

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

A measurement of the denominators of $R(D^{(*)})$ with 189 fb^{-1} of Belle II data.

Hannah Marie Wakeling

December 2022

Department of Physics, McGill University, Montréal, Québec

©Hannah Marie Wakeling 2022

Abstract

Using a sub-sample of the Belle II data sample of luminosity $(189.8804 \pm 0.0131) \text{ fb}^{-1}$, the branching fraction $\mathcal{B}(B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell)$, where ℓ is either e or μ , is explored. This analysis first contributed to the published Belle II results of the hadronic tagged $B^0 \rightarrow D^{*+}[D^0 \pi^+] \ell^- \bar{\nu}$ branching fraction where $D^0 \rightarrow K^- \pi^+$ [1]. This study is expanded to use both B^- and $\bar{B}^0$ decays in addition to multiple exclusive D^* and D decay modes. This contributes to one of the first tests of lepton flavour universality at Belle II using exclusive measurements of the denominator of $R(D^{(*)}) = \mathcal{B}(B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau) / \mathcal{B}(B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell)$ where ℓ is either e or μ .

Résumé

En utilisant un sous-échantillon de données Belle II avec une luminosité de $(189.8804 \pm 0.0131) \text{ fb}^{-1}$, le taux de branchement $\mathcal{B}(B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell)$, où ℓ est soit e soit μ , est exploré. Cette analyse a d'abord contribué aux résultats publiés de Belle II sur le taux de branchement $B^0 \rightarrow D^{*+}[D^0 \pi^+] \ell^- \bar{\nu}$ étiquetée hadronique où $D^0 \rightarrow K^- \pi^+$ [1]. Cette étude est élargie pour utiliser à la fois les désintégrations B^- et $\bar{B}^0$ en plus de multiples modes exclusifs de désintégration D^* et D . Cela contribue à l'un des premiers tests d'universalité de la saveur des leptons à Belle II en utilisant des mesures exclusives du dénominateur de $R(D^{(*)}) = \mathcal{B}(B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau) / \mathcal{B}(B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell)$ où ℓ est e ou μ .

Contents

Abstract	ii
Abrégé	iii
List of Figures	viii
List of Tables	xix
Author’s Contribution	xxv
Acknowledgements	xxviii
1. Introduction	1
2. Theory	5
2.1. The Standard Model of Particle Physics	5
2.1.1. The Cabibbo-Kobayashi-Maskawa matrix	9
2.1.2. Effective field theories	11
2.2. Semileptonic B meson decays	13
2.2.1. Search for New Physics beyond the Standard Model	15
2.2.2. The potential for NP in $B \rightarrow D^* \tau^- \bar{\nu}_\tau$	15
2.2.3. $R(D)$ and $R(D^*)$	18
2.2.4. Review of measurements of $\mathcal{B}(B \rightarrow D^{(*)} l^- \bar{\nu}_l)$ and $R(D)$ and $R(D^*)$	19
3. The Belle II Experiment	22
3.1. The SuperKEKB Accelerator	22
3.2. The Belle II Detector	25
3.2.1. Vertex Detector	28
3.2.2. Central Drift Chamber (CDC)	30

Contents

3.2.3.	Time of Propagation (TOP) counter	32
3.2.4.	Aerogel Ring-Imaging CHerenkov detector (ARICH)	33
3.2.5.	Electromagnetic Calorimeter (ECL)	34
3.2.6.	Neutral long-lived Kaon (K_L^0) and Muon (μ) detector (KLM) . . .	36
3.2.7.	Triggering and Data Acquisition	37
4.	Analysis Tools	39
4.1.	Monte Carlo (MC)	39
4.2.	The Belle II Data Sample	41
4.3.	The Belle II Analysis Software Framework (basf2)	42
4.3.1.	basf2 Structure	43
4.4.	The Full Event Interpretation Algorithm (FEI)	44
4.4.1.	Hadronic FEI Skim	46
4.5.	Truth information and truth matching	52
4.5.1.	Topology analysis using TopoAna	53
5.	The measurement of $B \rightarrow D^{(*)}l^- \bar{\nu}_l$ at Belle II	56
5.1.	Motivation of the measurement of $B \rightarrow D^{(*)}l^- \bar{\nu}_l$ at Belle II	56
5.2.	The novel aspects of this analysis	59
6.	Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$	62
6.1.	The analysis procedure	62
6.2.	Online selections	64
6.2.1.	B_{tag} reconstruction	65
6.2.2.	Tracking	65
6.2.3.	π selections	65
6.2.4.	Slow π and γ selections	69
6.2.5.	K_S^0 selection	70
6.2.6.	$D^{(*)}$ reconstruction	70
6.2.7.	Lepton selection	72
6.2.8.	B_{sig} reconstruction	77

Contents

6.2.9. Best Candidate Selection	77
6.2.10. E_{ECL}	80
6.3. Offline selections and corrections	85
6.3.1. Continuum Suppression	85
6.3.2. ROE selection	86
6.3.3. $D^{(*)}$	89
6.3.4. PID corrections	89
6.3.5. $\Upsilon(4S)$	91
6.4. Multivariate analysis reducing specific peaking backgrounds	100
6.4.1. Training variables	102
6.4.2. Target variable	105
6.4.3. MVA Model	106
6.4.4. FastBDT	107
6.4.5. MVA Results	109
6.4.6. The results of the applied MVA	117
7. $R(D)$ and $R(D^*)$ denominator extraction	123
7.1. Fitting procedure and software	123
7.1.1. Fitting with <code>binfit</code>	124
7.2. Fit validation	126
7.3. Fit results	127
7.4. Branching fraction calculation	128
7.4.1. Signal efficiency	130
8. Systematic Uncertainties	131
8.1. $\epsilon_{\text{tag+sel}}$ systematic uncertainties	132
8.1.1. FEI Calibration	132
8.1.2. Tracking Efficiency	133
8.1.3. Slow π data/MC efficiency	134
8.1.4. π^0 efficiency	134
8.1.5. MVA	134

Contents

8.2. N_s systematic uncertainties	135
8.2.1. The fit	135
8.2.2. Branching fraction uncertainties of charm decays	137
8.2.3. Particle ID and Fake Rates	137
8.2.4. Form Factors (FFs)	138
8.3. Total relative systematic uncertainties	139
9. Results and discussion	141
9.1. The “golden mode”	142
9.2. The D^{**} suppression	143
9.3. The future of this analysis	144
10. Conclusion	149
Appendix A. The CKM parametrisation	151
Appendix B. Distribution of variables used for the arbitration approach	153
Appendix C. E_{ECL}	156
Appendix D. The charmed meson D^{**}	159
Appendix E. Supplementary MVA resources	160
Appendix F. Suppression of D^* cross-feed for $B \rightarrow D\ell^-\bar{\nu}_\ell$	169
Appendix G. Alternate binning plots	173
Appendix H. Systematic uncertainties appendices	176
H.1. Tracking efficiency systematics	176
H.2. MVA systematics	176
List of Acronyms	181
Bibliography	189

List of Figures

1.1.	A depiction of a possible collision event with the decay of an $\Upsilon(4S)$ into a B_{tag} and a B_{sig} . The B_{sig} decays into $B \rightarrow D\ell^-\bar{\nu}_\ell$	3
2.1.	The SM of elementary particles. There are twelve fermions and five gauge bosons. For each fermion there is a corresponding anti-particle that is not depicted. The highlighting loops in the background indicate which gauge bosons couple to which fermions. This infographic is correct as of the publication of the 2019 Particle Data Group report [5]. Image sourced from [6].	6
2.2.	The colour charge combinations of baryons and mesons. Image obtained from Reference [7].	7
2.3.	A Feynman diagram depicting two up quarks exchanging a gluon g	8
2.4.	A unitary triangle of the CKM matrix, highlighting the CKM element relevant for this analysis, $ V_{cb} $. Here the $*$ indicates the conjugate of that element.	10
2.5.	A Feynman diagram depicting the decay $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$, with quark composition shown. Here ℓ represents an electron e or muon μ	12
2.6.	A modified Feynman diagram of the SM $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ decay with a charged Higgs as the mediator boson.	17
2.7.	A modified Feynman diagram of the SM $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ decay with a leptoquark as the mediator boson.	17

List of Figures

2.8.	The averages of the $R(D)$ (left) and $R(D^*)$ (right) of the experimental measurements and their averages compared to theoretical SM predictions. This was calculated and compiled by the HFLAV semileptonic group, last updated in May 2019 [11].	21
2.9.	The average SM prediction and experimental measurements of $R(D)$ against $R(D^*)$ by the HFLAV semileptonic group, last updated in May 2019 [11]. .	21
3.1.	The e^+e^- CM energy versus the cross-section e^+e^- to hadrons. The Υ resonances are indicated at each respective peak.	23
3.2.	The instantaneous (peak) luminosity before IR upgrade (red), after IR upgrade (orange) and integrated luminosity (blue) of the SuperKEKB accelerator as a function of time.	25
3.3.	A scaled model of the Belle II detector from the side depicting the relative size of the detector to humans, the coordinate system and the various sub-detectors. Figure sourced from Reference [43].	27
3.4.	The fully upgraded Belle II detector as viewed from the top.	28
3.5.	A schematic of the VDX with the Be beam pipe at the centre. The inner two layers are the PXD layer, whilst the outer four layers are the SVD.[42]	29
3.6.	A schematic comparing the wire configuration of the Belle (‘Present’) and Belle II (‘Upgrade’) drift chambers. [3]	31
3.7.	The specific ionisation of charged particles and their separation at Belle II as a function of momentum. Figure sourced from Reference [44].	32
3.8.	A schematic of the TOP counter consisting of a side view of totally internally reflecting Cherenkov photons, with a separation between K and π mesons shown. The labelled ‘photon detector’ is the PMT sensor plane discussed in the text [3].	33
3.9.	The concept of the ARICH detector at Belle II [3].	34
3.10.	The concept of the proximity focusing ARICH using 2 aerogel layers of different refractive indexes [42].	34

List of Figures

4.1.	The online luminosity of Belle II; the data used for this analysis. The total integrated daily luminosity is indicated in the (blue) histogram, while the cumulative integrated luminosities are indicated by the (red) line. Figure sourced from internal Belle II Confluence website - image created by Stefano Lacaprara [43].	42
4.2.	The basf2 data process chain. Image created by Andreas Moll [47].	43
4.3.	The FEI hierarchical structure showing the progression of the detector information of tracks (and displaced vertices) and neutral clusters (energy deposits in the ECL and KLM) to the final tagged B^+ or B^0 . Figure sourced from Reference [55].	45
4.4.	The FEI calibration factors required to correct for the difference in data-MC efficiency as a function of $\mathcal{P}_{\mathcal{B}_{\text{tag}}}$. The calibrations are separated into charged and neutral B mesons, and then electrons, muons and combined electrons and muons. This is a preliminary result, updated from Reference [56].	50
4.5.	The lepton momentum in the CM frame for $B \rightarrow X\ell^-\bar{\nu}_\ell$ used to extract the FEI calibration factor, in this case using the B_{tag} and μ . Generic MC is classed as $B \rightarrow X\ell^-\bar{\nu}_\ell$, and background. Data is overlaid as points with statistical uncertainties shown as error bars. This result is updated from Reference [56].	51
4.6.	A schematic depicting two topology diagrams of processes occurring at e^+e^- colliders. Diagram (a) is the $e^+e^- \rightarrow J/\Psi$ decay, while diagram (b) contains one possible iteration of the decays studied in this analysis: an $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ decay where the $\bar{B}^0$ decays to $D^{*+}\mu^-\bar{\nu}_\mu$ [55].	54

List of Figures

5.1.	The expected total uncertainty of various branching fraction ratios at Belle II over time. Time represents the expected dataset size and is subject to change. The pessimistic scenario (left) represents a prediction using current statistical and systematic errors that contain the irreducible 3% uncertainty. The optimistic scenario (right) lowered the irreducible systematic to 0.5% and assumes a 50% increase of the effectiveness of the exclusive tagging algorithms. The grey bands represent data collection downtime. The relevant uncertainties for this analysis are the $R(D)$ and $R(D^*)$ (had FEI, lep τ) ratios in green. Figure obtained from Reference [58].	59
5.2.	The projection of the $R(D)$ and $R(D^*)$ measurements at Belle II for 5 ab^{-1} (left) and 50 ab^{-1} (right). Figure obtained from Reference [42].	60
6.1.	Distributions of M_{bc} for $B \rightarrow D^{(*)}l^-\bar{\nu}_l$. Generic MC is classed as $B^0\bar{B}^0$, B^+B^- and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The red arrows depict the bins where the residual pull is more than 1σ away from data-MC agreement.	66
6.2.	Distributions of ΔE of the B_{tag} in the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ decay. Generic MC is classed as $B^0\bar{B}^0$, B^+B^- and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The red arrows depict the bins where the residual pull is more than 1σ away from data-MC agreement. . .	67
6.3.	Distributions of the base 10 logarithm of the $\mathcal{P}_{B_{\text{tag}}}$ for $B \rightarrow D^{(*)}l^-\bar{\nu}_l$. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as $B^0\bar{B}^0$, B^+B^- and continuum. Data is overlaid as points with statistical uncertainties shown as error bars.	68
6.4.	The invariant mass of the B_{sig} daughter D mesons used either as candidates in the $B \rightarrow Dl^-\bar{\nu}_l$ decay channels or as candidates for D^* daughters in the $B \rightarrow D^*l^-\bar{\nu}_l$ decay channels. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as $B^0\bar{B}^0$, B^+B^- and continuum. Data is overlaid as points with statistical uncertainties shown as error bars.	72

List of Figures

- 6.5. The mass difference between the B meson daughter and granddaughter particles. With restrictions, this variable should represent the mass difference between D^* meson and it's daughter D meson from the $B \rightarrow D^* l^- \bar{\nu}_l$ decay, and would peak around 0.145 GeV for the D^{*+} decay. 73
- 6.6. The lepton momentum spectrum for the $B \rightarrow D^* l^- \bar{\nu}_l$ decay channels. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. 74
- 6.7. The mass difference between the B meson daughter and granddaughter particles. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection. 80
- 6.8. The invariant mass of the B_{sig} daughter D mesons used either as candidates in the $B \rightarrow D l^- \bar{\nu}_l$ decay channels or as candidates for D^* daughters in the $B \rightarrow D^* l^- \bar{\nu}_l$ decay channels. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection. 81
- 6.9. The m_{miss}^2 distributions for $B \rightarrow D^{(*)} l^- \bar{\nu}_l$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)} l^- \bar{\nu}_l$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection. 82
- 6.10. The distribution of the number of tracks in the ROE for $B \rightarrow D^{(*)} l^- \bar{\nu}_l$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)} l^- \bar{\nu}_l$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection. 83

List of Figures

6.11. The m_{miss}^2 distributions for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection.	84
6.12. Distributions of R_2 for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. An offline cut of $R_2 < 0.3$ is projected on the figures.	87
6.13. The E_{ECL} distributions after the E_{ECL} beam background suppression is applied. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. An offline cut of $E_{\text{ECL}} < 1.0 \text{ GeV}$ is projected onto the figures.	88
6.14. The D meson invariant mass for the $B \rightarrow D^*\ell^- \bar{\nu}_\ell$ decay channels. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), D^{**} background, other backgrounds and continuum. The D^* optimisation selections have been applied.	90
6.15. The m_{miss}^2 after online and offline selections and corrections are applied. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.	92
6.16. The m_{miss}^2 after online and offline selections and corrections are applied, separated by lepton channel. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.	93

