Searches for Two-Body Charmless Baryonic B^0 Decays at LHCb

John K S Beddow

University of Glasgow
Department of Physics and Astronomy

Submitted in fulfiment of the requirements
for the degree of Doctor of Philosophy

2016

Abstract

The results of two separate searches for the rare two-body charmless baryonic decays $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ at the LHCb experiment are reported in this thesis. The first analysis uses a data sample, corresponding to an integrated luminosity of 0.9 fb^{-1}, of pp collision data collected by the LHCb experiment at a centre-of-mass energy of 7 TeV. An excess of $B^0 \to p\overline{p}$ candidates with respect to background expectations is seen with a statistical significance of 3.3 standard deviations. This constitutes the first evidence for a two-body charmless baryonic B^0 decay. No significant $B_s^0 \to p\overline{p}$ signal was observed. However, a small excess of $B_s^0 \to p\overline{p}$ events allowed the extraction of two sided confidence level intervals for the $B_s^0 \to p\overline{p}$ branching fraction using the Feldman-Cousins frequentist method. This improved the upper limit on the $B_s^0 \to p\overline{p}$ branching fraction by three orders of magnitude over previous bounds.

The 68.3% confidence level intervals on the branching fractions were measured to be

$$\begin{aligned} \mathcal{B}(B^0 \to p\overline{p}) &= (1.47\,^{+0.62\ +0.35}_{-0.51\ -0.14}) \times 10^{-8}\,, \\ \mathcal{B}(B_s^0 \to p\overline{p}) &= (2.84\,^{+2.03\ +0.85}_{-1.68\ -0.18}) \times 10^{-8}\,, \end{aligned}$$

where the first uncertainty is statistical and the second is systematic.

The second analysis followed on from the first LHCb result and included the full 2011 and 2012 samples of proton-proton collision data at centre of mass energies of 7 and 8 TeV, corresponding to a total integrated luminosity of 3.122 fb^{-1}. An excess of $B^0 \to p\overline{p}$ candidates with respect to background epectations is seen with a statistical significance of 5.9 standard deviations. This corresponds to a discovery of this decay and is the rarest hadronic B^0 meson decay ever observed.

The 68.27% and 90% confidence level intervals on the branching fraction of $B^0 \to p\overline{p}$ and the upper limit of $B_s^0 \to p\overline{p}$ are determined to be, from an unbinned maximum likelihood fit,

$$\begin{aligned}
\mathcal{B}(B^0 \to p\overline{p}) &= (1.18\,^{+0.27}_{-0.24}\,^{+0.12}_{-0.08}) \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0 \to p\overline{p}) &= (1.18\,^{+0.46}_{-0.37}\,^{+0.24}_{-0.13}) \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\overline{p}) &< 0.82 \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\overline{p}) &< 1.32 \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad .
\end{aligned}$$

where the first errors are statistical and the second are systematic.

No significant $B^0_s \to p\overline{p}$ signal is observed and an upper limit to its branching fraction of 1.3×10^{-8} at 90% CL is obtained, including all statistical and systematic uncertainties.

Acknowledgements

The completion of this thesis and the years of study it documents have been an immensely challenging yet rewarding experience. It takes a village to raise a physicist and as such there are a number of people I would like to acknowledge for their help and support throughout this process.

First of all, I would like to thank my supervisors at the University of Glasgow, Paul Soler and Lars Eklund for everything they have done for me and their continual help and guidance since my very first day with the group. I would particularly like to thank Paul for his great patience, advice and assistance during the writing of this thesis.

I would like to acknowledge the great amount of work and vital contributions made to the analyses presented in this thesis by Eduardo Rodrigues, from whom I have learned a tremendous amount. I would also like to thank my other collaborators Paul Sail and Christoph Hombach for their contributions.

During my time at CERN I was fortunate to meet many amazing people who helped to make my time on LTA a truly great experience. Special thanks go to Duncan, Ian, Simon, Mark and Josh for their friendship and camaraderie both at home and abroad.

Returning to Glasgow, thanks go to the students and academics, past and present, of the LHCb and neutrino groups with whom I've worked alongside over the years. In particular, the various officemates I've had during my time in the department: Daniel, Paul (again), Michael, Stephen and John. And to the current crop of students, Iain, Cameron, Murdo, Patrik and Lauren, I would like to wish you all good luck in the rest of your studies.

Special thanks go to Sarah for your support and companionship over the past few years. As you progress towards completing your own studies, I hope to be able to help you as much as you have helped me during mine.

Finally, I owe a huge thank you to my family. Without your unwavering support none of this would have been possible.

Declaration

The research results presented in this thesis are the product of my own work. Appropriate references are provided when results of third parties are mentioned. The research presented here was not submitted for another degree in any other department or university.

John K. S. Beddow

Preface

Chapter 1 includes a review of the available theory of the Standard Model, Flavour Physics and CP violation, with particular emphasis on a review of two-body charmless baryonic decays of B mesons, the topic of this thesis. The review was carried out by the author, with appropriate references to the relevant literature.

Chapter 2 is a summary of the details of the LHCb experiment and is mostly a review of the relevant technical literature regarding the construction and performance of the LHCb detector and its software. Plots and technical details have been extracted from the relevant detector and performance papers.

The work in Chapter 3 was carried out by the author in collaboration with colleagues from the Universities of Glasgow and Manchester (Dr. Lars Eklund, Dr. Eduardo Rodrigues, Dr. Paul Sail and Prof Paul Soler). This chapter is a summary of the analysis of the data taken with the LHCb detector in 2011 to find first evidence for the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ decay channels. This chapter follows closely the publication in Reference [1]. The author contributed to the signal selection, background determination, mass fits and confidence level calculations.

Chapters 4, 5 and 6 describe the search for $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ decays using data from the LHCb 2011 and 2012 data-taking runs. The analysis work described in these chapters, including all calculations and plots, was the author's work, with the appropriate supervision from Prof. Paul Soler and Dr. Lars Eklund, the author's supervisors at the University of Glasgow, and Dr. Eduardo Rodrigues from the University of Manchester (now at the University of Cincinnati).

Contents

List of Tables

List of Figures

Chapter 1

The Standard Model of Particle Physics and Flavour Theory

1.1 Introduction

The field of particle physics concerns itself with the study of the most fundamental building blocks of the Universe. The Standard Model (SM) of particle physics forms the backbone of our current understanding of all known fundamental particles and their interactions. The discoveries of a large number of new particles in the previous twenty years led to the development of the SM during the early 1960s. The mathematical theory of the SM was developed over the following decade into its current form, by the collaborative efforts of many theoretical physicists.

The SM has proved to be incredibly robust and has been shown to successfully predict the existence and interactions of all currently observed fundamental particles. However, despite its many successess, the SM is unable to explain certain fundamental properties of the observed Universe and is therefore currently believed to be a low energy "effective theory" of a higher energy theory. A major omission from the SM is the gravitational force, which is assumed to be one of the fundamental forces. However, gravity is known to be by far the weakest of the fundamental forces. From observations in cosmology two further issues with the SM arise. Firstly, the SM provides no candidate particle for the dark matter content of the Universe. Secondly, the SM provides no explanation for the observed asymmetry between matter and antimatter. This observed matter-antimatter asymmetry is far larger than any potential asymmetries that can be extracted from the SM. The SM also is unable to explain how neutrinos have non-zero mass, as indicated by

the observation of neutrino oscillations.

The main goals of modern high energy particle physics, such as those at the Large Hadron Collider (LHC), are to test the SM at its limits. The physics conditions producible at the LHC allow for rigorous tests of the SM and searches for evidence of further theories such as supersymmetry at energies never-before attained by particle physics experiments. The analyses presented in this thesis used data collected by the LHCb experiment. The LHCb experiment, which is described in detail in Chapter 2, is one of the major experiments at the LHC with its main physics aims being high precision measurements of Heavy Flavour physics processes in the SM and the investigation of possible new physics processes arising from discrepancies between experimental results and the predictions of the SM.

This chapter presents a description of the SM and the main physics processes concerning the analyses presented in this thesis. Section 1.2 details the particle content and mathematical formalism of the Standard Model. Following this, Section 1.3 presents a description of main areas of particle physics which are of interest to the LHCb physics program. Finally, Section 1.4 provides a description of the theory behind the specific physics processes studied in the analyses presented in Chapters 3 to 6.

1.2 The Standard Model

1.2.1 Particle content of the Standard Model

The fundamental particles making up the SM have been shown experimentally to have no observed substructure down to the scale of $\sim 10^{-19}$ m [2]. The specific individual properties of each particle in the SM can be described by a set of quantum numbers. Each particle has a corresponding partner or "antiparticle" of equal mass but with internal quantum numbers reversed. One quantum number inherent to all (anti)particles is the intrinsic angular momentum, or "spin". The Spin Statistics Theorem [3] states that all particles have either half- or full-integer spins. Full-integer spin particles, known as bosons, have fully symmetric wavefunctions under the exchange of two identical particles. However, half-integer spin particles, known as fermions, have anti-symmetric wavefunctions under the exchange of two identical particles. As a result, fermions obey the Pauli exclusion principle: that no two identical fermions may occupy the same quantum state. It is convenient to group the particles in the SM by their spin properties: fermions with half-integer spins and bosons with full integer spins

Standard Model Fermions

The fermions of the Standard Model can be divided into two groups, quarks and leptons. Quarks interact via the electroweak and strong forces. There are six types or "flavours" of quarks: three "up-type" with electric charge $+\frac{2}{3}$e, up (u), charm (c) and top (t) and three "down-type" with electric charge $-\frac{1}{3}e$, down (d), strange (s) and bottom (b). Table 1.1 lists these six quarks along with their masses and all relevant quantum numbers. Along with electric charge, quarks carry colour charge in one of three states, red, green or blue (antiquarks carry antired, antigreen or antiblue). In the SM, free particles can only exist in colour-neutral states, therefore individual free quarks are not observed in nature. Rather, quarks exist in bound states with zero net colour charge, either as a quark-antiquark pair ($q\bar{q}$) known as a meson, or as three quarks (qqq) known as a baryon. As a bound state containing quarks and antiquarks, mesons decay via $q\bar{q}$ annihilation and electroweak decay and therefore have characteristically short lifetimes. However, Long-lived baryons can be formed. The two lightest baryons are the proton (containing uud valence quarks) and the neutron (udd). Protons and neutrons combine to form the atomic nuclei of all of the visible atomic matter in the Universe. Quarks also carry flavour quantum numbers. The flavour quantum number for u and d quarks is I_3, the z component of the quark isospin. The remaining four flavours of quark each have their own respective quantum number: charm, strangeness, topness and bottomness. The sign of each flavour is equal to the sign of the corresponding quark's electric charge. The quark flavour quantum numbers are not conserved in weak interactions.

Leptons in the SM also exist in 6 different flavours: three massive particles with negative charge, $-e$, electron (e^-), muon (μ^-) and tau (τ^-) and three electrically neutral particles, or *neutrinos*, electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). Table 1.2 lists these six SM leptons along with their masses and all relevant quantum numbers. Charged leptons have corresponding antiparticles with positive charge $+e$ labelled as e^+, μ^+ and τ^+. The corresponding antiparticles of the neutrinos are labelled as $\bar{\nu_e}$, $\bar{\nu_\mu}$ and $\bar{\nu_\tau}$. It is so far undetermined experimentally whether or not neutrinos are their own antiparticles and thus known as "Majorana" neutrinos [4]. Neutrino masses are known to be non-zero as they have been observed to spontaneously change between flavour states, via a process known as neutrino oscillation [5]. However, presently only upper limits have been determined for the absolute neutrino masses.

Leptons interact exclusively via the electroweak force, therefore, they do not carry any colour charge and are permitted to exist as free particles. Leptons also carry flavour

Quark Type	Mass [MeV/c^2]	Electric Charge [e]	Generation	I_z	I_3	C	S	T	B
Up, u	$2.3^{+0.7}_{-0.5}$	$+\frac{2}{3}$	I	$+\frac{1}{2}$	$+\frac{1}{2}$	0	0	0	0
Down, d	$4.8^{+0.5}_{-0.3}$	$-\frac{1}{3}$	I	$-\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	0
Charm, c	$1,275 \pm 0.025$	$+\frac{2}{3}$	II	$+\frac{1}{2}$	0	1	0	0	0
Strange, s	95 ± 5	$-\frac{1}{3}$	II	$-\frac{1}{2}$	0	0	-1	0	0
Top, c	$173,210 \pm 510 \pm 710$	$+\frac{2}{3}$	III	$+\frac{1}{2}$	0	0	0	1	0
Bottom, b	$4,180 \pm 30$	$-\frac{1}{3}$	III	$-\frac{1}{2}$	0	0	0	0	-1

Table 1.1: Quark content of the Standard Model, all values obtained from [2]. I_z is the z component of the weak isospin of the left-handed field. I_3 is the z component of the quark's isospin. C, S, T and B are the "charm", "strangeness", "topness" and "bottomness" quantum numbers. All quarks have spin equal to $+\frac{1}{2}$ and carry either r, g or b colour charge. Antiquarks have equal masses and spins to their quark equivalents but with the signs of all remaining quantum numbers reversed.

quantum numbers. These quantum numbers are: electron number, muon number and tau number. Table 1.2 lists the six SM leptons along with their masses and all relevant quantum numbers. Lepton flavour number is conserved in all interactions, with the exception of the aforementioned neutrino oscillations. All SM calculations are performed assuming neutrinos are massless as neutrino masses are known to be far lower than the masses of all other massive SM particles.

Lepton Type	Mass [MeV/c^2]	Electric Charge	Generation	I_z	e No.	μ No.	τ No.
Electron, e	0.511	-1	I	$-\frac{1}{2}$	$+1$	0	0
e Neutrino, ν_e	$< 2 \times 10^{-6}$	0	I	$+\frac{1}{2}$	$+1$	0	0
Muon, μ	105.66	-1	II	$-\frac{1}{2}$	0	$+1$	0
μ Neutrino, ν_μ	< 0.9	0	II	$+\frac{1}{2}$	0	$+1$	0
Tau, τ	$1,776.86 \pm 0.12$	-1	III	$-\frac{1}{2}$	0	0	$+1$
τ Neutrino, ν_τ	< 18.2	0	III	$+\frac{1}{2}$	0	0	$+1$

Table 1.2: Lepton content of the Standard Model, all values obtained from [2]. I_z is the z component of the weak isospin of the left-handed field. e, μ and τ numbers are flavour quantum numbers for the lepton generations. All leptons have spin equal to $+\frac{1}{2}$.

Both quarks and leptons can be grouped into three generations which reflect the relative strengths of the interactions between different flavours in each case. Each quark generation contains one up type and one down type quark and are arranged in ascending order of mass: (u, d), (c, s), and (t, b). The three lepton generations each contain a charged lepton and a neutrino and are grouped by lepton flavour quantum number: (e^-, ν_e), (μ^-, ν_μ), and (τ^-, ν_τ). In all cases, antiparticles exist in the same generations as their particle equivalents.

Standard Model Bosons

The remaining group of SM particles are the integer-spin bosons, which are listed in Table 1.3. The SM bosons can be separated by spin into two groups: vector bosons with non-zero spin and scalar bosons, which have zero spin. There are four fundamental forces in nature: electromagnetic, weak, strong and gravity. The SM provides a mathematical theory for the interactions of particles via three of these, the electromagnetic, weak and strong forces. Gravity, being far weaker than the SM forces, so far has no quantum theory experimentally proven to describe its interactions. Gravity can therefore be ignored when describing the interactions of particles at high energies, such as those described in this thesis. For each of the three remaining forces there are vector bosons which act as gauge bosons mediating their interactions.

The photon γ is the gauge boson of the electromagnetic force. Photons couple to all electrically charged particles, i.e. all fermions except neutrinos. The photon itself is massless and not charged and does not decay into other particles or have any coupling to itself. This lack of self-coupling combined with the photon's zero mass results in the electromagetic force having an infinite range for interactions.

The weak force is mediated by the $W^\pm$ and Z^0 gauge bosons, which couple to all elementary fermions. The $W^\pm$ and Z^0 masses are large, $M_W = 80.4\,\mathrm{GeV}/c^2$ and $M_{Z^0} = 91.2\,\mathrm{GeV}/c^2$, which results in a short interaction range of $\sim 10^{-18}$ m.

The strong force is mediated by the gluon g which couples to colour charge. Like the photon, gluons are massless and electrically neutral. However, unlike photons they carry colour charge and self-interact. While quarks carry a single colour charge: red, green or blue. Gluons carry one of eight non-colour neutral combinations of colour and anticolour, e.g. red-antiblue, green-antired etc. In the SM, gluons couple to quarks, as they carry colour charge, but not leptons. As is hinted by its name, the strong force is by far the strongest of the fundamental forces, approximately 10^2 times stronger than the

electromagnetic force. As a result of its strength, gluons can only travel a short distance before interacting. Thus, the range of the strong force is limited to $\sim 10^{-15}$ m.

The last of the SM bosons is the scalar Higgs boson, H^0. In the simplest approximation of the SM all particles are massless and interactions of fermions are identical for all different generations. However, it is well established that this is not the case in the observable universe, and that fermions and the $W^{\pm}$ and Z^0 gauge bosons have mass and interact differently according to these masses. The Higgs boson was proposed as a component of the SM [6] to break its symmetry and give SM particles their masses. The experimental discovery of the Higgs boson was confirmed in 2012 by the ATLAS and CMS collaborations [7][8]. Following its discovery, further measurements have been made placing the Higgs boson mass at $M_{H^0} = 125.09 \pm 0.24\,\mathrm{GeV}/c^2$ [2].

Boson Type	Mass [MeV/c^2]	Electric Charge	Spin	I_z	Colour Charge
Photon, γ	0	0	1	0	0
$W^{\pm}$	$80,385 \pm 15$	± 1	1	± 1	0
Z^0	$91,187.6 \pm 2.1$	0	1	0	0
Gluon, g	0	0	1	0	8 combinations
Higgs, H^0	$125,090 \pm 240$	0	0	0	0

Table 1.3: Boson content of the Standard Model, all values obtained from [2].

1.2.2 The Mathematical Formalism of the Standard Model

The Standard Model is formulated as a Quantum Field Theory (QFT) which obeys gauge symmetries. Gauge symmetry, or gauge invariance, is a property of a field where measurable quantities of the field are invariant under a complete tranformation of the field. Each fundamental force in the SM is described by a gauge group with the choice of gauge group driven by the results of experimental observation. The interactions of particles in the SM are themselves manifestations of the symmetries of the SM gauge groups. Gauge groups describe the symmetries and permitted processes of the SM. A Special Unitary gauge group of order n ($SU(n)$) has $n^2 - 1$ generators, and therefore describes an interaction with $n^2 - 1$ gauge bosons. The strong force, with its eight gluons, can be described using the $SU(3)$ group, the weak force is described by $SU(2)$ and the electromagnetic

force by the $U(1)$ group. The overall gauge theory of the SM is therefore described by the combination of these gauge groups $SU(3) \times SU(2) \times U(1)$.

In classical mechanics, the dynamics of any general system can be described by its Lagrangian, L,

$$L = T - V, \tag{1.1}$$

where T and V are, respectively, the total kinetic energy and the total potential energy of the system. The dynamics of the Standard Model can also be described in terms of Lagrangians, though in this case it is more convenient to use the Lagrangian field density, $\mathcal{L}$, which is related to L via integration over the spatial component $\mathrm{d}^3\vec{x}$,

$$L = \int \mathcal{L} \, \mathrm{d}^3\vec{x}. \tag{1.2}$$

To extract the equations of motion for a field, ϕ, from its Lagrangian density, $\mathcal{L}$, one inserts the Lagrangian into the Euler-Lagrange equation.

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) = 0, \tag{1.3}$$

where ∂_μ is the covariant derivative running over the four space-time coordinates.

A key property of Lagrangian dynamics is that the Lagrangian of a system of many different processes can be defined simply as the sum of the Lagrangians of its individual components. Therefore, in the case of the SM, its complete Lagrangian can be written as the sum of the Lagrangians for all of its constituent particles and interactions:

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\text{boson masses}} + \mathcal{L}_{\text{fermion masses}} + \mathcal{L}_{\text{fermion kinetic}} + \mathcal{L}_{\mathrm{Higgs}}. \tag{1.4}$$

This Lagrangian defines the entirety of the interactions between fundamental particles in the Standard Model.

1.3 Flavour Physics and CP violation

In this section the formalism of the interactions between quarks in the SM is presented.

1.3.1 Additional Generations

The SM formalism laid out in Section 1.2.2 describes a model with only one generation of quarks (u and d) and leptons (e^- and ν_e). However, as was described in Section 1.2.1, the full SM contains three generations of fermions all with identical properties except for the masses of their constituent particles.

The additions of extra generations to the SM bring with them added complexities, especially from the phenomena of mixing between flavours and generations. Measurements of the mixings between quarks are a fundamental part of the physics programme of the LHCb experiment and therefore the theoretical framework for quark mixing in the SM is detailed in the remainder of this section.

To build up to the full three generation description of quark mixing, it is convenient to first consider the effects of adding only a second generation containing the charm and strange quarks. The properties of the charm (strange) quark are identical to those of the up (down) with the exceptions that it has a larger mass and a different flavour quantum number. In 1963, Cabibbo postulated [9] that the conservation of strangeness could be violated in weak decays. At this point only the u, d and s quarks had been discovered experimentally, though the quark model had not yet been theorised. Cabibbo proposed that the weak eigenstate d' can be represented as a rotation of the flavour eigenstates d and s such that

$$\begin{pmatrix} u \\ d' \end{pmatrix} = \begin{pmatrix} u \\ d\cos\theta_c + s\sin\theta_c \end{pmatrix}, \tag{1.5}$$

where θ_c, known as the Cabibbo angle, determines the amount of rotation of the weak eigenstate's coupling to the W boson and therefore the amount of mixing between generations. Current measurements give a value for θ_c of 13.02° [2]. The introduction of the c quark as the up-type partner of the s quark brings additional mixing terms such that a new 2×2 matrix, $\mathbf{V_C}$, describing the mixing of the two generations of quarks can be written as:

$$\mathbf{V_C} = \begin{pmatrix} V_{ud} & V_{us} \\ V_{cd} & V_{cs} \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \tag{1.6}$$

where $|V_{ij}|^2$ is the probability of quark j decaying into quark i. As particle number must be conserved in all quark interactions, $\mathbf{V_C}$ must be unitary. The resulting expression of the mixing of the weak eigenstates d' and s' with the flavour states d and s becomes:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \mathbf{V_C} \begin{pmatrix} d \\ s \end{pmatrix}. \tag{1.7}$$

This reveals that all of the mixing between the first two generations of quarks in the SM can be described by one parameter, θ_c. With the knowledge that $\theta_c \ll 45^\circ$ it becomes clear that there is a hierarchy within quark decays with on-axis, Cabibbo favoured, decays having large probabilities proportional to $\cos^2\theta_c$ and off-axis, Cabibbo suppressed, decays having much smaller probabilities, proportional to $\sin^2\theta_c$. Another notable property of the mixing matrix, $\mathbf{V_C}$, is that it contains no direct mixing terms, at the lowest order, for flavour changing interactions between d and s quarks or u and c quarks. These decays are known as Flavour Changing Neutral Currents (FCNCs) and are predicted to be forbidden in the SM. This prediction is known as the GIM mechanism [10] after its discoverers Glashow, Iliopoulos and Maiani. FCNCs have been observed to occur via second order processes. One such decay, $K_L \to \mu^+\mu^-$, which proceeds via a second order "box" diagram, actually provided evidence for the existence of the c quark before it was directly observed.

1.3.2 The CKM Matrix

With the introduction of a third generation of quarks (t and b), the mixing matrix becomes:

$$\mathbf{V_{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{1.8}$$

which is known as th Cabibbo-Kobayashi-Maskawa (CKM) matrix [11]. For the same reasons as the requirements for $\mathbf{V_C}$, $\mathbf{V_{CKM}}$ must also be unitary. For any $N \times N$ unitary matrix, the number of free parameters of the matrix is $(N-1)^2$, and the number of complex phases is $(N-1)(N-2)/2$. Therefore the 3×3 mixing matrix, $\mathbf{V_{CKM}}$, can be parameterised using three free parameters, which in this case are three mixing angles, θ_{ij}, analogous to θ_c in the two generation mixing matrix, and one complex phase, δ. The parameterisation of $\mathbf{V_{CKM}}$ using these four parameters is known as the standard

parameterisation [12]:

$$\mathbf{V_{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\mathrm{e}^{-\mathrm{i}\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & -c_{12}c_{23} - s_{12}s_{23}s_{13}\mathrm{e}^{\mathrm{i}\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.9}$$

where $\theta_{12} = \theta_C$ and $c_{ij} \equiv \cos\theta_{ij}$, $s_{ij} \equiv \sin\theta_{ij}$. The current world best measurements of θ_{ij} and δ are [2]:

$$\theta_{12} = (13.04 \pm 0.05)^\circ, \tag{1.10}$$
$$\theta_{13} = (0.201 \pm 0.011)^\circ, \tag{1.11}$$
$$\theta_{23} = (2.38 \pm 0.06)^\circ, \tag{1.12}$$
$$\delta = 1.20 \pm 0.08 \text{ rad.} \tag{1.13}$$

While quark behaviour is determined by $\mathbf{V_{CKM}}$, the mixing of antiquarks is determined by $\mathbf{V_{CKM}^\dagger}$. Therefore, any non-zero value for the complex phase, δ, would imply that the behaviour of antiquarks will differ from the behaviour of quarks. In physical terms, this would allow for the phenomenon of CP-violation, which is discussed in Section 1.3.4. As $\mathbf{V_C}$ contains no complex phase parameters, it is not possible to incorporate CP-violation via the two generation mixing matrix. Indeed, it was the desire to incorporate a mechanism for CP-violation in the model of quark interactions that motivated Kobayashi and Maskawa to develop the three generation quark model, before experimental observations of either of the third generation quarks had been made [11].

The values of the $\mathbf{V_{CKM}}$ matrix elements V_{ij} are not predicted by the theory and must be measured experimentally. The complete list of experimental measurements which have been used to contribute to the determinations of values of V_{ij} is too long to repeat in detail here. However, a detailed review of the magnitudes of the elements of $\mathbf{V_{CKM}}$ can be found in Reference [2]. The world-average experimental values for the magnitudes of V_{ij} are

$$|\mathbf{V_{CKM}}| = \begin{pmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.0011}_{-0.0012} & 0.99914 \pm 0.00005 \end{pmatrix}. \tag{1.14}$$

The unitarity requirement for $\mathbf{V_{CKM}}$ imposes constraints on the individual CKM

matrix elements such that

$$\sum_i V_{ij}V_{ik}^* = \delta_{jk}, \tag{1.15}$$

$$\sum_j V_{ij}V_{kj}^* = \delta_{ik}. \tag{1.16}$$

With three generations of quarks, there are six configurations where δ_{jk} or δ_{ik} evaluate to zero. These can be represented as triangles in a complex plane. As they are born out of the unitarity requirements of the CKM matrix, these are known as unitarity triangles. All six unitarity triangles have the same area. However, only two have the lengths of their three sides all within the same order of magnitude. Of these, the most frequently used is derived from the relation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1.17}$$

This triangle is used so frequently that it is often referred to as **the** unitary triangle. Dividing through each term in Equation 1.17 by $V_{cd}V_{cb}^*$ leaves us with a triangle with sides of length $\left|\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right|$, 1 and $\left|\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right|$, and internal angles α, β and γ where

$$\alpha = \arg\left(-\frac{\mathrm{V_{td}V_{tb}^*}}{\mathrm{V_{ud}V_{ub}^*}}\right), \tag{1.18}$$

$$\beta = \arg\left(\frac{\mathrm{V_{cd}V_{cb}^*}}{\mathrm{V_{td}V_{tb}^*}}\right), \tag{1.19}$$

$$\gamma = \arg\left(\frac{\mathrm{V_{ud}V_{ub}^*}}{\mathrm{V_{cd}V_{cb}^*}}\right). \tag{1.20}$$

Figure 1.1 shows the current experimental status of the CKM unitarity triangle fit along with limits on the CKM matrix parameters [13].

1.3.3 Symmetries

Noether's theorem [14] states that for any symmetry inherent in a system described by a Lagrangian, there is a corresponding conserved quantity. In the SM, we observe that the continuous symmetries in spatial translation, spatial rotation and time translation give rise to the universal conservations of linear momentum, angular momentum and energy. Furthermore, within the SM additional conserved quantities arise from the continuous

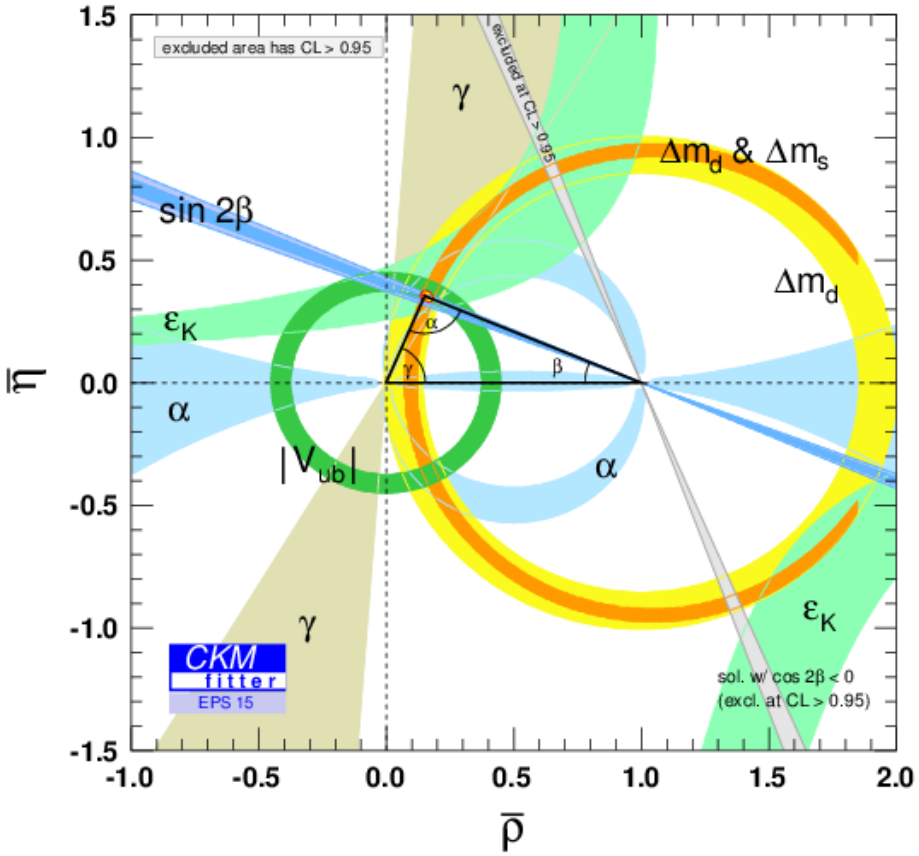


Figure 1.1: Current results of the CKM unitarity triangle fit with individual parameter constraints shown. Reproduced from [13].

symmetries of the SM Lagrangian under the transformations induced by the force carrying gauge bosons. Namely, for each fundamental force gauge boson the charge to which it couples is universally conserved in all interactions, i.e. colour charge, weak isospin and electric charge.

Discrete symmetries also give rise to additional conserved quantities in the SM. Of particular importance are the symmetries of Charge conjugation C, Parity P and Time reversal T:

- The charge conjugation transformation, C, reverses the signs of all internal quantum numbers of a particle thereby transforming particle into antiparticle, and vice versa.
- The parity transformation, P, reverses the sign of all spatial coordinates, effectively changing the chirality of a particle .

- Time reversal, T, transforms the direction of time propagation such that $t \to -t$ thereby reversing linear and angular momentum.

In order to satisfy Lorentz invariance, a Lagrangian quantum field theory must be invariant under the full combination of C, P and T in all interactions [15]. However, each symmetry in isolation, or combinations of two, may be broken. It is observed that, individually, C and P are always conserved in strong and electromagnetic interactions. In weak interactions, C and P are maximally violated as the weak gauge bosons couple only to left-handed fermions and right-handed antifermions. While the C and P symmetries are always violated in weak interactions, the combined transformation of the two, CP, is mostly conserved. However, CP violation is known to occur in some weak interactions and was first observed in 1964 through the decays of neutral kaons [16]. A discussion of the mechanisms for CP violation in the SM follow in the next section.

1.3.4 CP Violation

Since its discovery in neutral kaon decays in 1964, CP violation has been observed in a number of systems.

When considering the phenomenology of CP violation within the quark sector it is convenient to reparameterise the CKM matrix with an alternative representation. The Wolfenstein parameterisation [17] parameterises the CKM matrix in terms of $\sin\theta_{12}$ such that

$$\mathbf{V_{CKM}} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right), \tag{1.21}$$

where

$$\lambda \equiv \sin\theta_{12} \simeq 0.23, \tag{1.22}$$

$$A \equiv \frac{\sin\theta_{23}}{\sin^2\theta_{12}} \simeq 0.81, \tag{1.23}$$

$$\rho - i\eta \equiv \frac{\sin\theta_{13}e^{-i\delta}}{\sin\theta_{12}\sin\theta_{23}} \simeq 0.14 - 0.35i. \tag{1.24}$$

This parameterisation gives a clearer indication of which quark transitions should be most affected by CP-violation. At this order, the CP-violating phase, η, only appears in

transitions between the first and third generations with very little CP violation expected in transitions involving c and s quarks. Also of note is that the upper left 2×2 elements are just the expansions for sine and cosine, therefore showing that at $\mathcal{O}(\lambda^3)$ these elements of the CKM matrix are equal to V_{ij} of the Cabibbo matrix, $\mathbf{V_C}$.

Returning to the unitarity triangle defined in Equation 1.17, the internal angles of this triangle can be rewritten in terms of the Wolfenstein parameters as

$$\alpha \simeq \arg\left(-\frac{1-\rho-\mathrm{i}\eta}{\rho+\mathrm{i}\eta}\right), \tag{1.25}$$

$$\beta \simeq \arg\left(\frac{1}{1-\rho-\mathrm{i}\eta}\right), \tag{1.26}$$

$$\gamma \simeq \arg\left(\rho+\mathrm{i}\eta\right). \tag{1.27}$$

The area of this unitarity triangle, and all other CKM matrix unitarity triangles, before normalising by $V_{cd}V_{cb}^*$, is equal to half of the value of the parameter J, known as the Jarlskog invariant [18]. The Jarlskog invariant is defined by

$$\mathrm{Im}\left[V_{ij}V_{kl}V_{il}^*V_{kj}^*\right] = J\sum_{mn}\epsilon_{ikm}\epsilon_{jlm}, \tag{1.28}$$

whereby

$$J = c_{12}c_{23}c_{13}^2 s_{12}s_{23}s_{13}\sin\delta \tag{1.29}$$

$$\approx \lambda^6 A^2 \eta. \tag{1.30}$$

The Jarlskog invariant gives a parameterisation independent measure of the global CP violation in the CKM matrix model of quark interactions. It is notable that, while the existence of CP violation in the SM is dependent on the complex phase δ, the amount of CP violation is far more dependent on the mixing angles than it is dependent on δ. This can be seen clearly in Equation 1.30, where the value of J is shown to be proportional to λ^6.

Measuring CP violation in the decays of neutral mesons is a major area of activity in the LHCb physics program. Neutral mesons such as B^0, B_s^0, K^0 and D^0 can transform into their corresponding antiparticle state via charged weak interactions. In general terms, the amplitudes for the decay of a particle P, or antiparticle $\overline{\mathrm{P}}$, to a multibody final state

f or charge conjugate state $\overline{f}$ can be written as

$$A_f = \langle f|\mathcal{H}|P\rangle,\ A_{\overline{f}} = \langle \overline{f}|\mathcal{H}|P\rangle, \tag{1.31}$$

$$\overline{A}_f = \langle f|\mathcal{H}|\overline{P}\rangle,\ \overline{A}_{\overline{f}} = \langle \overline{f}|\mathcal{H}|\overline{P}\rangle, \tag{1.32}$$

where $\mathcal{H}$ is the Hamiltonian governing the weak interactions of the system.

CP violation manifests itself in the decays of mesons in three different ways:

- Direct CP violation. this has been observed in both charged and neutral meson decays and occurs when a neutral meson decay and its CP conjugate decay have different amplitudes. i.e.

$$\left|\frac{A_f}{\overline{A}_{\overline{f}}}\right| \neq 1. \tag{1.33}$$

  In this scenario, the decays P $\rightarrow$ f and $\overline{P} \rightarrow \overline{f}$ will have different rates. The amount of direct CP violation for a given decay can be measured by comparing these decay rates

$$\mathcal{A}_{\mathrm{CP}} = \frac{\mathrm{N}\left(\mathrm{P} \rightarrow f\right) - \mathrm{N}\left(\overline{\mathrm{P}} \rightarrow \overline{f}\right)}{\mathrm{N}\left(\mathrm{P} \rightarrow f\right) + \mathrm{N}\left(\overline{\mathrm{P}} \rightarrow \overline{f}\right)}. \tag{1.34}$$

- CP violation through mixing: this is an indirect form of CP violation which occurs when the mass eigenstates of a neutral meson are not CP eigenstates. This causes the mixing rate for P $\rightarrow \overline{\mathrm{P}}$ to be different from the rate of $\overline{\mathrm{P}} \rightarrow$ P.

- CP violation through interference: this form of CP violation can occur in neutral meson decays and is caused by the interference between mixing and decay. If a given final state f is accessible to both P and $\overline{\mathrm{P}}$, then the mixing process P $\rightarrow \overline{\mathrm{P}} \rightarrow f$ can interfere with the direct process P $\rightarrow f$. To test for CP violation in this case, one can measure the time dependent asymmetry

$$a_{CP}\left(t\right) = \frac{\Gamma_f - \overline{\Gamma_f}}{\Gamma_f + \overline{\Gamma_f}} \tag{1.35}$$

$$= A_{CP}^{dir}\cos\left(\Delta M t\right) + A_{CP}^{mix}\sin\left(\Delta M t\right), \tag{1.36}$$

  where Γ_f is the decay rate of the process P $\rightarrow f$, and M can be defined as

$$|M|^2 = \frac{1}{2}\left(|A_f|^2 + \left|\overline{A_f}\right|^2\right). \tag{1.37}$$

Following this procedure allows the extraction of separate terms for both the direct and mixing-induced CP asymmetries, A_{CP}^{dir} and A_{CP}^{mix}.

1.4 Two-body Charmless Baryonic B decays

1.4.1 Review of Two-body Hadronic Decays of B Mesons

The study of two-body charmless hadronic B decays is an important area of the physics programmes of modern flavour physics experiments. Within LHCb, the most commonly studied decays of this type involve the decay of a B-hadron to a two body final state consisting of charged kaons, pions and protons. The LHCb detector is well equipped to cleanly detect decays of this type, with excellent kaon, pion and proton particle identification performance. A range of analyses have been performed by the LHCb experiment on $H_b \to h^+h^-$ decays including:[1]

- Measurements of CP violation in $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$ [19].
- Lifetime measurements of dominant modes such as $B_s^0 \to K^+K^-$, $B^0 \to K^+\pi^-$ and $B_s^0 \to \pi^+K^-$ [20, 21].
- Searches and branching fraction measurements of rare decays such as $B^0 \to K^+K^-$ [22].

Charmless $b \to u$ decays are suppressed compared to $b \to c$ by $|V_{ub}/V_{cb}|^2 \sim (0.1)^2$. The decays of B^0 and B_s^0 mesons to two charged mesons (K^+K^-, $\pi^+\pi^-$ or $K^\pm\pi^\mp$), henceforth referred to as $B \to h^+h^-$ decays, have branching fractions typically of the order 10^{-5} to 10^{-7} [2]. These decays have contributions from both tree level $b \to u$ processes and $b \to sg$ penguin diagrams. Figure 1.2 shows the general topologies of tree and penguin contributions to these $B \to h^+h^-$ decays where one can see that the decays are interchangeable under the exchange of d and s quarks. Theoretically, $B \to h^+h^-$ decays are relatively well understood with theoretical accuracy only affected by non-factorisable U-spin breaking effects.

[1] Here we define H_b as either B^0, B_s^0 or Λ_b^0, and h as either a kaon, K, pion, π, or proton, p.

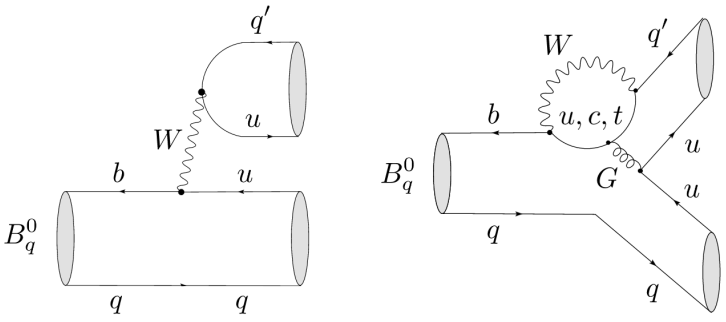


Figure 1.2: Tree and penguin topologies contributing to $B^0 \to \pi^+\pi^-$, $B_s^0 \to K^+K^-$, $B^0 \to K^+\pi^-$ and $B_s^0 \to \pi^+K^-$ decays, $(q, q' \in d, s)$. Reproduced from [23].

1.4.2 Theoretical models of Two-body Charmless Baryonic B decays

In contrast to the $B \to h^+h^-$ decays discussed in the previous section, charmless decays of B hadrons to a baryon and an antibaryon are poorly understood theoretically. While topologically similar to $B \to h^+h^-$ decays, with a B hadron decaying to two charged hadrons, baryonic B decays require a quark-antiquark pair to be pulled from the vacuum to make up the final state. Even in the event of the required extra $q\bar{q}$ being pulled from the vacuum, this is far more likely to result in a final state consisting of three mesons rather than two baryons. This makes them greatly suppressed compared to $B \to h^+h^-$, and $B \to h^+h^-h^0$ decays. Two-body baryonic B decays are also suppressed relative to three-body decays of the form $B \to \text{Baryon}^+\text{Baryon}^-\text{Meson}^{\pm,0}$ due to differences in the kinematics of the gluon emission required to create the extra $q\bar{q}$ pair in the final state. In two-body decays, an energetic $q\bar{q}$ pair must be emitted back to back by a highly off mass-shell gluon. This hard gluon process suppresses the two-body decay amplitudes by a factor of α_s/t, where α_s is the strong coupling constant and t is the four-momentum square transferred through the gluon. In contrast, for three-body decays, the baryon and antibaryon can be produced colinearly, opposing the direction of the final state meson. In this case the $q\bar{q}$ pair can be produced by a gluon much closer to the $q\bar{q}$ invariant mass shell and are therefore not affected by the same suppression factor as the two-body decays. This also provides an explanation for the observed "threshold enhancement" in three-body decays, where the baryon-antibaryon invariant mass is seen to peak near the low mass threshold [24].

A number of searches for two-body charmless baryonic B decays have been performed

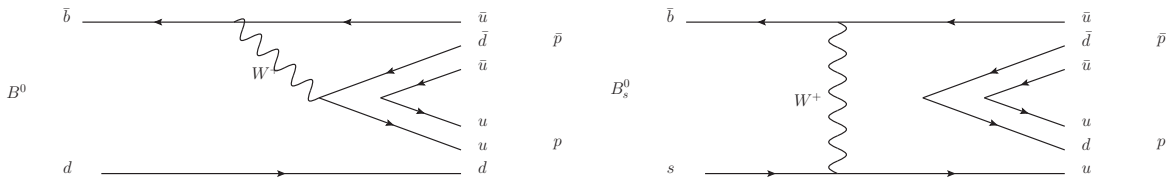
with the CLEO detector and at the SLAC and KEK B-factories as well as the LHCb experiment [25, 26, 27]. Thus far, only one such B decay has been observed, $B^+ \to p\bar{\Lambda}(1520)$, which was discovered at LHCb in 2013 [28]. The analyses presented in this thesis give details of the searches for the decays $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ performed using proton-proton collision data collected by the LHCb experiment during Run I. Table 1.4 lists the the best upper limits on the branching fractions of several two-body charmless baryonic decay modes prior to Run I of the LHC. The experimental 90% confidence level (CL) upper limit on the $B^0 \to p\overline{p}$ branching fraction, 1.1×10^{-7}, was dominated by the latest BELLE search [27], which used 414 fb^{-1} of data.

Decay Channel	BELLE UL [27]	BABAR UL [26]	CLEO UL [25]	ALEPH UL [29]
$B^0 \to p\overline{p}$	1.1×10^{-7}	2.7×10^{-7}	7.0×10^{-6}	
$B^0 \to \Lambda\overline{\Lambda}$	3.2×10^{-7}		3.9×10^{-6}	
$B^+ \to p\overline{\Lambda}$	3.2×10^{-7}		2.6×10^{-6}	
$B_s^0 \to p\overline{p}$				5.9×10^{-5}

Table 1.4: Experimental upper limits on the branching ratios of different charmless two-body baryonic B decays prior to Run I of the LHC. Limits shown correspond to a 90% confidence level.

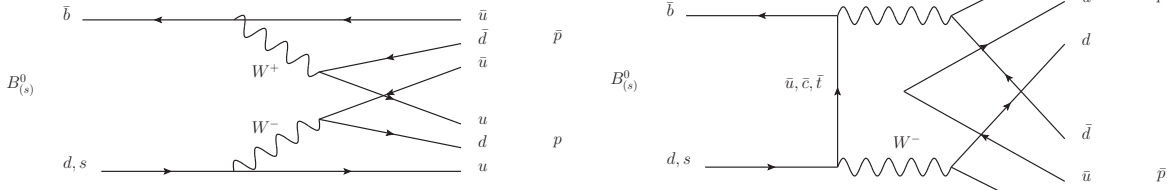
The $B_{(s)}^0 \to p\overline{p}$ decay modes are expected to be the simplest two-body charmless baryonic B decays to search for experimentally as the decay products are two stable, charged, particles. All other charmless baryon final states involve either neutral particles, such as neutrons, or shorter-lifetime particles which decay within a detector to multibody final states. The excellent B hadron vertexing performance and proton particle identification capabilities of the LHCb detector make $B_{(s)}^0 \to p\overline{p}$ decays a natural choice for investigation.

The dominant decay amplitude for $B^0 \to p\overline{p}$ is expected to be, in the Standard Model (SM), the $b \to u$ tree-level process shown in Fig. 1.3(a). Penguin annihilation and other electroweak processes such as W-exchange can also contribute, but should have a rather small influence; Fig. 1.3 presents a few examples. Unlike $B^0 \to p\overline{p}$, the $B_s^0 \to p\overline{p}$ mode cannot decay via a simple tree-level diagram and its decay is therefore predicted to be further suppressed. Several loop-level diagrams are shown in Fig. 1.3. As mentioned previously, all relevant (main contributing) topologies require a quark-antiquark pair to pop out of the vacuum.

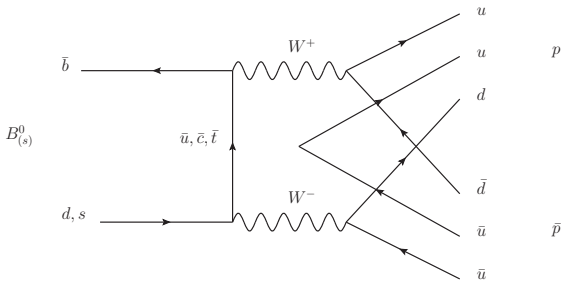


(a) Tree-level

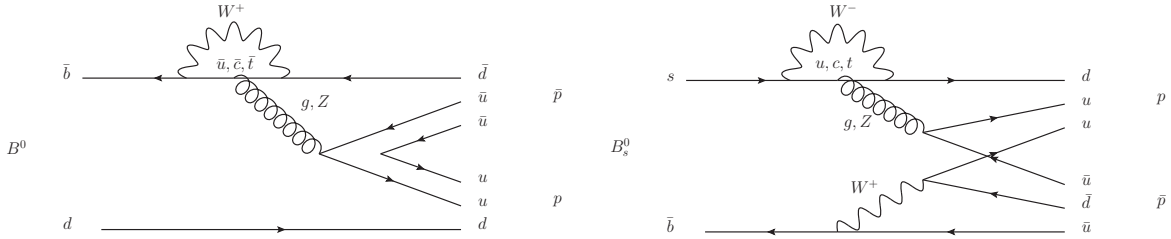
(b) W-exchange



(c) "Double W-emission"

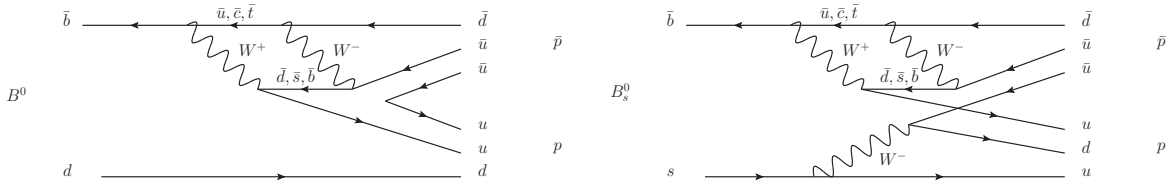


(d) "Half-box"



(e) Penguin

(f) Penguin



(g) "Box"

(h) "Box"

Figure 1.3: A selection of Feynman diagrams contributing to the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ decays.

Direct calculations of the branching fractions for such two-body baryonic decays within the SM are difficult to perform as they are generally not factorisable. Instead, calculations must be made using models [30]. Such branching fraction predictions vary depending on the method of calculation used, *e.g.* QCD sum rules, diquark model, pole model. Prior to the publication of the results produced by the main analyses discussed in this thesis, the predicted branching fractions were thought to be of the order 10^{-7}–10^{-6} [31, 32, 33, 34, 35], see Tab. 1.5.

Decay channel	QCD sum rule [31]	Diquark model[32]	Pole models	
			Harmonic oscillator model [33, 34]	MIT Bag Model [35]
$B^0 \to p\overline{p}$	1.2×10^{-6}	2.9×10^{-6}	7.0×10^{-6}	1.1×10^{-7}
$B^0 \to \Lambda\overline{\Lambda}$			2×10^{-7}	0
$B^+ \to p\overline{\Lambda}$	$\lesssim 3 \times 10^{-6}$			2.2×10^{-7}

Table 1.5: Theoretical predictions for the branching fractions of different baryonic two-body B decays. A blank entry indicates that the branching fraction was not calculated using that particular model.

In contrast to the state of two-body charmless baryonic B decays, decays of B hadrons to two charmed baryons, or one charmed and one charmless baryon, have been observed with typical branching fractions of the order $10^{-3} - 10^{-5}$. In general terms, one should be able to express the branching fractions of a two body charmless baryonic B decays in terms of the branching fraction of its charm containing equivalent. Using $B^0 \to p\overline{p}$ as an example, its branching fraction can be related to the branching fraction of the decay $B^0 \to p\bar{\Lambda}_c^-$ by

$$\begin{aligned} \mathcal{B}(\overline{B^0} \to p\overline{p}) &= \mathcal{B}(\overline{B^0} \to \Lambda_c^+ \overline{p}) \times |V_{ub}/V_{cb}|^2 \times f_{dynamic} \\ &\sim 2.7 \times 10^{-7} \times f_{\text{dynamic}}, \end{aligned} \tag{1.38}$$

where f_{dynamic} is a dynamic suppression factor.

To calculate the value of f_{dynamic}, further experimental observations of two-body charmless baryonic B decays are required. Progress on the theoretical calculations of two-body charmless baryonic B decays has been directly affected by the lack of experi-

mental observations of the modes under investigation.

Figure 1.4 shows a comparison of the statuses of theoretical predictions and 90% CL experimental limits on the $B^0 \to p\overline{p}$ branching fraction prior to the publications of the results presented in this thesis. The plotted values are taken from the results presented in Tables 1.4 and 1.5. It is interesting to note that the most recent BELLE result had already ruled out all but one of the existing theoretical predictions.

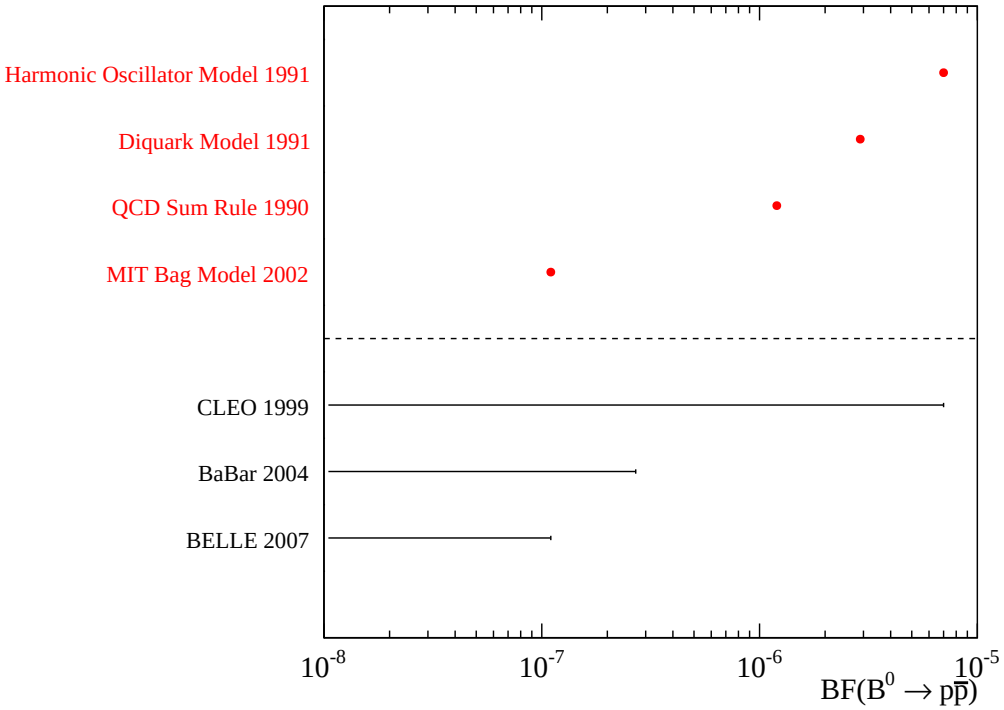


Figure 1.4: Theoretical and experimental limits of $\mathcal{B}\left(B^0 \to p\overline{p}\right)$ prior to Run I of the LHC. Theoretical predictions are shown in red and are taken from the values listed in Table 1.5. The uncertainties on the theoretical predictions were not provided. Experimental results are shown in black and are taken from the 90% CL values listed in Table 1.4.

1.4.3 Recent Activity

As will be described in detail in the following chapters, branching fraction measurements were made for $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ by the LHCb experiment using 0.9 fb^{-1} of proton-proton collision data at $\sqrt{s} = 7\,\text{TeV}$. These results provided the first evidence for $B^0 \to p\overline{p}$, with a statistical significance of 3.3 standard deviations, as documented in

References [36, 1]. This result was carried out by the author, in a collaboration between the Universities of Glasgow and Manchester, and is reported in Chapter 3 of this thesis. While no significant $B^0_s \to p\overline{p}$ signal was observed, two-sided confidence level intervals were placed on the $B^0 \to p\overline{p}$ and $B^0_s \to p\overline{p}$ branching fractions for the first time:

$$\mathcal{B}(B^0 \to p\overline{p}) = (1.47\,^{+0.62}_{-0.51}\,^{+0.35}_{-0.14}) \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL},$$
$$\mathcal{B}(B^0_s \to p\overline{p}) = (2.84\,^{+2.03}_{-1.68}\,^{+0.85}_{-0.18}) \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL},$$

where the first errors are statistical and the second are systematic. These results ruled out all of the existing published predictions for the $B^0_{(s)} \to p\overline{p}$ branching fractions by at least an order of magnitude.

Following the publication of these results, new theoretical predictions have been made to try and explain this suppression of the $B^0_{(s)} \to p\overline{p}$ decay modes and to expand further and predict the branching fractions of all the two-body baryonic modes of ground state octet and decuplet baryons [30]. Two theories have been put forward so far to explain the smallness of $\mathcal{B}(B^0 \to p\overline{p})$, which are able to place $\mathcal{B}(B^0_{(s)} \to p\overline{p})$ of the order 10^{-8}. Namely that the axial-vector current is not asymptotically conserved [37], and that internal W-emission is partially cancelled by Fierz transformation [38].

Hsiao and Geng [37] put forward the hypothesis that the partial conservation of axial-vector current (PCAC) is violated at the GeV scale. This could provide an explanation for the suppression of charmless two-body baryonic B decays. The violation of PCAC would allow the annihilation mechanism to be applied to all two-body baryonic $B^0_{(s)}$ decays. Applying modified form factors, Hsiao and Cheng are able to reproduce the measured $B^0_{(s)} \to p\overline{p}$ and $D^+_s \to p\overline{n}$ branching fractions with this method. Expanding further, they make predictions for the branching fractions of a number of charmless two-body baryonic B decays. The predicted branching fractions for the decays $B^+ \to p\overline{\Lambda}$, $\left(3.5^{+0.7}_{-0.5}\right) \times 10^{-8}$, and $B^0_s \to \Lambda\overline{\Lambda}$, $\left(5.3^{+1.4}_{-1.2}\right) \times 10^{-8}$, are larger than that of $B^0 \to p\overline{p}$. Searches for these decays would be within the current capabilities of the LHCb experiment and would provide further tests of the annihilation mechanism.

