A Flavour Search for Invisible Particles at Belle

Daniel Marcantonio

ORCID: 0000-0002-1315-8646

Submitted in total fulfillment
of the requirements of the degree of

Doctor of Philosophy

School of Physics
The University of Melbourne

March 2025

THE UNIVERSITY OF MELBOURNE

Abstract

School of Physics

Doctor of Philosophy

by Daniel Marcantonio
ORCID: 0000-0002-1315-8646

This thesis presents searches for five B meson decays of the form $B \to ha$, where h is a Standard Model hadron ($h \in \pi^{\pm}, K^{\pm}, p^{\pm}, D_s^{\pm}, \overline{D}^0$) and a is an invisible particle with a narrow decay width. There are several extensions to the Standard Model that predict this type of decay, and yet, to date, there have been few direct searches for such processes.

Every fundamental particle predicted by the Standard Model since its inception has been found through a series of particle colliders with ever-increasing discovery power. From the 1950s to the 2010s, a new fundamental particle was discovered in every decade. The discovery of the Higgs boson in 2012 marked the end of this era, thus completing the Standard Model. However, the unknown particle nature of dark matter, the strong CP problem, the baryon asymmetry of the universe, and the fact that neutrinos have mass constitute incredibly compelling evidence that our current understanding of the universe is incomplete. Therefore, the Standard Model is not a complete description of nature, and a new theory or an extension of the Standard Model is required. The discovery of any new particle not explained by the Standard Model would constitute evidence for new physics. Since the discovery of the Higgs boson there has been a proliferation of theories and opinions surrounding the most likely place to find the next fundamental particle, one that may bridge the gap between the sector of known particles and the unknown.

This work is the first-ever search for most of these decays. This analysis finds no evidence for new particles in any of the five decay channels studied. Upper limits on the branching fractions of these decays have been established. When mapped onto the couplings between known particles and particles from new physics models, these branching fraction limits provide the most stringent constraints to date.

Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD, except where indicated in the preface;
2. due acknowledgement has been made in the text to all other material used; and
3. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies, and appendices.

Daniel Marcantonio, March 2025

Preface

This thesis was written under the supervision of Professor Phillip Urquijo. Chapters 1–3 contain literature reviews regarding physics theory, details of the Belle detector, and other experimental methods pertinent to the analysis. Chapter 5 begins with an overview of the literature that describes many of the statistical methods used in the analysis.

The bulk of Chapters 4–9 contain entirely original work. In late 2022, Dr. Emi Kou suggested the inclusion of two charm meson channels in the search. Early versions of the sensitivity estimates presented in Chapter 7 were shown in 2023 [1] and later published in Physical Review D [2] in January 2024. This included sensitivity estimates to visibly decaying ALPs prepared by Dr. Yi Zhang.

Chapters 2–8 of this thesis are heavily based on the internal Belle document, or *Belle note*, for the analysis. Dr. Akimasa Ishikawa and Professor Chris Hearty provided detailed feedback on the construction of this Belle note, and their suggested improvements on both the development of the work and the communication of the analysis should be noted. The analysis material is currently under internal review within the Belle II collaboration and is to be submitted for publication.

This work was made possible by the Research Training Program scholarship granted by the Australian Government. Conference travel during the course of the degree was funded by Australian Research Council Discovery Project No. DP210101900 and the University of Melbourne's Laby PhD Travelling Scholarship.

Note added for the printed edition: More than a year after this thesis was submitted, the bulk of the analysis presented here was published as the Belle and Belle II Collaborations, "Search for Feebly Interacting Particles in B Decays with Missing Energy at Belle," Phys. Rev. Lett. ***136****, 231802 (2026) [3]. The associated data release is available via HEPData at* `https://doi.org/10.17182/hepdata.179653`.

Acknowledgements

I do not believe that a PhD can be completed without a strong support network surrounding the candidate, and as such there are a number of people I would like to thank.

First, I would like to thank Phillip Urquijo, who has been my supervisor since I started my MSc in 2019. Phill, your careful guidance and seemingly limitless knowledge of the field have been an immense help to me over the last five years. You should be proud of creating such a positive environment for postgraduate students within your group, and have my sincere gratitude for giving me the opportunity to work with you.

To the rest of my supervisory panel, Elisabetta Barberio and David Simpson, thank you for providing me with valuable advice and for taking an interest in my academic development over the course of my candidature.

I have been fortunate to have worked alongside and shared an office with many fantastic students. Cate, with whom I originally shared this work, Daniel, Marco, Eiasha, Paolo, Han, and Isabelle — thank you for the endless discussions about the intricacies of our work and the shared commiserations over computing cluster issues.

By far the best part of doing a PhD was meeting such excellent people, many of whom I am sure will be my lifelong friends. Lachlan, Marcel, Mike, Michael, Fred, Alex, Ayo, Oliver, Lukas, Francis, and Alexei — thank you for making the prospect of turning up to the School of Physics so fun. I am particularly grateful to Marcel for generously offering to proofread this thesis, and to Lachlan, who shared so much of this PhD journey with me.

To my parents, Fab and Helen, and to my brother Oscar — thank you for your love and unwavering support for me as I pursued this path. For as long as I can remember, you fostered an environment in which education and curiosity were inherently good, worthwhile things. This thesis is the result of that.

I would like to acknowledge the traditional owners and custodians of the lands on which this work was completed, the Wurundjeri people of the Kulin nation, and pay my respects to Indigenous Elders past and present. Sovereignty has never been ceded. It is, was, and always will be, Aboriginal land.

Contents

Abstract i

Declaration ii

Preface iii

Acknowledgements iv

List of Figures viii

List of Tables xi

Glossary of Abbreviations xii

1 Introduction 1

2 The Standard Model and beyond 4
2.1 The Standard Model . . . 4
2.1.1 Shortcomings of the Standard Model . . . 6
2.2 Physics beyond the Standard Model . . . 7
2.2.1 Axions and axion-like particles (pseudoscalar portal) . . . 8
2.2.1.1 The strong CP problem . . . 8
2.2.1.2 ALP models . . . 9
2.2.1.3 ALP production in quark annihilation . . . 11
2.2.2 Dark baryons . . . 11
2.2.3 Vector portal . . . 13
2.2.4 Scalar portal . . . 15
2.2.5 Other models . . . 18
2.3 Experimental landscape . . . 18
2.3.1 ALP searches . . . 18
2.3.2 Dark scalar searches . . . 20
2.3.3 Dark baryon searches . . . 22
2.4 Form factors . . . 24

3 The Belle experiment 29
3.1 The KEKB accelerator . . . 29
3.2 The Belle experiment . . . 31

3.2.1 Silicon vertex detector . . . 32
3.2.2 Central drift chamber . . . 33
3.2.3 Particle identification detectors . . . 34
3.2.4 Electromagnetic calorimeter . . . 35
3.2.5 K_L^0 and muon detector . . . 36
3.2.6 Trigger and data acquisition . . . 36
3.3 Analysis software . . . 36
3.3.1 Monte Carlo simulations . . . 37
3.3.2 The `b2bii` package and Belle II software . . . 37
3.3.3 Particle identification . . . 38
3.3.4 Full event interpretation . . . 38

4 Analysis overview, event reconstruction, and selection 41
4.1 Signal selection outline . . . 41
4.2 Signal events and background types . . . 43
4.3 The datasets . . . 44
4.3.1 Monte Carlo data . . . 44
4.3.1.1 Signal samples . . . 44
4.3.1.2 Background samples . . . 45
4.3.2 The Belle dataset . . . 48
4.4 Event reconstruction . . . 48
4.4.1 Rest of event . . . 49
4.4.2 Preselection criteria . . . 49
4.5 Event selection . . . 53
4.5.1 Boosted decision trees . . . 59
4.5.2 Background suppression with boosted decision trees . . . 60
4.5.3 Beam background and hadronic split-off . . . 67
4.5.4 Optimising selection criteria . . . 68
4.5.5 Irreducible backgrounds . . . 69
4.5.6 Peaking backgrounds . . . 70
4.6 Selection efficiency . . . 74
4.7 Sensitivity to long-lived particles . . . 76

5 Statistical analysis 81
5.1 Statistical methods in high-energy physics . . . 81
5.1.1 Parameter estimation . . . 81
5.1.2 Hypothesis testing and limit setting . . . 83
5.1.3 Significance and the look-elsewhere effect . . . 83
5.1.4 Implementation of methods . . . 85
5.2 Fit method . . . 85
5.2.1 Signal PDF . . . 86
5.2.1.1 Parameterisation of fit parameters . . . 87
5.2.1.2 Signal resolution . . . 91
5.2.2 Background PDF . . . 91
5.2.2.1 Background PDF cross-checks . . . 93
5.2.3 Combined fit . . . 97
5.2.4 Mass scan . . . 98

5.2.5 Veto regions . . . 99
5.3 Fit validation . . . 100

6 Validation in data and systematic uncertainties 105
6.1 Control modes . . . 105
6.1.1 Charged B control mode: $B^+ \to J/\psi K^+$ . . . 106
6.1.2 Neutral B control mode: $B^0 \to \overline{D}^0\pi^0$ . . . 110
6.1.3 Comparison to ratios derived from $B \to D^{(*)}\ell\nu_\ell$ decays . . . 111
6.2 Cross-checks with off-resonance data sample . . . 113
6.2.1 Continuum momentum correction . . . 114
6.3 Systematic uncertainties . . . 118
6.3.1 Modelling uncertainties . . . 118
6.3.2 Signal selection efficiency . . . 120

7 Signal extraction and results 122
7.1 Expected sensitivity . . . 122
7.2 Results . . . 123
7.2.1 Signal extraction . . . 123
7.2.2 Branching fraction upper limits . . . 125
7.3 Comparisons to other searches . . . 129

8 New physics interpretations 130
8.1 ALP couplings (pseudoscalar portal) . . . 130
8.2 Dark scalar couplings (scalar portal) . . . 136
8.3 B-mesogenesis parameter space . . . 140
8.4 RPV supersymmetry . . . 142

9 Conclusion 145

Bibliography 149

List of Figures

1.1 Voyager 2's view of Neptune 2

2.1 A representation of the elementary particles in the Standard Model 5
2.2 Feynman diagrams for two ALP production processes 10
2.3 Feynman diagrams for two quark annihilation ALP production processes . 11
2.4 A diagram of the decay $B^+ \to p\psi_D$, mediated by the colour-triplet scalar Y. 12
2.5 The $B^\pm \to K^\pm Z_D$ branching fraction as predicted by Equation (2.16) . . 14
2.6 Decay width of a Z_D boson assuming the dark sector consists of some massless dark fermions χ 16
2.7 Parameter space exclusions for the ALP-W coupling g_{aW} 21
2.8 Dark scalar parameter space for invisibly decaying scalars 23
2.9 Dark scalar parameter space for visibly decaying scalars 24
2.10 Current experimental constraints on B-mesogenesis channels 25
2.11 Current experimental constraints on B-mesogenesis with Λ_c 26
2.12 Form factors for $B \to \pi$ and $B \to K$ as a function of q^2, determined using the parameterisation and results in Equation (2.29) and Table 2.3. 28

3.1 A diagram of the KEKB collider 30
3.2 An isometric cutaway of the Belle detector 32
3.3 A schematic of the Belle SVD. 33
3.4 A schematic of a Belle CDC cell 34
3.5 A schematic of the Belle electromagnetic calorimeter 35
3.6 Examples of particle species separation in the CDC and TOF detectors . 39
3.7 A schematic overview of the FEI 40

4.1 $B^\pm \to \pi^\pm a$ histograms showing distributions of key selection variables . . 54
4.2 $B^\pm \to K^\pm a$ histograms showing distributions of key selection variables . . 55
4.3 $B^\pm \to pa$ histograms showing distributions of key selection variables . . . 56
4.4 $B^\pm \to D_s^\pm a$ histograms showing distributions of key selection variables . . 57
4.5 $B^0 \to \overline{D}^0 a$ histograms showing distributions of key selection variables . . 58
4.6 Event shape comparison between $B\overline{B}$ and continuum events 60
4.7 Pion channel BDT responses 64
4.8 Kaon channel BDT responses 64
4.9 Proton channel BDT responses 65
4.10 $D_s^\pm$ channel BDT responses 65
4.11 $\overline{D}^0$ channel BDT responses 66
4.12 BDT response comparison between data and simulation in the two-track sideband 66

4.13 E_{ECL} and Punzi FoM distributions with hadronic split-off suppression . . 68
4.14 Punzi FoM example 69
4.15 Reconstruction efficiencies for all five channels as a function of m_a. 75
4.16 Predicted ALP branching fractions with sizeable couplings to gluons . . . 76
4.17 Reconstruction efficiencies for several different visible final states for two benchmark lifetimes. 77
4.18 E_{ECL} distributions for signal MC with long-lived particles 78
4.19 Reconstruction efficiencies for all channels for a range of lifetimes. 80

5.1 A selection of signal fits 88
5.2 Signal fit parameter parameterisation 90
5.3 Signal peak resolution dependence on the FEI signal probability selection 91
5.4 Background-only fit quality for different amounts of continuum background 95
5.5 Background-only fits in the sideband $1.5 < E_{\mathrm{ECL}} < 6.0\,\mathrm{GeV}$ 96
5.6 Example of the combined fit model on a dataset with 20 signal events injected 98
5.7 Mean and standard deviation of the extracted significance for 500 toy signal extractions at and around the mass where 20 signal events were injected for four different mass points. 100
5.8 Significance vs. number of injected peaking background events for the decays $B^{\pm} \to \pi^{\pm}\omega$ and $B^{\pm} \to \pi^{\pm} a_0(1450)^0$ 101
5.9 S_{fit} value and pull distributions for 12,500 toy experiments 102
5.10 A selection of two linearity tests 102
5.11 Gradients of linearity tests across the entire search range for three channels 103
5.12 Gradients of linearity tests across the entire search range for two channels 104

6.1 Distributions of variables used for control channel selection and fitting . . 108
6.2 Background suppression BDT responses in data and MC in the two-track sideband 109
6.3 Fits to the $B^{\pm} \to J/\psi\, K^{\pm}$ peak in MC and data 109
6.4 Data-MC comparisons for four variables in the E_{ECL} sideband 111
6.5 Fits to the $B^0 \to \overline{D}^0\pi^0$ peak in MC and data 112
6.6 A comparison of the p_h^S distributions in off-resonance data and MC. . . . 115
6.7 Data-MC ratios for the off-resonance dataset in two distinct fit windows . 116
6.8 Momentum correction from the off-resonance sample for every scan point 117

7.1 Signal significance results 124
7.2 Branching fraction exclusions 126
7.3 Long-lived particles branching fraction exclusions 127
7.4 Long-lived particles branching fraction exclusions at the 95% CL 128

8.1 Updated limits on g_{aW} using results from this work 132
8.2 The g_{aW} parameter space up to $m_a = 800\,\mathrm{MeV}/c^2$ assuming $a \to \gamma\gamma$ is the primary ALP decay mode. 133
8.3 ALP-photon coupling 134
8.4 Limits on ALP-quark couplings 135
8.5 Dark scalar decay length $c\tau_S$ in the mass-mixing angle plane 137

8.6 Exclusion limits and model predictions for branching fractions as a function of scalar lifetime . . . 138
8.7 Exclusion regions for the dark scalar model . . . 139
8.8 Ratios of two-body to N-body branching fractions for the four possible dark baryon decays of the charged B meson. . . . 141
8.9 B-mesogenesis parameter space constraints . . . 142
8.10 Semileptonic asymmetries overlaid with the minimum branching fraction required for successful baryogenesis. . . . 143
8.11 90% CL upper limits on the RPV coupling λ''_{113} . . . 144

9.1 Belle II luminosity plots . . . 146

List of Tables

2.1 Results from pQCD computations of $\mathcal{B}(B \to ha)g_q^{-2}$ 12
2.2 The lightest final states for all B-mesogenesis operators 13
2.3 Form factor fit results 28

3.1 A breakdown of the e^+e^- cross section at the $\Upsilon(4S)$ resonance. 31
3.2 A concise comparison of B-factory tagging methods. 39

4.1 Mass difference between the B meson and the other signal-side hadron. . 45
4.2 Decay modes generated for the D_s^+ decay 46
4.3 Decay modes generated for the $\overline{D}^0$ decay 46
4.4 Mass hypotheses used in signal MC generation 47
4.5 A summary of the MC samples used in this analysis that were generated within the Belle collaboration 47
4.6 Preselection criteria common to all channels. 51
4.7 Channel-specific preselection criteria. 52
4.8 Summary of offline event selection criteria used in the analysis 53
4.9 The sets of variables used for training the two background suppression BDTs. 63
4.10 `XGBoost` parameters used for all BDT training. 67
4.11 Two-body decays present in the background for the $B^\pm \to \pi^\pm a$ search . . 71
4.12 Two-body decays present in the background for the $B^\pm \to K^\pm a$ search . . 72
4.13 Two-body decays present in the background for the $B^\pm \to pa$ search . . . 73
4.14 Two-body decays present in the background for the $B^\pm \to D_s^\pm a$ search . . 73
4.15 Two-body decays present in the background for the $B^0 \to \overline{D}^0 a$ search . . 73

5.1 Highest mass hypotheses used in the search for each channel 87

6.1 Event yields in data and MC for semileptonic decays from the Belle measurement of $B \to D^{**}\ell\nu_\ell$ 112
6.2 Overview of systematic uncertainties considered in this analysis. Each is treated as a nuisance parameter in the statistical model. 119
6.3 Summary of the systematic uncertainties 121

7.1 Comparison between our branching fraction exclusion results and the reaches of similar searches 129

Glossary of Abbreviations

ALP Axion-Like Particle; a pseudoscalar particle appearing in many new physics models

CDC Central Drift Chamber; the subdetector at Belle that was responsible for measuring the trajectories of charged particles through the detector

CERN European Organisation for Nuclear Research (Conseil Européen pour la Recherche Nucléaire); located on the border of France and Switzerland, near Geneva, CERN contains many particle physics experiments

CP Charge-Parity; a symmetry that, if obeyed, means that particles and antiparticles behave in the same way

ECL Electromagnetic Calorimeter; the subdetector at Belle that measured the energy of electromagnetically-interacting particles

EDM Electric Dipole Moment; the measure of separation between opposite charges in a system

EW Electroweak; the unified description of the electromagnetic and weak interactions

FEI Full Event Interpretation; an algorithm that reconstructs B mesons using thousands of different decay chains

FIP Feebly-Interacting Particle; a subatomic particle that only feebly couples to Standard Model particles

HEP High-Energy Physics; synonymous with particle physics

HER High-Energy Ring; the electron ring at KEK

HFLAV Heavy Flavour Averaging group; an international collaboration that averages results from many flavour physics experiments

IP Interaction Point; the point where beams cross each other at a collider

KDE Kernel Density Estimator; a non-parametric function constructed by summing kernel (generally Gaussian) distributions

KEK The High Energy Accelerator Research Organisation (Kō Enerugī Kasokuki Kenkyū Kikō)

KEKB KEK's electron-positron collider that provided collision events to Belle

KLM K_L^0 and Muon detector; the outermost subdetector of Belle

LER Low-Energy Ring; the positron ring at KEK

LHC The Large Hadron Collider; a high-energy hadron collider located at CERN

LINAC Linear Accelerator; a type of particle accelerator that operates in a straight line

MC Monte Carlo; a computational method based on random sampling. Also used to describe datasets simulated with this method

MVA Multivariate Analysis; a catch-all term for multi-variable machine learning classifiers

NP New Physics; any physics beyond what is described by the Standard Model

PDF Probability Density Function; a continuous function of a random variable whose value at a given point indicates the relative likelihood that the variable is equal to that value

PDG Particle Data Group; an international group of physicists that compiles and averages published results

PID Particle Identification; used to describe both hardware and software tools that distinguish between particle species

PMT Photomultiplier Tube; a type of light detector used in many particle physics experiments

QCD Quantum Chromodynamics; the theory that describes the action of the strong force

SM Standard Model of particle physics; the relativistic quantum field theory that describes the elementary particles and their interactions via the strong, weak, and electromagnetic forces

1

Introduction

What natural phenomena remain unexplained, and could they be attributed to unknown, invisible subatomic particles? How and why do we search for such particles at electron-positron colliders? The new particle search in this thesis aims to answer these broad questions.

In 1821, Alexis Bouvard published astronomical tables describing the orbit of Uranus, and predicted its future position based on Newton's laws and the contemporary understanding of gravitation. Later observations of Uranus' orbit deviated from Bouvard's predictions, leading him to hypothesise that another unknown celestial body was present in the solar system, influencing Uranus' orbit [4]. Over twenty years later, a new planet now known as Neptune was discovered thanks to the combined efforts of John Couch Adams and Urbain Le Verrier's predictions for its location, and Johann Gottfried Galle's observations [5].

For decades, astronomers have observed stars' anomalous rotation around the centres of galaxies, large-scale structure within the universe that defies conventional models, and stronger-than-expected gravitational lensing of light through distant galaxies. These phenomena can all be explained by the presence of some additional mass nearby; however, we cannot see any — at least not with telescopes reliant on the electromagnetic spectrum. This has led physicists to postulate the existence of dark matter: a new type of matter that does not interact electromagnetically, but does interact gravitationally. In a sense, Bouvard's new planet hypothesis is just as relevant today as it was almost two centuries ago. Observations tell us there is something out there; the challenge that remains is to determine what form it takes.

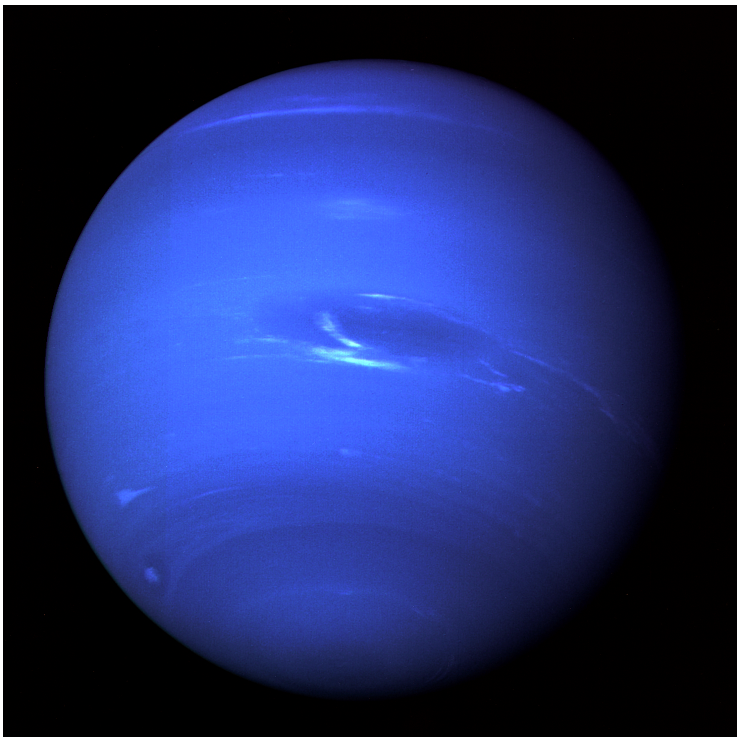


Figure 1.1: Voyager 2's view of Neptune, taken in 1989 [6].

The Standard Model (SM) of particle physics is one of the most rigorously tested scientific theories [7]; however, under this tight scrutiny, physicists have noticed a number of discrepancies between it and observations of the natural world. The strong evidence that dark matter [8] accounts for 85% of the mass of the universe [9] is perhaps the biggest indication that the SM fails to account for all the elementary particles present in the cosmos. The leading hypothesis for dark matter is that it is composed of particles [10], as opposed to some alternative theories that posit that the evidence for dark matter arises from a misunderstanding of gravitational dynamics [11]. Unlike dark matter, these alternative theories fail to explain the cosmic microwave background power spectrum and large-scale structures in the universe [7]. Particle theories of dark matter are generally presented as extensions to the relativistic quantum field theory that is the SM, and if dark matter interacts with any SM particles at all, it could theoretically be produced at particle colliders. In this thesis, SM particles are often referred to as the *visible sector* and a theoretical zoo of dark matter particles as the *dark* or *hidden sector*. Any particle that can mediate interactions between the visible and hidden sectors can be referred to as a *portal*. While dark matter provides a strong motivation for searching for new particles, other unexplained phenomena also point to gaps in our current theoretical framework.

This thesis presents a search for an invisible particle, a, that belongs to the hidden sector and possibly connects it to the visible sector. This search is model independent, in that it is agnostic to the spin and the details of the production mechanism of the invisible

particle. We motivate, and later interpret, the search result in the context of a number of new physics (NP) models, all of which connect a to the dark sector in some way. The a signal is searched for in rare decays of B mesons, produced in e^+e^- collisions at the Belle experiment in Tsukuba, Japan.

The data analysed in this thesis were collected between 1999 and 2010. The invisible particle a is searched for in five different decay channels: $B^+ \to \pi^+ a$, $B^+ \to K^+ a$, $B^+ \to p a$, $B^+ \to D_s^+ a$, and $B^0 \to \overline{D}^0 a$. Unless otherwise specified, charge-conjugate decays and particle labels are implied throughout this work.

This thesis is organised as follows:

- Chapter 2 introduces key Standard Model concepts and discusses new physics models that motivate the search for invisible particles.
- Chapter 3 describes the KEKB collider, the Belle detector, and essential software tools, including the Belle II software framework.
- Chapter 4 outlines the analysis methodology and provides a detailed description of the datasets used.
- Chapter 5 reviews statistical methods relevant to this study and explains their application in the search for invisible particles.
- Chapter 6 examines the differences in analysis performance between simulated and real collision events.
- Chapter 7 presents the experimental results of the search.
- Chapter 8 interprets the results from Chapter 7 within the theoretical framework established in Chapter 2.
- Chapter 9 summarises the results and discusses future prospects.

2

The Standard Model and beyond

This chapter establishes the motivation for searching for invisible particles in this thesis. It begins with a summary of the Standard Model and its limitations, then introduces key theoretical models that extend beyond it. A review of relevant experimental searches is also included.

2.1 The Standard Model

The Standard Model of particle physics is a theory that describes the electromagnetic, weak, and strong forces, and classifies the 30 fundamental particles and antiparticles discovered to date (shown in Fig. 2.1). A testament to its success in modelling nature is the fact that it predicted the existence of the $W^{\pm}$, Z, and Higgs bosons and the top quark many years before they were discovered. Within the SM, elementary particles are classified by their intrinsic spin (in units of $\hbar$), which is integer for bosons and half-integer for fermions.

Fundamental fermions can be further divided into two categories: quarks (u, d, c, s, t, b), which interact via all of the fundamental forces, and leptons (e, ν_e, μ, ν_μ, τ, ν_τ), which interact with all but the strong force as they lack colour charge. The fermions are sorted into three generations, with each generation being a copy of the previous generation, albeit with a higher mass. The quarks are charged under the $SU(3)_c$ gauge group, and these charges are called *colours*. The dynamics of quarks and gluons and the bound systems they are found in are described by a theory called quantum chromodynamics (QCD). Particles that do not interact under the strong force have no colour charge. Each generation contains one positive *up-type* quark with electric charge $+\frac{2}{3}e$ and one

negative *down-type* quark with electric charge $-\frac{1}{3}e$, where e is the elementary charge [12, 13]. As fundamental fermion types are numerous, like types of ice cream, the term *flavour* has been used to describe fermion type [14]. The branch of particle physics that studies interactions between particle species, particularly across generations, is *flavour physics*.

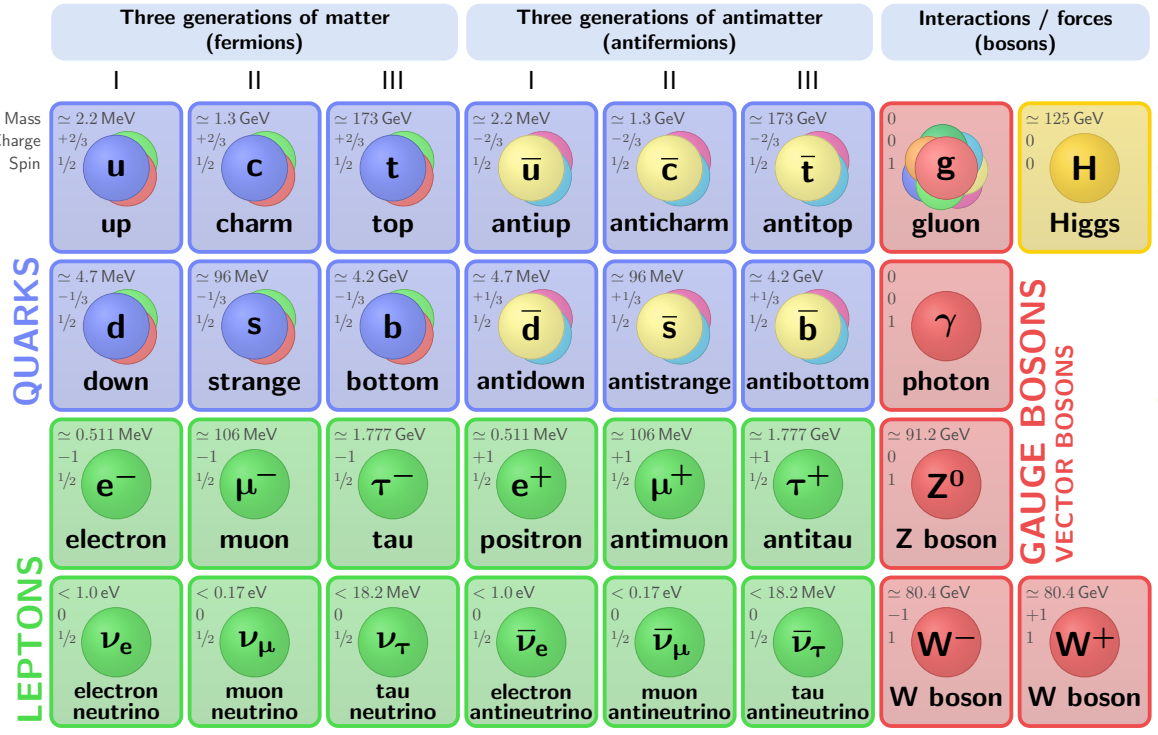


Figure 2.1: A representation of the elementary particles in the Standard Model, based on the tables in Refs. [15, 16].

Each generation of leptons contains a particle with one elementary electric charge (e, μ, τ) and one electrically neutral particle, or neutrino (ν_e, ν_μ, ν_τ), whose flavours correspond to those of the charged leptons. As the neutrinos are electrically neutral, they can only interact through the weak force.

In the SM, the bosons act as force carriers for the three fundamental forces. The gluon g mediates the strong interaction, and is the only boson that has colour charge. The photon γ mediates the electromagnetic force, while the massive Z and $W^\pm$ bosons mediate the weak force. The $W^\pm$ bosons are the only electrically charged bosons. All of these force-carrier particles are spin-1, and are therefore called *vector bosons*. There is one spin-0 *scalar boson* in the SM: the Higgs boson H. The heaviest of the fundamental bosons, the Higgs does not mediate a force, but the spontaneous symmetry breaking of its field leads to the elementary particles gaining their masses.

The genesis of the SM was in the unification of the $SU(3)_c$ gauge theory of the strong interaction, quantum chromodynamics (QCD) [17], and the $SU(2)_L \otimes U(1)_Y$ electroweak

theory [18–20] in the 1960s and '70s. The structure of its interactions follows from Lorentz invariance and the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ gauge symmetry where c denotes colour, L indicates that the weak isospin group acts only on left-handed fermions, and Y denotes hypercharge. While the SM possesses unrivalled predictive power at energies below the TeV scale, it provides no fundamental explanation for the values of its many free parameters, requiring a degree of fine-tuning to reproduce the universe as we observe it today.

2.1.1 Shortcomings of the Standard Model

Despite the success of the SM, physicists have made several observations that it cannot explain. Since the 1930s, astronomers have observed galactic rotation velocity profiles that deviate from predictions based on gravitational interactions involving only luminous matter [21–23]. This discrepancy led physicists to propose the existence of dark matter, an unseen form of matter that does not interact with electromagnetic radiation. Systematic measurements conducted decades later confirmed these unexpected dynamics, solidifying the case for dark matter [24]. There are no SM particles that match all the properties we require of dark matter, and so we can conclude from this that the SM does not provide us with a complete picture of the fundamental particles and interactions in our universe. The abundance of dark matter in the universe is one of the most important constraints to consider for NP models. Generally, SM extensions rely on the freeze-out mechanism to generate the density of dark matter in the universe today, or relic abundance [25]; however, this does not preclude other mechanisms [26]. In the simplest dark matter models, the hidden sector consists of some Dirac fermion χ that has a non-zero annihilation cross section to SM particles. In the early universe, dark matter is kept in thermal equilibrium with the visible sector, as the rate at which it self-annihilates into visible sector particles is equal to the rate of the reverse process. Freeze-out occurs when the annihilation rate drops below the expansion rate of the universe. The temperature-scaled abundance of a particle is defined as

$$Y = \frac{N}{VT^3} \tag{2.1}$$

where N is the number of particles, V is the volume, and T is the temperature of the universe. As the universe expands, it cools, and when the temperature drops below the dark matter mass, it cannot be produced by the SM bath.[1] Eventually, as the universe cools and expands, the decreasing annihilation rate stabilises, leaving a constant abundance of dark matter. This abundance depends on the annihilation cross section, which determines how frequently dark matter particles interact with the visible sector.

[1] The term *bath* refers to the pool of thermal energy provided by the environment, which in the case of the early universe, is a hot, dense plasma of particles.

Of course, the simple model summarised above is not necessarily correct or even the most likely NP scenario. It is possible that more complex scenarios exist, many of which could alter the relic abundance [27].

The SM provides no mechanism through which the strong force should be invariant under a combined charge C and parity P transformation, yet experiments consistently show that it preserves CP symmetry. This discrepancy, known as the strong CP problem, is one of the most interesting outstanding issues in physics today [28, 29].

We live in a matter-dominated universe, despite the fact that the universe most likely consisted of equal parts matter and antimatter during the Big Bang [30]. While the SM does allow for particle decay processes that occur at different rates for particles and their antiparticles, the resulting matter-antimatter asymmetry is vastly smaller than what is observed today. To address this discrepancy, numerous new theories have been proposed. The mechanism through which these theories generate matter-antimatter asymmetry is called *baryogenesis*. Generally, these mechanisms proceed via interactions that occur at different rates for baryons and antibaryons, not conserving CP symmetry.

There are myriad other phenomena that are not described by the SM, and so the goal of particle physics as a field today is to generate models and experimentally test for evidence of physics beyond the SM. Particles predicted by NP models have not yet been detected, and so a catch-all term for the ensemble of these particles is the hidden sector, to contrast with the visible sector of the SM particle zoo.

2.2 Physics beyond the Standard Model

Historically, experimental searches for physics beyond the Standard Model (SM) have focused on particles with sizeable couplings to the SM at high masses, typically at or above the electroweak (EW) scale, $\mathcal{O}(10^2 - 10^3)$ GeV. Collider experiments, such as those at the LHC, have pushed to increasingly higher energies to probe these high-mass scales, significantly constraining NP at the so-called *energy frontier*. However, due to the lack of evidence for NP at these higher energies, attention has shifted towards models that have remained undetected due to the small couplings between the visible and hidden sectors.

In these models, non-SM particles can have masses on a similar scale to most SM particles, within the few-MeV/c^2 to few-GeV/c^2 mass range. Particles in these hidden sectors are often classified as *feebly interacting particles* (FIPs) [31, 32]. This classification includes couplings via the weak interaction but also allows for new interactions mediated by forces beyond the three fundamental interactions described in the SM. The decays of B mesons provide an ideal environment for searching for such particles, as their mass

of 5.28 GeV/c^2 allows decays into FIPs with masses up to a few GeV/c^2, depending on the specific decay channel.

In the remainder of this chapter, general types of NP models are introduced, with a focus on those that could generate feebly interacting particles in B-meson decays. Experimentally, one can constrain the rates at which NP processes occur. For several interesting classes of models, formulae for the branching fractions are provided, describing both the production of invisible particles and the decays of NP particles into visible and invisible final states. Additionally, the experimental status of each class of models is reviewed.

Many NP models propose that only a small subset of particles in the hidden sector interact with the visible sector. These particles, which mediate interactions between the two sectors, are often referred to as *portals* and are typically classified by their intrinsic spin. This classification gives rise to vector, scalar, and pseudoscalar portals.

2.2.1 Axions and axion-like particles (pseudoscalar portal)

2.2.1.1 The strong CP problem

The best motivation for studies of axion-like particles (ALPs) is a discussion of the strong CP problem, which motivated the solution that generated the original QCD axion. The QCD Lagrangian includes a so-called θ-term that violates P and T symmetry (and hence CP),

$$\mathscr{L} = \theta \frac{g_s^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \tag{2.2}$$

where $G^a_{\mu\nu}$ is the gluon field strength tensor, g_s is the strong interaction coupling strength, and θ is a dimensionless coupling constant. Naïvely, one might expect the action of this Lagrangian term S_θ to evaluate to zero over all space, which would make the term unphysical, negating any potential CP-violating effects [33]. However, it was realised in the 1970s that the correct choice of boundary values for the gluon field gives rise to non-zero values of the action

$$S_\theta = \theta q \tag{2.3}$$

where $q \in \mathbb{Z}$ [34, 35]. For non-zero q then, Equation (2.2), and therefore θ itself, must influence physical quantities.

The low-energy theory of nucleons and pions should be impacted by the non-zero θ-term in QCD. Importantly, the neutron electric dipole moment (EDM) is dependent on this term [36]

$$d_n \sim \theta \times 10^{-16}\ e \text{ cm}. \tag{2.4}$$

Measurements of the neutron EDM have become incredibly precise in recent years, with the most recent result [37]

$$|d_n| \leq 1.8 \times 10^{-26}\ e\ \text{cm} \tag{2.5}$$

being seven orders of magnitude more precise than the first measurement in the 1950s. This upper limit directly implies an upper limit on θ of

$$\theta \lesssim 10^{-10}. \tag{2.6}$$

In a nutshell, the strong CP problem is the mystery of why θ is so minuscule, given that, as an angle between 0 and 2π, it would a priori be expected to be $\mathcal{O}(1)$. This is generally considered the strongest evidence for the lack of CP violation by the strong interaction, and puts the strictest numerical bound on θ. However, it is also important to note that there is no evidence of it in other strong processes either.

2.2.1.2 ALP models

A new particle, the axion, was proposed by Weinberg [38] and Wilczek [39] in response to the solution to the strong CP problem published by Peccei and Quinn [40, 41]. Essentially, this solution modifies Equation (2.2) to include a pseudoscalar field a

$$\mathscr{L} = \left(\theta + \frac{a}{f_a}\right)\frac{g_s^2}{32\pi^2}G^a_{\mu\nu}\tilde{G}^{a\mu\nu}. \tag{2.7}$$

The mechanics of this solution to the strong CP problem are covered fairly exhaustively in the literature, and are not directly relevant to the content of this thesis beyond being a strong motivator for the experimental work. A recent review of modern axion physics gives a good description of the mechanism [33].

Axions specifically address the strong CP problem, while axion-like particles (ALPs) share similar properties but may not solve this problem [42–59]. Regardless of whether they are axions or axion-like, these particles could mediate interactions with a hidden sector or be dark matter candidates themselves. Furthermore, ALP models can help to explain the so-called muon $g-2$ anomaly [60]. A simple model where ALPs can couple to the SM $SU(2)_W$ gauge boson fields has the Lagrangian

$$\mathscr{L}_a = \frac{1}{2}(\partial_\mu a)^2 - \frac{1}{2}m_a^2 a^2 + \frac{g_{aW}}{4}aW_{\mu\nu}\tilde{W}^{\mu\nu}, \tag{2.8}$$

where g_{aW} is the coupling strength between the ALP and the W.

In the simple case where ALPs exclusively couple to the $SU(2)_W$ gauge bosons shown in Ref. [61], they can facilitate $b \to s$ transitions leading to decays such as $B^\pm \to K^\pm a$,

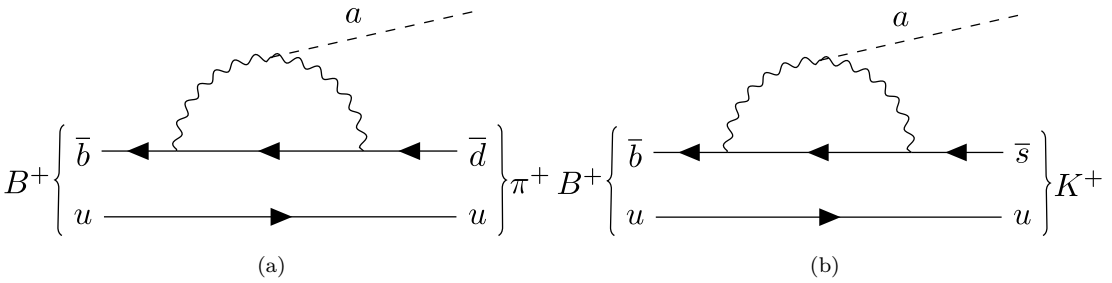


Figure 2.2: Feynman diagrams for two flavour-changing neutral current ALP production processes described in Ref. [61].

the decay rate of which is given by

$$\begin{aligned}\Gamma(B^{\pm} \to K^{\pm} a) &= \frac{M_B^3}{64\pi}|g_{abs}|^2\left(1-\frac{M_K^2}{M_B^2}\right)^2 f_0^2(M_a^2)\lambda_{Ka}^{\frac{1}{2}},\\ g_{abs} &= -\frac{3\sqrt{2}G_F M_W^2 g_{aW}}{16\pi^2}\sum_{\alpha\in c,t} V_{\alpha b}V_{\alpha s}^* f\left(\frac{M_\alpha^2}{M_W^2}\right),\\ \lambda_{Ka} &= \left(1-\frac{(M_a+M_K)^2}{M_B^2}\right)\left(1-\frac{(M_a-M_K)^2}{M_B^2}\right),\end{aligned} \tag{2.9}$$

where $f_0(q)$ is a form factor from the hadronic matrix elements [62], V_{ij} are elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix, λ_{Ka} is a phase space term from two-body kinematics, and

$$f(x) = \frac{x[1+x(\log x - 1)]}{(1-x)^2}. \tag{2.10}$$

Analogous expressions can be written for $B^{\pm} \to \pi^{\pm} a$ by replacing M_K with M_π and s with d as the $\pi^{\pm}$ and $K^{\pm}$ belong to the same meson octet. ALPs that generate these flavour-transition decays are sometimes dubbed *axiflavons* [63, 64]; however, for the rest of this work the more generic term *ALP* is used.

The possible ALP mass range spans many orders of magnitude, and excluding the presence of ALPs requires very different methods depending on the mass scale. If ALPs couple to SM particles, then they can be produced at colliders at mass scales up to the collider energy scale. Collider probes of ALPs commonly explore the ALP-photon coupling via processes that exclusively involve electrons and photons. In meson decays, however, the ALP to W field coupling g_{aW} and the mass m_a are of great interest. There are several experimental searches that can be used to constrain the areas in which ALPs can exist in this coupling-mass parameter space. These are presented in Fig. 2.7, which is, for the most part, a recreation of Figure 2 in Ref. [61] with some revisions and updates.

2.2.1.3 ALP production in quark annihilation

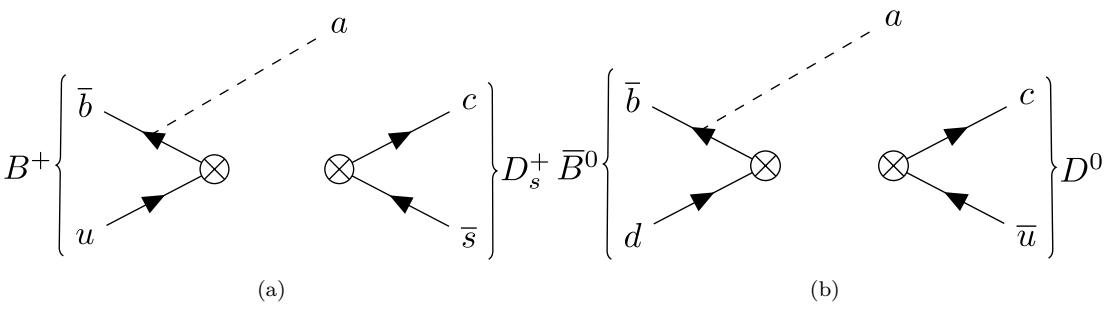


Figure 2.3: Feynman diagrams for two quark annihilation ALP production processes described in Ref. [2].

Recently, new ALP production mechanisms in the flavour sector have been investigated [2]. While the majority of literature discussing ALPs in B decays concentrates on $b \to s$ transitions, this new process allows for ALP production in conjunction with a wider variety of mesons, such as the decays shown in Fig. 2.3. In this treatment, we consider ALP couplings defined by the Lagrangian

$$\mathscr{L} = g_q a \bar{q} \gamma_5 q, \tag{2.11}$$

where g_q is the coupling between the ALP and quark, and $q \in u, d, s, c, b$. Using form factors computed with perturbative QCD (pQCD), the product $\mathcal{B}(B \to ha) g_q^2$ can be calculated, shown in Table 2.1. Smaller values of $\mathcal{B}(B \to ha) g_q^2$ indicate that weaker couplings g_q can be probed, making the channel more sensitive to ALP interactions. For example, with a branching fraction sensitivity of 10^{-6} for both $B^\pm \to \pi^\pm a$ and $B^\pm \to K^\pm a$, we would be able to set a limit on g_q a factor of $\sim \sqrt{40}$ lower for the channel containing the pion. This motivates searches for ALPs in several different channels, as different mesons in the final state allow for the best limits of different couplings g_q.

2.2.2 Dark baryons

To explain the observed baryon asymmetry of the universe, models have been proposed in which B mesons decay into a dark sector antibaryon ψ_D [65], transferring baryon number to the dark sector in the early universe. This could explain the matter-dominated universe we observe today. Explicit models that realise this mechanism can have signatures at collider experiments [66]. In these models, a TeV-scale colour-triplet scalar Y mediates an interaction between the b quark, an up-type quark, a down-type quark, and a dark baryon ψ_D, allowing B decays such as the one shown in Fig. 2.4. In this thesis, we separate this model from the generic scalar portal discussed in Section 2.2.4 as Y carries baryon number, whereas the more generic scalar portal does not. This decay is

Table 2.1: Results from pQCD computations of $\mathcal{B}(B \to ha)g_q^{-2}$ from Ref. [2].

h	$\mathcal{B}(B \to ha)g_q^{-2} \times 10^{-6}$				
	up	down	strange	charm	bottom
D^0	0.270	3.700	< 0.001	0.062	0.004
$D^*(2007)^0$	0.155	3.462	< 0.001	0.011	< 0.001
D_s	3.573	< 0.001	0.305	0.129	0.002
D_s^*	3.403	< 0.001	0.066	0.025	< 0.001
$K^\pm$	2.860	< 0.001	1.659	< 0.001	0.001
$K^{*\pm}$	0.765	< 0.001	1.370	< 0.001	< 0.001
K^0	< 0.001	2.220	1.510	< 0.001	0.0012
K^{*0}	< 0.001	1.090	1.590	< 0.001	< 0.001
$\pi^\pm$	2.080	0.320	< 0.001	< 0.001	0.000025
$\rho^\pm$	2.810	0.175	< 0.001	< 0.001	0.000062

governed by the operator $\mathcal{O}_{\psi bud}$, with the subscript denoting the particles involved in the interaction. The other possible operators are $\mathcal{O}_{\psi bcd}$, $\mathcal{O}_{\psi bus}$, and $\mathcal{O}_{\psi bcs}$. We tabulate several other decays implied by this interaction, involving different sets of quarks, in Table 2.2.

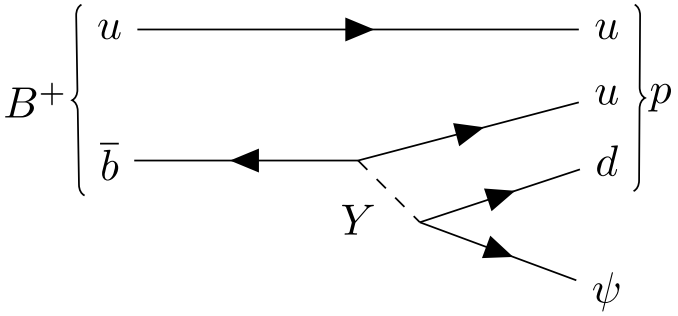


Figure 2.4: A diagram of the decay $B^+ \to p\psi_D$, mediated by the colour-triplet scalar Y.

To ensure proton stability in this SM extension, the dark baryon must satisfy

$$m_{\psi_D} > m_p - m_e \approx 938\,\mathrm{MeV}/c^2. \tag{2.12}$$

Additionally, the decay $B \to \psi_D p$ is allowed only if

$$m_{\psi_D} < m_B - m_p \approx 4.34\,\mathrm{GeV}/c^2. \tag{2.13}$$

These constraints indicate that, if the model is correct, searches for the dark baryon ψ_D should focus on the mass range of 0.94 to 4.34 GeV/c^2. The B-mesogenesis parameter space is directly sensitive to the branching fraction of $B \to \psi_D \mathcal{B}\mathcal{M}$ where $\mathcal{M}$ is one or

more mesons and $\mathcal{B}$ is a baryon. The relation between this multi-body decay rate and the two-body decay rate is discussed in Ref. [66].

Table 2.2: The lightest final states possible for B^0 and B^+ decays for all B-mesogenesis operators, adapted from Ref. [66]. Decays of B_s mesons, as well as b-containing baryons, are also possible, but omitted from this table as they are not relevant to the present work.

Operator	B^0 decay	B^+ decay
$\mathcal{O}_{\psi bud}$	$B^0 \to \psi n$	$B^+ \to \psi p$
$\mathcal{O}_{\psi bus}$	$B^0 \to \psi \Lambda$	$B^+ \to \psi \Sigma^+$
$\mathcal{O}_{\psi bcd}$	$B^0 \to \psi \Lambda_c^+ \pi^-$	$B^+ \to \psi \Lambda_c^+$
$\mathcal{O}_{\psi bcs}$	$B^0 \to \psi \Xi_c^0$	$B^+ \to \psi \Xi_c^+$

2.2.3 Vector portal

ALPs and dark baryons were the main motivations for the experimental search described in this thesis. However, many other models can also be probed in two-body B decays with one invisible daughter. By introducing an additional gauge group for dark matter, $U(1)_d$, which mixes with the $U(1)_Y$ group of the SM [67, 68], a new dark vector boson can be generated. There has been a renewed focus on this class of models in the context of the flavour sector of late, as the excess in $B^\pm \to K^\pm \nu\overline{\nu}$ seen by Belle II could be explained with vector portal models [69].

Typically, dark vector boson interactions are described by the Lagrangian

$$\mathscr{L}_{\text{gauge}} = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \frac{1}{2} \frac{\epsilon}{\cos\theta_W} B_{\mu\nu} Z_d^{\mu\nu} - \frac{1}{4} Z_{d\mu\nu} Z_d{}^{\mu\nu} \tag{2.14}$$

where ϵ is a dimensionless kinetic mixing parameter and the weak mixing angle is θ_W. The kinetic mixing parameter encodes the interaction between the $B_{\mu\nu}$ and $Z_{\mu\nu}$ gauge fields. After spontaneous symmetry breaking, a dark vector boson Z_D mixes with the SM Z, with the mass eigenstates represented by a linear combination of weak eigenstates, parameterised by the mixing angle ξ:

$$\begin{aligned} Z &= Z^0 \cos\xi - Z_D^0 \sin\xi \\ Z_D &= Z^0 \sin\xi + Z_D^0 \cos\xi. \end{aligned} \tag{2.15}$$

From ξ, we define a mass mixing parameter $\epsilon_Z = \frac{1}{2}\tan 2\xi$, which encodes the mixing between the two gauge bosons Z and Z_D. In the special case where the $U(1)_d$ symmetry is broken by a scalar SM singlet, $\xi = 0$ and no mixing occurs between the SM Z boson and the dark vector boson. The vector bosons generated by this special case are called dark photons. For non-zero mixing, the new vector boson Z_D is instead referred to as a dark Z boson.

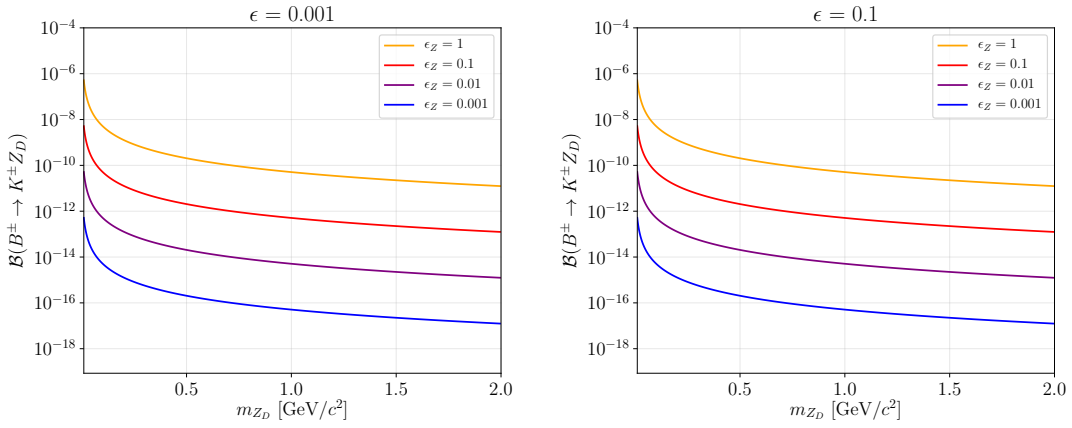
The production of the Z_D (or dark photon) in B meson decays is discussed in Ref. [70], which provides the branching fraction

$$\mathcal{B}(B \to K Z_D) = \frac{|g_V^\nu|^2}{64\pi} \frac{m_B^3 \lambda_{BKZ_D}^{\frac{3}{2}}}{m_B^3 m_{Z_D}^2 \Gamma_B} [f_+^{BK}(m_{Z_D}^2)]^2, \tag{2.16}$$

where the $B \to K$ form factor is denoted as f_+^{BK} [62],

$$\lambda_{xyz}(q^2) \equiv m_x^4 + m_y^4 + q^4 - 2(m_x^2 m_y^2 + m_y^2 q^2 + m_x^2 q^2), \tag{2.17}$$

and g_V^ν is the effective interaction strength of the sbZ_D couplings, as described in Ref. [71]. The interaction strength g_V^ν is related to the Wilson coefficient $(E_{c1,c2}^{0,Z})_Z$ [72], which depends linearly on ϵ_Z, as well as some SM particle masses and CKM matrix elements.



(a) (b)

Figure 2.5: The $B^\pm \to K^\pm Z_D$ branching fraction as predicted by Equation (2.16) for two different values of the kinetic mixing parameter ϵ.

The decays of the Z_D can produce SM leptons, hadrons, or other dark sector particles. The branching fractions to SM leptons are

$$\mathcal{B}(Z_D \to \nu\overline{\nu}) = \frac{e^2 m_{Z_D} \epsilon_Z^2}{96\pi c_W^2 s_W^2}, \tag{2.18}$$

$$\begin{aligned}\mathcal{B}(Z_D \to \ell^+\ell^-) = \frac{e^2}{96\pi m_{Z_D} c_W^2 s_W^2 \Gamma_{Z_D}} \sqrt{1 - 4\frac{m_\ell^2}{m_{Z_D}^2}} \Big(& 8 c_W^2 s_W^2 \epsilon^2 (2m_\ell^2 + m_{Z_D}^2) \\ & - 4 c_W s_W (4 s_W^2 - 1) \epsilon \epsilon_Z (2m_\ell^2 + m_{Z_D}^2) \\ & + \epsilon_Z^2 (m_\ell^2 (16 s_W^4 - 8 s_W^2 - 1) + m_{Z_D}^2 (8 s_W^4 - 4 s_W^2 + 1)) \Big),\end{aligned} \tag{2.19}$$

where e is the elementary charge, and c_W and s_W are the cosine and sine of the Weinberg angle θ_W respectively.

The branching fractions to hadrons are more complex but were recently calculated in Ref. [73], based on the ratio of the e^+e^- cross section to hadronic final states $\mathcal{H}$ and dimuon final states

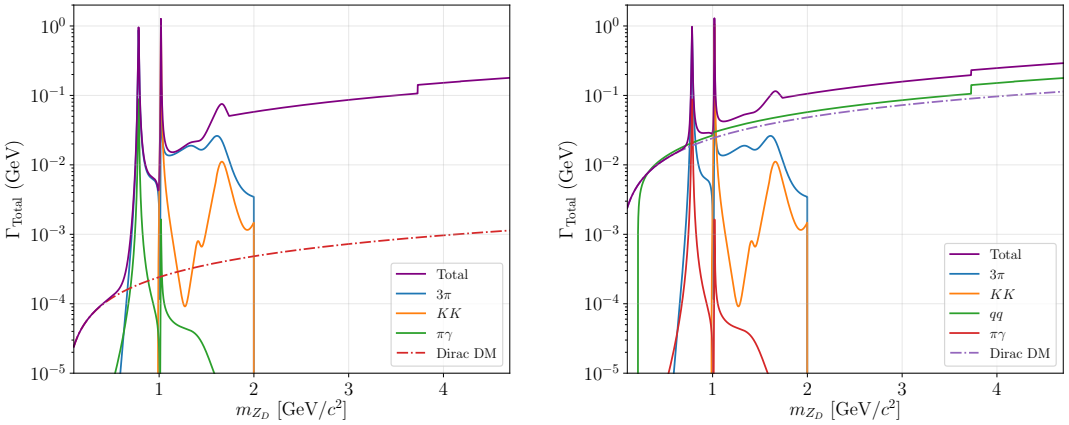
$$\Gamma(Z_D \to \mathcal{H}) = \Gamma(Z_D \to \mu^+\mu^-) \times R_\mu^{\mathcal{H}}, \tag{2.20}$$

where $R_\mu^{\mathcal{H}} = \sigma(e^+e^- \to \mathcal{H})/\sigma(e^+e^- \to \mu^+\mu^-)$. The authors of Ref. [73] implemented the above hadronic decay rates in the Python package `DeLiVeR`, which we use to calculate relevant Z_D decay rates later on in this work. The `DeLiVeR` package also allows one to calculate decay rates for different hypothetical dark sectors, which can greatly influence the total $Z_D \to$ invisible rate, as $\mathcal{B}(Z_D \to \nu\overline{\nu})$ is very small ($< \mathcal{O}(10^{-6})$). With large couplings to dark matter, and depending on the construction of the dark sector, the branching fraction of a vector mediator to hidden sector final states can reach as high as 40–50% for $m_{Z_D} > 2\,\mathrm{GeV}/c^2$, and nearly 100% for masses below the pion mass. Using this framework for the decays of a dark vector mediator, we find that the only scenario where $Z_D \to$ invisible decays have a substantial branching fraction is if there is a large coupling between the vector mediator and the dark sector. The lifetime of Z_D is generally on the nanosecond scale or shorter, depending on the specific coupling strength scenario, such as the two shown in Fig. 2.6.

While B decays involving the vector portal are possible, establishing a benchmark case for easy comparisons between experiments — such as those presented for the pseudoscalar and scalar portals [61, 74] — is more challenging. Because of this, the interpretation of results in Chapter 8 does not include the vector portal. However, one should keep in mind that there are models that allow vector mediator production in the flavour sector, and the final states of these decays are dependent upon the model and dark sector.

2.2.4 Scalar portal

The only fundamental scalar particle known to date is the Higgs boson, but there could be others. A second scalar could possibly mix with the SM Higgs boson [75, 76] and be much lighter than the SM Higgs [77], which can lead to solutions to outstanding problems in physics. Light new scalars can themselves be dark matter candidates [78], or they may mediate interactions with a hidden sector [79]. Through cosmological relaxation (whereby the EW scale is set in the early universe) they could generate the electroweak hierarchy [80], explaining why the EW scale and Higgs boson mass are so much smaller than the Planck scale. Baryogenesis can also be achieved with new scalar models [81–83]. It is also possible for new scalars to be inflatons in the early universe [84]. Furthermore, there has been a renewed interest in dark scalars in flavour-changing neutral current processes that can explain the $B^\pm \to K^\pm\nu\overline{\nu}$ excess reported by Belle II



(a) (b)

Figure 2.6: Total decay width of a Z_D boson assuming the dark sector consists of some massless dark fermions χ, for two different values of the dark sector coupling strength, calculated with the `DeLiVeR` package. Below $m_{Z_D} = 2\,\text{GeV}/c^2$, the package does the calculation by summing the decay rates to exclusive final states, such as $Z_D \to 3\pi$, $Z_D \to KK$. Above 2 GeV/c^2, it uses an inclusive $Z_D \to qq$ rate.

[85]. The phenomenology of the scalar portal in the context of B decays is detailed in Ref. [74], including several predicted sensitivities.

