

Measurement of the CKM angle γ and B meson lifetimes at the LHCb detector

Vladimir Vava Gligorov
St. John's College Oxford

Thesis submitted in fulfillment of the requirements for the degree of Doctor
of Philosophy at the University of Oxford

CERN-THESIS-2008-044
28/05/2008



Michaelmas Term, 2007

Abstract

LHCb is the dedicated B physics experiment at the Large Hadron Collider (LHC) at CERN. It will make precision measurements of CP violating effects in the B_d and B_s systems, as well as making precision measurements of the lifetimes of all flavours of B hadrons. In this thesis, two possible measurements of the CKM angle γ are evaluated: from the decay mode $B_d^0 \rightarrow D^- \pi^+$, and from the combined analysis of the decay modes $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$ under the conditions of U-spin symmetry. Also, a Monte Carlo independent method of measuring the lifetimes of B hadrons is described.

The reconstruction of the decay mode $B_d^0 \rightarrow D^- \pi^+$ is studied using the LHCb simulation software, and a general method for categorising background at LHCb is developed. The decay mode $B_d^0 \rightarrow D^- \pi^+$ is found to have a yearly yield of 1340k events, and a signal to background ratio of ~ 5 . It is shown that the analysis of time dependent decay rate asymmetries in $B_d^0 \rightarrow D^- \pi^+$ can result in a precision of $\sim 8^\circ$ on γ after one year of data taking. The combined extraction of γ from the decay modes $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$ is seen to allow an unambiguous extraction of γ , with a precision of a few degrees after one year of data taking.

The LHCb trigger uses the lifetime of reconstructed particles to discriminate between signal and background, and therefore introduces a bias in the lifetime distributions of B hadrons. A method for measuring this bias, and hence the lifetimes of B hadrons, without recourse to Monte Carlo information, is developed. This method can be applied to any decay mode reconstructed at LHCb which is expected to suffer from a lifetime bias, in particular hadronic decay channels. It is shown that LHCb can expect to match the current world average precision on the B_s^0 lifetime with a few weeks of data taking in the decay mode $B_s^0 \rightarrow D_s^- \pi^+$, while the world average precision on the B_d^0 lifetime can be matched with a few months of data taking in the decay mode $B_d^0 \rightarrow D^- \pi^+$.

For Milica and Peter.

Sanjam, sanjam da si tu
ko da letim, na zlatnom ćilimu.
Sanjam, bog nas prokleo,
sve nam dao, a sreću nam je uzeo.

—Piloti, Zaboravljeni

I don't know how to do statistics, but it doesn't matter
because I didn't have data.

—Scott Adams, Dilbert

Acknowledgements

It is traditional to begin with a word of thanks to the long suffering martyrs who tutored the author, but I hope they will forgive me if I begin by thanking my parents, Ljubica Strncevic and Vladimir Gligorov. They funded this thesis and the rest of my education; mum, dad, I hope it was worth it! That having been said, I am indebted to my thesis supervisors, Guy Wilkinson and Jonas Rademacker, for their support and for teaching me all the practically useful things I know about physics. Their meticulous proofreading has improved this thesis immeasurably; the remaining errors are all mine.

Various aspects of this thesis, especially concerning error estimation, owe much to discussions with Jim Libby. The CDF group at Oxford, and in particular Farrukh Azfar, provided valuable advice about the studies of B meson lifetimes. The design of the background categorization tool in Chapter 3 benefitted from discussions with Patrick Koppenburg, while Patrick Spradlin gave me the idea of using a genetic algorithm to optimize the event selection in the same chapter. Stuart Paterson taught me how to use the GRID and submit jobs efficiently (DIRAC's the word); without him, the studies in this thesis would likely have taken much longer to complete. I would also like to thank Olivier Callot for discussions about the LHCb trigger which informed the online-offline track association in Chapter 6; Gloria Corti for answering all my numerous questions about LHCb software without once letting on how stupid they were; Charlotte Newby for helping me in my first year and showing me how to use ROOT; Tomas Schietinger for writing the unofficial LHCb ROOT style file used to create most of the plots in this thesis; and David Bailey for writing the latex style file used to write the thesis itself.

None of this work would have been possible without the support of the secretarial staff at Oxford and CERN. I would like to thank Sue Cannon, Sue Geddes, Nathalie Grub, Monique Grygiel, Kim Marsh, Beverly Roger, and Mary-Elizabeth Shewry for their professionalism and dilligence which shielded me from the worst of the paperwork. I would also like to thank Neville Harnew for his support during my studies, and in particular for giving me all the financial support he could.

Finally, a special thank you to Andy Smith, who taught me how to program in Python and made my stays at CERN much less boring than they might otherwise have been. You made the acknowledgements mate, how crazy is that?

Contents

Introduction	1
1 CP violation in heavy flavour physics	3
1.1 Introduction	3
1.2 Symmetries in particle physics	4
1.2.1 C, P, T operators	4
1.2.2 Observations of symmetry violation	5
1.3 Model independent description of the B^0 system	5
1.3.1 $B_d^0 - \bar{B}_d^0$ mixing	6
1.3.2 B_d^0 decays	7
1.3.3 CP violation in B_d^0 decays	8
1.3.4 CP violation in B_s^0 decays	10
1.4 CP violation in the Standard Model	10
1.4.1 The CKM Matrix	11
1.4.2 Current constraints on the CKM matrix	13
1.4.3 Constraint on γ	15
1.5 Measuring the angle γ from $B_d^0 \rightarrow D^+\pi^-$ decay rate asymmetries	15
1.5.1 Value of $ \eta $	17
1.5.2 U-spin symmetry	17
1.6 Factorization	18
1.7 Heavy Quark Expansion (HQE)	19
1.7.1 Operator product expansion	19
1.7.2 Current status of lifetime measurements	20
2 The LHCb Detector	22
2.1 LHCb at the Large Hadron Collider	22
2.2 LHCb design	23
2.2.1 $b\bar{b}$ production at the LHC	23
2.2.2 Design luminosity of LHCb	24
2.3 Vertex detection	24
2.3.1 Velo sensor layout	25
2.3.2 Integration with the LHC vacuum	27
2.3.3 VELO sensor geometry	27
2.3.4 Sensor design and readout	28
2.3.5 Vertex resolutions	30
2.3.6 Pile up veto	30

2.4	Tracking system	30
2.4.1	The LHCb magnet	32
2.4.2	The trigger tracker (TT)	32
2.4.3	The inner tracker (IT)	33
2.4.4	The outer tracker (OT)	34
2.4.5	Tracker occupancy	34
2.4.6	Reconstruction and tracking performance	35
2.5	Hadron identification	37
2.5.1	Measuring particle masses	38
2.5.2	Radiator configuration	39
2.5.3	Hybrid photon detectors (HPD)	39
2.5.4	Mirrors	41
2.5.5	Precision on the Cherenkov angle	42
2.5.6	Pattern recognition	42
2.5.7	RICH performance	43
2.5.8	Electron and muon identification	43
2.6	Flavour tagging	45
2.6.1	Same side tagging	45
2.6.2	Opposite side tagging	46
2.6.3	Combining tagging decisions and performance summary	47
2.7	Trigger system	48
2.7.1	L0 hardware trigger	48
2.7.2	L1 trigger	49
2.7.3	High level trigger	51
2.7.4	The new trigger	52
2.8	LHCb software	52
2.8.1	Event simulation and reconstruction	53
2.8.2	GRID computing	54
3	Reconstruction of the Channel $B_d^0 \rightarrow D^+\pi^-$	56
3.1	Introduction	56
3.2	Datasets and preselection	57
3.2.1	Signal Dataset	57
3.2.2	Background Dataset and Preselection Cuts	57
3.3	Background Classification	62
3.3.1	Background Categories	62
3.3.2	Background Categories After the Preselection	65
3.4	Optimizing the Event Selection	65
3.4.1	Genetic Algorithm	67
3.4.2	The Optimized Cut Values	68
3.4.3	Nature of the surviving background	72
3.4.4	Reconstruction efficiencies	75
3.4.5	Mass and Time Resolutions	76
3.4.6	Lifetime acceptance function	77
3.5	Conclusion	78

4	Extraction of the angle γ from the channel $B_d^0 \rightarrow D^- \pi^+$	79
4.1	Overview	79
4.2	Tagging performance	79
4.3	Fitting method	80
4.4	Including detector effects	81
4.4.1	Time resolution	82
4.4.2	Acceptance function	82
4.4.3	Mistag	82
4.4.4	Background	82
4.4.5	The full asymmetry expressions	83
4.4.6	Calculating the error on γ	85
4.5	Fit results	86
4.5.1	Competing measurements of γ	87
4.5.2	Precision on γ and ω	87
4.6	Ambiguities in the extracted value of γ	89
4.7	Statistical and systematic errors	90
4.7.1	Statistical error	90
4.7.2	Systematic errors	92
4.7.3	Analyzing $B_d^0 \rightarrow D^- \pi^+$ and $B_d^0 \rightarrow D^{*-} \pi^+$	95
4.8	Conclusion	96
5	Extraction of the angle γ through the U-spin analysis of the channels $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$	97
5.1	Motivation	97
5.2	Overview of U-spin method	97
5.2.1	U-spin formalism	98
5.2.2	Time dependent rate asymmetries	98
5.2.3	Reducing the ambiguity through factorization	100
5.2.3	Exploiting $A_{\Delta\Gamma}$	100
5.2.4	Combined extraction of γ	101
5.2.6	Summary of combined extraction	103
5.3	Fitting to the CP-observables	103
5.4	Sensitivity to γ	104
5.4.1	Error estimation	105
5.4.2	$\gamma = 60^\circ$ and $\delta_{s,d} = 60^\circ$	106
5.4.3	$\gamma = 60^\circ$ and $\delta_{s,d} = 10^\circ$	107
5.4.4	Performance summary	109
5.5	Conclusion	110
6	A Monte Carlo independent measurement of the B_d^0 lifetime	112
6.1	Introduction	112
6.1.1	Motivation for measuring B lifetimes at LHCb	112
6.1.2	The channel $B_d^0 \rightarrow D^-(\pi^-\pi^-K^+)\pi^+$	114
6.2	The LHCb lifetime dependent trigger	114
6.2.1	Trigger architecture	115

6.2.2	Sub-detectors	115
6.2.3	The L1 trigger	115
6.3	The likelihood for signal events	116
6.3.1	Notation	117
6.3.2	Overview	117
6.3.3	Including measurement errors	119
6.3.4	Online track finding efficiency	123
6.4	The likelihood including background events	124
6.4.1	The problem	124
6.4.2	The full likelihood	125
6.4.3	The background model	127
6.5	The Fisher discriminant	127
6.5.1	Discriminating between signal and background events	128
6.5.2	Representing the acceptance function as a vector	128
6.5.3	Example Fisher distribution	129
6.6	Accessing the online data	129
6.6.1	Associating online and offline tracks	129
6.6.2	Trigger emulation	131
6.7	Fitting to signal events	131
6.7.1	Full Monte Carlo fit	132
6.7.2	A toy Monte Carlo for $B_d^0 \rightarrow D^- \pi^+$	132
6.7.3	Toy Monte Carlo fit	135
6.8	Including background	136
6.8.1	Full Monte Carlo fit	137
6.8.2	Toy Monte Carlo pulls	137
6.9	Expected statistical precision	141
7	Conclusion	142
A	Current direct constraints on CKM matrix elements	144
B	Background category decision tree	146
C	Additional U-spin analysis scenarios	148
C.1	$\gamma = 60^\circ$ and $\delta_{s,d} = 30^\circ$	148
C.2	$\gamma = 60^\circ$ and $\delta_{s,d} = 85^\circ$	148
C.3	$\gamma = 40^\circ$ and $\delta_{s,d} = 40^\circ$	151
C.4	$\gamma = 80^\circ$ and $\delta_{s,d} = 40^\circ$	152
C.5	$\gamma = 60^\circ$ and $\delta_{s,d} = 140^\circ$	152
D	Derivation of the full lifetime likelihood	155
E	The Fisher discriminant for lifetime fitting	158
E.1	Discriminating between signal and background events	158
E.2	Calculating $ (\overline{m}_s - \overline{m}_b)\rangle$	160
E.3	Calculating S_w	160
	Bibliography	161

List of Figures

1.1	The $B_d^0 \rightarrow \overline{B}_d^0$ mixing diagram.	9
1.2	The CKM Unitarity Triangle, taken from [ea06c]. $\phi_{1,2,3}$ are equivalent to the angles α, β, γ , in the notation used by BELLE. The bottom of the triangle has unit length by convention.	13
1.3	Current constraints on the Unitarity Triangle. Diagram taken from [ea05].	14
1.4	The one-loop Feynman diagram for the decay $B_d^0 \rightarrow \pi\pi$	16
1.5	The tree-level Feynman diagram for the decay $B_d^0 \rightarrow \pi\pi$	16
1.6	The Feynman graph for the decay $B_d^0 \rightarrow D^-\pi^+$	16
1.7	The Feynman graph for the decay $\overline{B}_d^0 \rightarrow D^-\pi^+$	16
1.8	The interference between the decays $B_d^0 \rightarrow D^+\pi^-$ and $B_d^0 \rightarrow \overline{B}_d^0 \rightarrow D^+\pi^-$	16
1.9	Current status of experimental and HQE ratios of B hadron lifetimes.	20
2.1	The LHCb detector.	23
2.2	Polar angles of hadrons formed from $b\bar{b}$ pairs calculated by PYTHIA.	24
2.3	Momenta of B mesons decaying through the channel $B_d^0 \rightarrow D^+\pi^-$	24
2.4	Probabilities for having a number of inelastic pp interactions at a given luminosity, for a total inelastic cross section of $500 \mu\text{b}$	25
2.5	The VELO subdetector, showing the vacuum vessel containing the sensors in the centre, and readout electronics on either side of the beamline.	26
2.6	The layout of the VELO sensor stations in the $\mathbf{x-z}$ plane. The red lines are the R sensors, the blue lines are the ϕ sensors.	26
2.7	A close up of the RF foil enclosing the VELO's secondary vacuum.	27
2.8	Cluster shapes for the depleted and underdepleted cases for a simple segmented p-on-n sensor, (a) and (b), and for a sensor with a double metal layer, (c) and (d). At full depletion the charge is focused on the diodes, while at underdepletion the clusters spread (b) and lose charge to the double metal layer (d). Figure and text of caption reproduced from [col01].	29
2.9	The single hit resolutions for VELO R and ϕ sensors. Binary refers to the resolution that would be achieved if the only information read out was whether or not the sensor registered a hit. The blue fit is made with data from the analog readout.	31
2.10	The VELO primary vertex x resolution.	31
2.11	The VELO primary vertex z resolution.	31
2.12	The VELO resolution on the B decay vertex in the x direction.	31
2.13	The VELO resolution on the B decay vertex in the z direction.	31

2.14	A schematic of the LHCb tracking system, showing the VELO, trigger tracker, T1-T3 tracking stations, and the different types of tracks reconstructed at LHCb. The LHCb B field in the y direction, at each point in z , is shown for reference.	32
2.15	A schematic of the LHCb magnet. The quoted lengths are in mm.	33
2.16	Front view of a tracker station, with dimensions in cm.	34
2.17	Side view of a tracker station, with dimensions in cm. The stereo detector layers are labelled u and v	34
2.18	Performance of the long track reconstruction. a) shows the reconstruction efficiency with momentum, b) the ghost rate as a function of a momentum cut, and c) the ghost rate as a function of a transverse momentum cut. . . .	36
2.19	Resolution on the reconstructed track parameters at the production vertex of the track: (a) momentum resolution as a function of track momentum, (b) impact parameter resolution as a function of $1/p_t$. For comparison, the momentum and transverse-momentum spectra of B-decay particles are shown in the lower part of the plots.	36
2.20	Momentum resolution and impact parameter precision for B-decay tracks. The momentum resolution is fitted with a single Gaussian.	37
2.21	The reconstructed mass of $B_d^0 \rightarrow \pi\pi$ in the absence of hadron identification.	38
2.22	The reconstructed mass of $B_d^0 \rightarrow \pi\pi$ using hadron identification.	38
2.23	RICH 1 schematic in the $\mathbf{x}$ - $\mathbf{z}$ plane.	40
2.24	RICH 2 schematic in the $\mathbf{x}$ - $\mathbf{z}$ plane.	40
2.25	A schematic of the principle behind the HPD, with a close up of the silicon sensor in the bottom right. Picture taken from [Gys].	40
2.26	A typical LHCb event as seen by RICH 1. Fitted rings are superimposed over the HPD hits, with solid lines for long tracks and dashed lines for other track types.	43
2.27	A typical LHCb event as seen by RICH 2. Fitted rings are superimposed over the HPD hits, with solid lines for long tracks and dashed lines for other track types.	43
2.28	Difference in log-likelihood between the kaon and pion hypotheses from the RICH system, for (a) kaons and (b) pions, in a sample of $B_s^0 \rightarrow D_s K^+$ events; the shaded histograms are for low momentum tracks. Caption and figure taken from [col00b].	44
2.29	The overall (RICH1 and RICH2) kaon identification efficiency and pion misidentification rate as a function of momentum, for two different values of the $\Delta \ln \Lambda_{K\pi}$ cut.	44
2.30	The pion-kaon separation in sigma as a function of particle momentum for pions from the decay $B_d^0 \rightarrow \pi\pi$. The solid line represents the average separation at each momentum.	44
2.31	The new HLT, as used in the DC06 LHCb Monte Carlo.	52
2.32	A schematic of the LHCb computing model, reproduced from [Bro04]	52
3.1	The 2D impact parameter of a track with respect to a vertex.	58
3.2	Background and signal distributions of cut quantities	60

3.3	The mass distributions for the various background categories after the preselection. The plots are unnormalized, the vertical axis represents the number of observed events in each bin of mass (width ~ 8 MeV). The dashed vertical lines demarcate the signal region.	66
3.4	Invariant mass of candidates passing selection on the generic $b\bar{b}$ sample, showing both the signal peak and the surviving background.	71
3.5	The reconstructed mass of $B_d^0 \rightarrow D^+ K^-$ decays misidentified as $B_d^0 \rightarrow D^+ \pi^-$ (dashed), shown against the reconstructed mass of true $B_d^0 \rightarrow D^+ \pi^-$ decays (solid). The total number of misidentified and true events plotted is the same.	74
3.6	The reconstructed mass of $\Lambda_b \rightarrow \Lambda_c (K\pi p) \pi$ decays misidentified as $B_d^0 \rightarrow D^+ \pi^-$ (dashed), shown against the reconstructed mass of true $B_d^0 \rightarrow D^+ \pi^-$ decays (solid). The total number of misidentified and true events plotted is the same.	74
3.7	The mass resolution of signal reconstructed B_d^0 mesons.	77
3.8	The lifetime resolution of signal reconstructed B_d^0 mesons.	77
3.9	The mass of the reconstructed D^+ from the decay $B_d^0 \rightarrow D^+ \pi^-$	77
3.10	Fitted lifetime acceptance for signal $B_d^0 \rightarrow D^+ \pi^-$ events passing the trigger and offline selection.	77
4.1	$A_{m,\eta}(\tau)$ with detector effects included.	84
4.2	$A_{m,\eta}(\tau)$ in the absence of any detector effects.	84
4.3	The generated asymmetry (solid red) and the fitted asymmetry (dashed blue).	84
4.4	Errors on γ as a function of $\phi_{\text{mix}} - \gamma$ and the error on $ \eta $	88
4.5	A plot of the ambiguous solutions for γ and Δ_{qcd} for true $\gamma = 60^\circ$ and $\Delta_{\text{qcd}} = 40^\circ$	89
4.6	The error on $\text{Im}(\eta)$ as a function of the dilution and angle $\phi_{\text{mix}} - \gamma$	91
4.7	Pull plot on u_+ for an absolute bias on the mistag of 1.0%.	92
4.8	Pull plot on u_- for an absolute bias on the mistag of 2.0%.	92
4.9	Pull plot on u_+ when fitting with the mistag as a free parameter.	93
4.10	The value of the fitted mistag, for an input value of 0.35.	93
4.11	Pull plot on u_+ when fitting with a time resolution of 40 fs for 360 one year measurements.	95
4.12	Pull plot on u_+ when fitting with a time resolution of 40 fs for 360 five year measurements.	95
4.13	Pull plot on u_+ when fitting with an error on $ \eta $ of 10% for 360 one year measurements.	95
5.1	Contours in $\gamma - \text{U-spin}$ breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 60^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.	106

5.2	Contours in $\gamma - \text{U-spin}$ breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 60^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.	107
5.3	Contours in $\gamma - \text{U-spin}$ breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 10^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.	108
5.4	Contours in $\gamma - \text{U-spin}$ breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 10^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.	108
6.1	Proper lifetimes of reconstructed signal $B_d^0 \rightarrow D^- \pi^+$ events passing the L1 trigger, divided by the exponential distribution of B_d^0 lifetimes. This is the trigger lifetime acceptance function.	113
6.2	Impact parameter cuts translate to lifetime cuts	118
6.3	Re-establishing the direct link between online impact parameter cuts and measured lifetime (measured from offline data). [ABH ⁺ 03]	121
6.4	The average difference between 2D online and offline impact parameters, versus the measured offline 2D impact parameter.	122
6.5	The average difference between 3D online and offline impact parameters, versus the measured offline 3D impact parameter.	122
6.6	The average of the 2D offline impact parameter multiplied by the difference between 2D online and offline impact parameters, versus the measured offline 2D impact parameter.	122
6.7	The average of the 3D offline impact parameter multiplied by the difference between 3D online and offline impact parameters, versus the measured offline 3D impact parameter.	122
6.8	The 2D online track finding efficiency versus the 3D impact parameter of the offline track.	123
6.9	The online-offline IP resolution for 3D tracks.	123
6.10	The track finding efficiency of 3D online tracks matched to 2D online tracks passing the 2D IP cut, versus the 2D impact parameter of the matching 2D track.	124
6.11	The track finding efficiency of 3D online tracks matched to 2D online tracks passing the 2D IP cut, versus the 3D impact parameter of the matching offline track.	124
6.12	This illustrates that it is necessary to take the means (marked with a cross) and spread of each variable into account when finding the direction of best separation[ABH ⁺ 03]	128

6.13	Fisher scalar distribution for a sample of $\sim 3,000$ signal (shaded) and $\sim 3,000$ background (unshaded) events.	129
6.14	The signal probability associated with a value of the scalar Fisher discriminant in Figure 6.13	129
6.15	The difference between the measured online and offline transverse momentum of final state tracks from the decay $B_d^0 \rightarrow D^-\pi^+$	130
6.16	Lifetime fit for 10571 full Monte Carlo signal events.	133
6.17	The fit residual in Figure 6.16.	133
6.18	The L1 lifetime acceptance function (data points), with the fitted acceptance function (solid red line) overlaid.	133
6.19	The pre-trigger (solid blue) and post-trigger (dashed red) lifetime distributions for the toy Monte Carlo.	135
6.20	The pre-trigger (solid blue) and post-trigger (dashed red) lifetime distributions for the full Monte Carlo.	135
6.21	Lifetime fit to 60,000 toy Monte Carlo signal events.	136
6.22	Pull plot of 323 Toy MC lifetime fits of 4.9k signal events each.	136
6.23	Lifetimes fit to signal and background full Monte Carlo events.	138
6.24	Lifetimes fit to signal and background full Monte Carlo events, with a log scale.	138
6.25	The mass fit for signal and background events.	139
6.26	A pull plot for 296 one day experiments, with 4900 signal (toy MC) and 3240 background (full MC) events each	140
6.27	The fitted lifetimes of 296 one day experiments	140
6.28	The errors on the fitted lifetimes of 296 one day experiments	140
B.1	The background Category algorithm decision tree.	147
C.1	Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 30^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	149
C.2	Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 30^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	149
C.3	Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 85^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	150
C.4	Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 85^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	150

C.5	Contours in $\gamma - \text{U-spin}$ breaking space for one year of data taking, with $\gamma = 40^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	151
C.6	Contours in $\gamma - \text{U-spin}$ breaking space for five years of data taking, with $\gamma = 40^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	152
C.7	Contours in $\gamma - \text{U-spin}$ breaking space for one year of data taking, with $\gamma = 80^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	153
C.8	Contours in $\gamma - \text{U-spin}$ breaking space for five years of data taking, with $\gamma = 80^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	153
C.9	Contours in $\gamma - \text{U-spin}$ breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 140^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	154
C.10	Contours in $\gamma - \text{U-spin}$ breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 140^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).	154
E.1	This illustrates that it is necessary to take the means (marked with a cross) and spread of each variable into account when finding the direction of best separation[ABH ⁺ 03]	159

List of Tables

1.1	Summary of fit results	5
2.1	Performance numbers for VELO sensor prototypes	30
2.2	Tagging performance	48
2.3	L0 trigger performance	50
2.4	L1 trigger performance	51
3.1	Program versions	57
3.2	Preselection Cuts for $B_d^0 \rightarrow D^+\pi^-$	61
3.3	Background Category Tabulation	63
3.4	Background composition	65
3.5	Optimized cut values for $B_d^0 \rightarrow D^+\pi^-$	69
3.6	Trigger efficiencies	70
3.7	Numbers of generated and selected events	76
3.8	Reconstruction efficiencies	76
4.1	Fit Parameters	85
4.2	Precisions on observables in $B_d^0 \rightarrow D^-\pi^+$	86
4.3	Ambiguous solutions	89
5.1	Precisions on observables in $B_d^0 \rightarrow D^-\pi^+$	104
5.2	Combined analysis input parameters	104
5.3	Combined analysis results	110
6.1	Summary of TISTOSTOB	131
6.2	Summary of fit results	141

Introduction

LHCb is a dedicated B physics experiment at the Large Hadron Collider (LHC) at CERN, which is scheduled to begin taking data in the autumn of 2008 and will make precision measurements of CP violating effects in the B system. The research in this thesis is concerned with three physics analyses at LHCb:

1. The extraction of the CKM angle γ from the time-dependent CP asymmetries in the decay mode $B_d^0 \rightarrow D^- \pi^+$;
2. The unambiguous extraction of the CKM angle γ from the combined analysis of the decay modes $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D^- K^+$, performed under the assumption of U-spin symmetry;
3. The measurement of B hadron lifetimes without recourse to Monte Carlo information in decay modes whose lifetime distributions are biased by the LHCb trigger.

The document is structured as follows:

- **Chapter 1** : The formalism for describing the CP violation in the B^0 system is introduced, and the status of CP violation in the Standard Model is reviewed. The precision measurement of CP violation parameters in the B^0 system at LHCb is motivated, with particular reference to the measurement of the CKM angle γ . The concepts of factorization, U-spin symmetry, and Operator Product Expansion are introduced as theoretical tools for studying the weak decays of B hadrons. The measurement of B meson lifetimes at LHCb is motivated with reference to the testing of QCD models such as the Heavy Quark Expansion.
- **Chapter 2** : The LHCb detector is introduced, and an overview of the major sub-detectors is given. The tagging of B^0 flavour at production is described. The LHCb trigger structure and performance is discussed in detail. An overview is given of the LHCb simulation software, and the LHCb distributed computing model is presented.
- **Chapter 3** : The reconstruction of the decay mode $B_d^0 \rightarrow D^- \pi^+$ is discussed. A general method of categorizing background in LHCb Monte Carlo events is described. An offline selection for $B_d^0 \rightarrow D^- \pi^+$ is developed and optimized using a genetic algorithm technique. Reconstruction performance, such as the mass, vertex, and lifetime resolutions, is shown.

- **Chapter 4** : The feasibility of an extraction of the CKM angle γ from the decay mode $B_d^0 \rightarrow D^- \pi^+$ is studied. The optimized selection developed in Chapter 3 is used to construct a toy Monte Carlo for the extraction of γ , and the precision on γ achievable at LHCb from this analysis is estimated.
- **Chapter 5** : An alternative extraction of γ , from the combined analysis of the decay modes $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D^- K^+$, is discussed. The concept of U-spin symmetry is developed in the context of these decay modes, and U-spin relations between them are derived. The precision achievable on γ at LHCb from this analysis is estimated, and compared to the precision shown in Chapter 4.
- **Chapter 6** : The problem of measuring B lifetimes in decay channels whose lifetime distributions are biased by the LHCb trigger is discussed. A method for measuring these biases without recourse to Monte Carlo data is introduced and developed in the context of LHCb. The precisions achievable on B_d^0 and B_s^0 lifetimes at LHCb is estimated.
- **Chapter 7** : The results of the previous chapters are summarized.

Chapters 3, 4, 5, and 6 contain original research carried out by the author. All webpages referred to in this thesis exist at the time of writing.

Chapter 1

CP violation in heavy flavour physics

1.1 Introduction

The Standard Model has been the cornerstone of particle physics for the past forty years. Its account of the physics of subatomic particles has withstood considerable experimental scrutiny, accurately predicting the existence and properties of the Z and W bosons among other successes. Nonetheless, various aspects of the Standard Model remain relatively poorly constrained by experiment¹, prominent among which is the precise nature of CP violation. The LHCb detector, due to begin taking data in late 2008 at the Large Hadron Collider (LHC) in CERN has been built to study this problem.

Although the Standard Model is compatible with CP violation in the mixing and decay of neutral mesons, there are reasons to believe that LHCb will observe new physics effects associated with CP violation in the decays of B mesons which are incompatible with the Standard Model. As noted by Sakharov [Sak67], CP violation is required for baryogenesis under the Big Bang model of the universe, yet the CP violation allowed under the Standard Model is not sufficient to explain the matter-antimatter asymmetry observed in the universe today. Another motivation for new physics is the hierarchy problem, in particular the fact that the quantum corrections to the mass of the Higgs boson are divergent at high energies in the Standard Model, whereas the Standard Model Higgs is expected to have a mass of order hundred GeV [CMPP79]. Therefore, if the Standard Model is valid even at high energies, a “fine tuned” cancellation of these divergent corrections would be needed to keep the mass of the Higgs boson low. Since such a solution is quite unnatural, there is good reason to expect new physics effects (for example Super Symmetry) to appear at TeV energy scales, which would remove the need this fine-tuning. Such physics need not be CP conserving, or transform under CP in the same manner as the Standard Model. By the same argument, failing to observe any new physics at the TeV scale could be indicative of a new fundamental symmetry responsible for the low Higgs mass.

While studies of CP violation seek to probe the theoretical description of the weak force,

¹As well as conceptual and philosophical difficulties such as the hierarchy problem, or the various parameters which have to be inserted into the theory by hand and have no a priori theoretical motivation.

the B hadrons are bound together by the strong force, and predictions of Quantum Chromodynamics (QCD) are often needed to relate experimental measurements in the B meson sector to CP violating parameters. The large numbers of B hadrons produced at the LHC will therefore also provide an opportunity to test QCD based predictions. In particular, the Heavy Quark Expansion (HQE) model predicts a precise pattern of ratios of B hadron lifetimes [Shi00, Ura98, Pet04, GOP04]. Any deviation from the HQE predictions observed at LHCb could provide valuable clues about the structure of QCD.

Since this thesis is concerned with studying the decays of B mesons, an overview of the theory of B meson propagation, mixing, and decay will lay the groundwork for the subsequent analysis. Natural units ($\hbar = c = 1$) are used throughout the thesis.

1.2 Symmetries in particle physics

One of the most fundamental principles in physics is the connection between conservation laws and symmetries of nature. Noether's theorem [Noe18] shows that if a physical system is invariant under a continuous transformation, there exists an associated conserved quantity; the conservation of energy follows from the invariance of physical laws under translations in time, the conservation of momentum follows from invariance under translations in space, and angular momentum is conserved because the laws of physics are invariant under spatial rotations.

Unlike continuous symmetries, discrete symmetries, such as parity transformation, are found to be broken in certain physical interactions. For particle physics, the relevant discrete symmetries are:

- C : charge conjugation changes the sign of all additive quantum numbers. With specific reference to the decay of a sub-atomic particle, charge conjugation consists of swapping every particle in the decay for its antiparticle.
- P : parity is the operation of reflection in one of the three spatial axes.
- T : the time reversal operator reverses the direction of motion by reflection in the time axis.

Although each of these three discrete symmetries is broken in weak interactions, the combined symmetry CPT is an exact symmetry in any local Lagrangian field theory.

C, P, T operators

For completeness sake, the effect of the C , P , and T operators on the spin 0, spin $\frac{1}{2}$, and spin 1 fields of the Standard Model is listed in Table 1.1. Since each of these operators is its own inverse, their eigenvalues are ± 1 .

Table 1.1: The C , P , and T operators acting on Standard Model fields.

Spin of field	Field	Charge conjugation	Parity	Time reversal
0	$\phi(t, \vec{x})$	$\phi^\dagger(t, \vec{x})$	$\phi(t, -\vec{x})$	$\phi(-t, \vec{x})$
1/2	$\psi(t, \vec{x})$	$i\gamma^2\gamma^0\bar{\psi}^T(t, \vec{x})$	$\gamma^0\psi(t, -\vec{x})$	$\gamma^1\gamma^3\psi(-t, \vec{x})$
1	$A_\mu(t, \vec{x})$	$-A^\mu(t, \vec{x})$	$A^\mu(t, -\vec{x})$	$A^\mu(-t, \vec{x})$

1.2.2 Observations of symmetry violation

Although most interactions are invariant under each of the C , P , and T symmetry transformations, it has long been known that weak interactions violate both C and P symmetry [LY56]. Furthermore, although CP is a good symmetry in most weak interactions, CP violation was first observed in the neutral kaon system [CCFT64], and has recently been seen in the B system as well [ea04b, ea01a]. Details of these measurements are given in Section 1.4.2. Invariance under CPT is a requirement of all relativistic field theories, and is constrained to $< 10^{-18}$ from measurements of the $K^0-\bar{K}^0$ mass difference[ea06c].

Model independent description of the B^0 system

This section will develop the mixing and CP violation formalism in the B^0 system, but equivalent relations will be quoted at the end for the B_s^0 meson, since the decay rate asymmetries in $B_s^0 \rightarrow D_s^- K^+$ decays are needed in the combined extraction of the CKM angle γ presented in Chapter 5.

B_d^0 mesons produced at the LHC are created in one of two distinct flavour eigenstates

$$B_d^0 = (\bar{b}, d), \bar{B}_d^0 = (b, \bar{d}). \quad (1.1)$$

However, their propagation must be described using mass eigenstates which do not correspond to these flavour eigenstates. In particular, the two mass eigenstates $|B_L\rangle$ and $|B_H\rangle$ can be written as a linear combination of the two flavour eigenstates:

$$|B_{L,H}\rangle = p |B_d^0\rangle \pm q |\bar{B}_d^0\rangle. \quad (1.2)$$

Since the CP operator simply transforms one flavour eigenstate into another, the mass eigenstates in the absence of CP violation are equivalent to the eigenstates of the CP operator. Also note that $|p| + |q| = 1$.

$B_d^0 - \overline{B}_d^0$ mixing

The Schrödinger equation for the time evolution of a superposition of flavour eigenstates, $p|B_d^0\rangle \pm q|\overline{B}_d^0\rangle$, is given by

$$i\frac{d}{dt}\begin{pmatrix} p \\ q \end{pmatrix} = \mathbf{H}\begin{pmatrix} p \\ q \end{pmatrix}. \quad (1.3)$$

In principle, $\mathbf{H}$ would be Hermitian, if it were not for the fact that the B_d^0 will decay. Hence it is useful to rewrite $\mathbf{H}$ as a sum of two Hermitian matrices $\mathbf{M}$ and $\mathbf{\Gamma}$

$$\mathbf{H} = \mathbf{M} - \frac{i}{2}\mathbf{\Gamma}. \quad (1.4)$$

Furthermore, $\mathbf{M}$ and $\mathbf{\Gamma}$ can be written as:

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{12}^* & m_{11} \end{pmatrix}, \mathbf{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{11} \end{pmatrix}, \quad (1.5)$$

where the diagonal elements of the matrices have been set equal to each other because of CPT invariance, which implies that

$$\langle B_d^0 | \mathbf{H} | B_d^0 \rangle = \langle \overline{B}_d^0 | \mathbf{H} | \overline{B}_d^0 \rangle. \quad (1.6)$$

The task is now to find the eigenvalues of $\mathbf{H}$ in terms of the elements of $\mathbf{M}$ and $\mathbf{\Gamma}$. The eigenvalues are given by

$$\lambda_{H,L} = m_{11} - \frac{i}{2}\Gamma_{11} \pm \sqrt{\left(m_{12} - \frac{i}{2}\Gamma_{12}\right)\left(m_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}, \quad (1.7)$$

which can be rewritten as

$$\lambda_{H,L} = \left(m_{11} \pm \frac{\Delta m}{2}\right) - \frac{i}{2}\left(\Gamma_{11} \pm \frac{\Delta \Gamma}{2}\right) \quad (1.8)$$

where Δm and $\Delta \Gamma$ refer to the mass and lifetime difference between the two mass eigenstates, and can also be related by the expression

$$\Delta m - \frac{i}{2}\Delta \Gamma = 2\sqrt{\left(m_{12} - \frac{i}{2}\Gamma_{12}\right)\left(m_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}. \quad (1.9)$$

From the definition of the mass eigenstates in Equation 1.2, it follows that the eigenvector equation is

$$\mathbf{H}\begin{pmatrix} p \\ \pm q \end{pmatrix} = \lambda_{H,L}\begin{pmatrix} p \\ \pm q \end{pmatrix}, \quad (1.10)$$

which has the solutions

$$\frac{q}{p} = -\frac{\Delta m - \frac{i}{2}\Delta \Gamma}{2\left(m_{12} - \frac{i}{2}\Gamma_{12}\right)}. \quad (1.11)$$

Present measurements [ea06c] put the mass difference in the B_d^0 system at the same order of magnitude as the width:

$$x_d = \frac{\Delta m_{B_d^0}}{\Gamma_{B_d^0}} = 0.776 \pm 0.008. \quad (1.12)$$

In the Standard Model, the ratio of the lifetime difference to the mass difference is expected to be proportional to m_b^2/m_t^2 , and hence very small. The precise size of this difference is at present only weakly constrained by experiment [ea03]:

$$\Delta\Gamma_{B_d^0}/\Gamma_{B_d^0} = -0.008 \pm 0.037 \text{ (stat)} \pm 0.018 \text{ (syst)}. \quad (1.13)$$

Setting $\Delta\Gamma = 0$ in Equation 1.11, and squaring both sides:

$$\frac{p^2}{q^2} = \frac{\Delta m^2}{4 \left(m_{12} - \frac{i}{2}\Gamma_{12}\right) \left(m_{12}^* - \frac{i}{2}\Gamma_{12}^*\right)}, \quad (1.14)$$

whereupon Equation 1.9 gives us $|p| = |q|$. The small value of $\Delta\Gamma$ makes this an excellent approximation in the B_d^0 system, and hence CP violation is expected to be negligible in B_d^0 mixing.

B_d^0 decays

In order to describe the decay of a B_d^0 , it is first of all necessary to describe the time evolution of the flavour eigenstates. In order to simplify notation, $\Delta\Gamma = 0$ will be assumed throughout, and Γ will simply refer to the common decay width of the two mass eigenstates. Now, for a particle created as a B_d^0 at $t = 0$,

$$|B_d^0(t)\rangle = \frac{1}{2p} \left(|B_H\rangle e^{-i(M_H - \frac{i}{2}\Gamma)t} + |B_L\rangle e^{-i(M_L - \frac{i}{2}\Gamma)t} \right). \quad (1.15)$$

Substituting Equation 1.2 for the mass eigenstates,

$$|B_d^0(t)\rangle = \frac{1}{2} e^{-\frac{\Gamma t}{2}} \left[\left(|B_d^0\rangle - \frac{q}{p} |\overline{B_d^0}\rangle \right) e^{-iM_H t} + \left(|B_d^0\rangle + \frac{q}{p} |\overline{B_d^0}\rangle \right) e^{-iM_L t} \right]. \quad (1.16)$$

The amplitude for a decay to a final state $|f\rangle$, at time t , is now given by

$$\begin{aligned} \langle f | \mathbf{H}_I | B_d^0(t) \rangle &= \frac{1}{2} e^{-\frac{\Gamma t}{2}} \langle f | \mathbf{H}_I | B_d^0 \rangle \times \\ &\times \left[\left(1 - \frac{q}{p} \frac{\langle f | \mathbf{H}_I | \overline{B_d^0} \rangle}{\langle f | \mathbf{H}_I | B_d^0 \rangle} \right) e^{-iM_H t} + \left(1 + \frac{q}{p} \frac{\langle f | \mathbf{H}_I | \overline{B_d^0} \rangle}{\langle f | \mathbf{H}_I | B_d^0 \rangle} \right) e^{-iM_L t} \right], \end{aligned} \quad (1.17)$$

where $\mathbf{H}_I$ is the interaction Hamiltonian. It is useful to define

$$\eta \equiv \frac{q}{p} \frac{\langle f | \mathbf{H}_I | \overline{B_d^0} \rangle}{\langle f | \mathbf{H}_I | B_d^0 \rangle}, \quad (1.18)$$

and

$$\overline{\eta} \equiv \frac{p}{q} \frac{\langle \overline{f} | \mathbf{H}_I | B_d^0 \rangle}{\langle \overline{f} | \mathbf{H}_I | \overline{B_d^0} \rangle}, \quad (1.19)$$

which allow the decay rates to be calculated for all four decay modes $B_d^0, \overline{B}_d^0 \rightarrow f, \overline{f}$. They are:

$$\begin{aligned} \left| \langle f | \mathbf{H}_I | B_d^0(t) \rangle \right|^2 &= \frac{1}{2} \left| \langle f | \mathbf{H}_I | B_d^0 \rangle \right|^2 \times \\ &\times e^{-\Gamma t} \left\{ (1 + |\eta|^2) + (1 - |\eta|^2) \cos(\Delta m t) - 2\text{Im}(\eta) \sin(\Delta m t) \right\}; \quad (1.20) \end{aligned}$$

$$\begin{aligned} \left| \langle f | \mathbf{H}_I | \overline{B}_d^0(t) \rangle \right|^2 &= \frac{1}{2} \left| \langle f | \mathbf{H}_I | B_d^0 \rangle \right|^2 \times \\ &\times e^{-\Gamma t} \left\{ (1 + |\eta|^2) - (1 - |\eta|^2) \cos(\Delta m t) + 2\text{Im}(\eta) \sin(\Delta m t) \right\}; \quad (1.21) \end{aligned}$$

$$\begin{aligned} \left| \langle \overline{f} | \mathbf{H}_I | \overline{B}_d^0(t) \rangle \right|^2 &= \frac{1}{2} \left| \langle \overline{f} | \mathbf{H}_I | \overline{B}_d^0 \rangle \right|^2 \times \\ &\times e^{-\Gamma t} \left\{ (1 + |\overline{\eta}|^2) + (1 - |\overline{\eta}|^2) \cos(\Delta m t) - 2\text{Im}(\overline{\eta}) \sin(\Delta m t) \right\}; \quad (1.22) \end{aligned}$$

$$\begin{aligned} \left| \langle \overline{f} | \mathbf{H}_I | B_d^0(t) \rangle \right|^2 &= \frac{1}{2} \left| \langle \overline{f} | \mathbf{H}_I | \overline{B}_d^0 \rangle \right|^2 \times \\ &\times e^{-\Gamma t} \left\{ (1 + |\overline{\eta}|^2) - (1 - |\overline{\eta}|^2) \cos(\Delta m t) + 2\text{Im}(\overline{\eta}) \sin(\Delta m t) \right\}. \quad (1.23) \end{aligned}$$

CP violation in B_d^0 decays

CP violation in B_d^0 decays manifests itself through the decay rate asymmetries between the above four rates, where the CP violating phases enter through the parameters $\eta, \overline{\eta}$. In the most general case, two time dependent asymmetries can be formed:

$$A_\eta(t) = \frac{\Gamma(B_d^0 \rightarrow f)(t) - \Gamma(\overline{B}_d^0 \rightarrow f)(t)}{\Gamma(B_d^0 \rightarrow f)(t) + \Gamma(\overline{B}_d^0 \rightarrow f)(t)}, \quad (1.24)$$

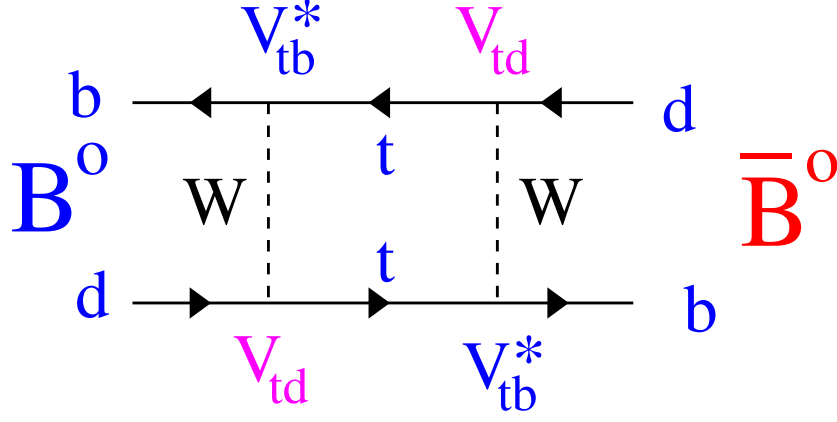
$$A_{\overline{\eta}}(t) = \frac{\Gamma(\overline{B}_d^0 \rightarrow \overline{f})(t) - \Gamma(B_d^0 \rightarrow \overline{f})(t)}{\Gamma(B_d^0 \rightarrow \overline{f})(t) + \Gamma(\overline{B}_d^0 \rightarrow \overline{f})(t)}, \quad (1.25)$$

corresponding to the asymmetry between the decays rates of the B_d^0 and $\overline{B}_d^0$ into the final state $f, \overline{f}$.

In decays to CP eigenstates

If the final state f is a CP eigenstate, the four decay rates reduce to two, and $|\eta| = 1$ when $p/q = 1$. Therefore the only CP violating parameter is the phase δ_{CP} , which parameterizes the phase difference between the two decay paths $\overline{B}_d^0 \rightarrow f$ and $B_d^0 \rightarrow f$, according to:

$$\eta = \pm e^{i\delta_{CP}}, \quad (1.26)$$

Figure 1.1: The $B_d^0 \rightarrow \overline{B}_d^0$ mixing diagram.

where the two values $\eta = +e^{i\delta_{CP}}$ and $\eta = -e^{i\delta_{CP}}$ correspond to the CP even and odd cases. The asymmetry is given² by:

$$A(t) = \pm \sin(\delta_{CP}) \sin(\Delta mt). \quad (1.27)$$

A well known case is the asymmetry in the decay $B_d^0 \rightarrow J/\psi K_S$, where the B_d^0 can either directly decay to $J/\psi K_S$, or oscillate to $\overline{B}_d^0$ and then decay to $J/\psi K_S$. One of the $B_d^0 \rightarrow \overline{B}_d^0$ oscillation box diagrams is shown in Figure 1.1. This interference between the mixed and unmixed decay amplitudes causes a CP violating asymmetry, the measurement of which was the first observation of CP violation in the B meson system [ea02, ea01b].

In decays to non- CP eigenstates

If f is not a CP eigenstate, all four decay rates are distinct, provided that B_d^0 and $\overline{B}_d^0$ can both decay to the same final state. In this case, there is an additional complication in the form of a possible strong phase difference, Δ_{qcd} , which does not violate CP , but nonetheless contributes to the total phase difference between the decay paths $\overline{B}_d^0 \rightarrow f$ and $B_d^0 \rightarrow f$. The parameters $\eta, \overline{\eta}$ are given by:

$$\eta = |\eta| e^{i(\Delta_{qcd} + \delta_{CP})}, \quad (1.28)$$

$$\overline{\eta} = |\eta| e^{i(\Delta_{qcd} - \delta_{CP})}, \quad (1.29)$$

while the asymmetries now reduce to:

$$A_\eta(t) = \frac{1 - |\eta|^2}{1 + |\eta|^2} \cos(\Delta mt) - \frac{2\text{Im}(\eta)}{1 + |\eta|^2} \sin(\Delta mt), \quad (1.30)$$

$$A_{\overline{\eta}}(t) = \frac{1 - |\eta|^2}{1 + |\eta|^2} \cos(\Delta mt) - \frac{2\text{Im}(\overline{\eta})}{1 + |\eta|^2} \sin(\Delta mt). \quad (1.31)$$

These equations are particularly relevant for this thesis, since they define the asymmetries for the $B_d^0 \rightarrow D^\mp \pi^\pm$ decays studied in Chapters 4 and 5.

²This assumes that there is only one decay diagram, which may not be true if the decay channel receives significant loop contributions which carry a different weak phase.

CP violation in B_s^0 decays

An equivalent derivation to the one presented so far can be performed in the B_s^0 system, but here the non-vanishing lifetime difference $\Delta\Gamma_{B_s^0}$ must be accounted for. The asymmetries are given by:

$$A_\eta(t) = \frac{C \cos(\Delta m_s t) + S \sin(\Delta m_s t)}{\cosh(\Delta\Gamma_s t/2) - A_{\Delta\Gamma} \sinh(\Delta\Gamma_s t/2)}, \quad (1.32)$$

$$A_{\bar{\eta}}(t) = \frac{\bar{C} \cos(\Delta m_s t) + \bar{S} \sin(\Delta m_s t)}{\cosh(\Delta\Gamma_s t/2) - A_{\Delta\Gamma} \sinh(\Delta\Gamma_s t/2)}, \quad (1.33)$$

where:

$$C = \frac{1 - |\eta|^2}{1 + |\eta|^2}, \quad (1.34)$$

$$\bar{C} = -\frac{1 - |\eta|^2}{1 + |\eta|^2}, \quad (1.35)$$

$$S = \frac{2\text{Im}(\eta)}{1 + |\eta|^2}, \quad (1.36)$$

$$\bar{S} = \frac{2\text{Im}(\bar{\eta})}{1 + |\bar{\eta}|^2}, \quad (1.37)$$

the parameter $\Delta\Gamma_s$ is the lifetime difference between B_s^0 and $\bar{B}_s^0$, and

$$A_{\Delta\Gamma} = -\frac{2\text{Re}(\eta)}{1 + |\eta|^2}, \quad (1.38)$$

is a CP -observable related to the width difference, and specific to the B_s^0 system. These parameters are related by:

$$|C|^2 + |S|^2 + |A_{\Delta\Gamma}|^2 = 1. \quad (1.39)$$

CP violation in the Standard Model

Weak interactions are described within the Standard Model through the $SU(2)$ group, and are observed to couple to left handed fermions only. This can be accounted for by choosing different $SU(2)$ representations for right handed and left handed fermions, whereby the right handed fermions are put into $SU(2)$ singlets with weak isospin 0, and left handed fermions are put into doublets with weak isospin $\pm 1/2$.

The method can be illustrated using the case of two quark generations, where the two isodoublets are

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L. \quad (1.40)$$

The d' and s' states are related to the physical d, s quarks through

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \mathbf{V}_C \begin{pmatrix} d \\ s \end{pmatrix}, \quad (1.41)$$

where $\mathbf{V}_C$ is a unitary 2×2 matrix which performs the rotation from one quark basis to another. The most general $\mathbf{V}_C$ will have four parameters, and can be written as

$$\mathbf{V}_C = \begin{pmatrix} e^{-i\phi_1} & \\ & 1 \end{pmatrix} \begin{pmatrix} \cos \theta_C & \sin \theta_C \\ -\sin \theta_C & \cos \theta_C \end{pmatrix} \begin{pmatrix} e^{-i\phi_2} & \\ & e^{-i\phi_3} \end{pmatrix}. \quad (1.42)$$

The crucial point here is that the three phases $\phi_{1,2,3}$ can be eliminated by rotating the basis of the u , d , and s quarks respectively, and in general for n quarks, the mixing matrix will be of order $\frac{n}{2} \times \frac{n}{2}$, with $n^2/4$ parameters, $n - 1$ of which will be phases that can be eliminated, leaving $n^2/4 - n + 1$ free parameters.

For the case of three quark generations, there are 6 quarks in total, leading to three doublets

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L, \quad (1.43)$$

and a 3×3 matrix relating the physical d, s, b quarks to the d', s' , and b' states coupling to the weak interaction

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (1.44)$$

better known as the Cabibbo-Kobayashi-Maskawa, or CKM, matrix [KM73].

1.4.1 The CKM Matrix

Complex phases in the CKM matrix

The CKM matrix parameterizes the mixing between three quark generations, and as a result it has nine parameters, five of which are relative phases between the quark states which can be factored out. This leaves four independent parameters, of which only three can be real, and hence leaves the CKM matrix for the case of three generations with one complex phase which fully parameterizes the CP violation within the Standard Model. Indeed it was precisely the fact that no CP violation was possible with only two quark generations that led Kobayashi and Maskawa to propose a third quark generation in 1973, as a way of allowing CP violation in the Standard Model [KM73].

The Wolfenstein parameterization

The Wolfenstein parameterization makes use of an observed hierarchy between the elements of the CKM matrix, whose orders of magnitude are structured as

$$\begin{pmatrix} 1 & \lambda & \lambda^3 \\ \lambda & 1 & \lambda^2 \\ \lambda^3 & \lambda^2 & 1 \end{pmatrix}, \quad (1.45)$$

to rewrite the matrix elements as a power series in λ , given here up to $O(\lambda^3)$

$$\begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho + i\eta) & -A\lambda^2 & 1 \end{pmatrix}, \quad (1.46)$$

where $\lambda = 0.22$ is the sine of the Cabibbo angle θ_C which parametrizes the quark mixing for the case of two generations above. The parameters A, ρ, λ, η are the four independent parameters for the case of six quarks, where the imaginary parameter η quantifies the CP violation in the Standard Model. This η is **not** the same as the η seen earlier in the decay rate asymmetries.

The CKM unitarity triangles

The CKM matrix is unitary, and this unitarity imposes nine conditions on the matrix elements, six of which can be represented as triangles in the complex plane, also known as “unitarity triangles”. The six conditions are given below, with the length of each side being given in square brackets alongside in orders of λ

$$V_{ud}^* V_{us} [O(\lambda^1)] + V_{cd}^* V_{cs} [O(\lambda^1)] + V_{td}^* V_{ts} [O(\lambda^5)] = 0, \quad (1)$$

$$V_{cd}^* V_{ud} [O(\lambda^1)] + V_{cs}^* V_{us} [O(\lambda^1)] + V_{cb}^* V_{ub} [O(\lambda^5)] = 0, \quad (2)$$

$$V_{us}^* V_{ub} [O(\lambda^4)] + V_{cs}^* V_{cb} [O(\lambda^2)] + V_{ts}^* V_{tb} [O(\lambda^2)] = 0, \quad (3)$$

$$V_{cd}^* V_{td} [O(\lambda^4)] + V_{cs}^* V_{ts} [O(\lambda^2)] + V_{cb}^* V_{tb} [O(\lambda^2)] = 0, \quad (4)$$

$$V_{ud}^* V_{td} [O(\lambda^3)] + V_{us}^* V_{ts} [O(\lambda^3)] + V_{ub}^* V_{tb} [O(\lambda^3)] = 0, \quad (5)$$

$$V_{ub}^* V_{ud} [O(\lambda^3)] + V_{cb}^* V_{cd} [O(\lambda^3)] + V_{tb}^* V_{td} [O(\lambda^3)] = 0. \quad (6)$$

The advantage of describing CP violation through the unitarity triangles is that any change in the parameterization of the CKM matrix merely rotates the triangles in the complex plane, while leaving their sides and internal angles unchanged. On the other hand, only two of the triangles, 5 and 6, are suitable for experimentally testing the CKM description of CP violation, as they are the only ones whose sides are of comparable length. As the two triangles coincide up to order λ^3 in the Wolfenstein parameterization, it is conventional to refer to triangle number 6 as the Unitarity Triangle. Figure 1.2 shows this triangle, along with the usual CKM angles α, β , and γ , sometimes referred to as ϕ_1, ϕ_2 , and ϕ_3 . The three angles relate to the complex components in the CKM matrix, and are given by:

$$\alpha = \arg \left(-\frac{V_{td} V_{tb}^*}{V_{ud} V_{ub}^*} \right) = \arg \left(\frac{(1 - \lambda^2/2)(i\eta - \rho)}{1 - \rho - i\eta} \right), \quad (1.47)$$

$$\beta = \arg \left(-\frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*} \right) = \arg \left(\frac{1}{1 - \rho - i\eta} \right), \quad (1.48)$$

$$\gamma = \arg \left(-\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*} \right) = \arg [(1 - \lambda^2/2)(\rho - i\eta)]. \quad (1.49)$$

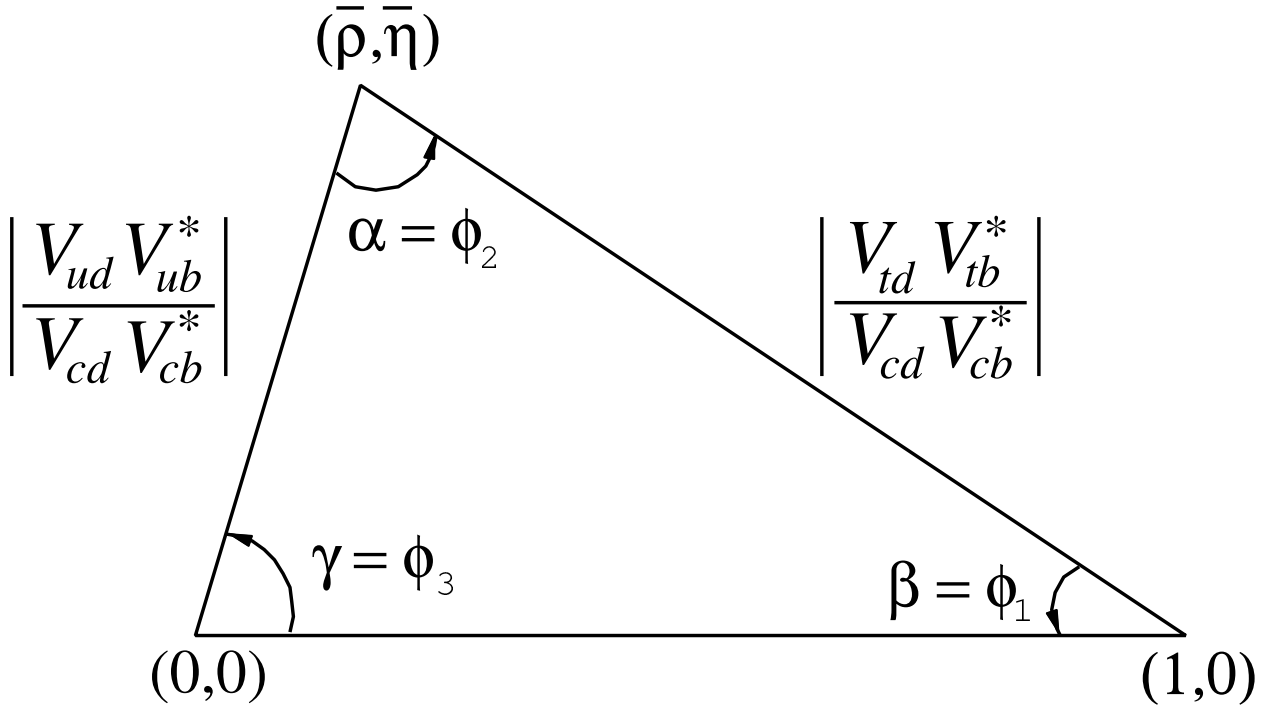


Figure 1.2: The CKM Unitarity Triangle, taken from [ea06c]. $\phi_{1,2,3}$ are equivalent to the angles α, β, γ , in the notation used by BELLE. The bottom of the triangle has unit length by convention.

1.4.2 Current constraints on the CKM matrix

At the time of writing, direct experimental constraints existed for all the magnitudes of all nine components of the CKM matrix, listed in Appendix A. The focus in this section will be on the constraints most directly relevant for the Unitarity Triangle, and the results presented have been taken from [ea05] unless otherwise cited.

Figure 1.3 shows the latest constraints on the Unitarity triangle, as combined by the CKMFitter collaboration [ea05]. The important direct constraints are on the quantities $\sin(2\beta)$, α , the B_d^0 and B_s^0 mixing parameters $\Delta m_{d,s}$, and the ratio $|V_{ub}/V_{cb}|$. Taken together, they constrain the top of the Unitarity Triangle in the Wolfenstein parametrization, which is located at the point $(\bar{\rho}, \bar{\eta}) \equiv (\lambda^2/2 - 1)(\rho, -i\eta)$ in the imaginary plane. If the combined constraint were to disfavour or exclude this value, it would be evidence of new physics, or the inconsistency of the individual measurements. At present, the angle γ is the least well directly measured element of the CKM matrix useful for constraining the Unitarity Triangle.

Constraint on $\sin(2\beta)$

The current world average is $\sin(2\beta) = 0.678 \pm 0.025$, obtained by the HFAG [ea06a] group by averaging measurements from time dependent asymmetry analyses of the $J/\psi K^0$, $\phi(2S) K_S$, $\xi_{C1} K_S$, $\eta_C K_S$, and $J/\psi K^{*0}$ decay modes of the B_d^0 at BaBar and BELLE.

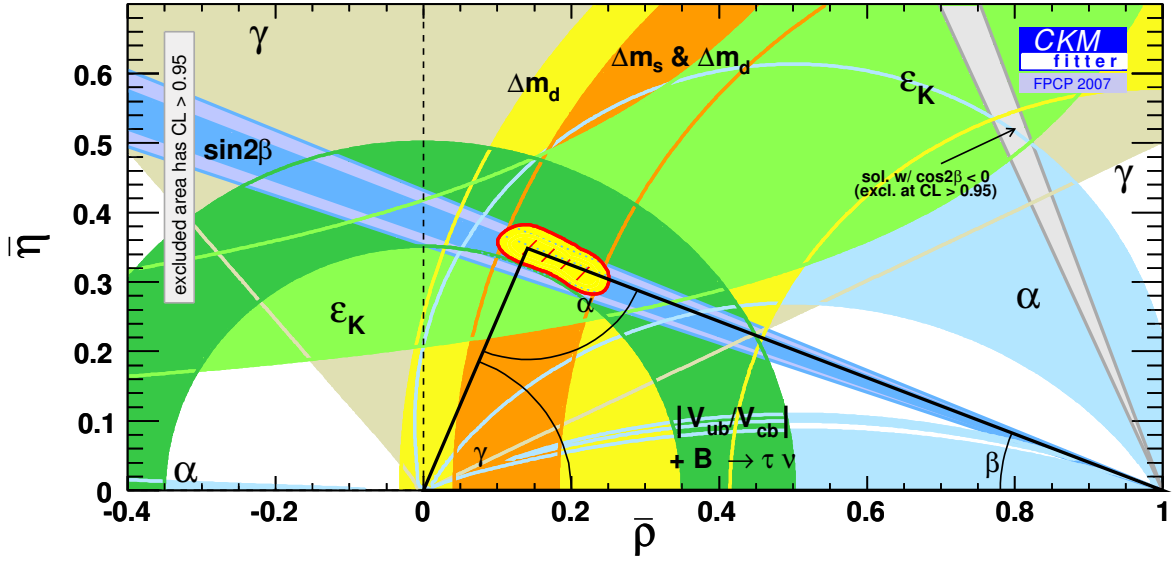


Figure 1.3: Current constraints on the Unitarity Triangle. Diagram taken from [ea05].

