respective efficiencies introduced. The collected data is thus split evenly between the two polarities.

3.3.4 Silicon and Straw tube tracking stations (TT, IT and OT)

The general tracking information is acquired with the Tracker Turicensis, TT, a silicon strip detector sitting downstream of RICH1 and upstream of the magnet and the three additional tracking stations, T1-T3, immediately downstream of

the magnet. The tracking stations consist of the Inner Tracker (IT [82]), a silicon strip detector located close to the beam pipe with small acceptance, but covering the high occupancy region, and the Outer Tracker (OT [83]), which is a straw tube detector with larger acceptance, but sparser resolution.

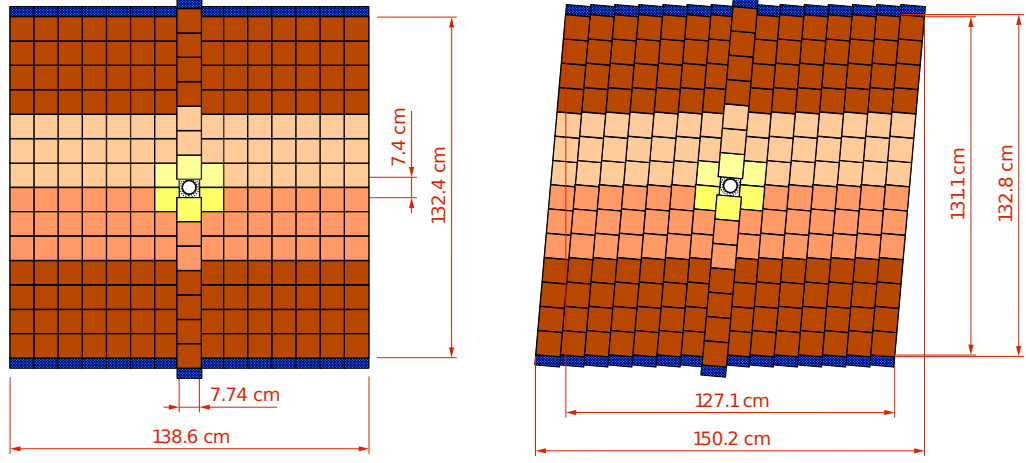


Figure 3.10 *The first two layers ('x' and 'u' types) of the Tracker Turicensis, showing the vertical and $+5^\circ$ rotated orientations [66].*

The Tracker Turicensis (TT) and the Inner Tracker (IT) are collectively referred to as the Silicon Tracker (ST) because they use the same silicon strip sensor technology. The TT is essential for reconstructing tracks originating outside the VELO, such as those beginning at the decay vertex of particles, which are longer-lived than the B, generally K_s^0 or Λ . Each ST station (the TT and the three IT stations) consist of four layers of silicon microstrip sensors with a pitch of 183 nm and 196 nm for the TT and the IT, respectively. The strips in the first and last layers ('x' layers) are oriented vertically, while the middle two layers ('u' and 'v' layers) are rotated by -5° and $+5^\circ$ from the vertical. The TT has a separation of ~ 27 cm between each layer to improve spatial resolution, while the separation for the IT is only ~ 4 cm in each station. The sensors are all held in light-tight, electrically and thermally insulated boxes. The temperature inside is kept below 5° , and nitrogen gas is continuously flushed to prevent condensation.

The TT and IT share their readout, control and monitoring systems. The readout is done by a hybrid connected by a wire-bonded Kapton^{TM7} ribbon cable to the silicon sensor. Each layer of the TT is split in half-modules with seven sensors

⁷the commercial name of a series of polyimide films developed first by DuPont in the late 1960s, with new variants still being produced. The leading valuable quality is that it remains stable across a wide range of temperatures, from -269 to $+400^\circ\text{C}$ [84]

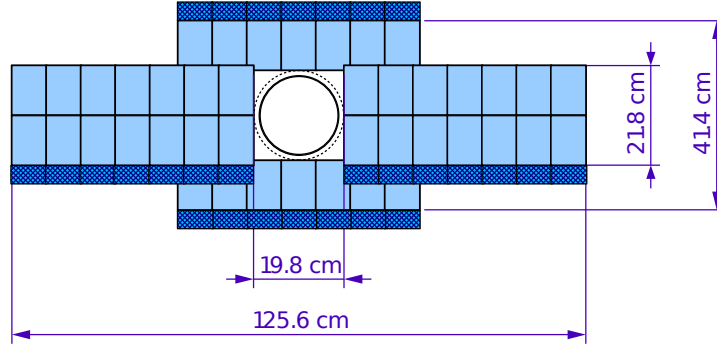


Figure 3.11 *Layout of an ‘x’ detection layer in the second IT station [66].*

each, above and below the beam pipe, consisting of fourteen sensors per column. Depending on typical occupancy, the sensors are read out in groups of different sizes (1, 2, 3 or 4 sensors per group). The IT, covering only the high occupancy region, has at most two sensors sharing a hybrid.

By having alternate modules displaced in the z -axis, adjacent modules of the TT are made to overlap by a few mm to avoid gaps in acceptance. Similarly to the TT, the boxes of the IT are staggered in z and overlap for the same reason. Furthermore, the TT readout, cooling system and structural supports are placed on the end of each module (above or below the TT), remaining outside the detector acceptance. The IT is a bit different in this regard, being the inner part of the T stations, its readout, cooling system and structural supports are inevitably inside the acceptance.

The OT is a drift-chamber detector surrounding the IT in the T stations and has four modules per station. Each of its modules is made of two z -staggered and overlapping layers of straw tubes. The OT covers the lower occupancy region not covered by the IT. Similarly to the TT and IT, the tubes in the modules are arranged as follows: the first and last modules are vertically aligned, while the second and third are rotated -5° and $+5^\circ$ from the vertical, respectively (Figure 3.12).

The straw tubes are gas-tight drift chambers made of electrically conductive material with an internal diameter of 4.9 mm. They are filled with a mixture of 70 % argon and 30 % carbon dioxide gas. The tube is made of two layers: the outer, made from polyimide-aluminium laminate, providing structure, gas-tightness and shielding; the inner, made from carbon-doped polyimide, which serves as the cathode. The anode is a thin gold-plated tungsten rod running

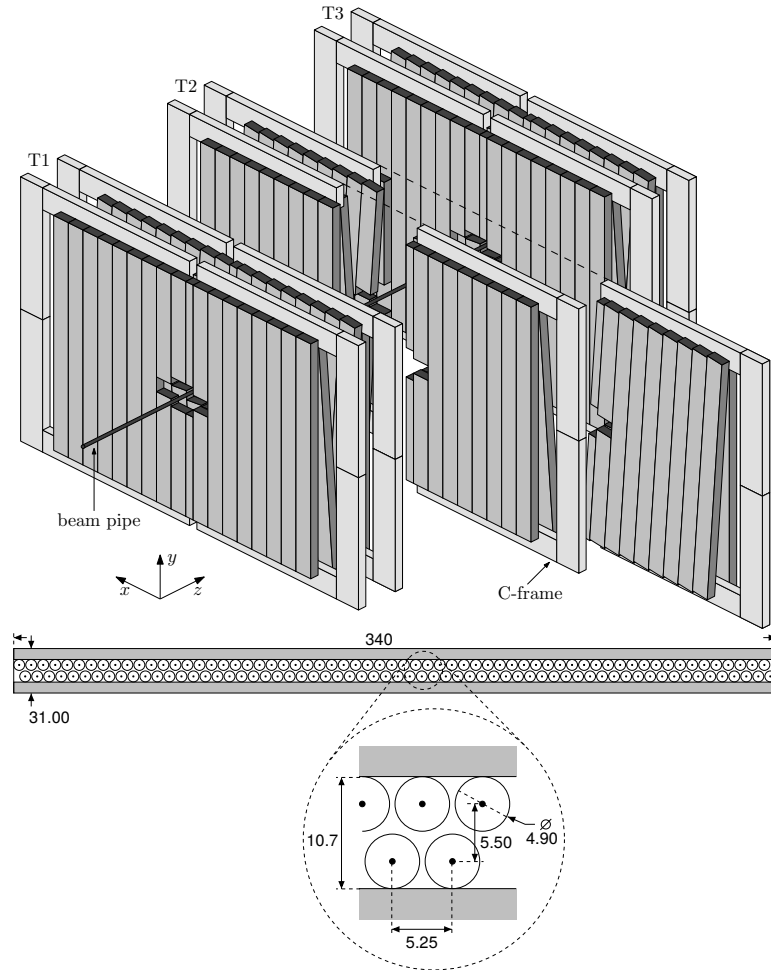


Figure 3.12 *Arrangement of OT straw tube modules in layers and stations with part of T2 retracted showing the orientations of individual tubes (top). Cross-section of a straw tube module (bottom) [83].*

through the middle of each tube.

The tubes provide a maximum drift time of 35 ns limiting the spillover to two bunch crossings. However, given particle time-of-flight through the tubes and signal propagation and readout, a 75 ns window is read out at a time. Figure 3.13 shows the comparison of measured drift times in three different bunch spacing runs.

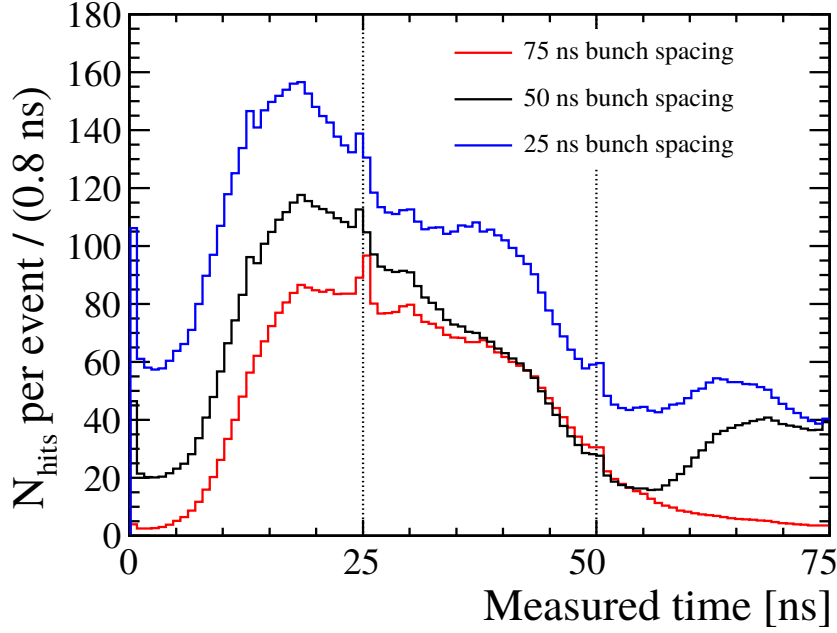


Figure 3.13 *Drift time distribution in module 8, inner most to the beam, for 75, 50, 25 ns bunch crossing spacing. The vertical lines at 64 and 128 TDC counts correspond to 25 and 50 ns, respectively. The distributions correspond to all hits in a 3000-event run for each bunch crossing spacing, recorded with an average number of overlapping events of $\mu = 1.2$ (75 ns), 1.4 (50 ns) and 1.2 (25 ns) [83].*

3.3.5 Tracking reconstruction

A track at LHCb is an abstract object describing the trajectory⁸ of a (charged) particle followed through the detector. Tracks are reconstructed by combining the hits in the VELO, TT and T stations. They are classified according to where the tracks have been observed to pass through (Figure 3.14). For example, a VELO track consists solely of hits within the VELO and does not include any hits from other subdetectors, while an upstream track has to have hits from the VELO and the TT associated with it. The primary physics analysis in this thesis (Chapter 6) deals mainly with tracks of kaons and muons, which are accepted as long tracks only.

Long tracks need to have associated components in all VELO, TT and T stations. Their reconstruction starts by identifying hits associated with straight trajectories inside the VELO. The hits must include at least three of each r and ϕ sensors to

⁸a trajectory is represented as a collection of states (straight line segments), each combined with a covariance matrix

be considered valid VELO tracks. Then two other algorithms, namely “forward tracking” and “track matching”, add complementary hits from the TT and T stations.

Given a VELO track, “forward tracking” tries to combine it with a hit in any T station. Once a suitable candidate is found, an overall trajectory is established. That allows to search and add further hits from all T stations belonging to the same trajectory. Before the algorithm has finished, it checks if the constructed object satisfies specific quality metrics (cuts), and if so, the object is saved as a valid track.

The algorithm of “track matching” takes a set of VELO tracks and a set of track segments reconstructed from hits in the T stations and tries to match elements of each. It extrapolates tracks from each set through the magnet and determines compatibility based on parameters such as trajectory position, gradient and number of hits. The final result is also a set of tracks (duplicate combinations are removed) where compatible hits from the TT are added to improve the measured momentum.

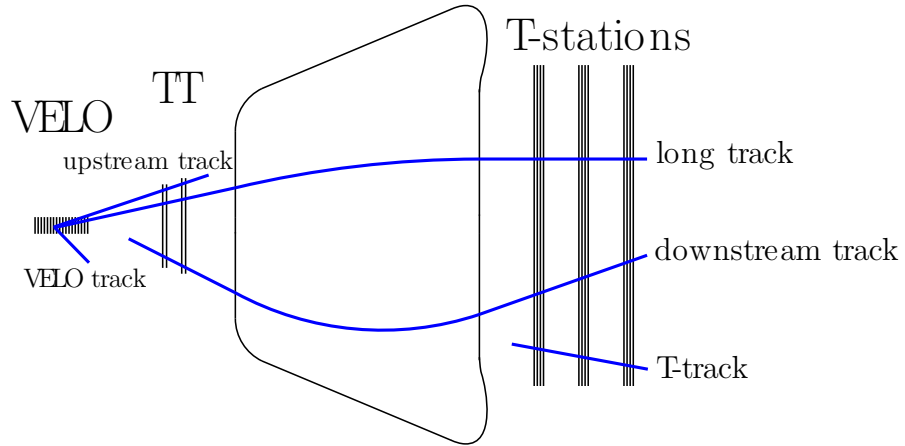


Figure 3.14 *LHCb track types [4].*

After track creation algorithms have finished, the final step is to improve track qualities by fitting each track with a Kalman filter [85]. Most of the improvements during this procedure are due to including corrections for multiple scattering in the detector material and ionisation energy losses. The fit quality is evaluated based on the χ^2/ndf [§] metric and assigned as the overall track quality.

[§] χ^2/ndf : χ^2 per degree of freedom

Apart from valid tracks, produced are also so-called “ghost” tracks. These are tracks that get created by the algorithms but do not correspond to any actual particle trajectories. Generally, these are due to wrongly matching the VELO and T station tracks together. The rate of “ghost” tracks is between 6.5 % to 20 % in minimally biased events. To constrain the problem and filter out “ghost” tracks more efficiently, a pre-trained neural network is used to assign a number to each track signifying its probability of being a “ghost” one. The network’s input is the track’s fit result, kinematics, the difference between measured and expected hits. Later algorithms and analyses themselves can decide to accept only tracks that have been assigned a probability of being a “ghost” below a certain threshold.

3.3.6 Tracking performance

Quantifying how well track reconstruction algorithms perform is done with a few metrics collectively called tracking performance. Practically, it determines the efficiency associated with track creation, reconstruction and assignment (tracking efficiency) and resolution on measured track kinematics (track momentum resolution).

At LHCb, the efficiency is measured using a “tag-and-probe” method using a plentiful and reliable event signature – the $J/\psi \rightarrow \mu^+\mu^-$ decay [70]. The J/ψ (mother) is reconstructed from two daughter muons. One of the muons (the ‘tag’) is fully reconstructed. It needs to pass strict cuts for track quality, kinematics, etc., while the other (the ‘probe’) need only be partially reconstructed and have an associated signature in the muon subdetectors. Then muon pairs of tag and probe particles must pass an invariant mass cut to be in the range of the J/ψ mass. Finally, the tracking efficiency is calculated as the ratio of probe muons which can be matched to actual fully reconstructed muon tracks divided by the total number of probe muons included in the J/ψ events.

Figure 3.15 shows the estimated tracking efficiency distribution comparing the years 2011, 2012, and 2015. Predominantly the efficiency is above 96 % and also improved significantly during the Run 2 data-taking. However, it is clear that track multiplicity is a major factor given it contributes to the amount of ghost tracks. That is understandable as the hit and thus track separation decreases with more tracks, resulting in hit association and it becomes increasingly more difficult to maintain the same ghost track rate.

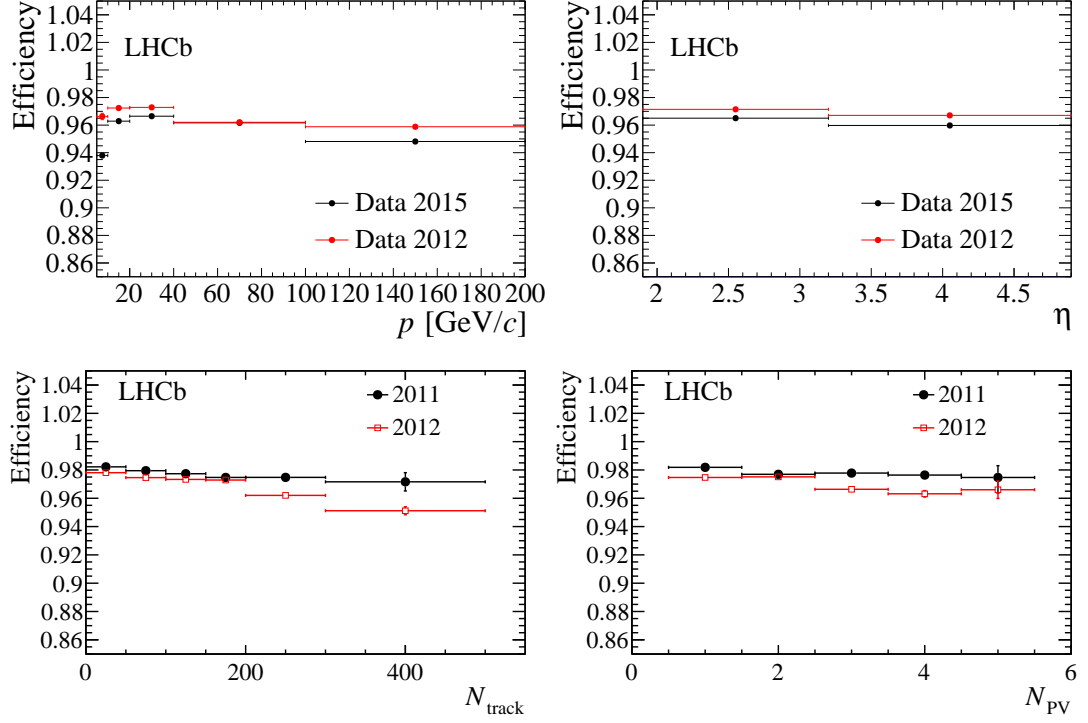


Figure 3.15 *Track efficiency in terms of momentum (top left), pseudorapidity (top right), track multiplicity (bottom left) and number of primary vertices in 2011, 2012 and 2015 data [4, 70].*

The momentum resolution on tracks is also measured using the same decay type. However, in this case, it is determined using the resulting invariant mass resolution of the J/ψ because the momentum resolution on the muons dominates the J/ψ mass resolution. Given the process is a two-body decay, it is safe to assume that the momentum of the muons will be similar and that their mass can be ignored⁹, giving an expression for the momentum resolution ($\delta p/p$) like so:

$$\left(\frac{\delta p}{p}\right)^2 = 2\left(\frac{\sigma_m}{m}\right)^2 - 2\left(\frac{p\sigma_\theta}{mc\theta}\right)^2$$

where m is the invariant mass, σ_m is the Gaussian width of the invariant mass fit, θ is the opening angle of the muon pair, and σ_θ is the error on the opening taken from the track fit of the muon tracks. On Figure 3.16 are shown the relative momentum resolution and the relative mass resolution. The relative mass resolution is determined from an invariant mass fit performed separately on six $\mu^+\mu^-$ resonances – J/ψ , $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and Z^0 .

Furthermore, to compare the absolute mass resolution between Run 1 and Run 2,

⁹ $m(\mu) \ll p(\mu)$ at LHCb

equivalent samples of $J/\psi \rightarrow \mu^+\mu^-$ events from 2012 and 2016 data-taking periods are used [4]. In this case, the J/ψ originates from a b hadron decay. Its invariant mass distribution is modelled using a Crystal Ball function [86–88], where the resolution is estimated using a weighted mean of the standard deviations of the two Gaussian components. The resolution for the 2012 sample is $12.4 \text{ MeV}/c^2$, while for the 2016 sample, it is $12.7 \text{ MeV}/c^2$. The difference comes from a slightly higher momentum spectrum in 2016 due to the higher beam energy in Run 2 and a slight degradation in the performance due to a simplified description of the detector geometry throughout Run 2 [4].

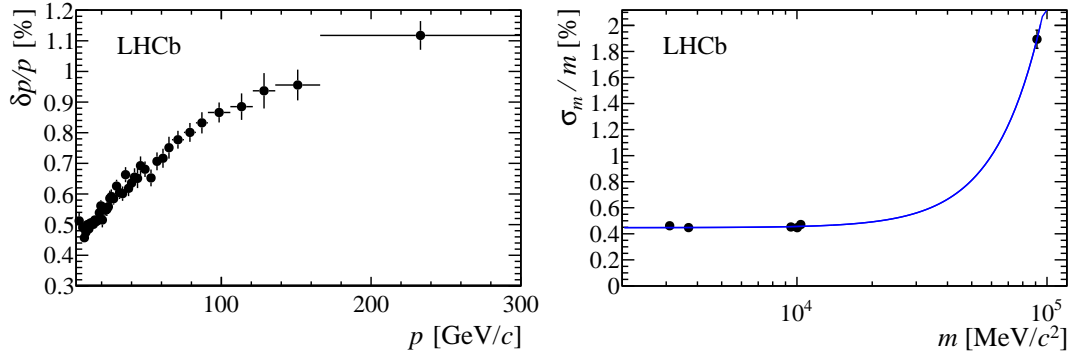


Figure 3.16 *Left: Relative track momentum resolution in terms of muon track momentum in $J/\psi \rightarrow \mu^+\mu^-$ decays. Right: Relative invariant mass resolution estimated from several $\mu^+\mu^-$ resonances in terms of total invariant mass. The solid blue line is a power-law function fit to the data [70].*

3.4 Particle Identification

Particle identification (PID) at LHCb is performed with the Ring-imaging Čerenkov (RICH), the calorimeters, and the muon systems. The RICH system employs Ring-imaging Čerenkov detectors to differentiate charged long-lived hadrons (namely kaons, pions and protons) from other particles and identify them. The calorimetry system is responsible for identifying and measuring the energy of photons, electrons and hadrons. The muon system identifies and measures the momentum of muons.

3.4.1 RICH system

B meson decays have access to a plethora of modes and final states. Consequently, many topologically similar decays can only be distinguished by precisely determining the final state particles. For example, the reconstruction of $B_s^0 \rightarrow J/\psi(\phi \rightarrow K^+K^-)$ decay, which is used to constrain the ϕ_s phase, can be polluted by $\Lambda_b^0 \rightarrow J/\psi pK$ and $B_d^0 \rightarrow J/\psi K^+\pi^-$ as misidentified backgrounds. As well, albeit more rarely it can also be reconstructed from $B_s^0 \rightarrow J/\psi \pi^+\pi^-$ when both kaons are misidentified. Therefore, distinguishing the decay modes puts direct requirements on adequately identifying the final states correctly.

3.4.2 Čerenkov effect

Charged particles traversing a dielectric medium of refractive index n with speed greater than the local speed of light c/n polarise it faster than the local phase velocity. Thus the electromagnetic field cannot respond in time (Figure 3.17). The resulting spontaneous emission of photons forming a wavefront with a cone's shape along the charged particle's direction (Figure 3.18) is called the Čerenkov effect [21, 89].

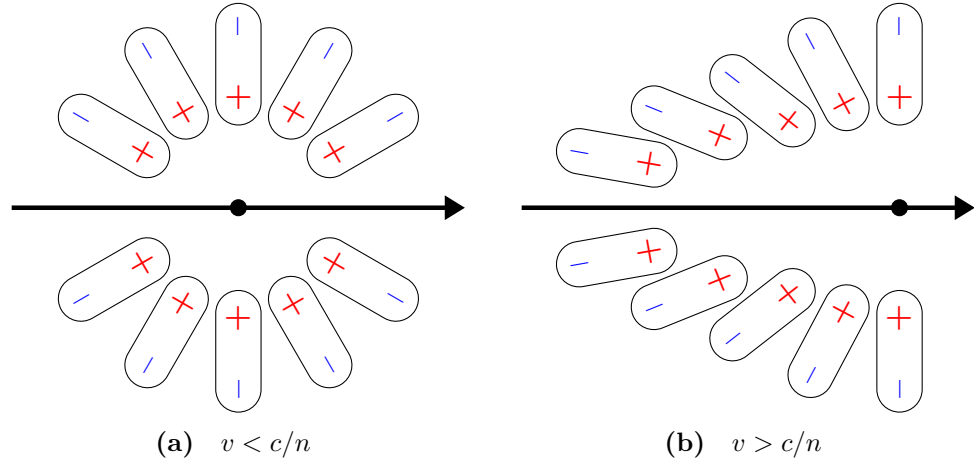


Figure 3.17 Čerenkov emission due to a charged particle in a dielectric. (a): As $v < c/n$, the symmetric structure of the induced dipoles results in no net dipole. (b): When $v > c/n$, a lack of causal connection between induced dipoles creates a net dipole. As the charged particle leaves, the medium returns to equilibrium through spontaneous emission of photons – called the Čerenkov effect.

The nature of emission is such that knowing just the refractive index n and

measuring the angle of the cone vertex θ_c^* determines the speed of the charged particle. More precisely¹⁰:

$$\cos \theta_c = \frac{1}{n\beta} \implies \beta = \frac{1}{n \cos \theta_c} \quad (3.4)$$

where $\beta = v/c$. The threshold for this effect is when $v = c/n$ exactly, which results in the photons being emitted directly along the particle's path (*i.e.* $\theta_c = 0$), while below this speed, the effect does not occur.

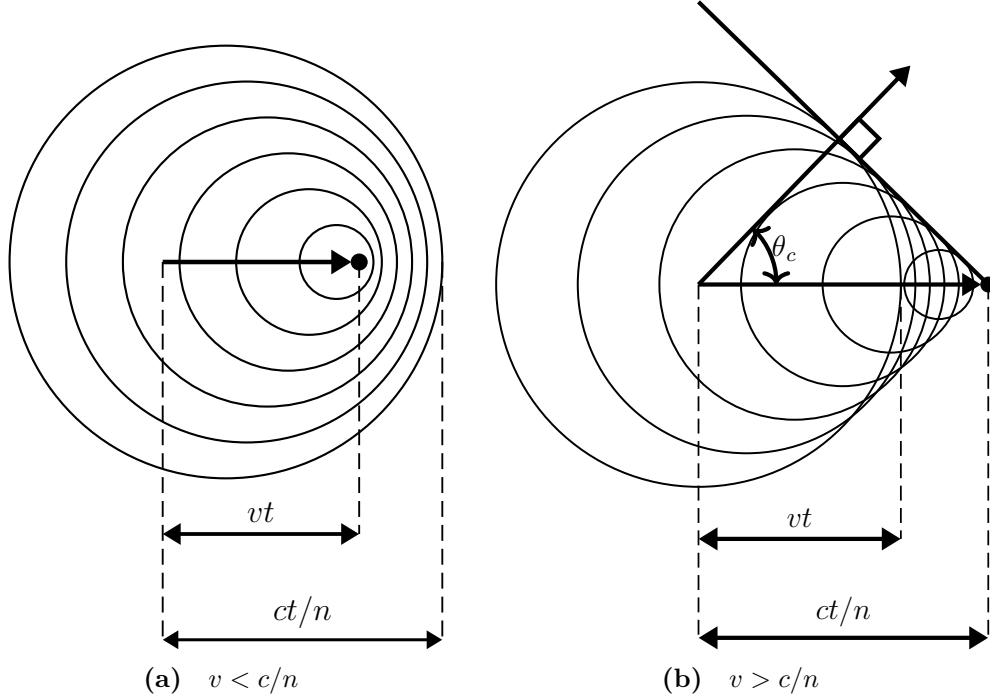


Figure 3.18 *Propagation of a charged particle in a medium with refractive index, n , and its connection to Čerenkov angle. (a): The particle moves with a locally subluminal velocity, and thus the generated waves are coherent. (b): The particle moves with locally superluminal velocity, and the generated waves are interfering constructively, forming a wavefront propagating at the Čerenkov angle, θ_c [91].*

The spectrum and intensity of photons is given by the Frank-Tamm formula [91]:

$$\frac{dN_\gamma}{dE} = \left(\frac{\alpha}{\hbar c} \right) Z^2 L \sin^2 \theta_c \quad (3.5)$$

where N_γ is the number of photons in a small energy range $\Delta E \in (E, E + dE)$, α is the fine structure constant, Z is the charge of the particle, and L is the average

^{*}known as the Čerenkov angle[21, 89]

¹⁰ignoring terms of order $\mathcal{O}(n^2)$ and higher, see Section B.1.1

path length of the particle (usually the width of the medium).

3.4.3 Ring-imaging Čerenkov (RICH) detectors

The Čerenkov effect can be a powerful tool. However, to be helpful, a way to easily measure the θ_c angle is needed. Typically, the associated practical issues are related to the fact that a 3-dimensional shape needs to be reconstructed, which means photon detectors need to be placed in all possible propagation directions and have excellent time resolution – cone structure is time-dependent. One can imagine all this becoming increasingly complicated when trying to handle not just one but multiple particles traversing the medium. Thus, a “naïve” Čerenkov detector will indeed be very complex and suffer from poor resolution.

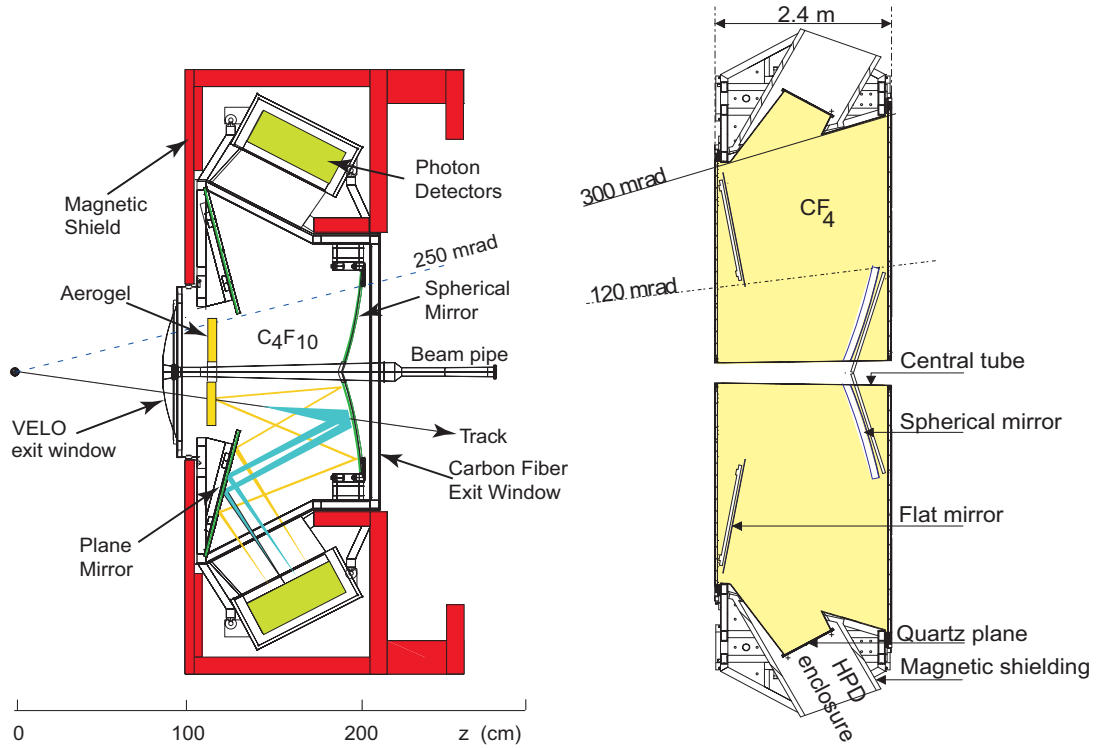


Figure 3.19 *Left: Side-view diagram of RICH1 with Čerenkov light paths. Right: Top-view diagram of RICH2. [21]*

Consequently, the Ring-imaging Čerenkov detector was invented [92]. The concept of this type of detector is to remove the timing aspect of the measurement altogether. The main idea is to collect the emitted light and focus it optically, causing the optical image to become a section of the Čerenkov light cone where the focus plane intersects it inside the medium. Instead of a 3-dimensional cone, only a 2-dimensional circle (ring, hence the name) needs to be reconstructed.

The ring radius is related to the Čerenkov angle, which in turn determines β^* . This method replaces accurate timing with accurate and precise optics, while the multiplicity problem becomes a pattern-finding issue.

Moreover, integrating Equation 3.5 over the energy bandwidth (photon spectral linewidth) assuming constant β gives the number of emitted photons:

$$N = \left(\frac{\alpha}{\hbar c}\right) Z^2 L \sin^2 \theta_c \int_E^{E+\Delta E} dE \quad (3.6)$$

Of course, in real detectors, the number of detected photons will differ, modified by the detector response N_0 :

$$N_0 = \left(\frac{\alpha}{\hbar c}\right) \int_E^{E+\Delta E} Q(E)T(E)R(E)dE \quad (3.7)$$

where Q , T and R are the quantum, transmission and mirror reflection efficiencies, respectively, all dependent on energy E . Finally, this gives the number of detected photons as:

$$N = N_0 Z^2 L \sin^2 \theta_c \quad (3.8)$$

$$= N_0 Z^2 L [1 - \cos^2 \theta_c] \quad (3.9)$$

$$= N_0 Z^2 L [1 - (1/n\beta)^2] \quad (3.10)$$

by substituting from Equation 3.4. Overall, this is significant because one must appropriately select the Čerenkov medium (radiator) used based on the expected particle momenta to maximise detector efficiency.

At LHCb, the RICH system consists of two subdetectors, RICH1 and RICH2, with two different radiators to fully cover the momentum acceptance [21]. Both subdetectors use spherical mirrors¹¹ to focus the light and a different set of flat mirrors to direct it onto an array of photon detectors (photon detector plane) located outside the acceptance. The overall geometry is shown in Figure 3.19, and their position in the overall LHCb experiment is shown in Figure 3.3. RICH1 and RICH2 both consist of a gas enclosure containing the gas radiators – C₄F₁₀ with $L = 85$ cm, $n = 1.0014$ and CF₄ with $L = 167$ cm, $n = 1.0005$ for RICH1 and RICH2 respectively. During Run 1, there was also a layer of silica aerogel [93] inside RICH1. It was used to cover the lower end of the momentum spectrum, but it was removed before Run 2 because it was severely degraded by particle

*See Section B.1.2

¹¹See Section B.1.3

radiation and was polluting the gas radiator.

The angular coverage of RICH1 is 25-250 mrad in the vertical plane and 25-300 mrad in the horizontal. By design, it is targeting the low-momentum spectrum of 2-60 GeV/ c . RICH2's angular coverage is 15-100 mrad and 15-120 mrad in the vertical and horizontal planes, respectively and targets the momentum range of 15-100 GeV/ c .

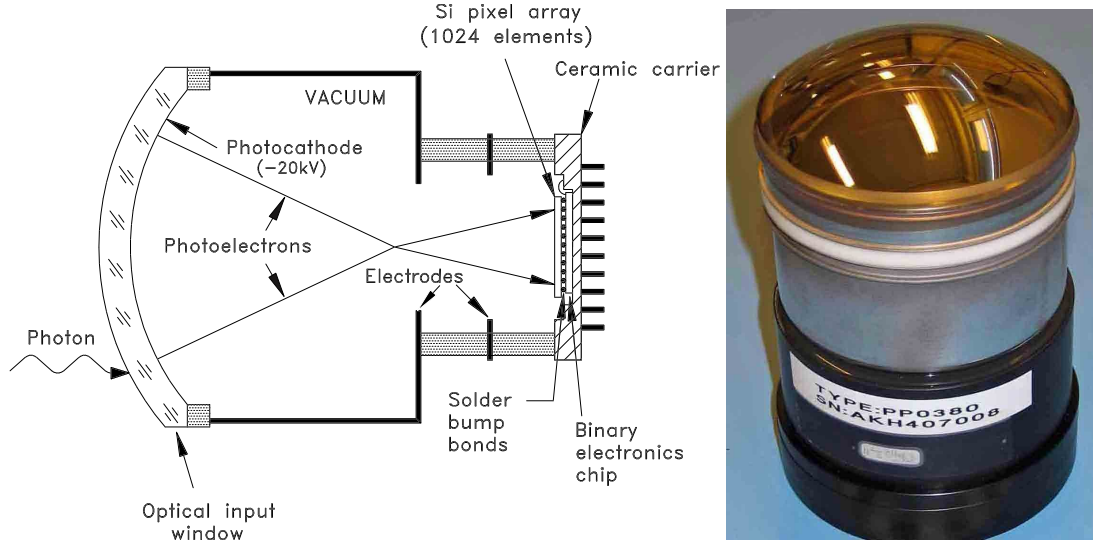


Figure 3.20 *Hybrid photon detector (HPD) schematic diagram (left) and photograph (right). [21]*

The photon detectors of choice to detect the Čerenkov light are the Hybrid Photon Detectors (HPDs). Figure 3.20 shows a schematic diagram and an image of an HPD. An HPD is a vacuum photon detector combining a photocathode, an accelerating high-voltage electric field (commonly referred to by “high voltage” or “HV”) and a silicon pixel sensor at the anode end. In principle, a Čerenkov photon hits the photocathode and gets converted to an electron, which in turn is accelerated by the HV (typically ~ 20 kV) towards the reverse-biased silicon sensor. The electron generates electron-hole pairs as it loses energy inside the silicon and the reverse-bias voltage (commonly referred to by “low voltage” or “LV”) creates a small current to fill them, which is used to identify the signal.

In more detail, the front part of an HPD is made of a single-piece optical quartz window in the shape of a section of a sphere. Quartz is chosen for its light transmission properties – it is transparent for wavelengths that satisfy the quantum efficiency requirements of the photocathode, namely between 200 nm and 600 nm [94, 95]. The spherical shape is chosen mainly for structural integrity

considerations, given the need for keeping a good quality vacuum. The inside wall of the quartz windows is coated with a thin multi-alkali layer serving as the photocathode. The tube’s HV potential is shaped in such a way as to demagnify the incoming photoelectrons by a factor of five¹², effectively “focusing” them onto the silicon sensor. On average, a photoelectron produces an electron-hole pair for each 3.6 eV of energy it deposits into the sensor. With an HV of -20 kV this means each hit generates about 5000 electron-hole pairs.

The silicon sensor is $300\text{ }\mu\text{m}$ thick and divided into 8192×8192 pixels, each with a size of $62.5\times 62.5\text{ }\mu\text{m}$. The sensor is reverse-biased to ~ 80 V and bonded to a binary readout chip. The readout chip keeps track of all the accumulated charge above a preset threshold that accumulates in a specific short time interval within the detector trigger time window and converts that to a binary signal hit. The readout has two modes of operation – a complete, granular, but slow readout of each pixel in the sensor and a faster mode that combines 8×8 neighbouring pixels by logical OR to form a lattice of 1024×1024 pixels (also-called “super”-pixels) of $500\times 500\text{ }\mu\text{m}$ in size. LHCb uses the fast readout mode. The demagnification is fivefold, which results in an effective spatial resolution of 2.5×2.5 mm at the photocathode.

The HPDs are all housed in columns supplying data connection, power and cooling. The columns, in turn, are enclosed in iron boxes to reduce the effects of external stray magnetic fields, such as the one generated by LHCb’s bending magnet, which would otherwise interfere with the paths of photoelectrons. Moreover, each HPD is also protected by an individual magnetic shield of mu-metal¹³. All of this allows the correct operation of HPDs in fields as strong as 50 mT.

3.4.4 RICH system performance

Typically, a single event in the LHCb detector contains hundreds of particles that pass through the RICH subdetectors. These particles create multiple overlapping Čerenkov rings in the image planes, which complicates their reconstruction greatly. Sifting through the rings is done by the pattern matching and reconstruction algorithms in the HLT and later in the Offline Reconstruction

¹²demagnification is accomplished by way of shaping the electric field with extra electrodes inside the HPD body

¹³mu-metal is a nickel-iron alloy with high magnetic susceptibility

computer farm. Since the most abundant particle in pp collisions are pions, once a track and a ring have been associated, the particle is assumed to be pion [96]. Then a likelihood minimisation procedure constructs several competing likelihood hypotheses for each particle type: e , μ , π , K and p [96]. Each hypothesis is constructed by first assuming the mass of the particle type of interest and then recalculating the likelihoods using information from the two RICH detectors combined with the rest of the subdetectors of LHCb. This process allows distinguishing between particle types based on the logarithm of the likelihood difference between the desired type and the other probable type. For example, in the case of a kaon, the logarithm of the likelihood difference between a particle being a kaon or being a pion or proton is calculated, *e.g.* $\Delta \log \mathcal{L}(K - \pi/p)$.

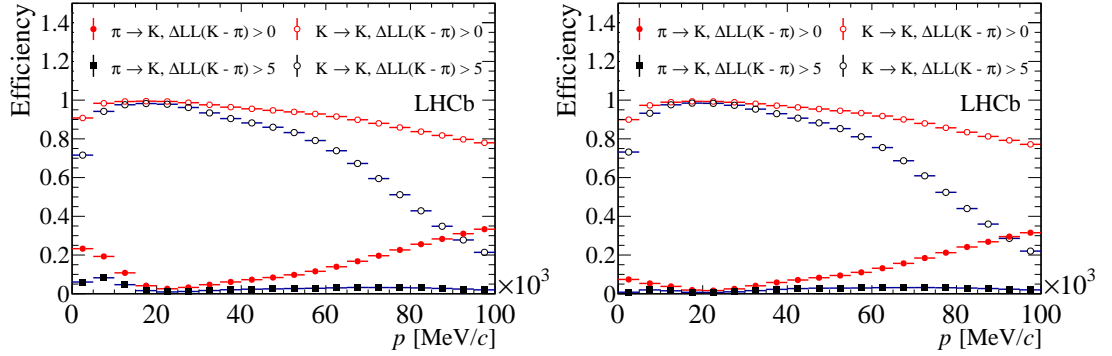


Figure 3.21 *Kaon identification efficiency ($K \rightarrow K$) and pion-as-kaon misidentification rate ($\pi \rightarrow K$) as a function of track momentum in two different boundary conditions on $\Delta \log \mathcal{L}(K - \pi)$: > 0 (red) and > 5 (black). [4]*

Thus, the RICH system is graded on its likelihood hypothesis separation capabilities. Its performance when it comes to PID is measured using control samples of data. Data is selected using only kinematic properties in decay channels, which are easier to identify and contain the particle types of interest only. Namely, the channels used are $K_s^0 \rightarrow \pi^+\pi^-$, $\Lambda \rightarrow p\pi^-$, $D^{*+} \rightarrow (D^0 \rightarrow K^-\pi^+)\pi^+$ and $J/\psi \rightarrow \mu^+\mu^-$. The kaon, pion and proton identification and misidentification rates are estimated using the latter decays. Figure 3.21 shows the kaon identification efficiency, and the pion-to-kaon misidentification rate is shown for a range of track momenta and two PID log-likelihood requirements.

3.4.5 Calorimeters

The calorimetry system aims to distinguish between electrons, photons and hadrons, and provide energy and position measurements [97]. The system also provides crucial information for the transverse energy measurements used in the hardware trigger (L0 in Section 3.5.1) decisions.

When it comes to the detector layout, LHCb has assumed the conventional approach of using an electromagnetic calorimeter (ECAL) followed by a hadronic calorimeter (HCAL). The ECAL is supplemented by an additional set of two specialised detectors: the Scintillating Pad Detector (SPD) and the Preshower detector (PS). They are placed in front of the ECAL. Their purpose is to reduce the background contribution from the large number of pions that would otherwise dominate the hardware electron trigger by improving the electron identification as well as help the reconstruction of B decays containing a γ or a π^0 . This feat is accomplished by ensuring different particle types leave different signatures across all subdetectors. As shown in Figure 3.22, the SPD identifies charged particles and distinguishes e^- from γ/π^0 , while the PS identifies electromagnetic particles helping to separate hadrons¹⁴.

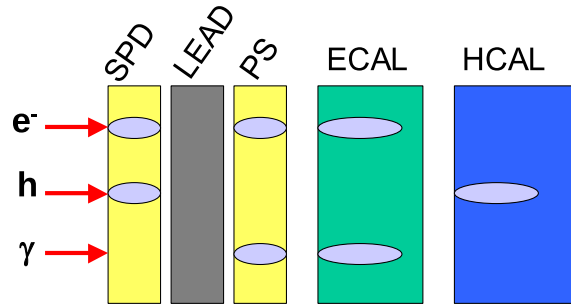


Figure 3.22 Particle shower (light grey area) signature of electrons (e^-), hadrons (h) and photons (γ) in SPD, PS, ECAL and HCAL [98].

All the subdetectors in the calorimeter system share a working principle – wavelength-shifting¹⁵ (WLS) fibres transfer scintillation light to photomultipliers (PMTs). The fibres from the SPD and PS are read out by MaPMTs, while individual phototubes read out the ECAL and HCAL fibres.

¹⁴Hadron showers have a longer shower profile.

¹⁵a wavelength shifting material is such that it can absorb higher frequency photons and re-emit lower frequency photons. This is normally used to shift scintillation light into the frequency range of a receiving photomultiplier.

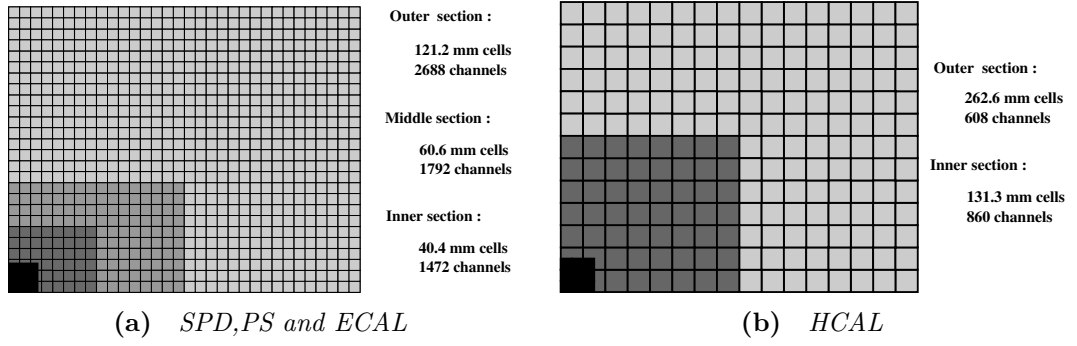


Figure 3.23 segmentation of (a) SPD, PS, ECAL, and (b) HCAL [97].

The thickness of the ECAL is 25 radiation lengths on top of the 2.5 of the SPD and PS, which allows it to contain high energy electromagnetic showers. The HCAL, on the other hand, is only 5.6 interaction lengths as it is space-limited. The calorimeter system has an angular acceptance between 25-250 mrad vertically and 25-300 mrad horizontally to match that of RICH1.

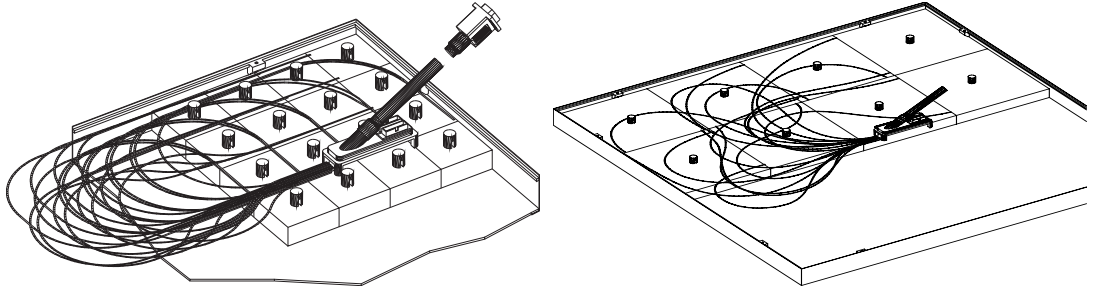


Figure 3.24 Schematic of SPD/ps inner (left) and outer (right) module [97].

The particle flux and thus the hit density decreases with the distance from the beam pipe¹⁶. To ensure the calorimeter provides roughly constant angular resolution, the subdetectors are variably segmented in the x - y plane with channel density increasing closer to the beam (as shown in Figure 3.23). Furthermore, the SPD, PS and ECAL channel size is scaled in such a way as to achieve almost projective symmetry with origin the interaction point. For example, the channels of the PS are roughly the size of the projection of the SPD channels in the x - y plane when looking from the interaction vertex. This means the segmentation of the SPD and PS in θ and ϕ matches that of the ECAL.

The SPD and PS are two almost identical planes of 15 mm-thick scintillator pads with a 12 mm-thick lead sheet sandwiched between them. Each scintillator pad

¹⁶hit density is about two orders of magnitude lower at the edge of the calorimeter compared to the innermost part closest to the beam

has a groove that houses a helicoidal WLS fibre accepting the light generated in the scintillator, shifting its frequency down and directing it to an MaPMT for readout (Figure 3.24).

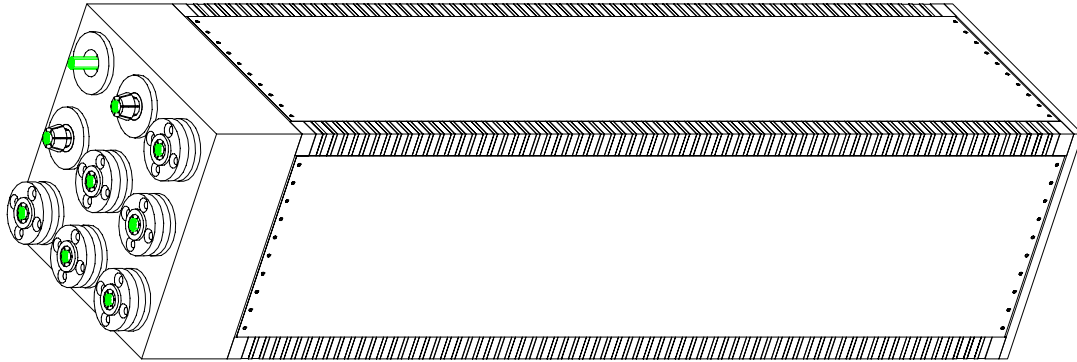


Figure 3.25 *Inner ECAL module [97].*

The ECAL is a series of modules consisting of layers of 2 mm-thick lead and 4 mm-thick scintillator tiles wrapped in TYVEK^{TM17} paper for light-tightness. WLS fibres are threaded through the scintillating tiles and are also read by a PTM. Figure 3.25 shows an ECAL module schematic diagram. The modules closest to the beam are segmented into nine cells, the middle region has modules segmented into four cells, and the farthest modules are not segmented. All the fibres running through each cell are bundled together and fed to one PMT per cell.

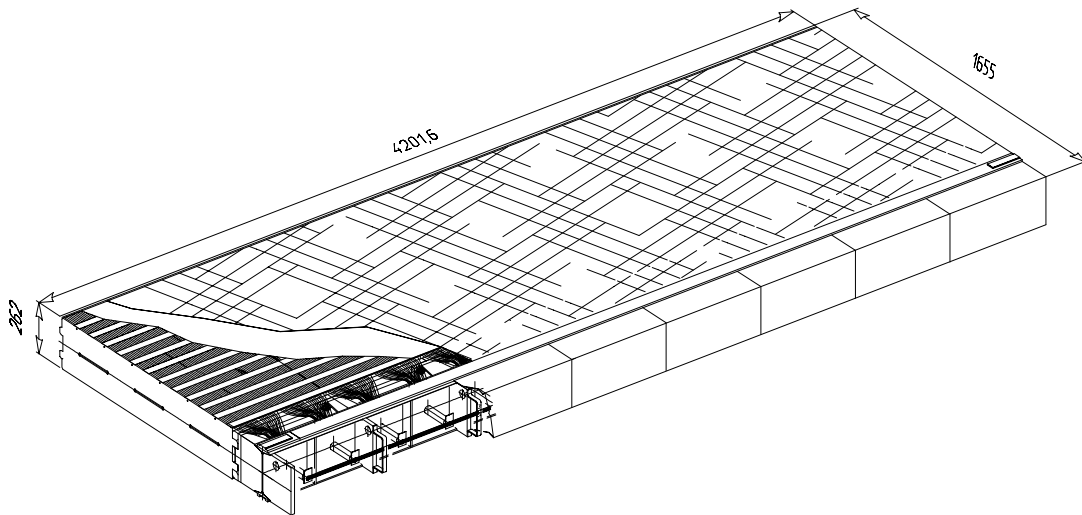


Figure 3.26 *HCAL module schematic [97].*

¹⁷commercial name of a synthetic material made from non-woven high-density polyethylene fibres, trademarked and produced by DuPont, which is opaque among other valuable qualities.

The HCAL design is quite different from other calorimeters. It is made from layers of iron and scintillating tiles, but they are oriented parallel to the beam pipe. The scintillating tiles layer contains the tiles with 10-mm-long sections of the iron layer protruding in-between so that the combined length of the iron protrusion and the tile is the same as the hadron interaction length in steel. The tiles and protrusions form a periodic structure with 20 mm thickness and 216 periods – a module. The scintillating light is taken up by WLS fibres and shown onto PMTs. A module is shown in Figure 3.26. It can be seen how fibres are bundled together to form cells. The cell segmentation is shown on Figure 3.23b. The inner region (closest to beam pipe) is made of square cells with a side of 131.3 mm, while the outer region cells have a side of 262.6 mm.

The energy resolution of the ECAL and HCAL modules was determined with data from a test beam. The parametrisation $\sigma_E/E = a/\sqrt{E} \oplus b \oplus c/E$ (E in GeV) is used, where a, b and c stand for the stochastic, constant and noise terms, respectively.¹⁸ Depending on the type of module and test beam conditions, the stochastic and constant terms were measured to be $a = (9 \pm 0.5) \%$ and $b \approx 0.8 \%$ for ECAL and $a = (69 \pm 5) \%$ and $b = (9 \pm 2) \%$ for HCAL with noise term isolated by fit to data [66].

3.4.6 Muon chambers

Detecting muon particles is essential as they are present in the final state of many B meson decays and, of course, the head channels of LHCb: $B_s^0 \rightarrow J/\psi \phi$ and $B_s^0 \rightarrow \mu^+ \mu^-$. Furthermore, muons are a crucial part of flavour tagging using semileptonic B decays. Also, the muon system, together with the calorimeters, is part of the L0 trigger (Section 3.5.1). The muon subdetector (MUON) system [99] is illustrated in Figure 3.27. It has five distinct stations (M1-M5). The first station (M1) is located upstream of the calorimeter to improve the transverse momentum measurement used in the trigger. M2-M5 are located downstream of the HCAL and are interspersed with 80-cm-thick iron absorbers. This is about 20 times the interaction length and set a minimum required momentum for muon to traverse the whole of LHCb to about 6 GeV.

The MUON detectors are segmented similarly to the calorimeters so that the channel occupancy is relatively constant. Each MUON station is split into

¹⁸ \oplus is used to denote the direct sum between relative error functions.

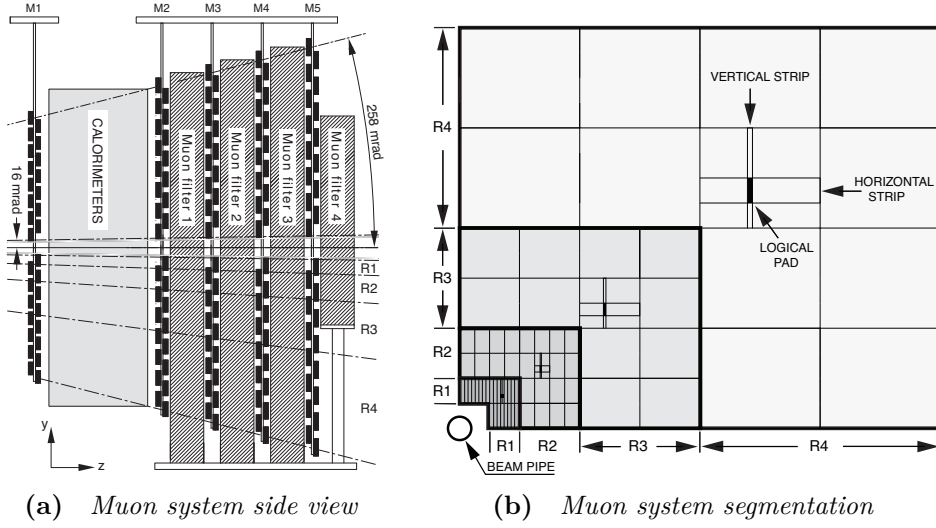


Figure 3.27 Schematic diagram of (a) muon system side view and (b) station segmentation [99].

four regions. Each region is made up of muon-detecting chambers split into groups of pads. The pad density changes between regions, where the innermost region has the highest pad density and the outermost the lowest, as shown in Figure 3.27b. Furthermore, stations vary their average pad density as per performance requirements. For example, M1 has half the horizontal pad density of M2 and M3, while M4 and M5 have half that of M1. The first three stations require higher spatial resolution to provide better transverse momentum resolution (about 20 %) for muon tracks. This is because transverse momentum is determined from the slope of the track in the horizontal plane, knowing the kick of the magnet and assuming the particle originated at the interaction point. At the same time, the last two stations are only used for muon identification and do not require good momentum resolution. The detecting technology is mainly Multi-Wire Proportional Chambers (MWPCs) for all pads except for the inner region of M1, where particle flux exceeds safety limits for ageing. There, triple Gas Electron Multipliers (triple-GEMs) are used.

