
Studies on Inclusive $|V_{ub}|$:
A Measurement of $|V_{ub}|/|V_{cb}|$ at the Belle Experiment
and
A Search for the Weak Annihilation Contribution to
Charmless Semileptonic B Decays at the Belle II Experiment

by
Marcel Stefan Hohmann
ORCID : 0000-0001-5147-4781

Dissertation submitted with the aim of achieving a jointly awarded doctoral degree
at the
School of Physics
of the University of Melbourne
and the
Faculty of Mathematics, Informatics and Natural Sciences
Department of Physics
of the Universität Hamburg

Hamburg
February 2025

Gutachter/innen der Dissertation (Hamburg):	Prof. Dr. Kerstin Tackmann Prof. Dr. Phillip Urquijo
Gutachter/innen der Dissertation (Extern):	Prof. Dr. Martina Artusa Dr. Marcello Rotondo
Zusammensetzung der Prüfungskommission:	Prof. Dr. Sven Moch Prof. Dr. Kerstin Tackmann Prof. Dr. Freya Blekman Prof. Dr. Phillip Urquijo Prof. Dr. Martina Artusa Dr. Marcello Rotondo
Vorsitzender der Prüfungskommission:	Prof. Dr. Sven Moch
Datum der Disputation:	19 Dezember 2024
Universität Hamburg	
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Markus Drescher
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter
University of Melbourne	
Head of School of Physics	Prof. Dr. Harry Quiney
Dean of the Faculty of Science	Prof. Dr. Moira O'Bryan

Abstract

Two analyses related to the inclusive determination of the magnitude of the Cabbibo-Kobayashi-Maskawa matrix element $|V_{ub}|$ are presented. The first analysis presents the first measurement of the ratio of partial branching fractions of the inclusive charmless, $\bar{B} \rightarrow X_u \ell \bar{\nu}$, to the inclusive charmed, $\bar{B} \rightarrow X_c \ell \bar{\nu}$, semileptonic B decay, where ℓ represents an electron or muon, and X a hadronic system. The measurement is performed at the Belle experiment, using the full dataset collected at the $\Upsilon(4S)$ resonance, corresponding to 711 fb^{-1} . We measure:

$$\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} = (1.99 \pm 0.17_{\text{stat}} \pm 0.16_{\text{syst}}) \times 10^{-2},$$

in the phase space region of lepton energies in the B -frame, $E_\ell^{B_{\text{sig}}}$, above 1 GeV, where the uncertainties are statistical and systematic, respectively. From the partial branching fraction ratio we extract $|V_{ub}|/|V_{cb}|$, finding

$$\frac{|V_{ub}|}{|V_{cb}|}^{\text{BLNP}} = (9.81 \pm 0.42_{\text{stat}} \pm 0.38_{\text{syst}} \pm 0.51_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2},$$

where the third and fourth uncertainties are from the partial decay rates of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$, respectively. This value is in excellent agreement with inclusive and exclusive world averages of $|V_{ub}|/|V_{cb}|$. We additionally report this measurement broken down by B charge, determining a limit on isospin breaking in $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays, as well as by lepton flavour, finding no evidence for lepton flavour universality violation in either the charmed or charmless inclusive semileptonic B decay. Furthermore, we report the differential ratio of partial branching fractions in momentum transfer squared to the lepton-neutrino system, q^2 , and $E_\ell^{B_{\text{sig}}}$.

The second analysis presented is a search for the poorly understood weak annihilation contribution to the inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay rate, which is a source of systematic uncertainty to inclusive $|V_{ub}|$ determinations. The analysis is performed with the Belle II data sample (364 fb^{-1}) collected at the $\Upsilon(4S)$ resonance before 2023. As little is known about the shape of the weak annihilation contribution, we search for structure at high q^2 , where it is expected to be localised, using a broad range of hypothesised models. The analysis must still undergo internal review; we therefore investigate the sensitivity of the analysis to simulated signal samples. Depending on the shape and magnitude of the simulated weak annihilation sample we expect to place 90% upper limits between 1.5–6.8% for the weak annihilation contribution to neutral B decays, and between 1.4–9.1% for charged B decays, competitive with existing estimates of the upper limit for the admixture. This will be the first direct search for the weak annihilation contribution to separately constrain the magnitude of the contribution to the charged and neutral B systems.

In addition to the physics analyses, we present improvements to the identification of leptons at Belle II by exploiting calorimeter information through machine learning. Reductions in pion-to-electron and pion-to-muon misidentification rates at low momentum of up to factors 10 and 1.6, respectively, are demonstrated.

Zusammenfassung

Es werden zwei Analysen im Zusammenhang mit der inklusiven Bestimmung der Größe des Cabbibo-Kobayashi-Maskawa-Matrixelement $|V_{ub}|$ vorgestellt. Die erste Analyse stellt die erste Messung des Verhältnisses der partiellen Verzweungsverhältnisse des inklusiven semileptonischen B -Zerfalls ohne ($\bar{B} \rightarrow X_u \ell \bar{\nu}$) und mit ($\bar{B} \rightarrow X_c \ell \bar{\nu}$) charm Quark vor, wobei ℓ ein Elektron oder Myon und X ein hadronisches System repräsentiert. Die Messung wird am Belle-Experiment durchgeführt und verwendet den vollständigen Datensatz von 711 fb^{-1} , der auf der $\Upsilon(4S)$ -Resonanz gesammelt wurde. Wir messen:

$$\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} = (1.99 \pm 0.17_{\text{stat}} \pm 0.16_{\text{syst}}) \times 10^{-2},$$

im Phasenraum der Leptonenergien im Inertialsystem des B -mesons, $E_\ell^{B_{\text{sig}}}$, oberhalb von 1 GeV , wobei die Unsicherheiten statistisch und systematisch sind. Aus dem Verhältnis der partiellen Verzweungsverhältnisse extrahieren wir $|V_{ub}|/|V_{cb}|$, und finden:

$$\left. \frac{|V_{ub}|}{|V_{cb}|} \right|_{\text{BLNP}} = (9.81 \pm 0.42_{\text{stat}} \pm 0.38_{\text{syst}} \pm 0.51_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2},$$

wobei die dritte und vierte Unsicherheit aus den partiellen Zerfallsraten von $\bar{B} \rightarrow X_u \ell \bar{\nu}$ bzw. $\bar{B} \rightarrow X_c \ell \bar{\nu}$ stammen. Dieser Wert stimmt ausgezeichnet mit den weltweiten Durchschnittswerten von $|V_{ub}|/|V_{cb}|$ überein, sowohl für inklusive als auch exklusive Messungen. Wir berichten zusätzlich über diese Messung, aufgeschlüsselt nach B -Ladung, um eine Grenze für Isospin-Verletzung in $\bar{B} \rightarrow X_u \ell \bar{\nu}$ -Zerfällen zu bestimmen. Des weiteren schlüsseln wir nach Leptonen-Flavour auf, wobei wir keinen Hinweis auf eine Universalitätsverletzung des Leptonen-Flavours finden, weder im $\bar{B} \rightarrow X_u \ell \bar{\nu}$ noch $\bar{B} \rightarrow X_c \ell \bar{\nu}$ Zerfall. Darüber hinaus berichten wir über das differentielle Verhältnis der Verzweungsverhältnisse beim Impuls-Transfer im Quadrat zum Lepton-Neutrino-System, q^2 , und $E_\ell^{B_{\text{sig}}}$.

Die zweite vorgestellte Analyse ist eine Suche nach dem schlecht verstandenen schwachen Annihilationsbeitrag zur inklusiven $\bar{B} \rightarrow X_u \ell \bar{\nu}$ -Zerfallsrate, der eine Ursache systematischer Unsicherheit bei inklusiven $|V_{ub}|$ -Bestimmungen ist. Die Analyse wird mit dem Belle II Datensatz (364 fb^{-1}) durchgeführt, der vor 2023 auf der $\Upsilon(4S)$ -Resonanz gesammelt wurde. Da nur wenig über die Form des schwachen Annihilationsbeitrags bekannt ist, suchen wir nach einer Struktur bei hohem q^2 , wo dieser voraussichtlich lokalisiert ist, unter Verwendung eines breiten Spektrums von Hypothesenmodellen. Die Analyse muss noch einer internen Überprüfung unterzogen werden; daher untersuchen wir die Sensitivität der Analyse an simulierten Signalproben. Abhängig von der Form und Größe des simulierten schwachen Annihilationsbeitrags erwarten wir 90% obere Ausschlussgrenzen zwischen 1,5–6,8% für den schwachen Annihilationsbeitrag zu neutralen B -Zerfällen und zwischen 1,4–9,1% für geladene B -Zerfälle, die konkurrenzfähig mit existierenden Ausschlussgrenzen sind. Dies wird die erste direkte Suche nach dem schwachen Annihilationsbeitrag sein, die separat die Größe des Beitrags zu den geladenen und neutralen B -Systemen einschränkt.

Zusätzlich zu den Analysen präsentieren wir Verbesserungen bei der Identifizierung von Leptonen bei Belle II, indem wir Kalorimeterinformationen mit maschinellem Lernen ausnutzen. Es wird gezeigt, dass die Häufigkeit der Fehlidentifikation von Pionen als Elektronen oder Myonen bei niedrigem Impuls um Faktoren von bis zu 10 bzw. 1,6 reduziert wird.

Declarations

I, Marcel Hohmann, declare that this thesis titled, ‘Studies on Inclusive $|V_{ub}|$: A Measurement of $|V_{ub}|/|V_{cb}|$ at the Belle Experiment and A Search for the Weak Annihilation Contribution to Charmless Semileptonic B Decays at the Belle II Experiment’ and the work presented in it are my own. I confirm that:

- The thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the preface;
- due acknowledgement has been made in the text to all other material used; and
- the thesis is fewer than the maximum word limit in length, exclusive of tables, maps, bibliographies and appendices as approved by the Research Higher Degrees Committee of the University of Melbourne.

13 February 2025
Date | *Datum*

Hamburg
Place | *Ort*


Signature | *Unterschrift*

I hereby declare and affirm that this doctoral dissertation is my own work and that I have not used any aids and sources other than those indicated.

If electronic resources based on generative artificial intelligence (gAI) were used in the course of writing this dissertation, I confirm that my own work was the main and value-adding contribution and that complete documentation of all resources used is available in accordance with good scientific practice. I am responsible for any erroneous or distorted content, incorrect references, violations of data protection and copyright law or plbeam breakdown over longer distances using HiPACE++ (0.5-1m etc.) agiarism that may have been generated by the gAI.

13 February 2025
Date | *Datum*

Hamburg
Place | *Ort*


Signature | *Unterschrift*

Statement of Contribution

Modern collider-based experimental particle physics is a highly collaborative effort. Hundreds to thousands of people contribute to the design, construction, and operation of the detectors and accelerators, the software to analyse and manage the recorded data, and detailed studies of the detector performance. The works presented in this thesis are only possible due to the tremendous efforts of the members of the KEKB, SuperKEKB, Belle and Belle II collaborations.

The following provides an overview of the contributions made to the work presented within this thesis. Unless otherwise stated, all work was performed by the author.

Chapters 2, 3, and 4 present original reviews of CKM flavour theory — motivating the presented studies, the Belle and Belle II detectors, and statistical methods.

The work on boosted-decision-tree based identification of charged particles using the calorimeter of the Belle II detector presented in Ch. 5 and Ch. 6 was carried out with the support of Dr. Marco Milesi. The convolutional neural network discussed in Sec. 5.10.2 was trained by Anja Novosel. A paper discussing both methods is currently under internal collaboration review.

The discussion of the modelling of key signal and background channels in Ch. 7 is an original review interspersed with details of the implementation by the author in the presented measurements. The modelling of the weak annihilation decay is entirely original.

The measurement of $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ detailed in Part III.A was performed under the supervision of Prof. Phillip Urquijo. It grew out of a preliminary study performed for the authors' Master's degree, also supervised by Prof. Urquijo, with the support of Dr. Francesco Tenchini. Dr. Lu Cao provided assistance for the hybrid modelling of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay, and the modelling of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ decay. The analysis additionally benefited from the comments of the internal reviewers, Prof. Maria Rozanska and Dr. Oskar Hartbrich. The manuscript describing the analysis is in revision following review by Physical Review D.

The search for weak annihilation presented in Part III.B has been performed under the supervision of Prof. Kerstin Tackmann and Prof. Phillip Urquijo as part of the Belle II inclusive $|V_{ub}|$ task force. The $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ form factor reweighting, generator level hadronic mass threshold for $\bar{B} \rightarrow D_{\text{broad}}^{**} \ell \bar{\nu}$ decays, slow pion efficiency corrections, and systematic uncertainty for $s\bar{s}$ production in the X_u system, were implemented by Tommy Martinov. The analysis has also greatly benefited from the feedback of Prof. Florian Bernlocher and Dr. Lu Cao, the studies of Merle Graf-Schreiber, and interesting discussions with Dr. Zoltan Ligeti and Asst. Prof. Keri Vos.

The study of low-momentum hadron identification at Belle presented in appendix B was carried out jointly with Dr. Jo-Frederik Krohn.

None of this work would have been possible without the Research Training Program scholarship provided by the Australian Government.

Acknowledgements

This work was made possible by the invaluable advice and support of many people over the last four years. First and foremost I must acknowledge my two supervisors, Phillip Urquijo and Kerstin Tackmann. They are both brilliant physicists and I have learnt a lot about physics and how to approach problems from both of them.

Phill, it has been almost exactly seven years since I submitted the report for my little research project on a K_S^0 classifier. Although that classifier was never used — and I removed it from the Belle II software a few years ago — that project sparked an interest in particle physics that led to this thesis. Thank you for letting me join your group as a Bachelor’s student all those years ago, and for inviting me to come back two more times. Thank you also for your endless guidance, especially when I was quite down on the analysis.

Kerstin, thank you for putting up with the arduous process of initiating a joint Ph.D. programme for a student on the other side of the planet during a global pandemic. Thank you for your support and invaluable guidance.

As part of this joint-Ph.D. I have been lucky enough to work with not one but two groups of excellent physicists, and, much more importantly, excellent people. Unfortunately this also means there are too many people I should thank for me to name them all, but I am indebted to each and every one of them.

I am grateful to all my Melbourne colleagues who have made the years in Melbourne so enjoyable. Thanks especially to Mike Mews, Michael Virgato, Oliver Anagnostou, Lachlan Milligan, Daniel Marcantonio, and Francis Pham, for the coffee runs, games of snooker, and Friday evenings at the Clyde. Michael also undertook the difficult task of trying to help me understand theory papers far outside his own field. My deep gratitude also to Marco Milesi for his guidance and insights when it comes to lepton identification, and to Jo-Frederik Krohn for the enjoyable collaboration on hadron identification.

I am also grateful to all the past and present members of the DESY Belle II group. I hope to have more Landgang bingo nights and barbecues with many of them. A few special mentions in no particular order: Tommy Martinov, with whom I have worked more closely than anyone else, for giving a tremendous amount of feedback on drafts of this thesis, and letting me bounce most of the silly ideas for the weak annihilation analysis off him; Sascha Dreyer, for managing to drag me further and further into the world of speciality coffee; Christian Wessel, for teaching me a lot about tracking in Belle II, and being an excellent office mate besides that; and Filippo Dattola, for always being willing to answer my questions about pyhf, but never being willing to do anything on the weekend.

A large thank you also to the members of the DESY-Bonn inclusive $|V_{ub}|$ task force: Tommy, Merle Graf-Schreiber, Munira Khan, Martin Angelsmark, Lu Cao, Kerstin and Florian Bernlocher. It has been extremely rewarding to work on this topic with you, and our workshops were always enjoyable.

My deepest gratitude to all those who proofread parts of this work: Phill, Kerstin, Tommy, Sascha, Merle, Filippo, Christian, Daniel and Alberto Martini. Given the state of the early drafts of this thesis, this was not an easy task.

Lastly I want to thank two groups of people outside of physics: my friends in Melbourne for keeping me sane during the COVID years, and my parents for their unwavering support.

10 May 2024

*I am going to write a little book for Murray on orchids,
and today I hate them worse than everything.*

Charles Darwin, 1861

Contents

1	Introduction	28
I	Background	32
2	Theory and state of semileptonic B-decays	33
2.1	Charged weak interactions and the Cabibbo–Kobayashi–Maskawa matrix . . .	34
2.2	Semileptonic B Decays	37
2.3	Experimental status of $ V_{cb} $ and $ V_{ub} $	44
2.4	Weak annihilation	49
2.5	Role of the analyses	53
3	Producing and observing B-mesons	54
3.1	Production	56
3.2	Detection	57
3.3	Software	71
3.4	Simulation	71
4	Overview of methods and tools	73
4.1	Machine learning	73
4.2	Full event interpretation	76
4.3	Parameter estimation	78
II	Improving charged particle identification in the Belle II calorimeter	82
5	Machine learning based charged particle identification in the ECL	83
5.1	Introduction	83
5.2	The ECLChargedPIDMVA module	84
5.3	Scheme overview	85
5.4	Input variables	85
5.5	Training Sample	88
5.6	Training	89
5.7	Calibration and likelihood extraction	90
5.8	Performance in simulation	94

5.9	Stability against beam background levels	96
5.10	Future developments	97
6	ECL particle identification: validation in data	101
6.1	Reconstruction	101
6.2	Feature variable modelling	104
6.3	Performance	106
III	Inclusive semileptonic B decays	112
7	Modelling	113
7.1	Charmed semileptonic decays	113
7.2	Charmless semileptonic decays	116
7.3	Weak annihilation	119
7.4	Remaining categories	123
A	Measurement of inclusive V_{ub} / V_{cb} at Belle	124
8	Event reconstruction and selection	125
8.1	Samples	125
8.2	Event reconstruction	125
8.3	Event selection	130
8.4	Tag calibration	134
8.5	Signal region enrichment	137
9	Measurement of ratios of partial branching fractions and CKM elements	148
9.1	Normalisation of non-semileptonic backgrounds	148
9.2	$\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction	149
9.3	$\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction	155
9.4	Statistical uncertainty	160
9.5	Corrections and systematic uncertainty estimation	160
9.6	Subsample results	169
9.7	Missing mass squared threshold scan	172
9.8	$ V_{ub} $ and $ V_{ub} / V_{cb} $	174
10	Ratio of spectra	178
10.1	Unfolding	178
10.2	Ratio of spectra	184
10.3	Effect of lepton energy threshold	188
B	Search for Weak Annihilation at Belle II	190

11 Event reconstruction and selection	191
11.1 Samples	191
11.2 Event reconstruction	192
11.3 Event selection	200
11.4 Control region study	212
12 Fitting procedure	216
12.1 Corrections and systematic uncertainties	218
12.2 Fit validation	229
12.3 Fit uncertainty	230
12.4 Expected limits	235
12.5 Outlook	239
12.6 Disclaimer	239
IV Conclusion	240
13 Conclusion	241
References	243
V Appendices	262
A Additional material for machine learning based particle identification in the calorimeter	263
A.1 Optimised hyperparameters	263
A.2 Feature importances	263
B Performance of low momentum kaon identification at Belle	268
B.1 Pion-to-kaon misidentification	268
B.2 Kaon identification	270
C Additional material for the search for weak annihilation contributions to semileptonic charmless B decays	277

List of Figures

2.1	The unitarity triangle and global fit to the position of its apex using all experimental constraints available before summer 2023 [10].	36
2.2	Feynman diagram of the decay $\bar{B} \rightarrow X \ell \bar{\nu}$	37
2.3	Kinematic distributions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (<i>green</i>) and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (<i>purple</i>). Regions in which $\bar{B} \rightarrow X_c \ell \bar{\nu}$ is kinematically forbidden are shaded. The distributions correspond to the nominal MC sample for the first measurement presented within this thesis, as described in Ch. 7. The $\bar{B} \rightarrow X_c \ell \bar{\nu}$ distribution is modelled as a sum of exclusive channels, the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ distribution as a combination of exclusive and inclusive channels. The steps in the distributions are caused by the combination of different channels. . .	41
2.4	Comparison of lepton energy and hadronic mass spectra predicted by different inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ models, GGOU, BLNP, DGE, and ADFR, as of 2010, using common inputs [46]. Also appears in [47]. For GGOU three lines are shown using different functional forms for the distribution functions as described in the text. The shaded areas indicate common selection thresholds in analysis $E_\ell^B > 1.9$ or 2.0 GeV, or $M_X < 1.55$ or 1.7 GeV. . . .	45
2.5	World averages of exclusive and inclusive $ V_{cb} $ determinations as reported by the Particle Data Group over the past 20 years [65]. The most recent averages are additionally shown as bands. The $ V_{cb} $ value determined by a global fit to the unitarity triangle without semileptonic inputs is given as a hatched grey band [10].	46
2.6	Measurements entering the inclusive $ V_{ub} $ world average using the prescriptions of ADFR, DGE, BLNP and GGOU, and the kinematics region in which the measurements were performed. From top to bottom these correspond to Refs. [25], [71], [28], [74], [72]×4, [73]. The solid horizontal line divides untagged and tagged measurements. The last entry gives the averages as determined by the Heavy Flavour Averaging Group [66]. The inner error bars are experimental, the outer experimental and theoretical. . . .	47
2.7	World averages of exclusive and inclusive $ V_{ub} $ determinations as reported by the Particle Data Group over the past 20 years [65]. The most recent averages are additionally shown as bands. The $ V_{ub} $ value determined by a global fit to the unitarity triangle without semileptonic inputs is given as a hatched grey band [10].	48

2.8	Feynman diagram of the valent (<i>left</i>) and nonvalent (<i>right</i>) $\bar{B} \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$ decay.	50
3.1	Schematics of the KEKB [113] (<i>left</i>) and SuperKEKB [114] (<i>right</i>) accelerator complexes. Not to scale.	56
3.2	Longitudinal cross section of the Belle detector [117].	58
3.3	Longitudinal cross section of the Belle II detector [118]. The sparse blue lines to the left of the CDC correspond to cabling and readout electronics for the CDC, not active detector components.	59
3.4	<i>Left</i> : Cross section of the Belle SVD2 [123]. <i>Right</i> : Cross section of the Belle II SVD [124]. The sensitive areas are in light grey.	60
3.5	<i>Left</i> : Side view of the full PXD [126]. <i>Right</i> : Front view of the PXD as installed in all early data taking [127]. The grey surfaces are the sensitive regions.	61
3.6	<i>Left</i> : Side view of the axial (top) and stereo (bottom) configuration. The skew of the stereo wires has been exaggerated [128]. <i>Right</i> : Simulated signals in the Belle II CDC in the x – y plane. Hits caused by a cosmic ray muon traversing the detector are given in purple, hits from beam background in black [129].	62
3.7	Schematic of the ECL crystal arrangement [135].	63
3.8	Schematic of an electromagnetic shower (<i>left</i>) [136] and a hadronic shower (<i>right</i>) [137].	65
3.9	Schematic of the interactions of different particle types with the Belle II subdetectors [141]. The subdetectors are not to scale.	66
3.10	<i>Left</i> : Schematic of the TOP operation [148]. <i>Right</i> : Schematic of the ARICH operation [149]. Note that the pion and kaon labels refer to photons originating from pions or kaons passing through the material.	68
4.1	Schematic of the FEI algorithm [84, 190]. Note that baryonic modes are not included in the diagram. For these modes we also identify protons and reconstruct intermediate Λ^0 , Σ , and Λ_c baryons.	77
5.1	Reconstructed momentum distributions, of the negatively charged training sample for low (<i>left</i>), medium (<i>centre</i>), and high momentum (<i>right</i>) in the barrel region.	89

5.2	<i>Left:</i> Example losses for the training and validation samples as a function of the number of trees added. While the loss of the validation sample flattens out the performance continues to asymptotically improve and thus does not trigger the early stopping condition. <i>Right:</i> Validation loss for each of the 500 trials in the hyperparameter optimisation process, the best trial is marked by the purple X. The optimisation quickly finds the region of minimal loss with subsequent trials only fine tuning the hyperparameters. These plots correspond to the training of the medium momentum negatively charged region in the barrel.	92
5.3	Distributions of five of the most powerful discriminating variables over the low and medium momentum barrel regions. One-dimensional distributions are shown on the diagonal, two-dimensional distributions on the off-diagonal, given by contours at 25% and 75% of the maximum density. The upper triangular shows medium-momentum particles, the lower triangular plots show low-momentum particles, which are represented by lighter colours than their medium-momentum counterparts. The samples contain equal numbers of positively and negatively charged particles. Note that, as per the discussion in the text, the energies of the ten most energetic crystal associated with the cluster, E_i, are included as features only for classifiers targeting low-momentum particles.	93
5.4	Example diagonal plots for electrons and pions with and without temperature calibration ($T = 0.996$) applied for the training sample (blue), the sample used to derive the temperature value (green) and an independent test sample (yellow). For each plot the score is divided into 50 equidistant bins. For a well calibrated model the purity of the sample within each bin is expected to be equivalent to the central score of the bin.	94
5.5	ROC curves for electron identification versus pion-to-electron misidentification in a simulated $B\bar{B}$ sample when enforcing selection thresholds on $P_{e,\text{BDT}}^{\text{ECL}}$ (blue), $P_{e,E/ \vec{p} }^{\text{ECL}}$ (black) or $\mathcal{P}_e^{\text{ECL}}$ (grey). We consider nine regions of phase space, divided into three polar angle regions: the forward endcap (<i>left column</i>), barrel (<i>center column</i>), and backwards endcap (<i>right column</i>), and three momentum regions: low (<i>top row</i>), medium (<i>center row</i>) and high (<i>bottom row</i>). The high-momentum region is truncated at 3 GeV due to the limited kinematic reach of the tracks produced in $B\bar{B}$ events. In each region, we integrate over the kinematic phase-space and track charge. As all methods rely on track-cluster matching, only tracks that are successfully matched to a track are considered. The area under the respective $1 - \varepsilon_\pi = f(\varepsilon_e)$ curve for each method is provided in the legend. .	95

5.6	ROC curves for muon identification vs. pion-to-muon misidentification in a simulated $B\bar{B}$ sample when enforcing selection thresholds on $P_{\mu,\text{BDT}}^{\text{ECL}}$ (green), $P_{\mu,E/ \vec{p} }^{\text{ECL}}$ (black) or $\mathcal{P}_{\mu}^{\text{ECL}}$ (grey). We consider nine regions of phase space, divided into three polar angle regions: the forward endcap (<i>left column</i>), barrel (<i>center column</i>), and backwards endcap (<i>right column</i>), and three momentum regions: low (<i>top row</i>), medium (<i>center row</i>) and high (<i>bottom row</i>). The high-momentum region is truncated at 3 GeV due to the limited kinematic reach of the tracks produced in $B\bar{B}$ events. In each region, we integrate over the kinematic phase-space and track charge. As all methods rely on track-cluster matching, only tracks that are successfully matched to a track are considered. The area under the respective $1 - \varepsilon_{\pi} = f(\varepsilon_{\mu})$ curve for each method is provided in the legend.	96
5.7	Comparison of electron pion (<i>left</i>) and muon pion (<i>right</i>) separation under different beam background conditions. The first sample given in the legend denotes the beam background level in the generation of the BDT training samples, either nominal (NP3) or early (EP3) phase 3. The second denotes the background conditions used to generate the $B\bar{B}$ sample on which the comparison was performed. The two sets of BDTs shown have been trained following the procedure outlined within this document on the same variable set.	98
5.8	Comparison of electron pion (<i>left</i>) and muon pion (<i>right</i>) separation achieved using ECL information in simulated tracks originating from $B\bar{B}$ events. The comparison includes the univariate $E/ \vec{p} $ based algorithm (black), the analysis level BDT (grey), the nominal BDT with (yellow) and without (blue) considering also the hadron energy fraction as computed from the pulse shape, and a model utilising the CNN with (pink) and without (green) considering the hadron energy fraction. As the kinematic coverage of the CNN is limited, algorithms that use it default to the BDT outside the covered region (the central barrel for tracks with momenta below 1 GeV). These algorithms are therefore labelled CNN + BDT. No requirement of a track-cluster match is made for this figure.	100
6.1	Di-track invariant masses and kinematic distributions of tracks from the Bhabha (<i>top</i>), radiative di-muon (<i>center</i>), and K_s^0 (<i>bottom</i>) samples. The dotted line in the lower plots gives the ratio of integrated data yield to integrated MC yield. The MC is normalised to the integrated luminosity. .	105
6.2	Distributions of the five most important feature variables in the K_s^0 , Bhabha and radiative di-muon control samples. The MC is normalised to the integrated luminosity.	107

6.3	Comparison of ECL only lepton identification performance, and data MC agreement, in the barrel region. In each 100 MeV momentum bin thresholds are placed on the output of the three particle identification algorithms under consideration, the likelihood ratios computed with the BDT and univariate $E/ \vec{p} $ based algorithms, and the analysis level BDTs, such that a lepton identification efficiency of 90% is achieved in MC. From top to bottom, the panels show the lepton identification efficiency, the ratio of lepton identification efficiencies in data over MC, the pion-to-lepton misidentification rate, and the ratio of pion misidentification rates in data over MC. See the text for details of the samples used in the comparison.	109
6.4	Comparison of global lepton identification performance, and data MC agreement, in the barrel region. In each 100 MeV momentum bin thresholds are placed on the output of the two particle identification algorithms under consideration, the likelihood ratios computed with the ECL contributions from the BDT and univariate $E/ \vec{p} $ based algorithms, such that a lepton identification efficiency of 90% is achieved in MC. The analysis level BDT is excluded from the comparison as it includes information from the TOP and SVD. From top to bottom, the panels show the lepton identification efficiency, the ratio of lepton identification efficiencies in data over MC, the pion-to-lepton misidentification rate, and the ratio of pion misidentification rates in data over MC. See text for details of the samples used in the comparison.	110
7.1	Projections of the hybrid model of semileptonic charmless decays of B^+ , as used for the Belle II analysis, in M_X , E_ℓ^B and q^2 . The DFN model is used for the hybrid scheme.	117
7.2	Kinematic distributions of the 70 considered weak annihilation models in q^2 , $E_\ell^{B_{\text{sig}}}$, and M_X . Five example distributions have been coloured to demonstrate the relations between the distributions.	121
8.1	Schematic comparison of the event shape of $B\bar{B}$ and continuum events. Originally appeared in Ref. [238].	131
8.2	Importances of the feature variables to the four continuum suppression classifiers, as measured by normalised average loss reduction per split.	133
8.3	Data-MC agreement of the off-resonance sample in the three most important variables for the continuum suppression classifiers, $\cos \theta_{\vec{T}_B \vec{T}_O}$, $\mathcal{H}_{20}^{\text{so}}$, and $\mathcal{H}_{10}^{\text{so}}$, and for the log-transformed classifier response. The MC has been scaled to the data yield, an increase of approximately 15%.	133
8.4	Receiver operator characteristic curves for the four continuum suppression classifiers, for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events against continuum events. The imposed thresholds are given as crosses.	134

8.5	Calibration fits to the ratio of off-resonance data to MC yields in the electron (<i>left</i>) and muon channels (<i>right</i>). The uncertainties combine the data and MC statistical uncertainties.	135
8.6	Example calibration fits to M_{bc} for two neutral and two charged B tag channels. The continuum component is fixed. The calibration weights, given by the ratio of pre-fit to post-fit yields of good and bad tags are given in the legend. The uncertainties are statistical.	137
8.7	Pre-calibration (<i>left</i>) and post calibration (<i>right</i>) M_{bc} spectrum of the calibration sample. The good tags have been split into two subcategories corresponding the strictly well-reconstructed tags, and tags in which one photon has been incorrectly added, labelled ‘Loose Good Tag’. The uncertainties are statistical.	138
8.8	The hadronic mass spectrum of reconstructed $\bar{B} \rightarrow X\ell\bar{\nu}$ events. Excellent Data-MC agreement is observed for the bulk of the distribution. Some disagreement is observed at high hadronic mass which may originate either from mismodelling of secondary and fake lepton events or of $\bar{B} \rightarrow D^{**}\ell\bar{\nu}$ events.	138
8.9	Simulated distributions of $M_{\text{Miss}}^2(D^*)$ and $\cos\theta_{D^*\ell}$ by status of the identified slow pion and result of the figure-of-merit scan. Note that events in which no slow pion is found are not shown in the figure.	140
8.10	Comparison of data and simulation in M_{Miss}^2 for the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ extraction sample.	141
8.11	Comparison of data and simulation in $ M_{\text{Miss}}^2 $ for the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ enriched (<i>left</i>) and depleted (<i>right</i>) subsamples of the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ extraction sample.	142
8.12	Optimisation of the $ M_{\text{Miss}}^2 $ selection. The expected statistical uncertainty was estimated by varying the pseudo-data yields in each step of the measurement by drawing from a Gaussian of width $\sqrt{N}$, for each bin in each histogram under consideration. This neglects the event-overlap between samples, however this is expected to have only a small impact on the optimisation. The estimates of the expected statistical uncertainty are noisy, however sufficient to note that the expected uncertainty is approximately uniform in the region about 0.4 GeV^2	143
8.13	Comparison of data and simulation for the number of reconstructed K^+ , and K_S^0 candidates. All selections defining the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ extraction sample are applied.	144
8.14	Reconstruction efficiency of $\bar{B} \rightarrow X_u\ell\bar{\nu}$ events after all selections as a function of $E_\ell^{B_{\text{sig}}}$, q^2 and M_X . The errors are statistical. Due to the reconstruction level selection of $E_\ell^{B_{\text{sig}}} > 1.0\text{ GeV}$ the efficiency below the shown 1 GeV boundary in $E_\ell^{B_{\text{sig}}}$ rapidly drops to zero.	146

9.1	Fit to the secondary and fake lepton control region ($0.7 < E_\ell^{B_{\text{sig}}} < 1.8 \text{ GeV}$, $M_X > 2.0 \text{ GeV}$) in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enhanced (<i>left</i>) and depleted (<i>right</i>) subsamples for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction sample. The four components are: correctly reconstructed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events (<i>green</i>), correctly reconstructed $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events (<i>purple</i>), continuum events (<i>blue</i>), and events in which either a hadron has been misidentified as a lepton or the lepton originates from a secondary decay (<i>red</i>).	150
9.2	Residuals of $E_\ell^{B_{\text{sig}}}$ and q^2 for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events, corresponding to events selected in the $\bar{B} \rightarrow X \ell \bar{\nu}$ sample and the enriched $\bar{B} \rightarrow X_u \ell \bar{\nu}$ sample respectively.	151
9.3	Binning structure for two dimensional fit to $q^2 : E_\ell^{B_{\text{sig}}}$. The sample is first divided into four equal $\bar{B} \rightarrow X_c \ell \bar{\nu}$ frequency q^2 bins. Each q^2 bin is then subdivided into four equal frequency $E_\ell^{B_{\text{sig}}}$ bins.	152
9.4	Reconstructed $E_\ell^{B_{\text{sig}}}$ and q^2 spectra, and the two-dimensional binning used for the fit for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enhanced (<i>left</i>) and depleted (<i>right</i>) subsamples. The error bands of simulated samples incorporate the full set of systematic uncertainties discussed in section 9.5. The secondaries and fakes calibration is not applied.	153
9.5	Double ratios of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields in data to MC in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched to depleted subsample ($R_E^{X_c \ell \bar{\nu}}/R_D^{X_c \ell \bar{\nu}}$) in the kinematic variables $E_\ell^{B_{\text{sig}}}$ and q^2 , and the two-dimensional binning of the two used in the analysis. The yield of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ in data is calculated as the data yield subtracting the MC expectation of $\bar{B} \rightarrow X_u \ell \bar{\nu}$, continuum and secondary and fake lepton events. The secondaries and fakes calibration is not applied.	154
9.6	Distribution of off-resonance events passing the selections of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched (<i>left</i>) and depleted (<i>right</i>) subsamples in the fit binning. The uncertainty band includes MC sample size, lepton identification uncertainties, and the continuum calibration uncertainties.	154
9.7	Fit to the $q^2 : E_\ell^{B_{\text{sig}}}$ distribution. The four components are: correctly reconstructed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events (<i>green</i>), correctly reconstructed $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events (<i>purple</i>), continuum events (<i>blue</i>), and events in which either a hadron has been misidentified as a lepton or the lepton originates from a secondary decay (<i>red</i>).	155
9.8	Fit to the secondary and fake lepton control region ($0.7 < E_\ell^{B_{\text{sig}}} < 1.8 \text{ GeV}$, $M_X > 2.0 \text{ GeV}$) in the kaon-depleted (<i>left</i>) and kaon-rich (<i>right</i>) subsamples for the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction sample.	156
9.9	The determined background compared with all events in the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction sample.	157
9.10	Distributions of residuals of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields measured in the toy test. Both are found to be consistent with zero indicating the procedure is unbiased.	159

9.11	Result of the linearity tests on $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ varying the input branching fractions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$, given by the scale factors $\mu_{\bar{B} \rightarrow X_{u/c} \ell \bar{\nu}}$. The values on the y -axis have been scaled to one for scale factors of one. The recovered scaling behaviour is consistent with expectation.	160
9.12	Generated q^2 spectrum for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events assigned to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples in the four bins of reconstructed q^2 considered in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction. Events assigned to the enriched subsample are biased towards a higher reconstructed q^2 than those in the depleted subsample due to the missing momentum of K_L^0	164
9.13	Ratio of unfolded yields in data and MC in bins of q^2 for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events within the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted subsample, and the fourth-order polynomial fit to them. The x -position of each bin is given by the mean position of the bin contents. Only the statistical uncertainties are considered as the systematic uncertainties are expected to mostly cancel, and we consider the full difference to the nominal result as the systematic uncertainty.	165
9.14	Measured ratios of partial branching fractions for the combined sample and orthogonal lepton flavour and B flavour subsamples.	169
9.15	Fit to the $q^2 : E_\ell^{B_{\text{sig}}}$ distribution for the electron (<i>left</i>) and muon (<i>right</i>) subsamples.	170
9.16	Fit to the $q^2 : E_\ell^{B_{\text{sig}}}$ distribution for the charged B (<i>left</i>) and neutral B (<i>right</i>) subsamples.	172
9.17	<i>Left:</i> Measured $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ values as a function of $ M_{\text{Miss}}^2 $ threshold, the inner error bar gives the statistical uncertainty, the outer the total uncertainty. The nominal selection is shown in red. <i>Right:</i> pairwise pull between the different thresholds.	174
9.18	Comparison of ratio of $ V_{ub} ^{\text{GGOU}}/ V_{cb} $, as measured in this work, with the averages for inclusive and exclusive $ V_{ub} $ and $ V_{cb} $ as in Fig. 66 of Ref. [66]. Note that to be consistent with Ref. [66], these are the 2021 averages of the Heavy Flavor Averaging Group. For $ V_{cb} $ these differ slightly from the 2023 averages of the Particle Data Group discussed in Ch. 2.	176
9.19	Comparison of the measured $ V_{ub} $ value with the world average [66] and measurements that currently enter the world average [25, 28, 71–74]. The inner error bars are experimental, the outer experimental and theoretical. The uncertainties have been naively added in quadrature.	177

10.1	Background subtracted spectra (black markers) of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (<i>top</i>) and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (<i>bottom</i>) in $E_\ell^{B_{\text{sig}}}$ and q^2 . The spectra are compared with the expected distributions in simulation (green for $\bar{B} \rightarrow X_u \ell \bar{\nu}$, and purple for $\bar{B} \rightarrow X_c \ell \bar{\nu}$), which have been normalised to the respective data yield. For the markers, the inner error bar gives the statistical uncertainties, the outer error bar the sum in quadrature of the statistical and systematic uncertainties. For the histogram, the lighter shaded area gives the systematic uncertainty on the shape.	179
10.2	Nominal migration matrices for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (<i>top</i>) and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (<i>bottom</i>) in $E_\ell^{B_{\text{sig}}}$ and q^2	180
10.3	Figure for merit scans for the SVD unfolding regularisation parameter k . From top to bottom the figures correspond to the unfolding of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ $E_\ell^{B_{\text{sig}}}$ and q^2 spectra, and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ $E_\ell^{B_{\text{sig}}}$ and q^2 spectra. From left to right the figures give the sum of biases ($\sum_i b_i$), sum of absolute biases ($\sum_i b_i $), and the sum of absolute biases scaled by the uncertainties ($\sum_i b_i /\sqrt{C_{ii}^{\text{unf}}}$). The solid horizontal lines give the figures of merit evaluated for unfolding via unregularised matrix inversion.	183
10.4	Unfolded ratios of partial branching fractions in the phase space $E_\ell^{B_{\text{sig}}} > 1 \text{ GeV}$. The simulated unfolded yields have been scaled to the data yields before taking the ratio. The uncertainty on the simulated bands includes all modelling uncertainties discussed in section 9.5 and the uncertainty due to the limited size of the MC sample.	185
10.5	Global correlation matrix of the ratio of partial branching fractions in $E_\ell^{B_{\text{sig}}}$ and q^2 . Values above the diagonal give the correlation matrix of the statistical uncertainty, and values below the diagonal give the correlation matrix of the systematic uncertainty.	185
10.6	Comparison of the unfolded differential ratio of partial branching fractions in lepton energy for the nominal setup and including an additional sacrificial bin below the $E_\ell^{B_{\text{sig}}}$ threshold. The results are found to be consistent. . . .	188
11.1	Validation of the calibration of the BDT trained for continuum modelling. .	197
11.2	Variable importance for the data-driven continuum reweighting.	197
11.3	Distribution of the nominal classifier response for the classifier trained to distinguish data from Monte Carlo (<i>top</i>) and distributions of $\mathcal{H}_{20}^{s0}$ and R_2 , before (<i>left</i>) and after (<i>right</i>) the data-driven reweighting procedure in the off-resonance test sample. For the calculation of the χ^2 value, bins with data yields or expected yields less than 5 events are merged into adjacent bins algorithmically. The χ^2 value is calculated using the statistical and systematic uncertainty taking into account correlations.	199

11.4	Expected distribution of the selected sample in lepton energy and hadronic mass. The selected thresholds are shown as vertical orange line. The normalisation of the overlaid weak annihilation samples (in orange and pink) is arbitrary.	201
11.5	Expected distribution of the selected sample in $M_{\text{Miss}}^2(D^* \rightarrow D\pi^+)$ and $M_{\text{Miss}}^2(D^* \rightarrow D\pi^0)$. Events above the selected thresholds, shown as vertical orange lines, are removed. The normalisation of the overlaid weak annihilation samples (in orange and pink) is arbitrary. All other selections described in Sec. 11.3 are applied.	202
11.6	Expected distribution of the missing mass squared. All other selections described in Sec. 11.3 are applied.	203
11.7	Comparison of the correlation of the trained continuum suppression networks to $E_\ell^{B_{\text{sig}}}$ (<i>top</i>), M_X (<i>center</i>), and q^2 (<i>bottom</i>) with different strengths of the distance correlation loss term, λ . The distribution of the admixture of WA models is shown in orange. For comparison contours of the distributions of the continuum (grey), $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (blue) and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (purple) samples at 20, 40, 60 and 80% of the respective maximum densities are overlaid.	207
11.8	Variable importance for the continuum suppression network.	208
11.9	Correlation of the log transformed continuum suppression classifier responses against the flatness variables as a function of the strength of the distance correlation loss term, λ	208
11.10	Log transformed distribution of the continuum suppression network output. We reject events below vertical dashed line.	209
11.11	Optimisation of the Punzi-loss. The optimiser rapidly converges on near optimal parameters. The best trial is shown as a pink cross.	210
11.12	One dimensional scans of the Punzi figure of merit about the selected thresholds in missing mass squared (<i>top</i>) and the continuum suppression variable (<i>bottom</i>) for the nine example weak annihilation models. For each model the figure of merit for the charged B component is given as a solid line, for the neutral B component as a dotted line. The dashed vertical lines give the selected thresholds. The optimised thresholds, which are optimised for all models simultaneously, are found to be near optimal for each individual model.	211
11.13	Data-MC agreement in the continuous event selection variables. For the calculation of the χ^2 value, bins with data yields or expected yields less than 5 events are merged into adjacent bins algorithmically. The χ^2 value is calculated using the statistical and systematic uncertainty taking into account correlations.	213

11.14	Data-MC agreement of the rest-of-event particle content based selection variables $M_{\text{Miss}}^2(D^* \rightarrow D\pi^+)$, $M_{\text{Miss}}^2(D^* \rightarrow D\pi^0)$, and, number of kaon, K_S^0 , and lepton candidates in the hadronic system. For the kaon and lepton counting variables, only the vetoed regions, $N_K \geq 1$, $N_{K_S^0} \geq 1$, $N_\ell \geq 1$, are shown.	214
11.15	Data-MC comparison of the weak annihilation signal strength extraction observable, q^2 , in the medium lepton energy control region, split by tag B charge. For the calculation of the χ^2 value, bins with data yields or expected yields less than 5 events are merged into adjacent bins algorithmically. The χ^2 value is calculated using the statistical and systematic uncertainty taking into account correlations.	215
12.1	Expected q^2 distribution in the signal region with linear (<i>top</i>), and logarithmic scales (<i>bottom</i>). The binning is the same as used for the fits.	217
12.2	Post-fit distribution of the fit used to estimate the FEI cross-feed rates using an Asimov sample.	228
12.3	Example pull test (<i>left</i>) and linearity test (<i>right</i>) for the weak annihilation signal strength in the neutral B channel, $\mu_{\text{WA}}^{B^0}$, using the fit configuration $M_{\ell\bar{\nu}}^2 = 22.1 \text{ GeV}^2$, $\Gamma_{\ell\bar{\nu}} = 0.02 \text{ GeV}$. The linearity tests setting $\bar{\mu}_{\text{WA}}^{B^+}$ to 0.01 and 3 have been offset in the x -axis for visualisation. The full set of pull and linearity tests for $\mu_{\text{WA}}^{B^0}$ and $\mu_{\text{WA}}^{B^+}$ are available in appendix C.	230
12.4	Breakdown of the total uncertainty on the parameters of interest, the weak annihilation signal strengths in the neutral and charged B channels, $\mu_{\text{WA}}^{B^0}$ (<i>left</i>) and $\mu_{\text{WA}}^{B^+}$ (<i>right</i>), respectively. For each fit configuration, an Asimov sample containing injected weak annihilation signal generated with the same parameters as used for the fit is used. The injected signal is normalised to 1% of the charged and neutral $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate.	231
12.5	Post-fit impact of the leading 30 nuisance parameters on the parameter of interest, for the weak annihilation signal strength in the charged B channel, $\mu_{\text{WA}}^{B^+}$, for four test fit configurations. For each fit configuration, an Asimov sample containing injected weak annihilation signal generated with the same parameters as used for the fit is used. The injected signal is normalised to 1% of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate. If a source of systematic uncertainty is associated with multiple nuisance parameters, the index of the nuisance parameter is given in square brackets. For the uncertainty associated to the limited size of the MC sample (MC Stat.) this corresponds to the bin index in q^2	233

12.6	Post-fit impact of the leading 30 nuisance parameters on the parameter of interest, for the weak annihilation signal strength in the neutral B channel, $\mu_{\text{WA}}^{B^0}$, for four test fit configurations. For each fit configuration, an Asimov sample containing injected weak annihilation signal generated with the same parameters as used for the fit is used. The injected signal is normalised to 1% of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate. If a source of systematic uncertainty is associated with multiple nuisance parameters, the index of the nuisance parameter is given in square brackets. For the uncertainty associated to the limited size of the MC sample (MC Stat.) this corresponds to the bin index in q^2	234
12.7	Observed and expected upper limits on $\mu_{\text{WA}}^{B^0}$ (<i>top row</i>), and $\mu_{\text{WA}}^{B^+}$ (<i>bottom row</i>) recovered on simulation, when injecting three example weak annihilation signals. The parameters of the injected signals are provided at the top of each plot.	237
A.1	Importances of the feature variables to electron identification for the 18 nominal reconstruction level ECL charged particle identification BDTs. The importance of each variable is measured by the mean absolute Shap value. .	265
A.2	Importances of the feature variables to muon identification for the 18 nominal reconstruction level ECL charged particle identification BDTs. The importance of each variable is measured by the mean absolute Shap value. .	266
A.3	Importances of the feature variables to pion identification for the 18 nominal reconstruction level ECL charged particle identification BDTs. The importance of each variable is measured by the mean absolute Shap value. .	267
B.1	Example fit to determine the pion-to-kaon misidentification rate. <i>Top</i> : Fit to determine the parameters of the signal model in truth-matched simulated events. <i>Bottom</i> : Simultaneous fit to events that pass or fail the selection $P_{K/\pi} > 0.6$	270
B.2	Example fit to determine the kaon identification efficiency. <i>Top</i> : Fit to determine the parameters of the signal model in truth-matched simulated events. <i>Bottom</i> : Simultaneous fit to events that pass or fail the selection $P_{K/\pi} > 0.6$	272
B.3	Pion-to-kaon misidentification rate for data collected with and MC simulated with the SVD1 configuration of the Belle detector.	273
B.4	Pion-to-kaon misidentification rate for data collected with and MC simulated with the SVD2 configuration of the Belle detector.	274
B.5	Kaon identification rate for data collected with and MC simulated with the SVD1 configuration of the Belle detector.	275
B.6	Kaon identification rate for data collected with and MC simulated with the SVD2 configuration of the Belle detector.	276

C.1	Pull tests for $\mu_{\text{WA}}^{B^0}$ and $\mu_{\text{WA}}^{B^+}$ for nine example weak annihilation models. The recovered means and widths of the fitted Gaussians generally agree well with zero and one, indicating the fits are able to accurately extract the signal strengths without bias and correctly estimate the uncertainty.	278
C.2	Linearity tests for $\mu_{\text{WA}}^{B^0}$ and $\mu_{\text{WA}}^{B^+}$ for the example regions. The results setting the normalisation of the WA component not under study to 0.01 and 3 have been offset in the x -axis for visualisation.	279

List of Tables

3.1	Production cross sections at electron-positron collisions at $\sqrt{s} = M_{\Upsilon(4S)}$ [84, 160–164]. The line separates hadronic and non-hadronic production. . .	71
4.1	FEI decay modes. The line divides purely mesonic decay modes from baryonic decay modes.	78
5.1	Hyperparameters of the XGBoost model. For parameters tuned by Bayesian optimisation, the range investigated by the optimiser is provided in the “Value” column. All parameters not listed are left at their default value as per version 1.7.0 of XGBoost.	91
7.1	Assumed branching fractions and their uncertainties for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ decays. See text for details.	115
7.2	Assumed branching fractions and their uncertainties for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays. See text for details. Non-resonant $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays are denoted $\bar{B} \rightarrow x_u \ell \bar{\nu}$	118
8.1	Percentage of selected events falling within the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples by event category.	144
8.2	Cumulative efficiency, and number of expected candidates, of the event-selection post reconstruction. The efficiency is given in terms of the number of candidates. Before the single-lepton requirement multiple candidates can be reconstructed within one event. The numbers in brackets correspond to candidates in which a lepton has been correctly reconstructed and the tag meets the definition of good tag used for the FEI calibration. Note that we apply the continuum and FEI calibrations to all candidates. This is expected to slightly overestimate the continuum yield before continuum suppression.	145
8.3	Reconstruction efficiency of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events by channel. The errors are statistical.	147
9.1	Summary of the central value, statistical, and systematic uncertainties for the ratio of partial branching fractions. The uncertainties are given as relative values on the central value in percent.	158

9.2	Summary of the central value (R), statistical, and systematic uncertainties for the subsample based results. Columns marked e, μ, B^+, B^0 give the ratio of partial branching fractions of the charmed to charmless modes for the electron and muon, and charged and neutral B subsamples, respectively. The uncertainties are given as relative values on the central value in percent.	173
10.1	Summary of the central values, statistical, and systematic uncertainties for the differential ratio of partial branching fractions as a function of $E_\ell^{B_{\text{sig}}}$ bin, defined by the boundaries [1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.7] GeV. The uncertainties are given as relative values on the central value in percent.	186
10.2	Summary of the central values, statistical, and systematic uncertainties for the differential ratio of partial branching fractions as a function of q^2 bin, defined by the boundaries [0, 2, 4, 6, 8, 10, 26] GeV ² . The uncertainties are given as relative values on the central value in percent.	187
11.1	Hyperparameters of the continuum suppression neural network. All hyperparameters not mentioned are left at their default values as of <code>tensorflow v2.10.0</code> .	205
12.1	Comparison of world averaged inclusive particle production rates [65], the production rates as generated in the official Belle II MC, and the production rates after the $D^{(*)}$ decay reweighting procedure.	224
12.2	Weakest 90% upper limit recovered on $\mu_{\text{WA}}^{B^0}$ amongst models in the fiducial model parameter space when injecting various weak annihilation signals with various models at a range of signal strengths.	238
12.3	Weakest 90% upper limit recovered on $\mu_{\text{WA}}^{B^+}$ amongst models in the fiducial model parameter space when injecting various weak annihilation signals with various models at a range of signal strengths.	238
A.1	Optimal hyperparameters of the nominal set of BDTs developed for ECL charged particle identification. The first three columns specify the phase-space region of the BDT. For details of the parameters and the optimisation procedure, see Ch. 5.	264

Chapter 1

Introduction

Over the course of the past century, humanity's understanding of the natural world has rapidly evolved. From unravelling the structure of DNA to imaging the black hole at the centre of our galaxy, dedicated research efforts have advanced our comprehension by leaps and bounds in almost all fields. Particle physics, the study of the fundamental constituents of the universe, is no exception. At the turn of the previous century, the study of matter focused on the categorisation and cataloguing of the properties of different elements, such as their optical spectra, without any understanding of the fundamental particles or forces endowing them with these properties. The first three decades of the 20th century brought the development of quantum mechanics and with it the first glimpses of understanding. Combining quantum mechanics with classical field theory and special relativity gave rise to quantum field theory, a powerful framework in which the next fifty years saw piece by piece the Standard Model (SM) of particle physics taking shape.

The SM describes the fundamental particles of matter and their interactions through three of the four fundamental forces, the strong, weak, and electromagnetic forces. It has been thoroughly tested since its inception and has so far withstood all challenges with experimental results in excellent agreement with theoretical predictions. Despite this robustness, there are several unanswered questions which lie beyond the scope of the SM.

- What is the origin of neutrino masses?
- What is dark matter?
- Why is there such a significant matter-antimatter asymmetry in the universe?
- How can gravity, the fourth force, be included in the model?

Additionally, there are several aspects of the SM which, while accommodated within the model, have no a priori justification.

- Why is the Cabbibo-Kobayashi-Maskawa matrix, which we will introduce shortly, almost diagonal?
- Why are there three generations of the fundamental matter particles of the universe?
- Why is there an apparent fine-tuning of the bare Higgs mass to cancel large corrections?
- Are the 18 free parameters of the SM related by a deeper theory?
- Why is there no violation of charge-parity symmetry in the strong interaction?

To answer these questions it is necessary to introduce physics beyond the standard model (BSM), commonly also referred to as new physics. It is the hunt for new physics, which may appear in any number of ways, that currently occupies the efforts of particle physicists. This search is approached from two directions: direct searches for new particles predicted by new physics models, and precision measurements of SM quantities, which may be modified by new physics processes through loop-level couplings. One area particularly suited for such indirect searches is the flavour sector, which describes the three generations of matter.

Within the SM, quarks, the fundamental particles that comprise hadronic matter, can ‘change their flavour’, transitioning between the generations through charged weak interactions. This is possible as the mass eigenstates, which propagate, do not align with the weak eigenstates, which interact. The rotation between these bases is given by the Cabbibo-Kobayashi-Maskawa (CKM) matrix, a 3×3 unitary matrix.

The CKM is heavily over-constrained, there are nine entries but due to its unitary nature only four free parameters, three rotation angles, and one complex phase which is responsible for all charge-parity symmetry violation within the SM. Precisely this over-constrained nature of the CKM lends itself to searches for subtle signs of BSM physics via precision measurements of parameters associated with the CKM matrix. Three parameters are expected to be free of new-physics effects due to being measured at tree level, and thus serve as an anchor against which the remaining measurements can be compared. They are the magnitudes of the elements $|V_{ub}|$ and $|V_{cb}|$, which govern the transitions between bottom- and up- or charm-type quarks, respectively, and the angle ϕ_1 which arises from the unitarity relations of the CKM. For such comparisons to be effective, these three parameters must be precisely determined.

This thesis details measurements related to the determination of $|V_{ub}|$ at two experiments: the Belle experiment and its successor, the Belle II experiment. Both experiments are B -factories producing millions of $B\bar{B}$ pairs. The rate at which these B mesons, containing a bottom quark and an up- or down-type antiquark, decay to charmless final states is directly proportional to $|V_{ub}|^2$. At first order such measurements are therefore conceptually simple: one measures the number of B mesons, and hence bottom quarks, produced at the experiment and counts the number which decay to a charmless final state, indicating that a $b \rightarrow uW$ transition occurred. In practice, the measurements are complicated for two reasons. First, due to quark confinement, one must also consider the dynamics of the b quark within the initial B meson and strong interactions between the final states, which modify the decay rate and kinematic distributions. To simplify the problem, $|V_{ub}|$ (and analogously $|V_{cb}|$) are extracted from semileptonic decays, $\bar{B} \rightarrow X_u \ell \bar{\nu}$, where in the theoretical description the leptonic current factorises from the hadronic current. Second, as $|V_{cb}|^2 \sim 100|V_{ub}|^2$, the charmed decay, $b \rightarrow cW$, dominates, creating an overwhelming background from which the number of charmless decays which have occurred must be carefully extracted.

There are two theoretically and experimentally complementary approaches to the meas-

urement of $\bar{B} \rightarrow X_u \ell \bar{\nu}$: one can search for a specific final state, such as $\bar{B} \rightarrow \pi \ell \bar{\nu}$, or impose no requirements on the hadronic system, thus performing what is referred to as an inclusive measurement. Theoretically, the second approach relies on quark-hadron duality, the idea that once a sufficient fraction of hadronic final states are considered in the measurement, quark level calculations, with some non-perturbative corrections, accurately describe the physics. Experimentally, one can search for an energetic charged lepton and perform a measurement in a region of phase space in which the dominant $\bar{B} \rightarrow X_c \ell \bar{\nu}$ decay, which shares a similar event topology with one high-energy lepton, is kinematically forbidden due to the mass of the charm quark. However, the Fermi motion of the b quark within the B meson has a large impact on the fraction of events expected within this limited region. As this is poorly understood such measurements carry a large theoretical uncertainty. To reduce this uncertainty the measurements must push into the charm dominated regime. This thesis presents a measurement at the Belle experiment that does so, covering approximately 87% of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate. This is achieved through an experimental technique known as tagging, which allows for an inclusive reconstruction of the kinematics of the hadronic system and thus for the kinematics of the neutrino, which does not interact with the detector to be inferred. We measure the partial branching fraction of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ relative to the partial branching fraction of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ to extract $|V_{ub}|/|V_{cb}|$. The substructure of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ component, which dominates, is not well understood. We therefore utilise a data-driven description of this component.

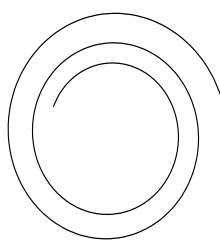
The second measurement presented within this thesis is performed at the Belle II experiment and is concerned with what is referred to as *weak annihilation*. The inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay is theoretically described as an operator product expansion in the heavy quark effective theory. At next-to-next-to-leading order in this expansion a term arises that describes a contribution roughly analogous to purely leptonic decays $B^- \rightarrow \ell^- \bar{\nu}$. The initial quarks annihilate via the weak interaction, hence the name, with an additional hadronic system originating from gluon emission. Theoretical models of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ spectra largely do not consider this term, as it enters at higher order than what the calculations have been performed to. Thus, the measurements of $|V_{ub}|$ are subject to an additional uncertainty, the size of the weak annihilation contribution. At present, this uncertainty is subleading compared to other theoretical and experimental uncertainties. However, with Belle II aiming to collect a data sample fifty times the size of its predecessor, many of these uncertainties are expected to reduce; weak annihilation is thus expected to become one of the larger sources of systematic uncertainty within the next decade. Little is known about the shape of the weak annihilation contribution beyond that it is expected to be concentrated toward the momentum transfer endpoint, where most of the energy of the initial B is concentrated in the lepton-neutrino system with a soft hadronic system. We therefore explore this region, searching for structure, in order to place limits on the size of the contribution. This search will be the first to constrain weak annihilation in the charged and neutral B meson decays separately.

In addition to these two measurements we present developments of machine learning

based particle identifications within the calorimeter of the Belle II experiment. These developments will improve future semileptonic measurements at the experiment.

The remainder of this thesis is organised as follows. Part I, provides a brief, contextual, overview of relevant background information. Chapter 2 introduces the theory and experimental state of semileptonic decays. Chapter 3 discusses the colliders and detectors whose data samples are analysed in this thesis. Chapter 4 introduces several methods and key techniques used in the work presented within this thesis. Part II discusses the implementation, training, and validation of the machine learning based particle identification in the Belle II calorimeter. The remaining chapters detail the experimental measurements performed. Part III discusses the modelling of the key signal and background channels in both measurements. Part III.A details the measurement of the partial branching fraction of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ relative to the partial branching fraction of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ at the Belle experiment. Finally, part III.B details the search for weak annihilation contributions to the inclusive semileptonic B decay rate at the Belle II experiment.

Throughout this thesis natural units, $\hbar = c = 1$, are used. Unless otherwise stated, the inclusion of charge-conjugate mode decays is implied, and decays without clear indication of charge, ie. $\bar{B} \rightarrow X \ell \bar{\nu}$, indicate the admixture of charged and neutral modes. Unless otherwise stated, or clear from context, the term lepton is taken to include only electrons and muons and their antiparticles ($\ell = e, \mu$).



Part I

Background

Theory and State of Semileptonic B -decays

We introduce the Standard Model of particle physics, describing the particle content and fundamental forces of the universe with a focus on weak interactions and the Cabibbo–Kobayashi–Maskawa matrix. Next, we review the theoretical descriptions of inclusive semileptonic weak decays and where weak annihilation enters the game. This theoretical overview is accompanied by a summary of the state of relevant experimental measurements.

The Standard Model (SM) is formulated as a relativistic quantum field theory, defined by the invariance of the SM Lagrangian under local gauge transformations of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge groups. Invariance under $SU(3)_C$ gives rise to the strong force described by Quantum Chromodynamics (QCD). Invariance under $SU(2)_L \times U(1)_Y$ yields the unified electroweak force whose symmetry is spontaneously broken by the Higgs mechanism, into the weak and electromagnetic forces, the latter described by Quantum Electrodynamics (QED). The symmetry breaking additionally generates masses for the massive fermions and bosons.

Within the SM, matter is composed of fermions, with spin $1/2$, grouped into charged and neutral leptons (also known as neutrinos), and up- and down-type quarks. All fermions interact via the weak force¹, charged leptons and quarks via the electromagnetic force, and only quarks via the strong force. Experimentally, three generations of fermions have been observed². The three generations of up-type quarks are called up (u), charm (c) and top (t), the three down-type quarks down (d), strange (s) and bottom (b), and the three charged leptons electron (e), muon (μ) and tauon (τ). The three neutrinos are known by their corresponding charged lepton (ν_e, ν_μ, ν_τ). Each generation of charged leptons and up- and down-type quarks is more massive than the previous generation, but otherwise identical. The three generations of neutrinos are nominally massless, although the observation of neutrino oscillations [4] indicates that they possess non-zero masses. The ordering of their masses and the mechanism by which they acquire them are not yet known.

¹With the caveat that only left handed particles interact with charged weak bosons.

²There is no a priori reason for there to be three generations. At least three are necessary to introduce CP violation into the SM; however, the amount of CP violation introduced in this manner is insufficient to explain the observed matter dominance in the universe. Searches for additional generations have placed stringent bounds on a fourth generation, requiring quark masses in excess of 1 TeV, a charged lepton mass of 100 GeV, and a neutrino mass of 60 GeV [1–3].

Forces are mediated by gauge bosons, with spin 1. The electromagnetic interaction by the massless photon (γ), the weak interaction by three massive bosons, two electromagnetically charged ($W^\pm$) and one electromagnetically neutral (Z), and the strong force by eight massless gluons (g).

2.1 Charged weak interactions and the Cabibbo–Kobayashi–Maskawa matrix

In the SM Lagrangian the interaction between quarks and leptons and W bosons is given by

$$\mathcal{L}_{\text{SM}} \supset -\frac{g_2}{\sqrt{2}} (\bar{\nu}_i \gamma^\mu P_L \ell_i + \bar{u}_i \gamma^\mu P_L V_{ij}^{\text{CKM}} d_j) W_\mu^+ + h.c., \quad (2.1)$$

where g_2 is the coupling constant of the weak interaction and $P_L \equiv \frac{1}{2}(1 - \gamma_5)$ is the left-handed projection operator selecting out the left-handed components of the fermion fields, reflecting the chiral nature of the weak interaction. The first term gives the interaction of charged leptons and neutrinos with W^+ bosons³, the second term being the interaction of the W^+ with up- and down-type quarks. Within the SM Lagrangian quarks are usually considered in their mass basis, defined by the eigenstates of the kinetic term in the Lagrangian arising from interactions with the Higgs field post symmetry breaking. The weak interaction, however, interacts with quarks in their weak eigenstates, which are rotated from the mass basis. The term V_{ij}^{CKM} is an entry of the Cabibbo–Kobayashi–Maskawa matrix (CKM) [6, 7] which encodes the rotation of the weak quark eigenstates from the mass eigenstates. It is given by:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{\text{Weak}} = \underbrace{\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}}_{V_{\text{CKM}}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{Mass}}. \quad (2.2)$$

The weak interaction, and more specifically, the charged weak interaction, thus does not conserve quark flavour, unlike the strong and electromagnetic interactions. For many particles, the only allowed decay channel is, therefore, via the weak interaction. Additionally, by decomposing the weak Lagrangian into vector and axial-vector terms, it can be shown that the vector part is invariant under parity transformations while the axial-vector is not, introducing the only known source of charge-parity (CP) violation in the SM.

The CKM matrix is unitary, and it can be shown that, due to this unitary nature and the freedom to absorb some parameters into the phases of the quarks, there are four free parameters within the matrix. These are fundamental parameters of the SM and must be measured. The standard parameterisation [8] consists of three real angles, θ_{12}, θ_{13} , and

³In the minimal SM, as presented here, leptons cannot mix. To introduce neutrino oscillations, one must include the PMNS matrix [5] analogous to the CKM in function, if not in origin.

θ_{23} , and one complex phase, δ , and is given by

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \quad (2.3)$$

where s_{ij}, c_{ij} are the trigonometric functions $\sin(\theta_{ij})$ and $\cos(\theta_{ij})$.

Experimentally, it has been determined that quarks strongly prefer to transition within their generation, leading to an alternate parameterisation exploiting the approximately diagonal nature of the CKM. This is known as the Wolfenstein [9] parameterisation with free parameters λ, ρ, η, A , based on an expansion in $\lambda = |V_{us}|$:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (2.4)$$

The unitarity of the CKM-matrix can be expressed as a set of nine relations given by

$$\sum_i V_{ij} V_{ik}^* = \delta_{jk} : \quad (2.5)$$

three relations state that the sum of coupling of any up-type quark to all down-type quarks is unity, and six that correspond to triangles in the complex plane. One of the triangles is of particular interest as all sides are of the same order and all related observables accessible in experiments. It is conventionally referred to as the ‘unitarity triangle’ and given by

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \quad (2.6)$$

In practice, the triangle is normalised by $V_{cd}V_{cb}^*$, the only unknown is therefore the position of the apex, given by $(\bar{\rho}, \bar{\eta})$, where $\bar{\rho} = \rho(1 - \lambda^2/2)$ and $\bar{\eta} = \eta(1 - \lambda^2/2)$. The sides of the triangle are given by the relations:

$$R_u = \left| \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{\bar{\rho}^2 + \bar{\eta}^2} = (1 - \lambda^2/2) \frac{1}{\lambda} \frac{|V_{ub}|}{|V_{cb}|}, \quad (2.7)$$

$$R_t = \left| \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{(1 - \bar{\rho})^2 + \bar{\eta}^2} = \frac{1}{\lambda} \frac{|V_{td}|}{|V_{cb}|}, \quad (2.8)$$

and the angles⁴ by:

$$\phi_1 = \arg \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right), \phi_2 = \arg \left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*} \right), \text{ and } \phi_3 = \arg \left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right). \quad (2.9)$$

It should also be noted that the area of the unitarity triangle provides a measurement of the amount of CP violation present within the SM quark sector.

Various processes can be used to constrain the apex of the unitarity triangle. The constraining parameters, and the processes used to measure them include [1]:

⁴An alternative convention exists labelling $\phi_1 \equiv \beta$, $\phi_2 \equiv \alpha$, and $\phi_3 \equiv \gamma$.

- ϵ_K , describing the size of indirect CP violation in the $K^0 - \bar{K}^0$ system;
- Δm_d and Δm_s , the mass splitting of $B_{d/s}$ mesons, extracted from $B_{d/s}$ oscillations;
- $\sin 2\phi_1$, extracted from time-dependent measurements of CP asymmetries in decays of $B^0, \bar{B}^0$ to a common final state containing a charm quark, such as $B \rightarrow J/\psi K_S^0$;
- ϕ_2 , extracted from time-dependent measurements of CP asymmetries in $b \rightarrow u\bar{u}d$ decays;
- ϕ_3 , extracted from measurements of CP violation in $B \rightarrow DK$ decays; and
- $|V_{ub}|$ and $|V_{cb}|$, extracted primarily from semileptonic B decays.

A graphical representation of the unitarity triangle and current experimental constraints on the apex is shown in Fig. 2.1. Constraints fall into two categories, those measured in SM tree-level decays ($|V_{ub}|$, $|V_{cb}|$, ϕ_3), and the remainder, which are measured through loop-level processes. The three tree-level constraints on the unitarity triangle are expected to be free from BSM effects, and thus fix the apex to its SM value. New-physics effects are expected to appear as either heavy particles running within the loop or new tree-level processes with small coupling constants such that SM tree-level processes would be dominant, but SM loop-level processes may occur with similar magnitudes. Comparisons of loop-level constraints with the SM apex thus serve as powerful tests for new physics. At present, global fits to all constraints, and subsets of constraints, show good consistency [10, 11].

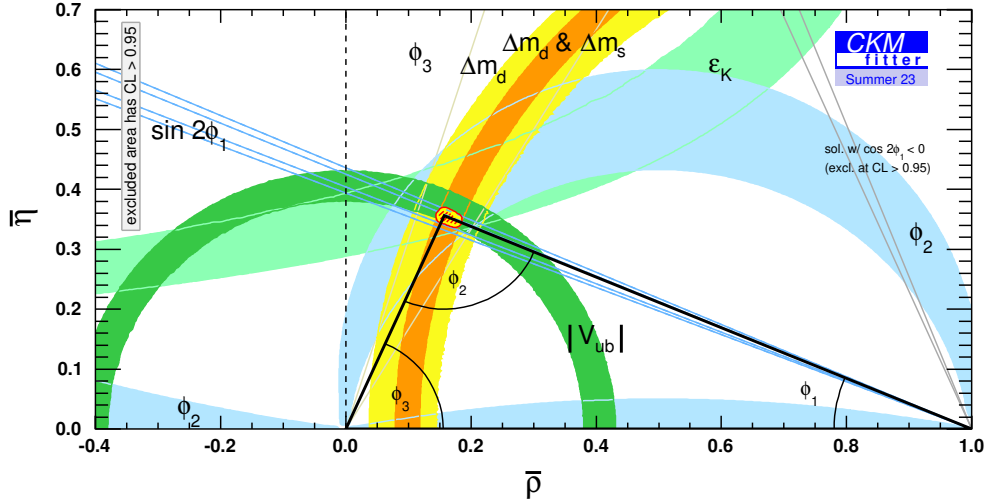


Figure 2.1: The unitarity triangle and global fit to the position of its apex using all experimental constraints available before summer 2023 [10].

2.2 Semileptonic B Decays

Both $|V_{ub}|$ and $|V_{cb}|$ can be extracted from the decay of b quarks, the rate of which is directly proportional to the square of the CKM element magnitude. Due to colour confinement however, no free quarks exist, and thus both quantities are conventionally extracted from semileptonic decays of B mesons⁵. To a good approximation, such decays can be modelled as the decay of a free quark with a non-interacting spectator quark. To fully capture the dynamics of the decay, especially near the allowed kinematic endpoints, the importance of which will be expanded upon below, it is, however, necessary to model the interactions with the spectator quark. A Feynman diagram of the $\bar{B} \rightarrow X \ell \bar{\nu}$ decay, where X is the child hadronic system, illustrating a set of possible strong interactions within the hadronic system is shown in Fig. 2.2.

The choice of semileptonic decays is beneficial for both experimental and theoretical reasons. Experimentally, the final-state charged lepton provides an easily identifiable signature. Theoretically, as the lepton-neutrino pair does not interact strongly, it is possible to factorise the leptonic and hadronic currents in the decay, greatly simplifying the calculation. For the following overview, we broadly follow Refs. [12, 13].

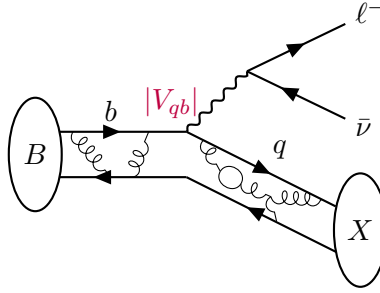


Figure 2.2: Feynman diagram of the decay $\bar{B} \rightarrow X \ell \bar{\nu}$.

As the decay occurs at an energy scale of M_B and $M_W \gg M_B$, it is possible to take the limit as $M_W \rightarrow \infty$, giving an effective four-fermion coupling. The decay is described by the effective weak Hamiltonian:

$$\mathcal{H}_W = \sqrt{2}G_F V_{qb} (\bar{q} \gamma^\mu P_L b) (\bar{\ell} \gamma_\mu P_L \nu) + h.c., \quad (2.10)$$

where G_F is the Fermi coupling constant. The transition matrix element of a semileptonic decay to a specific final state meson, X , can then be written as:

$$\mathcal{M}(\bar{B} \rightarrow X \ell \bar{\nu}) = \langle X \ell \bar{\nu} | \mathcal{H}_W | B \rangle, \quad (2.11)$$

$$\mathcal{M}(\bar{B} \rightarrow X \ell \bar{\nu}) = \sqrt{2}G_F V_{qb} \langle X | \bar{q} \gamma^\mu P_L b | B \rangle \langle \ell \bar{\nu} | \bar{\ell} \gamma_\mu P_L \nu | 0 \rangle, \quad (2.12)$$

⁵As B mesons are the lightest bound state that contains a b quark, they can only decay weakly, and are thus extremely suitable for measuring $|V_{ub}|$ and $|V_{cb}|$.

where the transition matrix element has been factorised to a hadronic current $H_\mu = \langle X | \bar{q} \gamma^\mu P_L b | B \rangle$, and a leptonic current $L^\mu = \langle \ell \bar{\nu} | \bar{\ell} \gamma_\mu P_L \nu | 0 \rangle$. The leptonic current is free of QCD effects and can be calculated with good accuracy.

The hadronic current is more difficult to calculate due to the running of α_s , the QCD coupling constant. Due to the need to resum ultraviolet divergences in a process known as renormalisation α_s increases at lower energies. The precise strength depends on the energy scale of renormalisation Q , and is given by:

$$\alpha_s(Q) = \frac{12\pi}{33 - 2N_f \ln(Q^2/\Lambda_{\text{QCD}}^2)}, \quad (2.13)$$

where $N_f = 6$ is the number of quark flavours. As α_s increases, calculations based on perturbation theory in α_s break down, and non-perturbative effects begin to contribute significantly. Formally, the divergence occurs at Λ_{QCD} , which is experimentally found to be approximately 200 MeV, but is itself dependent on the energy scale of renormalisation. It should also be noted that as parameters of the QCD Lagrangian the quark masses also have to be renormalised and their value thus depends on the renormalisation scheme, of which there are several with varying benefits, and the renormalisation scale.

At energies typical of bound hadronic states such as in a B meson, α_s is of $\mathcal{O}(1)$. The strong interactions within the hadronic system, which are dominated by soft gluon exchange, thus cannot be described perturbatively. There are two approaches to describe these interactions, depending on whether one is describing a specific meson final state, known as exclusive, or whether one sums over all possible final states, known as inclusive. The approaches are theoretically, and to a large extent experimentally, complementary.

2.2.1 Exclusive decays

Although the hadronic current cannot be directly calculated due to its non-perturbative nature, its general structure can be inferred from studying the Lorentz structure of specific transitions. For a transition of a pseudoscalar such as a B meson to another pseudoscalar, labelled P , it can thus be shown that

$$H_\mu = \langle P | \bar{q} \gamma^\mu P_L b | B \rangle = f_+(q^2)(p_B + p_P)_\mu + f_-(q^2)(p_B - p_P)_\mu, \quad (2.14)$$

where q is the four momentum transfer to the lepton-neutrino system, and p_B and p_P are the four-momenta of the B meson and pseudoscalar, respectively. The functions $f_+(q^2)$ and $f_-(q^2)$, are *form factors*, encoding the low-energy dynamics of the transition. Once contracted with the leptonic current, in the limit of massless leptons – a good approximation for electrons and muons – the term proportional to $f_-(q^2)$ vanishes. The total decay rate is thus given by

$$\frac{d\Gamma}{dq^2} = \frac{1}{24\pi^2} G_F^2 |V_{qb}|^2 |\vec{p}_P^B|^3 f_+^2(q^2), \quad (2.15)$$

where $|\vec{p}_P^B|$ is the magnitude of the pseudoscalars momentum in the B rest frame. It is evident that it is not possible to extract $|V_{ub}|$ or $|V_{cb}|$ from the decay of B mesons to pions

or D mesons, respectively, without external knowledge of f_+ , or a simultaneous determination. The form factor can be calculated at $q^2 \sim q_{\text{max}}^2$ in lattice QCD, and at $q^2 \sim 0$, using so-called light cone sum rules. These points, and measured q^2 spectra of the decays, are fit simultaneously using an ansatz of the form factor, usually a polynomial expansion [14, 15], to determine the magnitudes of the CKM elements. Transitions to vector final states (D^*, ρ, ω) can be parameterised similarly, with four form factors contributing. In the limit that the bottom and charm masses are infinite, known as the heavy quark limit, additional symmetries arise which can be used to constrain the form factors in $b \rightarrow c$ decays [16, 17].

2.2.2 Inclusive decays

The treatment of inclusive decays relies on the idea of quark-hadron duality, which states that if all, or a large enough fraction of, hadronic final states are considered, in a sufficiently large region of phase space, parton-level calculations accurately describe the hadronic transitions [18]. The treatment differs for heavy-to-heavy decays, $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and heavy-to-light decays, $\bar{B} \rightarrow X_u \ell \bar{\nu}$, where the subscript u or c indicates either a $b \rightarrow u \ell \bar{\nu}$ or $b \rightarrow c \ell \bar{\nu}$ transition has occurred. We will first discuss the heavy-to-heavy transition and then the additional considerations necessary for the heavy-to-light transition.

2.2.2.1 Heavy-to-heavy

To fully describe the decay, three independent kinematic variables are needed. One possible choice is the momentum transfer squared and the energies of the lepton and neutrino. The choice is, however, not unique so we will proceed without specifying a particular choice. In the rest frame of the B meson, the triple differential decay rate is given by

$$d^3\Gamma = \frac{1}{M_B} \sum_{X_c} \sum_{s_\ell, s_{\bar{\nu}}} |\langle X_c \ell \bar{\nu} | \mathcal{H}_W | B \rangle|^2 \delta^4(p_B - p_{X_c} - p_\ell - p_{\bar{\nu}}), \quad (2.16)$$

where the sums run over all possible hadronic final states, and the spins of the lepton and neutrino, $s_\ell, s_{\bar{\nu}}$, which are not determined experimentally, and thus averaged. The expression can again be factored into a leptonic and hadronic part, this time in the form of the leptonic tensor $L^{\alpha\beta}$ and the hadronic tensors $H_{\alpha\beta}$ giving

$$d^3\Gamma = \frac{1}{4\pi^3} G_F^2 |V_{cb}|^2 L^{\alpha\beta} H_{\alpha\beta}, \quad (2.17)$$

where the leptonic tensor is easily calculable. The hadronic tensor is given by

$$H_{\alpha\beta} = \frac{(2\pi)^3}{2M_B} \sum_{X_c} \langle B | h_\alpha^\dagger | X_c \ell \bar{\nu} \rangle \langle X_c \ell \bar{\nu} | h_\beta | B \rangle \delta^4(p_B - p_{X_c} - p_\ell - p_{\bar{\nu}}). \quad (2.18)$$

It is then possible to relate the hadronic tensor to the amplitude of the forward scattering of the B -meson through the optical theorem as

$$H_{\alpha\beta} = \frac{-1}{2\pi M_B} \text{Im} \int d^4x e^{-iq \cdot x} \langle B | T[h_\alpha^\dagger(x), h_\beta(0)] | B \rangle, \quad (2.19)$$

where $T[h_\alpha^\dagger(x), h_\beta(0)]$ is the time-ordered product. It is possible to write the time ordered product as an operator product expansion known as the Heavy Quark Expansion (HQE), [19–21]. It is given by

$$T[h_\alpha^\dagger(x), h_\beta(0)] = \sum_{n,i} \frac{1}{M_b^n} C_{n,i} O_{n+3,i}, \quad (2.20)$$

where $O_{n+3,i}$ is a set of non-perturbative operators of dimension $n+3$ and $C_{n,i}$ are their respective Wilson coefficients encoding the perturbative QCD. The Wilson coefficients are themselves expansions in $\alpha_s(M_b)$ which can be calculated in heavy quark effective theory. The decay rate is then an expansion in $1/M_b$, more precisely in Λ_{QCD}/M_b , taking the form

$$\Gamma = \Gamma_0 + \frac{1}{M_b} \Gamma_1 + \frac{1}{M_b^2} \Gamma_2 + \frac{1}{M_b^3} \Gamma_3 + \dots \quad (2.21)$$

The leading operators, at $n=0$, can be related to the matrix element $\langle B | \bar{b} \gamma_\mu b | B \rangle$, which denotes the decay rate of the free b quark, and hence Γ_0 gives the decay rate of the free quark. All operators at order $n=1$ can be expressed as $n=2$ operators using the equation of motion of the b quark and hence $\Gamma_1 = 0$. At order $n=2$ two non-perturbative contributions appear which can be written in terms of matrix elements:

$$2M_b \mu_\pi^2 = -\langle B | \bar{b}_\nu (iD)^2 b_\nu | B \rangle, \quad (2.22)$$

$$2M_b \mu_G^2 = \langle B | \bar{b}_\nu \sigma_{\alpha\beta} (iD^\alpha) (iD^\beta) b_\nu | B \rangle, \quad (2.23)$$

where D^α is the covariant derivative, $\sigma_{\alpha\beta} = \frac{i}{2} [\gamma_\alpha, \gamma_\beta]$, and the field b_ν is defined as $b_\nu = e^{iM_b(\nu \cdot x)} b(x)$. This definition originates from heavy quark effective theory where it is convenient to decompose the momentum of the b quark to $M_b \nu + k$, where $M_b \nu$ arises from the motion of the B meson with four-velocity ν , and k is the residual momentum describing the motion within the B meson and is of order Λ_{QCD} . The covariant derivative acting on this field thus yields the residual momentum k and therefore μ_π^2 is proportional to k^2 , the kinetic energy of the b quark inside the meson. The second term μ_G gives the chromomagnetic moment.

At order $n=3$, a further two matrix elements enter given by

$$2M_b \rho_D^3 = \langle B | \bar{b}_\nu (iD_\alpha) (i\nu D) (iD^\alpha) b_\nu | B \rangle, \quad (2.24)$$

$$2M_b \rho_{LS}^3 = \langle B | \bar{b}_\nu \sigma_{\alpha\beta} (iD^\alpha) (i\nu D) (iD^\beta) b_\nu | B \rangle, \quad (2.25)$$

where ρ_D^3 and ρ_{LS}^3 are known as the Darwin and spin-orbit terms. At higher order the number of matrix elements rapidly increases with nominally nine elements contributing at order $1/M_b^4$ [22].

The non-perturbative parameters, $\mu_\pi^2, \mu_G^2, \rho_D^3, \rho_{LS}^3, \dots$ cannot be calculated and must be extracted from data. This is usually done by fitting the theoretical predictions to experimentally determined moments of the lepton energy spectrum and the hadronic mass spectrum, with the fit extracting the non-perturbative parameters, and M_b , and $|V_{cb}|$. See Ref. [23] for an example of such a fit.

2.2.2.2 Heavy-to-light

In principle it is possible to compute the inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ transition rates using the heavy quark expansion as for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (see for example Ref. [24]). Experimentally, however, it is difficult to access $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events under the large background of $\bar{B} \rightarrow X_c \ell \bar{\nu}$, favoured by a factor of $|V_{cb}|^2/|V_{ub}|^2 \sim 100$. For early inclusive measurements of $|V_{ub}|$ [25–29] this has required restricting the measurement to small regions of phase space in which due to the mass of the charm $\bar{B} \rightarrow X_c \ell \bar{\nu}$ is kinematically forbidden. This corresponds to regions such as low hadronic mass, the endpoints of the lepton momentum or momentum transfer squared distributions, or combinations thereof. These regions are demonstrated in Fig. 2.3. In such restricted regions the operator product expansion is no longer valid, and calculations of the inclusive rates must take a different approach.

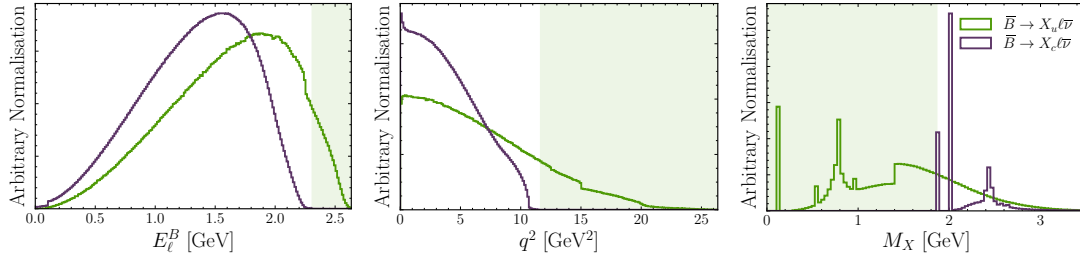


Figure 2.3: Kinematic distributions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (green) and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (purple). Regions in which $\bar{B} \rightarrow X_c \ell \bar{\nu}$ is kinematically forbidden are shaded. The distributions correspond to the nominal MC sample for the first measurement presented within this thesis, as described in Ch. 7. The $\bar{B} \rightarrow X_c \ell \bar{\nu}$ distribution is modelled as a sum of exclusive channels, the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ distribution as a combination of exclusive and inclusive channels. The steps in the distributions are caused by the combination of different channels.

The triple differential decay rate, here expressed in terms of the lepton energy and the light cone variables $P_+ = E_X - |\vec{p}_X|$, and $P_- = E_X + |\vec{p}_X|$, which depend on the energy and momentum magnitude of the hadronic system in the B frame is given by

$$\frac{d^3\Gamma}{dP_+ dP_- dE_\ell} = \frac{1}{192\pi^3} G_F^2 |V_{ub}|^2 \int dk_+ C(P_+, P_-, E_\ell, k_+) \mathcal{F}(k_+) + \mathcal{O}(\Lambda_{\text{QCD}}/M_b), \quad (2.26)$$

where k_+ is the light cone momentum and the choice of kinematic variable is again not unique. The function $C(P_+, P_-, E_\ell, k_+)$ is perturbatively calculable. The function $\mathcal{F}(k_+)$ is known as a *shape function* [30]. The shape function is non-perturbative and describes the Fermi motion of the b quark within the meson and smears the triple differential distribution, especially near the parton level kinematic endpoint. The moments of the leading-order shape function can be related to the HQE parameters:

$$\mathcal{F}(k_+) = \delta(k_+) + \frac{\mu_\pi^2}{6} \delta''(k_+) + \frac{\rho_D^3}{18} \delta'''(k_+) + \dots, \quad (2.27)$$

where the notation δ' indicates the derivative of the Dirac delta distribution. The leading-order shape function is known to be universal for all heavy-to-light transitions and can

thus be extracted from the photon energy spectrum of $\bar{B} \rightarrow X_s \gamma$ decays [31, 32], where, being a nominally two-body decay, the shape function is more easily constrained as it is solely responsible for the width of the distribution. At present, none of the existing models directly incorporate an experimentally determined shape function, instead use a variety of functional approximations. Ref. [33] proposes simultaneous fits to the spectra of $\bar{B} \rightarrow X_s \gamma$ and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ to fully exploit this information. A fit to just the $\bar{B} \rightarrow X_s \gamma$ photon energy spectrum has been performed in Ref. [34]. Alternatively, Ref. [35] proposes to approximate the shape function using neural networks based on both the information of the photon energy spectrum and the constraints on the moments of the distribution.

The treatment of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ differs in the available models. We therefore provide a brief summary of the most commonly considered models in the following (see also Ref. [36]).

DeFazio and Neubert

The model of DeFazio and Neubert (DFN) [37] calculates the triple differential decay rate up to $\mathcal{O}(\alpha_s)$ to be:

$$\begin{aligned} \frac{d^3\Gamma}{dx dz d\hat{\rho}^2} = 12\Gamma_0[(1 + \bar{x} - z)(z - \bar{x} - \hat{\rho}^2)\frac{M_b^2}{2}W_1 + (1 - z + \hat{\rho}^2)\frac{M_b}{2}W_2 \\ + (\bar{x}(z - \bar{x}) - \hat{\rho}^2)\frac{M_b}{4}(W_3 + 2M_b W_4 + M_b^2 W_5)], \end{aligned} \quad (2.28)$$

where $x = 2E_\ell/M_b$, $\bar{x} = 1 - x$, $\hat{\rho}^2 = p^2/M_b^2$, $z = 2(v \cdot p)/M_b$, v is the b quark velocity, p the momentum of the light degrees of freedom, and W_i are so called structure functions which originate from a general Lorentz invariant decomposition of the hadronic tensor. Each structure function is then expanded in a perturbation series in α_s as:

$$W_i(z, \hat{\rho}^2) = W_i^{(0)}(z, \hat{\rho}^2) + \frac{\alpha_s}{3\pi} W_i^{(1)}(z, \hat{\rho}^2) + \mathcal{O}(\alpha_s^2), \quad (2.29)$$

for the full expressions see Ref. [37]. These expressions are then convoluted with a shape function using an exponential ansatz

$$\mathcal{F}(k_+) = N(1 - \frac{k_+}{\bar{\Lambda}})^a e^{(1+a)\frac{k_+}{\bar{\Lambda}}}, \quad (2.30)$$

where $\bar{\Lambda} = M_B - M_b$, and a and N are set to meet the conditions of Eq. 2.27, yielding $\frac{1}{3}\mu_\pi^2 = \bar{\Lambda}^2/(1 + a)$.

DFN is generally considered to have been superseded by the GGOU and BLNP models, described later. It is however still used in the generation of Monte Carlo simulated samples as will be discussed in Ch. 7.

Bosch, Lange, Neubert, and Paz

The model of Bosch, Lange, Neubert, and Paz [38, 39] considers the triple differential decay rate in the kinematic variables P_+ , P_- and $P_\ell = M_B - 2E_\ell$, finding it to be

$$\begin{aligned} \frac{d^3\Gamma}{dP_+dP_-dP_\ell} = \frac{G_F^2|V_{ub}|^2}{16\pi^3} (M_B - P_+) & [(P_- - P_\ell)(M_B - P_- + P_\ell - P_+) \tilde{W}_1 \\ & + (M_B - P_-)(P_- - P_+) \frac{\tilde{W}_2}{2} \\ & + (P_- - P_\ell)(P_\ell - P_+) \left(\frac{y}{4} \tilde{W}_3 + \tilde{W}_4 + \frac{1}{y} \tilde{W}_5 \right)], \end{aligned} \quad (2.31)$$

where $y = (P_- - P_+)/ (M_B - P_+)$ and $\tilde{W}_i$ are structure functions. They can be written as an expansion in $1/M_b$ as

$$\tilde{W}_i \sim H \cdot J \otimes S + \frac{1}{M_b} \sum_k h \cdot J \otimes s^k + \frac{1}{M_b} \sum_k h \cdot j^k \otimes S + \mathcal{O}(\frac{1}{M_b^2}), \quad (2.32)$$

where the key technique is to factorise physics at different scales, such that H , J and S are the leading hard, jet, and shape functions, h , j , and s their subleading counterparts, and $\otimes$ indicates a convolution. The hard term describes physics at a scale of M_b , the jet term at $\sqrt{M_b\Lambda_{\text{QCD}}}$ and the shape function at Λ_{QCD} . The hard and jet terms are calculable with perturbation theory. This expression holds near the kinematic endpoint, in the so called shape function region, and smoothly transitions to the HQE based expression once a large enough region of phase space is considered.

At present the model considers corrections to the hard and jet functions at $\mathcal{O}(1/M_b)$ to $\mathcal{O}(\alpha_s)$, subleading shape functions at $\mathcal{O}(1/M_b)$, and terms of the OPE up to $\mathcal{O}(1/M_b^2)$. Three analytic forms of the leading-shape function are considered, an exponential, a Gaussian, and a hyperbolic form. The parameters of these are tuned to match the moments.

Ref. [40] has calculated NNLO corrections to the hard and jet functions, finding them to decrease the estimated value of $|V_{ub}|$ by up to 8%. These corrections have not yet been implemented in the BLNP model. Several additional advances in the knowledge of H and J are summarised in Ref. [41].

Gambino, Giordano, Ossala, and Uraltsev

The model of Gambino, Giordano, Ossala, and Uraltsev [42] considers the triple differential decay rate in terms of the lepton energy, E_ℓ , the momentum transfer to the leptonic system squared q^2 , and the total energy of the leptonic system q_0 . The rate is given by:

$$\begin{aligned} \frac{d^3\Gamma}{dE_\ell dq^2 dq_0} = \frac{G_F^2|V_{ub}|^2}{8\pi^3} & \left[q^2 W_1 - \left[2E_\ell^2 - 2q_0 E_\ell + \frac{q^2}{2} \right] W_2 + q^2 (2E_\ell - q_0) W_3 \right] \\ & \times \theta \left(q_0 - E_\ell - \frac{q^2}{4E_\ell} \right) \theta(E_\ell) \theta(q^2) \theta(q_0 - \sqrt{q^2}), \end{aligned} \quad (2.33)$$

where θ are Heaviside functions, and W_i are again structure functions, of which only the three relevant in the limit of massless charged leptons, a suitable approximation for electrons and muons, are considered. The structure functions are factorised through the use of a hard Wilsonian cutoff at scale μ as

$$W_i(q_0, q^2) \propto \int dk_+ F_i(k_+, q^2, \mu) W_i^{\text{pert}}(q_0 + \frac{k_+}{2}(1 - \frac{q^2}{M_B M_b}), q^2, \mu). \quad (2.34)$$

where W_i^{pert} contains the perturbatively calculable contributions above μ , and the distribution functions, F_i , the non-perturbative contributions, including the Fermi motion. The distribution functions absorb all subleading corrections such that they are no longer universal, and for each structure function, there is a unique distribution function. The distribution functions are, like the shape functions, parameterised in terms of a set of possible functional forms (exponential, Gaussian, hyperbolic, or Roman [43]). The first two moments of $W_i(q_0, q^2)$ in q_0 , are computed in the HQE and used to constrain the parameters of the distribution functions. All corrections to $\mathcal{O}(1/M_b^3)$ and $\mathcal{O}(\alpha_s^2 \beta_0)$ are included in the calculation.

Other models

The models discussed so far are, in terms of their treatment of the shape or distribution functions, considered to be model independent. Alternative models exist which avoid introducing these parameterised functions. These include the dressed-gluon exponentiation (DGE) of Andersen and Gardi [44] and the model of Aglietti, Di Lodovico, Ferrera and Ricciardi (ADFR) [45]. DGE considers an on-shell b quark and uses resummed (using dressed-gluon exponentiation) perturbation theory to compute the triple differential decay spectrum. The resummation of soft gluons naturally describes the smearing without the need for a shape function. ADFR likewise performs a resummation of soft-gluon effects and pairs this with an analytic effective QCD coupling without a Landau pole to perturbatively calculate the decay spectra. For both models, no systematic uncertainty has been assigned to their underlying assumptions, they are thus typically viewed to be less reliable than BLNP or GGOU.

A comparison of the lepton energy and hadronic mass spectra predicted by different models is shown in Fig. 2.4, BLNP and GGOU are found to generally agree well in the lepton energy spectrum, while ADFR and DGE predict harder spectra.

2.3 Experimental status of $|V_{cb}|$ and $|V_{ub}|$

Experimentally $|V_{ub}|$ and $|V_{cb}|$ are primarily extracted from measurements of semileptonic channels at B -factory experiments, which will be introduced in more detail in Ch. 3. Measurements can be tagged or untagged. For tagged measurements the presence of the second B -meson produced in the $\Upsilon(4S) \rightarrow B\bar{B}$ decay studied at the B -factories is exploited to infer information of the signal B and select clean samples at high cost to

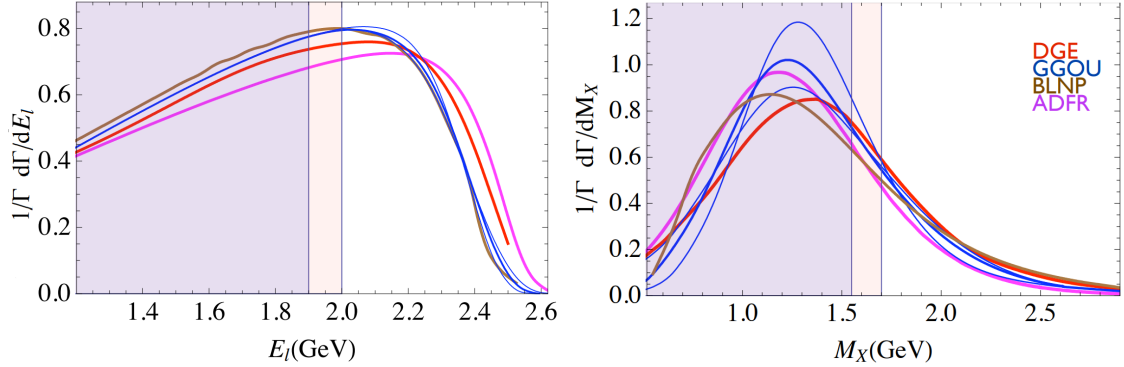


Figure 2.4: Comparison of lepton energy and hadronic mass spectra predicted by different inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ models, GGOU, BLNP, DGE, and ADFR, as of 2010, using common inputs [46]. Also appears in [47]. For GGOU three lines are shown using different functional forms for the distribution functions as described in the text. The shaded areas indicate common selection thresholds in analysis $E_\ell^B > 1.9$ or 2.0 GeV, or $M_X < 1.55$ or 1.7 GeV.

the reconstruction efficiency. Untagged measurements, in contrast, consider only particles from the decay of the signal B meson, thus have access to less information but at greater reconstruction efficiencies. These techniques are further discussed in Sec. 4.2.