Constraint on α

The angle α can be directly measured from time dependent CP asymmetries in $b \rightarrow u\bar{u}b$ dominated modes. The single best measurement, $\alpha = 88^\circ \pm 17^\circ$, comes from the $B_d^0 \rightarrow \rho^+ \rho^-$ decay mode, which is believed to have small penguin contributions compared to other modes that measure α , but suffers from a twofold ambiguity. Since the angles α , β , and γ are related by:

$$\alpha + \beta + \gamma = \pi, \quad (1.50)$$

a measurement of α is equivalent to a measurement of $\beta + \gamma$. It would therefore be possible to discuss the constraints on the Unitarity Triangle simply in terms of measurements of β and γ . However, it is currently conventional to classify measurements of CKM angles as measurements of α , β , and γ , and we will follow this convention throughout the thesis.

Constraint on B_d^0 and B_s^0 mixing parameters

The B_d^0 mass difference Δm_d has been measured by the ALEPH, BaBar, BELLE, CDF, D0, DELPHI, L3, and OPAL collaborations using a variety of methods and decay channels. The world average [ea06a] is:

$$\Delta m_d = 0.507 \pm 0.004 \text{ ps}^{-1}. \quad (1.51)$$

The B_s^0 mass difference Δm_s has recently been measured by CDF [TCcG06], using a variety of semileptonic and hadronic decay modes:

$$\Delta m_s = 17.77 \pm 0.10 \pm 0.07 \text{ ps}^{-1}, \quad (1.52)$$

and this measurement dominates the world average.

Constraint on the ratio $|V_{ub}/V_{cb}|$

Both of these CKM matrix elements can be determined from inclusive or exclusive semileptonic B decays; the former suffers uncertainties from charm background such as $B \rightarrow X_c l \bar{\nu}$, the latter from the theoretical uncertainties associated with lattice QCD calculations of the semileptonic form factors. The current world average [ea06a] is dominated by inclusive measurements:

$$\begin{aligned} |V_{ub}| &= 0.00431 \pm 0.00030, \\ |V_{cb}| &= 0.0416 \pm 0.0006. \end{aligned}$$

Constraint on γ

Assuming that the Standard Model provides the complete description of CP violation, it is possible to obtain an indirect constraint on the Standard Model value of γ by combining the directly known precisions on the Unitarity Triangle parameters above with the Standard Model value for the top of the Unitarity Triangle. This indirect precision is currently given by [TUCB06]:

$$\gamma_{\text{indirect}} = 64.1^\circ \pm 4.6^\circ, \quad (1.53)$$

which can be compared to the latest direct constraints on γ by BaBar and Belle [botBc07]:

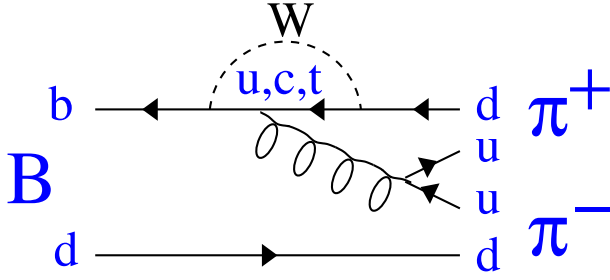
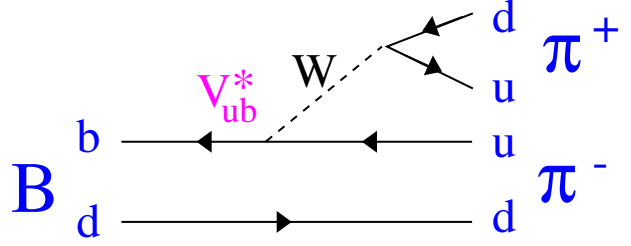
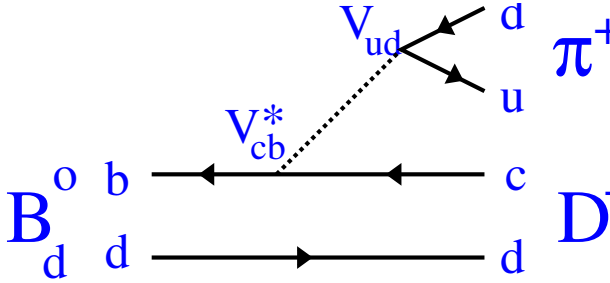
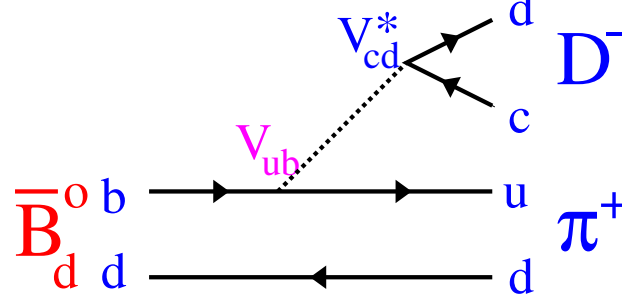
$$\gamma = (92 \pm 41 \pm 11 \pm 12)^\circ \text{ (BaBar)} \quad (1.54)$$

$$\gamma = (53_{-18}^{+15} \pm 3 \pm 9)^\circ \text{ (Belle)} \quad (1.55)$$

where BaBar constrain γ through a Dalitz analysis of $B^- \rightarrow DK^-$ and $B^- \rightarrow D^*K^-$, while the Belle constraint is obtained from a Dalitz analysis of $B^- \rightarrow DK^-$, $B^- \rightarrow D^*K^-$, and $B^- \rightarrow DK^{*-}$. Achieving a direct precision on γ comparable to the indirect precision from the other Unitarity Triangle measurements is therefore a key priority for LHCb, since it will be a powerful test of the Standard Model description of the Unitarity Triangle, and provide possible evidence of new physics.

Measuring the angle γ from $B_d^0 \rightarrow D^+\pi^-$ decay rate asymmetries

In order for an LHCb measurement of γ to provide evidence of new physics, it will also be important to have a measurement of γ in which Standard Model contributions are dominant, and which is therefore unaffected by any new physics. New physics is expected to enter the B_d^0 decays via penguin loop diagrams, an example of which is shown in Figures 1.4 and 1.5 for the case of $B_d^0 \rightarrow \pi\pi$. The channel $B_d^0 \rightarrow D^+\pi^-$ is a tree level dominated decay, which is not expected to receive any penguin loop contributions. The value of γ extracted from the decay rate asymmetries in this channel will be compared with the value of γ extracted from channels with new physics sensitivity, and any discrepancy will be evidence of new physics.

Figure 1.4: The one-loop Feynman diagram for the decay $B_d^0 \rightarrow \pi\pi$.Figure 1.5: The tree-level Feynman diagram for the decay $B_d^0 \rightarrow \pi\pi$.Figure 1.6: The Feynman graph for the decay $B_d^0 \rightarrow D^- \pi^+$ Figure 1.7: The Feynman graph for the decay $\bar{B}_d^0 \rightarrow D^- \pi^+$

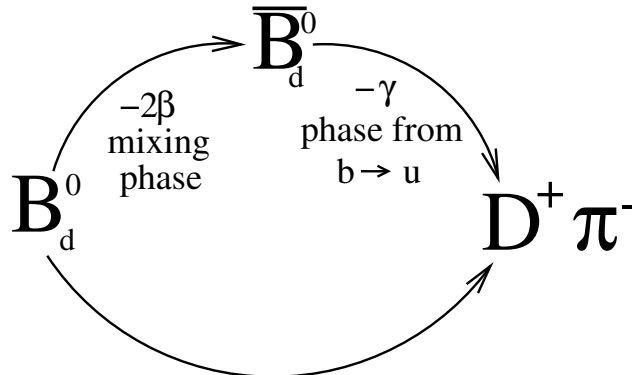
Figures 1.6 and 1.7 show the decay diagrams for B_d^0 and $\bar{B}_d^0$ to decay to the same final state, $D^- \pi^+$. Because both B_d^0 and $\bar{B}_d^0$ can decay to the same state, the CP violation in this channel will be a result of the interference between the decay $B_d^0 \rightarrow D^+ \pi^-$, and the decay $B_d^0 \rightarrow \bar{B}_d^0 \rightarrow D^+ \pi^-$, as illustrated by Figure 1.8. The angle γ is observable in this channel because of the V_{ub} transition in the second diagram, which makes the amplitude of this decay dependent on γ . Rewriting Equation 1.28 for this channel:

$$\eta = |\eta| e^{i(\Delta_{qcd} - 2\beta - \gamma)}, \quad (1.56)$$

$$\bar{\eta} = |\eta| e^{i(\Delta_{qcd} + 2\beta + \gamma)}, \quad (1.57)$$

where the δ_{CP} in Equation 1.28 is given by $2\beta + \gamma$ for this channel.

In order to extract γ , it will be necessary to have external knowledge of β , which is already

Figure 1.8: The interference between the decays $B_d^0 \rightarrow D^+ \pi^-$ and $B_d^0 \rightarrow \bar{B}_d^0 \rightarrow D^+ \pi^-$

well known, and can also be measured at LHCb in channels such as $B_d^0 \rightarrow J/\psi K_S$. Even if β is affected by new physics, it will be affected in the same way for both channels because the mixing diagram is the same in both cases. Since penguin contributions to $B_d^0 \rightarrow J/\psi K_S$ have the same phase as the tree level diagram they will not significantly change the extracted value of β .

Current measurements of γ from this channel have assumed $\Delta_{\text{qcd}} = 0$, in accordance with the factorization prediction, but no independent measurement of Δ_{qcd} exists at the time of writing.

Value of $|\eta|$

It is important to estimate the size of interference between the different tree-level diagrams in the decays $(B, \overline{B})_d^0 \rightarrow D^\mp \pi^\pm$, which is parameterized by the factor $|\eta|$. Assuming that any hadronic effects in the interference between the decay diagrams cancel, $|\eta|$ will be equal to the ratio of the CKM elements involved in the decay:

$$|\eta| = \frac{|V_{ub}| |V_{cd}|}{|V_{cb}| |V_{ud}|}, \quad (1.58)$$

which results in $|\eta| = 0.024 \pm 0.002$, using the values of the CKM elements listed in Appendix A. This value of $|\eta|$ will be assumed throughout this thesis. As will be seen in Chapter 4, this interference is too small to fit directly from the CP asymmetries, and will therefore have to be externally constrained in order to extract γ .

The statistical uncertainty on $|\eta|$ under the above assumption is therefore around 8%, and this number can be expected to improve before LHCb starts taking data. The theoretical uncertainty which stems from assuming that the hadronic effects cancel, i.e. perfect SU(3) symmetry, is harder to estimate. Recent BaBar [ea04a, ea06b] analyses of the decays $B_d^0 \rightarrow D^\mp \pi^\pm$ have estimated $|\eta|$ from the relation:

$$|\eta| = \tan(\theta_C) \sqrt{\frac{B(B^0 \rightarrow D_s^+ \pi^-)}{B(B^0 \rightarrow D_s^- \pi^+)}} \frac{f_D}{f_{D_s}}, \quad (1.59)$$

where θ_C is the Cabibbo angle, and the term $\frac{f_D}{f_{D_s}}$ is the ratio of decay constants for the D and D_s . This relation assumes SU(3) symmetry and it also ignores (small) W-exchange decay diagrams, and it results in the value

$$|\eta| = 0.019 \pm 0.004, \quad (1.60)$$

where 0.004 contains both the experimental and theoretical uncertainties, and is equivalent to an error of $\sim 20\%$ on $|\eta|$. Combined in quadrature with our previous calculation, the theoretical uncertainty can be estimated at $\sim 20\%$.

1.5.2 U-spin symmetry

It will be shown in Chapter 5 that, instead of trying to determine the value of $|\eta|$ under the assumption of perfect SU(3) symmetry, it is possible to eliminate this parameter from

the fits altogether by combining the analysis of the channels $(B, \overline{B})_d^0 \rightarrow D^\mp \pi^\pm$ with that of $(B, \overline{B})_s^0 \rightarrow D_s^\mp K^\pm$. This analysis relies on U-spin symmetry, which is an SU(2) subgroup of the flavour SU(3) symmetry. Under U-spin symmetry, the (d,s) pair of quarks is a doublet, in a similar way to (u,d) under isospin. Perfect U-spin symmetry implies that parameters associated with the strong force, such as the phases Δ_{qcd} in Section 1.4.2, have the same value for decays which are related by the interchange of a d quark for an s quark.

Continuing the analogy with isospin, U-spin is broken by the difference between the s and d quark masses, in the same way that isospin is broken by the difference between the d and u quark masses. Since the s quark is heavier than the u quark, U-spin breaking effects are greater than isospin breaking effects. On the other hand, there is no charge difference between the s and d quarks, so that U-spin symmetry, unlike isospin, is not broken by electroweak penguins. In addition, it has been argued in [SS07] that U-spin is expected to be a better symmetry than SU(3) when analyzing neutral B meson decays, because it does not require any assumptions about the relative amplitudes of different (tree, penguin, annihilation) decay topologies.

1.6 Factorization

Factorization is a mechanism for separating the effects of long-distance and short-distance forces when calculating the decay rates of “heavy” hadrons, and it will be used when analyzing the CP asymmetries for the decays $B_d^0 \rightarrow D^+ \pi^-$ and $B_s^0 \rightarrow D_s^+ K^-$ in Chapter 5. As an example, consider the decay rate amplitude

$$A\left(\overline{B}_d^0 \rightarrow D^+ \pi^-\right) = \frac{G_F}{\sqrt{2}} \overline{V} \overline{M}, \quad (1.61)$$

where $\overline{V}$ is the CKM factor for the decay in question, in this case $V_{ud}^* V_{cb}$. The hadronic matrix element for this decay can be written as

$$\overline{M} = \left\langle D^+ \pi^- \left| \overline{Q}_1 C_1(\mu) + \overline{Q}_2 C_2(\mu) \right| \overline{B}_d^0 \right\rangle, \quad (1.62)$$

where $\overline{Q}_{1,2}$ are the current-current operators parameterizing the weak (long range) interactions in this decay:

$$\overline{Q}_1 = (\overline{d}_\alpha u_\beta)_{V-A} (\overline{c}_\beta b_\alpha)_{V-A} \quad (1.63)$$

$$\overline{Q}_2 = (\overline{d}_\alpha u_\alpha)_{V-A} (\overline{c}_\beta b_\beta)_{V-A} \quad (1.64)$$

while the corresponding Wilson coefficients $C_{1,2}$ parametrize the strong (short range) interactions in the decay. In order to calculate this matrix element, it is helpful to separate the short-range interactions, which can be calculated using perturbative methods, from the long-range interactions which cannot.

Considered at the quark level, the decay $\overline{B}_d^0 \rightarrow D^+ \pi^-$ consists of the decay $b \rightarrow c(d\overline{u})$, where the c quark combines with the spectator $\overline{d}$ to form the D^+ . The $(d\overline{u})$ combination, which forms the π^- , has a low virtual mass compared to its momentum, and, because of

relativistic time-dilation, the formation time of the pion will therefore be long. The factorization argument consists of saying that, because the pion is formed a significant distance away from the original B_d^0 , it will not be influenced by the colour fields in the neighbourhood of this decay. This allows the matrix element to be rewritten as:

$$\overline{M} = \text{strong factor} \otimes \text{weak factor}, \quad (1.65)$$

where the “strong factor” depends entirely on the Wilson coefficients $C_{1,2}$, while the “weak factor” depends entirely on the current-current operators $Q_{1,2}$. More detailed discussion of factorization can be found in [Bjo89, Ben01], and a proof of factorization for the decay $\overline{B}_d^0 \rightarrow D^+ \pi^-$, in the limit where the heavy quark mass approaches infinity, can be found in [BPS01].

1.7 Heavy Quark Expansion (HQE)

Heavy quark expansion [Shi00, Ura98, Pet04, GOP04] is a mechanism for understanding the weak decays of heavy flavour hadrons, which relies on factorization arguments in the form of the Wilson operator product expansion (OPE). HQE makes precise predictions of the ratios of B hadron lifetimes, which can be subjected to precision tests at LHCb. This section will give a brief overview of HQE/OPE, together with the current status of lifetime measurements in the B hadron sector, and the motivation for measuring B hadron lifetimes at LHCb.

1.7.1 Operator product expansion

The operator product expansion (OPE) refers to a way of writing the effective weak Hamiltonian as a series of “effective vertices” multiplied by “effective coupling constants”. It is an application of factorization, since the aim is again to separate short-distance and long-distance contributions to the Hamiltonian. The “effective vertices” are represented by operators Q_i , such as the current-current operators $Q_{1,2}$ encountered above, while the “effective coupling constants” are just the Wilson coefficients C_i . The Hamiltonian can then be written as:

$$\mathcal{H}_{eff} = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i C_i(\mu) Q_i. \quad (1.66)$$

The couplings C_i are a function of the energy scale μ , and this scale is chosen so as to separate the short-distance effects (at an energy $E > \mu$), and the long-distance effects (at an energy $E < \mu$). When studying B hadron decays, it is conventional to choose μ to be of the same order as the mass of the decaying hadron. This means that the decay rate $A(M \rightarrow F)$, which is given by:

$$A(M \rightarrow F) = \langle F | \mathcal{H}_{eff} | M \rangle = \left\langle F \left| \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i C_i(\mu) Q_i(\mu) \right| M \right\rangle, \quad (1.67)$$

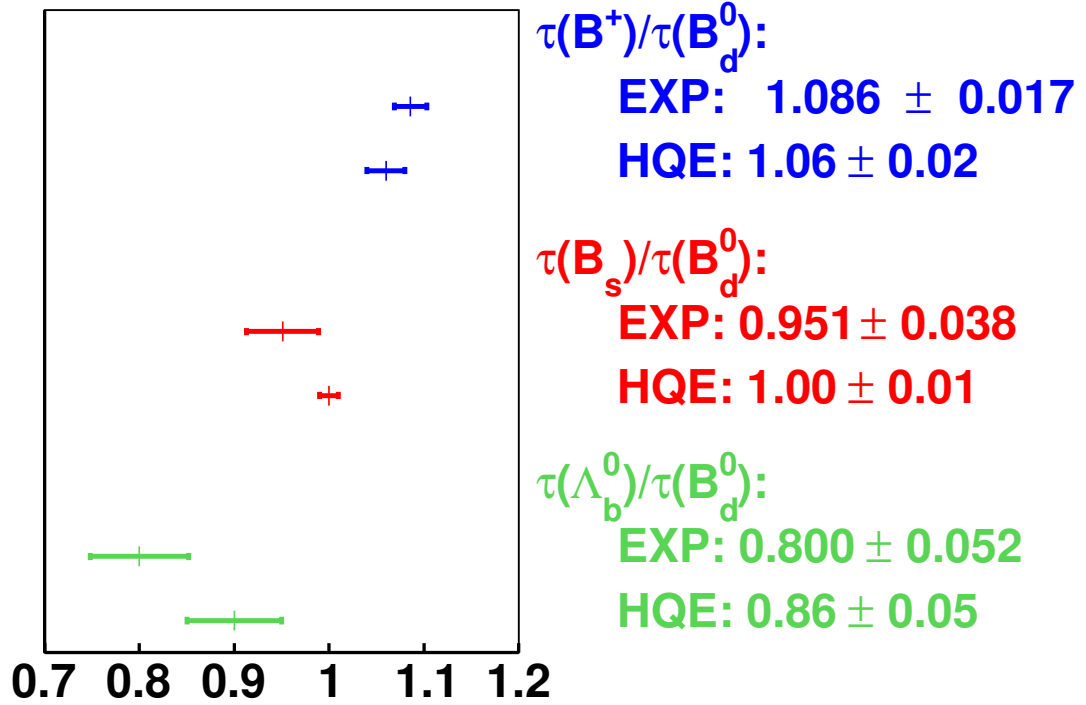


Figure 1.9: Current status of experimental and HQE ratios of B hadron lifetimes.

can be rewritten as:

$$A(M \rightarrow F) = \langle F | \mathcal{H}_{eff} | M \rangle = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i C_i(\mu) \langle F | Q_i(\mu) | M \rangle. \quad (1.68)$$

The major advantage of the rewritten formula is that it separates the calculation of $C_i(\mu)$, which can be done by perturbative methods, from that of $Q_i(\mu)$, which cannot be easily estimated theoretically. A good review of OPE and its application to HQE can be found in [Bur00].

1.7.2 Current status of lifetime measurements

Although the terms $Q_i(\mu)$ cannot be easily calculated, it turns out that they cancel when the **ratios** of B hadron lifetimes are considered. HQE makes predictions about these ratios by computing the widths of the hadron states using the above OPE in powers of $1/m$, where m is the mass of the “heavy” quark in question. Usually this has been taken to be the bottom quark, since the charm quark is not considered to be “heavy” from a theoretical point of view. A comparison of the current state of experimental results and HQE predictions is shown in Figure 1.9. In particular, the ratio of the Λ_b and B_d lifetime is relatively poorly measured, and shows the potential for disagreement with the precise HQE prediction. The ratio of the B_s^0 and B_d^0 lifetimes also offers a possibility of disagreement, and LHCb is expected to gather the necessary sample sizes to overtake the current world average on the

B_s^0 lifetime within the first few months of data taking, so that testing this ratio could be a good early measurement for LHCb. The predictions for the ratios are not unity because of next-to-leading order corrections to the OPE, which include contributions from the “light” spectator quark, as shown in [CFLM02].

In order to fully exploit its potential to constrain HQE, LHCb will have to make use of all decay channels of the B hadrons, so as to maximize the statistics available for lifetime measurements.

Chapter 2

The LHCb Detector

2.1 LHCb at the Large Hadron Collider

The Large Hadron Collider [Gro95] is a proton–proton collider which will operate at a centre-of-mass energy of 14 TeV, with a maximum luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Reusing the 27 km long LEP tunnel, LHC will house four experiments dedicated to exploring the limits of the Standard Model and possible observation of new physics: ALICE, ATLAS, CMS, and LHCb. The present timetable calls for the machine to be commissioned to the full running energy by the winter of 2008, at a luminosity of $10^{30} \text{ cm}^{-2}\text{s}^{-1}$, followed by a pilot physics run and gradual increase in the luminosity. At the maximum design luminosity, each beam will consist of 2808 bunches of $1.15 \cdot 10^{11}$ particles. The bunch crossing interval will be 25 ns.

LHCb is a dedicated B physics experiment, and will aim to exploit the unprecedented quantity of b hadrons produced at the LHC. Thanks to the $b\bar{b}$ cross section of $\sim 500 \mu\text{b}$, LHCb will have yields that are orders of magnitude higher than previous B physics experiments, giving it unique access to all flavours of B mesons and baryons. The intense operating environment of the LHC places special demands on the radiation hardness of the detector and triggering performance, which had to be addressed during the development of LHCb.

The present chapter will give an overview of those aspects of LHCb most relevant to the studies presented in the remainder of this thesis. It is primarily based on a series of technical design reports published by the LHCb collaboration, namely the Reoptimized Detector Technical Design Report [col03a], the RICH TDR [col00b], the Trigger TDR [col03b], the Magnet TDR [col00a], and the VELO TDR [col01]. All data and figures presented should be assumed to come from these documents unless otherwise referenced. As the detector design and performance have continually evolved during its development, the figures quoted in this chapter do not necessarily correspond exactly to the current state of the detector. They are, however, representative of it.

A diagram of LHCb, on which the coordinate system is labelled, can be seen in Figure 2.1. The interaction point is at the origin of this coordinate system. It is a single arm spectrometer, with an angular acceptance of 250 mrad in the horizontal (non-bending) plane, and 300 mrad in the vertical (bending) plane. LHCb is designed to run at a luminosity of

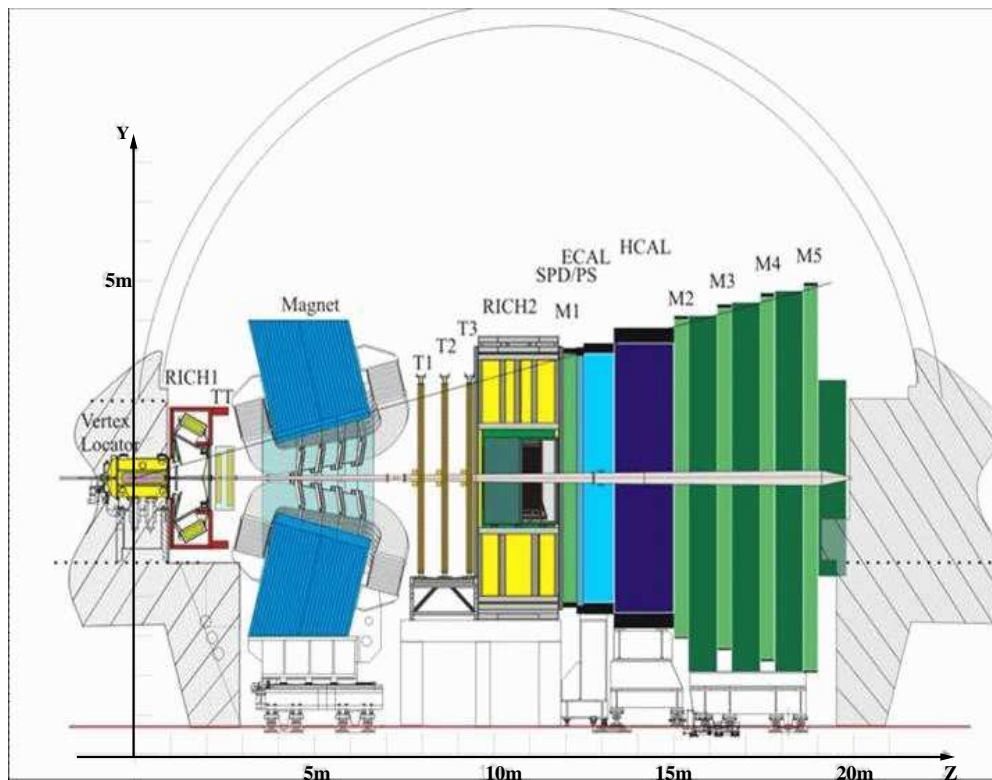


Figure 2.1: The LHCb detector.

$10^{32} \text{ cm}^{-2}\text{s}^{-1}$, 100 times below the LHC's design luminosity, and will consequently operate at its full design specification before either ATLAS or CMS.

2.2 LHCb design

$b\bar{b}$ production at the LHC

Proton-proton collisions at the LHC produce $b\bar{b}$ pairs through parton interactions. At LHC energies, the parton distribution functions of the protons are such that the interacting partons will have unequal energies. This means that the produced B hadrons will be highly boosted, and that the $b\bar{b}$ pairs will be produced in the same forward or backward cone. A PYTHIA simulation of the polar angle distributions of $b\bar{b}$ pairs in 14 TeV collisions is shown in Figure 2.2. This characteristic allows the single arm design of LHCb, since in the majority of events, both B hadrons will be travelling in the same z direction, and both will be within the detector acceptance. Another important advantage of the single arm design is that it allows excellent coverage of events close to the beamline. Figure 2.3 shows the momentum distribution of B mesons from $B_d^0 \rightarrow D^+\pi^-\pi^-$ decays. At a mean momentum of 120 GeV, a B with mean lifetime of 1.5 ps will have a decay length of ~ 1 cm. With suitable tracking and vertex detection technology, the secondary vertex associated with the b hadron can be clearly distinguished from the primary vertex, allowing b events to be identified and the proper lifetime to be measured.

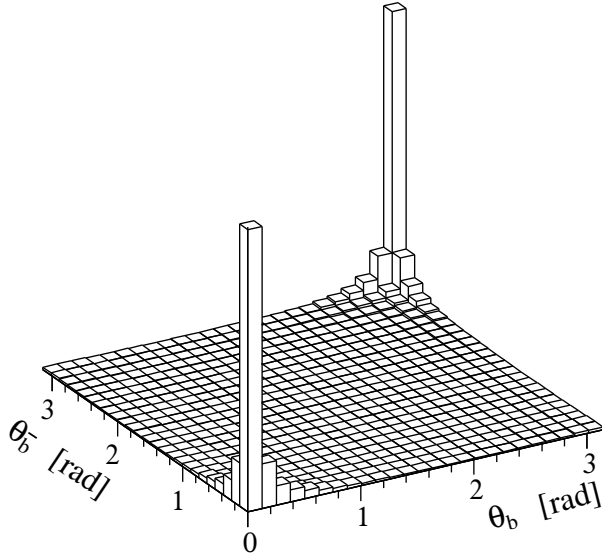


Figure 2.2: Polar angles of hadrons formed from $b\bar{b}$ pairs calculated by PYTHIA.

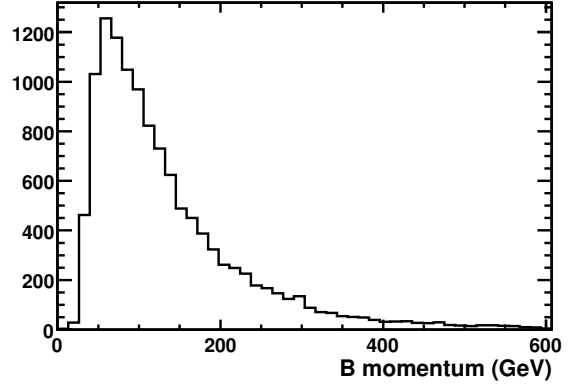


Figure 2.3: Momenta of B mesons decaying through the channel $B_d^0 \rightarrow D^+\pi^-$.

2.2.2 Design luminosity of LHCb

Proton proton collisions are inherently “dirty”, producing not only the particles of interest (in this case, B hadrons), but also many others within each event. For example, at the LHCb design luminosity, there will be an average of 72 charged tracks per event, of which only 4–8 will come from the B decays of interest. In order to facilitate the precise reconstruction of event parameters such as the B decay lifetime, where a decay vertex must be unambiguously associated with the correct production vertex, it is desirable for each event to involve only one proton-proton interaction. Figure 2.4 shows that already at a luminosity of $10^{33} \text{ cm}^{-2}\text{s}^{-1}$, the majority of events will contain more than one interaction per event, and this proportion will be even greater at the maximum LHC luminosity. Another disadvantage of a higher luminosity is the significantly higher radiation damage to the vertex detector, which would necessitate its frequent replacement. An operational luminosity of $10^{32} \text{ cm}^{-2}\text{s}^{-1}$ is chosen for these reasons, yielding 10^{12} $b\bar{b}$ pairs per year (10^7 s) of running.

2.3 Vertex detection

Almost all b physics analyses require the reconstruction of both the primary (production) and secondary (decay) vertices. As well as being important in background suppression, precise vertex reconstruction allows the proper time of the b hadron to be reconstructed. This is a necessary observable in many measurements, such as time-dependent analyses of CP violation, as well as being an interesting measurement in its own right. The vertex reconstruction is performed by software, using information from the Vertex Locator (VELO), a silicon strip detector lying close to the beamline. The impact parameter of a track with respect to the primary vertex, defined as the perpendicular distance between this vertex and the track, is an important cut variable in the LHCb trigger, and the resolution on this

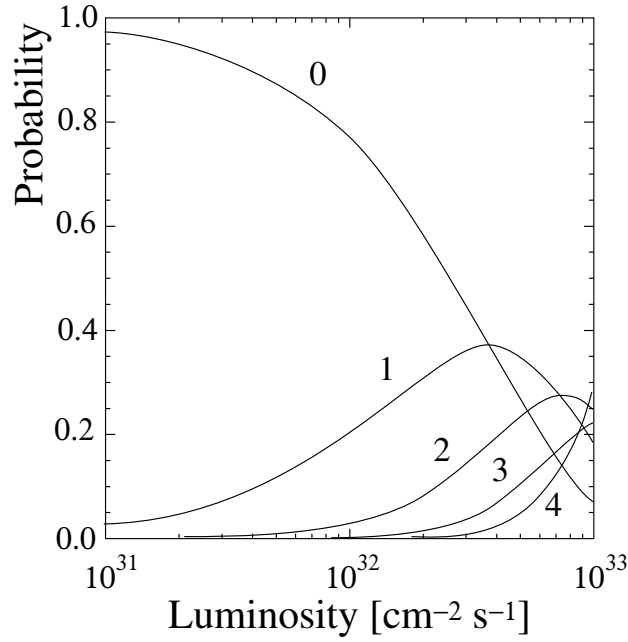


Figure 2.4: Probabilities for having a number of inelastic pp interactions at a given luminosity, for a total inelastic cross section of $500 \mu\text{b}$.

variable is determined by the VELO. Parts of the VELO serve as a dedicated pile-up veto, which rejects events with multiple primary vertices at the trigger level. On the one hand, such events enhance the statistical reach of LHCb, while on the other hand, they are more difficult to reconstruct accurately and worsen signal purity. Whether or not this veto will be used when taking data is still a matter of debate within LHCb.

2.3.1 Velo sensor layout

A cutaway of the VELO sub-detector is shown in Figure 2.5. It consists of two halves in the y plane, each containing 21 semi-circular sensor stations, and two additional stations reserved for use by the pile-up VETO. The z location of each hit is defined by identifying which station registered it, and $z = 0$ is defined as the nominal beam crossing point. The first station is at $z = -17.5$ cm, and the last at $z = 75$ cm, where the upstream stations are necessary for proper coverage of the interaction region and primary vertex reconstruction, since the σ_z of the LHC beam interaction point is ~ 5 cm. The number of downstream stations is chosen to allow a minimum of three VELO sensor hits to be registered for any charged track. Figure 2.6 shows the layout of the VELO sensors, and it can be seen that the two halves of the detector are staggered along the z axis, in order to provide overlapping areas of sensitivity.

The sensitive region begins at a radius of ~ 8 mm outside the beamline, and extends to a radius of ~ 42 mm. Although the LHC beam has a $\sigma_{x,y}$ of $100 \mu\text{m}$ during physics running conditions, its aperture increases during the injection phase, when the VELO must be retracted by 3 cm from the beamline. For this reason, the two detector halves are designed to maneuver independently of each other. The corrugated structure of the detector RF foils

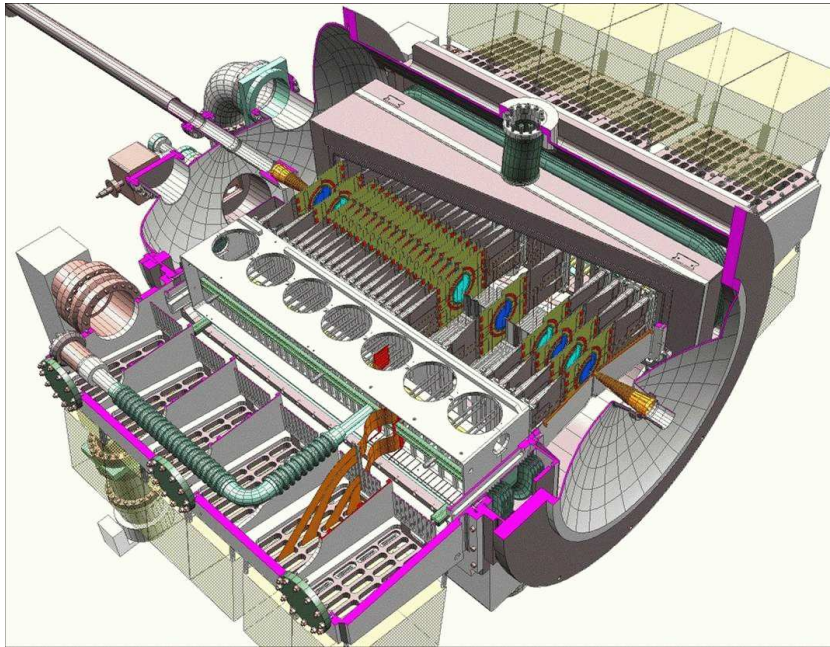


Figure 2.5: The VELO subdetector, showing the vacuum vessel containing the sensors in the centre, and readout electronics on either side of the beamline.

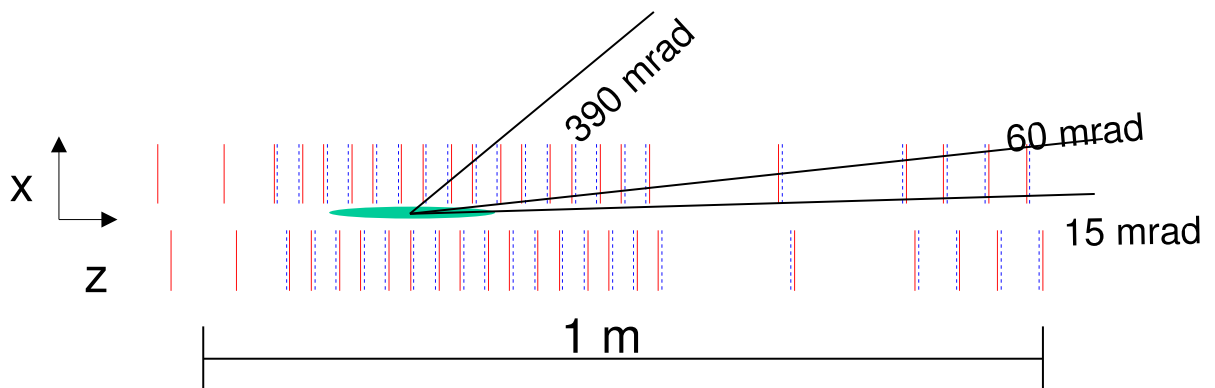


Figure 2.6: The layout of the VELO sensor stations in the x - z plane. The red lines are the R sensors, the blue lines are the ϕ sensors.

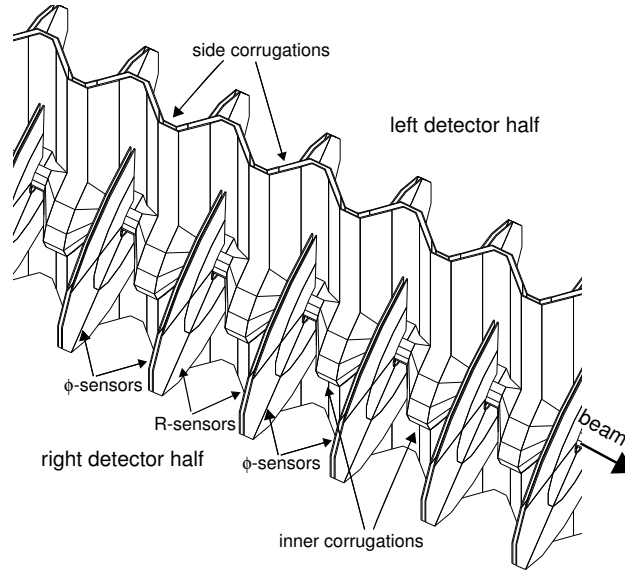


Figure 2.7: A close up of the RF foil enclosing the VELO's secondary vacuum.

allows for an overlap between the two halves, which helps with the detector alignment in the wake of a retraction and subsequent re-positioning.

2.3.2 Integration with the LHC vacuum

For optimal performance, it is desirable that the VELO sensors are as close as possible to the beamline, and that they have the smallest possible amount of material in the detector acceptance. This is best achieved by operating the sensors inside the beampipe, and hence the entire VELO is enclosed in a vacuum vessel, as seen in Figure 2.5. In order to protect the primary LHC vacuum, the sensors (seen in the middle of the figure) are themselves enclosed in a thin walled, aluminium, secondary vacuum container. This vessel also serves as a Radio-Frequency (RF) shield, protecting the detector modules from the high frequency fields generated by the LHC beams. Figure 2.7 shows a close-up of this shield, whose shape (in particular, the corrugations) is motivated by the need to provide protection to the primary LHC vacuum, while at the same time adding the minimum amount of necessary material to the detector acceptance. In addition to being thick enough to afford the requisite RF protection, the shield is also made thick enough so that the pressure differences between the primary and secondary vacuums don't cause inelastic deformations of the secondary vacuum vessel.

2.3.3 VELO sensor geometry

Each of the VELO sensor stations contains two sets of silicon strips: one for measuring in the R (radial) direction, and one for measuring in the ϕ direction. The R strips are subdivided into four 45° sectors each, providing an approximate ϕ measurement, while the ϕ strips similarly have a stereo angle of $10^\circ - 20^\circ$ which provides a limited measurement

capacity in the R direction. Along with a 2.6 mm overlap between R sensors in opposing VELO halves, these features help reduce ambiguities in the tracking. The strips can be read out independently of each other, an important feature in the trigger since it allows a reconstruction of tracks to be carried out in the $\mathbf{r-z}$ plane, with limited ϕ information from the ID of the R strip sector hit. This “2D” reconstruction is considerably quicker than if the ϕ sensors had to be read out and used in the reconstruction algorithm; it is used in the trigger to select tracks which will be reconstructed in 3D later on.

2.3.4 Sensor design and readout

Because of its proximity to the beamline, the VELO is particularly vulnerable to damage from the high radiation levels expected at the LHC. It is therefore important that the VELO sensors maintain their performance and signal-to-noise (S/N) ratios even after being exposed to a substantial amount of radiation, and this requirement influenced two key design choices in the VELO sensors: the choice of n-on-n silicon technology, and the choice of a thick silicon sensor with a double metal layer.

Radiation damage results in the underdepletion of silicon sensors, and Figure 2.8 shows the effect of underdepletion on a p-on-n sensor with and without a double metal layer. Because the irradiated sensor depletes from the n side, the underdepleted region acts as an insulator, causing the field lines to widen and the signal to spread over a greater range of strips. The addition of a double metal layer can cause an additional charge loss in the underdepleted case, further degrading resolution. On the other hand, a double metal layer is desirable because it allows the readout geometry to be decoupled from the sensor geometry, and hence allows the readout electronics to be placed out of the detector acceptance, reducing their vulnerability to radiation damage. Unlike a p-on-n sensor, an n-on-n sensor does not suffer from these problems, because the depleted layer is on the same side as the strips and hence does not act as an insulator. Table 2.1 shows the performance tests performed on the PR01 (n-on-n) and PR02 (p-on-n) prototypes, and it can be seen that after a dose corresponding to two years of LHCb running, the n-on-n prototype has a significantly lower depletion and higher efficiency than the p-on-n prototype.

An analogue readout is chosen for the VELO sensors because it improves resolution, as seen in Figure 2.9. A further benefit of this choice is that the radiation damage to the VELO is expected to be highly non-uniform, and this will be easier to monitor with an analogue readout. The readout system accepts data at the LHCb crossing rate of 40 MHz, and holds it in an analogue pipeline for 4 μs while waiting for the L0 decision. The read out of the data is performed at an average rate of 1 MHz, and sent via twisted pair cable to the software trigger systems in a radiation-free room 60 m away. The LHCb trigger system will be discussed further in Section 2.7.

The annual radiation dose will vary between the VELO stations, with the upstream stations receiving a higher amount of radiation than downstream stations. The expected annual dose for station 7 is $1.3 \cdot 10^{14} \text{ } \eta_{eq}\text{cm}^{-2}$, and a bias voltage of 400 V will be required to fully deplete the irradiated sensors after two years, while a voltage of 600 V will be required after three years. Since the maximum bias voltage that can be applied to the chip

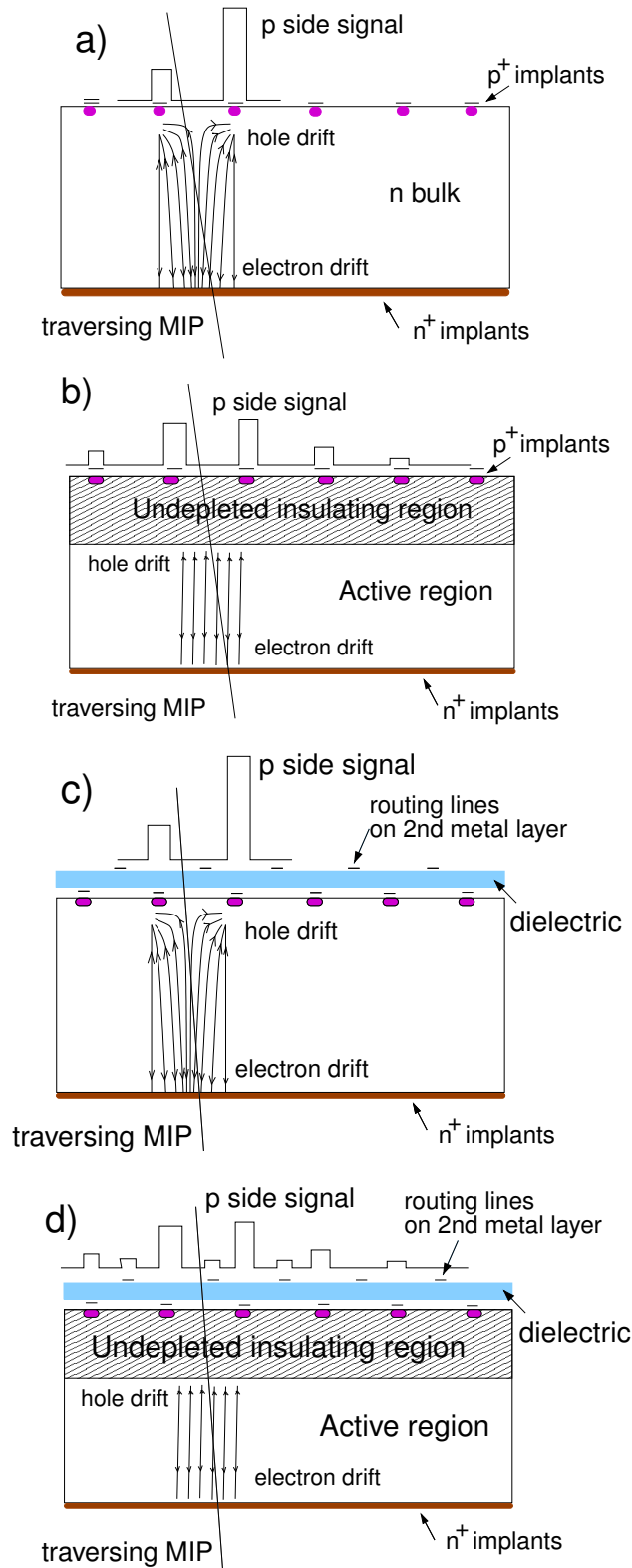


Figure 2.8: Cluster shapes for the depleted and underdepleted cases for a simple segmented p-on-n sensor, (a) and (b), and for a sensor with a double metal layer, (c) and (d). At full depletion the charge is focused on the diodes, while at underdepletion the clusters spread (b) and lose charge to the double metal layer (d). Figure and text of caption reproduced from [col01].

	PR01- ϕ	PR02- ϕ	PR02- ϕ
Name	h-14- ϕ	1832-9-a	1968-17-c
Thickness (μm)	300	200	300
Oxygenated	No	No	$2.5 \times 10^{17} \text{ atoms/cm}^3$
Electronics	SCT128A	SCT128A	SCT128A
S/N for a MIP	58:2.7	30:2.0	—
V_{dep} after $2 \times 10^{14} p/\text{cm}^2$	50 V	300 V	150 V
Efficiency at $V_{\text{bias}} = 200 \text{ V}$ after $2 \times 10^{14} p/\text{cm}^2$	99%	75%	—

Table 2.1: Performance numbers for the irradiated PR01 (n-on-n) and PR02 (p-on-n) prototypes. The S/N is given in ADC counts, where one count is equivalent to 380 electrons. Reproduced from [col01].

is 500 V, and the prototype n-on-n sensors showed no significant performance loss even at 40% underdepletion, the VELO is expected to have a lifetime of 3-4 years at the nominal LHCb running conditions.

2.3.5 Vertex resolutions

Figures 2.10 and 2.11 show the x ($9 \mu\text{m}$) and z ($47 \mu\text{m}$) resolutions on the primary vertex, and Figures 2.12 and 2.13 show the same ($\sigma_x = 17 \mu\text{m}$, $\sigma_z = 230 \mu\text{m}$) on the B_d^0 vertex in reconstructed signal decays of the type $B_d^0 \rightarrow D^+ \pi^-$. Figure 3.8 shows the resolution on the B_d^0 lifetime in the same channel. A double Gaussian fit results in a core width of 34 fs, or $\sim 2\%$ of the mean B_d^0 lifetime. The resolution on the impact parameters of tracks is shown in Figure 2.19.

2.3.6 Pile up veto

The two VELO stations located furthest upstream are used for the pile up veto counter. These stations only have R sensors, which are used to estimate the position of the primary vertices in the event. The information is fed into the L0 trigger, which then rejects events with multiple primary vertices. Simulation studies show that this system rejects 80% of events with two primary vertices, and correctly identifies 90% of events with only one primary vertex.

2.4 Tracking system

The LHCb tracking system is used together with the VELO detector to reconstruct charged particle tracks within the detector acceptance. Figure 2.14 shows a diagram of the tracking system together with magnetic flux density at the corresponding point in the detector. The tracking system consists of four parts: the magnet, the trigger tracker (TT), the inner tracker

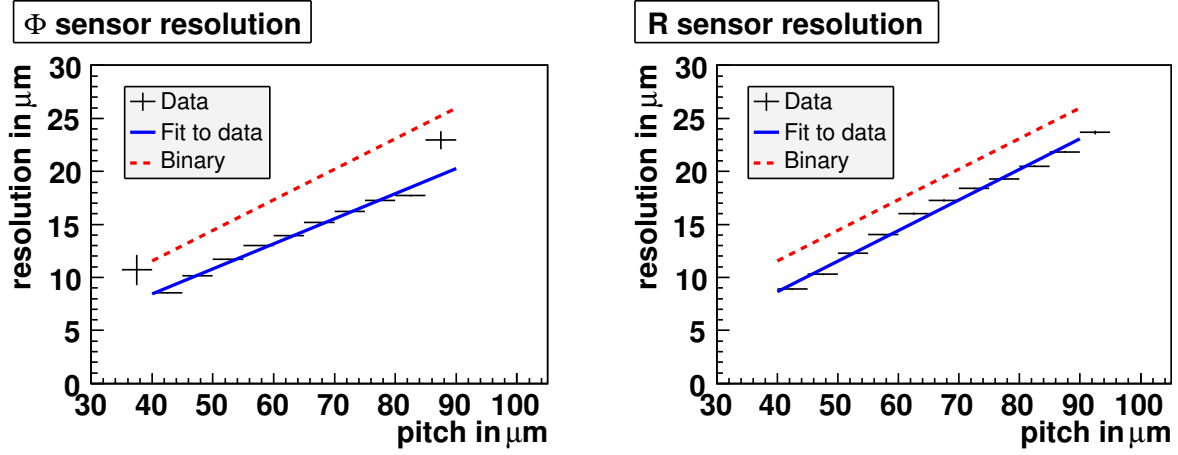


Figure 2.9: The single hit resolutions for VELO R and ϕ sensors. Binary refers to the resolution that would be achieved if the only information read out was whether or not the sensor registered a hit. The blue fit is made with data from the analog readout.

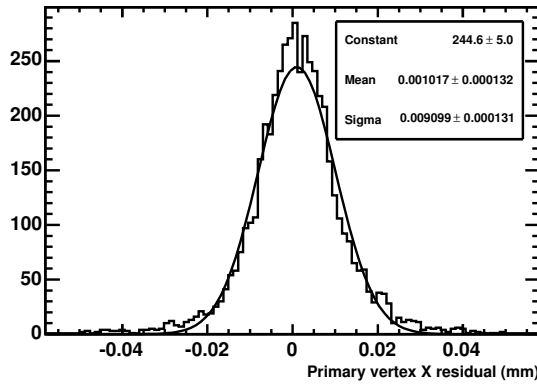


Figure 2.10: The VELO primary vertex x resolution.

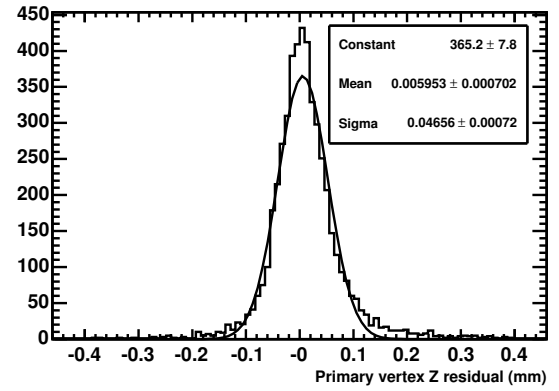


Figure 2.11: The VELO primary vertex z resolution.

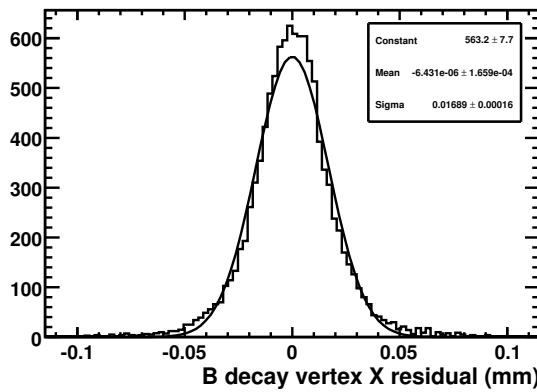


Figure 2.12: The VELO resolution on the B decay vertex in the x direction.

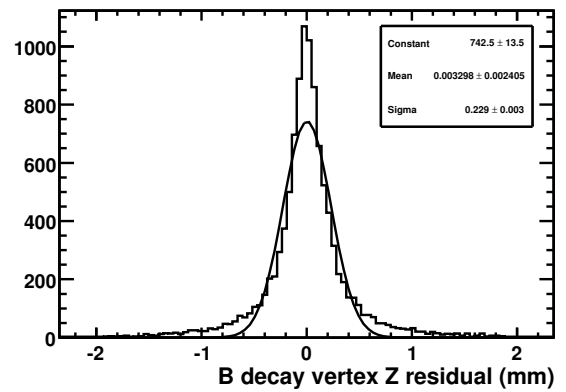


Figure 2.13: The VELO resolution on the B decay vertex in the z direction.

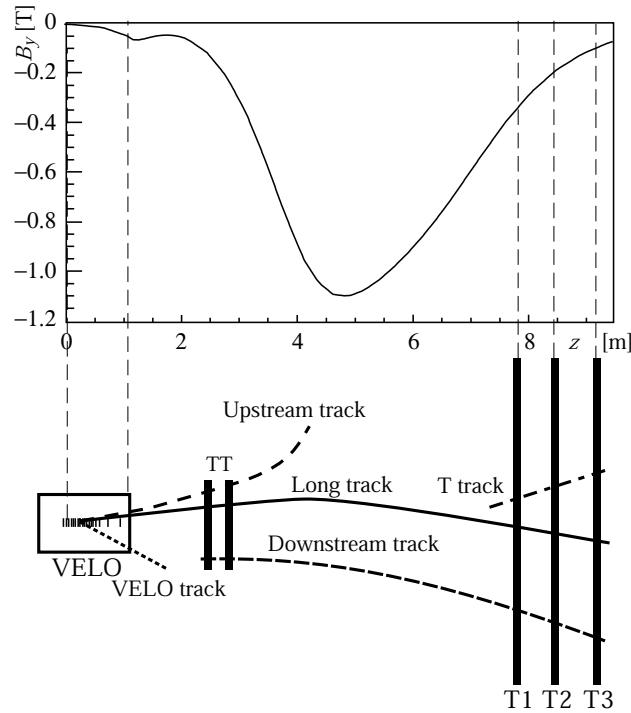


Figure 2.14: A schematic of the LHCb tracking system, showing the VELO, trigger tracker, T1-T3 tracking stations, and the different types of tracks reconstructed at LHCb. The LHCb B field in the y direction, at each point in z , is shown for reference.

(IT), and the outer tracker (OT).

2.4.1 The LHCb magnet

The design of the LHCb magnet was driven by the requirement to provide a momentum resolution of approximately 0.4% for momenta up to 200 GeV, which in turn necessitates an integrated field of 4 Tm for tracks originating at the interaction point. A warm dipole magnet design was chosen as cheaper and easier to implement than a superconducting magnet. In addition, it will be desirable to regularly reverse the magnetic field, in order to minimize systematic errors caused by asymmetries in the detector construction, as well as an inhomogeneous magnetic field, and this is easier with a warm magnet. A picture of the magnet can be seen in Figure 2.15. The pole faces are curved to follow the experimental acceptance in order to reduce the power requirements, but this is not found to significantly impact physics performance. The power required to operate the magnet is 4.2 MW, and the water cooling system will operate at a flow of $125 \text{ m}^3\text{h}^{-1}$.

2.4.2 The trigger tracker (TT)

The TT is a silicon strip detector located between the RICH 1 and the LHCb magnet. Together with the VELO it allows the momentum of tracks used in the L1 trigger to be measured. It also allows the reconstruction of neutral particles that decay outside the fiducial

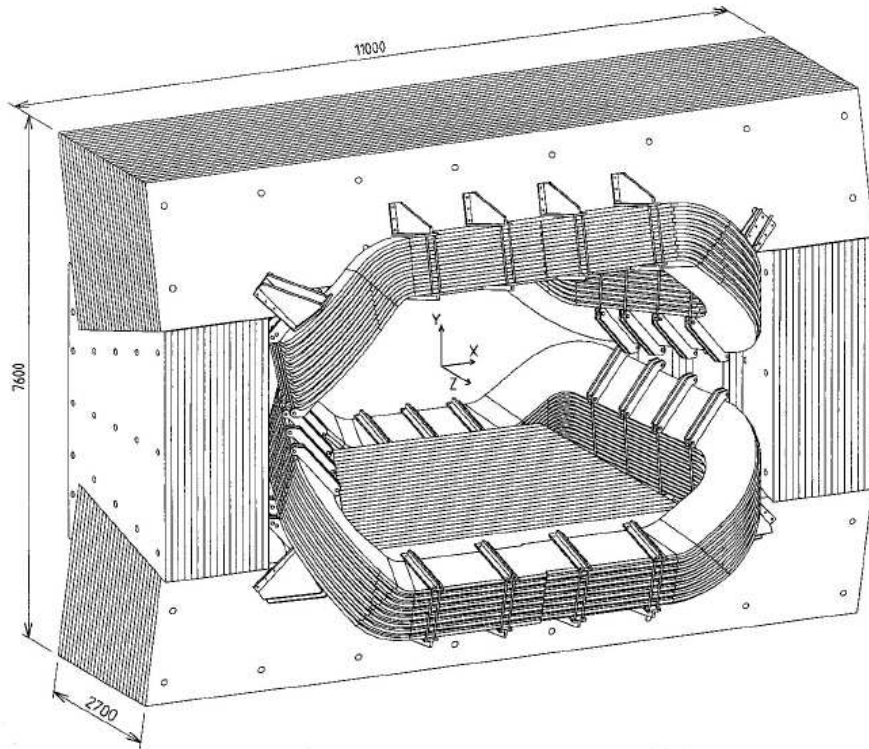


Figure 2.15: A schematic of the LHCb magnet. The quoted lengths are in mm.

volume of the VELO but inside the tracking volume (for example, the K_S), and allows the reconstruction of low momentum tracks which are bent out of the detector acceptance by the magnet before they traverse the later tracking systems. The TT subdetector has four detection layers, of which the first and last have vertical readout strips, while the middle detection layers have readout strips rotated by a stereo angle of $+5^\circ$ and -5° respectively, allowing the transverse component of the momentum to be determined.

2.4.3 The inner tracker (IT)

The IT is a silicon strip detector covering the 1.3% of the LHCb acceptance closest to the beamline, which contains 20% of all charged particles produced in the LHCb acceptance. This area requires special treatment because it has particle fluxes approximately 3 orders of magnitude greater than the rest of the acceptance. The inner tracker is divided into three tracking stations located downstream of the LHCb magnet and labeled T1-T3 in Figure 2.14. Each station has four detector boxes, located above, below, and to each side of the beamline, and each detector box has four detection layers arranged in the same manner as the layers in the TT. Figures 2.16 and 2.17 show front and side views of a tracking station, where the outer tracker elements are blue (on the outside in the front profile) and the inner tracker elements are orange.

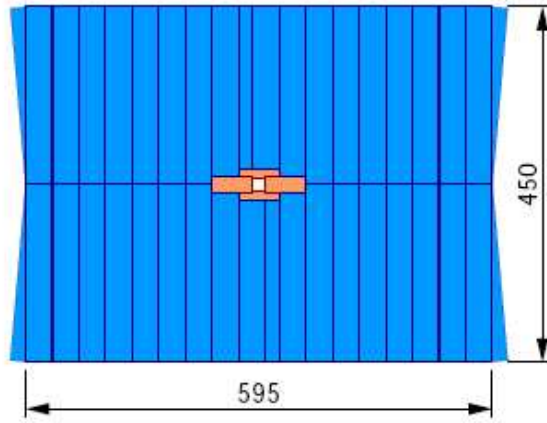


Figure 2.16: Front view of a tracker station, with dimensions in cm.

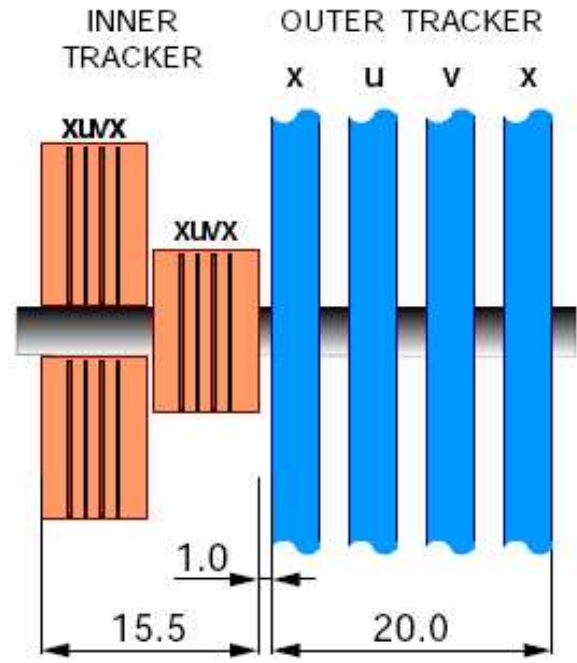


Figure 2.17: Side view of a tracker station, with dimensions in cm. The stereo detector layers are labelled u and v .