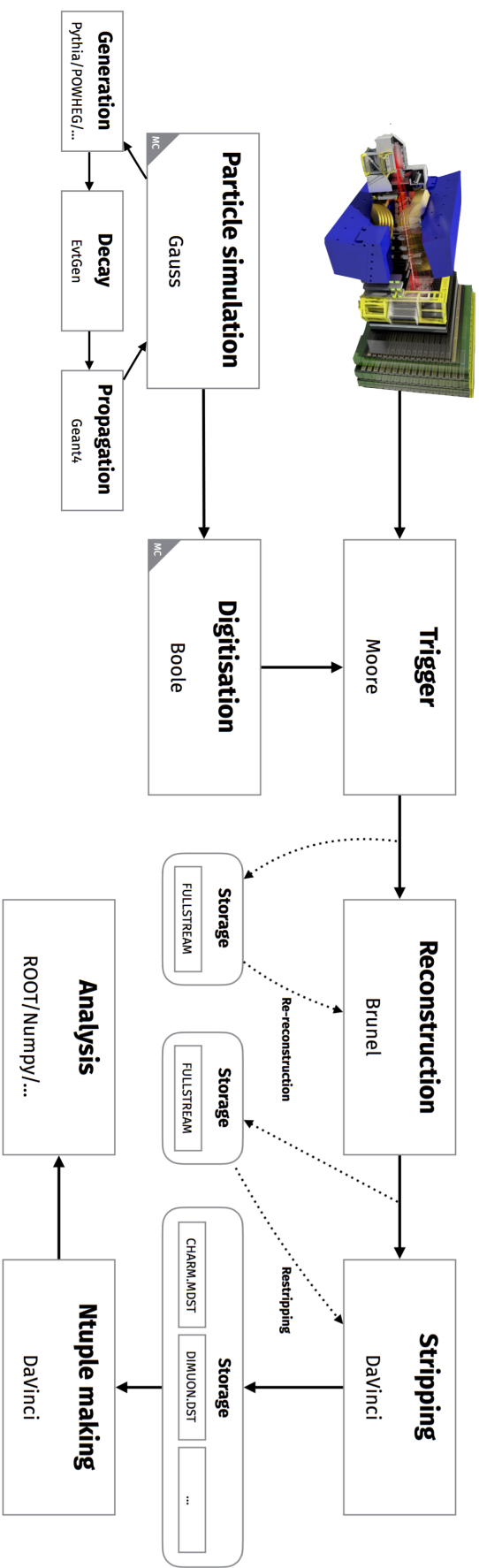


Figure 3.28 *LHCb data flow showing the associated data and simulation processing steps and relevant software [100].*

3.5 Trigger, Reconstruction and Data Processing

LHCb has several mechanisms to ensure it can collect valuable data, accurately record and distil helpful information, and present it in a meaningful way for later analysis. The trigger system is responsible for operating at the high interaction frequencies of the LHC, selecting events for detailed detector readout and later processing. The reconstruction algorithms are responsible for processing the raw data from the full detector readout and converting it into meaningful individual measurements and the overall structure of the observed events. The data processing includes the final steps in preparing the data and presenting the information in an accessible way for analysis. The overall LHCb data flow is shown in Figure 3.28.

3.5.1 Trigger (L0 and HLT)

The LHCb trigger is designed to operate at the full LHC bunch crossing rate of 40 MHz. This is, of course, taking into account that not all bunch crossing interactions are visible to the detector and that LHCb employs luminosity levelling (Section 3.1). In essence, at $\sqrt{s} = 13$ TeV and Run 2 average luminosity of $\mathcal{L} = 0.4 \text{ nb}^{-1} \text{ s}^{-1}$, the frequency of visible events is around 11 MHz. An event is considered visible to LHCb only if there are at least two charged particle tracks with enough hits in the VELO and tracking stations for the algorithms to cope with their reconstruction. The frequency of these events will be limited by some significant factors, such as the existence of “abort gaps”¹⁹ and many beam interactions resulting in elastic or diffractive scatters.

The maximum design rate at which LHCb detectors can be fully read out is around 1.1 MHz [4, 103]. Also, luminosity levelling keeps the maximum number of visible events per bunch crossing allowed on average at $\mu = 0.4$. However, during Run 2, that was increased to $\mu = 1.1$ to collect larger integrated luminosity, which only allowed the detector to be read out at a maximum rate of 1 MHz due to the increased data rate.

Thus, a system of field-programmable gate arrays (FPGAs) with a fixed latency

¹⁹normally about $3 \mu\text{s}$ to accommodate the abort-kicker-magnet rise time [101]

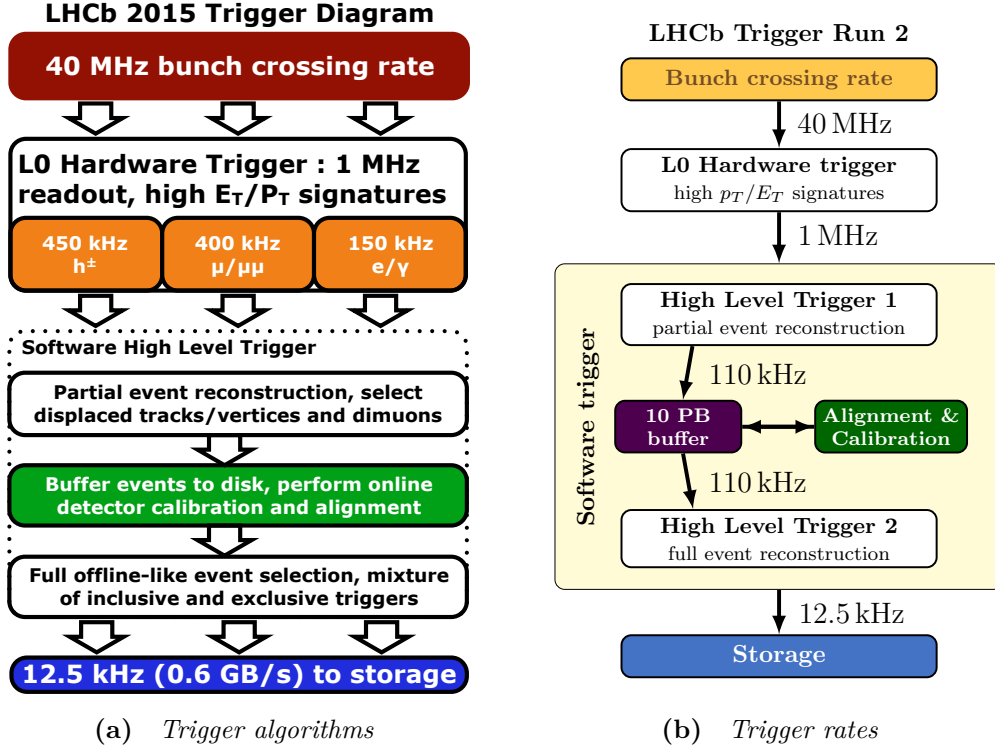


Figure 3.29 Algorithm workflow (a) and individual trigger component rates (b) for the LHCb L0 and HLT triggers during Run 2 [4, 102].

of $4\ \mu\text{s}$, called the L0, is used to limit the number of events triggering full detector readout based on specific criteria. This system represents the hardware trigger (L0) of LHCb and is a crucial step in data-taking, making it at all possible to have detailed measurements of interaction events.

Once L0 makes a decision to select (triggers on) an event, all subdetectors are read out. The information is stored in a small data buffer, and the next stage of the trigger is invoked. This stage is the flexible software trigger, called the High Level Trigger (HLT), which does not operate under a fixed frequency, but makes more complex decisions asynchronously. Its purpose is to reduce further the event rate that will be written to permanent storage. Given operating conditions of the LHC (bunch spacing $\in (25, 50)\text{ ns}$, number of bunches $\in (1260, 2250)$) and the LHCb settings ($\mu \in (1.1, 2.0)$), the data rate at 1 MHz can be between 40 and 70 GB/s [104]. Back of the envelope calculation shows this will result in an unmanageable amount of 1.8 EB over the course of a data-taking year. The HLT uses clever algorithms to reduce that rate to a manageable data rate of about 0.6 GB/s.

The criteria (or conditions) required to be satisfied to incur a trigger is called

a “trigger line”. Trigger lines are used in both L0 and HLT to specify sets of conditions corresponding to distinct types of interesting events. Once the conditions of a trigger line are met, it is said the event passed that trigger line, and its related information is passed along the next stage of processing or storage.

Figure 3.29 shows the overall design of the trigger and an overview of its internal structure and rates. Given that the average rate of $b\bar{b}$ pairs produced and contained inside the acceptance of LHCb is around 100 kHz, that only about 15 % of them have all the decay products also within the acceptance, and that most of the relevant physical processes have branching fractions below 10^{-3} , it is clear that the triggers have to be very efficient at selecting the small fraction of interesting events while rejecting as much of background as possible.

The L0 trigger is thus divided into two independent parts – the muon and calorimeter trigger systems. They both send data into a joint decision unit which ultimately decides whether to pass the event to the next stage (HLT). The decision time is limited to $2\mu\text{s}$ out of the $4\mu\text{s}$ due to cable lengths and particle flight time. The unit allows for prescaling and condition overlap. To

L0 decision	E_T/p_T threshold			SPD multiplicity threshold
	2015	2016	2017	
Hadron	$> 3.60\text{ GeV}$	$> 3.70\text{ GeV}$	$> 3.46\text{ GeV}$	< 450
Photon	$> 2.70\text{ GeV}$	$> 2.78\text{ GeV}$	$> 2.47\text{ GeV}$	< 450
Electron	$> 2.70\text{ GeV}$	$> 2.40\text{ GeV}$	$> 2.11\text{ GeV}$	< 450
Muon	$> 2.80\text{ GeV}$	$> 1.80\text{ GeV}$	$> 1.35\text{ GeV}$	< 450
Muon high p_T	$> 6.00\text{ GeV}$	$> 6.00\text{ GeV}$	$> 6.00\text{ GeV}$	none
Dimuon $p_{T1} \times p_{T2}$	$> 1.69\text{ GeV}^2$	$> 2.25\text{ GeV}^2$	$> 1.69\text{ GeV}^2$	< 900

Table 3.1 *The L0 thresholds for the different trigger lines used to take the majority of the data for each indicated year [4].*

make such quick decisions, the architecture of the trigger takes advantage of the fact that normally b decay products exhibit high transverse energy, E_T , and momentum, p_T . From the point of view of the trigger, the detector components are segmented into clusters of 2×2 cells in the ECAL and HCAL. The calorimeter trigger decision is based on the deposited transverse energy into a cluster, defined as:

$$E_T = \sum_{i=0}^4 E_i \sin \theta_i \quad (3.11)$$

where E_i is the energy deposited in cell i , while θ_i is the angle between the

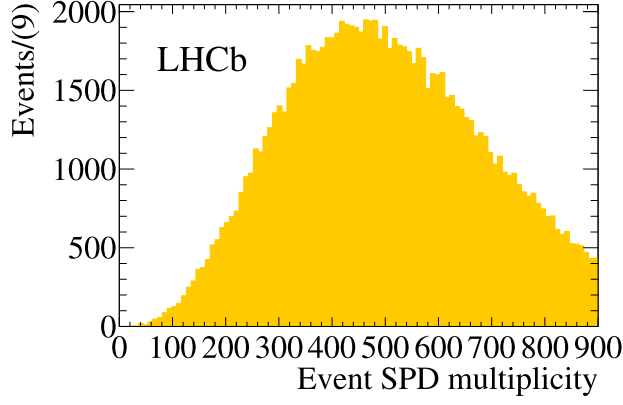


Figure 3.30 *SPD multiplicity for events containing $B^+ \rightarrow \bar{D}^0 \pi^+$ [4].*

z -axis and the line between the centre of that cell and the interaction point. Furthermore, information from the SPD and PS is used to differentiate between electron, photon and hadron associated clusters. This includes the signatures shown in Figure 3.22 as well as E_T thresholds listed in Table 3.1.

The muon trigger searches for muon tracks in the muon stations. The search is done by looking for straight tracks through all muon detector stations associated with originating at the interaction point. The two muons with the highest p_T are selected. The decision to trigger is based on either of two conditions: the highest p_T muon has p_T above a specific (**L0Muon**) threshold or the product $p_{T1} \times p_{T2}$ for the two muons is above another (**L0DiMuon**) threshold. The specific requirements are also listed in Table 3.1.

The per-event SPD multiplicity requirements are included in most L0 trigger lines to reduce the reconstruction complexity and improve reconstruction time in the later stages of the trigger. While the use of SPD requirements selects simpler and faster-to-reconstruct events, it does not result in a significant loss of absolute signal efficiency compared to using only energy and momentum thresholds [4]. That is due to the fact L0 signal-background discrimination deteriorates rapidly with event complexity except for the dimuon and electroweak trigger lines [4]. Thus, the high- p_T **L0Muon** trigger line does not have an SPD requirement to remove systematic uncertainties from the determination of its efficiency.

The following step of the HLT is split into two subsequent parts – the HLT1 and HLT2. HLT1 is responsible for partial event reconstruction and the precise reconstruction of the primary vertex, while HLT2 is the full event reconstruction. The HLT algorithms run on machines in the Event Filter Farm (EFF). The EFF

contains around 1700²⁰ computing nodes with about 27000 physical computing cores. In total, the EFF can accommodate around ≈ 50000 simultaneously running single-threaded processes since its physical cores employ simultaneous multithreading (SMT) technology called hyper-threading, which allow them to execute tasks requiring different computational units side by side. The HLT software is implemented in the MOORE [19] application, which is written in the same framework (GAUDI [105]) as the one used for the Offline Reconstruction of events for physics analyses. This combined with the increased capacity of the EFF and the improvements in the software itself allows the Offline Reconstruction steps to be performed in the HLT during Run 2. The Offline Reconstruction is done with a package called BRUNEL [19], also written using the GAUDI framework.

The HLT1 step uses information passed by the L0 decision unit about which trigger lines have been passed and uses this information to decide which algorithms to run to refine reconstructed charged particle tracks and vertices appearing in the tracking stations and VELO. Tracks that have traversed the complete LHCb tracking system (long tracks) are kept only when their reconstructed p_T is larger than 500 MeV/ c . The trigger timing constraints do not allow most of the PID algorithms to be run. The only exception is muon identification because of its clean signatures. Some of the HLT1 lines are specifically crafted such that their output (a specifically selected subset of events) serves as the input to the alignment and calibration tasks²¹. The primary vertex is reconstructed using VELO tracks only. This is because, first and foremost, the full set of VELO tracks is available in HLT1. Moreover, the full momentum information on long tracks is only available later in the HLT. However, neglecting that information does not introduce degradation in resolution and provides an early consistent PV position.

The HLT2 step takes care of charged particle track reconstruction, the reconstruction of neutral particles and PID information. The complete information from the tracking subdetectors is used to perform additional pattern recognition procedures, not possible in HLT1 due to time constraints. This results in finding high-quality long and downstream tracks and the most precise momentum estimate achievable. Similarly, the most precise algorithms for neutral particle clusters are also executed in HLT2. On top of that, after the HLT1 step for muon PID, further information is added from the RICH and calorimeters.

²⁰800 of the total 1700 computing nodes were added for Run 2, almost doubling the capacity.

²¹see Section B.2.1

3.5.2 Reconstruction

While full reconstruction is possible in the HLT and most of the tracks, clusters and PID information are fully reconstructed, some more complicated reconstruction steps can take quite a long time and require further processing during the Offline Reconstruction step. These include complex multi-body decay topologies, vertex, track and cluster likelihood fits which converge too slowly to be processed in the HLT, and full event reconstruction checks done using all the detector information anew. The events used in the analyses presented in this thesis have been fully reconstructed within the HLT as the reconstruction quality does not suffer from any ill effects [4, 106].

The reconstructed data is saved to files suitable for analysis, but they are rarely accessed directly (Section 3.5.4). Data are filtered further into pre-defined selections called STRIPPING. The STRIPPING selections are grouped into streams by similarity to save disk space and speed up access. So, only relevant streams of events will be processed in the analysis steps compared to processing all events each time for any analysis.

3.5.3 Simulation

Event simulation at LHCb is done with two pieces of software – GAUSS [107] and BOOLE [19]. GAUSS is the fundamental physics simulating application that generates the initial particles and simulates their traversal of LHCb. Then, BOOLE simulates the detector’s response and converts that information into the same format as if it comes from the LHCb detector readout.

GAUSS runs in three consecutive steps – event generation, decay and detector response simulation. The first step is accomplished by a dedicated event generator, PYTHIA 8 [108], which provides a generic user interface. PYTHIA is used to generate high-energy-physics events like those having sets of outgoing particles produced in the pp collisions at the LHC. The generator focuses on particle collisions with $\sqrt{s} \geq 10 \text{ GeV}$. Below those energies the abundance of hadronic resonances conflicts with some of the approximations needed to make PYTHIA work so well at higher energies. Thus, once produced at the interaction point, the decay of initial particles (B mesons) is simulated with EVTGEN [109]. EVTGEN was adapted by the LHCb collaboration to simulate

B mesons originating from pp collisions from its original version designed by BaBar collaboration for $e^+e^- \rightarrow \Upsilon \rightarrow B\bar{B}$ collisions. The output of EVTGEN is commonly referred to as “generator-level” events. The detector response uses the GEANT4 [110] package to simulate the physical interactions between detector components and the propagating particles. This step essentially generates hits in the sensors.

Finally, a digitisation step is needed, which converts the simulated hits from the previous step into the same output as the readout electronics. This is accomplished by the BOOLE application, which tries to reproduce the subdetector response, the readout and the L0 trigger hardware. It also considers beam bunch spacing and electronic readout effects, such as channel noise, cross-talk, and non-functioning (dead) channels. Then the HLT step is performed with the MOORE application equivalently to actual data-taking.

The reconstruction is also equivalent to that of real data, apart from one more procedure called “truth-matching”. BRUNEL tries to associate reconstructed tracks with the “generator-level” particles that created them. To associate a track with a generator-level particle, the fraction of reconstructed clusters used to build it coming from the particle must be larger than a pre-set threshold. Otherwise, the track is labelled a “ghost” track. If more than one track is associated with the same generator-level particle, those are labelled “clone” tracks.

3.5.4 Analysis

The DAVINCI [111] application is used to prepare the data and simulation samples in a meaningful way for further analysis. It allows access to information in a transparent and uniform format, handy for analysts at the LHCb Collaboration. DAVINCI software builds from the perspective of an event selection. Essentially, a search is defined for a specific entity (*e.g.* a set of processes, topologies or even bare particles), which DAVINCI uses to loop over all relevant input events. It gets all relevant information and can calculate quantities not readily available to check if all aspects of the selection are satisfied and if so, the event is accepted. As it is also written in the GAUDI framework, it can, for example, assign new PID hypotheses, and run constrained vertex fitting to realign PVs or decay vertices based on custom user conditions. It can also (re-)run flavour tagging algorithms, calculate particle decay angles and rescale or smear track momentum based on new user-supplied calibration.

3.5.5 Distributed computing (Dirac and the Grid)

The Worldwide LHC Computing Grid (WLCG) was created in 2002 [112] to facilitate all the computing power and accommodate the computational, storage, access, and processing needs of CERN and its users. It is a global collaboration of computing centres that links a large number of computers and storage systems in over 170 centres across 41 countries. This means it is organised as a distributed computing platform, where resources and computing time are shared, and infrastructure is interconnected seamlessly with ease-of-access in mind. The computer centres are arranged in “tiers” where each tier provides a different service. Services range from long-term storage and supercomputing sites through local processing clusters to user access terminals. The Grid, as it is commonly referred to, is capable of providing near-real-time access to LHC data from virtually any point on the globe.

This way of structuring the computing infrastructure allows for several advantages, such as data duplication (fast access and redundancy), eliminating single points of failure, spreading the computing load over several time zones thus ensuring round-the-clock monitoring and expert support and, of course, spreading the overall costs as per the availability.

Chapter 4

The LHCb Upgrade of the High Level Trigger (HLT)

The LHCb detector, including all of its subdetectors, will be upgraded to cope with an increase in luminosity from $\mathcal{L} = 4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ to a higher $\mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ during the Long Shutdown 2 (LS2) [20]. One fundamental change will be removing the hardware trigger system (L0) altogether and replacing it with a fully software-based trigger. The new trigger will run in a large computer farm and perform a full detector readout at 40 MHz rather than 1 MHz. These changes put incredible demands on the computing resources available and shift focus on the type of algorithms that could handle the incoming data.

4.1 Introduction

The purpose of the L0 trigger is to reduce the data flow to a manageable level for the HLT to handle. The HLT itself does the same, reducing a data flow of about 70 GBps (incoming at 1 MHz) to about a hundredth of that 0.6 GBps (outgoing at 110 kHz) [102]. After removing the hardware trigger, the software-based trigger will have to perform a similar reduction of data flow. However, its incoming rate will be about 40 MHz rather than the current 1 MHz. Coupled with the fivefold luminosity increase and the reduced time for triggering (by a factor of 40), the new trigger will have to make use of novel ideas to cope with operating in real-time conditions¹.

¹Real-time systems guarantee a response within specified time constraints.

Simply increasing the computing resources available linearly will not compensate for the increase in computational demand. This imbalance is partly due to algorithm execution performance not scaling linearly with the number of computational units. For example, most algorithms in the original GAUDI framework are not fully parallelised and are executed sequentially [17]. Increasing the number of CPU² cores available will, thus, have limited effect on the time it takes such an algorithm to complete. Another reason is that some algorithms depend on the result of others which cannot be trivially parallelised.

There are other such considerations, but one final important for this work is data parallelisation [17]. It is hard to put data in multiple places and process it in parallel for different purposes because of the amount of data, the real-time requirements and the latency of data I/O³.

4.2 LHCb HLT Upgrade

The most considerable inefficiencies in the trigger at LHCb come from the L0 decision, especially for hadronic decays [19]. One of the main objectives of the LHCb Upgrade is to remove this bottleneck and implement a readout system that can resolve the inelastic collision rate at the LHC of 30 MHz [19]. Such a system placed before the trigger will allow the software-based trigger to process all available information fully.

4.2.1 LHCb HLT Upgrade requirements

As per the physics focus of LHCb, the main objective of the software trigger is to select beauty and charm particles decaying into a large variety of final states with the highest possible efficiency and purity. The new software-based trigger should, thus, optimise the physics related metrics as this is the leading goal of the system. A good strategy is to aim the decisions and selections made to adhere as much as possible to the ones applied in the current offline analyses. That will result in maximising trigger efficiencies and minimising systematic uncertainties [19].

²central processing unit (CPU) is the part of the computer that executes the instructions needed to run an algorithm.

³I/O stands for input/output operations related to reading, writing or transfer of data.

Event Rate	40 MHz
Input Rate	30 MHz
Instantaneous luminosity	$2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$
Pile-up	7.6
Maximum processing time per event	4 ms
CPU nodes	< 4000

Table 4.1 *Boundaries and requirements for fully online software trigger [19].*

Some of the operational boundaries and performance requirements of the fully online software trigger are listed in Table 4.1. The design of LHCb’s new data centre, the Event Filter Farm (EFF), will limit the maximum number of CPU nodes because of its power, cooling and space constraints. Similarly, the input rate is bounded by the maximum number of proton bunches able to fit in the machine per beam, *i.e.* 2808, due to gaps associated with the rise time of the PS/SPS/LHC injection kickers and the LHC dump kicker [19]. Finally, the maximum processing time is estimated by the data rate, which is about [19, 113]: $1/40 \text{ MHz}/4000 \text{ nodes}/40 \text{ threads per node} = 4 \text{ ms}$.

4.2.2 Core Software

The core software design and architecture will have to adapt to support the advances of modern computing hardware and cope with the increased demands. Specifically, even though the hardware (CPUs) has continued to raise the number of computational units inside it, the single-threading performance has not followed suit in being ever-growing [17]. Instead, the additional computational power goes into extra computational threads to allow CPUs to perform more and more operations in parallel.

Given that most of the LHCb core software is implemented in the GAUDI [105] framework, this framework must evolve as to allow for the new HLT software to succeed. The specific details of this redesign are explained fully in the LHCb Upgrade Software and Computing Technical Design Report [17]. However, the crucial notes relevant for this work are encapsulated in the following key points. The GAUDI framework, and thus the HLT by extension, suffer from pivotal limitations, such as poor handling of computationally blocking operations due to other dependent operations, *e.g.* memory/disk/network I/O. Offloading heavy calculations to dedicated processors, such as general-purpose graphical processing

units (GPGPUs), is also not readily available.

Thus, the redesign of GAUDI is planned to address these limitations by concentrating the efforts on allowing it to take advantage of simultaneous multithreading as much as possible. The redesign also comes with the switch to create a computational queue where required processes/algorithms will be put on and scheduled. Then a multithreading scheduler algorithm will be responsible for assigning the tasks in the queue as efficiently as possible without being limited by the previous sequential model. The hope is to modernise the framework to allow for data parallelisation, inter and intra-event concurrency [17]. Finally, this will eventually allow for intra-task concurrency. In that sense, operations that previously prevented the continuity of computations because of latency or unfulfilled dependencies will only affect individual computation threads. These will be delegated to the queue, which will prioritise processing operations that can continue unobstructed. That way, required computations can be completed *just in time* (JIT).

4.2.3 HLT

There has been much work to improve upon the points raised in Section 4.2.2 by various proponents, including implementing real-time reconstruction operated in HLT2 during Run 2 [4]. Further work towards real-time calibration and alignment of the LHCb detector during operation has also been done [114]. Many of these approaches rely on saving only a subset of raw data collected from the detector itself. In contrast, most stored data are the trigger and reconstruction processing results like particles, tracks, algorithm and selection responses, and further higher-level information.

This approach is effective in significantly reducing the amount of data to be stored and processed during the initial trigger stages. However, it may prove dangerous when the raw data is required to re-process events for algorithmic improvements or in case of faults or bugs in the trigger software. As such, it has its advantages and disadvantages.

This chapter discusses an alternative approach, which relies on machine learning (ML) algorithms, to expand the trigger capabilities and reduce its latency.

4.2.4 Machine learning algorithms

Machine learning algorithms are becoming increasingly important for various purposes, specifically at LHCb for data set filtering, event selection, data and simulation corrections, and more (examples in Chapter 6). The power of ML is multipronged. Firstly, ML algorithms are primarily implemented nowadays by artificial neural networks (ANN or NN for short). The prevalent implementation of ANNs adheres to algorithmic and data parallelisation — meaning CPUs can process the data in chunks and in parallel. At the same time, the processing itself is typically achieved by taking advantage of parallelisable computing operations, such as matrix multiplication. These are very important for designing the new software trigger, as discussed in the previous sections.

Moreover, ML algorithms are much more flexible than standard algorithms to identify interesting event topologies and differentiate signal distributions from background noise. It has been shown that even some of the simplest ML algorithms can sufficiently approximate essentially any function [115]. This principle would enable the software trigger to be more capable of filtering events and allow for greater freedom of physics search space.

Furthermore, ML algorithms can potentially allow for significant improvements over the efficiency losses at the trigger. ML algorithms can be finely tuned to balance efficient event filtering and trigger evaluation time by combining the flexibility and execution latency reduction over sequential algorithms.

4.2.5 Machine learning algorithms at LHCb

More than 600 unique event signatures are being searched for in parallel in real-time at LHCb [1]. Each of these comes with its own set of requirements and inputs. However, only a few take advantage of any ML algorithms.

The specific reasons for this are unclear, but it is not for the lack of available tools. A large ecosystem of ML frameworks is open to any analyst to create various types of ML classifiers and other types of MVA algorithms. TMVA [116], Neurobayes [117] have been available for a long time. More recent examples include Scikit-Learn [118] and Keras [119], as well as TensorFlow [120] and PyTorch [121], just to name a few.

Furthermore, it has been shown in many LHC analyses that ML classifiers can account for differences in training variables between signal and background events while also preserving the intrinsic correlations adequately. An example of this is also the analysis presented in Chapters 6 and 7. Despite that fact, most searches of exciting and increasingly challenging to filter signatures are performed without ML.

Quite often, the reason for this is that the creation and training of a preferred ML algorithm is a task that requires a conceptually different set of skills than the task to translate this same algorithm into a GAUDI-compliant algorithm, the C++/PYTHON framework used at the LHCb trigger [4, 105].

Another more relevant to the trigger performance reason is that methods designed for fast evaluation, such as Bonsai Boosted Decision Trees (BBDTs) [122], do not quite meet the speed requirements of the full software trigger. In particular, BBDTs try to improve upon the general BDT performance by discretising the input variables. Input variables are not considered continuous but made discrete. Thus, they only allow for a limited number of tree nodes to be learned. However, that decision reduces BBDT granularity and does nothing to address the issue of scaling. Increasing the number of input variables grows the BBDT exponentially causing it to use too much memory and making its traversal impractical.

4.3 NNDrone

A novel framework that supports modern machine learning algorithms of various architectures is needed to address the issues mentioned above. It should provide suitable tools for the use and deployment of ML at the future fully software-based real-time trigger of LHCb. As described below, such a framework was developed, deployed, and assessed at the current HLT2 environment. This new framework allows analysts to train a *drone* neural network that can learn the critical features directly from a target ML model of any given package of preference, be it Scikit-Learn or Keras. The resultant drone network and its parameters are fed into a C++ algorithm that can run at any HEP production environment. Thus is also suitable for LHCb's future software trigger. The framework and the drone neural network that it produces and the configurable GAUDI compatible algorithms are collectively combined under a common toolkit called NNDrone [1].

The idea behind this is simple but powerful. Usually, complex ML architectures are left as one of the final steps of the selection process. They come after all of the L0, HLT, offline reconstruction, analyst defined selection in DAVINCI and other manual analysis specific steps. And there are good reasons for this. ML classifiers require relatively clean samples of events with the signal and background items clearly labelled. They are also cumbersome to train, test, validate, and are prone to unexpected biases and variances. One might say that an ML algorithm is only as good as its training and testing samples⁴. However, once arrived at a suitable classifier that passes all relevant metrics and scrutinies about efficiency, accuracy, precision, and minimal biases and variance, there are few reasons not to apply it everywhere. As already mentioned, these are primarily about the complexity of software implementation and the latency in execution performance. The proposed idea is to relieve these last two remaining hurdles and enable powerful analyst-trained fine-tuned machine-learning to be transferred to the earliest possible stages of event filtering to improve selection efficiency and overall data quality.

4.3.1 Drone neural network training

The principle behind creating and training a drone neural network that can successfully approximate a given model relies on the fact that it is possible and practical for the drone to find the same minimum of the loss function as in the original model [123]. The procedure starts with a simple feed-forward neural network [124] that initially has very few layers and few hyperparameters in each layer. The input and the output layers are, of course, adapted to suit the input and output spaces of the model being approximated, respectively. Then the training proceeds by probing the original model extensively in the parameter space where the accuracy of the drone response is desired. Subsequently, the hyperparameter space is iteratively expanded until the desired accuracy and performance are reached. In other words the drone can sufficiently mimic the responses of the original model. As mentioned briefly, a neural network's ability with a continuous, bounded, non-constant activation function to approximate arbitrary functions to an arbitrary degree has been known for a while [115](1991).

A simple feed-forward network is initiated with only one hidden layer to demonstrate the concept. Drones (drone neural networks) are usually initiated with the smallest possible complexity. The weights initialisation is of no

⁴not discussing unsupervised algorithms

consequence, and the $\mathcal{N}(0, 1)$ ⁵ is sampled to get the starting weights. The hidden layer is chosen to start with at least one more neuron than the output layer dimensions. The output layer is one-dimensional in this specific case, so the hidden layer must have at least two neurons. If it only has one, then it will be equivalent to the output layer. It is arbitrarily set to 5 neurons.

The activation function is chosen to be a sigmoid, $f(x) = 1/\exp(-x)$, which is defined everywhere on the real number line, has an *S* shape around zero, and its domain is between -1 and 1 to adhere to Reference [115].

Then the drone model is made to approximate an arbitrary input model. The drone has the same number of input variables and the same number of outputs. The training of the drone is using a supervised-learning approach, but not in the usual sense. Typically, the drone is given the same inputs labelled with the same desired outputs. However, in this case, a different strategy is placed. The drone is given the output of the original model for each input data point. Thus, the drone is set to mimic the original model and its responses while being *punished* for deviating. The loss function, \mathcal{L} , looks like so:

$$\mathcal{L} = \sum_i [F(\vec{x}_i) - G(\vec{x}_i)]^2 \quad (4.1)$$

where $F(\vec{x}_i)$ and $G(\vec{x}_i)$ are the outputs of the original and drone models on data point i of the mini-batch, respectively. The training is done using standard stochastic-gradient descent in batches.

In this case, the difference and the advantage is per-event equivalence between the drone and the original model, not just measured performance. In contrast, the original model might include complexity beyond its training data, which transcends into a better model overall: not just defined on the current dataset, but also perform well on future unseen data. This *knowledge* will not be captured by repeating the drone training with the same data but will be preserved if the drone successfully mimics the original model response.

Furthermore, another feature of this approach is that the drone never sees the original labels of the dataset. Instead, the drone extracts the properties of the original classifier directly. In a sense, the drone is empirically learning from other networks. Also, based on that, it is maybe capable of combining this empirical *knowledge* as well.

⁵ $\mathcal{N}(0, 1)$ stands for the normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$

4.3.2 Drone mutation and morphing during learning

It is important to remember that one of the main goals is to keep the evaluation latency of the model as low as possible. As mentioned briefly, drone networks' implementation takes advantage of parallelisation and does not suffer from branching penalties at the CPU level⁶. However, it is still vital to keep the hyperparameter space small and only add additional degrees of freedom when absolutely necessary. At the same time, the initial size of the drone network will seldom be enough to accommodate accurately approximating the original model.

For this reason, it is required to grow the drone until it can accurately reproduce the original model's response but keep a delicate balance with the number of hyperparameters and layer complexity to achieve low latency. The procedure to do this is, thus, called drone mutation or drone morphing. Essentially, the drone starts small. It is trained against the original model until no significant improvement in accuracy is observed. Then if the approximation is not as precise as needed, the drone is mutated by adding a new neuron to the hidden layer. This approach increases the learning capabilities of the drone to better match the original model.

Furthermore, the specific target CPU architecture may be considered by limiting the size a hidden layer can grow to before it becomes more latency-favourable to add a further hidden layer than to add more neurons to the current one. The following conditions are required to be met for a drone mutation to be triggered during the j^{th} epoch [1]:

$$\delta_j \equiv |\mathcal{L}_j - \mathcal{L}_{j-1}| / \mathcal{L}_j < \kappa \quad (\text{Condition 1}) \quad (4.2)$$

$$\sigma_j \equiv m(1 - e^{-b(\hat{t}+n)})\delta_j\mathcal{L}_j$$

$$\mathcal{L}_j < \hat{\mathcal{L}} - \sigma_j \quad (\text{Condition 2}) \quad (4.3)$$

where κ is the required threshold to be reached, σ is the required minimum improvement of the loss function, and $\hat{\mathcal{L}}$ is the value of the loss function when the hidden layer was last mutated. The required improvement starts from a minimum at n , increases with epoch number after the previous mutation, labelled by \hat{t} , with steepness b until a maximum at m . The precise values of the parameters κ , n , m , b are not of particular importance. Instead, the topology described by

⁶Branching occurs when particular instructions, like an if-statement, cause the CPU to deviate from or split the default instruction sequence.

Equations 4.2 and 4.3 is crucial. The relative loss function improvement, δ_j , can never realistically be greater than one. The limit, κ , at which no significant improvement is deemed to occur is acceptably set at 0.02 (smaller than 2σ standard deviations). The descent in loss space, $\hat{\mathcal{L}} - \mathcal{L}_j$, is additionally required to be significantly large, minimising the chance of getting stuck in isolated local minima. The function, σ_j , is chosen to increase this requirement with each epoch for two reasons. Firstly, it is bounded and can approach its asymptote arbitrarily fast. Moreover, it scales δ_j such that the loss descent must be significant before a mutation is triggered. Since δ_j is expected to decrease with epoch number, the minimum and maximum values of σ_j are chosen as such [1]:

$$\sigma_j(\hat{t} = 0) \equiv \min(\sigma_j) \equiv 2.5\delta_j\mathcal{L}_j \implies 5\sigma \text{ std.dev.} \quad (4.4)$$

$$\sigma_j(\hat{t} = \infty) \equiv \max(\sigma_j) \equiv 25\delta_j\mathcal{L}_j \implies 50\sigma \text{ std.dev.} \quad (4.5)$$

The steepness, b , is chosen such that the σ_j transition from its minimum to its maximum takes on average 50 epochs. This choice ensures a change cannot be triggered immediately after a previous one and the learning can still proceed if more freedom is required. Furthermore, it allows the network to stabilise after a significant change following mutation or escape any shallow local minimum.

When the conditions in Equations 4.2 and 4.3 are met, the model is updated to extend the weights matrices and bias vectors to accommodate the new hidden layer architecture. The associated neurons are initialised with zero weight to ensure the loss function value continuity between epochs.

4.3.3 High energy physics applications

Demonstrating the functionality and usefulness of the drone networks is best served with concrete examples. Initially, it is prudent to have the drone performance compared against an existing valid model. Moreover, it is best to evaluate the drone in multiple separate situations. Two disparate datasets and corresponding models are, thus, constructed for this purpose.

Firstly, a classifier between a desired (signal) B meson decay mode and a similar decay mode obscuring the signal one, called the background. Considered are the $B_s^0 \rightarrow (J/\psi \rightarrow \mu^+\mu^-)(\phi \rightarrow K^+K^-)$ mode as the signal and the $D^0 \rightarrow \pi^+\pi^-\pi^+\pi^-$ as the background. The other classifier is dedicated to separating different kinds of jets identified during reconstruction.

For the $B_s^0 \rightarrow J/\psi \phi$ classifier, two simulated data samples of 10 000 events each are generated with the RapidSim [125] package. Then a classifier based on the Keras [119, 126] framework is trained to distinguish between the signal and background. The Keras model is based on a one-dimensional, locally connected convolutional applied at the input. Locally connected convolutional layers do not share the weights between different filters, meaning each filter can easily learn features (or correlations between separate variables) of different types. This choice is necessary because there is no reason to expect the same feature will be presented between different combinations of inputs. A global pooling layer follows to downsample the convolutional layer and improve *local translational invariance* of features in the first hidden layer. Translational invariance means that the neural network will still be able to identify the features even if the input is translated. While local translational invariance means that subtle local changes to parts of the input will not prevent the neural network to still work as expected. Finally, a few dense (fully-connected) layers are put in place to reduce the dimensionality to the desired output. The output is a single probability a given input is either signal or not. The exact implementation looks like [1]:

```

1  #!/bin/environment python
2  ...
3  classifier = Sequential()
4  classifier.add(
5      LocallyConnected1D(
6          filters = 90,
7          kernel_size = 2,
8          activation = 'sigmoid',
9          input_shape = (len(sigTrain[0]),1)
10     )
11 )
12 classifier.add(GlobalMaxPooling1D())
13 classifier.add(Dense(30, activation = 'sigmoid'))
14 classifier.add(Dense(1, activation = 'sigmoid'))
15 classifier.compile(
16     optimizer = 'adam',
17     loss = 'binary_crossentropy',
18     metrics = ['accuracy']
19 )

```

Listing 4.1 $B_s^0 \rightarrow (J/\psi \rightarrow \mu^+ \mu^-)(\phi \rightarrow K^+ K^-)$ Keras classifier implementation

The training variables are related to the kinematic properties of the signal and background. The inputs to the network include transverse momentum of the B_s^0 , $p_T(B_s^0)$, its pseudorapidity, $\eta(B_s^0)$, the minimum and maximum of the final state particles transverse momentum and pseudorapidity, $\min p_T(\mu^\pm, K^\pm, \pi^+ \pi^-)$, $\min \eta(\mu^\pm, K^\pm, \pi^+ \pi^-)$, $\max p_T(\mu^\pm, K^\pm, \pi^+ \pi^-)$, and $\max \eta(\mu^\pm, K^\pm, \pi^+ \pi^-)$. The background, of course, only has pions as final state particles, so all combinations

of pion p_T and η are injected too. Furthermore, the locally connected layer combined with the pooling layer also takes care of any combinatorial differences. Input distributions for signal and background are shown in Figure 4.1.

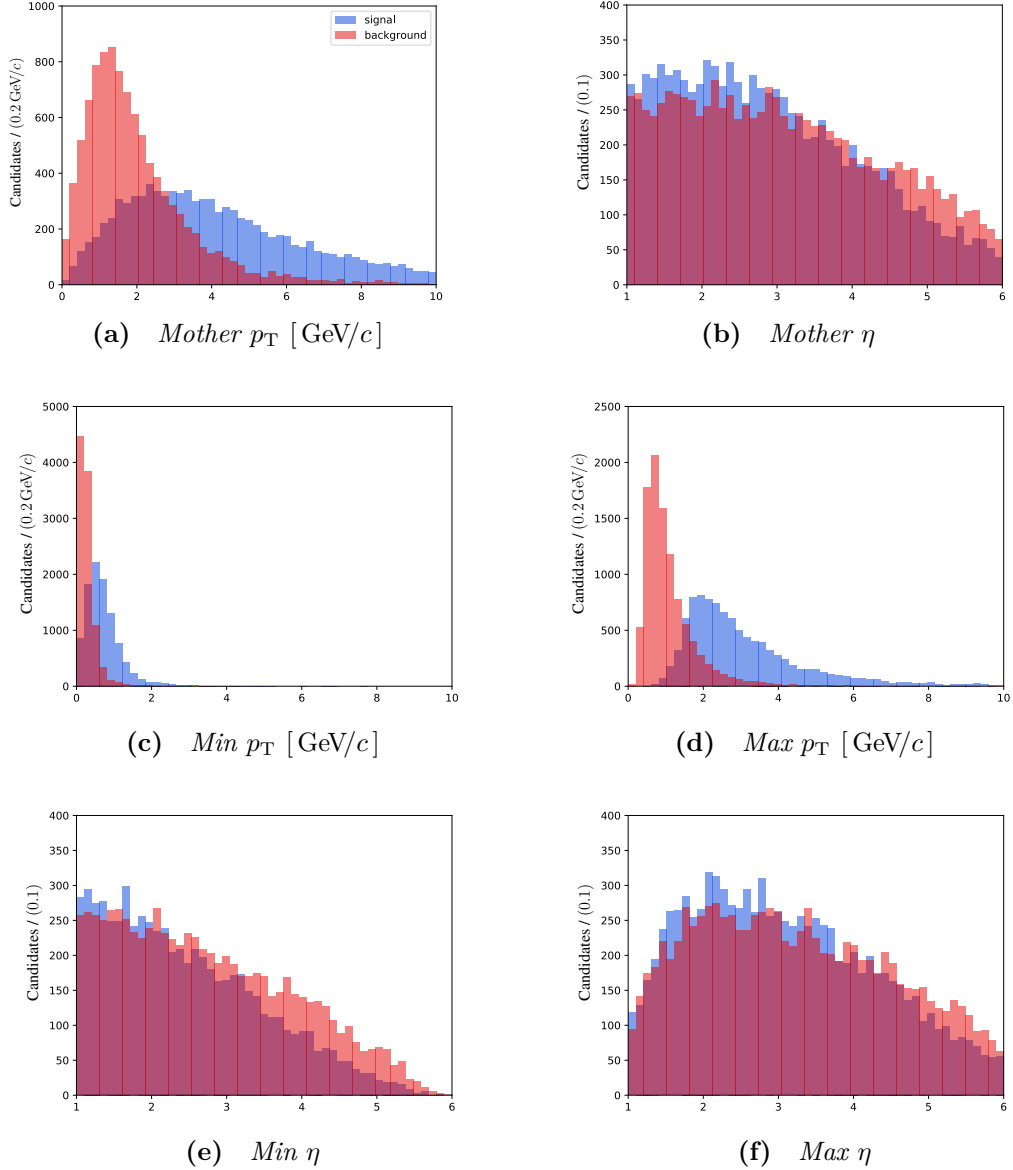


Figure 4.1 *Input distributions for simulated signal (blue) and background (red) events used to train the $B_s^0 \rightarrow J/\psi \phi$ Keras classifier [1].*

The simulated samples are split into training and testing sets to train the $B_s^0 \rightarrow J/\psi \phi$ classifier. The training and testing sets contain half of each simulated sample for signal and background with the corresponding labels. The testing sample is then used to evaluate the model for overtraining and variance.

The jet separation classifier, on the other hand, uses a different data sample. Events are simulated using the PYTHIA [127] event generator to produce pp collisions at a centre-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$. The resulting jets are reconstructed using the Rivet [128] analysis framework and then identified using the FastJet [129] package using the K_T algorithm [130]. The exact definitions of the K_T variable and the jet reconstruction algorithms can be found in References [131] and [132], respectively. Apart from a requirement of minimum jet $p_T > 20 \text{ GeV}/c$, all other parameters are the default given Rivet version 2.5.4 [1]. The signal and background samples are taken from Rivet analyses examples MC_WJETS and MC_QCD corresponding to $qg \rightarrow Wq$ and $gg \rightarrow gg$, respectively. The background rejection of the latter types of jets is of particular interest in many analyses [133].

The classifier to separate the jets is constructed in a similar way to the one for $B_s^0 \rightarrow J/\psi \phi$ decay [1]:

```

1  #!/bin/environment python
2  ...
3  classifier = Sequential()
4  classifier.add(LocallyConnected1D(filters = 32, kernel_size = 2,
5      activation = 'relu', input_shape = sig_data[0].shape))
6  classifier.add(MaxPooling1D(pool_size = 3, strides = 1))
7  classifier.add(Dropout(0.25))
8  classifier.add(Conv1D(filters = 32, kernel_size = 3,
9      activation = 'relu'))
10 classifier.add(MaxPooling1D(pool_size = 2, strides = 2))
11 classifier.add(Dropout(0.25))
12 classifier.add(Conv1D(filters = 32, kernel_size = 2,
13     activation = 'relu'))
14 classifier.add(MaxPooling1D(pool_size = 2, strides = 2))
15 classifier.add(Dropout(0.5))
16 classifier.add(Flatten())
17 classifier.add(Dense(50, activation = 'relu'))
18 classifier.add(Dense(1, activation = 'sigmoid'))
19 classifier.compile(optimizer = 'adam',
20     loss = 'binary_crossentropy', metrics = ['accuracy'])
21 earlystop = EarlyStopping(patience = 3)
22 model.fit(setTrain, labels, batch_size = batchSize,
23     epochs = epochNum, validation_data = (setTest, labels),
24     callbacks = [earlystop])

```

Listing 4.2 *Jet separation Keras classifier implementation*

The training variables are chosen based on measured quantities on each jet. There are a total of 17 input variables to the classifier. These include azimuthal angle, ϕ , pseudorapidity, η , of the jet, the spread of neutral and hadronic contributions to the jet in ϕ and η , and the average and energy-weighted kinematic variables [1]. The comparison of the signal and background distributions are

shown in Figures 4.2, 4.3 and 4.4. The training is performed equivalently as for the B_s^0 decay classifier. Finally, two drones, one for each of the $B_s^0 \rightarrow J/\psi \phi$ and jet separation models, are trained following the procedures outlined in Sections 4.3.1 and 4.3.2. Each drone is trained against its corresponding original model for 1500 epochs using a stochastic gradient descent method with a learning rate of 0.05 [1]. The threshold value, κ (Equation 4.2), is chosen to be 0.02, while the nuisance parameters⁷ b and m are set to 0.04 and 50, respectively [1].

The drone loss and the loss convergence as functions of epoch are shown in Figure 4.5 and Figure 4.6, respectively. The latter shows the difference between the current loss and the loss of the previous epoch. Furthermore, the epochs which trigger a drone mutation are clearly labelled. While the drone mutation, and thus hyperparameter extension, was triggered only ten times for the jet separation model vs thirty-eight times for the $B_s^0 \rightarrow J/\psi \phi$ model, the total parameters in the final drones are 121 and 286, respectively. This variance in mutation occurrence is mainly because of input space differences.

In conclusion, the training algorithm shows it can adequately adapt the drone network regardless of input dimensionality and is more in tune with maintaining performance and accuracy than mutation. Moreover, it can be seen in Figure 4.6 that additional hyperparameters do indeed allow the drone to learn faster after mutation. This indicates the network architecture is suitable for the approximation of the input model. The performance of the drones is compared to the original classifiers in Figure 4.7. In both examples, the drone is able to produce the desired accuracy using smaller hyperparameter space and by extension potentially less computing resources. Thus, the neural network architecture and mutation training algorithm show promise towards being directly pertinent to a wide class of high-energy-physics applications.

⁷Section C.3

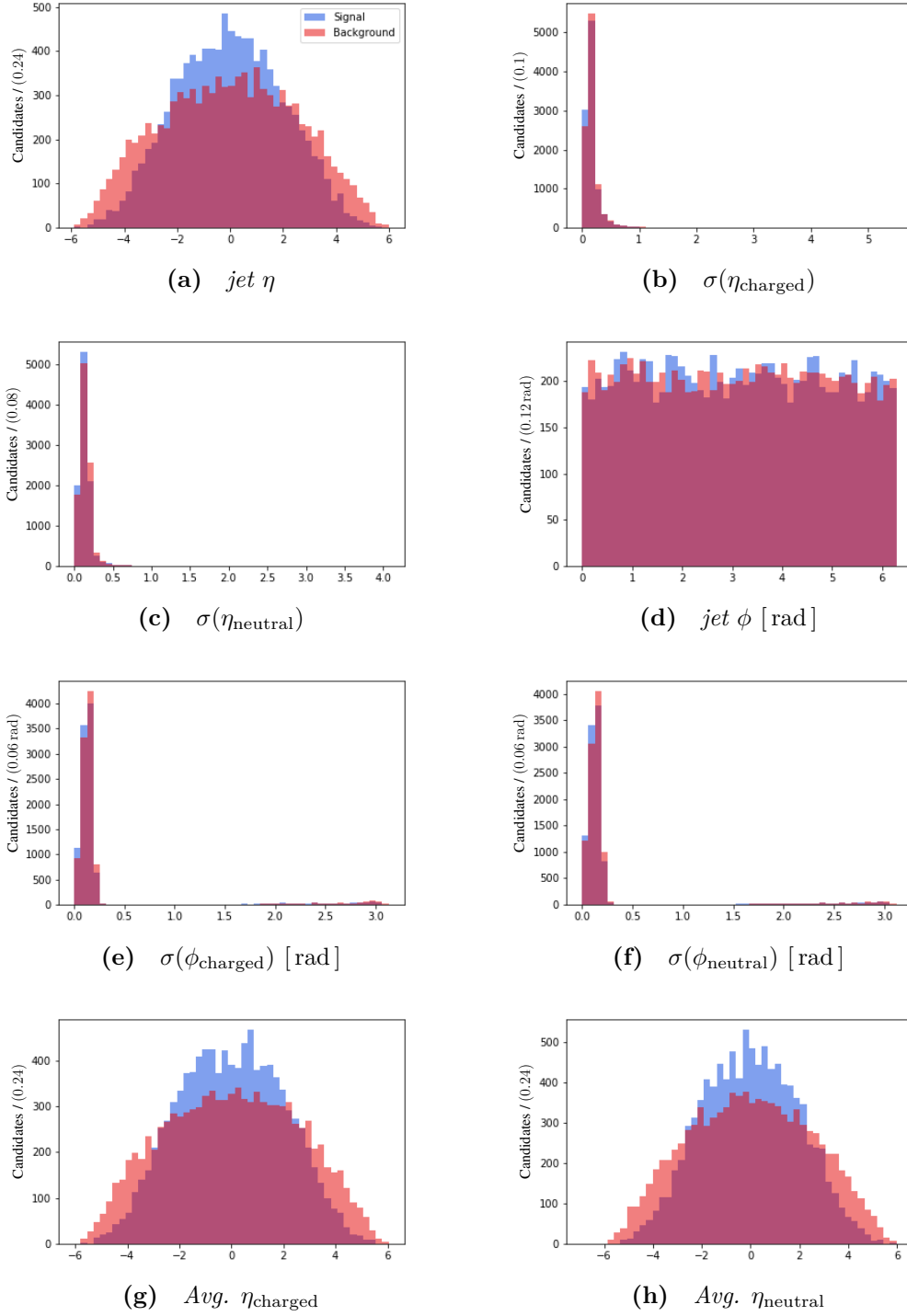


Figure 4.2 *Input distributions for simulated signal (blue) and background (red) events used to train the Keras jet-separation classifier [1].*

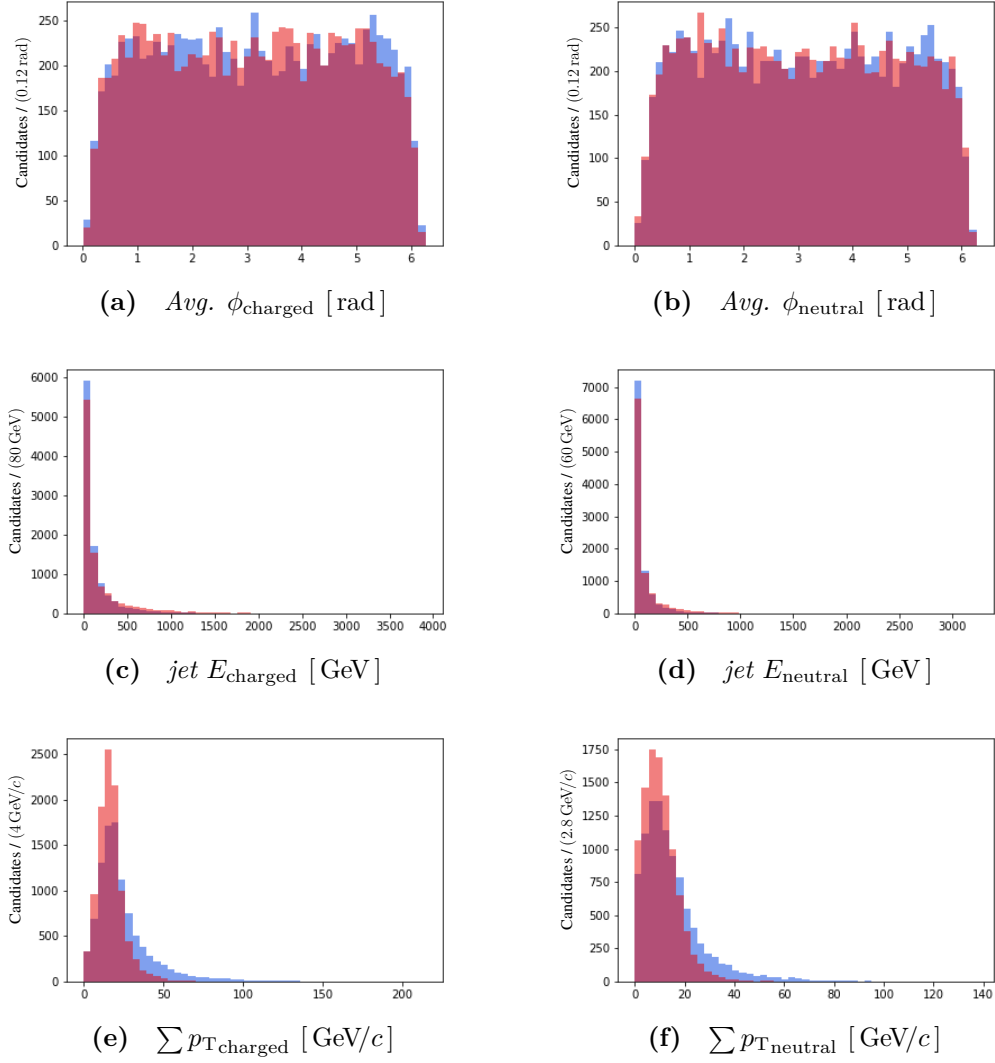


Figure 4.3 *Input distributions for simulated signal (blue) and background (red) events used to train the Keras jet-separation classifier [1].*

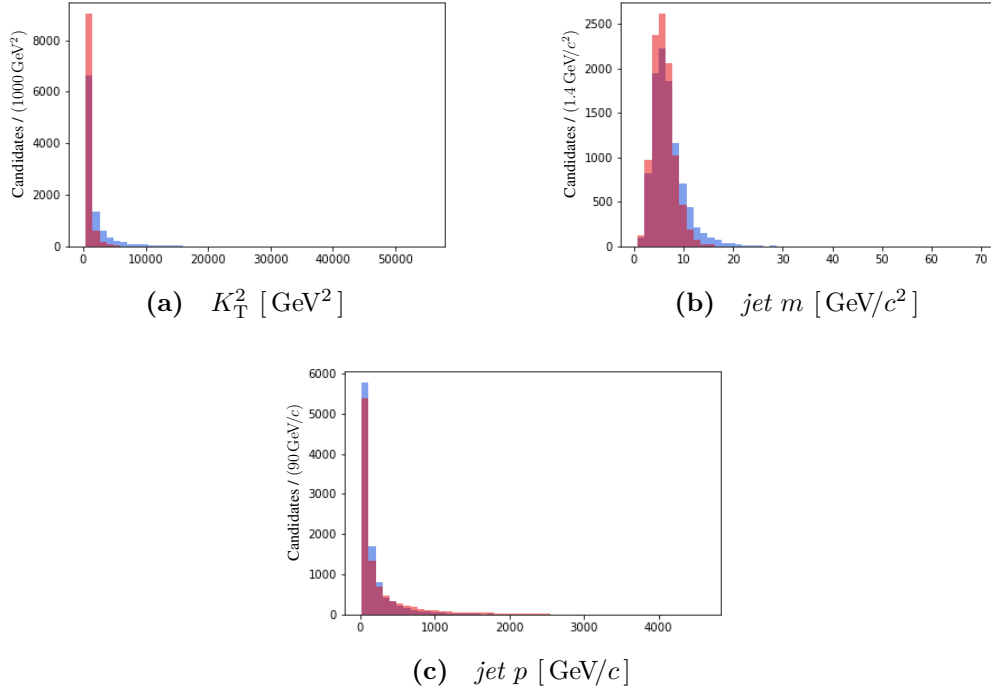


Figure 4.4 *Input distributions for simulated signal (blue) and background (red) events used to train the Keras jet-separation classifier [1].*

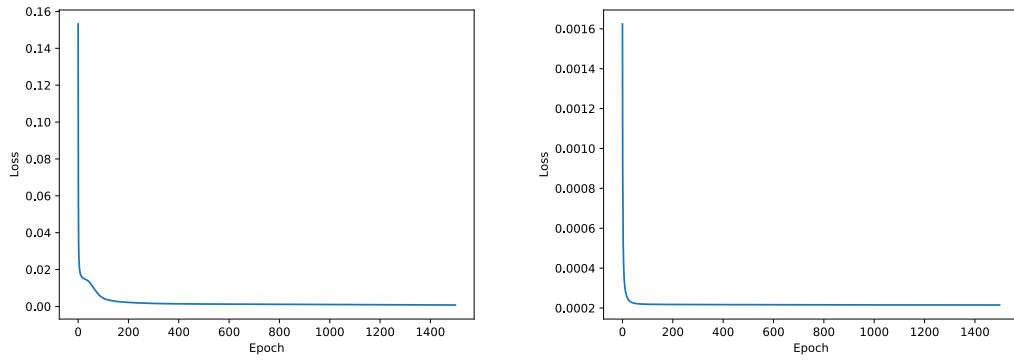


Figure 4.5 *Drone loss as a function of training epoch for the $B_s^0 \rightarrow J/\psi \phi$ decay (left) and jet separation (right). [1]*

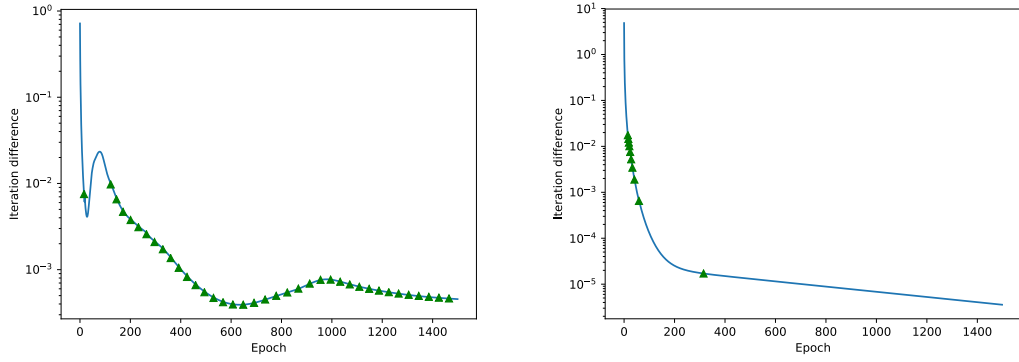


Figure 4.6 Drone loss convergence as a function of training epoch for the $B_s^0 \rightarrow J/\psi \phi$ decay (left) and jet separation (right). The green triangles show the epochs at which the drone was mutated. [1]

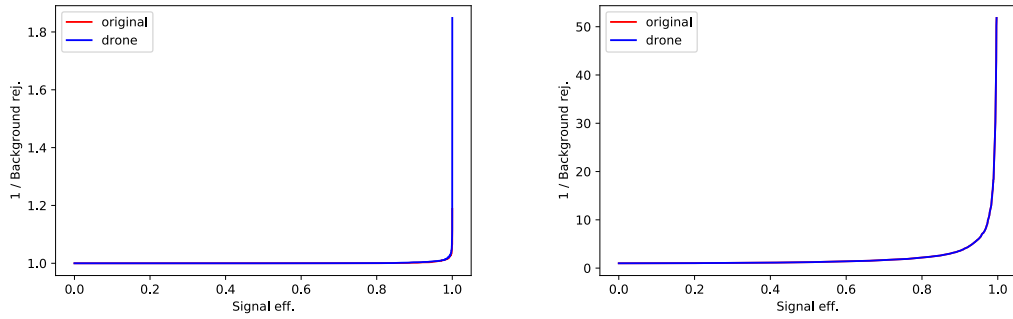


Figure 4.7 Signal efficiency vs background rejection comparison between the drone (blue) and original model (red) for the $B_s^0 \rightarrow J/\psi \phi$ (left) and jet separation (right). [1]

4.3.4 Transferability to low latency environments

For the framework to fulfil its purpose, the trained drones must be portable. This means there should be an easy way to store the hyperparameters and the relevant neural network structure. For this purpose, the JSON format was selected as storage and mediator. Firstly, JSON is human-readable. It is also natively supported and easily accessible in PYTHON, the preferred programming language interface for many analysts. Furthermore, an already available C++ framework, `lwtnn` [134], makes it easily pluggable into modern high-performance, low-latency production environments.

	original model	drone
$B_s^0 \rightarrow J/\psi \phi$	4111	121
jet separation	7081	286

Table 4.2 *Hyperparameter number comparison between original classifier and drone [1].*

Thus, a tool is provided to export a trained drone and save it to a JSON file that preserves all the information needed to reconstruct the drone elsewhere: input and output structure, layers, layer types, activation, hyperparameters, architecture, *etc.* Moreover, a separate tool is provided implemented using the GAUDI framework directly as part of LHCb core software. It can read in the information about the drone stored in the JSON file and then reconstruct it as a C++ class object based on the Keras model [1]. This tool also comes with a flexible class member structure that implements various layer types, activation and loss functions, allowing virtually any drone architecture to be reproduced. Furthermore, it can be just as easily put forth as a low-latency trigger algorithm or a high-precision later-stage reconstruction and analysis tool.

	original model	drone
$B_s^0 \rightarrow J/\psi \phi$	$3.87 \times 10^{-4} \text{ s}$	$4.8 \times 10^{-5} \text{ s}$
jet separation	$4.79 \times 10^{-4} \text{ s}$	$6.2 \times 10^{-5} \text{ s}$

Table 4.3 *Processing time comparison between original classifier and drone [1].*

Considering the identical performance of the drone and original classifier (Figure 4.7), it might be interesting to see if the goal of reducing latency has been achieved. For this, the drones are tested against the original classifiers using

a modern desktop workstation using an Intel Core i7-4770 processor. The results are shown in Table 4.2. The evaluation time is reduced by order of magnitude without compromising accuracy.