Let us introduce a basic model following Ref. [74], which introduces a new scalar field ϕ that interacts with the Higgs field, as well as fermions χ that belong to some hidden sector. This model can be described by the Lagrangian

$$\mathscr{L} = -\frac{1}{2}m_\phi^2\phi^2 - \lambda_3|H|^2\phi - y_\chi\chi\bar{\chi}\phi - \frac{1}{2}\,\mathrm{m}_\chi\chi\bar{\chi}. \tag{2.21}$$

Following electroweak symmetry breaking, the scalar ϕ undergoes mixing with the Higgs field, resulting in a dark scalar S and the SM Higgs boson h, governed by a mixing angle θ. The experimental signature left by the dark scalar very much depends on this mixing angle as it governs the proportion of the dark scalar decays to visible and hidden sector final states

$$\Gamma_S = s_\theta^2\Gamma_{SM} + c_\theta^2\Gamma_{\chi\chi}, \tag{2.22}$$

where s_θ and c_θ are shorthand for $\sin\theta$ and $\cos\theta$ respectively. For $m_S < 2m_\chi$, decays to the hidden sector are kinematically forbidden, so the dark scalar decay must be to a visible final state. Invisible decays dominate above this kinematic limit, as long as $y_\chi c_\theta > m_\ell s_\theta/v$, which is generally satisfied even for small dark fermion couplings y_χ since the Higgs field vacuum expectation value v is much larger than the mass of charged leptons. The branching fractions of the dark scalar to dark fermions, leptons,

and hadrons are:

$$\mathcal{B}(S \to \chi\chi) = \frac{y_\chi^2 c_\theta^2 m_S}{8\pi\Gamma_S}\left(1 - \frac{4m_\chi^2}{m_S^2}\right)^{3/2}; \tag{2.23}$$

$$\mathcal{B}(S \to \ell\ell) = \frac{m_\ell^2 s_\theta^2 m_S}{8\pi v\Gamma_S}\left(1 - \frac{4m_\ell^2}{m_S^2}\right)^{3/2}; \tag{2.24}$$

$$\mathcal{B}(S \to \pi\pi) = \frac{3s_\theta^2 G_F}{16\sqrt{2}\pi m_S\Gamma_S}\left(1 - \frac{4m_\pi^2}{m_S^2}\right)^{1/2}\left|\frac{7}{9}\Gamma_\pi + \frac{7}{9}\Delta_\pi + \frac{2}{9}\Theta_\pi\right|^2; \tag{2.25}$$

and

$$\mathcal{B}(S \to KK) = \frac{s_\theta^2 G_F}{4\sqrt{2}\pi m_S\Gamma_S}\left(1 - \frac{4m_K^2}{m_S^2}\right)^{1/2}\left|\frac{7}{9}\Gamma_K + \frac{7}{9}\Delta_K + \frac{2}{9}\Theta_K\right|^2; \tag{2.26}$$

where $v = 246\,\text{GeV}$ is the Higgs field vacuum expectation value and Γ, Δ, and Θ are form factors described in Ref. [86], from which we take all the dark scalar branching fraction predictions. In that work, Equations (2.25) and (2.26) are used for $m_S < 2$ GeV/c^2, while for $m_S > 2$ GeV/c^2, an inclusive decay rate to hadrons is used.

If one assumes that dark fermions χ are very light, then the decay rate of the dark scalar is, as previously mentioned, dominated by $S \to \chi\bar{\chi}$, which results in a missing energy signature for $B^\pm \to K^\pm S$ production at colliders.

For heavy dark fermions, the decay $S \to \chi\bar{\chi}$ is forbidden, and S will exclusively decay to visible final states. However, as S has a finite lifetime that scales with s_θ^{-2}, it can become long-lived at detector scales. If the lifetime is large enough, the dark scalar could escape a detector before it decays, meaning that the missing energy signature can also be used to set limits even with the assumption that there are no invisible final states. A missing energy search is therefore more powerful for longer lifetimes, i.e. smaller values of s_θ. At the same time, however, the dark scalar production rate scales with s_θ^2. As a result, the regions of parameter space that are most accessible experimentally (based on the decay length of the scalar alone) are suppressed by this production rate dependency. In a collider experiment, the lab-frame decay length of a particle also depends on its Lorentz boost, which, for a given momentum, will be higher for lighter particles. This means that missing energy searches are better at constraining long-lived particles the lighter they are.

The production rate of the dark scalar from B decays is

$$\begin{aligned}\mathcal{B}(B^+ \to K^+ S) =& \frac{\sqrt{2}G_F|C_{bs}|^2(m_b + m_s)^2}{64\pi\Gamma_{B^+}m_B^3(m_b - m_s)^2} f_0^2(m_S^2)(m_B^2 - m_K^2)^2 \\ &\times \left((m_B^2 - m_K^2 - m_S^2)^2 - 4m_K^2 m_S^2\right)^{1/2}\end{aligned} \tag{2.27}$$

where the Wilson coefficient C_{bs} is, to leading order in m_S/m_W,

$$C_{bs} = \frac{3\sqrt{2}G_F m_t^2}{16\pi^2} V_{tb}V_{ts}^* s_\theta + \mathcal{O}\left(\frac{m_S^2}{m_W^2}\right), \tag{2.28}$$

$f_0(q^2)$ is a form factor for $B \to K$ transitions [87], s_θ is the sine of the mixing angle θ and G_F is the Fermi constant.

2.2.5 Other models

In the R-parity violating (RPV) supersymmetry model detailed in Ref. [88], the process $B^+ \to \tilde{\chi}_1^0 p$ is possible, where $\tilde{\chi}_1^0$ is a light neutralino. This neutralino would be invisible to detectors, meaning this model could be constrained with the same measurement as the B-mesogenesis model.

2.3 Experimental landscape

Several experimental approaches can be used to constrain the parameter spaces of the models described above. Many of these constraints come from collider and beam-dump experiments. Regions in the mass-coupling plane that have been ruled out experimentally are referred to as *excluded regions*. Each exclusion region is associated with a confidence level (CL), typically 90% or 95%, as reported by experiments.

2.3.1 ALP searches

ALPs could have couplings to a number of visible particles and possible masses ranging across several orders of magnitude. Because of this, detection methods for ALPs are extremely varied [89, 90]. Some of the most imaginative ALP experiments are the so-called *light shining through a wall* experiments, where a high-intensity laser is shot at a wall with a strong magnetic field on both sides. If the laser photons convert to ALPs before the wall, they can pass through, and subsequently interact via the inverse Primakoff process in the magnetic field and convert back to detectable photons. Other experiments rely on natural sources of ALPs, like the Sun, and also aim to detect these inverse Primakoff photons. Such detectors are classed as *helioscopes*. As some ALPs are dark matter candidates, they could be the constituents of our galaxy's dark matter halo. The experiments that search for these ALPs are *haloscopes*. These experiments are generally sensitive to masses below the mass regions that traditional particle physics experiments can access. Collider experiments serve as probes for particles that many axion physicists might consider to have very high masses, certainly above the mass scale

of the QCD axion. An experiment can probe different ALP couplings depending on the ALP-detector interaction. These couplings commonly include, but are not limited to, the ALP coupling to photons, electrons, nucleons, and W bosons.

To simplify the analysis of ALP measurements, it is useful to consider one or a subset of these couplings at a time. Traditionally, the most common coupling of interest is the coupling to photons, $g_{a\gamma\gamma}$. At colliders, this is best constrained by looking for visible decays of ALPs. However, in the flavour sector in particular, the coupling to the W field g_{aW} is also important to consider. The experiments that have performed searches relevant to the coupling g_{aW} were described in Ref. [61], and are presented below.

- The E787 experiment and its successor the E949 experiment were beam-dump experiments at Brookhaven National Laboratory in the USA. They searched for the rare decay $K^+ \to \pi^+\nu\overline{\nu}$ [91–93], which can be reinterpreted as $K^+ \to \pi^+ a$. This limit can be related to limits on the g_{aW} coupling under the assumption that the invisible part of the kaon decay comes from an invisible ALP rather than neutrinos. At the time that Ref. [61] was published, these were the most stringent exclusions of ALP parameter space in the mass ranges they could access, but these limits have since been superseded.

- The NA62 experiment, located on the French side of CERN, was designed to study the rare decays of kaons using the dumped beam of the Super Proton Synchrotron (SPS) accelerator [94]. In 2021, NA62 published the most sensitive search for the decay $K^+ \to \pi^+\nu\overline{\nu}$ to date [95], beating previous limits from E787 and E949. In Fig. 2.7, this 2021 branching fraction limit is mapped to g_{aW}.

- The BaBar experiment was a B-factory much like Belle, located at the SLAC National Accelerator Laboratory in the USA. During its operational period from 1999 to 2008, it recorded 514 fb^{-1} of data. As ALPs could be directly produced via processes such as $e^+e^- \to \gamma a$ if the ALP couples to γ or Z, electron-positron colliders can be powerful probes for them. BaBar looked for invisible ALPs produced in processes such as these with a search for events with a single photon final state. At masses above 4.7 GeV/c^2, these are the most stringent limits on g_{aW}. BaBar has also published numerous searches for the decay $B \to K\nu\overline{\nu}$. The most widely cited of these is Ref. [96], as it is the lowest branching fraction exclusion (apart from the search that it replaced [7]), despite newer analyses (some with larger datasets) not setting such strict branching fraction upper limits.

- Belle II recently reported a hint of an excess in $B \to K\nu\overline{\nu}$ events [97], inconsistent with the branching fraction exclusion set by BaBar. The SM expectation is $\mathcal{B}(B^+ \to K^+\nu\overline{\nu}) = (5.58 \pm 0.37) \times 10^{-6}$ [98], and yet the Belle II result points to a rate of around twice this value. The inclusive tag result is already excluded by

the BaBar measurement; however, the excess from the hadronic tag analysis is not yet ruled out.

Furthermore, one can project limits onto the parameter space for future searches. It is easier to compare experiments like BaBar and Belle II, particularly when Belle II has collected its full dataset, as both are e^+e^- colliders. The sensitivity of the $B \to K\nu\overline{\nu}$ and mono-photon searches will be compared. Branching fraction sensitivity typically scales with the square root of the luminosity. Belle II's target luminosity is about 100 times larger than BaBar's, so one would expect a factor of 10 enhancement in branching fraction or cross section sensitivity. Additionally, Belle II has dedicated triggers for low-multiplicity events, which were lacking in Belle, further enhancing its ability to perform the mono-photon final state search. Under these simple assumptions, Belle II should be sensitive to at least $\mathcal{B}(B^+ \to K^+a) \sim 10^{-6}$. The BaBar mono-photon search was carried out with a comparatively small dataset of 23 fb^{-1} collected at the $\Upsilon(3S)$ resonance, which makes comparisons slightly harder considering Belle II has no plans to take a similar dataset. However, if a search for similar topologies at the $\Upsilon(4S)$ resonance is comparable in sensitivity to that of the one at the $\Upsilon(3S)$ resonance (a reasonable assumption), then Belle II could exclude $\mathcal{B}(\Upsilon(4S) \to \gamma a) \sim 10^{-8}$ thanks to the much larger dataset.

In the few MeV/c^2 to few GeV/c^2 ALP mass region that is the focus of this thesis, the coupling g_{aW} for invisibly decaying ALPs is constrained by a measurement of $B \to K\nu\overline{\nu}$ by BaBar. The current state of the parameter space for the coupling g_{aW} is shown in Fig. 2.7. At lower masses, collider and fixed-target experiments become less relevant, and bounds are dominated by dedicated axion experiments and limits derived from astrophysical observations [99].

2.3.2 Dark scalar searches

Many of the measurements that constrain the ALP parameter space can also be mapped onto the scalar portal parameter space.

- The E949 experiment's search for the process $K^+ \to \pi^+\nu\overline{\nu}$ [93] can also be interpreted as a search for a dark scalar being produced and going undetected in charged kaon decays. The phenomenology of $s \to d$ transitions is described in Ref. [74], and is analogous to that of $b \to s$ transitions discussed in the context of B decays.
- At the NA62 experiment, two relevant searches were performed. One is the aforementioned search for $K^+ \to \pi^+\nu\overline{\nu}$ [95] that supersedes the E949 result for most values of m_S, and the other is a search for invisible decays of the π^0 [100]. The

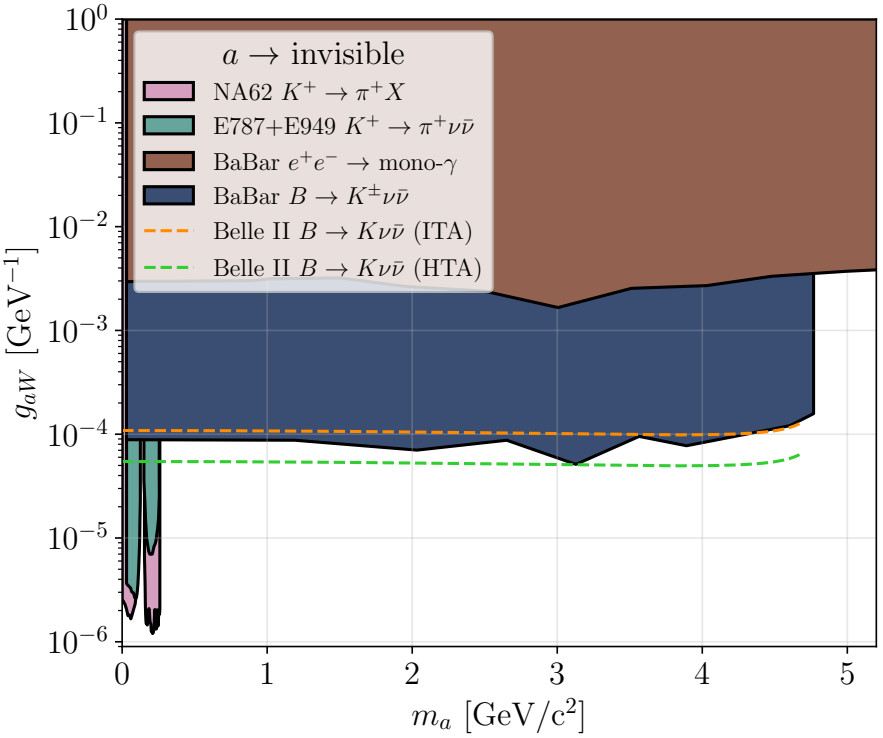


Figure 2.7: A recreation of Fig. 3 from Ref. [61], with some updated measurements added, including a search for $K^+ \to \pi^+ X$ at NA62. The branching fractions for $B^+ \to K^+ \nu\bar{\nu}$ measured by Belle II with hadronic (HTA) and inclusive (ITA) tagging methods are shown in dotted lines.

latter is only relevant to invisible particles with masses very close to the π^0 mass. The NA62 limits are the most stringent limits on the scalar portal parameter space below a new scalar mass of about 300 MeV/c^2, apart from a window around the π mass.

- PS191 was another CERN beam-dump experiment designed to search for sterile neutrinos in the decays of mesons. The lack of signal was interpreted as a limit on the scalar portal parameter space in 2021 [101], despite the fact that the last results from the experiment were published in 1988 [102].

- The CERN-Hamburg-Amsterdam-Rome-Moscow (CHARM) Collaboration, another experiment from the 1980s, searched for pairs of displaced leptons [103] and found no signal. The search was reinterpreted in the context of the scalar portal more recently [104].

- MicroBooNE, a liquid argon time projection chamber located in a neutrino beamline at Fermilab in the USA, has searched for dark scalars decaying to electron-positron pairs [105]. It relies on kaon decays to produce the dark scalars. The

limits from MicroBooNE have no veto around the π mass, and in that small region of parameter space provide the best limits on the dark scalar mixing angle.

- KTeV, another Fermilab experiment, was designed primarily to probe CP violation in the kaon sector. It searched for $K_L^0 \to \pi^0 \mu^+ \mu^-$ [106], the results of which were reinterpreted in Ref. [86] in the context of the scalar portal.
- The L3 experiment was one of the four main detectors at the Large Electron-Positron Collider (LEP), the collider that preceded the Large Hadron Collider (LHC) at CERN. It searched for new scalar particles with masses below 70 GeV/c^2 in the process $e^+e^- \to Z^*S$ (where the * marks the Z boson as virtual) [107]. Jets were used to reconstruct the final states in this search, as at these relatively high masses, many decays to hadronic final states are accessible to S.
- In 2015, BaBar presented a search for long-lived particles in e^+e^- collisions [108]. The search looked for pairs of oppositely charged tracks with a displaced vertex for several different species of charged particles. The results were reinterpreted in the context of dark scalars in Ref. [74].
- LHCb is one of the four main LHC experiments. They have published direct searches for channels with displaced muon pairs $B^+ \to K^+ S(\to \mu^+\mu^-)$ [109] and $B^0 \to K^{*0} S(\to \mu^+\mu^-)$ [110]. The limits corresponding to these searches are reported in Ref. [86]. Projections for the sensitivities of future LHCb searches are described in Ref. [111]. For new scalar masses from 300 MeV/c^2 to a few tens of GeV/c^2, these are the most stringent limits to date.
- Another LHC experiment, the Compact Muon Solenoid (CMS) experiment, used a special trigger for low-momentum muons to search for long-lived particles [112]. This search was rather agnostic to the long-lived particle production environment, so to date there is no literature that maps these results onto the scalar portal parameter space that is shown in this work. However, it would be an incomplete discussion without mentioning it, as CMS is sensitive to processes with $b \to sS$ transitions.
- Belle II recently published a dedicated search for long-lived particles in a number of final states in processes containing $b \to sS$ [113].

The regions of the m_S–θ plane they exclude are shown in Figures 2.8 and 2.9, which are adapted from Refs. [74, 114].

2.3.3 Dark baryon searches

In the original paper proposing dark baryon searches at colliders [66], B decays to dark baryons are constrained by reinterpreting an ALEPH search for b-flavoured hadron

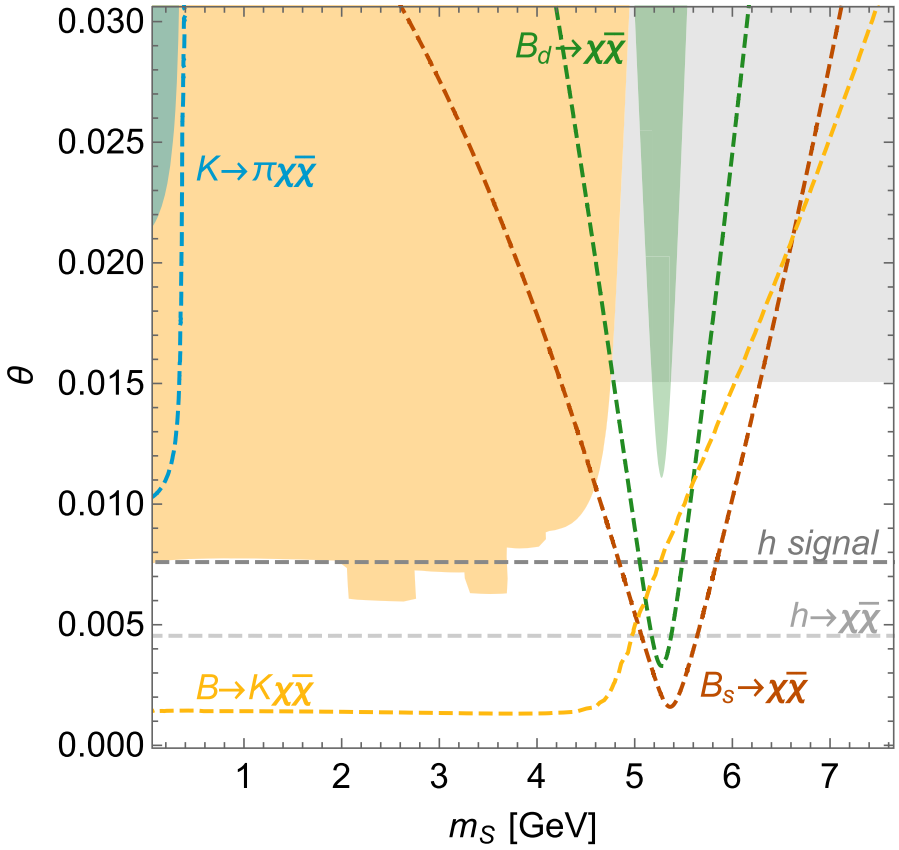


Figure 2.8: Dark scalar parameter space for invisibly decaying scalars, adapted from Ref. [74]. The shaded areas in the figure show the excluded regions of the scalar parameter space, while the areas bounded by dashed lines show the projected exclusion power of a number of upcoming experiments.

decays and large missing energy [115]. However, the ALEPH search did not require a baryonic decay of the b-containing hadron. The first dedicated measurement of a dark baryon decay channel was performed by BaBar in 2023 [116] in the channel $B^0 \to \Lambda\psi_D$, and followed up by a search for $B^+ \to p\psi_D$ later that year [117]. Belle also searched for the decay $B^0 \to \Lambda\psi_D$ [118]. The search for $B^+ \to p\psi_D$ excludes branching fractions compatible with B-mesogenesis for $m_{\psi_D} \gtrsim 3\,\mathrm{GeV}/c^2$ for the operator $\mathcal{O}_{\psi bud}$; however, the exclusion power varies rapidly as a function of m_{ψ_D} due to the cut-and-count method used.[2] The BaBar search for $B^0 \to \Lambda\psi_D$ shows similar variation, whereas the Belle search is far more constant as a function of m_{ψ_D}, but does not exclude branching fractions as small as the BaBar search. Both searches for $B^0 \to \Lambda\psi_D$ exclude branching fractions compatible with B-mesogenesis for $m_{\psi_D} \gtrsim 3.5\,\mathrm{GeV}/c^2$, with the BaBar search also providing some sporadic exclusions at lower masses. Recently, a BaBar search for $B^+ \to \Lambda_c^+\psi_D$ was published [119]. The branching fraction limits for these searches are

[2]The *cut-and-count* method is a straightforward approach used to search for new physics processes. It involves applying a series of selection criteria (cuts) to data to isolate a potential signal from background noise. After applying the selection criteria, the number of remaining events is counted and compared to the expected background prediction from the SM. A significant excess of events may indicate new physics, while consistency with predictions sets constraints on possible beyond-SM scenarios.

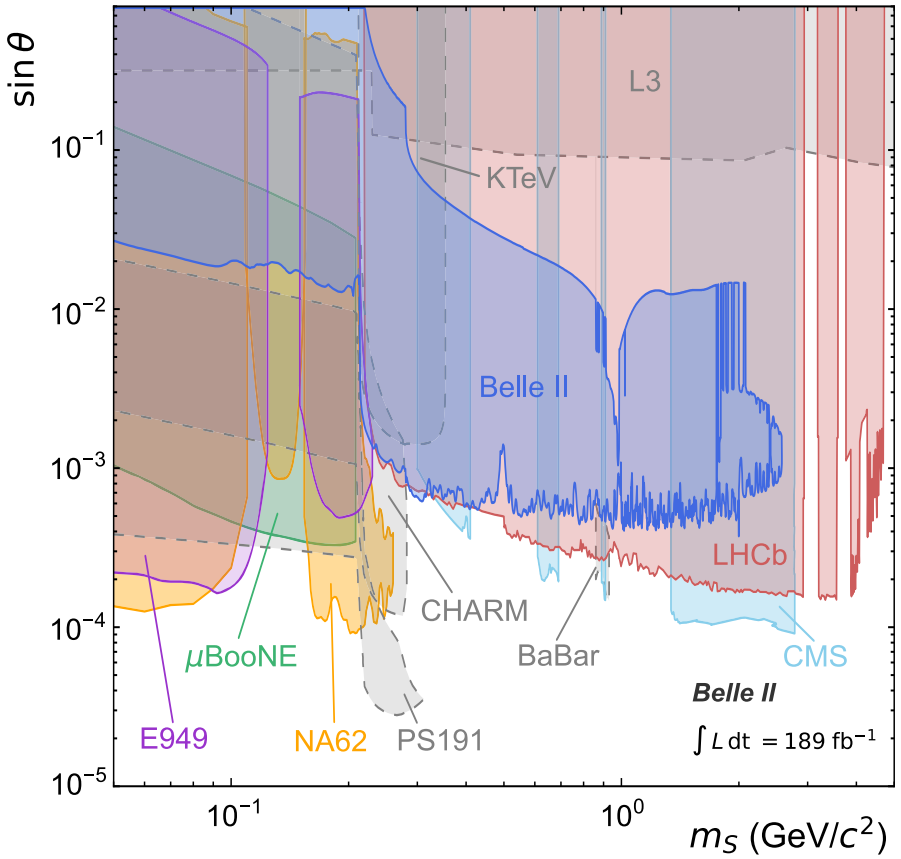
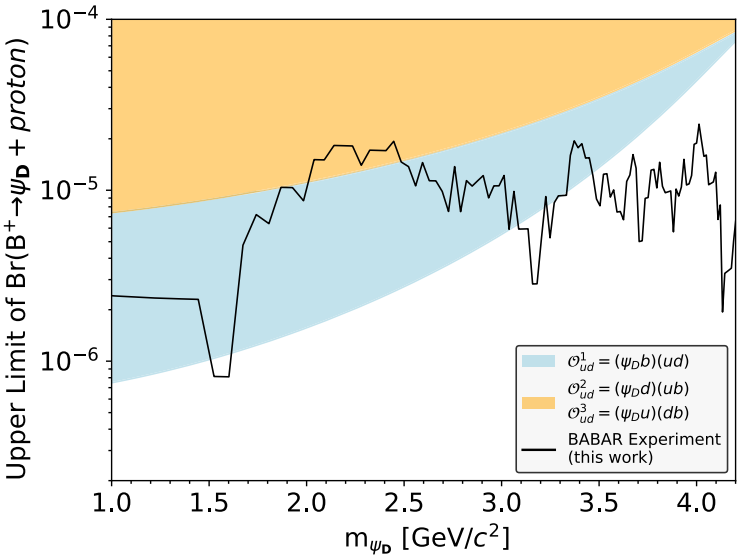


Figure 2.9: Dark scalar parameter space for visibly decaying scalars, adapted from Ref. [114]. The shaded areas in the figure show the excluded regions of the scalar parameter space, while the areas bounded by dashed lines show the projected exclusion power of a number of upcoming experiments.

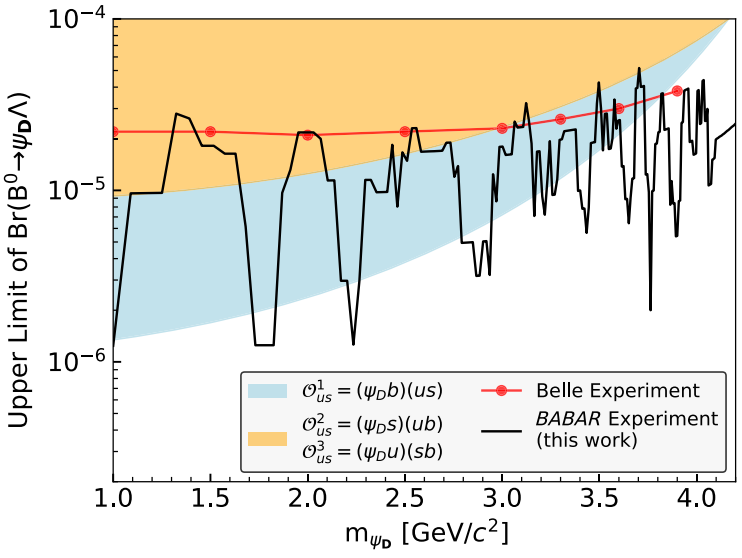
shown in Figures 2.10 and 2.11. Currently, the $\mathcal{O}_{\psi bud}$ and $\mathcal{O}_{\psi bus}$ operators are only compatible with B-mesogenesis at masses above about 3 GeV/c^2, so a search that is more sensitive at higher masses is required. The $\mathcal{O}_{\psi bcd}$ operator is still compatible with B-mesogenesis, with the BaBar search not having reached the sensitivity to exclude it. In all of the exclusion plots in Figures 2.10 and 2.11, the B-mesogenesis parameter space is defined as the area where the inclusive branching fraction of $B \to \psi_D \mathcal{BM}$ is greater than 10^{-4}.

2.4 Form factors

Form factors are a recurring concept in this work. Essentially, they describe the quantum-mechanical overlap between the initial- and final-state wavefunctions of particles in an interaction. These are impossible to determine analytically due to the infinite number of gluonic interactions that can occur within a hadronic current. The literature surrounding form factors is mostly concentrated on modelling SM processes so that decay rates



(a)



(b)

Figure 2.10: Current experimental constraints on B-mesogenesis channels. (a) The branching fraction limits for $B^+ \to p\psi_D$ from BaBar [117]. (b) The branching fraction limits for $B^0 \to \Lambda\psi_D$ from BaBar [116] and Belle [118]. The regions where B-mesogenesis is allowed are shaded in colour.

such as $\Gamma(B^+ \to K^+\ell^+\ell^-)$ can be accurately predicted. However, as the form factor deals with the hadronic part of the decay, these results can be generalised to NP decays such as $B^+ \to K^+ a$. We have already seen that they play a role in determining the theoretical production rates for ALPs, dark vector bosons, and dark scalars in Equations (2.9), (2.16) and (2.27). Furthermore, these form factors impact rate predictions for SM processes, and so their reliable calculation is essential in distinguishing real experimental evidence for NP. For example, different form factor calculations shifted the predicted value of $\mathcal{B}(B^+ \to K^+\nu\overline{\nu})$ by over 10% within the last decade [98, 120, 121].

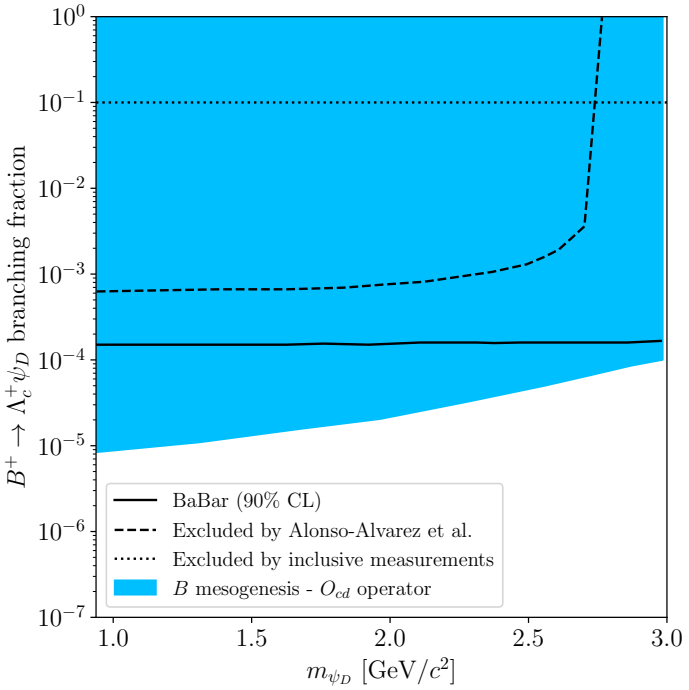


Figure 2.11: Branching fraction limits for $B^+ \to \Lambda_c^+ \psi_D$ from BaBar [119]. The region where the B-mesogenesis mechanism is allowed is shaded in blue.

In general, the value of a form factor is dependent on the momentum transfer q^2 between the initial- and final-state mesons. As this work is concerned with NP processes such as $B^+ \to K^+ a$, the form factors are generally evaluated at $q^2 = m_a^2$. Form factors appear in expressions for decay rates, and always contain a number of free parameters that need to be determined by theoretical calculations. There are a number of methods that are used to calculate them, all of which rely on some set of simplifying assumptions to make the problem calculable.

Lattice quantum chromodynamics (LQCD) simplifies the form factor calculation problem by discretising space and time into a four-dimensional lattice with defined spacing [122]. The result is extrapolated to a spacing of zero to recover the case of physical continuous QCD. Computationally, lattice simulations become more complex the smaller the overlap between the initial- and final-state particle wavefunctions is, meaning LQCD is generally more precise in phase space regions with large q^2.

Light-cone sum rules (LCSR) are another method for determining form factors, first used to calculate the branching fraction $\mathcal{B}(\Sigma^+ \to p\gamma)$ [123]. The method is an extension of QCD sum rules [124], which deal with two-point correlation functions of quark currents in a vacuum, to quark currents between a final-state meson and the vacuum, which are expanded around their light cone. LCSR calculates form factors in a nonphysical region of phase space where there is negative momentum transfer, and then extrapolates them to the low (positive) q^2 region.

In perturbative quantum chromodynamics (pQCD), calculations exploit the fact that at high momentum transfer, the strong coupling constant α_s becomes small due to asymptotic freedom [125]. This allows the use of perturbation theory to calculate the short-distance hard scattering amplitudes, which can then be convolved with long-distance parton distribution functions.

The variety of approaches taken to make the form factor calculation problem tractable leads to a landscape where one has to consider the validity or appropriateness of competing methods. This becomes particularly important when there is tension between their results. LQCD and pQCD both rely on extrapolating results from high q^2 to the low q^2 region of interest, while LCSR relies on an extrapolation from unphysical phase space (where $q^2 < 0$) to a physical region ($q^2 > 0$). It is important to remember that these methods are approximations of the true physics process, introducing inherent limitations. Each method has its strengths and weaknesses: LQCD provides non-perturbative first-principles calculations but is computationally intensive and limited at low q^2, whereas LCSR offers more flexibility in evaluating q^2-dependent quantities but requires careful treatment of the extrapolation process. Similarly, pQCD excels at high q^2 but becomes unreliable in the low q^2 regions where confinement effects dominate.

For studies of NP, such as those involving the rare decays discussed in this work, the precision of the form factor calculation is less critical than its consistency across different predictions. This is because the signal of NP typically manifests as a deviation from predicted rates or distributions, which can be identified even when absolute predictions have some uncertainty. Nevertheless, when competing form factor models predict significantly different branching fractions for the same process, the choice of model can introduce systematic uncertainties that propagate through the analysis. For this reason, cross-checking results against multiple approaches or adopting averaged form factor predictions, as commonly done in phenomenological studies, is a prudent strategy. Otherwise, using the same form factor model consistently across all predictions can help to mitigate the impact of these uncertainties.

The most relevant form factors in this work are those that describe $B^+ \to K^+$ and $B^+ \to \pi^+$ decays. These are parameterised as a function of q^2 based on results from LCSR calculations in Ref. [62]:

$$f_0(q^2) = \frac{r_2}{1 - q^2/m^2_{\text{fit}}}, \tag{2.29}$$

where r_2 and m^2_{fit} are fit parameters that take on different values depending on the mesons involved. The results from Ref. [62] that are relevant to this work are shown in Table 2.3. The form factor values as a function of q^2 are presented in Fig. 2.12.

Table 2.3: Form factor fit results from Ref. [62] for form factors relevant to this analysis.

Form factor	r_2	m^2_{fit}
f_0^π	0.258	33.81
f_0^K	0.330	37.46

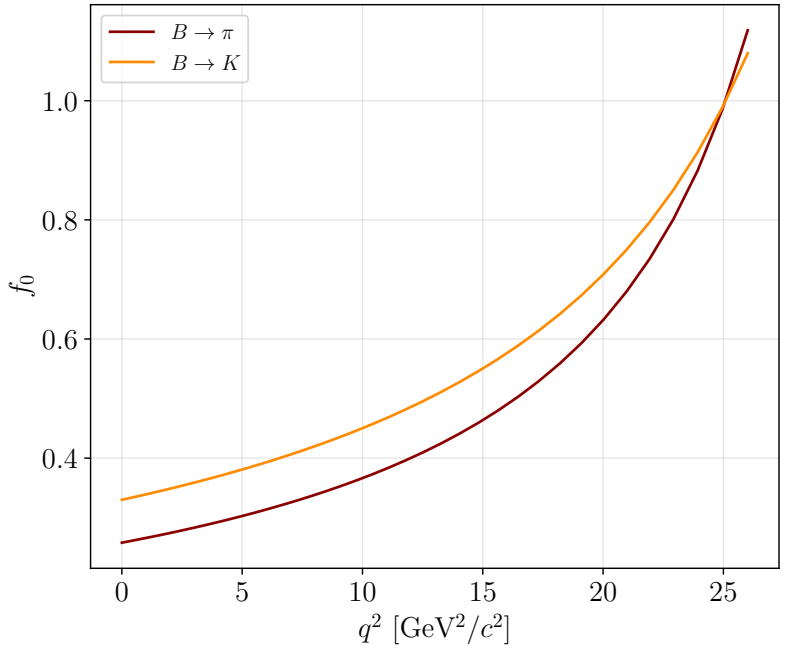


Figure 2.12: Form factors for $B \to \pi$ and $B \to K$ as a function of q^2, determined using the parameterisation and results in Equation (2.29) and Table 2.3.

3

The Belle experiment

This chapter describes the KEKB collider and the Belle detector, with details on each of its subdetectors. Software tools that are fundamental to analysing Belle data are also introduced.

Searches for hidden sector particles at collider experiments are most effective when the kinematics of collision events are fully understood. As electrons and positrons are fundamental particles, their collisions typically produce events with fewer particles than those at hadron colliders. These types of colliders can also reach very high instantaneous luminosities. The high-precision environment and large datasets offered by the so-called *B-factories* make them a natural choice for performing searches for invisible particles.

3.1 The KEKB accelerator

The KEKB particle accelerator provided the Belle experiment e^+e^- collisions for the duration of its data-taking period, from 1999 to 2010. A schematic of the KEKB collider is shown in Fig. 3.1. Electrons were produced by a photocathode radio frequency gun and accelerated through linear accelerator stages to energies of about 3.7 GeV. Some of these were diverted onto a tungsten target to produce positrons [127] at the linear injector stage. The electrons were further accelerated in the linear accelerator (LINAC) to 8 GeV for the high-energy ring (HER), while the positrons stayed at 3.5 GeV for the low-energy ring (LER). The collider was tuned to a centre-of-mass energy of the $\Upsilon(4S)$ resonance (10.58 GeV), which decays to a pair of B mesons more than 96% of the time [7]. KEKB's capacity to produce approximately 1 million $\Upsilon(4S)$ mesons per day for the Belle experiment, along with PEP-II's comparable output for the BaBar experiment,

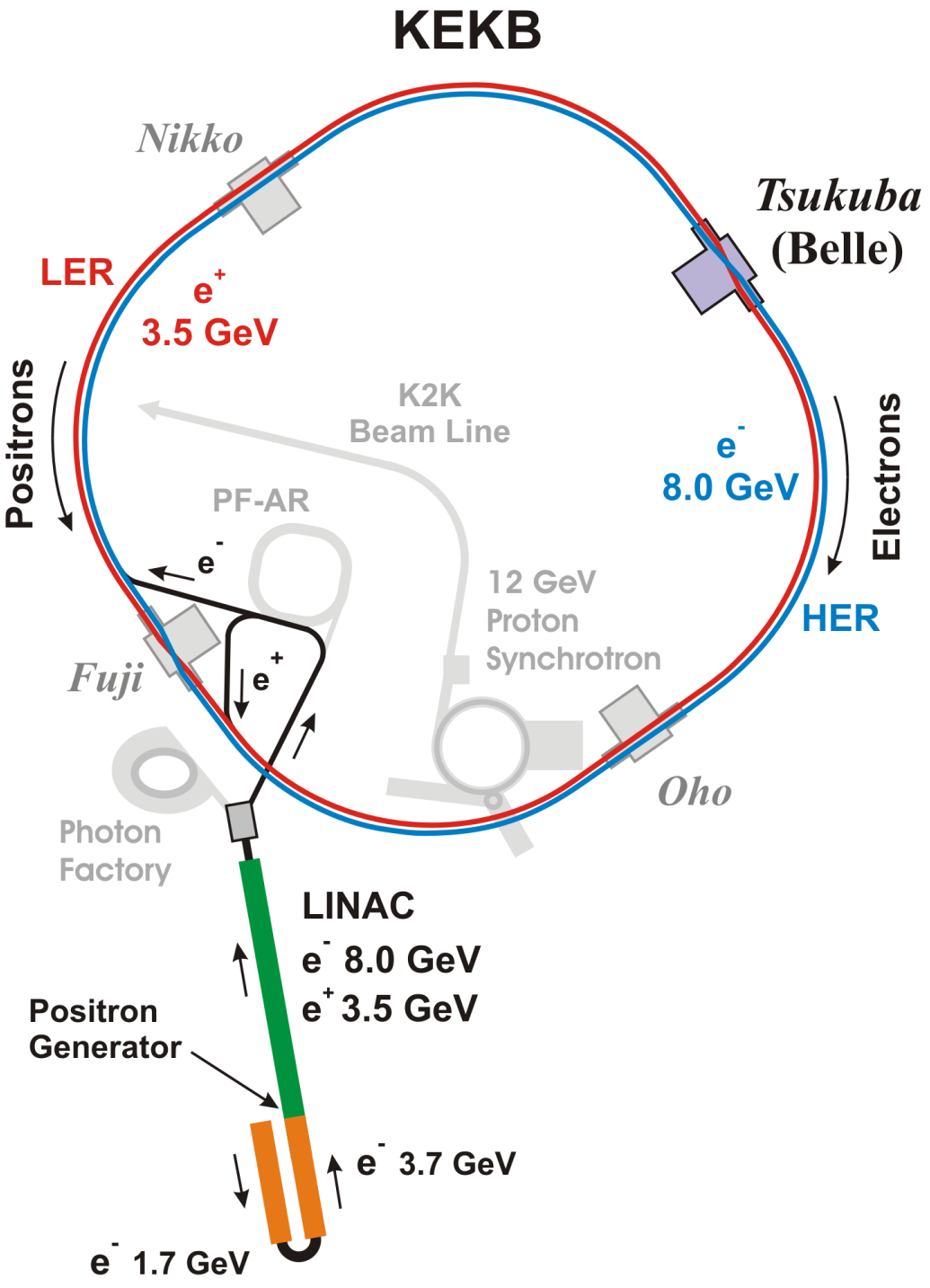


Figure 3.1: A diagram of the KEKB collider from Ref. [126].

earned these two facilities the nickname B-factories. The electron and positron beams had different energies so that the B mesons produced were boosted, making world-leading time-dependent CP-violation measurements possible. KEKB delivered a world-record instantaneous luminosity of 2.11×10^{34} cm^{-2}s^{-1} on June 7, 2009. Over its operational period, it also delivered a record time-integrated luminosity of 1041 fb^{-1}.

Of this total time-integrated luminosity, 711 fb^{-1} were collected at the $\Upsilon(4S)$ resonance, with smaller datasets taken at other energy levels, such as the 121 fb^{-1} taken at the $\Upsilon(5S)$ resonance. KEKB's instantaneous luminosity record has since been broken by its successor, SuperKEKB, which achieved its record luminosity 5.1×10^{34} $\text{cm}^{-2}\text{s}^{-1}$ in late 2024 [128].

As shown in Table 3.1, the e^+e^- cross section at the $\Upsilon(4S)$ is dominated by $e^+e^- \to q\overline{q}$ (where $q \in u,\ d,\ s,\ c$), or *continuum* events. Therefore, it is important for the experiment to understand these events well, as they are a substantial background for many analyses. To study continuum events, a data sample was taken with centre-of-mass energy 60 MeV below the $\Upsilon(4S)$ resonance, known as the *off-resonance* dataset.[1] The dataset collected by the Belle experiment corresponds to 772×10^6 $\Upsilon(4S)$ events. The $\Upsilon(4S)$ resonance decays to neutral and charged B meson pairs at slightly different rates. These rates are often written as $f_{00} = \mathcal{B}(\Upsilon(4S) \to B^0\overline{B}^0)$ and $f_{+-} = \mathcal{B}(\Upsilon(4S) \to B^+B^-)$. In the past, measurements of the ratio $R^{+/0} \equiv f_{+-}/f_{00}$ were consistent with 1, but made assumptions about isospin symmetry between the two decay channels. In 2005, the first direct measurement of f_{00} was performed [129]. This was followed by a number of subsequent measurements, which are averaged by the Heavy Flavour Averaging Group (HFLAV). They report values for these fractions as $f_{00} = 0.4861^{+0.0074}_{-0.0080}$ and $f_{+-} = 0.5113^{+0.0073}_{-0.0108}$ [130].

Table 3.1: A breakdown of the e^+e^- cross section at the $\Upsilon(4S)$ resonance.

Process	cross section [nb]
$e^+e^- \to \Upsilon(4S)$	1.1
$e^+e^- \to u\overline{u}$	0.7
$e^+e^- \to d\overline{d}$	0.7
$e^+e^- \to s\overline{s}$	0.7
$e^+e^- \to c\overline{c}$	1.3
$e^+e^- \to \tau^+\tau^-$	0.92

3.2 The Belle experiment

The Belle technical design report [131] gives a detailed description of the Belle detector and all of its systems. Belle uses a cylindrical coordinate system where the positive z-axis is aligned with the electron beam direction at the interaction point (IP). A polar angle θ is measured away from the positive z-axis, and a radial coordinate r is defined pointing outward from the z-axis. There is also an azimuthal angle ϕ, but as the detector was

[1] More specifically, these processes make up the quark continuum, as there are other processes such as $e^+e^- \to \tau^+\tau^-$ that make up the continuum. In this work, however, only the quark continuum is a relevant background, and hence we simply refer to it as *continuum* for the rest of this work.

more or less cylindrically symmetric, this angle rarely played a role in physics analyses. The Belle detector was split into three main regions: the barrel region, normal to the r coordinate, and the forward and backward endcap regions, normal to the positive and negative z-axes. The arrangement of subdetectors was designed to minimise the radiation length of material that particles had to traverse before reaching outer detectors. The subdetectors from the inside to the outside of the detector were: the silicon vertex detector, the central drift chamber, the time-of-flight counter (in the barrel region), the aerogel Cherenkov counter (in the barrel and forward regions), the electromagnetic calorimeter, and the K_L^0-muon system. The ordering and a sense of relative scale between each subdetector and the whole Belle detector system can be seen in the isometric cutaway diagram shown in Fig. 3.2. In the following sections, the key technologies and roles of each of Belle's subdetectors are discussed in detail.

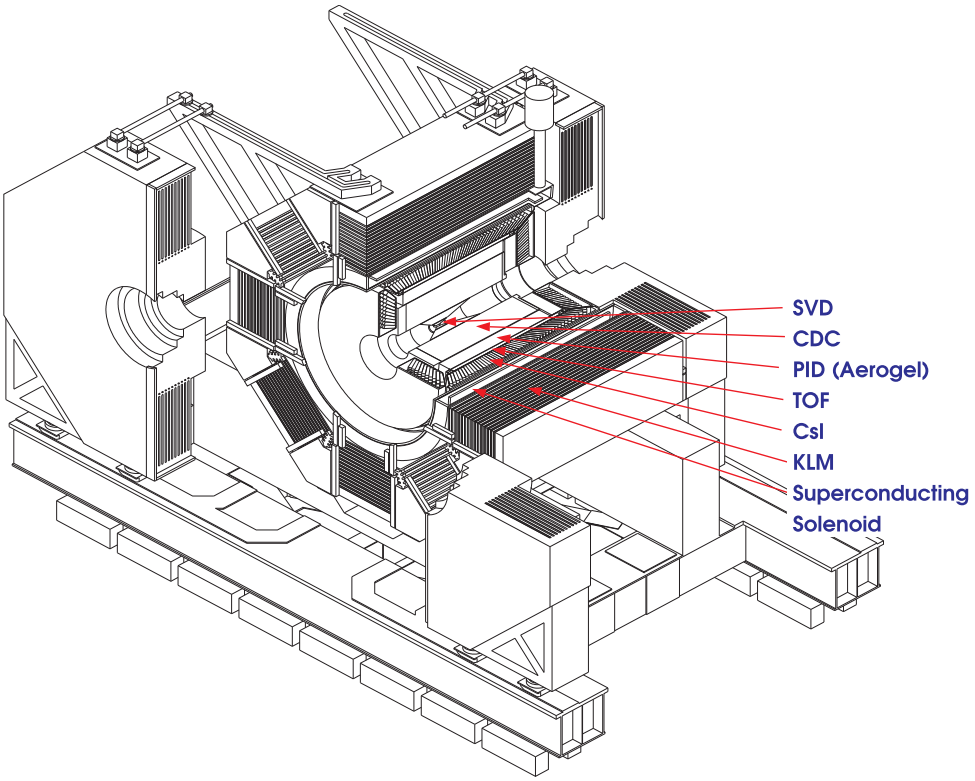


Figure 3.2: An isometric cutaway of the Belle detector taken from Ref. [132].

3.2.1 Silicon vertex detector

The Silicon Vertex Detector (SVD) was the closest subdetector to the interaction region and precisely detected the trajectories of charged particles to infer the position at which they were produced, or their *vertex*. During its lifetime, Belle used two SVDs made of double-sided silicon strip detectors (DSSDs) that were originally designed for the DELPHI micro-vertex detector [133]. Each DSSD had 1280 sense strips with 640 readout pads. The first, SVD1, consisted of three layers of silicon in the barrel region at $r = 30$,

45.5, and 60.5 mm. In total, SVD1 used 102 DSSDs. The layout of this SVD is shown in Fig. 3.3. The DSSDs were placed in a ladder structure, with each ladder containing a boron nitride support (BN) in between carbon fibre reinforced plastic (CFRP) to connect two sub-ladders, containing the DSSDs. Readout was done by VA1 integrated circuit chips, which were designed specifically for the readout of silicon vertex detectors [134, 135]. The ladders were arranged to overlap in the $r - \phi$ plane to aid the internal alignment among the sensors. The second, SVD2, consisted of four layers, the innermost of which was closer to the interaction point than SVD1's first layer [131]. The majority of the Belle dataset was recorded with SVD2. The layers were located at $r = 20$, 44, 70 and 88 mm and covered the full angular acceptance of Belle, $17° < \theta < 150°$. The intrinsic resolutions of SVD2 were 12 μm in the $r-\phi$ plane and 19 μm in the z direction [136].

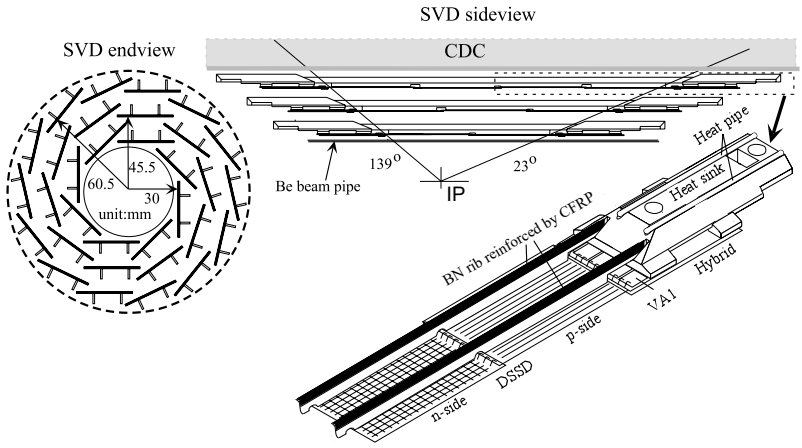


Figure 3.3: A schematic of the Belle SVD.

3.2.2 Central drift chamber

The Belle detector operated within a 1.5 T magnetic field, causing charged particles to follow curved trajectories as they passed through the detector. Since the momentum of a charged particle in a magnetic field is uniquely determined by the field strength and the radius of curvature of its path, a dedicated subdetector was essential for measuring these curves. This role was fulfilled by the Central Drift Chamber (CDC). It consisted of 50 cylindrical layers of wires immersed in a gaseous mixture of helium and ethane. There were two types of wire: field wires that produced an accelerating electric field, and sense wires that collected released charge. When a charged particle passed through the CDC volume, it ionised the gas, producing electrons. The electrons then drifted due to the electric field and deposited their charge in a sense wire, generating an electrical signal that could be read out. The polar angle coverage of the CDC was the same as SVD2, and it covered radii from 80 mm to 880 mm. Series of hits in the CDC were reconstructed in software as *tracks*.

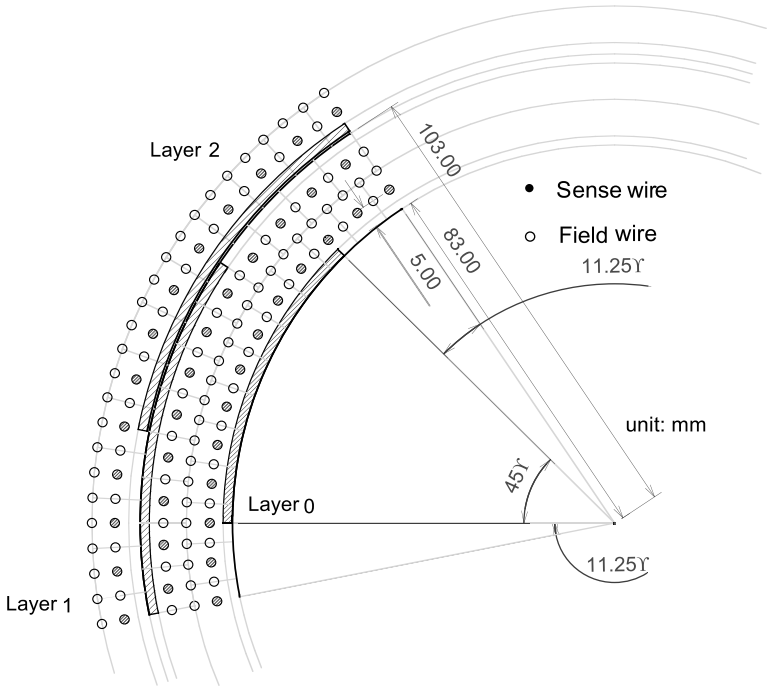


Figure 3.4: A schematic of a Belle CDC cell taken from Ref. [131].

The CDC also provided particle identification power for charged particles, as the energy loss of a particle per unit distance dE/dx differs depending on the type of charged particle that left the track, as shown in Fig. 3.6(a). This was particularly useful in identifying particles when their momentum was too low to reach the dedicated particle identification subdetectors. For tracks, the spatial resolution of the CDC was 130 μm. The dE/dx resolution for minimum-ionising particles such as $\mu^{\pm}$ was 7%.

3.2.3 Particle identification detectors

The subdetectors responsible for identifying hadron species at Belle were the Aerogel Cherenkov Counter (ACC) and the Time-of-Flight (TOF) system. The ACC consisted of silica aerogel tiles that were oriented towards the IP. The refractive index of the ACC tiles ranged from 1.01 to 1.03, and was chosen to maximise separation power between pions and kaons with a momentum greater than 1.2 GeV/c. The range of refractive indices was chosen to account for the hadron momentum spectrum's softening with polar angle. As the charged particles passed into the tiles, they entered a medium with a refractive index large enough for the particles to be faster than the speed of light in that medium. This resulted in the emission of Cherenkov radiation, which was detected by photomultiplier tubes (PMTs). There is a different photoelectron distribution for each refractive index in the ACC. The likelihood that a signal comes from a particle species is determined from these photoelectron distributions. The Cherenkov threshold for pions in a refractive index of $n = 1.03$ is about 0.5 GeV/c, and so the ACC did not

provide any information for pions slower than that. The thresholds are even higher for heavier hadrons, such as kaons and protons.

The TOF detector measured the time taken for particles to traverse the subdetector, achieving a temporal resolution of 100 ps. It was made up of 64 plastic scintillator modules coupled to a PMT at either end. This detector takes advantage of the fact that, for a given momentum, heavier particles travel slower than lighter particles, as shown in Fig. 3.6(b). At Belle, this detection technique was most effective at lower momenta, below 1 GeV/c. At higher momenta, the velocity differences between particles are smaller, making it hard to differentiate between them due to the finite timing resolution. However, combined with the ACC, Belle was able to have very good particle identification (PID) for a wide range of hadron momenta.

3.2.4 Electromagnetic calorimeter

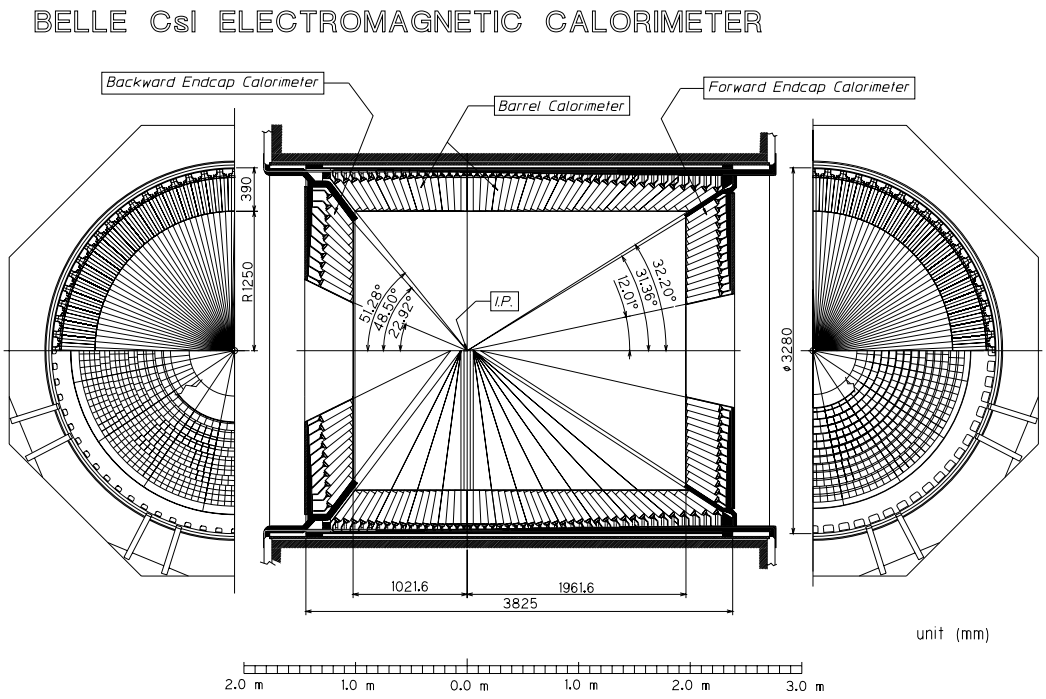


Figure 3.5: A schematic of the Belle electromagnetic calorimeter taken from Ref. [131].

The Electromagnetic Calorimeter (ECL) consisted of 8736 thallium-doped caesium iodide (CsI(Tl)) crystals and covered both the barrel and endcap regions of the detector [131]. The inner radius of the ECL was 1.25 m, and the forward and backward endcaps were located at $z = 2.0$ m and $z = -1.0$ m respectively. Each CsI(Tl) crystal was about 30 cm long, which corresponds to about 16 radiation lengths for photons and electrons, and was oriented in such a way that its long axis pointed to the IP. Photodiodes on the outer end of each crystal read out the scintillation light produced by particles hitting

the crystal, which was then used to infer the energy deposited in the crystal. The energy resolution of the ECL was dependent on the energy deposited; it ranged from 4% at 100 MeV down to 1.6% at 8 GeV. The ECL covered polar angles from 12.01–31.36 degrees in the forward endcap region, 32.01–128.72 degrees in the barrel region, and 131.50–157.08 degrees in the backward endcap region. The gaps between the three regions were necessary to accommodate support structures for the inner detectors. These gaps cause a signal loss of 3%. Despite their small size, it is important to consider the gaps between the barrel and endcap regions, as photons passing through the gaps would be missed and contribute to the missing energy of an event.

3.2.5 K_L^0 and muon detector

The K_L^0 and muon detector (KLM) was the outermost subdetector at Belle. It consisted of layers of iron plates interspersed with glass electrode resistive plate chambers (RPCs). The iron layers provided a dense material (3.9 radiation lengths) for normally incident particles. This large amount of ferrous material also served as a magnetic return yoke for the superconducting solenoid that provided Belle with its 1.5 T magnetic field [131, 137]. The penetrating power and deflection of a particle as it travelled through the KLM could be used to distinguish muons from other particles. For muons, the angular resolution of the KLM was 30 mrad in both the polar and azimuthal directions, and the momentum resolution was 0.6 GeV/c. Hits in the KLM that were unassociated with a track, or that did not undergo any transverse deflection, could be identified as a K_L^0.

3.2.6 Trigger and data acquisition

Without dedicated triggers to select events for storage, the collider could have overwhelmed the Belle data acquisition system, which had a maximum event rate of 0.5 kHz. A hardware trigger used subdetectors and global decision logic to flag hadronic events ($e^+e^- \to \Upsilon(4S)(\to B\overline{B})/q\overline{q}$), Bhabha and dimuon events ($e^+e^- \to e^+e^-/\mu^+\mu^-$), and others. A software trigger then processed the hadronic-flagged events to further reduce the event rate. It checked whether the tracks in the event were of high quality, and if any further criteria could be applied to the event to make it more useful for specific analyses.

3.3 Analysis software

In any modern physics analysis, the software plays a vital role. In the following sections, some key pieces of software used in this work are described.

3.3.1 Monte Carlo simulations

The invisible particle search presented in this thesis proceeds as a blind analysis. This means that the entire search procedure is developed and fine-tuned on simulated data based on our understanding of the physics processes occurring in the experiment. This precludes the introduction of biases in the analysis technique that arise from the real data recorded by the experiment. The simulated data are generated with a Monte Carlo (MC) method. For the remainder of this thesis, *MC* will be used to refer to this simulated dataset, while *data* will be used only to refer to real collision events recorded by Belle.

