Study of the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ with the LHCb experiment

By

SRISHTI BHASIN

School of Physics
UNIVERSITY OF BRISTOL

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

APRIL 2021

Word count: thirty thousand one hundred and five

Abstract

The decay $B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-$ is studied using data from proton-proton collisions collected by the LHCb experiment. The decay is observed for the first time, with 297 ± 14 signal candidates observed in data collected during 2011, 2012 and 2016. The branching fraction is measured relative to that of a control channel $B^0 \to D^{*-} D^0 K^+$, which is topologically similar and has the same final state particles. The ratio of branching fractions is measured to be

$$\mathcal{R} = (14.2 \pm 1.1 \pm 1.0)\%,$$

where the first uncertainty is statistical and the second is systematic. Using the most precise result published for the branching fraction of the control channel, the absolute branching fraction of $B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-$ decays is thus determined to be

$$\mathcal{B}(B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-) = (3.50 \pm 0.27 \pm 0.26 \pm 0.30) \times 10^{-4},$$

where the third uncertainty is due to the branching fraction of the control channel. Studies towards a subsequent amplitude analysis of this decay, using additional data collected in 2017 and 2018, are also presented. The analysis of this decay mode will provide insights on hadronic uncertainties in golden modes at LHCb such as $B^0 \to K^{*0} \mu^+ \mu^-$, and provide opportunities for spectroscopy to observe new and known states. Furthermore, this thesis outlines the development of the TORCH time-of-flight detector and calibration studies performed on its front-end electronics. This novel detector will enhance the particle identification capabilities of future upgrades of the LHCb experiment.

Dedication

I would firstly like to thank my supervisor, Kostas Petridis. You have been a source of constant inspiration and knowledge, and I cannot thank you enough for your enthusiasm and patience in leading me through this PhD. It has been a privilege to have you as my teacher and guide these past few years.

I would also like to thank my second supervisor, Jonas Rademacker. Your insights have been invaluable in this project, and I will always appreciate your persistent cheerleading. And my unofficial third supervisor, Mark Whitehead - all of your help made a big difference in my final year, and I really thank you for your time and support. My thanks also extends to the whole Bristol group for their support and camaraderie.

To my friends, the ones made during my PhD and LTA, and the ones from way before, I am grateful for the many happy memories that kept me going through both the highs and the lows.

To Lucas, for not only being the best partner in life, but also a sounding board and source of unfailing encouragement whenever it was most needed. You made completing a PhD during a pandemic unbelievably enjoyable.

And finally, thank you to my family for supporting this journey from the very beginning and always cheering me on.

AUTHOR'S DECLARATION

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

SIGNED: ..... DATE: .....

Author's contribution

The author's contributions are detailed by the following list

- Chapter 4: The author contributed to the set-up and data-taking of testbeam campaigns, described in Section 4.5, and the subsequent analysis is performed by collaborators. The calibration studies presented in Section 4.4 are performed by the author.
- Chapter 7: The analysis and results in this chapter are conducted by the author. The results and some figures have also been published in Ref. [1]. The data analysed in this chapter is collected using the LHCb detector, by the LHCb collaboration. The simulation samples are produced by the LHCb simulation group. Both the data and simulated samples had the LHCb software applied to it which is written by subgroups of the collaboration.
- Chapter 9: The studies and results presented in this chapter and the related Appendices A.2 and A.3 are the result of the author's own work. The data and simulation samples are from the LHCb collaboration as above, except where noted otherwise.

TABLE OF CONTENTS

List of Tables

List of Figures

Figure **Page**

Abbreviations

AUC	area under the curve
$B_{\text{selection}}$	classifier to select B^0 candidates
BDT	boosted decision tree
BF	branching fraction
BSM	Beyond the Standard Model
BW	Breit-Wigner
CP	charge-parity
$D_{\text{from}B}$	classifier to select D^0 candidates originating from a B^0 meson
DAQ	data acquisition
DIRA	the direction of an angle
DLL_{ab}	difference in log-likelihood between particle a and b hypotheses
DOCA_{ab}	distance of closest approach between two particles a and b
DSCB	double-sided Crystal Ball (function)
DTF	DecayTreeFitter
FCNC	flavour-changing neutral current
FD	flight distance (of a particle before decaying)
$\text{FD}\chi^2$	the change in the χ^2 of a vertex fit when two vertices are combined
FPGA	field-programmable gate array
FPR	false positive rate
HA	helicity amplitude
HLT1	High Level Trigger 1 — the first-level software trigger
HLT2	High Level Trigger 2 — the second-level software trigger
HPTDC	high-performance time-to-digital converter
IP	impact parameter — the minimum distance of a track to a vertex
L0	Level 0 — the hardware trigger
LFU	lepton flavour universality
LHC	Large Hadron Collider
MC	Monte Carlo
MCP	multi-channel plate
ML	machine learning
MLP	multi-layer perceptron
MVA	multivariate analysis
NN	neural network
NP	New Physics

NR	non-resonant
PDF	probability density function
PDG	Particle Data Group — values from Ref. [2] represent known results summarised across experimental measurements
PHSP	phase-space
PID	particle identification
PMT	photomultiplier tube
pp	proton-proton
ProbNN	probability of a particle hypothesis based on the output of a neural network
p_{T}	transverse momentum
PV	primary vertex
QCD	quantum chromodynamics
QED	quantum electrodynamics
ROC	receiver operating characteristic
SM	Standard Model
TOS	trigger on signal
TORCH	Time Of internally Reflected Cherenkov light
TOT	time-over-threshold
TIS	trigger independent of signal
TPR	true positive rate
vev	vacuum expectation value
χ^2/ndf	χ^2 divided by the number of degrees of freedom
χ^2_{IP}	IP χ^2— the change in the χ^2 of a vertex fit with and without the track under consideration

CHAPTER 1

INTRODUCTION

The Standard Model (SM) of particle physics represents the best description of all known elementary particles and their interactions to date. Its predictions have been thoroughly tested by experimental data, often to a very high precision. However, the SM is known to be incomplete, primarily due to evidence of as-yet unexplained phenomena. Notably, it does not incorporate gravitational effects, it does not account for the existence of dark matter, nor for the matter anti-matter asymmetry in the Universe. Furthermore, the requirement of many, seemingly arbitrary, numerical constants and the contrasting mass scales of different particles bring into question whether the SM is indeed a description of a fundamental theory. The aim of high energy physics experiments is to attempt to provide answers to these open questions, ideally through the discovery of new processes that will allow the theory to be expanded. This can be done by directly detecting Beyond the Standard Model (BSM) processes, or by conducting high precision measurements that may be sensitive to New Physics (NP) phenomena.

This thesis examines the flavour sector of the Standard Model via the study of b-hadron decay processes, using data from the Large Hadron Collider beauty (LHCb) experiment. Specifically, the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ is studied, which gives access to two key areas of research where LHCb measurements are expanding the current theoretical understanding. Most notably, a host of measurements of $b \to s\ell^+\ell^-$ transitions have shown tensions with the SM, with evidence for the violation of lepton universality most recently reported in $B^+ \to K^+\ell^+\ell^-$ decays in Ref. [3]. Theoretical uncertainties, as well as the choice of potential NP models, will be impacted by the determination of charm

loop contributions to these processes, accessible through amplitude analyses of $b \to s c \bar{c}$ decays such as $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$. Studies of the amplitude structure of this mode also offer a great opportunity for hadron spectroscopy. A large number of hadrons have been discovered by LHC experiments in their ten years of operation, as listed in Ref. [4]. The list includes many unexpected states, with recent highlights in Refs. [5, 6], which aid the development of theoretical models of strong interactions. These models in turn help to reduce hadronic uncertainties in searches for New Physics.

Precision measurements are performed at LHCb in order to test the robustness of the SM. Planned upgrades to the experiment, with a larger dataset, will provide a significant increase in precision and expand the scope of measurements to ever rarer modes, thereby further improving the sensitivity to potential NP effects. One of the key requirements for precision flavour physics is accurate particle identification (PID), and a substantial improvement to the PID capabilities of LHCb will be provided by the proposed TORCH detector in the next major upgrade. This thesis outlines the design and development of this novel detector, which will extend the momentum range of particle identification available to LHCb using time-of-flight measurements. Studies are performed to improve the timing resolution of TORCH, in order to achieve excellent separation of particles and greatly benefit a host of measurements in the LHCb physics programme.

This thesis is organised as follows. An overview of the Standard Model, with emphasis on the flavour structure, is given in Chapter 2, along with a discussion of its shortcomings. Chapter 3 introduces the LHCb experiment and its detector systems, and Chapter 4 describes the development of the TORCH detector, and calibration studies performed on its readout electronics. The following chapters describe studies of the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ using LHCb data, which form the bulk of the research completed in this thesis. The theoretical and phenomenological motivations for studying $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays are presented in Chapter 5, and some commonly used analysis techniques are introduced in Chapter 6. Chapter 7 details the first analysis of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays, which constitutes the first observation of the decay and a measurement of its branching fraction, as published in Ref. [1]. The subsequent amplitude analysis is presented in Chapter 8 and Chapter 9. The former establishes the formalism used in the amplitude analysis, and the latter demonstrates sensitivity studies performed within the context of the full LHCb data samples used in the analysis. Concluding remarks are given in Chapter 10, highlighting future prospects of the research presented in this thesis.

CHAPTER 2

THEORY

This chapter presents the theoretical background which underpins most of the work in this thesis, namely the Standard Model of particle physics. Particularly, the flavour sector of the model will be discussed, which is the focus of measurements made by the LHCb experiment. Outstanding questions within our current understanding will also be highlighted. The main physics analysis in this thesis is the study of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays, and a more detailed discussion of the phenomenology of these decays is given in Chapter 5.

The discussion in this section is an amalgamation of information from various lecture courses on particle physics and the Standard Model from a flavour perspective, e.g. Refs. [7, 8]. Specific theories and results are cited within.

2.1 The Standard Model

The Standard Model is a quantum field theory which describes the interactions of fundamental particles under the strong, weak and electromagnetic forces. General Relativity is the other main fundamental theory of nature which describes gravitational forces, but will not be discussed here.

Figure 2.1 shows all the fundamental particles described by the Standard Model. They are characterised by their interactions with the fundamental forces and their spin. Spin-1 vector bosons consist of the $W^{\pm}$, Z, and photon which mediate the electroweak force, and the gluon of the strong force. There is one spin-0 scalar boson, the Higgs boson,

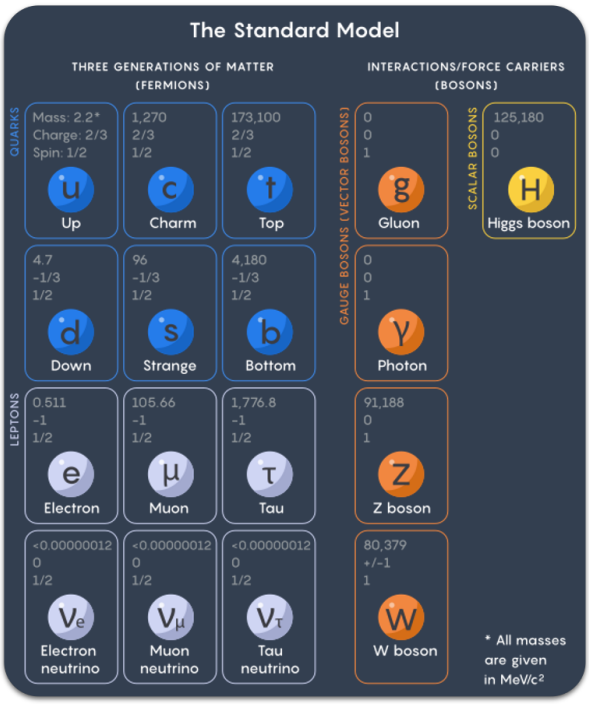


Figure 2.1: Particles of the Standard Model. Figure from Ref [13].

which was the final piece of the model discovered in 2012 [9, 10] after being postulated in 1964 [11, 12] to describe the mechanism by which particles acquire mass. Spin-$\frac{1}{2}$ fermions are split into two families - leptons and quarks. These are conventionally split into three generations, in increasing order by mass. Each fermion also has an anti-particle which is identical, except with its electric charge and other quantum numbers flipped. All fermions carry weak charge so interact via the weak force, and those with electric charge also interact via the electromagnetic force. Quarks have an additional charge, referred to as colour, such that they feel the strong force too. Interactions in the SM are determined by local gauge symmetries under the group

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y . \tag{2.1}$$

The group $\mathrm{SU}(3)_C$ dictates interactions between gluons and quarks, and results in the conservation of colour charge (C). The theory of the strong interaction is described by quantum chromodynamics (QCD), expressed with eight gluon fields $A_\mu^g (g = 1...8)$. Gluons also have colour charge, and their self-coupling leads to two important effects. Firstly, quarks are always found in bound colourless states, since the energy required to separate them is so large that it is favourable to create another colour-singlet state. This is known as *confinement*, and applies at low energies, or equivalently large distances. *Asymptotic*

freedom on the other hand, applies at high energies where quarks and gluons behave as free particles which may be detected as jets of hadrons. Mesons and baryons are the most commonly observed colourless bound states of quarks, consisting of a quark-antiquark pair or three quarks, respectively. However, states with four or five quarks are expected [14], with notable observations in Refs. [15, 16].

The group $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ dictates interactions of the unified electroweak force, resulting in the conservation of weak hypercharge (Y), and weak isospin which acts only on left-handed fields, L. There are four associate gauge fields, $W^a_\mu\,(a = 1,2,3)$ and B_μ. Details of electroweak unification are described further in Section 2.2.2.

The matter fields are represented by Dirac spinors which can be decomposed into left- and right-handed chiral states, *i.e.*

$$\psi = P_L\psi_L + P_R\psi_R, \tag{2.2}$$

where

$$P_R = \frac{(1+\gamma^5)}{2}\,\text{and}\,P_L = \frac{(1-\gamma^5)}{2}, \tag{2.3}$$

and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ for the Dirac gamma matrices γ^μ. This chiral representation is chosen since the weak force interacts with chiral particle states. In the massless limit, chirality is equivalent to helicity, defined by the projection of a particle's spin onto its direction of motion. Since the chiral states behave differently under gauge transformations, they are expressed separately as three left-handed doublets and three right-handed singlets for each generation. For leptons,

$$\begin{aligned} \ell^i_L &= \begin{pmatrix} \nu^i_L \\ e^i_L \end{pmatrix} \\ \ell^i_R &\in \{e_R, \mu_R, \tau_R\}, \end{aligned} \tag{2.4}$$

where right-handed neutrinos are not included in the Standard Model, and neutrinos are treated as massless. The index $i = 1,2,3$ represents the three generations. Similarly for quarks, there are left-handed doublets consisting of up-type and down-type quarks for each flavour

$$q^i_L = \begin{pmatrix} u^i_L \\ d^i_L \end{pmatrix}, \tag{2.5}$$

and right-handed singlets of both types

$$\begin{aligned} u^i_R &\in \{u_R, c_R, t_R\} \\ d^i_R &\in \{d_R, s_R, b_R\}. \end{aligned} \tag{2.6}$$

2.2 The SM Lagrangian

The Lagrangian density of the Standard Model is a mathematical representation of the observed fields and the symmetries under the group $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. It may be expressed as a sum of three terms,

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{Kinetic}} + \mathcal{L}_{\mathrm{Gauge}} + \mathcal{L}_{\mathrm{Higgs}}. \tag{2.7}$$

The first term describes the kinetic energy of the fields ψ, given by

$$\mathcal{L}_{\mathrm{Kinetic}} = i\overline{\psi}\gamma^{\mu}\mathcal{D}_{\mu}\psi. \tag{2.8}$$

The term $\mathcal{D}_{\mu}$ is the covariant derivative defined as

$$\mathcal{D}_{\mu} = \partial_{\mu} + igA^{g}_{\mu}T^{g} + ig_1 B_{\mu}Y + ig_2 W^{a}_{\mu}\tau^{a}, \tag{2.9}$$

where g, g_1 and g_2 are coupling constants. T^g, τ^a and Y are the generators of $\mathrm{SU}(3)_C$, $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$, respectively. Thus the first term of Equation 2.8 is the purely kinetic term, and the rest describe the interactions of the fermions to the $W^{\pm}$, Z and γ.

The gauge fields are described by

$$\mathcal{L}_{\mathrm{Gauge}} = -\frac{1}{2}\mathrm{Tr}F_{\mu\nu}F_{\mu\nu} = -\frac{1}{4}F^{a}_{\mu\nu}F^{a}_{\mu\nu}, \tag{2.10}$$

where $F^{a}_{\mu\nu}$ is the gauge field strength tensor, such that this term describes the kinetic energies and self-interactions of the gauge fields.

2.2.1 The Higgs mechanism

Only the gluon and photon are expected to be massless in the SM, as confirmed by experiment, as including a mass term for the fermions and vector boson fields does not leave the Lagrangian density invariant. This introduces the need for the Higgs field as proposed by Brout, Englert and Higgs in Refs. [11, 12]. They proposed that mass terms could be included by spontaneously breaking the gauge symmetry through the inclusion of scalar fields with a non-zero vacuum expectation value (vev). This is achieved through the introduction of a complex scalar field

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}, \tag{2.11}$$

and a related term in the Lagrangian,

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}^{\mu}\phi)^{\dagger}(\mathcal{D}_{\mu}\phi) - V(\phi) + \mathcal{L}_{\text{Yukawa}}. \tag{2.12}$$

The first term is the kinetic term, the second is a potential term, and $\mathcal{L}_{\text{Yukawa}}$ describes the interactions with fermions. The scalar potential takes the form

$$V(\phi) = \lambda^2(\phi^{\dagger}\phi)^2 - \mu^2\phi^{\dagger}\phi, \tag{2.13}$$

where μ is related to the scalar mass and λ represents its self-coupling. The minimum of this potential for $\mu^2 < 0$ results in a non-zero vacuum expectation value, ϕ_0 given by

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \tag{2.14}$$

Considering fluctuations around the vev as

$$\phi = \begin{pmatrix} h_+ + i\chi_+ \\ \frac{v}{\sqrt{2}} + h_0 + i\chi_0 \end{pmatrix}, \tag{2.15}$$

where $h_{+,0}$, $\chi_{+,0}$ are real scalars, results in only one mass term upon expansion of the scalar potential, given by

$$2\lambda v^2 h_0^2 \equiv m_H^2 h_0^2, \tag{2.16}$$

where h_0 is identified as the Higgs boson with mass

$$m_H = \sqrt{2\lambda}v = \sqrt{2}|\mu|. \tag{2.17}$$

Experimentally, m_H has been determined to be 125.38 ± 0.14 GeV/c^2 [17]. The remaining three fields represent massless Goldstone bosons, named after Goldstone's theorem which states that for n scalar fields, a broken symmetry will result in $n-1$ massless fields. These are interpreted as the longitudinal part of the massive gauge bosons, as explained below.

2.2.2 Electroweak unification

Since flavour-changing interactions are at the heart of decays such as $B^0 \to D^0\overline{D}^0K^+\pi^-$, and these are only possible through the weak force, it is worth briefly discussing the concept of electroweak unification here. A key quantity is weak isospin I, and the third component labelled $I_3 = \pm\frac{1}{2}$ for fermions. There may be three types of transitions between

fermions, resulting in three bosons W^+ ($I_3 = +\frac{1}{2} \to -\frac{1}{2}$), W^- ($I_3 = -\frac{1}{2} \to +\frac{1}{2}$), and W^0 (no change in I_3). Together with the hypercharge generator Y of the $\mathrm{U}(1)_Y$ group, the electric charge is defined by

$$Q = I_3 + \frac{1}{2}Y. \tag{2.18}$$

The model for unification, developed by Glashow, Weinberg and Salam, is based on the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ group, where SU(2) acts on the weak isospin and U(1) acts on the weak hypercharge. Starting with massless gauge fields $W^1_\mu, W^2_\mu, W^3_\mu$ and B_μ, the physical bosons, *i.e.* the ones that are detected, are expressed as superpositions of these gauge fields. The charged vector bosons $W^\pm$ are given by

$$W^\pm{}_\mu = \frac{W^1_\mu \mp i W^2_\mu}{\sqrt{2}}. \tag{2.19}$$

The neutral Z boson and photon are given by linear combinations of W^3_μ and B_μ,

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}, \tag{2.20}$$

where θ_W is called the weak mixing angle. However, the vector bosons are not massless, and they acquire mass through the Higgs mechanism. The introduction of the Higgs field with a non-zero vev breaks the symmetry as $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_{QED}$, giving massive vector bosons $W^\pm$ and Z, and a massless photon, as shown below. It is often said that the Goldstone bosons are "eaten" by the gauge bosons. This refers to the creation of massive vector fields from the gauge fields W^a_μ and B_μ, and the subsequent introduction of an additional degree of freedom in the longitudinal direction, which is given by the Goldstone bosons.

The masses of the gauge bosons are found by considering the covariant derivative of the Higgs field, defined as

$$\mathcal{D}_\mu \phi = (\partial_\mu + i g_2 W^a_\mu \frac{\sigma^a}{2} + i \frac{g_1}{2} B_\mu)\phi, \tag{2.21}$$

where the sigma matrices are related to the generators of SU(2), $\tau^a = \sigma^a/2$, and the hypercharge of the Higgs field has been substituted as $Y = \frac{1}{2}$. The coupling constants are related to the weak mixing angle as

$$\begin{aligned} \sin\theta_W &= \frac{g_1}{\sqrt{g_1^2 + g_2^2}}, \\ \cos\theta_W &= \frac{g_2}{\sqrt{g_1^2 + g_2^2}}. \end{aligned} \tag{2.22}$$

Substituting Equation 2.19 into Equation 2.21, for the Higgs vev ϕ_0, allows the derivation of the $W^{\pm}$ boson masses as

$$m_W = \frac{g_2 v}{2}. \tag{2.23}$$

Similarly the mass of the Z can be extracted and is found to be related to m_W as

$$m_Z = \frac{m_W}{\cos\theta_W}. \tag{2.24}$$

The coupling of A_μ to ϕ_0 is zero, representing an unbroken generator, and revealing that A_μ is massless and consistent with the photon described by QED. The various couplings and mass parameters defined here are consistently measured through experiment, with the world average measurements of the boson masses given by $m_W = 80.379 \pm 0.012\,\text{GeV}/c^2$ and $m_Z = 91.1876 \pm 0.0021\,\text{GeV}/c^2$ [2].

2.2.3 Yukawa interactions

The Yukawa term in the Lagrangian density describes the interactions between the fermion Dirac fields and the Higgs scalar field. It can be split into the lepton and quark contributions, with the lepton term given by

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -y^{\ell}_{ij}\overline{\ell^i_L}\phi\ell^j_R + h.c., \tag{2.25}$$

where y^{ℓ}_{ij} are the lepton couplings, $\ell^i_{L,R}$ are the chiral states defined in Equation 2.4, and $h.c.$ denotes the Hermitian conjugate. Inserting the Higgs vev $\langle\phi\rangle$ into this returns terms of the form

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -\frac{1}{\sqrt{2}}y^{\ell}_{ij}v(\overline{e^i_L}\ell^j_R + \overline{\ell^i_R}e^j_L) + \ldots, \tag{2.26}$$

which are the Dirac mass terms. Similarly, the quark term is given by

$$\mathcal{L}^{q}_{\text{Yukawa}} = -y^{d}_{ij}\overline{q^i_L}\phi d^j_R - y^{u}_{ij}\overline{q^i_L}\tilde{\phi}u^j_R + h.c., \tag{2.27}$$

where $\tilde{\phi} \equiv i\sigma_2\phi^*$. Inserting the Higgs vev here returns terms as

$$\mathcal{L}^{q}_{\text{Yukawa}} = -\frac{y^{d}_{ij}v}{\sqrt{2}}(\overline{d^i_L}d^j_R - \overline{d^i_R}d^j_L) - \frac{y^{u}_{ij}v}{\sqrt{2}}(\overline{u^i_L}u^j_R - \overline{u^i_R}u^j_L) + \ldots \tag{2.28}$$

These terms are expressed in the flavour basis, where the couplings $y^{\ell,u,d}_{ij}$ are elements of the corresponding coupling matrices $Y^{\ell,u,d}$. These matrices can instead be

diagonalised by a unitary transformation $Y \to VYV^\dagger$. The transformation can be chosen to project the fields into the mass basis, *i.e.*.

$$Y^\ell_{diag} = V^{\ell\dagger} Y^\ell V^\ell = \frac{\sqrt{2}}{v} \begin{pmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{pmatrix} \tag{2.29}$$

$$Y^d_{diag} = V^{d\dagger} Y^d V^d = \frac{\sqrt{2}}{v} \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix}, \tag{2.30}$$

$$Y^u_{diag} = V^{u\dagger} Y^u V^u = \frac{\sqrt{2}}{v} \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix}. \tag{2.31}$$

Since $V^d \neq V^u$, one can define the Cabibbo–Kobayashi–Maskawa (CKM) matrix [18, 19] as

$$V_{\text{CKM}} = V^{u\dagger} V^d. \tag{2.32}$$

This matrix describes the interactions between the physical mass states and the weak interactions, for example for the down-type quarks

$$\begin{pmatrix} d^1 \\ d^2 \\ d^3 \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.33}$$

where on the left there are the weak interaction states, and on the right the CKM matrix is multiplied by the mass eigenstates. The matrix is unitary and has a hierarchical structure visualised in Fig. 2.2. There are a few important consequences of the form of the CKM matrix. Firstly, it is not diagonal, thus allowing for transitions between different families of quarks. Secondly, the hierarchy implies that transitions within a family are favoured while transitions between different families are disfavoured. There are several ways of parameterising the CKM matrix, such as the Wolfenstein parameterisation [20] in which the terms are expressed as an expansion of the parameter $\lambda \equiv V_{us} \approx 0.22$, shown in Equation 2.34. There are also two other real parameters, A and ρ, and a complex coupling η which encapsulates the CP-violation in the Standard Model. This is the violation of combined charge-parity symmetry, where a particle is effectively swapped with its antiparticle. Recent measurements of the various CKM parameters can be found

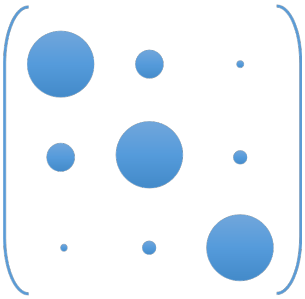


Figure 2.2: Structure of the CKM matrix, where the area of each circle represents the magnitude of the corresponding element.

in Ref. [21].

$$V_{\text{CKM}} = \begin{pmatrix} 1-\frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho-i\eta) \\ -\lambda & 1-\frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \tag{2.34}$$

The transitions between different families, *i.e.* flavour-changing transitions, occur through the exchange of charged bosons. This can be seen by considering the kinetic term in Equation 2.8 and the part of the covariant derivative (Equation 2.9) concerning the charged bosons, $W_\mu^\pm$. In contrast, the part concerning neutral bosons does not contain terms mixing different families (see *e.g.* Ref. [22] for a detailed derivation). Therefore, flavour-changing neutral currents (FCNCs) are forbidden at tree-level in the SM, and they must instead proceed via loop-level processes, which are further suppressed via the GIM mechanism [23]. This makes them an ideal place to look for potential New Physics contributions, as new types of couplings, for example, may contribute at a rate similar to that of the suppressed SM process. Such investigations are detailed further in Section 5.4.

2.3 Resonant states

This section outlines the basic properties of mesons, based on the quark model developed by Godfrey and Isgur [24]. The definitions and nomenclature given here are needed to describe the resonant states under study in the amplitude analysis in Chapter 8 and Chapter 9. While the focus is on mesons here, the description applies to all intermediate hadronic resonances which decay to quasi-stable particles. Contributions to the decays considered in this thesis include both light and heavy states, containing u, d, s and c

quarks. There are several key quantum numbers which define each state, determined by the constituent quark anti-quark pair, as listed below.

1. The spin S is 0 or 1 for anti-parallel or parallel quark spins, by considering two $S = \frac{1}{2}$ vectors being unaligned or aligned.

2. The orbital angular momentum L, due to the constituent quarks orbiting each other, determines the intrinsic parity of the state as $P = (-1)^{L+1}$. Together with the total spin, it also defines the charge parity for neutral mesons as $C = (-1)^{L+S}$.

3. The total angular momentum, J, can take values between $|L-S|$ and $|L+S|$. States are considered to have *natural (spin-)parity* when $P = (-1)^J$, *i.e.* when $S = 1$ and $CP = +1$. The total angular momentum is often expressed in terms of the SPDF notation of electron orbitals, where a $J = 0$ state is called S-wave, a $J = 1$ state is called P-wave, *etc.*

4. The isospin $I = 0, \frac{1}{2}, 1$ for mesons with 0, 1, or 2 constituent u or d quarks.

Resonant states are usually classified using J^P notation, and mesons are often called "spin-J" states, labelled by their total angular momentum rather than the spin quantum number. Quasi-stable particles represent the ground state of a diquark system, *i.e.* a meson with $J = L = S = 0$. Since these have odd parity, with $J^P = 0^-$, they are referred to as pseudoscalar mesons, and include the lightest charged pions and kaons, $\pi^\pm$ and $K^\pm$. In the key mode of this thesis, $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$, all of the initial and final state particles are pseudoscalar mesons. Also of importance to this discussion as highlighted in Section 5.4.1, are vector mesons with $J^P = 1^-$ which have natural parity. These encompass many of the $c\overline{c}$, *i.e. charmonium*, states which decay to $D\overline{D}$, and well known states such as the $K^*(892)^0$.

2.4 Beyond the Standard Model

The Standard Model as described above has been rigorously tested and validated by experimental evidence. However, there are some gaping holes in the theory compared to the Universe we observe. Firstly, the fundamental force of gravity cannot yet be described in a way compatible with quantum field theory. Secondly, there is an abundance of astrophysical evidence for the existence of a type of matter that interacts with gravity but does not radiate electromagnetically or exhibit strong interactions. Although there are searches for dark matter using existing and proposed particle physics experiments,

no particle has been observed with such characteristics. Furthermore, an open question remains as to the mechanism that drives the matter asymmetry in the Universe. As previously mentioned, there is some charge-parity violation entering the SM in the quark sector, however this is insufficient to produce the abundance of matter over anti-matter that was needed to create the Universe we see today, by several orders of magnitude. An aspect where experiment has shown clear evidence of an extension to the Standard Model is in the neutrino sector, where the observation of neutrino mixing necessitates them to have mass. The Yukawa interaction does not result in massive neutrinos, unlike the other fermions, so a different mechanism is needed to explain how they acquire mass. Neutrino oscillation experiments have also shown indications of CP-violation in the lepton sector, although there is no clear evidence this is greater than in the quark sector [25]. Additionally, there is growing evidence from heavy flavour decays for new types of couplings or possible violation of lepton flavour universality. This is discussed further in Section 5.4.

From a theoretical viewpoint, there are problems with the framework of the Standard Model brought about by ad hoc features, motivating the search for a deeper understanding. For example, the hierarchy problem refers to the large difference in masses of different particles, such as the up quark compared to the top quark. Similarly, the mass of the Higgs is far lower than the Planck scale, which means that a very finely tuned cancellation is needed to counteract its quantum corrections from fermion loops. Proposed theories to account for this generally introduce New Physics to appear at the 1 TeV scale. On the other hand, the flavour structure of the SM, though unnatural, has been measured very precisely, and applies constraints on any new particles at $> 10^4$ TeV. Therefore a TeV-scale solution would not be consistent with current experimental limits, and this is referred to as the "New Physics flavour problem" [26]. These considerations mean that the Standard Model as we know it can be considered an effective theory, *i.e.* one which is applicable at a certain limit, given by the energy scales we have probed so far [27]. Searches for Beyond the Standard Model physics aim to find the truer, fundamental theory that will solve these problems.

CHAPTER 3

THE LHCB EXPERIMENT AT THE LHC

The LHCb experiment is one of the four major particle physics experiments located at the Large Hadron Collider (LHC), at the Centre for European Nuclear Research (CERN). It is designed to study heavy flavour physics, *i.e.* processes involving beauty and charm hadrons. This is made possible by the LHCb Collaboration, which constitutes physicists, engineers, software developers and other experts from around the world who collect and analyse data from the LHC. Section 3.1 describes the facilities at the LHC, and Section 3.2 describes the LHCb experiment in further detail. The following sections outline the detector subsystems that make up LHCb, as well as the software framework used to simulate the detector, and a discussion of the future upgrades to the experiment. *In addition to the references cited in this chapter, further details of the LHCb detector can be found in Ref. [28].*

3.1 The LHC

The LHC is the largest and most powerful particle accelerator in the world, and is the biggest component of the accelerator complex at CERN. It is a circular collider with a 27 km circumference, located in an underground tunnel that spans the French and Swiss borders. This tunnel was previously used by the Large Electron-Positron Collider.

Inside the accelerator, two proton beams travelling in opposite directions are brought to collision at a target centre-of-mass energy ($\sqrt{s}$) of 14 TeV. The collision points are the locations of particle physics experiments such as LHCb. During first phase of collisions,

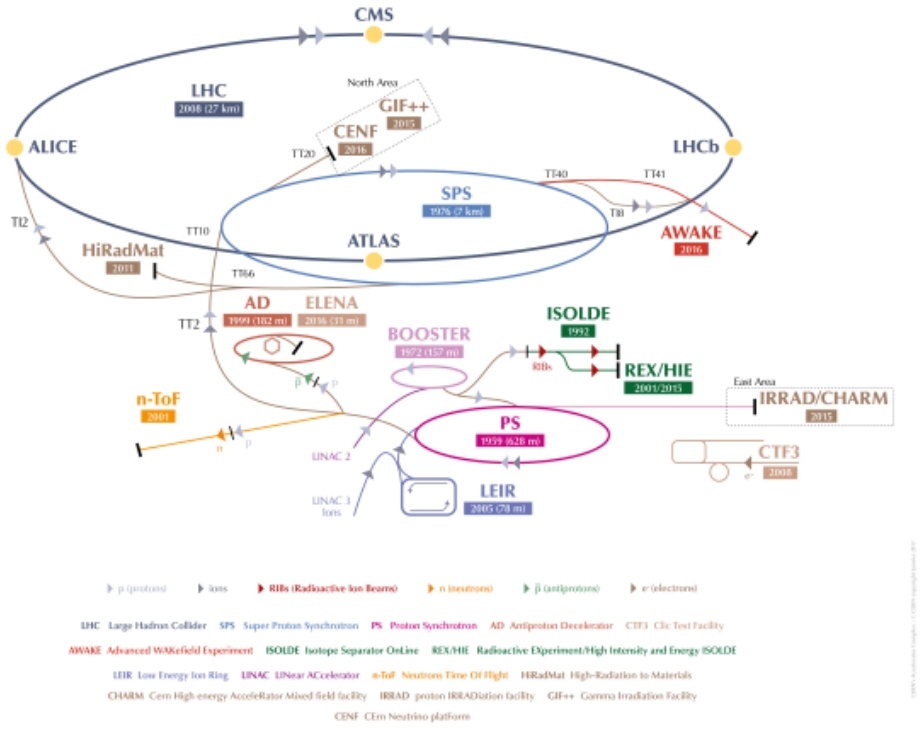


Figure 3.1: The CERN accelerator complex. A series of linear accelerators and synchotrons feed into the final LHC ring in dark blue. Figure from Ref. [29].

referred to as Run 1, the beams were collided at reduced energies of 7 TeV in 2011 and 8 TeV in 2012. The following phase ran from 2015-2018 at $\sqrt{s}$ =13 TeV. A number of smaller machines are used in order to boost the energy of the protons before entering the LHC ring. Firstly, protons, sourced from a bottle of hydrogen, are accelerated to 50 MeV in a linear accelerator known as Linac 2. These are passed to the Booster, which is the smallest synchrotron in the complex, to be accelerated to 1.4 GeV. There are two further synchrotrons of increasing size, the PS and SPS, which accelerate the protons to 25 GeV and 450 GeV respectively. Bunches of these protons are finally delivered to the LHC, where they are accelerated for 20 minutes up to the target energy of 6.5 TeV in each direction. Collisions can then take place for several hours, after which the beam is directed out of the main ring and into the beam dump facility. The full accelerator complex at CERN is shown in Figure 3.1. The locations of LHCb and the other main experiments on the LHC ring are displayed, along with numerous other experiments whose beams are provided by the smaller machines.

3.2 The LHCb detector

The Large Hadron Collider Beauty experiment is optimised to study interactions of b- and c-hadrons. The aim is to build upon previous measurements by the B Factories and the Tevatron, with higher statistics and improved precision, to test the Standard Model and probe New Physics phenomena. Unlike the general purpose detectors which surround the pp interaction point, LHCb is often referred to as a single-arm, forward spectrometer. This is because it is oriented in the forward direction in order exploit the maximum production of $b\overline{b}$ pairs, as demonstrated in Figure 3.2. The acceptance of the detector, given by the azimuthal beam angle, is $10 < \theta < 250\,\text{mrad}$, or $2 < \eta < 5$ where the pseudorapidity η is defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{3.1}$$

Additionally, proton beams from the LHC are tuned such that a uniform luminosity is delivered to LHCb. This allows cleaner signals to be seen and reduces radiation damage to the detector components. The tuning process involves a set of corrector magnets which adjust the beam overlap at the LHCb interaction point in real-time.

The layout of the detector is shown in Figure 3.3 in right-handed coordinates (as the orientation in Figure 3.2), where the z-axis is along the beam. Polar coordinates (r,θ,ϕ) where the range of θ indicates the angular acceptance may also be used. The pp interaction point is at the origin of the coordinate system, and the various subdetectors which make up the experiment are labelled. The main components are as follows. The interaction point is surround by the VErtex LOcator (VELO) which provides high resolution identification of vertices, described in more detail in Section 3.3. Next are the first of two Ring Imaging CHerenkov (RICH) modules, RICH1, and an upstream tracking station, the Tracker Turicensis (TT). Then there is the dipole magnet which is used to measure the momentum of charged particles by their track curvature. Directly downstream of the magnet reside three tracking stations T1-T3, which can be reached by particles with a minimum momentum of $1.5\,\text{GeV}/c$. Next is RICH2, and further details of the particle identification performed by the RICH detectors is found in Section 3.4. Following this are the muon systems and the calorimetry systems. The muon stations M1-M5 provide identification of muons and feed into the hardware level trigger. The LHCb trigger system reduces the event rate to a level that can be readout and processed; this is discussed further in Section 3.5. The calorimeter system consists of a Scintillating Pad Detector (SPD), a Preshower (PS), an electromagnetic calorimeter (ECAL), and a hadronic calorimeter (HCAL). It allows for the measurement of energies and positions of

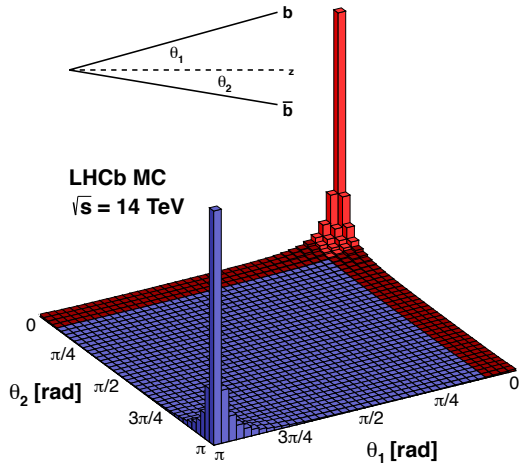


Figure 3.2: The production of $b\overline{b}$ quark pairs as a function of opening angle from LHCb simulations at $\sqrt{s}$ =14 TeV. θ_1 is the opening angle of the b and θ_2 is the opening angle of the $\overline{b}$. The red area indicates the acceptance of the LHCb detector. Figure from Ref. [30].

electrons, photons and hadrons. The algorithms used to simulate events in the LHCb detector are described in Section 3.6.

The main analysis in this thesis concerns decays of $B^0 \to D^0\overline{D}^0K^+\pi^-$. As particles traverse the detector, they deposit energy, and these 'hits' can be fitted to construct composite track objects containing tracking and vertexing information. In a given collision event, one or more $B^0 \to D^0\overline{D}^0K^+\pi^-$ candidates may be identified as composite track objects with the correct mass and particle identification hypotheses. In some cases, the candidate will not be a genuine $B^0 \to D^0\overline{D}^0K^+\pi^-$ decay, but some background process. A simple tracking and reconstruction is performed initially, followed by a more sophisticated process which is done offline on stored data.

A future upgrade to the LHCb detector is underway since the end of 2018, to prepare for the next phase of data-taking in Run 3. This will involve major changes to the subdetector systems, and is described in Section 3.7.

3.3 The tracking systems

Precise measurements of the decays of b- and c-hadrons are made possible by high resolution tracking in the LHCb detector.

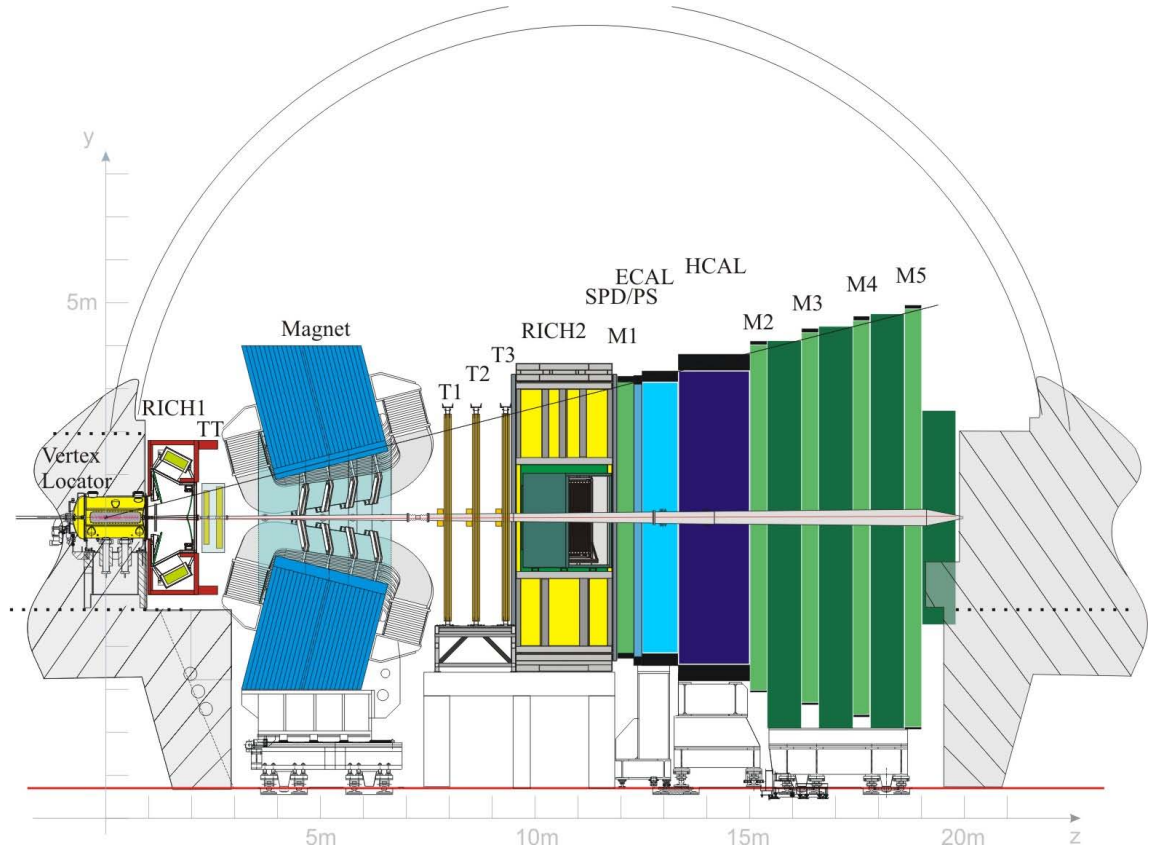


Figure 3.3: Cross-section of the LHCb detector. The main subdetector systems are indicated. Figure from Ref. [28].

3.3.1 The VELO

The innermost tracking system is the VErtex LOcator (VELO) which surrounds the pp interaction point, such that the primary vertex (PV) can be accurately located. The distinctive lifetime of b-hadron decays means they travel on the order of millimetres before decaying at the secondary vertex (SV). Thus the VELO can resolve the flight distance (FD) of the b-hadron to this displaced SV. Therefore the closest distance between the PV and the direction vector of the b-hadron, known as the impact parameter (IP), can also be measured. An example decay topology with the kinematic variables is shown in Figure 3.4.

The VELO is made up of 42 silicon modules, each of which provides a measurement in the r and ϕ directions along the beam, using strip sensors known as R and Φ sensors. These sensors are positioned 7 mm from the LHC beam, and therefore they are required to be retractable during beam injection. This is achieved by arranging the modules in two halves that are automatically closed and aligned during operation. Routing lines across

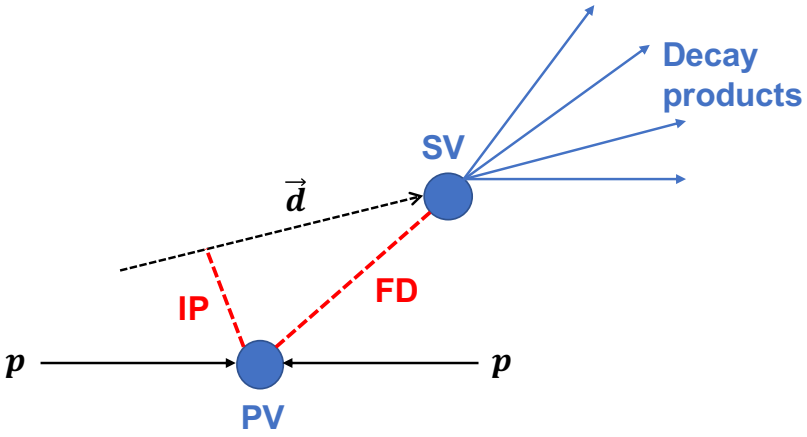


Figure 3.4: Schematic topology of a decay of a b-hadron from the pp collision point. The direction vector of the decaying particle is given by $\vec{d} = \vec{p}/|\vec{p}|$.

the sensor allow readout of the strips to chips around the circumference. An illustration of the sensors is shown in Figure 3.5.

The aim of the VELO is to provide accurate vertex reconstruction and precise impact parameter resolution. The PV resolution is correlated to the number of tracks in the event. For each event, a subsample of tracks can be used to create a PV, and these positions can be compared to find the resolution. Carrying out this procedure on 2011 data, it was concluded that a 25-track vertex has a transverse resolution (in x, y) of 13 μm, and a resolution in the z-direction of 71 μm. The impact parameter resolution is inversely related to the particle's momentum transverse to the beam direction, p_{T}. It can be calculated by measuring the width of the IP distributions in the x and y directions. Using data and simulation from 2012, it was determined that particles with $p_{\mathrm{T}} > 1\,\mathrm{GeV}/c$ achieve an IP resolution better than 35 μm. Overall the VELO has a track reconstruction efficiency of over 98%. [31]

3.3.2 The magnet and tracking stations

The dipole magnet in LHCb has an integrated magnetic field of 4 Tm, and was designed to provide a high field density between the VELO and the tracking stations, *i.e.* for tracks over a length of 10 m [28]. The curvature of charged particle tracks passing the magnet provides information on their momenta. During data-taking, the direction of the field is periodically flipped so that potential asymmetrical effects in the detector are reduced.

The silicon tracker encompasses both the TT and the inner component of the tracking stations T1-T3, known as the Inner Tracker (IT). The sensors used are silicon microstrips

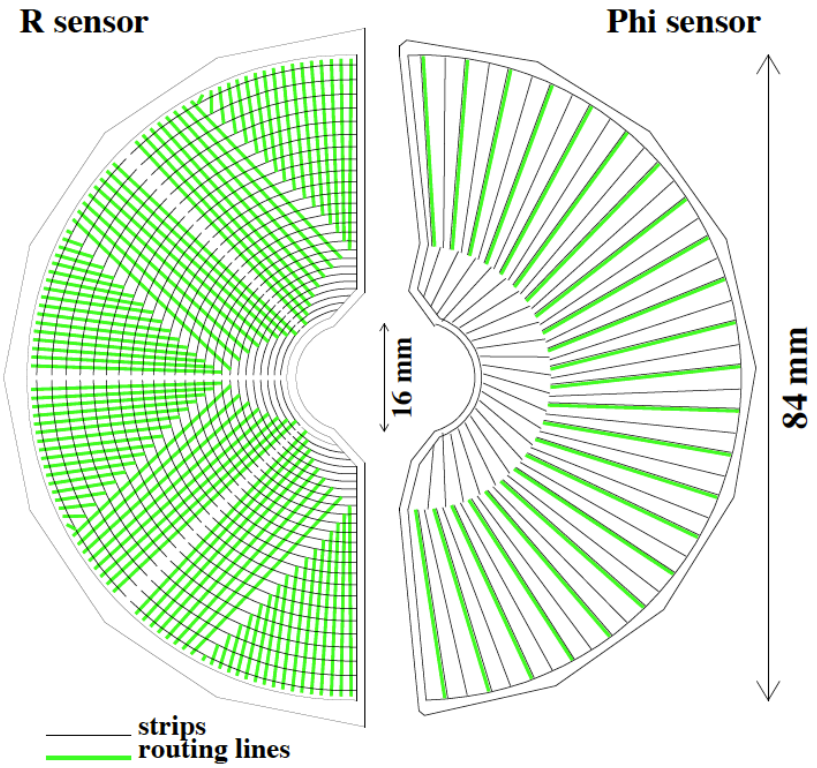


Figure 3.5: Schematic representation of a VELO module, showing the R and Φ sensors. Figure from Ref. [31].

with a separation, or pitch, less than 200 μm, to provide high spatial resolution. The resolution of a single hit was determined to be between 50 μm and 55 μm using 2011 and 2012 data. The outer component of the tracking stations is known as the Outer Tracker (OT), which is a drift-tube detector with a counting gas of Ar, CO_2 and O_2. The OT uses straw-tubes instead of silicon for reduced costs, to measure drift times of charged particles which ionise the gas contained in the ~ 200 modules. The single hit resolution for the OT is close to 200 μm. The muon stations also provide tracking, which is used as input to the trigger decisions used in this thesis. Muon candidates are built by aligning hits found in each of the muon stations M1-M5. These hits are detected using multi-wire proportional chambers in each station, and gas-electron multipliers in the highest occupancy region of M1. [32]

3.3.3 Track reconstruction

Depending on the hits in the VELO, TT, IT and OT, various types of tracks may be defined, as illustrated in Figure 3.6.

- Long tracks have hits in both the VELO and the tracking stations T1-T3. Thus long tracks are associated to particles that have traversed the full detector, and are the most useful for physics analyses.

- Upstream tracks have hits in only the VELO and TT, implying they originate from a low momentum particle that does not travel past the magnet.

- Downstream tracks have hits in only the TT and T stations, and so are usually associated with long-lived particles which decay downstream of the VELO.

- VELO tracks have hits in the VELO only, typically due to having a large angle or pointing backwards. These are nevertheless used for the reconstruction of the PV.

- T tracks have hits in the T stations T1-T3 only. These are associated with secondary interactions of particles.

Tracking algorithms are used to reconstruct all possible tracks in an event based on the hits in the detector. Each track definition has specific requirements on the hits in the relevant stations and the extrapolated trajectories thereafter. The tracks are then fitted with a Kalman filter [33] which takes into account scattering and energy losses. The quality of the reconstructed tracks are then given by the χ^2/ndf of this fit. The track reconstruction efficiency is greater than 96% for $5\,\text{GeV}/c < p < 200\,\text{GeV}/c$, and slightly less in events where the number of tracks exceeds 200. A measurement of the momentum of each track is then made, aided by the dipole magnet. The momentum resolution ($\delta p/p$) can also be calculated using decays of $J/\psi \to \mu^+\mu^-$, illustrated in Figure 3.7. This decay is chosen because the muons traverse the full detector, and can have a wide range of momenta. A relative momentum resolution of 0.5% at low momentum, and 1.0% at $200\,\text{GeV}/c$ is achieved. [32]

3.4 Particle Identification

The calorimeter, RICH detectors and muon stations provide particle identification in LHCb. Accurate determination of the identities of different tracks is a vital requirement in achieving the goals of LHCb analyses. In this analysis, the predominant focus is on the identification of hadrons, specifically kaons and pions.

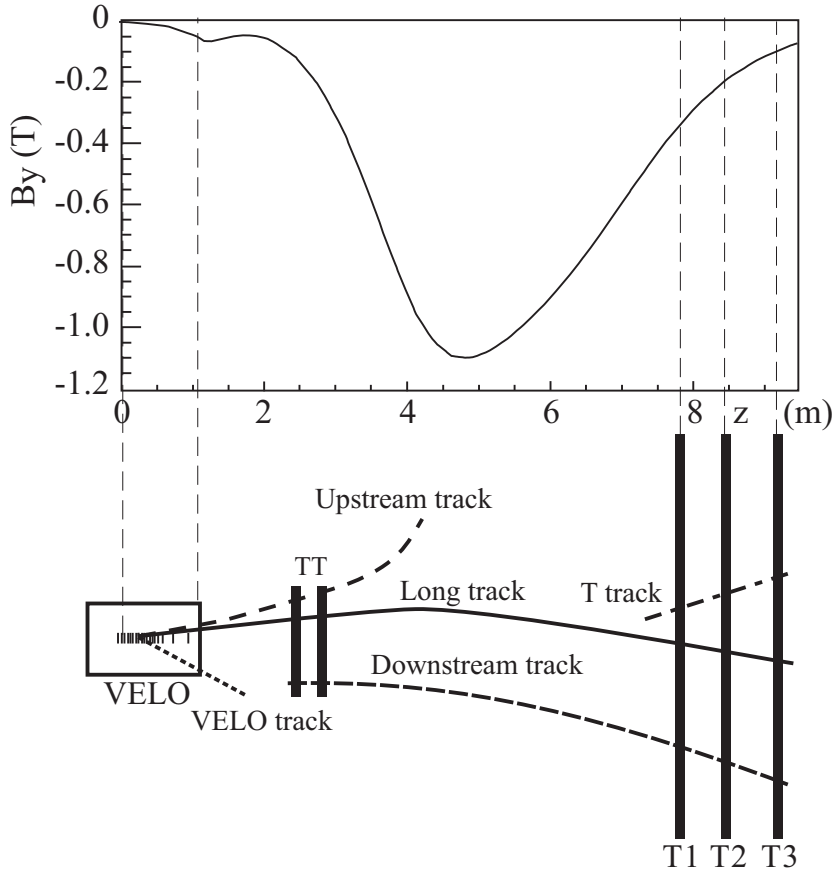


Figure 3.6: Illustration of track types in LHCb. The main magnetic field component, B_y, along the beam axis is plotted above for reference. Figure from Ref. [32].

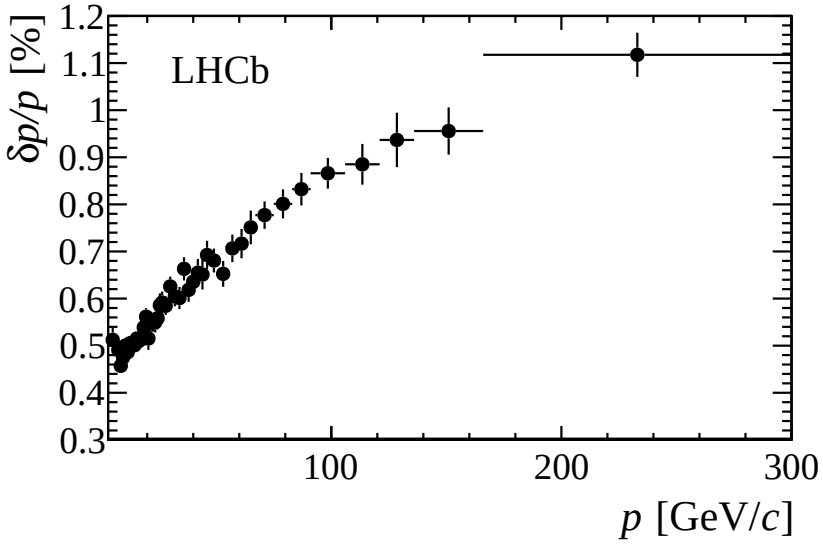


Figure 3.7: Momentum resolution against momentum obtained from reconstructed J/ψ candidates. Figure from Ref. [32].

3.4.1 The calorimeters

The calorimeter system is mainly used to identify photons, electrons and neutral pions. Both the ECAL and HCAL are scintillating sampling detectors made from lead and iron, respectively. The ECAL, which deals with electrons and photons only, requires clusters of energy deposits used in conjunction with information from the SPD and PS. Neutral pions are also reconstructed by looking for pairs of photons in the ECAL. For the identification of electrons, information from the HCAL is also used, to build a likelihood for the electron hypothesis based on the expected energy and momentum distributions. A neutral particle is distinguished from a charged one by the absence of a track associated to the cluster of hits in the calorimeters.

3.4.2 The RICH detectors

The RICH system aims to identify and separate charged pions, kaons and protons. It also provides information for the identification of charged leptons such as electrons and muons.

Cherenkov radiation is exploited for charged particle identification. When charged particles travel with a speed v which is faster than the speed of light in a given material, they emit Cherenkov radiation. Due to the direction of travel, these photons are produced in a cone, and projected onto a plane such that they are observed as rings. The Cherenkov angle of the photons, θ_C, may be expressed in terms of the momentum (p) of the particle as

$$\cos(\theta_C) = \frac{1}{n\beta} = \frac{\sqrt{m^2c^2 + p^2}}{np}. \tag{3.2}$$

Here n is the refractive index of the material, m is the mass of the particle, and $\beta = v/c$. Therefore the Cherenkov angle directly relates to a particle's momentum.