4.3.5 NNDrone at LHCb

A further test of performance is done using the LHCb HLT2 algorithms. A pre-trained drone is deployed at HLT2 in search of $B_s^0 \rightarrow \gamma\gamma$ decays and compete with pre-existing such algorithms [135]. Three separate trigger lines are created for each type of γ reconstruction possibility:

- `Hlt2RadiativeB2GammaGamma` for $B_s^0 \rightarrow \gamma\gamma$ without converted electrons
- `Hlt2RadiativeB2GammaGammaLL` for $B_s^0 \rightarrow \gamma(\gamma \rightarrow e^+e^-)$ with converted electrons as a long track
- `Hlt2RadiativeB2GammaGammaDD` for $B_s^0 \rightarrow \gamma(\gamma \rightarrow e^+e^-)$ with converted electrons as a downstream track

If both photons are detected as converted electrons of either type, they can be identified by logical combinations of the above trigger lines.

The old trigger lines are removed and replaced with the new ones based on the tool discussed in Section 4.3.4. Then a test is performed to judge the inclusive and exclusive rates in terms of event yields per second (Hz). The results are shown in Table 4.4. Moreover, the execution time of the trigger lines is reduced compared

***	Line	Incl. [kHz]	Excl. [kHz]
049	<code>Hlt2RadiativeB2GammaGamma</code>	0.01	0.01
022	<code>Hlt2RadiativeB2GammaGammaLL</code>	0.02	0.02
008	<code>Hlt2RadiativeB2GammaGammaDD</code>	0.05	0.05

Table 4.4 *Statistics from Moore test on HLT2 [136]. Uncertainties are not quotes as values are indicative.*

to using the new algorithms from 3.665 s to 0.011 s [136], and the selection yields are increased due to improved signal efficiency (to $\approx 75\%$ [136]).

4.3.6 Results

It has been demonstrated in the previous sections that for the case of high energy physics, particularly LHCb, the **NNDrone** toolkit can accurately approximate and learn the features of a neural network with a different structure and improve upon its execution performance [1]. Furthermore, the proposed method allows the drone networks to learn without accessing the original training data.

It is also shown that the equivalence of the drone to the original model allows an analyst to treat both similarly and apply the drone network easily in any environment, be it as a toy model or in the high-performance production environment of the LHCb trigger.

Chapter 5

The LHCb Upgrade of the Ring-imaging Čherenkov (RICH) system

As already briefly mentioned, particle identification (PID) is crucial for many physics analyses undertaken at LHCb. PID is specifically necessary for flavour-physics studies and their success. At LHCb, the hadron identification, especially differentiating between kaons and pions, is provided by the Ring-imaging Čherenkov (RICH) system (Section 3.4.1). It is pivotal for isolating different charmless b hadron decays and helping LHCb achieve the first observation of CP violation in the B_s^0 system [137]. PID is also mandatory in flavour-tagging, which is necessary for time-dependent CP violation studies [20], such as the one described in Chapters 6 and 7.

Chapter 4 briefly mentioned the LHCb upgrade and its transition to entirely remove the L0 hardware trigger and increase the full detector readout frequency to 40 MHz. However, the RICH system during Run 1 and Run 2 at LHCb was equipped with hybrid photon detectors (HPDs), which have a fixed 1 MHz readout rate in the electronics encapsulated within their tube. Thus, the RICH subdetector readout will have to be upgraded, and its photon detectors would have to be replaced entirely.

5.1 Introduction

The LHCb RICH detector will be upgraded during the LHCb Upgrade 1a to satisfy the new operating requirements [2]. The HPDs will be replaced with Multi Anode Photomultiplier Tubes (MaPMTs) with external readout electronics [20]. The entire upstream RICH detector (RICH 1) will be equipped with the Hamamatsu¹ R13742 MaPMT (64 channels, a custom modification of the R11265 sensor) [138]. The central high-occupancy area of the downstream RICH detector (RICH 2) will also be equipped with the R13742 MaPMT. However, the Hamamatsu R13743 (64 channels, a custom modification of the R12699 sensor) [138] will accommodate the low occupancy regions as they require lower granularity. As a result, some 3100 units of R13742 and 450 units of R13743 will be used in the upgraded RICH system (including spares). The Photon Detector Quality Assurance (PDQA) programme was created to ensure the operation qualities of the new MaPMTs. It extended over two years, between 2016 and 2018, with the main aim of characterising the MaPMT units. Its purpose is to assure minimum specifications, gather initial calibration variables, and pre-select tubes with similar characteristics to be grouped inside the RICH detectors to optimise their performance [2].

5.2 Photon Detector Quality Assurance

The LHCb RICH PDQA programme focused on the properties of MaPMT, such as PMT gain, dark-count rate, PMT channel cross-talk, PMT signal photoelectron thresholds and other such relevant properties. As such, it had high requirements for reliable and reproducible testing. Also, testing consistency and speed were key considerations since the programme had a tight schedule and time constraints. About 4000 units needed to be tested in two years. So, to complete it in time, at minimum $4000/(365 \times 2) \approx 5$ MaPMTs need to be tested each day, excluding test failures, holidays, defective or missing units and other factors, like human error. As such, it was determined that significantly more units need to be tested per day. Not only that, but the campaign ran in two sites: in Edinburgh, UK, and Padova, Italy [2]. Finally, it also required automating the testing procedures to standardise the tests and minimise the human factor.

¹Hamamatsu Photonics: <http://www.hamamatsu.com>

5.2.1 Custom test bench

For the above reasons, a custom test bench was created to characterise the MaPMTs. It used a custom readout chain based on the MAROC3² [139] platform, developed by Omega Microelectronics³. It is capable of integrating and discriminating the charge pulses of 64 MaPMT channels in parallel. As such, it suits the MaPMTs exactly. Using the data collected by the MAROC3, the signal spectra of the MaPMTs can be obtained using a multiplexed charge measurement. This method allows the extraction of qualities such as MaPMT gain, loss in efficiency, and dark-count rate and quantifies the single-photon detection capability of each of the MaPMT channels [2, 140]. Information from the MAROC3 DAQ chip is extracted via a custom digital readout board based on the Chimaera Reconfigurable Functional Unit [141, 142] specification. The Chimaera-based board is equipped with a configurable FPGA⁴ and a configurable clock, also used as a common trigger for device synchronisation. This digital readout board is attached to the MAROC3 and translates data to a PC over the network using a custom data format, all instrumented using the FPGA's custom

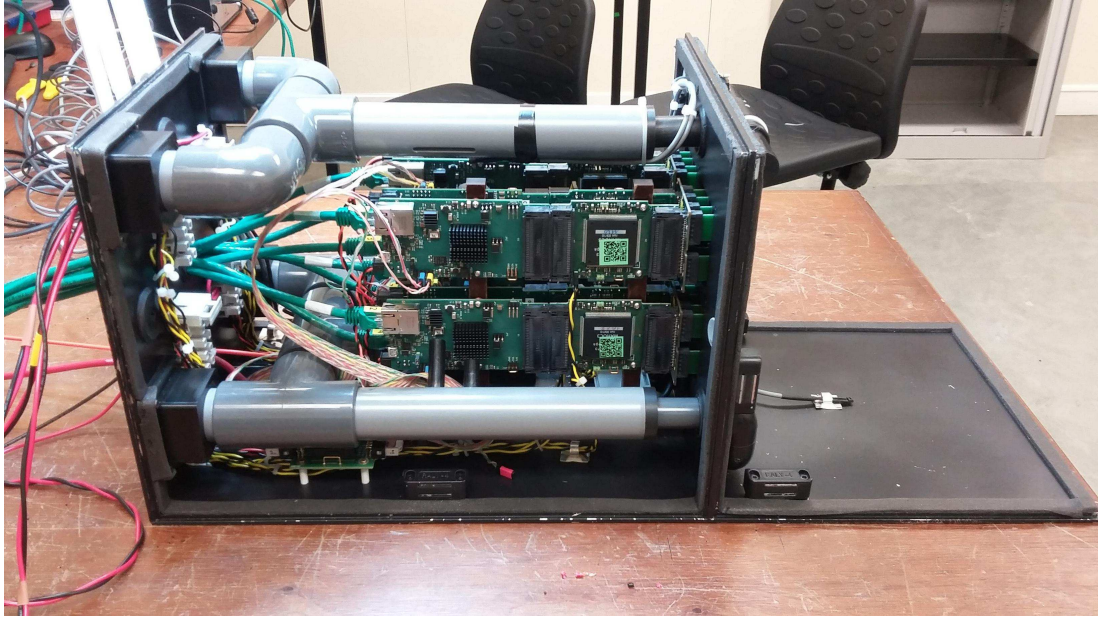


Figure 5.1 *PDQA test box with fully integrated front-end readout, data acquisition, cooling and environment monitoring. [2]*

firmware.

²MAROC3: Multi-Anode ReadOut Chip version 3

³omega.in2p3.fr

⁴Field-programmable gate array

Moreover, a custom *test box* was created to contain the electronics and provide a standardised setup between different testing labs. The test box comprises two compartments, separated by a light-tight wall [2]. The back compartment houses the front-end electronics and the digital readout. The front chamber contains the mounting points, the voltage divider chain, and the mounted MaPMTs themselves. The test box is equipped with a cooling system because temperature fluctuations affect the operation of MaPMTs [138]. The temperature inside the box is kept stable at around 25 °C by forced-air cooling, which is compatible with the expected conditions in the final detector [2]. The test box is shown in Figure 5.1.

The test is conducted by illuminating a mirror, placed at the back of the front compartment of the test box, by using optical fibre and a pulsed LED as a light source. The light is reflected from the mirror onto the array of MaPMTs to increase the illumination uniformity. The actual light source consists of two separate pulsed LEDs (470 nm wavelength) coupled with two separate optical fibres conducting the light pulses to two oppositely-mounted ends in the test box’s front compartment [2]. The light pulse duration is set at 10 ns, such that it is much faster than the common system clock, ensuring photon propagation and detection within the clock cycle. The fibres have a numerical aperture of $\approx 11^\circ$ and, thus, generate a cone of light on the flat mirror inside the test box. The relevant geometry ensures the photocathode plane of the mounted MaPMTs is provided with uniform illumination.

Furthermore, a separate control unit, called *slow control*, is created using a customised Aria-G25 FPGA⁵. It provides and orchestrates the high voltage (HV) supplied to the MaPMTs voltage divider. It also delivers and adjusts the low voltage (LV) to the readout electronics and a separate LV line modifying the light intensity of the LEDs. The slow control monitors these parameters, including the temperature and humidity sensors in the test box. However, the power supplies driving the slow control and other relevant components are controlled by the central system described in Section 5.2.3.

⁵<https://www.acmesystems.it/aria>

5.2.2 Goals

The test setup is designed to test $16 \times \text{R13742}$ or $4 \times \text{R13743}$ sensors in parallel depending on the exact test box configuration in the front compartment [2]. The PDQA procedure aims to verify the specification provided by Hamamatsu for each MaPMT [140]:

- average gain: $> 10^6 \text{ e}$
- dark-count rate: $< 16 \text{ kHz}$
- gain uniformity (gain variation between lowest and highest):
 - 1 : 4 for R13742
 - 1 : 3 for R13743
- peak-to-valley ratio: ≤ 3 channels with a ratio $< 3 : 1$ for single photons

The gain is defined as the converted charge response in the PMT to a signal photon arriving at the anode of the channel. The dark-count rate is defined as the frequency of responses a PMT channel produces while active in a non-illuminated environment — essentially, a measure of background noise due to spontaneous excitations of the anode. Gain uniformity measures the spread of gain values of all channels in a PMT. Finally, the peak-to-valley ratio is a measure of the separation between the channel response to signal photons and its response to background excitations. This is clearly visible in Figure 5.2, the dark-count excitations have a defined response value in the highest peak to the left, while the single-photon response peak is separated to the right, producing typically higher converted charge. The ratio between the height of the single-photon peak and the height of the valley is measured to represent how well a particular PMT channel can separate background from signal.

Furthermore, long periods of operation in the LHCb environment will degrade the MaPMT performance over the years [2]. This ageing will result in a gradual loss of gain. One strategy to compensate for this is to increase the supplied HV to counteract the gain-loss systematically. Therefore, an essential part of PDQA testing is to study the gain response for MaPMTs as a function of HV. This study will also determine the minimum required HV needed to achieve single-photon detection.

Finally, the PDQA test protocol is set to be executed in the following order [2]:

- MaPMTs are mounted in the test box in the late afternoon
- test box is closed, and MaPMTs are left to settle in the dark at nominal HV= 1000 V overnight
- a burn-in test is performed at maximum HV= 1100 V
- data collection at different HV values starting from HV= 1100 V down to HV= 850 V in steps of 50 V

The burn-in test lasts about an hour. This test is necessary to make sure PMTs will continue to operate normally as their HV is varied to account for gain losses. There is also a settling period for each new HV setting to counteract any excitation fluctuations in the anode. A dark-count rate measurement is performed at regular intervals during the overnight settling of MaPMTs, each HV setting, and one after the whole procedure. This is done to ensure the continued performance of the MaPMTs is stable and different operating modes do not introduce more noise in PMT channels. Temperature and humidity are recorded every 5 seconds. That relates PMT response to operating conditions.

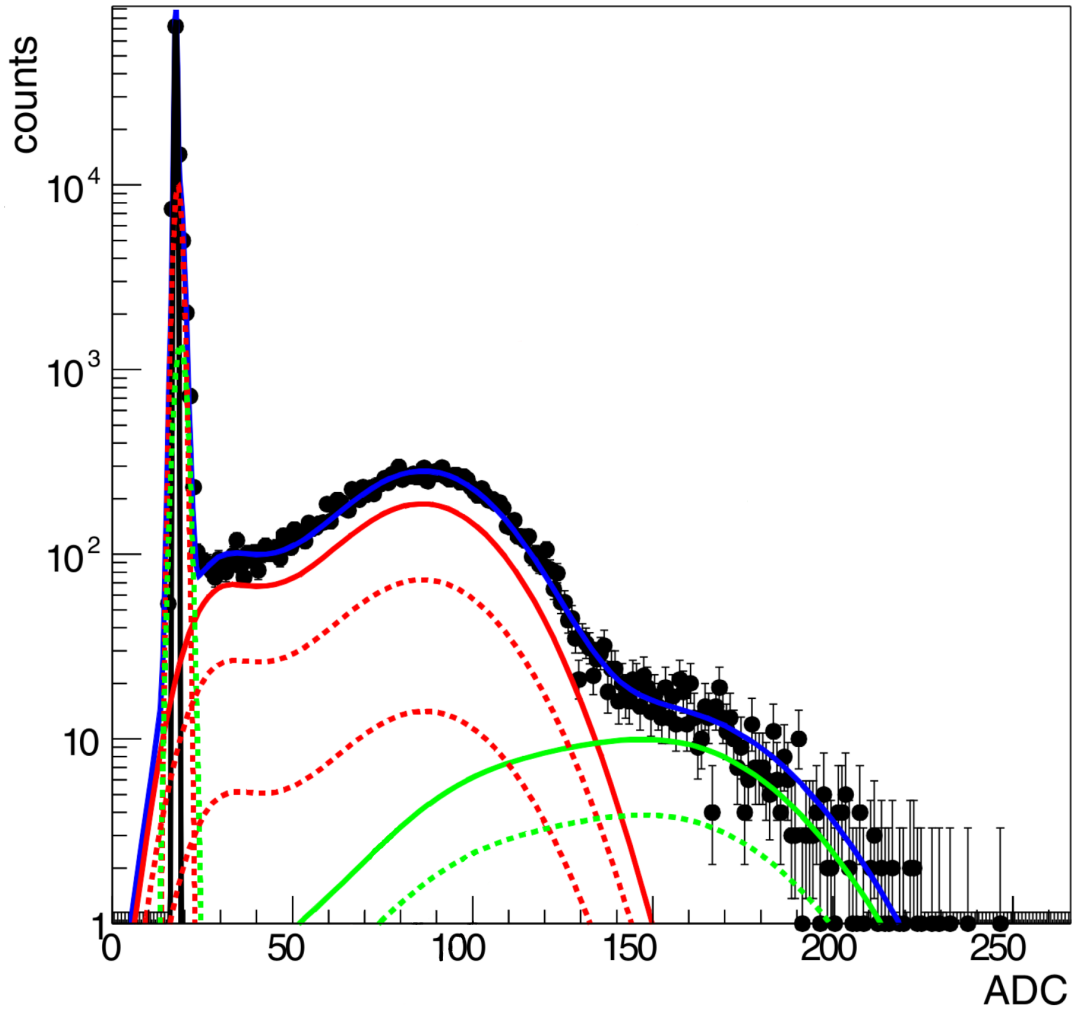


Figure 5.2 *Single-channel response of a R13742 PMT [143]. Showing the analog-to-digital conversion (ADC) response of the PMT. The highest peak to the left is the dark-count rate, the broad peak immediately to the right (red solid line) is the single-photon response. The single-photon response includes an ad-hoc model describing the charge loss at the dynode (lower bump in the broad peak). The size and depth of the valley between the dark-count and single-photon peaks is an indication of the separation of signal and background response. This is measured as the peak-to-valley ratio in the PDQA procedure. The green solid line models the double-photon response. The dashed lines model the cross-talk between neighbouring channels of the PMT.*

5.2.3 System automation

The number of units to be validated, the number of measurements, their complexity and tight time constraints necessitate very slim margins for error. To achieve this and to be able to run tests continuously over twenty-four hours day after day, a high level of automation was required.

Control software was developed to carry out the PDQA programme testing procedure with a requirement of minimal manual intervention [2]. It also included a robust system for handling data acquisition, MaPMT spectra fitting, speeding up results interpretation and measurement validation [2].

The full PDQA test station consists of the test box (Section 5.2.1), LED driver, slow control, power supplies and a workstation computer [2]. The control software needs to be able to operate the test station fully and autonomously for long periods to satisfy the requirements for reliability, reproducibility, and consistency. Furthermore, the software should monitor the environment, and dependent on conditions, prepare and perform each test independently. Thus, it should make simple decisions and take appropriate steps to complete the testing procedure without human operator intervention. Following this train of thought, there is an additional requirement that the system as a whole should be able to cope with most error conditions, then try to recover as best as possible and continue normal operation.

The required logical structure becomes quite intricate and quickly becomes impractical to implement, considering the total complexity of such a problem, the separate devices involved, their configuration and control, and their mutual interaction. A possible solution to this problem is to simplify the space of possibilities by making the correct assumptions and factoring out what is possible into smaller manageable components.

5.2.4 Microservices

For example, conventional monolithic type software (Figure 5.3) suffer from trying to take care of everything altogether as a single entity. They have a singular clear hierarchy that controls execution, logic and control [2]. However, that requires that the same application has the full information about the whole system, its operation and possible things that can go wrong. For complex systems

coupled with their operation in an environment where multiple things can happen simultaneously, monolithic applications quickly become overwhelmingly difficult to construct and operate. They tend to leave an increasing number of mishandled edge cases in their logic and undefined behaviour situations. The most crucial concern for the problem is that monolithic applications cannot easily recover their state from a fatal fault⁶ nor continue the testing procedure afterwards [2].

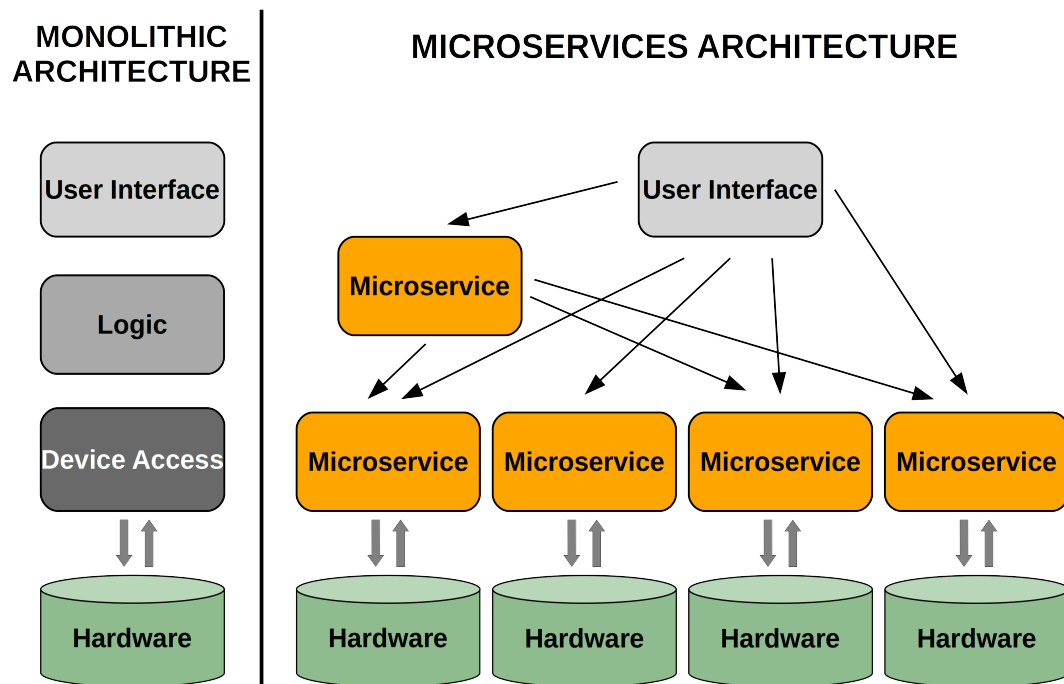


Figure 5.3 *Model representation of microservices architecture compared to a monolithic architecture. [2]*

For this reason, more modern approaches exist in the overall architecture and design of software that aim to tackle these problems. Specifically, it was chosen to accept the paradigms of the so-called Microservices [144] architecture. This architecture splits the monolithic software application into smaller independent and mostly autonomous units, called microservices. These have minimal responsibilities compared to the single application in the monolithic case. Each microservice executes a task, has its logic structure, but all microservices are connected over the same communication protocol (Figure 5.3) [2]. In a sense, a microservice provides only a single *service* that the rest of the units can utilise.

⁶A fatal fault is an error that causes a program to abort its execution and return control to the computer's operating system.

Other units can poll the state of the service in question, request its start or termination, get updates or results.

For example, a service can be given the responsibility of controlling the high voltage supply. Thus, it should know everything there is to know about the HV, but nothing else. Other microservices can talk to it and it should respond. It should provide information such as the status of the HV (on or off), the target HV value, or accommodate requests for changing HV. Also, given no new instructions, if the HV is not at the correct requested state, it should emit a warning when a service that depends on it asks about the HV. Moreover, the HV service should know how to operate the HV driver and deal with any possible faults as best it can. This whole setup limits the scope of possibilities and dramatically simplifies the implementation and operation of such applications.

5.2.5 Finite state machine

The power and simplicity of the microservices architecture can be alluring. However, there remains the problem of cooperation and achieving a common goal. To facilitate the execution of more complicated actions, still, some hierarchy must exist between the services to organise them [2].

Continuing the last example, this could be another service that knows nothing about operating the HV. Instead, it understands the conditions under which the HV must be set at a specific value and switched on or off. Thus, it follows its internal logic and polls its relevant dependent services for information until the its logic conditions are met. At this point, it asks the HV-controlling service to execute a particular task. Thus, it only needs to know how to communicate with other services, not operate the HV module itself [2].

Taking this concept further, a structure akin to a network of units can be built, capable of executing very complicated tasks while keeping the complexity of its building blocks low [2]. Additionally, this inherently grants multithreaded capabilities and simultaneous operation of multiple concurrent tasks as part of the design without any extra effort needed [2]. Ultimately, it addresses one of the main concerns about continued uninterrupted operation and recovery of the system. Individual units do not store all the information about the whole system and can, thus, fail, abort or be restarted as required by the rest of the units. In such cases, they can safely and fully recover relevant data from other units and

the environment.

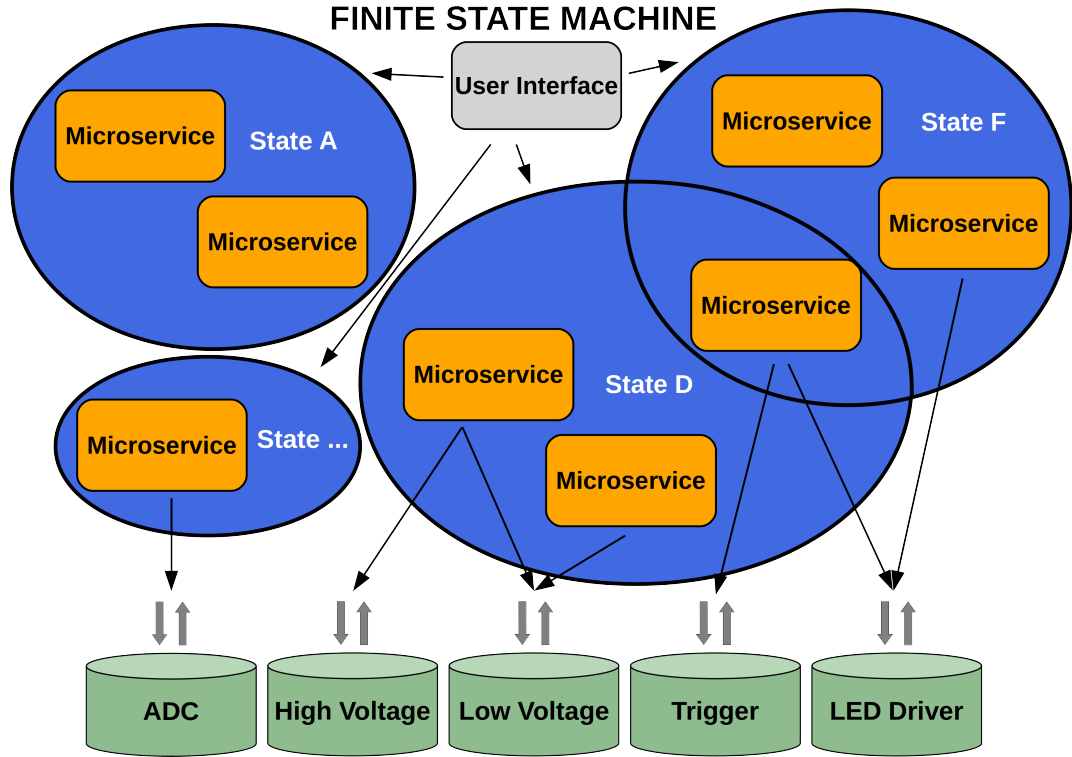


Figure 5.4 *Structure of the finite state machine automating the PDQA testing procedure. [2]*

The complete control software refines this concept by combining these premises into a finite state machine (FSM) with a graphical user interface (GUI) where each logical state is operated by a microservice [2]. The schematic diagram of how this works is shown in Figure 5.4. Then the logical-state microservices depend on other microservices to accomplish a specific task. The GUI itself is also implemented as several microservices such that it can remain responsive and accept input from or display information to the user without delay. This dramatically reduces perceived latency and means that the required tasks can be configured as the user expects reducing the likelihood of mistakes.

Moreover, the FSM can transition to executing the configured procedure if and when it has all the required details and not before. Thus, tasks can be performed safely and consistently. Furthermore, failure and recovery are guided and handled using the states of the FSM with a clear conditional structure based on state and dependencies. Finally, to make it all work together, an application programming interface (API) is designed and deployed that all microservices share and can use to communicate freely. This API can also be used to add more units, plug in

external software or gather statistics about the system.

The exact implementation of the control software uses the C++ [145] language for the lower-level and high-performance units, LABVIEW [146] for the GUI, ROOT [147] for data analysis and XML [148] for human-readable settings and logs [2].

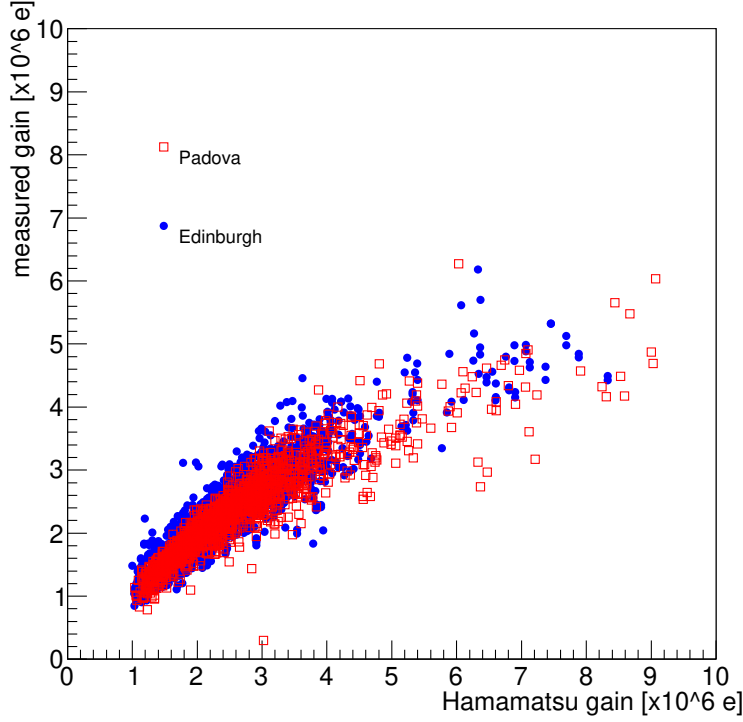


Figure 5.5 *Comparison of the R13742 (1in) PMT gain (measured at 1kV) between the Hamamatsu datasheet values and the result from the PDQA testing procedure. [2, 140]*

5.2.6 Results

The PDQA Automation [149] software was deployed in two labs and four testing stations in total and performed as expected [2].

Each station operated with minimal human intervention, limited to mounting the MaPMT units and starting the procedure [2]. The control system handled the PDQA testing procedure entirely. It performed the tests, data acquisition, spectra fit and basic analysis, allowing streamlined and standardised results. These enabled accepting or rejecting units based on comparison with the contractual

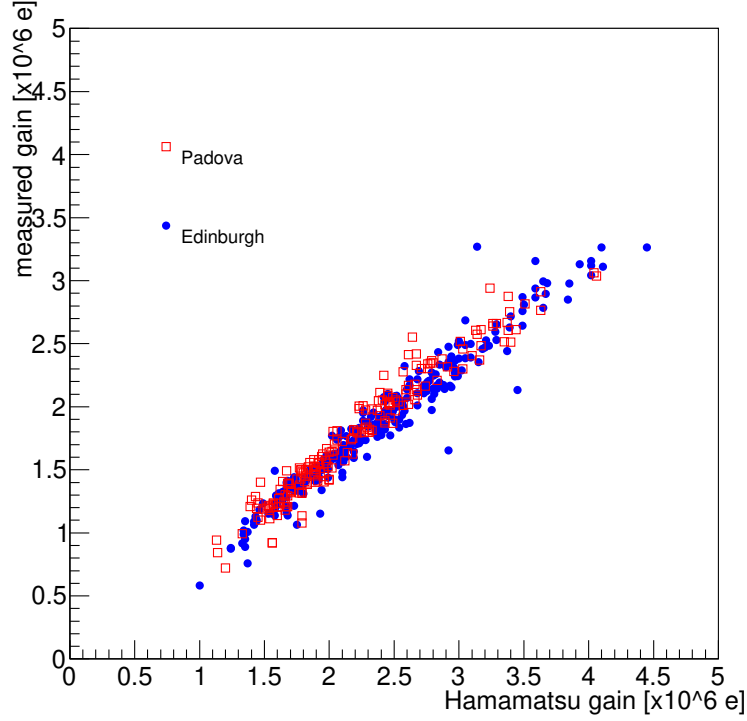


Figure 5.6 *Comparison of the R13743 (2in) PMT gain (measured at 1kV) between the Hamamatsu datasheet values and the result from the PDQA testing procedure. [2, 140]*

specifications with Hamamatsu.

A selection of the results for measured gains of the R13742 and R13743 MaPMT variants are shown in Figures 5.5 and 5.6, respectively. The measured average gain is compared to the value provided by Hamamatsu [2]. The majority of the difference between Hamamatsu and PDQA measured gain is due to the different test setup. At the factory, Hamamatsu subjects PMTs to a uniform high illumination and then measures produced current as an estimate of average PMT gain. These operating conditions are strictly different from the single-photon operation expected at LHCb and applied in these tests. In the Hamamatsu validation procedure the dynode chain is constantly excited while here long periods of darkness leave the PMT channels to settle. This difference also results in some PMTs presenting a major difference in average gain between the two environments due to dynode chain material qualities uniquely suited to one situation or the other. The other relevant metrics described in Section 5.2.2 were fully measured and validated against the manufacturer’s provided data. More details can be found in Reference [140].

There is broad agreement between the gain listed by the manufacturer and that measured in the labs, even considering the different environments of measurement. Furthermore, the results from the two separate testing facilities are consistent with each other. Finally, the requirement for average MaPMT gain to be above 10^6 e is satisfied.

Chapter 6

Simulation and event selection of $B_s^0 \rightarrow J/\psi \phi$ decays with LHC Run 2 data

6.1 Introduction

To be able to conduct the time-dependent angular analysis of the $B_s^0 \rightarrow J/\psi \phi$ decay mode and extract the value of ϕ_s , the LHCb dataset need to be properly selected for signal events. At its point of publication, the presented analysis culminated in the most precise determination of the CP -violating phase, ϕ_s . Some of the key ingredients in achieving that is taking advantage of several significant improvements in dataset distillation procedures, including a novel approach in machine-learning data selection techniques, data-driven corrections to simulated samples and the management of systematic effects using control samples are discussed.

6.2 Data and simulation samples

The analysis discussed in this chapter uses data collected by the LHCb Collaboration using the LHCb detector with the help of the computing facilities and software as described in Chapter 3. Furthermore, simulation samples are created using the software discussed in Section 3.5.3 to match the desired processes in

question. These are taken through equivalent processing steps, as would the actual collision data, to evaluate any effects on the quality of the data introduced by the detector hardware and software.

6.2.1 Dataset

The analysis uses collision data collected at the LHCb experiment at the LHC during Run 2. The dataset corresponds to a total integrated luminosity of $\int \mathcal{L} dt = 1.9 \text{ fb}^{-1}$. The collision data for the first $\int \mathcal{L} dt = 0.3 \text{ fb}^{-1}$ were gathered during 2015, while the larger set $\int \mathcal{L} dt = 1.6 \text{ fb}^{-1}$ during 2016, all at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$.

For the 2015 data, the software was configured using STRIPPING version **24r1**, while the 2016 sample was processed using STRIPPING version **28r1**. The version number, **XXrY**, signifies the major release by the **XX** part and the minor “fixes and improvements” release by the **Y** part. However, specifically relating to the decay channels discussed, both **Stripping v24r1** and **Stripping v28r1** provide completely equivalent event selection and differ primarily in unrelated areas, such as underlying implementation details, independent selections and accommodating external software.

The selection is provided by applying the constraints included in a pre-configured ruleset, a STRIPPING line, to the DIMUON stream, which comprises only data corresponding to events involving a pair of muons associated together, a dimuon pair (as discussed in Section 3.5.1). The STRIPPING lines **StrippingBetaSBs2JpsiPhiDetached**, **StrippingBetaSBd2JpsiKstDetached** and **StrippingBetaSBu2JpsiKDetached** are used to select events related to $B_s^0 \rightarrow J/\psi \phi$, $B_d^0 \rightarrow J/\psi K^{*0}$ and $B^+ \rightarrow J/\psi K^+$ decays respectively. The latter two are control samples related to decay time and angular corrections discussed further in Sections 7.3 and 7.5.

6.2.2 Simulation samples

At LHCb, simulations of pp collisions are created using the PYTHIA event generator combined with LHCb-specific configuration inside the GAUSS application [150]. The EVTGEN package simulates hadron decays while the radiative decays are handled by the PHOTOS package [151]. As discussed in Section 3.5.1,

the simulation samples go through all the steps that real data does. The difference here is that the STRIPPING algorithms can flag correlations or disparities between the true¹ and reconstructed quantities. The simulation configurations themselves are also developed and released in versions. The simulation samples discussed here use version **Sim09b**, where **SimXXy** is labelled with **XX** for the major release and **y** for the minor release.

Simulated samples can be split into several categories depending on their application. Simulated signal samples mimic how the detector and its software systems would observe and record real $B_s^0 \rightarrow J/\psi \phi$ signals. This is useful in correctly selecting relevant data for analysis and evaluating specific parameters, such as angular acceptance.

On the other hand, simulated background samples emulate how unwanted data related to other topologically very similar processes can pollute the analysis sample by appearing as $B_s^0 \rightarrow J/\psi \phi$ events. Such samples can also be used to constrain specific parameters which are more or less similar between different decay modes, but the $B_s^0 \rightarrow J/\psi \phi$ mode is not ideally suited for their determination. One such example is the decay-time detector acceptance and resolution. The STRIPPING configuration used to collect the best quality signal events also introduces a bias specific to decay time. In this case, a similar but separate decay is used to determine the decay-time observation parameters as they are primarily due to equivalent detector effects between similar decays (Sections 7.3 and 7.4).

Simulated control samples are also employed in the analysis. As the name would imply, they are mainly used for validation and cross-checking. However, there is another critically important use when it comes to this analysis. As discussed in Chapter 2, it is needed to know the flavour of the B_s^0 meson at creation to perform the ϕ_s measurement. Given that the $J/\psi \phi$ final state is its own charge conjugate, the algorithms responsible for establishing the B_s^0 flavour need to be calibrated further to the specific features of the decay in question (Section 7.6).

A breakdown of all simulated samples used is shown in Table 6.1. It is generally accepted to refer to the simulation samples as Monte Carlo (MC) samples pointing to the namesake method by which they are generated.

¹the true quantities here refer to the values generated by the simulation software

Decay mode	Options	Year	Number of events
Signal modes			
$B_s^0 \rightarrow J/\psi \phi$	dG=0,DecProdCut,S24	2015	4 M
$B_s^0 \rightarrow J/\psi \phi$	dG=0,DecProdCut,S26	2016	25 M
$B_s^0 \rightarrow J/\psi \phi$	DecProdCut,S24	2015	4 M
$B_s^0 \rightarrow J/\psi \phi$	DecProdCut,S26	2016	20 M
$B_s^0 \rightarrow J/\psi \phi$	DecProdCut,S28	2016	10 M
$B_s^0 \rightarrow J/\psi K^+ K^-$	DecProdCut,S26	2016	7 M
$B_s^0 \rightarrow J/\psi K^+ K^-$	DecProdCut,S28	2016	20 M
Background modes			
$\Lambda_b^0 \rightarrow J/\psi p K^-$	DecProdCut,S24	2015	15 M
$\Lambda_b^0 \rightarrow J/\psi p K^-$	DecProdCut,S26	2016	10 M
Inclusive J/ψ	DecProdCut,S24	2015	2 M
Inclusive J/ψ	DecProdCut,S26	2016	43 M
Control modes			
$B_d^0 \rightarrow J/\psi K^{*0}$	DecProdCut,S24	2015	10 M
$B_d^0 \rightarrow J/\psi K^{*0}$	DecProdCut,S26	2016	15 M
$B^+ \rightarrow J/\psi K^+$	DecProdCut,S24	2015	10 M
$B^+ \rightarrow J/\psi K^+$	DecProdCut,S26	2016	14 M
$B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi)\pi^+$	BcVegPy,DecProdCut,S20	2012	2 M
Inclusive J/ψ	DecProdCut,S24	2015	2 M
Inclusive J/ψ	DecProdCut,S26	2016	43 M

Table 6.1 *List of Monte-Carlo simulation samples used in the analysis. The “Options” field specifies relevant details for each sample: dG=0 means the MC was generated assuming $\Delta\Gamma_s = 0$; DecProdCut means that a generated event is only valid/saved if all of its decay products appear in the acceptance of the detector; SXX labels the major version of STRIPPING used; BcVegPy means that the generator was configured for hadronic production of B_c^+ ($gg \rightarrow B_c^{+[*]} b \bar{c}$)*

6.2.3 Momenta corrections

The data and the simulation samples are processed using the same version of the DAVINCI software package, namely DAVINCI version **v44r4**. The DAVINCI package simplifies many tasks related to selecting specific data based on certain criteria, applying initial corrections to said data and even running custom procedures on the samples.

One such correction is on the track momentum. To improve the track fitting

Parameter	Value
Δm_s	17.8 ps^{-1}
$\Delta \Gamma_s$	0.08543 ps^{-1}
Γ_s	0.6614 ps^{-1}
ϕ_s	-0.03 rad
$ A_0(0) ^2$	0.5242
$ A_{\parallel}(0) ^2$	0.2256
$ A_{\perp}(0) ^2$	0.2500
$\delta_{\parallel} - \delta_0$	3.26 rad
$\delta_{\perp} - \delta_0$	3.08 rad

Table 6.2 *Decay model parameters for Sim09b Monte-Carlo simulation samples used in the analysis.*

procedures discussed in Section 3.5, each data-taking year, LHCb uses about a third of its data to produce momentum calibration constants which are applied to the momentum scaling algorithms in DAVINCI [152, 153]. The momentum of reconstructed tracks is scaled depending on the corresponding data-taking period and specific track calibration constants. The procedure has the flexibility of adequately updating the track momentum to the latest, most accurate estimate possible after reconstruction and prior to data analysis. This has the effect of improving both invariant mass resolution and particle identification efficiency. For this analysis, momentum scaling also helps improve the angular resolution.

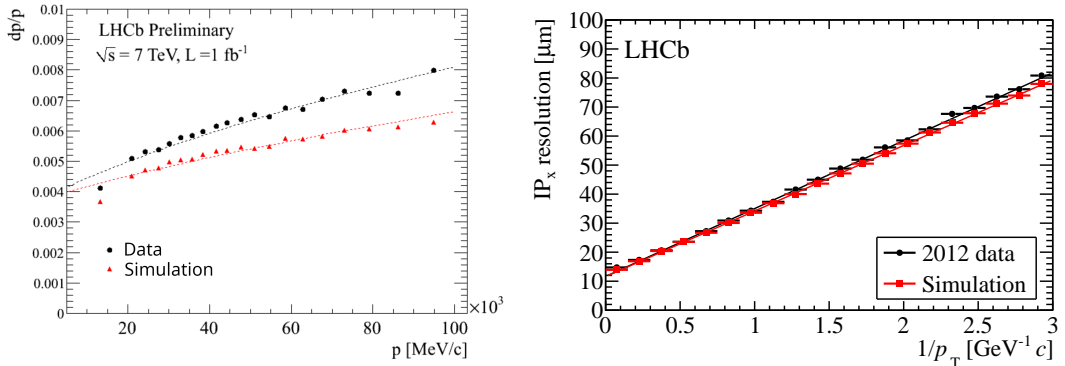


Figure 6.1 *Left: The momentum resolution as a function of momentum for data and simulation before corrections. Right: The resolution of impact parameter in the x-direction as a function of the inverse transverse momentum compared to simulation. [70, 154]*

Furthermore, simulated events usually do not perfectly reproduce the correct distribution for momentum resolution, as shown in Figure 6.1. This is because

simulated particles “interact” with a simulated detector, which is by definition better calibrated than the real one. The LHCb detector must be aligned and calibrated at the start of every run to account for many physical effects: subdetector relative spatial displacement, thermal expansion, air pressure variations, *etc.* The simulated detector is a much simpler approximation, and thus, other methods are employed to improve relative differences with data.

In this case, a technique called “momentum smearing” is applied. In simulated events, the reconstructed track momenta are “smeared” by introducing an effective Gaussian resolution factor into the track fit. The input to the track reconstruction algorithms are convolved with the resolution function, and thus, the result is more realistic. The parameters of this smearing factor are fitted to match comparable data and simulation samples. Similarly to momentum scaling, this correction can also be applied on-demand to take advantage of the latest improvements.

All samples of data and simulation in this analysis have momentum scaling and momentum smearing applied, respectively.

6.3 Event selection

The theoretical model of the $B_s^0 \rightarrow J/\psi \phi$ decay discussed in Chapter 2 can only describe a pure sample of data containing events only resulting from the decay in question. However, no experiment could ever hope to achieve such data purity due to the nature of particle colliders, specifically in this case of pp colliders. As such, the data sample collected at LHCb contains both signal and background decay candidates.

This section aims to detail the techniques used to remove as much of the data related to background processes (referred to as background henceforth) while at the same time keeping as much of the data associated with the signal decay mode (referred to as signal hereafter).

Since the amount of signal and background in the data sample is unknown, estimates have uncertainties associated with each. The theoretical model is fit to a sample where the background distribution is statistically subtracted from the total distribution. Thus, it is imperative to reduce the background content in the full distribution as much as possible to reduce the effects of inflating

the statistical uncertainties of the fit and hence the estimates of the physical parameters themselves. The theoretical background allowing this procedure is explained in Sections C.1 and C.2.

As such, this section will discuss the selections applied to the samples during the different processing stages. Firstly, online selections such as hardware (L0) and software (HLT) trigger selections are outlined, then offline selections such as the STRIPPING and later-applied more specific DAVINCI algorithms are detailed. Finally, the analysis specific selection steps which are applied lastly are also included.

6.3.1 Trigger selection

At LHCb, as described in Section 3.5.1, collision events are first required to pass a trigger [155]. The hardware L0 trigger configuration is as listed in Table 3.1, and events are required to have passed at least one of the trigger lines configured. The trigger comprises two stages: a hardware trigger (L0) based only on information from the calorimeters and muon system, followed by a software stage, which does a full event reconstruction.

L0 selection

The combination of L0 triggers must be considered before any other selection of candidate events corresponding to a $B_s^0 \rightarrow J/\psi \phi$ decay. Previous work on the topic has identified the following condition gives a good signal-to-background ratio: `(L0HadronDecision || L0MuonDecision)`, *i.e.* requiring the event to pass the hardware trigger selecting muons or hadrons with large transverse momentum [15].

However, with the increased collision energy in Run 2, the production cross-section of b-hadrons also increased by a factor of 2^\dagger [73]. This means the L0 conditions should be adapted to the new energy.

Particularly, it can be helpful to look at the distributions of relevant trigger lines' responses in data and compare them to those in signal MC simulation. For example, in Figure 6.2, a comparison is made between the distributions of the L0 triggers selecting single muons or pairs of muons (dimuons). Simulation shows a

[†]The ratio of total production cross-section of b-hadrons at $\sqrt{s} = 7$ GeV and $\sqrt{s} = 13$ GeV integrated over the acceptance of LHCb is $2.00 \pm 0.02 \pm 0.26$ [73].

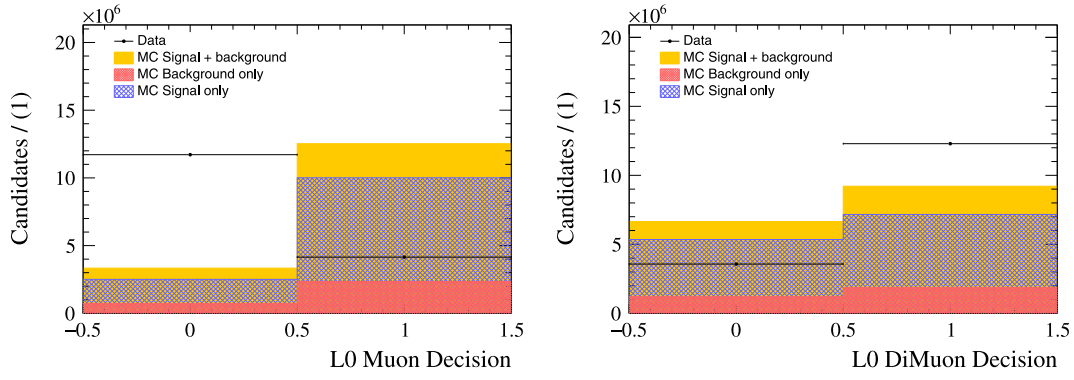


Figure 6.2 *The distribution of L0MuonDecision (left) and L0DiMuonDecision (right) trigger line in data and signal MC.*

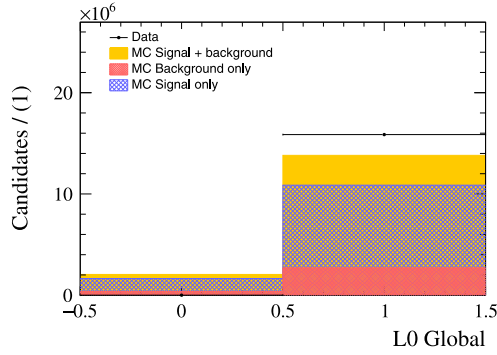


Figure 6.3 *The distribution of L0 decision for Global trigger line in data and MC. The condition is satisfied when any of the L0 individual triggers has fired for the event.*

significant number of signal events that do not trigger either category.

Thus, using either of the triggers separately will have the effect of removing a large percentage of signal candidate events, which is undesirable. The `L0MuonDecision` is only $78.97 \pm 0.03\%$ efficient on signal MC. The combination of the triggers is not much of an improvement as the efficiency of `L0DiMuonDecision` is just $58.06 \pm 0.03\%$, and it also rejects signal events that triggered other L0 lines but not the muon one, such as the hadron line.

Therefore, it was chosen to preserve as much of the signal events at the cost of background retention at this early selection stage to give a chance to more advanced techniques in later stages explained in the following sections. As such, the events are only required to pass any of the L0 triggers (pass `L0Global`).

The trigger selection, as outlined above, results in signal efficiency of $86.8 \pm 0.03\%$, which is a 7% increase in signal yield in 2016 and about 13% in 2015 with respect

to only using the information from the muon system. The difference in signal yield between the years here is due to the L0 thresholds in 2015 limiting the ingress of events as per Table 3.1.

HLT1 selection

Most of the HLT1 triggers have very low efficiency. For example, the trigger line `Hlt1MuonsJpsi2MuMuDecision` is only suitable for calibration purposes and feeds into `Hlt1CalibMuonAlignJpsiDecision`.

From the dimuon variants, only the `HighMass` mass variant requires the dimuon combination invariant mass to be: $m(\mu^+\mu^-) > 2700 \text{ MeV}$. Since the dimuon signal is relatively sharp around the J/ψ resonance mass, using this trigger is a chance to significantly reduce the combinatorial background. Furthermore, the `Hlt1DiMuonHighMassDecision` trigger also has an almost uniform efficiency as a function of B_s^0 decay time and, thus, will later be referred to as *unbiased* [5].

That leaves only four remaining trigger lines to consider, all of which are `Track` based, *i.e.* they deal with the quality of the reconstructed particle tracks. Including `Hlt1TrackMVA` and `Hlt1TrackMuonMVA` results in a very small signal gain: only $\approx 0.4 \pm 0.08\%$. Since all the `Track` triggers already include requirements for track qualities supplied from some form of an MVA (for example, `TRACK_GhostProb` < 0.2 [156, 157]), the additional requirement here does not provide significant extra discrimination power.

Therefore, the final HLT1 trigger criteria are selected to be the logical *OR* of the `Hlt1DiMuonHighMassDecision` applied to the J/ψ , the `Hlt1TrackMuonDecision` applied to all muon candidates in the event and `Hlt1TwoTrackMVADecision` applied to the dimuon pair part of the reconstructed B_s^0 candidate. The latter is kept as being the only two-track trigger available. These are selected for in a TOS configuration (Section B.3.1) to ensure the trigger decision was driven by the signal candidate itself.

However, it is necessary to note that the thresholds on `Hlt1TrackMuonDecision` and `Hlt1TwoTrackMVADecision` triggers changed between 2015 and 2016 data-taking. Also, `Hlt1TrackMuonDecision` looks at individual muon tracks, requires at least one muon with a $p_T > 1 \text{ GeV}/c$ and large IP significance², while `Hlt1TwoTrackMVADecision` uses an MVA response with a dependence on the

²the difference between the IP from a fit with and without the track in question

Used triggers
LOGlobal (B_s^0)
Hlt1DiMuonHighMassDecision TOS (J/ψ)
Hlt1TrackMuonDecision TOS (B_s^0)
Hlt1TwoTrackMVADecision TOS (B_s^0)
Hlt2DiMuonDetachedJPsiDecision TOS (J/ψ)

Table 6.3 *The analysis-relevant triggers.*

distance of the decay vertex from the PV. All of these introduce a non-trivial dependence in the selection efficiency on the B_s^0 decay time [5]. As such, any trigger combination that includes them will later be referred to as *biased*.

HLT2 selection

Most of the relevant lines in HLT2 are combinations of several HLT1 trigger decisions. So, the only trigger line left to work with is the specifically developed `Hlt2DiMuonDetachedJPsiDecision` for the $B_s^0 \rightarrow J/\psi \phi$ decay mode. The triggers used in the selection are listed in Table 6.3. The total trigger efficiency estimated on simulation is $84.5 \pm 0.01\%$. The result is visualised in the trigger distribution comparison for the J/ψ responses between collision data and signal MC simulation in Figure 6.4. The data distributions are almost identical to the signal MC distributions after the full trigger selection.

Even though the trigger selection is kept quite minimalist and the conditions are chosen with care to not introduce any undesirable effects onto the uniformity of the data, there are some caveats. The two HLT1 track trigger lines and the HLT2 line introduce a decay time acceptance effect for decay times on the order of $\tau(B_s^0) < 6$ ps because of some biasing requirements included in them. Specifically, the HLT1 lines require the χ_{IP}^2 to be greater than 35^\ddagger , while the HLT2 line requires that the J/ψ has a flight distance significance greater than 3. The latter is also referred to as the decay length significance and is the estimated decay length divided by its estimated error. This is done to discard events with uncertain decay length, but in this case, suppresses particles that decay promptly. The effects and the procedure to correct for the above is discussed in Section 7.4.

[‡] χ_{IP}^2 is the χ^2 of the impact parameter or the smallest perpendicular distance of a track to a vertex

The final trigger decision selection is:

$$\begin{aligned}
 & (\text{LOGlobal } B_s^0) \cap (\text{Hlt2DiMuonDetachedJPsiDecision TOS J/}\psi) \cap \\
 & \quad [(\text{Hlt1DiMuonHighMassDecision TOS J/}\psi) \cup \\
 & \quad (\text{Hlt1TrackMuonDecisionTOS } B_s^0) \cup \\
 & \quad (\text{Hlt1TwoTrackMVADecision TOS } B_s^0)] \quad (6.1)
 \end{aligned}$$

It is essentially the intersection of two sets of events. One is triggering the LOGlobal line in L0, and the Hlt2DiMuonDetachedJPsiDecision line on J/ψ candidates reconstructed in HLT2 as part of the signal decay mode. The other set comprises the union between events triggering Hlt1DiMuonHighMassDecision on J/ψ candidates reconstructed as part of the signal and events triggering Hlt1TrackMuonDecision and Hlt1TwoTrackMVADecision on any of the candidates taking part in the reconstruction of the B_s⁰ candidate in HLT1.