Exclusive $|V_{cb}|$ is extracted from the hadronic recoil spectra, closely related to q^2 , of exclusive $\bar{B} \rightarrow D \ell \bar{\nu}$ decays [48–50], or from the hadronic recoil spectra and angular information of $\bar{B} \rightarrow D^* \ell \bar{\nu}$ decays [49, 51–57], reaching 2.5% and 2% precision, respectively. As previously discussed inclusive $|V_{cb}|$ is determined from fits [23, 58] to the moments of the kinematic distributions of inclusive $\bar{B} \rightarrow X_c \ell \bar{\nu}$ [59–64]. The Particle Data Group [65] reports world averages of

$$|V_{cb}|_{\text{exc.}} = (39.8 \pm 0.6) \times 10^{-3}, \quad (2.35)$$

$$|V_{cb}|_{\text{inc.}} = (42.2 \pm 0.5) \times 10^{-3}. \quad (2.36)$$

An approximate 3σ discrepancy is observed between the inclusive and exclusive results. This has become known as the $|V_{cb}|$ puzzle. This disagreement, with varying significance, is long standing, Fig. 2.5 presents the inclusive and exclusive world averages reported by the Particle Data Group over the previous 20 years. The result of a global fit to the unitarity triangle without semileptonic B inputs finds $|V_{cb}|_{\text{CKM.Fitter}} = (41.1^{+0.7}_{-0.4}) \times 10^{-3}$ [10], favouring the inclusive result.

Exclusive $|V_{ub}|$ can be extracted from $\bar{B} \rightarrow \pi \ell \bar{\nu}$, $\bar{B} \rightarrow \rho \ell \bar{\nu}$, and $\bar{B} \rightarrow \omega \ell \bar{\nu}$, however, the pion channel is known with significantly greater precision than the vector-meson channels. Therefore, the exclusive $|V_{ub}|$ value is extracted [66] from the averaged $\bar{B} \rightarrow \pi \ell \bar{\nu}$ q^2 spectra of Refs. [67–70].

The inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ measurements can be grouped into two categories: studies of the region above or near the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ kinematic endpoint [25, 28, 71], which are

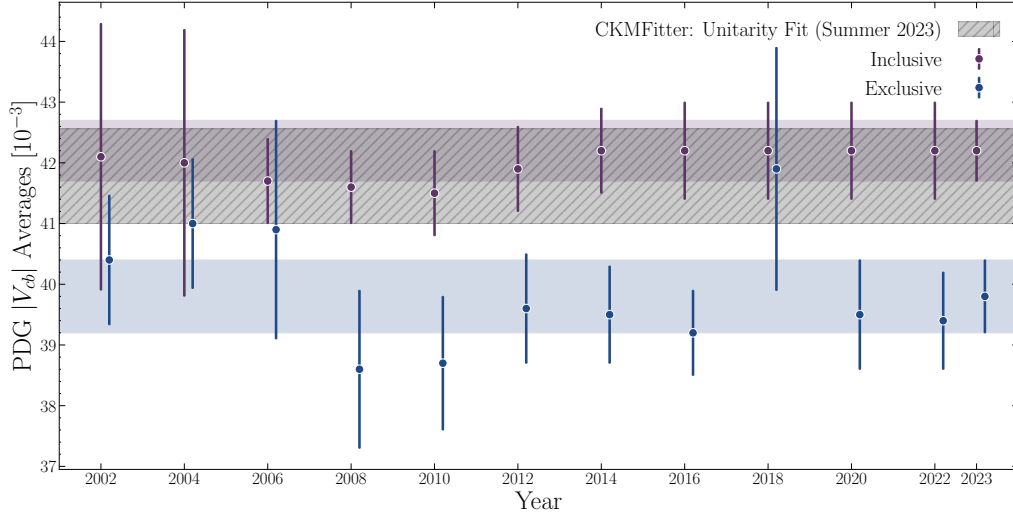


Figure 2.5: World averages of exclusive and inclusive $|V_{cb}|$ determinations as reported by the Particle Data Group over the past 20 years [65]. The most recent averages are additionally shown as bands. The $|V_{cb}|$ value determined by a global fit to the unitarity triangle without semileptonic inputs is given as a hatched grey band [10].

generally untagged measurements, searching for highly energetic leptons; or studies which reach into the charm-dominated regime using tagged techniques [72, 73], and generally measure several sub-regions of phase space that are of theoretical interest. Fig. 2.6 presents the inclusive $|V_{ub}|$ values determined in each measurement that currently enters the world average. Measurements are usually performed using the predictions of a single model to extract the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ partial branching fraction and then translated into $|V_{ub}|$ using all available models. Ref. [74], an untagged study of the electron momentum spectrum in $\bar{B} \rightarrow X e \bar{\nu}$ for $E_\ell^{B_{\text{sig}}} > 0.8 \text{ GeV}$, is unique in that the measurement is performed using four $\bar{B} \rightarrow X_u \ell \bar{\nu}$ models. As can be seen in Fig. 2.6, compared to the other measurements, this study reports greater differences between the value of $|V_{ub}|$ determined with each model, highlighting the additional dependence on modelling that arises in estimating the shape and reconstruction efficiency of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ component.

The Particle Data Group reports world averages of

$$|V_{ub}|_{\text{exc.}} = (3.70 \pm 0.10 \pm 0.12) \times 10^{-3}, \quad (2.37)$$

$$|V_{ub}|_{\text{inc.}} = (4.13 \pm 0.12^{+0.13}_{-0.14} \pm 0.18) \times 10^{-3}. \quad (2.38)$$

where the first errors are experimental, the second theoretical and for the inclusive measurement the third error is given by the spread of results using different theoretical models. The result of a global fit to the unitarity triangle without semileptonic B inputs finds $|V_{ub}|_{\text{CKM.Fitter}} = (3.64^{+0.07}_{-0.07}) \times 10^{-3}$ [10], favouring the exclusive result. The averages for the last 20 years are presented in Fig. 2.7, where, similar to the $|V_{cb}|$ determinations, the inclusive result has been consistently larger than the exclusive. At present the two res-

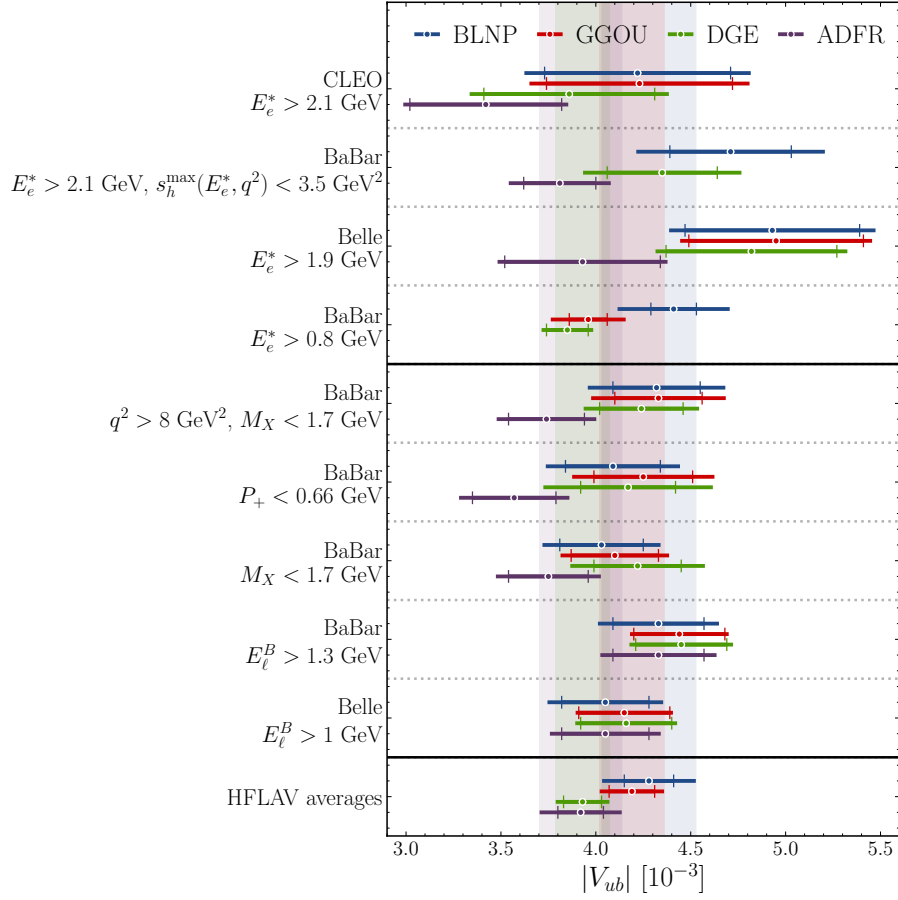


Figure 2.6: Measurements entering the inclusive $|V_{ub}|$ world average using the prescriptions of ADFR, DGE, BLNP and GGOU, and the kinematics region in which the measurements were performed. From top to bottom these correspond to Refs. [25], [71], [28], [74], [72]×4, [73]. The solid horizontal line divides untagged and tagged measurements. The last entry gives the averages as determined by the Heavy Flavour Averaging Group [66]. The inner error bars are experimental, the outer experimental and theoretical.

ults agree within 1.4σ . Recently the Belle collaboration also performed a simultaneous measurement of inclusive and exclusive $|V_{ub}|$, finding compatible results [75].

Finally, considering the ratio $|V_{ub}|/|V_{cb}|$, which directly constrains the length of the unitarity triangle side R_u . The world averages are:

$$|V_{ub}|/|V_{cb}|_{\text{exc.}} = (9.3 \pm 0.4) \times 10^{-2}, \quad (2.39)$$

$$|V_{ub}|/|V_{cb}|_{\text{inc.}} = (9.8 \pm 0.6) \times 10^{-2}, \quad (2.40)$$

which are in excellent agreement.

The LHCb collaboration has also directly extracted $|V_{ub}|/|V_{cb}|$ from $B_s \rightarrow (D_s, K)\ell\bar{\nu}$

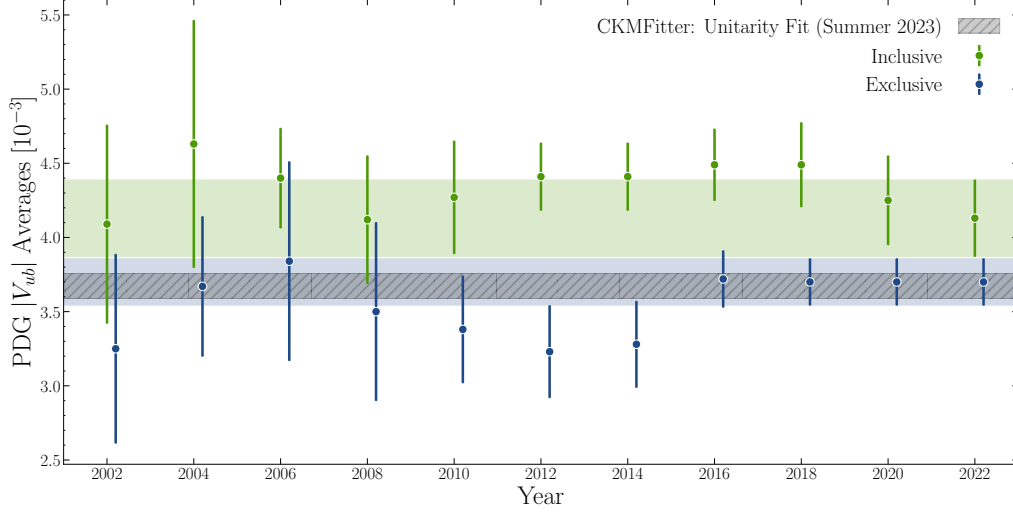


Figure 2.7: World averages of exclusive and inclusive $|V_{ub}|$ determinations as reported by the Particle Data Group over the past 20 years [65]. The most recent averages are additionally shown as bands. The $|V_{ub}|$ value determined by a global fit to the unitarity triangle without semileptonic inputs is given as a hatched grey band [10].

and $\Lambda_b \rightarrow (\Lambda_c, p)\ell\bar{\nu}$. For the baryonic modes the collaboration finds

$$|V_{ub}|/|V_{cb}| = (7.9 \pm 0.4 \pm 0.4) \times 10^{-2}, \quad (2.41)$$

where the first error is experimental, and the second arises from lattice QCD calculations of the relevant form factors at high q^2 [76]. For the $B_s \rightarrow (D_s, K)\ell\bar{\nu}$ mode, $|V_{ub}|/|V_{cb}|$ is extracted in two q^2 regions, using light cone sum rule based calculations of the form factors at low q^2 , and lattice QCD calculations at high q^2 . They find

$$|V_{ub}|/|V_{cb}|_{q^2 < 7 \text{ GeV}^2} = (6.07 \pm 0.21 \pm 0.30) \times 10^{-2}, \quad (2.42)$$

$$|V_{ub}|/|V_{cb}|_{q^2 > 7 \text{ GeV}^2} = (9.46 \pm 0.41 \pm 0.68) \times 10^{-2}, \quad (2.43)$$

where the first errors include all experimental uncertainties and uncertainties due to external inputs, and the second errors give the uncertainty due to the form factor calculations [77]. The Flavor Lattice Averaging Group has performed a fit for the relevant form factors to all available lattice QCD information and extrapolates the form factors to low q^2 . They find

$$|V_{ub}|/|V_{cb}|_{q^2 < 7 \text{ GeV}^2} = (8.19 \pm 0.29 \pm 0.72) \times 10^{-2}, \quad (2.44)$$

$$|V_{ub}|/|V_{cb}|_{q^2 > 7 \text{ GeV}^2} = (8.60 \pm 0.38 \pm 0.37) \times 10^{-2}, \quad (2.45)$$

observing excellent agreement between the low and high q^2 regions [78].

2.3.1 A brief note on leptonic decays

Purely leptonic decays, $B^- \rightarrow (e^-, \mu^-, \tau^-)\bar{\nu}$, provide a theoretically clean alternative to semileptonic decays for the extraction of $|V_{ub}|$, with decay rates given by:

$$\Gamma(B^- \rightarrow \ell^- \bar{\nu}) = \frac{G_F^2}{8\pi} |V_{ub}|^2 f_B^2 M_B M_\ell^2 \left(1 - \frac{M_\ell^2}{M_B^2}\right)^2, \quad (2.46)$$

where f_B , is the B meson decay constant. Leptonic decays are however helicity suppressed⁶ with expected branching fractions for electrons, muons, and taus of $\mathcal{O}(10^{-12})$, $\mathcal{O}(10^{-7})$ and $\mathcal{O}(10^{-5})$, respectively [10]. The low branching fractions, combined with the difficulty of measuring the tauonic channel due to the presence of multiple neutrinos, make such measurements uncompetitive for extracting $|V_{ub}|$ at present. The most recent measurements, using the full Belle and BaBar data samples, find evidence of the muonic channel with a 2.8σ significance [79], and the tauonic channels with between 2.3σ and 3.8σ using hadronic and semileptonic tagging techniques [80–83]. Combined, the tauonic measurements allow a determination of $|V_{ub}|$ with an uncertainty of 16% [78]. With the expected full Belle II data sample of 50 ab^{-1} , the tauonic channel is expected to reach experimental precisions of 2.5% [84], providing a theoretically clean third avenue to determine $|V_{ub}|$.

2.3.2 Projections to the full Belle II sample

Projections of the expected precision of exclusive and inclusive determinations of $|V_{ub}|$ for the full 50 ab^{-1} sample expected at Belle II have been made in Ref. [85]. Exclusive determinations are expected to reach precisions between 2–3% depending on improvements in lattice QCD inputs to the form factor. Inclusive determinations are expected to reach precisions of 3–4.8%, with theoretical uncertainties of 2.5–4.5%, dominating the expected 1.7% experimental uncertainty.

2.4 Weak annihilation

In the operator product expansion for $\bar{B} \rightarrow X_u \ell \bar{\nu}$, additional operators enter at $\mathcal{O}(1/M_b^3)$ that are not present in the expansion for $\bar{B} \rightarrow X_c \ell \bar{\nu}$. They correspond to four-quark interactions, annihilating the b quark with a $\bar{u}$ giving rise to the Feynman diagram in Fig. 2.8. The $\bar{u}$ can be a valent quark of a B^- meson, or it can be a non-valent quark, originating from strong interactions within the B -meson system [86]. In the context of inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ analyses, this contribution is commonly labelled weak annihilation (WA) and has been studied in Refs. [87–91].

⁶As the B meson has spin 0, the lepton and anti-neutrino must be emitted with opposite spins. The weak interaction couples only to left-handed particles and right-handed anti-particles. As the anti-neutrino is massless, its right-handed chirality translates to right-handed helicity, where helicity is defined as the projection of the particle's spin onto its momentum. To preserve a net spin of 0 the lepton must also have right-handed helicity. As leptons have mass, the left-handed lepton can have small right-handed helicity components, giving rise to the M_ℓ^2 term in Eq. 2.46.

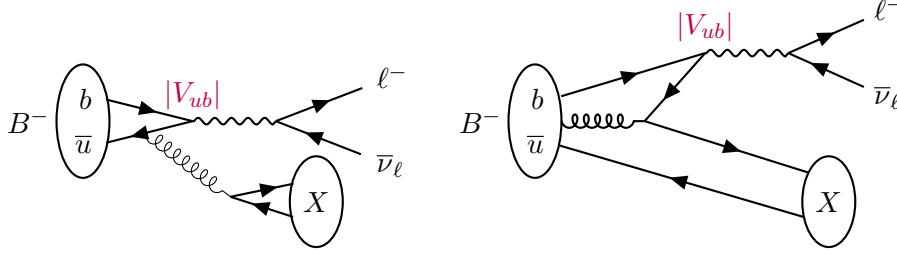


Figure 2.8: Feynman diagram of the valent (*left*) and nonvalent (*right*) $\bar{B} \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$ decay.

The WA contribution to the total inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate can be expressed as

$$\delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu}) = \frac{1}{12\pi} G_F^2 |V_{ub}|^2 f_B^2 M_B^2 M_B (B_2 - B_1), \quad (2.47)$$

where we note an enhancement of $16\pi^2$ compared to all other dimension 5 and 6 operators, and where B_1, B_2 are so called ‘bag constants’ parameterising the matrix elements of four quark operators [89, 90]. These are given by:

$$\frac{1}{8} f_B^2 M_B B_1 = \langle B | \bar{b}_\nu \gamma^\mu P_L u \bar{u} \gamma_\mu P_L b_\nu | B \rangle, \quad (2.48)$$

$$\frac{1}{8} f_B^2 M_B B_2 = \langle B | \bar{b}_\nu P_L u \bar{u} P_R b_\nu | B \rangle, \quad (2.49)$$

where $P_R \equiv \frac{1}{2}(1 + \gamma_5)$ is the right-handed projection operator. In the naive factorisation limit in which the matrix elements are saturated by the vacuum insertion, $|0\rangle\langle 0|$, both constants are equal, their value depending on the flavour of the spectator quark, $B_1 = B_2 = 1$ for B^+ , $B_1 = B_2 = 0$ for B^0 . In this limit, the weak annihilation contribution thus vanishes. Factorisation is however expected to be violated at up to the 10% level, allowing for a weak annihilation contribution of approximately 3% of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay rate [89, 92].

Formally, little is known about the shape of the weak annihilation contribution beyond an approximate location in phase space, near the maximum possible values of lepton energy and momentum transfer, with the contribution smearing about $\delta(q^2 \sim M_b^2)$ [89] and $\delta(E_\ell^B \sim M_b/2)$ [88]. The form of the smearing is largely unknown but is expected to smear the distributions in a range of order Λ_{QCD} . The hadronic system is thus expected to be soft [88]. Ref.[90] studied the contribution in the lepton energy, noting that unlike the shape functions previously discussed, the first moment of the structure function describing the smearing does not vanish and thus the mean of the distribution may shift about the quark level endpoint.

Weak annihilation can be considered in two parts: the nonvalent term, taken to contribute equally to the charged and neutral B mesons, and the valent contribution, responsible for the difference between the two B flavours [93, 94]. The valent contribution is affected

by helicity suppression and thus is nominally suppressed by M_ℓ^2/M_B^2 , however the suppression is expected to be partially lifted by the presence of the soft hadronic system [88]. The non-valent contribution does not suffer from helicity suppression but is expected to be suppressed by factors associated with the additional gluon exchange required to connect the annihilating light quark with the remainder of the diagram [92].

Most of the inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ models previously discussed do not incorporate weak annihilation. Its effects are assigned as a systematic uncertainty, taken to increase the reconstructed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yield in measurements and thus provide a unidirectional systematic uncertainty towards lower values of $|V_{ub}|$. GGOU mixes the WA operator with the Darwin term [42, 95], removing a singularity in the description of the high q^2 tail; however, the scale and shape of the WA contribution are left as free parameters. As weak annihilation is located near the kinematic endpoint, the experimental kinematic requirements used to reject $\bar{B} \rightarrow X_c \ell \bar{\nu}$ additionally enhance the weak annihilation contributions. The relative size of the effect therefore depends on the applied selections. Being non-perturbative the magnitude of the contribution, and thus the size of the systematic uncertainty assigned to $|V_{ub}|$, must be extracted from data. Three strategies have been pursued. They are summarised below.

Indirect constraints

Recent tagged inclusive $|V_{ub}|$ studies at Belle [73] and Babar [72] have reported partial branching fractions of $B^+ \rightarrow X_u^0 \ell^+ \nu$ and $B^0 \rightarrow X_u^- \ell^+ \nu$ in addition to that of the admixture $\bar{B} \rightarrow X_u \ell \bar{\nu}$. By considering the ratio of these partial branching fractions,

$$R_{\text{iso}} = \frac{\tau_{B^0}}{\tau_{B^+}} \frac{\Delta\mathcal{B}(B^+ \rightarrow X_u^0 \ell^+ \nu)}{\Delta\mathcal{B}(B^0 \rightarrow X_u^- \ell^+ \nu)}, \quad (2.50)$$

it is possible to estimate the contribution of weak annihilation as

$$\frac{\Gamma_{\text{WA}}}{\Gamma} = \frac{f_u}{f_{\text{WA}}} (R_{\text{iso}} - 1), \quad (2.51)$$

where Γ is the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay width and f_u, f_{WA} give the fraction of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and weak annihilation rates, respectively, included in the phase space Δ considered in the measurement of R_{iso} . As weak annihilation is expected to be confined to the high q^2 region, $q^2 \approx M_B^2$, and the measurements are sufficiently inclusive in q^2 , they assume $f_{\text{WA}} = 1$. Both collaborations report limits on the weak annihilation contribution to the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay rate, setting 90% confidence levels at $-0.14 < \frac{\Gamma_{\text{WA}}}{\Gamma} < 0.17$ and $-0.17 < \frac{\Gamma_{\text{WA}}}{\Gamma} < 0.19$, by the Belle and BaBar collaborations, respectively.

Such limits rely on the incorrect assumption that only the valent quark contributes and that the neutral B is free of weak annihilation effects. It is more accurate to consider these limits to be limits on the difference of weak annihilation contributions to the semileptonic decays of the charged and neutral B systems, ie. the valent contribution.

Direct measurement

The CLEO collaboration has published a direct search for weak annihilation in inclusive semileptonic B decays [96, 97]. The analysis used a simplified weak annihilation model, adding a soft hadronic system to a $\bar{B} \rightarrow \ell \bar{\nu}$ decay to search for structure at high q^2 . Variations of the model, changing the properties of the hadronic system, are scanned setting a model-dependent 90% confidence limit at $\frac{\Gamma_{\text{WA}}}{\Gamma} < 7.4\%$.

Extrapolation from charm decays

By heavy quark symmetry, the same matrix element governing weak annihilation in B decays governs weak annihilation in D decays. It is therefore possible to extract information from weak annihilation in D decays, where it is enhanced by a factor of M_b^3/M_c^3 , and extrapolate to B decays. This approach has been considered in Refs. [92, 93, 98] based on measurements of inclusive semileptonic decays of D^0, D^+, D_s from CLEO-c. Gambino and Kamenik [98] study the moments of the semileptonic D^0, D^+, D_s decay lepton energy spectra, and estimate the maximum contribution of weak annihilation to the B decay rate at 2%. Ligeti, Luke, and Manohar [92] study the semileptonic charm decay branching fractions, and estimate a weak annihilation contribution of around 1% to charmless semileptonic B decays. This estimate increases to 2.5% using an alternate extrapolation. Finally, Bigi, Mannel, Turczyk, and Uraltsev [93], consider the D_s and D_0 decay rates to estimate the valent weak annihilation contribution. They estimate a contribution to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate from the valent contribution of between -0.5% to 1% . It should be noted that they refer to the extrapolation from charm to bottom as ‘semi-quantitative at best’ due to insufficient heaviness of the charm. Furthermore, the three analyses indicate that, contrary to naive expectations, the contribution of weak annihilation to neutral B meson decays is greater than the contribution to charged B meson decays. This may indicate a destructive interference effect with the nominal $b \rightarrow u \ell \bar{\nu}$ transition [93].

BES-III has recently reported a new measurement of inclusive semileptonic D_s decays [99] which has not yet been considered but is in excellent agreement with the measurement of CLEO [100] and is not expected to shift these estimates.

2.4.1 Importance

At present the Heavy Flavour Averaging Group assigns uncertainties due to weak annihilation effects of -0.7% , -1.1% , and -1.1% to the world average determinations of $|V_{ub}|$ using the prescriptions of BLNP, DGE, and GGOU, respectively [66]. The total uncertainties are 5.7%, 3.2% and 3.9%, respectively, dominated by the uncertainty of the b quark mass, other non-perturbative parameters, and choice of matching scales. Weak annihilation is thus at present a subleading uncertainty. If inclusive $|V_{ub}|$ determinations are to reach 1% accuracy it is however necessary to greatly improve the understanding of weak annihilation and have a robust data-driven measurement of the size of the contribution in

the B system⁷.

2.5 Role of the analyses

Two experimental measurements are presented within this thesis. The first is a measurement of the ratio of inclusive partial branching fractions of charmless to charmed semileptonic B decays, $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$. One of the experimental challenges of such a measurement is the modelling of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ contributions. Charmed events are usually modelled as a sum of exclusive channels, $\bar{B} \rightarrow D \ell \bar{\nu}$, $\bar{B} \rightarrow D^* \ell \bar{\nu}$ and $\bar{B} \rightarrow D^{**} \ell \bar{\nu}$, which is discussed in more depth in Ch. 7. The sum of these channels does not fully saturate the known inclusive $\bar{B} \rightarrow X_c \ell \bar{\nu}$ rate, and poorly describes the measured inclusive moments of the kinematic distributions [101]. Recent measurements by the Belle and BaBar collaborations have also observed an excess of events with high lepton energies and high momentum transfer [63, 72, 73].

The modelling of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events, which dominate over the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ signal, thus introduces large systematic uncertainties and a potential source of bias in measurements of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays. We therefore adopt a data driven description of this background, extracting its shape from a control region when extracting $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yields in order to better control this potential bias. Furthermore, since the theoretical descriptions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ share many commonalities, and thus correlated systematics, not least of which a dependence on M_b^5 [45], we directly extract the ratio of partial branching fractions and thus $|V_{ub}|/|V_{cb}|$. This also cancels several experimental systematic uncertainties related to the tagging procedures and the identification of leptons. Additionally, we extract the differential ratio in lepton energy and momentum transfer which may be useful to further reduce modelling related uncertainties.

The second is preparation for a direct search for weak annihilation contributions to the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay rate by searching for structure at high q^2 in inclusive semileptonic B decays. The search is expected to be significantly more sensitive than that of CLEO, and will be the first to directly constrain the contribution to the charged and neutral B meson decay simultaneously by exploiting information gained through tagging.

⁷Alternatively, it has been proposed to measure $\bar{B} \rightarrow X_u \ell \bar{\nu}$ far away from the endpoint, enforcing a selection of $q^2 < M_B^2 - M_D^2$, to remove the weak annihilation contribution [39, 42]. As this region is enriched in $\bar{B} \rightarrow X_c \ell \bar{\nu}$, the expected experimental uncertainties associated with such a measurement make it uncompetitive at present and in the foreseeable future.

Producing and observing B -mesons

We introduce the experimental apparatus used in the measurements presented within this thesis. We briefly review KEKB and SuperKEKB, the colliders accelerating electrons and positrons to sufficient energies for their collisions to produce pairs of B mesons, and the Belle and Belle II detectors designed to observe the decay products of these mesons. The overview focuses on operational principles and selected technical specifications. While introducing the hardware, we sketch an outline of their respective reconstruction softwares, converting low-level detector information to high-level objects suitable for physics analyses.

In 1977 the E288 experiment at Fermilab reported an excess in the di-muon mass spectrum produced at by highly energetic protons impacting on copper and platinum targets [102]. This resonance, at approximately 9.5 GeV, would later become known as the $\Upsilon(1S)$. It provided the first evidence for a bound $b\bar{b}$ state and the b quark. Within a year, observations of the $\Upsilon(2S)$ and $\Upsilon(3S)$ states followed [103], and their existence was confirmed by the PLUTO and DASP collaborations at the DORIS electron-positron collider [104–106]. Just three years later, the CLEO collaboration, operating at the CESR electron-positron collider, observed the $\Upsilon(4S)$ at 10.55 GeV as an excess of hadronic events [107]. The first $b\bar{b}$ state massive enough to decay to two B mesons. In 1981 CLEO provided evidence for the B -meson with an analysis of the single-electron spectrum, following with the first full reconstruction of B mesons in four exclusive channels in 1983 [108]. The era of B physics had begun. In the forty years since, the study of B mesons and their decays has, and continues to, provide a wealth of insights into the fundamental nature of the universe.

The decay of B mesons, as a quantum mechanical process, is fundamentally random. To study rare decays it is therefore necessary to produce enormous amounts of B mesons. To this end, several generations of accelerators and detectors have been built. The detectors fall into two categories, those studying B mesons produced in proton-proton or proton-electron(positron) collisions, such as CDF, HERA-B, or LHCb currently active at the Large Hadron Collider, or those operating at electron-positron machines, such as CLEO, ARGUS, BaBar, Belle, and Belle II.

The operating conditions, and hence the design and goals, of experiments in the two categories are fundamentally different. Experiments operating at proton-proton or proton-electron(positron), collisions have access to orders of magnitude more B mesons than those

at electron-positron colliders¹. However, experiments at electron-positron colliders have several advantages.

- The environment is extremely clean, with a low particle multiplicity spread over the full solid angle of the detector, allowing for precise and highly efficient reconstruction of charged and neutral particles and the reconstruction of intermediate resonances that decay in part or in full to photons, such as π^0 , η , η' , ω , and ρ .
- The initial state is precisely known, both kinematically and in particle contents. This makes analyses of decay channels with missing particles, and/or that utilise the full reconstruction of the second B meson to infer information of the signal B meson, such as those presented within this thesis, possible.
- The $e^+e^- \rightarrow \Upsilon(4S)$ cross section, $\sigma_{\Upsilon(4S)}$, at approximately 1.1 nb [84], is large relative to the cross section of lighter quark anti-quark pairs, $\sigma_{q\bar{q}}$, known as continuum processes, $\sigma_{\Upsilon(4S)}/\sigma_{q\bar{q}} \sim 0.3$. Such processes can be a dominant background source for many analyses.
- As the collider operates just above the $\Upsilon(4S)$ threshold, reducing the collider energy by 60 MeV allows for the collection and analysis of a clean continuum background sample.
- Using asymmetric beam energies, a boost can be created, inducing spatial separation in the decay vertices of the two B mesons allowing for the measurement of time-dependent CP -violation.
- The cross sections of $e^+e^- \rightarrow \tau\tau$ and $e^+e^- \rightarrow c\bar{c}$ are within 20% of $\sigma_{\Upsilon(4S)}$ at the centre-of-mass energy $\sqrt{s} = \Upsilon(4S)$ [84] (see Tab. 3.1). Accelerators and detectors operating at the $\Upsilon(4S)$ resonance are therefore multipurpose machines, allowing for detailed studies of the τ and charm.

The study of B mesons accelerated with the arrival of the B factories, BaBar and Belle, so named as their respective accelerator complexes, PEP-II and KEKB, were capable of producing a million $B\bar{B}$ pairs per day of operation. The experiments collected data from 1999 to 2008 and 2010, respectively, achieving integrated luminosities of 514 fb^{-1} and 1041 fb^{-1} [110, 111]. The two collaborations published approximately 1200 physics papers across a wide field of topics, from precision CKM measurements to dark sector searches. Many of them have had a seminal impact on our understanding of the natural world, including the observation of CP violation in neutral B mesons, confirming the theory of Kobayashi and Maskawa, who were subsequently awarded the 2008 Nobel prize. Physics analyses of the data continue to this day, as demonstrated by the study of inclusive $|V_{ub}|/|V_{cb}|$ presented within this thesis.

¹At a centre-of-mass energy of 14 TeV the $b\bar{b}$ cross section at the LHC is $\sim 500 \mu\text{b}$ producing 10^{12} $b\bar{b}$ pairs each operating year at a luminosity of $2 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ [109]. At design luminosity, SuperKEKB is expected to produce 10^{10} $B\bar{B}$ pairs each operating year.

With the conclusion of the Belle physics run work started on SuperKEKB and Belle II, representing significant upgrades to the accelerator and detector. The first collisions of the beams occurred in April 2018. In early 2022, a maximum instantaneous luminosity of $4.65 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ was recorded [112], the highest luminosity ever achieved by a particle accelerator, a crown long held by KEKB. Belle II collected 424 fb^{-1} of integrated luminosity before a long shutdown in 2022. This initial data sample is used for the study of weak annihilation presented within this thesis. The experiment is expected to collect data until 2035, with the aim of collecting 50 ab^{-1} of integrated luminosity, exceeding its predecessor by a factor of 50.

For convenience, and following the official recommendation of the Belle collaboration, we discuss KEKB and Belle in the present tense alongside SuperKEKB and Belle II.

3.1 Production

The analyses presented in this thesis are carried out with data from the Belle and Belle II experiments at the High Energy Accelerator Research Organisation (KEK) in Tsukuba, Japan. These experiments study collisions produced by the KEKB accelerator, and its upgrade the SuperKEKB accelerator, respectively. Schematics of the complexes are shown in Figure 3.1.

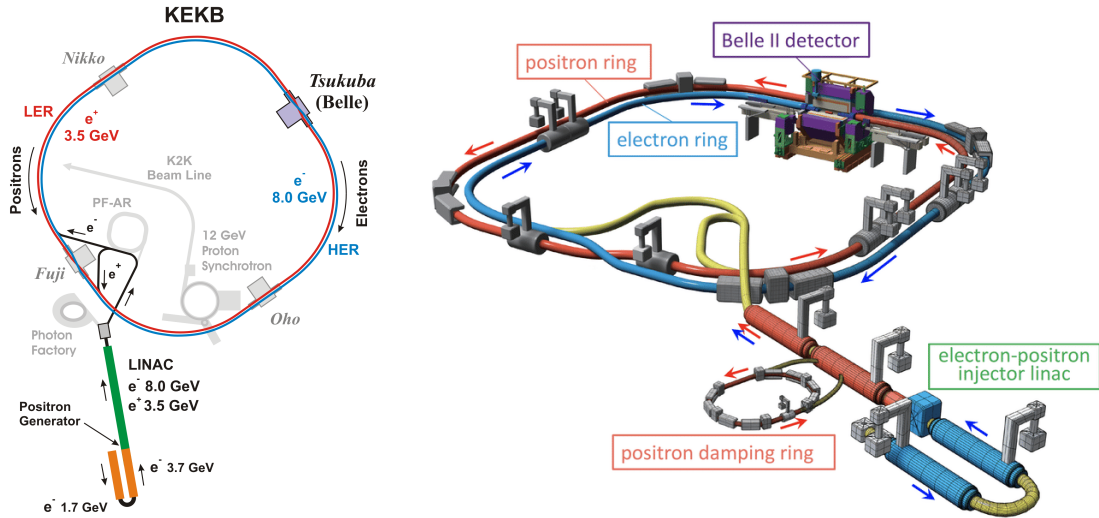


Figure 3.1: Schematics of the KEKB [113] (*left*) and SuperKEKB [114] (*right*) accelerator complexes. Not to scale.

For both accelerators, electrons are produced by a photocathode radio frequency gun and accelerated by a linear accelerator before being directed onto a tungsten target to produce positrons. The electrons and positrons are then accelerated to their nominal energies, 8 and 7 GeV for the electron beam, respectively in KEKB and SuperKEKB, and 3.5 and 4 GeV for the positron beam, respectively, before electrons are injected into

the high-energy and positrons into the low-energy storage rings. The two beams counter rotate within the 3 km circumference rings, colliding at the centre of the Belle or Belle II detectors, at a point known as the interaction point (IP). The energies of the beams can be tuned to create centre-of-mass energies from the $\Upsilon(1S)$ to the $\Upsilon(5S)$. The majority of operating time is spent at the $\Upsilon(4S)$ threshold.

SuperKEKB has been designed to achieve an ultimate luminosity of $8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$, 40 times higher than that achieved by KEKB [110]². This increase is to be achieved through two main upgrades. Firstly, the beam currents are increased by a factor of two, which directly translates to a twofold increase in luminosity. Secondly, a novel interaction scheme for the beams is employed, known as the nano-beam scheme [115, 116], which squeezes the interaction region, expected to achieve a factor 20 increase in luminosity.

The asymmetry in the beam energies was decreased to reduce particle losses from the expected increase in Touschek scattering at higher luminosity [84].

3.2 Detection

Surrounding the interaction points of the two colliders, are the Belle and Belle II detectors, respectively. These are general purpose detectors, designed to perform a range of analyses across low-multiplicity and dark-sector searches, and B , τ , and charm physics. There are several key design features of the detectors.

- A large solid angle acceptance, almost 4π .
- Ability to reconstruct the momentum of charged particles from $\sim 100 \text{ MeV}$ to 7 GeV , with high efficiency, accuracy, and precision.
- Accurate energy reconstruction of neutral particles (γ and n), which do not interact with the tracking system used to reconstruct charged particles.
- Ability to distinguish charged particles. Specifically, to identify the six charged particle species with long enough lifetimes to traverse the detector and therefore be considered *stable*. These are: electrons, muons, pions, kaons, protons, and deuterons, as well as their antiparticles.

To achieve these goals, Belle and Belle II consist of several subdetectors, each optimised for a specific goal and for its position relative to the interaction point. From the inside out, Belle is composed of the silicon vertex detector, the central drift chamber, the time of flight detector in the barrel region, the aerogel Cherenkov counter in the barrel and forwards region, the electromagnetic calorimeter and the K_L^0 -muon system. See Figure 3.2 for a schematic. Belle II is composed of the pixel detector, the silicon vertex detector, the central drift chamber, the time of propagation detector in the barrel region and its counterpart in the forward region – the aerogel ring imaging Cherenkov counter, the electromagnetic calorimeter, and the K_L^0 -muon system. See Figure 3.3 for a schematic.

²Recent projections target a maximum luminosity of $6.5 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$.

Before we discuss the subdetectors in detail, it is useful to define the coordinate system. The origin is taken to be the nominal interaction point; the z axis follows the detector symmetry axis, which approximately follows the direction of the incoming electron beam, deviating by 11 mrad in Belle and 41.5 mrad in Belle II; the y axis is taken to point upward towards the roof of the experiment hall, and the x axis points radially outward from the centre of the accelerator ring. The polar angle, θ , is defined from the positive z axis. The azimuthal angle, ϕ , gives the angle in the $x - y$ plane defined from the positive x axis. With the exception of a chimney through the K_L^0 -muon system to direct liquid cooling to the solenoid, the detectors are approximately symmetric in the $x - y$ plane. Therefore, it is common to integrate over ϕ when considering analysis level selections or detector performance. Due to the boost caused by the asymmetric beam energies, the particles produced are biased toward the positive z direction. Therefore, the detectors are not symmetric in polar angle, containing additional sub-detectors in the forward (positive z) direction that are not found in the backward direction, and the geometry of the sub-detectors is adjusted to improve performance in the forward direction.

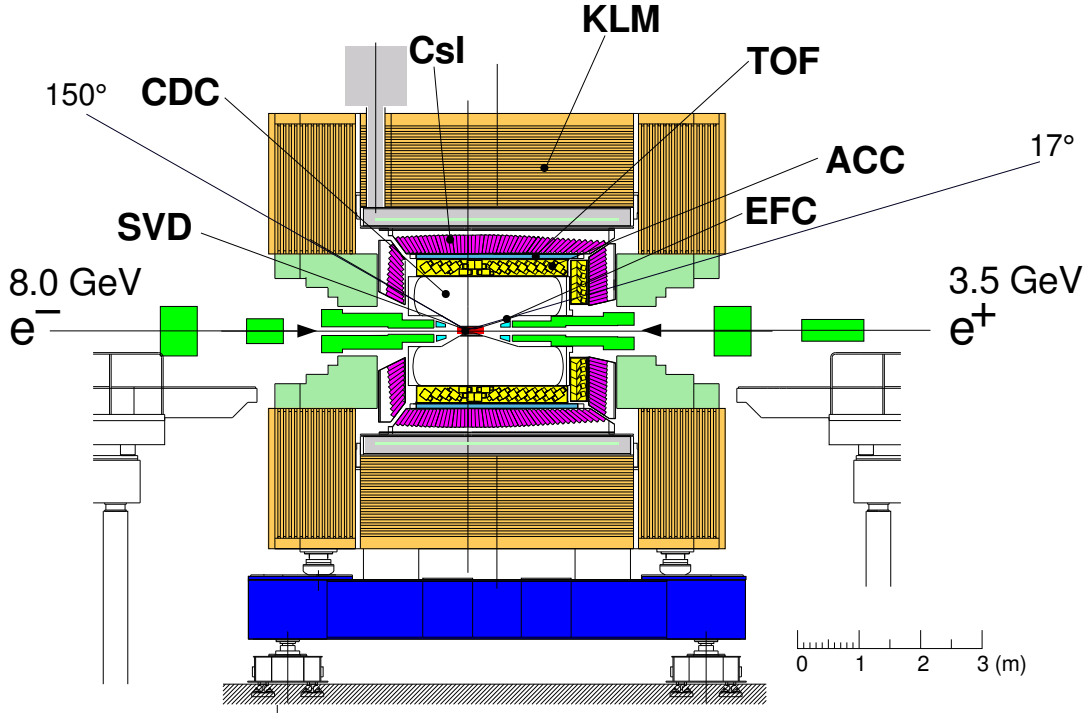


Figure 3.2: Longitudinal cross section of the Belle detector [117].

3.2.1 Tracking and vertexing

The subdetectors closest to the interaction point, immediately surrounding the beam-pipe, are designed to track charged particles with high precision to reconstruct their momentum

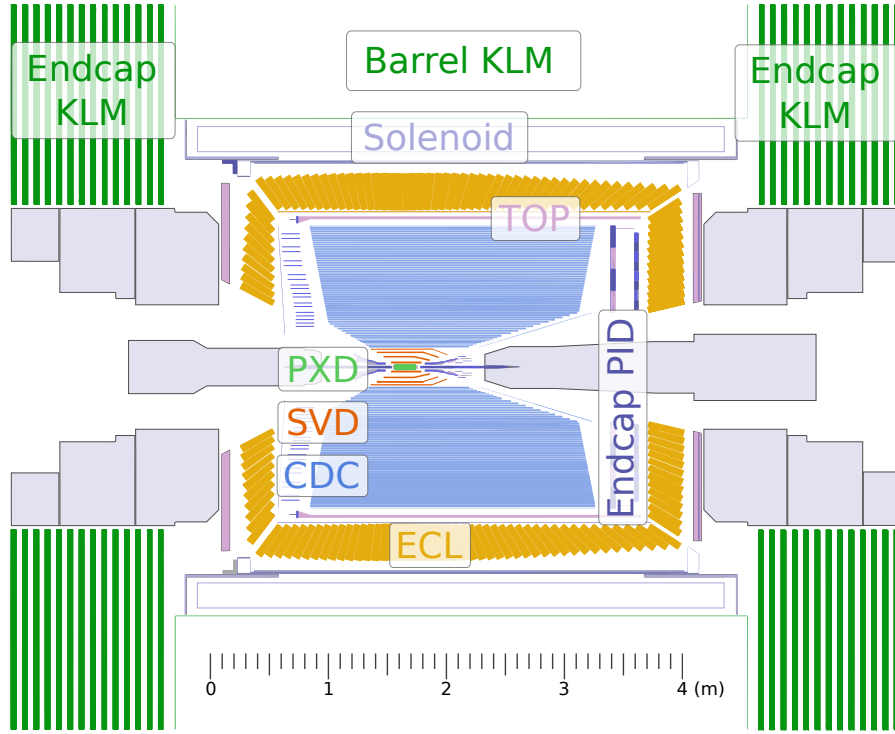


Figure 3.3: Longitudinal cross section of the Belle II detector [118]. The sparse blue lines to the left of the CDC correspond to cabling and readout electronics for the CDC, not active detector components.

and the point at which they were created, known as the production vertex. In the Belle detector this is the responsibility of the silicon vertex detector (SVD) and the central drift chamber (CDC). In Belle II these two sub-detectors are augmented by the pixel detector (PXID). Each subdetector works on the same physical principle: that energetic electromagnetically charged particles ionise material they traverse. Each, however, utilises different technologies based on the conditions the subdetector is subject to, and whether the subdetector aims at precise vertex reconstruction or tracking over a large spatial volume.

3.2.1.1 Silicon vertex detector

The innermost detector of Belle is the silicon vertex detector, designed to precisely measure the vertices of B meson decays. The SVD consists of double-sided silicon strip sensors, where strips on the long side run parallel to the z axis and strips on the short side run transverse. When a charged particle passes through a sensor it produces electron-hole pairs. The electrons drift towards one side while the holes drift to the other due to an applied voltage. The current is detected in strips on both sides of the sensor. The position can thus be constrained to the overlap area of the two orthogonal strips [119].

Initially, the Belle SVD consisted of three layers at radial distances 3.0, 4.54 and 6.05 cm from the IP, with a polar angle acceptance of $23\text{--}140^\circ$ [117]. In late 2003 the SVD was replaced with a four-layer design, known as SVD2, to improve the radiation tolerance of the electronics and reduce the dead time. The four layers were placed at radial distances of 2.0, 4.35, 7.0 and 8.8 cm from the IP with an improved polar angle acceptance of $17\text{--}150^\circ$ [120]. A schematic of SVD2 is shown in Figure 3.4.

In Belle II, due to the expected high occupancy, silicon strip detectors are not suitable for use extremely close to the IP. Thus, a new type of detector, the pixel detector, is used, immediately surrounding the beam pipe. This subdetector is discussed in more detail below. Surrounding this detector, is a four-layer SVD with layers at 3.9, 8.0, 10.4, and 13.5 cm [121, 122]. In the forward direction the three outer layers have trapezoidal sensors tilted towards the beam line to improve acceptance for tracks in the very forward direction. Acceptance extends from $17\text{--}150^\circ$ degrees in polar angle [119]. A cross section of the Belle II SVD can be seen in Figure 3.4.

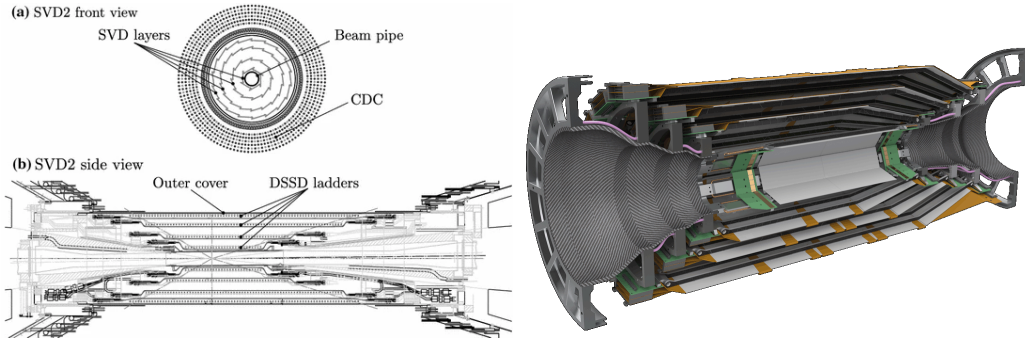


Figure 3.4: *Left:* Cross section of the Belle SVD2 [123]. *Right:* Cross section of the Belle II SVD [124]. The sensitive areas are in light grey.

3.2.1.2 Pixel detector

The PXD consists of two layers of silicon pixel sensors of depleted p-type field effect (DEPFET) sensors located at 1.4 and 2.2 cm radially from the interaction point. The layers nominally consist of eight and twelve ladders; however, due to production delays, all data collected before the first long shutdown, that is, all data utilised in the analysis presented within this thesis, was collected with just two ladders in the second layer. As precise vertexing does not play an important role in the presented analysis, the impact of this is minimal. A render of the full and partial PXD is shown in Figure 3.5.

Each ladder consists of two modules, attached end to end. Each is subdivided into 250×768 pixels of varying sizes, from $50 \times 50 \mu\text{m}^2$ at the centre of the inner layer to $50 \times 85 \mu\text{m}^2$ in the outer layer [119]. When a charged particle travels through the n^- doped bulk of the sensor it creates electron-hole pairs. The electrons drift towards an internal gate embedded into the bulk below a standard MOSFET source-gate-drain structure.

The accumulated electrons modulate the current between the source and drain. This modulation is read out, and the pixel reset through application of an external voltage [125].

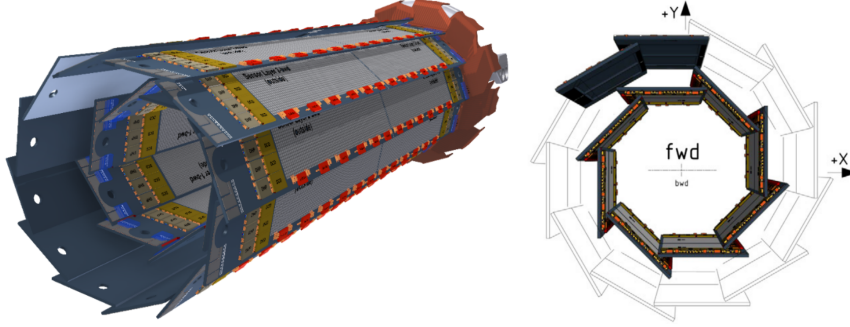


Figure 3.5: *Left:* Side view of the full PXD [126]. *Right:* Front view of the PXD as installed in all early data taking [127]. The grey surfaces are the sensitive regions.

3.2.1.3 Central drift chamber

The respective central drift chambers (CDC) of Belle and Belle II are the main tracking detectors of the experiments. They consist of wires strung across the detector in a helium ethane atmosphere. The wires act as either field wires, generating an electric field due to applied voltage, or sense wires, which are read out. They are arranged such that eight field wires form a square around each sense wire, forming a drift cell, with field wires shared between cells. Traversing charged particles ionise the gas, electrons then drift towards the sense wire, causing secondary ionisations. The current caused by the avalanche of electrons is detected as a hit in the CDC. Wires are grouped into superlayers, alternating in alignment between pointing along the axial direction (axial) and being slightly skewed (stereo) to improve resolution in the z -direction. Figure 3.6 illustrates the wire arrangement and gives an example of the signal left by a particle traversing the Belle II CDC.

The Belle CDC consists of 8400 drift cells with a radial coverage from 10.35 cm to 86.3 cm [117]. Acceptance extends from 17 – 150° in polar angle [117]. The Belle II CDC covers a larger volume, occupying the detector volume radially from 16.8 to 111.4 cm [119] with the same polar angle acceptance. A total of 56576 wires are used, forming 14336 drift cells.

An average charged particle is recorded in tens of hits in the drift chambers (about 70 hits in the Belle II CDC [130]). Combined these give a large lever arm, through which the momenta of charged particles can be precisely measured, achieving sub-percent accuracy [131].

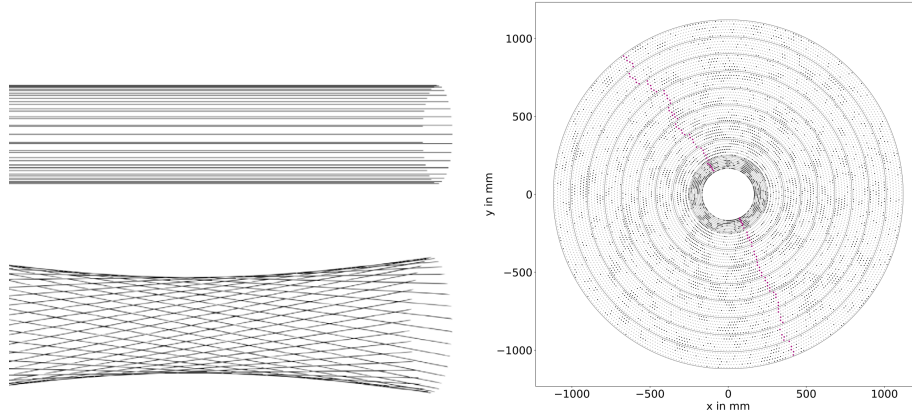


Figure 3.6: *Left:* Side view of the axial (top) and stereo (bottom) configuration. The skew of the stereo wires has been exaggerated [128]. *Right:* Simulated signals in the Belle II CDC in the $x-y$ plane. Hits caused by a cosmic ray muon traversing the detector are given in purple, hits from beam background in black [129].

3.2.1.4 Track reconstruction

Recorded hits from tracking detectors are filtered and grouped into collections of hits consistent with the signal of a charged particle traversing the detectors. In Belle tracks are searched for in the CDC and extrapolated into the SVD. Hits in the SVD that match the extrapolated trajectory are added to the track. In Belle II, tracks are searched for in the CDC then extrapolated into the SVD, remaining SVD hits are separately combined into tracks and extrapolated to the CDC. Both types of tracks are extrapolated to the PXD where any matching hits are added before performing a final track fit. An overview of the Belle II track finding algorithms is provided in Ref. [128].

For both Belle and Belle II the entire inner detector is subject to a spatially quasi-uniform 1.5 T magnetic field created by a solenoid situated between the calorimeter and the K_L^0 -muon system. The magnetic field causes the trajectory of charged particles to bend in the $x-y$ plane. Neglecting energy losses, they therefore follow helical paths, with the radius of the helix determined by the transverse momentum of the particle. The track fit therefore provides a measurement of the transverse momentum, and the point of closest approach of each track to the coordinate system origin, d_0 and z_0 , or to the interaction point dr and dz , where d_0 and dr are the respective radial distances, and z_0 and dz the respective longitudinal distance.

Tracks serve as a starting point for physics analyses. Tracks and neutral particles are combined to form resonances. The point of closest approach of the tracks to each other gives the decay vertex of the resonance. Vertex fitting algorithms [132, 133] can adjust track parameters by their uncertainty to originate from a single position.

3.2.2 Calorimeter

The electromagnetic calorimeter (ECL) is designed to measure the energy of charged and neutral particles. It consists of 8736 homogeneous thallium-doped caesium iodide (CsI(Tl)) crystals and was reused in Belle II with upgraded readout electronics. Each crystal is the shape of a truncated pyramid with an average cross section of $6 \times 6 \text{ cm}^2$ and a depth of 30 cm, corresponding to 16.2 radiation lengths³, sufficient to absorb the most highly energetic electrons and photons produced at KEKB and SuperKEKB with minimal leakage [134]. In comparison, the amount of material between the interaction point and the ECL is between 0.3–0.8 radiation lengths [111]. The crystals are wrapped in reflective material with a pair of photodiodes on the back recording the electromagnetic shower produced in the crystals, giving a measurement of the deposited energy. The ECL is split into three regions, the barrel, whose surface is at a radial distance of 125 cm from the z axis, and the two end caps that consist of concentric circles of crystals transverse to the z axis, at $z = -102$ and 192 cm [119]. With the end caps, the ECL has a polar angle acceptance of 12.4 – 155.1° . The small gaps between the barrel and endcaps, necessary for the support structure for the inner detectors, cause a 3% signal loss [119]. A diagram of this structure is presented in Figure 3.7.

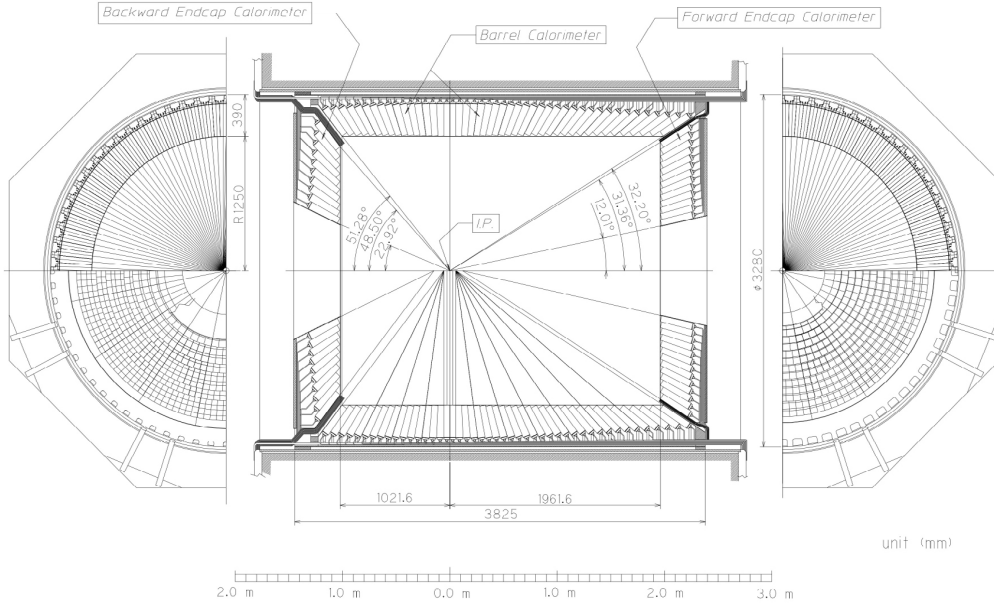


Figure 3.7: Schematic of the ECL crystal arrangement [135].

Particles that traverse the calorimeter lose energy through a variety of processes. All charged particles lose energy through ionisation according to the Bethe-Bloch equation [1].

³One radiation length gives the thickness of material in which the energy of a traversing electron is reduced by $1/e$.

For electrons, energy loss due to ionisation is dominated by energy losses through electromagnetic interactions with the atomic nuclei, emitting photons through bremsstrahlung. These photons as well as photons produced in the collision lose energy primarily through electron-positron pair production, with low-energy photons losing their remaining energy through Compton scattering and the photoelectric effect. Combined, bremsstrahlung and pair production lead electrons and photons to produce a cascade of photons and electrons/positrons when traversing a dense material. This is termed an electromagnetic shower, and is illustrated in Fig. 3.8. The longitudinal depth of these showers depends logarithmically on the energy of the initial particles. The radius of the shower perpendicular to the initial momentum of the incident particle is given by the Moliere radius, defined as covering 90% of the energy of the shower in a cylinder of radius $3.5R_M$. For CsI(Tl) $R_M = 3.53$ cm [1], a shower is therefore expected to deposit energy in multiple crystals.

In addition to ionisation, hadrons can undergo strong interactions with the atomic nucleons of the crystal, or any dense material in front of the calorimeter. Such interactions are complex, with a large number of possible processes contributing. Depending on the initial energy of the particle, additional charged hadrons, π, K, p, d , or neutral hadrons, π^0, K_L^0, K_S^0, n , can be produced, starting a hadronic shower. This is shown in Fig. 3.8. The charged particles produced will undergo the same interactions as described so far. Neutral pions will decay to a photon pair which will initiate an electromagnetic shower. Short-lived K_S^0 particles will decay after traversing a short distance, producing further charged or neutral pions. Long-lived neutral particles, K_L^0, n , can travel far without leaving an ionisation trail. If the direction of travel is approximately transverse to the surface of the ECL a secondary hadronic shower can be created, which is, during event reconstruction, not associated with the primary energy deposition. This is called a hadronic split-off cluster. As the ECL has been optimised for electromagnetic showers, it is only ~ 0.8 nuclear interaction lengths, the average length before a hadron undergoes a hadronic shower, in depth. Due to this and the energy carried away by n, ν and K_L^0 particles produced in the shower, the energy recorded in a hadronic shower is expected to be less than the energy of the incident particle. Measurements of the energy in the calorimeter, compared with the momentum of a track as measured in the tracking systems, thus provide an avenue to identify electrons.

The showers produced by an incident particle spread over several crystals; the measured energies must therefore be combined. As the showers of several incident particles can overlap, connected regions are identified and algorithmically split into clusters around local maxima. Single crystals can fractionally contribute to several clusters. Up to 25 crystals can be added to a cluster in the Belle clustering algorithm, corresponding to a 5×5 grid. The Belle II clustering algorithm allows for up to 21 crystals, excluding the crystals at the corner of the grid [138]. A matching algorithm checks if a track can be extrapolated to the cluster, in which case the energy measured in the cluster is associated to the track. If no track is extrapolated to the cluster, the cluster is assumed to originate from a neutral particle, assigned a photon or neutron hypothesis, and used alongside tracks

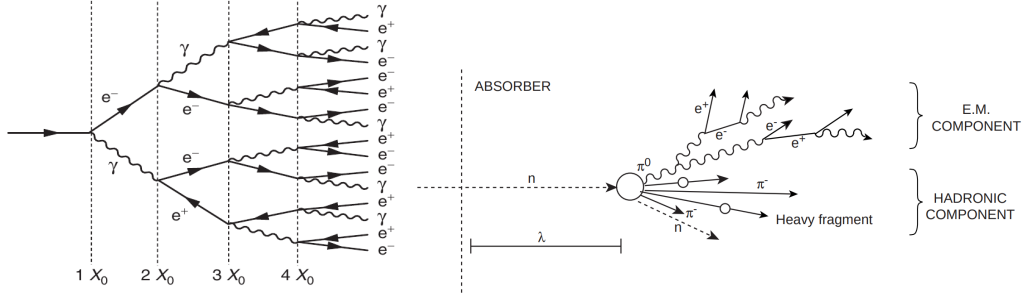


Figure 3.8: Schematic of an electromagnetic shower (*left*) [136] and a hadronic shower (*right*) [137].

as the two fundamental objects for physics analyses.

The ECL achieves relative energy resolutions between 4% for 100 MeV photons and 1.6% for 5 GeV photons [135, 139]. As neutral particles (γ , n , and K_L^0) have no ionising effect, they do not interact with the tracking detectors; thus the centre of the cluster position relative to the IP, combined with the measured energy, also provides the three-momentum information for neutral particles.

3.2.3 K_L^0 and muon detector

The outermost subsystem of Belle and Belle II is the K_L^0 -muon system (KLM). It is designed to measure and tag the two stable particle species most likely to pass through the inner detectors, K_L^0 , which only rarely interact with the ECL, and muons, which as minimum ionising particles punch through the ECL and thus can be tagged by their presence in the KLM.

The operating principle of the KLM is simple, provide a large amount of dense material in which the particles can interact and shower, interspersed with detectors. For the KLM, the dense material is a set of 4.7 cm thick iron plates⁴, equivalent to 3.9 interaction lengths, compared to the 0.8 interaction lengths of the ECL [135]. Like the ECL, the KLM is split into a barrel and end-cap structure. In the barrel there are 14 iron layers and 15 detector layers, in the end caps there are 14 layers of detector and iron plates [117, 119, 140]. In the Belle KLM the iron plates are interspersed with resistive plate chambers. In Belle II the plate chambers in the end caps were replaced with scintillators to better cope with the higher occupancy caused by the increased beam backgrounds due the higher luminosity at SuperKEKB [117, 140]. The KLM has a polar angle acceptance between 20–150°.

3.2.4 Particle identification

The Identification of charged particle species is a key requirement of Belle and Belle II, needed to reconstruct signal events and reduce the combinatorial background from misid-

⁴which serve also as the return yoke of the solenoid.

entifying charged particles. Within the analyses presented in this thesis, the identification of high-momentum electrons and muons, to identify semileptonic events, plays a key role.

Each subdetector discussed so far, except the PXD in Belle II, contributes to particle identification: electrons are primarily identified by the ECL, muons the KLM, and hadron species (pions and kaons) can be identified by measurements of the energy losses (dE/dx) in the SVD and CDC. However, at high momenta, particles can no longer be discriminated by dE/dx measurements; therefore, two additional subdetectors optimised for hadron identification across a broad momentum range are placed between the CDC and the ECL. In Belle these are the time of flight (TOF) detector in the barrel region and aerogel Cherenkov counter (ACC) in the barrel and forwards region, in Belle II they are the time of propagation (TOP) detector in the barrel region, and the aerogel ring imaging Cherenkov counter (ARICH) in the forward region. As an example, Fig. 3.9 schematically shows the interactions of different particle species with the Belle II subdetectors.

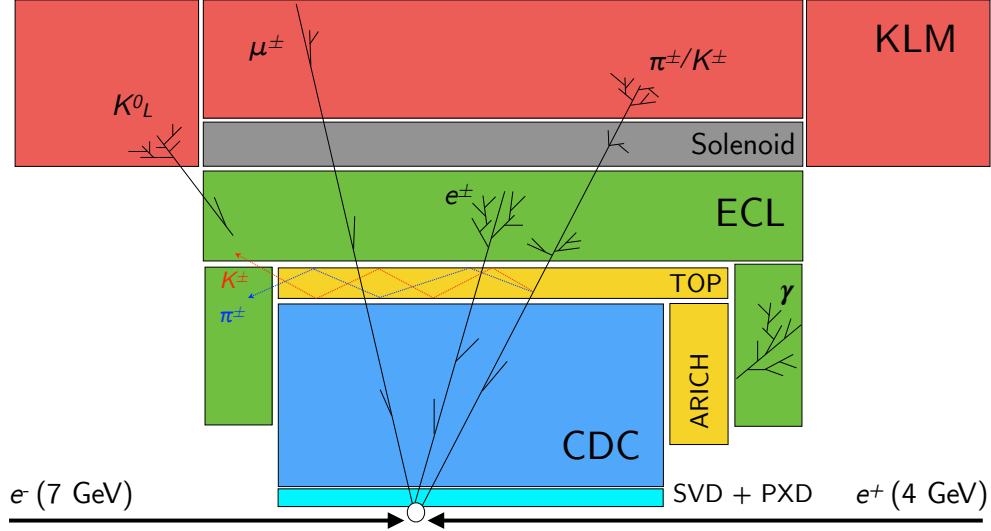


Figure 3.9: Schematic of the interactions of different particle types with the Belle II subdetectors [141]. The subdetectors are not to scale.

3.2.4.1 Time of flight detector

The Belle TOF consists of 64 modules, each consisting of two trapezoidal plastic scintillation bars separated from a thin trigger bar by 1.5 cm. Modules are placed 1.2 m from the IP, covering the face of the barrel region of the ECL with a polar angle acceptance of 34–120° [117]. When charged particles traverse the TOF modules, the material scintillates, emitting light, which is recorded by photomultiplier tubes at the ends of the modules. This signal is measured with a timing resolution of 100 ps [142]. Combining this timing information with event timing information, estimated by averaging over all TOF bars, and flight length as measured by the tracking system gives a measurement of the velocity of

the traversing particle. From this and the momentum as measured by the tracking system the mass of the particle can be estimated, efficiently discriminating between pions, kaons, and protons. The TOF is optimised for particle identification below momenta of 1.2 GeV [143].

3.2.4.2 Aerogel Cherenkov counter

The aerogel Cherenkov counter, located immediately beyond the CDC in Belle, extends the particle identification capability to the high-momentum regime, covering particles with momenta between 1 – 4.0 GeV [144]. The ACC operates on the Cherenkov effect [145], detecting light emitted by particles travelling through a medium at speeds that exceed the speed of light within that medium. Each module of the ACC consists of five aerogel tiles, acting as radiators, stacked in a $12 \times 12 \times 12$ cm aluminium box that is coupled to one or two photomultipliers. The ACC covers the barrel and the forward endcap, with an angular acceptance from 17–127°, the refractive index of the aerogel varies depending on the θ position of the module to optimise the performance. In the forward endcap, as there is no TOF, a higher refractive index is selected to aid lower momentum performance. Particle discrimination is achieved by comparing the number of recorded photoelectrons with expectations for each particle type at a given momentum [143, 144].

3.2.4.3 Time of propagation detector

Both the Belle II TOP and Belle II ARICH also rely on the Cherenkov effect. Instead of measuring the number of photoelectrons these detectors consider the opening angle, θ_C , of the emitted light, which depends inversely on the velocity of the particle. For two species with the same momenta, the angle therefore differs due to the masses of the particles. By combining measurements of the photon angle with the measured track momenta, particle species can be discriminated.

The TOP has 16 modules, each consisting of two fused quartz bars, acting as radiators, of dimension $125 \times 45 \times 2$ cm parallel to the axis of symmetry. At the forward end, the bar is mirrored. The emitted light internally reflects along the bar until it reaches the backward end, where an array of fast readout microchannel plate photomultiplier tubes is coupled to the detector to record the light [146, 147]. A schematic of the operating principle is shown in Figure 3.10.

Approximately 20 – 40 photons are detected for each charged particle [148], with a single-photon timing resolution of 35 ps [147]. This is sufficient to distinguish kaons from pions, the light from which has an arrival time difference of 200 ps for 2 GeV particles.

3.2.4.4 Aerogel ring imaging Cherenkov counter

The ARICH is a proximity-focusing ring-imaging Cherenkov detector, in a toroidal structure surrounding the beamline in the forward direction with an inner radius of 42 cm and an outer radius of 114 cm [150].

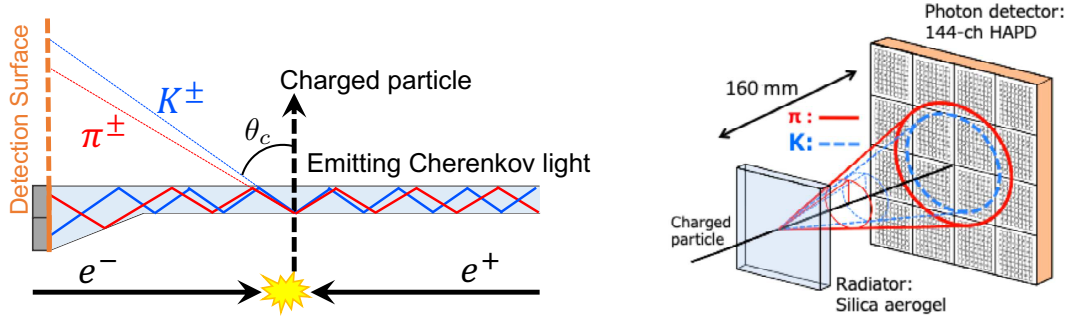


Figure 3.10: *Left:* Schematic of the TOP operation [148]. *Right:* Schematic of the ARICH operation [149]. Note that the pion and kaon labels refer to photons originating from pions or kaons passing through the material.

The detector consists of two 2 cm layers of aerogel tiles of different refractive indices acting as radiators, a 16 cm expansion space, and 8 cm deep set of hybrid avalanche photo detectors (HAPD). The operating principle of the ARICH is depicted in Figure 3.10. The light emitted in the radiators propagates to the photon detectors, where it is recorded as a circular pattern from which the Cherenkov angle can be inferred. The ARICH has demonstrated a 94% kaon identification efficiency at a pion-to-kaon misidentification rate of 11% [149].

3.2.4.5 Belle particle identification

Within the Belle software, information from several sub-detectors is combined to produce a set of discriminating variables with which, by requiring a minimum threshold, an analyst can select particles likely to belong to a particular species.

This is split into a general particle identification scheme, using the CDC, ACC and TOF information and specific variables aimed at electron and muon identification. The general scheme uses the expected distributions of light yield in the ACC, expected time of flight in the TOF and dE/dx measurements to build likelihoods for each of the six charged stable particle hypothesis. The likelihoods from the three detectors are multiplied together, forming a single likelihood for each particle hypothesis (i), $\mathcal{L}^{\text{CDC, TOF, ACC}}(\vec{\sigma}|i)$, where $\vec{\sigma}$ is the vector of observables entering the likelihood calculations. From these likelihoods a set of binary particle identification variables is created:

$$P_{i/j} = \frac{\mathcal{L}^{\text{CDC, TOF, ACC}}(\vec{\sigma}|i)}{\mathcal{L}^{\text{CDC, TOF, ACC}}(\vec{\sigma}|i) + \mathcal{L}^{\text{CDC, TOF, ACC}}(\vec{\sigma}|j)}, \quad (3.1)$$

where i and j are two particle hypothesis [143].

For electron identification, several additional variables originating from the calorimeter are considered. The full list of all considered variables is [151]:

- the ratio of the energy deposited in the ECL crystals to the track momentum measured in the CDC, $E/|\vec{p}|$;

- the quality of the matching of a track extrapolated to the ECL surface and a cluster in the ECL;
- the ratio of energy deposited in the central 9 crystals of an ECL shower to the energy deposited in the central 25 — an estimate of the shape of the shower;
- the photoelectron yield of the ACC, noting that electrons require only a few MeV of energy to produce a measurable light yield; and
- the ionisation losses measured in the CDC, dE/dx .

For each of these variables, likelihoods are calculated under an electron hypothesis and a ‘not electron’ hypothesis, consisting of a cocktail of hadrons as simulated for generic $B\bar{B}$ decays. The likelihoods are combined to form a binary likelihood ratio of electron vs. not electron, hereafter labelled P_e [151].

Muons are identified exclusively by information from the KLM⁵. Two variables are considered, the difference between the measured and expected range of the track in the planes of the KLM, and the goodness of fit of the transverse deviations of all hits associated with the track. The two variables are assumed to be uncorrelated and likelihoods for the muon, pion and kaon hypotheses are calculated using joint probability density functions constructed from single-particle simulations [152]. From the likelihoods, a muon identification variable, P_μ , is defined as the trinary likelihood ratio:

$$P_\mu = \frac{\mathcal{L}^{\text{KLM}}(\vec{\sigma}|\mu)}{\mathcal{L}^{\text{KLM}}(\vec{\sigma}|\mu) + \mathcal{L}^{\text{KLM}}(\vec{\sigma}|\pi) + \mathcal{L}^{\text{KLM}}(\vec{\sigma}|K)}. \quad (3.2)$$

3.2.4.6 Belle II particle identification

The default particle identification scheme at Belle II always considers all subdetectors except for the PXD. Each subdetector, if the particle interacted with it, contributes a likelihood for each of the six stable charged particle hypotheses. The likelihoods are computed either analytically or by comparison with a probability density function extracted from simulation. The observables that feed into the particle identification are:

- particle energy loss by ionisation, dE/dx , measured in the SVD and CDC;
- optical responses of the TOP and ARICH;
- the ratio of the energy deposited in the ECL crystals to the track momentum measured in the CDC and SVD, $E/|\vec{p}|$; and
- measurements of the penetration range in the KLM.

⁵Note that this implies only muons energetic enough to reach the KLM can be identified. This translates into an approximate requirement of $p > 700$ MeV.

The per subdetector likelihoods are combined into a global likelihood under the assumption that the observables under consideration are independent across subdetectors. The global likelihood is given by:

$$\mathcal{L}(\vec{o}|i) = \prod_{\text{det}=\text{dets}} \mathcal{L}^{\text{det}}(\vec{o}|i), \quad (3.3)$$

where $\text{dets} = \{\text{SVD}, \text{CDC}, \text{TOP}, \text{ARICH}, \text{ECL}, \text{KLM}\}$. Finally, the likelihood of each particle hypothesis is normalised by the sum of likelihoods:

$$P_i = \frac{\mathcal{L}(\vec{o}|i)}{\sum_j \mathcal{L}(\vec{o}|j)}. \quad (3.4)$$

To improve the lepton-hadron separation at low momentum, an alternative charged-particle identification scheme has been developed in Ref. [153]. Available non-ECL sub-detector likelihood information is combined with a more comprehensive set of ECL-based inputs using machine learning, yielding significant improvements at low-momentum lepton identification. This method is further discussed in Chapter 5.

3.2.5 Trigger

So far, we have discussed three production channels at B -factories that are of interest for physics analyses, $e^+e^- \rightarrow \Upsilon(4S), c\bar{c}$, and $\tau^+\tau^-$. These are, however, not the only processes that occur. Bhabha scattering, $e^+e^- \rightarrow e^+e^-(\gamma)$ has a production cross section of 300 nb, dominating these channels which have production cross sections $\mathcal{O}(1 \text{ nb})$. Several other low-multiplicity channels, listed in Table 3.1, also have significant production cross sections. In order to keep the data rate below the 500 Hz and 30 kHz that can be processed by the respective data acquisition systems of Belle and Belle II, a trigger system is installed that identifies and rapidly discards most of the uninteresting events. A small fraction is kept for calibration and performance studies.

Both the Belle and Belle II triggers operate in two stages. A low-level hardware trigger considers coarse information from the CDC, TOF or TOP, ECL and KLM [154–156] to rapidly veto simple-to-identify bhabha events. In Belle a software trigger then searches for at least one track originating from the interaction point using a fast tracking algorithm [117]. In Belle II the high-level software trigger runs the full Belle II reconstruction software [157, 158], with a range of software triggers optimised for a wide range of analyses, further reducing the event rate by a factor of three [159].

Trigger selections and performance can have a significant impact on analysis of low-multiplicity channels, such as dark-sector searches or τ analyses. For the analyses of semileptonic B decays discussed in this thesis, it suffices to note that the triggers are $\sim 100\%$ efficient.

Process	Cross Section [nb]
$\Upsilon(4S)$	1.11
$c\bar{c}$	1.30
$s\bar{s}$	0.38
$u\bar{u}$	1.61
$d\bar{d}$	0.40
$\tau^+\tau^-$	0.919
$\mu^+\mu^-(\gamma)$	1.148
$e^+e^-(\gamma)$	300
$\gamma\gamma$	4.99
$e^+e^-\mu^+\mu^-$	18.9
$e^+e^-e^+e^-$	39.7

Table 3.1: Production cross sections at electron-positron collisions at $\sqrt{s} = M_{\Upsilon(4S)}$ [84, 160–164]. The line separates hadronic and non-hadronic production.

3.3 Software

Raw data from the Belle subdetectors, and simulated samples, are processed with the Belle analysis software (**basf**) [165] to reconstruct high-level physics objects suitable for data analysis, such as tracks and clusters in the ECL. These high-level objects are stored in a space-saving reduced data format **mDST**⁶, made available to analysts.

Belle II data and simulated samples are processed using the Belle II analysis software (**basf2**) [157, 158]. Additionally, **basf2** contains a conversion software package to allow analysis of Belle **mDST** files with the powerful tools available in the **basf2** analysis package [166]. The analyses of Belle data presented in this thesis makes use of this conversion layer.

For the discussion in Chapter 5 it is useful to briefly touch on the structure of **basf2**. The framework has been developed to process large data sets in a highly flexible manner. Data processing algorithms are encapsulated within modules that can interact with a central data store. Multiple modules can act on the data in turn in a linear order. The chain of modules applied within one processing acts on each collision event processed in turn⁷ [157]. Conditional branching is supported on the basis of the properties of a single event or candidate.

3.4 Simulation

Simulated samples play an important role in modern particle physics analyses; both to prepare analyses, with selection criteria developed and optimised on simulated signal and

⁶mini Data Storage Table

⁷If a module is marked safe for parallel processing it can act on multiple events in parallel.

background samples, and often to extract the signal yields through the use of simulation-based probability density functions in fitting. For the simulation of physics processes we turn to Monte Carlo methods. For the decay of $\Upsilon(4S)$ -mesons events are generated via the **EvtGen** [167] package. Decays are randomly selected from a list of known channels with specified probabilities. Each decay is associated with a model that governs the kinematic distributions of the produced particles. Produced short-lived particles are recursively decayed until only particles with long enough life times to interact with the detector material are produced.

Continuum events in Belle, certain particles for which the decay channels are not known, or fractions of the total decay rate of a particle that are unknown are simulated via **PYTHIA** [168, 169] which hadronises the particles. In the Belle II simulation, $q\bar{q}$ events are generated with the two-fermion generator **KKMC** [160, 170], before being hadronised by **PYTHIA**. Final state radiation, the emission of photon by charged final-state particles, is simulated with the **PHOTOS** [171] package for both experiments. The interactions of particles with the Belle and Belle II detectors are simulated with **GEANT3** [172] and **GEANT4** [173–175], respectively.

Overview of methods and tools

We briefly introduce some statistical tools and methods used within this thesis. Machine learning through boosted decision trees and neural networks, and parameter estimation through fitting.

4.1 Machine learning

In recent years, machine learning has come to play a significant role in high-energy particle physics. Being employed to search for often rare decays hidden amongst overwhelming background by identifying and exploiting complex multidimensional patterns across many variables. Where each variable individually may only be able to weakly distinguish the signal. In addition to high-level classification tasks, machine learning is employed at all stages of data processing and event reconstruction, from monitoring and tuning beam conditions, filtering electronic noise in detector signals, to clustering and tracking.

Machine learning is an extremely broad field, with a large number of algorithms aimed at a wide range of tasks. The following discussion aims to serve as a brief introduction to supervised classification algorithms, as used within the work presented within this thesis.

4.1.1 Boosted decision trees

Perhaps the simplest approach to separate signal and background events using a set of N variables is to consider each variable in turn, selecting events that fall below or above a threshold, such that after considering all variables a purified sample has been selected. The values of the thresholds are selected to maximise the separation between signal and background using a sample in which it is known which events are signal and which are background. In particle physics, simulated samples usually fill this role. Such a linear series of selections, however, may reject a signal rich region of phase space in an early selection which could be identified by considering an alternate variable. Such regions can be retained by instead constructing decision trees. The construction of such a tree, also referred to as training, follows the same approach but at each selection further subdivides both the signal-like and background-like samples. In this manner a branching tree like structure is formed, with branch-divisions terminating at either fixed depths, or set minimum number of training events within a branch. When applying the classifier to events, an event is assigned a score based on the purity of the terminating node in training, mapping the

high-dimensional input space to a one-dimensional prediction score which can be used to select signal-like events.

In practice, large single decision trees have been found to be highly susceptible to over-training, learning the statistical noise within the training sample rather than the underlying structure of the distributions. To combat this problem, ensemble methods were developed that combine many weak classifiers. One such method is boosted decision trees (BDT), which combines many shallow trees. Each tree commonly no deeper than four or five decisions. The term boosting refers to how trees are added to the total classifier. Trees added after the first are trained not just to best separate events, but to minimise the error of preceding trees. This can be done by increasing the weight of incorrectly classified events [176, 177], or by minimising a loss function. The loss function, $L(y, f(\vec{x}))$, compares the model output, $f(\vec{x})$, where $\vec{x}$ is the vector of input features, to the desired output, y . A common loss function for binary classification is the cross-entropy. Given by:

$$\text{BCE}(y, f(\vec{x})) = y \ln(f(\vec{x})) + (1 - y) \ln(1 - f(\vec{x})). \quad (4.1)$$

The second approach is known as gradient boosted decision trees, and is the algorithm used when we refer to boosted decision trees within this thesis. We use two implementations, **FastBDT** [178] and **XGBoost** [179]. The algorithm works as follows [180]. The model is initialised to a constant response, $f_0(x) = c$, then, when adding the m th tree, the pseudo residuals are calculated from the response of all preceeding trees as

$$-g_m(\vec{x}_i) = - \left[\frac{\partial L(y_i, f(\vec{x}_i))}{\partial f(\vec{x})} \right]_{f(\vec{x})=f_{m-1}(\vec{x})}, \quad (4.2)$$

The pseudo residuals are then used to train the tree in place of the true labels, y . The tree is then added to the ensemble with a weight γ such that $\sum_i L(y_i, f_{m-1}(\vec{x}_i) + \gamma h_m(\vec{x}_i))$ is minimised, where h_m is the new tree. To slow the rate of the learning process, making it more robust against overtraining, the tree is added to the ensemble with a weight, η , known as the shrinkage or learning rate, ie.

$$f_m(\vec{x}) = f_{m-1}(\vec{x}) + \eta \gamma_m h_m(\vec{x}). \quad (4.3)$$

This continues until the desired number of trees have been added. Additionally, each tree can be trained with only a subsample of example events, introducing a stochastic nature to the training, which further prevents overtraining. To further prevent overtraining, the XGBoost algorithm adds an additional term to the loss function, penalising complex trees which have a large relative difference in weights between the leaves of the tree. It also requires a minimum improvement in loss before splitting a node.

4.1.2 Neural networks

The majority of machine-learning applications in this thesis use boosted decision trees; however, we also use a feed-forward neural network for continuum suppression in the search for weak annihilation. Thus, we briefly outline the operating principle.

Neural networks are based on perceptrons, inspired by the transmission of electrical impulses in neurons [181]. Perceptrons combine information from a set of input neurons, perform a calculation, and pass the result to a further layer of neurons. The output of the j^{th} neuron is given by

$$x_j = f\left(\sum_i x_i w_{ij} + b_j\right), \quad (4.4)$$

where the sum runs over all the input neurons, w_{ij} is a set of weights which modify the input values, b_j is a bias, and $f(x)$ a non-linear activation function. The weights and bias are trainable parameters. A neural network contains layers of such perceptrons, with the initial layer receiving the values of the feature variables, and the output of this layer being passed forward through the network, layer by layer. For binary classification problems, the final layer consists of a single neuron using a sigmoid activation function to map the output to a probability. Networks are trained by propagating batches of labelled example events through the network and, after each batch is processed, adjusting the learnable parameters of the model via backward propagation of the error in the network output to minimise a loss function [182]. The complete set of training events can be passed through the network multiple times, each iteration being known as an epoch.

Several approaches exist to combat overtraining in neural networks. We adopt two of these: dropout and batch normalisation layers. Dropout randomly removes a fraction of connections between neurons during training, increasing the robustness of the model. Batch normalisation normalises the value of the inputs and passed activation weights to reduce the impact of outliers.

4.1.3 Hyperparameter optimisation

Each machine learning algorithm learns the value of its parameters, choice of variable and threshold at a tree node or, weights and biases of a neuron, during training. The performance of the model, however, depends not only on these parameters, but also on parameters that determine the structure of the model and details of the learning processes. Such parameters are known as hyperparameters and are external to the training of the model. Many approaches exist to optimise the value of these hyperparameters. Most split the training sample in two, using one subsample to train variations of the model, and the other independent subsample to evaluate the performance of each variation, such that the best can be selected. The simplest technique is an N dimensional grid search, where N is the number of hyperparameters being considered. Models are trained using hyperparameter values from a regular N -dimensional grid and the best performing is selected. If the number of hyperparameter dimensions grows large, a grid search becomes infeasible due to the computational expense of training each model. Thus in addition to grid search, in this thesis we use tree-structured Parzen estimators (TPE) [183, 184], a form of Bayesian optimisation, as implemented in Refs. [185, 186] for hyperparameter optimisation. Bayesian optimisation assumes a flat prior over hyperparameter space, randomly samples points to be evaluated, and after assessing the performance of models trained

with these hyperparameter sets updates the prior, and samples new points. This is done in a way that balances exploration of unexplored phase space and optimisation based on existing knowledge. TPE methods extend Bayesian optimisation to support conditional dimensions, such as may be encountered when adding additional hidden layers to a neural network, and have empirically demonstrated the ability to find well-performing sets of hyperparameters within limited iterations [183].

4.1.4 Variable importances

While machine learning algorithms are powerful tools for separating signal from background, it is difficult to explain how they make their predictions once they grow beyond a few trees or a few hidden nodes. As some features are more robustly modelled than others, it is important to know which variables dominate the classification. We estimate this in two ways. For models trained with FastBDT, we consider the internal feature importance, calculated as the average gain at each node splitting in which a variable is used, weighted by the number of training events in the node and the importance of the tree to the total ensemble. For XGBoost models and neural networks we consider the SHAP (SHapely Additive exPlanation) [187] values. These are based on a game theory concept to fairly measure the contribution of a player to the outcome of a game by measuring their marginal contribution over all possible subsets of players [188]. Applied to machine learning models, the importance of a particular feature to a classifier can be estimated by averaging the difference in predictions of a model with and without the feature variable. In the case of boosted decision trees, this can be computed exactly [189]. For neural networks, an approximation is made by sampling values of the variable to be removed. Locally, the result of the model can thus be given as a linear function

$$f(\vec{x}_i) = \phi_0 + \sum_j \phi_j z'_{ij}, \quad (4.5)$$

where i runs over the feature variables, z'_{ij} are a simplified binary transformation of the input variables, and ϕ_j is the SHAP value associated with variable j . The output of the model is thus given by the sum of SHAP values. The importance of the i^{th} feature variable to the classification of N events can then be estimated as the mean absolute contribution, $\frac{1}{N} \sum_k |\phi_{ik}|$.

4.2 Full event interpretation

Two of the strengths of B -factory experiments are the hermeticity of the detectors and the clean initial states, consisting of a pair of B mesons. These strengths are exploited by making use of the companion B , i.e. the B that has not been identified as decaying in the channel currently under study, to provide constraints on the signal B . This can be done loosely, in what is referred to as an *untagged* or *inclusively tagged* study. In this approach, the sum of kinematic information of any particles present in the event but not used to

reconstruct the signal B is checked for general consistency with the decay of a B -meson. Alternatively, one can explicitly reconstruct the companion B . Explicit reconstruction provides precise information about the companion B , allowing strong constraints to be placed on the signal B , at the cost of efficiency due to the reconstruction efficiency of the companion B . Usually a distinction is made depending on the reconstructed mode of the companion B . Semileptonic tags, due to the missing neutrino have only partial kinematic information available, but have reconstruction efficiencies of $\mathcal{O}(1)\%$ due to the abundance of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events. Hadronic tags, reconstructing only decay chains that are fully hadronic, and fully visible in the detector, give access to the full information of the companion B , including the flavour, charge, and momentum, but at much lower efficiencies, $\mathcal{O}(0.1)\%$.

Within the Belle II software, the tag reconstruction for both semileptonic and hadronic modes is performed by the Full Event Interpretation (FEI) [190] algorithm. For both analyses performed in this thesis we use hadronic tagging of FEI, which allows us to identify the missing momentum as the momentum of the signal-side neutrino. FEI is a hierarchical six-stage algorithm. It begins by selecting tracks and ECL clusters from which it forms final-state particles, which are combined to form intermediate particles ($\pi^0, K_S^0, J/\psi, \Lambda^0$). The intermediate particles are in turn combined with the final-state particles to heavier intermediate particles ($D, D^*, \Sigma, \Lambda_c$), which are finally combined with other particles to form B meson candidates. This is shown schematically in Fig. 4.1.

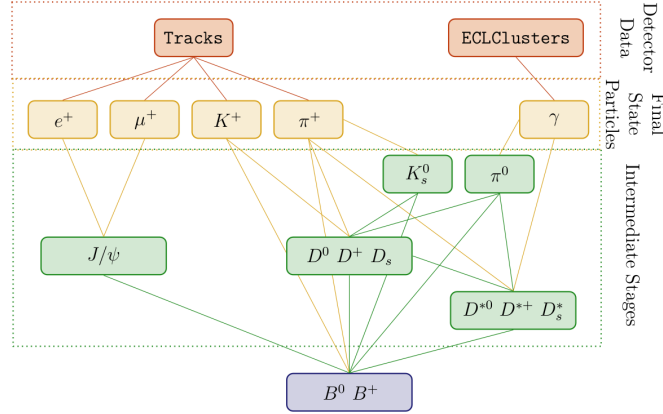


Figure 4.1: Schematic of the FEI algorithm [84, 190]. Note that baryonic modes are not included in the diagram. For these modes we also identify protons and reconstruct intermediate Λ^0, Σ , and Λ_c baryons.

At each stage of the reconstruction, a loose selection is performed to remove poor candidates and boosted decision tree classifiers are used to assess the likelihood that the particle has been correctly identified or reconstructed, using a variety of information where appropriate: particle identification likelihoods, distance of the point of closest approach to the interaction point, four-momenta, vertex information, angles between children, in-

variant mass compared to the nominal mass of the particle, the output of the classifiers for each child particle, and more. The final output score for a reconstructed B meson, $\mathcal{P}_{\text{FEI}}$, thus contains information from each stage of the reconstruction and can be used as a measure of the quality of the reconstructed B -meson candidate.

FEI reconstructs $\mathcal{O}(10000)$ total decay chains [190], culminating in 36 charged B modes, and 32 neutral B modes which are presented in Tab. 4.1. The version of FEI used for the measurement of the ratio of partial branching fractions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ to $\bar{B} \rightarrow X_c \ell \bar{\nu}$ presented in Part III. A did not include the reconstruction of baryonic modes (all modes below the line in Tab. 4.1), thus only 29 charged and 26 neutral B modes were reconstructed.

B^+		B^0	
$B^+ \rightarrow \bar{D}^0 \pi^+$	$B^+ \rightarrow \bar{D}^{*0} \pi^+$	$B^0 \rightarrow D^- \pi^+$	$B^0 \rightarrow D^{*-} \pi^+$
$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0$	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0$	$B^0 \rightarrow D^- \pi^+ \pi^0$	$B^0 \rightarrow D^{*-} \pi^+ \pi^0$
$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0 \pi^0$	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0 \pi^0$	$B^0 \rightarrow D^- \pi^+ \pi^0 \pi^0$	$B^0 \rightarrow D^{*-} \pi^+ \pi^0 \pi^0$
$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^-$	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^-$	$B^0 \rightarrow D^- \pi^+ \pi^+ \pi^-$	$B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^-$
$B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^- \pi^0$	$B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^- \pi^0$	$B^0 \rightarrow D^- \pi^+ \pi^+ \pi^- \pi^0$	$B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^- \pi^0$
$B^+ \rightarrow \bar{D}^0 D^+$	$B^+ \rightarrow D_s^{*+} \bar{D}^0$	$B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$	$B^0 \rightarrow D_s^{*+} D^-$
$B^+ \rightarrow \bar{D}^0 D^+ K_S^0$	$B^+ \rightarrow D_s^{*+} \bar{D}^{*0}$	$B^0 \rightarrow D^- D^0 K^+$	$B^0 \rightarrow D_s^{*+} D^{*-}$
$B^+ \rightarrow \bar{D}^{*0} D^+ K_S^0$	$B^+ \rightarrow \bar{D}^0 K^+$	$B^0 \rightarrow D^- D^{*0} K^+$	$B^0 \rightarrow D_s^{*+} D^{*-}$
$B^+ \rightarrow \bar{D}^0 D^{*+} K_S^0$	$B^+ \rightarrow D^- \pi^+ \pi^+$	$B^0 \rightarrow D^{*-} D^0 K^+$	$B^0 \rightarrow J/\psi K_S^0$
$B^+ \rightarrow \bar{D}^{*0} D^{*+} K_S^0$	$B^+ \rightarrow D^- \pi^+ \pi^+ \pi^0$	$B^0 \rightarrow D^{*-} D^{*0} K^+$	$B^0 \rightarrow J/\psi K^+ \pi^-$
$B^+ \rightarrow \bar{D}^0 D^0 K^+$	$B^+ \rightarrow J/\psi K^+$	$B^0 \rightarrow D^- D^+ K_S^0$	$B^0 \rightarrow J/\psi K_S^0 \pi^+ \pi^-$
$B^+ \rightarrow \bar{D}^{*0} D^0 K^+$	$B^+ \rightarrow J/\psi K^+ \pi^+ \pi^-$	$B^0 \rightarrow D^{*-} D^+ K_S^0$	
$B^+ \rightarrow \bar{D}^0 D^{*0} K^+$	$B^+ \rightarrow J/\psi K^+ \pi^0$	$B^0 \rightarrow D^- D^{*+} K_S^0$	
$B^+ \rightarrow \bar{D}^{*0} D^{*0} K^+$	$B^+ \rightarrow J/\psi K_S^0 \pi^+$	$B^0 \rightarrow D^{*-} D^{*+} K_S^0$	
$B^+ \rightarrow D_s^{*+} \bar{D}^0$		$B^0 \rightarrow D_s^{*+} D^-$	
$B^+ \rightarrow \Lambda_c^- p \pi^+ \pi^0$	$B^+ \rightarrow D^+ p \bar{p} \pi^+ \pi^-$	$B^0 \rightarrow \Lambda_c^- p \pi^+ \pi^-$	$B^0 \rightarrow \bar{D}^0 p \bar{p} \pi^+ \pi^-$
$B^+ \rightarrow \Lambda_c^- p \pi^+ \pi^- \pi^+$	$B^+ \rightarrow D^{*+} p \bar{p} \pi^+ \pi^-$	$B^0 \rightarrow \bar{D}^0 p \bar{p}$	$B^0 \rightarrow \bar{D}^{*0} p \bar{p} \pi^+ \pi^-$
$B^+ \rightarrow \bar{D}^0 p \bar{p} \pi^+$	$B^+ \rightarrow \Lambda_c^- p \pi^+$	$B^0 \rightarrow D^- p \bar{p} \pi^+$	
$B^+ \rightarrow \bar{D}^{*0} p \bar{p} \pi^+$		$B^0 \rightarrow D^{*-} p \bar{p} \pi^+$	

Table 4.1: FEI decay modes. The line divides purely mesonic decay modes from baryonic decay modes.

4.3 Parameter estimation

The main goal of experimental analyses is to extract parameters of interest, for example, the branching fraction of $\bar{B} \rightarrow X_u \ell \bar{\nu}$, from the distribution of observed data. This is achieved through maximum likelihood estimation, fitting a theoretical model that depends on the parameters of interest to the observed data.

Given a probability density function (PDF), $f(\vec{x}|\vec{\omega})$, of observables $\vec{x}$ and parameters $\vec{\omega}$ we can evaluate the likelihood of the PDF with respect to an observed data point $\vec{x}_i$

$$\mathcal{L}(\vec{\omega}) = f(\vec{x}_i|\vec{\omega}). \quad (4.6)$$

Considering instead many independent events we can define the likelihood to be

$$\mathcal{L}(\vec{\omega}) = \prod_i f(\vec{x}_i|\vec{\omega}). \quad (4.7)$$

If the number of observed events, n , depends on the parameters of the model, it is possible to extend the concept to include the normalisation in the fit, known as an extended maximum likelihood [191]. If n follows a Poisson distribution with mean ν the likelihood becomes

$$\mathcal{L}(\vec{\omega}) = \frac{\nu^n}{n!} \exp(-\nu) \prod_i f(\vec{x}_i|\vec{\omega}). \quad (4.8)$$

The best parameters, $\hat{\vec{\omega}}$, are then given by maximising the likelihood, or more commonly by minimising twice the negative log-likelihood via a gradient descent algorithm.