As mentioned, there are two RICH detectors, each covering a specific momentum range. RICH1, which sits upstream of the magnet, covers particles with momenta in the range 2-40 GeV/c. It covers the full angular acceptance of LHCb, and uses C_4F_{10} radiator gas which has a refractive index of 1.0014 at 400 nm. RICH1 is divided vertically about the beam pipe, and a side view of it is shown in Figure 3.8(a). Since RICH2 is downstream of the magnet it is optimised to measure higher momentum tracks, from 15-100 GeV/c, as these are less affected by the bending from the magnet. It also has a smaller angular acceptance, of 15-120 mrad, and uses a CF_4 radiator which has n = 1.0005 at 400 nm.

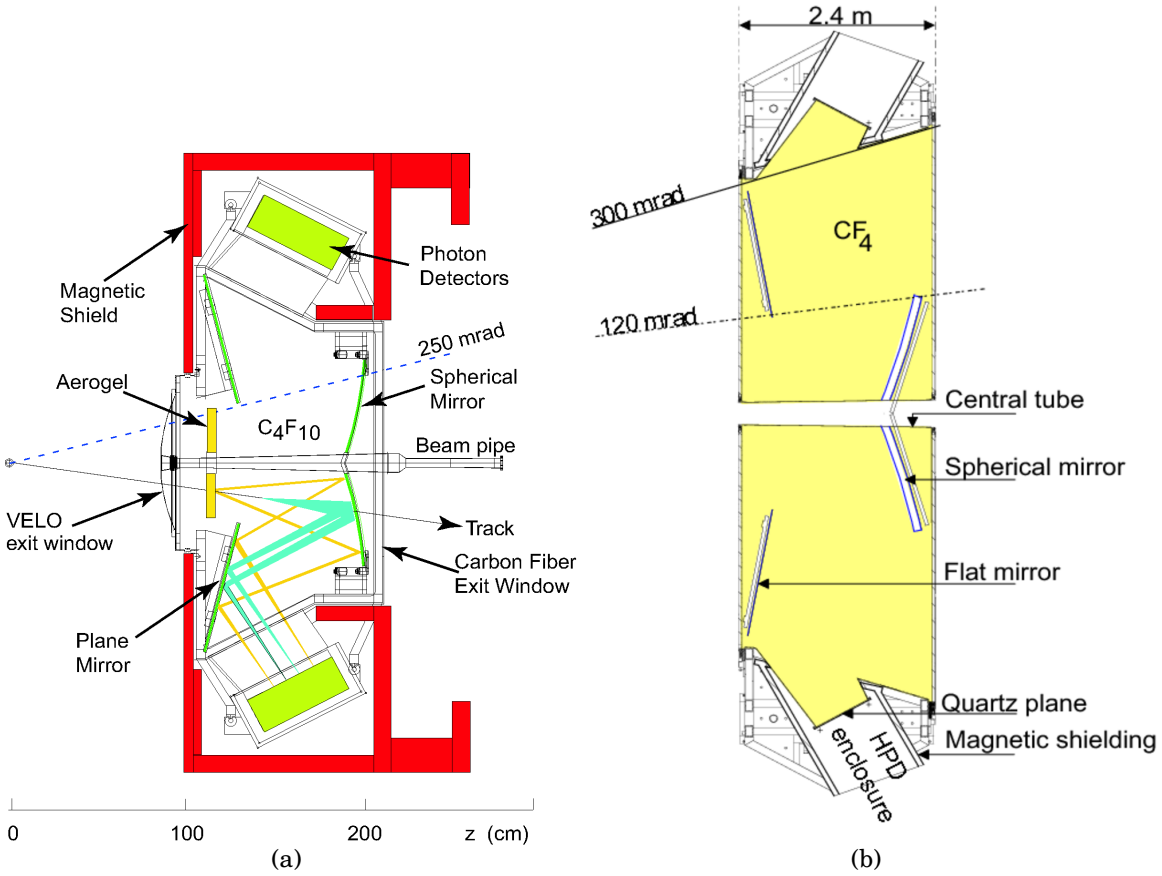


Figure 3.8: Schematics of the RICH detectors. a) RICH1 from a side view and b) RICH2 from a top view. Figures from Ref. [28].

RICH2 is divided horizontally about the beam axis, and a view of it from the top is shown in Figure 3.8(b). [34]

A specially designed optical system allows the RICH detectors to measure the Cherenkov angles produced by traversing particles. As illustrated in Figure 3.8(a), the cone of light emitted by a track is firstly focused on to a spherical mirror, and then reflected onto a plane mirror. This is then reflected onto a hybrid photon detector (HPD) specifically designed for the RICH system. From these HPD hits, a Cherenkov ring can be reconstructed using a pattern recognition algorithm described in Ref. [35]. These photon candidates can then be associated with a candidate track using information from the tracking systems in LHCb, and finally the Cherenkov angle may be computed. The resolution on θ_C is found to be 1.62 mrad and 0.68 mrad for RICH1 and RICH2, respectively, using data from 2011. [34]

3.4.3 PID strategy and performance

Using Equation 3.2, it is clear to see that combining the measured Cherenkov angle and the reconstructed particle momentum allows one to test the mass hypothesis of the particle. This is illustrated in Figure 3.9(a), where clear separation between particles is seen. This idea is used to compute the difference in log-likelihoods (DLL) of two particle hypotheses. For example, to compare a kaon and pion hypothesis, $\mathrm{DLL}_{K\pi}$ is computed, defined as

$$\mathrm{DLL}_{K\pi} = \log\mathcal{L}(\theta_C, p|K) - \log\mathcal{L}(\theta_C, p|\pi). \tag{3.3}$$

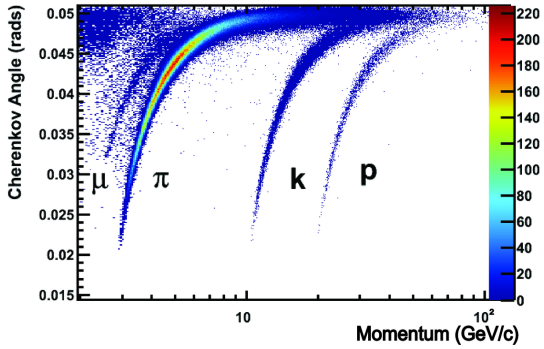
An analogous quantity can also be computed for protons, $\mathrm{DLL}_{p\pi}$. The likelihoods are calculated from a combination of information from the RICH, the calorimeters and the muon systems. For muon identification, the DLL has an additional dependence on the hits in the muon stations, and can be similarly expressed as

$$\mathrm{DLL}_{\mu\pi} = \log\mathcal{L}\left(\theta_C, p, D^2|\mu\right) - \log\mathcal{L}\left(\theta_C, p, D^2|\pi\right), \tag{3.4}$$

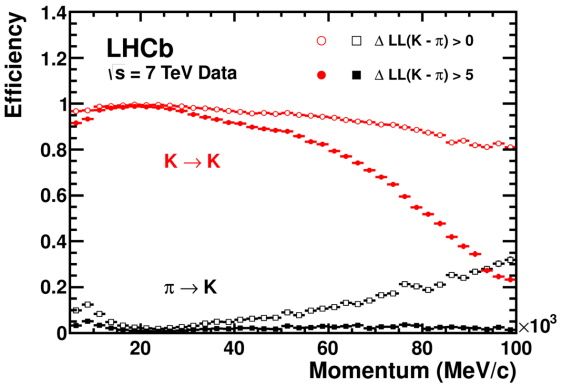
where D^2 is the average squared distance significance of the hits in the muon stations with respect to the track found by extrapolation from the tracking system [36].

The performance of K-π separation is displayed in Figure 3.9(b). For a requirement of $\mathrm{DLL}_{K\pi} > 0$, the kaon efficiency is around 95%, and the pion misidentification rate is around 10%. The kaon efficiency worsens under stricter requirements on the PID variable and at higher momenta. However the stricter requirement vastly reduces the rate of pions misidentified as kaons [34]. This performance is equally as good, and in some cases better, in Run 2 [37].

A multivariate particle identification variable is also used in LHCb, based on the response of an artificial neural network. Since this response is normalised between 0 and 1, it is called ProbNN. A wide range of input variables are used to train the neural network from all major subdetector components, in an effort to create a powerful variable that takes into account correlations between the subdetectors and all the DLL variables. The training samples are based on simulation which has been resampled to match the event occupancy in data. Separate networks are trained for pions, kaons, electrons, muons, protons, and 'ghost' or fake tracks, referred to as ProbNNπ, *etc.* More information on the use of these variables can be found in Ref. [38].



(a) Cherenkov angle distribution shown for isolated rings that can cleanly be associated to a specific track. Clear separation between muons, pions, kaons and protons is visible.



(b) The empty and filled markers represent two different requirements imposed on $\mathrm{DLL}_{K\pi}$.

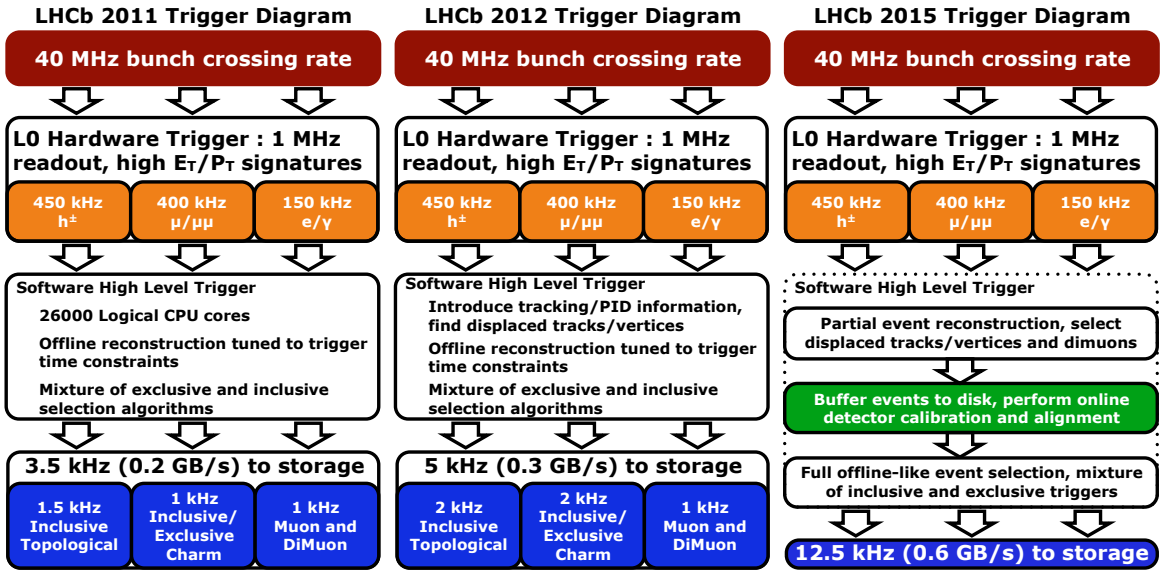
Figure 3.9: Left shows the reconstructed Cherenkov angle as a function of the track momentum for RICH1. Right shows the efficiency of kaon identification (red) and the rate of pion misidentification (black) as a function of track momentum. Figures from Ref. [34].

3.5 The trigger system

Due to the high collision rate at the LHC, reading out and reconstructing every event would be infeasible. This is due to constraints on the electronics of the data acquisition (DAQ) systems, computing time, and storage capacity. The LHCb trigger system reduces this rate by filtering events and preserving a sample that is rich in $b\overline{b}$ events. The trigger system is implemented in two steps, first in hardware where relatively rudimentary decisions may be applied. Events passing the first step are then given to the next step which is implemented in software such that more computationally intensive decisions may be applied. Overall, the trigger system reduces the 40 MHz bunch crossing rate from the LHC to 3.5 kHz in 2011, 5 kHz in 2012, and 12.5 kHz in Run 2 (2015 onwards). The rates increase with time through improvements in the DAQ systems and efficiency of the trigger systems. The rates and further details for each period are shown in Figure 3.10. Further details of the hardware and software level triggers are described below.

3.5.1 Hardware level trigger

The hardware level trigger, referred to as Level 0 (L0), reduces the data rate to 1 MHz which is the limit of the LHCb readout. It is instrumented in a system of field-programmable gate arrays (FPGAs). The L0 trigger relies on information from the calorimeter and the muon stations. Optimised thresholds on the transverse momentum



(a) (b) (c)

Figure 3.10: Trigger schemes for (a) 2011, (b) 2012 and (c) Run 2. Figures from Ref. [39].

are set to select hadrons, muons, electrons, and photons. The thresholds vary for each particle type, as well as changing year-on-year. More information on the L0 implementation and thresholds can be found in Refs. [40–42]. With the rate reduction achieved by the hardware trigger, the full detector response, including particle identification and tracking information, can be read out.

For the principal analysis in this thesis of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays, the relevant L0 triggers are `L0Hadron` and `L0Muon`. When a track which is associated with the signal decay causes the trigger to fire, *i.e.* fulfils the L0 conditions, this is referred to as Trigger-On-Signal, or TOS. Since the signal decay contains only hadrons, the relevant trigger line is named `L0HadronDecision_TOS`. In some cases, an event can fulfil the L0 conditions due to particles that do not constitute the signal decay. This is referred to as Trigger-Independent-of-Signal, or TIS. The trigger decisions used in this case are `L0HadronDecision_TIS` and `L0MuonDecision_TIS`. The muon decision is included as the muon trigger is designed to be highly efficient in LHCb. The TISTOS method [40, 43] may be used to derive the TOS efficiency as

$$\epsilon = \frac{N(\text{TOS}\,\&\,\text{TIS})}{N(\text{TIS})}, \tag{3.5}$$

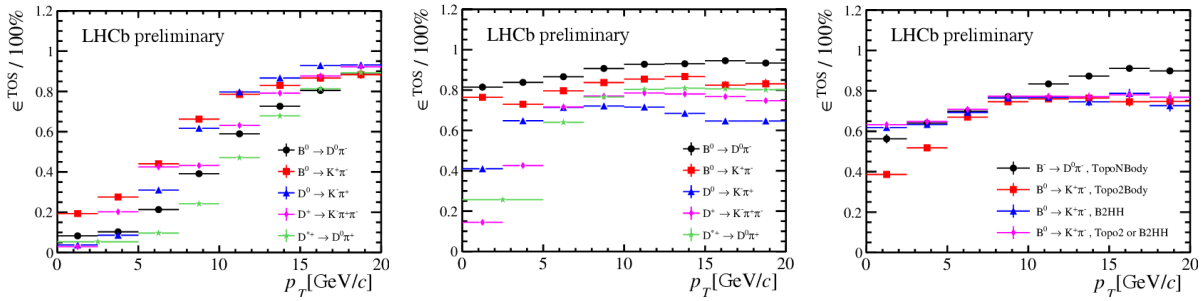
where $N(\text{TIS})$ is the number of TIS events in the sample and $N(\text{TOS}\,\&\,\text{TIS})$ is the number of events which are both TOS and TIS. The L0 hadron trigger performance in 2012, given by the TOS efficiency, is displayed in Figure 3.11(a).

3.5.2 Software level trigger

Events which are accepted by L0 are passed to a computing farm of close to 30,000 CPU cores. Here, the software trigger algorithms, known as the high level trigger (HLT), run additional reconstruction and selection to further reduce the data rate. This is done in two stages, known as HLT1 and HLT2, each with an increasing amount of complexity.

A partial event reconstruction is carried out in HLT1, using information from the VELO and the tracking stations, including the muon system for muon identification. The pattern recognition algorithm aims to select tracks with high momenta which are displaced from the PV, *i.e.* with a high impact parameter. This is done by extrapolating VELO tracks first to the TT and then to the T stations, and then fitting with a Kalman filter. The fitted tracks can then be used to determine the vertex positions. An inclusive beauty and charm trigger line called `Hlt1TrackAllL0` selects good quality tracks with a threshold of $p_{\mathrm{T}} > 1.6\,\mathrm{GeV}/c$ [40, 41]. This is the principal HLT1 trigger line used in this thesis, and its efficiency in 2012 is shown in Figure 3.11(b). In Run 2, there are two inclusive lines used instead, for single tracks and two tracks. These are re-optimisations of the aforementioned selections in Run 1, based on multivariate selection criteria, see Ref. [42] for further information. An event is required to contain at least one track which passes the relevant trigger line, and which is not rejected as fake by the fitting algorithm.

A full event reconstruction can be performed at HLT2, and the outputs based on a mixture of exclusive and inclusive selections, can be written to permanent storage. This is achieved by including all the available detector information, including particle identification from the RICH. Relevant to this thesis are the inclusive trigger lines in HLT2 which are optimised to search for a topology compatible with the decay of a b-hadron. These lines, named `Hlt2Topo(N)Body`, look for two-, three-, or four-track vertices with high transverse momentum and a significant displacement from the PV. A Bonsai-boosted Decision Tree (BBDT) classifier is used to implement the selection in an efficient and robust manner. Details of the algorithm can be found in Ref. [44], and its usage in the HLT in Run 1 in Refs. [40, 41]. The efficiency of inclusive trigger lines for various b-hadron decays, as well as select exclusive decays, are shown in Figure 3.11(c). In Run 2, through the optimisation of the selection algorithms and the exploitation of improved computing resources, the event reconstruction of HLT2 has a much improved performance. This is also aided by the implementation of real-time alignment and calibration of the detector subsystems, outlined in Figure 3.10. This means that events may be used directly from this step, without the need for additional offline reconstruction, in the so-called Turbo stream [42].



(a) L0 hadron trigger performance (b) `Hlt1TrackAllL0` performance (c) HLT2 beauty trigger performance

Figure 3.11: Trigger performance in 2012 at the three levels, (a) L0, (b) HLT1, (c) HLT2. TOS efficiency for various channels are shown as a function of the parent particle's transverse momentum. Figures from Ref. [41].

All candidates that are reconstructed offline, and thus used for physics analysis, are required to have been selected by the trigger. The specific trigger requirements used in this thesis for the selection of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays are summarised in Section 7.3.2.

3.6 The simulation framework

Simulated data is used extensively in LHCb analyses in order to understand detector performance and other experimental conditions, where no data driven alternatives are available. The LHCb simulation framework is based on the GAUSS simulation application which generates particles and simulates the resulting response of the detector [45], using a number of open-source software packages. The pp collisions are generated by PYTHIA [46], producing $b\overline{b}$ pairs quark pairs to be hadronised. The decay of the b-hadron of interest is then described by EVTGEN [47]. The interactions of these particles with the detector are then simulated based on a detailed detector model implemented in the GEANT4 package [48,49]. The digitisation of the subdetector systems is performed by the BOOLE application, so the format imitates the output of the DAQ system in LHCb. At this stage, the simulated data is processed in the same manner as real data, reconstructed by the BRUNEL package and further analysed with DAVINCI. These applications form the core LHCb software, using the LHCb event model and detector description as inputs, and are all based on the GAUDI framework [50,51].

3.7 Upgrades to the LHCb detector

To date the LHCb experiment has enabled a broad range of heavy flavour physics measurements to be performed at world-leading levels of precision. This was made possible by the vast numbers of b-hadron decays detected within the LHCb acceptance in the Run 1 and Run 2 data-taking periods. As the experiment is designed to operate at a constant luminosity which is lower than the luminosity delivered by the LHC, this is the limiting factor for data collection. The LHCb Upgrade [52] aims to modify and improve the detector such that it can operate at a higher luminosity and continue to expand its physics programme. Currently, the LHC and the detector is undergoing a long shutdown where the first upgrade, Upgrade I, will be installed for data collection in Run 3. The goal is to collect up to $50\,\text{fb}^{-1}$ in 5 years of data-taking at a luminosity of $2 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$. Following on from this, there will be an intermediate upgrade, and finally Upgrade II will be installed to operate at ten times higher the luminosity of Upgrade I, in order to exploit the High Luminosity LHC (HL-LHC) era in the next decade. Details of the physics goals of the upgrades can be found in the Letter of Intent for Upgrade I [53] and the Physics Case for Upgrade II [54], respectively.

Upgrade I involves updating many of the detector subsystems as well as, most notably, switching to a fully software trigger system. The tracking system will be fully updated, with the new VELO being instrumented with hybrid pixel detectors and sitting even closer to the beam pipe, at a distance of 5.1 mm. The TT will also be replaced with an Upstream Tracker (UT) which has a finer granularity to account for the increase in particle density. Additionally, the inner and outer trackers, IT and OT, will be replaced by a single scintillating fiber tracker known as the SciFi. Concerning the particle identification systems, RICH1 will be upgraded with improved mirrors and new electronics to increase the read-out rate. Similarly, the readout electronics of the calorimeter and muon systems will be upgraded. These upgrades will allow for a trigger-less readout system, thus allowing for the removal of the L0 hardware trigger bottleneck. Preparation and implementation of the upgrade trigger scheme is underway, with a goal of writing 2-5 GB/s to storage [55, 56]. A proposed time-of-flight detector TORCH [57], is also under development for a future upgrade, and is the topic of Chapter 4.

CHAPTER 4

DEVELOPMENT OF THE TORCH DETECTOR

Accurate particle identification is a key requirement of the LHCb physics objectives. As detailed in Chapter 3, charged hadron identification is provided by the RICH detectors [28]. The TORCH (Time Of internally Reflected CHerenkov light) detector is a proposed addition to LHCb which would complement the RICH systems by providing PID of pions, kaons and protons in the low momentum range 2-10 GeV/c [57]. The system would exploit Cherenkov radiation to measure time-of-flight differences between particles. The TORCH project is currently under development, with the possibility of installation into LHCb for Upgrade II. The following sections detail the physics and design of the proposed detector, calibration studies for the front-end electronics, and testbeam campaigns performed with prototype modules.

4.1 Physics of the TORCH detector

Charged particle identification is required for separating decay signals as well as tagging the flavour of neutral parent particles measured by LHCb. As described in Section 3.4, Ring Imaging Cherenkov techniques can be used in conjunction with tracking information to determine the charge and mass of a particle, thus identifying it. The separation power of a detector measuring Cherenkov radiation is determined by the refractive index n of the radiator and the particle momentum p. The RICH systems in LHCb have a lower limit on the kaon momentum of $\sim 10\,\text{GeV}/c$, such that low momentum kaons can only be identified by an absence of photons. This method is clearly prone to high backgrounds

from fake tracks, so an additional radiator would be required to achieve positive kaon identification under 10 GeV/c [57].

Another technique for PID is by time-of-flight measurements, since the time t taken by a particle of mass m to travel a path length L is given by

$$t = \frac{L}{c}\sqrt{1+\left(\frac{m}{p}\right)^2} \approx \frac{L}{c}\left[1+\frac{1}{2}\left(\frac{m}{p}\right)^2\right]. \tag{4.1}$$

This gives an expected time difference between kaons and pions, for example, of

$$t_K - t_\pi \approx \frac{L}{c}\frac{1}{2p^2}\left[m_K^2 - m_\pi^2\right]. \tag{4.2}$$

This principle can be combined with the measurement of Cherenkov angles in the following way. Charged particles passing through an optically dense quartz radiator will emit Cherenkov photons that travel to the edges of the plane by total internal reflection. The time of propagation of the photon through the radiator is then added to the time of flight of the particle up to the emission point, and used to determine the particle species using the relations outlined above [57]. The design of TORCH is based on these principles, inspired by the Belle II TOP [58] and PANDA DIRC [59] systems.

4.2 Design of the TORCH detector

This section will outline the design of the TORCH detector modules as well as the choice of optics and photon detectors. The detector is proposed to be installed into LHCb in front of the RICH2 system, which is about 10 m from the interaction point (see Figure 3.3). In order to achieve a 3σ separation between kaon and pion hypotheses, with a flight path of 10 m and momentum around 10 GeV/c, a timing resolution of 10-15 ps is required per track. Assuming 30 photons detected per track, this translates to a single photon timing resolution of 70 ps [60]. This resolution is associated to two main sources of uncertainty, the first of which is the accuracy of estimating the time of propagation of the Cherenkov photons through the radiator. The second source of uncertainty is the resolution of the photon detectors and readout electronics. To achieve 70 ps resolution overall, each of these two components can have a resolution of about 50 ps.

The TORCH detector uses a thin quartz plate as the radiator, and will be arranged as eighteen identical $660 \times 2500 \times 10\,\text{mm}^3$ modules in a 9×2 grid around the beampipe in LHCb. One such module is displayed in Figure 4.1, where the orientation is also defined. Cherenkov photons travel to the edges of the plate by total internal reflection, where

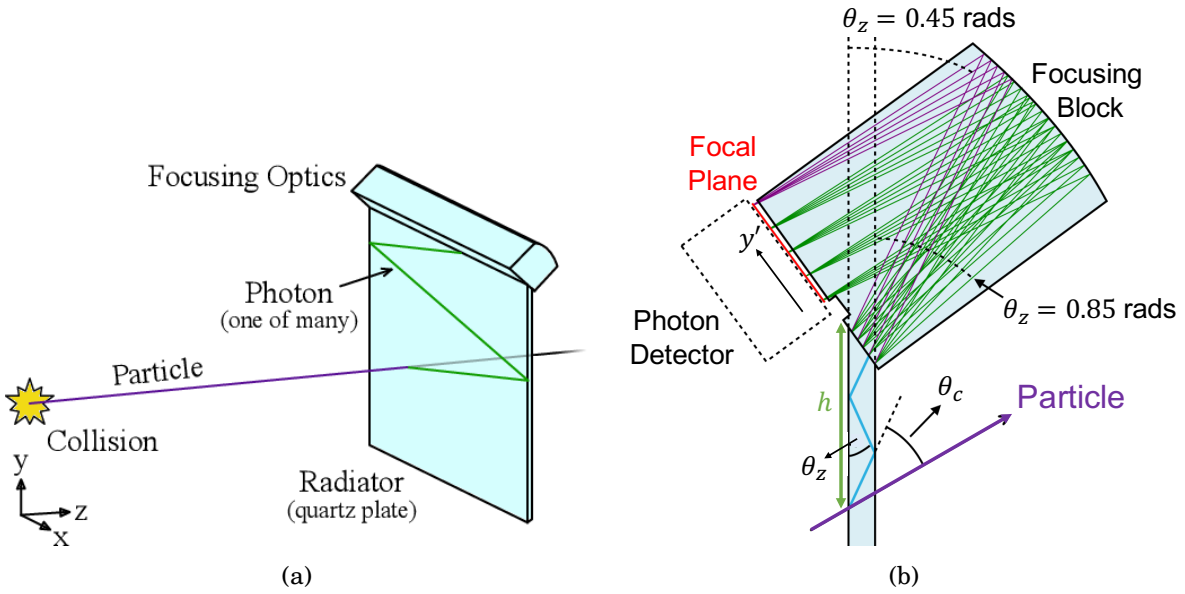


Figure 4.1: Schematic of a TORCH module. (a) Demonstrates a single Cherenkov photon propagating through the radiator via total internal reflection. (b) Shows the photon being reflected by the focusing optics onto the detector plane, for a number of possible photon angles, allowing the Cherenkov angle θ_C to be determined. Figures adapted from Ref. [60].

there are focusing optics, namely a block of quartz with a cylindrical mirrored surface. The block focuses each photon onto a plane of Micro-Channel Plate Photomultiplier Tubes (MCP-PMTs), where its position is measured. The distance along the focal plane, L, is related to the angle θ_z in the $y-z$ plane, according to

$$L = \frac{h}{\sin\theta_z}, \tag{4.3}$$

where h is the distance between the point of emission and the focusing block as indicated in Figure 4.1. The angle in the $x-z$ plane can also be determined from the direction of travel of the emitting particle. Altogether, this information allows the Cherenkov angle θ_C to be measured. For a 50 ps timing resolution, the angular resolution is required to be about 1 mrad [57].

The MCP photomultiplier tubes have been developed specifically for TORCH in collaboration with industrial partners Photek[1]. The tube design is driven by requirements on long lifetime, high spatial resolution, and fast timing [61, 62]. Photomultiplier tubes work through the principle of the photoelectric effect and secondary emission. When a photon hits the photocathode, an electron is emitted and then accelerated across the

[1]Photek Ltd., 26 Castleham Road, St Leonards on Sea, TN38 9NS, UK, `http://www.photek.com/`.

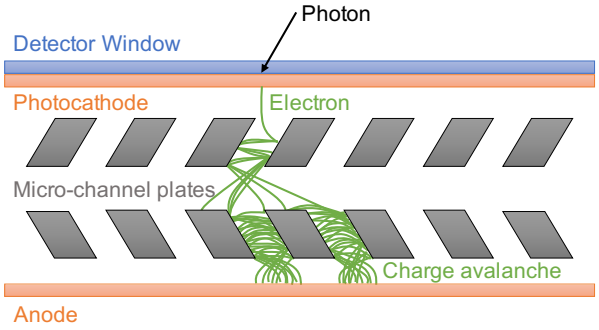


Figure 4.2: Schematic overview of a micro-channel plate PMT. A photon (black) hits the detector window, causing the emission of an electron (green) from the photocathode. A charge avalanche is then created by secondary emission from two planes of micro-channel plates, and collected at the anode.

tube by application of a potential difference. In this case, the electrons are accelerated towards two planes of micro-channel plates, where they are multiplied via secondary emission. The plates are arranged in a chevron formation as shown in Figure 4.2. This compact formation maximises the amplification, as well as achieving a fast response. The so-called charge avalanche is collected at the anode, where the gain is determined by the voltage bias across the tube.

An R&D programme has been underway to deliver MCP-PMTs which meet the requirements of TORCH within the LHCb environment, detailed in Refs. [61–63]. Firstly, an extended lifetime is required for the expected integrated charge on the anodes of $5\,\mathrm{C\,cm^{-2}}$. This is achieved with the use of a coating on the MCPs through Atomic Layer Deposition (ALD) [64], so that the detector can last several years of operation. Secondly, tubes of increasing granularity are developed. In order to achieve the desired resolution, an effective granularity of 8×128 pixels is needed, on an active area of $53 \times 53\,\mathrm{mm^2}$, shown in Figure 4.3. The x-direction, as indicated in Figure 4.1, has a coarser granularity requirement since the modules have a large width, such that they can be instrumented with a 6 mm pitch. Conversely, the y' direction along the photodetector plane requires a pitch finer than 0.4 mm, and therefore greater pixellisation. The final multi-anode design uses a 64×64 channel device, to keep electronic connection density low and reduce the chance of leaks during manufacturing. In the coarse direction, an external printed circuit board (PCB) is used to connect each set of 8 neighbouring anodes into single readout channels. In the other direction, charge sharing between adjacent anodes can be exploited to effectively operate as 128 pixels. This exploits a centroid algorithm, where

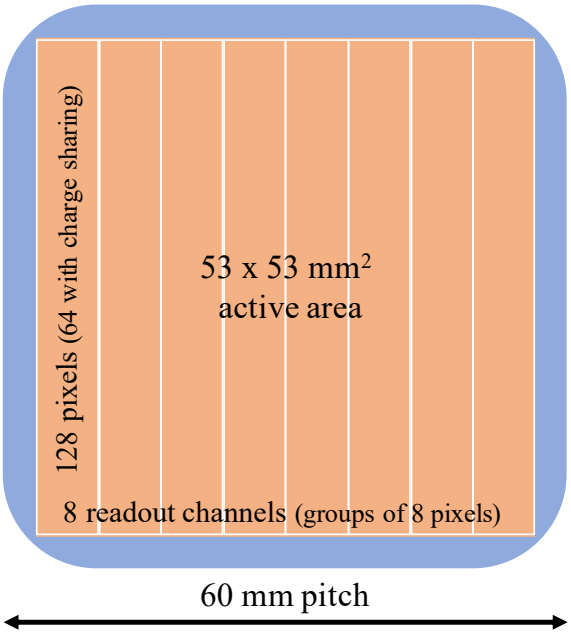


Figure 4.3: Diagrammatic view of the TORCH MCP-PMT design. The tubes have a 60 mm pitch, with an active area of $53 \times 53\,\mathrm{mm}^2$. The device has a 64×64 channel physical layout, and operates as 8×128 effective pixels through charge sharing and grouped readout electronics.

the spatial position of a photon is given by

$$X_\gamma = \frac{\sum_i x_i q_i}{\sum_i q_i}, \tag{4.4}$$

where i is an index running over the number of readout channels, x_i is the position of a readout anode, and q_i is the charge collected by that anode [61].

4.3 Readout electronics

A set of custom electronics has been designed to readout the TORCH MCP-PMTs. To date, the systems have been deployed and tested in testbeam campaigns, described further in Section 4.5. Four main components may be identified, a NINO board [65] for discrimination, an HPTDC board [66] for digitisation, a backplane, and a readout board. The backplane is simply used to connect the HPTDC to the readout boards, which are separated to allow flexibility for an independent readout framework to be developed in the future, in line with the LHCb data acquisition system. For the purposes of the testbeam, the readout boards transmit the signals to a DAQ PC which uses LabView DAQ software [67]. Figure 4.4 shows the main flow of data through the electronics system.

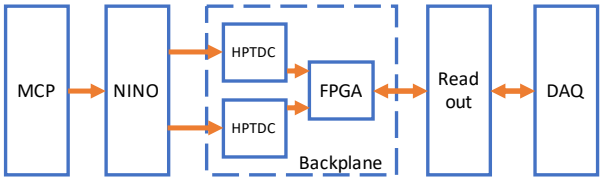


Figure 4.4: Diagram representing the flow of data through the TORCH readout system, showing the main components.

The NINO[2] is an application-specific integrated circuit (ASIC) chip which uses the time-over-threshold (TOT) method to measure an input charge and deliver an output pulse. The input charge can thus be inferred by the width of the output pulse. The TOT principle is illustrated in Figure 4.5, where the threshold above which charges are measured may be set by the user. The characteristic output pulse is referred to as a low voltage differential signal, or LVDS, which serves as a pre-amplification stage for the signal. High-performance time-to-digital converter (HPTDC) chips then measure the time of arrival of the NINO pulses and digitise them in 100 ps bins for readout. Both the leading and trailing edges of the NINO pulses are recorded this way. The latest version of the NINO boards operate with 128 channels, consisting of 4 chips with 16 channels each on the top of the board, and the same on the bottom of the board. This board requires 2 HPTDC chips for readout, with 64 channels each. The electronic system layout allows eight boards to be used for a single MCP-PMT, such that the full 8×128 channel readout is achieved. Data from the HPTDCs are read and buffered by an on-board FPGA, which also allows control of the NINO threshold settings. Finally, the readout board provides clock and trigger signals via HDMI connections, which in this case uses the AIDA Trigger Logic Unit (TLU) [68]. A description of the setup and further details can be found in Ref. [63].

There are two important effects to be considered with the time-over-threshold method used by the NINO. The relationship between charge and pulse width needs to be measured, discussed in detail in Section 4.4. Secondly, variations in pulse amplitude result in timing differences known as timewalk, indicated by Δt in Figure 4.5. This is due to the fact that smaller signals will take longer to exceed the threshold. Therefore a calibration is required to derive a timewalk correction for the leading edge of the input pulse, as a function of the NINO pulse width. A data-driven correction was derived in Ref. [69], and then improved upon as described in Ref. [60]. A single photon produces a cluster of hits in the MCP-PMTs, resulting in a response in multiple channels of the NINO. A

[2]NINO is a pseudo-acronym.

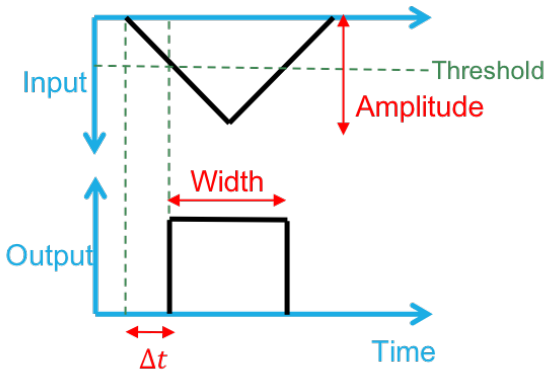


Figure 4.5: Diagram of NINO response using the time-over-threshold principle. An input pulse is received, and while its amplitude exceeds the NINO threshold, an output pulse of the corresponding width is produced.

difference between the time recorded in any given pair of channels in that cluster is therefore the result of timewalk. These differences are measured during data-taking and used to correct the arrival time of each pixel hit in the MCPs.

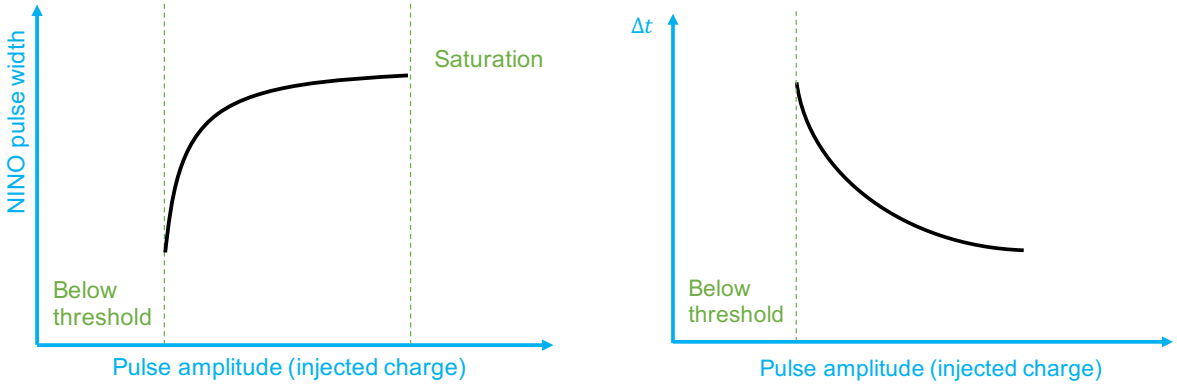
A further correction is required to account for the fact that the bins used to digitise the data by the HPTDC are not uniform in time [70]. This is referred to as the integral non-linearity (INL) correction, and is derived from dedicated high-statistics data samples using established methods as outlined in Refs. [60, 71].

4.4 Calibration studies

4.4.1 Setup

In addition to the timewalk effect already discussed, a calibration of the NINO is also required to characterise the width of its pulse as a function of the charge received. The expected behaviour of both of these effects is shown in Figure 4.6. An accurate measurement of the incident charge, made possible by a charge-to-width calibration, is necessary for the charge sharing technique required to achieve the ideal spatial resolution. Thus it will improve the overall timing resolution of the TORCH detector. This section describes the setup and presents preliminary results of a calibration system developed for this purpose. The goal of the system is to calibrate all NINO boards before they are deployed in future TORCH modules, though the calibration can also be performed on boards used in past testbeams (described in Section 4.5).

The laboratory setup is shown in Figure 4.7 with the main components highlighted.



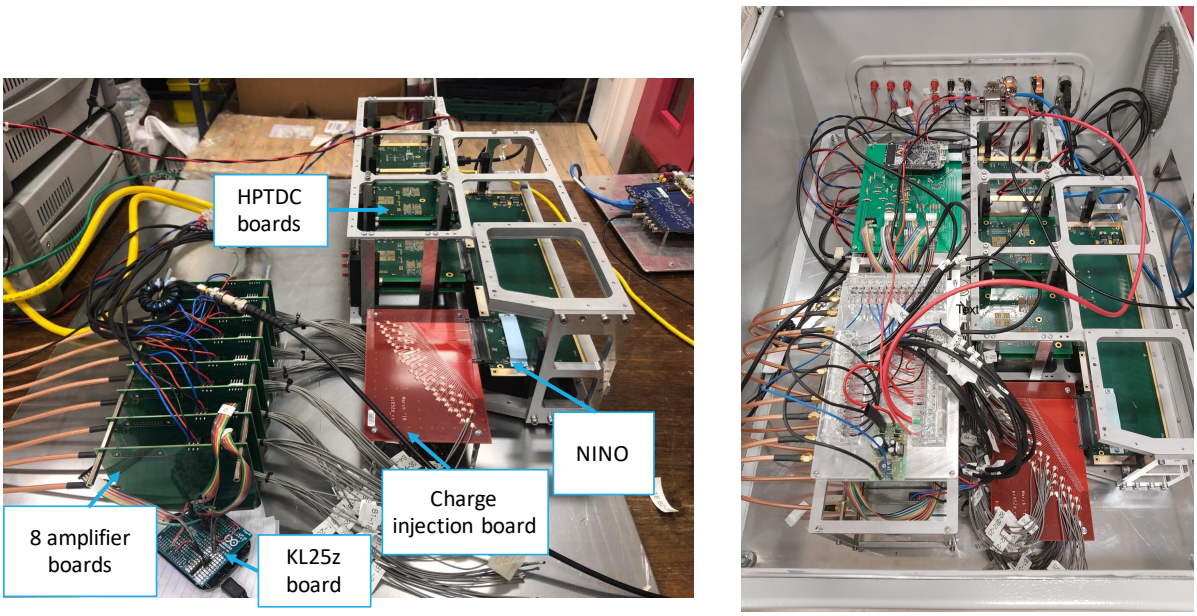
(a) (b)

Figure 4.6: Representations of the characteristic a) NINO pulse width and b) timewalk, as a function of the injected charge. No response is seen below the threshold, and both curves flatten at high amplitudes.

A pulse generator is fed into a set of 8 amplifier boards with 16 channels each, with a one-to-one mapping to the NINO channels. A microprocessor chip, named KL25z, is loaded with firmware that allows it to configure the channels in the amplifier boards. This means that a single channel under test can be turned on with the rest of the channels kept off. Next, a charge injection board routes the input pulses into the NINO, with a geometry designed to work with the electronics layout in TORCH. The NINO board is then connected to HPTDC boards and read out to a DAQ PC running LabView software, as in the full TORCH electronics design. A single NINO board with 128 channels is mounted in the calibration system, which requires two HPTDC boards for readout. The system subsequently moved to an enclosure to reduce suspected noise and interference from the surrounding area, as discussed below.

4.4.2 Data collection

Data acquisition for the TORCH electronics is implemented in LabView. The DAQ used for testbeam campaigns records the timestamps of leading and trailing edges of the NINO pulse, and serialises them by the HPTDC channels which digitise those timestamps. A mapping can then be used to identify the NINO channel which produced the pulse. This software is used with modifications to allow calibration data to be recorded. Both the KL25z microprocessor and the pulse generator can be controlled through LabView commands. Serial commands are received via USB connection to the microprocessor, and the pulse generator is controlled with a General Purpose Interface Bus (GPIB)



(a) (b)

Figure 4.7: Picture of the calibration system setup in the laboratory. a) Main components of the system are labelled and described in the text. Not shown is a pulse generator that connects into the amplifier boards on the left hand side. b) Calibration system mounted in an electromagnetically screened enclosure for noise protection. Note that this enclosure was not available for the studies described in this section.

connection.

In order to acquire a charge-to-width calibration for the NINO boards, data must be recorded for a range of input amplitudes for each channel. The algorithm implemented in LabView is shown as a data flow diagram in Figure 4.8. Firstly the range of channels and range of amplitudes is defined. At the start of the loop, the first channel is turned on using the microprocessor and the starting amplitude is set using the LabView GPIB interface to the pulse generator. These two variables specify the dataset, and data is recorded for a defined time. Next, the voltage is stepped up, and this loop is repeated until the highest amplitude is recorded. Then this loop is repeated for the next NINO channel, and so on, until data is recorded for all channels and all amplitudes. The time for which data is saved is determined by the number of hits needed to have reasonable statistics for the subsequent analysis. A higher amplitude will result in more hits in the same period, so in future the recording time can be varied as a function of the amplitude for more efficient data taking.

An additional calibration is performed to convert the amplitude from the pulse generator to an injected charge into the NINO. The pulse generator is connected to an

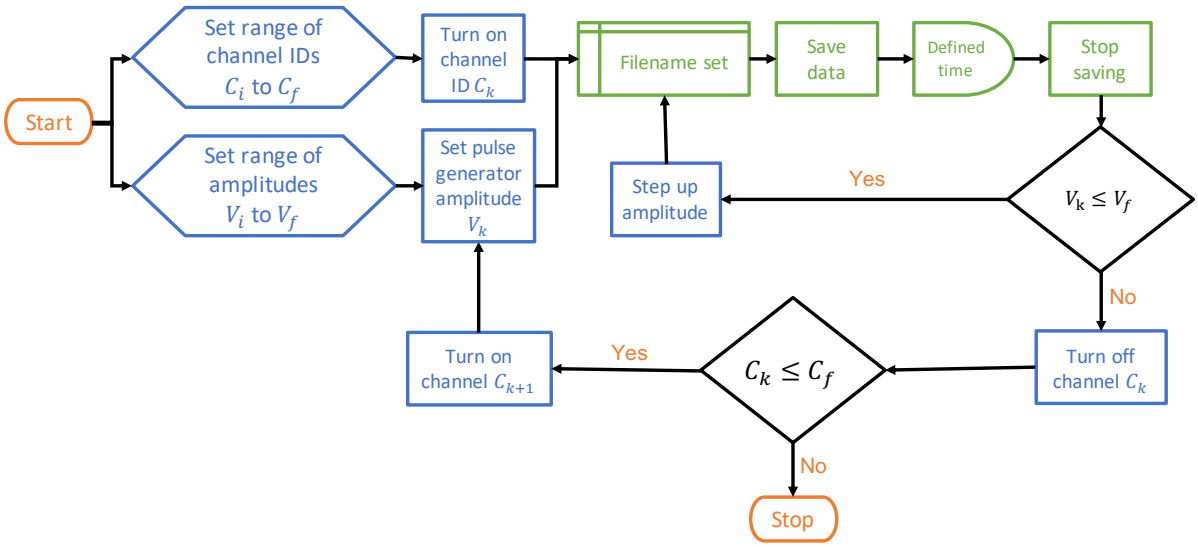


Figure 4.8: Data flow diagram representing the algorithm used for recording charge-to-width calibration data. Every channel of the NINO board is looped through, and within each channel, data is recorded over a range of amplitudes.

oscilloscope, and the charge is extracted as the area under the curve recorded by the scope, divided by the resistance which equals 50Ω.

4.4.3 Results

Distributions of the pulse width for a range of input amplitudes are shown in Figure 4.9 for one particular channel. A long tail is seen for low amplitudes, which is not unexpected, considering the shape of the curve in Figure 4.6(a). At lower amplitudes the curve is much steeper, so the expected error on the width will be higher. For low amplitudes the distributions are fitted with a Crystal Ball function [72] to model the tails, and for high amplitudes, a simple Gaussian function is employed. The fitted mean of each distribution is taken as the pulse width for that amplitude, and the error on the pulse width is given by the standard deviation of the Gaussian.

Figure 4.10 shows the result of the calibration of the pulse generator to extract the charge. Data is collected using the oscilloscope for three separate channels for amplitudes in the range 240-3500 mV. Thus the injected charge into the NINO can be determined from the input amplitude. A divergence in the response between channels is seen above 1000 mV, which is attributed to an issue with the pulse generator. This has since been replaced, but for the purposes of the results shown here, the average of the three channels is used as the corresponding charge, and the error is taken as the largest difference in

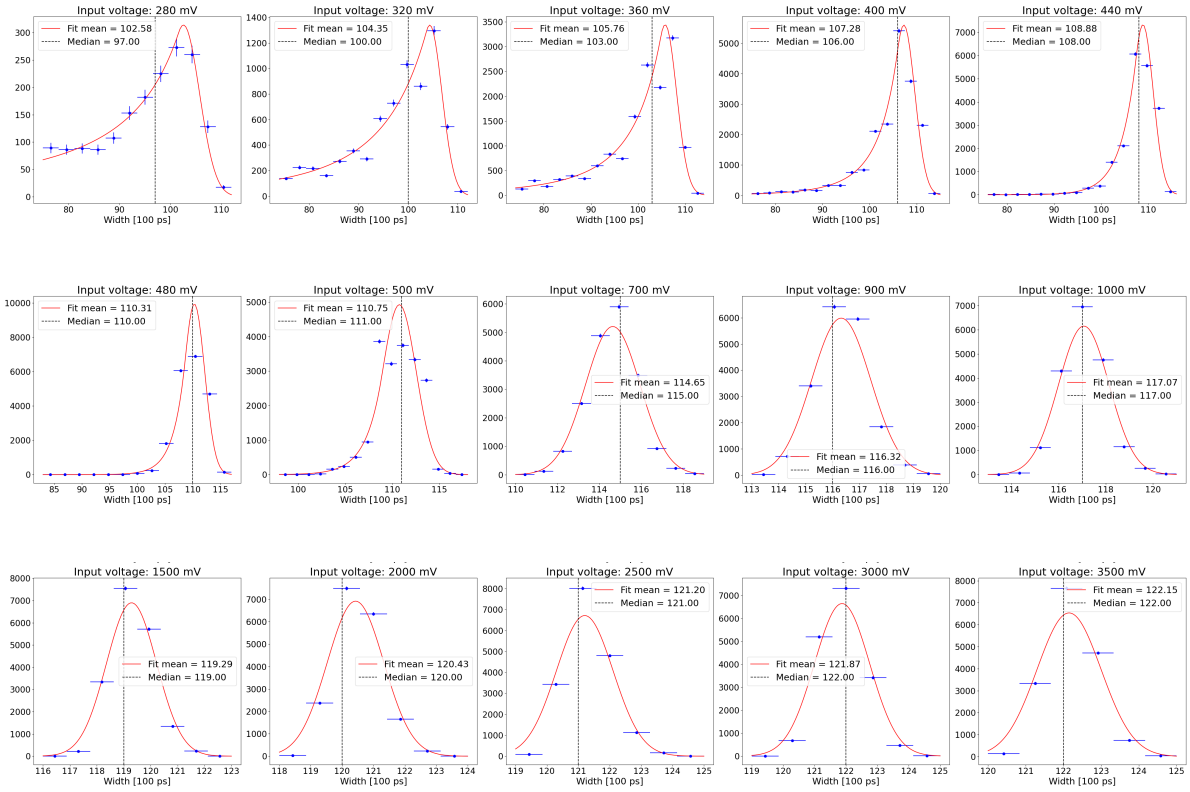


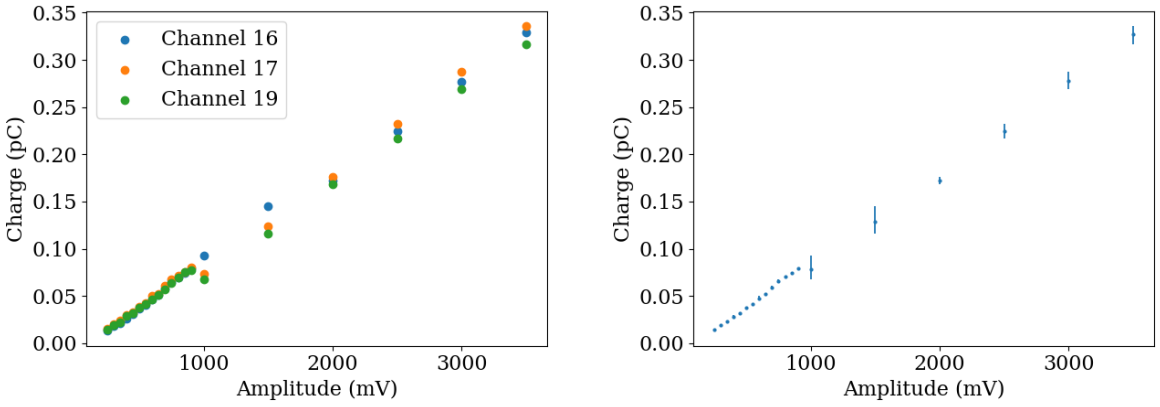
Figure 4.9: Distributions of NINO pulse width for a range of amplitudes, shown for a single channel. The blue points are data, with statistical errorbars, and the red line is the Crystal Ball or Gaussian fit projection. The dashed black line indicates the median of the distribution.

the readings, as shown in Figure 4.10(b).

These two measurements allow charge-to-width curves to be plotted for all the NINO channels under test. Curves for a number of channels from two NINO chips are shown in Figure 4.11. Where data is available up to high amplitudes, such that the linear section of the curve is accessible, the data is fitted with the function

$$ax + b - \frac{c}{x}, \tag{4.5}$$

where x represents the charge in pC, a and b are parameters describing a linear function, and c describes the curvature. For channels where data is collected for a lower range of amplitudes, the linear region cannot be determined and a is set to 0. The mean of the fitted curves is shown in Figure 4.12, along with the 95% confidence interval region. Moreover, the distributions of parameters from all the fitted channels are shown in Figure 4.13. The region of highest curvature is sufficiently populated such that the

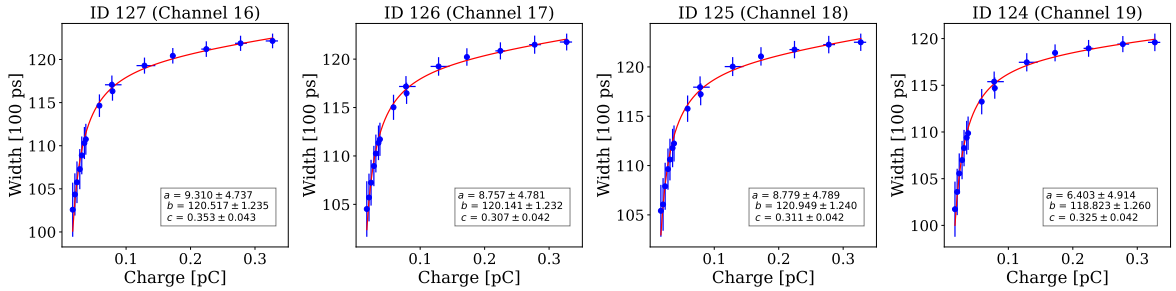


(a) (b)

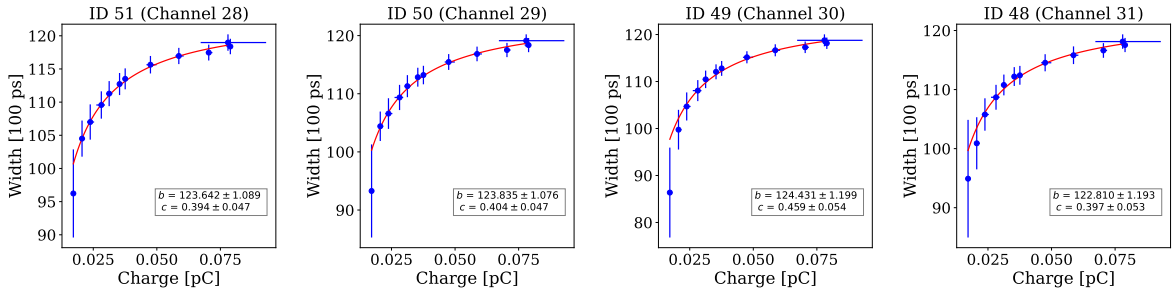
Figure 4.10: Relationship between pulse generator amplitude and resulting charge. a) Data measured for three NINO channels, and b) average result and errors used for calibration.

variation in c across the channels is at the same level as the error on the individual fits. For the linear parameters, more data is required at higher amplitudes to comment on the channel-by-channel variation. Overall, 87 channels are fitted successfully, out of 96 instrumented channels. The difference is due to a number of channels displaying faulty behaviour, as well as two channels used as time references.

Channels with sufficient data collected show the expected characteristic width as a function of the input charge, serving as proof of concept for the calibration system. The automation of the system also allows for efficient data collection. However, the large number of channels showing inconsistencies point towards a high level of noise, likely due to electromagnetic interference. This is demonstrated in Figure 4.14 which shows distributions of pulse arrival times with respect to the widths of the pulses. Pulse arrival time is defined as the time difference between the leading edge of the time reference pulse, t_{ref}, and the pulse from the channel under test, t_{NINO}. The time resolution can then be extracted as the width of the arrival time distributions, also shown in Figure 4.14. For the data taken with the setup described here, the time resolution is greater than 100 ps for most of the channels under test. Previous similar measurements, such as in Ref. [73], are closer to 50 ps. Abnormally poor resolution is likely to be due to noise in the system.



(a)



(b)

Figure 4.11: Charge-to-width calibration curves for various channels of the NINO board. The errors in charge reflect the calibration of the pulse generator shown in Figure 4.10. The red line is the fit projection, where in b) the parameter $a = 0$.

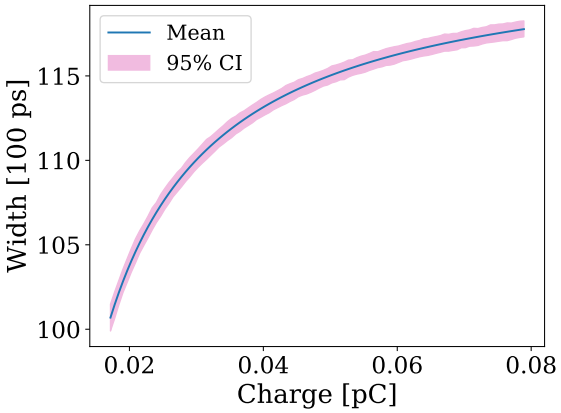


Figure 4.12: Mean fitted charge-to-width calibration curve across all channels. The highlighted band represents the 95% confidence interval across the fitted parameters. Shown for the low charge region where the parameter $a = 0$.