List of Figures

6.18. The m_{miss}^2 after online and offline selections and corrections are applied, separated by D^* decay channel. Generic MC is classed as the normalisation channels ($B \rightarrow D^* \ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background, other backgrounds and continuum.	95
6.19. The m_{miss}^2 after online and offline selections and corrections are applied, separated by D decay channel. Generic MC is classed as the normalisation channels ($B \rightarrow D \ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background, $B \rightarrow D \ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.	96
6.21. The m_{miss}^2 after online and offline selections and corrections are applied, separated by D decay channel. Generic MC is classed as the normalisation channels ($B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$), $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background, $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.	98
6.23. The distributions of some of the MVA training variables for the $\bar{B}^0 \rightarrow D^{*+} l^- \bar{\nu}_l$. Signal MC is separated using TopoAna into the signal, normalisation and D^{**} backgrounds. ‘Weighted’ refers to the signal MC being weighted to the 189 fb^{-1} of data.	104
6.25. The mean ROC of the preliminary machine learning techniques tested on the $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ decays.	107
6.26. The classifier output for the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ MVA. The red sample represents the backgrounds and the blue sample represents the signal and normalisation samples. The filled histograms are the training data, whilst the points with statistical error bars represent the test data. p represents the pull between testing and training data.	109
6.27. The classification result the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ MVA. The purity (left plot, red) and efficiency are shown (left plot, blue), and the true positive (right plot, blue) and false positive rates (right plot, red) are shown (right), all as a function of classifier output cut.	110
6.28. The ROC test result (red) in comparison to the training ROC (blue) for the $\bar{B}^0 \rightarrow D^{*+} l^- \bar{\nu}_l$ MVA. AUC values are shown and indicate good agreement.	111

List of Figures

6.29. The percentage drop in MC type as a function of MVA output classifier cut. The dotted coloured lines represent dedicated signal MC. The solid coloured lines represent generic MC split into the respective TopoAna selected $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decays, background decays and continuum.	113
6.30. The data-MC ratio as a function of classifier cut for the $\bar{B}^0 \rightarrow D^{*+}l^- \bar{\nu}_l$ decay channel. Errors shown are statistical uncertainties.	114
6.31. The output classifiers for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ MVAs. Where appropriate, the y-axes have been optimized for visualisation of backgrounds. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.	115
6.32. The post-MVA classifier cut m_{miss}^2 for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ MVAs. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum. Errors shown are statistical uncertainties. . .	118
6.33. The post-MVA classifier cut m_{miss}^2 for each of the $\bar{B}^0 \rightarrow D^{(*)+}\ell^- \bar{\nu}_\ell$ MVAs separated by lepton candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum. Errors shown are statistical uncertainties.	119
6.34. The post-MVA classifier cut m_{miss}^2 for each of the $B^- \rightarrow D^{(*)0}\ell^- \bar{\nu}_\ell$ MVAs separated by lepton candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum. Errors shown are statistical uncertainties.	120
6.35. The post-MVA classifier cut m_{miss}^2 for each of the $B \rightarrow D^*\ell^- \bar{\nu}_\ell$ MVAs separated by each D^* candidate decay type. Generic MC is classed as the normalisation channels ($B \rightarrow D^*\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, other backgrounds and continuum. Errors shown are statistical uncertainties. . .	121

List of Figures

7.1.	The pre-fit m_{miss}^2 distribution for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ channels. Generic MC is classed as the normalisation channel $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ and background. Errors shown are statistical uncertainties. The pull shows the standard deviation of data from MC, with red arrows indicating a difference of more than 1σ	125
7.2.	The results of the test of linearity of the fit on the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ decay channel. Generic MC is classed as the normalisation channel $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ and background. The errors shown are statistical uncertainties.	127
7.3.	The Asimov test of the fit model on the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ decay channel. The errors shown are statistical uncertainties.	128
7.4.	The post-fit m_{miss}^2 distributions for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ channels. Generic MC is classed as the normalisation channel $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ and background. Errors shown are statistical uncertainties. The pull shows the standard deviation of data from MC.	129
8.1.	The ratio of MC to data as a function of classifier cut for $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$.	135
8.2.	The denominator of $R(D^*)$ as a function of classifier cut for $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$.	136
B.1.	Distributions of a selection of variables used for ranking $\bar{B}^0 \rightarrow D^{(*)+}\ell^- \bar{\nu}_\ell$ events by B_{sig} candidate. The $B^0\bar{B}^0$ generic MC sample is classified using TopoAna and the truth-matched signal-side lepton.	154
B.2.	Distributions of a selection of variables used for ranking $B^- \rightarrow D^{(*)0}\ell^- \bar{\nu}_\ell$ events by B_{sig} candidate. The B^+B^- generic MC sample is classified using TopoAna and the truth-matched signal-side lepton.	155
C.1.	The E_{ECL} distributions before the E_{ECL} beam background suppression is applied. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars.	157

List of Figures

C.2.	Distributions of the E_{ECL} for $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ signal MC before (left) and after (right) the E_{ECL} beam background suppression is applied. The signal MC is classified using TopoAna as selected $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ events from generated $B^0 \rightarrow X_C \ell^- \bar{\nu}_\ell$. Correct reconstruction depends on the truth matched signal-side lepton in the event. The other $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays show similar changes in their E_{ECL} distributions.	158
E.1.	The classifier output for the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ MVA. The red sample represents the backgrounds and the blue sample represents the signal and normalisation samples. The filled histograms are the training data, whilst the points with statistical error bars represent the test data.	160
E.1.	The classifier output for each of the $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ MVAs. The red sample represents the backgrounds and the blue sample represents the signal and normalisation samples. The filled histograms are the training data, whilst the points with statistical error bars represent the test data.	161
E.2.	The classification result for the $B \rightarrow D \ell^- \bar{\nu}_\ell$ MVA. The purity (left plot, red) and efficiency are shown (left plot, blue), and the true positive (right plot, blue) and false positive rates (right plot, red) are shown (right), all as a function of classifier output cut.	162
E.2.	The classification result for each of the $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ MVAs. For the two decay channels the purity (left plot, red) and efficiency are shown (left plot, blue), and the true positive (right plot, blue) and false positive rates (right plot, red) are shown (right), all as a function of classifier output cut.	163
E.3.	The ROC test result (red) in comparison to the training ROC (blue) for each of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ MVAs. AUC values are shown and indicate good agreement.	164
E.4.	The data-MC ratio as a function of classifier cut for the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decay channel. Errors shown are statistical uncertainties.	168
F.1.	A 2D distribution of m_{miss}^2 versus p_{lep} of the candidate $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ events.	170

List of Figures

F.2.	The separated 2D distributions of m_{miss}^2 versus p_{lep} of the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ candidate events using TopoAna truth information. The $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ distribution (left) can be clearly seen to be the vertical distribution, while the contributing D^* cross-feed (right) has a diagonal distribution.	171
F.3.	The separated 2D distributions of m_{miss}^2 versus p_{lep} of the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ decay using TopoAna truth information. The $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ distribution (left) was minimally affected by the cut, while the contributing D^* cross-feed (right) has a largely reduced contribution to the $B \rightarrow D \ell^- \bar{\nu}_\ell$ signal region.	171
F.4.	The resulting m_{miss}^2 distribution for $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ after the suppression of D^* cross-feed events. The study was performed with 62.8 fb^{-1}	172
G.1.	The post-MVA classifier cut m_{miss}^2 for each of the $\bar{B}^0 \rightarrow D^{(*)+} \ell^- \bar{\nu}_\ell$ MVAs.	174
G.2.	The post-MVA classifier cut m_{miss}^2 for each of the $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ MVAs. .	175
H.1.	The ratio of MC to data as a function of classifier cut. At higher classifier cuts there is failure due to low statistics in certain bins leading to a non-accurate covariance.	179
H.2.	The denominator of $R(D^*)$ as a function of classifier cut. At higher classifier cuts there is failure due to low statistics in certain bins leading to a non-accurate covariance.	180

List of Tables

2.1.	The average of the measurements of each of the different $B \rightarrow D^{(*)}l^{-}\bar{\nu}_l$ decays. All branching fractions are averages taken from the 2020 update of the PDG Review of Particle Physics [5].	20
2.2.	A summary of the averages of the values measured/predicted for $R(D)$ and $R(D^*)$ reported by HFLAV, updated in May 2019 [11].	20
3.1.	The total production cross-sections from the $e^{-}e^{+}$ processes occurring at $\Upsilon(4S)$ CM energy $\sqrt{s} = 10.58$ GeV. Errors that are given are Monte Carlo (MC) statistics. Information obtained from Table 18 of Reference [42]. . . .	26
3.2.	The retention (%) for the Hadron skims on Run 5710 experiment 12 (FILTER-OFF) and Run 5157 experiment 10 (FILTER-ON), shown with the efficiencies (%) for generic MC using <code>basf2</code> software <code>release-04-02-05</code>	37
4.1.	The production cross-sections of the $e^{+}e^{-}$ processes at the $\Upsilon(4S)$ CM energy used for the 14 th run-independent MC simulation campaign. The corresponding Belle II MC name is shown for each process simulated in generic MC. Data for this table was collected from internal Belle II calculations and production notes [46].	40
4.2.	A summary of the dataset used for this analysis.	41

List of Tables

4.3.	The PDG branching fractions of the decays used in the hadronic FEI tag with corresponding assigned decay mode IDs [5]. No branching fraction in the PDG thus the following branching fraction values are the ones implemented in MC14 production: [i] value kept from Belle. [ii] removed from MC production. [iii] calculated from Isospin: $\mathcal{B}(J/\psi K^+ \pi^0) = \mathcal{B}(J/\psi K^0 \pi^+)/2$. [iv] branching fraction is the sum of both $B^0 \rightarrow D^{*-} D^+ K_S^0$ and $D^- D^{*+} K_S^0$. [v] resonant component is subtracted.	47
4.4.	The centrally validated and applied hadronic FEI cuts.(*omitted decay modes 23 for B^+ and 25 for B^0 which are listed in Table 4.3).	48
4.5.	The retention of the Hadronic FEI skim by MC type. The generic MC number of events generated corresponds to 1 ab^{-1} of simulated data. The different signal MC samples used in this analysis have the Belle II signal MC code stated for reproducibility.	49
4.6.	The recently updated FEI efficiency corrections for MC14ri_a calibrated for $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ and $B \rightarrow X \ell^- \bar{\nu}_\ell$	52
5.1.	A summary of the D^* and D meson decay channels and their corresponding branching fractions selected for the analysis. Values obtained from Reference [5].	61
6.1.	A summary of the software and relevant releases/versions used for this analysis.	63
6.2.	A summary of the online signal selections applied in this analysis. An issue was discovered with the probability density function for the electron ID of the TOP, which is present in release-05 of basf2 ; the TOP electron ID was removed for this analysis.	75
6.3.	The π^0 standard particle list with 40% data-MC efficiency defined in May 2020.	75
6.4.	The π^0 standard particle list with 50% data-MC efficiency defined in May 2020.	76
6.5.	A summary of the π^0 correction factors and corresponding systematic uncertainties as a function of momentum.	76

6.6.	A summary of the offline signal selections applied in this analysis. Here ‘else’ refers to the remaining $D^{(*)}$ decays in the mode selections.	86
6.7.	The size of the D^{**} background yield in generic MC compared to a corresponding COI yield, represented as a percentage. Yields are after all selections prior to the MVA are applied.	100
6.8.	The list of various training variables used in the four MVAs. Variables that are not used to train all four MVAs have their specific usage indicated. . .	103
6.9.	The yield for the $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**}\ell^-\bar{\nu}_\ell$ background is a subset of the background MC in this instance. Note that the number of events is rounded to an integer value (weighting renders events non-integer values).	112
6.10.	The percentage drop in yield for the $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**}\ell^-\bar{\nu}_\ell$ background is a subset of the background MC in this instance. The percentage drop in yields shown are calculated from the non-rounded yields in Table 6.9.	112
6.11.	The yields per generic MC grouping before and after the MVA selection is applied to each of the $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ channels. The $B \rightarrow D^{**}\ell^-\bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values). The percentage drop in yields shown are calculated from the non-rounded values.	116
6.12.	The expected MC signal yield (N_{MC}^{sig}), background yield (with continuum included as a background) (N_{MC}^{bkgd}) and sum of the two (N_{MC}^{total}), compared to the final measured yield in data (N_{data}) of the $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decay channels. The yields of the separated electron and muon are included. Errors shown are individually calculated statistical uncertainties.	122

List of Tables

7.1.	The fitted MC signal yield ($N_{\text{MC,fit}}^{\text{sig}}$), fitted background yield ($N_{\text{MC,fit}}^{\text{bkgd}}$) and total fit yield ($N_{\text{MC,fit}}^{\text{total}}$), compared to the final measured yield in data (N_{data}) of each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels. Errors shown are statistical uncertainties.	130
7.2.	The signal efficiency of each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels calculated using signal MC. The errors shown are statistical uncertainties.	130
8.1.	The average number of tracks per event for each of the four $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ normalisation decay modes and the associated systematic uncertainty. . . .	133
8.2.	The systematic uncertainties as a result of the tracking scale factor for each of the four $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ normalisation decay modes.	134
8.3.	The total fit yield $N_{\text{MC,fit}}^{\text{total}}$ with statistical and systematic uncertainties included.	137
8.4.	The relative systematic errors of the branching fractions on the normalisation for $D^{(*)}$ meson decays. Values are calculated from Table 5.1 [5].	138
8.5.	A summary of the relative systematic uncertainties on the measurements of the denominators of $R(D)$ and $R(D^*)$, separated into the four relevant normalisation decay channels.	140
9.1.	The measured branching fraction of each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels. The current world average branching fractions are given for comparison [5].	141
9.2.	The comparison of systematic errors for the D^{**} of the selected $D^{(*)}$ meson decays with and without application of the D^{**} MVA.	145
10.1.	The measured branching fraction of each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels.	150

List of Tables

- E.1. The yield for the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values). 165
- E.2. The percentage drop in yield for the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^+ \tau^- \bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. The percentage drop in yields shown are calculated from the non-rounded yields in Table E.1. . 165
- E.3. The yield for the $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^{*0} \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values). 166
- E.4. The percentage drop in yield for the $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^{*0} \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. The percentage drop in yields shown are calculated from the non-rounded yields in Table E.3. . 166
- E.5. The yield for the $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^0 \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values). 167
- E.6. The percentage drop in yield for the $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^0 \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. The percentage drop in yields shown are calculated from the non-rounded yields in Table E.5. . 167

List of Tables

H.1.	The fraction of each $D^{(*)}$ decay channel present in $\bar{B}^0 \rightarrow D^{(*)+} \ell^- \bar{\nu}_\ell$ signal MC, and the associated number of signal side tracks per event.	177
H.2.	The fraction of each $D^{(*)}$ decay channel present in $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ signal MC, and the associated number of signal side tracks per event.	178

Author's Contribution

Belle II is an international collaboration consisting of over 1000 members. Therefore the hardware and software tools used for the detector construction, operation, simulation and data collection are the culmination of many thousands of hours of person-power and expertise, spanning decades. This analysis uses publicly available and Belle II specific software to manipulate the data collected by the Belle II detector.

The study of the $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ decay is performed by many analysts within and outside of Belle II. The specific study of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ using Full Event Interpretation (FEI) Hadronic Tagging is ongoing at Belle II, but the refinements of the analysis presented in this thesis were worked on by a group of analysts including: Doctor Racha Cheaib, Professor Florian Bernlochner, William Suttcliffe, Michael Eliachevitch, Kyle Amerie and Maximillian Graf. With these people, the author contributed to the first $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ measurement published by Belle II at the International Conference of High Energy Physics 2020 [1]. This analysis is also contributing towards a current analysis producing Belle II's first measurement of $R(D^*)$ being worked on by the author, Kazuki Kojima, Qi-Dong Zhou and Doctor Racha Cheaib. The author's main contribution of this analysis is the optimization of the signal selection and of a multivariate analysis to suppress D^{**} backgrounds. These members have all contributed towards the development of the analysis framework and the validation of the analysis methods used.