If the number of events under consideration is large, as is the case in the analyses presented within this thesis, it is convenient to instead fit the binned distributions. If the total number of events is fixed, then the data can be described as a multinomial distribution, where the probability that an event will be found in a particular bin is a function of the model parameters $\vec{\omega}$. If the number of events is a free parameter, the data is described as a product of Poisson distributions where the expected number of events depend on $\vec{\omega}$.

The uncertainties of the fitted parameters due to statistical fluctuations can be estimated from the covariance of the fitted parameters or by examining the change in the profile likelihood (which we will define below) as the parameter is varied about its fitted value.

In general, measurements are subject not only to statistical uncertainties but also to systematic effects. The treatment of systematic uncertainties varies: for Belle and BaBar analyses, the impact of systematic effects was usually estimated by varying each systematic effect independently, repeating the analysis procedure and studying the changes to the fitted parameters. This is the approach taken in the analysis of ratio of partial branching fractions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ to $\bar{B} \rightarrow X_c \ell \bar{\nu}$ at Belle presented within this thesis. Alternatively, systematic effects can be incorporated directly into the fit model through the addition of nuisance parameters. The likelihood is then given by

$$\mathcal{L}(\vec{\omega}, \vec{\theta}) = f(\vec{x}|\vec{\omega}, \vec{\theta}) \cdot c(\vec{y}|\vec{\theta}), \quad (4.9)$$

where $\vec{\theta}$ are the nuisance parameters, $\vec{y}$ represent estimates of the nuisance parameters usually extracted from control measurements, and $c(\vec{y}|\vec{\theta})$ is a constraint function, usually taken to be Gaussian, constraining the nuisance parameters with our existing knowledge. We have also assumed that $\vec{x}$ and $\vec{y}$ are statistically independent.

In many cases it can be useful to consider a likelihood that is purely a function of the parameter of interest, or a set of parameters of interest. This can be done by profiling the nuisance parameters. The profile likelihood is given by

$$\mathcal{L}(\vec{\omega}, \hat{\vec{\theta}}) = \arg \max_{\vec{\theta}} \mathcal{L}(\vec{\omega}, \vec{\theta}), \quad (4.10)$$

where the double hat indicated these parameters have been profiled over.

4.3.1 Hypothesis testing and limit setting

Many analyses in particle physics concern themselves with searching for new particles that have not previously been measured. Such searches naturally involve hypothesis testing, a method of statistical inference used to determine whether to accept or reject a hypothesis, such as that a new particle with particular properties exists, in the face of a null hypothesis, that the particle does not exist.

Although weak annihilation is a standard model process, the specifics of the process are unknown, and it remains unmeasured. We therefore treat the search for weak annihilation contributions to inclusive charmless semileptonic decays presented in this thesis as a hypothesis test, testing the null hypothesis

$$H_0 : \text{weak annihilation does not exist}, \quad (4.11)$$

against the alternative hypothesis

$$H_1 : \text{weak annihilation exists}. \quad (4.12)$$

In general, one can assign a p value which reflects the agreement of the observed data and a hypothesis via a frequentist significance test. The p value is given by

$$p = \int_{t_{\text{obs}}}^{\infty} f(t|H) dt, \quad (4.13)$$

where t is a test statistic encoding agreement of the data with the hypothesis, and here defined such that a large value indicates poor agreement. If the p value is sufficiently low, the hypothesis can be excluded. It is common to set a threshold of 2.87×10^{-7} , equivalent to a 5σ significance, in order to consider the background-only hypothesis rejected.

More specifically, in the weak annihilation search, the p value is given by

$$p_{\mu}(\vec{\theta}) = \int_{t_{\text{obs}}}^{\infty} f(t|\mu, \vec{\theta}) dt, \quad (4.14)$$

where $\vec{\theta}$ are nuisance parameters as before, μ is the strength of the weak annihilation signal. A signal strength of zero corresponds to the background only hypothesis. In order to reject the hypothesis of background plus signal at a particular signal strength, μ , the p value must be smaller than the threshold for all values of the nuisance parameters. To

include this directly in the calculation of the p value, we can consider the profile likelihood ratio as the test statistic,

$$\lambda_p(\mu) = \frac{\mathcal{L}(\mu, \hat{\vec{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})}. \quad (4.15)$$

In the limit of large data samples, $-2 \ln \lambda_p(\mu)$ behaves as a χ^2 distribution depending only on μ [192]. It is this form that is usually used as the test statistic.

If no significant signal is observed, either because the signal process does not exist, or the analysis is not sufficiently sensitive, it is possible to set an upper limit on the signal strength parameter through the CL_s method [193]. The CL_s value is defined as

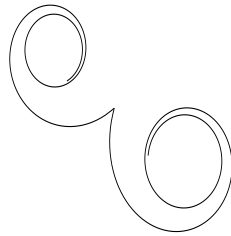
$$\text{CL}_s = \frac{p_{s+b}}{1 - p_b}, \quad (4.16)$$

where the correction by $1/(1 - p_b)$ ensures that significant downward fluctuations in background yields do not cause the test to exclude signal strengths to which the analysis is not sensitive.

As we consider weak annihilation to only contribute positively to the measured event rate, we use the test statistic $\tilde{q}_\mu$, specifically its asymptotic approximation [194], defined by:

$$\tilde{q}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0, \\ \frac{(\mu - \hat{\mu})^2}{\sigma^2} & 0 \leq \hat{\mu} < \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \quad (4.17)$$

where the third condition ensures that, when setting an upper limit, data with $\hat{\mu} > \mu$ is not considered as less compatible with μ than the obtained data.



Part II

Improving charged particle identification in the Belle II calorimeter

Machine learning based charged particle identification in the ECL

We review the need for improved lepton identification at low momentum within Belle II, the machine learning based scheme currently available, and motivate switching to machine learning based methods directly in the core reconstruction software. We then discuss the implementation of a set of boosted decision tree classifiers for particle identification in the calorimeter, and their performance on simulated samples.

5.1 Introduction

A key topic in the Belle II physics programme [84] is the study of semileptonic B decays to test lepton flavour universality, focusing on precise measurement of $R(D)$, $R(D^*)$ and $R(X)$; the ratios of semitauonic to light lepton semileptonic B decays with an accompanying D or D^* meson or inclusively. Combined, $R(D)$ and $R(D^*)$ currently exhibit a 3.3σ tension with the SM expectation [66]. These decays are largely studied using leptonic tauon decays, $\tau \rightarrow \ell \nu_\ell \nu_\tau$, which, due to the high multiplicity of the decay, produce soft leptons. To improve the precision of such measurements Belle II must therefore be able to identify low-momentum leptons with high accuracy. Excellent low-momentum performance will benefit many measurements, including the study of $|V_{ub}|$, and $|V_{cb}|$, allowing such measurements as presented within this thesis to push into lower lepton momentum regions and improving vetoes of secondary leptons.

In chapter 3 we discussed the global likelihood scheme for charged particle identification at Belle II. The set of observables entering this scheme provides excellent separation of leptons from hadrons at lab-frame particle momenta above 1 GeV. At low momentum, however, the separation power diminishes. Muons with transverse momenta less than 0.6 GeV do not reach the KLM, the dominant subdetector for muon-pion separation. Low-momentum electrons lose a significant fraction of their energy to bremsstrahlung before reaching the ECL, degrading the usefulness of $E/|\vec{p}|$ as a separation variable, the only information entering from the ECL.

To improve lepton-hadron separation at low momentum, an alternative charged-particle identification scheme has been developed in Ref. [153]. Using machine learning through a set of multiclass boosted decision tree classifiers (BDTs), the available non-ECL subdetector likelihood information is combined with a more comprehensive set of ECL-based

inputs. In addition to $E/|\vec{p}|$, variables that describe the structure of energy deposition in the crystals, which differs between electromagnetically showering and hadronically showering particles, are included.

Through the inclusion of the additional ECL-based variables, the BDT-based particle identification has been able to demonstrate significant improvements, reporting a decrease in pion-to-electron and pion-to-muon misidentification rates at low momentum of factor ten and two, respectively, in single-particle simulated samples [153]. This has also been partially attributed to the ability of the BDT to better incorporate non-trivial correlations between input variables than the naive likelihood model described before, as well as being less sensitive to miscalibration of subdetector likelihoods.

The model described above has several limitations. As discussed in Sec. 3.3, to prevent large computational overhead, the majority of event reconstruction is done centrally by the Belle II data production group. Only a reduced data format (**mDST**) is available to analysts, which contains only a limited set of information, mostly high-level variables, such as subdetector likelihoods and fitted track four-momenta, and selected lower-level information. The BDTs developed in Ref. [153] are limited to the information included in this reduced data format, much of which has been included for its ability to separate photon clusters from neutral hadrons, rather than for the purpose of charged-particle identification. Furthermore, each analyst must run the BDTs during the course of their analysis, introducing computationally expensive overhead. In the following, this set of classifiers will be referred to as the *analysis-BDT* or A.BDT and its predictions labelled $\mathcal{P}_{x,\text{A.BDT}}^{(\text{ECL})}$, where x is one of $\{e, \mu, \pi, K\}$, the four stable charged particle hypotheses that the analysis-BDTs are trained to distinguish, and the superscript ECL indicates the set of BDTs using information only from the ECL. The lack of superscript indicates the inclusion of the likelihoods of the other subdetectors in the feature variables of the BDTs.

To further improve the low-momentum lepton identification performance, it is therefore necessary to integrate such machine learning based methods directly into the central event reconstruction, gaining access to additional low-level variables describing the energy deposition in the ECL. This further allows for extracting likelihoods from the BDTs and integrating the response directly into the global likelihood scheme.

5.2 The ECLChargedPIDMVA module

To integrate machine learning-based charged particle identification in the ECL into the reconstruction software, the **ECLChargedPIDMVA** module has been developed as part of this thesis work. The module is designed with a high degree of flexibility, supporting three key features.

- The subdivision of phase space into arbitrary N dimensional binning. In each bin a separate classifier can be used.
- Support for all machine learning frameworks supported by the **basf2 mva** package.

This includes all currently popular frameworks such as **TMVA** [195], **XGBoost** [179], **TensorFlow** [196], and **pyTorch** [197].

- Support for arbitrary variable sets via the **basf2** variable manager.

For each event, the module operates on all tracks, finding and evaluating the classifier trained to cover the phase space region within which the track falls. From the classifier responses, a likelihood is extracted for each of the six stable charged particle hypothesis, assigned to the track, and stored in the **mDST**.

The classifiers presented in this and the following chapter have been adopted as the default particle identification contribution from the ECL, since late 2023.

5.3 Scheme overview

Having defined the **ECLChargedPIDMVA** module, responsible for evaluating the classifiers within **basf2**, we now discuss the classifiers themselves. We split the phase space into 18 regions, corresponding to all combinations of

- two charge regions (positive and negative charged tracks),
- three theta regions (the forward endcap (17.0–32.2°), barrel (32.2–128.7°), and backward endcap (128.7–150°) of the ECL),
- three lab-frame momentum regions (low momentum (0.2–0.6 GeV), medium momentum (0.6–1.0 GeV), and high momentum (1.0–7.0 GeV)).

As the polar angle selection is made to correspond to different physical regions of the ECL, we use θ_{cl} , the theta value associated with the ECL cluster position, not the track momentum. The training is segmented in momentum, as low-momentum particles are expected to be stopped in the ECL and exhibit significant track curvature before reaching the ECL, medium-momentum particles are on the threshold of muons punching through the ECL to reach the KLM, and high-momentum particles reach the ECL without large track curvatures.

For each region, we train a multiclass boosted decision tree classifier on relevant ECL based variables to distinguish the stable charged particle species. In contrast to the binary classification discussed in Ch. 4, multiclass classifiers are trained to separate more than two classes. For neural networks, this is achieved by having multiple output nodes, one per class. For boosted decision trees, this is achieved by training a set of one-vs-rest classifiers for each class. During application, any tracks with a cluster match that fall outside these boundaries are treated as belonging to the closest region.

5.4 Input variables

We consider a total of 75 input variables which fall within three categories: variables available in the **mDST** and thus used also in training the analysis BDT, Zernike moments

(defined below), and per crystal level information.

5.4.1 mDST variables

Nine variables related to the energy deposited by each track in the ECL, both in amplitude and shape, are available within the mDST format. Four of these will be discussed separately and the remaining five are as follows.

$E/|\vec{p}|$ The ratio of the energy of the cluster associated to each track measured in the ECL over the momentum of the track as measured by the tracking detectors. At present, the particle ID contribution from the ECL to the global likelihood scheme is based solely on this variable.

$E_{1/9}$ The ratio of the energy recorded in the most energetic crystal associated with the cluster to the sum of the crystals in a 3×3 grid about the centre.

$E_{9/21}$ The ratio of the sum of the energy recorded in the central 9 crystals in a cluster to the sum of the crystals in a 5×5 grid about the centre minus the 4 corners of the grid.

ΔL An estimate of the longitudinal shower depth [198]. The estimate is calculated by comparing the point of closest approach between the extrapolated track of the particle and the line passing through the centroid of the cluster and pointing in the average direction of the crystals associated with the cluster.

f_{LAT} The lateral energy distribution of the cluster defined as:

$$f_{\text{LAT}} = \frac{\sum_{i=2}^n w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1) r_0^2 + \sum_{i=2}^n w_i E_i r_i^2}, \quad (5.1)$$

where $r_0 \approx 5$ cm, i runs over the crystals associated with the cluster sorted in decreasing order of energy (E), w_i is the weight with which each crystal is associated with the cluster, and r_i is the distance of the crystal i to a projection of the cluster onto a plane perpendicular to the photon trajectory.

5.4.2 Zernike Moments

Zernike polynomials are a set of orthogonal polynomials on the unit disk. By calculating the moments of the energy distribution amongst the crystals associated with each cluster with respect to the polynomials we create a set of variables describing the shape of the cluster [199]. The shape is expected to differ between particles interacting electromagnetically or hadronically. The Zernike moments are calculated as:

$$|Z_{nm}| = \frac{n+1}{\pi} \frac{1}{\sum_i w_i E_i} \left| \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} w_i E_i \right|, \quad (5.2)$$

where n, m determines the moment rank with $m \leq n$, E_i and w_i are the energy and cluster-weight of the i -th crystal in the cluster, ρ_i is the radial distance of the i -th crystal from the cluster centre in the plane perpendicular to the cluster, and R_{nm} is defined as:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s}. \quad (5.3)$$

We consider 20 Zernike moments from $|Z_{10}|$ to $|Z_{55}|$. Higher order Zernike moments ($n \geq 6$) are known to be subject to noise and are thus discarded.

Two of the Zernike moments, $|Z_{40}|$ and $|Z_{51}|$, were found to be especially useful in separating neutral hadrons from photons. Furthermore, a classifier has been trained to separate neutral hadrons from photons by considering eleven Zernike moments calculated under both a photon and neutral hadron hypothesis [199]. The two selected moments and the classifier are available in the `mDST`. These three variables are also included in the analysis BDT.

5.4.3 Per-crystal properties of ECL clusters

For the 10 most energetic crystals in an ECL cluster, ranked by weighted energy ($w_i \cdot E_i$), we consider a set of per-crystal variables.

E_i The energy of the crystal (E_i). This is included only for classifiers targeting the low momentum region ($0.2 - 0.6$ GeV), the only category where it is found to substantially improve particle identification. As the variable does not have a sizeable discrimination power at higher momenta, and potentially introduces a dependence on distribution of particle energies used in the training, it is not included in the medium ($0.6 - 1.0$ GeV) or high momentum ($1.0 - 7.0$ GeV) regions.

$E_i / \sum_j E_j$ The ratio of the energy recorded in the crystal to the total energy of the cluster.

w_i The weight of the association of the crystal to the cluster (w_i). As multiple clusters can overlap the ECL reconstruction algorithms break the contribution of each crystal up amongst any potentially overlapping clusters. This is done by assigning the weight w_i .

r_i The distance of the centre of the crystal from the weighted centre of the cluster.

$\cos \theta_i$ The cosine of the polar angle between the crystal and the cluster centre.

The number of crystals associated with each cluster is highly variable, depending on the charge and species of the interacting particle. Therefore, in contrast to the previous variables, we must consider the treatment of missing values when a cluster has fewer than ten crystals associated to it. For all variables, we adopt a default value of zero.

5.5 Training Sample

The classifiers are trained on single-particle simulated samples. A particle gun ‘shoots’ a single track of a given species from the interaction point with momentum sampled from defined kinematic distributions. In each of the 18 (charge, momentum, polar angle) categories, we generate events for each of the six charged final state particles ($e^\pm, \mu^\pm, \pi^\pm, K^\pm, p^\pm, d^\pm$), uniformly distributed in $p - \theta$ within the bounds of each category. To avoid edge effects in the distribution of reconstruction level p and θ , we broaden the phase space in which particles are generated by 0.1 GeV in p and 5° in θ in both directions in each region.

Deuterons produced at the interaction point with low momenta, $p < 0.6$ GeV, do not reach the ECL, therefore, we consider only five of the six species within the categories corresponding to these regions. As it is not possible to return only likelihoods for a subset of the stable particle species within the Belle II particle identification system, within this region we set the deuteron likelihoods equal to those calculated for the proton hypothesis. Alternatively, one could assign a likelihood of zero, however, as deuterons could be produced at displaced vertices, or in material interactions, setting a likelihood of zero would cause such tracks to always be assigned a P_d value of zero. We therefore adopt the proton likelihood, which is expected to interact similarly to the deuteron in the ECL.

Electrons will lose momentum due to bremsstrahlung. To ensure that this does not bias the training sample we correct for bremsstrahlung by searching for photons emitted within a cone of 5° around the initial momentum direction of the track and summing their four-momentum to that of the track¹.

To perform the training, we select only reconstructed particles that have been matched to a generated track of the correct species and have reconstructed kinematic quantities within the defined bin boundaries. We additionally require that the particles have been matched to an ECL cluster, such that we can access the cluster-based variables used within the method. An alternative method which does not rely on track-cluster matching is discussed in section 5.10.2. For each species, in each category, we generate 180000 tracks that meet these requirements. Half of the generated sample is used for training; the remainder is further divided into thirds, which are used as a validation sample for early stopping conditions, to perform calibration, and as an independent testing sample.

While broadening the generated phase space ensures that there are no non-physical effects due to the simulation, there are some features to be considered in the generated distributions caused by the requirement of a matched cluster to the track. In particular, for low-momentum samples, we observe that for the heavier particle species a ‘turn-on’ curve is present in momentum. Figure 5.1 provides an example of the reconstructed momentum distributions of the training sample. For future developments, it may be beneficial to

¹For all variables used in the training, we consider the quantities calculated without the bremsstrahlung correction, as during final application the inference occurs before any possible bremsstrahlung correction can be applied.

introduce a reweighting procedure to ensure uniform distributions in $p - \theta$.

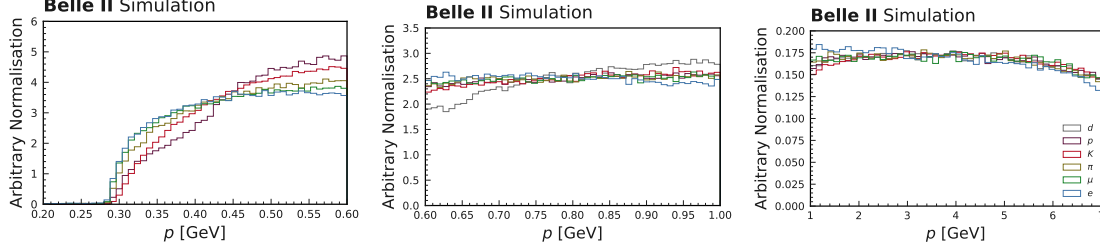


Figure 5.1: Reconstructed momentum distributions, of the negatively charged training sample for low (*left*), medium (*centre*), and high momentum (*right*) in the barrel region.

5.6 Training

For each region, the hyperparameters are optimised independently. We adopt an exponentially decaying learning rate (l) to allow the model to quickly learn to separate the samples while allowing later estimators to fine-tune the separation. To prevent a vanishing learning rate, we enforce a minimum, $l_{\min}$. The learning rate for tree n is thus given by

$$l(n) = \max(l_0(1 - r)^n, l_{\min}), \quad (5.4)$$

where l_0 , $l_{\min}$ and the decay rate r are hyperparameters. Optimisation is performed via a Bayesian optimisation process as implemented in Ref. [200] running for 500 trials. The complete set of hyperparameters and the allowed ranges are detailed in Table 5.1, and the optimum values found are given in appendix A.1.

Memory consumption during processing is a significant concern for `basf2`, with the requirement that total consumption remain below 4GB to efficiently balance CPU and memory resources. To ensure that the memory consumption of the trained set of classifiers is of a reasonable size, we limit the maximum size of each model by imposing a strict relationship between the maximum depth of a model, d , and the maximum number of estimators in the model, N . This is given by

$$N = 2^{12-d}. \quad (5.5)$$

This allows a model with trees of depth three to grow to 512 trees. Here, we note that the true total number of estimators is given by $N \cdot M$, where M is the number of classes learnt by the classifier, as multiclass BDT algorithms grow a series of one-vs.-rest classifiers. With this restriction, the memory consumption of the set of all 18 BDTs is approximately 180 MB. Allowing the trees to grow beyond this size was found to give only marginal improvement in classification performance at large increases in memory consumption. Given the tight memory restrictions of the experiment, it may be necessary to further limit the size of the classifiers in future.

To prevent the growth of needlessly complex models, we additionally allow for an early stopping condition. If the loss measured on the independent validation sample does not improve in an interval of 20 trees, training is terminated and the best model returned. In practice, we observe that given the strongly restricted maximum size of the model, early stopping is rarely triggered. Figure 5.2 provides an example of the loss curve on the training and validation samples and the evolution of the loss for each hyperparameter optimisation trial.

The full set of optimised hyperparameters for the nominal models is provided in appendix A.1. In general, the response in the loss is approximately flat to small perturbations of the hyperparameters about their best value; however, we note a few general trends.

- In each region the ‘loss guided’ training, growing trees node-by-node rather than level-by-level, provided the best result.
- Both the event-based subsampling and variable subsampling tended to be near one. Either overtraining did not affect performance in the validation sets or the regularisation parameters α , λ and γ were preferred to limit overtraining.
- The learning rate decay value tended to be small, indicating that the model preferred a flat learning rate rather than the exponentially decaying ansatz.
- Models preferred to have fewer trees with more nodes, indicating complex multidimensional relationships in the input features.

We evaluate the importance of the feature variables to lepton-hadron separation through their shapely values [187] as discussed in Sec. 4.1.4. The importances of the variables for electron, muon, and pion identification are given in Appendix A.2. As expected, the ratio of the recorded cluster energy to track momentum, $E/|\vec{p}|$, dominates for all species. Almost all variables are found to contribute to electron identification, while muon identification relies on only a few variables, with large contributions from ΔL and E_0 . One and two dimensional distributions of five of the most important variables for lepton hadron separation in the low and medium momentum barrel region, are shown in Fig. 5.3.

5.7 Calibration and likelihood extraction

Machine learning methods tend to be highly accurate but poorly calibrated, returning scores that are able to effectively separate signal from background, but do not reflect the probability of correctness. Although this predominantly affects deep neural networks rather than the simpler boosted decision-tree models considered here, we rely on the score returned by the model to assign a likelihood for each charged-particle hypothesis. Therefore, we adopt the temperature calibration method of Ref. [201] to ensure the best possible model calibration. Temperature calibration introduces the temperature parameter, T , to scale the response of the model. In our case, we consider the classifier responses, s_i . The

Parameter	Value	Description
Max Depth (d)	[1, 11]	Maximum depth of each decision tree. Note the maximum number of trees permitted to the model is dependent on the depth of each tree. If <i>lossguide</i> is adopted as the grow policy we instead limit the tree to 2^d nodes.
Initial Learning Rate (l_0)	[0.01, 1.0]	Initial learning rate of model. See text for more details.
Minimum Learning Rate ($l_{\min}$)	[0.001, 0.1]	Minimum learning rate of model. See text for more details.
Learning Rate Decay (r)	$[10^{-6}, 0.05]$	Decay of the learning rate. See text for more details.
Loss	<i>multi:soft-prob</i>	The figure of merit optimised by the model.
Alpha (α)	$[10^{-5}, 10^4]$	L1 regularisation term on the weights. Can help to prevent over-fitting.
Lambda (λ)	[0.05, 1000]	L2 regularisation term on the weights. Can help to prevent over-fitting.
Gamma (γ)	[0.001, 2]	Minimum reduction in loss required to further split a node of the tree. Can help to prevent over-fitting.
Subsample	[0.5, 1.0]	Fraction of training events sampled used to train each tree. Can help to prevent over-fitting.
Column Subsample	[0.5, 1.0]	Fraction of variables sampled used in training each tree. Can help to prevent over-fitting by forcing the model to rely on a broader range of variables.
Minimum Child Weight	10	Minimum number of training events required in each node after splitting a sample. Can prevent over-fitting by preventing extremely targeted selections.
Grow Policy	[<i>depthwise</i> , <i>lossguide</i>]	Whether new nodes are added to a tree in order of closest to the root node or in order of maximum reduction in loss. If <i>lossguide</i> is selected rather than limiting the maximum depth of the tree we limit the maximum number of nodes of each tree.
Tree Method	<i>gpu_hist</i>	The method used by XGBoost to train the trees. The <i>hist</i> methods pre-bin continuous features into histograms and base their node splitting on the bin indices of the histograms. This strategy is also employed by FastBDT [178].
nBins	512	Number of bins used to bin continuous features.

Table 5.1: Hyperparameters of the XGBoost model. For parameters tuned by Bayesian optimisation, the range investigated by the optimiser is provided in the “Value” column. All parameters not listed are left at their default value as per version 1.7.0 of XGBoost.

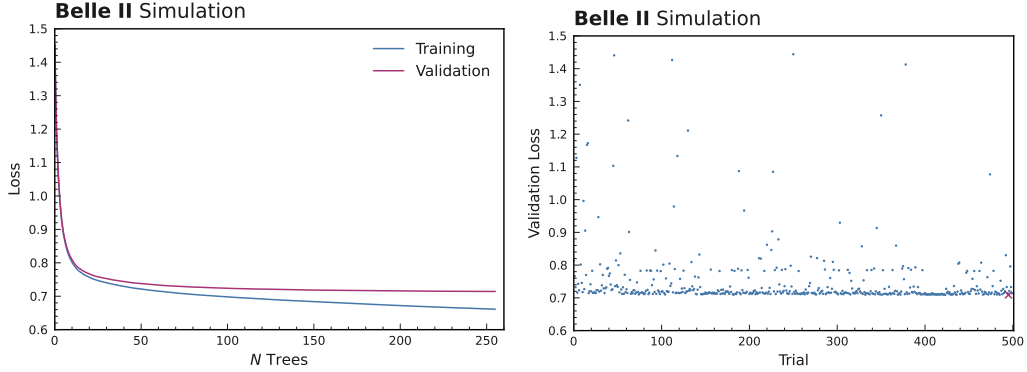


Figure 5.2: *Left:* Example losses for the training and validation samples as a function of the number of trees added. While the loss of the validation sample flattens out the performance continues to asymptotically improve and thus does not trigger the early stopping condition. *Right:* Validation loss for each of the 500 trials in the hyperparameter optimisation process, the best trial is marked by the purple X. The optimisation quickly finds the region of minimal loss with subsequent trials only fine tuning the hyperparameters. These plots correspond to the training of the medium momentum negatively charged region in the barrel.

calibrated scores, $\tilde{s}_i$, are then given by:

$$\tilde{s}_i = \frac{\exp(\log(s_i)/T)}{\sum_j \exp(\log(s_j)/T)} \quad (5.6)$$

where the value of T is optimised by minimising the cross-entropy on the dedicated calibration sample. For the BDTs considered within this work, we find that they are well calibrated, finding T values of approximately 1. Figure 5.4 provides an example of the impact of this calibration. Note that while for the training set we see some evidence of over-training, the model output is found to be well calibrated on the calibration and testing sets even before the temperature calibration. This holds for all 18 regions considered.

Having calibrated the model scores, we can interpret the classifier responses, $\tilde{s}_i$, as probabilities, $\tilde{s}_i = P(i|\vec{o})$, where i is one of the charged particle hypothesis considered, and $\vec{o}$ is the vector of observables considered by the model. We thus have:

$$\tilde{s}_i = P(i|\vec{o}) = \frac{\mathcal{L}(\vec{o}|i)P(i)}{\sum_j \mathcal{L}(\vec{o}|j)P(j)}, \quad (5.7)$$

As we ensure that the training sample is balanced between all hypotheses, we consider the priors to be flat, that is $P(i) = \frac{1}{6}$. The expression thus reduces to

$$\tilde{s}_i = \frac{\mathcal{L}(\vec{o}|i)}{\sum_j \mathcal{L}(\vec{o}|j)}. \quad (5.8)$$

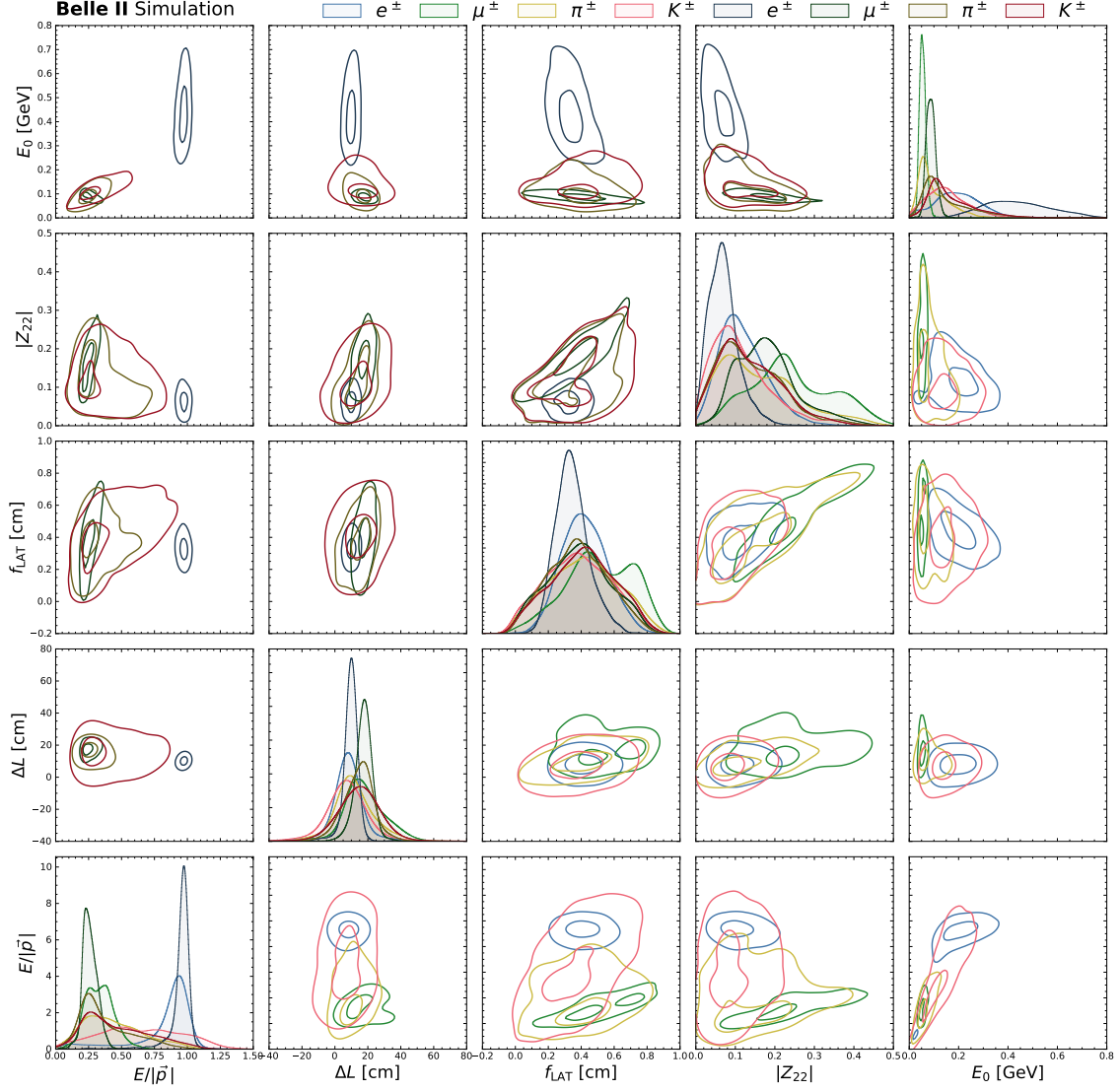


Figure 5.3: Distributions of five of the most powerful discriminating variables over the low and medium momentum barrel regions. One-dimensional distributions are shown on the diagonal, two-dimensional distributions on the off-diagonal, given by contours at 25% and 75% of the maximum density. The upper triangular shows medium-momentum particles, the lower triangular plots show low-momentum particles, which are represented by lighter colours than their medium-momentum counterparts. The samples contain equal numbers of positively and negatively charged particles. Note that, as per the discussion in the text, the energies of the ten most energetic crystal associated with the cluster, E_i , are included as features only for classifiers targeting low-momentum particles.

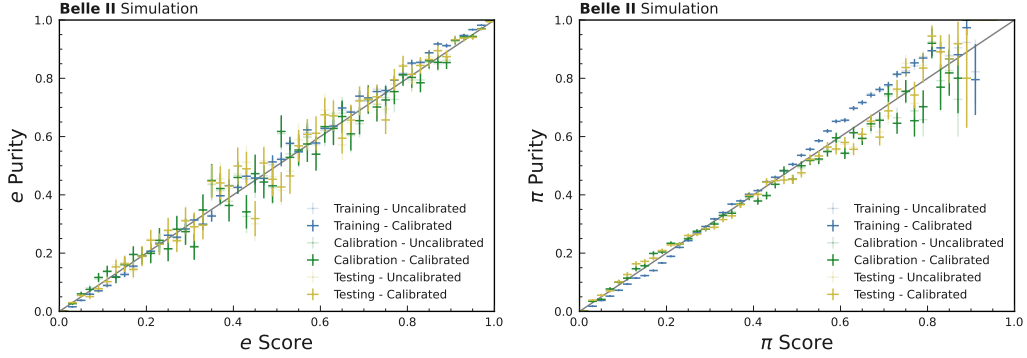


Figure 5.4: Example diagonal plots for electrons and pions with and without temperature calibration ($T = 0.996$) applied for the training sample (blue), the sample used to derive the temperature value (green) and an independent test sample (yellow). For each plot the score is divided into 50 equidistant bins. For a well calibrated model the purity of the sample within each bin is expected to be equivalent to the central score of the bin.

The log-likelihood of hypothesis i is therefore given by:

$$\log \mathcal{L}(\vec{o}|i) = \log \tilde{s}_i + \log\left(\sum_j \mathcal{L}(\vec{o}|j)\right). \quad (5.9)$$

We now note that for each track $\log(\sum_j \mathcal{L}(\vec{o}|j))$ is constant for all hypotheses. For particle identification, only comparisons of likelihood values, either through the difference in log-likelihood for two hypotheses or the likelihood ratio, are significant quantities. For both quantities, the aforementioned constant contribution cancels. Therefore, we consider the calibrated classifier response, $\tilde{s}_i$, as a likelihood to be used in the global likelihood scheme.

5.8 Performance in simulation

The performance of the models has been studied both in the independent simulated single-particle testing sample and in particles from simulated $B\bar{B}$ events. As no large differences were observed, we discuss only the performance on $B\bar{B}$ events. We consider 80000 simulated $B\bar{B}$ events, half simulated as B^+B^- , half as $B^0\bar{B}^0$ events. The Receiver Operator Characteristic (ROC) curves for ECL only electron and muon identification efficiency vs pion-to-lepton misidentification rates are shown in Figs. 5.5 and 5.6, respectively. The efficiencies ($\varepsilon_e, \varepsilon_\mu$) and misidentification-rates (ε_π) of a selection on one of the particle identification variables are defined as the fraction of the simulated particles of that type (e, μ or π) that pass the selection. We observe large reductions in pion misidentification rates when selecting electrons with BDT-based likelihoods compared to the $E/|\vec{p}|$ -based likelihoods or analysis BDT. A reduction of approximately factor 10 is observed in the barrel at a reference electron efficiency of 90%. The choice of 90% is arbitrary but lies within the regime commonly used by physics analyses. Smaller but still significant reductions in misidentification rate are observed in the end-caps. For muon identification, we observe

reductions in the misidentification rate of up to 50% at low momentum, when selecting muons with a 90% efficiency using $P_{e,\text{BDT}}^{\text{ECL}}$ as compared to selections based on $P_{e,E/|\vec{p}|}^{\text{ECL}}$. At high momentum, only small improvements are observed, with performance, as measured by the area under the ROC curve (AUC), even marginally degrading in the barrel.

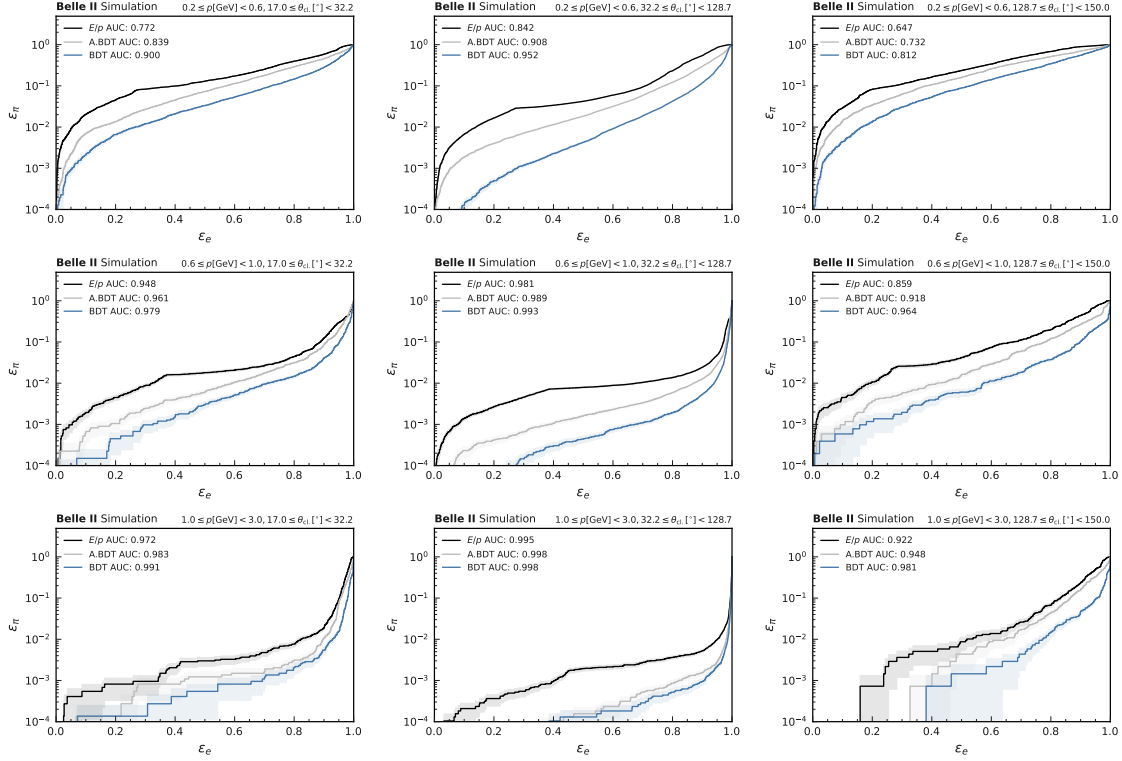


Figure 5.5: ROC curves for electron identification versus pion-to-electron misidentification in a simulated $B\bar{B}$ sample when enforcing selection thresholds on $P_{e,\text{BDT}}^{\text{ECL}}$ (blue), $P_{e,E/|\vec{p}|}^{\text{ECL}}$ (black) or $\mathcal{P}_e^{\text{ECL}}$ (grey). We consider nine regions of phase space, divided into three polar angle regions: the forward endcap (*left column*), barrel (*center column*), and backwards endcap (*right column*), and three momentum regions: low (*top row*), medium (*center row*) and high (*bottom row*). The high-momentum region is truncated at 3 GeV due to the limited kinematic reach of the tracks produced in $B\bar{B}$ events. In each region, we integrate over the kinematic phase-space and track charge. As all methods rely on track-cluster matching, only tracks that are successfully matched to a track are considered. The area under the respective $1 - \varepsilon_\pi = f(\varepsilon_e)$ curve for each method is provided in the legend.

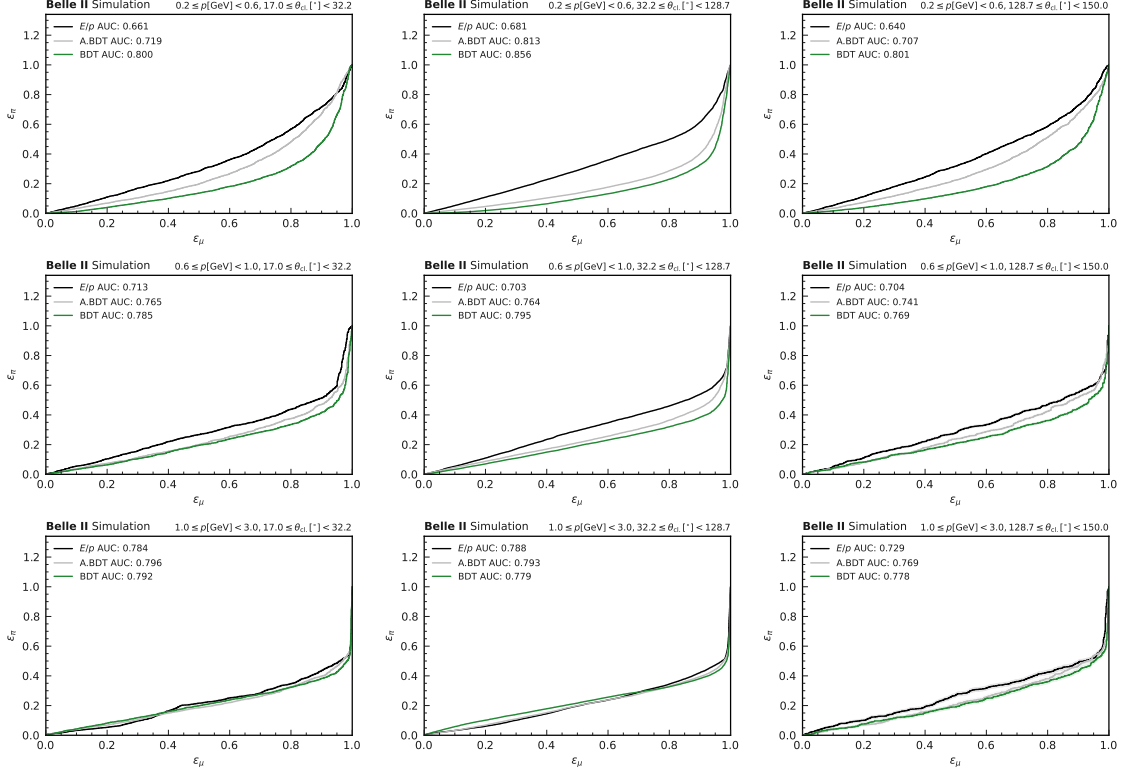


Figure 5.6: ROC curves for muon identification vs. pion-to-muon misidentification in a simulated $B\bar{B}$ sample when enforcing selection thresholds on $P_{\mu, \text{BDT}}^{\text{ECL}}$ (green), $P_{\mu, E/|\vec{p}|}^{\text{ECL}}$ (black) or P_{μ}^{ECL} (grey). We consider nine regions of phase space, divided into three polar angle regions: the forward endcap (*left column*), barrel (*center column*), and backwards endcap (*right column*), and three momentum regions: low (*top row*), medium (*center row*) and high (*bottom row*). The high-momentum region is truncated at 3 GeV due to the limited kinematic reach of the tracks produced in $B\bar{B}$ events. In each region, we integrate over the kinematic phase-space and track charge. As all methods rely on track-cluster matching, only tracks that are successfully matched to a track are considered. The area under the respective $1 - \varepsilon_{\pi} = f(\varepsilon_{\mu})$ curve for each method is provided in the legend.

5.9 Stability against beam background levels

In addition to the signals generated by the decay products of the primary interaction of interest, each event recorded by Belle II contains signals left by what is referred to as ‘beam background’. Where beam background is a collective term for processes arising from beam-beam interactions which scale with the instantaneous luminosity, (Bhabha scattering, and two-photon processes, $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$) or from interactions within each beam, or between the beam and residual gas remaining in the beam pipe such a

Coulomb scattering, Bremsstrahlung, or the Touschek effect. These effects and mitigating efforts, such as beam collimator schemes and shielding, are described in detail in Ref. [202]. Such beam backgrounds deposit additional energy in the ECL. If the deposits overlap with a cluster from a track of interest, the energy will be erroneously assigned to the track, altering the magnitude and shape of energy deposition and thus potentially degrading the particle identification performance. As the amount of beam background scales with luminosity and beam currents², the stability of particle identification performance against different run conditions must be investigated.

To test the stability of the trained classifiers against beam background levels, we consider the two simulated beam background levels that are available. These are, in internal Belle II jargon, ‘early phase 3’ corresponding to an approximate luminosity of $L = 1.06 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, roughly the average luminosity achieved before the first long shutdown, and ‘nominal phase 3’ corresponding to an approximate luminosity of $L = 8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$, the design luminosity of SuperKEKB. All simulated Belle II samples discussed elsewhere within this thesis are simulated with early phase 3 background³. We repeat the training and optimisation procedure with the nominal beam background samples and compare the performance of both sets of classifiers on $B\bar{B}$ samples generated at each background level. This is shown in Fig. 5.7. A performance drop can be observed in most phase space regions when applying BDTs trained with early phase 3 backgrounds to the nominal phase 3 sample. This performance loss is mostly recoverable by training with the nominal phase 3 sample. This suggests that the BDTs should be re-trained regularly to match the beam background conditions of the collected data. We also observe that when training with nominal phase 3 backgrounds, performance on the early phase 3 sample matches that of the MVAs trained on the early phase 3 sample. We postulate two contributions to this: that the higher beam background levels in the particle gun sample match the busier environment in the $B\bar{B}$ test sample, and that the higher background training sample contains examples of clusters both with and without significant beam background contributions, and thus learns to identify particles under both conditions, while the early phase 3 training sample does not contain events with high background levels and thus performs poorly in a high-background environment.

5.10 Future developments

Two additional developments are in preparation and expected to become available in late 2024⁴, the inclusion of pulse shape information as discriminating variables, and the use of

²Neglecting ongoing developments to background mitigation.

³We use what is referred to as run-independent samples, generated with a nominal detector condition and beam background levels. For Belle analyses, simulated samples are run-dependent, matching detector conditions, and using randomly triggered samples to overlay real beam background signals over the simulated signal events. The transition to run-dependent simulations is ongoing within the Belle II experiment.

⁴For `release-9` of `basf2`.

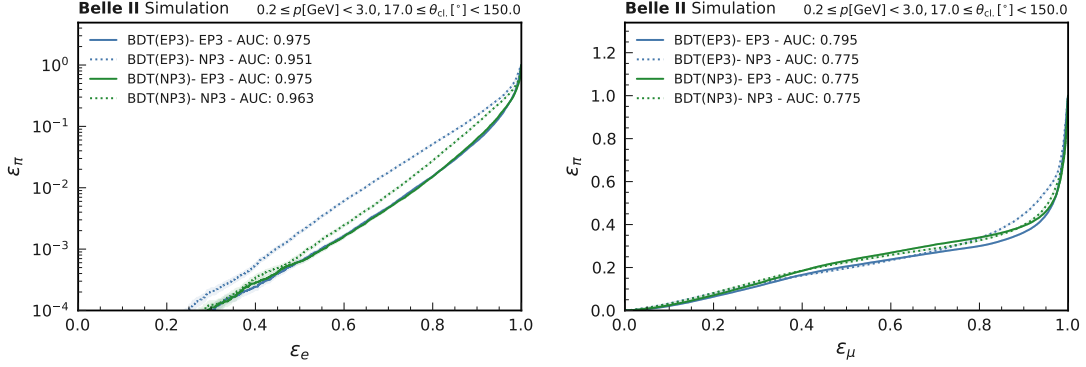


Figure 5.7: Comparison of electron pion (*left*) and muon pion (*right*) separation under different beam background conditions. The first sample given in the legend denotes the beam background level in the generation of the BDT training samples, either nominal (NP3) or early (EP3) phase 3. The second denotes the background conditions used to generate the $B\bar{B}$ sample on which the comparison was performed. The two sets of BDTs shown have been trained following the procedure outlined within this document on the same variable set.

convolutional neural networks within the central barrel regions of the ECL.

5.10.1 Pulse shape discrimination

The scintillation response of the CsI(Tl) crystals of the calorimeter differs for hadronic showers and electromagnetic showers. Highly ionising particles, such as protons that can be produced in the hadronic shower, create signals with decay times faster than those of photons and less ionising particles [203]. In addition to reporting the total recorded energy and timing information, the readout electronics of the Belle II calorimeter are capable of recording the waveform of the signal pulse for each crystal, sampled at 31 points at a frequency of 1.7669 MHz. To maintain a manageable data rate for the data acquisition systems, waveforms are recorded only for crystals with recorded energy exceeding 30 MeV. Each waveform is fitted to extract the fraction of the total energy emitted by the hadronic component.

The use of pulse shape information for particle identification has been studied in Ref. [203] in the context of $\gamma - K_L^0$ separation. We investigate its efficacy for lepton identification by including the hadron energy fraction as additional per-crystal feature variable for the ten most energetic crystals to the BDTs. If no waveform fit is available for a crystal, a default value of zero is adopted for the hadron energy fraction. The performance of the models with and without pulse shape information is compared in Fig. 5.8. The figure shows $1 - \text{AUC}$ in bins of track momentum, where AUC is the area under the curve of the receiver-operator characteristic curves in each momentum bin. This figure of merit highlights the momentum dependence of the particle identification performance

of each algorithm, while providing a single summary statistic for each bin. We integrate over polar angle and charge. The inclusion of the hadron energy fraction is found to improve the electron-pion separation, while having only a nonnegligible effect on the lowest momentum bin for muon-pion separation.

Before late 2023 (**basf2 release-8**) the waveform fit required manual calibration, which was only performed for every major reprocess. Therefore, pulse shape information was unusable for data collected since the last major reprocessing. Thus, it was excluded from consideration. In **release-8** this was remedied by an automated calibration procedure; therefore, the performance of the ECL particle identification performance in **release-9**, expected to become available in late 2024, can be further improved by considering pulse shape information.

5.10.2 Convolutional-neural networks

Our nominal model, the set of boosted decision tree classifiers, uses human-engineered variables that describe the shape of the shower. A powerful alternative to this approach is convolutional neural networks (CNN), deep learning models designed for image recognition [204]. CNNs can directly consider the energy deposited in a grid of crystals, analogous to pixels in an image, avoiding the assumptions and potential biases introduced by human-engineered features. Additionally, because of the convolutional nature of the network, locally combining information from adjacent crystals, the per-crystal information is exploited more efficiently than by the BDT.

The use of convolutional neural networks for muon-pion and electron-pion separation in the ECL was demonstrated in Refs. [205, 206], using crystal grids of 7×7 or 11×11 , centred on the crystal hit by the extrapolated track, to form an image where each pixel contains the energy of one crystal. As the method is not dependent on cluster formation or track-cluster matching, it is capable of providing particle identification regardless of the success of these algorithms. Recently, a set of multiclass convolutional networks has been developed and integrated into the **ECLChargedPIDMVA** module [207]. The network considers 11×11 images for low-momentum tracks and 7×7 images for medium-momentum tracks. In addition to the image, the track momentum and azimuthal and polar angle information for the crystal to which the track is extrapolated are considered by the network. No network is trained for the high-momentum region, as a strong dependence on momentum was observed within this region. Likelihoods are extracted from the networks using the same procedure as for the BDTs.

As the network requires a regular grid of crystals (pixels), which is not available in the end-caps, it is restricted to a sub-region of the barrel with the requirement that a full image can be formed. For the following comparison (and in application) where no network is available, the BDT based likelihoods are used. The hadron energy fraction for each crystal, determined from pulse-shape analysis, can be included in the network through the use of a second channel in the convolution.

Figure 5.8 presents a comparison of the CNN method to the BDT based method with

and without including pulse shape information. The CNN based algorithm is demonstrated to outperform the BDT based algorithm for both muon and electron identification. Validation of performance in data is ongoing at the time of writing.

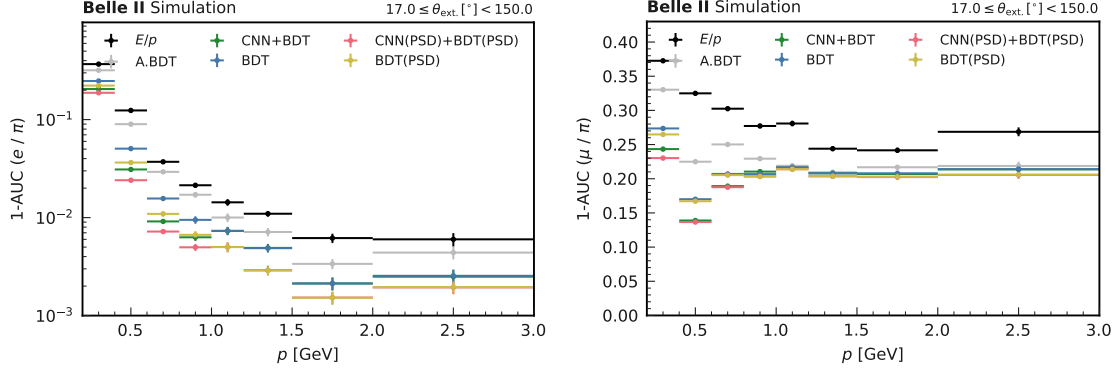


Figure 5.8: Comparison of electron pion (*left*) and muon pion (*right*) separation achieved using ECL information in simulated tracks originating from $B\bar{B}$ events. The comparison includes the univariate $E/|\vec{p}|$ based algorithm (black), the analysis level BDT (grey), the nominal BDT with (yellow) and without (blue) considering also the hadron energy fraction as computed from the pulse shape, and a model utilising the CNN with (pink) and without (green) considering the hadron energy fraction. As the kinematic coverage of the CNN is limited, algorithms that use it default to the BDT outside the covered region (the central barrel for tracks with momenta below 1 GeV). These algorithms are therefore labelled CNN + BDT. No requirement of a track-cluster match is made for this figure.

ECL particle identification: validation in data

We compare the performance of the new ECL particle identification algorithms with the previously available models.

We assess the modelling of input features and validate the performance of the new ECL charged particle identification algorithm in data. Three high-purity ‘standard candle’ channels, $K_s^0 \rightarrow \pi^+\pi^-$, radiative di-muon ($e^+e^- \rightarrow \mu^+\mu^-\gamma$), and Bhabha scattering $e^+e^- \rightarrow e^+e^-(\gamma)$, are used to probe performance on pions, muons, and electrons, respectively. The considered data sample consists of 11.6 fb^{-1} collected at the $\Upsilon(4S)$ resonance in late 2021 which has been made available by the Belle II data production group for this purpose. Due to the high rate of Bhabha scattering and K_s^0 production, these channels have been pre-scaled by factors of 200 and 4, respectively, reducing the effective data size. For each channel, the data sample is compared with simulated samples. For K_s^0 , 2.5 fb^{-1} equivalent of $B\bar{B}$ and $q\bar{q}$ samples are used. For the di-muon channel, 100 fb^{-1} equivalent of $e^+e^- \rightarrow \mu\mu(\gamma)$. For the bhabha sample 0.1 fb^{-1} of $e^+e^- \rightarrow e^+e^-(\gamma)$ events and between 2 to 10 fb^{-1} equivalent of other low multiplicity channels that can be incorrectly reconstructed as bhabha events, $e^+e^- \rightarrow \tau\tau, \gamma\gamma, eeee, ee\mu\mu, ee \rightarrow \pi\pi$.

To further purify the samples, we adopt a tag-and-probe approach in all three channels. For each channel, one of the two tracks is subjected to strict particle identification requirements, effectively tagging the event. The particle identification performance is probed on the other track.

6.1 Reconstruction

To select and reconstruct the three standard-candle channels, we closely follow the reconstruction procedures of the lepton identification performance group. As the study aims to compare particle identification methods which require track-cluster matching, we consider only candidates in which the probe track has been matched to a cluster.

6.1.1 $e^+e^- \rightarrow e^+e^-(\gamma)$

To reconstruct Bhabha events, we search for events with exactly two tracks that pass the requirement, $|dz| < 5\text{ cm}$ and $dr < 2\text{ cm}$. We identify these tracks as electrons and correct them for bremsstrahlung by searching for signals in the ECL consistent with photons

originating from the intersection of the electron tracks and the layers of the vertex detectors. The four momenta of these photons are summed to the four momentum of the electron candidate. We require that the square of the recoil mass against the dielectron sample does not exceed 10 GeV^2 . We select events that were triggered by the `lm11` trigger, requiring an ECL cluster in the barrel region with a minimum energy of 2 GeV , and do not pass the `hie` trigger requirements, which requires a minimum total energy of 1 GeV to be deposited in the ECL, and that the event fails a three dimensional Bhabha veto¹. The second requirement notably improves the agreement of simulation with data for track candidates with momenta between 3 and 4 GeV . We further require that the event passes the high level trigger, `selectee`, which identifies Bhabha events by requiring the invariant mass of the track pair to exceed 5.29 GeV and that either one or both tracks are associated with an ECL cluster with an energy exceeding 1.5 GeV . For the studied data sample, the `lm11` trigger was pre-scaled by a factor of 2 , and the high level trigger selection by a factor of 100 to reduce the total data rate. To veto $e^+e^- \rightarrow \gamma\gamma$ events, where one or two of the photons convert to electrons, we require $M_{ee} > 0.4 \text{ GeV}$.

The data used within this study was further skimmed by the Belle II lepton identification performance group. To keep the retention of the skim at a manageable rate, a selection requiring that at least one of the tracks has a high electron identification rate is included. For data, the skim was produced centrally by the Belle II data production group. For the simulated sample the skim was performed during the event reconstruction described here, therefore, the updated bdt-based ECL particle identification introduced within this work is used. There is thus a mismatch in which algorithm contributed the ECL contribution to this selection. We therefore enforce the selection on both versions of the algorithm, $P_{e,\text{BDT}}^{\text{TOP}} > 0.95$, and $P_{e,E/|\vec{p}|}^{\text{TOP}} > 0.95$, where the subscripts BDT and $E/|\vec{p}|$ identify the respective algorithm. As denoted by the overline, the TOP likelihoods are excluded on recommendation of the Belle II performance group, which observed that inclusion of the TOP degraded electron identification. Furthermore, in data processed with `basf2 release-6` the performance group identified issues with the SVD likelihoods, which were added for the first time in this release. It is therefore recommended to exclude the SVD contributions from lepton identification. This includes all Belle II data presented within this thesis. However, the definition of the skim used for Bhabha selection includes the SVD contribution in the tag side selection, degrading electron identification performance on data but not on simulation. As this applies to the tag electron, it is not expected to bias the probe electron selection.

To avoid potential bias from the particle identification requirement and the cluster energy requirements of the trigger selections, we enforce that the tag track passes all selections. The performance can thus be studied on the probe track without bias.

¹The veto searches for two high energy clusters, both with energies above 3 GeV , at least one with energy above 4.5 GeV . The azimuthal opening angle of the clusters must be between 160 – 200° in the c.m.s. system, and the sum of their polar angles between 165 – 190° .

6.1.2 $e^+e^- \rightarrow \mu^+\mu^-\gamma$

To reconstruct radiative di-muon events we begin with events identified by the high-level trigger radiative di-muon selection [208]. This selection is primarily designed to select a clean di-muon sample, with a wide momentum range, for calibration of the CDC, but is also used for calibration of the muon identification performance. It searches for tracks that originate near the interaction point, requiring $d_0 < 2$ cm and $|z_0| < 4$ cm, where d_0 and z_0 are the point of closest approach of the track to the coordinate system origin in the radial and z direction. Each track is required to have a minimum transverse momentum of 200 MeV, have a polar angle within the CDC acceptance, and have been seen in the CDC. Furthermore, each track is required to be associated with a cluster of at most 400 MeV of energy. This selection captures almost all muons which, as minimally ionising particles, are expected to deposit approximately 200 MeV in the ECL. If the track was seen in the KLM the cluster energy requirement is tightened to 250 MeV and the track is required to have a momentum less than 2 GeV.

The event is required to contain at most three tracks, with exactly two meeting these requirements. The two tracks must be separated by at least 90° in azimuthal angle. The higher momentum track must have at least 1 GeV of energy, with the lower momentum track having at most 3 GeV. At least one of the tracks must have a cluster energy of less than 250 MeV and at least one must not be seen in the KLM. In addition, it is required that the recoil momentum against the di-muon system be at least 100 MeV, consistent with a radiated photon.

We tighten the skim selection, requiring exactly two tracks. Furthermore, we enforce the standard impact parameter selections used in lepton identification studies, $|dz| < 2$ cm and $|dr| < 5$ cm, although the effect of this selection is expected to be negligible given the selections within the skim. Both tracks are required to have momenta between 0.2 and 6.5 GeV in the lab frame. We further require the presence of a photon with an energy in excess of 1 GeV, found in the ECL with a polar angle between 17° and 150° , such that if this cluster was created by a charged particle, the charged particle would be within the acceptance of the CDC. The photon must have deposited energy in at least two of the ECL crystals. Finally, the di-muon system must have an invariant mass between 10.2 and 10.8 GeV.

For the tag track, we require $P_{\mu, \text{BDT}}^{\text{SVD}} > 0.9$ to ensure that a pure sample of probe tracks is selected.

6.1.3 $K_S^0 \rightarrow \pi^+\pi^-$

We preselect events passing the high level trigger **hadron**, requiring at least three tracks that originate close to the interaction point with at least 200 MeV of transverse momentum, and vetoing Bhabha events by looking for high-energy back-to-back tracks. We form K_S^0 candidates from two oppositely charged tracks that originate near the interaction point, requiring each track to originate or pass close to the interaction point, $|dz| < 5$ cm and $dr <$

2 cm, and that the invariant mass of the two track system lies between 450 and 550 MeV. Due to the long lifetime of the K_S^0 , true K_S^0 candidates decay at measurable displacements from the interaction point ($c\tau \approx 2.7$ cm), we perform a vertex fit to find the decay vertex and improve the mass resolution, requiring a p value above 0.05. This requirement is somewhat at odds with the requirement that the tracks pass close to the interaction point. However, as we are interested in the performance of the particle identification systems for all tracks in $B\bar{B}$ events, of which almost all originate at the interaction point, we enforce the impact parameter requirements to select such tracks, while requiring consistency of the vertex position and momentum vector to select true K_S^0 decays.

We then tighten the invariant mass requirement, requiring an invariant mass within 10 MeV of the world average K_S^0 mass, and require that the cosine of the angle between the momentum vector of the reconstructed K_S^0 and the vector connecting the interaction point to the decay vertex of the K_S^0 is greater than 0.998. The reconstruction is inclusive, besides the skim selections, no requirements are made on particles, or lack of particles, other than the signal K_S^0 . For the tag track, we require $P_{\pi, \text{BDT}}^{\text{SVD}} > 0.99$.

6.1.4 Kinematic distributions

The aim of this work is to improve low-momentum lepton identification; therefore, for the purposes of this thesis, we restrict our consideration to tracks with momenta between 0.2 and 1.0 GeV. We also consider only tracks in the barrel region, since the Bhabha sample is highly asymmetric, with trajectories of electrons lying predominantly in the forward direction and positrons in the backward direction. In the barrel region approximately equivalent numbers of positrons and electrons are recorded. It is expected that the lepton identification performance is approximately uniform across the ECL barrel; therefore, the polar angle asymmetry of the Bhabha sample within this region is taken to be negligible for the following performance study, which is intended to serve only as a preliminary demonstration of the improvement in lepton identification due to the BDT based likelihood calculation. Performance in the end caps and at higher momentum is observed to match the trends seen within the low momentum barrel region.

The invariant mass distributions of the selected di-track systems, the momentum distributions of the probe tracks, and the polar angle distributions of the clusters associated with the probe tracks are shown in Fig. 6.1. The MC is normalised to the integrated luminosity. We observe some normalisation differences, which are expected to originate (in part or full) from the lack of particle identification efficiency correction on the tag-side track. The distributions of tracks in momentum and polar angle are observed to be well modelled.

6.2 Feature variable modelling

We study the modelling of the input feature variables in the three control samples with the aim of excluding any variables that exhibit notably poor agreement between data

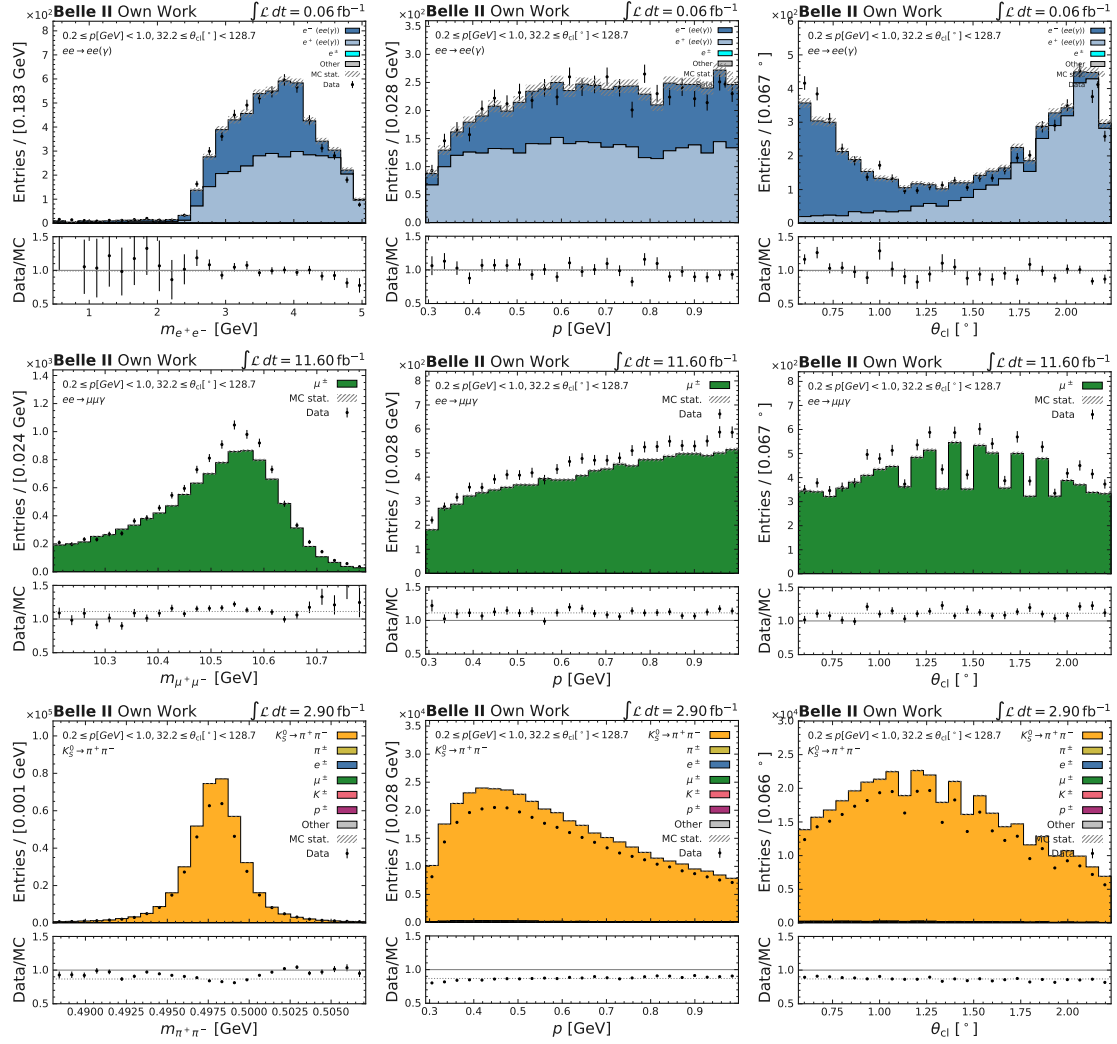


Figure 6.1: Di-track invariant masses and kinematic distributions of tracks from the Bhabha (*top*), radiative di-muon (*center*), and K_S^0 (*bottom*) samples. The dotted line in the lower plots gives the ratio of integrated data yield to integrated MC yield. The MC is normalised to the integrated luminosity.

and simulation. We observe that almost all variables show some level of disagreement, with several of the most important feature variables, such as $E/|\vec{p}|$, showing large shifts between the data and MC distributions. The distributions for the five most important input variables are given in Fig. 6.2. Given that $E/|\vec{p}|$ dominates ECL-based particle identification, we cannot exclude it from consideration without significantly degrading the performance of the lepton identification. However, ECL-based lepton identification based on $E/|\vec{p}|$ has demonstrated good performance, with studies of lepton efficiency in data and simulation agreeing within $\sim 5\%$ [209]. Therefore, we reluctantly consider the data-MC disagreements observed across all variables, the largest of which are of similar magnitudes

to that observed in $E/|\vec{p}|$ as acceptable. Future efforts must be made by the collaboration to improve the modelling of ECL showers.

6.3 Performance

We perform a quantitative comparison of the pion-to-lepton misidentification rates at fixed lepton identification efficiencies. For each algorithm, we impose threshold selections on either the ECL only or global likelihood ratios (see Eq. 3.4), or the response of the analysis level BDT, with the thresholds selected such that in MC 90% of true leptons are selected. Given the purity of the Bhabha and radiative di-muon samples, the corresponding lepton efficiency in data is estimated by simply counting the candidates above and below the threshold. Although the K_S^0 sample is of similar purity, leptons that contaminate the sample will pass the selection, whereas pions are largely rejected. The sample passing the selection can thus have a non-negligible fraction of true leptons. For this proof-of-method study, we subtract this contribution based on MC expectation, assigning a 100% uncertainty to the subtraction². For almost all selections studied, the expected lepton contamination in the pion sample is expected to be below 5%, reaching higher fractions only when the misidentification rate lies below 10^{-3} , as is the case with selections on $P_{e,\text{BDT}}^{\text{SVD, TOP}}$.

The procedure is performed in momentum bins of 100 MeV to reduce the impact of the different momentum distributions of the three samples and to highlight the momentum dependence of the lepton identification performance. Fig. 6.3 presents the efficiencies and misidentification rates for ECL-only particle identification. The reconstruction level BDT achieves a consistent factor ten reduction in the pion-to-electron misidentification rate compared to the $E/|\vec{p}|$ based likelihoods, and factor three compared to the analysis level BDT. For muon identification, we observe that the misidentification rate from pion to muon is reduced by factors between 2 and 1.4 in the momentum regime studied, compared to the $E/|\vec{p}|$ based particle identification. A small improvement, of approximately 5 – 15% reduction in misidentification rate, is observed compared to the analysis level BDT. The agreement between simulation and data in the pion-to-lepton misidentification rates is observed to be worse for the reconstruction level BDT, than the $E/|\vec{p}|$ based particle identification. Approximately doubling over the studied phase-space.

Considering global particle identification, as presented in Fig. 6.4, we note that the reduction in pion-to-electron misidentification rate moving from the $E/|\vec{p}|$ based to the BDT based likelihoods is approximately one order of magnitude. This is consistent with the reduction observed in the ECL-only identification and the importance of the ECL to electron identification. For muons, we observe reductions in misidentification rate between 25–40% at low momentum, with the performance of the methods equalising at

²When performing the official calibration of lepton identification performances, the Belle II charged particle identification group takes a more sophisticated approach, fitting the K_S^0 mass peak. Estimates of the misidentification rate using the K_S^0 sample are additionally combined with estimates from three-prong τ decays.

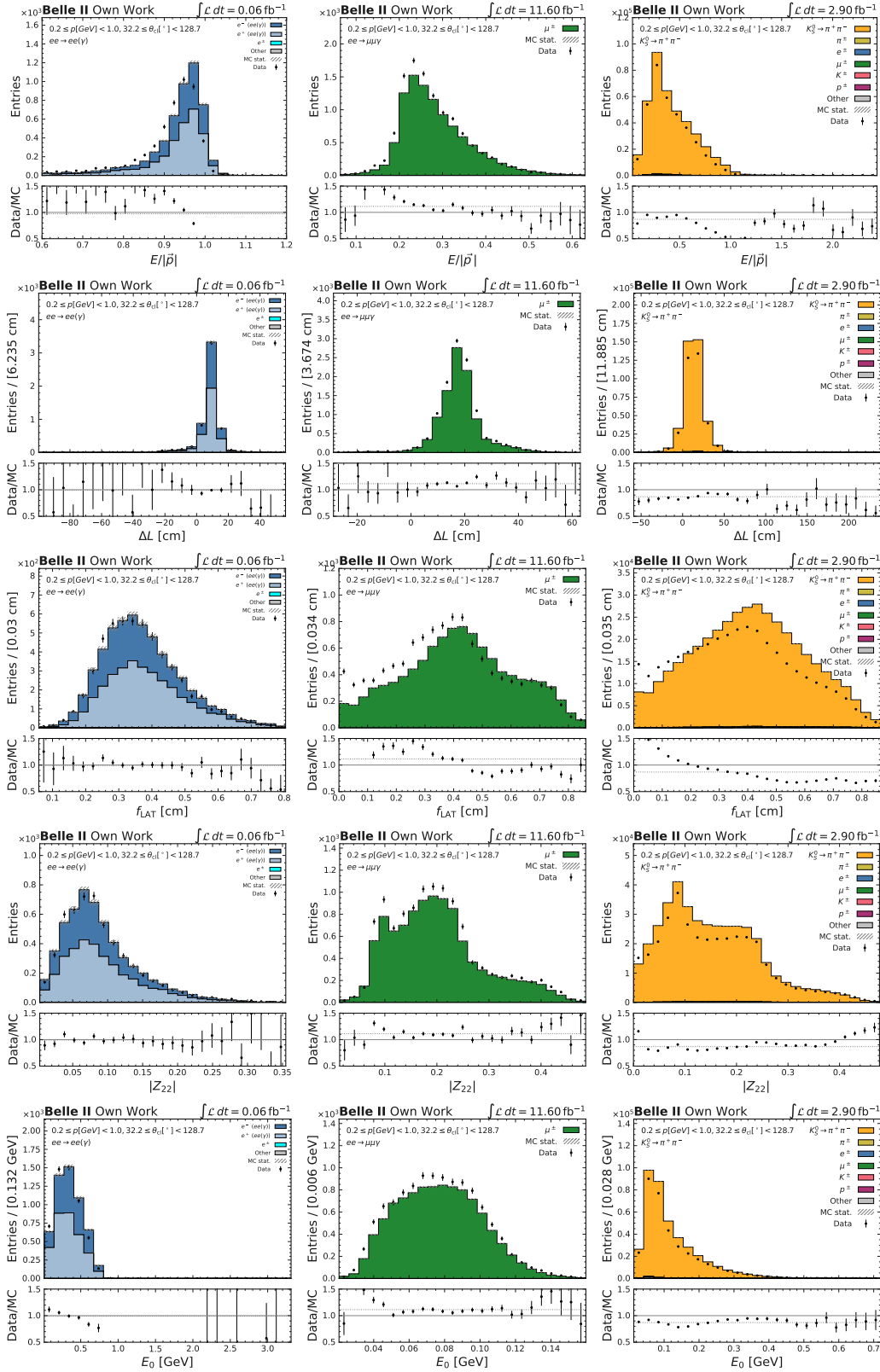


Figure 6.2: Distributions of the five most important feature variables in the K_S^0 , Bhabha and radiative di-muon control samples. The MC is normalised to the integrated luminosity.

higher momentum where the contribution of the KLM dominates particle identification. Finally, we observe that the performance of the global particle identification is similar in data and MC using both methods, indicating that data-MC disagreement in other subdetectors dominates the disagreement. Thus, while the BDT based method introduces larger disagreement in the ECL-only performance, for physics analyses, which use the global particle identification, significant reductions in pion-to-lepton misidentification rates are achieved without significant changes to the data-MC agreement. Similar performance improvements are expected in the kaon-to-lepton misidentification rates.

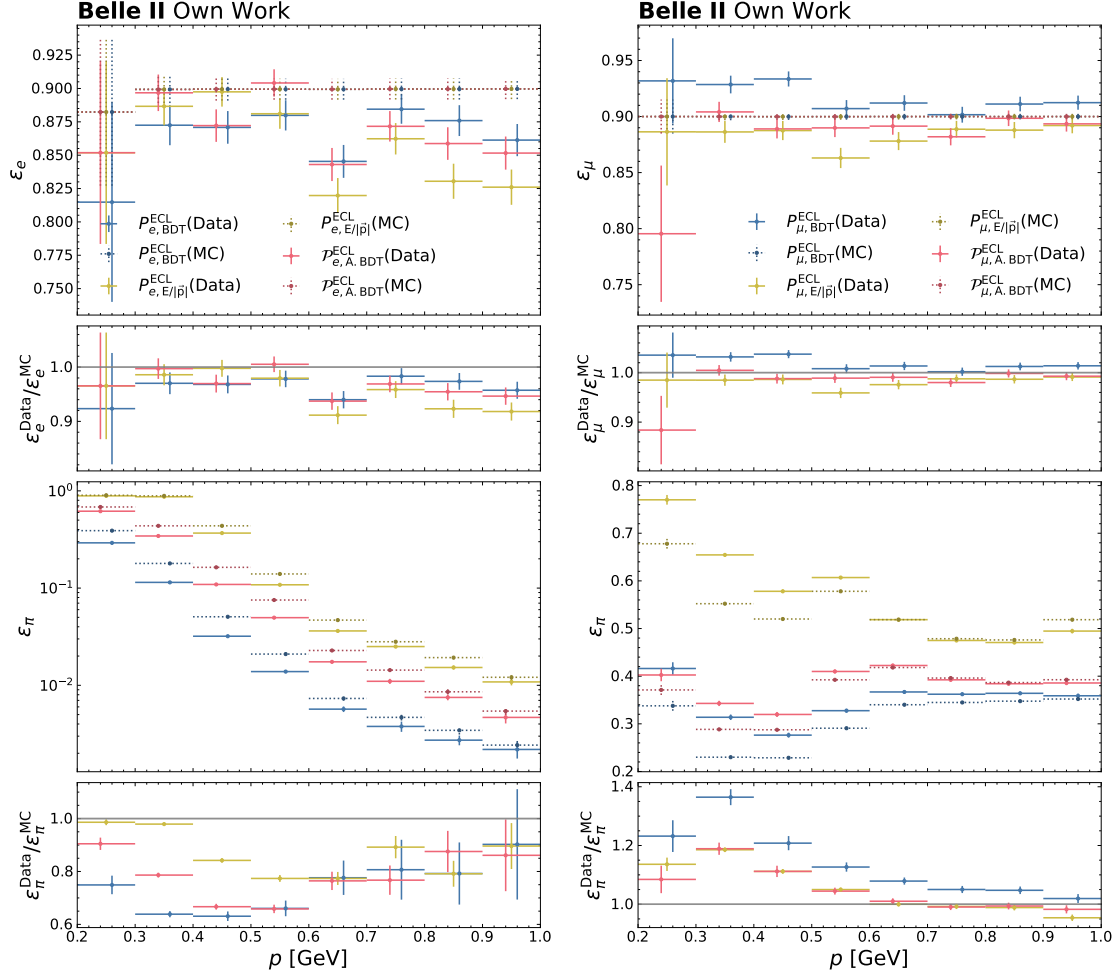


Figure 6.3: Comparison of ECL only lepton identification performance, and data MC agreement, in the barrel region. In each 100 MeV momentum bin thresholds are placed on the output of the three particle identification algorithms under consideration, the likelihood ratios computed with the BDT and univariate $E/|\vec{p}|$ based algorithms, and the analysis level BDTs, such that a lepton identification efficiency of 90% is achieved in MC. From top to bottom, the panels show the lepton identification efficiency, the ratio of lepton identification efficiencies in data over MC, the pion-to-lepton misidentification rate, and the ratio of pion misidentification rates in data over MC. See the text for details of the samples used in the comparison.

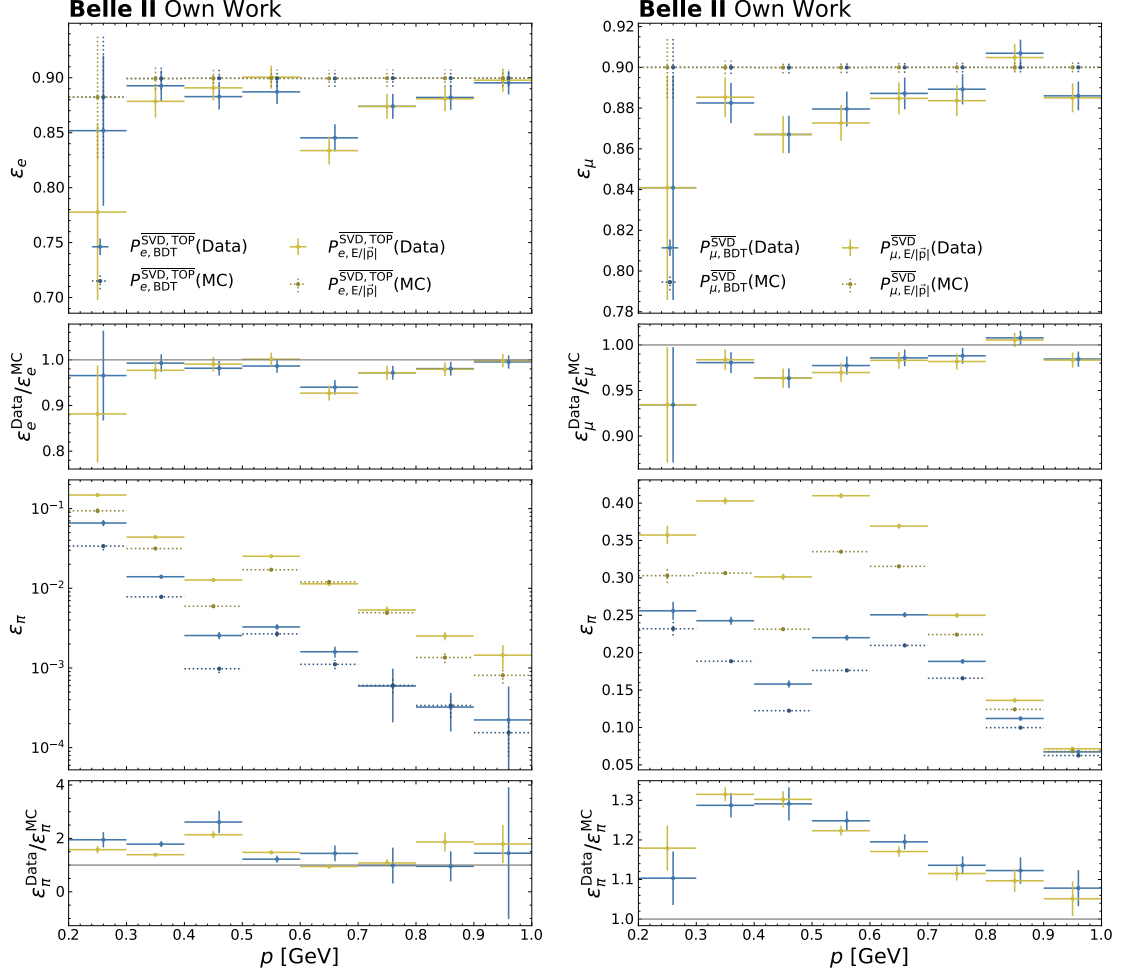
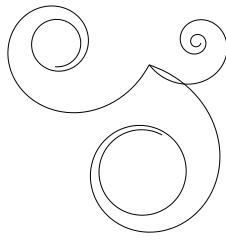


Figure 6.4: Comparison of global lepton identification performance, and data MC agreement, in the barrel region. In each 100 MeV momentum bin thresholds are placed on the output of the two particle identification algorithms under consideration, the likelihood ratios computed with the ECL contributions from the BDT and univariate $E/|\vec{p}|$ based algorithms, such that a lepton identification efficiency of 90% is achieved in MC. The analysis level BDT is excluded from the comparison as it includes information from the TOP and SVD. From top to bottom, the panels show the lepton identification efficiency, the ratio of lepton identification efficiencies in data over MC, the pion-to-lepton misidentification rate, and the ratio of pion misidentification rates in data over MC. See text for details of the samples used in the comparison.



Part III

Inclusive semileptonic B decays

Modelling

We discuss the modelling of the key signal and background channels in the analyses: $\bar{B} \rightarrow X_u \ell \bar{\nu}$, $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and weak annihilation. Additionally, this chapter serves to introduce the event categorisation that will appear in figures and discussions for the remainder of this thesis.

Simulated samples of the signal and background channels play an important role in most particle physics analyses, being used to prepare selections, estimate the reconstruction efficiency, and extract the signal yields. Thus, it is important that the modelling of these channels in the analyses represents the most up-to-date knowledge. As both analyses presented within this thesis consider $\bar{B} \rightarrow X \ell \bar{\nu}$ decays, with significant overlaps in modelling, we present the models used for both analyses in tandem. For brevity we refer to the analyses by the name of the experiment at which they are performed.

The following discussion examines the nominal models considered. Discussions on variations of the modelling as systematic uncertainties and technical details as to how the central and varied models are implemented within the analyses are deferred to the respective discussions on corrections and systematic uncertainties for the analyses.

7.1 Charmed semileptonic decays

Charmed semileptonic decays, $\bar{B} \rightarrow X_c \ell \bar{\nu}$, being Cabbibo-favoured over the charmless decay, are the dominant component of $\bar{B} \rightarrow X \ell \bar{\nu}$. We model such decays as the sum of exclusive channels. For the Belle analysis we closely follow the modelling (and MC samples) of Refs. [63, 73]. This modelling, with small variations, has been adopted as the standard model for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ decays by the Belle II collaboration [210], and is thus also used in the search for weak annihilation.

7.1.1 Branching fraction

The inclusive $\bar{B} \rightarrow X_c \ell \bar{\nu}$ branching fraction has been measured in Ref.[59], finding $\mathcal{B}(B^+ \rightarrow X_c^0 \ell^+ \nu) = 10.8 \pm 0.4\%$ and $\mathcal{B}(B^0 \rightarrow X_c^- \ell^+ \nu) = 10.1 \pm 0.4\%$. The exclusive channels that make up $\bar{B} \rightarrow X_c \ell \bar{\nu}$, are known with differing levels of precision. The branching fractions of $\bar{B} \rightarrow D \ell \bar{\nu}$ and $\bar{B} \rightarrow D^* \ell \bar{\nu}$ are well known and combined represent approximately 73% of the inclusive rate. For the Belle analysis we adopt the world average branching fractions [211] of these channels, for the Belle II analysis we adopt the isospin-averaged branching fractions.

Semileptonic decays to the four orbitally excited charmed mesons (D_0^*, D_1', D_1, D_2^*), where the D_0 and D_1' are the broad $j_\ell^P = \frac{1}{2}^+$ doublet, and the D_1 and D_2^* are the narrow $j_\ell^P = \frac{3}{2}^+$ doublet, comprise an additional 16% of the inclusive rate. Hereafter, these states are collectively denoted as D^{**} . We adopt the world average branching fractions of Ref. [212]. The recent measurements by the Belle collaboration [213] of $\bar{B} \rightarrow D_1(\rightarrow D\pi^+\pi^-)\ell\bar{\nu}$, $\bar{B} \rightarrow D^{(*)}\pi\ell\bar{\nu}$ and $\bar{B} \rightarrow D^{(*)}\pi^+\pi^-\ell\bar{\nu}$ are not yet included in these averages, but are mostly compatible with them. The exception is the branching fraction of $\bar{B} \rightarrow D_0(\rightarrow D\pi)\ell\bar{\nu}$ which is found to be significantly smaller than the present world average. The branching fractions are corrected to account for missing known and isospin conjugated decay modes, following the prescription of Ref. [214] (see Ref. [73] for details). The exact structure of these channels has a negligible impact on the weak annihilation measurement; therefore, we fix the ratios of $\mathcal{B}(D_2 \rightarrow D^*\pi)$ to $\mathcal{B}(D_2 \rightarrow D\pi)$ and $\mathcal{B}(D_1 \rightarrow D^*\pi)$ to $\mathcal{B}(D_1 \rightarrow D\pi)$ in the Belle II measurement.

The decays $\bar{B} \rightarrow D_s K \ell \bar{\nu}$ and $\bar{B} \rightarrow D_s^* K \ell \bar{\nu}$ have been measured with a combined branching fraction of $(5.9 \pm 1.0) \times 10^{-4}$ [211, 215, 216]. We neglect them in the Belle measurement as no simulated samples of these modes are available. Due to the small branching fraction and the expected soft kinematic distribution of the lepton — a result of the four-body nature of the decay — the contribution of such events to the region of interest is expected to be negligible. In Belle II, these channels are included in the generic MC cocktail, we therefore include them at their nominal branching fractions¹.

The decay of B mesons to $D^{(*)}\pi\pi\ell\bar{\nu}$ has been observed by the BaBar collaboration [217], and, as mentioned above, recently by the Belle collaboration [213]. The Belle result is not considered in the following. Most of the observed $\bar{B} \rightarrow D\pi\pi\ell\bar{\nu}$ rate is covered by production via an intermediate D_1 : $\bar{B} \rightarrow D_1(\rightarrow D\pi\pi)\ell\bar{\nu}$. The remainder, and the $\bar{B} \rightarrow D^*\pi\pi\ell\bar{\nu}$ events, are assumed to be generated in equal quantity through two intermediate broad resonances with masses and widths identical to the broad D^{**} . Although there is currently no evidence for decays of the broad D^{**} to these final states, such a model, generating an initial three-body decay, is considered to provide a more physical description of the decay kinematics than a five-body model based on the equi-distribution of the final states in phase space (ie. PHSP within EvtGen).

The listed decays saturate approximately 92% of the nominal inclusive rate, with a deviation of approximately $1.4 - 1.8\sigma$. It is not clear what, if anything², fills the remaining gap between the sum of exclusive branching fractions and the inclusive branching fraction. Measurements of $\mathcal{B}(\bar{B} \rightarrow D^{(*)}\pi\ell\bar{\nu})$ [211] indicate that the rate is fully saturated by production via intermediate D^{**} resonances, leaving no room for nonresonant production. Therefore, we fill the gap with equal parts $\bar{B} \rightarrow D^*\eta\ell\bar{\nu}$ and $\bar{B} \rightarrow D\eta\ell\bar{\nu}$ decays, for which there is no evidence, but also few constraints. The events are again generated via two

¹To avoid adding an insignificant event category, we absorb $\bar{B} \rightarrow D_s^{(*)} K \ell \bar{\nu}$ decays into the $\bar{B} \rightarrow D^{**} \ell \bar{\nu}$ category in plots and discussions on sample compositions.

²Preliminary results from Belle II find a lower $\bar{B} \rightarrow X_c \ell \bar{\nu}$ branching fraction, $(9.75 \pm 0.03_{\text{stat}} \pm 0.47_{\text{syst}})\%$ [218], reducing the size of the gap.

facsimiles of the broad D^{**} resonances. In the Belle II analysis, a 100% gaussian uncertainty is associated to these channels, in Belle, the prior is taken to be uniform between zero and twice the nominal branching fractions. We categorise the $\bar{B} \rightarrow D^{(*)}\pi\pi\ell\bar{\nu}$ and $\bar{B} \rightarrow D^{(*)}\eta\ell\bar{\nu}$ events, collectively, as $\bar{B} \rightarrow D_{\text{gap}}^{**}\ell\bar{\nu}$.

Fits of the exclusive branching fractions to moments of the kinematic distributions of the inclusive measurement also do not provide a solution to the gap problem [101]. The fits find generally poor agreement between the exclusive descriptions and the inclusive measurements, with the fit results pulling the $\bar{B} \rightarrow D^*\ell\bar{\nu}$ branching fraction to 5 – 10% above its precisely measured value.

The assumed branching fractions of $\bar{B} \rightarrow X_c\ell\bar{\nu}$ decays are summarised in Table 7.1.

Decay Channel	Belle		Belle II	
	$B^+ [\times 10^{-3}]$	$B^0 [\times 10^{-3}]$	$B^+ [\times 10^{-3}]$	$B^0 [\times 10^{-3}]$
$\bar{B} \rightarrow X_c\ell\bar{\nu}$	108 ± 4.0	101 ± 4.0	108 ± 4.0	101 ± 4.0
$\bar{B} \rightarrow D\ell\bar{\nu}$	23.5 ± 0.9	23.1 ± 1.0	24.1 ± 0.7	22.4 ± 0.7
$\bar{B} \rightarrow D^*\ell\bar{\nu}$	56.6 ± 2.2	50.5 ± 1.4	55.0 ± 1.1	51.1 ± 1.1
$\bar{B} \rightarrow D_0(\rightarrow D\pi)\ell\bar{\nu}$	4.2 ± 0.8	3.9 ± 0.7	4.2 ± 0.8	3.9 ± 0.7
$\bar{B} \rightarrow D'_1(\rightarrow D^*\pi)\ell\bar{\nu}$	4.2 ± 0.8	3.9 ± 0.8	4.2 ± 0.9	3.9 ± 0.8
$\bar{B} \rightarrow D_1(\rightarrow D^*\pi)\ell\bar{\nu}$	4.2 ± 0.3	3.9 ± 0.3	6.6 ± 1.1	6.2 ± 1.0
$\bar{B} \rightarrow D_1(\rightarrow D\pi\pi)\ell\bar{\nu}$	2.4 ± 1.0	2.3 ± 0.9		
$\bar{B} \rightarrow D_2(\rightarrow D^*\pi)\ell\bar{\nu}$	1.2 ± 0.1	1.1 ± 0.1	2.9 ± 0.3	2.7 ± 0.3
$\bar{B} \rightarrow D_2(\rightarrow D\pi)\ell\bar{\nu}$	1.8 ± 0.2	1.7 ± 0.2		
$\bar{B} \rightarrow D_s K\ell\bar{\nu}$	-	-	0.3 ± 0.1	-
$\bar{B} \rightarrow D_s^* K\ell\bar{\nu}$	-	-	0.3 ± 0.2	-
$\bar{B} \rightarrow D\pi\pi\ell\bar{\nu}$	0.6 ± 0.6	0.6 ± 0.6	0.6 ± 0.9	0.6 ± 0.8
$\bar{B} \rightarrow D^*\pi\pi\ell\bar{\nu}$	2.2 ± 1.0	2.0 ± 1.0	2.2 ± 1.0	2.0 ± 1.0
$\bar{B} \rightarrow D\eta\ell\bar{\nu}$	3.6 ± 2.0	4.0 ± 2.0	3.8 ± 3.8	4.1 ± 4.1
$\bar{B} \rightarrow D^*\eta\ell\bar{\nu}$	3.6 ± 2.0	4.0 ± 2.0	3.8 ± 3.8	4.1 ± 4.1

Table 7.1: Assumed branching fractions and their uncertainties for $\bar{B} \rightarrow X_c\ell\bar{\nu}$ decays. See text for details.

7.1.2 Form factors

For the Belle analysis, the $\bar{B} \rightarrow D\ell\bar{\nu}$ decays are modelled using the parameterisation of Boyd, Grinstein, and Lebed (BGL) [15] with parameter central values and uncertainties taken from the fit in Ref. [48], truncated at $N = 3$. For $\bar{B} \rightarrow D^*\ell\bar{\nu}$ decays we use the implementation of BGL proposed by Refs. [219, 220] with parameter central values and uncertainties from the BGL(2,2,2) fit of Ref. [57]. Semileptonic decays to the D^{**} are modelled using heavy-quark symmetry based form factors as proposed by Leibovich, Ligeti, Stewart, and Wise (LLSW) [17, 221], using fitted parameter values of Ref. [214]

(approximation C).

For the Belle II analysis, the decays of $\bar{B} \rightarrow D\ell\bar{\nu}$ and $\bar{B} \rightarrow D^*\ell\bar{\nu}$ are modelled using the description of Bernlocher, Ligeti, Papucci, Prim, Robinson, and Xong [222], with parameter values given by the first column of Table 7 within Ref. [222]. Decays of $\bar{B} \rightarrow D^{**}\ell\bar{\nu}$ are modelled using the descriptions of Bernlocher, Ligeti, and Robinson [214, 223], which generalises the LLSW model to include new physics effects. The parameter values of Table 5 in Ref. [223] are adopted. The decays $\bar{B} \rightarrow D_s^{(*)}K\ell\bar{\nu}$ are modelled via equi-distribution in phase space.

7.2 Charmless semileptonic decays

Modelling of the charmless semileptonic decay $\bar{B} \rightarrow X_u\ell\bar{\nu}$ poses a significant challenge. The inclusive models discussed in Ch. 2 predict the triple differential decay rate in lepton energy, momentum transfer, and hadronic mass, but do not provide any description of the makeup of the hadronic system. To simulate such events, we therefore follow a two-step process. The kinematics of the lepton, neutrino, and hadronic system are generated following the inclusive theory model, then the hadronic system is hadronised via PYTHIA. The generation requires a minimum hadronic mass of 280 MeV to allow PYTHIA to produce at least two pions.

The inclusive models predict smooth distributions; however, specific resonance decays are well measured and contribute significant fractions of the total rate, while being localised at specific hadronic masses. We therefore use the *hybrid* approach of Ref. [224] as used by Babar [72], recent Belle $\bar{B} \rightarrow X_u\ell\bar{\nu}$ and $B^- \rightarrow \mu\bar{\nu}$ studies [73, 79, 225], and is widely used in Belle II studies of charmless semileptonic decays. In the hybrid model, the inclusive predictions are considered to be correct in broad phase-space bins, and resonances are taken to be finer scale corrections. This is achieved by adding generated resonance events, corresponding to the decays $\bar{B} \rightarrow \pi\ell\bar{\nu}$, $\bar{B} \rightarrow \rho\ell\bar{\nu}$, $\bar{B} \rightarrow \eta\ell\bar{\nu}$, $\bar{B} \rightarrow \omega\ell\bar{\nu}$, $\bar{B} \rightarrow \eta'\ell\bar{\nu}$, and scaling down the contribution of inclusively generated events within the same broad phase-space bin such that the sum matches the prediction of the inclusive model. The semileptonic decays of B -mesons to heavier unflavoured resonances, $f_0, a_0, h_1, b_1, a_1, f_2, \dots$, are poorly known, we therefore absorb these channels into the inclusive rate.