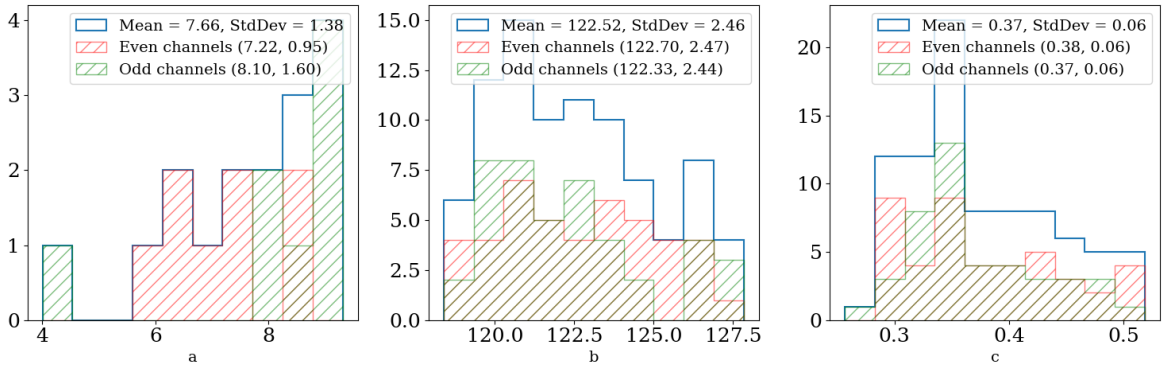


Figure 4.13: Distribution of fit parameters from charge-to-width curves. The parameter a is extracted from 16 channels only, and b and c are extracted from all 87 available channels. The mean and standard deviation of each parameter is shown in the legends. Odd and even channels are isolated to check for systematic differences.

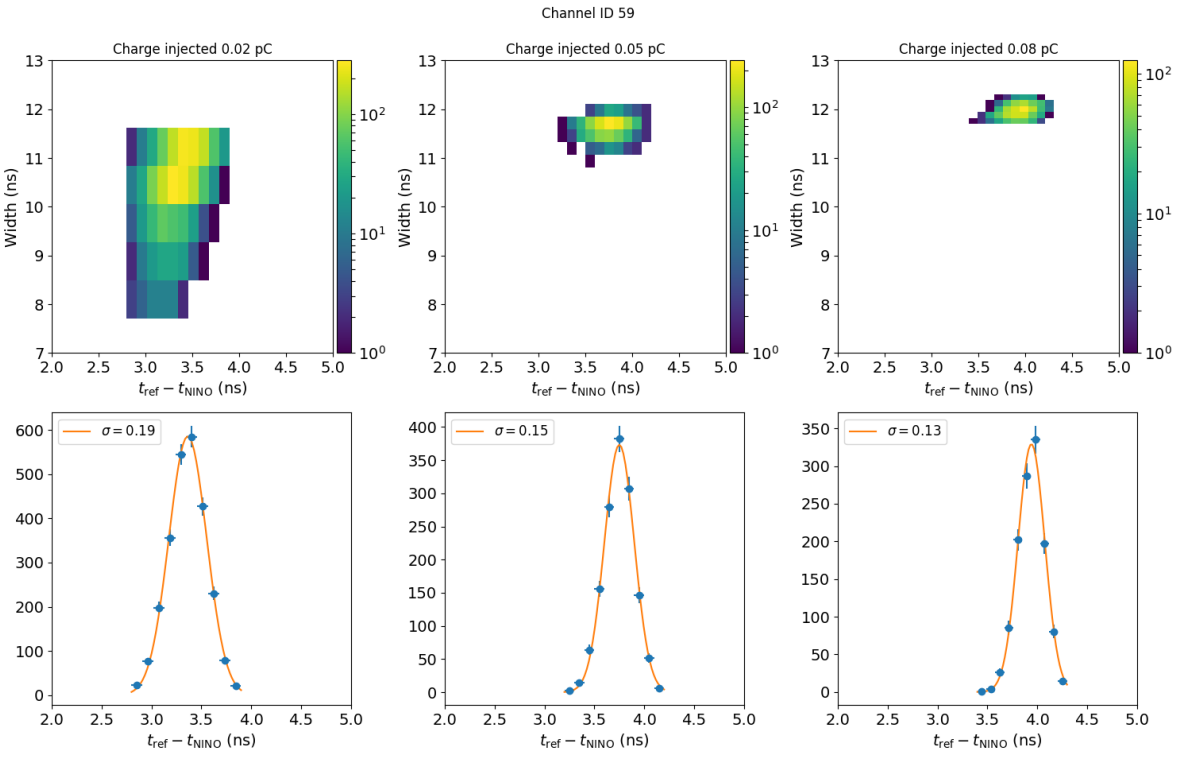


Figure 4.14: Arrival time of NINO pulses for a range of injected charges in a single channel. Top row shows distributions of pulse width against arrival time. Bottom row shows the distributions of pulse arrival times, fitted with a Gaussian distribution. The fitted widths are displayed in the legends.

4.4.4 Next steps

Once the system is further developed, the variation across channels as demonstrated by Figure 4.12, can inform the need for a channel-by-channel calibration. For example, if it is seen that the curves are very stable across different channels, it may be possible to simply take the values from a subset of channels, with the appropriate uncertainty applied. This would of course reduce the time required for charge-to-width calibrations.

Other laboratory tests indicated different behaviour of odd and even channels. This is minimally examined in Figure 4.13, which does not show a clear difference in response. The NINO threshold is also set separately for odd and even channels. Previous studies have shown that the threshold settings result in varying thresholds across channels and chips [73], and this will clearly also impact the measured pulse widths. Thus further studies should be pursued to calibrate the ideal threshold settings required to produce a stable response across all channels.

The data-driven timewalk correction discussed earlier can also be replaced by a dedicated timewalk calibration using the same setup as described here. Some of the readout channels are implemented as time reference channels, such that they record the clock time which is synchronised with the trigger signals. A signal can be sent to the time reference channel and a channel under test. The signal can be controlled with an attenuator to change the amplitude supplied to the test channel with respect to the time reference channel. The difference in recorded times of the two channels thus gives a measurement of timewalk as a function of the input charge.

As displayed in Figure 4.7, an enclosure can be used to reduce electromagnetic interference producing noise in the calibration system. Data should be collected with the system mounted in the enclosure, and the analysis described above should be repeated. Once a full NINO board is calibrated, the charge measurements can be used in the centroid algorithm (Equation 4.4), to improve the spatial resolution of each cluster of hits resulting from an incident photon in the MCP-PMTs. This resolution will then propagate to the overall timing resolution of TORCH, so that the ideal 70 ps single photon timing resolution may be achieved.

4.5 Testbeam campaigns

Several testbeam campaigns have been run to test and develop prototypes of the TORCH detector. The most recent campaigns took place during 2017 and 2018. Firstly, a small module named Mini-TORCH, measuring $120 \times 350 \times 10\,\mathrm{mm}^3$, was deployed with two

MCP-PMTs. The full setup and and subsequent data analysis from two testbeam campaigns with Mini-TORCH are described in Ref. [60]. Secondly, Proto-TORCH, a module measuring the full width and half the height of the baseline LHCb modules, was deployed in October 2018. This used two improved MCP-PMTs, with 8×128 effective pixellisation. The setup and a preliminary analysis of results is presented in Ref. [74]. A short summary of the testbeam conditions and findings is outlined here.

TORCH detector prototypes are installed in the East Hall T9 facility at CERN, where a beam is received from the Proton Synchrotron facility. This beam contains a mixture of protons and pions, with a maximum momentum of 15 GeV/c. Cherenkov counters in the beam area, consisting of barrels filled with pressurised CO_2, provide PID capabilities as pions will emit Cherenkov radiation in them while protons are above the threshold for emission. Multiple timing stations are also used in the beamline, both for the trigger and for PID. A coincidence of signals in more than one timing station is required in the trigger to reduce noisy signals. Combining timing information from stations that are spaced ~ 11 m apart in the area also gives independent PID for protons and pions propagating over that distance. The TORCH modules are mounted on a translation table such that data can be collected with the beam positioned on different areas of the radiator plate. The distribution of photon arrival times recorded from incident pions on Proto-TORCH is shown in Figure 4.15. This is displayed as a function of the pixel number in the y' axis as defined in Figure 4.1. Distinct bands corresponding to different photon paths through the radiator are seen, showing good agreement with simulation.

Studies of Mini-TORCH show a single-photon timing resolution of ~ 100 ps [60], and the analysis of data from Proto-TORCH, which is still ongoing, measures resolutions in the range 60-130 ps [75]. This is an encouraging sign that the goal resolution of 70 ps is within reach, which will also be aided by the completion of the electronics calibration described in Section 4.4.

4.6 Simulation and physics performance

Simulation and studies of the impacts on physics analysis of the proposed TORCH detector are underway, with a recent update presented in Ref. [76]. Highlights of measurements which will benefit from the inclusion of TORCH are outlined in the Physics case for Upgrade II of LHCb [54]. These primarily involve high-multiplicity or low-momentum final states, especially those containing protons, in studies of CP-violation and exotic searches. Integration of the TORCH detector in the full LHCb detector simu-

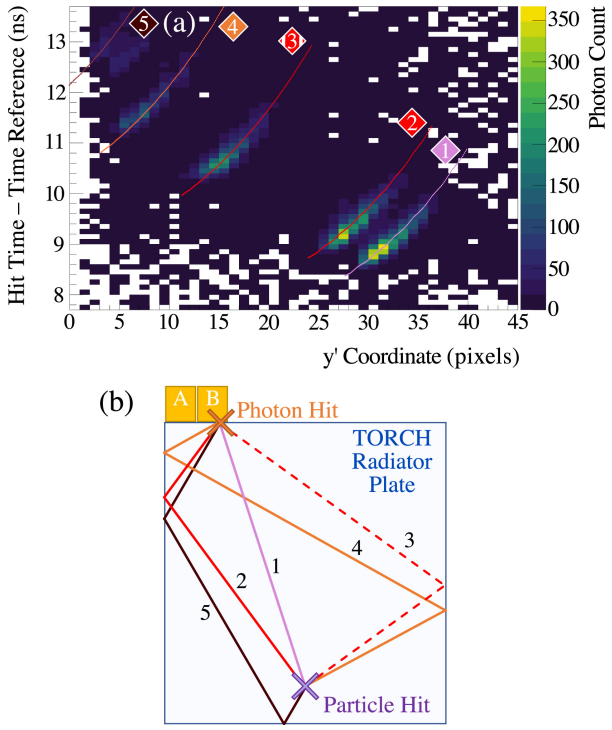


Figure 4.15: Photon arrival times as a function of the position on the MCP-PMT plane (y' axis), shown for pions detected by Proto-TORCH. The overlaid lines in a) correspond to the expectation from simulation for the photon paths indicated in b). Figure from Ref. [74].

lation is executed in GAUSS [45], and the pattern of Cherenkov photons are translated into particle hypotheses using an analogous algorithm as in the RICH systems [35]. The current status of the PID performance is shown in Figure 4.16, displaying excellent separation of particles at low momentum, and good separation of protons and kaons even up to 20 GeV/c. Leveraging this performance is also shown to significantly increase yields of interesting decay modes as well as suppressing backgrounds [76]. Together with the physical developments to the detector, these studies demonstrate that TORCH can be expected to provide significant improvements to LHCb physics.

4.7 Summary

This chapter presents the development of the TORCH detector to provide fast-timing and particle identification in the low momentum range in LHCb. The design and instrumentation of the detector is described, and calibration studies for the readout electronics are presented in detail. A system for charge-to-width calibration of the discriminator chips used in TORCH is developed, and data is collected with an automated DAQ system.

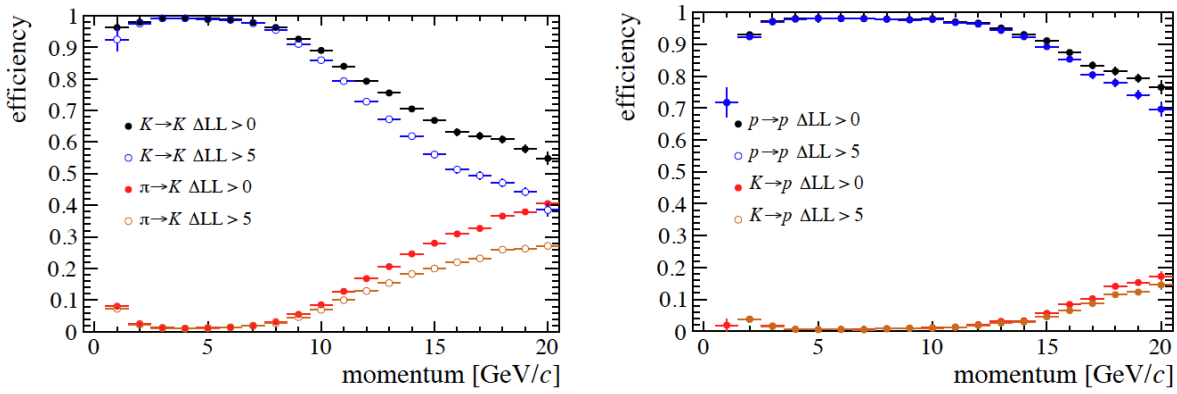


Figure 4.16: The efficiency of particle identification for (left) kaons and (right) protons by TORCH in LHCb, simulated for an instantaneous luminosity of $\mathcal{L} = 2\times10^{33}$ cm^{-2} s^{-1}. Figure from Ref. [76].

Analysis of the results shows the expected behaviour, as well as consistent responses across working channels. Improvements to the system are underway, and suggestions for further work are outlined. Additionally, three testbeam campaigns are performed, and the results summarised. The single-photon timing resolution of the TORCH detector is improving, and the calibration of NINO boards using the setup described here is hoped to improve this further in order to achieve the target resolution of 70 ps.

CHAPTER 5

PHENOMENOLOGY OF DOUBLY-CHARMED B^0 DECAYS

Doubly-charmed decays of b-hadrons offer a rich area of research, especially with the LHCb detector providing the opportunity to study ever rarer modes with the largest dataset of heavy flavour decays to date. The decays are of the form $B \to D\overline{D}h$ where B represents any b-hadron, D represents any c-hadron, and h can be any light or heavy hadron. These decays provide a lot of scope for spectroscopy studies, which involve investigating intermediate resonant structures in the decay paths. The analysis in this thesis focuses on the decay $B^0 \to D^0\overline{D}^0K^+\pi^-$, which has additional implications for non-local contributions, so-called *charm loops*, in rare $b \to s\ell^+\ell^-$ transitions [77]. This chapter discusses the phenomenological motivations of such studies as well as the opportunities for spectroscopy and charm loop studies provided by studying $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays.

Section 5.1 introduces the framework of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays, including an illustration of the possible Feynman diagrams, followed by a description of the current experimental status in Section 5.2. Spectroscopy of traditional and exotic states is discussed in Section 5.3, and Section 5.4 gives an introduction to the effective field theory approach of loop-level processes involving $b \to s\ell^+\ell^-$ transitions, and the significance of long-distance contributions therein.

5.1 Theory of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays

Doubly-charmed decays of b-hadrons with a kaon in the final state occur via the transition $b \to cW^-(\to \overline{c}s)$. This transition is favoured due to the relative size of the CKM element

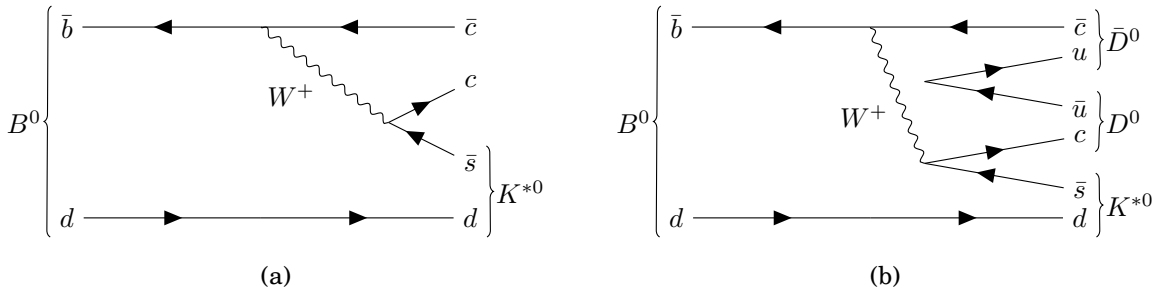


Figure 5.1: Feynman diagrams with internal W emission, resulting in final states with charmonium (left) or $D^0\overline{D}^0$ (right) states and a K^{*0} meson.

V_{cb}, which governs $b \to c$ transitions. Note that charge conjugation is implied throughout. There are two distinct topologies of the decay, with internal and external W emission, corresponding to colour-suppressed and colour-favoured amplitudes respectively. The colour suppression arises from the requirement on the arrangement and colours of the quarks to form colour singlet states.

A number of Feynman diagrams contributing to $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays may be considered as the following. Figure 5.1 shows the internal W emission diagram where the final state arises from a $D^0\overline{D}^0K^{*0}$ state, or more generally a $c\bar{c}K^{*0}$ state. Conversely, diagrams with external W emission are displayed in Figure 5.2. These include modes with $c\bar{s}$ final states as in $B^0 \to D^+_{sJ}\overline{D}^0\pi^-$, as well as the decay used as a control mode in this thesis, $B^0 \to D^{*-}D^0K^+$. Here the D^+_{sJ} is used generally to represent charm-strange mesons which are also often excited states, and specified by their spin, J. A highly suppressed diagram is also shown in Figure 5.3, which occurs via the transition $b \to uW^-(\to \overline{u}s)$. This suppression has two sources, first of which is the creation of a $c\bar{c}$ pair, which is far less energetically favourable than the $u\overline{u}$ pair in the earlier diagrams. Secondly, there is a suppression from the CKM elements, displayed in Figure 5.3, given by the factor

$$\left(\frac{|V_{ub}V_{us}|}{|V_{cb}V_{cs}|}\right)^2 \approx 4 \times 10^{-4}. \tag{5.1}$$

5.2 Experimental status

Historical interest in $b \to c\bar{c}s$ decays arises from the "charm counting puzzle" [78]. This was an apparent discrepancy between the semileptonic and charmed branching fractions of B decays. In order to describe the measurements of semileptonic decays, a larger than

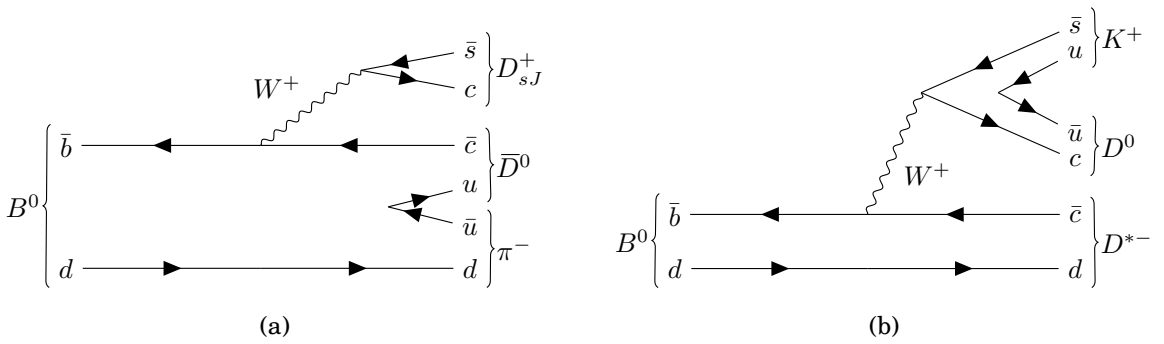


Figure 5.2: Feynman diagrams with external W emission, showing $B^0 \to D_{sJ}^+ \overline{D}^0 \pi^-$ (left) and $B^0 \to D^{*-} D^0 K^+$ (right) decays.

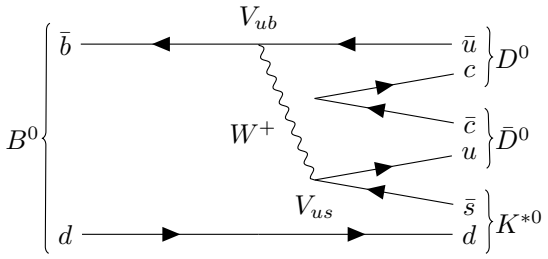


Figure 5.3: Highly suppressed diagram of $B^0 \to D^0 \overline{D}^0 K^{*0}$ decays with the disfavoured CKM elements indicated.

expected charmed branching fraction was needed. In 1995, it was proposed that a neglected and potentially significant fraction could be due to $B \to D\overline{D}Kh$ decays [79]. This puzzle is resolved nowadays, with measurements involving $B \to D^{(*)}\overline{D}^{(*)}K$ decays performed by the ALEPH, BaBar, Belle and LHCb collaborations [80–85], with spectroscopy studies as the driving factor. However, no measurements involving $B \to D^{(*)}\overline{D}^{(*)}K\pi$ transitions have been performed until now. The $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ branching fraction can be estimated based on considerations of similar decay modes (discussed later as in Equation 7.12), and is expected to be $\mathcal{O}(10^{-4})$. The product of the branching fractions including $D^0 \to K^- \pi^+$ charm meson decays is then much smaller, at the level of $\mathcal{O}(10^{-7})$.

The kinematic diagram of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays is shown in Figure 5.4. This displays the topology of the decay in the context of the LHCb detector, which also applies to similar modes such as $B^0 \to D^{*-} D^0 K^+$, which is used as a control mode in the following analysis.

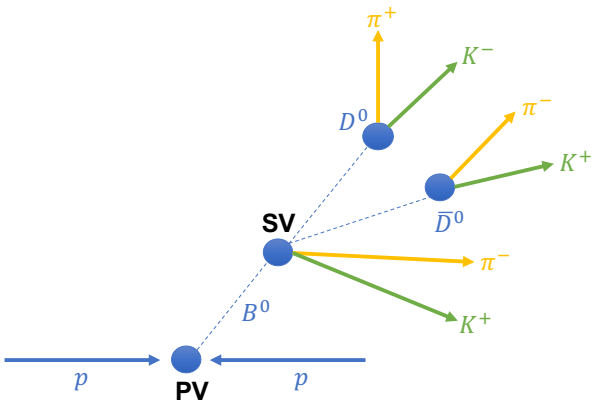


Figure 5.4: Schematic topology of the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ from the pp collision point. The same diagram can also represent the decay of $B^0 \to D^{*-} D^0 K^+$, since at the decay vertex of the B^0 (SV), the decay $D^{*-} \to \overline{D}^0 \pi^-$ may occur.

5.3 Spectroscopy

Hadron spectroscopy refers to the measurement of masses and other intrinsic properties of hadrons through studying their production mechanisms and their decays. This involves the study of the strong interaction, governed by QCD, which poses significant challenges. At short distances, asymptotic freedom applies, and perturbative calculations may be performed. However, at long distances, or equivalently low energy, the coupling exceeds unity such that perturbation theory cannot be applied. The scale that separates these two regimes is given by $\Lambda_{\text{QCD}} \simeq 200\,\text{MeV}$. The challenge faced by heavy flavour experiments and theorists is to disentangle these hadronic contributions from the quark-level interactions governed by the weak force. Reviews of the theoretical framework of hadron spectroscopy can be found *e.g.* in Refs. [86, 87].

Heavy hadrons containing a charm or bottom quark have opened up a vast arena of research, due to increasingly precise experimental measurements as well as progressing theoretical models. These hadrons, where the heavy quark is labelled Q, allows the exploitation of the heavy quark limit, $\text{m}_Q \to \infty$, since its mass $m_Q \gg \Lambda_{\text{QCD}}$. In this regime, a simplified framework called the Heavy Quark Effective Theory (HQET) [88], is used to predict properties of heavy hadrons. Considering a heavy meson as $Q\overline{q}$, the interaction between the heavy quark and the light degrees of freedom (the light anti-quark and the gluons) is proportional to $1/m_Q$, which dies off in this limit. Thus, the spin and parity quantum numbers are separately conserved for the heavy quark and the light degrees of freedom in strong interactions, and the properties of the meson can be determined solely from the total spin of the light system, j_q [87, 89]. As more,

sometimes unexpected, heavy hadrons are observed through experiment, theoretical models continue to be developed to improve our understanding of these non-perturbative QCD dynamics, some of which are highlighted in Refs. [86, 90, 91].

A number of experimental results demonstrate the existence of exotic particles that do not fit the conventional meson ($q\overline{q}$) or baryon (qqq) model. These particles, containing four or five quarks, satisfy the requirement for colour singlet states, but are coined exotic because they have only been observed through relatively recent experimental evidence. Most notably, the observation of pentaquark states by LHCb [16, 92] and the observation of the $X(3872)$ and $Z(4430)^-$ tetraquark candidates by Belle [93, 94]. There is considerable debate about the nature of these states, for example, whether the tetraquark candidates are better described as compact four-quark [95] or molecular meson-meson [96] objects. The search for exotic particles is of interest to the study of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays because of discoveries made in similar decay modes, as detailed below.

Studying $B \to D\overline{D}h$ decays provides access to the charmonium spectrum, and the charm-strange spectrum for kaons in the final state. Charmonium refers to states containing $c\overline{c}$ quarks, and an important definition is the open charm threshold, where 'open' refers to a single charm quark instead of a 'hidden' charm-anticharm pair. This threshold is given by the invariant mass of the $D\overline{D}$ system, such that states above the threshold can decay to a $D\overline{D}$ pair. Below this threshold, all observed states are in excellent agreement with the conventional quark model. Above the threshold however, a number of exotics referred to as XYZ states have been observed with tetraquark signatures or unexpected quantum numbers. A review of these states can be found in Ref. [97], and they are indicated on the charmonium spectrum displayed in Figure 5.5. Decays to $D^0\overline{D}^0$ states can help to understand the exact nature of such exotic states above the open charm threshold.

Of similar interest is the charm-strange spectrum, of D_{sJ} resonances consisting of $c\overline{s}$ mesons. Recent experimental and theoretical reviews of D_{sJ} spectroscopy can be found in Refs. [98, 99], highlighting the uncertain assignments of several observed states and the need for further measurements. Most recently, open charm tetraquark candidates dubbed $X(2900)$ were observed by LHCb in the D^-K^+ mass spectrum of $B^+ \to D^+D^-K^+$ decays [5, 100]. With minimal $\overline{c}d\overline{s}u$ quark content, these are manifestly exotic states, falling into neither of the charmonium or charm-strange categories described above. This ignited a flurry of activity in the theory community, for example Refs. [101, 102], opening up a new area of open flavour spectroscopy.

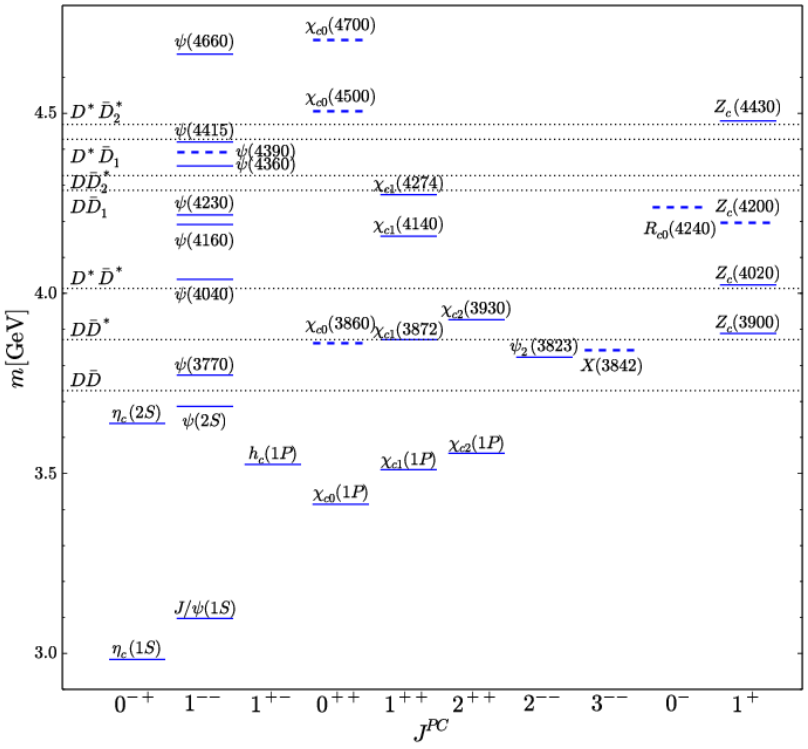
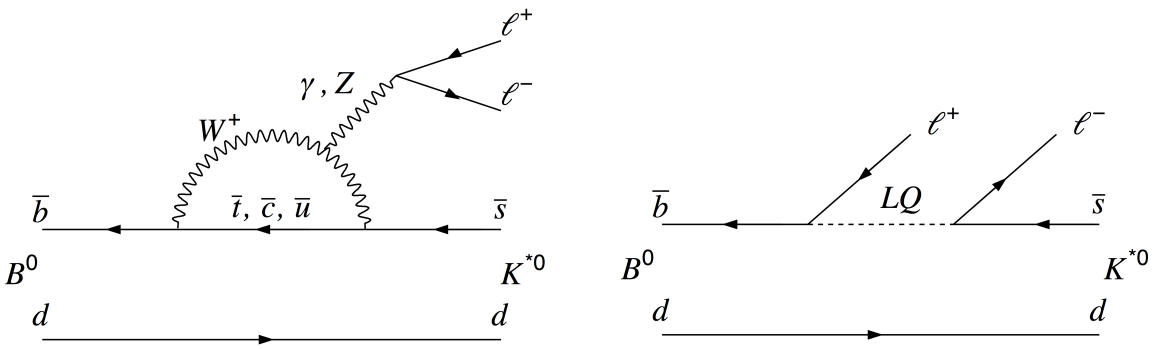


Figure 5.5: Spectrum of charmonium-like states as of July 2019. Solid blue lines correspond to established states, while dashed blue lines correspond to first observations. Dotted black lines are also shown at various energy thresholds. Figure from Ref. [97].

The decay $B^0 \to D^0\overline{D}^0K^+\pi^-$ provides the opportunity to study and search for $c\bar{c}$ and $c\bar{s}$ final states, as demonstrated by the Feynman diagrams in Figures 5.1 and 5.2, as well as potential exotic states. Examples of interesting searches include the charmonium states $\chi_{c(0,2)}(3930)$, where there is some debate over the spin assignment [100]. The D^0K^+ mass spectrum gives access to D_{sJ}^+ states, and $\overline{D}^0K^+$ could access exotic states akin to the aforementioned $X(2900)$, such as a potential isospin partner. The $Z(4430)^-$ has minimal $c\bar{c}d\bar{u}$ content, so could be probed with the $D^{*-}D^0$ spectrum in $B^0 \to D^{*-}D^0K^+$ decays, or a search for its isospin partner could be pursued in the $D^0\overline{D}^0$ spectrum. Along with conventional charmonium measurements, it is clear that an amplitude analysis of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays has exciting prospects for charm and exotic spectroscopy studies.



(a) Electroweak penguin diagram of $B^0 \to K^{*0}\ell^+\ell^-$.

(b) Tree-level diagram of $B^0 \to K^{*0}\ell^+\ell^-$ with a New Physics leptoquark coupling.

Figure 5.6: Leading order SM diagram (left) of the FCNC decay involving the transition $b \to s\ell^+\ell^-$. Example of a BSM diagram (right) for the same decay. Figures from Ref. [103].

5.4 Charm loop contributions to $b \to s\ell^+\ell^-$ transitions

A core part of the LHCb physics programme is to study rare processes which can test the limits of the Standard Model, and thus search for hints of physics Beyond the Standard Model. In particular, flavour-changing neutral current (FCNC) processes, wherein a quark changes flavour via mediation of a neutral boson *i.e.* Z, γ, such as the quark-level transition $b \to s\ell^+\ell^-$ are the topic of much study. These transitions are suppressed in the SM as detailed in Section 2.2, proceeding via an electroweak "penguin" loop or box diagram. This means they are highly sensitive probes of New Physics (NP), as a tree-level BSM process could occur at rates comparable to the SM loop processes. The SM diagram of these processes as well an example of equivalent BSM processes are shown in Figure 5.6. Therefore, studying the branching fractions and angular distributions of decays involving these transitions, such as the decay $B^0 \to K^{*0}\mu^+\mu^-$, offers the exciting prospect of probing potential NP scenarios. In addition, since electroweak interactions between leptons and gauge bosons are independent of flavour in the SM, lepton flavour universality (LFU) can also be tested.

Measurements of decays involving $b \to s\ell^+\ell^-$ transitions have been performed by many experiments, which study the differential branching fractions as well as angular observables, as a function of the invariant mass squared of the dilepton system, $q^2 \equiv m^2_{\ell\ell}$. The decay $B^0 \to K^{*0}\mu^+\mu^-$ is considered a golden mode because the polarisation of the K^* allows the angular distribution of the decay to be precisely determined, thus making

it a rigorous test of SM predictions [104]. A range of these measurements have shown tensions with Standard Model expectations, *e.g.* Refs [105–116], further motivating the search for NP processes as well as explanations for the anomalies such as new gauge bosons or leptoquarks [117–119]. The potential diagram with a leptoquark is shown in Figure 5.6(b), where the proposed 'LQ' particle describes a new coupling between leptons and quarks. However, there is considerable debate whether the theoretical uncertainties associated with hadronic effects, especially long-distance contributions from charm loops, could alleviate these tensions [120–123].

Similarly, measurements of ratios of decay rates of $B \to H l^+ l^-$ where H represents a hadronic state, i.e.

$$R_H = \frac{\int \frac{\mathrm{d}\Gamma(B \to H\mu^+\mu^-)}{\mathrm{d}q^2}\mathrm{d}q^2}{\int \frac{\mathrm{d}\Gamma(B \to He^+e^-)}{\mathrm{d}q^2}\mathrm{d}q^2}, \tag{5.2}$$

are effective probes of LFU violation. Measurements of the ratios R_K and R_{K^*} by the LHCb collaboration are found in Refs. [3,103], which consistently show values lower than the expected SM value of close to unity. In fact the measurement of R_K with the full LHCb dataset shows evidence of LFU violation with a significance of 3.1σ. These measurements are strong tests of the SM as the aforementioned hadronic uncertainties cancel in the ratio, so these observables are theoretically clean. Nevertheless, as the largest theoretical uncertainty, a better understanding of charm loop contributions is essential in informing the interpretations of the anomalies in terms of New Physics models [123]. Experimental measurements of the R_H ratios along with the measurements of branching fractions and angular observables mentioned above, allow global calculations to be performed for various NP models. An example of a global fit for particular Wilson coefficients (defined in Section 5.4.1) is shown in Figure 5.7, which displays a deviation from the SM expectation at a level greater than 6σ. Moreover, considering only the theoretically clean observables and the most recent experimental results as of March 2021, BSM predictions in Ref. [124] are preferred at the level of 4.7σ. Therefore, it is clear that these so-called b-anomalies are a crucial area of NP searches, and controlling hadronic uncertainties will further consolidate the results as well as directing investigations of NP models. The framework of measurements of $b \to s\ell^+\ell^-$ processes is outlined below, including a discussion of the need to estimate hadronic uncertainties with experimental input from $b \to c\bar{c}s$ processes such as $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays.

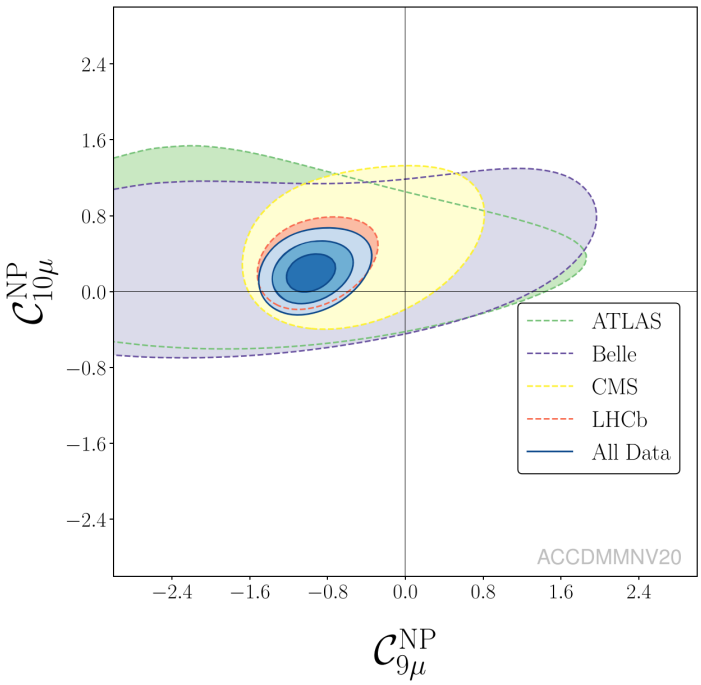


Figure 5.7: Allowed regions of the shift in the vector and axial-vector couplings due to New Physics, $C^{\rm NP}_{9\mu}$ and $C^{\rm NP}_{10\mu}$, given the measurements of muon decay modes performed by the indicated experiments, updated up to June 2020. The origin corresponds to the SM expectation, in the absence of NP. Figure from Ref. [125].

5.4.1 Effective field theory

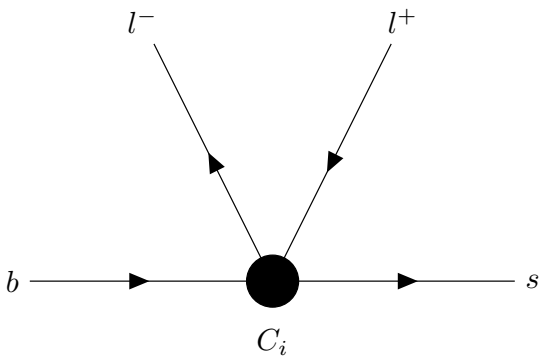
In order to probe NP in a model-independent fashion, an effective field theory approach is employed. The effective Hamiltonian of $b \to s\ell^+\ell^-$ decays is given by

$$\mathcal{H}_{\rm eff} = \frac{-4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_{i=1}^{10} \mathcal{C}_i \mathcal{O}_i, \tag{5.3}$$

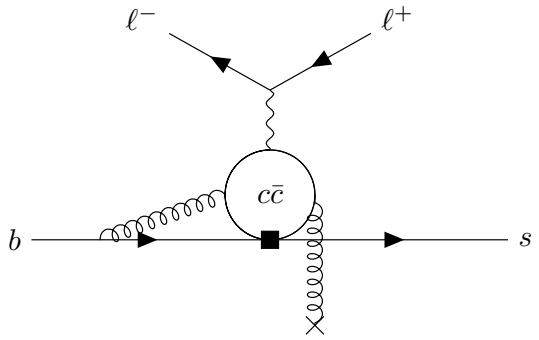
where G_F is the Fermi constant, $V_{tb,s}$ are elements of the CKM matrix for $b \to t$ and $t \to s$ transitions, $\mathcal{C}_i$ and $\mathcal{O}_i$ are the *Wilson coefficients* and operators as defined subsequently. This approach allows the separation of short-distance effects of the weak interaction from long-distance QCD effects. The former, describing the weak flavour-changing interaction, are described by the Wilson coefficients $\mathcal{C}_i$. The latter are described by the local operators $\mathcal{O}_i$, for the various contributions to the $b \to s\ell^+\ell^-$ transition such as scalar, vector and axial currents. The decay amplitude of $B^0 \to K^{*0}\mu^+\mu^-$, for example, is then given by

$$\mathcal{A}(B^0 \to K^{*0}\mu^+\mu^-) = \langle K^{*0}\mu^+\mu^- | \mathcal{H}_{\rm eff} | B^0 \rangle, \tag{5.4}$$

where the $B^0 \to K^{*0}$ matrix element can be separated out and is parameterised by hadronic form factors. Further detail of this framework can be found *e.g.* in Ref. [104].



(a) Schematic of effective $b \to s\ell^+\ell^-$ transition as a four-fermion interaction.



(b) Intermediate charm loop contribution to $b \to s\ell^+\ell^-$ processes, illustrating hard and soft gluon emission.

Figure 5.8: Diagrams of $b \to s\ell^+\ell^-$ transitions showing the effective field theory approach (left) and charm loop contributions (right).

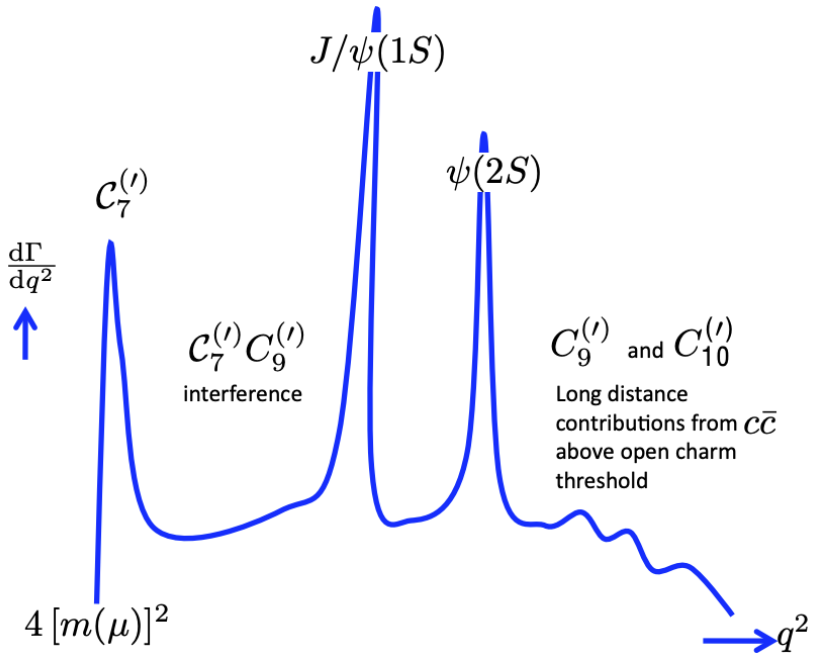


Figure 5.9: Sketch of the differential decay rate of $B^0 \to K^{*0}\mu^+\mu^-$ as a function of the dilepton invariant mass squared. Various regions are highlighted, and labelled with the corresponding dominant Wilson coefficient contributions. Figure from Ref. [126].

The full quark-level $b \to s\ell^+\ell^-$ transition is shown in Figure 5.6(a), while the effective theory description, as in Fermi theory [127], is displayed in Figure 5.8(a). For convenience, the Wilson coefficients are usually expressed as *effective* Wilson coefficients which account for combinations of $\mathcal{C}_i$. The dominant contributions are then simply given by $\mathcal{C}^{\text{eff}}_{7,9,10}$, and the equivalent right-handed currents represented in primed notation. Figure 5.9 displays the regions in q^2 over which the various contributions come into play. Relevant to this discussion are diagrams represented in Figure 5.8(a) which are mediated by vector ($\mathcal{C}_9$) and axial-vector ($\mathcal{C}_{10}$) currents. As indicated in Figure 5.9, in the high-q^2 region, long-distance $c\bar{c}$ states can interfere with these contributions. An intermediate charm loop state coupled to the lepton system via the virtual photon mimics a q^2-dependent contribution to $\mathcal{C}_9$, *i.e.*

$$\mathcal{C}_9^{\text{eff}} = \frac{4\pi}{\alpha_s}\mathcal{C}_9 + Y(q^2), \tag{5.5}$$

where $Y(q^2)$ contains charm loop contributions represented in Figure 5.8(b). The loop is not factorisable due to soft gluon emission, as illustrated in the figure, and the perturbative nature of these effects is not valid at high q^2 [77]. Therefore, these contributions are very difficult to estimate theoretically, and the current uncertainties assigned to them in the experimental measurements may be incorrect.

The part of the effective Hamiltonian in Equation 5.3 which describes non-local charm contributions can be expressed as

$$\mathcal{H}_{\text{eff}} = \mathcal{H}(0) + q^2\left[\sum_{J/\psi,\psi(2S),\ldots}(BW) + \int_{4m_D^2}^{\infty} \frac{\rho(s)}{s(s-q^2-i\epsilon)}ds\right], \tag{5.6}$$

where the first term is the isolated loop, the second term is a sum over resonant states modelled as Breit-Wigner functions (labelled BW), and the final term describes broad charm resonances and continuum states [77]. This is a dispersion relation, which can be used since the matrix element is analytic in q^2, and applies in the limit of $\epsilon \to 0$. The broad contributions, expressed in terms of a spectral density $\rho(s)$, are not known or measured, especially above the open charm threshold ($s > 4m_D^2$). Due to the width and overlapping nature of these broad states, such as the $\psi(3770)$, $\psi(4040)$, $\psi(4160)$, $\psi(4415)$, *etc.*, as well as non-resonant $c\bar{c}$ contributions, they cannot be excluded by vetoes in q^2 either. Measurements of the interference of the short- and long-distance amplitudes have been carried out using LHCb data in $B^+ \to K^+\mu^+\mu^-$ decays such as in Ref. [128]. However, the K^{*0} mode poses additional challenges because the different helicity states

of the vector meson can each have different relative phases. Additionally, the $K\pi$ S-wave contribution is being accounted for more carefully in recent measurements, *e.g.* Ref. [105]. Direct measurements of charm resonances above the open charm threshold in $b \to sc\bar{c}$ transitions can instead provide experimental constraints on the spectral density $\rho(s)$ [77, 120]. This is experimentally achievable through amplitude analyses of decays such as $B^0 \to D^0\overline{D}^0K^+\pi^-$, by extracting the $c\bar{c}$ components with the appropriate spin-parity. This is given by the $J^P = 1^-$ $D^0\overline{D}^0$ component, broken down into its various helicity contributions. Thus, the measurement and subsequent amplitude analysis of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays presented in this thesis will provide crucial input on charm loop contributions in existing and future measurements of $b \to s\ell^+\ell^-$ processes, as well as informing theoretical considerations of related New Physics models.

CHAPTER 6

ANALYSIS TECHNIQUES

A number of tools and techniques used at various stages of the analyses in this thesis are presented in this chapter. Section 6.1 describes a kinematic fit method used in LHCb, Section 6.2 presents the fit model used to describe the mass distribution of the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$, and Section 6.3 outlines the commonly used *sPlot* technique. Section 6.4 details machine learning techniques used for classification and regression tasks.

6.1 Kinematic fit

A kinematic fit is a procedure through which certain constraints are applied in order to improve the best estimate of track parameters of final state particles in a decay. This is accomplished by the DecayTreeFitter (DTF) package [129] in the LHCb software. There are two main types of constraints that can be applied, either on the mass or the vertex requirements of the particle. A mass constraint requires the decay products of a particle to form its nominal invariant mass. A vertex constraint requires a candidate to point towards its production vertex. In the case of the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$, a kinematic fit updates the four-momenta of the final state particles by applying constraints on the D^0 masses, and on the B^0 decay vertex. The main result of this fit is an improvement of the resolution of the reconstructed B^0 mass.

6.2 Fit model

A model is constructed to fit the mass distribution of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays in data. The signal shape is modelled as a double-sided Crystal Ball (DSCB) function [72], which consists of a Gaussian core and power-law tails. The combinatorial background is modelled by an exponential function, such that the full probability density function (PDF) is given by

$$N_S f_S(x) + N_B(\lambda e^{-\lambda x}), \tag{6.1}$$

where $N_{S,B}$ are the signal and background yields, and f_S is DSCB function, defined as

$$f_S(x;\alpha_L,\alpha_R,n_L,n_R,\mu,\sigma) = \begin{cases} \left(\frac{n_L}{|\alpha_L|}\right)^{n_L} e^{-\frac{|\alpha_L|^2}{2}} \left(\frac{n_L}{|\alpha_L|} - |\alpha_L| - \frac{x-\mu}{\sigma}\right)^{-n_L}, & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha_L \\ e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, & \text{for } -\alpha_L < \frac{x-\mu}{\sigma} < \alpha_R \\ \left(\frac{n_R}{|\alpha_R|}\right)^{n_R} e^{-\frac{|\alpha_R|^2}{2}} \left(\frac{n_R}{|\alpha_R|} - |\alpha_R| + \frac{x-\mu}{\sigma}\right)^{-n_R}, & \text{for } \frac{x-\mu}{\sigma} \geq \alpha_R\,. \end{cases} \tag{6.2}$$

The usual Gaussian mean and width are given by μ and σ, and the power law parameters are given by $n_{L,R}$ for the tails on the left and right of the mean. The boundary points between the Gaussian distribution and the power-law distributions on the left and right are given by $\alpha_{L,R}$, expressed as a number of standard deviations, σ. Both signal and background PDFs are normalised to the bounds of the mass distribution. For the branching fraction analysis presented in Chapter 7, the DSCB PDF and its analytic integral are manually implemented. For the amplitude analysis, the `zFit` [130] implementation is used, as presented in Chapter 9.

6.3 *sPlot*

The *sPlot* method [131] is a commonly used tool to statistically separate different contributions to a distribution in data. In this thesis, the usage is limited to the mass distribution, which consists of a signal and a background component. By fitting the mass with a PDF for both signal and background shapes, as outlined in Section 6.2, the *sPlot* technique calculates an event-by-event weighting for each component. The weights, called *s*Weights, are derived using the PDFs, the fitted signal and background yields, and the correlation matrix of the likelihood fit. The sum of the *s*Weights for any given event is equal to one, and weighting the data results in a signal only distribution. The *s*Weights

may be applied to distributions of events which are independent of the distribution used to obtain the weights, *i.e.* the mass distribution in this case.

This method is used to allow comparison between *s*Weighted data and simulation, which by construction contains only signal events. Most importantly, for the amplitude analysis in Chapter 9, the fit can be performed on *s*Weighted events such that there is only a single signal contribution in the amplitude fit.

6.4 Multivariate techniques

This section describes multivariate analysis (MVA) techniques used in this thesis. Such tools are widely used in particle physics, often in the form of a classifier to identify signal and background components in a data sample. A variety of algorithms are used, with the most common being Boosted Decision Trees (BDTs) [132, 133] and Artificial Neural Networks (ANNs, or simply NNs). In the analyses presented in Chapter 7 and 9, a neural network classifier is used as part of the event selection procedure in order to improve the signal purity. In this section, a theoretical introduction to neural networks is presented, along with the relevant tools and metrics employed in this analysis. A brief description of decision tree algorithms is also given, in the context of reweighting algorithms used in this analysis.

6.4.1 Neural Networks and Machine Learning methods

There are numerous references relating to machine learning and neural networks, however the discussion in this section is mainly based on Refs. [134, 135], and the further citations in the text.

The development of neural networks has been motivated by the superior capabilities of the human brain in areas such as pattern recognition, compared to conventional algorithms. The brain is able to swiftly perform complex and nonlinear computations using a system of neurons. Just as the brain develops through experiences or learning, Machine Learning (ML) models aim to recognise criteria based on learning in the form of training with data. A neural network mimics the concept of a brain by using simple processing units, or neurons, and weights connecting them, such as the synapses, to store and learn information. In the context of particle physics analyses, one of the main advantages of using a neural network is the fact that it can learn nonlinear relationships between many features, compared to cut-based selections for example, which may more

accurately describe the underlying physical relations. The following discussion focuses on using neural networks for classification tasks through supervised learning.

This analysis uses a feed-forward neural network, also known as a multi-layer perceptron (MLP). For a given set of inputs x_i, the output is given by

$$y_j = f_j\left(b_j + \sum_{i=1}^{m} w_{ji} x_i\right). \tag{6.3}$$

That is, the inputs are summed, weighted by a factor at each synapse or link between input and output, w_{ji}, added to a constant bias b_j, and subject to an activation function $f(\cdot)$. This calculation of the output from the inputs is referred to as a forward pass, and is displayed diagrammatically in Figure 6.1(a). As in Figure 6.1(b), if an additional layer is included between input and output, known as a hidden layer, these operations occur in sequence, such that the output becomes

$$y_k = f_k\left(b_k + \sum_j w_{kj} f_j\left(b_j + \sum_i w_{ji} x_i\right)\right). \tag{6.4}$$

This concept can be extended to many hidden layers, forming a fully-connected or Deep Neural Network (DNN), as used in this analysis. The architecture of the network, *i.e.* the number of layers (depth of the network) and the width of each layer, can be tuned depending on the input and feature dimensions. Similarly, other hyperparameters of the network such as the starting value of the bias should also be tuned accordingly. These quantities are known as hyperparameters to distinguish from the real parameters of the network, which are the weights w_{ji} and biases b_j.

The activation function is a element-wise operation, and there are many functions that can be used depending on the problem at hand. In a classification task, the final output is usually produced by a sigmoid activation function, defined as

$$\sigma(x) = \frac{1}{1+\mathrm{e}^{-x}}, \tag{6.5}$$

which varies smoothly from 0 to 1. For the hidden layers, the optimal activation function can be tuned, and in this analysis the Exponential Linear Unit (ELU) is used. This is defined as

$$f(\alpha, x) = \begin{cases} x, & x >= 0 \\ \alpha(\mathrm{e}^x - 1), & x < 0. \end{cases} \tag{6.6}$$

Neural networks learn via a technique known as back-propagation [136, 137]. The network is provided a training sample which specifies the correct output for every

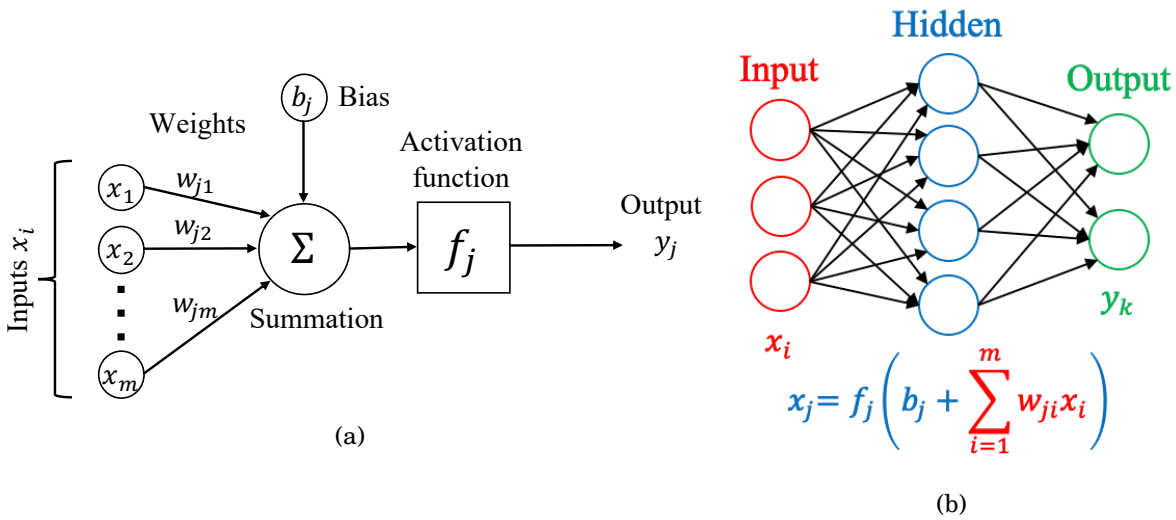


Figure 6.1: Diagrammatic representations of a) a forward pass in a single neuron and b) a fully-connected feed-forward neural network with one hidden layer.

input. When the network trains, there is an iterative calculation of the error, *i.e.* the difference, between its output and the true value. This is quantified by a loss function L, for example a mean squared error. The gradient of this function ∇L can then be computed with respect to each weight and bias. This information is propagated back through the network and the parameters adjusted by an amount given by the learning rate, which determines the size of the step along the gradient. This repetitive process eventually minimises the loss function. With the increased use of neural networks and other ML methods, including an increase in data sizes, this procedure is optimised with Stochastic Gradient Descent (SGD) [138]. Rather than calculating the full gradient at every step, it is estimated from a random subset of the data. This can be further optimised using algorithms, such as Adam which is used in this analysis, which adapt the learning rate according to previous gradient calculations [138]. During training, the network will see the full dataset multiple times, known as the number of training epochs. It is possible to update the learning rate at the start of each epoch so that the steps gradually decrease in size as the minimum is approached. This results in faster training times and higher performance.

6.4.1.1 Enhancing training and performance metrics

Many further techniques are employed in order to achieve the best training performance and safeguard against overfitting. During training, a subset of the data is kept aside as a

testing or validation sample. At every iteration, the loss is evaluated on the test sample in order to compare the results with the training sample. In the case of overfitting, after the initial period of convergence, the loss of the test sample will begin to increase. This is when the network is too closely fitted to the training sample, to the extent that it learns to fit to the inherent statistical noise in the sample. A common method to prevent this, and is also employed here, is Dropout [139]. This is when a number of processing units and their links are randomly dropped from the network. The rate of dropout can be tuned along with a weight constraint, such as max-norm, where an upper bound is set on the magnitude of the weight vector.

In general, a larger training sample will result in better performance of the NN. In order to not lose out on the potential training power of the separate test sample, k-fold cross-validation [140] may be used. Setting $k = 10$, the data is split into 10 subsets where 9 are used for training and 1 for testing. This is done recursively so that all possible data is used for both training and validation. The final model can then be chosen as the average result across all folds, or the one with the best performance as long as overfitting is not observed. An additional hyperparameter that can be optimised is the batch size, which defines the number of samples of the training set that are seen by the network before the parameters (weights and biases) are updated. The batch size represents a balance between the frequency of updating the weights and the accuracy of the gradient estimate.