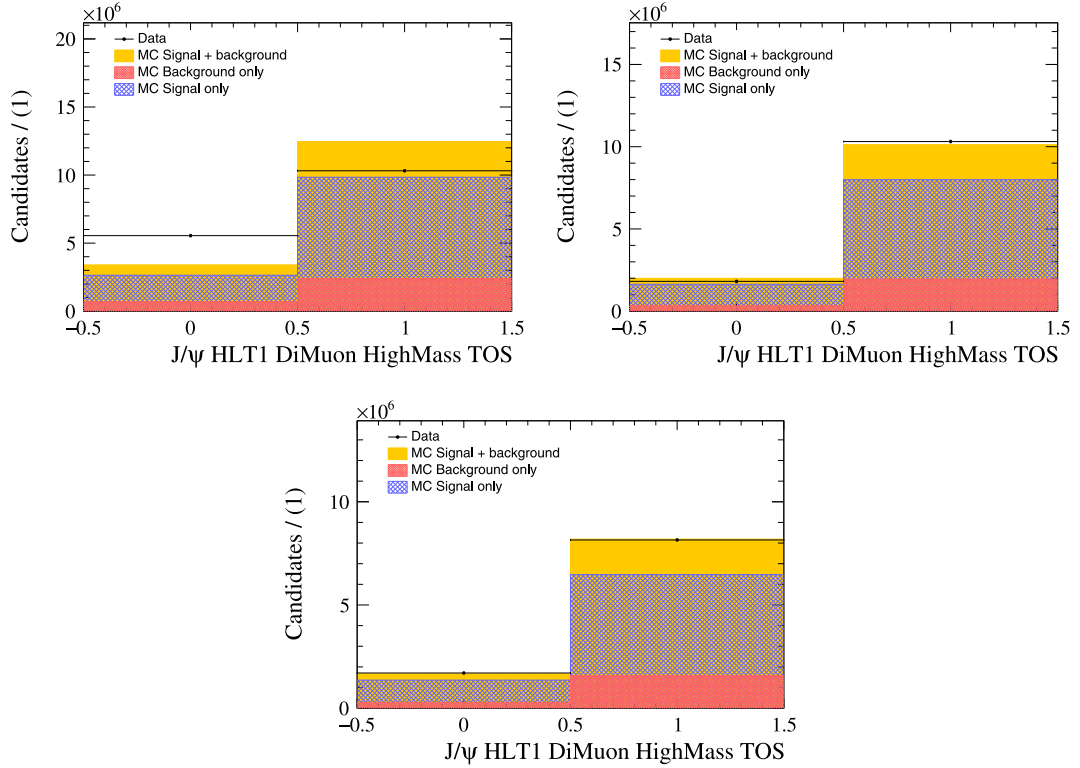


Figure 6.4 Comparison of the distributions of J/ψ trigger response between collision data and signal MC simulation sample for just the L0 requirement satisfied (top left), the L0 and HLT1 requirements combined (top right) and L0, HLT1 and HLT2 requirements combined (bottom).

6.3.2 STRIPPING selection

The later stages of collision data processing are also automated similarly to the trigger. As described in Section 3.5.2, the offline reconstruction step is not as limited in processing time as the trigger and can take much longer to process the incoming events. That gives it the advantage to apply slower but potentially more precise and accurate algorithms because of the reduced number of time-saving assumptions and approximations employed.

The offline reconstruction data flow of LHCb splits the data into streams for later ease of access. To filter the incoming events from the trigger and select them correctly into the appropriate stream, specialised STRIPPING lines are applied to each event. If the event passes the conditions in a specific line, the event gets added to the stream with which the line is associated. Then when analyses want to retrieve data for a particular process or decay mode, they can select the relevant stream and even specify the STRIPPING line of interest, which significantly reduces the computational requirements and time needed to select the appropriate data.

Based on previous experience [15], the LHCb working group focusing on B decays to charmonium states has designed the appropriate STRIPPING lines to select event candidates related to the analyses of the $B_s^0 \rightarrow J/\psi \phi$ decay. There are four crucial lines used in this analysis:

- **StrippingBetaSBs2JpsiPhiDetachedLine:** selects $B_s^0 \rightarrow J/\psi \phi$ candidates with a J/ψ vertex spatially detached from the primary vertex
- **StrippingBetaSBs2JpsiPhiPrescaledLine:** selects $B_s^0 \rightarrow J/\psi \phi$ candidates without spatial requirements on J/ψ vertex
- **StrippingBetaSBd2JpsiKstDetachedLine:** selects $B_d^0 \rightarrow J/\psi K^{*0}$ candidates with a J/ψ vertex spatially detached from the primary vertex
- **StrippingBetaSBu2JpsiKDetachedLine:** selects $B^+ \rightarrow J/\psi K^+$ candidates with a J/ψ vertex spatially detached from the primary vertex

The *detached- J/ψ* condition selects events that have been reconstructed as coming from a B_s^0 meson decay displaced from the vertex by a distance $d > 0.0598$ mm. For a decaying particle, like a B_s^0 meson, this is equivalent to a time-of-flight or a lifetime exceeding $\tau(B_s^0) > 0.2$ ps. The distance ensures that the J/ψ decay vertex is far enough away from the primary vertex to reduce the reconstruction

combinatorial possibilities significantly. As a result the algorithms arrive at a J/ψ reconstruction with a higher statistical significance. The selection criteria used by the `BetaSBs2JpsiPhiDetached` line are listed in Table 6.4.

Entity	Variable	Criteria
all tracks	χ^2/ndf	< 4
$J/\psi \rightarrow \mu^+ \mu^-$	$\Delta \ln \mathcal{L}_{\mu\pi}(\mu^\pm)$	> 0
	$p_T(\mu^\pm)$	$> 500 \text{ MeV}/c$
	$m(\mu^+ \mu^-)$	$\in [3017, 3077] \text{ MeV}/c^2$
	χ_{DOCA}^2	< 20
	$\chi_{\text{vtx}}^2/\text{ndf}$	< 16
$\phi \rightarrow K^+ K^-$	$\Delta \ln \mathcal{L}_{K\pi}(K^\pm)$	> 0
	$p_T(\phi)$	$> 500 \text{ MeV}/c$
	$m(K^+ K^-)$	$\in [980, 1060] \text{ MeV}/c^2$
	χ_{DOCA}^2	< 30
	$\chi_{\text{vtx}}^2/\text{ndf}$	< 25
$B_s^0 \rightarrow J/\psi \phi$	$m(J/\psi K^+ K^-)$	$\in [5150, 5550] \text{ MeV}/c^2$
	$\chi_{\text{vtx}}^2/\text{ndf}$	< 20
	τ	$> 0.2 \text{ ps}$

Table 6.4 *Criteria used to select $B_s^0 \rightarrow J/\psi \phi$ decay candidate events during STRIPPING processing.*

The pre-scaled line uses the same selection criteria apart from the displaced J/ψ vertex requirement. That means it eliminates the decay time bias but makes it quite tricky for the trigger and reconstruction algorithms to isolate genuine J/ψ candidates from all the combinatorial background near the collision vertex. Essentially, it becomes almost impossible to distinguish between two random muons originating from the collision vertex and two muons coming from a decaying J/ψ that was promptly created at the collision vertex. For this reason, the pre-scaled line gets its name — from the artificially applied pre-scale randomly choosing which events get saved and which are discarded based on a fixed average save rate. For the line in question, only 10 % of all events passing the `BetaSBs2JpsiPhiPrescaled` line get saved, or in other words, the *prescale* is 0.1.

Because if it fewer requirements, the pre-scaled line is suited for decay time resolution studies. Since there are less conditions which can bias the decay time estimation, the reconstructed events also span a larger decay time interval.

After the automated steps described above, both the collision data and simulation passing the relevant STRIPPING selection are processed and retrieved for later analysis using DAVINCI version 44r1. Initially, the analysis was using a previous version of the DAVINCI framework. However, after discovering the TISTOS bug (Section C.4), it shifted to a more recent version, which resolved these issues. Furthermore, it was revealed internally that the versions of VELO track fitting algorithms were different between data and simulation (Section C.5). That impacted the resolution of the tracks in data significantly. At the time, there existed no newer version of DAVINCI, which corrected the issue. Thus, close collaboration with the internal VELO experts resulted in appropriate adjustments to the relevant code to have the data and simulation VELO tracks re-fitted with the latest version and correct for the resolution discrepancy.

6.4 Corrections to simulated samples

Simulation samples are essential to the success of the analysis. They are used in numerous places, from evaluating the detector response through estimating the impact of the reconstruction software, all the way to the analysis specific tasks. In particular, the latter category includes improving the final event selection, which relies on multivariate methods (Section 6.5 and 6.5.1), determination of angular efficiency (Section 7.5) and the decay time acceptance (Section 7.4). Simulation is also required to study the effects of different backgrounds polluting the data sample (Section 7.2.2 and 7.8.1) might have on the measurement. The final fit to data cannot be performed without accurate estimates of the above.

Consequently, it is imperative that the simulation reproduces the data distributions as closely as possible. Therefore, several steps are taken to correct the simulation samples where deficiencies are identified and reduce or even eliminate any disagreement between data and MC. The following sections describe in detail the procedures applied.

Furthermore, to improve on the angular resolution, a PV-constraint is introduced into the selection and the reconstruction step is re-done including it. That is

because angular resolution is dependent on the momentum resolution, and the final-state-particle momentum resolution improves with constraining the PV for all daughter particles. So the final selection uses the decay time of the PV-constrained B_s^0 candidates.

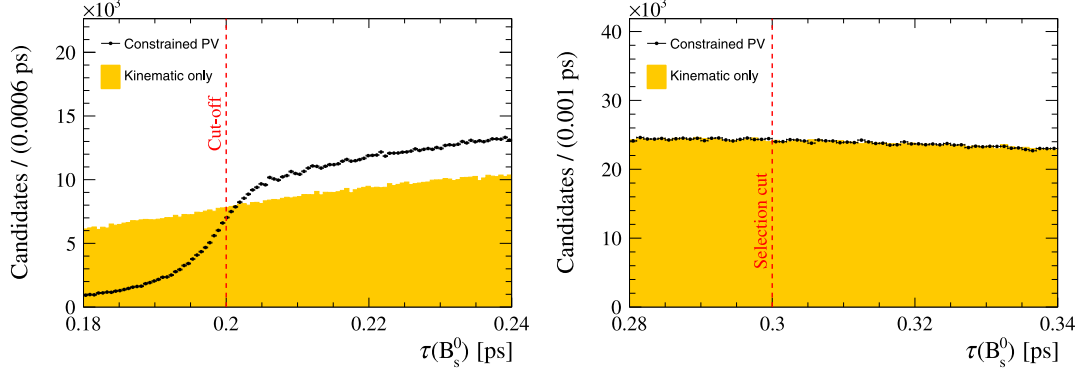


Figure 6.5 *Boundary effect of trigger cut on B_s^0 decay time. Comparing the distribution of reconstructed $\tau(B_s^0)$ near the trigger cut-off (left) and the final selection cut (right). Candidates reconstructed using kinematic information only are shown in yellow, while those reconstructed including a PV constraint are shown in black dots.*

However, due to the fact that the STRIPPING step selects $B_s^0 \rightarrow J/\psi \phi$ candidates based on a fit using kinematic information only (DecayTreeFit [85]), the PV-constrained decay-time distribution is affected for values near the selection cut-off of $\tau(B_s^0) > 0.2$ ps. Figure 6.5 shows that there is a boundary effect introduced when considering STRIPPING-selected $B_s^0 \rightarrow J/\psi \phi$ candidates later reconstructed using a PV constraint. Thus, to eliminate the boundary effect, it was decided to consider only candidates with decay times far away from those values for which the boundary effect is apparent. Figure 6.5 shows that a selection cut of $\tau(B_s^0) > 0.3$ ps is enough to achieve this.

It is also important to note that the most considerable impact from the MC corrections is evident in the multivariate algorithm (MVA) training, the selection and the fit of the B_s^0 invariant mass distribution. And since the decay-time-acceptance control channel shows similar data-simulation differences as the $B_s^0 \rightarrow J/\psi \phi$ decay mode, it is not necessary to re-design the corrections. The same procedures can be applied to these control channel samples as well. That approach has the added benefit of avoiding the introduction of new systematic effects.

6.4.1 Corrections of particle PID

The simulation samples cannot be expected to reproduce the data distributions as already discussed. That is especially applicable for particle PID as its estimation depends on multiple measurements and their combined interplay. PID is estimated using several methods represented in several different variables constituting the result of each technique.

For example, there are two Boolean variables labelled **HasRich** and **IsMuon**. The first is the result of information from the RICH subsystem, and the second is produced by data from the MUON stations [158, 159]. The **HasRich** variable is constructed by associating a track with the pattern matching algorithms that identify RICH rings. The **IsMuon** is mainly used to distinguish other charged tracks from charged muon tracks because pions and kaons can decay in flight to muons and leave traces in the MUON stations. However, the pion and kaon signatures differ from muon signatures by the number of stations with hits and the respective track momentum. Based on these metrics, **IsMuon** is set to true for tracks matching a predefined muon signature.

There are also more advanced methods, such as the Delta Log-Likelihood (DLL), *i.e.* a function-based approach relying on the difference between the likelihood estimations of whether a given track is compatible with a muon hypothesis or not:

$$\text{DLL}_p = \ln(\mathcal{L}_p) - \ln(\mathcal{L}_{\text{non-}p}) = \ln\left(\frac{\mathcal{L}_p}{\mathcal{L}_{\text{non-}p}}\right) \quad (6.2)$$

where p is a placeholder for the particle whose DLL is being calculated (*e.g.* μ), $\mathcal{L}_{p/\text{non-}p}$ is the likelihood a given track is compatible with the particle hypothesis or the non-particle hypothesis given the experimental parameters. For a muon, that would look like:

$$\text{DLL}_\mu = \ln(\mathcal{L}_\mu) - \ln(\mathcal{L}_{\text{other}}) = \ln\left(\frac{\mathcal{L}_\mu}{\mathcal{L}_{\text{other}}}\right) \quad (6.3)$$

However, in the end, a further iteration is preferred in practical use:

$$\text{PID}_p = \text{DLL}_p - \text{DLL}_\pi = \Delta \ln \mathcal{L}_{p\pi} \quad (6.4)$$

where $p \neq \pi$, is a particle hypothesis different from the pion hypothesis. In the

case for a μ that is:

$$\text{PID}_\mu = \text{DLL}_\mu - \text{DLL}_\pi = \Delta \ln \mathcal{L}_{\mu\pi} \quad (6.5)$$

Finally, more sophisticated methods are concerned with a broader view of each event and its associated tracks. These employ the response of dedicated pre-trained artificial neural networks and assign that as the probability that a given track is from a kaon (**ProbNNk**), pion (**ProbNNpi**), electron (**ProbNNe**), muon (**ProbNNmu**) or proton (**ProbNNp**) [70]. The neural networks are trained using several input variables containing information from the RICH subdetectors, the tracking system, muon detectors and the calorimeters.

Type Requirement	K	μ
other	HasRich $p_T > 500.0$ $\chi^2_{\text{track}}/\text{ndf} < 25.0$	HasRich IsMuon $p_T > 500.0$ $\chi^2_{\text{track}}/\text{ndf} < 25.0$
PID	$\Delta \ln \mathcal{L}_{K\pi} > 0.0$	$\Delta \ln \mathcal{L}_{\mu\pi} > 0.0$

Table 6.5 PIDCALIB *efficiency conditions per resampled track type.*

Thus, any individual differences between data and simulation in any of the associated variables and any differences in the way these correlate with each other will affect the resulting PID response. To combat these discrepancies between data and simulation, LHCb has developed a package with the goal to “correct” the presumably incorrect PID variables in simulation. The PIDCALIB [160] package is used to construct PDFs correlating the individual **ProbNN** responses for tracks in data with associated variables external to the **ProbNN** algorithms themselves but still correlated as part of the event. The PDFs are then used to resample the PID variables in simulation based on the corresponding external variables there.

Essentially, PIDCALIB applies an explicit re-weighting of the PID distributions in simulation [160]:

(...) which requires no information from the signal data, but simply a simulated signal sample with no PID selection applied. Using a subset of the variables η , p , p_T , and the track multiplicity in the

event, the calibration sample is split into regions, and in each region a distribution of the PID variable in question is constructed. Using the same subset of variables, each track in the signal simulation is matched to one of the regions defined for the calibration sample. The distribution of the PID variable from the matched region is then used as a PDF to randomly draw a new PID value. In this way, the efficiency of the selection can then be extracted from the signal simulation itself. This procedure has the effect of improving the agreement of the PID distributions between simulation and data. However, it only preserves the correlations of the sampled variables (up to three in the present implementation in PIDCALIB) and breaks any others, the effects of which have to be carefully evaluated on a case-by-case basis.

Since the strategy is to use these variables in a later MVA training and selection, it is desirable to have them better represent the correlations present in data between the kinematic and multiplicity variables. Given the statistical nature of MVAs, it is much more important to preserve dataset correlations if the choice is between those or individual values.

It is beneficial to resample the contributions in simulation while controlling the efficiency of the reweighting to preserve the correlations best. For this purpose, the distributions of p , p_T and the track multiplicity are selected to construct a partitioning scheme for the calibration data of PIDCALIB, such that the resampling efficiency, ε_{PID}^{eff} , is kept high for the corresponding track selection criteria. The overall resampling efficiency is the effective efficiency of the PID resampling of the **ProbNN** variables in simulation depending on the partitioning scheme of the calibration data and the PID related cuts of the data sample, whose distributions the signal simulation aims to reproduce.

Given a scheme, the overall effective PID efficiency takes the form:

$$\varepsilon_{PID}^{eff} = \frac{\sum_i \varepsilon_i N_i}{\sum_i N_i} \quad (6.6)$$

where ε_i is the efficiency in each segment of the partition:

$$\varepsilon_i = \frac{N'_i}{N_i} \quad (6.7)$$

where N_i is the number of calibration events in the i^{th} partition segment before the

Track type	Low momentum ($p < 200$ GeV)	High momentum ($p \geq 200$ GeV)
e^\pm	$B^+ \rightarrow (J/\psi \rightarrow e^+e^-)K^+$	
μ^\pm	$B^+ \rightarrow (J/\psi \rightarrow \mu^+\mu^-)K^+$	$J/\psi \rightarrow \mu^+\mu^-$
π^\pm	$K_s^0 \rightarrow \pi^+\pi^-$	$D^{*+} \rightarrow (D^0 \rightarrow K^-\pi^+)\pi^+$
K^\pm	$D_s^+ \rightarrow (\phi \rightarrow K^+K^-)\pi^+$	$D^{*+} \rightarrow (D^0 \rightarrow K^-\pi^+)\pi^+$
p, \bar{p}	$\Lambda \rightarrow p\pi^-$	$\Lambda \rightarrow p\pi^-, \Lambda_c^+ \rightarrow pK^-\pi^+$

Table 6.6 *The decay modes used to select the calibration samples for PIDCALIB in Run 2 [161].*

PID requirements are applied, and N'_i is their number after the PID requirements are applied. The equivalence $\varepsilon_{PID}^{eff} \equiv N'/N$ for the calibration samples is valid in general. However, for events containing multiple tracks, the individual ε_i are calculated by the ensemble probability:

$$\varepsilon_i = \left(\prod_j^n \varepsilon_i^j \right)^{\frac{1}{n}} \quad (6.8)$$

where n is the number of tracks in the event.

The PIDCALIB calibration samples are produced for each charged particle track of interest and provide reweighting for each ProbNN variable. A list of decay modes in data used to collect the calibration samples for each track type is shown in Table 6.6. The relevant PIDs for this analysis are for μ (ProbNNmu) and K tracks (ProbNNk, ProbNNpi and ProbNNp). Kaons require the different treatment of ProbNNpi and ProbNNp since these variables will later be used to estimate peaking backgrounds in data and their statistical subtraction (Section 7.2.2).

The selection criteria used for charged tracks in $B_s^0 \rightarrow J/\psi \phi$ as discussed in Section 6.3.2 (Table 6.4) are highlighted in Table 6.5, with the track and PID requirements separated. These are used in an iterative procedure that tries to come up with a calibration data partitioning which maximises ε_{PID}^{eff} .

The procedure utilises a tool that takes the desired variables (p , p_T and track multiplicity), applies the tracking selection requirements and then tries to produce an isometric partitioning scheme: each partition having the same number of events. This has two effects. Firstly, it makes sure it keeps statistical power relatively high and even across partition segments by avoiding to create slices with very few events for the sake of efficiency. Secondly, this results in an even coverage of the whole distribution space of reweighting variables.

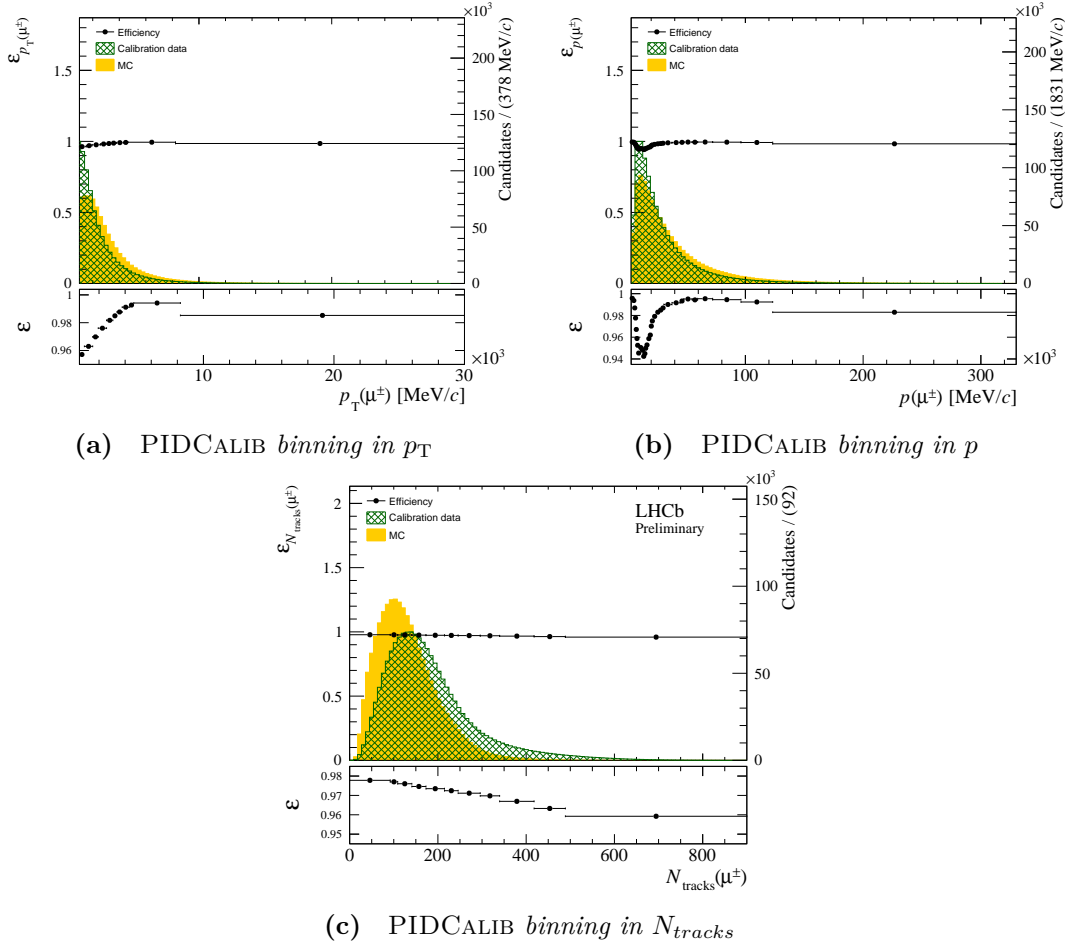


Figure 6.6 Projections of the binning scheme used to construct the distribution to resample $\text{ProbNNmu}(\mu)$ response in simulation. Left axis shows binning scheme efficiency, right axis shows $B_s^0 \rightarrow J/\psi \phi$ candidates in data and simulation (normalised to data). Showing the projection in $p_T(\mu)$ (a), the projection in $p(\mu)$ (b), and the projection in $N_{track}(\mu)$ (c) [160].

The tool takes a desired initial number of segments (bins) and creates a scheme. It starts by defining the boundaries of each bin in each variable. Then the content in each bin is evaluated and compared by the number of contained events and efficiency after PID cuts. The bin boundaries are then adjusted to improve upon both the efficiency and equalising the bin event content. If that is not possible, a new bin is created or removed depending on the particular case.

The iterative procedure varies the initial conditions regarding the number of bins, PID cuts and extremal limits of the variables. The latter has the requirement that the resulting distribution is a superset of the equivalent distribution in simulation.

³Gradient Boosted (GB) reweighting, where the event weights are produced by a gradient boosted decision tree.

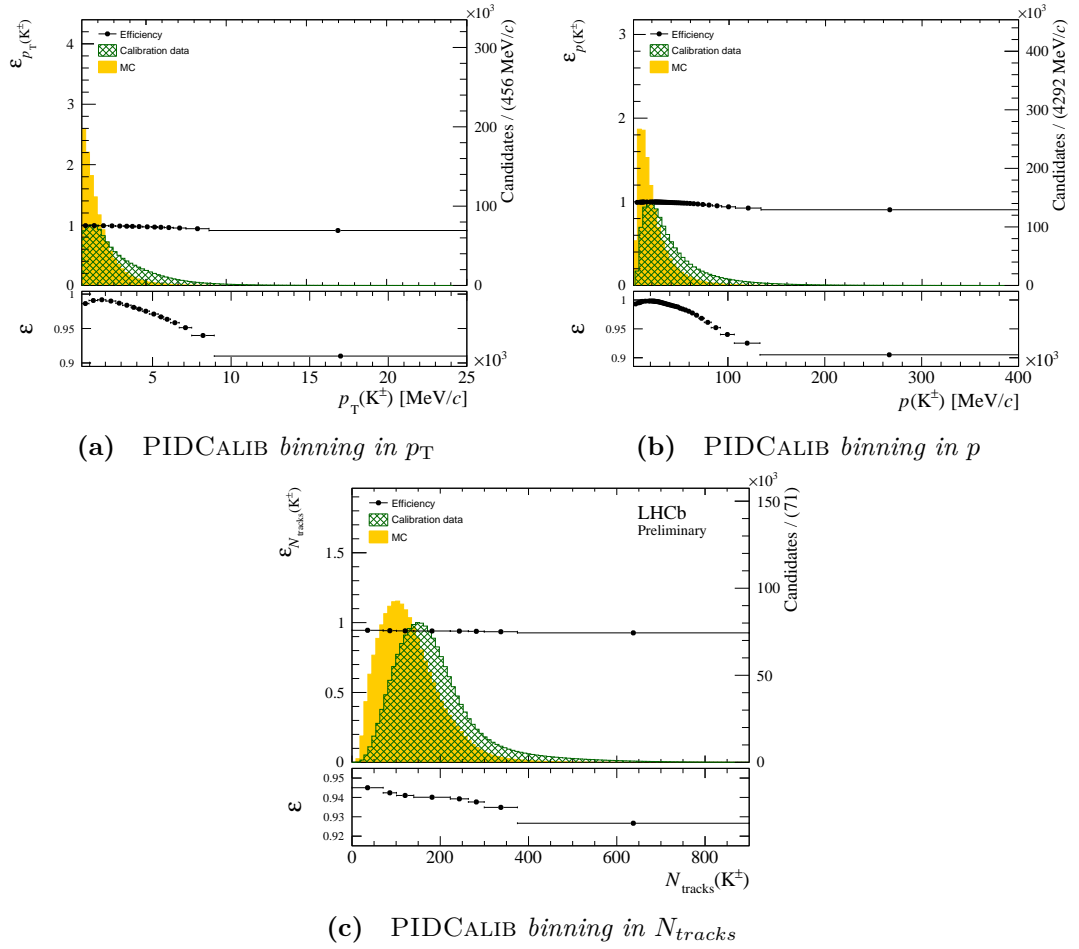


Figure 6.7 *Projections of the binning scheme used to construct the distribution to resample ProbNNk(K) response in simulation. Left axis shows binning scheme efficiency, right axis shows $B_s^0 \rightarrow J/\psi \phi$ candidates in data and simulation (normalised to data). Showing the projection in $p_T(K)$ (a), the projection in $p(K)$ (b), and the projection in $N_{track}(K)$ (c) [160].*

The procedure culminates in creating two separate partitioning schemes in p , p_T and track multiplicity: one for muon tracks and one for kaon tracks. Effectively, these can be viewed as a PDF, $F = F(p, p_T, \mathbf{nTracks})$, where the value of F specifies the distribution of ProbNN responses for all tracks in the calibration sample with corresponding p , p_T and track multiplicity from the particular partitioning bin. For muons, F is essentially a function $F : \mathbb{R}^3 \rightarrow \mathbb{R}^1$, where the image is a series of distributions of ProbNNmu. For kaons, on the other hand, three separate sub-schemes $F^i : \mathbb{R}^3 \rightarrow \mathbb{R}^1$, $i \in [1, 2, 3]$, had to be created, each having the image being a series of ProbNNk, ProbNNpi or ProbNNp. That ensures the correlations between the binning variables and between the resulting ProbNN responses are preserved for all types.

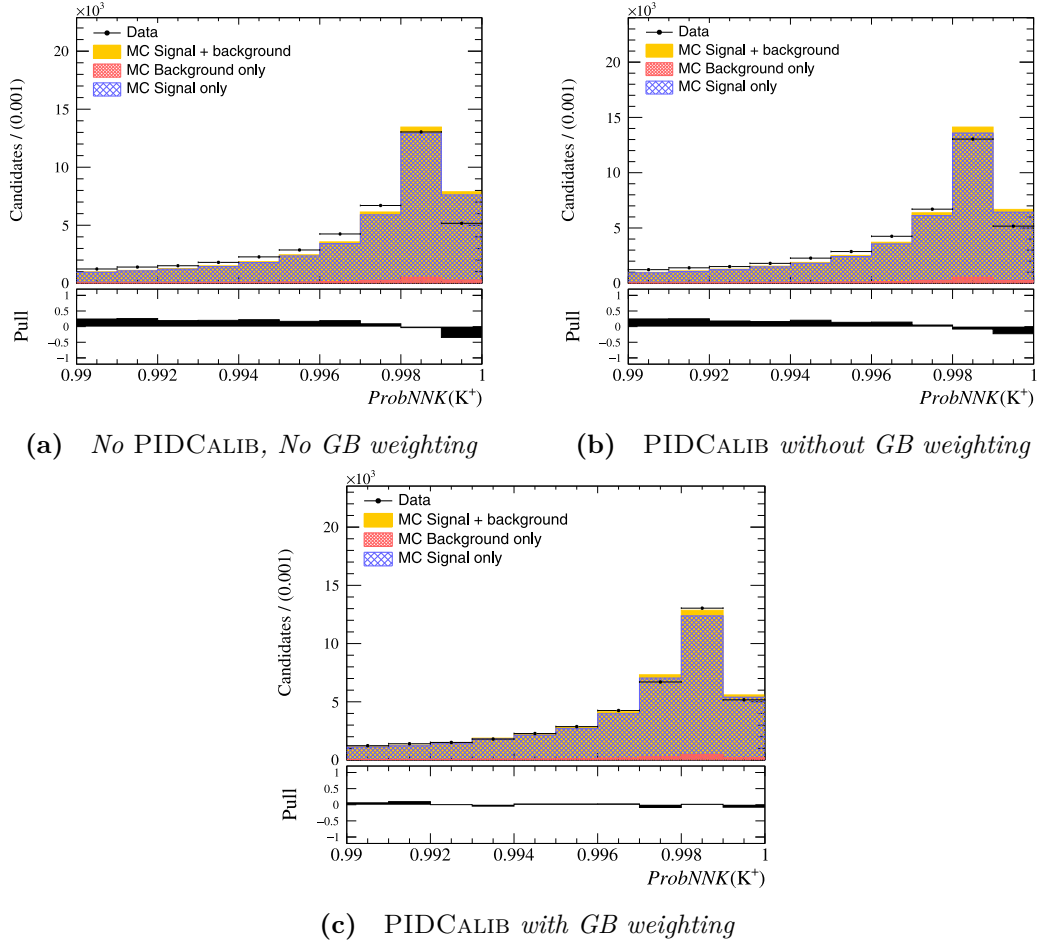


Figure 6.8 Corrections to simulation PID response for K^+ showing before any corrections (a), after PID resampling (b), and after PID resampling combined with GB reweighting³ to signal distributions in data (b) [160, 162].

The projections of the resulting F for muons ($\text{ProbNN}\mu$) and kaons ($\text{ProbNN}k$) in each of the variables are shown in Figure 6.6 and Figure 6.7, respectively, accompanied by the efficiency per bin integrated over the other two dimensions. The final resampling efficiencies are:

- $\varepsilon_{eff}^{PID}(\mu) = 0.975 \pm 0.003$
- $\varepsilon_{eff}^{PID}(K) = 0.932 \pm 0.004$

where the kaon efficiency is the result of the sub-partitioning efficiencies combined further using the ensemble rule from Equation 6.8 with $n = 3$ for the number of sub-schemes.

Afterwards, the distributions of ProbNN in the simulation are resampled using

their corresponding F scheme. A track from the simulation is taken with its p , p_T and associated track multiplicity. These are used to select the particular bin in F and the respective **ProbNN** distribution is randomly sampled to produce a new **ProbNN** value for the simulation track. The procedure results for **ProbNNmu** and **ProbNNk** on the muon and the kaon tracks, respectively, are shown in Figure 6.10b and Figure 6.8b. For comparison, the same are shown before the procedure in Figure 6.10a and Figure 6.8a, respectively.

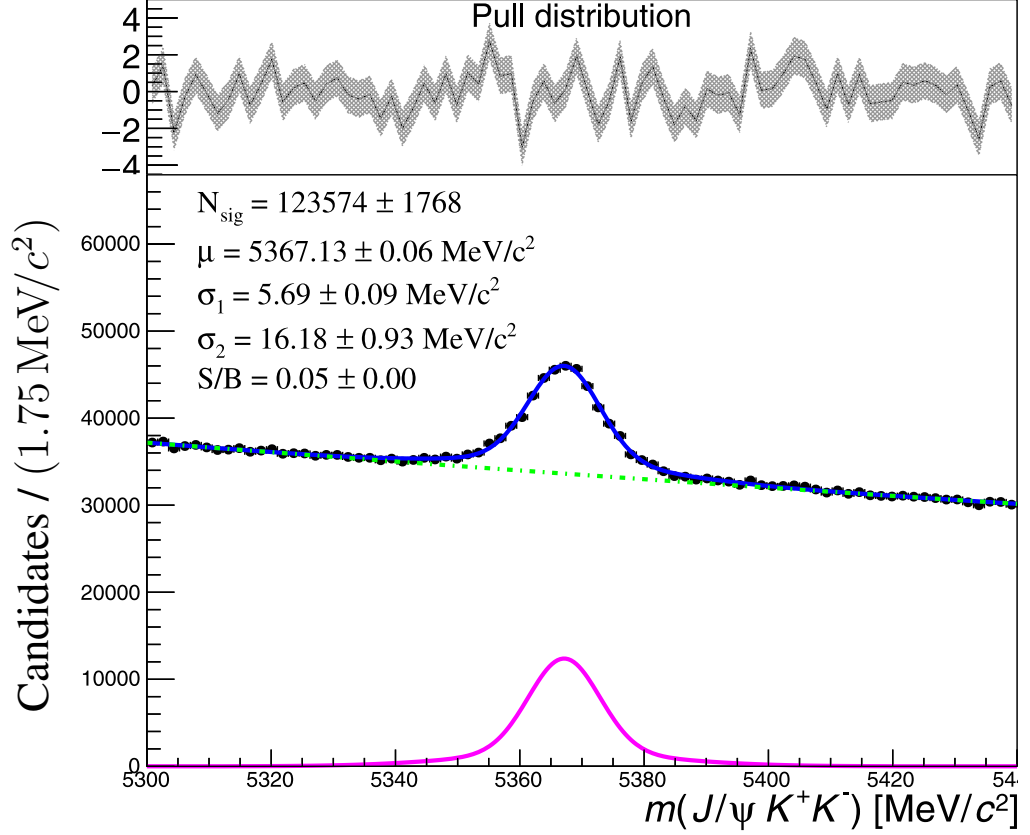


Figure 6.9 *Distribution of $m(J/\psi K^+ K^-)$ after the STRIPPING and trigger selection in 2016 data. The solid blue line shows the total fit with the signal and combinatorial background described by the pink and green lines, respectively.*

The pull distributions compare simulation against $_{\text{s}}\text{Weight-ed}$ data⁴. For this purpose, the data sample is processed in the following way:

- all requirements discussed in Section 6.3.1 and Section 6.3.2 are applied

⁴ $_{\text{s}}\text{Weight-ed}$ data is data that is weighted using the $_{\text{s}}\text{Plot}$ method described in Section C.2

- the data are required further to include only events with an invariant mass of the B_s^0 candidate $m(B_s^0) \in (5320, 5420) \text{ MeV}/c^2$
- the resulting sample is fit with a double Gaussian PDF for the signal and an exponential function for the background

The resulting s Plot for the invariant mass of the $J/\psi K^+ K^-$ combination is shown on Figure 6.9.

6.4.2 Reweighting of simulation kinematic, event and selection variables

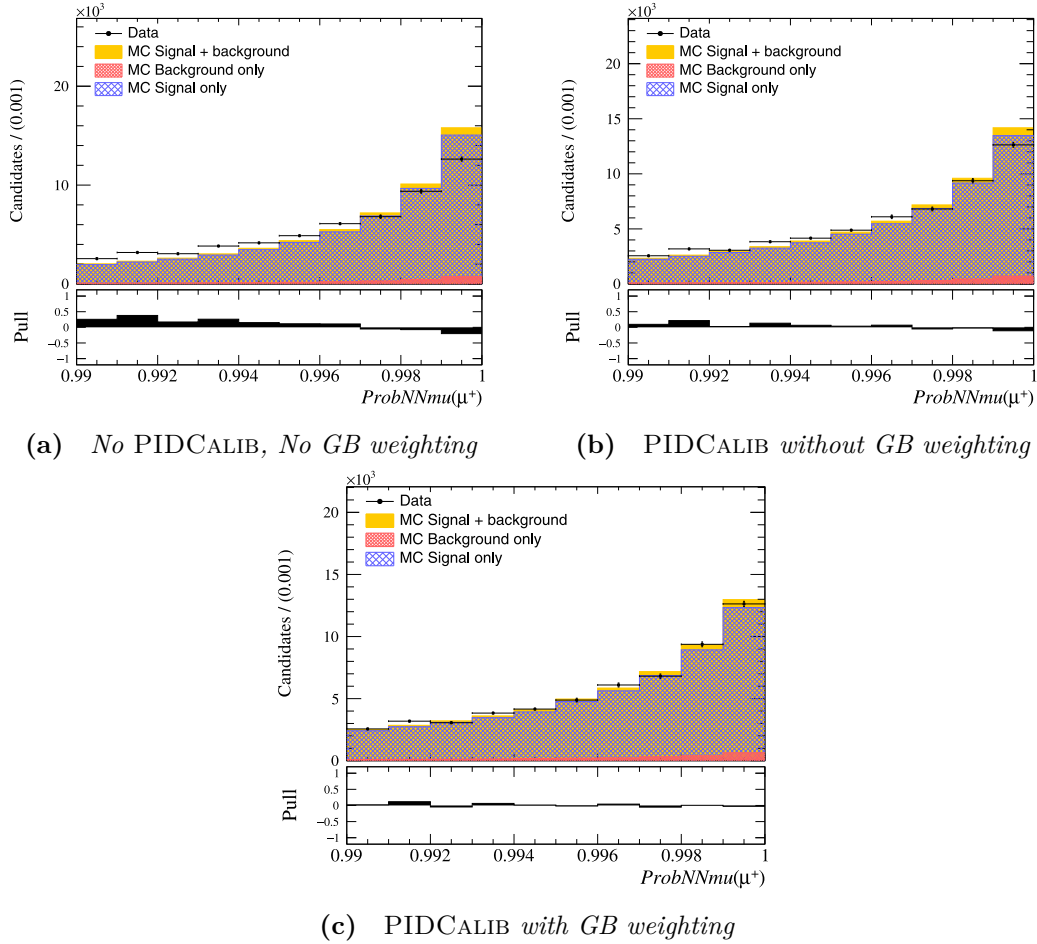


Figure 6.10 Corrections to simulation PID response for μ^+ showing before any corrections (a), after PID resampling (b), and after PID resampling combined with GB reweighting⁵ to signal distributions in data (b) [160, 162].

Having a closer look at the results of the PIDCALIB resampling, one might suppose there was little to no effect or that the procedure might have even had a negative contribution to the data-simulation agreement. However, the pull distributions are reduced overall after the resampling is applied. But it is also apparent that the results are not entirely satisfactory. A clue might be found by looking at the superimposed distributions of the partitioning variables in data and simulation (Figure 6.6 and Figure 6.7). These clearly show that the kinematic and multiplicity distributions do not agree very well from the start. A further correction is needed to address this discrepancy in the form of another reweighting of the relevant variables in the simulation.

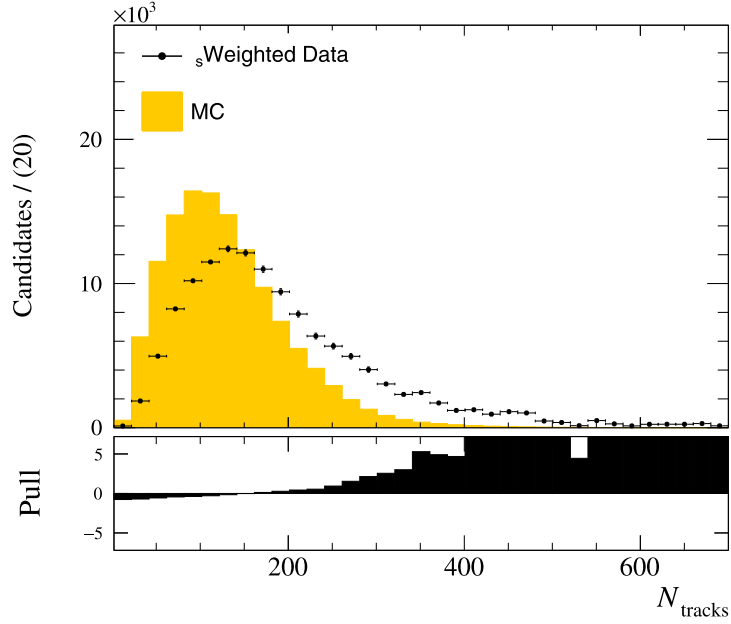


Figure 6.11 *Distributions of track multiplicity between s Weight-ed normalisation sample (black) and simulation (yellow).*

The typical approach is to split the distribution on a per-variable basis into multiple bins and then define a weight per bin to compensate for the discrepancy between data and simulation. In simple cases and low dimensionality, this may be enough and can be viewed as histogram division. However, the MVA will use a multitude of variables and try to take advantage of their intrinsic correlations. That means the binning approach will not suffice. The reason for that is that when many variables are used, the multidimensional binning has a higher and higher chance of leaving bin regions empty with each additional dimension due to the limited statistics of the samples. In practice, only one to two variables can successfully be reweighted, and the choice of which ones is complex and hides risks of introducing disagreement in other variables. Furthermore, the sample

size needed to estimate a density grows exponentially with each variable added for reweighting, because of the increase in dimensionality and each bin’s volume growth.

For this purpose, Gradient Boosted (GB) Reweighting is applied to solve this problem [162]. It uses an MVA approach, namely Boosted Decision Trees (BDTs), to find the optimal regions for reweighting. It is essential to distinguish these types of regions from the “rectangular” bins of the above-mentioned usual method. BDTs can split the multidimensional distribution into regions based on conditions. These regions are defined based on expressions (functions) of one or more dimensions. Thus, the distribution is split according to its statistical properties, and the region’s volume is no longer required to increase with the number of dimensions. Essentially, the significance of each region split is evaluated as part of the algorithm, and thus each variable takes part only in its relevant condition. That ensures the splitting is problem-specific and optimised.

Initially, the sample is split into a few vast regions using randomly generated conditions. The regions are passed to a decision tree which assigns a leaf node to each region. A process is then started to optimise the split conditions and the number of leaves (regions). That is achieved through a loss function and a multidimensional stochastic gradient descent. In this specific case, the loss function is χ^2 statistic on the difference between the data and simulation. The loss function is quantitatively estimating the χ^2 difference between the simulation (original) and data (target) distributions in total and in each region. It can be used to select regions more significant than others for the reweighting by looking at where its value is large. These regions’ weights are adjusted in the appropriate direction towards an agreement with data by small increments. That is the basis of the BDT’s learning algorithm. Moreover, to speed up the procedure and ensure the correct minimum is reached, the BDT utilises gradient boosting on its gradient descent to minimise the loss function [163].

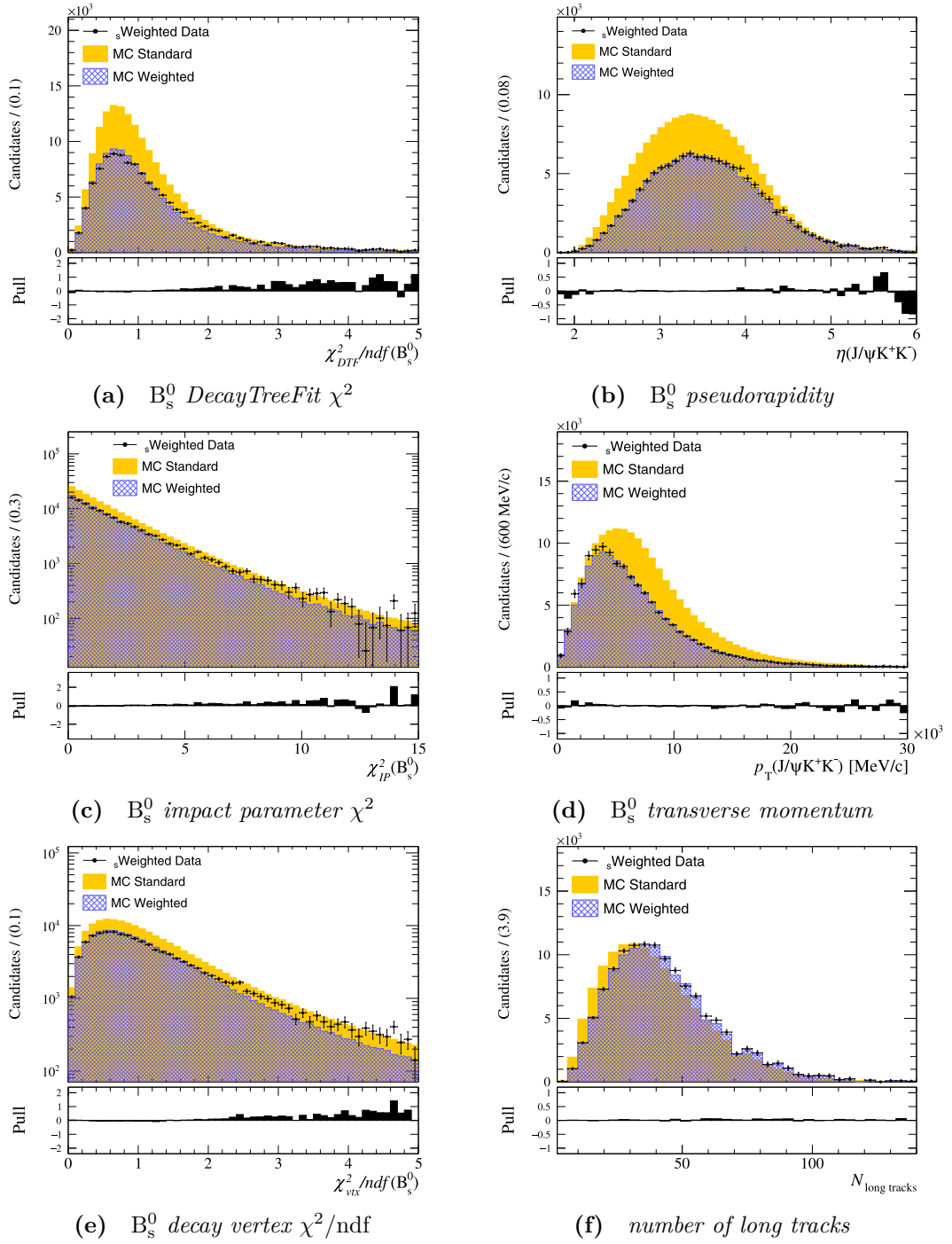


Figure 6.12 Distributions of kinematic, event and selection variables for the s Weight-ed normalisation sample (black) and simulation before (yellow) and after (blue) the gradient boosted reweighting corrections are applied [162].

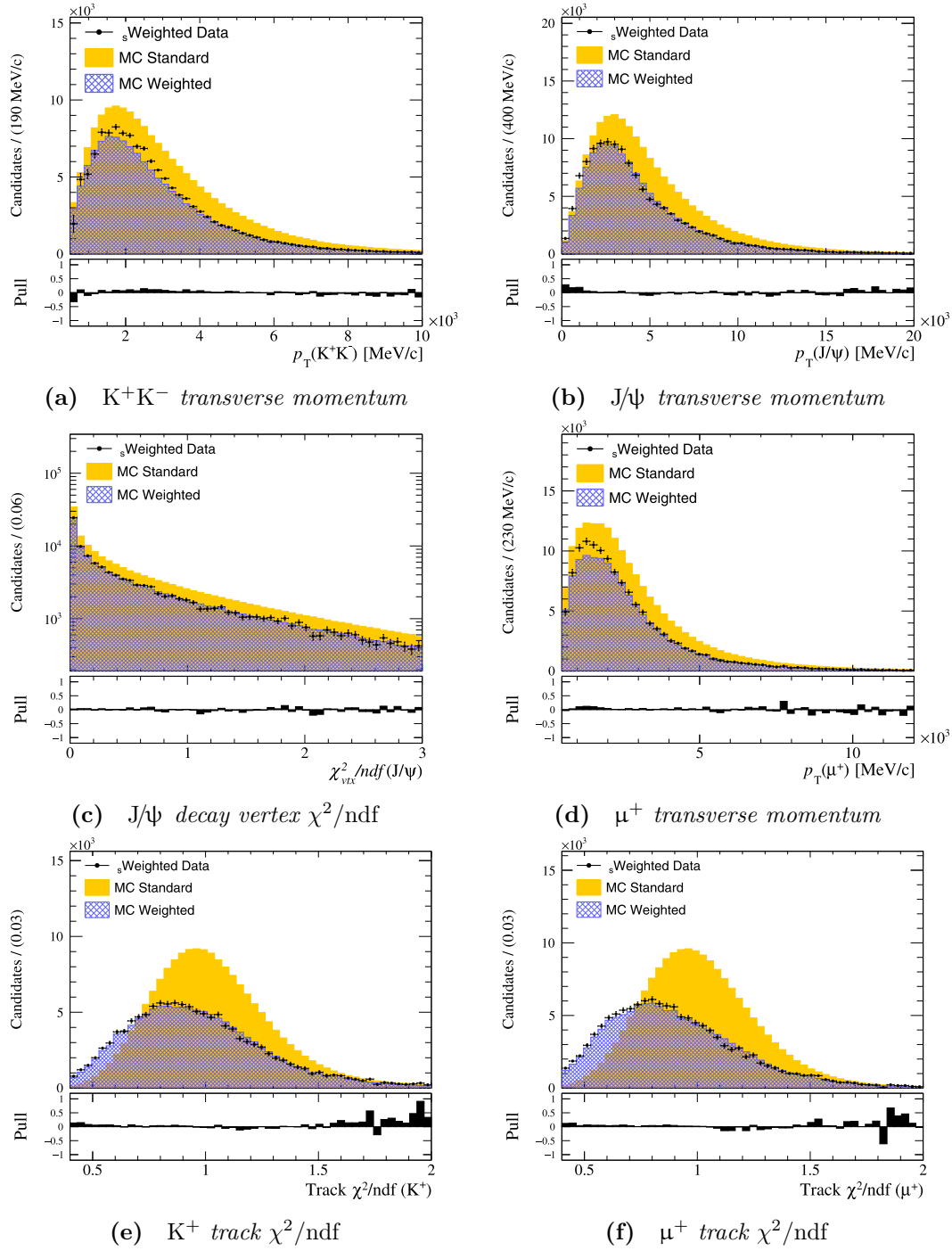


Figure 6.13 Distributions of kinematic, event and selection variables for the s Weight-ed normalisation sample (black) and simulation before (yellow) and after (blue) the gradient boosted reweighting corrections are applied [162].

Moreover, the benefits of this procedure do not end with simply the optimisation of region splitting. As the distribution is handled based on its properties and reweighted to match the reference based on variable-specific conditions, internal correlations are automatically handled. The result is a genuinely multidimensional reweighting of the distribution in question. It not only agrees to the reference in its 1-D variable projections but also agrees as multidimensional distribution.

The reweighting procedure is performed starting with the preparation of a normalisation sample from data. It is prepared as described in the previous Section 6.4.1. Namely, data is selected using only the requirements discussed in Section 6.3.1 and Section 6.3.2 and includes only events with an invariant mass of the B_s^0 candidate $m(B_s^0) \in (5320, 5420) \text{ MeV}/c^2$. This sample is then `_sWeight-ed` through an `_sPlot` fit to the invariant mass of the $J/\psi K^+ K^-$ combination. The fit is shown in Figure 6.9. As discussed in Section C.2, this technique allows extracting the shape of the signal distribution present in data by statistically subtracting the background distribution.

However, it has its limitations. Specifically, the accuracy and precision of the signal and background models are crucial for a successful `_sPlot` fit. The approach is also limited by only being a fit to a 1-D projection of the multidimensional distribution in invariant mass. Nevertheless, it is sufficient for the reweighting procedure as the most relevant distinction of interest is the approximate shape of the signal present in the data. Attempting to come up with more accurate and precise models of signal or background distributions in data at this stage will be detrimental to the final goal as the background population in data is overwhelmingly larger by two orders of magnitude.

With the normalisation sample ready, the variables for reweighting were chosen to be:

- μ^\pm track χ^2/ndf
- K^\pm track χ^2/ndf
- B_s^0 transverse momentum, $p_T(B_s^0)$
- B_s^0 pseudorapidity, $\eta(B_s^0)$
- number of long tracks in the event, `nLongTracks`

The latter was chosen instead of whole event track multiplicity, **nTracks**, due to significantly different distributions between data and simulation and limited statistics for high multiplicity in MC (Figure 6.11). The final set of variables was chosen after considering several different options and conducting several trials. The main aim was to reach the broadest possible agreement between the multidimensional distributions with the fewest variables, explicitly targeting the variables later used to train the MVA (Section 6.5).

The reweighting is performed using data collected during 2016 for creating the normalisation sample and S26 simulation sample for the same year with 20 M events in total. The resulting reweighting BDT is then applied to both 2015 and 2016 simulation samples. That is done to avoid the limiting factor of low statistics (event yields in other samples). Figures 6.12 and 6.13 show the various projections of the normalisation sample and the simulation before and after reweighting. It can be seen that distributions of the simulation after the GB reweighting procedure are in good agreement with those in the _sWeight-ed data control sample, while they did not agree beforehand. Furthermore, the multidimensional distributions agree in many more variables than those employed during the reweighting procedure, confirming the initial premise.

GB reweighting shrinks the effective simulation sample size by 30.46 %. It is trained on the $B_s^0 \rightarrow J/\psi \phi$ signal simulation sample, which contains 20 M events. That means that the effective size of the sample after reweighting is about $20(1 - 0.3046) = 13.91$ M events. To quantify the effect of this reduction, one can refer to the way the simulation sample is later used in the MVA training and the resulting model. The MVA selection performance is estimated using a fit to the $J/\psi K^+K^-$ invariant mass spectrum. The relevant metrics are signal yield, background yield and the mean and standard deviation of the fit.

Thus, it can be estimated what effect this reduction in sample size would have on the accuracy of the fit. In the simple case where the signal is only modelled with a double Gaussian PDF and the background as an exponential, the standard deviation, σ , is taken as:

$$\sigma = \sqrt{f\sigma_1^2 + (1-f)\sigma_2^2} \quad (6.9)$$

where f and $(1-f)$ are the fraction coefficients of the two Gaussian PDFs resulting in the total double Gaussian PDF. The signal population standard deviation of

the mean scales with the effective sample size:

$$\sigma_j = \frac{\sqrt{\sum_i^n (x_i - \mu)^2}}{\sqrt{n}} \quad (6.10)$$

where σ_j is the standard deviation of the mean, μ , of the j^{th} Gaussian component and n is the effective sample size. That essentially means σ_j can be affected by the reduction of the sample size used to estimate it. Since the signal simulation sample trains the MVA to identify its distribution in data, the reduction in effective sample size will increase the uncertainty associated with its size.

However, since the reweighting procedure improves upon reproducing the data distributions, this effect should only be statistical rather than biasing the MVA selection in any way. In that case, given the initial and final sample size is 20 M and 13.91 M events, respectively, the relative error introduced is around 0.05 %. It is important to note that the threshold for the relative error exceeding 0.5 % is when the simulation sample size shrinks by 89 %.

6.5 Multivariate selection

After the initial loose trigger and STRIPPING selections, the data still contains large amounts of combinatorial background contributions. To keep as much signal as possible, the selection discussed in the previous sections has barely removed any unwanted backgrounds, as shown by the signal-to-background ratio of $S/B = 0.05$ in Figure 6.9.

The strategy to improve upon this is to utilise multivariate algorithms (MVAs) to select the signal events and discard the backgrounds. As already discussed in Section 6.4.2, MVAs can provide much better performance in signal retention and background rejection than simple “rectangular” cuts. These can efficiently remove combinatorial backgrounds and further remove polluting data that could appear signal-like in terms of particle content and properties but disagree with the overall signal signature in a combined multidimensional model.

Several MVA algorithms are initially trained and compared based on several metrics to select the most appropriate that fits the purpose. The TMVA [116] software package, provided by the ROOT analysis framework, is used to prepare, configure, train and test the MVA algorithms.