The PDG currently lists 639 possible decay channels for the B^+ meson, and many of these decays produce particles which themselves have hundreds of possible decay channels. The software packages `EvtGen` [138] and `PYTHIA` [139] simulate some of these decay chains. In order to simulate particles not in the SM, `EvtGen` allows the user to introduce new particles and specify their mass, spin, charge, lifetime, and decay width. Final state electromagnetic radiation is simulated using `PHOTOS` [140]. Once these packages generate the full decay chains, including the four-momenta of all particles present, the event's interaction with the Belle detector is simulated with `GEANT3` [141]. A drawback of simulating the detector interaction in this way is that simulating custom particles' interactions with the detector is non-trivial; however, since this analysis only requires invisible particles beyond the SM, we circumvent this shortcoming. The detector simulation step is far more computationally expensive than the initial event generation, and so for some preliminary studies that were performed at the start of this work, the detector simulation step was skipped. This helped refine the range of masses and lifetimes for the particles generated before producing the full suite of MC data with detector simulation included.

3.3.2 The `b2bii` package and Belle II software

The analysis described in this thesis was conducted entirely on Belle MC and data but uses software tools developed for Belle's successor, Belle II. This allows the analysis to take advantage of software techniques developed for the new experiment, using them on old data. Despite the similarities between Belle and Belle II, including using the same experimental hall, ECL crystals, and KLM components, the format with which events were saved at Belle differs from the method used at Belle II. The `b2bii` package [142] was developed to convert Belle events to the format used by Belle II so that they can be used within the Belle II Analysis Software Framework (`basf2`) [143]. This conversion process allows modern algorithms like the Full Event Interpretation (FEI) algorithm

(which will be discussed in more detail later in this work) to be used on a dataset that is over a decade old.

3.3.3 Particle identification

For the analyst in any modern physics experiment, selecting different particle species is done in software. At Belle, charged particle candidates have likelihoods assigned to them for five particle species: e, μ, π, K, and p. These likelihoods are defined as the products of individual likelihoods specific to each subdetector. Each subdetector's own likelihood is calculated differently. For instance, the likelihood from the CDC for the i^{th} particle species is defined as

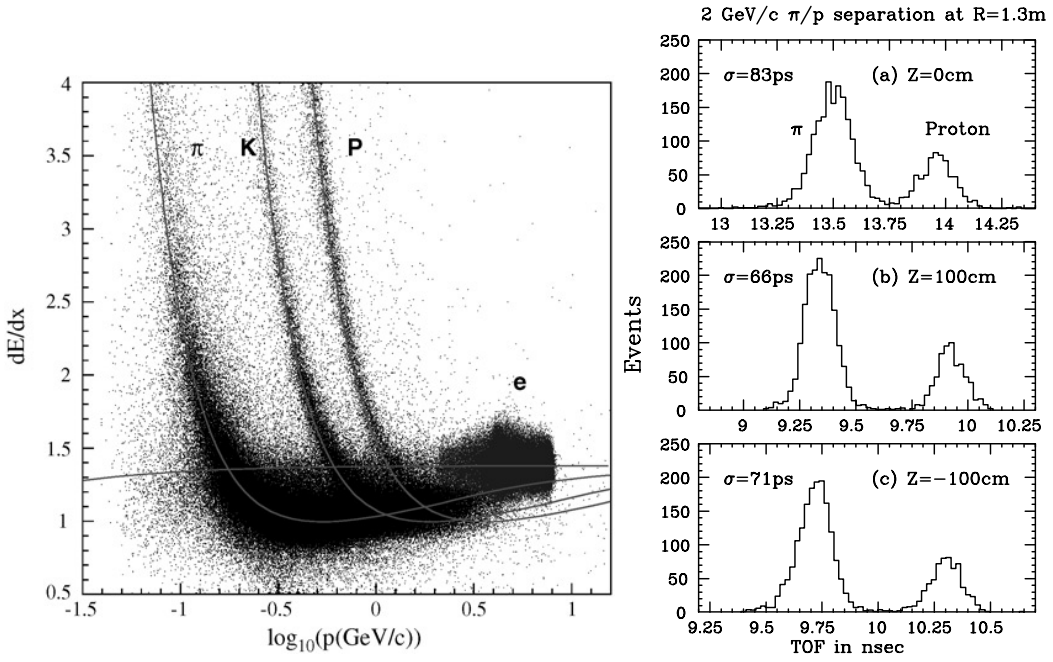
$$P_i^{\text{CDC}} = \frac{e^{-\chi_i^2/2}}{\sqrt{2\pi}\sigma_{\mathrm{d}E/\mathrm{d}x}} \tag{3.1}$$

where χ_i^2 and $\sigma_{\mathrm{d}E/\mathrm{d}x}$ are the squared residual and resolution of $\mathrm{d}E/\mathrm{d}x$, respectively.

Different particle species use information from different subdetectors. For example, pions and kaons are identified with a combination of likelihoods from the CDC, TOF, and ACC, while muons are identified with information from the KLM [144]. At Belle, analysts can select particles using a global likelihood, defined as the ratio of the likelihood for particle species i to the sum of likelihoods for all species, or a binary likelihood, which is the ratio of the likelihood for species i to the sum of the likelihoods for species i and j. In this thesis, only binary likelihoods are used, and are denoted as P_{ij}.

3.3.4 Full event interpretation

Two great advantages of the B-factories over hadron collider experiments are the clean initial states and the hermeticity of the detectors. The e^+e^- initial state means that B mesons must be produced in pairs with a well-defined initial state, and the almost 4π solid angle coverage of the detector means that all detectable particles leave a signal in the detector. Because of these advantages, it is often useful to use the non-signal B meson (i.e. the B meson that is not decaying via the process that is being measured in a given study) to constrain the signal B meson. In general, there are two analysis methods one can use to provide this constraint: *untagged* and *tagged*. In the former, the kinematic information of all particle candidates not used to reconstruct the signal B meson is naïvely summed. This method is also known as *inclusive* tagging. In the latter, the non-signal side B meson is explicitly reconstructed, which yields much more precise information and therefore stronger constraints on the signal side, at the cost of much poorer efficiency. Tagging methods are generally categorised by the type of decays used to explicitly reconstruct the B meson. Semileptonic tags are the most efficient of these methods, with reconstruction efficiencies of $\mathcal{O}(1\%)$ because of the high $\overline{B} \to X_c \ell \overline{\nu}$



(a) CDC dE/dx distribution (b) TOF π/proton separation

Figure 3.6: Examples of particle species separation in (a) the CDC, taken from Ref. [144], and (b) the TOF detector, taken from Ref. [145].

branching fraction. However, due to the presence of the neutrino in the final state, this method offers the poorest quality of kinematic information of the tagging methods. Hadronic tags, on the other hand, reconstruct purely hadronic decays, resulting in much better kinematic information as there are no invisible hadronic final states, albeit with a lower reconstruction efficiency of around $\mathcal{O}(0.1\%)$. Key differences between these tagging methods are displayed in Table 3.2.

Tag type	Advantage	Reconstruction efficiency
Exclusive hadronic	High purity, precise kinematics	$< 1\%$
Exclusive semileptonic	High purity	$\sim 1\%$
Inclusive	High reconstruction efficiency	$> 10\%$

Table 3.2: A concise comparison of B-factory tagging methods.

The tagging algorithm that is now used by both Belle and Belle II is the Full Event Interpretation (FEI) algorithm [146], the successor to the Full Reconstruction algorithm previously used by Belle [147]. The FEI can reconstruct B mesons from both hadronic and semileptonic decays. As searching for invisible particles relies on building a full picture of the kinematics of the event, the analysis described in this thesis uses the hadronic tag only.

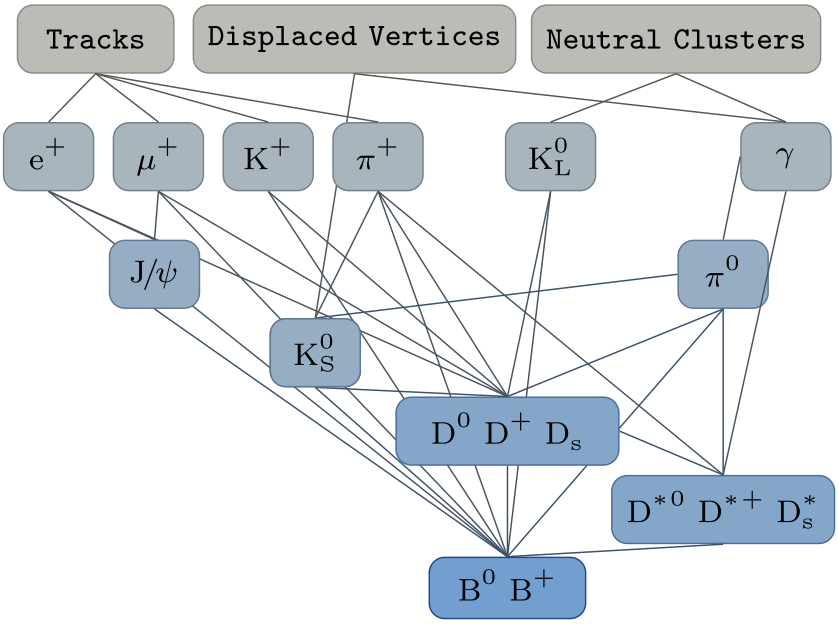


Figure 3.7: A schematic overview of the FEI taken from Ref. [146]. The algorithm first operates on objects reconstructed by the Belle II software (tracks, displaced vertices, and neutral clusters). From these basic objects, final-state particles $e^{\pm}, \mu^{\pm}, K^{\pm}, \pi^{\pm}, K_L^0$ and γ are built. From the final state particles, other composite particles are built, which are further combined into B mesons.

An overview of the FEI algorithm is shown in Fig. 3.7. First, final state particles are created from the detector data (tracks and clusters) reconstructed by the Belle II software.[2] These are combined into composite particles by the FEI, such as $\pi^0, K_S^0, J/\psi$, and Λ^0. Lighter composite particles are then recombined to make heavier composite particles, such as D, D^*, Σ, and Λ_c. In the last step, the particles from any of the previous steps are combined into B mesons. For instance, a B^0 candidate can be formed by combining a D^- and a π^+. Between each stage of the FEI reconstruction, selections are performed to get rid of candidates that are unlikely to be correctly reconstructed. This selection relies on boosted decision trees (BDTs) trained on information such as four-momenta, vertex information, angles between daughters, invariant mass compared to the nominal mass of the particle, particle identification likelihoods, distance of the point of closest approach to the interaction point, and the output of the BDTs for each daughter particle. There are over 100 explicit B meson decay channels that are reconstructed by the FEI, which means $\mathcal{O}(10000)$ unique decay chains are considered.

[2] At Belle, the term *final-state particles* refers collectively to particles that are stable on the length scale of the detector: $e^{\pm}, \mu^{\pm}, K^{\pm}, \pi^{\pm}, K_L^0$, and γ.

4

Analysis overview, event reconstruction, and selection

This chapter begins with a brief overview of the analysis strategy and a description of the real and simulated datasets used in the analysis. It then gives a detailed explanation of how signal-like events are selected out of these very large datasets, and touches on what kind of background events can mimic the signal we are searching for.

4.1 Signal selection outline

The search for invisible particles described in this thesis is a blind analysis [148], meaning that the experimental data was not looked at (unblinded) until the analysis strategy had been more or less decided upon. This approach is taken to avoid any experimental bias introduced by defining selection criteria or creating PDFs based on statistical fluctuations present in the experimental dataset that could lead to a false signal. To avoid the problem of statistical fluctuations biasing these things, the MC dataset used to prepare this analysis consists of between 6 and 50 times the number of events present in the experimental data, depending on the type of process that was simulated.

The event reconstruction is done by fully reconstructing a B meson that decays via a hadronic decay channel, which we will call the B_{tag} hereafter. The other B meson (which we will call B_{sig}) is partially reconstructed, leaving out its invisible daughter, which is of course impossible to detect. These B mesons are combined to form an $\Upsilon(4S)$ candidate.

The choice of final states $h = \pi^{\pm}, K^{\pm}, p, D_s^{\pm}, \overline{D}^0$ was guided by both theoretical motivation and practical considerations. In this thesis, h is used to refer to any of the aforementioned signal-side hadrons. For ALPs coupling to the W boson and dark scalars [61, 74], searching for the decay $B^+ \to K^+ a$ can provide very strong constraints. This is complemented by a search for $B^+ \to \pi^+ a$, which, although more CKM-suppressed, can probe higher ALP masses. These searches, along with the searches for the decays $B^+ \to D_s^+ a$ and $B^0 \to \overline{D}^0 a$, constitute the four channels most sensitive to probing the various ALP-quark coupling scenarios described in Ref. [2]. In the detector, these missing energy events look very similar to the process $B \to p\psi_D$, and so despite being motivated by a different model, it was natural to include the search for the channel containing the proton in this analysis as well.

The selection optimisation proceeds in a step-wise manner, where the procedure is first developed entirely using MC simulations. The selection is then validated on data samples that exclude the signal region. These datasets, referred to as *sidebands*, typically adhere to nearly the same selection criteria as the signal-region data, with only a single modification to ensure the selected sample is devoid of signal events while still retaining background events representative of those in the signal region. In some analyses, multiple selection criteria are adjusted to access sidebands corresponding to specific channels. However, as a general rule, more extensive modifications to the selection criteria result in greater discrepancies between the sideband and the signal region.

In Chapter 2, several models containing invisible particles were discussed. In all of these models, the mass of the invisible particle is a free parameter. The only model with strict bounds on allowed masses is the B-mesogenesis model, which requires dark baryons to be more massive than protons. Given this, the search for invisible particles is carried out as a scan across the entire mass range accessible to the experiment. For a decay of the form $B \to ha$, where h is some SM hadron and a is some invisible particle, the allowed mass of a is between 0 and $m_B - m_h$. As we cannot measure the mass of the invisible particle, we must use our understanding of the event kinematics to constrain it. We precisely know the four-momentum of the beams at Belle, and if the full reconstruction of the B_{tag} was done correctly, then we should know its four-momentum as well. As the B_{sig} is not fully reconstructed, we do not have access to all of its kinematic information; however, we can infer it, as the four-momentum of the beam must be conserved,

$$P_{B_{\text{sig}}} = P_{\text{beam}} - P_{B_{\text{tag}}}. \tag{4.1}$$

We will use P to denote four-momentum and p for three-momentum throughout this thesis. As we are searching for invisible particles produced in two-body decays, we can rely on the fact that in the rest frame of the mother particle, the daughter particles will

be produced back to back, with the same three-momentum magnitude,

$$|p_1| = |p_2| = \frac{[(M^2 - (m_1 + m_2)^2)(M^2 - (m_1 - m_2)^2)]^{1/2}}{2M} \tag{4.2}$$

where M is the mass of the mother particle and m_1 and m_2 are the masses of the daughter particles. Putting this together, we boost the visible daughter of the B_{sig} to the frame where the B_{sig} is at rest. Its momentum magnitude should then be uniquely defined by its mass, the B meson's mass, and the mass of the other particle produced in the decay. Henceforth in this thesis, the momentum of the signal-side hadron h in the B_{sig} rest frame is denoted p_h^S. Events containing $B \to ha$ decays will cause a peaking structure in the p_h^S distribution. The fit method, detailed in Chapter 5, focusses on extracting this peaking shape while correctly modelling the non-peaking structures caused by background processes.

The heavier a is, the smaller the amount of kinetic energy available to the hadron h is. At the kinematic limit of $m_a = m_B - m_h$, the only momentum available for h is the few hundred MeV/c that the mother B meson has in the lab frame. As the Belle detector cannot resolve charged particles very well at lab-frame momenta of less than 200 MeV/c, the highest invisible particle mass accessible to the experiment is slightly less than this kinematic limit.

In the case where the invisible particle is undetected because it decays outside of the detector volume, both the lifetime and the mass of the particle play a role. A lighter particle has a greater boost in the detector, and so will have a larger lab-frame decay length than a more massive particle with the same lifetime.

The signal selection is based on specific thresholds which are applied to a few variables, guided by our understanding of how the signal process should appear in the detector. Additional background suppression is achieved using machine learning methods. The selection criteria for the outputs of these multivariate techniques are optimised using a multi-dimensional figure of merit (FoM).

4.2 Signal events and background types

The five-channel search presented in this thesis is based on events with pairs of B mesons, where one decays via a SM channel and the other decays to a two-body final state with an invisible daughter. The final state in the detector then looks like that of a single B meson with an extra track or set of tracks from the other B meson's SM decay product. Anything that can mimic this type of event when we reconstruct events is classed as *background*. There are some general classes of background that will be discussed here briefly to contextualise the event selection discussion later in this work.

The full treatment of these backgrounds is also described later in the work. The main classes of background present in this analysis are:

- Continuum events: $e^+e^- \to q\bar{q}$ where $q \in \{u, d, s, c\}$. A fraction of the time, the tagging algorithm will find a suitable B candidate using final state particles from these events. As the production cross section of these events is high, they are a substantial portion of the background in this search.
- Three- (or more-) body B decays: These events are $B\bar{B}$ events with one correctly tagged B and a signal-side B that is incorrectly reconstructed. These kinds of backgrounds are responsible for the largest fraction of the background, and, along with the continuum background, dictate the broad shape of the background distribution of p_h^S, but do not contain signal-like structures.
- Peaking backgrounds: These events are $B\bar{B}$ events with one correctly tagged B and a signal-side B that has one correctly reconstructed final state particle, and a second final-state particle that goes undetected. This is a particularly important background type to understand. Because the signal extraction relies on two-body kinematics, this class of background will produce a signal-like peak in the p_h^S distribution.

4.3 The datasets

The datasets used in this work are either simulated MC datasets, or real data recorded by the experiment. Below is a brief description of the samples used, as well as a description of how they were made or recorded.

4.3.1 Monte Carlo data

Particle physics experiments have relied on Monte Carlo (MC) techniques to simulate decay processes for decades [149, 150]. Particle decay simulation relies on computing multidimensional phase space integrals. The advantage of MC techniques over other sampling approaches is that they scale well with a high-dimensional space. There is a wealth of software that is used to generate events. The MC datasets used in this work are generated with software described in Section 3.3.1.

4.3.1.1 Signal samples

Signal Monte Carlo was generated using the Belle Analysis Software Framework (`basf`) [151] and the `EvtGen` generator [138]. The full Belle detector's response is simulated

in `GEANT3` [141]. For each of the B decays searched for in this work ($B \to ha, h \in \pi^{\pm}, K^{\pm}, p, D_s^{\pm}, \overline{D}^0$), 1 million events are generated at a number of a mass hypotheses. The mass of the new particle is kinematically limited by the difference between the B mass and the mass of the other signal-side hadron, as shown in Table 4.1. The lowest mass generated for all channels is 1 MeV/c^2. The highest mass generated depends on the channel, but is always just below the limit that is kinematically allowed. Half of the events are generated with $B^+(B^0)$ decaying via the signal channel, while the other half feature $B^-(\overline{B}^0)$ decaying via the charge-conjugated signal channel. The non-signal-side B meson is allowed to decay generically according to the default Belle scheme. Signal MC for the channel $B^{\pm} \to D_s^+ a$ only has the decay $D_s^+ \to K^+K^-\pi^+$, since it is the only channel used to reconstruct $D_s^{\pm}$. The intermediate decay chains generated to reach this final state are summarised in Table 4.2. Similarly, in the $B^0 \to \overline{D}^0 a$ sample, only three $\overline{D}^0$ final states are generated, according to the three channels it is reconstructed in, as described in Table 4.3.

Many of the models that inspired this work can produce not only new invisible particles, but also particles that decay to visible final states after some unknown lifetime. Whether the final states are invisible or visible generally depends on the value of the free parameter(s) in the model, of which we have no prior knowledge. To assess the sensitivity of the nominal search for invisible particles to particles that decay visibly, we generate signal MC samples where a decays to several visible final states, with several lifetimes, in addition to the samples with invisibly decaying a. These additional samples are not used for the nominal analysis but are later employed in this work to map the sensitivity of the analysis to NP scenarios involving long-lived particles.

Channel	$m_B - m_h$ [MeV/c^2]
$B^{\pm} \to \pi^{\pm} a$	5139
$B^{\pm} \to K^{\pm} a$	4785
$B^{\pm} \to p^{\pm} a$	4341
$B^{\pm} \to D_s^{\pm} a$	3311
$B^0 \to \overline{D}^0 a$	3414

Table 4.1: Mass difference between the B meson and the other signal-side hadron.

A full list of the generated a masses for $B \to ha$ is shown in Table 4.4.

4.3.1.2 Background samples

Several centrally produced MC samples for Belle are used in this analysis, containing $B\overline{B}$ and $q\overline{q}$ decays. The *charged* dataset contains $\Upsilon(4S) \to B^+B^-$ events, while the *mixed* dataset contains $\Upsilon(4S) \to B^0\overline{B}^0$ events in which $B^0 - \overline{B}^0$ mixing can occur.

Table 4.2: Decay modes generated for the D_s^+ decay, along with assigned decay ID and the nominal branching fractions and their uncertainties as reported by the Particle Data Group (PDG) [7].

ID	Decay mode	Branching fraction
1	$D_s^+ \to K^+K^-\pi^+$	$0.95 \pm 0.03\%$
2	$D_s^+ \to \phi(\to K^+K^-)\pi^+$	$2.25 \pm 0.05\%$
3	$D_s^+ \to K^+\overline{K}^{*0}(\to K^-\pi^+)$	$2.61 \pm 0.05\%$
4	$D_s^+ \to K^+\overline{K}^*(1410)^0(\to K^-\pi^+)$	$0.179 \pm 0.026\%$
5	$D_s^+ \to f_0(980)(\to K^+K^-)\pi^+$	$1.12 \pm 0.19\%$
6	$D_s^+ \to f_0(1370)(\to K^+K^-)\pi^+$	$0.072 \pm 0.030\%$
7	$D_s^+ \to f_0(1710)(\to K^+K^-)\pi^+$	$0.068 \pm 0.028\%$
	Total	$7.25 \pm 0.21\%$

Table 4.3: Decay modes generated for the $\overline{D}^0$ decay, along with assigned decay ID and the nominal branching fractions and their uncertainties as reported by the Particle Data Group (PDG) [7].

ID	Decay mode	Branching fraction
1	$\overline{D}^0 \to K^-\pi^+$	$3.95 \pm 0.03\%$
2	$\overline{D}^0 \to K^-\pi^+\pi^0$	$14.40 \pm 0.60\%$
3	$\overline{D}^0 \to K^-\pi^+\pi^-\pi^+$	$8.22 \pm 0.14\%$
	Total	$26.57 \pm 0.62\%$

The charged and mixed on-resonance B samples correspond to a dataset ten times larger than the total Belle dataset, while the on-resonance $q\overline{q}$ (continuum) samples correspond to a dataset six times larger. *On-resonance* refers to events occurring at the $\Upsilon(4S)$ resonance. Additional samples, known as the *charged rare* and *mixed rare* samples, correspond to 50 times the total Belle dataset. They include processes with relatively small branching fractions compared to those in the main samples, and are entirely independent of the main charged and mixed samples. These rare decays are particularly important to examine closely, as they include decays that most closely resemble the signal channel. There are some known issues with the Belle MC. For example, many of the branching fractions used in MC production have been remeasured and updated since the samples were generated. Indeed, some of the branching fraction values for decays in the rare samples are based on educated estimates rather than precise measurements. Furthermore, the same cross sections are used to produce continuum events for all collision energy values. In reality, the cross section should scale as $1/E^2$, meaning the total amount of continuum MC for both on- and off-resonance samples is not entirely accurate [152]. The background samples are used to optimise selection criteria and extract efficiencies and uncertainties, and to define the PDF models used to fit the background. By generating MC samples that are multiple times larger than the total

Table 4.4: Mass hypotheses used in signal MC generation. Mass hypotheses were chosen to ensure even spacing when transformed into the recoiling momentum of the signal-side hadron in the B_{sig} frame, p_h^S. This made it easier to parameterise fit parameters with respect to p_h^S. Additional reference points were chosen at round numbers such as 1 and 100 MeV/c^2.

Channel	Mass hypotheses [MeV/c^2]
$B^\pm \to \pi^\pm a$	1, 10, 50, 100, 284, 300, 491, 500, 750, 1000, 1160, 1565, 1885, 2158, 2401, 2620, 2823, 3190, 3358, 3519, 3672, 3819, 3960, 4096, 4228, 4355, 4478, 4598, 4713, 4824, 4930, 5027, 5107
$B^\pm \to K^\pm a$	1, 10, 50, 100, 128, 300, 500, 700, 750, 1000, 1050, 1479, 1808, 2086, 2330, 2550, 2753, 2940, 3116, 3281, 3438, 3587, 3728, 3863, 3991, 4113, 4229, 4338, 4533, 4616, 4686, 4740, 4775
$B^\pm \to p^\pm a$	1, 10, 50, 100, 250, 300, 500, 750, 1000, 1035, 1441, 1753, 2016, 2247, 2454, 2643, 2818, 2980, 3132, 3274, 3407, 3532, 3649, 3759, 3860, 3954, 4039, 4115, 4181, 4237,4282, 4315, 4335
$B^\pm \to D_s^\pm a$	1, 10, 50, 100, 300, 425, 500, 750, 942, 1000, 1257, 1503, 1709, 1889, 2049, 2193, 2324, 2444, 2553, 2654, 2746, 2831, 2908, 2978, 3041, 3097, 3146, 3189, 3226, 3256, 3279, 3296
$B^0 \to \overline{D}^0 a$	1 , 10, 50, 100, 291, 300, 500, 750, 911, 1000, 1250, 1511, 1728, 1917, 2085, 2236, 2373, 2499, 2614, 2720, 2817, 2906, 2987, 3061, 3128, 3187, 3240, 3286, 3324, 3356, 3381, 3399, 3410

Table 4.5: A summary of the MC samples used in this analysis that were generated within the Belle collaboration. The luminosity is given in multiples of the luminosity of the Belle dataset.

Event type	MC name	No. events	$\mathscr{L}_{\text{int}}$ [fb^{-1}]
$e^+e^- \to \Upsilon(4S) \to B^+B^-$	charged	3.76×10^9	10×711
$e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$	mixed	3.76×10^9	10×711
$e^+e^- \to \Upsilon(4S) \to B^+B^-$	charged rare	1.56×10^8	50×711
$e^+e^- \to \Upsilon(4S) \to B^0\overline{B}^0$	mixed rare	1.17×10^8	50×711
$e^+e^- \to u\overline{u}, d\overline{d}, s\overline{s}$	uds	8.66×10^9	6×711
$e^+e^- \to c\overline{c}$	charm	5.39×10^9	6×711

Belle dataset and then scaling down the resulting PDFs, we reduce relative statistical uncertainties during the development of the analysis method. This is important for this analysis as achieving strong background rejection is crucial to the experimental sensitivity. A summary of the background samples used in the analysis is provided in Table 4.5.

4.3.2 The Belle dataset

The Belle dataset consists of 711 fb^{-1} of e^+e^- collisions recorded by the Belle detector between 1999 and 2010 at the $\Upsilon(4S)$ resonance. The cross section $\sigma(e^+e^- \to B\overline{B}) =$ 1090 pb, while $\sigma(e^+e^- \to q\overline{q}, q \in u, d, s) = 2090$ pb and $\sigma(e^+e^- \to c\overline{c}) = 1300$ pb. Thus, for every B event, there are approximately three continuum events. To better study this common background, an additional dataset was recorded at a centre-of-mass energy just below the $\Upsilon(4S)$ resonance, where only continuum events are produced. This dataset is known as the *off-resonance* dataset. Data taking periods were grouped into *experiments*, which could last from several days (especially in the early data-taking periods) to as long as a few months. During a single experiment, detector conditions remain relatively stable. Experiments are further divided into *runs*, where one run is a single period of uninterrupted data-taking from when the beam is turned on to when it is turned off (or is lost due to some issue occurring). The most significant subdetector change during data-taking was driven by the low radiation tolerance of the original SVD (SVD1), which was quickly replaced with an upgraded model (SVD2) [153]. The Belle dataset is often categorised into two periods: before and after this replacement, referred to as SVD1 and SVD2 data, respectively. The vertexing performance was very important for CP-violation analyses that were the flagship physics goals of the experiment, and the different performance of the two SVDs is a key consideration in such analyses. However, for the new particle search presented in this thesis, vertexing performance is less critical, and no special treatment beyond the standard Belle corrections is applied. All MC samples described in Section 4.3.1 use experiment-dependent detector conditions so that the simulated events reflect the changes in the detector conditions over time.

4.4 Event reconstruction

To reconstruct events containing the signal process, the signal-side hadron is reconstructed first. If this hadron is a $\pi^\pm, K^\pm$, or p, then this simply means finding a track left by a charged particle whose particle ID matches the desired species. For $D_s^\pm$ and $\overline{D}^0$, this means combining the 4-momenta of final state particles identified as $\pi^\pm, K^\pm$, or γ (which are combined to form π^0 candidates) and keeping the candidates that have invariant mass consistent with the $D_s^\pm$ or $\overline{D}^0$. Since the invisible particle cannot be reconstructed, accurately measuring its momentum requires the reconstruction of all other particles in the event. The non-signal-side B meson (B_{tag}) is reconstructed from hadronic decays using the previously described FEI algorithm.

4.4.1 Rest of event

When partial signal-side reconstruction occurs, we save the information for the Rest of Event (ROE). The ROE contains the charged particle tracks and energy deposit clusters in the ECL that have not been used in the reconstruction of either B candidate.

In `basf2`, the ROE is built by inputting the partially reconstructed $\Upsilon(4S)$ candidate (reconstructed from the B_{tag} and partially reconstructed B_{sig}). Objects must satisfy some selection criteria to be included in the ROE. In this analysis, we require that the ECL cluster energies are greater than 50, 100 and 150 MeV in the barrel, forward, and backward regions respectively. These are standard requirements for ROEs in missing energy analyses at Belle. They are implemented to suppress beam background and low-energy hadronic split-off. We also require that the track impact parameters dr and $|dz|$ are less than 2 and 5 cm respectively. This ensures that the tracks in the ROE originate from near the IP.

4.4.2 Preselection criteria

Each event can be reconstructed in a number of ways due to the combinatorics of having so many final state particles. Therefore, some selection criteria are applied at the reconstruction level to reduce the output file sizes to something manageable. There are a number of useful variables used at this stage to select correctly reconstructed events, which are listed below.

- Beam-constrained mass $M_{\text{bc}} = \sqrt{s/4 - |p_B^*|^2}$ where $\sqrt{s}$ is the Mandelstam variable that denotes the centre-of-mass (CMS) energy of the collision [154] and p_B^* is the reconstructed B-meson candidate momentum in the CMS frame. This variable powerfully discriminates between true B meson events and continuum backgrounds. The M_{bc} distribution for B events peaks at m_B, while continuum backgrounds follow an ARGUS distribution [155].
- Energy difference $\Delta E = E_B^* - \sqrt{s}/2$ where E_B^* is the CMS frame energy of the reconstructed B meson candidate and s is the same Mandelstam variable as above. This also powerfully separates correctly reconstructed B mesons from continuum background, as well as from candidates that are misreconstructed due to the assignment of an incorrect mass hypothesis to a final-state particle. The ΔE distribution peaks around zero for correctly reconstructed B meson candidates, while true B meson candidates with a misidentified final-state particle also peak, but offset from zero. Combinatorial backgrounds have a linearly falling ΔE distribution.
- dr and dz, the so-called *impact parameters*, are the distances between the IP and the point of closest approach of a track in the r and z directions respectively.

Restricting these to a few cm ensures that selected tracks originate from near the IP.

- FEI signal probability is the output from the FEI B classifier. Higher values indicate a greater chance of the B meson being correctly tagged.
- E_{ECL} is the sum of energy from ECL clusters that pass the ROE selection criteria.
- Number of ROE tracks refers to the number of tracks that pass the ROE selection criteria.
- p^* and p^{lab} refer to a particle's momentum in the CMS and lab frame respectively. The particle whose momentum is being referred to is denoted by a subscript. For example, lab-frame pion momentum is p_π^{lab}.
- m_h is the invariant mass of particle h. This is assigned based on the known value for final-state particles, but is an outcome of reconstructing final-state particles for composite particles such as D^0 and $D_s^\pm$.

The selection criteria used during reconstruction differ slightly from channel to channel, but generally follow the same logic:

- Selection criteria on so-called *impact parameters* dr and dz to ensure that tracks originate from near the interaction point.
- Minimum momentum requirements to reject hard-to-reconstruct tracks that do not travel far through the detector.
- Particle ID requirements so that there is some confidence that the mass hypothesis being assigned to a track is the correct one.
- Tag-quality selection to reject events that the FEI thinks have a low chance of being correctly reconstructed.
- Selection criteria to reject events where the ROE has a lot of energy.
- Selection criteria on the number of tracks in the ROE.
- Selection criteria on M_{bc} and ΔE related to the B_{tag}.

Certain selection criteria are *rectangular selections*, which keep rectangular areas of two-dimensional space in two variables. In general, on two variables x and y, these selections look like $a < x < b$ and $c < y < d$. Pairs of rectangular selections are combined with *or* statements to only exclude background-rich regions. These selections are particularly useful for reducing file sizes and preserving interesting background regions that can be

Table 4.6: Preselection criteria common to all channels.

Selection criteria	Description
Track $dr < 2$ cm and $\lvert dz\rvert < 3$ cm	Impact parameter selection ensuring tracks originate from near the IP
B_{tag} $\lvert\Delta E\rvert < 0.1\,\text{GeV}$	Ensure B_{tag} ΔE is close to 0
$\log_{10}(\text{FEI signal probability}) > -2.3$ and B_{tag} $M_{\text{bc}} > 5.24$ GeV/c^2 or $\log_{10}(\text{FEI signal probability}) > -8$ and B_{tag} $M_{\text{bc}} > 5.27$ GeV/c^2	Rectangular selection on B_{tag} M_{bc} and $\log_{10}$(FEI signal probability)
E_{ECL}<1.5 GeV and 2 $\leq$ ROE tracks or E_{ECL}<6 GeV and 0 ROE tracks	Rectangular selection on E_{ECL} and number of ROE tracks

used for later calibration studies while rejecting areas where the background is uninteresting. For example, requiring [E_{ECL}<1.5 GeV and 2 $\leq$ ROE tracks] or [E_{ECL}<6 GeV and 0 ROE tracks] selects the signal region (0 ROE tracks and $E_{\text{ECL}} < 1.5\,\text{GeV}$), a sideband with two ROE tracks and $E_{\text{ECL}} < 1.5\,\text{GeV}$, and a sideband with $E_{\text{ECL}} < 6\,\text{GeV}$ and zero ROE tracks. The background-rich regions with two or more ROE tracks and $E_{\text{ECL}} > 1.5\,\text{GeV}$ are discarded. The full suite of preselection criteria is presented in Tables 4.6 and 4.7.

Table 4.7: Channel-specific preselection criteria.

Channel	Selection criteria	Description
$B^{\pm} \to \pi^{\pm} a$	$p^*_{\pi^{\pm}} > 100$ MeV/c	CMS-frame pion momentum selection
	$P_{\pi/K} > 0.6$	PID selection to separate pions from kaons
$B^{\pm} \to K^{\pm} a$	$p^*_{K^{\pm}} > 100$ MeV/c	CMS-frame kaon momentum selection
	$P_{K/\pi} > 0.6$	PID selection to separate kaons from pions
$B^{\pm} \to p a$	$p^*_p > 100$ MeV/c	CMS-frame proton momentum selection
	$P_{p/\pi} > 0.1$	PID selection to separate protons from pions
	$P_{p/K} > 0.1$	PID selection to separate protons from kaons
$B^{\pm} \to D_s^{\pm} a$	$P_{\pi/K(K/\pi)} > 0.6$	PID selection for well-identified final state $\pi^{\pm}(K^{\pm})$
	$p^*_{D_s^{\pm}} > 100$ MeV/c	CMS-frame $D_s^{\pm}$ momentum selection
	$1.919 < m_{D_s^{\pm}} < 2.019$ GeV/c^2	$D_s^{\pm}$ invariant mass is close to nominal mass
$B^0 \to \overline{D}^0 a$	$E_\gamma > 50(100)$ MeV	Energy threshold for γ candidates used in π^0 reconstruction in the barrel (endcap) regions
	$115 < m^0_\pi < 152$ MeV/c^2	π^0 candidate invariant mass is close to nominal mass
	$p^{\rm lab}_{\pi^0} > 200$ MeV/c	Lab-frame π^0 momentum selection
	$P_{\pi/K(K/\pi)} > 0.6$	PID selection for well-identified final state $\pi^{\pm}(K^{\pm})$
	$p^*_{\overline{D}^0} > 100$ MeV/c	Signal-side $\overline{D}^0$ momentum selection
	$1.795 < m_{\overline{D}^0} < 1.935$ GeV/c^2	$\overline{D}^0$ invariant mass is close to nominal mass

Table 4.8: Summary of offline event selection criteria used in the analysis. The exact selection criteria on BDT outputs are channel-dependent.

Selection	Description
B_{tag} $M_{\text{bc}} > 5.27\,\text{GeV}/c^2$	Select events with M_{bc} consistent with correctly reconstructed B mesons
$\log_{10}(\text{FEI signal probability}) > -2.3$	Select well-tagged B candidates
$E_{\text{ECL}} < 1.5\,\text{GeV}$	Select events with low extra ECL energy
0 tracks in ROE	Select events with no extra tracks
BDT0 selection	Further remove continuum background
BDT1 selection	Further remove B background

4.5 Event selection

The preselection criteria detailed above are the requirements we impose on events before they are saved for further analysis (or *offline* analysis). They are deliberately looser than the final selection criteria used for the analysis because we want to retain some sidebands for control mode studies. Furthermore, it would be impossible to fully optimise the selection criteria on events without any preselection criteria as the datasets would be too large to work with effectively. As it stands, the data files saved after the preselection criteria use several hundred gigabytes of storage. The final selection criteria are shown in Table 4.8. The exact numerical values of the BDT output selection differ for each channel, but the other selection criteria are the same for all channels. A more detailed description of the BDTs is given in Sections 4.5.1 and 4.5.2. Histograms of B_{tag} M_{bc}, $\log_{10}(\text{FEI signal probability})$, E_{ECL}, and numbers of ROE tracks for all channels are shown in Figures 4.1 to 4.5. These histograms are all $N-1$ plots, meaning that they show only the events that survive the selection criteria on all other selection variables, save for the one being displayed, and the BDT-based selection.

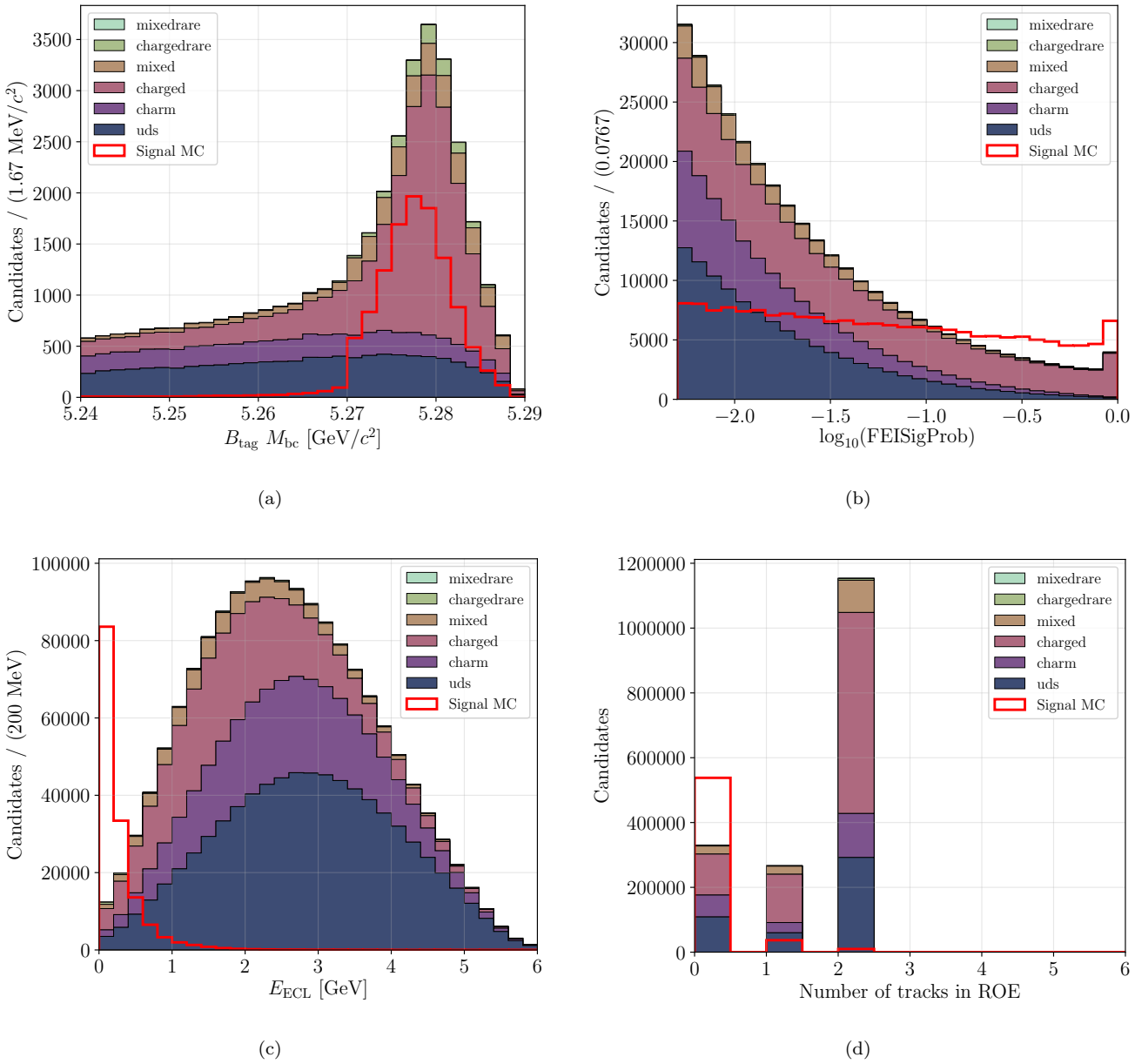


Figure 4.1: $B^{\pm} \to \pi^{\pm} a$ histograms showing distributions of key selection variables after all other selection criteria, apart from BDT-based selection, have been applied.

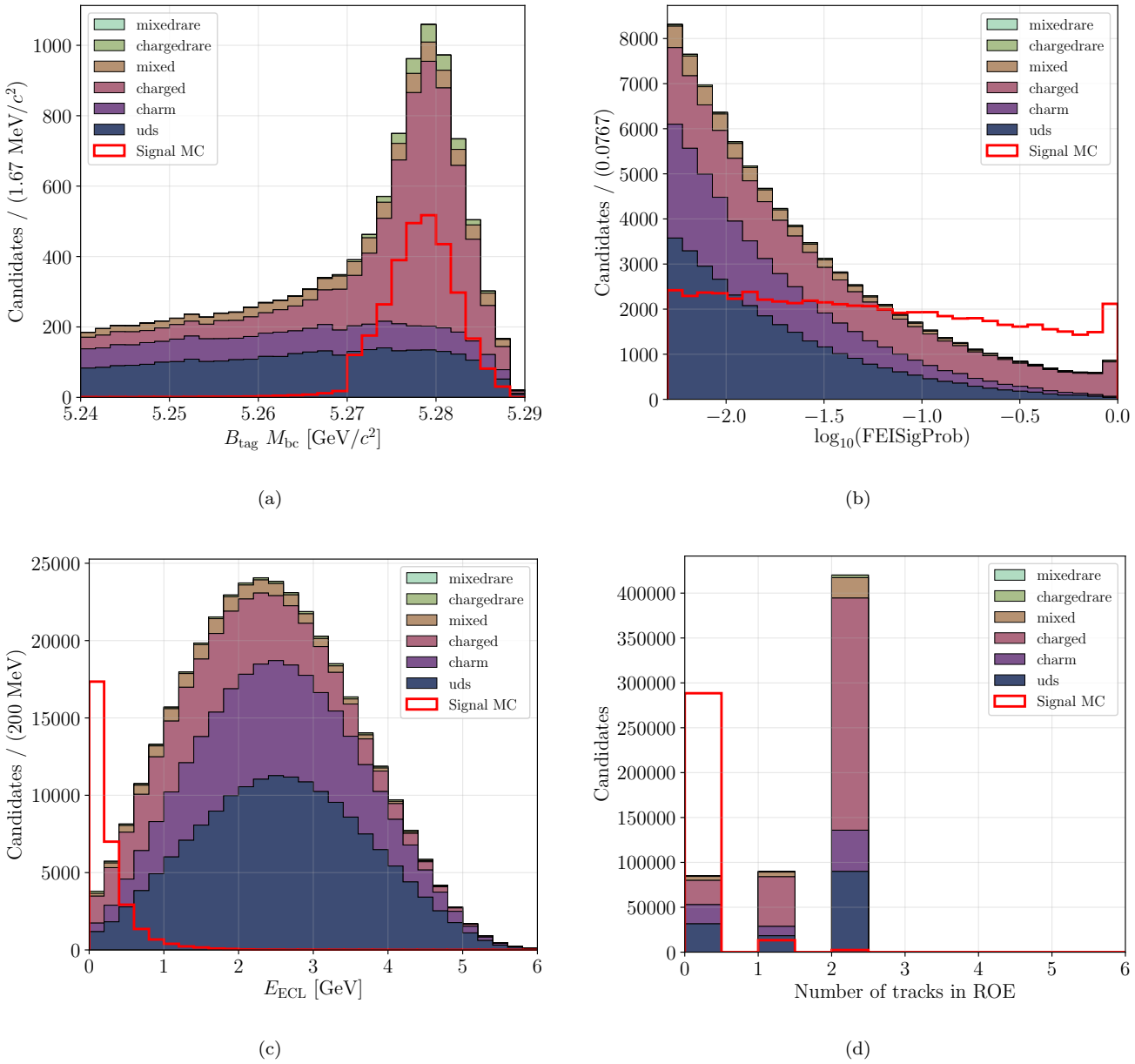


Figure 4.2: $B^{\pm} \to K^{\pm}a$ histograms showing distributions of key selection variables after all other selection criteria, apart from BDT-based selection, have been applied.

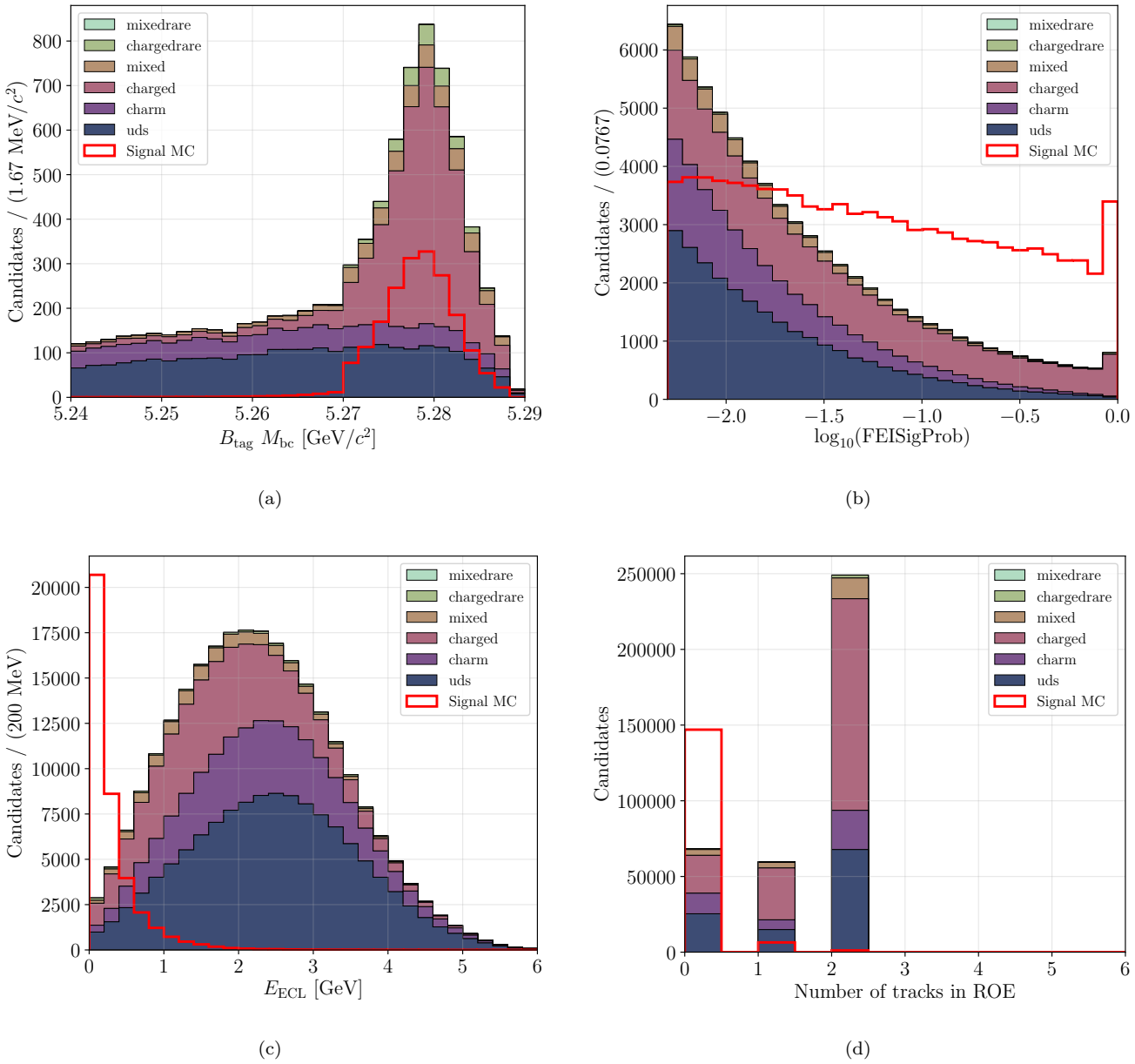


Figure 4.3: $B^{\pm} \to pa$ histograms showing distributions of key selection variables after all other selection criteria, apart from BDT-based selection, have been applied.

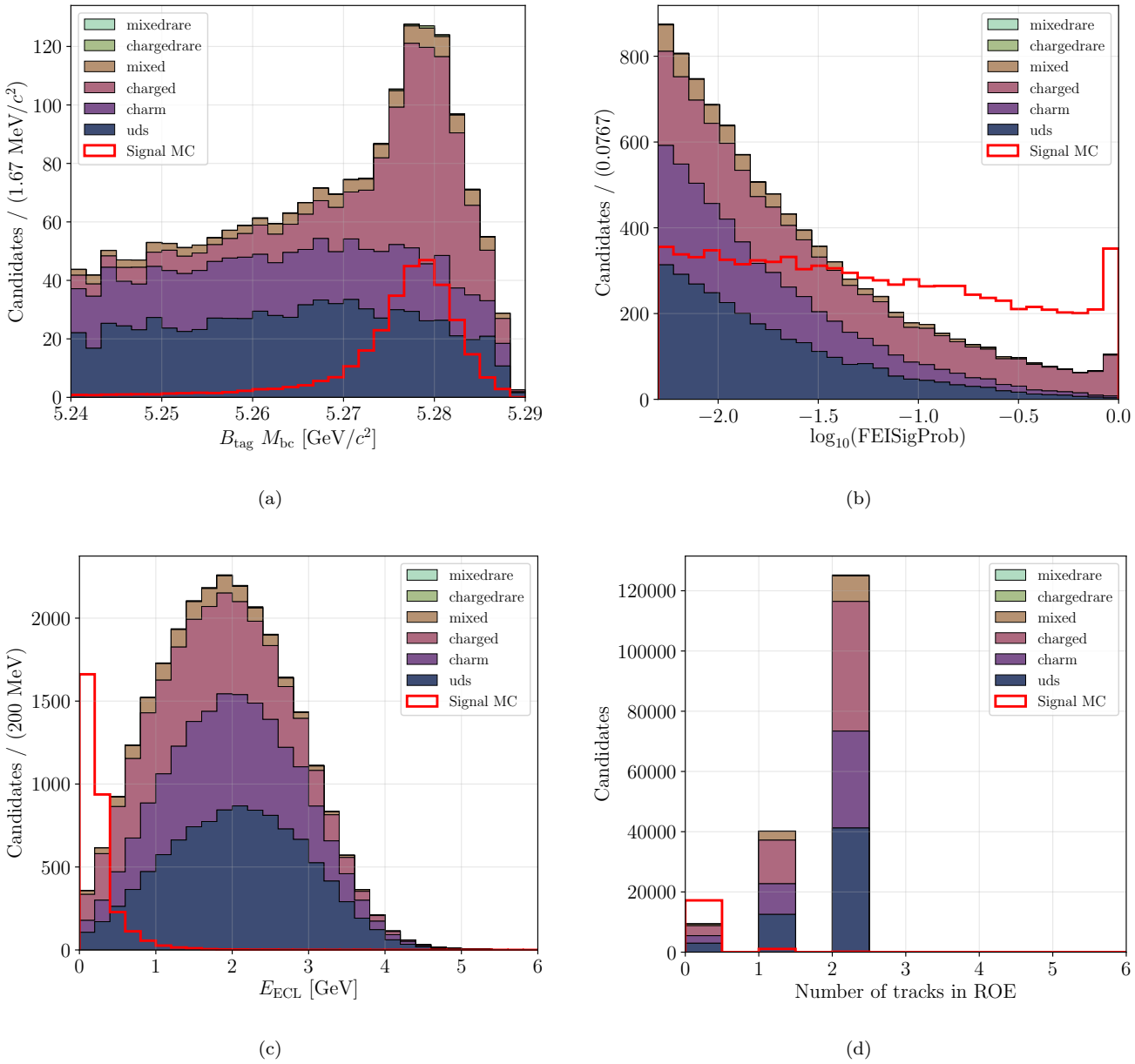


Figure 4.4: $B^{\pm} \to D_s^{\pm} a$ histograms showing distributions of key selection variables after all other selection criteria, apart from BDT-based selection, have been applied.

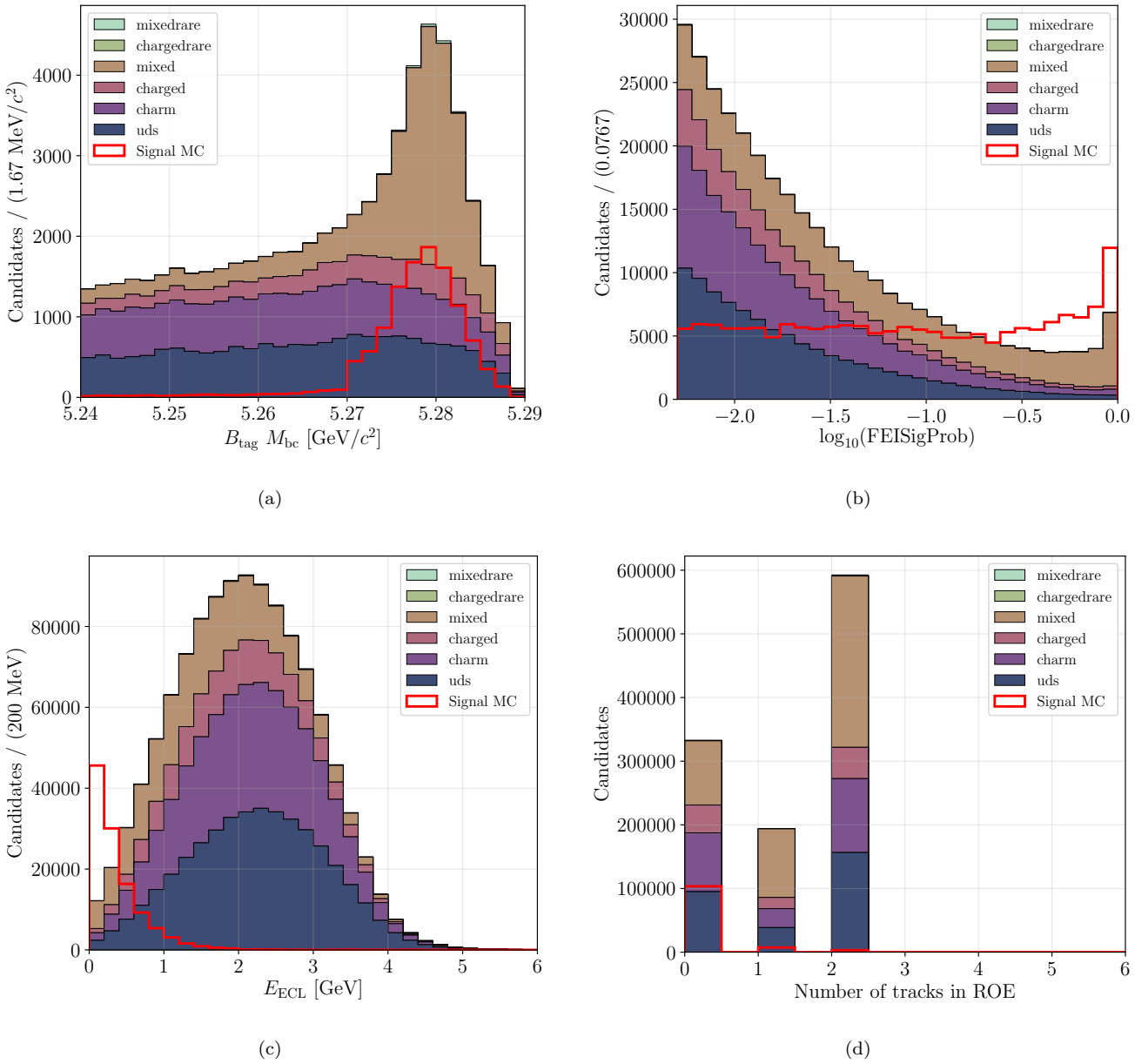


Figure 4.5: $B^0 \to \overline{D}^0 a$ histograms showing distributions of key selection variables after all other selection criteria, apart from BDT-based selection, have been applied.

4.5.1 Boosted decision trees

Traditionally, analyses at collider experiments have approached the separation of signal and background events by considering sets of variables and, variable by variable, retaining events that fall on either side of a defined threshold. The complete set of selection criteria is then intended to create a purer sample of signal; that is, to maximise the signal yield while minimising the background rate. However, this binary selection process can potentially exclude signal-rich regions that might have been retained under alternative criteria. Using decision trees instead of hierarchical selection bypasses this issue. Building the decision tree, or training the model, follows a similar methodology to hierarchical selection, but each decision point further divides the data into signal-like and background-like subsets, forming a branching data structure, or tree. The branching continues either until a fixed depth is reached, or until the number of data points in a branch falls below a specified minimum.

When a decision tree classifier is applied to new events, it assigns a score based on the signal purity of the terminal node where the event concludes its path through the tree. This process maps the arbitrarily dimensioned input space to a single score value, which can then be utilised by the analyst.

Real data is unfortunately subject to statistical noise, which large decision trees can end up modelling instead of the underlying structure of the data. This is known as over-training, and is undesirable in any machine learning model. Ensembles of classifiers can be employed to mitigate the over-training issue. One such method is the boosted decision tree, which combines many smaller, shallower trees. Trees that are added to the model after the first tree are not just trained to optimally separate signal and background events, but also to minimise the error of preceding decision trees. There are many methods through which this can be achieved [156, 157].

The classifiers in this work are gradient boosted decision trees [158]. XGBoost is used to directly separate signal and background [159], and FastBDT [160–162] is used within the FEI. The algorithm relies on comparing the model output $f(\vec{x})$, where $\vec{x}$ is the vector of input features, to the correct output y, using some differentiable loss function $L(y, f(\vec{x}))$. It is initialised with a constant response $f_0(\vec{x}) = c$, and then is iterated upon M times. For the m^{th} tree, pseudo-residuals are calculated as

$$r_{im} = -\left[\frac{\partial L(y_i, f(\vec{x}_i))}{\partial f(\vec{x}_i)}\right]_{f(\vec{x})=f_{m-1}(\vec{x})}. \tag{4.3}$$

A weight γ_m is then computed such that it minimises

$$L(y_i, f_{m-1}(\vec{x}_i) + \gamma h_m(\vec{x}_i)) \tag{4.4}$$

with respect to γ where h_m is the new tree. To slow the rate of the training process in order to avoid over-training, the new tree is added to the model with a weight η, which is known as the learning rate

$$f_m(\vec{x}) = f_{m-1}(\vec{x}) + \eta\gamma_m h_m(\vec{x}). \tag{4.5}$$

This repeats M times until the maximum number of trees has been added to the model. To add further robustness against over-training, each tree can be trained with a subset of the total training dataset. The XGBoost implementation of this algorithm adds an additional term to the loss function that penalises complex trees to further combat over-training.

4.5.2 Background suppression with boosted decision trees

Background processes at Belle can generally be split into two groups: generic $B\overline{B}$ background and $q\overline{q}$ or continuum background. As the $\Upsilon(4S)$ resonance is very close to $2m_B$, B mesons have limited momenta in the CMS frame, around 300 MeV/c. In this scenario, their decay products are produced somewhat isotropically. Continuum events, which contain only lighter quarks, do not have the same isotropy. Due to momentum conservation, all e^+e^- decay products are produced back-to-back in the CMS frame. The light quark jets of continuum events are therefore far more directional, as their momentum is much higher than that of B mesons. As such, the event shapes of these two kinds of events differ, with $B\overline{B}$ events having a spherical shape and continuum events being more jet-like as shown in Fig. 4.6.