The $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay as the denominator of the ratio $R(D^{(*)}) \equiv \mathcal{B}(B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau) / \mathcal{B}(B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell)$, with both a neutral and a charged B meson decaying to multiple $D^{(*)}$ decay channels, has not been validated before for the Belle II experiment. Thus, this is the contribution of the author to distinct contributions to knowledge. The author contributed to the development, improvement and validation of the Belle II hadronic FEI skim. The author examined the use of a two-dimensional cut for the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ branching fraction measurement. The author optimized the signal selection for $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ and $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ with consideration for the final goal measurement of $R(D)$ and $R(D^*)$. The author developed a D^{**} suppression multivariate analysis. The author measured the branching fraction of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ as the denominator of $R(D^*)$ and evaluated the

contributing systematics. Not included in this thesis is the author's contribution towards Belle II's first $R(D)$ and $R(D^*)$ measurement, predicted to be published in 2023.

Chapter 1: The author provides an overview of the study of the denominator of the $R(D)$ and $R(D^*)$ variable, including the motivation for the study.

Chapter 2: The author introduces the concepts of particle physics necessary to understand the context of this study.

Chapter 3: The author summarizes the Belle II experiment and the components used for collecting the data used in this thesis.

Chapter 4: The author summarizes the software tools used of in this study. The author contributed directly to the development and validation of the leptonic data reduction referred to as 'skims' for the Belle II collaboration including the hadronic FEI skim used in this thesis.

Chapter 5: The author describes the motivation for the measurement of the denominator of the $R(D)$ and $R(D^*)$ variable ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$) at Belle II, and discusses the novel aspects of the analysis.

Chapter 6: The author details the full analysis method. The author contributed directly to the selection of the exclusive decay modes, to the development and selections of the on-line and offline cuts, the best candidate selection and to the development, study and use of the MultiVariate Analysis (MVA). The author was one of the main authors of the paper "Measurement of the semileptonic $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ branching fraction with fully reconstructed B meson decays and 34.6 fb⁻¹ of Belle II data" [1]. The author also contributed to the paper^[i] "Identification of Beam Background Calorimeter Clusters and Fake Photons at Belle II Using BDT Classifiers" [2]. In this the author's contribution was the application and study of the effect of a separately developed beam background and fake photon suppression MVA on the MVA developed by the author in this thesis for the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay.

^[i]In review for publication.

Author's Contribution

Chapter 7: The author extracts the $R(D)$ and $R(D^*)$ denominators from the obtained fitted yields.

Chapter 8: The author evaluates the systematic uncertainties of the measurement of the $R(D)$ and $R(D^*)$ denominators.

Chapter 9: The author presents the final original results and discusses the future potential of the analysis. This analysis contributes towards the first Belle II measurements of $R(D)$ and $R(D^*)$, which shall be published in the future. The author compares the thesis result to the work in Reference [1], which was produced with co-authors: Doctor Racha Cheaib, Professor Florian Bernlochner, Doctor William Suttcliffe, Michael Eliachevitch, Kyle Amerie and Maximillian Graf.

Chapter 10: The author summarizes the method and results of the work presented in this thesis.

Appendices: The author supplies supplementary information that is not included in the main thesis text with the intention of a comprehensive yet succinct thesis.

Acknowledgements

First and foremost, to my supervisor, Professor Steven Robertson. Thank you for this opportunity, for your support and encouragement in the equity, diversity and inclusion work, and the environmental work beyond my research.

Secondly, to Doctor Racha Cheaib, thank you for the hours of hands-on training and help, for inspiring me and re-enforcing my love of analysis.

To the other ever changing members of my working group, for their support, friendships, and challenging questions that made me the researcher that I am today. In particular, Doctor Raynette van Tonder who helped navigate new concepts, and Professor Andreas Warburton, Andrea Fodor, and Trevor Shillington who helped understand software and physics conundrums many a time. To Professor Florian Bernlochner, William Suttcliffe, Michael Eliachevitch, Kyle Amerie and Maximillian Graf for the collaborative work on the branching fraction measurement of $B \rightarrow D\ell^-\bar{\nu}_\ell$. To Kazuki Kojima-san and Qi-Dong Zhou for their collaboration towards Belle II's first $R(D^*)$ measurement. To the other members of Working Group 1 and the Semitauonic group, who provided help and feedback throughout my PhD.

To my family; Mum, Dad, Paul and Rebecca, Grandma, Nanny Beryl, John Pooh, and my aunts, uncles, and cousins. For always watching over me, for your bottomless love, and for grounding me with daily video-calls and messages during the pandemic and beyond.

To Justin, for singing my praises throughout the toughest parts of this thesis. For loving, for understanding, for compromising and – undoubtedly – putting up with it all.

To Siobhan and Lisa, your daily messages meant the world. To my dear friends Bryony, Flo, Aurélie, Megan, Emilie, Soud, Lia, Jess, Yuna, Alicia, Jeremy, Claudia, Christina, Melissa, Chris, Emma, Ryan, Clare, Alice, Alana, Ariane, Joe; I wish I could name you all here, and I know I've missed people. If you're reading this, I likely mean you too. So to you, my friends, old and new, for being my chosen family. Every one of your messages of encouragement helped me in this, and got me one step closer to the finish line.

1. Introduction

The Standard Model (SM) of particle physics has been developed for over half a century into a relatively robust model of the way the universe works: through elementary particles and the interactions between them. However, this model contains assumptions and misses components, falling short of a truly unified theory. There are still aspects of the SM that do not quite encompass all of the features of the Universe that we observe today. With the recent discovery of non-zero mass neutrinos, the SM has 25 a priori unknown free parameters that need to be determined by experiment. This is where high energy collider experiments come in, to reinforce the SM through the determination of these free parameters. With these parameter inputs, there are various methods available to test the SM.

Within the SM, there is a universal coupling to the fermion generations called the electroweak gauge coupling. Between leptons, a flavour symmetry is created that can only be broken by the SM interactions that generate the charged lepton masses (called Higgs Yukawa interactions). Thus, the SM predicts lepton flavour universality (LFU). LFU has recently sparked interest in the particle physics community due to experimental hints at physics beyond the SM (BSM) due to contention with theoretical predictions. This thesis considers one particular facet of the LFU sector that is currently in contention with experimental measurements: the observables called $R(D^{(*)})$.

B mesons are particles called hadrons that consist of a quark-antiquark pair, in which one of the two is an (anti-) b quark^[i]. This is introduced in more detail in Chapter 2. B mesons can decay semileptonically i.e. into a combination of hadrons and leptons. These semileptonic decays can be used as exceptional tests of the SM due to their abundance and

^[i]Charge conjugate states are implied throughout this thesis, unless otherwise specified.

1. Introduction

their decay mechanisms.

The value of $R(D^{(*)})$ is the ratio between the rate of decay or “branching fraction” of two B meson decays, $B \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ and $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ ^[ii]. $R(D^{(*)})$ currently shows a discrepancy with the SM, indicating potential lepton flavour universality violation (LFUV). However, the significance of the discrepancy between experimental measurements and theoretical predictions is insufficient to conclude that non-SM contributions are present. An observation of LFUV would confirm physics BSM and could ultimately provide insight into a whole new sector of particle physics. It could also contribute to the discovery of dark matter candidates or additional necessary contributions to the origin of the matter-antimatter asymmetry in the early universe. The results of this study also contribute towards further future measurements of these decays and SM tests such as $|V_{cb}|$, and will eventually contribute to measurements of four free parameters (the three Cabibbo-Kobayashi-Maskawa (CKM) mixing angles and the CKM Charge-Parity (CP) violation phase) of the SM through the optimization of signal selection criteria and analysis techniques.

The Belle II experiment [3] is a unique environment in which to study these semileptonic B decays. The Belle II experiment is based in Tsukuba, Japan and consists of the SuperKEKB accelerator and Belle II detector. Here, 7 GeV electrons e^- and 4 GeV positrons e^+ are collided^[iii] at a centre of mass energy of 10.580 GeV. The e^+e^- produces an Υ meson, which is a bottomium state and thus consists of a bottom quark and its antiparticle. At this specific energy the Υ meson is excited to the $\Upsilon(4S)$ resonance, which is the first $b\bar{b}$ resonance with enough energy to decay into a $B\bar{B}$ meson pair. This pair is produced almost at rest in the $\Upsilon(4S)$ centre of mass (CM) frame. Due to Belle II’s large production rate of $B\bar{B}$ pairs, it is given the title “ B -factory”. Chapter 3 considers the Belle II experiment in detail, including a breakdown of the sub-detector components.

Alongside the unique environment at Belle II is the highly advanced software used in all analyses. Chapter 4 highlights the various software packages specific to the analysis in

^[ii] Throughout this thesis, l will represent the electron e , muon μ or tau lepton τ , whereas ℓ will only represent e and μ .

^[iii] Throughout this thesis the convention of “natural units” is used, in which $\hbar = c = 1$ unless explicitly stated.

1. Introduction

this thesis. The **Belle II Analysis Software Framework**, henceforth referred to as **basf2** is the Belle II specific software used for simulation and reconstruction of the measured final state particles in the detector. Due to the production of a $B\bar{B}$ pair at rest in the $\Upsilon(4S)$ CM frame, full reconstruction of one B meson allows the constraint of the flavour and momentum of the other B meson in the decay. A machine learning algorithm called Full Event Interpretation (FEI) is used to reconstruct and tag one of the B mesons (termed the B_{tag}). This is then combined with the signal decay of the other B meson (B_{sig}) to reconstruct the entire $\Upsilon(4S)$ decay. Figure 1.1 indicates a possible collision event, in which the particles are reconstructed from measured final state particles. The **TopoAna** topological mapping software is then used to compare the original generated collision events to the reconstructed collision events, and develop the analysis.

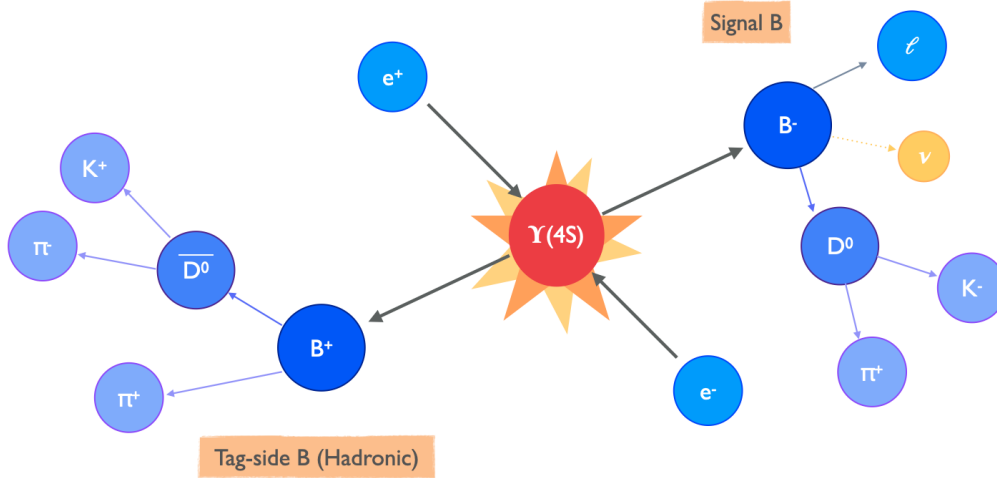


Figure 1.1.: A depiction of a possible collision event with the decay of an $\Upsilon(4S)$ into a B_{tag} and a B_{sig} . The B_{sig} decays into $B \rightarrow D\ell^-\bar{\nu}_\ell$.

This thesis aims to optimize the analysis of the denominators of $R(D)$ and $R(D^*)$, and to procure measurements of the relevant branching fractions. Initially, this analysis contributed to the first published Belle II results of hadronically tagged $B^0 \rightarrow D^{*+}[D^0\pi^+]\ell^-\bar{\nu}$ branching fraction where $D^0 \rightarrow K^-\pi^+$ [1]. This thesis expands on this measurement by including more exclusive decays of the D^* and D mesons in the B^0 decay, and further includes the charged B meson, B^+ (listed later in Table 5.1). This is the first measurement of its kind at the Belle II experiment.

1. Introduction

In Chapter 5, the motivation behind measuring the denominators of $R(D)$ and $R(D^*)$ at Belle II is discussed, with the combination of features that make this analysis unique. Of note is the Boosted Decision Tree (BDT) multi-variate analysis (MVA) that is developed to reduce specific backgrounds that peak in the signal region of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays. The analysis methods and selections optimized in this study are discussed in Chapter 6.

The extraction of the $R(D)$ and $R(D^*)$ denominator measurements is performed in Chapter 7, with in-depth analysis of the systematic uncertainties presented in Chapter 8.

Finally, in Chapter 9, the results of this analysis are compared to the first published results, and the success of the developed MVA is discussed alongside the future of this analysis.

2. Theory

This chapter introduces the key concepts of the SM of particle physics necessary for comprehension of the motivation behind the measurements of the semileptonic decays of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$. The current measurements of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ will be discussed, leading to the status of the $R(D)$ and $R(D^*)$ test of the SM for LFU.

2.1. The Standard Model of Particle Physics

The SM is the governing theoretical framework of particle physics. It is a consequence of the fundamental particles and their interactions. Gravity is not explicitly included in the SM as its effects are negligible in the interactions between the small particle masses at the distances and energies considered.

The fundamental particles consist of particles that mediate fields called bosons and matter particles called fermions, of integer and half integer spin, respectively. Figure 2.1 displays these elementary particles. There are the fermions - quarks and leptons - and the bosons - gluons, photons, Z bosons, W bosons and the recently discovered Higgs boson [4]. Within the SM there are also hadrons, which are composite particles of multiple quarks. Hadrons can be categorized into mesons and baryons which are composed of two and three quarks, respectively.