6.5.1 MVA training and optimisation

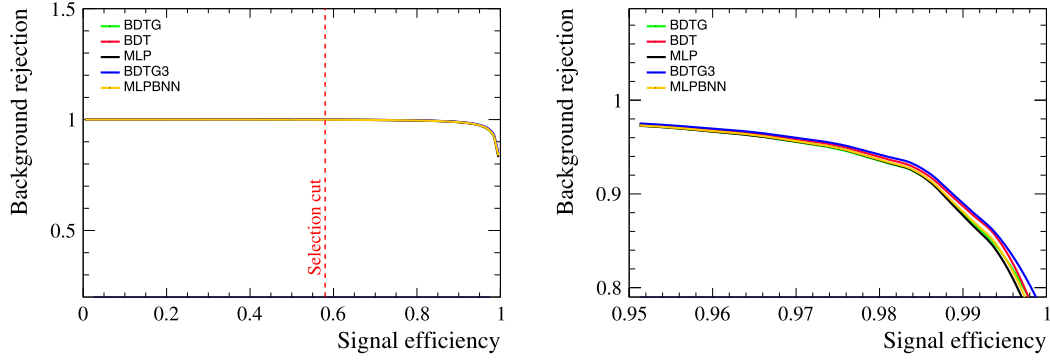


Figure 6.14 *Background rejection versus signal efficiency of the different MVAs trained to select $B_s^0 \rightarrow J/\psi \phi$ events. On the left, the final selection cut is shown for visual aid. On the right, the higher signal efficiency region of the classifiers is shown to display their differences. The naming scheme of the individual MVA algorithms is explained in Section 6.5.1.*

A few competing MVA algorithms are chosen to compare. Two primary types are compared: boosted decision trees (BDTs) and artificial neural networks (ANNs or NNs for short). The BDTs and NNs are split further based on their specific underlying configuration. In particular, BDTG is a label used to specify a BDT that utilises a gradient boosting algorithm to improve its learning rate. Also, all the NNs used here are of the multilayer perceptron (MLP) type. These are a class of the so-called feed-forward neural networks that consist of one, but more frequently, more than one layer of neurons (linear functions with multiple inputs and one output).

Five configurations of such algorithms are prepared. Three BDTs and two MLPs. They are as follows:

- BDT — BDT with 850 nodes with a maximum depth of 3 and the AdaBoost[164] learning algorithm
- BDTG and BDTG3 — BDT with 1000 nodes with a maximum depth of 3 and 6, respectively, and the gradient boosting learning algorithm
- MLP — MLP with a minimum of 5 hidden layers with tanh activation, stochastic gradient descent learning algorithm and no regularisation
- MLPBNN — MLP with a minimum of 5 hidden layers with tanh activation, BFGS[165, 166] learning algorithm and Bayesian regularisation

Both the AdaBoost and gradient boosted BDTs take advantage of essentially the same technique of using “weak” prediction models (decision trees) and combine them in an ensemble to form a much better overall classifier. The standard gradient boosting is achieved by iteratively adjusting the decisions of each tiny decision tree by following the slope from misclassified to correctly classified data. That is a sort of regression to minimise some loss function. However, the AdaBoost algorithm is a bit more efficient in the sense that it gives larger weights to wrongly labelled events before passing it on to the decision tree down the chain and thus improves the learning rate.

The MLPs compared here are all using a similar overall structure, however, with notable differences. The most simple MLP with stochastic gradient descent without regularisation is compared to a more advanced neural network and learning algorithm. Both use a sigmoid activation function on all layers as those are known to be good approximators of non-linear functions[115]. That makes each neuron exhibit the behaviour as a kind of switch, either being active or inactive, as sigmoid functions have a characteristic *S*-shape and their values are within $\in (0, 1)$ or $\in (-1, 1)$. However, the more advanced MLPBNN uses the Broyden-Fletcher-Goldfarb-Shanno [167–170] (BFGS) algorithm to improve the learning rate by adjusting the gradient using further information about the curvature supplied by the Hessian. That results in faster convergence and decreases the probability of arriving at shallow or local minima as well as multidimensional saddles.

On top of that, a Bayesian regularisation [171] is applied. It assigns significance to each weight in the neurons in all layers and prevents weight biases from occurring. For example, a huge individual weight in a neuron with many inputs may bias it so that when the backpropagation adjustments become small around a minimum, the particular neuron can no longer learn properly.

Plotting the signal efficiency (retention) versus the background rejection for various thresholds of the MVA response defines a curve that illustrates its performance. This curve is shown in Figure 6.14. That is equivalent to the more widely used term of a receiver operating characteristic (ROC) curve, where the background rejection is replaced by the false positive rate, or in other words, the background retention. Both can be viewed as a reflection of each other since the background rejection and retention sum up to 1 by definition. Thus, the term will be used interchangeably and distinction will be made only when necessary.

Before applying training and applying the MVAs, the following selection requirements are made. The $J/\psi K^+ K^-$ invariant mass distribution is restricted to the range $5200 \text{ MeV}/c^2 < m(J/\psi K^+ K^-) < 5550 \text{ MeV}/c^2$ and the $K^+ K^-$ invariant mass distribution is restricted to the range $990 \text{ MeV}/c^2 < m(K^+ K^-) < 1050 \text{ MeV}/c^2$. Moreover, as in the Run 1 analysis[15], the uncertainty on decay time, returned by the DecayTreeFit (DTF) used to determine the decay time, is required to be smaller than 0.15 ps to remove events with a very poor determination of the decay time. The MVA training is performed using 2016 samples only. Namely, the 2016 corrected simulated sample is used as the signal sample while 2016 data candidates with $5450 \text{ MeV}/c^2 < m(J/\psi K^+ K^-) < 5550 \text{ MeV}/c^2$ are used for the background sample. Special care was taken to avoid variables that could affect angular or decay time efficiencies, like impact parameter χ_{IP}^2 of final state particles, the direction angle of the B_s^0 (DIRA) or transverse momentum of final state particles. Ten variables were distilled to train the MVAs after trying several others and gradually excluding each for specific reasons:

- the maximum of the kaon track $\chi_{\text{track}}^2/\text{ndf}$
- the maximum of the muon track $\chi_{\text{track}}^2/\text{ndf}$
- the minimum of the logarithm of the kaons **ProbNNk**
- the minimum of the logarithm of the muons **ProbNNmu**
- the logarithm of the J/ψ decay vertex $\chi_{\text{vtx}}^2/\text{ndf}$
- the ϕ transverse momentum, $p_T(\phi)$
- the B_s^0 transverse momentum, $p_T(B_s^0)$
- the B_s^0 decay vertex $\chi_{\text{vtx}}^2/\text{ndf}$
- the logarithm of the B_s^0 impact parameter χ_{IP}^2
- the logarithm of the B_s^0 DecayTreeFit χ_{DTF}^2

The latter uses a DTF in which the J/ψ mass is constrained to the known value [172], and there is also a constraint on the position of the PV.

Initially the variables included the minimum transverse momentum of either the muon or kaon candidates and the minimum of the track ghost probabilities assigned by the tracking algorithms. However, it was discovered that using

the momenta of any final-state particle as a discriminating variable was biasing the angular spectrum of the resulting distribution. Furthermore, the track ghost probability was introducing inconsistency between validation epochs. The latter effect was possibly due to the the ghost probability distributions' highly non-linear and discontinuous structure, making it more probable for an MVA to converge into isolated local minima. Figure 6.5.1 shows the results of the

Rank	Variable	Importance
1	the logarithm of the B_s^0 DecayTreeFit χ_{DTF}^2	1.499×10^{-1}
2	the minimum of the logarithm of the kaons ProbNNk	1.499×10^{-1}
3	the B_s^0 decay vertex $\chi_{\text{vtx}}^2/\text{ndf}$	1.094×10^{-1}
4	the logarithm of the B_s^0 impact parameter χ_{IP}^2	9.843×10^{-2}
5	the maximum of the kaon track $\chi_{\text{track}}^2/\text{ndf}$	9.476×10^{-2}
6	the ϕ transverse momentum, $p_T(\phi)$	9.461×10^{-2}
7	the maximum of the muon track $\chi_{\text{track}}^2/\text{ndf}$	8.687×10^{-2}
8	the B_s^0 transverse momentum, $p_T(B_s^0)$	7.586×10^{-2}
9	the logarithm of the J/ψ decay vertex $\chi_{\text{vtx}}^2/\text{ndf}$	7.465×10^{-2}
10	the minimum of the logarithm of the muons ProbNNmu	6.557×10^{-2}

Table 6.7 *BDTG3 variable importance ranking (top variable is best ranked).*

ROC curve comparison. It is clear that each of the MVA algorithms performs sufficiently and almost equivalently. As such, the preference falls towards the most straightforward algorithm that performs well when its selection threshold cut is optimised. The figure of merit (FOM) used to optimise the MVA response is given by:

$$\text{FOM} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} \quad (6.11)$$

where w_i are the event weights (specifically signal ${}_s\text{Weights}$) obtained by an ${}_s\text{Plot}$ to the $m(J/\psi K^+ K^-)$ distribution with the same model from Section 6.4.1: a double Gaussian PDF for the signal mode and an exponential PDF for the background. As can be seen from Equation C.16 and Equation C.18, the FOM increases with the absolute yield of the signal ${}_s\text{Weights}$ and is inverse proportional to the size of the fit uncertainty. Thus, the FOM optimises both for larger signal model yield and lower uncertainty due to background contributions.

It was determined that the BDTG3 variant was sufficient and performing well in these metrics. It was selected preferentially for its simplicity to reduce the possibility of introducing unseen biases in selection. The signal and background distributions of the variables used for the training are shown in Figure 6.17. The

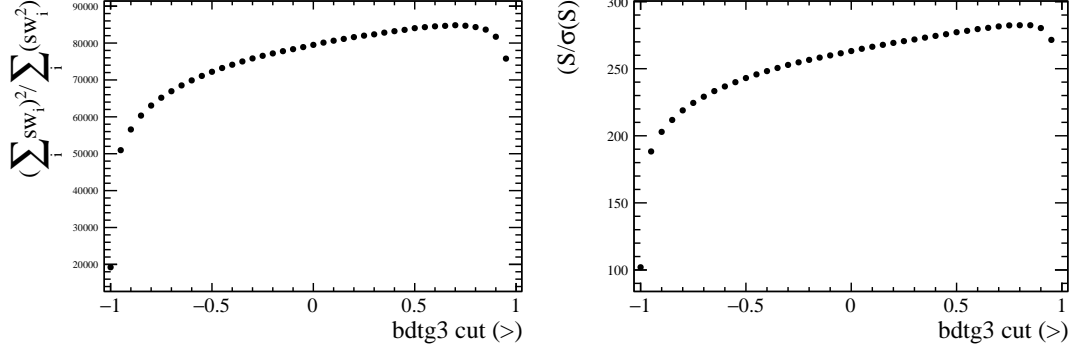


Figure 6.15 *Distribution of the figure of merit (Equation 6.11) used to optimise the threshold cut on the BDTG3 response (left) and distribution of signal yield divided by its uncertainty (right).*

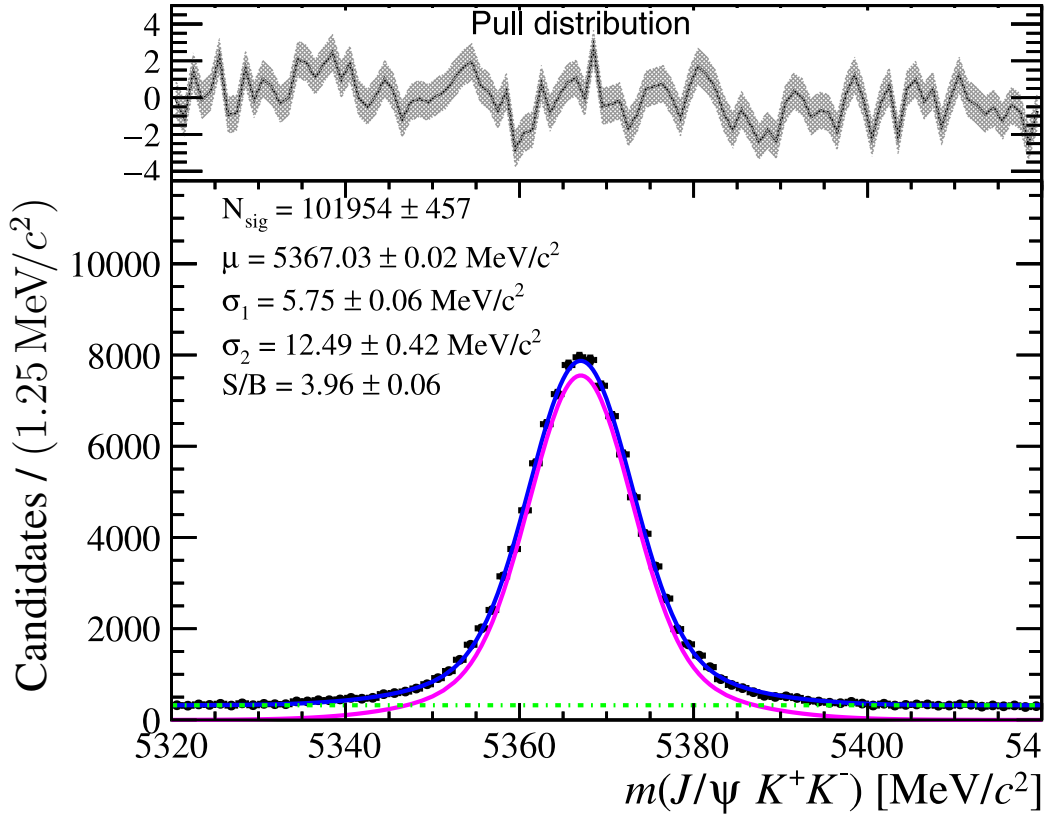


Figure 6.16 *Distribution of $m(J/\psi K^+ K^-)$ in 2016 data after applying all selections, including the MVA's optimised cut. The solid blue line shows the total fit with the signal and combinatorial background described by the pink and green lines, respectively. The parameters of the fit are not constrained.*

output of the BDTG3 classifier is shown in Figure 6.18, where a Kolmogorov-Smirnov[173] test is also performed between the labelled training and testing

sets. Figure 6.15 shows the FOM optimisation curve for BDTG3, where the optimum cut was estimated to be BTDG3 response > 0.58 . Table 6.7 shows the ranked performance of each of the variables in the final training. Figure 6.19 gives a comparison of the BDTG3 output in simulation and $_s$ Weight-ed data. A good match is observed, indicating that the various corrections have worked and subsequent efficiencies evaluated from the simulated samples can be trusted.

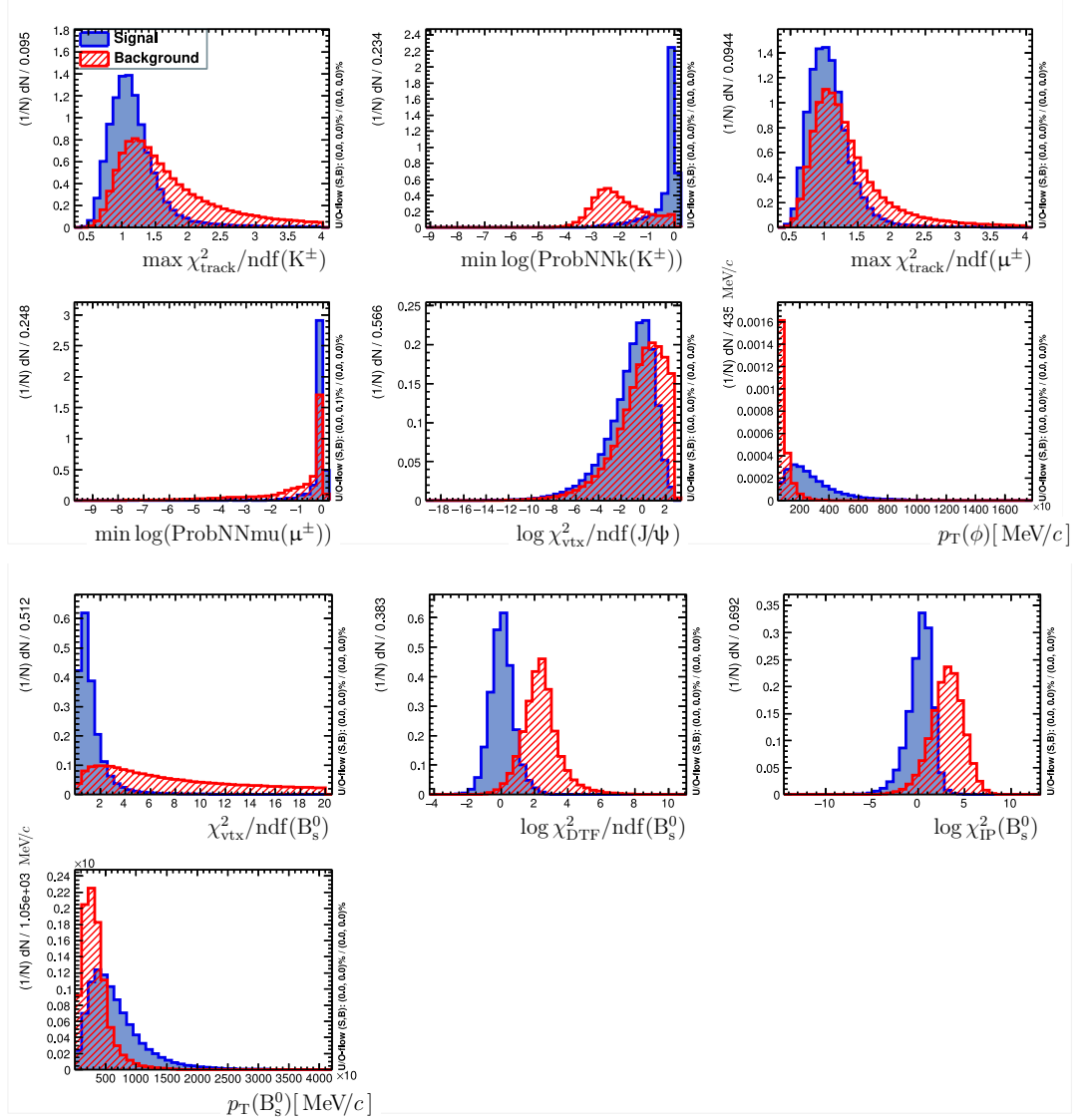


Figure 6.17 *Distributions of MVA training variables in the signal (blue) and background (red) samples.*

TMVA overtraining check for classifier: BDTG3

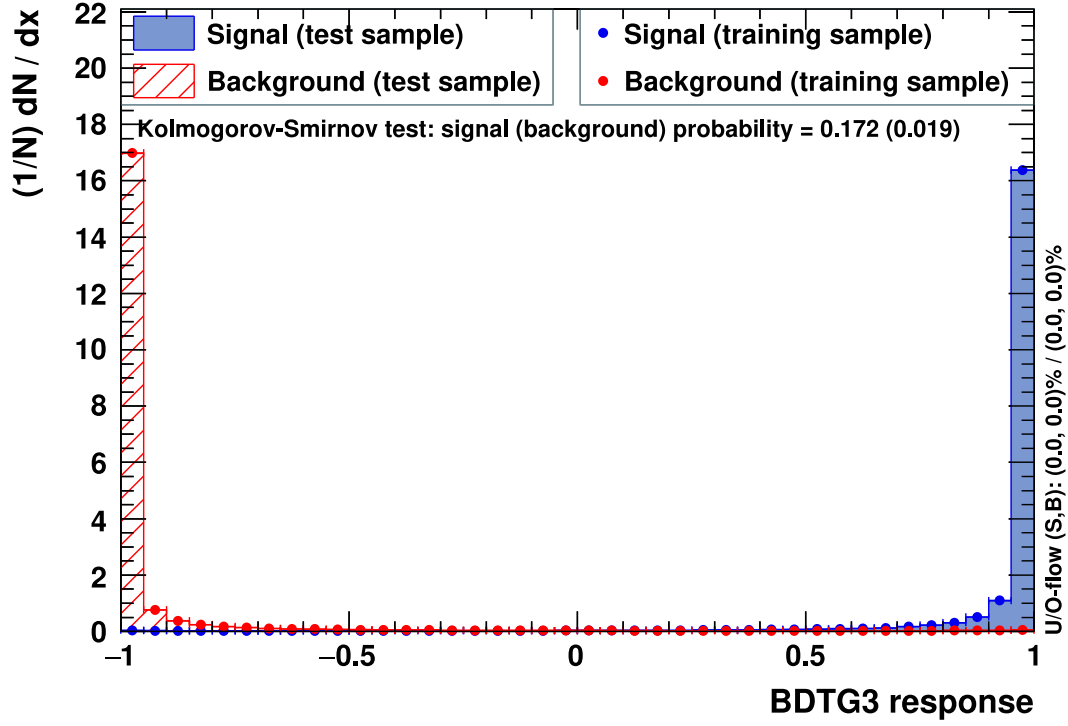


Figure 6.18 A two-sample Kolmogorov-Smirnov[173] between the training and testing samples of the BDTG3 MVA algorithm.

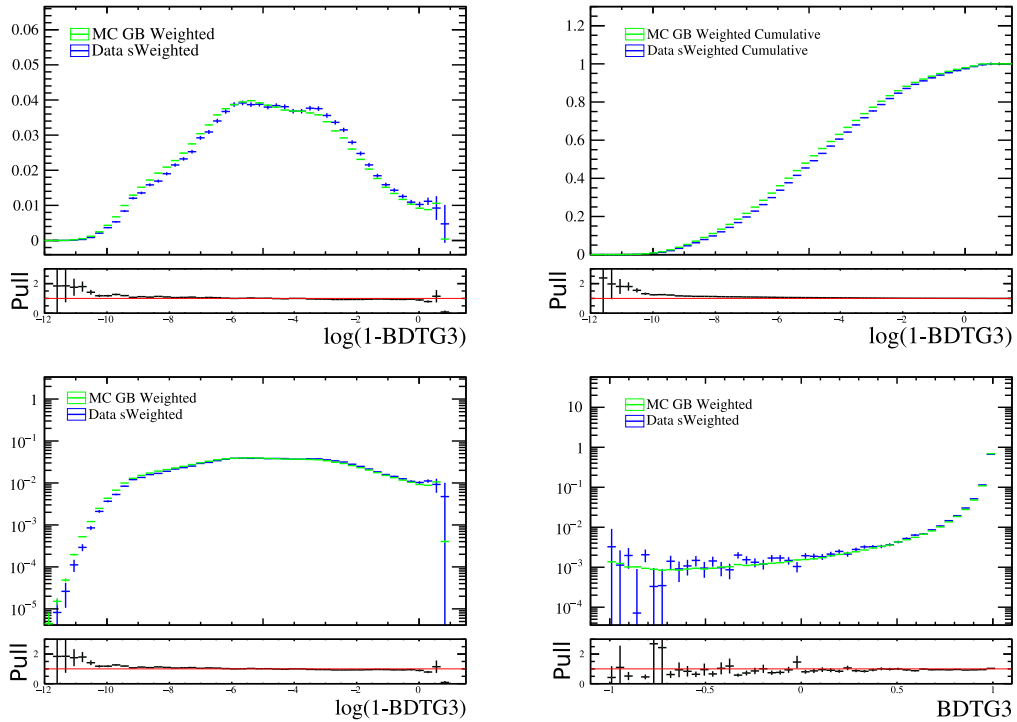


Figure 6.19 Comparison of BDTG3 response distribution in simulation (green) and s Weight-ed data (blue).

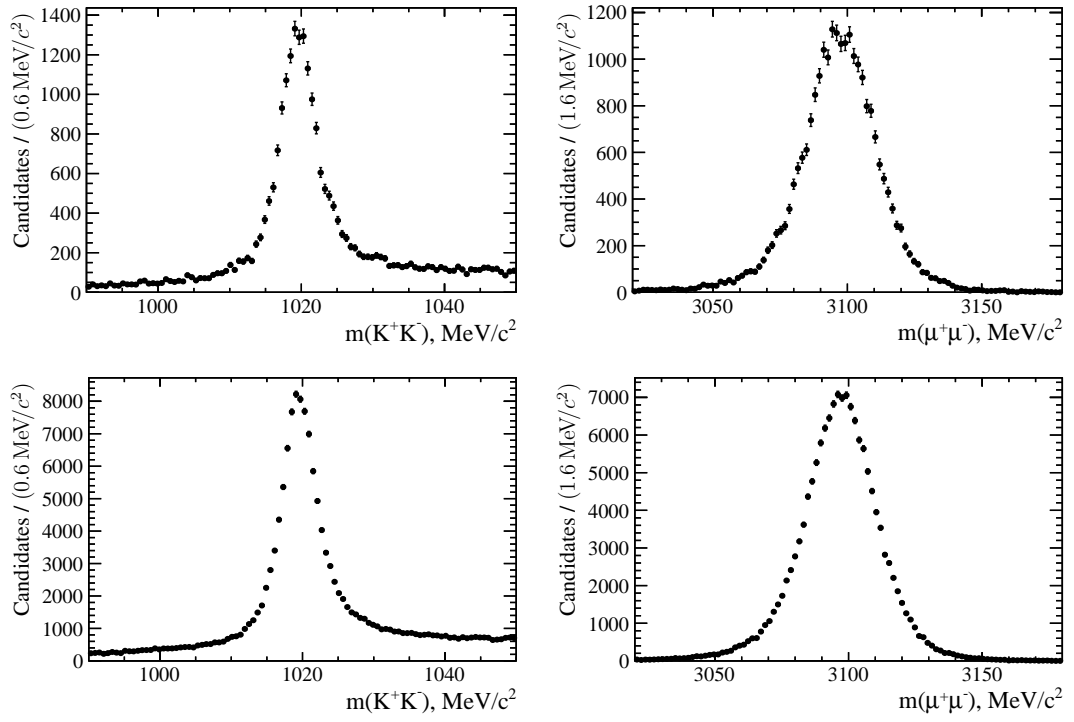


Figure 6.20 *Distributions of the $m_{K^+K^-}$ (left) and the $m_{\mu^+\mu^-}$ (right) obtained in 2015 (top) and 2016 (bottom) data after applying all selections, including MVA's optimised cut.*

Chapter 7

Measurement of the CP -violating phase ϕ_s using $B_s^0 \rightarrow J/\psi \phi$ decays with LHC Run 2 data

7.1 Introduction

The Standard Model is currently the most accurate and precise description of the interactions fundamental to the Universe, excluding gravitation. As discussed in Chapter 2, the parameter ϕ_s is a suitable test for the precision of the Standard Model predictions. Thus, any significant deviations from the ϕ_s^{SM} value of $-2\beta_s$ would indicate the influence of Beyond the Standard Model (BSM) physical effects. In turn, the experimental value of ϕ_s , depending on the precision of its measurement, could be used to pinpoint possible sources of such BSM Physics and differentiate between the accuracy of new models.

This realisation puts a heavy focus on the precision of the ϕ_s determination. However, its sensitivity, which generally would be considered an advantage towards being a suitable SM test, contributes to the complexity and difficulty of its experimental measurement. Multiple purely experimental effects have to be taken into account. Examples are the efficiency in selecting signal-like events from all the data and removing data pollution from background sources (Section 6.3), the uncertainties related to measurements of the decay time (Section 7.3) as well as precision in determining the acceptance of the detector (Section 7.5) among others.

A time-dependent angular analysis of the $B_s^0 \rightarrow J/\psi \phi$ decay is presented below to fulfil the requirements of such a precision measurement. Furthermore, The methods used to fit the model to the multi-dimensional dataset and extract the value for ϕ_s are detailed.

The work has been published in Reference [5], and the results are discussed in Section 7.9.

7.2 Invariant mass fit

As mentioned in Section 2.4.4, the adopted approach¹ for extracting the physics parameters of interest has been to fit the signal PDF to the unbinned decay-time and angular distributions (Equation 2.55). The events are weighted to statistically subtract the background components using an `_sPlot`[174] with the invariant mass of the $J/\psi K^+ K^-$ combination as the discriminating variable (Section C.2). Then the time-dependent angular log-likelihood fit is performed on the weighted data. The invariant mass model used to describe the signal, combinatorial background and reflection backgrounds are detailed in the following sections.

There have been two significant changes to the mass model used in this analysis compared to previous ones of this type¹. Firstly, the invariant mass of the $J/\psi K^+ K^-$ combination, $m(J/\psi K^+ K^-)$, was previously estimated using a DTF where the J/ψ mass is constrained to the known value [172], but in this analysis, a further constraint is added. As the latter can potentially bias the decay time and angles estimation, a study was performed to judge the effects of including this constraint. The second significant change in the mass model is that a new signal PDF is used to model the signal, namely, a *Crystal Ball* [86–88] (CB) function with the per-event mass error of the $m(J/\psi K^+ K^-)$ as a conditional observable.

7.2.1 Primary vertex constraint in J/ψ mass estimation

To study the effect of adding the PV constraint in the mass estimation, the data sample was split into several sets based on B_s^0 decay time. Then the results of a fit to the mass in each is investigated for any differences. The bin edges in B_s^0 decay

¹See previous iterations of this analysis[14, 15]

time are the following: [0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1, 2, 4, 15] ps. The mass model used is simpler as the absolute mass shape is less important than the relative differences between its shape in each decay time bin. The mass is fit with a double Gaussian for the signal, where the fraction between the two Gaussians is common between the bins and an exponential for the background. The values

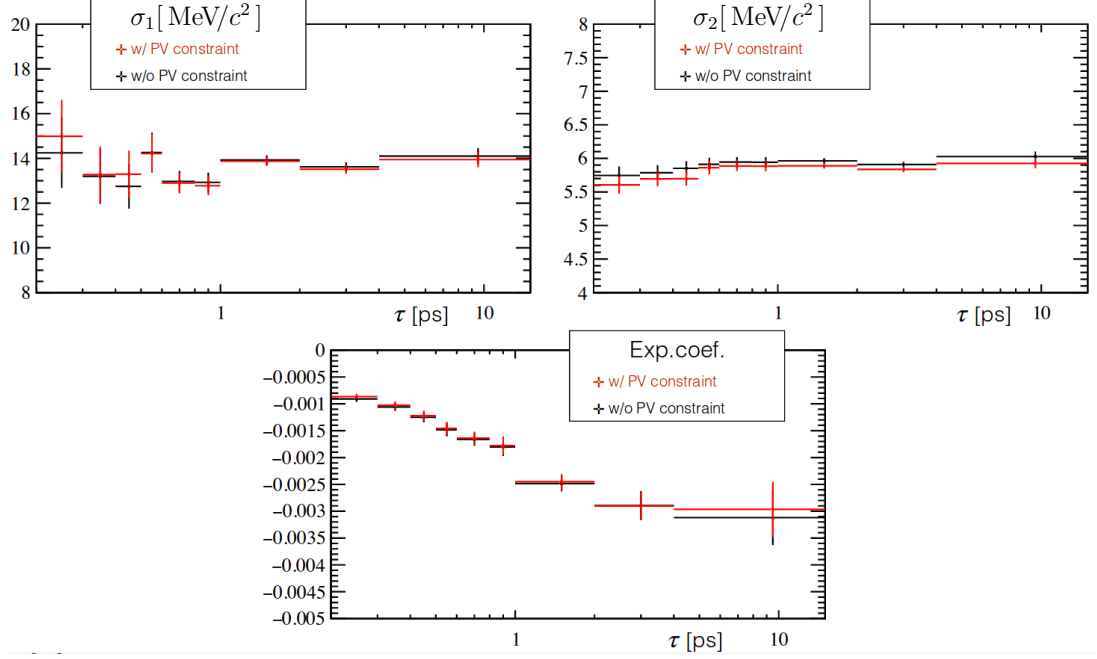


Figure 7.1 Values of the resolutions of the two signal Gaussians σ_1 (left) and σ_2 (right) and the coefficient of the exponential used to model the background (bottom) obtained by fitting the $m(J/\psi K^+ K^-)$ in bins of B_s^0 decay time. In red are the values obtained including a PV constraint, and in black are the values without [175].

of the resolutions of the two Gaussians, σ_1 and σ_2 , and the coefficient of the exponential distribution as a function of B_s^0 decay time are shown in Figure 7.1. It can be seen that the resolutions of the Gaussians are relatively constant, while the exponential coefficient varies significantly. That will necessitate a systematic uncertainty to be assigned to account for the correlation between mass and decay time. However, the variation is entirely consistent between the PV constrained fit and the one not including the constraint. That indicates the PV constraint should not have any significant impact on biasing the data sample. Thus, it was decided to include the PV constraint in the final fit to have better mass and angular resolutions (Section 6.4).

7.2.2 Peaking backgrounds and their subtraction

While the MVA and selection techniques discussed in Section 6.3 and Section 6.5 are very efficient in removing most of the combinatorial background, other relevant backgrounds come from sources more closely mimicking the signal spectrum of events. That is because their particle content and kinematics either completely or partially overlap with $B_s^0 \rightarrow J/\psi K^+ K^-$. For example, $B_d^0 \rightarrow J/\psi K^+ K^-$ and $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$. Both include the $J/\psi K^+ K^-$ final state, but their physical characteristics will be different to what the current analysis concentrates upon, namely a pair of B_s^0 mesons at production and a mostly isolated system, given the back-to-back centre-of-mass production. A

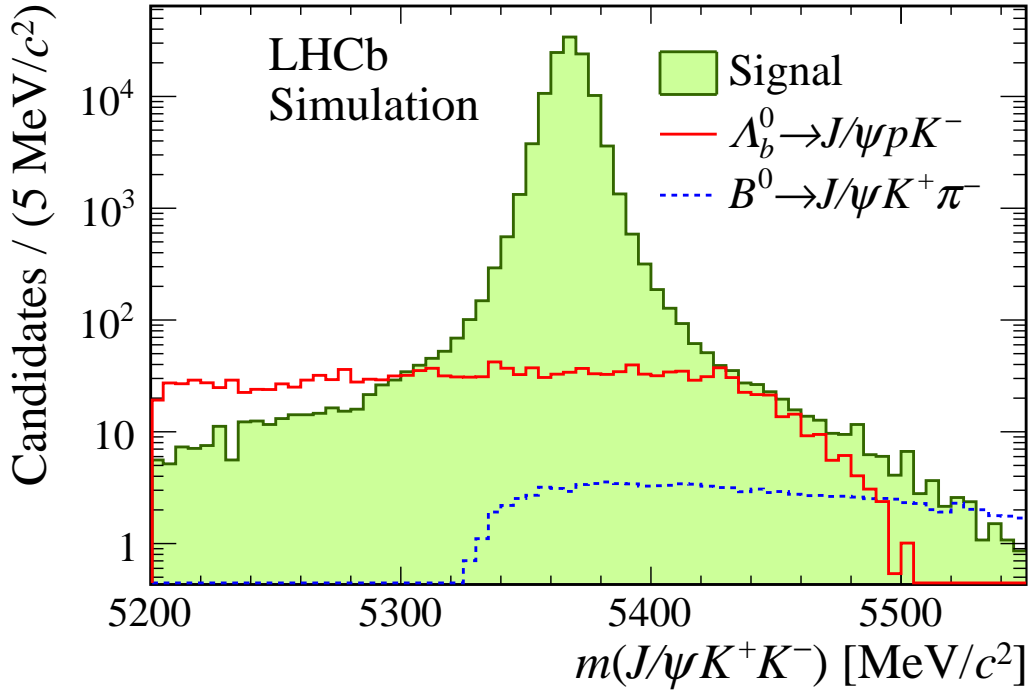


Figure 7.2 *Distribution of the invariant mass of B_s^0 candidates, selected from simulated $B_s^0 \rightarrow J/\psi K^+ K^-$ (green filled area), $\Lambda_b^0 \rightarrow J/\psi p K^-$ (solid red line) and $B_d^0 \rightarrow J/\psi K^+ \pi^-$ (dotted blue line) decays. The distributions are weighted to correct differences in the kinematics and the resonance content between simulation and data [5].*

systematic uncertainty is assigned because of $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ particular case (Section 7.8.1), while the $B_d^0 \rightarrow J/\psi K^+ K^-$ is discussed further below.

Other such contributions come from decay modes which contain a single misidentified final state particle and appear as $J/\psi K^+ K^-$ in the data. Specifically, it was determined that significant contributions come from $B_d^0 \rightarrow J/\psi K^{*0}$ and

$\Lambda_b^0 \rightarrow J/\psi p K^-$ where either a pion or a proton are misidentified as a kaon. These mass spectra appear near the B_s^0 mass, as seen in Figure 7.2. The remaining number of events after the MVA selection are 5200 and 350 from $\Lambda_b^0 \rightarrow J/\psi p K^-$ and $B_d^0 \rightarrow J/\psi K^{*0}$, respectively.

Thus, special care is taken to events with $m(J/\psi K^+ K^-)$ invariant mass within 15 MeV of the known values for Λ_b^0 and B_d^0 [172]. A further requirement is added on such events to be accepted in the final sample of events. Namely, more strict requirements on the **ProbNN** values returned from the PID algorithms. Specifically, it was determined that events falling within the above-mentioned mass regions would be discarded (vetoed) if either of their kaons does not meet the following criteria:

- $m(J/\psi K^+ K^-) \in (5619.51 \pm 15.00 \text{ MeV})$:
 - $\text{ProbNNp}(K_i) > 0.7$
 - $\text{ProbNNp}(K_i) > \text{ProbNNp}(K_j)$, where $i \neq j$
- $m(J/\psi K^+ K^-) \in (5279.63 \pm 15.00 \text{ MeV})$:
 - $\text{ProbNNpi}(K_i) > 0.7$ or $\text{ProbNNk}(K_i) < 0.35$
 - $\text{ProbNNpi}(K_i) > \text{ProbNNpi}(K_j)$, where $i \neq j$

In the case of mis-ID events coming from Λ_b^0 , either kaon is checked whether its assigned probability of being a proton, $\text{ProbNNp}(K_i)$, is above a threshold and is also higher than the same probability of the other kaon in the event. The latter is done to prevent legitimate events from being vetoed. Given the kinematics of the decay mode, the chance of both kaons having a high probability of being protons, and thus, both being misidentified, is negligible. Similarly, for the B_d^0 case, the kaons are tested using thresholds for their **ProbNNk** and **ProbNNpi**. At the same time, care is also taken to exclude instances with both kaons being considered as misidentified. In both cases for B_s^0 and B_d^0 , the latter also has a side benefit of making sure no events can be processed twice when evaluating the veto on the data sample. The **ProbNN** thresholds were determined from simulation studies [175].

After removing events matching the veto criteria, it was estimated that in the sample continue to remain 120 decays coming from B_d^0 and around 1600 ± 160 coming from Λ_b^0 , given veto efficiencies. The B_d^0 contribution is neglected, and

a systematic uncertainty is assigned. At the same time, the Λ_b^0 background is statistically subtracted by injecting $_{\text{s}}\text{Weight}$ -ed simulated decays with negative weights into the data sample, whose sum of weights has been normalised to the predicted yield of 1600. Moreover, this is done before the $_{\text{s}}\text{Plot}$ fit for the invariant mass. The simulated events have been reweighted to match the data properties similarly to the procedure described in Section 6.4. Using this approach by injecting simulated events also ensures that the correlations between the mass and angular variables are preserved.

As for the peaking background related to events from $B_d^0 \rightarrow J/\psi K^+ K^-$ decay mode, the strategy is to include it in the final mass model as a background during the fitting procedure. Given its small yield (biased 2016: 68 ± 0.41 , unbiased 2016: 211 ± 0.89 [175]), it is modelled only with a Gaussian. The mean is fixed based on the measured difference between the B_s^0 and B_d^0 mass [172]. Namely, the difference between the B_s^0 and B_d^0 masses is kept constant during the fit, while the resolution is fixed to $7 \text{ MeV}/c^2$, obtained by a fit to the $B_d^0 \rightarrow J/\psi K^{*0}$ control channel mass (Figure 7.6).

7.2.3 Per-event-error mass model

The $_{\text{s}}\text{Plot}$ method requires that the variable used for background subtraction is not correlated to any of the variables to which the $_{\text{s}}\text{Weights}$ are applied [174]. However, it is found that such a correlation exists between the signal mass shape and the cosine of the muon helicity angle, $\cos \theta_\mu$. That is because the estimated invariant-mass resolution is observed to depend on the measured muon transverse momentum [175]. Thus, it was decided to include the mass resolution parameter as a per-event parameter of the likelihood fit. Meaning that the likelihood calculations for each event will no longer share a common mass resolution value (a fixed parameter). Instead, each will have an independent mass resolution, referred to as the per-event mass error as a conditional observable.

Essentially, the aim is to account for this correlation in the overall mass model by using the estimated mass resolution from data as a proxy for the mass shape dependence on muon momentum. As such, the $_{\text{s}}\text{Plot}$ can natively assign correct $_{\text{s}}\text{Weights}$ to events based on their likelihood as the likelihoods will be adjusted by the mass resolution.

The final mass model thus fits the signal distribution of $m(J/\psi K^+ K^-)$, where the

mass of the J/ψ and the PV of the final products are constrained. The fit uses a CB PDF with the mass-error of that mass estimation as a conditional observable on a per-event basis. Furthermore, it was found by simulation studies that a quadratic dependence of the total mass shape uncertainty, σ_{CB} , on the per-event mass error, σ_m , is best suited to describe the signal [175]:

$$\sigma_{\text{CB}}(\sigma_m) = a_1 \sigma_m + a_2 \sigma_m^2 \quad (7.1)$$

where a_1 and a_2 are free parameters determined from the fit to data. The tail parameters of the CB distribution are fixed from fits to simulated events. The sample is divided into 24 independent subsamples corresponding to 6 bins in $m(K^+K^-)$ with boundaries at 990, 1008, 1016, 1020, 1024, 1032, 1050 MeV/c^2 , to the biased and unbiased trigger categories (Section 6.3.1), and the year of data-taking. The splitting is done to accommodate trigger differences between data-taking years, decay-time biasing and non-biasing triggers as well as the fact that the K^+K^- S-wave is assumed to be relatively flat in each of the $m(K^+K^-)$ bins. The fit is performed in each of the 24 subsamples and the invariant mass probability density function (PDF) is independent for each subsample. The parameters of the PDF between the subsamples are not correlated. An exponential is used to model the combinatorial background.

After the full selection (Chapter 6) is applied, the fit yield is about 102 000 and 15 000 $B_s^0 \rightarrow J/\psi \phi$ decays in 2016 and 2015 data, respectively (Figure 7.3).

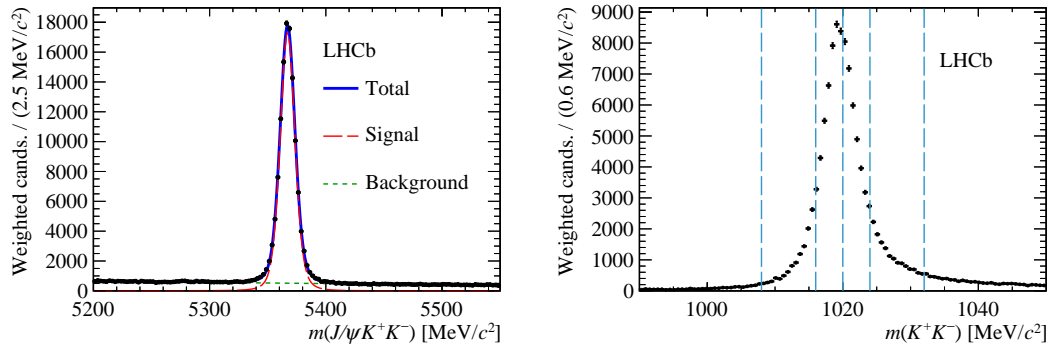


Figure 7.3 *Left: Distribution of the invariant mass of B_s^0 candidates selected from $B_s^0 \rightarrow J/\psi K^+K^-$ data. The signal component is shown with a long-dashed red line, the background is shown with a dashed green line, and the total fit is the solid black line. Right: Distribution of K^+K^- invariant mass where the background is statistically subtracted using the $_s\text{Plot}$ method. The dashed blue line shows the boundaries of the six $m(K^+K^-)$ bins. [5]*

7.3 Decay time resolution

The magnitude of the decay-time resolution and its uncertainty are key factors constraining the precise estimation of ϕ_s . The effect of insufficient decay-time resolution will result in being increasingly difficult to determine the precise time of B_s^0 mesons' decay, which will produce a sort of *smearing* of the observed waveform in time. Thus, it is crucial to understand and estimate the decay-time resolution and its effects on the data as best as possible.

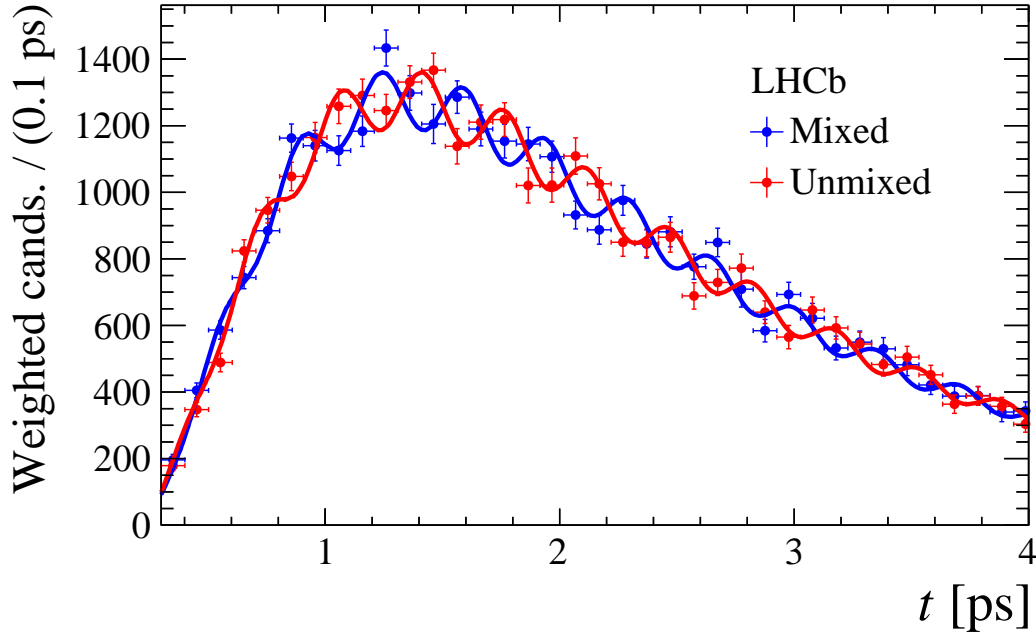


Figure 7.4 *Distribution of the decay time of $B_s^0 \rightarrow D_s^- \pi^+$ candidates tagged as mixed and unmixed with the projection of the fit result as described in Section 7.7.*

7.3.1 Model and parametrisation

The decay-time resolution is modelled as a Gaussian distribution centred at zero with a width σ_{eff} . The width, σ_{eff} , is determined by using a sample constructed from combinations of J/ψ , K^+ and K^- reconstructed candidates that originate predominantly in the primary interaction. This sample later referred to as the *prompt component* or *prompt events*. The sample of the combined candidates $J/\psi K^+ K^-$ is, thus, the prompt sample and is selected as described in Chapter 6 with the exception of the lower bound on decay time. Also,

the candidates are formed from data passing a different trigger line, namely `Hlt2DiMuonJPsiDecision`, which is heavily prescaled². The trigger is applied during the HLT2 stage on the J/ψ candidate.

By definition, the prompt component has zero decay time, so any non-zero decay time in the data is most likely due to detector effects such as resolution. Thus, given a satisfactory model considering the external contributions to a manageable degree, the detector decay-time resolution can be estimated.

The observed distribution of the prompt component decay time around zero is modelled by a Dirac delta function [176]. The other contributions to the total distribution are modelled separately. There is a contribution to later decay times from J/ψ mesons coming from b hadron decays. Also, there is a small contribution from events reconstructed with a mis-identified (wrong) PV assigned, which results in a background to this model because the estimated decay times are incorrect. An event could have its PV wrongly assigned due to LHCb PV fitting algorithms. For example, multiple pp interactions or 5 or 6 prong b-hadron decays can sometimes be reconstructed as a PV and some tracks might end up wrongly associated with them [177]. The yield of these events is found to be about 0.5 % in the prompt sample (See Appendix C.8).

To estimate the decay-time resolution, the sum of the prompt and b hadron components is convolved with a triple-Gaussian function resolution function:

$$\mathcal{R}(t) = \sum_{i=1}^3 f_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-\frac{(t - \mu)^2}{2\sigma_i^2} \right] \quad (7.2)$$

where $\sum_i f_i = 1$, μ describes a bias from zero in the decay-time measurement, and σ_i are the individual Gaussian widths. In this analysis, μ is fixed to zero and assigned a systematic uncertainty. The calibration sample is then divided into 11 subsamples based on per-candidate decay time, δ_t , and the model is fit to it. That allows extracting the parameters of the decay-time resolution function.

The effective dilution of the B_s^0 - \bar{B}_s^0 oscillation amplitude is estimated in each bin of δ_t as:

$$\mathcal{D} = \sum_{i=1}^3 f_i \exp \left[-\sigma_i^2 \Delta m_s^2 / 2 \right] \quad (7.3)$$

²only 10 % of candidates passing the requirements are kept, chosen randomly

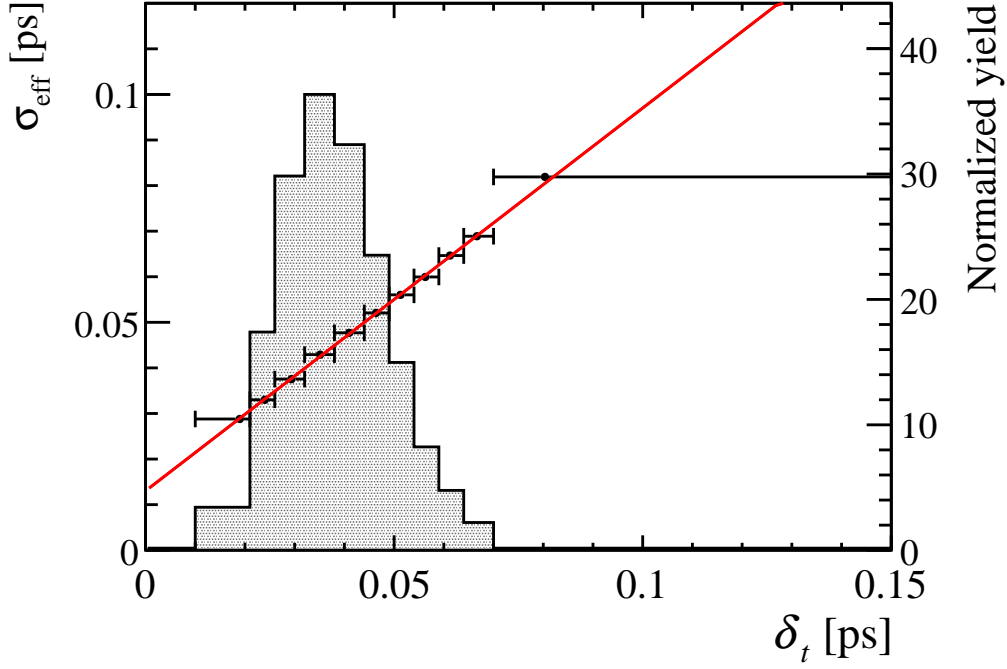


Figure 7.5 *Standard deviation of the effective single-Gaussian decay time resolution, σ_{eff} , as a function of the estimated per-candidate decay-time uncertainty, δ_t , obtained from a prompt $J/\psi K^+ K^-$ sample. The red line shows the result of a linear fit, while the grey area shows the distribution of δ_t in the background-subtracted $B_s^0 \rightarrow J/\psi \phi$ data sample.*

which is later used to evaluate an effective single-Gaussian width:

$$\sigma_{\text{eff}} = \sqrt{(-2/\Delta m_s^2) \ln \mathcal{D}} \quad (7.4)$$

The effective single-Gaussian resolution function constructed with σ_{eff} has the same dampening effect on the B_s^0 oscillation as the triple-Gaussian model [5]. Figure 7.5 shows the variation of σ_{eff} in terms of δ_t .

7.4 Decay time efficiency

The reconstruction and selection efficiencies depend on B_s^0 decay time. That is due to secondary vertex displacement requirements (Chapter 6) applied to signal tracks and an observed decrease in reconstruction efficiency for tracks with large IP with respect to the beamline [178]. In this analysis, a novel approach is taken towards estimating the decay time efficiency compared to the previous

iteration [14].

Specifically, the decay mode $B_d^0 \rightarrow J/\psi K^+ \pi^-$ is taken as a control sample. The decay mode is kinematically similar to the signal $B_s^0 \rightarrow J/\psi K^+ K^-$ decay. Since the decay width difference, $\Delta\Gamma_d$, of the B_d^0 system is measured to be consistent with zero [172], this mode is a great candidate for decay-time efficiency estimation. It allows ignoring any effects of B_d^0 oscillation during reconstruction. The candidates reconstructed as $B_d^0 \rightarrow J/\psi K^+ \pi^-$ are assumed to have a purely exponential decay time distribution with lifetime $\tau_{\text{data}}^{B_d^0} = 1.520$ ps [179].

However, there is no reason to expect a direct relationship between the decay-time efficiency for $B_s^0 \rightarrow J/\psi K^+ K^-$ decays, $\varepsilon_{\text{data}}^{B_s^0}$, and the efficiency for $B_d^0 \rightarrow J/\psi K^+ \pi^-$, $\varepsilon_{\text{data}}^{B_d^0}$. On the other hand, since the reconstruction and selection for both decays are the same in data and simulation, one can expect that the following is approximately true:

$$\frac{\varepsilon_{\text{data}}^{B_s^0}}{\varepsilon_{\text{sim}}^{B_s^0}} = \frac{\varepsilon_{\text{data}}^{B_d^0}}{\varepsilon_{\text{sim}}^{B_d^0}} \quad (7.5)$$

where $\varepsilon_{\text{sim}}^{B_s^0}$ and $\varepsilon_{\text{sim}}^{B_d^0}$ are the decay-time efficiencies in simulated events and reconstruction of $B_s^0 \rightarrow J/\psi K^+ K^-$ and $B_d^0 \rightarrow J/\psi K^+ \pi^-$ decays, respectively. The ratios between data and simulation on both sides cancel out any decay-specific differences and leave only the differences due to the real and simulated detector scenarios. The expression can be rearranged to extract the $B_s^0 \rightarrow J/\psi K^+ K^-$ decay-time efficiency in data:

$$\varepsilon_{\text{data}}^{B_s^0} = \varepsilon_{\text{data}}^{B_d^0} \times \frac{\varepsilon_{\text{sim}}^{B_s^0}}{\varepsilon_{\text{sim}}^{B_d^0}} \quad (7.6)$$

which are all functions of decay time, as mentioned previously. Thus, the B_s^0 decay-time efficiency is determined by a simultaneous fit to background-subtracted data and simulation with the following constraint:

$$\varepsilon_{\text{data}}^{B_s^0}(t) = \varepsilon_{\text{data}}^{B_d^0}(t) \times \frac{\varepsilon_{\text{sim}}^{B_s^0}(t)}{\varepsilon_{\text{sim}}^{B_d^0}(t)} \quad (7.7)$$

The B_d^0 samples are reconstructed and selected similarly to the signal channel (Chapter 6) to control any relevant effects. The data is reweighted to match the B_s^0 data in p and p_T . The B_d^0 data sample is also investigated for significant peaking backgrounds. However, after applying a separate requirement to remove misidentified pions, only a small sample of $B_s^0 \rightarrow J/\psi K^+ \pi^-$ is left [175]. Figure 7.6

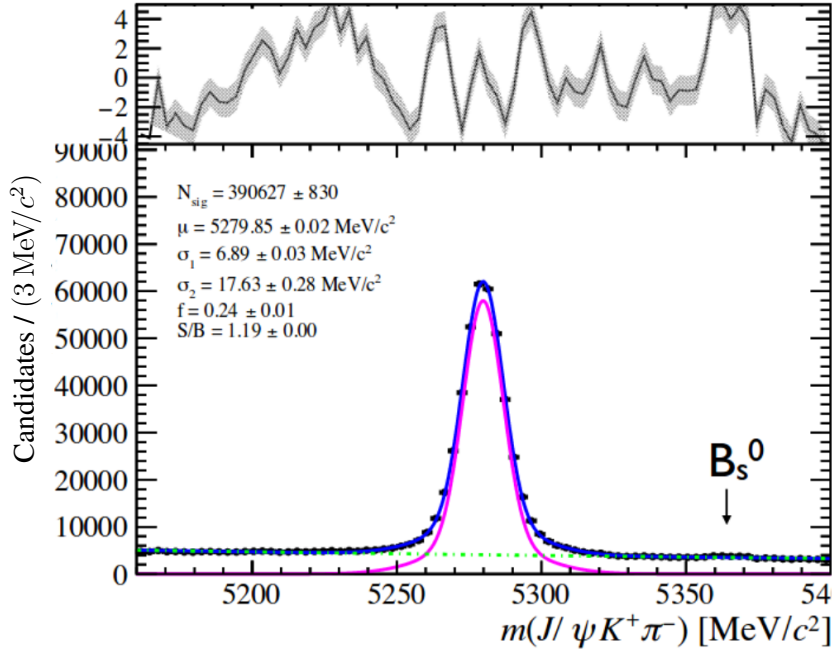


Figure 7.6 *Distribution of $m(J/\psi K^+ \pi^-)$ in 2016 data after applying all selections, including MVA's optimised cut. The solid blue line shows the total fit with the signal and combinatorial background described by the pink and green lines, respectively. The parameters of the fit are not constrained.*

shows the invariant mass fit to $B_d^0 \rightarrow J/\psi K^+ \pi^-$ data after all selection. As can be seen, the B_s^0 component is barely visible. To remove it entirely, all events with $m(J/\psi K^+ \pi^-) > 5350 \text{ MeV}/c^2$ are excluded. The result of the fit is used to background-subtract the B_d^0 data.

The efficiency functions are all modelled with cubic splines with the first node fixed at zero. The other nodes are chosen in order to ensure uniform yield in each subsample: nodes are at 0.00, 0.30, 0.58, 0.91, 1.35, 1.96, 3.01, 7.00 ps. The PDF used for the maximum-likelihood fit comprises the efficiency functions multiplied by a single exponential, modelling the purely exponential decay. The exponential decay model is also convolved with a single Gaussian centred at zero to include decay-time resolution (Section 7.3). As a cross-check, a separate study is performed using a B_s^0 simulated sample generated with $\Delta\Gamma_s = 0 \text{ ps}^{-1}$ to adhere to the single exponential decay model. The difference to the nominal fit is used to assign as a systematic. The decay time acceptance is obtained separately for the data-taking periods 2015 and 2016 and two different trigger paths.

7.5 Angular efficiency

Similarly to decay time efficiency, the LHCb detector's reconstruction and subsequent selection introduce an efficiency in terms of the angular distributions (helicity angles: $\theta_K, \theta_\mu, \phi_h$). This efficiency is modelled as a function of Ω : $\varepsilon_\Omega \equiv \varepsilon_a(\Omega) \equiv \varepsilon_a(\theta_K, \theta_\mu, \phi_h)$, where the subscript a signifies that the efficiency is a function of the helicity angles, or later referred to as an *angular efficiency*.