To quantify the performance of a network, one can use the accuracy during training, which is the percentage of predicted values that match the true values. This is also calculated on the validation sample. A more meaningful metric is the receiver operating characteristic (ROC) curve [141, 142]. In a binary classification problem, the outcome can either be positive or negative, corresponding to labels 1 and 0, respectively. Thus one can define the true positive rate (TPR), also called the recall, as the number of correctly classified positives divided by the total positives in the sample. Similarly, the false positive rate (FPR) is the number of incorrectly classified negatives divided by the total number of negatives. A ROC graph plots the true positive rate against the false positive rate, as demonstrated in Figure 6.2. A discrete classifier applied on a validation sample will produce a single point in ROC space given by (FPR, TPR). Therefore anything along the diagonal $y = x$ in Figure 6.2, *i.e.* equal expectation of true or false positive, is equivalent to a random guess. In the case of a neural network, the output varies between 0 and 1, which can be thought of as a threshold above which the result is positive and below which it is negative. Thus a ROC curve can be generated by scanning across all

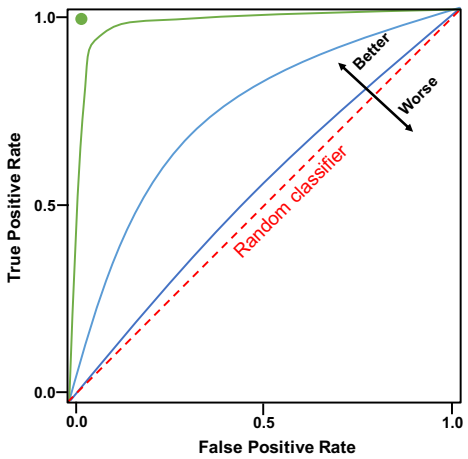


Figure 6.2: Example of a ROC graph, showing true positive rate against false positive rate, for various classifiers. The red dashed line represents a random classifier, with an AUC of 0.5. The curves with increasing areas represent classifiers with increasing performance. The green circle at (0, 1) represents a perfect classifier.

possible thresholds and building up a set of points over the ROC space. An ideal classifier would simply be a line joining (0, 0), (0, 1) and (1, 1). In such a case, the area under the curve (AUC) is equal to unity. Therefore the area under the ROC curve, or ROC AUC, can be used as a diagnostic metric, where the closer the value to one, the better the performance of the classifier.

6.4.2 Gradient boosted corrections

Decision trees are a form of multivariate algorithm that involve a sequence of steps, represented as leaf nodes in a tree, that, in the case of regression, seek to accurately predict a target continuous variable. Gradient boosting [143] involves an iterative process where poorly estimated events are weighted more highly and used to train another tree, thereby building up an ensemble, or forest, of boosted decision trees (BDTs). In this thesis, a gradient boosted *reweighting* algorithm is employed, described in Section 7.5.2. In this context, reweighting refers to correcting a sample by finding weights which make it look like a target sample. The common use in particle physics is to correct simulation to match (*s*Weighted) data.

Just as the neural networks described above, the reweighting algorithm has hyperparameters which can be optimised. The main hyperparameters of interest are the number of trees and the maximal depth of each tree. The number of trees defines the number of iterations of boosting to perform. The depth of the trees defines the number of nodes in

each tree. The optimal set of hyperparameters is determined by applying a simple BDT on the derived weights and the target data *s*Weights, and selecting the set that produces a ROC AUC closest to 0.5. This signifies that the discriminating BDT cannot separate the derived weights and the target, so the reweighting algorithm has successfully reproduced the target.

CHAPTER 7

BRANCHING FRACTION MEASUREMENT OF $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ DECAYS

This chapter describes the branching fraction (BF) ratio measurement of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays with respect to the $B^0 \to D^{*-} D^0 K^+$ mode, utilising data collected by the LHCb experiment in 2011, 2012 (Run 1) and 2016. The results have been published in Ref. [1].

Section 7.1 describes the analysis strategy and Section 7.2 summarises the data samples used in this analysis. In Sections 7.3 and 7.4 the selection procedure is described for both decay modes. Section 7.5 specifies the Monte Carlo (MC) simulation samples used in this analysis, and Section 7.6 details the calculation of various selection efficiencies. Section 7.7 describes the treatment of backgrounds. In Section 7.8, the mass fits across both modes and data-taking periods are presented. Systematic uncertainties are described in Section 7.9. Finally Section 7.10 details the calculation of the ratio of branching fractions and the measurement of $\mathcal{B}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-)$.

7.1 Analysis strategy

This analysis makes the first observation of the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$, referred to as the signal mode. The decay $B^0 \to D^{*-} D^0 K^+$ is referred to as the control mode, and is chosen because it has a similar topology to the signal mode. A kinematic diagram of this topology is shown in Figure 5.4, and the leading Feynman diagrams are shown in

Section 5.1. The control mode has a known BF of $(2.47 \pm 0.21) \times 10^{-3}$ [84]. This is used as the normalisation factor in the branching fraction measurement of the signal mode, in order to reduce the impact of systematic uncertainties. After the decay of the D^{*-} meson via the strong interaction, the signal and control modes present the same final-state particles $D^0 \overline{D}^0 K^+ \pi^-$, with the D^0 mesons decaying as $D^0 \to K^- \pi^+$. Throughout this analysis, the charge conjugate process is included unless otherwise stated.

It is worth noting that the LHCb collaboration has recently performed a measurement of the ratio of branching fractions [85]

$$\frac{\mathcal{B}(B^0 \to D^{*-} D^0 K^+)}{\mathcal{B}(B^0 \to D^- D^0 K^+)} = 1.754 \pm 0.028 \pm 0.016 \pm 0.035, \tag{7.1}$$

where the first uncertainty is statistical, the second systematic, and the third is due to the uncertainties on the D-meson branching fractions. However, the current precision on the branching fraction of the decay $B^0 \to D^0 D^- K^+$ [2]

$$\mathcal{B}(B^0 \to D^0 D^- K^+) = 1.07 \pm 0.07 \pm 0.09, \tag{7.2}$$

does not yet allow for a more precise measurement of $\mathcal{B}(B^0 \to D^{*-} D^0 K^+)$.

The signal decay is reconstructed as $B^0 \to D^0 \overline{D}^0 (K\pi)_{K^{*0}}$, however only a loose kinematic requirement is applied on $m(K\pi)$ (see Section 7.3). The data is assumed to contain $K^*(892)^0$, as well as S- and D-wave $K\pi$, and a possible reflection from the decay $B^0 \to D_s^{*+} D^0 \pi^-$. Further expected resonance states are discussed in the results (Section 7.10), and will be confirmed by a full amplitude analysis study which is the topic of Chapter 9. Thus, the full kinematically-allowed range, outside the D^{*-} resonance, is considered. For labelling purposes, the kaon and pion from the B^0 decay are referred to as K_{K^*} and π_{K^*} respectively.

First, the ratio of branching fractions is measured as

$$\mathcal{R} = \frac{\mathcal{B}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-)}{\mathcal{B}(B^0 \to D^{*-} D^0 K^+)}, \tag{7.3}$$

and then the known BF of the control mode is used to calculate $\mathcal{B}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-)$. The branching fraction of the signal mode is defined in terms of the number of signal candidates $\mathcal{N}$ as follows:

$$\mathcal{B}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-) = \frac{\mathcal{N}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-)}{\mathcal{L} \times \sigma_{b\bar{b}} \times 2 \times f_d \times \epsilon^{\text{sig}} \times \mathcal{B}(D^0 \to K\pi)^2}. \tag{7.4}$$

Similarly for the control mode,

$$\mathcal{B}(B^0 \to D^{*-} D^0 K^+) = \frac{\mathcal{N}(B^0 \to D^{*-} D^0 K^+)}{\mathcal{L} \times \sigma_{b\bar{b}} \times 2 \times f_d \times \epsilon^{\text{cont}} \times \mathcal{B}(D^{*-} \to \overline{D}^0 \pi^-) \times \mathcal{B}(D^0 \to K\pi)^2}, \tag{7.5}$$

where ϵ^{mode} is the overall selection efficiency of the relevant decay mode. The other parameters used in Equation 7.4 and Equation 7.5 are defined as follows:

- $\mathcal{L}$ is the integrated luminosity, detailed in Section 7.2 for the different years of data-taking in LHCb.
- $\sigma_{b\bar{b}}$ is the cross-section of $b\bar{b}$ production in pp collisions at the relevant centre-of-mass energy for the different data-taking runs in LHCb [144].
  - $\sigma_{b\bar{b}} = 295\mu b$ for Run 1;
  - $\sigma_{b\bar{b}} = 560\mu b$ for Run 2.
- f_d is the probability for a b-quark to hadronise into a B^0 meson, estimated as 40% [2]. A factor of 2 is included to account for $\overline{B}^0$ mesons.
- The branching fraction $\mathcal{B}(D^0 \to K^-\pi^+) = \mathcal{B}(\overline{D}^0 \to K^+\pi^-) = (3.93 \pm 0.03) \times 10^{-2}$ [2].
  - This BF has been updated to $(3.999 \pm 0.006 \pm 0.031 \pm 0.032)\%$ [145] to account for improvements in the modeling of final state radiation. Nevertheless, the stated value is used as no effect is seen on the optimal multivariate classifier cut point where the value is utilised, see Section 7.4.
- The branching fraction $\mathcal{B}(D^{*-} \to \overline{D}^0\pi^-) = 0.667 \pm 0.005$ [2].

Substituting Equations 7.4 and 7.5 into Equation 7.3 gives the final ratio to be measured as

$$\mathcal{R} = \frac{\mathcal{N}(B^0 \to D^0\overline{D}^0K^+\pi^-)}{\mathcal{N}(B^0 \to D^{*-}D^0K^+)} \times \mathcal{B}(D^{*-} \to D^0\pi^-) \times \frac{\epsilon^{\text{cont}}}{\epsilon^{\text{sig}}}. \tag{7.6}$$

Therefore, the signal branching fraction can be extracted by measuring this ratio and multiplying it by the known control channel branching fraction. Data and simulation from the control channel are also used to verify and derive various corrections, detailed later.

7.2 Data samples

This analysis utilises data from pp collisions collected by the LHCb experiment in two data-taking periods. The first period, referred to as Run 1, includes $1.0\,\text{fb}^{-1}$ of data collected in 2011 at a centre-of-mass energy of 7 TeV, and $2.0\,\text{fb}^{-1}$ of data collected in

2012 at $\sqrt{s} = 8\,\text{TeV}$. The second period includes $1.7\,\text{fb}^{-1}$ of data collected in 2016 at $\sqrt{s} = 13\,\text{TeV}$, which is a subset of the full Run 2 dataset. The simulation samples used in this analysis are described in Section 7.5. The data and simulation have several stages of selections applied, detailed in Sections 7.3 and 7.4.

7.3 Event selection

This section describes selections made to identify $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ and $B^0 \to D^{*-} D^0 K^+$ candidates. The first part of this procedure relies on cut-based selections described in Sections 7.3.1 and 7.3.2, which aim to significantly reduce the size of the datasets whilst retaining a high fraction of signal candidates. Further so-called *preselections* are also applied to reduce combinatorial background events, described in Section 7.3.3. The combinatorial background consists of events where one or more of the particles in the decay is combined with random particles from other b-hadron or c-hadron decays, to mimic a signal candidate. Next, specific background processes which can be present in the signal data are removed or estimated through a variety of different methods, outlined in Section 7.7. Finally a multivariate analysis (MVA), where a classifier is trained to distinguish signal and background events, is used to further reduce the combinatorial background, see Section 7.4.

Variables used in selection decisions and elsewhere in this analysis are described below, using the abbreviations defined at the start of this thesis.

- χ^2_{track} is the χ^2/ndf of the fit to the track associated to the particle, using a Kalman filter [33, 146].

- χ^2_{IP} (PV) is the significance of the impact parameter with respect to the primary vertex. A larger value implies the track is more likely to not originate from the pp interaction point, *i.e.* it is not a prompt track.

- $Ghost_{\text{track}}$ is the probability of the track being fake. A ghost track is produced by an incorrect combination of hits that do not originate from a single particle. The algorithm used for this probability determination is described in Ref. [147].

- $X_Y\,\text{DLL}_{K\pi}$ is the $\text{DLL}_{K\pi}$ of the particle X from the parent particle Y (and similar for `ProbNN` variables).

- IP_{PV} is the value of the impact parameter with respect to the primary vertex.

- χ^2_{vtx}/ndf is the quality of the fit of the reconstructed decay vertex of the particle.
- $\chi^2_{\text{vtx-PV}}$ is the significance of the distance between the reconstructed decay vertex and the associated PV. A larger value relates to a greater displacement from the PV.
- DIRA_{PV} is the cosine of the angle between the momentum of the particle, and the vector joining its vertex to the PV.
- τ_{PV} is the lifetime of the particle with respect to the PV.
- `BBDT` is a Bonsai-boosted Decision Tree for B decays to charm, see Section 3.5.1.
- `nTracks` is the number of long tracks, *i.e.* tracks reconstructed downstream of the magnet. Setting an upper limit on this reduces the number of high-multiplicity events.
- m_{PDG} is the nominal mass of a particle, where the value is taken as the world average reported in Ref. [2].

7.3.1 Stripping

Stripping describes a number of selections relating to the kinematic, PID, and track-quality properties of candidates. The $B^0 \to D^0\overline{D}^0K^+\pi^-$ and $B^0 \to D^{*-}D^0K^+$ candidates are subject to a dedicated set of selections, or *stripping line*, for the specific topology and kinematics of these decays. The cuts applied in this stripping line are summarised in Table 7.1 and 7.2, where the variables are defined in detail above. The first relates to the final-state kaons and pions, and the second relates to the intermediate particles and the B^0 meson. In most cases, the selections are identical across run periods, and the few differences are indicated as such. The stripping also includes basic selections based on the high-level trigger, not detailed here.

7.3.2 Trigger

The trigger selection requires the high-level triggers firing on the signal B^0 candidate. Descriptions of these trigger lines mentioned in this section can be found in Section 3.5. For HLT1 in Run 1, events are triggered by the `Hlt1TrackAllL0Decision` line, and at HLT2, the `Hlt2Topo{2,3,4}BodyBBDTDecision` lines which use a decision tree to reduce combinatorial background. In 2016, the equivalent lines in HLT1 are `Hlt1TrackMVADecision`

Table 7.1: Stripping selections applied on the final state particles. Values in brackets represent changes made for the 2016 period.

Particle	Cut
$(\pi,K)_{D^0,\bar{D}^0,K^{*0}}$	$\chi^2_{\text{track}} < 3.0\,(4.0)$
$(\pi,K)_{D^0,\bar{D}^0,K^{*0}}$	$p_{\text{T}} > 100\,\text{MeV}/c$
$(\pi,K)_{D^0,\bar{D}^0}$	$p > 1000\,\text{MeV}/c$
$(\pi,K)_{K^{*0}}$	$p > 2000\,\text{MeV}/c$
$(\pi,K)_{D^0,\bar{D}^0,K^{*0}}$	$\chi^2_{\text{IP}}(\text{PV}) > 4$
$(\pi,K)_{D^0,\bar{D}^0,K^{*0}}$	$Ghost_{\text{track}} < 0.4$
$\pi_{D^0,\bar{D}^0}$	$\text{DLL}_{K\pi} < 20\,(10)$
$K_{D^0,\bar{D}^0}$	$\text{DLL}_{K\pi} > -10\,(-5)$
at least 1 final-state particle	$\chi^2_{\text{track}} < 2.5$
at least 1 final-state particle	$p_{\text{T}} > 500\,\text{MeV}/c$
at least 1 final-state particle	$p > 5000\,\text{MeV}/c$
at least 1 final state	$p_{\text{T}} > 1.7\,\text{GeV}/c$
at least 1 final state	$p > 10\,\text{GeV}/c$
at least 1 final state	$\text{IP}_{PV} > 0.1\,\text{mm}$

Table 7.2: Stripping selections applied on the intermediate particles and B^0 candidate.

Particle	Cut
$D^0,\bar{D}^0$	$p_{\text{T}} > 1800\,\text{MeV}/c$
K^{*0}	$p_{\text{T}} > 1000\,\text{MeV}/c$
$D^0,\bar{D}^0$	$M \in [m_{\text{PDG}}(D^0) \pm 100\,\text{MeV}/c^2]$
$D^0,\bar{D}^0,K^{*0}$	$\text{DOCA}_{K\pi} < 0.5\,\text{mm}$
$D^0,\bar{D}^0$	$\chi^2_{\text{vtx}}/\text{ndf} < 10$
K^{*0}	$\chi^2_{\text{vtx}}/\text{ndf} < 16$
$D^0,\bar{D}^0$	$\chi^2_{\text{vtx-PV}} < 36$
K^{*0}	$\chi^2_{\text{vtx-PV}} < 16$
$D^0,\bar{D}^0$	$\text{DIRA}_{\text{PV}} > 0$
B^0	$M \in [4750, 6000]\,\text{MeV}/c^2$
B^0	$p_{\text{T}} > 5000\,\text{MeV}/c$
B^0	$\chi^2_{\text{vtx}}/\text{ndf} < 10$
B^0	$\tau_{\text{PV}} > 0.2\,\text{ps}$
B^0	$\chi^2_{\text{IP}}(\text{PV}) < 25$
B^0	$\texttt{BBDT} > 0.05$
Event	$\texttt{nTracks} < 500$

and `Hlt1TwoTrackMVADecision`, and the HLT2 lines are renamed to `Hlt2Topo{2,3,4}BodyDecision`.

Events are then split into two categories, TOS and TIS, based on the hardware level trigger, L0. Trigger-On-Signal (TOS) is defined by any of the final-state particles firing the `L0HadronDecision` line, except for $\pi_{K^{*0}}$, which is excluded due to possible differences between the kinematics of the $\pi_{K^{*0}}$ in the signal and control modes. Trigger-Independent-of-Signal (TIS) is defined by the `L0HadronDecision` or `L0MuonDecision` lines firing independently of the signal B^0 candidate. This is defined exclusively to TOS, *i.e.* by requiring the coincidence with the `L0HadronDecision` line *not* firing due to a signal B^0 candidate.

7.3.3 Offline selections

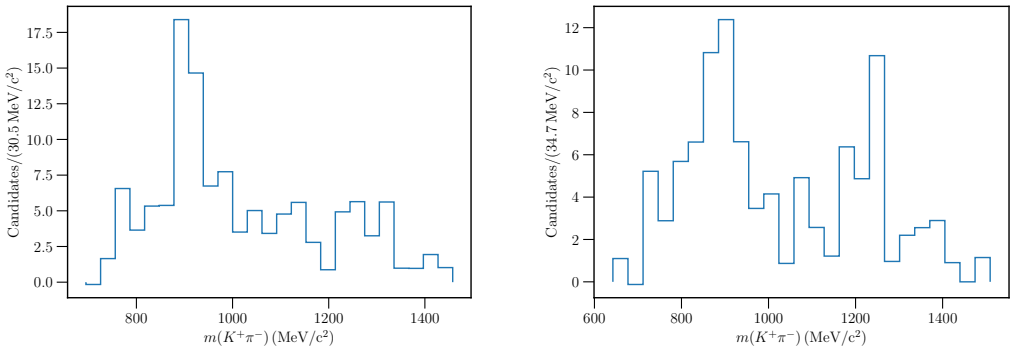
Further cuts, referred to as preselections, are applied before training a classifier for background reduction. The kinematic fit, as described in Section 6.1, with the D^0 masses constrained, is required to have converged. A loose requirement is applied on the K^{*0} mass, $m(K\pi) < 1600\,\text{MeV}/c^2$, where $m(B^0) - 2m(D^0) + 50\,\text{MeV}/c^2 = 1600\,\text{MeV}/c^2$; the data does not contain a purely P-wave sample, see Figure 7.1, so a cut is not applied around the $K^*(892)^0$ mass. The significance of the flight distance of the D mesons is required to be > 0, which is used to fight charmless backgrounds, see Section 7.7. The D^0 meson candidates are selected within a window of $\pm 30\,\text{MeV}/c^2$ around their nominal masses. Loose PID selections are applied as $\pi_{K^{*0}}\text{DLL}_{K\pi} < 10$ and $K_{K^{*0}}\text{DLL}_{K\pi} > -10$. An additional cut is applied on the ProbNN variable $K_{K^{*0}}\,\texttt{ProbNN}K > 0.05$ in order to reduce mis-modelling of data by MC in this region. Finally, a cut is placed on the vertex quality of the $K\pi$ system, requiring the χ^2 of the reconstructed vertex to be < 8.

The data is separated into the signal and control mode by specifying the control mode with the requirement

$$\Delta_{D^*} < (4 \times 0.724)\,\text{MeV}/c^2, \tag{7.7}$$

$$\Delta_{D^*} = |m(D_{D^*}\pi) - m(D_{D^*}) - [m_{PDG}(D^{*-}) - m_{PDG}(D^0)]|, \tag{7.8}$$

where D_{D^*} is the D^0 or $\overline{D}^0$, depending on the sign of the pion. This ensures that the π^- and the $\overline{D}^0$ (π^+ and D^0 for the conjugate mode) combine to give a mass that is within 4σ of the reconstructed D^{*-} (D^{*+}) mass. Here σ is found by fitting a Gaussian to the invariant mass spectrum of the D^{*-} decay products. The signal mode is defined with the inverted requirement. The signal mode also has vetoes applied to combat certain peaking backgrounds, see Section 7.7.



(a) Run 1 (b) 2016

Figure 7.1: Invariant mass distributions of $K^+\pi^-$ for sWeighted data in the signal region ($|m_{B^0} - m_{\text{PDG}}| < 50\,\text{MeV}/c^2$), after all selections and application of the Neural Network classifier (see Section 7.4) for the TOS category. a) shows Run 1 and b) shows 2016.

The invariant mass spectra of the B^0 candidates out of the stripping and after these selections are shown in Figure 7.2, as well as the subtracted mass quantity Δ_{D^*} defined in Equation 7.8.

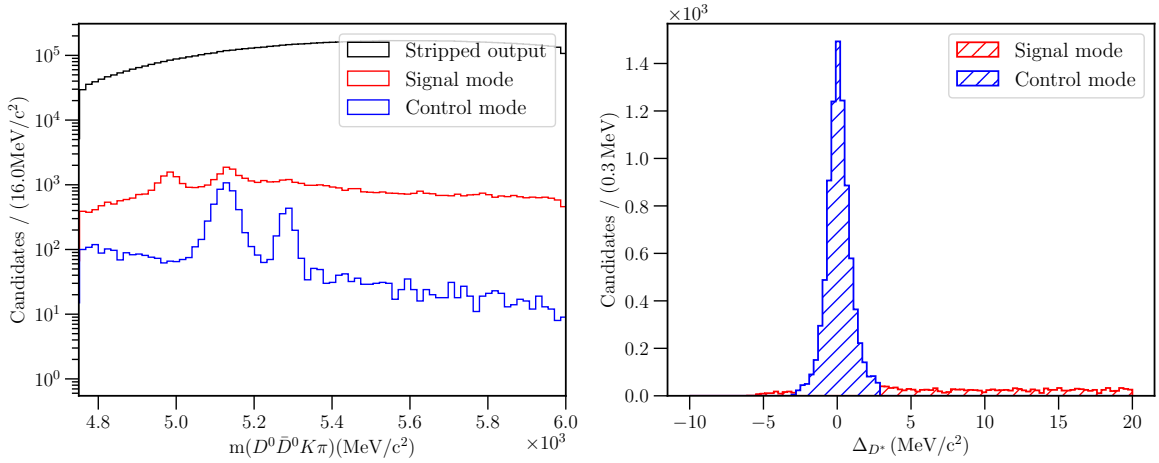
7.3.3.1 Multiple candidates and further checks

After all selections, including the classifier described in the next Section 7.4, a few events $\sim 2\%$ remain with multiple candidates. Within each event, one candidate is chosen at random, and the rest are removed. Moreover, within each candidate there is a small chance that tracks are clones, for example if the bachelor kaon and pion are in fact the same track. It is confirmed this does not occur by examining the opening angle between all possible pairs of particles, where the opening angle θ between particles labelled a, b is defined as

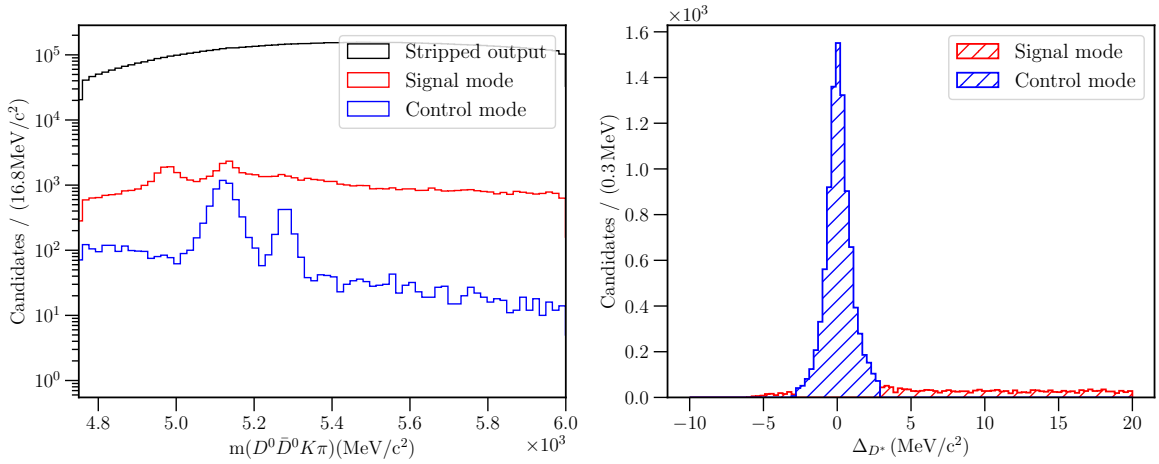
$$\cos\theta = \frac{p_x^a p_x^b + p_y^a p_y^b + p_z^a p_z^b}{p^a p^b}, \tag{7.9}$$

and no peaks are seen at very low values of the opening angle.

A further check is carried out to ensure that the wrong tracks are not combined to form the D^0 mesons, by checking the various two-body invariant mass combinations. The probability of this occurring is already reduced due to the D^0 mass windows and the requirement on $m(K\pi)$, and it is verified that no peaks are seen in the wrong combination of $K\pi$ pairs.

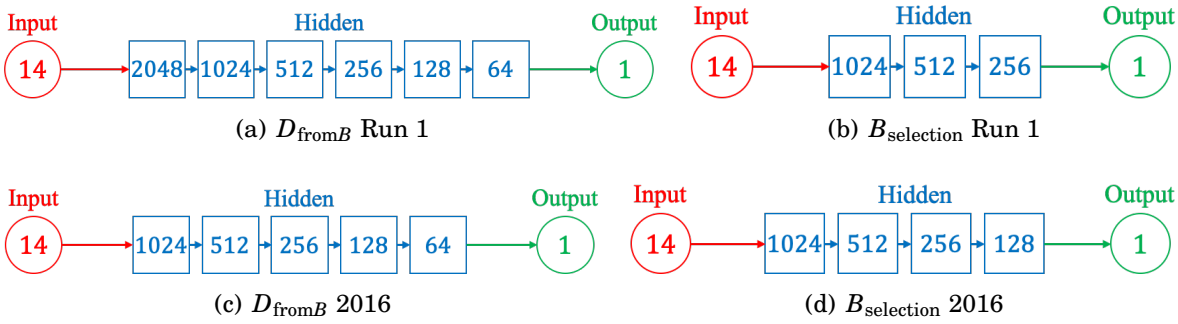


(a) Run 1



(b) 2016

Figure 7.2: On the left, the invariant mass distribution of B^0 meson after the stripping selections (black), and after the offline selections for $B^0 \to D^0\bar{D}^0K^{*0}$ (red) and $B^0 \to D^{*-}D^0K^+$ (blue). On the right, the requirement on the invariant mass distribution of Δ_{D^*} (Equation 7.8), zoomed in to $< 20\,\text{MeV}/c^2$ to show the disentanglement of the signal and control modes. a) shows Run 1 data and b) shows 2016 data.



(a) $D_{\text{from}B}$ Run 1 (b) $B_{\text{selection}}$ Run 1

(c) $D_{\text{from}B}$ 2016 (d) $B_{\text{selection}}$ 2016

Figure 7.3: Architectures of the networks for each stage and data-taking period, as indicated in the subcaptions. The architecture is kept constant for each trigger category within the subsamples.

7.4 Classifier selection

This section describes the use of MVA techniques, namely a neural network classifier, used in the event selection procedure of this analysis. The optimisation and training of the classifier is laid out, followed by the method used to choose the optimal cut value.

7.4.1 Classifier training

A neural network based on the MLP model described in Section 6.4 is used to further reduce combinatorial background after the preselections. It is implemented in Keras [148] using TensorFlow [149] as the backend. The architectures of the various NNs are shown in Figure 7.3, where the final output from a sigmoid activation function results in a response between 0 and 1, corresponding to background-like and signal-like events respectively. The starting width of the first hidden layer is given by 2^n where n decreases by 1 at each subsequent layer, and the total number of hidden layers, is optimised. All such hyperparameters are tuned via a comprehensive grid-search of the space to achieve the highest accuracy possible, whilst accounting for overfitting. The number of epochs is set large enough that the curves of accuracy and loss are sufficiently flat by the end of training. The learning rate is halved every ten epochs. The weight constraint is also tuned for a constant rate of dropout. The training uses k-fold cross-validation with $k = 10$. The model from the training fold with the best performance is saved to be applied to data afterwards, which can be done in the case of no overfitting, discussed below.

There are two classifiers trained in stages, with the first stage aiming to select D^0 mesons originating from a B^0 called $D_{\text{from}B}$, using kinematic and topological variables related to its decay products. These classifiers for both the D^0 and $\overline{D}^0$ (with no cuts

placed) are then fed into a second stage along with additional variables aiming to select B^0 candidates, called $B_{\text{selection}}$. This staged approach is employed because the two stages aim to combat different types of background. The $D_{\text{from}B}$ classifier aims to reject prompt charm candidates, *i.e.* D^0 mesons that originate from the pp collision rather than a B^0 decay. The $B_{\text{selection}}$ classifier then focuses on combinatorial B^0 background which, as mentioned, consists of B^0 candidates formed by a random combination of particles rather than the specific signal decay.

Training is carried out separately for the trigger categories TOS and TIS, as well as being separated into the two stages and the two data-taking periods, such that there are 8 classifiers overall. The signal mode is used for training, and the trained classifier is then applied to both decay modes. The variables used in the classifiers are listed in Table 7.3 and Table 7.4. In both cases there are fourteen input variables or features, which determine the input dimension of the network as in Figure 7.3. The particle identification variables are transformed as described in Section 7.5.1. The distributions of the training variables are shown in Figure 7.4 and 7.5 for the $D_{\text{from}B}$ and $B_{\text{selection}}$ classifiers, respectively. These distributions are shown only for Run 1 in the TOS category for illustration purposes.

The signal proxy sample used to train the classifiers consists of simulated data after all the selections described in Section 7.3. The simulated events have corrections applied as detailed in Section 7.5, including a transformation of the particle identification variables, and an event-by-event weighting. For the $D_{\text{from}B}$ classifier, the background proxy sample is taken from data in the D^0 sidebands ($|m(D^0) - m_{\text{PDG}}(D^0)| > 40\,\text{MeV}/c^2$) and all other preselections applied except with a looser window on the $\overline{D}^0$ of $40\,\text{MeV}/c^2$ and an additional requirement on the B^0 signal region ($|m(B^0) - m_{\text{PDG}}(B^0)| < 100$). For the $B_{\text{selection}}$ classifier, the background proxy sample is data in the B^0 upper mass sideband ($m(B^0) - m_{\text{PDG}}(B^0) > 200\,\text{MeV}/c^2$), with all preselections applied except for the D^0 and $\overline{D}^0$ mass windows. In some cases, there is an imbalance between the signal and background samples by a significant factor that reduces performance as one sample is favoured during training. One method for combating this is to use the SMOTE algorithm [150] to over-sample the lesser dataset. Additionally, a constant weight can be applied to one of the classes (signal or background) during training. The $D_{\text{from}B}$ classifier in Run 1 (TOS and TIS) has a factor ~ 15 imbalance, *i.e.* the background sample is ~ 15 times smaller than the signal sample. Therefore SMOTE is used to over-sample by a factor of 5, and a training weight of 3 is applied to the background sample. The $D_{\text{from}B}$ classifier in 2016, and the $B_{\text{selection}}$ classifier in Run 1 have negligible imbalances in both trigger categories.

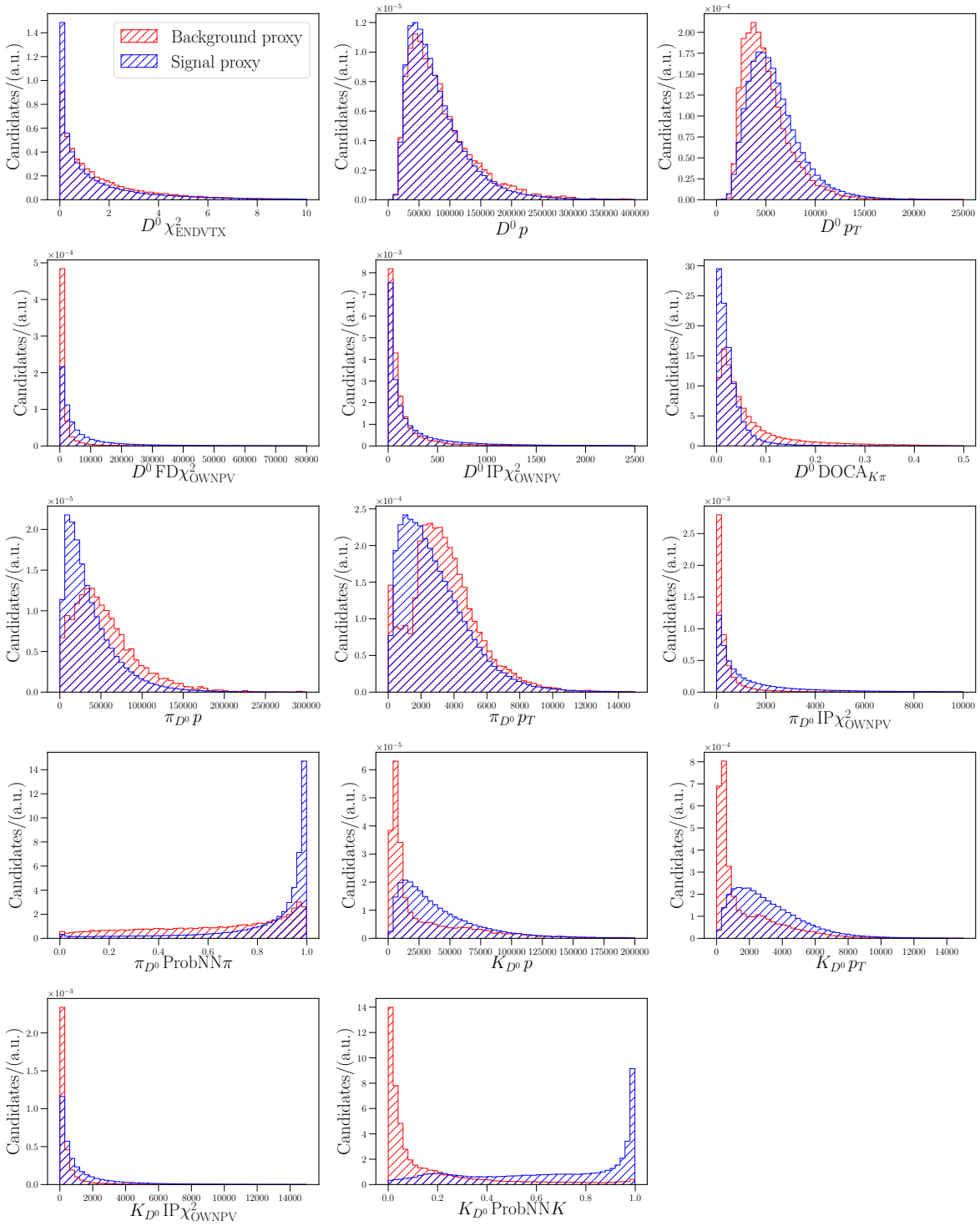


Figure 7.4: Distributions of signal and background proxy samples used to train D_{fromB}, shown here for Run 1 TOS only.

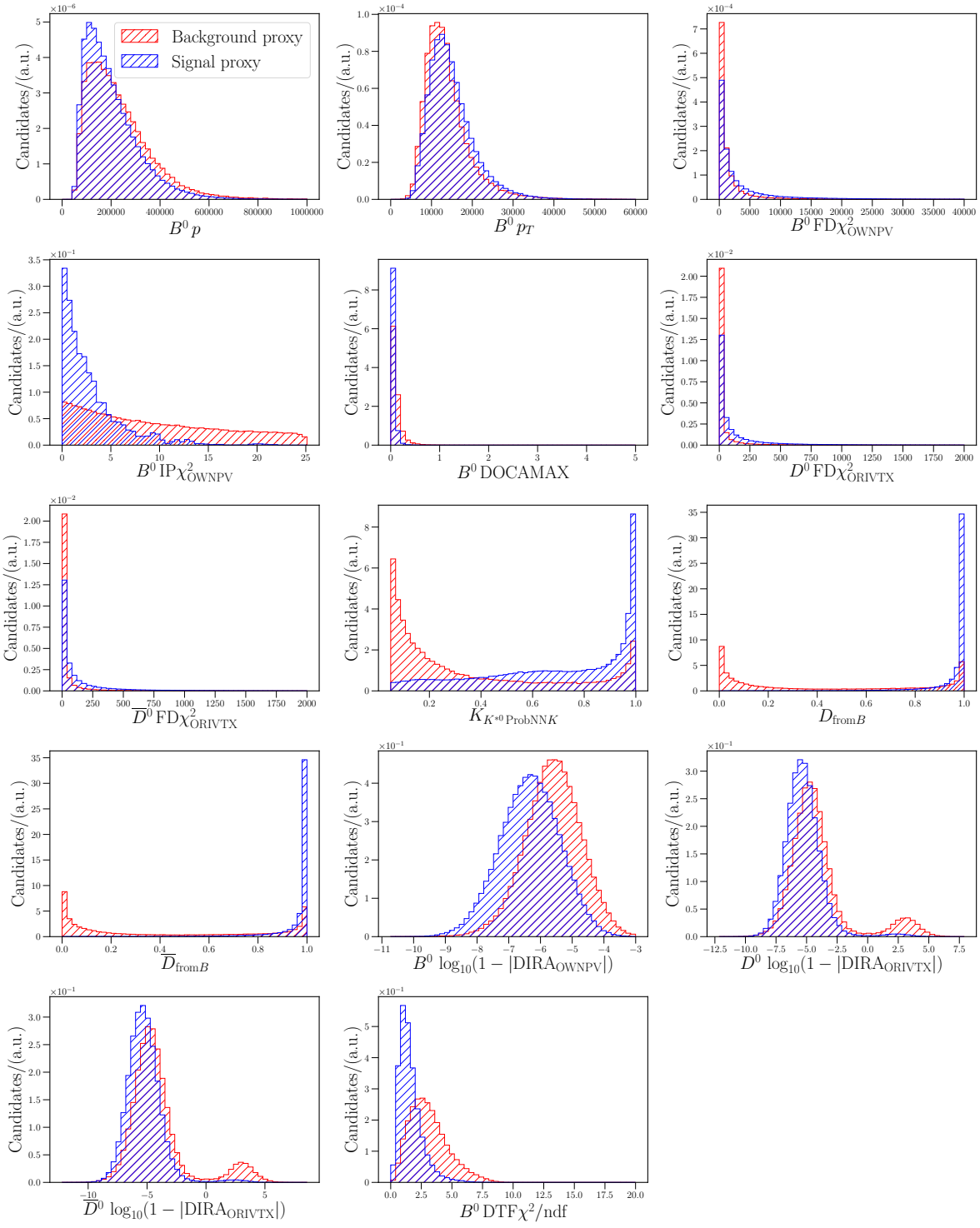


Figure 7.5: Distributions of signal and background proxy samples used to train $B_{\text{selection}}$, shown here for Run 1 TOS only.

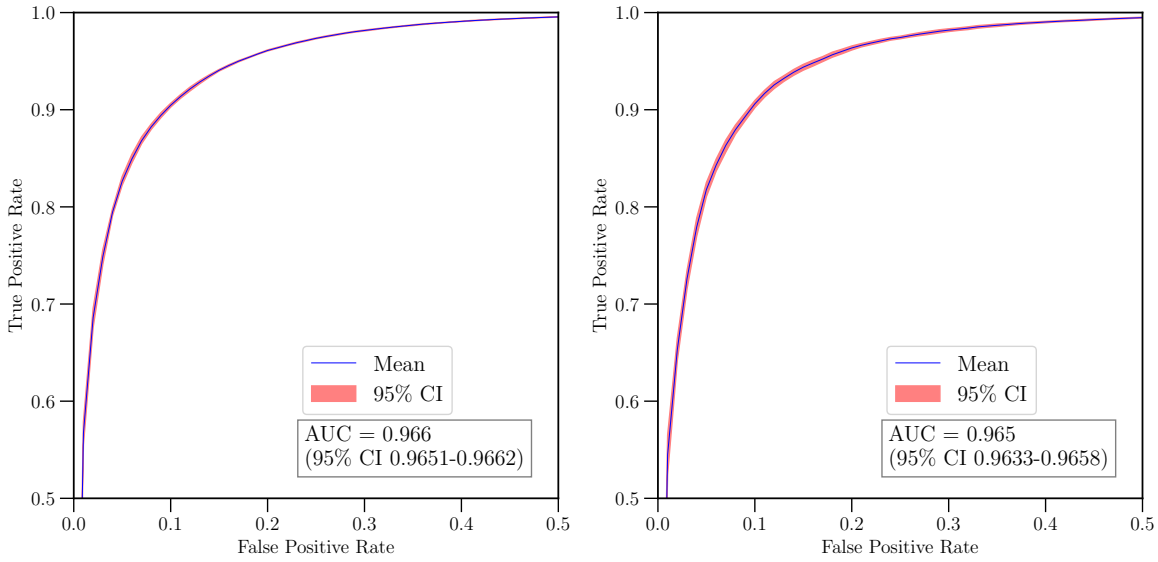
Table 7.3: $D_{\text{from}B}$ input variables. The subscript OWNPV means the quantity is defined with respect to the primary vertex. The subscript ENDVTX means the quantity describes the reconstructed vertex of the relevant particle.

Particle	**Variable**
D^0	χ^2_{ENDVTX}
	p
	p_T
	$\text{FD}\chi^2_{\text{OWNPV}}$
	$\text{IP}\chi^2_{\text{OWNPV}}$
	$\text{DOCA}_{K\pi}$
π_{D^0}	p
	p_{T}
	$\text{IP}\chi^2_{\text{OWNPV}}$
	$\text{ProbNN}\pi$
K_{D^0}	p
	p_{T}
	$\text{IP}\chi^2_{\text{OWNPV}}$
	$\text{ProbNN}K$

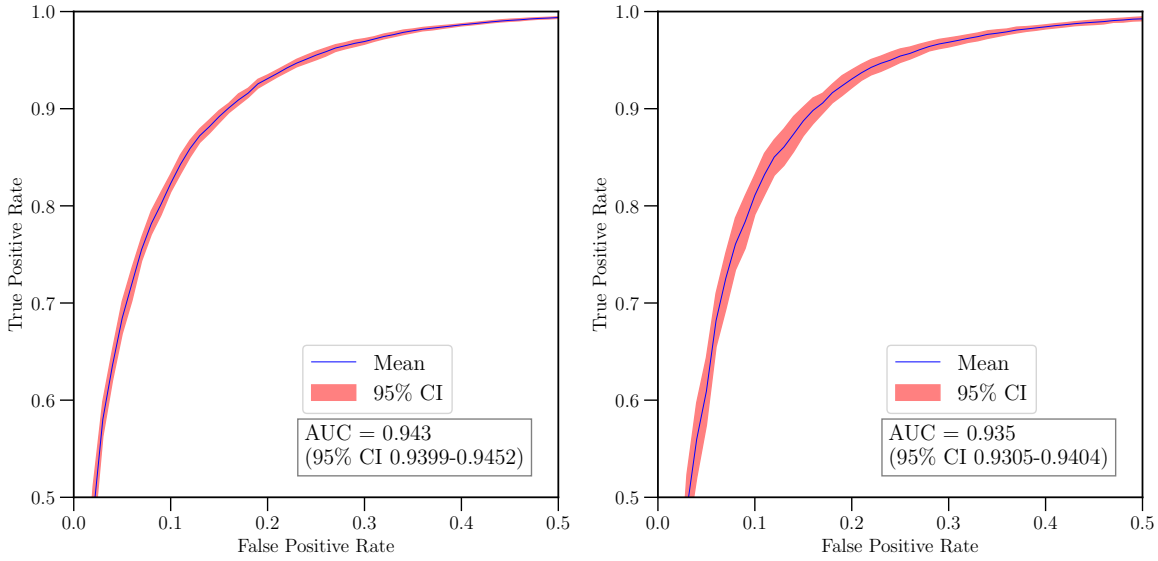
The $B_{\text{selection}}$ classifier in 2016 (TOS and TIS) has a signal training sample that is ~ 5 times smaller than the background sample, so a training weight is applied accordingly. All training data is standardised (*i.e.* mean shifted to 0 and variance scaled to unity) using `scikit-learn` [151].

Receiver operating characteristic curves, can be seen for both periods in Figure 7.6 for $D_{\text{from}B}$, and in Figure 7.7 for $B_{\text{selection}}$. The ROC curves and AUC values are taken as the mean across the 10 k-folds, and the 95% confidence intervals are calculated using bootstrapping [152]. A larger spread is seen in 2016 compared to Run 1 because it has a smaller training sample due to the number of simulation events.

In order to check for overfitting, the classifier response on both training and validation data for each fold of training is checked, and a subset of folds are shown in Figure 7.8 and Figure 7.9 for each classifier respectively. The overlap of training and validation responses signifies no overfitting is taking place, as well as very good separation between signal and background samples. The responses of the classifiers are also checked on the control mode data, which is not used for training, after background subtraction using the *sPlot* method (described in Section 6.3). The integrated distributions of the $B_{\text{selection}}$ responses are shown in Figure 7.10 and Figure 7.11 for both periods comparing MC and sWeighted data to verify that they largely agree. A systematic uncertainty on the

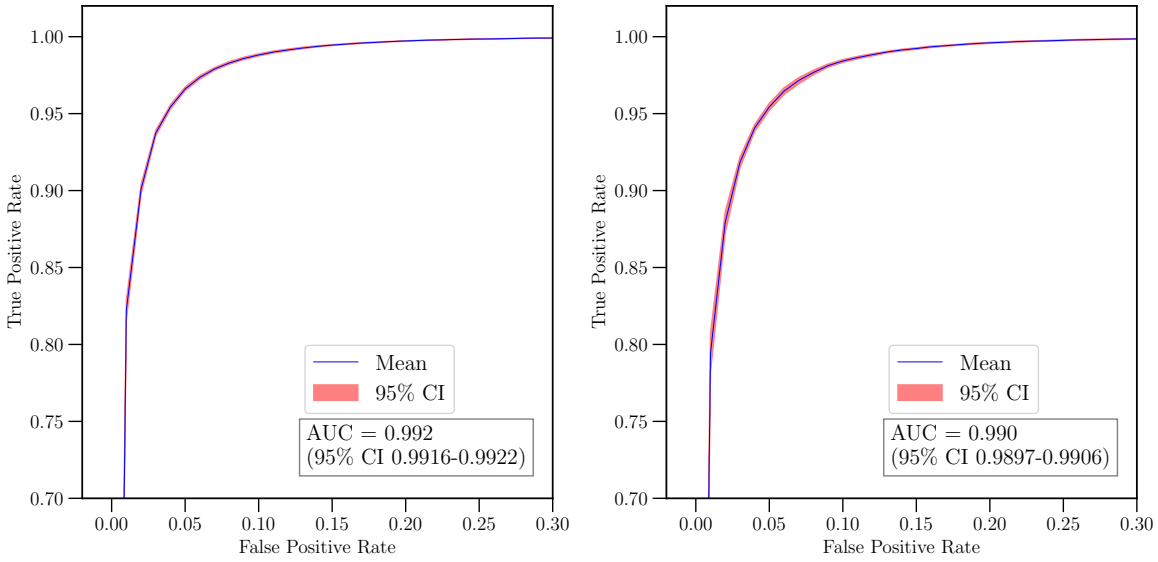


(a) $D_{\text{from}B}$ Run 1: (left) TOS, (right) TIS

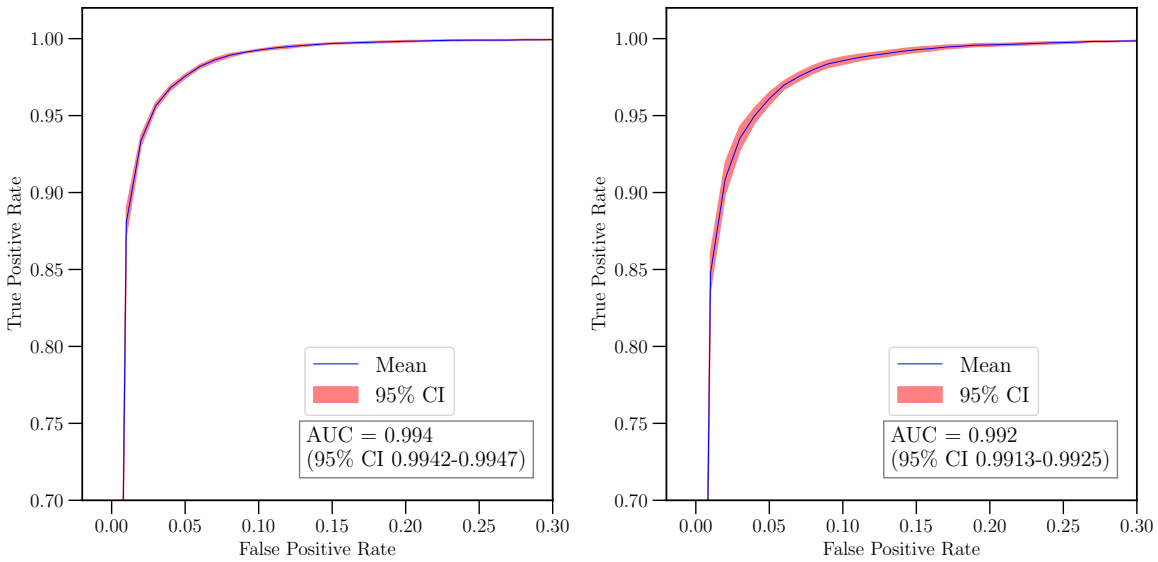


(b) $D_{\text{from}B}$ 2016: (left) TOS, (right) TIS

Figure 7.6: Average ROC curves and confidence intervals of $D_{\text{from}B}$ classifiers (zoomed into area of interest). a) shows Run 1 and b) shows 2016.



(a) $B_{\text{selection}}$ Run 1: (left) TOS, (right) TIS



(b) $B_{\text{selection}}$ 2016: (left) TOS, (right) TIS

Figure 7.7: Average ROC curves and confidence intervals of $B_{\text{selection}}$ classifiers (zoomed into area of interest). a) shows Run 1 and b) shows 2016.

Table 7.4: $B_{\text{selection}}$ input variables. The subscript OWNPV means the quantity is defined with respect to the primary vertex. The subscript ORIVTX means the quantity is defined with respect to the B^0 decay vertex.

Particle	**Variable**
B^0	p
	p_{T}
	DTFχ^2/ndf
	FDχ^2_{OWNPV}
	IPχ^2_{OWNPV}
	DOCAMAX
	$\log_{10}(1-\lvert\text{DIRA}_{\text{OWNPV}}\rvert)\cdot sign(\text{DIRA}_{\text{OWNPV}})$
D^0 &$\overline{D}^0$	$D_{\text{from}B}$ (1^{st} stage classifier)
	FDχ^2_{ORIVTX}
	$\log_{10}(1-\lvert\text{DIRA}_{\text{ORIVTX}}\rvert)\cdot sign(\text{DIRA}_{\text{ORIVTX}})$
K^{*0}	$K_{K^{*0}}$ProbNNK

classifier response is considered in Section 7.9. In this case, the sWeights are calculated before applying the final classifier cut, so they can be applied over a wider range of classifier responses.

7.4.2 Classifier cut optimisation

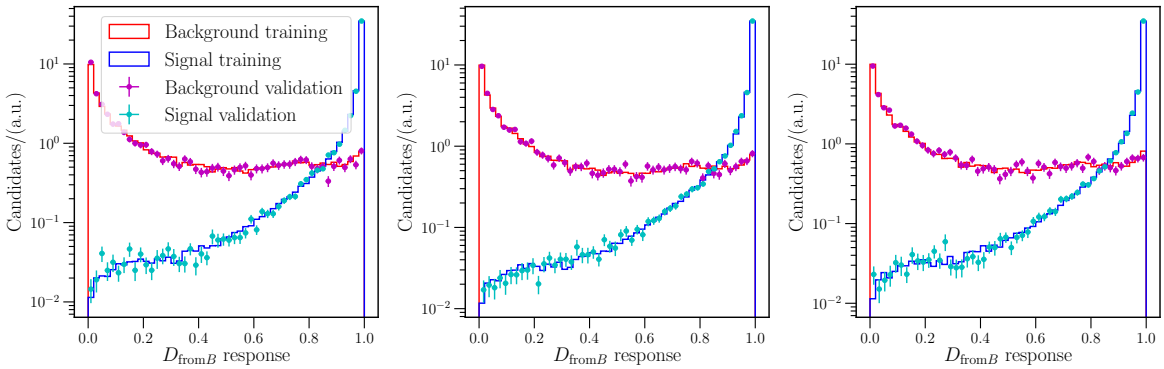
A scan is made over the cut values of the $B_{\text{selection}}$ classifier output, and the corresponding expected number of signal and background events is calculated in order to find the maximum significance, defined as $\mathcal{S} = \frac{\mathcal{N}_S}{\sqrt{\mathcal{N}_S+\mathcal{N}_B}}$. $\mathcal{N}_S$ is the expected number of signal events, defined for the signal mode $B^0 \to D^0\overline{D}^0K\pi$ by rearranging Equation 7.4 as follows,

$$\mathcal{N}_S^{\text{sig}} = \mathcal{L}\times\sigma_{b\bar{b}}\times 2\times f_d\times\epsilon^{\text{sig}}_{\text{geom}}\times\epsilon^{\text{sig}}_{\text{strip}}\times\epsilon^{\text{sig}}_{\text{presel}}\times\epsilon^{\text{sig}}_{\text{trigger}}\times\epsilon^{\text{sig}}_{\text{NN}} \times\mathcal{B}(B^0\to D^0\overline{D}^0K^{*0})\times\mathcal{B}(D^0\to K\pi)^2. \quad (7.10)$$

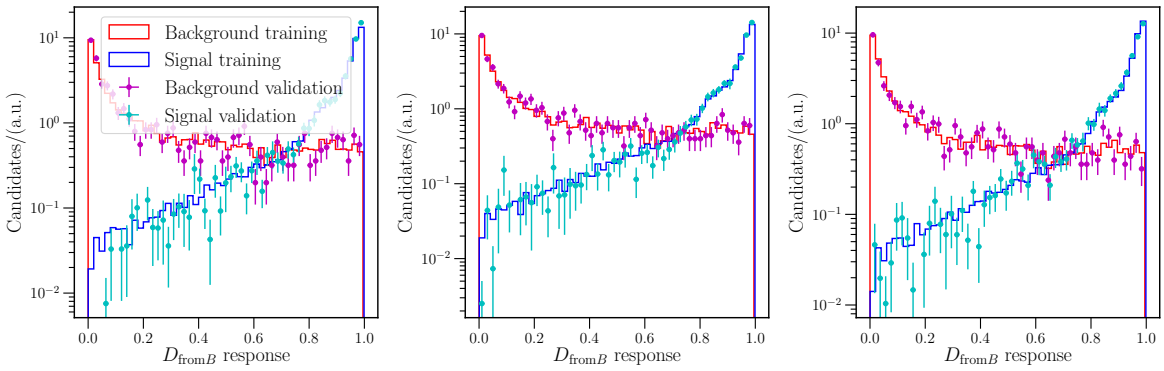
The expected number of signal events in $B^0 \to D^{*-}D^0K^+$ is defined by rearranging Equation 7.5 as follows,

$$\mathcal{N}_S^{\text{cont}} = \mathcal{L}\times\sigma_{b\bar{b}}\times 2\times f_d\times\epsilon^{\text{cont}}_{\text{geom}}\times\epsilon^{\text{cont}}_{\text{strip}}\times\epsilon^{\text{cont}}_{\text{presel}}\times\epsilon^{\text{cont}}_{\text{trigger}}\times\epsilon^{\text{cont}}_{\text{NN}} \times\mathcal{B}(B^0\to D^{*-}D^0K^+)\times\mathcal{B}(D^0\to K\pi)^2\times\mathcal{B}(D^{*-}\to\overline{D}^0\pi^-). \quad (7.11)$$

The parameters used in Equation 7.10 and Equation 7.11 are mostly defined in Section 7.1, and the rest are as follows,



(a) Run 1



(b) 2016

Figure 7.8: 1st stage $D_{\text{from}B}$ classifier response on training (line) and validation (points) samples for signal (red) and background (blue) samples in the TOS category, shown for a random selection of training k-folds. a) shows Run 1 and b) shows 2016.