2.4.4 The outer tracker (OT)

The OT is a straw tube detector which complements the IT by covering the remaining LHCb acceptance. It is similarly laid out, with each of the three tracking stations having four straw tube OT layers associated with it, arranged in the same manner (vertical, stereo, stereo, vertical) as those of the TT. In order for the pattern recognition to meet performance requirements, the drift time across the straw tubes had to be under 50 ns, corresponding to two bunch crossings. For this reason, a diameter of 5 mm was chosen for the straw tubes, which are mounted at a pitch of 5.25 mm. A gas mixture of $Ar/CF_4/CO_2$ was chosen, in line with the choice made by HERA-B and COMPASS, in order to minimize the risk of cathode surface degradation and anode wire ageing which can occur in the presence of fluorine radicals and hydrogen, for example with a mixture such as $Ar/CF_4/CH_4$.

2.4.5 Tracker occupancy

The fraction of channels in a detector region hit within the event time window, in this case 50 ns, is defined as the detector occupancy. This is a vital index of detector performance because higher occupancies make pattern recognition more difficult and increase the rate of ghost¹ tracks. The occupancy depends on the density of tracks and the detector granularity, where the density of tracks in the plane perpendicular to the beam scales approximately as

¹A ghost is a falsely reconstructed track which had no corresponding particle, e.g. when hits deposited by two or more particles are combined into a single track, or when a track is formed from background hits.

r^{-2} , and r is the distance to the beam line. For this reason, the tracking system consists of a fine-grained inner tracker in the region where the track density is highest, and the more coarsely grained outer tracker in the lower density region. Simulations show that for acceptable pattern recognition performance, an occupancy $< 15\%$ is required in the “hottest” parts of the tracking system. This is another feature limiting the luminosity at which LHCb can run, since higher luminosities correspond to higher numbers of tracks produced at the interaction point.

2.4.6 Reconstruction and tracking performance

Referring back to Figure 2.14, tracks reconstructed at LHCb are classified into five different types:

- **Long tracks** traverse all detector subsystems from the VELO to the T3 tracking station. They form the basis of all LHCb analyses.
- **VTT tracks**, also known as upstream tracks, exit the detector acceptance after the TT tracking station. Although they are not reconstructed with the same quality as long tracks, VTT tracks give information on low momentum particles upstream of the magnet, which is useful for calibrating the RICH 1 detector.
- **Downstream tracks** are first detected at the TT tracking station, and are used to reconstruct K_S^0 and Λ particles.
- **VELO tracks**, which exit the detector acceptance after the VELO, are used in primary vertex reconstruction.
- **T tracks** are first detected at the T1 tracking station, and are useful in understanding the contribution of downstream secondary particles in the RICH2 particle reconstruction.

Long tracks have the highest reconstruction quality and momentum resolution, and most physics analyses are performed using only long tracks. Figure 2.18 shows the long track finding performance of the tracking system. Efficiency is seen to rise rapidly with momentum and exceed 90% for momenta above 10 GeV, while ghosts can be reduced sharply by applying a cut on the momentum and transverse momentum of the reconstructed tracks. Figure 2.19 shows the momentum and impact parameter resolutions for long tracks, and Figure 2.20 shows the momentum resolution and impact parameter precision for all B-decay tracks. The average momentum resolution of the tracks is 0.37%, while the average impact parameter resolution of the tracks is 40 μm .

The track reconstruction strategy at LHCb is based on the principle of “track seeds”, formed in those regions where the magnetic field is weakest, which can be reconstructed as straight line segments and combined with other detector hits. For example, long track reconstruction begins by forming the seeds from VELO hits, and combining these with all T station hits in turn. A VELO seed and a single T station hit uniquely determine the

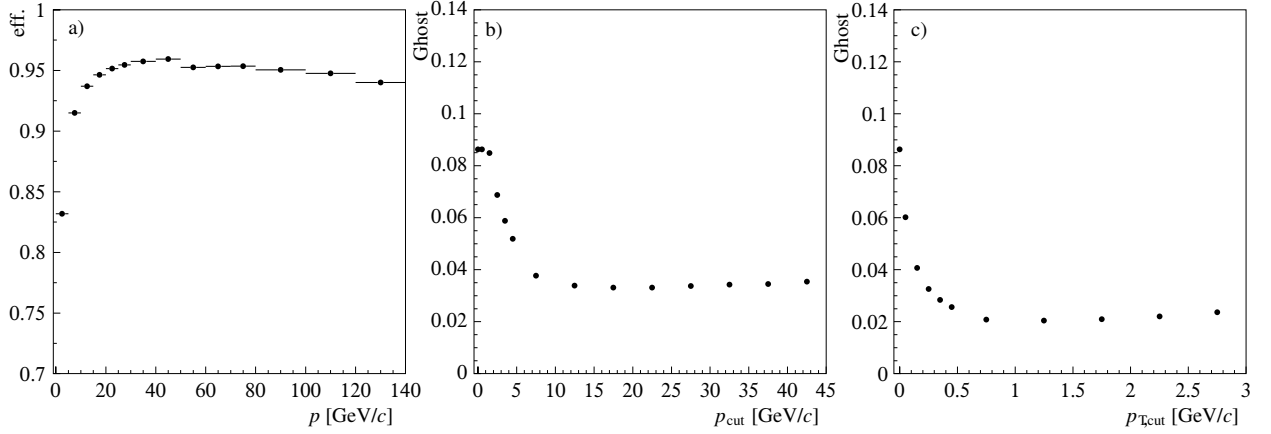


Figure 2.18: Performance of the long track reconstruction. a) shows the reconstruction efficiency with momentum, b) the ghost rate as a function of a momentum cut, and c) the ghost rate as a function of a transverse momentum cut.

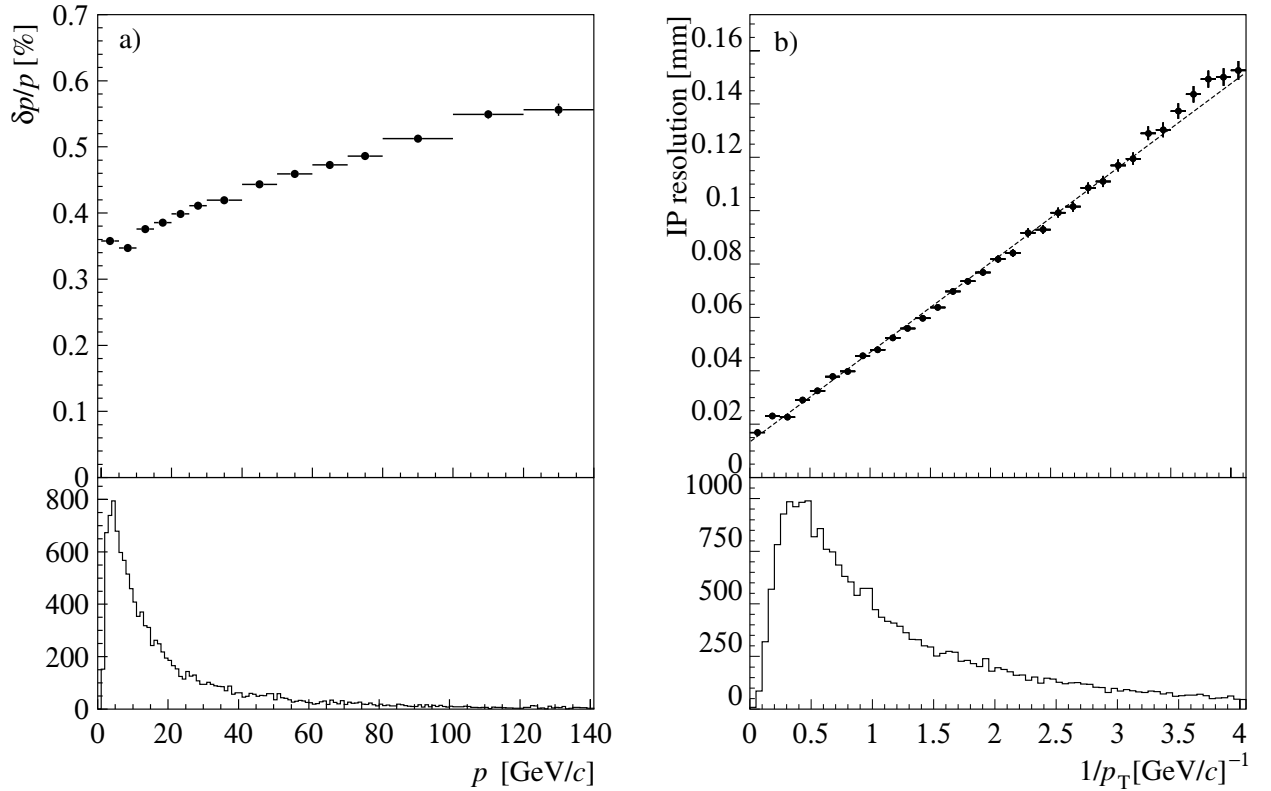


Figure 2.19: Resolution on the reconstructed track parameters at the production vertex of the track: (a) momentum resolution as a function of track momentum, (b) impact parameter resolution as a function of $1/p_t$. For comparison, the momentum and transverse-momentum spectra of B-decay particles are shown in the lower part of the plots.

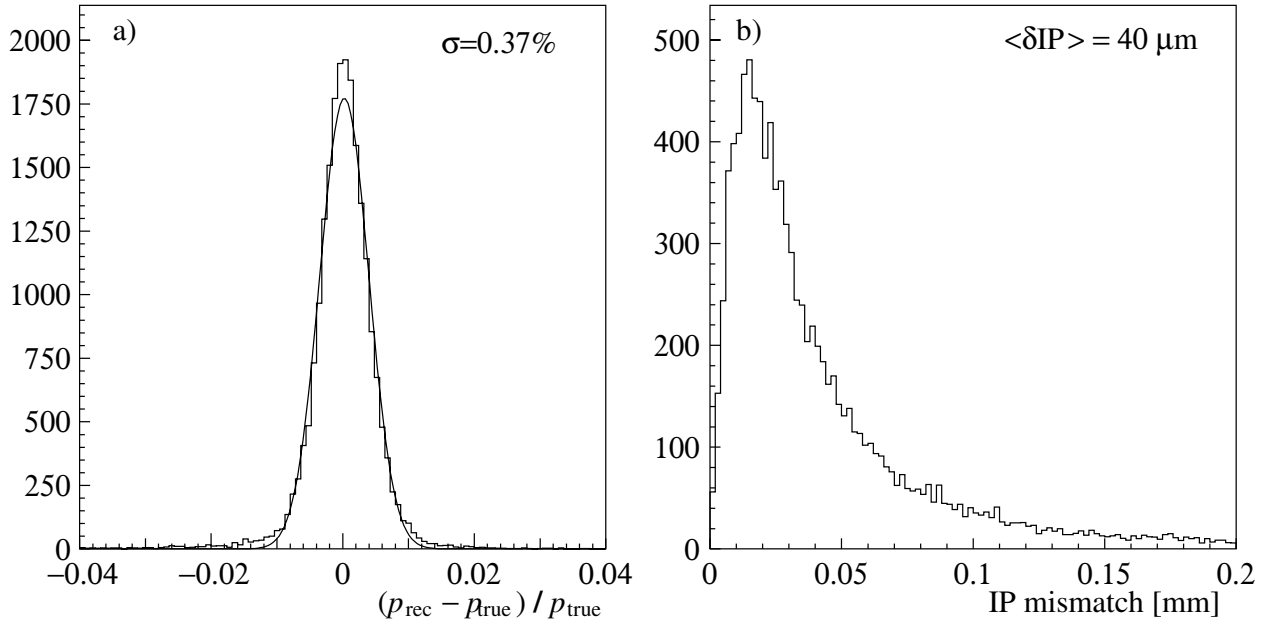


Figure 2.20: Momentum resolution and impact parameter precision for B-decay tracks. The momentum resolution is fitted with a single Gaussian.

track's trajectory through the detector, and other detector hits are then combined with this trajectory in an attempt to form a complete track. If a sufficient number of hits is found close to this trajectory, a long track is reconstructed and the hits associated with this track are not used in the reconstruction of other long tracks.

2.5 Hadron identification

In order to distinguish between topologically identical decays of B mesons, it is necessary to have a method for identifying the individual daughter particles in the decay. The canonical LHCb example is the decay $B_d^0 \rightarrow \pi\pi$, important for measuring the CKM angle γ , which must be separated from topologically identical decays such as $B_d^0 \rightarrow K\pi$. Figure 2.21 shows the reconstructed B mass for $B_d \rightarrow \pi\pi$ and its topological equivalent backgrounds—the background sits underneath the signal peak and makes disentangling the two impossible. Hadron identification allows the pions to be separated from other hadrons. A further help in distinguishing the various mass peaks is LHCb's excellent mass resolution, which is a result of the 0.37% momentum resolution seen in the previous section, with 20 MeV being a typical mass resolutions for B mesons. The isolated signal peak for $B_d^0 \rightarrow \pi\pi$, using hadron identification, is seen in Figure 2.22. The blue events on the bottom left of the signal peak are the $B_d^0 \rightarrow K\pi$ events which are misidentified as $B_d^0 \rightarrow \pi\pi$.

Hadron identification is also vital for flavour tagging, which identifies the flavour of B mesons at production. As will be discussed below, an important method of tagging relies on identifying kaons produced in the B_s^0 fragmentation chain, since their flavour will be correlated with that of the B_s^0 . Since a typical LHCb event has an order of magnitude more pions than kaons in an event, kaon tagging would be impossible without a method of

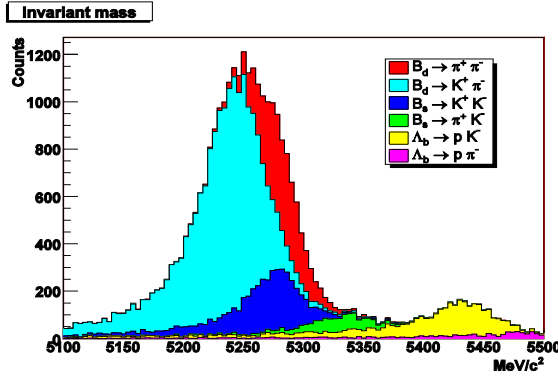


Figure 2.21: The reconstructed mass of $B_d^0 \rightarrow \pi\pi$ in the absence of hadron identification.

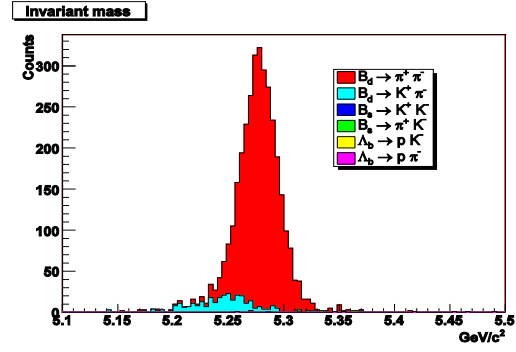


Figure 2.22: The reconstructed mass of $B_d^0 \rightarrow \pi\pi$ using hadron identification.

identifying kaon tracks from the pion background. Since the tagging kaons typically have momenta of a few GeV, it is also vital that the method of hadron identification performs well at low momenta.

2.5.1 Measuring particle masses

The identification of hadrons at LHCb is accomplished by measuring their mass, using two Ring Imaging CHerenkov (RICH) detectors. Cherenkov light is produced by particles travelling with speed β above the speed of light in a dielectric medium, and spreads out in a cone with opening angle θ_C according to:

$$\cos(\theta_C) = \frac{1}{n\beta}, \quad (2.1)$$

where n is the refractive index of the medium. In terms of the particle momentum, p and mass m :

$$\cos(\theta_C) = \frac{1}{n} \sqrt{1 + \left(\frac{m}{p}\right)^2}. \quad (2.2)$$

In principle, then, the momentum of the track and opening angle are measured, the refractive index is known, and hence the mass can be extracted. Since signal tracks at LHCb can have momenta from 1 GeV to > 100 GeV, it will be necessary to provide the required discrimination over the full momentum range, and this cannot be done with a single dielectric medium. Following the derivation in [Rad01], assume that θ_C is small and $p \gg m$:

$$1 - \frac{\theta_C^2}{2} = \frac{1}{n} \left(1 + \frac{m^2}{2p^2}\right). \quad (2.3)$$

Therefore, if $\theta_{1,2}^2$ are the Cherenkov angles of two particles with unequal masses and equal momenta:

$$\theta_1^2 - \theta_2^2 = -\frac{1}{n} \frac{\Delta(m^2)}{p^2}. \quad (2.4)$$

Rewriting

$$\Delta\theta_c \cdot (\theta_1 + \theta_2) = -\frac{1}{n} \frac{\Delta(m^2)}{p^2}, \quad (2.5)$$

and making another approximation of large momenta, using Equation 2.3,

$$(\theta_1 + \theta_2) = 2\sqrt{2(n-1)}. \quad (2.6)$$

The difference between the Cherenkov angles of two particles with unequal masses and equal momenta can therefore be written as:

$$\Delta\theta_c = -\frac{1}{2n\sqrt{2(n-1)}} \frac{\Delta(m^2)}{p^2}, \quad (2.7)$$

and the significance of this separation, n_σ , given in terms of the Gaussian error on each angle measurement, σ_θ , is

$$n_\sigma = -\frac{1}{2n\sigma_\theta\sqrt{2(n-1)}} \frac{\Delta(m^2)}{p^2}. \quad (2.8)$$

Assuming that a minimum separation of 3σ is required to separate the two tracks, it is possible to calculate the maximum momentum at which this separation is possible for a given mass difference and refractive index n . On the other hand, the minimum operating momentum is given by the threshold for Cherenkov light emission:

$$p_{\min} = \frac{m}{\sqrt{n^2 - 1}}. \quad (2.9)$$

In order to convince ourselves that more than one radiator is required, it will be instructive to calculate the momentum coverage of one of LHCb's radiators: silicon aerogel. Aerogel has a refractive index of $n = 1.03$. For a pion mass of $m_\pi = 140$ MeV, this gives a minimum operating momentum of $p_{\min} \approx 600$ MeV. For the same refractive index, a pion-kaon separation of 3σ can be achieved up to a momentum of $p_{\max} \approx \frac{400}{\sqrt{\sigma_\theta}}$ MeV. Assuming a Cherenkov ring resolution of around a twentieth of the size of each angle, $p_{\max} \approx 4$ GeV, and since the momentum range that needs to be covered is between 1 GeV and > 100 GeV, it can be seen that other radiators will be needed.

2.5.2 Radiator configuration

LHCb will use three radiators in two RICH detectors, RICH 1 and RICH 2, shown in Figures 2.23 and 2.24. RICH 1 uses the silicon aerogel radiator with refractive index of 1.03, and gaseous C_4F_{10} , with a refractive index of 1.0014, and covers the momentum range between 1 GeV and 70 GeV. For momenta > 70 GeV, RICH 2 uses a CF_4 radiator with a refractive index of 1.0005. In addition, RICH 2 is designed to have a better intrinsic resolution on the Cherenkov angles by making the detector substantially larger, which also improves high momentum performance. RICH 1 and 2 also cover different acceptances, which optimizes performance in light of the strong correlation between track momentum and polar angle at LHCb. RICH 1 covers polar angles between 25 mrad and 300 mrad, while RICH 2 covers angles between 15 mrad and 120 mrad.

2.5.3 Hybrid photon detectors (HPD)

The photodetectors used in the RICH systems are HPDs [BCF⁺01, Eas01, CFG⁺98], a type of photodetector developed by LHCb in collaboration with industry. The HPD combines a

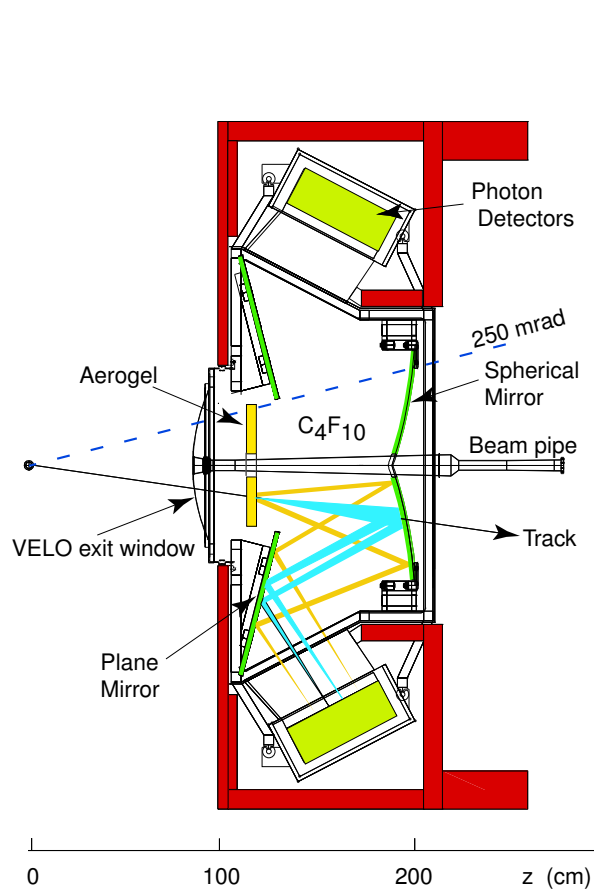


Figure 2.23: RICH 1 schematic in the x - z plane.

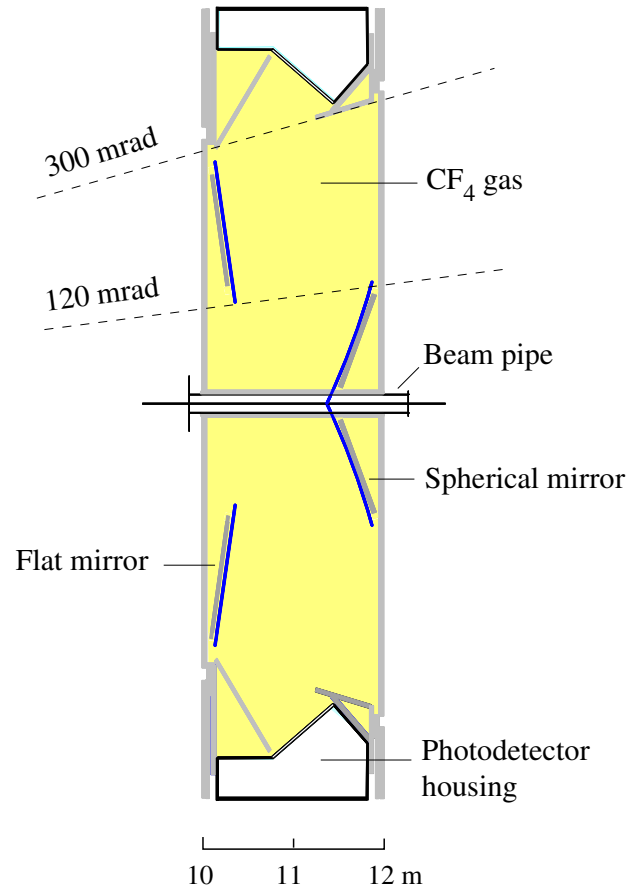


Figure 2.24: RICH 2 schematic in the x - z plane.

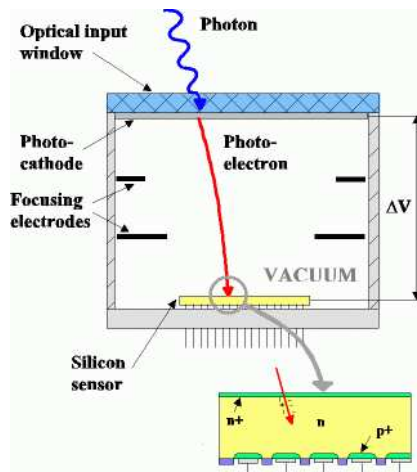


Figure 2.25: A schematic of the principle behind the HPD, with a close up of the silicon sensor in the bottom right. Picture taken from [Gys].

vacuum photo cathode with a silicon sensor to detect the Cherenkov photons created in the RICH radiators. Figure 2.25 shows a schematic of this system. A Cherenkov photon striking the photo-cathode results in the creation of a photo-electron, which is then accelerated by an applied voltage of 12 to 20 kV through a vacuum onto a reverse-biased silicon detector. The photo-electron deposits its energy onto the silicon surface, resulting in the creation of thousands of electron hole pairs and therefore an electrical signal corresponding to the incident photon energy. The focusing electrodes shown in the diagram focus the photo-electron onto the part of the silicon corresponding to the area hit on the photo-cathode, and are nearly 100% efficient. The effective pixel size of the silicon detector is 2.4 mm, and the peak quantum efficiency of 26% is achieved at a photoelectron energy of 4.5 eV.

An important measure of RICH performance is the number of photo-electrons detected per track, also referred to as the photon yield. This number will depend on HPD parameters such as:

- The anode efficiency;
- The effective input radius of the HPD;
- Mirror reflectivity.

In the latest LHCb simulation [Jon06a, Jon06b, Col08], the photon yields for the three radiators are:

- **Aerogel** – an average of 5.0 photons per track, with an RMS of 2.5;
- **C₄F₁₀** – an average of 30.3 photons per track, with an RMS of 6.1;
- **CF₄** – an average of 21.9 photons per track, with an RMS of 6.0;

where the photon yields are described by a Poisson distribution, and are valid for tracks with $\beta \sim 1$, which traverse the whole radiator. The aerogel photon yield is much lower than that of the fluoro-carbons because the aerogel is much thinner: approximately 5 cm compared to 1 m for the fluoro-carbons. The aerogel thickness was dictated by the amount of Raleigh scattering in the material, which increases with tile thickness.

2.5.4 Mirrors

As seen in the schematics of the two RICH detectors, both RICH 1 and RICH 2 will use a combination of spherical and flat mirrors to focus the Cherenkov photons onto the HPDs outside the detector acceptance. In the case of RICH 1, four spherical carbon fibre mirrors are used, each one divided into four rectangular segments. These focus the photons onto 16 rectangular plane mirrors, divided into two halves above and below the beam, which in turn direct the photon signal onto the HPD boxes. For RICH 2, the spherical mirrors are divided into 56 hexagonal mirror segments, and the flat mirrors into 40 square segments. Both types of RICH 2 mirrors are made of 6 mm thick glass, polished, and coated with

900 nm of aluminium and 200 nm of quartz for protection. The wavelength range of the Cherenkov photons in which the HPDs have sensitivity is $200 \text{ nm} < \lambda < 600 \text{ nm}$, and mirror reflectivity in this range is expected to exceed 85%.

2.5.5 Precision on the Cherenkov angle

A number of different effects contribute to the precision on the Cherenkov angle:

- **Emission point error:** because the spherical mirrors are tilted, photons emitted at different points on the track will have different image positions. The pattern recognition treats all photons as coming from the midpoint of the track in the radiator, which causes a smearing of the Cherenkov rings.
- **Chromatic dispersion:** the radiator chromatic dispersion means that the Cherenkov angle depends on the energy of the emitted photon.
- **Detector granularity:** the finite size of the photon detectors places a limit on the position resolution of hits in the detector plane, and hence the Cherenkov angle resolution.
- **Tracking:** an error in the measured track direction will lead to an error in the measured Cherenkov angle.

The single photon precisions, for each of the three radiators, are:

- **Aerogel** – 2.6 mrad;
- **C₄F₁₀** – 1.5 mrad;
- **CF₄** – 0.7 mrad.

2.5.6 Pattern recognition

A global RICH pattern recognition is performed in the reconstruction software, using information from all three radiators. Each track found in the event is assigned a particle hypothesis and Cherenkov rings are calculated for it based on this hypothesis in both the RICH 1 and 2 detectors. The likelihood of these Cherenkov rings corresponding to the actual photon hits in the HPDs is calculated, and a maximum likelihood fit is performed to determine the best hypothesis for all tracks. Figures 2.26 and 2.27 show examples of pattern recognition for RICH 1 and RICH 2 detectors respectively. In order to discriminate between different hypotheses for the same track, a log-likelihood difference (DLL) cut is used. For example, the DLL between the kaon and pion hypotheses for a track is given by:

$$\Delta \ln \Lambda_{K\pi} = \ln \Lambda_K - \ln \Lambda_\pi, \quad (2.10)$$

where $\ln \Lambda_K$ and $\ln \Lambda_\pi$ are the log-likelihoods of the kaon and pion hypothesis respectively. A full description of the LHCb pattern recognition is given in [FS98].

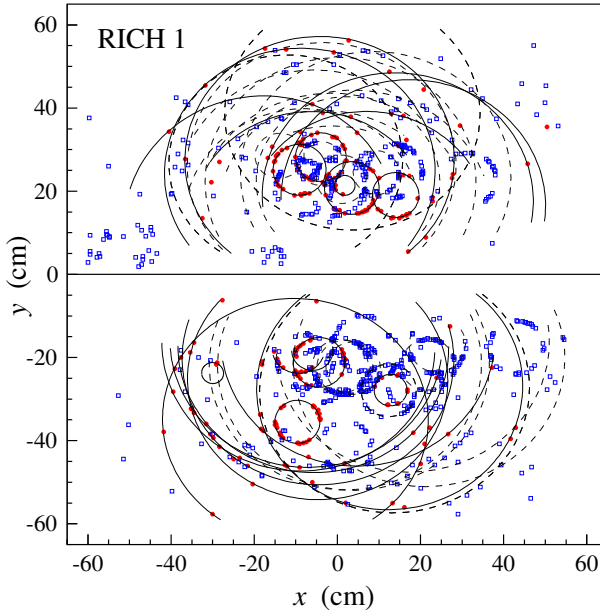


Figure 2.26: A typical LHCb event as seen by RICH 1. Fitted rings are superimposed over the HPD hits, with solid lines for long tracks and dashed lines for other track types.

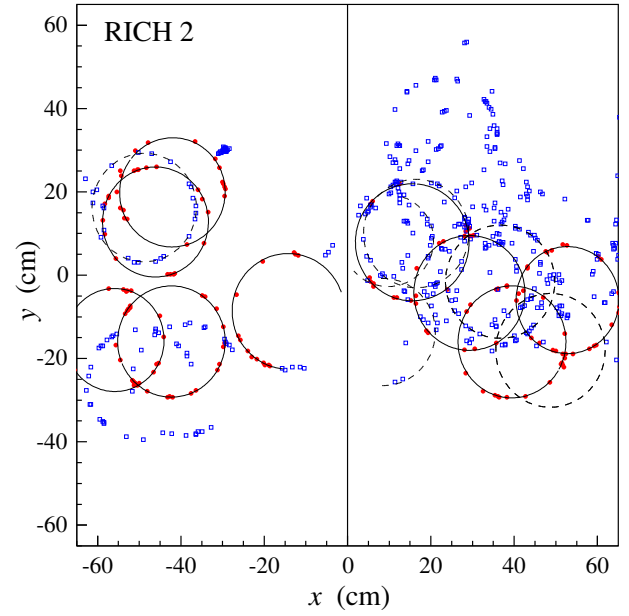


Figure 2.27: A typical LHCb event as seen by RICH 2. Fitted rings are superimposed over the HPD hits, with solid lines for long tracks and dashed lines for other track types.

2.5.7 RICH performance

As an example of RICH performance, consider the important case of correctly identifying kaons in an event, which is critical both in flavour tagging and the reconstruction of specific B decays. Figure 2.28 shows the $\Delta \ln \Lambda_{K\pi}$ distribution for true kaon and pion tracks, where the kaons are seen to have generally positive values of $\Delta \ln \Lambda_{K\pi}$, and pions are seen to have generally negative ones. Figure 2.29 shows the efficiency for identifying kaon tracks as kaons, and the rate of misidentifying the pions as kaons, when a cut is applied on the $\Delta \ln \Lambda_{K\pi}$ as a function of momentum. Higher values of the cut reduce the rate at which kaons are misidentified as pions, at the expense of reducing the efficiency of kaon identification. The overall kaon-pion separation performance is seen to be best in the momentum range 15-40 GeV. Finally, Figure 2.30 shows the pion-kaon separation in sigma as a function of pion momentum, for pions from the decay $B_d^0 \rightarrow \pi\pi$. Although the best separation is achieved in the momentum range 10-30 GeV, it is in general good between 2 and 100 GeV.

2.5.8 Electron and muon identification

Electron identification is achieved via the LHCb calorimeter system, which consists of a scintillating photo-detector, an electron calorimeter and a hadron calorimeter. The electron calorimeter has a radiation length of $25X_0$ and an energy resolution of $\sigma/E \sim 10\%/\sqrt{E}$. The hadron calorimeter has an energy resolution of $\sigma/E \sim 80\%/\sqrt{E}$. The electron is identified according to the χ^2 match between its track and the calorimeter shower, as well as by the showers of Bremsstrahlung photons emitted by the electron. The efficiency to identify electrons is 95%, with a pion misidentification rate of 0.7%.

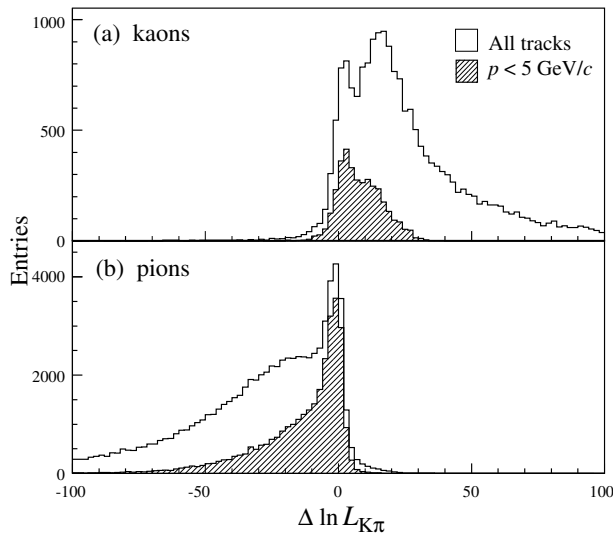


Figure 2.28: Difference in log-likelihood between the kaon and pion hypotheses from the RICH system, for (a) kaons and (b) pions, in a sample of $B_s^0 \rightarrow D_s K^+$ events; the shaded histograms are for low momentum tracks. Caption and figure taken from [col00b].

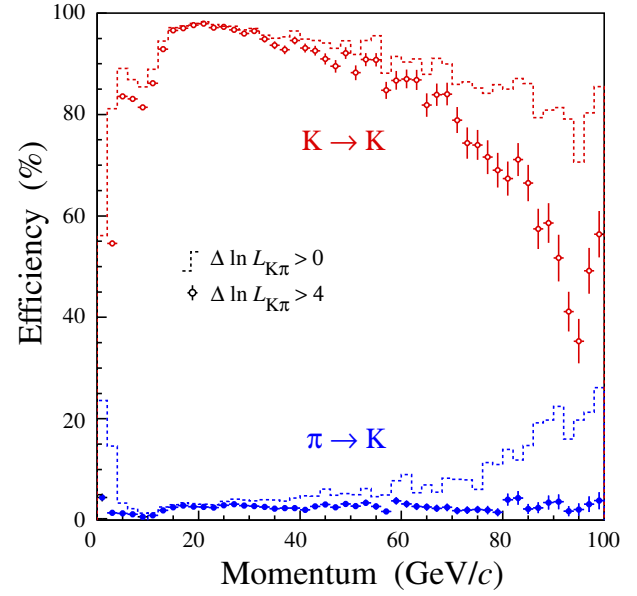


Figure 2.29: The overall (RICH1 and RICH2) kaon identification efficiency and pion misidentification rate as a function of momentum, for two different values of the $\Delta \ln \Lambda_{K\pi}$ cut.

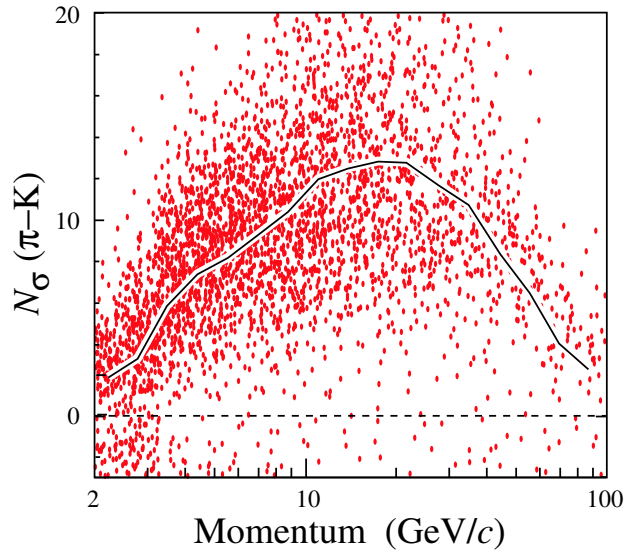


Figure 2.30: The pion-kaon separation in sigma as a function of particle momentum for pions from the decay $B_d^0 \rightarrow \pi\pi$. The solid line represents the average separation at each momentum.

The muon identification system consists of five multi-wire proportional chambers. Muon identification is performed by extrapolating tracks with $p > 3$ GeV within the muon detection system, and searching for hits in the neighbourhood of the extrapolated track. The efficiency to identify muons is 94%, with a pion misidentification rate of 1%.

2.6 Flavour tagging

The measurement of time dependent CP asymmetries at LHCb relies on being able to correctly identify the flavour of the B meson at production, since neutral B mesons can oscillate into their anti-particle between production and decay. There are two strategies for tagging at LHCb: “same side” tagging and “opposite side” tagging. Same and opposite side tagging refer to whether an event is tagged using information associated with the production of the signal B (same side), or from the decay chain of the “other” B (opposite side). This terminology bears no relation to LHCb geometry.

In all cases, the reduction in statistical power due to the mistagging of events is given by

$$P_{\text{tag}} = \epsilon_{\text{tag}} (1 - 2\omega_{\text{tag}})^2, \quad (2.11)$$

where ϵ_{tag} is the efficiency to find a tag, and ω_{tag} the mistag rate, i.e. the probability to wrongly tag the B flavour. For a typical channel with $P_{\text{tag}} \sim 10\%$, a sample of 100k untagged events will therefore have, when tagged, the same statistical power as a sample of 10k perfectly tagged events. All values and tables in this section are taken from [CLM07].

2.6.1 Same side tagging

Same side tagging uses correlations in the $b\bar{b}$ fragmentation chain or decays of excited states to tag the flavour of the B at production. If a B_s^0 is produced, there is an extra $\bar{s}$ left over, which forms a charged kaon about 50% of the time. This can be used to tag the B_s^0 , by attempting to find a charged kaon from the primary vertex whose pseudorapidity is close to that of the reconstructed signal B . The list of cuts applied on same side tagging kaon candidates is:

1. An impact parameter significance with respect to the primary vertex of $\frac{IP}{\sigma_{\text{IP}}} < 2.5$;
2. A difference in pseudorapidity with respect to the reconstructed B_s^0 of $|\Delta\eta| < 1$;
3. A difference in ϕ angle of $|\Delta\phi| < 1.1$ rad;
4. A difference in mass of the $B_s^0 K$ combination and the B_s^0 of $\Delta m < 1.5$ GeV;
5. A momentum of $p > 4$ GeV;
6. A transverse momentum of $p_t > 0.4$ GeV.

The same procedure can be applied to B_d^0 mesons and pions. Additionally, in the B_d^0 case, the tagging pion can also come from a $B^{**} \rightarrow B_d^0 \pi^+ \pi^0$ decay. The list of cuts applied on the same side tagging pion candidate is:

1. An impact parameter significance with respect to the primary vertex of $\frac{IP}{\sigma_{IP}} < 3$;
2. A difference in mass of the $B_d^0 \pi$ combination and the B_d^0 of $\Delta m < 3$ GeV;
3. A momentum of $p > 2$ GeV;
4. A transverse momentum of $p_t > 0.2$ GeV.

The pion with the highest p_t which survives these cuts is chosen as the tag, subject to an additional cut of $\Delta m < 1$ GeV.

Same side taggers primarily suffer from mistag because the wrong pion/kaon, or another particle not originating in the signal decay chain, is associated to the B . As there are many more pions than kaons in a typical LHCb event, the same side pion tag suffers from a greater probability of selecting an incorrect tagging partner than the same side kaon tag, and is therefore less powerful.

2.6.2 Opposite side tagging

Opposite side tagging infers the flavour of the signal B meson at production by observing the flavour of the “opposite side” B . The opposite side B is required to have decayed via a semileptonic or $b \rightarrow c \rightarrow s$ decay chain, at which point the muon, electron, or kaon daughter of the opposite side B tags its mother’s flavour at decay. The cuts applied for muon tags are:

1. A momentum of $p > 5$ GeV;
2. A transverse momentum of $p_t > 1.2$ GeV.

These cuts reduce the contribution of muons from $b \rightarrow c \rightarrow \mu$ decays, which would mistag the B . In addition, a DLL cut on the muon-pion separation of $DLL_{\mu-\pi} > 3$ is applied. The cuts applied for kaon tags are:

1. An impact parameter significance with respect to the primary vertex of $\frac{IP}{\sigma_{IP}} < 3.5$;
2. A momentum of $p > 3$ GeV;
3. A transverse momentum of $p_t > 0.4$ GeV.

These cuts preferentially select kaons from B decays to those from the primary vertex. To help reduce the background from misidentified pions and protons, DLL cuts on the kaon-pion separation of $DLL_{K-\pi} > 3$ and the kaon-proton separation of $DLL_{K-p} > -4$ are applied. The opposite side electron tag has the following cuts applied to it:

1. A momentum of $p > 5$ GeV;
2. A transverse momentum of $p_t > 1$ GeV;
3. A cut on the ratio E/p of the electron energy measured in the calorimeter compared to the momentum of the electron of $E/p > 0.8$;
4. A DLL cut on the electron-pion separation of $\text{DLL}_{e-\pi} > 3$.

The above tags are collectively referred to as single-particle tags, because they identify the flavour of a B based on a single tagging particle. There exists another type of opposite side tagger, the vertex charge tag, which performs an inclusive reconstruction of secondary vertices, and determines the flavour of the B from the charge of the reconstructed secondary vertex. The weighted charge of the vertex is determined by adding up the charge of the tracks used to make this vertex, weighted by the transverse momentum of the charges raised to the power of a value which has been empirically chosen to maximise performance:

$$Q_{\text{vtx}} = \frac{\sum_i p_{ti}^{0.4} Q_i}{\sum_i p_{ti}^{0.4}}. \quad (2.12)$$

In order for an event to be described as tagged, a cut of $|Q_{\text{vtx}}| > 0.3$ must be passed. Opposite side taggers suffer mistag because of misidentified daughters of the opposite side B , misreconstructed secondary vertices, and the fact that they tag the opposite side B at decay, whereas this B could have oscillated in the meantime.

2.6.3 Combining tagging decisions and performance summary

Since a single event may be tagged in more than one manner, a procedure is required to combine these tags into a final tagged B flavour. Each tagging decision q_i is assigned a probability p_i of being correct, based on the kinematic properties of the tag, and these probabilities are then combined into an overall probability $P(b)$ that the signal meson contains a b -quark:

$$P(b) = \frac{p(b)}{p(b) + p(\bar{b})}, \quad (2.13)$$

where $p(b)$ and $p(\bar{b})$ are given by

$$p(b) = \prod_i \left(\frac{1 - q_i}{2} + q_i p_i \right), \quad (2.14)$$

and

$$p(\bar{b}) = \prod_i \left(\frac{1 + q_i}{2} - q_i p_i \right). \quad (2.15)$$

The flavour of the B is decided by which of $P(b)$, $P(\bar{b})$ is the larger, with the caveat that events with $0.45 < P(b) < 0.55$ are considered untagged. This procedure assumes that the individual taggers are completely independent, which is not strictly true, especially for the

vertex charge tagger which is found to be highly correlated with other tags. These correlations will cause a loss of discriminating power, but are not expected to cause systematic errors in the tagging. The probabilities p_i are calculated using a neural network, which will be trained on control channels with similar kinematic properties to the decay channel being tagged. Examples of such control channels are $B_s^0 \rightarrow D_s^+ \pi^-$ for $B_s^0 \rightarrow D_s^+ K^-$, or $B_d^0 \rightarrow J/\psi K^{*0}$ for $B_d^0 \rightarrow J/\psi K_S^0$. Table 2.2 is taken from [CLM07], and lists the overall tagging performance in several representative LHCb channels.

Table 2.2: Performance of the combined tag for different signal decays passing trigger and offline cuts. The listed uncertainties are statistical.

Channel	$\epsilon_{\text{tag}} (\%)$	$\omega (\%)$	$\epsilon_{\text{eff}} (\%)$
$B^0 \rightarrow \pi^+ \pi^-$	52.48 ± 0.19	34.09 ± 0.25	5.31 ± 0.23
$B^0 \rightarrow K^+ \pi^-$	53.24 ± 0.17	33.50 ± 0.22	5.79 ± 0.22
$B^0 \rightarrow J/\psi (\mu\mu) K_S^0$	52.95 ± 0.08	35.37 ± 0.10	4.53 ± 0.09
$B^0 \rightarrow D^{*-} \mu^+ \nu_\mu$	56.22 ± 0.50	35.35 ± 0.63	4.82 ± 0.60
$B^+ \rightarrow J/\psi (\mu\mu) K^+$	54.05 ± 0.11	35.31 ± 0.15	4.66 ± 0.13
$B_s^0 \rightarrow D_s^\pm K^\pm$	61.11 ± 0.34	30.02 ± 0.42	9.76 ± 0.56
$B_s^0 \rightarrow D_s^- \pi^+$	60.86 ± 0.21	30.07 ± 0.26	9.67 ± 0.35
$B_s^0 \rightarrow J/\psi (\mu\mu) \phi$	54.46 ± 0.16	31.97 ± 0.21	7.08 ± 0.23

2.7 Trigger system

At the time the research within this thesis was being conducted, the LHCb trigger consisted of three separate stages: the L0 hardware trigger, the L1 software trigger, and the High Level Trigger (HLT) software trigger. Since then the L1 and HLT triggers have been merged into a new HLT, while the L0 has remained as before. This merger was for the most part accomplished by absorbing the L1 functionality within the HLT. The description below refers to the old trigger, since it is the relevant system for the physics studies presented in subsequent chapters. Some brief comments on the new trigger are included at the end of the section.

2.7.1 L0 hardware trigger

The L0 is a hardware trigger, whose purpose is to produce an event rate of 1 MHz for a bunch crossing rate of 40 MHz. It uses information from the calorimeter and muon subdetectors, as well as the pile-up veto. The L0 primarily triggers on high- p_t single particles, detected according to high energy deposits in the calorimeters. In order to improve the efficiency of finding J/ψ decays, there is one special kind of event for the L0: dimuon events, where two

muon tracks are reconstructed with a total $p_t > 1.3$ GeV, are accepted regardless of any other criteria. For all other events, the following cuts must be satisfied:

- The pile-up veto must accept the event;
- The total energy deposited in the hadron calorimeter (HCAL) must be > 5 GeV;
- The event must contain at least one calorimeter hit with a transverse energy above the relevant threshold
 - $E_T(\text{hadron}) > 3.6$ GeV;
 - $E_T(e) > 2.8$ GeV;
 - $E_T(\gamma) > 2.6$ GeV;
 - $E_T(\mu) > 1.1$ GeV;
 - $E_T(\pi^0 \text{ merged}) > 4.5$ GeV;
 - $E_T(\pi^0 \text{ resolved}) > 4.0$ GeV.

The “merged” and “merged” π^0 thresholds refer to the fact that the L0 distinguishes between two kinds of π^0 signals: those where both photons are registered on the same L0 readout card (merged), and those where the two photons register hits on neighbouring cards (resolved). Table 2.3 shows the L0 efficiencies, measured on events passing the full offline reconstruction, for a variety of LHCb signal channels. The table is dominated by hadronic and leptonic decay modes, none of which were used to calibrate the L0 trigger performance.

2.7.2 L1 trigger

The L1 software trigger uses data from the L0 decision unit, the VELO, the TT station and the calorimeters to reduce the data rate from 1 MHz to 40 kHz. It consists of a “generic” trigger component, which looks for a pair of tracks with high impact parameters and transverse momenta, and a “bonus” trigger component, which boosts the score from the generic trigger for an event according to the particles involved in the event’s L0 decision. The impact parameter of a track is defined as the perpendicular distance between a primary vertex and that track’s point of closest approach to this primary vertex, and is calculated with respect to every primary vertex in the event, with the biggest of these impact parameters being used in the trigger decision.

The first stage of the L1 trigger consists of 2D track reconstruction (in the $\mathbf{r}\text{-}\mathbf{z}$ plane) from the VELO R sensors. 2D tracks are then used to reconstruct 2D primary vertices, and a 2D impact parameter cut is then applied on those tracks not used to reconstruct the primary vertices. In order for a track to be selected for reconstruction in 3D, it must have a 2D impact parameter between 0.1 mm and 3.0 mm with respect to the primary vertex. The surviving tracks are reconstructed in 3D, using all VELO sensors, and are assigned a transverse momentum by the TT tracking station. These 3D tracks must have a 3D impact parameter cut between 0.15 mm and 3.0 mm with respect to the primary vertex in order

Table 2.3: L0 efficiencies, relative to the offline selection, for a variety of signal channels, with the inclusive efficiencies for the hadron, electron, and muon L0 triggers shown.

Decay Channel	ϵ_{L0} (%)	$\epsilon_{\text{hadron}}^{\text{L0}}$ (%)	$\epsilon_{\text{electron}}^{\text{L0}}$ (%)	$\epsilon_{\text{muon}}^{\text{L0}}$ (%)
$B_d^0 \rightarrow K^+ \pi^-$	54.1 ± 0.8	48.3 ± 0.8	12.3 ± 0.5	7.2 ± 0.4
$B_s^0 \rightarrow K^- \pi^+$	56.5 ± 1.1	51.2 ± 1.1	13.2 ± 0.7	6.7 ± 0.5
$B_s^0 \rightarrow K^- K^+$	51.8 ± 0.3	46.0 ± 0.3	11.6 ± 0.2	6.5 ± 0.2
$B_d^0 \rightarrow \pi^+ \pi^- \pi^0$	77.2 ± 1.6	39.4 ± 1.9	66.2 ± 1.8	7.9 ± 1.1
$B_d^0 \rightarrow D^{*-} (\overline{D^0} \pi^-) \pi^+$	49.0 ± 1.1	41.7 ± 1.1	14.0 ± 0.8	8.4 ± 0.6
$B_d^0 \rightarrow \overline{D^0} (K^+ \pi^-) K^{*0} (K^+ \pi^-)$	53.0 ± 1.4	45.3 ± 1.4	13.9 ± 0.9	8.1 ± 0.7
$B_d^0 \rightarrow \overline{D^0} (K^+ K^-) K^{*0} (K^+ \pi^-)$	50.7 ± 1.2	43.4 ± 1.1	13.6 ± 0.8	8.4 ± 0.6
$B_d^0 \rightarrow J/\psi (\mu^+ \mu^-) K^{*0} (K^+ \pi^-)$	91.2 ± 0.3	23.1 ± 0.4	9.3 ± 0.3	88.6 ± 0.3
$B_u^+ \rightarrow J/\psi (\mu^+ \mu^-) K^+$	90.3 ± 0.4	26.2 ± 0.5	9.1 ± 0.4	87.1 ± 0.4
$B_s^0 \rightarrow J/\psi (e^+ e^-) \phi (K^+ K^-)$	49.0 ± 0.6	22.9 ± 0.5	38.3 ± 0.5	7.0 ± 0.3
$B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \eta (\gamma \gamma)$	92.1 ± 0.8	19.2 ± 1.2	37.2 ± 1.5	88.4 ± 1.0
$B_s^0 \rightarrow \eta_c (4\pi, 2K2\pi) \phi (K^+ K^-)$	47.0 ± 3.0	41.5 ± 2.9	12.5 ± 1.9	8.0 ± 1.6
$B_s^0 \rightarrow \phi (K^+ K^-) \phi (K^+ K^-)$	41.8 ± 0.9	28.7 ± 0.9	9.7 ± 0.6	8.6 ± 0.5
$B_d^0 \rightarrow \mu^+ \mu^- K^{*0} (K^+ \pi^-)$	93.6 ± 0.7	24.9 ± 1.2	10.3 ± 0.8	91.8 ± 0.7
$B_s^0 \rightarrow \phi (K^+ K^-) \gamma$	69.6 ± 1.6	33.1 ± 1.6	65.8 ± 1.7	7.7 ± 0.9
$B_c^+ \rightarrow J/\psi (\mu^+ \mu^-) \pi^+$	92.6 ± 0.5	29.4 ± 0.8	9.9 ± 0.5	89.5 ± 0.6

to participate in the generic trigger decision. The generic trigger score is calculated for all pairings of surviving tracks according to:

$$\text{Generic L1 score} = \ln \left(p_t^{(1)} \right) + \ln \left(p_t^{(2)} \right), \quad (2.16)$$

where a track is assigned a $p_t = 400$ MeV if it did not have a transverse momentum measured by the trigger tracker.

The generic score is then boosted using bonus trigger decisions. For example, an event with an invariant dimuon mass within a ± 500 MeV mass window of the J/ψ or the B mass, or above the latter, is given a bonus large enough to automatically pass the L1 trigger irrespective of any other considerations. Similarly, events can gain a bonus for having high E_T electrons or photons. The event passes the L1 trigger if:

$$\text{Generic L1 score} + \text{Dimuon bonus} + \text{Electron bonus} + \text{Photon bonus} > \text{threshold}, \quad (2.17)$$

where the threshold is empirically calculated by requiring an output rate of 40 kHz.

Because its generic component cuts on the impact parameters of tracks, the L1 trigger preferentially selects long lived events, a point which will be considered more closely in Chapter 6. Table 2.4 shows the L1 efficiencies for a number of LHCb signal channels, for events passing the full offline reconstruction and the L0 trigger.

Table 2.4: L1 efficiencies for a variety of signal channels, normalized to events passing the L0 trigger.

Decay Channel	ϵ_{L1} (%)
$B_d^0 \rightarrow \pi^+ \pi^-$	62.7 ± 0.5
$B_d^0 \rightarrow K^+ \pi^-$	61.5 ± 1.0
$B_s^0 \rightarrow K^- \pi^+$	65.0 ± 1.4
$B_s^0 \rightarrow K^- K^+$	60.0 ± 0.4
$B_d^0 \rightarrow \pi^+ \pi^- \pi^0$	46.6 ± 2.2
$B_d^0 \rightarrow D^{*-} (\overline{D^0} \pi^-) \pi^+$	56.0 ± 1.6
$B_d^0 \rightarrow \overline{D^0} (K^+ \pi^-) K^{*0} (K^+ \pi^-)$	66.7 ± 1.8
$B_d^0 \rightarrow \overline{D^0} (K^+ K^-) K^{*0} (K^+ \pi^-)$	61.6 ± 1.6
$B_s^0 \rightarrow D_s^- (K^+ K^- \pi^-) \pi^+$	63.0 ± 0.9
$B_s^0 \rightarrow D_s^- (K^+ K^- \pi^-) K^+$	62.6 ± 0.4
$B_d^0 \rightarrow J/\psi (\mu^+ \mu^-) K_S^0 (\pi^+ \pi^-)$	67.7 ± 0.9
$B_d^0 \rightarrow J/\psi (e^+ e^-) K_S^0 (\pi^+ \pi^-)$	54.9 ± 1.4
$B_d^0 \rightarrow J/\psi (\mu^+ \mu^-) K^{*0} (K^+ \pi^-)$	76.8 ± 0.3
$B_u^+ \rightarrow J/\psi (\mu^+ \mu^-) K^+$	76.0 ± 0.5
$B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \phi (K^+ K^-)$	71.4 ± 0.2
$B_s^0 \rightarrow J/\psi (e^+ e^-) \phi (K^+ K^-)$	57.2 ± 0.8
$B_s^0 \rightarrow J/\psi (\mu^+ \mu^-) \eta (\gamma \gamma)$	70.3 ± 1.5
$B_s^0 \rightarrow \eta_c (4\pi, 2K2\pi) \phi (K^+ K^-)$	59.0 ± 4.0
$B_s^0 \rightarrow \phi (K^+ K^-) \phi (K^+ K^-)$	60.3 ± 1.5
$B_d^0 \rightarrow \mu^+ \mu^- K^{*0} (K^+ \pi^-)$	78.5 ± 1.1
$B_d^0 \rightarrow K^{*0} (K^+ \pi^-) \gamma$	51.9 ± 1.4
$B_s^0 \rightarrow \phi (K^+ K^-) \gamma$	49.3 ± 2.0
$B_c^+ \rightarrow J/\psi (\mu^+ \mu^-) \pi^+$	65.6 ± 0.9

2.7.3 High level trigger

The high level trigger (HLT) is a software trigger which reduces the data rate from the L1 output of 40 kHz to the final output rate of 2 kHz. It has access to information from all subdetectors. It consists of a generic trigger, a set of exclusive selections, and a set of inclusive selections. The generic trigger is also the first stage of the HLT, at which the L1 trigger decision is re-confirmed using the higher quality reconstruction available to the HLT, in particular by adding information from the three tracking stations T1-T3 to the TT station in order to improve the momentum resolution. This step alone halves the data rate to 20 kHz. Subsequently, a series of inclusive and exclusive preselection algorithms is run over the event, and the event passes as long as one of these preselections selects it. HLT efficiencies vary from channel to channel – for example, the channel $B_d^0 \rightarrow D^+ \pi^-$ has an efficiency of 94% in the generic HLT for events passing the L0 and L1 triggers.

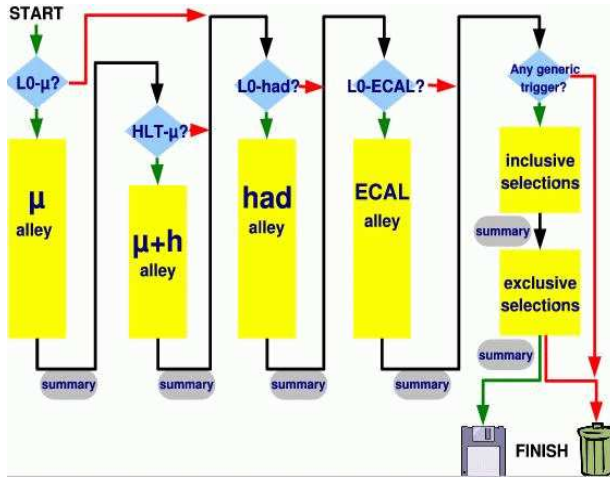


Figure 2.31: The new HLT, as used in the DC06 LHCb Monte Carlo.

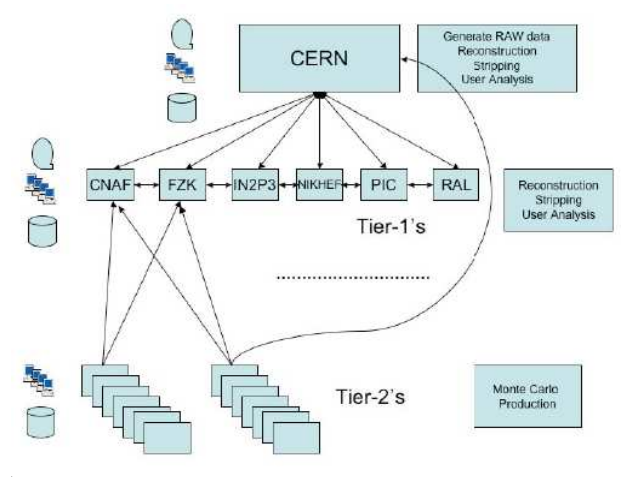


Figure 2.32: A schematic of the LHCb computing model, reproduced from [Bro04]

2.7.4 The new trigger

A recent redesign of the trigger has merged the L1 and HLT into a single software based HLT, which will reduce the data rate from the L0 output of 1 MHz to a rate of 2 kHz, which will be written to tape. The trigger is now divided into “alleys”, as seen in Figure 2.31. The hadron alley is based on the old L1 generic trigger (two high impact parameter and transverse momentum tracks), while the “Any generic trigger?” stage represents the point at which the old generic HLT is applied. A major difference between the new trigger scheme and the trigger described in this chapter, and a motivating factor in the change, is that all detectors are read out at 1 MHz after the L0 hardware trigger, allowing for greater flexibility in the design of the trigger decisions after L0. The new trigger allows any cut used in an offline selection to be applied in the trigger—for example, studies have recently begun on using particle ID information instead of impact parameter cuts in the early stages of the trigger. At the time of writing, the new trigger has yet to be fully implemented in the LHCb software, and no studies with it were possible. Approximately 900 Hz of the 2 kHz output rate of the new trigger is expected to be taken up by a single muon trigger, which identifies events with a single high p_T muon. This is the LHCb equivalent of a “minbias” trigger, and will be used for data mining.

2.8 LHCb software

All results in this thesis have been obtained using PYTHIA [SMS06] generated Monte Carlo data, processed by the full LHCb detector simulation. This software is important not only for designing analyses before the detector is switched on, but will also be used to help in detector calibration and the analysis of certain channels once the LHC is running. The LHCb software is documented online [col07c], and the following summary draws on this documentation. All software version numbers listed are those used for the studies in this thesis.

2.8.1 Event simulation and reconstruction

The simulation of events at the LHC consists of five phases: event generation, detector response simulation, digitization, event reconstruction, and event selection and analysis, which are outlined in turn.

- **Event generation** : the LHCb Monte Carlo events are generated using PYTHIA and EvtGen [Lan01]. PYTHIA simulates the production of $b\bar{b}$ pairs in pp collisions through the processes of gluon fusion, gluon splitting and flavour excitation. The EvtGen package then decays the resulting B mesons from a table of predetermined decay channels and rates. The full LHCb Monte Carlo used in this thesis was centrally created using PYTHIA 6.226.2 and a private (to the LHCb collaboration) adaptation of EvtGen.
- **Detector response simulation**: once an event has been generated, it is necessary to simulate the passing of particles through the detector and their interactions with it. The package GEANT 4 [All06] is used for this purpose, controlled through the LHCb simulation program GAUSS. The LHCb detector is simulated in detail, including:
 - **Passive materials** such as supports, frames, shielding elements (including the VELO RF foil and vacuum vessels), and the LHC beam pipe are simulated, and secondary interactions of particles are tracked down to 10 MeV for hadrons and 1 MeV for electrons and photons.
 - **The LHCb magnet** is simulated using a field map measured on the magnet which will be used in the running detector.
 - **The spill over** of particles from one event into another, is simulated. The probability of neighbouring bunch crossings producing particles is calculated from the instantaneous luminosity.
 - **Individual subdetectors** are fully simulated. For example, effects such as charge sharing, noise, and cross talk are included in the simulation of silicon sensors.

GAUSS version v15r19 was used to simulate the data analyzed in this thesis.

- **Digitization**: once the detector response to the event has been simulated, it is necessary to simulate the response of the detector and the readout electronics to the hits in the detector. This includes simulating the propagation of electrical signals through the detector, the response of the readout hardware to these signals, and the analogue to digital conversions performed in the readout modules and L0 trigger boards. The digitization is performed using the package BOOLE, version v6r5. BOOLE produces digitized data in the same format as that which will be produced by the LHCb hardware once the machine is running.
- **Reconstruction**: the LHCb reconstruction software BRUNEL creates tracks and RICH Cherenkov rings out of the digitized output provided by BOOLE. BRUNEL version v24r6 was used to reconstruct the data in this thesis, and it contains the full

pattern recognition and tracking software which will be used to reconstruct data taken during LHCb running.