7.5.1 Angular efficiency model and parametrisation

The angular efficiency, $\varepsilon_a(\Omega)$, is a three-dimensional function making it increasingly difficult to explicitly model or directly include in the final fit. Thus, a different strategy is undertaken to account for $\varepsilon_a(\Omega)$ by utilising *angular normalisation weights* [54, 180].

Since the final signal PDF (Equation 7.19) is split in terms of its individual angular contributions, each term, f_k , should in principle be multiplied by the angular efficiency, $\varepsilon_a(\Omega)$, in the final maximum likelihood fit. Let us consider the general case of an efficiency in both time and angles, $\varepsilon(t, \Omega)$, and the condition to find the parameters, Θ , which would solve the maximum likelihood fit condition of the signal PDF, $s(t, \Omega|\Theta)$:

$$-\frac{d}{d\Theta}(\ln \mathcal{L}) = -\frac{d}{d\Theta} \sum_e \ln \frac{s(t_e, \Omega_e|\Theta) \varepsilon(t_e, \Omega_e)}{\iint s(t, \Omega|\Theta) \varepsilon(t, \Omega) d\Omega dt} = 0 \quad (7.8)$$

Details about this definition and maximum likelihood fits can be found in Appendix C. By Equation 2.55, the signal PDF factorises as $s(t, \Omega|\Theta) = \mathcal{A}_k(t|\Theta) f_k(\Omega)$. Furthermore, the term $\varepsilon(t_e, \Omega_e)$ does not depend on the parameters Θ , thus, it is irrelevant to the determination of the maximum likelihood:

$$-\frac{d}{d\Theta} \sum_e \ln \frac{\mathcal{A}_k(t|\Theta) f_k(\Omega)}{\iint \mathcal{A}_k(t|\Theta) f_k(\Omega) \varepsilon(t, \Omega) d\Omega dt} = 0 \quad (7.9)$$

The normalisation integral of the PDF does include the angular efficiency, and it cannot be factored out. For each term in Equation 7.19, a term like $\iint \varepsilon(\Omega, t) f_k(\Omega) d\Omega dt$ needs to be considered in the normalisation. Notice that, in general, the angular acceptance can depend on decay time and flavour tagging.

It is also not guaranteed to factor in time and angles. However, it is found that $\varepsilon(\Omega, t) = \varepsilon_a(\Omega) \times \varepsilon(t)$ is a good approximation, studied separately in Reference [175].

As such, the only place where this efficiency appears in the final fit is in the normalisation, which can be determined only once. The value of these terms is labelled the normalisation weight for each angular term and expressed as:

$$\xi_k = \int \varepsilon_a(\Omega) f_k(\Omega) d\Omega \quad (7.10)$$

where ξ_k is the normalisation weight corresponding to the f_k angular term in Equation 7.19. The weights are extracted from simulated events by comparing the known distributions at event generation to the distributions reconstructed in the simulated detector with the following expression [54]:

$$\xi_k(t_e, q_e) = \frac{1}{N_{gen}} \sum_e^{N_{obs}} \frac{f_k(\Omega_e)}{\mathcal{P}_{gen}(\Omega_e | t_e, q_e)} \quad (7.11)$$

where N_{gen} and N_{obs} are the simulated and reconstructed event yield, respectively. \mathcal{P}_{gen} is the PDF used to generate the simulated events. This method also allows deriving statistical uncertainties, which can later be propagated to the relevant parameters.

Of course, for this to work, the simulated samples need to reproduce the distributions in data very closely. However, simulated samples containing the K^+K^- S-wave contributions were not available for the exact detector reconstruction used in the data. Thus, the available simulation samples were only used to cross-check and validate the procedure [175]. That necessitates the approach to be undertaken with specific extra steps to correct and adjust the differences between data and simulation (Chapter 6). On top of the selection and corrections applied, the simulated samples of $B_s^0 \rightarrow J/\psi \phi$ decays are reweighted further using the same technique from Section 6.4.2 a second time to match the distributions of p_T , p and $m(K^+K^-)$ in data. That is to ensure simulation samples can reproduce the S-wave fraction observed in data.

Several cross-checks are also made using $B_d^0 \rightarrow J/\psi K^+ \pi^-$ data and simulation, selected and reweighted to match as above. Then the angular normalisation weights are computed. Afterwards, the weights are used to extract the polarisation amplitudes of the decay through an unbinned fit to the data. Finally, they are compared to the known values [181] with which they agree. A further

test is done using $B^+ \rightarrow J/\psi K^+$ and the θ_μ helicity angle. Before the angular corrections from simulation, the helicity angle distribution is found to deviate from the expected shape of $1 - \cos^2 \theta_\mu$ by as much as 30 %. Still, after applying the procedure from above, the expected shape is fully recovered to a precision of 0.1 %.

7.6 Flavour Tagging

Flavour tagging [182, 183] is the act of determining the initial flavour of the B_s^0 meson at creation. As already discussed, the time evolution of the B_s^0 - \bar{B}_s^0 system is highly dependent on the flavour state at creation. Thus, tagging the B_s^0 meson becomes essential for the analysis of the system. To tag the B_s^0 flavour, LHCb uses multiple algorithms split into two main classes — same-side (SS) and opposite-side (OS) tagging (Figure 7.7).

The OS tagger takes advantage of the fact that b and \bar{b} quarks are almost exclusively produced in pairs in pp collisions. That assumption allows tagging the flavour of the signal B_s^0 candidate at production by identifying the flavour of the other b hadron in the same event. The OS tagger looks at several pieces of information. It tries to find a semileptonic b decay involving muons or electrons and uses the information about the lepton charge. It adds the knowledge about the charge of a kaon coming from a $b \rightarrow c \rightarrow s$ transition-type decay. Then more data is added about any charged tracks converging into a secondary vertex associated with the other b and the charge of a secondary charm hadron. These data are combined into a weighted average, depending on the transverse momenta of the associated tracks, to form an appropriate pre-tuned response for the B_s^0 signal candidate flavour.

The SS tagger uses a different but somewhat analogous strategy. It concentrates on information involving the hadronisation of the \bar{b} (b) forming the signal B_s^0 (\bar{B}_s^0) candidate instead. Particularly, in the case of a B_s^0 , the s quark needed to hadronise comes from an $s\bar{s}$ pair, which often results in the production of a charged kaon. Depending on the flavour of the s (\bar{s}), a differently charged kaon, K^- (K^+), is produced, and its charge can then be used to extract information about the complementary \bar{s} (s) forming a \bar{B}_s^0 (B_s^0) hadron.

Both the SS and OS taggers have been revisited and improved since Run 1,

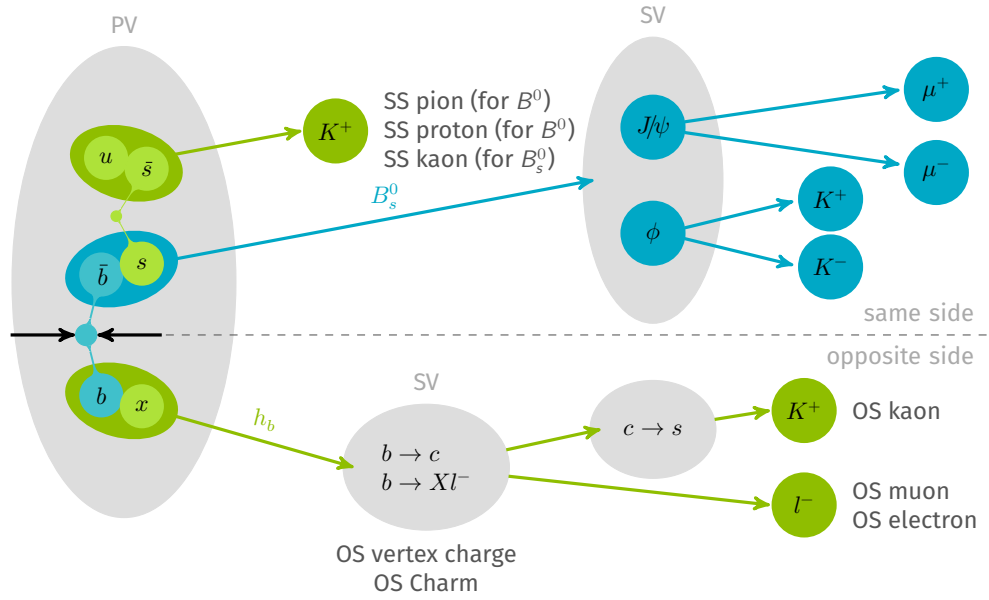


Figure 7.7 Overview of LHCb flavour tagging algorithms showing the decision paths of the same-side (SS) and opposite-side (OS) tagger configurations [184].

obtaining a significantly higher combined tagging performance [185].

Each tagging algorithm provides a response in terms of two quantities: the tag, $q \in (-1, 0, 1)$ and an estimate, η , of the probability the tag is incorrect (*mistag*) [5]. The tag is +1 (−1) for each tagged B_s^0 (\bar{B}_s^0) and 0 for untagged candidates, where the algorithms cannot make a decision. The mistag probability is only defined in the range $\eta \in (0, 0.5)$, because $\eta(q) > 0.5 \equiv 1 - \eta(-q) < 0.5^{\S}$. By this definition, it is clear that in the situation when $\eta(q) = \eta(-q) = 0.5$, the candidate can only be assigned $q = 0$ since both tagging hypotheses are equally likely and no tagging decision can be made.

The tagging algorithms are implemented as BDTs. The SSKaon (Figure 7.7) tagger is trained and optimised on large dedicated simulated samples, and the OS tagger is prepared using a large set of $B^+ \rightarrow J/\psi K^+$ decays. As such, the responses of the BDTs are only an approximation of the true flavour and mistag probability. However, given that the mistag is essentially the predictive factor for the assigned tag, it also directly attenuates (scales) the observed B_s^0 oscillation amplitude as larger tagging uncertainty means less of a significant difference between the B_s^0 and \bar{B}_s^0 hypotheses (Section 7.7.2).

For this reason, the mistag for each tagger is further calibrated with dedicated

[§] $\eta > 0.5$ on q corresponds to opposite decision, $-q$, with $1 - \eta$

collision data samples for tagger-specific decay channels. That is done to relate the estimated mistag probability to the true mistag probability, ω . Each tagger is assigned a tagging power, $\epsilon_{\text{tag}} D^2$, where ϵ_{tag} is the efficiency of the tagger in terms of the fraction of successfully tagged candidates and $D = 1 - 2\omega$ is defined as the dilution induced on the observed B_s^0 oscillation amplitude³.

The OS tagger is calibrated using a sample of $B^+ \rightarrow J/\psi K^+$ events selected in the same way as Chapter 6 and background-subtracted using the already discussed $_s\text{Plot}$ method. The $B^+ \rightarrow J/\psi K^+$ sample is weighted similarly to the procedure described in Section 6.4.2 to match the kinematic distributions of the signal $B_s^0 \rightarrow J/\psi \phi$ sample. The B_s^0 flavour for such candidates is determined by the charge of the kaon and is thus an independent tag. This information is used to relate η and ω by linear regression:

$$\begin{aligned}\omega(\eta) &= \left(p_0 + \frac{\Delta p_0}{2}\right) + \left(p_1 + \frac{\Delta p_1}{2}\right)(\eta - \langle\eta\rangle) \\ \bar{\omega}(\eta) &= \left(p_0 - \frac{\Delta p_0}{2}\right) + \left(p_1 - \frac{\Delta p_1}{2}\right)(\eta - \langle\eta\rangle)\end{aligned}\quad (7.12)$$

where $\omega(\eta)$ and $\bar{\omega}(\eta)$ are the calibrated mistag probabilities for B^+ and B^- mesons, respectively. $\Delta p_{0,1}$ are mistag asymmetries and $\langle\eta\rangle$ is the average estimated mistag probability of the $B^+ \rightarrow J/\psi K^+$ sample. The calibration parameters are determined by a maximum-likelihood fit to η of the following PDF:

$$\mathcal{P}(a|\eta) = (1-a)\bar{\omega}^{(-)}(\eta) + a(1-\bar{\omega}^{(-)}(\eta)) \quad (7.13)$$

where a is a discrete parameter with possible values of 0 or 1 for an incorrect or correct tagging decision, respectively.

The SS tagger is calibrated using a flavour specific channel, $B_s^0 \rightarrow D_s^+ \pi^+$, where a fit to data resolves the B_s^0 - \bar{B}_s^0 oscillation. As already mentioned, the amplitude of the observed (resolved) B_s^0 - \bar{B}_s^0 oscillation is related to the average mistag probability, $\tilde{\omega}$, via the PDF of the decay time distribution of the flavour-tagged $B_s^0 \rightarrow D_s^+ \pi^+$ decays:

$$\begin{aligned}\mathcal{P}(t) &= \epsilon(t) [\Gamma(t) \otimes R(t-t')], \\ \Gamma(t) &= \Gamma_s e^{-\Gamma_s t} \left[\cosh(\Delta\Gamma_s t/2) + q^{\text{mix}}(1 - 2\tilde{\omega}(\eta)) \cos(\Delta m_s t) \right]\end{aligned}\quad (7.14)$$

³the factor of 2 in front of ω is because it applies two times in the amplitude attenuation, *flattening* both mixed and unmixed waveforms and reducing their difference on both sides

where t' and t are true and reconstructed decay time of the B_s^0 meson, respectively, and $\Gamma(t)$ is the B_s^0 decay rate.

7.7 Time-dependent angular fit to data

The fitting procedure and maximum likelihood fit are similar to the ones used by previous analyses [14]. However, a fundamental change in this analysis is that instead of using the absolute value of Γ_s as a parameter, the fit estimates the $\Gamma_s - \Gamma_d$ difference. Consequently, also the decay-time efficiency is treated differently in the fit as well. Specifically, it has been shown by simulated experiments that given a decay-time efficiency obtained using Γ_d as an input parameter, the estimated value of $\Gamma_s - \Gamma_d$ and its associated uncertainty are independent of the Γ_d uncertainty. Thus, the advantage here is that the estimated value of $\Gamma_s - \Gamma_d$ can be updated with new, more precise measurements of Γ_d to obtain better estimates of Γ_s and Γ_s/Γ_d as well.

7.7.1 Statistical background subtraction

Using the $_s$ Plot technique and the method described in Section 7.2, each event candidate is given a weight, $W_i = {}_s\text{Weight}_i$. When applied, the weighting effectively subtracts all backgrounds based on the $m(J/\psi K^+ K^-)$ discriminating variable. Any angular correlations are accounted for by the conditional σ_m variable, leaving only the signal shape present in the sample. As such, only a signal model PDF is needed to fit the data.

7.7.2 The $B_s^0 \rightarrow J/\psi \phi$ probability density function

A weighted maximum likelihood fit is performed (Section C.2) on data in the B_s^0 decay time and helicity-angle distributions using only a signal model PDF. The data model is fit in 24 subsamples as outlined in Section 7.2, and the corresponding log-likelihood in each subsample, j , is scaled by a factor α_j :

$$\alpha_j = \frac{\sum_i W_i^j}{\sum_i (W_i^j)^2} \quad (7.15)$$

which is done to account for scaling of parameter uncertainties (Section C.2, Equation C.19).

As described in Section 2.4.4, the distributions of decay time and angles of a B_s^0 meson created at time $t = 0$ take the form as shown in Equation 2.55 — a sum of ten terms, corresponding to the four polarisation amplitudes ($0, \parallel, \perp, S$) squared and their interference terms. The h_k terms explicitly used in the signal PDF are of the form in Equation 2.48:

$$h_k(t|B_s^0) = \frac{3}{4\pi} e^{-\Gamma_s t} \left[a_k \cosh\left(\frac{1}{2}\Delta\Gamma_s t\right) + c_k \cos(\Delta m_s t) + b_k \sinh\left(\frac{1}{2}\Delta\Gamma_s t\right) + d_k \sin(\Delta m_s t) \right] \quad (7.16)$$

where the parameters a_k, b_k, c_k, d_k, N_k and the functions f_k are specified in Table 7.1. For a \bar{B}_s^0 at production ($t = 0$), the signs of c_k and d_k are reversed (Section 2.4.3).

7.7.3 K^+K^- S-wave interference

The interference between the P-wave and K^+K^- S-wave contributions in PDF terms when $k \in [8, 9, 10]$, for the A_S state discussed in Section 2.4.4, is handled by an effective multiplicative (coupling) factors, C_{SP} [5]. The factors are calculated by integrating the interference of P-wave and K^+K^- S-wave in each of the six $m(K^+K^-)$ bins, as done in previous analysis iterations [14]. When $k < 8$, the C_{SP} factors are unity given the respective PDF terms contain either no A_S or only A_S terms. Their overall effect on the fit is small.

7.7.4 Inclusion of resolution, detector acceptance and flavour tagging

So far, the PDF expression only considers the physics parameters relevant to the system's evolution but has not yet addressed experimental effects. Thus, the PDF is modified to include detector resolutions, acceptances and the efficiency and precision of the LHCb flavour tagging algorithms.

Four separate tagging scenarios are considered for all possible tagging algorithm responses: OS-only tagged candidates, SSK-only tagged candidates, OS and

SSK tagged candidates, and completely untagged candidates. The appropriate parametrisation, including the values of the specific tag and its associated uncertainty, are defined to scale the likelihood on a per-candidate basis:

$$\begin{aligned} \mathcal{Q}(q^{\text{OS}}, q^{\text{SSK}}, \eta^{\text{OS}}, \eta^{\text{SSK}}) &= \left[1 + q^{\text{OS}} \left(1 - 2\omega(\eta^{\text{OS}})\right)\right] \\ &\quad \times \left[1 + q^{\text{SSK}} \left(1 - 2\tilde{\omega}(\eta^{\text{SSK}})\right)\right] \end{aligned} \quad (7.17)$$

$$\begin{aligned} \overline{\mathcal{Q}}(q^{\text{OS}}, q^{\text{SSK}}, \eta^{\text{OS}}, \eta^{\text{SSK}}) &= \left[1 + q^{\text{OS}} \left(1 - 2\overline{\omega}(\eta^{\text{OS}})\right)\right] \\ &\quad \times \left[1 + q^{\text{SSK}} \left(1 - 2\tilde{\omega}(\eta^{\text{SSK}})\right)\right] \end{aligned} \quad (7.18)$$

As such, the full PDF, taking into account the detector response effects, is conditional on the mistag probability and the estimated decay-time uncertainty:

$$\begin{aligned} \mathcal{P}(t, \Omega | q^{\text{OS}}, q^{\text{SSK}}, \eta^{\text{OS}}, \eta^{\text{SSK}}, \delta_t) \\ \propto \sum_{k=1}^{10} C_{\text{SP}}^k N_k f_k(\Omega) \varepsilon_{\text{data}}^{\text{B}_s^0}(t) \\ \cdot \left\{ \left[\mathcal{Q}(q^{\text{OS}}, q^{\text{SSK}}, \eta^{\text{OS}}, \eta^{\text{SSK}}) h_k(t | \text{B}_s^0) \right. \right. \\ \left. \left. + \overline{\mathcal{Q}}(q^{\text{OS}}, q^{\text{SSK}}, \eta^{\text{OS}}, \eta^{\text{SSK}}) h_k(t | \overline{\text{B}}_s^0) \right] \otimes \mathcal{R}(t - t' | \delta_t) \right\} \end{aligned} \quad (7.19)$$

where a normalisation constant is not explicitly included. With the above PDF, a simultaneous fit is made in the 24 subsamples in terms of $m(\text{K}^+\text{K}^-)$, trigger category and data-taking year. The physics parameters are unconstrained in the fit and are shared across the subsamples, except for the S-wave fractions and the $\delta_S - \delta_\perp$ phase difference, which are independent for each $m(\text{K}^+\text{K}^-)$ bin.

k	f_k	N_k	a_k	b_k	c_k	d_k
1	$c_K^2 s_l^2$	$ A_0 ^2$	$\frac{1}{2}(1 + \lambda_0 ^2)$	$- \lambda_0 \cos(\phi_0)$	$\frac{1}{2}(1 - \lambda_0 ^2)$	$ \lambda_0 \sin(\phi_0)$
2	$\frac{1}{2} s_K^2 (1 - c_\phi^2 s_l^2)$	$ A_\parallel ^2$	$\frac{1}{2}(1 + \lambda_\parallel ^2)$	$- \lambda_\parallel \cos(\phi_\parallel)$	$\frac{1}{2}(1 - \lambda_\parallel ^2)$	$ \lambda_\parallel \sin(\phi_\parallel)$
3	$\frac{1}{2} s_K^2 (1 - s_\phi^2 s_l^2)$	$ A_\perp ^2$	$\frac{1}{2}(1 + \lambda_\perp ^2)$	$ \lambda_\perp \cos(\phi_\perp)$	$\frac{1}{2}(1 - \lambda_\perp ^2)$	$- \lambda_\perp \sin(\phi_\perp)$
4	$s_K^2 s_l^2 s_\phi c_\phi$	$ A_\perp A_\parallel $	$\frac{1}{2} \left[\sin(\delta_\perp - \delta_\parallel) - \lambda_\perp \lambda_\parallel \right]$ $\sin(\delta_\perp - \delta_\parallel - \phi_\perp + \phi_\parallel)$	$\frac{1}{2} \left[\lambda_\perp \sin(\delta_\perp - \delta_\parallel - \phi_\perp) \right]$ $+ \lambda_\parallel \sin(\delta_\parallel - \delta_\perp - \phi_\parallel)$	$\frac{1}{2} \left[\sin(\delta_\perp - \delta_\parallel) + \lambda_\perp \lambda_\parallel \right]$ $\sin(\delta_\perp - \delta_\parallel - \phi_\perp + \phi_\parallel)$	$-\frac{1}{2} \left[\lambda_\perp \cos(\delta_\perp - \delta_\parallel - \phi_\perp) \right]$ $+ \lambda_\parallel \cos(\delta_\parallel - \delta_\perp - \phi_\parallel)$
5	$\sqrt{2} s_K c_K s_l c_l c_\phi$	$ A_0 A_\parallel $	$\frac{1}{2} \left[\cos(\delta_0 - \delta_\parallel) + \lambda_0 \lambda_\parallel \right]$ $\cos(\delta_0 - \delta_\parallel - \phi_0 + \phi_\parallel)$	$-\frac{1}{2} \left[\lambda_0 \cos(\delta_0 - \delta_\parallel - \phi_0) \right]$ $+ \lambda_\parallel \cos(\delta_\parallel - \delta_0 - \phi_\parallel)$	$\frac{1}{2} \left[\cos(\delta_0 - \delta_\parallel) - \lambda_0 \lambda_\parallel \right]$ $\cos(\delta_0 - \delta_\parallel - \phi_0 + \phi_\parallel)$	$-\frac{1}{2} \left[\lambda_0 \sin(\delta_0 - \delta_\parallel - \phi_0) \right]$ $+ \lambda_\parallel \sin(\delta_\parallel - \delta_0 - \phi_\parallel)$
6	$-\sqrt{2} s_K c_K s_l c_l s_\phi$	$ A_0 A_\perp $	$-\frac{1}{2} \left[\sin(\delta_0 - \delta_\perp) - \lambda_0 \lambda_\perp \right]$ $\sin(\delta_0 - \delta_\perp - \phi_0 + \phi_\perp)$	$\frac{1}{2} \left[\lambda_0 \sin(\delta_0 - \delta_\perp - \phi_0) \right]$ $+ \lambda_\perp \sin(\delta_\perp - \delta_0 - \phi_\perp)$	$-\frac{1}{2} \left[\sin(\delta_0 - \delta_\perp) + \lambda_0 \lambda_\perp \right]$ $\sin(\delta_0 - \delta_\perp - \phi_0 + \phi_\perp)$	$-\frac{1}{2} \left[\lambda_0 \cos(\delta_0 - \delta_\perp - \phi_0) \right]$ $+ \lambda_\perp \cos(\delta_\perp - \delta_0 - \phi_\perp)$
7	$\frac{1}{3} s_l^2$	$ A_S ^2$	$\frac{1}{2}(1 + \lambda_S ^2)$	$ \lambda_S \cos(\phi_S)$	$\frac{1}{2}(1 - \lambda_S ^2)$	$- \lambda_S \sin(\phi_S)$
8	$\frac{2}{\sqrt{6}} s_K s_l c_l c_\phi$	$ A_S A_\parallel $	$\frac{1}{2} \left[\cos(\delta_S - \delta_\parallel) - \lambda_S \lambda_\parallel \right]$ $\cos(\delta_S - \delta_\parallel - \phi_S + \phi_\parallel)$	$\frac{1}{2} \left[\lambda_S \cos(\delta_S - \delta_\parallel - \phi_S) \right]$ $- \lambda_\parallel \cos(\delta_\parallel - \delta_S - \phi_\parallel)$	$\frac{1}{2} \left[\cos(\delta_S - \delta_\parallel) + \lambda_S \lambda_\parallel \right]$ $\cos(\delta_S - \delta_\parallel - \phi_S + \phi_\parallel)$	$\frac{1}{2} \left[\lambda_S \sin(\delta_S - \delta_\parallel - \phi_S) \right]$ $- \lambda_\parallel \sin(\delta_\parallel - \delta_S - \phi_\parallel)$
9	$-\frac{2}{\sqrt{6}} s_K s_l c_l s_\phi$	$ A_S A_\perp $	$-\frac{1}{2} \left[\sin(\delta_S - \delta_\perp) + \lambda_S \lambda_\perp \right]$ $\sin(\delta_S - \delta_\perp - \phi_S + \phi_\perp)$	$-\frac{1}{2} \left[\lambda_S \sin(\delta_S - \delta_\perp - \phi_S) \right]$ $- \lambda_\perp \sin(\delta_\perp - \delta_S - \phi_\perp)$	$-\frac{1}{2} \left[\sin(\delta_S - \delta_\perp) - \lambda_S \lambda_\perp \right]$ $\sin(\delta_S - \delta_\perp - \phi_S + \phi_\perp)$	$-\frac{1}{2} \left[- \lambda_S \cos(\delta_S - \delta_\perp - \phi_S) \right]$ $+ \lambda_\perp \cos(\delta_\perp - \delta_S - \phi_\perp)$
10	$\frac{2}{\sqrt{3}} c_K s_l^2$	$ A_S A_0 $	$\frac{1}{2} \left[\cos(\delta_S - \delta_0) - \lambda_S \lambda_0 \right]$ $\cos(\delta_S - \delta_0 - \phi_S + \phi_0)$	$\frac{1}{2} \left[\lambda_S \cos(\delta_S - \delta_0 - \phi_S) \right]$ $- \lambda_0 \cos(\delta_0 - \delta_S - \phi_0)$	$\frac{1}{2} \left[\cos(\delta_S - \delta_0) + \lambda_S \lambda_0 \right]$ $\cos(\delta_S - \delta_0 - \phi_S + \phi_0)$	$\frac{1}{2} \left[\lambda_S \sin(\delta_S - \delta_0 - \phi_S) \right]$ $- \lambda_0 \sin(\delta_0 - \delta_S - \phi_0)$

Table 7.1 Angular and time-dependent functions used in the fit to the data. Abbreviations used include $c_K = \cos \theta_K$, $s_K = \sin \theta_K$, $c_l = \cos \theta_l$, $s_l = \sin \theta_l$, $c_\phi = \cos \phi$ and $s_\phi = \sin \phi$.

7.8 Systematics

Effects resulting in systematic uncertainties or biases in the measured physics parameters arise from different sources. Beginning with data-taking, reconstruction and selection, efficiencies estimation, *etc.*, they can appear in every step of the analysis.

There is a multitude of systematics taken under consideration and evaluated as part of the complete analysis. Most prominently, these are associated with the mass model of $m(J/\psi K^+ K^-)$ and the calculation of s Weights later used to subtract the background in the data samples. Furthermore, systematic uncertainties are assigned to account for any potential mis-modelling of the $\Lambda_b^0 \rightarrow J/\psi p K^-$ peaking background contribution as the main physics background. Moreover, systematic effects are considered where events with multiple B_s^0 candidates are reconstructed. Several studies also consider possible biases in the fitting procedure and the potential biases in the fit results. One such systematic is discussed in detail in Section 7.8.1.

Additionally, systematic uncertainties are associated with the choices made when determining angular and decay-time efficiencies and potential biases in their final values and their subsequent effects in the analysis. Finally, no sizeable systematic effects on the results are observed when repeating the same analysis on subsets of the data split by LHCb magnet polarity, year of data-taking, number of primary vertices, bins of $p_T(B_s^0)$, pseudorapidity (η) or decay-time uncertainty. The results of the systematic uncertainties are summarised in Table 7.2.

7.8.1 Contribution from B_c^+ decays

LHCb has observed the $B_c^+ \rightarrow B_s^0 \pi^+$ decay mode and has measured $\sigma(B_c^+)/\sigma(B_s^0) \times B(B_c^+ \rightarrow B_s^0 \pi^+) = [2.37 \pm 0.31 \pm 0.11] \times 10^{-3}$ [186]. As such, a small fraction of $B_s^0 \rightarrow J/\psi \phi$ signal candidates may come from the decay of a B_c^+ meson. This fraction is estimated as:

$$f = \frac{\sigma(B_c^+)}{\sigma(B_s^0)} \times B(B_c^+ \rightarrow B_s^0 \pi^+) \times \frac{\varepsilon_{B_c^+}}{\varepsilon_{B_s^0}} = 0.0015 \pm 0.0003 \quad (7.20)$$

where $\varepsilon_{B_c^+} = \varepsilon_{B_c^+}^{\text{gen}} + \varepsilon_{B_c^+}^{\text{trig+sel}}$ is the total efficiency for selecting $B_s^0 \rightarrow J/\psi \phi$ decays from a simulated sample of $B_c^+ \rightarrow B_s^0 (\rightarrow J/\psi \phi) \pi^+$ decays. The efficiency $\varepsilon_{B_s^0}$ is

defined the same way but using a simulated sample of $B_s^0 \rightarrow J/\psi \phi$. The fraction is multiplied by a factor of 3.7 to account for $B(B_c^+ \rightarrow B_s^0 X)/B(B_c^+ \rightarrow B_s^0 \pi^+) \approx 60\%/16\%$ estimated in Reference [187]. That leads to the estimate of a 0.5 % contribution of B_s^0 -from- B_c^+ decays in the signal sample. The effect of ignoring this contribution in the final fit is evaluated using 1984 simulated experiments containing 30000 tagged and 60000 untagged $B_s^0 \rightarrow J/\psi \phi$ signal events. A 0.5 % contribution of B_s^0 -from- B_c^+ decays is added by randomly sampling events from the fully simulated and selected sample of $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi)\pi^+$ decays used. Time and angular acceptances and decay-time resolution are included. The $B_s^0 \rightarrow J/\psi \phi$ signal events are generated using the same physics parameter values as those used in the generation of the B_s^0 decay in the $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi)\pi^+$ sample. A tagged fit to the merged dataset is performed, and the results show that ignoring the B_c^+ component leads to no significant biases. Figures 7.8, 7.9, 7.10, 7.11 and 7.12 show a Gaussian function fit to the resulting distributions produced as follows:

$$P(p) = \frac{p^{\text{incl.}B_c^+} - p^{\text{excl.}B_c^+}}{\sigma_p^{\text{excl.}B_c^+}} \quad (7.21)$$

where P is the resulting distribution for a parameter p of the tagged angular and time-dependent fit as outlined in Equation 2.55 and Table 7.1. Also, $p^{\text{incl.}B_c^+}$ and $p^{\text{excl.}B_c^+}$ are the values of this parameter produced as a result of an individual simulated experiment comparing a fit including B_c^+ -originating signal events and excluding them, respectively; while $\sigma_p^{\text{excl.}B_c^+}$ is the estimated standard deviation of the parameter p produced with the fit excluding B_c^+ events. The Gaussian fits are used to compare the impact of each fit parameter by including and excluding the B_c^+ contribution and performing the 1984 individual simulated experiments. It can be seen that no significant directional biases are present and all relevant are well behaved. Moreover, the standard deviations are negligible, with the largest being 0.09, while some parameters do not display any measurable differences with the given statistics.

Source	ϕ_s [rad]	$ \lambda $	$\Gamma_s - \Gamma_d$ [ps]	$\Delta\Gamma_s$ [ps]	Δm_s [ps]	$ A_\perp ^2$	$ A_0 ^2$	$\delta_\perp - \delta_0$ [rad]	$\delta_\parallel - \delta_0$ [rad]
Mass: width parametrisation	—	—	—	0.0002	0.001	0.0004	0.0006	—	0.003
Mass: decay-time and angles dependence	0.004	0.0037	0.0007	0.0022	0.016	0.0005	0.0002	0.05	0.009
Multiple candidates	0.0011	0.0011	0.0003	0.0001	0.001	0.0001	0.0001	0.01	0.002
Fit bias	0.0010	—	—	0.0003	0.001	0.0006	0.0001	0.02	0.033
C_{SP} factors	0.0010	0.0010	—	0.0001	0.002	0.0001	—	0.01	0.005
Time resolution: model applicability	—	—	—	—	0.001	—	—	—	0.001
Time resolution: t bias	0.0032	0.0010	0.0002	0.0003	0.005	—	—	0.08	0.001
Time resolution: wrong PV	—	—	—	—	0.001	—	—	—	0.001
Angular efficiency: simulated sample size	0.0011	0.0018	—	—	0.001	0.0004	0.0003	—	0.004
Angular efficiency: weighting	0.0022	0.0043	0.0001	0.0002	0.001	0.0011	0.0020	0.01	0.008
Angular efficiency: clone candidates	0.0005	0.0014	0.0002	0.0001	—	0.0001	0.0002	—	0.002
Angular efficiency: t and σ_t dependence	0.0012	0.0007	0.0002	0.0010	0.003	0.0012	0.0008	0.03	0.006
Decay-time efficiency: statistical	—	—	0.0012	0.0008	—	0.0003	0.0002	—	—
Decay-time efficiency: kinematic weighting	—	—	0.0002	—	—	—	—	—	—
Decay-time efficiency: PDF weighting	—	—	0.0001	0.0001	—	—	—	—	—
Decay-time efficiency: $\Delta\Gamma_s = 0$ simulation	—	—	0.0003	0.0005	—	0.0002	0.0001	—	—
Length scale	—	—	—	—	0.004	—	—	—	—
Quadratic sum of syst.	0.0061	0.0064	0.0015	0.0026	0.018	0.0019	0.0023	0.10	0.036

Table 7.2 *Summary of systematic uncertainties [5].*

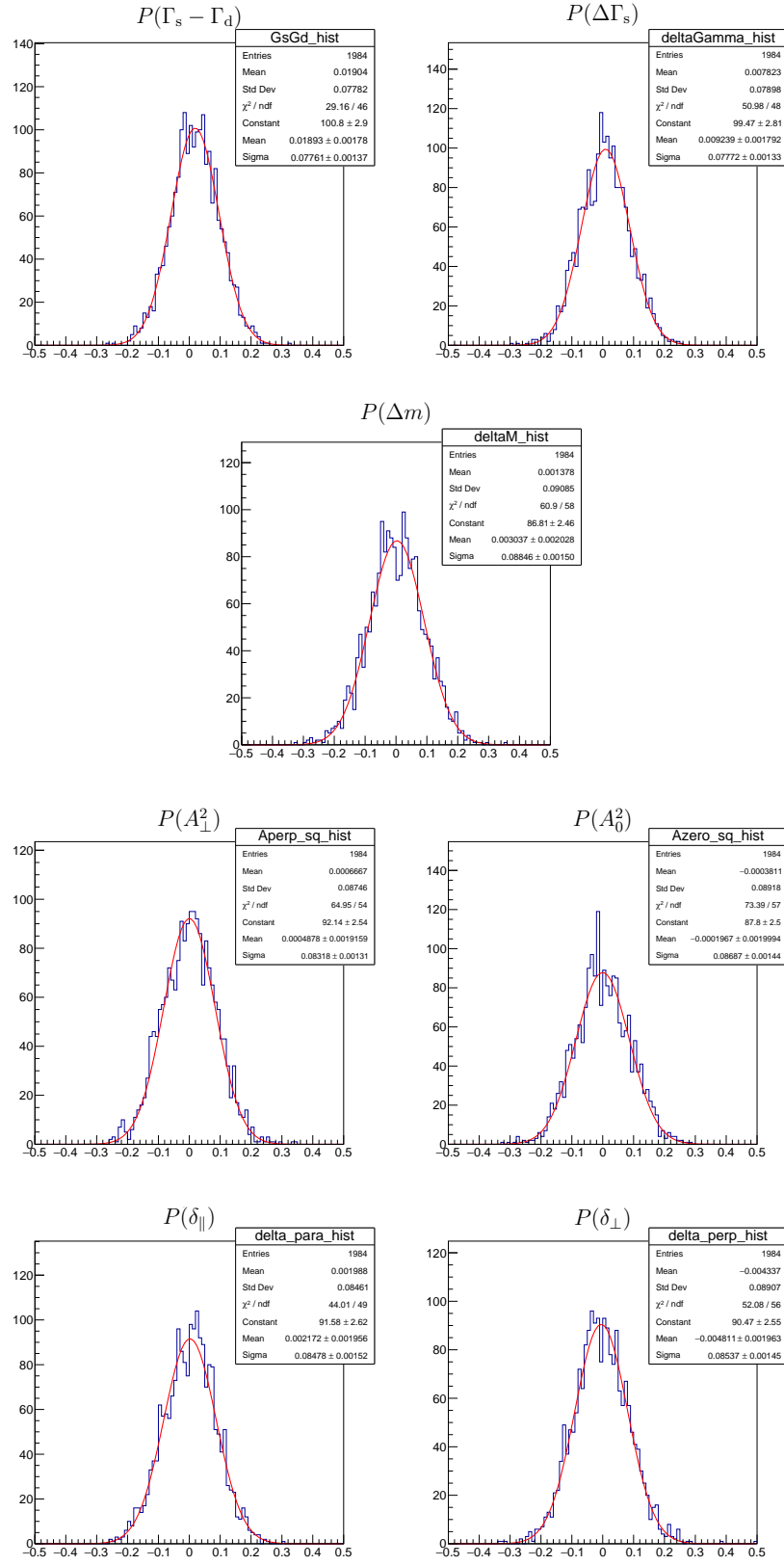


Figure 7.8 Single Gaussian fit to the P distributions from Eq. 7.21 of different parameters obtained from fitting simulated experiments with $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ contribution included.

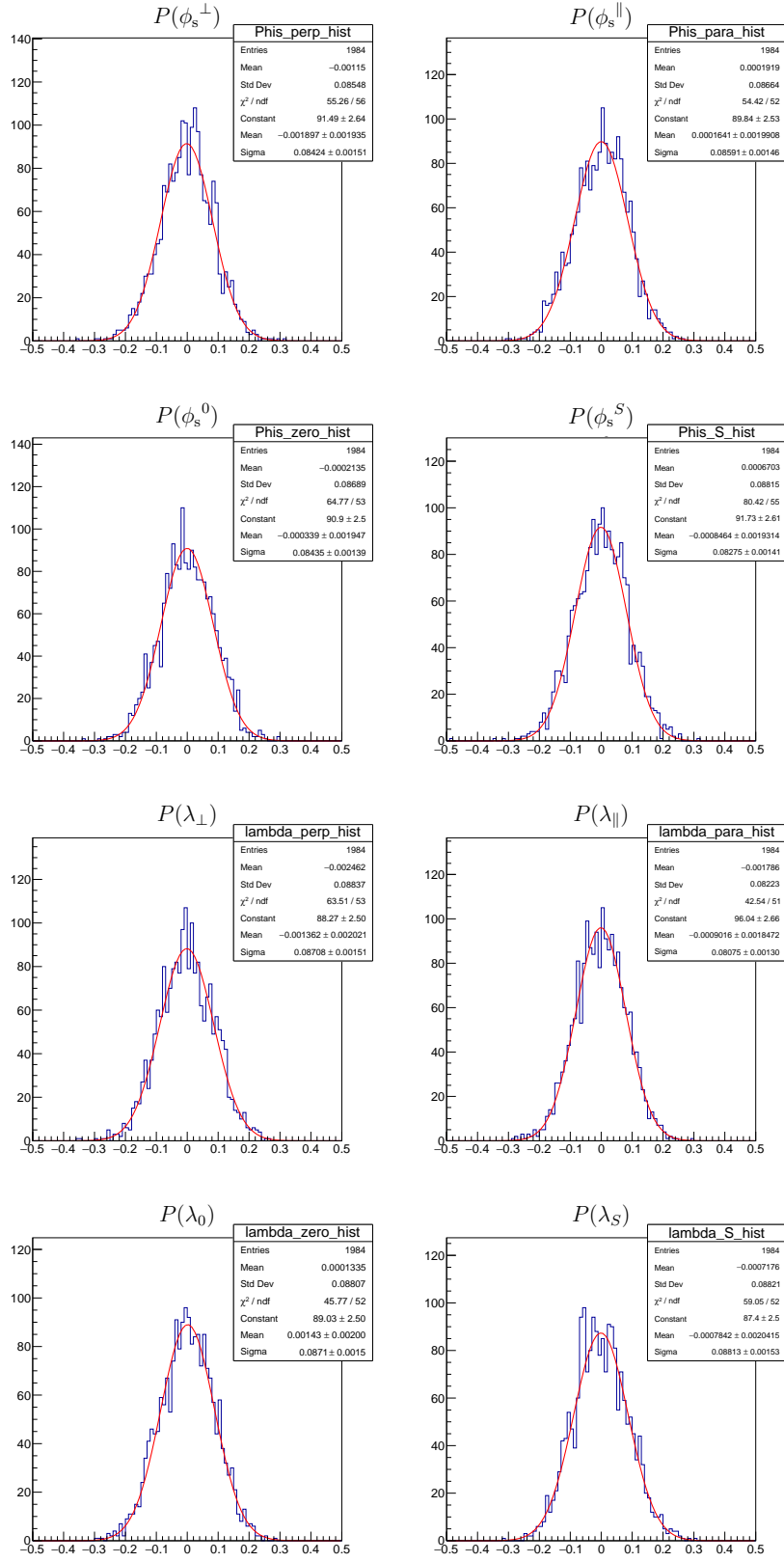


Figure 7.9 Single Gaussian fit to the P distributions from Eq. 7.21 of different parameters obtained from fitting simulated experiments with $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi\phi)\pi^+$ contribution included.

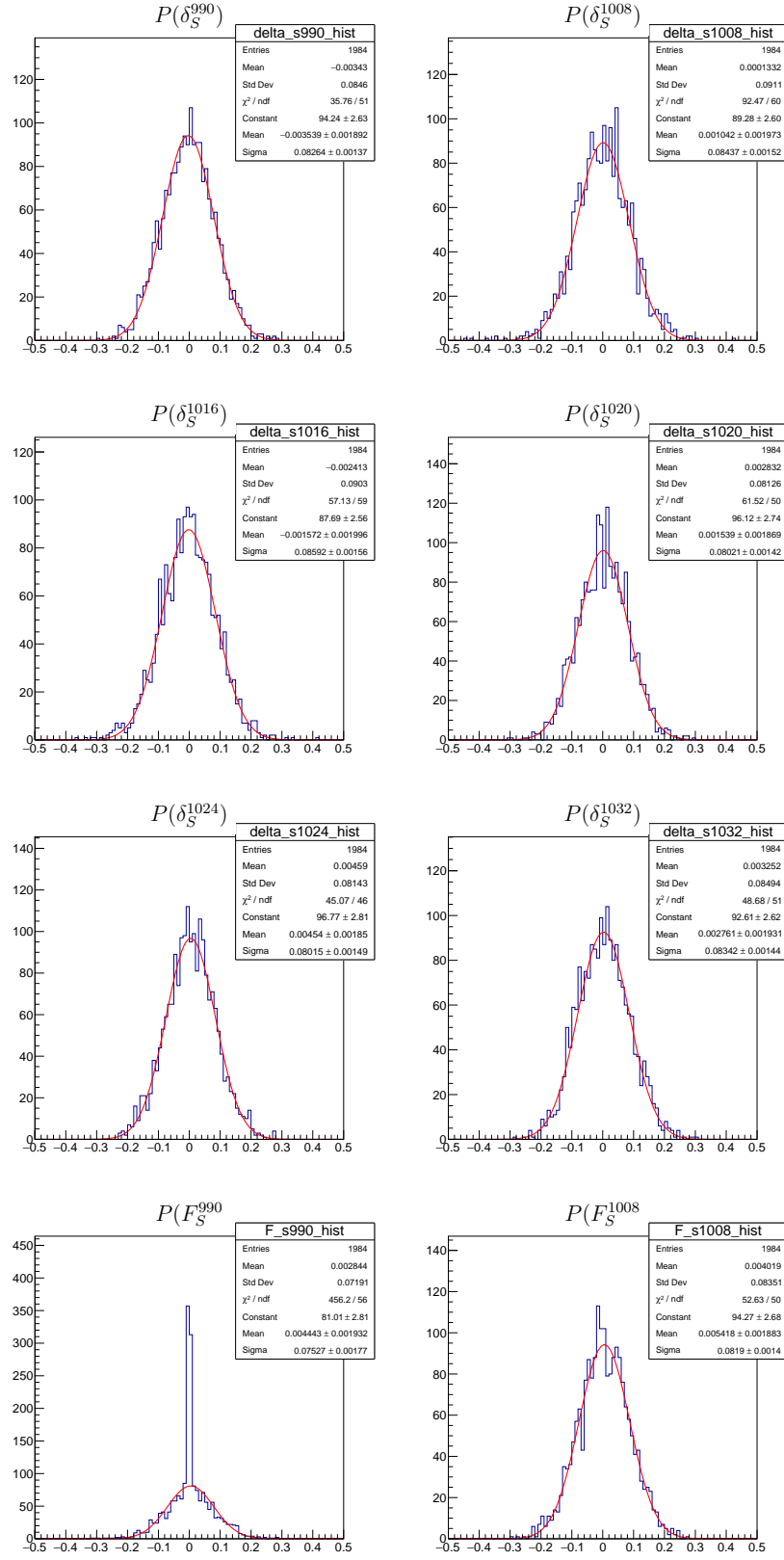


Figure 7.10 Single Gaussian fit to the P distributions from Eq. 7.21 of different parameters obtained from fitting simulated experiments with $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ contribution included.

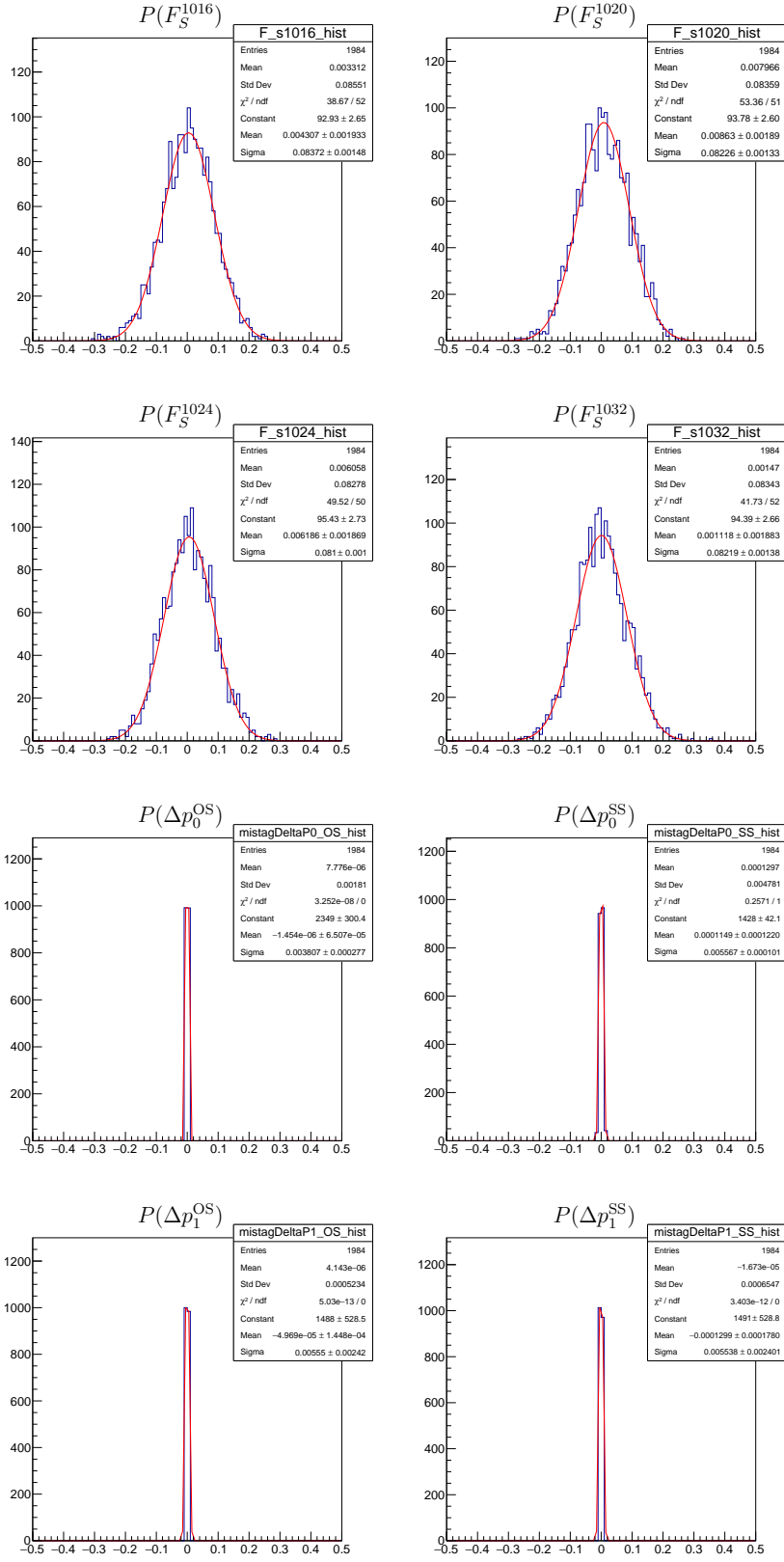


Figure 7.11 Single Gaussian fit to the P distributions from Eq. 7.21 of different parameters obtained from fitting simulated experiments with $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ contribution included.

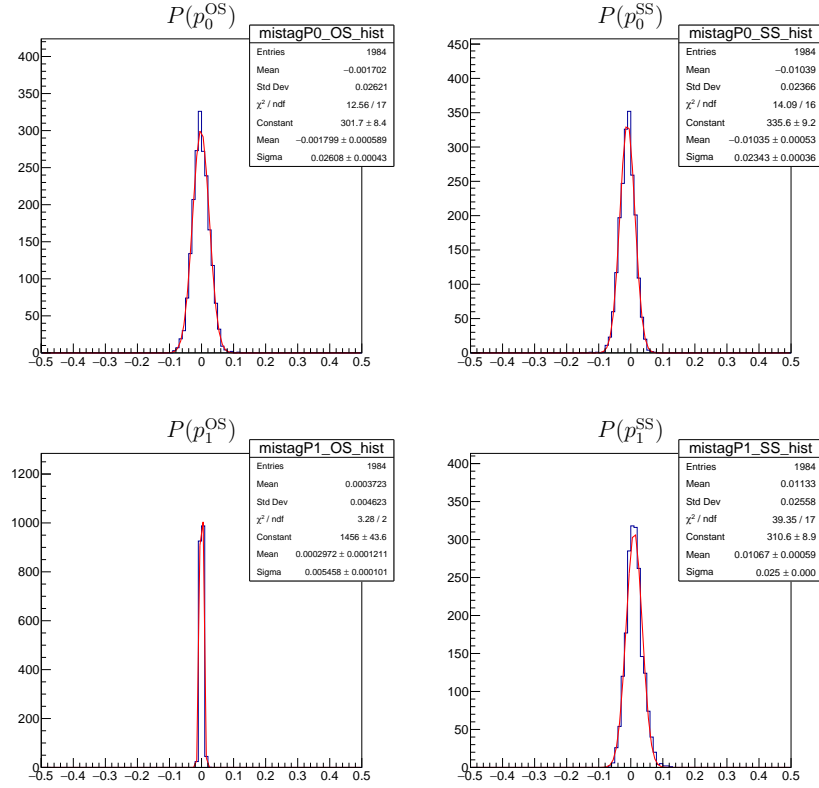


Figure 7.12 Single Gaussian fit to the P distributions from Eq. 7.21 of different parameters obtained from fitting simulated experiments with $B_c^+ \rightarrow B_s^0(\rightarrow J/\psi \phi) \pi^+$ contribution included.

7.9 Results

The results of the maximum likelihood fit described in Section 7.7 including uncertainties are as follows:

$$\begin{aligned}
\phi_s &= -0.083 \pm 0.04 \pm 0.006 \text{ rad} \\
|\lambda| &= 1.012 \pm 0.016 \pm 0.006 \\
\Gamma_s - \Gamma_d &= -0.0041 \pm 0.0024 \pm 0.0016 \text{ ps}^{-1} \\
\Delta\Gamma_s &= 0.077 \pm 0.008 \pm 0.003 \text{ ps}^{-1} \\
\Delta m_s &= 17.703 \pm 0.059 \pm 0.018 \text{ ps}^{-1} \\
|A_\perp|^2 &= 0.2456 \pm 0.0040 \pm 0.0019 \\
|A_0|^2 &= 0.5186 \pm 0.0029 \pm 0.0024 \\
\delta_\perp - \delta_0 &= 2.64 \pm 0.13 \pm 0.10 \text{ rad} \\
\delta_\parallel - \delta_0 &= 3.06^{+0.08}_{-0.07} \pm 0.04 \text{ rad}
\end{aligned}$$

and the background-subtracted data distributions with fit projections are shown in Figure 7.13. All results are listed with the statistical first and the systematic uncertainty second. The results are in good agreement with the previous LHCb

	ϕ_s	$ \lambda $	$\Gamma_s - \Gamma_d$	$\Delta\Gamma_s$	Δm_s	$ A_\perp ^2$	$ A_0 ^2$	$\delta_\perp - \delta_0$	$\delta_\parallel - \delta_0$
ϕ_s	1.00	0.16	-0.05	0.01	-0.02	0.01	0.00	0.03	0.00
$ \lambda $		1.00	0.07	-0.09	0.06	0.04	-0.02	0.04	0.01
$\Gamma_s - \Gamma_d$			1.00	-0.46	0.06	0.35	-0.24	-0.01	0.03
$\Delta\Gamma_s$				1.00	-0.05	-0.64	0.46	-0.02	0.00
Δm_s					1.00	0.01	0.01	0.55	-0.01
$ A_\perp ^2$						1.00	-0.64	0.01	0.07
$ A_0 ^2$							1.00	0.01	-0.02
$\delta_\perp - \delta_0$								1.00	0.25
$\delta_\parallel - \delta_0$									1.00

Table 7.3 *Correlation matrix including the statistical and systematic correlations between the parameters [5].*

measurement [14], while the results of ϕ_s , $\Delta\Gamma_s$, $\Gamma_s - \Gamma_d$ were the most precise at the time of publication of the analysis article (Ref. [5]). Given the current statistics, the results also agree with the SM expectations [188–191] and, as such, indicate no additional CP violation in $B_s^0 \rightarrow J/\psi K^+ K^-$ decays within current precision. The value of Δm_s is in good agreement with the world average [172], suggesting B_s^0

oscillation is well reproduced in this analysis. The parameter correlation matrix is shown in Table 7.3.

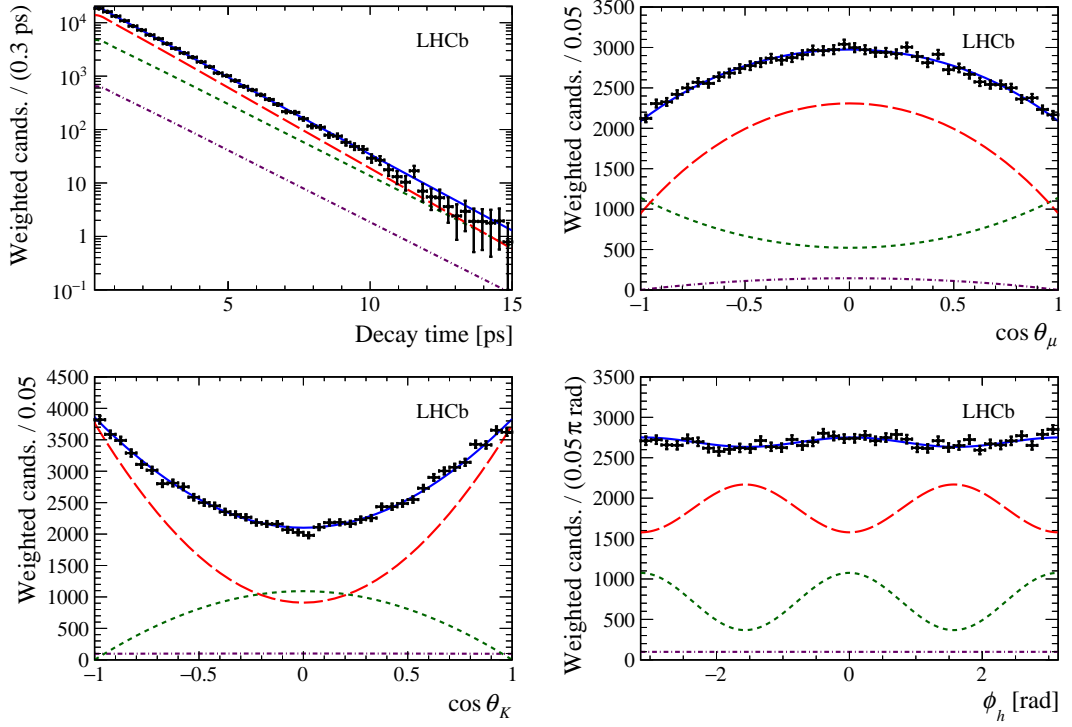


Figure 7.13 *Decay-time and helicity-angle distributions for background-subtracted $B_s^0 \rightarrow J/\psi K^+ K^-$ decays (data points) with the one-dimensional projections of the PDF. The solid blue line shows the total signal contribution, which contains CP -even (long-dashed red), CP-odd (short-dashed green) and S-wave (dotted-dashed purple) contributions. Data and fit projections for the different samples considered (data-taking year, trigger and tagging categories, $m(K^+ K^-)$ bins) are combined [5].*

Chapter 8

Conclusion

This thesis presents three original achievements. Firstly, a novel method to handle collision data is demonstrated using an efficient machine-learning algorithm to approximate arbitrary models and execute them in the LHCb trigger. Secondly, a highly automated control system is developed and used to successfully validate a total of 3100 R13742 and 450 R13743 MaPMTs later used in the LHCb RICH upgrade. Lastly, a flavour-tagged decay-time-dependent angular analysis of $B_s^0 \rightarrow J/\psi \phi$ decays is performed using $\int \mathcal{L} dt = 1.9 \text{ fb}^{-1}$ of pp collision data recorded by the LHCb experiment during the 2015 and 2016 runs of the LHC.