Gauge bosons are the force carrying vector particles that are responsible for the interactions between SM particles. Thus the SM is a gauge theory, based on local gauge symmetries. The SM is described by the gauge groups $SU(3)_c \times SU(2)_L \times U(1)_Y$. These three gauge groups represent the strong and electroweak forces of the SM. Under electroweak symmetry breaking effected by the Higgs mechanism [5], the electroweak force

Standard Model of Elementary Particles

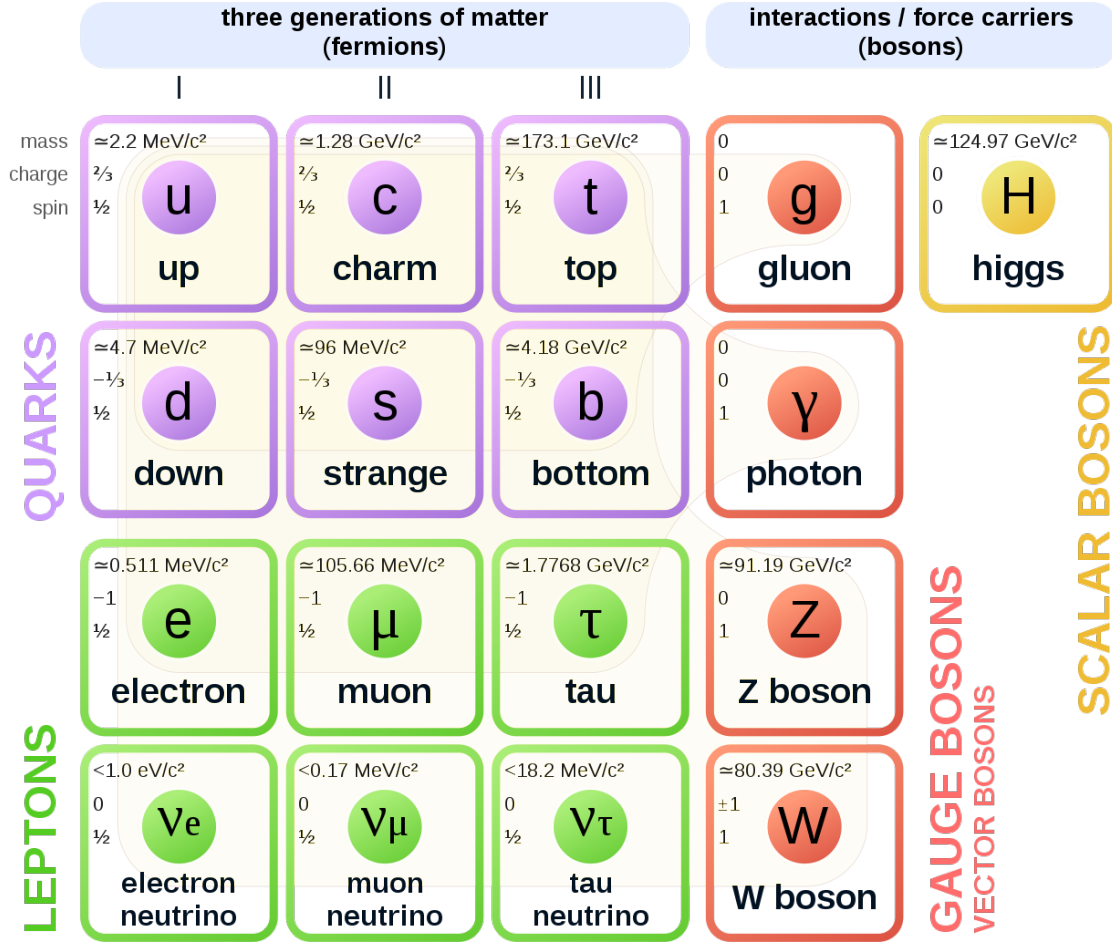


Figure 2.1.: The SM of elementary particles. There are twelve fermions and five gauge bosons. For each fermion there is a corresponding anti-particle that is not depicted. The highlighting loops in the background indicate which gauge bosons couple to which fermions. This infographic is correct as of the publication of the 2019 Particle Data Group report [5]. Image sourced from [6].

2. Theory

becomes the electromagnetic force and the weak force. The electromagnetic force $U(1)_{em}$ occurs between charge-carrying particles and is mediated by the photon γ . The weak force $SU(2)_L$ occurs between fermions and is mediated with the uniquely massive $W^\pm$ or Z bosons. The weak force is discussed further in Section 2.1.1. The strong force $SU(3)_C$ occurs only between quarks q and gluons g . $SU(3)_C$ symmetry manifests a conserved charge, a quantum property called ‘colour charge’. There are three possible charges in this property, which are distinguished and labelled for ease as ‘red’, ‘blue’ and ‘green’ colour charges. These can be combined into baryons or mesons to create either a colour singlet or net zero colour charge (a ‘colourless’ state). Figure 2.2 shows the possible combinations of colour for mesons and baryons. Furthermore, the addition of a colour and its corresponding anti-

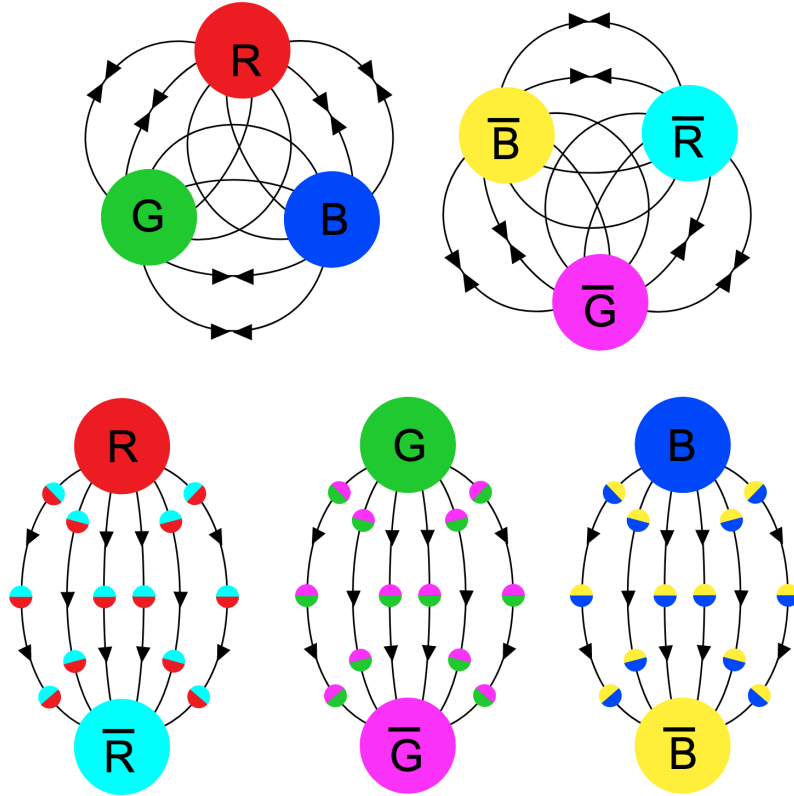


Figure 2.2.: The colour charge combinations of baryons and mesons. Image obtained from Reference [7].

colour (such as red and anti-red) also results in a colourless state. Each (anti-)quark can carry a (anti-)colour charge. A colourless state is required for hadrons, however gluons are required to carry an overall (anti-)colour charge, and therefore cannot be colourless. As a

2. Theory

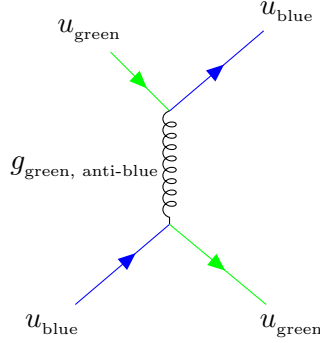


Figure 2.3.: A Feynman diagram depicting two up quarks exchanging a gluon g .

direct result of this, gluons can not only mediate the strong force, but can participate in it too. Figure 2.3 depicts an example strong interaction with a mediating gluon carrying an overall colour charge.

Finally, there is the Higgs boson. This particle is a scalar boson and is responsible for all massive fundamental particles acquiring mass.

The fermions of the SM are leptons and quarks. The leptons are divided into three generations: the electronic, muonic and tauonic leptons. Each generation consists of an electrically charged lepton l and an associated neutrino ν ; each of which have a corresponding anti-particle. The charged leptons experience both the electromagnetic and weak forces, whereas neutrinos are neutrally charged and thus only experience the weak force. Neutrinos were long assumed to be massless particles until evidence of neutrino oscillation between generations in 1998 indicated they definitely have a non-zero mass [8].

Each lepton generation is distinguished by ‘flavour’, which is a property or ‘quantum number’ that carries a value relating to the generation and particle. Therefore there are six lepton flavour quantum numbers: electron (e) number, muon (μ) number, tau (τ) number and their analogous neutrino flavour numbers.

Likewise the SM quarks are composed of three generations, each of an up-type and down-type quark. The first generation consists of the up u and down d quark; the second of the charm c and strange s quark; and the third of the top t and bottom b quark. Each quark has an electric charge, a colour charge, a spin value of $\frac{1}{2}$, and a corresponding anti-quark with opposite charge, but same mass and spin. Quarks experience all of the three fundamental forces of the SM.

2. Theory

The relatively recent experimental discoveries of particles such as the top quark (1995) [9], the tau neutrino (2000) [10], and the Higgs (2012) [4], were predicted by the SM many years before these particles were observed. This indicates the reliability and strength of the SM. Currently within the SM there are no further theorized particles left to detect experimentally. Today, with technological advancement, the SM is scrutinized with ever more precise measurements. The Heavy Flavour Averaging Group (HFLAV) [11] and the Particle Data Group (PDG) [5] collect, update, collate, average and publish experimental results from all over the world, regularly.

However, the SM has some shortcomings that show that it is not a complete theory of everything. Some of these theoretical deficiencies could be unrelated to particle physics, such as dark matter – there is no explanation of the mass-energy distribution of the universe. Yet there are other problems that are intrinsic to the SM that indicate that it is not fundamental. These shortcomings of the SM include necessary components such as: the theory for gravitational interaction – there is no gauge boson responsible for gravitation; and an explanation for the matter-antimatter asymmetry of the initial conditions of the universe – there is no component that perfectly describes our matter dominated universe. There are also questions as to *why* things are the way they are. For example, it is still unknown as to *why* all fermions appear in three identical copies except for their mass.

2.1.1. The Cabibbo-Kobayashi-Maskawa matrix

This section discusses the ability to measure 4 of the free parameters of the SM, the CKM mixing angles θ_{12} , θ_{23} , θ_{13} and the CKM CP violation phase δ .

In the SM the weak force alone has the ability to change a quark from one flavour to another. Flavour changing charged-current (FCCC) weak interactions, mediated by the $W^\pm$ boson, are the only processes which violate flavour and CP conservation in the SM. In these interactions quarks and leptons weak eigenstates are different from their mass eigenstates. This means that if a quark couples to a weak boson, the quark is not in its mass eigenstate but in a superposition of the three generations of quark mass eigenstates.

A FCCC interaction can occur through both same-generational and cross-generational mixing of mass eigenstates. The strength of these decays varies by the type of flavours that

2. Theory

are involved in each interaction, and this can be described using a unitary 3×3 matrix called the CKM matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.1)$$

Each of these flavour changes are represented in the CKM matrix by an element V_{ij} that has a subscript indicating the kind of transition, i.e. $|V_{ub}|$ represents a quark flavour change of $u \rightarrow b$ or $b \rightarrow u$. The left-hand side of Equation 2.1 shows the weak eigenstates of down-type quarks while the furthest right column vector on the right-hand side displays the corresponding mass eigenstates.

The CKM matrix can be parametrized by 4 free parameters: 3 mixing angles (θ_{12} , θ_{23} , θ_{13}) and 1 CP violating phase δ . The relations between the CKM matrix and these parameters are shown explicitly in Appendix A. The mixing angles can be constrained by a CKM unitary triangle, an example of which is shown in Figure 2.4. Unitary triangle elements can be measured by experiment to then determine these free parameters. CP violation and the SM can be tested through measurement of the CKM matrix due to its contributions of CP violating parameters.

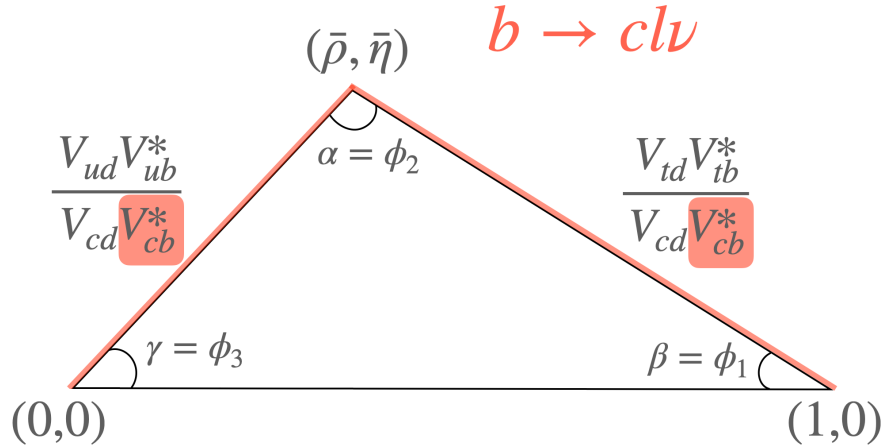


Figure 2.4.: A unitary triangle of the CKM matrix, highlighting the CKM element relevant for this analysis, $|V_{cb}|$. Here the * indicates the conjugate of that element.

2. Theory

The current measured values of the CKM matrix values are [5]:

$$V_{\text{CKM}} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.0011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix}. \quad (2.2)$$

A hierarchy in the diagonal elements is observed in these matrix elements. This mysterious hierarchy could be indicative of a hidden mechanism at a higher energy scale, hinting at physics BSM, for example, a new flavour symmetry.

2.1.2. Effective field theories

When considering a single particle, the calculation of a decay rate is relatively straightforward using Fermi's Golden Rule [12]. Intrinsic to the Golden Rule is the matrix element $\mathcal{M}$, which is the coupling term for an initial state transitioning to a final state. For a free b quark decaying to a c quark this can be written as:

$$\mathcal{M} = -i \frac{G_F}{\sqrt{2}} |V_{cb}| L^\mu H_\mu \quad (2.3)$$

where G_F is the Fermi constant, L^μ is the leptonic current $L^\mu = \bar{u}_l \gamma^\mu (1 - \gamma_5) v_\nu$ and $H_\mu = \bar{c}_l \gamma^\mu (1 - \gamma_5) b$ is the quark current (where γ_5 and γ^μ are the Dirac gamma matrices).

These calculations quickly increase in complexity when considering real, confined quarks due to the need to consider quark-quark interactions. As discussed in Section 2.1, quark-quark interactions are also mediated by the strong force. These cannot be calculated from first principles due to the strong force's highly non-linear nature and its large coupling constant at lower energies, however it is possible to model particles computationally using Effective Field Theories (EFT). EFTs can use just one of the many length scales that exist in quantum field theories to simplify calculations. For example, they can allow for perturbative calculations of interactions with energy scales that are large compared to the quantum chromodynamics (QCD) energy scale $\Lambda_{\text{QCD}} \approx \mathcal{O}(100 \text{ MeV})$. This is beneficial in the consideration of b and c quarks (due to the B meson) because their masses are large in comparison to Λ_{QCD} , thus the size of relevant strong force interactions between them are

2. Theory

necessarily large.

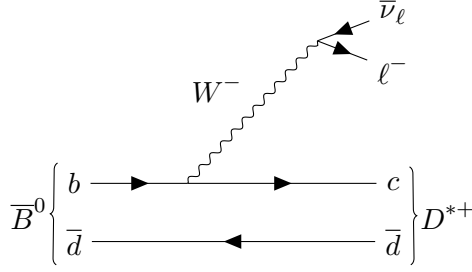


Figure 2.5.: A Feynman diagram depicting the decay $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$, with quark composition shown. Here ℓ represents an electron e or muon μ .

Figure 2.5 depicts an example decay involving a B meson to a D^* meson and two other products that are a lepton ℓ (e or μ) and corresponding neutrino ν_ℓ . This is the equivalent to the quark level process of $b \rightarrow c \ell \bar{\nu}_\ell$. In this diagram the hadronic current is modified to:

$$H^\mu = \langle D | \bar{c}_l \gamma^\mu (1 - \gamma_5) b | B \rangle = \langle D | V^\mu | B \rangle - \langle D | A^\mu | B \rangle, \quad (2.4)$$

where $V^\mu = \bar{c} \gamma^\mu b$ is a vector current and $A^\mu = \bar{c} \gamma^\mu \gamma_5 b$ is an axial current. As this matrix calculation cannot be calculated with ease, it can instead be expressed using Form Factors (FF). FFs are scalar functions of the momentum transfer whose values impact the branching fraction of a decay. They describe the properties of a particle interaction without including the explicit physics behind it. They must either be determined experimentally or computed using lattice QCD [13]. EFTs and their resulting FFs allow precise estimations of branching fractions that would be too difficult to calculate from first principles. In this example case the FFs are functions of the leptonic momentum transfer q^2 , and we can express the hadronic current as:

$$\langle D(p') | V^\mu | B(p) \rangle = f_+ (p + p')^\mu + f_- (p - p')^\mu, \quad (2.5)$$

where f_+ and f_- are the FFs, and

$$\langle D(p') | A^\mu | B(p) \rangle = 0, \quad (2.6)$$

2. Theory

such that

$$H^\mu = f_+ (p + p')^\mu + f_- (p - p')^\mu, \quad (2.7)$$

where p and p' are the momenta of the B and D meson, respectively.