- **Analysis:** The tracks and RICH information output by BRUNEL are the inputs to the LHCb analysis package, DAVINCI, which uses them to reconstruct the B mesons in the decay. The version of DAVINCI used in this thesis actually contained the software triggers (L1 and HLT) as packages, but the software trigger will eventually be separated from the offline analysis software. Although the analysis software is written in C++, there have been recent moves to standardize the tools used in the offline analysis of events, and most end users will perform offline reconstruction through so-called “options files”, written in a shell scripting language. It is foreseen that these options files will eventually be migrated to python for the greater flexibility it affords the end user. DAVINCI versions v12r14, v12r16, and v12r18 have been used in this thesis.

2.8.2 GRID computing

The amount of data collected and processed by the LHC detectors will far exceed any previous high energy physics experiment, and this fact has necessitated the development of new computing strategies, in particular the GRID distributed computing model. It is estimated that LHCb will require (for offline processing) 12.97 MSI2k of computing power, 3.3 PB of disk storage space and another 3.4 PB of mass storage in the first year of running. By comparison, a modern desktop with a 3 GHz quad core processor provides 12 kSI2k of computing power, meaning that LHCb alone² would require 1000 such computers. Moreover, such requirements will rise over time, so that in its third year of running, LHCb will require 17.9 MSI2k of computing power, 4.7 PB of disk space and 11.6 PB of mass storage space.

The GRID distributed computing project has been developed to meet the above demands, and the GRID based LHCb computer model is detailed in [Bro04]. A schematic of the model can be seen in Figure 2.32. The GRID serves two functions: to quickly distribute raw data from the LHC to processing centres around the world, and to dynamically allocate computing resources as data processing jobs are initiated. Raw data will be produced at CERN and distributed among the Tier-1 computing centres, where it will be reconstructed and “stripped” into a smaller data set ready for physics analysis. Raw data will be stored at CERN, with another copy distributed across the other Tier-1 centres, while each year’s stripped data will be distributed across all Tier-1 centres to balance the analysis load. Jobs are submitted to the GRID as a whole, and are allocated to whatever processing centre has spare capacity at that time, in a manner transparent to the user. The Tier-2 centres serve primarily for Monte Carlo production, although they are expected to provide modest disk storage capacity and may also be used to run certain analysis jobs.

The signal and background events in the channel $B_d^0 \rightarrow D^+ \pi^-$ analyzed in this thesis were reconstructed on the GRID. A proprietary python script written³ by the author was

²The computing requirements of ATLAS and CMS will be many times greater because of their higher running luminosity.

³Based on a shell script written for the same purpose by A. Powell.

used to translate DAVINCI options files into a format readable by the DIRAC software used for GRID job submission on LHCb. Additional python scripts were written to allow the bulk submission and automated retrieval of jobs from the GRID, and all scripts used DIRAC as the interface to GRID services.

Chapter 3

Reconstruction of the Channel

$$B_d^0 \rightarrow D^+ \pi^-$$

3.1 Introduction

The CP asymmetry accessible in the $B_d^0 \rightarrow D^+ \pi^-$ decay modes¹ depends on the combination of CKM phases $2\beta + \gamma$. The γ sensitivity comes from tree level diagrams, and hence allows for a theoretically clean extraction of the angle γ , which is expected to be robust against new physics [DS88, Fle03]. Such a measurement of γ can then be used as a benchmark comparison for other channels, like $B_{d,s}^0 \rightarrow D_{d,s} D_{d,s}$, which are expected to be sensitive to new physics. Its large branching fraction also makes this channel a prime candidate for a measurement of the B_d^0 lifetime. As discussed in Chapter 6, such a measurement will be important in detector calibration, as well as helping to constrain theoretical models such as Heavy Quark Expansion theory [Shi00, Ura98, Pet04, GOP04].

The channel in question is reconstructed using the the exclusive decay chain

$$B_d^0 \rightarrow D^+ (K^- \pi^+ \pi^+) \pi^-$$

as this exploits the most abundant decay mode of the D^+ that does not involve neutral particles.

Section 3.2 presents an overview of the techniques for separating $B_d^0 \rightarrow D^+ \pi^-$ events from the background. Section 3.3 discusses a tool developed for the categorization of the background events surviving the selection. Section 3.4 describes a genetic algorithm developed in order to optimize the event selection, and lists the resulting selection cuts.

All studies within this thesis were performed using the DC04 [col07a] versions of the LHCb Monte Carlo and detector simulation. Table 3.1 lists the version numbers of the programs used to create datasets for this study. The final DC04 release of the LHCb analysis framework, DaVinci v12r18, was used to analyze the Monte Carlo data.

¹Unless otherwise specified, this label should be taken to refer to the following four decay modes : $B_d^0 \rightarrow D^+ \pi^-$, $B_d^0 \rightarrow D^- \pi^+$, $\overline{B}_d^0 \rightarrow D^+ \pi^-$, $\overline{B}_d^0 \rightarrow D^- \pi^+$.

Table 3.1: The program versions used in the creation and reconstruction of the Monte Carlo data for this study.

Package Name	Signal Dataset	Background Dataset
Brunel	v24r5	v23r7
Boole	v6r5	v5r9
Gauss	v15r19	v15r13

3.2 Datasets and preselection

In order to estimate the expected yields and signal to background ratios in this channel, signal and background datasets were created. A year of LHCb running is expected to contain $10^7 s$ of data-taking at a luminosity of $2 \times 10^{32} \text{cm}^{-2} \text{s}^{-1}$. Assuming a $b\bar{b}$ production cross section of $500 \mu\text{b}$, this corresponds to $10^{12} b\bar{b}$ events.

3.2.1 Signal Dataset

350,000 signal events were generated, each of which contains at least one decay of the type $B_d^0 \rightarrow D^+ \pi^-$, where the D^+ decays according to $D^+ \rightarrow K^- \pi^+ \pi^+$. The momentum vector of these B_d^0 particles was generated within a 400 mrad forward cone, and simulation of the detector shows that 34.7% of all b quarks produced within 4π survive such a cut.

The branching ratio assumed for the channel $B_d^0 \rightarrow D^+ (K^- \pi^+ \pi^+) \pi^-$ is $(2.55 \pm 0.15) \cdot 10^{-4}$ [ea06c]. The hadronization fraction for the B_d^0 is approximately 40% [ea06c], so that 64% of all $b\bar{b}$ events will contain at least one B_d^0 . Taking into account the signal acceptance of 34.7%, 350,000 signal events therefore correspond to $12.6 \cdot 10^9 b\bar{b}$ events, or ~ 36 hours of LHCb data taking.

3.2.2 Background Dataset and Preselection Cuts

33.9 million $b\bar{b}$ events were generated for background studies. The B hadrons were hadronized and decayed inclusively in a geometric acceptance of 43%. A set of $b\bar{b}$ events is used because the LHCb trigger [col03b] is expected to effectively suppress events not containing a B hadron, and inclusive B events are therefore expected to provide the dominant source of background. These events correspond to 13.1 minutes of LHCb data taking.

In order to speed up analysis, this data set is stripped using a set of loose preselection cuts. During the stripping, preselection algorithms for all channels under consideration at LHCb are run in parallel, and an event is saved if it is selected by at least one preselection algorithm. 15,536 events are selected only by the $B_d^0 \rightarrow D^+ \pi^-$ preselection during the stripping², and these are subsequently used to optimize the yield and signal to background level. As a

²Events which are selected by more than one preselection are likely to be noisy, and are hence discarded.

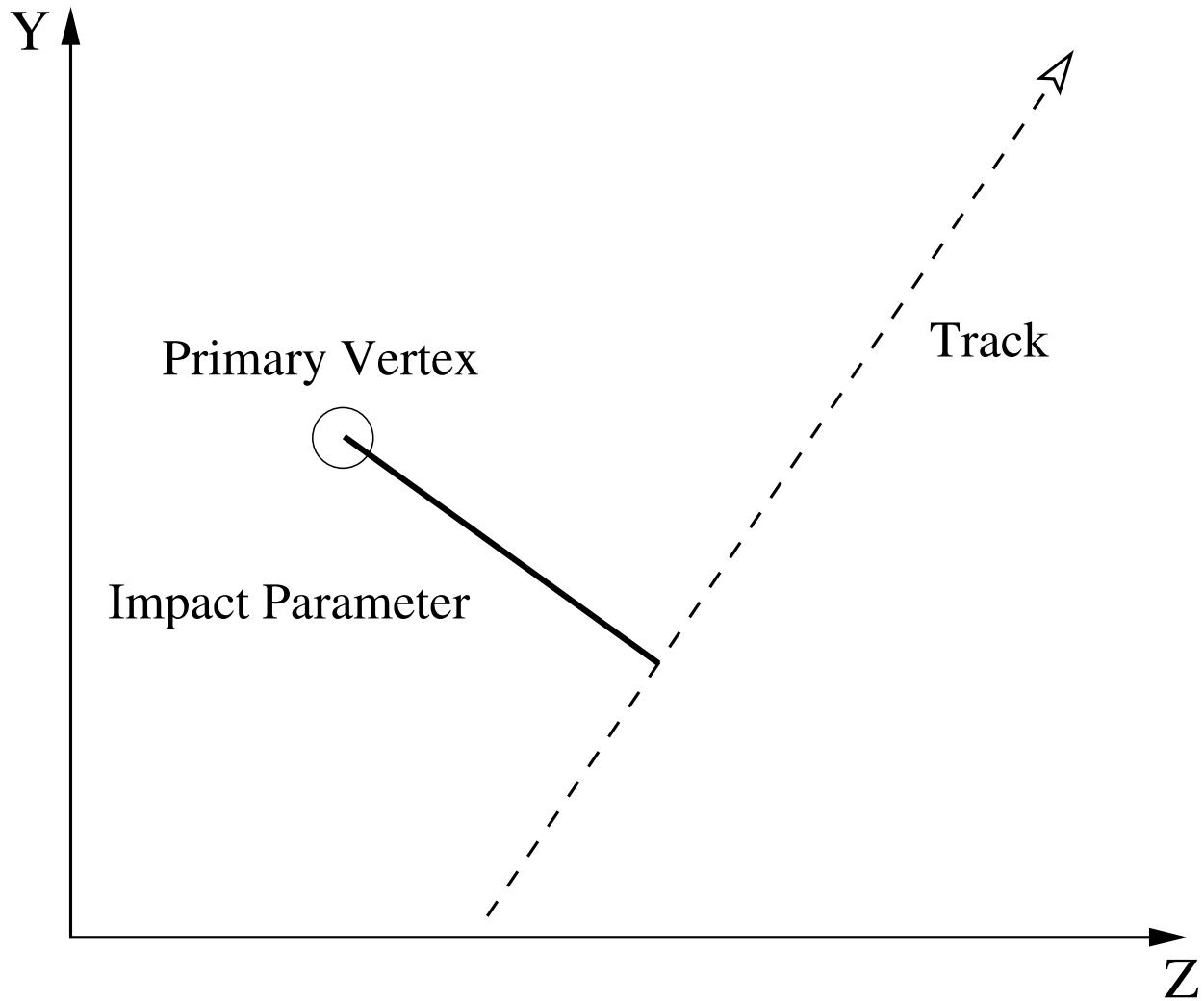


Figure 3.1: The 2D impact parameter of a track with respect to a vertex.

consequence, the preselection cuts define the minimum cut values for the optimization, and no cut should be used in the preselection if there is any doubt that it can be safely used in the final selection. The preselection was required to accept at most $\frac{1}{2000}$ of the $b\bar{b}$ events, in order to limit the output data size of the stripping; in addition, it was desired to standardize as much as possible the selections of $B_d^0 \rightarrow D^+\pi^-$ and $B_s^0 \rightarrow D_s^+h^-$, where the h^- stands for a pion or a kaon, since a homogeneous selection is desirable for the extraction of γ exploiting the U-spin symmetry [Fle03], as discussed in Chapter 5. For this reason, the preselection for $B_d^0 \rightarrow D^+\pi^-$ was based wherever possible on the existing $B_s^0 \rightarrow D_s^+h^-$ preselection [BNSVH07]. The full list of preselection cuts used is given in Table 3.2. The term “bachelor pion” refers to the pion originating directly from the B , which is especially useful in the selection of this channel owing to its high transverse momentum relative to the daughter pions and kaons of the D . Some of the less obvious cuts are defined below:

- **Impact parameter :** The impact parameter of a track with respect to a vertex is defined as the perpendicular distance between the track and the vertex. Figure 3.1 shows this graphically in two dimensions.

- **DLL $K - \pi$ separation** : As explained in Chapter 2, the RICH [col00b] sub-detector is used to provide particle identification, and in the context of this channel it is used to distinguish between pions and kaons. The Delta Log Likelihood (DLL) cut is the difference in the log-likelihoods of the RICH reconstructions of a track under different particle hypotheses, in this case the pion hypothesis and the kaon hypothesis. Because pions are considerably more abundant than kaons in the event, this cut is applied on the kaon candidate in the decay.
- **D^+ and B_d^0 mass windows** : Wide mass windows are used in the preselection to provide “sidebands” to allow for a study of the background level and makeup. These windows correspond to approximately 30σ in the case of the B_d^0 and 7σ in the case of the D^+ . σ is the resolution of the signal peak: ~ 17 MeV for the B_d^0 , and ~ 7 MeV for the D^+ . Tight mass windows, corresponding to 3σ in both cases, are used for the final $\frac{S}{B}$ estimate. Neither the B nor the D are mass constrained during the vertex fitting; rather, the mass window cut is applied on the reconstructed mass of the B or D candidate after the vertex fit.
- **D^+ minimum flight separation (abbreviated as FSPV)** : The D^+ is a relatively long lived particle (with a mean life of ~ 1.0 ps) and is a daughter of the B_d^0 , itself a long-lived particle. Hence a cut on the significance of the D^+ spatial separation from the primary vertex, given in equation 3.1, is an excellent discriminant against background. In events which contain more than one primary vertex, the primary vertex with respect to which the B_d^0 has the smallest impact parameter is used for this cut. This cut is also applied to the B_d^0 , but only in the final selection:

$$FSPV = \frac{z_{D^+} - z_{PV}}{\sqrt{\sigma_{z,D^+}^2 + \sigma_{z,PV}^2}}, \quad (3.1)$$

where z_{D^+} is the z-axis position of the D^+ decay vertex, z_{PV} is the z-axis position of the primary vertex, and $\sqrt{\sigma_{z,D^+}^2 + \sigma_{z,PV}^2}$ is the error on these parameters added in quadrature.

- **Cut on the B_d^0 direction ($\cos \theta$)** : The direction of the B_d^0 momentum should be consistent with the B_d^0 flight direction as calculated from the vector joining the primary vertex and the B_d^0 decay vertex. The angle θ is defined as $\cos \theta = \frac{\mathbf{P}_B \cdot (\mathbf{D}\mathbf{V}_B - \mathbf{P}\mathbf{V})}{|\mathbf{P}_B| |\mathbf{D}\mathbf{V}_B - \mathbf{P}\mathbf{V}|}$, where $\mathbf{D}\mathbf{V}_B$ is the coordinates of the B_d^0 decay vertex, and $\mathbf{P}\mathbf{V}$ is the coordinates of the primary vertex. An event at LHCb can have more than one reconstructed primary vertex. The primary vertex with respect to which the B_d^0 has the smallest impact parameter is used for this cut.
- **Cut on the impact parameter significance (IPs)** : The daughters of the B_d^0 and D^+ come from detached vertices, and as a result they are expected to have high impact parameter significances with respect to the primary vertex. This cut is applied with respect to all primary vertices in the event.

Figure 3.2 shows the signal and background distributions for some of the cut quantities used in this selection.

Figure 3.2: The signal (solid red) and background (dashed blue) distributions of some cut quantities used in the selection. Signal and background are both normalized to unit area. “IPs” refers to the 3D unsigned impact parameter significance.

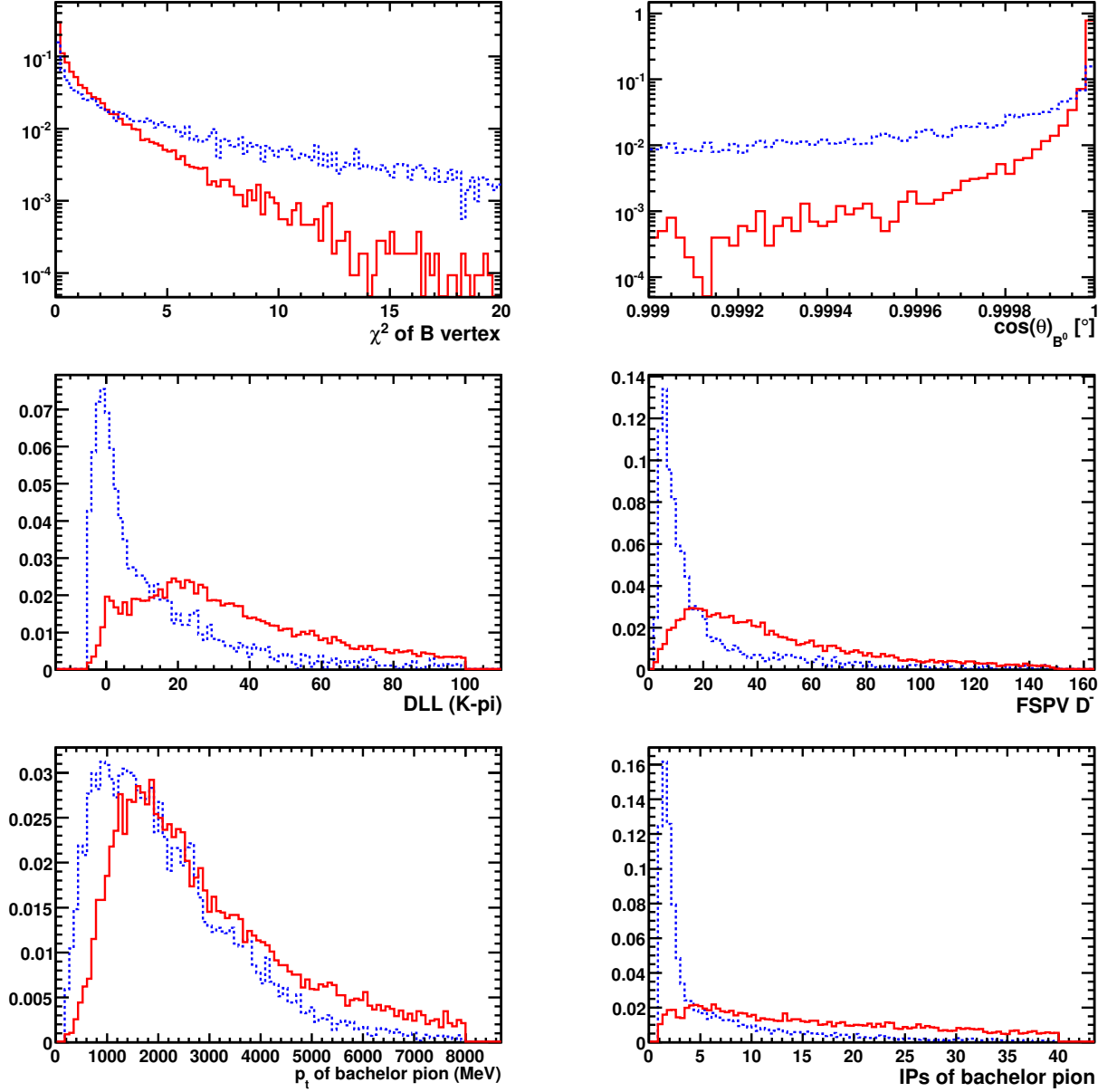


Table 3.2: The list of cuts used in the stripping preselection and a summary of the number of events surviving the cuts.

Cut	Cut Value
All particle momenta	$> 2000 \text{ MeV}$
DLL K-Pi	> -5.0
D^+ daughter p_t	$> 300 \text{ MeV}$
D^+ p_t	$> 2000 \text{ MeV}$
D^+ daughter IPs	> 1
D^+ IPs	> 2
D^+ mass window	$\pm 50 \text{ MeV}$
D^+ vertex χ^2	< 20
D^+ FSPV	> 4.5
Bachelor π p_t	$> 500 \text{ MeV}$
Bachelor π IPs	> 2.5
B_d^0 mass window	$\pm 500 \text{ MeV}$
B_d^0 vertex χ^2	< 20
B_d^0 IPs	< 6
$B_d^0 \cos \theta$	> 0.9999
Input signal events	148,000
Surviving signal events	9901
Surviving background events	1267
$\frac{S}{B}$	~ 1.3

3.3 Background Classification

While the overall signal to background ratio directly affects the expected statistical precision on the physics parameters measured, the composition of the background can be equally important. In particular, it must be established whether the background under the peak is qualitatively equivalent to the background in the sidebands for the parameters under consideration, since the sidebands are what will be used to judge the overall background composition with real data. In the case of lifetime measurements, the question is whether the background has the same lifetime profile, while in the case of measuring γ , it is whether the background has the same CP asymmetry. Although it is in principle possible to scan through the Monte Carlo “truth dumps” in order to establish the background composition, this is a time intensive procedure which should be applied to as few events as possible. Therefore, a tool has been developed which performs a classification of the background into such categories as combinatoric, ghost, mis-identification, etc. This classification gives the user an idea of the background composition, which can guide the selection of cuts and selection optimization. The classification described here, and implemented by the author³, has been adopted as a standard approach at LHCb.

3.3.1 Background Categories

The background categories can be broadly split into two groups: those due to “ghost” tracks or combinatorics, which do not therefore involve the reconstruction of anything physically meaningful, and those caused by the reconstruction of all or part of an actual physical process (a B hadron decay, for example). A ghost track is a reconstructed track which does not correspond to any unique physical particle, and is an artifact of the reconstruction software. A detailed description of each type of background is given below. In addition, Table 3.3 gives the numerical codes associated with each type of background which the tool outputs when running. The classifications listed are mutually exclusive, and Appendix B contains a flowchart showing the order in which the categories are considered.

- **Signal :** For a reconstructed decay to be defined as signal, it is first of all necessary to associate each final state particle in the decay with a Monte Carlo truth particle (referred to as “MCParticle” from here on). A decay is defined as signal if the following conditions are satisfied:
 - Each final state particle has an MCParticle associated to it with the same PID;
 - All such MCParticles have a common MCParticle mother;
 - There is a one-to-one correspondence between the final state particles and final state MCParticle daughters of the common MCParticle mother;
 - The common MCParticle mother has the same PID as the reconstructed decay mother.

³This method of classification was first proposed in [SK05].

Table 3.3: Numerical codes associated with each background category.

Category	Background Type
0	Signal
20	Fully reconstructed physics background
30	Reflection
40	Partially reconstructed physics background
50	Low mass background
60	Ghost background
70	Primary vertex background
80	Badly reconstructed primary vertex
100	Pileup background
110	$b\bar{b}$ background
120	$c\bar{c}$ background
130	uds background

A final state particle is associated to an MCParticle if 70% of their hits (in all the experiment's subdetectors) match. This classification does not require that any intermediate resonances have been reconstructed. For example, the decay $D^+ \rightarrow K^- \pi^+ \pi^+$ (non-resonant) is considered indistinguishable from the decay $D^+ \rightarrow \bar{K}^*(892)^0 \pi^+$ where the $\bar{K}^*(892)^0$ decays to $K^- \pi^+$.

- **Physics Backgrounds :** The following backgrounds are caused by the reconstruction of all or part of an actual physics decay.
 - **Fully reconstructed physics background :** An example of this background is a true $B_d^0 \rightarrow K^+ \pi^-$ decay reconstructed as $B_s^0 \rightarrow K^+ \pi^-$. The decay topology is fully and correctly reconstructed, but the mother particle is incorrectly identified.
 - **Reflection :** This background is caused by the misidentification of a final state particle, for instance that of a pion as a kaon. An example is $D^0 \rightarrow K^- \pi^+$, which has a branching ratio approximately an order of magnitude greater than that of $D^0 \rightarrow K^- K^+$ and can be mistaken for the latter mode if a pion is misidentified as a kaon. Reflections are protected against by using tight RICH particle identification cuts and mass cuts.
 - **Partially reconstructed physics background :** This label is applied to cases where a fragment of a decay occurring in the event is incorrectly identified as a signal decay. An example would be the decay $B_d^0 \rightarrow D^+ \pi^- \pi^- \pi^+$ where two of the pions are missed out in the reconstruction and it is then misidentified as a $B_d^0 \rightarrow D^+ \pi^-$ decay. Such a background is not in principle dangerous, because the reconstructed B_d^0 is likely to have an unphysically low mass. However, if this background involves not only a missing particle but also a misidentified particle (as in the reflection example above), the reconstructed B_d^0 mass could be pushed

back under the mass peak. Such backgrounds are very dangerous and difficult to protect against because they peak under the signal.

- **Low mass background :** This is a special case of the partially reconstructed physics background, in which there is no misidentification and the reconstructed particle is found to have a mass systematically below the signal peak. This background is not dangerous insofar as it does not fall under the peak, however the presence of such background means that the sideband below the mass peak cannot safely be used to model the background under the signal.
- **Technical and Combinatoric Backgrounds :** The following backgrounds are caused by technical problems (ghost tracks) and combinatorics. They involve the reconstruction of a particle which did not really exist in the event, unlike physics background, whose reconstructed particles did exist in the event, but were either misidentified or did not originate from the decay being reconstructed.
 - **Ghost background :** Any reconstructed decay in which one or more of the final state particles is found to have no associated MCParticle is classified as ghost. As with most backgrounds in this section, ghost backgrounds are best protected against by kinematic or lifetime cuts, for example those on transverse momentum or impact parameter, or by track quality cuts.
 - **Primary vertex background :** Any reconstructed decay in which one or more of the final state particles are found to come from the primary vertex is classified in this category. Impact parameter cuts protect against this type of background.
 - **Badly reconstructed primary vertex :** It is occasionally possible for the primary vertex reconstruction to fail to reconstruct a primary vertex correctly. Particles originating from such a vertex can seem to be coming from a secondary vertex, mimicking a detached B or D decay vertex, which makes them a dangerous source of background. In this case, it is necessary to apply a vertex isolation cut, demanding that no particles apart from those used in the reconstruction of the candidate particle point to its decay vertex. An example of the use of such a cut is detailed in [FS06].
 - **Pileup background :** Any reconstructed decay in which the final state particles are found to come, directly or indirectly, from more than one primary vertex is classified in this category.
 - **$b\bar{b}$ background :** Any background which does not fit into any of the previous categories, but in which at least one of the final state particles has a mother with bottom content, is classified in this category.
 - **$c\bar{c}$ background :** Any background which does not fit into any of the previous categories, but in which at least one of the final state particles has a mother with charm content, is classified in this category.
 - **uds background :** Any background which does not fit into any of the previous categories is classified here.

Table 3.4: The composition of background events after the preselection and the final selection as optimized for a best $\frac{S}{\sqrt{S+B}}$. All numbers are as measured on the $b\bar{b}$ tapes.

Background category	After preselection (w/o trigger)	After final selection (with trigger)
Signal	186	59
Reflection	136	8
Partially reconstructed physics background	58	1
Low mass background	277	0
Ghost background	573	1
Primary vertex background	386	0
Badly reconstructed primary vertex	454	0
Pileup background	704	0
$b\bar{b}$ background	160	2

3.3.2 Background Categories After the Preselection

Table 3.4 shows the makeup of the background after the stripping preselection has been applied, as well as after the final selection described later. The branching fraction of the $B_d^0 \rightarrow D^+(K^-\pi^+\pi^+)\pi^-$ decay in the LHCb Monte Carlo was $2.91 \cdot 10^{-4}$. Note that the preselection is run without the trigger, whereas the final selection is run after the full trigger sequence. The background after the preselection is dominated by combinatorics, with substantial contributions from the low mass background category and reflections. We also note that $\sim 6\%$ of the events selected by the preselection are signal. Figure 3.3 shows the mass distributions of the various observed background category types after the preselection. It can be seen that the “low mass” category will not contribute to the background under the signal peak, and events of this type will be ignored in all background calculations from now on.

3.4 Optimizing the Event Selection

Studies anticipated for this channel, such as the extraction of the angle γ from a study of the time dependent asymmetries, will require both high yields and signal purity. It is therefore crucial that the final event selection for the channel is optimized. However, the number of cuts involved makes a brute force search through the “phase space” of selection variables impractical. With 13 cuts to optimize (neglecting the mass cuts and the global cut on particle momenta, but adding a FSPV cut on the B_d^0), there are n^{13} points in phase space to scan through, where n is the number of discrete values each cut can take, which is assumed to be same for all cuts. The measured speed of scanning through the phase space is 500 grid points per second, so that even with a coarse grid, say $n = 5$, it would take 1.9 years to fully cover all possible cuts. It is of course possible to use various tricks to speed up such

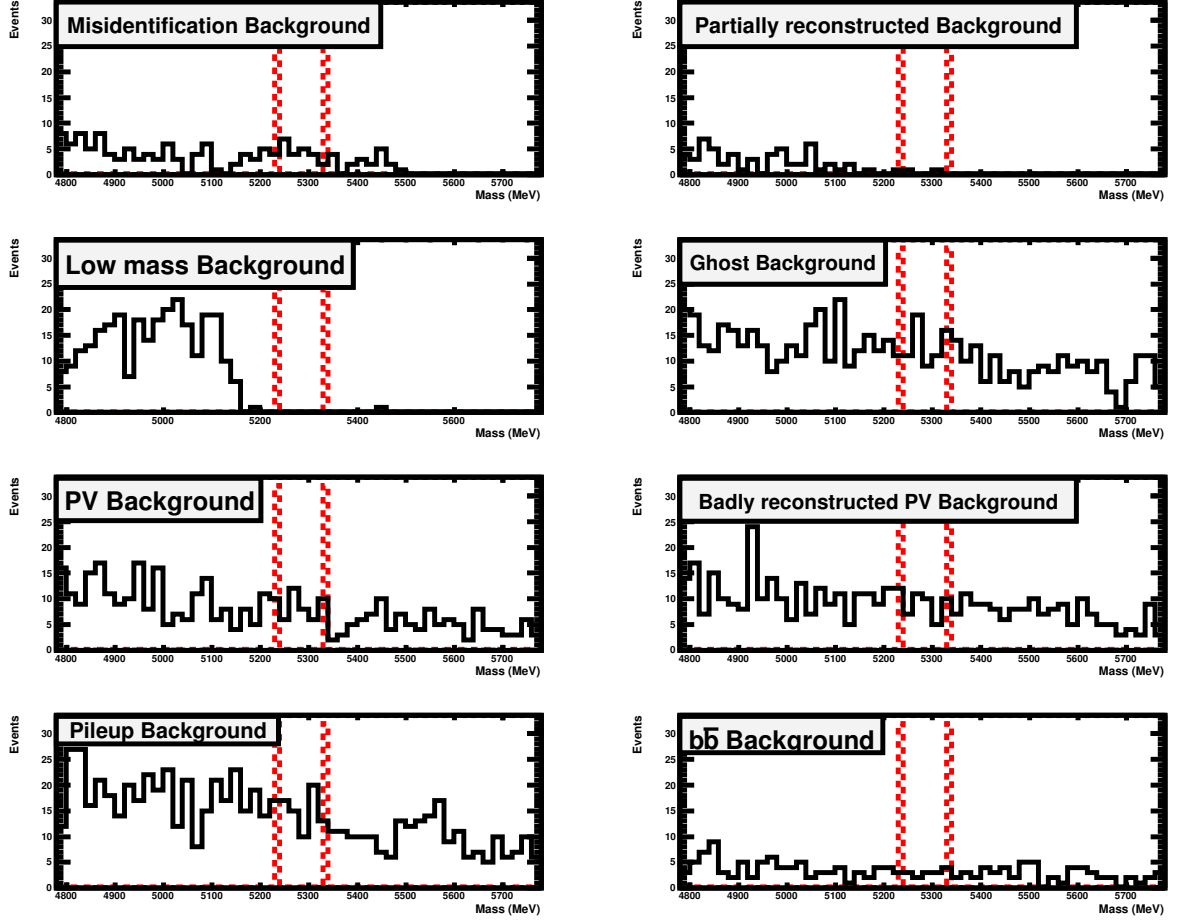


Figure 3.3: The mass distributions for the various background categories after the preselection. The plots are unnormalized, the vertical axis represents the number of observed events in each bin of mass (width ~ 8 MeV). The dashed vertical lines demarcate the signal region.

a search: for example, scanning through the cuts in groups of 3 or 4, assuming that their optimal values are not correlated. On the other hand, it is valid to ask whether a method exists which would allow the phase space to be scanned through at a fraction of the time of a brute force search, and still be robust against local minima and cut correlations. A genetic algorithm which achieves precisely this, converging within ~ 1 minute while simultaneously scanning all cuts, is detailed here. It follows the approach described in [GJ03].

3.4.1 Genetic Algorithm

As the name suggests, the optimization algorithm works on the basis of mimicking the process of natural selection. The starting point for the algorithm is a population of cut-sets, each of which contains some values for the cuts to be optimized. In this case there are 13 cuts to optimize, hence a cut-set refers to a set of 13 cut values, one for each cut in the set. Each cut-set has an associated value for the parameter to be optimized, most typically the $\frac{S}{B}$ ratio or $\frac{S}{\sqrt{S+B}}$. These cut-sets are then “reproduced” to produce new cut-sets. Finally the population is culled by removing the cut-sets with the worst value of the optimization parameter. The process is repeated until the algorithm converges.

Starting Population

There are two obvious ways to select the starting values for the cut-sets: distribute them evenly through the phase space, or populate each cut-set randomly. Both approaches were tested and both converged, although an even distribution can be preferable if the starting population is very small. The suggested population size in [GJ03] is between 10 and 20 cut sets; values between 10 and 100 were tested, and convergence was always observed. The method proved to be robust, and converged regardless of the initial choice of cuts or other running conditions. For the purposes of the results presented later, a population size of 20 was used, and the starting points in phase space were selected randomly.

Reproduction

In order to produce the next generation of cut-sets, the original population of cut-sets is divided randomly into pairs⁴, and each pair of cut-sets then produces two new offspring cut-sets. In particular, for each cut within the cut-set, there is a 50% chance that the cut will be swapped between the parent cut-sets when creating the offspring. Therefore each offspring will, on average, contain half the cuts from one parent and half from the other. At this point, “mutation” is introduced into the population in order to protect against local minima, so that each cut within the offspring cut-sets has a fixed probability of changing to a random value. Higher mutation rates lead to slower convergence; for the purposes of this analysis, a mutation rate of 10% was chosen. The total population (parents+offspring)

⁴The population is kept as a list, and for the sake of simplicity cuts are paired off with those adjacent to them. In order to prevent excessive homogeneity in the population, the list is scrambled before the cuts are paired off.

is then ordered according to the value of the parameter to be optimized, and the worst half of the population is discarded. The remaining cut-sets become the new generation and are themselves reproduced, and this process continues until one is satisfied that the algorithm has converged.

Convergence

Convergence could in principle be defined according to the rate of change in the average value of the optimization parameter within the population; for the purposes of this study, however, the algorithm was simply run for a given number of generations and the output monitored by eye. A few thousand generations were found to be sufficient for the optimal cut values to be found, with no change occurring subsequently even when run over 100,000 generations or more. The algorithm took ~ 3 seconds per 1000 generations with a population size of 20 to run. In order to be sure that nothing was missed, the optimizer was run for 10,000 generations for the results presented below. One of the ancillary benefits of this method is that the population of cut-sets at convergence can contain degenerate cuts, which correspond to different ways of cutting on the event sample to produce the same background discrimination. The user can then choose between these degenerate cut-sets using non-statistical criteria, for example to minimize systematic biases associated with a particular type of cut. Another quality of degenerate cut-sets is that they make it obvious which cuts are critical to the optimal background discrimination obtained and which are not. The former will stay the same, or nearly so, between degenerate cut-sets, while the latter will vary between degenerate cut-sets for which the majority of other cuts stay the same. This allows the user to judge the position of the non-critical cuts correctly, instead of relying on luckily landing at the optimal value. In particular it protects against artificially high values for heavily correlated groups of cuts. Finally, note that all cuts obtained through this optimization are rounded off to the nearest appropriate value.

3.4.2 The Optimized Cut Values

Before an optimized set of cuts is found, a decision has to be taken which parameter should be used as the figure of merit. The statistical error on γ as extracted from this channel will vary with $\frac{S}{\sqrt{S+B}}$. On the other hand, a poorly understood background can lead to large systematic errors, which can potentially render a low statistical error irrelevant. Hence it was decided to optimize separately for $\frac{S}{\sqrt{S+B}}$ and $\frac{S}{B}$. The values of the cuts thus obtained are listed in Table 3.5. In order to produce a realistic set of cuts, the optimization is performed only on those events which pass all triggers, as described in Section 2.7. A breakdown of trigger efficiency is given in Table 3.6. The trigger system used for this study is expected to differ significantly from the trigger configuration which will be used once the detector is running. The L1 efficiency is measured on events which pass the L0, and the HLTGeneric efficiency on events which pass both the L0 and L1.

Throughout the following, the terms “annual” and “per year” refer to a nominal LHCb year of running, or 2 fb^{-1} of data.

Table 3.5: The list of optimized cuts for three different scenarios. In the leftmost column, the preselection cuts are listed again for ease of reference. In the third column, the cuts are optimized for $\frac{S}{\sqrt{S+B}}$. In the fourth column, they are optimized for the best $\frac{S}{B}$, while demanding a minimum yield of roughly 80% of the best possible yield. In the rightmost column, they are optimized for the best $\frac{S}{B}$, while demanding a minimum yield of roughly 50% of the best possible yield. The background events listed in the table are selected in the narrow mass window, as scaled from the appropriate events in the wide mass window. The signal events are those surviving the relevant selection on the **signal** tapes, and must be scaled appropriately before the quoted $\frac{S}{B}$ ratios can be reached.

Cut	Preselection	Best $\frac{S}{\sqrt{S+B}}$	Best $\frac{S}{B}$ (1)	Best $\frac{S}{B}$ (2)
All particle momenta	> 2000 MeV	> 2000 MeV	> 2000 MeV	> 2000 MeV
Kaon candidate DLL K-Pi	> -5.0	> 0.0	> 1.0	> 1.0
D^+ daughter p_t	> 300 MeV	> 300 MeV	> 300 MeV	> 350 MeV
D^+ p_t	> 2000 MeV	> 2000 MeV	> 2000 MeV	> 2800 MeV
D^+ daughter IPs	> 1.0	> 2.0	> 3.5	> 3.5
D^+ IPs	> 2.0	> 3.0	> 3.0	> 3.0
D^+ mass	± 50 MeV	± 21 MeV	± 21 MeV	± 21 MeV
D^+ vertex χ^2	< 20	< 20	< 10	< 10
D^+ FSPV	> 4.5	> 10	> 11.5	> 20
Bachelor π p_t	> 500 MeV	> 500 MeV	> 750 MeV	> 750 MeV
Bachelor π IPs	> 2.5	> 2.5	> 3.0	> 6.0
B_d^0 mass wide	± 500 MeV	± 500 MeV	± 500 MeV	± 500 MeV
B_d^0 mass tight	± 50 MeV	± 50 MeV	± 50 MeV	± 50 MeV
B_d^0 vertex χ^2	< 20	< 10	< 10	< 10
B_d^0 IPs	< 6.0	< 5.5	< 3.0	< 2.5
B_d^0 FSPV	> 0.0	> 2.5	> 5.0	> 7.0
B_d^0 $\cos \theta$	> 0.9999	> 0.9999	> 0.9999	> 0.9999
Signal Events	9901	2804	2206	1533
Background Events	1267	12	3.3	1.4
Annual Yield	4.7M $\pm$ 0.3M	1.34M $\pm$ 80k	1.05M $\pm$ 70k	724k $\pm$ 47k
$\frac{S}{B}$	~ 1.3	4.6 $\pm$ 1.4	12.6 $\pm$ 2.4	20 $\pm$ 6

Table 3.6: The efficiency for a given trigger level, as measured on signal passing the preselection. Each trigger is run in sequence, so the L1 efficiency is measured on events which pass the L0, and the HLT efficiency is measured on events which pass both the L0 and L1 triggers.

Trigger Level	Efficiency
L0	47.3%
L1	76.1%
HLTGeneric	93.9%
All triggers	33.8%

Events used for the optimization

As mentioned above, 350,000 signal and 33.9 million $b\bar{b}$ inclusive events were available for the optimization. Under normal circumstances, the signal events would be used to select signal and the $b\bar{b}$ inclusive events would be used to select background. A normalization factor proportional to the signal branching ratio would then be used to compute the S/B ratio and the signal significance, $\frac{S}{\sqrt{S+B}}$. However, the branching ratio of $B_d^0 \rightarrow D^\pm \pi^\mp$ is large enough that this procedure is not necessary – enough signal decays exist in the $b\bar{b}$ inclusive events to allow an optimization using $b\bar{b}$ inclusive events alone. This has the added advantage of closely resembling the procedure of analysing real data. Once the cuts have been optimized in this manner, the 350,000 signal events are used to calculate the expected signal yield with a greater precision.

Optimizing for $\frac{S}{\sqrt{S+B}}$

In the case of minimizing with respect to $\frac{S}{\sqrt{S+B}}$, 2804 events are selected on the signal tapes. On the background tapes, 12 events are selected in the narrow mass window. The distribution of background categories was given in Table 3.4, and the majority of the background events surviving are reflection. A more detailed discussion of the nature of this surviving background will be given in Section 3.4.3.

The expected signal yield is given by

$$\text{Signal} = \frac{N_{sel}}{N_{tot}} \cdot \eta_\theta \cdot 0.80 \cdot B.R. \cdot 10^{12}. \quad (3.2)$$

Here, $N_{sel} = 2804$ is the number of events selected on the signal tape; $N_{tot} = 148,000$ is the total number of signal events processed; $\eta_\theta = 0.347$ is the geometric acceptance for signal; 0.80 accounts for the fact that with a hadronization factor of 40%, 80 B_d^0 mesons are expected in every 100 $b\bar{b}$ events; $B.R. = 2.55 \cdot 10^{-4}$ is the branching ratio for this channel [ea06c]; 10^{12} is the number of $b\bar{b}$ events expected in one year of LHCb running. This corresponds to an annual yield of $(1.34 \pm 0.08) \cdot 10^6$ signal events.

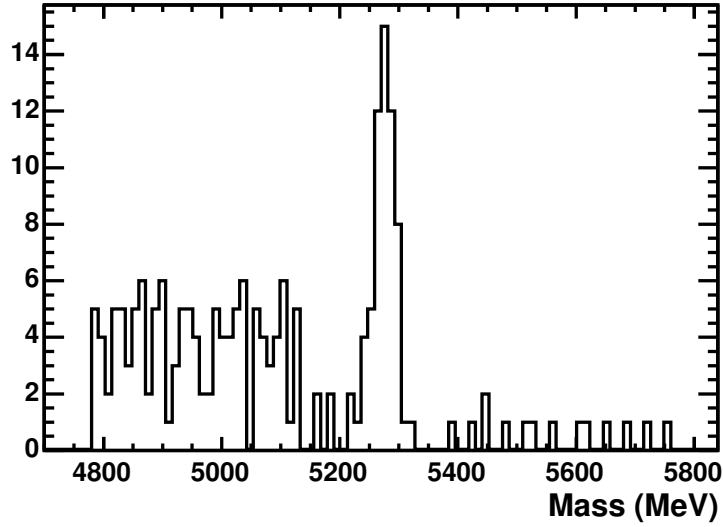


Figure 3.4: Invariant mass of candidates passing selection on the generic $b\bar{b}$ sample, showing both the signal peak and the surviving background.

The expected background yield is given by

$$\text{Background} = \frac{N_{sel}}{N_{tot}} \cdot \eta_{\theta} \cdot 10^{12}. \quad (3.3)$$

Here, $N_{sel} = 12$ is the number of background events selected on the background tape; $N_{tot} = 16,950,000$ is the total number of background events processed; $\eta_{\theta} = 0.43$ is the geometric acceptance for background; 10^{12} is the number of $b\bar{b}$ events expected in one year of LHCb running. This corresponds to an annual yield of $(300 \pm 87) \cdot 10^3$ background events.

These yields correspond to a signal to background ratio of $\frac{S}{B} = 4.6 \pm 1.4$. A mass plot in the wide mass window, obtained through this selection on the background tapes, is given in Figure 3.4. The optimal value of the signal significance, $\frac{S}{\sqrt{S+B}}$, is 0.91.

We can also confirm these numbers by referring to Table 3.4, where 59 signal events are listed as surviving the selection on the background tapes after the optimized final selection. Using Equation 3.2, the expected signal yield is $(1.48 \pm 0.19) \cdot 10^6$ events per annum. However, it is also necessary to correct for the fact that the LHCb Monte Carlo uses a branching fraction of $2.91 \cdot 10^{-4}$ for the channel $B_d^0 \rightarrow D^-(K^+\pi^-\pi^-)\pi^+$, which is higher than the current PDG value. Hence the actual expected signal yield from the background tapes would be $(1.30 \pm 0.17) \cdot 10^6$ events per annum after all triggers, which agrees with the previous estimate.

Optimizing for $\frac{S}{B}$

The above selection will be optimal with respect to statistical errors when this channel is used to study the extraction of γ and fitting of lifetimes. In a scenario where the background is discovered to contribute a significant systematic error to the results, however, it may be necessary to come up with a purer selection. Therefore the selection has been re-optimized,

this time with respect to $\frac{S}{B}$, with two modifications from the above procedure

- Because the background statistics are necessarily very low in the narrow mass window, these selections are optimized within the wide mass window of the B_d^0 . It will be assumed that the background level in the narrow mass window can be obtained by scaling the number of surviving events of all background categories except for the low mass background (category 50) in the wide mass window by a factor of 10, corresponding to the ratio of the sizes of the two mass windows.
- The values of the cuts in the selection optimized for $\frac{S}{\sqrt{S+B}}$ will be used as the starting point for this optimization, so as to ensure that the selections are subsets of each other.

Two scenarios are considered, one in which the yield is constrained to at least $\sim 80\%$ of the optimal yield of $(1.34 \pm 0.80) \cdot 10^6$ tagged signal events a year, and one in which yield is constrained to at least $\sim 50\%$ of this optimal yield. They are labelled scenario (1) and (2) in Table 3.5.

When optimizing for scenario (1), 2206 events are selected on the signal tapes. On the background tapes, 83 events are selected in the wide mass window. Of these, 50 are of type low-mass background, and are hence not expected to feature in the tight mass window. This leaves 33 background events, and since the tight mass window is 10 times narrower than the wide mass window, 3.3 background events are expected in the narrow mass window. In fact, 6 background events are observed in this window, which is not inconsistent with the statistically more precise indirect estimate.

Using Equation 3.2, the expected annual signal yield is $(1.05 \pm 0.07) \cdot 10^6$ events. Using Equation 3.3, and assuming 3.3 background events in the narrow mass window, the expected annual background yield is $(83 \pm 15) \cdot 10^3$ events per year. These yields correspond to a signal to background ratio of $\frac{S}{B} = 12.6 \pm 2.4$.

When optimizing for scenario (2), 1533 events are selected on the signal tapes. On the background tapes, 45 events are selected in the wide mass window. Of these, 31 are of type low-mass background, and are hence not expected to feature in the tight mass window. This leaves 14 background events, and since the tight mass window is 10 times narrower than the wide mass window, 1.4 background events are expected in the narrow mass window. In fact, 3 background events are observed in this window, which is again consistent with the statistically more precise indirect estimate.

Using Equation 3.2, the expected annual signal yield is $(724 \pm 47) \cdot 10^3$ events. Using Equation 3.3, and assuming 3.3 background events in the narrow mass window, the expected annual background yield is $(36 \pm 10) \cdot 10^3$ events per year. These yields correspond to signal to background ratio of $\frac{S}{B} = 20 \pm 6$.

3.4.3 Nature of the surviving background

In order to gain a deeper insight into the composition of that background which survives the final selection, and in particular the 12 events which survive when the selection is optimized

for best $\frac{S}{\sqrt{S+B}}$, it is necessary to consult the Monte Carlo “truth” content of the event. This was done by printing out the MC truth dumps for the entire event, and then comparing the MCParticles associated to the reconstructed final state particles with the MC truth dumps in order to determine the true character of the background events which were reconstructed as $B_d^0 \rightarrow D^+\pi^-$. The character of the 12 surviving events is as follows:

1. Four events of the type $B_d^0 \rightarrow D^+K^-$, with a correctly reconstructed D^+ and a kaon misidentified as a pion.
2. One event of the type $B_s^0 \rightarrow D_s^+ (K^+K^+\pi^-) K^-$, with two kaons misidentified as pions.
3. One event of the type $B_d^0 \rightarrow J/\psi (\mu\mu) K^* (K\pi)$. Here both muons were misidentified as pions, one was then combined with the $K\pi$ combination to form a D^- , while one was reconstructed as the bachelor pion.
4. One event of the type $\Lambda_b \rightarrow D_s^* (D_s (KK\pi) \gamma) p^+$, where the proton is misidentified as a pion, and combined with the D_s , a kaon having been misidentified as a pion to reconstruct the D_s as a D^- . The photon is ignored.
5. One event of the type $\Lambda_b \rightarrow \Lambda_c (K\pi p) \pi$, where the proton is mis-identified as a pion.
6. One event of the type $\Lambda_b \rightarrow \Lambda_c (K\pi p) K$, where the proton is mis-identified as a pion and the bachelor kaon is mis-identified as a pion.
7. One event in which a true D^- coming from the decay $B_d^0 \rightarrow D^* (D^- (K\pi\pi) \pi^0 (\gamma\gamma)) e^+ \nu_e$ is combined with a random pion from elsewhere in the event.
8. One event which is a combination of taking the D from a $B^+ \rightarrow D^* (2010) D_s (KK\pi) \pi$ decay, where one of the kaons in the D_s decays has been mis-identified as a pion so as to make a D^- , and a random pion from elsewhere in the collision.
9. One event which is in fact a signal event, but is not identified as such because of associator failure. This event has been counted as background in all subsequent analyses, leading to a slightly conservative estimate of the offline selection purity.

The greatest single source of background is composed of $B_d^0 \rightarrow D^+K^-$ events, where the bachelor kaon has been misidentified as a pion. Such background will not peak directly under the signal because of the mass differences between the misidentified particles, but it will peak in the lower mass sideband and protrude into the signal region. Figure 3.5 shows a comparison between the reconstructed mass of a $B_d^0 \rightarrow D^+K^-$ decay which has been misidentified as $B_d^0 \rightarrow D^+\pi^-$, and the reconstructed mass of a true $B_d^0 \rightarrow D^+\pi^-$ decay. The background peaks at around 5240 MeV and extends throughout the bottom half of the signal region. It will be especially difficult to estimate this background from the sidebands because the correctly reconstructed D^+ offers no help. Other sources of background which will show a peaking structure are:

- $B_s^0 \rightarrow D_s^+K^-$ events, which will peak around 5250 MeV, but which can be controlled by studying the reconstructed D mass sidebands, since the misidentified D_s will have a nearly flat distribution compared to the signal D^+ mass.

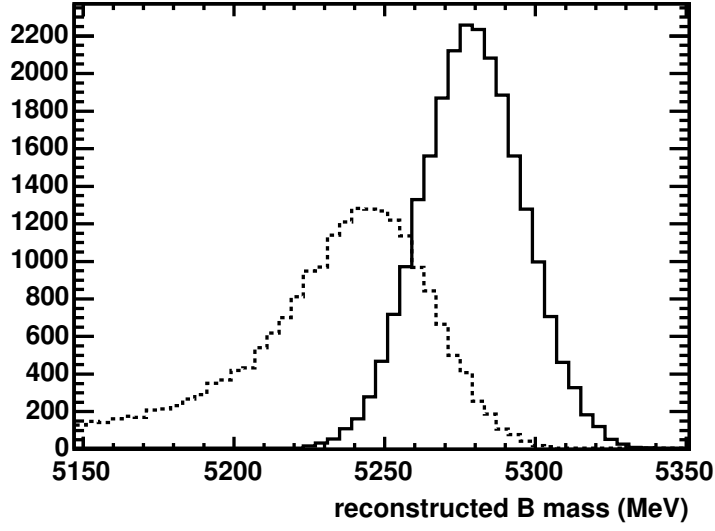


Figure 3.5: The reconstructed mass of $B_d^0 \rightarrow D^+ K^-$ decays misidentified as $B_d^0 \rightarrow D^+ \pi^-$ (dashed), shown against the reconstructed mass of true $B_d^0 \rightarrow D^+ \pi^-$ decays (solid). The total number of misidentified and true events plotted is the same.

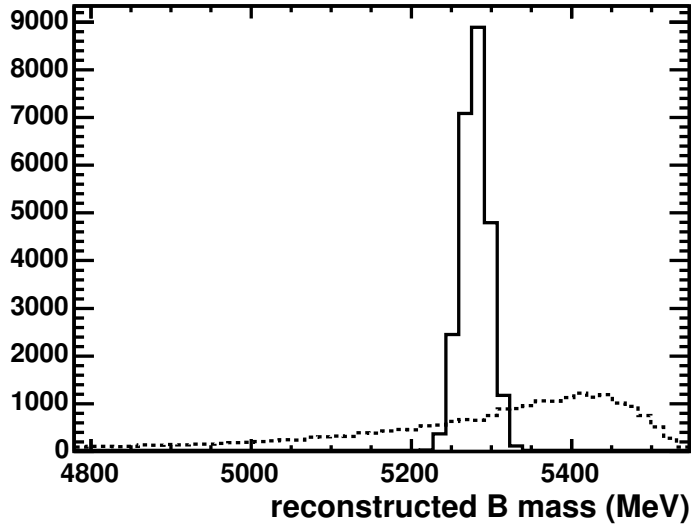


Figure 3.6: The reconstructed mass of $\Lambda_b \rightarrow \Lambda_c (K \pi p) \pi$ decays misidentified as $B_d^0 \rightarrow D^+ \pi^-$ (dashed), shown against the reconstructed mass of true $B_d^0 \rightarrow D^+ \pi^-$ decays (solid). The total number of misidentified and true events plotted is the same.

- $\Lambda_b \rightarrow \Lambda_c (K\pi p) \pi$ events, whose mass distribution is seen in Figure 3.6. Although this distribution is not flat, it does not peak in the signal region, and can be reasonably extrapolated from the sidebands,
- $\Lambda_b \rightarrow \Lambda_c (K\pi p) K$ events, which will peak around 5350 MeV, but which can be controlled by studying the D mass sidebands.

Since the majority of the surviving background involves another particle being identified as a pion (for example a kaon, muon or proton), future studies could seek to improve the purity of this selection by applying DLL cuts on the reconstructed pions, for example on the pion-kaon or muon-kaon separation. An upper cut on the kaon momentum could also help, since the RICH does not effectively discriminate between kaons and pions with momenta above 100 GeV.

3.4.4 Reconstruction efficiencies

It is desirable to compare the reconstruction efficiencies of the various selections presented above, and to attempt to disentangle the efficiency of the offline selection from the intrinsic efficiency associated with the detector and trigger system. The Decay of Interest (DOI) is the decay being reconstructed. Defining:

- Reconstructible event: an event in which it is possible to reconstruct every final state track from the DOI, as defined in [col07b];
- Reconstructed event: an event in which every final state track from the DOI has an associated MC-truth track;
- Reconstructed and reconstructible event: an event which satisfies both the above conditions.

The relevant numbers on the signal Monte Carlo sample are listed in Table 3.7. In order to separate the efficiencies associated with the detector and trigger from those associated with the offline selection, it is useful to define the following:

- $\epsilon_{\text{trg}/\text{sel}} = \frac{N_{\text{trg}}}{N_{\text{sel}}}$ is the fraction of DOI selected by the offline selection which pass the full trigger sequence, L0×L1×HLTGeneric.
- $\epsilon_{\text{sel}/\text{rec}} = \frac{N_{\text{sel}}}{N_{\text{rec'ed}}}$ is the fraction of reconstructed DOI selected by the offline selection.
- $\epsilon_{\text{rec}/\text{det}} = \frac{N_{\text{rec'ible/ed}}}{N_{\text{rec'ible}}}$ is the fraction of reconstructible DOI that are reconstructed.
- $\epsilon_{\text{off}} = \epsilon_{\theta} \cdot \frac{N_{\text{sel}}}{N_{\text{tot}}}$ is the efficiency of the offline selection: the fraction of total DOI selected by the offline selection, multiplied by the detector acceptance.
- $\epsilon_{\text{det}} = \frac{\epsilon_{\text{off}}}{\epsilon_{\text{rec}/\text{det}} \cdot \epsilon_{\text{sel}/\text{rec}}}$ represents the detector efficiency, including the geometric acceptance.

Table 3.7: The numbers of generated, reconstructible, reconstructed and selected Decays of Interest (DOI). Note that the preselection is run without executing the trigger, while the other three selections are run after the trigger.

Number of events	148,000
Generated DOI	148,013
Reconstructible DOI	21,741
Reconstructed DOI	20,318
Reconstructible and reconstructed DOI	16,649
DOI selected by preselection, before triggers	9,901
DOI selected by $\frac{S}{\sqrt{S+B}}$ selection and triggers	2,804
DOI selected by $\frac{S}{B}$ selection (1) and triggers	2,206
DOI selected by $\frac{S}{B}$ selection (2) and triggers	1,533

- $\epsilon_{\text{tot}} = \epsilon_{\text{det}} \cdot \epsilon_{\text{rec/det}} \cdot \epsilon_{\text{sel/rec}} \cdot \epsilon_{\text{trg/sel}}$ is the total efficiency, representing the fraction of all DOI in that are selected in 4π .

The efficiencies are listed in Table 3.8, along with the associated yields and $\frac{S}{B}$ ratios.

Table 3.8: Summary of signal reconstruction efficiencies, yields, and $\frac{S}{B}$ ratios.

Efficiency	Preselection	$\frac{S}{\sqrt{S+B}}$ selection	$\frac{S}{B}$ selection (1)	$\frac{S}{B}$ selection (2)
ϵ_{det}	6.2%	6.2%	6.2%	6.2%
$\epsilon_{\text{rec/det}}$	76.6%	76.6%	76.6%	76.6%
$\epsilon_{\text{sel/rec}}$	48.7%	38.2%	29.6%	19.4%
$\epsilon_{\text{trg/sel}}$	33.8%	36.1%	36.7%	39.1%
ϵ_{tot}	0.78%	0.66%	0.52%	0.36%
Annual Yield	4.7M $\pm$ 0.3M	1.34M $\pm$ 80k	1.05M $\pm$ 70k	724k $\pm$ 47k
$\frac{S}{B}$	~ 1.3	4.6 $\pm$ 1.4	12.6 $\pm$ 2.4	20 $\pm$ 6

3.4.5 Mass and Time Resolutions

The mass resolution of the reconstructed signal events is shown in Figure 3.7. A Gaussian fit to the mass distribution gives a width of 14.6 MeV. For the time dependent measurements envisaged with this channel, it is crucial that the time resolution is better than the B_d^0 oscillation period of 13.3 ps. The resolution is shown in Figure 3.8, and a double Gaussian fit to the distribution gives a core width of 33.9 fs, with a secondary width of 70.3 fs and a core fraction of 62.6%, which is certainly good enough. There is a small offset observed in

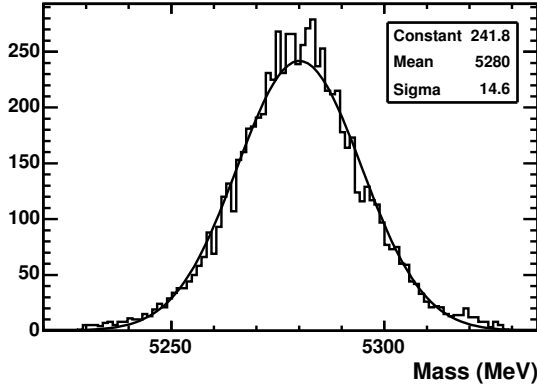


Figure 3.7: The mass resolution of signal reconstructed B_d^0 mesons.

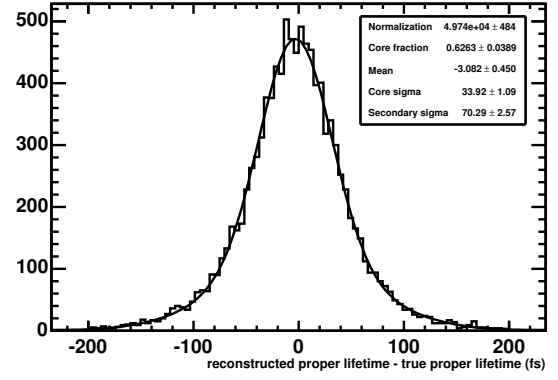


Figure 3.8: The lifetime resolution of signal reconstructed B_d^0 mesons.

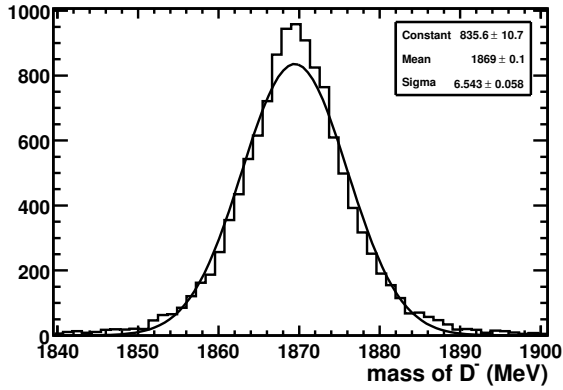


Figure 3.9: The mass of the reconstructed D^+ from the decay $B_d^0 \rightarrow D^+\pi^-$.

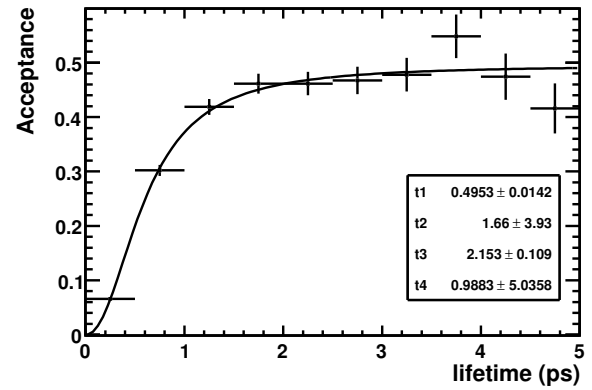


Figure 3.10: Fitted lifetime acceptance for signal $B_d^0 \rightarrow D^+\pi^-$ events passing the trigger and offline selection.

this distribution, however, which is caused by a biased primary vertex reconstruction. The ultimate source of this bias is the inclusion of tracks from secondary vertices when reconstructing the primary vertex [Xie05]. A plot of the fitted D^+ mass is shown in Figure 3.9. The mass resolution is 6.5 MeV.