The machine-learning algorithm was integrated into a toolkit, NNDRone. The toolkit enabled analysts to approximate arbitrary machine-learning algorithms and improve on the time it takes to execute them while running at the LHCb trigger. It was demonstrated that a drone neural network can accurately and precisely learn the features of a competing model when it comes to event selection, even when the competing model has a different structure. Furthermore, approximating models can happen without ever having access to the training data, or indeed any data, but solely with appropriate questioning of the original model.

Moreover, the equivalence between the outputs of the drone and the original model enables treating both the original and the drone in the same way. The standardised algorithm to produce a drone allows the usage of any desired machine-learning package. It is shown possible to use any package to initially isolate a decay signature, which in turn enables the creation of a classifier guaranteed to be suitable for execution in C++ real-time data selection frameworks, such as

the upgraded LHCb fully software-based trigger.

The highly automated control system was deployed to assist in the efforts of the LHCb RICH Photon Detector Quality Assurance (PDQA) programme. During the PDQA, the system successfully validated a total of 3100 R13742 and 450 R13743 MaPMTs and gathered crucial data about the properties of each MaPMT unit to be later used to determine its best placement in the upgraded RICH system. The control system was deployed on four testing stations in two separate facilities and enabled consistent characterisation of $16 \times \text{R13742}$ or $4 \times \text{R13743}$ MaPMTs per station per day. The results showed excellent gain, uniformity and single-photon resolution [140]. Finally, the automation allowed for thorough consistency between the two labs and an overall efficiency increase by removing human-error systematic effects.

Finally, the thesis details the results of the flavour-tagged decay-time-dependent angular analysis of $B_s^0 \rightarrow J/\psi K^+ K^-$ decays intending to measure CP -violating quantities. The analysis has been performed using pp collision data collected at LHCb during the 2015 and 2016 years of data-taking, accumulating to a dataset size of 1.9 fb^{-1} . Approximately 117 000 signal decays have been selected with a decay-time resolution of 45 fs and tagging power of about 4.7 %. The CP -violating phase, ϕ_s , is measured to be $-0.083 \pm 0.04 \pm 0.006 \text{ rad}$. The decay-width difference between the B_s^0 mass eigenstates is measured to be $\Delta\Gamma_s = 0.077 \pm 0.008 \pm 0.003 \text{ ps}^{-1}$, and the difference of the average decay widths of B_s^0 and B_d^0 mesons is obtained as $\Gamma_s - \Gamma_d = -0.0041 \pm 0.0024 \pm 0.0016 \text{ ps}^{-1}$. Using the known value of the B_d^0 meson lifetime, $1.520 \pm 0.004 \text{ ps}$ [13], the ratio of decay widths between B_s^0 and B_d^0 is measured to be $\Gamma_s/\Gamma_d = 0.9938 \pm 0.0036 \pm 0.0023$. These are the most precise values ever obtained at the time of this thesis. Furthermore, the mass difference between B_s^0 mass eigenstates is measured as $\Delta m_s = 17.703 \pm 0.059 \pm 0.018 \text{ ps}^{-1}$. All of these results are consistent with Standard Model (SM) theoretical predictions [188–191].

The results are further combined with those from Run 2 using $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ [192], those from Run 1 using $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ [193], $B_s^0 \rightarrow J/\psi K^+ K^-$ with $m(K^+ K^-) > 1.05 \text{ GeV}/c^2$ [194], $B_s^0 \rightarrow \psi(2S)\phi$ [195], and $B_s^0 \rightarrow D_s^+ D_s^-$ [196]. The combined results are: $\phi_s = -0.041 \pm 0.025 \text{ rad}$, $|\lambda| = 0.993 \pm 0.010$, $\Gamma_s = 0.6562 \pm 0.0021 \text{ ps}^{-1}$, $\Delta\Gamma_s = 0.0816 \pm 0.0048 \text{ ps}^{-1}$. These values are consistent with SM theoretical predictions [188, 189]. The ϕ_s value is consistent with the non-zero CP violation predicted by the SM. It is consistent with no CP violation in the interference between B_s^0 - \bar{B}_s^0 mixing and decay. Finally, the $|\lambda|$ parameter is consistent with

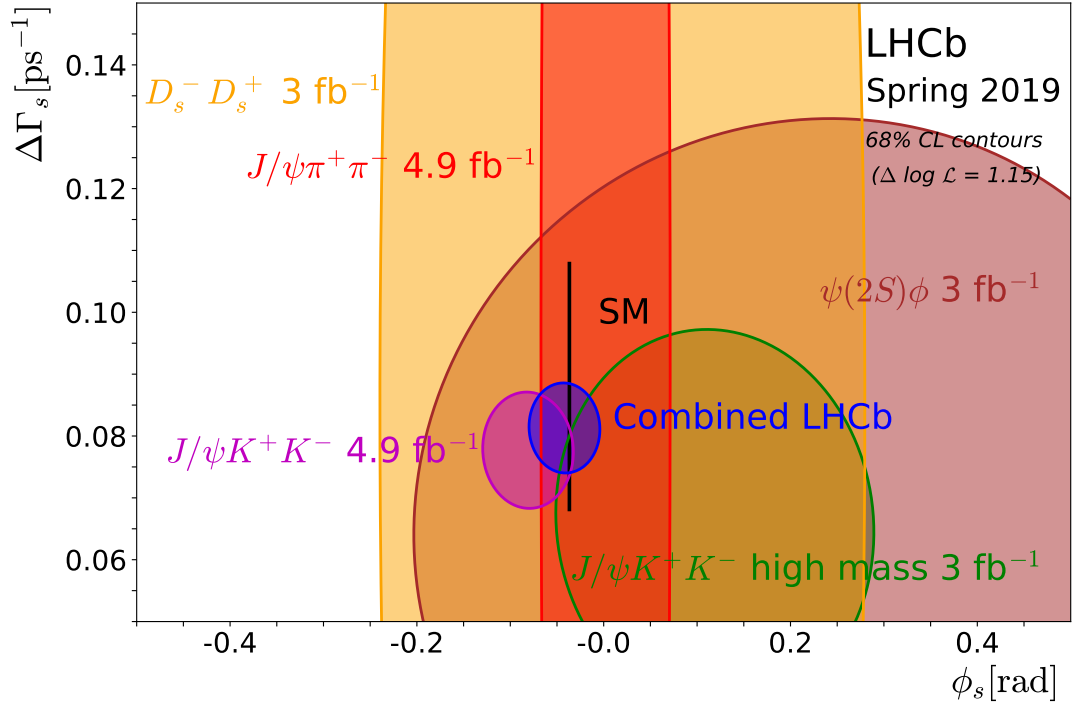


Figure 8.1 *Regions of 68% confidence level in the ϕ_s - $\Delta\Gamma_s$ plane for the individual LHCb measurements and a combined contour (in blue). The $B_s^0 \rightarrow J/\psi K^+ K^-$ (magenta) and $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ [192] (red) contours show the Run 1 and Run 2 combined numbers. The ϕ_s [188] and $\Delta\Gamma_s$ [190] predictions are indicated by the thin black rectangle.*

unity, signifying no evidence for direct CP violation in $B_s^0 \rightarrow J/\psi K^+ K^-$ decays. The combined results are shown in Figure 8.1, with their 68 % confidence regions as a coloured area around their central point. The results of this thesis are published in Ref. [5].

Appendix A

Further detail on some parts of Chapter 2

A.1 The Standard Model

A.1.1 The unconventional distinction used to portray chirality

For illustrative purposes, a special case is made in the specific paragraph that led the reader here. It is made solely to give detail by example. That distinction is clarified below.

The objects labelled here as the “electron” and “positron” (and their antipartners) are different from the objects in the usual terminology, and they do not represent the “physical” electron in, say, an atom. However, the ideas discussed here are connected. The key point is that in the present local variation from convention, there are four distinct particles:

- electron: left-handed, charge -1 , weak isospin $-\frac{1}{2}$, interacts weakly
- antielectron: right-handed, charge $+1$, weak isospin $+\frac{1}{2}$, interacts weakly
- positron: left-handed, charge $+1$, weak isospin 0 , no weak interaction
- antipositron: right-handed, charge -1 , weak isospin 0 , no weak interaction

The left- and right-helicity electrons and antipositrons act as the four components of the Dirac spinor for the object, which is conventionally understood as the electron (in the mass basis). Similarly, the left- and right-helicity antielectrons and positrons for the conjugate Dirac spinor represent the conventional positron (in the mass basis).

The names “electron” and “positron” are used here to distinguish between the particles that couple to the W and those that don’t. The conventional language in particle physics calls these the left-handed (chirality) electron and the right-handed (chirality) electron. The idea here is to use a different convention to highlight that the above are not related by parity (space inversion, frame change or reversing angular momentum).

A.2 *CP* Violation

A.2.1 Time-reversal operator

The time-reversal operator, T , is anti-unitary. This choice is necessary for T to be involved in complex conjugation to preserve observables like positive energy and unchanged spatial coordinates. An anti-unitary operator, like T , still keeps the canonical commutators:

$$\begin{aligned} [r_i, p_j] &= i\hbar\delta_{ij} \\ T[r_i, p_j] &= [r_i, -p_j] = -(-i)\hbar\delta_{ij} = i\hbar\delta_{ij} \end{aligned} \tag{A.1}$$

even though it does not act on r_i . It also must preserve that energy is always positive. Thus:

$$T \exp(-iEt) \neq \exp(-iE(-t)) = \exp(iEt) \text{ [non - physical]} \tag{A.2}$$

but instead:

$$T \exp(-iEt) = \exp(-(-i)E(-t)) = \exp(-iEt) \tag{A.3}$$

A.2.2 Weak decay Hamiltonian construction

In the SM, the weak effective Hamiltonian takes the generic form [197]:

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i C_i(\mu) Q_i \quad (\text{A.4})$$

where G_F is the Fermi constant, μ is the renormalisation scale, V_{CKM}^i and C_i are the Cabibbo-Kobayashi-Maskawa factors[45, 46] and the Wilson coefficients [198, 199] respectively, describing the strength with which each local decay operator Q_i enters the Hamiltonian.

A.2.3 Two-mode decay amplitude

$$\begin{aligned} |A_f|^2 &= A_f A_f^* = |A_1|^2 + |A_2|^2 + |A_1||A_2|e^{i(\delta_1-\delta_2)}e^{i(\varphi_1-\varphi_2)} \\ &\quad + |A_1||A_2|e^{-i(\delta_1-\delta_2)}e^{-i(\varphi_1-\varphi_2)} \\ &= |A_1|^2 + |A_2|^2 + |A_1||A_2|(e^{i(\Delta\delta)}e^{i(\Delta\varphi)} + e^{-i(\Delta\delta)}e^{-i(\Delta\varphi)}) \\ &\quad \text{by Euler's identity} \\ &= |A_1|^2 + |A_2|^2 + |A_1||A_2|[(\cos(\Delta\delta) + i\sin(\Delta\delta))(\cos(\Delta\varphi) + i\sin(\Delta\varphi)) \\ &\quad + (\cos(\Delta\delta) - i\sin(\Delta\delta))(\cos(\Delta\varphi) - i\sin(\Delta\varphi))] \\ &= |A_1|^2 + |A_2|^2 + |A_1||A_2|[\cos(\Delta\delta)\cos(\Delta\varphi) - \sin(\Delta\delta)\sin(\Delta\varphi) \\ &\quad + i\sin(\Delta\varphi)\cos(\Delta\delta) + i\sin(\Delta\delta)\cos(\Delta\varphi) \\ &\quad + \cos(\Delta\delta)\cos(\Delta\varphi) - \sin(\Delta\delta)\sin(\Delta\varphi) \\ &\quad - i\sin(\Delta\delta)\cos(\Delta\varphi) - i\cos(\Delta\varphi)\sin(\Delta\delta)] \\ &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2|[\cos(\Delta\delta)\cos(\Delta\varphi) - \sin(\Delta\delta)\sin(\Delta\varphi)] \\ &\quad \text{and similarly} \\ |\bar{A}_{\bar{f}}|^2 &= |A_1|^2 + |A_2|^2 + 2|A_1||A_2|[\cos(\Delta\delta)\cos(\Delta\varphi) + \sin(\Delta\delta)\sin(\Delta\varphi)] \quad (\text{A.5}) \end{aligned}$$

A.2.4 CP in mixing

Concerning only mixing, $B_s^0 \rightarrow \bar{B}_s^0$ oscillations can be modelled with a 2×2 effective Hamiltonian as:

$$i \frac{d}{dt} \begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} = (\mathbf{M} - \frac{i}{2}\mathbf{\Gamma}) \begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} \quad (\text{A.6})$$

where \mathbf{M} and $\mathbf{\Gamma}$ are Hermitian matrices and $B_s^0(t)$ is the time evolution of a flavour eigenstate produced as a B_s^0 at $t = 0$. In contrast, $\bar{B}_s^0(t)$ is the time evolution of the oppositely produced flavour eigenstate. Each of the mass eigenstates is a mixture of the flavour eigenstates:

$$\begin{aligned} |B_L\rangle &\approx p\sqrt{1-z}|B_s^0\rangle + q\sqrt{1+z}|\bar{B}_s^0\rangle \\ |B_H\rangle &\approx p\sqrt{1+z}|B_s^0\rangle - q\sqrt{1-z}|\bar{B}_s^0\rangle \end{aligned} \quad (\text{A.7})$$

where p and q are equivalent complex parameters as in Equation 2.19, while z is only non-zero when CPT is violated. Assuming CPT is conserved and $z = 0$, the amplitudes are normalised with the condition of $|p|^2 + |q|^2 = 1$. Given CPT conservation and the fact that \mathbf{M} and $\mathbf{\Gamma}$ are Hermitian, then:

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^* & M_{11} \end{pmatrix} \quad \mathbf{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{11} \end{pmatrix} \quad (\text{A.8})$$

where the elements M_{11} is labelled M_s and is the mass of the B_s^0 . Similarly, Γ_{11} is the decay width of the B_s^0 , Γ_s . Thus the flavour state time evolution is specified by a Hamiltonian of the form:

$$\mathbf{H} \equiv \begin{pmatrix} H_0 & H_{12} \\ H_{21} & H_0 \end{pmatrix} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} \equiv \begin{pmatrix} M_s & M_{12} \\ M_{12}^* & M_s \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma_s & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_s \end{pmatrix} \quad (\text{A.9})$$

Here, the off-diagonal elements of \mathbf{M} and $\mathbf{\Gamma}$ are specifying the oscillations. The M_{12} element is responsible for the dispersive part of H_{12} , while Γ_{12} is the absorptive part. The dispersive contribution is given rise by the intermediate states in the mixing diagram and is dominated by diagrams involving virtual t quarks (Figure 2.5). On the other hand, the absorptive part is due to the possible final states into which B_s^0 and \bar{B}_s^0 can decay. However, the mass eigenstates are decoupled with defined mass ($M_{L,H}$) and lifetime ($\Gamma_{L,H}$). Therefore, there exists a transformation \mathbf{T} which diagonalises \mathbf{H} resulting in a diagonal \mathbf{H}^m (specifying

mass eigenstate evolution):

$$\mathbf{H}^{\mathbf{m}} = \mathbf{T}^{-1} \mathbf{H} \mathbf{T} \quad \text{and} \quad \begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} = \mathbf{T} \begin{pmatrix} |B_L(t)\rangle \\ |B_H(t)\rangle \end{pmatrix} \quad \text{such that} \quad (\text{A.10})$$

$$\mathbf{H}^{\mathbf{m}} = \begin{pmatrix} M_L & 0 \\ 0 & M_H \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma_L & 0 \\ 0 & \Gamma_H \end{pmatrix}$$

The resulting $\mathbf{H}^{\mathbf{m}}$ can also be expressed in the eigenvalues of \mathbf{H} , while \mathbf{T} can be constructed from its eigenvectors:

$$\mathbf{H}^{\mathbf{m}} = \begin{pmatrix} [H_0 - \sqrt{H_{12}H_{21}}] & 0 \\ 0 & [H_0 + \sqrt{H_{12}H_{21}}] \end{pmatrix} \quad (\text{A.11})$$

$$\mathbf{T} = \begin{pmatrix} \begin{bmatrix} \sqrt{H_{12}} \\ -\sqrt{H_{21}} \end{bmatrix} & \begin{bmatrix} \sqrt{H_{12}} \\ +\sqrt{H_{21}} \end{bmatrix} \end{pmatrix}$$

Combining Equations A.9, A.10 and A.11 yields the following relationships:

$$H_0 = \frac{1}{2}(M_H + M_L) - \frac{i}{4}(\Gamma_L + \Gamma_H) \quad (\text{A.12})$$

$$\sqrt{H_{12}H_{21}} = \frac{1}{2}(M_H - M_L) + \frac{i}{4}(\Gamma_L - \Gamma_H)$$

and the following useful definitions:

$$m_s = \frac{1}{2}(M_H + M_L)$$

$$\Delta m_s = M_H - M_L$$

$$\Gamma_s = \frac{1}{\tau_s} = \frac{1}{2}(\Gamma_L + \Gamma_H) \quad (\text{A.13})$$

$$\Delta \Gamma_s = \Gamma_L - \Gamma_H$$

Observe that all m_s , Δm_s , Γ_s and $\Delta \Gamma_s$ are all positive real numbers. More interestingly, the expression of the transformation matrix \mathbf{T} is revealed:

$$\mathbf{T} = \begin{pmatrix} p & p \\ q & -q \end{pmatrix} \quad \text{and so} \quad (\text{A.14})$$

$$\begin{pmatrix} |B_L\rangle \\ |B_H\rangle \end{pmatrix} = \mathbf{T}^T \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix} = \begin{pmatrix} p & q \\ p & -q \end{pmatrix} \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix}$$

Moreover, the complex parameters p and q are connected to the oscillation parameters:

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}} \quad (\text{A.15})$$

which allows obtaining some valuable identities:

$$\Delta m_s = 2|M_{12}| \quad \Delta \Gamma_s = 2|\Gamma_{12}| \cos \phi \quad \frac{q}{p} = e^{-i(\phi_M - \delta_n)} \quad (\text{A.16})$$

where the mixing phase, ϕ_M , is set as $\arg(M_{12})$, $\phi = \arg(-M_{12}/\Gamma_{12})$, and δ_n is the unobservable phase introduced in Equation 2.12. Finally, the time evolution of the system can be expressed:

$$\begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} = e^{-i\mathbf{H}t} \begin{pmatrix} |B_s^0(0)\rangle \\ |\bar{B}_s^0(0)\rangle \end{pmatrix} = \mathbf{T} \begin{pmatrix} |B_L(t)\rangle \\ |B_H(t)\rangle \end{pmatrix} = \mathbf{T} e^{-i\mathbf{H}^m t} \begin{pmatrix} |B_L(0)\rangle \\ |B_H(0)\rangle \end{pmatrix} \quad (\text{A.17})$$

and by the definitions in Equation A.10:

$$\mathbf{T} e^{-i\mathbf{H}^m t} \begin{pmatrix} |B_L(0)\rangle \\ |B_H(0)\rangle \end{pmatrix} = \mathbf{T} e^{-i\mathbf{H}^m t} \mathbf{T}^{-1} \begin{pmatrix} |B_s^0(0)\rangle \\ |\bar{B}_s^0(0)\rangle \end{pmatrix} \quad (\text{A.18})$$

Expanding further gives:

$$\begin{aligned} e^{-i\mathbf{H}t} &= \mathbf{T} e^{-i\mathbf{H}^m t} \mathbf{T}^{-1} \\ &= \begin{pmatrix} p & p \\ q & -q \end{pmatrix} \begin{pmatrix} e^{-i(M_L - i\Gamma_L/2)t} & 0 \\ 0 & e^{-i(M_H - i\Gamma_H/2)t} \end{pmatrix} \frac{1}{2pq} \begin{pmatrix} q & p \\ q & -p \end{pmatrix} \\ &= \begin{pmatrix} g_+(t) & \frac{p}{q}g_-(t) \\ \frac{q}{p}g_-(t) & g_+(t) \end{pmatrix} \end{aligned} \quad (\text{A.19})$$

where $g_{\pm}(t)$ are the following functions:

$$\begin{aligned} g_{\pm}(t) &= \frac{1}{2} \left(e^{-i(M_L - i\Gamma_L/2)t} \pm e^{-i(M_H - i\Gamma_H/2)t} \right) \\ &= \frac{1}{2} e^{-iM_s t} e^{-i\Gamma_s t/2} \left(e^{i\Delta m_s t/2} e^{-\Delta\Gamma_s t/4} \pm e^{-i\Delta m_s t/2} e^{\Delta\Gamma_s t/4} \right) \end{aligned} \quad (\text{A.20})$$

A.3 On-shell and off-shell particles

On-shell particles also referred to as particles on their mass shell, are particles whose kinematics satisfy the energy-momentum relation of special relativity. As

such, their kinematic observables satisfy the mass shell (the hyperboloid) equation in energy-momentum space defined as:

$$g^{\mu\nu} p_\nu p_\mu = p^\mu p_\mu = -m^2 \quad (\text{A.21})$$

where $g^{\mu\nu} = \eta^{\mu\nu}$ is the Minkowski space metric connection with signature $(-, +, +, +)$ and using natural units ($c = 1$), p_μ and m are the four-momentum and the invariant mass of the particle, respectively. More familiarly, this hyperboloid can also be expressed as:

$$E^2 - |\mathbf{p}|^2 c^2 = m^2 c^4 \quad (\text{A.22})$$

where E and \mathbf{p} are the total energy and the three-momentum of the particle, respectively. Off-shell particles do not obey the energy-momentum relation and do not necessarily sit on their mass shell, *e.g.* virtual particles.

A.4 Expected ϕ_s statistical uncertainty against amount of collected data at LHCb

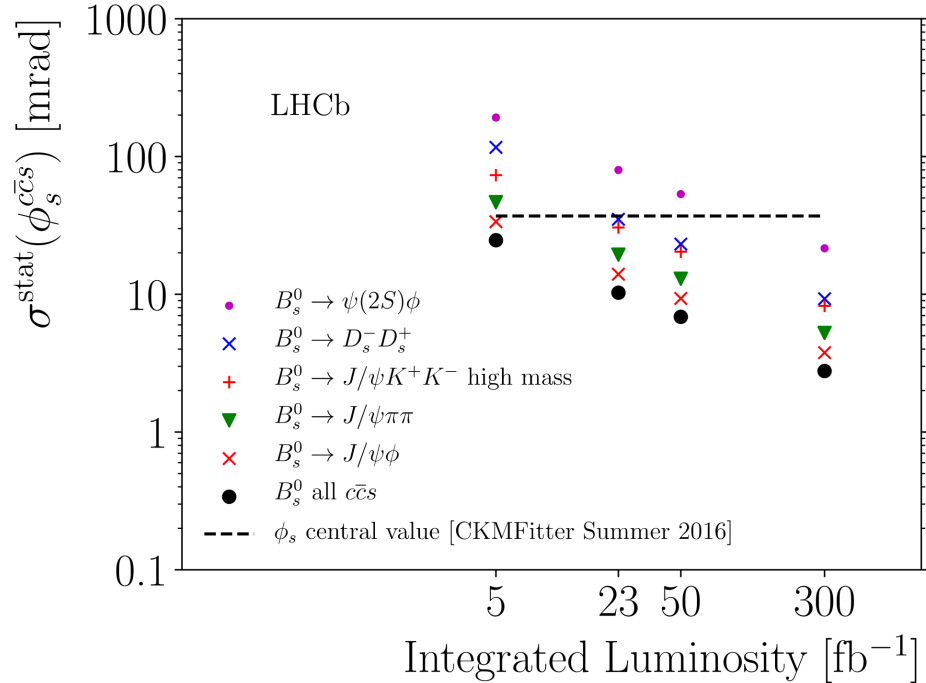


Figure A.1 Expected ϕ_s statistical uncertainty against LHCb integrated luminosity [51, 200].

A.5 P2VV angular differential decay rate in the helicity formalism

The $B_s^0 \rightarrow J/\psi \phi$ decay can be viewed as three separate two-body decays. Firstly, the B_s^0 decays into a J/ψ and a ϕ . Then the J/ψ decays into $\mu^+\mu^-$ and the ϕ decays into a pair of K^+K^- . The spin vectors of the decaying particles can be projected onto their momentum direction to give the helicity of each particle. As such, an angular dependence arises between the rotations of the spin vectors into the direction of their momenta [201].

In the case of $B_s^0 \rightarrow J/\psi \phi$ decay, the relevant diagram on Figure 2.7 shows the centre-of-mass frames and angles required to construct the amplitudes. Generally, in the centre-of-mass frame of the two decay products, the momentum direction of one of the products defines a helicity axis. In contrast, the direction opposite the mother particle momentum specifies the z -axis.

The sum of the spin projections of each of the decay-product particles onto the chosen helicity axis is equivalent to the difference of their helicities. Given that the helicity of the decaying particle (the spin projection of the mother onto the z -axis) is known (definite), the probability amplitude of measuring the required spin projections along the helicity axis for each decay product is the basis of the angular dependence. More concretely, this probability amplitude is specified by the complex conjugate of a Wigner D-matrix, a function of three Euler angles needed to identify the rotations of the spin projection from the z -axis to the helicity axis. The D-matrix is parametrised by a real-valued rotation matrix (d-matrix) and two complex phases:

$$D_{mn}^k(\varphi, \theta, \varphi') = e^{-im\varphi} d_{mn}^k(\theta) e^{-in\varphi'} \quad (\text{A.23})$$

where k is the spin of the decaying mother, m is the projection of its spin onto the z -axis, and n is the projection of its spin onto the helicity axis. The angles φ and θ are the spherical coordinates of the spin vector of the mother in the mother rest frame. φ' is the angle that orients the decay product coordinate system around the helicity axis. Spin projections along the helicity axis are independent of any boosts along it because we are not working directly with the helicities of the particles. Consequently, the projection of the spin of a particle along its helicity axis is the same as the projection of the spin along the z -axis in its rest

frame. This equivalency relates the mother and daughter frames by particular projections.

Appendix B

Further detail on some parts of Chapter 3

B.1 Čerenkov effect

B.1.1 Čerenkov angle

Currents:

$$\begin{aligned}\text{charged particle} - I &= t\beta c \\ \text{photon} - I &= tc/n\end{aligned}$$

Čerenkov angle:

$$\cos \theta_c = \frac{1}{n\beta} + \frac{\hbar k}{2p} \left(1 - \frac{1}{n^2}\right) \quad (\text{B.1})$$

where:

p – momentum of charged particle

$\hbar k$ – momentum of photon ($k = 2\pi/\lambda$)

$\hbar k \ll p$ and thus the correction for recoil is usually not needed.

Energy threshold (γ_s):

$$\gamma_s = \frac{E_s}{m_0 c^2} = \frac{1}{\sqrt{1 - \beta_s^2}} = \frac{1}{\sqrt{1 - 1/n^2}} \quad (\text{B.2})$$

measurements of γ_s allow mass determination if energy is known.

B.1.2 Čerenkov ring radius

Čerenkov photons emitted from a radiator with refractive index n when focused by an appropriately sized spherical mirror (given the small-angle approximation) form a ring of radius:

$$r = F \cdot \tan \theta_c = \frac{R}{2} \cdot \tan \theta_c \quad (\text{B.3})$$

where r is the ring radius, F is the mirror focal length, R is the radius of the spherical mirror, and θ_c is the Čerenkov angle [202]. Equation B.3 can be expanded further to reveal a method for extracting particle properties as follows:

$$r = F \sqrt{2 - \frac{2}{n} \sqrt{1 + \frac{m^2 c^2}{p^2}}} \quad (\text{B.4})$$

The angle separation between two particles of invariant masses m_1 and m_2 is then given by:

$$\theta_c \Delta \theta_c = \frac{m_1^2 - m_2^2}{2p^2} \quad (\text{B.5})$$

B.1.3 RICH Optics

The best optically focusing mirrors are based on the paraboloid shape and are commonly referred to as parabolic mirrors. However, parabolic mirrors are hard to manufacture and become increasingly expensive with size and desired optical quality. In place of this, experiments rely on the sufficient approximation that a small section of a large sphere is almost indistinguishable from the middle section of an appropriately sized paraboloid. This approximation, combined with other practical aspects related to focus calibration, is why LHCb and many others use spherical and not parabolic mirrors in their RICH detectors.

B.2 Alignment and calibration

B.2.1 Data samples and tasks

The computing power available in the Run 2 EFF allows for automated alignment and calibration tasks, providing offline quality information to the trigger reconstruction and selections, as described in Ref. [114, 203]. Dedicated samples selected by HLT1 are used to align and calibrate the detector in real-time. The alignment and calibrations are performed at regular intervals, and the resulting alignment and calibration constants are updated only if they differ significantly from the current values [4]. The primary detector alignment and calibration tasks consist of:

- the VELO alignment, followed by the alignment of the tracking stations;
- the MUON alignment;
- alignment of the rotations around various local axes in both RICH detectors of the primary and secondary mirrors;
- global time calibration of the OT
- RICH gas refractive-index calibration;
- RICH Hybrid Photon Detectors calibration;
- ECAL LED (relative) and π^0 (absolute) calibrations.

Each task has a dedicated HLT1 trigger line that supplies it with the types of events required. When the required sample sizes have been collected, the selected events are saved to the disk buffer of the EFF, and calibration and alignment tasks are performed in parallel within the EFF.

B.3 Trigger

B.3.1 TIS and TOS

LHCb trigger decisions are classified into several categories [204]:

- trigger independently of signal (TIS)
 - for simple candidate [*e.g.* a track]: trigger hits do not have overlap with user-selected signal hits
 - for composite trigger candidate [*e.g.* particle with many daughters (tracks)]: all subcomponents are TIS
 - for trigger selection: at least one subcomponent is TIS
- trigger on signal (TOS)
 - for simple candidates: trigger hits are contained within signal hits
 - for composite trigger candidate: all subcomponents are TOS
 - for trigger selection: at least one trigger candidate is TOS
- trigger on both (TOB)
 - TIS and TOS
- trigger partially on signal (TPS)
 - simple candidate: trigger hits are partially contained within signal hits [or in simple terms neither TIS nor TOS]
 - for composite candidates:
 - * default: either one subcomponent is TPS, or the subcomponents are individually exclusively TIS or TOS
 - * composite TPS via partial TOS only: subcomponents are individually exclusively TIS or TOS
 - selection: at least one candidate is TPS
- trigger used signal (TUS)
 - by definition: trigger somehow used signal hits or components
 - not the same as not TIS
 - * firing trigger producing no candidates are not TIS by definition but do not constitute TUS
 - * for selections requiring composite TPS via partial TOS only TUS has a narrower meaning than not TIS

Appendix C

Further detail on some parts of Chapter 6 and Chapter 7

C.1 Maximum likelihood estimation (fit)

Maximum likelihood estimation (MLE) is a statistical method for determining the parameters of the statistical model by maximising the likelihood that under the assumption of the model, the observed data is most probable or, in other words, such that the model best describes the underlying data sample. In this case, the statistical model is a probability density function (PDF), and the likelihood is a separate function that measures how well the model reproduces the data.

The likelihood for a single data example (a data point), in this case, a single particle decay, is given by evaluating the PDF at the values for this data example, *i.e.* the recorded decay time, reconstructed helicity angles and so on. The total likelihood represents the joint probability of the individual likelihoods. It is, thus, the product of the likelihoods of each data point in the case where data points are independent measurements.

The likelihood function becomes apparent by varying the parameters of the PDF. Thus, each likelihood function value represents the likelihood of the data for a given set of PDF parameter values. As such, the total likelihood is also a function. For a given set of data, $X = \{x_1, x_2, x_3, \dots\}$, and a probability density function,

f , with parameter set, $\Theta = \{\theta_1, \theta_2, \theta_3, \dots\}$, the total likelihood function is:

$$\mathcal{L}(\Theta|X) = \prod_i f(x_i|\Theta) \quad (\text{C.1})$$

where $f(x_i|\Theta)$ is the conditional probability of the data point, x_i , given the parameter set Θ . The global maximum of the total likelihood is the point at which the data sample is most probable, given the PDF parameters used to evaluate the likelihood. In other words, the global maximum of the total likelihood is where the current parameters represent the data best. These values are chosen to estimate the true parameter values.

However, in practice, due to computational convenience¹, the negative logarithm of the likelihood (NLL) is minimised to obtain the point at which the maximum likelihood occurs and determine the parameters. In particular, the logarithm function has the effect of scaling very small numbers and allows computers to deal better with normalised probabilities (real numbers in the range $(0, 1) \in \mathbb{R}$)². The logarithm likelihood function can be defined as $\ell(\Theta|X)$, and thus the NLL becomes:

$$\begin{aligned} \text{NLL}(\Theta|X) &= -\ell(\Theta|X) = -\ln [\mathcal{L}(\Theta|X)] \\ &= -\ln \left(\prod_i f(x_i|\Theta) \right) = -\sum_i \ln [f(x_i|\Theta)] \end{aligned} \quad (\text{C.2})$$

When talking about optimisation, it is also customary to speak of the likelihood in terms of the objective (or cost) function:

$$\begin{aligned} \hat{\text{NLL}}(\Theta|X) &= -\hat{\ell}(\Theta|X) = -\ln [\hat{\mathcal{L}}(\Theta|X)] \\ &= -\ln \left(\left[\prod_i f(x_i|\Theta) \right]^{\frac{1}{n}} \right) = -\frac{1}{n} \sum_i \ln [f(x_i|\Theta)] \end{aligned} \quad (\text{C.3})$$

Furthermore, it is worth introducing the concept of Fisher information [206]. Fisher pioneered the use of likelihood for statistical models and inference, specifically applicable here for inferring the PDF parameters from the available data.

¹It is a convention that software packages optimise for the minimum of a given function and not its maximum.

²a single-precision floating-point number has limited precision and is rounded to zero below a value of 2^{-126} [205]

C.1.1 Fisher information

Fisher information³ is essentially a measure of how much information an observable random variable, x , provides about an unknown parameter, θ , of a PDF, $f(x|\theta)$, that models x . Intuitively, if $f(x|\theta)$ has a sharp peak in θ , meaning it changes significantly with minor changes in the value of Θ , then x provides a lot of information about θ . Thus the parameters can be more easily estimated. Conversely, if $f(x|\theta)$ is very flat that means there are multiple parameter values with similar likelihoods that “predict” the data. One might suppose that this can simply be given by the log-likelihood gradient, $\frac{\partial}{\partial \theta} \ell(\theta|x)$, also called the “score”. However, at the extrema of the likelihood, the expectation value of the score is zero evaluated at the true parameter value. Thus, the Fisher information is more formally defined as the variance of the score:

$$\mathcal{I}(\theta) = \mathbb{E}_{\theta} \left\{ \left[\frac{\partial}{\partial \theta} \ell(\theta|X) \right]^2 \right\} = \int \left[\frac{\partial}{\partial \theta} \ln f(x|\theta) \right]^2 f(x|\theta) dx \geq 0 \quad (\text{C.4})$$

If the second-order derivative of $\ell(\theta|x)$ with respect to θ exists then this definition is also equivalent to [207]:

$$\mathcal{I}(\theta) = -\mathbb{E}_{\theta} \left\{ \frac{\partial^2 \ell(\theta|X)}{\partial \theta^2} \right\} \quad (\text{C.5})$$

This definition allows seeing the Fisher information as the curvature of the log-likelihood graph. Moreover, Cramér-Rao Inequality states that the Fisher information sets a lower bound on the variance of an unbiased estimator [208]:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)} \quad (\text{C.6})$$

An unbiased estimator, in this case, is when the expectation value, $\hat{\theta}$, is indeed the true value of θ .

These concepts can be extended to the multivariate case. Take the observable random variable now to be a vector $X = \{x_1, x_2, x_3, \dots\}$ and the parameter set of $\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_N\}$ specify the PDF, $f(X|\Theta)$, then the Fisher information becomes an $N \times N$ matrix with elements:

$$\mathcal{I}_{ij}(\Theta) = \mathbb{E}_{\Theta} \left[\left(\frac{\partial}{\partial \theta_i} \ell(\Theta|X) \right) \left(\frac{\partial}{\partial \theta_j} \ell(\Theta|X) \right) \right] = -\mathbb{E}_{\Theta} \left[\frac{\partial^2 \ell(\Theta|X)}{\partial \theta_i \partial \theta_j} \right] \quad (\text{C.7})$$

³also referred to in this context only as information

Similarly, the Fisher information is the negative of the expectation value of the Hessian matrix (a matrix of the second-order partial derivatives):

$$\begin{aligned}\mathbf{H}_{ij}(\ell(\Theta|X)) &= \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell(\Theta|X) \implies \\ \mathcal{I}_{ij}(\Theta) &= -\mathbb{E}_{\Theta}[\mathbf{H}_{ij}(\ell(\Theta|X))]\end{aligned}\tag{C.8}$$

Consider a more general case where there is a possibly biased estimator, $T(X)$. Its expectation value does not coincide with the true value of Θ but instead is its function, $\psi(\Theta)$. Then the Cramér-Rao inequality becomes:

$$\text{Cov}_{\Theta}(T(X)) \geq \frac{\partial\psi(\Theta)}{\partial\Theta}\mathcal{I}^{-1}(\Theta)\left[\frac{\partial\psi(\Theta)}{\partial\Theta}\right]^T\tag{C.9}$$

where $\partial\psi(\Theta)/\partial\Theta$ is the Jacobian matrix of ψ given by $\mathbf{J}_{ij} = \partial\psi_i(\Theta)/\partial\theta_j$ and the matrix inequality, $\mathbf{A} \geq \mathbf{B}$ is to mean that matrix $\mathbf{A} - \mathbf{B}$ is positive semidefinite⁴. Now, if the true value of Θ is estimated very closely, $\psi(\Theta) \approx \hat{\Theta}$, then the above becomes:

$$\text{Cov}_{\Theta}(T(X)) \geq \mathcal{I}^{-1}(\Theta)\tag{C.10}$$

Arriving at this situation is essentially the goal of maximising the likelihood and getting the best estimate of the parameters. As already mentioned, the Hessian of the log-likelihood function is the negative of the observed Fisher information. However, it follows that in the case of minimising the negative log-likelihood, the returned Hessian is equivalent to the observed Fisher information matrix and not its negative.

C.1.2 Minimising the negative logarithm of the likelihood

Minimising the NLL provides a convenient way of approximating its shape around the minimum. Since at the minimum the NLL vanishes by definition, the Taylor series expansion up to $\mathcal{O}(\Theta^2)$ is:

$$\text{NLL}(\Theta) = -\ell(\hat{\Theta}|X) + 0 + \frac{1}{2!}[\Theta - \hat{\Theta}]^T \mathbf{H}(-\ell(\hat{\Theta}|X))[\Theta - \hat{\Theta}]\tag{C.11}$$

where $\hat{\Theta}$ is the set of values for the set of parameters Θ , which minimise the NLL. As per Section C.1.1, the Hessian at the minimum negative log-likelihood

⁴ \mathbf{M} is positive semidefinite if the scalar $z^T \mathbf{M} z \geq 0$ for any column vector z with compatible dimensions.

estimate is simply the observed Fisher information, $I(\hat{\Theta})$. As such, it is nothing but the inverse of the covariance matrix:

$$\text{NLL}(\Theta) = -\ell(\hat{\Theta}|X) + \frac{1}{2}[\Theta - \hat{\Theta}]^T \Sigma^{-1}(\hat{\Theta})[\Theta - \hat{\Theta}] \quad (\text{C.12})$$

Thus, the approximation for any given parameter, θ_i , becomes:

$$\text{NLL}(\theta_i) \approx \frac{1}{2}[\theta_i - \hat{\theta}_i]^T \Sigma_{ii}^{-1}(\hat{\theta})[\theta_i - \hat{\theta}_i] + C \quad (\text{C.13})$$

where $\text{NLL}(\theta_i)$ is the single parameter function given by fixing all other parameters, $\Theta \setminus \theta_i = \mathbf{constant}$, and data, X , and allowing only θ_i to vary. This function is called the “profile likelihood” of θ_i . The constant, C , is specified by the particular value at the minimum and is the same for all $\text{NLL}(\theta_i)$ functions. It is inconsequential to the shape of the likelihood, the parameter and parameter error estimation. Furthermore:

$$\text{NLL}(\theta_i) \approx \frac{1}{2\sigma_i^2}(\theta_i - \hat{\theta}_i)^2 + C \quad (\text{C.14})$$

since the term $\Sigma_{ii} = \sigma_i^2$ is the variance of the particular parameter. So minimising the NLL gives a natural way to estimate the statistical uncertainties for the parameters. Therefore, the shape of the likelihood around the minimum resembles an upside-down bell and σ_i is an estimate for the width of the bell-shaped dip.

How well the second-order Taylor expansion approximates the NLL depends on its actual shape, and thus how close is σ_i to the true statistical uncertainty. One of the most critical factors is including larger data sets in the PDF fitting procedure. This, combined with the assumption that the parameters and data are well correlated (Section C.1.1), allows the shape of the likelihood to come closer and closer to the “smoothness” of the approximation around the minimum.

In the cases where this assumption is well reproduced, σ_i is also a good measure of the true statistical uncertainty of the parameter θ_i as it represents the standard deviation of the possible values around $\hat{\theta}_i$. For example, suppose θ_i is truly normally distributed (is described by a Gaussian distribution) when estimating it from different measurements. In that case, then the interval between $\hat{\theta}_i - \sigma_i$ and $\hat{\theta}_i + \sigma_i$ will contain the true value of θ_i in 68 % of the measurements. This represents a confidence interval with 68 % coverage in this specific case.

However, suppose the shape of the likelihood around the minimum is not as well

behaved. In that case, it could be that σ_i and $\hat{\theta}_i$ are still defined but no longer represent the same reasonable estimate of the true statistical uncertainty or a confidence interval with 68 % coverage. In general, this is commonly expected, and more involved approaches are needed to determine the standard deviation, statistical uncertainty and confidence interval.

Widely accepted method for determining these is to start at the ideal situation of a Gaussian and realise that the value of the likelihood rises by half at the edges of the 68 % confidence interval, in other words:

$$\text{NLL}(\hat{\theta}_i - \sigma_i) = \text{NLL}(\hat{\theta}_i + \sigma_i) = \frac{1}{2} + C \quad (\text{C.15})$$

Then define σ_i to be the point at which that value of the likelihood is reached. Of course, based on the exact likelihood function, the standard deviation value above and below the minimum can be different. In that case, two values exist, σ_i^+ and σ_i^- , and the confidence interval is $[\hat{\theta}_i - \sigma_i^-; \hat{\theta}_i + \sigma_i^+]$, which will approximately still have a coverage of 68 %.

Sometimes, rather inconveniently, the likelihood exhibits multiple minima or an extensive shallow minimum or a combination of these. The estimate at the point $\hat{\theta}_i$ no longer represents the distribution nicely. In such cases, confidence intervals are better suited for the task. As such, the intervals are constructed by extending the method of $\frac{1}{2} + C$ value from above. Similarly, the intervals should include the points between which the NLL values are smaller than a specified value. Commonly these are chosen to be the values that a parabola (second-order Taylor approximation) will reach at $\hat{\theta}_i \pm n \cdot \sigma_i$, which correspond to a value of $\frac{1}{2}n^2 + C$ since a parabola is symmetric around its minimum. These are known as the n -sigma intervals.

C.2 Fit to weighted data

As mentioned in Section 6.3, the data sample used in the analysis of $B_s^0 \rightarrow J/\psi \phi$ decays contains events from both signal sources and background sources. The final model fit to data is performed by fitting the signal PDF to the sample with statistically subtracted background event distribution. In particular, the latter procedure introduces event candidates from background sources into the data sample with negative weights. Then when fitting for the minimum NLL with the

sample, each event's contribution is multiplied by its corresponding weight:

$$\text{NLL} = - \sum_i \ln(w_i \cdot \mathcal{L}_i) \quad (\text{C.16})$$

where \mathcal{L}_i is the contribution to the likelihood for the i^{th} event and w_i is the corresponding weight. This aims to effectively make the likelihood independent of the background present in data assuming the final sum of contributions from background present and background introduced is zero.

This approach minimises the NLL relevant only to the signal distribution present in the data sample. As such, the position of the NLL minimum should represent a good estimate of the relevant parameters. However, the uncertainties to the values of the parameters can no longer be correctly determined from looking the shape of the NLL around the minimum only. This can be illustrated by taking a look at a fit in intervals of some variable. In such a case, the observed number of events in each interval can be modelled. The uncertainties of the model parameters will be related to the uncertainties of the number of events in each interval.

In the particular case of the data sample being mainly concerned with the observed number of events, the Poisson distribution is most relevant. It models the number of times an event distributed according to $f(X|\Theta)$ occurs in a specific interval. Thus it would modify the likelihood as such:

$$\mathcal{L} = \frac{\nu^n}{n!} e^{-\nu} \prod_i^n f(x_i|\Theta) = \frac{e^{-\nu}}{n!} \prod_i^n \nu f(x_i|\Theta) \quad (\text{C.17})$$

where ν is both the mean (expected value) and the variance of the number of observed events, while n is their actual observed value.

Then an estimate of the expected number of events in each interval is given by n . Consequently, the uncertainty on that estimate is given by the square root of the variance: \sqrt{n} . For weighted sets, the number of events is replaced by the sum of weights instead, $W = \sum w_i$. Then it would follow that the uncertainty is given now by \sqrt{W} . However, consider a sort of scaling factor, w_s , which is the same for all weights. Then the expected number of events will scale with this factor, but the absolute uncertainty will scale with the square root of it. This implies a contradiction since no new information was added to the data and, thus, the relative uncertainty should remain constant. To resolve it, the correct uncertainty

should instead be the square root of the sum of the squared weights:

$$\sigma = \sqrt{\sum w_i^2} \quad (\text{C.18})$$

which would scale correctly with w_s . It is convenient to define a correction factor α_c :

$$\alpha_c = \frac{\sum w_i}{\sum w_i^2} \quad (\text{C.19})$$

that can be used to obtain the correct uncertainty from the NLL shape by multiplying the profile likelihood as:

$$\text{NLL}(\theta_i) \approx \frac{\alpha_c}{2\sigma_i^2} (\theta_i - \hat{\theta}_i)^2 + C \quad (\text{C.20})$$

without the need to split the dataset in intervals. This would also preserve the position of the minimum.

Moreover, as multiple species of data are merged and analysed together, it is worth considering an extended NLL definition with one or more species of data representing analysis relevant sources (signal) and one or more species of background, data from analysis irrelevant sources:

$$\text{NLL} = -\ell = - \left(\sum_i^N \left\{ \ln \left[\sum_s^{N_s} n_s f(X|\Theta) \right] \right\} - \sum_s^{N_s} n_s + C \right) \quad (\text{C.21})$$

where N is the total number of events in the data sample, N_s is the number of species of events in the data sample, and n_s is the number of events expected in the s^{th} species on average [174]. Using this method of constructing the likelihood allows unfolding the contributions of each species of events in the data using a tool like `sPlot` [174]. This combined with the correction factor α_c enables the determination of event yields of signal and background, weights used in the statistical background subtraction and correct uncertainties on the parameters Θ from the fit.

C.3 Nuisance parameters

Nuisance parameters, $\Theta_S = \{\theta_1, \theta_2, \dots, \theta_S\} \subsetneq \Theta$ are required parameters of either the PDF or NLL but which are not specifically of physical, observational or other direct significance. Hence, these are of no particular interest but arise as

part of the model and need to be accounted for nonetheless. The most intuitive example are the parameters describing the background model in a fit: they must be determined to disambiguate the signal parameters but are purely dataset dependent quantities and do not persist in the physically relevant parameters.

C.4 Detailed information about the so-called TISTOS bug

In February of 2017, while doing some summary data plots to include in a publicity campaign, it was discovered that most of the dimuon pairs reconstructed by the offline algorithms failed to trigger-on-signal (TOS, Section B.3) correctly. The problem was soon after tracked down to an unforeseen interaction between a new and experimental at the time **Turbo** stream and the established logic of the HLT2 tool that assigns the TIS or TOS decisions called the **Hlt2TriggerTisTos** tool.

- Original internal report sparking further investigation: Click here or visit: <https://its.cern.ch/jira/browse/LHCBPS-1702>.
- Actual software issue identified: Click here or visit: <https://its.cern.ch/jira/browse/LHCBPS-1746>.

The problem affected charged particle tracks only in reconstructed Run 2 data samples. This was unfortunate for analyses that had both charge and neutral particles in their final state as any mitigating techniques applied would pollute the correctly labelled neutral particle information as the tool is applied to all candidates on a per event basis.

The core of the problem had to do with the way the HLT2 decisions interact with the algorithm responsible for keeping track of event information — keeping a record of how an event was selected by the trigger, which trigger lines fired on it, which algorithms have already processed it, what are the different responses by those algorithms and so on and so forth. This information is handled by **Hlt2SelReports** tool. One key piece of information that is kept is the so-called **CALO LHCbID**. Historically, **CALO LHCbID** was assigned to a particle with no associated track. This was done to help **Hlt2SelReports** algorithm to avoid overlapping event selections and store events only once. However, with the

introduction of the new **Turbo** stream and its associated algorithms, it could happen that a particle that was processed early by the **Turbo** algorithms will have a **CALO LHCbID** and an associated track since these algorithms store information differently. In such cases, **Hlt2SelReports** will recognise that the event is already processed and will not create its report anew, meaning later algorithms will see both a **CALO LHCbID** and an associated track. This meant that a particle's **TOS** decision depended on whether it was also a candidate in a **Turbo** selection. The bug resulted in the **Hlt2TriggerTisTos** finding too many differences between triggered and reconstructed candidates and being unable to associate them, it failed and did not mark the candidates as **TOS** correctly.

The problem was apparent in the significant decrease of **TOS** labelled events and there were also uncharacteristic distortions in decay time acceptance. For analyses using only charged particles in their final state, such as $B_s^0 \rightarrow J/\psi \phi$, it was discovered that forcing some of the inputs to the **TISTOS** to be unphysical can cause it to recalculate the correct information from scratch and circumvent the issue. Specifically, the following lines of code added to the configuration of the **DAVINCI** options restored the correct responses:

```
#!/bin/environment python
# DaVinci v42r5p1
...
# change TOSFracMu for B
from Configurables import TriggerTisTos
tuple_B2jpsiphi.B.TupleToolTISTOS.addTool(TriggerTisTos())
tuple_B2jpsiphi.B.TupleToolTISTOS.TriggerTisTos.TOSFracMuon = 0.
tuple_B2jpsiphi.B.TupleToolTISTOS.TriggerTisTos.TOSFracEcal = 0.
tuple_B2jpsiphi.B.TupleToolTISTOS.TriggerTisTos.TOSFracHcal = 0.

# Change TOSFracMu for Jpsi
tuple_B2jpsiphi.Jpsi.TupleToolTISTOS.addTool(TriggerTisTos())
tuple_B2jpsiphi.Jpsi.TupleToolTISTOS.TriggerTisTos.TOSFracMuon = 0.
tuple_B2jpsiphi.Jpsi.TupleToolTISTOS.TriggerTisTos.TOSFracEcal = 0.
tuple_B2jpsiphi.Jpsi.TupleToolTISTOS.TriggerTisTos.TOSFracHcal = 0.
...
```

Listing C.1 *DAVINCI configuration needed to mitigate the TISTOS bug*

For versions of **DAVINCI** later than **v42r6p1**, the workaround was not necessary.

C.5 Impact of VELO track fit mismatch between data and simulation

On 30th March 2017, LHCb discovered internally that the VELO track fit had a mismatch in the parametrisation of its error estimation between data and simulation [209]. Afterwards, it was confirmed that this discrepancy affected the $B_s^0 \rightarrow J/\psi \phi$ analysis (Figure C.1). This prompted the analysis to switch to

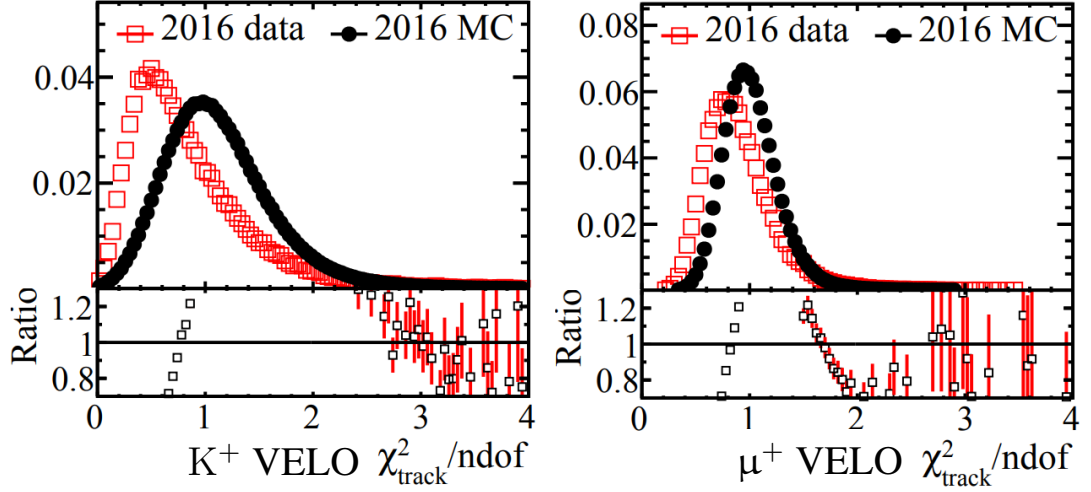


Figure C.1 *Comparison of the distributions of K (left) and μ (right) VELO $\chi^2_{\text{track}}/\text{ndf}$ between collision data and simulation for the 2016 data-taking year [210].*

DAVINCI v44r1 and applying the appropriate fixes to synchronise the algorithms used on data and simulation. This ultimately resolved the discrepancy.

C.6 Flavour Tagging

Tagger	OS	SSK
p_0	$0.3890 \pm 0.0007 \pm 0.0028$	$0.4325 \pm 0.0108 \pm 0.0030$
p_1	$0.849 \pm 0.006 \pm 0.027$	$0.92 \pm 0.13 \pm 0.02$
Δp_0	0.0090 ± 0.0014	0.00 ± 0.03
Δp_1	0.014 ± 0.012	0.00 ± 0.03
$\langle \eta \rangle$	0.360	0.417

Table C.1 Calibration parameters for the OS and SSK taggers. Where given, the first uncertainty is statistical, and the second is systematic [5].

Category	$\epsilon_{\text{tag}}(\%)$	D^2	$\epsilon_{\text{tag}} D^2(\%)$
OS only	11.4	0.078	0.88 ± 0.04
SSK only	42.6	0.032	1.38 ± 0.30
OS and SSK	23.8	0.104	2.47 ± 0.15
Total	77.8	0.061	4.73 ± 0.34

Table C.2 Overall tagging performance for $B_s^0 \rightarrow J/\psi K^+ K^-$. The uncertainty on $\epsilon_{\text{tag}} D^2$ is obtained by varying the tagging calibration parameters within their statistical and systematic uncertainties summed in quadrature [5].

C.7 Details for combined fit between Run 1, 2015 and 2016 results

	ϕ_s	$ \lambda $	$\Gamma_s - \Gamma_d$	$\Delta\Gamma_s$	Δm_s	$ A_\perp ^2$	$ A_0 ^2$	$\delta_\perp - \delta_0$	$\delta_\parallel - \delta_0$
ϕ_s	1.00	0.10	-0.02	-0.03	0.02	0.01	-0.01	0.07	0.00
$ \lambda $		1.00	0.04	-0.04	-0.05	0.03	-0.02	-0.04	0.03
$\Gamma_s - \Gamma_d$			1.00	-0.35	0.04	0.28	-0.17	0.01	0.01
$\Delta\Gamma_s$				1.00	-0.01	-0.62	0.40	-0.05	-0.01
Δm_s					1.00	0.01	-0.01	0.62	0.02
$ A_\perp ^2$						1.00	-0.67	0.03	0.01
$ A_0 ^2$							1.00	-0.06	-0.06
$\delta_\perp - \delta_0$								1.00	0.28
$\delta_\parallel - \delta_0$									1.00

Table C.3 *Correlation matrix for the results taking into account correlated systematics between Run 1 and the 2015 and 2016 results [5].*

	ϕ_s	$ \lambda $	Γ_s	$\Delta\Gamma_s$
ϕ_s	1.00	0.06	-0.01	-0.03
$ \lambda $		1.00	0.03	-0.02
Γ_s			1.00	-0.17
$\Delta\Gamma_s$				1.00

Table C.4 *Correlation matrix for the results obtained taking into account correlated systematics between the considered analyses: $B_s^0 \rightarrow J/\psi K^+ K^-$ Run 1 [14], $B_s^0 \rightarrow J/\psi K^+ K^-$ (2015 and 2016) [5], $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ Run 1 [193], and $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ (2015 and 2016) [211].*

C.8 Wrong PV component in decay time resolution studies

The wrong PV component in the prompt sample is handled by simulating the effect of incorrect PV assignment during reconstruction and extracting the resultant shape in decay time. As LHCb reconstruction cannot directly associate a decay vertex with its true PV, a different approach is taken, called event mixing [175]. For each reconstructed event, a different PV is injected from a randomly chosen separate event. The new PV is flagged as such for later reference⁵. The best PV is determined anew by repeating the whole

⁵normal reconstruction algorithms cannot see that flag

reconstruction process from the beginning. Then the observed decay time distribution for the reconstructed events with the flagged wrong PV is fitted with a double-sided double-exponential [175].

C.9 S-wave fractions parameters

The results from the maximum likelihood fit described in Section 7.7 include the S-wave fractions and phase differences in each $m(K^+K^-)$ bin [5], summarised in Table C.5.