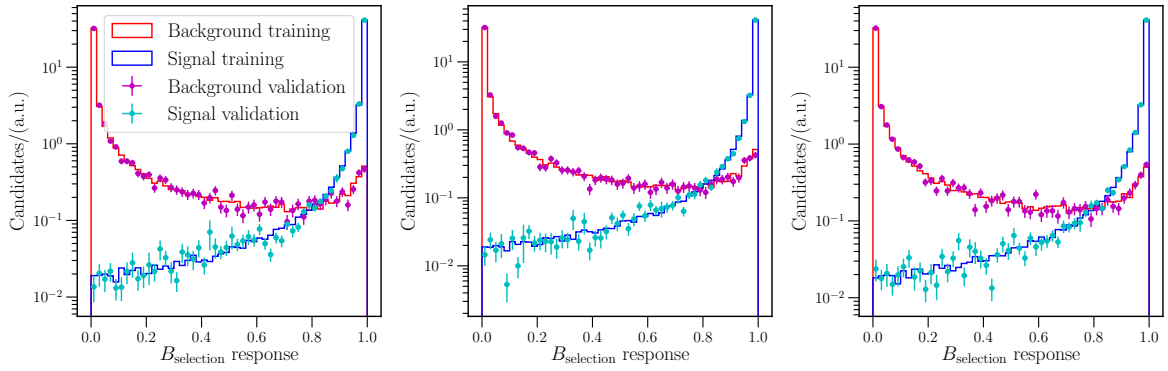
- ϵ_X^{mode} are the various selection efficiencies for the signal or control mode, defined in Section 7.6.

- $\mathcal{B}(B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-) = 2.4 \times 10^{-4}$ is estimated using

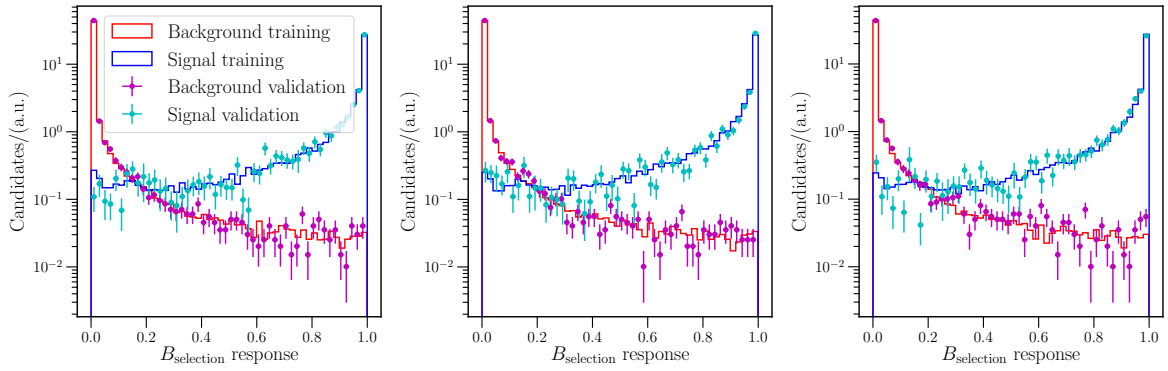
$$\frac{\mathcal{B}(B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-)}{\mathcal{B}(B^0 \to D^0 \overline{D}{}^0 K^0)} = \frac{\mathcal{B}(B^0 \to D^0 K^{*0})}{\mathcal{B}(B^0 \to D^0 K^0)}, \tag{7.12}$$

  where all the other branching fractions are known [2]:

  - $\mathcal{B}\,(B^0 \to D^0 \overline{D}{}^0 K^0) = (0.27 \pm 0.10) \times 10^{-3}$;
  - $\mathcal{B}\,(B^0 \to D^0 K^{*0}) = (4.5 \pm 0.6) \times 10^{-5}$;
  - $\mathcal{B}\,(B^0 \to D^0 K^0) = (5.2 \pm 0.7) \times 10^{-5}$.



(a) Run 1



(b) 2016

Figure 7.9: 2nd stage $B_{\text{selection}}$ classifier response on training (line) and validation (points) samples for signal (red) and background (blue) samples in the TOS category, shown for a random selection of training k-folds. a) shows Run 1 and b) shows 2016.

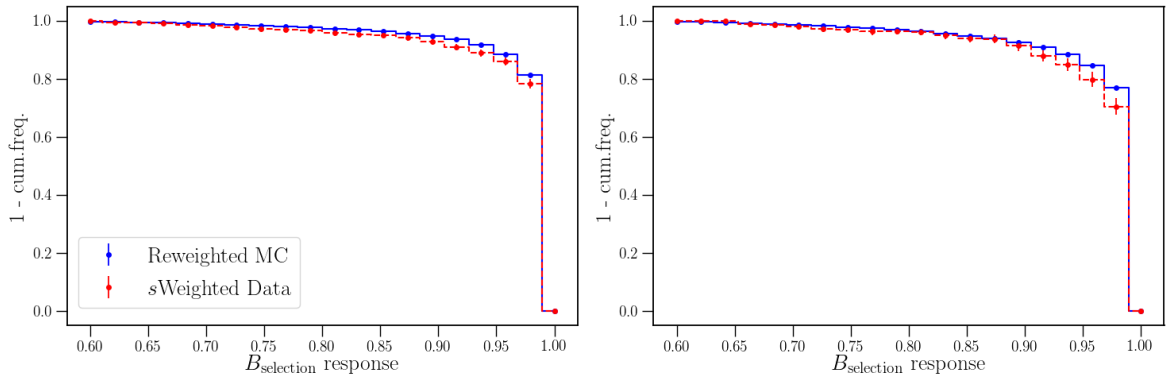


Figure 7.10: Cumulative frequency histograms of the $B_{\text{selection}}$ classifier response for reweighted MC and sWeighted data for Run 1 (left) TOS and (right) TIS.

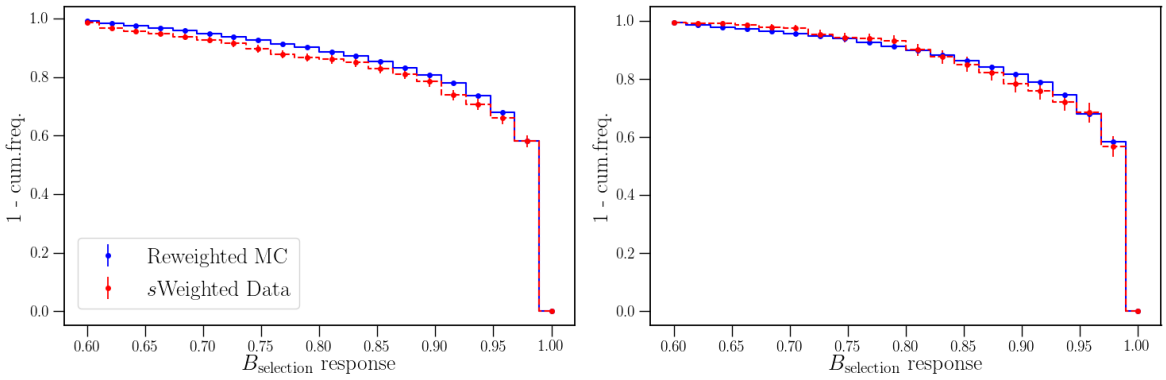


Figure 7.11: Cumulative frequency histograms of the $B_{\text{selection}}$ classifier response for reweighted MC and *s*Weighted data for 2016 (left) TOS and (right) TIS.

This estimate is derived heuristically, since there is no obvious method of calculating $\mathcal{B}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-)$ from other measured decay modes. Additionally, this value is only used to find the optimal cut point of the classifier, and enters the equation as a constant, so the precise value is inconsequential.

The number of background events is estimated by fitting the upper mass sideband region $m(B^0) \in [5380, 5800]$ with a straight line, and extrapolating this back to the signal region, defined as $\pm 50\,\text{MeV}/c^2$ around the nominal B^0 mass, *i.e.*

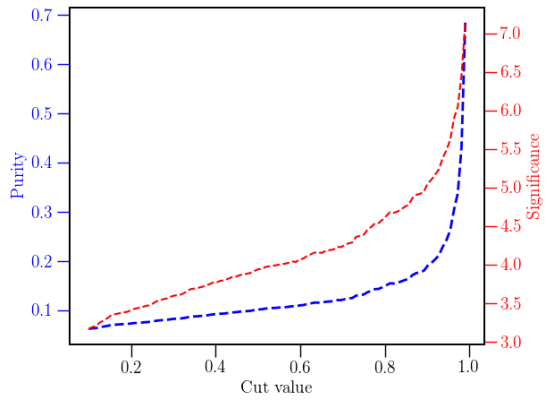
$$\mathcal{N}_B = \int_{m_{B^0}-50}^{m_{B^0}+50} c_0 + c_1 x \, \mathrm{d}m. \tag{7.13}$$

The optimisation curves are shown in Figure 7.12, showing both the figure of merit, significance $\mathcal{S}$, and the purity $\mathcal{P} = \frac{\mathcal{N}_S}{\mathcal{N}_B}$. These are shown for Run 1 TOS only as illustration, displaying the full range of responses as well as a zoomed in region around the optimal point where a finer scan is carried out. The expected significance at the optimal cut points are shown in Table 7.5.

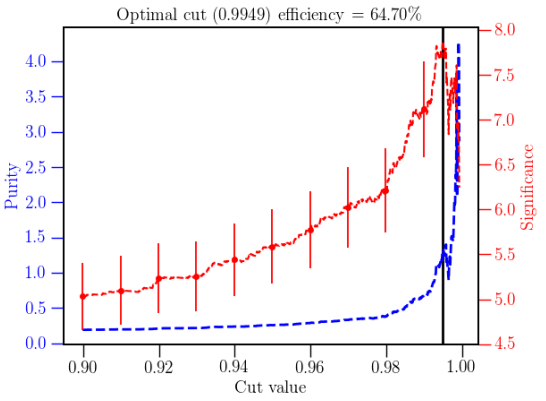
7.5 Simulation samples

Simulated events are needed for three main purposes in this analysis:

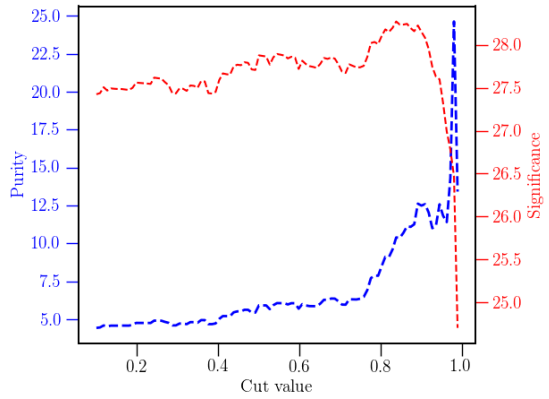
1. Providing a signal sample for training the neural network classifier (Section 7.4.1);
2. Calculating selection efficiencies (Section 7.6);
3. Obtaining the signal shape of the invariant mass of the B^0 (Section 7.8).



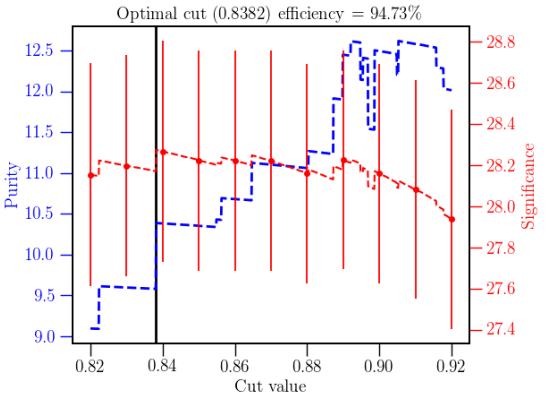
(a) Signal mode



(b) Signal mode



(c) Control mode



(d) Control mode

Figure 7.12: Scans of classifier response against significance (red) and purity (blue), shown for Run 1 TOS. b) and d) show the scans over a zoomed in region around the optimal cut point. Errors are shown for the significance to highlight that the level of variation is small on this scale

Table 7.5: Significance ($\mathcal{S}$) and purity ($\mathcal{P}$) for the optimal classifier cut in each subsample.

Mode	**Period**	**Trigger**	$\mathcal{S}$	$\mathcal{P}$
$B^0 \to D^0 \overline{D}^0 K^+ \pi^-$	Run 1	TOS	7.86	1.31
		TIS	4.55	0.95
	2016	TOS	9.16	4.69
		TIS	6.03	1.15
$B^0 \to D^{*-} D^0 K^+$	Run 1	TOS	28.2	10.4
		TIS	16.8	11.1
	2016	TOS	29.8	14.4
		TIS	17.5	19.5

The simulation samples are generated as described in Section 3.6. Methods used to address the differences between data and simulation are described below. First, the particle identification (PID) variables are transformed, with tools implemented in the PIDCalib package [153, 154] using software developed within LHCb, see Section 7.5.1. Secondly, an event-by-event weight is derived to correct for observed discrepancies in the control channel as detailed in Section 7.5.2.

For the signal decay two samples are used, $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ simulated decays and $B^0 \to D^0 \overline{D}^0 K^{*0}$ simulated decays. The first is a phase-space (PHSP) sample used to calculate selection efficiencies for the final derivation of $\mathcal{R}$, henceforth denoted with the superscript 'phsp'. This uses a simple phase-space decay model where all spins of the initial and final state particles are averaged. The second is used as the signal simulation sample throughout, henceforth denoted with the superscript 'sig'. The control mode uses a simulated sample of $B^0 \to D^{*-} D^0 K^+$ decays, denoted with the superscript 'cont'. Samples of simulated background events are also used to perform studies detailed in Section 7.7. In all cases, the D^0 mesons decay according to phase-space constraints only, and the K^{*0} and D^{*-} decays are modelled as vector mesons decaying to a pair of scalar particles, produced with the correct distributions of the decay angles. As mentioned in Section 3.6, these decay models are implemented in the EvtGen package.

As with the data selections described in Section 7.3, the MC samples undergo stripping, then the preselections, then are split into the two separate trigger categories, and finally subject to the MVA selection.

7.5.1 PID transformation

Differences are often observed in particle identification variables between data and simulation, due to the difficulty in simulating the detectors involved in PID. This primarily involves simulating the RICH subdetector, which depends on the particle kinematics, the event multiplicity, the detector alignment, and properties of the gas. Thus a data-driven method is used to correct the variables in MC using large calibration samples. For the PID of kaons and pions, samples of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decays are used. In this analysis, the PID variables are used in the training of the NN classifier, so it is important that the correlations between variables for a given track are preserved. This is achieved using a kernel density estimation (KDE) approach, implemented in the `Meerkat` [155] package. A KDE is utilised in place of a functional representation of the 3-dimensional distribution of the momentum, pseudorapidity and event multiplicity, ($p_{\mathrm{T}}, \eta, \mathtt{nTracks}$),

of the calibration sample. The variables in simulation are then transformed using this KDE so that they are distributed as in data.

The effect of this transformation is verified using the control channel. Figures 7.13 and 7.14 show the PID variables before and after the transformation, compared to *s*Weighted data, in the TOS category for Run 1 and 2016, respectively. As intended, the transformed variables have better agreement with data.

7.5.2 MC reweighting

Two further variables are seen to have differences in data and MC, namely the track multiplicity `nTracks`, and the χ^2_{IP} of the B^0 with respect to the primary vertex, $\text{IP}\chi^2_{\text{OWNPV}}$. A 2-dimensional weighting is derived using an ensemble of gradient-boosted decision trees, as described in Section 6.4.2, implemented in the `hep_ml` package [156]. This reweighting is derived in the control channel and then applied to the signal channel, and is done separately for each data subsample. The *s*Weights in this case are extracted before the application of the NN classifier. This can be done because the control mode has a relatively low level of combinatorial background after preselections, mostly due to the requirement on Δ_{D^*} (defined in Equation 7.7) to select a D^{*-} candidate.

The 1-dimensional projections of the reweighting variables ($B^0\,\text{IP}\chi^2_{\text{OWNPV}}$, `nTracks`), are displayed in Figure 7.15. It can be seen that the agreement with data improves significantly with the reweighted MC.

7.6 Selection efficiencies

This section describes the extraction of selection efficiencies from simulated events, including a procedure to correct for the non-uniformity of the efficiencies as a function of the decay phase-space. These efficiencies are required for the final calculation of the ratio of branching fractions, as well as in the optimisation of the classifier cut described in Section 7.4.2. An efficiency is defined for each stage of the event selection process sequentially as follows:

- ϵ_{geom}: the geometrical efficiency is the probability of all the decay products being produced with the LHCb detector acceptance, given a generated event. [1]

[1] At the time of this analysis, ϵ_{geom} for $B^0 \to D^0\overline{D}^0K^+\pi^-$ MC in Run 1 was not available to be extracted from the LHCb simulation logs. The value from $B^0 \to D^0\overline{D}^0K^{*0}$ MC is used as they are seen to be very similar in 2016.

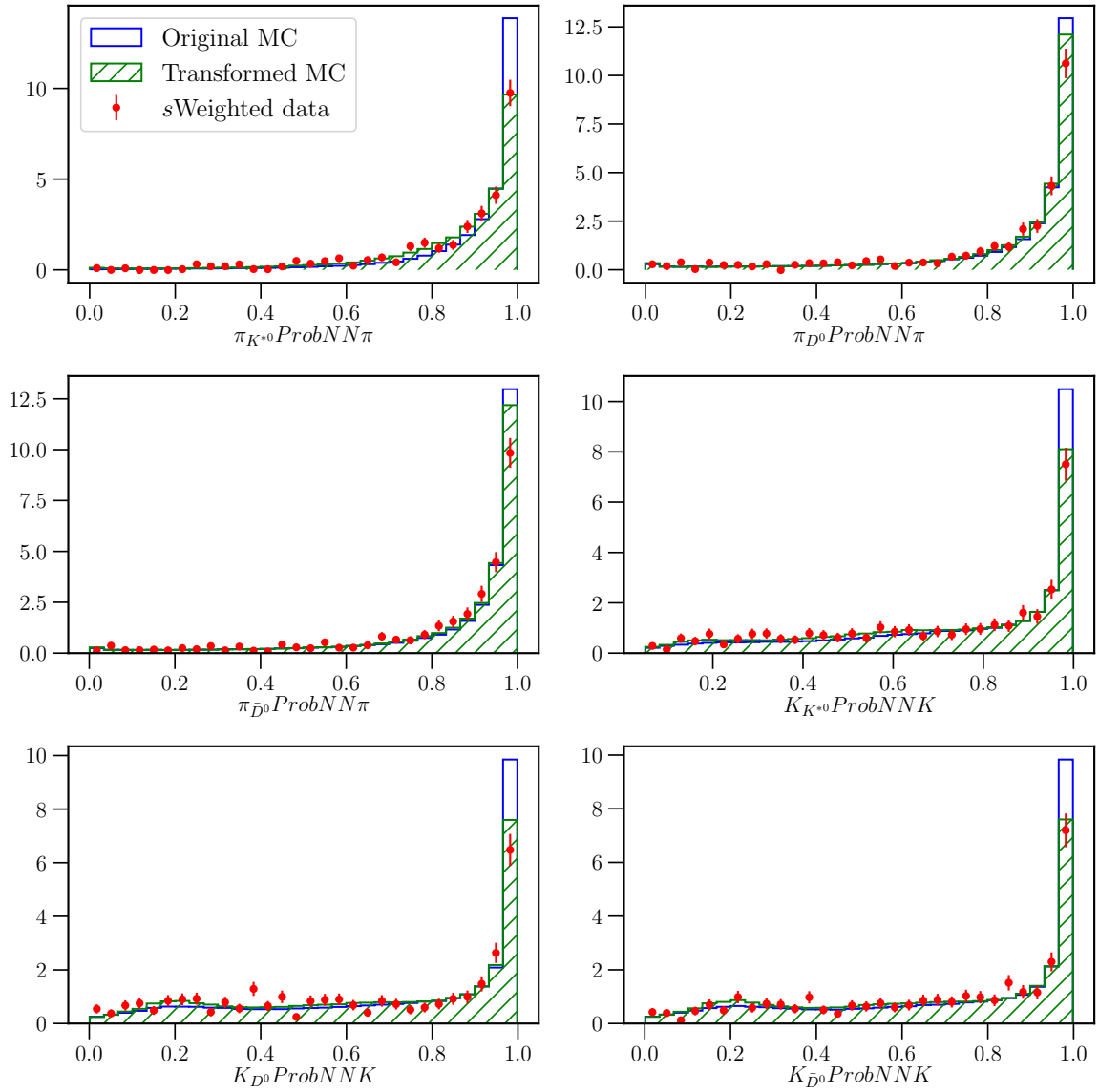


Figure 7.13: Comparison of PID variables between data and MC in Run 1 for the TOS category in the control mode. The original MC distribution is shown in blue, and the transformed distribution in green. *s*Weighted data points are in red. Both MC distributions are shown with the 2D reweighting applied.

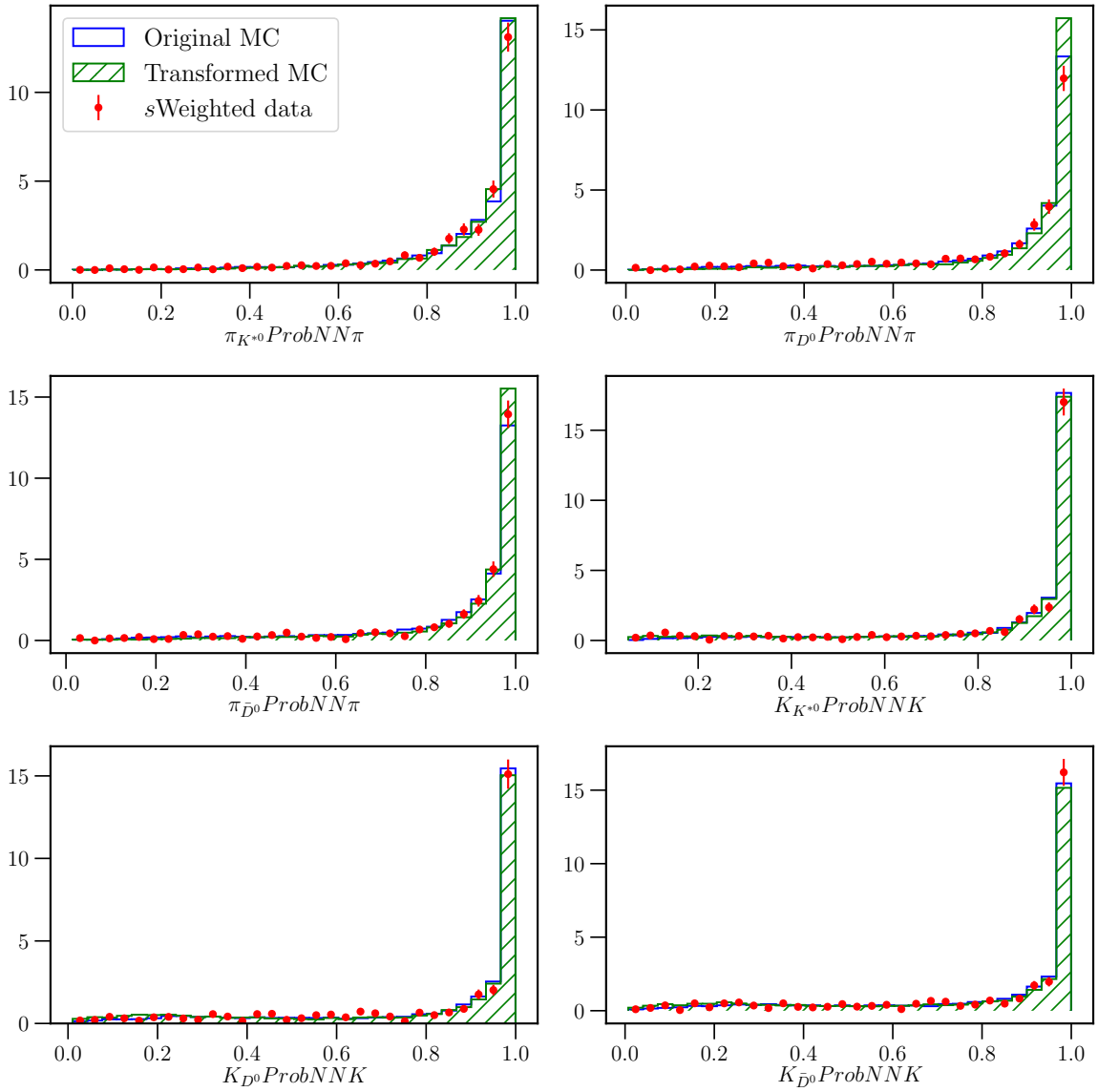
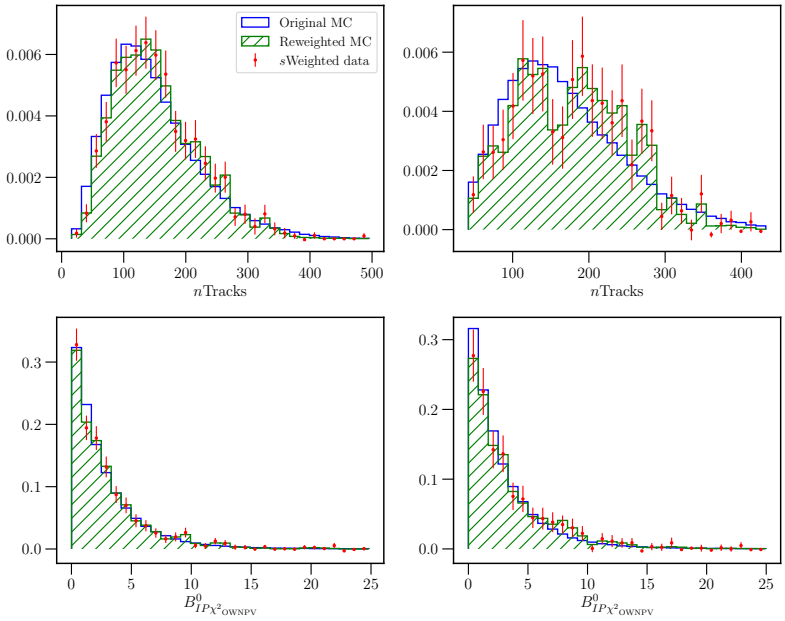
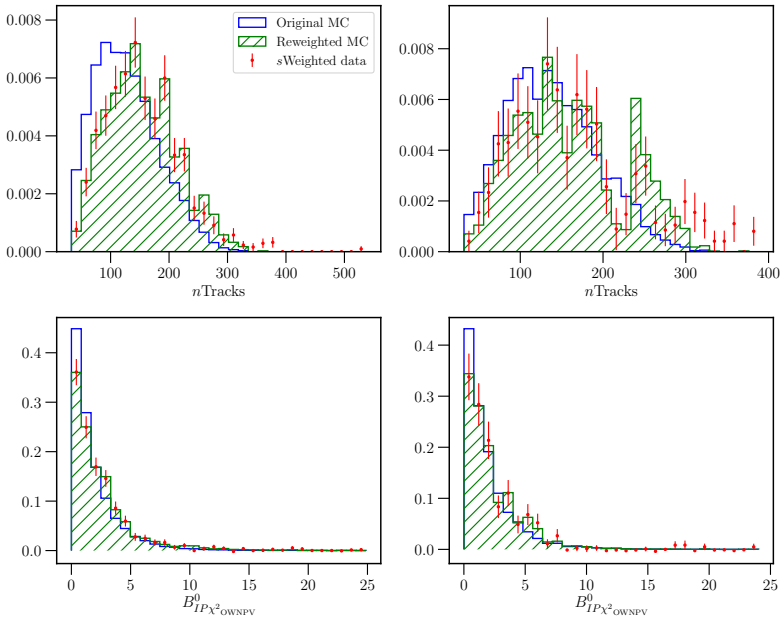


Figure 7.14: Comparison of PID variables between data and MC in 2016 for the TOS category in the control mode. The original MC distribution is shown in blue, and the transformed distribution in green. *s*Weighted data points are in red. Both MC distributions are shown with the 2D reweighting applied.



(a) Run 1: (left) TOS, (right) TIS



(b) 2016: (left) TOS, (right) TIS

Figure 7.15: Projected distributions showing the 2D reweighting of [B^0 IPχ^2_{OWNPV}, nTracks] in MC to match *s*Weighted data in the control mode. The original MC distribution is shown in blue, and the reweighted distribution in green. *s*Weighted data points are in red. a) shows Run 1 and b) shows 2016.

- ϵ_{strip}: the stripping and reconstruction efficiency is the probability that an event will pass the requirements of the stripping line, detailed in Section 7.3.1, and be reconstructed correctly, given that it is produced in the LHCb acceptance.
- ϵ_{presel}: the preselection efficiency is the probability that an event will pass the offline selections described in Section 7.3.3, given that it passed the stripping selection.
- $\epsilon_{\text{trigger}}$: the trigger efficiency is the probability that an event will pass the trigger requirements described in Section 7.3.2 for either TOS (ϵ_{TOS}) or TIS (ϵ_{TIS}), given that it passed the preselections.
- ϵ_{NN}: the neural network classifier efficiency is the probability that an event will pass the optimised classifier cut (see Section 7.4.2), given that it passed the relevant trigger requirements.

The values of the various efficiencies, as well as the numbers of events used to calculate them, are displayed in Tables 7.6 and 7.7 for Run 1 and 2016. $\mathcal{N}_{\text{accepted}}$ is the total number of generated simulation events found in the LHCb acceptance, used to calculate ϵ_{geom}. The overall selection efficiency for each mode can be calculated as the number of events passing all selections out of the number that were generated, *i.e.* $\frac{\mathcal{N}_{\text{selected}}}{\mathcal{N}_{\text{generated}}}$. This is equivalent to multiplying all the efficiencies at each stage of the selection together.

The ratios of efficiencies between the signal and control mode, which are used to calculate the branching fraction ratio $\mathcal{R}$, are shown in Table 7.8. The stripping efficiency varies between Run 1 and 2016 in both modes due to a number of changes implemented in the stripping configuration. In 2016, the PID cuts used to select the final state particles are tightened, and the track χ^2 requirement is loosened. Most importantly, the hardware-level L0 thresholds are lowered, and the software-level trigger is improved. The inclusion of the topological trigger requirement in the stripping means that ϵ_{strip} reflects this change between data-taking periods. The efficiency of the classifier also varies slightly across periods, which is not unexpected due to training on different sample sizes and separate optimisation of cut points.

7.6.1 Efficiency correction

The efficiencies calculated above are global values, however, there may be a dependence on the decay phase-space. For example, regions of the phase-space that have higher occu-

Table 7.6: MC statistics and selection efficiencies (%) for the two signal mode samples. All uncertainties are statistical.

	$B^0 \to D^0 \overline{D}^0 K^{*0}$		$B^0 \to D^0 \overline{D}^0 K^+ \pi^-$	
	Run 1	2016	Run 1	2016
$\mathcal{N}_{\text{accepted}}$	80,393,445	3,957,400	79,384,248	17,649,951
ϵ_{geom}	15.17 ± 0.02	16.26 ± 0.04	15.17 ± 0.02	16.27 ± 0.04
ϵ_{strip}	0.964 ± 0.001	1.595 ± 0.006	0.972 ± 0.0011	1.615 ± 0.003
ϵ_{presel}	70.7 ± 0.1	71.6 ± 0.2	69.8 ± 0.1	70.9 ± 0.1
ϵ_{TOS}	62.2 ± 0.1	55.5 ± 0.2	61.0 ± 0.1	54.1 ± 0.1
ϵ_{TIS}	27.5 ± 0.1	27.6 ± 0.2	28.4 ± 0.1	28.3 ± 0.1
$\epsilon_{\text{NN,TOS}}$	64.7 ± 0.1	54.0 ± 0.3	64.2 ± 0.1	53.8 ± 0.2
$\epsilon_{\text{NN,TIS}}$	56.8 ± 0.1	72.7 ± 0.4	56.2 ± 0.1	72.0 ± 0.2

Table 7.7: MC statistics and selection efficiencies (%) for the control mode. All uncertainties are statistical.

	$B^0 \to D^{*-} D^0 K^+$	
	Run 1	2016
$\mathcal{N}_{\text{accepted}}$	97,715,939	5,126,005
ϵ_{geom}	15.40 ± 0.02	16.43 ± 0.04
ϵ_{strip}	0.613 ± 0.001	0.928 ± 0.004
ϵ_{presel}	77.6 ± 0.1	78.1 ± 0.2
ϵ_{TOS}	68.7 ± 0.1	63.9 ± 0.2
ϵ_{TIS}	23.2 ± 0.1	22.4 ± 0.2
$\epsilon_{\text{NN,TOS}}$	94.7 ± 0.0	97.6 ± 0.1
$\epsilon_{\text{NN,TIS}}$	98.8 ± 0.0	94.0 ± 0.3

Table 7.8: Ratios of selection efficiencies for the $B^0 \to D^{*-}D^0K^+$ control sample with respect to the two signal mode samples, $B^0 \to D^0\overline{D}^0K^{*0}$ and $B^0 \to D^0\overline{D}^0K^+\pi^-$.

	$\frac{\epsilon^{\text{cont}}}{\epsilon^{\text{sig}}}$		$\frac{\epsilon^{\text{cont}}}{\epsilon^{\text{phsp}}}$	
	Run 1	2016	Run 1	2016
ϵ_{geom}	1.015 ± 0.002	1.010 ± 0.003	1.015 ± 0.002	1.010 ± 0.003
ϵ_{strip}	0.636 ± 0.001	0.582 ± 0.004	0.631 ± 0.001	0.575 ± 0.003
ϵ_{presel}	1.098 ± 0.001	1.091 ± 0.004	1.111 ± 0.001	1.101 ± 0.003
ϵ_{TOS}	1.106 ± 0.002	1.151 ± 0.007	1.127 ± 0.002	1.181 ± 0.005
ϵ_{TIS}	0.843 ± 0.003	0.813 ± 0.010	0.819 ± 0.003	0.794 ± 0.008
$\epsilon_{\text{NN,TOS}}$	1.464 ± 0.002	1.807 ± 0.011	1.476 ± 0.002	1.814 ± 0.005
$\epsilon_{\text{NN,TIS}}$	1.740 ± 0.004	1.294 ± 0.008	1.758 ± 0.004	1.305 ± 0.005

pancy may have higher efficiency and vice-versa. In the following discussion, corrections to the global efficiency values are considered.

The efficiency can be plotted as a function of the decay phase-space, in bins of various combinations of the invariant mass planes of the decays. The variation is seen to be large in the control mode, shown in Figure 7.16. The phase-space is parameterised by the invariant mass distributions $m^2(D^{*-}K)$ and $m^2(D^0K)$, known as the Dalitz plot [157] (discussed further in Section 8.1). A correction is defined as

$$\epsilon^{corr} = \frac{\sum_i w_i \epsilon_i(m^2(D^{*-}K^+), m^2(D^0K^+))}{\mathcal{N}_S}, \tag{7.14}$$

where the index i runs over all the candidates in the fitted control mode data, w_i is the signal sWeight, ϵ_i is the efficiency for candidate i as a function of its position in the phase-space, and $\mathcal{N}_S$ is the signal yield. Figure 7.16 displays the binning and values of ϵ_i, where the binning scheme is chosen to be similar to Ref. [85]. The values are overlaid onto a finely binned distribution in order to more clearly display the kinematic boundaries of the Dalitz plot. A sample of events is generated using `RapidSim` [158] in 4π geometry, in order to obtain the phase-space distribution of events before the LHCb acceptance cut. This sample is scaled to the number of events generated in the LHCb Monte Carlo simulation samples. The sWeights and signal yields are derived from the fits to data described in Section 7.8, and thus the correction is measured separately for each category of run period and trigger decision, and ranges from $5-10\%$.

For the signal mode, a single global efficiency derived from the PHSP MC sample $B^0 \to D^0\overline{D}^0K^+\pi^-$ is used, as this has the most accurate kinematic coverage of the signal data. A systematic study is conducted to test the dependence on this model, by applying

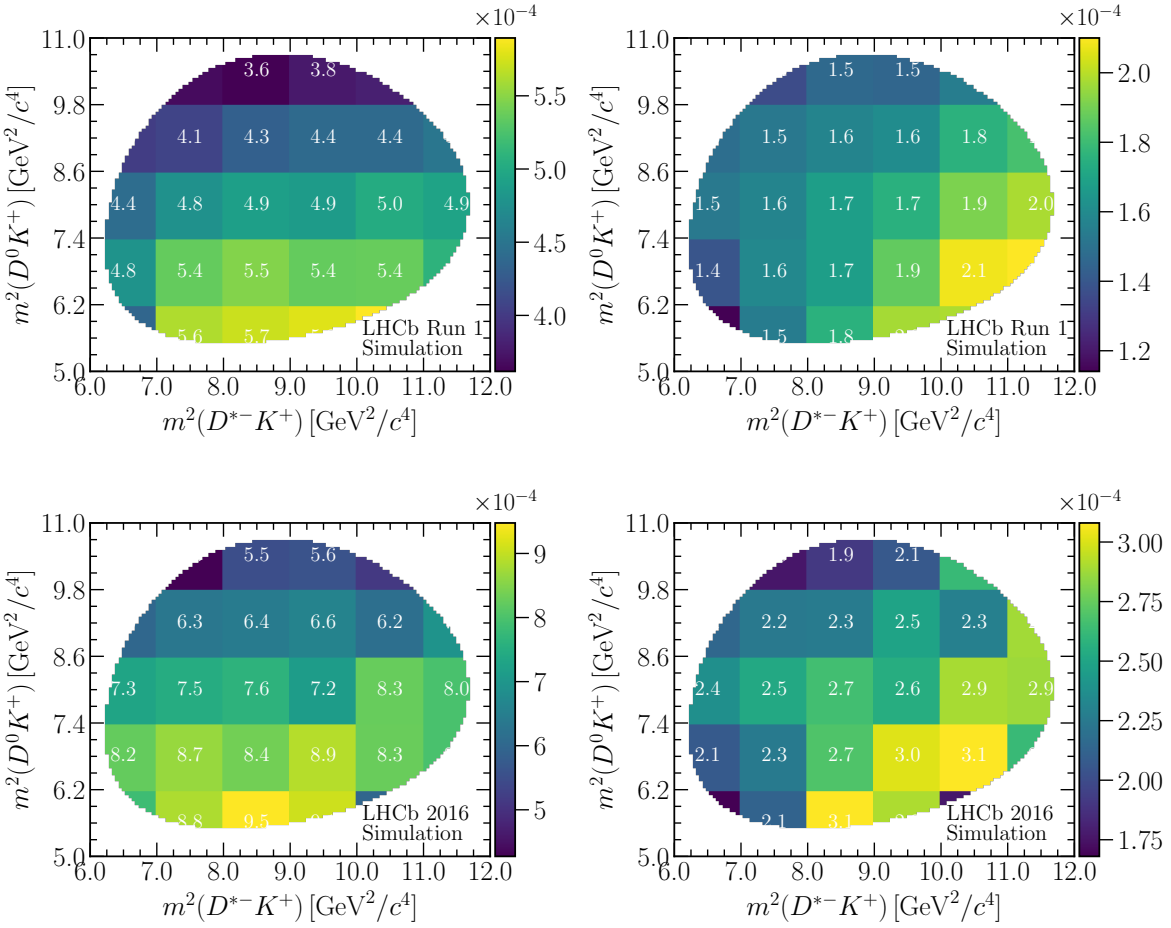


Figure 7.16: Overall selection efficiency as a function of position in the two-body squared invariant mass plane of the control mode $B^0 \to D^{*-} D^0 K^+$. Top row shows Run 1 (left) TOS and (right) TIS, bottom row shows 2016 (left) TOS and (right) TIS.

a similar correction as in the control mode, see Section 7.9. A full correction is not considered, however, as the signal mode would require a fully 5-dimensional paramterisation of the phase-space. This would require a far greater sample of MC, and an arbitrary determination of the binning scheme.

7.7 Background treatment

This section describes the treatment of three types of common backgrounds that may be present in the signal data. These backgrounds persist despite the selections already described. They can be summarised as follows,

1. **Partially reconstructed backgrounds**: These backgrounds occur if the B^0 decays to excited D^0 mesons. These excited mesons will radiate a neutral particle or photon that does not get reconstructed in the $D^0\overline{D}^0K^+\pi^-$ data. The resulting events will have a lower invariant mass than the signal, but could still leak into the region under the B^0 mass peak.

2. **Charmless backgrounds**: This is a possible peaking background that arises when the reconstructed D^0 candidates are not sufficiently well separated from the decay vertex of the B^0 mesons. Thus the decay can proceed via $K\pi$ states that are arranged to form any resonant structure, *i.e.* the decays $B^0 \to {}^{(}\overline{D}{}^{)}K\pi K\pi$ or $B^0 \to K\pi K\pi K\pi$.

3. **Peaking backgrounds**: These backgrounds arise when a final-state particle of a certain decay gets mis-identified, and the event is incorrectly reconstructed as a signal candidate under the $D^0\overline{D}^0K^+\pi^-$ mass hypothesis. They are referred to as such due to the peak of the distribution lying within or near the signal window of $B^0 \to D^0\overline{D}^0K^+\pi^-$.

There is also a combinatorial background component as previously mentioned, which is when random combinations of particles result in $D^0\overline{D}^0K^+\pi^-$ candidates. This background is reduced through selections and finally modelled as described in Section 7.8.

7.7.1 Partially reconstructed backgrounds

In order to estimate the size of the contribution of $B^0 \to D^{(*)0}\bar{D}^{(*)0}K^+\pi^-$ decays (where at least one D^0 is excited), MC samples are generated using a fast simulation package known as `RapidSim` [158]. These samples include the effect of smearing of the momentum distributions of final state charged particles due to detector effects. For the cases with $D^{*0} \to D^0\pi^0$, or equivalently $D^{*+} \to D^0\pi^+$, the mass of the pion dictates that the peak will be separated from the signal peak to below the fit region ([5235, 5600] MeV/c^2). Therefore only the decay $D^{*0} \to D^0\gamma$ is considered here. Figure 7.17 shows the invariant mass distributions of the three generated `RapidSim` samples, overlaid onto signal $B^0 \to D^0\overline{D}^0K^+\pi^-$ MC for Run 1. It is apparent that any partially reconstructed (part-reco) component can be treated as negligible. Since $D^*(2007)^0 \to D^0\gamma$ is a 2-body decay, the γ will carry some energy due to the difference in masses $m(D^*(2007)^0) - m(D^0) = 142\,\text{MeV}/c^2$, so the shift in signal peaks is understood. Additionally, a study is carried out to fit the partially

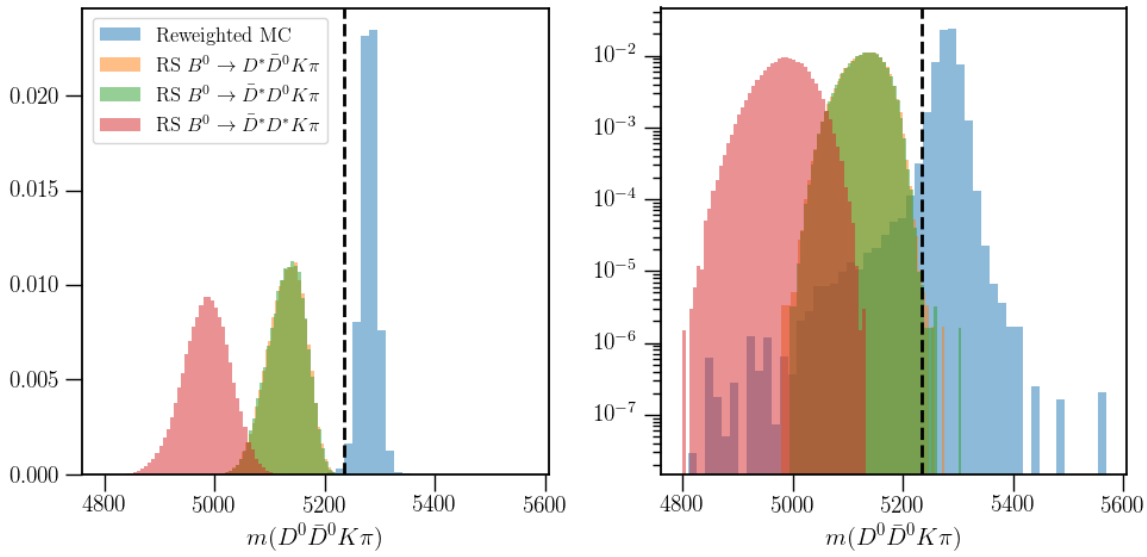


Figure 7.17: Invariant mass distributions of `RapidSim` (RS) samples of $B^0 \to D^{(*)0} \bar{D}^{(*)0} K^+ \pi^-$ (where $\overline{D}^{(*)}(2007)^0 \to \overline{D}^0 \gamma$) and signal MC after all selections. Plotted in log-scale on the right. Black dashed line marks the lower bound of the mass fit range.

reconstructed peaks, as a cross-check and to confirm the separation with the signal region, see Section 7.8.5.

7.7.2 Charmless backgrounds

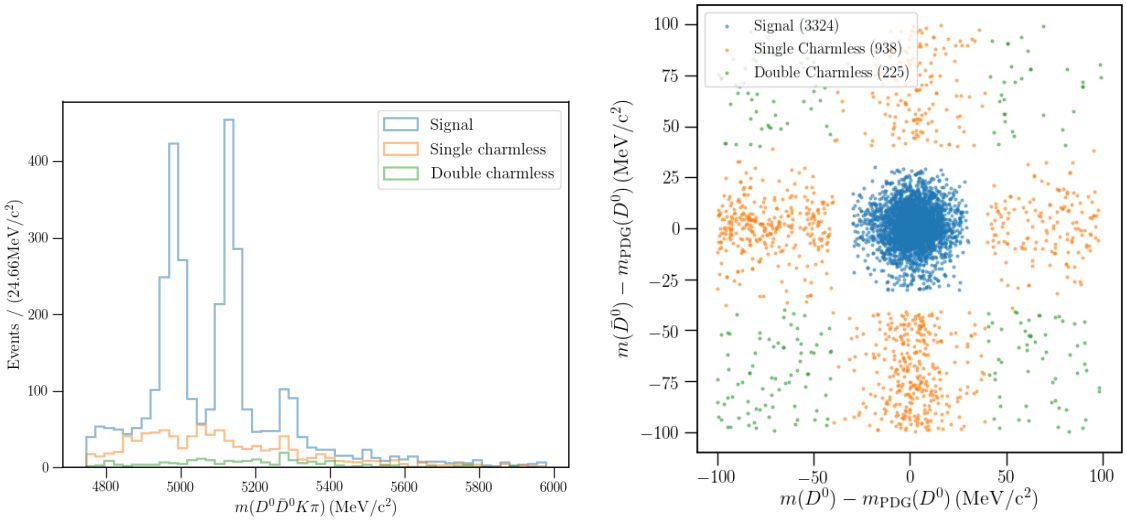
There is a potential background from $B^0 \to \overline{D}^{(0)} K\pi K\pi$ or $B^0 \to K\pi K\pi K\pi$ decays, referred to as single and double charmless respectively. As mentioned in Section 7.3.3, there is a cut on the significance of the flight distance of the D^0 mesons to fight this background. Here a study is carried out to estimate how much charmless background is still present, using the signal mode in the Run 1 TOS category as an illustration. For the purposes of this study, the signal selection requiring $|\, m(D^0) - m_{PDG}(D^0) \,| < 30\,\text{MeV}/c^2$ is not included.

In double charmless backgrounds, both D^0 mesons will be in the sidebands i.e.,

$$40\,\text{MeV}/c^2 < |\, m(\overline{D}^{(0)}) - m_{\text{PDG}}(D^0) \,| < 100\,\text{MeV}/c^2.$$

In the single charmless case, only one of the D^0 mesons will fulfil this condition. The three regions are illustrated in Figure 7.18.

The contamination of single and double charmless backgrounds, ie. the amount of these backgrounds present in the signal D^0 region, is estimated as follows. Firstly the number of combinatorial background events in the signal B^0 region is found by fitting



(a) (b)

Figure 7.18: Single and double charmless background distributions for Run 1 TOS. On the left, the B^0 mass distribution without the D^0 mass windows, with signal and charmless backgrounds separated. On the right, a 2D mass distribution of D^0 against $\overline{D}^0$ with respect to the nominal D mass. The numbers of events in each selection are listed in the legend.

to the B^0 mass without the D^0 mass windows, and integrating the background model $\pm 50\,\text{MeV}/c^2$ around the nominal mass to find $N^{\text{sig}}_{\text{comb}}$. The $D^0\overline{D}^0$ mass plane in the B^0 signal region and the B^0 upper mass sideband is then examined. The number of events in the upper mass sideband are scaled up to $N^{\text{sig}}_{\text{comb}}$, and subtracted from all the events in the B^0 signal region, so that only signal and charmless events are left, and no combinatorial background. These events can then be used to calculate the amount of contamination into the signal region by charmless backgrounds, $N^{\text{contam}}_{\text{charmless}}$, with the relations in Equations 7.15 and 7.16. These relations rely on the ratios between different regions of the $D^0\overline{D}^0$ mass plane such as $\frac{\mathcal{A}_{\text{signal}}}{\mathcal{A}_{\text{charmless}}}$ (where $\mathcal{A}_{\text{region}}$ is the area of the defined region), shown in Figure 7.18(b). In Equation 7.15, the contamination of double charmless events in the single charmless region has been subtracted. This procedure is demonstrated in Figure 7.19 for Run 1 TOS data, and the expected contamination of charmless backgrounds in each category is listed in Table 7.9. The overall contribution summed across all categories is 10 ± 7 candidates. These values are subtracted from the signal yields in the final mass fits for the calculation of $\mathcal{R}$, see Section 7.10.

$$N^{\text{contam}}_{\text{single}} = \left(N_{\text{single}} - N_{\text{double}}\frac{\mathcal{A}_{\text{single}}}{\mathcal{A}_{\text{double}}}\right)\frac{\mathcal{A}_{\text{signal}}}{\mathcal{A}_{\text{single}}}. \tag{7.15}$$

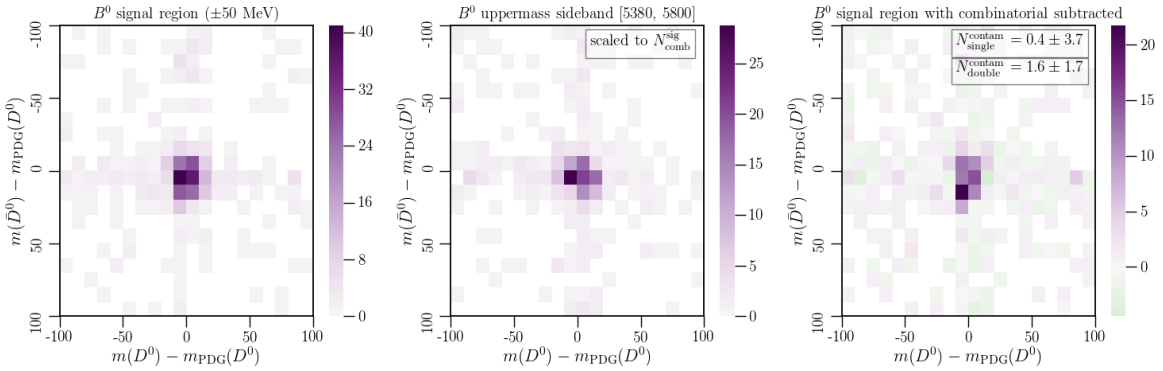


Figure 7.19: 2D histograms of D^0 against $\overline{D}^0$ with respect to the nominal D^0 mass for Run 1 TOS. Left: data in the B^0 signal region. Middle: data in the B^0 uppermass sideband region, scaled up to the number of combinatorial background events in the signal region, $N^{\text{sig}}_{\text{comb}}$. Right: data in the B^0 signal region with the combinatorial events subtracted, *i.e.* the middle plot subtracted from the leftmost plot. The remaining signal and charmless events are used to calculate the contamination of charmless backgrounds using Equations 7.15 and 7.16.

Table 7.9: Number of single and double charmless events expected in the signal region for each category. The bottom row shows the total expected charmless contamination for each category, with statistical uncertainties.

	Run 1		2016	
	TOS	TIS	TOS	TIS
$N^{\text{contam}}_{\text{single}}$	0.4 ± 3.7	-2.5 ± 2.4	-0.1 ± 3.2	2.7 ± 3.1
$N^{\text{contam}}_{\text{double}}$	1.6 ± 1.7	1.7 ± 1.1	3.9 ± 1.3	2.6 ± 1.5
$N^{\text{contam}}_{\text{charmless}}$	2.0 ± 4.0	-0.8 ± 2.6	3.7 ± 3.5	5.3 ± 3.4

$$N^{\text{contam}}_{\text{double}} = N_{\text{double}} \frac{\mathcal{A}_{\text{signal}}}{\mathcal{A}_{\text{double}}}. \tag{7.16}$$

7.7.3 Peaking backgrounds

This section examines a number of other possible peaking backgrounds. All plots in this section are shown for data in the Run 1 TOS category, and with a preliminary set of selections applied. In calculations of the relative yield of background modes with respect to the signal mode, the signal branching fraction is taken as $\mathcal{B}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-) = (3.6 \pm 0.4) \times 10^{-4}$ and the overall selection efficiency as $\epsilon_{D^0 \overline{D}^0 K^+ \pi^-} = (4.65 \pm 0.2) \times 10^{-4}$, which is a preliminary result using only Run 1 data selected in the TOS category (prior to the full selection chain being finalised). Note that

this value of the BF is different from the estimate used in Equation 7.12 because the NN classifier must be applied before any measurable signal peak can be observed.

7.7.3.1 $B^+ \to D^0\overline{D}^0K^+$

This is the case where a $B^+ \to D^0\overline{D}^0K^+$ decay is reconstructed with an additional random pion. $B^+ \to D^0\overline{D}^0K^+$ MC from 2012 is used to estimate the expected relative yield of this component. The MC is processed with the same stripping and other configurations as the signal mode. In this case, PID transformation is not carried out. All the selections are then applied to this sample, giving an efficiency of $\epsilon_{D^0\bar{D}^0K^+} = (1.72 \pm 0.25) \times 10^{-5}$. The relative yield is calculated as

$$R_{D^0\bar{D}^0K^+} = \frac{\mathcal{B}(B^+ \to D^0\overline{D}^0K^+)}{\mathcal{B}(B^0 \to D^0\overline{D}^0K\pi)} \frac{\epsilon_{D^0\bar{D}^0K^+}}{\epsilon_{D^0\bar{D}^0K\pi}} r_{f_{\text{had}}} \tag{7.17}$$

where $\mathcal{B}(B^+ \to D^0\overline{D}^0K^+) = (1.45 \pm 0.33) \times 10^{-3}$ [2], and the factors for $\mathcal{B}(D^0 \to K\pi)$ cancel with the signal mode. $r_{f_{\text{had}}} = 1$ since B^0 and B^+ have the same fragmentation fraction. This gives an expected relative yield of $R_{D^0\bar{D}^0K^+} = (14.8 \pm 4.4)\%$. Therefore, this background is removed using the following veto, illustrated in MC and data in Figures 7.20 and 7.21 respectively:

$$(5220 < m(D^0\overline{D}^0K^+) < 5340)\,\text{MeV}/c^2 \;\&\; m(D^0\overline{D}^0K\pi) > 5380\,\text{MeV}/c^2 \tag{7.18}$$

This veto is 99.996% efficient in Run 1, and 99.998% in 2016, where both efficiencies are derived from signal $B^0 \to D^0\overline{D}^0K^{*0}$ MC. The impact of this veto on the combinatorial background model is addressed in Section 7.8.1. The excited mode $B^+ \to D^0\overline{D}^0K^{*+}$ is expected to be negligible as it will be further suppressed by phase-space requirements.

7.7.3.2 $B^+ \to D^{*0}\overline{D}^0K^+$ or $B^+ \to D^0\overline{D}^{*0}K^+$

This is the partially reconstructed decay of $B^+ \to D^0\overline{D}^0K^+$, described above. This is a particularly dangerous background, as a π^0 or photon is missed and a random pion is added, such that the peak will lie directly under the signal region. This background is removed with the following veto:

$$(5050 < m(D^0\overline{D}^0K^+) < 5200)\,\text{MeV}/c^2 \tag{7.19}$$

This veto is 92.73% efficient, calculated from signal MC combined across all categories. The impact of this veto on the combinatorial background model is addressed in Section 7.8.1.

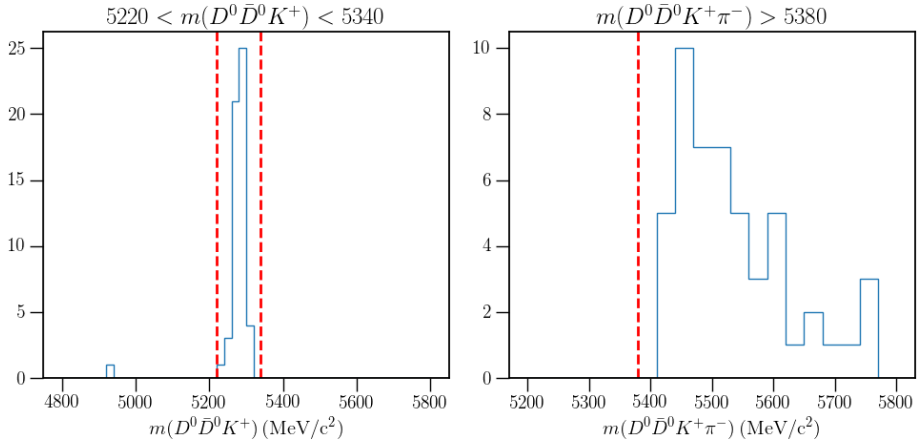


Figure 7.20: Invariant mass distributions of $m(D^0\overline{D}^0K^+)$ (left) and $m(D^0\overline{D}^0K\pi)$ (right) in $B^+ \to D^0\overline{D}^0K^+$ MC after all signal selections, showing the veto regions defined in Equation 7.18. The veto shown on the left plot is applied on the events in the right plot, and vice versa.