3.4.6 Lifetime acceptance function

The L1 trigger contains impact parameter cuts, which will bias the distribution of selected B mesons against those with low lifetimes. This lifetime function, as measured on the signal Monte Carlo events passing the offline selection and the triggers is shown in Figure 3.10. With real data, the acceptance function will be measured using the methods presented in Chapter 6, with no recourse to Monte Carlo data. Nevertheless, a simple parametrization of the acceptance function, which can be used in the toy Monte Carlo studies in Chapters 4 and 5 is needed. The fitted parameterization is:

$$P_A(\tau) = 0.5 * \frac{(1.7\tau)^{2.2}}{1 + (1.7\tau)^{2.2}}. \quad (3.4)$$

The upper impact parameter cuts used in the trigger mean that such a parameterization will necessarily be inferior to a Monte Carlo free fit of the acceptance function on real data. Nonetheless, it was felt adequate for a toy Monte Carlo study, in particular because it describes the acceptance well in the region below three times the mean B_d^0 lifetime. Finally, since the time dependent trigger will almost certainly undergo significant changes before the experiment starts taking data (as well as during the running of the experiment), reproducing the current trigger acceptance function precisely is not crucial.

3.5 Conclusion

The channel $B_d^0 \rightarrow D^+\pi^-$ has been reconstructed using LHCb Monte Carlo data. A method for classifying the background surviving the selection cuts has been designed and implemented within the LHCb analysis framework. A genetic algorithm has been implemented to optimize the selection in a robust manner, and three selections have been developed with different yields and purities, allowing the channel to be used in analysis scenarios requiring high purity as well as those requiring high yields. As optimized for maximal $\frac{S}{\sqrt{S+B}}$, the channel is expected to have a yield of $1.34 \text{ M} \pm 80 \text{ k}$ fully triggered events in 2 fb^{-1} , with a signal to background ratio of 4.6 ± 1.4 . When optimized for purity, the channel is expected to have a yield of $724 \text{ k} \pm 47 \text{ k}$ fully triggered events in 2 fb^{-1} , with a signal to background ratio of 20 ± 6 .

Chapter 4

Extraction of the angle γ from the channel $B_d^0 \rightarrow D^- \pi^+$

4.1 Overview

As discussed in Section 1.4.2, the angle γ is currently the least well measured of the three angles in the CKM unitary triangle. The channel $B_d^0 \rightarrow D^- \pi^+$ allows for a theoretically clean extraction of γ , which is expected to be robust against new physics and thus provide a baseline measurement for new physics sensitive channels. Discrepancies between these measurements of γ will be a signature of non-Standard Model processes. Chapter 3 discussed the reconstruction of this channel at LHCb; Table 3.8 gives the selection efficiencies, yields and purities.

This chapter presents a “toy” Monte Carlo study performed to show the viability of the extraction of γ from this channel at LHCb and estimate the precision of such an extraction. Standard techniques are used for generating toy Monte Carlo events.

The method used to fit γ follows that used in [Rad01] to extract γ from the channel $B_d^0 \rightarrow D^{*-} \pi^+$, and the substantial points of the fitting method will be set out again here for the reader’s convenience.

4.2 Tagging performance

Measuring the CP asymmetry requires knowledge of the flavour of the decaying B_d^0 meson at birth, and this is obtained through flavour tagging. The tagging performance for this channel has not been explicitly measured with the LHCb simulation, but similar channels [CLM07] whose flavour tagging performance has been measured have a tagging efficiency of 50% and a mistag rate of 35%. It is expected that the mistag rate will be known to better than 1% (absolute) with 2 fb^{-1} of data. These values will be assumed throughout this study.

4.3 Fitting method

The angle γ can be extracted from the time dependent decay rate asymmetries:

$$A_\eta(\tau) = \frac{\Gamma(B_d^0 \rightarrow D^- \pi^+) - \Gamma(\bar{B}_d^0 \rightarrow D^- \pi^+)}{\Gamma(B_d^0 \rightarrow D^- \pi^+) + \Gamma(\bar{B}_d^0 \rightarrow D^- \pi^+)}, \quad (4.1)$$

and

$$A_{\bar{\eta}}(\tau) = \frac{\Gamma(\bar{B}_d^0 \rightarrow D^+ \pi^-) - \Gamma(B_d^0 \rightarrow D^+ \pi^-)}{\Gamma(\bar{B}_d^0 \rightarrow D^+ \pi^-) + \Gamma(B_d^0 \rightarrow D^+ \pi^-)} \quad (4.2)$$

as introduced in Section 1.2.2 . The asymmetries are defined such that the differences in the detection efficiencies of the two final states (if any) will cancel. The decay rates are given by:

$$\begin{aligned} \Gamma(B_d^0 \rightarrow D^- \pi^+) &\equiv R_{D^-}(\tau) = e^{-\Gamma\tau} \{ (1 + |\eta|^2) - (1 - |\eta|^2) \cos \Delta m \tau + 2 \text{Im}(\eta) \sin \Delta m \tau \}; \\ \Gamma(\bar{B}_d^0 \rightarrow D^- \pi^+) &\equiv \bar{R}_{D^-}(\tau) = e^{-\Gamma\tau} \{ (1 + |\eta|^2) + (1 - |\eta|^2) \cos \Delta m \tau - 2 \text{Im}(\eta) \sin \Delta m \tau \}; \\ \Gamma(\bar{B}_d^0 \rightarrow D^+ \pi^-) &\equiv \bar{R}_{D^+}(\tau) = e^{-\Gamma\tau} \{ (1 + |\bar{\eta}|^2) - (1 - |\bar{\eta}|^2) \cos \Delta m \tau - 2 \text{Im}(\bar{\eta}) \sin \Delta m \tau \}; \\ \Gamma(B_d^0 \rightarrow D^+ \pi^-) &\equiv R_{D^+}(\tau) = e^{-\Gamma\tau} \{ (1 + |\bar{\eta}|^2) + (1 - |\bar{\eta}|^2) \cos \Delta m \tau + 2 \text{Im}(\bar{\eta}) \sin \Delta m \tau \}. \end{aligned}$$

Consequently the asymmetries, in the absence of any detector effects or background, are given by:

$$A_\eta(\tau) = -\frac{(1 - |\eta|^2)}{(1 + |\eta|^2)} \cos(\Delta m \tau) + \frac{2 \text{Im}(\eta)}{(1 + |\eta|^2)} \sin(\Delta m \tau); \quad (4.3)$$

$$A_{\bar{\eta}}(\tau) = -\frac{(1 - |\eta|^2)}{(1 + |\eta|^2)} \cos(\Delta m \tau) - \frac{2 \text{Im}(\bar{\eta})}{(1 + |\eta|^2)} \sin(\Delta m \tau). \quad (4.4)$$

Because the magnitude of η assumed in Section 1.4.2, $|\eta| = 0.024$, is so small, the cosine terms will have amplitudes very close to 1, while the sine terms will be small. The dependence on γ comes through the parameters η and $\bar{\eta}$, which are given by:

$$\begin{aligned} \eta &= |\eta| e^{i(\Delta_{\text{qcd}} + (\phi_{\text{mix}} - \gamma))}, \\ \bar{\eta} &= |\eta| e^{i(\Delta_{\text{qcd}} - (\phi_{\text{mix}} - \gamma))}. \end{aligned}$$

For fitting purposes, it will be useful to define the parameters:

$$\phi_{\pm} = \Delta_{\text{qcd}} \pm (\phi_{\text{mix}} - \gamma), \quad (4.5)$$

and

$$u_+ \equiv \sin \phi_+ = \frac{\text{Im}(\eta)}{|\eta|}, \quad (4.6)$$

$$u_- \equiv \sin \phi_- = \frac{\text{Im}(\bar{\eta})}{|\eta|}. \quad (4.7)$$

In order to extract γ from the asymmetries, it is necessary to determine the value of $|\eta|$, $\text{Im}(\eta)$, and $\text{Im}(\bar{\eta})$. However, $|\eta|$ is too small to be fitted successfully with the number of

events expected at LHCb. $|\eta|$ will therefore have to be constrained from external sources, and the likely constraints on $|\eta|$ were discussed in Section 1.4.2. An unbinned log-likelihood fit is performed to the asymmetries using ROOT [BR97] and the minimization package MINUIT [JR75]. The log-likelihood function to be maximized is:

$$\begin{aligned} \ln L = & \sum_{\text{all } B_d^0 \rightarrow D^- \pi^+} \ln(p_\eta(\tau_i)) + \sum_{\text{all } \bar{B}_d^0 \rightarrow D^- \pi^+} \ln(1 - p_\eta(\tau_i)) \\ & + \sum_{\text{all } \bar{B}_d^0 \rightarrow D^+ \pi^-} \ln(p_{\bar{\eta}}(\tau_i)) + \sum_{\text{all } B_d^0 \rightarrow D^+ \pi^-} \ln(1 - p_{\bar{\eta}}(\tau_i)). \end{aligned} \quad (4.8)$$

$p_\eta(\tau_i)$ is the probability to find a $B_d^0 \rightarrow D^- \pi^+$ event at time τ_i rather than a $\bar{B}_d^0 \rightarrow D^- \pi^+$ event, and $p_{\bar{\eta}}(\tau_i)$ is the probability to find a $\bar{B}_d^0 \rightarrow D^+ \pi^-$ event at time τ_i rather than a $B_d^0 \rightarrow D^+ \pi^-$ event. In terms of the asymmetries, p_η and $p_{\bar{\eta}}$ are given by:

$$p_{\eta, \bar{\eta}}(\tau) = \frac{A_{\eta, \bar{\eta}}(\tau) + 1}{2}. \quad (4.9)$$

As stated above, the value of $|\eta|$ will have to be constrained during the fit. In order to account for the uncertainty in the knowledge of $|\eta|$, this constraint is implemented by multiplying the likelihood function with a Gaussian:

$$L_{\text{constrained}} = L \cdot e^{-\frac{(|\eta|_{\text{estim}} - |\eta|_{\text{fit}})^2}{2\sigma_{|\eta|}^2}} \quad (4.10)$$

where $|\eta|_{\text{estim}}$ is the assumed value of $|\eta|$, $|\eta|_{\text{fit}}$ is the fit parameter, and $\sigma_{|\eta|}$ is the uncertainty associated with $|\eta|_{\text{estim}}$.

4.4 Including detector effects

In order to add realism to the study, the following effects were considered:

- Time resolution: The reconstructed decay time in this channel will be subject to an experimental resolution. The resolution measured in Chapter 3 was a double Gaussian with a core resolution of 33 fs. For the sake of simplicity, however the resolution in this toy Monte Carlo study was modelled by a single Gaussian with a resolution of 40 fs.
- Acceptance function: The impact parameter and flight separation cuts used in the trigger and the offline channel selection will introduce a bias against events with a low B lifetime, referred to as a proper time acceptance.
- Mistag: The mistagging of B_d^0 mesons as $\bar{B}_d^0$ and vice versa dilutes the asymmetry. In the majority of time-dependent CP analyses, the mistag rate is determined from control channels, which can introduce systematic uncertainties in the fitting. It will be shown that, in our analysis, the mistag rate can be fitted from the asymmetries themselves, thus eliminating this potential source of error.

- Background: A signal to background level of $\frac{S}{B} = 4.6$ is assumed within a 3σ mass window, taken from the optimized selection in Chapter 3.

These are now discussed in turn.

4.4.1 Time resolution

Although the time resolution is a double Gaussian, for the purposes of the toy Monte Carlo analysis here it will be modelled by a single Gaussian corresponding to the core resolution. After an event has been generated, the time resolution is applied to its proper lifetime. If the smeared lifetime becomes negative, the event is discarded. During the fitting, the time resolution is ignored, enabling us to estimate any systematic biases associated with it.

4.4.2 Acceptance function

The acceptance function used to generate events in the toy Monte Carlo is:

$$P_A(\tau) = 0.5 \cdot \frac{(1.7\tau)^{2.2}}{1 + (1.7\tau)^{2.2}}.$$

as shown in Section 3.4.6. The acceptance function is not used when fitting, since it cancels out in the expression for the asymmetries.

4.4.3 Mistag

A mistag fraction ω will lead to B_d^0 mesons being identified as $\overline{B}_d^0$ mesons, and hence the decay rates for these two flavours of B^0 to decay into the same final state will be mixed together. Hence the apparent decay rate, $R_{\omega D^-}$, of the decays tagged as B_d^0 into $D^-\pi^+$ is given by:

$$R_{\omega D^-}(\tau) = (1 - \omega)R_{D^-}(\tau) + \omega\overline{R}_{D^-}(\tau), \quad (4.11)$$

and similarly for the other rates. The measured asymmetry is now diluted:

$$A_{\omega}(\tau) = (1 - 2\omega)A(\tau), \quad (4.12)$$

compared to $A(\tau)$, the asymmetry for the case of perfect tagging.

4.4.4 Background

The background in this study is assumed to be independent of the decay considered, so that a background $D^-\pi^+$ or $D^+\pi^-$ combination is equally likely to be identified as coming from a B_d^0 as it is to be identified as coming from a $\overline{B}_d^0$. It is assumed that the acceptance function will be the same for signal and background, and that the background will follow the $e^{-\Gamma t}$ distribution. These assumptions are justified by the fact that the major component of the

background in this channel is the decay $B_d^0 \rightarrow D^\mp K^\pm$, with a correctly identified $D^\mp$, as seen in Chapter 3. In addition, even if these assumptions were to prove less than perfect, the high purity of this channel means that the background is not likely to be a significant source of statistical error in this extraction. Therefore, in the presence of a background to signal ratio of $\frac{B}{S}$, the apparent decay rate, $R_{\frac{B}{S}D^-}(\tau)$, of events tagged as B_d^0 into $D^-\pi^+$ is given by:

$$R_{\frac{B}{S}D^-}(\tau) = R_{D^-}(\tau) + \frac{B}{S} \cdot \frac{1}{2} (\bar{R}_{D^-}(\tau) + R_{D^-}(\tau)), \quad (4.13)$$

and similarly for the other rates. The measured asymmetry becomes:

$$A_{\frac{B}{S}}(\tau) = \frac{1}{1 + B/S} A(\tau), \quad (4.14)$$

compared to $A(\tau)$, the asymmetry without background.

4.4.5 The full asymmetry expressions

In the presence of the proper time acceptance, mistag, and background, the measured decay rates, $R_{mD^-}(\tau)$ etc., are:

$$\begin{aligned} R_{mD^-}(\tau) &= P_A(\tau) \cdot \left((1 - \omega) R_{D^-}(\tau) + \omega \bar{R}_{D^-}(\tau) + \frac{B}{S} \cdot \frac{1}{2} (\bar{R}_{D^-}(\tau) + R_{D^-}(\tau)) \right); \\ \bar{R}_{mD^-}(\tau) &= P_A(\tau) \cdot \left((1 - \omega) \bar{R}_{D^-}(\tau) + \omega R_{D^-}(\tau) + \frac{B}{S} \cdot \frac{1}{2} (\bar{R}_{D^-}(\tau) + R_{D^-}(\tau)) \right); \\ R_{mD^+}(\tau) &= P_A(\tau) \cdot \left((1 - \omega) R_{D^+}(\tau) + \omega \bar{R}_{D^+}(\tau) + \frac{B}{S} \cdot \frac{1}{2} (\bar{R}_{D^+}(\tau) + R_{D^+}(\tau)) \right); \\ \bar{R}_{mD^+}(\tau) &= P_A(\tau) \cdot \left((1 - \omega) \bar{R}_{D^+}(\tau) + \omega R_{D^+}(\tau) + \frac{B}{S} \cdot \frac{1}{2} (\bar{R}_{D^+}(\tau) + R_{D^+}(\tau)) \right). \end{aligned}$$

It should be noted from the above that the proper time acceptance, $P_A(\tau)$, will cancel out of any asymmetry. Nonetheless, it is important to understand the shape of the acceptance function, because as will be shown in Section 4.7, it will affect the precision with which the fit parameters can be measured. It is also possible that the acceptance function as measured on real data could be different for signal and background events, in which case it will have to be accounted for in the fit. The measured asymmetry now becomes:

$$A_m(\tau) = \frac{1 - 2\omega}{1 + B/S} A(\tau) = D A(\tau), \quad (4.15)$$

where $A(\tau)$ is again the asymmetry in the absence of any detector effects, and

$$D = \frac{1 - 2\omega}{1 + B/S}, \quad (4.16)$$

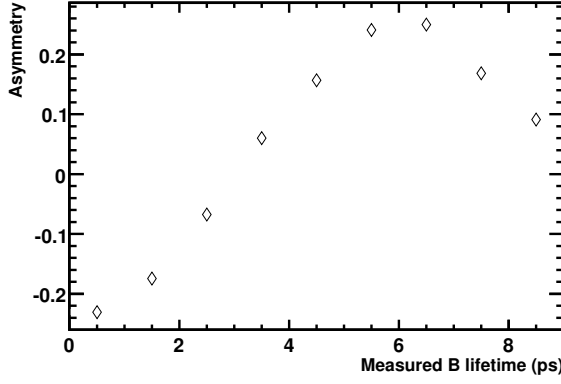


Figure 4.1: $A_{m,\eta}(\tau)$ with detector effects included.

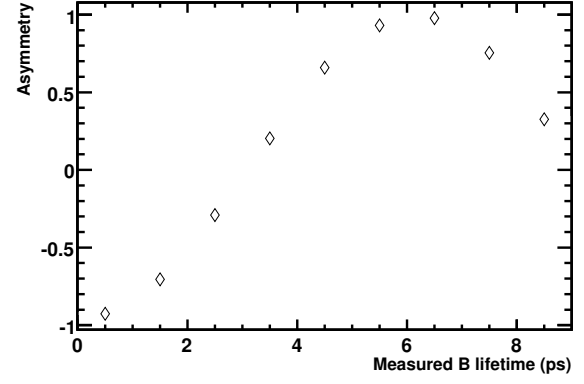


Figure 4.2: $A_{m,\eta}(\tau)$ in the absence of any detector effects.

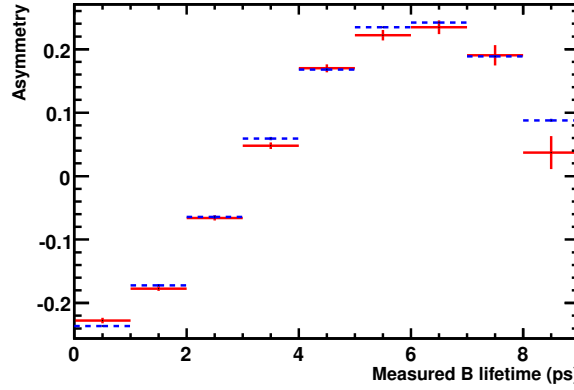


Figure 4.3: The generated asymmetry (solid red) and the fitted asymmetry (dashed blue).

represents the combined dilution of the asymmetry from the mistag and background. Written out explicitly, using Equation 4.3, this is:

$$A_{m,\eta}(\tau) = D \left(-\frac{(1 - |\eta|^2)}{(1 + |\eta|^2)} \cos(\Delta m \tau) + \frac{2\text{Im}(\eta)}{(1 + |\eta|^2)} \sin(\Delta m \tau) \right); \quad (4.17)$$

$$A_{m,\bar{\eta}}(\tau) = D \left(-\frac{(1 - |\eta|^2)}{(1 + |\eta|^2)} \cos(\Delta m \tau) - \frac{2\text{Im}(\bar{\eta})}{(1 + |\eta|^2)} \sin(\Delta m \tau) \right). \quad (4.18)$$

In order to fit correctly in the presence of these detector effects, it is necessary to determine the background and mistag levels in this channel. The background is normally determined from the mass sidebands, and the mistag from control channels, but in this case the dilution due to detector effects, given by the ratio $\frac{1-2\omega}{1+B/S}$, can be fitted directly alongside the CP violating parameters, since the four decay rates provide four observables and $|\eta|$ is constrained from external sources. It is not possible to fit $\text{Im}(\eta)$, $|\eta|$, and the dilution at the same time because the dilution would be maximally correlated with the sine and cosine terms in the asymmetry, but fixing the cosine term allows the dilution to be disentangled from the magnitude of the asymmetry sine term. In principle this means that the systematic errors associated with poor external knowledge of the mistag rate, and a badly understood

Table 4.1: Values of parameters and constants used in the toy Monte Carlo.

Parameter	Value
ω	35%
$ \eta $	0.024
$\frac{S}{B}$	5.8
Δm	0.507 ps ⁻¹
Γ	0.653 ps ⁻¹
$P_A(\tau)$	measured on MC signal events

background under the mass peak, can be avoided. Because $B_d^0 \rightarrow D^+\pi^-$ is expected to have a high purity, the background level was assumed to be well understood in this study, allowing the mistag rate to be fitted alongside $\text{Im}(\eta)$ and $\text{Im}(\bar{\eta})$.

It is also possible to do the reverse, and fit to the background level in the circumstances where the mistag rate is well understood from external sources. Both approaches offer potential gains in other analyses. For example, comparing the background level under the peak as extrapolated from the sidebands to the fitted background level from the asymmetries could offer valuable hints on the presence of peaking background under the signal, which could cause a systematic error in the fitted B_d^0 lifetime. It was shown in Chapter 3 that a major component of the surviving background are misidentified $B_d^0 \rightarrow D^+K^-$ decays, which will peak in the mass sideband below the signal and protrude into the signal region. If the mistag is sufficiently well understood from the control channels, it may be possible to divide the signal region in bins of mass, and fit to the background level in each bin separately from them, thus building the shape of the background without any recourse to Monte Carlo data. Alternatively, if the background level can be well understood through other means, a fit to the mistag may allow $B_d^0 \rightarrow D^-\pi^+$ to be used as a control channel for other channels with similar kinematics.

The final fit parameters chosen are (ω, u_+, u_-) . Table 4.1 lists the values of the various constants and parameters used as input in the toy simulation. Figures 4.1 and 4.2 show the asymmetry with and without the detector effects listed above. Figure 4.3 shows the generated and fitted asymmetry, which agree well with each other.

Calculating the error on γ

In the case when $|\eta|$ is known exactly, u_+ and u_- are uncorrelated, otherwise they have a covariance proportional to the error on $|\eta|$. The relevant results from [Rad01] are reproduced below:

$$\text{cov}(u_+, u_-) = u_+ u_- \left(\frac{\sigma_{|\eta|}}{|\eta|} \right)^2, \quad (4.19)$$

Table 4.2: A summary of the expected statistical errors on the observables u_+, u_- , and ω in the channel $B_d^0 \rightarrow D^- \pi^+$, for one and five years of data. The relevant pull plots are unbiased in all cases, with a width of 1.

	ω	u_+	u_-
σ_{1y}	0.1%	0.200	0.200
σ_{5y}	0.05%	0.089	0.089
σ_{1y} without fit to mistag	N/A	0.196	0.196
σ_{5y} without fit to mistag	N/A	0.088	0.088

and

$$\sigma_\gamma = \frac{1}{2} \left(\left(\frac{\sigma_{u_+}}{\cos \phi_+} \right)^2 + \left(\frac{\sigma_{u_-}}{\cos \phi_-} \right)^2 - 2 \tan \phi_+ \tan \phi_- \frac{\sigma_{|\eta|}}{|\eta|} \right)^{\frac{1}{2}}. \quad (4.20)$$

The error on γ is seen to be dependent on the values of $\phi_\pm$, and hence on the values of Δ_{qcd} and $\phi_{\text{mix}} - \gamma$. It should be noted that in regions of parameter space where ϕ_+ or ϕ_- approach 0, the error on γ can no longer be calculated using this formula, and if the real values of Δ_{qcd} and $\phi_{\text{mix}} - \gamma$ do indeed correspond to such a region of parameter space, another method for evaluating the error will be required.

4.5 Fit results

Three fitting scenarios were considered:

1. One year and five years of data taking with an exact knowledge of $|\eta|$.
2. One year and five years of data taking with an uncertainty on $|\eta|$ of 10%, corresponding to the current precision on the ratio of the associated CKM matrix elements, and perfect cancellation of the hadronic contributions.
3. One year and five years of data taking with an uncertainty on $|\eta|$ of 30%.

Table 4.2 summarizes the expected precisions on the observables u_+ , u_- , and ω after one and five years of data taking, for the case where the mistag is fitted, and where the mistag is fixed to its generator value. Note that leaving the mistag as a free parameter in the fit leaves the statistical precision on the variables u_+, u_- almost unchanged. Figure 4.4 shows the error on $\phi_{\text{mix}} - \gamma$ as a function of the value of $\phi_{\text{mix}} - \gamma$, for different values of Δ_{qcd} , where the results from scenarios 1, 2, and 3 above have been overlaid to demonstrate the degradation in performance due to an uncertainty in the knowledge of $|\eta|$. The error on $\phi_{\text{mix}} = -2\beta$, as calculated from the error given on $\sin(2\beta)$ in Chapter 2 is:

$$\sigma_{\phi_{\text{mix}}} \sim 2.0^\circ \quad (4.21)$$

The precision on this parameter will improve further by the time LHCb starts, and precision measurements of ϕ_{mix} will also be made at LHCb itself. As a result, the error on ϕ_{mix} will be considered negligible, so that $\sigma_\gamma = \sigma_{\phi_{\text{mix}} - \gamma}$ in the following discussion. It can be seen from this figure that the error on γ is highly dependent on the values of γ and Δ_{qcd} , and that certain areas of parameter space are significantly more favourable for this analysis than others. Before we discuss the precision that can be expected on γ , however, a word on the other ways of measuring γ with which this analysis will be competing once LHCb starts taking data.

Competing measurements of γ

Among tree level measurements of γ at LHCb, perhaps the most prominent is the analysis of the charged B meson decays $B^\pm \rightarrow D^0 K^\pm$. It exploits the interference between the tree level diagrams for the decays $B^\pm \rightarrow D^0 K^\pm$ and $B^\pm \rightarrow \bar{D}^0 K^\pm$, where the D^0 and $\bar{D}^0$ are required to decay into a common state. CP eigenstates such as KK , non-CP eigenstates such as $K\pi$, and self-conjugate states have all been considered. The extraction of γ using this strategy has been studied extensively at LHCb [Wil05, Pow07, Pat07, AG07], and is expected to provide a sensitivity of 5° or better on γ with 2 fb^{-1} of data, depending on parameters such as the ratios of interfering diagrams between the decay channels $B^\pm \rightarrow D^0 K^\pm$ and $B^\pm \rightarrow \bar{D}^0 K^\pm$.

Precision on γ and ω

First consider the case where $\gamma = 60^\circ$, or $\phi_{\text{mix}} - \gamma \sim 74^\circ$. For large values of Δ_{qcd} , assuming a 10% uncertainty on the value of $|\eta|$, it will be possible to achieve a precision of 8° with 2 fb^{-1} of data, or 4° with 10 fb^{-1} . Even allowing for a 30% uncertainty on the value of $|\eta|$, it will be possible to achieve a precision of 10° with 2 fb^{-1} of data, or 8° with 10 fb^{-1} . Here it can be seen that the error on $|\eta|$ saturates the error on γ . A value of Δ_{qcd} close to 0° would, on the other hand, make the extraction uncompetitive for the currently favoured Standard Model value of $\gamma = 60^\circ$. Note that the infinities seen in Figure 4.4 are not physical but an artefact of the procedure used to calculate the error on γ .

The error on ω is independent of the values of $\phi_{\text{mix}} - \gamma$ and Δ_{qcd} . After one year it will be:

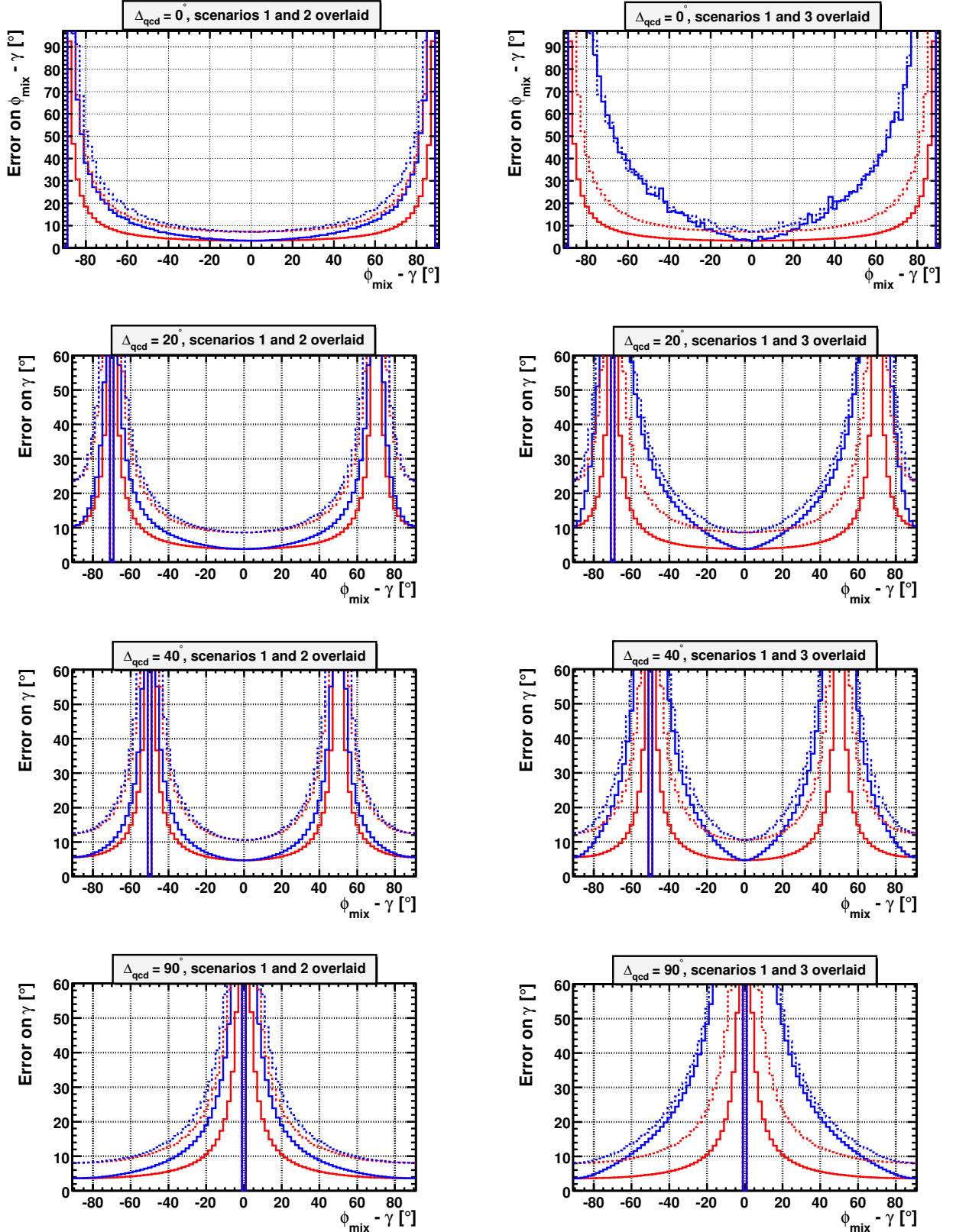
$$\sigma_\omega = 0.1\%, \quad (4.22)$$

and after five years:

$$\sigma_\omega = 0.05\%. \quad (4.23)$$

The error on ω is also independent of the error on $|\eta|$. Although the nature of systematic errors associated with extracting the mistag with an imperfect knowledge of the background are unknown at present, the statistical precision is rather competitive compared with the expected precision from control channels quoted in Section 4.2.

Figure 4.4: The errors on $\phi_{\text{mix}} - \gamma$ plotted as a function of $\phi_{\text{mix}} - \gamma$, for different combinations of scenarios 1, 2, and 3. In all the plots, broken lines show one year of data taking, solid lines show five years of data taking. Red (lighter) lines are for perfect knowledge of $|\eta|$, blue (darker) lines are for fits with the appropriate error on $|\eta|$. The plots for $\Delta_{\text{qcd}} = 0^\circ$ were created with the toy Monte Carlo simulation, while the remaining plots were generated from Equations 4.19 and 4.20.



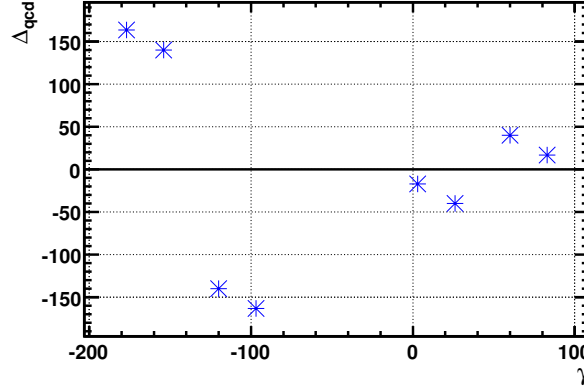


Figure 4.5: A plot of the ambiguous solutions for γ and Δ_{qcd} for true $\gamma = 60^\circ$ and $\Delta_{\text{qcd}} = 40^\circ$

Table 4.3: A list of the ambiguous solutions for γ and Δ_{qcd} , in the case when the “true” values are $\gamma = 60^\circ$ and $\Delta_{\text{qcd}} = 40^\circ$.

γ	-177°	-154°	-120°	-97°	3°	26°	60°	83°
Δ_{qcd}	163.5°	140°	-140°	-163°	-17°	-40°	40°	17°

Ambiguities in the extracted value of γ

Equations 4.5 and 4.6, define the relationship between the fitted parameters $\frac{\text{Im}(\eta)}{|\eta|}$, $\frac{\text{Im}(\bar{\eta})}{|\eta|}$ and the value of $\phi_{\text{mix}} - \gamma$, which is being extracted. Ambiguities arise in this extraction from the fact that it is necessary to eliminate the hadronic phase, Δ_{qcd} , from these equations, so as to extract the weak phase information in isolation. Writing:

$$\langle u \rangle_+ = \frac{u_+ + u_-}{2} = \sin(\Delta_{\text{qcd}}) \cos(\phi_{\text{mix}} - \gamma), \quad (4.24)$$

$$\langle u \rangle_- = \frac{u_+ - u_-}{2} = -\cos(\Delta_{\text{qcd}}) \sin(\phi_{\text{mix}} - \gamma), \quad (4.25)$$

the term $\sin(\phi_{\text{mix}} - \gamma)$ can be extracted from:

$$\sin^2(\phi_{\text{mix}} - \gamma) = \frac{1}{2} \left[(1 + \langle u \rangle_-^2 - \langle u \rangle_+^2) \pm \sqrt{(1 + \langle u \rangle_-^2 - \langle u \rangle_+^2)^2 - 4 \langle u \rangle_-^2} \right]. \quad (4.26)$$

Hence the formula corresponds to a fourfold solution for $\sin^2(\phi_{\text{mix}} - \gamma)$, and therefore an eightfold solution for γ itself. Figure 4.5 shows the situation for an input value of $\gamma = 60^\circ$, $\Delta_{\text{qcd}} = 40^\circ$, with the ambiguous solutions for γ and Δ_{qcd} listed in Table 4.3. In principle, external knowledge of the value of Δ_{qcd} could help to resolve some of the ambiguities, but notice that some of the solutions are separated by a distance in “ γ -space” similar to the one year precision on γ of 10° . Such solutions would be impossible to resolve, leading to a greater error on γ than that associated with each individual solution.

4.7 Statistical and systematic errors

The statistical precision on γ will depend not only on the number of reconstructed events, but also on the level of background present, the shape of the lifetime acceptance function, the mistag level and other detector or model effects. These detector effects can also induce systematic errors in the fit, which must be understood in order to correctly interpret the fitted value of γ .

4.7.1 Statistical error

At this point, it is traditional to quote the dependence of the statistical error on the detector effects which contribute to it. If the reader will indulge a brief digression, it will be demonstrated that the form of this dependence usually quoted is a simplification, and that it will only hold for certain values of the mistag and B/S ratio.

Number of events

The statistical error varies in the first instance with the number of reconstructed events:

$$\sigma_\gamma \propto \frac{1}{\sqrt{N_{S+B}}}, \quad (4.27)$$

where N_{S+B} is the total number of signal and background events reconstructed.

Background

According to Equation 4.14, in the presence of background, the asymmetry becomes:

$$A_{\frac{B}{S}}(\tau) = \frac{1}{1 + B/S} A(\tau).$$

Combined with Equation 4.27, the following dependence is obtained for the statistical error:

$$\sigma_\gamma \propto \sqrt{\frac{1 + B/S}{S}}, \quad (4.28)$$

where N_S is the number of reconstructed signal events.

Mistag

According to Equation 4.12, the mistag rate modifies the asymmetry:

$$A_\omega(\tau) = (1 - 2\omega)A(\tau).$$

Combined with Equation 4.28, the dependence of the statistical error on mistag and background usually quoted in the literature is arrived at:

$$\sigma_\gamma \propto \frac{1}{1 - 2\omega} \sqrt{\frac{1 + B/S}{S}}. \quad (4.29)$$

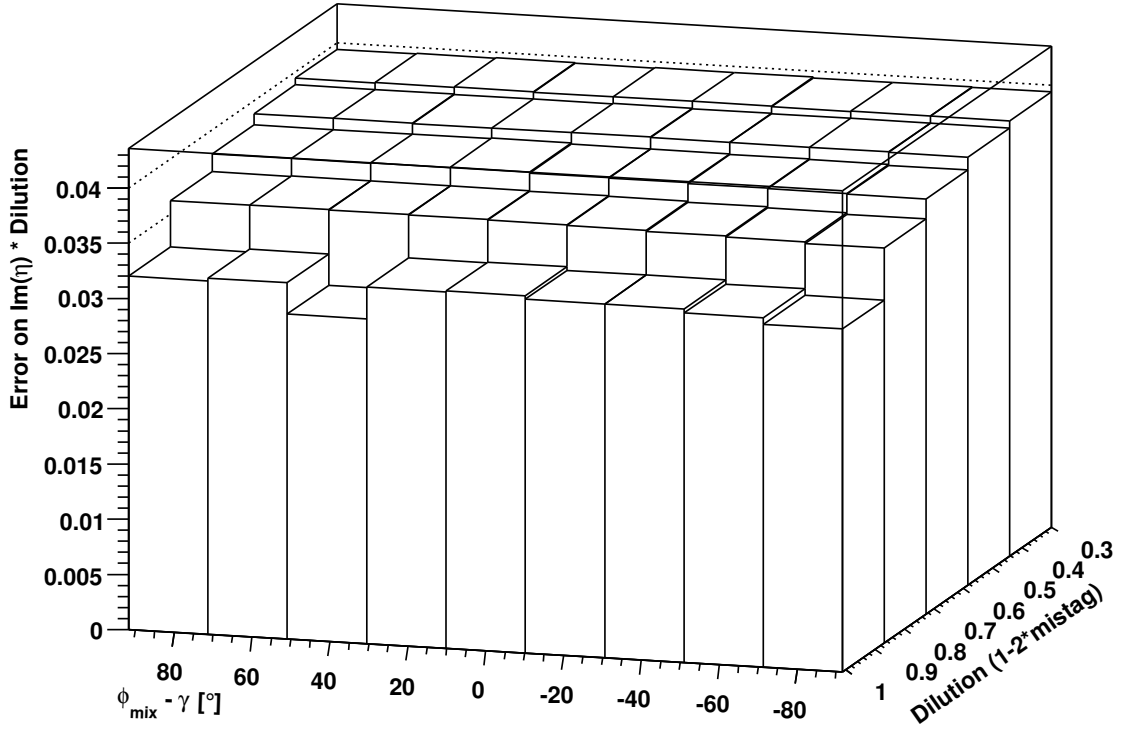


Figure 4.6: The error on $\text{Im}(\eta)$ as a function of the dilution and angle $\phi_{\text{mix}} - \gamma$.

This formula is a simplification, however, which will only hold for high values of the mistag. To see this, consider the covariance error matrix which is inversely proportional to the second derivative of the log-likelihood:

$$E_{ij} = - \left(\frac{\partial^2 \log L}{\partial p_i \partial p_j} \right)^{-1}. \quad (4.30)$$

Here, $p_{i,j}$ are the parameters varied in the fit. If the asymmetries given in Equation 4.17 are used to calculate this error matrix, the resulting error on γ is seen to have a non-trivial dependence on the dilution D , defined in Equation 4.16. The analytic form of this dependence is difficult to untangle, since it depends on the sum of the log-likelihoods over the reconstructed events. It is, however, possible to plot the variation of this error as a function of $\phi_{\text{mix}} - \gamma$ and the dilution D , by performing numerous toy Monte Carlo experiments. Figure 4.6 shows this variation. The vertical axis is the error multiplied by the dilution, and if Equation 4.29 held exactly, the plot would be flat, with a z-value given by the value for a dilution of 1 on the current plot. Since the purity in this channel is so high, the term $\sqrt{1 + B/S}$ will be approximately equal to 1, and hence will not affect the error. Therefore, it can be seen that Equation 4.29 holds only for high values of the mistag, but across all values of ϕ_{mix} . The likely values of the mistag at LHCb are high enough that Equation 4.29 will hold in their vicinity.

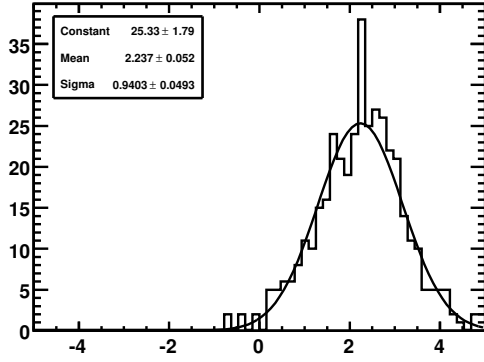


Figure 4.7: Pull plot on u_+ for an absolute bias on the mistag of 1.0%.

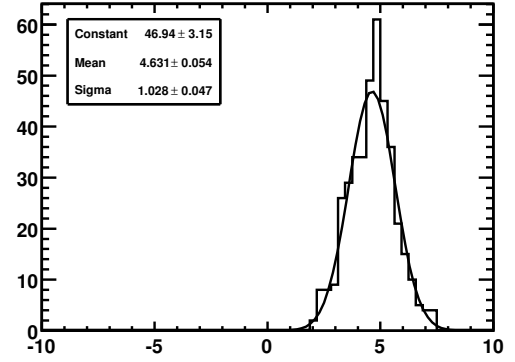


Figure 4.8: Pull plot on u_- for an absolute bias on the mistag of 2.0%.

Lifetime acceptance

The lifetime acceptance biases the selection against low-lifetime events. These events are where the cosine terms in the asymmetries dominate, but the sensitivity to γ comes from the sine terms in the asymmetries, which are ordinarily small in magnitude compared to the cosine terms. It follows that the acceptance will reduce the difference in magnitudes between the sine and cosine terms in the asymmetries, and this is seen to improve the precision with which the parameters u_+, u_- can be fitted for a given number of events. The errors in the presence of an acceptance function, relative to the errors when no acceptance function is present, are 10% smaller in the absence of mistag and background, and 30% smaller if mistag and background are included in the fit. We conclude that lifetime-biasing cuts should be used wherever possible when creating a sample of $B_d^0 \rightarrow D^- \pi^+$ events for the extraction of γ . If it is possible to achieve the same yield and purity using either lifetime-biasing or lifetime-unbiased cuts, for example with impact parameter rather than transverse momentum cuts, the selection using lifetime-biasing cuts will provide a B_d^0 sample with an intrinsically higher sensitivity to γ .

4.7.2 Systematic errors

As well as the above contributions to the statistical error, there may be systematic errors associated with detector effects, which will now be discussed.

Mistag

Examining Equation 4.17, which gives the asymmetry in the presence of background and mistag, it can be seen that an imperfect knowledge of the mistag and background levels will lead to systematic biases in the fitted values of u_+, u_- , and hence γ . Specifically, any change in the dilution term requires corresponding changes in the values of u_+, u_- in order to balance the overall asymmetry. This bias is relevant if the mistag level is to be fixed from external sources, which are expected to have a statistical error of 1% after one year of running. If it

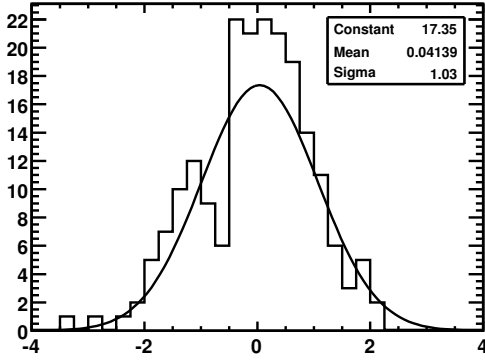


Figure 4.9: Pull plot on u_+ when fitting with the mistag as a free parameter.

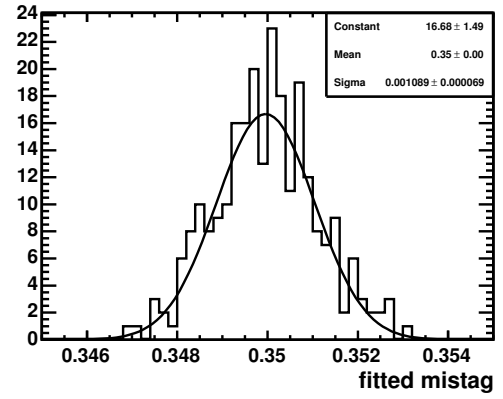


Figure 4.10: The value of the fitted mistag, for an input value of 0.35.

is assumed that there are no additional systematic errors in the measurement of the mistag itself, the expected systematic bias can be established by fitting to the asymmetries using a value of the mistag 1σ away from the value used to generate the toy Monte Carlo, while assuming that the background is perfectly understood. A systematic bias can be shown on a pull plot, where the pull of observable x with fitted value x_{fit} and fitted error of σ_x is given by:

$$\text{pull}_x = \frac{x - x_{\text{fit}}}{\sigma_x}. \quad (4.31)$$

For correctly fitted x , the pull plot should have a mean of 0 and a width of 1. A deviation from 0 in the mean indicates a systematic bias in the fitting, while a width different from 1 indicates that the fitting is not estimating the errors correctly. Figures 4.7 and 4.8 show the pull plots on the fitted parameter u_+ for a bias on the mistag of 1.0% and 2.0% respectively. The systematic offset is linear with the error on the mistag, and fitting with a mistag rate which is 1% (absolute) away from the true value translates into a bias of $2.2\sigma_{\text{stat}}$ on the parameters u_+ , u_- after one year of running, given in terms of σ_{stat} , the statistical error on the fitted parameters. Hence we see that even a 1% bias on the mistag rate can produce a systematic offset considerably larger than the statistical precision after only one year of running.

Fitting the mistag rate directly from the asymmetries eliminates such a bias. Figure 4.9 shows the pull plots on the fitted parameters u_+ , u_- in the case when the mistag is fitted, without any external input, from the asymmetries; the pull plots are clearly unbiased. Figure 4.10 shows the fitted value of the mistag. The one year precision on the value of the mistag of 0.1%, as given in Equation 4.22, underlines the potential of this channel to itself serve as a mistag control channel for other analyses. In the event that the background of this channel is poorly understood it is always possible to fit to the overall dilution instead, losing the mistag information while protecting against systematic errors on the extracted value of γ .

Time resolution and acceptance function

The systematic error caused by ignoring the time resolution when fitting to the asymmetries was shown in [Rad01] to be approximated by:

$$\sigma_{syst,\sigma_t} \approx \frac{\Delta m \Gamma \sigma_\tau^2}{2|\eta|} \approx 1.0\%, \quad (4.32)$$

as a fraction of the fitted parameter itself. Figure 4.11 shows the pull plot on the fitted parameters u_+, u_- , when fitting without taking into account the time resolution, after one year of data taking. Figure 4.12 shows the same quantity after five years of data taking. All detector effects were included in the fit, and the value of $|\eta|$ is assumed to be known exactly. The time resolution is introduced in the manner described earlier in the chapter, but no attempt is made to account for it when fitting, and the asymmetry fit is performed as if there were no finite time resolution present. The mean value of the one year fit is biased by $(-0.08 \pm 0.04) \sigma_{1y}$, while that of the five year fit is biased by $(-0.35 \pm 0.04) \sigma_{5y}$, where $\sigma_{1y,5y}$ is the appropriate statistical error. Since the statistical error decreases after five years of data taking, it is expected that the bias will increase as a fraction of the statistical error, and the results are consistent with a constant absolute bias due to the time resolution. These plots show that the systematic error resulting from ignoring the time resolution in the fits will not be negligible when compared to the statistical error on u_+, u_- .

It is worth noting that the worst case scenario is presented here, since it is in principle possible to account for the presence of time resolutions in the fit¹. It is empirically observed that this bias is reduced by approximately a factor of 2 in the absence of an acceptance function.

A bias of around one fifth of the statistical precision on u_+ will propagate into an equivalent bias on the value of γ . For those areas of parameter space where the error on γ is around 10° , this translates into a bias of ~ 2 -3 percent compared to the value of γ , in rough agreement with Equation 4.32.

Imperfect knowledge of $|\eta|$

The precision with which γ will be measured from $B_d^0 \rightarrow D^\mp \pi^\pm$ depends on how well the ratio of CKM suppressed to CKM favoured decay modes, $|\eta|$, is known. As shown in Section 4.5, any error on $|\eta|$ will saturate the error on γ , so that the statistical error does not improve with additional data except in a narrow region of phase space. Larger errors on $|\eta|$ will lead to a quicker saturation, and with an error of 30% on $|\eta|$ the error on γ will not improve between one year of data taking and five years of data taking. The extra error on γ induced by an imperfect knowledge of $|\eta|$ is highly dependent on the area of parameter space being considered. In particular, for the currently favoured value of $\gamma + 2\beta \sim 15^\circ$, it can be seen that this additional error is smallest if the strong phases in the decay channels $B_d^0 \rightarrow D^\mp \pi^\pm$ are zero.

¹See, for example, a similar fit in the channel $B_s^0 \rightarrow D_s^\mp K^\pm$ [CMR07].

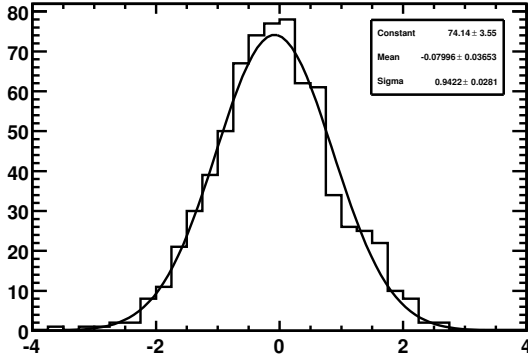


Figure 4.11: Pull plot on u_+ when fitting with a time resolution of 40 fs for 360 one year measurements.

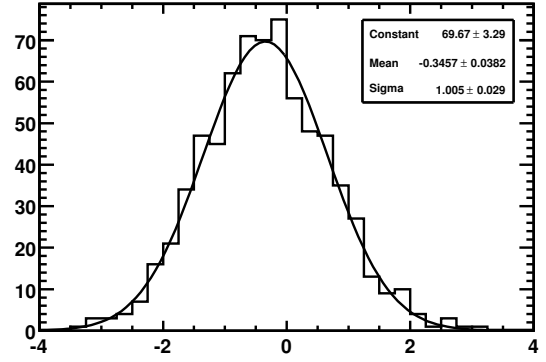


Figure 4.12: Pull plot on u_+ when fitting with a time resolution of 40 fs for 360 five year measurements.

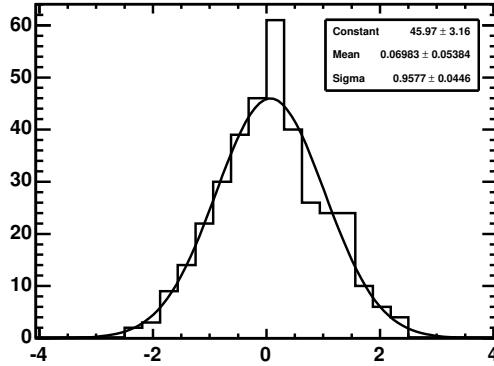


Figure 4.13: Pull plot on u_+ when fitting with an error on $|\eta|$ of 10% for 360 one year measurements.

A pull plot was produced for the fit parameters u_+, u_- , when fitting with an external constraint on the value of $|\eta|$ for one year of data taking, shown in Figure 4.13. The pull plot is unbiased – any bias in the fitted value of u_+, u_- would translate into a bias on γ according to Equation 4.20. A 10% error on $|\eta|$ was assumed.

Analyzing $B_d^0 \rightarrow D^- \pi^+$ and $B_d^0 \rightarrow D^{*-} \pi^+$

It was noted at the beginning of the chapter that this extraction of γ is substantially the same as that used in [Rad01] to extract γ from $B_d^0 \rightarrow D^{*-} \pi^+$. There are several reasons why it is desirable to make an extraction using both channels:

- Combined analysis : The channel $B_d^0 \rightarrow D^- \pi^+$ can be used to perform a combined U-spin extraction of γ , eliminating most of the ambiguous solutions for γ , as will be shown in the following chapter.
- Sensitivity dependence on the strong phase : As was shown in [Rad01], and will be reaffirmed in Section 4.5, the expected sensitivity to γ has a dependence on the value

of the strong phase, Δ_{qcd} . Since it is unlikely that both channels will have the same strong phase, extracting γ from both channels protects against a value of the strong phase which has a low sensitivity to γ associated with it.

- Knowledge of $|\eta|$: The value of $|\eta|$ may be different and/or better known in one channel compared with the other.
- Statistics : It is desirable to measure γ from as many channels as possible, in order to arrive at the optimal precision and search for signs of new physics.

4.8 Conclusion

The angle γ can be extracted with a precision which depends significantly on the values of the parameters Δ_{qcd} and $\phi_{\text{mix}} - \gamma$, as well as the assumed error on the knowledge of $|\eta|$. It can be as good as 8° for 2 fb^{-1} of data taking in certain sectors of parameter space. It has been shown that the value of the mistag in this channel can be fitted from the asymmetries independently of any external information about its value, which eliminates one significant source of systematic error in the fit. It has also been shown that the achievable precision on γ depends on the lifetime distribution of the selected B_d^0 mesons, with samples richer in longer lived B_d^0 particles affording a better sensitivity to γ . The precision achievable on the value of γ depends significantly on the knowledge of the parameter $|\eta|$, which must be established from external sources.

The extraction of γ presented in this chapter also suffers an eightfold ambiguity on the extracted value of γ . The next chapter will present an alternative method for extracting γ which aims to resolve this ambiguity, as well as requiring no a priori knowledge of the parameter $|\eta|$.

Chapter 5

Extraction of the angle γ through the U-spin analysis of the channels $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$

5.1 Motivation

The last chapter showed that the conventional extraction of γ from the channel $B_d^0 \rightarrow D^- \pi^+$ suffers from two weaknesses:

1. An eightfold ambiguity on the extracted value of γ ,
2. The requirement that $|\eta|$ is well constrained from external sources.

This chapter will discuss a method of measuring γ by using the U-spin symmetry between the channels $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$, proposed in [Fle03], which offers the potential for an unambiguous extraction of γ and does not require any knowledge of $|\eta|$. Although it is not part of this study, an equivalent extraction can be performed for the channels $B_d^0 \rightarrow D^{*-} \pi^+$ and $B_s^0 \rightarrow D_s^{*-} K^+$, or indeed $B_d^0 \rightarrow D^- \rho^+$ and $B_s^0 \rightarrow D_s^- K^*$, and the mathematical formalism presented in this chapter holds for those decay modes as well.

5.2 Overview of U-spin method

For the reader's benefit, the combined U-spin analysis of the channels $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$ is reviewed in this section. It is natural to adopt the notation in [Fle03] for these purposes; since, however, this notation differs from that used in the previous chapter, some attention will be paid to relating the two.

5.2.1 U-spin formalism

As explained in Section 1.5.2, U-spin is an SU(2) subgroup of the flavour SU(3) symmetry, under which the (d, s) pair of quarks is a doublet, similarly to (u, d) under isospin. As an example of this mechanism, recall that under isospin, the pions are considered as isospin eigenstates of the same hadron:

$$\begin{aligned}\pi^+ (u\bar{d}) &\equiv I_3 = +1; \\ \pi^0 (u\bar{u} - d\bar{d}) / \sqrt{2} &\equiv I_3 = 0; \\ \pi^- (\bar{u}d) &\equiv I_3 = -1.\end{aligned}\tag{5.1}$$

Although the I_3 component of isospin varies, $I = 1$ for all three pions. It is empirically observed that the strong interactions of the pions are almost identical. Since the three pions are related by the exchange of u and d quarks, isospin symmetry implies that strong interactions are unchanged under the exchange of u and d quarks. U-spin symmetry functions in the same way, but between d and s quarks. Hence the $B_s^0 (b\bar{s})$ and $B_d^0 (b\bar{d})$ mesons, which are related by the exchange of a d quark for an s quark, are identical under U-spin, and their strong interactions are postulated to be identical also.

5.2.2 Time dependent rate asymmetries

The time dependent asymmetries take the same form for B_s and B_d^0 decays, and can therefore be developed in a general way, for a decay from state B_q to $D_q\bar{u}_q$, as:

$$\begin{aligned}\frac{\Gamma(B_q^0 \rightarrow D_q\bar{u}_q)(t) - \Gamma(\bar{B}_q^0 \rightarrow D_q\bar{u}_q)(t)}{\Gamma(B_q^0 \rightarrow D_q\bar{u}_q)(t) + \Gamma(\bar{B}_q^0 \rightarrow D_q\bar{u}_q)(t)} = \\ \frac{C(B_q^0 \rightarrow D_q\bar{u}_q) \cos(\Delta M_q t) + S(B_q^0 \rightarrow D_q\bar{u}_q) \sin(\Delta M_q t)}{\cosh(\Delta\Gamma_q t/2) - A_{\Delta\Gamma}(B_q \rightarrow D_q\bar{u}_q) \sinh(\Delta\Gamma_q t/2)},\end{aligned}\tag{5.2}$$

where q stands for the d or s quark, and u_q is a pion or kaon. The new terms in the notation are C, S , and $A_{\Delta\Gamma}$, referred to as “CP-observables” from now on. $\Delta\Gamma_q$ represents the decay width difference between the “heavy” and “light” B_q mass eigenstates—it is negligibly small for the B_d^0 system, whose asymmetry therefore simplifies to:

$$\begin{aligned}\frac{\Gamma(B_d^0 \rightarrow D\pi)(t) - \Gamma(\bar{B}_d^0 \rightarrow D\pi)(t)}{\Gamma(B_d^0 \rightarrow D\pi)(t) + \Gamma(\bar{B}_d^0 \rightarrow D\pi)(t)} = \\ C(B_d^0 \rightarrow D\pi) \cos(\Delta M_d t) + S(B_d^0 \rightarrow D\pi) \sin(\Delta M_d t),\end{aligned}\tag{5.3}$$

which can be compared to Equation 4.3 in the previous chapter. The CP-observables can be written as:

$$C(B_q^0 \rightarrow D_q\bar{u}_q) \equiv C_q = \frac{1 - |\xi_q|^2}{1 + |\xi_q|^2},\tag{5.4}$$

$$S(B_q^0 \rightarrow D_q \bar{u}_q) \equiv S_q = \frac{2\text{Im}(\xi_q)}{1 + |\xi_q|^2}; \quad (5.5)$$

$$A_{\Delta\Gamma}(B_q^0 \rightarrow D_q \bar{u}_q) = \frac{2\text{Re}(\xi_q)}{1 + |\xi_q|^2}. \quad (5.6)$$

The term ξ is related to the η of the previous chapter, and is given by:

$$\xi = -(-1)^L e^{-i(\phi_q + \gamma)} \left[\frac{1}{x_q e^{i\delta_q}} \right]. \quad (5.7)$$

Here $\phi_q \equiv \phi_{\text{mix}}$ and $\delta_q \equiv -\Delta_{\text{qcd}}$, while $x_q \equiv |\eta|$. Furthermore, the terms x_q and δ_q are related to the matrix elements $M_q, \bar{M}_q$ of the decays $\bar{B}_q^0 \rightarrow D_q \bar{u}_q$ and $\bar{B}_q^0 \rightarrow \bar{D}_q u_q$ through:

$$a_q e^{i\delta_q} \propto \frac{M_q}{\bar{M}_q}, \quad (5.8)$$

where a_q is given by:

$$a_s \equiv \frac{x_s}{R_b}, \quad (5.9)$$

$$a_d \equiv -\left(\frac{1 - \lambda^2}{\lambda^2}\right) \frac{x_d}{R_b}, \quad (5.10)$$

and R_b is given by:

$$R_b \equiv \left(1 - \frac{\lambda^2}{2}\right) \frac{1}{\lambda} \left| \frac{V_{ub}}{V_{cb}} \right|. \quad (5.11)$$

The term $(-1)^L$, where L is the angular momentum of the $D_q \bar{u}_q$ system, determines the relative sign of the sine and cosine terms in the asymmetry, and can be used to eliminate certain ambiguities as will be discussed later. For the decays considered in this study, $L = 0$. There are analogous CP-observables for the other asymmetry, from a state B_q to $\bar{D}_q u_q$. These will be labeled $\bar{C}_q, \bar{S}_q$ and $\bar{A}_{\Delta\Gamma}$, but are governed by the same three equations above, with $\bar{\xi}_q$ replacing ξ_q :

$$\bar{\xi}_q = -(-1)^L e^{-i(\phi_q + \gamma)} [x_q e^{i\delta_q}]. \quad (5.12)$$

Finally, the CP-observables C and S can be written out explicitly in this notation:

$$C_q = -\left[\frac{1 - x_q^2}{1 + x_q^2} \right]; \quad (5.13)$$

$$\bar{C}_q = +\left[\frac{1 - x_q^2}{1 + x_q^2} \right]; \quad (5.14)$$

$$S_q = (-1)^L \left[\frac{2x_q \sin(\phi_q + \gamma + \delta_q)}{1 + x_q^2} \right]; \quad (5.15)$$

$$\bar{S}_q = (-1)^L \left[\frac{2x_q \sin(\phi_q + \gamma - \delta_q)}{1 + x_q^2} \right]. \quad (5.16)$$

It will prove convenient to define orthogonal combinations of the CP-observables S, C :

$$\begin{aligned}
\langle C_q \rangle_+ &\equiv \frac{\overline{C}_q + C_q}{2} = 0; \\
\langle C_q \rangle_- &\equiv \frac{\overline{C}_q - C_q}{2} = \frac{1 - x_q^2}{1 + x_q^2}; \\
\langle S_q \rangle_+ &\equiv \frac{\overline{S}_q + S_q}{2} = +(-1)^L \left[\frac{2x_q \cos \delta_q}{1 + x_q^2} \right] \sin(\phi_q + \gamma); \\
\langle S_q \rangle_- &\equiv \frac{\overline{S}_q - S_q}{2} = -(-1)^L \left[\frac{2x_q \sin \delta_q}{1 + x_q^2} \right] \cos(\phi_q + \gamma).
\end{aligned} \tag{5.17}$$

These quantities are related to the terms $\langle u_\pm \rangle$ used in the conventional extraction of γ as follows:

$$\langle u \rangle_- = (-1)^L \left[\frac{1 + x_q^2}{2x_q} \right] \langle S_q \rangle_+, \tag{5.18}$$

$$\langle u \rangle_+ = -(-1)^L \left[\frac{1 + x_q^2}{2x_q} \right] \langle S_q \rangle_-. \tag{5.19}$$

5.2.3 Reducing the ambiguity through factorization

One technique for eliminating ambiguous solutions is to use factorization arguments, the concept of which was introduced in Chapter 2. Here they can be used to fix the sign of $\cos \delta_q$ in Equation 5.17, thus enabling the sign of $\sin(\phi_q + \gamma)$ to be fixed also. Factorizing the four-quark matrix operators $M_q, \overline{M}_q$ in Equation 5.8 into the product of two quark currents allows the sign of their ratio to be determined, which in turn allows the sign of δ_q and therefore $\sin(\phi_q + \gamma)$ to be determined. The details of the calculation are given in [Fle03], and here we will only quote the result:

$$\delta_q|_{\text{fact}} = 0^\circ. \tag{5.20}$$

Hence the sign of $\sin(\phi_q + \gamma)$ is seen to be positive, reducing the number of ambiguous solutions by half. It should be noted that the factorization assumption may receive sizable corrections for $\overline{B}_q^0 \rightarrow \overline{D}_q u_q$ decays, where the spectator quark q ends up in the “light” u_q meson. Nonetheless, it is expected that these corrections will not change the sign of $\cos \delta_q$.