In Reference [38], the authors consider the effects of applying the Fierz transformation, which effectively replaces colour and flavour indices within a given Feynman diagram, for tree-dominated charmless two-body baryonic B decays. The authors propose that for a subset of Feynman diagrams contributing to internal W-emission decays, the decay amplitude of the tree operator O_i cancels with the amplitude of its corresponding Fierz-transformed operator O'_i, therefore causing further suppression of the decay. The authors also note that this partial cancellation is not expected to occur in two-body baryonic

B decays to final states involving charm baryons, and so provides a potential source for f_{dynamic}, the dynamic suppression of $B^0 \to p\overline{p}$ compared to $B^0 \to p\bar{\Lambda}_c^-$. Using the results of the measured $B^0 \to p\overline{p}$ branching fraction, these authors have made predictions of the branching fractions of a number of charmless two-body baryonic B decay modes (see Table 1 in Reference [30]). Despite only having $\mathcal{B}(B^0 \to p\overline{p})$ as an experimental value to put into their calculations, all of the resulting branching fraction predictions are consistent with current measured upper limits on other charmless two-body baryonic B decay modes. The predicted branching fractions for the decays $B^+ \to p\overline{\Lambda}$, $\left(10.03^{+14.14+42.79}_{-6.62-9.91} \pm 0.05\right) \times 10^{-8}$, and $B_s^0 \to \Lambda\overline{\Lambda}$, $\left(6.33^{+8.71+26.02}_{-4.11-6.27} \pm 0.27\right) \times 10^{-8}$, are larger than that of $B^0 \to p\overline{p}$. These predictions are consistent with the predicted branching fractions given by Hsiao and Geng, albeit with much larger uncertainties.

1.5 Summary

This chapter has presented an overview of the Standard Model of particle physics with particular focus on the theories governing the interactions of quarks and the motivations for the main physics analyses presented in this thesis. The particle content and mathematical formalism of the Standard Model are presented in Section 1.2. Following this, Section 1.3 provides a description of the CKM mechanism and the interactions between quarks with particular attention given to the phenomenon of CP violation. Section 1.4 covers the theory outlining the decays of B hadrons to two-body charmless baryonic final states. The current status of theoretical predictions for the branching fractions of this family of decays is described and compared to existing experimental results. This section also provides motivations for the analyses presented in Chapters 3 to 6.

Chapter 2

The LHCb Detector

2.1 The LHC

The Large Hadron Collider (LHC) [39], based at the European Centre for Nuclear Research (CERN) near Geneva, Switzerland, is both the largest and most powerful particle accelerator in the world. Situated at a mean depth of 100 m, the collider sits within a 26.7 km circumference tunnel previously used by the Large Electron-Positron (LEP) collider. The LHC is a proton-proton collider with a nominal centre-of-mass energy ($\sqrt{s}$) of 14 TeV (7 TeV per beam) and frequency of 40 MHz. During the intial running of the collider in 2011-2012 the machine was run at maximum energies of $\sqrt{s} = 7$ TeV in 2011 and $\sqrt{s} = 8$ TeV in 2012, with a maximum frequency of 20 MHz.

The process of accelerating protons within the LHC involves several stages and utilises a number of existing accelerators. Figure 2.1 shows the full acceleration chain of the LHC. Protons used in the $p-p$ collisions of the LHC start off as hydrogen atoms which are ionised and then passed through to a Linear Accelerator (LINAC2) where they are accelerated up to an energy of 50 MeV. Following the LINAC2 stage, the protons pass through a series of proton synchrotrons: the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). The protons exit the SPS having reached an energy of 450 GeV at which point they can be injected into the LHC ring. Within the LHC ring protons are injected in bunches into two separate beam pipes, which circulate in opposite directions, and are accelerated to their ultimate energies: 3.5 TeV in 2011 and 4 TeV in 2012.

Figure 2.1 also shows the positions of the four main experiments studying LHC collisions:

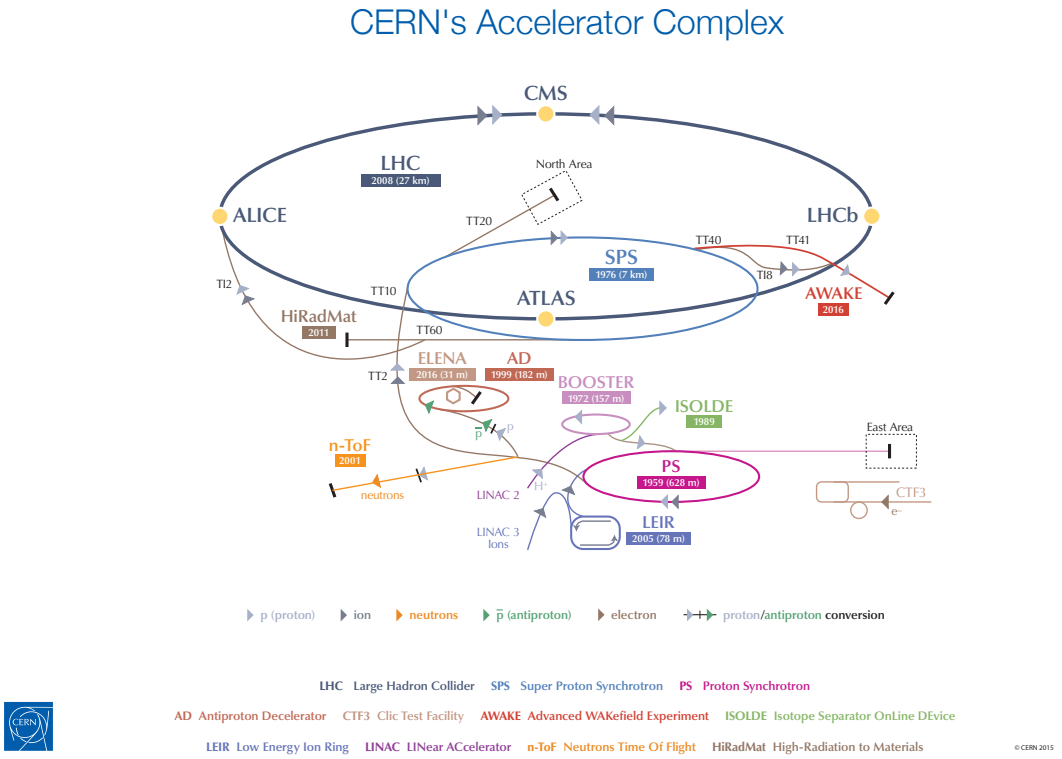


Figure 2.1: The LHC accelerator complex, reproduced from [40].

- ALICE[41] ALICE is a heavy ion collision experiment focusing on the study of quark-gluon plasma, mimicking the conditions of the universe shortly after the big bang.

- ATLAS[42] and CMS[43] are general purpose detectors (GPDs) with different designs but with common, wide ranging physics programmes with major interests in searches for the Higgs boson and new physics (NP), such as supersymmetry as well as precision tests of the parameters of the SM.

- LHCb[44] focuses on flavour physics with specific interest in searching for indirect evidence of new physics through CP violation and rare decays of beauty and charm hadrons. The data analysed and presented in this thesis was collected by the LHCb detector and thus detailed descriptions of the LHCb experiment and detector are given in the remainder of this chapter.

2.2 The LHCb Detector

The main focus of the LHCb experiment is to study the decays of hadrons containing b and c quarks. Proton-proton collisions, such as those occuring at the LHC, produce b quarks in $b\bar{b}$ pairs. At the high $\sqrt{s}$ energies of the LHC, $b\bar{b}$ pairs are produced at very small polar angles with respect to the beam direction with the b and $\bar{b}$ characteristically produced in the same forward or backward cone, as shown in Figure 2.2. As a result, the LHCb detector has a forward facing geometry with an angular coverage of approximately 15 mrad to 300 (250) mrad in the horizontal (vertical) plane [44]. This corresponds to approximately only 2 % of the full solid angle but accepts approximately 30 % of b quarks produced from LHC collisions.

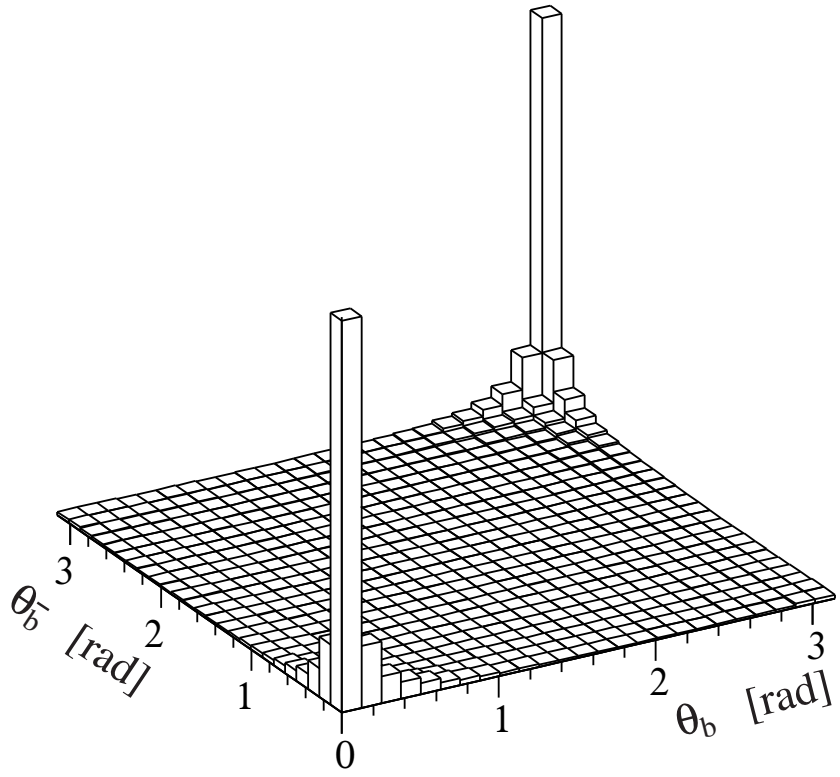


Figure 2.2: Simulated $b\bar{b}$ production angles from $\sqrt{s} = 14$ TeV proton-proton collisions, relative to the beam direction. Reproduced from [45].

Analyses involving the decays of hadrons containing b and c quarks rely on the accurate reconstruction of displaced decay vertices. To aid in the reconstruction of such displaced vertices, it is beneficial for the detector to be exposed to events with relatively low pile-up μ_{vis}, defined as the average number of reconstructable interactions per bunch crossing. The maximum luminosities delivered by the LHC to the ATLAS and CMS experiments

during Run I corresponded to average μ_{vis} values of $\sim$ 10 in 2011 and $\sim$ 20 in 2012. These amounts of pile-up would be prohibitive for adequate performance of the LHCb detector. To combat this, the luminosities delivered to LHCb were reduced from the nominal LHC luminosities by changing the beam focus at the LHCb interaction point. Figure 2.3 shows the average number of visible interactions per bunch crossing and the instantaneous luminosities measured at the LHCb interaction point during Run I of the LHC. The figure shows that during 2011 the experiment ran at instantaneous luminosities of up to $4 \times 10^{32}\text{cm}^{-2}\text{s}^{-1}$, which is double the design luminosity, corresponding to $\sim$ 1.5 visible interactions per beam crossing. During 2011 a method was developed to actively control the instantaneous luminosity at the interaction point throughout each fill. This is achieved by adjusting the transverse overlap of the colliding beams, bringing the beams closer together as time increases during a fill. This luminosity levelling procedure was fully adopted throughout 2012 data taking, the results of which can be seen in Fig 2.3 where the LHC consistently delivered a peak luminosity of $4 \times 10^{32}\text{cm}^{-2}\text{s}^{-1}$ with an average μ_{vis} of 1.7 visible events per bunch crossing.

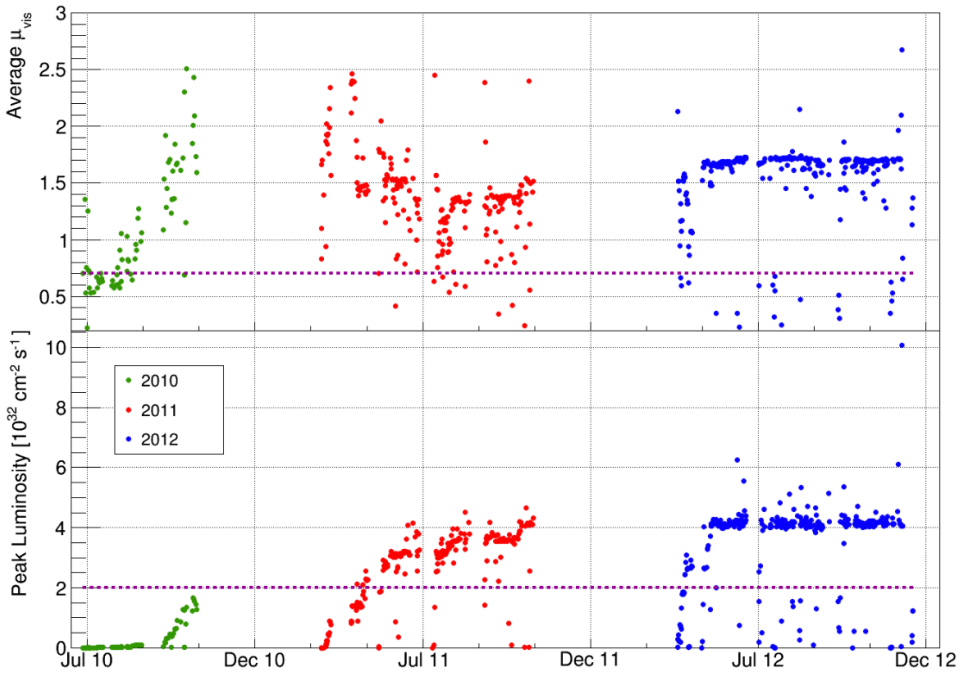


Figure 2.3: LHCb Run I operating conditions showing average number of interactions per bunch crossing (top) and instantaneous luminosity (bottom). Dotted lines show design specifications. Reproduced from [46].

A cross-sectional view of the LHCb detector is shown in Figure 2.4. The geometry

of the detector is described using a right-handed coordinate system with the z direction orientated along the beam axis, y in the vertical and x in the horizontal with the positive x direction going into the page. The LHCb detector is comprised of several subdetector systems (Figure 2.4) each designed to play a specific role in the reconstruction and analysis of physics events. In order of proximity to the interaction point they are: the Vertex Locator (VELO), the first Ring Imaging Cherenkov detector (RICH 1), the Tracker Turicensis (TT), the dipole magnet, the tracking stations (T1-T3), the second Ring Imaging Cherenkov detector (RICH 2), the first muon station (M1), the Scintillator Pad Detector and Preshower (SPD/PS), the Electromagnetic Calorimeter (ECAL), the Hadronic Calorimeter (HCAL) and the four muon stations (M2-M5). Each of these subdetector systems are described in detail in the remaining sections of this chapter.

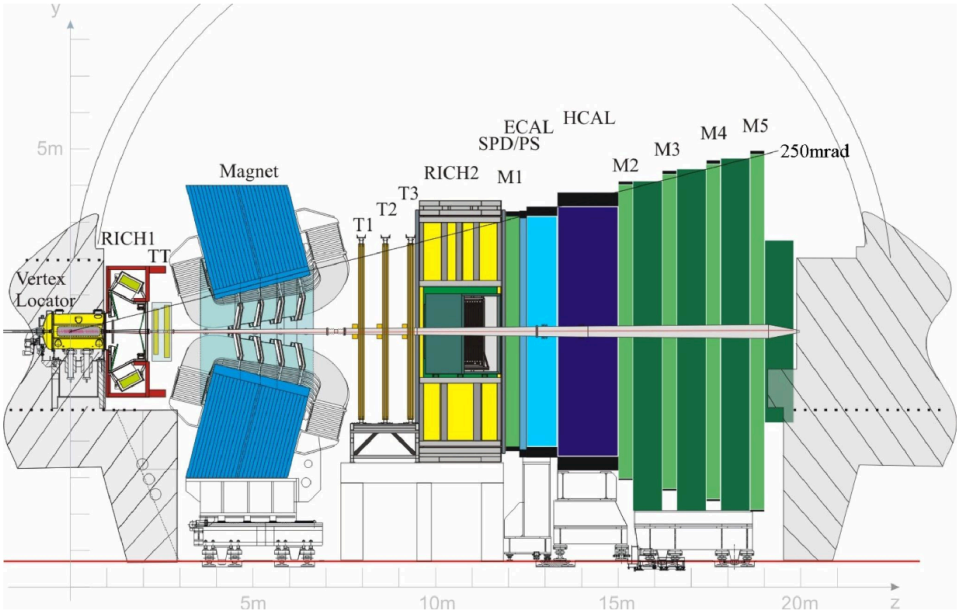


Figure 2.4: View of the LHCb detector, Reproduced from [44].

2.2.1 Summary of LHCb Run I Data Taking

During Run I of proton-proton collisions at the LHC, the LHCb experiment recorded over 3 fb^{-1} of proton collison data at centre of mass energies of 7 and 8 TeV. Table 2.1 shows a summary of the data taking conditions at the LHCb during Run I. With the $b\bar{b}$ cross section, $\sigma_{b\bar{b}}$, measured as $\sigma_{b\bar{b}} = (284 \pm 20 \pm 49)$ µb [47] (at $\sqrt{s} = 7$ TeV), this corresponds to a $b\bar{b}$ production rate at the LHC of approximately 100,000 per second. If we assume

that $\sigma_{b\bar{b}}$ scales linearly with $\sqrt{s}$ then this rate increases to approximately 114,000 $b\bar{b}$ pairs per second during 2012 running conditions.

Figure 2.5 shows the time evolution of the delivered and recorded integrated luminosities at LHCb in 2010, 2011 and 2012. The average operational efficiency, defined as the ratio of recorded and delivered integrated luminosities, across Run I was 93%. In 2012 alone, the operational efficiency was 95%. The two analyses presented in this thesis were performed using data recorded by the LHCb experiment in 2011 and 2012, corresponding to integrated luminosities of 0.90 fb^{-1} and 3.12 fb^{-1} respectively.

Year	$\sqrt{s}$ (TeV)	Instantaneous Luminosity ($\text{cm}^{-2}\text{s}^{-1}$)	μ_{vis}	Integrated Luminosity (fb^{-1})
2010	7	1×10^{32}	$0.5 - 2.5$	0.04
2011	7	3.5×10^{32}	1.5	1.11
2012	8	4×10^{32}	1.7	2.08

Table 2.1: Summary of LHCb proton-proton collision data taking conditions for the years 2010 to 2012. Values taken from [46].

2.3 Vertex Locator

The VELO forms part of the LHCb tracking system and is responsible for providing precise tracking measurements close to the interaction point [48]. The main job of the VELO is to reconstruct the primary vertices (PVs) of proton-proton collisions. With characteristic lifetimes on the order of 1 ps, b hadrons produced at the interaction point travel a short but measurable distance within the detector before decaying into their final states. It is also the job of the VELO to reconstruct these displaced or secondary vertices (SVs). The precise reconstruction of primary and secondary vertices is of great use in discriminating heavy flavour decays of b and c hadrons from light flavour processes.

2.3.1 VELO Design and Construction

The design requirements of the VELO demand that it provides measurements of charged particle tracks with excellent spatial resolution and efficiency in an extreme radiation environment. There are also tight requirements on the angular acceptance of the VELO to match the angular acceptances of the downstream subdetectors.

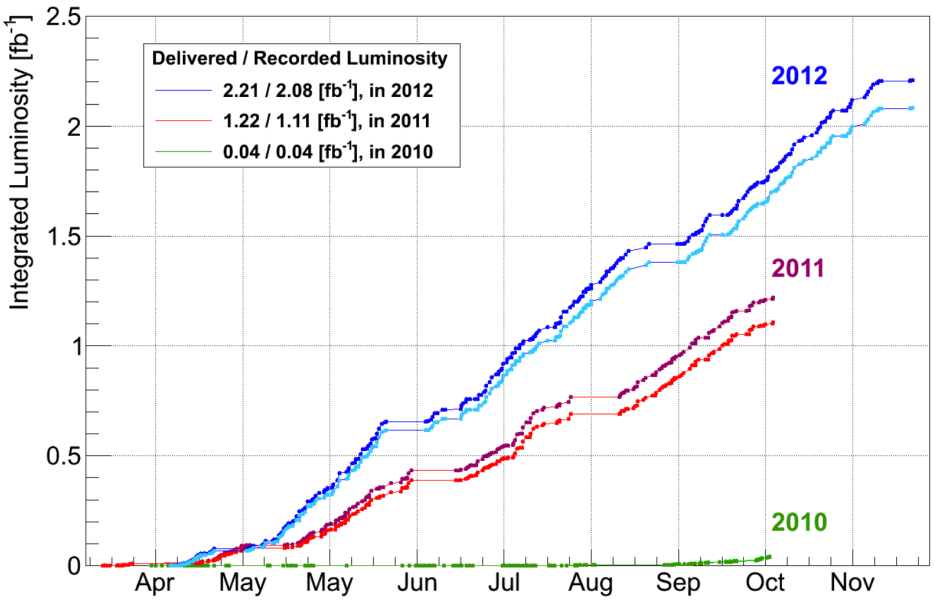


Figure 2.5: Integrated luminosity as a function of time for the LHCb experiment during 2010, 2011 and 2012. The dark coloured lines show the delivered integrated luminosities while the lighter coloured lines show the integrated luminosities recorded by the experiment. From [46].

To enable precise vertex measurements, the VELO needs to be located as close to the interaction point as possible to minimise the extrapolation distance between the first hit of a reconstructed track and the interaction point. When taking data, the VELO is positioned at a radial distance of 8 mm from the LHC beam axis, this is smaller than the width of the LHC proton beams during injection. Therefore the VELO subdetector system is designed as two retractable halves that are each moved to a safe distance of 30 mm from the beam axis during injection. When stable beams are achieved, the two halves are closed together until each is approximately 8 mm from the beam axis. At this distance, VELO material is actually inside the radius nominally covered by the beam pipe. Therefore, to maintain the vacuum conditions within the beampipe, the VELO detectors are mounted within an evacuated vessel (known as the *RF-box*) with a thin layer of aluminium foil (known as *RF-foil*) in place to separate it from the beam vacuum. Figure 2.6 shows an exploded view of the layout of one half of the VELO with the *RF-box* pulled away.

The VELO is comprised of twenty-one silicon detector stations laid out along the z-axis. Each station consists of two modules, one either side of the beam pipe, both containing two sensors: one measuring the radial position (R) and one measuring the

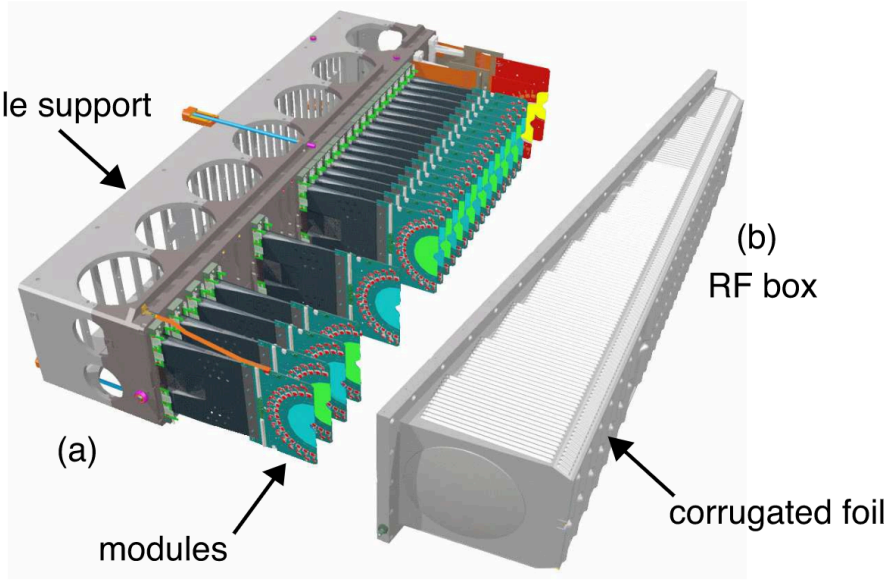


Figure 2.6: An illustration of the modules and module support for one half of the fully assembled VELO subdetector. Reproduced from [44].

azimuthal angle (ϕ). The R- and ϕ-sensors within each station are placed back-to-back and have an angular coverage of $\sim 182°$. Figure 2.7 shows the layout of the VELO detector stations in z (top) and the front faces of the first modules in their closed (left) and open (right) positions. In addition to the twenty-one $R\phi$ stations, there are two R-sensor pileup VETO stations upstream of the interaction point, which provide information for the Level 0 trigger (see Sec. 2.9.1) helping to reject events with excessively large numbers of tracks. As is illustrated in the figure, there is an offset of 15 mm in z between the positions of the two halves of each detector station. Combined with the $\sim 182°$ angular coverage of each sensor, this offset allows the two detector halves to overlap when brought together for data taking and provide full azimuthal acceptance.

The VELO R and ϕ sensors are made of 300 μm thick silicon wafers each covered with 2048 aluminium strips. The silicon detector material is constructed of n-implants on n-bulk material with a p-spray applied to allow strip isolation. The layout and pitches of the R and ϕ sensor strips are shown in Figure 2.8. The R-sensor strips are divided into four subdivisions each covering a 45° angle with the strips in each subdivision laid out in concentric semi-circles. The particle flux at a given radius, r, from the beam is proportional to $\frac{1}{r^2}$ and so R-sensor strip pitch sizes increase linearly from 38 μm at the innermost radius up to 101.6 μm at the outer radius of 41.9 mm. The strips on the ϕ-sensors are arranged radially in order to provide an orthogonal position measurement

to the R-sensor. To reduce the strip lengths, ϕ-sensors are divided into an inner region and an outer region. The inner region is defined from the innermost radius of the detector up to a radial distance of 17.25 mm. The strip pitch in the inner region ranges linearly from 38 μm at the innermost radius to 78.3 μm at 17.25 mm with a skew for each strip of $\sim 20°$ to the radial at the innermost radius of the sensor. This skew is introduced to improve pattern recognition and is reversed in adjacent ϕ-sensors to aid with the discrimination of ghost hits with respect to true hits. The outer region covers the detector area at radii greater than 17.5 mm. Strip pitches in the outer region range from 39.3 μm at 17.5 mm to 97 μm at the outer radius with the skew on each strip set to $\sim -10°$ to the radial at a radius of 17 mm. The skews of the inner and outer regions are reversed, as shown in the figure.

Signal readout from each VELO sensor is performed by sixteen Beetle chips [44] which read out at a rate of 1 MHz.

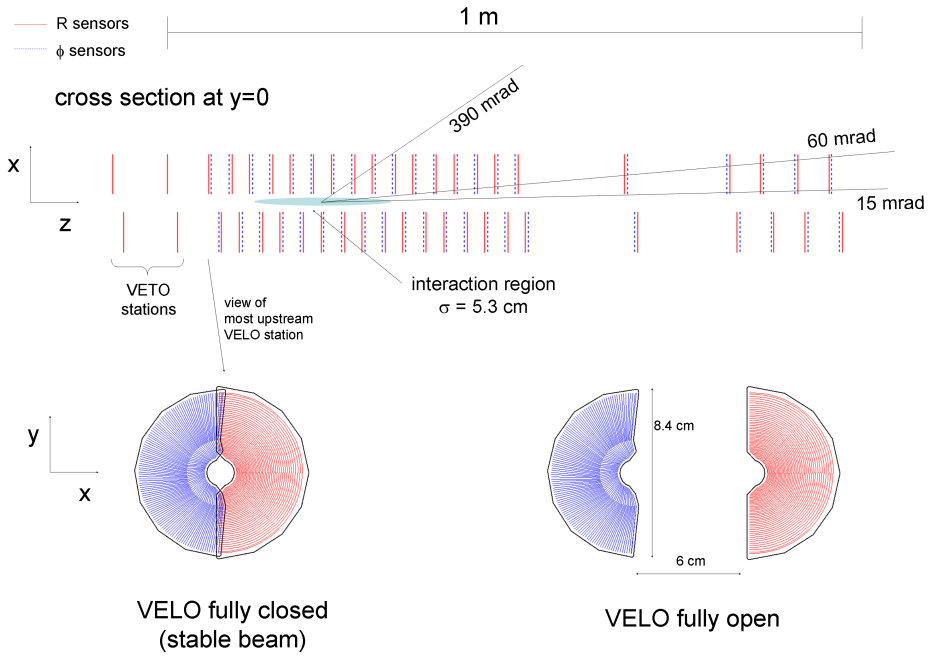


Figure 2.7: Overview of the layout of LHCb VELO modules and sensors. Radial distance, R, sensors are shown in red and azimuthal angle, ϕ, sensors are shown in blue. Reproduced from [44].

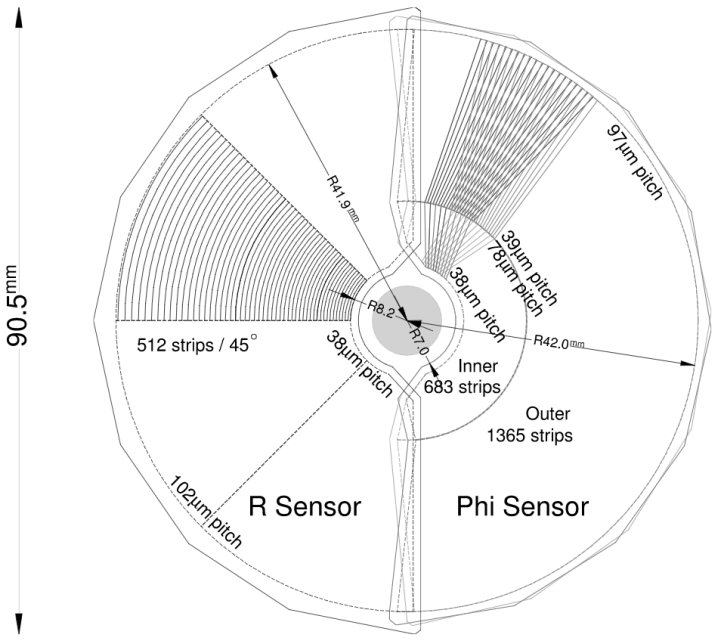


Figure 2.8: An illustration of the strip layouts on VELO R and ϕ sensors. Reproduced from [44].

2.4 Ring Imaging Cherenkov Detectors

A fundamental requirement of LHCb is accurate particle identification (PID). For the analyses presented in this thesis in particular, efficient identification and separation of protons, pions and kaons from B hadron decays is vital. Within LHCb, the PID system consists of two Ring Imaging Cherenkov (RICH) detectors. The RICH detectors provide PID information via the detection of Cherenkov radiation produced by charged particles. Cherenkov radiation is produced when a charged particle travels through a dielectric medium at a velocity greater than the speed of light in that medium. These Cherenkov photons are radiated in a cone at an angle, θ_c, to the particle's trajectory dependent on its velocity, v:

$$\cos\theta_c = \frac{c}{nv}. \tag{2.1}$$

where c is the speed of light in a vacuum and n is the refractive index of the medium. For a track with a given measured momentum, p, the radius of the Cherenkov photon cone differs depending on the species of charged particle. Figure 2.9 shows an example from LHCb of the dependence of θ_c on particle momentum for various particle species.

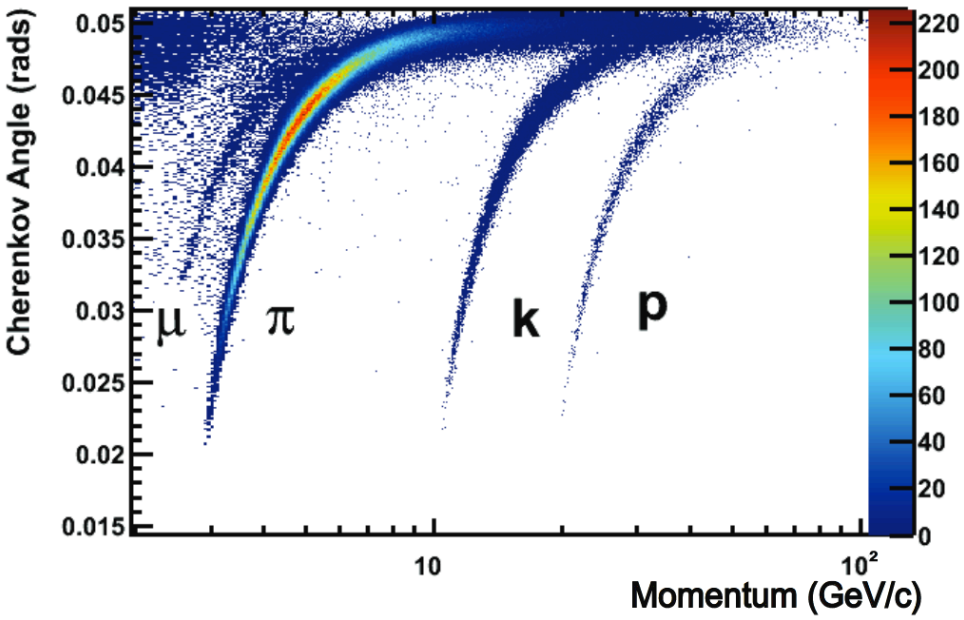


Figure 2.9: Reconstructed Cherenkov angle over a range of particle momenta within RICH 1, reproduced from [49].

2.4.1 Design

The LHCb contains two RICH detectors, RICH 1[50] and RICH 2[51], which perform complimentary roles. A schematic view of both RICH detectors is shown in Figure 2.10. RICH 1 is located immediately downstream of the VELO and covers the full angular acceptance of the detector, 25 mrad to 300 (250) mrad in the horizontal (vertical) plane. It contains two types of radiator material, aerogel ($n \approx 1.03$) and C_4F_{10} gas ($n = 1.0014$) which give it a resulting momentum range of ~ 1 GeV/c to ~ 70 GeV/c. RICH 2 is located after the T3 tracking station, much further downstream than RICH 1, and is designed to perform PID for high momentum tracks. RICH 2 uses CF_4 gas ($n \approx 1.0005$) as a radiator which gives it a momentum range of ~ 15 GeV/c to ~ 100 GeV/c. High momentum tracks within LHCb are predominantly produced at angles close to the beamline. Therefore, RICH 2 covers a smaller angular region, 15 mrad to 100 (120) mrad in the vertical (horizontal) plane, than the full angular acceptance of the entire detector.

Detection of radiated Cherenkov photons in the two RICH detectors is made using an arrangement of mirrors which reflect onto arrays of Hybrid Photon Detectors (HPDs). As shown as a schematic in Figure 2.11, an HPD is a vacuum photon detector in which incident photons interact with a photocathode releasing a photoelectron. This photoelectron is accelerated by a 20 kV applied voltage onto a silicon pixel detector consisting of 8192

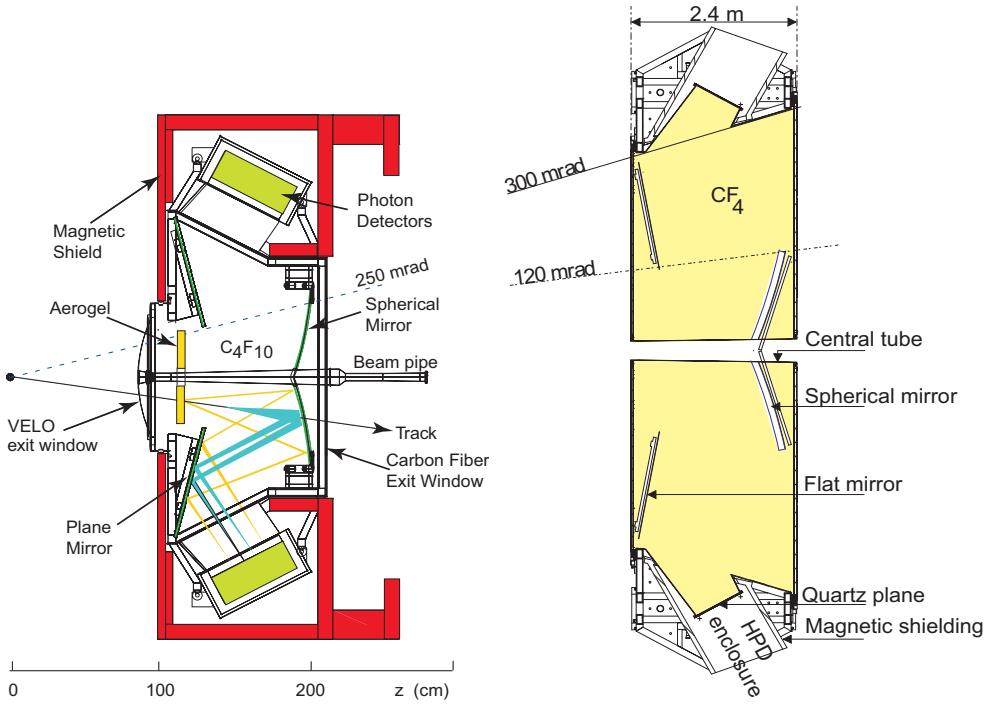


Figure 2.10: Schematic view of the LHCb RICH 1 (left) and RICH 2 (right) detectors. reproduced from [44].

individual pixels that are logically ORed into 1024 logical pixels in a 32×32 configuration. The design of the HPDs allows them to efficiently detect single photons.

2.4.2 Particle Identification and Performance

For a given charged track in the RICH detectors, its PID response is calculated under five different mass hypotheses (e, μ, π, K, p). Under each mass hypothesis the corresponding θ_c value is calculated from the track's momentum, as measured by the tracking system. The expected radius of the Cherenkov photon ring corresponding to θ_c is then compared to a fit of the detector responses of the RICH detectors simultaneously. Under each mass hypothesis the log-likelihood of the RICH detector response fit compared to the expected photon ring radius is calculated. As pions are by far the most abundantly produced particle within LHCb, the PID variables returned by the PID algorithm are the difference in log-likelihood (DLL) between the specific mass hypothesis (K, p, e, μ) and the pion hypothesis.

This method does leave the possibility for misidentifications of particles to occur. To test the efficiencies and misidentification rates of the PID algorithm, PID selections are

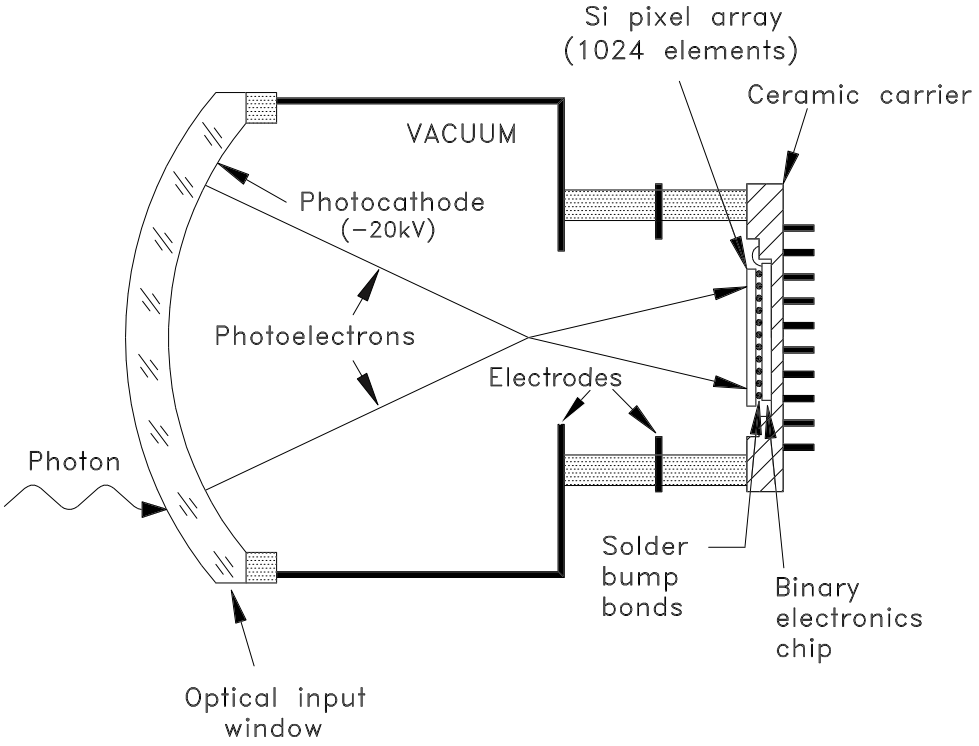


Figure 2.11: Schematic view of a Hybrid Photon Detector (HPD). Reproduced from [44].

applied to high statistics samples of kaons, pions and protons from decays which can be selected without the use of RICH information. Figure 2.12 shows the efficiencies, across a range of momenta, of correctly identifying a kaon as a kaon and incorrectly identifying a pion as a kaon for a loose cut of $\mathrm{DLL}_{\mathrm{K}\pi} > 0$ and a tighter cut of $\mathrm{DLL}_{\mathrm{K}\pi} > 5$. The figure shows that, as expected, correct PID performance decreases at high momentum which reflects what is shown in Figure 2.9, that as momentum increases, θ_c tends towards the same value for each mass hypothesis. However, overall the PID performance is shown to be excellent, with $> 90\%$ correct PID rates and $< 10\%$ mis-ID rates for kaons, pions and protons across a wide range of momenta.

2.5 Magnet

The LHCb detector apparatus contains a dipole magnet [52], which is used in the measurement of the momenta of charged particles. A perspective view of the magnet is shown in Figure 2.13. Rather than being super-conducting, it is a warm magnet design consisting of two identical saddle-shaped coils of Aluminium Al-99.7 (225 turns each) symmetri-

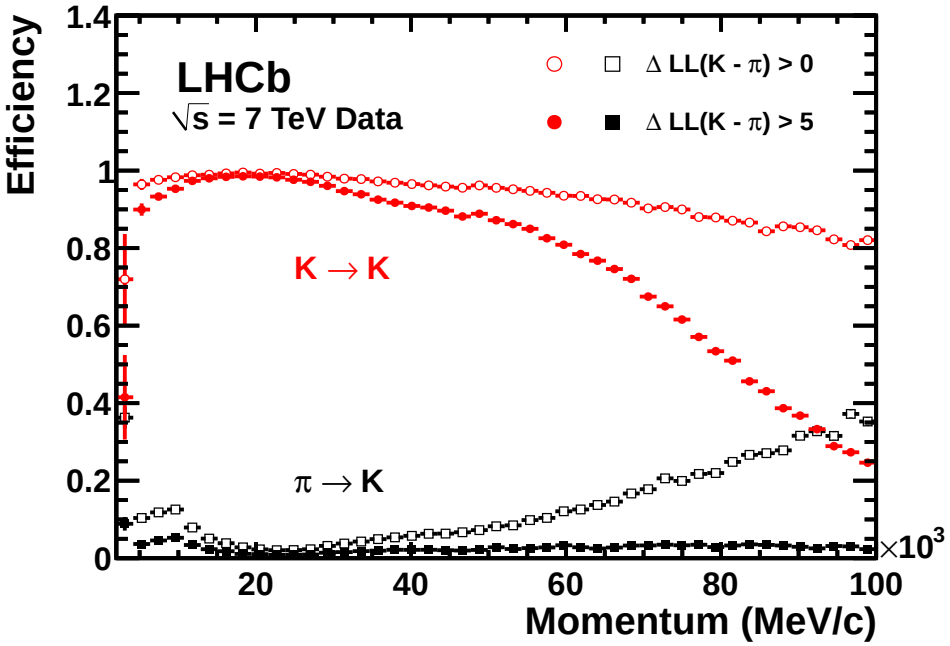


Figure 2.12: Efficiency, as a function of track momentum, of correctly identifying a kaon as a kaon and incorrectly identifying a pion as a kaon within the RICH, using 2011 data. Reproduced from [49].

cally positioned above and below the beam axis and within a 1450 ton low-carbon steel window-frame yoke. The magnet is situated between the two RICH detectors with the centre line of the two coils lying at a distance of $z = 5.3$ m from the interaction point.

At full power, the magnet produces a magnetic field integral $\int Bdl$ of 4 Tm for tracks 10 m in length with the principle field component along the y axis. The polarity of the magnet is reversed regularly as a measure to reduce systematic uncertainties in charged particle tracking.

The desired momentum resolution of charged particles within the detector requires that the value of the magnetic field integral be known to a precision $\mathcal{O}(10^{-4})$. A measurement of the magnetic field across the detector was performed using an array of Hall probes. Measuring over a fine grid of $80 \times 80 \times 100$ mm^3, the magnetic field was scanned from the interaction point to the RICH 2 detector. By comparing the measured values of the field strength for different Hall probes at the same location, the precision of the measurement was obtained and found to be approximately 4×10^{-4} across the measured volume. The results of this measurement were compared to a computer simulation of the detector. Over the majority of the detector the difference between measurement and computer simulation

was less than 1%, with an area upstream of the magnet registering a difference of 3.5% due to the precision of the simulation and the nearby presence of iron reinforcements within the structure of the cavern [44].

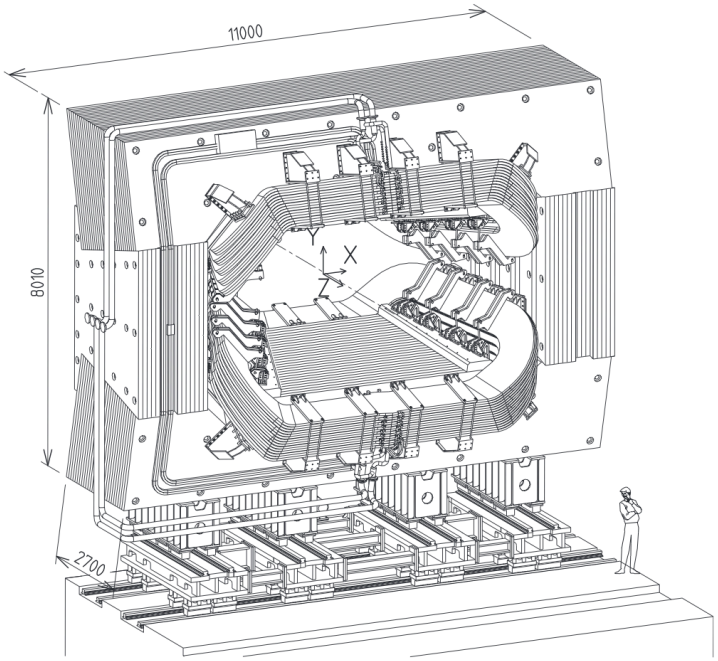


Figure 2.13: Perspective view of the LHCb dipole magnet, reproduced from [44].

2.6 Tracking System

In addition to the VELO, the tracking system of the LHCb experiment consists of two further elements, the Silicon Tracker (ST) and the Outer Tracker (OT). These elements are combined within four tracking stations placed between the RICH 1 and RICH 2 detectors, one upstream of the dipole magnet (TT) and three downstream (T1-3), each covering an area from the beam pipe to the edge of the acceptance, in x and y, of the detector. The ST and OT both perform the same function, measuring the trajectories of charged particles such that their momenta can be calculated, however they differ in their constructions and location. Particle flux is very high close to the beam pipe and so a high level of radiation hardness is required for any detector components in this area. The tracking system also requires high granularity to keep hit occupancy at a low enough level to allow adequate pattern recognition. It is in this region that the Silicon Tracker is located, the components of which are described in Sec. 2.6.1. Further out, radially, from

the beam pipe the OT provides the remaining coverage for the tracking stations T1 to T3. The particle flux in this region is significantly lower compared to that covered by the ST and, as such, 'straw tube' drift-time sensors can be used which provide a benefit of much reduced cost compared to the ST components. Details of the Outer Tracker components are given in Sec. 2.6.2.

2.6.1 The Silicon Tracker

The Silicon Tracker comprises one tracking station upstream of the dipole magnet, known as the Tracker Turicensis (TT) [44, 53], and the inner parts of the three stations (T1-3) downstream between the magnet and the RICH 2 detector, collectively known as the Inner Tracker (IT).

The TT is a planar tracking station with an area 150 $\times$ 130 cm that covers the full acceptance of the experiment. A schematic view of the TT is shown in Figure 2.14. Within the TT there are four planar detector layers in an arrangement where the outer layers are aligned vertically and the the two central layers are aligned at an angle of -5° and $+5^\circ$ to the vertical respectively. Having the central two layers aligned at an angle removes hit placement ambiguities and allows reconstruction of the hit placement in three dimensions with a single hit spatial resolution of 50 μm. The individual silicon sensors used in the TT are single sided p^+-on-n sensors each measuring 9.64 cm in width and 9.44 cm in length. Each sensor carries 512 readout strips with a pitch of 183 μm.

The IT stations consist of four layers configured in a similar $(x-u-v-x)$ layout to the TT, again with a $\pm 5^\circ$ stereo angle rotation in the alignment of the u and v layers. Each IT layer is arranged in a cross shape measuring 120 cm wide and 40 cm high as shown in Figure 2.15. The silicon detectors used in the IT have a strip pitch of 193 μm and, overall, the IT has a spatial resolution of approximately 50 μm.

2.6.2 The Outer Tracker

The OT consists of an array of straw tube drift-time sensor modules. The OT sensor modules surround the three tracking stations of the IT, as shown in Figure 2.16. Each drift tube has an internal diameter of 4.9 mm and contains a mixture of 70% Argon and 30% CO_2. This gas mixture provides a drift time of < 50 ns and a drift coordinate resolution of 200 μm. Each OT module is comprised of a double layer of drift tubes, 256 straws in total, and each OT station is comprised of four OT modules arranged in the same $(x-u-v-x)$ configuration as the TT and IT. The complete OT system consists

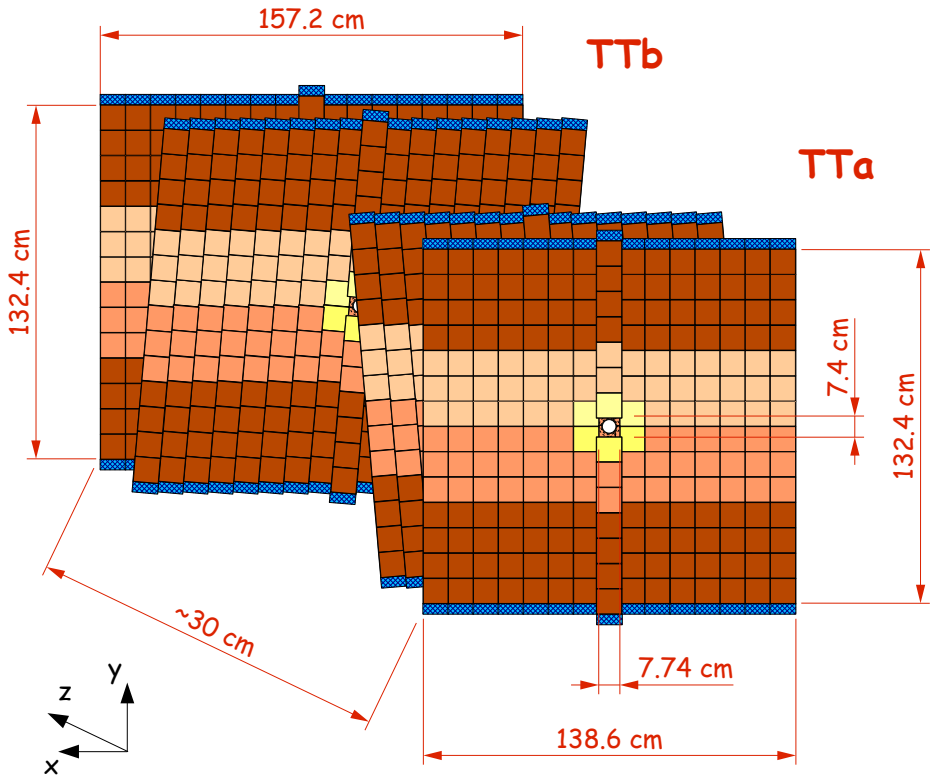


Figure 2.14: Schematic view of the LHCb Tracker Turicensis, reproduced from [54].

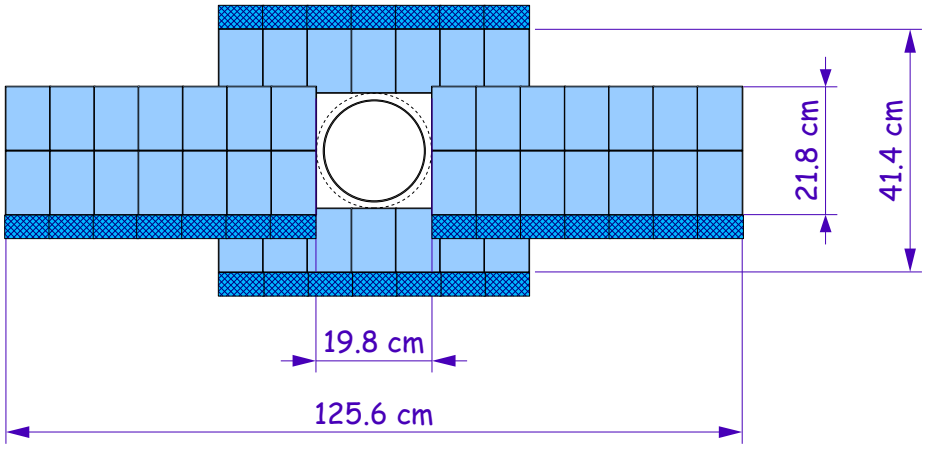


Figure 2.15: Layout of LHCb Inner Tracker x-layer, reproduced from [54].

of approximately 55000 individual, single straw-tube channels. Each OT station covers a total active area of 5971 x 4850 mm^2 from the outer edges of the IT to the the edge of the detector acceptance, thereby covering an angular acceptance of up to 300 mm (250 mm) in the bending (non-bending) plane.

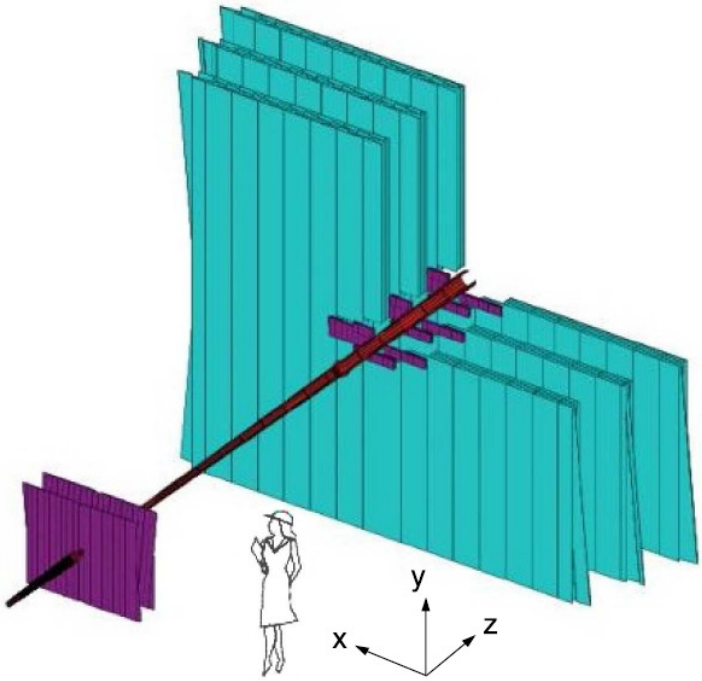


Figure 2.16: Perspective view of the LHCb tracking system with the Silicon Tracker, shown in purple, and the Outer Tracker, shown in blue, reproduced from [44].

2.6.3 Performance of the Tracking System

Measurements of key tracking system performance parameters were made throughout Run I of the LHC. The hit efficiency of a silicon sensor is defined as the ratio between the number of hits found and the number of hits expected for a given sector, and is an important descriptor of tracking performance [46]. A similarly important performance measurement is the detector hit resolution, which is determined from the unbiased residuals between the measured position of a track hit and its expected position. Track hit efficiencies and resolutions in the TT and IT were measured in 2011 and 2012. The hit efficiency measurements are made using daughter tracks from clean samples of $J/\psi \rightarrow \mu^+\mu^-$ decays. These tracks are required to have momentum greater than 10 GeV/c and also have

additional track quality cuts applied. Table 2.2 shows a summary of the TT and IT hit efficiency and resolution measurements for 2011 and 2012 data. Also shown are the corresponding expected hit efficiencies and resolutions from simulation. The measured hit efficiencies across all conditions are greater than 99% with excellent agreement between data and simulation. The measured hit resolutions are all between 50 μm and 55 μm and agree with the results from simulation.

Detector	Measurement	2011 Data	2012 Data	2011 MC	2012 MC
TT	Hit Efficiency	99.7%	99.8%	99.9%	99.9%
	Hit Resolution	52.6 μm	53.4 μm	47.8 μm	48.0 μm
IT	Hit Efficiency	99.8%	99.9%	99.9%	99.9%
	Hit Resolution	50.3 μm	54.9 μm	53.8 μm	53.9 μm

Table 2.2: Measured TT and IT hit efficiencies and resolutions for 2011 and 2012 data and MC simulation. Reproduced from [46].

The hit efficiency in the OT modules was measured as function of horizontal module distance from the beam pipe. For hits within $|r| < 1.25$ mm, where r is the radius of the straw tube drift-time sensor, the single hit efficiency is measured to be greater than 99% for all but the outermost modules, on either side, where the efficiency is 98% [55].

The single hit resolution of the OT is measured using "good quality" tracks, which are required to have momentum greater than 10 GeV/c, at least 16 hits in the OT and a track-fit $\chi^2/nDoF < 2$ (calculated excluding the hit under study). Following a similar method to the IT and TT hit resolution measurements, the measured track drift time and hit position are compared to the predicted values for the track. From fits to the resulting drift time and hit position distributions, the drift time and hit resolutions are found to be approximately 3 ns and 200 μm, respectively, which are consistent with the design values for the detector [55].

Excellent momentum and mass resolutions of the detector are key requirements for the LHCb physics program. The detector momentum resolution, δ_p, in data is measured using long tracks from $J/\psi \rightarrow \mu^+$ μm decays. Figure 2.17 shows the relative momentum resolution, δ_p/p, as a function of momentum, p, for tracks with momenta up to 300 GeV/c. The momentum resolution ranges from approximately 0.5 % at 20 GeV/c to 0.8 % for tracks at 100 GeV/c.