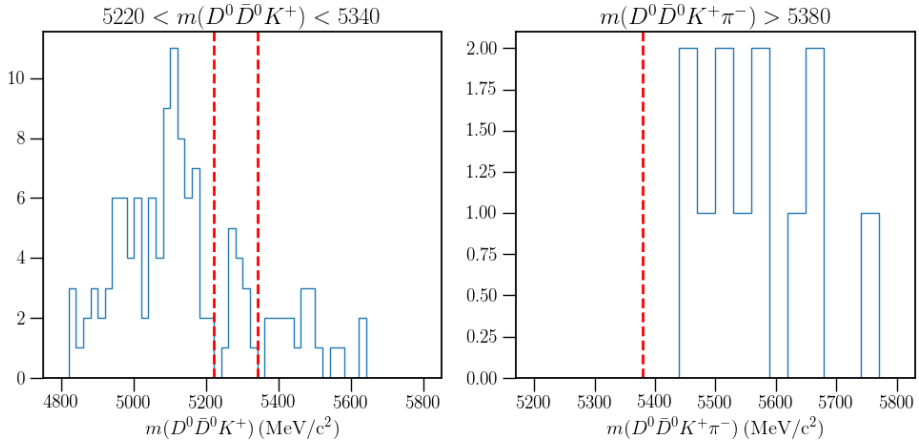


Figure 7.21: Invariant mass distributions of $m(D^0\overline{D}^0K^+)$ (left) and $m(D^0\overline{D}^0K\pi)$ (right) in signal data after all selections in Run 1 TOS, showing the veto regions defined in Equation 7.18. The veto shown on the left plot is applied on the event in the right plot, and vice versa.

7.7.3.3 $B^0 \to D^0\overline{D}^0 K\pi$

This is the case where the pion and kaon are both mis-identified as the other. This background is reduced to a negligible level by the requirements on the PID variables in the preselections, $\pi_{K^*}\text{DLL}_{K\pi} < 10$ and $K_{K^*}\text{DLL}_{K\pi} > -10$ (see Section 7.3.3).

7.7.3.4 $B_s^0 \to D^0\overline{D}^0\rho(770)$

This is the case when a π^+ is reconstructed as the $K^+_{K^*}$, so the decay could proceed via $\rho(770) \to \pi^+\pi^-$. This decay is suppressed by the CKM factors [2],

$$\left|\frac{V_{cd}}{V_{cs}}\right|^2 = \left(\frac{0.220}{0.995}\right)^2 \sim \frac{1}{25} \quad \text{or} \quad \left|\frac{V_{td}}{V_{ts}}\right|^2 = \left(\frac{0.0081}{0.0394}\right)^2 \sim \frac{1}{25}. \tag{7.20}$$

If the rate of mis-identification of $\pi \to K$ is (over-)estimated to be 10%, the contribution of $B_s^0 \to D^0\overline{D}^0\rho(770)$ decays is then expected to be $\ll 1\%$ and therefore neglected.

7.7.3.5 $B_s^0 \to D^0\overline{D}^0\phi$

This is the case when a K^- is reconstructed as the $\pi^-_{K^*}$, so the decay could proceed via $\phi \to K^+K^-$. $B_s^0 \to D^0\overline{D}^0\phi$ MC from 2012 is used to estimate the expected relative yield of this component. The MC is processed with the same stripping and other configurations as the signal mode. In this case, PID transformation is not carried out. All the selections are then applied to this sample, giving an efficiency of $\epsilon_{D^0\overline{D}^0\phi} = (1.02 \pm 0.07) \times 10^{-4}$. The relative yield is calculated as

$$R_{D^0\overline{D}^0K^+} = \frac{\mathcal{B}(B_s^0 \to D^0\overline{D}^0\phi) \times \mathcal{B}(\phi \to K^+K^-)}{\mathcal{B}(B^0 \to D^0\overline{D}^0K\pi)} \frac{\epsilon_{D^0\overline{D}^0\phi}}{\epsilon_{D^0\overline{D}^0K\pi}} r_{f_{\text{had}}} \tag{7.21}$$

where $\mathcal{B}(\phi \to K^+K^-) = (0.492 \pm 0.005)$ [2], and $\mathcal{B}\,(B_s^0 \to D^0\overline{D}^0\phi)$ is not known, so the signal mode BF is used. $r_{f_{\text{had}}} = 0.267^{+0.021}_{-0.020}$ [159]. This gives an expected relative yield of $R_{D^0\overline{D}^0\phi} = (2.9 \pm 0.5)\%$. Therefore, this background is removed using the following veto, illustrated in MC and data in Figures 7.22 and 7.23 respectively:

$$\begin{gathered}(5321 < m(D^0\overline{D}^0K^+(\pi^- \to K^-)) < 5411)\,\text{MeV}/c^2 \\ \text{and either} \\ (1010 < m(K^+(\pi^- \to K^-)) < 1030)\,\text{MeV}/c^2\ \&\ \pi\text{DLL}_{K\pi} > -10 \\ \text{or} \\ (1030 < m(K^+(\pi^- \to K^-)) < 1075)\,\text{MeV}/c^2\ \&\ \pi\text{DLL}_{K\pi} > 10\end{gathered} \tag{7.22}$$

This veto is 99.97% efficient in Run 1, and 99.98% in 2016, where both efficiencies are derived from signal $B^0 \to D^0\overline{D}^0K^{*0}$ MC.

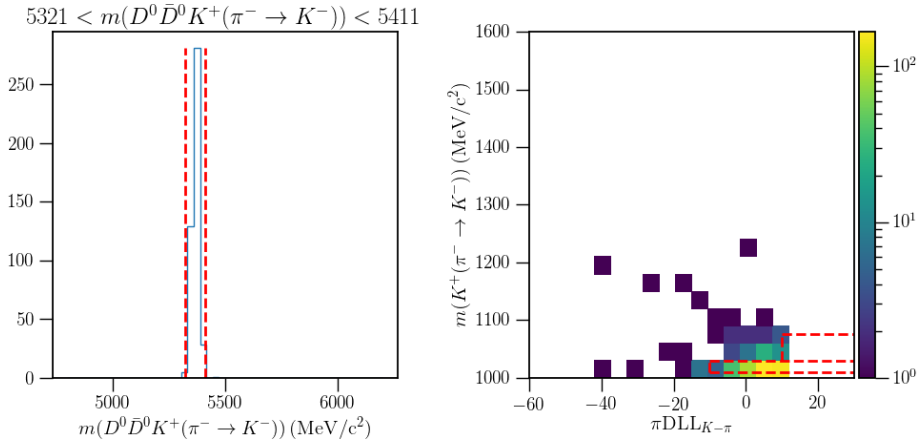


Figure 7.22: On the left, invariant mass distribution of the swapped hypothesis $m(D^0\overline{D}^0K^+(\pi^- \to K^-))$ in $B_s^0 \to D^0\overline{D}^0\phi$ MC after all signal selections. On the right, a 2D distribution of $m(K^+(\pi^- \to K^-))$ against the pion $\mathrm{DLL}_{K-\pi}$. The red lines show the veto regions defined in Equation 7.22. The veto shown on the left plot is applied on the events in the right plot, and vice versa.

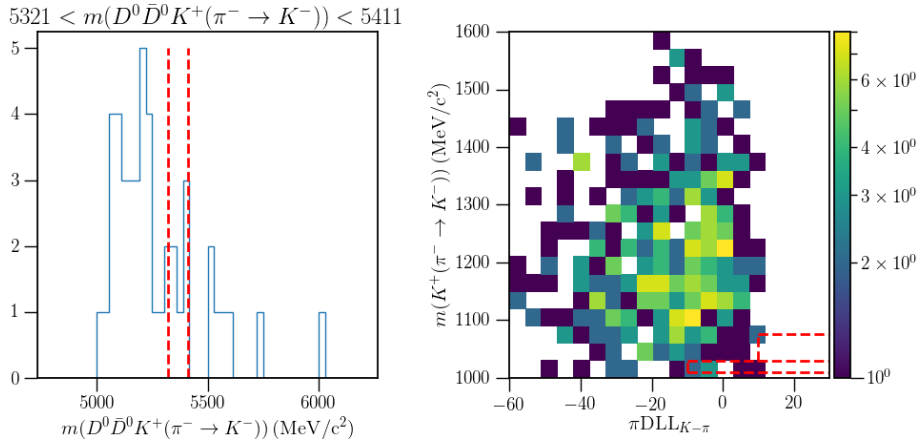


Figure 7.23: On the left, invariant mass distribution of the swapped hypothesis $m(D^0\overline{D}^0K^+(\pi^- \to K^-))$ in signal data after all selections in Run 1 TOS. On the right, a 2D distribution of $m(K^+(\pi^- \to K^-))$ against the pion $\mathrm{DLL}_{K-\pi}$. The red lines show the veto regions defined in Equation 7.22. The veto shown on the left plot is applied on the events in the right plot, and vice versa.

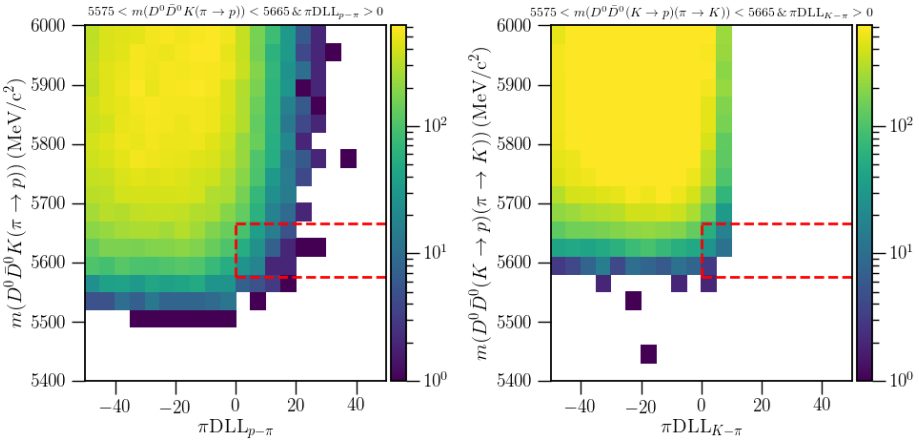


Figure 7.24: 2D distributions showing the reconstructed invariant mass against the DLL in signal MC after all selections for the $\pi \to p$ swap on the left, and the $\pi \to K$ and $K \to p$ swap on the right. The red lines show the veto regions defined in Equations 7.23 and 7.24.

7.7.3.6 $\Lambda_b^0 \to D^0\bar{D}^0pK$

This is the case when a p is reconstructed as either the π_{K^*} or K_{K^*}. In order to remove this background, a 2-dimensional veto is applied for both cases. For the $\pi \to p$ case, events are removed if they have an invariant mass after the change in mass hypothesis which is consistent with a Λ_b^0 and the pion has a proton-like DLL, *i.e.*

$$\begin{gathered}(5575 < m_{D^0\bar{D}^0K(\pi\to p)} < 5665)\,\mathrm{MeV}/c^2 \\ \pi\mathrm{DLL}_{p\pi} > 0\end{gathered} \tag{7.23}$$

For the $K \to p$ case, events are removed if they have an invariant mass after the change in both mass hypotheses which is consistent with a Λ_b^0 and the pion has a kaon-like DLL, *i.e.*

$$\begin{gathered}(5575 < m_{D^0\bar{D}^0(\pi\to K)(K\to p)} < 5665)\,\mathrm{MeV}/c^2 \\ \pi\mathrm{DLL}_{K\pi} > 0\end{gathered} \tag{7.24}$$

These vetoes in Equations 7.23 and 7.24 are illustrated in MC and data in Figures 7.24 and 7.25 respectively. Overall the vetoes for $\Lambda_b^0 \to D^0\bar{D}^0pK$ are 99.7% efficient in Run 1, and 99.8% in 2016, where both efficiencies are derived from signal $B^0 \to D^0\bar{D}^0K^{*0}$ MC.

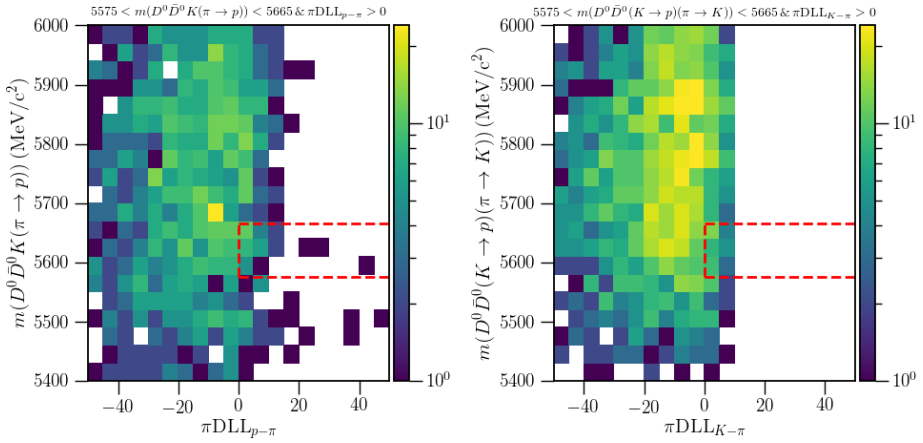


Figure 7.25: 2D distributions showing the reconstructed invariant mass against the DLL in signal data after all selections in Run 1 TOS, for the $\pi \to p$ swap on the left, and the $\pi \to K$ and $K \to p$ swap on the right. The red lines show the veto regions defined in Equations 7.23 and 7.24.

7.7.3.7 $B_s^0 \to D^0 \overline{D}^0 \overline{K}^{*0}$

This decay occurs via a $b \to c\bar{c}d$ transition instead of $b \to c\bar{c}s$, so it encounters suppression in two ways:

$$\left|\frac{V_{cd}}{V_{cs}}\right|^2 = \left(\frac{0.220}{0.995}\right)^2 \sim \frac{1}{25} \tag{7.25}$$

$$\frac{f_s}{f_d} \sim \frac{1}{4} \tag{7.26}$$

So this background is expected to be $\sim 1\%$ of the signal mode, and peaks at a higher mass, so is treated as negligible.

7.8 Mass fits

This section describes the mass fitting procedure, and displays the fits to simulation and data samples. Fits are carried out to the mass of the B^0 after all selections, under D^0 mass constraints, by maximising an extended negative log-likelihood function. This is done in an unbinned procedure, for the fit range $(5235, 5600)\,\mathrm{MeV}/c^2$, where the mass variable is labelled $B^0{}_{\mathrm{Only}D_{\mathrm{M}}}$.

7.8.1 Fit model

The nominal fit model is given by a double-sided Crystal Ball for the signal and an exponential function for the background shape, as detailed in Section 6.2. The invariant mass distribution of the combinatorial background is shaped in a non-trivial way by the vetoes applied to remove $B^+ \to D^0\overline{D}^0K^+$ and $B^+ \to D^{*0}\overline{D}^0K^+$ peaking backgrounds, described in Equations 7.18 and 7.19. A correction for this shaping is derived as follows. A large number of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays are generated using `TGenPhasespace` in the ROOT package [160], where the B^0 mass is sampled uniformly between 5240 and 5600 MeV/c^2. In total 340 mass points are sampled and for each point 10,000 decays are produced. In order to ensure that this phase-space sample is representative of the combinatorial background kinematics in the data, the $m(D^0\overline{D}^0)$ and $m(K^+\pi^-)$ distributions of the phase-space sample integrated in the B^0 mass are corrected to match those from the upper mass sideband of the signal data. The ratio of the B^0 mass distribution (c_i), corrected using the aforementioned procedure, before and after the vetoes are applied is shown in Figure 7.26[2]. The effect of the phase-space correction is also shown. The background model $f_B = \lambda e^{-\lambda x}$ is then corrected as

$$f_B^{corr}(x) = f_B(x)c_i(x), \tag{7.27}$$

where $c_i(x)$ is the correction factor in bin i corresponding to the reconstructed B^0 mass x (represented by the red points in Fig. 7.26).

7.8.2 Fit to MC

Simulation is used to determine the signal shape, in the region $(5235 < B^0{}_{\mathrm{Only}D_M} < 5320)\,\mathrm{MeV}/c^2$. Fits to the signal mode in the TOS category are shown in Figures 7.27 and 7.28. The TIS category and the control mode fits are not shown for brevity. In the TOS category, the parameter n_R is fixed to improve the fit stability, since it is highly correlated to α_R which remains floating. The figures also show the pull distributions, where the pull is defined as the difference between the data and the fit result, with respect to the expected error. The fit result in each bin is calculated by integrating the function over the bin. The expected error is the statistical uncertainty of the fit result in each bin. It is worth noting that some of the pulls look large in the fits to Run 1 MC, where the sample sizes are very large as indicated by the scale of the

[2]The `TGenPhasespace` weight is applied both to the numerator and denominator of the correction factor.

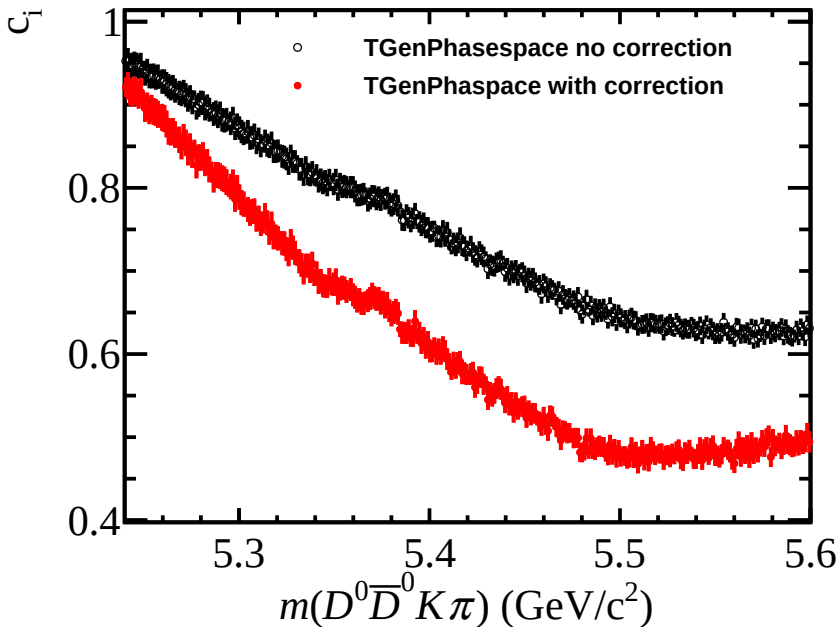


Figure 7.26: Correction of combinatorial background shape as a function of the reconstructed B^0 mass. The black points show the B^0 mass distribution, corrected by the distributions of $m(D^0\overline{D}^0)$ and $m(K^+\pi^-)$ in the signal data. The red points show the same distribution, after also applying the vetoes for $B^+ \to D^0\overline{D}^0K^+$ and $B^+ \to D^{*0}\overline{D}^0K^+$ background decays.

y-axes. However this does not pose a problem, as the level of mismodelling is far smaller than the statistical uncertainty of the real data.

7.8.3 Fit to data

The Gaussian core parameters of the signal shape, the exponential decay parameter of the background shape, and the signal and background yields (N_S, N_B) are extracted from fits to the data. The Crystal Ball tail parameters are fixed to the values extracted from the fits to MC. For the control mode, the value of μ^{cont} is fixed to the value obtained from MC, and the data is used to find the shift in the mean, $\mu^{\text{cont}}_{\text{shift}}$. Then in the signal mode, μ^{sig} is fixed to MC, and $\mu^{\text{sig}}_{\text{shift}}$ is fixed to the value found in control mode data, plus any shift between signal mode MC and control mode MC. The width σ^{sig} is fixed to the control mode data, and a scale factor, $\sigma^{\text{sig}}_{\text{scale}}$ is fixed to the ratio between the width of the signal mode MC and control mode MC. The values of $\sigma^{\text{sig}}_{\text{scale}}$ are shown in Table 7.12. Table 7.10 and 7.11 summarise the free and fixed parameters in the control and signal mode, respectively.

The fits to the control mode data are shown in Figures 7.29 and 7.30, and Figure 7.31

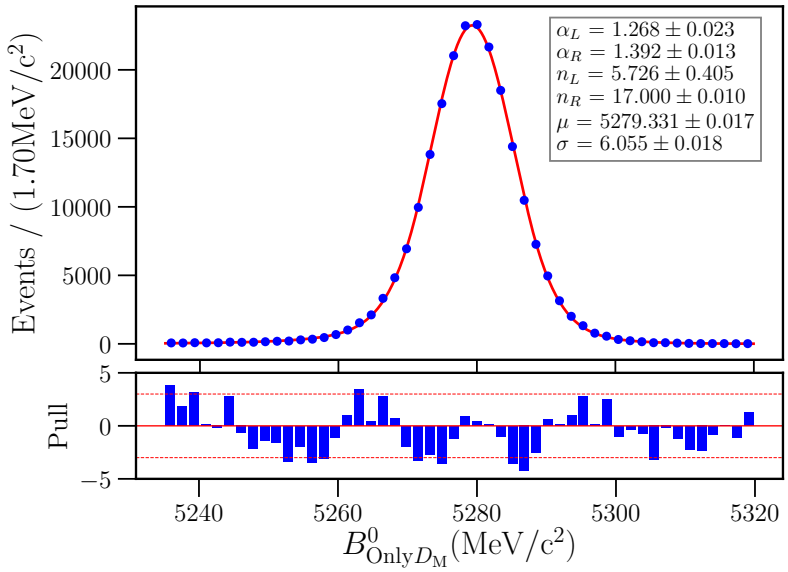


Figure 7.27: Fits to the reconstructed B^0 mass in simulation for the signal mode in Run 1 in the TOS category. The blue points represent MC and the red line represents the fit model.

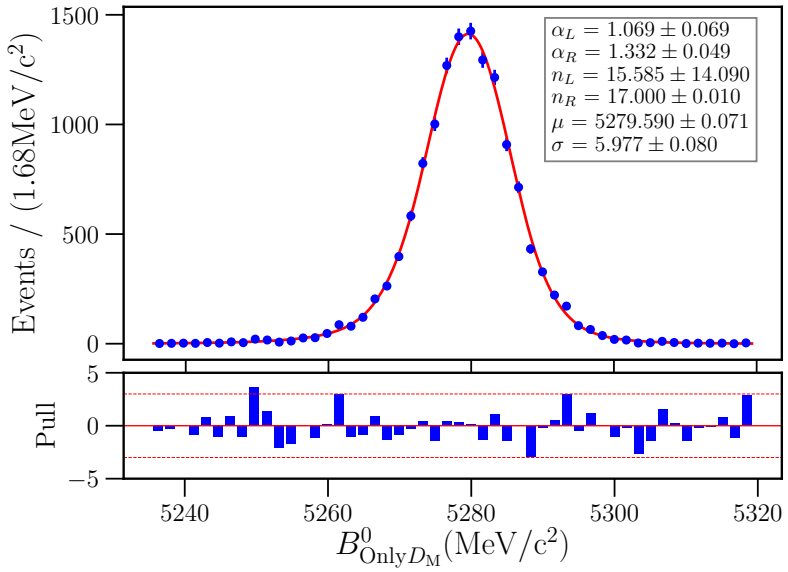


Figure 7.28: Fits to the reconstructed B^0 mass in simulation for the signal mode in 2016 in the TOS category. The blue points represent MC and the red line represents the fit model.

Table 7.10: Status of parameters in the fit to control mode data. Parameters are either fixed to the results of the MC fits, or they are floating (freely varying). The DSCB tail paramters, $\alpha_L, \alpha_R, n_L, n_R$ are fixed to MC as standard.

Parameter	Fixed to MC	Floating
μ^{cont}	✓	
$\mu^{\text{cont}}_{\text{shift}}$		✓
σ^{cont}		✓
λ		✓
N_S		✓
N_B		✓

Table 7.11: Status of parameters in the fit to signal mode data. Parameters are either fixed to the results of the MC fits, fixed to the results of the control mode data fits, or they are floating (freely varying). The DSCB tail paramters, $\alpha_L, \alpha_R, n_L, n_R$ are fixed to MC as standard.

Parameter	Fixed to MC	Fixed to control data	Floating
μ^{sig}	✓		
$\mu^{\text{sig}}_{\text{shift}}$		✓	
σ^{sig}		✓	
$\sigma^{\text{sig}}_{\text{scale}}$	✓		
λ			✓
N_S			✓
N_B			✓

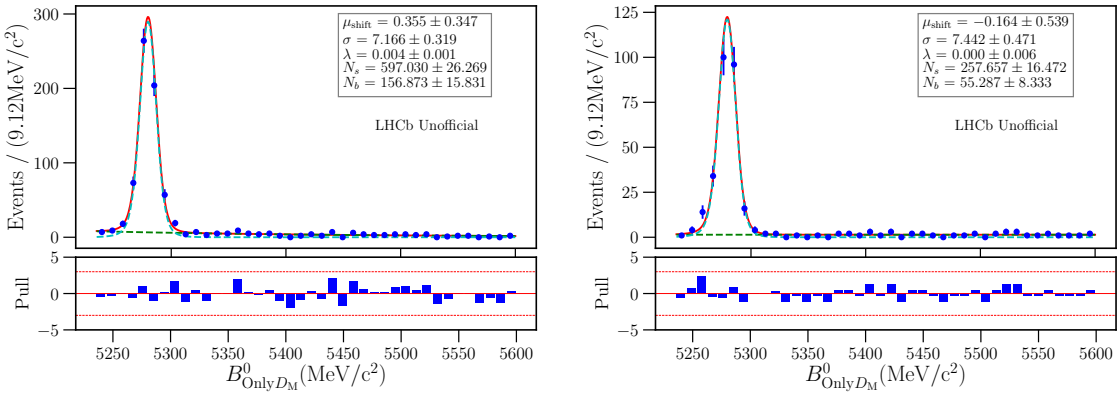
shows all the control data subsamples and fit projections combined. The fits to the signal mode data are shown in Figures 7.32 and 7.33. The floating parameters in each fit are displayed in the legends. The signal yields for the signal and control modes in the two trigger categories are shown in Table 7.13.

7.8.4 Simultaneous fit

In order to obtain a combined result of $\mathcal{R}$ across the 2 data-taking periods and 2 subsequent trigger categories, the subsamples in the signal mode are fitted simultaneously. The signal yield $\mathcal{N}(B^0 \to D^0 \overline{D}^0 K^+ \pi^-)$ is parameterised in terms of $\mathcal{R}$, by rearranging

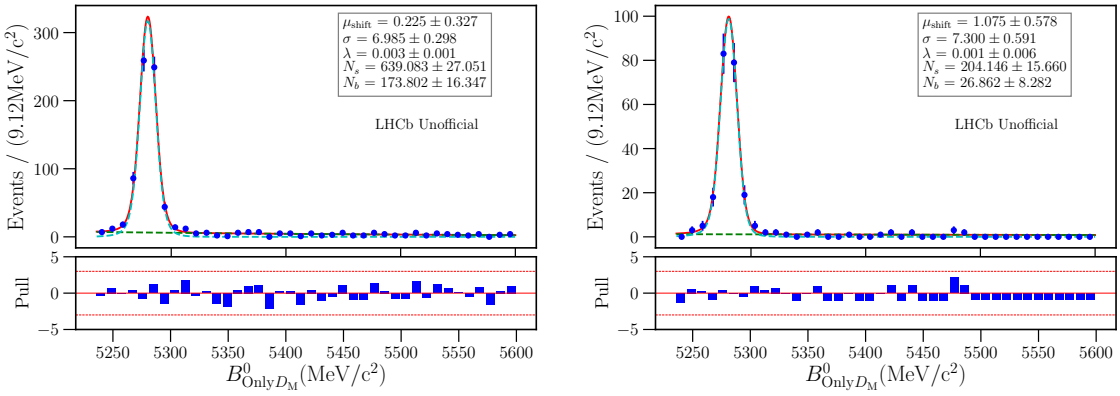
Table 7.12: Ratio of Gaussian width σ between signal mode and control mode MC, used to scale the width in the fit to signal mode data.

Category	$\sigma^{\text{sig}}_{\text{scale}}$
Run 1 TOS	0.95
Run 1 TIS	0.94
2016 TOS	0.92
2016 TIS	0.87



(a) TOS (b) TIS

Figure 7.29: Fits to the reconstructed B^0 mass in data for the control mode in Run 1. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.



(a) TOS (b) TIS

Figure 7.30: Fits to the reconstructed B^0 mass in data for the control mode in 2016. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.

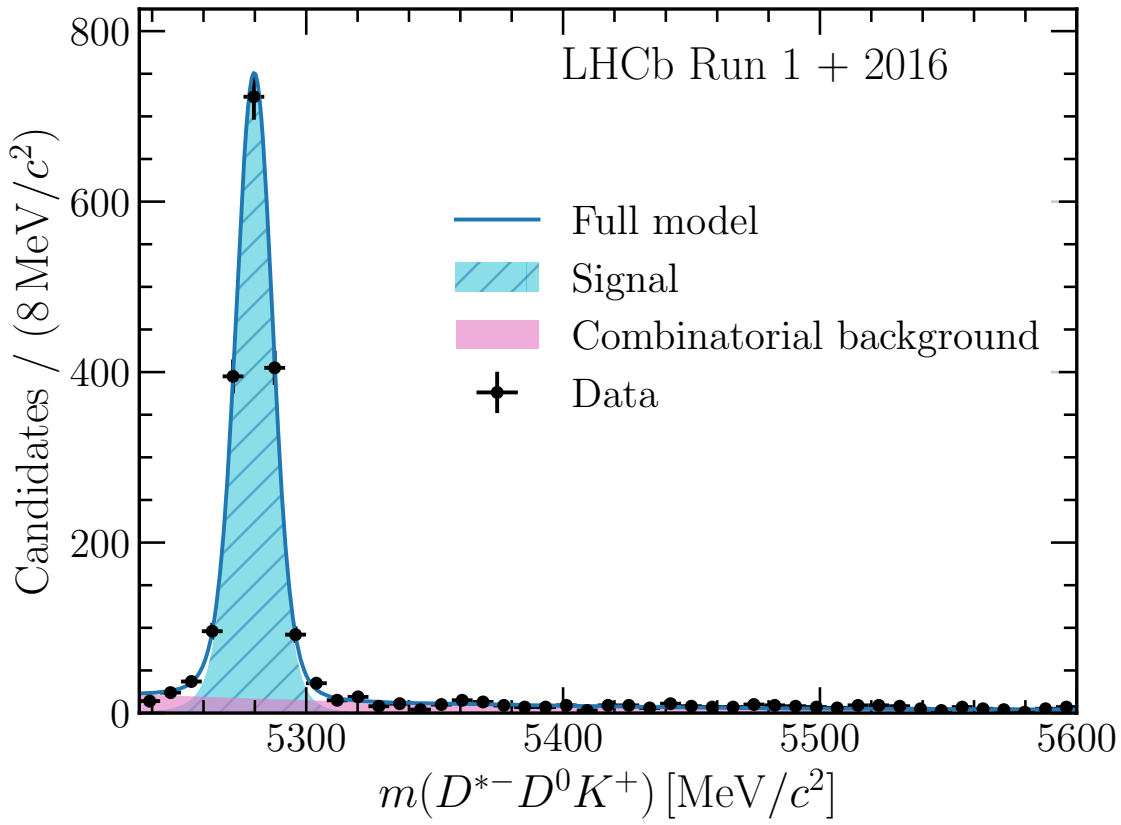
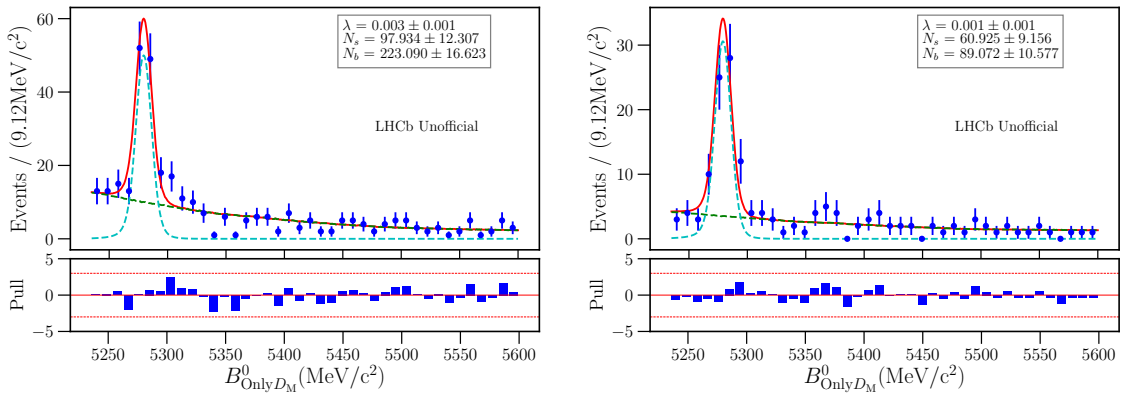


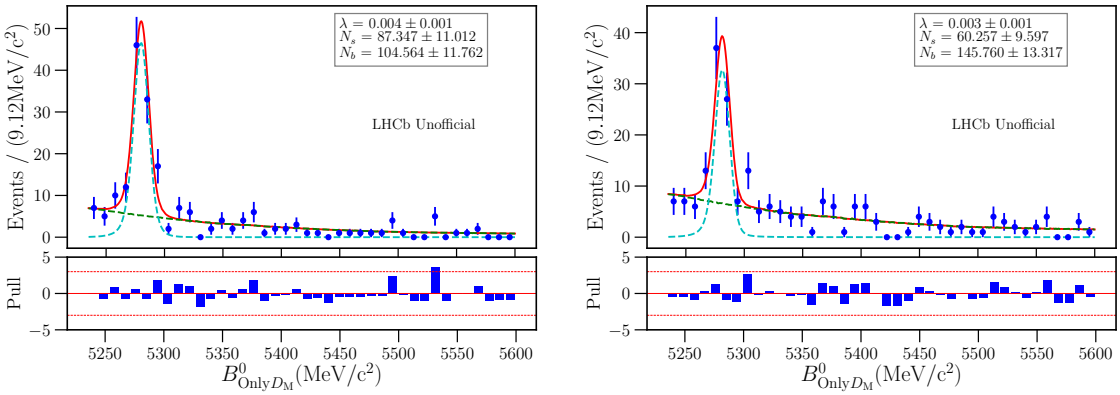
Figure 7.31: Invariant mass distributions for B^0 candidates in the control mode for all subsamples combined. The data are shown as black points with error bars and the fit components are as described in the legend.



(a) TOS

(b) TIS

Figure 7.32: Fits to the reconstructed B^0 mass in data for the signal mode in Run 1. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.



(a) TOS (b) TIS

Figure 7.33: Fits to the reconstructed B^0 mass in data for the signal mode in 2016. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.

Table 7.13: Signal yields from fits to $B^0 \to D^0\overline{D}^0K^+\pi^-$ and $B^0 \to D^{*-}D^0K^+$ split by data taking period and trigger category. The signal mode yields contain the remaining charmless contributions.

	$B^0 \to D^0\overline{D}^0K^+\pi^-$		$B^0 \to D^{*-}D^0K^+$	
	Run 1	2016	Run 1	2016
$\mathcal{N}_S$ (TOS)	98 ± 12	87 ± 11	597 ± 26	639 ± 27
$\mathcal{N}_S$ (TIS)	61 ± 9	60 ± 10	258 ± 16	204 ± 16

Equation 7.6 as

$$\mathcal{N}_S^{\text{sig},i} = \mathcal{R} \times \frac{\mathcal{N}_S^{\text{cont},i}}{\mathcal{B}(D^{*-} \to D^0\pi^-)} \times \frac{\epsilon^{\text{sig},i}}{\epsilon^{\text{cont},i}}, \tag{7.28}$$

where i is an index labelling the various categories, ϵ^{mode} is the overall selection efficiency for the relevant decay mode, and $\mathcal{N}_S^{\text{sig}}$ has had the expected charmless contribution subtracted (see Section 7.7.2). As detailed in Section 7.6.1, the efficiency in the control mode ϵ^{cont} is corrected based on the position of the events in the 2D phase-space, and the efficiency in the signal mode is derived from the PHSP $B^0 \to D^0\overline{D}^0K^+\pi^-$ MC sample.

The ratio of selection efficiencies is fixed for each category, and the the control mode yields from the individual fits $\mathcal{N}_S^{\text{cont},i}$ listed in Table 7.13, are passed to the likelihood as a Gaussian constraint. This is achieved by adding an additional term to the log-likelihood, given by $\log\mathcal{L}_{Gauss}$ where $\mathcal{L}_{Gauss}$ is a Gaussian distribution with mean equal to the appropriate control mode yield, and width equal to the error on that yield. The DSCB

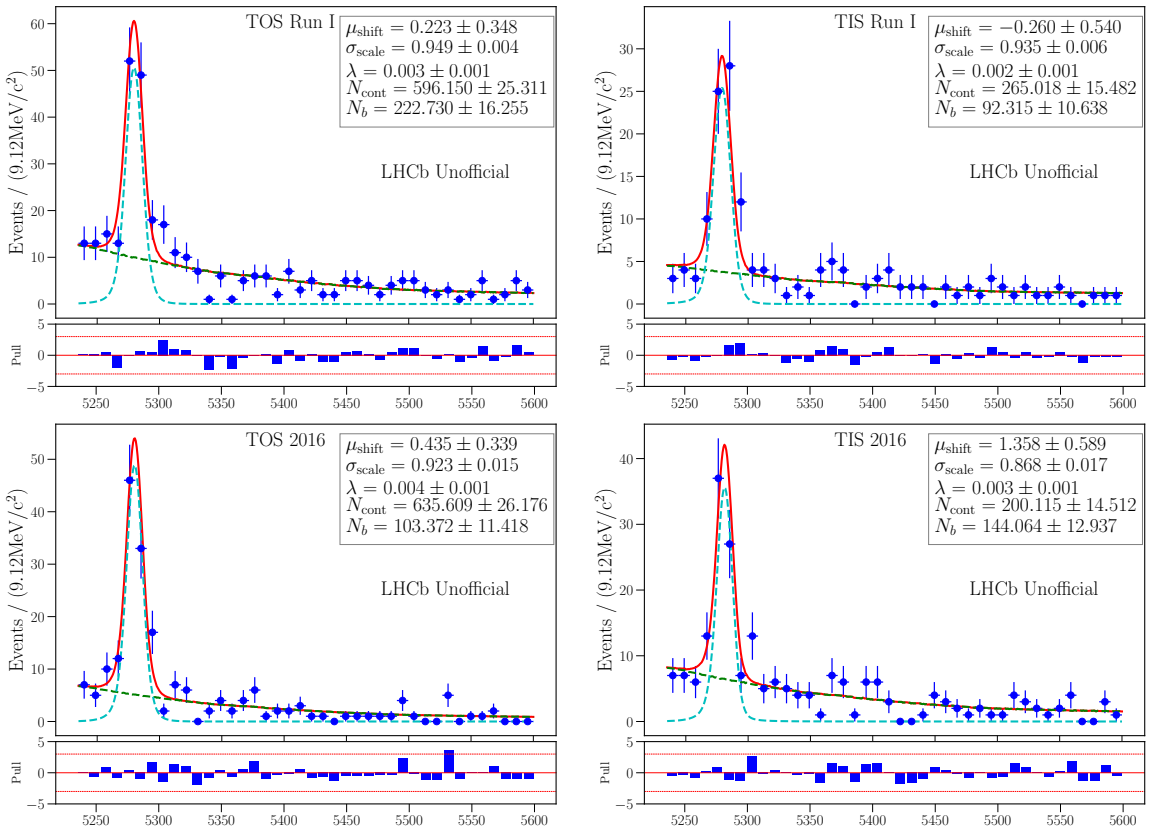


Figure 7.34: Simultaneous fit across signal mode categories. Upper left: Run 1 TOS, upper right: Run 1 TIS, lower left: 2016 TOS, lower right: 2016 TIS. The fit to $\mathcal{R}$ is done simultaneously across the categories.

Gaussian mean and width are fixed using MC and the control mode data as in the individual fits. Otherwise the same parameters as before, *i.e.* the exponential decay parameter and the yields N_S^{sig} and N_B^{sig} are floating in each fit, and $\mathcal{R}$ is shared.

The simultaneous fit is shown in Figure 7.34, and the resulting value of $\mathcal{R}$ is

$$\mathcal{R} = (14.2 \pm 1.1)\%, \tag{7.29}$$

where the uncertainty is statistical. This value gives signal yields for each category that are compatible with the individual fit yields in Table 7.13. Figure 7.35 shows the invariant mass distribution and overlaid fit projections for all subsamples in the signal data combined.

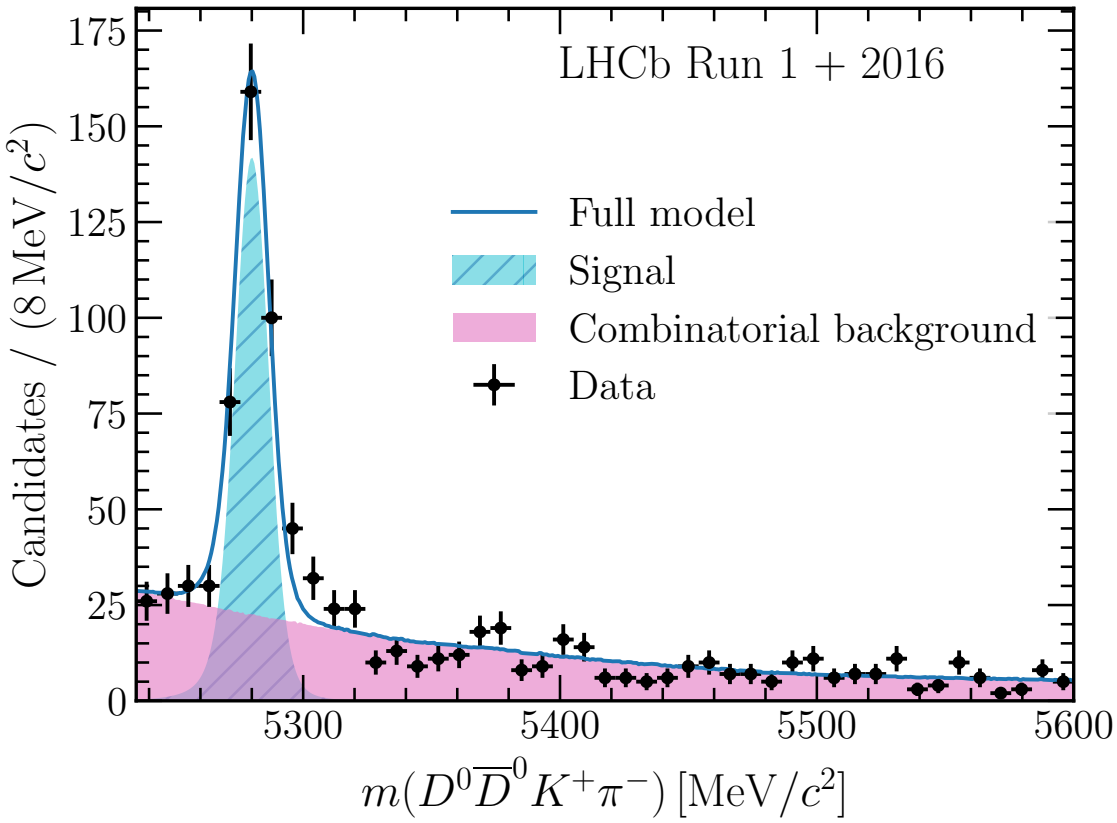


Figure 7.35: Invariant mass distributions for B^0 candidates in the signal mode for all subsamples in the simultaneous fit combined. The data are shown as black points with error bars and the fit components are as described in the legend. The small charmless contamination is included in the signal component in the plot and subtracted statistically.

7.8.5 Fit to extended range

For illustration purposes, and to provide a cross-check, the fit to the signal mode is also carried out in an extended range (4900, 5600) MeV/c^2 to see the separation with the partially reconstructed peaks of $B^0 \to D^{0(*)}\overline{D}^{0(*)}K^+\pi^-$ decays. This fit is displayed in Figure 7.36. Both partially reconstructed peaks are modelled by Crystal Ball functions, where the shapes are initially floating in fits to the individual categories, and subsequently fixed in the simultaneous fit. Otherwise the fitting procedure is as described above. The fitted values of the exponential decay parameter λ in each category is in good agreement with the results from the nominal fit. The value of $\mathcal{R}$ obtained from this fit over the extended range is also compatible with the nominal result to within one-third of the statistical uncertainty. The calculation of the relative uncertainties between the nom-

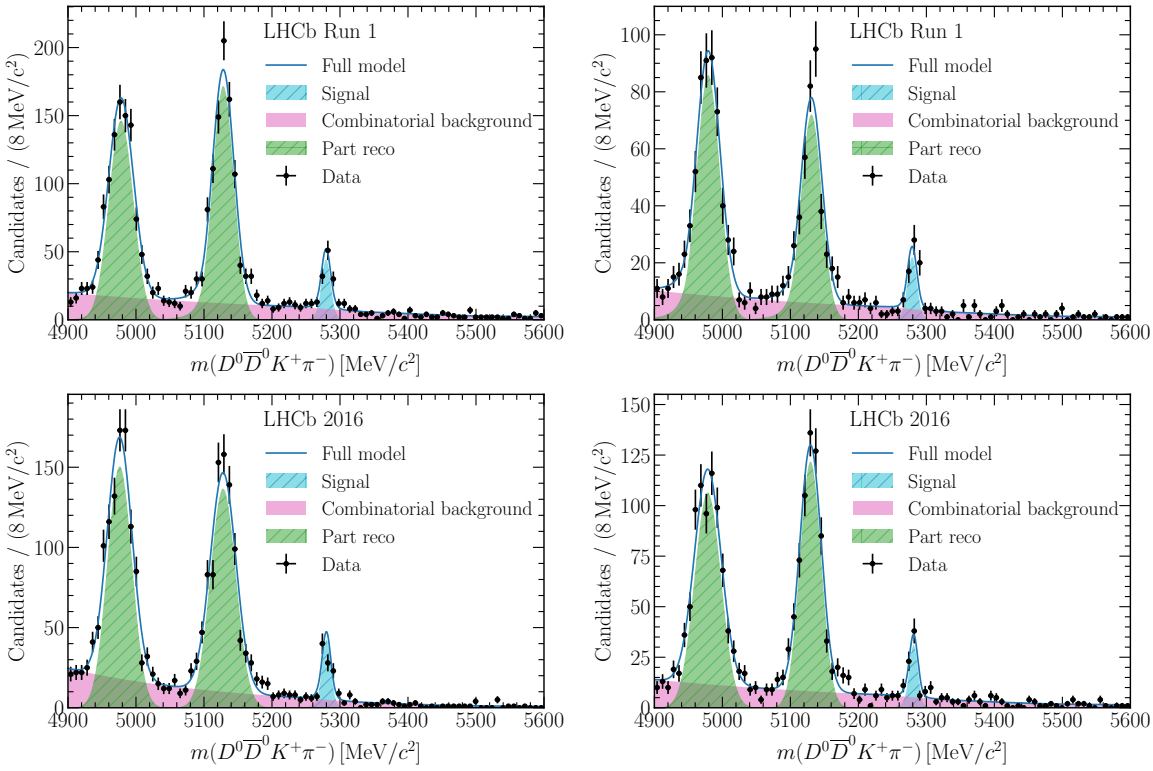


Figure 7.36: Simultaneous fit across signal mode categories in an extended range. Upper left: Run 1 TOS, upper right: Run 1 TIS, lower left: 2016 TOS, lower right: 2016 TIS. The data are shown as black points with error bars and the fit components are as described in the legends.

inal fit and this fit over the extended range took into account the correlations between the two datasets.

7.9 Systematic uncertainties

Sources of systematic uncertainty are considered in this section. Since this analysis calculates a ratio of branching fractions, many uncertainties cancel out between the signal and control mode. Further modelling effects that contribute significant systematic uncertainties are detailed below. A summary is provided in Table 7.14, where the uncertainties are expressed as a relative fraction of the branching fraction ratio $\mathcal{R}$. The total systematic uncertainty, calculated by summing the individual sources in quadrature, is 7.3% relative to $\mathcal{R}$. A common strategy that is employed here is to generate toys or pseudodata, in this case from the nominal model and fit results. This pseudodata can then be fitted with the nominal model and an alternative model, depending on which effect

Table 7.14: Sources of systematic uncertainty expressed as a percentage of the branching fraction ratio $\mathcal{R}$. Statistical uncertainty from mass fits also included for comparison.

Source	Uncertainty
Signal model	5.0
Simulation sample	2.5
Background model	2.0
Fixed parameters	2.0
Charmless backgrounds	2.0
MC reweighting	2.0
NN classifier	2.0
PID transformation	1.2
Stripping efficiency	0.6
Total	7.3
Statistics	7.7

is being measured, and the results compared. The samples are repeatedly produced to generate 500 pseudoexperiments, such that an average difference or root-mean-squared deviation can be calculated. They are generated with high statistics, of order 10^6 events, so that statistical uncertainties can be treated as negligible. All the uncertainties listed below are expressed as a percentage of $\mathcal{R}$.

Fit bias

Pseudoexperiments are used to test whether the fitted value of $\mathcal{R}$ is unbiased. Toy data are generated, where the number of events is taken from Gaussian distributions of the signal and background yields added together for each category. Each sample is fitted for $\mathcal{R}$ using the procedure described in Section 7.8.4. The resulting pull distribution is shown in Figure 7.37, where the mean and error are compatible with 0 and unity, confirming that the fit to $\mathcal{R}$ is not biased. Additionally, to check the consistency and stability of the final fit to the signal mode data, the results from the simultaneous fit can be compared to the individual subsample fits. Table 7.16 shows $\mathcal{R}$ derived from each category, and these are all consistent with the result from the simultaneous fit.

Signal fit model

The dependence on the signal model, based on $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ MC, is measured in two ways, listed below. This is the dominant source of systematic uncertainty of 5%, which is

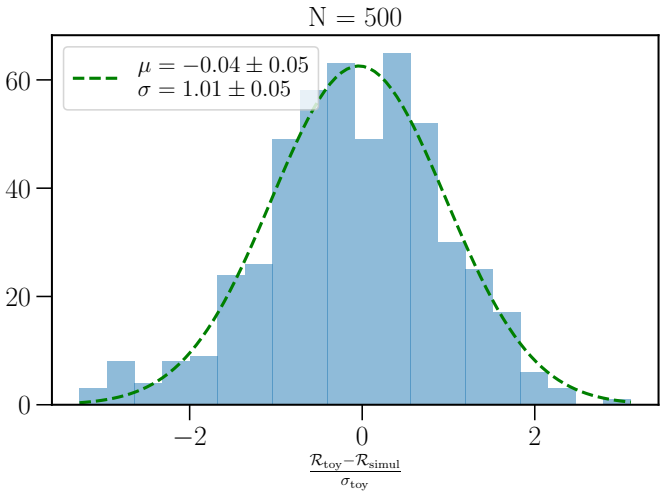


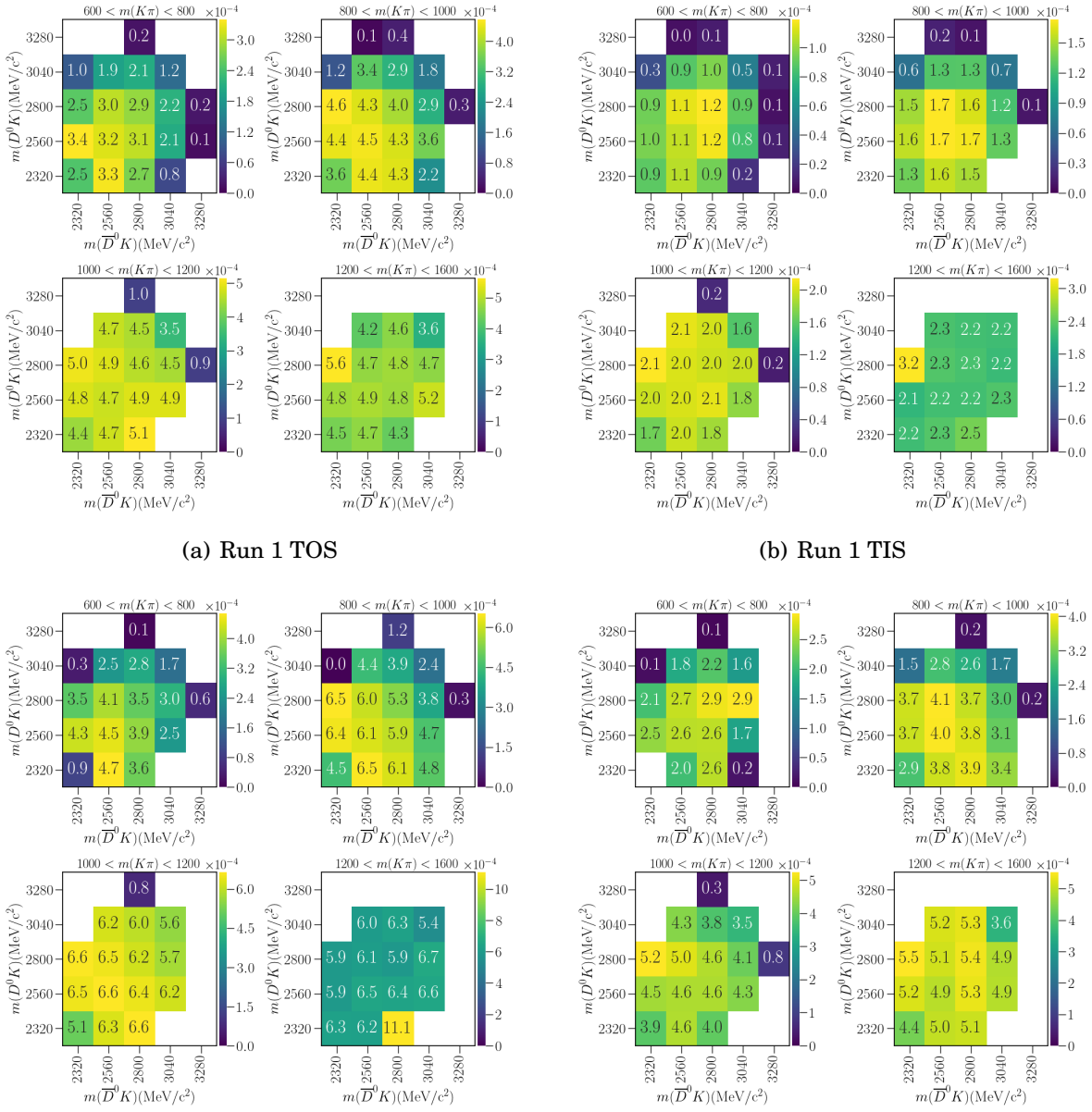
Figure 7.37: Pull distribution showing the difference in the fitted value of $\mathcal{R}$ from $N = 500$ toys and the value obtained from the simultaneous fit $\mathcal{R}_{\text{simul}}$, with respect to the error in the fit. The dashed line represents a Gaussian distribution fitted to the pulls, with the fit results shown in the legend.

nevertheless smaller than the statistical uncertainty on $\mathcal{R}$.

Firstly, the fit to data is repeated, using selection efficiencies derived from $B^0 \to D^0 \overline{D}^0 K^{*0}$ simulation samples. This is compared to the nominal result, and a difference of 0.8% is measured. Secondly, an event-by-event efficiency correction is considered, as an alternative to the global efficiency derived from the phase-space MC sample. Similar to the efficiency correction of the control mode in Section 7.6.1, for the signal mode, the position-dependence is based on a 3-dimensional parameterisation of the phase-space, as

$$\epsilon^{corr} = \frac{\sum_i w_i \epsilon_i(m(D^0K), m(\overline{D}^0K), m(K\pi))}{\mathcal{N}_S}, \tag{7.30}$$

where w_i are the sWeights extracted from the nominal fit to signal mode data, and $\mathcal{N}_S$ is the signal yield. This is done separately in each category, and the binning scheme can be seen in Figure 7.38, where $m(K\pi)$ is split into the regions [(600, 800), (800, 1000), (1000, 1200), and (1200, 1600)] MeV/c^2. Applying this correction and re-doing the fits results in a value of $\mathcal{R}$ that is shifted by 5.0%. An alternative subspace of the full 5D phasespace is also considered, as a function of the position in $(m(D^0K\pi), m(\overline{D}^0K\pi), m(K\pi))$. This alternative correction results in a shift of 2.2%. To be conservative, the larger uncertainty of 5.0% on $\mathcal{R}$ is used.



(a) Run 1 TOS

(b) Run 1 TIS

(c) 2016 TOS

(d) 2016 TIS

Figure 7.38: Overall selection efficiency in the signal mode in bins of $m(D^0K)$ and $m(\overline{D}^0K)$, and subsequently in slices of $m(K\pi)$ as written in subplot titles.

Simulation sample size

The finite size of the simulation samples introduces a statistical uncertainty on the selection efficiencies. These efficiencies are fixed in the mass fit, so their errors introduce a systematic uncertainty on $\mathcal{R}$. The contribution from the two decay modes are considered independently, and the combined uncertainty on $\mathcal{R}$ is 2.5%. For the signal mode, the simultaneous fit is repeated to pseudoexperiments, where ϵ^{sig} is varied within its errors for each fit. The root-mean-square (RMS) deviation from these fits gives an uncertainty of 2.5% on $\mathcal{R}$. For the control mode, the MC sample is sampled via bootstrapping [152] in order to repeatedly derive new corrections to the efficiency, ϵ^{corr} (Equation 7.14). The RMS deviation from this procedure is measured to be 0.5%.