Exploiting $A_{\Delta\Gamma}$

There is another observable yet to be exploited in Equation 5.3, $A_{\Delta\Gamma}$:

$$A_{\Delta\Gamma} = \frac{2\text{Re}\xi_q}{1 + |\xi_q|^2}. \tag{5.21}$$

Within the asymmetry, $A_{\Delta\Gamma}$ is multiplied by the term $\sinh(\Delta\Gamma_q t/2)$, where $\Delta\Gamma_q$ is the difference in widths of the B_q mass eigenstate. This width is negligible in the B_d^0 sector,

however, it may play a substantial role in the B_s sector, where $\Delta\Gamma_q/\Gamma_q$ has been estimated at the Tevatron [TDCA07][TCCA05] as $|\Delta\Gamma_q/\Gamma_q| \sim 10\%$. It is therefore worth briefly reviewing its importance with respect to the decay $B_s^0 \rightarrow D_s^- K^+$, with reference to the latest LHCb Monte Carlo studies in this channel [CMR07].

What makes $A_{\Delta\Gamma}$ such a useful observable is that it can be extracted from the untagged rate $\Gamma(B_q^0(t) \rightarrow D_q \bar{u}_q)$, i.e. the combined rate for B_q^0 and $\bar{B}_q^0$ events to decay into the same final state:

$$\Gamma(B_q^0 \rightarrow D_q \bar{u}_q)(t) = \left[\Gamma(B_q^0 \rightarrow D_q \bar{u}_q) + \Gamma(\bar{B}_q^0 \rightarrow D_q \bar{u}_q) \right] \times [\cosh(\Delta\Gamma_q t/2) - A_{\Delta\Gamma} \sinh(\Delta\Gamma_q t/2)] e^{-\Gamma_q t}. \quad (5.22)$$

Here, Γ_q refers to the average decay width, $\Gamma_q = (\Gamma_H^q + \Gamma_L^q)/2$. This substantially increases the number of events available to the analysis, since the tagging power in B_s channels is only expected to be around 8%, and provides valuable information on γ through the constraint:

$$[C]^2 + [S]^2 + [A_{\Delta\Gamma}]^2 = 1, \quad (5.23)$$

It has been shown [CMR07] that including these untagged events in the study improves the precision with which $\gamma + \phi_s$ can be measured in the channel $B_s^0 \rightarrow D_s^- K^+$ from $\sim 12^\circ$ to $\sim 10^\circ$ in 1 year of LHCb running in a conventional analysis. In the context of the combined U-spin analysis, using untagged events results in an improved precision on the parameters $S_s, \bar{S}_s$, and helps reduce the eightfold ambiguity to a twofold one.

Combined extraction of γ

The combined extraction of γ relies on the U-spin symmetry between the channels $B_d \rightarrow D_d \bar{u}_d$ and $B_s \rightarrow D_s \bar{u}_s$, which are obtained by replacing all down quarks in the decay by strange quarks. From Equation 5.17 it follows that:

$$\left[\frac{a_s \cos \delta_s}{a_d \cos \delta_d} \right] R = -(-1)^{L_s - L_d} \left[\frac{\sin(\phi_d + \gamma)}{\sin(\phi_s + \gamma)} \right] \left[\frac{\langle S_s \rangle_+}{\langle S_d \rangle_+} \right], \quad (5.24)$$

and similarly

$$\left[\frac{a_s \sin \delta_s}{a_d \sin \delta_d} \right] R = -(-1)^{L_s - L_d} \left[\frac{\cos(\phi_d + \gamma)}{\cos(\phi_s + \gamma)} \right] \left[\frac{\langle S_s \rangle_-}{\langle S_d \rangle_-} \right], \quad (5.25)$$

where

$$R \equiv \left(\frac{1 - \lambda^2}{\lambda^2} \right) \left[\frac{1 + x_d^2}{1 + x_s^2} \right]. \quad (5.26)$$

Here, λ is the sine of the Cabbibo angle. For the decays considered here, $L_s = L_d = 0$, and exact U-spin symmetry implies that:

$$|a_s| = |a_d|, \quad \delta_s = \delta_d. \quad (5.27)$$

Substituting these into Equations 5.24 and 5.25, it is possible to extract γ unambiguously in a number of ways, depending on what additional assumptions are made, as described in [Fle03]. Three scenarios are considered in turn: a strong U-spin assumption, a phase U-spin assumption, and an amplitude U-spin assumption.

Strong U-spin assumption

In this scenario, it is assumed that $\delta_s = \delta_d$ and $a_s = a_d$. Equations 5.24 and 5.25 are used to constrain γ , and their theoretical uncertainties are determined by U-spin breaking corrections to

$$a_s \cos(\delta_s) = a_d \cos(\delta_d),$$

and

$$a_s \sin(\delta_s) = a_d \sin(\delta_d)$$

respectively.

Phase U-spin assumption

In this scenario it is assumed that $\delta_s = \delta_d$, but no assumption is made about the relative values of a_s and a_d . Equations 5.24 and 5.25 imply the following relation:

$$\frac{\tan \phi_d + \gamma}{\tan \phi_s + \gamma} = \left[\frac{\tan \delta_d}{\tan \delta_s} \right] \left[\frac{\langle S_s \rangle_-}{\langle S_s \rangle_+} \right] \left[\frac{\langle S_d \rangle_+}{\langle S_d \rangle_-} \right], \quad (5.28)$$

which can then be used to extract γ unambiguously from the four observables $\langle S_{s,d} \rangle_{+,-}$. The theoretical uncertainty in this extraction results from U-spin breaking corrections to $\delta_s = \delta_d$ only.

Amplitude U-spin assumption

In this scenario, it is assumed that $a_s = a_d$, but no assumption is made about the values of δ_s or δ_d . The exact relation:

$$\left(\frac{a_s}{a_d} \right) R = \sigma \left| \frac{\sin(2\phi_d + 2\gamma)}{\sin(2\phi_s + 2\gamma)} \right| \sqrt{\frac{\langle S_s \rangle_+^2 \cos(\phi_s + \gamma)^2 + \langle S_s \rangle_-^2 \sin(\phi_s + \gamma)^2}{\langle S_d \rangle_+^2 \cos(\phi_d + \gamma)^2 + \langle S_d \rangle_-^2 \sin(\phi_d + \gamma)^2}}, \quad (5.29)$$

which follows from Equation 5.17 and the definitions of a_q and R given before, is used to extract γ . The parameter σ governs the sign of the right hand side, so that:

$$\sigma = -\text{sgn} [\langle S_s \rangle_+ \langle S_d \rangle_+ \sin(\phi_d + \gamma) \sin(\phi_s + \gamma)], \quad (5.30)$$

if $\cos \delta_s$ and $\cos \delta_d$ are assumed to have the same sign, while:

$$\sigma = -\text{sgn} [\langle S_s \rangle_- \langle S_d \rangle_- \cos(\phi_d + \gamma) \cos(\phi_s + \gamma)], \quad (5.31)$$

if $\sin \delta_s$ and $\sin \delta_d$ are assumed to have the same sign.

5.2.6 Summary of combined extraction

The combined U-spin analysis of the channels $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$ allows for an unambiguous extraction of γ under a variety of theoretical assumptions. It also has a great advantage over the “conventional” extraction of not requiring $\langle x_d \rangle$ to be resolved, since $\langle x_d \rangle$ only enters the expressions for γ as a second order correction through the term $1 + \langle x_q \rangle^2$, which is negligible.

Even if this combined analysis is outperformed by other analyses at measuring γ , it can instead be used to learn about the scale of U-spin symmetry breaking. For example, under perfect U-spin symmetry, the untagged rate analysis and the U-spin analysis can be expected to return the same value of γ , since the experimental systematics will be broadly similar in the B_s and B_d systems¹. Any observable difference between the fitted values of γ can therefore be used to constrain the U-spin breaking parameters. Finally, if this analysis allows a limit to be set on a_s/a_d , it may be possible to use the expression:

$$\frac{a_s}{a_d} = \left(\frac{\lambda^2}{1 - \lambda^2} \right) \left| \frac{x_s}{x_d} \right|, \quad (5.32)$$

to make an indirect measurement of x_d , since x_s will be large enough to measure from the asymmetries in $B_s^0 \rightarrow D_s^- K^+$ alone.

5.3 Fitting to the CP-observables

In the present study, fitting to the CP-observables is done using the same toy Monte Carlo as in the previous chapter. The fitted parameters are $S_d, \overline{S}_d$, as defined in Equation 5.15 and 5.16, while the parameters $C_d, \overline{C}_d$ are fixed to be ± 1 in the fit, which is an excellent approximation. Because C_d is fixed, no knowledge of the parameter x_d is required, so that the fitted precisions on $S_d, \overline{S}_d$ only depend on the number of events and any systematic effects from the time resolution and imperfect knowledge of the mistag and background. As before, the mistag rate is left as a free parameter in the fit, thus eliminating that source of systematic error. For the channel $B_s^0 \rightarrow D_s^- K^+$, the precisions were determined in [CMR07], and Equation 5.23 has been used to provide an additional constraint on the parameters $S_s, \overline{S}_s$, improving the precision with which they are extracted. Table 5.1 shows the precisions on the CP-observables, mistag, and x_s , which do not depend on the values of γ , ϕ_d or δ_d .

The CP-observables $S, \overline{S}$ are typically of order $S_d \sim O(0.01)$ and $S_s \sim O(0.1)$. $S_d, \overline{S}_d$ have been measured by B-factories in the B_d^0 sector, with a current world precision of ± 0.04 [ea06c]. This precision will improve by the time LHCb starts taking data, and B-factory data will be combined with data from LHCb to yield the best possible constraint on γ . At the time of writing, there has been no observation of CP violation in the decays $B_s^0 \rightarrow D_s^- K^+$.

¹In particular, the mistag rate can be fitted from the asymmetries in both $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$, the proper time resolutions are expected to be similar, as are the lifetime acceptance functions.

Table 5.1: A summary of the statistical precisions on the observables $S_q, \overline{S}_q$, the mistag rate ω_q , and x_s in the channels $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$ for one and five years of running, corresponding to 2 fb^{-1} and 10 fb^{-1} of data respectively.

	ω_d	S_d	$\overline{S}_d$	ω_s	S_s	$\overline{S}_s$	x_s
σ_{1y}	0.1%	0.0076	0.0076	0.3%	0.11	0.10	0.066
σ_{5y}	0.05%	0.0034	0.0034	0.13%	0.049	0.045	0.029

Table 5.2: Input parameter values to the combined analysis toy Monte Carlo.

Parameter	x_s	x_d	λ	ϕ_d	ϕ_s	R_b
Value	0.37	0.024	0.22	47°	0°	17.2

Sensitivity to γ

Since the extraction of γ can proceed in any one of the three ways reviewed in Section 5.2.3, it will be necessary to perform a case study for each of these extraction scenarios. In particular, it is useful to concentrate on the sensitivity of these extraction scenarios to U-spin breaking effects, as well as their sensitivity to the input values of the parameters γ and $\delta_{s,d}$.

The inputs to the combined analysis are given in Table 5.2. Although the CP-observables will have associated systematic errors as described in Chapter 4, the dominant systematic errors within the combined analysis are expected to come from U-spin breaking effects. Therefore, only the statistical errors on the CP-observables will be considered in this study. Also, it is assumed that ϕ_s , ϕ_d , and R_b are known exactly. In the case of R_b , this assumption is justified because the precisions on x_s from the channel $B_s^0 \rightarrow D_s^- K^+$ imply a relative error of $\sim 4\%$ on R_b after one year of data taking. This error is both negligible compared to the errors on the CP-observables, especially in the B_s sector, and it is also likely to be negligible compared to U-spin breaking effects. Effects such as time resolution, mistag, background, and the lifetime acceptance have been taken into account when determining the statistical errors on $S_q, \overline{S}_q$. Finally, since the channels in question are expected to be insensitive to new physics, the usual assumption that γ lies between 0° and 180° degrees will be used throughout.

In principle, it is possible to use all four Equations 5.24, 5.25, 5.28, and 5.29 in order to extract γ unambiguously. These four extractions are complementary, not only because they allow for the extraction of γ under different assumptions, but also because they afford varying precisions and discrimination between ambiguities depending on the values of γ and $\delta_{s,d}$. The contour plots shown in this section are to be read as follows:

- The vertical axis describes the level of U-spin breaking, with 1 corresponding to perfect U-spin symmetry for the quantities shown.
- The thick curved line is the value of γ which would be measured given the appropriate level of U-spin symmetry breaking, assuming that the observables $S_q, \overline{S}_q$ were perfectly known.
- The thick horizontal line represents the scenario for which U-spin symmetry is perfectly observed. It crosses with the thick curved line at the input value of γ , in this case 60° .
- The dashed straight lines enclose the region of $\pm 30\%$ U-spin breaking. Although there is no precise estimate of the likely level of U-spin breaking, 30% is a conservative assumption, since U-spin is expected to be a better symmetry than SU(3) in these decays. The expected level of SU(3) breaking was discussed in Section 1.4.2, and given as 10%.
- The dashed curved lines are the 1σ contours, representing the uncertainty on the observables $S_q, \overline{S}_q$.
- The triangle marks the location of the input value of γ .
- The upside down triangles mark the ambiguous solutions coming from the conventional $B_d^0 \rightarrow D^+\pi^-$ analysis.

Note that not all of the plots have the same horizontal (γ) scale, because the four different Equations will in principle constrain different areas of parameter space. Each plot was zoomed to best show the area of parameter space constrained and the width of the associated 1σ intervals.

Two particular scenarios, corresponding to large and small strong phase differences, will be considered in detail, followed by a more general overview of the performance. Additional discussion of other interesting areas of parameter space can be found in Appendix C.

5.4.1 Error estimation

In the examples given, the errors are estimated from the intersection of the 1σ contours with horizontal lines indicating a $\pm 30\%$ level of U-spin symmetry breaking. In particular, consider Figure 5.2, and the contours for Equation 5.25. The 1σ contours and $\pm 30\%$ U-spin breaking lines define a trapezoid, whose upper right corner corresponds to the maximum positive error (systematic+statistical), while its lower left corner corresponds to the maximum negative error for the 1σ contours, assuming that U-spin breaking is limited to $\pm 30\%$. The negative error is taken from that contour for which it is smallest (in this case Equation 5.25), while the positive error is independently taken from that contour in which it is smallest. In this case that is also Equation 5.25, but in principle that need not be the case.

Such an error estimation procedure is not optimal, because it does not use all the available information. For one thing, the U-spin breaking is assumed to contribute a flat error, when

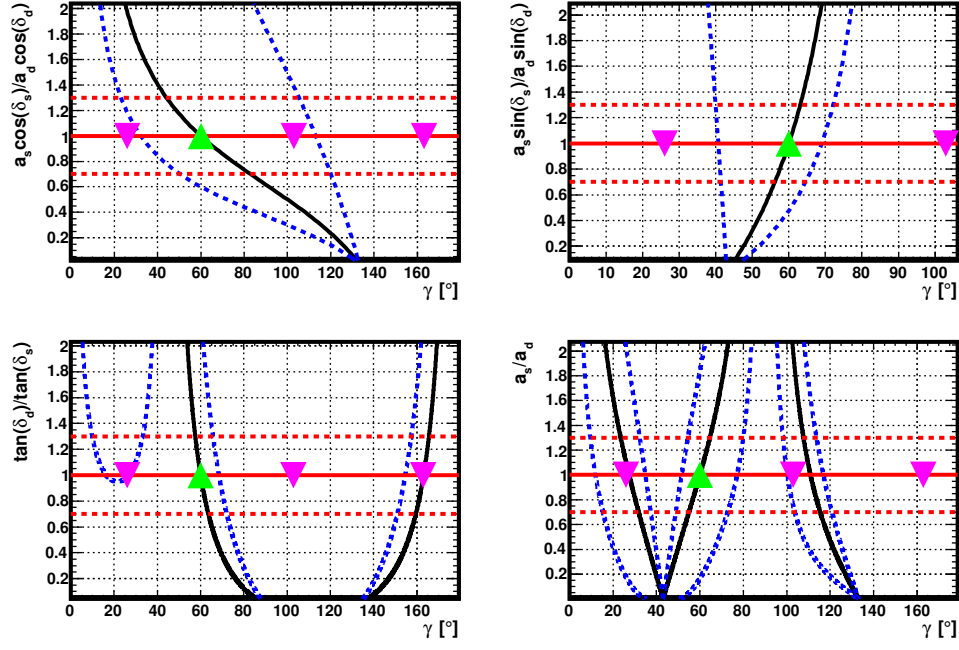


Figure 5.1: Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 60^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.

in reality it should be estimated by a gaussian of appropriate width. In addition, each of the four sets of contours is considered independently, when in reality they should be considered together. Instead of drawing 1σ bands for each U-spin assumption, the correct procedure would be to calculate the PDF in γ –(U-spin breaking) space using all four assumptions and accounting for the correlations between them. Tak as an example the contours associated with Equations 5.24 and 5.25: these are produced from orthogonal combinations of independent (ignoring correlations) variables, and are themselves independent. Therefore they could be multiplied together to produce a joint PDF in γ –(U-spin breaking) space. The 1σ contours in γ would then be calculated by projecting the resulting PDF onto the γ axis. The current procedure is akin to projecting the PDF associated with each assumption before combining them, which is sub-optimal.

$\gamma = 60^\circ$ and $\delta_{s,d} = 60^\circ$

The contour plots for this scenario are shown in Figure 5.1 for one year of data taking, and in Figure 5.2 for five years of data taking. After one year of data taking, the strong U-spin assumption of Equation 5.25 already disfavors all ambiguous solutions, even allowing for a 30% U-spin breaking correction. The statistical precision is $-15^\circ, +7^\circ$, while a 30% U-spin breaking correction introduces an additional $\pm 3^\circ$ systematic error. The phase U-spin assumption provides an upper statistical error on γ of 8° . Although the ambiguous solution at $\sim 160^\circ$ is not disfavoured by this assumption, it is disfavoured for most scenarios by the other three relations. The amplitude U-spin assumption provides a statistical precision of

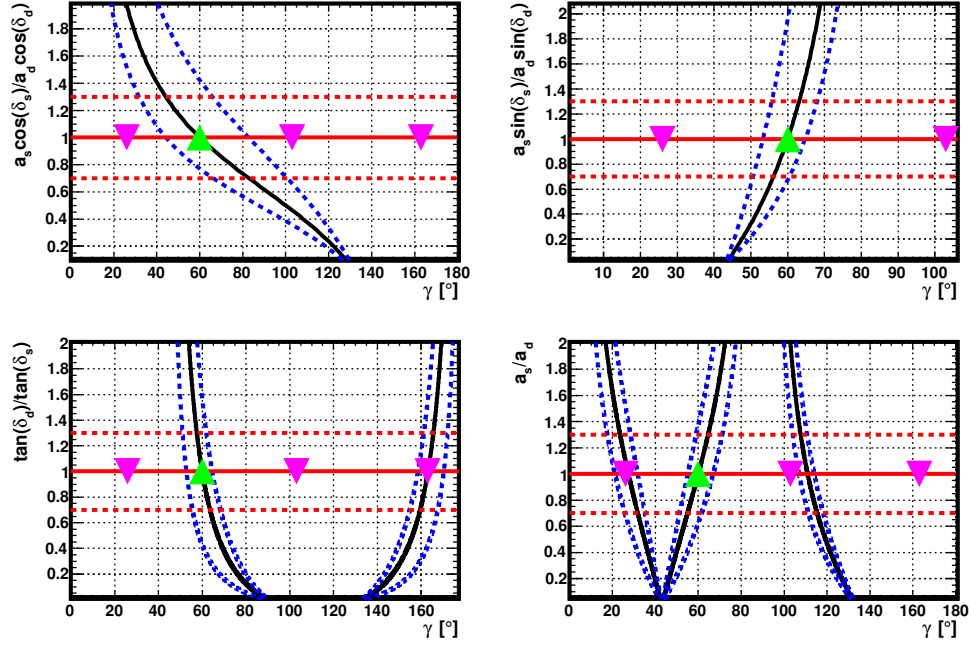


Figure 5.2: Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 60^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.

$-8^\circ, +14^\circ$, and a 30% U-spin breaking correction introduces an additional $\pm 4^\circ$ systematic error. Once again, ambiguous solutions are disfavoured when all four relations are considered together.

After five years, the strong U-spin assumption of Equation 5.25 gives a statistical precision of $-7^\circ, +5^\circ$, while a 30% U-spin breaking correction introduces an additional $\pm 2^\circ$ systematic error. All ambiguous solutions are now heavily disfavoured for any U-spin breaking correction for which $a_s \sin(\delta_s)$ has the same sign as $a_d \sin(\delta_d)$. The phase U-spin assumption now also gives a statistical precision of $-7^\circ, +5^\circ$, while a 30% U-spin breaking correction introduces an additional $-2^\circ, +3^\circ$ systematic error. The amplitude U-spin assumption affords a statistical precision of $-5^\circ, +7^\circ$, while a 30% U-spin breaking correction introduces a systematic of $\pm 3^\circ$. When all four relations are considered together, it is clear that all ambiguous solutions are excluded after five years of data taking.

$\gamma = 60^\circ$ and $\delta_{s,d} = 10^\circ$

Now consider the same value of γ , but much smaller strong phases. The contour plots are shown in Figure 5.3 for one year of data taking, and in Figure 5.4 for five years of data taking. After one year, none of the four relations provide effective discrimination against ambiguities. The best resolution comes from using the strong U-spin assumption of Equation 5.24, which reduces the ambiguity on γ to a twofold one, and gives a statistical precision of $-17^\circ, +25^\circ$. After five years, the strong U-spin assumption of Equation 5.24 gives a statistical

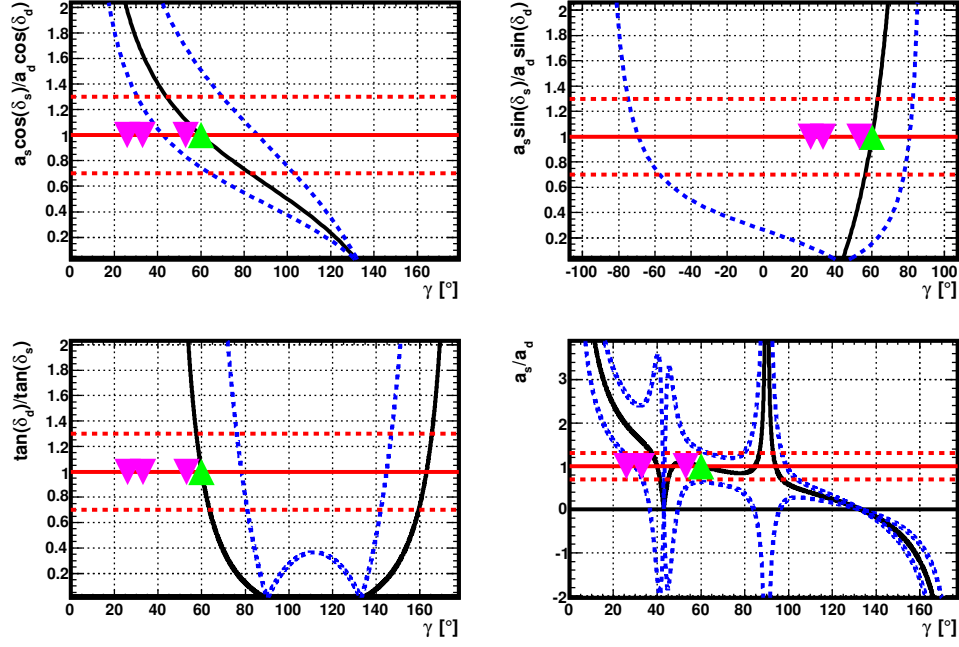


Figure 5.3: Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 10^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.

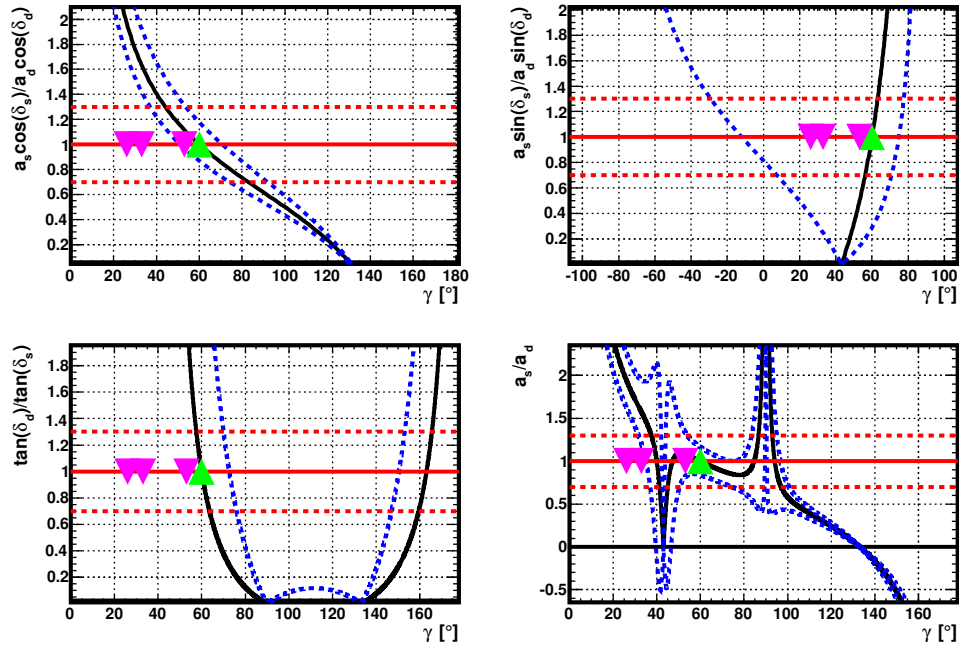


Figure 5.4: Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 10^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin). Note that the four plots have different scales on the horizontal axis.

precision of $-8^\circ, +10^\circ$, while the phase U-spin assumption gives an upper statistical precision of $+12^\circ$, however it is much less sensitive to U-spin breaking effects, with an upper systematic error of $+4^\circ$ associated with a 30% level of U-spin breaking. Equation 5.24 is now beginning to disfavour the solution at $\sim 53^\circ$ under exact U-spin symmetry. In general, Equation 5.24 is the only relation which provides any discrimination when the strong phases are small, and this area of parameter space is where the combined extraction performs least well.

5.4.4 Performance summary

The specific values of γ and the strong phases affect the combined extraction in two important ways. Firstly, certain combinations of $\gamma, \delta_s, \delta_d$ can lead to ambiguous γ solutions separated by only a few degrees, which will be impossible to distinguish even after many years of running. Secondly, the width of the error contours in U-spin/ γ space varies with the values of $\gamma, \delta_s, \delta_d$. Consider Equation 5.25, modified to take into account the errors on $\langle S_{s,d} \rangle_-$:

$$\left[\frac{a_s \sin \delta_s}{a_d \sin \delta_d} \right] R = -(-1)^{L_s - L_d} \left[\frac{\cos(\phi_d + \gamma)}{\cos(\phi_s + \gamma)} \right] \left[\frac{\langle S_s \rangle_- \pm \sigma_{\langle S_s \rangle_-}}{\langle S_d \rangle_- \pm \sigma_{\langle S_d \rangle_-}} \right]. \quad (5.33)$$

If the error on $\langle S_{s,d} \rangle_-$, $\sigma_{\langle S_{s,d} \rangle_-}$, is greater than the value of $\langle S_{s,d} \rangle_-$ itself, the sign of this observable will be badly defined, and the corresponding error contour will not provide a useful constraint on γ . An example of this effect was seen in the scenario with small strong phases above, where it is precisely this effect that prevents Equation 5.25 from usefully constraining γ , even after five years of running.

This uncertainty on the sign of $\langle S_{s,d} \rangle_\pm$ also causes the achievable precision on γ to be limited by the errors on $\langle S_d \rangle_\pm$, not $\langle S_s \rangle_\pm$. Although the errors on $\langle S_d \rangle_\pm$ are much smaller than those on $\langle S_s \rangle_\pm$, the fact that $x_s \gg x_d$ means that $|\langle S_s \rangle_\pm| \gg |\langle S_d \rangle_\pm|$, and the parameters $\langle S_d \rangle_\pm$ are therefore more susceptible to being measured with an incorrect sign. For example, when $\gamma = 60^\circ$ and $\delta_{s,d} = 60^\circ$, the observables are:

$$\begin{aligned} \langle S_d \rangle_- &= -0.012, \\ \langle S_d \rangle_+ &= 0.023, \\ \langle S_s \rangle_- &= 0.28, \\ \langle S_s \rangle_+ &= -0.28, \end{aligned}$$

while the errors are given by

$$\begin{aligned} \sigma_{\langle S_s \rangle_\pm} &= 0.11, \\ \sigma_{\langle S_d \rangle_\pm} &= 0.0076. \end{aligned}$$

Notice the doubling the error on $\langle S_s \rangle_\pm$ will not change the sign of $\langle S_s \rangle_\pm$, but doubling the error on $\langle S_d \rangle_\pm$ will change the sign of $\langle S_d \rangle_\pm$, rendering Equation 5.25 once again useless in constraining γ . It is important not to allow the small errors on $\langle S_d \rangle_\pm$ to obscure the fact that it is these parameters that need to be known most precisely for the combined analysis to fulfill its potential.

Table 5.3: One and five year precisions on γ for the scenarios considered in this thesis, giving the statistical error and the systematic error associated with a 30% U-spin breaking.

	$\sigma_{1y}^{\text{stat}}$	$\sigma_{1y}^{\text{syst}}$	$\sigma_{5y}^{\text{stat}}$	$\sigma_{5y}^{\text{syst}}$
$\gamma = 60^\circ, \delta_{s,d} = 60^\circ$	$-8^\circ, +8^\circ$	$-4^\circ, +3^\circ$	$-5^\circ, +5^\circ$	$\pm 3^\circ$
$\gamma = 60^\circ, \delta_{s,d} = 10^\circ$	$-17^\circ, +25^\circ$	$-10^\circ, +22^\circ$	$-8^\circ, +12^\circ$	$-15^\circ, +4^\circ$
$\gamma = 60^\circ, \delta_{s,d} = 30^\circ$	$-20^\circ, +11^\circ$	$-10^\circ, +3^\circ$	$-13^\circ, +6^\circ$	$-1^\circ, +4^\circ$
$\gamma = 60^\circ, \delta_{s,d} = 85^\circ$	$-7^\circ, +8^\circ$	$\pm 3^\circ$	$\pm 4^\circ$	$\pm 3^\circ$
$\gamma = 60^\circ, \delta_{s,d} = 140^\circ$	$-12^\circ, +11^\circ$	$-2^\circ, +3^\circ$	$-9^\circ, +5^\circ$	$-2^\circ, +4^\circ$
$\gamma = 40^\circ, \delta_{s,d} = 40^\circ$	$-15^\circ, +14^\circ$	$-7^\circ, +4^\circ$	$-10^\circ, +8^\circ$	$-7^\circ, +2^\circ$
$\gamma = 80^\circ, \delta_{s,d} = 40^\circ$	$-20^\circ, +6^\circ$	$-20^\circ, +1^\circ$	$-14^\circ, +3^\circ$	$-2^\circ, +1^\circ$

Table 5.3 presents the errors for all the individual scenarios considered in this chapter and Appendix C. The upper error quoted in any one γ, δ scenario is the smallest upper error in any one of the four U-spin assumptions, and similarly for the lower error. For example, it is possible for the quoted upper error in a scenario to come from the strong U-spin assumption, while the lower error comes from the phase U-spin assumption. The statistical and systematic error are always quoted from the same U-spin assumption.

It turns out that $\gamma = 60^\circ$ is particularly favourable for this extraction, because none of the four observables $\langle S_{s,d} \rangle_\pm$ is then particularly small. Looking back at Equation 5.17, the $\sin(\phi_q + \gamma), \cos(\phi_q + \gamma)$ terms control the size of $\langle S_{s,d} \rangle_\pm$. For $\gamma = 60^\circ$, $\phi_d + \gamma = 107^\circ$ and $\phi_s + \gamma = 60^\circ$, so that all four sine and cosine terms are substantial, but for values of γ substantially higher or lower than 60° these terms may approach 0, limiting precision regardless of the values of the strong phases. Finally, the performance for negative strong phases will be exactly the same as for positive strong phases of the same magnitude, as long as δ_s and δ_d have the same sign.

5.5 Conclusion

A combined analysis of the channels $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$ has been presented, which aims to extract γ in an unambiguous manner from tree level diagrams, thus providing a Standard Model baseline for new physics sensitive measurements of γ . The method's viability has been demonstrated through a toy Monte Carlo incorporating significant amounts of realism, including background, mistag, proper time resolution, and the proper time acceptance. The precision on γ , and the viability of excluding ambiguous solutions, is found to have a significant dependence on the precise input values of γ and the strong phases $\delta_{d,s}$. With the preferred Standard Model value $\gamma = 60^\circ$ and large strong phases, it is possible to achieve a statistical precision of $< 5^\circ$ after five years of data taking, while excluding all am-

biguous solutions. Other scenarios, such as small strong phases, or low values of γ , make an unambiguous extraction more difficult, and reduce the statistical precision achievable after five years of data taking to $\sim 10^\circ$, with a systematic error of $\sim 15^\circ$.

Chapter 6

A Monte Carlo independent measurement of the B_d^0 lifetime

6.1 Introduction

This chapter describes a Monte Carlo independent method for measuring the lifetime acceptance function of the LHCb lifetime biased trigger, which will allow the measurement of B meson lifetimes in decay channels biased by this trigger. The reconstruction of “hadronic” decay channels, such as $B_d^0 \rightarrow D^- \pi^+$ or $B_s^0 \rightarrow D_s^- \pi^+$, is a recognized strength of LHCb. However, it is precisely these channels which suffer a lifetime bias because of the trigger. Accounting for this bias allows the statistical power of LHCb to be fully exploited for lifetime measurements. The method presented here, and the “UHTFitter” software used, were originally developed at CDF. The basic idea of the acceptance correction is described in [Rad07]. The details of the method, especially the treatment of background, are described in [ABH⁺03], which has been developed in the context of CDF. For the reader’s convenience, the method will be described once again in detail within this chapter, with particular attention paid to LHCb-specific matters.

The remainder of this section will motivate the study. Section 6.2 describes those parts of the LHCb trigger which are relevant to this study. Sections 6.3, 6.4, and 6.5 describe the method for measuring the acceptance function. Section 6.7 shows fits to signal data. Section 6.8 shows fits to signal and background data. Section 6.9 discusses the expected statistical precision on lifetimes and possible biases arising from the background.

Motivation for measuring B lifetimes at LHCb

As discussed in Chapter 2, LHCb will considerably enhance the statistics available for the study of heavy flavour physics, in particular in the B_s^0 and b -baryon sectors. The measurement of B lifetimes will be important both for early detector calibration studies, and as a test of theoretical models such as Heavy Quark Expansion theory [Shi00, Ura98, Pet04, GOP04]. A particularly interesting and important measurement is that of the lifetime difference [LN07]

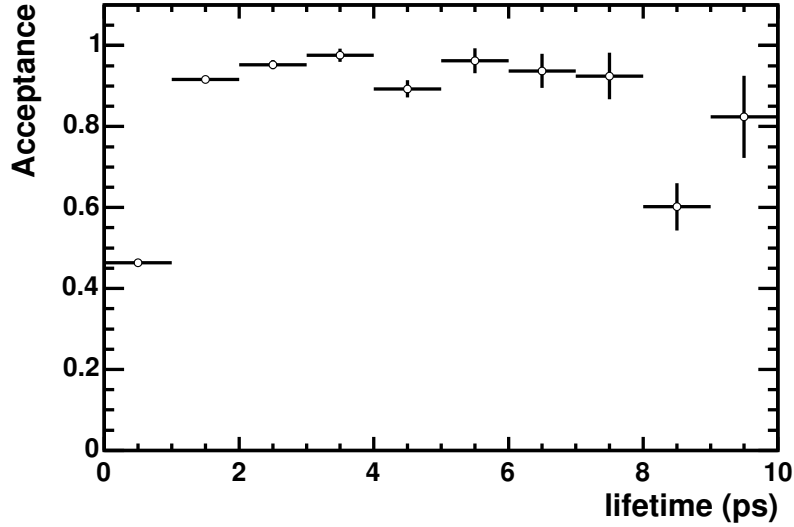


Figure 6.1: Proper lifetimes of reconstructed signal $B_d^0 \rightarrow D^-\pi^+$ events passing the L1 trigger, divided by the exponential distribution of B_d^0 lifetimes. This is the trigger lifetime acceptance function.

in the B_s^0 system. LHCb is expected to quickly improve on the current world lifetime average.

High yield hadronic channels can make significant contributions to the knowledge of lifetimes, but suffer from the lifetime acceptance bias introduced by the hadronic trigger. B particles are long lived compared to uds background or background from charm decays, and their decay products will therefore have high impact parameters (IP) with respect to the primary vertex. The LHCb trigger, described more fully in Section 6.2, cuts on the IP information, and in doing so it introduces an acceptance bias in the lifetime distributions. The lifetime acceptance function for the LHCb L1 trigger, as measured in the full LHCb Monte Carlo, is shown in Figure 6.1. The requirement that final state particles have a high IP with respect to the primary vertex is responsible for the lower efficiency below ~ 1 ps, while upper IP cuts, used in the trigger to limit background from secondary interactions and K_S^0 decays, reduce the efficiency for lifetimes above ~ 4 ps.

A possible method for taking these acceptance effects into account is to measure an acceptance function in Monte Carlo events. This is susceptible to systematic effects arising from possible differences between the simulation and real data. An alternative way that has been proposed is to apply a lower lifetime cut and assume that beyond that lifetime cut, the acceptance is flat. This assumption is not valid for the LHCb trigger at the time of writing this thesis because the upper impact parameter cuts introduce a downward slope in the acceptance. However, even if this upper impact parameter cut were removed, and the acceptance were in fact flat beyond a certain point, this approach would lose a significant number of events below the cut. Here a Monte Carlo independent method for correcting the trigger bias is proposed, which does not require a lifetime cut-off, or a flat acceptance curve beyond a certain point, and which can cope with both lower and upper impact parameter or lifetime cuts.

At this point, the reader may well be wondering why we are discussing only the trigger,

when the offline $B_d^0 \rightarrow D^-\pi^+$ selection shown in Chapter 3 also makes use of lifetime-dependent information, and therefore has a lifetime acceptance function of its own. The reason is that the lifetime acceptance of the offline selection can be trivially measured, since the events which fail the offline selection remain available for study, whereas events which fail the trigger are irrevocably lost. It would of course also be possible to use the method presented in this Chapter to measure the lifetime acceptance of the offline selection, and comparing the fitted offline selection acceptance with its “known” shape could potentially help in the study of systematic errors.

The channel $B_d^0 \rightarrow D^-(\pi^-\pi^-K^+)\pi^+$

This channel was selected to test the fitting method because its high yield and purity make it a prime candidate among hadronic channels for the actual measurement of the B_d^0 meson lifetime. It also has the same topology as the channel $B_s^0 \rightarrow D_s^-(\pi^-K^+K^+)\pi^+$, which is the prime candidate among hadronic B_s^0 channels for a measurement of the B_s^0 lifetime. The reconstruction of the channel $B_d^0 \rightarrow D^-(\pi^-\pi^-K^+)\pi^+$ has been described in Chapter 3, and the data set used in this study is the same one used there. The decay mode will have an annual yield of 1340k fully triggered events, with a signal to background ratio of ~ 5 .

All fits were performed using ROOT [BR97] v4.02/00 and the fitting package MINUIT [JR75].

6.2 The LHCb lifetime dependent trigger

The current study was performed using the DC04 versions of DaVinci¹ [col07a], the LHCb analysis software, and the corresponding trigger architecture. During this phase of Monte Carlo studies, aimed at understanding detector performance, there were changes to specific cut values in the trigger, but the overall architecture of the lifetime dependent trigger remained unchanged. In this section, the lifetime dependent part of the trigger will be described. A method of associating the particles used in the offline reconstruction to the online tracks used in the trigger, which was developed for this study, will also be described.

It is important to distinguish between the information seen by the trigger, and the information seen by the offline reconstruction. Because it operates under time constraints, the trigger necessarily performs only a partial reconstruction of the event, and quantities such as the momenta and impact parameters of the final state particle tracks are therefore measured with a poorer resolution than that of the offline selection. The term “online” will refer to quantities or tracks as measured by the LHCb trigger, unless otherwise specified. Similarly, the term “offline” will refer to quantities or tracks as measured by the offline event reconstruction. Throughout this document, the term “final state particle” will refer to the three D^- daughters and the π daughter of the B_d^0 .

¹References to the trigger throughout the rest of this chapter should be taken to mean the DC04 implementation of the trigger, unless otherwise stated

6.2.1 Trigger architecture

The trigger architecture relevant to this study [col03b] consists of three parts, as described more fully in Chapter 2. The L0 hardware trigger does not exploit lifetime dependent information: although the pile-up veto could in principle be lifetime biasing, it is empirically observed that this is not the case. Hence the L0 does not concern us further. The L1 trigger is a software trigger, which performs a partial event reconstruction and searches for high transverse momentum and impact parameter signatures in those events passed by the L0 trigger, and it will therefore bias the lifetime distribution of selected B hadrons. Events selected by the L1 are then passed to the high-level trigger (HLT). This is a software trigger which performs a fuller reconstruction of the event, and is in principle able to cut on any quantity which can be cut on in the offline selection. It is likely that the HLT will also bias the lifetime distribution of selected B hadrons, although its cuts have not been finalized at the time of writing this thesis. For the purposes of this thesis, the L1 trigger has been studied as a proof of principle; the HLT is expected to be a straightforward extension of this problem. It is worth noting that the L1 trigger has been absorbed into the HLT in the latest versions of the LHCb detector simulation, and this configuration is expected to be used when the LHC starts up.

6.2.2 Sub-detectors

LHCb subdetectors of particular importance to the L1 trigger are the VELO and the TT (trigger tracker) station, which were described in Chapter 2. The VELO is used for the reconstruction of primary vertices, secondary vertices, and tracks in the L1 trigger. In particular, the L1 trigger makes use of the VELO's r and ϕ sensors to reconstruct tracks, first in 2D², and then, after a cut is applied on the 2D impact parameter of the track, those tracks passing this IP cut are further reconstructed in 3D. This two stage reconstruction speeds up the decision making, since 2D tracks can be reconstructed far quicker than 3D tracks. The TT station provides limited momentum information on the 3D reconstructed tracks to the L1 trigger and allows it to discriminate between tracks based on their transverse momentum.

6.2.3 The L1 trigger

The LHCb L1 trigger was described in Section 2.7.2; here, we will restrict ourselves to discussing those parts of the L1 trigger which exploit lifetime dependent information, specifically the “generic” trigger component which searches for two tracks with high transverse momenta and high impact parameters with respect to the primary vertex. In order to operate at the required speed, the event reconstruction proceeds in two stages. First of all, tracks with hits in the VELO are reconstructed in two dimensions in the $\mathbf{r}$ - $\mathbf{z}$ plane, where

²The two dimensions selected are $\mathbf{r}$ and $\mathbf{z}$, where $\mathbf{r}$ information comes from the VELO r sensors, and $\mathbf{z}$ information comes from the VELO station registering the hit. In addition, the VELO ϕ sensors are subdivided into 8 sectors, and the identity of the sector registering the track's hits is used to provide limited information about the ϕ position of the track.

the $\mathbf{z}$ axis is defined along the LHC beamline, and $\mathbf{r}$ is the perpendicular distance to the beamline. These 2D tracks are then required to pass an impact parameter cut with respect to the primary vertex:

$$0.1 \text{ mm} < IP_{2\text{D track}} < 3.0 \text{ mm}, \quad (6.1)$$

and those that do are reconstructed in 3D. Information is included at this point from the TT station, which provides a measurement of the track's transverse momentum. In order to pass the trigger, the event must contain at least two 3D tracks which pass an impact parameter cut:

$$0.15 \text{ mm} < IP_{3\text{D track}} < 3.0 \text{ mm}, \quad (6.2)$$

and whose transverse momenta satisfy the equation

$$\ln(p_t^{(1)}/\text{MeV}) + \ln(p_t^{(2)}/\text{MeV}) > 13.915, \quad (6.3)$$

where $p_t^{(1)}$ and $p_t^{(2)}$ refer to the transverse momenta of the two highest transverse momentum tracks in the event. This decision is complicated in several ways for events where there is more than one primary vertex:

1. The impact parameters have to be evaluated independently with respect to all primary vertices.
2. All tracks which have an impact parameter less than 0.15 mm with respect to any primary vertex are discarded, even if their impact parameter with respect to the primary vertex currently being considered is greater.
3. The highest transverse momentum sum pair, as defined in Equation 6.3 is used to evaluate whether the event passes the trigger or not. The primary vertex associated with this pair is the one which yields the smallest impact parameter for the higher p_t track.

Since this is very much a proof-of-principle study, what was important was to reproduce the lifetime dependence of the trigger, not details relating to the minority of events which will have more than one primary vertex. Approximately 72% of reconstructed $B_d^0 \rightarrow D^-\pi^+$ events will have one primary vertex, and cutting out events with more than one primary vertex is not found to affect the purity of the selection. For these reasons it was decided to select only those events for which both the trigger and offline reconstruction reconstruct exactly one primary vertex. This has the added benefit of making the association of offline and online primary vertices trivial.

6.3 The likelihood for signal events

The acceptance function and biased B lifetime will be fitted using a multivariate unbinned log-likelihood fit, whose log-likelihood formula will be derived in this section. This method was originally developed at CDF, and has previously been described in detail in [ABH⁺03] and [Rad07].

6.3.1 Notation

This and subsequent sections are largely concerned with manipulating probabilities, and the following notation is adopted throughout:

- $P(x)$ the probability of x ;
- $P(x, y)$ the probability of x and y ;
- $P(x|y)$ the probability of x given y ;
- $|x\rangle$ a column vector;
- $\langle x|$ a row vector.

Bayes' theorem is frequently used and is restated for convenience:

$$P(x|y) = \frac{P(x) P(y|x)}{P(y)}, \quad (6.4)$$

which leads to the relation:

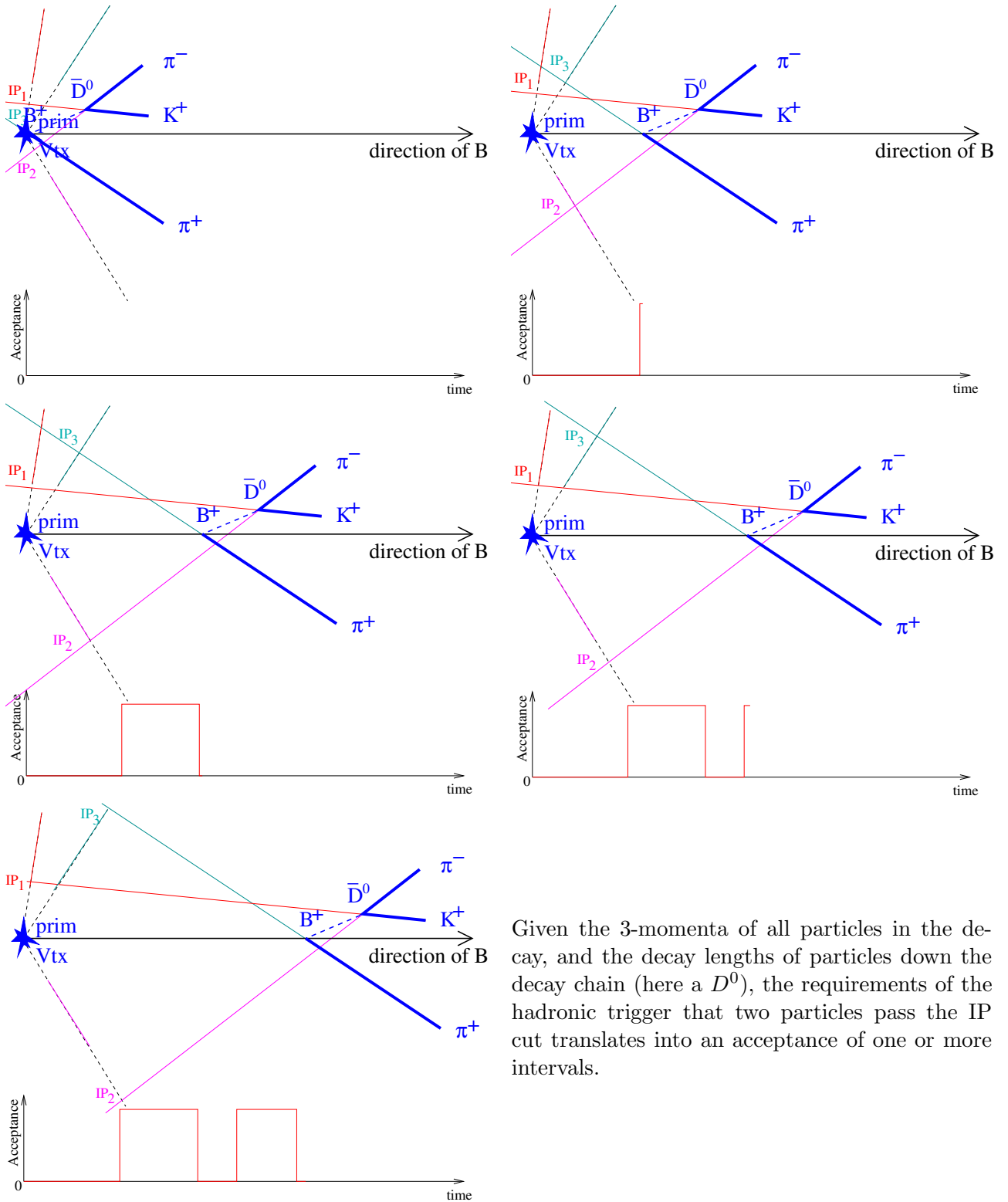
$$P(x, y) = P(x) P(y|x). \quad (6.5)$$

6.3.2 Overview

A B meson or baryon originates at the primary vertex of the event, and its decay position and kinematics determine the subsequent position of its daughter tracks. Specifically, for any given set of kinematics, there is a one-to-one correspondence between the decay vertex of the B and the impact parameter of the final state particle tracks. It follows that for any given set of event kinematics, it is possible to build up the trigger acceptance function for that event by “swimming” the B along its direction of momentum, adjusting the positions of the final state tracks accordingly, and evaluating the trigger response each time. Note that what is actually moved is the primary vertex of each event, not the B or any of the tracks associated with its final state particles. This is done in order to avoid the difficulties which would arise when moving individual detector hits, owing to the discrete nature of the VELO.

Figure 6.2 demonstrates the principle of this swimming. The dashed lines represent the perpendicular of a track with respect to the primary vertex, and they are solid in those areas where the track passes the impact parameter cut. For any given position of the B decay vertex, the event either passes the trigger or it does not; as a result, the value of the acceptance function is always either 1 (accept) or 0 (reject), and the acceptance function is a series of top-hats. What makes the acceptance function and likelihood simple is that the kinematics which translate the impact parameter cuts into a lifetime cut do not depend on the lifetime of the B . As a result, there is a one-to-one correspondence between the kinematics

Figure 6.2: Illustration of the link between impact parameter requirements and cuts on the decay time. Whenever 2 tracks pass the IP requirements, the acceptance is set to one, otherwise zero. [ABH⁺03]



Given the 3-momenta of all particles in the decay, and the decay lengths of particles down the decay chain (here a D^0), the requirements of the hadronic trigger that two particles pass the IP cut translates into an acceptance of one or more intervals.

of a B decay and the lifetime acceptance function for that decay. The probability of finding an event with a given lifetime and acceptance function is:

$$P(t, \text{acc}) = P(t|\text{acc})P(\text{acc}). \quad (6.6)$$

The term $P(\text{acc})$ is the probability of finding the given acceptance function, which only depends on the kinematics of the B decay in this event. It does not depend on the mean lifetime of the B meson because the decay kinematics are independent of the lifetime, and it will factor out of the eventual likelihood. The term $P(t|\text{acc})$ represents the probability of finding an event with lifetime t given the measured acceptance function, which constrains the lifetimes between $t_{\min}$ and $t_{\max}$. Expanding this term in the equation, and assuming no measurement errors for now:

$$P(t, \text{acc}) = \frac{\frac{1}{\tau} e^{-\frac{t}{\tau}}}{e^{-\frac{t_{\min}}{\tau}} - e^{-\frac{t_{\max}}{\tau}}} P(\text{acc}), \quad (6.7)$$

where τ is the mean B lifetime, and $P(t, \text{acc}) = 0$ if t is not between $t_{\min}$ and $t_{\max}$. This equation is correct for an acceptance function which consists of only a single top-hat time interval. For a more complicated event, it is necessary to sum over all the top-hats:

$$P(t, \text{acc}) = \frac{\frac{1}{\tau} e^{-\frac{t}{\tau}}}{\sum_{i=1}^n \left(e^{-\frac{t_{\min}}{\tau}} - e^{-\frac{t_{\max}}{\tau}} \right)} P(\text{acc}). \quad (6.8)$$

Therefore, for a set of N events, in the absence of any background or detector effects, the log-likelihood is given by:

$$\log(L) = -N \log(\tau) - \sum_{i=1}^N \left[\frac{t_i}{\tau} + \sum_{j=1}^{n_i} \log \left(e^{-\frac{t_{\max_{ij}}}{\tau}} - e^{-\frac{t_{\min_{ij}}}{\tau}} \right) \right], \quad (6.9)$$

where the term involving $P(\text{acc})$ has been omitted since it does not depend on τ , and hence will not influence the fit.

6.3.3 Including measurement errors

In practice, the simple likelihood quoted above will have to be modified to take into account errors on both the measured lifetimes and impact parameters. In particular, the fact that the trigger cuts on online quantities, while the swimming is performed using offline information, must be accounted for. In principle, any given track in the event is reconstructed online and offline, as are the primary vertices, and since the two reconstructions are performed with different resolutions the online and offline impact parameters will be different for the same track. Therefore, there is no longer a direct correspondence between the B lifetime measured using the **offline** reconstruction and the impact parameters measured using the **online** reconstruction. The **offline** B lifetime directly corresponds to the **offline** impact parameters, and a correspondence will have to be established between the **offline** and **online** impact parameters.

Before this is dealt with, the probability of finding an event needs to be modified to include the error on the measurement of that lifetime. Since the trigger cuts on the measured lifetime, the probability will have to be integrated over all true event lifetimes. Assuming a Gaussian resolution function, the probability of measuring a time t_m given a true decay time t and a lifetime resolution of σ_t is:

$$P(t_m|t, \sigma_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-t_m)^2}{2\sigma_t^2}}. \quad (6.10)$$

In principle, non-Gaussian resolution functions can be accounted for, however the resolution function in the LHCb detector simulation is Gaussian, and hence that is the function used here³.

Writing $A(t_m)$ for the acceptance as a function of the measured decay time, the probability of observing this measured decay time, with the given time resolution and overall acceptance function, is given by:

$$P(t_m, \sigma_t, \text{acc}) = \frac{\int_0^{+\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-t_m)^2}{2\sigma_t^2}} A(t_m) dt}{\int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-t_m)^2}{2\sigma_t^2}} A(t_m) dt dt_m}. \quad (6.11)$$

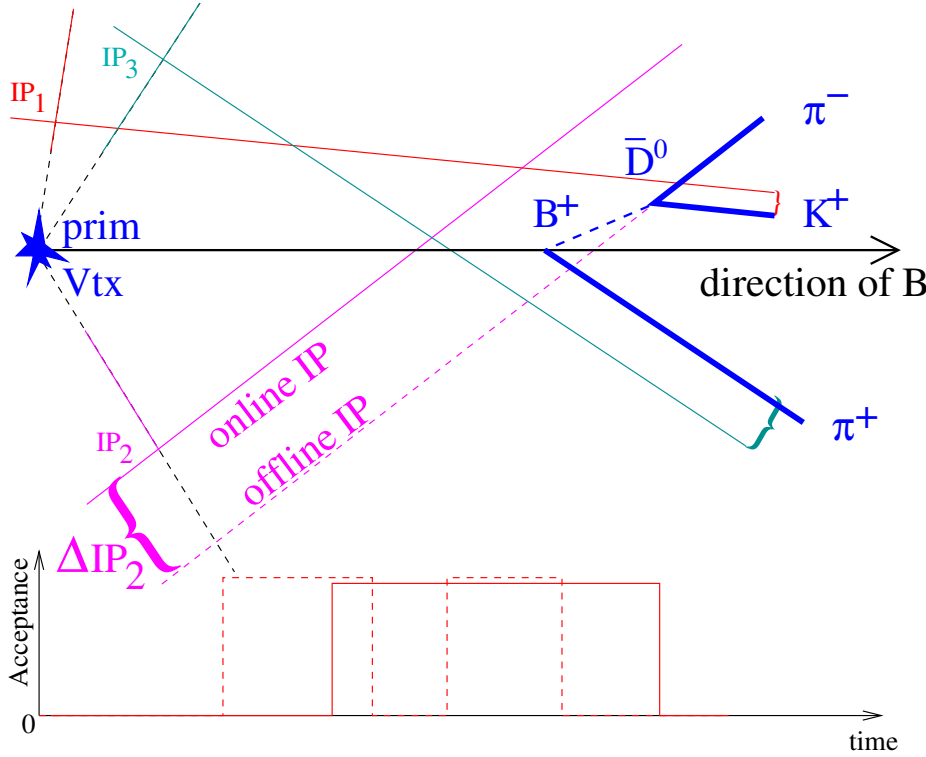
The difference between the acceptance as a function of the measured lifetime and the acceptance as a function of the true lifetime is simply that the acceptance intervals are calculated from the measured impact parameters - the acceptance remains a series of step functions. The measured impact parameter is the impact parameter as measured by the trigger, which will in principle differ from the impact parameter as measured by the offline reconstruction, and this difference needs to be accounted for. So long as the average difference between the online and offline impact parameters is independent of the B lifetime, this difference can be accounted for by measuring it at the **reconstructed** B lifetime, and subtracting the measured difference from the offline impact parameters to get the online impact parameters during the swimming.

Figure 6.3 illustrates the principle. The difference between the online and offline impact parameters is measured at the one point at which it is known, which is the actual decay position of the B . The swimming is then performed using the offline quantities, and the offline impact parameter is converted into an online impact parameter at each point by adjusting for the measured difference.

As described above, there are two kinds of online impact parameter within the LHCb lifetime dependent trigger: 2D and 3D. For the method to work, both quantities need to have a difference with respect to the offline impact parameter which is independent of the lifetime. Figures 6.4 and 6.5 show the average impact parameter difference plotted against the offline measured impact parameter. The differences are essentially flat with impact parameter, although some deviation is observed in the region around 0 for the both the 2D and 3D impact parameters. It should be noted, however, that such deviation is of the order of 0.005 mm, while the minimum 2D IP of a track passing the trigger is 0.1 mm and the

³Figure 3.8 in Chapter 3 shows a plot of the LHCb Monte Carlo lifetime resolution. Note that, if necessary, a more complicated resolution function could be built by adding together Gaussians.

Figure 6.3: Re-establishing the direct link between online impact parameter cuts and measured lifetime (measured from offline data). [ABH⁺03]



minimum 3D IP is 0.15 mm. Nonetheless, this is a potentially important matter, not least because it occurs precisely in the area around the lower impact parameter cut for 2D tracks. Since the reasons for this behaviour are not currently understood, it was decided to approach the problem in a pragmatic fashion, and attempt to find a relation between the online and offline impact parameters which produces a flat function usable for swimming. As it turns out, it is enough to multiply the impact parameter difference by the measured offline impact parameter – the resulting distributions, shown in Figures 6.6 and 6.7 are now quite flat with impact parameter.

It is now possible to derive the full probability to measure an event at time t_m , following [ABH⁺03]. The acceptance, as a function of the measured time t_m , is given by:

$$A(t_m) = \sum_{i=\text{all intervals}} (\theta(t_m - t_{\min i}) - \theta(t_m - t_{\max i})), \quad (6.12)$$

where $\theta(t)$ is the Heaviside function, and “all intervals” refers to a sum over all the top-hats in the acceptance function. Inserting this in Equation 6.11:

$$P(t_m, \sigma_t, \text{acc}) = \frac{\int_0^{+\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi i \sigma_t}} e^{-\frac{(t-t_m)^2}{2\sigma_t^2}} dt}{\sum_{i=\text{all intervals}} \int_{t_{\min i}}^{t_{\max i}} \int_0^{+\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-t_m)^2}{2\sigma_t^2}} dt dt_m}. \quad (6.13)$$

As this probability must be calculated often in the fitting code, it is profitable to rewrite it

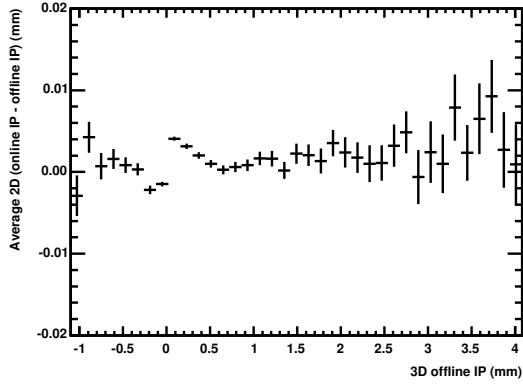


Figure 6.4: The average difference between 2D online and offline impact parameters, versus the measured offline 2D impact parameter.