To model the FFs in this analysis, we must use an EFT that considers the interactions between heavy Q and light q quarks. Heavy quarks are defined as quarks with a mass m_Q that is much larger than Λ_{QCD} . Here Q is a b or c quark and q is a u or d quark. Heavy Quark Effective Theory (HQET) is one such EFT. It approximates m_Q as infinitely large, imposes heavy quark flavour symmetry (interactions are unchanged under flavour exchange) and spin symmetry (interactions are unchanged under arbitrary spin transformations). This gives the effect of allowing the strong force interactions within the decay to be considered negligible with respect to those of the weak decay. HQET is crucial in the undertaking of the measurements of semileptonic decays including that of the CKM matrix elements $|V_{cb}|$ and $|V_{ub}|$. It has thus become an essential model for B meson physics.

2.2. Semileptonic B meson decays

The B meson is a meson of quark composition $\bar{b}q$ where q represents u for a charged B^+ , d for a neutral B^0 , s for a strange B_s^0 or c for a charmed B_c^+ .

Hadronic decays are particle decays that contain solely hadronic products i.e. $B \rightarrow D\pi$; leptonic decays only contain leptonic products i.e. $B \rightarrow \tau\nu$; semileptonic decays are such that there is at least one lepton produced i.e. $B \rightarrow X\tau\nu$, where X is a meson, as in Figure 2.5.

The B meson can decay leptonically or semileptonically in the SM via the weak interaction, and can be modelled well using EFTs. In contrast, hadronic B decays are much harder to describe. Simply due to the fact that there are more quarks in a hadronic decay, there are more QCD interactions to model, in particular between initial and final state quarks.

As the W boson mediator of the weak interaction is much heavier in comparison to all

2. Theory

quarks except the top quark, all CKM matrix elements are accessible through (semi)leptonic decays^[i] except those of $|V_{tq}|$.

Leptonic B processes are simpler to calculate theoretically as they occur via the annihilation of the b quarks with the light quark in the B meson. However, they also have a low production rate and are therefore much harder to measure with precision in experiments. Semileptonic B processes are more complex due to their dependence on multiple FFs and the uncertainties that come with them. However, the benefits of semileptonic decays outweigh the limitations of the FFs due to their high branching ratios compared to that of leptonic decays and the fact that one can largely cancel out the uncertainties of the form factors by measuring the ratio of two branching fractions (Section 2.2.3).

Leptonic and semileptonic decays that involve the τ lepton allow investigation into the third lepton generation of particle physics. As this generation is not yet measured as precisely as the first and second generation, these decays can provide further tests of the SM[5]. This is discussed in more detail in Section 2.2.2. These decays can also contribute to the measurements of CP asymmetries in B decays. Any possible new physics (NP) contributions to these decays are expected to be highly suppressed [14]. The masses of μ and e are negligible in comparison to the B and $D^{(*)}$ mesons in a semileptonic decay, whereas the mass of the τ is the same order of magnitude as them [5]. This differentiates the τ decay significantly from the other two leptonic decays. At this stage we shall split the definition of semileptonic decays into that of semileptonic decays (decays including the e or μ such as $B \rightarrow Xe\nu_e$ or $B \rightarrow X\mu\nu_\mu$) and semitauonic decays (decays including a τ lepton such as $B \rightarrow X\tau\nu_\tau$) for clarity in the future sections of this thesis. Here X represents a hadronic system consisting of one or more mesons with a c or a u quark and light spectator quark, a u or a d .

As mentioned in Section 2.1.1, the precision measurements of $|V_{cb}|$ and $|V_{ub}|$ are necessary to validate the SM. Such measurements can be made through charged-current semileptonic or semitauonic transitions of B mesons. These decays are predominantly tree-level W boson mediated decays, and are expected to provide excellent measurements of these fundamental parameters of the SM, including that of the masses of the b and c quarks. As semileptonic

^[i]Not all are accessible via B decays.

2. Theory

and semitauonic decays predominantly decay to the charmed version of the meson, the two higher priority decays are the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ and $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ decays, where $D^{(*)}$ is either a D or D^* ($D^*(2007)^0$ and $D^*(2010)^+$) meson and where ℓ refers to the leptons e or μ . In this thesis a distinction shall be made between $l = e, \mu, \tau$ and $\ell = e, \mu$ to indicate that $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ does not include semitauonic $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ decays. The large samples of B meson data provided by B factories in the recent history of the field have enabled many precision studies of these decays.

2.2.1. Search for New Physics beyond the Standard Model

There are discrepancies between SM theory and experiment which hint at possible physics beyond the SM. In particular, there are fundamental questions about the flavour sector of quarks and leptons that require investigation. For example LFU is an accidental symmetry of the SM where electroweak gauge bosons couple with equal strength to the three different generations of leptons. Leptons alone have a flavour symmetry within the SM that can be broken by Higgs Yukawa interactions. This means that the SM predicts that LFU exists such that symmetries occur between lepton flavours that are only broken by the charged lepton masses. LFU has recently sparked interest in the particle physics community with growing experimental hints at physics BSM due to contention of measurements with theory [15]. These discrepancies are naturally limited by the uncertainties involved in theoretical calculations and the precision of experimental measurements, however with each improvement in calculation, data sample and analysis method, these uncertainties are reduced in size.

Whether or not there is physics BSM is yet to be confirmed. No concrete evidence of NP has been measured directly or indirectly to date, but various NP scenarios have been postulated to explain that which the SM cannot.

2.2.2. The potential for NP in $B \rightarrow D^*\tau^- \bar{\nu}_\tau$

In high-energy collider experiments such as Muon g-2, LHCb and Belle II, there have been interesting hints at the possibility of NP in the form of three particular flavour observables. Firstly, the current world average of the measurements of $B \rightarrow D^*\tau^- \bar{\nu}_\tau$ are in tension with

2. Theory

the SM by 3σ [5]. Secondly, the branching fraction measurements of $B \rightarrow K\ell^+\ell^-$ are in 3σ tension with the SM [16, 17, 18]. Finally, the anomalous magnetic moment of the muon was recently measured in experiment and found to be 4.2σ away from SM prediction [19]. Each of these three deviations from the SM are cases of LFUV. If these three LFUV measurements are related, there may also be new physics effects present in $B \rightarrow D^*\mu\bar{\nu}_\mu$ decays [20].

This thesis considers one facet of the LFUV sector that is currently in contention with experimental measurements: an observable related to $b \rightarrow c\ell\bar{\nu}_\ell$ and $b \rightarrow c\tau\bar{\nu}_\tau$ called $R(D)$ and $R(D^*)$, discussed in Section 2.2.3.

In the search for the possible causes of the measured deviations of experiment from the SM predictions, extensions to the SM in the form of new mediator particles have been put forward. In the quest for providing excellent measurements of fundamental parameters of the SM, semitauonic decays could provide insight into BSM physics as, in addition to the W boson mediator, there could be other mediator particles that could contribute to the $B \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ branching fraction.

In particular, extended Higgs sector models such as those with charged Higgs bosons would potentially be detectable. One such model is the Type II Two Higgs Doublet Model (2HDM) [21], which is the minimal extension of the Higgs sector in the SM from one Higgs doublet to two. This model consists of 5 Higgs bosons: the scalar Higgs h^0 and H^0 , the pseudo-scalar Higgs A^0 , and the charged Higgs $H^\pm$. The presence of the charged Higgs as a mediator particle would modify semitauonic differential distributions like the momentum transfer to the lepton pair $q^2 = (p_\tau + p_\nu)^2 = (p_B - p_{D^{(*)}})^2$, where p represents each labelled particle's momentum. Though there has been a gradual increase of experimental constraints imposed on the 2HDM, the recent muon g-2 result [19] has caused a resurgence in interest in the model [22]. A discovery of a charged Higgs contribution to the $B \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ branching fraction depicted in Figure 2.6 would be evidence of physics BSM.

In terms of the benefit of measuring $B \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$, not only would it be sensitive to the charged Higgs contributions at tree level, it also decays at a rate 100 times more than the other NP sensitive tauonic decay $B^- \rightarrow \tau^-\bar{\nu}_\tau$. This and the fact that in $B^- \rightarrow \tau^-\bar{\nu}_\tau$

2. Theory

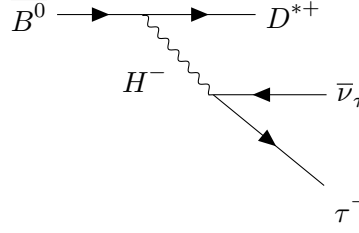


Figure 2.6.: A modified Feynman diagram of the SM $\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau$ decay with a charged Higgs as the mediator boson.

decays there are significant backgrounds and a large uncertainty on $|V_{ub}|$ make it an unlikely prospect for measurements of NP for the time being. Furthermore, the additional $D^{(*)}$ meson in the semitauonic decay helps to constrain signal events and to differentiate them from background, as will be shown in Chapter 6.

Another possible extension to the SM is the leptoquark, of which many models have been proposed [23, 24, 25, 26]. It is also possible that the leptoquark model is the only NP model that seems to be compatible with the discrepancies observed [27]. Leptoquarks are elementary particles that carry both a quantum quark number and lepton number, and therefore can couple to both quarks and leptons. Leptoquarks are assumed to be invariant under SM gauge transformations. It is also assumed that the leptoquark can mediate $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ without introducing a right-handed neutrino, which is therefore more compatible with the SM. Figure 2.7 shows the possible modified Feynman diagram of SM $\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau$ with the leptoquark L as the mediator particle. Note that here the L can violate LFU since the neutrino flavour is not necessarily related to the flavour of the τ .

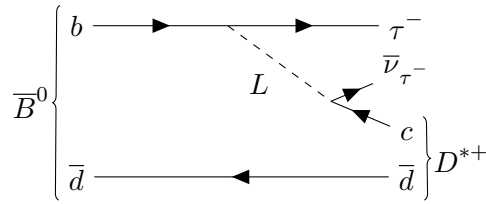


Figure 2.7.: A modified Feynman diagram of the SM $\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau$ decay with a leptoquark as the mediator boson.

2. Theory

2.2.3. $R(D)$ and $R(D^*)$

The measurement of the branching ratios of a decay of interest is an important step in the understanding of the SM. However, as discussed in Section 2.1.2, the theoretical calculations of a semileptonic branching fraction rapidly gain complexity with the presence of quark-quark interactions. A ratio between two measured branching fractions where only the lepton flavour varies has the benefit of cancelling or largely cancelling some uncertainties. The largest uncertainties in the measurement of branching fractions of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ are the $|V_{cb}|$ contribution and the hadronic matrix elements. Thus, it is advantageous to instead measure the ratio of the branching fractions of $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ to $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ as:

$$R(D^*) \equiv \frac{\mathcal{B}(B \rightarrow D^* \tau^- \bar{\nu}_\tau)}{\mathcal{B}(B \rightarrow D^* \ell^- \bar{\nu}_\ell)} \quad (2.8)$$

and

$$R(D) \equiv \frac{\mathcal{B}(B \rightarrow D \tau^- \bar{\nu}_\tau)}{\mathcal{B}(B \rightarrow D \ell^- \bar{\nu}_\ell)}. \quad (2.9)$$

Here, ℓ refers to the electron and muon only. $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ is often known as the ‘signal mode’ and $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ is known as the ‘normalisation mode’. The $R(D)$ and $R(D^*)$ ratios are independent of $|V_{cb}|$ and the hadronic matrix elements as they cancel in the ratio, therefore the theory of these variables are very precise.

In a ratio such as this, where only the lepton differs, one might expect that due to LFU the ratio should be equal to unity. This ratio is in fact subject to the non-negligible τ mass relative to the rest of the decay, thus the values of $R(D^*)$ are expected to be a fraction of unity. The ratios of branching fractions that only include the e and μ leptons such as in $b \rightarrow c \ell \nu_\ell$ and $b \rightarrow s \ell^+ \ell^-$ decays are indeed very close to unity, theoretically [28], but show experimental evidence of deviation from that, hinting at LFUV.

One difficulty in the measurement of $R(D)$ specifically is that it has a large cross-feed contribution. Cross-feed occurs when, for example, two signal channels are reconstructed, and each appear as a background to the other channel. In the case of $R(D)$, a D^* meson

2. Theory

that is not correctly reconstructed could be incorrectly reconstructed as a D meson. This is referred to as “feed down”. Here the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ decays can easily contribute to the $B \rightarrow Dl^-\bar{\nu}_l$ signal sample due to the decay of the D^* to a D meson, therefore it is a large source of background for $R(D)$. The reverse case in $R(D^*)$, where D mesons could be misreconstructed as D^* mesons (“feed up”), is much less likely due to the requirements for D^* reconstruction, i.e. it is easier to miss a π or γ and mis-reconstruct a D meson than it is to find one with the correct properties to mis-reconstruct a D^* . Thus, the $R(D^*)$ measurement is expected to be the cleaner measurement.

2.2.4. Review of measurements of $\mathcal{B}(B \rightarrow D^{(*)}l^-\bar{\nu}_l)$ and $R(D)$ and $R(D^*)$

Now decommissioned experiments such as Belle and Babar [29] were B meson-focused, accelerator-based particle experiments called B factories, which were tremendously successful in their measurements of the free parameters of the SM, the aforementioned CKM matrix elements. They, alongside the still active experiments of Belle II and LHCb [30], contribute to the results and analysis techniques that are reported and used in this analysis.

There are two possible methods in which to study the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ decays: inclusive or exclusive methods. In an inclusive measurement, all possible decay modes are considered. In an exclusive measurement, one specific decay chain is selected for a study, such as a final state that involves a specific meson (i.e. $D^{(*)}$). There will be an emphasis on exclusive analyses throughout this thesis as an exclusive set of B meson decays are utilized.

The first exclusive measurement of $B \rightarrow D\tau^-\bar{\nu}_\tau$ was made in 2007 by the Belle collaboration with a 5.2σ significance [31]. Until then only inclusive and semi-inclusive measurements by Large Electron-Positron (LEP) experiments [32, 33, 34, 35] contributed to an average measurement of the branching fraction of $\mathcal{B}(B \rightarrow X\tau\nu_\tau) = (2.48 \pm 0.26)\%$, where X is a hadron. The current average measurements of the branching fractions for the various exclusive $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ decays are available in Table 2.1.

Today, there have been multiple measurements of $R(D)$ and $R(D^*)$ and numerous revisions and proposals for the SM prediction of the values of $R(D)$ and $R(D^*)$. The current measurements of $R(D)$ and $R(D^*)$ are in contention with the SM and show deviations of $\sim 1 - 3\sigma$ as listed in Table 2.2. The measurements that contributed towards these averages

2. Theory

Decay	Experimentally Measured Fraction (Γ_i/Γ)
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	$(2.35 \pm 0.09)\%$
$B^- \rightarrow D^0 \tau^- \bar{\nu}_\tau$	$(0.77 \pm 0.25)\%$
$B^- \rightarrow D^*(2007)^0 \ell^- \bar{\nu}_\ell$	$(5.66 \pm 0.22)\%$
$B^- \rightarrow D^*(2007)^0 \tau^- \bar{\nu}_\tau$	$(1.88 \pm 0.20)\%$
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	$(2.31 \pm 0.10)\%$
$\bar{B}^0 \rightarrow D^+ \tau^- \bar{\nu}_\tau$	$(1.08 \pm 0.23)\%$
$\bar{B}^0 \rightarrow D^*(2010)^+ \ell^- \bar{\nu}_\ell$	$(5.05 \pm 0.14)\%$
$\bar{B}^0 \rightarrow D^*(2010)^+ \tau^- \bar{\nu}_\tau$	$(1.57 \pm 0.09)\%$

Table 2.1.: The average of the measurements of each of the different $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays. All branching fractions are averages taken from the 2020 update of the PDG Review of Particle Physics [5].