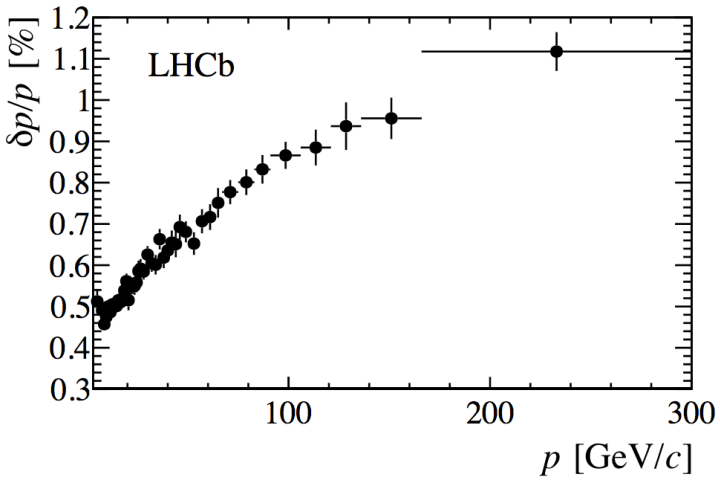


Figure 2.17: Measured relative momentum resolution, δ_p/p, versus track momentum, p, for data tracks from J/ψ decays, reproduced from [46].

The mass resolution, σ_m, of a dimuon resonance is directly related to the momentum resolutions of the two daughter muons. The mass resolution of the detector is measured as a function of dimuon resonance invariant mass by comparing the mass resolutions of six different dimuon resonances: J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and Z^0. For each resonance, the invariant mass distribution is fitted with a double Gaussian function and σ_m is calculated as the root mean square of the double Gaussian. Figure 2.18 shows the measured relative mass resolution, σ_m/m, versus invariant mass for the six studied resonances. The relative mass resolution is shown to be pretty constant at approximately 0.5% up to the Υ masses (~ 10 GeV/c^2).

2.7 Calorimeters

Calorimetry within the LHCb detector is primarily split between two subdetector systems: the Electromagnetic CALorimeter (ECAL) and the Hadronic CALorimeter (HCAL). The ECAL is located immediately after the first muon tracking station and provides energy and position measurements for electrons and photons. The HCAL is positioned downstream of the ECAL and provides similar position and energy measurements for hadrons. In addition to energy and position measurements, calorimeter information is also used for photon, electron and hadron particle identification and as part of the first level trigger (L0) to select high transverse energy photon, electron and hadron candidates.

In front of the ECAL sit two detector systems, a scintillator pad detector (SPD) fol-

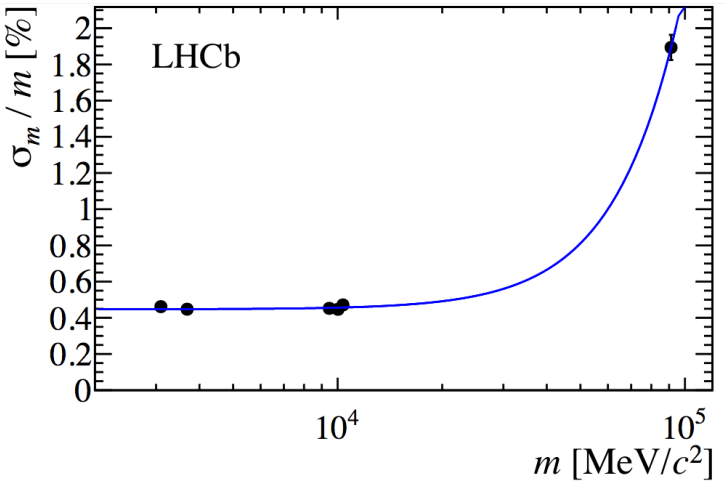


Figure 2.18: Measured relative mass resolution, σ_m/m, as a function of dimuon resonance invariant mass, m, for six different dimuon resonances. Reproduced from [46].

lowed by a preshower detector (PS). The SPD identifies charged particles entering the ECAL and helps with the rejection of π^0 backgrounds, while the PS suppresses contamination from charged pion backgrounds. The SPD and PS detectors are both made up of a layer of lead convertor 2.5 radiation lengths thick (15mm) sandwiched between two scintillator pads that read out to a 64 channel Multi-Anode Photo-Multiplier (MaPMT) via individual wavelength-shifting (WLS) fibres. The two detectors are almost identical except the SPD detector dimensions are $\approx 0.45\%$ smaller than those of the PS. Hit densities across the SPD/PS/ECAL vary by two orders of magnitude, therefore these modules are segmented into three sections, as shown in Figure 2.19.

Energy deposits in the ECAL are made by electromagnetic showers from Bremsstrahlung or pair production of particles due to interactions between incoming electrons and photons and the material of the calorimeter. The detector configuration of the ECAL is made up of alternating layers of 2 mm thick Lead, 120 μm reflecting TYVEK paper and 4 mm thick scintillator tiles along the Z direction. The energy resolution requirements of the ECAL demand that showers from high energy photons are fully contained within the detector, which leads to the full thickness of the ECAL covering 25 radiation lengths with a Moliere radius of 3.5 cm. The ECAL design achieves an energy resolution of $\frac{8\%}{\sqrt{E(\text{GeV})}} \bigoplus 0.8\ \%$.

The HCAL comprises of alternating layers of iron absorber material and scintillating tiles aligned parallel to the beam axis, as shown in Figure 2.20. The total height and width of the HCAL covers an area 6.8 m wide by 8.4 m high. The length of each iron

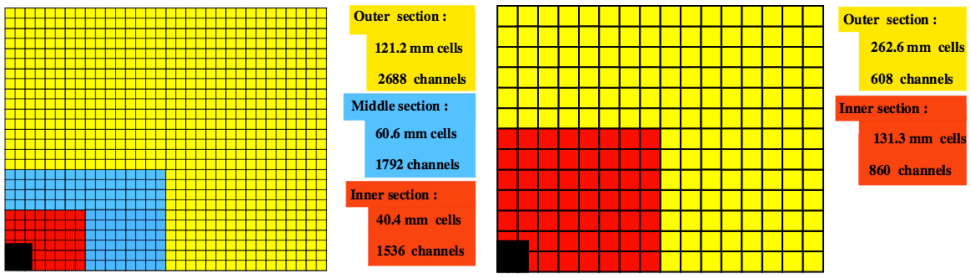


Figure 2.19: Diagrams showing the segmentations of the LHCb SPD/PS/ECAL (left) and HCAL (right), reproduced from [44].

layer, in the z direction, is equal to the interaction length of hadrons λ_I in steel. Within each scintillator layer are three separate scintillator tiles spread out along the length of the HCAL, each reading out via WLS fibres to PMTs. In a similar fashion to the SPD, PS and ECAL, the granularity of the HCAL modules varies radially from the beam line, with the HCAL modules divided into two segments, as shown in Figure 2.19. The energy resolution achieved in the HCAL is $\frac{69\%}{\sqrt{E(\text{GeV})}} \bigoplus 9\ \%$.

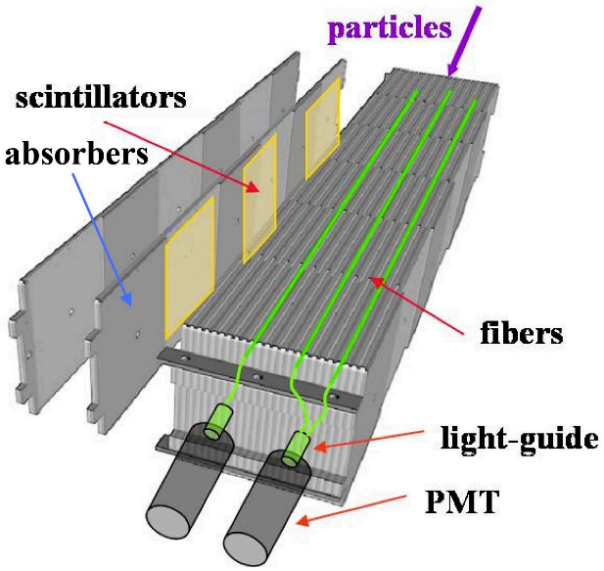


Figure 2.20: LHCb HCAL module layout, reproduced from [44].

2.7.1 Calorimeter Performance

Calorimeter performance was monitored and calibrated throughout Run I. The ECAL and HCAL are each equipped with an array of light emitting diodes (LEDs) which are used to monitor the performance of the calorimeter PMTs. The PMT response to the

LED signals was recorded at various times throughout the Run I data-taking period. The responses of the calorimeters to the LED signals were found to have good agreement with the calorimeter responses to real particles. Further ECAL monitoring is performed using invariant mass distributions of $\pi^0 \to \gamma\gamma$ decays, which are used correct the energy-scales of the individual ECAL detector cells. This improved the ECAL $\pi^0 \to \gamma\gamma$ mass resolution from 8-10% to approximately 2%.

High precision HCAL calibration was performed outside of data-taking periods using two samples (one per detector half) of 10 mCi radioactive caesium-137, ^{137}Cs. These samples were moved through the HCAL with the resulting currents in the PMTs measured. Comparing these currents to values obtained during test beam studies allowed cell-to-cell calibrations to be performed resulting in an intercalibration level of 2-3% [56].

Over the course of Run I, the performances of the ECAL and HCAL were observed to be affected by ageing effects. Radiation damage was observed to affect scintillator performance with effects proportional to the experienced particle flux. Additionally, a decrease in PMT gains was observed due to the degradation of the dynode system when subjected to high integrated anode currents over the course of 2011. Detailed studies were performed of the HCAL light yield degradation as a function of delivered luminosity during the ^{137}Cs HCAL calibration procedures. Figure 2.21 shows the average relative decrease in light yield for the 44 most central HCAL cells as a function of delivered luminosity across 2011 and 2012 showing the increase in degradation over the course of Run I data taking.

The particle reconstruction and identification performance of the calorimeter system was also carefully measured. Electron identification performance of the ECAL was evaluated using a tag-and-probe method on samples of 2011 $J/\psi \to e^+e^-$ data. The average electron ID performance was found to be $(91.9 \pm 1.3)\%$ with a misidentifiaction rate of $(4.54 \pm 0.02)\%$ [46]. High-energy photon reconstruction performance was measured using invariant mass fits of selected $B^0 \to K^{*0}\,(K^+\pi^-)\,\gamma$ candidates and the ECAL photon mass resolution was found to be 93 MeV/c^2 [57]. The ECAL photon and neutral pion identification efficiencies were also measured using $B^0 \to K^{*0}\gamma$ decays. Using a neural network classifier, it was possible to select photons with an efficiency of 95% while rejecting 45% of π^0 tracks incorrectly reconstructed as photons. Neutral pion reconstruction performance was measured separately for low and high transverse momentum, p_T, candidates, where 2 GeV/c is considered the cut off between low and high p_T. Below 2 GeV/c, the π^0 decays into two well separated photons which can be resolved in the ECAL to give a mass resolution of 8 MeV/c^2. However, above 2 GeV/c, the two photons are not well separated in

the detector and cannot be resolved as individual clusters, thus negatively affecting the mass resolution. A method was developed to identify overlapping, or "merged", ECAL clusters resulting in a mass resolution of approximately 30 MeV/c^2 for merged, high p_T, π^0 candidates.

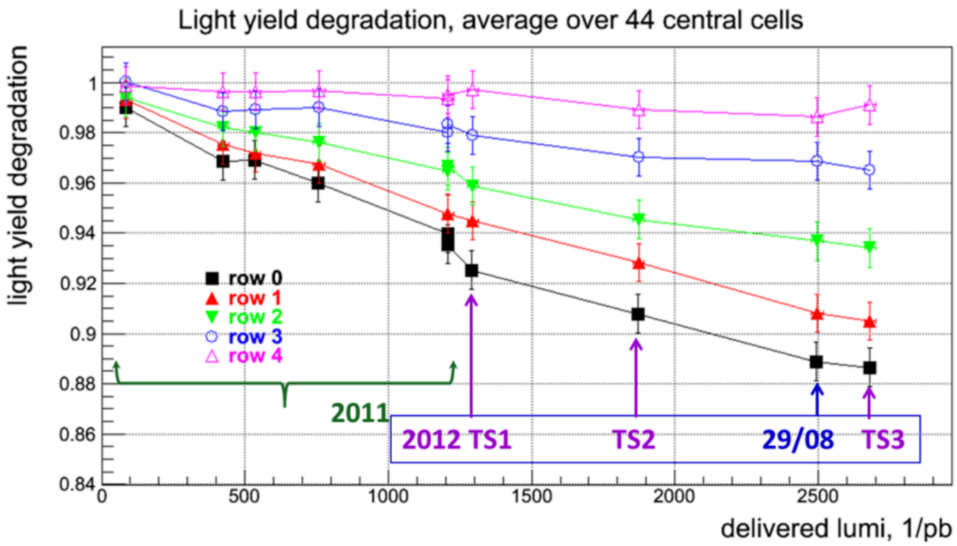


Figure 2.21: Relative decrease in light yield of the LHCb HCAL module as a function of the delivered luminosity for five different layers of scintillator tiles. Reproduced from [56].

2.8 Muon System

Due to their much larger mass, muons emit far less energy through Bremsstrahlung than electrons and can easily penetrate through the length of the LHCb detector, depositing only a fraction of their total energy within the calorimeter system. As decays involving muons contribute greatly to the core physics programme of the LHCb experiment, an accurate muon detection system is a vital component of the detector.

The LHCb muon tracking system [44, 58] is designed to detect any muons passing through the detector and beyond the calorimeter. The muon system is made up of five stations (M1-M5) with M1 positioned immediately before the SPD and PS modules and M2-M5 located after the HCAL. The muon stations increase in size as they get further from the interaction point in order to maintain an angular acceptance of 20 (16) mrad to 306 (258) mrad in the bending (non-bending) plane respectively. The M2-M5 stations are separated by 80 cm thick iron absorber layers to select penetrating muons. This

results in only muons with momenta of at least 6 GeV/c reaching the M5 station. The position of the M1 station is chosen to improve the measurement of muon p_T for the muon-specific trigger. The sensors in the muon stations are all multiwire proportional chambers (MWPCs) with the exception of the M1 station, which being much closer to the interaction point, experiences a much greater particle flux than M2-M5 and so requires greater radiation hardness and thus uses triple-Gas Electron Multiplier (GEM) detectors in its innermost region. The M1-M3 stations have high spatial resolutions along the bending plane and are used to measure the muon candidate track direction and p_T with a resolution of 20%. The M4 and M5 stations have lower spatial resolutions and are used to confirm the detection of high momentum muons.

Similarly to the calorimeter systems, the muon stations are segmented into regions with different granularities moving outwards from the beam pipe in the $x-y$ plane. Each muon station is divided into four regions (R1-R4) with the dimensions of the regions scaled to the ratio 1:2:4:8 with particle flux and channel occupancy expected to be equal across each region. A schematic view of a quadrant of M1 muon station is shown in Figure 2.22.

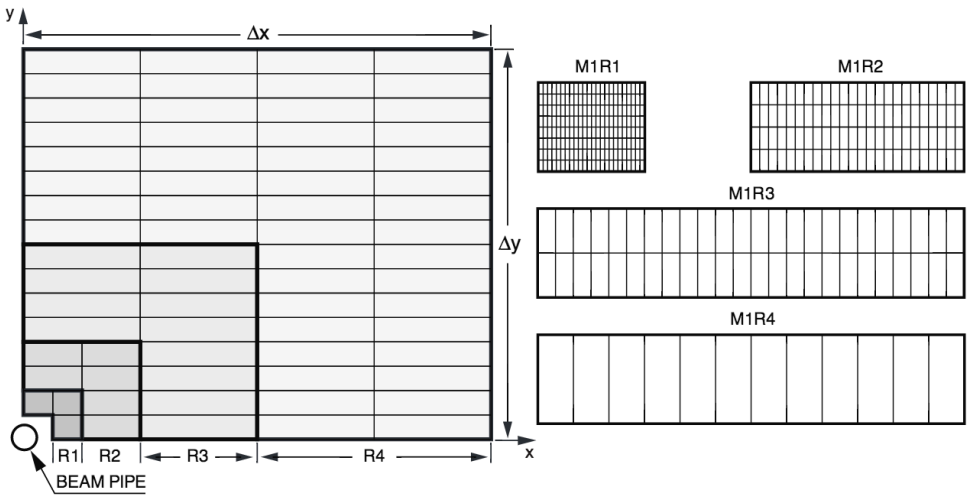


Figure 2.22: LHCb M1 muon station quadrant layout (left), showing the segmentation of the detector into four regions. Each rectangle represents one chamber with example segmentations of each chamber shown (right). Reproduced from [44].

2.8.1 Muon System Performance

The performance of LHCb muon identification was measured using samples of muons (from $J/\psi \to \mu^+\mu^-$ decays), protons (from $\Lambda^0 \to p\pi^-$) and kaons and pions (both from $D^{*+} \to \pi^+ D^0 (\to K^-\pi^+)$ from 2011 data. For each particle sample, muon selection

(mis)identification efficiencies were measured both as a function of momentum and PID selection cut value. Average muon identification efficiencies of 98% were obtained with pion, kaon and proton misidentification rates less than 1% at high p_T [59].

2.9 Trigger System

The LHC operates with a maximum bunch collision frequency of 40 MHz; however, the rate of bunch crossings containing interactions of interest to the LHCb detector is approximately 10 MHz. An event rate of 10 MHz would be unmanageable for the detector and storage systems to process. It is the job of the trigger system to apply judicious selection criteria to reduce the event rate down to a manageable storage rate of 5 kHz. The trigger system [60] is split into three levels, an intital, level-0 (L0), hardware trigger running synchronously with the LHC bunch crossing frequency, and two software High Level Trigger systems (HLT1, HLT2). The trigger systems run consecutively with HLT1 only processing events which pass the L0 trigger and likewise HLT2 only processes events passing HLT1. Only events passing all three trigger stages are sent to permanent storage, all other events are discarded.

2.9.1 Level 0 trigger

The L0 trigger system is required to reduce the 10 MHz visible event rate down to a rate of 1 MHz at which the entire detector can be read out. Hardware information gathered from the VELO, calorimeter and muon systems is fed through to a custom built level-0 Decision Unit (L0DU) which evaluates the final L0 decision within a maximum time of 1 μs.

An L0 trigger decision is made having considered three separate criteria:

- Charged Particle Energy: Due to their large masses, heavy flavour baryons and mesons typically decay into particles with high transverse energies and momenta. The L0 trigger uses information from the calorimeters to identify high E_T or p_T hadrons, electrons and photons and information from the muon system to identify pairs of high p_T muons.

- Pile up: Information from the VELO veto systems is used to reject events with more than one primary interaction.

- Track Multiplicity: SPD and VELO pile-up system information is used to estimate the number of charged tracks within an event and reject events with large track multiplicities, since B decay products become increasingly difficult to reconstruct as the number of tracks in the event increases.

Events with an electron, photon or hadron with sufficient E_T or pair of muons with p_T above the required threshold and low enough track multiplicities and event pile up are accepted by the L0DU and passed on to the HLT1.

2.9.2 High Level Trigger

Events passing the L0 trigger have their full detector response information sent through to the HLT1 at an output rate of 1 MHz. The HLT1 provides a software decision using about two-thousand computing nodes contained within the Event Filter Farm (EFF). The main focus of the HLT1 is to reconstruct specific charged particles from VELO and Tracker station information or confirm the absence of charged particles in the case of π^0 or photon decisions. Events meeting one or more HLT1 trigger requirements are then fed on to the HLT2. The HLT1 system reduces the 1 MHz output rate of the L0 trigger down to approximately 30 kHz.

The HLT2 software trigger takes the particle decisions from HLT1 and performs a full event reconstruction, which has been made possible by the reduction in rate down to 30 kHz by the L0 and HLT1 decisions. The HLT2 trigger stage consists of a range of independent selection algorithms designed to select both inclusive decays, which require only a partially reconstructed mother particle e.g. $B \to h^+h^-$, and exclusive decays where a fully reconstructed mother particle is required e.g. $B^0 \to K^+\pi^-$. The final HLT2 trigger is the logical OR of the inclusive and exclusive selections which further reduces the retention rate down to the manageable rate of 5 kHz at which data can be written to permanent storage.

2.10 Data Processing and Simulation

To process the raw data from the detectors into useable formats for analyses, LHCb uses a series of software packages. To perform the offline reconstruction of triggered events, the BRUNEL [61] software package is used. This takes information from the individual subdetector system outputs to reconstruct the tracks and vertices in the event and then assign e, μ, π, K, p PID likelihoods for each track. The fully reconstructed tracks from

the BRUNEL output can then be combined using the DAVINCI [62] software package to perform searches for specific decays. The DAVINCI software can combine tracks under specific hypotheses that they were created by the daughter particles with a shared mother.

The full dataset of triggered and reconstructed events is far too large for analyses to attempt to use in its entirety. In order to streamline the available data to provide analysts with only the subset of the full data relevant to their specific searches, a set of selection algorithms are run through DAVINCI on the reconstructed events in a process known as stripping. These stripping algorithms reconstruct the events under a specific decay hypothesis and apply selection cuts to reinforce the decisions made by the trigger systems and remove background events from the specific areas of interest. The DAVINCI software can then be used to select these stripped events and provide data samples for analysts to use containing any variables or parameters they wish to implement or investigate.

Chapter 3

The Search for the Decays $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ with 2011 Data

3.1 Introduction

In this chapter, a search for the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ rare decay modes at LHCb is presented. The work in this chapter was carried out by the author in a collaboration between the Universities of Glasgow and Manchester. This work was performed at the beginning of the authors PhD to try and find first evidence for the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ decay channels. The work in this first analysis was collaborative and published in Reference [1]. The author contributed to the signal selection, background determination, mass fits and confidence level calculations. The full Run 1 search for $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ decays using data from the LHCb 2011 and 2012 data-taking runs, which is the main topic of this thesis, was led by the author and will be documented in Chapters 4 to 6.

3.2 Analysis Strategy

The analysis strategy largely follows the ideas developed for the first LHCb sensitivity studies performed in 2009 [63]. In short, the strategy is based on a relative branching fraction measurement measured with respect to the branching fraction of the $B^0 \to K^+\pi^-$ decay mode. This method is chosen in order to cancel or minimise systematic effects. Uncertainties in the b-quark cross-section, b hadronisation and luminosity cancel. Also, by choosing a normalisation channel topologically identical to the signal channel, the ratios

of efficiencies entering the branching fraction calculation cancel to a large extent, the differences in efficiencies between signal and normalisation channel being due to small differences in decay kinematics and the different daughter particle types. The inclusion of charge-conjugate processes is implied throughout.

In general, the number of events triggered and selected can be calculated as

$$N = \int \mathcal{L}\,\mathrm{d}t \cdot \sigma_{b\bar{b}} \cdot 2 \cdot f_{d,s} \cdot \mathcal{B}_{\mathrm{vis}} \cdot \epsilon_{\mathrm{tot}} \quad , \tag{3.1}$$

where $\int \mathcal{L}\,\mathrm{d}t$ is the integrated luminosity, $\sigma_{b\bar{b}}$ is the $b\bar{b}$ cross section, the factor 2 accounts for both the b and the $\bar{b}$, $f_{d,s}$ represents the b hadronisation probability to the relevant hadron, $\mathcal{B}_{\mathrm{vis}}$ stands for the visible branching fraction and ϵ_{tot} contains the product of all efficiencies for the signal daughters to be in the LHCb acceptance, for triggering, reconstruction, stripping and final selection

Of the possible normalisation channels, the two-body charmless modes are a natural choice. This analysis opted for $B^0 \to K^+\pi^-$ as this mode provides the largest statistics of any hadronic two-body decay and its branching fraction is the most precisely known, $\mathcal{B}\left(B^0 \to K^+\pi^-\right) = 19.55^{+0.54}_{-0.53} \times 10^{-6}$ [64]. Both of the features are optimal as far as this method is concerned, see Eq. 3.2 below.

Using the $B^0 \to K^+\pi^-$ normalisation channel, the $B^0 \to p\bar{p}$ branching fraction can be extracted from

$$\mathcal{B}(B^0 \to p\bar{p}) = \frac{N(B^0 \to p\bar{p})}{N(B^0 \to K^+\pi^-)} \cdot \frac{\epsilon_{B^0 \to K^+\pi^-}}{\epsilon_{B^0 \to p\bar{p}}} \cdot \mathcal{B}(B^0 \to K^+\pi^-) \quad . \tag{3.2}$$

For $B^0_s \to p\bar{p}$ an extra factor f_d/f_s appears on the righthand side of the equation above.

The analysis is done in a blind approach, ignoring the $B^0 \to p\bar{p}$ and $B^0_s \to p\bar{p}$ signal regions throughout the whole process; the definition of the signal region is detailed later in this chapter. The selection optimisation takes as a starting point signal Monte Carlo (MC) events passing the dedicated stripping line as well as 10% of the full 2011 sidebands data sample. To avoid introducing a potential bias to the results, the 10% data sample used for selection optimisation is omitted from the final selection. In other words only 90% of the full 2011 data sample, *i.e.* 0.92 fb^{-1}, is actually used to perform the search. The $B^0 \to K^+\pi^-$ selection follows that of the LHCb $B \to h^+h^-$ lifetime analysis [21], with trivial changes in particle identification requirements.

3.3 Candidate selection

The selection requirements of both signal modes and the normalisation channel exploit the characteristic topology of two-body decays and their kinematics. All daughter tracks tend to have larger transverse momenta, p_T, compared to generic tracks from light-quark background owing to the high B mass, therefore a minimum p_T requirement is imposed for all daughter candidates. Furthermore, the two daughters form a secondary vertex (SV) displaced from the primary vertex (PV) due to the relatively long B lifetime. The reconstructed B momentum vector points to its production vertex, the PV, which results in the B meson having a small impact parameter (IP) with respect to the PV. This is in contrast with the daughters, which tend to have a large IP with respect to the PV as they originate from the SV, therefore a minimum $\chi^2_{\rm IP}$ with respect to the PVs is imposed on the daughters. The condition that the B candidate comes from the PV is further reinforced by requiring that the angle between the B candidate momentum vector and the line joining the associated PV and the B decay vertex (B direction angle) is close to zero.

To avoid potential biases, $p\overline{p}$ candidates with invariant mass within $\pm 50\,\mathrm{MeV}/c^2$ ($\approx 3\sigma$) around the known B^0 and B^0_s masses, specifically the region $[5230, 5417]\,\mathrm{MeV}/c^2$, are not examined until all analysis choices are finalised. The final selection of $p\overline{p}$ candidates relies on a boosted decision tree (BDT) algorithm [65] as a multivariate classifier to separate signal from background. Additional preselection criteria are applied prior to the BDT training.

The BDT is trained with simulated signal samples and data from the sidebands of the $p\overline{p}$ mass distribution as background. Of the $1.0\,\mathrm{fb}^{-1}$ of data recorded in 2011, 10% of the sample is randomly selected and exploited for the training of the $B^0_{(s)} \to p\overline{p}$ selection, and 90% for the actual search. The BDT training relies on an accurate description of the distributions of the selection variables in simulated events. The agreement between simulation and data is checked on the $B^0 \to K^+\pi^-$ proxy decay with distributions obtained from data using the *sPlot* technique [66] which provides a method to unfold the various signal and background components from a data sample via a maximum likelihood fit. No significant deviations are found, giving confidence that the inputs to the BDT yield a nearly optimal selection. The variables used in the BDT classifier are properties of the B candidate and of the B daughters, *i.e.* the proton and the antiproton. The B candidate variables are: the χ^2 per number of degrees of freedom of the measured decay vertex; the decay vertex $\chi^2_{\rm IP}$; the direction angle; the distance in z (the direction of the interacting

proton beams) between its decay vertex and the related PV; and the p_T asymmetry within a cone around the B direction defined by $A_{p_T} = ({p_T}^B - {p_T}^{\text{cone}})/({p_T}^B + {p_T}^{\text{cone}})$, with ${p_T}^{\text{cone}}$ being the p_T of the vector sum of the momenta of all tracks measured within the cone radius $R = 0.6$ around the B direction, except for the B-daughter particles. The cone radius is defined in pseudorapidity and azimuthal angle (η, ϕ) as $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The BDT selection variables on the daughters are: their distance of closest approach; the minimum of their p_T; the sum of their p_T; the minimum of their χ^2_{IP}; the maximum of their χ^2_{IP}; and the minimum of their cone multiplicities within the cone of radius $R = 0.6$ around them, the daughter cone multiplicity being calculated as the number of charged particles within the cone around each B daughter.

The cone-related discriminators are motivated as isolation variables. The cone multiplicity requirement ensures that the B daughters are reasonably isolated in space. The A_{p_T} requirement further exploits the isolation of signal daughters in comparison to random combinations of particles.

The figure of merit suggested in Reference [67] is used to determine the optimal selection point of the BDT classifier

$$\text{FoM} = \frac{\epsilon^{\text{BDT}}}{a/2 + \sqrt{B_{\text{BDT}}}}\,, \tag{3.3}$$

where ϵ^{BDT} is the efficiency of the BDT selection on the $B^0_{(s)} \to p\overline{p}$ signal candidates, and is determined from simulation. B_{BDT} is the expected number of background events within the (initially excluded) signal region, estimated from the data sidebands. The term $a = 3$ quantifies the target level of significance in units of standard deviation. With this optimisation the BDT classifier is found to retain 44% of the $B^0_{(s)} \to p\overline{p}$ signals while reducing the combinatorial background level by 99.6%.

The kinematic selection of the $B^0 \to K^+\pi^-$ decay is performed using individual requirements on a set of variables similar to that used for the BDT selection of the $B^0_{(s)} \to p\overline{p}$ decays, except that the cone variables are not used. This selection differs from the selection used for signal modes and follows from the synergy with contemporary LHCb analyses on two-body charmless B decays, in particular the measurement of the effective $B^0_s \to K^+K^-$ lifetime (see Reference [68]).

The particle identification (PID) criteria applied in addition to the $B^0_{(s)} \to p\overline{p}$ BDT classifier are also optimised via Eq. 3.3. In this instance, the signal efficiencies are determined from data control samples owing to known discrepancies between data and simulation for the PID variables. Proton PID efficiencies are tabulated in bins of p, p_T

and the number of tracks in the event from data control samples of $\Lambda^0 \to p\pi^-$ decays that are selected solely using kinematic criteria. Pion and kaon efficiencies are likewise tabulated from data control samples of $D^{*+} \to D^0(\to K^-\pi^+)\,\pi^+$ decays. The kinematic distributions of the simulated decay modes are then used to determine an average PID efficiency.

Specific PID criteria are separately defined for the two signal modes and the normalisation channel. The PID efficiencies are found to be approximately 56% for the $B^0_{(s)} \to p\overline{p}$ signals and 42% for $B^0 \to K^+\pi^-$ decays.

The ratio of efficiencies of $B^0_{(s)} \to p\overline{p}$ with respect to $B^0 \to K^+\pi^-$, $\epsilon_{B^0_{(s)}\to p\overline{p}}/\epsilon_{B^0\to K^+\pi^-}$, including contributions from the detector acceptance, trigger, selection and PID, is 0.60 (0.61). After all selection criteria are applied, 45 and 58009 candidates remain in the invariant mass ranges $[5080, 5480]\,\text{MeV}/c^2$ and $[5000, 5800]\,\text{MeV}/c^2$ of the $p\overline{p}$ and $K\pi$ spectra, respectively.

Possible sources of background to the $p\overline{p}$ and $K\pi$ invariant mass spectra are investigated using simulation samples. These include partially reconstructed backgrounds with one or more particles from the decay of the b hadron escaping detection, and two-body b-hadron decays where one or both daughters are misidentified.

3.4 Signal yield determination

After the full selection, signal and background candidates are separated using unbinned maximum likelihood fits to the invariant mass spectra.

The $K\pi$ mass spectrum of the normalisation mode is described with a series of probability density functions (PDFs) for the various components, similar to Reference [69]: the $B^0 \to K^+\pi^-$ signal, the $B^0_s \to \pi^+K^-$ signal, the $B^0_s \to K^+K^-$, $B^0 \to \pi^+\pi^-$ and the $\Lambda^0_b \to p\pi^-$ misidentified backgrounds, partially reconstructed backgrounds, and combinatorial background. Any contamination from other decays is treated as a source of systematic uncertainty.

The $B \to h^+h^-$ signal mass distributions are modelled using a double Crystal Ball (DCB) function, which comprises two separate single Crystal Ball (CB) functions [70]. There is a fractional component of each CB in the total mass fit, with one describing the low mass tail and the other the high mass tail. The general single CB function is described by

$$f(x;\alpha,n,\mu,\sigma) = N \cdot \begin{cases} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\mu}{\sigma})^{-n} & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha \end{cases} \tag{3.4}$$

where μ is the mean of the CB distribution and σ is the width,

$$\begin{aligned} A &= (\tfrac{n}{|\alpha|})^n \cdot \exp(-\tfrac{|\alpha|^2}{2}) \quad , \\ B &= \tfrac{n}{|\alpha|} - |\alpha| \quad . \end{aligned} \tag{3.5}$$

The peak values and the widths of the two CB components are constrained to be the same. All CB tail parameters and the relative normalisation of the two CB functions are fixed to the values obtained from simulation whereas the signal peak value and width are free to vary in the fit to the $K\pi$ spectrum. The $B_s^0 \to \pi^+K^-$ signal width is constrained to the fitted $B^0 \to K^+\pi^-$ width such that the ratio of the widths is identical to that obtained in simulation.

The invariant mass distributions of the misidentified $B_s^0 \to K^+K^-$, $B^0 \to \pi^+\pi^-$ and $\Lambda_b^0 \to p\pi^-$ backgrounds are determined from simulation and modelled with non-parametric PDFs. The fractions of these misidentified backgrounds are related to the fraction of the $B^0 \to K^+\pi^-$ signal in the data via scaling factors that take into account the relative branching fractions [2, 71], b-hadron production fractions f_q [72, 73], and relevant misidentification rates. The latter are determined from calibration data samples.

Partially reconstructed backgrounds represent decay modes that can populate the spectrum when misreconstructed as signal with one or more undetected final-state particles, possibly in conjunction with misidentifications. The shape of this distribution is determined from simulation, where each contributing mode is assigned a weight dependent on its relative branching fraction, f_q and selection efficiency. The weighted sum of these partially-reconstructed backgrounds is shown to be well modelled with the sum of two exponentially-modified Gaussian (EMG) functions

$$\mathrm{EMG}(x;\mu,\sigma,\lambda) = \frac{\lambda}{2} e^{\frac{\lambda}{2}(2x+\lambda\sigma^2-2\mu)} \cdot \mathrm{erfc}\Big(\frac{x+\lambda\sigma^2-\mu}{\sqrt{2}\sigma}\Big) , \tag{3.6}$$

where $\mathrm{erfc(x)} = 1 - \mathrm{erf(x)}$ is the complementary error function. The signs of the variable x and parameter μ are reversed compared to the standard definition of an EMG function. The parameters defining the shape of the two EMG functions and their relative weight are determined from simulation. The component fraction of the partially-reconstructed backgrounds is obtained from the fit to the data, all other parameters being fixed from

simulation. The mass distribution of the combinatorial background is found to be well described by a linear function whose gradient is determined by the fit.

The fit to the $K\pi$ spectrum, presented in Fig. 3.1, determines seven parameters, and yields $N(B^0 \to K^+\pi^-) = 24\,968 \pm 198$ signal events, where the uncertainty is statistical. The full list of $B^0 \to K^+\pi^-$ fit parameters is shown in Table 3.1

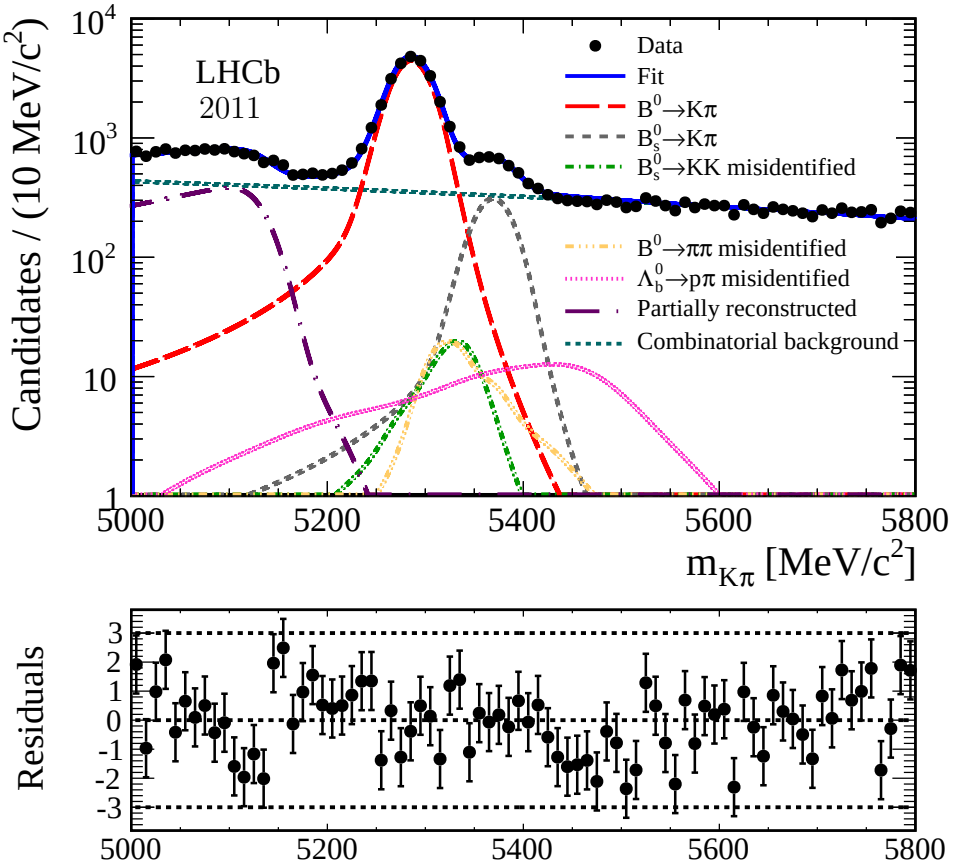


Figure 3.1: Invariant mass distribution of $K\pi$ candidates after full selection. The fit result (blue, solid) is superposed together with each fit model component as described in the legend. The normalised fit residual distribution is shown at the bottom.

The $p\overline{p}$ spectrum is described by PDFs for the three components: the $B^0 \to p\overline{p}$ and $B^0_s \to p\overline{p}$ signals, and the combinatorial background. In particular, any contamination from partially reconstructed backgrounds, with or without misidentified particles, is treated as a source of systematic uncertainty.

Potential sources of non-combinatorial background to the $p\overline{p}$ spectrum are two- and three-body decays of b hadrons into protons, pions and kaons, and many-body b-baryon

Parameter	Value		Origin	Description
f_{B^0}	0.4304 ± 0.0029		Data	B^0 signal fraction
$f_{B_s^0}$	0.03053 ± 0.00141		Data	B_s^0 signal fraction
f_{misID}^{KK}	$(2.5839 \pm 0.0171) \cdot 10^{-3}$		Data/MC	$B_s^0 \to K^+K^-$ misID fraction
$f_{misID}^{\pi\pi}$	$(2.9545 \pm 0.0196) \cdot 10^{-3}$		Data/MC	$B^0 \to \pi^+\pi^-$ misID fraction
$f_{misID}^{p\pi}$	$(5.952 \pm 0.040) \cdot 10^{-3}$		Data/MC	$\Lambda_b^0 \to p\pi^-$ misID fraction
f_{part}	0.0847 ± 0.0030		Data	Partially reconstructed fraction
μ_{B^0}	5285.600 ± 0.047	MeV/c^2	Data	B^0 signal peak mean
$\mu_{B_s^0}$	5369.19 ± 1.30	MeV/c^2	Data	B_s^0 signal peak mean
σ_{B^0}	20.955 ± 0.168	MeV/c^2	Data	B^0 signal width
$\sigma_{B_s^0}$	21.332 ± 0.171	MeV/c^2	MC	B_s^0 signal width (from B^0 width)
$\alpha_{Low}^{B^0}$	-1.161 ± 0.091		MC	Low CB part boundary (B^0)
$\alpha_{High}^{B^0}$	1.806 ± 0.059		MC	High CB part boundary (B^0)
$n_{Low}^{B^0}$	1.850 ± 0.070		MC	Low CB part tail (B^0)
$n_{High}^{B^0}$	8.080 ± 1.400		MC	High CB part tail (B^0)
$frac_{Low}^{B^0}$	0.2844 ± 0.0401		MC	MC fraction of low CB (B^0)
$\alpha_{Low}^{B_s^0}$	-0.9989 ± 0.1460		MC	Low CB part boundary (B_s^0)
$\alpha_{High}^{B_s^0}$	1.960 ± 0.0090		MC	High CB part boundary (B_s^0)
$n_{Low}^{B_s^0}$	2.036 ± 0.135		MC	Low CB part tail (B_s^0)
$n_{High}^{B_s^0}$	5.777 ± 1.154		MC	High CB part tail (B_s^0)
$frac_{Low}^{B_s^0}$	0.2265 ± 0.0531		MC	MC fraction of low CB (B_s^0)
$\mu_{K\pi}^{EMG_1}$	5137.3 ± 1.3	MeV/c^2	MC	EMG peak
$\sigma_{K\pi}^{EMG_1}$	19.95 ± 0.95	MeV/c^2	MC	EMG smearing
$\lambda_{K\pi}^{EMG_1}$	$(4.16 \pm 0.55) \cdot 10^{-3}$	$(\text{MeV}/c^2)^{-1}$	MC	EMG tail
$\mu_{K\pi}^{EMG_2}$	5166.7 ± 7.7	MeV/c^2	MC	EMG peak
$\sigma_{K\pi}^{EMG_2}$	43.6 ± 4.3	MeV/c^2	MC	EMG smearing
$\lambda_{K\pi}^{EMG_2}$	$(7.3 \pm 7.4) \cdot 10^{-5}$	$(\text{MeV}/c^2)^{-1}$	MC	EMG tail
$R_{K\pi}^{EMG_{12}}$	11.7		MC	Ratio of the two EMGs
$\nabla_{comb}^{K\pi}$	$(-1.089 \pm 0.052) \cdot 10^{-6}$	$(\text{MeV}/c^2)^{-1}$	Data	Gradient of the combinatorial

Table 3.1: Description and fitted values of the parameters of the mass fit to the $K\pi$ spectrum. The "Origin" column states if the value is determined from MC, and therefore fixed in the fit, or from the fit to the data.

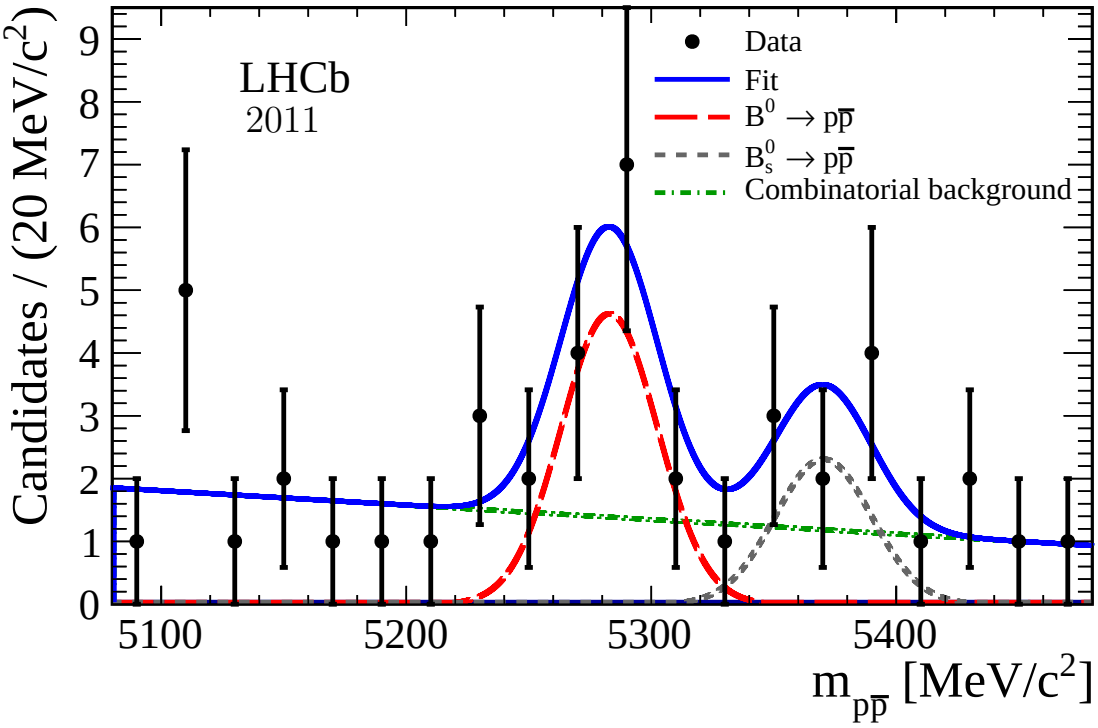


Figure 3.2: Invariant mass distribution of $p\bar{p}$ candidates after full selection. The fit result (blue, solid) is superposed with each fit model component: the $B^0 \to p\bar{p}$ signal (red, dashed), the $B_s^0 \to p\bar{p}$ signal (grey, dotted) and the combinatorial background (green, dot-dashed).

modes partially reconstructed, with one or multiple misidentifications. It is verified from extensive simulation studies that the ensemble of specific backgrounds do not peak in the signal region but rather contribute to a smooth mass spectrum, which can be accommodated by the dominant combinatorial background contribution. The most relevant backgrounds are found to be $\Lambda_b \to \Lambda_c(\to p\overline{K}^0)\pi^-$, $\Lambda_b \to \overline{K}^0 p\pi^-$, $B^0 \to K^+K^-\pi^0$ and $B^0 \to \pi^+\pi^-\pi^0$ decays. Calibration data samples are exploited to determine the PID efficiencies of these decay modes, thereby confirming the suppression with respect to the combinatorial background by typically one or two orders of magnitude. Henceforth physics-specific backgrounds are neglected in the fit to the $p\bar{p}$ mass spectrum.

The $B^0_{(s)} \to p\bar{p}$ signal mass shapes are verified in simulation to be well described by a single Gaussian function. The widths of both Gaussian functions are assumed to be the same for $B^0 \to p\bar{p}$ and $B^0_s \to p\bar{p}$; a systematic uncertainty associated to this assumption is evaluated. They are determined from simulation with a scaling factor to account for differences in the resolution between data and simulation; the scaling factor is determined

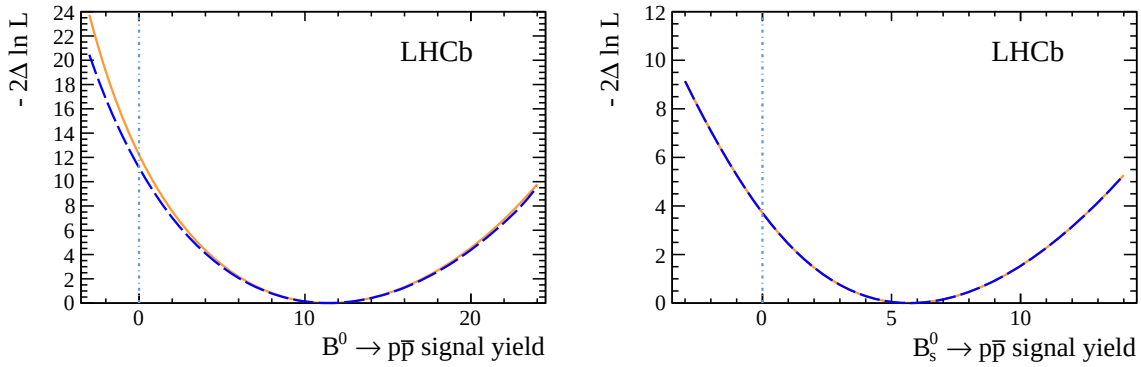


Figure 3.3: Negative logarithm of the profile likelihoods as a function of (left) the $B^0 \to p\overline{p}$ signal yield and (right) the $B^0_s \to p\overline{p}$ signal yield. The orange solid curves correspond to the statistical-only profiles whereas the blue dashed curves include systematic uncertainties.

from the $B^0 \to K^+\pi^-$ data and simulation samples. The mean of the $B^0_s \to p\overline{p}$ Gaussian function is constrained according to the B^0_s–B^0 mass difference [74]. The mass distribution of the combinatorial background is described by a linear function.

The fit to the $p\overline{p}$ mass spectrum is presented in Fig. 3.2. The full list of $B^0_{(s)} \to p\overline{p}$ fit parameters are shown in Table 3.2 The yields for the $B^0_{(s)} \to p\overline{p}$ signals in the full mass range are $N(B^0 \to p\overline{p}) = 11.4^{+4.3}_{-4.1}$ and $N(B^0_s \to p\overline{p}) = 5.7^{+3.5}_{-3.2}$, where the uncertainties are statistical only.

Parameter	Value		Origin	Description
f_{B^0}	0.254 ± 0.087		Data	B^0 signal fraction
$f_{B^0_s}$	0.127 ± 0.073		Data	B^0_s signal fraction
μ_{B^0}	5282.90 ± 6.38	MeV/c^2	Data	B^0 signal peak mean
Δm	87.35	MeV/c^2	PDG	B^0_s-B^0 mass difference
σ_B	19.69	MeV/c^2	MC	Signal mass resolution
$\nabla^{p\overline{p}}_{comb}$	$(-4.13 \pm 4.08) \times 10^{-6}$	$(\,\text{MeV}/c^2)^{-1}$	Data	Gradient of the combinatorial

Table 3.2: Description and fitted values of the parameters of the mass fit to the $p\overline{p}$ spectrum. The "Origin" column states if the value is determined from MC or the PDG, and therefore fixed in the fit, or from the fit to the data.

The statistical significances of the $B^0_{(s)} \to p\overline{p}$ signals are computed, using Wilks' theorem [75], from the change in the mass fit likelihood profiles when omitting the signal under scrutiny, namely $\sqrt{2\ln(L_{\text{S+B}}/L_{\text{B}})}$, where $L_{\text{S+B}}$ and L_{B} are the likelihoods from the

baseline fit and from the fit without the signal component, respectively. The statistical significances are $3.5\,\sigma$ and $1.9\,\sigma$ for the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ decay modes, respectively. Each statistical-only likelihood curve is convolved with a Gaussian resolution function of width equal to the systematic uncertainty on the signal yield (discussed below). The resulting likelihood profiles are presented in Fig. 3.3. The total signal significances are 3.3σ and 1.9σ for the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ modes, respectively. We observe an excess of $B^0 \to p\overline{p}$ candidates with respect to background expectations; the $B_s^0 \to p\overline{p}$ signal is not considered to be statistically significant.

3.5 Systematic uncertainties

The sources of systematic uncertainty are minimised by performing the branching fraction measurement relative to a decay mode topologically identical to the decays of interest. They are summarised in Table 3.3.

Table 3.3: Relative systematic uncertainties contributing to the $B^0_{(s)} \to p\overline{p}$ branching fractions. The total corresponds to the sum of all contributions added in quadrature.

Source	Value (%)		
	$B^0 \to p\overline{p}$	$B_s^0 \to p\overline{p}$	$B^0 \to K^+\pi^-$
$B^0 \to K^+\pi^-$ branching fraction	–	–	2.8
Trigger efficiency relative to $B^0 \to K^+\pi^-$	2.0	2.0	–
Selection efficiency relative to $B^0 \to K^+\pi^-$	8.0	8.0	–
PID efficiency	10.6	10.7	1.0
Yield from mass fit	6.8	4.6	1.6
f_s/f_d	–	7.8	–
Total	15.1	16.3	3.4

The branching fraction of the normalisation channel $B^0 \to K^+\pi^-$, $\mathcal{B}(B^0 \to K^+\pi^-) = (19.55 \pm 0.54) \times 10^{-6}$ [71], is known to a precision of 2.8%, which is taken as a systematic uncertainty. For the measurement of the $B_s^0 \to p\overline{p}$ branching fraction, an extra uncertainty arises from the 7.8% uncertainty on the ratio of fragmentation fractions $f_s/f_d = 0.256 \pm 0.020$ [73].

The trigger efficiencies are assessed from simulation for all decay modes. The simulation describes well the ratio of efficiencies of the relevant modes that comprise the

same number of tracks in the final state. Neglecting small p and p_T differences between the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ modes, the ratios of $B^0 \to K^+\pi^-/B^0_{(s)} \to p\overline{p}$ trigger efficiencies should be consistent within uncertainties. The difference of about 2% observed in simulation is taken as systematic uncertainty.

The $B^0 \to K^+\pi^-$ mode is used as a proxy for the assessment of the systematic uncertainties related to the selection; $B^0 \to K^+\pi^-$ signal distributions are obtained from data, using the *sPlot* technique, for a variety of selection variables. From the level of agreement between simulation and data, a systematic uncertainty of 8% is derived for the $B^0_{(s)} \to p\overline{p}$ selection efficiencies relative to $B^0 \to K^+\pi^-$.

The PID efficiencies are determined from data control samples. The associated systematic uncertainties are estimated by repeating the procedure with simulated control samples, the uncertainties being equal to the differences observed betweeen data and simulation, scaled by the PID efficiencies estimated with the data control samples. The systematic uncertainties on the PID efficiencies are found to be 10.6%, 10.7% and 1.0% for the $B^0 \to p\overline{p}$, $B_s^0 \to p\overline{p}$ and $B^0 \to K^+\pi^-$ decay modes, respectively. The large uncertainties on the proton PID efficiencies arise from limited coverage of the proton control samples in the kinematic region of interest for the signal.

Systematic uncertainties on the fit yields arise from the limited knowledge or the choice of the mass fit models, and from the uncertainties on the values of the parameters fixed in the fits. They are investigated by studying a large number of simulated datasets, with parameters varying within their estimated uncertainties. Combining all sources of uncertainty in quadrature, the uncertainties on the $B^0 \to p\overline{p}$, $B_s^0 \to p\overline{p}$ and $B^0 \to K^+\pi^-$ yields are 6.8%, 4.6% and 1.6%, respectively.

3.6 Results and conclusion

The branching fractions are determined relative to the $B^0 \to K^+\pi^-$ normalisation channel according to

$$\begin{aligned} \mathcal{B}(B^0_{(s)} \to p\overline{p}) &= \frac{N(B^0_{(s)} \to p\overline{p})}{N(B^0 \to K^+\pi^-)} \cdot \frac{\epsilon_{B^0 \to K^+\pi^-}}{\epsilon_{B^0_{(s)} \to p\overline{p}}} \cdot f_d/f_{d(s)} \cdot \mathcal{B}(B^0 \to K^+\pi^-) \\ &= \alpha^{2011}_{d(s)} \cdot N(B^0_{(s)} \to p\overline{p})\,, \end{aligned} \tag{3.7}$$

where $\alpha^{2011}_{d(s)}$ are the single-event sensitivities equal to $(1.31 \pm 0.18) \times 10^{-9}$ and $(5.04 \pm 0.81) \times 10^{-9}$ for the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ decay modes, respectively; their uncertainties

amount to 14% and 16%, respectively.

The Feldman-Cousins (FC) frequentist method [76] is chosen for the calculation of the branching fractions. The determination of the 68.3% and 90% CL bands is performed with simulation studies relating the measured signal yields to branching fractions, and accounting for systematic uncertainties.

The FC approach naturally determines one- or two-sided confidence intervals, meaning upper limits (ULs) on branching fractions or branching fraction measurements in our specific case. Given the "mix" of expectations there are for the B^0 and B_s^0 decays, this is an excellent feature to exploit.

The determination of the 68.27% and 90% confidence level (CL) intervals is done with toys for both the B^0 and the B_s^0 modes. In the event of a significant excess in either mode, the corresponding 68.27% CL interval will then translate into the usual 1σ error measurement. The confidence intervals with systematic errors are calculated following the method outlined in [76]. The FC confidence bands are formed by scanning across the likelihood ratio

$$R(x) = P(x|\mu)/P(x|\mu_{\rm best}) \quad , \tag{3.8}$$

where $P(x|\mu)$ is the probability density function for measuring a value for an observable x from the value of a parameter μ. In this case, μ is $\mathcal{B}(B^0_{(s)} \to p\overline{p})$ and x is a sample value for the $B^0_{(s)} \to p\overline{p}$ yield, $N^{\rm test}_{\rm fit}$. $P(x|\mu)$ is here described by a Gaussian,

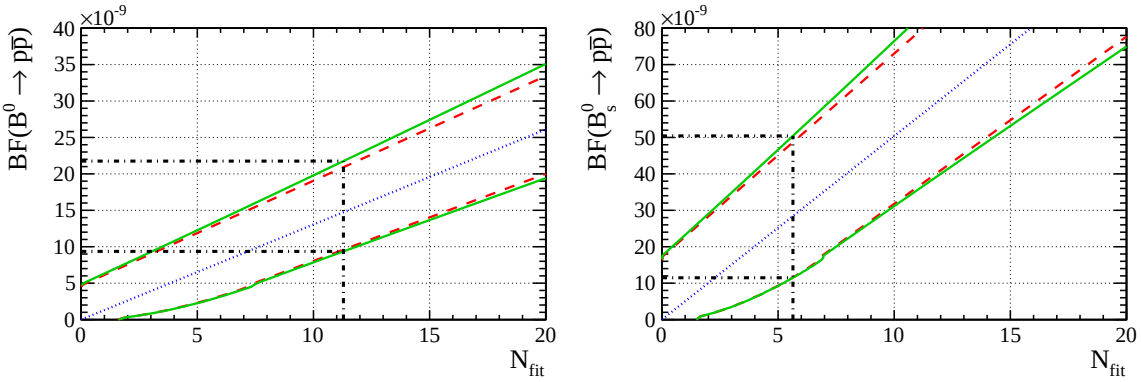
$$P(N^{\rm test}_{\rm fit}|\mathcal{B}(B^0_{(s)} \to p\overline{p})) = G(N^{\rm test}_{\rm fit}; N_{\rm fit}, \sigma_{N_{\rm fit}}) \quad , \tag{3.9}$$

where $N_{\rm fit}$ is the $B^0_{(s)} \to p\overline{p}$ yield calculated from $\mathcal{B}(B^0_{(s)} \to p\overline{p})$ via Eq. 6.7. For each value of $N_{\rm fit}$, $N^{\rm test}_{\rm fit}$ is sampled across the range $N_{\rm fit} \pm 5\sigma_{N_{\rm fit}}$. The value $\mu_{\rm best}$ is that which maximises $P(x|\mu_{\rm best})$ for each value of x, for $N_{\rm fit} \geq 0$, as is the case for both $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$, $\mu_{\rm best}$ is the peak of the Gaussian defined in Eq. 3.9. With $P(x|\mu)$ described by a Gaussian and $x \geq 0$, the likelihood ratio $R(x)$ is found to be given by

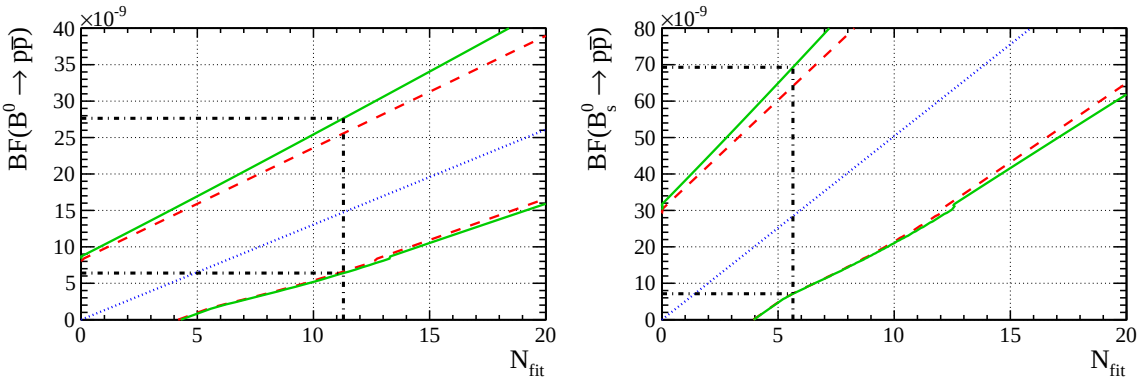
$$R(x) = P(x|\mu)/P(x|\mu_{\rm best}) = \exp(-(x-\mu)^2/2\sigma(\mu)^2) \quad . \tag{3.10}$$

For a chosen $\mathcal{B}(B^0_{(s)} \to p\overline{p})$, $P(N^{\rm test}_{\rm fit}|\mathcal{B}(B^0_{(s)} \to p\overline{p}))$ and $R(N^{\rm test}_{\rm fit})$ are calculated for each value of $N^{\rm test}_{\rm fit}$, with these $N^{\rm test}_{\rm fit}$ values ranked according to $R(N^{\rm test}_{\rm fit})$. To construct the required CL interval we scan over the values of $N^{\rm test}_{\rm fit}$ in order of ranking, summing $P(N^{\rm test}_{\rm fit}|\mathcal{B}(B^0_{(s)} \to p\overline{p}))$ until the total probability exceeds the desired value (68.27% or

90%). The confidence bands for the number of fitted signal events (N_{fit}) over the ranges $[0, 20]$ events for $B^0 \to p\bar{p}$ and $B_s^0 \to p\bar{p}$ are shown in Figure 3.4.