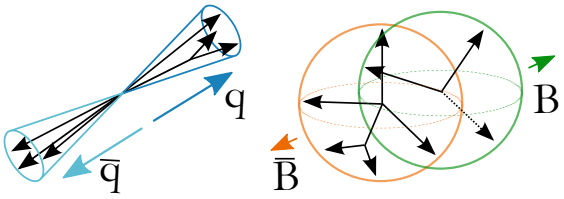


Figure 4.6: Comparison of event shapes between $B\overline{B}$ events, which are spherical, and continuum events, which are more jet-like. This figure is taken from Ref. [163].

BDT training is done on a collection of variables that describe the shape of an event. The continuum suppression variables are built around the tag-side B and use the partially reconstructed B_{sig} and previously defined ROE as the *rest of event* for continuum suppression. A description of the variables follows:

- Thrust-related variables: The *thrust axis* $\hat{T}$ of an event is the axis that maximises the sum of the projections of all the event's particles' momenta onto it in the CMS

frame. Using it, one can define a scalar quantity T known as the *thrust*, given by

$$T = \frac{\sum_{i=1}^{N} |\hat{T} \cdot \vec{p_i}|}{\sum_{i=1}^{N} |\vec{p_i}|} \tag{4.6}$$

where N is the number of particles in the event.

When these variables are calculated using only the B_{tag}, we get the tag-only thrust and thrust axis, T_B and $\hat{t}_B$, respectively. We can define θ_T as the angle between the thrust axis of the B candidate decay particles and the thrust axis of the ROE. Analogously, we can define $\theta_{T,z}$ as the angle between the thrust axis of the B candidate and the beam axis. For a $B\overline{B}$ event, the mesons are produced almost at rest in the CMS frame, so their decay particles are isotropically distributed, and in the range [0, 1], $\cos(\theta_T)$ should be a random uniform distribution. For continuum events, the momenta are more jet-like, following the hadronic jets in the event, which leads to $\cos(\theta_T)$ having a peaking structure. We use the cosines of these two angles, $\cos(\theta_T)$ and $\cos(\theta_{T,z})$ as continuum suppression variables.

- Kakuno Super Fox-Wolfram (KSFW) variables [164–166]: These are described in detail in the Belle II Physics Book [167] and The Physics of the B Factories [148]. In calculating these variables, all particles are categorised as belonging to the B_{tag} candidate (denoted with s) or to the ROE (denoted with o), and different orders of moments arise from the order of Legendre polynomial ℓ used to calculate them. They are calculated as terms in the Fisher discriminant

$$KSFW = \sum_{l=0}^{4} R_l^{so} + \sum_{l=0}^{4} R_l^{oo} + \gamma \sum_{n=1}^{N_t} |(P_t)_n|. \tag{4.7}$$

The moment R^{so} is split into three categories: charged c, neutral n, and missing m, and is defined as

$$R_l^{so} = \sum_{x \in c,n,m} \frac{\alpha_{xl} H_{xl}^{so}}{E_{\text{beam}}^{*} - \Delta E}, \tag{4.8}$$

where α_{xl} are Fisher coefficients and the missing mass squared mm^2 is defined as

$$mm^2 = \left(E_{\Upsilon(4S)} - \sum_{n=1}^{N} E_n\right)^2 - \left(\sum_{n=1}^{N} \vec{p}_n\right)^2 \tag{4.9}$$

where the energy and momentum of the n^{th} particle in the event are E_n and $\vec{p}_n$ respectively and $E_{\Upsilon(4S)}$ is the energy of the $\Upsilon(4S)$ in the rest frame of the event.

For odd l, we have

$$H_{nl}^{so} = H_{ml}^{so} = 0 \text{ and} \tag{4.10}$$

$$H_{cl}^{so} = \sum_{i} \sum_{jx} Q_i Q_{jx} |p_{jx}| P_l(\cos\theta_{i,jx}), \tag{4.11}$$

where the Q factors are the charges of the respective particles, p is the momentum and $P_l(\cos\theta_{i,jx})$ is the l^{th} order Legendre polynomial of the angle between particles i and jx (where x still indexes the ROE categories of charged, neutral, and missing).

For even l, we have

$$H_{xl}^{so} = \sum_i \sum_{jx} |p_{jx}| P_l(\cos\theta_{i,jx}). \tag{4.12}$$

The moment R^{oo} is defined similarly to R^{so}

$$R_l^{oo} = \frac{H_l^{oo}}{E_{\text{beam}}^* - \Delta E} \tag{4.13}$$

where for odd and even l,

$$H_l^{oo} = \sum_j \sum_k b_l^{oo} Q_j Q_k |p_j||p_k| P_l(\cos\theta_{jk}) \tag{4.14}$$

and

$$H_l^{oo} = \sum_j \sum_k b_l^{oo} |p_j||p_k| P_l(\cos\theta_{jk}) \tag{4.15}$$

respectively.

There are a number of free parameters in the definitions above. For R_l^{so} with odd l, there are two free parameters, $a_{xl} \cdot b_l^{so}$, while there are 9 for even l. For R_l^{oo}, there is one free parameter for each l.

- CLEO cones: These cones are created by dividing the space around the event thrust axis into polar angle intervals of 10° and performing a scalar sum of the lab-frame particle momenta in each cone, as first done by the CLEO Collaboration [168].

A multi-layered BDT approach is taken to separately classify continuum and B backgrounds. The first BDT layer, *BDT0*, was trained to separate events containing B mesons from continuum events. It was trained on 400 fb^{-1} of generic continuum and $B\overline{B}$ events. Two-thirds of this dataset was used for training, and the rest for validation. The second BDT layer, *BDT1*, was trained to separate signal events from background events that passed a selection on BDT0. This BDT was trained on events that satisfied selection criteria on BDT0 that retained 90% of signal events (with all other selection criteria applied). 80% of events were used for training, and 20% were reserved for validation. The signal sample used in training BDT1 was a mixture of all generated mass hypotheses.

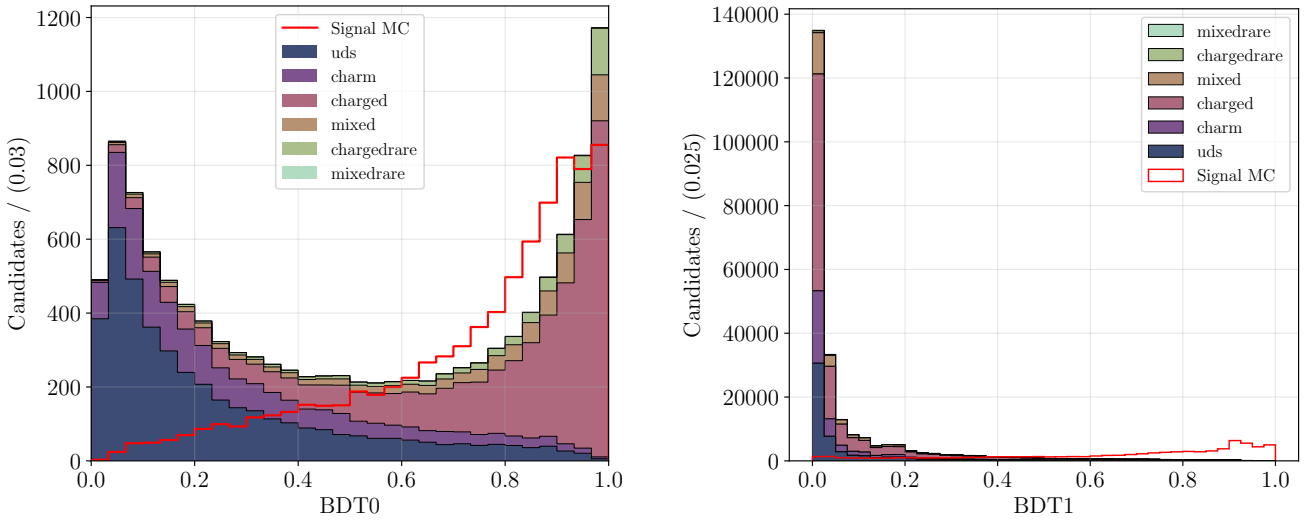
The variables used for training are the same for all channels, and are shown in Table 4.9. The variables relating to event shape, as well as E_{ECL} and B_{tag} ΔE, have already been

described. The flight distance of the B_{tag} is calculated based on the vertex fitted to the B_{tag}. Not all of the KSFW variables are used. Variables that are highly correlated with each other do not offer substantial new information for the classifier. H_{12}^{so} is highly correlated with E_{ECL}, so it is not used in training BDT1. In signal MC, the H_{2l}^{so} for even l are highly correlated with H_{0l}^{oo}, and H_{l}^{oo} for $l \in 0, 2, 4$ are highly correlated with KSFW missing mass squared, so they are dropped from the set of BDT1 training variables. None of the training variables are strongly correlated with p_h^S.

Table 4.9: The sets of variables used for training the two background suppression BDTs.

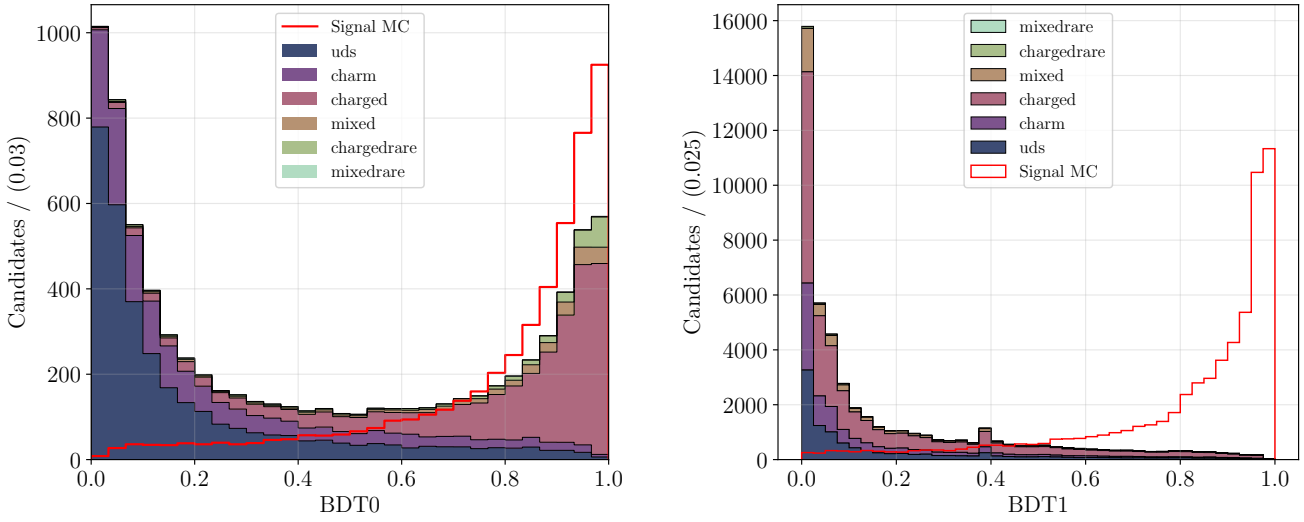
BDT	Variables
BDT0	B_{tag} ΔE, flight distance, R_2, $\cos(\theta_T)$, $\cos(\theta_{T,z})$, $\hat{T}(O)$, $\hat{T}(B)$, KSFW E_T, M^2_{miss}, H_{xl}^{so} for $x \in$ [1, 2, 3], $l \in$ [0, 2, 4], H_{l}^{oo} for $l \in$ [1, 2, 3, 4], CLEO cones $1-9$
BDT1	E_{ECL}, B_{tag} ΔE, flight distance, R_2, $\cos(\theta_T)$, $\cos(\theta_{T,z})$, $\hat{T}(O)$, $\hat{T}(B)$, KSFW E_T, M^2_{miss}, H_{0l}^{so}, $l \in$ [0, 2, 4], H_{1l}^{so}, $l \in$ [0, 4], H_{l}^{oo} for $l \in$ [1,3], CLEO cones $1-9$

Distributions of the BDT outputs are shown in Figures 4.7 to 4.11 with all non-BDT-based selection applied on the BDT0 plots, and with the BDT0 selection included on the BDT1 plots. Background MC is scaled to the size of the Belle dataset, and the overlaid signal MC distributions are arbitrarily scaled. In both of these histograms, all other selection criteria are applied to the samples, apart from the criteria on the BDT being plotted (with the exception of the datasets shown in (a), which also do not have any selection on BDT1 applied, so that the continuum suppression data-MC agreement could be shown independently from any effects of BDT1). In Section 6.1, we discuss in detail the use of $B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \ell^+\ell^-$ for $\ell \in e, \mu$ as a control channel to check agreement between data and MC. Fig. 4.12 demonstrates that the BDT output distributions show good agreement between data and MC in this control mode.



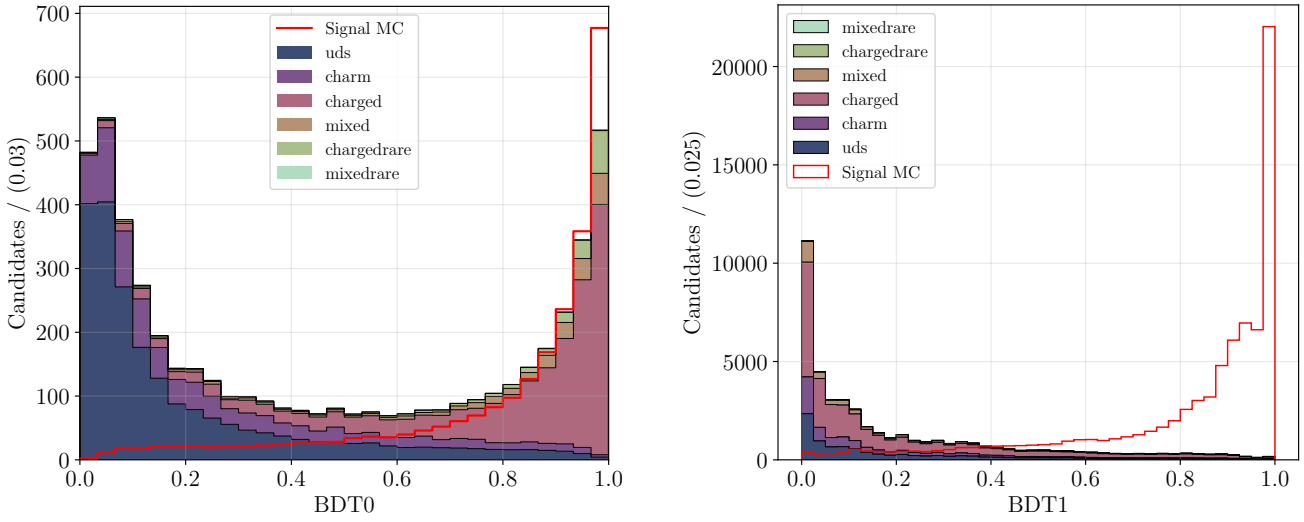
(a) (b)

Figure 4.7: Pion channel BDT responses to signal and background Monte Carlo (MC) simulations. The plots show the distribution of BDT0 (left) and BDT1 (right) outputs after applying all other selection criteria, highlighting the separation between signal and background events.



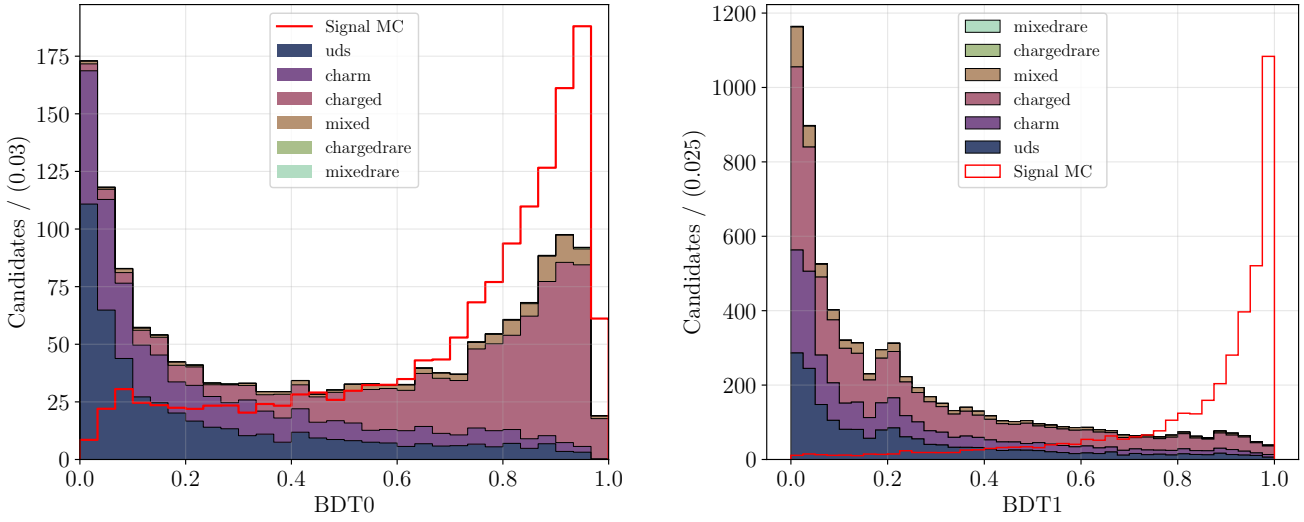
(a) (b)

Figure 4.8: Kaon channel BDT responses to signal and background Monte Carlo (MC) simulations with all other selection criteria applied. The left plot shows the BDT0 response, which is trained to separate B meson events from continuum events, while the right plot shows the BDT1 response, which is trained to separate signal events from B background events.



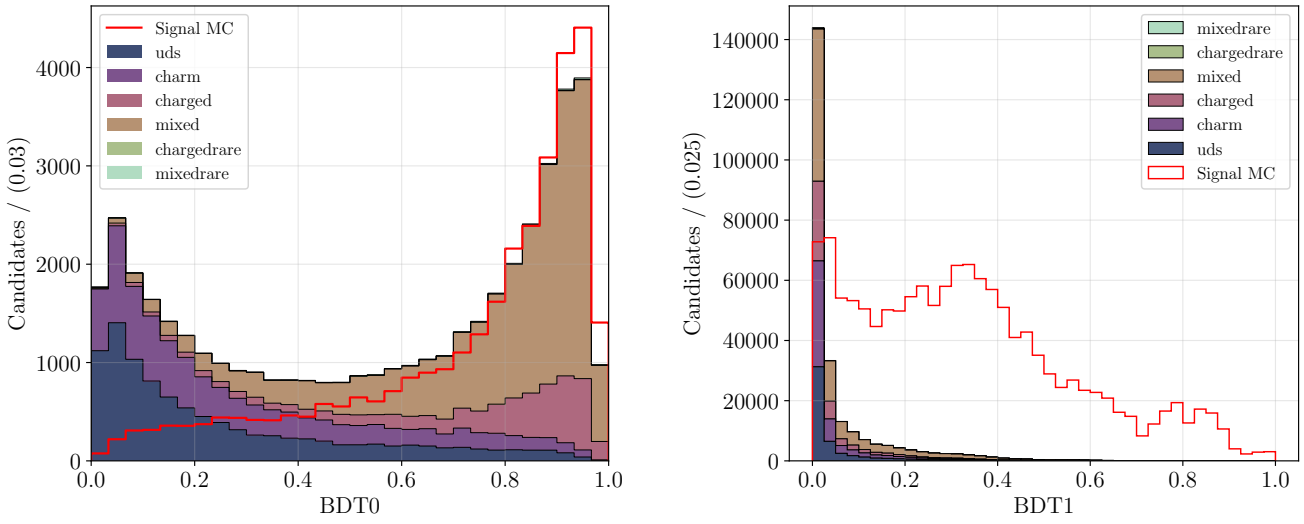
(a) (b)

Figure 4.9: Proton channel BDT responses to signal and background Monte Carlo (MC) simulations. The plots show the distribution of BDT0 (left) and BDT1 (right) outputs after applying all other selection criteria, highlighting the separation between signal and background events.



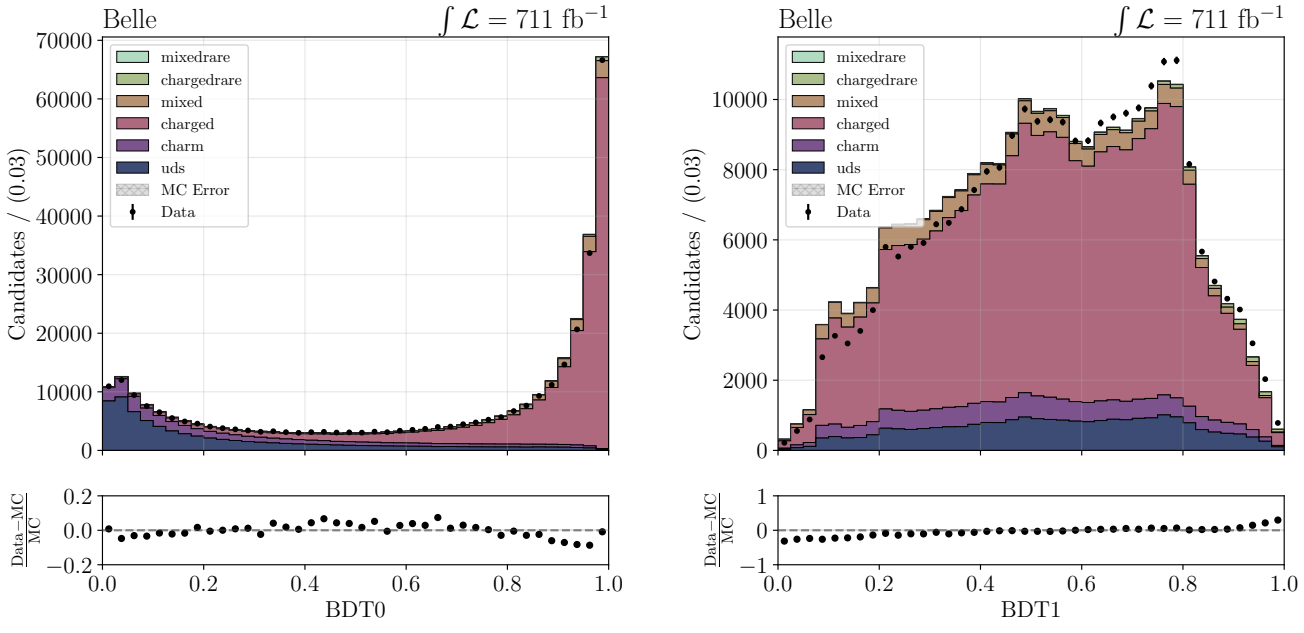
(a) (b)

Figure 4.10: BDT responses for the $D_s^{\pm}$ channel, showing the separation between signal and background Monte Carlo (MC) simulations. The left plot shows the BDT0 response, which is trained to separate B meson events from continuum events, while the right plot shows the BDT1 response, which is trained to separate signal events from B background events. All other selection criteria have been applied.



(a) (b)

Figure 4.11: BDT responses for the $\overline{D}^0$ channel, showing the separation between signal and background events in Monte Carlo simulations. The left plot displays the BDT0 output, which distinguishes B meson events from continuum events, while the right plot shows the BDT1 output, which separates signal events from B background events, with all other selection criteria applied.



(a) (b)

Figure 4.12: Comparison of data and Monte Carlo (MC) simulations for the kaon channel BDT responses in the sideband region with two tracks in the Rest of Event (ROE). The plots show the distribution of BDT0 (left) and BDT1 (right) outputs, highlighting the agreement between data and MC simulations.

Table 4.10: `XGBoost` parameters used for all BDT training.

Hyperparameter	Value
Number of estimators	2000
Learning rate	0.02
Maximum depth	3
Evaluation metric	logloss

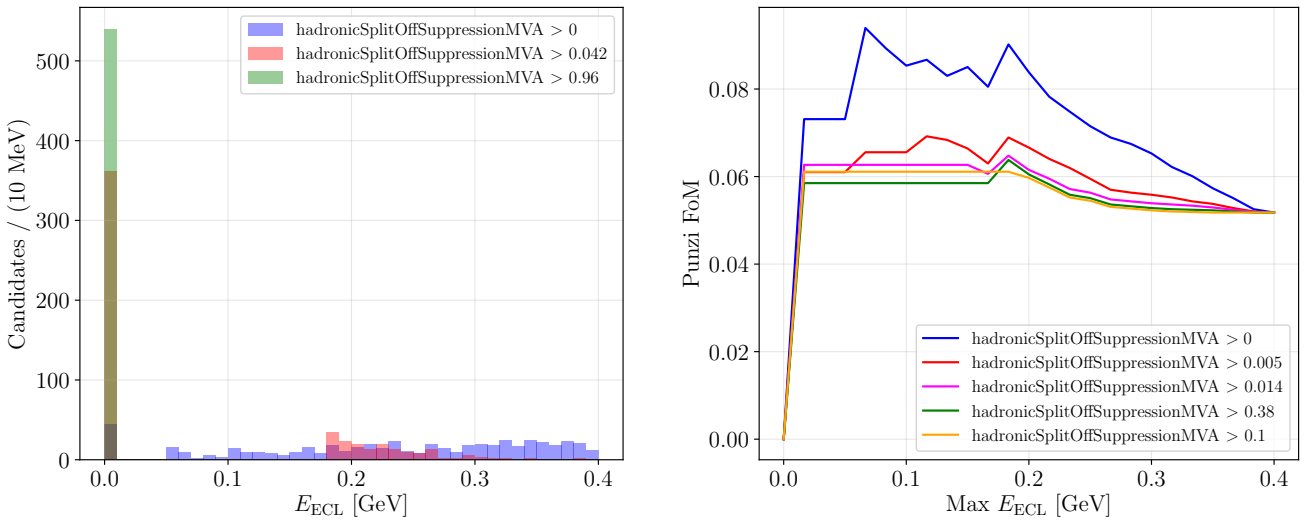
There are a number of hyperparameters in `XGBoost` that control the learning process, and a grid search was performed to search over different values of these hyperparameters to optimise the BDT models used. The package `Hyperopt` is used to tune the `XGBoost` parameters. We found that there was negligible performance gained between the default `XGBoost` settings and the optimised hyperparameters. The final training was done with the default settings, described in Table 4.10.

4.5.3 Beam background and hadronic split-off

A particularly powerful variable used to separate signal and background in this analysis is E_{ECL}, which we have defined as the sum of energy of all neutral clusters in the ECL from the ROE. If the signal process is reconstructed correctly, then we expect this to be zero; however, as seen in the E_{ECL} plots shown in Figures 4.1 to 4.5, there is a substantial fraction of signal events with non-zero E_{ECL}. Two sources of ECL clusters that are unrelated to the signal event are photons from beam background and photons from hadronic split-off. Hadronic split-off occurs when a hadron interacts with detector material and produces more hadrons. These hadrons can originate far from the interaction point, but could result in ECL clusters close to a track, especially if the hadronic split-off occurs close to the ECL surface.

In `basf2`, there is a multivariate classifier trained to discriminate between hadronic split-off and non-hadronic split-off ECL clusters. A study was performed to assess this classifier's impact on the invisible particle search. The hadronic split-off classifier output was used to select clusters, reducing the total number of ECL clusters in the ROE for most events. This shifted the E_{ECL} distribution towards zero, as shown in Fig. 4.13(a) for both signal and background events. We found that hadronic split-off suppression lowers the Punzi FoM [169] (see Fig. 4.13(b)) that this analysis has been optimised with, so we do not include any hadronic split-off suppression selection in the final selection criteria. In summary, Fig. 4.13 demonstrates that the Punzi FoM is maximised when no hadronic split-off suppression is applied, indicating that hadronic split-off suppression is not beneficial for this analysis. Belle has fewer variables saved that describe ECL clusters than Belle II, and as such these cluster classifiers have better performance in Belle II. For Belle, this cluster suppression mostly relied on the distance between the

cluster and the closest track's position on the surface of the ECL. Furthermore, these classifiers have recently shifted to do catch-all fake photon suppression, and their training has been focused on Belle II, which is of little use to this Belle analysis [170, 171].



(a) (b)

Figure 4.13: (a) E_{ECL} distributions for three different hadronic split-off MVA thresholds (one stream of generic MC, experiment 65). The plot shows how the hadronic split-off suppression affects the E_{ECL} distribution. (b) Punzi FoM curves for different hadronic split-off MVA thresholds as a function of the E_{ECL} selection criteria.

4.5.4 Optimising selection criteria

The selection criteria thresholds placed on the BDT outputs are optimised using a 2D grid search over the outputs for BDT0 and BDT1 for each mass hypothesis. The grid search tries to find the combination of two BDT thresholds that maximises a Punzi figure of merit [169]

$$\mathrm{FoM}_{\mathrm{Punzi}} = \frac{\varepsilon_{\mathrm{selection}}}{\frac{\mathrm{n}_\sigma}{2} + \sqrt{\mathrm{B}}} \tag{4.16}$$

where $\varepsilon_{\mathrm{selection}}$ is the total signal selection efficiency (before best candidate selection), n_σ is the level of significance the analysis is targeting (which we choose to be 3 for this analysis), and B is the number of background events. For our purposes, we restrict B to be the number of background events in a $3\sigma_{\mathrm{eff}}$ window around the mean of the signal peak for the mass hypothesis we are optimising for. As we want to use the same selection across any single BDT training region, we take the average threshold that the grid search finds for each region. For some channels, this is then slightly loosened manually until the selection efficiency is effectively flat for all values of m_a. Two examples of 2D scans in the Punzi FoM are shown in Fig. 4.14 for two different mass hypotheses, $m_a = 1\,\mathrm{MeV}/c^2$

and $m_a = 2.272\,\text{GeV}/c^2$, in the channel $B^\pm \to pa$. The optimal selection threshold for BDT0 varies depending on the mass hypothesis, with different values for 1 MeV/c^2 and 2.272 GeV/c^2. The final selection criterion applied to all events is 0.379, which falls between the optimal values for these two masses. However, as shown in Fig. 4.14, this choice does not significantly reduce the FoM.

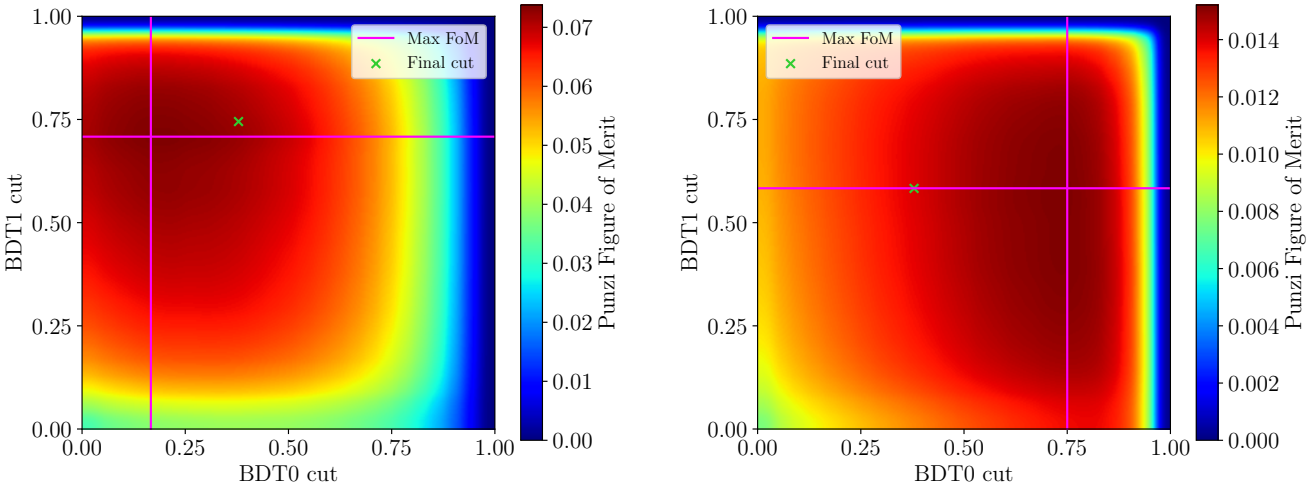


Figure 4.14: Punzi FoM for different selection combinations on BDT0 and BDT1 for the $B^\pm \to pa$ search. The point where the two magenta lines intersect shows the combination that optimises the Punzi FoM for the specific mass of a in the signal MC used to make each plot. The green cross shows the placement of the final selection criterion that is used for all masses in the relevant training region.

4.5.5 Irreducible backgrounds

While a lot of effort has gone into reducing the amount of background in this analysis, there are some processes which can mimic the signal process $B \to ha$ exactly and are therefore impossible to reject. The Belle detector has only finite acceptance, so particles that pass through uninstrumented parts of the detector can mimic invisible particles. Furthermore, if more than one particle goes undetected in this way, the missing momentum of the event will not necessarily point to these uninstrumented regions. Processes that produce neutral kaons can mimic the signal as well. The decay vertices of K^0_S are often significantly displaced from the IP, which reduces the efficiency of finding their charged particle daughters. Furthermore, the KLM detector does not detect K^0_L with perfect efficiency, so K^0_L often go completely missing in the Belle detector. The performance of the KLM is somewhat poorly understood, and its behaviour in MC is very different to its performance in real data. In fact, at lower momenta, the K^0_L

detection efficiency in data and simulation differs by about 50% [172]. A short study was performed to assess the feasibility of a K_L^0 veto using the KLM. We found that only about half the K_L^0-containing background could be removed. As this veto would still leave in problematic backgrounds such as $B^\pm \to \pi^\pm K_L^0$ and also introduce an additional source of systematic uncertainty, it is not used in the final analysis.

4.5.6 Peaking backgrounds

Any two-body B decay that satisfies the full selection criteria will cause a peaking structure in p_h^S. The backgrounds that are the hardest to reject usually have a K_L^0 in the final state, either directly produced in a B decay, or in the decay of a D meson on the signal side, where the other daughters of the D meson are lost as well. In this analysis, it is assumed that the decay width of a is small compared to the detector resolution so that the width of the observed peak is only determined by detector effects. However, there are many particles with significantly broader decay widths. For instance, $f_0(1370)$ has a width of 500 MeV/c^2, and would therefore cause a much broader peak in p_h^S than our assumed a with the same mass. In Section 5.2.5, we show that the search is less sensitive to broad resonances such as this. An exhaustive list of possible peaking backgrounds is presented in Tables 4.11 to 4.15. These tables detail all of the two-body B decays that are accepted by the full selection criteria, and include their branching fractions as reported by the PDG [7] in 2023, as well as the branching fractions used in generic Belle MC production. The masses and decay widths are the values the particles have in MC used by Belle. The number expected from MC is the total number of MC events that survive all selection criteria, including best candidate selection, scaled to match the Belle luminosity (but not rescaled based on differences between their production branching fractions in MC and as reported by the PDG). These decays are classified as either *veto* or *rare*. We do not attempt to perform signal extraction around the masses of the backgrounds labelled veto. Rare peaking backgrounds are treated as a source of systematic uncertainty. Their treatment is described in more detail in Section 6.3.

Table 4.11: Two-body decays present in the background for the $B^{\pm} \to \pi^{\pm} a$ search, listed in order of most frequently appearing in the MC background after selection.

Decay	Neutral hadron mass (GeV/c^2)	Neutral hadron decay width (MeV/c^2)	Number from MC	Branching fraction (PDG)	Branching fraction (Belle MC)	Type
$B^+ \to \pi^+ \overline{D}^0$	1.865	0.0	127	$(4.68 \pm 0.13) \times 10^{-3}$	4.84×10^{-3}	Veto
$B^+ \to \pi^+ D^*(2007)^0$	2.007	0.0	95	$(5.18 \pm 0.15) \times 10^{-3}$	5.2×10^{-3}	Veto
$B^+ \to \pi^+ K^0$	0.498	0.0	47	$(2.37 \pm 0.08) \times 10^{-5}$	2.94×10^{-3}	Veto
$B^+ \to \pi^+ D_0(2400)^0$	2.300	274.0	6.5	$(6.4 \pm 1.4) \times 10^{-4}$	8.3×10^{-4}	Veto
$B^+ \to \pi^+ K_0(1430)^0$	1.425	270.0	6.12	$(3.9^{+0.6}_{-0.5}) \times 10^{-5}$	1.07×10^{-5}	Veto
$B^+ \to \pi^+ D_2(2460)^0$	2.461	47.5	3.92	$< 1.3 \times 10^{-3}$	8.5×10^{-4}	Rare
$B^+ \to \pi^+ D_1(2430)^0$	2.427	384.0	3.12	$(3.5 \pm 0.6) \times 10^{-4}$	6.2×10^{-4}	Rare
$B^+ \to \pi^+ K(892)^0$	0.896	47.3	3.10	$(1.01 \pm 0.08) \times 10^{-5}$	3.56×10^{-6}	Rare
$B^+ \to \pi^+ D_1(2420)^0$	2.421	31.7	2.52	$< 6 \times 10^{-6}$	9.3×10^{-4}	Rare
$B^+ \to \pi^+ K^*(1410)^0$	1.414	232.0	1.18	$< 4.5 \times 10^{-5}$	9.9×10^{-7}	Rare
$B^+ \to \pi^+ K(1680)^0$	1.718	322.0	0.88	$< 1.2 \times 10^{-5}$	1.55×10^{-6}	Rare
$B^+ \to \pi^+ \pi^0$	0.135	0.0	0.42	$(5.5 \pm 0.4) \times 10^{-6}$	5.59×10^{-6}	Rare
$B^+ \to \pi^+ K^*(1430)^0$	1.432	109.0	0.50	$(5.6^{+2.2}_{-1.5}) \times 10^{-6}$	6.9×10^{-6}	Rare
$B^+ \to \pi^+ a_0(1450)^0$	1.474	265.0	0.44	Unknown	5.8×10^{-6}	Rare
$B^+ \to \pi^+ f_0(1370)$	1.370	500.0	0.38	$< 4.0 \times 10^{-6}$	3×10^{-6}	Rare
$B^+ \to \pi^+ f_2(1270)$	1.276	186.7	0.48	$(2.2^{+0.7}_{-0.4}) \times 10^{-6}$	8.2×10^{-6}	Rare
$B^+ \to \pi^+ J/\psi$	3.097	0.093	0.10	$(3.92 \pm 0.08) \times 10^{-5}$	4.8×10^{-5}	Rare
$B^+ \to \pi^+ a_1(1260)^0$	1.230	425.0	0.26	$(2.0 \pm 0.6) \times 10^{-5}$	2.04×10^{-5}	Rare
$B^+ \to \pi^+ b_1(1235)^0$	1.230	142.0	0.26	$(6.7 \pm 2.0) \times 10^{-6}$	2.0×10^{-6}	Rare
$B^+ \to \pi^+ \rho^0$	0.775	147.8	0.06	$(8.3 \pm 1.2) \times 10^{-6}$	8.7×10^{-6}	Rare
$B^+ \to \pi^+ \eta$	0.548	0.0	0.12	$(4.02 \pm 0.27) \times 10^{-6}$	4.4×10^{-6}	Rare
$B^+ \to \pi^+ \chi_{c1}(1P)$	3.511	0.8	0.12	$(2.2 \pm 0.5) \times 10^{-5}$	2.2×10^{-5}	Rare
$B^+ \to \pi^+ f_2'(1525)$	1.517	86.0	0.08	Unknown	1.0×10^{-7}	Rare
$B^+ \to \pi^+ \omega$	0.783	8.4	0.02	$(6.9 \pm 0.5) \times 10^{-6}$	6.9×10^{-6}	Rare
$B^+ \to \pi^+ a_2(1320)^0$	1.319	105.0	0.06	Unknown	2.6×10^{-6}	Rare
$B^+ \to \pi^+ f_1(1285)$	1.282	22.7	0.04	Unknown	2×10^{-6}	Rare
$B^+ \to \pi^+ \phi$	1.019	4.2	0.02	$(3.2 \pm 1.5) \times 10^{-8}$	1.2×10^{-7}	Rare
$B^+ \to \pi^+ b_1(1235)^0$	1.230	142.0	0.02	$(6.7 \pm 2.0) \times 10^{-6}$	2.0×10^{-6}	Rare
$B^+ \to \pi^+ \eta(1405)$	1.409	50.1	0.02	Unknown	2.0×10^{-6}	Rare
$B^+ \to \pi^+ \eta'$	0.958	0.188	0.02	2.7×10^{-6}	2.7×10^{-6}	Rare

Table 4.12: Two-body decays present in the background for the $B^{\pm} \to K^{\pm}a$ search, listed in order of most frequently appearing in the MC background after selection.

Decay	Neutral hadron mass (GeV/c^2)	Neutral hadron decay width (MeV/c^2)	Number from MC	Branching fraction (PDG)	Branching fraction (Belle MC)	Type
$B^+ \to K^+\eta_c$	2.984	32.0	15.16	$(1.09 \pm 0.08) \times 10^{-3}$	9.0×10^{-4}	Veto
$B^+ \to K^+D^*(2007)^0$	2.007	27.98	14.08	$(4.19^{+0.31}_{-0.28}) \times 10^{-4}$	4.2×10^{-4}	Veto
$B^+ \to K^+\overline{D}^0$	1.865	0.0	13.32	$(3.64 \pm 0.15) \times 10^{-4}$	4.0×10^{-4}	Veto
$B^+ \to K^+J/\psi$	3.097	0.1	12.14	$(1.020 \pm 0.019) \times 10^{-3}$	1.0×10^{-3}	Veto
$B^+ \to K^+\chi_{c1}$	3.511	0.8	6.42	$(2.1 \pm 0.7) \times 10^{-4}$	6.8×10^{-4}	Veto
$B^+ \to K^+\psi(2S)$	3.686	0.3	3.02	$(6.24 \pm 0.20) \times 10^{-4}$	6.9×10^{-4}	Rare
$B^+ \to K^+\chi_{c0}$	3.415	10.8	2.90	$(1.51 \pm 0.15) \times 10^{-4}$	6.0×10^{-4}	Rare
$B^+ \to K^+\eta_c(2S)$	3.638	11.3	2.72	$(4.4 \pm 1.0) \times 10^{-4}$	3.6×10^{-4}	Rare
$B^+ \to K^+K^0$	0.498	0.0	2.72	$(1.31 \pm 0.17) \times 10^{-6}$	1.36×10^{-6}	Rare
$B^+ \to K^+\pi^0$	0.135	0.0	2.22	$(1.29 \pm 0.05) \times 10^{-5}$	1.29×10^{-5}	Rare
$B^+ \to K^+f_2'(1525)$	1.517	86.0	2.14	$(1.8 \pm 0.5) \times 10^{-6}$	1.0×10^{-7}	Rare
$B^+ \to K^+f_0(1370)$	1.370	500.0	1.96	$< 1.07 \times 10^{-5}$	2.0×10^{-6}	Rare
$B^+ \to K^+\phi$	1.019	4.2	1.62	$(8.8^{+0.7}_{-0.6}) \times 10^{-6}$	1.5×10^{-3}	Rare
$B^+ \to K^+\eta'$	0.958	0.2	1.36	$(7.04 \pm 0.25) \times 10^{-5}$	7.02×10^{-5}	Rare
$B^+ \to K^+f_0(1500)$	1.506	112.0	1.32	Unknown	4.40×10^{-6}	Rare
$B^+ \to K^+f_0(1710)$	1.704	123.0	1.00	$(4.8^{+4.0}_{-2.6}) \times 10^{-7}$	3.2×10^{-7}	Rare
$B^+ \to K^+h_c$	3.525	0.7	0.96	$< 3.4 \times 10^{-6}$	2.0×10^{-4}	Rare
$B^+ \to K^+a_1(1260)^0$	1.230	425.0	0.90	Unknown	2.7×10^{-5}	Rare
$B^+ \to K^+\chi_{c2}$	3.556	1.97	0.68	$(1.1 \pm 0.4) \times 10^{-5}$	8.4×10^{-5}	Rare
$B^+ \to K^+K^*(892)^0$	0.896	47.3	0.60	$(5.9 \pm 0.8) \times 10^{-7}$	1.1×10^{-6}	Rare
$B^+ \to K^+a_0(1450)^0$	1.474	265.0	0.52	Unknown	2.5×10^{-6}	Rare
$B^+ \to K^+\eta$	0.548	1×10^{-3}	0.50	$(2.4 \pm 0.4) \times 10^{-6}$	1.5×10^{-3}	Rare
$B^+ \to K^+\eta(1405)$	1.409	50.1	0.36	$< 1.2 \times 10^{-6}$	3.0×10^{-6}	Rare
$B^+ \to K^+\phi(1680)$	1.680	150.0	0.30	$< 8 \times 10^{-7}$	8×10^{-7}	Rare
$B^+ \to K^+\omega(782)$	0.783	8.49	0.22	$(6 \pm 9) \times 10^{-6}$	6.7×10^{-6}	Rare
$B^+ \to K^+f_1(1285)$	1.282	22.7	0.20	$< 2.0 \times 10^{-6}$	2.0×10^{-6}	Rare
$B^+ \to K^+X(3872)$	3.871	0.0	0.14	Unknown	1.4×10^{-5}	Rare
$B^+ \to K^+\rho(1450)^0$	1.465	400.0	0.12	$< 1.17 \times 10^{-5}$	1.17×10^{-5}	Rare
$B^+ \to K^+\eta(1295)$	1.294	55.0	0.12	$(2.9 \pm 0.8) \times 10^{-6}$	3.0×10^{-6}	Rare
$B^+ \to K^+\rho(770)^0$	0.775	147.8	0.10	$(3.7 \pm 0.5) \times 10^{-6}$	4.25×10^{-6}	Rare
$B^+ \to K^+f_2(1270)$	1.276	186.7	0.08	$(1.07 \pm 0.27) \times 10^{-6}$	1.33×10^{-6}	Rare
$B^+ \to K^+a_2(1320)^0$	1.319	105.0	0.08	$< 1.1 \times 10^{-6}$	1.10×10^{-6}	Rare

Table 4.13: Two-body decays present in the background for the $B^{\pm} \to pa$ search, listed in order of most frequently appearing in the MC background after selection.

Decay	Neutral hadron mass (GeV/c^2)	Neutral hadron decay width (MeV/c^2)	Number from MC	Branching fraction (PDG)	Branching fraction (Belle MC)	Type
$B^+ \to p\Lambda(1520)^0$	1.519	16.0	1.26	$(3.1 \pm 0.6) \times 10^{-7}$	1.5×10^{-6}	Rare
$B^+ \to pn$	0.934	0.0	0.04	Unknown	3.0×10^{-7}	Rare

Table 4.14: Two-body decays present in the background for the $B^{\pm} \to D_s^{\pm} a$ search, listed in order of most frequently appearing in the MC background after selection.

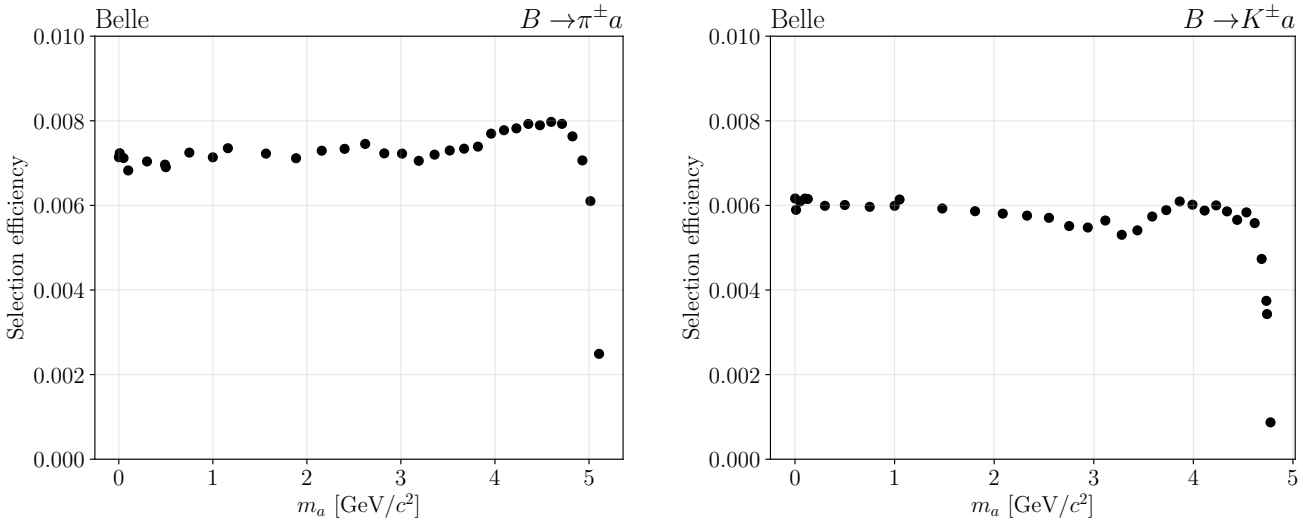
Decay	Neutral hadron mass (GeV/c^2)	Neutral hadron decay width (MeV/c^2)	Number from MC	Branching fraction (PDG)	Branching fraction (Belle MC)	Type
$B^+ \to D_s^+ \overline{D}^0$	1.865	0.0	2.72	$(9.0 \pm 0.9) \times 10^{-3}$	1.0×10^{-2}	Veto
$B^+ \to D_s^+ \overline{D}^*(2007)^0$	2.007	0.0	1.50	$(7.6 \pm 1.6) \times 10^{-3}$	8.2×10^{-3}	Veto
$B^+ \to D_s^+ D_2^*(2460)^0$	2.461	47.5	0.12	Unknown	1.81×10^{-3}	Rare
$B^+ \to D_s^+ D_1(2420)^0$	2.421	31.7	0.02	Unknown	3.4×10^{-4}	Rare
$B^+ \to D_s^+ \pi^0$	0.135	0.0	0.04	$(1.6 \pm 0.5) \times 10^{-5}$	1.08×10^{-5}	Rare
$B^+ \to D_s^+ \rho^0$	0.775	147.8	0.04	$< 3 \times 10^{-4}$	2.84×10^{-5}	Rare
$B^+ \to D_s^+ \omega$	0.783	8.49	0.02	$< 4 \times 10^{-4}$	2.16×10^{-5}	Rare
$B^+ \to D_s^+ a_1(1260)^0$	1.230	425.0	0.02	$< 1.8 \times 10^{-3}$	4.72×10^{-5}	Rare
$B^+ \to D_s^+ \eta'$	0.958	0.188	0.02	Unknown	7.2×10^{-6}	Rare
$B^+ \to D_s^+ \eta$	0.548	0.0013	0.02	$< 1.4 \times 10^{-5}$	8×10^{-6}	Rare

Table 4.15: Two-body decays present in the background for the $B^0 \to \overline{D}^0 a$ search, listed in order of most frequently appearing in the MC background after selection.

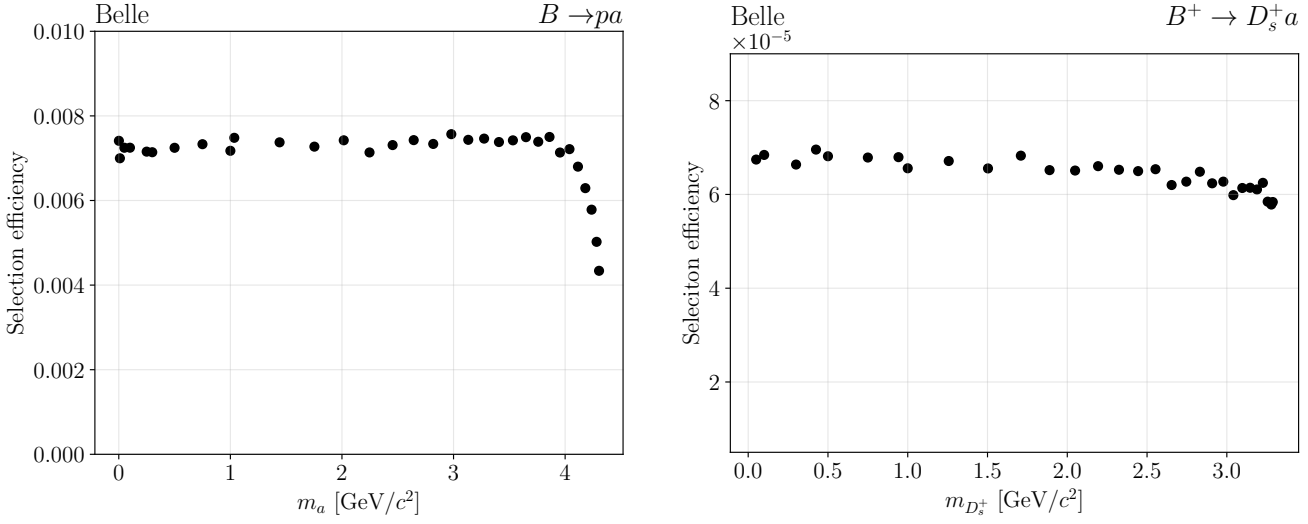
Decay	Neutral hadron mass (GeV/c^2)	Neutral hadron decay width (MeV/c^2)	Number from MC	Branching fraction (PDG)	Branching fraction (Belle MC)	Type
$B^0 \to \overline{D}^0 \rho^0$	0.775	147.8	0.40	$< 5.1 \times 10^{-4}$	2.9×10^{-4}	Rare
$B^0 \to \overline{D}^0 D^*(2007)^0$	2.007	0.0	0.06	$< 2.9 \times 10^{-4}$	2.4×10^{-5}	Rare

4.6 Selection efficiency

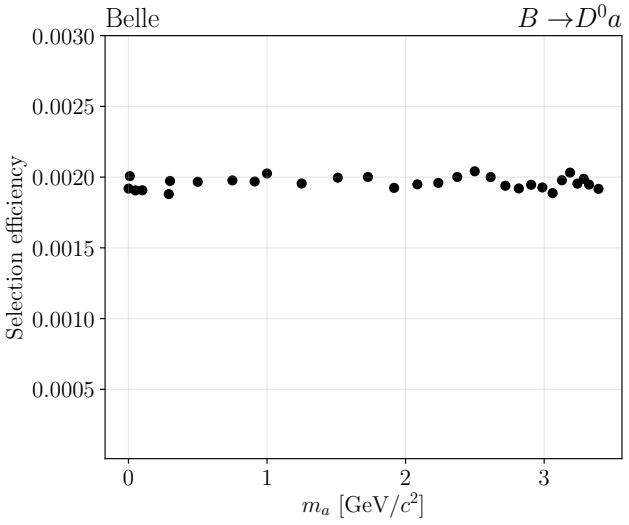
The signal selection efficiency is defined as the fraction of generated signal events that pass all of the selection criteria described in the sections above, including best candidate selection. Generally, reconstruction with the FEI leads to $\mathcal{O}(1\%)$ efficiency. The selection efficiency as a function of m_a is shown in Fig. 4.15 for all channels. For the three channels with a single track on the signal side ($h \in \pi, K, p$), the efficiency is around 0.7%. The channels with $D_s^{\pm}$ and D^0 on the signal side have lower efficiencies as only a subset of those mesons' decays are reconstructed. Each data point in Fig. 4.15 corresponds to a signal MC sample of 1 million events, so several thousand events pass the selection criteria. The uncertainty on each of these points is purely statistical, and hovers around the percent level. The selection efficiency drops off dramatically as m_a approaches $m_B - m_h$ because the momentum of the signal-side hadron is very low in this region. Later in the analysis, when the signal channel branching fraction is extracted, the efficiency is required as an input. The branching fraction extraction is performed for many more masses than those generated in the samples shown in Fig. 4.15, so an interpolation is done between these mass points so that we can find a value for any arbitrary mass. As the efficiency is mostly flat across phase space, the interpolated efficiency should not substantially deviate from the true efficiency.



(a) (b)



(c) (d)



(e)

Figure 4.15: Reconstruction efficiencies for all five channels as a function of m_a.

4.7 Sensitivity to long-lived particles

In most cases present in the literature, new particles could decay to visible final states. For example, Ref. [54] describes the scenario with hadronically decaying ALPs shown in Fig. 4.16, while Ref. [74] focuses on the leptonic decays of dark scalars. Furthermore, the total decay width and therefore lifetime of many of these new particles is somewhat free, often depending on an unconstrained coupling [173]. There has been some interest in probing these decays at Belle II [174]. Therefore, experimental results should specify the range of lifetimes for which they are valid and the impact of different assumed lifetimes.

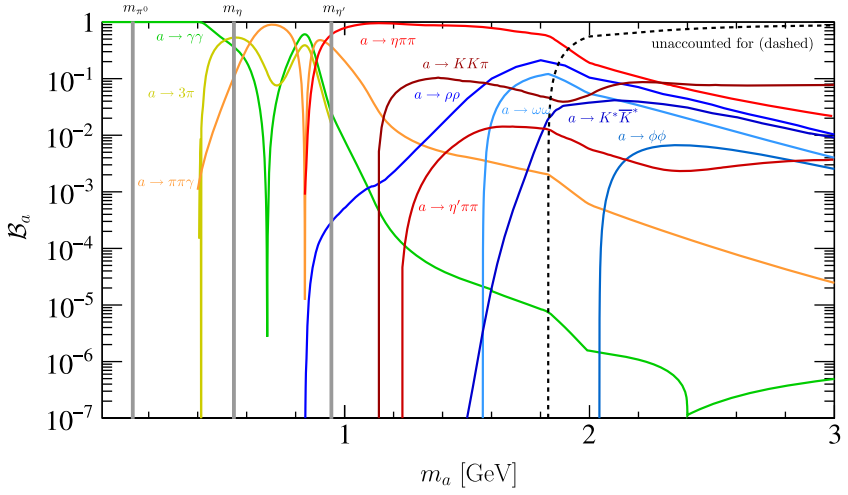
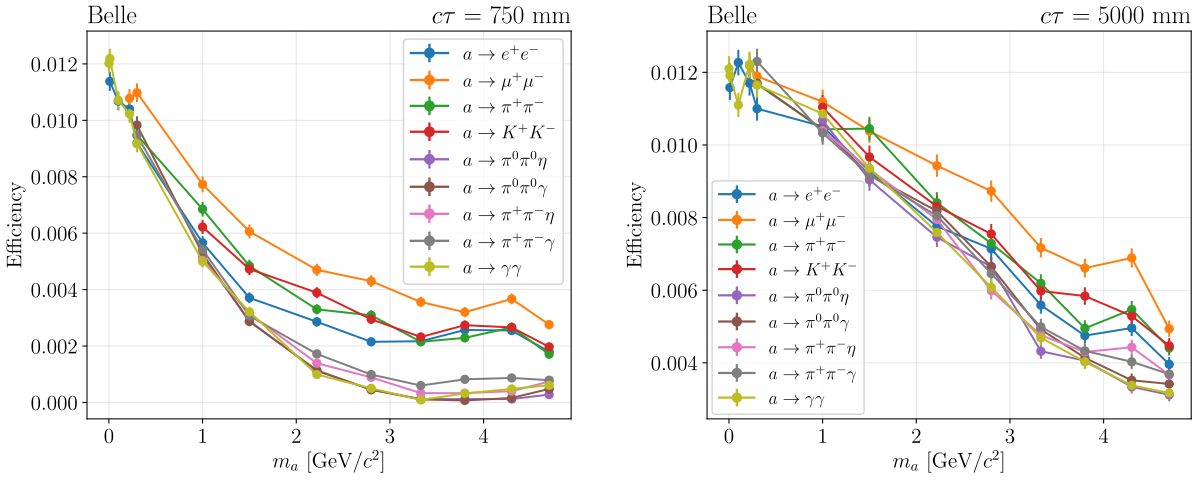


Figure 4.16: Predicted ALP branching fractions, assuming ALPs have sizeable couplings to gluons, from Ref. [54].

While the detection efficiencies of long-lived particles are often projected using some simple kinematics and knowledge of detector dimensions, it is more accurate to use Monte Carlo events with full detector simulation. Unfortunately, this is a computationally taxing approach — simply generating the 10 ALP final states shown in Fig. 4.16 increases the required CPU time and storage requirement of the MC generation by an order of magnitude. To simplify this problem, we checked the impact of the specific final state on the invisible particle search's sensitivity to long-lived particles. For a few masses and lifetimes, we generated 250,000 $B^+ \to K^+ a, a \to xx(\gamma)$ events where $xx(\gamma) \in (\gamma\gamma, e^+e^-, \mu^+\mu^-, \pi^+\pi^-, K^+K^-, \pi^0\pi^0\eta, \pi^0\pi^0\gamma, \pi^+\pi^-\eta, \pi^+\pi^-\gamma)$, and then performed the $B \to ha, a \to$ invisible reconstruction. We then compared the efficiency between different final states, and found that the diphoton final state generally had the lowest reconstruction efficiency. There were some samples where other final states had lower efficiency for a given mass; however, they were always within statistical uncertainty of the diphoton final state. The final states that were closest in efficiency to the $a \to \gamma\gamma$ also contained at least one photon, while the final states that were exclusively charged particles were more likely to make it past the selection criteria. This is because photons only need to hit the ECL to be detected, whereas a track produced in the middle of the

CDC has a lower chance of being reconstructed compared to a prompt track, despite being produced well before the ECL.