LFU observable	Experimental Measurement	SM Theoretical Prediction	Deviation (σ)
R_D	$0.340 \pm 0.027 \pm 0.013$	0.299 ± 0.003	1.4
R_{D^*}	$0.295 \pm 0.011 \pm 0.008$	0.258 ± 0.005	2.5

Table 2.2.: A summary of the averages of the values measured/predicted for $R(D)$ and $R(D^*)$ reported by HFLAV, updated in May 2019 [11].

are shown in Figure 2.8. From these figures and Figure 2.9 it is clear that both $R(D)$ and $R(D^*)$ are in contention with the SM. However, due to the large statistical uncertainties in these measurements, the fact that they all exceed the predictions of the SM is not yet enough to confirm to a high degree of certainty that there are effects of NP present. More recently, LHCb presented an updated $R(D)$ and $R(D^*)$ measurement contributing little change to the world average [36]. Many analyses have looked for NP using specific BSM models; at present no NP have been observed leading to the ruling out of certain models and the constraint of others [37, 38, 39]. The average measurements of $R(D)$ and $R(D^*)$ remain intriguing.

2. Theory

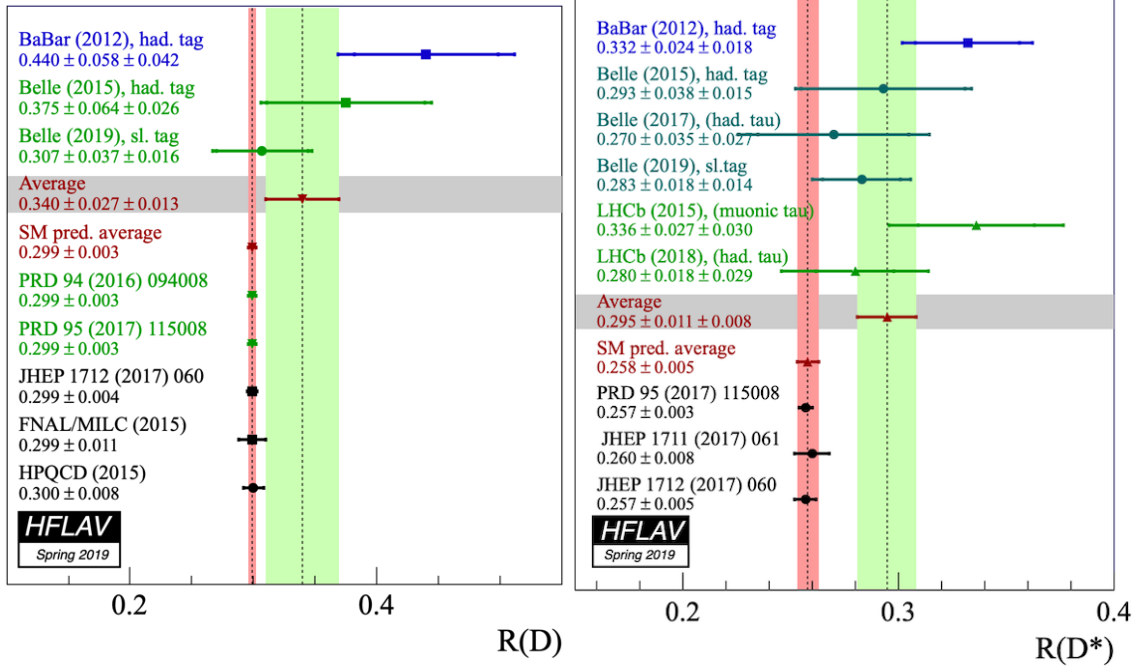


Figure 2.8.: The averages of the $R(D)$ (left) and $R(D^*)$ (right) of the experimental measurements and their averages compared to theoretical SM predictions. This was calculated and compiled by the HFLAV semileptonic group, last updated in May 2019 [11].

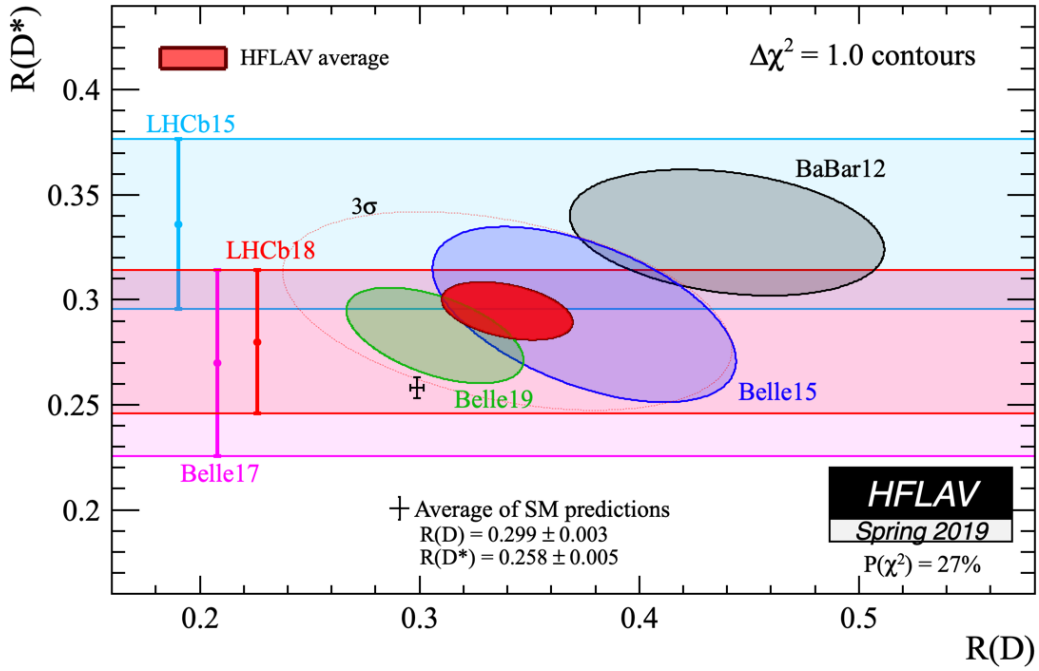


Figure 2.9.: The average SM prediction and experimental measurements of $R(D)$ against $R(D^*)$ by the HFLAV semileptonic group, last updated in May 2019 [11].

3. The Belle II Experiment

In 1997, the High Energy Accelerator Research Organization was founded in Tsukuba, Japan. This became the host of the Belle Experiment from 1998 to 2010 which, amongst decades of important studies and measurements of SM parameters and branching fractions, notably verified Kobayashi and Maskawa’s proposal that a single irreducible complex phase could contribute to the matter-antimatter asymmetry of the universe [40]. This led to their 2008 Nobel Prize in Physics [41]. After the cessation of Belle’s data collection, major upgrades were made to the whole Belle II Experiment [3]. In 2019, the first physics data was collected from the newly completed experiment: the SuperKEKB accelerator and the Belle II detector.

3.1. The SuperKEKB Accelerator

The SuperKEKB Accelerator is an approximately 3 km circumference electron-positron (e^-e^+) ringed particle accelerator based in Tsukuba, Japan. Bunches (collections) of e^- and e^+ are made to collide at a single location called the interaction region (IR). The bunches complete a full revolution of the rings every $10\mu s$, and are separated by 4 ns. The head-on collisions have a CM energy of $\sqrt{s}=10.580\text{ GeV}$, which corresponds to the mass of the $\Upsilon(4S)$ resonance. Figure 3.1 displays the CM energy of the $\Upsilon(4S)$ ‘on-resonance’ and other lower possible resonances of the Υ particle. The $\Upsilon(4S)$ resonance is important for the Belle II experiment in its unique ability to decay to a $B\bar{B}$ pair at a high rate of approximately 96%, and does so without any other decay products.

The $B\bar{B}$ mass is very close to the $\Upsilon(4S)$ energy, thus the two B mesons are produced nearly at rest in the CM frame. As the B meson only has a mean lifetime of order of

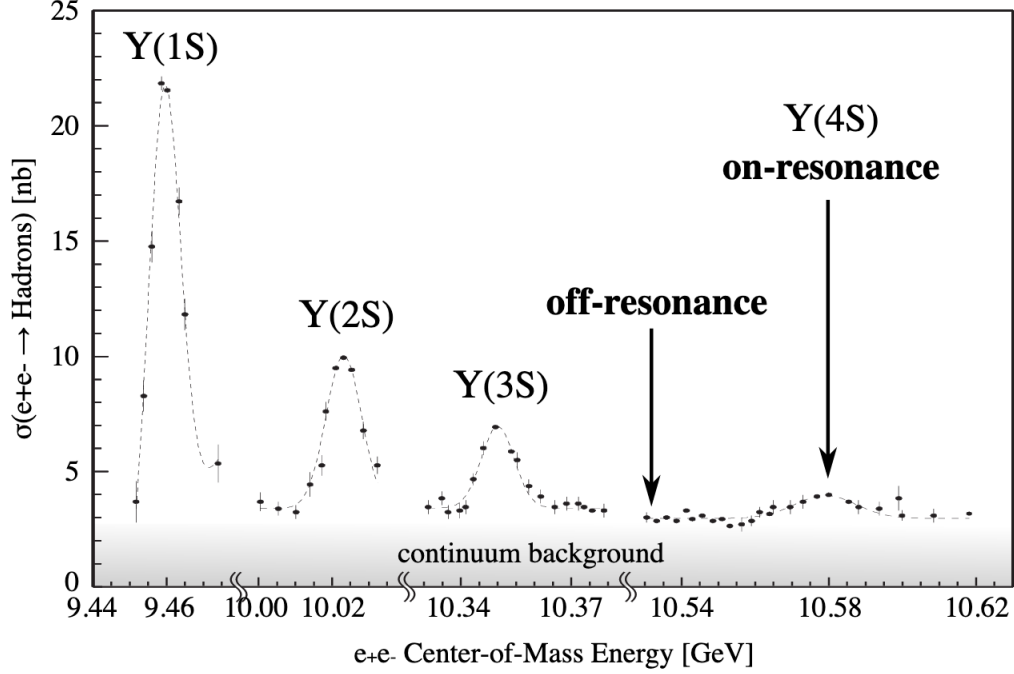


Figure 3.1.: The e^+e^- CM energy versus the cross-section e^+e^- to hadrons. The Υ resonances are indicated at each respective peak.

10^{-12} s, if the B mesons were to be produced at rest in the lab frame they would have too short a lifetime to be measured directly. Thus, the SuperKEKB accelerator is designed such that the energies of the e^- and e^+ beams are produced asymmetrically at 7 GeV and 4 GeV, respectively, to result in a boost of $\beta\gamma = 0.28$ along the z axis in the laboratory frame. This boost in the lab frame ensures that the B decay lengths are long enough for the time sequence of their decays to be measured and their decay vertices to be easily separated. Additionally, this beam energy asymmetry is slightly reduced compared to that of the previous accelerator KEKB, which improves the solid angle coverage of Belle II.

The high B meson production rate of about $1.1 \times 10^9 B\bar{B}$ pairs per 100 fb^{-1} , that gives the experiment the title ‘ B -Factory’, still requires record breaking amounts of data to be able to push boundaries of particle physics and produce results that could confirm NP beyond the SM with the conventional minimum required significance of five standard deviations (5σ).

In particle physics, the amount of data collected is referred to as the ‘integrated lumi-

3. The Belle II Experiment

osity'. Instantaneous luminosity is defined as

$$L = \frac{1}{\sigma} \frac{dN}{dt}. \quad (3.1)$$

Here, σ is the interaction cross-section, N is the number of collision events, and t is time. Luminosity is the number of events per second per cm^2 . In terms of particles, the unit of centimetres is too large, thus luminosity at particle detectors is measured in inverse barns, where $1 \text{ barn} = 10^{-24} \text{ cm}^2$. If one has an instantaneous luminosity of $4 \text{ nb}^{-1} \text{ s}^{-1}$, then an interaction cross-section of 1 nb would give one a rate of 4 events per second. One can then calculate the integrated luminosity (over time) to find the total number of inverse barns of data collected. Table 3.1 lists the cross-sections of some of the possible e^-e^+ processes produced in the experiment, and therefore hints at the importance of having such a high luminosity. The further decay of a B meson to a desired signal decay channel, and the fact that a reconstruction efficiency is never 100%, leads to the demand of millions or billions of $B\bar{B}$ meson pairs to be created and recorded for viable measurements.

A nanobeam scheme was chosen in the Belle II upgrade to reach the desired luminosity. In the nanobeam scheme, the beam size at the collision point was reduced by a factor of 20, from $1 \mu\text{m}$ to 50 nm , and the currents were increased from those of KEKB by a factor of 2. Thanks to this upgrade, Belle II expects a factor of ~ 20 increase in instantaneous luminosity, in addition to an ~ 1.5 factor increase due to the escalation of currents. This leads to an overall factor of ~ 30 gain in target instantaneous luminosity from the previous Belle experiment to $\sim 6.5 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$. In Spring 2021, SuperKEKB broke its own world's highest instantaneous luminosity record with $3.1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. This is still an order of magnitude below design luminosity. Figure 3.2 shows the projection for full integrated luminosity of 50 ab^{-1} to be attained by 2031.

The benefits of this considerable increase in luminosity does not come without its pitfalls. The beam background contributions of SuperKEKB substantially increase backgrounds that can occur both with and without collisions. These backgrounds are significant and are taken into account throughout the design of the Belle II detector to reduce damaging effects to its components and to reduce the chance of beam backgrounds (Touschek scattering,

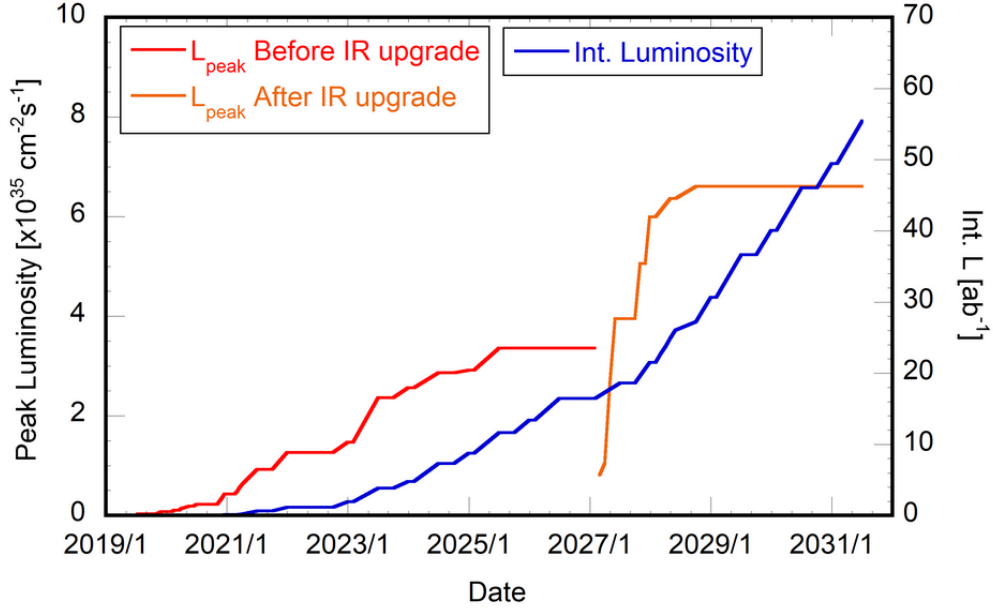


Figure 3.2.: The instantaneous (peak) luminosity before IR upgrade (red), after IR upgrade (orange) and integrated luminosity (blue) of the SuperKEKB accelerator as a function of time.

beam-gas scattering and synchrotron radiation) being falsely identified as signal decays.