(a) $B^0 \to p\bar{p}$ 68.27% CL **(b)** $B_s^0 \to p\bar{p}$ 68.27% CL



(c) $B^0 \to p\bar{p}$ 90% CL **(d)** $B_s^0 \to p\bar{p}$ 90% CL

Figure 3.4: FC confidence level intervals on the signal yield at 68.27% (a, b) and 90% (c, d) confidence levels for the $B_{(s)}^0 \to p\bar{p}$ signal modes. The blue dotted lines show the central value for N_{fit} used in the calculation. The red dotted lines show the lower and upper limits with only statistical uncertainties included, while the green solid lines show the lower and upper limits with statistical and systematic uncertainties included. The black dashed show the limits extracted from this analysis for the 11.29 observed $B^0 \to p\bar{p}$ events and 5.64 observed $B_s^0 \to p\bar{p}$ events.

From the results shown in Fig. 3.4 we obtain 68.27% and 90% CL intervals with full statistical and systematic uncertainties for $\mathcal{B}(B_{(s)}^0 \to p\bar{p})$ at the observed signal yields of 11.29 events ($B^0 \to p\bar{p}$) and 5.64 events ($B_s^0 \to p\bar{p}$):

$$\mathcal{B}(B^0 \to p\bar{p}) = (1.47\,^{+0.62\,+0.35}_{-0.51\,-0.14}) \times 10^{-8} \quad \text{at} \quad 68.3\% \quad \text{CL}\,,$$
$$\mathcal{B}(B^0 \to p\bar{p}) = (1.47\,^{+1.09\,+0.69}_{-0.81\,-0.18}) \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL}\,,$$
$$\mathcal{B}(B^0_s \to p\bar{p}) = (2.84\,^{+2.03\,+0.85}_{-1.68\,-0.18}) \times 10^{-8} \quad \text{at} \quad 68.3\% \quad \text{CL}\,,$$
$$\mathcal{B}(B^0_s \to p\bar{p}) = (2.84\,^{+3.57\,+2.00}_{-2.12\,-0.21}) \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL}\,,$$

where the first uncertainties are statistical and the second are systematic.

In summary, a search has been performed for the rare two-body charmless baryonic decays $B^0 \to p\bar{p}$ and $B^0_s \to p\bar{p}$ using a data sample, corresponding to an integrated luminosity of 0.9 fb^{-1}, of pp collisions collected at a centre-of-mass energy of 7 TeV by the LHCb experiment. The results allow two-sided confidence limits to be placed on the branching fractions of both $B^0 \to p\bar{p}$ and $B^0_s \to p\bar{p}$ for the first time. We observe an excess of $B^0 \to p\bar{p}$ candidates with respect to background expectations with a statistical significance of $3.3\,\sigma$. This is the first evidence for a two-body charmless baryonic B^0 decay. No significant $B^0_s \to p\bar{p}$ signal is observed and the present result improves the previous bound by three orders of magnitude.

Figure 3.5 shows a comparison of the 90% CL interval on $\mathcal{B}\,(B^0 \to p\bar{p})$ from this analysis with the previous experimental limits and existing theory predictions. The measured $B^0 \to p\bar{p}$ branching fraction is incompatible with all published theoretical predictions by one to two orders of magnitude and motivates new and more precise theoretical calculations of two-body charmless baryonic B decays. An improved experimental search for these decay modes at LHCb with the full 2011 and 2012 dataset will help to clarify the situation, and is the topic of chapters 4, 5 and 6 of this thesis.

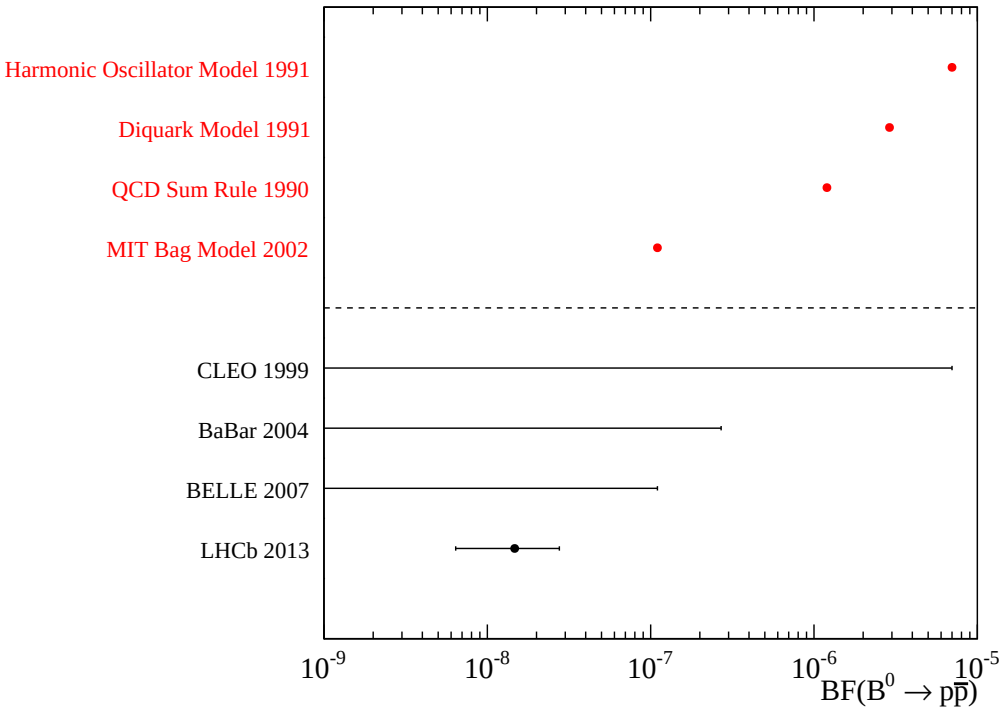


Figure 3.5: Theoretical and experimental limits of $\mathcal{B}\left(B^0 \to p\overline{p}\right)$ including the result presented in this chapter (LHCb 2013). Theoretical predictions are shown in red and are taken from the values listed in Table 1.5. The uncertainties on the theoretical predictions were not provided. Experimental results are shown in black and are taken from the 90% CL interval presented in this chapter and the values listed in Table 1.4.

Chapter 4

The Search for the Decays $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ with Combined 2011 and 2012 Data

4.1 Analysis Strategy

The strategy chosen for this update analysis largely follows that of the search based on the 2011 data sample, as described in the previous chapter (see also References [36, 1]). In particular, the $B^0_{(s)} \to p\overline{p}$ signal branching fractions are again measured relative to the branching fraction of the normalisation mode $B^0 \to K^+\pi^-$, in order to cancel or minimise systematic effects, via Equation 3.2, which is repeated here for convenience:

$$\mathcal{B}(B^0_{(s)} \to p\overline{p}) = \frac{N(B^0_{(s)} \to p\overline{p})}{N(B^0 \to K^+\pi^-)} \cdot \frac{\epsilon_{B^0 \to K^+\pi^-}}{\epsilon_{B^0_{(s)} \to p\overline{p}}} \cdot \frac{\int \mathcal{L}\, \mathrm{d}t^{B^0 \to K^+\pi^-}}{\int \mathcal{L}\, \mathrm{d}t^{B^0_{(s)} \to p\overline{p}}} \cdot \mathcal{B}(B^0 \to K^+\pi^-) \cdot \frac{f_d}{f_{d(s)}} \,. \quad (4.1)$$

where, once again, ϵ contains the product of all efficiencies for the signal daughters to be in the LHCb acceptance, for triggering, reconstruction, stripping and final selection. For $B^0_s \to p\overline{p}$ the extra factor f_d/f_s appears on the righthand side of the equation, to take into account the fragmentation of the B^0 and B^0_s mesons..

The analysis adopts a blind approach, ignoring the $B^0 \to p\overline{p}$ and $B^0_s \to p\overline{p}$ signal regions throughout the whole process. The signal region is defined as $[5230, 5417]\,\mathrm{MeV}/c^2$ which corresponds to $\pm 50\,\mathrm{MeV}/c^2$ from the nominal B^0 and B^0_s masses.

A certain number of important changes are worth listing:

- Tighter $B^0_{(s)} \to p\overline{p}$ stripping selection to increase the background rejection earlier in the selection chain

- For the multivariate analysis classifier, a Multilayer Perceptron Artificial Neural Network (MLP ANN) is now used whereas a Boosted Decision Tree (BDT) was implemented in the previous analysis.

- Two multivariate analysis (MVA) discriminators will in fact be used, each on 50% of the data. In the previous analysis 10% of the data was used to optimise the selection and was then removed from the final analysis dataset. By splitting the data we are able to fully exploit the available sample without introducing a bias to the analysis.

- The PID selections applied to the $B^0_{(s)} \to p\overline{p}$ candidate daughter tracks are now independent with separate PID selection cuts applied to proton and antiproton candidate tracks.

- The final selection of the signals and the normalisation mode will be made as similar as possible, using the same multivariate analysis methods in each selection, to minimise the systematic uncertainties coming from the selection determined in the 2011 analysis.

- The invariant mass fits of the $B^0 \to K^+\pi^-$ normalisation channel are split into two separate fits of the individual $K\pi$ charge states.

4.2 Event Samples

4.2.1 Data

This analysis uses data recorded by the LHCb experiment in 2011 (recorded at a centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$) and 2012 ($\sqrt{s} = 8\,\text{TeV}$) for a total combined data set of 3.122 fb^{-1}. Table 4.1 lists the data sets used for this analysis. It is important to note that the data set used for the $B^0 \to K^+\pi^-$ selection contains $0.005\,\text{fb}^{-1}$ less recorded luminosity in 2012 than the data set used for the $B^0_{(s)} \to p\overline{p}$ events. This is due to slightly different stripping versions being used for the signal and normalisation channel selections. This is taken into account in the branching fraction calculation.

Selection	$\int \mathcal{L}$ 2011 (fb^{-1})	$\int \mathcal{L}$ 2012 (fb^{-1})
$B^0_{(s)} \to p\overline{p}$	1.078 ± 0.013	2.044 ± 0.024
$B^0 \to K^+\pi^-$	1.078 ± 0.013	2.039 ± 0.024

Table 4.1: Data sets used in this analysis.

4.2.2 Monte Carlo Simulation

Several Monte Carlo (MC) simulation samples with the MC2012 configuration were used in the analysis, notably for selection, normalisation and background studies. The full list of samples used is shown in Table 4.2.

4.3 Event Selection

4.3.1 Trigger

At the L0 level, events are required to pass either a specific hadron "TOS" or a global "TIS" trigger selection, where TOS and TIS denote "Trigger On Signal" and "Trigger Independent of Signal", respectively. Following L0, at the HLT level only hadronic decay TOS trigger lines are required, specifically a hadron tracking seelction trigger at the HLT1 and either a specific two-body hadronic B decay trigger (B2HH) or a topological two-body B decay trigger which utilises a BDT algorithm. at HLT2. Table 4.3 lists the L0, HLT1 and HLT2 trigger line requirements imposed on signal candidates with their approximate $B^0_{(s)} \to p\overline{p}$ efficiencies with respect to the previous trigger level (or stripping level in the case of the L0 efficiency). These trigger requirements are identical to those used in the previous analysis.

4.3.2 Stripping Selection

Stripping selection criteria have been derived to exploit the characteristic topologies and kinematics of $B^0_{(s)} \to p\overline{p}$ decays and have been refined between this analysis and the previous search for $B^0_{(s)} \to p\overline{p}$ published in Reference [1]. Table 4.4 lists the $B^0_{(s)} \to p\overline{p}$ stripping selection cuts used in this analysis with the stripping selection used in the previous analysis also listed for comparison. For this analysis, compared to the previous, the stripping selection has tighter requirements on the daughter particle minimum IP χ^2, cosine of the B direction angle, B vertex and IP χ^2 and the B transverse momentum.

Decay mode	MC Sample Size		$\mathcal{B}(\times 10^{-6})$	Used as bkg.
	2011	2012		
$B^0 \to p\overline{p}$	1126997	2075494	$1.47\,^{+0.71}_{-0.53} \times 10^{-2}$ [1]	-
$B_s^0 \to p\overline{p}$	1122999	2064491	$2.84\,^{+2.20}_{-1.69} \times 10^{-2}$ [1]	-
$B^0 \to K^+\pi^-$	775498	1529495	$19.57\,^{+0.53}_{-0.52}$ [71]	$p\overline{p}$
$B_s^0 \to \pi^+K^-$	1514494	3071739	5.38 ± 0.76 [77]	$p\overline{p}$
$B_s^0 \to \pi^+\pi^-$	1024500	2030741	0.73 ± 0.14 [71]	$p\overline{p}$
$B^0 \to K^+K^-$	1027248	2035242	0.11 ± 0.78 [77]	$p\overline{p}$
$\Lambda_b^0 \to pK^-$	775995	1513745	6.2 ± 1.9 [2]	$p\overline{p}$
$\Lambda_b^0 \to p\pi^-$	764750	1509492	4.0 ± 1.3 [2]	$p\overline{p}$
$B^0 \to K^+\pi^-\pi^0$	825745	1540497	37.8 ± 3.2 [71]	$p\overline{p}$, $K\pi$
$B^0 \to \pi^+\pi^-\pi^0$	1296245	2554495	$50^{(*)}$ (< 720 @ 90% CL [2])	$p\overline{p}$, $K\pi$
$B^0 \to K^+K^-\pi^0$	1334245	2554490	2.17 ± 0.65 [71]	$p\overline{p}$, $K\pi$
$B_s^0 \to K^-\pi^+\pi^0$	1328745	2523492	$5^{(*)}$	$p\overline{p}$, $K\pi$
$\Lambda_b \to (\Lambda_c \to pK_S^0)\pi^-$	2565742	2519745	151 ± 19 [2]	$p\overline{p}$
$B_s^0 \to K^+K^-\pi^0$	591999	1013498	$20^{(*)}$	$p\overline{p}$, $K\pi$
$B^0 \to \pi^+\pi^-$	1527244	3067742	5.11 ± 0.22 [71]	$p\overline{p}$
$B_s^0 \to K^+K^-$	1532248	3052242	25.4 ± 3.7 [71]	$p\overline{p}$
$B^+ \to p\overline{p}\pi^+$	778249	1525246	$1.60\,^{+0.18}_{-0.17}$ [71]	$p\overline{p}$, $K\pi$
$B^+ \to p\overline{p}K^+$	530500	1038747	5.48 ± 0.34 [71]	$p\overline{p}$, $K\pi$
$B^+ \to \pi^+\pi^-K^+$	519000	1020995	16.3 ± 2.0 [71]	$p\overline{p}$, $K\pi$
$B^+ \to \pi^+K^-K^+$	509998	1024197	5.0 ± 0.7 [71]	$p\overline{p}$, $K\pi$
$B^+ \to K^+K^-K^+$	1035498	2039993	32.5 ± 1.5 [71]	$p\overline{p}$, $K\pi$
$B^+ \to \pi^+\pi^-\pi^+$	512998	1014198	15.2 ± 1.4 [71]	$p\overline{p}$, $K\pi$
$\Lambda_b \to (\Lambda_c \to pK_S^0(\pi^0 \to \gamma\gamma))\pi^-$	519999	1015749	209 ± 33 [2]	$p\overline{p}$
$B^0 \to K_S^0\pi^+\pi^-$	1007496	4058986	25.9 ± 1.0 [71]	$p\overline{p}$, $K\pi$
$\Lambda_b \to \Lambda K^+K^-$	1020499	4014238	$9.97^{(*)}$	$p\overline{p}$, $K\pi$
$\Lambda_b \to \Lambda\pi^+\pi^-$	504249	2074741	$25.9^{(*)}$	$p\overline{p}$, $K\pi$
$\Lambda_b \to \Lambda p\overline{p}$	511998	2048494	$1.33^{(*)}$	$p\overline{p}$, $K\pi$
$\Lambda_b \to (D^0 \to K_S^0 \to \pi^+\pi^-)pK^-$	520247	1521496	1.37 ± 0.27 [2]	$p\overline{p}$, $K\pi$
$\Lambda_b \to p\pi^-K^-\pi^+$	1092942	2004286	$0.1^{(*)}$	$p\overline{p}$
$B^+ \to p\overline{\Lambda}\gamma$	917750	2017261	$2.4\,^{+0.5}_{-0.4}$ [2]	$p\overline{p}$
$B^+ \to p\overline{\Lambda}\pi^0$	525999	2052670	$3.0\,^{+0.61}_{-0.62}$ [71]	$p\overline{p}$
$B^+ \to (J/\psi \to p\overline{p})\pi$	541498	1020495	40.4 ± 1.7 [71]	$p\overline{p}$
$B^0 \to p\pi^-\overline{\Sigma^0}$	779748	1522745	3.8 (< 3.8 @ 90% CL [2])	$p\overline{p}$
$B^0 \to \Lambda\overline{\Lambda}$	2541994	5056486	0.32 (< 0.32 @ 90% CL [2])	$p\overline{p}$

Table 4.2: MC samples used during the analysis for selection, normalisation and background studies. Decays marked with $^{(*)}$ indicate that the branching fraction is not known and a value estimated from a similar decay ($B^0 \to K_S^0 h^+ h^-$) was used. Note that the given branching fractions are not visible, *e.g.* the K_S^0 branching fraction needs to be accounted for. The last column indicates whether the mode has been used for background studies of $B_{(s)}^0 \to p\overline{p}$ and/or $B^0 \to K^+\pi^-$. Sample sizes are the number of generated MC events before any trigger or stripping selection.

There is also now included a cut on the daughters ghost probability, which was previously included as a preselection cut prior to the MVA selection. The mass range has also been

Trigger stage	Required lines	Approx. eff. 2011 (%)	Approx. eff. 2012 (%)
L0	Hadron TOS	43	39
	Global TIS	32	30
	`OR` of both L0 lines	59	56
HLT1	Tracking TOS	77	76
HLT2	Two-body hadronic B TOS	64	64
	Topological two-body B BDT TOS	79	87
	`OR` of both HLT2 lines	86	89
$\epsilon_{\text{trig/strip}}$	`ALL`	39	38

Table 4.3: List of trigger line requirements imposed on signal candidates. The final two columns indicate the approximate $B^0 \to p\overline{p}$ efficiencies when going from stripping $\to$ L0 $\to$ HLT1 $\to$ HLT2 (calculated with respect to the previous trigger level) in 2011 and 2012 MC.

extended up to $200\,\text{MeV}/c^2$ above the B_s^0 mass to increase the size of the upper mass sideband. As part of the stripping selection stage for the $B^0_{(s)} \to p\overline{p}$ MC we also apply MC truth matching requirements where we require that for each event the mother particle is a true B^0 (B_s^0) and the daughter particles are a true protons and a true antiproton. We also require that the mother particle for each daughter is a true B^0 (B_s^0). All the following selection stages and efficiencies are determined using truth matched MC events.

As a comparison, this stripping selection retains $\sim 24\%$ of truth-matched 2011 $B^0 \to p\overline{p}$ MC events whereas the selection applied in the previous analysis had a $B^0 \to p\overline{p}$ signal retention efficiency of $\sim 59\%$, which reflects the overall tightening of the stripping selection criteria.

4.3.3 Preselection

Multivariate analysis (MVA) and particle identification (PID) selections are applied to events passing the stripping and trigger selections. Before the MVA and PID selection stages a small number of preselection cuts are applied to clean up the data:

- Fiducial momentum cuts on both daughter tracks, $5 < p < 150\,\text{GeV}/c$.
- Fiducial pseudorapidity cuts on both daughter tracks, $1.5 < \eta < 5.0$.
- $\ln(B\text{-decay vertex distance in } z \text{ (in mm) from the related primary vertex}) > -0.15$.

Variable	Cut value This analysis	Cut value 2011 analysis
Min. of daughters' p_T	$> 900\,\text{MeV}/c$	$> 900\,\text{MeV}/c$
Min. of daughters' min. IP χ^2 with respect to any PV	> 10	> 9
Max. of daughters' p_T	$> 2100\,\text{MeV}/c$	$> 2100\,\text{MeV}/c$
Max. of daughters' min. IP χ^2 with respect to any primary vertex	> 25	> 25
Daughters' $\text{DLL}_{p\pi}$	> -1	> 0
Daughters' $\text{DLL}_{p\text{K}}$	> -2	> -2
Daughters' Ghost Probability	< 0.4	N/A
Daughters' track fit χ^2/ nDoF	N/A	< 5
$\lvert(B^0 \lvert\lvert B_s^0)$ mass - reference (PDG) mass$\rvert$	$< 200\,\text{MeV}/c^2$	N/A
$\lvert B^0$ mass - reference (PDG) mass$\rvert$	N/A	$< 200\,\text{MeV}/c^2$
B p_T	$> 1100\,\text{MeV}/c$	$> 1000\,\text{MeV}/c$
B vertex χ^2	< 9	< 16
B IP χ^2	< 16	< 36
$\cos(B$ direction angle$) \equiv \cos(\text{DIRA})$	> 0.9997	> 0.9995

Table 4.4: $B^0_{(s)} \to p\overline{p}$ stripping line cuts used in this analysis, with those used in the previous analysis for comparison. Note that $\text{DLL}_{p\text{K}} = \text{DLL}_{p\pi} - \text{DLL}_{\text{K}\pi}$, with DLL denoting the difference in the logarithm of the particle identification likelihood.

The daughter p and η cuts are a consequence of the PID calibration method used within the PID selection, which is detailed in Section 4.3.4.

4.3.4 PID and Multivariate Selections

As with the 2011 data analysis [1], MVA and PID selections are vital components of the $B^0_{(s)} \to p\overline{p}$ selection chain. In contrast to the previous analysis we now apply the PID selection before the MVA. The main intention of the MVA selection is to discriminate the $B^0_{(s)} \to p\overline{p}$ signals from combinatorial background events. Therefore, applying the PID selection before the final MVA should improve the performance of the MVA selection by providing a much purer sample of combinatorial background events for training than if only the loose stripping level PID selection was applied. The MVA selection is then also not being trained on events which would be removed by the PID selection anyway.

PID Selection

A PID selection is applied to both daughter tracks. Initial studies into the PID selection optimisation found that using all of the events passing the preselection, trigger and stripping cuts retains too many background events to allow an effective PID optimisation to be performed. Therefore it was decided to apply an additional set of selection cuts to signal and background events prior to the PID optimisation to reduce the initial background statistics and increase the efficiency of the optimisation procedure. These additional selection cuts are used solely within the PID selection optimisation procedure and are not included in the final analysis selection chain.

In order to keep the PID preselection as close to the final analysis selection as possible an MVA selection was trained and applied using a configuration very similar to that used in the final MVA selection for the full analysis, described in Section 4.3.4. The MVA selection is trained using $B^0 \to p\overline{p}$ 2012 MC as signal and 2012 sideband data as background with the sideband data in two samples split by even or odd event number. The end result is two separate multilayer perceptron (MLP) artificial neural network (ANN) selections trained on even and odd numbered sideband events respectively with each then applied on the other set of events, i.e. An even event trained MLP is applied to odd numbered events and vice versa. For each MLP selection, the $B^0 \to p\overline{p}$ signal MC events are split randomly into equally sized training and testing samples. The full details of the construction and training of these MLP selections are shown in Appendix C. MLP cuts were chosen to retain 50% of signal $B^0 \to p\overline{p}$ events, which retains 1018 total sideband events from 2012 data (0.52% retention rate in sideband data).

PID Calibration

A robust PID selection is a basic requirement of an analysis such as this: however, it is well established that the description of PID variables in MC simulation is poor when compared to real data. Therefore, as mentioned previously, $B^0_{(s)} \to p\overline{p}$ signal PID efficiencies are calculated using the `PIDCalib` package.

The `PIDCalib` package provides a method for calculating accurate charged track PID efficiencies using calibration samples from data. An explanation of the methods and software used in the package can be found in Ref. [78]. For proton PID, the package provides samples of calibration protons from $\Lambda^0 \to p\pi^-$ and hadronically triggered $\Lambda_c \to pK^-\pi^+$ decays which are all selected without using any PID information.

PID performance is known to be correlated with various kinematic and event proper-

ties, in particular track momentum, pseudorapidity, transverse momentum and the number of tracks in the event. The `PIDCalib` method requires that PID efficiencies be calculated over a defined region of kinematic phase-space within a sensible binning scheme. The choice of binning variables should attempt to accurately describe the full kinematic dependence of the LHCb detector PID response. The internal binning structure needs to cover as much of the $B^0_{(s)} \to p\overline{p}$ kinematics while retaining sufficient calibration event statistics within each bin. To minimise systematic uncertainties due to using the `PIDCalib` method, the binning schemes chosen need to minimise the variation in PID response within each bin, whilst taking into account the previously listed requirements. For a given PID selection, histograms are made of the calibration tracks before and after the PID cuts are applied, within the defined binning scheme. From these a performance histogram is made calculating the efficiencies of the PID selection for each bin. PID efficiencies are then calculated for the $B^0_{(s)} \to p\overline{p}$ MC samples on an event-by-event basis where each daughter track is assigned the PID efficiency value of the calibration performance histogram bin corresponding to the kinematic properties of the track.

The choices of kinematic variables and binning scheme were made following studies using calibration data and signal MC (Appendix D). For this analysis, a two-dimensional binning scheme is chosen, in which tracks are binned in momentum, p, and pseudorapidity, η. Though it is well established that there is also a PID dependence on the number of tracks in the event, $nTracks$, we choose not to include this variable in the binning. This is because the per-event PID efficiencies given by the output of the `PIDCalib` method are used during the MVA selection (see Section 4.3.4) to reweight the signal MC input variable distributions: however, the $nTracks$ response in simulation is known to poorly represent what is seen in data so it should not be used as part of this MVA input reweighting procedure. We assign a systematic uncertainty due to the choice of kinematic binning variables in Section 6.1.5. Table 4.5 shows the binning scheme used for protons, the upper bound in momentum at 150 GeV/c is due to insufficient calibration sample statistics above this value. Figure 4.1 shows the $p-\eta$ distributions of $B^0 \to p\overline{p}$ MC and $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ calibration data protons, where the $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ tracks are shown within the binning scheme listed in Table 4.5. From looking at the distributions shown in Figure 4.1, it is clear that the kinematics of the $B^0 \to p\overline{p}$ protons are much closer to those of the $\Lambda_c \to pK^-\pi^+$ decay than $\Lambda^0 \to p\pi^-$. However, the overall available statistics for $\Lambda_c \to pK^-\pi^+$ calibration protons is very low compared to $\Lambda^0 \to p\pi^-$. Therefore a combination of the two calibration samples is used.

To combine the calibration protons from $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ decays we use

Variable	Binning
Proton η	1.5 : 2 : 2.25 : 2.5 : 2.75 : 3 : 3.25 : 3.5 : 3.75 : 4 : 4.33333 : 4.66667 : 5
Proton p (MeV/c)	5000 : 8750 : 12500 : 16250 : 20000 : 28000 : 36000 : 44000 : 52000 : 60000 : 68000 76000 : 84000 : 92000 : 100000 : 110000 : 120000 : 130000 : 140000 : 150000

Table 4.5: Binning scheme used for proton PID calibration.

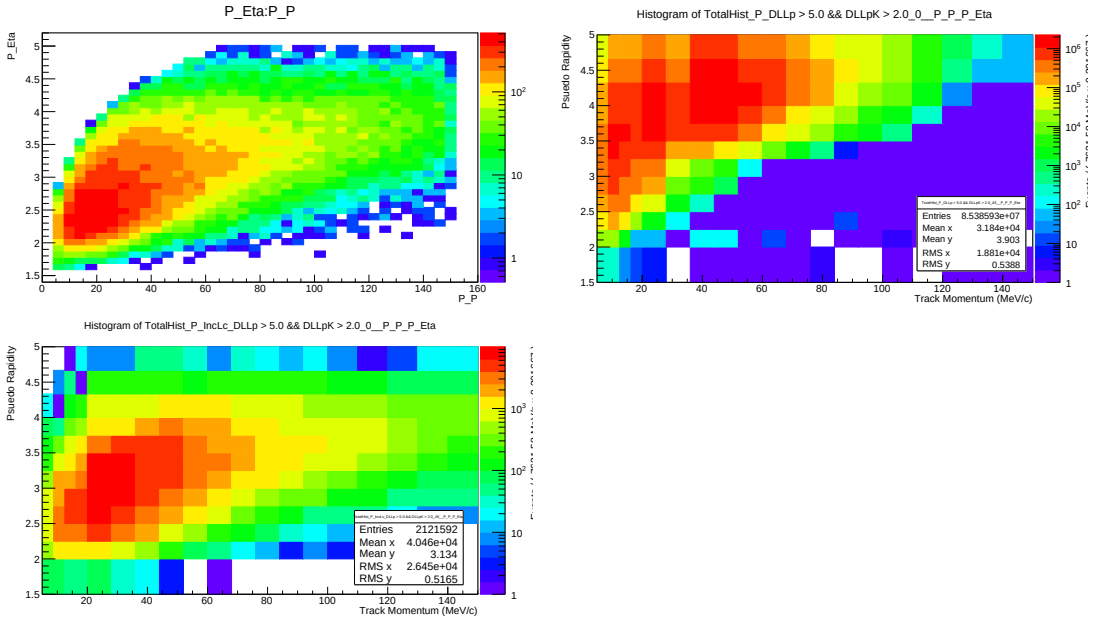


Figure 4.1: Two-dimensional distributions of p and η for 2012, magnet up $B^0 \to p\overline{p}$ MC (top, left) and calibration data samples $\Lambda^0 \to p\pi^-$ (top, right) and $\Lambda_c \to pK^-\pi^+$ (bottom). The $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ distributions are binned in the final calibration binning scheme.

a method which merges the efficiencies of the two samples via a weighted average. First a bin-by-bin standard deviation consistency test is made between the $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ performance histograms,

$$N_{\text{std.dev}} = \frac{|\epsilon_{\Lambda^0} - \epsilon_{\Lambda_c^{\text{inc}}}|}{\sqrt{\sigma^2_{\Lambda^0} + \sigma^2_{\Lambda_c^{\text{inc}}}}} \quad , \tag{4.2}$$

where ϵ_{Λ^0} ($\epsilon_{\Lambda_c^{\text{inc}}}$) and σ_{Λ^0} ($\sigma_{\Lambda_c^{\text{inc}}}$) are the PID efficiency and statistical uncertainty for the $\Lambda^0 \to p\pi^-$ ($\Lambda_c \to pK^-\pi^+$) proton calibration performance histogram bin. For any bin where $N_{\text{std.dev}} < 3$ the efficiencies of the two samples are merged via the weighted average,

$$\epsilon_{\text{weighted}} = \frac{(\epsilon_{\Lambda^0}/\sigma^2_{\Lambda^0}) + (\epsilon_{\Lambda_c^{\text{inc}}}/\sigma^2_{\Lambda_c^{\text{inc}}})}{1/\sigma^2_{\Lambda^0} + 1/\sigma^2_{\Lambda_c^{\text{inc}}}} \quad . \tag{4.3}$$

For bins where $N_{\text{std.dev}} > 3$ the efficiency of the sample with the largest yield in that bin after the chosen PID cut is used as the merged efficiency. The weights of any bins in either histogram with less than five events are set to zero. If two corresponding bins in the $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ performance histograms both have a weight of zero, the weighted average of the two efficiencies is automatically set to zero. Figure 4.2 shows an example merged performance histogram for stripping 20, magnet down, $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ protons for a PID selection of $\text{DLL}_{p\pi} > 14$ and $\text{DLL}_{p\text{K}} > 8$. The figure shows that the resulting binned efficiency distribution gives sensible results and does not feature large changes in efficiency between adjacent bins.

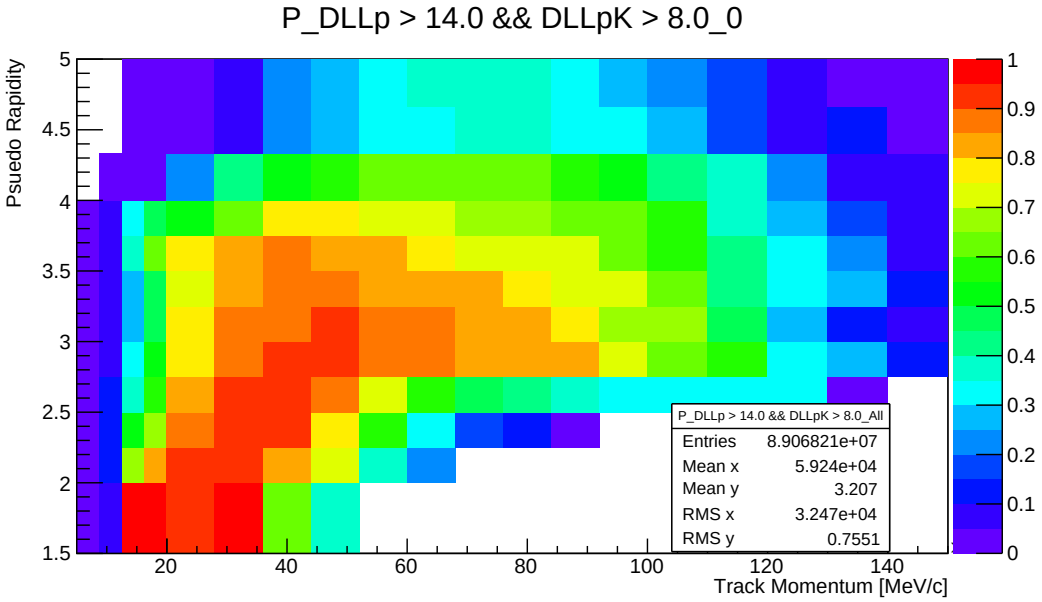


Figure 4.2: PID performance histogram of stripping 20, magnet down, merged $\Lambda^0 \to p\pi^-$ and $\Lambda_c \to pK^-\pi^+$ protons for a PID selection of $\text{DLL}_{p\pi} > 14$ and $\text{DLL}_{p\text{K}} > 8$.

PID Optimisation

The optimisation of the PID selection is performed on 2012 data and MC using the Punzi figure of merit (FoM) [67]

$$\text{FoM} = \frac{\epsilon^{\text{PID}}_{B^0 \to p\bar{p}}}{a/2 + \sqrt{B^{\text{PID/presel}}}} \quad , \tag{4.4}$$

where $\epsilon^{\text{PID}}_{B^0 \to p\bar{p}}$ is the selection efficiency of the PID cuts on 2012 $B^0 \to p\bar{p}$ MC, calculated

using the `PIDCalib` package, and $B^{\text{PID/presel}}$ is the expected number of 2012 background events passing the same set of cuts. The term a is the target signal significance in units of σ; for this analysis the value $a = 5$ is chosen. The expected background yield within the signal region for a specific PID selection, $B^{\text{PID/presel}}$, is estimated from the data sidebands.

A wide range of $\text{DLL}_{p\pi}$ and $\text{DLL}_{p\text{K}}$ cuts were studied with the FoM calculated at each set of cuts. In the previous analysis, identical PID cuts were applied to proton and antiproton tracks; however, for this analysis we allow the two daughter tracks to have different PID selections, effectively optimising the PID selection in four dimensions of PID "phase space": proton $\text{DLL}_{p\pi}$, proton $\text{DLL}_{p\text{K}}$, antiproton $\text{DLL}_{p\pi}$ and antiproton $\text{DLL}_{p\text{K}}$. Figure 4.3 shows the distributions of $\epsilon^{\text{PID}}_{B^0\to p\overline{p}}$ and $B^{\text{PID/presel}}$ used in the calculation of the PID selection figure of merit. The two plots shown in Figure 4.4 display the FoM responses over ranges of $\text{DLL}_{p\pi}$ and $\text{DLL}_{p\text{K}}$ cut values. The PID selection which yields the highest FoM value is proton $\text{DLL}_{p\pi} > 16$, $\text{DLL}_{p\text{K}} > 8$ and antiproton $\text{DLL}_{p\pi} > 15$, $\text{DLL}_{p\text{K}} > 4$. However, these cuts are very harsh and it can be seen from the plots in Figure 4.4 that the distribution of FoM values around this peak is rather broad, espcially for $\text{DLL}_{p\pi}$ cuts. Therefore we choose a looser selection of proton $\text{DLL}_{p\pi} > 14$, $\text{DLL}_{p\text{K}} > 8$ and antiproton $\text{DLL}_{p\pi} > 15$, $\text{DLL}_{p\text{K}} > 3$. This selection is applied to all $B^0_{(s)} \to p\overline{p}$ candidates and corresponds to mean PID efficiencies for $B^0 \to p\overline{p}$ events of $(48.81 \pm 0.08)\%$ (stat) in 2011 and $(46.47 \pm 0.06)\%$ (stat) in 2012 when applied to the full analysis sample. In comparison, the PID selection efficiency for the previous analysis, with different PID selection cuts, was $(43.97 \pm 0.22)\%$. This includes the caveat that the PID selection in that analysis was applied as the final stage of the offline selection chain whereas here we apply the PID selection before the MVA.

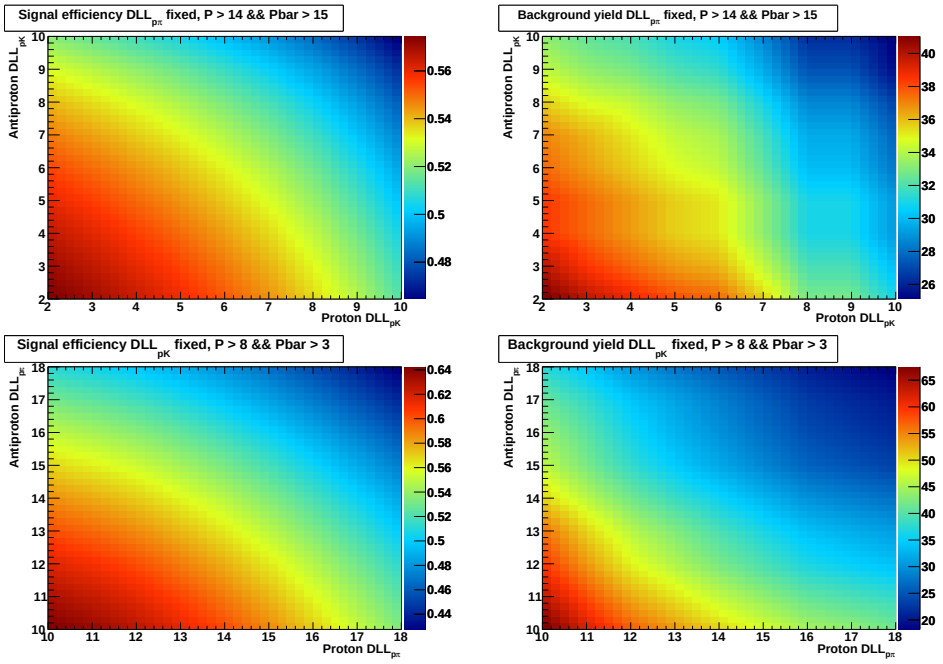


Figure 4.3: PID selection signal efficiencies (left) and estimated background yields (right) for fixed $\mathrm{DLL}_{p\pi}$ cuts (proton $\mathrm{DLL}_{p\pi} > 14$, antiproton $\mathrm{DLL}_{p\pi} > 15$, top) and fixed $\mathrm{DLL}_{p\mathrm{K}}$ cuts (proton $\mathrm{DLL}_{p\mathrm{K}} > 8$, antiproton $\mathrm{DLL}_{p\mathrm{K}} > 3$, bottom).

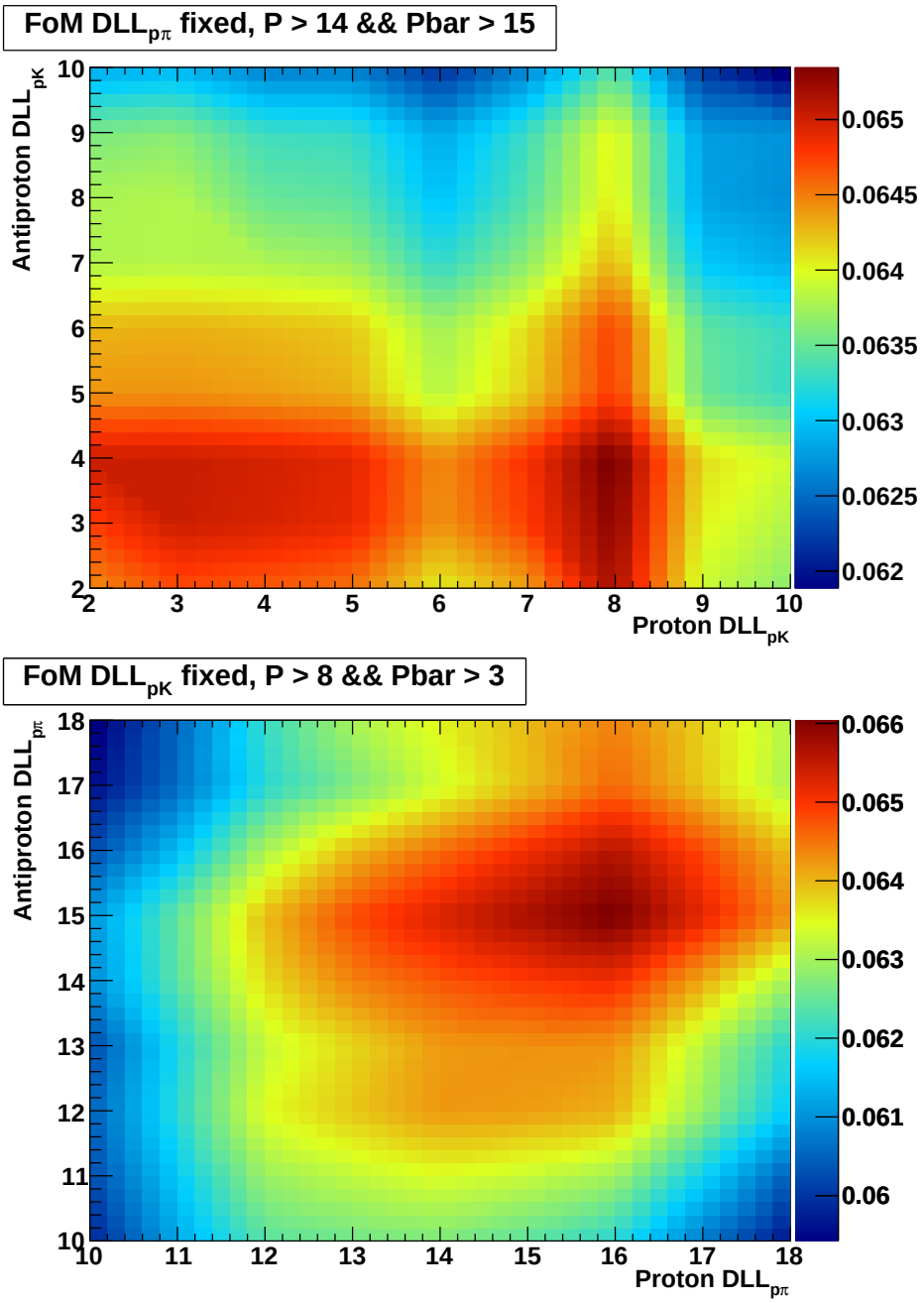


Figure 4.4: PID selection FoM for fixed $\mathrm{DLL}_{p\pi}$ cuts (proton $\mathrm{DLL}_{p\pi} > 14$, antiproton $\mathrm{DLL}_{p\pi} > 15$, top) and fixed $\mathrm{DLL}_{p\mathrm{K}}$ cuts (proton $\mathrm{DLL}_{p\mathrm{K}} > 8$, antiproton $\mathrm{DLL}_{p\mathrm{K}} > 3$, bottom).

Multivariate Analysis

The MVA selection is trained and applied after the PID selection as the final stage of the entire selection chain. Within this analysis the main aim of the MVA selection is to remove combinatorial background events whilst retaining as much of the $B^0_{(s)} \to p\overline{p}$ signals as possible. With these aims in mind, the MVA selection is trained using fully selected and MC truth matched $B^0 \to p\overline{p}$ MC events as the signal sample and sideband data as the background. The sideband data events are taken from across the entire $p\bar{p}$ invariant mass spectrum, excluding events within the blinded signal region, $[5230, 5417]\,\mathrm{MeV}/c^2$. To incorporate the effects of the PID selection in the signal MC samples, the per-event PID efficiencies obtained from the PIDCalib procedure are used to reweight the distributions of the signal MVA input variables. This is done instead of applying cuts to the MC PID variables which are known to be not accurately modelled in MC and would therefore result in inaccurate performance of the resulting MVA discriminator.

To validate the MVA procedure the input signal and background samples are each split into training and testing samples, where the MVA selection is trained using the training samples and then tested by applying the selection to the testing sample events. In the final analysis, training sample data events cannot be used within a selection which implements the MVA selection for which they were used to train. In the previous analysis 10% of the full, $1.078\,\mathrm{fb}^{-1}$, data set was used as the background training sample for the MVA selection. These events were then omitted from the final analysis data set leaving $0.92\,\mathrm{fb}^{-1}$ of data for the final analysis. For this analysis all data samples are split into two sets containing events with even or odd event numbers. Separate MVA selections are trained using either the odd- or even-numbered sideband samples and then applied to the opposite sample, i.e. An MVA selection trained on even event-numbered sideband data is then applied to odd event-numbered data and vice versa. With this method we retain 100% of the available data for the final analysis. The $B^0 \to p\overline{p}$ MC signal samples are also split randomly into two evenly sized samples for MVA training and testing. As well as training separate MVAs for odd- and even-numbered event numbers we also train separate MVAs for 2011 and 2012 events, giving in total four separate MVA selections used in this analysis, which are denoted 2011-Even, 2011-Odd, 2012-Even and 2012-Odd, where -Even or -Odd refers to odd or even the event numbers in the sideband data sample used for training.

Input Variables

Input variables for the MVA selection are chosen to maximise signal and background discrimination whilst avoiding strong correlations between each other. Variables with poor data and MC agreement are avoided. The chosen input variables are identical across the four MVA selections. Several potential input variables were considered, with ten variables making it into the final selections:

- ln(B Impact Parameter (IP) χ^2);
- ln(B vertex χ^2/ nDoF);
- ln(cosine(B direction angle)), known as the DIRA variable;
- A_{p_T}, the p_T asymmetry of the B within a cone of radius[1] $R = 1.0$ around the B, given by

$$A_{p_T} = \frac{{p_T}^B - {p_T}^{\text{cone}}}{{p_T}^B + {p_T}^{\text{cone}}} \quad , \tag{4.5}$$

  with ${p_T}^{\text{cone}}$ being the p_T of the vector sum of all tracks measured within the cone radius $R = 1.0$ around the B, except for the B-daughter particles;
- B-daughters' distance of closest approach (DOCA);
- Minimum of the daughters' ln(IP χ^2);
- Maximum of the daughters' ln(IP χ^2);
- Minimum of the daughters' p_T;
- Sum of the daughters' p_T;
- Minimum of the daughters' η.

Figures 4.5 and 4.6 show the distributions of the signal and background input variables used in the 2011-Even and 2012-Even MVA selections. Figures 4.7 and 4.8 show the corresponding distributions for the 2011-Odd and 2012-Odd selections. The signal sample distributions have been reweighted by their per-event PID efficiencies, extracted from the PID calibration procedure. Slight differences are observed between the 2011 and 2012 configurations mostly due to the increase in centre-of-mass energy. Figures 4.9 and 4.10 show the correlation matrices of the input variables for the four MVA selections.

[1]The cone radius is defined in (η, ϕ) as $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

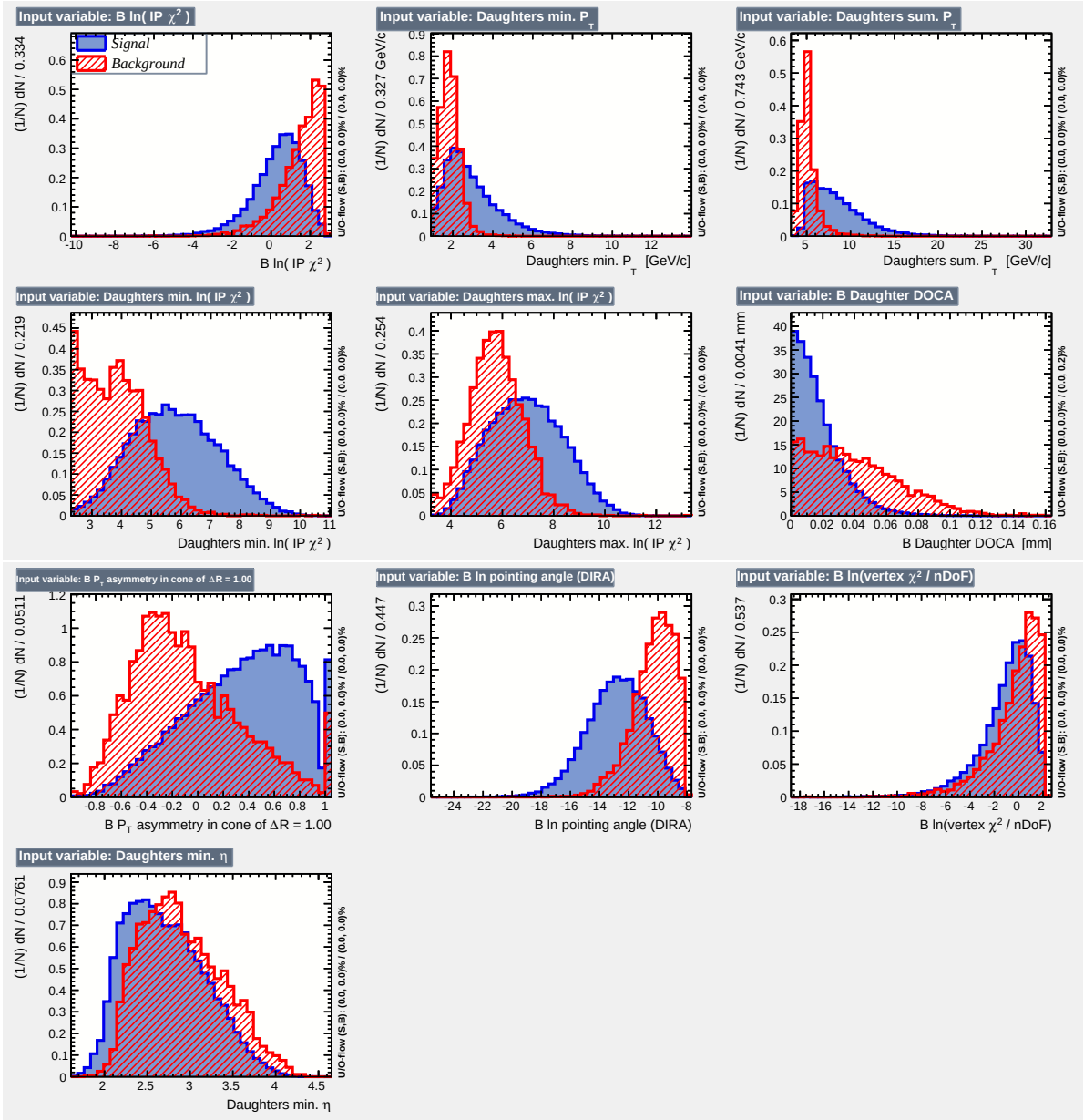


Figure 4.5: Distributions of the TMVA input discriminating variables for 2011 MC signal and even numbered event sideband data samples.

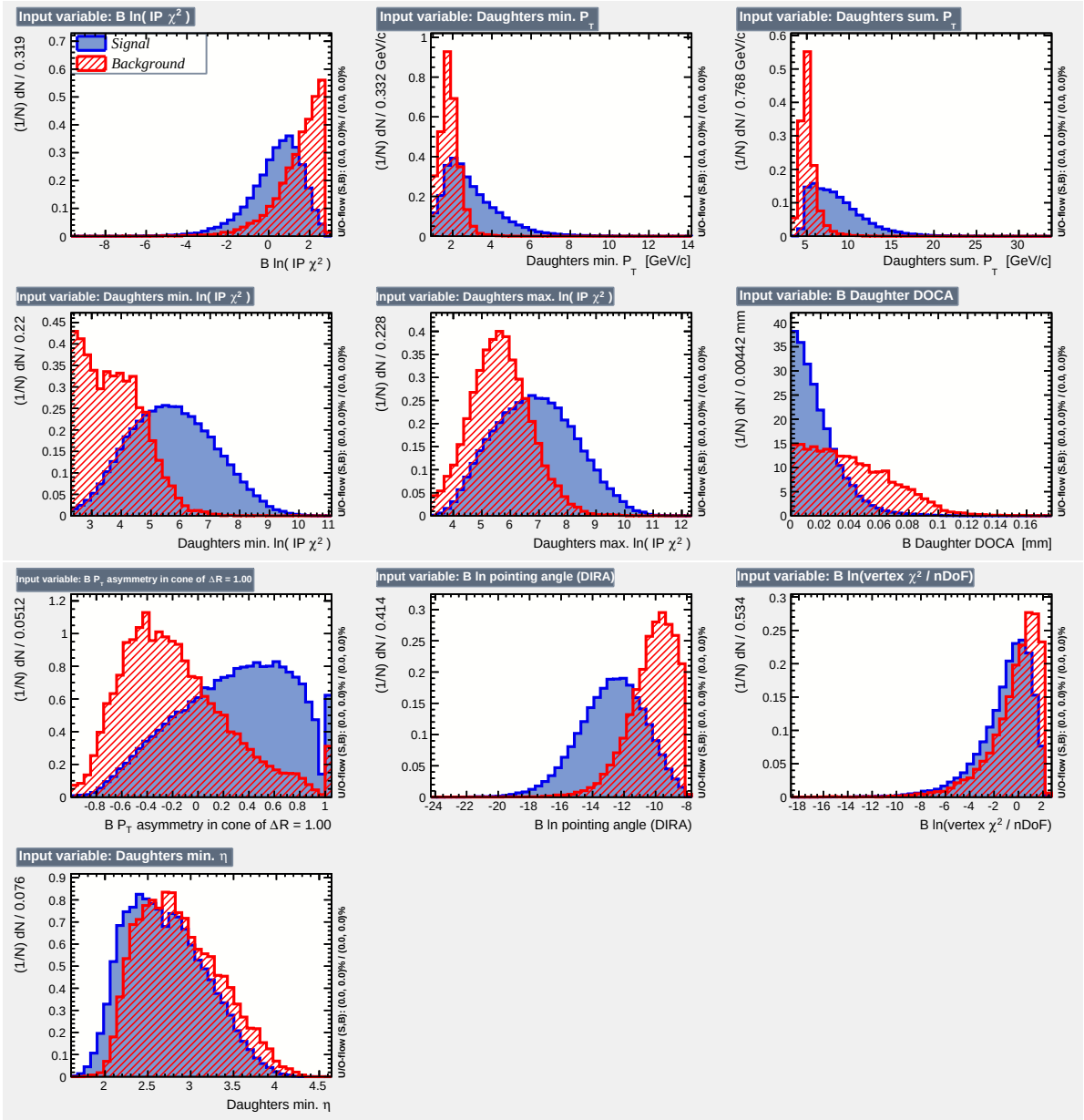


Figure 4.6: Distributions of the TMVA input discriminating variables for 2012 $B^0 \to p\overline{p}$ MC signal and even numbered event sideband data samples.