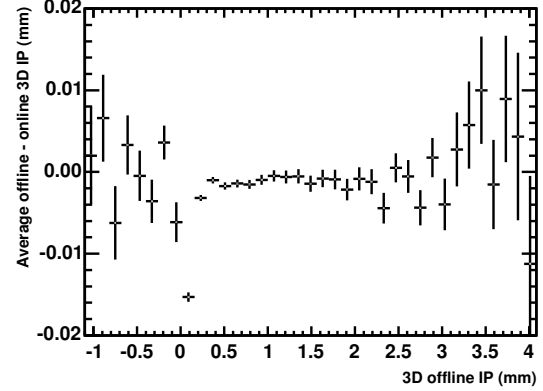


Figure 6.5: The average difference between 3D online and offline impact parameters, versus the measured offline 3D impact parameter.

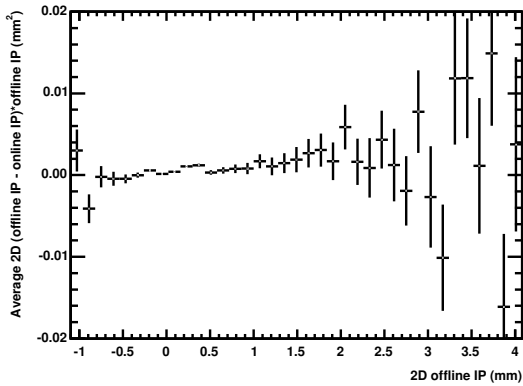


Figure 6.6: The average of the 2D offline impact parameter multiplied by the difference between 2D online and offline impact parameters, versus the measured offline 2D impact parameter.

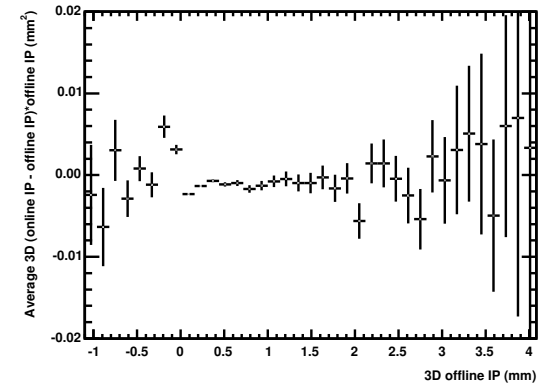


Figure 6.7: The average of the 3D offline impact parameter multiplied by the difference between 3D online and offline impact parameters, versus the measured offline 3D impact parameter.

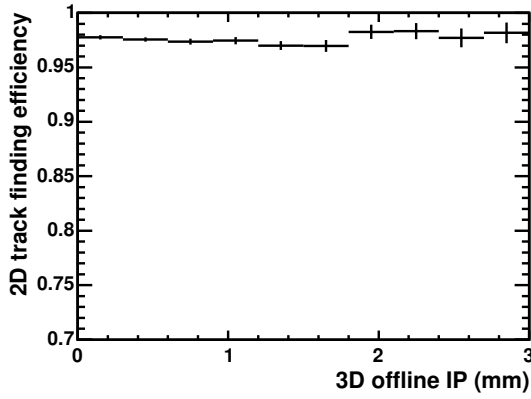


Figure 6.8: The 2D online track finding efficiency versus the 3D impact parameter of the offline track.

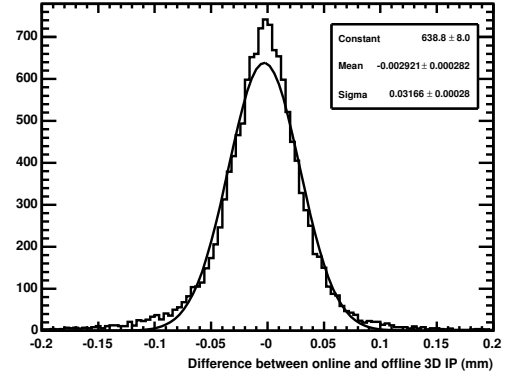


Figure 6.9: The online-offline IP resolution for 3D tracks.

using the frequency function:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy, \quad (6.14)$$

since there exist fast numerical algorithms for computing $F(x)$ within standard C++ libraries. Finally, using Equations 6.13 and 6.14:

$$P(t_m, \sigma_t, \text{acc}) = \frac{\frac{1}{\tau} e^{-\frac{t_m}{\tau}} F\left(\frac{t_m}{\sigma} - \frac{\sigma}{\tau}\right)}{\sum_{i=\text{all intervals}} \left[-e^{-\frac{t}{\tau}} F\left(\frac{t_m}{\sigma} - \frac{\sigma}{\tau}\right) + e^{-\frac{\sigma^2}{2\tau^2}} F\left(\frac{t}{\sigma}\right) \right]_{t=t_{\min i}}^{t=t_{\max i}}}. \quad (6.15)$$

6.3.4 Online track finding efficiency

So far, it has been implicitly assumed that the trigger will always find an online track corresponding to the offline reconstructed track, or at the very least that its probability to find an online track is independent of the impact parameter of that track. As Figure 6.8 shows, this is indeed true for 2D tracks. A difference in efficiency is observed between signal and background events, as a track from a signal event has approximately a 98% chance of having a matching 2D online track, and a track from a background event approximately a 92% chance. Since the efficiencies are so high, it was decided to only accept those events where all four offline tracks have a matching 2D online track, which has the added benefit of improving purity.

The situation seems more complicated for 3D tracks, because they are only reconstructed if the matching 2D track passes an impact parameter cut of its own. This can lead to a situation where a 2D reconstructed track fails the 2D IP cut at the decay lifetime of the B , and hence no 3D track is reconstructed, but the 2D track would have passed the IP cut at some point during the swimming. At this point, a 3D track would have been reconstructed and used in the trigger decision, but its online information is now unavailable. Our method requires that we construct a replacement for this missing online information. In order to simulate this missing online information, the IP and transverse momentum of the offline

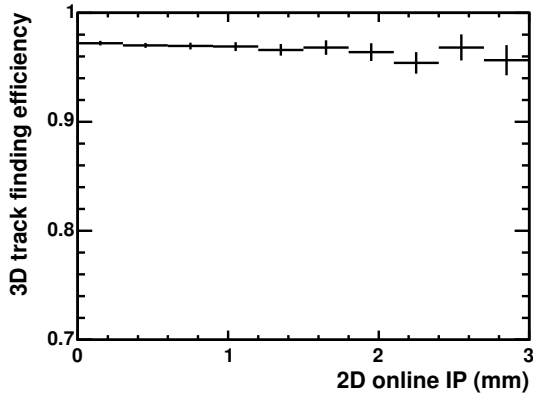


Figure 6.10: The track finding efficiency of 3D online tracks matched to 2D online tracks passing the 2D IP cut, versus the 2D impact parameter of the matching 2D track.

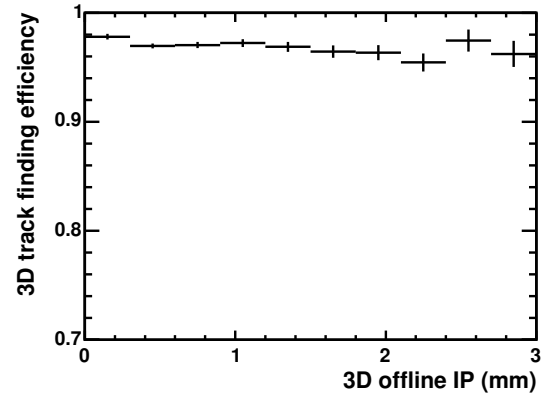


Figure 6.11: The track finding efficiency of 3D online tracks matched to 2D online tracks passing the 2D IP cut, versus the 3D impact parameter of the matching offline track.

track are smeared by the appropriate online-offline resolution, shown in Figure 6.9. The resolution exhibits a bias of $3 \mu\text{m}$, but as this is ten times smaller than the $30 \mu\text{m}$ width of the resolution, it will be ignored. The fitter therefore sees both a 2D and a 3D online track partner for every offline track, and the efficiency of finding a 3D track given that a 2D track passes the impact parameter cut is assumed to be 100%.

In order not to introduce a bias, the reconstruction efficiency for 3D tracks whose corresponding 2D track passes the impact parameter cut must be flat with respect to impact parameter. Figure 6.10 shows the 3D track finding efficiency versus the 2D impact parameter of the matching 2D track, while Figure 6.11 shows the 3D track finding efficiency versus the 3D impact parameter of the matching offline track, and both are indeed essentially flat. It is found that, as with 2D tracks, offline tracks from signal events are more likely to have a matching online 3D track than those from background events.

The unequal online track finding efficiencies for tracks from signal and background events are a potential source of error. When LHCb starts, it will be possible to evaluate the trigger performance and decide whether the track finding efficiencies need to be accounted for. For the moment, the fits are simplified by assuming the online track finding efficiencies to be 100%, without a detrimental effect on performance.

6.4 The likelihood including background events

6.4.1 The problem

The addition of background complicates the likelihood function because the term involving $P(\text{acc})$, representing the probability of finding an event with a given acceptance function calculated from its kinematics, no longer factors out. This can be seen by separating Equation

6.6 into signal and background parts:

$$P(t, \text{acceptance}) = P(s) P(t|\text{acc}, s) P(\text{acc}|s) + P(b) P(t|\text{acc}, b) P(\text{acc}|b). \quad (6.16)$$

Here, $P(s)$ is defined as the probability that an event is signal, and $P(b) = 1 - P(s)$ as the probability that an event is background. Previously, it was possible to argue that the probability of finding a given acceptance function depended only on the event kinematics, and hence factored out of the likelihood. With the addition of the background, however, what has to factor out is the probability of finding a given acceptance function given that the event is signal or background – i.e. signal and background events would have to, on average, share the same kinematics⁴. In order to see the problem more clearly, it is useful to rewrite this probability:

$$P(t, \text{acceptance}) = P(s|\text{acc}) P(t|\text{acc}, s) P(\text{acc}) + P(b|\text{acc}) P(t|\text{acc}, b) P(\text{acc}). \quad (6.17)$$

The factor $P(\text{acc})$ factors out as before, since it simply represents the overall probability of finding an acceptance function in the dataset. That leaves two kinds of terms to be accounted for in the probability:

1. $P(s|\text{acc})$ and $P(b|\text{acc})$ represent the probability that an event with a given set of kinematics is signal or background
2. $P(t|\text{acc}, s)$ and $P(t|\text{acc}, b)$ represent the probability of finding an event with lifetime t given the kinematics and the fact that the event is signal or background.

The first term represents a specific manifestation of a more general problem: discriminating between signal and background events based on the kinematic characteristics of the event. The method chosen in this fitter is that of Fisher discriminants, and will be described in Section 6.5. First, however, it will be necessary to rearrange the second term so that it can be calculated in the likelihood. The subsequent section draws heavily on the analogous derivation in [ABH⁺03].

6.4.2 The full likelihood

The likelihood derived in Section 6.3 will now have to be modified to take into account the presence of background, and in particular, a mass term will have to be added to the likelihood to help separate signal and background events. The log likelihood is given by:

$$\log(L) = \sum P(t, \text{acceptance}). \quad (6.18)$$

The following physical parameters are used in the fit:

- t_m – the measured lifetime;

⁴This is quite implausible, and would anyway have rather unfortunate implications for background discrimination if true.

- σ_t – the error on the measured lifetime;
- m – the mass of the reconstructed B ;
- acc – the acceptance function generated by swimming the event;

By definition, the only events which will be fitted are events that have passed the trigger – hence the required probability is:

$$P(t, \text{acceptance}) = P(t_m, \sigma_t, m, \text{acc} | \text{trigger}), \quad (6.19)$$

which can be broken up into signal and background parts:

$$P(t_m, \sigma_t, m, \text{acc} | \text{trigger}) = P(s, t_m, \sigma_t, m, \text{acc} | \text{trigger}) + P(b, t_m, \sigma_t, m, \text{acc} | \text{trigger}). \quad (6.20)$$

In the absence of background, $P(t_m, \sigma_t, \text{acc})$ is given by Equation 6.15; what is required, however, is a final probability which can be easily calculated from the observed quantities t_m , σ_t , m , and acc . Hence this probability must be rearranged in a non-trivial fashion, details of which are given in Appendix D. The two most important assumptions made during this derivation are:

1. The reconstructed mass, m , only depends on whether the event is signal or background, and is independent of the decay kinematics, measured lifetime, and the error on the measured lifetime. Any dependence of this mass on the decay kinematics will be negligible because of LHCb's excellent mass resolution.
2. The measured lifetime, t_m , is independent of the decay kinematics. This is a sensible assumption since the measured lifetime is the difference between the reconstructed decay vertex and reconstructed primary vertex positions, and those will not depend on the momenta or lifetimes of the B daughters.

The final probability, for signal, is given by:

$$P(s, t_m, \sigma_t, m, \text{acc} | \text{trigger}) = \frac{P(t_m | \sigma_t, s) P(m | s) P(\text{trigger} | \text{event}) P(s | \text{acc}, \sigma_t, \text{trigger})}{P(\text{trigger} | \text{acc}, \sigma_t, s)}. \quad (6.21)$$

The individual components are given by:

- $P(t_m | \sigma_t, s)$ – the probability of measuring a lifetime t_m for a signal event with mean lifetime τ and lifetime error σ_t is given by:

$$P(t_m | \sigma_t, s) = \frac{1}{\tau} e^{\frac{-t}{\tau} + \frac{\sigma_t^2}{2\tau^2}} F\left(\frac{t_m}{\sigma} - \frac{\sigma}{\tau}\right). \quad (6.22)$$

- $P(m | s)$ – the mass distribution is given by a Gaussian:

$$P(m | s) = \frac{1}{\sqrt{2\pi}\sigma_m} e^{\frac{-(m-\mu_m)^2}{2\sigma_m^2}}. \quad (6.23)$$

As with lifetime errors, a mean mass error is used for all events.

- $P(\text{trigger}|\text{acc}, \sigma_t, s)$ – this probability is obtained by integrating the probability of finding a measured lifetime over the acceptance function:

$$P(\text{trigger}|\text{acc}, \sigma_t, s) = \sum_{i=\text{all intervals}} \int_{t_{\min i}}^{t_{\max i}} \frac{1}{\tau} e^{\frac{-t}{\tau} + \frac{\sigma^2}{2\tau^2}} F\left(\frac{t_m}{\sigma} - \frac{\sigma}{\tau}\right) dt_m. \quad (6.24)$$

The probability for background events has the same form, but the mass distribution is of course different; it will be measured from the sidebands once the experiment is running.

6.4.3 The background model

In order to replicate the above derivation for background events, a model for the lifetime behaviour of background events in the absence of a trigger is required. The background lifetime distribution is described by a sum of exponentials with different mean lifetimes, which are subject to the same swimming procedure as the signal. A zero lifetime prompt component is also included to take account of any background from the primary vertex.

In order to discriminate between signal and background, the fitter needs to know what the background looks like in the absence of signal. The background can be modeled by events in the mass sidebands, and a simultaneous mass and lifetime fit is performed using this information to determine the probability that an event is signal. Such an approach assumes that the lifetime profile of the background under the mass peak is the same as that in the sidebands, and is therefore a possible source of systematic error in the fit. In particular, the lower mass sideband can contain events where a real B decay was partially reconstructed, leading to a systematically low mass of the reconstructed B meson. Because such events will not be found under the signal, the lower mass sideband cannot be used to model the background under the signal, while the higher mass sideband can. The expected makeup of the background was discussed in Chapter 3. The majority of the background will come from mis-identified or partially reconstructed decays of B mesons and baryons, and will therefore mimic the lifetime profile of the signal.

6.5 The Fisher discriminant

In order to calculate the full likelihood with background in Equation 6.21, the signal fraction needs to be parametrized as a function of the acceptance function, $P(s|\text{acc})$. To facilitate this parameterization a single number is associated to each acceptance function, allowing $P(s|\text{acc})$ to be written as $P(s|\text{number}(\text{acc}))$. To ensure that $P(s|\text{number}(\text{acc}))$ is in fact a good approximation to $P(s|\text{acc})$ this number needs to be associated in a way that minimizes the information loss regarding the “signal-ness” of the acceptance function, i.e. a number that discriminates well between signal and background.

Fisher discriminants are a reasonably simple measure of how likely any given event is to be signal (or background). A set of parameters describing each event is projected down to a single number, the Fisher scalar, that provides optimal signal-background separation. The

input parameters in this case are those describing the acceptance functions. The number of parameters describing the acceptance functions is unclear, because there can be 1, 2, ... n distinct top-hat functions. Therefore a brute-force approach is chosen, and the time axis split into a number of bins. The value of an acceptance function at a given bin is one parameter in the parameter set, the number of parameters is the number of bins. Note that in this analysis, the Fisher discriminant is used purely to obtain the signal fraction of each event – it is not used as a cut, and no events are thrown away, whatever their fisher discriminant value is.

6.5.1 Discriminating between signal and background events

Consider a sample of signal and background events with certain distributions in two variables, x and y , as illustrated in Figure 6.12(c). The means of x and y for the the distributions,

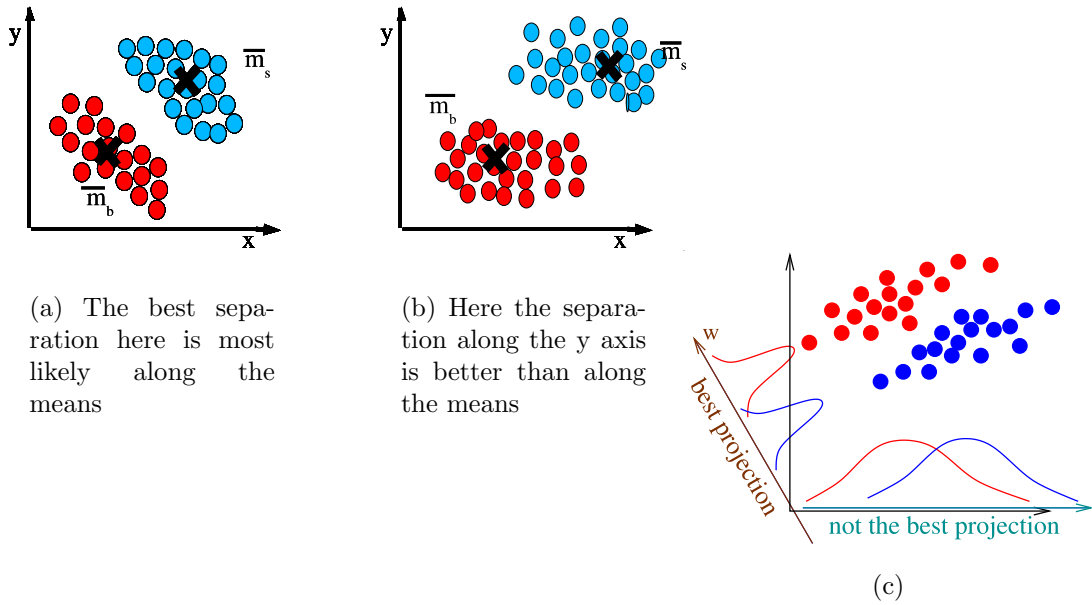


Figure 6.12: This illustrates that it is necessary to take the means (marked with a cross) and spread of each variable into account when finding the direction of best separation[ABH⁺03]

$\bar{m}_s = (\bar{x}_s, \bar{y}_s)$ and $\bar{m}_b = (\bar{x}_b, \bar{y}_b)$, are marked with crosses on the diagrams. In order to discriminate between signal and background, they need to be projected in such a direction that the separation of the means is large, while the spread around each individual mean is small. The mathematical details of this separation are shown in Appendix E.

6.5.2 Representing the acceptance function as a vector

The trigger acceptance function for an event is fully determined by the event kinematics, and can be represented as a histogram. The acceptance function is calculated using a binning which captures all features, i.e. those points where the acceptance function changes from 0 to 1 or vice-versa, since these fully define the acceptance function. The result is a vector

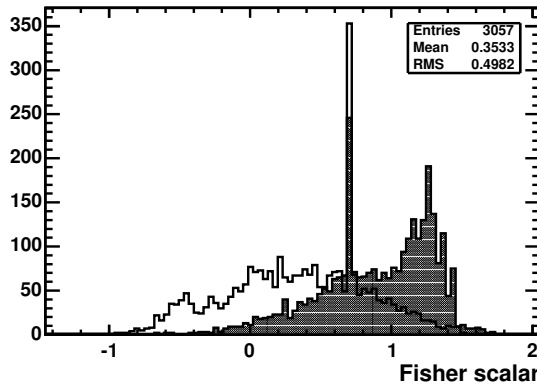


Figure 6.13: Fisher scalar distribution for a sample of $\sim 3,000$ signal (shaded) and $\sim 3,000$ background (unshaded) events.

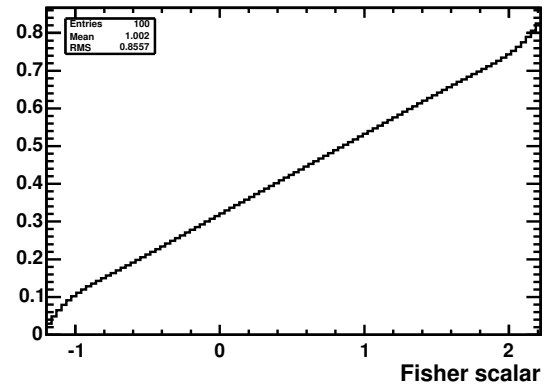


Figure 6.14: The signal probability associated with a value of the scalar Fisher discriminant in Figure 6.13

of zeroes or ones representing the acceptance function, with each vector component corresponding to a different lifetime bin. It is possible to use up to 20 bins for the final binning of the acceptance function, but 5 bins have typically been found sufficient in this study.

6.5.3 Example Fisher distribution

An example of a Fisher scalar distribution for a sample containing $\sim 3,000$ full LHCb Monte Carlo signal and $\sim 3,000$ full LHCb Monte Carlo background events can be seen in Figure 6.13. Figure 6.14 shows the probability of an event being signal fitted for the values of the scalar Fisher discriminant. The probability of an event being signal increases smoothly with the value of the scalar Fisher discriminant, and most of the events have high values of this discriminant, as would be expected in a sample in which the majority of events are signal. The spike which is seen in the middle of Figure 6.13 is not understood.

6.6 Accessing the online data

Having developed the method thus far, two matters remain: associating an online track to an offline track, and accessing the trigger decision when swimming the event.

6.6.1 Associating online and offline tracks

In order to correct for the resolution difference between online and offline impact parameters when swimming the event, an association has to be made between online and offline tracks in the event. In common with other LHCb associators, it was decided to match the tracks based on the number of hits they have in common in the VELO sub-detector: two tracks were considered a match if they shared more than 70% of hits in the VELO. A publicly available DaVinci tool was written which enabled this association to be performed, and is

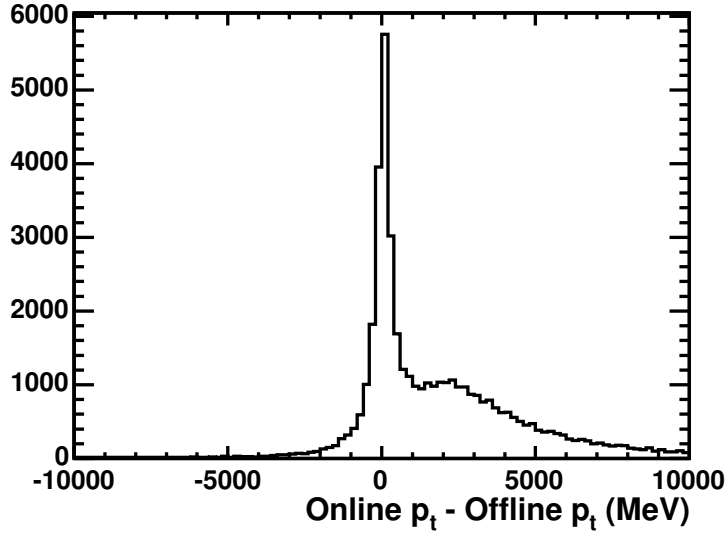


Figure 6.15: The difference between the measured online and offline transverse momentum of final state tracks from the decay $B_d^0 \rightarrow D^- \pi^+$.

documented more fully in [Gli05].

As discussed in Section 6.2, the L1 trigger uses two types of track: ones reconstructed in 2D, and ones reconstructed in 3D. Compared to offline tracks, which are reconstructed using the full detector information, 2D tracks are reconstructed using only the VELO R-segments. The 3D tracks add the VELO ϕ -sensors, and the four-layer TT station to the reconstruction. The limited amount of information available to the online reconstruction and the use of a “fast” pattern recognition both contribute to the resolution difference between the online and offline impact parameters.

Not all offline tracks with a 2D online partner will also have a 3D online partner, since not all the 2D online tracks will have passed the required impact parameter cut in the trigger. However, during the swimming, these tracks may be found to pass the impact parameter cut, and “online” information for the relevant 3D tracks will also be required. Therefore it is necessary to generate “online” impact parameters and momenta for these tracks, and this is done by smearing the relevant offline quantities by the measured distributions of the online-offline resolutions. For example, a histogram of the difference between the online transverse momentum and the offline transverse momentum is made for all offline tracks with a 3D online partner. A random value is then picked from this histogram, and used to generate “online” transverse momenta for offline tracks without a 3D online partner. It should also be noted that there is a finite probability of $\sim 20\%$ that the L1 reconstruction will not measure a transverse momentum even when a 3D online track exists. As in the L1 trigger, such tracks are assigned a default transverse momentum of 400 MeV. The difference between the online and offline transverse momenta for all tracks which had a measured online transverse momentum is shown in Figure 6.15. Its shape is not understood at the time of writing, but is indicative of significant biases in the online momentum estimation when compared to the measurement of offline momenta.

Table 6.1: Summary of TIS, TOS, and TOB events in the signal sample.

Event Category	Number of signal events in category
TIS	2107
TOS	11230
TOB	2119

6.6.2 Trigger emulation

Since it is not currently possible to re-run the online trigger in order to swim an event⁵, an offline emulation is required. This emulation has to reproduce only the lifetime dependent parts of the trigger. Following [Dji04, Wit], three types of event are distinguished:

- **TIS** (Triggered independently of signal): Events which pass the trigger without any final state products of the signal B being used in the trigger decision. These events are lifetime unbiased by definition, since swimming the signal has no impact on the trigger decision.
- **TOS** (Triggered on signal): Events which pass the trigger using only final state products of the signal B in the trigger decision, and which could not pass the trigger if the signal B and its decay products were not present in the event. These events are lifetime biased, and their acceptance functions have to be measured by swimming them.
- **TOB** (Triggered on both): Events which requires both final state tracks from the signal B and other tracks in the event in order to pass the trigger. These events cannot be accounted for at present, but they represent a much smaller fraction of the signal sample than either TIS or TOS events..

The trigger emulation runs on the offline final state particles for all TOS events, demanding that two tracks satisfy the L1 impact parameter cuts and satisfy 6.3. Because of the peculiar nature of 2D impact parameters, it has also been necessary to write routines calculating these for offline tracks (when swimming) in a manner analogous to that for online 2D tracks. Table 6.1 shows the number of TIS, TOS, and TOB events in our signal sample.

6.7 Fitting to signal events

This section shows results obtained by fitting the B_d^0 lifetime to a pure signal sample, both in the case of the full LHCb Monte Carlo for the channel $B_d^0 \rightarrow D^- \pi^+$, and in the case of a toy

⁵An interface which would allow this is currently under discussion and is expected to be available in the near future as part of the LHCb offline analysis framework.

Monte Carlo created for the same channel. The expected annual yield for this channel, using the optimal $\frac{S}{\sqrt{S+B}}$ selection developed in Chapter 3, is 1340k events once all triggers have been taken into account. However only around 20k fully reconstructed events are available from the full Monte Carlo once the L1 trigger is applied, and far fewer if the L0 and HLT triggers are also applied. This is why a toy Monte Carlo is needed to estimate the expected statistical precision of this method.

6.7.1 Full Monte Carlo fit

The signal fit in Figure 6.16 was created using a sample of 10571 signal events which passed the online LHCb L1 trigger. The generator value of the mean B_d^0 lifetime is 1.534ps. The fit result is:

$$\text{lifetime} = 1.532 \pm 0.017\text{ps},$$

$$\frac{\chi^2}{ndof} = 1.35,$$

$$\chi^2 \text{ probability} = 2.9\%,$$

which is in good agreement with the input value, and has an acceptable $\frac{\chi^2}{ndof}$. The fit residual is seen in Figure 6.17, and corresponds to a straight line with a slope and intercept which are compatible with 0. The L1 lifetime acceptance function is shown in Figure 6.18 with the fitted acceptance function overlaid.

A toy Monte Carlo for $B_d^0 \rightarrow D^- \pi^+$

Because the fitter measures the acceptance function on an event by event basis, any toy Monte Carlo has to contain all the information required to measure such a function for each event. This means generating not only the lifetime distribution, but also the other variables used in the fit, such as the momenta of the final state particles, decay vertex positions, B and D lifetimes, and so on. Moreover, these distributions should be a good approximation to the analogous distributions in the full Monte Carlo.

The chosen method starts by generating distributions for those variables in the event that can be assumed to be independent of each other. The B, D masses and lifetimes are generated according to the relevant distributions (Gaussian and exponential respectively), using the current world average values [ea06c], and resolutions measured from the full Monte Carlo data. The B lifetime used in the generation is 1.534ps. The rest of the independent variables are generated from their corresponding distributions as measured on the full Monte Carlo. All other variables in the event are then calculated from these independent variables. All the independent and dependent variables used in the toy Monte Carlo are listed below.

- Independent variables: Primary Vertex position, B lifetime, B mass, B momentum, B vertex χ^2 , D mass, D lifetime, D vertex χ^2 , errors on measured impact parameters.

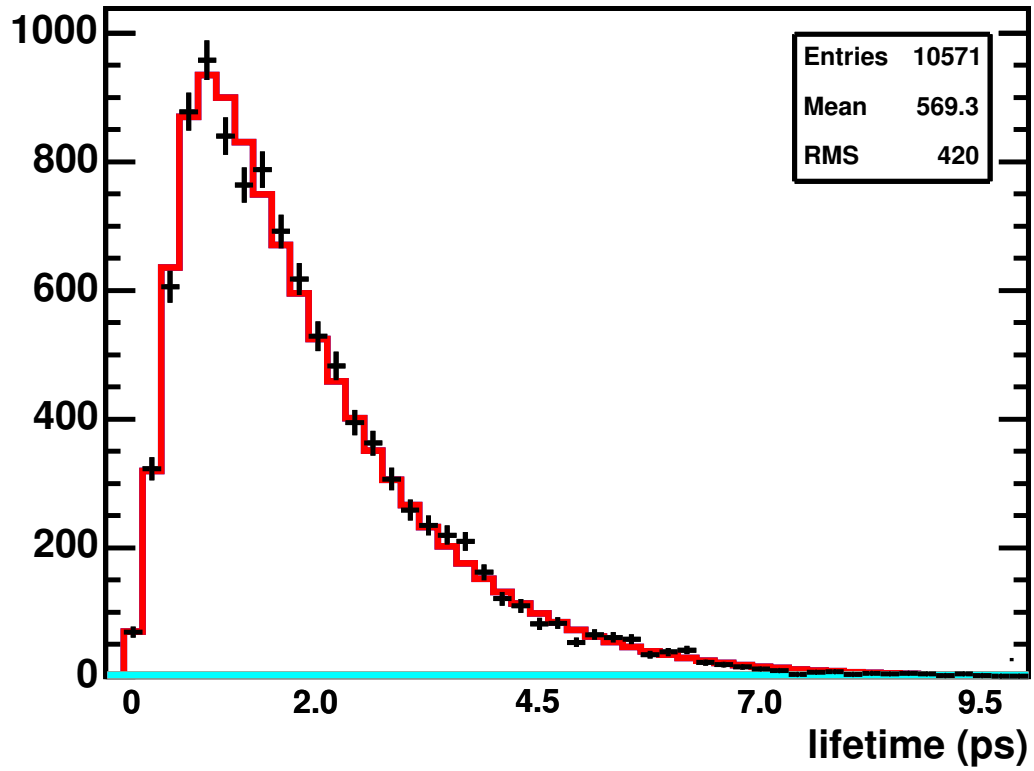


Figure 6.16: Lifetime fit for 10571 full Monte Carlo signal events.

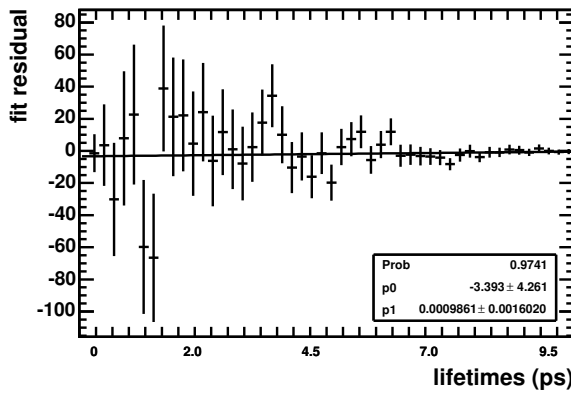


Figure 6.17: The fit residual in Figure 6.16.

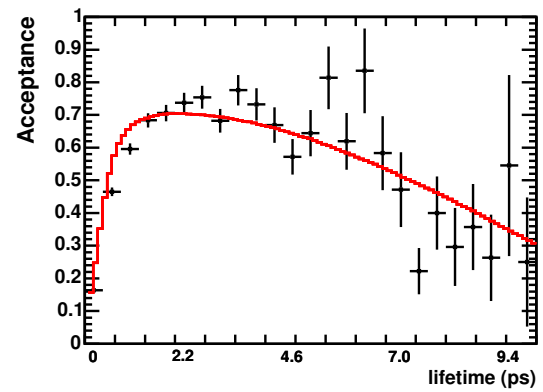


Figure 6.18: The L1 lifetime acceptance function (data points), with the fitted acceptance function (solid red line) overlaid.

- Dependent variables: B vertex position, D vertex position, D momentum, D impact parameter, final state particle momenta, final state particle impact parameters, online quantities (discussed further below).

The process is best illustrated by working our way through the generation of an event. The primary vertex position and B momentum, are generated from their measured distributions in the full Monte Carlo, while the B mass and lifetime are generated from Gaussian and exponential distributions respectively. Next, the D mass is generated from a Gaussian distribution, and the four momenta of the D and the bachelor pion from the B daughters are calculated by taking the B four momentum and decaying it with the ROOT package TGenPhaseSpace⁶. The B decay vertex is calculated from the primary vertex position, B lifetime and B momentum, and is assigned a χ^2 value based on the measured distribution in the full Monte Carlo. The D is assigned a lifetime according to an exponential distribution, and the D decay vertex position is similarly calculated. Finally the four momenta of the D daughters, and the impact parameters of all the final state particles are calculated. All impact parameters are assigned an error, which is generated from the measured distribution of impact parameter errors in the full Monte Carlo.

The fact that “online quantities” are listed as a dependent variable requires further explanation. In order to perform a realistic swimming, and mimic the L1 trigger decision, the online impact parameters and momenta for the final state particles in this decay must be generated. The first step is to measure the distribution of the difference between these online and offline quantities on the full Monte Carlo. The offline quantities generated for the toy Monte Carlo are then smeared by these distributions. As the L1 trigger only uses one tracking station, it has a probability of 79.2% of measuring an online momentum for a given online track. This probability is not lifetime dependent and is applied to the online momenta, with an online momentum of 400 MeV being assigned to those tracks which fail the test. This is the default momentum assigned to such tracks in the online L1 trigger.

In order to make the toy Monte Carlo more realistic, several detector resolution effects are included:

- Momentum resolution: The measured momenta are obtained by smearing the generator level momenta with a resolution of 0.37%, taken from [col03a].
- Charged track acceptance: For all generated final state particle momenta, the measured detector reconstruction efficiency [col03a] is applied to determine if the particle with this momentum was reconstructed by the detector. If not, the event is discarded.
- Kaon momentum cut: Any event with a kaon whose momentum exceeds 100GeV/c is discarded, in order to improve $K-\pi$ separation.
- Angular acceptance: All four final state particles are required to lie in the detector’s angular acceptance of 300 mrad [col03a].

⁶A ROOT C++ translation of an original CERNLIB package.

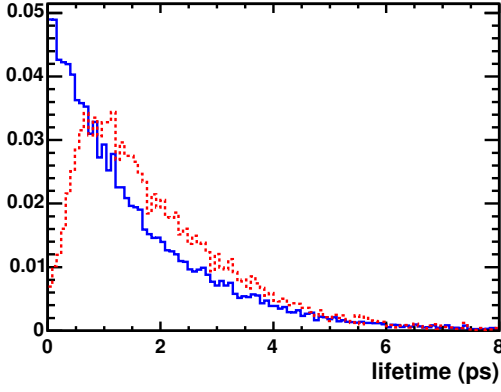


Figure 6.19: The pre-trigger (solid blue) and post-trigger (dashed red) lifetime distributions for the toy Monte Carlo.

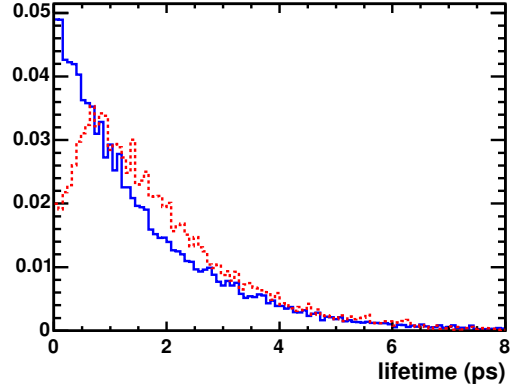


Figure 6.20: The pre-trigger (solid blue) and post-trigger (dashed red) lifetime distributions for the full Monte Carlo.

- **Vertex resolution:** The measured vertex positions are obtained by smearing the generator level vertex positions by the appropriate resolution. In this case, the following resolutions are used: primary vertex X resolution of $0.013 \mu\text{m}$, primary vertex Z resolution of $0.063 \mu\text{m}$, secondary vertex X resolution of $0.011 \mu\text{m}$, secondary vertex Z resolution of $0.090 \mu\text{m}$. For the purposes of this toy Monte Carlo, all secondary vertices are taken to have the same resolution.

Figures 6.20 and 6.19 show the lifetime distributions for signal events in the full and toy Monte Carlo datasets respectively. The toy Monte Carlo models the correct behavior in this variable, and the same is seen across other parameters.

6.7.3 Toy Monte Carlo fit

The fit in Figure 6.21 was created using 60k signal events. The fit result is:

$$\text{lifetime} = 1.528 \pm 0.007 \text{ ps}$$

$$\frac{\chi^2}{ndof} = 0.91$$

$$\chi^2 \text{ probability} = 71.7\%$$

The reason the fit was performed with 60k events and not a full one year sample of 1340k toy Monte Carlo events is the time required to generate and fit such a number of events. Comparing the statistical precision of this fit to the 0.017 ps precision of the full Monte Carlo fit in Figure 6.16, it is seen to scale as expected with the square root of the number of events. Hence the yearly statistical precision is expected to be approximately 0.0014 ps.

Figure 6.22 shows a pull plot obtained by generating 323 samples of 4900 signal events each, corresponding to approximately one day of running, and fitting to them. The plot is unbiased with a width of 1 (within error).

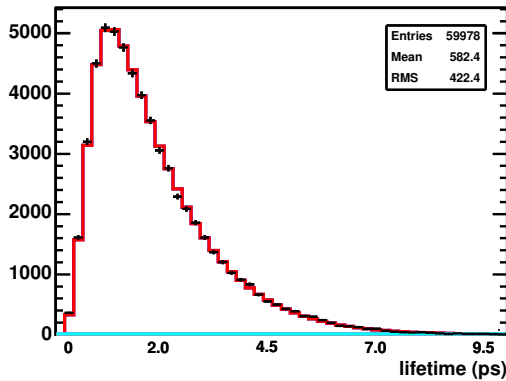


Figure 6.21: Lifetime fit to 60,000 toy Monte Carlo signal events.

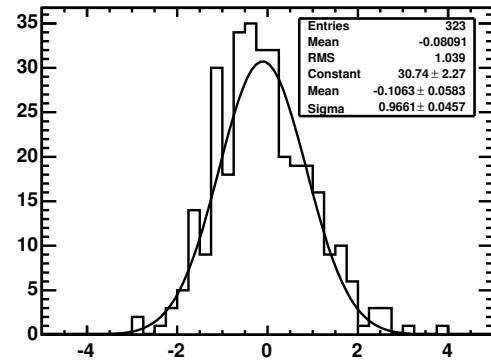


Figure 6.22: Pull plot of 323 Toy MC lifetime fits of 4.9k signal events each.

6.8 Including background

A sample of background events obtained from $b\bar{b}$ inclusive tapes is used in order to estimate the degradation in the statistical precision, as well as potential systematic errors, in the presence of background. However, in the full MC sample, only around 25 background events pass both the L1 trigger and the offline cuts, as opposed to $\sim 10,000$ signal events. Since a $\frac{S}{B}$ of around 5 is expected under the signal peak, for the optimal $\frac{S}{\sqrt{S+B}}$ selection developed in Chapter 3, more background is required in the Monte Carlo sample in order to be able to produce realistic fits.

To obtain a bigger sample, the offline cuts used in the selection have been loosened. The background after the full offline cuts have been applied is dominated by events with a misidentified pion or kaon and other partially reconstructed B decays, while the background after the trigger is dominated by combinatorics. It is therefore likely that the background used in this study will not be representative of the background encountered in real data, and that the major difference between these backgrounds will be qualitative rather than quantitative.

The background from partially reconstructed decays and mis-identification is especially dangerous because the fitter assumes the lifetime behaviour of the background is the same in the mass sidebands and under the peak. Clearly a background from real B mesons lying under the peak makes this assumption problematic. Any such background in the sidebands is also problematic, and it is expected to be found in the left sideband because partially reconstructing a B decay or mis-identifying one of its daughters will systematically decrease the mass of the reconstructed B . For this reason, the lower mass sideband is not used in the fit. As explained in Section 6.4, the background is described by a sum of exponentials, as well as a prompt component.

6.8.1 Full Monte Carlo fit

Figure 6.23 shows a fit to the same signal sample as in Figure 6.16, with approximately 2300 background events added, in the mass region from 5229 MeV to 5429 MeV. The background level corresponds to $\frac{S}{B} = 7$ within a 3σ window around the B mass⁷, where the quantity of background used is constrained by the amount of reconstructed background available. The fit result is:

$$\begin{aligned}\text{lifetime} &= 1.549 \pm 0.019 \text{ ps}, \\ \frac{\chi^2}{ndof} &= 1.49, \\ \chi^2 \text{ probability} &= 0.6\%,\end{aligned}$$

The different parts of the fit are labelled on Figure 6.23, while Figure 6.24 shows the same fit with a log scale on the axis. The background is modelled using two exponential components, as well as a prompt background component, although it can be seen that the fitter essentially prefers to fit the background with a single exponential. The poor $\frac{\chi^2}{ndof}$ of the fit indicates that this background model does not describe the background very well. Because the fitter prefers to use only one exponential component anyhow, adding more exponentials to the fit does not help. On the other hand, the fitted value is in good agreement with the input value. A pull study is required to determine whether the shift relative to the signal-only fit is systematic or statistical, which can only be performed with a toy Monte Carlo.

Figure 6.25 shows the fit to the B mass distribution which is performed simultaneously with the lifetime fit. The signal area was fitted using a single Gaussian, and the background by a straight line.

6.8.2 Toy Monte Carlo pulls

Although a toy Monte Carlo was developed for the background in a similar way as for the signal, it was decided not to use it for the pull study since it would have taken too long to fit a reasonable number of full one-year experiments. Additionally, studying biases in fits to small samples is interesting in its own right: one day of nominal LHCb running in the channel $B_d^0 \rightarrow D^- \pi^+$ corresponds to 4900 signal events, and such a sample would allow a precision of approximately 0.026 ps to be achieved. In particular, the decay $B_s^0 \rightarrow D_s \pi$ has the same topology as $B_d^0 \rightarrow D^- \pi^+$, and is hence expected to have equivalent statistical performance in lifetime fitting. Since this decay has a yearly yield of 140k events, 4900 events correspond to roughly two weeks of running time, and the precision of 0.026 ps is comparable to the current world average precision on the B_s^0 lifetime of 0.031 ps [ea06c]. Hence the measurement of the B_s^0 lifetime from this channel is an excellent candidate for an early LHCb analysis.

⁷These fits were made at a time when the signal to background ratio of the selection in this channel was estimated as 7, and could not be redone with a signal to background ratio of 5 for time reasons. This discrepancy is not expected to significantly affect the final results.

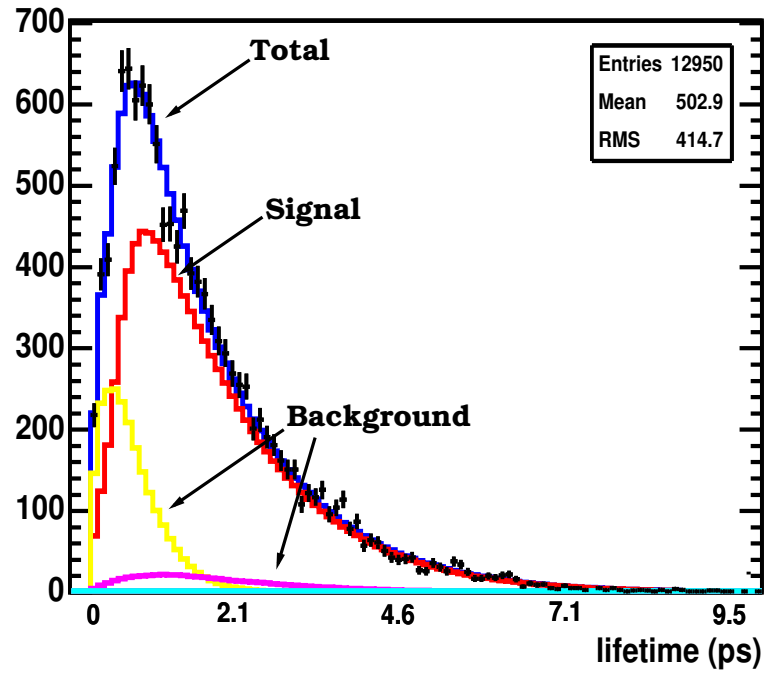


Figure 6.23: Lifetimes fit to signal and background full Monte Carlo events.

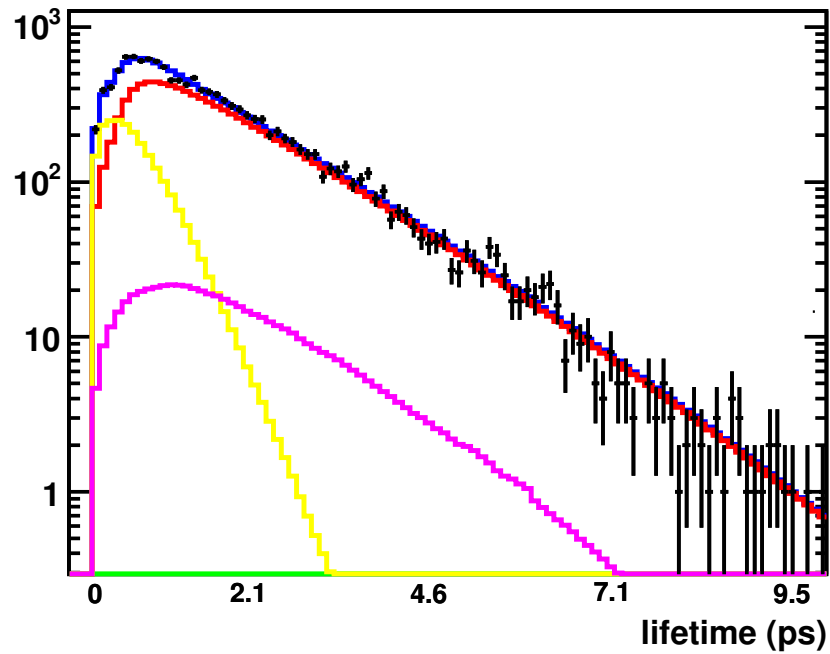


Figure 6.24: Lifetimes fit to signal and background full Monte Carlo events, with a log scale.

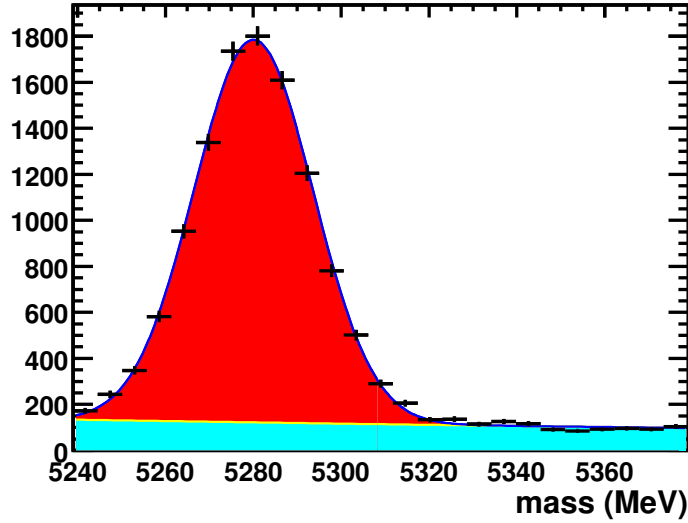


Figure 6.25: The mass fit for signal and background events.

For these reasons, it was decided to make the pull plots with 4900 signal events and 3240 background events in a mass window of 5245 MeV to 5629 MeV, corresponding to a S/B ratio of ~ 6 under the peak. The fits were performed with three exponential backgrounds each, and 10 bins allocated to the Fisher's discriminant. The pull distribution is shown in Figure 6.26, while Figures 6.27 and 6.28 present the fitted lifetimes, and the errors on these lifetimes. The pulls were calculated with an adjustment for the fact that the fitted error scales with the magnitude of the fitted lifetime:

$$\text{Pull} = \frac{t_{\text{fitted}} - t_{\text{generator}}}{\sigma_t \cdot \frac{t_{\text{generator}}}{t_{\text{fitted}}}}. \quad (6.25)$$

Out of the 340 fits submitted, 296 converged successfully, and the returned error was largely Gaussian. The fits which did not converge generally returned a sensible value for the mean lifetime, but failed to return a sensible error because the error matrix was not positive definite. The cause of this, as well as the tail on the right hand side of the error distribution in Figure 6.28, are not understood at present. The average fitted lifetime was 1.539 ± 0.026 ps, compared to the input lifetime of 1.534 ps, while the pull plot has a mean of 0.13 ± 0.06 and a width of 0.93 ± 0.05 . Although there is no clear evidence of a significant bias, it is worth considering the case where the non-zero mean reflects a real bias in the lifetime fit. Assuming a biased mean of 0.1σ after 4900 events, the systematic error would only saturate the statistical error after 490,000 events. This corresponds to over three years of running in the case of $B_s^0 \rightarrow D_s \pi$, and approximately four months of running in the case of $B_d^0 \rightarrow D^- \pi^+$. It also corresponds to a statistical error of approximately 0.0026 ps, around ten times better than the current B_s^0 world average, and around three times better than the current B_d^0 world average. Such a systematic bias should not, therefore, prevent a significant measurement of the B_d^0 and B_s^0 lifetimes with this method.

It should be noted once again that most of the background events used in this pull study were combinatoric background events, because of the small amount of full Monte Carlo

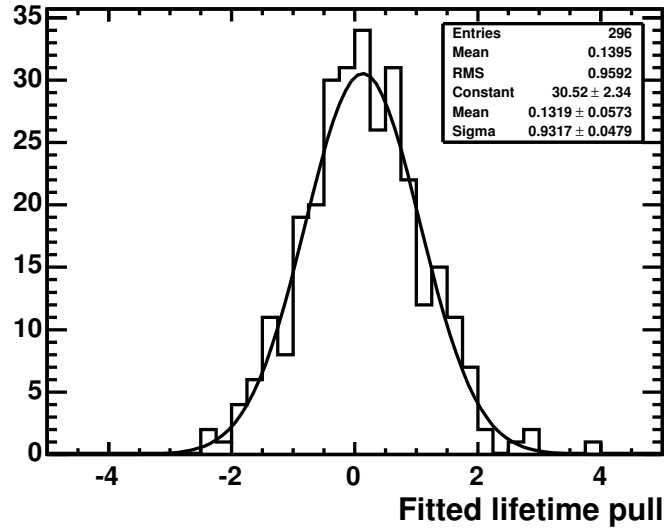


Figure 6.26: A pull plot for 296 one day experiments, with 4900 signal (toy MC) and 3240 background (full MC) events each

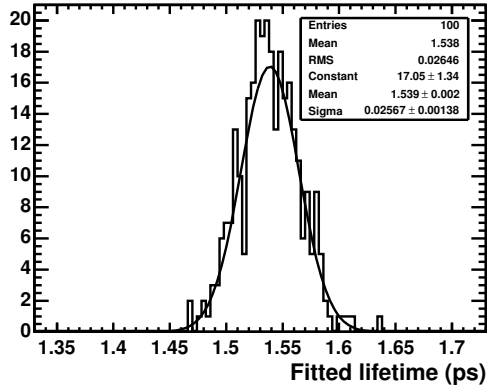


Figure 6.27: The fitted lifetimes of 296 one day experiments

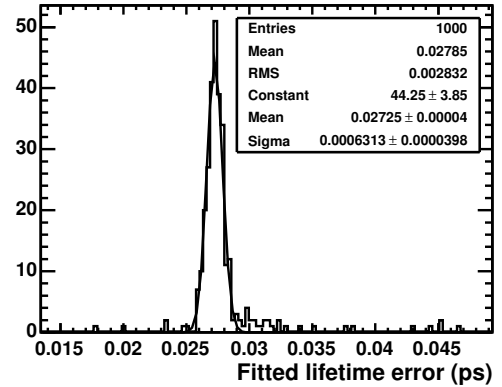


Figure 6.28: The errors on the fitted lifetimes of 296 one day experiments

background available for study. In particular, the background noted in Chapter 3 from misidentified B_s^0 and Λ_b decays could cause particular problems for the fitter, since it peaks near the signal region and will be difficult to estimate from the sidebands. If the mistag in this channel is well known, it will be possible to cross-check the level of such background from the CP asymmetries, as shown in Chapter 4. Set against these difficulties is the fact that the single greatest source of background in this channel comes from misidentified decays of the type $B_d^0 \rightarrow D^- K^+$, which will have the same lifetime profile as the signal and will therefore be as good as signal events from the point of view of a lifetime fit.

Table 6.2: A summary of the fit results presented in this study. A central lifetime value of 1.534ps was used for all fits in this table. FMC refers to full Monte Carlo events, TMC refers to toy Monte Carlo events.

Fit data	Fit result (ps)	Fit σ (ps)	χ^2 of fit	χ^2 prob.
10,571 signal (FMC)	1.532	0.017	1.35	2.9%
10,571 signal and 2300 background (FMC)	1.549	0.019	1.49	0.6%
60,000 signal (TMC)	1.528	0.007	0.91	71.7%

6.9 Expected statistical precision

A summary of the fit results in this study is shown in Table 6.2. The decay $B_d^0 \rightarrow D^- \pi^+$ has an expected yearly yield of 1340k fully triggered events, with a $\frac{S}{B}$ ratio of ~ 4.6 . Fitting the lifetime with 60k toy Monte Carlo signal events achieves a statistical precision of 0.007 ps, while fitting to 60k signal and 15k background events achieves a precision of 0.009ps. A similar result is seen in data generated with the full LHCb detector simulation. For 10,571 signal events the fit result is 1.532 ± 0.017 for an input value of 1.534 ps. Adding 2300 background events in a mass window of -50MeV, +150MeV ($\frac{S}{B} \sim 7$ within ± 3 sigma of the mass peak) a result of 1.549 ± 0.019 ps is obtained. It is seen that the presence of background does not lead to a major deterioration in performance. Moreover, the pull study performed with currently available signal and background Monte Carlo shows no significant bias due to the presence of background in the fits. Therefore, a statistical precision comparable with the current world average of 0.009ps [ea06c] can be expected within the first nominal year of running, corresponding to 2fb^{-1} of data.

In the B_s^0 sector, the decay $B_s^0 \rightarrow D_s \pi$ has the same topology and is hence expected to have equivalent statistical performance. A yearly yield of 140k $B_s^0 \rightarrow D_s \pi$ events and a $\frac{S}{B}$ of ~ 5 , as estimated in [GHVH⁺03], allows us to achieve a statistical precision comparable with the current world average of 0.031ps [ea06c] within the first couple of weeks of running at the LHCb design luminosity of $2 \times 10^{32} \text{cm}^{-2} \text{s}^{-1}$.

Chapter 7

Conclusion

With CP violation already established in the B_d^0 and B_s^0 systems, including measurements of ϕ_d and ϕ_s as well as the B_s mass difference Δm_s , LHCb's physics programme will concentrate heavily on the measurement of the CKM angle γ , the most significant part of the Standard Model CP violation picture yet to be constrained.

The decay mode $B_d^0 \rightarrow D^- \pi^+$ is well matched to LHCb's strengths, since the final state consists entirely of charged pions and kaons, allowing the excellent particle identification provided by the RICH detectors to come to the fore. The innovative aspects of the reconstruction developed in Chapter 3 are the mechanism for the categorization of background, and the fast genetic algorithm for optimizing the offline selection cuts. A detailed study of the background in this decay mode has shown that most background events surviving the optimized offline selection will consist of misidentified decays such as $B_d^0 \rightarrow D^- K^+$, and has highlighted the need to make more extensive use of the RICH particle identification in future studies of this channel.

The decay modes $B_d^0 \rightarrow D^- \pi^+$ and $B_s^0 \rightarrow D_s^- K^+$ offer the possibility of a theoretically clean measurement of γ which is not sensitive to new physics effects, and can be used as a benchmark measurement of γ for modes which **are** sensitive to new physics effects, such as $B_s^0 \rightarrow K^+ K^-$. Considered separately, either mode offers the potential of a one-year precision on γ of a few degrees, but the value of γ extracted in this fashion will suffer from an eightfold ambiguity. There is some hope of resolving ambiguities in the $B_s^0 \rightarrow D_s^- K^+$ analysis through the use of untagged events, but this relies on there existing a significant lifetime difference between the B_s^0 and $\overline{B}_s^0$ states, the magnitude of which is currently unclear. Meanwhile, the $B_d^0 \rightarrow D^- \pi^+$ analysis also suffers from theoretical uncertainties involved in estimating the strength of the interference $|\eta|$ between the different tree-level diagrams, which will be too small to fit from the CP asymmetries themselves. Since these decay modes are closely related through U-spin symmetry, it is natural to attempt a combined analysis, which can in principle lead to an unambiguous measurement of γ .

It was seen in Chapter 5 that this combined U-spin analysis can lead to a precision of $O(10^\circ)$ on γ after one year of data taking. However, this estimate is only realistic in the case when the strong phases in these decays are large. The likely performance degrades in the factorization limit, however, it may be possible to obtain additional information from

the related U-spin decay pairs:

- $B_d^0 \rightarrow D^{*-}\pi^+$ and $B_s^0 \rightarrow D_s^{*-}K^+$;
- $B_d^0 \rightarrow D^-\rho^+$ and $B_s^0 \rightarrow D_s^-K^{*+}$;

which would allow an unambiguous precision measurement of γ irrespective of the values of the strong phases in these decay channels.

In the case that this U-spin extraction of γ does not prove competitive with other tree-level γ measurements such as those from $B^+ \rightarrow D^0 K^+$, it can still be used as a probe of U-spin symmetry, and therefore constrain QCD models. Another probe of QCD models such as HQE will be provided by the precision measurement of B meson lifetimes at LHCb, which will exploit the unprecedented statistical power available, particularly in the B_s sector. In order to fully exploit LHCb's potential for measuring these lifetimes, it will be necessary to account for the lifetime biases introduced by LHCb's trigger. A method, originally developed at CDF, which measures this bias with no recourse to Monte Carlo information has been studied in the LHCb context. It has been shown that the channel $B_s^0 \rightarrow D_s^-\pi^+$ **alone** can overtake the present world average on the B_s^0 lifetime within the first **few weeks** of LHCb data taking, confirming the potential of these lifetime measurements to further constrain theoretical models such as HQE. The integration of this lifetime fitting package into the LHCb trigger is currently underway, and it is expected to provide a kind of “self monitoring” for the trigger once the experiment starts, by enabling the trigger biases to be measured in the course of normal operation.

Appendix A

Current direct constraints on CKM matrix elements

Unless otherwise stated, the below constraints were taken from [ea06c].