The weights applied to the inclusive sample are given by:

$$w_{ijk} = \frac{\Delta\mathcal{B}_{ijk}^{\text{inc}} - \Delta\mathcal{B}_{ijk}^{\text{exc}}}{\Delta\mathcal{B}_{ijk}^{\text{inc}}}, \quad (7.1)$$

where $\Delta\mathcal{B}_{ijk}^{\text{exc}}$ and $\Delta\mathcal{B}_{ijk}^{\text{inc}}$ are the partial branching fraction of the exclusive (resonant) and inclusive components in a phase-space bin, and the indices i, j, k define the three dimensional phase-space bin³, running over the one-dimensional bins in q^2, E_ℓ^B, M_X , respectively,

³As the resonances occur at low M_X , where the inclusive rates do not predict significant contributions, one very broad bin is used to contain all the resonances. In practice the hybridisation process is therefore essentially two-dimensional in q^2 and E_ℓ^B , with only a small tail of exclusive events leaking into the next M_X bin. We nevertheless refer to this as three-dimensional due to convention.

defined by the bin boundaries:

$$\begin{aligned} q^2 &: [0, 2.5, 5, 7.5, 10, 12.5, 15, 20, 25] \text{ GeV}^2, \\ E_\ell^B &: [0, 0.5, 1, 1.25, 1.5, 1.75, 2, 2.25, 3] \text{ GeV}, \text{ and} \\ M_X &: [0, 1.4, 1.6, 1.8, 2, 2.5, 3, 3.5] \text{ GeV}. \end{aligned}$$

The one-dimensional projections of the hybrid MC in these three variables, using the branching fractions and modelling for the inclusive and resonant components as used in the Belle II analysis, are presented in Fig 7.1.

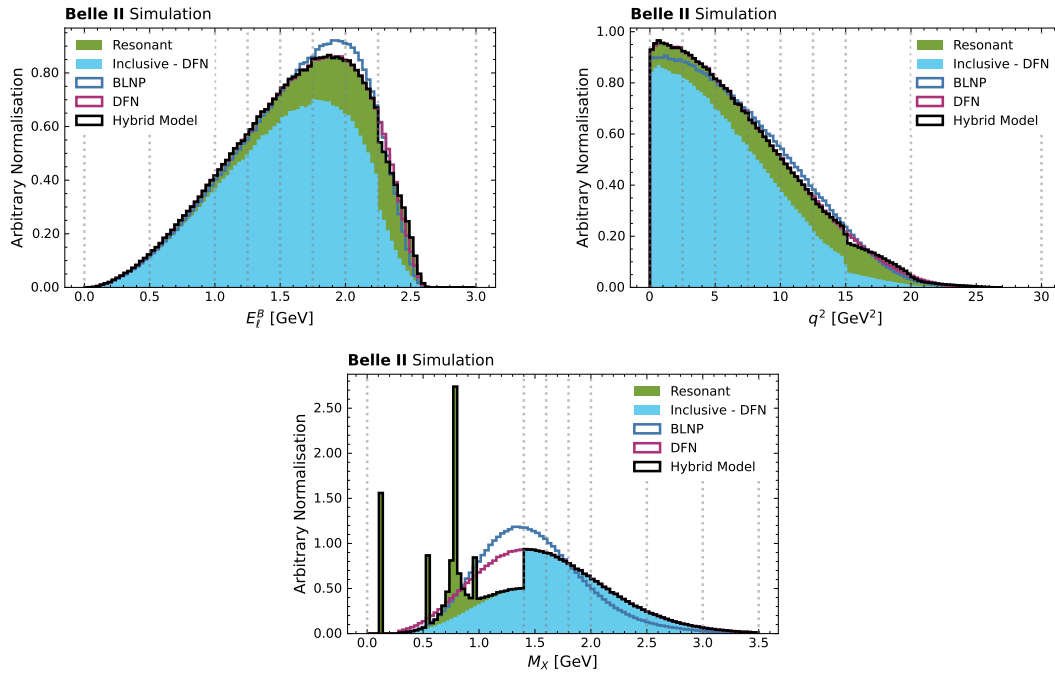


Figure 7.1: Projections of the hybrid model of semileptonic charmless decays of B^+ , as used for the Belle II analysis, in M_X , E_ℓ^B and q^2 . The DFN model is used for the hybrid scheme.

For the inclusive prediction, we adopt the model of De Fazio and Neubert [37] (DFN) in both analyses. For Belle, the choice of model is preordained, since only DFN is supported by the `EvtGen` version. For Belle II, DFN is adopted over the model of Bosch, Lange, Neubert and Paz [38, 39] (BLNP), which is supported by the newer `EvtGen` version, as DFN has demonstrated better agreement to data at the lepton-momentum endpoint [74], where weak annihilation events are expected to be located. The difference in shape between the predictions of DFN and BLNP is considered as a systematic uncertainty in both analyses, using separately generated events in Belle II and a reweighting technique for Belle. As discussed in Sec. 2.2.2.2, the free parameters of DFN are the b quark mass in the Kagan-Neubert scheme [226], $M_b^{\text{KN}} = (4.66 \pm 0.04) \text{ GeV}$ and a non-perturbative

parameter $a^{\text{KN}} = 1.3 \pm 0.5$. At leading order, a^{KN} is related to the average momentum squared of the b quark inside the B meson and determines the second moment of the shape function. It can be expressed as $a^{\text{KN}} = 3\bar{\Lambda}^2/\mu_\pi^{2,\text{KN}} - 1$, where $\Lambda \equiv M_B - M_b^{\text{KN}}$, is the binding energy, and $\mu_\pi^{2,\text{KN}}$ the kinetic energy parameter. The values of Λ and $\mu_\pi^{2,\text{KN}}$, and hence M_b^{KN} and a , were determined in Ref. [227] from a fit to moments of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ distribution in kinematic variables and the photon energy spectrum of $\bar{B} \rightarrow X_s \gamma$ decays.

For the Belle measurement, we adopt values of the semileptonic charmless branching fractions of charged and neutral B decays derived from the world averaged measurement of the mixture of both charges, $\mathcal{B}(\bar{B} \rightarrow X \ell \bar{\nu}) = (2.13 \pm 0.31) \times 10^{-3}$ [211], via

$$\mathcal{B}(B^+ \rightarrow X_u^0 \ell^+ \nu) = \tau_{+/0} \frac{1 + f_{+/0}}{1 + \tau_{+/0} f_{+/0}} \mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu}), \quad (7.2)$$

$$\mathcal{B}(B^0 \rightarrow X_u^- \ell^+ \nu) = \frac{1 + f_{+/0}}{1 + \tau_{+/0} f_{+/0}} \mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu}), \quad (7.3)$$

where $f_{+/0} = 1.058 \pm 0.024$ is the ratio of $\Upsilon(4S)$ decays to charged over neutral B pairs and $\tau_{+/0} = 1.076 \pm 0.004$ is the ratio of the lifetime of charged over neutral B mesons [211]. For the Belle II measurement, we adopt directly the measured branching fractions of Ref. [73]. For the considered resonant contributions, we adopt the world average branching fractions of Ref. [211], which are still current [1]. All assumed branching fractions are presented in Table 7.2.

Decay Channel	Belle		Belle II	
	$B^+ [\times 10^{-3}]$	$B^0 [\times 10^{-3}]$	$B^+ [\times 10^{-3}]$	$B^0 [\times 10^{-3}]$
$\bar{B} \rightarrow X_u \ell \bar{\nu}$	2.21 ± 0.31	2.05 ± 0.29	1.92 ± 0.24	1.76 ± 0.22
$\bar{B} \rightarrow \pi \ell \bar{\nu}$	0.078 ± 0.003	0.150 ± 0.006	0.078 ± 0.003	0.150 ± 0.006
$\bar{B} \rightarrow \rho \ell \bar{\nu}$	0.158 ± 0.011	0.294 ± 0.021	0.158 ± 0.011	0.294 ± 0.021
$\bar{B} \rightarrow \omega \ell \bar{\nu}$	0.119 ± 0.009	-	0.119 ± 0.009	-
$\bar{B} \rightarrow \eta \ell \bar{\nu}$	0.039 ± 0.005	-	0.039 ± 0.005	-
$\bar{B} \rightarrow \eta' \ell \bar{\nu}$	0.023 ± 0.008	-	0.023 ± 0.008	-
$\bar{B} \rightarrow x_u \ell \bar{\nu}$	1.79 ± 0.32	1.60 ± 0.30	1.50 ± 0.24	1.32 ± 0.22

Table 7.2: Assumed branching fractions and their uncertainties for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays. See text for details. Non-resonant $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays are denoted $\bar{B} \rightarrow x_u \ell \bar{\nu}$.

The decays $\bar{B} \rightarrow \pi \ell \bar{\nu}$, $\bar{B} \rightarrow \rho \ell \bar{\nu}$, $\bar{B} \rightarrow \omega \ell \bar{\nu}$ are modelled via the expansion of Bourrely, Caprini, and Lellouch (BCL) [14, 228]. For $\bar{B} \rightarrow \pi \ell \bar{\nu}$ we adopt the form factor parameter central values and uncertainties from the global fit of Ref. [229] for the Belle measurement and the central values of Table 57 of Ref. [78] for the Belle II measurement. For $\bar{B} \rightarrow \rho \ell \bar{\nu}$ and $\bar{B} \rightarrow \omega \ell \bar{\nu}$ decays we adopt the values of Ref. [228] for the Belle measurement and Ref. [230] for the Belle II measurement. The decays $\bar{B} \rightarrow \eta \ell \bar{\nu}$ and $\bar{B} \rightarrow \eta' \ell \bar{\nu}$ are modelled via the light-cone sum rule predictions of Ref. [231] in both analyses.

7.3 Weak annihilation

As discussed in Ch. 2, very little is known about the shape of the weak annihilation contribution: it is expected to be localised at the momentum transfer endpoint, smearing about $\delta(q^2 - M_b^2)$, and the hadronic system is expected to be light [88]. The shape of the smearing, and composition of the hadronic system are not known. For the search for weak annihilation at Belle II, we therefore consider a simplified model that captures the known features with sufficient flexibility to test variations of the model within the expected regime.

The decay, which we also label $\bar{B} \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$, is taken to correspond approximately to a two-body decay ($\bar{B} \rightarrow \ell \bar{\nu}$) with the hadronic system acting as a correction. Within the generator (EvtGen) this is achieved by means of nested two-body decays:

$$\bar{B} \rightarrow X_u W^*, \quad (7.4)$$

$$W^* \rightarrow \ell \bar{\nu}, \quad (7.5)$$

where $M_X \lll M_{W^*} \equiv M_{\ell \bar{\nu}} \lesssim M_B$, such that almost all energy of the B is transferred to the lepton neutrino system. The W^* is assigned a spin of 0 such that its leptonic decay matches the vector axial-vector structure of a leptonic B decay.

Only limited constraints exist on the composition of the hadronic system. As the form-factor based descriptions of the resonant $\bar{B} \rightarrow X_u \ell \bar{\nu}$ final states include all events with the given Lorentz structure, they include weak annihilation events. The measured $\bar{B} \rightarrow \pi \ell \bar{\nu}$ q^2 spectra can thus constrain the single-pion final state. Approximately 17.9% of the total $\bar{B} \rightarrow \pi \ell \bar{\nu}$ rate is found at $q^2 > 18 \text{ GeV}^2$ [66]. Attributing the full rate above this threshold to weak annihilation corresponds to a weak annihilation contribution to just the single-pion final state of 0.8% or 1.4% of the total charmless semileptonic rate for B^0 and B^+ decays, respectively. As the ‘standard’ non-weak annihilation decay will also contribute within this high q^2 regime, and is naively expected to dominate, these estimates represent loose upper bounds. For other resonant states, the q^2 spectra are not known with sufficient precision to make even such statements.

It should be noted that although the descriptions of the resonant $\bar{B} \rightarrow X_u \ell \bar{\nu}$ channels include the weak annihilation contribution, the modelling of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays within this analysis is nominally free of this contribution. The resonant channels are combined with the inclusive channels through the hybrid modelling procedure. In this process, the contribution of resonant events which originate from weak annihilation processes is absorbed into the shape predicted by the inclusive modelling. Thus, as weak annihilation is not included in the inclusive models, a consequence of the hybrid modelling is that the modelled $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events are free of weak annihilation contributions.

The Belle collaboration has also published a study of the $B^+ \rightarrow \pi^+ \pi^- \ell^+ \nu$ q^2 spectrum [232]. This study does not require that the $\pi^+ \pi^-$ system corresponds to a resonance and finds a partial branching fraction of $(2.59_{-0.49}^{+0.55}_{\text{stat}} \pm 0.47_{\text{syst}}) \times 10^{-5}$ for $17 < q^2 < 25 \text{ GeV}^2$. This corresponds to 1.34% of the nominal $B^+ \rightarrow X_u^0 \ell^+ \nu$ rate. Accounting for contributions

from $B^+ \rightarrow \rho^0 \ell^+ \bar{\nu}$ and non-resonant $\bar{B} \rightarrow X_u \ell \bar{\nu}$, only limited room remains for a weak annihilation contribution to this final state.

For lack of a better model we hadronise the hadronic system, X_u , via PYTHIA. The hadronisation implicitly requires at least two pions in the final state. Although the rate is expected to be limited, no constraint is made on the dipion final state, as such a constraint would require splitting the weak annihilation contribution into several components. Moreover, we do not add single-pion events as the relative contribution of such final states is unknown, and including this component would introduce a further degree of freedom. As the analysis procedure has been designed to have low sensitivity to the composition of the hadronic system, these simplifying assumptions are expected to have negligible impact.

The model has four free parameters, the masses and widths of the hadronic system X_u (M_X, Γ_X), and of the intermediate W^* boson ($M_{\ell\bar{\nu}}, \Gamma_{\ell\bar{\nu}}$). By varying the value of these parameters, especially those of the W^* , we generate a range of shapes that match the expected features of the weak annihilation contribution:

- weak annihilation is concentrated at the lepton endpoint,

$$\langle E_\ell^B \rangle = M_b/2 \pm 200 \text{ MeV}, \quad (7.6)$$

$$\langle q^2 \rangle = (M_b \pm 200 \text{ MeV})^2; \quad \text{and} \quad (7.7)$$

- the hadronic system is soft, where $M_X \sim \mathcal{O}(\Lambda_{\text{QCD}})$.

We consider 70 variations, corresponding to all possible pairwise combinations of ten values of $M_{\ell\bar{\nu}}$, spaced equidistantly between 4.3^2 and 5.0^2 GeV^2 , and seven values of $\Gamma_{\ell\bar{\nu}}$ at geometric intervals from 0.002 to 1 GeV. Values are selected to generate a wide range of models, covering from the expected contribution, a peak near $q^2 \sim M_b^2$, to extreme variations, both in being centred far from $q^2 \sim M_b^2$ and being extremely broad. The kinematic distributions of these models is given in Fig. 7.2.

Although this model allows us to generate a large set of models broadly matching the expected features of weak annihilation, it is important to acknowledge the limitations imposed by its simplified nature and whether the measurement is sensitive to them.

- It is not feasible to scan over the four free parameters of the model. We therefore consider variations to M_X and Γ_X as systematic uncertainties fixing them to $M_X = 420_{-130}^{+140} \text{ MeV}$ and $\Gamma_X = 280 \pm 140 \text{ MeV}$.⁴ This is consistent with the expectation that the hadronic system is soft.
- We couple the weak annihilation contribution across the charged and neutral B channels. That is, we always consider weak annihilation decays in both modes with common model parameters ($M_{\ell\bar{\nu}}, \Gamma_{\ell\bar{\nu}}, M_X, \Gamma_X$). The model thus implicitly assumes

⁴For technical reasons we consider 290 MeV as the lower variation for M_X . We require a minimum M_X of 280 MeV for PYTHIA to hadronise X_u into at least two pions. If the nominal value of M_X is equal to the minimum value, the event generation fails, necessitating a slight shift to 290 MeV.

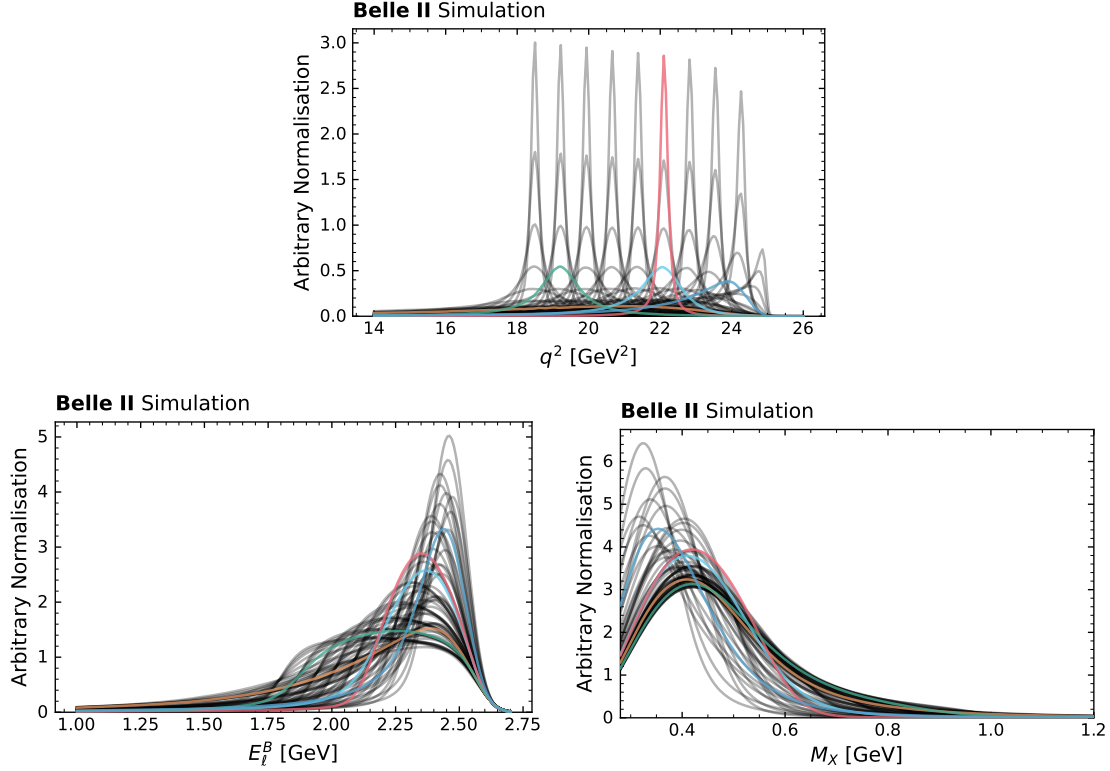


Figure 7.2: Kinematic distributions of the 70 considered weak annihilation models in q^2 , $E_\ell^{B_{\text{sig}}}$, and M_X . Five example distributions have been coloured to demonstrate the relations between the distributions.

that the valent and non-valent weak annihilation contributions have the same distribution. If this assumption is violated, it would enter by cross-feed of the B tagging algorithm, mistagging a charged B as neutral or vice versa, mixing the charged and neutral samples. As the cross-feed rate is low, the impact of such a violation is suppressed. Thus, the effect is expected to be small if the weak annihilation contributions are of similar magnitude across both channels.

- Given the nested two-body nature of the model, the distributions in E_ℓ^B can best be described as a plateau about $M_{\ell\bar{\nu}}$ when $M_{\ell\bar{\nu}}$ is significantly below M_B . Although this may be somewhat unphysical for what is nominally at minimum a three-body decay, we design the analysis to be insensitive to the exact distribution in lepton energy; selecting events with high-energy leptons with a loose requirement, $E_\ell^B > 1.5 \text{ GeV}$, far from the core of the distributions. For extremely broad states, there is a long tail towards lower E_ℓ^B values, causing up to percent level losses in signal efficiency through this selection. The sensitivity to such broad states is low; therefore, the loss is considered negligible. We also note that in this high energy regime the performance of the Belle II lepton identification systems is known to be excellent

and approximately uniform. Therefore, a change in distribution within this regime would not cause significant changes in reconstruction efficiency.

- As the lepton-neutrino system in the model arises from the decay of the W^* , the distribution of weak annihilation events in q^2 follows a non-relativistic Breit–Wigner distribution, as adopted for the mass of the W^* . This introduces a further model dependence; however, as the q^2 resolution is poor, $\mathcal{O}(1 \text{ GeV}^2)$, the measurement is not sensitive to the precise shape of the signal peak.
- A smooth distribution is assumed for the hadronic mass, with no resonant contributions. As the hadronic system is not explicitly reconstructed, the measurement is expected to have low sensitivity to the exact composition of the hadronic system.

7.3.1 Leptonic decays

Leptonic B decays, $B^- \rightarrow e^- \bar{\nu}(\gamma)$, $B^- \rightarrow \mu^- \bar{\nu}(\gamma)$ share a signature similar to the signal weak annihilation decays in the Belle II study⁵, a high-momentum lepton and a lepton and neutrino system with $q^2 \sim M_B^2$. They differ from the signal decay by the lack of a hadronic system; however, as we are performing an inclusive analysis, we make only minimal requirements on the hadronic system, and thus these events are not discriminated against.

The purely leptonic channels $B^- \rightarrow (e^-, \mu^-) \bar{\nu}$ are helicity suppressed with expected branching fractions [10] of

$$\mathcal{B}(B^- \rightarrow e^- \bar{\nu}) = (8.75_{-0.29}^{+0.41}) \times 10^{-12}, \quad (7.8)$$

$$\mathcal{B}(B^- \rightarrow \mu^- \bar{\nu}) = (3.75_{-0.15}^{+0.14}) \times 10^{-7}. \quad (7.9)$$

We thus expect only a negligible contribution of events from the muonic channel, equivalent to approximately 0.02% of the inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate, a factor 50 smaller than the theoretical limits of the weak annihilation rate. The addition of a photon to the final states, $\mathcal{B}(B^- \rightarrow \ell^- \bar{\nu} \gamma)$ lifts the helicity suppression, leading to more substantial contributions to the signal region ($\mathcal{O}(10^{-6})$). We simulate muonic decays following the V-A coupling of the charged weak current, and radiative events via the model of Ref. [233] and adopt the central values and uncertainties of the most recent measurements of all three channels [79, 234]:

$$\mathcal{B}(B^- \rightarrow \mu^- \bar{\nu}) = (5.3 \pm 2.12) \times 10^{-7}, \quad (7.10)$$

$$\mathcal{B}(B^- \rightarrow \mu^- \bar{\nu} \gamma) = (1.0 \pm 1.3) \times 10^{-6}, \quad (7.11)$$

$$\mathcal{B}(B^- \rightarrow e^- \bar{\nu} \gamma) = (1.7 \pm 1.7) \times 10^{-6}. \quad (7.12)$$

Given their small branching fractions, we neglect these contributions for the measurement of $|V_{ub}|/|V_{cb}|$ at Belle.

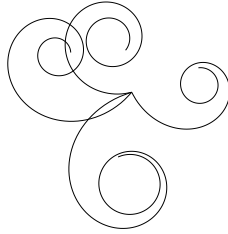
⁵One could also refer to these decays as weak annihilation, however as they do not produce a hadronic system we do not include them when referring to weak annihilation and instead refer to these channels as leptonic decays.

7.4 Remaining categories

The selections applied in the analyses are studied and optimised on simulated samples, consisting of events from the respective ‘generic’ simulated samples which contain thousands of possible decays and, where needed, supplemented with additional simulated samples of the decay channels discussed above. For these additional samples, one of the pair of produced B -mesons is set to decay via the specific channel, the other decaying generically.

The respective event reconstruction of both analyses begins by identifying a lepton candidate and combining it with a hadronic B meson, reconstructed by the Full Event Interpretation algorithm, the hadronic tagging algorithm introduced in Sec. 4.2. For most of the following discussions, events are categorised on the basis of the lepton candidate. If the lepton originated in any of the channels discussed within this chapter, the event is labelled as such. In addition to these channels, the reconstruction procedure will reconstruct events from several additional categories.

- Secondary leptons, where a candidate lepton has been correctly identified as the respective lepton but did not originate from a semileptonic B decay. Originating instead, in a secondary decay, such as semileptonic D decays or the decay of τ or J/ψ particles. We note that this category includes tauonic ($\bar{B} \rightarrow \tau \bar{\nu}$) and semitauonic events ($\bar{B} \rightarrow X \tau \bar{\nu}$) if the tauon decays leptonically ($\tau \rightarrow \ell \nu \bar{\nu}$).
- Fake leptons, where a hadron has been misidentified as a lepton. The hadron can originate directly from the decay of the B or a secondary decay.
- Continuum events, $e^+e^- \rightarrow q\bar{q}$, $q = \{u, d, s, c\}$. We group all reconstructed candidates originating in such decays, regardless of whether the reconstructed lepton corresponds to a true lepton.



Part III-A

Measurement of inclusive $|V_{ub}|/|V_{cb}|$ at Belle

Event reconstruction and selection

We discuss how the basic physics objects observed in the detector, tracks and calorimeter clusters, are combined to reconstruct semileptonic B decays, and the selections imposed on these events to purify the sample.

Our aim is to measure the ratio of partial branching fractions, $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$, from which one can extract $|V_{ub}|/|V_{cb}|$. We begin by reconstructing $\bar{B} \rightarrow X \ell \bar{\nu}$ events by searching for well-identified leptons and then attempt to construct a B , that has decayed hadronically, from the remaining particles. After reconstruction, we impose a set of selection criteria on quantities that can be correlated to the quality of the event reconstruction and the particle contents of the event. The selections first aim to purify the semileptonic sample and then to further select $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events from within the purified sample.

8.1 Samples

The analysis is performed on all the data collected by the Belle experiment at the $\Upsilon(4S)$ resonance, 711 fb^{-1} [235], corresponding to approximately 770 million $B\bar{B}$ pairs. An additional 89.4 fb^{-1} collected 60 MeV below the $\Upsilon(4S)$ resonance is used to improve the modelling of the continuum background.

The simulated samples correspond to approximately ten times the expected yield for $B\bar{B}$ processes, and six times the expected yield for continuum processes. Corrections to these samples are described in Sections. 8.4 and 9.5.

8.2 Event reconstruction

The data and MC undergoes an initial selection to select hadronic events, known in Belle as the **HadronBJ** skim [111]. Events must have at least three tracks that originate near the interaction point, and at least 20% of the centre of mass energy must be observed. Additional selections on the amount and shape of energy deposited in the ECL and the mass of the heaviest jet in the event reject tau pair and QED processes, while maintaining high efficiency on events containing B , J/ψ and $\psi(2S)$ mesons.

8.2.1 Lepton

Leptons are identified by their likelihood ratio. We require $P_\mu > 0.9$ for a track to be considered a muon and $P_e > 0.8$ and $P_\mu < 0.9$ to consider it an electron, where the second requirement ensures an orthogonal selection to the muon sample. We additionally require that signal lepton candidates, i.e. those expected to originate in the semileptonic B decay, pass near the interaction point, $dr < 2$ cm and $|dz| < 4$ cm, and pass through the barrel of the detector, where the particle identification performance is highest and best understood, corresponding to an angular acceptance of $\theta \in (35^\circ, 125^\circ)$ to reach the ECL and $\theta \in (25^\circ, 145^\circ)$ to reach the KLM for electrons and muons, respectively. As leptons originating from semileptonic decays, especially charmless semileptonic decays, are expected to be energetic, we additionally require a centre-of-mass (c.m.) momentum exceeding 0.5 GeV.

Electrons, due to their low mass have a high probability of emitting final-state radiation, known as bremsstrahlung, reducing their reconstructed momenta. The probability of emitting final-state radiation reduces rapidly with the energy of the emitted photon, and the number of photons. To correct for this effect we search for photon candidates with an energy less than 1 GeV within a cone of 5° around the initial momentum direction of the electron track. The candidate with the highest energy is taken as radiative, and its momentum re-summed to the electron candidate.

8.2.2 Tag

Having selected a lepton, all remaining particles in the event are used by the Full Event Interpretation algorithm (Sec. 4.2) to construct hadronic B candidates, determining their charge, flavour, and four-momentum. During reconstruction we perform a loose preselection to reduce computational expense. To select reconstructed tags consistent with a B meson, we require that the candidates have a beam-energy constrained mass of

$$M_{bc} = \sqrt{E_{\text{beam}}^2 - |\vec{p}_{B_{\text{tag}}}^*|^2} > 5.24 \text{ GeV}, \quad (8.1)$$

where $E_{\text{beam}} = \sqrt{s}/2$ denotes half the c.m. energy and $\vec{p}_{B_{\text{tag}}}^*$ is the three momentum of the B_{tag} candidate in the c.m. frame. This condition is relaxed to $M_{bc} > 5.21$ GeV when analysing off-resonance data, corresponding to half the shift in centre-of-mass energy. We further consider the energy difference from nominal, ΔE , given by

$$\Delta E = E_{B_{\text{tag}}}^* - E_{\text{beam}}, \quad (8.2)$$

where $E_{B_{\text{tag}}}^*$ is the energy of the B_{tag} in the c.m. frame, and we impose the requirement, $|\Delta E| < 0.2$ GeV. Finally, each candidate B_{tag} is associated with a final FEI classifier score, $\mathcal{P}_{\text{FEI}}$. We require that $\mathcal{P}_{\text{FEI}} > 0.001$. Due to combinatorics, a large number of B candidates can be reconstructed, if more than seven candidates pass these selections, we keep only the seven with the highest $\mathcal{P}_{\text{FEI}}$.

8.2.3 Hadronic system

All remaining tracks and clusters, other than those which have been identified as the signal lepton or are used to reconstruct the tag B , are assigned to the *rest of event*. Summing the four vectors of these particles provides an inclusive reconstruction of the hadronic system X . In addition, studying the composition of these particles can help identify and suppress background processes. We impose additional loose selections on the particles assigned to the rest of event in order to suppress beam background processes and remove spurious tracks. This improves the resolution of quantities associated with the X system.

Photon candidates are reconstructed from energy depositions in the ECL that have not been matched to a track. To veto beam-background-induced photons, we discard photons that do not meet minimum energies of 100, 50, 150 MeV when found in the forward end-cap ($12.4^\circ \leq \theta < 31.4^\circ$), barrel ($32.2^\circ \leq \theta < 128.7^\circ$), and backward end-cap ($130.7^\circ \leq \theta < 155.1^\circ$) of the ECL, respectively.

For tracks, we require $|\vec{p}^*| < 3.2 \text{ GeV}$ to veto tracks outside the kinematic regime of tracks produced in B decays, where $\vec{p}^*$ is the c.m. frame three-momentum. We also enforce a transverse momentum dependent selection on the point of closest approach to the interaction point:

$$\begin{aligned} p_t < 0.25 \text{ GeV} : |dr| < 20 \text{ cm}, |dz| < 100 \text{ cm}, \\ 0.25 \leq p_t < 0.50 \text{ GeV} : |dr| < 15 \text{ cm}, |dz| < 50 \text{ cm}, \\ p_t \geq 0.50 \text{ GeV} : |dr| < 10 \text{ cm}, |dz| < 20 \text{ cm}. \end{aligned}$$

These requirements are loose to maintain high acceptance of pions originating from displaced K_S^0 decays. Both the photon and impact parameter selections are standard selections used in inclusive analyses at Belle [29, 236, 237].

Low momentum tracks can curl within the tracking detectors and be reconstructed as multiple tracks causing non-zero net event charges. We therefore use a gradient boosted decision tree [178] based algorithm to compare pairs of tracks and identify duplicates. The algorithm has been developed for this analysis and is discussed in detail in Ref. [238] (the authors' master's thesis). The variables considered are error-weighted differences of the five track helix parameters, the product of the charges of the track, the angle between the tracks, and error-weighted differences in the transverse and longitudinal momenta. If multiple tracks are identified to have originated from the same particle, they are ranked by $25|dr|^2 + |dz|^2$ with all but the lowest value track removed. The factor of 25 approximately corresponds to the squared difference in resolution of the impact parameters [239]. The classifier is trained separately for the early and late SVD geometry due to the significant changes in tracking hardware and software.

The existence of kaons in an event is indicative of a $b \rightarrow c \rightarrow s$ cascade decay. Kaons in the rest of event can thus be used to tag a charmed decay with high probability. We identify charged kaons based on binary likelihood ratios, requiring $P_{K/\pi} > 0.6$, $P_{K/e} > 0.2$, and

$P_{K/p} > 0.2$ to identify a track as a kaon¹. The first selection dominates the identification, with the second and third vetoing electron and proton to kaon fakes, which occur at small rates, respectively. If a track is identified as a kaon we assign it the kaon mass hypothesis when calculating the kinematic quantities of interest, improving the resolution of several of these quantities.

Candidate K_S^0 -mesons are reconstructed from pairs of charged pions in the rest of event, requiring that the invariant mass of the dipion system be approximately the K_S^0 mass, $468 \text{ MeV} < M_{\pi\pi} < 528 \text{ MeV}$. As a large number of combinatorial candidates are reconstructed, we further purify the sample with a multivariate NeuroBayes classifier [240]. Two sets of variables are used to identify a displaced vertex and to veto misidentified $\Lambda \rightarrow p\pi$ events [241]. The first includes the K_S^0 candidate momentum, flight length, angle between its momentum and flight direction, the distance between the helices of the two tracks in the z direction, point of closest approach of each track to the interaction point in the radial direction, the presence of hits recorded in the SVD and number of hits in the axial and stereo wires in the CDC for each track, and angle between the K_S^0 momentum in the lab frame to the pion momentum in the K_S^0 frame. The second relies on the binary likelihood ratio $P_{p/\pi}$ for each track, the magnitude of the momentum and the polar angle of each track, and the invariant mass of the candidate K_S^0 if the proton and pion mass hypotheses are used for the tracks. The candidates selected by the classifier all lie well inside the selected mass window.

Large differences have been observed in the performance of K_L^0 reconstruction between data and simulation. Candidates reconstructed using only KLM or KLM and ECL information are twice or three times as likely to be found in simulation than in data, respectively [242, 243]. This difference in performance is thought to originate from poor modelling of K_L^0 nuclear interactions in the KLM. Therefore, we make no attempt to reconstruct K_L^0 particles.

To further improve the resolution of quantities associated with the hadronic system, we search for proton candidates in the rest of event. The production of baryons from D mesons is kinematically forbidden; therefore, protons must arise either from baryonic B decays, which have been incorrectly reconstructed as semileptonic decays, or from baryonic decays on the tag side, which have been incorrectly reconstructed by the FEI in a nonbaryonic mode (we do not reconstruct baryonic modes in this analysis). For $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events, protons can also arise in the hadronic system, however due to their rarity ($\mathcal{B}(B^+ \rightarrow p \bar{p} \ell \bar{\nu}) = 5.8 \pm_{2.3}^{2.6} \times 10^{-6}$ [244]), almost all identified protons are expected to originate from the tag side. Baryonic B decay modes are poorly understood and poorly modelled. Inclusive measurements of proton production in the decay of the B^+/B^0 admixture by the ARGUS and CLEO collaborations find an average $p/\bar{p}$ multiplicity of $8 \pm 0.4\%$. Only a fraction of the decays that contribute to this multiplicity have been measured [1]. In

¹For kaons and protons, we additionally require that they fail the lepton identification requirements, ie. $P_e < 0.8$ and $P_\mu < 0.9$, however as we later reject all events in which additional leptons were found these selections are redundant.

the Belle MC, baryonic decays originate, almost entirely in the $\sim 36\%$ of B meson decays that are modelled via allowing PYTHIA to hadronise the quark content. As there may be a large disagreement in the proton production rate between the data and simulation, and the misidentification of protons as pions causes a shift in the missing mass of the event, an important variable for the analysis, we identify protons via the requirement $P_{p/K} > 0.6$, $P_{p/\pi} > 0.4$, and $p < 1.2$ GeV.

All remaining tracks are assigned a pion mass hypothesis.

8.2.4 Kinematic variables of interest

We define the lepton energy in the signal B meson rest frame as, $E_\ell^{B_{\text{sig}}} \equiv |\vec{p}_\ell^{B_{\text{sig}}}|$, neglecting the small contribution of the lepton mass to the energy, and inferring the signal B momentum from the tag B . In the c.m. frame this is given by

$$\vec{p}_{B_{\text{sig}}}^* = -\vec{p}_{B_{\text{tag}}}^*, \quad (8.3)$$

where $p_{B_{\text{tag}}}^*$ is the four vector of the B_{tag} in the c.m. frame. The four-momentum of the X system is calculated as

$$p_X = \sum_{i \in \{\text{tracks, photons}\}} (\sqrt{M_i^2 + |\vec{p}_i|^2}, \vec{p}_i) \quad (8.4)$$

where M_i is the nominal mass of the particle hypothesis assigned to track i , (one of M_π , M_K , or M_p), or zero for photons. The four-momentum of the neutrino is inferred to be the missing four-vector in the event, defined as:

$$p_{\text{Miss}} = p_{e^+e^-} - p_{B_{\text{tag}}} - p_\ell - p_X. \quad (8.5)$$

where $p_{e^+e^-}$, $p_{B_{\text{tag}}}$ and p_ℓ , are the four-vectors of the c.m. system, tag B , and lepton, respectively. We additionally calculate the momentum transfer squared as

$$q^2 = ((|\vec{p}_{\text{Miss}}|, \vec{p}_{\text{Miss}}) + p_\ell)^2, \quad (8.6)$$

where $\vec{p}_{\text{Miss}}$ is the missing momentum three-vector. This definition, which constrains the missing energy to the magnitude of the missing momentum, is adopted because the resolution on missing energy is poor compared to the resolution of missing momentum. Such a constraint is consistent with a true semileptonic event with only one missing neutrino. This definition is found to have a resolution, calculated as the standard deviation of the residuals, 16% smaller than a definition that also considers the missing energy for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events. Additionally, this definition is insensitive to incorrect mass assignment for tracks.

Finally, we consider the missing mass squared of the event reconstructed using the full X system defined as

$$M_{\text{Miss}}^2 = (p_{e^+e^-} - p_{B_{\text{tag, cons.}}} - p_\ell - p_X)^2, \quad (8.7)$$

where we constrain the energy of the B_{tag} in the c.m. frame to half the c.m. energy, marginally improving the resolution.

8.3 Event selection

To further purify events, we apply a sequence of selection criteria. We first require that the net event charge is consistent with zero, $Q_{B_{\text{tag}}} + Q_{\ell} + Q_X = 0$ where $Q_{B_{\text{tag}}}$, Q_{ℓ} , Q_X are the charge of the tag B , signal lepton and sum of charges in the rest of event, respectively. This requirement removes candidates with missing charged particles, improving the resolution of q^2 and M_{Miss}^2 .

Leptons originating from the secondary decay of a charm, $c \rightarrow s\ell\bar{\nu}$, will have charges opposite to leptons that originate from semileptonic B decays. To suppress candidate events in which a secondary lepton has been selected we require that the charge of the lepton candidate is consistent with a prompt semileptonic decay of B_{sig} . The flavour of B_{sig} is inferred from that of the B_{tag} . In inferring the B_{sig} flavour, we make no allowance for B^0 -mixing. Oscillated semileptonic events, $B^0 \rightarrow \bar{B}^0 \rightarrow X\ell^+\bar{\nu}$, suffer from large backgrounds from secondary decays $B^0 \rightarrow D(\rightarrow Y\ell^+\bar{\nu})X$, where Y is a hadronic system arising in the D decay, of the correct flavour B , and hence have a low purity. For neutral B decays, this restriction vetoes $18.58 \pm 0.11\%$ [211] of signal events.

We tighten the requirements on the quality of the reconstructed tag, selecting candidates with $M_{\text{bc}} > 5.27 \text{ GeV}$, $-0.15 < \Delta E < 0.1 \text{ GeV}$, and $\mathcal{P}_{\text{FEI}} > 0.01$. For off-resonance events, the M_{bc} threshold is set to 5.24 GeV . If, after the application of these selection criteria, multiple candidates are present, we select the candidate with the highest $\mathcal{P}_{\text{FEI}}$. Combined with the veto of events containing additional leptons, this ensures that only a single candidate is reconstructed per event. The highest ranked preselected tag candidate is selected in approximately 90% of events. The seventh ranked candidate is selected in less than 0.05% of $\bar{B} \rightarrow X\ell\bar{\nu}$ events, thus truncating the preselected sample to a maximum of seven candidates has minimal impact on the analysis.

If any tracks assigned to the X system have high lepton likelihoods, $P_e > 0.8$ or $P_{\mu} > 0.9$, the event is vetoed, as we cannot unambiguously ascertain which lepton originates from the primary B decay, and which originates from a secondary decay. Additionally, most events with a secondary lepton also contain an extra neutrino, worsening the resolution of quantities associated with the neutrino reconstruction.

As the low lepton energy region is still highly contaminated by continuum events, and events in which we have reconstructed a fake or secondary lepton, we restrict our consideration to $E_{\ell}^{B_{\text{sig}}} > 0.7 \text{ GeV}$.

8.3.1 Continuum suppression

Although the above selections and the implicit requirement that the FEI algorithm can reconstruct a suitable B meson candidate already suppress a large amount of background from continuum processes, it is possible to efficiently veto a large portion of the remaining continuum events by studying the shape of the event. As $B\bar{B}$ pairs are produced near their mass threshold, they have only small momenta in the c.m. frame, $\sim 300 \text{ MeV}$, leading to an isotropic distribution of decay products. In contrast, continuum events produce back-

to-back jets of light quarks with significant momenta. The difference is demonstrated schematically in Fig. 8.1. We exploit this shape difference to train boosted decision tree classifiers to identify continuum events.

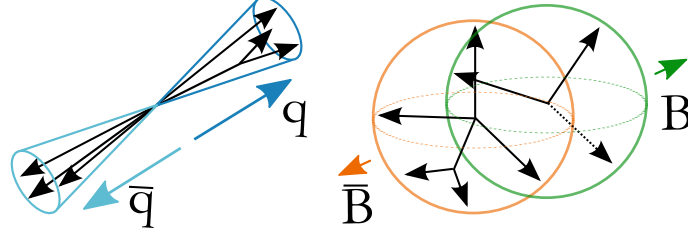


Figure 8.1: Schematic comparison of the event shape of $B\bar{B}$ and continuum events. Originally appeared in Ref. [238].

The classifier considers 30 variables, all calculated with respect to the tag B and the rest of event system plus lepton, which together make up the *other* B system. There are 13 variables related to the thrust, defined as:

$$\mathcal{T} = \sum_i^N |\vec{T} \cdot \vec{p}_i| / |\vec{p}_i|, \quad (8.8)$$

where, $\vec{T}$, is the thrust axis, the unit vector along which the projection of all momenta is maximised, and the sum runs over all considered particles. A value of 1 indicates a jet-like shape, 1/2 a spherical shape. The thrust is calculated separately for particles assigned to the tag B and other B , $\mathcal{T}_B$, and $\mathcal{T}_O$, respectively. The classifier also considers the angle between the two thrust axes, which is expected to be back-to-back for continuum events, $\cos \theta_{\vec{T}_B \vec{T}_O}$, and the angle between the thrust axis of the tag B and the z -axis $\cos \theta_{\vec{T}_B \vec{z}}$.

We also consider nine variables, known as CLEO cones [245] which are calculated by dividing the angle about the thrust axis of the tag B candidate into sections of 10° . The CLEO cones, denoted as $\mathcal{C}_1, \dots, \mathcal{C}_9$, give the sum of momenta in each region, folding the backward and forward regions together.

The remaining 17 variables are related to Fox-Wolfram moments [246], a set of rotationally invariant observables that describe two-particle correlations in terms of spherical harmonics. The moments are given by:

$$H_l = \sum_{i,j} |\vec{p}_i| |\vec{p}_j| P_l(\cos \theta_{i,j}), \quad (8.9)$$

where i and j run over selected particles, and $\vec{p}_i, \vec{p}_j$ are the momentum of the i^{th} and j^{th} particle, $\theta_{i,j}$ the angle between them and P_l the l^{th} order Legendre polynomial. Of particular interest is the ratio of the second to zeroth moment, $R_2 \equiv H_2/H_0$, as it effectively separates $B\bar{B}$ from continuum events. We therefore include it as a feature variable.

Fox-Wolfram moments have been extended to consider separately particles associated with the tag B (here denoted s) or other B (o), or consider combinations between the two

sets (*so*) [247], and to consider separately charged and neutral particles and the missing momentum vector [248], in what is referred to as Kakuno-Super-Fox-Wolfram (KSFw) moments, which we denote $\mathcal{H}$. The use of the moments originates from the study of charmless hadronic B decays, in which it is common to extract the signal yield from a simultaneous fit to $M_{bc}, \Delta E$, and the output of a discriminator based on KSFw moments. The moments calculated with respect to only particles originating from, what is in our context the tag B , but in the context of these analysis the signal B , were excluded to reduce the correlation to $M_{bc}, \Delta E$, and are not available within **basf2**². Fourteen KSFw moments are used as features: $\mathcal{H}_{00}^{so}, \mathcal{H}_{02}^{so}, \mathcal{H}_{04}^{so}, \mathcal{H}_{10}^{so}, \mathcal{H}_{12}^{so}, \mathcal{H}_{14}^{so}, \mathcal{H}_{20}^{so}, \mathcal{H}_{22}^{so}, \mathcal{H}_{24}^{so}, \mathcal{H}_0^{oo}, \mathcal{H}_1^{oo}, \mathcal{H}_2^{oo}, \mathcal{H}_3^{oo}$, and $\mathcal{H}_4^{oo}$. Ref. [248] additionally observed a dependence on the transverse energy of the event E_t , and the missing mass squared M_{Miss}^2 , both variables are therefore included as features.

To account for differences in reconstruction due to the different detector geometries, we train BDTs independently for SVD1 and SVD2. We further split the samples in half, training two classifiers for each SVD configuration, which are cross applied during inference. For data, one of the classifiers is chosen at random. The BDTs are trained to distinguish well-reconstructed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events from continuum events. For SVD1 (SVD2), a total of 23960 (111282) signal events, and 691924 (3007836) continuum events are used in training³. The samples are further divided, with 70% of events used for training and 30% used to optimise the hyperparameters via grid search. We train 500 trees of depths two and three for the early and late SVD configurations, respectively, using **FastBDT** [178]. The remaining two hyperparameters, the shrinkage and subsampling rate were left at their default parameters of 0.1 and 0.5, respectively, as it was found that varying these parameters had negligible impact.

The importance of the feature variables to the classifiers, as measured by normalised average loss reduction per node split, are presented in Fig. 8.2. We investigate the modelling of the input variables within the continuum sample, and classifier response, using the off-resonance sample, as a pure continuum sample. A comparison of the data to MC in the off-resonance sample for the three most important variables and the classifier response is shown in Fig. 8.3. We observe some mismodelling, this is addressed in Sec. 8.4.1.

The receiver operator characteristic curves, showing the continuum rejection as a function of acceptance of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ or $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events, are presented in Fig. 8.4. The selected thresholds are shown as crosses. We reject 92.8% of continuum events while maintaining high efficiency on $B\bar{B}$ events, selecting 94.0% and 89.0% of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events, respectively.

²The currently available KSFw variable set is therefore suboptimal for a tagged analysis such as those presented within this thesis. However, after rejecting events the classifier deems continuum-like, which has only a small effect on signal efficiency, continuum events are only a minor background. Therefore, the lack of moments calculated with respect to the tag B is essentially negligible.

³The training used an independent sample of inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events generated with slightly different parameters for the DFN model, due to an incorrect scheme conversion. The difference in the shape of the event is expected to be negligible.

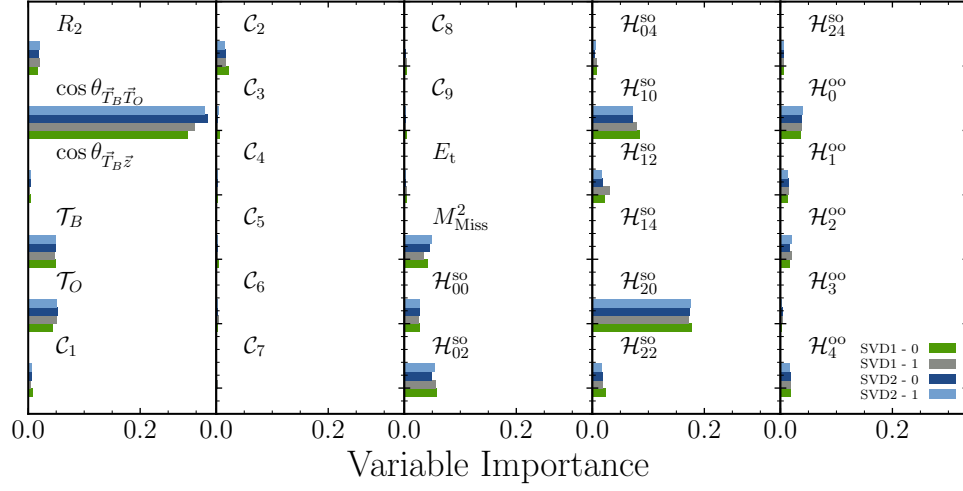


Figure 8.2: Importances of the feature variables to the four continuum suppression classifiers, as measured by normalised average loss reduction per split.

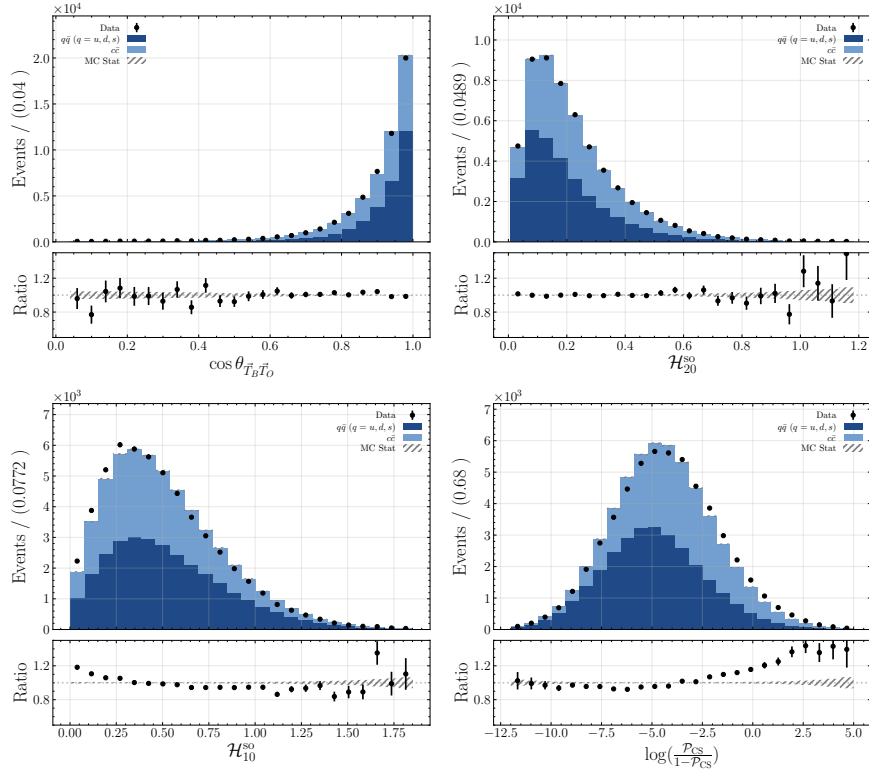


Figure 8.3: Data-MC agreement of the off-resonance sample in the three most important variables for the continuum suppression classifiers, $\cos \theta_{\vec{T}_B \vec{T}_O}$, $\mathcal{H}_{20}^{\text{so}}$, and $\mathcal{H}_{10}^{\text{so}}$, and for the log-transformed classifier response. The MC has been scaled to the data yield, an increase of approximately 15%.

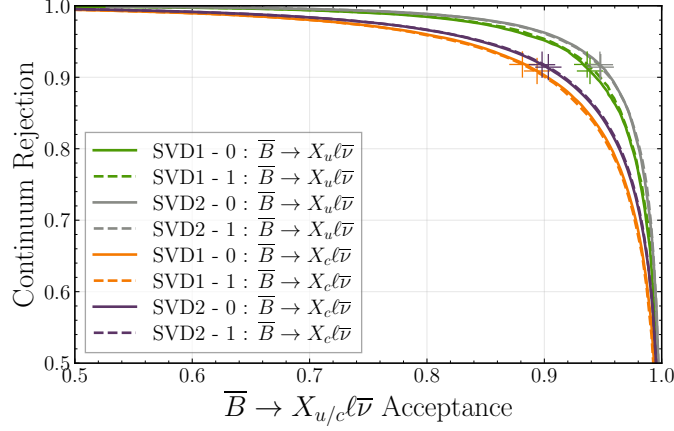


Figure 8.4: Receiver operator characteristic curves for the four continuum suppression classifiers, for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events against continuum events. The imposed thresholds are given as crosses.

8.4 Tag calibration

As discussed in Sec. 4.2 FEI is a complex algorithm, combining large amounts of low-level inputs in hundreds of BDTs to reconstruct $\mathcal{O}(10000)$ channels. Many of the inputs are poorly reproduced in the simulated samples, such as particle identification variables, and the modelling of considered channels; including their branching fractions and resonant substructures, many of which have never been measured. Thus, the efficiency of the FEI is expected to differ between data and MC, and as the FEI is trained on simulated samples, the efficiency is expected to be higher in simulation than data.

As we measure the ratio of $\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})$ to $\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ the overall normalisation of the MC, and thus estimated reconstruction efficiency of the FEI, is irrelevant with any constant factor cancelling in the ratio. However, the fraction of events in which the tag is correctly reconstructed (good tags), as compared to combinatorial background tags, impacts the resolution of reconstructed signal side quantities. In addition, knowledge of the continuum event reconstruction rate, relative to the $B\bar{B}$ rates is needed to correctly subtract this background.

8.4.1 Continuum

The number of reconstructed and selected continuum events differs between data and MC for three interconnected reasons: poor modelling of $q\bar{q}$ events in MC, differences in the FEI reconstruction efficiency, and differences in acceptance rates of the continuum suppression classifier as observed in Sec. 8.3.1.

Studying the off-resonance sample, we observe a linear dependence of the data to MC ratio in the lepton energy. Therefore, we adopt a linear calibration function of the form $f(E_\ell^{B_{\text{sig}}}) = mE_\ell^{B_{\text{sig}}} + c$, to correct for all three causes of mismodelling. The calibration

is performed independently for the electron and muon modes as large differences were observed in the required calibration functions. These differences are thought to originate from the different proportions of true to fake leptons in the two samples⁴. The data to MC ratios of the two samples, and fit results are presented in Fig. 8.5. The calibration is applied to the on-resonance simulated continuum sample.

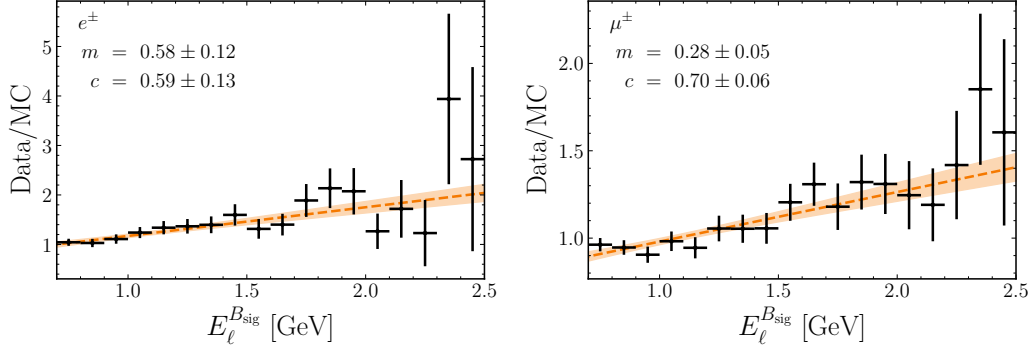


Figure 8.5: Calibration fits to the ratio of off-resonance data to MC yields in the electron (*left*) and muon channels (*right*). The uncertainties combine the data and MC statistical uncertainties.

8.4.2 Tag calibration

For $B\bar{B}$ events, we consider the tag side M_{bc} to allow us to distinguish well-reconstructed tags from combinatorial background. Well-reconstructed tags are expected to peak near the B meson mass, with combinatorial backgrounds following what is expected to be an ARGUS distribution [249] with a long tail towards lower M_{bc} values. To properly constrain the combinatorial background, we must access a region of phase space dominated by it, away from the B peak. We thus loosen the M_{bc} selections applied to the tag B , considering candidates with $M_{bc} > 5.24$ GeV. If multiple tag candidates are present in our nominal sample, we take the tag with the largest value of $\mathcal{P}_{\text{FEI}}$ after the M_{bc} selection. As we have varied the M_{bc} selection, we must also reconsider this selection. For the tag calibration, we take only candidates with the highest rank $\mathcal{P}_{\text{FEI}}$ after the loose online preselection. All other selections discussed in Sec. 8.3 are applied. The impact of this selection is discussed below.

We fit each of the 26 neutral and 29 charged FEI channels individually with a binned extended maximum likelihood fit to the M_{bc} distribution between 5.24 GeV and 5.29 GeV. As the number of reconstructed events in each channel differs by several orders of magnitude between channels, the number of bins used in the fit is determined algorithmically.

⁴One could therefore propose a more complicated correction procedure, correcting the true and fake lepton components individually. However, as continuum events are only a small background in the analysis, and the shapes of the two components in lepton energy and momentum transfer are similar, the simple approach is considered sufficient.

The number of bins is given by $n = \lfloor N^{1/3} \rfloor$ bins, where N is the yield of data events in each channel, and all bins are of equal width. If any bins contain no data events, the number of bins is iteratively reduced by one until at least one event is found in each bin.

Three components are considered by the fit: continuum, well-reconstructed tags, and combinatorial $B\bar{B}$ events. The continuum component has a similar shape to the combinatorial $B\bar{B}$ events, therefore, we fix the yield of this component to the MC expectation (after continuum calibration). To define well-reconstructed tags, we adopt a looser definition than standard for Belle analyses. We consider tags in which all particles have been correctly reconstructed (the standard definition) and tags in which one photon, originating either from beam background or the signal side, has been incorrectly added to the tag side, as well reconstructed. We additionally do not require that the full decay tree is correct, only that the final state particles are correctly assigned. This definition is adopted because tags in which one photon has been incorrectly included are observed to peak at M_B . In addition, they have been reconstructed with the correct charge and flavour and differ from the tags which meet the strict definition only by a slight degradation of their momentum resolution. We apply the ratio of pre-fit to post-fit yields for the well-reconstructed tag and combinatorial background as calibration weights to the MC sample. On average, in the nominal sample, a well-reconstructed tag is assigned a weight of $0.68 \pm 0.001_{\text{stat}} \pm 0.023_{\text{syst}}$ and a poorly reconstructed tag a weight of $0.87 \pm 0.002_{\text{stat}} \pm 0.041_{\text{syst}}$, where the first error is statistical and the second systematic. The systematic error is dominated by the branching fractions of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ components; for the evaluation of these uncertainties, see Sec. 9.5. Example fits are shown in Fig. 8.6, and a comparison of the M_{bc} spectrum before and after the calibration in Fig. 8.7. We note that there is still some data-MC disagreement in the high M_{bc} shoulder of the distribution. This disagreement is expected to be caused by small mismodelling of the beam energy in simulation. As the following discussion will clarify, a mismodelling of this scale has a negligible impact on the measured ratio of partial branching fractions.

As discussed above, the selection of tag candidates in the calibration sample varies from that of the nominal sample. Approximately 90% of the nominal sample, the candidates with the highest-ranked FEI signal probability after pre-selection, are present in both samples. We investigate the applicability of the calibration to the remaining 10% by comparing the data and MC yields for events in the nominal sample not found in the calibration sample. The yields are found to agree within 5%.

We test the impact of the precision of the tag calibration by varying the 110 tag calibration factors randomly about their nominal values by 5%, assuming normal distributions. The measured $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ ratios vary by 0.2% in simulation and data. Given the uncertainty of the measurement exceeds this by almost two orders of magnitude, the remaining mismodelling in the tag calibration is considered negligible.

Having calibrated for the tagging efficiency, we compare the hadronic mass spectrum of events in simulation with data in Fig. 8.8. Good agreement is observed between the data and the simulation.

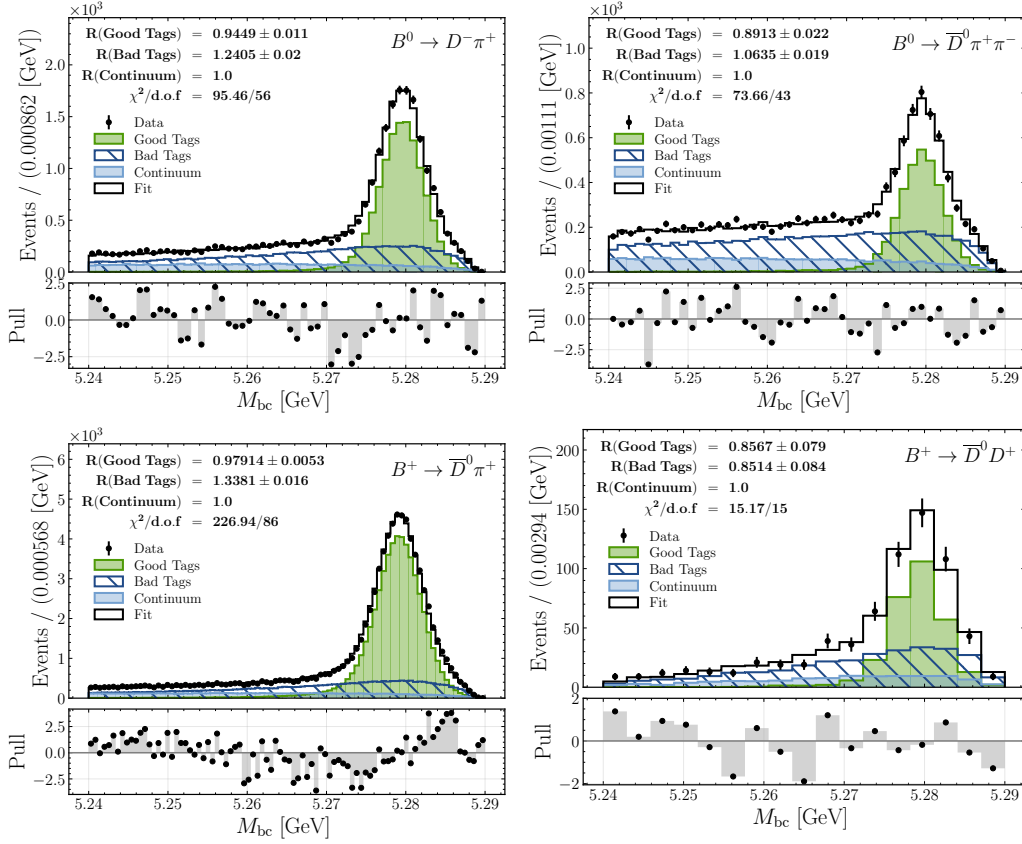


Figure 8.6: Example calibration fits to M_{bc} for two neutral and two charged B tag channels. The continuum component is fixed. The calibration weights, given by the ratio of pre-fit to post-fit yields of good and bad tags are given in the legend. The uncertainties are statistical.

8.5 Signal region enrichment

We further restrict our samples to events within the region of interest, requiring $E_{\ell}^{B_{\text{sig}}} > 1$ GeV. We further require that the number of charged kaons in the event be consistent with $|N_{K^+} - N_{K^-}| \leq 1$, allowing for a single kaon produced via a $b \rightarrow c \rightarrow s$ cascade decay and additional pairs produced via pair production. This selection has only a marginal impact. The sample selected according to all above requirements will be used to extract the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yield, and we now label the $\bar{B} \rightarrow X \ell \bar{\nu}$ sample.

From the $\bar{B} \rightarrow X \ell \bar{\nu}$ sample, we select a subsample with a higher $\bar{B} \rightarrow X_u \ell \bar{\nu}$ purity using selections on the missing mass of the event and a partial reconstruction of the D^* mesons. This sample is labelled the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction sample.

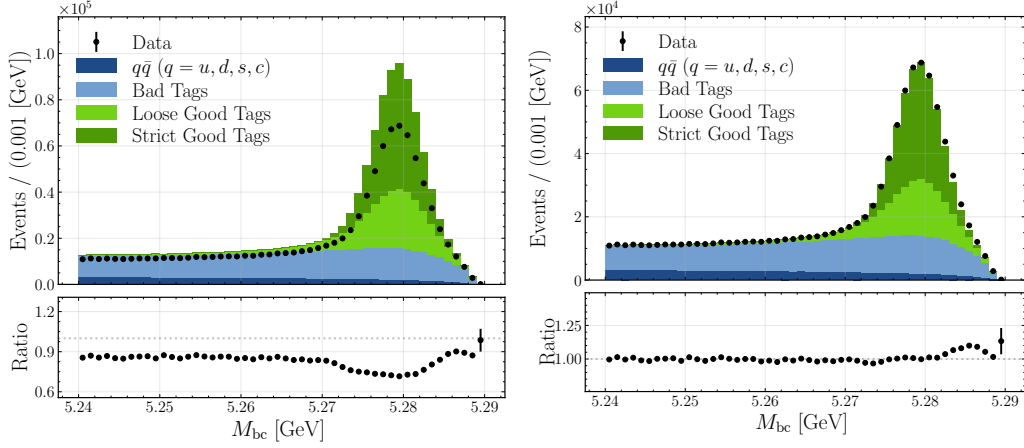


Figure 8.7: Pre-calibration (*left*) and post calibration (*right*) M_{bc} spectrum of the calibration sample. The good tags have been split into two subcategories corresponding to the strictly well-reconstructed tags, and tags in which one photon has been incorrectly added, labelled ‘Loose Good Tag’. The uncertainties are statistical.

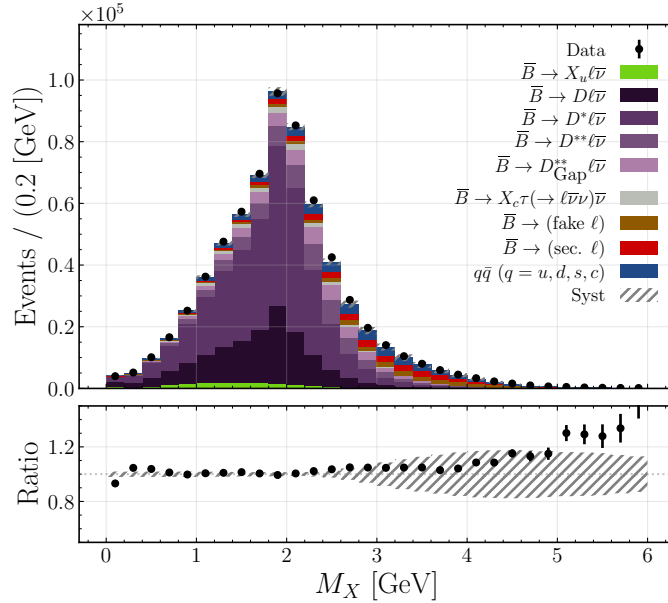


Figure 8.8: The hadronic mass spectrum of reconstructed $\bar{B} \rightarrow X \ell \bar{\nu}$ events. Excellent Data-MC agreement is observed for the bulk of the distribution. Some disagreement is observed at high hadronic mass which may originate either from mismodelling of secondary and fake lepton events or of $\bar{B} \rightarrow D^{**} \ell \bar{\nu}$ events.

8.5.1 Partial D^* reconstruction

Approximately two-thirds of D^{*+} mesons decay to $D^0\pi^+$, where due to the small mass difference of the D^{*+} and D^0 mesons, the pion is expected to be produced almost at rest; thus, we label it a *slow* pion. Identifying such a pion thus allows us to infer the presence of a D^{*+} meson, identifying the event as charmed.

We search for pions in the rest of event with momenta below 200 MeV in the c.m. frame. If such a pion is found, we inclusively reconstruct the D^* candidate. We assume the D^* momentum to lie along the flight direction of the slow pion in the lab frame, and estimate its energy to be

$$E_{D^*} = \frac{M_{D^*}}{M_{D^*} - M_{D^0}} E_\pi. \quad (8.10)$$

As pions are abundant in $B\bar{B}$ events, including $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events, identifying just a slow pion is not sufficient to veto an event. We therefore target specifically $\bar{B}^0 \rightarrow D^{*+}(\rightarrow D^0\pi^+)\ell\bar{\nu}$ events by calculating a missing mass squared from the estimated D^* four vector,

$$M_{\text{Miss}}^2(D^*) = (p_{e^+e^-} - p_{B_{\text{tag}}} - p_\ell - p_{D^*})^2. \quad (8.11)$$

If the slow pion originated from a D^{*+} decay which itself originated from a $\bar{B}^0 \rightarrow D^{*+}(\rightarrow D^0\pi^+)\ell\bar{\nu}$ decay, $M_{\text{Miss}}^2(D^*)$ is expected to be consistent with zero. If the slow pion does not originate from the decay of a D^* , the incorrectly estimated D^* four vector leads to negatively skewed $M_{\text{Miss}}^2(D^*)$ values. If multiple slow pion candidates are reconstructed, we keep the candidate with the lowest absolute $M_{\text{Miss}}^2(D^*)$. We further restrict the sample to slow pions with charges opposite to the signal lepton, to be consistent with $\bar{B}^0 \rightarrow D^{*+}(\rightarrow D^0\pi^+)\ell\bar{\nu}$ events, and with lab frame momenta between 50 and 200 MeV to select slow pions within the region in which the performance of slow pion reconstruction has been studied by the Belle collaboration.

Finally, we study the distribution of true $\bar{B}^0 \rightarrow D^{*+}(\rightarrow D^0\pi^+)\ell\bar{\nu}$ and combinatorial events in $M_{\text{Miss}}^2(D^*)$ and $\cos\theta_{D^*\ell}$, the cosine of the angle between the D^* and the lepton momenta, which is expected to peak near -1 for events with a true semileptonic $\bar{B}^0 \rightarrow D^{*+}(\rightarrow D^0\pi^+)\ell\bar{\nu}$ decay. We scan a two-dimensional figure of merit to find the optimal threshold for these variables. The chosen figure of merit is

$$\text{FoM} = \frac{\eta^{\bar{B} \rightarrow X_u \ell \bar{\nu}}}{\sqrt{\eta^{\bar{B} \rightarrow X_u \ell \bar{\nu}} + \eta^{\bar{B} \rightarrow X_c \ell \bar{\nu}} + \eta^{q\bar{q}} + \eta^{\text{Sec.Fakes}}}}, \quad (8.12)$$

where $\eta^{\bar{B} \rightarrow X_u \ell \bar{\nu}}$, $\eta^{\bar{B} \rightarrow X_c \ell \bar{\nu}}$, $\eta^{q\bar{q}}$, $\eta^{\text{Sec.Fakes}}$ are the expected yields of $\bar{B} \rightarrow X_u \ell \bar{\nu}$, $\bar{B} \rightarrow X_c \ell \bar{\nu}$, continuum, and secondary and fake lepton events after applying the veto selection. We find the optimal threshold to be at $M_{\text{Miss}}^2(D^*) > -2 \text{ GeV}^2$, with no requirement on $\cos\theta_{D^*\ell}$. The distribution of the two variables in simulation and the two dimensional figure of merit scan are shown in Fig. 8.9.

In principle, a similar selection could be performed using slow π^0 candidates, which could identify most of the remaining $\bar{B}^0 \rightarrow D^{*+} \ell \bar{\nu}$ and two-thirds of $B^+ \rightarrow D^{*0} \ell \bar{\nu}$. How-

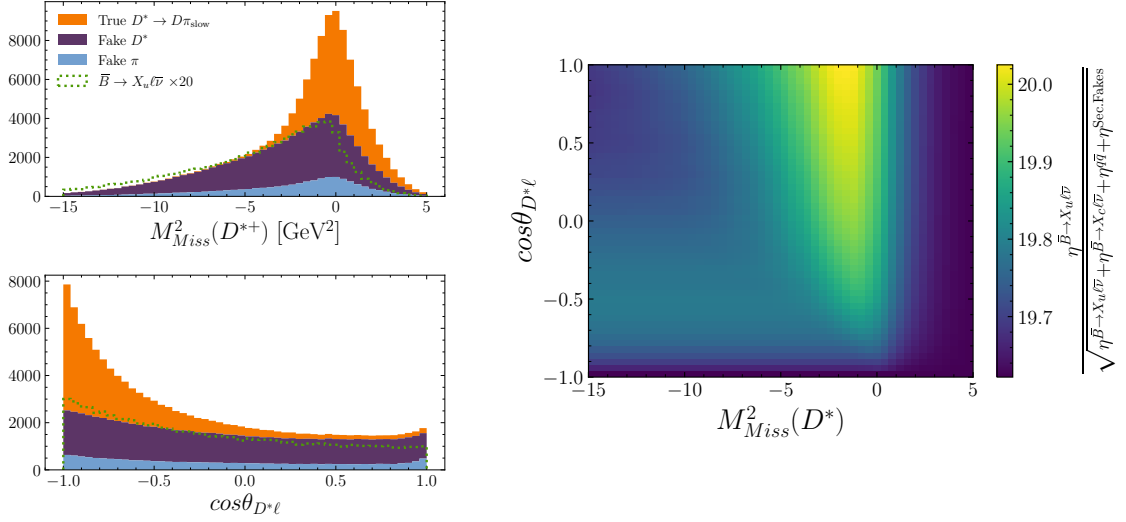


Figure 8.9: Simulated distributions of $M_{\text{Miss}}^2(D^*)$ and $\cos \theta_{D^* \ell}$ by status of the identified slow pion and result of the figure-of-merit scan. Note that events in which no slow pion is found are not shown in the figure.

ever, because of the abundance of photons in average $B\bar{B}$ events, a large number of slow-pion candidates are reconstructed, and the rejected sample would consist of mostly events in which combinations of random photons have been identified. We therefore consider only charged slow pions.

8.5.2 Missing mass squared

We consider the missing mass squared of the event, as defined in Eq. 8.7. For well-reconstructed true semileptonic $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events, this is expected to peak at zero, as only the massless neutrino is missing in the event. For $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events and other backgrounds, there is a skew towards higher values due to additional missing particles. These can be additional neutrinos, K_L^0 particles as they are not reconstructed, or simply particles which are not reconstructed as they fall outside the detector acceptance or do not leave clear enough signals to be reconstructed by the reconstruction algorithms. Due to the higher multiplicity of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events, this occurs more frequently in such events than in $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events. It is thus possible to increase the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ purity of the sample by rejecting events with high missing mass squared.

The missing mass squared distribution for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ sample is presented in Fig. 8.10. A deficit is observed in the simulation for $M_{\text{Miss}}^2 < 0 \text{ GeV}^2$. This deficit has been seen repeatedly in Belle studies and has been considered in detail in the context of a measurement of $\bar{B} \rightarrow X \tau \bar{\nu}$ in Ref. [250], which reported several key findings.

- The mismodelling is independent of the number of tracks reconstructed.

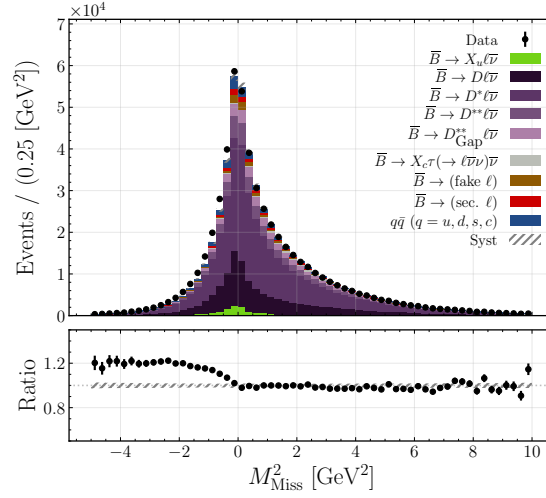


Figure 8.10: Comparison of data and simulation in M_{Miss}^2 for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction sample.

- The collision energy is known much too precisely to introduce such discrepancies.
- Placing more stringent requirements on the rest of event particles, requiring photons to have energies above 150 MeV and tracks to have transverse momenta above 100 MeV, does not resolve the mismodelling.
- A slight shift in the track momentum distribution was observed, with tracks in data on average carrying harder momenta than tracks in simulation. The observed mismodelling is reduced if the momentum of tracks in simulation is increased. Such a procedure however also shifts the reconstructed mass of intermediate particles K_S^0 , D , away from the correct values. The mismodelling of the track momenta must therefore be related to the fundamental modelling of the decays in terms of multiplicities and kinematics, rather than a uniform underestimation of track momenta.
- No relation was found between the mismodelling and reconstructed tag B , both in terms of reconstructed channel, and the presence of photons in the decay chain.
- In general, events in the region of $M_{\text{Miss}}^2 \sim 0 \text{ GeV}^2$ are well reconstructed, missing a single massless particle such as a photon or neutrino. In order for the squared missing mass to be negative, the magnitude of the missing momentum must exceed the missing energy. This can occur if a charged particle has been misidentified as a species heavier than it truly is, if a particle has been duplicated in the reconstruction, or if particles that did not originate in the $\Upsilon(4S)$ decay have been reconstructed, such as beam background or other noise.

The same deficit is observed in the off-resonance sample, and in events in which the lepton carries the wrong charge, ie. we reconstruct $\Upsilon(4S) \rightarrow B^+ \ell^+ X^{--}$, or $\bar{B}^0 \ell^+ X^-$. The

‘wrong-charge’ sample is dominated by hadronic B decays in which the child D -meson of the B has decayed semileptonically, $\bar{B} \rightarrow D(\rightarrow Y\ell^-\nu)X$, where X is a hadron or hadrons, and Y represents a hadronic system arising from the decay of the semileptonic decay of the D meson. The modelling of this and the off-resonance sample is thus to a large extent independent of the nominal sample. This indicates that the origin of the mismodelling cannot originate (solely) in the modelling of the physics process under consideration. It must originate in a source common to all three event categories, such as modelling of beam-background, detector noise, or detector responses.

As we do not fit the M_{Miss}^2 distribution in the analysis, the primary impact of mismodelling is an incorrect estimation of the efficiency. We therefore consider the absolute missing mass squared, $|M_{\text{Miss}}^2|$, which was observed to be better modelled in the off-resonance and wrong-lepton sidebands. The distribution of this variable near zero for events in the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ extraction sample, is shown in Fig. 8.11. The figure is split into the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ enriched and depleted subsamples, the criteria for which are described in Sec. 8.5.3. The agreement of data to simulation is found to be good, with perhaps a slight linear trend towards a deficit in simulation at higher $|M_{\text{Miss}}^2|$ values. In the region of interest for extracting $\bar{B} \rightarrow X_u\ell\bar{\nu}$ events $|M_{\text{Miss}}^2| \lesssim 0.5 \text{ GeV}^2$, good agreement is observed.

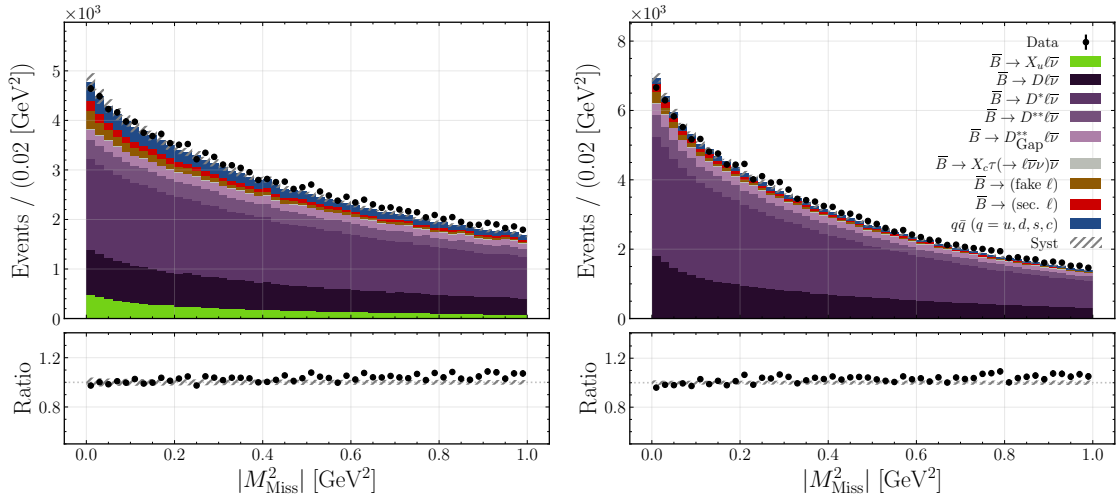


Figure 8.11: Comparison of data and simulation in $|M_{\text{Miss}}^2|$ for the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ enriched (left) and depleted (right) subsamples of the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ extraction sample.

The threshold imposed on $|M_{\text{Miss}}^2|$, $|M_{\text{Miss}}^2| < 0.43 \text{ GeV}^2$, is selected to minimise the expected total uncertainty on the measured ratio of partial branching fractions, $\Delta\mathcal{B}(\bar{B} \rightarrow X_u\ell\bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c\ell\bar{\nu})$, using the nominal fit procedure defined in the following chapter. The expected uncertainties as a function of threshold are given in Fig. 8.12.

In Sec. 9.7 we test the stability of the measurement against the value of the applied missing mass squared threshold.

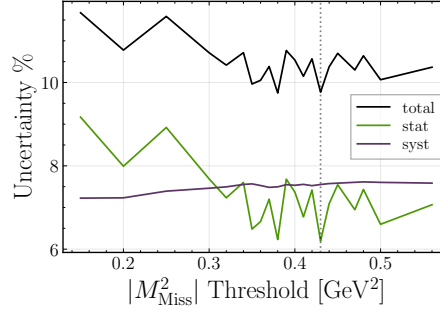


Figure 8.12: Optimisation of the $|M_{\text{Miss}}^2|$ selection. The expected statistical uncertainty was estimated by varying the pseudo-data yields in each step of the measurement by drawing from a Gaussian of width $\sqrt{N}$, for each bin in each histogram under consideration. This neglects the event-overlap between samples, however this is expected to have only a small impact on the optimisation. The estimates of the expected statistical uncertainty are noisy, however sufficient to note that the expected uncertainty is approximately uniform in the region about 0.4 GeV^2 .

8.5.3 Kaon counting

As mentioned during the discussion on the identification of tracks in the rest of event, the presence of kaons on the signal side is a strong indication of a $b \rightarrow c \rightarrow s$ cascade transition, effectively tagging charmed events. We thus select events in which we find no, or pairs of kaons, requiring

$$(N_K + N_{K_S^0}) \bmod 2 = 0, \quad (8.13)$$

where $N_K, N_{K_S^0}$ are the number of reconstructed charged kaons and K_S^0 candidates, respectively. We allow for the production of kaon pairs as during the hadronisation of the X_u system $s\bar{s}$ pairs can be produced, known as $s\bar{s}$ popping, leading to a pair (or multiple pairs) of kaons. As simulated, $s\bar{s}$ popping occurs in approximately 8% of inclusively generated events. The rate of $s\bar{s}$ popping is poorly understood, and further discussed in Sec. 9.5. The sample selected in this manner is named the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ *enriched* sample. Events that do not pass this selection are labelled as $\bar{B} \rightarrow X_u \ell \bar{\nu}$ *depleted* and offer insight into the background composition of events within the enriched sample. The efficiency of this selection for the considered event categories is given in Table 8.1.

A comparison of the N_K and $N_{K_S^0}$ distributions in data and simulation is given in Fig. 8.13. Excellent agreement between simulation and data is observed.

The cumulative efficiency of all discussed selections, is presented in Table 8.2.

Category	$\bar{B} \rightarrow X_u \ell \bar{\nu}$ Enriched [%]	$\bar{B} \rightarrow X_u \ell \bar{\nu}$ Depleted [%]
$\bar{B} \rightarrow X_u \ell \bar{\nu}$	93.0	7.0
$\bar{B} \rightarrow D \ell \bar{\nu}$	40.4	59.6
$\bar{B} \rightarrow D^* \ell \bar{\nu}$	39.2	60.8
$\bar{B} \rightarrow D^{**} \ell \bar{\nu}$	43.1	56.9
$\bar{B} \rightarrow D_{\text{Gap}}^{**} \ell \bar{\nu}$	42.1	57.9
$\bar{B} \rightarrow X_c \tau (\rightarrow \ell \bar{\nu} \nu) \bar{\nu}$	40.1	59.9
$\bar{B} \rightarrow (\text{fake } \ell)$	60.8	39.2
$\bar{B} \rightarrow (\text{sec. } \ell)$	53.8	46.2
$q\bar{q} \ (q = u, d, s, c)$	69.6	30.4

Table 8.1: Percentage of selected events falling within the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples by event category.

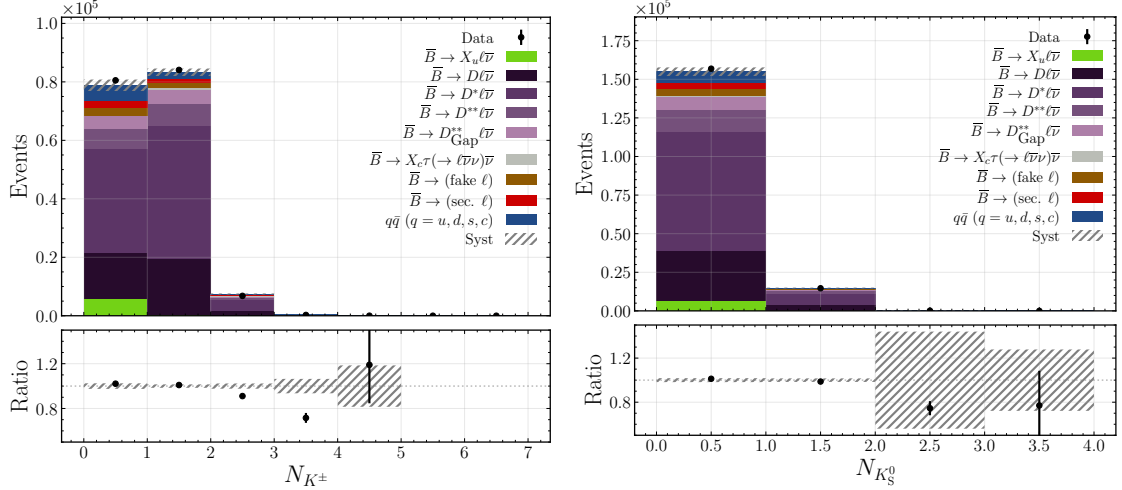


Figure 8.13: Comparison of data and simulation for the number of reconstructed K^+ , and K_S^0 candidates. All selections defining the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction sample are applied.

Selection	$\bar{B} \rightarrow X_u \ell \bar{\nu}$	$\bar{B} \rightarrow X_c \ell \bar{\nu}$	$\bar{B} \rightarrow X_c \tau (\rightarrow \ell \bar{\nu} \nu) \bar{\nu}$	$\bar{B} \rightarrow (f.l.)$	$\bar{B} \rightarrow (sec.l.)$	Continuum
Reconstructed [Candidates]	327592 (83669)	16331572 (5466146)	575031	7052846	10185583	78302507
$Q_{B_{\text{tag}}} + Q_{B_{\text{sig}}} = 0$ [%]	56.08(66.95)	50.25(56.11)	49.72	40.20	39.63	45.27
Lepton charge correlation [%]	48.56(61.92)	41.83(43.67)	40.41	23.73	17.91	26.80
$M_{bc} > 5.27$ GeV [%]	27.12(54.71)	23.44(38.56)	22.68	10.13	6.53	7.84
$-0.15 < \Delta E < 0.10$ GeV [%]	24.09(48.46)	20.91(34.19)	20.25	9.21	5.98	7.14
$\mathcal{P}_{\text{FEI}} > 0.01$ [%]	11.28(29.65)	9.54(21.32)	9.09	3.67	1.88	1.55
Best tag selection [%]	5.79(16.46)	4.78(11.78)	4.56	1.92	1.10	1.16
Single lepton [%]	5.50(15.49)	4.25(9.83)	4.06	1.07	0.88	0.95
$E_{\ell}^{B_{\text{sig}}} > 0.7$ GeV [%]	5.41(15.24)	4.13(9.22)	3.37	0.73	0.57	0.72
Continuum classifier [%]	5.09(14.59)	3.68(8.08)	2.55	0.34	0.29	0.05
$E_{\ell}^{B_{\text{sig}}} > 1.0$ GeV [%]	4.74(13.61)	3.25(6.90)	1.43	0.15	0.12	0.03
$ N_{K^+} - N_{K^-} < 2$ [%]	4.73(13.60)	3.24(6.88)	1.42	0.15	0.12	0.03
$\bar{B} \rightarrow X \ell \bar{\nu}$ sample [Candidates]	15509 (11378)	529113 (376087)	8189	10431	11943	22190
D^* veto [%]	96.59(97.28)	89.14(89.57)	88.93	89.16	89.13	92.72
$ M_{\text{Miss}}^2 < 0.43$ GeV ² [%]	41.16(44.57)	27.67(29.83)	7.62	45.93	34.66	35.35
$N_K + N_{K_S^0} \bmod 2 = 0$ [%]	38.29(41.75)	11.05(11.26)	2.95	27.85	18.61	24.31
$\bar{B} \rightarrow X_u \ell \bar{\nu}$ sample [Candidates]	5938 (4750)	58444 (42356)	241	2905	2223	5395

Table 8.2: Cumulative efficiency, and number of expected candidates, of the event-selection post reconstruction. The efficiency is given in terms of the number of candidates. Before the single-lepton requirement multiple candidates can be reconstructed within one event. The numbers in brackets correspond to candidates in which a lepton has been correctly reconstructed and the tag meets the definition of good tag used for the FEI calibration. Note that we apply the continuum and FEI calibrations to all candidates. This is expected to slightly overestimate the continuum yield before continuum suppression.

8.5.4 $\bar{B} \rightarrow X_u \ell \bar{\nu}$ reconstruction efficiencies

Before discussing the measurement procedure, it is useful to briefly discuss the reconstruction efficiency of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events as a function of the kinematic variables and specific resonant states. These are presented in Fig. 8.14 and Table 8.3, respectively. We observe events at high $E_\ell^{B_{\text{sig}}}$ and q^2 , and low M_X , to be reconstructed at rates several times higher than those at the other end of the respective spectrum⁵. Although the measurement probes the phase space $E_\ell^{B_{\text{sig}}} \geq 1 \text{ GeV}$, it should thus be noted that the sensitivity of the measurement will be driven by these regions of phase space, and the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ channels, with highest efficiency. These regions also correspond to the regions with the lowest $\bar{B} \rightarrow X_c \ell \bar{\nu}$ contamination, further driving the sensitivity.

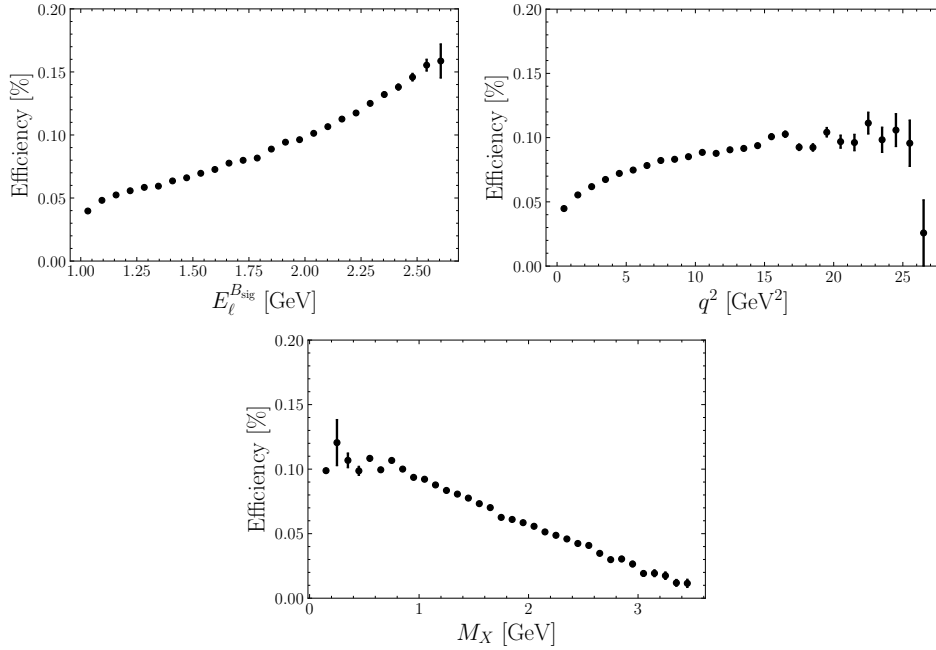


Figure 8.14: Reconstruction efficiency of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events after all selections as a function of $E_\ell^{B_{\text{sig}}}$, q^2 and M_X . The errors are statistical. Due to the reconstruction level selection of $E_\ell^{B_{\text{sig}}} > 1.0 \text{ GeV}$ the efficiency below the shown 1 GeV boundary in $E_\ell^{B_{\text{sig}}}$ rapidly drops to zero.

⁵Only one previous study published the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ reconstruction efficiency as a function of the kinematic variables: the differential measurement of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ spectrum by the Belle collaboration [225]. This study pursued an aggressive background rejection strategy and hence observes even larger dependence on the kinematic variables. However as this is a differential measurement, such a dependence is less impactful.

Decay Channel	B^+ [%]	B^0 [%]
$\bar{B} \rightarrow \pi \ell \bar{\nu}$	0.1408 ± 0.0027	0.0755 ± 0.0015
$\bar{B} \rightarrow \rho \ell \bar{\nu}$	0.1383 ± 0.0019	0.0772 ± 0.0011
$\bar{B} \rightarrow \eta \ell \bar{\nu}$	0.1143 ± 0.0034	-
$\bar{B} \rightarrow \omega \ell \bar{\nu}$	0.1197 ± 0.0020	-
$\bar{B} \rightarrow \eta' \ell \bar{\nu}$	0.0850 ± 0.0039	-
$\bar{B} \rightarrow x_u \ell \bar{\nu}$	0.08656 ± 0.00040	0.04397 ± 0.00032

Table 8.3: Reconstruction efficiency of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events by channel. The errors are statistical.

Measurement of ratios of partial branching fractions and CKM elements

In the preceding chapter, we discussed the reconstruction and selection of $\bar{B} \rightarrow X\ell\bar{\nu}$ events, and the selection of a $\bar{B} \rightarrow X_u\ell\bar{\nu}$ enriched subsample. In this chapter, we discuss how we extract the yield of $\bar{B} \rightarrow X_u\ell\bar{\nu}$ and $\bar{B} \rightarrow X_c\ell\bar{\nu}$ events from these samples and thus measure the ratio $\Delta\mathcal{B}(\bar{B} \rightarrow X_u\ell\bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c\ell\bar{\nu})$ and $|V_{ub}|/|V_{cb}|$.

The measurement of $\Delta\mathcal{B}(\bar{B} \rightarrow X_u\ell\bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c\ell\bar{\nu})$ and extraction of $|V_{ub}|/|V_{cb}|$ proceeds in four steps.

1. We first correct the normalisation of the remaining non-semileptonic events, those in which we have selected a secondary lepton or a hadron misidentified as a lepton, in the samples.
2. We extract the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ yield from the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ enriched subsample, using the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ depleted subsample to derive a data driven model of $\bar{B} \rightarrow X_c\ell\bar{\nu}$ events.
3. We extract the $\bar{B} \rightarrow X_c\ell\bar{\nu}$ yield through a simple background subtraction.
4. Finally, we correct for the reconstruction efficiencies of the two samples and combine the corrected yields to form the ratio of partial branching fractions.

9.1 Normalisation of non-semileptonic backgrounds

The sample selected in the previous chapter consists of four components: $\bar{B} \rightarrow X_u\ell\bar{\nu}$, $\bar{B} \rightarrow X_c\ell\bar{\nu}$, continuum, and $B\bar{B}$ backgrounds. In order to extract the $\bar{B} \rightarrow X_u\ell\bar{\nu}$ and $\bar{B} \rightarrow X_c\ell\bar{\nu}$ events the contribution of the backgrounds must be calibrated such that they can be accurately subtracted. The continuum contribution has already been calibrated. We calibrate the contribution of $B\bar{B}$ background by considering a sample defined by the selection $0.7 < E_\ell^{B_{\text{sig}}} < 1.8 \text{ GeV}, M_X > 2.0 \text{ GeV}$. This sample is enriched in background originating from leptons from a secondary decay of the hadronic system, leptonic decays of tauons from semitauonic events $\bar{B} \rightarrow X\tau(\rightarrow \ell\nu\bar{\nu})\bar{\nu}$, and purely hadronic decays where a hadron has been incorrectly identified as a lepton. These three backgrounds have similar

shapes in the kinematic variables of interest and are thus grouped into a single component, secondary and fake leptons. Additionally, the selection removes the endpoint region to reduce sensitivity to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ modelling. The calibration is performed for each subsample ($\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted) separately, as the calibration factor may differ between the samples.

For each subsample we perform a binned maximum likelihood fit in $E_\ell^{B_{\text{sig}}}$ floating the fraction of events assigned to the secondary and fake lepton or $\bar{B} \rightarrow X_c \ell \bar{\nu}$ components. The continuum component and the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ component, which is heavily suppressed in this region of phase space, are kept fixed to expectation. The post-fit distributions are shown in Fig. 9.1. For the remainder of this work, unless otherwise stated we scale the expected yield of the combined secondary and fake lepton contribution by the ratio of the fitted yield to the expectation in MC. The fits to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples have $\chi^2/\text{ndf} = 10.9/9$, and $\chi^2/\text{ndf} = 4.5/9$, respectively, where

$$\chi^2 = \sum_{i,j} (N_i^{\text{obs}} - N_i^{\text{exp}}) (C_{\text{stat}}^{\text{obs}} + C_{\text{syst}}^{\text{exp}})^{-1}_{ij} (N_j^{\text{obs}} - N_j^{\text{exp}}), \quad (9.1)$$

N_i^{obs} and N_i^{exp} are the observed yields in data and fitted yields per bin, and $C_{\text{stat}}^{\text{obs}}$ and $C_{\text{syst}}^{\text{exp}}$ are the statistical covariance matrix of the data (which is diagonal), and the systematic covariance matrix of the fitted yields. For details of the evaluation of the statistical and systematic uncertainties, as well as their covariance matrices, see Sec. 9.5. The measured calibration factors for the secondaries and fakes component are $0.922 \pm 0.02_{\text{stat}} \pm 0.16_{\text{syst}}$ and $0.856 \pm 0.03_{\text{stat}} \pm 0.11_{\text{syst}}$, where the first error is statistical and the second systematic, for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples, respectively.