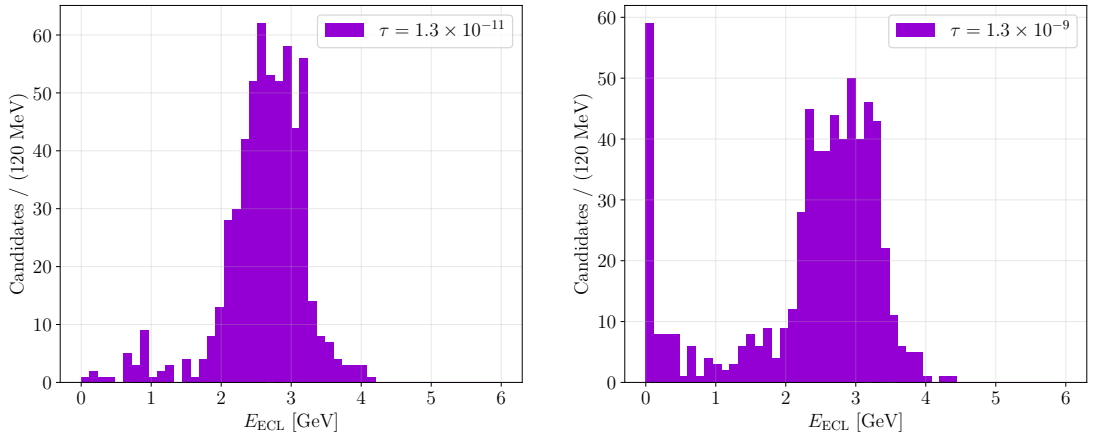
(a) $c\tau = 750$ mm (b) $c\tau = 5000$ mm

Figure 4.17: Reconstruction efficiencies for several different visible final states for two benchmark lifetimes.

About 95% of the photon daughters of a in the LLP MC samples are produced in the ECL acceptance region. Photon detection efficiency at Belle for photons with energy greater than 30 MeV is about 90%, as the almost 4π solid angle calorimeter has gaps around the beampipe and between the barrel and endcaps. Below 30 MeV, the energy resolution is very poor due to a dominant contribution of electronics noise. Across the generated values of m_a, at least 97% of photons produced in $B \to Ka, a \to \gamma\gamma$ have energy greater than 30 MeV, and this number approaches 100% for decays of a where m_a is greater than about 750 MeV/c^2. Thus, for most samples, photon detection is primarily limited by the ECL acceptance. However, because of this limited efficiency, the reconstruction efficiency of visible decays to photons should only reduce to $\mathcal{O}(1\%)$ of the efficiency of the invisible search shown in Fig. 4.15, as each photon in the pair has an approximate 10% chance of going undetected.

Long-lived particles may decay within the ECL itself, reducing the photon detection efficiency. However, as the ECL crystals are 16 radiation lengths long, most of the energy is deposited in the first few centimetres. This means that an invisible particle could decay several radiation lengths into an ECL crystal without substantially altering the detected photon energy. Furthermore, due to the spread of lab-frame momenta, particles with the same lifetime can have very different decay lengths spread over values much greater than the few centimetres required to achieve an ECL energy deposit. Because of these two factors, for a given finite lifetime, the extra ECL energy can take three different forms. For a particle whose lifetime is sufficiently large so that it always

decays outside of the detector volume, E_{ECL} will only peak at zero, much like the signal MC distributions shown in Figures 4.1 to 4.5. For very short-lived particles, the photon daughters will almost always hit the ECL, and so there are few or no events with no ECL energy deposits. This results in E_{ECL} distributions like the plot on the right in Fig. 4.18. When the lifetime is somewhere around the scale of the detector, the long-lived particle can decay before, while in, or after passing through the ECL. This results in a bimodal distribution such as the one shown on the left in Fig. 4.18. The key point is that complications arising from particles decaying inside the ECL crystals and having only partial energy detection are rare. In general, simulations show that long-lived particles decaying into two photons either deposit substantial energy in the ECL ($\mathcal{O}(1\,\text{GeV})$) or very little to no energy ($< \mathcal{O}(100\,\text{MeV})$).



(a) (b)

Figure 4.18: E_{ECL} distributions for a long-lived particle with $m_a = 1$ GeV/c^2 produced in the channel $B \to Ka, a \to \gamma\gamma$ for two different lifetimes; (a) $\tau_a = 13$ ps, (b) $\tau_a = 1300$ ps.

As $a \to \gamma\gamma$ is the final state most likely to interact with the detector, it is also the one most likely to be rejected by the selection criteria optimised to search for invisible final states. We therefore base our sensitivity estimates for any visible final state on the sensitivity to $a \to \gamma\gamma$ decays, which makes the estimate inherently conservative. This allows us to bypass the issue of generating an extremely large MC sample containing many different final states.

Long-lived particle MC is generated for 30 mass hypotheses and 30 lifetimes for 5 channels. For each unique combination, 250,000 events are generated. The total sample size is 1.125 billion events. As the reconstruction efficiency is at the sub-percent level, such a sample is required so that at least $\mathcal{O}(1000)$ events pass the selection criteria in the best-case scenario lifetimes. Even so, statistical fluctuation means that for some samples, the selection efficiency is higher than the selection efficiency on the invisible a sample as shown in Fig. 4.17. To avoid overstating the sensitivity to finite-lifetime

particles, we replace the efficiency in such cases with the reconstruction efficiency from the invisible sample. Furthermore, as lifetime decreases, efficiency should decrease, and so if for some mass and lifetime hypothesis the efficiency is greater than for the next largest generated lifetime, we set them to have the same efficiency when we parameterise the efficiency as a function of mass for different lifetimes. This means that the efficiency contours for shorter lifetimes such as those shown in Fig. 4.19 will never exceed that of samples with longer lifetimes at the same mass. This is particularly important when lifetimes get very short, as the statistical uncertainty in that region is significant due to the very low efficiency.

The selection efficiency for a long-lived particle decaying to two photons is compared to the efficiency for invisible particles in Fig. 4.19 for all channels. As lighter particles are more boosted in the detector, their decay products are more likely to be produced outside of the detector volume, and have an invisible signature. This can be seen for all channels in Fig. 4.19 where for a given lifetime, the efficiency is always highest at the lowest generated mass. The lowest generated mass for all long-lived particle samples is 100 keV/c^2. While many models introduce FIPs that can be lighter than that, lower masses are often very highly constrained by other sources, such as the astrophysical bounds on ALPs. To set limits on decays of the form $B \to ha, a \to$ visible where a has some finite lifetime, the nominal analysis efficiency in Fig. 4.15 is replaced with the efficiency in Fig. 4.19.

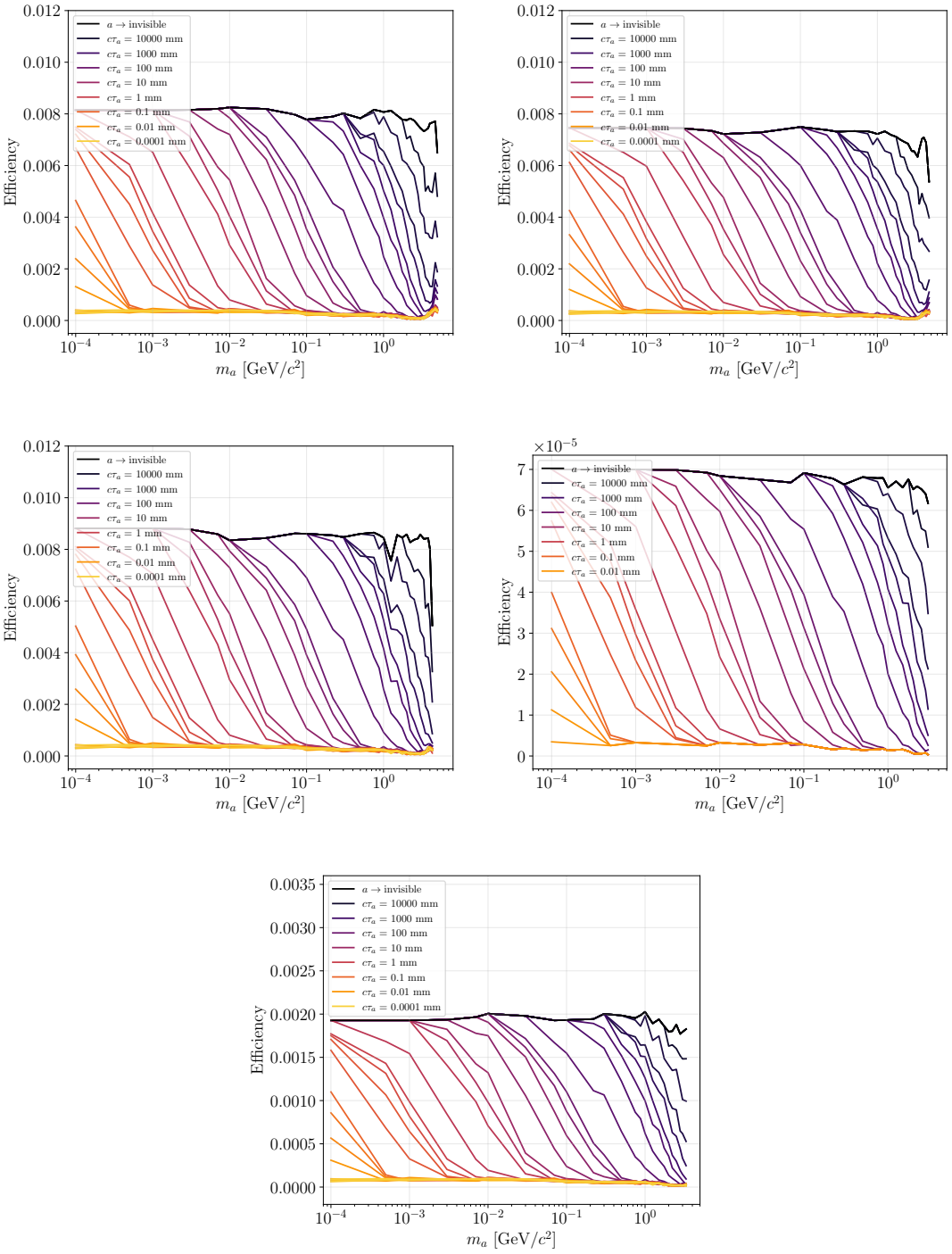


Figure 4.19: Reconstruction efficiencies for all channels for a range of lifetimes.

5

Statistical analysis

Various advanced statistical and computational techniques are utilised to extract meaningful insights from the datasets outlined earlier. This chapter begins by providing an overview of these methods, followed by a detailed description of the specific approach employed to search for new physics.

5.1 Statistical methods in high-energy physics

The search for a recoiling invisible particle will be carried out by searching for a peak in the distribution of p_h^S. Loosely speaking, the size of this peak determines the signal strength. In this chapter, we will introduce the concepts of parameter estimation, hypothesis testing, and limit setting in the context of the physics analysis, and describe how these are used in the search for invisible particles.

5.1.1 Parameter estimation

The aim of most experimental particle physics studies is to extract a value for a parameter of interest. For instance, this could be the number of events consistent with the decay $B^{\pm} \to K^{\pm}a$ in the distribution of some variable in a dataset. This estimation is done using maximum likelihood estimation, where a theoretical model distribution that depends on the parameter of interest is fit to the observed data. For a given probability density function (PDF) of observables $\vec{x}$ and parameters ω, $f(\vec{x}|\vec{\omega})$, and some observed data point $\vec{x_i}$, the likelihood of the PDF can be calculated as

$$\mathcal{L}(\vec{\omega}) = f(\vec{x}_i|\vec{\omega}). \tag{5.1}$$

When there are many independent events in the dataset, the likelihood becomes a product of the individual likelihoods of the single events:

$$\mathcal{L}(\vec{\omega}) = \prod_i f(\vec{x}_i|\vec{\omega}). \tag{5.2}$$

For an analysis such as a search for an excess of particles, or for measuring a quantity of a certain type of decay, the parameter of interest will depend on the number of observed events in some way. One can extend the concept in the equation above to include the overall normalisation of the fit, known as an extended maximum likelihood function [175]. For n following a Poisson distribution with a mean of ν, the extended likelihood becomes

$$\mathcal{L}(\vec{\omega}) = \frac{\nu^n}{n!}\exp(-\nu)\prod_i f(\vec{x}_i|\vec{\omega}). \tag{5.3}$$

When this likelihood is maximised, then the best parameters $\hat{\vec{\omega}}$ have been found. In practice, and as is done in this work, the problem is often solved by minimising twice the negative log-likelihood (NLL).[1]

Both the statistical and systematic uncertainties of fitted parameters of interest can be estimated. Statistical uncertainties can be estimated using the covariance of the fitted parameters (in the case of multiple parameters), or by looking at the variation in the profile likelihood (defined in Equation (5.5)) as the parameter floats about its fitted value. Systematic uncertainties can be incorporated directly into the fit through the use of nuisance parameters. These change the likelihood so that it becomes

$$\mathcal{L}(\,\vec{\omega},\vec{\theta}\,) = f(\,\vec{x}\,|\,\vec{\omega},\vec{\theta}\,)\cdot c(\,\vec{y}|\,\vec{\theta}\,) \tag{5.4}$$

where $\vec{\theta}$ is the vector of nuisance parameters, $\vec{y}$ is the vector of nuisance parameter value estimates, and $c(\,\vec{y}|\vec{\theta}\,)$ is a (typically Gaussian) constraint function. These typically constrain a source of systematic uncertainty, such as an efficiency correction, to the error on its estimation. Profiling over the set of nuisance parameters yields the profile likelihood of a parameter of interest (or set of parameters)

$$\mathcal{L}(\,\vec{\omega},\hat{\vec{\theta}}\,) = \max_{\vec{\theta}}\ \mathcal{L}(\,\vec{\omega},\vec{\theta}\,). \tag{5.5}$$

[1] There are several reasons to minimise the NLL rather than simply maximising the likelihood. As the likelihoods are often very small numbers, performing calculations with them directly can lead to numerical instability. It also connects directly to well-known statistical tests. For large samples, $-2\log\mathcal{L}$ follows a χ^2 distribution. Furthermore, many optimisation frameworks are designed for minimisation, so converting the maximisation problem to a minimisation problem is advantageous.

5.1.2 Hypothesis testing and limit setting

All searches for new particles are really just tests of competing hypotheses — does the data contain a new particle or not? Therefore, searches are naturally about hypothesis testing, a statistical method used to infer whether to accept or reject some hypothesis.

It is natural to assign the SM expectation as the null hypothesis, and any variation on it as the alternative hypothesis. We label the null hypothesis H_0: no new particle exists (i.e. the data is *background-only*), and an alternative hypothesis H_1: a new particle exists (i.e. the data has *signal* + *background* components).

In hypothesis testing, a p value is used to determine the degree to which the observed data and the hypotheses agree via a frequentist significance test. The p value is given by

$$p = \int_{t_{\text{obs}}}^{\infty} f(t|H)dt, \tag{5.6}$$

where t is a test statistic that encodes the agreement between the data and the hypothesis. For low p-values, the hypothesis is ruled out.

We can directly calculate the significance of an excess in events using the difference in the negative log-likelihood from fits with signal + background and background-only hypotheses, giving significance Z

$$Z = \sqrt{2(\text{NLL}_{\text{s+b}} - \text{NLL}_{\text{b}})}. \tag{5.7}$$

In particle physics, a discovery threshold of 5σ is commonly used, which corresponds to a p value of about 2.8×10^{-7} (this is discussed more extensively in Section 5.1.3). In the absence of a significant signal, one instead aims to set an upper limit on the signal strength parameter (in the case of this search, the number of signal events) using the CL_{s} method [176]. The CL_{s} value is defined using the signal + background and background-only p-values

$$\text{CL}_s = \frac{p_{\text{s+b}}}{1 - p_{\text{b}}}. \tag{5.8}$$

In a nutshell, this method applies a correction to the p_{s+b} value of $1/(1-p_b)$, which makes the statistical test robust against significant downward fluctuations in background yields, which could otherwise cause the test to exclude signal strengths beyond its sensitivity.

5.1.3 Significance and the look-elsewhere effect

In general, particle physics has a very strict threshold for what is considered a discovery. While many fields might state a discovery with p-values at the percent level, the discovery threshold in particle physics is a p-value of about 2.8×10^{-7}. To facilitate easier discussion

and more readable papers, the field generally converts this to a significance in units of σ. For a normalised Gaussian with width σ, the significance is the multiple of σ away from the centre of the distribution that leaves an area equal to the p-value on one side. This area is given by the definite integral between the σ threshold and infinity. So for a Gaussian centred at zero with a width of 1, the area bounded from the left at 5σ is

$$\int_5^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 2.8 \times 10^{-7}, \tag{5.9}$$

which is the previously mentioned discovery threshold p-value. This is why the discovery threshold is more commonly referred to as 5σ, rather than *a p-value of three by ten to the negative seven.* In the past, many seemingly observed phenomena that corresponded to 3 or 4σ effects disappeared after the collection of more data [177–179], and so 5σ has been the standard in the field for several decades now. However, this unified statistical treatment of so-far undiscovered phenomena is somewhat reductionist; some suggest that the threshold should be lower for less surprising phenomena [180]. For instance, discoveries that are fundamentally inconsistent with our modern understanding of physics, such as faster-than-light neutrinos, should be held to a higher discovery significance threshold than phenomena predicted by the SM, such as observing $B_s^0 \to \mu^+\mu^-$ decays. Furthermore, the dataset size and the treatment (or non-treatment) of systematic uncertainties by the analyst can have a large effect on the significance [181, 182].

We derive the local significance of a new physics signal based on a fit at a fixed mass point, but as our signal can appear at any kinematically allowed mass, and our signal peak is much narrower than the entire range, an excess of events could occur at any place within that range. This means that rather than merely accounting for the chance that any excess or rejection of the null hypothesis occurred due to statistical fluctuations in the background at a specific mass, we need to consider the fact that this kind of fluctuation could happen at any masses that we search for new particles in. This is known as the *look-elsewhere effect*, or LEE [183, 184]. Put plainly, there is a greater chance of seeing an upwards fluctuation that looks like a signal the more places you look for a signal.

The most conservative method [185] to account for the LEE is using a trial factor method [186] where one converts a local p value to a global p value. This accounts for the probability that a significant result will occur due to upwards fluctuations in background given the number of search points. The global p-value is given by

$$p_{\text{global}} = p_{\text{local}} \cdot N_{\text{trial}}, \tag{5.10}$$

where the trial factor is

$$N_{\text{trial}} = 1 + \frac{1}{p_{\text{local}}} \langle n_{\text{up}}(Z_{\text{test}}) \rangle \exp\left(\frac{Z_{\text{test}}^2 - Z_{\text{local}}^2}{2}\right), \tag{5.11}$$

where $\langle n_{\text{up}}(Z_{\text{test}}) \rangle$ is the mean number of upcrossings past a test significance Z_{test} one would expect from the predicted background. Here, an upcrossing is defined as when the significance crosses the test significance with a positive gradient. Therefore, a single maximum in significance above Z_{test} will result in 1 upcrossing. By resampling the background-only MC, this mean number of upcrossings can be estimated. In this study, we choose $Z_{\text{test}} = 1$ when calculating the mean number of upcrossings. The average number of upcrossings should be greater for searches with more scan points, so we would expect a greater difference between the global and local p-values for channels with a larger region of accessible phase space, all other things being equal. The global p value can be converted into a significance σ_{global}, and for a new particle search such as this to claim the 5σ *discovery threshold*, we would have to reach $5\sigma_{\text{global}}$ to account for the LEE.

5.1.4 Implementation of methods

Most of the statistical work presented in this thesis is carried out with the `ROOT` software package [187], with limit setting done within the `RooStats` framework [188, 189]. Some more basic regression analysis, such as the fits to linearity test results shown in Fig. 5.10, is done using `SciPy` [190].

All events, real and simulated, are stored as `.mdst` files. The `basf2` scripts used to reconstruct events read these files and write their outputs as `.root` ntuples. To save storage space and facilitate easier interfacing with Python, the ntuples are converted to Parquet files [191] and 64-bit floats are downcast to 32-bit floats. Data manipulation, including resampling, creating selection criteria, and plotting, is done with commonly used Python packages `numpy` [192], `pandas` [193, 194], and `matplotlib` [195].

5.2 Fit method

As we are testing for the presence of new particles, we perform an extended maximum likelihood fit where the parameter of interest is the yield of the PDF that describes the expected shape of signal events, or *signal PDF*. The SM component is described by a *background PDF*.

5.2.1 Signal PDF

As previously discussed, the signal events will cause a peak in the distribution of p_h^S depending on m_a. While the width of the hypothetical invisible particle is negligible compared to the detector resolution, the detector's finite resolution will smear out the peak, leading to a wider shape. The width is momentum dependent. We find that choosing the sum of two Gaussians to fit the signal peak is appropriate. While a more complex functional form could better match the simulated signal distribution, the exact peaking shape is less important in the final fit. This is because we expect the signal process to be only a handful of events, and therefore the exact shape of the PDF to be less important. Finding a simple parameterisation for the peaking structure is a suitable choice. The signal PDF can be written as

$$\mathrm{PDF}_{\mathrm{sig}}(x) = \frac{f}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_1}\right)^2} + \frac{(1-f)}{\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_2}\right)^2} \tag{5.12}$$

where μ is the common mean that both Gaussians have, σ_1 and σ_2 are the widths of the two Gaussians respectively, and f encodes the relative contribution of each Gaussian to the total PDF. From these parameters, we also define an effective resolution σ_{eff} that is the weighted sum of the individual resolutions in quadrature

$$\sigma_{\mathrm{eff}} = \sqrt{f\sigma_1^2 + (1-f)\sigma_2^2}. \tag{5.13}$$

This parameter is used to define the number of steps in the mass scan.

A selection of signal PDF fits is shown in Fig. 5.1. In general, the resolution, defined as a weighted sum of the two component Gaussians' widths, ranges from near 10 MeV/c to about 40 MeV/c for all channels. The three channels with single final-state particles have poorer resolution at high momentum but better resolution at low momentum than the $D_s^{\pm}$ and $\overline{D}^0$ channels. A number of different signal models were considered in the early phases of this work, including triple Gaussians, bifurcated Gaussians, and Johnson's S_U distribution [196]. While some of these PDFs did actually model the signal MC samples more accurately, the increase in fit parameters complicated the fit model. If a signal is present in the data, we expect its yield to be less than 100 events, as any processes more common than that should have been found by past searches. For a statistically limited signal size, the added complexity of alternative signal PDF options does not yield a better fit result.

There was some concern that when m_a was close to its maximum allowed value for a given channel, the signal peak would become distorted. This would only impact the analysis if distortion occurred at p_h^S values greater than about 0.2 GeV/c, as this is the lower limit on our background PDF definition (and in the lab frame, the lowest allowed momentum for signal-side tracks). The highest mass we search for is the highest mass

for which at least 95% of the area under the curve lies above this 0.2 GeV/c cutoff (i.e. the mean of the PDF is at least 2σ higher than 0.2 GeV/c). The highest masses we search for in each channel are shown in Table 5.1. No meaningful deviation in fit quality is observed between the highest mass point for each channel compared to the variation between other mass points.

Table 5.1: Highest mass hypotheses used in the search for each channel.

Channel	Maximum mass (MeV/c^2)
$B^{\pm} \to \pi^{\pm} a$	5015
$B^{\pm} \to K^{\pm} a$	4733
$B^{\pm} \to p a$	4304
$B^{\pm} \to D_s^{\pm} a$	3288
$B^0 \to \overline{D}^0 a$	3392

5.2.1.1 Parameterisation of fit parameters

For each channel in the analysis, all 32 signal MC samples are fitted. However, it is important to know how the fit parameters depend on the mass of the sample so that a well-defined signal PDF can be used for arbitrary values of m_a. To that end, the fit parameters σ_1, σ_2, and f as described in Equation (5.12) are parameterised as a function of the mean μ of the Gaussians in Equation (5.12). The effective resolution σ_{eff} is also parameterised in a similar manner, as it is used to determine the spacing between scan points in the bump hunt. All parameters are fitted with fourth-order Chebyshev polynomials. The values of these parameters from every fit, as well as their parameterisation as a function of μ, are shown in Fig. 5.2.

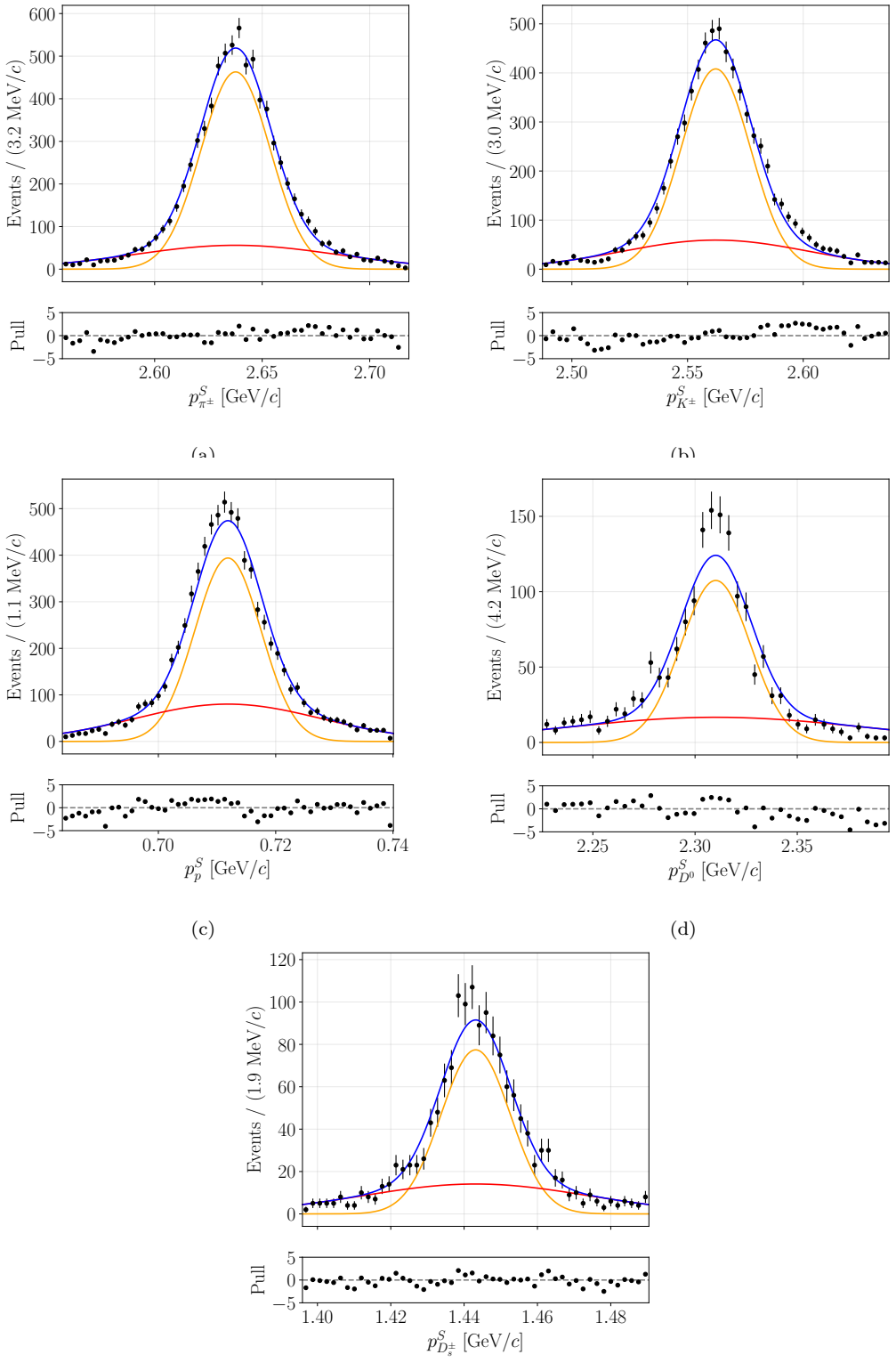
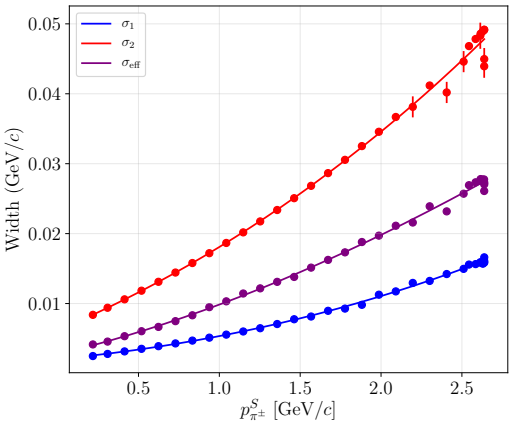
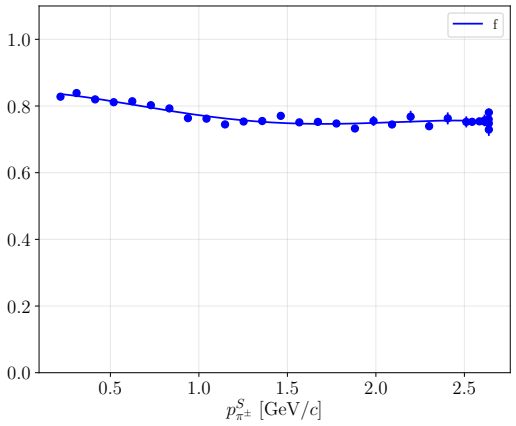


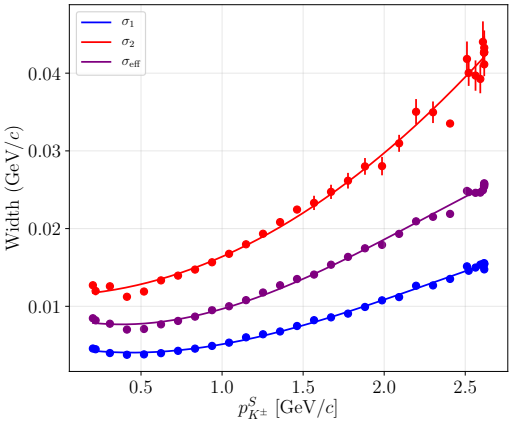
Figure 5.1: A selection of fits to p_h^S for different signal MC samples. The two component Gaussians are shown in red and orange, with their sum shown in blue.



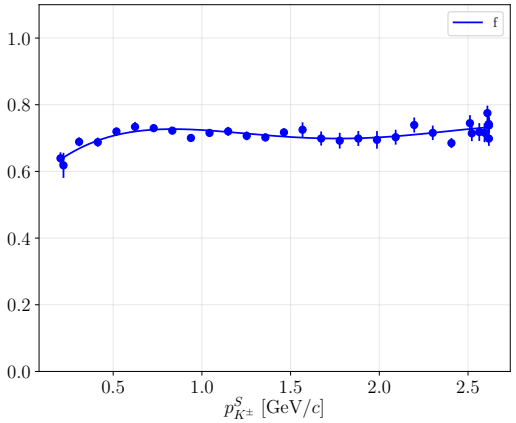
(a) Widths for $B^\pm \to \pi^\pm a$ signal fits



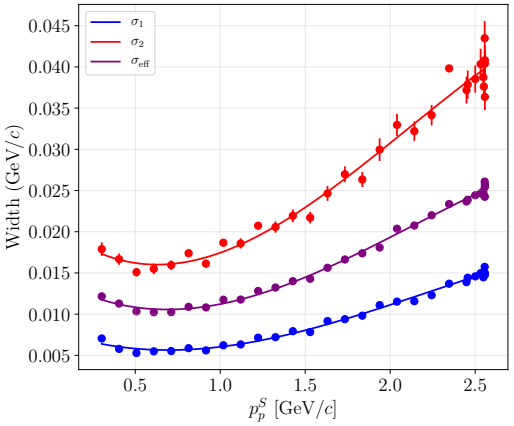
(b) f parameter for $B^\pm \to \pi^\pm a$ signal fits



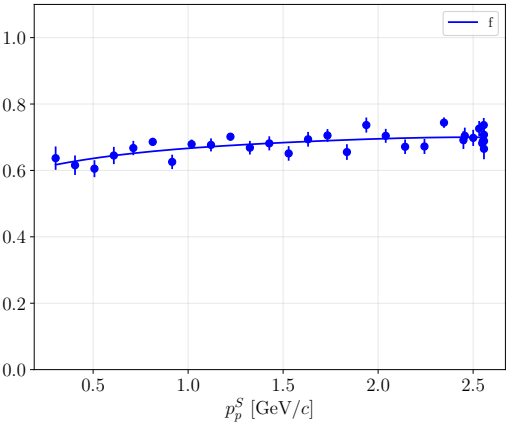
(c) Widths for $B^\pm \to K^\pm a$ signal fits



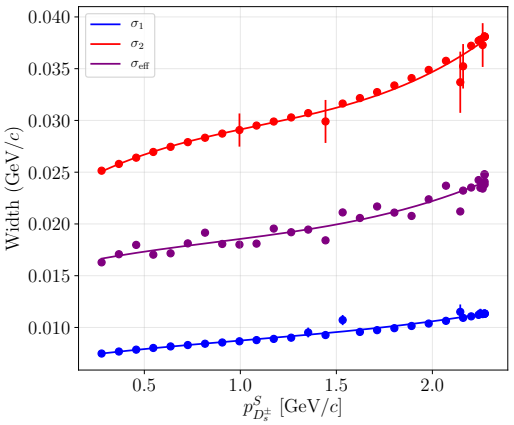
(d) f parameter for $B^\pm \to K^\pm a$ signal fits



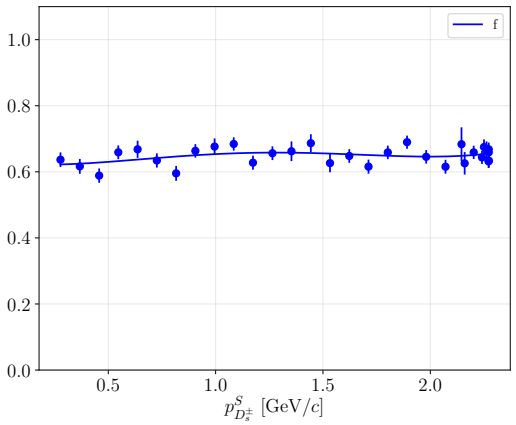
(e) Widths for $B^\pm \to pa$ signal fits



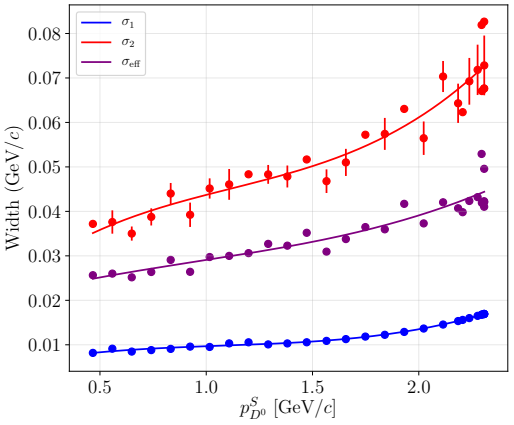
(f) f parameter for $B^\pm \to pa$ signal fits



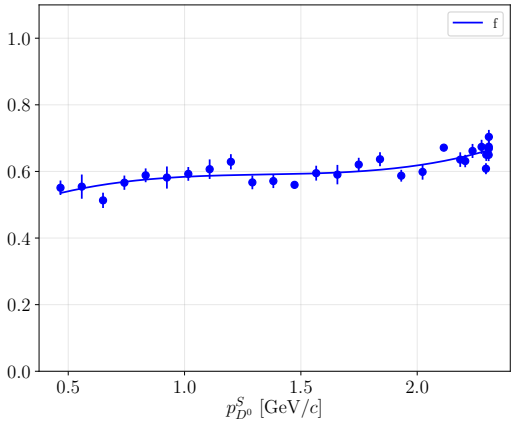
(g) Widths for $B^\pm \to D_s^\pm a$ signal fits



(h) f parameter for $B^\pm \to D_s^\pm a$ signal fits



(i) Widths for $B^0 \to \overline{D}^0 a$ signal fits

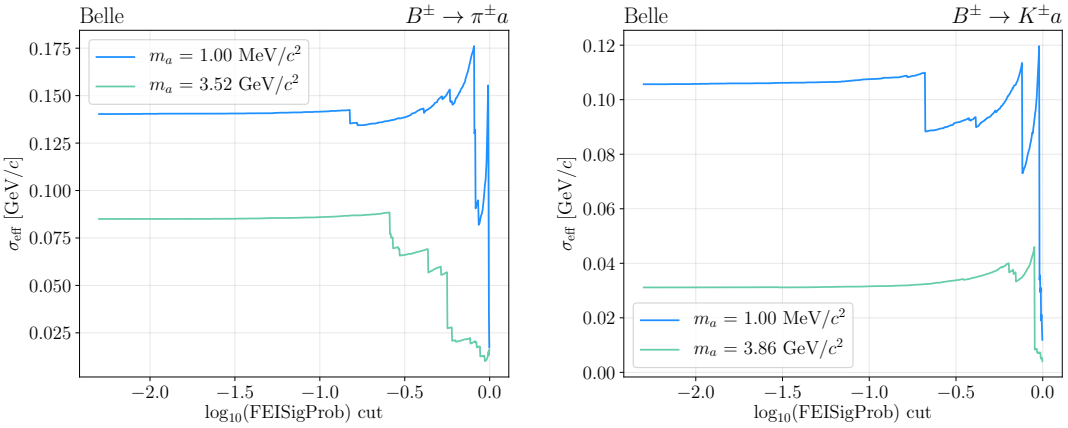


(j) f parameter for $B^0 \to \overline{D}^0 a$ signal fits

Figure 5.2: Signal fit parameters, parameterised as a function of the mean of the Gaussian μ.

5.2.1.2 Signal resolution

The resolution of the peak should depend on the quality of the kinematic information in the event. Because of this, one would expect that the resolution would improve if the events had a higher purity of correctly reconstructed B_{tag}. To test the effect of B_{tag} purity on the resolution, we performed a short study whereby the FEI signal probability selection was progressively tightened and the remaining sample was refitted. We found that unless the selection criterion on the FEI signal probability was fairly tight (tighter than $\log_{10}(\text{FEI signal probability}) > -0.9$), the effective resolution σ_{eff} only varied by a few percent. The evolution of σ_{eff} at different FEI output thresholds is shown in Fig. 5.3. Tight selection like this reduces the signal efficiency by 70–90% depending on the channel and m_a. As the branching fraction sensitivity is directly proportional to the signal selection efficiency, and the signal-to-background ratio has already been optimised, we opted to keep the selection at $\log_{10}(\text{FEI signal probability}) > -2.3$ and deal with the marginally worse signal peak resolution. Furthermore, there is some non-trivial structure in the resolution as a function of the FEI signal probability cut that can be seen in Fig. 5.3 that we do not understand. Because of this, keeping the cut in the region without structure seems to be the more reasonable approach.



(a) (b)

Figure 5.3: Dependence of the effective signal peak resolution on the FEI signal probability selection for (a) the $\pi^\pm$ channel and (b) the $K^\pm$ channel, at two benchmark masses per channel.

5.2.2 Background PDF

The composition of the background to this invisible particle search varies greatly across phase space, with continuum events contributing much of the background at lower values of p_h^S, while rare two-body B decays tend to dominate at very high values. The

most obvious features in the background come from vetoed peaking backgrounds, which are discussed in Section 4.5.6. Outside of these veto regions, the background shape is generally flat, or has some gentle curve. Rather than choosing a specific model or sets of models to describe different background curves, we choose to use a non-parametric method of defining the background PDF. We use a kernel density estimator or KDE [197] to parameterise the background shape. A KDE constructs a non-parametric PDF by adding a kernel to the PDF for each data point it is being constructed from, and in general is given by

$$\hat{f}_0(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - t_i}{h}\right) \tag{5.14}$$

where $\{t_i\}$ represents the dataset with n entries and h is the bandwidth, or smoothing parameter, which controls the width of the kernels, and K is the kernel function. The classic choice of kernel is a Gaussian function:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{5.15}$$

with $\mu = 0$ and $\sigma = 1$, as the function is defined over an infinite domain, positive definite, and infinitely differentiable. This ensures that these desirable properties are shared by the KDE function itself.

The smoothing parameter h is a trade-off between capturing the true distribution as accurately as possible and parameterising statistical fluctuations in the data. Ideally, we would minimise smoothing where the dataset is dense, and increase it for sparse datasets to smooth out statistical fluctuations. As the density of data depends on the composition of the background, this problem is certainly important to consider in the context of this analysis. In Ref. [197], a method for adaptive kernel estimation is introduced, which considers the local density of data when setting the bandwidth term for a given kernel. The adaptive bandwidth parameter for the i^{th} data entry is

$$h_i = h/\sqrt{f(t_i)}. \tag{5.16}$$

In Ref. [197], it is shown that the mean integrated squared error of $\hat{f}_0$ is minimised when

$$h^* = \left(\frac{4}{3}\right)^{1/5} \sigma n^{-1/5} \tag{5.17}$$

where h^* is the optimal bandwidth. By replacing the bandwidth h in Equation (5.14) with the adaptive bandwidth, we get the expression for an adaptive KDE

$$\hat{f}_1(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x - t_i}{h_i}\right) \tag{5.18}$$

where the optimal bandwidth for each entry i is

$$h_i^* = \rho\left(\frac{4}{3}\right)^{1/5} \sqrt{\frac{\sigma}{\hat{f}_0(t_i)}} n^{-1/5}, \tag{5.19}$$

and

$$\rho = \sqrt{\frac{\sigma_{\text{local}}}{\sigma}}. \tag{5.20}$$

ρ is typically 1 as long as the local structure of the data is similar to the standard deviation of the entire dataset. This almost entirely removes the KDE's dependence on the original choice of bandwidth. The adaptive kernel estimate $\hat{f}_1$ is essentially a second iteration of the first estimate $\hat{f}_0$.

A drawback of constructing PDFs in this way is that the local density of data is naturally lower at the edges of the range of the observable of interest. When naïvely constructing KDEs without taking this into account, the PDF will underestimate the data at the edge of the allowed range. There are a number of methods that can be used to overcome this issue, such as introducing boundary kernels or mirroring the dataset below the edge of the range. However, we have taken a simpler approach, and created the KDE across a wider range than what is required for the fit, and then performed the combined fit over a reduced range. This means we can avoid any systematic bias introduced by dealing with the boundary effects in some parametric way while still mitigating any issues that they may cause.

We consider $B\overline{B}$ and continuum background separately when defining background KDEs as we want the background modelling to be robust against situations where there are different amounts of continuum events relative to the number of $B\overline{B}$ events. Indeed, due to the difference between continuum production cross sections discussed in Section 4.3.1.2, we do not expect the ratio of B to continuum events to have perfect agreement between data and simulation. Therefore, we create two KDEs, $\text{KDE}_{B\overline{B}}$ and $\text{KDE}_{q\overline{q}}$ to describe the two background types separately.

5.2.2.1 Background PDF cross-checks

The PDF we use to describe the background distribution of p_h^S is a sum of two KDEs: one for $B\overline{B}$ and one for continuum background. The rationale behind defining two separate shapes is to have a fit that is robust against different ratios of continuum to $B\overline{B}$ background between data and MC. To test the robustness of this method, we perform fits to data with the background-only PDFs in the sideband where $1.5 < E_{\text{ECL}} < 6.0\,\text{GeV}$. The shapes of $\text{KDE}_{\text{q}\overline{\text{q}}}$ and $\text{KDE}_{\text{B}\overline{\text{B}}}$ are defined with MC, and then fitted to the real data. We define a quantity $r_{B\overline{B}}^{\text{MC}}$ as the fraction of the total number of background events in MC that we expect to come from $B\overline{B}$ events, calculated by simply counting events (not

using any fit results). We define another quantity, $r_{B\overline{B}}^{\text{data}}$, as the fraction of observed events in data that are fitted by $\text{KDE}_{\text{B}\overline{\text{B}}}$. We then define a ratio of these two quantities

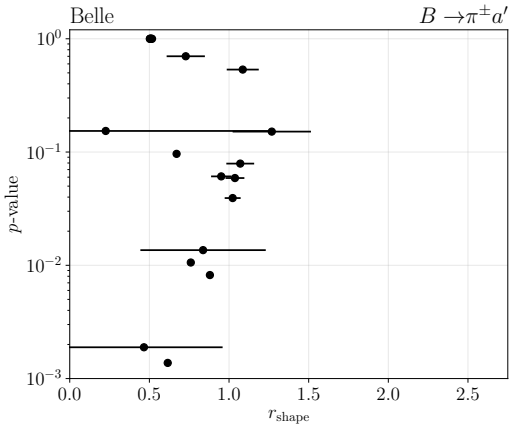
$$r_{\text{shape}} = \frac{r_{B\overline{B}}^{\text{data}}}{r_{B\overline{B}}^{\text{MC}}}. \tag{5.21}$$

If $r_{\text{shape}} = 1$, then the ratio of continuum to $B\overline{B}$ events in the MC is the same as the ratio determined by the fit to data, and therefore the overall shape of the distributions is the same. The more this value deviates from 1, the greater the difference in the shape. As we want to check that the fit is good even if the shape is different, we calculate a p-value for fits to the sideband distribution for different mass points and check its dependence on r_{shape}.

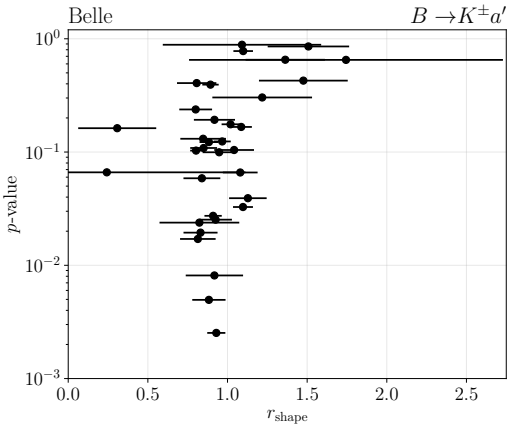
Fits are performed in regions corresponding to fit windows for 35 evenly spaced mass hypotheses (with fit windows defined as $\pm 15\sigma_{\text{eff}}$, for σ_{eff} set by the mass hypothesis). The p-value of the fit is calculated as $1 - P(\chi^2, \text{d.o.f})$ where $P(\chi^2, \text{d.o.f})$ is the cumulative distribution function of the chi-square distribution and the χ^2 is calculated from the fit distribution and the data after binning. The number of degrees of freedom (d.o.f.) is the number of bins minus 2 as there is one free fit parameter. The number of bins is usually 40, but for mass windows where the background rate is very low, the bin number is lowered until there are no empty bins.

We find that the vast majority of fits in this study result in p-values > 0.01, indicating no major discrepancies between the data and the model, regardless of the change in shape between data and MC. The poorest fits are all in regions where some peaking backgrounds are present. A summary of fit p-values with respect to the r_{shape} value for the E_{ECL} sideband defined above is shown in Fig. 5.4. The uncertainty on the r_{shape} value varies substantially due to the varying number of background events and therefore varying statistical uncertainty between mass windows.

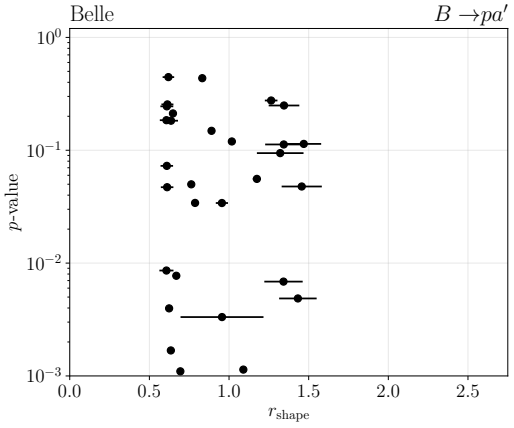
The worst fit p-values shown in Fig. 5.4 are generally in fit windows where a rare peaking background is present. The production branching fractions in the MC do not necessarily reflect the most up-to-date measurements in the PDG, as shown in Tables 4.11 to 4.15. We do not expect the peaking backgrounds present in this sideband to appear at the same rate in the signal region, and we are not testing to check how well they are modelled here either. Because of this, we do not take issue with the fact that their presence causes worse fits in this test. We have still shown that the non-peaking background components can be well modelled with our fit approach, and the method is robust against higher- or lower-than-expected continuum background rates. The fits corresponding to the three lowest p-values are shown in Fig. 5.5.



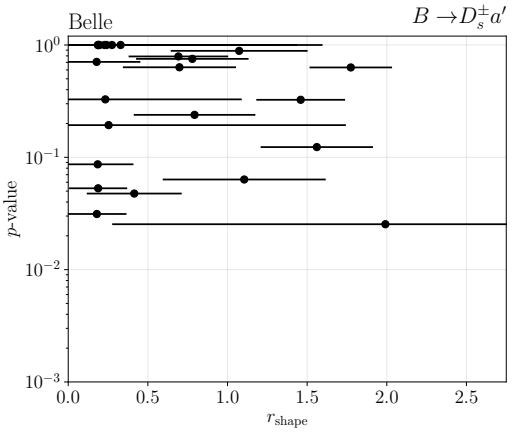
(a) $B^{\pm} \to \pi^{\pm} a$



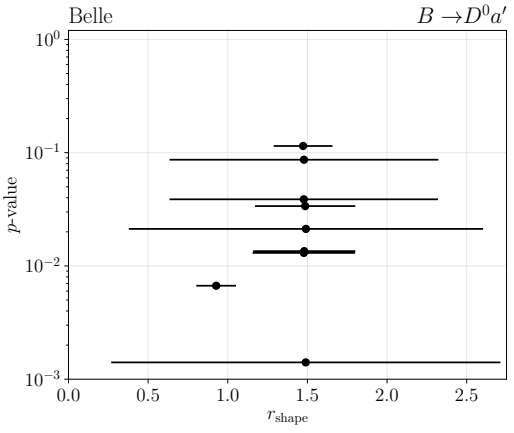
(b) $B^{\pm} \to K^{\pm} a$



(c) $B^{\pm} \to pa$

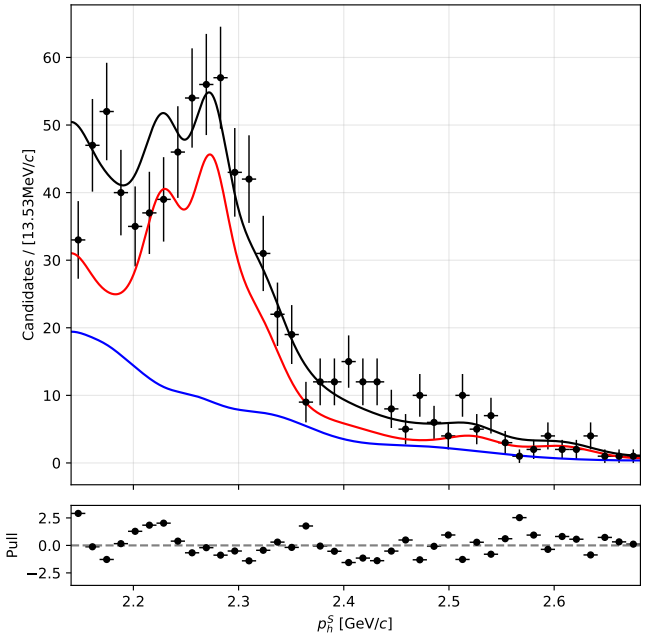


(d) $B^{\pm} \to D_s^{\pm} a$

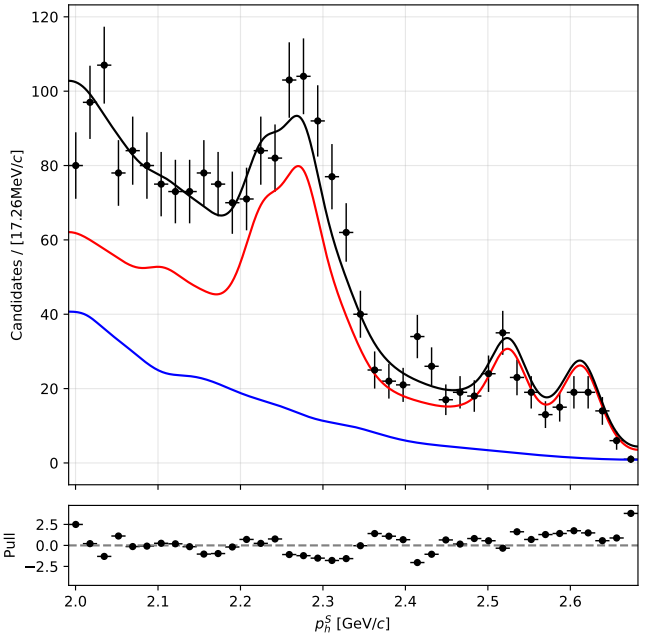


(e) $B^0 \to \overline{D}^0 a$

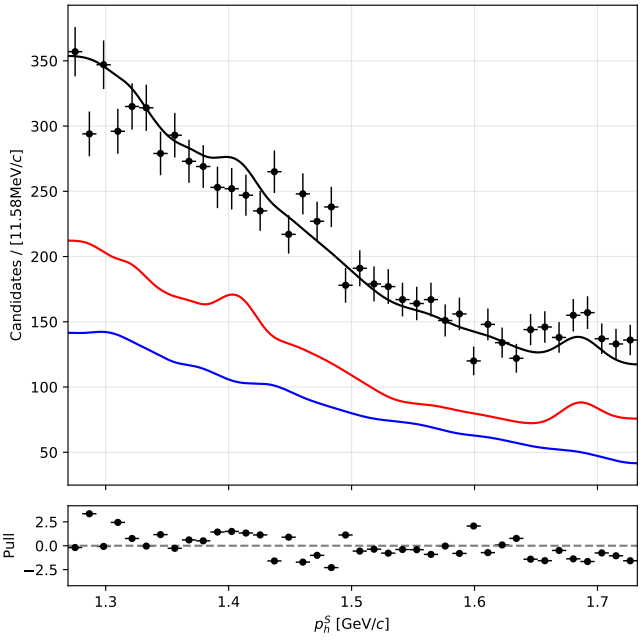
Figure 5.4: Background-only fit p-values for different values of r_{shape} for all 5 channels. Error bars show the statistical error returned by the minimiser during the fit.



(a) Fit centered around $m_a = 0.712\,\mathrm{GeV}/c^2$



(b) Fit centered around $m_a = 1.555\,\mathrm{GeV}/c^2$



(c) Fit centered around $m_a = 3.380\,\mathrm{GeV}/c^2$

Figure 5.5: Background-only fits in the sideband $1.5 < E_{\mathrm{ECL}} < 6.0\,\mathrm{GeV}$ corresponding to the three fit p-values lower than 1% in Fig. 5.4(b). All three distributions contain obvious peaking structures from the decays $B^\pm \to K^\pm\pi^0$, $B^\pm \to K^\pm K^0$, and $B^\pm \to K^\pm J/\psi$ respectively.

5.2.3 Combined fit

Signal extraction is done in a $\pm 10\sigma_{\text{eff}}$ window in p_h^S, and the KDE is defined over an extended range of $\pm 15\sigma_{\text{eff}}$ to mitigate the KDE edge effects. The continuum and $B\overline{B}$ background KDEs are added with some relative weighting that is allowed to float in the combined fit. The number of signal events S_{fit} and number of background events B are also floating parameters in the combined fit. Furthermore, for the possible peaking backgrounds listed in Tables 4.11 to 4.15, a PDF using the same template as the signal PDF is added, denoted $\text{PDF}_{\text{peak},i}$ in Equation (5.22). The yield of each of these peaking background components is governed by some nuisance parameter n_i, where i indexes the peaking background channel. The combined fit function can be written as

$$f(x) = S_{\text{fit}}\text{PDF}_{\text{sig}}(x) + B\Big[r_{q\overline{q}}\text{KDE}_{q\overline{q}}(x) + (1 - r_{q\overline{q}})\text{KDE}_{B\overline{B}}(x)\Big] + \sum_i n_i \text{PDF}_{\text{peak},i}(x) \tag{5.22}$$

where $r_{q\overline{q}}$ is the fraction of background events that $\text{KDE}_{q\overline{q}}$ contributes to the overall normalisation, $\text{PDF}_{\text{sig}}(x)$ is as it is defined in Equation (5.12), and the KDEs are adaptive kernel density estimators as defined in Equation (5.18). The vast majority of peaking backgrounds are found in the pion and kaon channels for m_a of about 2 GeV/c^2 or less, and so in these regions, there are many associated nuisance parameters in the fit. However, there are also many fit windows with no peaking backgrounds. Simulated events that come from peaking backgrounds are removed from the datasets used to define $\text{KDE}_{B\overline{B}}$ to avoid double-counting these events in the fit.

The key parameter of interest in this search is not simply the number of signal events, but the branching fraction $\mathcal{B}(B \to ha)$. These quantities are related to each other by

$$\mathcal{B} = \frac{N}{\varepsilon N_{B\overline{B}}} \tag{5.23}$$

where ε is the total signal selection efficiency, $N_{B\overline{B}}$ is the number of B meson pairs in the dataset, and N is the number of signal events observed. While Equation (5.23) obviously describes the relationship between the number of signal events seen and the a production branching fraction when 100% of the a and h decay width are reconstructed, it is important to remember that for some channels, secondary branching fractions need to be taken into account. For $h \in \pi^{\pm}, K^{\pm}, p$,

$$\mathcal{B} = \mathcal{B}(B^{\pm} \to ha) \times \mathcal{B}(a \to \text{invisible}), \tag{5.24}$$

while for $h \in D_s^{\pm}, \overline{D}^0$,

$$\mathcal{B} = \mathcal{B}(B^{\pm} \to ha) \times \mathcal{B}(\text{Secondary branching fractions}) \times \mathcal{B}(\text{a} \to \text{invisible}). \tag{5.25}$$

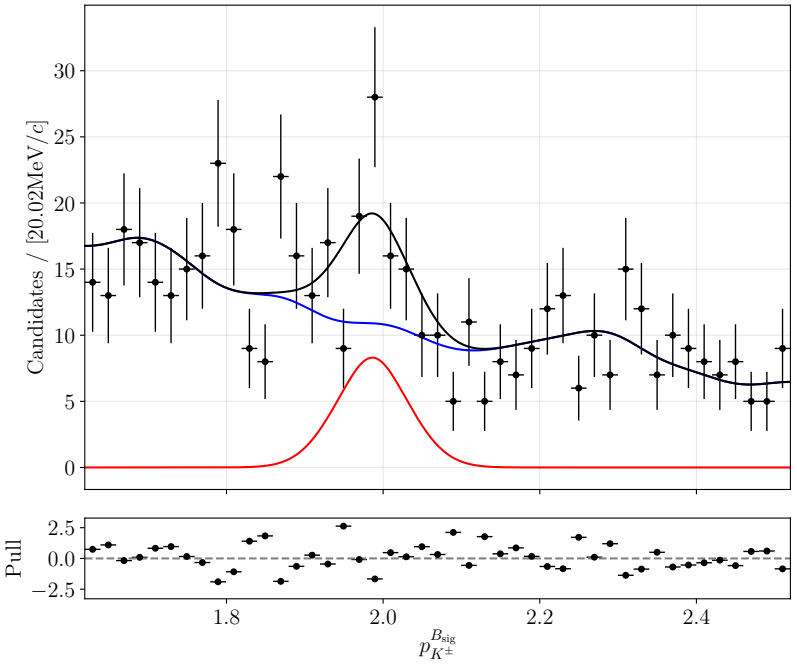


Figure 5.6: Example of the combined fit model on a dataset with 20 signal events injected. The KDE defined on the background is shown in blue, the signal PDF in red, and the combined PDF in black.

We extract $\mathcal{B}$ directly from the fit using Equation (5.23) to relate it to the signal PDF yield, and include the relevant systematic uncertainties on the efficiency and dataset size as nuisance parameters in this full fit model. A more thorough description of the treatment of systematic uncertainties can be found in Section 6.3.

Fig. 5.6 presents an example fit to the distribution from simulated events with an injected signal for the channel $B^\pm \to K^\pm a$, where $m_a = 2.55\,\text{GeV}/c^2$. The data includes 20 signal events combined with the background within the relevant mass region. The fit yields 25.3 ± 7.1 signal events and a local significance of 4.2σ.

5.2.4 Mass scan

As we have no a priori knowledge of m_a, we must scan every mass that the energy scale of the experiment affords us. Of course, with m_a being a continuous parameter, there are a theoretically infinite number of masses to scan. However, with finite detector resolution and statistics, there must be some limit as to the mass difference between two masses that are experimentally distinguishable. Therefore, to search all masses across the entire available mass range, we can scan in discrete mass points. The spacing between points is set to half of the effective resolution $\sigma_{\text{eff}}(p_h^S)$, following the approach established by other searches [97].