3.2. The Belle II Detector

Belle II is the particle detector, placed around the interaction point (IP) of SuperKEKB. It is a 4π spectrometer and consists of several layers of sub-detectors in a cylindrical geometry about the IP. As depicted in Figure 3.3, the x -axis is defined in the horizontal direction, the y -axis is defined in the vertical direction, and the cylinder is oriented along the z -axis and along the beam. This is known as a right-handed orientation. The polar angle θ subtends from the positive z -axis and the azimuthal angle ϕ in the $x - y$ plane subtends from the positive x -axis. It is pragmatic here to also define the transverse momentum p_T , which is momentum perpendicular to the z -axis. It is a convenient variable that represents physics processes that occurred at the IP, in contrast to any momentum along the z -axis, which may be an artefact of beam particles.

One of the main goals of the combined SuperKEKB accelerator and Belle II experiment

3. The Belle II Experiment

Physics process ($e^-e^+ \rightarrow$)	Cross section [nb]
$\Upsilon(4S)$	1.110 ± 0.008 (MC stat.)
$\Upsilon(4S) \rightarrow B^+B^-$	0.5654
$\Upsilon(4S) \rightarrow B^0\bar{B}^0$	0.5346
$u\bar{u}(\gamma)$	1.61
$d\bar{d}(\gamma)$	0.40
$s\bar{s}(\gamma)$	0.38
$c\bar{c}(\gamma)$	1.30
$e^+e^-(\gamma)$	300 ± 3 (MC stat.)
$\gamma\gamma(\gamma)$	4.99 ± 0.05 (MC stat.)
$\mu^+\mu^-(\gamma)$	1.148
$\tau^+\tau^-(\gamma)$	0.919
$\nu\bar{\nu}(\gamma)$	0.25×10^{-3}
$e^+e^-e^+e^-$	39.7 ± 0.1 (MC stat.)
$e^+e^-\mu^+\mu^-$	18.9 ± 0.1 (MC stat.)

Table 3.1.: The total production cross-sections from the e^-e^+ processes occurring at $\Upsilon(4S)$ CM energy $\sqrt{s} = 10.58$ GeV. Errors that are given are Monte Carlo (MC) statistics. Information obtained from Table 18 of Reference [42].

is to search for NP BSM through measurements of B meson decays. The collaboration will do this by collecting 50 times more data than their predecessors, KEKB and Belle. By the early 2030s, Belle II aims to collect 50 ab^{-1} of data, equivalent to 55 billion $B\bar{B}$ events. This goal requires Belle II to operate at a 40 times higher luminosity leading to 40 times the Belle event rates. This has the knock on effect of background rates increasing by an anticipated factor of 40 [42]. To mitigate the effects of the increased background levels whilst maintaining a good performance of the spectrometer, there were many considerations taken into account, including the need for the ability to tolerate an increase in occupancy and radiation hardness, which shall be discussed. This section introduces a breakdown of the Belle II components at the point of data collection in 2020, which is referred to as the phase 3 geometry stage.

Belle II is separated into various regions and sub-detectors. Each sub-detector has particular roles for measuring particle kinematics and properties. It is useful to identify two regions of the Belle II detector, visible in the full schematic found in Figure 3.4. Firstly, the ‘barrel’ region is the region of the detector that encompasses the beam pipe cylindrically. Secondly, the ‘endcaps’ are the two ends of the detector walls that are perpendicular to

3. The Belle II Experiment

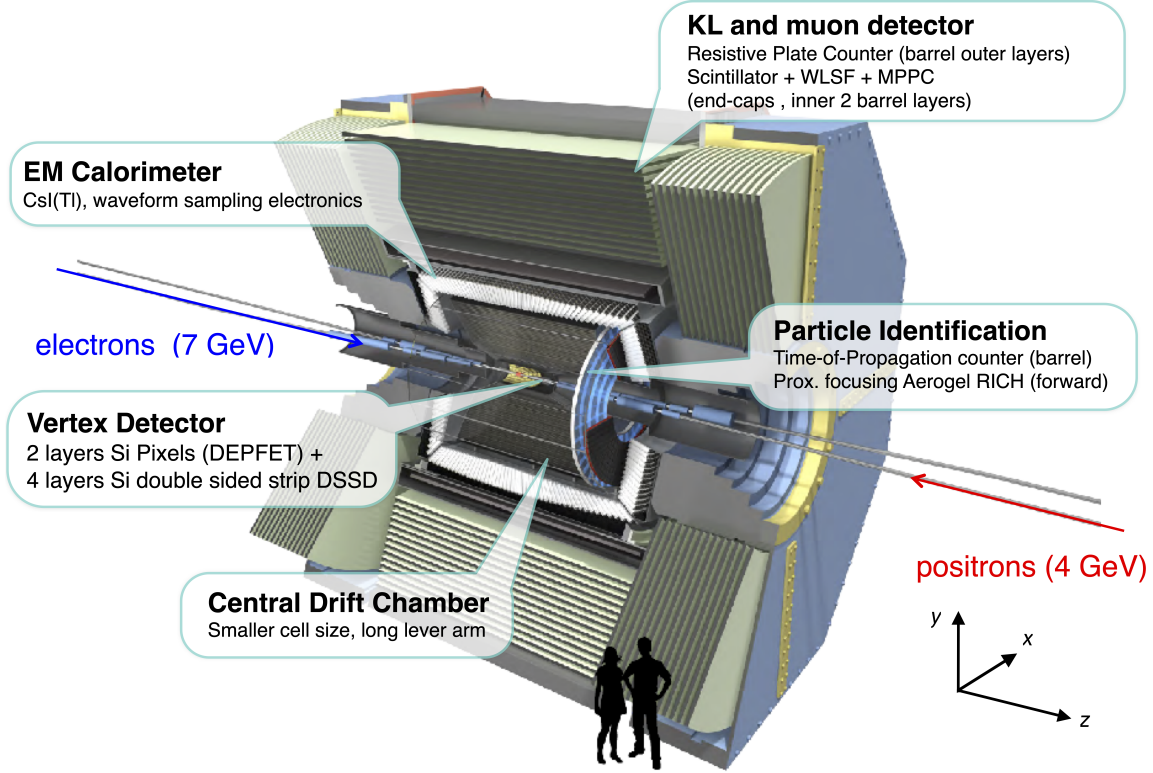


Figure 3.3.: A scaled model of the Belle II detector from the side depicting the relative size of the detector to humans, the coordinate system and the various sub-detectors. Figure sourced from Reference [43].

the z -axis, and close the cylindrical detector into near hermetic coverage. The two end-caps are distinguished using the asymmetric property of the beam, such that the ‘forward’ endcap is the endcap on the right of the diagram, oriented towards the electron beam and consequent lab-frame boost (i.e. in the positive z direction). The other endcap is termed the ‘backwards’ endcap.

Belle II has state-of-the-art tracking detectors which consist of the pixel vertex detector (PXD), the silicon vertex detector (SVD) and the central drift chamber (CDC). These are together placed around the 10 mm radius Be beam pipe and are responsible for tracking the trajectories of charged particles. These trajectories are intentionally curved using a solenoid with a 1.5 T magnetic field (oriented to the z axis) to allow for particle momentum to be measured. These trajectories also allow for precise determination of the decay vertex of their parent particles, which is particularly important for analysis of B meson decays.

Belle II also has an excellent particle identification (PID) system for the discrimination

3. The Belle II Experiment

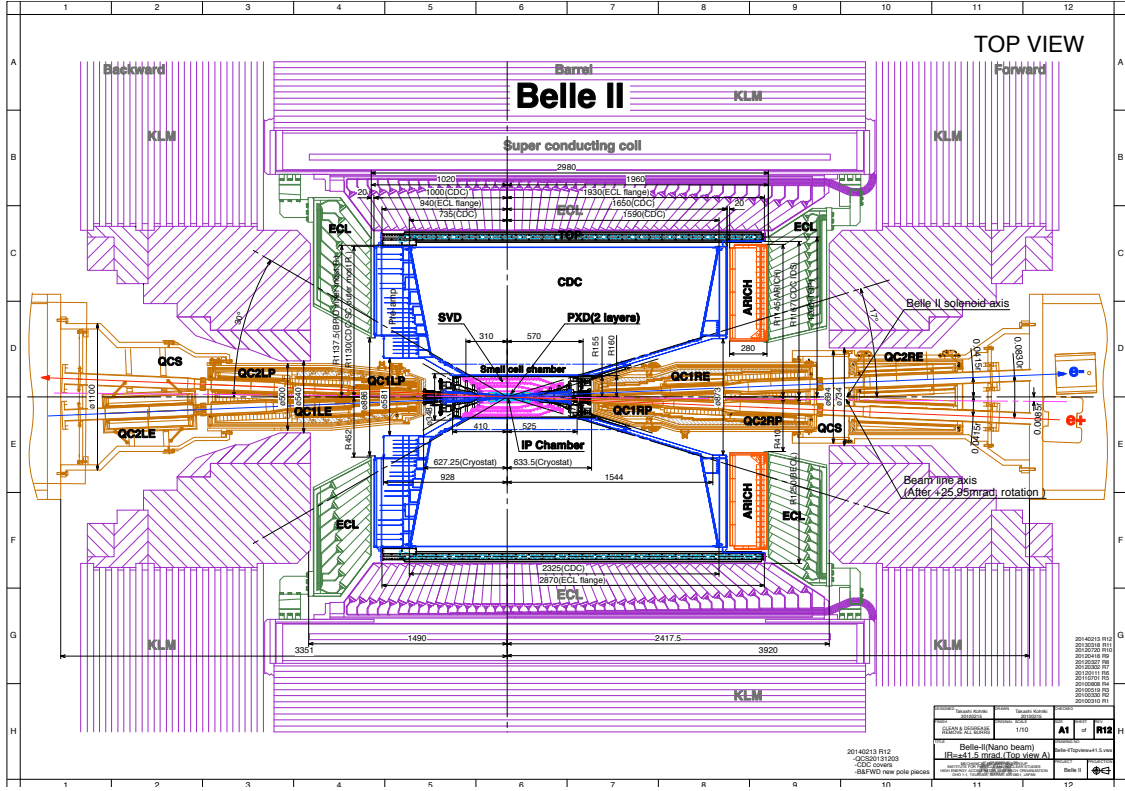


Figure 3.4.: The fully upgraded Belle II detector as viewed from the top.

of charged pions and kaons, comprising of the time of propagation (TOP) and aerogel ring imaging Cherenkov (ARICH) detectors.

The Belle II detector is not yet in its final form and expects a few more upgrades in the coming years. The TOP detector will also be enhanced with more radiation hard photomultipliers. Then in 2026, the final focusing superconducting magnets will be moved further away from the beam crossing point to reduce the chance of quenches and allow for Belle II to reach its final target instantaneous luminosity.

The following sections will introduce each sub-detector and are a summary of the in-depth information recorded in Reference [3].

3.2.1. Vertex Detector

The vertex detector is the innermost sub-detector of the Belle II experiment with the primary goal of reconstructing decay vertices, especially those of short-lived particles. It uses position-sensitive silicon layers to measure the trajectories of charged particles. The vertex

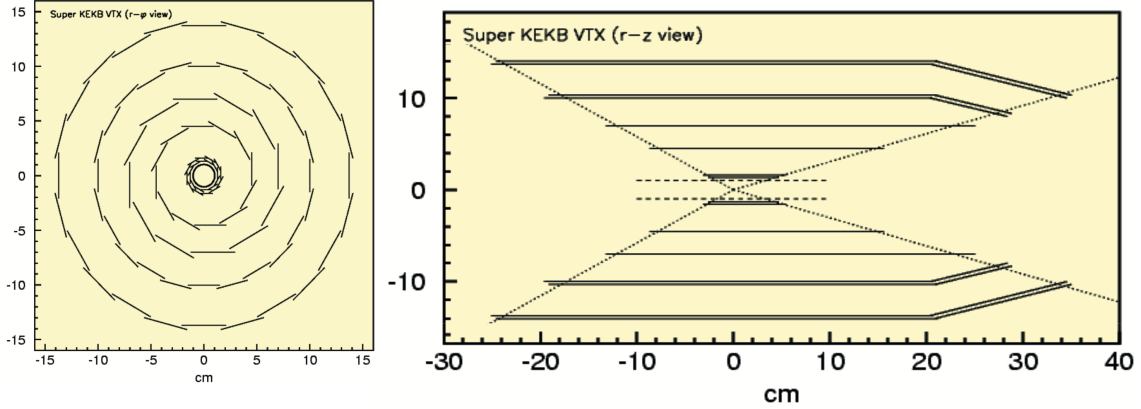


Figure 3.5.: A schematic of the VDX with the Be beam pipe at the centre. The inner two layers are the PXD layer, whilst the outer four layers are the SVD.[42]

detector consists of two inner layers of silicon pixel sensors, the PXD, and four outer layers of silicon microstrip sensors, the SVD, visible in Figure 3.5. These two sub-detectors are responsible for stand-alone tracking of low transverse momentum charged particles that cannot escape the vertex detector region, as well as for providing the innermost tracking space-points for higher transverse momentum tracks which traverse the central drift chamber (Chapter 3.2.2). The positioning of these layers is significant when compared to the original Belle vertex detector as the beam pipe and SVD are closer to the IP and the outermost layer of the PXD is at a considerably larger radius. This has the effect of improved vertex resolution, as well as improved reconstruction efficiency of any incidences of the $K_S^0 \rightarrow \pi^+ \pi^-$ decay within the vertex detector.

PiXel Detector (PXD)

The PXD is the vertex detector situated closest to the beam-pipe. It includes two layers of a total 8 million silicon DEPLETED Field Effect Transistor (DEPFET) pixels of 50 by 50 μm each, situated at radii of 14 mm and 22 mm. These small radii compensate for the smaller separation between $B\bar{B}$ pairs than at Belle due to the lower energy beam asymmetry. Ionising particles travelling through the PXD produce electron-hole pairs that propagate and accumulate charge to produce a signal in the semi-conductor DEPFET pixels. The benefits of using DEPFET technology are the low power consumption such that air cooling

3. The Belle II Experiment

is sufficient for active cooling, radiation hardness, and the ability to be constructed at 50 microns [3]. Thus, each layer is 0.19% of a radiation length and allows the preservation of vertex resolution through the reduction of the likelihood of charged particles multiple scattering in the detector medium. The PXD has a spatial resolution of $10\ \mu\text{m}$.

Silicon Vertex Detector (SVD)

The SVD consists of four layers of double-sided silicon microstrip sensors, placed at 38 mm, 80 mm, 115 mm and 140 mm. Each have dimensions of approximately 125 by 45 mm². These give two 1D position measurements in the region just outside of the PXD. The material thickness is naturally larger than in the PXD and constitutes a radiation length of 0.57% per layer. As the double-sided layers are orthogonal to each other in the z and $r - \phi$ directions, the resolution of the SVD is strip dependent. Depending on the angle of incidence, the z layers have a resolution of $2 - 5\ \mu\text{m}$, while the $r - \phi$ have a resolution of $7 - 27\ \mu\text{m}$. The SVD is the last sub-detector responsible for measuring the vertexes of any K_S^0 mesons that escape the PXD.

3.2.2. Central Drift Chamber (CDC)

The CDC is a large-volume small-cell drift chamber of outer radius 1130 mm. For charged particles the CDC supports the PXD and SVD through precise measurements of momenta and charge, gives particle identification information and gives efficient and reliable trigger signals.

It has a large radius, with the inner cylinder starting at 160 mm. After Belle, the CDC was redesigned for Belle II to contain a total of 14,336 sense wires arranged in 56 layers either aligned with the solenoidal magnetic field (termed ‘axial’) or in ‘stereo’ i.e. orientated at a slant with respect to the axial wires, and 42,240 field wires. Figure 3.6 shows the increase in density of the inner wire layers to account for the higher beam backgrounds that will be experienced at Belle II.