9.2 $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction

As discussed in Ch. 7, little is known about significant portions of the total $\bar{B} \rightarrow X_c \ell \bar{\nu}$ decay rate, with the gap between known exclusive channels and the inclusive rate having to be filled with best-guess channels in our modelling. Given the relative abundance of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ mismodelling of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ component introduced by this lack of knowledge has the potential to introduce large biases in measurements of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate. To reduce such biases we adopt a data-driven description of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ shape, using the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted subsample to derive a template that is used to model the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ shape when extracting the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yield from the enriched subsample by means of a template maximum likelihood fit. The template (T) is given as

$$T_i = \tau_i (N_{i,D}^{\text{Data}} - a \eta_{i,D}^{X_u \ell \bar{\nu}} - \eta_{i,D}^{q\bar{q}} - \eta_{i,D}^{\text{Sec.Fakes}}), \quad (9.2)$$

where $N_{i,D}^{\text{Data}}$ is the data yield in bin i for the depleted sample and $\eta_{i,D}^{X_u \ell \bar{\nu}}, \eta_{i,D}^{q\bar{q}}, \eta_{i,D}^{\text{Sec.Fakes}}$ are the MC yields for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$, $q\bar{q}$, and secondary and fake lepton components, respectively. The coefficient a is initially set to 1; τ_i is a transfer factor from the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted to enriched subsample given by the ratio of MC expectations in the enriched and

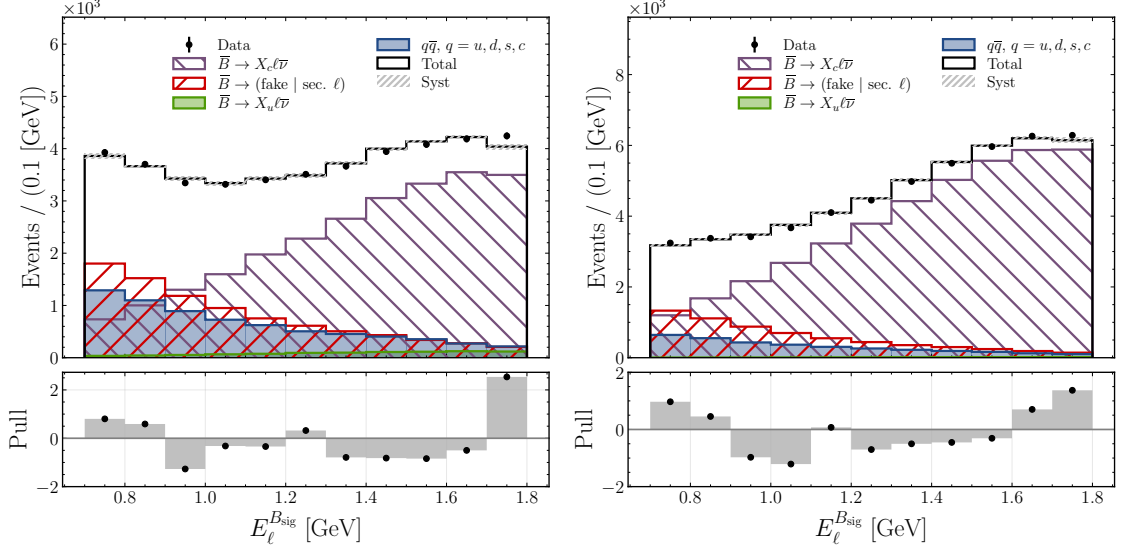


Figure 9.1: Fit to the secondary and fake lepton control region ($0.7 < E_\ell^{B_{\text{sig}}} < 1.8 \text{ GeV}$, $M_X > 2.0 \text{ GeV}$) in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enhanced (left) and depleted (right) subsamples for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction sample. The four components are: correctly reconstructed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events (green), correctly reconstructed $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events (purple), continuum events (blue), and events in which either a hadron has been misidentified as a lepton or the lepton originates from a secondary decay (red).

depleted subsamples, $\eta_{i,E}^{X_c \ell \bar{\nu}}$ and $\eta_{i,D}^{X_c \ell \bar{\nu}}$, respectively,

$$\tau_i = \frac{\eta_{i,E}^{X_c \ell \bar{\nu}}}{\eta_{i,D}^{X_c \ell \bar{\nu}}}, \quad (9.3)$$

and incorporates expected differences in the shape of the signal enriched and depleted spectra. These differences arise from correlations between the event selection criteria, such as the $|M_{\text{Miss}}^2|$ selection, and the kaon counting criteria used to split the subsamples.

We extract the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yield by means of a two-dimensional binned fit in $q^2 : E_\ell^{B_{\text{sig}}}$ to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched subsample. The residuals in these two variables in simulation, defined as the generated quantity minus the reconstructed, for the selected $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events are given in Fig. 9.2. The figure also gives the residuals for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events in the $\bar{B} \rightarrow X \ell \bar{\nu}$ sample, as this is relevant for the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction.

The data driven template for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extracted in the side band is valid for the description in $E_\ell^{B_{\text{sig}}}$. This is because the hadronic system and the presence and identification of kaons within it, is independent of the lepton energy. For q^2 , the data-driven approach is partially valid. The used definition of q^2 is insensitive to incorrect mass assignment for tracks in the hadronic system, and thus the misidentification of charged kaons, and to missed reconstruction of K_s^0 candidates, as the four vectors of the child pions are included. Events that are assigned to the enriched subsample due to these factors have the

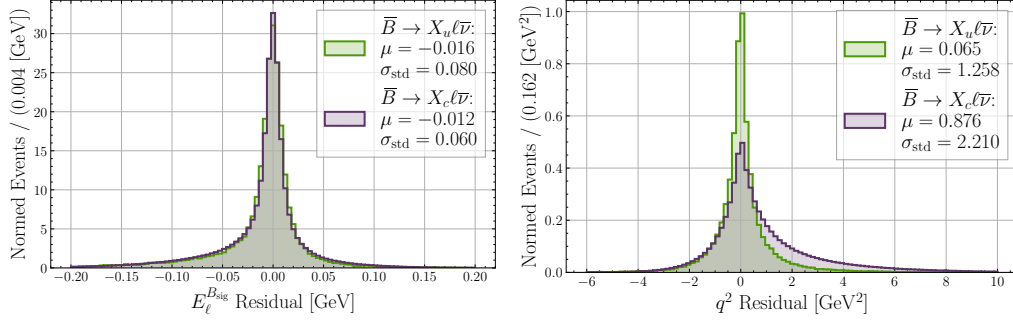


Figure 9.2: Residuals of $E_\ell^{B_{\text{sig}}}$ and q^2 for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events, corresponding to events selected in the $\bar{B} \rightarrow X \ell \bar{\nu}$ sample and the enriched $\bar{B} \rightarrow X_u \ell \bar{\nu}$ sample respectively.

same distribution as they would have if the kaons were correctly identified and the events were assigned to the depleted subsample. Events in which kaons are missed due to not being reconstructed, such as K_L^0 , charged kaons in which the track is not reconstructed, or K_S^0 in which the child pions tracks or clusters are not reconstructed, are assigned to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched sample. Such events will be biased towards higher q^2 as the momentum of the missed particle is absorbed into the reconstructed neutrino momentum. The $\bar{B} \rightarrow X_c \ell \bar{\nu}$ template derived in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted sample (Eq. 9.2) is therefore not a fully valid description of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ shape in q^2 in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched sample. As the bias is small however, and the expected precision of the two-dimensional fit was found to significantly improve on a one-dimensional fit to $E_\ell^{B_{\text{sig}}}$, the two-dimensional fit is used. The effect of the bias in q^2 is considered as a systematic uncertainty and discussed in Sec. 9.5.1.3.

The fit uses 16 bins, with the binning chosen such that each bin is expected to have equal $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yield, dividing the phase space first into four bins of q^2 , with the events in each subdivided into four bins in $E_\ell^{B_{\text{sig}}}$. This binning is demonstrated in Fig. 9.3 for the four components under consideration (secondary and fake leptons, continuum, $\bar{B} \rightarrow X_u \ell \bar{\nu}$, $\bar{B} \rightarrow X_c \ell \bar{\nu}$). The binning ensures that a large sample of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events are present in each bin, providing sufficient statistical power for the data-driven $\bar{B} \rightarrow X_c \ell \bar{\nu}$ template. A large fraction of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events collect in the final broad $E_\ell^{B_{\text{sig}}}$ and q^2 bins, $q^2 > 7.5 \text{ GeV}^2$, $E_\ell^{B_{\text{sig}}} \gtrsim 1.8 \text{ GeV}$ reducing the exposure to $\bar{B} \rightarrow X_u \ell \bar{\nu}$ modelling in the endpoint region.

A comparison of data and simulation in the two variables of interest and in their two-dimensional binning is shown in Fig. 9.4. To test if the observed mismodelling is consistent between the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples, we consider the double ratio of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields in data to MC in the enriched to depleted subsample ($R_E^{X_c \ell \bar{\nu}} / R_D^{X_c \ell \bar{\nu}}$). The yield of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ in data in each subsample is calculated as the data yield subtracting the MC expectation of $\bar{B} \rightarrow X_u \ell \bar{\nu}$, continuum, and secondary and fake lepton events. These are presented in Fig. 9.5. The mismodelling is found to be consistent between the two subsamples and is attributed to the modelling of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events.

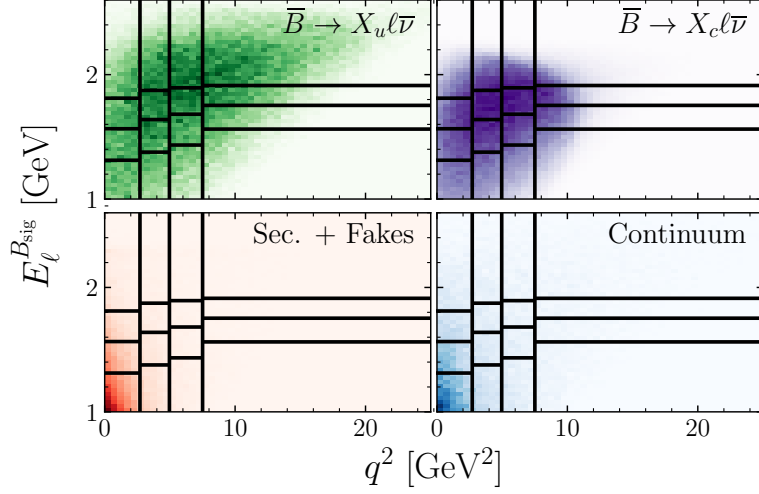


Figure 9.3: Binning structure for two dimensional fit to $q^2 : E_\ell^{B_{\text{sig}}}$. The sample is first divided into four equal $\bar{B} \rightarrow X_c \ell \bar{\nu}$ frequency q^2 bins. Each q^2 bin is then subdivided into four equal frequency $E_\ell^{B_{\text{sig}}}$ bins.

As the contribution of continuum events is fixed to MC expectation, we further test that the calibration derived from the total off-resonance sample is applicable to the sub-samples under consideration. Fig. 9.6 shows the distribution of the off-resonance data and simulated events that pass the requirements. Due to the smaller size of the off-resonance sample and the imposed continuum rejection criteria only a few hundred events are reconstructed. The comparison is thus noisy but reasonable agreement is observed.

The fit is performed as a maximum likelihood fit where the number of expected events in each bin of the enhanced sample is given by

$$H_i = N_E^{\text{Data}} \left[\frac{\eta_E^{\text{Sec.Fakes}}}{N_E^{\text{Data}}} h_{i,E}^{\text{Sec.Fakes}} + \frac{\eta_E^{q\bar{q}}}{N_E^{\text{Data}}} h_{i,E}^{q\bar{q}} + f_E^{X_u \ell \bar{\nu}} h_{i,E}^{X_u \ell \bar{\nu}} \right. \\ \left. + (1 - f_E^{X_u \ell \bar{\nu}} - \frac{\eta_E^{\text{Sec.Fakes}} + \eta_E^{q\bar{q}}}{N_E^{\text{Data}}}) \frac{T_i}{\sum_j T_j} \right], \quad (9.4)$$

where $h_{i,E}^{X_u \ell \bar{\nu}}, h_{i,E}^{q\bar{q}}, h_{i,E}^{\text{Sec.Fakes}}$ are the fraction of events of the $\bar{B} \rightarrow X_u \ell \bar{\nu}, q\bar{q}$, and secondary and fake lepton components reconstructed in bin i , respectively, as determined by the MC simulation. The total data yield is given by N_E^{Data} , and $\eta_E^{q\bar{q}}, \eta_E^{\text{Sec.Fakes}}$ are the expected yield of the continuum and secondary and fake lepton component in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enhanced sample, respectively. The parameter $f_E^{X_u \ell \bar{\nu}}$ is the fraction of events assigned to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ component and is the only parameter floated in the fit.

From simulation we expect 7% of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events to contaminate the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted sample. To reduce the dependence on the assumed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ branching fraction

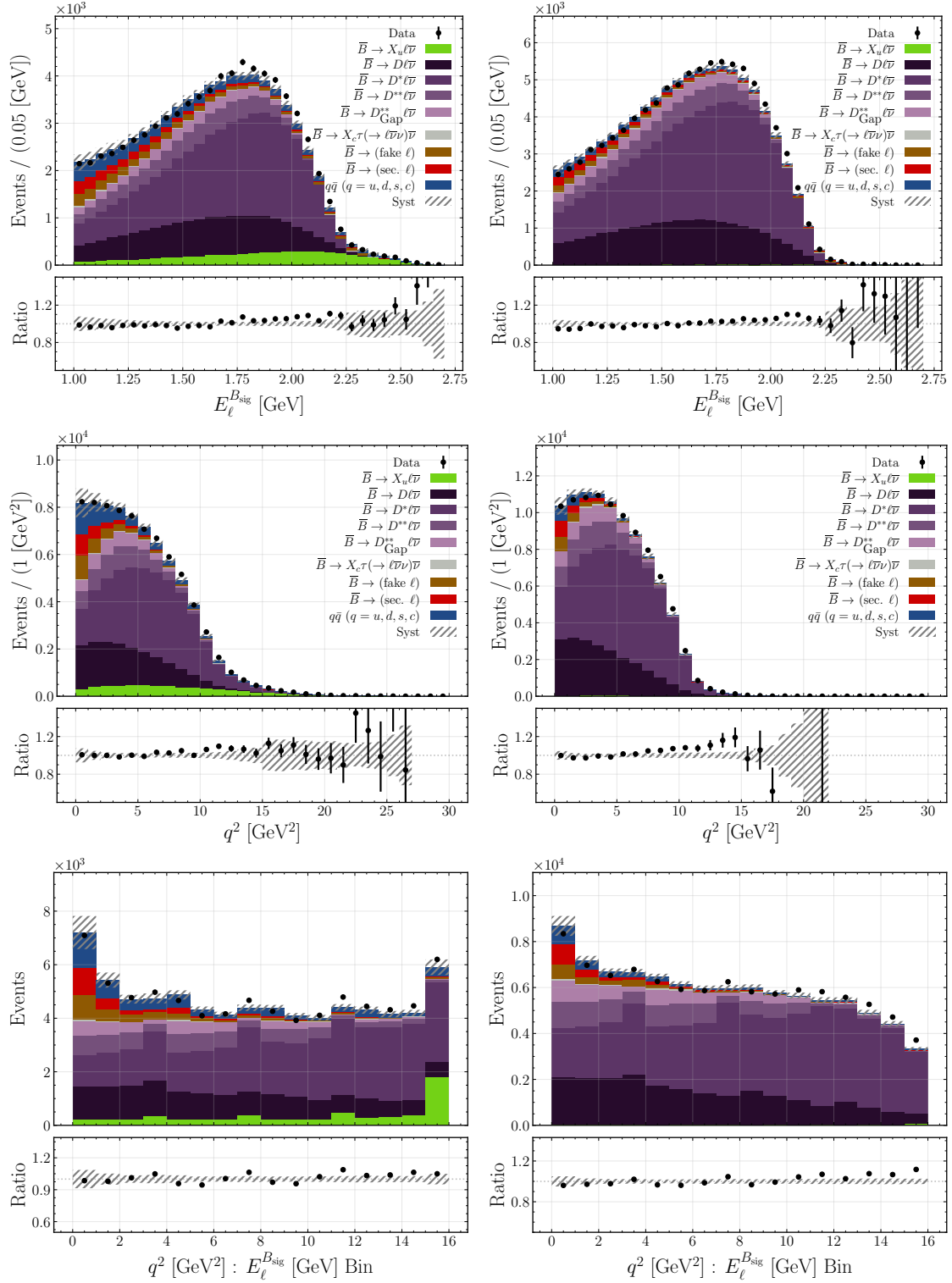


Figure 9.4: Reconstructed $E_\ell^{B_{\text{sig}}}$ and q^2 spectra, and the two-dimensional binning used for the fit for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enhanced (left) and depleted (right) subsamples. The error bands of simulated samples incorporate the full set of systematic uncertainties discussed in section 9.5. The secondaries and fakes calibration is not applied.

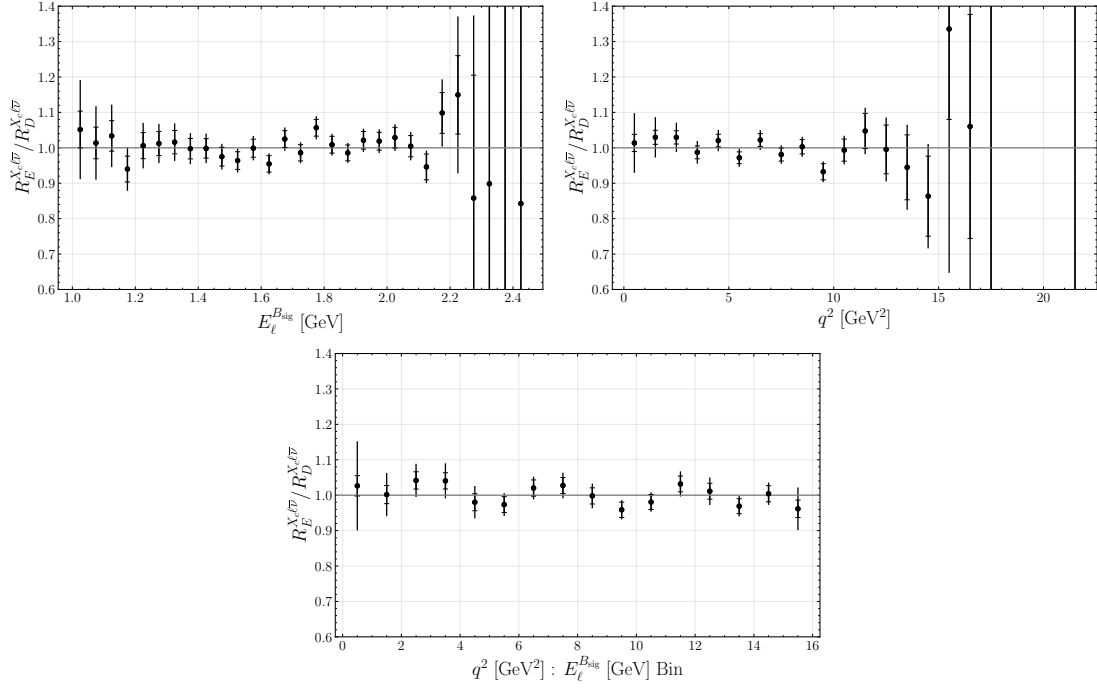


Figure 9.5: Double ratios of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields in data to MC in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched to depleted subsample ($R_E^{X_c \ell \bar{\nu}} / R_D^{X_c \ell \bar{\nu}}$) in the kinematic variables $E_\ell^{B_{\text{sig}}}$ and q^2 , and the two-dimensional binning of the two used in the analysis. The yield of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ in data is calculated as the data yield subtracting the MC expectation of $\bar{B} \rightarrow X_u \ell \bar{\nu}$, continuum and secondary and fake lepton events. The secondaries and fakes calibration is not applied.

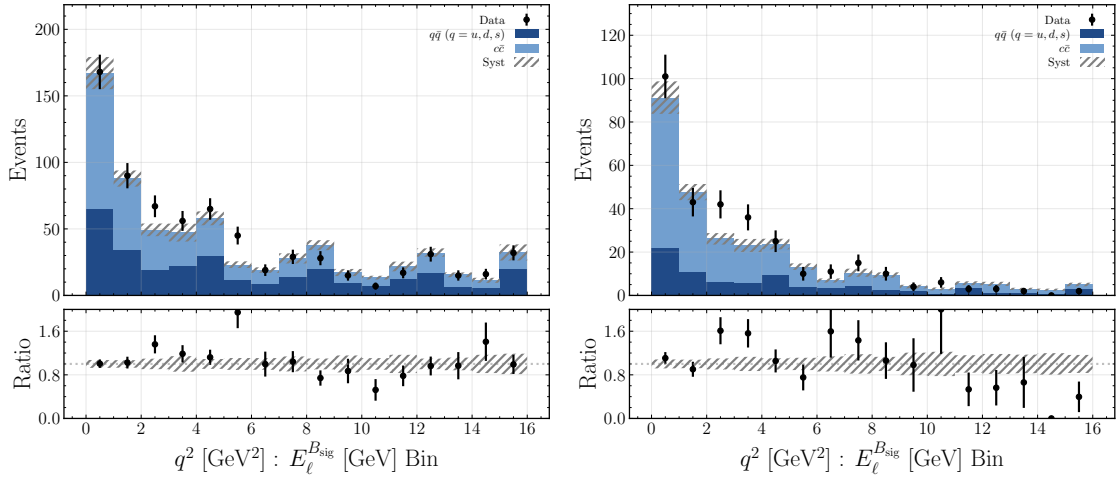


Figure 9.6: Distribution of off-resonance events passing the selections of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched (left) and depleted (right) subsamples in the fit binning. The uncertainty band includes MC sample size, lepton identification uncertainties, and the continuum calibration uncertainties.

in MC we repeat the fit 20 times updating the coefficient a as

$$a = \frac{f_E^{X_u \ell \bar{\nu}} N_E^{\text{Data}}}{\eta_E^{X_u \ell \bar{\nu}}}, \quad (9.5)$$

on each iteration. Convergence, defined loosely as a change in a less than 10^{-4} between iterations, is generally observed within 4 iterations. The templates of the four components, an the fit result are shown in Fig. 9.7. The fit yields $N^{X_u \ell \bar{\nu}} = f_E^{X_u \ell \bar{\nu}} N_E^{\text{Data}} = 5430 \pm 450_{\text{stat}} \pm 350_{\text{syst}}$ $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events with a $\chi^2/\text{ndf} = 15.1/14$.

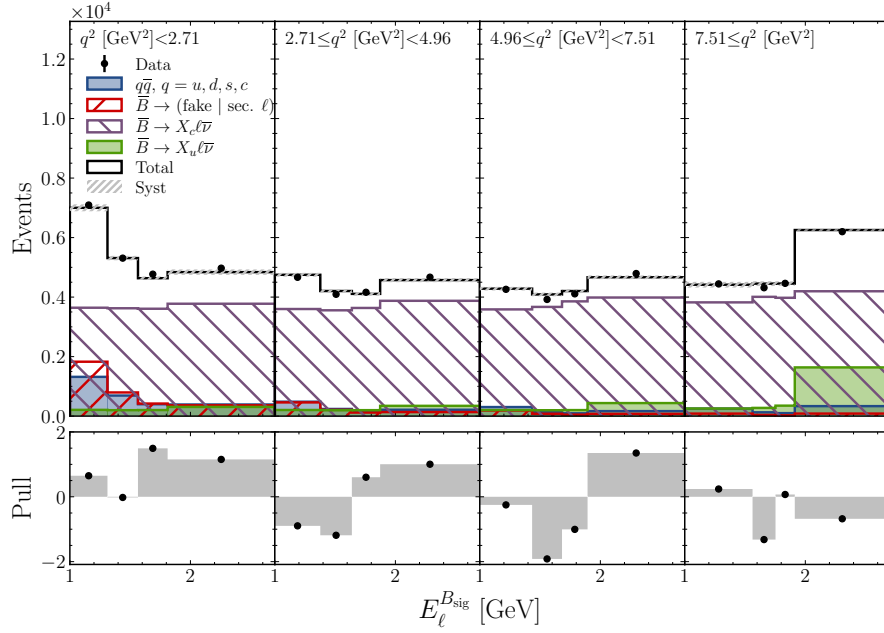


Figure 9.7: Fit to the $q^2 : E_\ell^{B_{\text{sig}}}$ distribution. The four components are: correctly reconstructed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events (*green*), correctly reconstructed $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events (*purple*), continuum events (*blue*), and events in which either a hadron has been misidentified as a lepton or the lepton originates from a secondary decay (*red*).

9.3 $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction

To extract the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yield we consider the $\bar{B} \rightarrow X \ell \bar{\nu}$ sample, selected before the D^* veto and $|M_{\text{Miss}}^2|$ requirements. The secondary and fake lepton component is normalised following the procedure established for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction sample. We maintain the split into the ‘enriched’ and ‘depleted’ subsample based on the number of kaons to allow for comparison of the found calibration factors with those for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ subsample.

The fits are presented in Fig. 9.8 and have $\chi^2/\text{ndf} = 24.1/9$ for the enriched subsample and $\chi^2/\text{ndf} = 16.0/9$ for the depleted subsample. The found calibration factors are $0.869 \pm 0.01_{\text{stat}} \pm 0.15_{\text{syst}}$ and $0.81 \pm 0.02_{\text{stat}} \pm 0.11_{\text{syst}}$ for the enriched and depleted subsamples,

respectively, lower than those found for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction sample, but exhibiting the same relative trend. These large χ^2/ndf values are dominated by the contribution from the final bin in $E_\ell^{B_{\text{sig}}}$ as the peak of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ distribution appears to be shifted between data and MC. To test the impact of this mismodelling we repeat the whole procedure removing the final bin from all four secondary and fake lepton normalisation fits, finding $\chi^2/\text{ndf} = 2.8/8(3.4/8)$ and $7.1/8(14.0/8)$ for the enhanced (depleted) subsamples of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction samples, respectively. The central value of $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ shifts by $+0.46\%$, which is considered negligible.

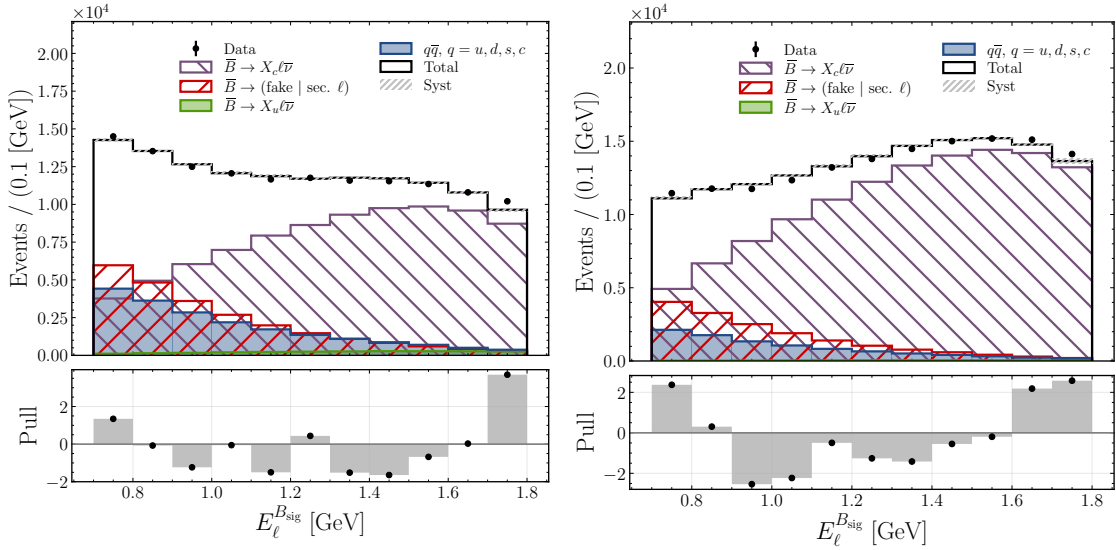


Figure 9.8: Fit to the secondary and fake lepton control region ($0.7 < E_\ell^{B_{\text{sig}}} < 1.8 \text{ GeV}$, $M_X > 2.0 \text{ GeV}$) in the kaon-depleted (*left*) and kaon-rich (*right*) subsamples for the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction sample.

The reconstructed $E_\ell^{B_{\text{sig}}}$ spectrum of the combined enhanced and depleted subsamples is shown in Fig. 9.9. Given the high purity of the sample, in excess of 90%, a simple background subtraction suffices to extract the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yield with a negligible uncertainty in comparison to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction. The yield is given as

$$N^{X_c \ell \bar{\nu}} = N^{\text{Data}} - \eta^{q\bar{q}} - \eta^{\text{Sec.Fakes}} - a\eta^{X_u \ell \bar{\nu}}, \quad (9.6)$$

where N^{Data} is the data yield in the $\bar{B} \rightarrow X \ell \bar{\nu}$ selection, and $\eta^{q\bar{q}}, \eta^{\text{Sec.Fakes}}$, and $\eta^{X_u \ell \bar{\nu}}$ are the MC expectations of the continuum, secondary and fake lepton background, and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ component. We measure $N^{X_c \ell \bar{\nu}} = 544900 \pm 1400_{\text{stat}} \pm 2400_{\text{syst}}$.

9.3.1 $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$

The ratio of partial branching fractions is given by

$$\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} = \frac{\epsilon^{X_c \ell \bar{\nu}} N^{X_u \ell \bar{\nu}}}{\epsilon^{X_u \ell \bar{\nu}} N^{X_c \ell \bar{\nu}}}, \quad (9.7)$$

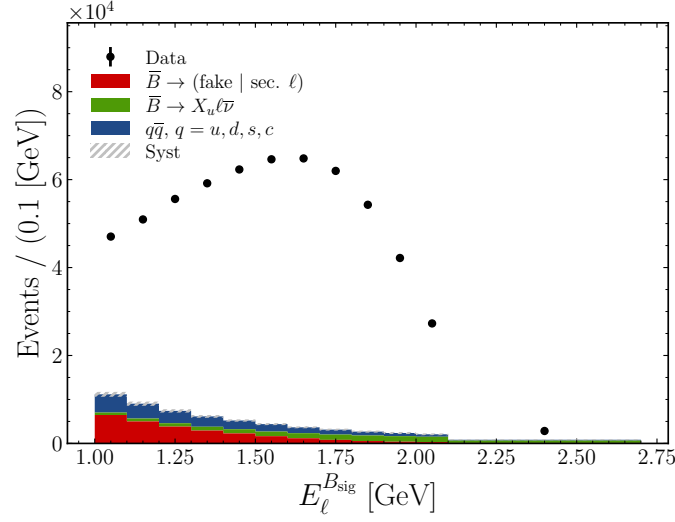


Figure 9.9: The determined background compared with all events in the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction sample.

where $\epsilon^{X_u \ell \bar{\nu}}$, $\epsilon^{X_c \ell \bar{\nu}}$ are the reconstruction efficiencies of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events, respectively, and are estimated from MC simulation as the ratio of reconstructed events in MC with $E_\ell^{B_{sig}} > 1 \text{ GeV}$ to the number generated with $E_{\ell, gen}^{B_{sig}} > 1 \text{ GeV}$ to be:

$$\epsilon^{X_u \ell \bar{\nu}} = (1.043 \pm 0.002_{\text{stat}} \pm 0.042_{\text{syst}}) \times 10^{-3}, \quad (9.8)$$

$$\epsilon^{X_c \ell \bar{\nu}} = (2.084 \pm 0.003_{\text{stat}} \pm 0.067_{\text{syst}}) \times 10^{-3}. \quad (9.9)$$

As the resolution in $E_\ell^{B_{sig}}$ is good, there is only negligible migration about the 1 GeV threshold, which is implicitly corrected from the MC expectation of the migration using the above definition. We measure

$$\frac{\Delta \mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta \mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} = (1.99 \pm 0.17_{\text{stat}} \pm 0.16_{\text{syst}}) \times 10^{-2}, \quad (9.10)$$

where the first error is statistical and the second systematic. A breakdown of the systematic uncertainties is given in Table 9.1, the estimation of the systematic uncertainties is described in Sec. 9.5.

9.3.2 Validity checks

To test the extraction method for bias we perform a toy test. The analysis is performed on 10^4 pseudodata sets generated from the simulated sample. For each repetition, a sample statistically equivalent to expectation in data is drawn using bootstrapping [251].

Central Value $\times 100$	1.99
Stat. Error (Data)	8.5
$\mathcal{B}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta' \ell \bar{\nu})$	0.2
$\mathcal{FF}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta' \ell \bar{\nu})$	0.3
$\mathcal{B}(\bar{B} \rightarrow x_u \ell \bar{\nu})$	0.8
Hybrid Model (BLNP)	1.0
DFN ($M_b^{\text{KN}}, a^{\text{KN}}$)	5.1
$N_{g \rightarrow s \bar{s}}$	0.4
$\mathcal{B}(\bar{B} \rightarrow D \ell \bar{\nu})$	0.1
$\mathcal{B}(\bar{B} \rightarrow D^* \ell \bar{\nu})$	0.8
$\mathcal{B}(\bar{B} \rightarrow D^{**} \ell \bar{\nu})$	0.3
$\mathcal{B}(\bar{B} \rightarrow D^{(*)} \eta \ell \bar{\nu})$	0.2
$\mathcal{B}(\bar{B} \rightarrow D^{(*)} \pi \pi \ell \bar{\nu})$	0.2
$\mathcal{FF}(\bar{B} \rightarrow D \ell \bar{\nu})$	0.2
$\mathcal{FF}(\bar{B} \rightarrow D^* \ell \bar{\nu})$	0.9
$\mathcal{FF}(\bar{B} \rightarrow D^{**} \ell \bar{\nu})$	0.4
Sec.Fakes. Composition	3.8
In-situ q^2 Calibration	2.8
ℓ ID Efficiency	0.1
ℓ ID Fake Rate	0.3
$K\pi$ ID Efficiency	1.0
$K\pi$ ID Fake Rate	0.6
K_S^0 Efficiency	0.2
π_{slow} Efficiency	< 0.1
Tracking	0.1
Continuum Calibration	0.4
N_{BB}	< 0.1
$f_{+/0}$	< 0.1
Stat. Error (MC)	2.6
Total Syst.	7.8

Table 9.1: Summary of the central value, statistical, and systematic uncertainties for the ratio of partial branching fractions. The uncertainties are given as relative values on the central value in percent.

The distributions of the differences¹ between the measured $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields and the true yields in the bootstrapped samples are given in Fig. 9.10. To each distribution, we fit a Gaussian and find the means to be consistent with zero, indicating that the extraction is unbiased.

Furthermore, we validate that the analysis procedure remains unbiased if the normalisation of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ contributions differs from the assumed. We vary the normalisation of the two components separately, varying the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yield in a range between $\pm 45\%$, and the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields in a range between $\pm 15\%$, equivalent to varying each by about three times the uncertainty of the world averages of their branching fractions. We test 11 points within these ranges, generating 10^3 pseudo-data samples for each point and repeating the fit procedure. For each tested normalisation, a Gaussian is fit to the measured ratio of partial branching fractions distributions. A linear trendline is fitted to the means of the Gaussians to test that the recovered ratio of partial branching fractions scales as expected when varying the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ contributions in the pseudo-data. The fits are presented in Fig. 9.11 and are consistent with the expected linear behaviour.

It should be noted that both of these tests do not extend to the tag calibration procedure. As the tag calibration is dominated by $\bar{B} \rightarrow X_c \ell \bar{\nu}$ the measured $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yield is expected to be near the assumed, and calibrated to, value. As the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ contribution is orders of magnitude smaller than the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ contribution to the tag calibration, differences to the assumed branching fractions will introduce only a negligible non-linearity.

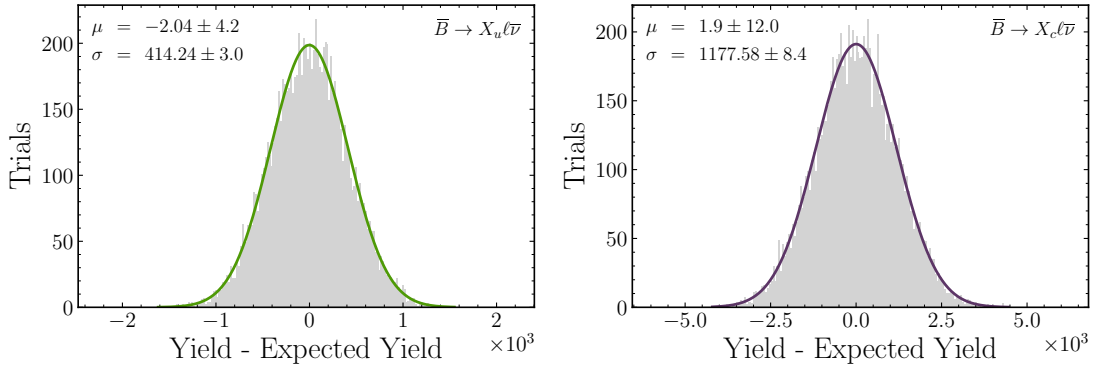


Figure 9.10: Distributions of residuals of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields measured in the toy test. Both are found to be consistent with zero indicating the procedure is unbiased.

¹It is usual to divide the residuals by the measured uncertainty, performing what is known as a pull test, allowing for validation of the estimated error of the fit as well as testing for biases. Given the overlap in events between the analysis steps, the errors estimated directly from the fit are not valid and thus it is not possible to perform a pull test. The statistical uncertainty of the results is estimated via bootstrapping.

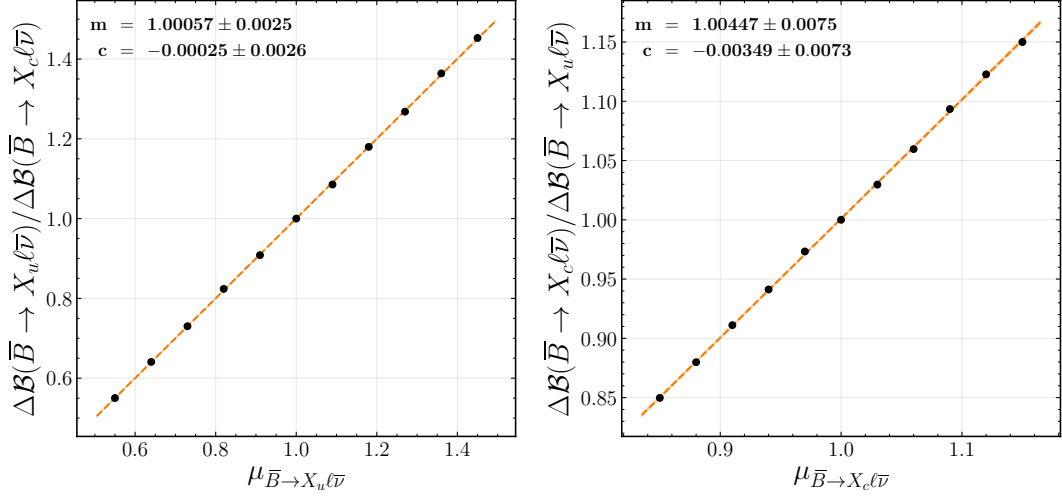


Figure 9.11: Result of the linearity tests on $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ varying the input branching fractions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$, given by the scale factors $\mu_{\bar{B} \rightarrow X_{u/c} \ell \bar{\nu}}$. The values on the y -axis have been scaled to one for scale factors of one. The recovered scaling behaviour is consistent with expectation.

9.4 Statistical uncertainty

As the different steps of the analysis consider overlapping data samples, it is not valid to simply propagate the uncertainties as estimated from the fitting procedure. We therefore determine the statistical uncertainty and correlations using a Poisson bootstrapping procedure [251]: 2000 ensembles of the collision data set are created and the entire analysis procedure is repeated, beginning with the calibration of the tagging algorithm. Half the inter-percentile range, that is the interval between the 15.87% and 84.13% percentiles of the distributions of the result measured in the toy studies, is taken as the uncertainty. No indication of a second minimum is observed in the distribution of the measured ratios of partial branching fractions.

The correlation between observables, such as subsample measurements or the yield in bins of a distribution, is estimated as the Pearson correlation coefficient from the central values of the trials.

9.5 Corrections and systematic uncertainty estimation

Several systematic uncertainties affect the measured quantities. The most important sources of systematic uncertainties for the main result, the ratio of partial branching fractions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ to $\bar{B} \rightarrow X_c \ell \bar{\nu}$, are associated with the modelling of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ component and the composition of the secondary and fake lepton component. Each systematic effect is varied independently and the analysis procedure repeated. The evaluation of the systematic uncertainties fall within three categories.

- Systematic error sources estimated via large sets of pseudo-experiments. The uncertainties are estimated to be half the inter-percentile range of the distribution of the results measured in the pseudo-experiment studies. The correlation between observables is estimated to be the Pearson correlation coefficient of the central values of the trials. Unless otherwise stated 200 pseudo-experiments are used for each source of systematic uncertainty.
- Systematic error sources estimated as the difference between a nominal and alternate model. For such systematic error source the full difference to the nominal result is taken as the uncertainty and assumed to be symmetric about the nominal result. The covariance between bins or measurements is then given by

$$C = \vec{\sigma} \otimes \vec{\sigma}, \quad (9.11)$$

where $\otimes$ represents the outer product and $\vec{\sigma}$ is the signed vector of estimated uncertainties in the bins or measurements under consideration. All bins are thus considered to be fully correlated or fully anti-correlated.

- Systematic error sources estimated by varying an input one standard deviation about the nominal value. The covariance of such systematics is given by

$$C = \frac{\vec{\sigma}_+ \otimes \vec{\sigma}_+ + \vec{\sigma}_- \otimes \vec{\sigma}_-}{2}, \quad (9.12)$$

where $\vec{\sigma}_+$ and $\vec{\sigma}_-$ are the signed differences of the up and down variation with respect to the nominal result, respectively. The uncertainty is thus given by the quadratic mean of the difference of the result between the up and down variations to the nominal result.

All systematic uncertainties are taken as uncorrelated and summed in quadrature for the total systematic uncertainty. The total systematic covariance $C_{\text{syst}}^{\text{exp}}$ is given by

$$C_{\text{syst}}^{\text{exp}} = \sum_x C_{\text{syst}}^x, \quad (9.13)$$

where x runs over the individual sources of systematic uncertainty under consideration. Systematic variations that affect signal side normalisation: $\bar{B} \rightarrow X_c \ell \bar{\nu}$ branching fractions, lepton identification efficiencies, continuum calibration, secondary and fake lepton normalisation, the number of $B\bar{B}$ pairs produced and the fraction of $\Upsilon(4S)$ decaying to B^+B^- or $B^0\bar{B}^0$, are fully propagated to the tag calibration. Uncertainty due to the tag calibration is therefore not separately identified. Furthermore, as already discussed, to a good approximation, the tag calibration cancels in the ratio measurements.

9.5.1 Modelling

9.5.1.1 $\bar{B} \rightarrow X_u \ell \bar{\nu}$ modelling

As the simulation of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events is a hybrid composition of low-mass resonant and high-mass non-resonant states the relative contributions of the different states will impact

the reconstruction efficiency, and shape of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ template. We evaluate the uncertainty by increasing and decreasing the assumed branching fractions of the resonant decays, $\bar{B} \rightarrow (\pi, \rho, \omega, \eta, \eta') \ell \bar{\nu}$, by one standard deviation.

The assumed shape of the exclusive decays will likewise impact the measurement. These events have been generated using the SLPOLE model [167] for $\bar{B} \rightarrow (\pi, \rho, \omega) \ell \bar{\nu}$ decays and the ISGW2 [252] model for $\bar{B} \rightarrow (\eta, \eta') \ell \bar{\nu}$ decays. We update these models to the nominal models discussed in Ch. 7 using the **eFFORT** [253] framework which assigns event wise weights as

$$w_i = \frac{\Gamma_{\text{old}}}{\Gamma_{\text{new}}} \frac{d\Gamma_{\text{new}}(q^2)/dq^2}{d\Gamma_{\text{old}}(q^2)/dq^2}, \quad (9.14)$$

where Γ_{new} and Γ_{old} are the decay rates of the updated, and generated model respectively. The same procedure is used to vary the parameters of the form-factors by one standard deviation along the eigen-directions of their covariance matrix to estimate the modelling uncertainty for $\bar{B} \rightarrow (\pi, \rho, \omega) \ell \bar{\nu}$ decays. For $\bar{B} \rightarrow \eta \ell \bar{\nu}$ and $\bar{B} \rightarrow \eta' \ell \bar{\nu}$ decays we replace the nominal model with the alternate description of Ref. [254] and assign the full difference as the uncertainty.

For the non-resonant channels we vary M_b^{KN} and a^{KN} by their uncertainties in the eigen-directions of their covariance matrix. This is achieved by reweighting the nominal sample to alternate descriptions based on their three-dimensional generator-level distributions in E_ℓ^B, M_X, q^2 in the binning given in Eq. 7.2. For each variation, one million events were generated. To estimate the impact of the choice of model we reweight the nominal sample, generated following the prescription of DFN, to BLNP [39] using the same procedure as for the DFN variations. The parameters used in the BLNP model are $M_b^{\text{SF}} = 4.61 \text{ GeV}$ and $\mu_\pi^{2\text{SF}} = 0.2 \text{ GeV}^2$. The full difference to the nominal result is taken as the uncertainty and is assumed to be symmetric. We vary the assumed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yield by $\pm 1\sigma$.

For each variation investigated, for either the exclusive or inclusive component, the hybrid weights are recalculated following Eq. 7.1.

The cross-feed fraction of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events into the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ depleted sample depends on the production rate of K^+ and K_S^0 in the fragmentation of the X_u system. We vary the relative weight of simulated $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events generated with kaon pairs, ie. $g \rightarrow s\bar{s}$, by $\pm 25\%^2$.

9.5.1.2 $\bar{B} \rightarrow X_c \ell \bar{\nu}$ modelling

If the nominal branching fraction of a $\bar{B} \rightarrow D^{(*)} \ell \bar{\nu}$ decay differs from the value used in generating the MC we reweight the events accordingly. As the total $\bar{B} \rightarrow X_c \ell \bar{\nu}$ rate

²The ratio of kaon production in the hadronisation of the inclusive system is controlled by the **PYTHIA** parameter γ_s . The production rate of $e^+e^- \rightarrow s\bar{s}$ has been studied in Refs. [255, 256] at centre-of-mass energies of 12 and 36 GeV finding $\gamma_s = 0.35 \pm 0.05$ and $\gamma_s = 0.27 \pm 0.06$. Within the Belle simulation a value of 0.3 is used, consistent with both measurements. Ref. [73] adopts an uncertainty of ± 0.09 on this value. Events generated varying γ_s by this uncertainty are found to have varied kaon pair productions of -24% , $+27\%$, consistent with the variation considered within this analysis.

has decreased since the Belle MC sample was generated, we also update the relative contribution of double semileptonic events, in which both B mesons decay semileptonically.

We vary the values of the branching fractions of $\bar{B} \rightarrow D^{(*)} \ell \bar{\nu}$ about the nominal value in the simulation, assuming a Gaussian error profile. For the unmeasured gap channels, $\bar{B} \rightarrow D_{\text{Gap}}^{**} (\rightarrow D^{(*)} \eta) \ell \bar{\nu}$, we assume a uniform error profile between zero and twice the nominal branching fraction. For both cases, the uncertainty is taken to be half the 15.9 – 84.1% inter-percentile range of the obtained distribution.

The decays of $\bar{B} \rightarrow D^{(*)} \ell \bar{\nu}$ are modelled by the parameterisation of Caprini, Lellouch and Neubert [257] within the Belle MC. We update these to the nominal models discussed in Ch. 7 using the **eFFORT** framework. For $\bar{B} \rightarrow D \ell \bar{\nu}$ events the reweighting is one-dimensional in the hadronic recoil variable w . For $\bar{B} \rightarrow D^* \ell \bar{\nu}$ the reweighting is four-dimensional considering also θ_ℓ , the angle between the lepton and the B in the W frame, θ_ν , the angle between the child D of the D^* and the B in the D^* rest frame, and χ , the angle between the two planes formed by the decay products of the W and D^* .

Semileptonic decays to the orbitally excited D^{**} mesons are replaced in the generic MC cocktail with updated samples using masses and widths of the D^{**} from Ref. [211]. The events are generated following the ISGW2 model [252], as this is the only model supported within the Belle software. To update the generated events to the nominal model [221], we adopt the reweighting procedure of Refs. [258, 259] as the decays are not supported within the **eFFORT** framework. The procedure uses samples of the four channels generated with the Belle II software to produce either one, two, or three dimensional histograms in w (D_0^*), w and $\cos \theta_\ell$ (D_1^*, D_2^*), or w , $\cos \theta_\ell$, and $\cos \theta_\nu$ (D_1) using the same model as used to generate the Belle MC, and using the updated model and parameters. The distribution is then updated by means of weights taken from radial basis function based interpolations of the ratio of these histograms.

For each channel, we repeat the procedure, varying the form factors up and down by one standard deviation along the eigen-directions of the covariance matrix. The uncertainties of all eigen-directions are summed in quadrature.

9.5.1.3 Bias of data driven $\bar{B} \rightarrow X_u \ell \bar{\nu}$ template in q^2

As we do not reconstruct K_L^0 mesons, the momentum they carry away is attributed to the neutrino, biasing the reconstructed q^2 towards higher values. Charmed semileptonic events in which a K_L^0 is produced within the hadronic system are predominantly assigned to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched sample. Events in the enriched and depleted subsamples reconstructed within the same q^2 bin thus do not map to the same true q^2 distribution, causing a small bias in the data-driven template correction. This is illustrated in Fig. 9.12. The same applies to events in which the charged kaon or the children of the K_S^0 were not reconstructed, although such events are suppressed by the net event charge requirement.

We estimate the impact of this bias via an in-situ calibration of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ q^2 spectra. After calibration, the data-driven template must only correct the lepton energy, which, as previously discussed, is independent of the kaon identification and reconstruc-

tion.

The calibration is performed by unfolding the q^2 distribution of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted subsample after background subtraction, following the procedure which will be established in Ch. 10. A fourth-order Chebyshev polynomial is fit to the ratio of the normalised unfolded data yield to the normalised unfolded yield in simulation. The fitted polynomial is given in Fig. 9.13. Correction weights are applied to the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ simulated events during the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction by evaluating the polynomial at the generated q^2 of each event. The full difference to the nominal result is taken as a systematic uncertainty.

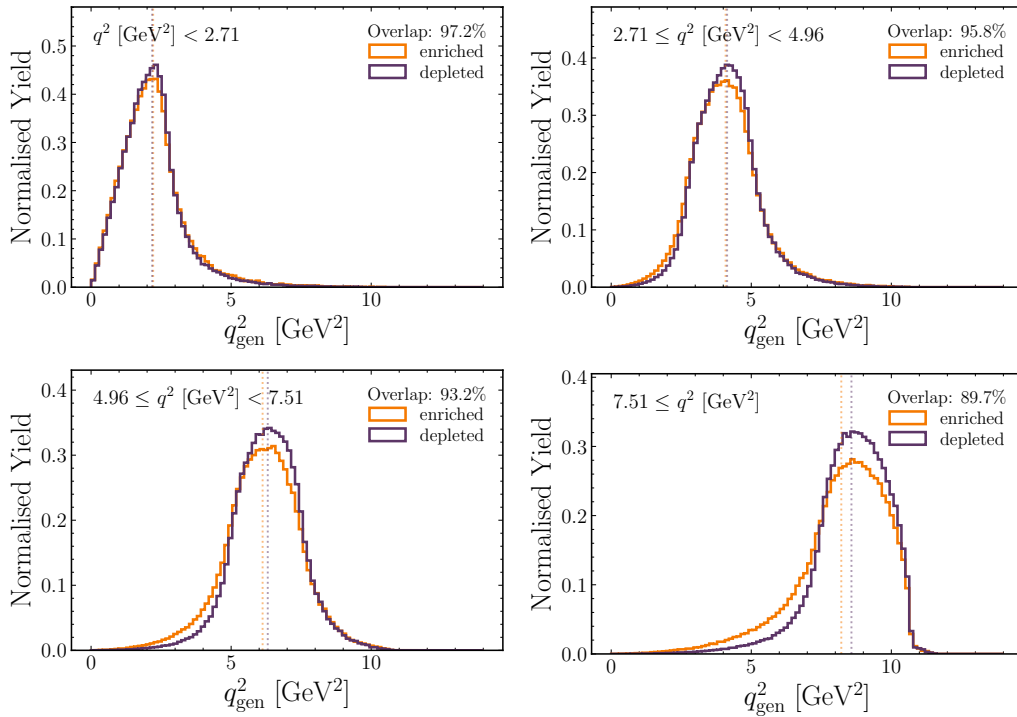


Figure 9.12: Generated q^2 spectrum for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events assigned to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples in the four bins of reconstructed q^2 considered in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction. Events assigned to the enriched subsample are biased towards a higher reconstructed q^2 than those in the depleted subsample due to the missing momentum of K_L^0 .

The method has been validated in simulation by injecting a data-MC disagreement in pseudo-data which approximates the observed disagreement in data. This is achieved by weighting $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events according to their generator-level quantities as

$$w = (0.15(E_{\ell, \text{gen}}^B - 1.3) + 1) \times (0.015(q_{\text{gen}}^2 - 4) + 1). \quad (9.15)$$

The $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yield recovered with the nominal procedure differs from the expected by 2.1%. Applying the in-situ calibration, the measured yield is within 0.4% of the ex-

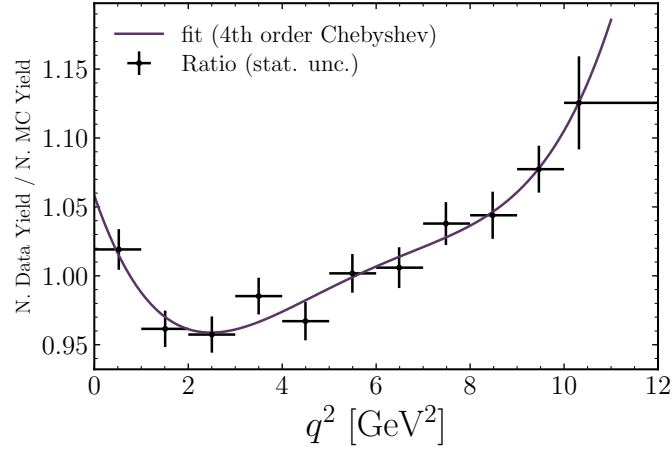


Figure 9.13: Ratio of unfolded yields in data and MC in bins of q^2 for $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events within the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted subsample, and the fourth-order polynomial fit to them. The x-position of each bin is given by the mean position of the bin contents. Only the statistical uncertainties are considered as the systematic uncertainties are expected to mostly cancel, and we consider the full difference to the nominal result as the systematic uncertainty.

pected value, having been overcorrected. The remaining non-closure is attributed to the polynomial interpolation.

9.5.1.4 Secondary and fake lepton normalisation

Contributions from events in which either a secondary lepton has been selected or a hadron has been misidentified as a lepton and has been selected represent a large source of background events in the analysis. The size and shape of these contributions are subject to two sources of uncertainty. The first is the performance of the lepton identification systems, which is discussed separately. The second is related to the production rate of the relevant channels. The events fall within three categories.

- The first is events in which the signal lepton is a misidentified hadron (fake leptons). The hadrons can originate from essentially all decay channels of the B meson. Many of these are poorly known, with significant fractions modelled by `PYTHIA`. There is thus uncertainty related to the number of hadrons produced and their kinematic distributions.
- The second category is events in which the signal lepton originates from a semileptonic D decay or leptonic J/ψ decay. The J/ψ production rate is small, and such events are suppressed by the additional lepton veto. As most B decay chains produce D mesons, the production rate is assumed to be relatively well modelled. The semileptonic D decay rates are measured to be $\mathcal{B}(D^+ \rightarrow X e \bar{\nu}) = (16.1 \pm 0.3)\%$,

$\mathcal{B}(D^+ \rightarrow X\mu\bar{\nu}) = (17.6 \pm 3.2)\%$, $\mathcal{B}(D^0 \rightarrow Xe\bar{\nu}) = (6.49 \pm 0.11)\%$, $\mathcal{B}(D^0 \rightarrow X\mu\bar{\nu}) = (6.8 \pm 0.6)\%$ [211]. The simulation assumes equal rates for the two lepton flavours using $\mathcal{B}(D^+ \rightarrow X\ell\bar{\nu}) = 16.1\%$ and $\mathcal{B}(D^0 \rightarrow X\ell\bar{\nu}) = 6.6\%$, in good agreement with the world averages. The rates at which individual channels contribute to the inclusive rates are generally well measured and well reproduced in the simulation. The decays are modelled via the prescription of Ref. [252].

- The third category is events in which the signal lepton originates from semitauonic decays, where the tauon has decayed leptonically, $\bar{B} \rightarrow X\tau(\rightarrow \ell\nu\bar{\nu})\bar{\nu}$. The production rate of $\bar{B} \rightarrow X\tau\bar{\nu}$ is not well known, with the simulation assuming a total decay rate of $\mathcal{B}(\bar{B} \rightarrow X\tau\bar{\nu}) = 2.7\%$. The world average branching fractions for the two sub-channels which have been measured, $\bar{B} \rightarrow D\tau\bar{\nu}$ and $\bar{B} \rightarrow D^*\tau\bar{\nu}$, are $(0.99 \pm 0.12)\%$ and $(1.50 \pm 0.08)\%$ [211], respectively. The $\bar{B} \rightarrow D^*\tau\bar{\nu}$ rate is in excellent agreement with the 1.48% assumed in simulation. The $\bar{B} \rightarrow D\tau\bar{\nu}$ rate is approximately 2σ above the 0.74% used in simulation. As the lepton energy spectrum is very soft, such events are largely rejected by the requirement $E_\ell^{B_{\text{sig}}} > 1 \text{ GeV}$.

Due to the similar shape of the three samples in the variables of interest, we combined all three samples into one component within the fitting procedure: the total yield of which is calibrated using the $M_X > 2 \text{ GeV}, 0.7 < E_\ell^{B_{\text{sig}}} < 1.8 \text{ GeV}$ control region. Due to this calibration, the measurement is insensitive to the assumed normalisation of the samples. It is, however, sensitive to the shape of the combined component. We assign a 30% uncertainty to the normalisation of each of the three contributing samples, effectively varying the shape of the total contribution, and estimate the uncertainty by performing pseudo-experiments. This uncertainty thus also considers the error in the extrapolation of the calibration from the control region to the $E_\ell^{B_{\text{sig}}} > 1 \text{ GeV}$ region.

9.5.1.5 $\Upsilon(4S)$ branching fraction

The rate at which the $\Upsilon(4S)$ decays to charged or neutral B pairs, $f_{+-} = \mathcal{B}(\Upsilon(4S) \rightarrow B^+B^-)/(\mathcal{B}(\Upsilon(4S) \rightarrow B^+B^-) + \mathcal{B}(\Upsilon(4S) \rightarrow B^0\bar{B}^0))$, can impact the measurement, as these channels are reconstructed with different efficiencies, and thus the composition of the ad-mixture can vary the shape and efficiency of the reconstruction. Therefore, we vary the rate by the PDG uncertainty, $f_{+-} = (51.4 \pm 0.6)\%$ [211]. The effect is expected to be absorbed into the tag calibration, up to the cross-feed contribution of charged events reconstructed as neutral B tags and vice versa.

9.5.1.6 Number of $B\bar{B}$ pairs

We vary the $B\bar{B}$ yield in MC by the uncertainty on the number of $B\bar{B}$ pairs observed at Belle, $(771 \pm 11) \times 10^6$. As we measure the ratio of two $B\bar{B}$ processes, the uncertainty is expected to cancel, barring small effects due to the change to the relative yield of $B\bar{B}$ and continuum events. The systematic thus mostly serves to sanity-check the procedure.

9.5.1.7 Continuum modelling

The uncertainty on the off-resonance calibration of the continuum contribution is estimated by varying the coefficients of the linear calibration functions within their uncertainty for 200 trials.

9.5.1.8 Limited MC sample size

The uncertainty due to the limited size of the MC samples used, which affects the shapes of the templates used in the fits, and the efficiency estimation is estimated via bootstrapping 200 trials.

9.5.2 Detector effects

Several systematic uncertainties are related to the performance of the detector and the difference in performance between data and simulated events.

9.5.2.1 Lepton identification

The performance of the lepton identification systems for electron and muon identification has been studied using $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ events in bins of lepton momentum magnitude and polar angle. We correct the efficiency of the signal lepton by assigning weights given by the ratio of measured efficiencies in data to simulation. We propagate the uncertainty of this correction by drawing 200 sets of additional corrections where the statistical uncertainty in each bin is considered uncorrelated, and the systematic uncertainty to be fully correlated across all bins for each lepton flavour. The systematic uncertainties includes fluctuations of the efficiencies across different data collection periods, and differences to efficiencies measured in the busier environment of $\bar{B} \rightarrow XJ/\psi(\rightarrow \ell^+\ell^-)$ events.

The hadron to lepton misidentification rate has been studied using a sample of $D^* \rightarrow D(\rightarrow K\pi)\pi_{\text{slow}}$ [55] in bins of track momentum magnitude and polar angle. We correct the misidentification rate in simulation and propagate the statistical uncertainty of the correction, which dominates, using 200 pseudo-experiments.

9.5.2.2 Kaon identification

Charged kaon identification and pion-to-kaon misidentification rates have been studied using a sample of $D^* \rightarrow D(\rightarrow K\pi)\pi_{\text{slow}}$ for the selection $P_{K/\pi} > 0.6$ in bins of polar angle and momentum magnitude. As these studies focus on medium and high momentum tracks, we perform a re-analysis of $D^* \rightarrow D(\rightarrow K\pi)\pi_{\text{slow}}$ events to measure the low-momentum kaon identification performance, and study a $K_S^0 \rightarrow \pi^+\pi^-$ sample to measure the low-momentum pion-to-kaon misidentification rate. These studies are discussed in appendix B.

We correct the rate in simulation by identifying bins in which the identification efficiency in simulation exceeds the efficiency in data and relabelling kaons as pions with

probability

$$P = \frac{\epsilon_{\text{MC}} - \epsilon_{\text{Data}}}{\epsilon_{\text{MC}}}, \quad (9.16)$$

where ϵ_{MC} and ϵ_{Data} are the identification efficiencies in simulation and data respectively. If the efficiency in data exceeds that in simulation we relabel kaons which were misidentified as pions in the simulation to kaons with probability:

$$P = \frac{\epsilon_{\text{Data}} - \epsilon_{\text{MC}}}{1 - \epsilon_{\text{MC}}}. \quad (9.17)$$

Once relabelled we update the total kaon count, thus migrating events between the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted samples, and recalculate the hadronic mass and missing mass squared of the event.

We propagate the uncertainty by drawing 200 sets of new ϵ_{Data} and ϵ_{MC} values, where again we consider the statistical uncertainty to be uncorrelated between bins and the systematic uncertainties to be fully correlated between bins.

The efficiency of the classifier used for K_S^0 selection has been studied using a sample of $D^{*+} \rightarrow D^0(\rightarrow K_S^0 \pi^0) \pi_{\text{slow}}^+$ events, providing efficiencies in bin of K_S^0 momentum. We correct the efficiency following a similar procedure as for charged kaons. We consider a true K_S^0 to be a false negative if both the child pions are assigned to the hadronic system but the K_S^0 has not been reconstructed or identified.

9.5.2.3 Slow pion tracking

The performance of the slow pion reconstruction has been studied using a sample of $B \rightarrow D^* \pi_{\text{slow}}^+$ events in six bins of pion momentum between 50 and 200 MeV. We correct the veto efficiency of the partial D^* reconstruction, and propagate the uncertainty on the slow pion reconstruction by performing 200 pseudo-experiments varying the correction factors by their uncertainty. The statistical uncertainty is taken to be uncorrelated between the six bins and the systematic uncertainty, which includes an uncertainty originating from the normalisation to the tracking efficiency for slow pion candidates with $p > 200$ MeV, to be fully correlated. As the performance depends strongly on the vertex detector and the inner layers of the tracking chamber we consider two sets of correction factors, one for data collected with SVD1, and one for data collected with SVD2. As the systematic uncertainty is dominated by the normalisation to the high momentum bin we consider the corrections for two subsamples to be fully uncorrelated.

As the reconstruction of a slow pion is independent of kaon identification, mismodelling of the identification rate is expected to affect the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ enriched and depleted subsamples similarly. Thus given the data-driven $\bar{B} \rightarrow X_c \ell \bar{\nu}$ template, the uncertainty on slow pion identification largely cancels.

9.5.2.4 Tracking

The efficiency of track finding has been studied using a sample of partially reconstructed $D^{*+} \rightarrow D^0(\rightarrow \pi^+ \pi^- K_S^0) \pi_{\text{slow}}^+$ events. Excellent agreement is observed between data and

simulation, with an uncertainty of 0.35% assigned to each track. We propagate the uncertainty by assigning event-wise weights of x^N , where N is the number of tracks in the hadronic system plus one for the signal lepton, and x is drawn from a normal distribution of width 0.35%, and mean of one. A total of 200 pseudo-experiments are performed. Tracks on the tag side are excluded, as the tag calibration corrects for tracking efficiency differences.

9.6 Subsample results

We split the sample into two sets of orthogonal subsamples, first by lepton flavour and secondly by B flavour, and repeat the analysis. This allows us to test lepton flavour universality and search for isospin-breaking effects, as well as acting as a cross-check on the analysis. The measured ratios of partial branching fractions for the subsamples are summarised in Fig. 9.14. The results split by lepton flavour are compatible within 1.3σ , the results split by B flavour within 1.6σ .

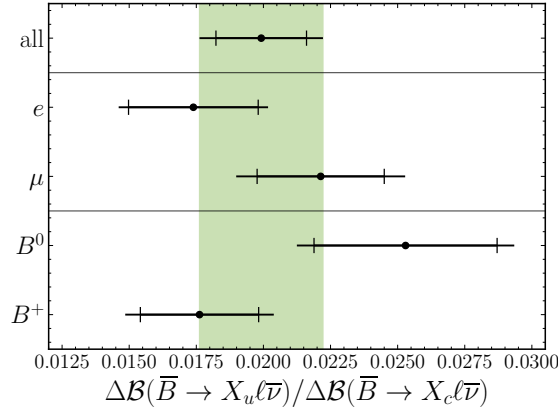


Figure 9.14: Measured ratios of partial branching fractions for the combined sample and orthogonal lepton flavour and B flavour subsamples.

9.6.1 Electron and muon subsamples

We test the expectation of lepton flavour universality by repeating the analysis procedure separately for the electron and muon modes. Systematic uncertainties are evaluated taking into account correlations between the lepton modes. The goodness of fits are $\chi^2/\text{ndf} = 15.3/14$ $\chi^2/\text{ndf} = 12.7/14$ for the electron and muon mode $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction fit, respectively, which are shown in Fig. 9.15. We obtain,

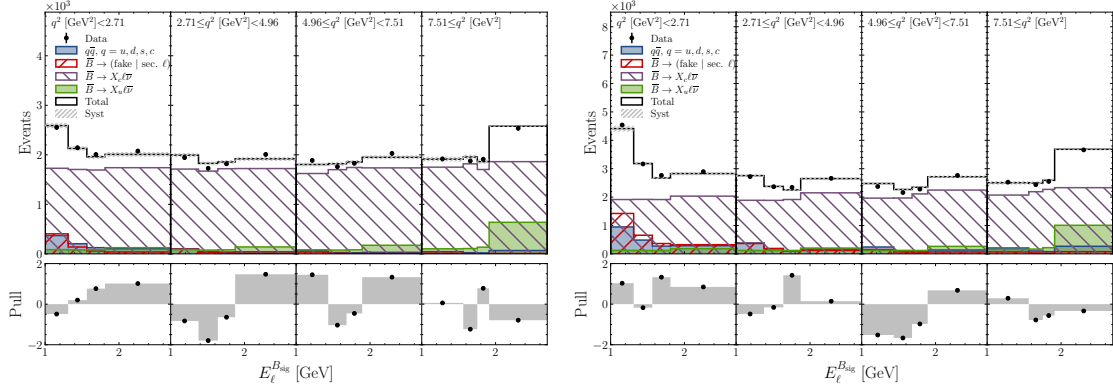


Figure 9.15: Fit to the $q^2 : E_\ell^{B_{\text{sig}}}$ distribution for the electron (*left*) and muon (*right*) subsamples.

$$\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u e \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c e \bar{\nu})} = (1.74 \pm 0.24_{\text{stat}} \pm 0.14_{\text{syst}}) \times 10^{-2}, \quad (9.18)$$

$$\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \mu \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \mu \bar{\nu})} = (2.21 \pm 0.24_{\text{stat}} \pm 0.21_{\text{syst}}) \times 10^{-2}, \quad (9.19)$$

where the expected cross-feed is negligible. We further compare directly the efficiency corrected yields of the charmed and charmless modes and find

$$R_{e/\mu}^{Xu} \equiv \frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u e \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \mu \bar{\nu})} = 0.79 \pm 0.13_{\text{stat}} \pm 0.07_{\text{syst}}, \quad (9.20)$$

$$R_{e/\mu}^{Xc} \equiv \frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_c e \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \mu \bar{\nu})} = 1.002 \pm 0.005_{\text{stat}} \pm 0.024_{\text{syst}}, \quad (9.21)$$

where the systematic uncertainty in the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ mode is dominated by lepton identification efficiency uncertainties and in the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ mode by the composition of the secondary and fake lepton component and MC statistics. The breakdown of the systematic uncertainties is given in Table 9.2. The normalisation of hadrons misidentified as leptons has been varied independently for the electron and muon channels. The charmless mode is compatible with unity within 1.4σ . We extrapolate the charmed mode to the full phase space via a correction factor of 1.00453 ± 0.00012 , extracted from simulation, finding

$$\frac{\mathcal{B}(\bar{B} \rightarrow X_c e \bar{\nu})}{\mathcal{B}(\bar{B} \rightarrow X_c \mu \bar{\nu})} = 1.007 \pm 0.005_{\text{stat}} \pm 0.025_{\text{syst}}, \quad (9.22)$$

which is in excellent agreement with the SM prediction, 1.006 ± 0.001 [260], and a recent measurement at Belle II that found $1.007 \pm 0.009_{\text{stat}} \pm 0.019_{\text{syst}}$ [261]. The systematic uncertainty of the Belle II measurement is also dominated by uncertainties in the performance of the lepton identification systems. This performance is somewhat better understood at the Belle II experiment than at its predecessor.

9.6.2 Isospin breaking effects

As discussed in Ch. 2, differences in the ratio of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ branching fractions in the charged and neutral B samples would indicate isospin breaking effects, such as weak annihilation, or more precisely, the valent weak annihilation contribution. We therefore consider the double ratio

$$R_{\text{iso}} = \frac{\Delta\mathcal{B}(B^+ \rightarrow X_u^0 \ell^+ \nu)}{\Delta\mathcal{B}(B^+ \rightarrow X_c^0 \ell^+ \nu)} \times \frac{\Delta\mathcal{B}(B^0 \rightarrow X_c^- \ell^+ \nu)}{\Delta\mathcal{B}(B^0 \rightarrow X_u^- \ell^+ \nu)}. \quad (9.23)$$

Isospin breaking effects in the charmed mode are expected to be negligible, with $\Gamma(B^0 \rightarrow X_c^- \ell^+ \bar{\nu}) = \Gamma(B^+ \rightarrow X_c^0 \ell^+ \nu)$ up to $\mathcal{O}(1/M_b^3)$ [262]. Taking this result, the above relation reduces to

$$R_{\text{iso}} = \frac{\tau_{B^0}}{\tau_{B^+}} \frac{\Delta\mathcal{B}(B^+ \rightarrow X_u^0 \ell^+ \nu)}{\Delta\mathcal{B}(B^0 \rightarrow X_u^- \ell^+ \nu)}. \quad (9.24)$$

FEI may reconstruct neutral tags in $B^+ B^-$ events, and charged tags in $B^0 \bar{B}^0$ events introducing a cross-feed between the samples. The measured $\bar{B} \rightarrow X_{u/c} \ell \bar{\nu}$ yields can be expressed in terms of the true yields as

$$N_{\text{measured}}^{X_q^0 \ell^+ \nu} = \mathcal{P}_{B^+ \rightarrow B^+}^q N_{\text{true}}^{X_q^0 \ell^+ \nu} + \mathcal{P}_{B^0 \rightarrow B^+}^q N_{\text{true}}^{X_q^- \ell^+ \nu}, \quad (9.25)$$

$$N_{\text{measured}}^{X_q^- \ell^+ \nu} = \mathcal{P}_{B^0 \rightarrow B^0}^q N_{\text{true}}^{X_q^- \ell^+ \nu} + \mathcal{P}_{B^+ \rightarrow B^0}^q N_{\text{true}}^{X_q^0 \ell^+ \nu}, \quad (9.26)$$

where $N_{\text{true}}^{X_q^- \ell^+ \nu}$ and $N_{\text{true}}^{X_q^0 \ell^+ \nu}$ denote the true $\bar{B} \rightarrow X_u \ell \bar{\nu}$ or $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields and $\mathcal{P}_{B^a \rightarrow B^b}^q$ the probability for the FEI algorithm to reconstruct a B meson of charge a when the true meson has charge b for either the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events or the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events. These probabilities are extracted from simulation, finding central values of:

$$\begin{aligned} \mathcal{P}_{B^0 \rightarrow B^0}^u &= 0.89, & \mathcal{P}_{B^+ \rightarrow B^+}^u &= 0.96, \\ \mathcal{P}_{B^0 \rightarrow B^0}^c &= 0.79, & \mathcal{P}_{B^+ \rightarrow B^+}^c &= 0.95. \end{aligned}$$

We correct for the cross-feed and assign the full difference as a systematic uncertainty. The measured ratios of partial branching fractions are:

$$\frac{\Delta\mathcal{B}(B^+ \rightarrow X_u \ell^+ \nu)}{\Delta\mathcal{B}(B^+ \rightarrow X_c \ell^+ \nu)} = (1.76 \pm 0.22_{\text{stat}} \pm 0.17_{\text{syst}}) \times 10^{-2}, \quad (9.27)$$

$$\frac{\Delta\mathcal{B}(B^0 \rightarrow X_u \ell^+ \nu)}{\Delta\mathcal{B}(B^0 \rightarrow X_c \ell^+ \nu)} = (2.53 \pm 0.34_{\text{stat}} \pm 0.22_{\text{syst}}) \times 10^{-2}, \quad (9.28)$$

$$R_{\text{iso}} = 0.70 \pm 0.13_{\text{stat}} \pm 0.07_{\text{syst}}, \quad (9.29)$$

where the dominant systematic on R_{iso} is due to the MC statistical error and we find $\chi^2/\text{ndf} = 13.8/14$ and $16.2/14$ for the charged and neutral B sample $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction fits, respectively. The post-fit distributions are shown in Fig. 9.16. The breakdown of the systematic uncertainties is given in Table 9.2. This value of R_{iso} is 2.0σ apart from the

expectation of equal semileptonic rates for both isospin states. The relative contribution of weak annihilation to the total can be defined as

$$\frac{\Gamma_{\text{WA}}}{\Gamma} = \frac{f_u}{f_{\text{WA}}}(R_{\text{iso}} - 1), \quad (9.30)$$

where Γ is the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay width and f_u, f_{WA} give the fraction of phase space considered in the measurement of R_{iso} for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and weak annihilation, respectively. Weak annihilation is expected to be confined to the high q^2 region, $q^2 \approx M_B^2$, we therefore assume $f_{\text{WA}} \approx 1$. The value of f_u is extracted from MC to fully propagate shape varying uncertainties. The central value is $f_u = 0.866$ following the prescription of the DFN model [37] utilized in the construction of the hybrid MC. We find a central interval of $-0.54 < \frac{\Gamma_{\text{WA}}}{\Gamma} < -0.02$ at 90% confidence level. This is consistent with recent results from Belle and BaBar which set 90% confidence levels at $-0.14 < \frac{\Gamma_{\text{WA}}}{\Gamma} < 0.17$ [73] and $-0.17 < \frac{\Gamma_{\text{WA}}}{\Gamma} < 0.19$ [72], respectively. We stress that set using this method, the limits only consider the valent weak annihilation contribution. However, a value of $R_{\text{iso}} < 1$, as observed, suggests that the weak annihilation contribution to the neutral B decay rate exceeds the contribution to the decay rate of charged B mesons. This is consistent with the studies of inclusive semileptonic D decays of Refs. [92, 98].

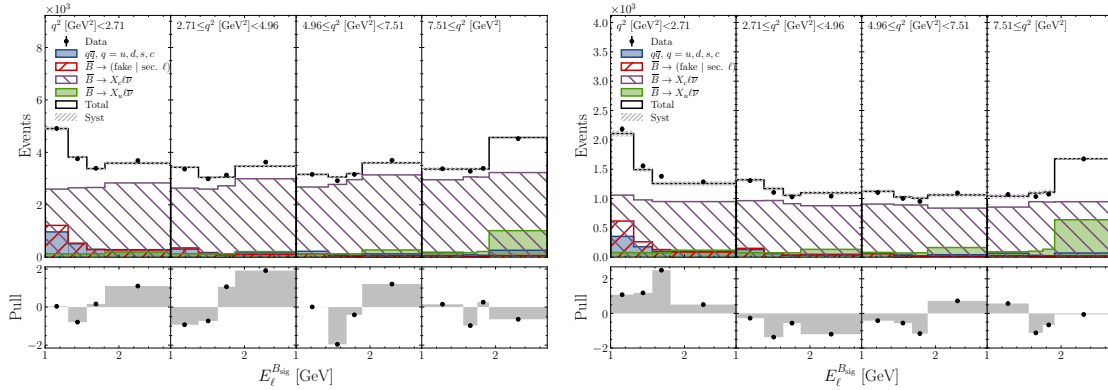


Figure 9.16: Fit to the $q^2 : E_\ell^{B_{\text{sig}}}$ distribution for the charged B (left) and neutral B (right) subsamples.

9.7 Missing mass squared threshold scan

As the modelling of missing mass squared, one of the most powerful selection variables for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction subsample, is poor, we test the stability of the measurement as a function of the threshold applied to the absolute missing mass squared. The measured $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ values, and the pairwise pulls of the results are presented in Fig. 9.17. For all thresholds, the in situ q^2 calibration systematic uncertainty is estimated using the weights derived at the nominal 0.43 GeV^2 threshold. The impact of this is

Result	e	μ	$R_{e/\mu}^{Xu}$	$R_{e/\mu}^{Xc}$	B^+	B^0	R_{iso}
Central Value $\times 100$	1.74	2.21	78.7	100.2	1.76	2.53	69.7
Stat. Error (Data)	13.9	10.7	16.6	0.5	12.5	13.5	19.3
$\mathcal{B}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta' \ell \bar{\nu})$	0.2	0.2	0.1	< 0.1	0.2	0.9	1.0
$\mathcal{FF}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta' \ell \bar{\nu})$	0.3	0.3	0.1	< 0.1	0.2	0.6	0.4
$\mathcal{B}(\bar{B} \rightarrow x_u \ell \bar{\nu})$	0.9	0.7	0.2	< 0.1	0.2	2.2	2.0
Hybrid Model (BLNP)	0.9	2.4	3.3	0.1	0.4	3.3	3.8
DFN ($M_b^{\text{KN}}, a^{\text{KN}}$)	5.1	5.4	0.3	< 0.1	5.3	4.8	0.6
$N_{g \rightarrow s\bar{s}}$	0.5	0.4	0.1	< 0.1	0.1	1.5	1.4
$\mathcal{B}(\bar{B} \rightarrow D \ell \bar{\nu})$	0.2	0.1	0.1	< 0.1	0.2	0.3	0.5
$\mathcal{B}(\bar{B} \rightarrow D^* \ell \bar{\nu})$	0.9	0.8	0.1	< 0.1	0.3	0.3	0.6
$\mathcal{B}(\bar{B} \rightarrow D^{**} \ell \bar{\nu})$	0.5	0.2	0.3	< 0.1	0.6	0.7	0.9
$\mathcal{B}(\bar{B} \rightarrow D^{(*)} \eta \ell \bar{\nu})$	0.3	0.2	0.4	< 0.1	0.2	0.6	0.8
$\mathcal{B}(\bar{B} \rightarrow D^{(*)} \pi \pi \ell \bar{\nu})$	0.3	0.1	0.4	< 0.1	0.3	0.4	0.5
$\mathcal{FF}(\bar{B} \rightarrow D \ell \bar{\nu})$	0.2	0.2	< 0.1	< 0.1	0.2	0.2	< 0.1
$\mathcal{FF}(\bar{B} \rightarrow D^* \ell \bar{\nu})$	1.0	0.8	0.2	< 0.1	1.2	0.3	0.9
$\mathcal{FF}(\bar{B} \rightarrow D^{**} \ell \bar{\nu})$	0.5	0.4	0.2	< 0.1	0.6	0.2	0.8
Sec.Fakes. Composition	1.4	5.6	5.3	0.4	5.1	2.0	3.1
In-situ q^2 Calibration	3.3	2.5	0.7	< 0.1	3.6	1.3	2.3
ℓ ID Efficiency	0.1	0.1	2.3	2.4	0.1	0.1	0.1
ℓ ID Fake Rate	0.1	0.4	0.4	0.1	0.4	0.2	0.2
$K\pi$ ID Efficiency	1.4	1.0	1.2	< 0.1	1.5	1.2	1.6
$K\pi$ ID Fake Rate	0.9	0.7	1.2	< 0.1	0.8	1.0	1.4
K_S^0 Efficiency	0.3	0.3	0.4	< 0.1	0.3	0.3	0.5
π_{slow} Efficiency	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Tracking	0.1	0.1	< 0.1	< 0.1	0.1	0.1	< 0.1
Continuum Calibration	0.2	0.6	0.7	0.2	0.6	0.2	0.6
N_{BB}	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
$f_{+/0}$	< 0.1	< 0.1	< 0.1	< 0.1	0.1	< 0.1	0.1
Stat. Error (MC)	4.1	3.3	5.8	0.2	3.4	3.9	5.2
FEI Crossfeed	-	-	-	-	2.3	2.8	5.3
Total Syst.	7.9	9.4	9.1	2.4	9.5	8.6	10.0

Table 9.2: Summary of the central value (R), statistical, and systematic uncertainties for the subsample based results. Columns marked e, μ, B^+, B^0 give the ratio of partial branching fractions of the charmed to charmless modes for the electron and muon, and charged and neutral B subsamples, respectively. The uncertainties are given as relative values on the central value in percent.

expected to be negligible. The nominal result lies slightly below the adjacent results; however, it agrees with all results within 2σ . The measured $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ values are approximately constant in the region of the nominal threshold, with only small variations on a scale much smaller than the measured uncertainties. Therefore, we consider the measured value of $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ to have no dependence on the applied $|M_{\text{Miss}}^2|$ threshold.

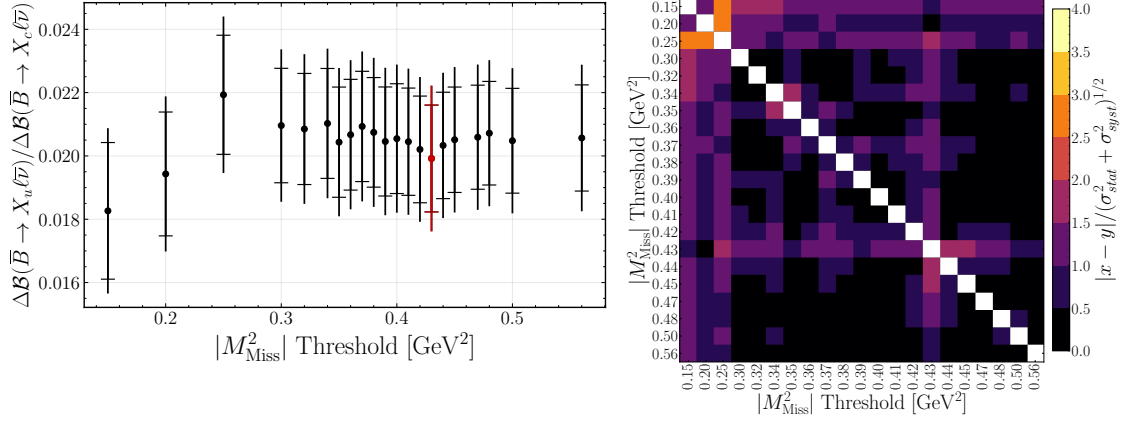


Figure 9.17: *Left:* Measured $\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})/\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ values as a function of $|M_{\text{Miss}}^2|$ threshold, the inner error bar gives the statistical uncertainty, the outer the total uncertainty. The nominal selection is shown in red. *Right:* pairwise pull between the different thresholds.

9.8 $|V_{ub}|$ and $|V_{ub}|/|V_{cb}|$

There are no calculations of the ratio of partial rates for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ to $\bar{B} \rightarrow X_c \ell \bar{\nu}$ decays that take into account the correlations in the input parameters, which would likely lead to partial cancellation of the uncertainties. In order to compare our result with the world average results, we therefore perform naive extractions of $|V_{ub}|/|V_{cb}|$ and $|V_{ub}|$.

The relationship between the measured ratio of partial branching fractions and the ratio of CKM element magnitudes is given by

$$\frac{|V_{ub}|}{|V_{cb}|} = \sqrt{\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} \frac{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})}}, \quad (9.31)$$

where $\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})$ and $\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})$ are the predictions of the partial rates for lepton energies above 1 GeV, omitting the CKM factors, of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ respectively.

For $\bar{B} \rightarrow X_u \ell \bar{\nu}$, we consider four predictions of the partial rate:

- those of BLNP, using the input parameters given in Sec. 9.5, $\Delta\Gamma^{\text{BLNP}}(\bar{B} \rightarrow X_u \ell \bar{\nu}) = 61.5^{+6.4}_{-5.1} \text{ ps}^{-1}$;

- those of GGOU, using as input $M_b^{\text{kin}} = 4.55 \pm 0.02 \text{ GeV}$ and $\mu_\pi^{2\text{kin}} = 0.45 \pm 0.08 \text{ GeV}^2$, finding $\Delta\Gamma^{\text{GGOU}}(\bar{B} \rightarrow X_u \ell \bar{\nu}) = 58.5_{-2.3}^{+2.7} \text{ ps}^{-1}$,
- those of DGE, using as input the b quark mass in the $\overline{\text{MS}}$ scheme, $M_b^{\overline{\text{MS}}} = 4.19 \pm 0.04 \text{ GeV}$, finding $\Delta\Gamma^{\text{DGE}}(\bar{B} \rightarrow X_u \ell \bar{\nu}) = 58.2_{-3.0}^{+3.6} \text{ ps}^{-1}$; and
- and those of ADFR, also taking $M_b^{\overline{\text{MS}}} = 4.19 \pm 0.04 \text{ GeV}$, finding $\Delta\Gamma^{\text{DGE}}(\bar{B} \rightarrow X_u \ell \bar{\nu}) = 61.5_{-5.1}^{+5.8} \text{ ps}^{-1}$.

These partial rates are consistent with those used in the most recent Belle result [73].

For $\bar{B} \rightarrow X_c \ell \bar{\nu}$, we extract the partial rate from the global fit to the moments of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ spectra in the kinetic scheme of Ref. [212], as

$$\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu}) = f_c \frac{\mathcal{B}^{\text{kin}}(\bar{B} \rightarrow X_c \ell \bar{\nu})}{\tau_B |V_{cb}^{\text{kin}}|^2} = 29.7 \pm 1.2 \text{ ps}^{-1}, \quad (9.32)$$

where $\tau_B = 1.579 \pm 0.004 \text{ ps}$ [212], $|V_{cb}^{\text{kin}}| = (42.19 \pm 0.78) \times 10^{-3}$, $\mathcal{B}^{\text{kin}}(\bar{B} \rightarrow X_c \ell \bar{\nu}) = (10.65 \pm 0.16)\%$, and we naively, and conservatively, assume $|V_{cb}^{\text{kin}}|$ and $\mathcal{B}^{\text{kin}}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ to be uncorrelated. The phase-space correction factor $f_c = 0.784 \pm 0.005$, is estimated from simulation. We measure

$$\begin{aligned} \frac{|V_{ub}|}{|V_{cb}|}^{\text{BLNP}} &= (9.81 \pm 0.42_{\text{stat}} \pm 0.38_{\text{syst}} \pm 0.51_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2}, \\ \frac{|V_{ub}|}{|V_{cb}|}^{\text{GGOU}} &= (10.1 \pm 0.43_{\text{stat}} \pm 0.39_{\text{syst}} \pm 0.23_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2}, \\ \frac{|V_{ub}|}{|V_{cb}|}^{\text{DGE}} &= (10.1 \pm 0.43_{\text{stat}} \pm 0.40_{\text{syst}} \pm 0.31_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2}, \\ \frac{|V_{ub}|}{|V_{cb}|}^{\text{ADFR}} &= (9.81 \pm 0.42_{\text{stat}} \pm 0.38_{\text{syst}} \pm 0.41_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2}, \end{aligned}$$

where the third and fourth uncertainties correspond to the uncertainty of the partial rate predictions for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$, respectively. The uncertainty on the partial rate predictions for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ has been symmetrised by taking the larger of the asymmetric uncertainties. These values are in excellent agreement with the world averages of both the inclusive and exclusive determinations of $|V_{ub}|/|V_{cb}|$ presented in Ch. 2. A comparison of the result determined using the GGOU partial rates, and the inclusive and exclusive averages as determined by the Heavy Flavor Averaging Group is given in Fig. 9.18, reproducing Fig. 66 of Ref. [66].

We also extract $|V_{ub}|$, with the relation

$$|V_{ub}| = \sqrt{\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} \frac{\Delta\mathcal{B}^{\text{ex.}}(\bar{B} \rightarrow X_c \ell \bar{\nu})}{\tau_B \Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})}}, \quad (9.33)$$

where $\tau_B = 1.579 \pm 0.004 \text{ ps}$ [212] is the average B meson lifetime, and $\Delta\mathcal{B}^{\text{ex.}}(\bar{B} \rightarrow X_c \ell \bar{\nu})$ is the externally measured partial branching fraction of $\bar{B} \rightarrow X_c \ell \bar{\nu}$. For the latter, we take

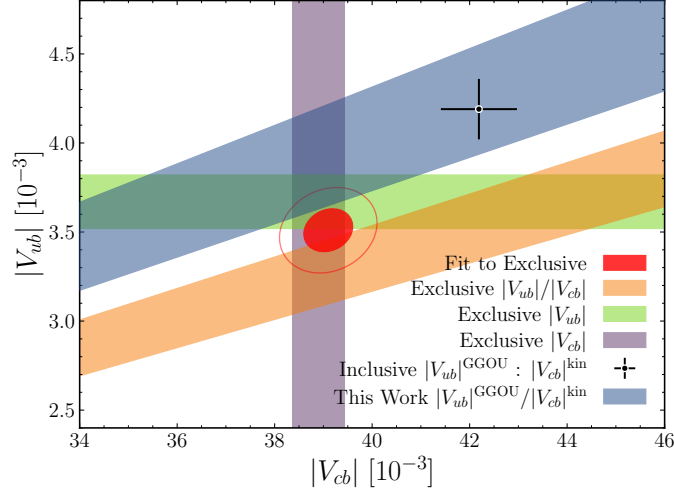


Figure 9.18: Comparison of ratio of $|V_{ub}|^{\text{GGOU}}/|V_{cb}|$, as measured in this work, with the averages for inclusive and exclusive $|V_{ub}|$ and $|V_{cb}|$ as in Fig. 66 of Ref. [66]. Note that to be consistent with Ref. [66], these are the 2021 averages of the Heavy Flavor Averaging Group. For $|V_{cb}|$ these differ slightly from the 2023 averages of the Particle Data Group discussed in Ch. 2.

the average of the values reported by the Belle and BaBar collaborations in Refs. [59, 62], assuming them to be fully uncorrelated, finding $\Delta\mathcal{B}^{ex.}(\bar{B} \rightarrow X_c \ell \bar{\nu}) = (8.55 \pm 0.13)\%$. We measure

$$\begin{aligned}
 |V_{ub}|^{\text{BLNP}} &= (4.19 \pm 0.18_{\text{stat}} \pm 0.17_{\text{syst}} \pm 0.22_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.03_{\Delta\mathcal{B}^{ex.}(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-3}, \\
 |V_{ub}|^{\text{GGOU}} &= (4.29 \pm 0.18_{\text{stat}} \pm 0.17_{\text{syst}} \pm 0.10_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.03_{\Delta\mathcal{B}^{ex.}(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-3}, \\
 |V_{ub}|^{\text{DGE}} &= (4.31 \pm 0.18_{\text{stat}} \pm 0.17_{\text{syst}} \pm 0.13_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.03_{\Delta\mathcal{B}^{ex.}(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-3}, \\
 |V_{ub}|^{\text{ADFR}} &= (4.19 \pm 0.18_{\text{stat}} \pm 0.17_{\text{syst}} \pm 0.17_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.03_{\Delta\mathcal{B}^{ex.}(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-3},
 \end{aligned}$$

which are in excellent agreement with the respective inclusive world average values, as shown in Fig. 9.19.

Finally, it is useful to compare the analysis with the most recent Belle study [73], as both analyses are performed on the full Belle data sample. If the two measurements are not highly correlated, it will be possible to combine the estimates of $|V_{ub}|$, improving the precision. In the same phase-space region, $E_\ell^{B_{\text{sig}}} > 1 \text{ GeV}$, Ref. [73] measures $\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu}) = (1.59 \pm 0.07_{\text{stat}} \pm 0.16_{\text{syst}}) \times 10^{-3}$, finding $1800 \pm 80_{\text{stat}} \pm 120_{\text{syst}}$ $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events compared to the $5430 \pm 450_{\text{stat}} \pm 350_{\text{syst}}$ found within this analysis. The statistical uncertainties of the two analyses are thus expected to have only limited correlation. The difference in yields is attributed to the more aggressive background rejection approach of Ref. [73] and the use of the Full Reconstruction algorithm [263], the hadronic tagging algorithm used in Belle before the more efficient FEI algorithm. The leading systematic uncertainties of both analyses are the uncertainties related to the DFN model paramet-

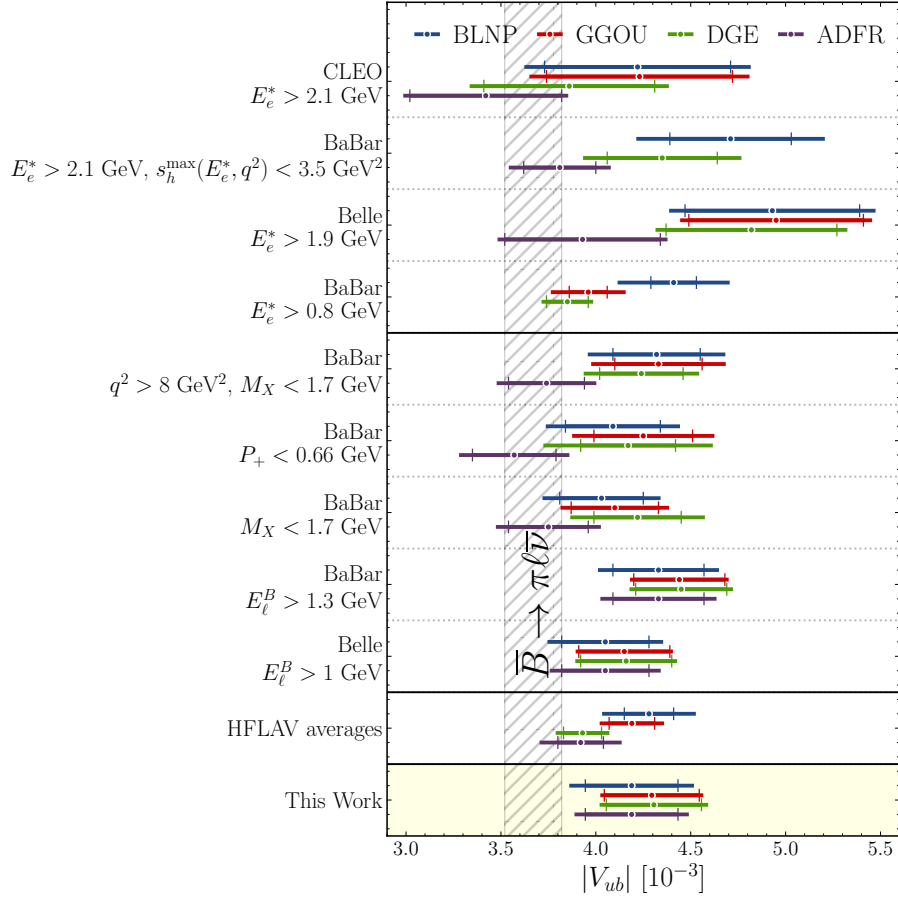


Figure 9.19: Comparison of the measured $|V_{ub}|$ value with the world average [66] and measurements that currently enter the world average [25, 28, 71–74]. The inner error bars are experimental, the outer experimental and theoretical. The uncertainties have been naively added in quadrature.

ers ($M_b^{\text{KN}}, a^{\text{KN}}$). For both measurements, the second and third leading uncertainties are combined approximately of the same size as the uncertainty associated to the DFN model parameters. For Ref. [73], these are the choice of inclusive model (DFN or BLNP), and the assumed $\bar{B} \rightarrow X_u \ell \bar{\nu}$ branching fraction and how it interacts with the hybrid model (labelled $\mathcal{B}(\bar{B} \rightarrow x_u \ell \bar{\nu})$ within this thesis). For the presented analysis, they are the composition of the secondaries and fake component, and the in-situ q^2 calibration. Both sets of uncertainties are small or not relevant in the other analysis. The systematic uncertainties between the two measurements are thus expected to have only low to moderate correlations. As both the statistical and systematic uncertainties are expected to be only moderately correlated between the two measurements, a statistical combination of the two results is expected to provide the most accurate inclusive determination of $|V_{ub}|$ by the Belle experiment.

Ratio of spectra

We project the samples into lepton energy and momentum transfer squared and extract the differential ratio of partial branching fractions.

As discussed in Ch. 2, differential measurements can be used to reduce the model dependence of inclusive $|V_{ub}|$ determinations, and can help constrain $\bar{B} \rightarrow X_u \ell \bar{\nu}$ models. Therefore, in addition to the ratio of partial branching fractions we report the differential ratio of partial branching fractions in q^2 and $E_\ell^{B_{\text{sig}}}$.

We measure the background subtracted yields of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ by projecting the respective samples into $E_\ell^{B_{\text{sig}}}$ and q^2 bins and subtracting all non $\bar{B} \rightarrow X_u \ell \bar{\nu}$ or $\bar{B} \rightarrow X_c \ell \bar{\nu}$ components, respectively. For the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ extraction, the bins are defined by the boundaries:

$$E_\ell^{B_{\text{sig}}} = [1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.7] \text{ GeV}, \quad \text{and} \\ q^2 = [0, 2, 4, 6, 8, 10, 12, 26] \text{ GeV}^2.$$

For the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extraction we merge the bins about the kinematic endpoint as $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events are kinematically forbidden in the final bins. The used bins are defined by the boundaries:

$$E_\ell^{B_{\text{sig}}} = [1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.7] \text{ GeV}, \quad \text{and} \\ q^2 = [0, 2, 4, 6, 8, 10, 26] \text{ GeV}^2.$$

All components (continuum, secondary and fake leptons, and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ or $\bar{B} \rightarrow X_u \ell \bar{\nu}$, respectively) are normalised to their respective yields during the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ extractions described in the previous chapter. No additional selection criteria are applied. For the measurement of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ differential yield, the shape of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ component is derived from the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ depleted subsample, following Eq. 9.2. The statistical and systematic uncertainties and correlations are estimated as described in the previous chapter. The background subtracted yields for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ in $E_\ell^{B_{\text{sig}}}$ and q^2 are given in Fig. 10.1.