We verify this spacing does not reduce sensitivity by performing the combined fit with varying fit mass m_a on a dataset injected with 20 signal events with fixed a mass m_{test} (which is not necessarily equal to the fit mass). The fit mass is varied in steps of $0.5\sigma_{\text{eff}}$. A relatively large number of signal events is chosen so the significance clearly peaks when $m_a = m_{\text{test}}$. The resampling and fit are performed 500 times. For this spacing to be appropriate, the significance of an extracted signal should be constant for $\mu(m_a) \pm 1\sigma_{\text{eff}}$ of m_{test} where μ is the mean of the signal PDF, as defined in Equation (5.12). It can be written as a function of m_a using Equation (4.2). The results for four of these tests are shown in Fig. 5.7. For the first two tests, where m_a is set to some arbitrary mass in the middle of the accessible phase space region, it is clear that the fit mass being $\pm 0.5\sigma_{\text{eff}}$ on either side of the true mass m_{test} does not meaningfully alter the significance of the fit result. However, if m_{test} is close to 0, the significance stays high even if the fit mass is shifted to lower masses (and therefore higher values of μ). This is because above the range of kinematically allowed momenta for the signal process, the background is essentially non-existent, so any PDF that can fit events in that region will provide a better fit result than the background-only PDF. An increase of the signal PDF yield for higher values of μ than what is kinematically allowed will therefore yield a better fit. This behaviour can be seen in the bottom two plots of Fig. 5.7, where the significance is asymmetric on either side of the origin. For negative μ offset, it behaves normally, but for positive μ offset, the significance does not decrease.

5.2.5 Veto regions

As shown in Tables 4.11 to 4.15, there is a large number of two-body decays present in the background to this search, many of which occur due to narrow neutral resonances which would mimic the narrow invisible signal we are searching for. We only consider vetoing around a mass region if the expected yield from the corresponding decay in the aforementioned tables exceeds the expected upper limit on the number of signal events in Fig. 7.2. As the limit on the branching fraction is directly extracted from the fit, we convert it to a number of signal events using Equation (5.23), and round down to the nearest integer number to be conservative. If the expected number of events is lower than the number of events corresponding to the expected branching fraction sensitivity, then the decay is too rare to be seen by the analysis (given the reconstruction efficiency for that decay).

For each decay in Tables 4.11 to 4.15, we cross-check the method for deciding on a veto region by carrying out the combined fit method in a $\pm 10\sigma_{\text{eff}}$ window, centred around the expected location of the peak caused by the background decay. The tests start at the expected yield of the two-body background, and increase the signal event yield in integer steps until the local significance exceeds 3σ. This is run as a toy MC study with

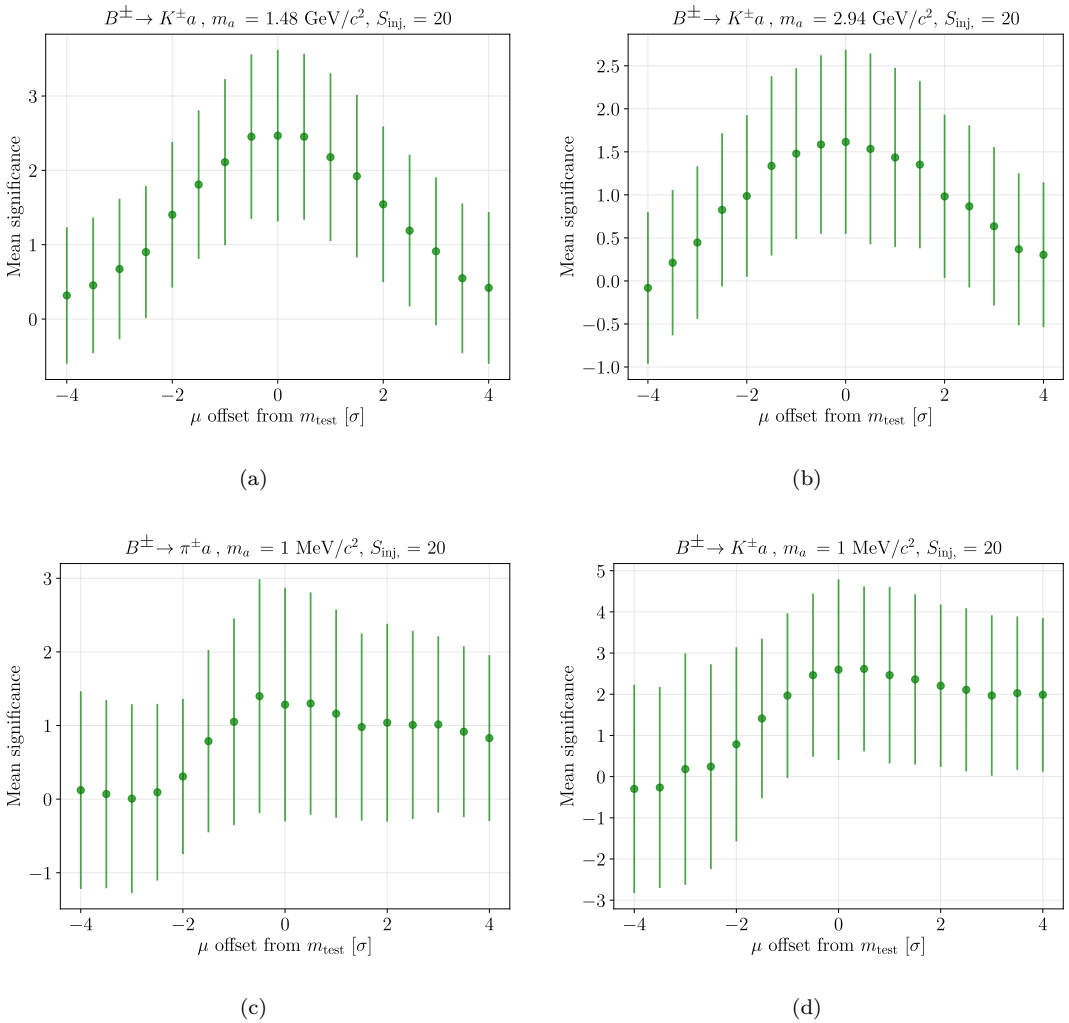
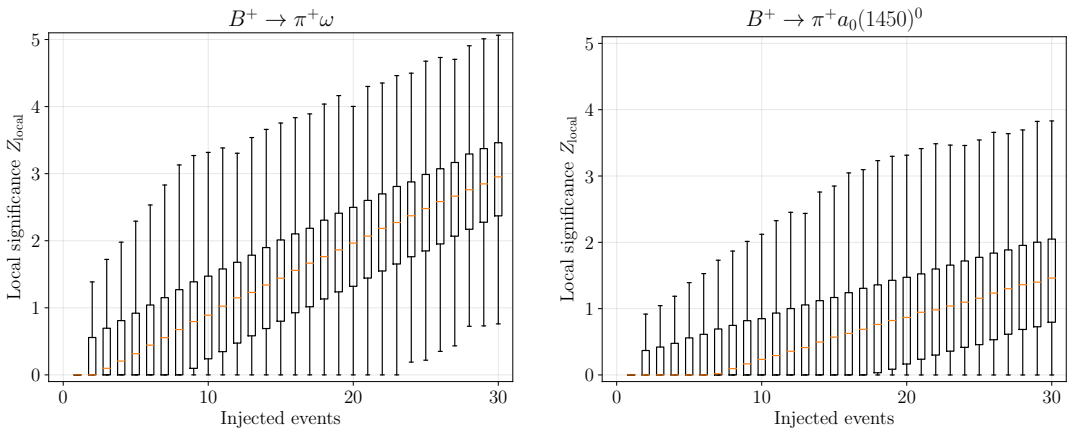


Figure 5.7: Mean and standard deviation of the extracted significance for 500 toy signal extractions at and around the mass where 20 signal events were injected for four different mass points.

500 toys for each peaking background. We found that the significance of the fit increases with injected events more rapidly for narrow resonances, while for broad resonances, the significance increases more slowly (as shown in Fig. 5.8 for the two decays $B^{\pm} \to \pi^{\pm}\omega$ and $B^{\pm} \to \pi^{\pm} a_0(1450)^0$). The widths of the ω and $a_0(1450)^0$ are 8.4 MeV/c^2 and 265 MeV/c^2 respectively.

5.3 Fit validation

Before proceeding to apply the fit method described in Section 5.2 to data, it is important to verify that it works as intended; that is, that it does not introduce bias to the measurement. A pull test is performed to check for any bias. For a fit parameter x, the



(a) (b)

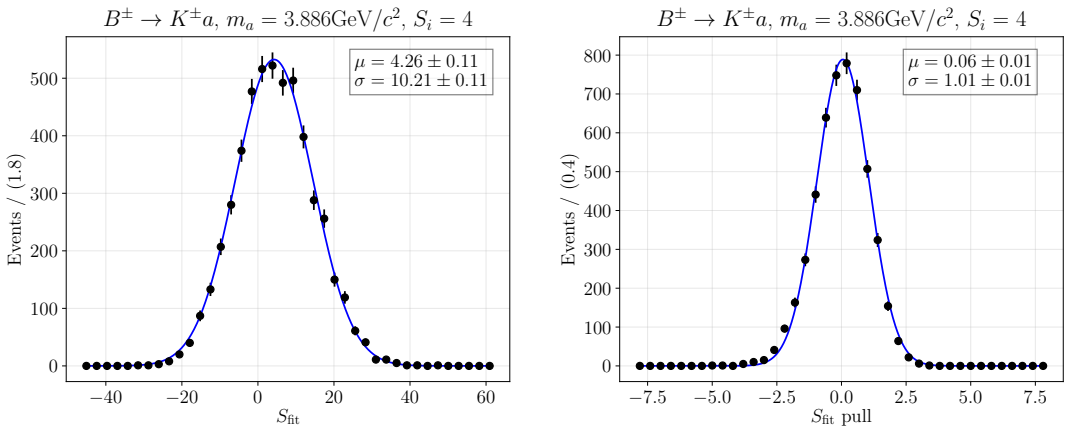
Figure 5.8: Significance vs. number of injected peaking background events for the decays $B^\pm \to \pi^\pm\omega$ and $B^\pm \to \pi^\pm a_0(1450)^0$. The widths of the ω and the $a_0(1450)^0$ are 8.4 MeV/c^2 and 265 MeV/c^2 respectively, and we see that the significance does not increase as rapidly for the $a_0(1450)^0$. The orange line on the plots represents the mean value of the toy MC studies. The boxes represent the area between the first and third quartile of the data, while the bars with caps are $\pm 1.5\times$ the interquartile range.

pull is defined as

$$\text{pull} = \frac{x_{\text{true}} - x_{\text{fit}}}{\sigma_{x,\text{fit}}} \tag{5.26}$$

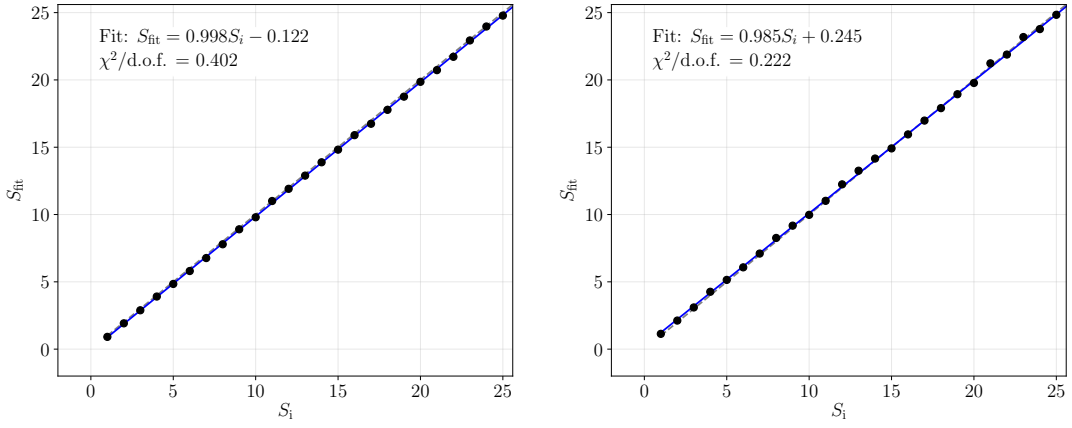
where x_{fit} and $\sigma_{x,\text{fit}}$ are the value and error of parameter x as extracted by the fit, and x_{true} is the value of the parameter that the dataset was generated with. In this search, the free parameter is the yield of the signal process $B \to ha$. If the fit perfectly yields the true value, then the pull will be zero. However, by repeating the fit on resampled datasets, we can check the pull distribution. The resampled datasets are generated from the signal and background PDF definitions, with yields that fluctuate around the expected yield with Poissonian error. We will sometimes use the term *pseudodataset* to refer to data that has been generated in this way. For each mass point, we repeat this resampling and fitting 12,500 times to obtain distributions of the signal yield and uncertainties, which are then used to calculate the pull distributions. We fit a Gaussian PDF to the pull distribution. If there is no bias in the method, the pull Gaussian should have a mean consistent with zero and a width of unity. These conditions respectively ensure that there is no bias introduced by the fit method and that the uncertainties are being properly evaluated. An example of these distributions is shown in Fig. 5.9.

The procedure above is repeated for every point in the mass scan with 0–35 injected signal events for each channel, amounting to a total of 2.25 million fits. We analyse the results of these fits to check the linearity of the method. If the fit approach is linear, then the number of injected signal events S_i will equal the number of signal events extracted



(a) (b)

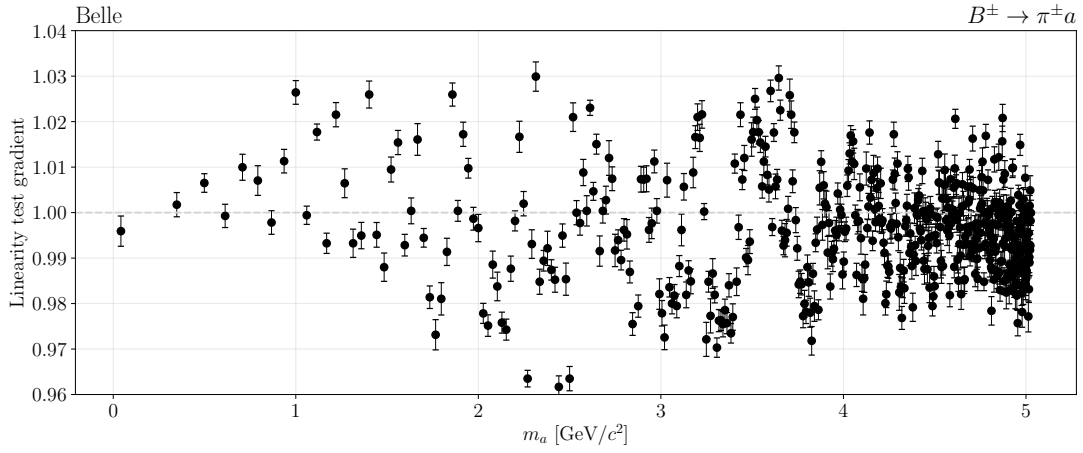
Figure 5.9: S_{fit} value and pull distributions for 12,500 toy experiments. The toy experiment shown is for $B^\pm \to K^\pm a$ with $m_a = 3.886\,\text{GeV}/c^2$.



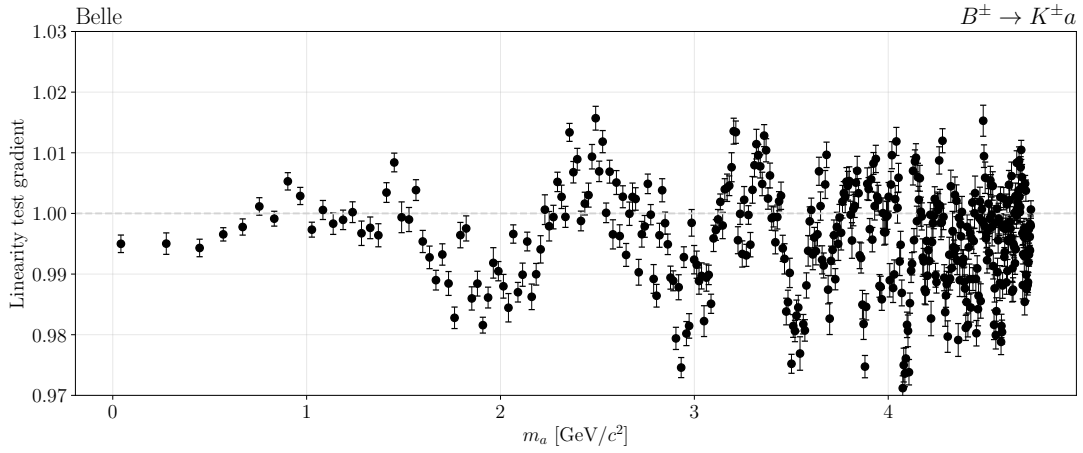
(a) $B^\pm \to K^\pm a$, $m_a = 670\,\text{MeV}/c^2$ (b) $B^\pm \to K^\pm a$, $m_a = 3.886\,\text{GeV}/c^2$

Figure 5.10: A selection of two linearity tests. Each data point corresponds to an MC study with 12,500 pseudodatasets. The corresponding fits are also shown.

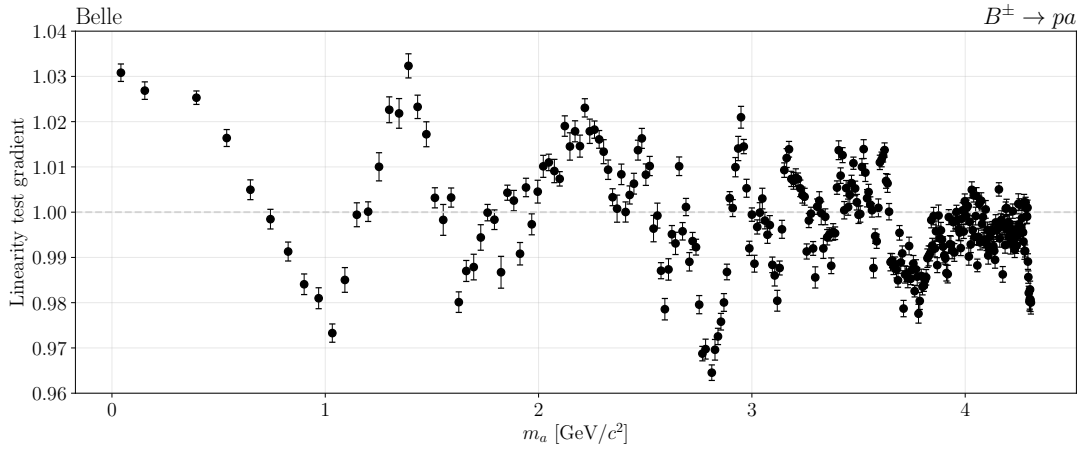
by the fit S_{fit}. For each point in the mass scan, we fit a straight line to S_i vs. S_{fit} and check the gradient to assess the linearity, as shown in Fig. 5.10. Here, a gradient consistent with unity indicates the fit extracts a parameter that scales linearly with the true value. The linearity gradients for all masses and channels are shown in Figures 5.11 and 5.12. The linearity gradients are within a few percent of unity, indicating a good level of linearity across the entire analysis. The data points in Figures 5.11 and 5.12 become more concentrated at higher masses due to improved resolution, resulting in a denser distribution of mass scan points.



(a)

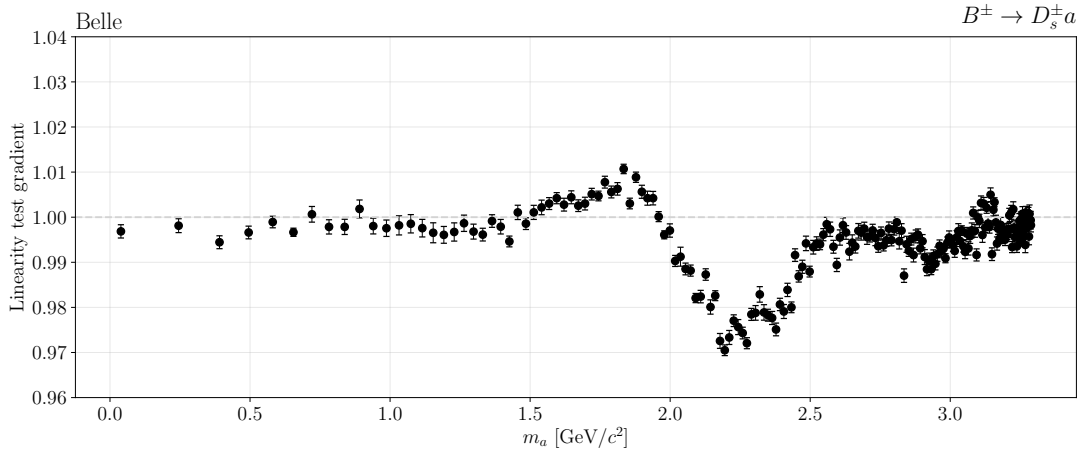


(b)

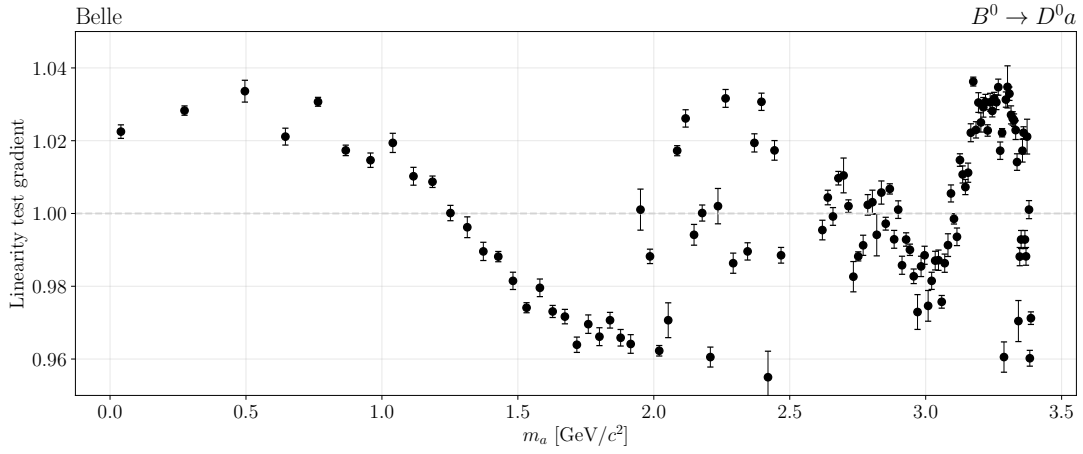


(c)

Figure 5.11: Gradients of linearity tests across the entire search range for $B^{\pm} \to \pi^{\pm} a$, $B^{\pm} \to K^{\pm} a$, and $B^{\pm} \to pa$. The full fit model's response to an injected signal is always linear to a few percent. Each data point on these plots is the result of a linear fit such as those shown in Fig. 5.10.



(a)



(b)

Figure 5.12: Gradients of linearity tests across the entire search range for $B^{\pm} \to D_s^{\pm} a$and $B^0 \to \overline{D}^0 a$.

6

Validation in data and systematic uncertainties

In the following chapter, the performance of the invisible particle search is validated on datasets orthogonal to the one that contains the signal channel, allowing the assessment of performance differences between simulation and real data. Furthermore, the key sources of systematic uncertainty in the search are summarised.

6.1 Control modes

Control modes are well-known SM decays that can be used to calibrate signal peak resolutions and overall efficiencies in the analysis, which could both exhibit some differences between data and MC. There are numerous inputs to the simulation that may not necessarily reflect the true function of the detector or the underlying physics processes produced by e^+e^- collisions. A good control mode is similar to the signal channel and has a high branching fraction to minimise statistical uncertainty during calibration. Of course, the nominal selection criteria are designed to reject many well-known SM processes, so the control modes used in this analysis are found in sidebands with selection orthogonal to the nominal signal selection criteria described in Section 4.5. As the calibration using the control modes requires looking at real data, the orthogonal selection ensures that the signal region remains blinded during this stage of the analysis process. Four channels in this analysis require the reconstruction of charged B mesons, while one requires B^0 reconstruction. As the two B meson types have completely different decay channels, we must look at a control mode for each of them.

The peaks in p_h^S for the control modes are fitted using the signal PDF templates, and a scale factor is extracted

$$\alpha_\sigma = \frac{\sigma^{\text{data}}}{\sigma^{\text{MC}}} \tag{6.1}$$

where σ^{MC} is the expected width of the signal Gaussian determined by fits to signal MC and σ^{data} is the actual width in the data. The scale factor α_σ will be multiplied with the σ parameters in the signal PDFs in the final fit, and will account for overall differences in resolution between MC and data. There is the possibility that α_σ has some dependence on the momentum of the signal-side hadron; however, we are limited in the number of control modes available to us. As such, we calculate it at fixed momenta for two control channels in the following sections and assume it is the same across the entire phase space.

These control modes can also be used for a comparison between data and MC by fitting a PDF and extracting a branching fraction, which can be compared to the nominal value provided by the PDG.

6.1.1 Charged B control mode: $B^+ \to J/\psi\, K^+$

We isolate events containing the decay $B^\pm \to K^\pm J/\psi$ by applying the same selection criteria as we do for the nominal analysis (except for the selection on BDT1) and with the number of ROE tracks required to be 2 rather than 0. This alternative selection is referred to as the *two-track sideband*. The selection on BDT1 is not applied as that classifier was trained to separate signal MC from generic B background MC, the latter of which contains the decay $B^\pm \to K^\pm J/\psi$. We show in Fig. 6.2 that the BDT1 response in $B^\pm \to K^\pm J/\psi$ events is much closer to the response we would expect from background events, rather than signal events. While the J/ψ decays to hadrons more than 87% of the time, the two-body decays to charged, stable final states ($\pi\pi$, KK, and $p\bar{p}$) all have branching fractions $\leq \mathcal{O}(10^{-3})$ [7]. In contrast, the branching fraction to electron and muon pairs is nearly 12%, making these leptonic decays the most common two-body final states of the J/ψ. We can assign mass hypotheses to the ROE tracks and calculate their invariant mass to select track pairs that are most likely from a J/ψ. We find that the invariant mass resolution of the two tracks in $B^\pm \to J/\psi\, K^\pm$ decays in the generic MC is about 57.5 MeV/c^2. Therefore, to further isolate this decay, we require the invariant mass of the ROE tracks to be within three times the resolution (about 172.5 MeV/c^2) of the nominal J/ψ mass of 3.097 GeV/c^2. The signal-side B is then fully reconstructed using the ROE tracks and ΔE is calculated for the B_{sig}. We require B_{sig} ΔE to be between -0.1 and 0.1 GeV. If multiple candidates remain, the candidate with the highest FEI signal probability is kept. These criteria retain 63% of $B^\pm \to K^\pm J/\psi$

events within the two-track sideband selection, while rejecting 98.7% of other events. The invariant mass, ΔE, and p_h^S distributions in MC and data are shown in Fig. 6.1.

The peak in p_K^S from the decay $B^\pm \to J/\psi K^\pm$ is fit with a double Gaussian PDF analogous to the PDF used to fit the signal channel peak. The background in the region 1.60 – 1.78 GeV/c is fit with a second-degree polynomial. The combined fit to data fixes all shape parameters as defined by the fit to MC, apart from the mean of the peak, and introduces a single floating fudge factor α_σ that varies the width of the signal Gaussian. The number of signal and background events are also floated in the fit to data. The final fits to MC and data are shown in Fig. 6.3. To derive an efficiency correction factor, we also correct for the difference between $\mathcal{B}(B^+ \to K^+ J/\psi)$ in Belle MC and the PDG. We find

$$r_{\text{charged}} = \frac{N_{\text{Data}}}{N_{\text{MC}}} \times \frac{\mathcal{B}_{\text{MC}}}{\mathcal{B}_{\text{Data}}} = \frac{283.3 \pm 9.5}{365.6 \pm 19.1} \times \frac{9.9 \times 10^{-4}}{(1.020 \pm 0.019) \times 10^{-3}} = 0.747 \pm 0.092 \quad (6.2)$$

where N_{Data} and N_{MC} are the yields of the double Gaussian in the fits to data and MC respectively, $\mathcal{B}_{\text{MC}}$ is the branching fraction in Belle MC (9.9×10^{-4}), and $\mathcal{B}_{\text{Data}}$ is the branching fraction according to the PDG ($(1.020 \pm 0.019) \times 10^{-3}$). The subscript *charged* denotes that the data-MC efficiency ratio r is derived using the FEI to reconstruct charged B mesons. This difference in efficiency between data and simulation is similar in magnitude to other studies using the same FEI training [198]. The ratio of the double Gaussian effective widths is

$$\alpha_{\sigma,\ \text{Charged}} = \frac{\sigma_{\text{Data}}}{\sigma_{\text{MC}}} = 1.21 \pm 0.05. \quad (6.3)$$

By allowing the mean of the peak to float in both fits, we check that the relationship between peak centre and mass in Equation (4.2) holds for the fit method in both MC and data. Using $m_{B^\pm} = 5.27934\,\text{GeV}/c^2$, $m_{J/\psi} = 3.0969\,\text{GeV}/c^2$, and $m_{K^\pm} = 493.667\,\text{MeV}/c^2$, we analytically calculate the mean for the peak to be $p_K^S = 1.684\,\text{GeV}/c$. The fit to MC returns a mean of $p_K^S = (1.683 \pm 0.002)\,\text{GeV}/c$, while the fit to data returns a mean of $p_K^S = (1.684 \pm 0.001)\,\text{GeV}/c$. At this momentum, the peak resolution is approximately 14 MeV/c, as shown in Fig. 5.2, so both fit results agree with the calculated value to well within $0.5\sigma_{\text{eff}}$.

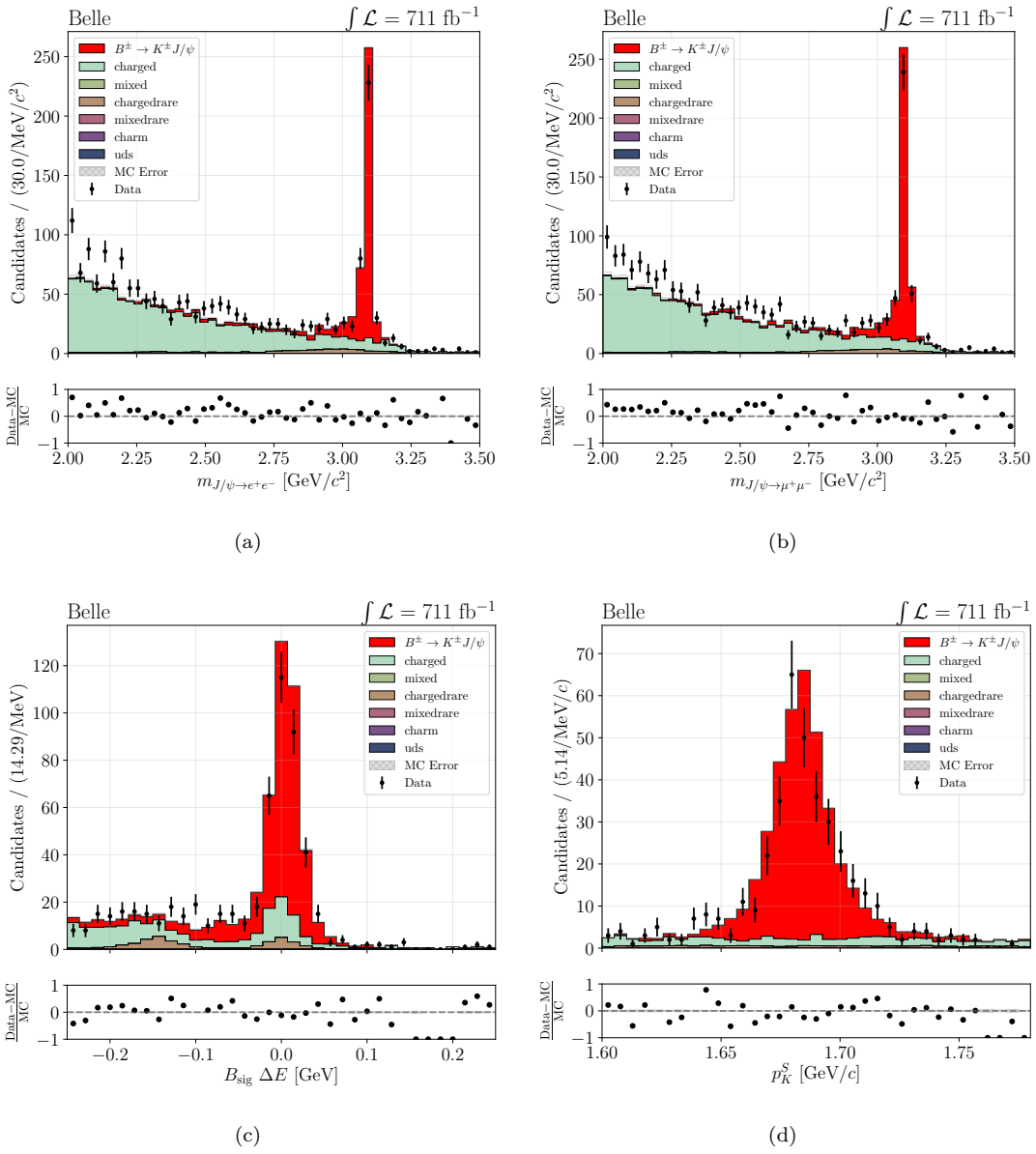
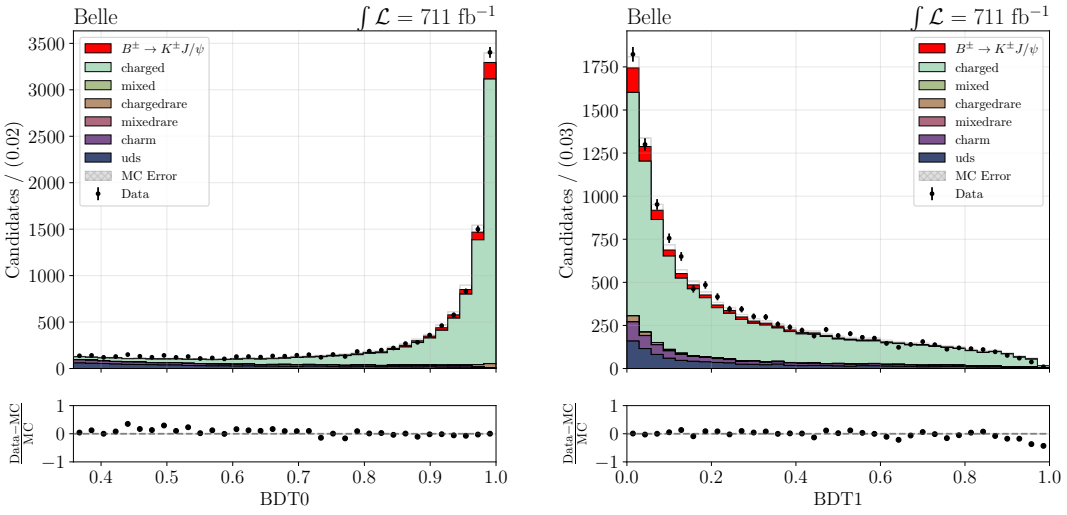
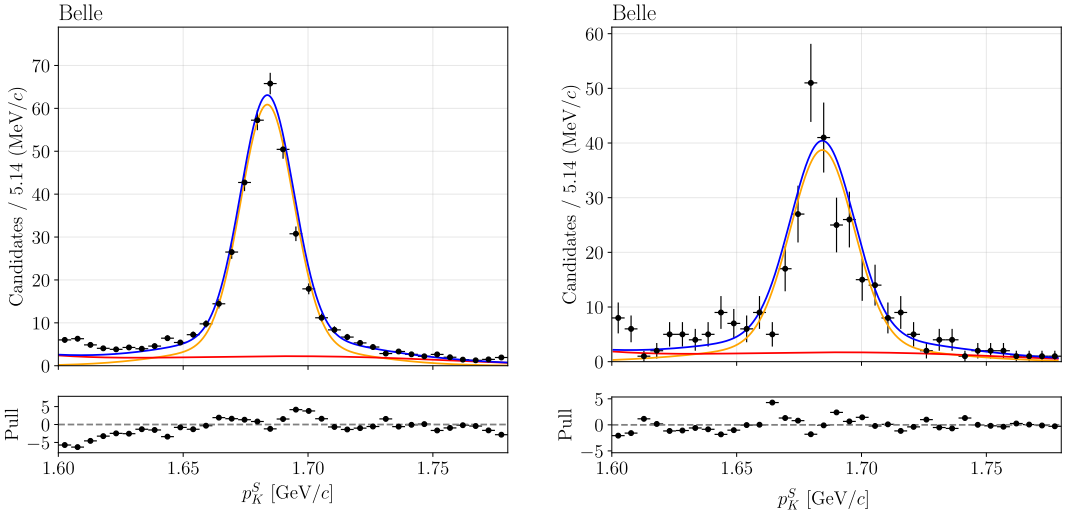


Figure 6.1: Distributions of variables used for control channel selection and fitting, with MC scaled by a factor of r_{charged}. We only consider events in the momentum range the J/ψ peak is fitted in (1.6–1.78 GeV/c). In (a) and (b) the J/ψ invariant mass distributions for electron and muon track mass hypotheses are shown. Both plots only show the distribution down to 2 GeV/c^2 as the background below that mass starts to dwarf the peak from the J/ψ. The histogram in (c) shows ΔE for the B meson fully reconstructed using the two ROE tracks. In (d) we show p_K^S in the region with the peak corresponding to the J/ψ after invariant mass and ΔE selection criteria are applied. MC in all four of these plots is scaled by the FEI correction factor of 0.785 described in Equation (6.6).



(a) (b)

Figure 6.2: (a) BDT0 (continuum suppression) and (b) BDT1 (B background suppression) responses in the two-track sideband dataset in MC and data, with $B^\pm \to J/\psi K^\pm$ events shown separately from the rest of the generic MC. Note that the BDT response to this channel is the same as to other events in the charged MC sample.



(a) (b)

Figure 6.3: Fits to the $B^\pm \to J/\psi K^\pm$ peak in (a) MC and (b) data. The small errors in (a) are because there are multiple streams in the MC, and so the yield is scaled down to match the luminosity in data.

6.1.2 Neutral B control mode: $B^0 \to \overline{D}^0\pi^0$

The easiest SM two-body decay that could be isolated with a B^0 decaying to a $\overline{D}^0$ was $B^0 \to \overline{D}^0\pi^0$. To isolate this channel, we chose a sideband where $2.25 < E_{\rm ECL} < 3.5\,{\rm GeV}$ and $2.2 < p^S_{D^0} < 2.45$, with all other selection criteria applied as normal, apart from BDT1, for the same reasons it is not applied in the $B^\pm \to J/\psi\, K^\pm$ sample. The isolation is still much worse than what can be achieved with $B^\pm \to J/\psi\, K^\pm$, but it is important to have a control mode for the neutral B meson, so we proceeded anyway. In Fig. 6.4 we show comparisons between data and MC in the distributions of the two background suppression BDTs, the invariant mass of the D^0 candidate, and $p^S_{D^0}$. The D^0 invariant mass distribution shows a slight shift towards lower values in data as compared to MC. As we have quite loose selection criteria on the mass of the D^0, any discrepancies in efficiency between data and MC will be quite small, especially when compared to the size of the disagreement between efficiencies in data and MC that arises from using the FEI.

The peak in $p^S_{\overline{D}^0}$ from the decay $B^0 \to \overline{D}^0\pi^0$ is fit with a double Gaussian PDF analogous to that which the signal channel is fit with. The background in the region 2.2 – 2.45 GeV/c is fit with a KDE, using the same method as the KDE used for the background in the signal region. Apart from the mean of the peak, the combined fit to data fixes all shape parameters as defined by the fit to MC, and introduces a single floating fudge factor α_σ that varies the width of the signal Gaussian. The number of signal and background events are also floated in the fit to data. The final fits to MC and data are shown in Fig. 6.5. To derive an efficiency correction factor, we also correct for the difference between $\mathcal{B}(B^0 \to \overline{D}^0\pi^0)$ in Belle MC and the PDG. We find

$$r_{\rm neutral} = \frac{N_{\rm Data}}{N_{\rm MC}} \times \frac{\mathcal{B}_{\rm MC}}{\mathcal{B}_{\rm Data}} = \frac{30.7 \pm 11.3}{35.6 \pm 6.0} \times \frac{2.60 \times 10^{-4}}{(2.67 \pm 0.09) \times 10^{-4}} = 0.839 \pm 0.341 \quad (6.4)$$

where the above parameters are defined analogously to those in Equation (6.2). The ratio of the double Gaussian effective widths for this control mode is

$$\alpha_{\sigma,\ \rm Neutral} = \frac{\sigma_{\rm Data}}{\sigma_{\rm MC}} = 1.08 \pm 0.23. \quad (6.5)$$

By allowing the mean of the peak to float in both fits, we check that the relationship between peak centre and mass in Equation (4.2) holds for the fit method in both MC and data. Using $m_{B^0} = 5.27966\,{\rm GeV}/c^2$, $m_{D^0} = 1.86584\,{\rm GeV}/c^2$, and $m_{\pi^0} = 134.979\,{\rm MeV}/c^2$, we analytically calculate the mean for the peak to be $p^S_{\overline{D}^0} = 2.3079\,{\rm GeV}/c$. The fit to MC returns a mean of $p^S_{\overline{D}^0} = (2.306 \pm 0.007)\,{\rm GeV}/c$, while the fit to data returns a mean of $p^S_{\overline{D}^0} = (2.320 \pm 0.008)\,{\rm GeV}/c$. At this momentum, the peak resolution is approximately

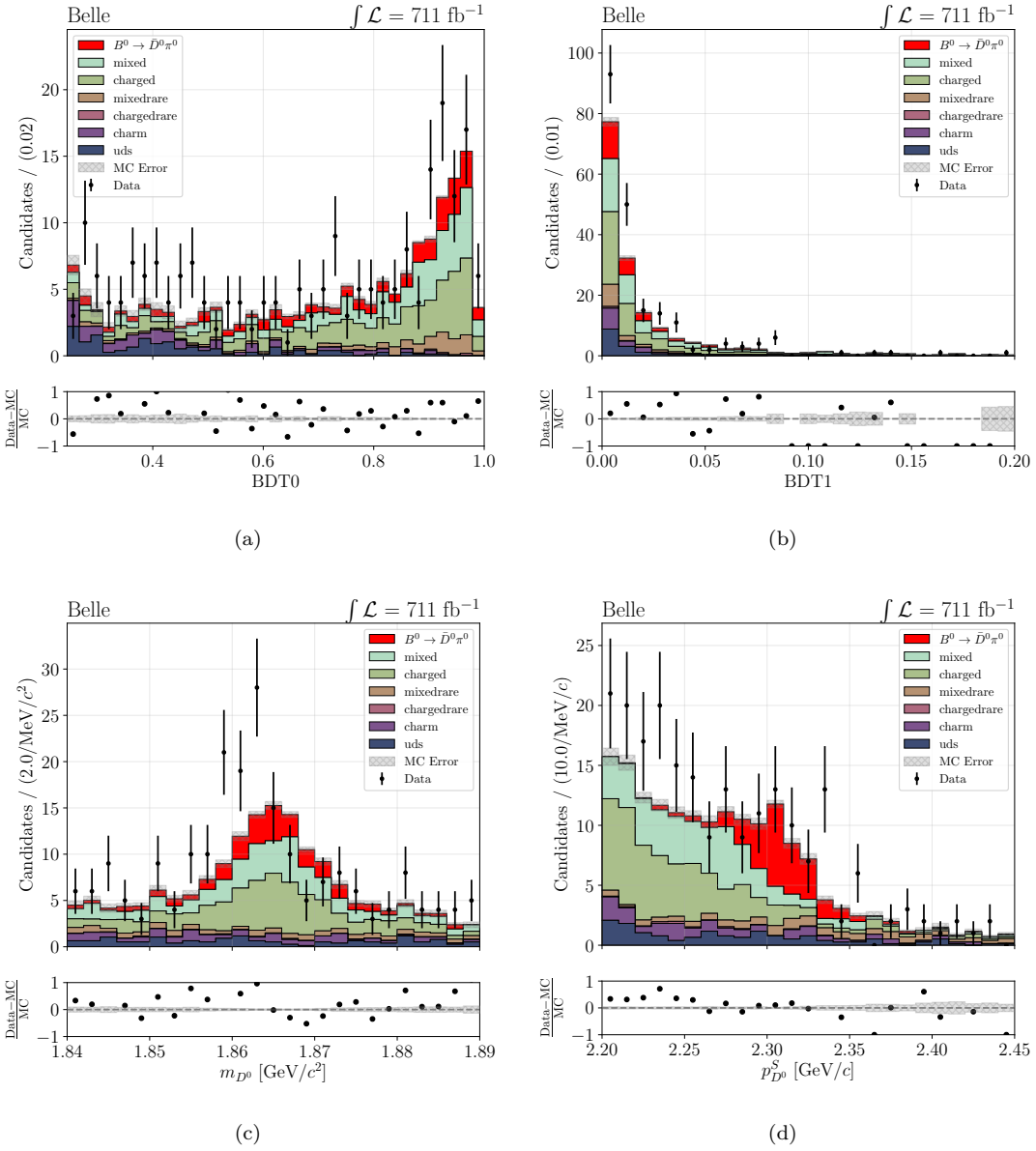
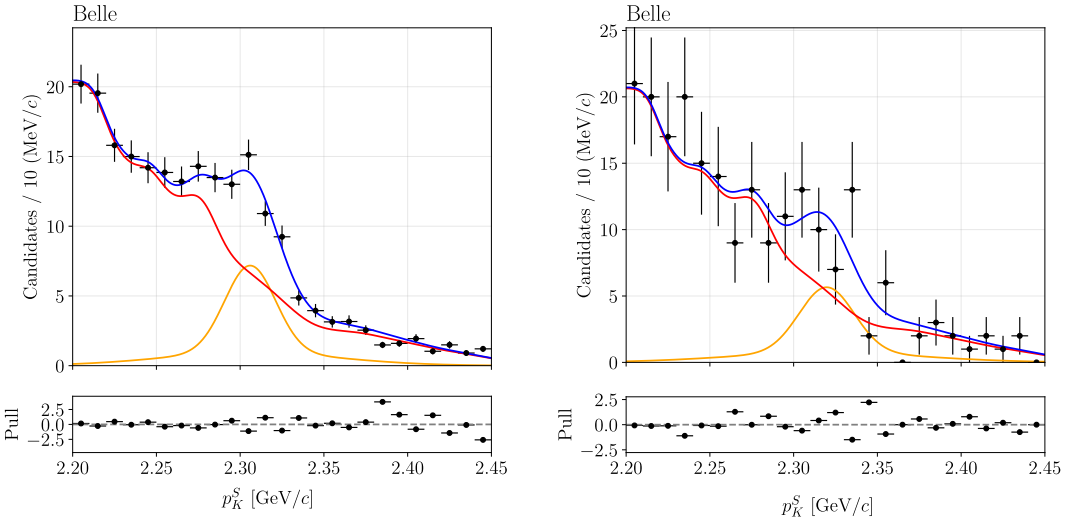


Figure 6.4: In (a) and (b), data-MC comparisons for the two background suppression BDTs' responses are presented, (c) shows the invariant mass of the D^0 candidate, and (d) the signal-B frame D^0 momentum. For all of these plots, the MC is scaled by a factor of r_{neutral} as defined in Equation (6.7).

30 MeV/c, as shown in Fig. 5.2, so while there is a shift between the calculated value and the fit to data, it is a shift of less than $0.5\sigma_{\text{eff}}$, and as such will have a negligible impact on the result.

6.1.3 Comparison to ratios derived from $B \to D^{(*)}\ell\nu_\ell$ decays

The Belle measurement of the $B \to D^{**}\ell\nu_\ell$ branching fractions detailed in Ref. [199] uses the same FEI training and selection as this analysis, and uses ratios of $B \to D^{(*)}\ell\nu_\ell$



(a) (b)

Figure 6.5: Fits to the $B^0 \to \overline{D}^0\pi^0$ peak in MC (a) and data (b). The small errors in the MC dataset are because there are multiple streams in the MC, and so the yield is scaled down to match the luminosity in data.

($\ell = e,\ \mu$) decays from which we can compare efficiency differences between MC and data. The larger branching fractions of these semileptonic decays allow for much higher statistics than the control samples detailed in the previous section. The event yields in MC and fit yields in data from this analysis are summarised in Table 6.1.

Table 6.1: Event yields in data and MC for semileptonic decays from Ref. [200], with data-MC ratios in the rightmost column. The ratios are generally higher for the channels containing muons rather than electrons.

Channel	Data yield	MC yield (6 streams)	PDG BF (%)	MC BF (%)	Ratio
$B^0 \to D^- e^+ \nu_e$	3154 ± 67	23635	2.31	2.13	0.741 ± 0.024
$B^0 \to D^- \mu^+ \nu_\mu$	2723 ± 60	20214	2.31	2.13	0.803 ± 0.026
$B^+ \to \overline{D}^0 e^+ \nu_e$	8974 ± 136	65874	2.35	2.31	0.744 ± 0.025
$B^+ \to \overline{D}^0 \mu^+ \nu_\mu$	7752 ± 124	56678	2.35	2.31	0.807 ± 0.027
$B^0 \to D^{*-} e^+ \nu_e$	6271 ± 102	49365	4.90	5.33	0.824 ± 0.011
$B^0 \to D^{*-} \mu^+ \nu_\mu$	5624 ± 91	44403	4.90	5.33	0.775 ± 0.010
$B^+ \to \overline{D}^{*0} e^+ \nu_e$	19940 ± 200	157980	5.66	5.79	0.690 ± 0.050
$B^+ \to \overline{D}^{*0} \mu^+ \nu_\mu$	18045 ± 183	142049	5.66	5.79	0.750 ± 0.050

From this analysis, we compute an efficiency ratio for both the charged and neutral FEI and find:

$$r_{\text{charged}} = \frac{\sum \frac{N_{\text{Data}}\mathcal{B}_{\text{MC}}}{\mathcal{B}_{\text{PDG}}}}{\sum N_{\text{MC}}} = 0.785 \pm 0.005 \pm 0.018 \tag{6.6}$$

where the sum is over the four charged B channels in Table 6.1 and

$$r_{\text{neutral}} = \frac{\sum \frac{N_{\text{Data}}\mathcal{B}_{\text{MC}}}{\mathcal{B}_{\text{PDG}}}}{\sum N_{\text{MC}}} = 0.779 \pm 0.006 \pm 0.015 \tag{6.7}$$

is calculated in an analogous manner, where the first and second errors are the statistical and systematic errors respectively. Note that these should be added in quadrature to get the total error on the ratios. For both charged and neutral B meson reconstruction with the FEI, we find that the data-MC efficiency ratios from semileptonic decays are in agreement with the yield ratios from the $B^{\pm} \to J/\psi K^{\pm}$ and $B^0 \to \overline{D}^0\pi^0$ control modes respectively, within error. The errors on the ratios in Equations (6.6) and (6.7) include the uncertainty on the yield in data, the uncertainty from MC statistics, as well as the systematic uncertainties related to D-meson reconstruction from Ref. [200].

We use r_{charged} to scale MC for all four $B^{\pm}$-containing channels and r_{neutral} for the one channel that reconstructs B^0.

6.2 Cross-checks with off-resonance data sample

We perform data-MC agreement checks on the off-resonance data sample to verify the ability of the background suppression BDTs to generalise to real data and to ensure that the continuum component of the background is properly normalised. In the full fit model, the normalisation of the continuum component is allowed to float, and based on linearity studies, we do not expect it to bias the results. However, significant deviations from expectations could indicate an issue with our understanding of the continuum rate. Furthermore, if this disagreement is momentum-dependent, it could lead to a distortion in the continuum background shape.

In Fig. 6.6, we show the distribution of signal-side hadron momentum in the B_{sig} frame for off-resonance data and MC, with the full selection criteria applied, except for the B_{tag} beam-constrained mass selection. This exclusion is necessary because the variable is shifted downward due to the lower beam energy used in the off-resonance sample. There is some obvious disagreement between the distributions in data and MC, although only a small number of events survive the selection. For all channels except $B^{\pm} \to D_s^{\pm}a$, the momentum distribution in data is noticeably skewed toward higher values compared to the MC. For the channel $B^{\pm} \to D_s^{\pm}a$, the limited number of events precludes any strong conclusions, unlike the other four channels.

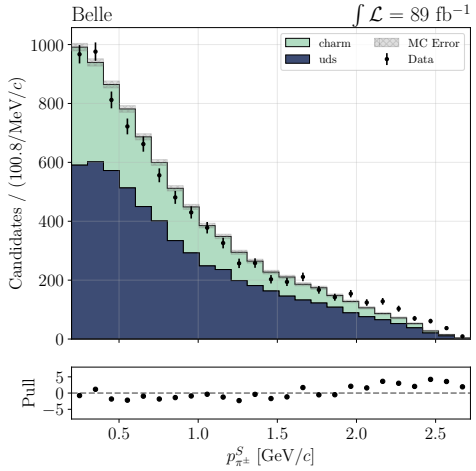
6.2.1 Continuum momentum correction

The fraction of simulated continuum events that hadronise into different excited states is not necessarily correct, which can result in inaccuracies in the meson momentum distributions, as is evident in Fig. 6.6. In these plots, the MC is scaled to match the luminosity of the data sample. No other rescaling is applied. Initially, we attempted to reweight all off-resonance MC using a fifth-order polynomial. However, the large uncertainties in the polynomial's fit parameters significantly reduced the sensitivity of the analysis.

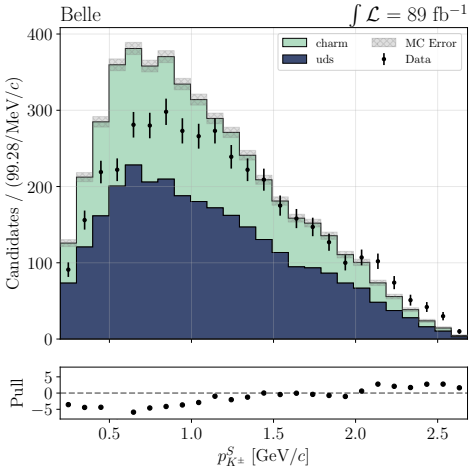
To address this issue, we implemented a simpler parameterisation using a linear function on a per-fit-window basis. Specifically, in each fit window, we define a linear correction function:

$$c(x) = a_0 + a_1 x \tag{6.8}$$

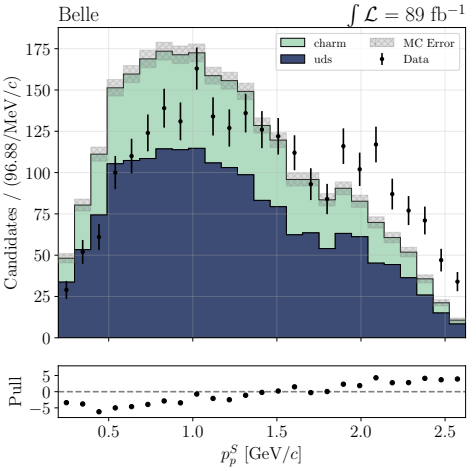
and then replace $\mathrm{KDE}_{q\bar{q}}(x)$ with $c(x)\mathrm{KDE}_{q\bar{q}}(x)$. In this parameterisation, the slope parameter a_1 is treated as a nuisance parameter in the combined fit and is allowed to float within its uncertainties. The intercept a_0, on the other hand, does not affect the shape of the continuum background PDF; it only influences the overall normalisation. Since the $r_{q\bar{q}}$ parameter already allows the continuum component's overall contribution to float during the fit, we fix a_0 in the combined fit to avoid introducing redundancy with $r_{q\bar{q}}$.



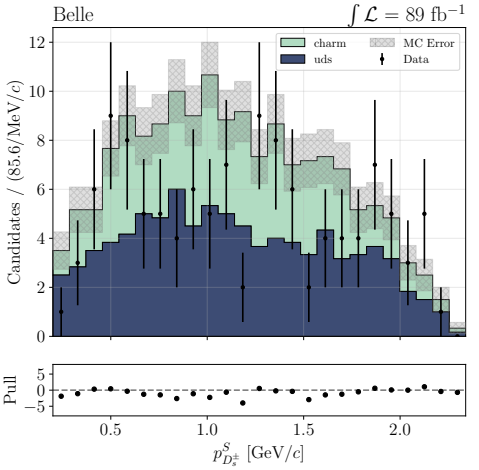
(a) $B^\pm \to \pi^\pm a$



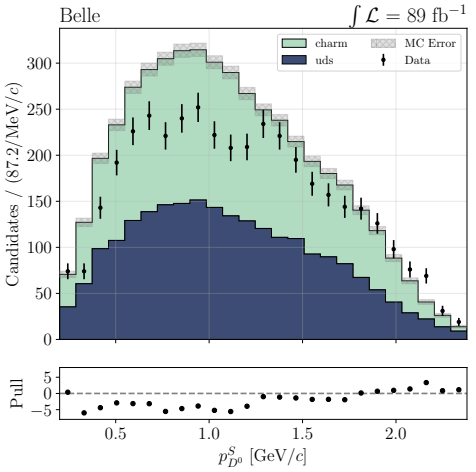
(b) $B^\pm \to K^\pm a$



(c) $B^\pm \to pa$

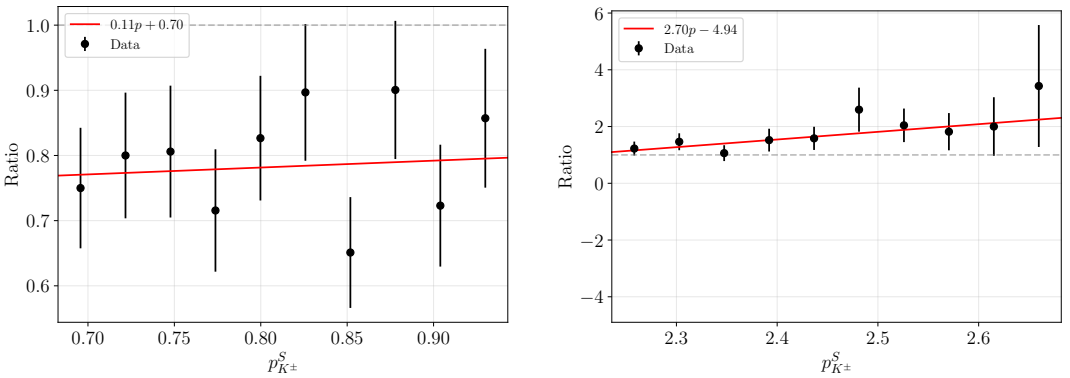


(d) $B^\pm \to D_s^\pm a$



(e) $B^0 \to \overline{D}^0 a$

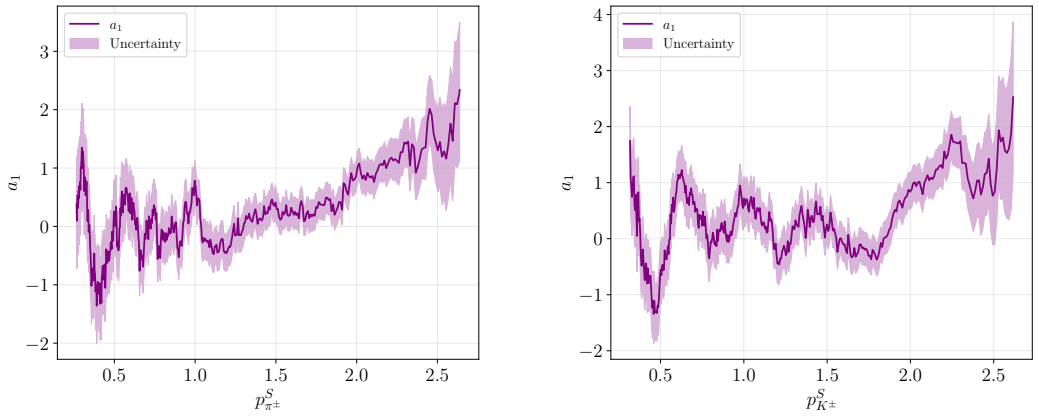
Figure 6.6: A comparison of the p_h^S distributions in off-resonance data and MC.



(a) $B^\pm \to K^\pm a$, fit window for $m_a = 4.251\,\mathrm{GeV}/c^2$

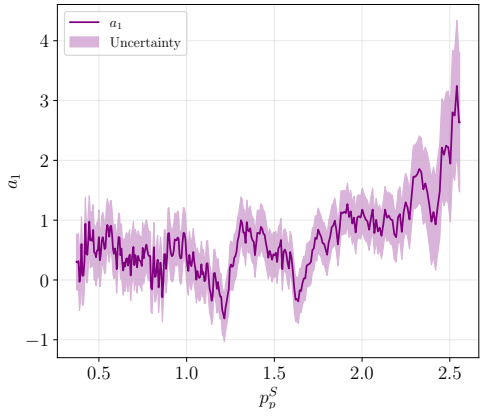
(b) $B^\pm \to K^\pm a$, fit window for $m_a = 0.042\,\mathrm{GeV}/c^2$

Figure 6.7: Data-MC ratios for the off-resonance dataset in two distinct fit windows, with the linear fits in red that we used to correct the continuum distribution. For the higher mass window in (a), the correction is flatter than what we see in the low-mass window in (b).