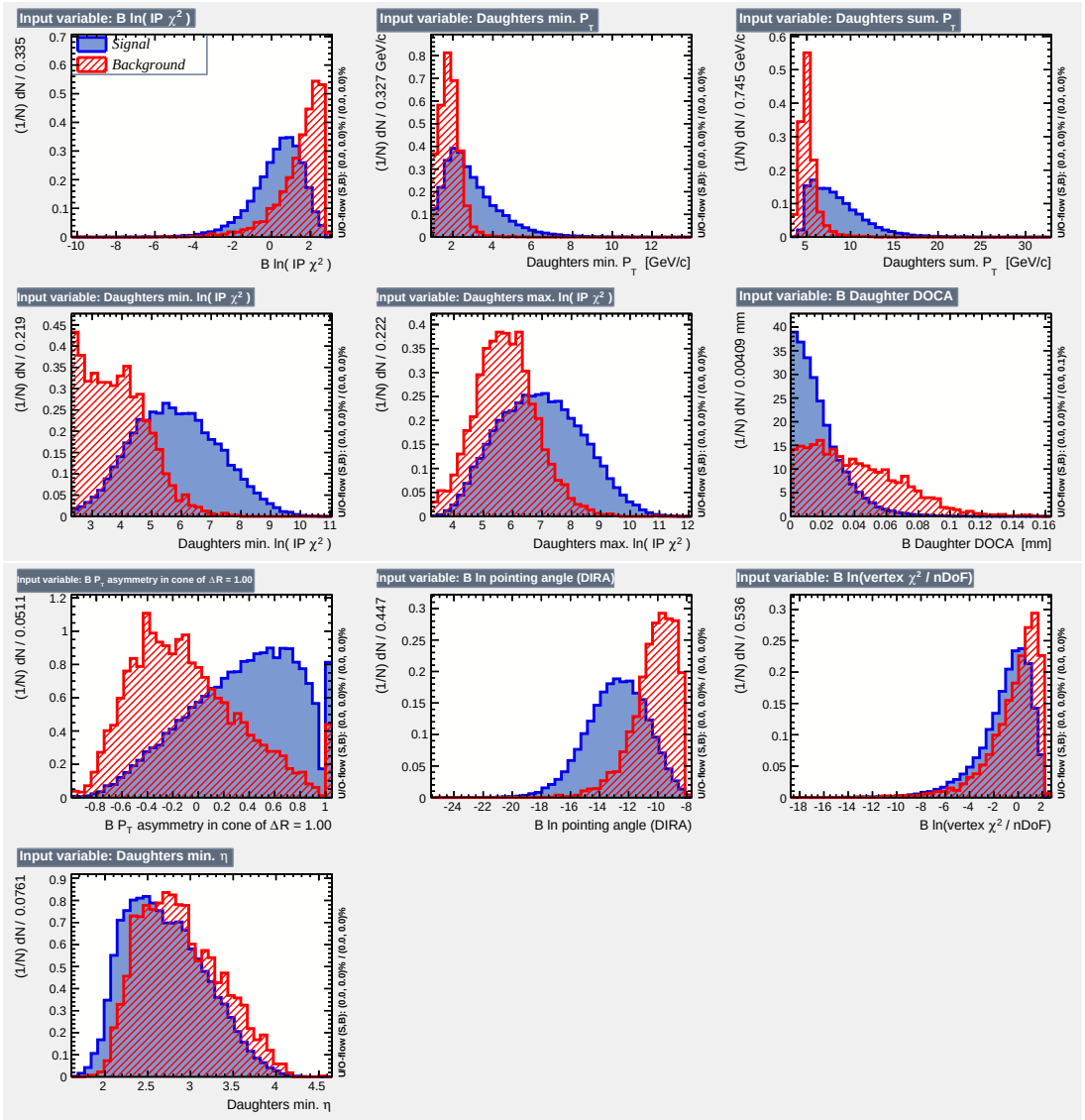


Figure 4.7: Distributions of the TMVA input discriminating variables for 2011 $B^0 \to p\bar{p}$ MC signal and odd numbered event sideband data samples.

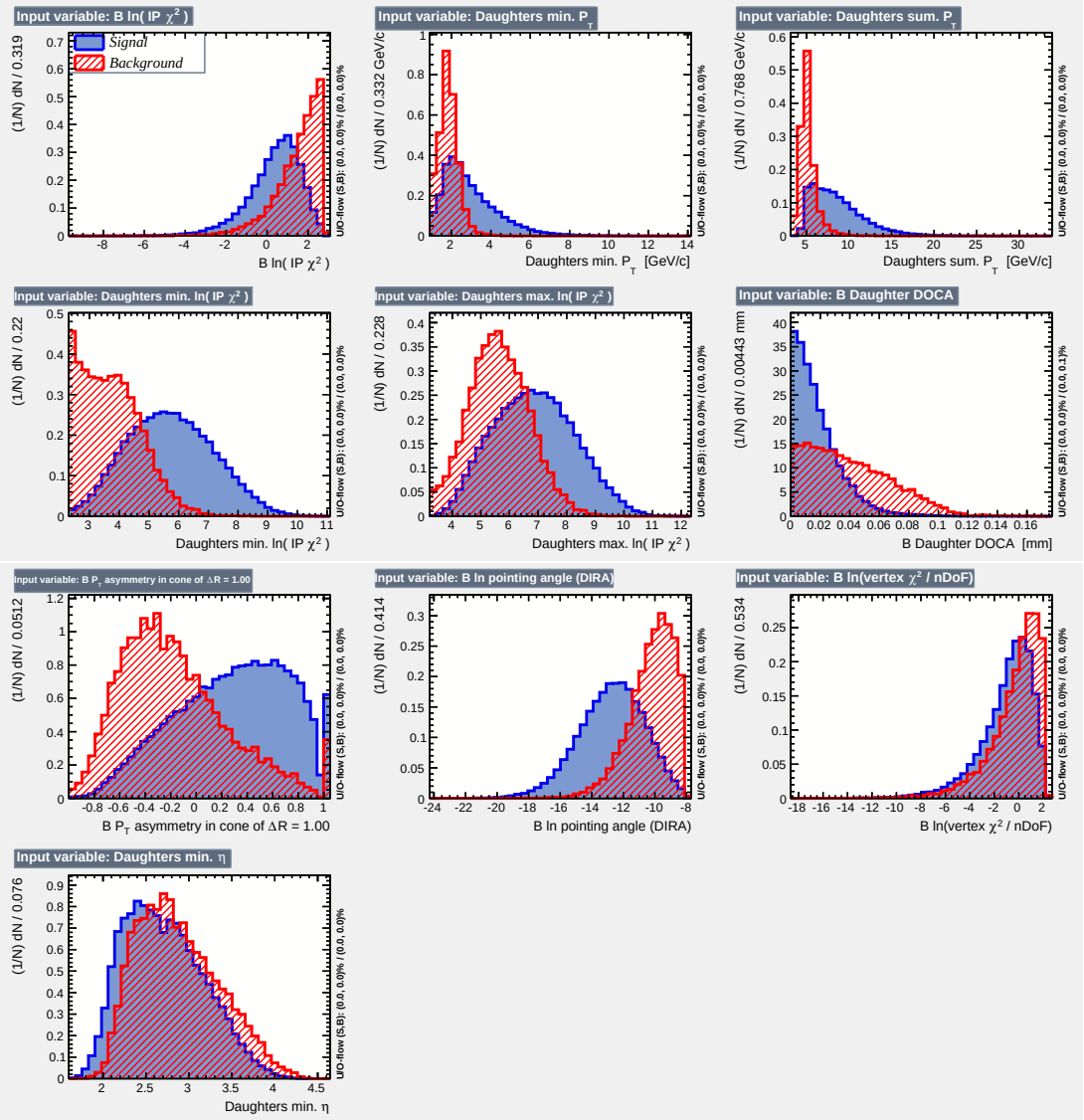


Figure 4.8: Distributions of the TMVA input discriminating variables for 2012 $B^0 \to p\bar{p}$ MC signal and odd numbered event sideband data samples.

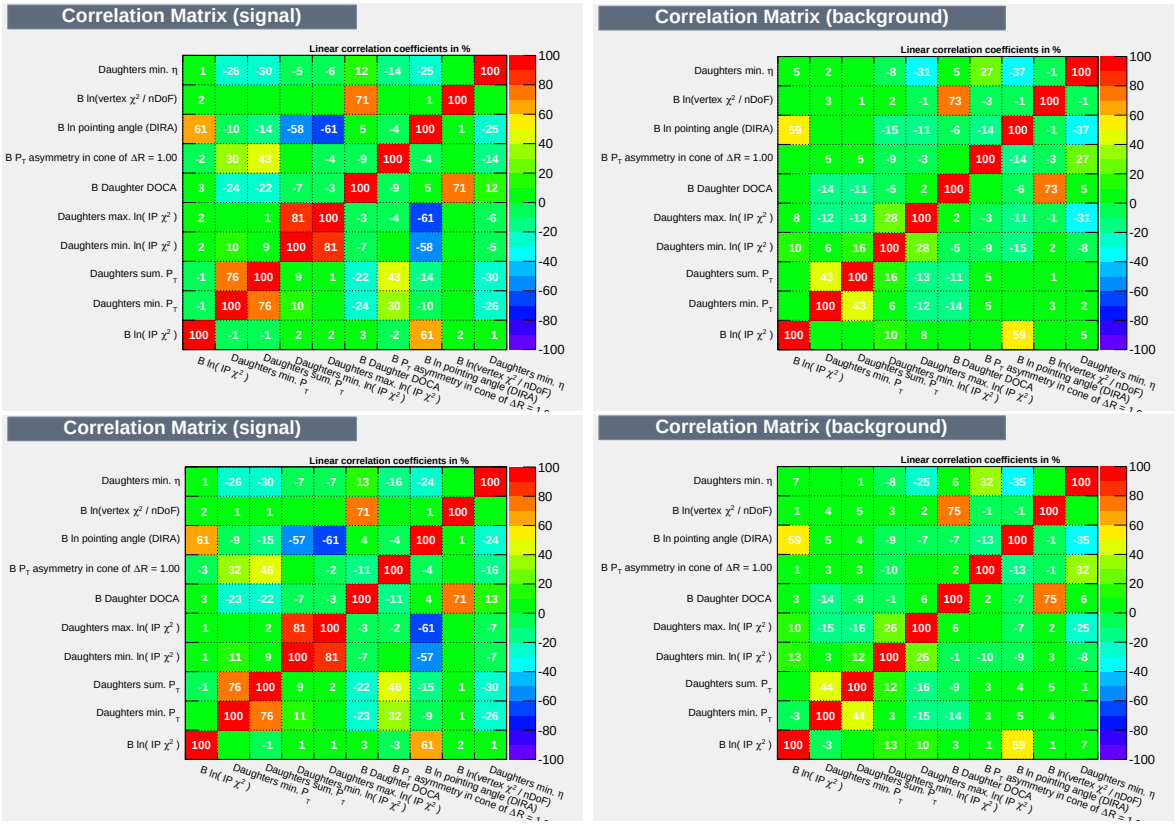


Figure 4.9: Correlation matrices of MVA input variables for $B^0 \to p\overline{p}$ signal MC (left column) and even numbered event sideband data (right column) samples for 2011 (top row) and 2012 (bottom row).

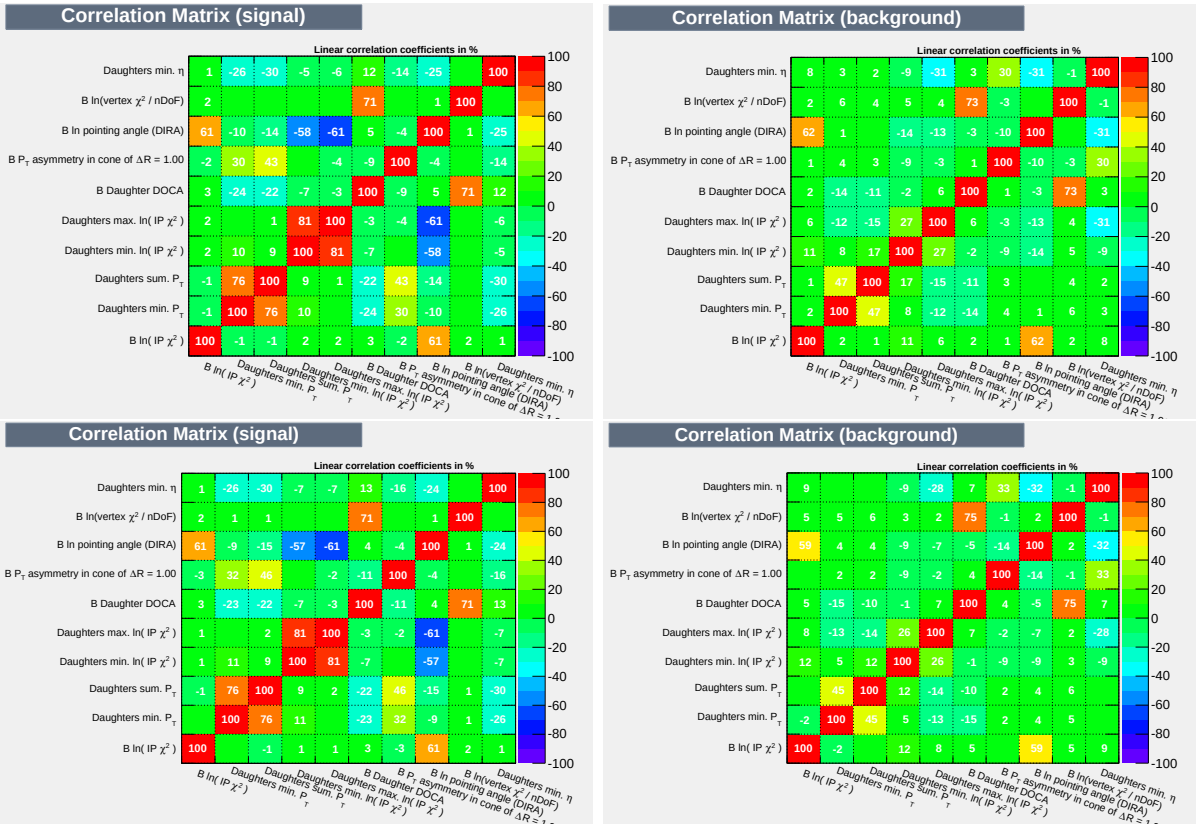


Figure 4.10: Correlation matrices of MVA input variables for $B^0 \to p\bar{p}$ signal MC (left column) and odd numbered event sideband data (right column) samples for 2011 (top row) and 2012 (bottom row).

MVA Training and Testing

The choice of MVA method was made after examining the signal efficiency versus background rejection power using the selected input variables. The aim is to select a method which retains the highest proportion of signal while rejecting as much background as possible. Figure 4.11 shows the background rejection rate versus signal efficiency for three of the best performing options available within TMVA for the 2012-Even configuration. The chosen method for all of the MVA selections in this analysis is the Multilayer Perceptron (MLP) Artificial Neural Network (ANN) option, as it offers the highest level of background rejection across a wide range of signal efficiencies relative to other available MVA methods.

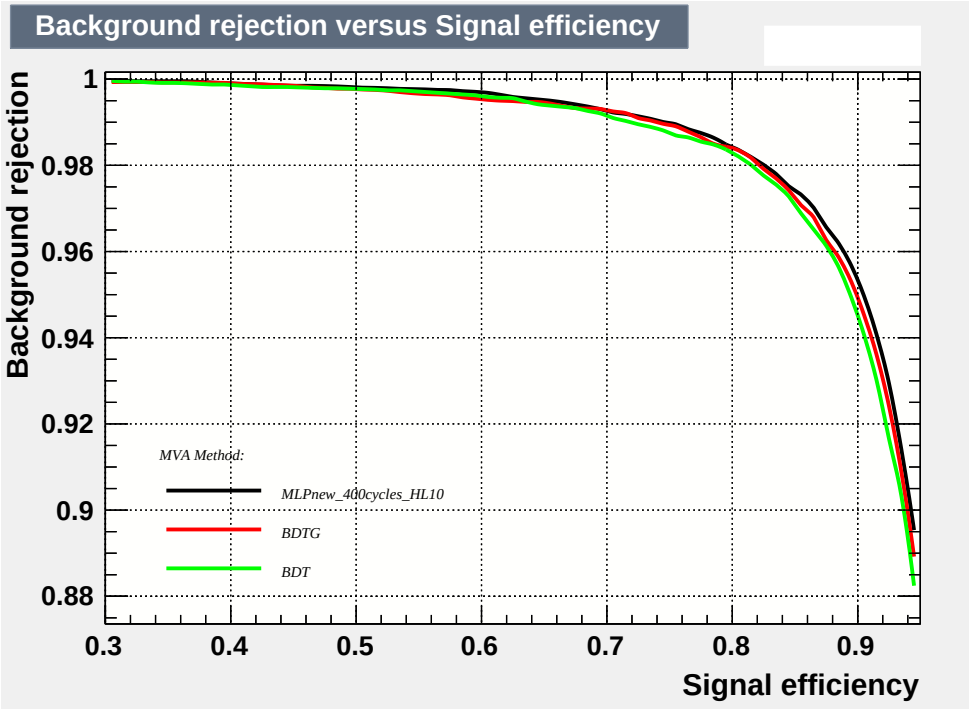


Figure 4.11: Background rejection versus signal efficiency curves for various MVA methods.

Figures 4.12 and 4.13 show comparisons between the MLP-responses of training and testing samples for the four MLP selections. No evidence of any significant overtraining is observed and the shapes of both signal and background distributions are consistent between odd and even trained selections within each year.

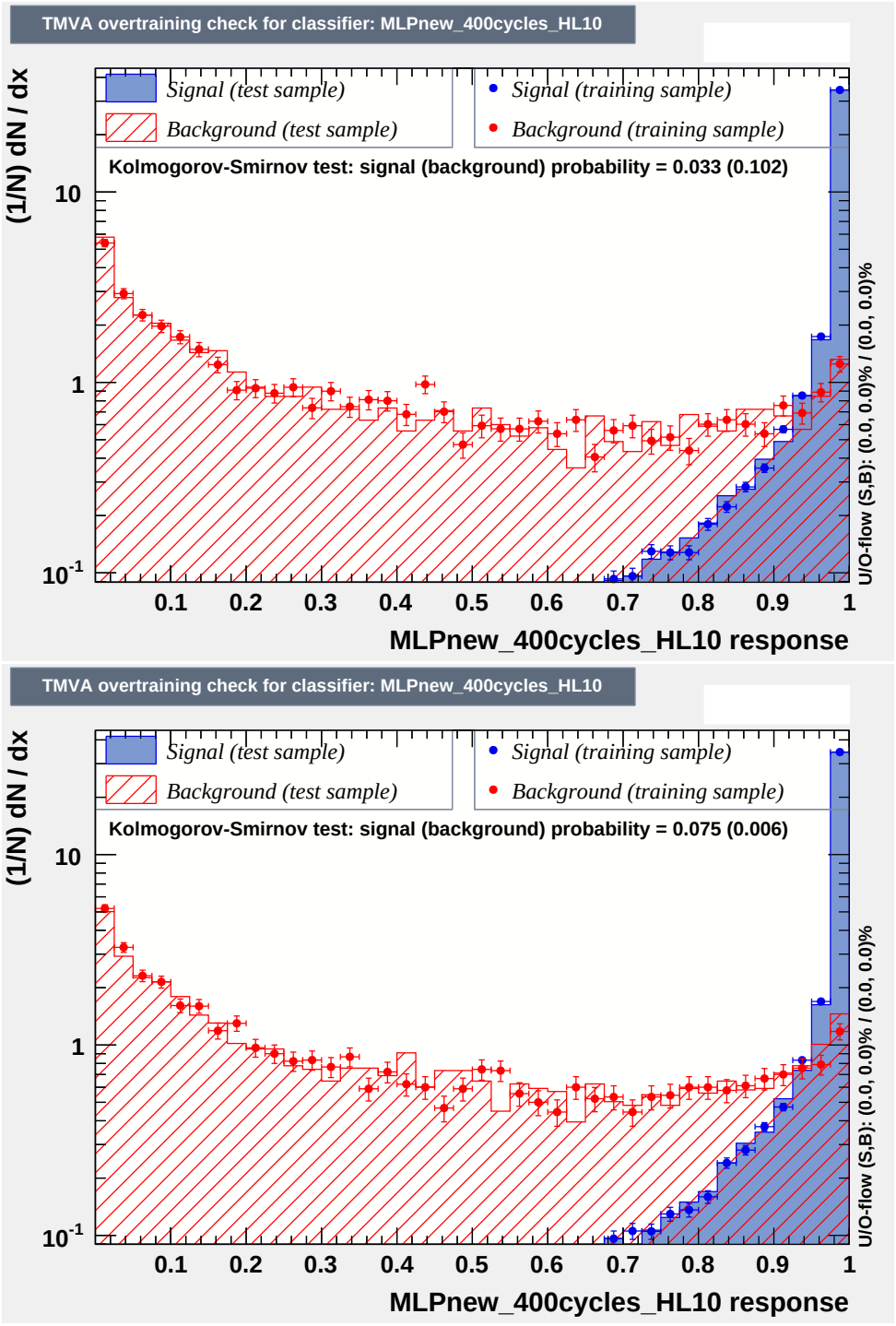


Figure 4.12: MLP ANN distributions and overtraining checks for $B^0 \to p\bar{p}$ MC signal and sideband data background samples for the training and testing samples for 2011-Even (top) and 2011-Odd (bottom).

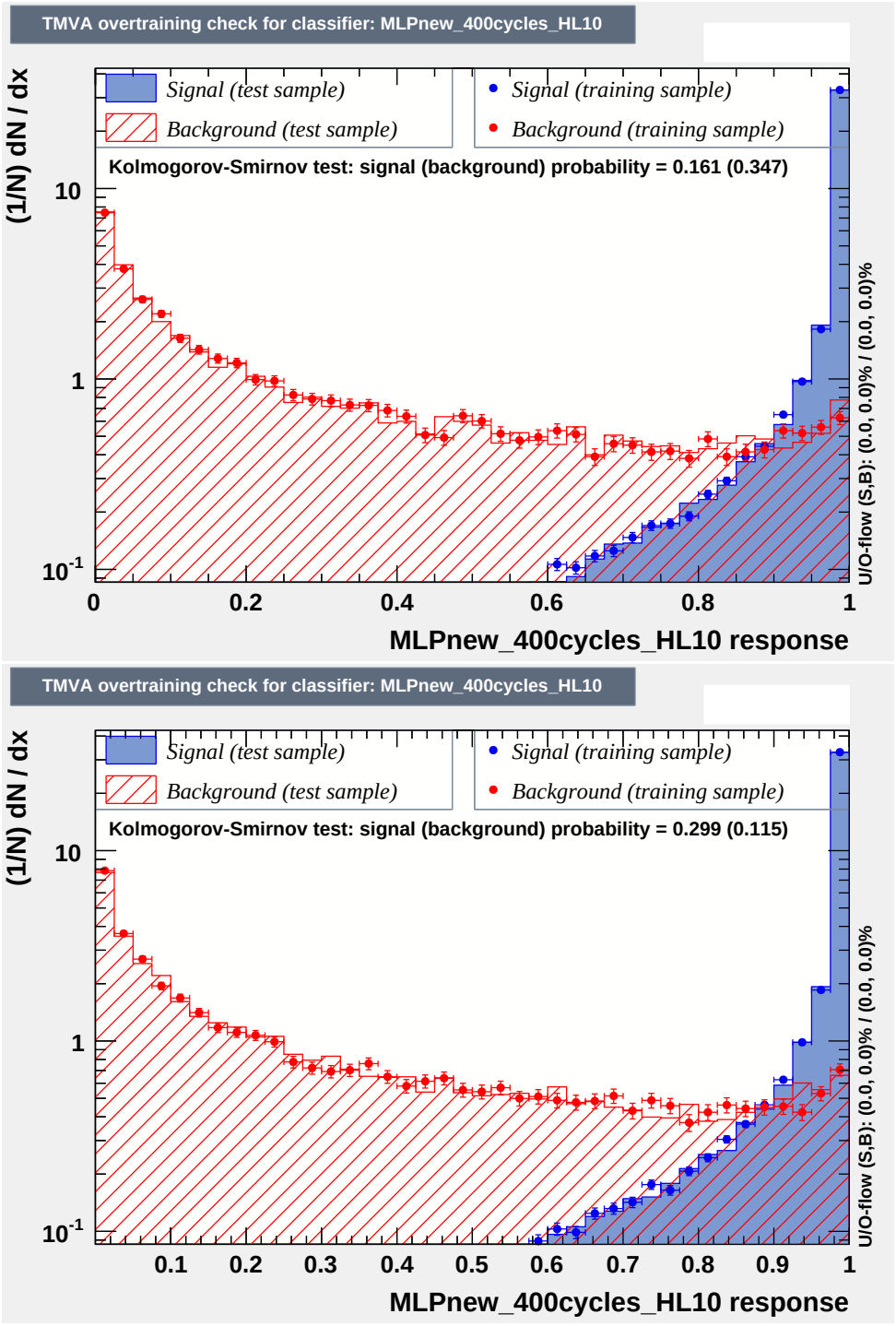


Figure 4.13: MLP ANN distributions and overtraining checks for $B^0 \to p\bar{p}$ MC signal and sideband data background samples for the training and testing samples for 2012-Even (top) and 2012-Odd (bottom).

MVA Optimisation

In a similar fashion to the PID optimisation, we optimise the MLP selection for selecting $B^0 \to p\overline{p}$ events using the Punzi figure of merit:

$$\text{FoM} = \frac{\epsilon^{\text{MVA/PID}}_{B^0 \to p\overline{p}}}{a/2 + \sqrt{B^{\text{MVA/PID}}}} \quad , \tag{4.6}$$

where $\epsilon^{\text{MVA/PID}}_{B^0 \to p\overline{p}}$ is the efficiency of the MLP selection cut with respect to $B^0 \to p\overline{p}$ MC events passing the PID selection and $B^{\text{MVA/PID}}$ is the estimated combinatorial background yield in the B^0 signal region after the MLP cut. Once again $a = 5$.

Separate optimisations are performed for 2011 and 2012 events; however, within each year, Odd and Even event trained MLP selections are combined to give a single FoM value. A range of MLP cut values are tested with FoM values calculated at each point, Figure 4.14 shows the distributions of FoM versus MLP cut value for the 2011 and 2012 samples. As the MLP response in the $B^0 \to p\overline{p}$ signal MC is shown to be dominated by events with values > 0.9, we choose to plot the MLP response as $|log_{10}(1 - MLP)|$ to better illustrate the region of interest for FoM studies. The optimal MLP cut values for 2011 and 2012 are found to be very close to one another at $|log_{10}(1 - MLP)| > 2.8$ and 2.7 respectively. These values are then used as the MLP selection requirements for this analysis and correspond to MLP response variable values of 0.998415 and 0.998005 in 2011 and 2012 respectively.

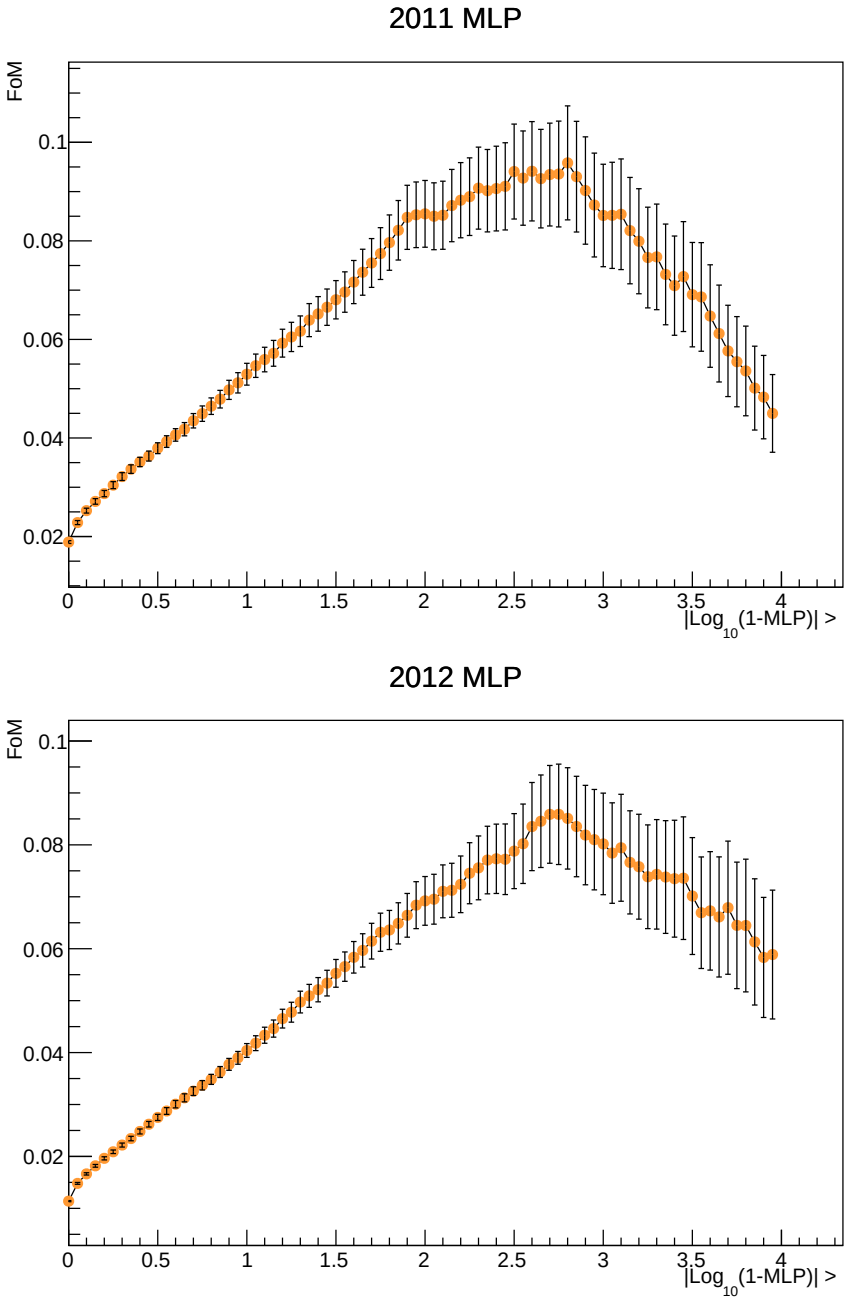


Figure 4.14: FoM versus MLP cut value for 2011 (top) and 2012 (bottom) events. Refer to the text for details.

4.3.5 Data After Full Selection

Figure 4.15 shows the $p\bar{p}$ invariant mass spectra for combined 2011 and 2012 sideband data events after each stage of the $B^0_{(s)} \to p\bar{p}$ selection. The blinded signal region $[5230, 5417]$ can be clearly seen, while Figure 5.11 shows the expected signal and background distribution after the full selection assuming $\mathcal{B}(B^0 \to p\bar{p}) = 1.47 \times 10^{-8}$ and $\mathcal{B}(B^0_s \to p\bar{p}) = 2.84 \times 10^{-8}$. Assuming these branching fractions, the estimated yields for $B^0 \to p\bar{p}$ and $B^0_s \to p\bar{p}$ after the full selection are:

$$\begin{aligned} N(B^0 \to p\bar{p}) &= 51.6 \pm 19.8 \text{ events}, \\ N(B^0_s \to p\bar{p}) &= 26.5 \pm 15.6 \text{ events}. \end{aligned}$$

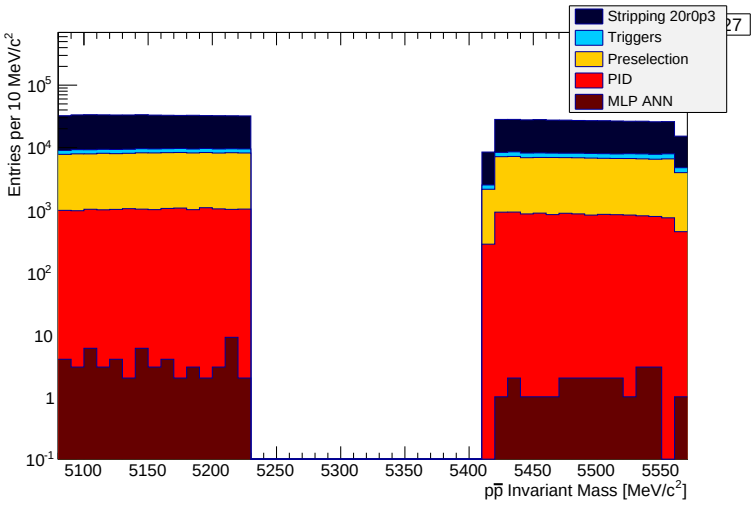


Figure 4.15: Sideband data $p\bar{p}$ invariant mass distribution after each selection stage.

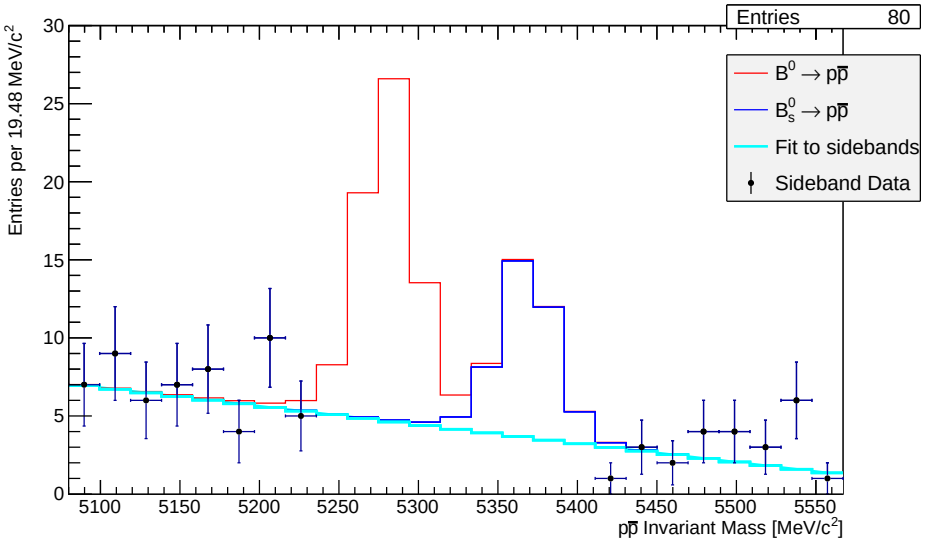


Figure 4.16: Invariant mass spectrum of $p\overline{p}$ sideband candidates after the full selection with fitted sideband mass shape and expected $B^0_{(s)} \to p\overline{p}$ signal yields ($B^0_{(s)} \to p\overline{p}$ shapes from MC).

4.4 Background Studies

In the previous analysis extensive background studies were performed to assess if there were any potential sources of background from specific physics decays which could peak within the $B^0_{(s)} \to p\overline{p}$ signal region. No such decays were found, with the combined sum of all studied backgrounds contributing a smooth mass spectrum which was accommodated within the combinatorial background shape. For this analysis it is necessary to reinvestigate the possible sources of background as the data taking, reconstruction and selection conditions are changed from the previous analysis.

As with the previous analysis, the groups of decays of most interest are two- and three-body hadronic b-hadron decays, $H_b \to h^+h'^-$ and $B \to hh'h''$, and many-body hadronic Λ^0_b decays. Table 4.6 lists the relevant background channels and their corresponding group. For each decay, the full selection, as detailed in Section 4.3, was applied to each MC sample in stages from the stripping selection through to the final MVA selection. After each stage, stacked plots were made of the full $p\overline{p}$ invariant mass spectrum with the integral, I_H, of each decay histogram set to

$$I_H = f_q \cdot \mathcal{B}_{vis} \cdot \epsilon_{\text{stage}} \tag{4.7}$$

where f_q represents the b hadronisation probability to the relevant hadron, $\mathcal{B}_{\text{vis}}$ stands for the visible branching fraction and ϵ_{stage} is the total selection efficiency for the decay after the given selection stage. This gives a simplified calculation of the expected yield, not taking into account the integrated luminosities or $b\overline{b}$ cross sections[2]. Figures 4.17 and 4.18 show the stacked $p\overline{p}$ invariant mass spectra after the offline preselection stage and the full $B^0_{(s)} \to p\overline{p}$ selection, respectively.

Figure 4.17 shows that a significant amount of specific background events are present within the $p\overline{p}$ invariant mass spectrum after the stripping, trigger and offline preselection cuts are applied. However, Figure 4.18 shows that when the PID and MVA selections are applied, these contaminating backgrounds become negligible compared to the expected $B^0_{(s)} \to p\overline{p}$ signals and show no evidence of any defined structure beyond a smooth, flat distribution across the full mass spectrum which is overwhelmed by the combinatorial background contribution. Therefore, all specific physics backgrounds will be considered to be modelled as a part of the comibinatorial background in the subsequent $p\overline{p}$ invariant mass fit.

[2]The full equation for calculating the expected yield of a b-hadron decay is shown in Eq. 3.1.

Signal Modes	$\mathcal{B}(\times 10^{-6})$
$B_s^0 \to p\overline{p}$	$2.84\,^{+2.20}_{-1.69} \times 10^{-2}$ [1]
$B^0 \to p\overline{p}$	$1.47\,^{+0.71}_{-0.53} \times 10^{-2}$ [1]

$H_b \to h^+h'^-$ *Modes*	$\mathcal{B}(\times 10^{-6})$
$B_s^0 \to K^+K^-$	25.4 ± 3.7 [71]
$B^0 \to K^+\pi^-$	$19.57\,^{+0.53}_{-0.52}$ [71]
$\Lambda_b^0 \to pK^-$	6.2 ± 1.9 [2]
$B_s^0 \to \pi^+K^-$	5.38 ± 0.76 [77]
$B^0 \to \pi^+\pi^-$	5.11 ± 0.22 [71]
$\Lambda_b^0 \to p\pi^-$	4.0 ± 1.3 [2]
$B_s^0 \to \pi^+\pi^-$	0.73 ± 0.14 [71]
$B^0 \to \Lambda\overline{\Lambda}$	0.32 (< 0.32 @ 90% CL [2])
$B^0 \to K^+K^-$	0.11 ± 0.78 [77]

$B \to hh'h''$ *Modes*	$\mathcal{B}(\times 10^{-6})$
$B^0 \to \pi^+\pi^-\pi^0$	$50^{(*)}$ (< 720 @ 90% CL [2])
$B^+ \to (J/\psi \to p\overline{p})\pi$	40.4 ± 1.7 [71]
$B^0 \to K^+\pi^-\pi^0$	37.8 ± 3.2 [71]
$B^+ \to K^+K^-K^+$	32.5 ± 1.5 [71]
$B^0 \to K_S^0\pi^+\pi^-$	25.9 ± 1.0 [71]
$B_s^0 \to K^+K^-\pi^0$	$20^{(*)}$
$B^+ \to \pi^+\pi^-K^+$	16.3 ± 2.0 [71]
$B^+ \to \pi^+\pi^-\pi^+$	15.2 ± 1.4 [71]
$B^+ \to p\overline{p}K^+$	5.48 ± 0.34 [71]
$B^+ \to \pi^+K^-K^+$	5.0 ± 0.7 [71]
$B_s^0 \to K^-\pi^+\pi^0$	$5^{(*)}$
$B^0 \to p\pi^-\overline{\Sigma^0}$	3.8 (< 3.8 @ 90% CL [2])
$B^+ \to p\overline{\Lambda}\pi^0$	$3.0\,^{+0.61}_{-0.62}$ [71]
$B^+ \to p\overline{\Lambda}\gamma$	$2.4\,^{+0.5}_{-0.4}$ [2]
$B^0 \to K^+K^-\pi^0$	2.17 ± 0.65 [71]
$B^+ \to p\overline{p}\pi^+$	$1.60\,^{+0.18}_{-0.17}$ [71]

Many-body Λ_b^0 *Decays*	$\mathcal{B}(\times 10^{-6})$
$\Lambda_b \to (\Lambda_c \to pK_S^0(\pi^0 \to \gamma\gamma))\pi^-$	209 ± 33 [2]
$\Lambda_b \to (\Lambda_c \to pK_S^0)\pi^-$	151 ± 19 [2]
$\Lambda_b \to \Lambda\pi^+\pi^-$	$25.9^{(*)}$
$\Lambda_b \to \Lambda K^+K^-$	$9.97^{(*)}$
$\Lambda_b \to (D^0 \to K_S^0 \to \pi^+\pi^-)pK^-$	1.37 ± 0.27 [2]
$\Lambda_b \to \Lambda p\overline{p}$	$1.33^{(*)}$
$\Lambda_b \to p\pi^-K^-\pi^+$	$0.1^{(*)}$

Table 4.6: Potential $B_{(s)}^0 \to p\overline{p}$ backgrounds, sorted by background type and branching fraction. Decays marked with $^{(*)}$ indicate that the branching fraction is not known and a value estimated from a similar decay ($B^0 \to K_S^0 h^+h^-$) was used.

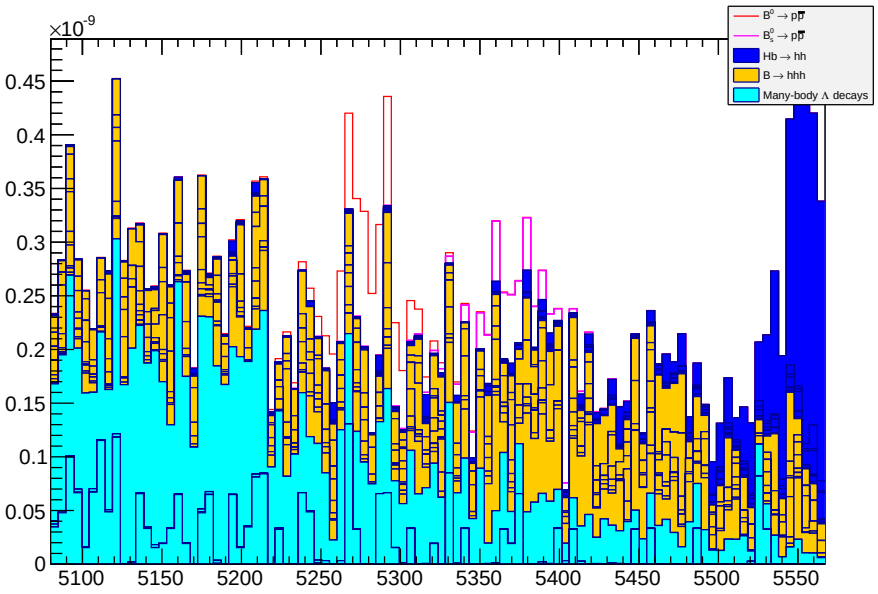


Figure 4.17: Stacked plot of MC $p\overline{p}$ invariant mass spectrum after the offline preselection stage, showing the expected relative "yields" for $B^0_{(s)} \to p\overline{p}$ signals and physics background processes which are colour grouped by decay type.

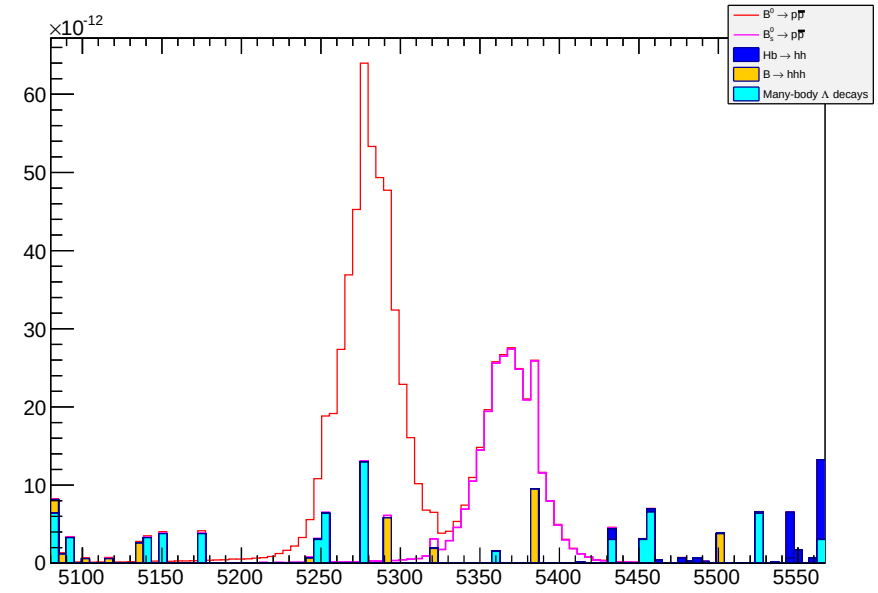


Figure 4.18: Stacked plot of MC $p\overline{p}$ invariant mass spectrum after the full $B^0_{(s)} \to p\overline{p}$ selection, showing the expected relative "yields" for $B^0_{(s)} \to p\overline{p}$ signals and physics background processes which are colour grouped by decay type.

4.5 $B^0 \to K^+\pi^-$ Normalisation Channel

As outlined in Section 4.1, the measurement of the branching fractions of $B^0_{(s)} \to p\overline{p}$ is made relative to that of $B^0 \to K^+\pi^-$, the normalisation channel. The normalisation channel samples are split by charge state resulting in separate samples of $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ candidates. This charge state splitting provides a natural method for calculating the normalisation channel selection efficiencies while taking into account the known $B^0 - \overline{B^0}$ production asymmetries [69]. The two charge-state separated samples are then fitted separately with the final $B^0 \to K^+\pi^-$ yield taken as the sum of the extracted $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ yields. Having separate fits for $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ candidates also has an added effect of improving the invariant mass fit performance relative to a single fit of the combined spectra.

The selections of $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ candidates are identical. We hereafter detail the selection of $B^0 \to K^+\pi^-$ candidates on a data sample of $3.117\,\text{fb}^{-1}$.

4.5.1 Trigger Selection

The same trigger requirements are imposed on the normalisation channel $B^0 \to K^+\pi^-$ as for the $B^0_{(s)} \to p\overline{p}$ signal channels, see Table 4.3. In this way the ratio of trigger efficiencies entering the relative branching fraction calculation cancel to a large extent. The trigger selection retains 48.9% of the $B^0 \to K^+\pi^-$ candidates that pass the stripping selection.

4.5.2 Stripping Selection

For selecting $B^0 \to K^+\pi^-$ events a two-body hadronic B^0 decay stripping line is used which reconstructs events under the $\pi^+\pi^-$ hypothesis. The list of cuts in the stripping line is shown in Table 4.7.

This exact stripping line is not present in the same stripping configuration as the $B^0_{(s)} \to p\overline{p}$ stripping line so the $B^0 \to K^+\pi^-$ data is taken from a different stripping configuration but one which has identical experimental conditions as the $B^0_{(s)} \to p\overline{p}$ data. In the 2012 recorded data there is a $0.005\,\text{fb}^{-1}$ difference in integrated luminosity between the two stripping generation data sets. This results in an overall integrated luminosity of $3.117\,\text{fb}^{-1}$ in the $B^0 \to K^+\pi^-$ candidate data sample compared to $3.122\,\text{fb}^{-1}$ for the stripping 20r0p3 $B^0_{(s)} \to p\overline{p}$ data. This $0.005\,\text{fb}^{-1}$ difference is taken into account in the relative branching fraction calculations.

Variable	Cut value
Daughters' track fit χ^2/ nDoF	< 3
Min. of daughters' p_T	> 1100 MeV/c
Max. of combined daughters' p_T	> 2700 MeV/c
Min. of daughters' IP	> 0.15 mm or $\chi^2 > 100$
Max. of daughters' IP	> 0.27 mm or $\chi^2 > 200$
Distance of closest approach of daughters (DOCA)	< 0.08 mm
Daughter Ghost Probability	< 0.8
B p_T	> 1200 MeV/c
B primary vertex IP	< 0.08 mm or $\chi^2 < 12$
B decay time τ	> 0.6 ps
B reconstructed mass	$4800 - 5800$ MeV/c^2

Table 4.7: $B^0 \to K^+\pi^-$ stripping selection cuts.

4.5.3 PID and Multivariate Selection

Events passing the stripping selection detailed in Section 4.5.2 are refitted under the $K\pi$ mass hypothesis using the `SubstitutePID` algorithm.

PID DLL cuts are used to identify K and π daughter tracks and are listed in Table 4.8. These cuts were optimised for 2012 data using the `PIDCalib` package. The optimisation method compares the PID selection efficiencies for $B^0 \to K^+\pi^-$ signal and $B_s^0 \to K^+K^-$, $B^0 \to \pi^+\pi^-$, $\Lambda_b^0 \to pK^-$ and $\Lambda_b^0 \to p\pi^-$ misidentified backgrounds, and finds a PID selection that maximises the $B^0 \to K^+\pi^-$ selection efficiency whilst minimising the efficiencies of the misidentified backgrounds.

π daughter	Cut value
$\mathrm{DLL_{K\pi}}$	< -4
$\mathrm{DLL}_{p\pi}$	$< +3$
$\mathrm{DLL}_{e\pi}$	< 0
$\mathrm{DLL}_{\mu\pi}$	< 1
K daughter	
$\mathrm{DLL_{K\pi}}$	$> +5$
$\mathrm{DLL_{K}}_p$	> 0
$\mathrm{DLL_{K}}_e$	> 0
$\mathrm{DLL_{K}}_\mu$	> -1

Table 4.8: $B^0 \to K^+\pi^-$ PID selection cuts.

An MVA selection is applied to $B^0 \to K^+\pi^-$ events using the same MLP ANN selections trained to select $B^0 \to p\overline{p}$ events and detailed in Section 4.3.4. Figure 4.19 shows the MLP responses for combined 2011 and 2012 $B^0 \to K^+\pi^-$ MC and high-mass ($m_{K\pi} > 5500\,\text{MeV}/c^2$) $B^0 \to K^+\pi^-$ "sideband" data, which is dominated by combinatorial background events. The figure shows a strong separation in MLP response shape between $B^0 \to K^+\pi^-$ signal events and the combinatorial background, and gives clear motivation to use the $B^0 \to p\overline{p}$ trained MLP selections in the $B^0 \to K^+\pi^-$ event selection.

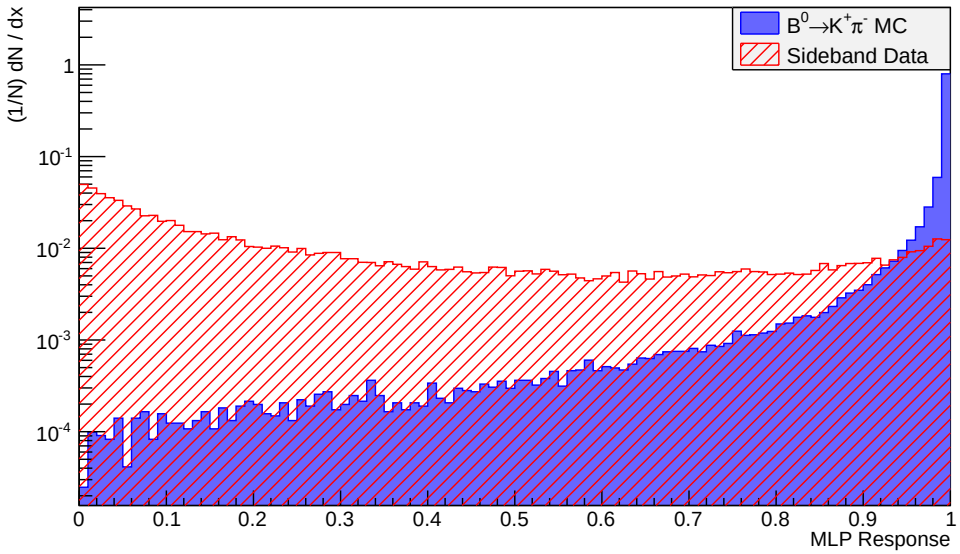


Figure 4.19: MLP response disributions for $B^0 \to K^+\pi^-$ MC (blue) and sideband data (red).

The choice of MLP cut for the $B^0 \to K^+\pi^-$ selection is optimised via invariant mass fits of the full $B^0 \to K^+\pi^-$ data set. Over a range of MLP cut values the $B^0 \to K^+\pi^-$ candidate data set is fitted over the $K\pi$ invariant mass range $[5200, 5830]$ MeV/c^2 with a fit consisting of three components: individual Gaussian functions for the $B^0 \to K^+\pi^-$ and $B_s^0 \to \pi^+K^-$ signals and a one-dimensional polynomial to describe the combinatorial background shape. For each cut value the statistical significance, $\frac{S}{\sqrt{S+B}}$, where S and B are the number of signal and background respectively, of the fitted $B^0 \to K^+\pi^-$ signal around the fitted B^0 mass mean ($\pm 50\,\text{MeV}/c^2$) is calculated with the highest significance found at MLP response > 0.855. A second set of fits over the MLP values is then performed this time with the $B^0 \to K^+\pi^-$ and $B_s^0 \to \pi^+K^-$ signal shapes modelled by double Crystal Ball (DCB) functions with parameters taken from fits to MC samples after applying the

highest-significance MLP cut from the Gaussian fit stage. The optimal cut is found at MLP response > 0.815 and this value is chosen as the $B^0 \to K^+\pi^-$ final selection MLP response cut. Figure 4.20 shows the $B^0 \to K^+\pi^-$ signal significance for a range of MLP response cut values.

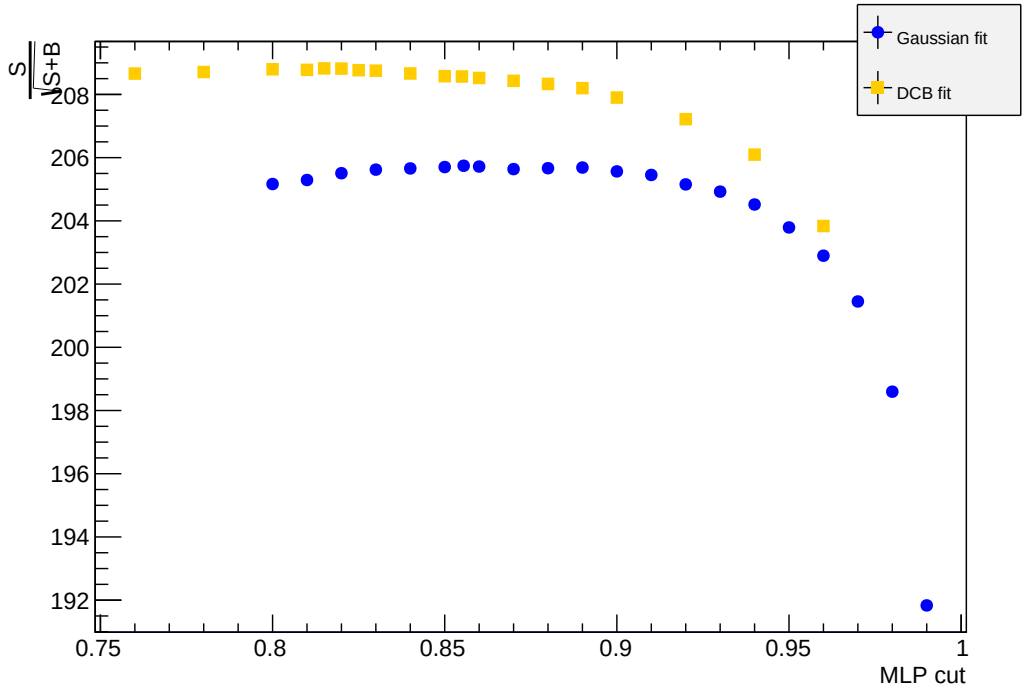


Figure 4.20: Statistical significance of fitted $B^0 \to K^+\pi^-$ signal across a range of MLP cut values for the $B^0 \to K^+\pi^-$ signal shape fitted with Gaussian (blue) and double Crystal Ball (yellow) PDFs.

Chapter 5

Selection Efficiencies and Invariant Mass Fits

5.1 Efficiency Calculations

The calculation of the $B^0_{(s)} \to p\overline{p}$ branching fractions or upper limits necessitates the determination of the ratio of efficiencies with respect to the normalisation channel, *cf.* Equation 3.2. This ratio can be explicitly decomposed in its "constituents" as

$$\frac{\epsilon^{tot}_{B^0\to K^+\pi^-}}{\epsilon^{tot}_{B^0_{(s)}\to p\overline{p}}} = \frac{\epsilon^{\text{gen}}_{B^0\to K^+\pi^-}}{\epsilon^{\text{gen}}_{B^0_{(s)}\to p\overline{p}}} \cdot \frac{\epsilon^{\text{strip+rec/gen}}_{B^0\to K^+\pi^-}}{\epsilon^{\text{strip+rec/gen}}_{B^0_{(s)}\to p\overline{p}}} \cdot \frac{\epsilon^{\text{trig/strip+rec}}_{B^0\to K^+\pi^-}}{\epsilon^{\text{trig/strip+rec}}_{B^0_{(s)}\to p\overline{p}}} \cdot \frac{\epsilon^{\text{presel/trig}}_{B^0\to K^+\pi^-}}{\epsilon^{\text{presel/trig}}_{B^0_{(s)}\to p\overline{p}}} \cdot \frac{\epsilon^{\text{PID/presel}}_{B^0\to K^+\pi^-}}{\epsilon^{\text{PID/presel}}_{B^0_{(s)}\to p\overline{p}}} \cdot \frac{\epsilon^{\text{MVA/PID}}_{B^0\to K^+\pi^-}}{\epsilon^{\text{MVA/PID}}_{B^0_{(s)}\to p\overline{p}}} \quad , \tag{5.1}$$

where the different terms refer subsequently to efficiencies on the LHCb acceptance, stripping plus reconstruction, trigger, preselection, PID and MVA criteria. The determination of the various contributing factors from MC is detailed in the subsections below.

Most efficiencies are extracted from simulated events, calculated using MC truth matched candidates, hence the method relies on a good agreement between the distributions in data and MC. No comparison can be made for the $B^0_{(s)} \to p\overline{p}$ decays as the signal region is blinded, so the distributions of the kinematically similar $B^0 \to K^+\pi^-$ decay are studied. Appendix B shows *sPlot*s extracted from data of key variables for the $B^0 \to K^+\pi^-$ decay compared with the MC distributions of the same variables. There is good agreement between data and simulated events. All efficiencies presented in this section are shown with statistical uncertainties only.