10.1 Unfolding

The measured yields are affected by the resolution and acceptance of the detector, which causes migration between the bins, and the efficiency of the event reconstruction and

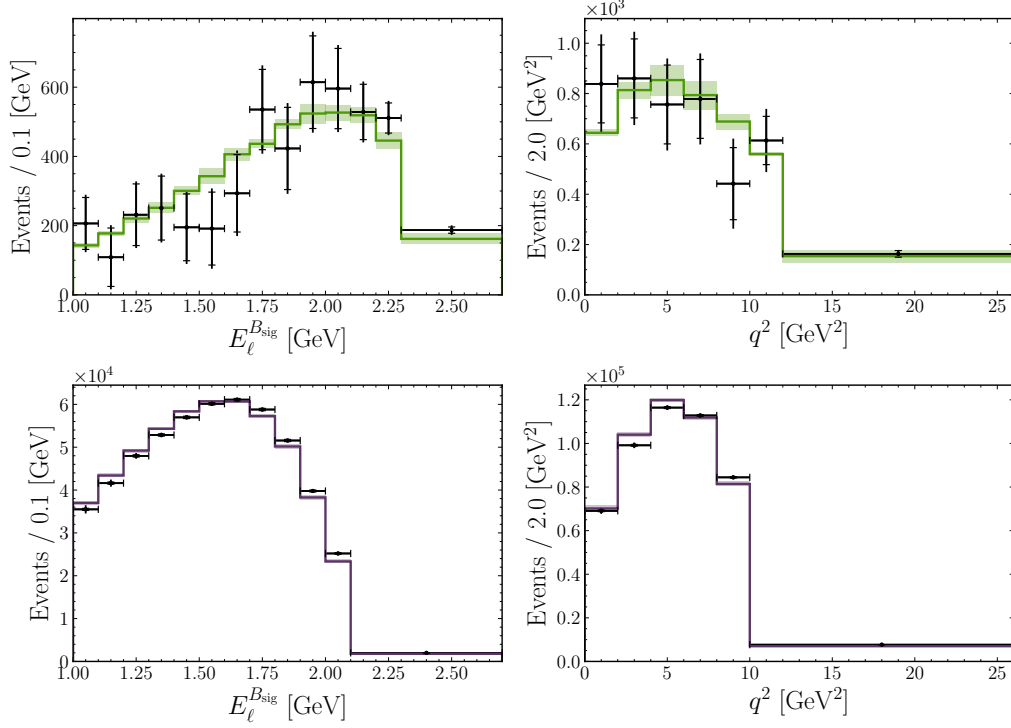


Figure 10.1: Background subtracted spectra (black markers) of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (top) and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (bottom) in $E_\ell^{B_{\text{sig}}}$ and q^2 . The spectra are compared with the expected distributions in simulation (green for $\bar{B} \rightarrow X_u \ell \bar{\nu}$, and purple for $\bar{B} \rightarrow X_c \ell \bar{\nu}$), which have been normalised to the respective data yield. For the markers, the inner error bar gives the statistical uncertainties, the outer error bar the sum in quadrature of the statistical and systematic uncertainties. For the histogram, the lighter shaded area gives the systematic uncertainty on the shape.

selection, which is not expected to be uniform. The yields in each bin thus cannot be compared directly with theory expectations or other experimental results. To do so, it is necessary to correct for the bin-to-bin migration in a process known as unfolding, and to correct for the reconstruction efficiency.

The measured yield in each bin of the reconstructed quantity, n_i , can be expressed as:

$$n_i = \sum_j A_{ij} y_j, \quad (10.1)$$

where y_j is the yield given in bins of the true variable, with index j , and A_{ij} an element of the migration matrix A , encoding the probability of events belonging in bin j to be reconstructed in bin i . In order to correct for the migration it is thus necessary to know the migration matrix. As it is not possible to extract this from data, it is extracted from simulation. The nominal migration matrices of the four spectra unfolded within this analysis are given in Fig. 10.2.

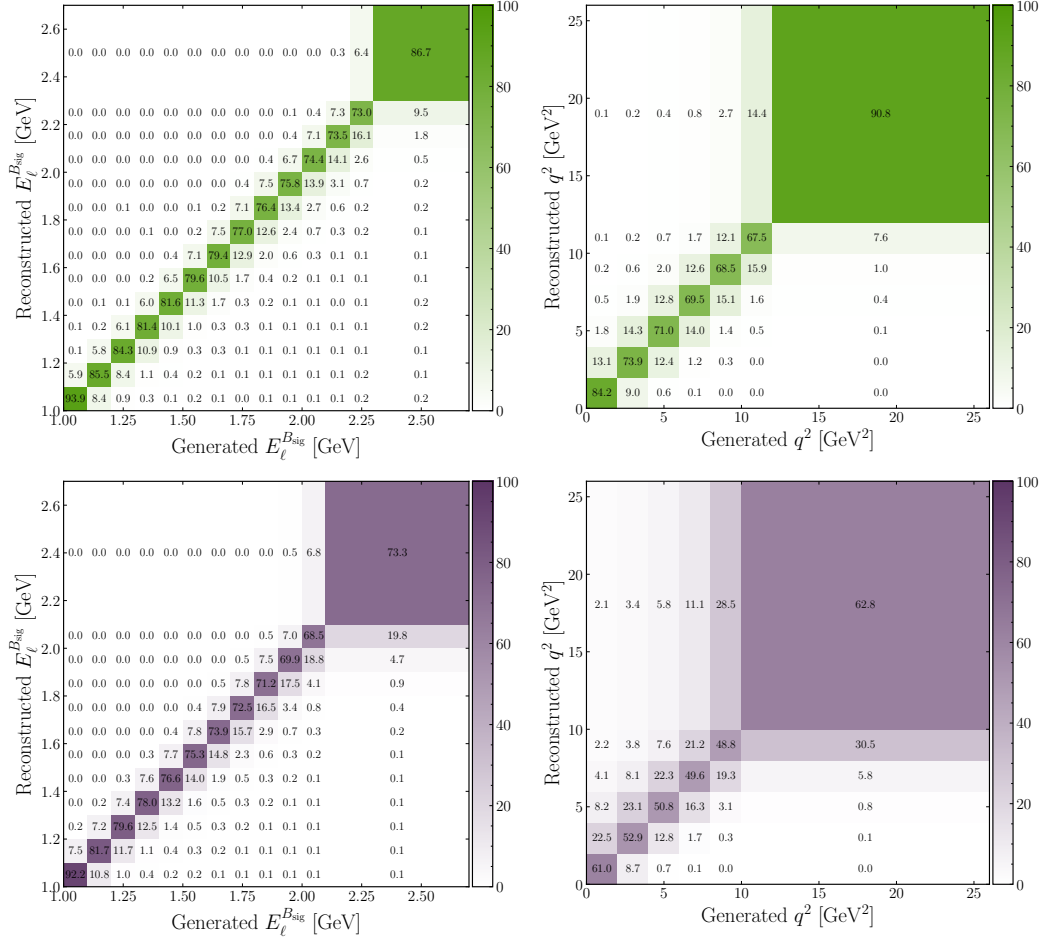


Figure 10.2: Nominal migration matrices for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (top) and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (bottom) in $E_\ell^{B_{\text{sig}}}$ and q^2 .

If the number of reconstructed and generated bins is equivalent, the simplest method to correct for the migration is to invert the migration matrix. The unfolded yields y are then given by:

$$y_j = \sum_i A_{ij}^{-1} n_i, \quad (10.2)$$

where A_{ij}^{-1} is an element of the inverted migration matrix. However, as the migration matrix can be badly conditioned, the unfolded distribution can exhibit large oscillations, with adjacent bins being almost fully anti-correlated, and next-to-adjacent bins almost fully correlated [264]. Several methods have been proposed that introduce regularisation, avoiding the oscillation by presupposing a smooth distribution. We use the method of Ref. [265, 266] known as SVD unfolding.

The SVD method minimises the expression

$$(\tilde{A}\vec{w} - \vec{n})^T C^{-1} (\tilde{A}\vec{w} - \vec{n}) + \tau (B\vec{w})^T (B\vec{w}), \quad (10.3)$$

where the first term treats Eq. 10.2 as a weighted least squares problem, and the second term introduces regularisation. The matrix C is the covariance matrix of the measured spectrum $\vec{n}$. The matrix $\tilde{A}$ and vector $\vec{w}$ are scaled versions of A and $\vec{y}$, given by:

$$\tilde{A}_{ij} = y_i^{\text{NC}} A_{ij}, \quad w_i = y_i / y_i^{\text{NC}}, \quad (10.4)$$

where y_i^{NC} is the expected distribution in simulation. This serves to scale the relative importance of the equations by the yields of the bins, rather than placing highly populated bins on equal footing to those with effectively no events. The strength of the regularisation is given by the parameter τ . The condition of the regularisation, that $\vec{w}$ should be smooth, is encoded in B , which takes the form

$$B = \begin{pmatrix} -1 & 1 & 0 & & \\ 1 & -2 & 1 & 0 & \\ 0 & 1 & -2 & 1 & 0 \\ & & \ddots & \ddots & \ddots \\ & & 0 & 1 & -2 & 1 \\ & & & 0 & 1 & -1 \end{pmatrix}. \quad (10.5)$$

One can further absorb the covariance matrix C into $\tilde{A}$ and $\vec{n}$ by means of singular value decomposition

$$\hat{A}_{ij} = \frac{1}{r_i} \sum_m Q_{im} \tilde{A}_{mj}, \quad \hat{n}_i = \frac{1}{r_i} \sum_m Q_{im} n_m, \quad (10.6)$$

where

$$C^{-1} = QR^{-1}Q^T, \quad r_i = \sqrt{R_{ii}}. \quad (10.7)$$

Minimising Eq. 10.3 leads to the expression

$$\begin{pmatrix} \hat{A}B^{-1} \\ \sqrt{\tau}I \end{pmatrix} B\vec{w} = \begin{pmatrix} \hat{\vec{n}} \\ 0 \end{pmatrix}. \quad (10.8)$$

Solving this expression gives

$$w = B^{-1}V\vec{z}, \quad (10.9)$$

where

$$z_i = \frac{d_i s_i}{s_i^2 + \tau}, \quad (10.10)$$

and V , d , and s originate from the singular value decomposition of $\hat{A}B^{-1} = USV^T$, where s_i are the diagonal elements of S , and $\vec{d} = U^T \hat{\vec{n}}$. For convenience S is ordered such that s_i are non-increasing as i increases.

In this form the effect of the regularisation is apparent. Terms at large i , which often correspond to quickly oscillating eigenvectors, without regularisation contribute to the unfolded spectrum as the inverse of s_i . With regularisation, they contribute as s_i/τ , if $s_i \ll \tau$, dampening their effects. The strength of the regularisation is usually set to the

square of the k th eigenvalue, $\tau = s_k^2$, where k gives the number of significant contributions. To choose k we perform a series of tests on pseudo-data, and examine the expected bias from the true distribution for each test. For the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ unfolding we consider five tests.

- u0 The nominal simulated distribution is used as pseudo-data and the nominal simulated migration matrix are used.
- u1 The simulated $\bar{B} \rightarrow X_u \ell \bar{\nu}$ distribution is re-weighted to the DFN parameters varied along the first eigen-direction, $\lambda_1 + \sigma_{\lambda_1}$.
- u2 The simulated $\bar{B} \rightarrow X_u \ell \bar{\nu}$ distribution is re-weighted to the prediction of BLNP.
- u3 The simulated $\bar{B} \rightarrow X_u \ell \bar{\nu}$ distribution is taken to be purely the inclusive component as modelled by the predictions of DFN, without including any resonant contributions.
- u4 The size of the off-diagonal entries of the migration matrix is increased by doubling the weight of simulated events with $|E_\ell^{B_{\text{sig}}} - E_{\ell, \text{gen}}^{B_{\text{sig}}}| > 0.1 \text{ GeV}$ or $|q^2 - q_{\text{gen}}^2| > 2 \text{ GeV}^2$ respectively. This probes differences in resolution between data and MC. The nominal $\bar{B} \rightarrow X_u \ell \bar{\nu}$ spectrum is used for testing.

The second and third test are variations within the expected uncertainty. The fourth and fifth tests are extreme variations, providing an insight into the stability of the unfolding against significant mismodelling. For the unfolding of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ spectra we consider three tests.

- c0 The nominal simulated distribution is used as pseudo-data and the nominal simulated migration matrix are used.
- c1 We double the contribution of the gap channels, $\bar{B} \rightarrow D^{(*)} \eta \ell \bar{\nu}$, in the pseudo-data $\bar{B} \rightarrow X_c \ell \bar{\nu}$ spectrum.
- c2 We double the weight of off-diagonal events, $|E_\ell^{B_{\text{sig}}} - E_{\ell, \text{gen}}^{B_{\text{sig}}}| > 0.1 \text{ GeV}$ or $|q^2 - q_{\text{gen}}^2| > 2 \text{ GeV}^2$ respectively, in the migration matrix. The nominal $\bar{B} \rightarrow X_c \ell \bar{\nu}$ spectrum is used for testing.

For each test, we generate 1000 toys, where the background subtracted yields are drawn from multinomial distributions centred on the MC expectations of the spectrum as varied for the test, using the covariance matrix of the nominal pseudo-data sample. The nominal covariance matrix is used as performing the full analysis chain for each variation is computationally expensive, and it is expected to be close to the covariance matrix that would be measured for each of the test samples. Tests varying each element of the covariance matrix by multiplicative factors drawn from normal distributions centred on 1 with width 0.25, were found to have only negligible impact on the measured bias of the unfolding.

For each test, three figures of merit are considered. The first two are the sum of bias, $\sum_i b_i$, and sum of absolute biases for all bins, $\sum_i |b_i|$, where the bias is defined as the difference between the unfolded yield in each bin for each toy and the true spectrum for the pseudo-data. The third figure of merit is the sum of absolute biases scaled by the inverse of the post-unfolding uncertainty of the event yield within the bin $\sum_i |b_i|/\sqrt{C_{ii}^{\text{unf}}}$, where C^{unf} is the covariance matrix of the unfolded bin yields. This gives an indication of the scale of the expected bias relative to the uncertainty on the yields. It however does not consider bin-to-bin correlations and thus should be taken only as an indication.

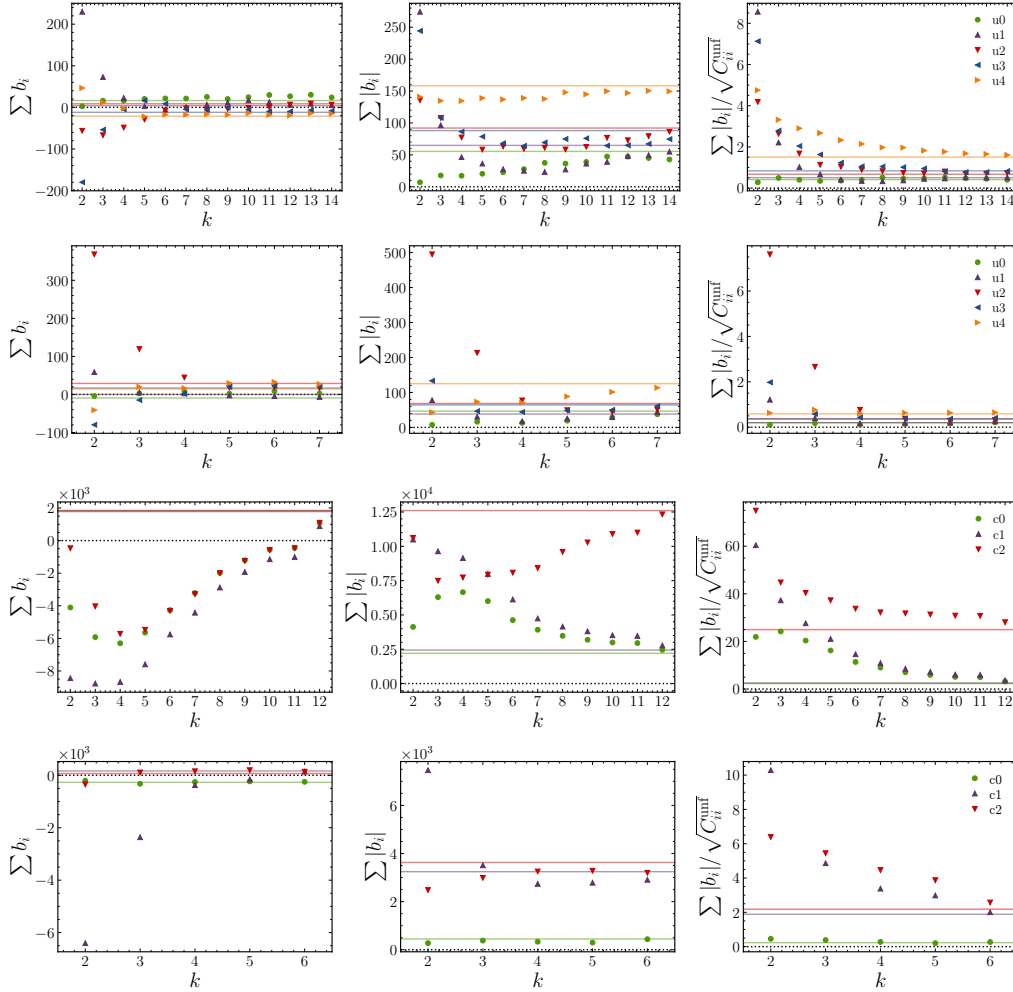


Figure 10.3: Figure for merit scans for the SVD unfolding regularisation parameter k . From top to bottom the figures correspond to the unfolding of the $\bar{B} \rightarrow X_u \ell \bar{\nu} E_\ell^{B_{\text{sig}}}$ and q^2 spectra, and $\bar{B} \rightarrow X_c \ell \bar{\nu} E_\ell^{B_{\text{sig}}}$ and q^2 spectra. From left to right the figures give the sum of biases ($\sum_i b_i$), sum of absolute biases ($\sum_i |b_i|$), and the sum of absolute biases scaled by the uncertainties ($\sum_i |b_i|/\sqrt{C_{ii}^{\text{unf}}}$). The solid horizontal lines give the figures of merit evaluated for unfolding via unregularised matrix inversion.

For the unfolding of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ spectra we select $k = 8$ and $k = 5$ for $E_\ell^{B_{\text{sig}}}$ and q^2 , respectively. These values correspond to the lowest k at which for all tests the sum of biases and sum of absolute biases approaches its steady state. For unfolding the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ q^2 spectrum we also select $k = 5$ with the same reasoning. For the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ $E_\ell^{B_{\text{sig}}}$ spectrum we observe that for the test in which the contribution of events with large residuals in $E_\ell^{B_{\text{sig}}}$ is increased (c2), the sum of absolute biases increases while the sum of biases approaches zero, indicating oscillatory behaviour. We select $k = 8$ to limit the oscillatory behaviour while allowing the expected total bias to approach zero.

Finally, as the nominal simulated sample represents the best knowledge of the decay processes under study, we consider the biases observed in unfolding the nominal spectra to be representative of the biases which will occur when unfolding the spectra measured in data. We therefore assign the relative size of the biases observed in tests $u0$ and $c0$ as a systematic uncertainty when unfolding the spectra measured in data. The biases are taken to be fully (anti-)correlated within an unfolded spectrum, and uncorrelated between the different considered spectra.

10.2 Ratio of spectra

We estimate the reconstruction efficiency in bins of the unfolded quantity, $\epsilon_i^{\text{unf.}X_u/c\ell\bar{\nu}}$, as the ratio of the number of reconstructed events in simulation with generated quantity falling within the bin, to the number of events generated within the bin. The differential ratio of partial branching fractions is then given by

$$\frac{\Delta_i \mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta_i \mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} = \frac{\epsilon_i^{\text{unf.}X_c\ell\bar{\nu}} N_i^{\text{unf.}X_u\ell\bar{\nu}}}{\epsilon_i^{\text{unf.}X_u\ell\bar{\nu}} N_i^{\text{unf.}X_c\ell\bar{\nu}}}, \quad (10.11)$$

where $N_i^{\text{unf.}X_u\ell\bar{\nu}}, N_i^{\text{unf.}X_c\ell\bar{\nu}}$ are the unfolded $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ yields in bin i . As previously discussed, the number of bins of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ spectra are not identical, with bins near the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ endpoint merged for the measurement of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ spectrum. To determine the ratio of spectra we therefore merge the efficiency corrected $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yields within these bins. The measured differential ratios of partial branching fractions are presented in Fig. 10.4, and their covariance in Fig. 10.5. A breakdown of the systematic uncertainties is given in Table 10.1 and Table 10.2. Statistical and systematic uncertainties are propagated via toy methods as discussed in the previous chapter. For each toy, in addition to all elements of the fit procedure discussed previously, we vary simultaneously (if applicable) the migration matrix. As the unfolding method requires the covariance matrix of the measured spectra (Eq. 10.3), and this cannot be calculated for each individual toy, we use the nominal covariance matrix for each toy. The effect of this is expected to be negligible.

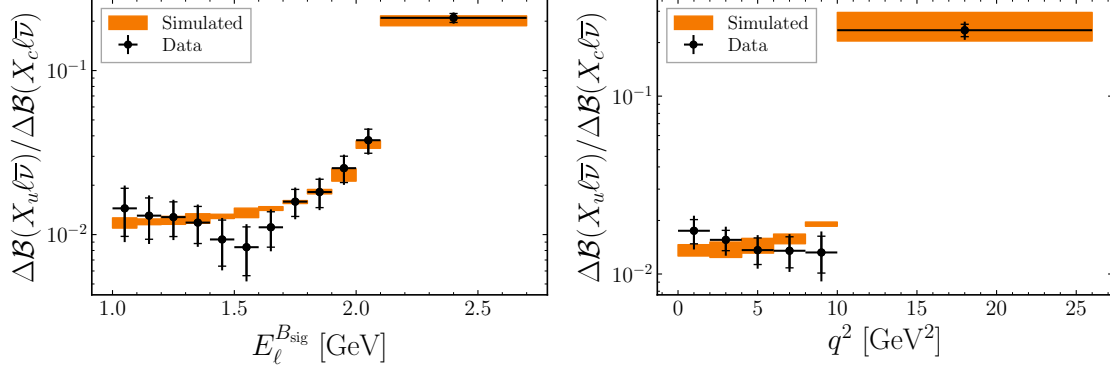


Figure 10.4: Unfolded ratios of partial branching fractions in the phase space $E_\ell^{B_{\text{sig}}} > 1 \text{ GeV}$. The simulated unfolded yields have been scaled to the data yields before taking the ratio. The uncertainty on the simulated bands includes all modelling uncertainties discussed in section 9.5 and the uncertainty due to the limited size of the MC sample.

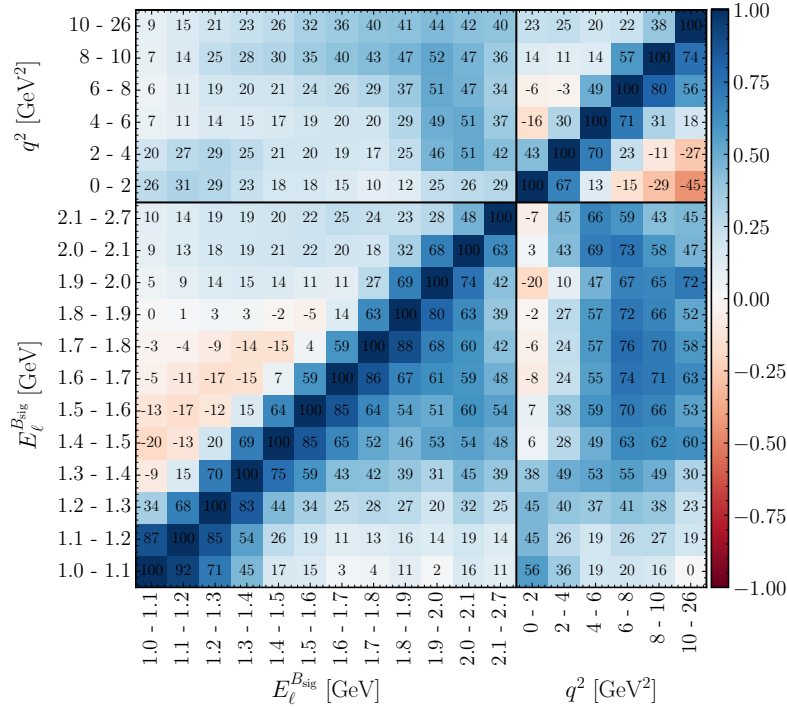


Figure 10.5: Global correlation matrix of the ratio of partial branching fractions in $E_\ell^{B_{\text{sig}}}$ and q^2 . Values above the diagonal give the correlation matrix of the statistical uncertainty, and values below the diagonal give the correlation matrix of the systematic uncertainty.

$E_\ell^{B_{\text{sig}}}$ Bin	0	1	2	3	4	5	6	7	8	9	10	11
Central Value $\times 100$	1.45	1.30	1.28	1.18	0.93	0.84	1.11	1.59	1.82	2.54	3.76	20.96
Stat. Error (Data)	32.5	28.2	23.9	25.5	31.3	33.0	24.3	18.8	19.7	18.3	16.7	6.3
$\mathcal{B}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta'\ell\bar{\nu})$	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2
$\mathcal{FF}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta'\ell\bar{\nu})$	0.3	0.3	0.4	0.4	0.6	0.7	0.6	0.4	0.4	0.3	0.3	0.1
$\mathcal{B}(\bar{B} \rightarrow x_u\ell\bar{\nu})$	1.4	1.4	1.2	1.2	1.3	1.2	0.9	0.7	0.7	0.7	0.5	0.5
Hybrid Model (BLNP)	6.4	2.1	3.3	3.8	1.7	1.2	1.8	1.0	0.5	4.0	0.2	0.2
DFN ($M_b^{\text{KN}}, a^{\text{KN}}$)	6.6	6.9	7.1	7.6	9.0	10.1	8.2	6.5	5.7	4.3	3.2	0.7
$N_{g \rightarrow s\bar{s}}$	0.3	0.6	0.6	0.8	1.6	2.1	1.3	0.7	0.7	0.4	0.4	0.1
$\mathcal{B}(\bar{B} \rightarrow D\ell\bar{\nu})$	0.4	0.4	0.3	0.3	0.2	0.5	0.5	0.4	0.6	0.5	0.2	0.1
$\mathcal{B}(\bar{B} \rightarrow D^*\ell\bar{\nu})$	0.3	0.4	0.6	0.7	1.0	1.3	1.2	1.1	1.1	1.2	1.0	0.6
$\mathcal{B}(\bar{B} \rightarrow D^{**}\ell\bar{\nu})$	0.6	0.7	0.8	0.6	0.6	0.5	0.4	1.1	1.3	0.9	0.8	0.2
$\mathcal{B}(\bar{B} \rightarrow D^{(*)}\eta\ell\bar{\nu})$	0.9	0.7	0.6	0.9	1.1	0.8	0.5	0.5	0.9	0.8	0.9	0.3
$\mathcal{B}(\bar{B} \rightarrow D^{(*)}\pi\pi\ell\bar{\nu})$	0.8	0.8	0.7	0.6	1.1	1.0	0.7	0.5	0.7	0.5	0.6	0.3
$\mathcal{FF}(\bar{B} \rightarrow D\ell\bar{\nu})$	0.3	0.3	0.3	0.3	0.4	0.4	0.3	0.2	0.2	0.2	0.2	0.2
$\mathcal{FF}(\bar{B} \rightarrow D^*\ell\bar{\nu})$	1.2	1.4	1.5	1.6	2.0	2.3	1.8	1.4	1.3	1.0	0.7	0.8
$\mathcal{FF}(\bar{B} \rightarrow D^{**}\ell\bar{\nu})$	1.3	1.6	1.3	1.1	0.9	0.6	0.7	1.2	1.9	1.9	1.1	0.2
Sec.Fakes. Composition	10.7	9.4	5.3	3.7	5.5	7.8	6.5	4.5	3.9	3.5	3.4	3.9
In-situ q^2 Calibration	4.2	4.6	4.5	5.0	6.5	7.6	5.9	4.1	3.8	2.9	2.0	0.6
ℓ ID Efficiency	0.5	0.4	0.4	0.3	0.3	0.2	0.1	0.1	0.2	0.2	0.2	0.1
ℓ ID Fake Rate	1.9	1.7	1.1	0.7	0.7	0.9	0.8	0.8	0.7	0.6	0.5	0.2
$K\pi$ ID Efficiency	3.4	2.8	2.2	2.2	2.8	3.1	2.6	1.9	2.0	1.8	1.6	0.6
$K\pi$ ID Fake Rate	3.1	2.2	1.9	2.1	2.4	2.8	2.1	1.5	1.7	1.4	1.4	0.5
K_S^0 Efficiency	0.7	0.6	0.5	0.5	0.7	0.8	0.6	0.5	0.5	0.4	0.4	0.1
π_{slow} Efficiency	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Tracking	0.2	0.2	0.2	0.1	0.1	< 0.1	< 0.1	< 0.1	0.1	0.1	0.1	0.1
Continuum Calibration	1.5	1.5	1.2	1.0	0.8	0.4	0.3	0.5	0.6	0.4	0.3	0.7
N_{BB}	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
$f_{+/0}$	0.1	0.1	0.1	0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Stat. Error (MC)	10.6	9.1	7.5	7.3	8.8	10.5	7.5	5.6	6.0	5.6	4.7	2.0
Unfolding Bias	0.9	2.7	1.8	1.0	0.5	0.4	0.3	0.1	0.2	0.3	< 0.1	0.4
Total Syst.	19.1	16.8	13.8	13.6	16.2	19.2	15.0	11.3	10.8	10.1	7.6	4.8

Table 10.1: Summary of the central values, statistical, and systematic uncertainties for the differential ratio of partial branching fractions as a function of $E_\ell^{B_{\text{sig}}}$ bin, defined by the boundaries [1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.7] GeV. The uncertainties are given as relative values on the central value in percent.

q^2 Bin	0	1	2	3	4	5
Central Value $\times 100$	1.75	1.56	1.36	1.35	1.32	23.47
Stat. Error (Data)	15.6	13.1	17.0	19.8	23.4	8.0
$\mathcal{B}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta'\ell\bar{\nu})$	0.2	0.2	0.2	0.2	0.3	0.2
$\mathcal{FF}(\bar{B} \rightarrow \pi/\eta/\rho/\omega/\eta'\ell\bar{\nu})$	0.2	0.3	0.4	0.4	0.5	0.3
$\mathcal{B}(\bar{B} \rightarrow x_u\ell\bar{\nu})$	1.1	0.9	0.8	0.8	0.8	0.6
Hybrid Model (BLNP)	8.4	7.9	3.8	0.4	2.8	5.5
DFN ($M_b^{\text{KN}}, a^{\text{KN}}$)	6.5	6.8	6.9	6.5	5.6	2.0
$N_{g \rightarrow s\bar{s}}$	0.4	0.1	0.5	0.7	1.0	0.6
$\mathcal{B}(\bar{B} \rightarrow D\ell\bar{\nu})$	0.8	0.5	0.2	0.8	1.2	0.3
$\mathcal{B}(\bar{B} \rightarrow D^*\ell\bar{\nu})$	0.2	0.1	0.8	1.6	2.2	1.3
$\mathcal{B}(\bar{B} \rightarrow D^{**}\ell\bar{\nu})$	0.7	0.4	0.5	0.7	0.9	0.3
$\mathcal{B}(\bar{B} \rightarrow D^{(*)}\eta\ell\bar{\nu})$	1.9	1.0	0.5	1.1	1.6	0.4
$\mathcal{B}(\bar{B} \rightarrow D^{(*)}\pi\pi\ell\bar{\nu})$	0.8	0.7	0.4	0.5	1.2	0.4
$\mathcal{FF}(\bar{B} \rightarrow D\ell\bar{\nu})$	0.1	0.2	0.4	0.3	0.2	0.1
$\mathcal{FF}(\bar{B} \rightarrow D^*\ell\bar{\nu})$	1.0	1.1	1.6	1.9	4.0	0.7
$\mathcal{FF}(\bar{B} \rightarrow D^{**}\ell\bar{\nu})$	1.9	1.2	1.0	1.3	1.8	0.5
Sec.Fakes. Composition	6.5	4.7	7.9	6.5	6.5	2.9
In-situ q^2 Calibration	3.6	3.9	< 0.1	6.4	16.0	4.6
ℓ ID Efficiency	0.3	0.1	0.2	0.2	0.2	0.2
ℓ ID Fake Rate	0.5	0.2	0.6	0.7	0.7	0.2
$K\pi$ ID Efficiency	1.9	1.7	1.6	2.4	2.9	0.9
$K\pi$ ID Fake Rate	1.5	1.1	1.4	1.6	1.7	0.5
K_S^0 Efficiency	0.6	0.5	0.4	0.7	0.9	0.2
π_{slow} Efficiency	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
Tracking	0.3	0.2	0.1	0.1	0.1	0.1
Continuum Calibration	2.4	0.9	0.7	0.9	0.8	0.3
N_{BB}	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
$f_{+/0}$	0.1	< 0.1	0.1	< 0.1	< 0.1	< 0.1
Stat. Error (MC)	5.3	4.2	5.5	5.8	6.8	2.3
Unfolding Bias	0.1	< 0.1	0.2	0.4	0.4	0.2
Total Syst.	14.8	13.2	12.9	13.5	20.7	8.6

Table 10.2: Summary of the central values, statistical, and systematic uncertainties for the differential ratio of partial branching fractions as a function of q^2 bin, defined by the boundaries $[0, 2, 4, 6, 8, 10, 26]$ GeV^2 . The uncertainties are given as relative values on the central value in percent.

10.3 Effect of lepton energy threshold

In the unfolding procedure, we have considered only events reconstructed with $E_\ell^{B_{\text{sig}}} \geq 1$ GeV. While the background subtracted spectrum includes all events reconstructed above the threshold, the migration matrix and efficiency correction consider only events that were both reconstructed and generated above the threshold. The unfolding procedure thus does not consider migration about the threshold, instead relying on the simulation to correctly model it. As can be seen from Fig. 9.2 and the migration matrices, the resolution of $E_\ell^{B_{\text{sig}}}$ is good, and events migrating about the threshold are expected to be found almost entirely within 100 MeV of the threshold and thus reconstructed in the adjacent bin. We therefore test the impact of the threshold on the final result by extending the unfolding procedure to consider an additional sacrificial bin below the threshold $0.9 \leq E_\ell^{B_{\text{sig}}} < 1.0$ GeV. The four components, continuum, secondaries and fakes, $\bar{B} \rightarrow X_c \ell \bar{\nu}$, and $\bar{B} \rightarrow X_u \ell \bar{\nu}$, are normalised to the nominal region, $E_\ell^{B_{\text{sig}}} \geq 1$ GeV, using the nominal procedure, and extrapolated to the sacrificial bin. For the unfolding, we set the regularisation parameter to $k = 10$. After unfolding, we discard the events in the sacrificial bin. Good agreement is observed between the two methods. This indicates that given the excellent resolution in $E_\ell^{B_{\text{sig}}}$, the choice of treatment of migration about the threshold represents a negligible impact on the analysis. Furthermore, the excellent agreement suggests that the unfolding procedure is robust against small changes to the regularisation.

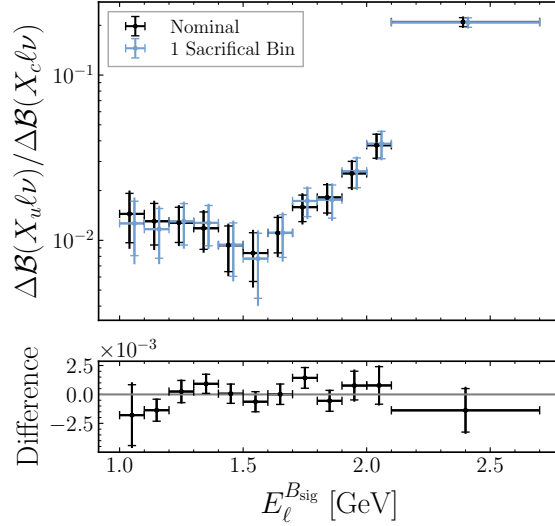
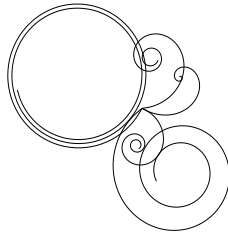


Figure 10.6: Comparison of the unfolded differential ratio of partial branching fractions in lepton energy for the nominal setup and including an additional sacrificial bin below the $E_\ell^{B_{\text{sig}}}$ threshold. The results are found to be consistent.



Part III-B

Search for Weak Annihilation at Belle II

Event reconstruction and selection

We describe the event reconstruction and event selection for the search for weak annihilation. Additionally, we discuss the calibration of continuum events.

As discussed in Ch. 2, the magnitude of the weak annihilation contribution to $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays is a source of a possible bias in the inclusive determination of $|V_{ub}|$. In order to constrain this component, we perform a search for a peaking structure at high q^2 within the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ spectrum, the expected signature of weak annihilation. The search is performed using data collected by the Belle II experiment. As we analyse $\bar{B} \rightarrow X \ell \bar{\nu}$ events, the reconstruction procedure of the analysis closely matches the of the previous analysis. There are, of course, some differences due to the hardware and software upgrades between Belle and Belle II, and improvements to the procedure; the reasoning driving the selections, however, remains the same. While the following description of selection criteria is complete, we thus approach it with some terseness, focusing on differences.

11.1 Samples

The analysis is developed and optimised assuming that it will be performed on the full data sample collected by Belle II between 2019 and 2022, before long shutdown one, corresponding to $364.09 \pm 0.02_{\text{stat}} \pm 2.33_{\text{syst}} \text{ fb}^{-1}$ collected on-resonance and a further $42.56 \pm 0.01_{\text{stat}} \pm 0.27_{\text{syst}} \text{ fb}^{-1}$ collected off-resonance.

We use 1 ab^{-1} of MC simulated generic $B\bar{B}$ and continuum events and 50 fb^{-1} of simulated off-resonance continuum events. The generic $B\bar{B}$ events, which contain a cocktail of thousands of decays are supplemented with resonant and nonresonant $\bar{B} \rightarrow X_u \ell \bar{\nu}$ events, $\bar{B} \rightarrow D_{\text{Gap}}^{**} \ell \bar{\nu}$ events and $\bar{B} \rightarrow \ell \bar{\nu}(\gamma)$ events, as described in Ch. 7. Between 5 and 25 million events are used per channel, which are weighted to their nominal branching fractions. For each of the 70 weak annihilation models we generate one million events, split equally between charged and neutral B decays and between the electron and muon channels. All samples are run-independent.

Additional MC samples are available for most channels, up to 2.8 ab^{-1} for generic $B\bar{B}$ events, however the utilised fraction provides a large enough sample to prepare the analysis. As the analysis must switch to run-dependent MC before publication¹, it was not considered necessary to process the full MC sample. Presented evaluations of the

¹Run-dependent samples were not yet available during preparation of this analysis.

expected systematic uncertainty due to the MC samples size are scaled to match the full run-independent sample.

11.2 Event reconstruction

We select events that pass the hadron trigger, which was discussed in Sec. 6.1.3. To minimise waste of computing resources, we also veto events in which it is highly unlikely that the FEI algorithm will be able to reconstruct a candidate B meson, requiring events to contain at least three clean tracks and clusters in the event, where clean refers to loose requirements to veto beam background, $p_t > 0.1$ GeV, $|z_0| < 2$ cm, and $|d_0| < 0.5$ cm for tracks and $E > 0.1$ GeV and $17 < \theta < 150^\circ$ for clusters, and a total visible energy of at least 4 GeV.

11.2.1 Tag

The tag candidate is reconstructed by FEI in one of 36 charged or 32 neutral hadronic B channels, listed in Table 4.1. As discussed in Sec. 8.4, the reconstruction efficiency of FEI is known to differ significantly between data and simulation. The Belle II performance group produces an official set of calibration factors which we adopt in this study (with some caveats that we will discuss in Sec. 12.1.2.6). In order for these calibrations to be valid, we match exactly the selections used in the calibration studies. We require the energy of the reconstructed B meson to be consistent with half the beam energy, $-0.15 < \Delta E < 0.10$ GeV, a beam constrained mass M_{bc} in excess of 5.24 GeV, and apply a loose selection on the response of the final FEI classifier, $\mathcal{P}_{\text{FEI}} > 10^{-3}$. This is one of two selections for which calibrations are available, and selects approximately 50% more events than the other available selection, which requires a tighter threshold of 10^{-2} . The additional events are more likely to be combinatorial background than those selected by the tighter threshold requirement; however, as we are searching for a rare process, the larger sample size is beneficial.

We additionally place a selection on the continuum suppression variable $\cos \theta_{\vec{T}_B \vec{T}_O}^{\text{cM}}$, the cosine of the angle between the thrust of the tag B meson and the remaining particles in the event, of $\cos \theta_{\vec{T}_B \vec{T}_O}^{\text{cM}} < 0.9$, to reduce the continuum contamination, which peaks at 1, at a cost of approximately 10% of the signal, which is nominally uniformly distributed. As discussed in Sec. 8.2.3 to calculate this variable we make some selections on the particles not used to reconstruct the tag B meson, in order to reject beam background. To select the particles that enter the calculation of $\cos \theta_{\vec{T}_B \vec{T}_O}^{\text{cM}}$, we adopt the same selections as used in studies of the FEI performance, such that it is valid to apply the calibrations derived in these studies. We consider only tracks that originate close to the interaction point with $dr < 2$ cm and $|dz| < 4$ cm, have a minimum transverse momentum of $p_t > 0.2$ GeV, and have momenta within the CDC acceptance. All tracks are considered to be pions. Clusters are required to have minimum energy of 80, 30, or 60 MeV respectively in the forward, barrel, and backwards regions of the ECL. Their reconstructed polar angle must

be between $17 < \theta_{\text{cl}} < 150^\circ$, vetoing the forward and backward edges of the ECL where performance is poor, and they must be detected within 200 ns of the event t_0 , consistent with photons created in the event.

After these selections, we perform a tag-side best candidate selection, selecting the reconstructed charged and neutral B mesons with the highest $\mathcal{P}_{\text{FEI}}$ respectively. We then tighten the M_{bc} requirement to 5.27 GeV. For off-resonance events we apply the selection to a scaled beam-constrained mass to account for the energy difference between on-resonance and off-resonance events,

$$M_{\text{bc}}^{\text{corr.}} = \sqrt{\frac{E_{\text{on}}^2}{4} - \left(\frac{E_{\text{on}}}{E_{\text{off}}}\right)^2 p_{B_{\text{tag}}}^{*2}}, \quad (11.1)$$

where E_{on} and E_{off} are the nominal beam energies for on and off-resonance data collection, 10.58 GeV and 10.52 GeV, respectively, and $p_{B_{\text{tag}}}^*$ is the centre-of-mass frame four momentum of the tag B candidate.

11.2.2 Lepton

We first select tracks that originate near the interaction point, requiring $|dr| < 1$ cm and $|dz| < 3$ cm, and that the tracks fall within the CDC acceptance in momentum polar angle and have at least one associated hit within the CDC. We identify electrons and muons with the selections $\mathcal{P}_e > 0.9$, and $P_\mu^{\text{SVD}} > 0.9$, respectively, where the choice of the BDT based electron identification, and likelihood ratio based muon identification follows the recommendation of the Belle II particle identification group. In addition for electron candidates we note that as the particle identification BDT was trained to only separate electrons, muons, pions and kaons, there is some contamination of protons into the electron sample. We therefore add a veto requirement that $P_p < 0.99$, which has near-unity efficiency for electrons while vetoing three quarters of the misidentified protons.

We additionally require that the leptons are found with momentum magnitudes and momentum polar angles in kinematic regions in which the performance of the particle identification has been studied and is well understood. For electrons, this corresponds to an additional requirement of $|\vec{p}| > 0.3$ GeV. For muons, this corresponds to requiring $|\vec{p}| > 0.3$ GeV, and tightening the polar angle requirements to $0.4 < \theta < 2.6$ rad.

We correct the four momentum of electron candidates for energy losses due to bremsstrahlung by searching for photons within a cone of 0.05 rad around the initial momentum direction of the electron track. The photons must have energies below 1 GeV, and otherwise meet the below defined requirements enforced on photon candidates within the hadronic system. The four momentum of any photons identified is summed to the four-momentum of the electron track. We allow for multiple radiative photons.

In addition to the semileptonic channels, true leptons may originate from the decay of the J/ψ , which decay to dileptons with a branching fraction of approximately 12%, or, in the case of electrons, from photon conversion in the detector materials. In both cases, the two-body nature of the decay allows for efficient vetoing of these contributions.

We combine the lepton candidate with all remaining oppositely charged tracks event, and select the pair with an invariant mass closest to the mass of the J/ψ or 0 GeV (for photon conversion) where the oppositely charged track is assigned the same mass hypothesis as the signal lepton. For the J/ψ veto we require the oppositely charged track to pass the selection imposed on tracks assigned to the hadronic system. For the photon conversion veto such a selection is not required as we expect a significant fraction of conversions to happen at the beam pipe, and applying strict dr and $|dz|$ requirements may miss such conversion events. Tracks used with an electron hypothesis undergo the same bremsstrahlung recovery procedure as the signal lepton. No particle identification selections are made on the second track to remain inclusive. We veto any candidates matching the following requirements.

- A dielectron mass, $M_{e^+e^-}$, less than 0.050 GeV, where the threshold has been selected to capture 95.45% of the true photon conversion tracks.
- A dielectron mass consistent with the J/ψ mass, $3.043 < M_{e^+e^-} < 3.129$ GeV, where the thresholds have been selected to capture 95.45% of true J/ψ candidates above the nominal J/ψ mass and 68.27% below the nominal J/ψ mass. The coverage in the lower mass direction has been reduced due to the long and broad tail caused by bremsstrahlung.
- A dimuon mass consistent with the J/ψ mass, $3.072 < M_{\mu^+\mu^-} < 3.122$ GeV, where the upper threshold has been selected to capture 95.45% of true J/ψ candidates above the nominal J/ψ mass. The mass peak has a small broad component towards the negative direction. To capture only the peak we symmetrise the upper threshold about the nominal J/ψ mass.

Selected lepton candidates are matched with the tag B meson candidates. We select candidates with $E_\ell^{B_{\text{sig}}} > 1$ GeV, where $E_\ell^{B_{\text{sig}}}$ is the energy of the lepton in the signal B rest frame which we infer as discussed in Sec. 8.2.4, with the notable differences that we include the rest mass of the lepton. If in the same event, lepton candidates can be successfully matched with both the selected charged and neutral tag candidates we select the tag with the highest $\mathcal{P}_{\text{FEI}}$. If multiple lepton candidates are matched with the selected tag, we select the lepton with the highest energy in the B meson frame.

Finally, we require the lepton to have a charge consistent with the decay of a signal B with a flavour inferred from the tag B , making no allowance for B mixing. As discussed for the previous analysis, this requirement is effective in removing events in which secondary leptons have been selected, but also vetoes $18.58 \pm 0.11\%$ [211] of neutral B signal events. As weak annihilation events are expected to have highly energetic leptons, and secondary leptons tend to be soft, secondary leptons can be vetoed by selections on the lepton energy, thus this requirement is expected to be needlessly strict. However, the FEI algorithm has only been calibrated for events in which the lepton charge is consistent with the tag B flavour. The charge matching requirement is therefore enforced for leptons matched to

charged and neutral B tag candidates, following the recommendations of the Belle II FEI performance group.

11.2.3 Hadronic system

The hadronic system is inclusively constructed from the remaining particles in the event by summing their four momenta. Such particles are required to pass basic selections in order to reduce contamination due to beam background, detector noise, and duplicated tracks.

For tracks, we require they originate close to the interaction point with $dr < 1$ cm and $|dz| < 3$ cm, must have recorded at least two hits in the CDC, and have a momentum polar angle between $0.31 < \theta < 2.61$ rad, slightly within the edges of the CDC, vetoing regions of poor agreement due to misalignment. Additionally, we apply the duplicate curl track rejection discussed in Sec. 8.2.3. The classifier used in this algorithm has been re-trained for Belle II. All tracks are assigned a pion mass hypothesis. As we later veto any events in which kaons or additional leptons are identified, assigning their mass hypothesis would be redundant. We do not attempt to identify protons, as we reconstruct baryonic modes with the FEI algorithm, reducing the cross-feed to the signal side.

For clusters, we require minimum energies of 70 MeV when recorded in the forward or backwards endcaps of the ECL, or 60 MeV in the barrel region, rejecting the majority of beam background photons, which tend to be soft. We further require $\mathcal{P}_{\text{Fake}\gamma} > 0.5$, where $\mathcal{P}_{\text{Fake}\gamma}$ is the response of a classifier trained to distinguish real photons from fake photons (hadronic split-off, noise, failures of the cluster-track matching algorithm) [267]. The classifier considers six variables: the cluster energy, timing, and polar angle, the distance between the cluster and nearest track extrapolated to the ECL surface, and the output of the Zernike and pulse shape discrimination classifiers², both trained to separate neutrons from photons as discussed in Sec. 5.4. Finally we require the photons to be recorded within 70 ms of t_0 , consistent with photons created in the event, and that they be recorded with polar angles between $0.29 < \theta_{\text{cl.}} < 2.65$ rad, removing small regions near the edge of the ECL which are poorly modelled.

11.2.4 Continuum calibration

We examine the modelling of the continuum background by comparing the off-resonance data with the simulated off-resonance sample. We observe large disagreements in normalisation and shape, in both the event-shape variables and the kinematic variables. The mismodelling is expected to be related to both the differences in FEI performance between data and MC, as we do not apply any calibration to continuum events, and from underly-

²As we noted in our previous discussion of pulse shape information (Sec. 5.10.1), the information is only partially available for the currently processed data set. Although this degrades the performance of the fake photon classifier in data, we observe an approximate match between data and MC distributions in sideband studies, and our selection is loose, reducing the impact of this mismodelling.

ing mismodelling of continuum events³. Therefore, we perform a data-driven calibration in two steps, correcting the normalisation and shape of the continuum sample, separately.

To improve the normalisation, we compare the number of off-resonance data events in the training sample (defined below) to the expected number of off-resonance events from the MC. This is done independently for the charged and neutral B_{tag} channels, since the rate at which false B mesons are reconstructed by the FEI algorithm may differ between both channels. Correction factors of 0.881 and 0.896 are found for the charged and neutral channels.

To improve the modelling of the continuum sample, we follow the approach proposed in Ref. [268], and used in recent searches for $B \rightarrow \mu\nu$ at Belle [79], and $B \rightarrow K\nu\nu$ at Belle II [269]. Our primary goal is to improve the modelling of the event shape variables, as these are subsequently used to train a continuum suppression classifier, and the kinematic variables, as we will fit the q^2 distribution and place further selections on M_X and $E_\ell^{B_{\text{sig}}}$.

We train a BDT on the event shape and kinematic variables to distinguish off-resonance data from off-resonance MC. For definitions of the event-shape variables see Sec. 8.3.1 noting that they are calculated with the track and cluster selections discussed in Sec. 11.2.3. This classifier is then applied to the on-resonance continuum MC, assigning a weight of

$$w_{\text{CW}} = \frac{\mathcal{P}_{\text{CW}}}{1 - \mathcal{P}_{\text{CW}}}, \quad (11.2)$$

where $\mathcal{P}_{\text{CW}}$ is the classifier response, which is larger as the event is more data-like. The ratio $\mathcal{P}_{\text{CW}}/(1 - \mathcal{P}_{\text{CW}})$ can be interpreted as an estimate of the likelihood ratio $\mathcal{L}^{\text{Data}}/\mathcal{L}^{\text{MC}}$, where $\mathcal{L}^{\text{Data}}, \mathcal{L}^{\text{MC}}$ are the likelihoods of the continuum event to be from data and simulation, respectively.

We reconstruct 49176 data and 71543 MC events in the off-resonance sample. These are split into three categories:

- training (50%), used directly to train the sample and derive the normalisation correction;
- validation (20%), used to prevent the growth of unnecessarily complex models by terminating training when performance on this sample, independent of the training sample, plateaus for 20 training iterations; and
- testing (30%), used to check classifier performance.

As the testing sample is not used in training the classifier, it remains statistically independent and can be used later to further constrain the continuum. As this is a simple classification task, we use a boosted decision tree as the classifier, via the `xgboost` package. We use the same hyperparameter optimisation procedure as for the ECL particle identification BDTs described in Sec. 5.6, running for 500 trials, optimising for the best cross-entropy loss on the validation sample.

³There are ongoing efforts at Belle II to tune the `PYTHIA8` parameters which enter into this modelling.

For the reweighting procedure to be valid, more specifically for the interpretation of the classifier response as a likelihood ratio to be valid, the classifier must be well calibrated. It's score should give an accurate estimation of the purity in a small window about the score [201] (recall the discussion in Sec. 5.7). Fig. 11.1 shows the purity in bins of classifier score for the training, validation and testing samples. The validation and testing samples are found to be well calibrated, indicating the reweighting procedure is valid. In the training sample we observe poor agreement, as the BDT has slightly over-trained on this sample in order to improve it's performance on the validation sample.

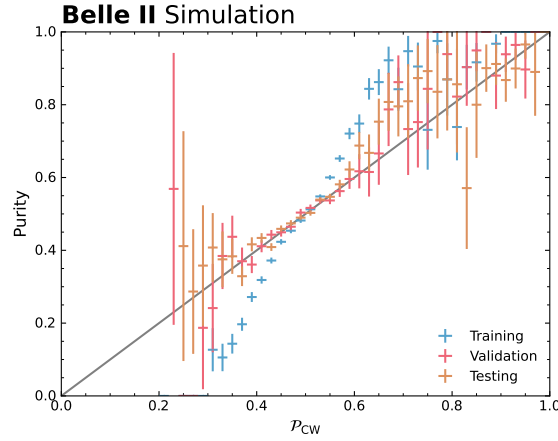


Figure 11.1: Validation of the calibration of the BDT trained for continuum modelling.

We estimate the importance of each variable to the classifier via their Shapley values (see Sec. 4.1.4). In this context, a higher importance indicates a greater sensitivity to mismodelling. The importances, as measured on the validation sample, are presented in Fig. 11.2.

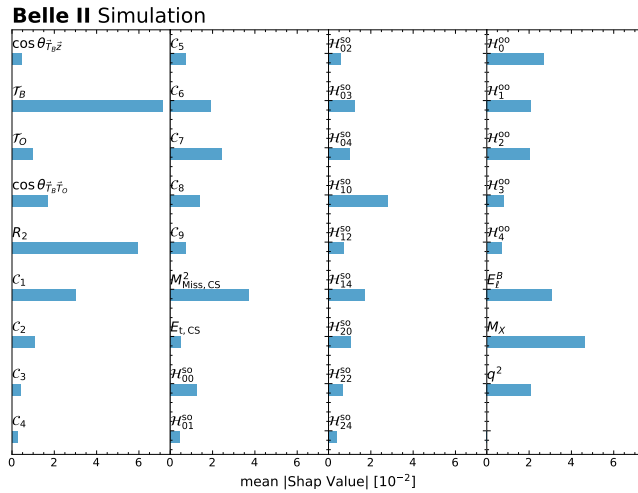


Figure 11.2: Variable importance for the data-driven continuum reweighting.

The response of the classifier to the testing sample is given in Fig. 11.3. The figure additionally gives an example of the pre-reweighting and post-reweighting distribution in $\mathcal{H}_{20}^{so}$, the most important variable for the continuum suppression classifier applied in the analysis, and R_2 one of the most mismodelled variables. This continuum suppression classifier is described in Sec. 11.3.4.

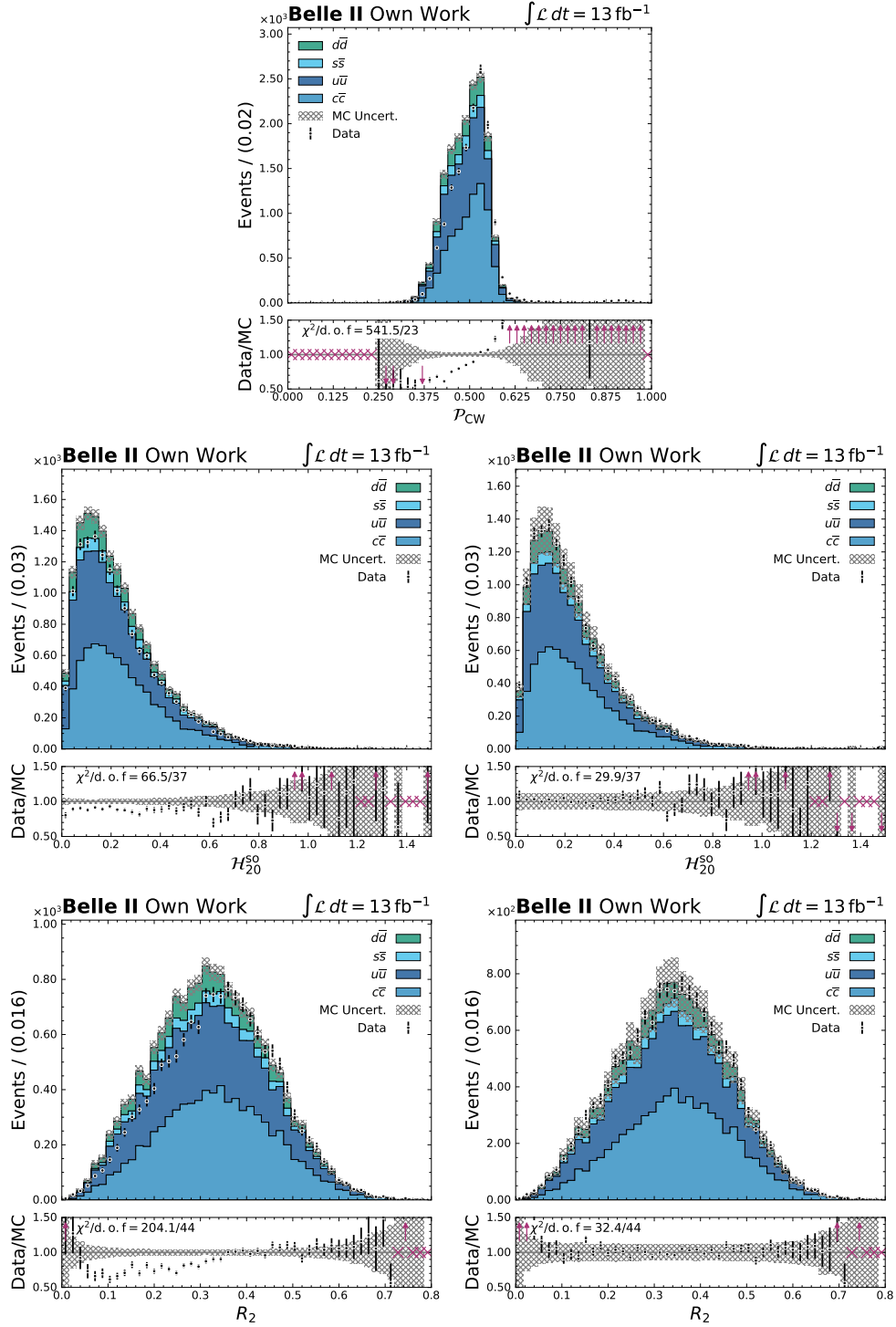


Figure 11.3: Distribution of the nominal classifier response for the classifier trained to distinguish data from Monte Carlo (*top*) and distributions of $\mathcal{H}_{20}^{so}$ and R_2 , before (*left*) and after (*right*) the data-driven reweighting procedure in the off-resonance test sample. For the calculation of the χ^2 value, bins with data yields or expected yields less than 5 events are merged into adjacent bins algorithmically. The χ^2 value is calculated using the statistical and systematic uncertainty taking into account correlations.

11.3 Event selection

Having reconstructed $\bar{B} \rightarrow X\ell\bar{\nu}$ events that contain highly energetic leptons, we now add further selections to veto background while maintaining a high weak annihilation reconstruction efficiency. Particular care is taken to reduce potential bias due to the modelling of the hadronic system in the weak annihilation contribution.

Unless otherwise stated the selections on all considered variables are optimised simultaneously using a multivariate optimisation procedure which we discuss in Sec. 11.3.5 after introducing all considered quantities.

11.3.1 Kinematic variables

We restrict the phase space to near the kinematic end point where the weak annihilation contribution is expected to be concentrated, requiring a low hadronic mass, $M_X < 2 \text{ GeV}$, and high lepton energy $E_\ell^{B_{\text{sig}}} > 1.5 \text{ GeV}$. These selections are kept loose to reduce sensitivity to the modelling of the hadronic system and lepton energy spectrum. As discussed in Sec. 7.3 the modelling of weak annihilation decays in both these variables is poorly understood. A tight selection may therefore veto weak annihilation events in data if their distribution differs significantly from the simulated samples. The expected distributions of these two variables are shown in Fig. 11.4. Note that for these figures, and all subsequent figures within this section, all selections discussed except for the selection applied to the variable under discussion are applied. The sample of events with leptons that have energies in the signal B frame between 1 and 1.5 GeV is used to check the modelling of the selection variables in Sec. 11.4.

To demonstrate the range of possible weak annihilation shapes, for each figure we overlay three weak annihilation models, corresponding to samples produced with:

$$\begin{aligned} (M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (19.2 \text{ GeV}^2, 0.521 \text{ GeV}), \\ (M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (22.1 \text{ GeV}^2, 0.141 \text{ GeV}), \\ (M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (24.3 \text{ GeV}^2, 0.020 \text{ GeV}), \end{aligned}$$

given by orange and pink dotted, solid and dashed lines, respectively. The normalisation of these samples is arbitrary.

11.3.2 Particle content vetoes

As discussed for the Belle analysis, the particle content of the hadronic system can identify charmed B decays, without needing to reconstruct specific decay chains. We search the hadronic system for kaons, indicative of $b \rightarrow c \rightarrow s$ decays, charged leptons, as they are likely to originate from semileptonic D meson decays, which make up 13% of D^0 decays, and 34% of D^+ decays, and for slow pions, originating from the decay of D^* mesons.

We identify charged kaons by the likelihood ratio requirement $P_K > 0.6$ and that they must have recorded at least 20 hits in the CDC. Candidate K_S^0 particles are reconstructed

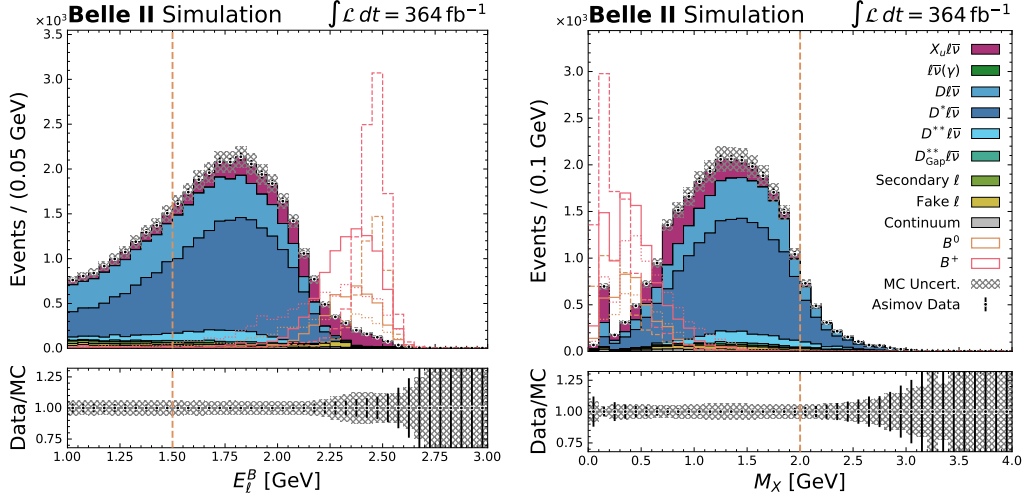


Figure 11.4: Expected distribution of the selected sample in lepton energy and hadronic mass. The selected thresholds are shown as vertical orange line. The normalisation of the overlaid weak annihilation samples (in orange and pink) is arbitrary.

from pairs of oppositely charged tracks under pion mass hypotheses. As K_S^0 particles are long-lived, their decay products do not originate strictly near the interaction point, we therefore do not enforce the requirements made on rest-of-event tracks above. We perform a vertex fit to find the decay vertex and improve the mass resolution. After the fit the K_S^0 candidates are required to have an invariant mass between 470 MeV and 530 MeV, consistent with the nominal K_S^0 mass (497.6 MeV), and the position of the vertex must be consistent with momentum of the K_S^0 candidate, requiring a cosine of the angle between the momentum vector and the vector connecting the interaction point with the decay vertex to be greater than 0.998. We make no attempt to reconstruct K_L^0 as the performance of the K_L^0 identification remains poorly understood⁴. Leptons are identified by the same particle identification requirements applied to the signal lepton, $\mathcal{P}_e > 0.9$ or $P_\mu^{\text{SVD}} > 0.9$. For charged kaons and leptons we additionally require that particles have momenta within phase space regions in which the identification efficiencies of the particle species, and pion-to-kaon or pion-to-lepton misidentification rates have been studied by the Belle II particle identification performance group.

Finally, we repeat the partial D^* reconstruction as outlined in Sec. 8.5.1. We consider both charged and neutral slow pions. For charged slow pions, we require momenta between 50 and 200 MeV in the lab frame, and apply the curl track rejection. We do not apply any impact parameter selections as the resolution of impact parameters is expected to be poor for slow pions, which are expected to record only a few hits in the vertex detectors

⁴Recently several improvements have been made to the K_L^0 reconstruction, centred about improving the matching of K_L^0 candidates between ECL and KLM only clusters. Future iterations of this study and inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ studies at Belle II are expected to see significant benefits from including a K_L^0 veto which will remove appreciable fractions of the remaining $\bar{B} \rightarrow X_c \ell \bar{\nu}$ background.

and inner layers of the CDC. To reconstruct neutral slow pions we select photons with minimum energies of 80, 30 and 60 MeV in the forward, barrel, and backward region of the calorimeter, respectively. They must also have been observed in at least two crystals of the calorimeter. The cluster must be located between $17 < \theta_{\text{cl}} < 150^\circ$. Pairs of photons are combined into a π^0 candidate, where we select candidates with a diphoton mass between $120 \text{ MeV} < m_{\gamma\gamma} < 145 \text{ MeV}$ and a momentum between 50 and 200 MeV in the laboratory frame. These selections match a set of selections used in official Belle II neutral slow pion performance studies, tuned to have an approximate 40% reconstruction efficiency.

From the slow pion candidates, we infer the D^* energy as per Eq. 8.10, again assume it to be co-linear to the pion in the lab frame and estimate a missing mass for the event as per Eq. 8.11. If the D^* was reconstructed from a charged slow pion, we require the charge of the slow pion to be opposite to that of the signal lepton. We perform separate best candidate selections for D^* candidates built from charged and neutral slow pions, selecting the respective candidate with the lowest $M_{\text{Miss}}^2(D^* \rightarrow D\pi^+)$ or $M_{\text{Miss}}^2(D^* \rightarrow D\pi^0)$, respectively. The expected distributions of these two variables for selected events in which slow pions are found, are shown in Fig. 11.5.

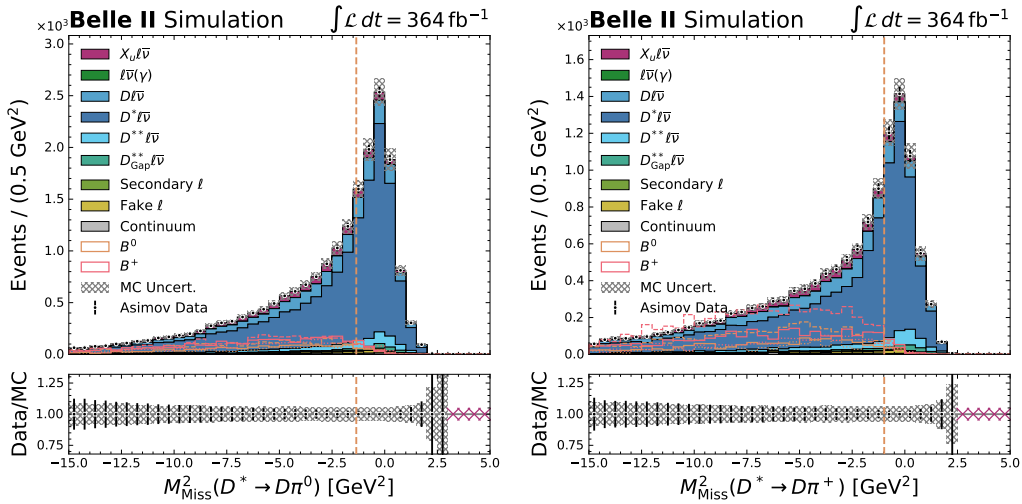


Figure 11.5: Expected distribution of the selected sample in $M_{\text{Miss}}^2(D^* \rightarrow D\pi^+)$ and $M_{\text{Miss}}^2(D^* \rightarrow D\pi^0)$. Events above the selected thresholds, shown as vertical orange lines, are removed. The normalisation of the overlaid weak annihilation samples (in orange and pink) is arbitrary. All other selections described in Sec. 11.3 are applied.

11.3.3 Missing mass

We further apply a selection on the missing mass squared of the event, M_{miss}^2 , which is expected to be insensitive to the precise modelling of the weak annihilation decay. As discussed in Sec. 8.2.4 for a well-reconstructed semileptonic event with only a single missing particle in the event, that being the neutrino, this corresponds to the neutrino

mass, peaking at 0 GeV^2 .

The selection serves mainly to veto $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and secondary and fake lepton events which have a tail toward larger values due to additional missing particles (K_L^0 , additional neutrinos, or simply a higher probability to miss a particle in an event due to the higher particle multiplicity). Fig. 11.6 gives the expected M_{Miss}^2 distribution.

As the dominant background to the weak annihilation signal in the endpoint region, $\bar{B} \rightarrow X_u \ell \bar{\nu}$ is also expected to have no additional missing particles, the distributions of this component and the weak annihilation signal are almost identical. The optimised selection is therefore relatively loose, maintaining high weak annihilation and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ selection efficiencies.

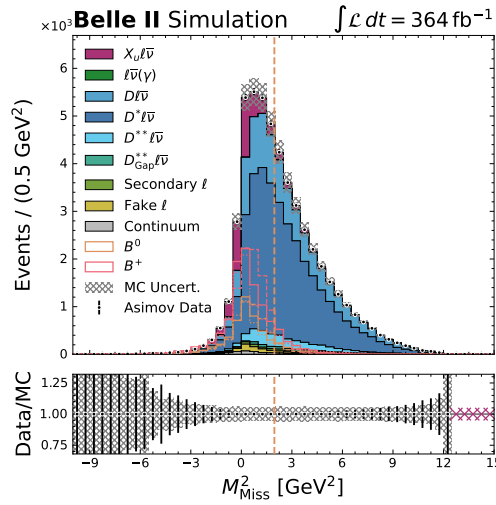


Figure 11.6: Expected distribution of the missing mass squared. All other selections described in Sec. 11.3 are applied.

11.3.4 Continuum suppression

Our selections so far, especially the requirements for low hadronic mass, high lepton energy, and that the FEI algorithm must be able to reconstruct a tag B meson, have already rejected a large fraction of continuum. The remaining continuum events peak at low q^2 , away from the signal; however, a long tail persists towards high q^2 . This tail leads to a large continuum contribution to the region of interest beyond the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ endpoint. To reduce this contamination, we add an additional continuum suppression classifier.

We split off 10% of the simulated WA samples, and 8% of the simulated continuum samples to use in training and then discard, corresponding to 72505 and 120821 events, respectively. These samples are further subdivided, with 20% used for validation and the remaining 80% used to train the model. For the signal sample we use an admixture of all WA models. The training sample is selected after the basic event reconstruction, without the additional selections described within this selection. This ensures a large training

sample and allows for the selections on the classifier response and the other selection variables to be optimised simultaneously.

As input variables we consider the event shape variables that we corrected in Sec. 11.2.4. The missing mass squared is excluded, as it is included as a separate selection. For the model, we use a simple feed forward neural network, implemented using `tensorflow` [196]. The model parameters are optimised via a Bayesian hyperparameter optimisation process [185]. Details of the floated parameters, their optimised values and allowed ranges are given in Table 11.1. The training of each model can take several minutes, even on gpu resources. Therefore, we run the hyperparameter optimisation procedure for only 200 iterations.

As we study a simplified weak annihilation model, any correlation of the continuum suppression to the kinematic variables, $M_X, E_\ell^{B_{\text{sig}}}$ and q^2 , or the number of tracks or photons produced in the hadronic system is highly undesirable, as this introduces additional model dependence. We therefore add an additional loss term to our model, penalising configurations whose output has a high distance correlation [272] to specific variables, which we refer to as *flatness* features. Distance correlation can measure non-linear correlations, and networks utilising it for decorrelation have been shown to equal or outperform more complicated adversarial networks [272].

The total loss (L) is given by:

$$L = \text{BCE}(\vec{y}, f(\vec{x})) + g(J) \frac{\lambda}{n} \sum_i \text{dCorr}_{y=1}^2(\vec{q}_i, \vec{y}) + T(\vec{y}, f(\vec{x})) \quad (11.3)$$

where BCE indicates the standard binary cross entropy loss function, dCorr^2 the distance correlation function, $f(\vec{x})$ is the prediction of the model and $\vec{y}$ are the true labels of the training data, zero or one for continuum and WA events, respectively. The index in the sum, i runs over the set of flatness features with the vector $\vec{q}_i$ giving the value of the variable for each event. The subscript $y = 1$ indicates that we only calculate the distance correlation for the signal class. The strength of the distance correlation loss term is governed by a single hyperparameter λ , which is normalised by the number of flatness features (n). This hyperparameter must be tuned separately to the hyperparameter optimisation procedure.

The term $T(\vec{y}, f(\vec{x}))$ is added for technical reasons. During development, it was found that a fraction of models tended to become stuck in a local minimum, where the model focused entirely on the decorrelation term, predicting essentially the same value for all events to ensure there was no correlation. We therefore add a term which is designed to introduce a large loss if the mean prediction of the signal and background class are almost identical but vanish if there is any separation between them. The term takes the form:

$$T(\vec{y}, f(\vec{x})) = (1 - |\text{mean}_{y=1}(f(\vec{x})) - \text{mean}_{y=0}(f(\vec{x}))|)^{100}, \quad (11.4)$$

To further aid the model in finding a minimum that balances the cross entropy and decorrelation loss, we gradually introduce the decorrelation loss term over the first 30

Parameter	Value	Range	Description
N Layers	5	[1, 5]	Number of layers in the feed forward network.
Layer Size	[32, 256, 128, 128, 256]	{8, 16, 32, 64, 128, 256}	Number of neurons in each layer of the network. The number is independently selected for each layer and is restricted to a power of two.
Activation Function	Swish	-	We use the swish [270] activation function of the form $f(x) = \frac{x}{1+\exp(-x)}$. Tests with ReLU showed similar performance, indicating the network is largely agnostic to the specific activation function.
Optimiser	AdamW	-	We adopt the AdamW [271] optimiser which computes adaptive learning rates for each parameter and has demonstrated rapid convergence in few epochs. Tests using the SGD optimiser converged on similar models.
Learning Rate	0.0059	$[10^{-5}, 10^{-2}]$	Learning rate of the model. Governs the rate at of gradient descent.
Weight Decay	7.6×10^{-7}	$[10^{-8}, 10^{-1}]$	Strength of the weight decay regularisation. Helps to avoid over-fitting by penalising overly complex models.
Batch Normalisation	<i>true</i>	<i>[true, false]</i>	If true, a batch normalisation layer is added after every dense layer.
Dropout Layers	<i>true</i>	<i>[true, false]</i>	If true, we add a dropout layer after each dense layer. If the batch normalisation is also used the dropout is added after the normalisation.
Dropout Rate	0.22	[0, 0.5]	Governs the rate at which the dropout layers randomly drop a fraction of input connections. The remaining connections are re-scaled to compensate.
Label Smoothing (<i>s</i>)	1.7×10^{-5}	$[10^{-5}, 0.05]$	Slightly smooths the labels by changing them from binary, 1 or 0 for signal and background, to $1 - s$ and s . This helps to prevent over-fitting, and prevents issues with many events being assigned to exactly 1 or 0 due to floating point precision.
Batch Size	8192	-	Size of a minibatch for each epoch of training. The batches are large to allow for accurate calculation of the DisCo loss term for each batch.

Table 11.1: Hyperparameters of the continuum suppression neural network. All hyperparameters not mentioned are left at their default values as of **tensorflow v2.10.0**.

epochs of training, via the coefficient $g(J)$, where J is the current epoch of training, of the form

$$g(J) = \min(1, J/30). \quad (11.5)$$

We again add an early stopping condition, terminating training if the performance of a model on the validation sample does not improve after 50 further epochs, returning the model as it was at its best epoch.

The two-dimensional distributions of the log-transformed classifier response against the kinematic variables are shown in Fig. 11.7 for models trained with $\lambda = 0, 20$, and 50. The Pearson correlation coefficient of the flatness features to the log-transformed classifier responses is shown in Fig. 11.9. Optimising λ is challenging, it requires a robust estimate of the range of possible weak annihilation distributions within the flatness variables, such that an appropriate balance in separation power between continuum backgrounds and weak annihilation samples and the reduction in model dependence can be struck. Given the simplified weak annihilation model, such robust estimates are missing. However, we observe the correlation of the classifier response to the flatness variables to reduce rapidly as the distance correlation term is introduced, with further reductions in correlation appearing asymptotic. Therefore, we select $\lambda = 20$, since models trained with this strength of distance correlation loss exhibit low correlations to the flatness variables, while maintaining excellent continuum rejection.

In Fig. 11.7 we observe a peak of events at zero in the log-transformed classifier responses. The origin of this peak is not understood and remains under investigation. It is not related to specific tag modes, the kinematics of the tag or lepton, or the multiplicities of the hadronic system.

The variable importances, as estimated using SHAP values, are shown in Fig. 11.8. The log-transformed expected distribution of the continuum suppression variable, $\mathcal{P}_{\text{CS}}^{\lambda=20}$, is shown in Fig. 11.10. The three overlaid WA distributions, representing broad and narrow models at different $M_{\ell\bar{\nu}}$, have similar shapes, further indicating that the decorrelation procedure is successful.

11.3.5 Optimisation procedure

We optimise all selections, except for the requirements on lepton energy and hadronic mass, which, as discussed, are kept deliberately loose, through a simultaneous optimisation procedure. We select nine representative weak annihilation models, corresponding to pairwise combinations of $M_{\ell\bar{\nu}}^2 = \{19.2, 22.1, 24.3\} \text{ GeV}^2$ and $\Gamma_{\ell\bar{\nu}} = \{0.02, 0.14, 0.52\} \text{ GeV}$.

We adopt the figure of merit proposed in Ref. [273], given by:

$$f_{\text{Punzi}}^{\text{WA}} = \frac{\epsilon_{\text{WA}}}{a/2 + \sqrt{B}}, \quad (11.6)$$

where ϵ_{WA} is the efficiency of reconstructing and selecting weak annihilation events, B is the yield of background events, and a corresponds to the desired number of standard

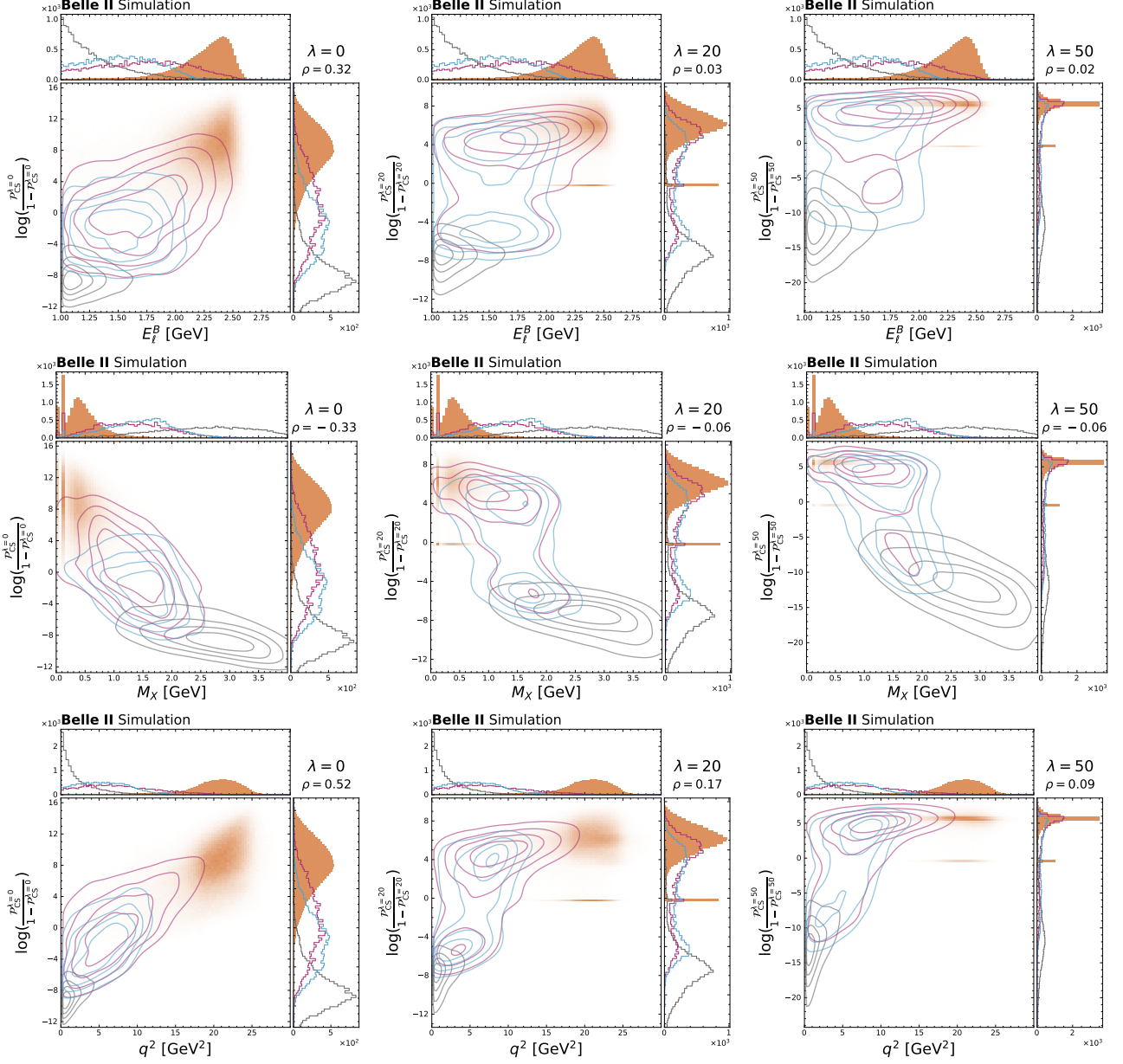


Figure 11.7: Comparison of the correlation of the trained continuum suppression networks to $E_\ell^{B_{\text{sig}}}$ (top), M_X (center), and q^2 (bottom) with different strengths of the distance correlation loss term, λ . The distribution of the admixture of WA models is shown in orange. For comparison contours of the distributions of the continuum (grey), $\bar{B} \rightarrow X_c \ell \bar{\nu}$ (blue) and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ (purple) samples at 20, 40, 60 and 80% of the respective maximum densities are overlaid.

Belle II Simulation

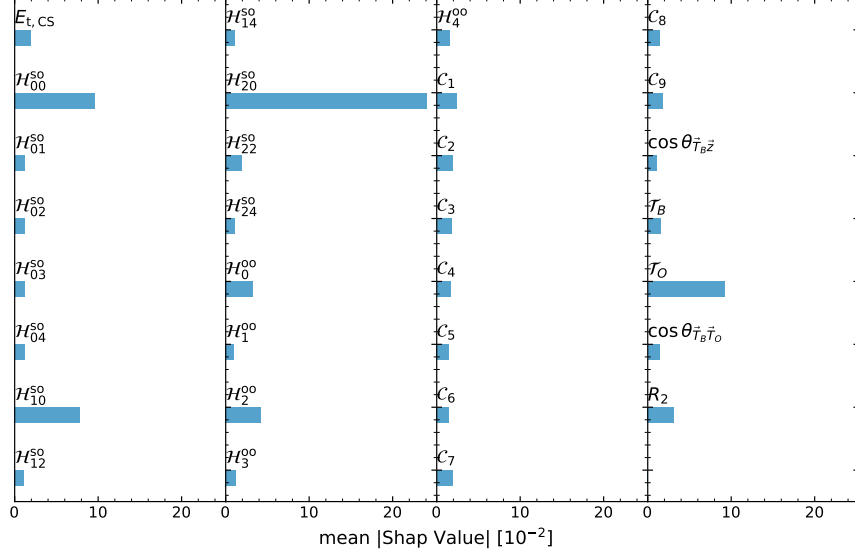


Figure 11.8: Variable importance for the continuum suppression network.

Belle II Simulation

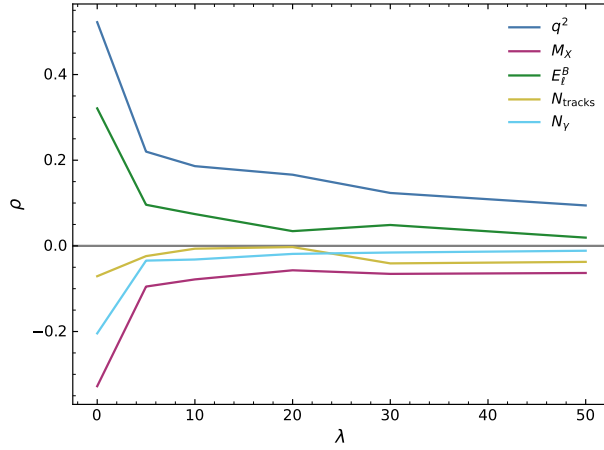


Figure 11.9: Correlation of the log transformed continuum suppression classifier responses against the flatness variables as a function of the strength of the distance correlation loss term, λ .

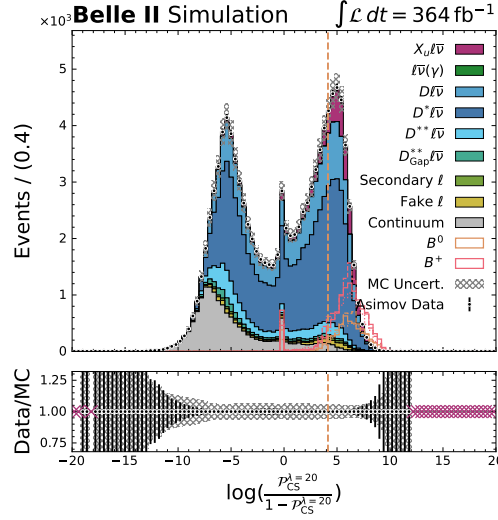


Figure 11.10: Log transformed distribution of the continuum suppression network output. We reject events below vertical dashed line.

deviations for a one sided Gaussian hypothesis test. Maximising this figure-of-merit maximises the sensitivity of counting experiments, considering only statistical uncertainties. As we expect the search for weak annihilation to be statistically limited, maximising this figure of merit thus maximises the sensitivity of our search. We note that as we expect B to be large, the specific choice of a is inconsequential⁵. We therefore set $a = 5$.

For each of the nine models, we evaluate $f_{\text{Punzi}}^{\text{WA}}$ separately for the charged and neutral WA contribution, considering background events reconstructed with charged or neutral B tags, respectively. We neglect FEI cross-feed contributions of the WA samples. As WA is concentrated at high q^2 , and there are large background contributions at low q^2 we evaluate the signal efficiencies and background yields in a restricted region of q^2 , chosen to contain 90% of the WA yield after the tested selections are applied.

We optimise the selections by Bayesian optimisation, via the implementation of Ref. [185]. We simultaneously optimise for all weak annihilation models, minimising [274]:

$$F_{\text{Punzi}} = \sum_{\{B^+, B^0\}} \sum_i 1/f_{\text{Punzi}}^i, \quad (11.7)$$

where i runs over the nine representative weak annihilation models. The optimisation sets the imposed thresholds on $\mathcal{P}_{\text{CS}}^{\lambda=20}$, M_{Miss}^2 , $M_{\text{Miss}}^2(D^* \rightarrow D\pi^+)$ and $M_{\text{Miss}}^2(D^* \rightarrow D\pi^0)$, and whether to veto events in which kaons or additional leptons are identified in the hadronic system. The optimisation procedure is run for 500 trials and converges to near-optimal

⁵As we expect the background yields to be much larger than the signal yields, our choice of figure-of-merit to maximise is functionally equivalent to the other common choice, the statistical significance, $S/\sqrt{S+B}$ where S is the signal yield. If $B \gg S$, this reduces to $S/\sqrt{B}$, and thus differs only by the constant scaling factor $1/S_0$, the generated number of signal events.

selections within approximately 100 trials, as shown in Fig. 11.11. The selected thresholds are:

$$\begin{aligned} \log\left(\frac{\mathcal{P}_{\text{CS}}^{\lambda=20}}{1 - \mathcal{P}_{\text{CS}}^{\lambda=20}}\right) &> 4.15, \\ M_{\text{Miss}}^2 &< 1.96 \text{ GeV}^2, \\ M_{\text{Miss}}^2(D^* \rightarrow D\pi^+) &< -0.98 \text{ GeV}^2, \\ M_{\text{Miss}}^2(D^* \rightarrow D\pi^0) &< -1.35 \text{ GeV}^2, \end{aligned}$$

and events with kaon or additional lepton candidates are vetoed. After all selections, the reconstruction efficiencies for weak annihilation events range between 0.43–0.62% and 0.26–0.35% for the charged and neutral B -meson decay channels, respectively.

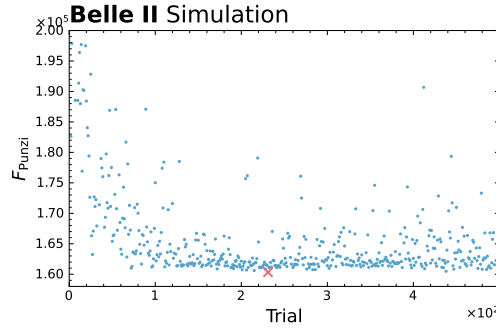


Figure 11.11: Optimisation of the Punzi-loss. The optimiser rapidly converges on near optimal parameters. The best trial is shown as a pink cross.

In this procedure, we have simultaneously optimised for all weak annihilation models, assuming that there are no significant differences in optimal thresholds between the models. We test the validity of this assumption by performing one-dimensional scans about the selected thresholds in $\log(\frac{\mathcal{P}_{\text{CS}}^{\lambda=20}}{1 - \mathcal{P}_{\text{CS}}^{\lambda=20}})$ and M_{Miss}^2 . The values of f_{Punzi} for each of the representative models are presented in Fig. 11.12. The selected thresholds are found to be near optimal for each model. The single optimised set of selections is thus valid for all models individually.

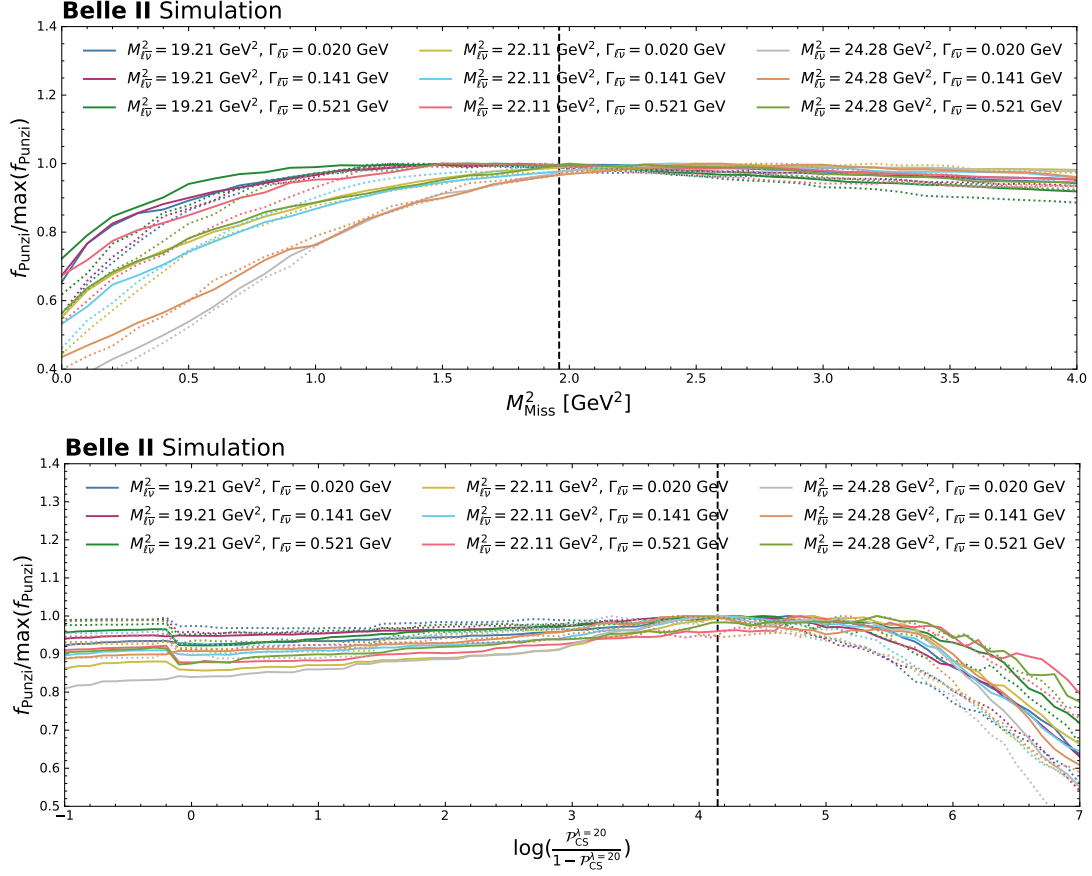


Figure 11.12: One dimensional scans of the Punzi figure of merit about the selected thresholds in missing mass squared (*top*) and the continuum suppression variable (*bottom*) for the nine example weak annihilation models. For each model the figure of merit for the charged B component is given as a solid line, for the neutral B component as a dotted line. The dashed vertical lines give the selected thresholds. The optimised thresholds, which are optimised for all models simultaneously, are found to be near optimal for each individual model.

11.4 Control region study

To validate the modelling of the dominant background components, we consider a control region expected to be free of weak annihilation events, given by the requirement $1.0 < E_\ell^{B_{\text{sig}}} < 1.5 \text{ GeV}$, with all other selections matching the signal region. Fig. 11.13 presents comparisons of data and simulation in $E_\ell^{B_{\text{sig}}}$, M_X , $\log(\frac{\mathcal{P}_{\text{CS}}^{\lambda=20}}{1-\mathcal{P}_{\text{CS}}^{\lambda=20}})$, and M_{Miss}^2 . For each plot, the selections on all other variables are applied. The modelling of $E_\ell^{B_{\text{sig}}}$ is found to be excellent. For M_X , a slight shift between the data and simulated spectra is observed; however, the agreement is within uncertainty. For $\log(\frac{\mathcal{P}_{\text{CS}}^{\lambda=20}}{1-\mathcal{P}_{\text{CS}}^{\lambda=20}})$, some structure is observed in the right shoulder. We also observe some disagreement in the modelling of M_{Miss}^2 , expected to be caused by similar factors as discussed in Sec. 8.5.2. Refs. [213, 259], studied the exclusive decays $\bar{B} \rightarrow D^{**} \ell \bar{\nu}$ and $\bar{B} \rightarrow D^{(*)} \ell \bar{\nu}$, and finding similar disagreement, smear the missing mass squared distribution using normal distributions or combinations of normal, Laplace and Cruijff distributions. Preliminary studies to smear the missing mass squared distribution observed within this study have not identified suitable smearing procedures; however, more complex procedures which also consider the particle multiplicities have not yet been investigated.

A comparison of the $M_{\text{Miss}}^2(D^* \rightarrow D\pi^+)$, and $M_{\text{Miss}}^2(D^* \rightarrow D\pi^0)$ distributions, and the observed number of charged kaons, K_S^0 , and leptons in the hadronic system for events vetoed by requirements on these quantities is shown in Fig. 11.14. The number of additional leptons is found to be in excellent agreement between data and simulation. The number of events in which two charged kaons are found is underestimated in simulation. However, this is found to be compensated for by a slight overestimation of the number of events with a single charged kaon. The number of K_S^0 candidates is underestimated in simulation, expected to be caused by the low $D^0 \rightarrow K^0 X$ rate in simulation, but agrees within 1.7σ .

Finally, we examine the distribution of data and MC in q^2 , from which we will extract the weak annihilation signal strengths. The details of this extraction are discussed in the following chapter. The comparison is shown in Fig. 11.15. Excellent agreement is observed between data and simulation, indicating that the dominant backgrounds to the analysis, $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$, are well modelled within this region of phase space.

In general, the agreement between data and simulation in the medium lepton momentum region is good. We therefore consider the expected fit results discussed in the following chapter, and especially the expected limits discussed in Sec. 12.4 to be an accurate estimate of the sensitivity of the measurement.