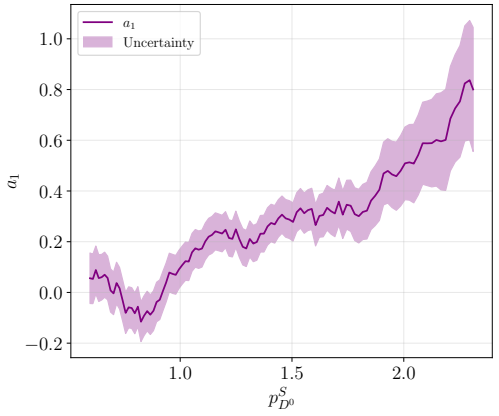


(a) $B^\pm \to \pi^\pm a$ test

(b) $B^\pm \to K^\pm a$ test



(c) $B^\pm \to pa$ test



(d) $B^0 \to \overline{D}^0 a$ test

Figure 6.8: Correction gradients a_1, as defined in Equation (6.8) for every scan point in all channels. To avoid showing hundreds of individual error bars on each plot (with the error on the gradient σ_{a_1} on each), the area bounded by $a_1 \pm \sigma_{a_1}$ is shaded in purple.

6.3 Systematic uncertainties

Systematic uncertainties reflect limitations in our understanding of an experiment. Typically, they arise from imperfect simulations or theoretical predictions, or from uncertainties in input parameters. They can also stem from uncertainties in corrections applied to simulations to better match real data. While statistical uncertainties on a dataset of size N scale with $1/\sqrt{N}$, systematic uncertainties do not, as they often arise from fixed experimental effects or correlated uncertainties that do not diminish with additional data. Reduction of systematic uncertainties comes from improving our understanding of the underlying processes, whatever they may be. A summary of the sources of systematic uncertainty present in this analysis is presented in Table 6.2.

Setting a limit on the branching fractions to the decays of interest in this analysis requires three main inputs: the limit on the number of signal events we are sensitive to, the signal selection efficiency, and the number of B mesons in the Belle dataset. The number of events the analysis is sensitive to depends on the fit model. All systematic uncertainties are treated as nuisance parameters during the statistical analysis, and are discussed in Section 6.3.1. The selection efficiency itself depends on a number of inputs which are discussed in Section 6.3.2. However, it should be noted here that for this analysis, the dominant uncertainty on the result comes from the statistical variation in the fit result rather than any systematic effect.

6.3.1 Modelling uncertainties

The first important systematic uncertainty we consider is the number of B meson pairs in the dataset. Belle recorded $(771.581 \pm 10.566) \times 10^6$ B meson pairs during its operation. This is key as the branching fraction sensitivity directly relies on knowing the number of B meson pairs in the dataset.

The signal PDF parameterisation described in Equation (5.12) may introduce some systematic bias in our analysis. As such, the three signal fit parameters f, σ_1, and σ_2 are allowed to vary as nuisance parameters in the final fit. The variation is constrained by a Gaussian centred about the expected value. The maximum difference between the parameterised value in Fig. 5.2 and the real fit result (i.e. the maximum difference between a data point and a curve in Fig. 5.2) is used as a conservative estimate of the width of this Gaussian constraint. This is done for all channels.

The number of events and the composition of the SM background we expect can depend on a number of parameters, some of which are shared with the signal efficiency, including the PID correction and tracking uncertainties. Uncertainties are calculated with an analogous approach to that in Table 6.3. The most important source of uncertainty

Source	Description
B_{tag} efficiency correction	Correction applied to match MC tagging efficiency to data. Dominant uncertainty for 1-track signal channels.
Number of $B\bar{B}$ pairs	Uncertainty on the total number of B mesons produced in the Belle dataset.
Signal PDF parameters	Gaussian-constrained nuisance parameters (f, σ_1, σ_2) modelling signal shape.
Peaking background yields	Signal-like backgrounds from rare decays, modelled explicitly with 100% yield uncertainty.
Background shape modelling	Uncertainty due to finite MC statistics and potential mismodelling of inclusive background shapes.
Tracking efficiency	0.35% per track added in quadrature. Larger effect for multi-track final states.
PID corrections	Data-driven corrections to PID efficiency; uncertainties from Belle standard tables.
π^0 and photon efficiency	Relevant to channels involving electromagnetic clusters; uncertainty propagated to selection efficiency.
Signal MC statistics	Uncertainty due to limited simulated events used to estimate efficiency.
Branching fraction inputs	PDG-reported uncertainties on decay modes used for reconstruction or normalisation.

Table 6.2: Overview of systematic uncertainties considered in this analysis. Each is treated as a nuisance parameter in the statistical model.

regarding the background shape is the impact of rare peaking backgrounds. We include signal-like PDFs for these backgrounds, and assign a conservative 100% uncertainty on their expected yield as listed in Tables 4.11 to 4.15. This only applies to backgrounds that are not common enough to be vetoed. The yield of any peaking backgrounds present in a fit window is treated as a nuisance parameter in the fit model. Some of these backgrounds have an expected yield of less than 1 event. A fluctuation based on a yield of 0.04 events obviously does not reflect the change in a real dataset, where event counts are always integers. To this end, we take a conservative approach and set the minimum fluctuation in the number of events for these peaking background yields to 2. The decay widths of many of these peaking backgrounds are substantial, and we have

already shown the relationship between peak width and signal sensitivity in Fig. 5.8. However, rather than choosing some arbitrary cutoff decay width at which we determine a two-body background to be non-peaking, we treat all peaking backgrounds as a source of systematic uncertainty. It is expected that broad resonances will not adversely alter the sensitivity compared to narrow resonances, as a relatively smaller fraction of those backgrounds will lie in any single signal region. To factor in the non-negligible decay widths, we change the signal-like PDFs for decays with broad resonances. For all two-body backgrounds in Tables 4.11 to 4.15 containing neutral hadrons with non-zero decay widths, we sample 50,000 values from a Breit-Wigner distribution described by their mass and decay width. This is converted to a p_h^S distribution with Equation (4.2). The width of this sampled momentum is first calculated, and then added in quadrature with the width of the signal-like PDF for each decay to get a peaking distribution corrected for the real decay width of the particle. The fit window corresponding to $m_a = 1.655\,\text{GeV}/c^2$ for the channel $B^\pm \to \pi^\pm a$ contains the largest number of peaking backgrounds, with 22 (and therefore 22 nuisance parameters added to the fit for these backgrounds).

6.3.2 Signal selection efficiency

The fit method directly extracts the branching fraction following Equation (5.23). This means that the signal selection efficiency needs to be treated as a nuisance parameter in the fit. There are a number of corrections applied to the raw efficiency on simulated events to get a more accurate estimate of the efficiency in real data. All of these uncertainties are summed in quadrature to find the total systematic uncertainty on the signal selection efficiency. The total signal selection efficiency is then allowed to float with a Gaussian constraint based on its total uncertainty. The contribution to the uncertainties from various sources is shown in Table 6.3.

The sources of systematic uncertainty on the signal selection efficiency are: the uncertainty on the B_{tag} efficiency correction (as detailed in Section 6.1), PID corrections (from standard tables), tracking uncertainty (0.35% per track), signal MC statistics, and, for the $\overline{D}^0$ channel only, the π^0 efficiency correction. The signal selection efficiency for channels containing D^0 and $D_s^\pm$ relies on the measured branching fractions of the modes they are reconstructed in ($D_s^+ \to K^+K^-\pi^+$ and as described in Table 4.3 for $\overline{D}^0$). We include the uncertainty on these measurements as reported by the PDG as a systematic uncertainty. The largest source of systematic uncertainty is the B_{tag} efficiency correction for all channels except $B^0 \to \overline{D}^0 a$, for which the secondary branching fractions dominate. The channels with D mesons also have larger tracking and signal MC statistics uncertainties. The channels $B^\pm \to D_s^\pm a$ and $B^0 \to \overline{D}^0 a$ have greater tracking uncertainties as the 0.35% uncertainty is added on a per-track basis. For $B^0 \to \overline{D}^0 a$, the value comes from the weighted average of the number of tracks on the signal side, based on

the secondary branching fractions used during reconstruction shown in Table 4.3. Signal MC statistics' systematic uncertainty could be reduced by generating larger signal MC samples in the future. The π^0 correction is derived from a previous internal Belle study of π^0 efficiency [201]. Each systematic uncertainty is calculated by varying the weights of the signal MC based on the uncertainty on the size of the correction, and then taking the average result across all mass hypotheses. The systematic effects show negligible variation across the range of mass hypotheses generated.

Table 6.3: Summary of the systematic uncertainties on the signal selection efficiency, with the total uncertainty calculated as the sum in quadrature of the individual uncertainties.

Source	Uncertainty (%)				
Channel / h	$\pi^\pm$	$K^\pm$	p	$D_s^\pm$	$\overline{D}^0$
B_{tag} efficiency	2.38	2.38	2.38	2.38	2.07
PID	0.20	0.16	0.53	0.36	0.33
Tracking	0.35	0.35	0.35	1.05	0.84
Signal MC statistics	0.61	0.59	0.67	1.32	1.16
Secondary branching fraction	-	-	-	1.86	2.33
π^0 selection	-	-	-	-	2.00
Total uncertainty	2.49	2.48	2.55	3.48	3.98

7

Signal extraction and results

With detailed descriptions of the physics motivation and experimental and statistical methods used to search for new particles, we now show the expected sensitivity of the search, as well as the results on the full Belle dataset. The results include the signal significance as a function of invisible particle mass, as well as upper limits on branching fractions to the search channels.

7.1 Expected sensitivity

Section 5.1 described the CL_s frequentist method [176] that was used to extract a limit on $\mathcal{B}(B \to ha)$. The method was implemented using the `RooStats` framework based on the fit model described in Section 5.2. Specifically, the asymptotic calculator strategy was used [202]. The parameter of interest in the limit setting is the branching fraction itself. To compute the expected sensitivity across all phase space, the limit calculation was performed for every $0.5\sigma_{\text{eff}}$ step, following the mass scan described in Section 5.2.4.

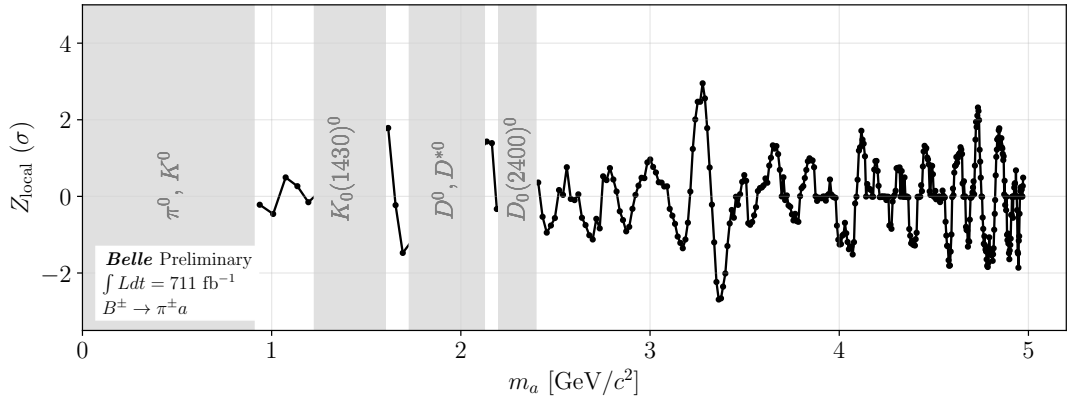
All of the limits shown in the main body of this thesis are computed at the 90% confidence level, as many of the other experimental results we intend to compare these to are calculated at that same level. It is important to compute the expected upper limit as a benchmark to compare the final observed upper limit with. Significant deviation observed from the expected limit across large areas of phase space indicates poor modelling of the analysis' background shape and rate. However, if this deviation is constrained to a narrow mass region, it could be an indication of some signal being present. The branching fraction sensitivity for the channels with a single track on the signal side, $B^\pm \to \pi^\pm a$, $B^\pm \to K^\pm a$, and $B^\pm \to pa$ is $\mathcal{O}(10^{-6})$. For $B^\pm \to D_s^\pm a$ it is $\mathcal{O}(10^{-4})$

and for $B^0 \to \overline{D}^0 a$ it is $\mathcal{O}(10^{-5})$. The error on the expected sensitivity depends on the mass hypothesis, much like the sensitivity itself. This error is generally substantial, often over 100%. This indicates that we should expect the observed limit to fluctuate fairly substantially either side of the expected limit. The large variation arises from the relatively small amount of hard-to-parameterise backgrounds present in the analysis.

7.2 Results

The results shown below are produced after the entire analysis approach was finalised based on the MC datasets. There were no alterations to the analysis technique after the unblinding of the full dataset. Prior to fully unblinding, a random sample of 10% of real events was unblinded to check that background rates were consistent with expectation. This fractional dataset showed some disagreements between data and simulation in the phase space region corresponding to the heaviest possible m_a values for the channels $B^\pm \to \pi^\pm a$ and $B^\pm \to K^\pm a$. This most likely corresponds to poor modelling of low-momentum pions and kaons in the background somewhere, and cannot come from the presence of the signal decay as the disagreement covers too wide an area of phase space. These phase space regions were ignored in the final mass scan, meaning that for the channel $B^\pm \to \pi^\pm a$, the highest mass in the scan is reduced from 5.048 GeV/c^2 to 4.977 GeV/c^2, and for $B^\pm \to K^\pm a$, reduced from 4.768 GeV/c^2 to 4.702 GeV/c^2.

7.2.1 Signal extraction



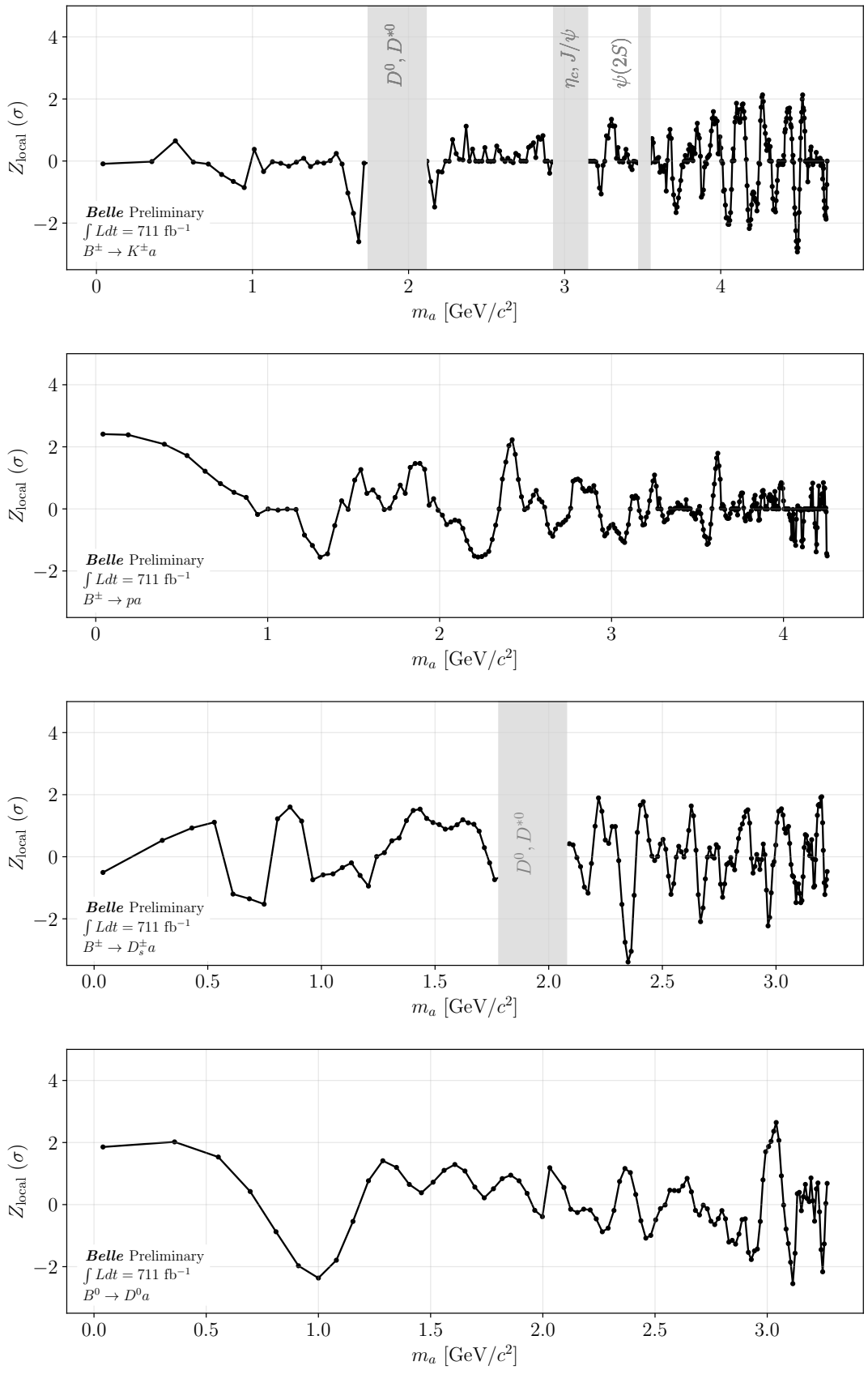


Figure 7.1: Local significance as a function of mass for all five channels.

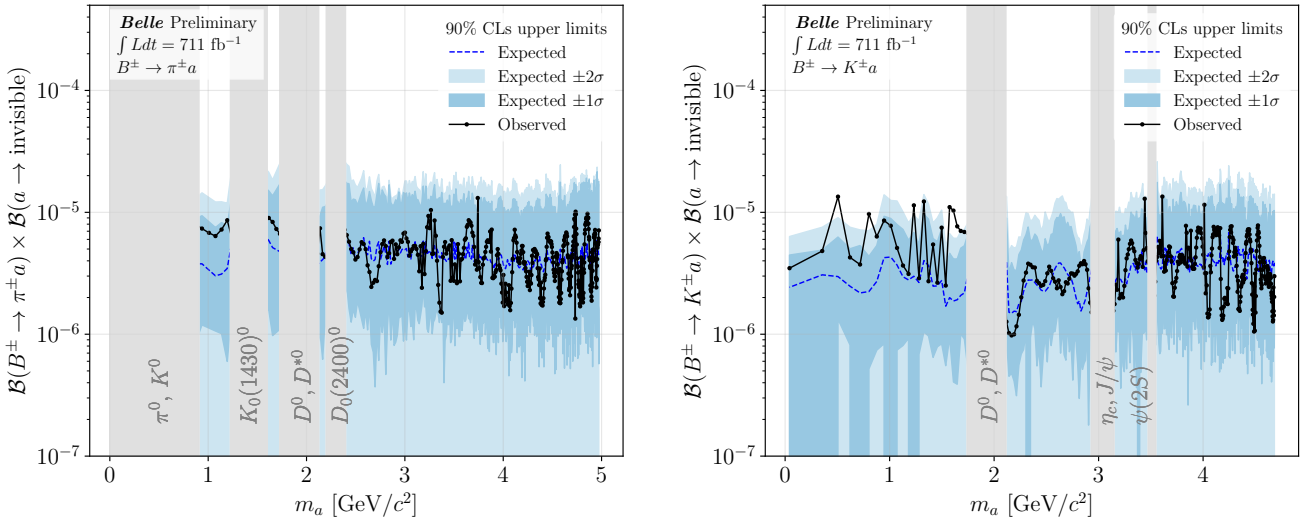
The mass scan is run over the unblinded dataset, and the local significance is extracted for each mass point. The results of this signal extraction are shown in Fig. 7.1. The largest positive local significance is $Z_{\text{local}} = 2.95\sigma$ and occurs at $m_a = 3.28\,\text{GeV}/c^2$ in the search for the process $B^\pm \to \pi^\pm a$. This corresponds to a global significance of $Z_{\text{global}} = 0.28\sigma$ — far, far below the discovery significance. Therefore, in all five channels, we see no significant excess over the expected SM background.

7.2.2 Branching fraction upper limits

The observed branching fraction upper limits are shown in Fig. 7.2, overlaid on top of the expected limits. The observed limit is higher than expected below $m_a \approx 2\,\text{GeV}/c^2$, which is where the largest mismodelling of continuum background occurs. Otherwise, the observed limits are all fairly consistent with the expected limits, with deviations mostly within 1σ of the expectation.

We can map the observed upper limits for the nominal $a \to$ invisible assumption in Fig. 7.2 to limits that assume $a \to$ visible decays by replacing the nominal efficiency used in Equation (5.23) with the long-lived particle efficiency from Fig. 4.19. This yields efficiencies that decrease as lifetime decreases. We compute this for lifetimes between $c\tau_a = 0.01$ mm and $c\tau_a = 10000$ mm. For lifetimes greater than the latter, decays within the detector are sufficiently rare that we can consider the a invisible.

Many other experimental results are shown at the 95% CL, rather than the 90% CL. The results of this search are also presented at the 95% CL in Fig. 7.4. The 95% CL upper limits are, as one would expect, slightly higher than the 90% CL results shown in Fig. 7.3, usually by a factor of between 1.2 and 1.5.



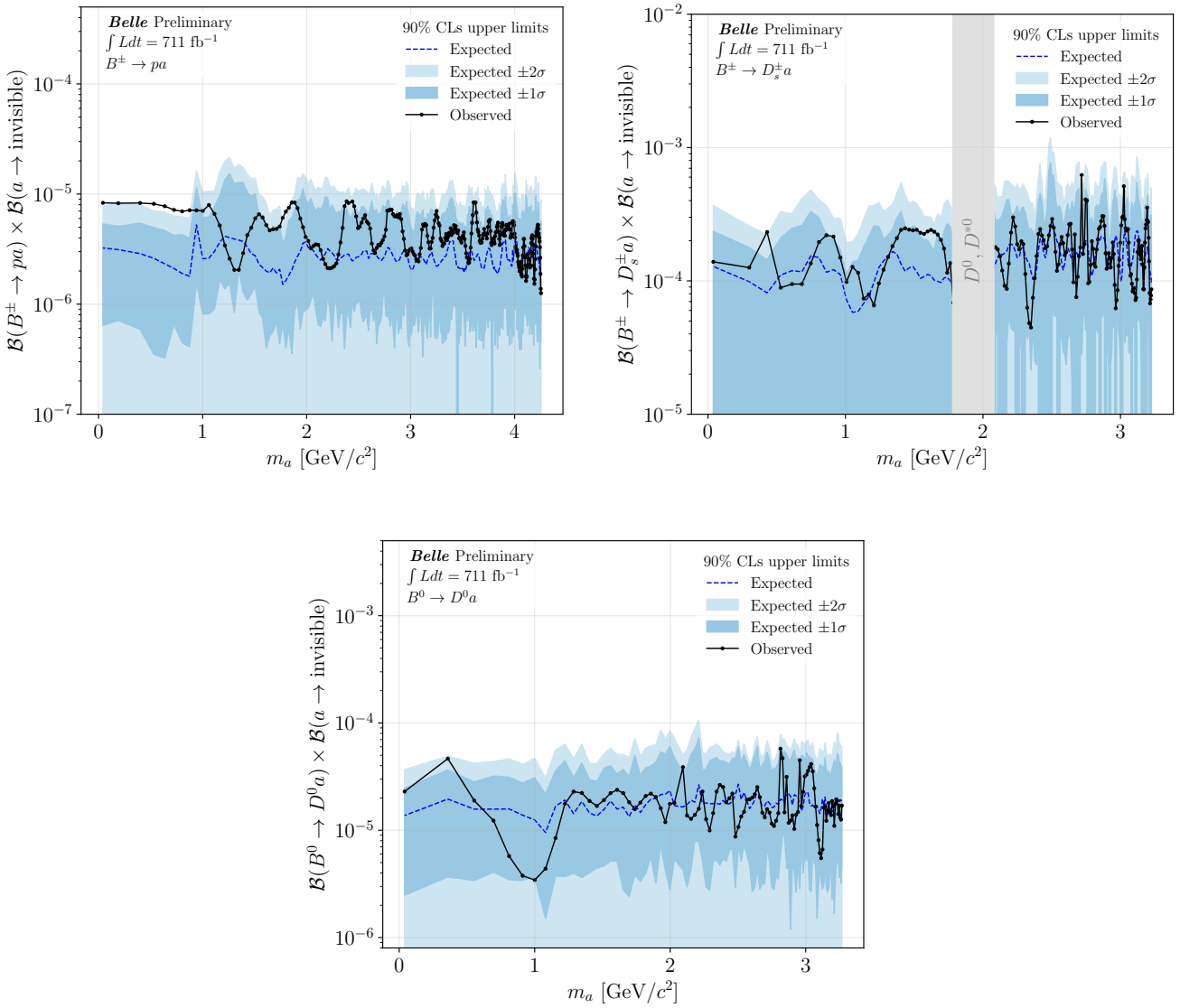


Figure 7.2: Expected (blue) and observed (black) 90% confidence level upper limits on the branching fractions of all five channels assuming a is invisible. The blue shaded areas indicate the 1 and 2σ regions around the expected limit.

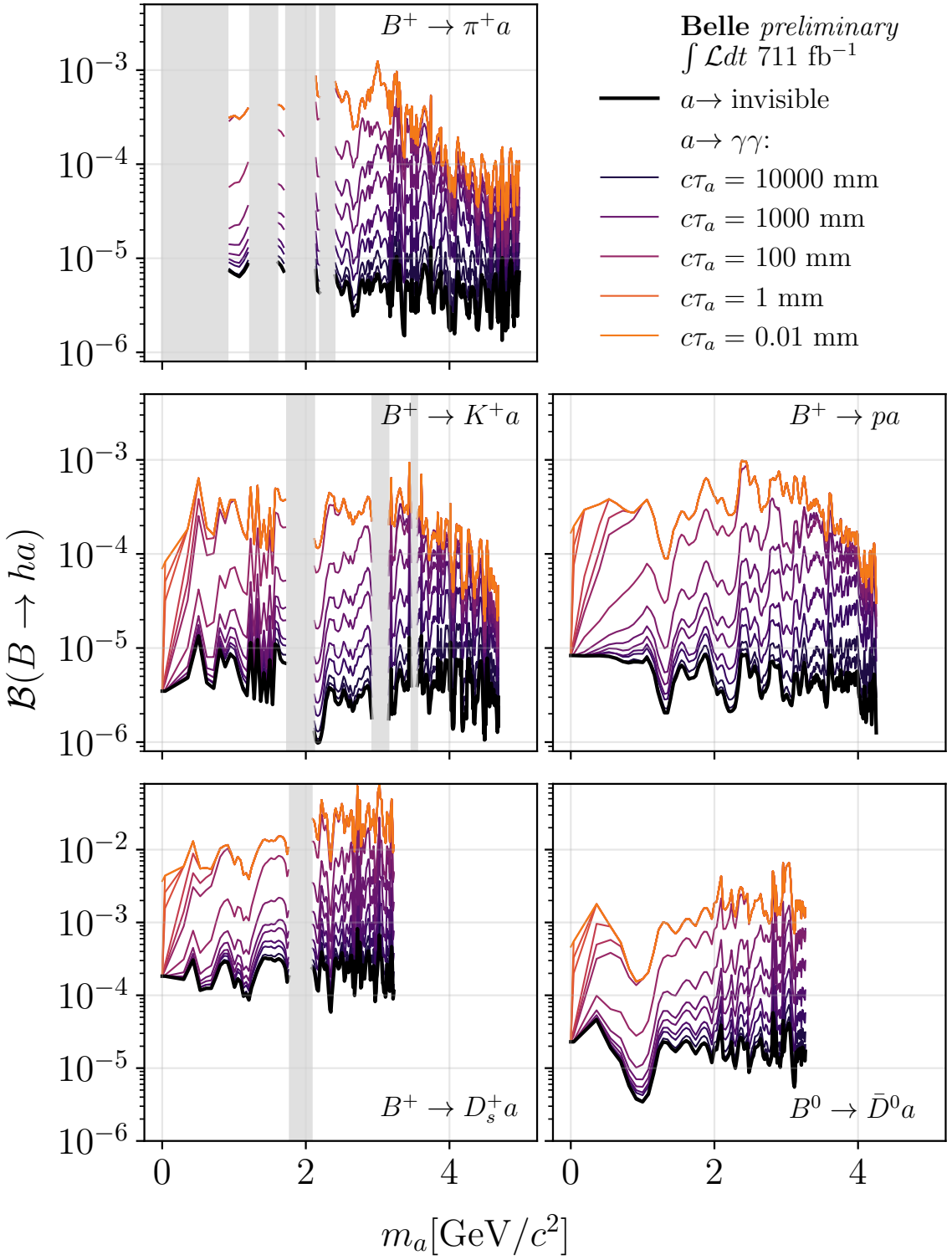


Figure 7.3: Observed 90% confidence level upper limit on the branching fractions of all five channels assuming a decays invisibly (black) or visibly and has some finite lifetime (coloured).

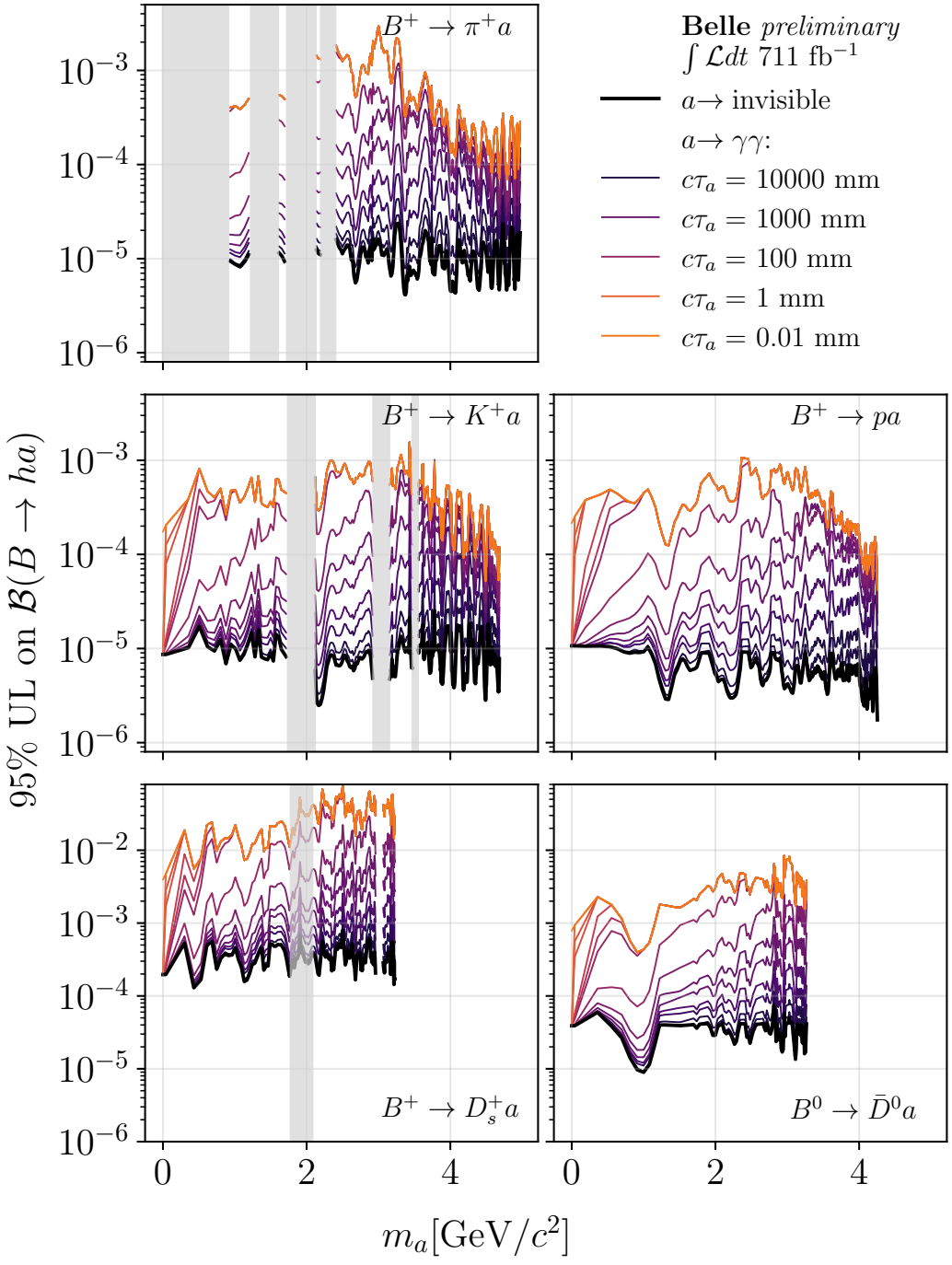


Figure 7.4: Observed 95% confidence level upper limit on the branching fractions of all five channels assuming a decays visibly and has some finite lifetime.

7.3 Comparisons to other searches

In Chapter 8, the implications of these experimental results will be discussed and compared to those from other experiments. In this section, we directly compare the results with measurements of the same channels, or at least the closest proxy to them. The comparison is summarised in Table 7.1. For $B^{\pm} \to \pi^{\pm} a$ and $B^{\pm} \to K^{\pm} a$, our result supersedes recast results from $B^{\pm} \to \pi^{\pm} \nu\overline{\nu}$ and $B^{\pm} \to K^{\pm} \nu\overline{\nu}$ searches from Belle [203, 204] and BaBar [96]. As these searches were both for three-body final states, no mass-dependent branching fraction exclusion is available.

The only experimental search for a two-body decay exactly like one of the channels searched for in this work is from a recent study from BaBar, searching for $B^{\pm} \to pa$[117]. That study's branching fraction exclusion was highly dependent on the mass of the invisible particle, with the strictest limit excluding branching fractions down to 8×10^{-7} (at the 90% CL) for invisible particle masses of between 1.5 and 1.6 GeV/c^2. For most of the rest of the accessible phase space, the BaBar search was far less sensitive than the search presented in this work, hovering between the $10^{-5} - 10^{-6}$ level.

Currently, no searches have been conducted for the channels $B^{\pm} \to D_s^{\pm} a$ and $B^0 \to \overline{D}^0 a$. The most closely related SM decays, $B^{\pm} \to D_s^{\pm} \nu\overline{\nu}$ and $B^0 \to \overline{D}^0 \nu\overline{\nu}$, are highly suppressed and have not been discussed in any literature. This work presents the world's first-ever experimental results for these two channels.

Table 7.1: Comparison between our branching fraction exclusion results and the reaches of similar searches. This study remains the sole search for $B^{\pm} \to D_s^{\pm} a$ and $B^0 \to \overline{D}^0 a$, while there are indirect limits on the pion and kaon channels from searches for the SM decays $B^{\pm} \to h\nu\bar{\nu}$ with $h \in \pi^{\pm}, K^{\pm}$. A recent search by BaBar shows a competitive result in the channel $B^{\pm} \to p^{\pm} a$.

Decay measured in this work	Previous study	Previous world best	This work
$B^{\pm} \to \pi^{\pm} a$	Belle $B^{\pm} \to \pi^{\pm} \nu\overline{\nu}$ [203]	1.4×10^{-5}	$\sim 10^{-6}$
$B^{\pm} \to K^{\pm} a$	Belle $B^{\pm} \to K^{\pm} \nu\overline{\nu}$ [203]	1.6×10^{-5}	$\sim 10^{-6}$
$B^{\pm} \to p^{\pm} a$	BaBar $B^{\pm} \to p^{\pm} a$ [117]	$10^{-5} - 10^{-6}$	$\sim 10^{-6}$
$B^{\pm} \to D_s^{\pm} a$	N/A	N/A	$\sim 10^{-4}$
$B^0 \to \overline{D}^0 a$	N/A	N/A	$\sim 10^{-5}$

8

New physics interpretations

So far, this thesis has presented a detailed description of the analysis procedure used to search for new-physics B meson decays. Here, we consider the physics models that inspired the experimental search and show the exclusions to their parameter space implied by the results. There are assumptions that must go into every two-dimensional exclusion plot, as there are always several free parameters. However, some benchmark assumptions found elsewhere in the literature have been chosen so that this search can be compared to other experiments in the context of several interesting models.

8.1 ALP couplings (pseudoscalar portal)

The brief literature review in Chapter 2 gives the detail needed to map the branching fraction limits on $B^{\pm} \to \pi^{\pm}/K^{\pm}a$ to the a-W coupling g_{aW}. This section will provide more detail on the calculation of the coupling g_{aW}, and how it relates to the ALP's couplings to photons and the Z boson. In Ref. [205], the Lagrangian introduced in Ref. [42] for an ALP coupled to the $\mathrm{SU}(2)_W$ gauge bosons W^a_μ and the $\mathrm{U}(1)_Y$ gauge boson B_μ is shown alongside couplings to the electroweak gauge bosons. It is written as

$$\mathscr{L} = \frac{1}{2}\partial_\mu a \partial^\mu a - \frac{1}{2}m_a^2 a^2 - c_{WW}\frac{a}{f_a}W^a_{\mu\nu}\tilde{W}^{a\mu\nu} - c_{BB}\frac{a}{f_a}B_{\mu\nu}B^{\mu\nu} \tag{8.1}$$

where the ALP mass and decay constant are m_a and f_a respectively, and are free parameters in the model, along with the coupling constants to the W^a_μ and B_μ fields c_{WW}

and c_{BB}. In this treatment, the gauge bosons are defined as

$$\begin{aligned} W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc} W^b_\mu W^c_\nu, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu. \end{aligned} \tag{8.2}$$

After electroweak symmetry breaking, the coupling constants for the ALP's interactions with the electroweak gauge bosons are related to the coupling constants to the fields in Equation (8.1) and the ALP decay constant f_a. The relations are

$$\begin{aligned} g_{a\gamma\gamma} &= \frac{4}{f_a}(s^2_{\theta_W} c_{WW} + c^2_{\theta_W} c_{BB}), \\ g_{aZ\gamma} &= \frac{2}{f_a}(c_{WW} - c_{BB}) s_{2W}, \\ g_{aZZ} &= \frac{4}{f_a}(c^2_{\theta_W} c_{WW} + s^2_{\theta_W} c_{BB}), \\ g_{aW} &= \frac{4}{f_a} c_{WW}. \end{aligned} \tag{8.3}$$

The top three couplings in Equation (8.3), under the assumption in Ref. [61] that $c_{BB} = 0$ (i.e. the ALP only interacts with the W field), can be rewritten in terms of g_{aW},

$$\begin{aligned} g_{a\gamma\gamma} &= g_{aW} s^2_{\theta_W}, \\ g_{aZ\gamma} &= \frac{g_{aW}}{2} s_{2\theta_W}, \\ g_{aZZ} &= g_{aW} c^2_{\theta_W}, \end{aligned} \tag{8.4}$$

where $s_{\theta_W} = \sin\theta_W$, $c_{\theta_W} = \cos\theta_W$, and θ_W is the Weinberg angle.

The decay rate for $B^\pm \to K^\pm a$ is given by Equation (2.9). To compute the couplings numerically, we use $b \to s$ and $b \to d$ form factors calculated from fits to light-cone sum rules in Ref. [62]. The form factor parameterisation has the form of Equation (2.29), with results for the fit parameters given in Table 2.3. To convert the branching fraction limits we presented in Fig. 7.2, we multiplied the branching fraction by the total decay width of the B meson in GeV ($\Gamma_{B,\ \mathrm{Total}} = 4.02 \times 10^{-13}\,\mathrm{GeV}$ using $\tau_B = 1638 \times 10^{-15}$ s [7] and $\Gamma = \hbar/\tau$) and rearranged Equation (2.9) for g_{aW}, yielding

$$g_{aW} = \left[\frac{64\pi \mathcal{B}(B^\pm \to K^\pm a)\Gamma_B}{M^3_{B^+}(1 - \frac{M^2_{K^+}}{M^2_{B^+}})^2 \lambda^{\frac{1}{2}}_{K^+,a}}\right]^{\frac{1}{2}} \times \left[\frac{3\sqrt{2} M^2_W G_F \sum_{\alpha\in c,t} V_{\alpha b} V_{\alpha s} f(M^2_\alpha/M^2_W)}{16\pi^2}\right]^{-1}. \tag{8.5}$$

These couplings can also be related to the rate $\Gamma(B^\pm \to \pi^\pm a)$, by replacing the kaon mass with the pion mass, $V_{\alpha s}$ with $V_{\alpha d}$, and the $B \to K$ form factor with the $B \to \pi$ form factor from Ref. [62]. Using this relation between the ALP production branching fraction and the coupling, we place 90% CL upper limits on g_{aW} assuming an invisible

ALP, shown in Fig. 8.1. For most of the mass range accessible by this search, $B^{\pm} \to K^{\pm}a$ provides the strictest limits on g_{aW} coupling, while $B^{\pm} \to \pi^{\pm}a$ provides the strictest limits in the mass region between 4.7 and 5.0 GeV/c^2.

The exclusion limits from this search do not have the reach of the kaon experiments E787, E949, and NA62; however, these experiments can only search for invisible ALPs up to $m_a = 260\,\text{MeV}/c^2$. The g_{aW} equivalent to the branching fraction excess seen by the Belle II search for $B^+ \to K^+\nu\bar{\nu}$ hadronic tag analysis (HTA) is almost completely excluded by our results.

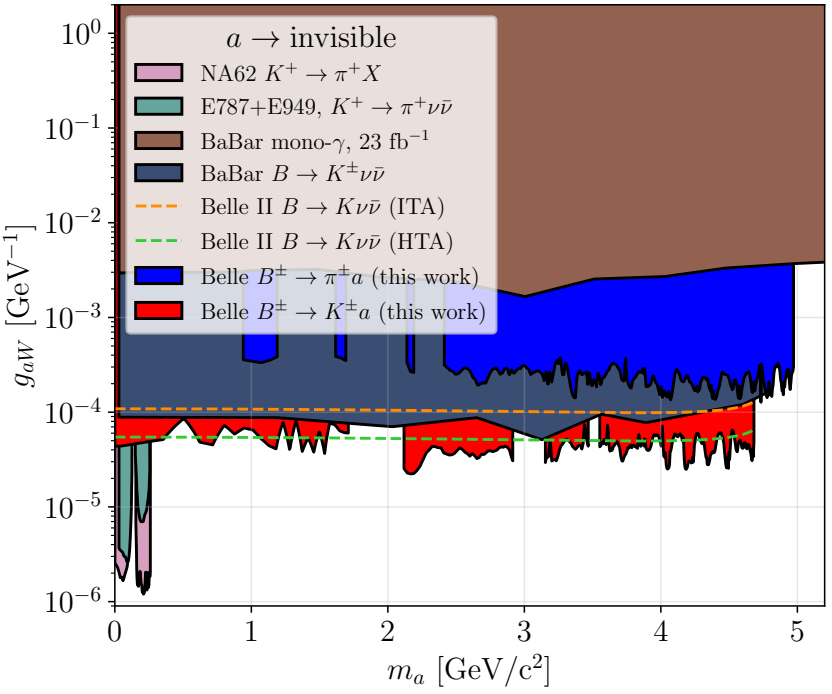


Figure 8.1: A recreation of Fig. 2.7, assuming $a \to$ invisible with the branching fraction exclusions from this work included.

Instead of the invisible ALP case, we can consider ALPs decaying to SM particles. For our analysis to be sensitive to these decays, the ALP must have a finite lifetime so that enough events escape detection. In this case, we assume the ALP decay width to be

$$\Gamma_a = \frac{g_{aW}^2 m_a^3 s_{\theta_W}^4}{64\pi}, \tag{8.6}$$

which follows the model used in a previous BaBar search [206] to set limits on g_{aW}. As the ALP decay width in Equation (8.6) is proportional to g_{aW}^2, the ALP lifetime is proportional to $1/g_{aW}^2$. Our sensitivity scales with the lifetime, so for shorter lifetimes, the limits here will not be as competitive as the invisible case. For the lifetime hypotheses that we generate signal MC for, we calculate the corresponding curves in the g_{aW}–m_a parameter space. Using the predicted sensitivity to $\mathcal{B}(B^{\pm} \to K^{\pm}a) \times \mathcal{B}(a \to \gamma\gamma)$ for

different lifetimes, like those shown in Fig. 7.3, we check to see where the experimental limit is lower than the prediction. The region where this is the case is our exclusion region. For any given value of g_{aW}, and therefore ALP proper lifetime, the lighter ALPs are more boosted, so our sensitivity is best for low masses as they are most likely to escape the detector. The exclusion plot for g_{aW} under these assumptions is shown in Fig. 8.2.

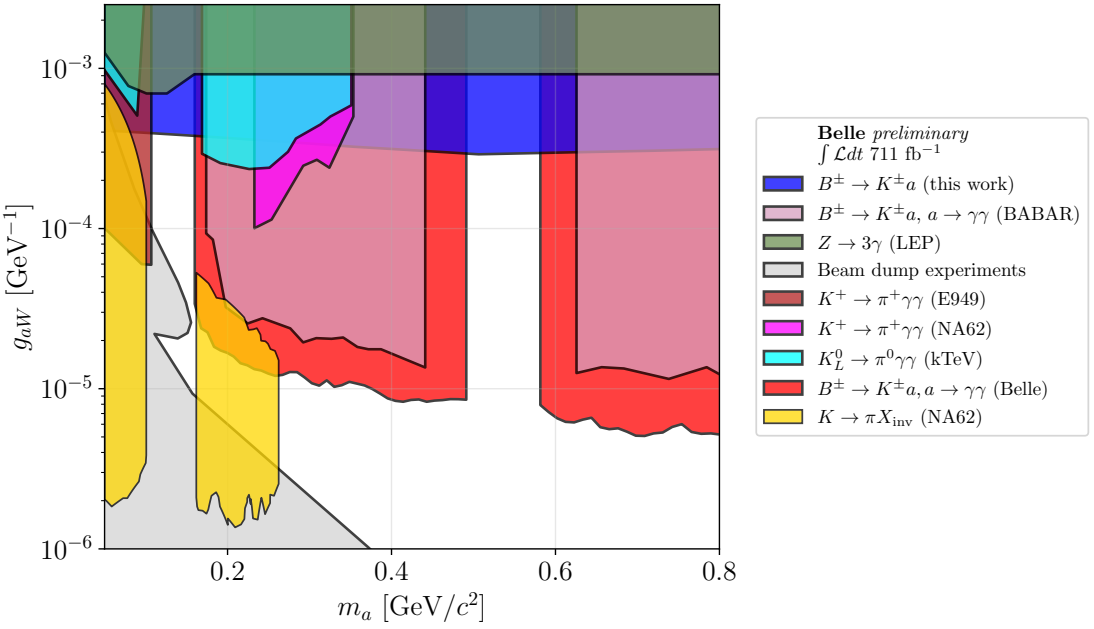


Figure 8.2: The g_{aW} parameter space up to $m_a = 800\,\text{MeV}/c^2$ assuming $a \to \gamma\gamma$ is the primary ALP decay mode.

The range of masses shown in Fig. 8.2 is smaller compared to that of Fig. 8.1 as the sensitivity drops off dramatically after $m_a = 1\,\text{GeV}/c^2$ due to the smaller Lorentz boost of the heavier ALPs. Given the assumption that the ALP decay rate is described by Equation (8.6), this search can exclude some small regions of parameter space that are unconstrained by other experiments, particularly in the gaps left by the π^0 and K^0 vetoes in BaBar's search for visible ALPs [206]. At low masses, beam-dump experiments dominate the visible ALP parameter space.

In the scenario where the ALP does not interact with the hypercharge gauge boson ($c_{BB} = 0$), the ALP-photon coupling $g_{a\gamma\gamma}$ is directly related to g_{aW} by the Weinberg angle, as shown in Equation (8.4). As $\sin^2\theta_W = 0.22305$ [207], g_{aW} exclusions are enhanced by a factor of about 4.5 when translated into $g_{a\gamma\gamma}$ exclusions. The advantage of using this coupling as a comparison is the ease with which one can contextualise a result, as the $m_a - g_{a\gamma\gamma}$ plane is by far the most commonly discussed ALP parameter space in the literature. We present this limit in Fig. 8.3, where other $g_{a\gamma\gamma}$ bounds come from the Axion Limits repository [99]. The limits in this mass range come from $B^\pm \to K^\pm a$ up to 4.702 GeV/c^2, and from $B^\pm \to \pi^\pm a$ from 4.702 GeV/c^2 to 4.977

GeV/c^2, which is the mass range only accessible to the pion channel. It should be noted that all other limits in Fig. 8.3 solely rely on the $g_{a\gamma\gamma}$ coupling for their ALP production, whereas the limits from the search presented in this thesis are indirectly inferred from g_{aW}. Despite this more contrived scenario, it does help to further contextualise the result with relation to the ALP parameter space — there are many more experimental results present in Fig. 8.3 than in Fig. 8.1.

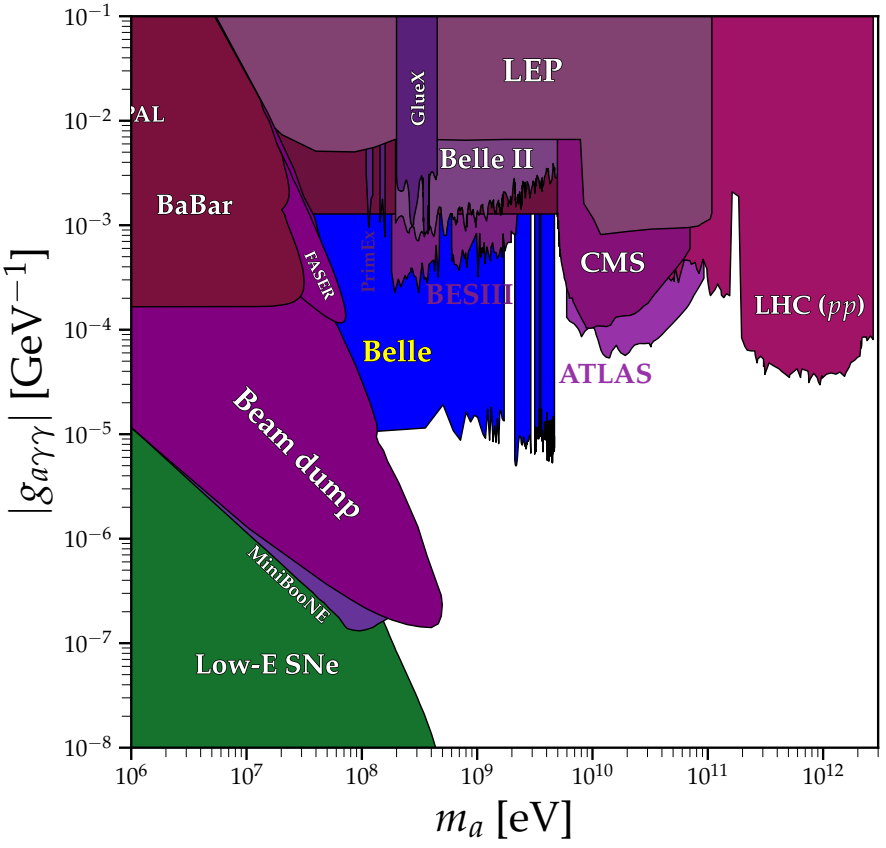


Figure 8.3: Heavy ALP parameter space from Ref. [99], with the limits from this thesis shown in Belle blue and yellow.

While discussion of the loop-level decays $B^\pm \to \pi^\pm a$ and $B^\pm \to K^\pm a$ is not uncommon in the literature, ALPs can be probed through other types of decay. There has been recent work conducted looking at ALP production by quark annihilation [2]. In these decays, the quarks in the B meson annihilate, and an ALP is emitted from one of the quarks in the interaction. There are two key advantages to looking for ALPs produced in these kinds of processes: one, that there are many final-state hadrons that can be produced in association with the ALP; and two, that the process can occur at tree level. The calculations relating to these decays are done with perturbative QCD (pQCD) in Ref. [2], and give numerical results that relate annihilation decay branching fractions to g_q, the ALP's coupling to some quark q. The results that are applicable to this study are summarised in Table 2.1. In Ref. [2], only $a \to$ invisible and $a \to \gamma\gamma$ decays are considered. Using the results in Table 2.1, we map the expected branching fraction

limits in Fig. 7.2 to the ALP-quark couplings for $q \in u, d, s, c, b$, shown in Fig. 8.4. This treatment of ALP decays highlights the advantage of searching for the 'new' channels $B^{\pm} \to \pi^{\pm} a$, $B^{\pm} \to D_s^{\pm} a$, and $B^0 \to \overline{D}^0 a$, as the most stringent limits for each of the g_q couplings come from these channels.

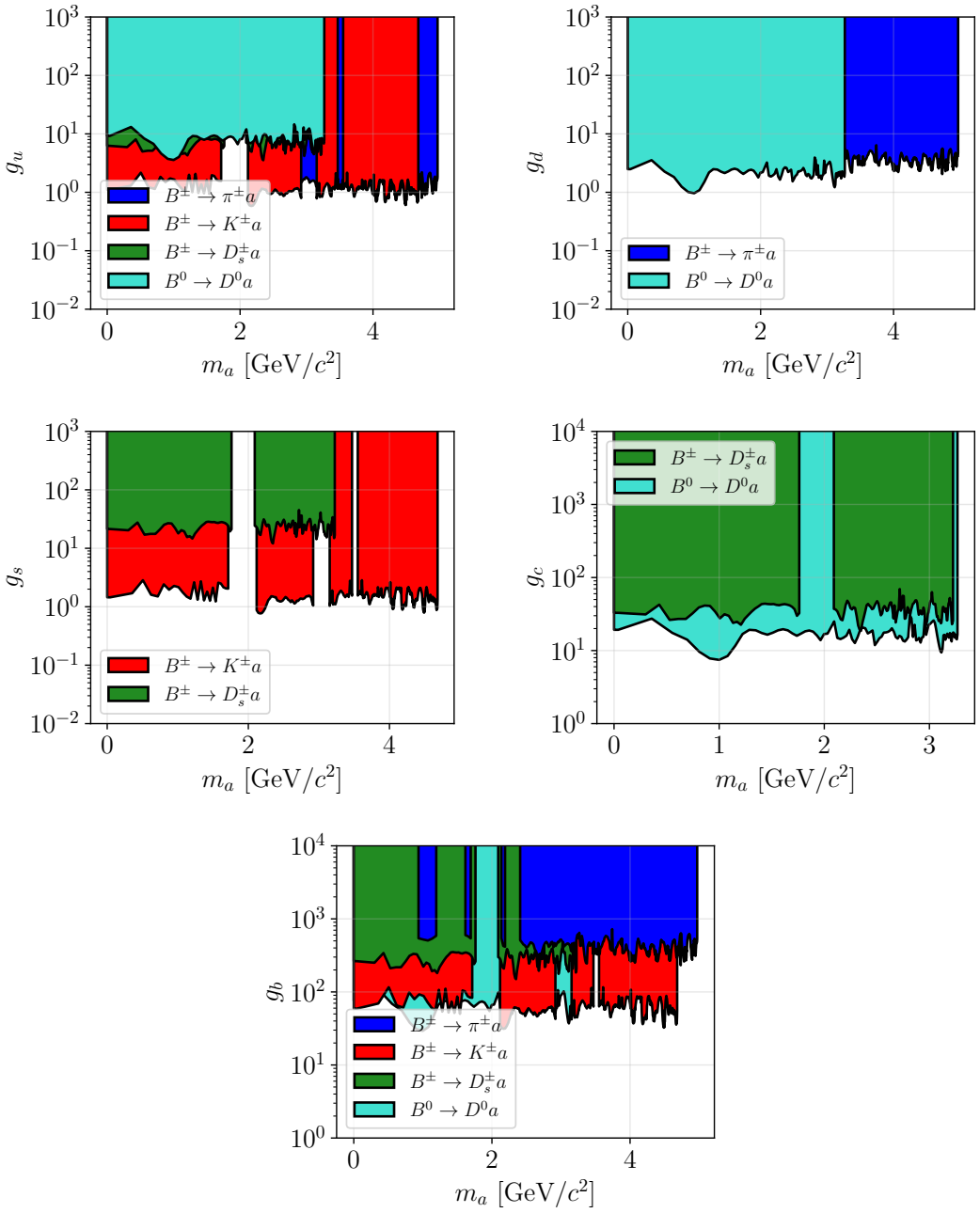


Figure 8.4: Limits on ALP-quark couplings following the ALP decay model in Ref. [2].

8.2 Dark scalar couplings (scalar portal)

The search for $B^{\pm} \to K^{\pm}a$ decays, where a is invisible, could also probe the decay $B^{\pm} \to K^{\pm}S$, where S is a hypothetical dark scalar particle, as proposed in Ref. [74]. This decay channel could appear invisible in two scenarios: first, if S has significant couplings to dark fermions that are kinematically accessible decay products, or second, if S decays to SM particles but does so outside the detector volume due to its long lifetime.

A key aspect of this dark scalar model is the mixing angle θ, which appears in Equation (2.22) and governs the fraction of decays to visible (SM) and hidden sector ($\chi\chi$) final states. The branching fraction for the production of dark scalars in B meson decays is given by Equation (2.27). Notably, both the production rate of dark scalars in B decays and their decay rate when $\Gamma_{\chi\chi} = 0$ are proportional to $\sin^2\theta$. The branching fractions to specific SM final states are documented in Refs. [74, 86]. For decays to dark fermions χ, the branching fraction is given by Equation (2.23), while the branching fractions to charged leptons, pions, and kaons are given by Equations (2.24) to (2.26).

We provide two separate exclusion plots in the $m_S - \theta$ plane, corresponding to the two benchmark cases used to make similar plots in Ref. [74]. The first case, which we will refer to as the *invisible* case, assumes that $m_\chi = 0$ and $y_\chi = 1$, which maximises the decay rate of S to the dark sector for all phase space. With these assumptions, the visible decay rate Γ_{SM} is always several orders of magnitude smaller than the invisible decay rate $\Gamma_{\chi\chi}$, meaning $\mathcal{B}(S \to \chi\chi) \approx 1$. Thus, the limit for $B^{\pm} \to K^{\pm} + \text{invisible}$ shown in Fig. 7.2 is directly interpreted as $\mathcal{B}(B^{\pm} \to K^{\pm}S) \times \mathcal{B}(S \to \chi\chi)$.