5.1.1 Generator Efficiencies

The generator-level efficiencies for all daughters to satisfy $10 < \theta < 400\,\text{mrad}$ – averaged over the magnet up and down samples – for all the relevant decay modes are collected in Table 5.1. They indicate that the efficiencies for the signal daughters to be in the LHCb acceptance are rather similar, though roughly 10% higher for the $p\overline{p}$ final states in both 2011 and 2012. This is consistent with what was seen in the previous analysis.

Decay mode	2011 $\epsilon_{\text{gen}}(\%)$	2012 $\epsilon_{\text{gen}}(\%)$
$B^0 \to p\overline{p}$	20.34 ± 0.05	20.62 ± 0.05
$B^0_s \to p\overline{p}$	20.30 ± 0.04	20.60 ± 0.04
$B^0 \to K^+\pi^-$	18.77 ± 0.05	18.73 ± 0.05
$\overline{B^0} \to \pi^+K^-$	18.77 ± 0.05	18.73 ± 0.05

Table 5.1: Generator efficiencies obtained from the MC samples.

5.1.2 Stripping and Reconstruction Efficiencies

The combined stripping plus reconstruction efficiencies for $B^0_{(s)} \to p\overline{p}$ and $B^0 \to K^+\pi^-$ are calculated from MC and are listed in Table 5.2. $B^0_{(s)} \to p\overline{p}$ samples are not split by charge state and therefore the truth matching requirements combine both charge conjugate states resulting in high truth matching efficiencies of 94%. As explained in Section 4.5, the normalisation channel samples are split by charge states which are truth matched separately resulting in much lower truth matching efficiencies than $B^0_{(s)} \to p\overline{p}$. The $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ samples are produced with a 10% production asymmetry applied between them, to mimic the observed $B^0 \to K^+\pi^-$ CP asymmetry. This asymmetry is the cause of the observed difference in stripping and reconstruction efficiencies between the two modes.

In all cases it is observed that the efficiencies for 2011 are slightly higher than those for 2012 ($<$ 10% difference) and the efficiency for $B^0_{(s)} \to p\overline{p}$ is greater than that of $B^0 \to K^+\pi^-$.

5.1.3 Trigger Efficiencies

The total trigger selection efficiencies for $B^0_{(s)} \to p\overline{p}$ and $B^0 \to K^+\pi^-$ events are shown in Table 5.3. The efficiencies are calculated using MC and show that the trigger efficiencies for $B^0 \to K^+\pi^-$ are higher than for $B^0_{(s)} \to p\overline{p}$, which is consistent with what was seen in

Decay mode	2011 $\epsilon_{\text{strip+rec/gen}}(\%)$	2012 $\epsilon_{\text{strip+rec/gen}}(\%)$
$B^0 \to p\overline{p}$	24.32 ± 0.05	22.51 ± 0.04
$B_s^0 \to p\overline{p}$	24.41 ± 0.05	23.14 ± 0.04
$B^0 \to K^+\pi^-$	11.44 ± 0.04	10.74 ± 0.03
$\overline{B^0} \to \pi^+K^-$	9.30 ± 0.03	8.68 ± 0.03

Table 5.2: Stripping plus reconstruction efficiencies with respect to generation, obtained from the MC samples.

the previous analysis. Again, the efficiencies for selecting 2011 events are higher than for 2012 across all decay channels.

Decay mode	2011 $\epsilon_{\text{trig/strip}}(\%)$	2012 $\epsilon_{\text{trig/strip}}(\%)$
$B^0 \to p\overline{p}$	39.18 ± 0.14	37.87 ± 0.11
$B_s^0 \to p\overline{p}$	39.45 ± 0.14	37.79 ± 0.12
$B^0 \to K^+\pi^-$	50.56 ± 0.29	47.97 ± 0.25
$\overline{B^0} \to \pi^+K^-$	50.95 ± 0.33	48.15 ± 0.28

Table 5.3: Trigger efficiencies with respect to stripping, obtained from the MC samples.

5.1.4 Preselection Efficiencies

The preselection efficiencies have been assessed with respect to triggered (and stripped) events, to mimic the sequence in which all sets of selection cuts are applied. All numbers for all the relevant decay modes are summarised in Table 5.4. While the preselection criteria applied to the signal and normalisation channels are identical, the observed differences between the efficiencies simply reflect the different momentum spectra of $B^0_{(s)} \to p\overline{p}$ and $B^0 \to K^+\pi^-$. The major impact from the preselection on the $B^0 \to K^+\pi^-$ efficiencies comes from the imposed lower-edge momentum cut of $> 5000\,\text{MeV}/c$ which removes a larger relative amount of events in $B^0 \to K^+\pi^-$ than $B^0_{(s)} \to p\overline{p}$.

5.1.5 PID Efficiencies

PID efficiencies are calculated using the PID calibration methods described in Section 4.3.4. The $B^0 \to K^+\pi^-$ efficiencies are calculated within a kinematic binning in p, η, the same binning variables as those used in the $B^0_{(s)} \to p\overline{p}$ PID calibration. However,

Decay mode	2011 $\epsilon_{\text{sel/trig}}(\%)$	2012 $\epsilon_{\text{sel/trig}}(\%)$
$B^0 \to p\overline{p}$	91.01 ± 0.41	89.40 ± 0.37
$B_s^0 \to p\overline{p}$	89.59 ± 0.40	87.49 ± 0.35
$\overline{B^0} \to K^+\pi^-$	86.53 ± 0.60	84.91 ± 0.53
$\overline{B^0} \to \pi^+K^-$	86.07 ± 0.66	84.80 ± 0.58

Table 5.4: Preselection efficiencies relative to events that passed the stripping and trigger requirements obtained on the MC samples.

the internal binning structure, listed in Table 5.5, differs greatly from the $B^0_{(s)} \to p\overline{p}$ PID binning scheme shown in Table 4.5 due to the different kinematics of the K and π tracks compared to protons. All PID efficiencies are summarised in Table 5.6.

Variable	Binning
Track η	1.5 : 2.08333 : 2.66667 : 3.25 : 3.83333 : 4.41667 : 5
Track p (MeV/c)	3000 : 9300 : 15600 : 19000 : 24400 : 29800 : 35200 : 40600 : 46000 : 51400 : 56800 : 62200
	67600 : 73000 : 78400 : 83800 : 89200 : 94600 : 100000 : 112500 : 125000 : 137500 : 150000

Table 5.5: Binning scheme in η and p used for $B^0 \to K^+\pi^-$ PID efficiency calculations.

Decay mode	2011 $\epsilon_{\text{PID}}(\%)$	2012 $\epsilon_{\text{PID}}(\%)$
$B^0 \to p\overline{p}$	48.81 ± 0.08	46.47 ± 0.06
$B_s^0 \to p\overline{p}$	48.63 ± 0.08	46.46 ± 0.06
$\overline{B^0} \to K^+\pi^-$	43.84 ± 0.11	43.76 ± 0.16
$\overline{B^0} \to \pi^+K^-$	42.99 ± 0.11	42.98 ± 0.16

Table 5.6: PID efficiencies with respect to events passing the preselection cuts.

5.1.6 MVA Efficiencies

MVA cut efficiencies calculated with respect to events passing the PID selection stage are listed in Table 5.7. The much higher selection efficiencies for $B^0 \to K^+\pi^-$ compared to $B^0_{(s)} \to p\overline{p}$ are due to the differing optimisations of the MVA selection in the signal and normalisation channel selections. The $B^0_{(s)} \to p\overline{p}$ MVA selection is optimised to maximise the Punzi FoM (see Equation 3.3) for the $B^0_{(s)} \to p\overline{p}$ signal (see Section 4.3.4), whereas the $B^0 \to K^+\pi^-$ MVA selection is optimised to maximise the statistical significance, $\frac{S}{\sqrt{S+B}}$, of the $B^0 \to K^+\pi^-$ signal.

Decay mode	2011 $\epsilon_{\text{MVA/PID}}(\%)$	2012 $\epsilon_{\text{MVA/PID}}(\%)$
$B^0 \to p\overline{p}$	57.26 ± 0.32	58.07 ± 0.25
$B_s^0 \to p\overline{p}$	60.05 ± 0.32	59.26 ± 0.25
$B^0 \to K^+\pi^-$	97.58 ± 0.77	95.84 ± 0.58
$\overline{B^0} \to \pi^+K^-$	97.42 ± 0.77	95.95 ± 0.58

Table 5.7: MVA selection efficiencies with respect to events passing the PID selections.

5.1.7 Total Efficiencies

Table 5.8 shows the total selection efficiencies for the signal and normalisation channels calculated for each year by combining the numbers from Tables 5.1–5.7. The combined efficiencies given in the right-hand column of Table 5.8 are the averages of the 2011 and 2012 efficiencies weighted by the integrated luminosity in each year and are calculated using the formula

$$\frac{1}{\epsilon_{\text{tot}}^{\text{comb}}} = \frac{\frac{\int \mathcal{L}\,\mathrm{d}t^{11}}{\epsilon_{\text{tot}}^{11}} + \frac{\int \mathcal{L}\,\mathrm{d}t^{12}}{\epsilon_{\text{tot}}^{12}}}{\int \mathcal{L}\,\mathrm{d}t^{11} + \int \mathcal{L}\,\mathrm{d}t^{12}} \tag{5.2}$$

where $\int \mathcal{L}\,\mathrm{d}t^{\text{year}}$ is the integrated luminosity of the data set.

Combining the total efficiencies from Table 5.8 with Equation 5.1 we obtain the ratios of efficiencies for $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ for 2011, 2012 calculated as,

$$\frac{\epsilon^{tot}_{B^0\to K^+\pi^-} + \epsilon^{tot}_{\overline{B^0}\to\pi^+K^-}}{\epsilon^{tot}_{B^0\to p\overline{p}}} = 1.459 \pm 0.012 \quad , \tag{5.3}$$

$$\frac{\epsilon^{tot}_{B^0\to K^+\pi^-} + \epsilon^{tot}_{\overline{B^0}\to\pi^+K^-}}{\epsilon^{tot}_{B_s^0\to p\overline{p}}} = 1.419 \pm 0.011 \quad . \tag{5.4}$$

Decay mode	2011 $\epsilon_{\text{tot}}(\%)$	2012 $\epsilon_{\text{tot}}(\%)$	Combined $\epsilon_{\text{tot}}(\%)$
$B^0 \to p\overline{p}$	0.493 ± 0.004	0.424 ± 0.003	0.446 ± 0.003
$B_s^0 \to p\overline{p}$	0.511 ± 0.004	0.434 ± 0.003	0.458 ± 0.003
$B^0 \to K^+\pi^-$	0.402 ± 0.005	0.344 ± 0.004	0.362 ± 0.004
$\overline{B^0} \to \pi^+K^-$	0.320 ± 0.004	0.274 ± 0.003	0.288 ± 0.004

Table 5.8: Total selection efficiencies.

While efforts have been made to keep the signal and normalisation mode selections as similar as possible, the final ratios of efficiencies show that there are clear differences in the signal and normalisation modes. They arise mainly from the differences in the trigger and offline selections, which are dominated by the differences in MVA selection efficiencies, see Tables 5.3 and 5.7. These differences make the estimation of the related systematic uncertainties slightly more delicate but do not harm the sensitivity of the analysis. The systematic uncertainties are determined in Section 6.1.

5.2 Mass Fits

The fits to the signals $B^0_{(s)} \to p\overline{p}$ and the normalisation mode $B^0 \to K^+\pi^-$ are both detailed in the following subsections after a brief description of the common fitting framework.

5.2.1 Fitter Framework

The mass fit is implemented as an unbinned maximum likelihood fit and utilises the same *G-Fact* analysis package as the previous analysis and several other LHCb analyses. *G-Fact* is a versatile fitter framework designed for use on LHCb analyses. The configuration used for this analysis performs unbinned maximum likelihood fits to the invariant mass spectra of the signal and normalisation channel candidates.

5.2.2 Fit to the $B^0 \to K^+\pi^-$ Normalisation Channel

The event selection of the normalisation channel $B^0 \to K^+\pi^-$ described in Section 4.5 is assumed to yield a data sample consisting of the following classes:

- The $B^0 \to K^+\pi^-$ signal;
- The $B^0_s \to \pi^+K^-$ signal;
- The $B^0_s \to K^+K^-$ misidentified background;
- The $B^0 \to \pi^+\pi^-$ misidentified background;
- The $\Lambda^0_b \to p\pi^-$ misidentified background;
- The $\Lambda^0_b \to pK^-$ misidentified background;
- Partially reconstructed backgrounds;

- Combinatorial background.

Any contamination from other decays is treated as a systematic uncertainty as described in Section 6.1.7.

Normalisation channel candidate events are split by charge conjugate states resulting in separate $K\pi$ invariant mass fits for $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ candidates. The signal yields from these fits are then combined to give the final normalisation channel yield.

The mass spectra are fitted with unbinned maximum likelihood fits including the signal and background classes listed above. The details of the line shapes of each of the classes are given in the following subsections.

$B \to h^+h^-$ Signal Classes

As with previous analysis, the $B \to h^+h^-$ signal mass distributions are modelled using a double Crystal Ball (DCB) function (see Equation 3.4). The fits to the $B^0 \to K^+\pi^-$ signal in the two charge conjugate samples share common DCB parameters which are obtained from fits to combined $K^+\pi^-$ and π^+K^- MC events. The separate fits to the $B_s^0 \to \pi^+K^-$ signal also share common DCB parameters with an identical method used to obtain them. The mass distributions for the $B^0 \to K^+\pi^-$ and $B_s^0 \to \pi^+K^-$ modes from MC simulation data sets, after applying the trigger, stripping and offline selections, are shown in Figures 5.1 and 5.2. The PID selection is implemented via cutting on MC PID variables, rather than reweighting the mass shapes using PIDCalib per-event efficiencies. Investigations were made into fitting PIDCalib reweighted events instead but the results when implemented in the fits to the $K\pi$ data were not as good as those from the mass shapes derived from MC PID variable selected samples. These distributions are fitted with the double CB function, which contains a composite of the high and low tail versions of Equation 3.4, where the high tail refers to the DCB shape in the high mass region where $x > \mu$ and the low tail refers to the low mass region, $x < \mu$. These DCBs have means μ_{B^0,B_s^0} and widths $\sigma_B, \sigma_{B_s^0}$, with tail components $\alpha_{Low}^{B^0,B_s^0}$, $\alpha_{High}^{B^0,B_s^0}$, $n_{Low}^{B^0,B_s^0}$ and $n_{High}^{B^0,B_s^0}$. The mean and the widths of the two components in the double CB function are set to be the same. When fitting to data, the $B_s^0 \to \pi^+K^-$ modes are treated slightly differently to the $B^0 \to K^+\pi^-$ in their respective fits: the $B^0 \to K^+\pi^-$ peak mean, μ_{B^0}, and width, σ_{B^0}, are free to float in the fit, however the mean, $\mu_{B_s^0}$, and width, $\sigma_{B_s^0}$, of the B_s^0 are fixed relative to the fitted $B^0 \to K^+\pi^-$ values. The value of $\sigma_{B_s^0}$ is fixed such that the ratio of the B_s^0 and B^0 widths is identical to that

from the fitted MC distributions ($\sigma^{K^+\pi^-}_{B^0_s} = (1.02083 \pm 0.00319) \times \sigma_{B^0\to K^+\pi^-}$ and $\sigma^{\pi^+K^-}_{B^0_s} = (1.01829 \pm 0.00395) \times \sigma_{\overline{B^0}\to\pi^+K^-}$). The value of $\mu_{B^0_s}$ is fixed such that the difference between the B^0_s and B^0 masses is equal to 87.35 MeV/c^2 [2]. The DCB tail parameters are extracted from MC simulation events. The fits to the $B^0 \to K^+\pi^-$ and $B^0_s \to \pi^+K^-$ mass shapes from simulation are given in Figures 5.1 and 5.2 and the fitted values are given in Table 5.11. The table also shows the parameter values extracted from data, in the mass fits shown in Figures 5.5 and 5.6.

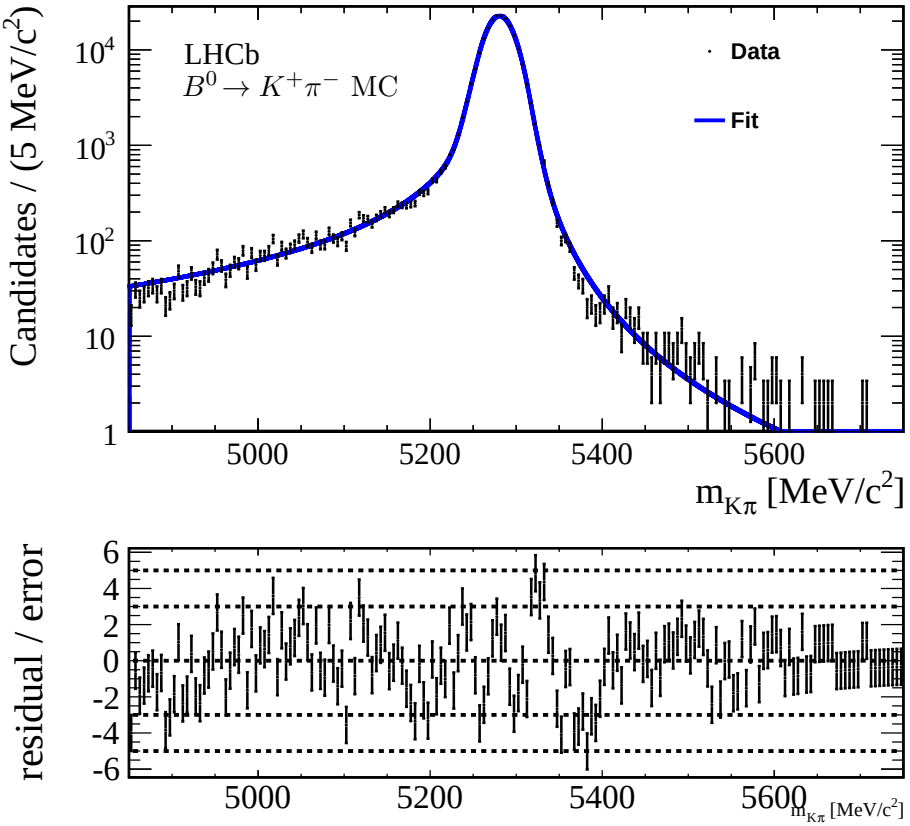


Figure 5.1: $B^0 \to K^+\pi^-$ mass distribution from MC where the trigger, stripping and offline selections (*cf.* Section 4.5) have been applied. The distributions are fitted with a double Crystal Ball shape [70], Equation 3.4. Refer to the text for details.

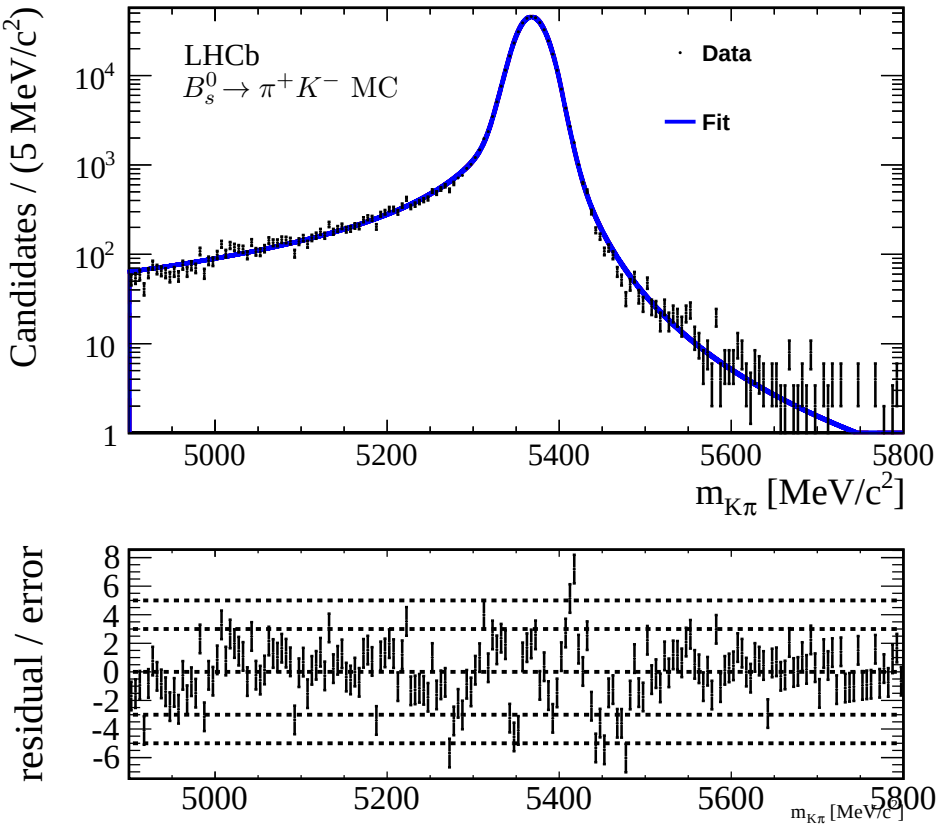


Figure 5.2: Fitted mass distributions of $B_s^0 \to \pi^+ K^-$ MC events where the trigger, stripping and offline selections (*cf.* Section 4.5) have been applied. The distributions are fitted with a double Crystal Ball shape [70], Equation 3.4

.

Misidentified Backgrounds

Misidentified $B \to h^+h^-$ and $\Lambda_b^0 \to ph^-$ events with expected yields greater than 0.1% of the $B^0 \to K^+\pi^-$ yield are included individually in the fit as misidentified backgrounds. The specific decays included are $B_s^0 \to K^+K^-$, $B^0 \to \pi^+\pi^-$, $\Lambda_b^0 \to p\pi^-$ and $\Lambda_b^0 \to pK^-$. The line shapes for these backgrounds are modelled with non-parametric distributions, as shown in Figure 5.3, extracted from MC samples reconstructed and selected under the $K\pi$ hypothesis. The selected distributions are smeared using the *kernel method*, which provides a method for producing a continuous distribution for a given varaible from a finite dataset [79]. This is achieved by having each event in the dataset be described by its own kernel function which serves to spread the contribution from the event over a range of values, rather than being confined to a specific bin. The overall probability density function for the dataset is made up from the sum of the individual kernel functions in the dataset.

The signal fractions of the misidentified peaks are linked to the signal fraction of the $B^0 \to K^+\pi^-$ peak in the data via the scaling factors

$$k^{KK}_{\text{misID}} = \frac{f_s}{f_d} \times \frac{\mathcal{B}(B_s^0 \to K^+K^-)}{\mathcal{B}(B^0 \to K^+\pi^-)} \times \frac{\omega_{K\pi}^{B_s^0 \to K^+K^-}}{\epsilon_{K\pi}^{B^0 \to K^+\pi^-}} \quad , \tag{5.5}$$

$$k^{\pi\pi}_{\text{misID}} = \frac{\mathcal{B}(B^0 \to \pi^+\pi^-)}{\mathcal{B}(B^0 \to K^+\pi^-)} \times \frac{\omega_{K\pi}^{B^0 \to \pi^+\pi^-}}{\epsilon_{K\pi}^{B^0 \to K^+\pi^-}} \quad , \tag{5.6}$$

$$k^{p\pi}_{\text{misID}} = \frac{f_{\Lambda_b}}{f_d} \times \frac{\mathcal{B}(\Lambda_b^0 \to p\pi^-)}{\mathcal{B}(B^0 \to K^+\pi^-)} \times \frac{\omega_{K\pi}^{\Lambda_b^0 \to p\pi^-}}{\epsilon_{K\pi}^{B^0 \to K^+\pi^-}} \tag{5.7}$$

and

$$k^{pK}_{\text{misID}} = \frac{f_{\Lambda_b}}{f_d} \times \frac{\mathcal{B}(\Lambda_b^0 \to pK^-)}{\mathcal{B}(B^0 \to K^+\pi^-)} \times \frac{\omega_{K\pi}^{\Lambda_b^0 \to pK^-}}{\epsilon_{K\pi}^{B^0 \to K^+\pi^-}} \quad , \tag{5.8}$$

where $\omega^d_{K\pi}$ is the misidentification rate of the decay d to the final state $K\pi$ and $\epsilon^d_{K\pi}$ is the efficiency of identifying the decay d with the final state $K\pi$. The signal fractions are given by $f^{KK}_{misID} = f_{B^0} \times k^{KK}_{misID}$, $f^{\pi\pi}_{misID} = f_{B^0} \times k^{\pi\pi}_{misID}$, $f^{p\pi}_{misID} = f_{B^0} \times k^{p\pi}_{misID}$ and $f^{pK}_{misID} = f_{B^0} \times k^{p\pi}_{misID}$, where f_{B^0} is the $B^0 \to K^+\pi^-$ signal fraction extracted from the $B^0 \to K^+\pi^-$ invariant mass fit. The kernelised histograms used in the individual charge state fits are identical but have different scalings relative to the B^0 peak corresponding to their differing PID efficiencies.

The parameters $\omega^d_{K\pi}$ and $\epsilon^d_{K\pi}$ are determined using the `PIDCalib` package as weighted

averages over 2011 and 2012, magnet up and magnet down MC samples. The relative efficiencies, compared to $B^0 \to K^+\pi^-$, of the selection steps prior to the PID selection (Generation, Stripping, Triggers etc.) are assumed to be 1 for each of the included misidentified backgrounds.

Decay	$\epsilon, \omega_{PID}^{K^+\pi^-}$ %	$k_{misID}^{K^+\pi^-}$	$\epsilon, \omega_{PID}^{\pi^+K^-}$ %	$k_{misID}^{\pi^+K^-}$
$B^0 \to K^+\pi^-$	43.23	1	42.97	1
$B^0 \to \pi^+\pi^-$	1.10	6.64×10^{-3}	1.10	6.66×10^{-3}
$B_s^0 \to K^+K^-$	0.92	7.15×10^{-3}	0.91	7.14×10^{-3}
$\Lambda_b^0 \to p\pi^-$	2.89	4.54×10^{-3}	2.88	4.55×10^{-3}
$\Lambda_b^0 \to pK^-$	1.40	3.41×10^{-3}	1.41	3.46×10^{-3}

Table 5.9: PID selection efficiencies and scaling factors for the $B^0 \to K^+\pi^-$ signal and $B \to h^+h^-$ and $\Lambda_b^0 \to ph^-$ misidentified backgrounds in both $K^+\pi^-$ and π^+K^- charge states.

The scaling factor for each misidentified background is a free parameter but its value is constrained within a Gaussian of mean k_{misID}^{decay} and width $\sigma_{k_{misID}^{decay}}$, where $\sigma_{k_{misID}^{decay}}$ is the statistical uncertainty of k_{misID}^{decay} and is calculated via the propagation of errors of the k_{misID}^{decay} calculation terms..

The only other free parameter of the template mass distributions is an offset that shifts the template in mass to accommodate for any shift in the LHCb mass scale. This offset is linked to the mean of the B^0 mass peak and the mass templates are shifted by the difference in fitted B^0 mass and the PDG value of the B^0 mass ($5279.5\,\mathrm{MeV}/c^2$).

Partially Reconstructed Backgrounds

The shape of the partially reconstructed backgrounds is determined from a cocktail of MC samples. The list of potentially contributing decays is given in Table 4.2. It is not exhaustive, but consists of known decays, with large $\mathcal{B}$s, which could be mis-reconstructed as signal with one lost final state particle, possibly in conjunction with a single misidentification. The list is obviously limited to decay modes that are available in MC but gives a representative sample.

All decay modes are reconstructed and selected as signal in the $K\pi$ mass hypothesis. Each mode is assigned a weight given by its relative branching fraction, the hadronisation fraction (f_q/f_d) and the relative sizes of the original MC samples. The hadronisation fractions are those collected in Table A.1 of Appendix A and the branching fractions and

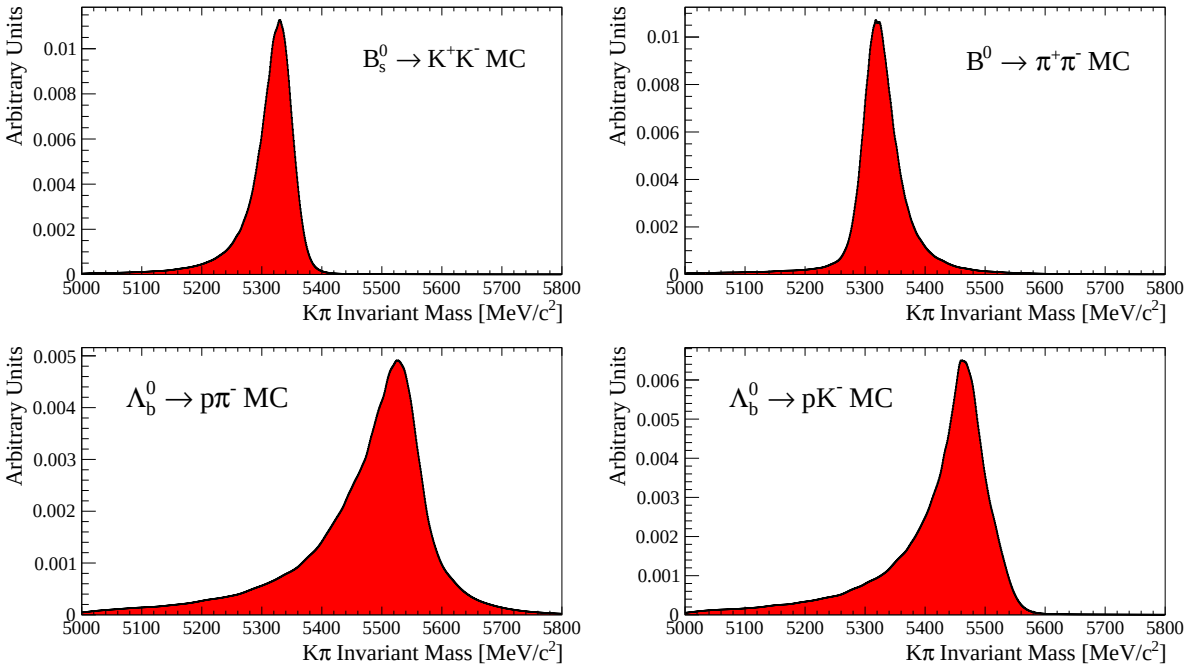


Figure 5.3: Kernelised histograms of MC events selected and reconstructed under the different hypotheses. These non-parametric distributions are used to model the mis-identified specific background events in the $B^0 \to K^+\pi^-$ control channel. The distributions of the background channels misidentified as $B^0 \to K^+\pi^-$ are: $B_s^0 \to K^+K^-$ (top, left) $B^0 \to \pi^+\pi^-$ (top, right), $\Lambda_b^0 \to p\pi^-$ (bottom, left) and $\Lambda_b^0 \to pK^-$ (bottom, right).

MC sample sizes are given in Table 4.2. A weighted sum of the reconstructed and selected spectra is formed as shown in Figure 5.4.

The ensemble of partially reconstructed background spectra is modelled with a linear combination of two exponential functions. The chosen function does not reproduce all features of the observed MC cocktails. However, the lower limit of the range of the mass fit is chosen to be 5150 MeV/c^2 so that only the high mass tail of the partially reconstructed backgrounds needs to be described as it is mainly this part of the distribution that is extending under the signal peaks which affects the measurement. There is some arbitrariness in these distributions since the statistics is limited by the amount of MC available, and the list of potentially contributing decays may not be complete. Hence the aim is to qualitatively model the shape of the distribution with this function. A systematic uncertainty will be derived from the sensitivity to the exact shape of the distribution.

The signal fraction of the partially reconstructed backgrounds is extracted from the fit to the data, all other parameters being fixed from MC.

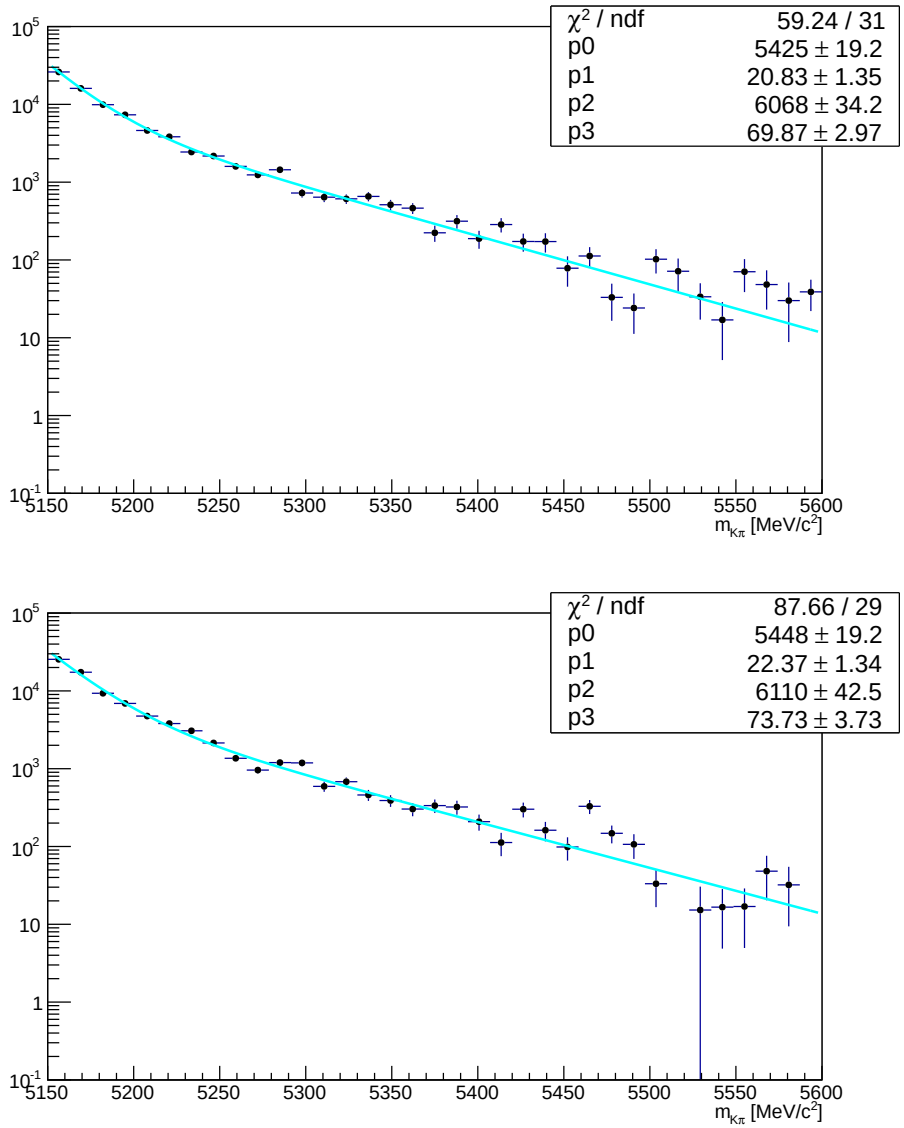


Figure 5.4: The reconstructed $K\pi$ mass distributions of the weighted sum of all contributing partially reconstructed background decays for the $K^+\pi^-$ (top) and π^+K^- (bottom) mass hypotheses. The list of decay modes is given in Table 4.2. The fitted linear combination of two exponential distributions is also shown. The listed parameters p1 and p3 give the gradients of the two exponential functions, ∇^1_{part} and ∇^2_{part}.

Combinatorial Background

In both fits the mass distribution of the combinatorial background is described by a first degree polynomial with the gradient ($\nabla^{K\pi}_{comb}$) as the only free parameter. The constant term is determined from the normalisation within the mass range. The combinatorial

background does not have an independent parameter for its fractional contribution to the fit. Since all the signal fractions should add up to one it is taken as $1 - \sum f_{class}$.

Fit Result

The range of the mass fits are $5150 - 5750$ MeV/c^2. The lower limit is chosen to include the high mass tail of the partially reconstructed backgrounds below the signal peak. The upper limit is chosen to provide an adequate range over which to determine the combinatorial background gradient.

The fits to the $K\pi$ spectra each have 25 parameters. The tail parameters of the double CB functions and the shape parameters of the partially reconstructed backgrounds are determined from fits to MC, with the remaining being extracted from data. The signal fractions of each signal class are left floating in the fit with the constraint that the sum should be equal to one. This gives 7 free parameters in the $K\pi$ fit. All parameters are listed in Table 5.11 (except the parameter that specifies the shift in mass for the misID background templates).

The signal fractions in the normalisation channel are extracted from unbinned maximum likelihood fits to the $K^+\pi^-$ and π^+K^- invariant mass spectra, shown in Figures 5.5 and 5.6, with the fit results listed in Table 5.11. The fitted $B^0 \to K^+\pi^-$ yields, shown in Table 5.10, are determined using the total number of events in each fit (with the associated Poissonian uncertainties) and the corresponding fitted $B^0 \to K^+\pi^-$ signal fraction.

Decay	Fitted Yield
$B^0 \to K^+\pi^-$	48689 ± 254 (stat)
$\overline{B^0} \to \pi^+K^-$	40272 ± 227 (stat)
Combined	88961 ± 341 (stat)

Table 5.10: Fitted $B^0 \to K^+\pi^-$ yields extracted from the individual fits to the $K^+\pi^-$ and π^+K^- invariant mass spectra.

The difference in extracted yields between the $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ fits corresponds to a raw asymmetry, defined as $\frac{N_{B^0\to K^+\pi^-} - N_{\overline{B^0}\to\pi^+K^-}}{N_{B^0\to K^+\pi^-} + N_{\overline{B^0}\to\pi^+K^-}}$, of approximately -9.5%. It is consistent with the raw $B^0 \to K^+\pi^-$ CP asymmetry of $A_{\text{raw}}(B^0 \to K^+\pi^-) = -9.1 \pm 0.6\%$ measured in Reference [69]. There is good agreement even though the reconstruction and selection conditions of this cited measurement were different from the conditions

presented in this analysis.

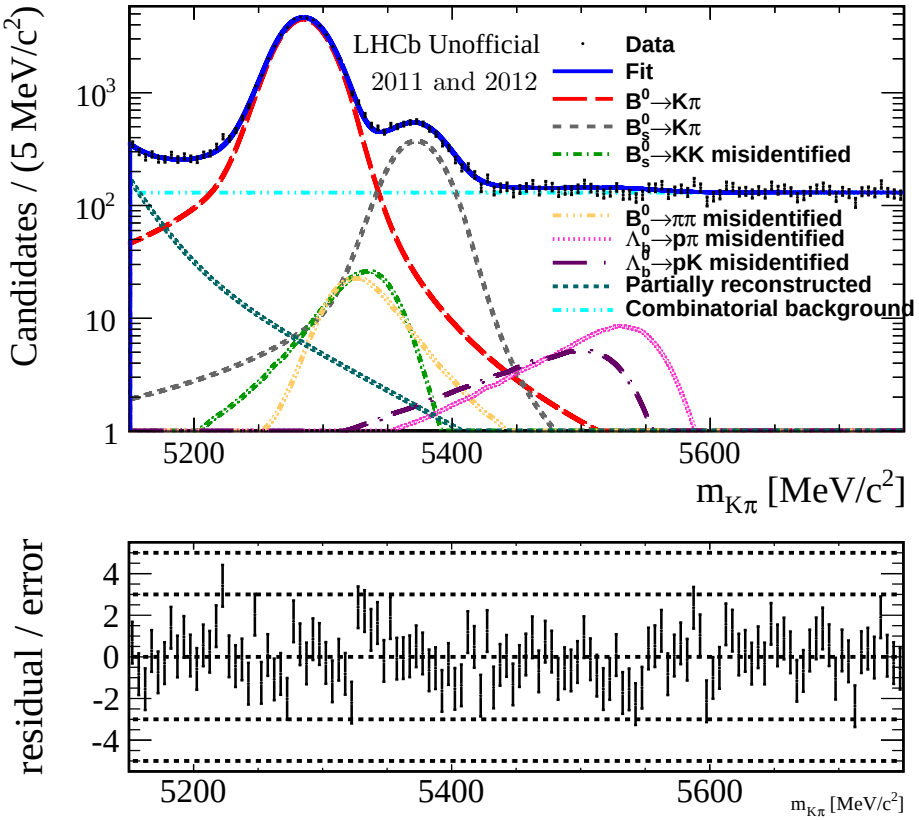


Figure 5.5: Mass fit to the $B^0 \to K^+\pi^-$ normalisation channel. The lower plot shows the binned residuals of the fit divided by the fit error. The parameters are extracted from an unbinned maximum likelihood fit, the binning of the data is only for illustration.

Parameter	Value $K^+\pi^-$	Value π^+K^-	Origin	Description
f_{B^0}	0.68906 ± 0.0025	0.67047 ± 0.0027	Data	B^0 signal fraction
$f_{B_s^0}$	0.05907 ± 0.00144	0.04639 ± 0.00149	Data	B_s^0 signal fraction
f_{misID}^{KK}	$(5.164 \pm 0.05) \cdot 10^{-3}$	$(5.137 \pm 0.05) \cdot 10^{-3}$	Data/MC	$B_s^0 \to K^+K^-$ misID fraction
$f_{misID}^{\pi\pi}$	$(4.730 \pm 0.05) \cdot 10^{-3}$	$(4.692 \pm 0.05) \cdot 10^{-3}$	Data/MC	$B^0 \to \pi^+\pi^-$ misID fraction
f_{misID}^{pK}	$(2.239 \pm 0.04) \cdot 10^{-3}$	$(2.211 \pm 0.04) \cdot 10^{-3}$	Data/MC	$\Lambda_b^0 \to pK^-$ misID fraction
$f_{misID}^{p\pi}$	$(2.978 \pm 0.04) \cdot 10^{-3}$	$(2.962 \pm 0.04) \cdot 10^{-3}$	Data/MC	$\Lambda_b^0 \to p\pi^-$ misID fraction
f_{part}	0.01578 ± 0.00137	0.01972 ± 0.0016	Data	Partially reconstructed fraction
μ_{B^0} (MeV/c^2)	5284.82 ± 0.114	5284.78 ± 0.125	Data	B^0 signal peak mean
Δm (MeV/c^2)	87.35	87.35	PDG	B_s^0-B^0 mass difference
σ_{B^0} (MeV/c^2)	20.561 ± 0.104	20.776 ± 0.108	Data	B^0 signal width
$\alpha_{Low}^{B^0}$	1.398 ± 0.0634	1.398 ± 0.0634	MC	Low CB part boundary (B^0)
$\alpha_{High}^{B^0}$	2.018 ± 0.0339	2.018 ± 0.0339	MC	High CB part boundary (B^0)
$n_{Low}^{B^0}$	1.433 ± 0.0313	1.433 ± 0.0313	MC	Low CB part tail (B^0)
$n_{High}^{B^0}$	3.040 ± 0.0961	3.040 ± 0.0961	MC	High CB part tail (B^0)
$frac_{Low}^{B^0}$	0.3572 ± 0.0359	0.3572 ± 0.0359	MC	MC fraction of low CB (B^0)
$\alpha_{Low}^{B_s^0}$	1.423 ± 0.0428	1.423 ± 0.0428	MC	Low CB part boundary (B_s^0)
$\alpha_{High}^{B_s^0}$	1.983 ± 0.0240	1.983 ± 0.0240	MC	High CB part boundary (B_s^0)
$n_{Low}^{B_s^0}$	1.399 ± 0.0201	1.399 ± 0.0201	MC	Low CB part tail (B_s^0)
$n_{High}^{B_s^0}$	3.281 ± 0.0810	$3.281. \pm 0.810$	MC	High CB part tail (B_s^0)
$frac_{Low}^{B_s^0}$	0.3694 ± 0.0252	0.3694 ± 0.0252	MC	MC fraction of low CB (B_s^0)
∇_{part}^1	20.83 ± 1.35	22.37 ± 1.34	MC	Partially reconstucted gradient 1
∇_{part}^2	$69.87 \pm 2.97.$	$73.73 \pm 3.73.$	MC	Partially reconstructed gradient 2
$\nabla_{comb}^{K\pi}$(MeV/c^2)$^{-1}$	$(0.407 \pm 16.355) \cdot 10^{-8}$	$(37.065 \pm 17.259) \cdot 10^{-8}$	Data	Gradient of the combinatorial
N_{Total}	70660	60065	Data	Total number of fitted events
$N_{K^+\pi^-,\pi^+K^-}$	48689 ± 254	40272 ± 227	Data	Extracted $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ yields

Table 5.11: Description and fitted values of the parameters of the mass fit to the $K\pi$ spectrum. The "Origin" column states if the value is determined from MC, and therefore fixed in the fit, or from the fit to the data.

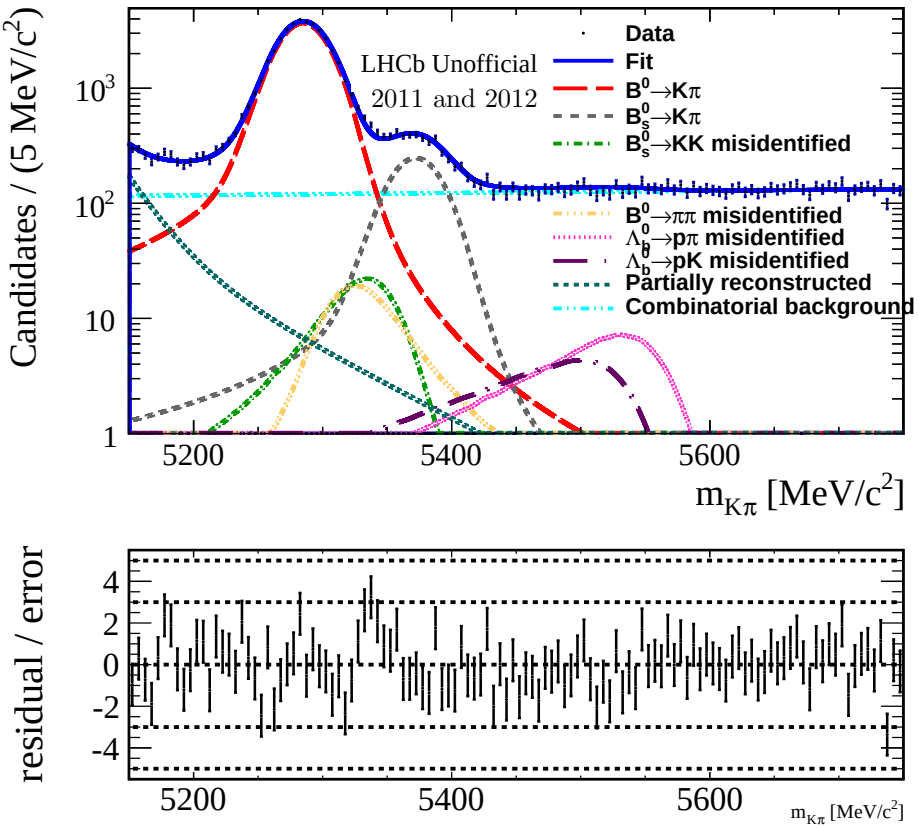


Figure 5.6: Mass fit to the $\overline{B^0} \to \pi^+ K^-$ normalisation channel. The lower plot shows the binned residuals of the fit divided by the fit error. The parameters are extracted from an unbinned maximum likelihood fit, the binning of the data is only for illustration.

5.2.3 Fit to the Signal Sample

The $B^0_{(s)} \to p\overline{p}$ event selection described in Section 4.3 is assumed to yield a data sample consisting of the following classes:

- The $B^0 \to p\overline{p}$ signal;
- The $B^0_s \to p\overline{p}$ signal;
- Combinatorial background.

In particular, any contamination from partially reconstructed backgrounds, with or without extra misidentified particles (see Section 4.4), is treated as a systematic uncertainty, as described in Section 6.1.7.

Again, the mass spectrum is fitted with a maximum likelihood fit including the signal and background classes listed above; details are given in the remainder of this section.

Signal Classes

The $B^0_{(s)} \to p\overline{p}$ signal mass shapes in MC, see Figures 5.7 and 5.8, are reasonably well described by a single Gaussian. Less than 100 $B^0 \to p\overline{p}$ and events and even fewer $B^0_s \to p\overline{p}$ events are expected, so the tails in Figures 5.7 and 5.8 need no detailed description. For this same reason the signal shapes should contribute little to the systematic uncertainty on the yields or branching fraction upper limits.

The widths of both Gaussians are assumed to be the same for $B^0 \to p\overline{p}$ and $B^0_s \to p\overline{p}$ with a systematic uncertainty to be assigned due to this assumption. They are taken from MC with a scaling factor to account for differences in the resolution between data and MC:

$$\begin{aligned} \sigma_B &= \sigma_B^{MC} \frac{\sigma_{K\pi}^{\text{Data}}}{\sigma_{K\pi}^{\text{MC}}} \\ &= 17.62 \times \frac{20.56}{17.58} \quad . \end{aligned} \tag{5.9}$$

In this equation, σ_B^{MC} is the weighted average of the 2011 and 2012 $B^0 \to p\overline{p}$ MC mass resolutions after the full selection with MC PID cuts applied. In other words, the width of both Gaussians is fixed to $\sigma_B = 20.60\,\text{MeV}/c^2$ in the fit. A systematic uncertainty will later be assigned to this value. The peak of the $B^0_s \to p\overline{p}$ Gaussian is constrained to the

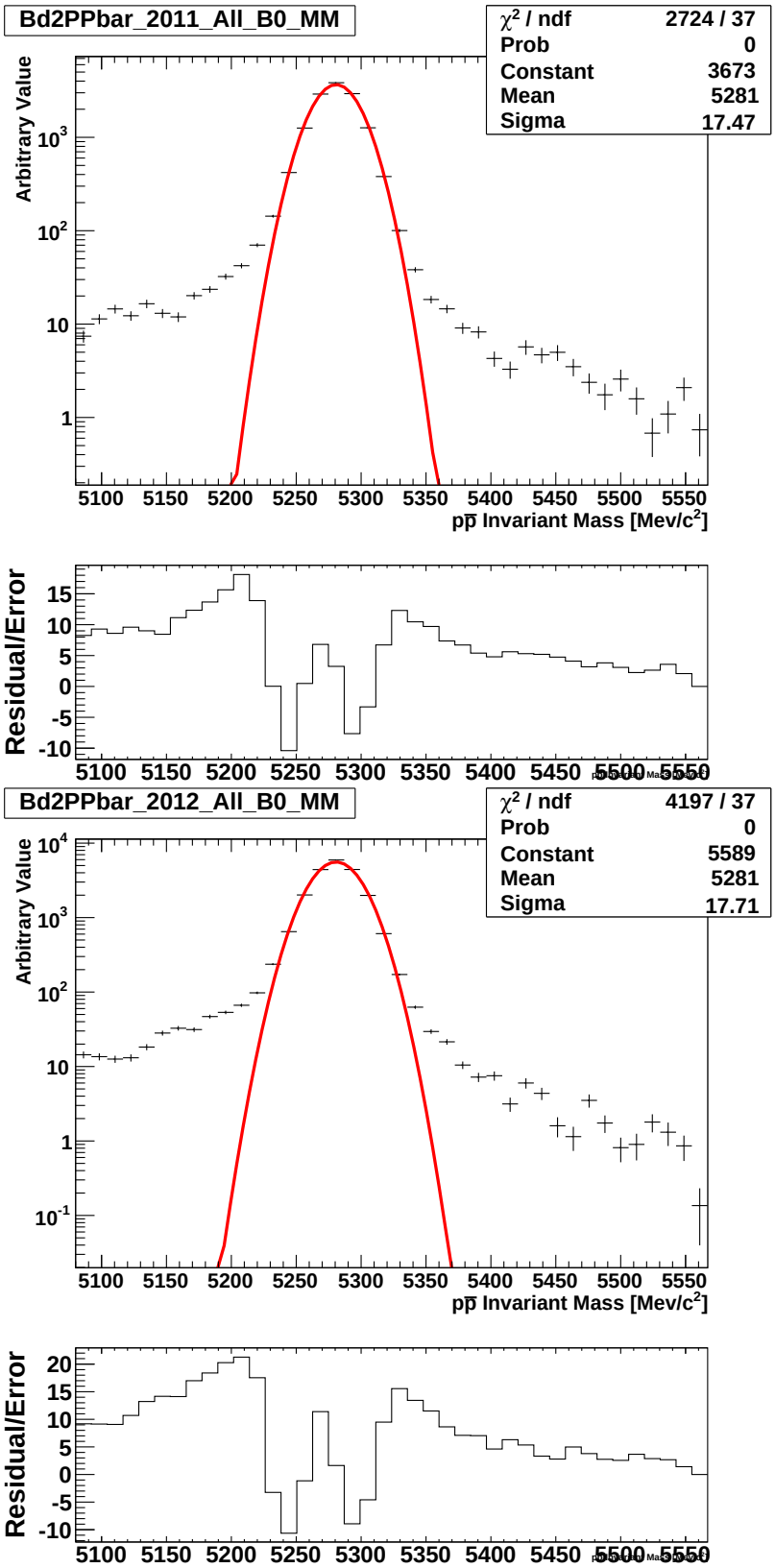


Figure 5.7: Invariant mass distributions for MC-matched $B_d^0 \to p\bar{p}$ events after the full selection.

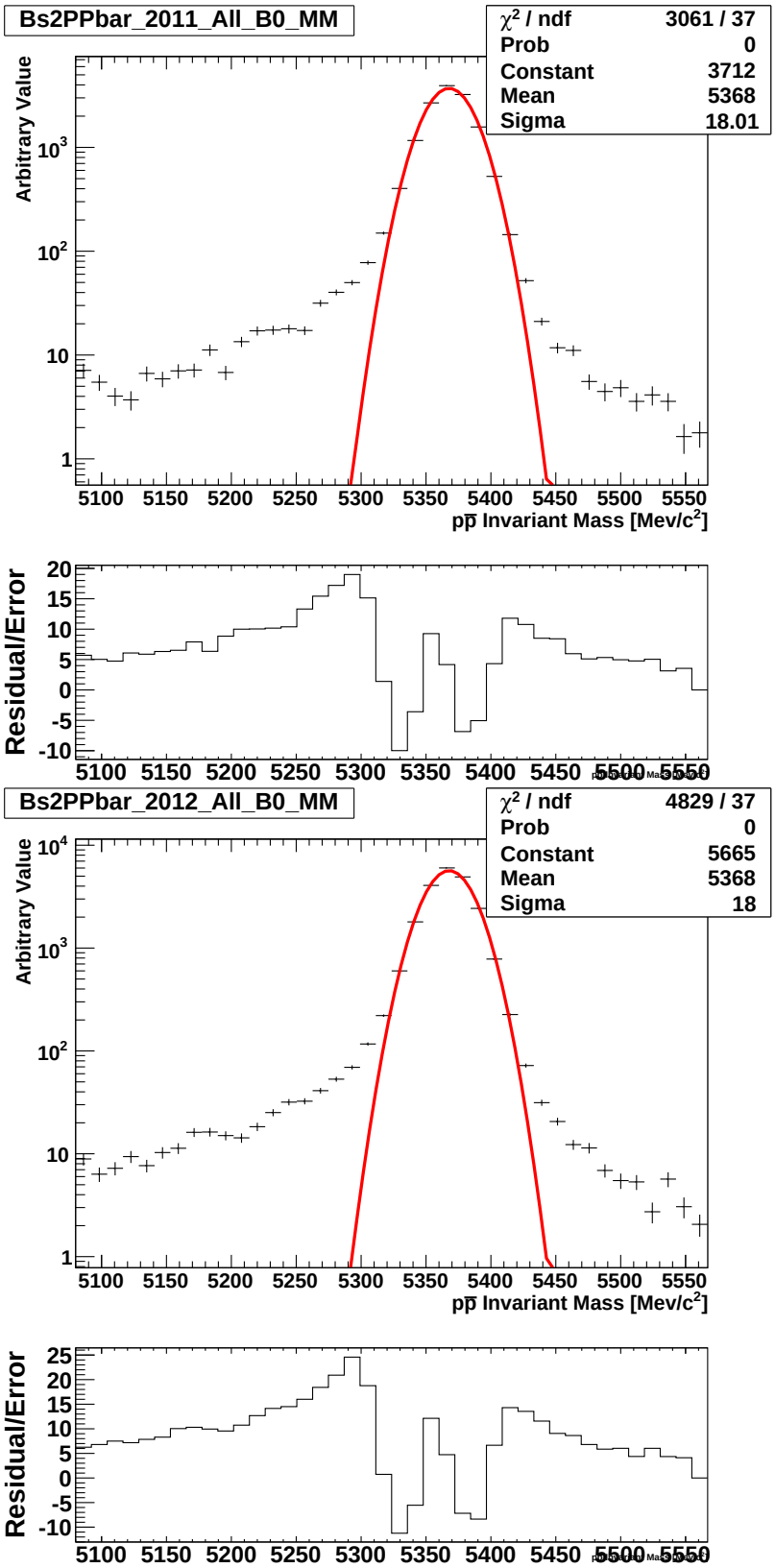


Figure 5.8: Invariant mass distributions for MC-matched $B_s^0 \to p\bar{p}$ events after the full selection.

B_s^0-B^0 mass difference of 87.35 MeV/c^2 [2]. All signal model parameters are collected in Table 5.12.

Sources of Background

As mentioned, contributions from specific physics backgrounds are assumed to be negligible after the full event selection and thus the sole background component included in the $p\overline{p}$ invariant mass fit is to describe the contribution from combinatorial background candidates. The mass distribution of this combinatorial background is described by a first degree polynomial with the gradient ($\nabla^{p\overline{p}}_{comb}$) as the only free parameter. The combinatorial background does not have an independent parameter for the signal fraction; since all the signal fractions should add up to one it is taken as $1 - \sum f_{class}$.

5.3 Mass Fit Toy Studies

Verification of the $B^0_{(s)} \to p\overline{p}$ mass fit is made using fits to toy MC samples. One thousand toy MC samples are produced comprising of $B^0 \to p\overline{p}$, $B^0_s \to p\overline{p}$ and combinatorial background events with relative fractions, 27.4 : 14.3 : 58.3, taken from the expected yields of each class after the full selection. The total number of events in each toy sample is sampled from a Poissonian distribution with a mean of 170.