Background fit model

The dependence on the fit model is tested by fitting to high statistics pseudoexperiments with a straight line function as the background model, and comparing this with the nominal exponential model. This is first done in each control mode subsample, and then simultaneously across the signal mode data. The average difference over repeated fits is 2.0%.

Fixed fit parameters

The systematic uncertainty of fixing various parameters in the fits to data is calculated. In the control mode, the tail parameters and the Gaussian mean is fixed from simulation. In the signal mode, the Crystal Ball parameters are fixed from simulation, and the shift in the Gaussian mean and width are fixed to the control mode results. These three sets of variables are sampled in turn from multi-variate Gaussian distributions, where the covariance matrices are taken from the original fits. The fits are then repeated using pseudoexperiments. The root-mean-squared deviation gives an uncertainty of 2.0% of $\mathcal{R}$.

Charmless backgrounds

A systematic uncertainty arises due to the statistical uncertainty in the charmless background component that is subtracted from the signal yield. This is calculated by varying the charmless contamination within its errors in pseudoexperiment fits. The RMS deviation from these fits gives an uncertainty of 2.0% of $\mathcal{R}$.

MC reweighting

The reweighting algorithm is affected by the statistical uncertainty of the simulation and data samples used. The effect of the statistical uncertainty of the simulation samples is already accounted for above. So the effect of the statistics of the data samples is considered by bootstrapping the data and recalculating the weights. This is done separately for each run period and trigger category. The systematic uncertainty on $\mathcal{R}$ is calculated as 2.0%, by repeating the simultaneous fit to pseudodata, using the efficiency ratios from each bootstrapped sample.

NN classifier

In Figures 7.10 and 7.11, a small difference is seen in the classifier response on sWeighted data compared to MC. In order to quantify this effect, a correction of the second stage classifier $B_{\text{selection}}$ response in MC is derived to match data in the control mode. This correction is then applied to the signal mode and a new efficiency ratio calculated in each category. The simultaneous fit to data is then repeated to a high statistic toy sample with these new efficiency ratios to re-derive $\mathcal{R}$. The relative difference is 2.0%, which is taken as the systematic uncertainty.

PID transformation

The biggest difference between the signal and control mode is expected to enter in the particle ID variables of the $\pi_{K^{*0}}$ and $K_{K^{*0}}$. The method of PID transformation which uses kernel density estimation (KDE) to correct the PID response of MC samples has a systematic uncertainty due to the parameterisation of the KDE. This is evaluated by transforming the variables with a modified kernel which is 50% wider, and assessing the difference. The nominal transformed variables [πProbNNπ,KProbNNK] are weighted to match these modified weights, and the resulting change in the ratio of efficiencies is calculated. The systematic uncertainty on $\mathcal{R}$ is calculated by repeating the simultaneous fit to toy data samples using these corrected efficiency ratios in each category. The relative difference in the ratio is 1.2%.

Efficiency of stripping selection

The ratio of stripping efficiencies between the signal and control modes deviates from unity for both Run 1 and 2016. This is expected to be mostly due to the cuts placed on

Table 7.15: Trigger efficiency (%) calculated using the TISTOS method in data and MC for the control mode. Uncertainties are statistical.

	Data		MC	
	Run 1	2016	Run 1	2016
ϵ_{TOS}	45.8 ± 6.2	48.0 ± 6.8	49.1 ± 0.2	52.1 ± 0.7
ϵ_{TIS}	33.8 ± 4.4	29.9 ± 4.0	32.7 ± 0.1	38.3 ± 0.5

the K^{*0} system, which will be tighter on the control mode which has a slow pion. A simulated sample with 1 million generated events in 2016 conditions is used for this study. Comparing the nominal stripping cuts with a sample where no cuts are applied on $Kp_{\text{T}} + \pi p_{\text{T}}$, and other flight distance and vertex requirements are loosened, shows an increase in the stripping efficiency of 11%. Since the signal mode decay nominally has a stripping efficiency that is $\sim 50\%$ higher than the control mode, it can be assumed that most of the remaining discrepancy is due to the underlying reconstruction efficiency.

To consider a systematic uncertainty for the pions that *are* reconstructed, the kinematics (p_T, η) of the slow pion in the control mode are reweighted to match background-subtracted data. This weighting is then applied to the MC sample with loosened cuts. The cut $Kp_{\text{T}} + \pi p_{\text{T}} > 1000\,\text{MeV}/c$ is then placed afresh, and the efficiency of this cut with and without the kinematic weights has a difference of 0.6%. This is taken as the systematic uncertainty due to the stripping selection.

Efficiency of trigger selection

The trigger efficiencies ($\epsilon_{\text{TOS,TIS}}$) of the control mode and signal mode (listed in Tables 7.6 and 7.7) are very similar, *i.e.* their ratio is close to unity. A further check is carried out in the control mode, using the `TISTOS` method [43]. This method allows a direct comparison of the efficiency of trigger selections in data and simulation. The trigger efficiencies calculated using the `TISTOS` method in Table 7.15 show good agreement between data and MC. Considering both these checks, the systematic uncertainty due to the estimation of trigger efficiencies is treated as negligible.

7.10 Results

This thesis presents the first observation of the decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$, published in Ref. [1], and its branching fraction relative to $B^0 \to D^{*-} D^0 K^+$ is measured using LHCb

Table 7.16: Ratio of branching fractions split by data taking period and trigger category, expressed as a percentage (%). Uncertainties are statistical only.

	Run 1		2016	
	TOS	TIS	TOS	TIS
$\mathcal{R}$	12.5 ± 1.8	16.9 ± 2.8	11.8 ± 1.7	12.3 ± 2.5

data from Run 1 and 2016. This measurement uses the full kinematically-allowed range of $B^0 \to D^0\overline{D}^0K^+\pi^-$ outside of the D^{*-} region. In total 297 ± 14 signal and 1697 ± 42 control mode decays are found. The signal yield quoted is extracted from the simultaneous fit to data across the four categories defined previously, such that the overall uncertainty is smaller than adding the individual fit yields in quadrature. The ratio of branching fractions $\mathcal{R}$ is defined in Equation 7.6, and is derived from a simultaneous fit across the signal mode data to be

$$\mathcal{R} = (14.2 \pm 1.1\,(\text{stat}) \pm 1.0\,(\text{syst}))\%, \tag{7.31}$$

where the first uncertainty is statistical and the second is systematic. Using the yields extracted from fits to the individual subsamples, minus the expected charmless contaminations, gives the separate $\mathcal{R}$ values listed in Table 7.16. These results are compatible with the combined result.

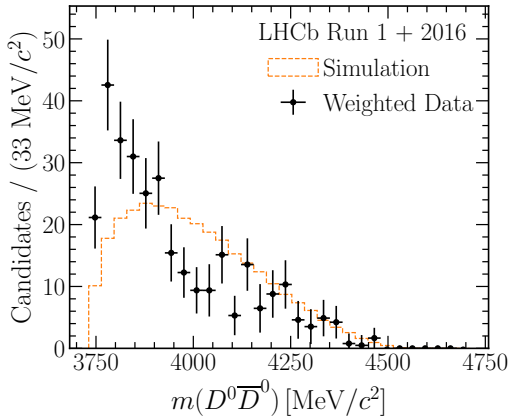
Substituting the known branching fraction of $B^0 \to D^{*-}D^0K^+$ [84], the absolute branching fraction of the signal mode is thus measured to be

$$\mathcal{B}(B^0 \to D^0\overline{D}^0K^+\pi^-) = (3.50 \pm 0.27\,(\text{stat}) \pm 0.26\,(\text{syst}) \pm 0.30(\mathcal{B}')) \times 10^{-4}, \tag{7.32}$$

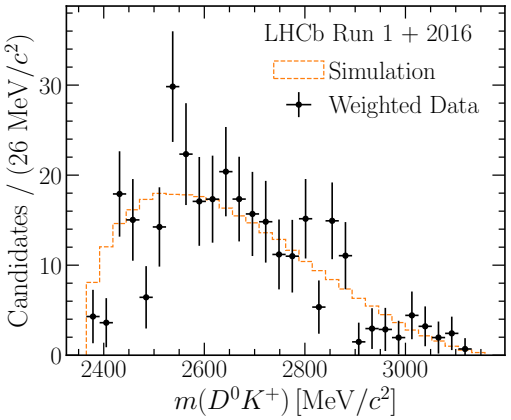
where the third uncertainty is due to the uncertainty on the control mode branching fraction (labelled by $\mathcal{B}'$).

7.10.1 Projections

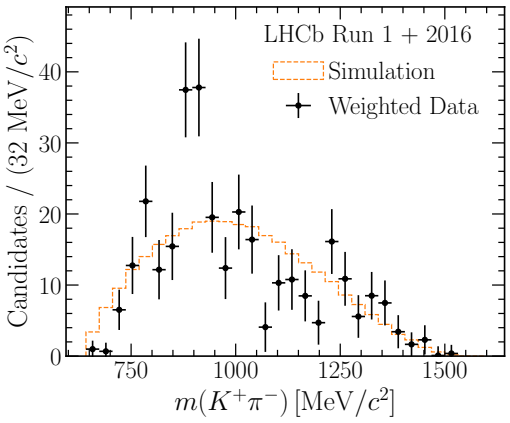
Resonances are expected to be seen in the data, most likely in the invariant mass spectra of $m(D^0\overline{D}^0)$, $m(D^0K^+)$ and $m(K^+\pi^-)$. These projections are shown in Figure 7.39. A phase-space model of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays is also overlaid, to represent what the spectra are expected to look like without the presence of any intermediate resonant states. Possible structures are visible at the masses of the states $\psi(3770)$ in $m(D^0\overline{D}^0)$, at $D^*_{s2}(2573)^+$ and $D^*_{s(1,3)}(2860)^+$ in $m(D^0K^+)$, and $K^*(892)^0$ in $m(K^+\pi^-)$. However, a full amplitude analysis is needed to verify the presence of these resonances as the structures



(a) Invariant mass distribution of $m(D^0\overline{D}^0)$



(b) Invariant mass distribution of $m(D^0K^+)$



(c) Invariant mass distribution of $m(K^+\pi^-)$

Figure 7.39: Projections of invariant mass distributions of sWeighted data (black points), combined across all categories of the signal mode, overlaid with $B^0 \to D^0\overline{D}^0K\pi$ MC (orange dashed line). The MC sample is scaled to match the data statistics.

could be due to reflections and other interference effects. An amplitude analysis will also allow for the extraction of the charm loop contribution above the open charm threshold, to be used as input to measurements of $b \to s\ell^+\ell^-$ processes. Chapter 9 presents work towards an amplitude analysis, which is the next step in the study of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays.

CHAPTER 8

AMPLITUDE ANALYSIS FORMALISM

An amplitude analysis is a study of the dynamics of a decay in order to determine properties of the processes that mediate the decay. The multi-body decay of a B meson to a final state f typically proceeds via a number of intermediate resonant states. An amplitude analysis measures the relative rates of the various possible paths to the final state, thereby allowing the study of the strong decays of the intermediate resonant states as well as their intrinsic properties. Relevant definitions related to resonant states are provided in Section 2.3. This chapter describes the formalism used in amplitude analyses, starting with important definitions and descriptions of multi-body decays, and then putting it into the context of the four-body decay $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$.

8.1 Multi-body decay kinematics

Amplitude analyses generally consider spinless particles in the initial and final states, *i.e.* the decay of a spin-0 particle, such as a pseudoscalar B meson, decaying to N (pseudo)scalar particles. The discussion here assumes this is the case throughout. The differential decay rate is given by

$$d\Gamma \propto |\mathcal{M}|^2 d\Phi, \tag{8.1}$$

where $\mathcal{M}$ is the matrix element, which encodes the dynamics of the decay, and $d\Phi$ represents the phase-space of the decay. The goal of an amplitude analysis is to find an

expression for $\mathcal{M}$, by fitting a model to experimental data that can describe the various contributions to the total decay amplitude.

The dimensionality of the phase-space is dependent on the number of final state particles. The 4-vectors of the daughters, p_i, give $4N$ degrees of freedom, while the knowledge of their masses imposes N constraints as $p_i^2 = m_i^2$. Conservation of energy and momentum, given by $p_B = \sum_i p_i$, imposes another 4 constraints. A further 3 degrees of freedom can be integrated out due to arbitrary rotations, since there is no dependence on the orientation in space. Therefore the number of degrees of freedom of an N-body decay, and therefore the number of independent variables required to describe the decay, is $3N - 7$.

Three-body decays have two degrees of freedom, which are conventionally chosen to be a pair of squared invariant masses. The two-dimensional distribution of these invariant masses is known as a Dalitz plot [157], and much of the information of the intermediate resonances can be visualised with it. For example, for the decay $B \to abc$, the sum of the squared invariant masses are constrained by

$$m_{ab}^2 + m_{ac}^2 + m_{bc}^2 = m_B^2 + m_a^2 + m_b^2 + m_c^2, \tag{8.2}$$

where $m_{ab}^2 = (p_a + p_b)^2$. In the absence of any resonances, the two-dimensional distribution will be uniform, contained within a boundary defined by the kinematically allowed region due to conservation of momentum. A contribution to the decay from a resonant state R, such as $B \to R(\to ab)c$, will produce a band in the m_{ab}^2 distribution at the value of m_R^2. This is seen as a band of increased density in the Dalitz plot. An example of such a Dalitz plot with a single resonant band is displayed in Figure 8.1. Four-body decays, which are the focus of this thesis, have five degrees of freedom, resulting in added complexity in both the choice of independent variables and the visualisation of the dynamics.

8.2 Lineshapes

There are two main factors to consider when formulating the lineshape of a resonant state, the barrier factors and the parameterisation of its mass distribution. In the following, the example decay $B^0 \to K^{*0}(\to K^+\pi^-)X$, where X is a placeholder state, will be used. The momentum of the K^{*0} in the B^0 rest frame is labelled p, and the momentum of the daughters (K^+ or π^-) in the K^{*0} rest frame is labelled q. For a meson of radius d, the decay products are limited by their linear momentum q, in generating

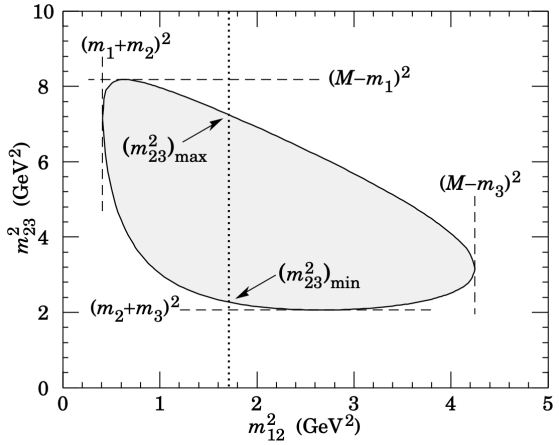


Figure 8.1: Schematic of a Dalitz plot for a three-body decay of a particle with mass M, such as $P_0 \to P_1P_2P_3$. The shaded area within the solid line is the kinematically allowed region of the decay phase-space, with the various boundaries indicated by dashed lines. The vertical dotted line represents an intermediate resonance such as $R \to P_1P_2$. Figure from Ref. [2].

sufficient angular momentum L, in order to conserve the spin of the resonance. This spin-dependent suppression is encoded by the Blatt-Weisskopf barrier factors [161], B'_L. These are given, up to a $L = 2$ (D-wave), in Equation 8.3, where q_0 is the momentum when $m_{K\pi} = m_0$ (in this case $m_0 = m_{K^{*0}}$) and $z = (|q|d)^2$.

$$\begin{aligned} B'_0(q,q_0) &= 1 \\ B'_1(q,q_0) &= \sqrt{\frac{1+z_0}{1+z}} \\ B'_2(q,q_0) &= \sqrt{\frac{(z_0-3)^2+9z_0}{(z-3)^2+9z}}\,. \end{aligned} \tag{8.3}$$

For isolated and narrow resonances, the most common parameterisation of the mass distribution is the relativistic Breit–Wigner (BW) function. The amplitude, for the $K\pi$ state for example, is given by

$$BW(m_{K\pi}) = \frac{1}{m_0^2 - m_{K\pi}^2 - i m_0 \Gamma(m_{K\pi})}, \tag{8.4}$$

where the mass-dependent width is

$$\Gamma(m_{K\pi}) = \Gamma_0 \left(\frac{q}{q_0}\right)^{2L_{K^{*0}}+1} \frac{m_0}{m} B'_{L_{K^{*0}}}(q,q_0)^2\,. \tag{8.5}$$

The overall lineshape of the $K\pi$ resonance is then given by

$$R(m_{K\pi}) = B'_{L_{B^0}}(p, p_0)\left(\frac{p}{m_{B^0}}\right)^{L_{B^0}} \times BW(m_{K\pi}) \times B'_{L_{K^{*0}}}(q, q_0)\left(\frac{p}{m_{K\pi}}\right)^{L_{K^{*0}}}, \tag{8.6}$$

where Blatt-Weisskopf factors for both the production and decay of the resonance are included.

8.3 Helicity basis

Various descriptions of the amplitude structure of a process differ by their definition of a particle's spin projection in its rest frame, and the choice of formalism dictates how decay amplitudes may be expanded. A brief description of the construction of helicity states is outlined here, such that the decay amplitudes in the helicity basis may be presented in the following section. A detailed explanation of the helicity formalism can be found in Refs. [162, 163].

The helicity formalism, developed by Jacob and Wick [164], is used in the treatment of decay amplitudes of states with definite angular momentum and helicity. The helicity operator is defined in terms of the spin and momentum operators as

$$\hat{h} = \frac{\hat{\mathbf{S}} \cdot \hat{\mathbf{p}}}{p}, \tag{8.7}$$

and is invariant under rotations as well as boosts along $\mathbf{p}$. These properties mean that the helicity formalism is useful for the description of angular decay distributions of $b \to s\ell^+\ell^-$ processes such as $B^0 \to K^{*0}\mu^+\mu^-$. Therefore, the amplitude analysis in this thesis will follow the same formalism.

A single particle state, with spin j and spin projection in the z-axis m, is conventionally specified by a canonical basis vector $|jm\rangle$. This may be acted upon by a rotation operator, defined as

$$R(\alpha, \beta, \gamma) = e^{-i\alpha J_z} e^{-i\beta J_y} e^{-i\gamma J_z}, \tag{8.8}$$

where (α, β, γ) are the standard Euler angles. The rotation of this state is given by

$$\hat{R}(\alpha, \beta, \gamma)|jm\rangle = \sum_{m'=-j}^{j} D^j_{m'm}(\alpha, \beta, \gamma)|jm'\rangle, \tag{8.9}$$

where the Wigner-D functions [165], $D^{j}_{m'm}$, are rotation matrices that form a unitary representation of the group SO(3). The Wigner-D functions are proportional to the well-known spherical harmonics, and are defined as

$$D^{j}_{m'm}(\alpha,\beta,\gamma) = \langle jm'|e^{-i\alpha J_z}e^{-i\beta J_y}e^{-i\gamma J_z}|jm\rangle. \tag{8.10}$$

A related transformation of the canonical state can give the connection to the helicity basis, $|j\lambda\rangle$. The helicity quantum number λ is the projection of a particle's spin along the direction of its momentum $\mathbf{p}$. The angular convention may be defined for any given particle decay, and for this general case the standard spherical coordinates are used, with polar angle θ and azimuthal angle ϕ. The helicity state (with its momentum specified) $|\mathbf{p},j\lambda\rangle$ is obtained by rotating the rest state $|0,j\lambda\rangle$ to point along $\hat{\mathbf{p}}(\theta,\phi)$, and then applying a Lorentz boost along $\hat{\mathbf{p}}(\theta,\phi)$. Therefore, the helicity state can be defined in terms of the canonical state as

$$|j\lambda\rangle = \sum_{m=-j}^{j} D^{j}_{m\lambda}(\phi,\theta,-\phi)|jm\rangle\,. \tag{8.11}$$

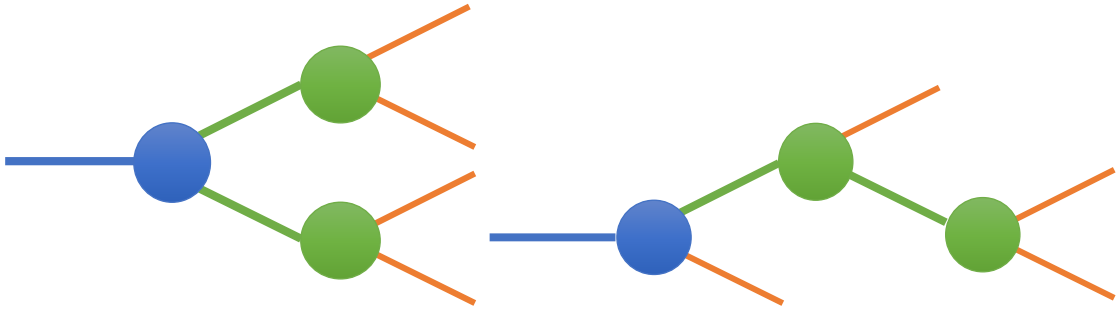
This formalism can be directly extended to a two-particle state, such as the decay $A \to B_1B_2$. The initial particle has quantum numbers J_A and M_A, and decays to two particles with helicities λ_1 and λ_2. In the centre-of-mass frame, the two-particle state can be defined in spherical coordinates p,θ,ϕ, where $p = |\mathbf{p_1}| = |\mathbf{p_2}|$ and (θ,ϕ) define the direction of $\mathbf{p_1}$. Therefore the helicity state of a two-particle state is defined by $|\theta,\phi,\lambda_1,\lambda_2\rangle$, where the momentum is suppressed for brevity. The relationship to the canonical basis is equivalent to the single particle case, with an additional normalisation factor, given by

$$|\theta,\phi,\lambda_1,\lambda_2\rangle = \sum_{J_A,M_A=-J_A}^{M_A=J_A} \sqrt{\frac{2J_A+1}{4\pi}}\, D^{J_A}_{M_A\lambda}(\phi,\theta,-\phi)|J_A,M_A,\lambda_1,\lambda_2\rangle, \tag{8.12}$$

where $\lambda = \lambda_1 - \lambda_2$.

8.4 Decay amplitudes

A multi-body decay amplitude is usually treated as a coherent sum over successive two-body decay amplitudes, which is known as the *isobar* model. This is an empirical approach in which any final-state interactions are considered negligible. For a four-body decay, an isobar approach takes the form of a quasi two-body decay, or a cascade



(a) Quasi two-body topology (b) Cascade topology

Figure 8.2: Diagrams of four-body decay topologies under the isobar model. Each state (circle) decays to two particles, with the green circles representing intermediate resonances.

decay. These decay topologies are shown in Figure 8.2. For the purposes of charm loop studies, it is instructive to consider the quasi two-body approach wherein the decay $B^0 \to D^0\overline{D}^0K^+\pi^-$ takes the form $B^0 \to K^*(\to K^+\pi^-)\psi'(\to D^0\overline{D}^0)$. Here K^* represents any $K\pi$ S-wave or P-wave state, and ψ' represents any $c\overline{c}$ state decaying to $D^0\overline{D}^0$. In this way, the formalism is analogous to the treatment of $B^0 \to K^{*0}\mu^+\mu^-$ decays as $B^0 \to K^*(\to K^+\pi^-)\gamma(\to \mu^+\mu^-)$. Considering two-body decays also allows direct use of the helicity basis described above. The helicity formalism applied to $b \to s\ell^+\ell^-$ processes is described in Ref. [166], and the following setup follows a similar procedure.

Each combination of resonances that contributes to the decay has a specific decay amplitude, $\mathcal{A}$, which can be written in the helicity basis as a product of

1. the two-body helicity amplitude, $\mathcal{H}_\lambda$,
2. the resonant lineshape,
3. the corresponding Wigner-D function.

An example of a single decay amplitude is shown in Equation 8.14. The overall matrix element of a decay is formed from a *coherent* sum of each decay amplitude, where each one is weighted by a complex coefficient, C_r, *i.e.*

$$\mathcal{M} = \sum_r C_r\mathcal{A}_r = \sum_r c_r \exp(i\phi_r)\mathcal{A}_r\,. \tag{8.13}$$

An amplitude analysis is concerned with fitting the two free parameters of the complex coefficient, the magnitude c_r and phase ϕ_r. They may also be expressed as the real and imaginary parts of C_r.

The two-body decay amplitude for the generic decay $K_{J_K} \to k_1k_2$ can be written as

$$\mathcal{A}_K = \sum_{\lambda_K} \mathcal{H}_{\lambda_{k_1}\lambda_{k_2}} R_K(m_{k_1k_2}) D^{J_K}_{\lambda_K,\lambda_{k_1}-\lambda_{k_2}}(\Omega_K), \tag{8.14}$$

where $\mathcal{H}_{\lambda_{k_1}\lambda_{k_2}}$ is the helicity coupling, R_K is the resonant lineshape including barrier factors as defined in Equation 8.6. $D^{J_K}_{\lambda_K,\lambda_{k_1}-\lambda_{k_2}}$ is the Wigner-D function dependent on the spin of the resonance J_K, defined in Equation 8.10. The angular information of the resonance is contained by Ω_K, which defines its direction. Accordingly, the generalised amplitude for the decay $B_{J_B} \to K_{J_K}(\to k_1 k_2)\psi_{J_\psi}(\to D_1 D_2)$ has contributions from each two-body decay, $B_{J_B} \to K\psi$, $K_{J_K} \to k_1 k_2$, and $\psi_{J_\psi} \to D_1 D_2$. There is no contribution from the lineshape of the B meson, as it is the parent particle, so there is no dependence on its mass. The overall amplitude is given by their product, *i.e.*

$$\begin{aligned}\mathcal{A}_{K\psi} = \sum_{\lambda_K\lambda_\psi} &\mathcal{H}_{\lambda_K\lambda_\psi} D^{J_B}_{\lambda_B,\lambda_K-\lambda_\psi}(\Omega_B)\times \\ &\mathcal{H}_{\lambda_{k_1}\lambda_{k_2}} R_K(m_{k_1k_2}) D^{J_K}_{\lambda_K,\lambda_{k_1}-\lambda_{k_2}}(\Omega_K)\times \\ &\mathcal{H}_{\lambda_{D_1}\lambda_{D_2}} R_\psi(m_{D_1D_2}) D^{J_\psi}_{\lambda_\psi,\lambda_{D_1}-\lambda_{D_2}}(\Omega_\psi). \end{aligned} \tag{8.15}$$

For the decay of interest in this analysis, $B^0 \to K^*(\to K^+\pi^-)\psi'(\to D^0\overline{D}^0)$, the generalised amplitude in Equation 8.15 simplifies due to knowledge of the particle spins, similarly to Ref. [166]. Firstly, $J_B = 0$ sets $\lambda_B = 0$, which implies $\lambda \equiv \lambda_{K^*} = \lambda_{\psi'}$, and the corresponding Wigner-D function equals unity. Furthermore, as all final state particles are scalar, the helicity amplitudes for the intermediate resonances reduce to scalar coupling constants. Therefore, the decay amplitude is given by

$$\begin{aligned}\mathcal{A}_{K^*\psi'} \propto \sum_{\lambda=-J_{\psi'}}^{J_{\psi'}} &\mathcal{H}^{B^0\to K^*\psi'}_{\lambda} \times \\ &R_{K^*}(m_{k_1k_2}) D^{J_{K^*}}_{\lambda,0}(\Omega_{K^*})\times \\ &R_{\psi'}(m_{D_1D_2}) D^{J_{\psi'}}_{\lambda,0}(\Omega_{\psi'}). \end{aligned} \tag{8.16}$$

The number of terms in the helicity sum is defined by the spin state of the resonances in the decay chain. In this setup ψ' is by definition a $J^P = 1^-$ vector meson, for example the $\psi(3770)$. This means all the associated decay amplitudes are formed by a sum over three helicity states, $\lambda = -1, 0, 1$, given by $\mathcal{A}_-, \mathcal{A}_0, \mathcal{A}_+$, respectively. While it is conventional to sum over the helicity states, it is also possible to measure each component, and this is key here for the application to $b \to s\ell^+\ell^-$ measurements. In order to better understand the impact of charm loop interferences in $B^0 \to K^{*0}\mu^+\mu^-$ measurements, the relative phase of each $c\overline{c}$ resonance helicity component needs to be measured in the amplitude analysis of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays.

There are additional decay topologies that may be considered, with different pairs of two-body decays. Firstly, the charmonium decay may not necessarily be from a vector

meson, for example the $\chi_{c2}(3930)$ which has $J^P = 2^+$. In this case, the decay amplitude would simply be taken as the sum over the helicity states. Secondly, there may be contributions from $D_{sJ} \to D^0 K^+$ decays, for an overall decay $B^0 \to D_{sJ}(\to D^0 K^+) NR(\to \overline{D}^0 \pi^-)$, where NR represents a non-resonant state. The associated decay amplitudes are also added coherently to the overall matrix element.

8.5 Implementation

The amplitude analysis in this thesis uses the `AmpGen` [167] package, which performs both generation and fitting of multi-body decays using the isobar model. `AmpGen` performs amplitude fits in an efficient and parallelised manner, allowing for complicated amplitude models of three and four-body decays. In addition to fitting for the amplitude couplings, parameters of the resonant lineshapes may also be fitted. The `AmpGen` software is integrated within the LHCb simulation framework and has been used in other LHCb four-body analyses, such as Ref. [168]. A few details of the implementation of an amplitude fit, in general and in `AmpGen`, are described here.

There are a number of additional lineshapes available beyond the relativistic Breit-Wigner. Since the $\psi(3770)$ mass is close to the $D^0\overline{D}^0$ threshold, a Flatté distribution may be used to better describe its lineshape. This is because the proximity to threshold distorts the lineshape from the simple BW form [169, 170]. For non-resonant contributions, such as a non-resonant $D^0\overline{D}^0$ component, a uniform shape may be used, which simply models the mass distribution uniformly, given the phase-space of the decay. An S-wave $K\pi$ component is also of interest to this analysis, and this can be modelled using the LASS parameterisation [171]. This parameterisation is chosen as it models both the broad resonance $K_0^*(1430)^0$, and the non-resonant spin-0 contribution [172] in the $K\pi$ spectrum.

The amplitude fit is carried out to the overall signal PDF, which is the modulus squared of the matrix element defined in Equation 8.13, with the amplitudes of different helicities added *incoherently*. The amplitude $\mathcal{A}_{K^*\psi'}$ for the decay $B^0 \to K^{*0}\psi'$ in Equation 8.16 can be generalised for any combination of two-body decays, $\mathcal{A}_r$, such that the signal PDF is given by

$$\mathcal{P}(x) = \sum_\lambda |\mathcal{M}|^2 = \sum_\lambda \left| \sum_r C_{r,\lambda} \mathcal{A}_{r,\lambda}(x) \right|^2 , \tag{8.17}$$

where x represents a point in the phase-space of the decay, and λ is the helicity index as defined in the previous section. `AmpGen` can use the canonical helicity formalism, with a small adjustment to the basis states. A set of transversity states, $\mathcal{A}_0, \mathcal{A}_\parallel, \mathcal{A}_\perp$, are defined

with the simple relations

$$\mathcal{A}_0 = \mathcal{A}_0\,, \tag{8.18}$$

$$\mathcal{A}_\parallel = \frac{\mathcal{A}_+ + \mathcal{A}_-}{\sqrt{2}}\,, \tag{8.19}$$

$$\mathcal{A}_\perp = \frac{\mathcal{A}_+ - \mathcal{A}_-}{\sqrt{2}}\,. \tag{8.20}$$

The absolute values of the magnitude and phases of the complex coefficients are dependent on the model and convention used, and are defined with respect to one another. One contribution must be fixed in the amplitude fit, for example the $B^0 \to K^{*0}(892)\psi(3770)$ contribution, with magnitude 1 and phase 0. Each amplitude has a CP-conjugate amplitude, $\overline{\mathcal{A}}_r(x) = \mathcal{A}_r(\overline{x})$, where $\overline{x}$ represents the CP-transformed point in phase-space, obtained by considering oppositely charged particles and reversed three-momenta. If the matrix element of the B^0 decay is written as

$$\mathcal{M}_{B^0}(x) = \sum_r C_r \mathcal{A}_r(x)\,, \tag{8.21}$$

then the conjugate decay process is given by

$$\mathcal{M}_{\bar{B}^0}(x) = \sum_r \overline{C}_r \mathcal{A}_r(\overline{x}) = \sum_r C_r \mathcal{A}_r(\overline{x})\,, \tag{8.22}$$

where the last equality comes from the assumption of CP-conservation. This assumption follows from the fact that the relevant CKM elements have highly suppressed contributions from the CP-violating phase in the SM. Therefore, the B^0 and $\bar{B}^0$ decays can be combined in the amplitude fit, as long as each particle is swapped to its antiparticle and the three-momenta have their signs flipped in the $\bar{B}^0$ decay.

As the amplitude model is dependent on the formalism, the information contained by the relative magnitudes and phases of each decay amplitude can also be expressed as *fit fractions*. Each contribution's fit fraction is given by

$$F_{r,\lambda} = \frac{\int \left|C_{r,\lambda}\mathcal{A}_{r,\lambda}(x)\right|^2 dx}{\int \mathcal{P}(x)dx}\,, \tag{8.23}$$

which can be interpreted as its relative branching fraction in the limit of no interference. When there is interference between contributions, the *interference fit fraction* describes the net constructive or destructive interference contribution to the total amplitude, defined as

$$F_{rs,\lambda} = \frac{\int 2\mathrm{Re}[C_{r,\lambda}C^*_{s,\lambda}\mathcal{A}_{r,\lambda}(x)\mathcal{A}^*_{s,\lambda}(x)]dx}{\int \mathcal{P}(x)dx}\,. \tag{8.24}$$

By definition, the fit fractions and interference fit fractions sum up to unity,

$$\sum_{r,\lambda} F_{r,\lambda} + \sum_{r<s,\lambda} F_{rs,\lambda} = 1. \tag{8.25}$$

8.5.1 Likelihood fit

The amplitude fit aims to perform the usual minimisation of the negative log-likelihood, where the likelihood is given by

$$\mathcal{L}(\Theta) = \prod_{x_i} \frac{\epsilon(x_i)\phi(x_i)\mathcal{P}(x_i;\Theta)}{I} \tag{8.26}$$

where Θ denotes the fit parameters, $\epsilon(x_i)$ is the efficiency and $\phi(x_i)$ is the phase-space density at point x_i, and I is the normalisation factor given by integrating the signal PDF over the phase-space,

$$I = \int dx\, \epsilon(x)\phi(x)\mathcal{P}(x;\Theta). \tag{8.27}$$

The factors of $\epsilon(x)\phi(x)$ factorise since they do not depend on the fit parameters Θ, such that they only enter into the normalisation integral. Monte Carlo integration is used to evaluate the normalisation, by considering the probability distribution of N integration events generated with a model labelled $\mathcal{P}'(x)$, which are processed and selected in an identical manner to the real data. A detailed derivation can be found in Appendix A.1, which shows that the normalisation integral simplifies to

$$I = \frac{1}{N}\sum_{i=0}^{N} \frac{\mathcal{P}(x_i)}{\mathcal{P}'(x_i)}, \tag{8.28}$$

with an error proportional to $\frac{1}{\sqrt{N}}$. This approach eliminates the need to parameterise the selection efficiency as a function of the position in phase-space, as long as an MC integration sample that has undergone the full reconstruction and selection procedure is used [173]. Any data-driven corrections applied to the MC on a per-event level, such as those described in Section 7.5, are absorbed into the definition of $\mathcal{P}'(x_i)$. In this analysis a phase-space MC sample of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays is used for the integration events, where no intermediate resonances are specified in order to have $\mathcal{P}'(x)$ as close to $\mathcal{P}(x)$ as possible, without imposing any prior knowledge of the decay model. Clearly, a large integration sample (large N) is advantageous to reduce the uncertainty in the fit result.

CHAPTER 9

AMPLITUDE ANALYSIS OF $B^0 \to D^0\overline{D}^0K^+\pi^-$ DECAYS

This chapter presents the amplitude analysis of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays, following on from the first observation and branching fraction measurement presented in Chapter 7. As outlined in Chapter 5, $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays offer the opportunity for spectroscopy and charm loop studies, through a full amplitude analysis. Initial sensitivity studies are presented in this thesis, and future plans are discussed.

The amplitude analysis will be performed with the full LHCb dataset from Run 1 and Run 2, and these samples are detailed in Section 9.1. Section 9.2 presents sensitivity studies performed with toy simulation samples, and plans for future work are outlined in Section 9.3.

9.1 Data samples

A detailed description of the event selection for $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays is presented in Section 7.3 in the context of the branching fraction measurement. The following amplitude analysis employs a very similar event selection strategy. In addition to the data from Run 1 and 2016 already presented, the amplitude analysis will make use of data collected by LHCb during 2017 and 2018. This corresponds to an additional 1.7 fb^{-1} and 2.2 fb^{-1} of data collected at a centre-of-mass energy $\sqrt{s}$ =13 TeV, bringing the total integrated luminosity used in the amplitude analysis to 8.7 fb^{-1}. Another major addition to the data sample in this analysis is the inclusion of modes where the D^0 mesons are reconstructed from the decay $D^0 \to K^-\pi^+\pi^+\pi^-$, also written as $D^0 \to K3\pi$. These modes

Table 9.1: Signal yields from fits to the B^0 mass in $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays, split by data taking period and trigger category. The signal mode yields contain the remaining charmless contributions.

	Run 1	2016	2017	2018
$\mathcal{N}_S$ (TOS)	100 ± 13	127 ± 15	149 ± 14	150 ± 14
$\mathcal{N}_S$ (TIS)	55 ± 9	55 ± 9	77 ± 11	69 ± 10

are expected to make a significant contribution to the overall yield since the branching fraction $\mathcal{B}(D^0 \to K^-\pi^+\pi^+\pi^-) = (8.23 \pm 0.14)\%$ [2], which is roughly twice the branching fraction of $D^0 \to K^-\pi^+$ decays. On the other hand, the increase in the number of tracks results in a decrease in the reconstruction efficiency for these decays. The mode with both D^0 mesons decaying to $K3\pi$ is not included due to the further decrease in reconstruction efficiency and the large increase expected in combinatorial background contamination. Therefore the complete decay modes included in this analysis are as follows

$$B^0 \to (K^-\pi^+)_{D^0}(K^+\pi^-)_{\overline{D}^0} K^+\pi^-, \tag{9.1}$$

$$B^0 \to (K^-\pi^+\pi^+\pi^-)_{D^0}(K^+\pi^-)_{\overline{D}^0} K^+\pi^-, \tag{9.2}$$

$$B^0 \to (K^-\pi^+)_{D^0}(K^+\pi^-\pi^+\pi^-)_{\overline{D}^0} K^+\pi^-. \tag{9.3}$$

In this thesis, results are shown for all years, but only for the $D^0 \to K^-\pi^+$ mode.

The phase-space sample of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ simulated decays is used throughout the amplitude analysis. As before, MC is used as the signal proxy in the classifier training, and it is used to obtain the signal shape of the B^0 invariant mass distribution. It is also used to study the efficiency effects during the amplitude fit procedure, as described in Section 8.5.

The event selection procedure is mostly unchanged from the branching fraction analysis, with a few adjustments detailed in Appendix A.2. The mass fits as in Section 7.8, are shown in Appendix A.3, for all run period and trigger categories. The signal yields from the mass fits are listed here in Table 9.1. The total yield of these samples, given by the sum of the sWeights, is 783 signal candidates. The $D^0 \to K3\pi$ modes are expected to double this yield, such that the full amplitude analysis will have ~ 1600 signal candidates.

9.1.1 Mass projections

It is instructive to look at the projections of the 1-dimensional invariant mass distributions with $8.7\,\text{fb}^{-1}$ of Run 1 and Run 2 data, now amounting to nearly 800 signal

candidates, accounting for only $D^0 \to K^-\pi^+$ decays. These are shown in Figure 9.1. The potential hints of resonances seen in $m(D^0\overline{D}^0)$, $m(D^0K^+)$ and $m(K^+\pi^-)$ will guide the development of a full amplitude model that best describes the data. As in Figure 7.39, it appears that a state is present at the mass of the $\psi(3770)$ in $m(D^0\overline{D}^0)$. With the more than twofold increase in data compared to the branching fraction analysis, there now also appears to be a state at the mass of the $\chi_{c2}(3930)$. The potential resonant states are indicated in the distributions at their world average mass values [2]. Figure 9.1 also shows the distribution of $m(\overline{D}^0K^+)$, where no resonances are expected. Evidently, however, there appears to be an excess observed in the bin near 2575 MeV/c^2. The subsequent amplitude analysis will be able to determine whether this excess represents a genuine resonant state.

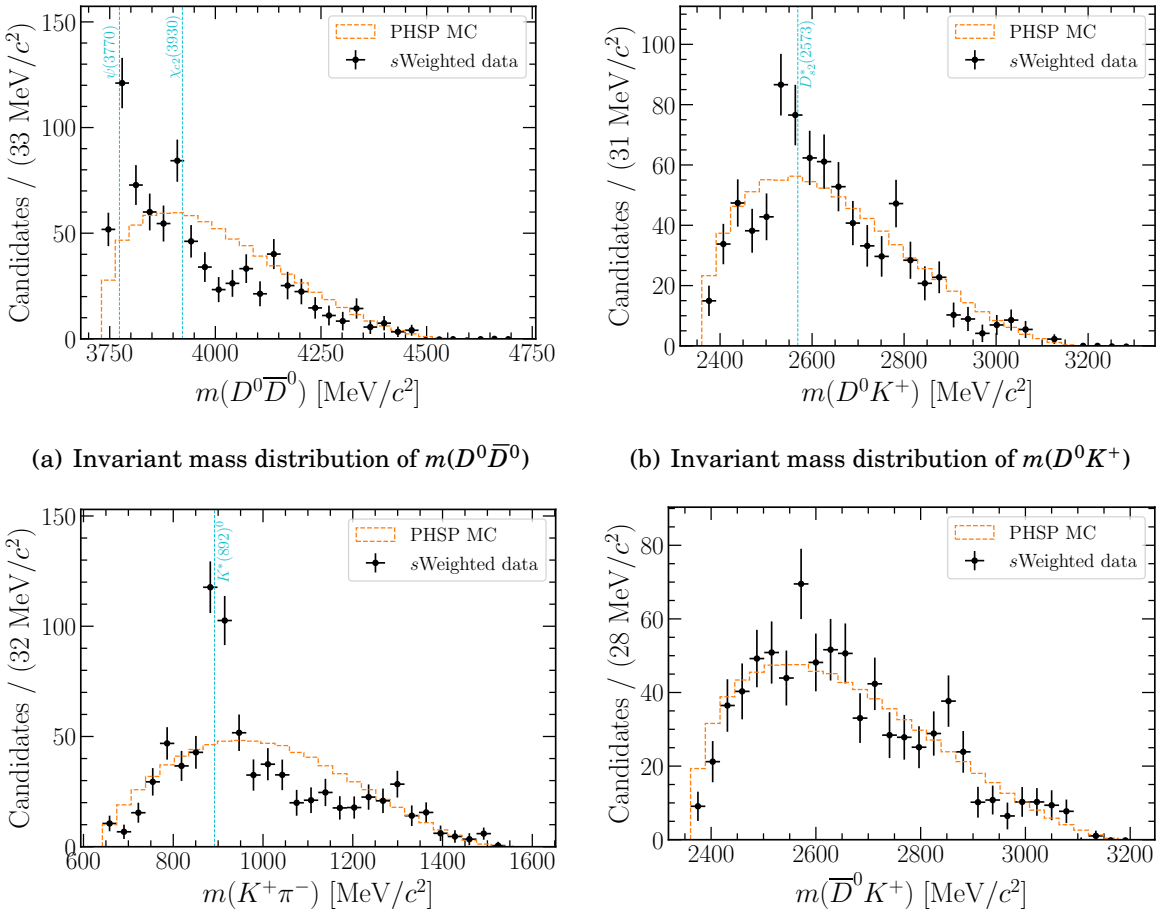
9.2 Toy studies

A toy study is performed in order to test the feasibility of performing an amplitude analysis of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays with 1600 signal candidates. A pseudodata, or *toy*, sample is generated according to a certain amplitude model, and then fitted back with the same model. This gives an indication of how well the various contributions to the model can be extracted from a similar fit to real data. The process is repeated N times to build up an ensemble of toys, called a toy study, and the distribution of fit parameters is examined. For an unbiased fit with correctly estimated uncertainties, the fit parameters will be normally distributed, for large N. The pull of a parameter, a, is given by

$$p_i = \frac{a_i - a_{\text{exp}}}{\sigma_i}, \tag{9.4}$$

where a_i is the result of fit i and σ_i is its uncertainty, and a_{exp} is the expected value, in this case the model input value. Therefore, the distribution of the pulls is expected to be a Gaussian distribution centered at zero with unit width. In the context of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays, it is useful to verify that the toys behave as expected under a relatively complicated amplitude model, with many different contributions floating in the fit. The results of a semi-realistic toy study are shown below.

The choice of amplitudes included in the toy model is motivated by the indications of resonances in Figure 9.1, and they are as follows. There are four resonance states considered in the decay to a $D^0\overline{D}^0$ pair, $\psi(3770), \psi(4040), \psi(4160)$ and $\chi_{c2}(3930)$. For the $J^P = 1^-$ vector states, ψ', the three helicity components are floated separately in the fit. A non-resonant S-wave $D^0\overline{D}^0$ contribution is also included. For D^0K^+, a single resonance



(a) Invariant mass distribution of $m(D^0\overline{D}^0)$

(b) Invariant mass distribution of $m(D^0K^+)$

(c) Invariant mass distribution of $m(K^+\pi^-)$

(d) Invariant mass distribution of $m(\overline{D}^0K^+)$

Figure 9.1: Projections of invariant mass distributions of sWeighted data (black points), combined across all categories of the signal mode, overlaid with $B^0 \to D^0\overline{D}^0K\pi$ MC (orange dashed line). The MC sample is scaled to match the data statistics. Potential resonant states are indicated in cyan at their world average mass values.

is included, the $D^*_{s2}(2573)^+$, and the $\overline{D}^0\pi^-$ system is modelled as a non-resonant S-wave state. Two $K\pi$ states are included, the $K^*(892)^0$, simply labelled K^{*0} henceforth, and an S-wave component. This accounts for both the $K^*_0(1430)^0$ and a non-resonant contribution, parameterised by the LASS lineshape as described previously. For each charmonium resonance, components for both $K\pi$ states are included. The full list of amplitudes is given in Table 9.2. Note that a D-wave $K^*_2(1430)^0$ state is not included in this instance, but should be included in future.

In order to build a semi-realistic toy sample, the relative contributions from each decay amplitude are derived using results from amplitude analyses of related modes. The $\lambda = 0$ helicity component of the $B^0 \to K^{*0}\psi(3770)$ is fixed with magnitude equal to 1 and phase equal to 0. The relative magnitudes and phases of the ψ' states are derived from the measurement of the phase difference between short- and long-distance amplitudes in the $B^+ \to K^+\mu^+\mu^-$ decays in Ref. [128]. Within each of the ψ' contributions, the relative helicity components are taken from the measurement of polarisation amplitudes in $B^0 \to J/\psi K^{*0}$ decays in Ref [174]. The size of the $\chi_{c2}(3930)$ and non-resonant $D^0\overline{D}^0$ contributions, relative to the $\psi(3770)$ resonance, are treated as in Ref. [100] which studied $B^+ \to D^+D^-K^+$ decays. Finally, the $D^*_{s2}(2573)^+$ contribution is derived from the Dalitz plot analysis of $B^+ \to D^0\overline{D}^0K^+$ decays in Ref. [175]. Clearly these modes have different rates and phase-space considerations compared to $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays, but these estimations are acceptable for the purposes of this semi-realistic toy model. The full model parameters are listed in Table 9.2. Invariant mass distributions of $m^2(D^0\overline{D}^0), m^2(D^0K^+)$, and $m^2(K^+\pi^-)$ are shown in Figure 9.2, from a single toy sample of 1600 events generated with this model. The result of a fit to this toy is also displayed, showing that the main resonant features are well modelled.

The toy study is performed using the `AmpGen` package, with details of its implementation given in Section 8.5. The yield generated in each toy sample is allowed to fluctuate around 1600 events according to a Poisson distribution. The transversity basis is used, with the components named in terms of one longitudinal and two transverse amplitudes. $\mathcal{A}_0$ is labelled 'Long', $\mathcal{A}_\parallel$ is labelled 'Tran1', and $\mathcal{A}_\perp$ is labelled 'Tran2'. The fit can be expressed in terms of the magnitude and phase of the complex coefficient of each associated decay amplitude, or in terms of its real and imaginary parts. Since the phase may be ill-defined when the magnitude is close to zero, the real and imaginary parts are used in this toy study. The pull distributions for all the floating parameters in the fit are shown in Figure 9.3, for 500 toys. A binned likelihood fit is performed to extract the Gaussian mean and width of each pull distribution. The distributions are verified

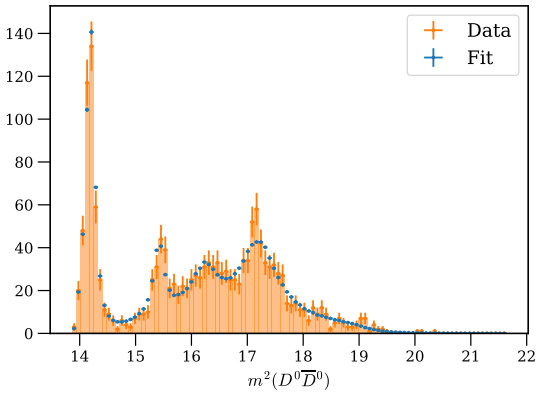
Table 9.2: Toy amplitude model for $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays, where each contribution, indexed r, has an associated complex coefficient $C_r = c_r exp(i\phi_r)$, and fit fraction F_r, quoted as a percentage. $K^*_{\rm LASS}$ refers to the S-wave $K\pi$ system parameterised with the LASS lineshape. The phases are given in degrees, and the magnitude and phase are calculated from the real and imaginary components used in the toy study. The transversity amplitudes for the contributions of the form $B^0 \to K^{*0}\psi'$ are indicated in square brackets.

r	c_r	ϕ_r	Re(C_r)	Im(C_r)	F_r(%)
$[K^{*0}\psi(3770)]_0$	1	0	1	0	4.75 ± 0.04
$[K^{*0}\psi(3770)]_\parallel$	0.63	-168	-0.616	-0.131	1.87 ± 0.02
$[K^{*0}\psi(3770)]_\perp$	0.593	168	-0.580	0.123	1.67 ± 0.02
$K^{*0}\chi_{c2}(3930)$	0.7	48	0.468	0.520	0.9 ± 0.3
$[K^{*0}\psi(4040)]_0$	0.551	-22	0.511	-0.206	1.54 ± 0.01
$[K^{*0}\psi(4040)]_\parallel$	0.347	170	-0.342	0.060	0.612 ± 0.006
$[K^{*0}\psi(4040)]_\perp$	0.327	146	-0.271	0.183	0.539 ± 0.005
$[K^{*0}\psi(4160)]_0$	1.385	13	1.35	0.312	7.47 ± 0.07
$[K^{*0}\psi(4160)]_\parallel$	0.872	-155	-0.790	-0.369	2.96 ± 0.03
$[K^{*0}\psi(4160)]_\perp$	0.821	-178	-0.821	-0.029	2.61 ± 0.02
$K^{*0}(D^0\overline{D}^0)_{\rm NR}$	1.29	-138	-0.959	-0.863	16.4 ± 0.1
$K^*_{\rm LASS}\psi(3770)$	0.8	-147	-0.671	-0.436	16.9 ± 0.2
$K^*_{\rm LASS}\chi_{c2}(3930)$	0.56	-100	-0.097	-0.551	8.3 ± 0.8
$K^*_{\rm LASS}\psi(4040)$	0.44	-170	-0.433	-0.076	4.0 ± 0.04
$K^*_{\rm LASS}\psi(4160)$	1.10	-134	-0.764	-0.791	16.6 ± 0.2
$D^*_{s2}(2573)^+(\overline{D}^0\pi^-)_{\rm NR}$	0.02	-17	0.019	-0.010	$(3.11 \pm 0.03) \times 10^{-3}$

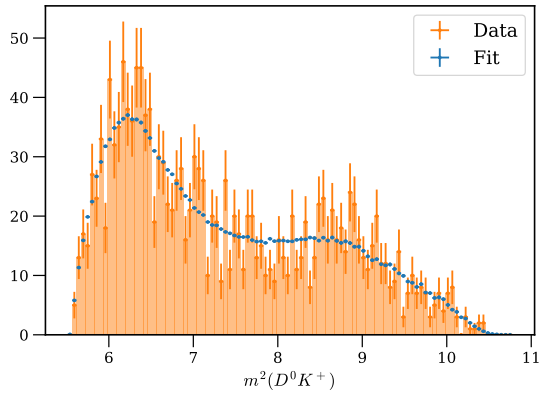
to resemble Gaussian distributions, and Figure 9.3 simply shows the fitted mean and widths. Overall, the pulls have widths compatible with unity, and only some small biases are seen. These results are encouraging for this semi-realistic toy model, and repeating the toy study for the amplitude model based on the full LHCb dataset will be discussed further in the next section.

9.3 Summary and future plans

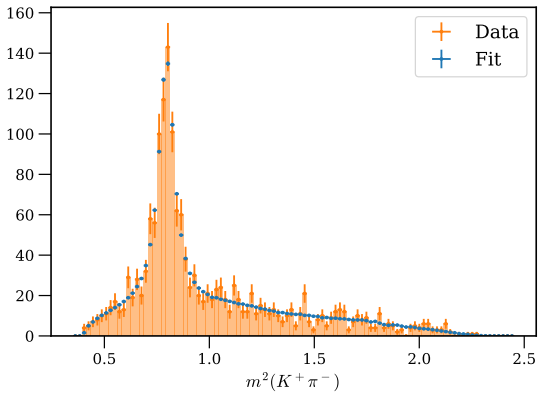
This chapter presents initial studies towards an amplitude analysis of $B^0 \to D^0 \overline{D}^0 K^+ \pi^-$ decays, based on the formalism introduced in Chapter 8. A semi-realistic toy study is performed and reveals that an amplitude analysis would be successful in extracting the various contributions to the overall decay amplitude, for the expected signal yield of all the decay modes under consideration. The ability to measure the relative helicity amplitudes of decays of the form $B^0 \to K^{*0}\psi'$ is encouraging for the study of charm loop



(a) Invariant mass distribution of $m^2(D^0\overline{D}^0)$



(b) Invariant mass distribution of $m^2(D^0K^+)$



(c) Invariant mass distribution of $m^2(K^+\pi^-)$

Figure 9.2: Projections of invariant mass distributions of generated data (orange) and fit results (blue) for the toy amplitude model.

contributions in $b \to s\ell^+\ell^-$ processes. Specifically, these results can provide input to the study of the golden mode $B^0 \to K^{*0}\mu^+\mu^-$, to assess the interference of the electroweak penguin decay with the various vector contributions.

With the event selection procedure established as described, the next steps are to include data from the $D^0 \to K3\pi$ modes and build an associated amplitude model with `AmpGen`. The resonant contributions in the toy model introduced here may be used as a baseline model for the real data, along with any further resonances apparent in the mass distributions of the full dataset. The final amplitude model will be determined using an unbiased approach based on the Least Absolute Shrinkage and Selection Operator (LASSO) procedure [176]. This method balances the complexity of the model and the fit quality by adding a penalty term to the negative log-likelihood, tuned to the

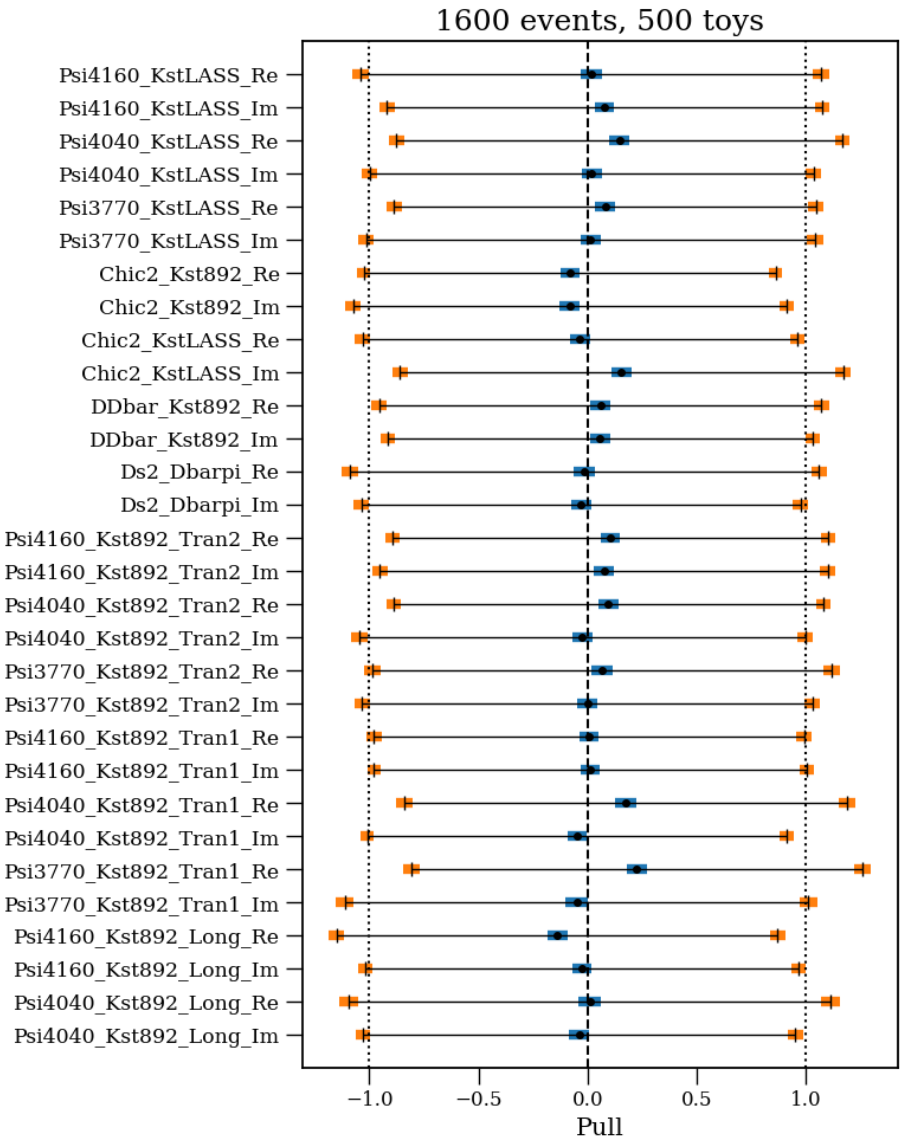


Figure 9.3: Binned likelihood fit results of the pulls of the fit parameters in the toy study. The real and imaginary parts of each decay amplitude in the toy model are shown along the vertical axis. Each horizontal line is centered on the fitted mean of the associated pull distribution, and the blue shaded area shows the error on the mean. The width of each line is given by the fitted width, and the orange shaded area is the error on the width.

data [177]. It will be especially interesting to see whether hints of resonances in the $m(\overline{D}^0 K^+)$ distribution (Figure 9.1) persist with the inclusion of more data, and whether the inclusion of new states are required in order to model these features. It will also be instructive to investigate the spin assignment of the state or states near the $\chi_{c2}(3930)$ resonance, to compare with the results presented in Ref. [100].