Parameter	Value
F_{S1}	$0.492 \pm 0.043 \pm 0.010$
F_{S2}	$0.041 \pm 0.008 \pm 0.006$
F_{S3}	$0.0044^{+0.0030}_{-0.0017} \pm 0.0014$
F_{S4}	$0.0069^{+0.0062}_{-0.0045} \pm 0.0016$
F_{S5}	$0.073 \pm 0.013 \pm 0.004$
F_{S6}	$0.152^{+0.019}_{-0.018} \pm 0.009$
$\delta_{S1} - \delta_{\perp} [\text{rad}]$	$+2.21^{+0.17}_{-0.20} \pm 0.20$
$\delta_{S2} - \delta_{\perp} [\text{rad}]$	$+1.56 \pm 0.29 \pm 0.05$
$\delta_{S3} - \delta_{\perp} [\text{rad}]$	$+1.09^{+0.47}_{-0.36} \pm 0.10$
$\delta_{S4} - \delta_{\perp} [\text{rad}]$	$-0.28^{+0.16}_{-0.26} \pm 0.12$
$\delta_{S5} - \delta_{\perp} [\text{rad}]$	$-0.54^{+0.09}_{-0.10} \pm 0.02$
$\delta_{S6} - \delta_{\perp} [\text{rad}]$	$-1.10^{+0.13}_{-0.16} \pm 0.11$

Table C.5 *Values of the S-wave parameters in each $m(K^+K^-)$ bin. The first uncertainty is statistical and the second systematic [5].*

Bibliography

- [1] Sean Benson and Konstantin Gizdov. ‘NNDrone: a toolkit for the mass application of machine learning in High Energy Physics’. In: (2017). DOI: 10.1016/j.cpc.2019.03.002. arXiv: 1712.09114 [hep-ex] (cit. on pp. v, 81, 82, 85–95, 97).
- [2] Konstantin Gizdov. ‘Strategy and automation of the quality assurance testing of MaPMTs for the LHCb RICH upgrade’. In: *Nucl. Instrum. Meth. A* (2019). DOI: 10.1016/j.nima.2019.05.046 (cit. on pp. v, 100–104, 106–111).
- [3] Konstantin Gizdov. ‘Measurement of CP Violation in $B_s^0 \rightarrow J/\psi\phi$ Decays’. In: *Proceedings, 16th Conference on Flavor Physics and CP Violation (FPCP 2018): Hyderabad, India, July 14-18, 2018*. Ed. by Anjan Giri and Rukmani Mohanta. Vol. 234. 2019, pp. 245–251. DOI: 10.1007/978-3-030-29622-3 (cit. on p. v).
- [4] R. Aaij et al. ‘Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC’. In: *JINST* 14.04 (2019), P04013. DOI: 10.1088/1748-0221/14/04/P04013. arXiv: 1812.10790 [hep-ex] (cit. on pp. v, 43, 51, 53, 54, 61, 69–72, 74, 80, 82, 201).
- [5] R. Aaij et al. ‘Updated measurement of time-dependent CP -violating observables in $B_s^0 \rightarrow J/\psi K^+ K^-$ decays’. In: *The European Physical Journal C* 79.8 (Aug. 2019), p. 706. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7159-8. arXiv: 1906.08356 [hep-ex]. URL: <https://doi.org/10.1140/epjc/s10052-019-7159-8> (cit. on pp. v, 35, 121, 122, 154, 156, 159, 162, 168, 171, 176, 182, 183, 187, 214–216).
- [6] Akiva Goldsman. ‘Et in Arcadia Ego, Part 2’. English. In: *Star Trek: Picard* 10. CBS (Mar. 2020). URL: <https://www.cbs.com/shows/star-trek-picard/> (visited on 30/03/2020) (cit. on p. vii).
- [7] Laurent Canetti, Marco Drewes and Mikhail Shaposhnikov. ‘Matter and antimatter in the universe’. In: *New Journal of Physics* 14.9 (Sept. 2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012. URL: <https://doi.org/10.1088%2F1367-2630%2F14%2F9%2F095012> (cit. on p. 1).

- [8] Andrei D Sakharov. ‘Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe’. In: *Soviet Physics Uspekhi* 34.5 (May 1991), pp. 392–393. DOI: 10.1070/pu1991v034n05abeh002497. URL: <https://doi.org/10.1070%2Fpu1991v034n05abeh002497> (cit. on pp. 1, 14, 22).
- [9] Glennys R. Farrar and M. E. Shaposhnikov. ‘Baryon asymmetry of the universe in the standard electroweak theory’. In: *Phys. Rev. D* 50 (1994), p. 774. DOI: 10.1103/PhysRevD.50.774. arXiv: hep-ph/9305275 [hep-ph] (cit. on p. 1).
- [10] M. B. Gavela et al. ‘Standard model CP violation and baryon asymmetry. Part 1: Zero temperature’. In: *Nucl. Phys. B* 430 (1994), pp. 345–381. DOI: 10.1016/0550-3213(94)00409-9. arXiv: hep-ph/9406288 [hep-ph] (cit. on p. 1).
- [11] Patrick Huet and Eric Sather. ‘Electroweak baryogenesis and standard model CP violation’. In: *Phys. Rev. D* 51 (1995), pp. 379–394. DOI: 10.1103/PhysRevD.51.379. arXiv: hep-ph/9404302 [hep-ph] (cit. on p. 1).
- [12] A. J. Bevan et al. ‘The Physics of the B Factories’. In: *Eur. Phys. J. C* 74 (2014), p. 3026. DOI: 10.1140/epjc/s10052-014-3026-9. arXiv: 1406.6311 [hep-ex] (cit. on p. 2).
- [13] Y. Amhis et al. ‘Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016’. In: *Eur. Phys. J. C* 77.12 (2017), p. 895. DOI: 10.1140/epjc/s10052-017-5058-4. arXiv: 1612.07233 [hep-ex] (cit. on pp. 3, 186).
- [14] Roel Aaij et al. ‘Precision measurement of CP violation in $B_s^0 \rightarrow J/\psi K^+ K^-$ decays’. In: *Phys. Rev. Lett.* 114.4 (2015), p. 041801. DOI: 10.1103/PhysRevLett.114.041801. arXiv: 1411.3104 [hep-ex] (cit. on pp. 2, 35, 154, 163, 170, 171, 182, 215).
- [15] R Aaij et al. ‘Measurement of CP violation and the B_s^0 meson decay width difference with $B_s^0 \rightarrow J/\psi K^+ K^-$ and $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ decays’. In: *Phys. Rev. D* 87.11 (2013), p. 112010. DOI: 10.1103/PhysRevD.87.112010. arXiv: 1304.2600 [hep-ex] (cit. on pp. 2, 25, 26, 29, 35, 119, 124, 146, 154).
- [16] C Bozzi and S Roiser and. ‘The LHCb software and computing upgrade for Run 3: opportunities and challenges’. In: *Journal of Physics: Conference Series* 898 (Oct. 2017), p. 112002. DOI: 10.1088/1742-6596/898/11/112002. URL: <https://doi.org/10.1088%2F1742-6596%2F898%2F11%2F112002> (cit. on p. 3).
- [17] CERN The LHCb Collaboration. *Upgrade Software and Computing*. Tech. rep. CERN-LHCC-2018-007. LHCB-TDR-017. Geneva: CERN, Mar. 2018. URL: <https://cds.cern.ch/record/2310827> (cit. on pp. 4, 78–80).

- [18] CERN (Meyrin) LHCb Collaboration. *Computing Model of the Upgrade LHCb experiment*. Tech. rep. CERN-LHCC-2018-014. LHCb-TDR-018. Geneva: CERN, May 2018. URL: <https://cds.cern.ch/record/2319756> (cit. on p. 4).
- [19] *LHCb Trigger and Online Upgrade Technical Design Report*. Tech. rep. CERN-LHCC-2014-016. LHCb-TDR-016. May 2014. URL: <https://cds.cern.ch/record/1701361> (cit. on pp. 4, 73, 74, 78, 79).
- [20] LHCb Collaboration. *LHCb PID Upgrade Technical Design Report*. Tech. rep. CERN-LHCC-2013-022. LHCb-TDR-014. Nov. 2013. URL: <https://cds.cern.ch/record/1624074> (cit. on pp. 4, 77, 99, 100).
- [21] S Amato et al. *LHCb RICH: Technical Design Report*. Technical Design Report LHCb. Geneva: CERN, 2000. URL: <https://cds.cern.ch/record/494263> (cit. on pp. 4, 55–59).
- [22] Ernest S. Abers and Benjamin W. Lee. ‘Gauge theories’. In: *Physics Reports* 9.1 (1973), pp. 1–2. ISSN: 0370-1573. DOI: [https://doi.org/10.1016/0370-1573\(73\)90027-6](https://doi.org/10.1016/0370-1573(73)90027-6). URL: <http://www.sciencedirect.com/science/article/pii/0370157373900276> (cit. on p. 7).
- [23] Sheldon L. Glashow. ‘Partial-symmetries of weak interactions’. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2). URL: <http://www.sciencedirect.com/science/article/pii/0029558261904692> (cit. on p. 7).
- [24] Steven Weinberg. ‘A Model of Leptons’. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264> (cit. on pp. 7, 10).
- [25] Abdus Salam. ‘Weak and Electromagnetic Interactions’. In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034 (cit. on p. 7).
- [26] F. Englert and R. Brout. ‘Broken Symmetry and the Mass of Gauge Vector Mesons’. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on pp. 7, 9).
- [27] Peter W. Higgs. ‘Broken Symmetries and the Masses of Gauge Bosons’. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on pp. 7, 9).
- [28] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble. ‘Global Conservation Laws and Massless Particles’. In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on pp. 7, 9).

- [29] H. Fritzsch, M. Gell-Mann and H. Leutwyler. ‘Advantages of the color octet gluon picture’. In: *Physics Letters B* 47.4 (1973), pp. 365–368. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(73\)90625-4](https://doi.org/10.1016/0370-2693(73)90625-4). URL: <http://www.sciencedirect.com/science/article/pii/S0370269373906254> (cit. on p. 7).
- [30] S. Chatrchyan et al. ‘Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC’. In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.021>. URL: <http://www.sciencedirect.com/science/article/pii/S0370269312008581> (cit. on p. 7).
- [31] G. Aad et al. ‘Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC’. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.020>. URL: <http://www.sciencedirect.com/science/article/pii/S037026931200857X> (cit. on p. 7).
- [32] Michela D’Onofrio and Kari Rummukainen. ‘Standard model cross-over on the lattice’. In: *Phys. Rev. D* 93 (2 Jan. 2016), p. 025003. DOI: 10.1103/PhysRevD.93.025003. URL: <https://link.aps.org/doi/10.1103/PhysRevD.93.025003> (cit. on pp. 7, 10).
- [33] Various contributors. *Standard Model of Elementary Particles*. Wikimedia Foundation, Inc., Sept. 2018. URL: https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg (visited on 30/09/2020) (cit. on p. 8).
- [34] H. Fritzsch and P. Minkowski. ‘Colored gluons, bjorken scaling and quark confinement’. In: *Physics Letters B* 49.5 (1974), pp. 462–464. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(74\)90636-4](https://doi.org/10.1016/0370-2693(74)90636-4). URL: <http://www.sciencedirect.com/science/article/pii/S0370269374906364> (cit. on p. 11).
- [35] Y. Fukuda et al. ‘Evidence for Oscillation of Atmospheric Neutrinos’. In: *Phys. Rev. Lett.* 81 (8 Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562> (cit. on p. 13).
- [36] Q. R. Ahmad et al. ‘Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory’. In: *Phys. Rev. Lett.* 87 (7 July 2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.87.071301> (cit. on p. 13).
- [37] Q. R. Ahmad et al. ‘Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory’. In: *Phys. Rev. Lett.* 89 (1 June 2002), p. 011301. DOI: 10.1103/PhysRevLett.89.011301. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.89.011301> (cit. on p. 13).

- [38] Emmy Noether. ‘Invariant Variation Problems’. In: *Gott. Nachr.* 1918 (1918), pp. 235–257. DOI: 10.1080/00411457108231446. arXiv: physics/0503066 (cit. on p. 13).
- [39] E. Noether. ‘Invariante Variationsprobleme’. ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: <http://eudml.org/doc/59024> (cit. on p. 13).
- [40] Peter E. Hydon and Elizabeth L. Mansfield. ‘Extensions of Noether’s Second Theorem: from continuous to discrete systems’. In: *Proceedings: Mathematical, Physical and Engineering Sciences* 467.2135 (2011), pp. 3206–3221. ISSN: 13645021. URL: <http://www.jstor.org/stable/23065657> (cit. on p. 13).
- [41] Anthony C. L. Ashton. ‘Conservation Laws and Non-Lie Symmetries for Linear PDEs’. In: *Journal of Nonlinear Mathematical Physics* 15.3 (2008), pp. 316–332. DOI: 10.2991/jnmp.2008.15.3.5. eprint: <https://doi.org/10.2991/jnmp.2008.15.3.5>. URL: <https://doi.org/10.2991/jnmp.2008.15.3.5> (cit. on p. 13).
- [42] R. D. Peccei and Helen R. Quinn. ‘CP Conservation in the Presence of Pseudoparticles’. In: *Phys. Rev. Lett.* 38 (25 June 1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.38.1440> (cit. on p. 14).
- [43] J. H. Christenson et al. ‘Evidence for the 2π Decay of the K_2^0 Meson’. In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.138> (cit. on pp. 14, 18).
- [44] The Nobel Foundation. *The Nobel Prize in Physics 1980*. URL: <https://www.nobelprize.org/prizes/physics/1980/summary> (visited on 26/10/2020) (cit. on pp. 14, 18).
- [45] Nicola Cabibbo. ‘Unitary Symmetry and Leptonic Decays’. In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.10.531> (cit. on pp. 18, 191).
- [46] Makoto Kobayashi and Toshihide Maskawa. ‘CP-Violation in the Renormalizable Theory of Weak Interaction’. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. eprint: <https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf>. URL: <https://doi.org/10.1143/PTP.49.652> (cit. on pp. 18, 191).
- [47] M.B. Gavela et al. ‘Standard model CP-violation and baryon asymmetry (II). Finite temperature’. In: *Nuclear Physics B* 430.2 (Nov. 1994), pp. 382–426. ISSN: 0550-3213. DOI: 10.1016/0550-3213(94)00410-2. URL: [http://dx.doi.org/10.1016/0550-3213\(94\)00410-2](http://dx.doi.org/10.1016/0550-3213(94)00410-2) (cit. on p. 22).

- [48] M. Shaposhnikov. ‘Is there a hot electroweak phase transition at large Higgs masses?’ In: *32nd Rencontres de Moriond: Electroweak Interactions and Unified Theories*. 1997, pp. 441–446 (cit. on p. 22).
- [49] M. B. GAVELA et al. ‘STANDARD MODEL CP-VIOLATION AND BARYON ASYMMETRY’. In: *Modern Physics Letters A* 09.09 (Mar. 1994), pp. 795–809. ISSN: 1793-6632. DOI: 10.1142/S0217732394000629. URL: <http://dx.doi.org/10.1142/S0217732394000629> (cit. on p. 22).
- [50] Patrick Huet and Eric Sather. ‘Electroweak baryogenesis and standard model CP violation’. In: *Physical Review D* 51.2 (Jan. 1995), pp. 379–394. ISSN: 0556-2821. DOI: 10.1103/physrevd.51.379. URL: <http://dx.doi.org/10.1103/PhysRevD.51.379> (cit. on p. 22).
- [51] J. Charles et al. ‘CP violation and the CKM matrix: assessing the impact of the asymmetric B factories’. In: *The European Physical Journal C* 41.1 (May 2005), pp. 1–131. ISSN: 1434-6052. DOI: 10.1140/epjc/s2005-02169-1. URL: <http://dx.doi.org/10.1140/epjc/s2005-02169-1> (cit. on pp. 22–24, 34, 195).
- [52] J. Charles et al. ‘Predictions of selected flavour observables within the Standard Model’. In: *Phys. Rev. D* 84 (2011), p. 033005. DOI: 10.1103/PhysRevD.84.033005. arXiv: 1106.4041 [hep-ph] (cit. on p. 27).
- [53] Jhovanny Mejia-Guisao. *Experimental status of CPV in B_s^0 mixing*. 2020. arXiv: 2008.08199 [hep-ex] (cit. on p. 27).
- [54] Tristan du Pree. ‘Search for a Strange Phase in Beautiful Oscillations’. Presented on 22 Oct 2010. 2010. URL: <https://cds.cern.ch/record/1299931> (cit. on pp. 30, 165, 166).
- [55] Sheldon Stone and Liming Zhang. ‘S-waves and the Measurement of CP Violating Phases in B_s Decays’. In: *Phys. Rev. D* 79 (2009), p. 074024. DOI: 10.1103/PhysRevD.79.074024. arXiv: 0812.2832 [hep-ph] (cit. on p. 31).
- [56] Xin Liu, Wei Wang and Yuehong Xie. ‘Penguin pollution in $B \rightarrow J/\psi V$ decays and impact on the extraction of the $B_s - \bar{B}_s$ mixing phase’. In: *Phys. Rev. D* 89.9 (2014), p. 094010. DOI: 10.1103/PhysRevD.89.094010. arXiv: 1309.0313 [hep-ph] (cit. on p. 31).
- [57] A. Lenz et al. ‘Anatomy of new physics in $B \rightarrow \bar{B}$ mixing’. In: *Physical Review D* 83.3 (Feb. 2011). ISSN: 1550-2368. DOI: 10.1103/physrevd.83.036004. URL: <http://dx.doi.org/10.1103/PhysRevD.83.036004> (cit. on p. 32).
- [58] J. Charles et al. ‘Current status of the standard model CKM fit and constraints on $\Delta F = 2$ new physics’. In: *Physical Review D* 91.7 (Apr. 2015). ISSN: 1550-2368. DOI: 10.1103/physrevd.91.073007. URL: <http://dx.doi.org/10.1103/PhysRevD.91.073007> (cit. on pp. 33, 34).

- [59] Roel Aaij et al. ‘Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+ \mu^-$ decays’. In: *Phys. Rev. Lett.* 118.19 (2017), p. 191801. DOI: 10.1103/PhysRevLett.118.191801. arXiv: 1703.05747 [hep-ex] (cit. on p. 35).
- [60] R. Aaij et al. ‘Test of Lepton Flavor Universality by the measurement of the $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ branching fraction using three-prong τ decays’. In: *Phys. Rev.* D97.7 (2018), p. 072013. DOI: 10.1103/PhysRevD.97.072013. arXiv: 1711.02505 [hep-ex] (cit. on p. 35).
- [61] R. Aaij et al. ‘Test of lepton universality with $B^0 \rightarrow K^{*0} \ell^+ \ell^-$ decays’. In: *JHEP* 08 (2017), p. 055. DOI: 10.1007/JHEP08(2017)055. arXiv: 1705.05802 [hep-ex] (cit. on p. 35).
- [62] Lyndon Evans and Philip Bryant. ‘LHC Machine’. In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001 (cit. on p. 35).
- [63] K. Aamodt et al. ‘The ALICE experiment at the CERN LHC’. In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002 (cit. on p. 36).
- [64] G. Aad et al. ‘The ATLAS Experiment at the CERN Large Hadron Collider’. In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003 (cit. on p. 36).
- [65] S. Chatrchyan et al. ‘The CMS Experiment at the CERN LHC’. In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on p. 36).
- [66] A. Augusto Alves Jr. et al. ‘The LHCb Detector at the LHC’. In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005 (cit. on pp. 36, 47, 48, 65).
- [67] Thomas Lenzi. ‘Development and Study of Different Muon Track Reconstruction Algorithms for the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors’. MA thesis. U. Brussels (main), 2013. arXiv: 1306.0858 [physics.ins-det] (cit. on p. 36).
- [68] Georges Aad et al. ‘Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC’. In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on p. 36).
- [69] L. Rossi. ‘The LHC superconducting magnets’. In: *Proceedings of the 2003 Particle Accelerator Conference*. Vol. 1. May 2003, 141–145 Vol.1. DOI: 10.1109/PAC.2003.1288863 (cit. on p. 37).
- [70] Roel Aaij et al. ‘LHCb Detector Performance’. In: *Int. J. Mod. Phys.* A30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex] (cit. on pp. 37, 43, 52–54, 117, 129).
- [71] F. Follin and D. Jacquet. ‘Implementation and experience with luminosity levelling with offset beam’. In: *Proceedings, ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders (BB2013): CERN, Geneva, Switzerland, March 18-22 2013*. [183(2014)]. 2014, pp. 183–187. DOI: 10.5170/CERN-2014-004.183. arXiv: 1410.3667 [physics.acc-ph] (cit. on p. 37).

- [72] Rolf Lindner. ‘LHCb layout_2. LHCb schema_2’. LHCb Collection. Feb. 2008. URL: <http://cds.cern.ch/record/1087860> (cit. on p. 38).
- [73] Roel Aaij et al. ‘Measurement of the b -quark production cross-section in 7 and 13 TeV pp collisions’. In: *Phys. Rev. Lett.* 118.5 (2017). [Erratum: *Phys. Rev. Lett.* 119, no. 16, 169901 (2017)], p. 052002. DOI: 10.1103/PhysRevLett.119.169901, 10.1103/PhysRevLett.118.052002. arXiv: 1612.05140 [hep-ex] (cit. on pp. 39, 119).
- [74] A. Abashian et al. ‘The Belle Detector’. In: *Nucl. Instrum. Meth.* A479 (2002), pp. 117–232. DOI: 10.1016/S0168-9002(01)02013-7 (cit. on p. 39).
- [75] D. Boutigny et al. ‘BaBar technical design report’. In: (Mar. 1995) (cit. on p. 39).
- [76] Jan Fiete Grosse-Oetringhaus and Klaus Reygers. ‘Charged-Particle Multiplicity in Proton-Proton Collisions’. In: *J. Phys.* G37 (2010), p. 083001. DOI: 10.1088/0954-3899/37/8/083001. arXiv: 0912.0023 [hep-ex] (cit. on p. 39).
- [77] Roel Aaij et al. ‘Measurement of charged particle multiplicities and densities in pp collisions at $\sqrt{s} = 7$ TeV in the forward region’. In: *Eur. Phys. J.* C74.5 (2014), p. 2888. DOI: 10.1140/epjc/s10052-014-2888-1. arXiv: 1402.4430 [hep-ex] (cit. on p. 39).
- [78] Christian Elsässer. $\bar{b}b$ production angle plots. URL: https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html (visited on 2019) (cit. on p. 39).
- [79] B. Muratori and T. Pieloni. ‘Luminosity levelling techniques for the LHC’. In: *Proceedings, ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders (BB2013): CERN, Geneva, Switzerland, March 18-22 2013*. [177(2014)]. 2014, pp. 177–181. DOI: 10.5170/CERN-2014-004.177. arXiv: 1410.5646 [physics.acc-ph] (cit. on p. 40).
- [80] P R Barbosa-Marinho et al. *LHCb VELO (VErteX LOcator): Technical Design Report*. Technical Design Report LHCb. Geneva: CERN, 2001. URL: <http://cds.cern.ch/record/504321> (cit. on pp. 40–42).
- [81] S Amato et al. *LHCb magnet: Technical Design Report*. Technical Design Report LHCb. Geneva: CERN, 2000. URL: <http://cds.cern.ch/record/424338> (cit. on pp. 45, 46).
- [82] P R Barbosa-Marinho et al. *LHCb inner tracker: Technical Design Report*. Technical Design Report LHCb. revised version number 1 submitted on 2002-11-13 14:14:34. Geneva: CERN, 2002. URL: <https://cds.cern.ch/record/582793> (cit. on p. 47).
- [83] P R Barbosa-Marinho et al. *LHCb outer tracker: Technical Design Report*. Technical Design Report LHCb. Geneva: CERN, 2001. URL: <http://cds.cern.ch/record/519146> (cit. on pp. 47, 49, 50).

- [84] X.-F. Navick et al. ‘Fabrication of ultra-low radioactivity detector holders for Edelweiss-II’. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 520.1 (2004). Proceedings of the 10th International Workshop on Low Temperature Detectors, pp. 189–192. ISSN: 0168-9002. DOI: 10.1016/j.nima.2003.11.290. URL: <http://www.sciencedirect.com/science/article/pii/S0168900203031991> (cit. on p. 47).
- [85] Wouter D. Hulsbergen. ‘Decay chain fitting with a Kalman filter’. In: *Nucl. Instrum. Meth. A* 552 (2005), pp. 566–575. DOI: 10.1016/j.nima.2005.06.078. arXiv: physics/0503191 (cit. on pp. 51, 127).
- [86] John Erthal Gaiser. ‘Charmonium spectroscopy from radiative decays of the J/ψ and ψ' ’. SLAC-R-255, p. 178. PhD thesis. Stanford, California 94305: Stanford University, Aug. 1982 (cit. on pp. 54, 154).
- [87] Mark Joseph Oreglia. ‘A study of the reactions $\psi' \rightarrow \gamma\gamma\psi$ ’. SLAC-R-236, Appendix D. PhD thesis. Stanford, California 94305: Stanford University, Dec. 1980 (cit. on pp. 54, 154).
- [88] Tomasz Skwarnicki. ‘A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances’. DESY-F31-86-02, DESY-F-31-86-02. PhD thesis. Cracow, INP, Apr. 1986 (cit. on pp. 54, 154).
- [89] P.A. Cherenkov. ‘Visible luminescence of pure liquids under the influence of γ -radiation’. In: *Dokl. Akad. Nauk SSSR* 2.8 (1934), pp. 451–454. DOI: 10.3367/UFNr.0093.196710n.0385 (cit. on pp. 55, 56).
- [90] John David Jackson. *Classical electrodynamics; 2nd ed.* New York, NY: Wiley, 1975. URL: <https://cds.cern.ch/record/100964> (cit. on p. 56).
- [91] Harvey B. Newman and Thomas Ypsilantis. *History of Original Ideas and Basic Discoveries in Particle Physics*. Vol. 352. Nato Science Series B. Springer US, 1996. ISBN: 978-0-306-45217-8. DOI: 10.1007/978-1-4613-1147-8 (cit. on p. 56).
- [92] J. Seguinot and T. Ypsilantis. ‘Photoionization and Cherenkov Ring Imaging’. In: *Nucl. Instrum. Meth.* 142 (1977), p. 377. DOI: 10.1016/0029-554X(77)90671-1 (cit. on p. 57).
- [93] Stanislas Jean Teichner and Gilbert Andre Nicolaon. ‘Method of preparing inorganic aerogels’. U.S. pat. US3672833A. 27th June 1972 (cit. on p. 58).
- [94] Stephan Eisenhardt. *Production and tests of Hybrid Photon Detectors for the LHCb RICH detectors*. Tech. rep. LHCb-PROC-2007-030. CERN-LHCb-PROC-2007-030. Geneva: CERN, July 2008. URL: <https://cds.cern.ch/record/1432233> (cit. on p. 59).
- [95] Massimiliano Fiorini. ‘The upgrade of the LHCb RICH detectors’. In: *Nucl. Instrum. Methods Phys. Res., A* 952 (2020), 161688. 3 p. DOI: 10.1016/j.nima.2018.12.003. URL: <https://cds.cern.ch/record/2715467> (cit. on p. 59).

- [96] M. Adinolfi et al. ‘Performance of the LHCb RICH detector at the LHC’. In: *The European Physical Journal C* 73.5 (May 2013). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2431-9. URL: <http://dx.doi.org/10.1140/epjc/s10052-013-2431-9> (cit. on p. 61).
- [97] S Amato et al. *LHCb calorimeters: Technical Design Report*. Technical Design Report LHCb. Geneva: CERN, 2000. URL: <https://cds.cern.ch/record/494264> (cit. on pp. 62–64).
- [98] Eduardo Picatoste Olloqui. ‘LHCb Preshower(PS) and Scintillating Pad Detector (SPD): commissioning, calibration, and monitoring’. In: *J. Phys.: Conf. Ser.* 160 (2009), 012046. 8 p. DOI: 10.1088/1742-6596/160/1/012046. URL: <https://cds.cern.ch/record/1293075> (cit. on p. 62).
- [99] P R Barbosa-Marinho et al. *LHCb muon system: Technical Design Report*. Technical Design Report LHCb. Geneva: CERN, 2001. URL: <https://cds.cern.ch/record/504326> (cit. on pp. 65, 66).
- [100] LHCb collaboration. *The LHCb data flow*. URL: <https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html> (visited on 25/08/2020) (cit. on p. 68).
- [101] Malika Meddahi et al. ‘LHC Abort Gap Monitoring and Cleaning’. In: *Conf. Proc. C* 100523 (2010). Ed. by Akira Noda et al., MOPEC009 (cit. on p. 69).
- [102] LHCb collaboration. *Trigger Schemes*. URL: <http://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html> (visited on 13/08/2020) (cit. on pp. 70, 77).
- [103] R Antunes-Nobrega et al. *LHCb trigger system: Technical Design Report*. Technical Design Report LHCb. revised version number 1 submitted on 2003-09-24 12:12:22. Geneva: CERN, 2003. URL: <https://cds.cern.ch/record/630828> (cit. on p. 69).
- [104] LHCb collaboration. *Beam Conditions*. URL: <https://twiki.cern.ch/twiki/bin/view/Main/LHCb-Facts> (visited on 13/08/2020) (cit. on p. 70).
- [105] G. Barrand et al. ‘GAUDI - A software architecture and framework for building HEP data processing applications’. In: *Comput. Phys. Commun.* 140 (2001), pp. 45–55. DOI: 10.1016/S0010-4655(01)00254-5 (cit. on pp. 73, 79, 82).
- [106] Francesca Dordei. ‘LHCb detector and trigger performance in Run II’. In: *EPJ Web Conf.* 164 (2017). Ed. by L. Bravina, Y. Foka and S. Kabana, p. 01016. DOI: 10.1051/epjconf/201716401016 (cit. on p. 74).
- [107] M Clemencic et al. ‘The LHCb Simulation Application, Gauss: Design, Evolution and Experience’. In: *Journal of Physics: Conference Series* 331.3 (Dec. 2011), p. 032023. DOI: 10.1088/1742-6596/331/3/032023. URL: <https://doi.org/10.1088%2F1742-6596%2F331%2F3%2F032023> (cit. on p. 74).

- [108] Torbjörn Sjöstrand et al. ‘An introduction to PYTHIA 8.2’. In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on p. 74).
- [109] D.J. Lange. ‘The EvtGen particle decay simulation package’. In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4 (cit. on p. 74).
- [110] S. Agostinelli et al. ‘GEANT4—a simulation toolkit’. In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (cit. on p. 75).
- [111] R Antunes-Nobrega et al. *LHCb computing: Technical Design Report*. Technical Design Report LHCb. Submitted on 11 May 2005. Geneva: CERN, 2005. URL: <https://cds.cern.ch/record/835156> (cit. on p. 75).
- [112] WLCG collaboration. *The Worldwide LHC Computing Grid*. URL: <https://home.cern/science/computing/worldwide-lhc-computing-grid> (visited on 26/08/2020) (cit. on p. 76).
- [113] André Günther. ‘LHCb’s software trigger and how to find track online (fast)’. LHCb Collection. Feb. 2008. URL: https://indico.cern.ch/event/882442/contributions/3717963/attachments/1997811/3333608/guenther_highrr_seminartalk.pdf (cit. on p. 79).
- [114] S. Borghi. ‘Novel real-time alignment and calibration of the LHCb detector and its performance’. In: *Nucl. Instrum. Meth. A* 845 (2017). Ed. by G. Badurek et al., pp. 560–564. DOI: 10.1016/j.nima.2016.06.050 (cit. on pp. 80, 201).
- [115] Kurt Hornik. ‘Approximation capabilities of multilayer feedforward networks’. In: *Neural Networks* 4.2 (1991), pp. 251–257. ISSN: 0893-6080. DOI: 10.1016/0893-6080(91)90009-T. URL: <http://www.sciencedirect.com/science/article/pii/089360809190009T> (cit. on pp. 81, 83, 84, 145).
- [116] Andreas Hocker et al. *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide*. *TMVA - Toolkit for Multivariate Data Analysis*. Tech. rep. TMVA-v4 Users Guide: 135 pages, 19 figures, numerous code examples and references. Geneva: CERN, Mar. 2007. URL: <https://cds.cern.ch/record/1019880> (cit. on pp. 81, 143).
- [117] M. Feindt and U. Kerzel. ‘The NeuroBayes neural network package’. In: *Nucl. Instrum. Meth. A* 559 (2006), pp. 190–194. DOI: 10.1016/j.nima.2005.11.166 (cit. on p. 81).
- [118] Fabian Pedregosa et al. ‘Scikit-learn: Machine Learning in Python’. In: *J. Machine Learning Res.* 12 (2011), pp. 2825–2830. arXiv: 1201.0490 [cs.LG] (cit. on p. 81).
- [119] François Chollet et al. *Keras*. <https://github.com/fchollet/keras>. 2015 (cit. on pp. 81, 87).

- [120] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <https://www.tensorflow.org/> (cit. on p. 81).
- [121] Adam Paszke et al. ‘PyTorch: An Imperative Style, High-Performance Deep Learning Library’. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf> (cit. on p. 81).
- [122] V. V. Gligorov and Mike Williams. ‘Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree’. In: *JINST* 8 (2013), P02013. DOI: 10.1088/1748-0221/8/02/P02013. arXiv: 1210.6861 [physics.ins-det] (cit. on p. 82).
- [123] Anna Choromanska et al. ‘The Loss Surface of Multilayer Networks’. In: *CoRR* abs/1412.0233 (2014). arXiv: 1412.0233. URL: <http://arxiv.org/abs/1412.0233> (cit. on p. 83).
- [124] G. Bebis and M. Georgiopoulos. ‘Feed-forward neural networks’. In: *IEEE Potentials* 13.4 (1994), pp. 27–31. DOI: 10.1109/45.329294 (cit. on p. 83).
- [125] G. A. Cowan, D. C. Craik and M. D. Needham. ‘RapidSim: an application for the fast simulation of heavy-quark hadron decays’. In: *Comput. Phys. Commun.* 214 (2017), pp. 239–246. DOI: 10.1016/j.cpc.2017.01.029. arXiv: 1612.07489 [hep-ex] (cit. on p. 87).
- [126] Diederik P. Kingma and Jimmy Ba. ‘Adam: A Method for Stochastic Optimization’. In: (2017). arXiv: 1412.6980 [cs.LG] (cit. on p. 87).
- [127] Torbjörn Sjöstrand, Stephen Mrenna and Peter Skands. ‘A brief introduction to PYTHIA 8.1’. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph] (cit. on p. 89).
- [128] Andy Buckley et al. ‘Rivet user manual’. In: *Comput. Phys. Commun.* 184 (2013), pp. 2803–2819. DOI: 10.1016/j.cpc.2013.05.021. arXiv: 1003.0694 [hep-ph] (cit. on p. 89).
- [129] Matteo Cacciari, Gavin P. Salam and Gregory Soyez. ‘FastJet User Manual’. In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph] (cit. on p. 89).
- [130] Gavin P. Salam and Gregory Soyez. ‘A Practical Seedless Infrared-Safe Cone jet algorithm’. In: *JHEP* 05 (2007), p. 086. DOI: 10.1088/1126-6708/2007/05/086. arXiv: 0704.0292 [hep-ph] (cit. on p. 89).
- [131] S. Catani et al. ‘Longitudinally-invariant k_{\perp} -clustering algorithms for hadron-hadron collisions’. In: *Nuclear Physics B* 406.1 (1993), pp. 187–224. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(93\)90166-M](https://doi.org/10.1016/0550-3213(93)90166-M). URL: <https://www.sciencedirect.com/science/article/pii/055032139390166M> (cit. on p. 89).

- [132] Ryan Atkin. ‘Review of jet reconstruction algorithms’. In: *Journal of Physics: Conference Series* 645 (Oct. 2015), p. 012008. DOI: 10.1088/1742-6596/645/1/012008. URL: <https://doi.org/10.1088/1742-6596/645/1/012008> (cit. on p. 89).
- [133] Patrick T. Komiske, Eric M. Metodiev and Matthew D. Schwartz. ‘Deep learning in color: towards automated quark/gluon jet discrimination’. In: *JHEP* 01 (2017), p. 110. DOI: 10.1007/JHEP01(2017)110. arXiv: 1612.01551 [hep-ph] (cit. on p. 89).
- [134] Daniel Hay Guest et al. *lwtnn/lwtnn: Lightweight Trained Neural Network*. Version v2.12.1. July 2021. DOI: 10.5281/zenodo.5082190. URL: <https://doi.org/10.5281/zenodo.5082190> (cit. on p. 95).
- [135] Sean Benson, Xabier Cid Vidal and Albert Puig. ‘ $\gamma\gamma$ HLT lines implementation status’. Presented at HLT weekly meeting (unpublished). 2019. URL: <https://indico.cern.ch/event/692131/> (cit. on p. 96).
- [136] Sean Benson and Rosen Matev. ‘Reoptimise $B_s^0 \rightarrow \gamma\gamma$ lines’. Presented at HLT weekly meeting (unpublished). URL: <https://its.cern.ch/jira/browse/LBHLT-386> (cit. on p. 96).
- [137] R Aaij et al. ‘First observation of CP violation in the decays of B_s^0 mesons’. In: *Phys. Rev. Lett.* 110.22 (2013), p. 221601. DOI: 10.1103/PhysRevLett.110.221601. arXiv: 1304.6173 [hep-ex] (cit. on p. 99).
- [138] L Cadamuro et al. ‘Characterization of the Hamamatsu R11265-103-M64 multi-anode photomultiplier tube’. In: *Journal of Instrumentation* 9.06 (June 2014), P06021–P06021. DOI: 10.1088/1748-0221/9/06/p06021. URL: <https://doi.org/10.1088/1748-0221/9/06/p06021> (cit. on pp. 100, 102).
- [139] Sylvie Blin, Pierre Barrillon and Christophe de La Taille. ‘MAROC, a generic photomultiplier readout chip’. In: *IEEE Nucl. Sci. Symp. Conf. Rec.* 2010 (2010). Ed. by Ron Keyser, pp. 1690–1693. DOI: 10.1109/NSSMIC.2010.5874062 (cit. on p. 101).
- [140] S. Gambetta. ‘First results from quality assurance testing of MaPMTs for the LHCb RICH upgrade’. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 876 (2017). The 9th international workshop on Ring Imaging Cherenkov Detectors (RICH2016), pp. 206–208. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2017.02.079>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900217302917> (cit. on pp. 101, 103, 110, 111, 186).
- [141] Scott Hauck et al. *The Chimaera Reconfigurable Functional Unit*. 1997 (cit. on p. 101).
- [142] Philip Garcia et al. ‘An Overview of Reconfigurable Hardware in Embedded Systems’. In: *EURASIP Journal on Embedded Systems* 2006.1 (Sept. 2006), p. 056320. ISSN: 1687-3963. DOI: 10.1155/ES/2006/56320. URL: <https://doi.org/10.1155/ES/2006/56320> (cit. on p. 101).

- [143] Konstantin Gizdov. ‘First results from production testing of 64-channel MaPMT R13742 (1 in) and R13743 (2 in) for the LHCb RICH Upgrade’. July 2018. URL: <https://cds.cern.ch/record/2632422> (cit. on p. 105).
- [144] J. Lewis and M. Fowler. *Microservices [WWW Document]*. 2012. URL: <https://martinfowler.com/articles/microservices.html> (visited on 10/09/2018) (cit. on p. 107).
- [145] ISO/IEC. *ISO International Standard ISO/IEC 14882:2014(E) — Programming Language C++*. Geneva, Switzerland: International Organization for Standardization (ISO), 2014. URL: <https://isocpp.org/std/status> (cit. on p. 110).
- [146] Rick Bitter, Taqi Mohiuddin and Matt Nawrocki. *LabVIEW: Advanced programming techniques*. Crc Press, 2006 (cit. on p. 110).
- [147] Rene Brun et al. *root-project/root: v6.18/02*. Version v6-18-02. Aug. 2019. DOI: 10.5281/zenodo.3895860. URL: <https://doi.org/10.5281/zenodo.3895860> (cit. on p. 110).
- [148] Tim Bray et al. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. 2008. URL: <https://www.w3.org/TR/xml/> (visited on 13/08/2021) (cit. on p. 110).
- [149] Konstantin Gizdov and Stefano Gallorini. *PDQA Automation*. 2018. URL: <https://gitlab.cern.ch/kgizdov/pdqa-automation> (visited on 11/09/2018) (cit. on p. 110).
- [150] I. Beiyaev et al. ‘Handling of the generation of primary events in Gauss, the LHCb simulation framework’. In: *IEEE Nuclear Science Symposium Medical Imaging Conference*. 2010, pp. 1155–1161. DOI: 10.1109/NSSMIC.2010.5873949 (cit. on p. 114).
- [151] Elisabetta Barberio, Bob van Eijk and Zbigniew Was. ‘PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays’. In: *Comput. Phys. Commun.* 66 (1991), pp. 115–128. DOI: 10.1016/0010-4655(91)90012-A (cit. on p. 114).
- [152] M Needham. *Momentum scale calibration using resonances*. Tech. rep. LHCb-2008-037. CERN-LHCb-2008-037. LPHE-2008-08. Geneva: CERN, July 2008. URL: <https://cds.cern.ch/record/1115072> (cit. on p. 117).
- [153] Matthew Needham. ‘Run 2 momentum scale calibration’. Presented at LHCb A&S Week organised by the LHCb Collaboration, CERN. Apr. 2017. URL: <https://indico.cern.ch/event/603961/contributions/2531597/attachments/1437967/2212042/momentumscale.pdf> (visited on 15/12/2020) (cit. on p. 117).
- [154] Matthew Needham. ‘Momentum smearing tools’. Presented at LHCb Tracking and Alignment Meeting organised by the LHCb Collaboration, CERN. Apr. 2013. URL: <https://indico.cern.ch/event/223871/contributions/465337/attachments/367679/511778/smearingtools.pdf> (visited on 19/02/2022) (cit. on p. 117).

- [155] R Aaij et al. ‘The LHCb Trigger and its Performance in 2011’. In: *JINST* 8 (2013), P04022. DOI: 10.1088/1748-0221/8/04/P04022. arXiv: 1211.3055 [hep-ex] (cit. on p. 119).
- [156] Roel Aaij et al. *Optimization of the muon reconstruction algorithms for LHCb Run 2*. Tech. rep. LHCb-PUB-2017-007. CERN-LHCb-PUB-2017-007. Geneva: CERN, Feb. 2017. URL: <https://cds.cern.ch/record/2253050> (cit. on p. 121).
- [157] Francesco Dettori, Diego Martinez Santos and Jessica Prisciandaro. *Low- p_T dimuon triggers at LHCb in Run 2*. Tech. rep. LHCb-PUB-2017-023. CERN-LHCb-PUB-2017-023. Geneva: CERN, Dec. 2017. URL: <https://cds.cern.ch/record/2297352> (cit. on p. 121).
- [158] Olli Lupton et al. *Calibration samples for particle identification at LHCb in Run 2*. Tech. rep. LHCb-PUB-2016-005. CERN-LHCb-PUB-2016-005. Geneva: CERN, Mar. 2016. URL: <https://cds.cern.ch/record/2134057> (cit. on p. 128).
- [159] G Lanfranchi et al. *The Muon Identification Procedure of the LHCb Experiment for the First Data*. Tech. rep. LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013. Geneva: CERN, Aug. 2009. URL: <https://cds.cern.ch/record/1202759> (cit. on p. 128).
- [160] Lucio Anderlini et al. *The PIDCalib package*. Tech. rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021. Geneva: CERN, July 2016. URL: <https://cds.cern.ch/record/2202412> (cit. on pp. 129, 132–134, 136).
- [161] Roel Aaij et al. ‘Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2’. In: *EPJ Tech. Instrum.* 6.1 (2019), p. 1. DOI: 10.1140/epjti/s40485-019-0050-z. arXiv: 1803.00824 [hep-ex] (cit. on p. 131).
- [162] A. Rogozhnikov. ‘Reweighting with Boosted Decision Trees’. In: *J. Phys. Conf. Ser.* 762.1 (2016). Ed. by Luis Salinas and Claudio Torres, p. 012036. DOI: 10.1088/1742-6596/762/1/012036. arXiv: 1608.05806 [physics.data-an] (cit. on pp. 134, 136, 138–140).
- [163] Yoav Freund and Robert E. Schapire. ‘Experiments with a New Boosting Algorithm’. In: *IN PROCEEDINGS OF THE THIRTEENTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING*. Morgan Kaufmann, 1996, pp. 148–156 (cit. on p. 138).
- [164] Yoav Freund and Robert E Schapire. ‘A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting’. In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139. ISSN: 0022-0000. DOI: <https://doi.org/10.1006/jcss.1997.1504>. URL: <https://www.sciencedirect.com/science/article/pii/S002200009791504X> (cit. on p. 144).

- [165] Dong C. Liu and Jorge Nocedal. ‘On the limited memory BFGS method for large scale optimization’. In: *Mathematical Programming* 45.1 (Aug. 1989), pp. 503–528. ISSN: 1436-4646. DOI: 10.1007/BF01589116. URL: <https://doi.org/10.1007/BF01589116> (cit. on p. 144).
- [166] R. Fletcher and Trond Steihaug. ‘Practical methods of optimization volume 1: Unconstrained optimization, Wiley, New York, 1980’. In: *Networks* 12.4 (1982), pp. 508–509. DOI: <https://doi.org/10.1002/net.3230120417>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/net.3230120417>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/net.3230120417> (cit. on p. 144).
- [167] C. G. BROYDEN. ‘The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations’. In: *IMA Journal of Applied Mathematics* 6.1 (Mar. 1970), pp. 76–90. ISSN: 0272-4960. DOI: 10.1093/imamat/6.1.76. eprint: <https://academic.oup.com/imamat/article-pdf/6/1/76/2233756/6-1-76.pdf>. URL: <https://doi.org/10.1093/imamat/6.1.76> (cit. on p. 145).
- [168] R. Fletcher. ‘A new approach to variable metric algorithms’. In: *The Computer Journal* 13.3 (Jan. 1970), pp. 317–322. ISSN: 0010-4620. DOI: 10.1093/comjnl/13.3.317. eprint: <https://academic.oup.com/comjnl/article-pdf/13/3/317/988678/130317.pdf>. URL: <https://doi.org/10.1093/comjnl/13.3.317> (cit. on p. 145).
- [169] Donald Goldfarb. ‘A Family of Variable-Metric Methods Derived by Variational Means’. en. In: *Math. Comp.* 24.109 (1970), pp. 23–26. ISSN: 0025-5718, 1088-6842. DOI: 10.1090/S0025-5718-1970-0258249-6 (cit. on p. 145).
- [170] D. F. Shanno. ‘Conditioning of Quasi-Newton Methods for Function Minimization’. en. In: *Math. Comp.* 24.111 (1970), pp. 647–656. ISSN: 0025-5718, 1088-6842. DOI: 10.1090/S0025-5718-1970-0274029-X (cit. on p. 145).
- [171] Frank Burden and Dave Winkler. ‘Bayesian Regularization of Neural Networks’. In: *Methods in molecular biology (Clifton, N.J.)* 458 (Jan. 2009), pp. 23–42. DOI: 10.1007/978-1-60327-101-1_3 (cit. on p. 145).
- [172] Particle Data Group et al. ‘Review of Particle Physics’. In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>. URL: <https://doi.org/10.1093/ptep/ptaa104> (cit. on pp. 146, 154, 157, 158, 163, 182).
- [173] Richard Simard and Pierre L’Ecuyer. ‘Computing the Two-Sided Kolmogorov-Smirnov Distribution’. In: *Journal of Statistical Software, Articles* 39.11 (2011), pp. 1–18. ISSN: 1548-7660. DOI: 10.18637/jss.v039.i11. URL: <https://www.jstatsoft.org/v039/i11> (cit. on pp. 148, 150).

- [174] Muriel Pivk and Francois R. Le Diberder. ‘SPlot: A Statistical tool to unfold data distributions’. In: *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083 (cit. on pp. 154, 158, 210).
- [175] Sean Benson et al. *Measurement of the CP-violating phase ϕ_s using $B_s^0 \rightarrow J/\psi K^+ K^-$ decays in Run 2*. LHCb Internal Note (unpublished). Apr. 2017. URL: <https://cds.cern.ch/record/2259530> (cit. on pp. 155, 157–159, 163, 166, 215, 216).
- [176] Paul Adrien Maurice Dirac. *The Principles of Quantum Mechanics*. en. Clarendon Press, 1981. ISBN: 978-0-19-852011-5 (cit. on p. 161).
- [177] Marcin Kucharczyk, Piotr Morawski and Mariusz Witek. *Primary Vertex Reconstruction at LHCb*. Tech. rep. Geneva: CERN, Sept. 2014. URL: <https://cds.cern.ch/record/1756296> (cit. on p. 161).
- [178] Roel Aaij et al. ‘Measurements of the B^+ , B^0 , B_s^0 meson and Λ_b^0 baryon lifetimes’. In: *JHEP* 04 (2014), p. 114. DOI: 10.1007/JHEP04(2014)114. arXiv: 1402.2554 [hep-ex] (cit. on p. 162).
- [179] Y. Amhis et al. ‘Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016’. In: *Eur. Phys. J. C* 77.12 (2017), p. 895. DOI: 10.1140/epjc/s10052-017-5058-4. arXiv: 1612.07233 [hep-ex] (cit. on p. 163).
- [180] R Aaij et al. ‘Tagged time-dependent analysis of $B_s^0 \rightarrow J/\psi K^+ K^-$ and $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ decays with 1.0 fb $^{-1}$ ’. In: (Apr. 2013). Linked to LHCb-PAPER-2013-002. URL: <https://cds.cern.ch/record/1459106> (cit. on p. 165).
- [181] R Aaij et al. ‘Measurement of the polarization amplitudes in $B^0 \rightarrow J/\psi K^*(892)^0$ decays’. In: *Phys. Rev. D* 88 (2013), p. 052002. DOI: 10.1103/PhysRevD.88.052002. arXiv: 1307.2782 [hep-ex] (cit. on p. 166).
- [182] R. Aaij et al. ‘Opposite-side flavour tagging of B mesons at the LHCb experiment’. In: *The European Physical Journal C* 72.6 (June 2012). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-2022-1. URL: <http://dx.doi.org/10.1140/epjc/s10052-012-2022-1> (cit. on p. 167).
- [183] The LHCb collaboration. ‘Bflavour tagging using charm decays at the LHCb experiment’. In: *Journal of Instrumentation* 10.10 (Oct. 2015), P10005–P10005. ISSN: 1748-0221. DOI: 10.1088/1748-0221/10/10/P10005. URL: <http://dx.doi.org/10.1088/1748-0221/10/10/P10005> (cit. on p. 167).
- [184] Julian Wishahi. *TikZ Flavour Tagging Scheme*. URL: <https://gitlab.cern.ch/lhcb-ft/TikZPics/> (visited on 28/06/2021) (cit. on p. 168).
- [185] Davide Fazzini. ‘Flavour Tagging in the LHCb experiment’. In: *PoS LHCP2018* (2018), p. 230. DOI: 10.22323/1.321.0230 (cit. on p. 168).

- [186] R. Aaij et al. ‘Observation of the Decay $B_c^+ \rightarrow B_s^0 \pi^+$ ’. In: *Physical Review Letters* 111.18 (Nov. 2013). ISSN: 1079-7114. DOI: 10.1103/physrevlett.111.181801. URL: <http://dx.doi.org/10.1103/PhysRevLett.111.181801> (cit. on p. 174).
- [187] V. V. Kiselev. ‘Decays of the B_c meson’. In: (Aug. 2003). arXiv: hep-ph/0308214 (cit. on p. 175).
- [188] J. Charles et al. ‘Current status of the Standard Model CKM fit and constraints on $\Delta F = 2$ New Physics’. In: *Phys. Rev. D* 91.7 (2015), p. 073007. DOI: 10.1103/PhysRevD.91.073007. arXiv: 1501.05013 [hep-ph] (cit. on pp. 182, 186, 187).
- [189] M. Bona et al. ‘The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of Delta m(s) and BR($B \rightarrow \tau \nu(\tau)$)’. In: *JHEP* 10 (2006), p. 081. DOI: 10.1088/1126-6708/2006/10/081. arXiv: hep-ph/0606167 (cit. on pp. 182, 186).
- [190] Marina Artuso, Guennadi Borissov and Alexander Lenz. ‘CP violation in the B_s^0 system’. In: *Rev. Mod. Phys.* 88.4 (2016). [Addendum: *Rev.Mod.Phys.* 91, 049901 (2019)], p. 045002. DOI: 10.1103/RevModPhys.88.045002. arXiv: 1511.09466 [hep-ph] (cit. on pp. 182, 186, 187).
- [191] M. Kirk, A. Lenz and T. Rauh. ‘Dimension-six matrix elements for meson mixing and lifetimes from sum rules’. In: *JHEP* 12 (2017). [Erratum: *JHEP* 06, 162 (2020)], p. 068. DOI: 10.1007/JHEP12(2017)068. arXiv: 1711.02100 [hep-ph] (cit. on pp. 182, 186).
- [192] Roel Aaij et al. ‘Measurement of the CP -violating phase ϕ_s from $B_s^0 \rightarrow J/\psi \pi^+ \pi^-$ decays in 13 TeV pp collisions’. In: *Phys. Lett. B* 797 (Mar. 2019). Submitted to *JHEP*, 134789. 24 p. DOI: 10.1016/j.physletb.2019.07.036. arXiv: 1903.05530. URL: <https://cds.cern.ch/record/2666948> (cit. on pp. 186, 187).
- [193] Roel Aaij et al. ‘Measurement of the CP -violating phase ϕ_s in $\bar{B}_s^0 \rightarrow J/\psi \pi^+ \pi^-$ decays’. In: *Phys. Lett. B* 736 (2014), pp. 186–195. DOI: 10.1016/j.physletb.2014.06.079. arXiv: 1405.4140 [hep-ex] (cit. on pp. 186, 215).
- [194] Roel Aaij et al. ‘Resonances and CP violation in B_s^0 and $\bar{B}_s^0 \rightarrow J/\psi K^+ K^-$ decays in the mass region above the $\phi(1020)$ ’. In: *JHEP* 08 (2017), p. 037. DOI: 10.1007/JHEP08(2017)037. arXiv: 1704.08217 [hep-ex] (cit. on p. 186).
- [195] Roel Aaij et al. ‘First study of the CP -violating phase and decay-width difference in $B_s^0 \rightarrow \psi(2S) \phi$ decays’. In: *Phys. Lett. B* 762 (2016), pp. 253–262. DOI: 10.1016/j.physletb.2016.09.028. arXiv: 1608.04855 [hep-ex] (cit. on p. 186).

- [196] Roel Aaij et al. ‘Measurement of the CP -violating phase ϕ_s in $\bar{B}_s^0 \rightarrow D_s^+ D_s^-$ decays’. In: *Phys. Rev. Lett.* 113.21 (2014), p. 211801. DOI: 10.1103/PhysRevLett.113.211801. arXiv: 1409.4619 [hep-ex] (cit. on p. 186).
- [197] Andrzej J. Buras. ‘Weak Hamiltonian, CP violation and rare decays’. In: *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*. June 1998, pp. 281–539. arXiv: hep-ph/9806471 (cit. on p. 191).
- [198] Kenneth G. Wilson. ‘Non-Lagrangian Models of Current Algebra’. In: *Phys. Rev.* 179 (5 Mar. 1969), pp. 1499–1512. DOI: 10.1103/PhysRev.179.1499. URL: <https://link.aps.org/doi/10.1103/PhysRev.179.1499> (cit. on p. 191).
- [199] Andrzej J. Buras. ‘Operator Product Expansion, Renormalization Group and Weak Decays’. In: *Quantum Field Theory*. Ed. by Peter Breitenlohner and Dieter Maison. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 65–85. ISBN: 978-3-540-44482-4 (cit. on p. 191).
- [200] I Bediaga et al. *Framework TDR for the LHCb Upgrade: Technical Design Report*. Tech. rep. Apr. 2012. URL: <https://cds.cern.ch/record/1443882> (cit. on p. 195).
- [201] M. Jacob and G.C. Wick. ‘On the general theory of collisions for particles with spin’. In: *Annals of Physics* 7.4 (1959), pp. 404–428. ISSN: 0003-4916. DOI: [https://doi.org/10.1016/0003-4916\(59\)90051-X](https://doi.org/10.1016/0003-4916(59)90051-X). URL: <http://www.sciencedirect.com/science/article/pii/000349165990051X> (cit. on p. 196).
- [202] Jurgen Engelfried. ‘Cherenkov Light Imaging: Fundamentals and recent Developments’. In: *Nucl. Instrum. Meth. A* 639 (2011). Ed. by R. Forty et al., pp. 1–6. DOI: 10.1016/j.nima.2010.10.023. arXiv: 1009.0052 [physics.ins-det] (cit. on p. 200).
- [203] Giulio Dujany and Barbara Storaci. ‘Real-time alignment and calibration of the LHCb Detector in Run II’. In: *J. Phys. Conf. Ser.* 664.8 (2015), p. 082010. DOI: 10.1088/1742-6596/664/8/082010 (cit. on p. 201).
- [204] T. Skwarnicki. *Update on TISTOS software*. URL: <https://indico.cern.ch/event/82230/contributions/1258763/attachments/1067057/1521782/TisTos-Nov-1-10.pdf> (visited on 25/08/2020) (cit. on p. 201).
- [205] ‘IEEE Standard for Floating-Point Arithmetic’. In: *IEEE Std 754-2019 (Revision of IEEE 754-2008)* (2019), pp. 1–84. DOI: 10.1109/IEEESTD.2019.8766229 (cit. on p. 204).
- [206] R. A. Fisher. ‘On the mathematical foundations of theoretical statistics’. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222 (594-604 Jan. 1922), pp. 309–368. DOI: 10.1098/rsta.1922.0009 (cit. on p. 204).

- [207] Alexander Ly et al. ‘A Tutorial on Fisher Information’. In: *J. Math. Psychol.* 80 (2017), pp. 40–55. ISSN: 0022-2496. DOI: <https://doi.org/10.1016/j.jmp.2017.05.006>. arXiv: 1705.01064 [math.ST] (cit. on p. 205).
- [208] Maarten Jansen and Gerda Claeskens. ‘Cramér–Rao Inequality’. In: *International Encyclopedia of Statistical Science*. Ed. by Miodrag Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 322–323. ISBN: 978-3-642-04898-2. DOI: 10.1007/978-3-642-04898-2%008197 (cit. on p. 205).
- [209] Mark Smith. ‘Summary of the new VELO error parametrisation’. Presented at LHCb T&A Meeting organised by the LHCb Collaboration, CERN. Mar. 2017. URL: https://indico.cern.ch/event/600190/contributions/2529764/attachments/1436679/2209543/TA_300317.pdf (visited on 16/01/2021) (cit. on p. 213).
- [210] Sevdal Esen. ‘Track χ^2 in 2016 and 2017 In the context of the $B_s^0 \rightarrow J/\psi\phi$ analysis’. Presented at LHCb Joint T&A and Simulation Meeting organised by the LHCb Collaboration, CERN. July 2018. URL: https://indico.cern.ch/event/693514/contributions/3085234/attachments/1692271/2722993/JPsiPhi_Tracking_2016vs2017.pdf (visited on 16/01/2021) (cit. on p. 213).
- [211] Roel Aaij et al. ‘Measurement of the CP -violating phase ϕ_s from $B_s^0 \rightarrow J/\psi\pi^+\pi^-$ decays in 13 TeV pp collisions’. In: *Phys. Lett. B* 797 (2019), p. 134789. DOI: 10.1016/j.physletb.2019.07.036. arXiv: 1903.05530 [hep-ex] (cit. on p. 215).