- $|V_{ud}|$ has been measured using superallowed $0^+ \rightarrow 0^+$ nuclear beta decays. The current world average is

$$|V_{ud}| = 0.97377 \pm 0.00027. \quad (\text{A.1})$$

- $|V_{us}|$ is measured via semileptonic kaon decay rates, and the current world average is

$$|V_{us}| = 0.2257 \pm 0.0021. \quad (\text{A.2})$$

- $|V_{cd}|$ is determined from neutrino and antineutrino interactions, specifically the ratio of double-muon to single-muon production in neutrino-antineutrino collisions. The current world average is

$$|V_{cd}| = 0.230 \pm 0.011. \quad (\text{A.3})$$

- $|V_{cs}|$ can be directly measured from semileptonic D decays, with theoretical uncertainties associated with lattice QCD calculations of the semileptonic form factors. The current world average is

$$|V_{cs}| = 0.957 \pm 0.017 \pm 0.093. \quad (\text{A.4})$$

- $|V_{cb}|$ can be determined from inclusive or exclusive semileptonic B decays, with the dominant constraint currently coming from the inclusive measurements. The current world average is

$$|V_{cb}| = 0.0416 \pm 0.0006. \quad (\text{A.5})$$

- $|V_{ub}|$ is also determined from inclusive or exclusive semileptonic B decays; the former suffers uncertainties from charm background such as $B \rightarrow X_c l \bar{\nu}$, the latter from the theoretical uncertainties associated with lattice QCD calculations of the semileptonic form factors. The current world average, dominated by the inclusive measurements, is

$$|V_{ub}| = 0.00431 \pm 0.00030. \quad (\text{A.6})$$

- $|V_{td}|$ is determined from the mass difference of the neutral B meson eigenstates under the assumption $|V_{tb}| = 1$. The current world average is

$$|V_{td}| = 0.0074 \pm 0.0008. \quad (\text{A.7})$$

- $|V_{ts}|$ is determined from $|V_{td}|$ and the ratio $\Delta m_d / \Delta m_s$. The best constraint at present comes from a CDF measurement of

$$|V_{td}|/|V_{ts}| = 0.208^{+0.008}_{-0.006}. \quad (\text{A.8})$$

- $|V_{tb}|$ is determined from the ratios of top decay rates. The current 95% CL lower limit is

$$|V_{tb}| > 0.78. \quad (\text{A.9})$$

- The angle α can be directly measured from time dependent CP asymmetries in $b \rightarrow u\bar{u}b$ dominated modes. The single best measurement, $\alpha = 88^\circ \pm 17^\circ$, comes from the $B_d^0 \rightarrow \rho^+ \rho^-$ decay mode, which is believed to have small penguin contributions compared to other modes that measure α . In combination with a Dalitz analysis of $B_d^0 \rightarrow \pi^+ \pi^- \pi^0$, it provides the current world average of

$$\alpha = (99^{+13}_{-8})^\circ. \quad (\text{A.10})$$

- The angle β is the best known angle of the unitarity triangle, having been measured through theoretically clean decays of the type $B_d^0 \rightarrow J/\psi K_{S,L}^0$, with a world average [ea06a] of

$$\sin(2\beta) = 0.678 \pm 0.025. \quad (\text{A.11})$$

- The angle γ has recently been constrained by BaBar and BELLE [botBc07]. The constraints are:

$$\gamma = (92 \pm 41 \pm 11 \pm 12)^\circ \text{ (BaBar)} \quad (\text{A.12})$$

$$\gamma = (53^{+15}_{-18} \pm 3 \pm 9)^\circ \text{ (Belle)} \quad (\text{A.13})$$

the BaBar constraint is obtained through a Dalitz analysis of $B^- \rightarrow DK^-$ and $B^- \rightarrow D^* K^-$, while the Belle constraint is obtained from a Dalitz analysis of $B^- \rightarrow DK^-$, $B^- \rightarrow D^* K^-$, and $B^- \rightarrow DK^{*-}$.

The above constraints and measurements can be used to check the unitarity of the CKM matrix and the unitarity triangle, and result in the following constraints

$$|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 0.9992 \pm 0.0011, \quad ,$$

$$|V_{cd}|^2 + |V_{cs}|^2 + |V_{cb}|^2 = 0.968 \pm 0.181, \quad ,$$

$$|V_{ud}|^2 + |V_{cd}|^2 + |V_{td}|^2 = 1.001 \pm 0.005, \quad ,$$

$$\alpha + \beta + \gamma = (184^{+20}_{-15})^\circ, \quad ,$$

all of which are consistent with unitarity.

Appendix B

Background category decision tree

The decision tree of the Background Category algorithm is shown overleaf. It is designed to deliver an unambiguous category for a given event, although there are events which could fall into several categories: for example, if the particles used to reconstruct the candidate B meson come partly from the primary vertex and partly from the decay of a D meson, the background could be classified as either primary vertex background or combinatoric background. In general, non-physics background take precedence over physics backgrounds, i.e. the algorithm begins by testing for the most egregious mistakes in the reconstruction (Do all the tracks used correspond to a physical particle? Do they all come from the same particle? etc.) and an event is classified as signal only if it does not fail any of the tests. The category priority list among non-physics backgrounds is:

1. Ghost background;
2. Primary vertex backgrounds;
3. $b\bar{b}$ combinatorics;
4. $c\bar{c}$ combinatorics;
5. Other combinatorics.

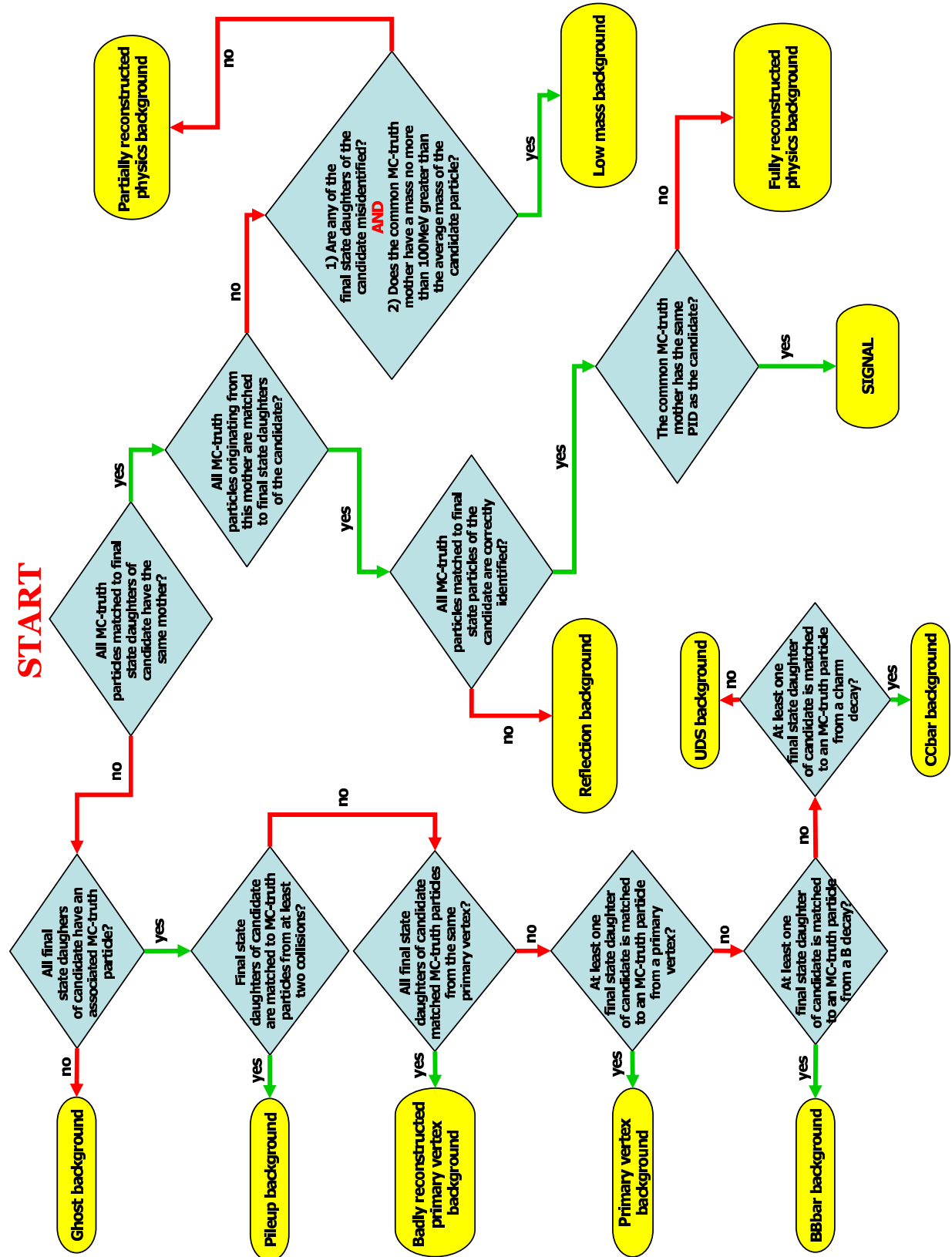


Figure B.1: The background Category algorithm decision tree.

Appendix C

Additional U-spin analysis scenarios

$$\gamma = 60^\circ \text{ and } \delta_{s,d} = 30^\circ$$

This scenario is interesting because it illustrates the lower limit (with respect to the size of the strong phases) at which Equation 5.25 gives competitive statistical precisions after five years. The contour plots are shown in Figure C.1 for one year of data taking, and in Figure C.2 for five years of data taking.

After one year of data taking, it is already possible to disfavour all ambiguous solutions by combining the strong and phase U-spin assumptions. The strong U-spin assumption of Equation 5.24 gives a statistical precision of $-20^\circ, +30^\circ$, with a systematic error of $-10^\circ, +15^\circ$ associated with a U-spin breaking level of 30%. The phase U-spin assumption leads to an upper error of $+11^\circ$ on γ , with an upper systematic error of $+3^\circ$ associated with a U-spin breaking level of 30%.

After five years of data taking, all ambiguous solutions are clearly excluded by the phase U-spin assumption. This assumption leads to a statistical precision of $-13^\circ, +6^\circ$ on γ , with a systematic error of $-1^\circ, +4^\circ$ associated with a U-spin breaking level of 30%.

$$\gamma = 60^\circ \text{ and } \delta_{s,d} = 85^\circ$$

The γ precision plots in Chapter 4 showed that the conventional extraction of γ suffers from very high errors for a γ around the currently favoured value of 60° when the strong phases are large, quite apart from the problem of ambiguity. It is therefore especially interesting to see that this is the area in which the combined extraction of γ performs best. The contour plots are shown in Figure C.3 for one year of data taking, and in Figure C.4 for five years of data taking.

After one year, the strong U-spin assumption of Equation 5.25 can be used to exclude all ambiguous solutions. Its statistical precision is $-16^\circ, +8^\circ$, while a U-spin breaking level of 30% introduces a systematic error of $-0^\circ, +3^\circ$. As a good example of a complementary measurement, the amplitude U-spin assumption can be used to give a lower statistical pre-

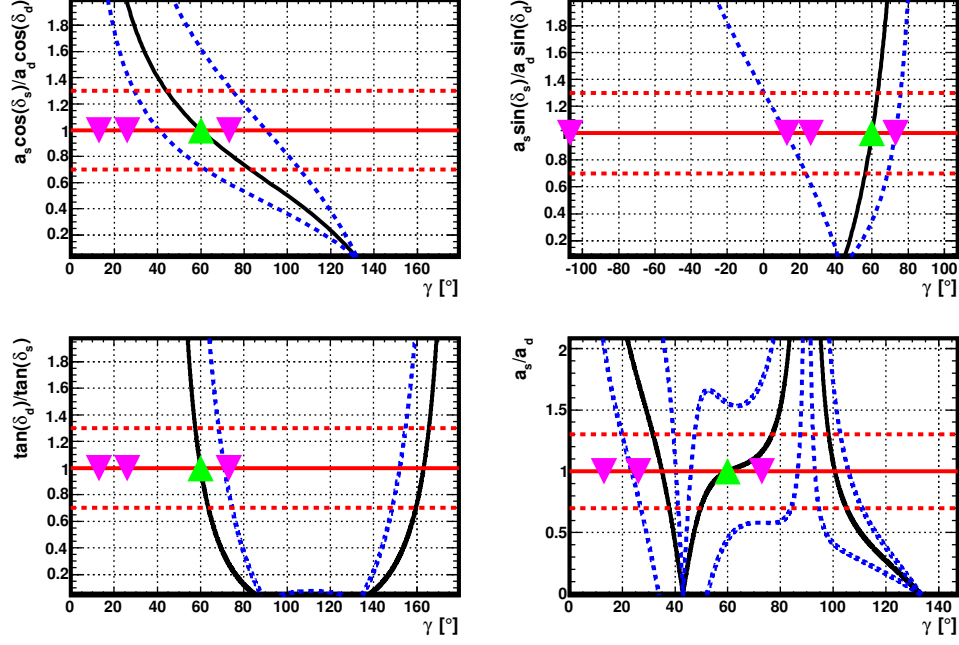


Figure C.1: Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 30^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

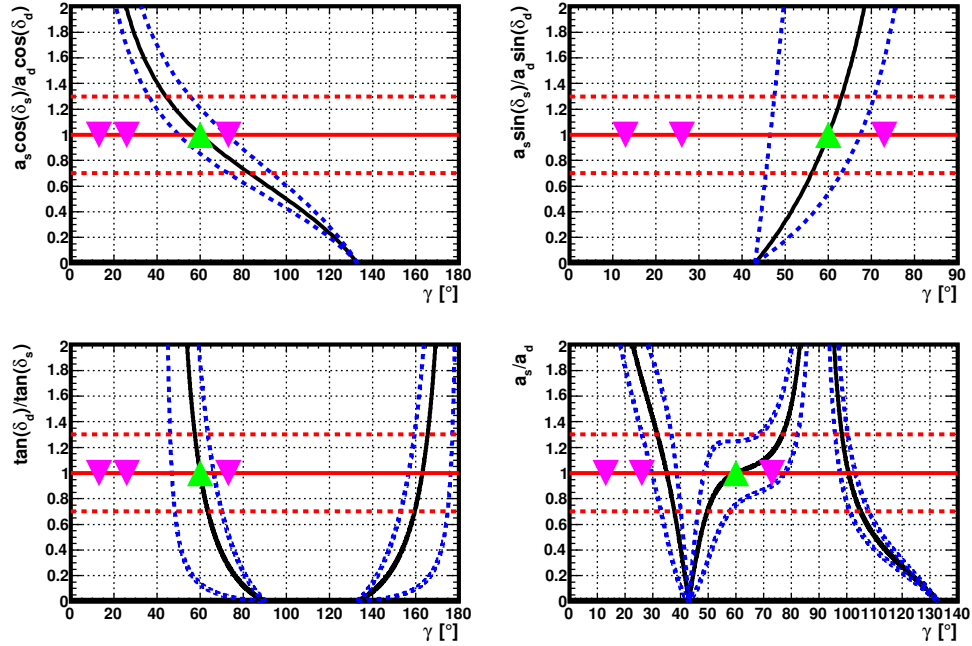


Figure C.2: Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 30^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

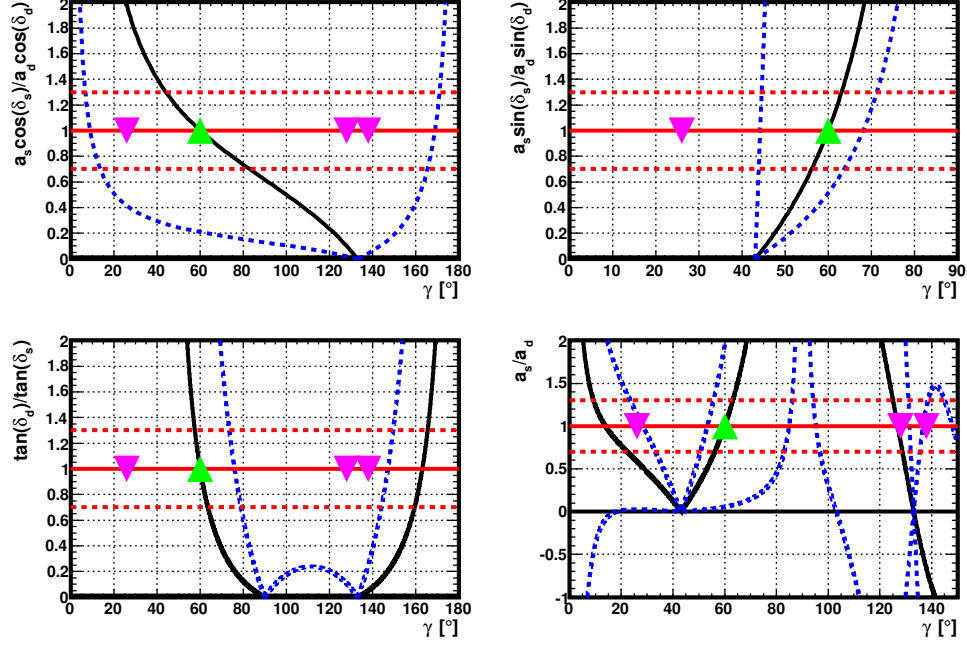


Figure C.3: Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 85^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

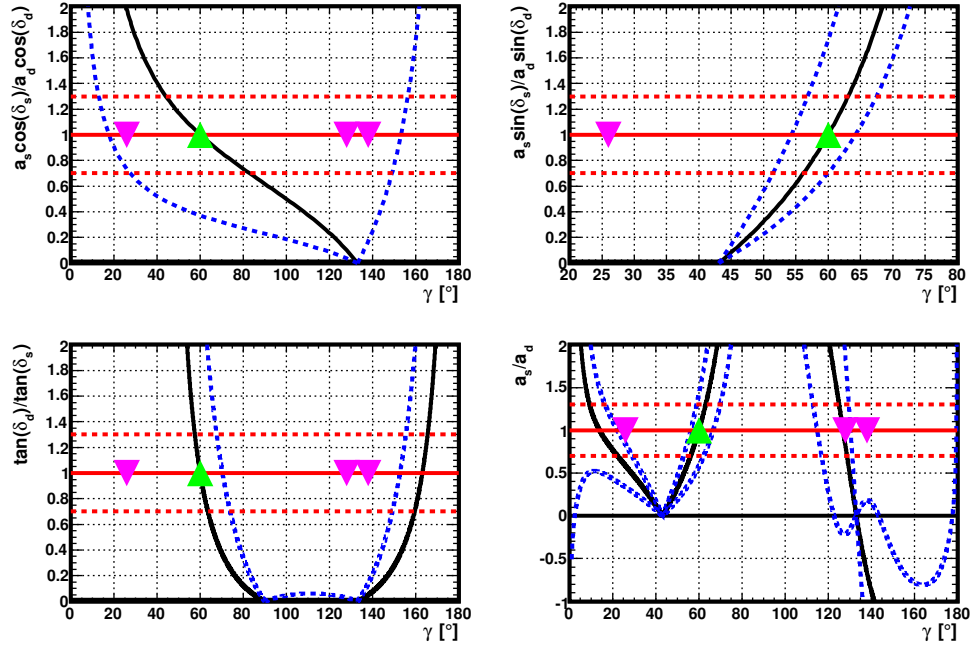


Figure C.4: Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 85^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

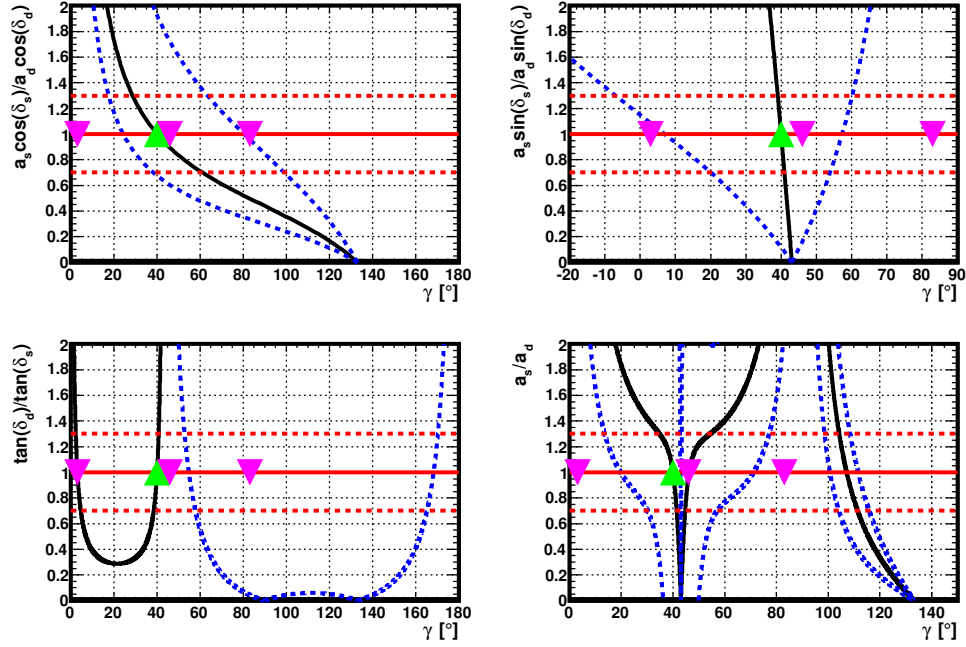


Figure C.5: Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 40^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

cision of -7° , with an associated systematic of -2° for a U-spin breaking level of 30%. The combined statistical precision is $-7^\circ, +8^\circ$, which is rather competitive for a first year measurement. While bearing in mind the caveat that this is a very specific area of parameter space, and the measurement relies on U-spin basically being a good symmetry (while being robust against reasonable levels of breaking), it is heartening to see such performance precisely in those areas of parameter space where the conventional extraction of γ is least useful. After five years, combining these assumptions again results in a statistical precision of $\pm 4^\circ$, with a systematic error of $\pm 3^\circ$ associated with a U-spin breaking level of 30%.

$\gamma = 40^\circ$ and $\delta_{s,d} = 40^\circ$

Until now, γ has been assumed to lie at its current preferred standard model value of 60° . After the last scenario, which demonstrates the potential of the combined extraction in a favourable area of parameter space, it is as well to consider a more difficult combination of parameters. The contour plots are shown in Figure C.5 for one year of data taking, and in Figure C.6 for five years of data taking.

After one year, none of the relations can exclude all ambiguities. The best result is achieved by a combination of the strong and phase U-spin assumptions, which can lead to a statistical precision of $-15^\circ, +14^\circ$, with associated U-spin breaking systematics of $-7^\circ, +4^\circ$. After five years, this combination leads to a statistical precision of $-10^\circ, +8^\circ$, with a twofold ambiguity on γ , and associated U-spin breaking systematics of $-7^\circ, +2^\circ$.

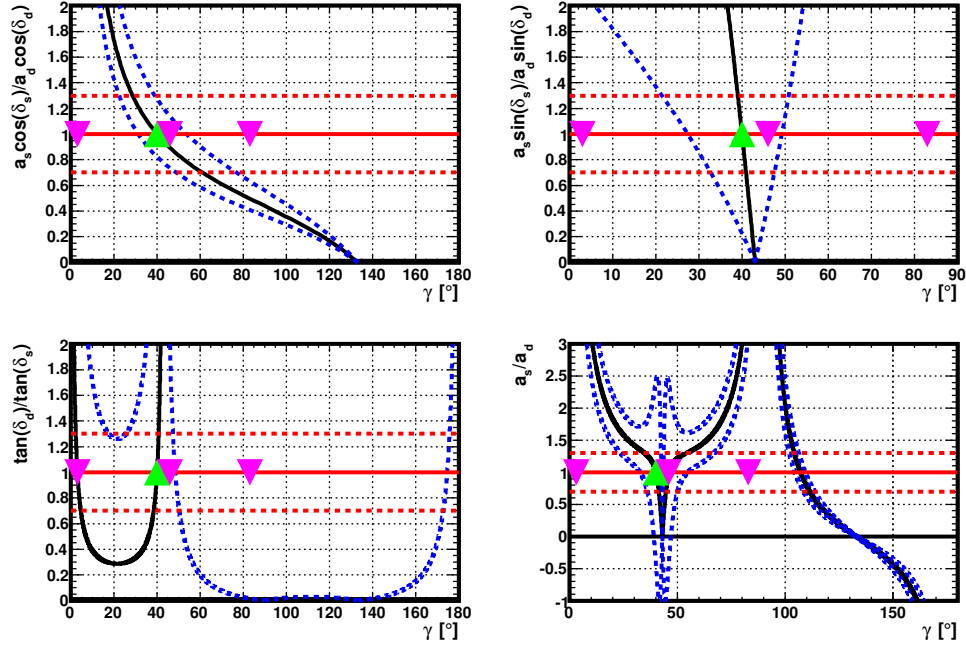


Figure C.6: Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 40^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

$\gamma = 80^\circ$ and $\delta_{s,d} = 40^\circ$

Sometimes, the ambiguous values of γ are almost degenerate, for example in the case where $\gamma = 80^\circ$ and $\delta_{s,d} = 40^\circ$. The contour plots are shown in Figure C.7 for one year of data taking, and in Figure C.8 for five years of data taking.

After one year, the only useful information is provided by the phase U-spin assumption, which affords an upper statistical precision of $+6^\circ$, and an upper systematic error of 1° for a 30% breaking of U-spin symmetry. After five years, even though this assumption can now give a statistical precision of $-14^\circ, +3^\circ$, it does not exclude the solution at $\gamma = 83^\circ$.

$\gamma = 60^\circ$ and $\delta_{s,d} = 140^\circ$

Finally, consider a scenario with strong phases $> 90^\circ$. The contour plots are shown in Figure C.9 for one year of data taking, and in Figure C.10 for five years of data taking.

After one year, combining the amplitude U-spin assumption with the strong U-spin assumption of Equation 5.25 already allows an unambiguous extraction of γ , with a statistical precision of $+11^\circ, -12^\circ$ and associated U-spin breaking systematics of $+3^\circ, -2^\circ$. After five years, all four equations provide useful limits on γ , with the best precision coming from the phase U-spin assumption: the statistical error is $+5^\circ, -9^\circ$, while the systematic uncertainty associated with a 30% breaking of U-spin symmetry is $+4^\circ, -2^\circ$.

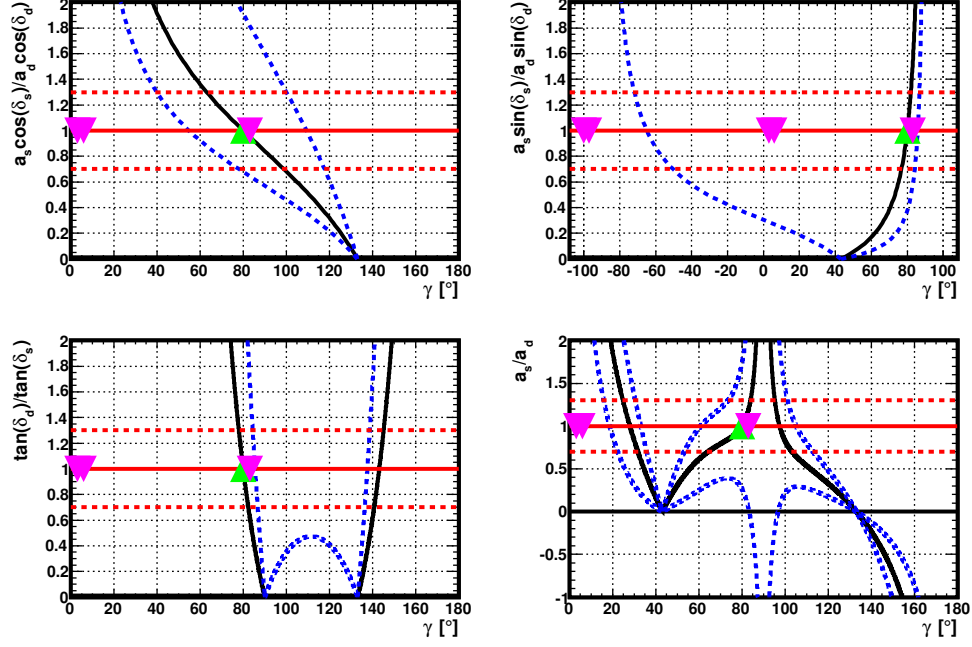


Figure C.7: Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 80^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

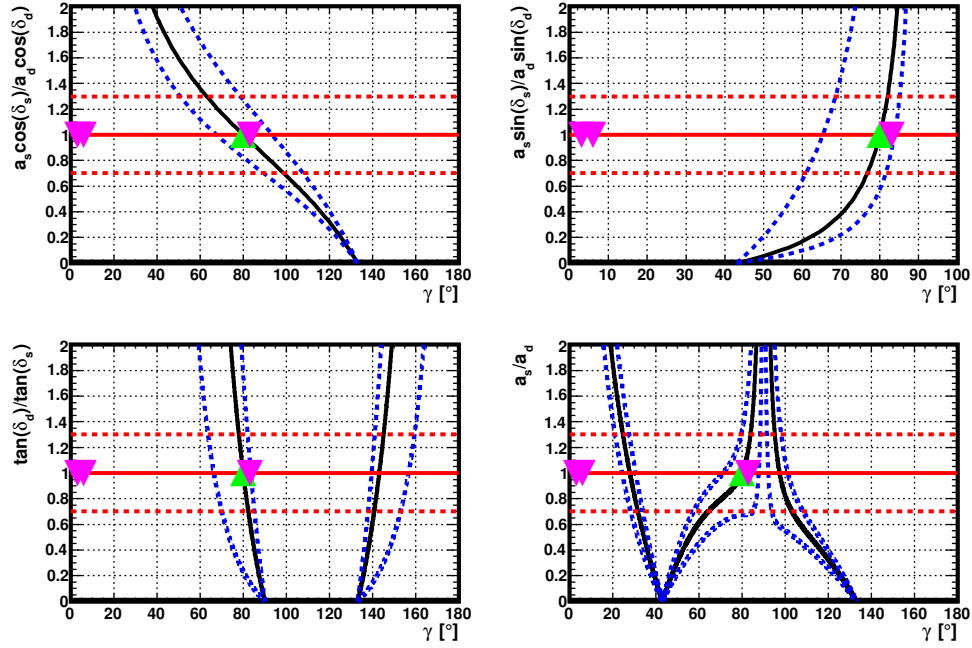


Figure C.8: Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 80^\circ$ and $\delta_{s,d} = 40^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

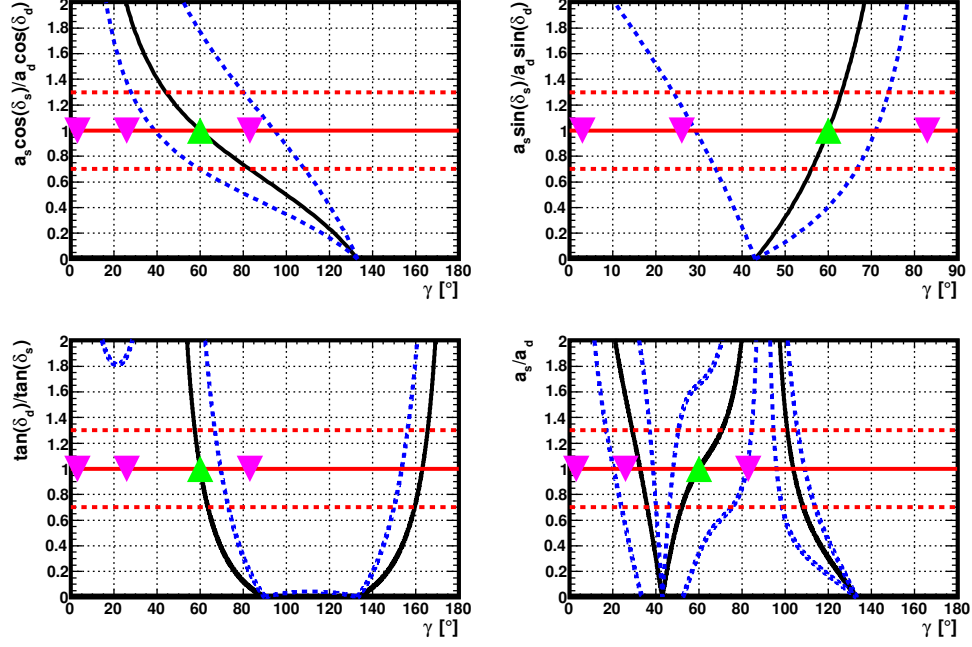


Figure C.9: Contours in γ – U-spin breaking space for one year of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 140^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

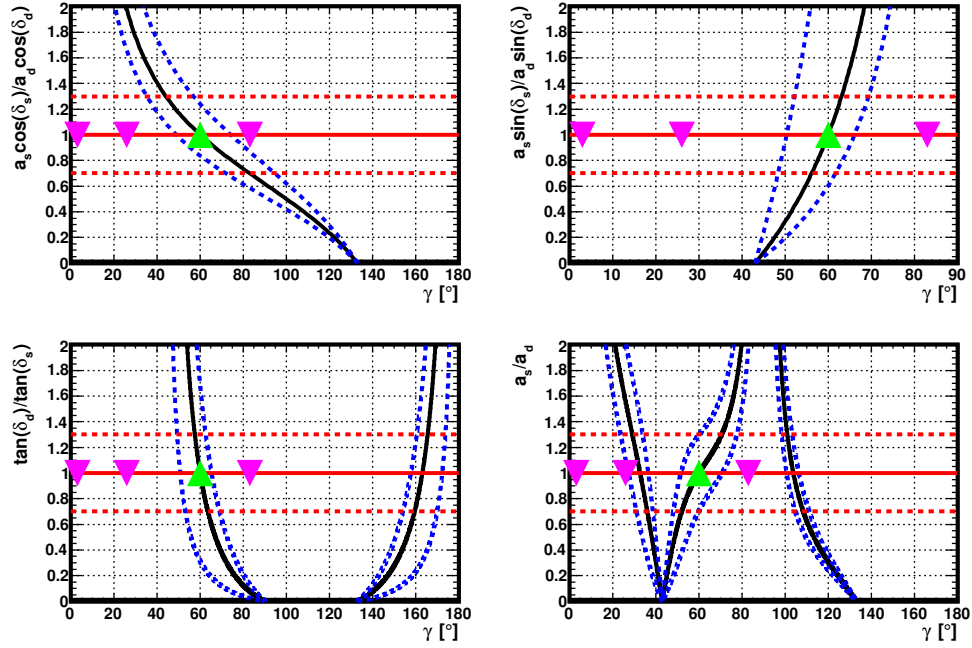


Figure C.10: Contours in γ – U-spin breaking space for five years of data taking, with $\gamma = 60^\circ$ and $\delta_{s,d} = 140^\circ$. The four scenarios shown correspond, clockwise from top left, to Equations 5.24 (strong U-spin), 5.25 (strong U-spin), 5.29 (amplitude U-spin), and 5.28 (phase U-spin).

Appendix D

Derivation of the full lifetime likelihood

In Section 6.4, the probability of measuring an event with a given lifetime, lifetime resolution, B mass, and acceptance function was given. This appendix gives the full derivation of this probability. As a reminder, the probability being calculated is:

$$P(t, \text{acceptance}) = P(t_m, \sigma_t, m, \text{acc} | \text{trigger}), \quad (\text{D.1})$$

which can be broken up into signal and background parts:

$$P(t_m, \sigma_t, m, \text{acc} | \text{trigger}) = P(s, t_m, \sigma_t, m, \text{acc} | \text{trigger}) + P(b, t_m, \sigma_t, m, \text{acc} | \text{trigger}). \quad (\text{D.2})$$

The probability that the event passed the trigger, $P(\text{trigger})$ is different from the probability of finding a given acceptance function $P(\text{acc})$, because it represents a summation over all possible acceptance functions as evaluated at the measured lifetime of the event, t_m . In other words, to find out the probability of triggering, it is necessary to multiply the probability of finding a given acceptance function by the value of the acceptance function at the measured lifetime (1 or 0, as always), and then sum over all acceptance functions. This is important because unlike $P(\text{acc})$, $P(\text{trigger})$ – and by extension, $P(\text{trigger} | \text{acc})$ – depends on the mean lifetime of the B .

What follows will deal with the signal term – the background term can be retrieved by swapping s for b in the subsequent equations. The goal of this derivation is simply to reduce the probability of finding a given triggered event into a form which can be readily calculated from the measured quantities in the event. Using Bayes' theorem:

$$P(s, t_m, \sigma_t, m, \text{acc} | \text{trigger}) = \frac{P(s, t_m, \sigma_t, m, \text{acc}) P(\text{trigger} | s, t_m, \sigma_t, m, \text{acc})}{P(\text{trigger})}, \quad (\text{D.3})$$

the term $P(\text{trigger} | s, t_m, \sigma_t, m, \text{acc})$ is just the probability of triggering, given the event, and is always 1 or 0, since the trigger either selects the event or it does not. It will be abbreviated to $P(\text{trigger} | \text{event})$ from now on. The term $P(\text{trigger})$ needs to be rewritten, and it is convenient to rewrite it in terms of the probability of triggering given an acceptance function:

$$P(\text{trigger}) = P(\text{trigger} | \text{acc}, \sigma_t, s) \frac{P(\text{acc}, \sigma_t, s)}{P(\text{acc}, \sigma_t, s | \text{trigger})}. \quad (\text{D.4})$$

$P(\text{trigger}|\text{acc}, \sigma_t, s)$ is the probability of triggering given an acceptance function, an error on the lifetime, and the event being signal. The error on the lifetime is included here in order to allow the use of event-by-event lifetime errors in the future – as stated above, an average lifetime error is used in this study. Rewriting:

$$P(s, t_m, \sigma_t, m, \text{acc}) = P(\text{acc}, \sigma_t, s) P(t_m, m|\text{acc}, \sigma_t, s), \quad (\text{D.5})$$

allows the term $P(\text{acc}, \sigma_t, s)$ to be cancelled out:

$$P(s, t_m, \sigma_t, m, \text{acc}|\text{trigger}) = \frac{P(t_m, m|\text{acc}, \sigma_t, s) P(\text{acc}, \sigma_t, s|\text{trigger}) P(\text{trigger}|\text{event})}{P(\text{trigger}|\text{acc}, \sigma_t, s)}. \quad (\text{D.6})$$

Taking stock, it can be seen that the denominator represents the probability of an event triggering given the kinematics and error on the measured lifetime. It can be calculated for any event by integrating over $\text{acc}(t) P(t|\tau)$, for a given mean lifetime τ . The term $P(\text{trigger}|\text{event})$ is 1 by definition for all events, since only events that pass the trigger are available for analysis. Rewriting the first factor in the numerator:

$$P(t_m, m|\text{acc}, \sigma_t, s) = P(t_m|\text{acc}, \sigma_t, s) P(m|t_m, \text{acc}, \sigma_t, s). \quad (\text{D.7})$$

Further simplification requires the following assumptions:

1. The reconstructed mass, m , only depends on whether the event is signal or background, and is independent of the decay kinematics, measured lifetime, and the error on the measured lifetime.
2. The measured lifetime, t_m , is independent of the decay kinematics. This is a sensible assumption since the measured lifetime is the difference between the reconstructed decay vertex and reconstructed primary vertex positions, and those will not depend on the momenta or lifetimes of the B daughters.

Rewriting D.7 once again:

$$P(t_m, m|\text{acc}, \sigma_t, s) = P(t_m|\sigma_t, s) P(m|s). \quad (\text{D.8})$$

The first term, $P(t_m|\sigma_t, s)$, is the probability of measuring a lifetime t_m given a signal event of mean lifetime τ and a lifetime error, and can be readily calculated as in Section 6.3.3. $P(m|s)$ is the probability of measuring a mass given a signal event, and is given by the usual Gaussian distribution. All that remains is to rewrite $P(\text{acc}, \sigma_t, s|\text{trigger})$ in a form convenient for subsequent calculations:

$$P(\text{acc}, \sigma_t, s|\text{trigger}) = P(\text{acc}, \sigma_t|\text{trigger}) P(s|\text{acc}, \sigma_t, \text{trigger}). \quad (\text{D.9})$$

The first factor is simply the probability of finding a given acceptance function and lifetime error given that the event passed the trigger, and will factor out of the likelihood since it does not depend on the mean lifetime of the B . The second term is the probability that the event is signal, given its acceptance function - it is analogous to the term $P(s|\text{acc})$ in Section

6.4.1, and will be dealt with in Section 6.5. It is possible to write the full probability for signal events now:

$$P(s, t_m, \sigma_t, m, \text{acc}|\text{trigger}) = \frac{P(t_m|\sigma_t, s) P(m|s) P(\text{trigger}|\text{event}) P(s|\text{acc}, \sigma_t, \text{trigger})}{P(\text{trigger}|\text{acc}, \sigma_t, s)}. \quad (\text{D.10})$$

The individual components are given by:

- $P(t_m|\sigma_t, s)$ – the probability of measuring a lifetime t_m for a signal event with mean lifetime τ and lifetime error σ_t is given by:

$$P(t_m|\sigma_t, s) = \frac{1}{\tau} e^{\frac{-t}{\tau} + \frac{\sigma^2}{2\tau^2}} F\left(\frac{t_m}{\sigma} - \frac{\sigma}{\tau}\right). \quad (\text{D.11})$$

- $P(m|s)$ – the mass distribution is given by a Gaussian:

$$P(m|s) = \frac{1}{\sqrt{2\pi}\sigma_m} e^{\frac{-(m-\mu_m)^2}{2\sigma_m^2}}. \quad (\text{D.12})$$

As with lifetime errors, a mean mass error is used for all events.

- $P(\text{trigger}|\text{acc}, \sigma_t, s)$ – this probability is obtained by integrating the probability of finding a measured lifetime over the acceptance function:

$$P(\text{trigger}|\text{acc}, \sigma_t, s) = \sum_{i=\text{all intervals}} \int_{t_{\min i}}^{t_{\max i}} \frac{1}{\tau} e^{\frac{-t}{\tau} + \frac{\sigma^2}{2\tau^2}} F\left(\frac{t_m}{\sigma} - \frac{\sigma}{\tau}\right) dt_m. \quad (\text{D.13})$$

Appendix E

The Fisher discriminant for lifetime fitting

This Appendix derives the calculation of the Fisher discriminant in Chapter 6. The purpose of the Fisher discriminant is to assign a probability to each event of being signal, based on the trigger acceptance function for that event.

E.1 Discriminating between signal and background events

Consider a sample of signal and background events with certain distributions in two variables, x and y , as illustrated in Figure E.1(c). The means of x and y for the distributions, $\overline{m}_s = (\overline{x}_s, \overline{y}_s)$ and $\overline{m}_b = (\overline{x}_b, \overline{y}_b)$, are marked with crosses on the diagrams. In order to discriminate between signal and background, they need to be projected in such a direction that the separation of the means is large, while the spread around each individual mean is small. For a direction of projection w , the square of the separation of the means is given by:

$$(\langle w | \overline{m}_s \rangle - \langle w | \overline{m}_b \rangle)^2 = \langle w | (\overline{m}_s - \overline{m}_b) \rangle \langle (\overline{m}_s - \overline{m}_b) | w \rangle, \quad (\text{E.1})$$

and

$$|(\overline{m}_s - \overline{m}_b) \rangle \langle (\overline{m}_s - \overline{m}_b)| = S_M. \quad (\text{E.2})$$

For a set of signal events $p_i = (x_i, y_i)$, the square of the spread can be similarly defined:

$$\begin{aligned} \text{Spread}_{\text{sig}}^2 &= \sum_{\text{all signal events}} (\langle w | \overline{p}_i \rangle - \langle w | \overline{m}_s \rangle)^2 \\ &= \sum_{\text{all signal events}} \langle w | (\overline{p}_i - \overline{m}_s) \rangle \langle (\overline{p}_i - \overline{m}_s) | w \rangle \\ &= \sum_{\text{all signal events}} \langle w | S_{\text{sig}} | w \rangle, \end{aligned} \quad (\text{E.3})$$

and an analogous expression holds for background events. Hence Fisher's criterion for optimal separation can be defined as the projection axis for which the ratio of the separation of

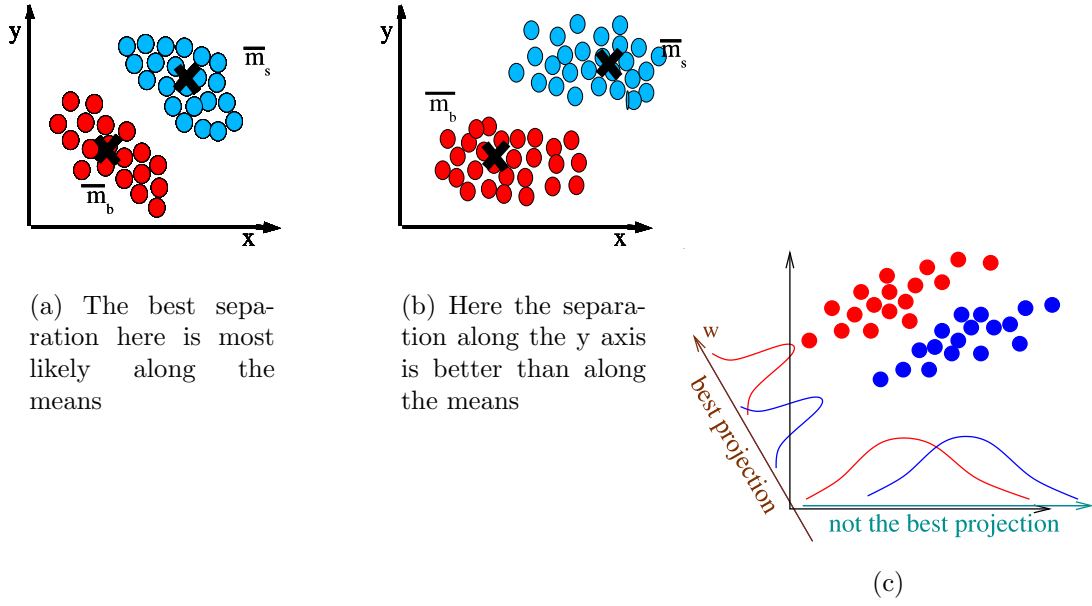


Figure E.1: This illustrates that it is necessary to take the means (marked with a cross) and spread of each variable into account when finding the direction of best separation[ABH⁺03]

the means and the spread around the means is maximized:

$$J(w) = \frac{\langle w | S_M | w \rangle}{\langle w | S_{\text{sig}} + S_{\text{back}} | w \rangle} = \frac{\langle w | S_M | w \rangle}{\langle w | S_w | w \rangle}. \quad (\text{E.4})$$

Differentiating:

$$\nabla_w (J(w)) = \frac{2S_M | w \rangle}{\langle w | S_w | w \rangle} - \frac{\langle w | S_M | w \rangle}{\langle w | S_w | w \rangle} \frac{2S_w | w \rangle}{\langle w | S_w | w \rangle} = 0. \quad (\text{E.5})$$

Since

$$\frac{\langle w | S_M | w \rangle}{\langle w | S_w | w \rangle} = \lambda, \quad (\text{E.6})$$

i.e. it is simply a number, it follows that

$$S_M | w \rangle - \lambda S_w | w \rangle = 0, \quad (\text{E.7})$$

and therefore:

$$S_w^{-1} S_M | w \rangle = \lambda | w \rangle. \quad (\text{E.8})$$

Since all that is required is the direction of the projection axis, there is no need to solve for the specific values of λ . Using Equation E.2:

$$S_M | w \rangle \propto |(\overline{m}_s - \overline{m}_b)\rangle \quad (\text{E.9})$$

Finally, the projection axis is given by

$$S_w^{-1} |(\overline{m}_s - \overline{m}_b)\rangle \propto |w\rangle. \quad (\text{E.10})$$

Although this example used only two discriminating variables, the method trivially extends to any number of these. Since signal and background have to be discriminated on the basis of the acceptance functions, the acceptance functions will have to be represented as a series of “variables”. This is done by transforming the acceptance function into a vector.

Calculating $|(\overline{m}_s - \overline{m}_b)\rangle$

In order to create this vector, a mass fit is used to define a signal and sideband region in the data. The signal region contains both signal and background events, the sideband region contains background events only, and it is assumed that the background events in the sideband region are representative of the background events in the signal region. $|\overline{m}_b\rangle$ can be found from the sideband region, while $|\overline{m}_s\rangle$ can be found from the signal region using the fraction of signal events in the signal region, f_s , obtained from the mass fit:

$$|\overline{m}_s\rangle = \frac{|\overline{m}_r\rangle - (1 - f_s) |\overline{m}_b\rangle}{f_s}. \quad (\text{E.11})$$

Here $|\overline{m}_r\rangle$ refers to the vector of means for all events in the signal region, both signal and background. Finally $|(\overline{m}_s - \overline{m}_b)\rangle$ is given by:

$$|(\overline{m}_s - \overline{m}_b)\rangle = \frac{1}{f_s} |(\overline{m}_r - \overline{m}_b)\rangle. \quad (\text{E.12})$$

Calculating S_w

Following a similar procedure, consider signal and background acceptance vectors, v_s and v_b . From Equation E.3, S_w can be written as:

$$S_w = \sum_{\text{signal events}} |(v_s - \overline{m}_s)\rangle \langle(v_s - \overline{m}_s)| + \sum_{\text{background events}} |(v_b - \overline{m}_b)\rangle \langle(v_b - \overline{m}_b)|. \quad (\text{E.13})$$

The sum over background events will in fact be a sum over the background events in the sideband region, which is trivially given by

$$S_{\text{bk}} = \sum_{\text{sideband events}} |(v_i - \overline{m}_b)\rangle \langle(v_i - \overline{m}_b)|, \quad (\text{E.14})$$

and the background events in the signal region, which can be found by using the signal fraction in the signal region, f_s . Defining another sum over the sideband events,

$$S_{\text{bkassig}} = \sum_{\text{sideband events}} |(v_i - \overline{m}_s)\rangle \langle(v_i - \overline{m}_s)|, \quad (\text{E.15})$$

and the analogous sum over the signal region events,

$$S_{\text{sg+b}} = \sum_{\text{signal region events}} |(v_i - \overline{m}_s)\rangle \langle(v_i - \overline{m}_s)|, \quad (\text{E.16})$$

the formula for S_w is given by:

$$S_w = S_{\text{sg+b}} - \frac{\# \text{ bkg in sideband}}{\# \text{ bkg in signal region}} S_{\text{bkassig}} + \frac{\# \text{ bkg in dataset}}{\# \text{ bkg in sideband}} S_{\text{bk}}. \quad (\text{E.17})$$

What has happened here is that the background in the signal region is added to the sum as if it were signal in $S_{\text{sg+b}}$, and this is corrected by subtracting it out using an appropriately weighted S_{bkassig} .

Bibliography

- [ABH⁺03] F. Azfar, J. Boudreau, T. Huffman, L. Lyons, S. Malde, N. Pounder, J. Rademacker, and A. Rahaman. A Monte Carlo Independent Method for Lifetime Fits in Data biased by the Hadronic Trigger. Technical report, CDF, 2003. CDF/ANAL/BOTTOM/CDFR/6756.
- [AG07] K. Akiba and M. Gandelman. Estimate of LHCb’s sensitivity to the CKM angle γ using $B^0 \rightarrow D^0 K^{*0}$ decays. Technical Report CERN-LHCb-2007-050, CERN, Geneva, 2007.
- [All06] J. et. al. Allison. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.
- [BCF⁺01] A. Braem, E. Chesi, F. Filthaut, A. Go, C. Joram, J. Sguinot, P. Weilhammer, and T. Ypsilantis. The Pad HPD as photodetector of the LHCb RICH detectors. Technical Report CERN-LHCb-2000-063, CERN, Geneva, 2001.
- [Ben01] M. Beneke. Conceptual aspects of QCD factorization in hadronic B decays. *Journal of Physics G: Nuclear and Particle Physics*, 27(6):1069–1079, 2001.
- [Bjo89] J. D. Bjorken. Topics in B-physics. *Nuclear Physics B Proceedings Supplements*, 11:325–341, 1989.
- [BNSVH07] J. Borel, L. Nicolas, O. Schneider, and J. Van Hunen. The $B_s^0 \rightarrow D_s^\mp \pi^\pm$ and $B_s^0 \rightarrow D_s^\mp K^\pm$ selections. Technical Report CERN-LHCb-2007-017, CERN, Geneva, 2007.
- [botBc07] V. Tisserand (On behalf of the Belle and BaBar collaborations). Measurements of the CKM angle ϕ_3/γ . Technical Report hep-ex/0706.2786, 2007.
- [BPS01] C. Bauer, D. Pirjol, and I. Stewart. Proof of factorization for $B \rightarrow D\pi$. *Phys. Rev. Lett.*, 87(20):201806, Oct 2001.
- [BR97] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.
- [Bro04] N. Brook. LHCb Computing Model. Technical Report CERN-LHCb-2004-119, CERN, Geneva, Dec 2004.

- [Bur00] A. Buras. Operator Product Expansion, Renormalization Group, and Weak Decays. In P. Breitenlohner and D. Maison, editors, *Quantum Field Theory*, volume 558 of *Lecture Notes in Physics*, Berlin Springer Verlag, page 65, 2000.
- [CCFT64] J. Christenson, J. Cronin, V. Fitch, and R. Turlay. Evidence for the 2π decay of the K_2^0 meson. *Phys. Rev. Lett.*, 13:138, 1964.
- [CFG⁺98] M. Campbell, F. Formenti, T. Gys, W. Snoeys, and K. Wyllie. Development of pixel hybrid photon detectors for the RICH counters of LHCb. Technical Report CERN-LHCb-98-035, CERN, Geneva, 1998.
- [CFLM02] M. Ciuchini, E. Franco, V. Lubicz, and F. Mescia. Next-to-leading order QCD corrections to spectator effects in lifetimes of beauty hadrons. *Nucl. Phys.*, B625:211–238, 2002.
- [CLM07] M. Calvi, O. Leroy, and M. Musy. Flavour tagging algorithms and performances in LHCb. Technical Report CERN-LHCb-2007-058, CERN, Geneva, May 2007.
- [CMPP79] N. Cabibbo, L. Maiani, G. Parisi, and R. Petronzio. Bounds on the fermions and Higgs boson masses in grand unified theories. *Nuc. Phys.*, B158:295–305, 1979.
- [CMR07] S. Cohen, M. Merk, and E. Rodrigues. $\gamma + \phi_s$ sensitivity studies from combined $B_s^0 \rightarrow D_s^\mp \pi^\pm$ and $B_s^0 \rightarrow D_s^\mp K^\pm$ samples at LHCb. Technical Report CERN-LHCb-2007-041, CERN, Geneva, 2007.
- [col00a] The LHCb collaboration. LHCb Magnet Technical Design Report. Technical report, CERN, Geneva, 2000.
- [col00b] The LHCb collaboration. RICH Technical Design Report. 2000.
- [col01] The LHCb collaboration. LHCb VELO Technical Design Report. Technical report, CERN, Geneva, 2001.
- [col03a] The LHCb collaboration. LHCb reoptimized detector design and performance Technical Design Report. Technical report, CERN, Geneva, 2003.
- [col03b] The LHCb collaboration. LHCb trigger system Technical Design Report. Technical report, CERN, Geneva, 2003.
- [col07a] The LHCb collaboration. Documentation for the LHCb DAVINCI project, 2007. <http://tinyurl.com/3bauda>.
- [col07b] The LHCb collaboration. The definitions of reconstructed and reconstructible events at LHCb, 2007. <http://tinyurl.com/3dquod>.
- [col07c] The LHCb collaboration. The LHCb software homepage, 2007. <http://lhcb-comp.web.cern.ch/lhcb-comp/>.
- [Col08] The LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 2008. To be published.

- [Dji04] H. Dijkstra. Buffer Tampering”: proposal and motivations, 2004. <http://tinyurl.com/2wr358>.
- [DS88] I. Dunietz and R. G. Sachs. Asymmetry between inclusive charmed and anticharmed modes in B^0 , $\bar{B}^0$ decay as a measure of CP violation. *Phys. Rev. D*, D37:3186 – 3192, 1988.
- [ea01a] The BaBar Collaboration (B. Aubert et. al.). Observation of CP violation in the B^0 meson system. *Phys.Rev.Lett.*, 87:091801, 2001.
- [ea01b] The Belle Collaboration (K. Abe et. al.). Observation of large CP violation in the neutral B meson system. *Phys. Rev. Lett.*, 87:091802, 2001.
- [ea02] The BaBar Collaboration (B. Aubert et. al.). A study of time dependent CP -violating asymmetries and flavor oscillations in neutral B decays at the $\Upsilon(4S)$. *Phys. Rev.*, D66:032003, 2002.
- [ea03] The BaBar Collaboration (B. Aubert et. al.). Limits on the lifetime difference of neutral B mesons and on CP , T , and CPT violation in B^0 - $\bar{B}^0$ mixing. Technical Report SLAC-PUB-9696, 2003.
- [ea04a] The BaBar Collaboration (B. Aubert et. al.). Measurement of time-dependent CP asymmetries in $B^0 \rightarrow D^{(*)\pm}\pi^\mp$ decays and constraints on $\sin(2\beta + \gamma)$. *Phys. Rev. Lett.*, 92:251802, 2004.
- [ea04b] The Belle Collaboration (Y. Chao et. al.). Evidence for Direct CP Violation in $B^0 \rightarrow K^\pm \pi^\pm$ decays. *Phys.Rev.Lett.*, 93:191802, 2004.
- [ea05] The CKMfitter Group (J. Charles et al.). updated results and plots available at: <http://ckmfitter.in2p3.fr>. *Eur. Phys. J.*, C41:1–131, 2005.
- [ea06a] Heavy Flavor Averaging Group (E. Barberio et. al.). Averages of b-hadron properties at the end of 2006, 2006. [hep-ex/0704.3575v1](http://arxiv.org/abs/hep-ex/0704.3575v1).
- [ea06b] The BaBar Collaboration (B. Aubert et. al.). Measurement of time-dependent CP asymmetries in $B^0 \rightarrow D^{*\mp}\pi^\pm$ and $B^0 \rightarrow D^\mp\rho^\pm$ decays. *Phys. Rev.*, D73:111101, 2006.
- [ea06c] W.-M.Yao et. al. Particle Data Group. *J. Phys.*, G33(1), 2006.
- [Eas01] S. Easo. Tests of full-scale Pixel HPD prototype. Technical Report CERN-LHCb-2000-070, CERN, Geneva, 2001.
- [Fle03] R. Fleischer. New strategies to obtain insights into CP violation through $B_s \rightarrow D_s^\pm K^\mp$, $D_s^{*\pm} K^\mp$, ... and $B_d \rightarrow D^\pm \pi^\mp$, $D^{*\pm} \pi^\mp$, ... decays. *Nucl. Phys.*, B657:459 – 482, 2003.
- [FS98] R. Forty and O. Schneider. RICH pattern recognition. Technical Report CERN-LHCb-98-040, CERN, Geneva, Apr 1998.

- [FS06] L. Fernandez and O. Schneider. *Exclusive trigger selections and sensitivity to the B_s - $\overline{B}_s$ mixing phase at LHCb*. PhD thesis, EPFL, Lausanne, Lausanne, 2006. Presented on 14 Jul 2006.
- [GHVH⁺03] A. Golutvin, R. Hierck, J. Van Hunen, M. Prokudin, and R. White. $B_s^0 \rightarrow D_s^\mp K^\pm$ and $B_s^0 \rightarrow D_s^\mp \pi^\pm$ event selection. Technical Report CERN-LHCb-2003-127, CERN, Geneva, 2003.
- [GJ03] G. Grohall and J. Jung. Multiple objective step function maximization with genetic algorithms and simulated annealing. *Reihe Ökonomie/Economics Series*, 141, 2003.
- [Gli05] V. V. Gligorov. Matching online and offline tracks via channel ID data, 2005. <http://tinyurl.com/3xagpn>.
- [GOP04] F. Gabbiani, A. Onishchenko, and A. Petrov. Spectator effects and lifetimes of heavy hadrons. Technical report, 2004.
- [Gro95] LHC Study Group. The Large Hadron Collider: conceptual design. Technical Report CERN-AC-95-05-LHC, CERN, Geneva, 1995.
- [Gys] T. Gys. The pixel HPD webpage. <http://tinyurl.com/32muh3>.
- [Jon06a] C. Jones. PID DC06 status, 2006. <http://tinyurl.com/2k93g7>.
- [Jon06b] C. Jones. Review of DC06 results, 2006. <http://tinyurl.com/2wfeac>.
- [JR75] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.
- [KM73] M. Kobayashi and T. Maskawa. CP Violation in the renormalizable theory of weak interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.
- [Lan01] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.
- [LN07] A. Lenz and U. Nierste. Theoretical update of B_s - $\overline{B}_s$ mixing. *JHEP*, 06:072, 2007.
- [LY56] T. Lee and C. Yang. Question of parity conservation in weak interactions. *Phys. Rev.*, 104:254, 1956.
- [Noe18] E. Noether. Invariante Variationsprobleme. *Nachr. d. König. Gesselsch. d. Wiss. zu Göttingen*, pages 235 – 257, 1918.
- [Pat07] M. Patel. Including $B^\pm \rightarrow D^* K^\pm$ decays in the determination of γ via the Atwood-Dunietz-Soni Method at LHCb. Technical Report CERN-LHCb-2007-043, CERN, Geneva, 2007.

- [Pet04] A. Petrov. Lifetimes of heavy hadrons. Technical Report WSU-HEP-0407, hep-ph/0408093, 2004.
- [Pow07] A. Powell. Selection studies of $B \rightarrow D(4H)K$ events at LHCb. Technical Report CERN-LHCb-2007-004, CERN, Geneva, 2007.
- [Rad01] J. Rademacker. Evaluation of the lhcb rich detectors and a measurement of the ckm angle gamma, 2001. CERN-THESIS-2001-017.
- [Rad07] J. Rademacker. Reduction of statistical power per event due to upper lifetime cuts in lifetime measurements. *Nucl. Instrum. Meth.*, A570:525–528, 2007.
- [Sak67] A. Sakharov. Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe. *Soviet Physics Journal of Experimental and Theoretical Physics (JETP)*, 5:24–27, 1967.
- [Shi00] M. Shifman. Quark-hadron duality. Technical Report TPI-MINN-00/44, hep-ph/0009131, 2000.
- [SK05] O. Schneider and P. Koppenburg. Proposal for background classification at LHCb, 2005. <http://tinyurl.com/2jdrfy>.
- [SMS06] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.
- [SS07] A. Soni and D. Suprun. Determination of gamma from charmless $B \rightarrow M^{(1)}M^{(2)}$ decays using U-spin. *Phys. Rev.*, D75:054006, 2007.
- [TCCA05] D. E. et. al.) The CDF Collaboration (Acosta. Measurement of the lifetime difference between B_s^0 mass eigenstates. *Phys. Rev. Lett.*, 94:101803, 2005.
- [TCcG06] S.) The CDF collaboration (Giagu. Observation of $B_s^0 - \bar{B}_s^0$ oscillations and measurement of Δm_s in CDF, 2006. hep-ex/0610044.
- [TDCA07] V. M. et. al.) The D0 Collaboration (Abazov. Lifetime difference and CP-violating phase in the B_s^0 system. *Phys. Rev. Lett.*, 98:121801, 2007.
- [TUCB06] M. et. al.) The UTFit Collaboration (Bona. The unitarity triangle fit in the standard model and hadronic parameters from lattice QCD: a reappraisal after the measurements of Δm_s and $\text{BR}(B \rightarrow \tau \nu_\tau)$. *JHEP*, 10:081, 2006.
- [Ura98] N. Uraltsev. Heavy-quark expansion in beauty and its decays. Technical Report UND-HEP-98-BIG1, hep-ph/9804275, 1998.
- [Wil05] G. Wilkinson. Measuring γ at LHCb in $B^- \rightarrow D^0 K^-$ decays with an ADS method. Technical Report CERN-LHCb-2005-066, CERN, Geneva, 2005.
- [Wit] M. Witek. The buffer tampering package homepage. <http://tinyurl.com/36m7fo>.
- [Xie05] Y. Xie. Making Vertex Fit Reliable. *LHCb reconstruction meeting 22nd August*, 2005. <http://tinyurl.com/2lm5y6>.