The second case, which will be referred to as the *long-lived particle* case, assumes that invisible decays are forbidden. This situation could be due to $S \to \chi\chi$ decays being kinematically forbidden ($m_S < 2m_\chi$) or S not coupling to the dark fermions ($y_\chi = 0$). Mapping the limit in Fig. 7.2 to the $m_S - \theta$ plane is less trivial here. For each mass point in the scan, we interpolate between the selection efficiencies for long-lived particles as shown in Fig. 4.19. The interpolation is used to update the branching fraction limits in Fig. 7.2 to the limits in Fig. 7.3 for long-lived particles. An interpolation between branching fraction sensitivities is shown as the blue lines on Fig. 8.6. The product of branching fractions $\mathcal{B}(B^{\pm} \to K^{\pm}S) \times \mathcal{B}(S \to xx)$ (for $x \in e, \mu, \pi, K$) is also a function of the lifetime and m_S. Thus, for each mass point, we calculate the predicted rate as a function of lifetime, shown as the purple lines on Fig. 8.6. Then, for each mass we have done this for, we check the range of lifetimes for which our experimental limit is smaller than what is predicted by the model. These ranges constitute our exclusion region. As the dark scalar lifetime τ_S is directly related to the mixing angle θ, we then convert these exclusions to the $m_S - \theta$ plane. The dark scalar lifetime is a function of its mass

and mixing parameter. In the phase space region where the scalar is only allowed to decay to electrons, τ_S can be written as

$$\tau_S = \frac{\hbar}{\Gamma_{\text{SM}}} = \frac{8\pi\hbar v}{m_e^2 s_\theta^2 m_S}\left(1 - \frac{4m_e^2}{m_S^2}\right)^{-3/2}, \tag{8.7}$$

where s_θ is the sine of the mixing angle θ and $v = 246\,\text{GeV}$ is the Higgs vacuum expectation value. In this region, the total decay width Γ_S is just the decay width to electrons, which can be read off Equation (2.24) for $\ell = e$ (ignoring the factor of Γ_S so it is just the partial width). It can be generalised by using the sum of all possible visible decay rates using Equations (2.24) to (2.26) in a similar way. The scalar lifetime is shown in the $m_S - \theta$ plane in Fig. 8.5. There are a few regions with distinct jumps in lifetime. These occur when new final states become available, such as the one at approximately $m_S = 2m_\mu$. There is also a slight discontinuity at $m_S = 2.0\,\text{GeV}/c^2$, where the total decay width calculation transitions from sums of exclusive final states to an inclusive rate to hadrons. The predicted branching fractions and exclusion limits as a

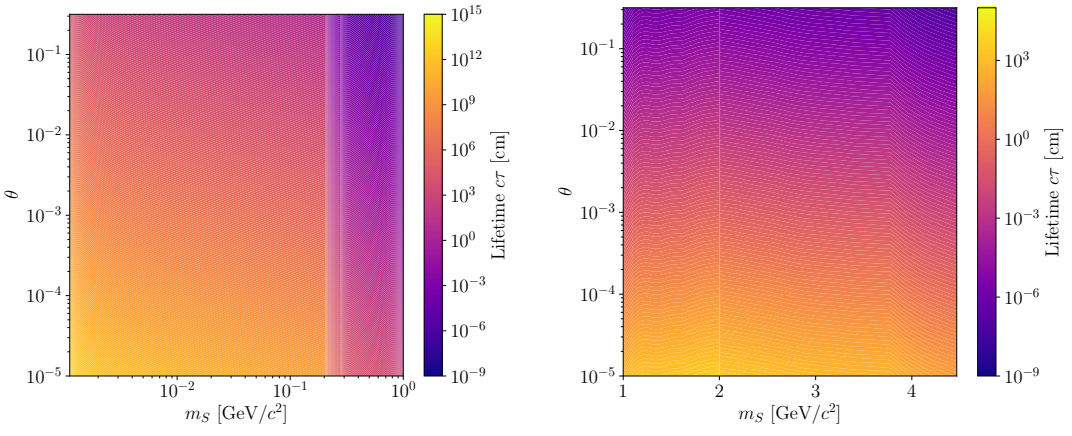
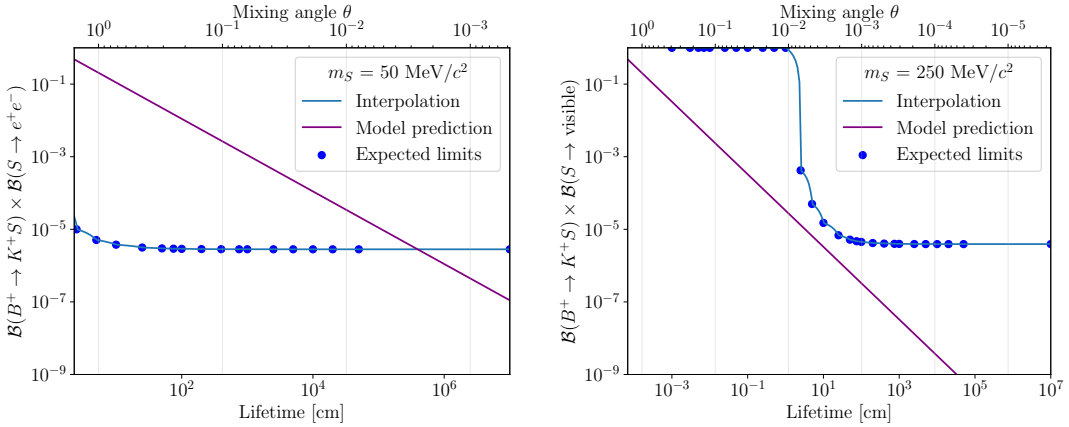


Figure 8.5: Dark scalar decay length $c\tau_S$ in the mass-mixing angle plane, (a) between 1.12 MeV/c^2 and 1.0 GeV/c^2 and (b) between 1.0 GeV/c^2 and 4.78 GeV/c^2. The low mass region is shown separately as the decay width in that region is much smaller, which leads to larger lifetimes.

function of lifetime are shown in Fig. 8.6 for two example masses, $m_S = 50\,\text{MeV}/c^2$ and $m_S = 250\,\text{MeV}/c^2$. Sensitivity is best at low masses, as the scalar is more boosted, while at higher masses, it decreases, with any sensitivity effectively lost past $m_S = 300\,\text{MeV}/c^2$. In the $m_S = 50\,\text{MeV}/c^2$ case, the only kinematically allowed scalar decay is $S \to e^+e^-$, while at $m_S = 250\,\text{MeV}/c^2$, both electron and muon final states are possible. The lifetime is expressed in centimetres on the bottom axis of Fig. 8.6, while the corresponding mixing angle θ is shown on the top axis. Since $\sin\theta$ is maximised at $\theta = \pi/2$, we impose a cut-off in the plots to prevent degenerate cases occurring when $\theta > \pi/2$. The excluded



(a) (b)

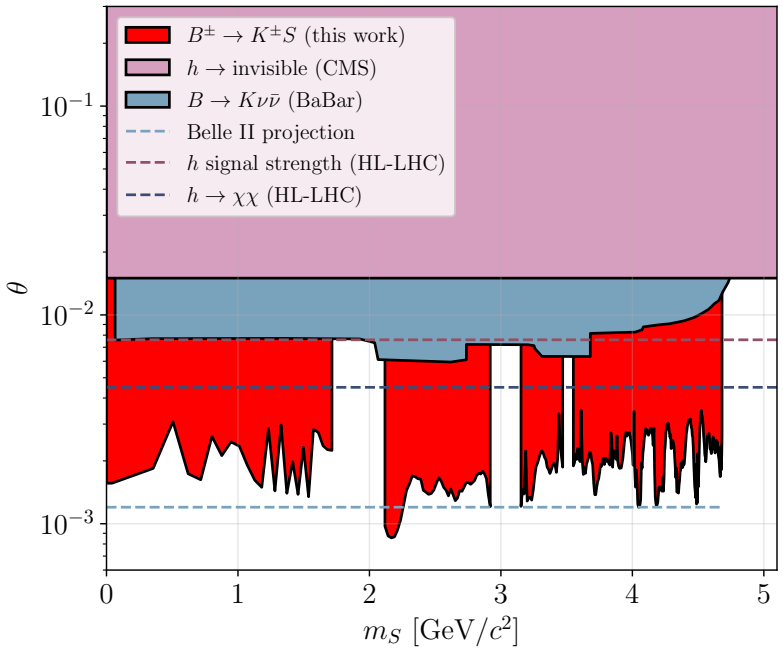
Figure 8.6: Exclusion limits (blue) and model predictions (purple) for branching fractions as a function of scalar lifetime, for (a) $m_S = 50\,\text{MeV}/c^2$ and (b) $m_S = 250\,\text{MeV}/c^2$. The excluded regions correspond to lifetimes where the model prediction exceeds the expected limit.

regions for a given mass correspond to the lifetimes where the model prediction, shown in purple in Fig. 8.6, exceeds the blue limit.

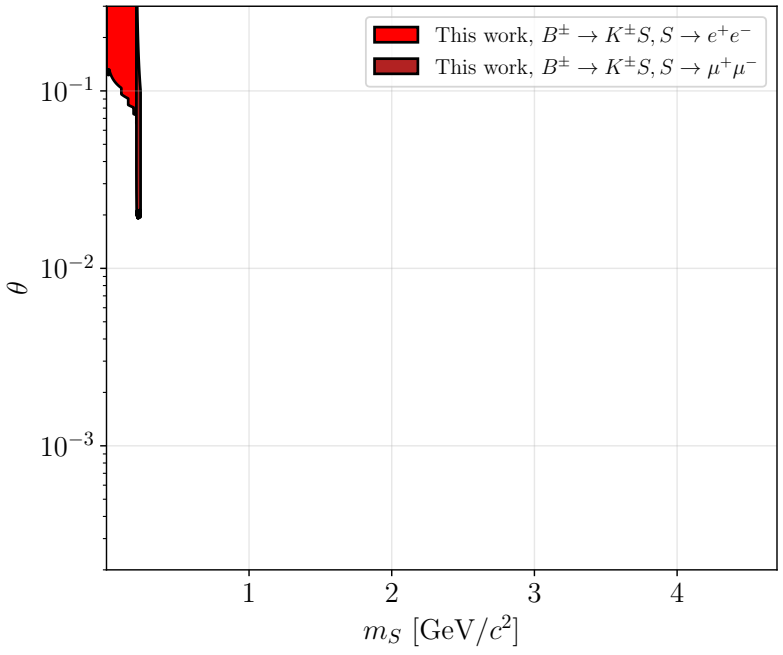
The exclusion limits are shown in Fig. 8.7 in a similar manner to the plots shown in Ref. [74], including some relevant results and predicted sensitivities from other experiments. For the invisible case, this work is more sensitive than the projected High-Luminosity LHC (HL-LHC) sensitivities that probe SM Higgs decays shown in Ref. [74], but falls short of the expected sensitivity Belle II can achieve with $50\,\text{ab}^{-1}$. The HL-LHC will produce a massive sample of Higgs bosons. As S should mix with the SM Higgs boson h, it induces invisible decays of h. Furthermore, this mixing also affects the Higgs coupling to all visible particles, suppressing the ratio

$$\frac{\sigma_h \times \mathcal{B}(h \to \text{visible})}{\sigma_h \times \mathcal{B}(h \to \text{visible})_{\text{SM}}}, \tag{8.8}$$

where σ_h is the Higgs production rate, by a factor of $\cos\theta$. As these searches are less direct with regard to the dark scalar, Belle remains very competitive. The visible case covers much less parameter space, and is best at low masses, where S is more boosted and therefore less likely to decay within the detector volume. We are sensitive to $2m_e < m_S < 2m_\pi$, the lower bound of which pushes to lighter masses than is shown in much of the literature; however, in general, it is not competitive with limits from dedicated long-lived particle searches.



(a)



(b)

Figure 8.7: Exclusion regions for the dark scalar model for (a) the invisible case and (b) the long-lived particle case. For the latter, we are only sensitive at low masses of less than $m_S \approx 230\,\mathrm{MeV}/c^2$.

8.3 B-mesogenesis parameter space

The B-mesogenesis mechanism relies on the interplay of B meson decays into dark baryons and the asymmetries between B and $\bar{B}$ mesons. The parameter space defined by the semileptonic asymmetries A^d_{SL} and A^s_{SL} is explored in Ref. [66]. These asymmetries are quantified with respect to B meson decay widths to some final states f as follows:

$$A^q_{\mathrm{SL}} = \frac{\Gamma(\overline{B}^0_q \to B^0_q \to f) - \Gamma(B^0_q \to \overline{B}^0_q \to f)}{\Gamma(\overline{B}^0_q \to B^0_q \to f) + \Gamma(B^0_q \to \overline{B}^0_q \to f)}. \tag{8.9}$$

To rule out the model, measurements must exclude parameter regions where branching fractions of B mesons to dark baryons align with semileptonic asymmetry values. These regions must also be consistent with the observed baryon asymmetry of the universe, $Y_{\mathcal{B}}$. The relation between these parameters is given by

$$Y_{\mathcal{B}} \simeq 8.7 \times 10^{-11} \frac{\mathrm{Br(B} \to \psi \mathcal{B} \mathcal{M})}{10^{-3}} \sum_q \alpha_q \frac{A^q_{\mathrm{SL}}}{10^{-3}} \tag{8.10}$$

where $\mathcal{B}$ denotes a baryon, $\mathcal{M}$ denotes one or more mesons, and the α_q are functions of the reheating temperature T_R. The reheating temperature is defined as the temperature of the early universe at which it becomes radiation-dominated. The baryon asymmetry of the universe is most favourably produced at temperatures $T_R \sim 10 - 20\,\mathrm{MeV}$. The details of how these α_q functions evolve with T_R are described in the Appendix of Ref. [66]. For the limits we produce in this work, we follow the same scenario used to set limits in Ref. [66] and assume $T_R = 20\,\mathrm{MeV}$.

In general, B-mesogenesis models propose effective interactions with vertices of the form $\bar{b} \to \psi_D ij$, where $i \in \{u, c\}$ and $j \in \{d, s\}$. These interactions are mediated by four operators, $\mathcal{O}_{\psi bij}$. For this analysis, only $\mathcal{O}_{\psi bud}$ is relevant, as the decay $B^\pm \to \psi p$ involves the vertex ψbud. However, decays involving dark baryons may proceed via any combination of the four operators, and B-mesogenesis could occur with just a single operator being turned on.

The ratio of branching fractions $\mathcal{B}(B \to \mathcal{B}_{ij}\psi)/\mathcal{B}(B \to \mathcal{B}_{ij}\psi\mathcal{M})$ is given in Ref. [66] and shown in Fig. 8.8. The limit on the branching fraction $\mathcal{B}(B^+ \to \psi p)$ presented in Fig. 7.2 can be mapped onto the N-body rate with all allowed final-state hadrons $\mathcal{B}(B \to \mathcal{B}_{ij}\psi\mathcal{M})$. This ratio depends on the *type* of operator chosen, which is discussed in more detail in Ref. [66]. *Type-1* operators have distinct phase space considerations compared to *type-2* and *type-3*. This distinction explains why the limits for type-1 are presented as separate contours. There is some choice in the b-quark mass used in a phase space integral when computing these ratios, so we use three choices from Ref. [66]: the nominal value is calculated with $m_b = m_b^{\mathrm{pole}} = 4.78\,\mathrm{GeV}/c^2$, with lower and upper bounds given by $\bar{m}_b = 4.18\,\mathrm{GeV}/c^2$ and $m_B = 5.279\,\mathrm{GeV}/c^2$. The ratios are

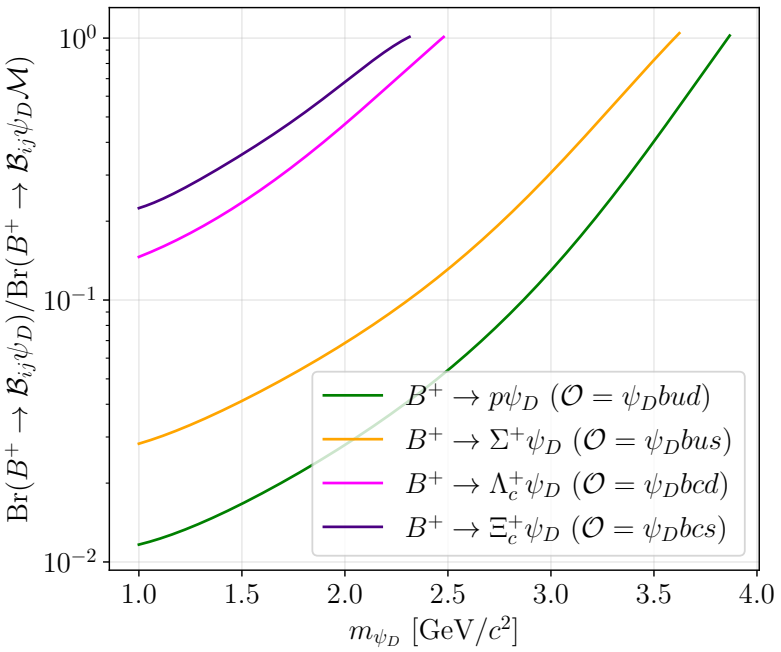


Figure 8.8: Ratios of two-body to N-body branching fractions for the four possible dark baryon decays of the charged B meson.

shown for an exhaustive list of b-flavoured hadrons in Figures 6 and 7 of Ref. [66]. The ratio $\mathcal{B}(B \to \mathcal{B}_{ij}\psi)/\mathcal{B}(B \to \mathcal{B}_{ij}\psi\mathcal{M})$ as a function of the dark baryon mass is shown in Fig. 8.8.

The limits this analysis places on the B-mesogenesis parameter space are shown in Fig. 8.9. Other literature has defined the area where B-mesogenesis is allowed as the area corresponding to where $\mathcal{B}(B \to \mathcal{B}_{ij}\psi\mathcal{M}) > 10^{-4}$.[1] The 2023 BaBar search for $B^+ \to p\psi_D$ excluded B-mesogenesis from dark baryons heavier than about 3 GeV/c^2. Our work builds on this, excluding the possibility that these decays are responsible for baryogenesis for dark baryon masses down to about 2 GeV/c^2.

The allowed B-mesogenesis region commonly used in the literature of $\mathcal{B}(B \to \mathcal{B}_{ij}\psi\mathcal{M}) > 10^{-4}$ is a conservative estimate that only roughly corresponds to the $\pm 2\sigma$ A_{SL}^q region. In Ref. [66], this parameter space is shown as a rectangle in the $A_{\mathrm{SL}}^s - A_{\mathrm{SL}}^d$ plane. Fig. 8.10 shows this parameter space, including the PDG averages for the semileptonic asymmetries [7]:

$$\begin{aligned} A_{\mathrm{SL}}^d &= (-2.1 \pm 1.7) \times 10^{-3}, \\ A_{\mathrm{SL}}^s &= (6.0 \pm 0.3) \times 10^{-4}, \end{aligned} \tag{8.11}$$

[1] *Allowed* in this context refers to regions of parameter space where these decays to dark baryons can be responsible for baryogenesis.

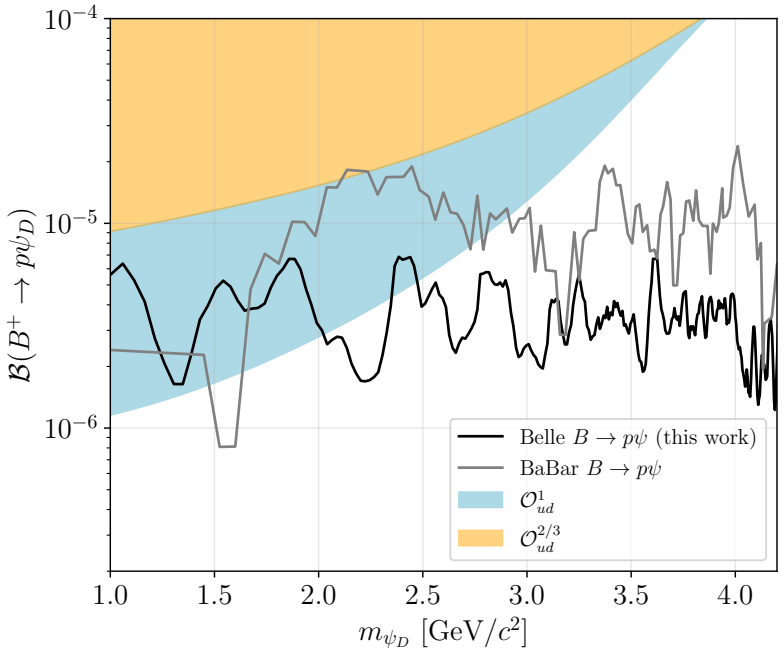


Figure 8.9: Our exclusion for $B^{\pm} \to pa$ with BaBar's limit in the same channel [117] and the region where B-mesogenesis is allowed overlaid.

and their predicted values based on the SM [208]:

$$\begin{aligned} A^d_{\mathrm{SL}} &= (-4.7 \pm 0.4) \times 10^{-4}, \\ A^s_{\mathrm{SL}} &= (2.1 \pm 0.2) \times 10^{-5}. \end{aligned} \tag{8.12}$$

There are also so-called *cosmological uncertainties* discussed in Ref. [66], which translate to uncertainties in mapping a branching fraction to the $A^s_{\mathrm{SL}} - A^d_{\mathrm{SL}}$ plane. This category of uncertainty arises from a number of sources, including the B meson fragmentation ratios, the neutral B meson charge distribution, and cosmological parameters in the early universe (such as the reheating temperature discussed above). Future studies that reduce the uncertainty on some of these parameters may help to more powerfully constrain the B-mesogenesis parameter space. Ignoring these cosmological uncertainties, the minimum branching fraction required for successful baryogenesis is $\mathcal{B}(B \to \mathcal{B}_{ij}\mathcal{M}\psi_D) = 1.7 \times 10^{-4}$.

8.4 RPV supersymmetry

The limits we found for the channel $B^+ \to p$ invisible can be directly used to set limits on couplings related to the process $B^+ \to p\tilde{\chi}^0_1$, where $\tilde{\chi}^0_1$ is the lightest neutralino in the model. Here, the couplings λ''_{ijk} give the coupling strengths of quarks in different

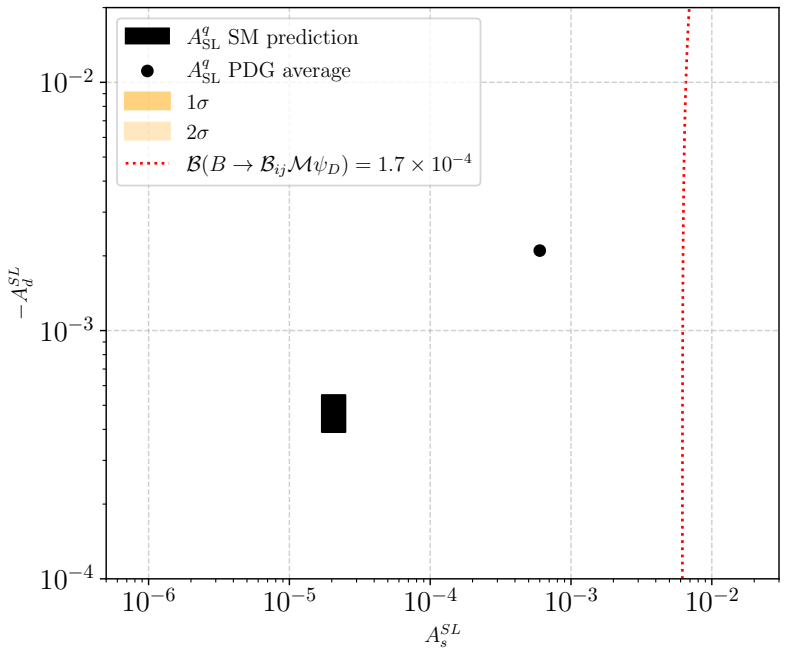


Figure 8.10: Semileptonic asymmetries overlaid with the minimum branching fraction $\mathcal{B}(B \to \mathcal{B}_{ij}\mathcal{M}\psi_D) = 1.7 \times 10^{-4}$ (ignoring cosmological uncertainties) required for successful baryogenesis. Smaller asymmetries require larger branching fractions to dark baryons for successful baryogenesis.

generations, where the indices ijk are the generation numbers. Analogous to the B-mesogenesis model, this analysis can probe the udb couplings, which correspond to $ijk = 113$. In Ref. [88], a direct relationship between the branching fractions and the relevant coupling is shown in Figure 2. We set limits on the RPV coupling λ''_{113} normalised by the squark mass shown in Fig. 8.11.

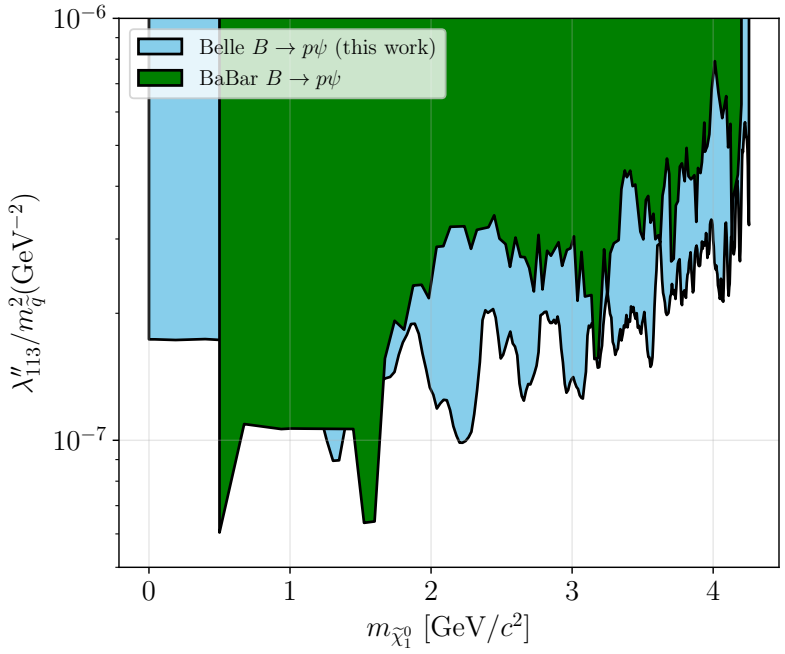


Figure 8.11: 90% CL upper limits on the RPV coupling λ''_{113}.

9

Conclusion

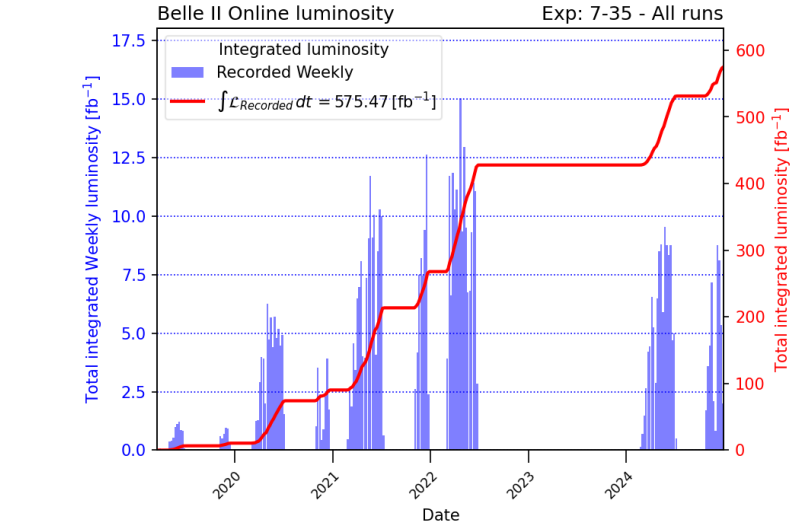
This thesis has presented a model-agnostic search for new invisible particles in five different B meson decay channels, using data from the Belle experiment. It is the first-ever search for the decays $B^{\pm} \to \pi^{\pm} a$, $B^{\pm} \to D_s^{\pm} a$, and $B^0 \to \overline{D}^0 a$, and the most sensitive search to date for the decays $B^{\pm} \to K^{\pm} a$ and $B^{\pm} \to p a$. The search utilises all of the data Belle collected at the $\Upsilon(4S)$ resonance, corresponding to approximately 771 million B meson pairs.

The analysis selection criteria are optimised on simulated events, and the statistical analysis method was fully developed before unblinding the dataset. The search is performed by looking for peaking structures in the B_{sig}-frame signal hadron momentum p_h^S. The signal distribution is modelled with a double Gaussian, while the background distribution is modelled with kernel density estimators.

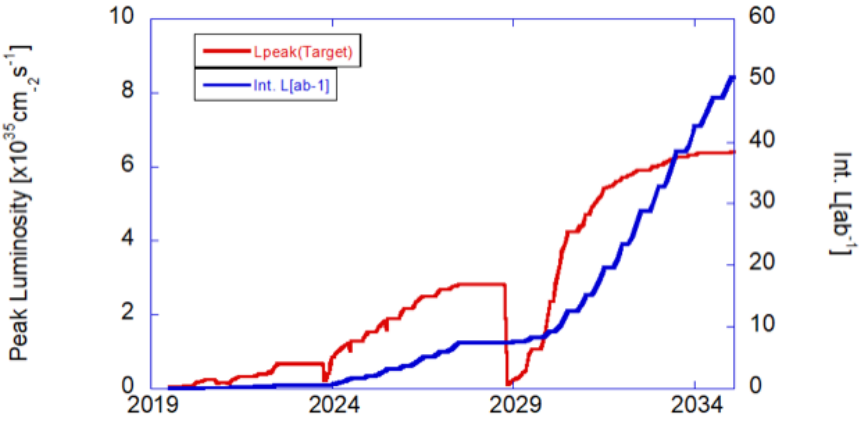
For every mass hypothesis in the scan, a combined (signal and background) fit and a background-only fit are performed, and from this, a local significance is calculated. This is then converted to a global significance to account for the look-elsewhere effect. No significant signal is observed.

Under the assumption that hidden sector particles such as ALPs and dark scalars decay invisibly, these results place the strictest limits on the ALP-W coupling g_{aW} and the dark scalar-Higgs mixing angle θ with the $B^{\pm} \to K^{\pm} a$ search, for masses beyond the kinematic reach of kaon decay experiments. Furthermore, the searches for $B^{\pm} \to \pi^{\pm} a$, $B^{\pm} \to D_s^{\pm} a$, and $B^0 \to \overline{D}^0 a$ are the world's first measurements of these decays. All of the decays probed set bounds on ALP-quark couplings. Future searches for the $B^{\pm} \to D_s^{\pm} a$ and $B^0 \to \overline{D}^0 a$ channels should aim to reconstruct more $D_s^{\pm}$ and $\overline{D}^0$ final states to significantly improve the reconstruction efficiency in those channels. The

baryonic decay channel $B^{\pm} \to pa$ sets the strictest limits on B-mesogenesis to date, almost completely excluding $bud\psi_D$ couplings as a source of baryogenesis.



(a)



(b)

Figure 9.1: Top: Total recorded integrated luminosity by Belle II up to January 6, 2025. Bottom: Belle II's plan for reaching 50 ab^{-1} by the mid-2030s. Both plots are from Belle II's public luminosity page [209], which is regularly updated.

Belle ceased data taking in 2010; however, its successor, Belle II, has a dataset that will soon supplant Belle's. In the early hours of Boxing Day 2024, the SuperKEKB collider broke its own world record for the highest instantaneous luminosity at a particle collider, reaching 5.1×10^{34} cm^{-2}s^{-1} [128]. The continuing accelerator improvements mean that Belle II's dataset will continually grow. The branching fraction sensitivity of an experiment scales with the square root of the luminosity of the experiment. Therefore, using the same analysis technique described in this thesis, Belle II's goal of 50 ab^{-1} will improve the branching fraction exclusions by an order of magnitude. This improvement will in turn be enhanced by Belle II's more advanced subdetectors, as well as improvements in tagging algorithms. Currently, the FEI has a B meson reconstruction efficiency

of about 1%; however, it already tags B mesons twice as efficiently as older methods, such as Belle's Full Reconstruction and BaBar's Semi-Exclusive Reconstruction [147, 210]. More advanced tagging methods in the future will compound with the full Belle II dataset to make the experiment an extremely powerful probe of new physics in missing energy searches.

Despite not finding any evidence for the particle nature of dark matter, this work manages to rule out large swaths of many parameter spaces. This helps to inform future experiments in their quest to continue the search. Baryogenesis and the strong CP problem remain mysteries.

Bibliography

[1] D. Marcantonio, "Rare B and B_s decays at Belle", in Proceedings of 20th International Conference on B-Physics at Frontier Machines (Nov. 2024), page 46.

[2] Y. Zhang et al., "Belle II observation prospects for axionlike particle production from B meson annihilation decay", Physical Review D **109**, 016008 (2024).

[3] Belle Collaboration and Belle II Collaboration (Belle and Belle II Collaborations), "Search for feebly interacting particles in B decays with missing energy at Belle", Physical Review Letters **136**, 231802 (2026), Associated data release: HEPData, doi:10.17182/hepdata.179653.

[4] The Editors of Encyclopaedia Britannica, "Alexis Bouvard", Encyclopaedia Britannica (2024).

[5] Ellis D. Miner and The Editors of Encyclopaedia Britannica, "Neptune", Encyclopaedia Britannica (2024).

[6] NASA/JPL, "Neptune Full Disk View", NASA Photojournal (1998).

[7] S. Navas et al. (Particle Data Group), "Review of Particle Physics", Physical Review D **110**, 030001 (2024).

[8] K. Garrett and G. Dūda, "Dark Matter: A Primer", Advances in Astronomy **2011**, 968283 (2011).

[9] N. Aghanim et al. (Planck Collaboration), "Planck 2018 results. VI. Cosmological parameters", Astronomy & Astrophysics **641** (2018).

[10] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: evidence, candidates and constraints", Physics Reports **405**, 279–390 (2005).

[11] M. Milgrom, "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis." The Astrophysical Journal **270**, 365–370 (1983).

[12] M. Gell-Mann, "A schematic model of baryons and mesons", Physics Letters **8**, 214–215 (1964).

[13] G. Zweig, "An SU_3 model for strong interaction symmetry and its breaking", Developments in the Quark Theory of Hadrons, 22–101 (1964).

[14] T. E. Browder et al., "New physics at a Super Flavor Factory", Reviews of Modern Physics **81**, 1887–1941 (2009).

[15] Izaak Neutelings, "Standard Model", TikZ.net.

[16] Wikipedia Contributors, "Standard Model", Wikipedia (2024).

[17] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the color octet gluon picture", Physics Letters B **47**, 365–368 (1973).

[18] S. L. Glashow, "Partial-symmetries of weak interactions", Nuclear Physics **22**, 579–588 (1961).

[19] S. Weinberg, "A Model of Leptons", Physical Review Letters **19**, 1264–1266 (1967).

[20] A. Salam, "Weak and electromagnetic interactions", in *World Scientific Series in 20th Century Physics*, Vol. 5 (World Scientific, May 1994), pages 244–254.

[21] J. H. Oort, "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems", Bulletin of the Astronomical Institutes of the Netherlands **6**, 249 (1932).

[22] F. Zwicky, "Republication of: The redshift of extragalactic nebulae", General Relativity and Gravitation **41**, 207–224 (2009).

[23] H. W. Babcock, "The rotation of the Andromeda Nebula", Lick Observatory Bulletin **498**, 41–51 (1939).

[24] V. C. Rubin and W. K. Ford Jr., "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions", The Astrophysical Journal **159**, 379 (1970).

[25] G. Steigman, B. Dasgupta, and J. F. Beacom, "Precise relic WIMP abundance and its impact on searches for dark matter annihilation", Physical Review D **86**, 023506 (2012).

[26] L. J. Hall et al., "Freeze-in production of FIMP dark matter", Journal of High Energy Physics **2010**, 80 (2010).

[27] J. Hisano, S. Matsumoto, and M. M. Nojiri, "Explosive dark matter annihilation", Phys. Rev. Lett. **92**, 031303 (2004).

[28] Ethan Siegel, "The 'Strong CP Problem' Is The Most Underrated Puzzle In All Of Physics", Forbes (2019).

[29] T. Mannel, "Theory and Phenomenology of CP Violation", Nuclear Physics B - Proceedings Supplements **167**, 170–174 (2007).

[30] A. D. Sakharov, "Violation of *CP* in variance, *C* asymmetry, and baryon asymmetry of the universe", Soviet Physics Uspekhi **34**, 392–393 (1991).

[31] P. Agrawal et al., "Feebly-interacting particles: FIPs 2020 workshop report", The European Physical Journal C **81**, 1015 (2021).

[32] C. Antel et al., "Feebly-interacting particles: FIPs 2022 Workshop Report", The European Physical Journal C **83**, 1122 (2023).

[33] L. Di Luzio et al., "The landscape of QCD axion models", Physics Reports **870**, 1–117 (2020).

[34] G. 't Hooft, "Symmetry Breaking through Bell-Jackiw Anomalies", Physical Review Letters **37**, 8–11 (1976).

[35] G. 't Hooft, "Computation of the quantum effects due to a four-dimensional pseudoparticle", Physical Review D **14**, 3432–3450 (1976).

[36] R. Crewther et al., "Chiral estimate of the electric dipole moment of the neutron in quantum chromodynamics", Physics Letters B **88**, 123–127 (1979).

[37] C. Abel et al. (PSI Collaboration), "Measurement of the Permanent Electric Dipole Moment of the Neutron", Physical Review Letters **124**, 081803 (2020).

[38] S. Weinberg, "A New Light Boson?", Physical Review Letters **40**, 223–226 (1978).

[39] F. Wilczek, "Problem of Strong P and T Invariance in the Presence of Instantons", Physical Review Letters **40**, 279–282 (1978).

[40] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of pseudoparticles", Physical Review Letters **38**, 1440–1443 (1977).

[41] R. D. Peccei and H. R. Quinn, "Constraints imposed by CP conservation in the presence of pseudoparticles", Physical Review D **16**, 1791–1797 (1977).

[42] H. Georgi, D. B. Kaplan, and L. Randall, "Manifesting the invisible axion at low energies", Physics Letters B **169**, 73–78 (1986).

[43] M. Bauer, M. Neubert, and A. Thamm, "Collider probes of axion-like particles", Journal of High Energy Physics **2017**, 44 (2017).

[44] D. S. M. Alves and N. Weiner, "A viable QCD axion in the MeV mass range", Journal of High Energy Physics **2018**, 92 (2018).

[45] I. Brivio et al., "ALPs effective field theory and collider signatures", The European Physical Journal C **77**, 572 (2017).

[46] C. Cornella, P. Paradisi, and O. Sumensari, "Hunting for ALPs with lepton flavor violation", Journal of High Energy Physics **2020**, 158 (2020).

[47] J. M. Camalich et al., "Quark flavor phenomenology of the QCD axion", Physical Review D **102**, 015023 (2020).

[48] M. Bauer et al., "Flavor probes of axion-like particles", Journal of High Energy Physics **2022**, 56 (2022).

[49] L. Merlo et al., "Revisiting the production of ALPs at B-factories", Journal of High Energy Physics **2019**, 91 (2019).

[50] A. Kivel, J. Laux, and F. Yu, "Axion couplings in gauged $U(1)'$ extensions of the standard model", Journal of High Energy Physics **2023**, 78 (2023).

[51] A. Bharucha et al., "Axion-like particles as mediators for dark matter: beyond freeze-out", Journal of High Energy Physics **2023**, 141 (2023).

[52] X. C. Vidal et al., "New Axion Searches at Flavor Factories", arXiv, `10.48550/arXiv.1810.09452` (2018).

[53] B. Döbrich et al., "Model-independent bounds on light pseudoscalars from rare B-meson decays", Physics Letters B **790**, 537–544 (2019).

[54] D. Aloni, Y. Soreq, and M. Williams, "Coupling QCD-Scale Axionlike Particles to Gluons", Physical Review Letters **123**, 031803 (2019).

[55] G. Alonso-Álvarez, M. B. Gavela, and P. Quilez, "Axion couplings to electroweak gauge bosons", The European Physical Journal C **79**, 223 (2019).

[56] J. Bonilla et al., "The cost of an ALP solution to the neutral B-anomalies", Journal of High Energy Physics **2023**, 138 (2023).

[57] M. B. Gavela et al., "Flavor constraints on electroweak ALP couplings", The European Physical Journal C **79**, 369 (2019).

[58] B.-F. Hou et al., "Deciphering the Belle II data on $B \to K\nu\bar{\nu}$ decay in the (dark) SMEFT with minimal flavour violation", arXiv, `10.48550/arXiv.2402.19208` (2024).

[59] F. Acanfora et al., "Fusing photons into nothing, a new search for invisible ALPs and Dark Matter at Belle II", arXiv, `10.48550/arXiv.2307.06369` (2023).

[60] J. Liu et al., "ALP explanation to the muon $(g-2)$ and its test at future tera- Z and Higgs factories", Physical Review D **107**, 095016 (2023).

[61] E. Izaguirre, T. Lin, and B. Shuve, "Searching for Axionlike Particles in Flavor-Changing Neutral Current Processes", Physical Review Letters **118**, 111802 (2017).

[62] P. Ball and R. Zwicky, "New results on $B \to \pi, K, \eta$ decay form factors from light-cone sum rules", Physical Review D **71**, 014015 (2005).

[63] L. Calibbi et al., "Minimal axion model from flavor", Physical Review D **95**, 095009 (2017).

[64] T. Alanne, S. Blasi, and F. Goertz, "Axiflavon-Higgs Unification", arXiv, `10.48550/arXiv.1905.07285` (2019).

[65] G. Elor, M. Escudero, and A. E. Nelson, "Baryogenesis and dark matter from B mesons", Physical Review D **99**, 035031 (2019).

[66] G. Alonso-Álvarez, G. Elor, and M. Escudero, "Collider signals of baryogenesis and dark matter from B mesons: a roadmap to discovery", Physical Review D **104**, 035028 (2021).

[67] B. Holdom, "Two U(1)'s and ϵ charge shifts", Physics Letters B **166**, 196–198 (1986).

[68] H. Davoudiasl, H.-S. Lee, and W. J. Marciano, ""Dark" Z implications for parity violation, rare meson decays, and Higgs physics", Physical Review D **85**, 115019 (2012).

[69] E. Gabrielli et al., "Explaining the $B^+ \to K^+\nu\bar{\nu}$ excess via a massless dark photon", The European Physical Journal C **84**, 460 (2024).

[70] K. Fuyuto, W.-S. Hou, and M. Kohda, "Z'-induced FCNC decays of top, beauty, and strange quarks", Physical Review D **93**, 054021 (2016).

[71] A. Datta et al., "Dark photon and dark Z mediated B meson decays", Journal of High Energy Physics **2023**, 108 (2023).

[72] A. V. Bednyakov and c S. H. Tanyıldızı, "A Mathematica package for calculation of one-loop penguins in FCNC processes", International Journal of Modern Physics C **26**, 1550042 (2015).

[73] A. L. Foguel, P. Reimitz, and R. Z. Funchal, "A robust description of hadronic decays in light vector mediator models", Journal of High Energy Physics **2022**, 119 (2022).

[74] A. Filimonova, R. Schäfer, and S. Westhoff, "Probing dark sectors with long-lived particles at Belle II", Physical Review D **101**, 095006 (2020).

[75] B. Patt and F. Wilczek, "Higgs-field Portal into Hidden Sectors", `10.48550/arXiv.hep-ph/0605188` (2006).

[76] D. O'Connell, M. J. Ramsey-Musolf, and M. B. Wise, "Minimal extension of the standard model scalar sector", Physical Review D **75**, 037701 (2007).

[77] F. Piazza and M. Pospelov, "Sub-eV scalar dark matter through the super-renormalizable Higgs portal", Physical Review D **82**, 043533 (2010).

[78] V. Silveira and A. Zee, "Scalar Phantoms", Physics Letters B **161**, 136–140 (1985).

[79] M. Pospelov, A. Ritz, and M. Voloshin, "Secluded WIMP dark matter", Physics Letters B **662**, 53–61 (2008).

[80] P. W. Graham, D. E. Kaplan, and S. Rajendran, "Cosmological Relaxation of the Electroweak Scale", Physical Review Letters **115**, 221801 (2015).

[81] J. Espinosa and M. Quirós, "The electroweak phase transition with a singlet", Physics Letters B **305**, 98–105 (1993).

[82] S. Profumo, M. J. Ramsey-Musolf, and G. Shaughnessy, "Singlet Higgs phenomenology and the electroweak phase transition", Journal of High Energy Physics **2007**, 010–010 (2007).

[83] D. Croon et al., "QCD baryogenesis", Physical Review D **101**, 055042 (2020).

[84] F. Bezrukov and D. Gorbunov, "Light inflaton hunter's guide", Journal of High Energy Physics **2010**, 10 (2010).

[85] D. McKeen, J. N. Ng, and D. Tuckler, "Higgs Portal Interpretation of the Belle II $B^+ \to K^+\nu\nu$ Measurement", arXiv, `10.48550/arXiv.2312.00982` (2023).

[86] M. W. Winkler, "Decay and detection of a light scalar boson mixing with the Higgs boson", Physical Review D **99**, 015018 (2019).

[87] J. A. Bailey et al. (Fermilab Lattice and MILC Collaborations), "$B \to Kl^+l^-$ decay form factors from three-flavor lattice QCD", Physical Review D **93**, 025026 (2016).

[88] C. O. Dib et al., "Probing R-parity violation in B-meson decays to a baryon and a light neutralino", Journal of High Energy Physics **2023**, 224 (2023).

[89] Yannis Semertzidis, "Introductory review on axion physics", in The Axion Quest 2024 conference - 20th Rencontres du Vietnam (Aug. 2024).

[90] I. G. Irastorza and J. Redondo, "New experimental approaches in the search for axion-like particles", Progress in Particle and Nuclear Physics **102**, 89–159 (2018).

[91] V. V. Anisimovsky et al. (E949 Collaboration), "Improved measurement of the $K^+ \to \pi^+\nu\bar{\nu}$ branching ratio", Physical Review Letters **93**, 031801 (2004).

[92] A. V. Artamonov et al. (E949 Collaboration), "New measurement of the $K^+ \to \pi^+\nu\bar{\nu}$ branching ratio", Physical Review Letters **101**, 191802 (2008).

[93] A. V. Artamonov et al. (E949 Collaboration), "Study of the decay $K^+ \to \pi^+\nu\bar{\nu}$ in the momentum region $140 < p_\pi < 199\text{MeV}/c$", Physical Review D **79**, 092004 (2009).

[94] E. C. Gil et al., "The beam and detector of the NA62 experiment at CERN", Journal of Instrumentation **12**, P05025–P05025 (2017).

[95] E. Cortina Gil et al. (NA62 Collaboration), "Measurement of the very rare $K^+ \to \pi^+\nu\bar{\nu}$ decay", Journal of High Energy Physics **2021**, 93 (2021).

[96] J. P. Lees et al. (BaBar Collaboration), "Search for $B \to K^{(*)}\nu\bar{\nu}$ and invisible quarkonium decays", Physical Review D **87**, 112005 (2013).

[97] I. Adachi et al. (Belle II Collaboration), "Evidence for $B^+ \to K^+\nu\bar{\nu}$ decays", Physical Review D **109**, 112006 (2024).

[98] W. Parrott, C. Bouchard, and C. Davies (HPQCD Collaboration), "Standard model predictions for $B \to Kl^+l^-$, $B \to Kl_1^+l_2^-$ and $B \to K\nu\bar{\nu}$ using form factors from $N_f = 2 + 1 + 1$ lattice QCD", Physical Review D **107**, 014511 (2023).

[99] C. O'Hare, "Cajohare/AxionLimits: AxionLimits", Zenodo, July 2020.

[100] E. Cortina Gil et al. (NA62 Collaboration), "Search for π^0 decays to invisible particles", Journal of High Energy Physics **2021**, 201 (2021).

[101] D. Gorbunov, I. Krasnov, and S. Suvorov (PS191 Collaboration), "Constraints on light scalars from PS191 results", Physics Letters B **820**, 136524 (2021).

[102] G. Bernardi et al. (PS191 Collaboration), "Further limits on heavy neutrino couplings", Physics Letters B **203**, 332–334 (1988).

[103] F. Bergsma et al. (CHARM Collaboration), "Search for axion-like particle production in 400 GeV proton-copper interactions", Physics Letters B **157**, 458–462 (1985).

[104] J. Beacham et al., "Physics beyond colliders at CERN: beyond the Standard Model working group report", Journal of Physics G: Nuclear and Particle Physics **47**, 010501 (2019).

[105] P. Abratenko et al. (MicroBooNE Collaboration), "Search for a Higgs Portal Scalar Decaying to Electron-Positron Pairs in the MicroBooNE Detector", Physical Review Letters **127**, 151803 (2021).

[106] A. Alavi-Harati et al. (KTeV Collaboration), "Search for the decay $K_L \to \pi^0\mu^+\mu^-$", Physical Review Letters **84**, 5279–5282 (2000).

[107] M. Acciarri et al. (L3 Collaboration), "Search for neutral higgs boson production through the process $e^+e^- \to Z^*H^0$", Physics Letters B **385**, 454–470 (1996).

[108] J. P. Lees et al. (BaBar Collaboration), "Search for long-lived particles in e^+e^- collisions", Physical Review Letters **114**, 171801 (2015).

[109] R. Aaij et al. (LHCb Collaboration), "Search for long-lived scalar particles in $B^+ \to K^+\chi(\mu^+\mu^-)$ decays", Physical Review D **95**, 071101 (2017).

[110] R. Aaij et al. (LHCb Collaboration), "Search for hidden-sector bosons in $B^0 \to K^{*0}\mu^+\mu^-$ decays", Physical Review Letters **115**, 161802 (2015).

[111] D. Craik et al., "LHCb future dark-sector sensitivity projections for Snowmass 2021", arXiv, `10.48550/arXiv.2203.07048` (2022).

[112] A. Tumasyan et al. (CMS Collaboration), "Search for long-lived particles decaying into muon pairs in proton-proton collisions at $\sqrt{s} = 13$ tev collected with a dedicated high-rate data stream", Journal of High Energy Physics **2022**, 62 (2022).

[113] I. Adachi et al. (Belle II Collaboration), "Search for a long-lived spin-0 mediator in $b \to s$ transitions at the belle ii experiment", Physical Review D **108**, L111104 (2023).

[114] S. S. Dreyer, "Search for a long-lived spin-0 particle in $b \to s$ quark transitions at the Belle II experiment", PhD thesis (Hamburg U., 2023).

[115] R. Barate, D. Decamp, and P. Ghez (ALEPH Collaboration), "Measurements of BR($b \to \tau^- \bar{\nu}_\tau X$) and BR($b \to \tau^- \bar{\nu}_\tau D^{*\pm} X$) and upper limits on BR($B^- \to \tau^- \bar{\nu}_\tau$) and BR($B \to s\nu\bar{\nu}$)", The European Physical Journal C - Particles and Fields **19**, 213–227 (2001).

[116] J. Lees et al. (BaBar Collaboration), "Search for B mesogenesis at BaBar", Physical Review D **107**, 092001 (2023).

[117] J. P. Lees et al. (BaBar Collaboration), "Search for Evidence of Baryogenesis and Dark Matter in $B^+ \to \psi_D + p$ Decays at BABAR", Physical Review Letters **131**, 201801 (2023).

[118] C. Hadjivasiliou et al. (Belle Collaboration), "Search for B^0 meson decays into Λ and missing energy with a hadronic tagging method at Belle", Physical Review D **105**, L051101 (2022).

[119] J. P. Lees et al. (BaBar Collaboration), "A search for baryogenesis and dark matter in $B^+ \to \Lambda_c^+ +$ invisible decays", Physical Review D **111**, `10.1103/PhysRevD.111.L031101` (2025).

[120] D. Du et al. (Fermilab Lattice and MILC Collaborations), "Phenomenology of semileptonic B-meson decays with form factors from lattice QCD", Physical Review D **93**, 034005 (2016).

[121] T. Felkl, S. L. Li, and M. A. Schmidt, "A tale of invisibility: constraints on new physics in $b \to s\nu\nu$", Journal of High Energy Physics **2021**, 118 (2021).

[122] K. G. Wilson, "Confinement of quarks", Physical Review D **10**, 2445–2459 (1974).

[123] I. I. Balitsky, V. M. Braun, and A. V. Kolesnichenko, "Radiative decay $\sigma^+ \to p\gamma$ in quantum chromodynamics", Nuclear Physics B **312**, 509–550 (1989).

[124] P. Colangelo and A. Khodjamirian, "QCD Sum Rules, a Modern Perspective", At The Frontier of Particle Physics, 1495–1576 (2001).

[125] Y.-Y. Keum, H.-n. Li, and A. I. Sanda, "Penguin enhancement and $B \to K\pi$ decays in perturbative QCD", Physical Review D **63**, 054008 (2001).

[126] Wikipedia Contributors, "KEKB (accelerator)", Wikipedia (2022).

[127] National Laboratory for High Energy Physics, "KEKB B factory design report", KEK report (1995).

[128] Institute of Particle and Nuclear Studies, "SuperKEKB/Belle II Complete 2024 Operations", https://www2.kek.jp/ipns/en/news/7015/, Jan. 2025.

[129] B. Aubert et al. (BaBar Collaboration), "Measurement of the branching fraction of $\Upsilon(4S) \to B^0 \bar{B}^0$", Physical Review Letters **95**, 042001 (2005).

[130] S. Banerjee et al. (HFLAV), "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023", arXiv, `10.48550/arXiv.2411.18639` (2024).

[131] A. Abashian et al. (Belle Collaboration), "The Belle detector", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **479**, 117–232 (2002).

[132] K. Abe et al. (Belle Collaboration), "Measurement of inclusive production of neutral pions from $\Upsilon(4S)$ decays", Physical Review D **64**, 072001 (2001).

[133] I. Stavitski, "Upgrade of DELPHI Microvertex detector in 1996", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Position-Sensitive Detectors Conference 1996 **392**, 165–168 (1997).

[134] E. Nygård et al., "CMOS low noise amplifier for microstrip readout Design and results", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **301**, 506–516 (1991).

[135] O. Toker et al., "VIKING, a CMOS low noise monolithic 128 channel frontend for Si-strip detector readout", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **340**, 572–579 (1994).

[136] R. Abe et al., "BELLE/SVD2 status and performance", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **535**, 379–383 (2004).

[137] T. Abe et al. (Belle II Collaboration), "Belle II Technical Design Report", arXiv, `10.48550/arXiv.1011.0352` (2010).

[138] D. J. Lange, "The EvtGen particle decay simulation package", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462**, 152–155 (2001).

[139] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", Journal of High Energy Physics **2006**, 026–026 (2006).

[140] E. Barberio and Z. Was, "PHOTOS - a universal Monte Carlo for QED radiative corrections: version 2.0", Computer Physics Communications **79**, 291–308 (1994).

[141] R. Brun et al., *GEANT3*, technical report CERN-DD-EE-84-1 (CERN, 1987).

[142] M. Gelb et al., "B2BII: Data Conversion from Belle to Belle II", Computing and Software for Big Science **2**, 9 (2018).

[143] T. Kuhr et al., "The Belle II Core Software: Belle II Framework Software Group", Computing and Software for Big Science **3**, 1 (2019).

[144] E. Nakano, "Belle PID", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **494**, 402–408 (2002).

[145] A. Satpathy, "Particle identification with BELLE", arXiv, 10.48550/arXiv.hep-ex/9903045 (1999).

[146] T. Keck et al., "The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment", Computing and Software for Big Science **3**, 6 (2019).

[147] M. Feindt et al., "A hierarchical neurobayes-based algorithm for full reconstruction of B mesons at B factories", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **654**, 432–440 (2011).

[148] A. J. Bevan et al., "The Physics of the B Factories", The European Physical Journal C **74**, 3026 (2014).

[149] F. James, "Monte Carlo theory and practice", Reports on Progress in Physics **43**, 1145–1189 (1980).

[150] B. R. Webber, "Monte Carlo Simulation of Hard Hadronic Processes", Annual Review of Nuclear and Particle Science **36**, 253–286 (1986).

[151] I. Adachi et al., "Computing System for the Belle Experiment", 10.48550/arXiv.physics/0306120 (2003).

[152] Yutaka Ushiroda, "Remarks on Generated Generic MC samples", Belle Collaboration website.

[153] H. Ishino, "The Belle Silicon Vertex Detector", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **549**, 16–19 (2005).

[154] S. Mandelstam, "Determination of the Pion-Nucleon Scattering Amplitude from Dispersion Relations and Unitarity. General Theory", Physical Review **112**, 1344–1360 (1958).

[155] H. Albrecht et al. (ARGUS Collaboration), "Exclusive hadronic decays of B mesons", Zeitschrift für Physik C Particles and Fields **48**, 543–551 (1990).

[156] Yoav Freund and Robert E. Schapire, "Experiments with a New Boosting Algorithm", in International Conference on Machine Learning, Vol. 96 (1996), pages 148–156.

[157] 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**, 119–139 (1997).

[158] J. H. Friedman, "Greedy function approximation: A gradient boosting machine." The Annals of Statistics **29**, 10.1214/aos/1013203451 (2001).

[159] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Aug. 2016), pages 785–794.

[160] T. Keck, "FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification", 2016.

[161] T. Keck, "FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment", Computing and Software for Big Science **1**, 2 (2017).

[162] T. Keck, "Machine learning algorithms for the Belle II experiment and their validation on Belle data", https://publikationen.bibliothek.kit.edu/1000078149, 2017.

[163] M. Hohmann, "An inclusive measurement of $|V_{ub}|$ at the Belle experiment", PhD thesis (The University of Melbourne, 2019).

[164] G. C. Fox and S. Wolfram, "Observables for the analysis of event shapes in annihilation and other processes", Physical Review Letters **41**, 1581–1585 (1978).

[165] K. Abe et al. (Belle Collaboration), "A measurement of the branching fraction for the inclusive $B \to X\gamma$ decays with the Belle detector", Physics Letters B **511**, 151–158 (2001).

[166] S. H. Lee et al. (Belle Collaboration), "Evidence for $B^0 \to \pi^0\pi^0$", Physical Review Letters **91**, 261801 (2003).

[167] E. Kou et al., "The Belle II Physics Book", Progress of Theoretical and Experimental Physics **2019**, 123C01 (2019).

[168] D. M. Asner et al. (CLEO Collaboration), "Search for exclusive charmless hadronic B decays", Physical Review D **53**, 1039–1050 (1996).

[169] G. Punzi, "Sensitivity of searches for new signals and its optimization", arXiv, `10.48550/arXiv.physics/0308063` (2003).

[170] Kyle Amirie et al., *Identification of Beam Background Calorimeter Clusters and Fake Photons at Belle II Using BDT Classifiers*, Belle II Note BELLE2-NOTE-TE-2022-015 (2022).

[171] Priyanka Cheema et al., *Development of FastBDT Classifiers to Suppress Beam Background Clusters and Fake Photons*, Belle II Note BELLE2-NOTE-TE-2022-015 (2022).

[172] S. Kumar, *KLM response to K_L, $\pi^\pm$ and $K^\pm$*, Belle Note BN351 (2000).

[173] D. Aloni, Y. Soreq, and M. Williams, "Coupling QCD-Scale Axionlike Particles to Gluons - Supplement", Physical Review Letters **123**, 031803 (2019).

[174] E. Bertholet et al., "Heavy QCD axion at Belle II: Displaced and prompt signals", Physical Review D **105**, L071701 (2022).

[175] R. Barlow, "Extended maximum likelihood", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **297**, 496–506 (1990).

[176] A. L. Read, "Presentation of search results: the CL_s technique", Journal of Physics G: Nuclear and Particle Physics **28**, 2693–2704 (2002).

[177] D. C. Hom et al., "Observation of High-Mass Dilepton Pairs in Hadron Collisions at 400 GeV", Physical Review Letters **36**, 1236–1239 (1976).

[178] J. Yoh, "The discovery of the b quark at Fermilab in 1977: The experiment coordinator's story", in Twenty beautiful years of bottom physics (1998), pages 29–42.

[179] S. W. Herb et al., "Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions", Physical Review Letters **39**, 252–255 (1977).

[180] L. Lyons, "Discovering the Significance of 5 sigma", arXiv, 10.48550/arXiv.1310.1284 (2013).

[181] Harold Jeffreys, "Theory of Probability", Second (Oxford At The Clarendon Press, 1939).

[182] D. V. Lindley, "A Statistical Paradox", Biometrika **44**, 187–192 (1957).

[183] L. Lyons, "Open statistical issues in Particle Physics", The Annals of Applied Statistics **2**, 10.1214/08-AOAS163 (2008).

[184] A. De Roeck, H. Prosper, and L. Lyons, "PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics", CERN Yellow Reports: Conference Proceedings, 10.5170/CERN-2008-001 (2008).

[185] S. Algeri et al., "On methods for correcting for the look-elsewhere effect in searches for new physics", Journal of Instrumentation **11**, P12010–P12010 (2016).

[186] E. Gross and O. Vitells, "Trial factors for the look elsewhere effect in high energy physics", The European Physical Journal C **70**, 525–530 (2010).

[187] R. Brun et al., "Root-project/root: v6.18/02", Zenodo, Aug. 2019.

[188] L. Moneta et al., "The RooStats project", in Proceedings of 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research — PoS(ACAT2010) (Feb. 2011), page 057.

[189] RooStats Team, "RooStats for Searches", in PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (2011), pages 199–208.

[190] P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python", Nature Methods **17**, 261–272 (2020).

[191] J. Le Dem, "Efficient Data Storage for Analytics with Apache Parquet 2.0", July 2014.

[192] C. R. Harris et al., "Array programming with NumPy", Nature **585**, 357–362 (2020).

[193] W. McKinney, "Data Structures for Statistical Computing in Python", in Python in Science Conference (2010), pages 56–61.

[194] The pandas development team, "Pandas-dev/pandas: Pandas", Zenodo, Sept. 2024.

[195] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering **9**, 90–95 (2007).

[196] N. L. Johnson, "Systems of Frequency Curves Generated by Methods of Translation", Biometrika **36**, 149 (1949).

[197] K. Cranmer, "Kernel estimation in high-energy physics", Computer Physics Communications **136**, 198–207 (2001).

[198] F. Abudinén et al. (Belle II Collaboration), "Exclusive $B \to X_u \ell \nu_\ell$ Decays with Hadronic Full-event-interpretation Tagging in 62.8 fb^{-1} of Belle II Data", `10.48550/arXiv.2111.00710` (2021).

[199] F. Meier et al. (Belle Collaboration), "First observation of $B \to \bar{D}_1(\to \bar{D}\pi^+\pi^-)l^+\nu_l$ and measurement of the $B \to \bar{D}^{(*)}\pi l^+\nu_l$ and $B \to \bar{D}^{(*)}\pi^+\pi^- l^+\nu_l$ branching fractions with hadronic tagging at Belle", Physical Review D **107**, 092003 (2023).

[200] F. Meier and A. Vossen, *Measurement of $B \to D^{**} l\nu_l$ branching fractions using Belle data and hadronic Full Event Interpretation*, Belle Note BN1569 (2022).

[201] S. Ryu, *Study of the π^0 efficiency using $\tau^- \to \pi^-\pi^0\nu_\tau$*, Belle Note BN1244 (Belle Collaboration, 2012).

[202] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", The European Physical Journal C **71**, 1554 (2011).

[203] J. Grygier et al. (Belle Collaboration), "Search for $B \to h\nu\bar{\nu}$ decays with semileptonic tagging at Belle", Physical Review D **96**, 091101 (2017).

[204] J. Grygier et al. (Belle Collaboration), "Publisher's note: search for $B \to h\nu\bar{\nu}$ decays with semileptonic tagging at Belle", Physical Review D **97**, 099902 (2018).

[205] M. Aiko and M. Endo, "Electroweak precision test of axion-like particles", Journal of High Energy Physics **2023**, 147 (2023).

[206] J. P. Lees et al. (BaBar Collaboration), "Search for an Axionlike Particle in B Meson Decays", Physical Review Letters **128**, 131802 (2022).

[207] Eite Tiesinga et al., "The 2022 CODATA Recommended Values of the Fundamental Physical Constants: weak mixing angle", https://physics.nist.gov/cgi-bin/cuu/Value?sin2th, 2022.

[208] A. Lenz and G. Tetlalmatzi-Xolocotzi, "Model-independent bounds on new physics effects in non-leptonic tree-level decays of B-mesons", Journal of High Energy Physics **2020**, 177 (2020).

[209] Belle II Collaboration, “Luminosity”, Belle II Collaboration website (2025).

[210] William Sutcliffe, “Full Event Interpretation at Belle II”, in Full Event Intepretation at Belle II, Vol. 19, International Workshop on Advanced Computing and Analysis Techniques in Physics Research (2019).