There are several important factors in the amplitude fit procedure that require careful consideration. The amplitude fit is likely to be performed on *s*Weighted data, and the uncertainties in the fit parameters from a weighted maximum likelihood fit are known to be poorly estimated. This is demonstrated in Ref. [178], from which the appropriate correction to the covariance of the fit results must therefore be applied. This procedure, as well as the overall fit bias, must then be verified with a toy study as above. In this case the pseudodata will be generated from the nominal amplitude model, and the biases may be corrected for, or assigned as systematic uncertainties on the fit parameters. The widths of the resulting pull distributions can also be used to verify or apply corrections to the errors derived from the fit. Performing an *s*Weighted fit also relies on the assumption that the signal and background PDFs factorise in the mass distribution. This must be explicitly verified here, or assigned a systematic uncertainty if the correlation between the various invariant mass projections and the B^0 mass distribution of the background is not negligible. A number of sources of systematic uncertainty also arise from experimental considerations, many of which are similar to the sources discussed in Section 7.9. For the amplitude fit itself, there are a number of model-dependent uncertainties to consider, such as the fixed lineshape parameters. Further considerations will depend on the developed model, such as assessing the impact of including or excluding specific amplitude contributions, or varying the intrinsic properties of lesser-known resonant states.

CHAPTER 10

CONCLUSIONS AND OUTLOOK

This thesis documents studies of the decay $B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-$ with data from the LHCb experiment at the Large Hadron Collider. The first observation of the decay is presented, as published in Ref. [1], using data corresponding to $4.7\,\text{fb}^{-1}$ of integrated luminosity collected during 2011, 2012 and 2016. In total, 297 ± 14 signal candidates are observed, as determined from a simultaneous fit to the reconstructed B^0 mass distribution across four categories, which are specified by the trigger selections and run periods. A measurement of the branching fraction is performed, with respect to $B^0 \to D^{*-} D^0 K^+$ decays, and is determined to be

$$\mathcal{B}(B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-) = (3.50 \pm 0.27\,(\text{stat}) \pm 0.26\,(\text{syst}) \pm 0.30(\mathcal{B}')) \times 10^{-4},$$

where the first uncertainty is statistical, the second is systematic, and the third is due to the uncertainty of the branching fraction of the normalisation channel (labelled by $\mathcal{B}'$). Preparatory studies for an amplitude analysis of $B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-$ decays are also presented, using additional data collected during 2017 and 2018, a total $8.7\,\text{fb}^{-1}$ of integrated luminosity. Amplitude fits are performed on toy simulation samples with the total expected signal yield of 1600 events, verifying the stability and statistical precision of a proposed four-body amplitude model in the helicity formalism. Plans for the full amplitude analysis, including building the amplitude model and considerations of potential sources of systematic uncertainty, are also given.

The research described in this thesis presents the crucial foundation upon which a full amplitude analysis of $B^0 \to D^0 \overline{D}{}^0 K^+ \pi^-$ decays may be performed. This decay mode

will provide important insights to spectroscopy studies and to the evaluation of charm loop contributions in rare $b \to s\ell^+\ell^-$ processes. The potential for spectroscopy of heavy hadrons is exciting, especially given the numerous unexpected meson discoveries in similar decay modes. Experimental measurements in this area improve theoretical understanding of the quark model and of challenging QCD effects. The results of the amplitude analysis will inform the expected sensitivity and potential new avenues of research into this and similar decay modes in LHCb data collected in Run 3 and beyond. Furthermore, the input to hadronic uncertainties in rare semileptonic decays will have a crucial impact, as the numerous b-anomaly measurements are the most intriguing area of LHCb physics at the moment. The measured discrepancies with the Standard Model and the mounting evidence for lepton flavour universality violation are the most promising signs of New Physics at the LHC in its ten years of data-taking. By providing experimental constraints on the impact of charm loops in $b \to s\ell^+\ell^-$ transitions, the forthcoming amplitude analysis of $B^0 \to D^0\overline{D}^0K^+\pi^-$ decays will reduce hadronic uncertainties in these measurements, as well as informing the choice of NP models which aim to consistently describe all of the anomalous results.

In addition to the analyses performed with LHCb data, this thesis outlines the development of the TORCH time-of-flight detector for future installation within the LHCb experiment. The proposed addition will benefit low-momentum particle identification, which has significant impact on future physics analysis. A calibration system for the front-end electronics of the prototype detector is developed, and integrated into the data acquisition software for the first time. A description of the setup is provided, along with first results, which demonstrate efficient data collection and an analysis of the charge-to-width calibration of test chips. Next steps for a complete set of calibration studies are detailed; these are essential for improving the overall timing resolution of TORCH, and demonstrating that it reaches its intended performance.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR AMPLITUDE ANALYSIS

This appendix provides additional information for the amplitude analysis presented in Chapters 8 and 9. Section A.1 provides a derivation for the Monte Carlo integration procedure described in Section 8.5.1. The selection procedure for the data samples used in the amplitude analysis is very similar to that presented in the branching fraction analysis (Chapter 7). The updates to the selection strategy are described in Section A.2, as well as presenting the mass fits for the data across all years in Section A.3.

A.1 Importance sampling

Monte Carlo integration is used to evaluate the normalisation factor of the likelihood in Equation 8.26, using the principle of *importance sampling*, see Ref. [179]. First consider the expectation of a function $f(x)$ where x is distributed according to $p(x)$,

$$\langle f(x) \rangle = \int dx\, f(x) p(x) \approx \frac{1}{N} \sum_i f(x_i). \tag{A.1}$$

This relation can be used to express the integral of $f(x)$ as

$$\int dx\, f(x) = \int dx\, \frac{f(x)}{p(x)} p(x) \approx \frac{1}{N} \sum_i \frac{f(x_i)}{p(x_i)}. \tag{A.2}$$

This method can be applied to the integral in Equation 8.27, by considering the probability distribution of N integration events labelled $g(x;\Theta)$, such that

$$\int dx\,\epsilon(x)\phi(x)\mathcal{P}(x;\Theta) \approx \frac{1}{N}\sum_{i=0}^{N}\frac{\epsilon(x_i)\phi(x_i)\mathcal{P}(x_i;\Theta)}{g(x_i;\Theta)}. \tag{A.3}$$

If the integration events are generated with a model, labelled $\mathcal{P}'(x;\Theta)$, and processed and selected in an identical manner to the real data, the generator distribution is then given by

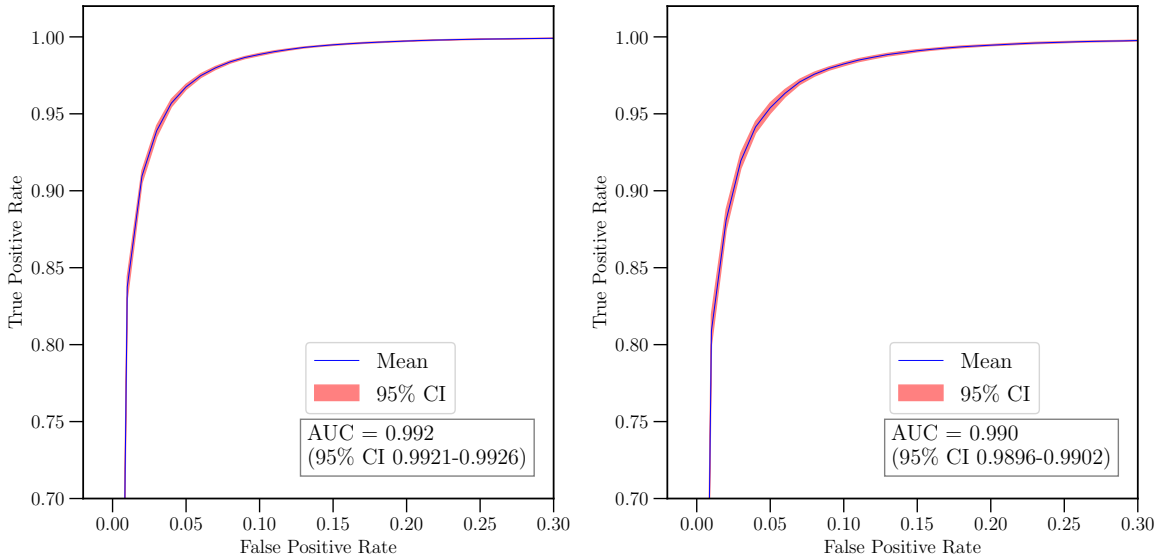
$$g(x;\Theta) = \epsilon(x)\phi(x)\mathcal{P}'(x;\Theta). \tag{A.4}$$

Therefore, the normalisation integral simplifies to

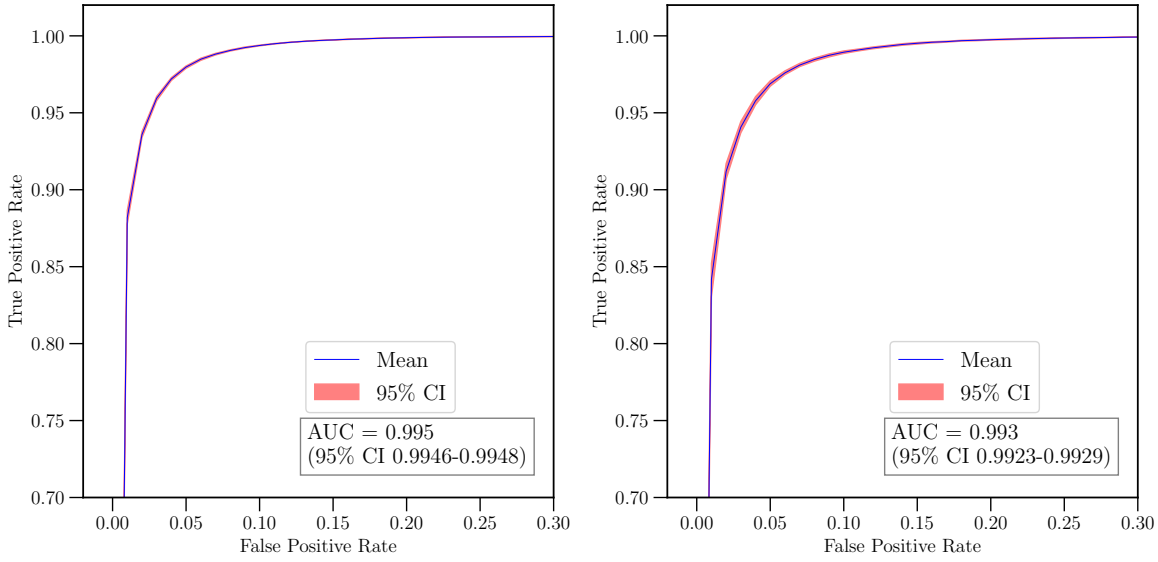
$$I = \frac{1}{N}\sum_{i=0}^{N}\frac{\mathcal{P}(x_i)}{\mathcal{P}'(x_i)}. \tag{A.5}$$

A.2 Event selection

The stripping, trigger and offline selections are unchanged from the branching fraction measurement. The training of the classifier is enhanced in several ways. Firstly, for the second stage $B_{\text{selection}}$ classifier, two additional variables are included in the training sample. These are the impact parameter significance, χ^2_{IP}, of the pion and kaon. Furthermore, the background proxy, which is defined by selection in the uppermass sideband of the B^0 mass, now uses a larger window. This is given by $m(B^0) - m_{\text{PDG}}(B^0) > 100\,\text{MeV}/c^2$, where the threshold is lowered from $200\,\text{MeV}/c^2$. This serves to both increase the size of the background proxy and to improve modelling of background contamination close to the signal peak. The training for each classifier is also combined for 2016-2018 to form a set of Run 2 classifiers. Other than this, the years are still kept separate, for example in terms of the mass fits described below. Similarly, the optimisation of the NN cut is still performed separately. In the BF analysis, the training is done using k-folding with $k = 10$, and then selecting the subset with the highest performance, given no evidence for overfitting. In this analysis, all ten networks are saved, and the resulting weights are applied only to the appropriate validation subsamples. This maximises the usefulness of the full training sample. For events not used in training, such as data in the signal region, the weight is calculated as an average over the weights from the ten distinct neural networks. The ROC curves for the Run 1 and 2 $B_{\text{selection}}$ classifiers are shown in Figure A.1. The AUC is comparable with the previous classifiers in Figure 7.7, and an



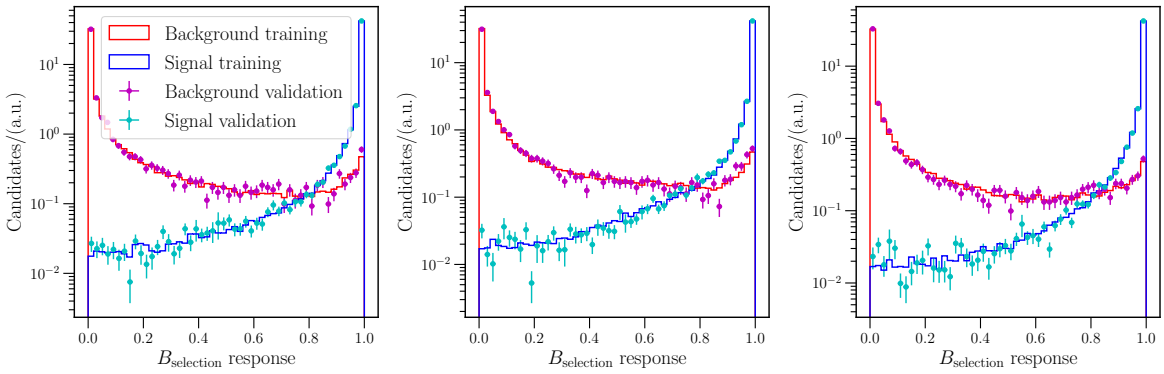
(a) Run 1: (left) TOS, (right) TIS



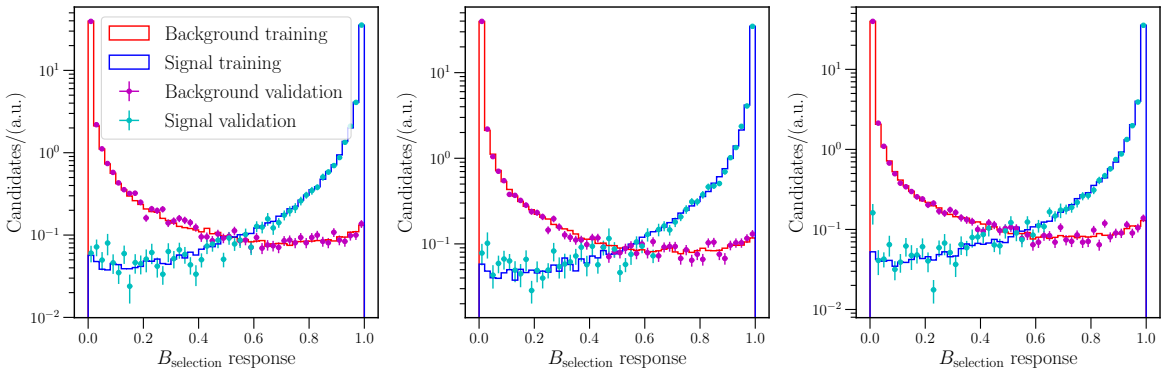
(b) Run 2: (left) TOS, (right) TIS

Figure A.1: Average ROC curves and confidence intervals of $B_{\text{selection}}$ classifiers (zoomed into area of interest).

improvement is seen in the Run 2 case compared to the 2016-only sample, as expected. The classifier response on the training and validation samples is also shown in Figure A.2 for a selection of training folds. As before, good separation is seen between signal and background, and good agreement is seen between training and validation sets.



(a) Run 1



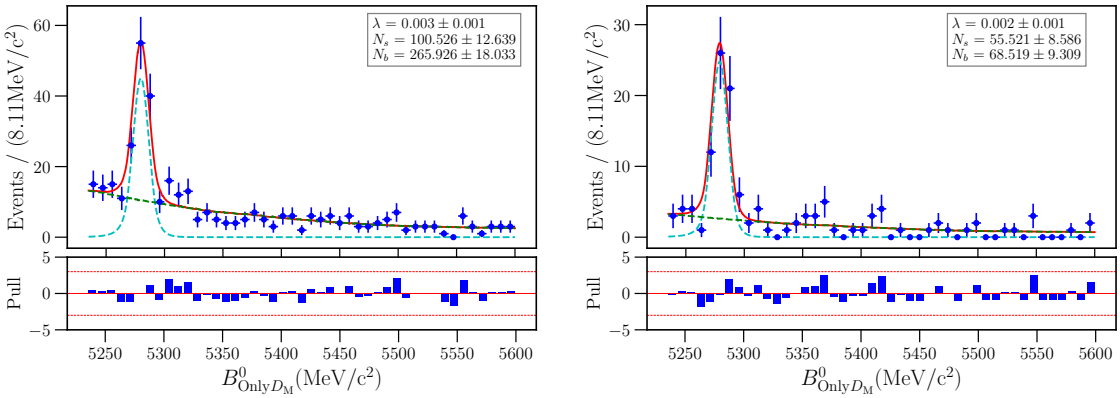
(b) Run 2

Figure A.2: 2nd stage $B_{\text{selection}}$ classifier response on training (line) and validation (points) samples for signal (red) and background (blue) samples in the TOS category, shown for a random selection of training k-folds.

Table A.1: Overall selection efficiency for the data subsamples, specified by the run period and trigger category. As indicated, the quoted numbers are on the order of 10^{-4}.

$\times 10^{-4}$	Run 1	2016	2017	2018
TOS	4.33 ± 0.02	8.25 ± 0.07	7.57 ± 0.07	5.30 ± 0.05
TIS	1.57 ± 0.01	3.83 ± 0.05	4.51 ± 0.05	2.85 ± 0.04

The overall selection efficiencies for the various categories, derived from simulation, are listed in Table A.1. The increase seen in 2016 and 2017 compared to Run 1 is due to the increase in the stripping efficiency as previously explained in Section 7.6, with the L0 thresholds lowered further in 2017. This increase is offset in the 2018 sample due to a relatively lower classifier efficiency.



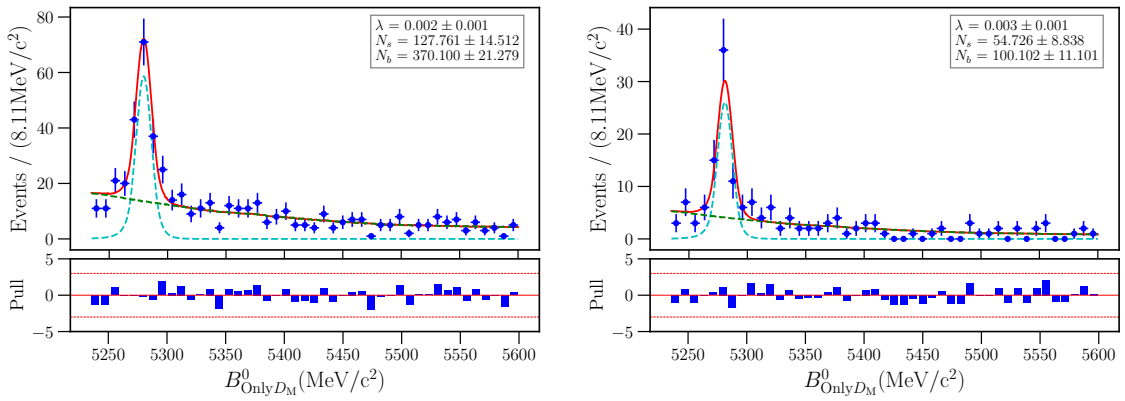
(a) TOS

(b) TIS

Figure A.3: Fits to the reconstructed B^0 mass in data for the signal mode in Run 1. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.

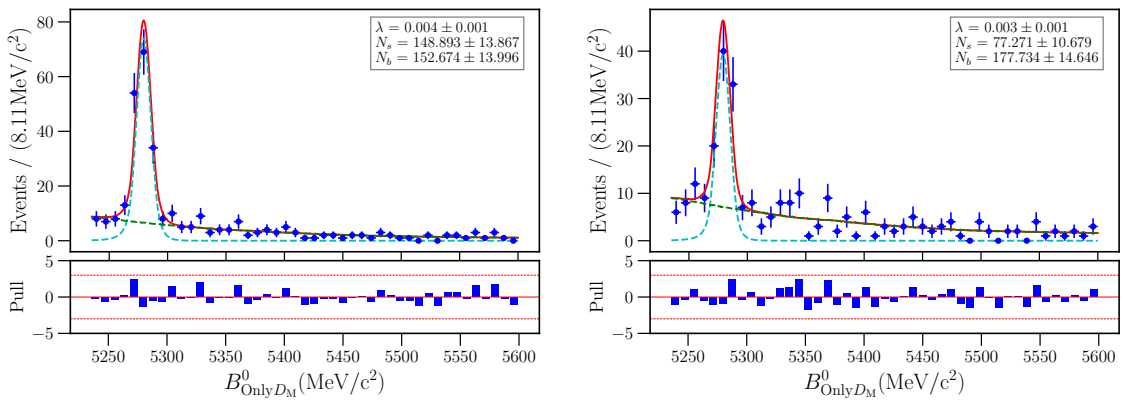
A.3 Mass fits

The fitting procedure is unchanged from the BF analysis, with the exception of the software packages used to implement the PDFs and the likelihood estimation. For the amplitude analysis, the fitting is performed with the most up-to-date packages supported by the Scikit-HEP project [180, 181]. The mass fits are performed using the package `zfit` [130, 182], which is built upon a TensorFlow backend. The derived *s*Weights from the fits are then extracted with the associated `hepstats` package. Mass fits to signal data for the eight independent categories are shown in Figures A.3, A.4, A.5, A.6.



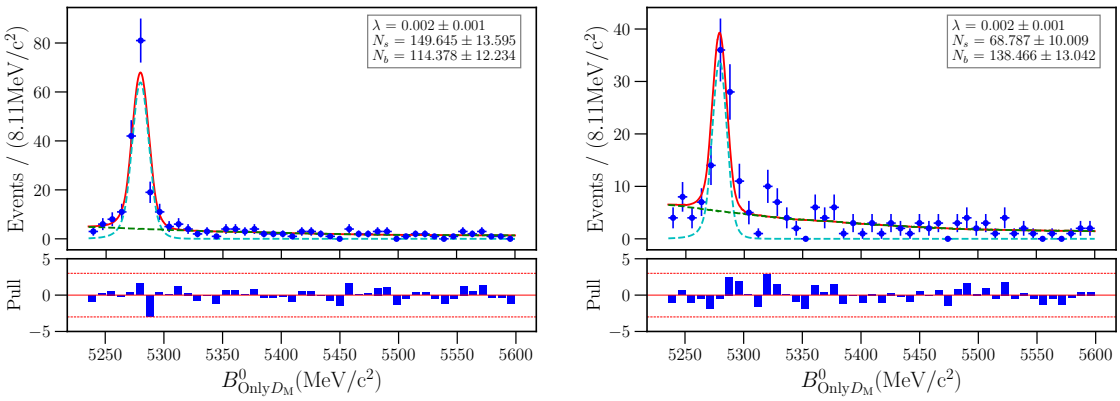
(a) TOS (b) TIS

Figure A.4: Fits to the reconstructed B^0 mass in data for the signal mode in 2016. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.



(a) TOS (b) TIS

Figure A.5: Fits to the reconstructed B^0 mass in data for the signal mode in 2017. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.



(a) TOS

(b) TIS

Figure A.6: Fits to the reconstructed B^0 mass in data for the signal mode in 2018. The blue points represent data and the red line represents the full fit model. The background and signal models are represented by green and cyan dashed lines, respectively.

BIBLIOGRAPHY

[1] LHCb collaboration, R. Aaij *et al.*, *First observation of the decay $B^0 \to D^0 \bar{D}^0 K^+ \pi^-$*, Phys. Rev. **D102** (2020) 051102, arXiv:2007.04280.
(Cited on pages 2, 71, 126, and 149.)

[2] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01.
(Cited on pages 9, 72, 73, 75, 88, 105, 107, 131, 140, and 141.)

[3] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769, Submitted to Nature Physics.
(Cited on pages 1 and 58.)

[4] LHCb collaboration, *List of hadrons observed at the LHC*, LHCB-FIGURE-2021-001, CERN, 2021.
(Cited on page 2.)

[5] LHCb collaboration, R. Aaij *et al.*, *A model-independent study of resonant structure in $B^+ \to D^+ D^- K^+$ decays*, Phys. Rev. Lett. **125** (2020) 242001, arXiv:2009.00025.
(Cited on pages 2 and 55.)

[6] LHCb collaboration, R. Aaij *et al.*, *Observation of structure in the J/ψ-pair mass spectrum*, Science Bulletin **65** (2020) 1983, arXiv:2006.16957.
(Cited on page 2.)

[7] Z. Ligeti, *TASI Lectures on Flavor Physics*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*, 2015, arXiv:1502.01372.
(Cited on page 3.)

[8] A. Lenz, *Heavy Flavour Physics and Effective Field Theories*, https://www.ippp.dur.ac.uk/~lenz/Lecture_Flavour_2017.pdf, 2017.
(Cited on page 3.)

[9] ATLAS collaboration, 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*, Phys. Lett. B **716** (2012) 1, arXiv:1207.7214.
(Cited on page 4.)

[10] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv:1207.7235.
(Cited on page 4.)

[11] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.
(Cited on pages 4 and 6.)

[12] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.
(Cited on pages 4 and 6.)

[13] N. Walchover, S. Velasco, and L. Reading-Ikkanda, *A New Map of All the Particles and Forces*, https://www.quantamagazine.org/a-new-map-of-the-standard-model-of-particle-physics-20201022/, 2020.
(Cited on page 4.)

[14] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8** (1964) 214.
(Cited on page 5.)

[15] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the $Z(4430)^-$ state*, Phys. Rev. Lett. **112** (2014) 222002, arXiv:1404.1903.
(Cited on page 5.)

[16] LHCb collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi p K^-$ decays*, Phys. Rev. Lett. **115** (2015) 072001, arXiv:1507.03414.
(Cited on pages 5 and 55.)

[17] CMS, A. M. Sirunyan *et al.*, *A measurement of the Higgs boson mass in the diphoton decay channel*, Phys. Lett. B **805** (2020) 135425, arXiv:2002.06398.
(Cited on page 7.)

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

(Cited on page 10.)

[19] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.
(Cited on page 10.)

[20] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.
(Cited on page 10.)

[21] CKMfitter Group, J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. C **41** (2005) 1, arXiv:hep-ph/0406184, updated results and plots available at: http://ckmfitter.in2p3.fr.
(Cited on page 11.)

[22] Y. Grossman, *Introduction to flavor physics*, in *2009 European School of High-Energy Physics*, 111–144, 2010, arXiv:1006.3534.
(Cited on page 11.)

[23] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285.
(Cited on page 11.)

[24] S. Godfrey and N. Isgur, *Mesons in a relativized quark model with chromodynamics*, Phys. Rev. D **32** (1985) 189.
(Cited on page 11.)

[25] T2K collaboration, K. Abe *et al.*, *Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations*, Nature **580** (2020) 339, arXiv:1910.03887, [Erratum: Nature 583, E16 (2020)].
(Cited on page 13.)

[26] Y. Grossman, *Beyond the standard model with B and K physics*, Int. J. Mod. Phys. A **19** (2004) 907, arXiv:hep-ph/0310229.
(Cited on page 13.)

[27] G. Isidori, *Flavor physics and CP violation*, in *2012 European School of High-Energy Physics*, 69–105, 2014, arXiv:1302.0661.
(Cited on page 13.)

[28] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

(Cited on pages 15, 19, 20, 25, and 33.)

[29] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, https://cds.cern.ch/record/2197559, 2016.
(Cited on page 16.)

[30] Christian Elsässer, *$\overline{b}b$ production angle plots*, https://lhcb.web.cern.ch/speakersbureau/html/bb_ProductionAngles.html.
(Cited on page 18.)

[31] LHCb collaboration, R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.
(Cited on pages 20 and 21.)

[32] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.
(Cited on pages 21, 22, and 23.)

[33] R. Frühwirth, *Application of kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444 .
(Cited on pages 22 and 74.)

[34] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.
(Cited on pages 25, 26, and 27.)

[35] R. Forty, *RICH pattern recognition for LHCb*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **433** (1999) 257 .
(Cited on pages 25 and 49.)

[36] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.
(Cited on page 26.)

[37] LHCb RICH collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II. Performance of the LHCb RICH detectors during the LHC Run II*, LHCB-PUB-2017-012. CERN-LHCb-PUB-2017-012, CERN, Geneva, 2017.
Updated authors' details and DOI, doi: 10.1016/j.nima.2017.03.009.
(Cited on page 26.)

[38] L. Anderlini *et al.*, *Computing strategy for PID calibration samples for LHCb Run 2*, LHCb-PUB-2016-020. CERN-LHCb-PUB-2016-020, CERN, Geneva, 2016.
(Cited on page 26.)

[39] LHCb collaboration, *Trigger schemes*, `https://lhcb.web.cern.ch/speakersbureau/html/TriggerScheme.html`.
(Cited on page 28.)

[40] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, `arXiv:1211.3055`.
(Cited on pages 28 and 29.)

[41] J. Albrecht, V. V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, Journal of Physics: Conference Series **513** (2014) 012001.
(Cited on pages 28, 29, and 30.)

[42] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, `arXiv:1812.10790`.
(Cited on pages 28 and 29.)

[43] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, 2014.
(Cited on pages 28 and 126.)

[44] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, Journal of Instrumentation **8** (2013) P02013.
(Cited on page 29.)

[45] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.
(Cited on pages 30 and 49.)

[46] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`.
(Cited on page 30.)

[47] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.
(Cited on page 30.)

[48] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.
(Cited on page 30.)

[49] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.
(Cited on page 30.)

[50] G. Corti *et al.*, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323.
(Cited on page 30.)

[51] G. Barrand *et al.*, *Gaudi - a software architecture and framework for building hep data processing applications*, Computer Physics Communications **140** (2001) 45 , CHEP2000.
(Cited on page 30.)

[52] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007. LHCb-TDR-12, CERN, Geneva, 2012.
(Cited on page 31.)

[53] LHCb collaboration, *Letter of Intent for the LHCb Upgrade*, CERN-LHCC-2011-001. LHCC-I-018, CERN, Geneva, 2011.
(Cited on page 31.)

[54] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, arXiv:1808.08865.
(Cited on pages 31 and 48.)

[55] LHCb collaboration, *LHCb Trigger and Online Upgrade Technical Design Report*, CERN-LHCC-2014-016. LHCB-TDR-016, CERN, 2014.
(Cited on page 31.)

[56] LHCb collaboration, M. P. Whitehead, *The upgrade of the LHCb trigger for Run III*, PoS **EPS-HEP2017** (2018) 528. 5 p.
(Cited on page 31.)

[57] M. J. Charles and R. Forty, *TORCH: Time of flight identification with Cherenkov radiation*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **639** (2011) 173 , Proceedings of the Seventh International Workshop on Ring Imaging Cherenkov Detectors.

(Cited on pages 31, 33, 34, and 35.)

[58] M. Akatsu *et al.*, *Time-of-propagation cherenkov counter for particle identification*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **440** (2000) 124 .
(Cited on page 34.)

[59] K. Fohl *et al.*, *The DIRC detectors of the PANDA experiment at FAIR*, Nucl. Instrum. Meth. A **595** (2008) 88.
(Cited on page 34.)

[60] S. Bhasin *et al.*, *Test-beam studies of a small-scale TORCH time-of-flight demonstrator*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **961** (2020) 163671.
(Cited on pages 34, 35, 38, 39, and 48.)

[61] T. M. Conneely *et al.*, *The TORCH PMT: a close packing, multi-anode, long life MCP-PMT for Cherenkov applications*, Journal of Instrumentation **10** (2015) C05003.
(Cited on pages 35, 36, and 37.)

[62] N. Harnew *et al.*, *Status of the TORCH time-of-flight project*, Nucl. Instrum. Meth. A **952** (2018) 161692, arXiv:1812.09773.
(Cited on pages 35 and 36.)

[63] R. Gao *et al.*, *Development of TORCH readout electronics for customised MCPs*, Journal of Instrumentation **11** (2016) C04012.
(Cited on pages 36 and 38.)

[64] S. M. George, *Atomic layer deposition: An overview*, Chemical Reviews **110** (2010) 111, PMID: 19947596.
(Cited on page 36.)

[65] F. Anghinolfi *et al.*, *NINO: an ultra-fast and low-power front-end amplifier/discriminator ASIC designed for the multigap resistive plate chamber*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **533** (2004) 183 , Proceedings of the Seventh International Workshop on Resistive Plate Chambers and Related Detectors.
(Cited on page 37.)

[66] M. Mota *et al.*, *A flexible multi-channel high-resolution time-to-digital converter asic*, in *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No.00CH37149)*, **2** 9/155–9/159 vol.2, 2000.
(Cited on page 37.)

[67] R. Bitter, T. Mohiuddin, and M. Nawrocki, *LabVIEW: Advanced programming techniques*, CRC Press, 2006.
(Cited on page 37.)

[68] P. Baesso, D. Cussans, and J. Goldstein, *The AIDA-2020 TLU: a flexible trigger logic unit for test beam facilities*, Journal of Instrumentation **14** (2019) P09019.
(Cited on page 38.)

[69] N. Brook *et al.*, *Testbeam studies of a TORCH prototype detector*, Nuclear Instruments & Methods in Physics Research Section A-accelerators Spectrometers Detectors and Associated Equipment **908** (2018) 256.
(Cited on page 38.)

[70] A. V. Akindinov *et al.*, *Design aspects and prototype test of a very precise TDC system implemented for the Multigap RPC of the ALICE-TOF*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **533** (2004) 178 , Proceedings of the Seventh International Workshop on Resistive Plate Chambers and Related Detectors.
(Cited on page 39.)

[71] S. Liu, F. Changqing, Y. Han, and A. Qi, *LUT-based non-linearity compensation for BES III TOF's time measurement*, Nuclear Science and Techniques **21** (2010) 49.
(Cited on page 39.)

[72] 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.
(Cited on pages 42 and 64.)

[73] L. Castillo Garcia, *Study of a prototype module of a precision time-of-flight detector for particle identification at low momentum*, PhD thesis, Ecole Polytechnique, Laussane, 2016, Presented 10 Jun 2016.
(Cited on pages 44 and 47.)

[74] T. H. Hancock *et al.*, *Beam tests of a large-scale TORCH time-of-flight demonstrator*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **958** (2020) 162060.
(Cited on pages 48 and 49.)

[75] N. Harnew *et al.*, *Status of the TORCH Project*, Journal of Instrumentation **15** (2020) C04031–C04031.
(Cited on page 48.)

[76] S. Bhasin *et al.*, *TORCH physics performance : improving low-momentum PID performance during Upgrade IB and beyond*, LHCb-PUB-2020-006. CERN-LHCb-PUB-2020-006, CERN, Geneva, 2020.
(Cited on pages 48, 49, and 50.)

[77] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, *Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$*, JHEP **09** (2010) 089, arXiv:1006.4945.
(Cited on pages 51, 61, and 62.)

[78] T. E. Browder, *Hadronic decays and lifetimes of B and D mesons*, in *High energy physics: Proceedings, 28th International Conference, ICHEP'96, Warsaw, Poland, July 25-31, 1996. Vol. 1, 2*, 1139–1142, 1996, arXiv:hep-ph/9611373.
(Cited on page 52.)

[79] G. Buchalla, I. Dunietz, and H. Yamamoto, *Hadronization of $b \to c\bar{c}s$*, Phys. Lett. B **364** (1995) 188, arXiv:hep-ph/9507437.
(Cited on page 53.)

[80] ALEPH collaboration, R. Barate *et al.*, *Observation of doubly charmed B decays at LEP*, Eur. Phys. J. C **4** (1998) 387.
(Cited on page 53.)

[81] BaBar collaboration, B. Aubert *et al.*, *Measurement of the branching fractions for the exclusive decays of B^0 and B^+ to $\bar{D}^{(*)}D^{(*)}K$*, Phys. Rev. D **68** (2003) 092001, arXiv:hep-ex/0305003.
(Cited on page 53.)

[82] Belle collaboration, G. Gokhroo *et al.*, *Observation of a Near-threshold $D^0\overline{D}^0\pi^0$ enhancement in $B \to D^0\overline{D}^0\pi^0 K$ decay*, Phys. Rev. Lett. **97** (2006) 162002, arXiv:hep-ex/0606055.
(Cited on page 53.)

[83] Belle collaboration, J. Brodzicka *et al.*, *Observation of a new D_{sJ} meson in $B^+ \to \overline{D}^0 D^0 K^+$ decays*, Phys. Rev. Lett. **100** (2008) 092001, arXiv:0707.3491.
(Cited on page 53.)

[84] BaBar collaboration, P. del Amo Sanchez *et al.*, *Measurement of the $B \to \overline{D}^{(*)} D^{(*)} K$ branching fractions*, Phys. Rev. D **83** (2011) 032004, arXiv:1011.3929.
(Cited on pages 53, 72, and 127.)

[85] LHCb collaboration, R. Aaij *et al.*, *Measurement of branching fraction ratios for $B^+ \to D^{*+} D^- K^+$, $B^+ \to D^{*-} D^+ K^+$, and $B^0 \to D^{*-} D^0 K^+$ decays*, JHEP **12** (2020) 139, arXiv:2005.10264.
(Cited on pages 53, 72, and 99.)

[86] J. L. Rosner, *Hadron Spectroscopy: Theory and Experiment*, J. Phys. G **34** (2007) S127, arXiv:hep-ph/0609195.
(Cited on pages 54 and 55.)

[87] P. Colangelo, F. De Fazio, F. Giannuzzi, and S. Nicotri, *New meson spectroscopy with open charm and beauty*, Phys. Rev. D **86** (2012) 054024, arXiv:1207.6940.
(Cited on page 54.)

[88] M. Neubert, *Heavy-quark symmetry*, Physics Reports **245** (1994) 259–395.
(Cited on page 54.)

[89] F. De Fazio, *Weak decays of heavy quarks*, At The Frontier of Particle Physics (2001) 1671–1717.
(Cited on page 54.)

[90] H.-X. Chen *et al.*, *A review of the open charm and open bottom systems*, Rept. Prog. Phys. **80** (2017) 076201, arXiv:1609.08928.
(Cited on page 55.)

[91] L. X. Gutiérrez-Guerrero, A. Bashir, M. A. Bedolla, and E. Santopinto, *Masses of Light and Heavy Mesons and Baryons: A Unified Picture*, Phys. Rev. D **100** (2019) 114032, arXiv:1911.09213.
(Cited on page 55.)

[92] LHCb collaboration, R. Aaij *et al.*, *Observation of a narrow $P_c(4312)^+$ state, and of two-peak structure of the $P_c(4450)^+$*, Phys. Rev. Lett. **122** (2019) 222001, arXiv:1904.03947.
(Cited on page 55.)

[93] Belle collaboration, S. K. Choi *et al.*, *Observation of a narrow charmonium - like state in exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ decays*, Phys. Rev. Lett. **91** (2003) 262001, arXiv:hep-ex/0309032.
(Cited on page 55.)

[94] Belle collaboration, S. K. Choi *et al.*, *Observation of a resonance-like structure in the $pi^{\pm}\psi'$ mass distribution in exclusive $B \to K\pi^{\pm}\psi'$ decays*, Phys. Rev. Lett. **100** (2008) 142001, arXiv:0708.1790.
(Cited on page 55.)

[95] L. Maiani, F. Piccinini, A. D. Polosa, and V. Riquer, *Diquark-antidiquarks with hidden or open charm and the nature of X(3872)*, Phys. Rev. D **71** (2005) 014028, arXiv:hep-ph/0412098.
(Cited on page 55.)

[96] C. Hanhart, Y. S. Kalashnikova, and A. V. Nefediev, *Interplay of quark and meson degrees of freedom in a near-threshold resonance: multi-channel case*, Eur. Phys. J. A **47** (2011) 101, arXiv:1106.1185.
(Cited on page 55.)

[97] N. Brambilla *et al.*, *The XYZ states: experimental and theoretical status and perspectives*, Phys. Rept. **873** (2020) 1, arXiv:1907.07583.
(Cited on pages 55 and 56.)

[98] A. Palano, *Experimental status of excited D_s^+ mesons*, arXiv:2009.07126.
(Cited on page 55.)

[99] S. Godfrey and K. Moats, *Properties of excited charm and charm-strange mesons*, Physical Review D **93** (2016) .
(Cited on page 55.)

[100] LHCb collaboration, R. Aaij *et al.*, *Amplitude analysis of the $B^+ \to D^+D^-K^+$ decay*, Phys. Rev. D **102** (2020) 112003, arXiv:2009.00026.
(Cited on pages 55, 56, 143, and 147.)

[101] M. Karliner and J. L. Rosner, *First exotic hadron with open heavy flavor: $cs\bar{u}\bar{d}$ tetraquark*, Phys. Rev. D **102** (2020) 094016, arXiv:2008.05993.
(Cited on page 55.)

[102] T. J. Burns and E. S. Swanson, *Discriminating among interpretations for the $X(2900)$ states*, Phys. Rev. D **103** (2021) 014004, arXiv:2009.05352.
(Cited on page 55.)

[103] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, arXiv:1705.05802.
(Cited on pages 57 and 58.)

[104] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond*, JHEP **01** (2009) 019, arXiv:0811.1214.
(Cited on pages 58 and 59.)

[105] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay*, Phys. Rev. Lett. **125** (2020) 011802, arXiv:2003.04831.
(Cited on pages 58 and 62.)

[106] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ decay*, arXiv:2012.13241, submitted to Phys. Rev. Lett.
(Cited on page 58.)

[107] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of angular parameters from the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. **B781** (2018) 517, arXiv:1710.02846.
(Cited on page 58.)

[108] ATLAS collaboration, M. Aaboud *et al.*, *Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **10** (2018) 047, arXiv:1805.04000.
(Cited on page 58.)

[109] Belle collaboration, S. Wehle *et al.*, *Lepton-flavor-dependent angular analysis of $B \to K^*\ell^+\ell^-$*, Phys. Rev. Lett. **118** (2017) 111801, arXiv:1612.05014.
(Cited on page 58.)

[110] BaBar collaboration, B. Aubert *et al.*, *Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$*, Phys. Rev. **D73** (2006) 092001, arXiv:hep-ex/0604007.
(Cited on page 58.)

[111] CDF collaboration, T. Aaltonen *et al.*, *Measurements of the angular distributions in the decays $B \to K^{(*)}\mu^+\mu^-$ at CDF*, Phys. Rev. Lett. **108** (2012) 081807, arXiv:1108.0695.
(Cited on page 58.)

[112] LHCb collaboration, R. Aaij *et al.*, *Angular moments of the decay $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ at low hadronic recoil*, JHEP **09** (2018) 146, arXiv:1808.00264.

(Cited on page 58.)

[113] LHCb collaboration, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay* $B_s^0 \to \phi\mu^+\mu^-$, JHEP **09** (2015) 179, arXiv:1506.08777.
(Cited on page 58.)

[114] LHCb collaboration, R. Aaij *et al.*, *Measurements of the S-wave fraction in* $B^0 \to K^+\pi^-\mu^+\mu^-$ *decays and the* $B^0 \to K^*(892)^0\mu^+\mu^-$ *differential branching fraction*, JHEP **11** (2016) 047, Erratum ibid. **04** (2017) 142, arXiv:1606.04731.
(Cited on page 58.)

[115] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of* $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ *decays*, JHEP **06** (2015) 115, Erratum ibid. **09** (2018) 145, arXiv:1503.07138.
(Cited on page 58.)

[116] LHCb collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of* $B \to K^{(*)}\mu^+\mu^-$ *decays*, JHEP **06** (2014) 133, arXiv:1403.8044.
(Cited on page 58.)

[117] S. Descotes-Genon, J. Matias, and J. Virto, *Understanding the* $B \to K^*\mu^+\mu^-$ *Anomaly*, Phys. Rev. **D88** (2013) 074002, arXiv:1307.5683.
(Cited on page 58.)

[118] W. Altmannshofer, S. Gori, S. Profumo, and F. S. Queiroz, *Explaining dark matter and B decay anomalies with an* $L_\mu - L_\tau$ *model*, JHEP **12** (2016) 106, arXiv:1609.04026.
(Cited on page 58.)

[119] C. Cornella *et al.*, *Reading the footprints of the B-meson flavor anomalies*, arXiv:2103.16558.
(Cited on page 58.)

[120] J. Lyon and R. Zwicky, *Resonances gone topsy turvy - the charm of QCD or new physics in* $b \to s\ell^+\ell^-$*?*, arXiv:1406.0566.
(Cited on pages 58 and 62.)

[121] M. Ciuchini *et al.*, $B \to K^*\ell^+\ell^-$ *decays at large recoil in the Standard Model: a theoretical reappraisal*, JHEP **06** (2016) 116, arXiv:1512.07157.
(Cited on page 58.)

[122] C. Bobeth, G. Hiller, and G. Piranishvili, *CP Asymmetries in $\bar{B} \to \bar{K}^*(\to \bar{K}\pi)\bar{\ell}\ell$ and Untagged $\bar{B}_s$, $B_s \to \phi(\to K^+K^-)\bar{\ell}\ell$ Decays at NLO*, JHEP **07** (2008) 106, arXiv:0805.2525.
(Cited on page 58.)

[123] T. Blake *et al.*, *An empirical model to determine the hadronic resonance contributions to $\overline{B}^0 \to \overline{K}^{*0}\mu^+\mu^-$ transitions*, Eur. Phys. J. C **78** (2018) 453, arXiv:1709.03921.
(Cited on page 58.)

[124] W. Altmannshofer and P. Stangl, *New Physics in Rare B Decays after Moriond 2021*, arXiv:2103.13370.
(Cited on page 58.)

[125] M. Algueró *et al.*, *Emerging patterns of New Physics with and without Lepton Flavour Universal contributions*, Eur. Phys. J. C **79** (2019) 714, arXiv:1903.09578, [Addendum: Eur.Phys.J.C 80, 511 (2020)].
(Cited on page 59.)

[126] D. Gerick, *Rare decays at LHCb: recent results*, https://indico.cern.ch/event/857473/contributions/4060371/, 2020.
Implications of LHCb measurements and future prospects.
(Cited on page 60.)

[127] E. Fermi, *Tentative di una Teoria Dei Raggi β*, Il Nuovo Cimento **11** (1934) .
(Cited on page 61.)

[128] LHCb collaboration, R. Aaij *et al.*, *Measurement of the phase difference between the short- and long-distance amplitudes in the $B^+ \to K^+\mu^+\mu^-$ decay*, Eur. Phys. J. **C77** (2017) 161, arXiv:1612.06764.
(Cited on pages 61 and 143.)

[129] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.
(Cited on page 63.)

[130] R. S. Coutinho, A. N. Puig, and J. Eschle, *zfit: scalable pythonic fitting*, 2019. doi: 10.5281/zenodo.3599416.
(Cited on pages 64 and 155.)

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

(Cited on page 64.)

[132] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, CRC Press, 2017.
(Cited on page 65.)

[133] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of Computer and System Sciences **55** (1997) 119 .
(Cited on page 65.)

[134] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
(Cited on page 65.)

[135] S. Haykin, *Neural networks: a comprehensive foundation*, Prentice Hall PTR, 1994.
(Cited on page 65.)

[136] H. J. Kelley, *Gradient theory of optimal flight paths*, Ars Journal **30** (1960) 947.
(Cited on page 66.)

[137] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533.
(Cited on page 66.)

[138] S. Ruder, *An overview of gradient descent optimization algorithms*, `arXiv:1609.04747`.
(Cited on page 67.)

[139] N. Srivastava *et al.*, *Dropout: A simple way to prevent neural networks from overfitting*, J. Mach. Learn. Res. **15** (2014) 1929–1958.
(Cited on page 68.)

[140] A. Blum, A. Kalai, and J. Langford, *Beating the hold-out: bounds for k-fold and progressive cross-validation*, in *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, COLT '99, (New York, NY, USA), 203, ACM, 1999.
(Cited on page 68.)

[141] C. E. Metz, *Basic principles of ROC analysis*, Seminars in Nuclear Medicine **8** (1978) 283 .
(Cited on page 68.)

[142] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters **27** (2006) 861.
(Cited on page 68.)

[143] J. H. Friedman, *Greedy function approximation: A gradient boosting machine.*, Ann. Statist. **29** (2001) 1189.
(Cited on page 69.)

[144] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, Erratum ibid. **119** (2017) 169901, arXiv:1612.05140.
(Cited on page 73.)

[145] HFLAV, Y. S. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, arXiv:1909.12524.
(Cited on page 73.)

[146] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Journal of Basic Engineering **82** (1960) 35.
(Cited on page 74.)

[147] M. Needham, *Identification of Ghost Tracks using a Likelihood Method*, LHCb-2008-026. CERN-LHCb-2008-026. LPHE-2008-004, CERN, Geneva, 2008.
(Cited on page 74.)

[148] F. Chollet *et al.*, *Keras*, 2015.
Software available from https://keras.io.
(Cited on page 80.)

[149] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015.
Software available from https://tensorflow.org.
(Cited on page 80.)

[150] K. W. Bowyer, N. V. Chawla, L. O. Hall, and W. P. Kegelmeyer, *SMOTE: synthetic minority over-sampling technique*, Journal Of Artificial Intelligence Research **16** (2002) 321, arXiv:1106.1813.
(Cited on page 81.)

[151] F. Pedregosa *et al.*, *Scikit-learn: Machine learning in Python*, J. Machine Learning Res. **12** (2011) 2825, arXiv:1201.0490, and online at http://scikit-learn.org/stable/.

(Cited on page 84.)

[152] B. Efron, *Bootstrap methods: Another look at the jackknife*, Ann. Statist. **7** (1979) 1.
(Cited on pages 84 and 124.)

[153] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021, 2016.
(Cited on page 92.)

[154] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2018) 1, arXiv:1803.00824.
(Cited on page 92.)

[155] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, JINST **10** (2015) P02011, arXiv:1411.5528.
(Cited on page 92.)

[156] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. Conf. Ser. **762** (2016) , arXiv:1608.05806, https://github.com/arogozhnikov/hep_ml.
(Cited on page 93.)

[157] R. H. Dalitz, *On the analysis of tau-meson data and the nature of the tau-meson*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1953) 1068.
(Cited on pages 99 and 130.)

[158] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, arXiv:1612.07489.
(Cited on pages 99 and 101.)

[159] LHCb collaboration, R. Aaij *et al.*, *Measurement of b-hadron production fractions in* 7 TeV *pp collisions*, Phys. Rev. **D85** (2012) 032008, arXiv:1111.2357.
(Cited on page 107.)

[160] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Nucl. Inst. & Meth. in Phys. Res. A **389** (1997) 81.
(Cited on page 111.)

[161] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics*, John Wiley & Sons (New York), 1952.
(Cited on page 131.)

[162] J. D. Richman, *An experimenter's guide to the helicity formalism*, CALT-68-1148, Calif. Inst. Technol., Pasadena, CA, 1984.
(Cited on page 132.)

[163] S. U. Chung, *Spin formalisms*, CERN-71-08, CERN, Geneva, 1970.
doi: 10.5170/CERN-1971-008.
(Cited on page 132.)

[164] M. Jacob and G. C. Wick, *On the general theory of collisions for particles with spin*, Annals of Physics **7** (1959) 404.
(Cited on page 132.)

[165] E. P. Wigner, *Group theory and its application to the quantum mechanics of atomic spectra*, Academic Press, 1931.
Translated into English by Griffin, J. J. (1959).
(Cited on page 133.)

[166] J. Gratrex, M. Hopfer, and R. Zwicky, *Generalised helicity formalism, higher moments and the $B \to K_{J_K}(\to K\pi)\bar{\ell}_1\ell_2$ angular distributions*, Phys. Rev. D **93** (2016) 054008, arXiv:1506.03970.
(Cited on pages 134 and 135.)

[167] Tim Evans, *et al.*, *AmpGen*, https://github.com/GooFit/AmpGen, 2020.
(Cited on page 136.)

[168] LHCb collaboration, Aaij, Roel *et al.*, *Studies of the resonance structure in $D^0 \to K^\mp\pi^\pm\pi^\pm\pi^\mp$ decays*, Eur. Phys. J. **C78** (2018) 443, arXiv:1712.08609.
(Cited on page 136.)

[169] S. M. Flatté, *Coupled - Channel Analysis of the $\pi\eta$ and $K\overline{K}$ Systems Near $K\overline{K}$ Threshold*, Phys. Lett. B **63** (1976) 224.
(Cited on page 136.)

[170] LHCb, R. Aaij *et al.*, *Study of the lineshape of the $\chi_{c1}(3872)$ state*, Phys. Rev. D **102** (2020) 092005, arXiv:2005.13419.
(Cited on page 136.)

[171] Aston, D. *et al.*, *A Study of $K^-\pi^+$ Scattering in the Reaction $K^-p \to K^-\pi^+n$ at 11-GeV/c*, Nucl. Phys. B **296** (1988) 493.
(Cited on page 136.)

[172] BaBar collaboration, Aubert, Bernard *et al.*, *Evidence for Direct CP Violation from Dalitz-plot analysis of $B^{\pm} \to K^{\pm}\pi^{\mp}\pi^{\pm}$*, Phys. Rev. D **78** (2008) 012004, arXiv:0803.4451.
(Cited on page 136.)

[173] P. d'Argent *et al.*, *Amplitude Analyses of $D^0 \to \pi^+\pi^-\pi^+\pi^-$ and $D^0 \to K^+K^-\pi^+\pi^-$ Decays*, JHEP **05** (2017) 143, arXiv:1703.08505.
(Cited on page 138.)

[174] LHCb collaboration, R. Aaij *et al.*, *Measurement of the polarization amplitudes in $B^0 \to J/\psi K^*(892)^0$ decays*, Phys. Rev. **D88** (2013) 052002, arXiv:1307.2782.
(Cited on page 143.)

[175] BaBar collaboration, J. P. Lees *et al.*, *Dalitz plot analyses of $B^0 \to D^-D^0K^+$ and $B^+ \to D^0\overline{D}^0K^+$ decays*, Phys. Rev. D **91** (2015) 052002.
(Cited on page 143.)

[176] R. Tibshirani, *Regression Shrinkage and Selection via the Lasso*, Journal of the Royal Statistical Society. Series B (Methodological) **58** (1996) 267.
(Cited on page 145.)

[177] B. Guegan, J. Hardin, J. Stevens, and M. Williams, *Model selection for amplitude analysis*, JINST **10** (2015) P09002, arXiv:1505.05133.
(Cited on page 147.)

[178] C. Langenbruch, *Parameter uncertainties in weighted unbinned maximum likelihood fits*, arXiv:1911.01303.
(Cited on page 147.)

[179] S. Weinzierl, *Introduction to Monte Carlo methods*, arXiv:hep-ph/0006269.
(Cited on page 151.)

[180] E. Rodrigues *et al.*, *The Scikit-HEP Project – overview and prospects*, EPJ Web Conf. **245** (2020) 06028.
(Cited on page 155.)

[181] E. Rodrigues *et al.*, *scikit-hep/scikit-hep: Version 2.0.0*, 2021.
doi: 10.5281/zenodo.4556007.
(Cited on page 155.)

[182] J. Eschle, A. N. Puig, R. Silva Coutinho, and N. Serra, *zfit: scalable pythonic fitting*, EPJ Web Conf. **245** (2020) 06025.
(Cited on page 155.)