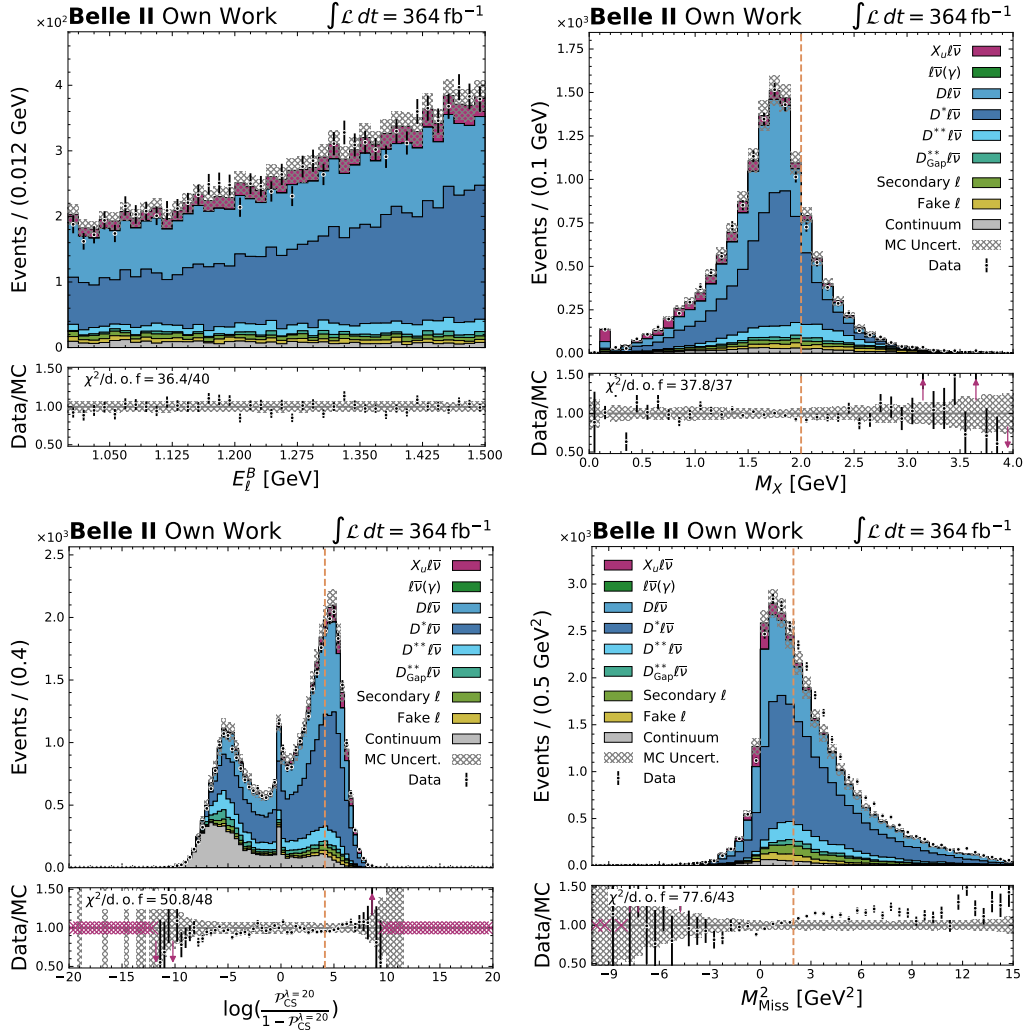


Figure 11.13: Data-MC agreement in the continuous event selection variables. For the calculation of the χ^2 value, bins with data yields or expected yields less than 5 events are merged into adjacent bins algorithmically. The χ^2 value is calculated using the statistical and systematic uncertainty taking into account correlations.

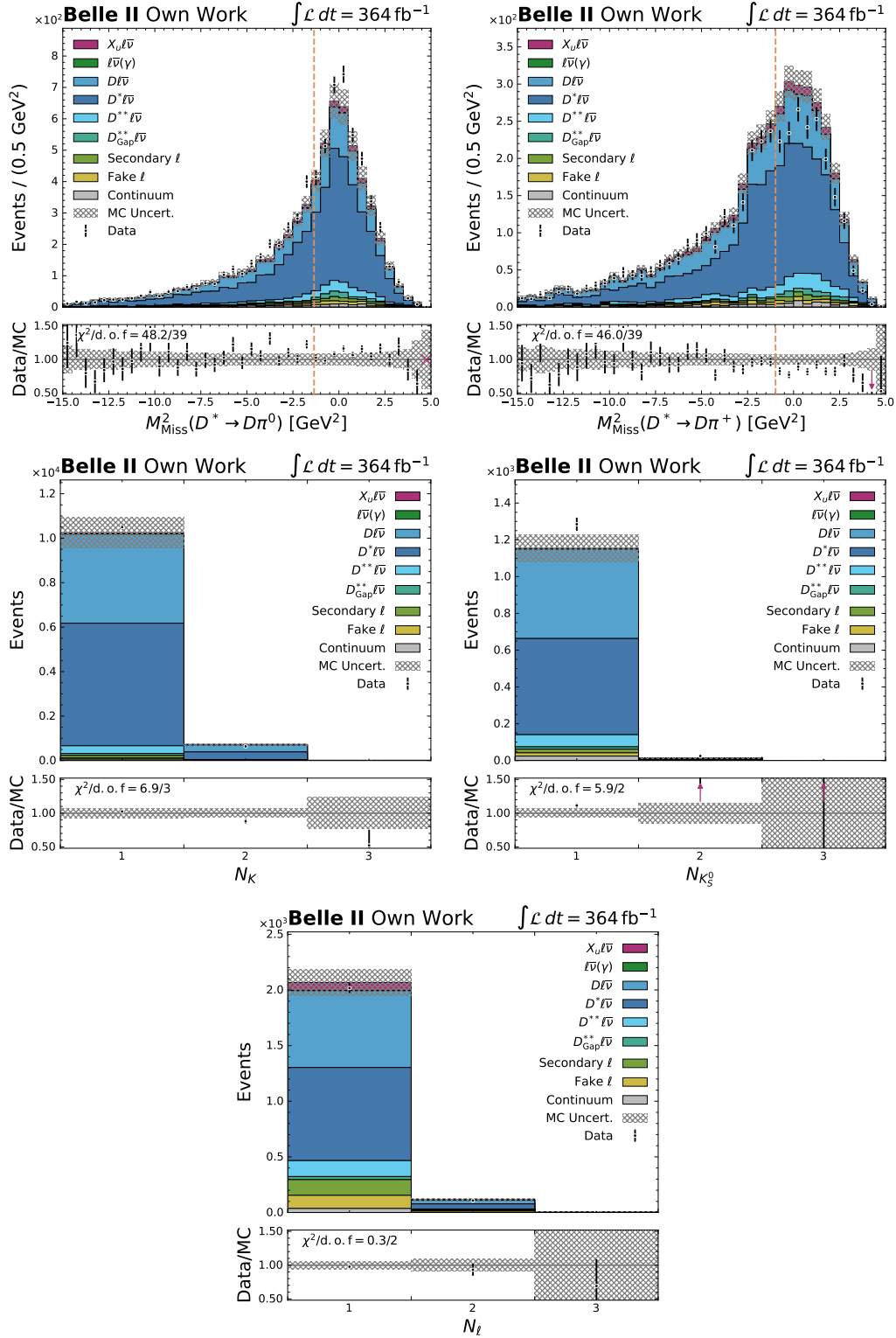


Figure 11.14: Data-MC agreement of the rest-of-event particle content based selection variables $M_{\text{Miss}}^2(D^* \rightarrow D\pi^+)$, $M_{\text{Miss}}^2(D^* \rightarrow D\pi^0)$, and, number of kaon, K_S^0 , and lepton candidates in the hadronic system. For the kaon and lepton counting variables, only the vetoed regions, $N_K \geq 1$, $N_{K_S^0} \geq 1$, $N_\ell \geq 1$, are shown.

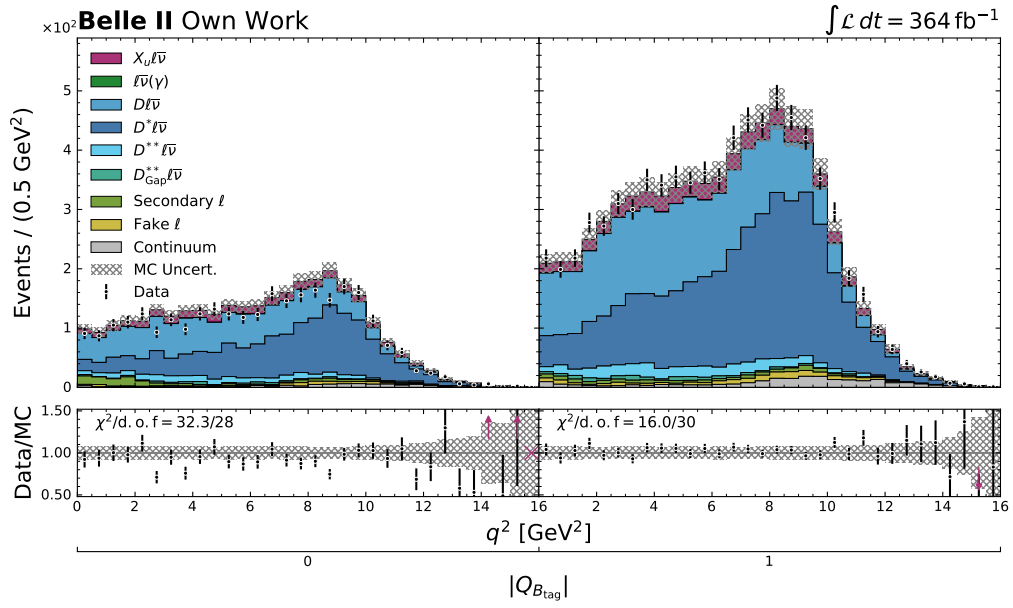


Figure 11.15: Data-MC comparison of the weak annihilation signal strength extraction observable, q^2 , in the medium lepton energy control region, split by tag B charge. For the calculation of the χ^2 value, bins with data yields or expected yields less than 5 events are merged into adjacent bins algorithmically. The χ^2 value is calculated using the statistical and systematic uncertainty taking into account correlations.

Fitting procedure

We present the fitting procedure which will be used in the search for weak annihilation. In addition to the statistical model, we discuss the considered systematic uncertainties and calibrations applied to the simulation. Finally, we discuss the expected sensitivity of the measurement. All presented results are based only on simulation.

We now turn our attention to the q^2 spectrum, from which we will extract the weak annihilation branching fractions by searching for a peak at high q^2 . The expected distribution of background events, with the three overlaid example weak annihilation models is shown in Fig. 12.1.

The extraction is performed using binned extended maximum likelihood fits, implemented via the `pyhf` package [275] following the prescriptions of Ref. [276]. The fits consider 13 bins of q^2 , 12 of width 2 GeV^2 , from 0 to 24 GeV^2 , and a wide bin between 24 and 30 GeV^2 . An alternative fit strategy which removes the low q^2 region, requiring $q^2 > 10 \text{ GeV}^2$, has been investigated but was not found to improve the sensitivity of the fit. This region is thus kept so that the fit may use it to constrain the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ and $\bar{B} \rightarrow X_u \ell \bar{\nu}$ backgrounds.

We simultaneously fit the charged and neutral B distributions, in the parlance of `pyhf` these are two channels, defined by the charge of the tag B . Fitting both B charges simultaneously allows the fit to constrain some shape and normalisation uncertainties using information from both samples, and to constrain the cross-feed contributions of weak annihilation events between the two samples.

Following the discussion in Sec. 4.3, the likelihood is defined as:

$$\mathcal{L} = \prod_c \prod_i \text{Pois}(N_{c,i} | \eta_{c,i}(\vec{\mu}, \vec{\theta})) \prod_j C(\theta_j), \quad (12.1)$$

where the index c runs over the considered channels, and the index i identifies the q^2 bins. The data yield in each bin is given by $N_{c,i}$, and the expected yield by $\eta_{c,i}(\vec{\mu}, \vec{\theta})$, which depends on the free floating parameters $\vec{\mu}$, and the constrained nuisance parameters $\vec{\theta}$. The term $C(\theta_j)$ gives the constraint on the j^{th} nuisance parameter.

Due to the number of decay channels that enter the selected sample, and the complex interactions with the detector, there is no analytic description of the shape of the signal or background components. We therefore turn to template methods, in which the distributions of the simulated sample are used to construct probability density functions. The expected number of events in each bin is then given by:

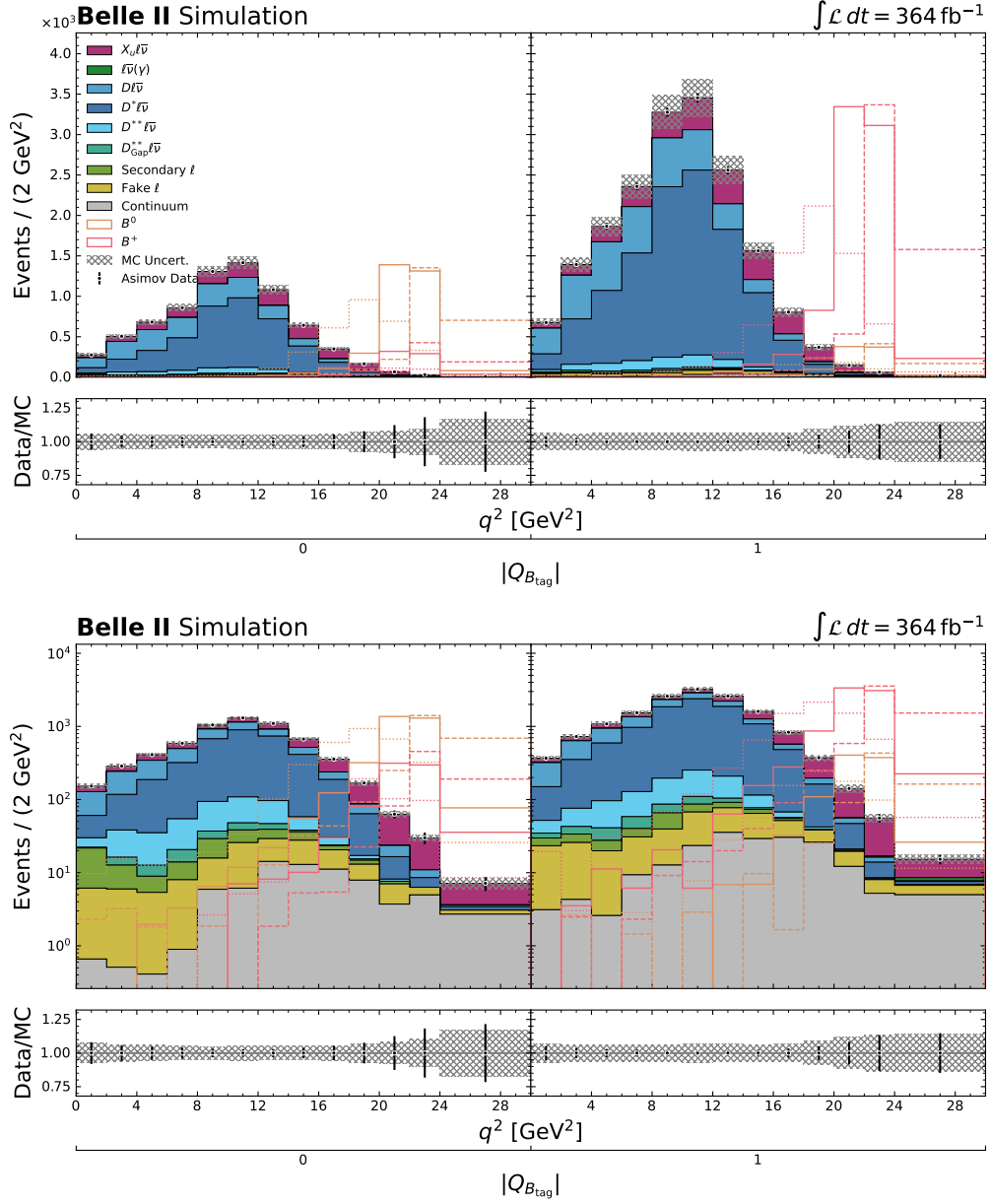


Figure 12.1: Expected q^2 distribution in the signal region with linear (*top*), and logarithmic scales (*bottom*). The binning is the same as used for the fits.

$$\eta_{c,i}(\vec{\mu}, \vec{\theta}) = \sum_s \eta_{c,i,s}(\vec{\mu}, \vec{\theta}), \quad (12.2)$$

where s represents a fit component and $\eta_{c,i,s}(\vec{\mu}, \vec{\theta})$ the yield of events for that component in bin c, i . We consider seven components. These are: $B^+ \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$, $B^0 \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$, $B^+ \rightarrow \ell \bar{\nu}(\gamma)$, $\bar{B} \rightarrow X_u \ell \bar{\nu}$, $\bar{B} \rightarrow X_c \ell \bar{\nu}$, secondary and fake leptons, and continuum events. For each component the yield is given by

$$\eta_{c,i,s}(\vec{\mu}, \vec{\theta}) = \left(\prod_{\kappa \in \boldsymbol{\kappa}} \kappa(\vec{\mu}, \vec{\theta}) \right) \left(\eta_{c,i,s}^0 + \sum_{\Delta \in \boldsymbol{\Delta}} \Delta(\vec{\theta}) \right), \quad (12.3)$$

where $\boldsymbol{\kappa}$ and $\boldsymbol{\Delta}$ are sets of multiplicative and additive modifiers, respectively, and $\eta_{c,i,s}^0$ the nominal yield of the component within the q^2 bin.

For each fit, there are four unconstrained parameters ($\vec{\mu}$). They modify the expected rate of the samples to which they are applied as $\kappa(\mu) = \mu$. Two are the parameters of interest, the signal strength of weak annihilation decays in the charged and neutral channels, $\mu_{\text{WA}}^{B^+}$ and $\mu_{\text{WA}}^{B^0}$, respectively. These samples are normalised such that their nominal yields correspond to weak annihilation contributions of 1% of the nominal $B^+ \rightarrow X_u^0 \ell^+ \nu$ and $B^0 \rightarrow X_u^- \ell^+ \nu$ branching fractions, respectively.

The other two floating parameters are the normalisation of $B\bar{B}$ events, given by μ^{B^+} and μ^{B^0} . These factors apply to all $B\bar{B}$ events reconstructed with charged and neutral B tags, respectively, including the weak annihilation contributions. We include these factors as the official FEI calibration, which we further discuss below, originates in part from studies of $\bar{B} \rightarrow X \ell \bar{\nu}$ events. As we analyse the $\bar{B} \rightarrow X \ell \bar{\nu}$ sample, the applied calibrations are therefore not independent of the sample we investigate. We thus float the normalisation of $B\bar{B}$ events in the fit.

In addition, approximately 165 nuisance parameters enter each fit, encoding the systematic uncertainties. The number varies as for some sources of systematic uncertainty, the number of considered nuisance parameters is truncated algorithmically.

12.1 Corrections and systematic uncertainties

Systematic uncertainties arise from a range of sources, including modelling of the signal and background components and detector performance. Each uncertainty can affect the normalisation of fit components, or their shape, or both. We first discuss the treatment of the systematics within the fit, before discussing the individual sources of systematic uncertainties.

To propagate the uncertainties into the fit, variations of the templates are constructed, corresponding to varying each source of systematic uncertainty by one standard deviation in the positive and negative direction. Each pair is associated with one nuisance parameter. If the fit pulls on the nuisance parameter, the template is morphed in the direction of the pull. Each nuisance parameter is associated with both a multiplicative modifier and

an additive modifier. The multiplicative modifier scales the total yields of the component, while the additive modifier shifts the shape of the component, without varying the total yield within a channel. Both modifiers use a polynomial interpolation between the two templates generated by varying the systematic uncertainty source by one standard deviation and extrapolate outside this range. For the multiplicative factor an exponential extrapolation is used, and for the additive a linear extrapolation is used.

The majority of systematic uncertainties that are considered naturally have two-sided variations. These correspond to systematic uncertainties such as branching fractions of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ channels in which there are auxiliary measurements from which we can take $\pm\sigma$ variations. If a source of systematic uncertainty depends on several correlated parameters, we first diagonalise the covariance matrix of the parameters. The parameters ($\vec{x}$) are then varied in the eigendirections of the covariance as:

$$\vec{x}_i^\pm = \vec{x}_i|_{\text{central}} \pm \sqrt{\lambda_i} \vec{e}_i. \quad (12.4)$$

where λ_i and $\vec{e}_i$ are the eigenvalues and eigenvectors of the matrix. For each pair of variations, one nuisance parameter is added to the fit.

Systematics in which we have only a one-sided variation, such as switching from the nominal inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ model (DFN) to the alternate (BLNP), are challenging to deal with, as there is no natural choice for the distribution of the prior for the nuisance parameter. Within this analysis, the choice of the inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ model is the most impactful nuisance parameters of this type. As the modelling of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ is in general not well known, we symmetrise such one-sided variations by mirroring about the nominal shape to create two-sided variations.

Several uncertainties, such as those associated to the statistical uncertainty on BDT based continuum shape corrections, or particle identification corrections, are propagated through the analysis as large sets of possible variations. To capture the bin-to-bin correlations of these uncertainties, we construct the covariance matrix of the binned MC samples. The covariance matrix is constructed in a global binning that assigns a unique bin index to the combination of {channel, component, q^2 bin}. We decompose the covariance matrix, S , as:

$$S = V \text{diag}(\vec{\lambda}) V^{-1}, \quad (12.5)$$

where V is the matrix of eigenvectors ($\vec{e}_i$) of S and $\text{diag}(\vec{\lambda})$ the diagonal matrix of eigenvalues, λ_i . The varied yields of the templates ($\vec{N}_i^\pm$) for variation i are then given by:

$$\vec{N}_i^\pm = \vec{N}_i|_{\text{central}} \pm \sqrt{\lambda_i} \vec{e}_i. \quad (12.6)$$

To capture the full covariance matrix, the number of variations, and hence nuisance parameters in the fit, would be equivalent to the number of bins in the global binning. Most of these nuisance parameters would be negligible. To avoid adding these, we truncate the

number of variations, taking the largest n variations such that the sum of their eigenvalues exceeds 99.5% of the sum of all eigenvalues and thus covers 99.5% of the variance.

We additionally consider a special case of the two-sided variations in which we build a covariance matrix from a large number of two-sided variations and follow the procedure above to reduce the number of nuisance parameters. This is used to reduce the number of nuisance parameters associated with the modelling of the charm decay from 232 to ~ 10 and to reduce the number of nuisance parameters associated with the FEI calibration.

Asymmetry about the nominal template in the upward and downward varied templates can cause biases in the fit. Most systematics we consider are naturally symmetric; however, some, such as form factor variations which depend non-linearly on the input parameters, variations which rely on alternate simulated samples, or variations after which the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ hybrid modelling procedure is repeated, can be asymmetric. To avoid bias in the fit, a symmetrisation procedure is adopted. We compute the distance of the varied templates from the nominal template in each bin and take the maximum envelope of the varied templates, maintaining relative sign differences between the variations.

12.1.1 Modelling

As we use template fits, modelling of the templates, especially for the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$ background components, is a crucial part of the analysis, with uncertainty in the modelling contributing significant systematic uncertainties to the extracted weak annihilation signal strengths.

12.1.1.1 Weak annihilation modelling

As discussed in Sec.7.3 we vary the width and mass of the generated hadronic system in weak annihilation decays about their nominal values, $M_X, \Gamma_X = (420_{-130}^{+140} \text{ MeV}, 280 \pm 140 \text{ MeV})$.

For each of the 70 WA models under consideration ($M_{\ell\bar{\nu}}, \Gamma_{\ell\bar{\nu}}$ pairs), we generate two million events for each variation of the M_X and Γ_X parameters. These events are binned into 1000 bins in $E_\ell^{B_{\text{sig}}}, m_x, q^2$ using a recursive equal frequency binning. The bins are selected such that events of the nominal model fall with equal frequency into 10 bins of $E_\ell^{B_{\text{sig}}}$. The events in each bin are then subdivided into 10 bins of M_x with equal frequency, which in turn are subdivided into 10 bins of q^2 . We generate weights as:

$$w_{ijk} = \frac{\Delta \mathcal{B}_{ijk}^{\text{varied}}}{\Delta \mathcal{B}_{ijk}^{\text{central}}}, \quad (12.7)$$

where i, j, k are the indices running over the three-dimensional binning, and $\mathcal{B}_{ijk}^{\text{varied/central}}$ gives the partial rate of $\bar{B} \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$ events falling within the bin for the varied or nominal sample, respectively. To propagate the variation through the reconstruction and analysis, the weights are applied to the nominal sample.

12.1.1.2 Leptonic decay modelling

The branching fractions of the leptonic channels are varied about their nominal values by their uncertainties as listed in Sec. 7.3.1.

12.1.1.3 Charmless semileptonic decay modelling

The branching fraction of the resonant channels are varied by the uncertainties listed in Table 7.2. For $\bar{B} \rightarrow \pi \ell \bar{\nu}$, $\bar{B} \rightarrow \rho \ell \bar{\nu}$, and $\bar{B} \rightarrow \omega \ell \bar{\nu}$ decays, we vary the input parameters of the form factors along the eigendirections of the parameter covariance matrix. This is performed using the **eFFORT** [253] framework which assigns event wise weights as

$$w_i = \frac{\Gamma_{\text{old}}}{\Gamma_{\text{new}}} \frac{d\Gamma_{\text{new}}/dq^2}{d\Gamma_{\text{old}}/dq^2}, \quad (12.8)$$

where Γ_{new} and Γ_{old} are the decay rates of the varied, and generated model respectively. The same procedure is used to update the sample to the nominal modelling if it differs from the model used to generate the events. For $\bar{B} \rightarrow \eta \ell \bar{\nu}$ and $\bar{B} \rightarrow \eta' \ell \bar{\nu}$ decays we replace the nominal model with the alternate description of Ref. [254] and assign the full difference as the uncertainty.

We vary the inclusive branching fractions by the uncertainties listed in Table 7.2. We vary the input parameters of the DFN model, M_b^{KN} and a^{KN} along the eigendirections of their covariance matrix. For each variation we generate event weights using Eq. 12.7 and a three dimensional binning in $E_\ell^{B_{\text{sig}}}$, M_X , q^2 . The bins are:

$$\begin{aligned} q^2 &: [0, 1.25, 2.5, \dots, 21.25, 22.5, 30] \text{ GeV}^2, \\ E_\ell^B &: [0, 0.125, 0.25, \dots, 2.5, 3.0] \text{ GeV}, \text{ and} \\ M_X &: [0, 0.3, 0.4, 0.5, \dots, 3.3, 3.5] \text{ GeV}, \end{aligned}$$

where the edges have been chosen to provide a finer binning than the bins used for the hybridisation procedure while having all bin edges of the hybrid binning as a subset of the edges within this binning.

As discussed in Sec. 7.2 there are several models describing inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ available, two of which are supported within EvtGen: DFN — our nominal model, and BLNP. We replace the nominal events with events generated following the BLNP description for an estimate of the systematic uncertainty associated with the choice of model. For the model parameters, we adopt the central values of Ref. [66]: $M_b^{\text{SF}} = 4.582 \text{ GeV}$, $\mu_\pi^{2,\text{SF}} = 0.202 \text{ GeV}^2$, $\mu_h = 0.4366 \text{ GeV}$, $\mu_i = 1.5 \text{ GeV}$, $\tilde{\mu} = 1.5 \text{ GeV}$. BLNP predicts a narrower inclusive event distribution than DFN (see Fig. 7.1), thus, at the nominal branching fractions, the predicted number of resonant events near the endpoint exceeds the number of inclusive events. To avoid the inclusive events being assigned negative weights, we merge the first two bins in M_X , and the respective final two bins in E_ℓ^B and q^2 .

The hadronisation of the hadronic system in nonresonant decays is controlled by several PYTHIA parameters. Of particular importance is the parameter known as γ_s , which

controls the suppression of s quark production relative to u or d . As we have vetoed events in which we have identified kaons, changes in the kaon production rate directly affect the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ reconstruction efficiency. The production rate of $e^+e^- \rightarrow s\bar{s}$ has been studied in Refs. [255, 256] at centre-of-mass energies of 12 and 36 GeV finding $\gamma_s = 0.35 \pm 0.05$ and $\gamma_s = 0.27 \pm 0.06$. We adopt 0.30 ± 0.09 ¹. We replace the nominal $\bar{B} \rightarrow X_u \ell \bar{\nu}$ sample with samples generated with γ_s varied by one standard deviation. We do not produce samples of $\bar{B} \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$ with varied γ_s , as the hadronic system is generally too soft for the production of kaon pairs.

For each discussed variation in branching fraction or modelling, we repeat the hybrid model procedure. Variations in the resonant component therefore do not change the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ branching fraction, only the relative rates of exclusive to inclusive events.

12.1.1.4 Charmed semileptonic decay modelling

We vary the branching fractions of the $\bar{B} \rightarrow X_c \ell \bar{\nu}$ channels by the uncertainties listed in Table 7.1. Analogously to the above discussion for resonant $\bar{B} \rightarrow X_u \ell \bar{\nu}$ channels, we vary the modelling of the channels by varying the input parameters in the eigendirections of their covariance matrix, using the HAMMER framework [277].

The HAMMER framework does not consider the widths of the broad D^{**} states directly, always assuming the central mass value. We follow the approach of Refs. [261, 278] and broaden the uncertainty on the form factors of these states, taking 1.5σ variations. Furthermore, for the semileptonic decays to broad D^{**} , an unphysical enhancement was observed at $w = 1$, where

$$w = \frac{M_B^2 + M_X^2 - q^2}{2M_B M_X}, \quad (12.9)$$

caused by events generated several widths away from the nominal mass. We adopt the procedure of Refs. [261, 278, 279] and add a generator level selection, requiring events to be at most 3 or 2.5 widths above the nominal D_0 or D_1' mass, respectively. The weight of the remaining events is increased to compensate for the removed events.

12.1.1.5 Simulated sample size

The limited size of the simulated sample used to generate the fit templates introduces statistical uncertainties to the shape of the templates and the expected yields. We propagate this uncertainty into the fit. As detailed in Sec. 11.1 we do not use the full available sample for the preparation for the measurement presented within this thesis. We thus scale the MC statistical uncertainties to be representative of a measurement performed with the full sample.

¹The default value used in Belle II is $\gamma_s = 0.217$, to avoid generating extra samples close to this value we use the samples with the default value as the -1σ variation.

12.1.1.6 $\Upsilon(4S)$ branching fraction

The rate at which $\Upsilon(4S)$ decays to charged or neutral B pairs, f_{+-} can impact the measurement due to cross-feed between the two reconstructed samples. We therefore vary the rate by the PDG uncertainty, $f_{+-} = 51.4 \pm 0.6\%$ [1].

12.1.1.7 Charm decay modelling

Inclusive studies at Belle II have reported poor agreement between data and MC in quantities related to the hadronic system (M_X , M_{Miss}^2 , q^2 , ...) [218, 261, 269, 278, 280]. One possible origin of the observed disagreement is the modelling of the charm decay. We therefore reweight the decay of the charm mesons to the most up to date values² [65], and more importantly, attempt to capture the uncertainty of these branching fractions. As the FEI calibrations have been measured without this update, and we have separately calibrated the continuum sample, we consider only the decay of D and D^* mesons which originate from the same B meson as the signal lepton.

There are several caveats to the reweighting procedure that should be noted.

- It is only possible to update channels that were simulated. Thus, there are channels, although usually very rare, which have been measured but are not reproduced in the simulated sample.
- When a decay has been measured through a decay channel of an intermediate resonance (eg. $D^0 \rightarrow K^- a_1^+ (a_1^+ \rightarrow \rho^0 \pi^+)$), then the branching fraction is corrected assuming the branching fraction of the intermediate resonance as set in the simulation ($\mathcal{B}(a_1^+ \rightarrow \rho^0 \pi^+) = 0.30875$). No uncertainty is associated to this branching fraction.
- If only a limit is available, the branching fraction is set to the limit with 100% uncertainty.
- Uncertainties are taken to be fully uncorrelated. This can lead to significantly inflated uncertainties when a poorly known resonance contribution is subtracted from a well known inclusive rate.
- The reweighting procedure only considers the branching fraction. It is not feasible to reweight the decay models, as for most channels no better model exists. A significant portion of charm decays are modelled via an equidistribution of particles in phase space, and thus likely to still be poorly modelled even with corrected branching fractions.

²For some decays we must refer back to the 2008 PDG [281] as subsequent versions no longer list all sub-modes of $D^0 \rightarrow K\pi\pi\pi$.

- The $D^0 \rightarrow K^- a_1^+$ rate, as measured through $a_1 \rightarrow \rho^0 \pi^+$, has increased several percentage points from 7.8% to 14%. Combined with all other measured channels at their respective central values, and zeroing out the unmeasured $D^0 \rightarrow K_{S/L} \pi^+ \pi^- \pi^0 \pi^0 \pi^0$, the total D^0 branching fraction is $\sim 102\%$. As the branching fractions of the a_1 are not well known, we decrease the $D^0 \rightarrow K^- a_1^+$ branching fraction to $11.94 \pm 2.3\%$, such that the total D^0 branching fraction correctly sums to unity. The difference to the nominal branching fraction is added as an additional uncertainty on this channel.
- For D^+ the opposite problem is encountered, with the measured branching fractions covering approximately 98% of the total decay rate. We therefore increase the branching fraction of the unmeasured channels ($D^+ \rightarrow \eta 5\pi$) by a factor of ~ 2.5 over what is currently used. The branching fractions of these channels are assigned an uncertainty of 100%.

Table 12.1 presents a summary of the of final state particle production rates before and after the reweighting in comparison with the world average inclusive measurements. The reweighting procedure has largely shifted the inclusive particle production rates further from the measured values. The rate of inclusive neutral kaon production in the D^0 channel shows a deviation of approximately 2.3σ from the world average rate. As the updated values are the most up-to-date branching fractions available, we nonetheless adopt them. To estimate the uncertainty associated with the decay of the charm mesons in the hadronic system, we vary the branching fractions of the 232 $D^{(*)0}$ and $D^{(*)+}$ decay channels by their uncertainties, without conserving probability.

Final State	D^0 [%]			D^+ [%]		
	PDG	Generated	Updated	PDG	Generated	Updated
$K_S^0 X$	-	20.44	18.48 ± 1.46	-	30.39	32.01 ± 0.78
$K_L^0 X$	-	20.12	18.50 ± 1.46	-	30.32	31.95 ± 0.78
$(K_S^0 \text{ or } K_L^0) X$	47.00 ± 4.00	40.07	36.46 ± 2.45	61.00 ± 5.00	59.70	63.32 ± 1.27
$K^+ X$	3.40 ± 0.40	3.36	3.24 ± 0.27	5.90 ± 0.80	5.60	4.40 ± 0.15
$K^- X$	54.70 ± 2.80	55.85	59.84 ± 4.33	25.70 ± 1.40	27.66	27.57 ± 1.20
$(K^+ \text{ or } K^-) X$	-	56.47	60.45 ± 4.33	-	31.41	30.23 ± 1.20
$\mu^+ X$	6.80 ± 0.60	6.27	6.96 ± 1.34	17.60 ± 3.20	16.26	15.27 ± 0.25
$e^+ X$	6.49 ± 0.11	6.54	7.20 ± 1.32	16.07 ± 0.30	16.17	16.11 ± 0.82

Table 12.1: Comparison of world averaged inclusive particle production rates [65], the production rates as generated in the official Belle II MC, and the production rates after the $D^{(*)}$ decay reweighting procedure.

12.1.1.8 Continuum reweighting

The full size of the normalisation corrections discussed in Sec. 11.2.4, are assigned as systematic uncertainties. The corrections for the charged and neutral tagged samples are taken to be uncorrelated.

Uncertainties on the BDT based shape correction arising from the limited statistics of the off-resonance data and simulated samples are propagated by training 200 additional classifiers [282] on bootstrapped samples of the off-resonance data and MC.

12.1.2 Detector Effects

Several detector effects impact the reconstruction, changing the shape of the templates and reconstruction efficiencies.

12.1.2.1 Charged particle identification

Identification and misidentification efficiencies for charged particles have been studied in dedicated control channels and have been characterised in bins of momentum magnitude, polar angle, and track charge. In each bin, the Belle II performance group provides calibration factors, given by the ratio of identification efficiencies in data over MC, and their associated statistical and systematic uncertainties, where the systematic uncertainties combine systematic uncertainties arising from the measurement of each control channel, and uncertainties assigned due to performance differences observed between control channels.

For lepton identification, the corrections originate from studies of $J/\psi \rightarrow \ell^+\ell^-$, $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$, $e^+e^- \rightarrow e^+e^-(\gamma)$ and $e^+e^- \rightarrow \mu^+\mu^-(\gamma)$. For hadron-to-lepton misidentification rates $K_S^0 \rightarrow \pi^+\pi^-$, $e^+e^- \rightarrow \tau^+\tau^-$, $D^* \rightarrow D^0(\rightarrow K\pi)\pi_{\text{slow}}$, and $\Lambda^0 \rightarrow p\pi^-$ events are used. For charged kaon identification efficiencies and pion to kaon misidentification rates $D^{*-} \rightarrow D^0(\rightarrow K^-\pi^+)\pi_{\text{slow}}^-$, and $K_S^0 \rightarrow \pi^+\pi^-$ events are used. For an example of some of these studies, see the control channel studies of the improvements to the ECL charged particle identification in Ch. 6. Note that the calibration studies use the full data sample and more complex procedures to extract the particle identification efficiencies.

We correct the respective efficiencies and misidentification rates using the nominal weights and propagate the uncertainties using toy Monte Carlo methods. We draw 200 variations of the calibration factors for lepton identification and kaon identification, using multivariate Gaussians, about the nominal values, where the statistical uncertainties are taken to be fully uncorrelated and the systematic for efficiency or misidentification correction to be fully correlated.

In a small number of cases, mostly $K \rightarrow \ell$ misidentification rates, the tracks fall outside the coverage of the correction tables. When this occurs, the correction factor is assumed to be one with an uncertainty that combines the average of the total uncertainties in all measured bins and the average difference of the measured correction factors from unity in quadrature.

When used to correct veto efficiencies, rather than a binary veto of events in simulation, we apply a factor of $(1 - w_k)$, where w_k is the ratio of identification efficiency in data to simulation, such that if identification in data is less efficient than in simulation, the simulated events are kept with a small weight. If identification in data is more efficient than in simulation, the simulated events are assigned a negative weight. This rarely occurs. When performing the binned fit, all bin yields have positive expectation values.

12.1.2.2 K_S^0 reconstruction

The reconstruction efficiency of K_S^0 particles has been studied in the control channel $D^{*-} \rightarrow D^0(\rightarrow K_S^0 \pi^+ \pi^-) \pi_{\text{slow}}^-$, finding a flight distance dependent overestimation of efficiency in simulation. We correct for the effect ($0.55\% \text{ cm}^{-1}$) and assign the full correction as the systematic uncertainty.

12.1.2.3 Tracking

The efficiency of track finding has been studied with $e^+e^- \rightarrow \tau^+\tau^-$ events, finding excellent agreement between data and simulation. We propagate a systematic uncertainty of 0.24% per medium and high momentum track, $p > 0.2 \text{ GeV}$, in the hadronic system and signal lepton by assigning event-wise weights of $(1 \pm 0.0024)^N$, where N is the number of tracks. Tracks on the tag side are excluded as the FEI calibration corrects for tracking efficiency differences.

12.1.2.4 Slow pion reconstruction

The reconstruction efficiencies of charged and neutral slow pions have been studied in three momentum bins, between 50 and 200 MeV using dedicated D^* control samples. For events vetoed due to the properties of the D^* inferred from slow pions, we correct the veto efficiency and propagate the uncertainties, which contain statistical and correlated systematic components. We additionally use the calibration to correct and propagate the uncertainty on the low momentum tracking efficiency.

12.1.2.5 Photon reconstruction efficiency

The photon reconstruction efficiency has been studied using an $e^+e^- \rightarrow \mu^+\mu^-\gamma$ control sample. For high-energy photons excellent agreement was found between data and MC. For low energy photons the reconstruction efficiency was found to be 3–5% higher in simulation than data. As the effect is small, affecting only low energy photons which do not contribute significantly to the four-momentum of the hadronic system, no correction is applied.

12.1.2.6 Full event interpretation efficiency

As discussed for the Belle measurement, the FEI exhibits large differences in tagging efficiencies between data and MC, caused by mismodelling of charm and hadronic B decay branching fractions, resonant substructures, and variables considered by the FEI classifiers. The Belle II performance group provides sets of calibration factors, in the form of ratios of reconstruction efficiency in data over MC, for the reconstructed B channels, providing individual calibration factors for the 11 or 10 most populated channels for charged and neutral B decays, respectively, and a further factor for each charge combining all remaining channels.

The calibrations are performed in two channels, inclusive $\bar{B} \rightarrow X\ell\bar{\nu}$ events with $E_\ell^{B_{\text{sig}}} > 1$ GeV, and a recoil method using the bachelor pion of $\bar{B} \rightarrow D^{(*)}\pi^+$ decays. The two methods are combined by a χ^2 fit, assigning additional systematic uncertainties to observed differences between the two calibrations. The combined calibration is dominated by the inclusive semileptonic measurement, due to its higher statistical precision.

The $\bar{B} \rightarrow X\ell\bar{\nu}$ sample is the same used within this analysis, the statistical and systematic uncertainties of the calibrations are thus expected to be highly correlated to the uncertainties of our study. Therefore, as described in the introduction of this section, we float the neutral and charged $B\bar{B}$ yields within the fit. As the fit is only able to float all channels simultaneously, the FEI calibrations are still applied, and their uncertainties propagated as a best estimate for the impact of the channel dependence of the FEI. The uncertainties, which are dominated by $\bar{B} \rightarrow X_c\ell\bar{\nu}$ branching fractions, are taken to be uncorrelated to any other uncertainties within the fit. This is expected to be a conservative simplification, as the $\bar{B} \rightarrow X_c\ell\bar{\nu}$ branching fractions are expected to be anticorrelated to the uncertainty of the FEI calibration.

12.1.2.7 Full event interpretation cross-feed

FEI cross-feed, which corresponds to mistagging a B^+B^- event as $B^0\bar{B}^0$ or vice versa, has the potential to be a significant systematic if the rates of weak annihilation in the neutral and charged channels differ significantly. Only preliminary studies of the cross-feed rates exist. We therefore aim to estimate these rates via a three-dimensional fit to the absolute tag B charge, the FEI signal probability $\log_{10} \mathcal{P}_{\text{FEI}}$ and the event shape variable, $\cos \theta_{\vec{T}_B \vec{T}_O}$, where the first two variables constrain the cross-feed rates and the third the continuum component, which shares a similar shape to the cross-feed sample in FEI signal probability. The fit will be performed in the medium lepton energy region, $1.0 \leq E_\ell^{B_{\text{sig}}} < 1.5$ GeV, to have a statistically independent sample to the nominal region. All other selections are identical to the nominal fit region. Four parameters are floated, the yields of the correctly reconstructed and cross-feed components in the charged and neutral B tagged samples, respectively.

All uncertainties discussed previously are included in the fit as nuisance parameters. As we are interested only in the cross-feed rate, and not efficiency differences, the templates

associated with the up and down variations of the nuisance parameters are normalised to the nominal yield, such that the nuisance parameters cause only shape variations. The post-fit distribution to a simulated sample is shown in Fig. 12.2. We expect to measure the purity of the charged and neutral bins with precisions of 2.5% and 2.8%, respectively.

The FEI cross-feed rate is expected to be signal-side channel dependent. As we are unable to measure it directly for the individual signal side channels $\bar{B} \xrightarrow{\text{WA}} X_u \ell \bar{\nu}$, $\bar{B} \rightarrow X_u \ell \bar{\nu}, \dots$, we must assume that while there may be an overall shift of cross-feed rates between data and MC, the difference between different channels is approximately correctly reproduced in MC. We therefore propagate the uncertainty by using the above fit result to derive two scale factors, the ratio of cross-feed events in the charged or neutral tagged sample found by the fit over the number of respective cross-feed events in MC. The scale factor is applied to the cross-feed events in each signal side channel; we then rescale the correctly reconstructed events such that the yield in the neutral and charged channels does not change.

As the FEI cross-feed calibration procedure requires knowledge of the relative rates of the signal side channel in charged and neutral decays B , we cannot perform this scaling for the weak annihilation sample. Therefore, we apply the scale factors found for $\bar{B} \rightarrow X_u \ell \bar{\nu}$ as the topologies are similar and hence the cross-feed rates are expected to be similar. This allows for the systematic uncertainty associated with this cross-feed to additionally have a small efficiency effect; however, this is negligible.

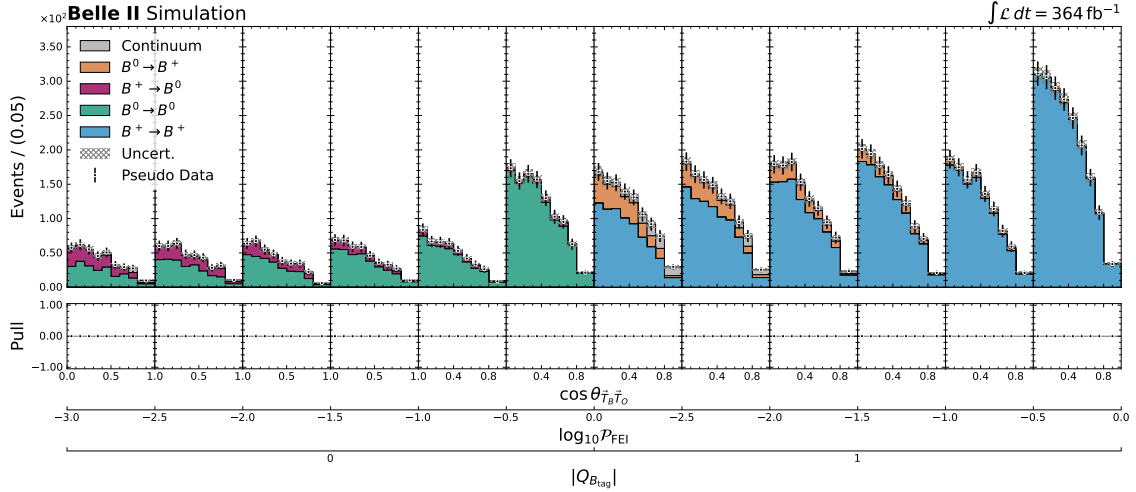


Figure 12.2: Post-fit distribution of the fit used to estimate the FEI cross-feed rates using an Asimov sample.

In Sec. 8.4 we observed that the difference in FEI performance between data and simulation depends on the quality of the tag B . Cross-feed tags are, by definition, poor tags. They are therefore expected to be distributed similarly in $\mathcal{P}_{\text{FEI}}$ to all other poor tags. As discussed, the available FEI calibrations consist of one factor per channel, without dependence on the tag quality. It is therefore not clear whether the above-proposed

procedure would correctly measure the cross-feed rate or whether the measured cross-feed rate would absorb the difference in calibration between good and bad tags. This remains under investigation. Alternatively, it may be sufficient to calibrate the FEI as a function of $\mathcal{P}_{\text{FEI}}$, which is therefore expected to improve the modelling of the cross-feed. For the remainder of this thesis, we assume the above procedure to be valid and estimate the expected sensitivities using the above-found cross-feed rates.

12.2 Fit validation

We validate the fitting procedure with toy MC methods, studying nine representative weak annihilation models. These correspond to all pairwise combinations of $M_{\ell\bar{\nu}}^2 = \{18.5, 22.1, 24.3\}$ GeV, $\Gamma_{\ell\bar{\nu}} = \{0.02, 0.141, 1.0\}$ GeV.

For each test, the content of each bin is varied about its expectation values following a Poisson distribution. Additionally, for each nuisance parameter, a new value of the auxiliary measurement constraining the nuisance parameter is drawn following a Gaussian distribution.

12.2.1 Pull test

To test the fit for biases and to ensure the estimated uncertainties are accurate, we perform a pull test. For each WA model under study, we generate 2000 toy pseudo experiments centred on $(\mu_{\text{WA}}^{B^+}, \mu_{\text{WA}}^{B^0}) = (1, 1)$. As the WA process is rare, we expect signal yields in the tens of events or below and the uncertainty to be asymmetric. Therefore, we define the pull, g as [283]:

$$g(\mu, \sigma) = \begin{cases} \frac{\bar{\mu} - \mu}{\sigma_+}, & \text{if } \mu < \bar{\mu} \\ \frac{\mu - \bar{\mu}}{\sigma_-}, & \text{otherwise} \end{cases} \quad (12.10)$$

where μ is the fitted value, $\bar{\mu}$ the true value, and σ_+, σ_- the asymmetric uncertainties on μ in the upward and downward directions, respectively. If the fitter measures the central value without bias and correctly estimates the uncertainties, the pull distribution is expected to be a Gaussian centred on zero with a width of one. An example of pull distribution is shown in Fig. 12.3. The complete set of pull distributions for $\mu_{\text{WA}}^{B^+}$ and $\mu_{\text{WA}}^{B^0}$ for the example samples are presented in Fig. C.1. The pull distributions are found to be consistent with Gaussians with means of zero and widths of one, indicating that the fitter is unbiased and correctly estimates the uncertainties.

12.2.2 Linearity test

To ensure that the fit is capable of accurately determining the signal strengths over the possible range, we perform what is known as a linearity test. We vary the input strength of the weak annihilation signal, testing 11 values, $\bar{\mu} = \{0.01, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 2.5, 3.0\}$. For each test value, we generate 1000 pseudo experiments. For each set of

toys, we extract the median measured signal strength. Finally, we fit a linear trend line to the medians against the input strengths.

As the strengths of the WA contribution in the neutral and charged channels are uncoupled, we additionally test whether the linear behaviour in one of the signal strengths is maintained as the magnitude of the other varies. Therefore, we repeat the procedure three times for the neutral and charged WA signal strengths independently, setting the signal strength of the other contribution to 0.1, 1.0 and 3.0. An example result of the linearity test is shown in Fig. 12.3. The results of the linearity tests for $\mu_{\text{WA}}^{B^0}$ and $\mu_{\text{WA}}^{B^+}$ are presented in Fig. C.2. The fits are found to accurately measure the input signal strengths without bias.

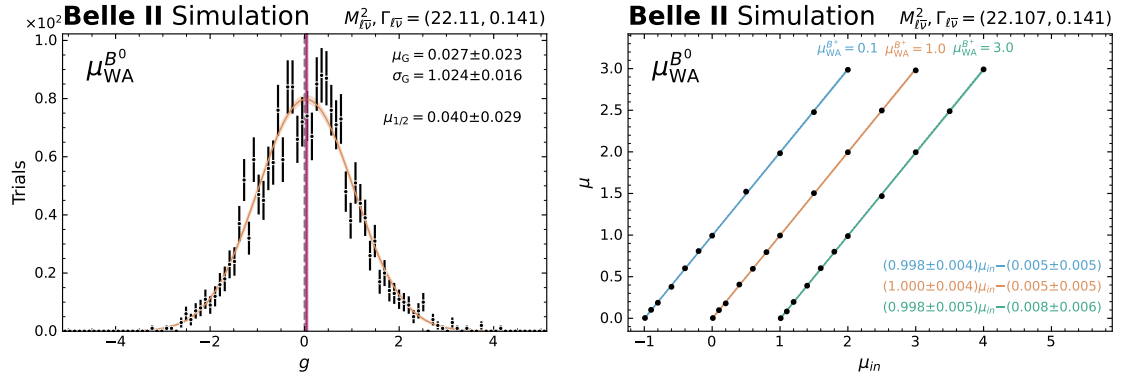


Figure 12.3: Example pull test (*left*) and linearity test (*right*) for the weak annihilation signal strength in the neutral B channel, $\mu_{\text{WA}}^{B^0}$, using the fit configuration $M_{\ell\bar{\nu}}^2 = 22.1 \text{ GeV}^2$, $\Gamma_{\ell\bar{\nu}} = 0.02 \text{ GeV}$. The linearity tests setting $\mu_{\text{WA}}^{B^+}$ to 0.01 and 3 have been offset in the x -axis for visualisation. The full set of pull and linearity tests for $\mu_{\text{WA}}^{B^0}$ and $\mu_{\text{WA}}^{B^+}$ are available in appendix C.

12.3 Fit uncertainty

The expected uncertainty and its breakdown into statistical and systematic components depends strongly on the signal distribution and normalisation in data.

An example of the breakdown for test fit configurations is presented in Fig. 12.4. For each of the 70 weak annihilation models under investigation, the fit is performed on an Asimov sample containing an injected weak annihilation signal generated with the same model as used in the fit. The injected signal is normalised to $\mu_{\text{WA}}^{B^0}, \mu_{\text{WA}}^{B^+} = (1, 1)$. For each fit, the statistical uncertainty is extracted by means of 4000 pseudo-experiments, varying the expected yield in each bin about its expectation according to a Poisson distribution. The uncertainty is taken as the asymmetric inter-percentile range. The systematic uncertainty is estimated to be the quadratic difference between the total and statistical uncertainty.

For the test configurations, the uncertainty on $\mu_{\text{WA}}^{B^0}$ is dominated by the statistical uncertainty. For $\mu_{\text{WA}}^{B^+}$, the statistical and systematic uncertainties are of similar scales.

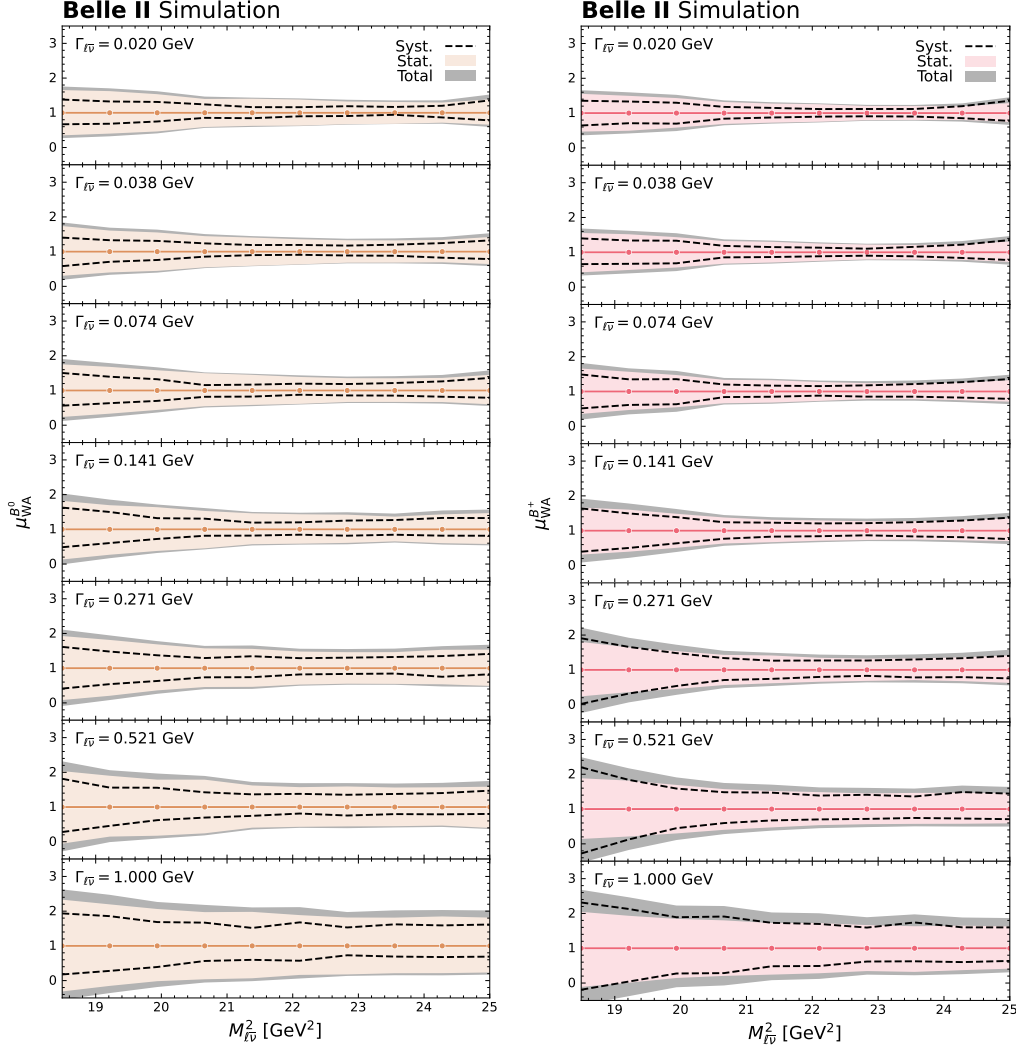


Figure 12.4: Breakdown of the total uncertainty on the parameters of interest, the weak annihilation signal strengths in the neutral and charged B channels, $\mu_{\text{WA}}^{B^0}$ (*left*) and $\mu_{\text{WA}}^{B^+}$ (*right*), respectively. For each fit configuration, an Asimov sample containing injected weak annihilation signal generated with the same parameters as used for the fit is used. The injected signal is normalised to 1% of the charged and neutral $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate.

12.3.1 Systematics impact

Some insight, such as which systematic uncertainties are expected to be dominant, can be gained from examining reference models. We consider five reference models. The models

cover high and low $M_{\ell\bar{\nu}}$ values, and narrow and wide distributions. They are:

$$\begin{aligned}(M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (19.9 \text{ GeV}^2, 0.141 \text{ GeV}), \\(M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (22.1 \text{ GeV}^2, 0.020 \text{ GeV}), \\(M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (22.1 \text{ GeV}^2, 0.541 \text{ GeV}), \\(M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (24.3 \text{ GeV}^2, 0.141 \text{ GeV}).\end{aligned}$$

For each fit configuration, we fit an Asimov sample, where the yields in each bin of the pseudo-data are set exactly to the expectation of the simulation. Each sample contains an injected weak annihilation signal generated with the same parameters as used for the fit. The injected signal is normalised to 1% of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate, ie. $\mu_{\text{WA}}^{B^0} = \mu_{\text{WA}}^{B^+} = 1$.

To determine which systematic uncertainties are important to the fit, we test the impact of varying each nuisance parameter about its fitted value, $\hat{\theta}_i$. Each nuisance parameter is shifted up and down by its post-fit uncertainty, $\Delta\hat{\theta}_i$. The parameter is then fixed and the fit is repeated. The impact of the nuisance parameter on the measurement is given by the change in the parameter of interest with respect to the nominal fit result, $\Delta\mu$. Moreover, the same procedure is repeated using the pre-fit uncertainty, $\Delta\theta_i$. A comparison of the observed shift in the parameter of interest when shifting the nuisance parameter by either its pre-fit or post-fit uncertainty gives an indication of the fits ability to constrain the nuisance parameter.

The 30 most impactful nuisance parameters for the determination of $\mu_{\text{WA}}^{B^+}$ and $\mu_{\text{WA}}^{B^0}$ for the example fit configurations are given in Fig. 12.5 and Fig. 12.6, respectively. Some general trends can be observed:

- Weak annihilation models centred at lower $M_{\ell\bar{\nu}}^2$ values, or that are broader, are subject to larger uncertainties related to the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ modelling, as their distributions become more similar. This includes the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ branching fractions, the DFN input parameters, and the choice of inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ model. Stronger selections on the kinematics and particle content of the final state, such as vetoing events in which the hadronic system is consistent with the decay products of a ρ or ω meson, may be able to reduce the size of the systematic uncertainty associated with the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ model. Such selections, however, will introduce additional dependence on the modelling of the hadronic system in weak annihilation, and are thus not considered. The $\bar{B} \rightarrow X_u \ell \bar{\nu}$ component thus represents an irreducible background.
- For narrow high $M_{\ell\bar{\nu}}^2$ models, the modelling of the hadronic system, and its impact on the shape of the q^2 distribution, is the dominant systematic.
- The size of the simulated sample (MC statistics) in the three highest q^2 bins are amongst the largest systematics for all fit configurations.

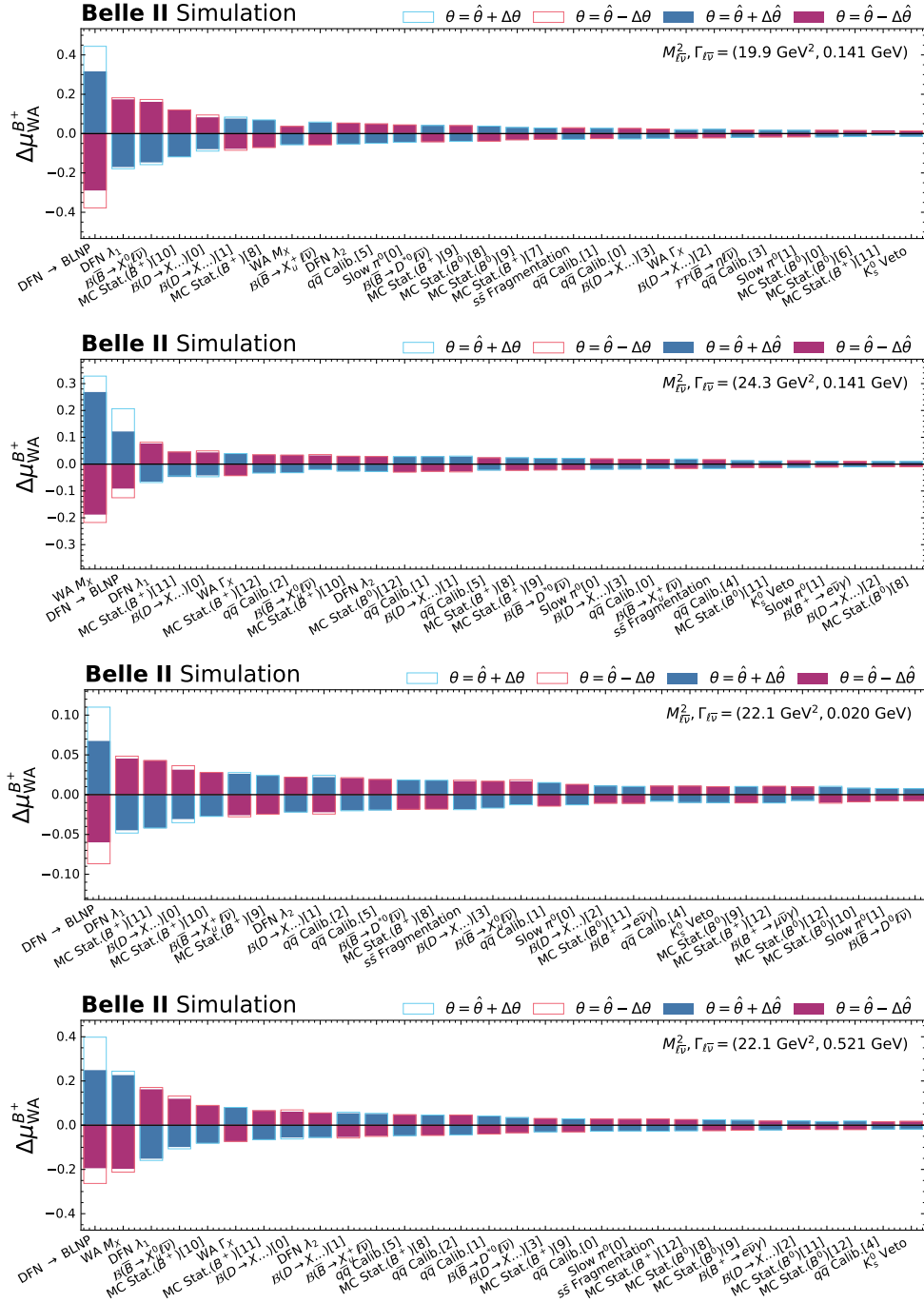


Figure 12.5: Post-fit impact of the leading 30 nuisance parameters on the parameter of interest, for the weak annihilation signal strength in the charged B channel, μ_{WA}^{B+} , for four test fit configurations. For each fit configuration, an Asimov sample containing injected weak annihilation signal generated with the same parameters as used for the fit is used. The injected signal is normalised to 1% of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate. If a source of systematic uncertainty is associated with multiple nuisance parameters, the index of the nuisance parameter is given in square brackets. For the uncertainty associated to the limited size of the MC sample (MC Stat.) this corresponds to the bin index in q^2 .

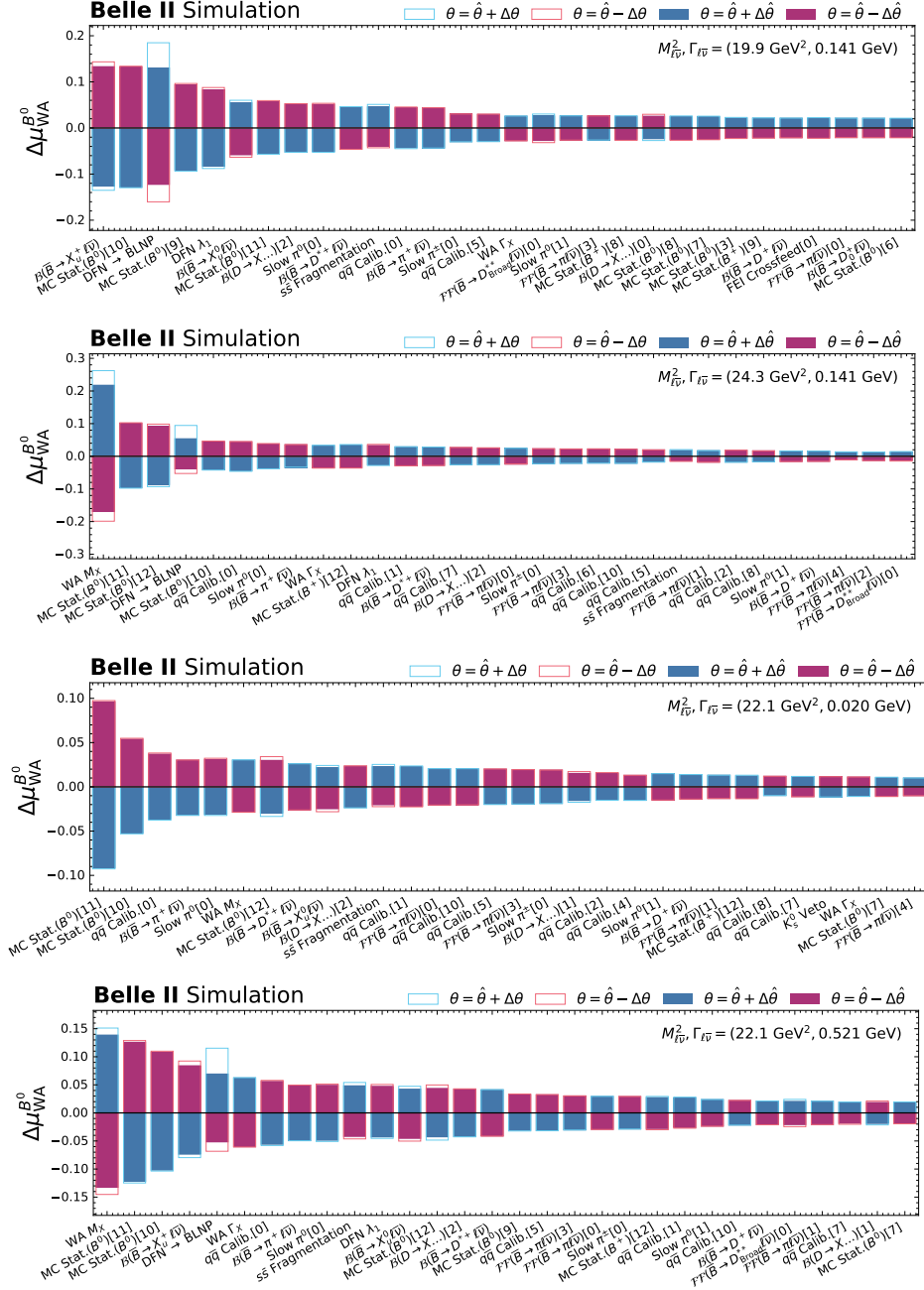


Figure 12.6: Post-fit impact of the leading 30 nuisance parameters on the parameter of interest, for the weak annihilation signal strength in the neutral B channel, $\mu_{\text{WA}}^{B^0}$, for four test fit configurations. For each fit configuration, an Asimov sample containing injected weak annihilation signal generated with the same parameters as used for the fit is used. The injected signal is normalised to 1% of the total $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate. If a source of systematic uncertainty is associated with multiple nuisance parameters, the index of the nuisance parameter is given in square brackets. For the uncertainty associated to the limited size of the MC sample (MC Stat.) this corresponds to the bin index in q^2 .

12.4 Expected limits

Finally, we estimate the 90% upper limits that the analysis will be able to set on the weak annihilation branching fractions for charged and neutral B mesons. As discussed in Sec. 4.3.1 we use the CL_s technique with the $\tilde{q}$ test statistic.

For each weak annihilation model tested, there are two parameters to scan, $\mu_{\text{WA}}^{B^0}$ and $\mu_{\text{WA}}^{B^+}$. As there is some crossfeed between the two samples, for the most rigorous results, one should perform a two-dimensional scan of the signal strengths. However, as the crossfeed between the channels is quite limited, and therefore the correlation between the measured signal strengths is expected to be low, we place limits on the two signal strengths independently. Additionally, performing a two-dimensional scan for each of the 70 models under consideration is computationally expensive, and the recovered limits difficult to interpret due to the high dimensionality. When scanning hypothesised values of one of the parameters of interest, the other is therefore profiled by the fit, allowed to float freely, including becoming negative.

To understand the behaviour of the measured limits, we perform a qualitative comparison of the limits recovered across all studied models when injecting WA signal events into the pseudo-data. Three test WA signals are injected, corresponding to models with

$$\begin{aligned}(M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (19.9 \text{ GeV}^2, 0.521 \text{ GeV}), \\ (M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (22.1 \text{ GeV}^2, 0.141 \text{ GeV}), \\ (M_{\ell\bar{\nu}}^2, \Gamma_{\ell\bar{\nu}}) &= (24.3 \text{ GeV}^2, 0.020 \text{ GeV}),\end{aligned}$$

with injected strengths of $\mu_{\text{WA}}^{B^0} = \mu_{\text{WA}}^{B^+} = 1$. The remainder of the pseudo-data corresponds to the Asimov sample of the non-WA channels.

The expected limits³ if there was no weak annihilation contribution and the observed limits for each test scenario are presented in Fig. 12.7. For injected models with broad distributions in q^2 , we observe that fits assuming narrow WA shapes are not sensitive to the injected events, recovering 90% upper limits on the signal strengths approximately equal to the expected limits. For injected narrow distributions, weak limits are recovered for fits assuming broad WA shapes, as the fits are able to construct a narrow peak by reducing the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ contribution and including a large weak annihilation contribution.

For a qualitative summary of the limits we expect to set with the analysis, we consider a fiducial set of models. The fiducial region corresponds to the 42 models within the model space: $19.94 \leq M_{\ell\bar{\nu}}^2 \leq 24.28 \text{ GeV}^2$ and $\Gamma_{\ell\bar{\nu}} \leq 0.521 \text{ GeV}$, excluding extremely broad models, and models far away from the parton-level endpoint, $M_{\ell\bar{\nu}}^2 = M_b^2 \sim 4.7^2 \text{ GeV}^2$, which are considered unlikely. In order to report a single limit for the weak annihilation contribution to the charged and neutral charmless semileptonic B decay, we consider the

³The expected limits are calculated by fitting the pseudo-data with the weak annihilation contributions fixed to zero, $\mu_{\text{WA}}^{B^0} = \mu_{\text{WA}}^{B^+} = 0$. The limit is then calculated from toys based on the fitted background yields. In this way, pulls on the nuisance parameters of the background contributions are incorporated into the expected limit when the analysis is performed on data.

weakest limit set with any model in the fiducial region. The weakest 90% upper limit on the signal strength of weak annihilation measured by these models, when injecting test signal samples at various strengths are given in Table 12.2 and Table 12.3.

If the weak annihilation contribution to charmless semileptonic decays is sufficiently suppressed in nature to be much less than 1% of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate we expect to recover 90% upper limits of 1.50% and 1.42% for the neutral and charged modes, respectively. If the true rate is 1% of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate, the measurement is expected to recover limits between 2.3–3.1% and 2.3–4.2% for the neutral and charged modes, respectively, and between 4.4–6.8% and 4.3–9.1%, respectively, if the true rate is 3% of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ rate. As discussed above, the weakest limits are recovered when the injected signal is narrow.

If the true weak annihilation rate is less than or equal to 1%, the limits on the weak annihilation contribution set in both channels are therefore expected to be competitive with the upper bounds estimated by extrapolating from inclusive semileptonic D decays ($\lesssim 2 - 3\%$ [13]). Here we note that the extrapolated estimates are subject to an inherent uncertainty in the extrapolation due to the limited heaviness of the charm. If the true rate is at the upper bound set by the extrapolation from charm decays, the measured upper limits in the two channels are expected to be at, or improve upon, the limit set by CLEO (7.4% [96, 97]) for the combined sample. Additionally, analysis of the fit quality with different weak annihilation models is expected to be able to provide some insight into the shape of the weak annihilation contribution.

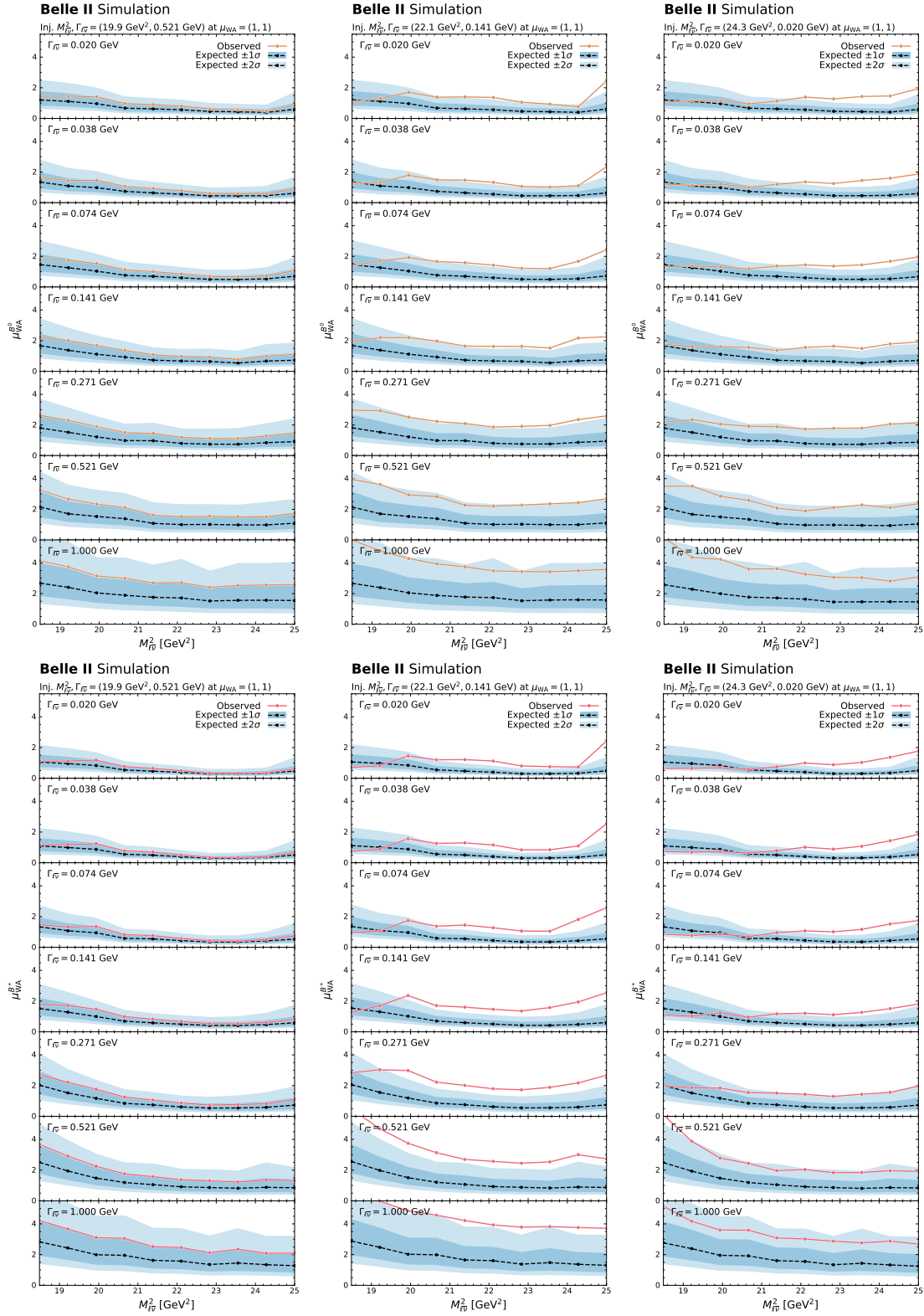


Figure 12.7: Observed and expected upper limits on $\mu_{WA}^{B^0}$ (top row), and $\mu_{WA}^{B^+}$ (bottom row) recovered on simulation, when injecting three example weak annihilation signals. The parameters of the injected signals are provided at the top of each plot.

Injected WA Model		Injected ($\bar{\mu}_{\text{WA}}^{B^0}, \bar{\mu}_{\text{WA}}^{B^+}$)					
$M_{\ell\bar{\nu}}^2$ [GeV ²]	$\Gamma_{\ell\bar{\nu}}$ [GeV]	(0, 0)	(1, 1)	(2, 2)	(3, 3)	(3, 0)	(0, 3)
19.94	0.02	1.50	2.91	4.86	6.84	6.20	1.84
19.94	0.14	1.50	2.66	4.20	5.81	5.53	1.67
19.94	0.52	1.50	2.34	3.36	4.44	4.39	1.58
22.11	0.02	1.50	3.05	4.45	6.14	6.55	1.24
22.11	0.14	1.50	2.94	4.27	5.77	6.15	1.29
22.11	0.52	1.50	2.72	3.98	5.12	5.34	1.52
24.28	0.02	1.50	2.85	4.31	5.57	5.67	1.61
24.28	0.14	1.50	2.74	3.91	4.89	5.52	1.31
24.28	0.52	1.50	2.62	3.72	4.70	5.09	1.41

Table 12.2: Weakest 90% upper limit recovered on $\mu_{\text{WA}}^{B^0}$ amongst models in the fiducial model parameter space when injecting various weak annihilation signals with various models at a range of signal strengths.

Injected WA Model		Injected ($\bar{\mu}_{\text{WA}}^{B^0}, \bar{\mu}_{\text{WA}}^{B^+}$)					
$M_{\ell\bar{\nu}}^2$ [GeV ²]	$\Gamma_{\ell\bar{\nu}}$ [GeV]	(0, 0)	(1, 1)	(2, 2)	(3, 3)	(3, 0)	(0, 3)
19.94	0.02	1.42	2.99	5.35	7.78	1.60	7.77
19.94	0.14	1.42	2.73	4.45	6.53	1.56	6.57
19.94	0.52	1.42	2.25	3.26	4.32	1.45	4.30
22.11	0.02	1.42	4.19	6.77	9.05	2.27	7.95
22.11	0.14	1.42	3.75	6.05	8.13	2.03	7.24
22.11	0.52	1.42	2.77	4.29	5.72	1.66	5.35
24.28	0.02	1.42	2.79	4.21	5.34	1.55	5.06
24.28	0.14	1.42	3.30	5.17	6.77	1.87	6.24
24.28	0.52	1.42	2.74	4.23	5.57	1.66	5.19

Table 12.3: Weakest 90% upper limit recovered on $\mu_{\text{WA}}^{B^+}$ amongst models in the fiducial model parameter space when injecting various weak annihilation signals with various models at a range of signal strengths.

12.5 Outlook

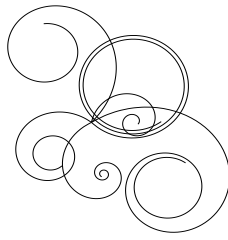
Before performing the analysis in the signal region, several steps are required.

- The analysis must switch to run-dependent MC, this is expected to improve the agreement between simulation and data.
- As discussed in Sec. 11.4, a procedure to correct the missing mass spectrum or assign a systematic uncertainty must be developed.
- The analysis must undergo an internal review within the collaboration.
- The look-elsewhere effect [284, 285] must be considered in case a significant deviation from the background-only hypothesis is seen.

None of these steps are expected to lead to significant changes in the analysis.

12.6 Disclaimer

Between submission of this thesis and the preparation of this final version, a bug was discovered within the Belle II software by the author. This bug affects the interaction between the Full Event Interpretation and the specific method used for signal sample generation in this analysis, leading to a 5%–10% lower reconstruction efficiency for simulated weak annihilation events. Preliminary tests with fixed samples, but using the selections optimised on the bugged samples, indicate that the impact on the analysis is minor. The weakest limits recovered in the fiducial space for $\mu_{\text{WA}}^{B^0}$ when injecting various weak annihilation signals, change by between -0.19 – 0.23 , -0.21 – 0.59 , and -0.18 – 0.68 for injected $(\bar{\mu}_{\text{WA}}^{B^0}, \bar{\mu}_{\text{WA}}^{B^+})$ of $(1, 1)$, $(2, 2)$, and $(3, 3)$, respectively, when compared with those reported in Table 12.2. For the analogous limits reported for $\mu_{\text{WA}}^{B^0}$, the limits improve by between $-(0.39$ – $0.05)$, $-(0.85$ – $0.04)$, and $-(1.18$ – $0.02)$, respectively, compared to those reported in Table 12.3.



Part IV

Conclusion

Conclusion

Precision studies in the flavour sector are one of the most intriguing avenues in the search for beyond Standard Model physics. This thesis presents two analyses related to the determination of the magnitude of the Cabbibo-Kobayashi-Maskawa matrix element $|V_{ub}|$. Precise knowledge of this magnitude is necessary to anchor the Standard Model determination of the position of the Unitarity Triangle apex, and thus improve the sensitivity of new physics searches. Two experimentally and theoretically complementary approaches to the measurement of $|V_{ub}|$ exist: exclusive, reconstructing a specific hadronic final state; or inclusive, summing over all hadronic final states. At present, the two approaches agree at the 1.4σ level. Both of the presented analyses are related to the inclusive determination.

The first analysis presented reports the first measurement of the ratio of partial branching fractions of the inclusive charmless to the inclusive charmed semileptonic B decay. The measurement is performed at the Belle experiment and uses the full dataset of 711 fb^{-1} collected at the $\Upsilon(4S)$ resonance. State-of-the-art software developed for the Belle II experiment is used. This includes the Full Event Interpretation tagging algorithm [190], which uses machine learning to reconstruct one of the B mesons from the $\Upsilon(4S)$ decay in thousands of fully hadronic channels. Using the reconstructed tag B , the kinematic properties of the signal B can be inferred through conservation laws, without having to explicitly reconstruct it. This allows for the inclusive reconstruction of the signal semileptonic B decay. The primary difficulty of the measurement lies in the treatment of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events in the extraction of the $\bar{B} \rightarrow X_u \ell \bar{\nu}$ yield. Such decays are only partially understood and dominate $\bar{B} \rightarrow X_u \ell \bar{\nu}$. A data-driven description of this component, extracted from a $\bar{B} \rightarrow X_c \ell \bar{\nu}$ enriched side band, is used. In the phase-space region $E_\ell^{B_{\text{sig}}} > 1.0\text{ GeV}$, we measure:

$$\frac{\Delta\mathcal{B}(\bar{B} \rightarrow X_u \ell \bar{\nu})}{\Delta\mathcal{B}(\bar{B} \rightarrow X_c \ell \bar{\nu})} = (1.99 \pm 0.17_{\text{stat}} \pm 0.16_{\text{syst}}) \times 10^{-2},$$

where the uncertainties are statistical and systematic, respectively. From the partial branching fraction ratio we extract $|V_{ub}|/|V_{cb}|$ using four theoretical calculations for the partial decay rate of $\bar{B} \rightarrow X_u \ell \bar{\nu}$, finding

$$\begin{aligned} \frac{|V_{ub}|}{|V_{cb}|}^{\text{BLNP}} &= (9.81 \pm 0.42_{\text{stat}} \pm 0.38_{\text{syst}} \pm 0.51_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2}, \\ \frac{|V_{ub}|}{|V_{cb}|}^{\text{GGOU}} &= (10.1 \pm 0.43_{\text{stat}} \pm 0.39_{\text{syst}} \pm 0.23_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2}, \\ \frac{|V_{ub}|}{|V_{cb}|}^{\text{DGE}} &= (10.1 \pm 0.43_{\text{stat}} \pm 0.40_{\text{syst}} \pm 0.31_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2}, \end{aligned}$$

$$\frac{|V_{ub}|}{|V_{cb}|}^{\text{ADFR}} = (9.81 \pm 0.42_{\text{stat}} \pm 0.38_{\text{syst}} \pm 0.41_{\Delta\Gamma(\bar{B} \rightarrow X_u \ell \bar{\nu})} \pm 0.20_{\Delta\Gamma(\bar{B} \rightarrow X_c \ell \bar{\nu})}) \times 10^{-2},$$

where the third and fourth uncertainties are from the partial decay rates of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ and $\bar{B} \rightarrow X_c \ell \bar{\nu}$, respectively. These values are in excellent agreement with inclusive and exclusive world averages of $|V_{ub}|/|V_{cb}|$ [65].

We additionally present this measurement broken down by B charge, determining a limit on isospin breaking in $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays, as well as by lepton flavour, finding no evidence for lepton flavour universality violation in either the charmed or charmless inclusive semileptonic B decay. Furthermore, we report the differential ratio of partial branching fractions in q^2 and $E_\ell^{B_{\text{sig}}}$.

The second analysis presented is a search for the weak annihilation contribution to the inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decay rate. In the context of inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$, weak annihilation describes the annihilation of the b quark with either a valent or nonvalent u quark, creating a lepton neutrino pair, and a hadronic system through gluon emission. Weak annihilation enters the theoretical description of inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ at next-to-next-to-next-to-leading order and is not accounted for in any of the existing theoretical descriptions of $\bar{B} \rightarrow X_u \ell \bar{\nu}$. The scale of this contribution is thus a systematic uncertainty on inclusive $|V_{ub}|$ determinations. At present, this uncertainty is subleading, but it must be understood if inclusive $|V_{ub}|$ determinations are to reach percent level precision.

The analysis is performed at the Belle II experiment, using the dataset collected at the $\Upsilon(4S)$ resonance before 2023, corresponding to 364 fb^{-1} . As little is known about the shape of the weak annihilation contribution, we search for structure at high q^2 , where it is expected to be localised, using a broad range of hypothesised models. The analysis must still undergo internal review; we therefore investigate the sensitivity of the analysis to simulated signal samples. Depending on the shape and magnitude of the simulated weak annihilation sample we expect to place 90% upper limits between 1.5–6.8% for the weak annihilation contribution to neutral B decays, and between 1.4–9.1% for charged B decays. This will be the first direct search for the weak annihilation contribution to separately constrain the magnitude of the contribution to the charged and neutral B systems.

In addition to the physics analyses, we present improvements to the identification of leptons at Belle II by exploiting calorimeter information through machine learning. The method is validated in data and demonstrates reductions in pion-to-electron and pion-to-muon misidentification rates at low momentum of up to factors 10 and 1.6, respectively. This will benefit all future studies of semileptonic decays at the experiment, especially studies reliant on the identification of low-momentum leptons such as measurements of $R(D^{(*)})$.

References

- [1] R. Workman, V. Burkert, V. Crede et al. (Particle Data Group), *Review of Particle Physics*, PTEP **2022**, 083C01 (2022).
- [2] A. Dighe et al., *Large mass splittings for fourth generation fermions allowed by LHC Higgs exclusion*, Phys. Rev. D **85**, 114035 (2012).
- [3] L. M. Carpenter and A. Rajaraman, *Revisiting Constraints on Fourth Generation Neutrino Masses*, Phys. Rev. D **82**, 114019 (2010).
- [4] T. S.-K. Collaboration and Y. F. et al (Super-Kamiokande), *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81**, 1562–1567 (1998).
- [5] Z. Maki, M. Nakagawa and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, Prog. Theor. Phys. **28**, 870–880 (1962).
- [6] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10**, 531–533 (1963).
- [7] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49**, 652–657 (1973).
- [8] L.-L. Chau and W.-Y. Keung, *Comments on the Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **53**, 1802 (1984).
- [9] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51**, 1945 (1983).
- [10] J. Charles et al. (CKMfitter Group), *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. C **41**, updated results and plots available at: <http://ckmfitter.in2p3.fr>, 1–131 (2005).
- [11] M. Bona, M. Ciuchini, E. Franco et al. (UTfit), *The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of Δm_s and $\mathcal{BR}(B \rightarrow \tau \nu_\tau)$* , JHEP **10**, 081 (2006).
- [12] J. D. Richman and P. R. Burchat, *Leptonic and semileptonic decays of charm and bottom hadrons*, Rev. Mod. Phys. **67**, 893–976 (1995).
- [13] J. Dingfelder and T. Mannel, *Leptonic and semileptonic decays of B mesons*, Rev. Mod. Phys. **88**, 035008 (2016).
- [14] C. Bourrely, I. Caprini and L. Lellouch, *Model-independent description of $B \rightarrow \pi \ell \nu$ decays and a determination of $|V_{ub}|$* , Phys. Rev. D **79**, [Erratum: Phys.Rev.D 82, 099902 (2010)], 013008 (2009).

-
- [15] C. Boyd, B. Grinstein and R. F. Lebed, *Constraints on form-factors for exclusive semileptonic heavy to light meson decays*, Phys. Rev. Lett. **74**, 4603–4606 (1995).
 - [16] N. Isgur and M. B. Wise, *Weak transition form factors between heavy mesons*, Phys. Lett. B **237**, 527–530 (1990).
 - [17] A. K. Leibovich et al., *Predictions for $B \rightarrow D_1(2420)\ell\bar{\nu}$ and $B \rightarrow D_2^*(2460)\ell\bar{\nu}$ at order $\Lambda_{QCD}/m_{c,b}$* , Phys. Rev. Lett. **78**, 3995–3998 (1997).
 - [18] I. Bigi and N. Uraltsev, *A Vademecum on Quark-Hadron Duality*, Int. J. Mod. Phys. A **16**, 5201–5248 (2001).
 - [19] J. Chay, H. Georgi and B. Grinstein, *Lepton energy distributions in heavy meson decays from QCD*, Phys. Lett. B **247**, 399–405 (1990).
 - [20] I. I. Bigi et al., *QCD predictions for lepton spectra in inclusive heavy flavor decays*, Phys. Rev. Lett. **71**, 496–499 (1993).
 - [21] T. Mannel, *Operator product expansion for inclusive semileptonic decays in heavy quark effective field theory*, Nucl. Phys. B **413**, 396–412 (1994).
 - [22] M. Fael, T. Mannel and K. K. Vos, *V_{cb} determination from inclusive $b \rightarrow c$ decays: an alternative method*, JHEP **02**, 177 (2019).
 - [23] P. Gambino and C. Schwanda, *Inclusive semileptonic fits, heavy quark masses, and V_{cb}* , Phys. Rev. D **89**, 014022 (2014).
 - [24] M. Brucherseifer, F. Caola and K. Melnikov, *On the $O(\alpha_s^2)$ corrections to $b \rightarrow X_u e \bar{\nu}$ inclusive decays*, Phys. Lett. B **721**, 107–110 (2013).
 - [25] A. Bornheim, E. D. Lipeles, S. Pappas et al. (CLEO), *Improved measurement of $|V_{ub}|$ with inclusive semileptonic B decays*, Phys. Rev. Lett. **88**, 231803 (2002).
 - [26] B. Aubert (BaBar), *Measurement of the Inclusive Charmless Semileptonic Branching Ratio of B Mesons and Determination of $|V_{ub}|$* , Phys. Rev. Lett. **92**, 071802 (2004).
 - [27] B. Aubert, R. Barate, D. Boutigny et al. (BaBar), *Measurement of the Inclusive Electron Spectrum in Charmless Semileptonic B Decays Near the Kinematic Endpoint and Determination of $|V_{ub}|$* , Phys. Rev. D **73**, 012006 (2006).
 - [28] A. Limosani, K. Abe, I. Adachi et al. (Belle), *Measurement of inclusive charmless semileptonic B -meson decays at the endpoint of the electron momentum spectrum*, Phys. Lett. B **621**, 28–40 (2005).
 - [29] I. Bizjak, K. Abe, H. Aihara et al. (Belle), *Determination of $|V_{ub}|$ from measurements of the inclusive charmless semileptonic partial rates of B mesons using full reconstruction tags*, Phys. Rev. Lett. **95**, 241801 (2005).
 - [30] M. Neubert, *QCD-based interpretation of the lepton spectrum in inclusive $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays*, Phys. Rev. D **49**, 3392–3398 (1994).

- [31] I. I. Bigi et al., *On the Motion of Heavy Quarks inside Hadrons: Universal Distributions and Inclusive Decays*, Int. J. Mod. Phys. A **9**, 2467–2504 (1994).
- [32] I. Bigi and N. Uraltsev, *On the expected photon spectrum in $B \rightarrow X_s + \gamma$ and its uses*, Int. J. Mod. Phys. A **17**, 4709–4732 (2002).
- [33] F. U. Bernlochner et al., *Status of SIMBA*, arXiv:1101.3310 [hep-ph] (2011).
- [34] F. U. Bernlochner et al. (SIMBA), *Precision Global Determination of the $B \rightarrow X_s \gamma$ Decay Rate*, Phys. Rev. Lett. **127**, 102001 (2021).
- [35] P. Gambino, K. J. Healey and C. Mondino, *Neural network approach to $B \rightarrow X_u \ell \nu$* , Phys. Rev. D **94**, 014031 (2016).
- [36] E. Gardi, *On the determination of $-V_{ub}$ — from inclusive semileptonic B decays*, Frascati Phys. Ser. **47**, 381–405 (2008).
- [37] F. De Fazio and M. Neubert, *$B \rightarrow X_u \ell \bar{\nu}_\ell$ decay distributions to order α_s* , JHEP **06**, 017 (1999).
- [38] S. W. Bosch et al., *Factorization and Shape-Function Effects in Inclusive B -Meson Decays*, Nucl. Phys. B **699**, 335–386 (2004).
- [39] B. O. Lange, M. Neubert and G. Paz, *Theory of charmless inclusive B decays and the extraction of V_{ub}* , Phys. Rev. D **72**, 073006 (2005).
- [40] C. Greub, M. Neubert and B. D. Pecjak, *NNLO corrections to $B \rightarrow X_u \ell \nu$ and the determination of $-V_{ub}$* , Eur. Phys. J. C **65**, 501–515 (2010).
- [41] P. Gambino, A. Kronfeld, M. Rotondo et al., *Challenges in semileptonic B decays*, Eur. Phys. J. C **80**, 966 (2020).
- [42] P. Gambino et al., *Inclusive semileptonic B decays and the determination of $|V_{ub}|$* , JHEP **10**, 058 (2007).
- [43] I. Bigi et al., *Heavy Quark Distribution Function in QCD and the AC^2M^2 Model*, Phys. Lett. B **328**, 431–440 (1994).
- [44] J. R. Andersen and E. Gardi, *Inclusive spectra in charmless semileptonic B decays by dressed gluon exponentiation*, JHEP **01**, 097 (2006).
- [45] U. Aglietti et al., *Inclusive measure of $|V_{ub}|$ with the analytic coupling model*, Eur. Phys. J. C **59**, 831–840 (2009).
- [46] P. Gambino, *Status and prospects of inclusive $b \rightarrow u \ell \nu$ theory*, Istituto Universitario di Studi Superiori - IUSS 1391, Ferrara, Mar. 2009.
- [47] M. Antonelli et al., *Flavor Physics in the Quark Sector*, Phys. Rept. **494**, 197–414 (2010).
- [48] R. Glattauer, C. Schwanda, A. Abdesselam et al. (Belle), *Measurement of the decay $B \rightarrow D \ell \nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$* , Phys. Rev. D **93**, 032006 (2016).

-
- [49] B. Aubert, M. Bona, Y. Karyotakis et al. (BaBar), *Measurements of the Semileptonic Decays $\bar{B} \rightarrow D\ell\bar{\nu}$ and $\bar{B} \rightarrow D^*\ell\bar{\nu}$ Using a Global Fit to $DX\ell\bar{\nu}$ Final States*, Phys. Rev. D **79**, 012002 (2009).
 - [50] B. Aubert, Y. Karyotakis, J. Lees et al. (BaBar), *Measurement of $|V_{cb}|$ and the Form-Factor Slope in $\bar{B} \rightarrow D\ell^-\bar{\nu}$ Decays in Events Tagged by a Fully Reconstructed B Meson*, Phys. Rev. Lett. **104**, 011802 (2010).
 - [51] B. Aubert, M. Bona, D. Boutigny et al. (BaBar), *Determination of the form-factors for the decay $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ and of the CKM matrix element $|V_{cb}|$* , Phys. Rev. D **77**, 032002 (2008).
 - [52] B. Aubert, M. Bona, Y. Karyotakis et al. (BaBar), *Measurement of the Decay $B^- \rightarrow D^{*0}e^-\bar{\nu}_e$* , Phys. Rev. Lett. **100**, 231803 (2008).
 - [53] W. Dungel, C. Schwanda, I. Adachi et al. (Belle), *Measurement of the form factors of the decay $B^0 \rightarrow D^{*-}\ell^+\nu$ and determination of the CKM matrix element $|V_{cb}|$* , Phys. Rev. D **82**, 112007 (2010).
 - [54] A. Abdesselam, I. Adachi, K. Adamczyk et al. (Belle), *Precise determination of the CKM matrix element $|V_{cb}|$ with $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle*, arXiv:1702.01521 [hep-ex] (2017).
 - [55] E. Waheed, P. Urquijo, D. Ferlewicz et al. (Belle), *Measurement of the CKM matrix element $|V_{cb}|$ from $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ at Belle*, Phys. Rev. D **100**, [Erratum: Phys.Rev.D 103, 079901 (2021)], 052007 (2019).
 - [56] I. Adachi, K. Adamczyk, L. Aggarwal et al. (Belle-II), *Determination of $|V_{cb}|$ using $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decays with Belle II*, Phys. Rev. D **108**, 092013 (2023).
 - [57] D. Ferlewicz, P. Urquijo and E. Waheed, *Revisiting fits to $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ to measure $|V_{cb}|$ with novel methods and preliminary LQCD data at nonzero recoil*, Phys. Rev. D **103**, 073005 (2021).
 - [58] F. Bernlochner et al., *First extraction of inclusive V_{cb} from q^2 moments*, JHEP **10**, 068 (2022).
 - [59] P. Urquijo, E. Barberio, K. Abe et al. (Belle), *Moments of the electron energy spectrum and partial branching fraction of $B \rightarrow X_c e \nu$ decays at Belle*, Phys. Rev. D **75**, 032001 (2007).
 - [60] C. Schwanda, K. Abe, I. Adachi et al. (Belle), *Moments of the Hadronic Invariant Mass Spectrum in $B \rightarrow X_c \ell \nu$ Decays at Belle*, Phys. Rev. D **75**, 032005 (2007).
 - [61] B. Aubert, R. Barate, D. Boutigny et al. (BaBar), *Measurement of the electron energy spectrum and its moments in inclusive $B \rightarrow X e \nu$ decays*, Phys. Rev. D **69**, 111104 (2004).
 - [62] B. Aubert, Y. Karyotakis, J. Lees et al. (BaBar), *Measurement and interpretation of moments in inclusive semileptonic decays $\bar{B} \rightarrow X_c \ell^-\bar{\nu}$* , Phys. Rev. D **81**, 032003 (2010).