The conditions and parameters of the fit are identical to those used in the fit to the $B^0_{(s)} \to p\overline{p}$ data described in Section 5.2.3. The invariant mass fit is performed on each toy MC sample and for each free parameter the fitted value is compared to the "true" value used in the generation of the toy samples. The pull distributions of the five free parameters are shown in Figure 5.9. No significant biases are observed in any of the fit parameters.

5.4 Statistical Significance Estimation

Using Wilks' theorem [75], fits to $B^0_{(s)} \to p\overline{p}$ toy samples used in Section 5.3 can be used to estimate the statistical significance for the $B^0 \to p\overline{p}$ signal achievable with this analysis. To calculate the estimated significance it is required to fit the $B^0_{(s)} \to p\overline{p}$ toy samples separately under a signal hypothesis and a null hypothesis. In this instance, the signal hypothesis requires that only the B^0 fraction is left as a free parameter in the fit, all other parameters in the fit are fixed to the values listed in Table 5.12 (with the exception that

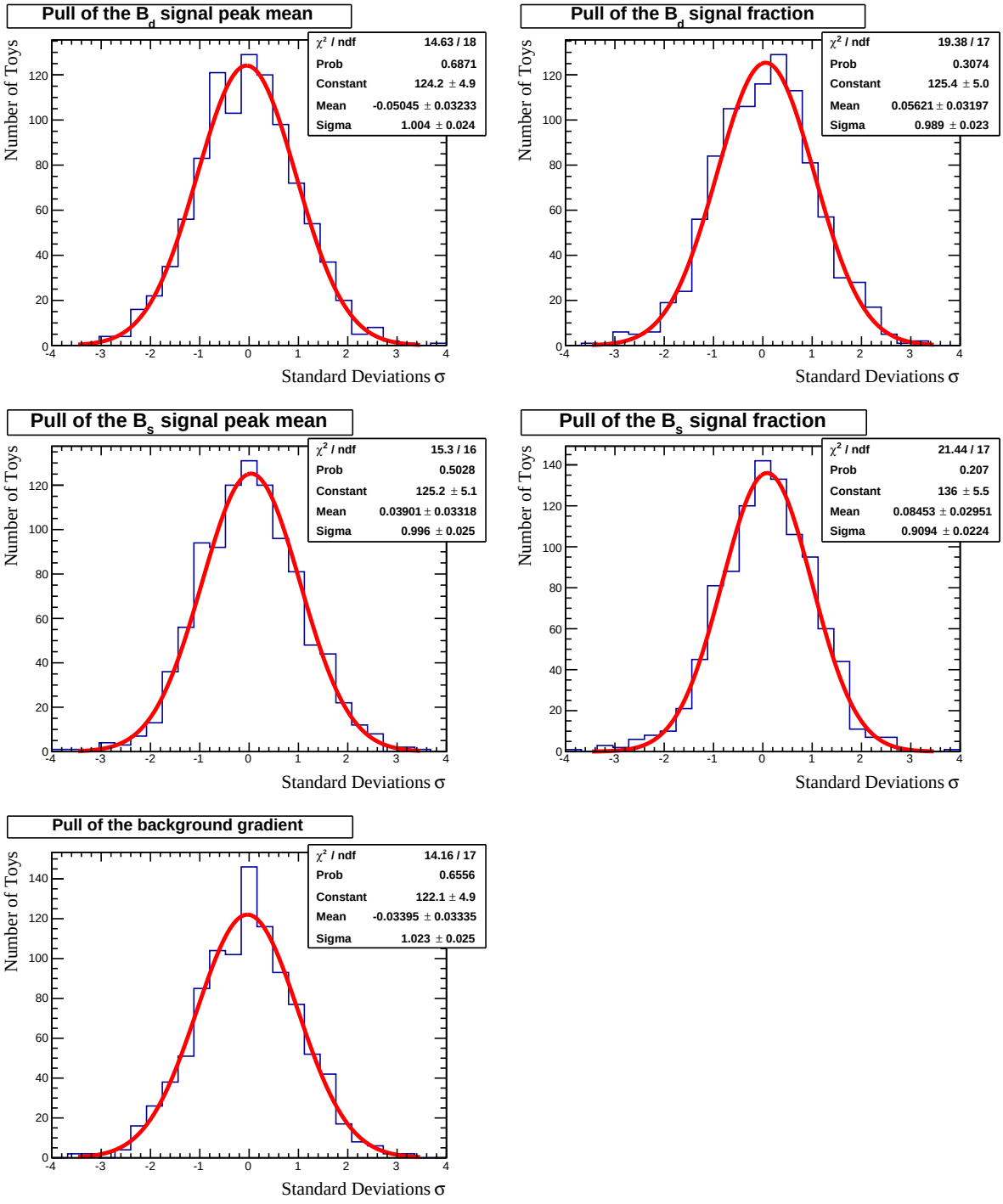


Figure 5.9: $B^0_{(s)} \to p\bar{p}$ toy MC pull distributions for the free fit parameters: $B^0 \to p\bar{p}$ signal peak mean, $B^0 \to p\bar{p}$ signal fraction, $B^0_s \to p\bar{p}$ signal peak mean, $B^0_s \to p\bar{p}$ signal fraction and combinatorial background gradient.

the B_s^0 fraction is set to 0). The null hypothesis then requires that all the parameters in the fit are fixed to the values in Table 5.12 but with both the B^0 and B_s^0 fractions set to 0. With the $B^0_{(s)} \to p\overline{p}$ toy samples fitted for both of these hypotheses, a test statistic, D, is calculated using the following formula:

$$D = 2(\Sigma_{LL}^{Null\ hyp.} - \Sigma_{LL}^{Sig.\ hyp.}) \quad , \tag{5.10}$$

where $\Sigma_{LL}^{Null\ hyp.}$ and $\Sigma_{LL}^{Sig.\ hyp.}$ are the sums of the logarithms of the likelihoods of the events fitted under the null and signal hypotheses, respectively. As the difference in the number of degrees of freedom between the signal and null hypotheses is 1, applying Wilks' theorem allows calculation of the $B^0 \to p\overline{p}$ significance as

$$\text{Significance} = \sqrt{D}. \tag{5.11}$$

Figure 5.10 shows the distributions of estimated $B^0 \to p\overline{p}$ significances for the $B^0_{(s)} \to p\overline{p}$ toy samples The figure shows a Gaussian distribution of significances around a mean value of $(6.94 \pm 0.04)\sigma$ suggesting that in the case of $\mathcal{B}(B^0 \to p\overline{p}) = 1.47 \times 10^{-8}$ there is a greater than 95% probability of obtaining a significance $> 5\sigma$ in the $B^0 \to p\overline{p}$ signal.

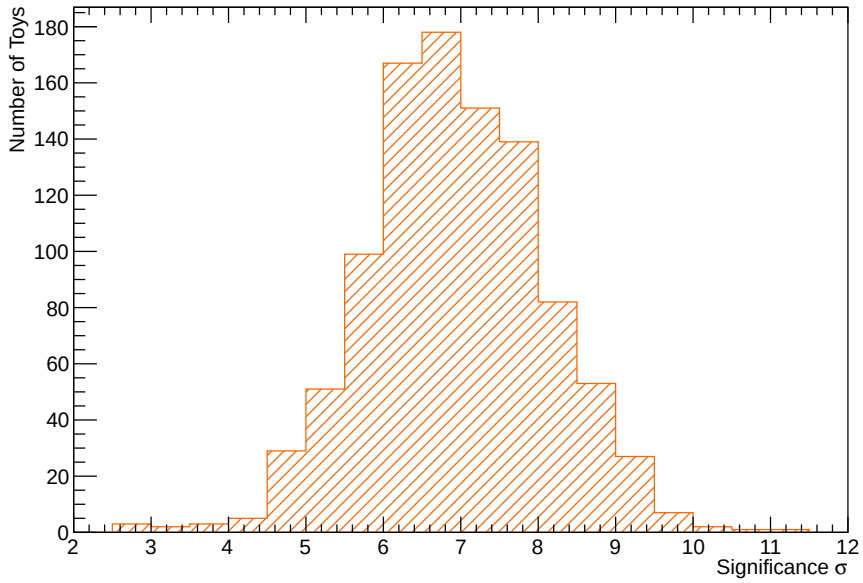


Figure 5.10: $B^0 \to p\overline{p}$ signal estimated statistical significance distributions, calculated using Wilks' theorem, from fits to 1000 $\mathcal{B}\left(B^0 \to p\overline{p}\right) = 1.47 \times 10^{-8}$ toy samples.

Fit Result

Parameter	Value		Origin	Description
f_{B^0}	0.226 ± 0.046		Data	B^0 signal fraction
$f_{B_s^0}$	0.008 ± 0.026		Data	B_s^0 signal fraction
μ_{B^0}	5289.94 ± 5.01	MeV/c^2	Data	B^0 signal peak mean
Δm	87.35	MeV/c^2	PDG	B_s^0-B^0 mass difference
σ_B	20.60	MeV/c^2	MC	Signal mass resolution
$\nabla_{comb}^{p\bar{p}}$	$(-4.47 \pm 1.14) \times 10^{-6}$	$(\,\text{MeV}/c^2)^{-1}$	Data	Gradient of the combinatorial

Table 5.12: Description and fitted values of the parameters of the mass fit to the $p\bar{p}$ spectrum. The "Origin" column states if the value is determined from MC or the PDG, and therefore fixed in the fit, or from the fit to the data.

With the mass fit successfully verified with the toy studies, the $B^0_{(s)} \to p\bar{p}$ invariant mass fit is performed on the full sample of 2011 and 2012 unblinded signal data. The fit to the $p\bar{p}$ mass spectrum, Figure 5.11, is performed in the range $[5080, 5567]$ MeV/c^2 defined directly from the spectrum of candidates passing the stripping selection. In total, four parameters are fitted, the width of the signal Gaussians being fixed from MC, see above, and the B_s^0-B^0 mass difference being constrained to the PDG value.

After unblinding, the full $p\bar{p}$ dataset contains 166 events in the range $[5080, 5567]$ MeV/c^2. The fit to this data returns a successful, converged fit, the results of which are shown, superposed to the data distribution, in Fig. 5.11, with all fitted model parameters collected in Tab. 5.12.

As with the $B^0 \to K^+\pi^-$ fit, from the results of the fit, the total number of $p\bar{p}$ candidate events (166 ± 12.88) is multiplied by the extracted $B^0_{(s)} \to p\bar{p}$ signal fractions to obtain the yields for the $B^0_{(s)} \to p\bar{p}$ signals in the full mass range $[5080, 5567]$ MeV/c^2:

$$\begin{aligned} N(B^0 \to p\bar{p}) &= 37.58 \pm 8.13\ (\text{stat}) \quad , \\ N(B_s^0 \to p\bar{p}) &= 1.29 \pm 4.34\ (\text{stat}) \quad , \\ N(comb) &= 127.13 \pm 13.10\ (\text{stat}) \quad . \end{aligned}$$

The signal significances are calculated using Wilks' theorem [75] from the change in the logarithm of the likelihood values obtained with and without each of the signal components in the fit in turn, via Equation 5.10. The inclusion of systematic uncertainties is postponed until Section 6.2.6. We thus obtain

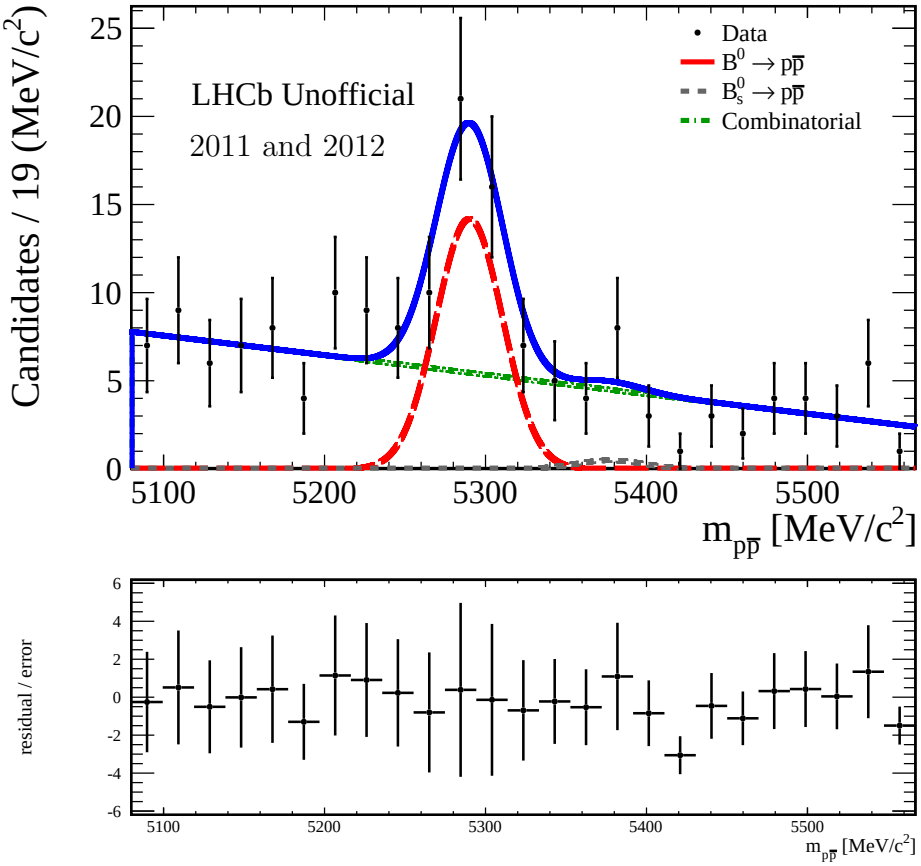


Figure 5.11: Invariant mass distribution of $p\overline{p}$ candidates after full selection. The fit result (blue, solid) is superposed with each fit model component: the $B^0 \to p\overline{p}$ signal (red, dashed), the $B_s^0 \to p\overline{p}$ signal (grey, dotted) and the combinatorial background (green, dot-dashed).

$$\begin{aligned} \text{Significance } (B^0 \to p\overline{p}) &= 5.94\sigma \text{ (stat only)} \quad , \\ \text{Significance } (B_s^0 \to p\overline{p}) &= 0.32\sigma \text{ (stat only)} \quad . \end{aligned}$$

In other words, we observe an excess of $B^0 \to p\overline{p}$ candidates with respect to background expectations corresponding to a significance of greater than five standard deviations from the null hypothesis, and we find no evidence for a $B_s^0 \to p\overline{p}$ signal.

Chapter 6

Systematic Uncertainties and Results

6.1 Systematic Uncertainties

6.1.1 $B^0 \to K^+\pi^-$ Branching Fraction

The branching fraction of $B^0 \to K^+\pi^-$ entering Equation 3.2,

$$\mathcal{B}(B^0 \to K^+\pi^-) = (19.57^{+0.53}_{-0.52}) \cdot 10^{-6} \quad , \tag{6.1}$$

is taken from the HFAG average [71]. Its relative uncertainty of 2.7% is taken as a systematic uncertainty on the ratio of branching fractions as given by Equation 3.2.

6.1.2 Relative Generator-level Efficiency

The MC is expected to describe very well the geometric acceptance. As such, no systematic uncertainty is set for the ratio of generator-level efficiencies $\epsilon^{\rm gen}_{B^0\to K^+\pi^-}/\epsilon^{\rm gen}_{B^0_{(s)}\to p\overline{p}}$.

6.1.3 Trigger Uncertainties

The trigger efficiencies were assessed on MC for all modes involved and we assume them to be reasonably realistic given that the p_T distributions are well described in MC and all final states comprise the same number of tracks. Neglecting small p and p_T differences between the $B^0 \to p\overline{p}$ and $B^0_s \to p\overline{p}$ modes, the ratios of $B^0 \to K^+\pi^-/B^0_{(s)} \to p\overline{p}$ trigger

efficiencies should be identical within errors. Taking the systematic uncertainty on the relative trigger efficiencies as

$$\Delta_{\text{trig}} = \frac{|\epsilon_{\text{trig}}\left(B^0 \to p\overline{p}\right) - \epsilon_{\text{trig}}\left(B^0_s \to p\overline{p}\right)|}{\epsilon_{\text{trig}}\left(K\pi\right)}, \tag{6.2}$$

yields values for Δ_{trig} of 0.53% for 2011 and 0.17% in 2012. Where $\epsilon_{trig}(B^0_{(s)} \to p\overline{p})$ are the $B^0_{(s)} \to p\overline{p}$ trigger selection efficiencies from Table 5.1.3 and $\epsilon_{trig}(K\pi)$ is the weighted average of the $B^0 \to K^+\pi^-$ and $\overline{B^0} \to \pi^+K^-$ trigger efficiencies. These uncertainties are much smaller than 1% and therefore we choose not to assign a specific systematic uncertainty due to the relative trigger efficiencies.

A further contribution to the trigger efficiency systematics may come from the imperfect description of the `L0_Hadron_TOS` trigger in the simulation, from which the trigger efficiencies are calculated. However this uncertainty should mostly cancel when taking the ratio of efficiencies. Work is currently ongoing to evaluate the systematic uncertainty from this source

6.1.4 Relative Selection Efficiency

As shown in Equation 5.1, for this analysis the full selection efficiency can be decomposed into the products of various selection efficiencies calculated along the selection chain. While the previous sections dealt with the systematic uncertainties concerning the generator level and trigger efficiencies, the uncertainties due to the remainder of the analysis selection steps still need to be accounted for. The PID selection systematic uncertainties are covered in the next section. This leaves the systematic uncertainties due to the reconstruction, stripping, preselection and MVA selections.

The reconstruction efficiency is expected to be very well described by the simulation with any discrepancies expected to cancel when taking the ratio of efficiencies and therefore we consider contributions from the reconstruction efficiency to the overall selection efficiency systematic to be negligible. Likewise, the preselection criteria applied to both signal and normalisation channels are very loose and cut on variables which are relatively well described in the simulation. As such, we consider any contributions from the preselection to the overall selection efficiency systematic to be negligible also.

The remaining systematic uncertainty contributions are therefore from the stripping and MVA selections. To assess the systematic uncertainties due to data and MC disagreement in these selections we follow a similar procedure to that laid out in the recent

LHCb search for the $B^+ \to p\overline{\Lambda}$ decay [80]. This method entails reweighting key MC variables to match the responses shown in $sWeighted$ data. For this analysis we use the $sWeighted$ $B^0 \to K^+\pi^-$ data from Appendix B as a proxy. These $sWeighted$ data are used to reweigh key MC stripping and MVA selection input variables which have distributions with relatively large differences to those in the data. These reweighted distributions are then used as inputs to retrain the $B^0 \to p\overline{p}$ MVA selections. Selection efficiencies are then calculated for these new MVA selections cutting at the same MVA response values used in the nominal analysis.

Work is currently ongoing to assess the final values for this systematic for the $B^0_{(s)} \to p\overline{p}$ search. However, the $B^+ \to p\overline{\Lambda}$ decay [80] search follows a very similar overall analysis method with an identical method for assigning the relative selection efficiency systematic uncertainty. For that analysis, the authors assign a systematic uncertainty of 2.5% for this systematic uncertainty which we also assign as the systematic uncertainty value here.

6.1.5 PID Uncertainties

The systematic uncertainties due to the $B^0_{(s)} \to p\overline{p}$ and $B^0 \to K^\pm\pi^\mp$ PID selections arise from the implementation of the `PIDCalib` calibration methods. The PID calibration systematics come from two main sources: the choice of kinematic binning variables used and the choice of internal binning structure.

In choosing the combination of binning variables, in this case track p and η, the assumption is made that these variables alone are sufficient to accurately model the PID performance of the RICH detector. However, as was mentioned previously, it is known that detector occupancy variables such as the number of tracks in an event, $nTracks$, have a direct effect on RICH performance independent of individual track kinematics [49]. It is also well established that these variables are currently poorly reproduced in MC and for the reasons explained in Section 4.3.4 we choose to not include $nTracks$ as a binning variable for the $B^0_{(s)} \to p\overline{p}$ and $B^0 \to K^\pm\pi^\mp$ PID calibration pricedures. To measure the difference in PID response between using the nominal $(p-\eta)$ binning scheme and one with $nTracks$ included, the PID calibration procedure is re-run using a $(p-\eta-nTracks)$ binning scheme. The p and η binning structure is kept identical to nominal PID calibration binning but with the bins in $nTracks$ being (0 - 60 - 120 - 200 - 300 - 400 - 500 - 600). To avoid additional effects due to low statistics, the calibration procedures are performed for a subset of $B^0_{(s)} \to p\overline{p}$ ($B^0 \to K^\pm\pi^\mp$) events around the peaks of the signal p and η distributions $p \in [16250, 44000]$ MeV/c and $\eta \in [2.0, 3.25]$ ($p \in [9300, 46000]$ MeV/c and

$\eta \in [2.1, 4.4]$). The PID calibration procedure is then performed with the resulting mean PID efficiency from the $nTracks$ binning scheme, $\epsilon^{\text{PID}}_{nTracks}$, compared with the nominal PID efficiency within the given $p - \eta$ range, $\epsilon^{\text{PID}}_{\text{nom}}$. The systematic uncertainty is then extracted as:

$$\Delta_{PID-nTracks} = \frac{|\epsilon^{\text{PID}}_{\text{nom}} - \epsilon^{\text{PID}}_{nTracks}|}{\epsilon^{\text{PID}}_{\text{nom}}}. \tag{6.3}$$

Table 6.1 lists the values of $\epsilon^{\text{PID}}_{\text{nom}}$, $\epsilon^{\text{PID}}_{\text{nTracks}}$ and $\Delta_{PID-nTracks}$ for the signal and normalisation channels.

Decay	ϵ_{nom} (%)	$\epsilon_{\text{nTracks}}$ (%)	$\Delta_{PID-nTracks}$ (%)
$B^0 \to p\overline{p}$	70.37	74.67	6.1
$B^0_s \to p\overline{p}$	70.45	75.75	7.5
$B^0 \to K^+\pi^-$	65.88	67.22	2.0
$\overline{B^0} \to \pi^+K^-$	65.78	67.18	2.0

Table 6.1: PID selection efficiencies with and without binning in nTracks and the corresponding systematic uncertainty calculated from the relative difference between the two efficiencies with respect to the nominal PID efficiency.

To assess the systematic uncertainties due to the choices of calibration binning schemes used, the $B^0_{(s)} \to p\overline{p}$ ($B^0 \to K^{\pm}\pi^{\mp}$) PID efficiencies are recalculated for the nominal $p\bar{p}$ ($K^{\pm}\pi^{\mp}$) selection using a different binning scheme. The structures of the new binning schemes are shown in Table 6.2. For $B^0_{(s)} \to p\overline{p}$ the binning scheme used is Scheme 13 from the PID binning studies described in Appendix D. This scheme was chosen as in the PID binning studies (see Appendix D) it gave the largest difference in results compared to the nominal scheme (Scheme 15, Table D.2) with small relative statistical uncertainties. For K and π PID efficiencies, two additional binning schemes were considered, as shown in Table 6.2, one with much finer binning than the nominal binning scheme and one with much coarser binning.

For each decay the weighted average of the efficiencies for 2011 and 2012, magnet up and down, with the new binning scheme, $\epsilon^{\text{PID}}_{\text{bin}}$, is compared to the nominal PID efficiency, $\epsilon^{\text{PID}}_{\text{nom}}$, with the systematic uncertainty given. by:[1]

$$\Delta_{PID-bin} = \frac{|\epsilon_{\text{nom}} - \epsilon_{\text{bin}}|}{\epsilon_{\text{nom}}}. \tag{6.4}$$

[1]Weighted average calculated using same formalism as Equation 5.2.

In the case of K and π, the systematic uncertainty is assigned from the binning scheme which gives the largest value of $\Delta_{PID-bin}$. Table 6.3 lists the values of ϵ_{nom}, ϵ_{bin} and $\Delta_{PID-bin}$ for the signal and normalisation channels.

The final PID selection systematic uncertainty is taken from the quadrature sum of $\Delta_{PID-bin}$ and

Particle	p Bins (MeV/c)	η Bins
Proton	5000 : 6944.44 : 8888.89 : 10833.3 : 12777.8 14722.2 : 16666.7 : 18611.1 : 20555.6 : 22500 24444.4 : 26388.9 : 28333.3 : 30277.8 : 32222.2 34166.7 : 36111.1 : 38055.6 : 40000 : 46000 52000 : 58000 : 64000 : 70000 : 75000 80000 : 85000 : 90000 : 95000 : 100000 105000 : 110000 : 115000 : 120000 : 150000	1.5 : 2 : 2.25 : 2.5 : 2.75 3 : 3.25 : 3.5 : 3.75 : 4 4.33333 : 4.66667 : 5
Kaon, Pion: Fine	3000 : 6150 : 9300 : 12450 : 15600 19000 : 23050 : 27100 : 31150 : 35200 39250 : 43300 : 47350 : 51400 : 55450 59500 : 63550 : 67600 : 71650 : 75700 79750 : 83800 : 87850 : 91900 : 95950 100000 : 106250 : 112500 : 118750 : 125000 131250 : 137500 : 143750 : 150000	1.5 : 1.85 : 2.2 : 2.55 : 2.9 : 3.25 :3.6 : 3.95 : 4.3 : 4.65 5
Kaon, Pion: Coarse	3000 : 9300 : 15600 : 19000 : 30571.4 42142.9 : 53714.3 : 65285.7 : 76857.1 : 88428.6 100000 : 125000 : 150000	1.5 : 2.375 : 3.25 : 4.125 : 5

Table 6.2: PIDCalib kinematic binning schemes used for determination of PID systematic uncertainties.

Decay	ϵ_{nom} (%)	ϵ_{bin} (%)	$\Delta_{PID-bin}$ (%)
$B^0 \to p\overline{p}$	47.23	47.78	1.2
$B_s^0 \to p\overline{p}$	47.12	47.69	1.1
$B^0 \to K^+\pi^-$	43.78	43.16	1.4
$\overline{B^0} \to \pi^+K^-$	42.97	43.51	1.3

Table 6.3: PID selection efficiencies for separate kinematic binning schemes and the corresponding systematic uncertainty calculated from the relative difference between the two efficiencies with respect to the nominal PID efficiency.

6.1.6 Hadronisation Probability

As mentioned in Sec. 4.1, an extra factor f_s/f_d enters Eq. 3.2 on the righthand side for the case of the $B_s^0 \to p\overline{p}$. Its relative error of 5.0% (see Appendix A) is taken as a systematic error.

6.1.7 Mass Fits

The systematic uncertainties coming from the fits to the $B^0_{(s)} \to p\overline{p}$ signal and $B^0 \to K^+\pi^-$ mass distributions are the last source of systematics. They arise from the limited knowledge or the choice of the mass models, and from the uncertainties on the values of the parameters fixed in the mass fits.

$B^0 \to K^+\pi^-$ Yield

To assess the systematic uncertainty on the $B^0 \to K^+\pi^-$ yield we rely on large sets of toy studies. The fit model parameters fixed in the nominal fit are all considered in turn. As for all floating quantities, the fit naturally accounts for their uncertainties.

The overall systematic uncertainty assigned to the $B^0 \to K^+\pi^-$ yield is taken as the sum in quadrature of the observed percentage shifts in the fitted B^0 fraction with respect to the nominal value for each of the considered parameters.

This is a very similar method to that which was performed for the 2011 $B^0_{(s)} \to p\overline{p}$ analysis. Final calculations of the $B^0 \to K^+\pi^-$ yield systematic uncertainties for this analysis are still ongoing. The final value is expected to be small and similar to the $B^0 \to K^+\pi^-$ yield systematic uncertainty from the 2011 analysis, which was 1.6%. Therefore we assign a systematic uncertainty of 1.6% to the $B^0 \to K^+\pi^-$ yield here.

$B^0_{(s)} \to p\overline{p}$ Yields

To calculate the systematic uncertainties relevant to the $B^0_{(s)} \to p\overline{p}$ mass fit we use a combination of toy studies and amended fits to the data. As with the $B^0 \to K^+\pi^-$ fit, systematic uncertainties on the $B^0_{(s)} \to p\overline{p}$ fit arise from the non-perfect knowledge of the parameters fixed in the fit – in this case the parameters Δm and σ_B – as well as the choice of models used to describe the signal PDFs in the fit (refer to Tab. 5.12 for the definition of the fit parameters). In total, three toy studies were performed to assess the impact of the assumptions and uncertainties related to Δm and σ_B. All sets of toys were produced with 2000 samples of 45 events each, with the B^0 and B_s^0 fractions and the combinatorial

background gradient set to their fitted values in the full data fit, as listed in Tab. 5.12; the parameter under consideration was changed according to needs, see below. The mean relative difference in $f_{B^0_{(s)}}$ between each set of toys and the results of the nominal fit to data is taken as the systematic uncertainty. The full list of toy studies are:

- The contribution from Δm is assessed by changing at generation time the Δm value. The latter is sampled from a Gaussian of mean $\mu = \Delta m$ and width $\sigma = 0.23\,\mathrm{MeV}/c^2$ corresponding to the uncertainty on Δm from the PDG.

- The $B^0_s \to p\overline{p}$ Gaussian width is shifted up by 1.8%, which is the amount by which the $B^0 \to K^+\pi^-$ and $B^0_s \to \pi^+K^-$ widths differ in MC, and approximately the amount by which the $B^0 \to p\overline{p}$ and $B^0_s \to p\overline{p}$ differ in MC after the full selection. This gives a measure of the systematic uncertainty due to our assumption that the B^0 and B^0_s widths are equal in the fit.

- The widths of the $B^0_{(s)} \to p\overline{p}$ Gaussian PDFs are varied by sampling σ_B from a Gaussian with $\mu = \sigma_B$ and $\sigma = 0.5\%$ of σ_B, where 0.5% corresponds to the errors on the fitted $B^0 \to K^+\pi^-$ peak sigmas. This provides a measure of the effect of our imperfect knowledge of the signal peak widths on the $B^0_{(s)} \to p\overline{p}$ fit results.

Source of systematic uncertainty	Shift from nominal f_{B^0} (%)	Shift from nominal $f_{B^0_s}$ (%)
B^0 and B^0_s mass difference, Δm	0.4	0.3
Difference in $B^0_{(s)} \to p\overline{p}$ mass widths	0.04	2.9
$B^0_{(s)} \to p\overline{p}$ mass width values	0.5	1.4
Total	0.6	3.2

Table 6.4: Summary of systematic uncertainties on the $B^0_d \to p\overline{p}$ and $B^0_s \to p\overline{p}$ fitted fractions.

Tab. 6.4 lists the contributions to the systematic uncertainties on the $B^0_{(s)} \to p\overline{p}$ fitted fractions.The total systematic uncertainty is taken as the quadrature sum of these contributions.

These parameter shifts are shown to have a greater effect on the $B^0 \to p\overline{p}$ fitted fraction than on the $B^0_s \to p\overline{p}$, with the main contribution to the overall systematics on the $B^0 \to p\overline{p}$ fraction coming from the width of the B^0 peak while for the $B^0_s \to p\overline{p}$ fraction the main contribution comes from the description of the combinatorial background shape.

6.1.8 Summary of Systematic Uncertainties

Table 6.5 summarises all systematic uncertainties.

Uncertainty origin	Value (%)		
	$B^0 \to p\overline{p}$	$B_s^0 \to p\overline{p}$	$B^0 \to K^+\pi^-$
$B^0 \to K^+\pi^-$ Branching fraction	–	–	2.7
Selection efficiency relative to $B^0 \to K^+\pi^-$	2.5	2.5	–
PID uncertainties	6.2	7.6	2.4
Yield from mass fit	0.6	3.2	1.6
f_s/f_d	–	5.0	–
Total	6.7	10.0	4.0

Table 6.5: Summary of the systematic uncertainties. The totals correspond to the quadratic sum of each column.

6.2 Results and Conclusion

6.2.1 Statistical Treatment of the Results

As was the case for the previous analysis, we opt to follow the Feldman-Cousins (FC) approach [76] for the extraction of the analysis results.

Prior to unblinding FC belts are constructed to test what the expected limits are in case of no observed signals (modulo systematics that can only be evaluated after unblinding). This is done for both the scenarios of pure background and signal+background.

The statistical significance of any observed signal is estimated from the change in likelihood with and without the signal component; details on this method, based on Wilks' theorem, are given in Section 5.4.

6.2.2 Single-event Sensitivities

As detailed in Section 4.1, the $B^0_{(s)} \to p\overline{p}$ branching fractions are extracted relative to the $B^0 \to K^+\pi^-$ normalisation channel. In detail,

$$\begin{aligned}
\mathcal{B}(B^0_{(s)} \to p\overline{p}) &= \frac{N(B^0_{(s)} \to p\overline{p})}{N(B^0 \to K^+\pi^-)} \\
&\quad \cdot \frac{\epsilon^{\text{total}}_{B^0 \to K^+\pi^-}}{\epsilon^{\text{total}}_{B^0_{(s)} \to p\overline{p}}} \\
&\quad \cdot f_d / f_{d(s)} \\
&\quad \cdot \frac{\int \mathcal{L}\, \mathrm{d}t^{B^0 \to K^+\pi^-}}{\int \mathcal{L}\, \mathrm{d}t^{B^0_{(s)} \to p\overline{p}}} \qquad (6.5) \\
&\quad \cdot \mathcal{B}(B^0 \to K^+\pi^-) \qquad (6.6) \\
&= \alpha_{d(s)} \cdot N(B^0_{(s)} \to p\overline{p}) \qquad (6.7)
\end{aligned}$$

where $\alpha_{d(s)}$ is a normalisation factor, the single-event sensitivity. The uncertainties on $\alpha_{d(s)}$, calculated by propagation of errors rather than from the sum of the squares of all errors on the relevant factors, are detailed in Table 6.6; they amount to 3.7% and 1.7% for the B^0 and B^0_s modes, respectively. Note that the systematic uncertainties on $\alpha_{d(s)}$ are calculated from Table 6.5, excluding the systematic uncertainties associated to the mass fit for the $B^0_{(s)} \to p\overline{p}$ channels and including the systematic uncertainties on the $B^0 \to K^+\pi^-$ fitted yield.

As will become evident in the subsequent sections, the final results on $\mathcal{B}(B^0_{(s)} \to p\overline{p})$ are obtained from the signal yields determined in $\pm 50\,\mathrm{MeV}/c^2$ windows around the fitted $B^0_{(s)} \to p\overline{p}$ mass peak means. The $\alpha_{d(s)}$ quantities above should therefore account for the efficiency of these implicitly defined mass-window cuts. The latter have been determined from MC to be 98.5% and the 1.5% corrections are hereafter neglected given the large statistical uncertainties on the $B^0_{(s)} \to p\overline{p}$ signal yields.

Quantity	$B^0 \to p\overline{p}$	$B^0_s \to p\overline{p}$
$\frac{\epsilon^{\mathrm{total}}_{B^0 \to K^+\pi^-}}{\epsilon^{\mathrm{total}}_{B^0_{(s)} \to p\overline{p}}}$	1.459 ± 0.014	1.419 ± 0.014
$f_{d(s)}/f_d$	1.0	0.259 ± 0.013
$\frac{\int \mathcal{L}\,\mathrm{d}t^{B^0 \to K^+\pi^-}}{\int \mathcal{L}\,\mathrm{d}t^{B^0_{(s)} \to p\overline{p}}}$	0.9984 ± 0.0231	
$\mathcal{B}(B^0 \to K^+\pi^-)$	$(19.57^{+0.53}_{-0.52}) \times 10^{-6}$	
$\mathrm{N}(B^0 \to K^+\pi^-)$	88961 ± 341	
α	$(3.20 \pm 0.12) \times 10^{-10}$	$(1.20 \pm 0.02) \times 10^{-9}$

Table 6.6: Summary of the factors and their combined statistical and systematic uncertainties entering the single-event sensitivities for $B^0 \to p\overline{p}$ and $B^0_s \to p\overline{p}$.

6.2.3 Example Construction of FC Confidence Intervals

Prior to unblinding and fitting the $B^0_{(s)} \to p\overline{p}$ data, the FC confidence belt was constructed, covering the range of expected event yields. The belt bands were constructed using the `TFeldmanCousins` class in ROOT [81], which takes as input values the total number of events observed (in the signal region) and the number of expected background events, returning lower and upper limits on the number of signal events at the desired CL. Figure 6.1 shows the 68.27% and 90% CL intervals on the number of signal events (N_{signal}) given an expected background yield of $N_{\mathrm{background}} = 26$ events, as a function of the number of observed events (N_{observed}); the background yield of 26 events is chosen as an estimate of the expected background within a range $\mu_{B^0_{(s)}} \pm 50\,\mathrm{MeV}/c^2$, which is defined as the $B^0_{(s)}$ signal regions (*i.e.* as regions $\pm 50\,\mathrm{MeV}/c^2$ around the relevant signal Gaussian peak means).

To extract the correct N_{observed} and $N_{\mathrm{background}}$ values from the fit to the data, the integrals of the three fitted PDF shapes (the two signal Gaussian PDFs and the combinatorial background polynomial PDF) are calculated within the $\mu_{B^0_{(s)}} \pm 50\,\mathrm{MeV}/c^2$ signal regions. N_{observed} is just the total number of events within the B^0 or B^0_s signal region while

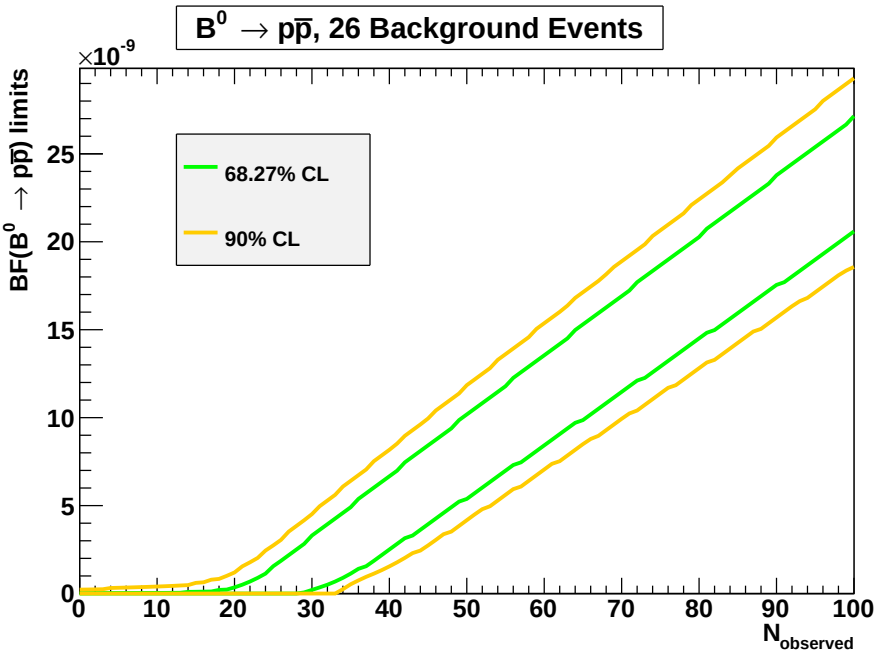


Figure 6.1: Feldman-Cousin belts on the $B^0 \to p\overline{p}$ branching fraction at 68.27% and 90% confidence levels for 26 expected background events and over a range of observed number of events.

$N_{\text{background}}$ within the B^0 (B_s^0) signal region is given by the integral of the combinatorial background plus the cross-feed contribution of the B_s^0 (B^0) tail.

6.2.4 Observed Yields and Limits Excluding Systematic Uncertainties

Following the unblinding of the $B^0_{(s)} \to p\overline{p}$ signal regions and with the data successfully fitted, as described in Section 5.4, the observed $B^0_{(s)} \to p\overline{p}$ signal region yields and confidence limits are extracted using the method outlined in Section 6.2.3, for now excluding systematic uncertainties. From the total fitted yields across the full mass range $[5080, 5567]$ MeV/c^2, see Section 5.4, the yield for each fit component is calculated within the $B^0_{(s)} \to p\overline{p}$ signal regions defined as $\mu_{B^0_{(s)}} \pm 50\,\text{MeV}/c^2$, *i.e.* as a region $\pm 50\,\text{MeV}/c^2$ around the relevant signal Gaussian peak mean. From Table 5.12 the fitted B^0 (B_s^0) mass peak mean is $\mu_{B^0} = 5289.94 \pm 5.01\,\text{MeV}/c^2$ ($\mu_{B_s^0} = 5377.29 \pm 5.01\,\text{MeV}/c^2$) from which we define the B^0 (B_s^0) signal region as $[5239.94, 5339.94]\,\text{MeV}/c^2$ ($[5327.29, 5427.29]\,\text{MeV}/c^2$). Tables 6.7 and 6.8 list the yields obtained for all fit components within the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$

signal regions, respectively. The correlation matrix provided by the mass fit is used to correctly determine N_{observed} and $N_{\text{background}}$.

Fit component	Yield within B^0 signal region
$N(B^0 \to p\overline{p})$	37.01 ± 8.01
$N(B_s^0 \to p\overline{p})$	0.05 ± 0.15
$N(comb)$	28.01 ± 2.89
N_{observed}	65.07 ± 7.75
$N_{\text{background}}$	28.06 ± 2.85

Table 6.7: Event yields determined within the $B^0 \to p\overline{p}$ signal region, with N_{observed} being the sum of all three fit components and $N_{\text{background}}$ the sum of the $B_s^0 \to p\overline{p}$ and combinatorial background yields.

Fit component	Yield within B_s^0 signal region
$N(B^0 \to p\overline{p})$	1.31 ± 0.28
$N(B_s^0 \to p\overline{p})$	1.27 ± 4.28
$N(comb)$	23.05 ± 2.38
N_{observed}	25.63 ± 4.20
$N_{\text{background}}$	24.36 ± 2.32

Table 6.8: Event yields determined within the $B_s^0 \to p\overline{p}$ signal region, with N_{observed} being the sum of all three fit components and $N_{\text{background}}$ the sum of the $B^0 \to p\overline{p}$ and combinatorial background yields.

Using these N_{observed} and $N_{\text{background}}$ yields, we construct statistical-only FC 68.27% and 90% CL intervals for the yields of the $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ signals following the method described in Section 6.2.3:

$$\begin{aligned}
N(B^0 \to p\overline{p}) &= [29.76, 45.76] \text{ at } 68.27\% \text{ CL} , \\
N(B^0 \to p\overline{p}) &= [25.41, 51.47] \text{ at } 90\% \text{ CL} , \\
N(B_s^0 \to p\overline{p}) &= [0, 6.5] \text{ at } 68.27\% \text{ CL} , \\
N(B_s^0 \to p\overline{p}) &= [0, 10.16] \text{ at } 90\% \text{ CL} .
\end{aligned}$$

From these intervals, and utilising the single-event sensitivities listed in Table 6.6, we obtain statistical-only 68.27% and 90% CL intervals

$$\begin{aligned}
\mathcal{B}(B^0 \to p\overline{p}) &= [9.52, 14.16] \, ([8.13, 16.47]) \times 10^{-9} \text{ at } 68.27\% \, (90\%) \text{ CL} , \\
\mathcal{B}(B_s^0 \to p\overline{p}) &= [0, 7.73] \, ([0, 12.19]) \times 10^{-9} \text{ at } 68.27\% \, (90\%) \text{ CL} .
\end{aligned}$$

6.2.5 Observed Yields and Limits Including Systematic Uncertainties

To calculate FC confidence intervals with the full statistical and systematic uncertainties included, it is necessary to parameterise the uncertainties into components which are dependent and independent of the signal yield[2]. With the $\mathcal{B}(B^0_{(s)} \to p\overline{p})$ calculated via Equation 6.7, the statistical uncertainties on the branching fractions come entirely from the statistical uncertainties on the $B^0_{(s)} \to p\overline{p}$ yields from the mass fit. The $B^0_{(s)} \to p\overline{p}$ yield statistical uncertainty's dependence on the yield is assessed via toy studies.

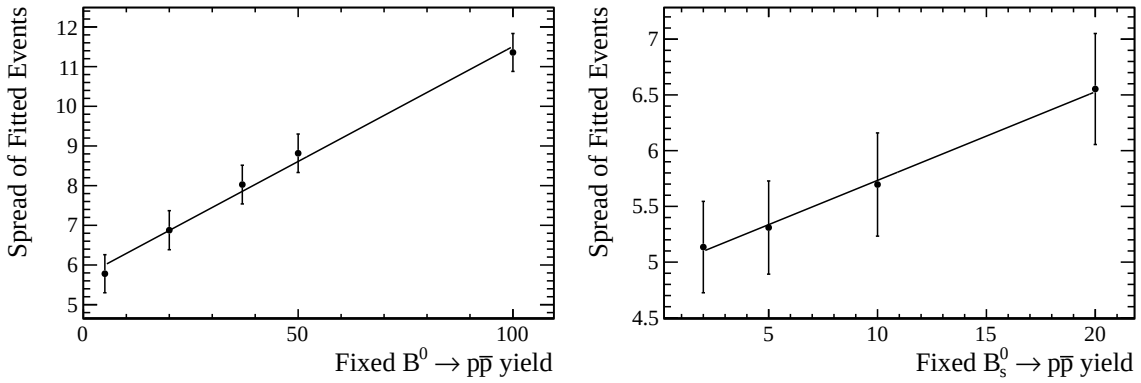
Toy samples are produced with fixed $B^0 \to p\overline{p}$ yields in the range $[0, 100]$, 1000 toys at each fixed yield with the $B^0 \to p\overline{p}$ yield sampled from a Poisson distribution on a toy-by-toy basis. The $B^0_s \to p\overline{p}$ (1 event) and combinatorial background (127 events) yields are kept constant in all the toy samples; these numbers correspond to the fit results presented in Section 5.4, rounded to integers. For each $B^0 \to p\overline{p}$ fixed yield, the toy samples are fitted with the nominal $B^0_{(s)} \to p\overline{p}$ fit and the mean error on the fitted $B^0 \to p\overline{p}$ yield within the signal region is calculated by fitting the distribution of the errors on the $B^0 \to p\overline{p}$ yield across the 1000 toy samples with a Gaussian shape with the mean $B^0 \to p\overline{p}$ yield error taken as the mean of the Gaussian and the error on this mean taken as the width of the Gaussian. The relationship between mean $B^0 \to p\overline{p}$ yield error and fixed $B^0 \to p\overline{p}$ yield is found to be linear. Similar fixed yield toy samples are produced for $B^0_s \to p\overline{p}$ with yields fixed over the range $[0, 20]$, the yields for $B^0 \to p\overline{p}$ and combinatorial background are kept constant in these toys at 38 and 127 events, respectively. Performing the same analysis on mean $B^0_s \to p\overline{p}$ yield error as a function of fixed $B^0_s \to p\overline{p}$ yield shows a similar linear relationship to that found for $B^0 \to p\overline{p}$. Figure 6.2 shows the spread of signal events as a function of the fixed signal yield, $N_{\rm fit}$, with the resulting straight line fit for the $B^0 \to p\overline{p}$ (left) and $B^0_s \to p\overline{p}$ (right) fixed yield toy studies. The results of the fits to the $B^0_{(s)} \to p\overline{p}$ fixed yield toys are

$$\sigma_{N_{\rm fit}} = 5.71 + 0.058 \times N_{\rm fit}(B^0 \to p\overline{p}) \quad , \tag{6.8}$$

$$\sigma_{N_{\rm fit}} = 4.94 + 0.079 \times N_{\rm fit}(B^0_s \to p\overline{p}) \quad . \tag{6.9}$$

The systematic uncertainty on $\mathcal{B}(B^0_{(s)} \to p\overline{p})$ is given by the combination of the sys-

[2]The author hereby states their appreciation to Tom Latham (University of Warwick) for very useful discussions and for sharing relevant code.



(a) $B^0 \to p\overline{p}$ **(b)** $B_s^0 \to p\overline{p}$

Figure 6.2: Mean statistical uncertainty on the $B^0_{(s)} \to p\overline{p}$ yields as a function of the $B^0_{(s)} \to p\overline{p}$ yields with corresponding fit results for $B^0 \to p\overline{p}$ (left) and $B^0_s \to p\overline{p}$ (right) toy studies. Refer to the text for details.

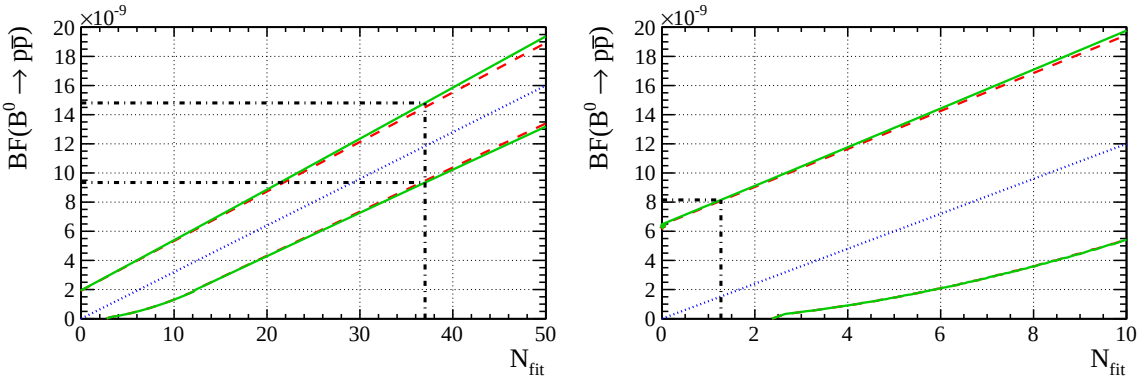
tematic uncertainties on $\alpha_{d(s)}$ and $N(B^0_{(s)} \to p\overline{p})$. The uncertainties on $\alpha_{d(s)}$ (described in Section 6.2.2) are all dependent on the $B^0_{(s)} \to p\overline{p}$ yield while the systematic uncertainties on the $B^0_{(s)} \to p\overline{p}$ yields from the fit are conservatively assumed to be independent of $B^0_{(s)} \to p\overline{p}$ yield. As such, we assign a fixed value of 0.22 events for $B^0 \to p\overline{p}$ and 0.04 events for $B^0_s \to p\overline{p}$ as the systematic uncertainty contributions from $N(B^0_{(s)} \to p\overline{p})$, these values correspond to the percentage values for the systematic uncertainty due to the mass fit (0.6% for $B^0 \to p\overline{p}$, 3.2% for $B^0_s \to p\overline{p}$, see Section 6.1.7) multiplied by the observed signal yield in the relevant signal region (37.01 events for $B^0 \to p\overline{p}$ and 1.27 events for $B^0_s \to p\overline{p}$).

The FC intervals with systematic uncertainties included are constructed using an identical method to that used in the previous analysis (see Chapter 3.6, Equation 3.9 onwards).

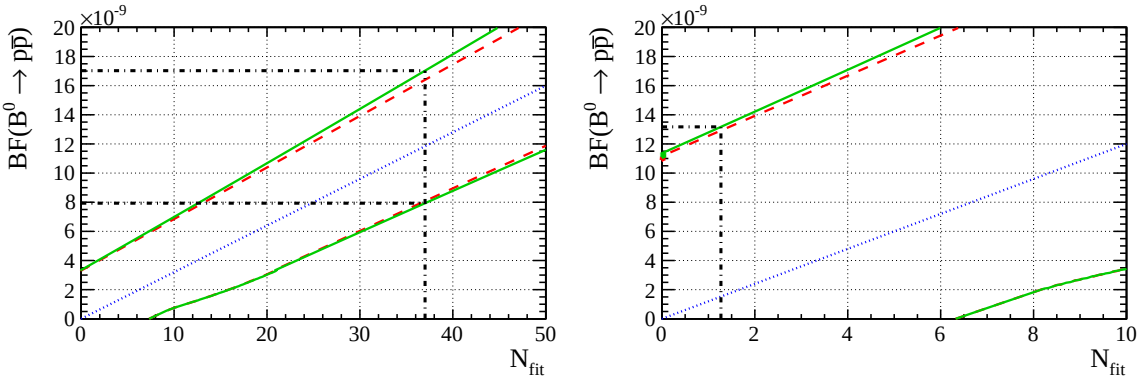
The confidence bands for N_{fit} over the ranges $[0, 50]$ events for $B^0 \to p\overline{p}$ and $[0, 20]$ for $B^0_s \to p\overline{p}$ are shown in Figure 6.3.

From the results shown in Figure 6.3 we obtain 68.27% and 90% CL intervals with full statistical and systematic uncertainties for $\mathcal{B}(B^0_{(s)} \to p\overline{p})$ at the observed signal yields of 37.01 events ($B^0 \to p\overline{p}$) and 1.27 events ($B^0_s \to p\overline{p}$):

$$\begin{aligned}
\mathcal{B}(B^0 \to p\overline{p}) &= (1.18\,^{+0.27\,+0.12}_{-0.24\,-0.08}) \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0 \to p\overline{p}) &= (1.18\,^{+0.46\,+0.24}_{-0.37\,-0.13}) \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\overline{p}) &< 0.82 \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\overline{p}) &< 1.32 \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad .
\end{aligned}$$



(a) $B^0 \to p\bar{p}$ 68.27% CL **(b)** $B^0_s \to p\bar{p}$ 68.27% CL



(c) $B^0 \to p\bar{p}$ 90% CL **(d)** $B^0_s \to p\bar{p}$ 90% CL

Figure 6.3: FC confidence level intervals on the signal yield at 68.27% (a, b) and 90% (c, d) confidence levels for the $B^0_{(s)} \to p\bar{p}$ signal modes. The blue dotted lines show the central value for $N_{\rm fit}$ used in the calculation. The red dotted lines show the lower and upper limits with only statistical uncertainties included, while the green solid lines show the lower and upper limits with statistical and systematic uncertainties included. The black dashed show the limits extracted from this analysis for the 37.01 observed $B^0 \to p\bar{p}$ events and 1.27 observed $B^0_s \to p\bar{p}$ events.

Adding all errors in quadrature yields

$$\begin{aligned}
\mathcal{B}(B^0 \to p\bar{p}) &= 1.18^{+0.30}_{-0.25} \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0 \to p\bar{p}) &= 1.18^{+0.52}_{-0.39} \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\bar{p}) &< 0.82 \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\bar{p}) &< 1.32 \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad .
\end{aligned}$$

6.2.6 Significance of Signals

To assess the statistical significance of the $B^0 \to p\overline{p}$ ($B_s^0 \to p\overline{p}$) signals after unblinding, accounting for fit systematic uncertainties, the fitter is first run over the data in the same settings as before except for the B^0 (B_s^0) fraction fixed at various values corresponding to $B_{(s)}^0 \to p\overline{p}$ yields across the range $[-10, 60]$ ($[-10, 20]$). At each "scanning point", *i.e.* for each B^0 or B_s^0 signal yield considered[3], the difference in the logarithm of the likelihoods between the nominal fit and that at the fixed yield is computed ($-\Delta\ln L$). The ensemble of scanning points composes the profile of the negative logarithm of the fit likelihood against the considered component fraction. Fit systematic uncertainties are incorporated convolving the likelihood itself (not its logarithm) with a Gaussian of width equal to the total systematic uncertainty on the signal yield.

Figure 6.4 shows the negative logarithm of the profile likelihoods against the $B_{(s)}^0 \to p\overline{p}$ signal yields with and without systematic uncertainties. In both sub-figures the $-\Delta\ln L$ curve reaches a minimum at the fitted yield from the nominal fit. The red solid curves correspond to the profile likelihoods calculated with statistical uncertainties only while the blue dashed curves include the systematic uncertainties.

Applying Equation 5.10 to Figure 6.4 we obtain the following $B_{(s)}^0 \to p\overline{p}$ signal statistical significances:

$$\begin{aligned} \text{Significance } (B^0 \to p\overline{p}) &= 5.92\sigma \quad , \quad 5.94\sigma \text{ (stat only)} \quad , \\ \text{Significance } (B_s^0 \to p\overline{p}) &= 0.32\sigma \quad , \quad 0.32\sigma \text{ (stat only)} \quad . \end{aligned}$$

6.2.7 Conclusion

We have presented a search for the rare two-body charmless baryonic B-decays $B^0 \to p\overline{p}$ and $B_s^0 \to p\overline{p}$ using a combined sample of 2011 and 2012 pp collision data from the LHCb detector, corresponding to an integrated luminosity of 3.122 fb^{-1}.

The 68.27% and 90% confidence level intervals on the branching fraction of $B^0 \to p\overline{p}$ and the upper limit of $B_s^0 \to p\overline{p}$ are determined to be, from an unbinned maximum likelihood fit,

[3]The transformation of signal fractions into yields is trivial.