- [63] R. van Tonder, L. Cao, W. Sutcliffe et al. (Belle), *Measurements of q^2 Moments of Inclusive $B \rightarrow X_c \ell^+ \nu_\ell$ Decays with Hadronic Tagging*, Phys. Rev. D **104**, 112011 (2021).
- [64] F. Abudinén, K. Adamczyk, L. Aggarwal et al. (Belle-II), *Measurement of Lepton Mass Squared Moments in $B \rightarrow X_c \ell \bar{\nu}_\ell$ Decays with the Belle II Experiment*, Phys. Rev. D **107**, 072002 (2023).
- [65] R. Workman, V. Burkert, V. Crede et al. (Particle Data Group), *Review of Particle Physics*, PTEP **2022**, and 2023 update, 083C01 (2022).
- [66] Y. S. Amhis, S. Banerjee, E. Ben-Haim et al. (HFLAV), *Averages of b -hadron, c -hadron, and τ -lepton properties as of 2021*, Phys. Rev. D **107**, 052008 (2023).
- [67] J. Lees, V. Poireau, V. Tisserand et al. (BaBar), *Branching fraction and form-factor shape measurements of exclusive charmless semileptonic B decays, and determination of $|V_{ub}|$* , Phys. Rev. D **86**, 092004 (2012).
- [68] A. Sibidanov, K. Varvell, I. Adachi et al. (Belle), *Study of Exclusive $B \rightarrow X_u \ell \nu$ Decays and Extraction of $|V_{ub}|$ using Full Reconstruction Tagging at the Belle Experiment*, Phys. Rev. D **88**, 032005 (2013).
- [69] H. Ha, E. Won, I. Adachi et al. (Belle), *Measurement of the decay $B^0 \rightarrow \pi^- \ell^+ \nu$ and determination of $|V_{ub}|$* , Phys. Rev. D **83**, 071101 (2011).
- [70] P. del Amo Sanchez, J. Lees, V. Poireau et al. (BaBar), *Study of $B \rightarrow \pi \ell \nu$ and $B \rightarrow \rho \ell \nu$ Decays and Determination of $|V_{ub}|$* , Phys. Rev. D **83**, 032007 (2011).
- [71] B. Aubert, R. Barate, D. Boutigny et al. (BaBar), *Determination of $|V_{ub}|$ from measurements of the electron and neutrino momenta in inclusive semileptonic B decays*, Phys. Rev. Lett. **95**, [Erratum: Phys.Rev.Lett. 97, 019903 (2006)], 111801 (2005).
- [72] J. Lees, V. Poireau, V. Tisserand et al. (BaBar), *Study of $\bar{B} \rightarrow X_u \ell \bar{\nu}$ decays in $B\bar{B}$ events tagged by a fully reconstructed B -meson decay and determination of $|V_{ub}|$* , Phys. Rev. D **86**, 032004 (2012).
- [73] L. Cao, W. Sutcliffe, R. Van Tonder et al. (Belle), *Measurements of Partial Branching Fractions of Inclusive $B \rightarrow X_u \ell^+ \nu_\ell$ Decays with Hadronic Tagging*, Phys. Rev. D **104**, 012008 (2021).
- [74] J. Lees, V. Poireau, V. Tisserand et al. (BaBar), *Measurement of the inclusive electron spectrum from B meson decays and determination of $|V_{ub}|$* , Phys. Rev. D **95**, 072001 (2017).
- [75] L. Cao, F. Bernlochner, K. Tackmann et al. (Belle), *First Simultaneous Determination of Inclusive and Exclusive $|V_{ub}|$* , Phys. Rev. Lett. **131**, 211801 (2023).
- [76] R. Aaij, B. Adeva, M. Adinolfi et al. (LHCb), *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Nature Phys. **11**, 743–747 (2015).

-
- [77] R. Aaij, C. A. Beteta, T. Ackernley et al. (LHCb), *First observation of the decay $B_s^0 \rightarrow K^- \mu^+ \nu_\mu$ and Measurement of $|V_{ub}|/|V_{cb}|$* , Phys. Rev. Lett. **126**, 081804 (2021).
 - [78] Y. Aoki, T. Blum, G. Colangelo et al. (Flavour Lattice Averaging Group (FLAG)), *FLAG Review 2021*, Eur. Phys. J. C **82**, 869 (2022).
 - [79] M. Prim, F. Bernlochner, P. Goldenzweig et al. (Belle), *Search for $B^+ \rightarrow \mu^+ \nu_\mu$ and $B^+ \rightarrow \mu^+ N$ with inclusive tagging*, Phys. Rev. D **101**, 032007 (2020).
 - [80] I. Adachi, K. Adamczyk, H. Aihara et al. (Belle), *Evidence for $B^- \rightarrow \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle*, Phys. Rev. Lett. **110**, 131801 (2013).
 - [81] B. Kronenbitter, M. Heck, P. Goldenzweig et al. (Belle), *Measurement of the branching fraction of $B^+ \rightarrow \tau^+ \nu_\tau$ decays with the semileptonic tagging method*, Phys. Rev. D **92**, 051102 (2015).
 - [82] B. Aubert, Y. Karyotakis, J. Lees et al. (BaBar), *A Search for $B^+ \rightarrow \ell^+ \nu_\ell$ Recoiling Against $B^- \rightarrow D^0 \ell^- \bar{\nu} X$* , Phys. Rev. D **81**, 051101 (2010).
 - [83] J. Lees, V. Poireau, V. Tisserand et al. (BaBar), *Evidence of $B^+ \rightarrow \tau^+ \nu$ decays with hadronic B tags*, Phys. Rev. D **88**, 031102 (2013).
 - [84] W. Altmannshofer et al. (Belle-II), *The Belle II Physics Book*, PTEP **2019**, edited by E. Kou and P. Urquijo, [Erratum: PTEP 2020, 029201 (2020)], 123C01 (2019).
 - [85] L. Aggarwal et al. (Belle-II), *Snowmass White Paper: Belle II physics reach and plans for the next decade and beyond*, arXiv:2207.06307 [hep-ex] (2022).
 - [86] D. Pirjol and N. Uraltsev, *Four fermion heavy quark operators and light current amplitudes in heavy flavor hadrons*, Phys. Rev. D **59**, 034012 (1999).
 - [87] I. I. Bigi and N. Uraltsev, *Gluonic enhancements in non-spectator beauty decays: An Inclusive mirage though an exclusive possibility*, Phys. Lett. B **280**, 271–280 (1992).
 - [88] I. I. Bigi and N. Uraltsev, *Weak annihilation and the endpoint spectrum in semileptonic B decays*, Nucl. Phys. B **423**, 33–55 (1994).
 - [89] M. Voloshin, *Nonfactorizable terms, heavy quark masses, and semileptonic decays of D and B mesons*, Mod. Phys. Lett. A **17**, 245–251 (2002).
 - [90] A. K. Leibovich, Z. Ligeti and M. B. Wise, *Enhanced Subleading Structure Functions in Semileptonic B Decay*, Phys. Lett. B **539**, 242–248 (2002).
 - [91] N. Uraltsev, *Theoretical uncertainties in $\Gamma_{(sl)}(b \rightarrow u)$* , Int. J. Mod. Phys. A **14**, 4641–4652 (1999).
 - [92] Z. Ligeti, M. Luke and A. V. Manohar, *Constraining weak annihilation using semileptonic D decays*, Phys. Rev. D **82**, 033003 (2010).

- [93] I. Bigi et al., *The Two Roads to ‘Intrinsic Charm’ in B Decays*, JHEP **04**, 073 (2010).
- [94] A. Lenz and T. Rauh, *D-meson lifetimes within the heavy quark expansion*, Phys. Rev. D **88**, 034004 (2013).
- [95] P. Gambino, G. Ossola and N. Uraltsev, *Hadronic mass and q^2 moments of charmless semileptonic B decay distributions*, JHEP **09**, 010 (2005).
- [96] Thomas Oliver Meyer, *Limits on Weak Annihilation In Inclusive Charmless Semileptonic B Decays*, PhD thesis (Cornell, Jan. 2005).
- [97] J. Rosner, N. Adam, J. Alexander et al. (CLEO), *Experimental limits on weak annihilation contributions to $b \rightarrow ul\nu$ decay*, Phys. Rev. Lett. **96**, 121801 (2006).
- [98] P. Gambino and J. F. Kamenik, *Lepton energy moments in semileptonic charm decays*, Nucl. Phys. B **840**, 424–437 (2010).
- [99] M. Ablikim, M. N. Achasov, P. A. Adlarson et al. (BESIII), *Measurement of the absolute branching fraction of inclusive semielectronic D_s^+ decays*, Phys. Rev. D **104**, 012003 (2021).
- [100] D. Asner, K. Edwards, J. Reed et al. (CLEO), *Measurement of absolute branching fractions of inclusive semileptonic decays of charm and charmed-strange mesons*, Phys. Rev. D **81**, 052007 (2010).
- [101] F. U. Bernlochner et al., *Constraints on exclusive branching fractions $\mathcal{B}_i(B^+ \rightarrow X_c^i l^+ \nu)$ from moment measurements in inclusive $B \rightarrow X_c l \nu$ decays*, Eur. Phys. J. C **74**, 2914 (2014).
- [102] S. Herb, D. Hom, L. Lederman et al. (E288), *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39**, 252–255 (1977).
- [103] W. R. Innes, J. Appel, B. Brown et al. (E288), *Observation of Structure in the Υ Region*, Phys. Rev. Lett. **39**, [Erratum: Phys.Rev.Lett. 39, 1640 (1977)], 1240–1242 (1977).
- [104] C. Darden, H. Hasemann, A. Krolzig et al. (DASP), *Evidence for a Narrow Resonance at 10.01-GeV in electron-positron Annihilations*, Phys. Lett. B **78**, 364–365 (1978).
- [105] C. Darden, H. Hasemann, A. Krolzig et al. (DASP), *Observation of a Narrow Resonance at 9.46 GeV in electron-positron Annihilations*, Phys. Lett. B **76**, 246 (1978).
- [106] C. Berger, W. Lackas, F. Raupach et al. (Pluto), *Observation of a Narrow Resonance Formed in e^+e^- Annihilation at 9.46 GeV*, Phys. Lett. B **76**, 243–245 (1978).
- [107] D. Andrews, K. Berkelman, R. Cabenda et al. (CLEO), *Observation of a Fourth Upsilon State in e^+e^- Annihilations*, Phys. Rev. Lett. **45**, 219 (1980).
- [108] S. Behrends, K. Chadwick, J. Chauveau et al. (CLEO), *Observation of Exclusive Decay Modes of B Flavored Mesons*, Phys. Rev. Lett. **50**, 881–884 (1983).

-
- [109] M. Pepe Altarelli and F. Teubert, *B physics at LHCb*, Int. J. Mod. Phys. A **23**, 5117–5136 (2008).
 - [110] T. Abe, K. Akai, N. Akasaka et al., *Achievements of KEKB*, PTEP **2013**, 03A001 (2013).
 - [111] A. Bevan, B. Golob, T. Mannel et al. (BaBar, Belle), *The Physics of the B Factories*, Eur. Phys. J. C **74**, 3026 (2014).
 - [112] Y. Funakoshi, T. Abe, K. Akai et al., *The SuperKEKB Has Broken the World Record of the Luminosity*, in JACoW, Vol. IPAC2022 (July 2022), MOPLXGD1.
 - [113] Wikimedia Commons (Krib), *A schematic of the KEKB accelerator complex*, <https://commons.wikimedia.org/w/index.php?curid=27905978> (visited on 27/11/2023).
 - [114] K. Akai, K. Furukawa and H. Koiso (SuperKEKB), *SuperKEKB Collider*, Nucl. Instrum. Meth. A **907**, 188–199 (2018).
 - [115] P. Raimondi, *Status of the SuperB Effort*, 2nd Workshop on Super \$B\$ Factory, INFN-LFN, Frascati, Mar. 2006.
 - [116] M. Bona, J. Garra Tico, E. Grauges Pous et al. (SuperB), *SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report*, arXiv:0709.0451 [hep-ex] (2007).
 - [117] A. Abashian, K. Gotow, N. Morgan et al. (Belle), *The Belle Detector*, Nucl. Instrum. Meth. A, Detectors for Asymmetric B-factories **479**, 117–232 (2002).
 - [118] Christian Pulvermacher, *dE/dx particle Identification and Pixel Detector Data Reduction for the Belle II Experiment*, Diplomarbeit (Karlsruhe Institute of Technology (KIT), Karlsruhe, June 2012).
 - [119] T. Abe, I. Adachi, K. Adamczyk et al. (Belle-II), *Belle II Technical Design Report*, arXiv:1011.0352 [physics.ins-det], 10.48550/arXiv.1011.0352 (2010).
 - [120] H. Aihara et al., *Belle SVD2 vertex detector*, Nucl. Instrum. Meth. A, New Developments in Radiation Detectors **568**, 269–273 (2006).
 - [121] F. Forti (Belle-II SVD), *The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector*, JINST **17**, P11042 (2022).
 - [122] G. Rizzo et al., *The Belle II Silicon Vertex Detector: Performance and Operational Experience in the First Year of Data Taking*, in Proceedings of the 29th International Workshop on Vertex Detectors (VERTEX2020), Vol. 34, JPS Conference Proceedings (Journal of the Physical Society of Japan, June 2021).
 - [123] Y. Ushiroda, *Belle silicon vertex detectors*, Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip. **511**, 6–10 (2003).
 - [124] INFN (Belle II Italian Collaboration), *Silicon Vertex Detector*, <https://web.infn.it/Belle-II/index.php/detector/svd> (visited on 03/08/2023).

- [125] F. Müller, *Characterization and optimization of the prototype DEPFET modules for the Belle II Pixel Vertex Detector*, PhD thesis (Ludwig-Maximilians-Universität München, July 2017).
- [126] M. Srebre et al., *Generation of Belle II Pixel Detector Background Data with a GAN*, EPJ Web Conf. **245**, 02010 (2020).
- [127] H. Ye, F. Abudinen, K. Ackermann et al. (Belle-II DEPFET, PXD), *Commissioning and performance of the Belle II pixel detector*, Nucl. Instrum. Meth. A **987**, 164875 (2021).
- [128] V. Bertacchi, T. Bilka, N. Braun et al. (Belle II Tracking Group), *Track finding at Belle II*, Comput. Phys. Commun. **259**, 107610 (2021).
- [129] M. Neu et al., *Real-time Graph Building on FPGAs for Machine Learning Trigger Applications in Particle Physics*, arXiv:2307.07289 [hep-ex] (2023).
- [130] T. Hauth (Belle II Tracking Group), *Belle II Track Reconstruction and Results from first Collisions*, EPJ Web Conf. **214**, 02032 (2019).
- [131] T. Dong et al., *Calibration and alignment of the Belle II central drift chamber*, Nucl. Instrum. Meth. A **930**, 132–141 (2019).
- [132] W. Waltenberger, W. Mitaroff and F. Moser, *RAVE—a Detector-independent vertex reconstruction toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, VCI 2007 **581**, 549–552 (2007).
- [133] J.-F. Krohn, P. Urquijo, F. Abudinén et al. (Belle-II analysis software Group), *Global decay chain vertex fitting at Belle II*, Nucl. Instrum. Meth. A **976**, 164269 (2020).
- [134] V. Aulchenko, A. Bobrov, A. Bondar et al., *Electromagnetic calorimeter for Belle II*, J. Phys. Conf. Ser. **587**, 012045 (2015).
- [135] I. Adachi et al. (Belle-II), *Detectors for extreme luminosity: Belle II*, Nucl. Instrum. Meth. A, Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn) **907**, 46–59 (2018).
- [136] M. Thomson, *Modern particle physics* (Cambridge University Press, New York, 2013).
- [137] N. Ellis and J. D. March-Russell, *1998 European School on High-Energy Physics: St. Andrews, UK 23 Aug - 5 Sep 1998. European School on High-Energy Physics* (CERN / CERN, Geneva, 1999).
- [138] Michael De Nuccio, *Tutorial: The Belle II Electromagnetic Calorimeter*, Tsukuba, Japan, June 2019.
- [139] B. A. Shwartz (Belle), *Performance and upgrade plans of the BELLE calorimeter*, in 10th Int. Conf. Calorim. High Energy Phys. CALOR 2002 (Mar. 2002), pp. 182–189.

-
- [140] T. Aushev, D. Besson, K. Chilikin et al., *A scintillator based endcap K_L and muon detector for the Belle II experiment*, Nucl. Instrum. Meth. A **789**, 134–142 (2015).
 - [141] Marco Milesi, *Particle identification at Belle II with the Electromagnetic Calorimeter (ECL)*, Tsukuba, Japan, Jan. 2020.
 - [142] H. Kichimi et al., *The BELLE TOF system*, Nucl. Instrum. Meth. A **453**, 315–320 (2000).
 - [143] E. Nakano, *Belle PID*, Nucl. Instrum. Meth. A, Proceedings of the 8th International Conference on Instrumentation for Colliding Beam Physics **494**, 402–408 (2002).
 - [144] T. Iijima et al., *Aerogel Cherenkov counter for the BELLE detector*, Nucl. Instrum. Meth. A **453**, 321–325 (2000).
 - [145] P. Cherenkov, *Visible luminescence of pure liquids under the influence of γ -radiation*, Dokl. Akad. Nauk SSSR **2**, 451–454 (1934).
 - [146] J. Fast (Belle-II Barrel Particle Identification Group), *The Belle II imaging Time-of-Propagation (iTOP) detector*, Nucl. Instrum. Meth. A, The 9th International Workshop on Ring Imaging Cherenkov Detectors (RICH2016) **876**, 145–148 (2017).
 - [147] K. Inami, *MCP-PMT development for Belle-II TOP counter*, Phys. Procedia, Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011) **37**, 683–690 (2012).
 - [148] K. Kojima (Belle II TOP group), *The operation and performance of the TOP detector at the Belle II experiment*, in PoS, Vol. EPS-HEP2021 (SISSA Medialab, Mar. 2022), p. 803.
 - [149] M. Yonenaga, I. Adachi, L. Burmistrov et al., *Performance evaluation of the aerogel RICH counter for the Belle II spectrometer using early beam collision data*, PTEP **2020**, 093H01 (2020).
 - [150] S. Nishida, I. Adachi, N. Hamada et al., *Aerogel RICH for the Belle II forward PID*, Nucl. Instrum. Meth. A, RICH2013 Proceedings of the Eighth International Workshop on Ring Imaging Cherenkov Detectors Shonan, Kanagawa, Japan, December 2-6, 2013 **766**, 28–31 (2014).
 - [151] K. Hanagaki et al., *Electron identification in Belle*, Nucl. Instrum. Meth. A **485**, 490–503 (2002).
 - [152] A. Abashian, K. Abe, K. Abe et al., *Muon identification in the Belle experiment at KEKB*, Nucl. Instrum. Meth. A **491**, 41693 (2002).
 - [153] M. Milesi, J. Tan and P. Urquijo, *Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers*, EPJ Web Conf. **245**, 06023 (2020).
 - [154] Y. Ushiroda et al. (Belle), *Development of the central trigger system for the BELLE detector at the KEK B-factory*, Nucl. Instrum. Meth. A **438**, 460–471 (1999).

- [155] Y. Iwasaki et al., *Level 1 trigger system for the Belle II experiment*, in 2010 17th IEEE-NPSS Real Time Conf. (May 2010), pp. 1–9.
- [156] S. Yamada et al., *Data Acquisition System for the Belle II Experiment*, IEEE Trans. Nucl. Sci. **62**, 1175–1180 (2015).
- [157] T. Kuhr et al. (Belle-II Framework Software Group), *The Belle II Core Software*, Comput. Softw. Big Sci. **3**, 1 (2018).
- [158] The Belle II Collaboration, *Belle II Analysis Software Framework (basf2): release-06-00-00*, Zenodo, Oct. 2021.
- [159] R. Itoh, N. Braun, C. Li et al., *The Performance of Belle II High Level Trigger in the First Physics Run*, EPJ Web Conf. **245**, 01040 (2020).
- [160] S. Jadach, B. Ward and Z. Was, *The Precision Monte Carlo event generator K K for two fermion final states in e^+e^- collisions*, Comput. Phys. Commun. **130**, 260–325 (2000).
- [161] C. Carloni Calame et al., *Large angle Bhabha scattering and luminosity at flavor factories*, Nucl. Phys. B **584**, 459–479 (2000).
- [162] F. A. Berends, P. Daverveldt and R. Kleiss, *Complete lowest-order calculations for four-lepton final states in electron-positron collisions*, Nucl. Phys. B **253**, 441–463 (1985).
- [163] F. A. Berends, P. Daverveldt and R. Kleiss, *Radiative Corrections to the Process $e^+e^- \rightarrow e^+e^-\mu^+\mu^-$* , Nucl. Phys. B **253**, 421 (1985).
- [164] F. A. Berends, P. Daverveldt and R. Kleiss, *Monte Carlo Simulation of Two Photon Processes. 1. Radiative Corrections to Multiperipheral $e^+e^-\mu^+\mu^-$ Production*, Comput. Phys. Commun. **40**, 271–284 (1986).
- [165] R. Itoh, *BASF - BELLE Analysis Framework*, in 9th Int. Conf. Comput. High-Energy Nucl. Phys. (Apr. 1997).
- [166] M. Gelb, T. Keck, M. Prim et al., *B2BII: Data Conversion from Belle to Belle II*, Comput. Softw. Big Sci. **2**, 9 (2018).
- [167] D. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines **462**, 152–155 (2001).
- [168] T. Sjostrand et al., *High-energy physics event generation with PYTHIA 6.1*, Comput. Phys. Commun. **135**, 238–259 (2001).
- [169] C. Bierlich, S. Chakraborty, N. Desai et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, SciPost Phys. **Codebases** **8**, 10 . 21468 / SciPostPhysCodeb.8 (2022).
- [170] S. Jadach, B. F. L. Ward and Z. Was, *Coherent Exclusive Exponentiation For Precision Monte Carlo Calculations*, Phys. Rev. D **63**, 113009 (2001).

-
- [171] E. Barberio, B. van Eijk and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays*, Comput. Phys. Commun. **66**, 115–128 (1991).
- [172] R. Brun et al., *GEANT3*, (1987).
- [173] S. Agostinelli et al. (GEANT4), *Geant4—a simulation toolkit*, Nucl. Instrum. Meth. A **506**, 250–303 (2003).
- [174] J. Allison et al., *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53**, 270 (2006).
- [175] J. Allison et al., *Recent developments in Geant4*, Nucl. Instrum. Meth. A **835**, 186–225 (2016).
- [176] Y. Freund and R. E. Schapire, *Experiments with a new boosting algorithm*, in Proc. Thirteen. Int. Conf. Int. Conf. Mach. Learn. ICML’96 (July 1996), pp. 148–156.
- [177] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55**, 119–139 (1997).
- [178] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, Comput. Softw. Big Sci. **1**, 2 (2017).
- [179] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (Aug. 2016), pp. 785–794.
- [180] J. H. Friedman, *Greedy function approximation: A gradient boosting machine*. Annals Statist. **29**, 1189–1232 (2001).
- [181] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychol. Rev. **65**, 386–408 (1958).
- [182] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, Nature **323**, 533–536 (1986).
- [183] J. Bergstra, D. Yamins and D. D. Cox, *Making a Science of Model Search*, Sept. 2012.
- [184] J. Bergstra et al., *Algorithms for Hyper-Parameter Optimization*, in Adv. Neural Inf. Process. Syst. Vol. 24, edited by J. Shawe-Taylor et al. (2011).
- [185] T. Akiba et al., *Optuna: A Next-generation Hyperparameter Optimization Framework*, arXiv:1907.10902 [cs.LG] (2019).
- [186] J. Bergstra et al., *Hyperopt: a Python library for model selection and hyperparameter optimization*, Comput. Sci. Discov. **8**, 014008 (2015).
- [187] S. M. Lundberg and S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, in Adv. Neural Inf. Process. Syst. Vol. 30 (2017).
- [188] L. S. Shapley, *A Value for N-Person Games*, tech. rep. (RAND Corporation, Mar. 1952).
- [189] S. M. Lundberg et al., *From local explanations to global understanding with explainable AI for trees*, Nat Mach Intell **2**, 56–67 (2020).

-
- [190] T. Keck, F. Abudinén, F. Bernlochner et al., *The Full Event Interpretation*, Comput. Softw. Big Sci. **3**, 6 (2019).
- [191] R. J. Barlow, *Extended maximum likelihood*, Nucl. Instrum. Meth. A **297**, 496–506 (1990).
- [192] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9**, 60–62 (1938).
- [193] A. L. Read, *Presentation of search results: the CLs technique*, J. Phys. G **28**, 2693–2704 (2002).
- [194] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71**, [Erratum: Eur.Phys.J.C 73, 2501 (2013)], 1554 (2011).
- [195] A. Hocker, P. Speckmayer, J. Stelzer et al. (TMVA), *TMVA - Toolkit for Multivariate Data Analysis*, arXiv:physics/0703039 (2009).
- [196] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015.
- [197] A. Paszke et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, in Advances in Neural Information Processing Systems 32 (Curran Associates, Inc., 2019), pp. 8024–8035.
- [198] D. N. Brown, J. Ilic and G. Mohanty, *Extracting longitudinal shower development information from crystal calorimetry plus tracking*, Nucl. Instrum. Meth. A **592**, 254–260 (2008).
- [199] A. Hershenhorn, T. Ferber and C. Hearty, *ECL shower shape variables based on Zernike moments*, Belle II Intern Note (2017).
- [200] J. Bergstra, D. Yamins and D. Cox, *Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures*, in Proc. 30th Int. Conf. Mach. Learn. (Feb. 2013), pp. 115–123.
- [201] C. Guo et al., *On Calibration of Modern Neural Networks*, Aug. 2017.
- [202] A. Natochii et al., *Beam background expectations for Belle II at SuperKEKB*, arXiv:2203.05731 [hep-ex] (2022).
- [203] S. Longo et al., *CsI(Tl) Pulse Shape Discrimination with the Belle II Electromagnetic Calorimeter as a Novel Method to Improve Particle Identification at Electron-Positron Colliders*, Nucl. Instrum. Meth. A **982**, 164562 (2020).
- [204] K. O’Shea and R. Nash, *An Introduction to Convolutional Neural Networks*, arXiv:1511.08458 [cs.NE] (2015).
- [205] A. N. Charan, *Particle identification with the Belle II calorimeter using machine learning*, J. Phys. Conf. Ser. **2438**, 012111 (2023).
- [206] A. Novosel et al., *Identification of light leptons and pions in the electromagnetic calorimeter of Belle II*, Nucl. Instrum. Meth. A **1056**, 168630 (2023).

-
- [207] A. Novosel et al., *ECL image-based charged particle identification using convolutional neural networks at Belle II*, Belle II Intern. Note (2023).
 - [208] Gaetano de Marino and Karim Tarabelsi, *HLT skims*, Belle II Intern. Note (2020).
 - [209] Belle II Charged PID Group (Belle II), *Muon and electron identification efficiencies and hadron-lepton mis-identification rates at Belle II for Moriond 2022*, (2022).
 - [210] *Correction of $\mathcal{B}(B \rightarrow X_c \ell \nu)$ in generic MC - Belle II - DESY Confluence*.
 - [211] P. Zyla, R. Barnett, J. Beringer et al. (Particle Data Group), *Review of Particle Physics*, PTEP **2020**, 083C01 (2020).
 - [212] Y. S. Amhis, S. Banerjee, E. Ben-Haim et al. (HFLAV), *Averages of b -hadron, c -hadron, and τ -lepton properties as of 2018*, Eur. Phys. J. C **81**, 226 (2021).
 - [213] F. Meier, A. Vossen, I. Adachi et al. (Belle), *First observation of $B \rightarrow \bar{D}_1(\rightarrow \bar{D} \pi^+ \pi^-) \ell^+ \nu_\ell$ and measurement of the $B \rightarrow \bar{D}^{(*)} \pi \ell^+ \nu_\ell$ and $B \rightarrow \bar{D}^{(*)} \pi^+ \pi^- \ell^+ \nu_\ell$ branching fractions with hadronic tagging at Belle*, Phys. Rev. D **107**, 092003 (2023).
 - [214] F. U. Bernlochner and Z. Ligeti, *Semileptonic $B_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(D^{**})$* , Phys. Rev. D **95**, 014022 (2017).
 - [215] P. del Amo Sanchez, J. Lees, V. Poireau et al. (BaBar), *Observation of the Decay $B^- \rightarrow D_s^{(*)+} K^- \ell^- \bar{\nu}_\ell$* , Phys. Rev. Lett. **107**, 041804 (2011).
 - [216] J. Stypula, M. Rozanska, I. Adachi et al. (Belle), *Evidence for $B^- \rightarrow D_s^+ K^- \ell^- \bar{\nu}_\ell$ and search for $B^- \rightarrow D_s^{*+} K^- \ell^- \bar{\nu}_\ell$* , Phys. Rev. D **86**, 072007 (2012).
 - [217] J. Lees, V. Poireau, V. Tisserand et al. (BaBar), *Observation of $\bar{B} \rightarrow D^{(*)} \pi^+ \pi^- \ell^- \bar{\nu}$ decays in $e^+ e^-$ collisions at the $\Upsilon(4S)$ resonance*, Phys. Rev. Lett. **116**, 041801 (2016).
 - [218] F. Abudinén, I. Adachi, R. Adak et al. (Belle-II), *Measurement of the inclusive semileptonic B meson branching fraction in 62.8 fb^{-1} of Belle II data*, arXiv:2111.09405 [hep-ex] (2021).
 - [219] D. Bigi, P. Gambino and S. Schacht, *A fresh look at the determination of $|V_{cb}|$ from $B \rightarrow D^* \ell \nu$* , Phys. Lett. B **769**, 441–445 (2017).
 - [220] B. Grinstein and A. Kobach, *Model-Independent Extraction of $|V_{cb}|$ from $\bar{B} \rightarrow D^* \ell \bar{\nu}$* , Phys. Lett. B **771**, 359–364 (2017).
 - [221] A. K. Leibovich et al., *Semileptonic B decays to excited charmed mesons*, Phys. Rev. D **57**, 308–330 (1998).
 - [222] F. U. Bernlochner et al., *Constrained second-order power corrections in HQET: $R(D^{(*)})$, $|V_{cb}|$, and new physics*, Phys. Rev. D **106**, 096015 (2022).

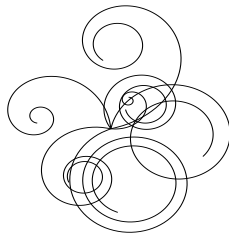
- [223] F. U. Bernlochner, Z. Ligeti and D. J. Robinson, *Model independent analysis of semileptonic B decays to D^{**} for arbitrary new physics*, Phys. Rev. D **97**, 075011 (2018).
- [224] C. Ramirez, J. F. Donoghue and G. Burdman, *Semileptonic $b \rightarrow u$ decay*, Phys. Rev. D **41**, 1496 (1990).
- [225] L. Cao, W. Sutcliffe, R. Van Tonder et al. (Belle), *Measurement of Differential Branching Fractions of Inclusive $B \rightarrow X_u \ell^+ \nu_\ell$ Decays*, Phys. Rev. Lett. **127**, 261801 (2021).
- [226] A. L. Kagan and M. Neubert, *QCD anatomy of $B \rightarrow X_s \gamma$ decays*, Eur. Phys. J. C **7**, 5–27 (1999).
- [227] O. Buchmuller and H. Flacher, *Fit to moment from $B \rightarrow X_c \ell \bar{\nu}$ and $B \rightarrow X_s \gamma$ decays using heavy quark expansions in the kinetic scheme*, Phys. Rev. D **73**, 073008 (2006).
- [228] A. Bharucha, D. M. Straub and R. Zwicky, *$B \rightarrow V \ell^+ \ell^-$ in the Standard Model from Light-Cone Sum Rules*, JHEP **08**, 098 (2016).
- [229] J. A. Bailey, A. Bazavov, C. Bernard et al. (Fermilab Lattice, MILC), *$|V_{ub}|$ from $B \rightarrow \pi \ell \nu$ decays and $(2+1)$ -flavor lattice QCD*, Phys. Rev. D **92**, 014024 (2015).
- [230] F. U. Bernlochner, M. T. Prim and D. J. Robinson, *$B \rightarrow \rho \ell \bar{\nu}$ and $\omega \ell \bar{\nu}$ in and beyond the Standard Model: Improved predictions and $|V_{ub}|$* , Phys. Rev. D **104**, 034032 (2021).
- [231] G. Duplancic and B. Melic, *Form factors of $B, B_s \rightarrow \eta^{(\prime)}$ and $D, D_s \rightarrow \eta^{(\prime)}$ transitions from QCD light-cone sum rules*, JHEP **11**, 138 (2015).
- [232] C. Beleño, A. Frey, I. Adachi et al. (Belle), *Measurement of the branching fraction of the decay $B^+ \rightarrow \pi^+ \pi^- \ell^+ \nu_\ell$ in fully reconstructed events at Belle*, Phys. Rev. D **103**, 112001 (2021).
- [233] G. P. Korchemsky, D. Pirjol and T.-M. Yan, *Radiative leptonic decays of B mesons in QCD*, Phys. Rev. D **61**, 114510 (2000).
- [234] M. Gelb, F. Bernlochner, P. Goldenzweig et al. (Belle), *Search for the rare decay of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ with improved hadronic tagging*, Phys. Rev. D **98**, 112016 (2018).
- [235] J. Brodzicka, T. Browder, P. Chang et al. (Belle), *Physics Achievements from the Belle Experiment*, PTEP **2012**, 04D001 (2012).
- [236] H. Kakuno, K. Abe, I. Adachi et al. (Belle), *Measurement of $|V_{ub}|$ Using Inclusive $B \rightarrow X_u \ell \nu$ Decays with a Novel X_u -Reconstruction Method*, Phys. Rev. Lett. **92**, 101801 (2004).
- [237] P. Urquijo, E. Barberio, I. Adachi et al. (Belle), *Measurement Of $|V_{ub}|$ From Inclusive Charmless Semileptonic B Decays*, Phys. Rev. Lett. **104**, 021801 (2010).
- [238] Marcel Hohmann, *An Inclusive Measurement of $|V_{ub}|$ at the Belle Experiment*, MscThesis (The University of Melbourne, Melbourne, Australia, Oct. 2019).

-
- [239] Ilija Bizjak, *Measurement of $|V_{ub}|$ Using Inclusive Semileptonic Decays on a Sample of Reconstructed B mesons with the Belle Detector*, PhD thesis (University of Ljubljana, Ljubljana, Slovenia, 2005).
 - [240] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. A, Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research **559**, 190–194 (2006).
 - [241] H. Nakano, A. Ishikawa, K. Sumisawa et al. (Belle), *Measurement of time-dependent CP asymmetries in $B^0 \rightarrow K_S^0 \eta \gamma$ decays*, Phys. Rev. D **97**, 092003 (2018).
 - [242] Koji Hara, *Calibration of low momentum K_L^0 efficiency for veto usage in missing E analyses*, Belle Intern. Note **1228**.
 - [243] Sven Einar Vahsen, *Measurement of CP Violation in $B^0 \rightarrow J/\psi K_L$ Decays*, PhD thesis (Princeton University, Princeton, USA, Nov. 2003).
 - [244] K.-J. Tien et al. (Belle), *Evidence for semileptonic $B^- \rightarrow p \bar{p} \ell^- \nu_\ell$ decays*, Phys. Rev. D **89**, 011101 (2014).
 - [245] D. Asner, M. Athanas, D. Bliss et al. (CLEO), *Search for exclusive charmless hadronic B decays*, Phys. Rev. D **53**, 1039–1050 (1996).
 - [246] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, Phys. Rev. Lett. **41**, 1581 (1978).
 - [247] K. Abe, K. Abe, I. Adachi et al. (Belle), *Measurement of branching fractions for $B \rightarrow \pi\pi, K\pi$ and KK decays*, Phys. Rev. Lett. **87**, 101801 (2001).
 - [248] S. Lee, K. Suzuki, K. Abe et al. (Belle Collaboration), *Evidence for $B^0 \rightarrow \pi^0 \pi^0$* , Phys. Rev. Lett. **91**, 261801 (2003).
 - [249] H. Albrecht et al. (ARGUS), *Search for hadronic $b \rightarrow u$ decays*, Phys. Lett. B **241**, 278–282 (1990).
 - [250] Jan Hasenbusch, *Analysis of inclusive semileptonic B meson decays with τ lepton final states at the Belle experiment*, PhD thesis (Rheinische Friedrich-Wilhelms-Universität Bonn, June 2018).
 - [251] N. Chamandy et al., *Estimating Uncertainty for Massive Data Streams*, Technical Report 43157 (Google, 2012).
 - [252] D. Scora and N. Isgur, *Semileptonic Meson Decays in the Quark Model: An Update*, Phys. Rev. D **52**, 2783–2812 (1995).
 - [253] Markus Prim, *eFFORT*, Belle II, May 2022.
 - [254] P. Ball and G. Jones, *$B \rightarrow \eta^{(\prime)}$ form Factors in QCD*, JHEP **08**, 025 (2007).
 - [255] W. Bartel, L. Becker, C. Bowdery et al. (JADE), *Charged particle and neutral kaon production in e^+e^- annihilation at PETRA*, Z. Phys. C - Particles and Fields **20**, 187–206 (1983).

- [256] M. Althoff, W. Braunschweig, F. Kirschfink et al. (TASSO), *A Detailed Study of Strange Particle Production in e^+e^- Annihilation at High-energy*, Z. Phys. C **27**, 27 (1985).
- [257] I. Caprini, L. Lellouch and M. Neubert, *Dispersive bounds on the shape of $\bar{B} \rightarrow D^{(*)}\ell\bar{\nu}$ form factors*, Nucl. Phys. B **530**, 153–181 (1998).
- [258] M. Welsch, *Measurement of the branching fraction of $B \rightarrow D^{**}(\rightarrow D^{(*)}\pi^0)\ell\nu$ in the context of a $R(D^*)$ analysis at the Belle experiment*, MA thesis (Karlsruhe Institute of Technology (KIT), 2019).
- [259] Felix Metzner, *Preparation of a Measurement of $R(D^{(*)})$ with Leptonic τ and Hadronic FEI Tag at the Belle Experiment*, PhD thesis (Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, Feb. 2022).
- [260] M. Rahimi and K. K. Vos, *Standard Model predictions for lepton flavour universality ratios of inclusive semileptonic B decays*, JHEP **11**, 007 (2022).
- [261] L. Aggarwal, H. Ahmed, H. Aihara et al. (Belle-II), *Test of Light-Lepton Universality in the Rates of Inclusive Semileptonic B -Meson Decays at Belle II*, Phys. Rev. Lett. **131**, 051804 (2023).
- [262] M. Gronau, Y. Grossman and J. L. Rosner, *Isospin in B decays and the $(B^0\bar{B}^0)/(B^+B^-)$ production ratio*, Phys. Rev. D **73**, 057501 (2006).
- [263] M. Feindt et al., *A Hierarchical NeuroBayes-based Algorithm for Full Reconstruction of B Mesons at B Factories*, Nucl. Instrum. Meth. A **654**, 432–440 (2011).
- [264] O. Behnke et al., eds., *Data analysis in high energy physics: A practical guide to statistical methods* (Wiley-VCH, Weinheim, Germany, 2013).
- [265] A. Hocker and V. Kartvelishvili, *SVD approach to data unfolding*, Nucl. Instrum. Meth. A **372**, 469–481 (1996).
- [266] K. Tackmann and A. Hoecker, *SVD-based unfolding: implementation and experience*, in PHYSTAT 2011 (2011), pp. 276–279.
- [267] P. Cheema et al., *Development of FastBDT Classifiers to Suppress Beam Background Clusters and Fake Photons*, Belle II Intern. Note (2023).
- [268] D. Martschei et al., *Advanced event reweighting using multivariate analysis*, J. Phys. Conf. Ser. **368**, 012028 (2012).
- [269] I. Adachi, K. Adamczyk, L. Aggarwal et al. (Belle-II), *Evidence for $B^+ \rightarrow K^+\nu\bar{\nu}$ Decays*, arXiv:2311.14647 [hep-ex] (2023).
- [270] P. Ramachandran, B. Zoph and Q. V. Le, *Searching for Activation Functions*, Oct. 2017.
- [271] I. Loshchilov and F. Hutter, *Decoupled Weight Decay Regularization*, in 7th Int. Conf. Learn. Represent. ICLR 2019 New Orleans USA May 6-9 2019 (2019).

-
- [272] G. Kasieczka and D. Shih, *Robust Jet Classifiers through Distance Correlation*, Phys. Rev. Lett. **125**, 122001 (2020).
 - [273] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908**, MODT002 (2003).
 - [274] P. Feichtinger et al., *Punzi-loss: A non-differentiable metric approximation for sensitivity optimisation in the search for new particles*, Eur. Phys. J. C **82**, 121 (2022).
 - [275] L. Heinrich et al., *Pyhf: pure-Python implementation of HistFactory statistical models*, J. Open Source Softw. **6**, 2823 (2021).
 - [276] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, New York, 2012.
 - [277] F. U. Bernlochner et al., *Das ist der HAMMER: Consistent new physics interpretations of semileptonic decays*, Eur. Phys. J. C **80**, 883 (2020).
 - [278] I. Adachi, K. Adamczyk, L. Aggarwal et al. (Belle-II), *First measurement of $R(X_{\tau/\ell})$ as an inclusive test of the $b \rightarrow c\tau\nu$ anomaly*, arXiv:2311.07248 [hep-ex], 10.48550/arXiv.2311.07248 (2023).
 - [279] H. Junkerkalefeld et al., *Test of lepton universality in inclusive, semileptonic and semitauonic B meson decays at Belle II*, Belle II Intern. Note (2023).
 - [280] F. Abudinén, I. Adachi, R. Adak et al. (Belle-II), *Measurement of Hadronic Mass Moments $\langle M_X^n \rangle$ in $B \rightarrow X_c \ell \nu$ Decays at Belle II*, arXiv:2009.04493 [hep-ex], 10.48550/arXiv.2009.04493 (2020).
 - [281] C. Amsler et al. (Particle Data Group), *Review of Particle Physics*, Phys. Lett. B, Review of Particle Physics **667**, 1–1340 (2008).
 - [282] L. Feld, *Continuum Reweighting - Data-Driven Improvement of the Continuum Monte Carlo Simulation for Belle and Belle II*, MA thesis (Karlsruhe Institute of Technology (KIT), 2020).
 - [283] L. D. L. Lyons (CDF), *Everything you always wanted to know about pulls*, tech. rep. CDF-5776 (Nov. 2001).
 - [284] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, Eur. Phys. J. C **70**, 525–530 (2010).
 - [285] O. Vitells and E. Gross, *Estimating the significance of a signal in a multi-dimensional search*, Astropart. Phys. **35**, 230–234 (2011).
 - [286] S. Nishida, *Study of Kaon and Pion Identification Using Inclusive D^* Sample*, Belle Intern. Note **779**.

REFERENCES



Part V

Appendices

Appendix A

Additional material for machine learning based particle identification in the calorimeter

A.1 Optimised hyperparameters

Table A.1 presents the optimised hyperparameters of the nominal set of BDTs developed for the identification of charged particles in the ECL.

A.2 Feature importances

The importances of the feature variables for lepton and pion identification for the 18 BDTs that comprise the ECL charged particle identification method developed in Ch. 5 are presented in Figs. A.1, A.2, A.3.

$p[\text{GeV}]$	$\theta_{\text{d},l} [^\circ]$	q	d	N	l_0	r	l_{min}	α	λ	γ	Evt.Subsample	Col.Subsample	Grow Policy
0.2-0.6	17.0-32.2	-1	3.0	511	0.73	9.7×10^{-5}	0.0015	3.6	22	0.29	0.92	0.98	lossguide
0.2-0.6	17.0-32.2	1	5.0	128	0.52	0.00040	0.045	14	0.67	0.0071	0.92	0.98	lossguide
0.2-0.6	32.2-128.7	-1	4.0	256	0.77	5.6×10^{-5}	0.017	11	16	0.044	1.0	0.95	lossguide
0.2-0.6	32.2-128.7	1	4.0	254	0.69	8.1×10^{-5}	0.0037	0.0012	1.0×10^2	0.0034	1.0	0.96	lossguide
0.2-0.6	128.7-150.0	-1	3.0	511	0.78	0.00027	0.013	0.0046	2.5×10^2	0.015	0.95	0.99	lossguide
0.2-0.6	128.7-150.0	1	3.0	512	0.64	1.5×10^{-6}	0.10	16	0.077	0.19	0.98	0.78	lossguide
0.6-1.0	17.0-32.2	-1	3.0	512	0.59	2.4×10^{-5}	0.0038	1.7×10^{-5}	87	0.15	0.99	0.90	lossguide
0.6-1.0	17.0-32.2	1	4.0	256	0.49	2.5×10^{-6}	0.0085	0.00097	37	0.13	1.0	0.96	lossguide
0.6-1.0	32.2-128.7	-1	4.0	256	0.47	0.00011	0.047	12	7.4	0.0063	0.99	0.94	lossguide
0.6-1.0	32.2-128.7	1	4.0	256	0.44	7.5×10^{-6}	0.0071	7.4	0.19	0.34	0.94	0.77	lossguide
0.6-1.0	128.7-150.0	-1	3.0	510	0.64	3.2×10^{-6}	0.0013	14	0.11	0.0081	0.91	0.93	lossguide
0.6-1.0	128.7-150.0	1	4.0	256	0.48	0.00023	0.017	5.5	0.23	0.089	0.98	0.93	lossguide
1.0-7.0	17.0-32.2	-1	3.0	512	0.36	6.8×10^{-6}	0.0015	4.6	14	0.84	0.99	0.97	lossguide
1.0-7.0	17.0-32.2	1	4.0	256	0.35	1.3×10^{-6}	0.0088	13	5.7	0.063	0.99	0.90	lossguide
1.0-7.0	32.2-128.7	-1	3.0	510	0.35	1.3×10^{-5}	0.0051	4.6	1.6×10^2	0.037	0.98	0.85	lossguide
1.0-7.0	32.2-128.7	1	6.0	62	0.37	0.00024	0.050	16	11	0.61	0.97	0.93	lossguide
1.0-7.0	128.7-150.0	-1	3.0	512	0.33	2.4×10^{-6}	0.0051	4.2	2.8	0.059	0.95	0.85	lossguide
1.0-7.0	128.7-150.0	1	5.0	128	0.38	0.00020	0.0027	0.0045	42	0.0040	0.99	0.87	lossguide

Table A.1: Optimal hyperparameters of the nominal set of BDTs developed for ECL charged particle identification. The first three columns specify the phase-space region of the BDT. For details of the parameters and the optimisation procedure, see Ch. 5.

Belle II Simulation

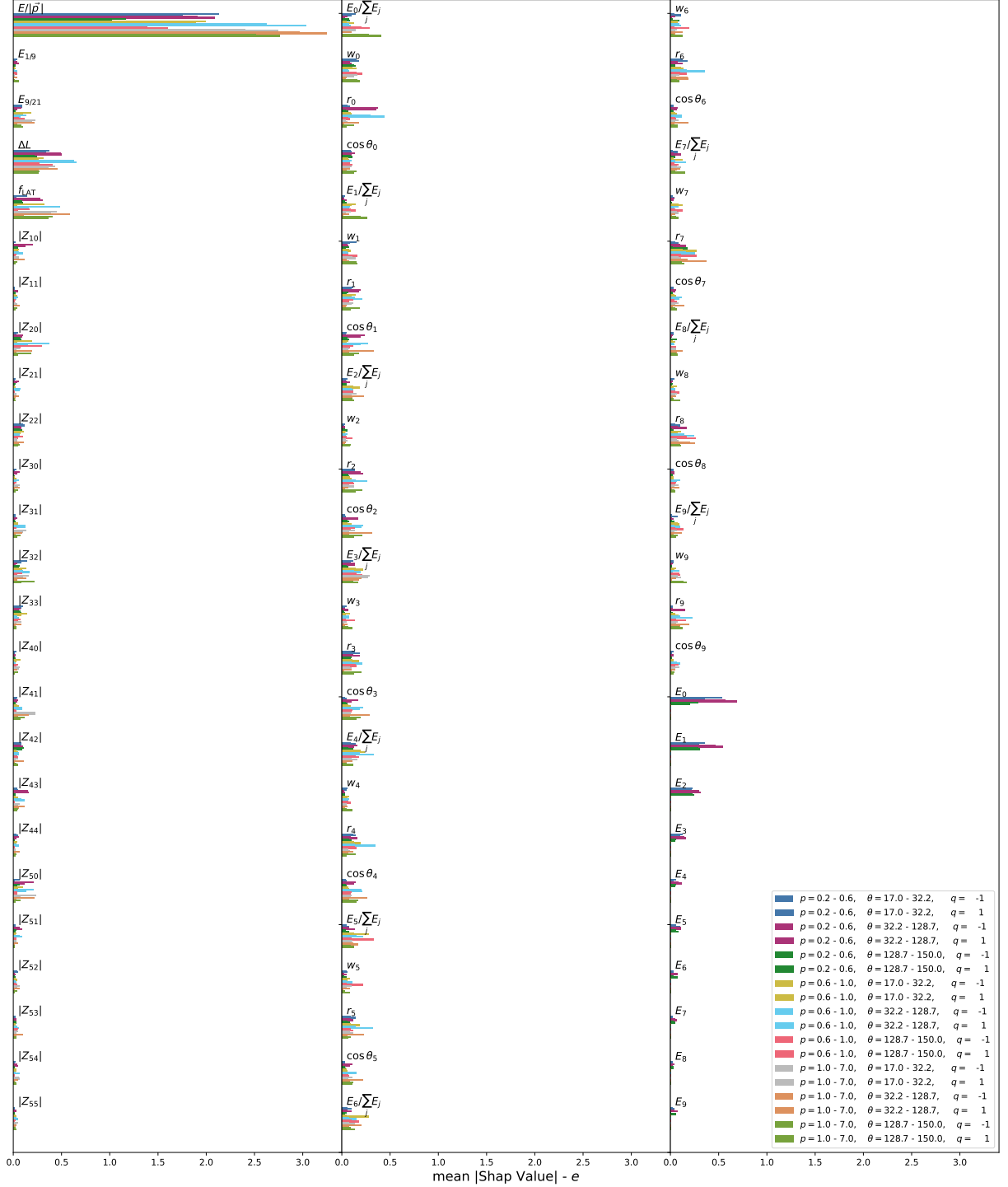


Figure A.1: Importances of the feature variables to electron identification for the 18 nominal reconstruction level ECL charged particle identification BDTs. The importance of each variable is measured by the mean absolute Shap value.

Belle II Simulation

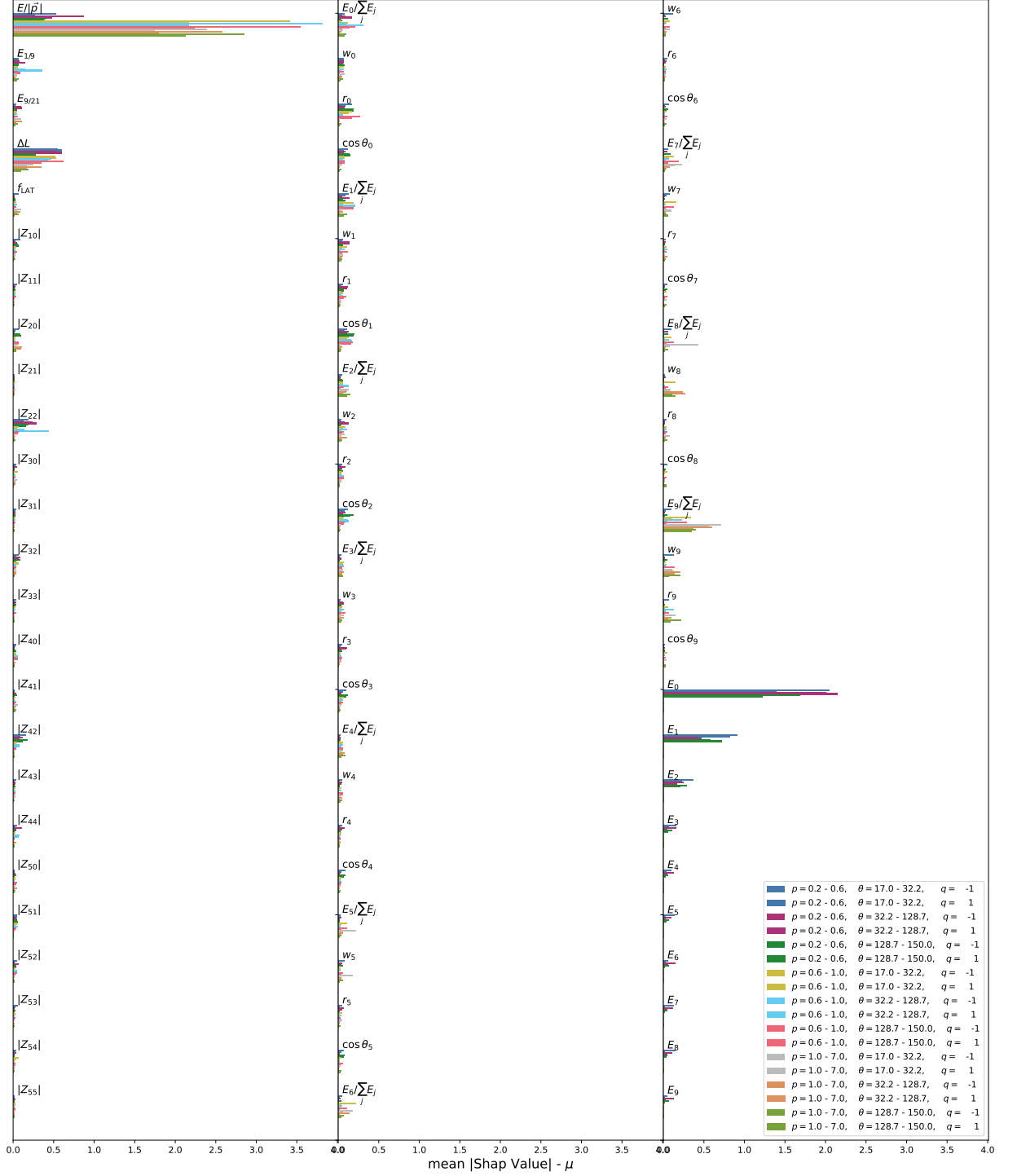


Figure A.2: Importances of the feature variables to muon identification for the 18 nominal reconstruction level ECL charged particle identification BDTs. The importance of each variable is measured by the mean absolute Shap value.

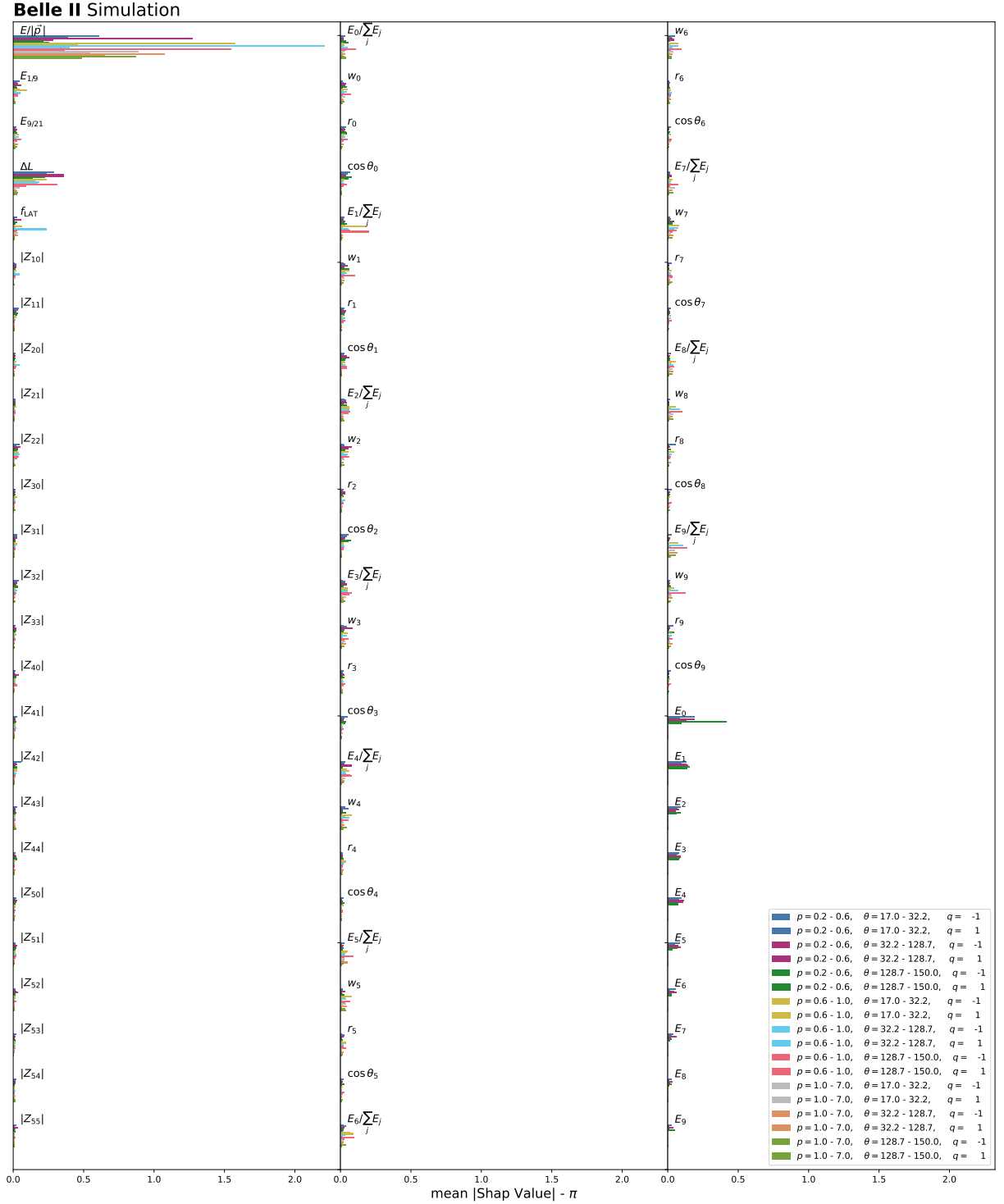


Figure A.3: Importances of the feature variables to pion identification for the 18 nominal reconstruction level ECL charged particle identification BDTs. The importance of each variable is measured by the mean absolute Shap value.

Appendix B

Performance of low momentum kaon identification at Belle

The performance of the Belle kaon identification systems, in terms of both the kaon identification efficiency and pion-to-kaon misidentification rate, has been studied using an inclusive sample of $D^{*+} \rightarrow D^0(\rightarrow K^+\pi^-)\pi_{\text{slow}}^+$ events [286]. The analysis was able to measure both quantities in broad regions of phase-space. Due to stringent background rejection criteria, the study was, however, unable to measure the performance for low-momentum forward particles.

In the inclusive semileptonic analysis presented in Part. III. A the identification of kaons within the rest-of-event is a powerful tag of $\bar{B} \rightarrow X_c \ell \bar{\nu}$ events. As for such events, the hadronic system tends to have a high multiplicity, the momenta of tracks within the rest of event tend to be soft. A significant fraction of tracks thus fall within the phase space in which the performance of the kaon identification is unknown. We therefore perform dedicated studies of the low-momentum kaon identification performance.

B.1 Pion-to-kaon misidentification

The pion-to-kaon misidentification rate is studied using an inclusive K_S^0 sample. The full on-resonance Belle data sample is used. The data collected with the early and late detector geometries are analysed independently, as the change in geometries is expected to affect the particle identification performance of the CDC. We use simulated samples equivalent to three times the expected data yield of $B\bar{B}$ and $q\bar{q}$ events for the early geometry and two times the expected yield for the late geometry.

The analysis is performed in `basf2` [157, 158] using the `b2bii` conversion model [166]. We reconstruct K_S^0 candidates by combining two oppositely charged tracks originating near the interaction point, $|dr| < 2$ cm and $|dz| < 5$ cm. The invariant mass of the di-track system under pion hypotheses is required to be between 400 and 600 MeV. We perform a vertex fit [133], requiring a p-value above 10^{-3} . After the vertex fit the mass selection is tightened to 450 and 550 MeV. Finally, we require that the cosine of the angle between the momentum vector of the reconstructed K_S^0 and the vector connecting the interaction point to the decay vertex of the K_S^0 is greater than 0.998. As discussed in Sec. 6.1.3 this requirement effectively vetoes most remaining combinatorial backgrounds.

The study is performed in bins of polar angle, θ , and track momentum p . The bins are defined by the boundaries:

$$\cos \theta = [-1, -0.612, -0.511, -0.300, -0.152, 0.017, \\ 0.209, 0.355, 0.435, 0.542, 0.692, 0.842, 1], \quad (\text{B.1})$$

$$p = [0, 0.2, 0.3, 0.4, 0.5, \dots, 1.5] \text{ GeV}. \quad (\text{B.2})$$

We follow the existing study and make no distinction between positive and negative tracks. Any K_S^0 candidates in which both child tracks are found within the same bin are excluded to avoid double counting the candidate.

The pion-to-kaon misidentification efficiency of the selection $P_{K/\pi} > 0.6$, is measured via a simultaneous extended maximum likelihood fit to the K_S^0 mass peak of events that pass or fail the selection. The signal is modelled as a sum of three Gaussians that share a common mean μ . The widths, $\sigma_1, \sigma_2, \sigma_3$, and relative contributions of the three Gaussians, $r_1, r_2, r_3 \equiv 1 - r_2 - r_3$, are first extracted by fitting true signal events in MC with no requirement on $P_{K/\pi}$. These six parameters are fixed for the fit to data. As there may be differences in resolution between data and simulation, we, however, introduce a fudge-factor, β , which simultaneously scales all three widths in the fit. All parameters of the signal model are shared for both subsamples (events which pass or fail the likelihood selection). The background in both subsamples is modelled as second degree Chebyshev polynomials. The parameters of these polynomials are independent across the subsamples. The efficiency of the likelihood selection in data, $\varepsilon_{\text{Data}}$, is extracted directly from the simultaneous fit. The fit functions are:

$$F_{P_{K/\pi} > 0.6} = \varepsilon_{\text{Data}} N_{\text{sig}}(r_1 G_1(\mu, \beta \sigma_1) + r_2 G_2(\mu, \beta \sigma_2) + r_3 G_3(\mu, \beta \sigma_3)) \\ + N_{\text{bkg1}} \text{Ch}^2_{\text{bkg1}}(a_{0,1}, a_{1,1}), \quad (\text{B.3})$$

$$F_{P_{K/\pi} < 0.6} = (1 - \varepsilon_{\text{Data}}) N_{\text{sig}}(r_1 G_1(\mu, \beta \sigma_1) + r_2 G_2(\mu, \beta \sigma_2) + r_3 G_3(\mu, \beta \sigma_3)) \\ + N_{\text{bkg2}} \text{Ch}^2_{\text{bkg2}}(a_{0,2}, a_{1,2}), \quad (\text{B.4})$$

where Ch^2 are the Chebyshev polynomials, and G_i the Gaussians. The efficiency of the selection in MC is calculated via truth-matching. However, the fit is performed to the simulated sample to evaluate the bias of the fitting procedure. To do this, the fit is performed 200 times, fluctuating the yields in each bin according to Poissonian statistics for each trial. The mean observed bias is treated as a systematic uncertainty on $\varepsilon_{\text{Data}}$. Furthermore, we vary the parameters $r_1, r_2, r_3, \sigma_1, \sigma_2, \sigma_3$ within their uncertainties for 200 trials and assign the observed spread in results as an additional systematic uncertainty.

In general we observe reasonable fits to data. As the sample has extremely high statistics, some residual pulls are observed in most bins. Tests adding additional degrees of freedom to the signal model, such as switching to bifurcated Gaussians, gave only negligible changes. A fit to an example region is shown in Fig. B.1. Comparisons of the measured misidentification rates with those measured in the previous analysis are shown in Figs. B.3 and B.4, for the early and late detector geometry, respectively. The efficiencies estimated by both methods are in good agreement.

For the analysis presented in Part III. A we use the factors measured on the K_S^0 sample where available.

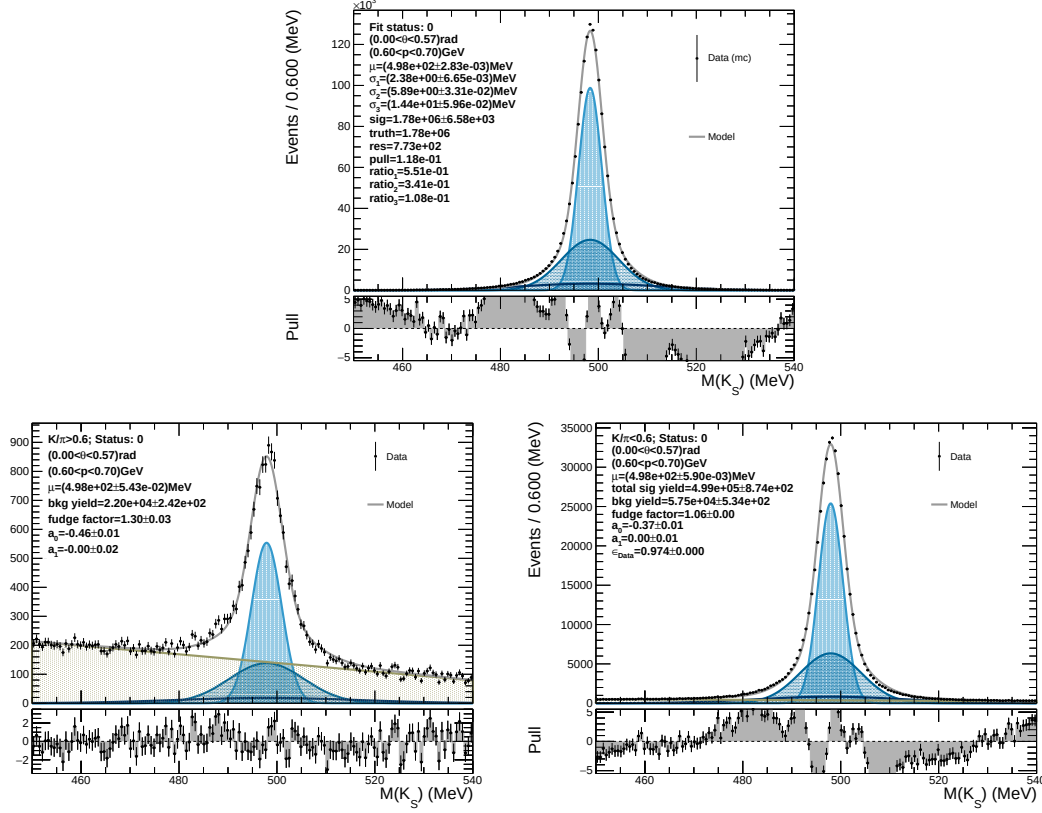


Figure B.1: Example fit to determine the pion-to-kaon misidentification rate. *Top:* Fit to determine the parameters of the signal model in truth-matched simulated events. *Bottom:* Simultaneous fit to events that pass or fail the selection $P_{K/\pi} > 0.6$.

B.2 Kaon identification

The kaon identification is studied by re-analysing the inclusive D^* sample. The procedure closely matches that of the analysis of the K_S^0 sample.

We search for three tracks originating near the interaction point, requiring $|dr| < 2$ cm and $|dz| < 4$ cm. One of these is required to have a momentum between 50 and 200 MeV and is identified as the slow pion. The other two are identified as a kaon and a pion and combined to form a D^0 candidate. We require that the invariant mass of the two tracks lies between 1.835 and 1.895 GeV. Additionally, we require that if the mass hypotheses of the two tracks are switched, the invariant mass falls outside this window. The D^0 candidate is combined with the slow pion to form a D^* and the entire decay chain is vertex fit [133], requiring a p-value above 10^{-3} . Finally, the D^* is required to have a centre-of-mass frame momentum above 0.8 GeV. The main differences to the existing

inclusive D^* study is a relaxation of the centre-of-mass frame momentum selection to include D^* candidates originating in $B\bar{B}$ events, and the removal of a selection on $|\cos\theta_D|$. The selection on $|\cos\theta_D|$, where θ_D is the angle between the kaon momentum in the D rest frame and the D momentum in the centre-of-mass frame, is effective at removing combinatorial background, however, does so by vetoing all events in which the kaon is found in the low-momentum forwards region.

The measurement is performed in bins of polar angle and momentum. The polar angle bins match those of the K_S^0 study. The momentum bins are given by the boundaries:

$$p = [0, 0.5, 0.6, 0.7, \dots 2.9, 3.0, 3.2, 3.4, 3.6, 4.0, 4.5, \infty) \text{ GeV}. \quad (\text{B.5})$$

We measure the efficiency in data and MC for the selection on the likelihood ratio $P_{K/\pi} > 0.6$. The MC efficiency is estimated from MC-truth information. To measure the efficiency in data, the $M_{D^*} - M_{D^0}$ mass difference is fit between 140 and 155 MeV. For each p, θ bin, the sample is split into two subsamples, events that pass or fail the likelihood ratio selection, which are fit simultaneously. The signal model is defined as the sum of two Gaussians and one bifurcated Gaussian. The widths, $\sigma_1, \sigma_2, \sigma_{3,l}, \sigma_{3,r}$, and relative yields, $r_1, r_2, r_b \equiv 1 - r_1 - r_2$ are first fixed by fitting to the truth matched MC sample with no likelihood selection applied. The bifurcated Gaussian is initialised to have the smallest yield, but is not constrained to remain so. When the fit is performed to the data sample, the widths are modified by a common fudge factor, β , to allow for resolution differences between MC and Data. The mean, μ , of the signal peaks is also shared between the two subsamples. For each subsample the combinatorial background is modelled as an empirical function given by:

$$D(x|x_0, a, b, c) = \left(1 - \exp\left(\frac{x - x_0}{c}\right)\right) \left(\frac{x}{x_0}\right)^a + b \left(\frac{x}{x_0} - 1\right). \quad (\text{B.6})$$

The parameters x_0, a, b, c , are independent in each subsample and floated in the fit. The efficiency, $\varepsilon_{\text{Data}}$, is extracted directly from the fits by parameterising the signal yields as $\varepsilon_{\text{Data}} N_{\text{sig}}$ and $(1 - \varepsilon_{\text{Data}}) N_{\text{sig}}$ where N_{sig} is the total signal yield. The combined model is thus:

$$H_{P_{K/\pi} > 0.6} = \varepsilon_{\text{Data}} N_{\text{sig}} (r_1 G_1(\mu, \beta\sigma_1) + r_2 G_2(\mu, \beta\sigma_2) + r_b \text{BG}(\mu, \beta\sigma_{b,l}, \beta\sigma_{b,r})) + N_{\text{bkg1}} D_{\text{bkg1}}(x_0, a_1, b_1, c_1), \quad (\text{B.7})$$

$$H_{P_{K/\pi} < 0.6} = (1 - \varepsilon_{\text{Data}}) N_{\text{sig}} (r_1 G_1(\mu, \beta\sigma_1) + r_3 G_2(\mu, \beta\sigma_2) + r_b \text{BG}(\mu, \beta\sigma_{b,l}, \beta\sigma_{b,r})) + N_{\text{bkg2}} D_{\text{bkg2}}(x_0, a_2, b_2, c_2), \quad (\text{B.8})$$

where G_i are Gaussians, BG is a bifurcated Gaussian, and the background yields in the two subsamples are denoted as N_{bkg1} and N_{bkg2} . Statistical uncertainties are extracted from the fit. Systematic uncertainties due to fixing the signal parameters are extracted via a toy method where the parameters are varied by their uncertainties for 200 trials. The observed spread in measured efficiencies is assigned as the uncertainty. The bias as measured on

MC is assigned as an additional systematic uncertainty following the procedure outlined for the measurement on the K_S^0 sample.

For most bins, we observe the fits to converge well and the measured efficiency to agree well with the efficiencies reported by the previous study. In rare cases, the fit has converged on a clearly incorrect solution. These bins could possibly be recovered by optimising the initial parameters for these bins. This, however, occurs only in bins in which the previous study reports values. As the current study is intended purely to supplement phase-space regions which did not have any correction factors, these poor fits can be safely ignored.

The fit to an example bin is given in Fig. B.2. Comparisons of the measured misidentification rates with those measured in the previous analysis are shown in Figs. B.5 and B.6, for the early and late detector geometry, respectively. The efficiencies estimated by both methods are in good agreement. For the analysis presented in Part III. A, we use the calibration factors for kaon identification measured in the previous study, where available. For phase space regions in which this study reports no calibration factors, the factors measured in this reanalysis of the D^* sample are used.

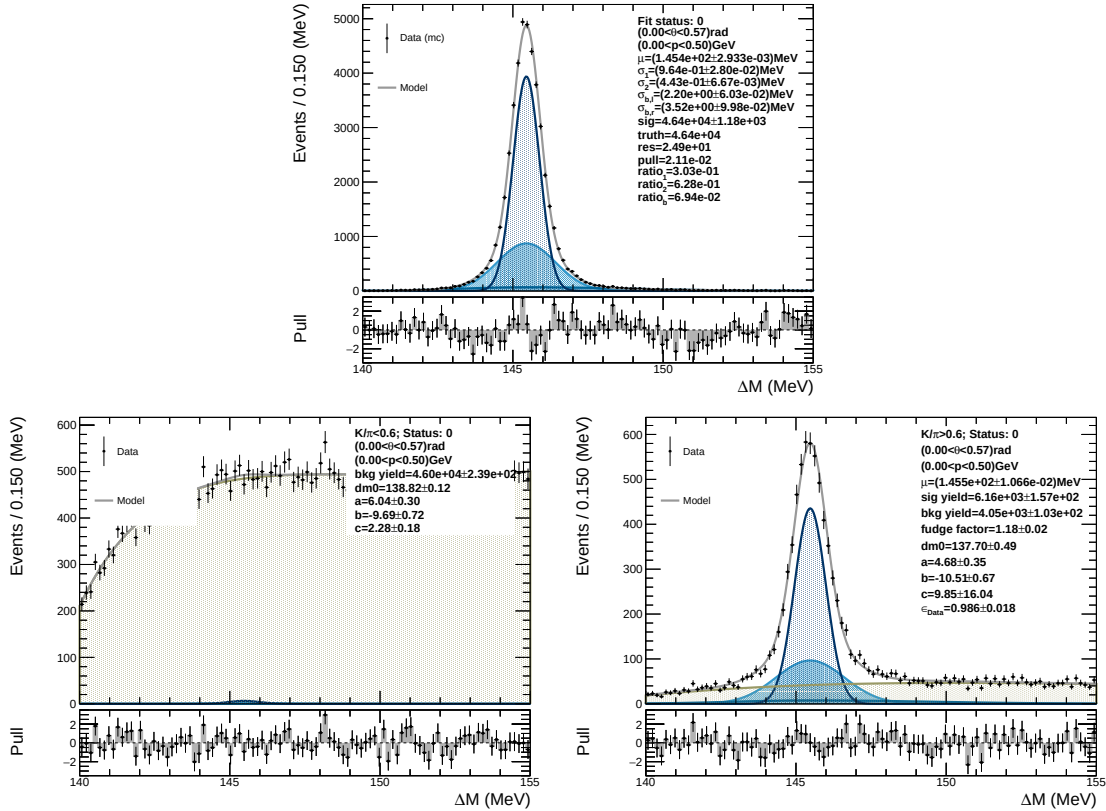


Figure B.2: Example fit to determine the kaon identification efficiency. *Top:* Fit to determine the parameters of the signal model in truth-matched simulated events. *Bottom:* Simultaneous fit to events that pass or fail the selection $P_{K/\pi} > 0.6$.

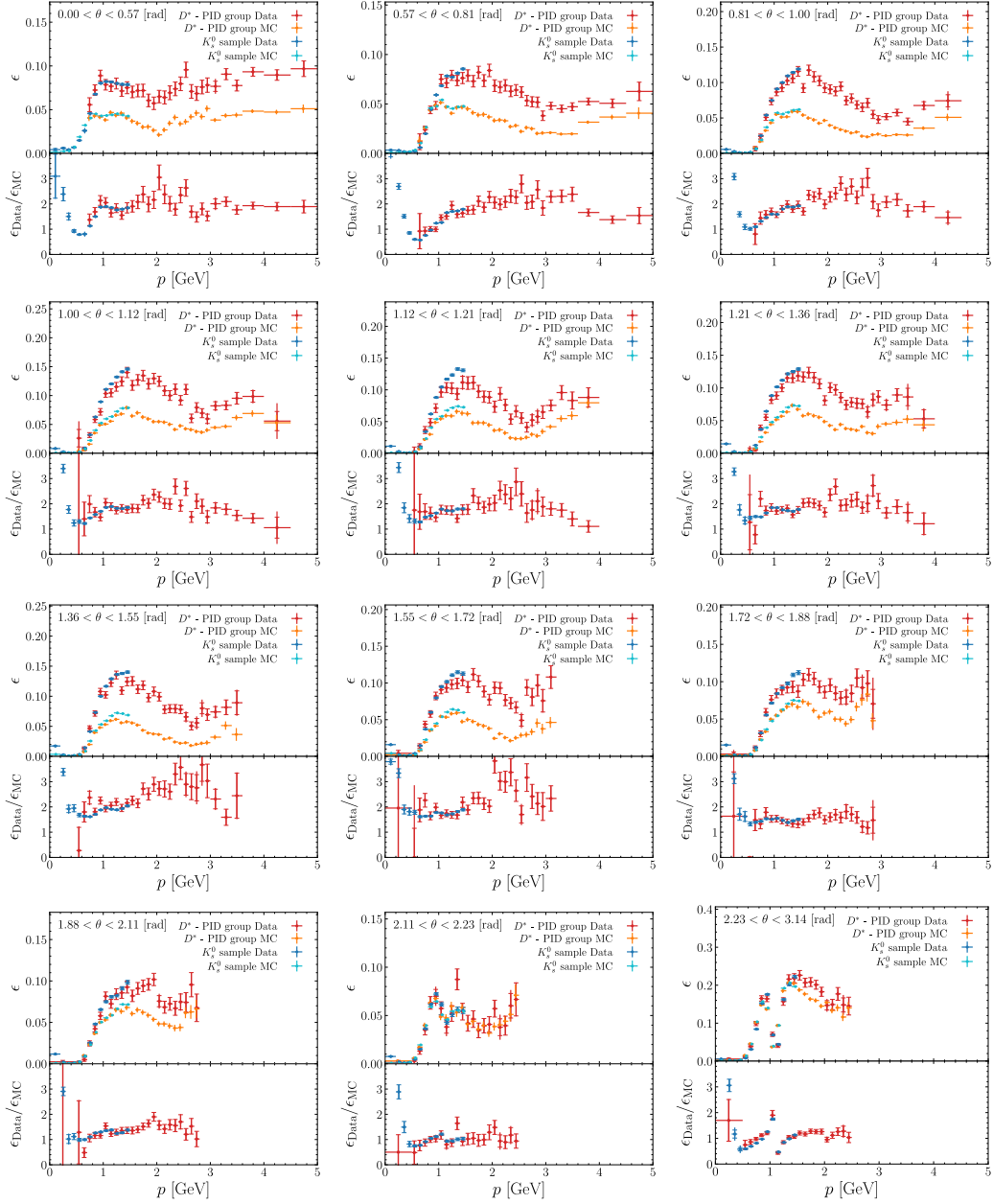


Figure B.3: Pion-to-kaon misidentification rate for data collected with and MC simulated with the SVD1 configuration of the Belle detector.

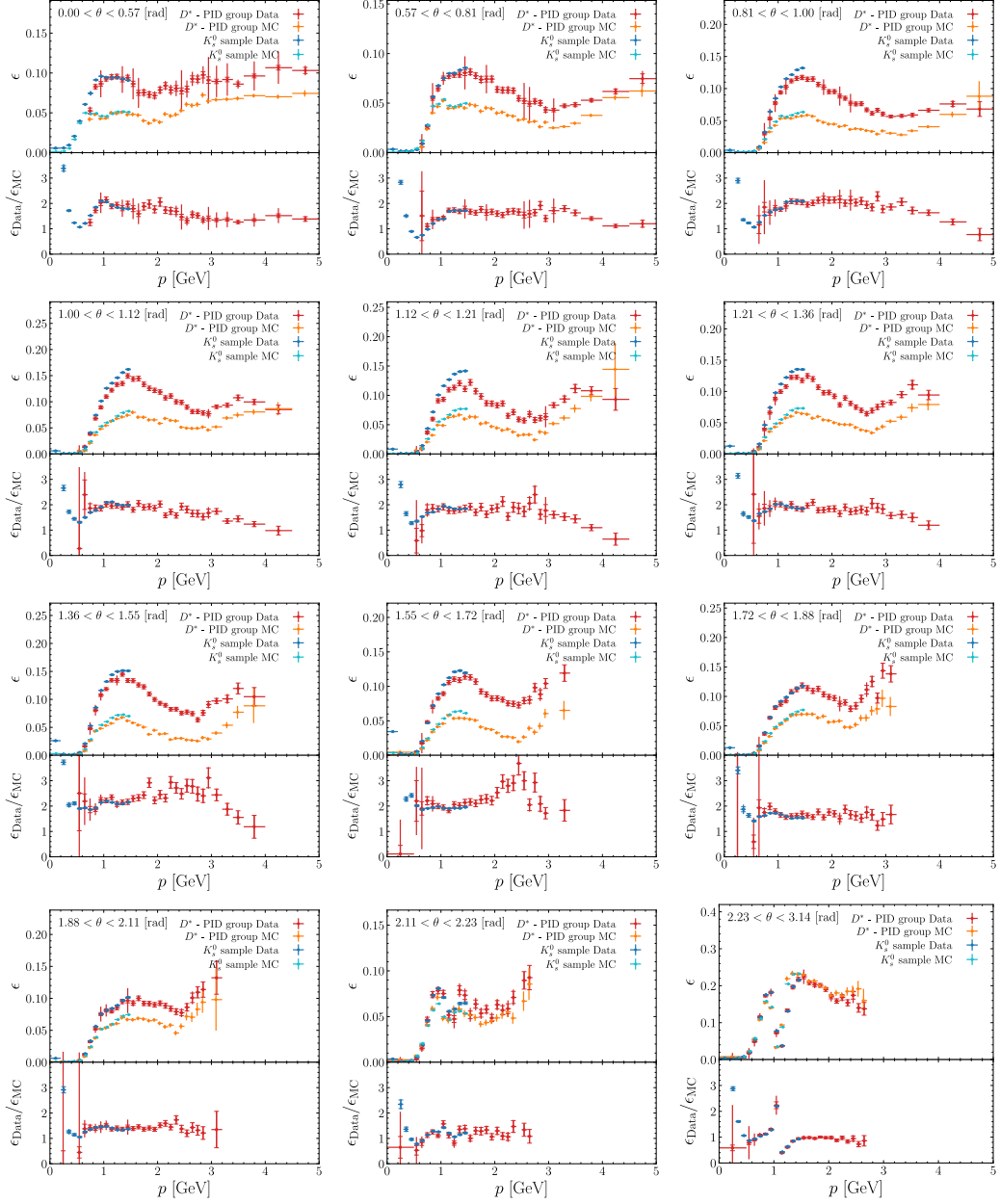


Figure B.4: Pion-to-kaon misidentification rate for data collected with and MC simulated with the SVD2 configuration of the Belle detector.

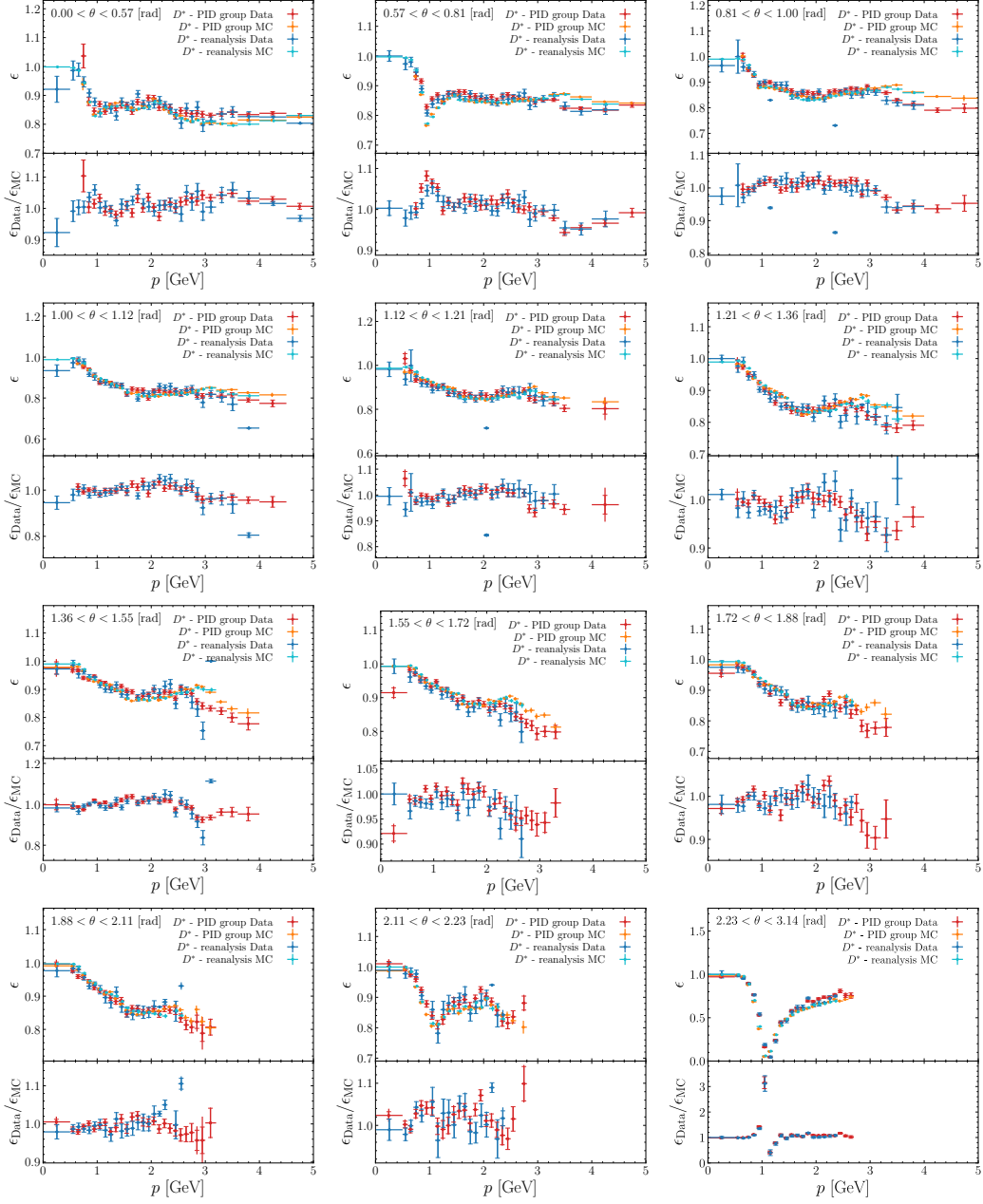


Figure B.5: Kaon identification rate for data collected with and MC simulated with the SVD1 configuration of the Belle detector.

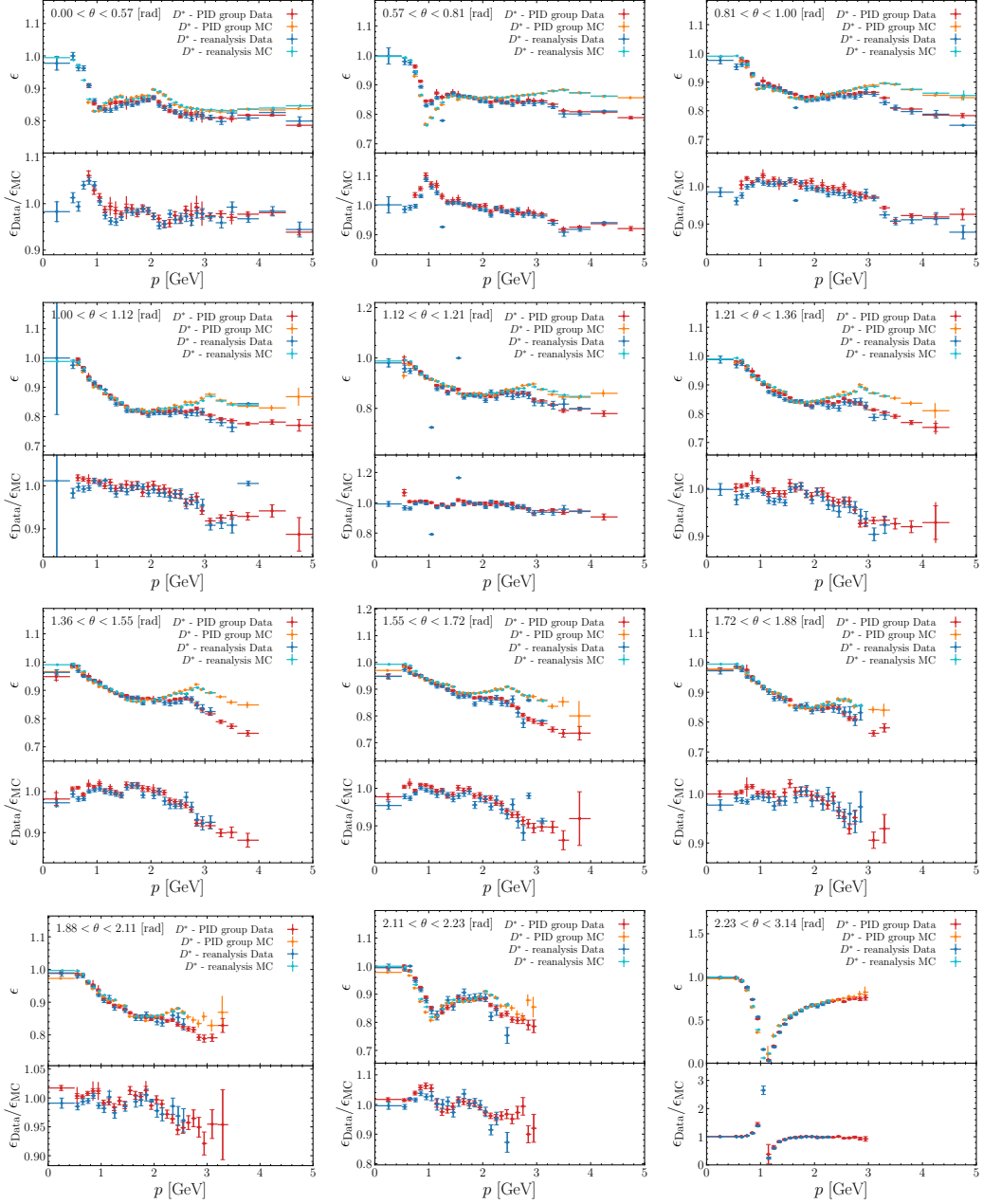


Figure B.6: Kaon identification rate for data collected with and MC simulated with the SVD2 configuration of the Belle detector.

Appendix C

Additional material for the search for weak annihilation contributions to semileptonic charmless B decays

The complete set of pull and linearity tests for all tested weak annihilation models are given in Figs. C.1 and C.2.

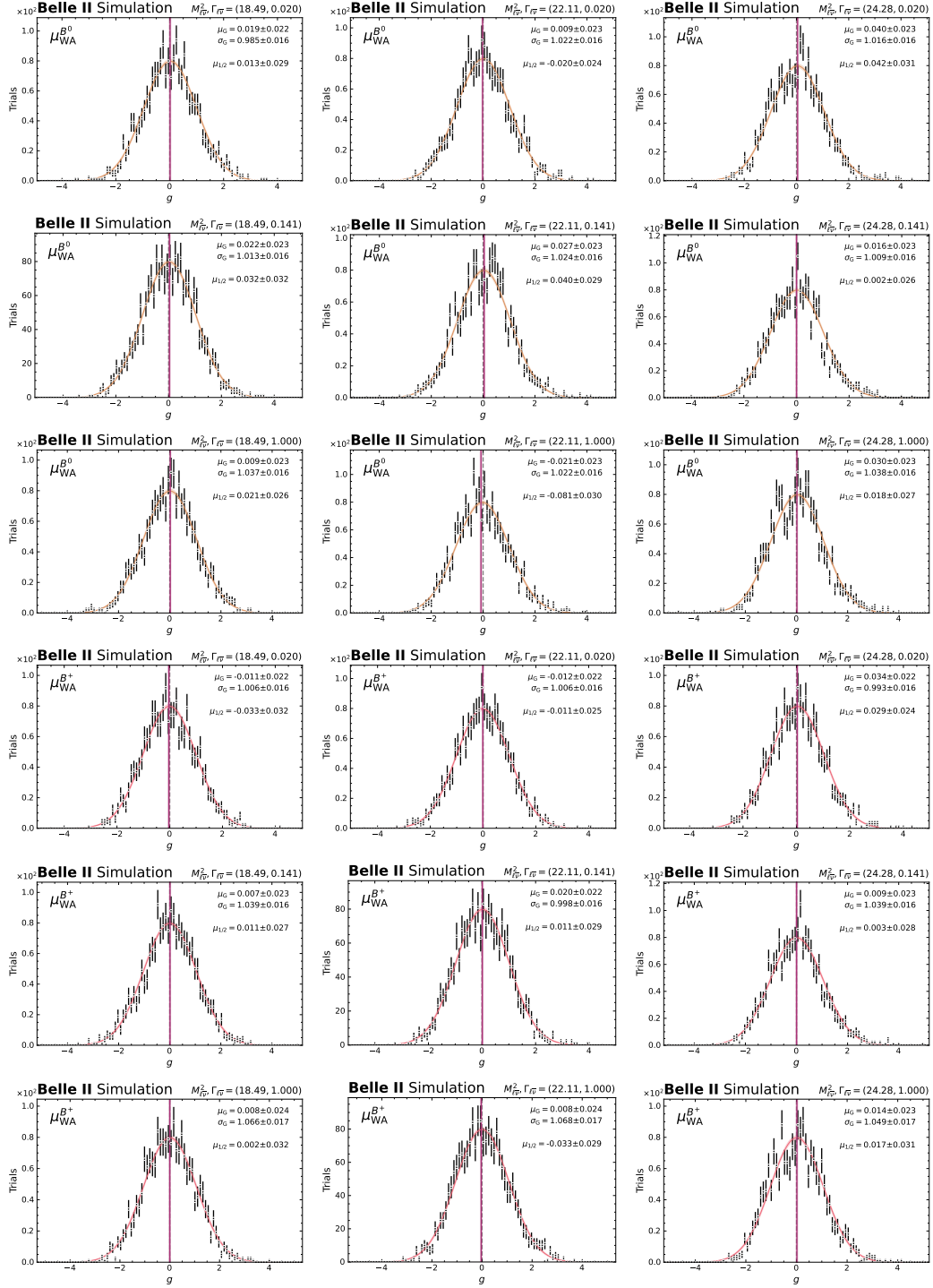


Figure C.1: Pull tests for $\mu_{\text{WA}}^{B^0}$ and $\mu_{\text{WA}}^{B^+}$ for nine example weak annihilation models. The recovered means and widths of the fitted Gaussians generally agree well with zero and one, indicating the fits are able to accurately extract the signal strengths without bias and correctly estimate the uncertainty.

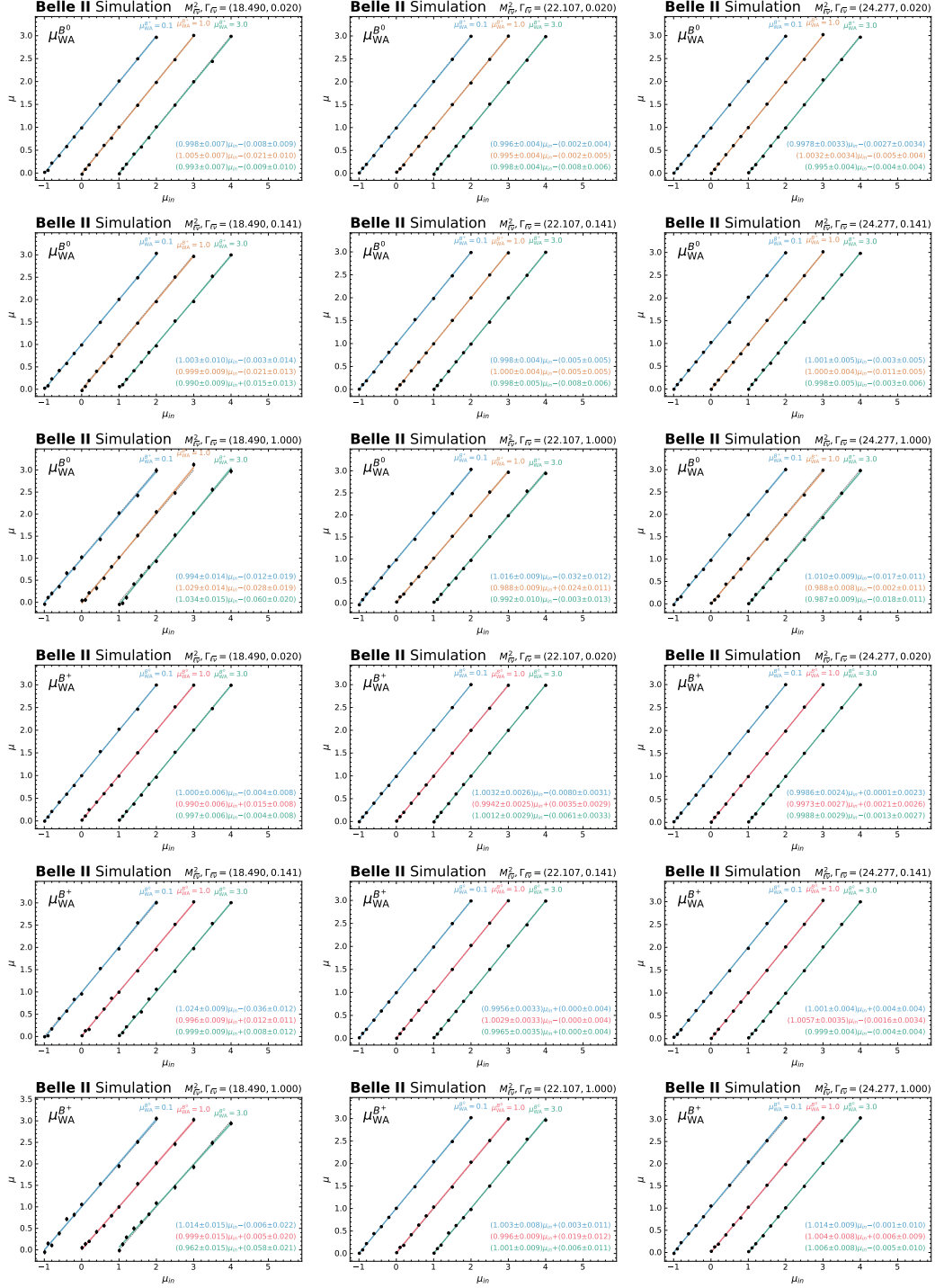


Figure C.2: Linearity tests for $\mu_{WA}^{B^0}$ and $\mu_{WA}^{B^+}$ for the example regions. The results setting the normalisation of the WA component not under study to 0.01 and 3 have been offset in the x -axis for visualisation.