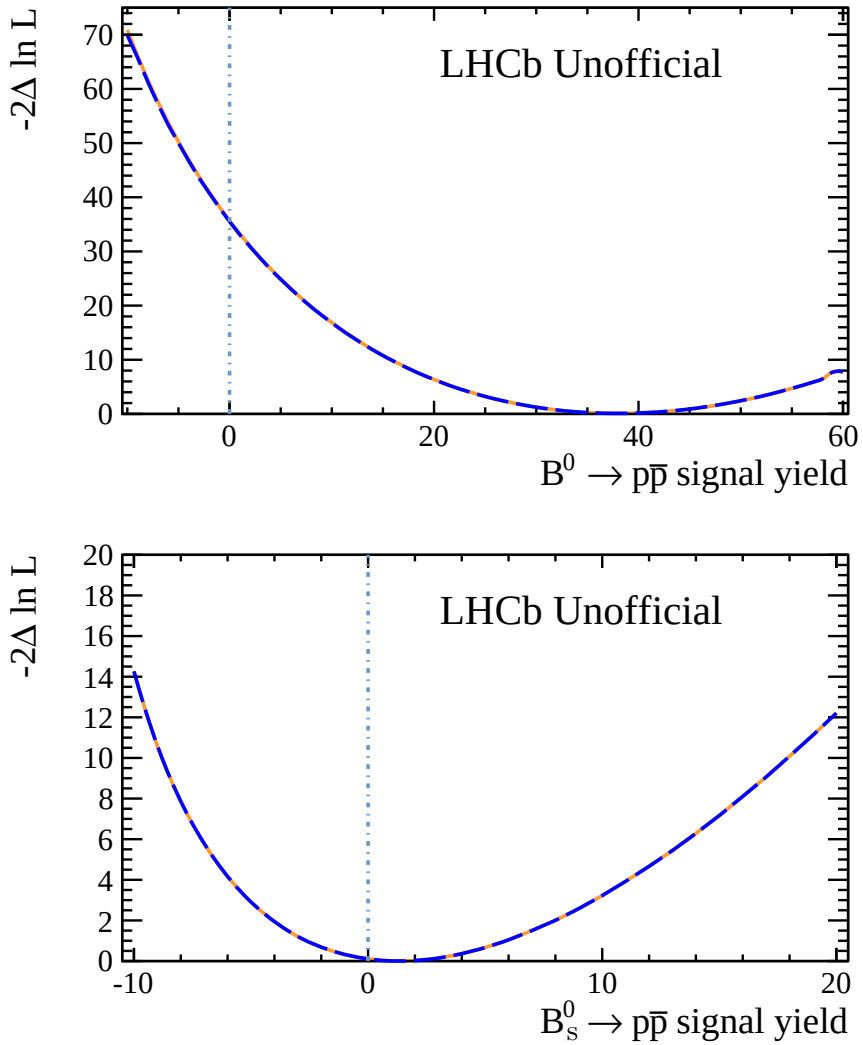


Figure 6.4: Negative logarithm of the profile likelihoods as a function of the $B^0 \to p\overline{p}$ signal yield (top) and the $B^0_s \to p\overline{p}$ signal yield (bottom). The orange solid curves correspond to the statistical-only profiles whereas the blue dashed curves include systematic uncertainties.

$$\begin{aligned}
\mathcal{B}(B^0 \to p\overline{p}) &= (1.18\,^{+0.27}_{-0.24}\,^{+0.12}_{-0.08}) \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0 \to p\overline{p}) &= (1.18\,^{+0.46}_{-0.37}\,^{+0.24}_{-0.13}) \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\overline{p}) &< 0.82 \times 10^{-8} \quad \text{at} \quad 68.27\% \quad \text{CL} \quad , \\
\mathcal{B}(B^0_s \to p\overline{p}) &< 1.32 \times 10^{-8} \quad \text{at} \quad 90\% \quad \text{CL} \quad .
\end{aligned}$$

where the first errors are statistical and the second are systematic.

We observe an excess of $B^0 \to p\overline{p}$ candidates with respect to background expectations

with a statistical significance of 5.92 standard deviations. With a significance above five standard deviations this would signify the first experimental observation the $B^0 \to p\overline{p}$ decay, and at the same time the first observation of a two-body charmless baryonic B^0 decay. It is also the smallest hadronic B^0 branching fraction ever measured. No significant $B_s^0 \to p\overline{p}$ signal is observed and an upper limit to its branching fraction of 1.3×10^{-8} at 90% CL is obtained. This corresponds to a factor of five improvement over the previous LHCb measurement [1].

Figure 6.5 shows a comparison of this latest measurement of $\mathcal{B}\left(B^0 \to p\bar{p}\right)$ with existing experimental results and published theoretical predictions. The measured $B^0 \to p\overline{p}$ branching fraction excludes all published theoretical predictions before 2014 and is consistent with the more recent predictions made in References [30] and [38]. The discovery of the $B^0 \to p\overline{p}$ decay gives motivation for further studies of baryonic B meson decays.

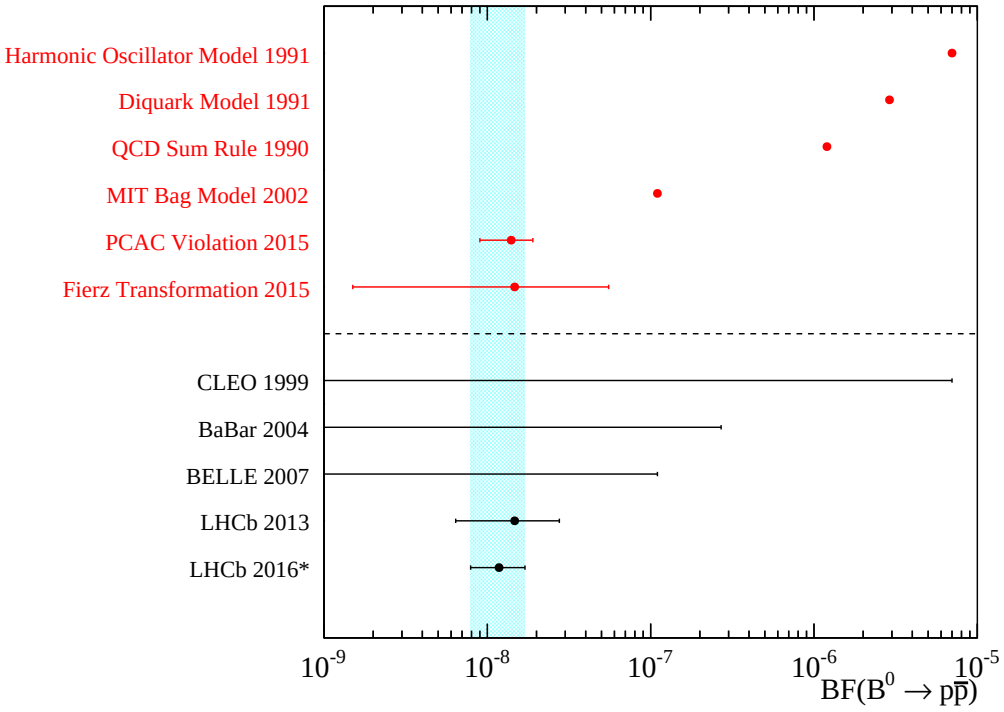


Figure 6.5: Theoretical and experimental limits of $\mathcal{B}\left(B^0 \to p\overline{p}\right)$ including the result from the combined 2011 and 2012 LHCb dataset. Theoretical predictions are shown in red and are taken from the values listed in Table 1.5 with the inclusion of the two theoretical predictions published since the publication of the previous LHCb $\mathcal{B}\left(B^0 \to p\overline{p}\right)$ measurement. Experimental results are shown in black and are taken from 90% CL values.

Appendix A

Hadronisation Fractions

The hadronisation fractions f_q represent the b hadronisation probabilities to the possible b-flavoured final states. In this analysis all f_q values have been taken from LHCb measurements or HFAG world averages [71] assuming isospin symmetry, namely $f_u/f_d = 1$. They are collected in Table A.1.

It is nevertheless important to stress that the values given in the table have only been used to determine expected signal and background yields, and to assess relative background contributions.

Measured quantity	Experimental result
f_u/f_d	1 (assumed in [71])
f_s/f_d	0.259 ± 0.013 [71]
$f_{\Lambda_b^0}/f_d$	$0.151 + \exp(-0.573 - 0.095 \times p_T[\,\text{GeV/c}])$ [82]
f_q	Used value
f_d	0.406 ± 0.005 [71]
f_u	0.406 ± 0.005 [71]
f_s	0.105 ± 0.005 [71]
f_{baryon}	0.083 ± 0.010 [71]

Table A.1: Hadronisation fractions used in the analysis. The quoted $f_{\Lambda_b^0}/f_d$ result was obtained assuming an average p_T of 10 GeV/c, see [82] for details.

Appendix B

Comparison of Distributions of Key Variables

This analysis relies heavily on the use of simulated MC events. It is therefore important that the available MC samples accurately represent what is seen in data. To verify the accuracy of the MC, comparisons are made of the distributions of key variables between MC and data. As the $B^0_{(s)} \to p\overline{p}$ data is blinded it is not possible to compare data and MC using this decay. Instead, these data/MC comparisons are made using the $B^0 \to K^+\pi^-$ normalisation channel. It is of particular importance to verify that the variables used as inputs to the anaysis MVA selection (see Section 4.3.4) are well modelled in the MC. To avoid any potential biases in the comparison results, the selections applied to the data and MC comparison datasets comprise the full $B^0 \to K^+\pi^-$ selection minus the MVA selection.

The signal distributions in data are extracted using the *sPlot* technique [66], using an invariant mass fit to extract the *sWeights*. The parameters of this fit to data are identical to those used in the analysis $B^0 \to K^+\pi^-$ invariant mass fit (see Table 5.11). Figure B.1 shows the result of the invariant mass fit used to extract the *sWeights*.

The distributions in MC are taken from combined samples of 2011 and 2012 $B^0 \to K^+\pi^-$ MC.

The variables chosen for comparison are:

- B-decay vertex distance in z (in mm) from the related primary vertex;
- B momentum, p;
- B transverse momentum, p_T;

- B Impact Parameter (IP) χ^2;
- B vertex χ^2/ nDoF;
- Cosine(B direction angle)
- A_{p_T}, the p_T asymmetry of the B within a cone of radius $R = 1.0$ around the B.
- B-daughters' distance of closest approach (DOCA);
- Daughter (K^+, π^-) IP χ^2;
- Daughter transverse momentum, p_T;
- Daughter momentum, p;
- Daughter pseudorapidity, η.

Figures B.2 and B.3 show the resulting data and MC distributions for the chosen mother and daughter particle comparison variables. In each figure, the histograms for the *sWeighted* data have had their integrals scaled to match the MC. No re-weighting is performed when the signal shapes are determined from MC, as the daughters' momentum distributions (Figures B.3(a) and B.3(b)) are reasonably well described in the simulation. Most of the remaining variables can be seen to be well described in the simulation. The only observed discrepancies between data and MC are in the mother's momentum (Figure B.2(a) and cone p_T-asymmetry (Figure B.2(c)). However, these discrepancies are small and will further be partially cancelled in the ratio of efficiencies between the $B^0 \to K^{\pm}\pi^{\mp}$ and $B^0_{(s)} \to p\overline{p}$ modes.

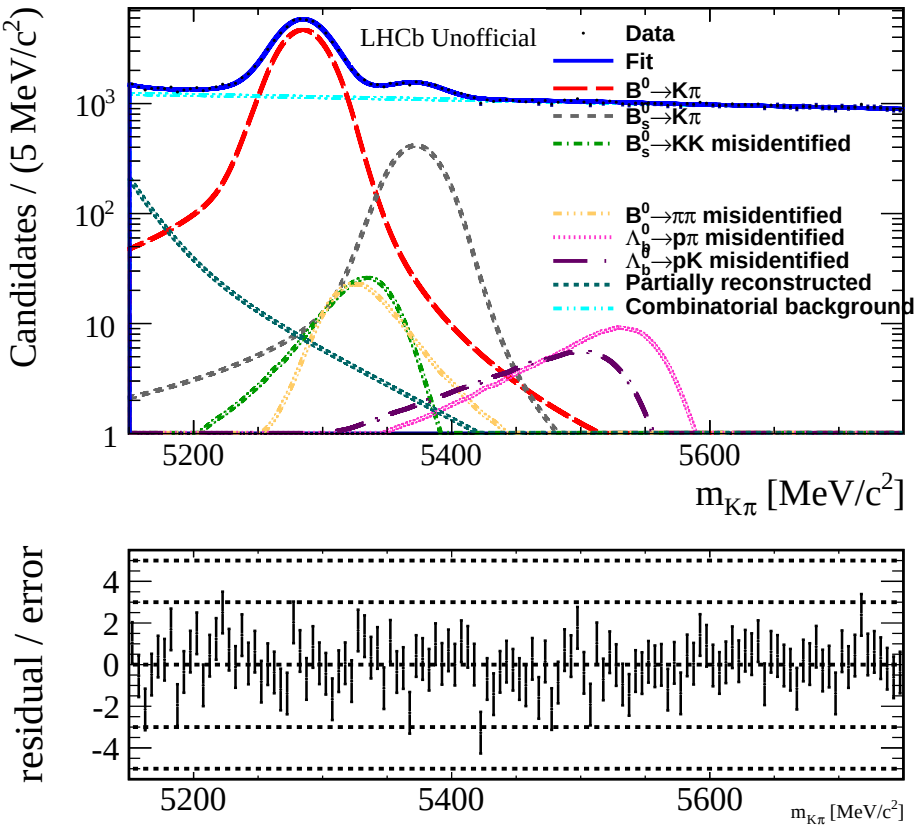
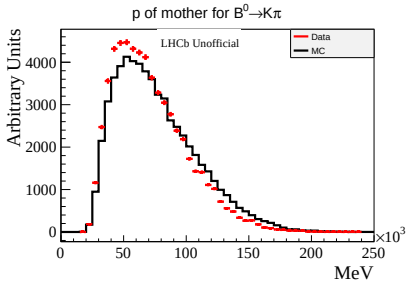
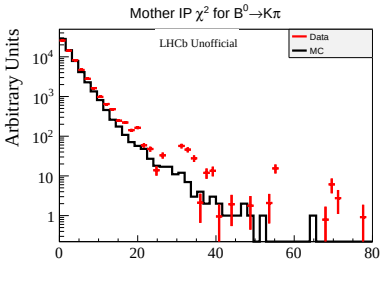


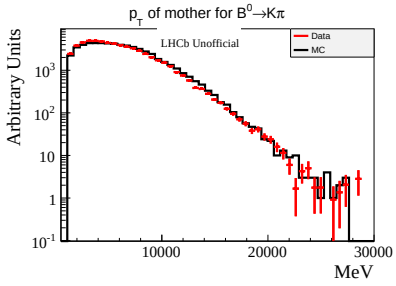
Figure B.1: Mass fit to the $B^0 \to K^+\pi^-$ normalisation channel before MVA selection. The lower plot shows the binned residuals of the fit divided by the fit error. The parameters are extracted from an unbinned maximum likelihood fit, the binning of the data is only for illustration.



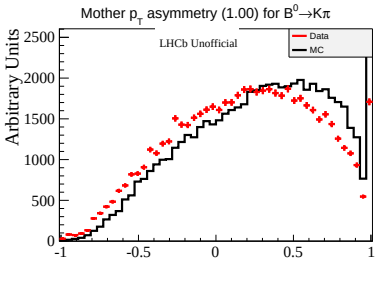
(a) B^0 Momentum, p



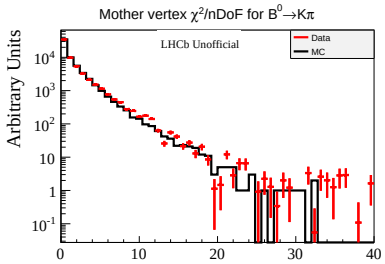
(b) B^0 IPχ^2



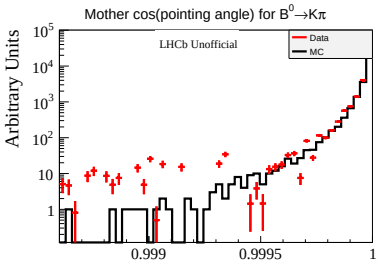
(c) B^0 Transverse Momentum, p_T



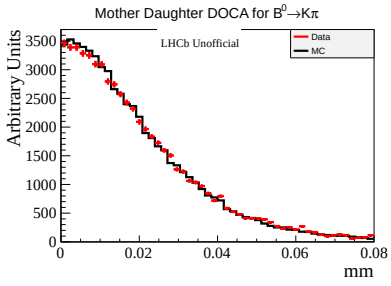
(d) B^0 p_T Asymmetry in cone of $\Delta R = 1.00$



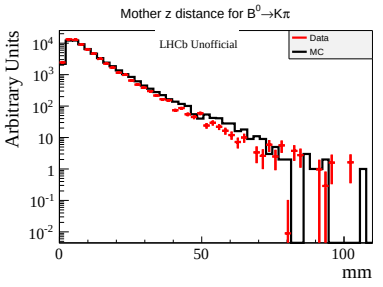
(e) B^0 Vertex $\chi^2/nDoF$



(f) B^0 cos(pointing angle)

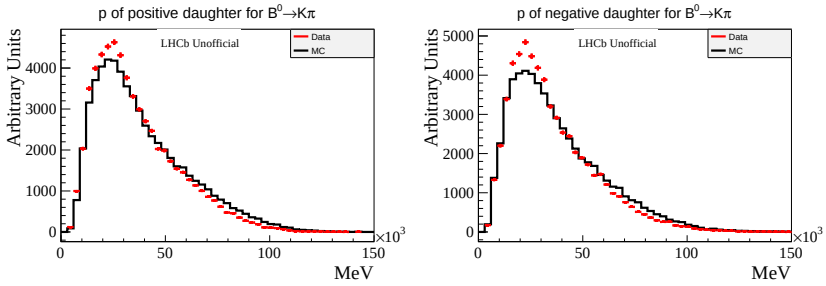


(g) B^0 daughter DOCA



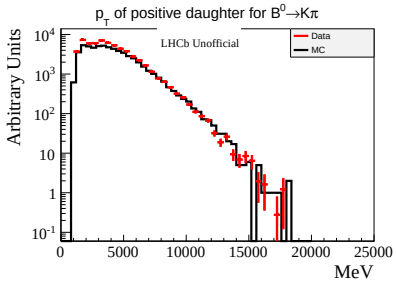
(h) B^0 decay vertex z distance to PV

Figure B.2: Distributions of key mother particle variables for $B^0 \to K^+\pi^-$ events. The distributions of data events are shown in red while MC simulation events are shown in black. The data histograms come from *sPlot*s of the $B^0 \to K^\pm\pi^\mp$ signal peak and the corresponding MC distributions come from the triggered, stripped and PID selected combined 2011 and 2012 MC simulation samples.

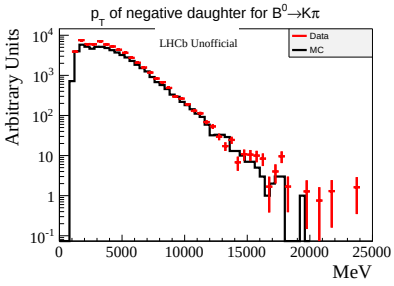


(a) K^+ Momentum, p

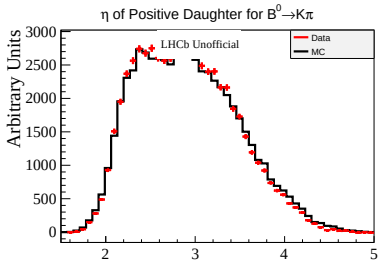
(b) π^- Momentum, p



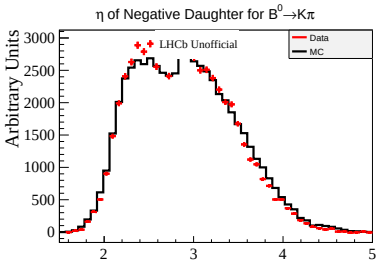
(c) K^+ Transverse Momentum, p_T



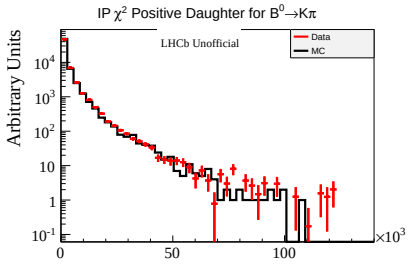
(d) π^- Transverse Momentum, p_T



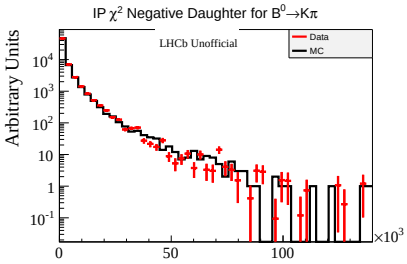
(e) K^+ Pseudorapidity, η



(f) π^- Pseudorapidity, η



(g) K^+ IPχ^2



(h) π^- IPχ^2

Figure B.3: Distributions of key daughter particle variables for $B^0 \to K^+\pi^-$ events. The distributions of data events are shown in red while MC simulation events are shown in black. The data histograms come from *sPlot*s of the $B^0 \to K^\pm\pi^\mp$ signal peak and the corresponding MC distributions come from the triggered, stripped and PID selected combined 2011 and 2012 MC samples.

Appendix C

MVA Preselection for PID Optimisation

Figures C.1 to C.5 show the MVA input variables, MVA method performance and signal and background MLP ANN responses for the MVA preselection applied during the PID selection optimisation described in Section 4.3.4. The choice of input variables and MVA method are identical to those used in the full analysis MVA selection detailed in Section 4.3.4.

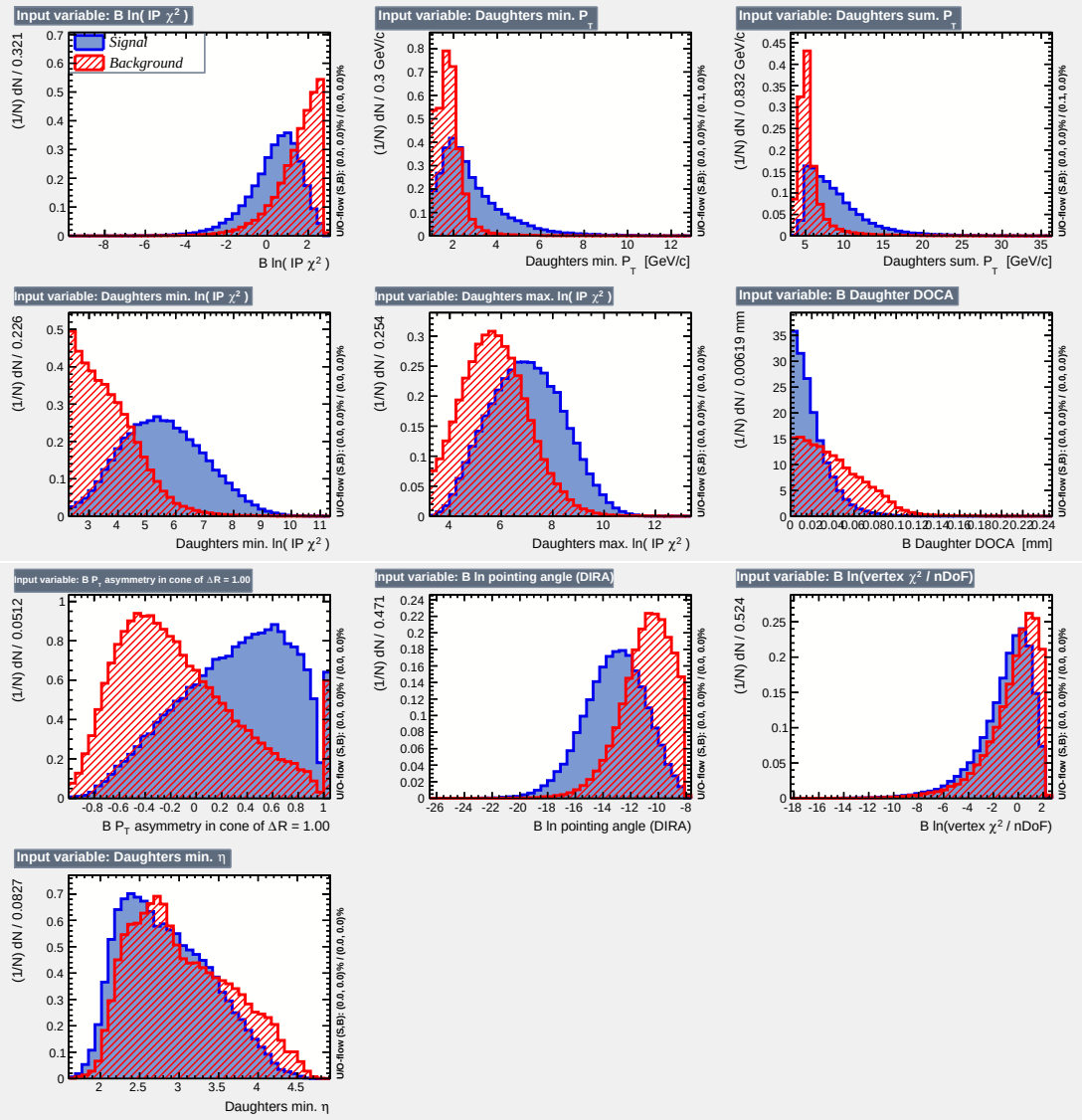


Figure C.1: Distributions of the TMVA input discriminating variables for MC signal and even numbered event sideband data samples used in the MVA preselection for the PID selection optimisation.

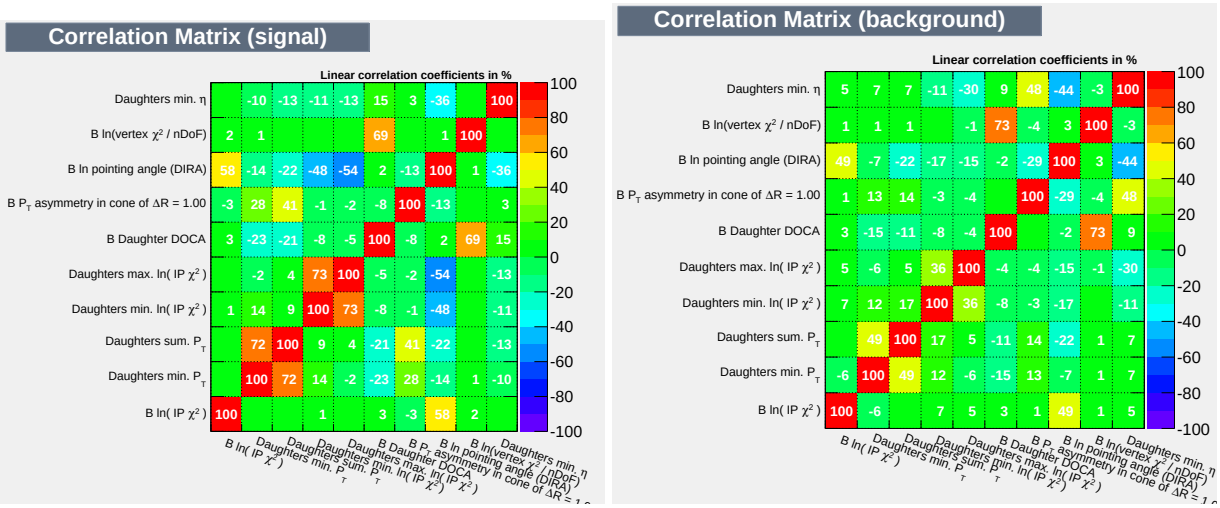


Figure C.2: Correlation matrices of MVA input variables for signal MC (left) and even numbered event sideband data (right) samples used in the MVA preselection for the PID selection optimisation.

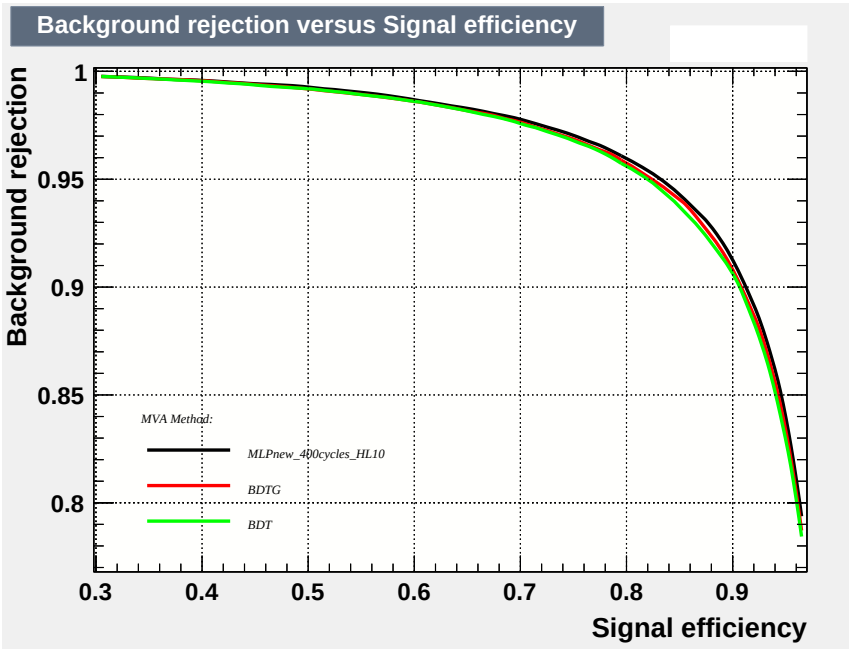


Figure C.3: Background rejection versus signal efficiency curves for MVA methods used in the MVA preselection for the PID selection optimisation.

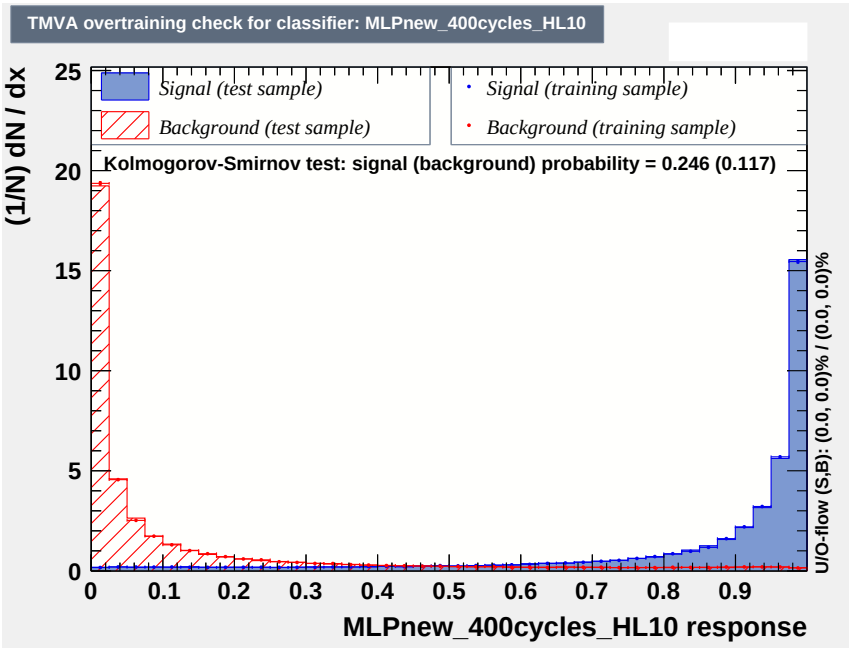


Figure C.4: MLP ANN distributions for signal and background samples for the training and testing samples used in the MVA preselection for the PID selection optimisation.

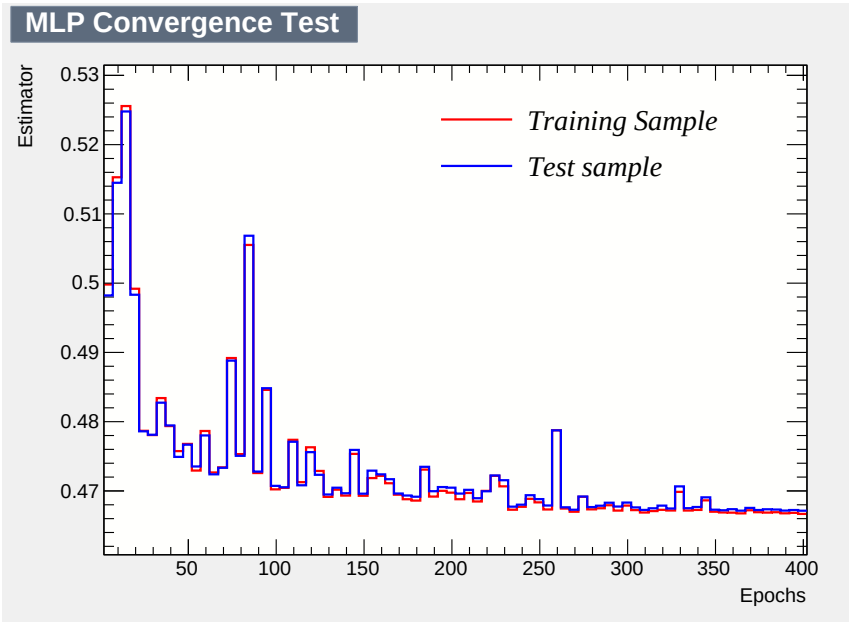


Figure C.5: MLP ANN convergence test of the MLP method used in the MVA preselection for the PID selection optimisation.

Appendix D

PID Binning Studies

To obtain accurate results with minimal systematic uncertainties while using `PIDCalib` calibration methods it is important to carefully choose an appropriate kinematic binning scheme. For this analysis, studies were conducted to determine the optimal binning variables and binning scheme for proton PID calibration. The choice of kinematic variables is intended to cover the full kinematic dependencies of the PID response which is known to depend on particle track p, p_T and η as well as event track multiplicity. As described in Section 4.3.4, track multiplicity is not considered as a binning variable for this analysis. Instead, several two-dimensional binning schemes in both $p-p_T$ and $p-\eta$ were considered. In total, fourteen individual binning schemes were studied, eight in $p-p_T$ phasespace and six in $p-\eta$. The binning structures of the fourteen studied binning schemes are listed in Tables D.1 and D.2 where schemes 1 to 8 are binned in $p-p_T$ and 11 to 16 in $p-\eta$. Each $p-p_T$ binning scheme covers the kinematic range $p \in [5000, 150000]$ MeV/c and $p_T \in [900, 15000]$ MeV/c and each $p-\eta$ scheme covers the range $p \in [5000, 150000]$ MeV/c and $\eta \in [1.5, 5]$.

For each binning scheme the `PIDCalib` multitrack method is used to calculate the overall mean efficiency, for both daughters, of a PID selection of $\mathrm{DLL}_{p\pi} > 15$ and $\mathrm{DLL}_{p\mathrm{K}} > 5$ for 2011 and 2012, magnet up and down, $B^0 \to p\overline{p}$ MC. Figure D.1 shows the resulting PID efficiencies (top) and relative uncertainties (bottom) as a function of scheme number, where the relative uncertainty, σ^{rel} is given by

$$\sigma^{rel} = \frac{\sigma_{PID}}{\epsilon_{PID}} \times 100. \tag{D.1}$$

Comparing the $p-p_T$ (1-8) and $p-\eta$ schemes (11-16) it's evident that the spread of efficiencies between schemes is smaller within the $p-\eta$ schemes compared to the

$p-p_T$ schemes (1-8) which show a large discrepancy in overall efficiency between the three schemes with the coarsest p_T binning (schemes 5, 6 and 7) and the remaining schemes. Also, the relative uncertainties for many of the $p-p_T$ schemes are higher than most of the $p-\eta$ schemes. With this all taken into consideration, we choose p and η as the kinematic binning variables for the proton PID calibration. We select scheme 15 as the basis for the proton PID calibration binning scheme as it has small relative uncertainties across all four MC samples whilst returning comparable PID efficiency values to the other $p-\eta$ schemes. The variations in efficiency between schemes when evaluating the $B^0_{(s)} \to p\overline{p}$ main analysis PID selection are included as a contribution to the PID selection systematic uncertainties which are detailed in Section 6.1.5.

Scheme	p Bins (MeV/c)	p_T Bins (MeV/c)
1	5000 : 6944.44 : 8888.89 : 10833.3 : 12777.8 : 14722.2 : 16666.7 : 18611.1 : 20555.6 : 22500 : 24444.4 : 26388.9 : 28333.3 : 30277.8 : 32222.2 : 34166.7 : 36111.1 : 38055.6 : 40000 : 46000 : 52000 : 58000 : 64000 : 70000 : 75000 : 80000 : 85000 : 90000 : 95000 : 100000 : 105000 : 110000 : 115000 : 120000 : 150000	900 : 1810 : 2720 : 3630 : 4540 : 5450 : 6360 : 7270 : 8180 : 9090 : 10000 : 11250 : 12500 : 13750 : 15000
2	5000 : 8888.89 : 12777.8 : 16666.7 : 20555.6 : 24444.4 : 28333.3 : 32222.2 : 36111.1 : 40000 : 50000 : 60000 : 70000 : 80000 : 90000 : 100000 : 110000 : 120000 : 150000	900 : 1810 : 2720 : 3630 : 4540 : 5450 : 6360 : 7270 : 8180 : 9090 : 10000 : 11250 : 12500 : 13750 : 15000
3	5000 : 6944.44 : 8888.89 : 10833.3 : 12777.8 : 14722.2 : 16666.7 : 18611.1 : 20555.6 : 22500 : 24444.4 : 26388.9 : 28333.3 : 30277.8 : 32222.2 : 34166.7 : 36111.1 : 38055.6 : 40000 : 46000 : 52000 : 58000 : 64000 : 70000 : 75000 : 80000 : 85000 : 90000 : 95000 : 100000 : 105000 : 110000 : 115000 : 120000 : 150000	900 : 1506.67 : 2113.33 : 2720 : 3326.67 : 3933.33 : 4540 : 5146.67 : 5753.33 : 6360 : 6966.67 : 7573.33 : 8180 : 8786.67 : 9393.33 : 10000 : 10833.3 : 11666.7 : 12500 : 13333.3 : 14166.7 : 15000
4	5000 : 8888.89 : 12777.8 : 16666.7 : 20555.6 : 24444.4 : 28333.3 : 32222.2 : 36111.1 : 40000 : 50000 : 60000 : 70000 : 80000 : 90000 : 100000 : 110000 : 120000 : 150000	900 : 1506.67 : 2113.33 : 2720 : 3326.67 : 3933.33 : 4540 : 5146.67 : 5753.33 : 6360 : 6966.67 : 7573.33 : 8180 : 8786.67 : 9393.33 : 10000 : 10833.3 : 11666.7 : 12500 : 13333.3 : 14166.7 : 15000
5	5000 : 6944.44 : 8888.89 : 10833.3 : 12777.8 : 14722.2 : 16666.7 : 18611.1 : 20555.6 : 22500 : 24444.4 : 26388.9 : 28333.3 : 30277.8 : 32222.2 : 34166.7 : 36111.1 : 38055.6 : 40000 : 46000 : 52000 : 58000 : 64000 : 70000 : 75000 : 80000 : 85000 : 90000 : 95000 : 100000 : 105000 : 110000 : 115000 : 120000 : 150000	900 : 2720 : 4540 : 6360 : 8180 : 10000 : 11666.7 : 13333.3 : 15000
6	5000 : 8888.89 : 12777.8 : 16666.7 : 20555.6 : 24444.4 : 28333.3 : 32222.2 : 36111.1 : 40000 : 50000 : 60000 : 70000 : 80000 : 90000 : 100000 : 110000 : 120000 : 150000	900 : 2720 : 4540 : 6360 : 8180 : 10000 : 11666.7 : 13333.3 : 15000
7	5000 : 8750 : 12500 : 16250 : 20000 : 28000 : 36000 : 44000 : 52000 : 60000 : 68000 : 76000 : 84000 : 92000 : 100000 : 110000 : 120000 : 130000 : 140000 : 150000	900 : 2720 : 4540 : 6360 : 8180 : 10000 : 11666.7 : 13333.3 : 15000
8	5000 : 8750 : 12500 : 16250 : 20000 : 28000 : 36000 : 44000 : 52000 : 60000 : 68000 : 76000 : 84000 : 92000 : 100000 : 110000 : 120000 : 130000 : 140000 : 150000	900 : 1506.67 : 2113.33 : 2720 : 3326.67 : 3933.33 : 4540 : 5146.67 : 5753.33 : 6360 : 6966.67 : 7573.33 : 8180 : 8786.67 : 9393.33 : 10000 : 10833.3 : 11666.7 : 12500 : 13333.3 : 14166.7 : 15000

Table D.1: Proton PIDCalib kinematic binning schemes in $(p - p_T)$ studied.

Scheme	p Bins (MeV/c)	η Bins
11	5000 : 6944.44 : 8888.89 : 10833.3 : 12777.8 : 14722.2 : 16666.7 18611.1 : 20555.6 : 22500 : 24444.4 : 26388.9 : 28333.3 : 30277.8 32222.2 : 34166.7 : 36111.1 : 38055.6 : 40000 : 46000 : 52000 58000 : 64000 : 70000 : 75000 : 80000 : 85000 : 90000 95000 : 100000 : 105000 : 110000 : 115000 : 120000 : 150000	1.5 : 1.71875 : 1.9375 : 2.15625 : 2.375 : 2.59375 2.8125 : 3.03125 : 3.25 : 3.46875 : 3.6875 4.5625 : 4.78125 : 3.90625 : 4.125 : 4.34375 : 5
12	5000 : 8888.89 : 12777.8 : 16666.7 : 20555.6 : 24444.4 : 28333.3 32222.2 : 36111.1 : 40000 : 50000 : 60000 : 70000 : 80000 90000 : 100000 : 110000 : 120000 : 150000	1.5 : 1.71875 : 1.9375 : 2.15625 : 2.375 2.59375 : 2.8125 : 3.03125 : 3.25 : 3.46875 3.6875 : 3.90625 : 4.125 : 4.34375 : 4.5625 4.78125 : 5
13	5000 : 6944.44 : 8888.89 : 10833.3 : 12777.8 : 14722.2 : 16666.7 18611.1 : 20555.6 : 22500 : 24444.4 : 26388.9 : 28333.3 : 30277.8 32222.2 : 34166.7 : 36111.1 : 38055.6 : 40000 : 46000 : 52000 58000 : 64000 : 70000 : 75000 : 80000 : 85000 : 90000 95000 : 100000 : 105000 : 110000 : 115000 : 120000 : 150000	1.5 : 2 : 2.25 : 2.5 : 2.75 3 : 3.25 : 3.5 : 3.75 : 4 4.33333 : 4.66667 : 5
14	5000 : 8888.89 : 12777.8 : 16666.7 : 20555.6 : 24444.4 : 28333.3 32222.2 : 36111.1 : 40000 : 50000 : 60000 : 70000 : 80000 90000 : 100000 : 110000 : 120000 : 150000	1.5 : 2 : 2.25 : 2.5 : 2.75 3 : 3.25 : 3.5 : 3.75 : 4 4.33333 : 4.66667 : 5
15	5000 : 8750 : 12500 : 16250 : 20000 : 28000 : 36000 44000 : 52000 : 60000 : 68000 : 76000 : 84000 : 92000 100000 : 110000 : 120000 : 130000 : 140000 : 150000	1.5 : 2 : 2.25 : 2.5 : 2.75 3 : 3.25 : 3.5 : 3.75 : 4 4.33333 : 4.66667 : 5
16	5000 : 8750 : 12500 : 16250 : 20000 : 28000 : 36000 44000 : 52000 : 60000 : 68000 : 76000 : 84000 : 92000 100000 : 110000 : 120000 : 130000 : 140000 : 150000	1.5 : 1.71875 : 1.9375 : 2.15625 : 2.375 2.59375 : 2.8125 : 3.03125 : 3.25 : 3.46875 3.6875 : 3.90625 : 4.125 : 4.34375 : 4.5625 4.78125 : 5

Table D.2: Proton PIDCalib kinematic binning schemes in $(p - \eta)$ studied.

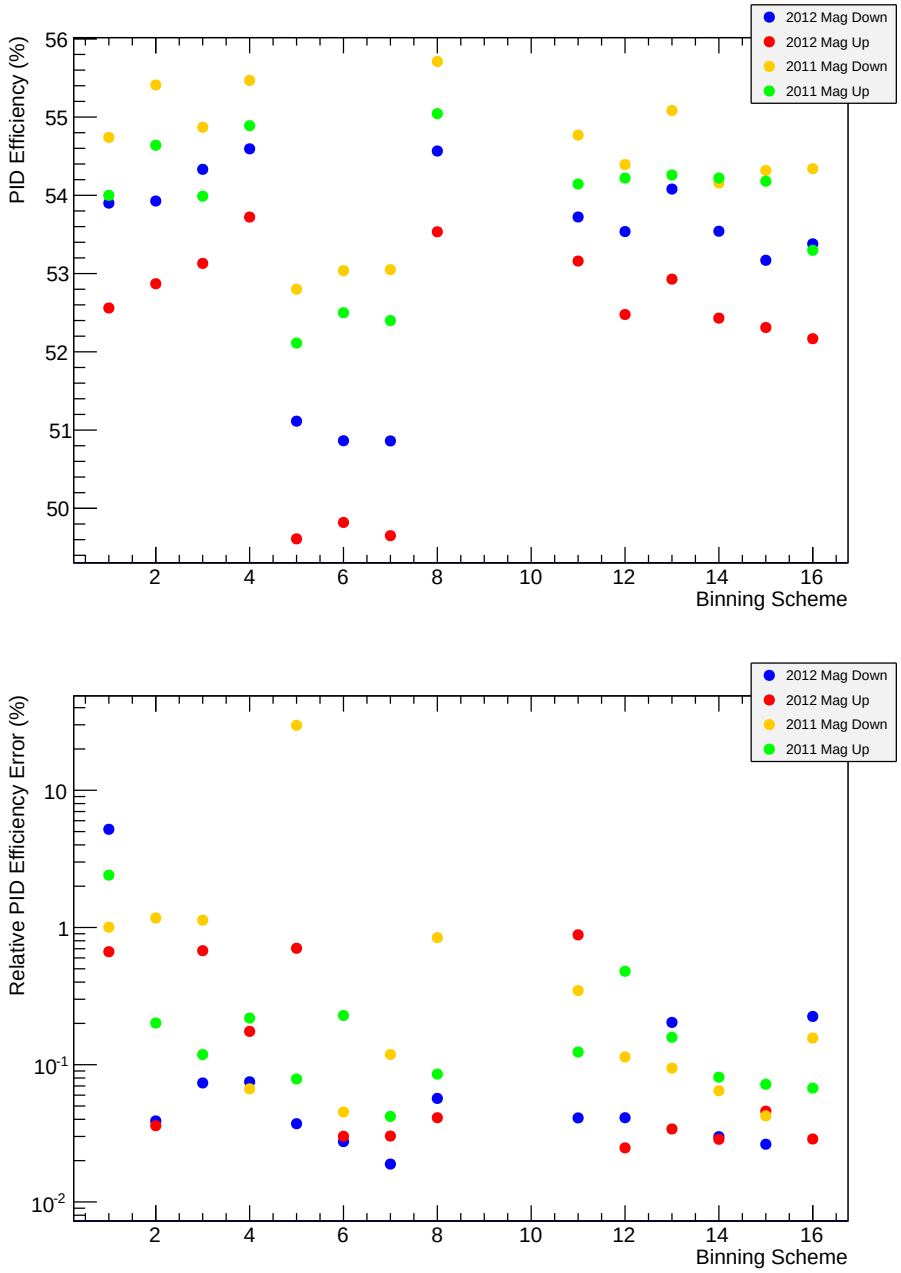


Figure D.1: PIDCalib $B^0 \to p\overline{p}$ PID efficiencies (top) and statistical uncertainties (bottom) for a range of different binning schemes. The efficiencies are evaluated for a PID selection of $p(\bar{p})$ $\mathrm{DLL}_{p\pi} > 15$ and $\mathrm{DLL}_{p\mathrm{K}} > 5$. The points along the x-axis correspond to different binning schemes which are detailed in Tables D.1 and D.2, with schemes 1-10 binned in $p - p_T$ phasespace and 11-16 binned in $p - \eta$.

Bibliography

[1] LHCb collaboration, R. Aaij *et al.*, *First evidence for the two-body charmless baryonic decay* $B^0 \to p\bar{p}$, J. High Energy Phys. **10** (2013) 005. 12 p.

[2] Particle Data Group, K. A. Olive *et al.*, *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.

[3] Pauli, W, *The Connection Between Spin and Statistics*, Phys. Rev. **58** (1940) 716.

[4] E. Majorana, *Teoria Simmetrica Dell'Elettrone e Del Positrone*, Nuovo Cim. **14** (1937) 171.

[5] Super-Kamiokande Collaboration, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562.

[6] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964), no. 2 132 .

[7] G. Aad *et al.*, *Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc*, Physics Letters B **716** (2012), no. 1 1 .

[8] CMS, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv:1207.7235.

[9] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10** (1963) 531.

[10] Glashow, S.L. and Iliopoulos, J. and Maiani, L, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285.

[11] M. Kobayashi and T. Maskawa, *CP Violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973) 652.

[12] Chau, Ling-Lie and Keung, Wai-Yee, *Comments on the Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **53** (1984) 1802.

[13] T. C. Group, *CP violation and the CKM matrix: assessing the impact of the asymmetric B factories*, Eur. Phys. J. **C41** (2005) 1.

[14] E. Noether, *Invariante Variationsprobleme*, Nachr. d. König. Gesselsch. d. Wiss. zu Göttigen (1918) 235.

[15] Schwinger, Julian S, *A Theory of the Fundamental Interactions*, Annals Phys. **2** (1957) 407.

[16] J. Christenson, J. Cronin, V. Fitch, and R. Turlay, *Evidence for the 2π decay of the K_2^0 meson.*, Phys. Rev. Lett. **13** (1964) 138.

[17] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[18] C. Jarlskog, *Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal* CP *nonconservation*, Phys. Rev. Lett. **55** (1985) 1039.

[19] LHCb collaboration, R. Aaij *et al.*, *First measurement of time-dependent CP violation in $B_s^0 \to K^+K^-$ decays*, J. High Energy Phys. **10** (2013) 183. 22 p.

[20] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the effective $B_s^0 \to K^+K^-$ lifetime*, Phys. Lett. B **716** (2012) 393.

[21] LHCb collaboration, R. Aaij *et al.*, *Effective lifetime measurements in the $B_s^0 \to K^+K^-$, $B^0 \to K^+\pi^-$ and $B_s^0 \to \pi^+K^-$ decays*, Phys. Lett. B **736** (2014) 446.

[22] LHCb Collaboration, R. Aaij *et al.*, *Measurement of b-hadron branching fractions for two-body decays into charmless charged hadrons*, J. High Energy Phys. **10** (2012) 037. 21 p.

[23] R. Fleischer, *$B_{s,d} \to \pi\pi, \pi K, KK$: Status and Prospects*, Eur. Phys. J. **C52** (2007) 267, arXiv:0705.1121.

[24] M. Suzuki, *Partial waves of baryon-antibaryon in three-body B meson decay*, J. Phys. **G34** (2007) 283, arXiv:hep-ph/0609133.

[25] CLEO Collaboration, T. Coan *et al.*, *Search for exclusive rare baryonic decays of B mesons*, Phys. Rev. **D59** (1999) 111101, arXiv:hep-ex/9810043.

[26] BaBar Collaboration, B. Aubert *et al.*, *Search for the decay* $B^0 \to p\bar{p}$, Phys. Rev. **D69** (2004) 091503, arXiv:hep-ex/0403003.

[27] Belle Collaboration, Y.-T. Tsai *et al.*, *Search for* $B^0 \to p\bar{p}, \Lambda\bar{\Lambda}$ *and* $B^+ \to p\bar{\Lambda}$ *at Belle*, Phys. Rev. **D75** (2007) 111101, arXiv:hep-ex/0703048.

[28] LHCb collaboration, R. Aaij *et al.*, *Studies of the decays* $B^+ \to p\bar{p}h^+$ *and observation of* $B^+ \to \bar{\Lambda}(1520)p$, Phys. Rev. D **88** (2013) 052015. 20 p.

[29] ALEPH Collaboration, D. Buskulic *et al.*, *Observation of charmless hadronic B decays*, Physics Letters B **384** (1996), no. 14 471 .

[30] C.-K. Chua, *Charmless Two-body Baryonic* $B_{u,d,s}$ *Decays*, PoS **EPS-HEP2015** (2015) 582.

[31] V. Chernyak and I. Zhitnitsky, *B-meson exclusive decays into baryons*, Nuclear Physics B **345** (1990), no. 1 pp.137.

[32] P. Ball and H. G. Dosch, *Branching ratios of exclusive decays of bottom mesons into baryon-antibaryon pairs*, Zeitschrift fr Physik C Particles and Fields **51** (1991) 445.

[33] M. Jarfi *et al.*, *Pole model of B-meson decays into baryon-antibaryon pairs*, Phys. Rev. D **43** (1991) 1599.

[34] M. Jarfi *et al.*, *Relevance of baryon-antibaryon decays of* B^0, $\overline{B}^0$ *in tests of CP violation*, Physics Letters B **237** (1990), no. 3-4 pp.513.

[35] H.-Y. Cheng and K.-C. Yang, *Charmless exclusive baryonic B decays*, Phys. Rev. D **66** (2002) 014020.

[36] J. Beddow *et al.*, *Search for the Rare Two-body Charmless Baryonic B-decays* $B^0 \to p\overline{p}$ *and* $B_s^0 \to p\overline{p}$, LHCb-ANA-2012-064.

[37] Y. K. Hsiao and C. Q. Geng, *Violation of partial conservation of the axial-vector current and two-body baryonic B and* D_s *decays*, Phys. Rev. **D91** (2015), no. 7 077501, arXiv:1407.7639.

[38] H.-Y. Cheng and C.-K. Chua, *On the smallness of Tree-dominated Charmless Two-body Baryonic B Decay Rates*, Phys. Rev. **D91** (2015), no. 3 036003, arXiv:1412.8272.

[39] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[40] *CERN Accelerators and Technology Sector*, `https://espace.cern.ch/acc-tec-sector/default.aspx`, February, 2016.

[41] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[42] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[43] CMS, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[44] LHCb Collaboration, A. A. Alves *et al.*, *The LHCb Detector at the LHC*, J. Instrum. **3** (2008), no. LHCb-DP-2008-001. CERN-LHCb-DP-2008-001 S08005.

[45] *LHCb : Technical Proposal*. Tech. Proposal. CERN, Geneva, 1998.

[46] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015), no. 07 1530022, arXiv:1412.6352.

[47] LHCb collaboration, R. Aaij *et al.*, *Measurement of* $\sigma(pp \to b\bar{b}X)$ *at* $\sqrt{s} = 7$ *TeV in the forward region*, Phys. Lett. **B694** (2010) 209, arXiv:1009.2731.

[48] LHCb Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb VELO (VErtex LOcator): Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001.

[49] LHCb RICH group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2012) 2431. 25 p.

[50] LHCb RICH Collaboration, N. Brook *et al.*, *LHCb RICH1 Engineering Design Review Report*, Tech. Rep. LHCb-2004-121. CERN-LHCb-2004-121, CERN, Geneva, Oct, 2005.

[51] M. Adinolfi *et al.*, *LHCb RICH 2 engineering design review report*, Tech. Rep. LHCb-2002-009, CERN, Geneva, Mar, 2002.

[52] LHCb Collaboration, S. Amato *et al.*, *LHCb magnet: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2000.

[53] LHCb Collaboration, R. Antunes-Nobrega *et al.*, *LHCb reoptimized detector design and performance: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2003.

[54] *ST Material for Publication*, `http://www.physik.unizh.ch/groups/lhcb/public/material/`, February, 2016.

[55] D. van Eijk *et al.*, *Radiation hardness of the LHCb Outer Tracker*, Nucl. Instrum. Methods Phys. Res. , A **685** (2012), no. LHCb-DP-2012-001. CERN-LHCb-DP-2012-001 62.

[56] P. Perret and X. Vilasis-Cardona, *Performance of the LHCb calorimeters during the period 2010-2012*, J. Phys. : Conf. Ser. **587** (2014) 012012. 6 p, see LHCb-TALK-2014-236.

[57] LHCb, R. Aaij *et al.*, *Measurement of the ratio of branching fractions* $BR(B_0 \to K^{\star 0}\gamma)/BR(B_{s0} \to \phi\gamma)$ *and the direct CP asymmetry in* $B_0 \to K^{\star 0}\gamma$, Nucl. Phys. **B867** (2013) 1, arXiv:1209.0313.

[58] LHCb Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb muon system: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001.

[59] LHCb MuonID group, F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, J. Instrum. **8** (2013) P10020. 17 p.

[60] LHCb Collaboration, R. Antunes-Nobrega *et al.*, *LHCb trigger system: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2003.

[61] The LHCb Collaboration, *Documentation for the LHCb BRUNEL project*, 2016. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/.

[62] The LHCb Collaboration, *Documentation for the LHCb DAVINCI project*, 2016. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/.

[63] L. Carson, *Performance of Hybrid Photon Detectors and Studies of Two-Body Hadronic B Decays at LHCb*, PhD thesis, Glasgow U., 2009.

[64] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton*, arXiv:1207.1158.

[65] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Wadsworth international group, Belmont, California, USA, 1984.

[66] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[67] G. Punzi, *Sensitivity of Searches for New Signals and Its Optimization*, in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology* (L. Lyons, R. Mount, and R. Reitmeyer, eds.), p. 79, 2003. arXiv:physics/0308063.

[68] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the effective $B_s^0 \to K^+K^-$ lifetime*, Phys. Lett. B **716** (2012) 393.

[69] LHCb, R. Aaij *et al.*, *First observation of CP violation in the decays of B_s^0 mesons*, Phys. Rev. Lett. **110** (2013), no. 22 221601, arXiv:1304.6173.

[70] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.

[71] Heavy Flavor Averaging Group (HFAG), Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2014*, arXiv:1412.7515.

[72] LHCb collaboration, R. Aaij *et al.*, *Measurement of b hadron production fractions in 7 TeV pp collisions*, arXiv:1111.2357.

[73] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **1304** (2013) 001, arXiv:1301.5286.

[74] Particle Data Group, J. Beringer *et al.*, *Review of particle physics*, Phys. Rev. **D86** (2012) 010001.

[75] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, The Annals of Mathematical Statistics (1938) 60.

[76] G. J. Feldman and R. D. Cousins, *A Unified Approach to the Classical Statistical Analysis of Small Signals*, Phys. Rev. **D57** (1998) 3873.

[77] LHCb collaboration, R. Aaij *et al.*, *Measurement of b-hadron branching fractions for two-body decays into charmless charged hadrons*, arXiv:1206.2794.

[78] *PIDCalib Package*, `https://twiki.cern.ch/twiki/bin/view/LHCb/PIDCalibPackage`.

[79] K. Cranmer, *Kernel Estimation in High Energy Physics.*, Comput. Phys. Commun. **136** (2001) 198.

[80] C. Hombach *et al.*, *Measurement of the rare Two-body charmless baryonic B-decay* $B^+ \to \overline{\Lambda}p$, LHCb-ANA-2014-094 (2014).

[81] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.

[82] LHCb collaboration, R. Aaij *et al.*, *Study of the kinematic dependences of* Λ_b^0 *production in pp collisions and a measurement of the* $\Lambda_b^0 \to \Lambda_c^+\pi^-$ *branching fraction*, J. High Energy Phys. **08** (2014) 143. 22 p.