The charged particles travel through a helium-ethane gas mixture and ionize the gas molecules, whilst field wires create an electric field to accelerate these ionized particles. Referring to Figure 3.4, the now free electrons accelerate towards the high voltage sensing

3. The Belle II Experiment

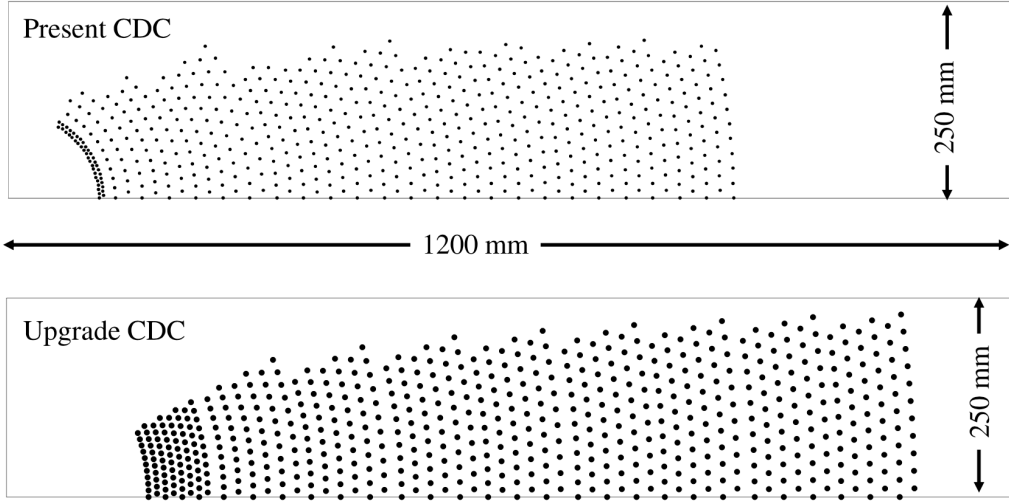


Figure 3.6.: A schematic comparing the wire configuration of the Belle ('Present') and Belle II ('Upgrade') drift chambers. [3]

wires that are stretched across the chamber. As these electrons travel towards the sense wires, they ionize other particles. Eventually through gas amplification a number of electrons travel at a high speed towards the sense wires, which registers the signal propagation and creates a pulse signal. The data acquisition (DAQ) system picks up this signal and processes it.

The outer solenoid at Belle II causes the trajectory of the charged particles to curve. A charged particle travelling through the CDC creates a trail of ionisation following the particle's trajectory and, through the combination of information from the axial and stereo wires, Belle II is able to reconstruct a full 3D helical track and determine the particle's momentum. From the timing of each wire signal, the drift time and therefore the distance at which the primary ionization was initialized can be inferred.

The particle's energy loss dE/dx is determined through the specific ionization calculated from the deposited charge on each wire. The CDC can then be used to identify the charged particle depending on its unique signature. Each type of charged particle has a characteristic distribution of kinetic energy loss as a function of distance in their tracks as illustrated in Figure 3.7. Thus even low momentum particles that cannot reach the higher radius sub-detectors for additional identification have a good probability of being identifiable. The CDC is expected to attain a momentum resolution of 0.2%. The $\frac{dE}{dx}$

3. The Belle II Experiment

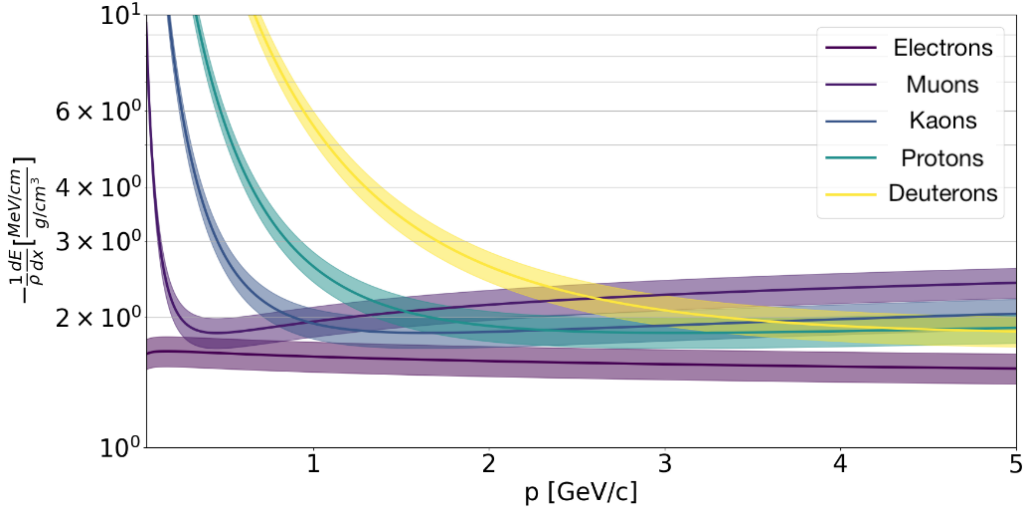


Figure 3.7.: The specific ionisation of charged particles and their separation at Belle II as a function of momentum. Figure sourced from Reference [44].

resolution varies in the range 8.5 % to 12 % due to the dependency of position and trajectory reconstruction resolutions on the incident angle of the charged particles.

3.2.3. Time of Propagation (TOP) counter

The TOP detector is a time-of-propagation detector which surrounds the outer walls of the CDC chamber in the barrel region and provides charged PID. 16 modules of 0.02 m thick, 2.7 m long, 0.45 m wide quartz radiator bars are distributed azimuthally around the inner components. Installed at the forward end of the bars is a mirror; installed at the backward end is a 10 cm prism. Attached to the prisms are micro-channel plate (MCP) photomultipliers (PMTs) to measure the incident photons. The charged particle passing through the quartz causes Cherenkov photons to be produced at an angle that is directly dependent on the particle's velocity. The photons are totally internally reflected. Upon interaction with the PMTs, the TOP can be used to reconstruct 2D information of Cherenkov rings through the time of arrival and impact position of the Cherenkov photons. The TOP has a time resolution of 100 ps.

Combining the velocity from the Cherenkov light angle with the particle's momentum measured in other sub-detectors allows PID through its mass. The separation power of the PID is more limited the lower the momentum, due to chromaticity of Cherenkov photons.

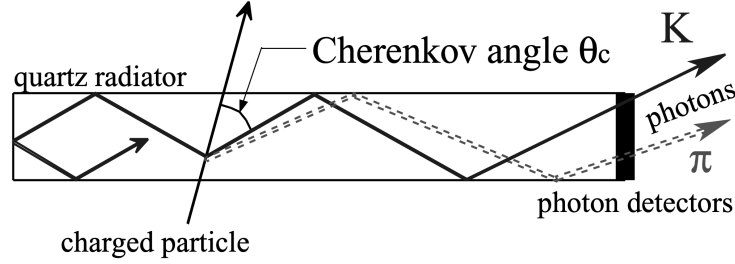


Figure 3.8.: A schematic of the TOP counter consisting of a side view of totally internally reflecting Cherenkov photons, with a separation between K and π mesons shown. The labelled ‘photon detector’ is the PMT sensor plane discussed in the text [3].

Therefore, the primary PID function is to separate between K and π particles. The K - π separation of 4.3σ at $p = 4 \text{ GeV}$ [45] due to difference in time-of-flight and detection position displayed in Figure 3.8. Due to the chromaticity, the time resolution broadens at lower momenta. To minimize this effect, a mirror focusing system was introduced allowing the division of the ring image depending on the wavelength of the Cherenkov photons.

3.2.4. Aerogel Ring-Imaging Cherenkov detector (ARICH)

The ARICH detector is a proximity focusing, aerogel, ring-imaging Cherenkov detector. It only exists in the forward region of the Belle II detector and provides K - π separation over the majority of their momentum spectrum (0.4 GeV to 4 GeV) alongside separation of low momentum ($< 1 \text{ GeV}$) π , μ and e . As shown in Figure 3.9, Cherenkov photons are produced in the aerogel by charged particles. These traverse across a 20 cm gap that allows development into rings which are detected at a position sensitive photon detector. One key aspect of the ARICH that needed to be considered during the design was the minimum number of photons required for a good ring image. To increase the number of photons produced by a charged particle, one can increase the thickness of the aerogel. However, with this increase, the resolution of the Cherenkov angle drops due to the gain in uncertainty of the photon origin. To improve the ring image without reducing resolution, a novel solution was found in using two 2 cm thick aerogel layers with different refractive indexes of 1.045 and 1.055. This creates this effect of a focusing lens, as displayed in Figure 3.10.

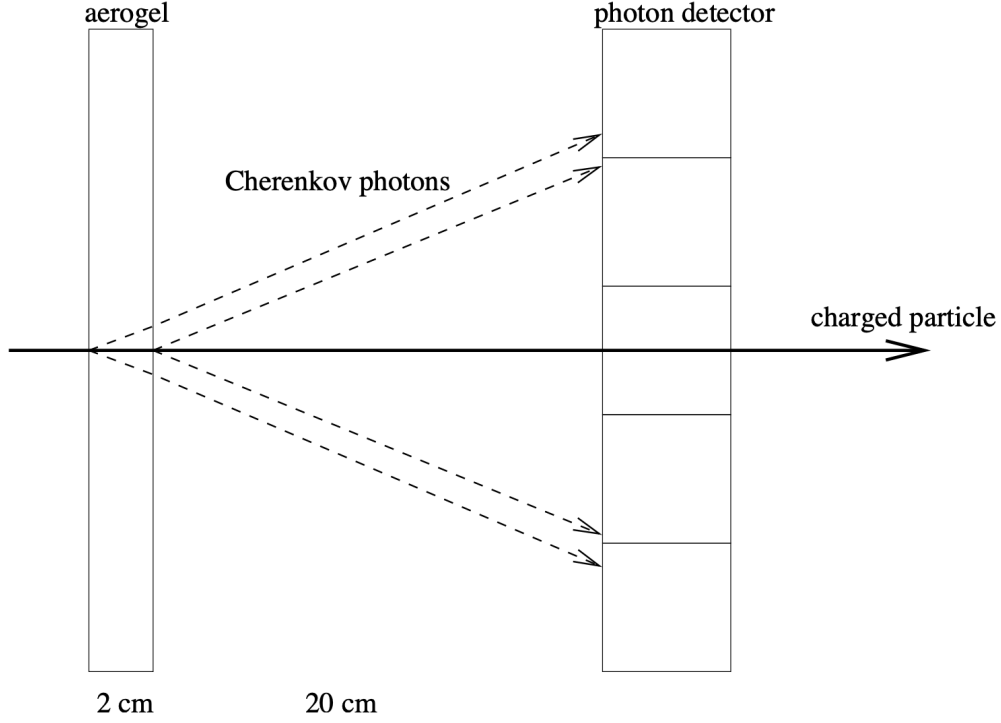


Figure 3.9.: The concept of the ARICH detector at Belle II [3].

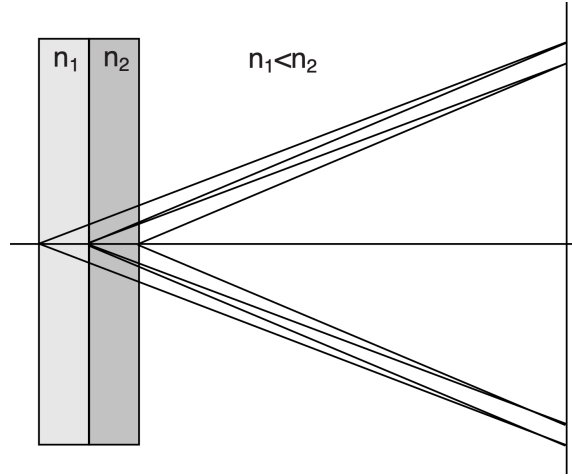


Figure 3.10.: The concept of the proximity focusing ARICH using 2 aerogel layers of different refractive indexes [42].

3.2.5. Electromagnetic Calorimeter (ECL)

The ECL is the first sub-detector layer in the Belle II detector that has the ability to detect neutral particles from their photon products and also to detect gamma-rays. The ECL – segmented in a projective geometry across both the barrel and endcap regions of

3. The Belle II Experiment

the Belle II detector – is wrapped around the TOP and ARICH sub-detectors and other inner aforementioned detectors. It is still placed within the solenoid as shown in Figure 3.4. As most sub-detectors in the Belle II detector only have the ability to measure charged particles, the ECL is the primary component in neutral particle detection. In addition to detection of photons, it is able to determine photon energies and spacial coordinates with good resolution. The ECL is a homogeneous, thallium-doped caesium iodide CsI(Tl) crystal scintillation electromagnetic calorimeter. CsI(Tl) was chosen as the scintillation medium in part for its high light output and relatively short radiation length. The forward and backward endcap regions span 0.42 m in the $\pm z$ directions and contain 2112 crystals. The barrel region is 3 m long with an inner radius of 1.25 m and outer radius of 1.62 m. The ECL has nearly complete azimuthal angle coverage and polar angle coverage from 12.1° to 157.1° minus 1° gaps at the joints between the barrel and endcaps.

As an ionising particle traverses the CsI(Tl) medium, bremsstrahlung photons are emitted. These photons, and any other incident photons, pair produce e^+e^- which then radiate bremsstrahlung photons. These photons then continue this cyclic process over and over, producing what is called an electromagnetic shower (until there is no longer sufficient energy to pair produce or emit more photons). The electromagnetic shower is absorbed by the CsI(Tl) crystals and re-emitted as scintillation photons. The photons are then detected via two 10 mm by 20 mm PIN photodiodes per crystal, converting them into electronic signals. The electronic signals are then amplified by preamplifiers placed at the back of each crystal. The scintillation light is proportional to the total ionization i.e. the total energy of the incident particle, allowing differentiation between incident charged particle types.

The ECL is also utilized to differentiate between electrons and hadrons (in particular pions) with high efficiency due to the difference in signature of electromagnetic and hadronic showers. The ability to destructively measure the energy of charged particles enables identification of electrons through energy deposits or total absorption, unlike muons which interact only via ionization and hence typically lose only a small fraction of their energy as they pass through the whole detector. Furthermore, the ECL has the important tasks of the online and offline luminosity measurements and K_L^0 PID alongside the neutral long-lived

3. The Belle II Experiment

kaon (K_L^0) and muon (μ) detector (KLM) (Chapter 3.2.6).

One third of B meson decay products are either π^0 or neutral particles that decay to photons of energies between 20 MeV to 4 GeV. Due to this reason, a high resolution ECL is necessary to detect photons.

Due to the increase in background levels compared to Belle, there is an increase of fake hits and pile-up noise (signal overlap when a crystal has not finished scintillating from a previous event). To mitigate this, the readout system of the scintillator photo-sensors is equipped with wave-form sampling read-out electronics. Overall, the timing resolution for the ECL is 12 ns (22 ns) in the barrel region (forward and backward endcap regions), the position resolution is $\sigma_x = 5 - 10$ mm and an energy resolution of $\sigma_E/E \sim 4\%$ at 100 MeV.

A uniform axial super-conducting solenoid magnet surrounds the calorimeter, filling the inner sub-detectors with the aforementioned 1.5 T magnetic field.

3.2.6. Neutral long-lived Kaon (K_L^0) and Muon (μ) detector (KLM)

The KLM detector is the outermost sub-detector of Belle II, located outside of the 1.5 T magnet. It serves as a hadronic calorimeter and enables identification of neutral K_L^0 mesons and muons. It comprises of layers of resistive-plate chambers in the barrel region, interspersed between iron plates. These plates also serve as the magnetic flux return for the field produced by the superconducting solenoid. In the endcap regions however, it is necessary to use a material with a shorter dead-time to be able to compete with the higher beam backgrounds at Belle II. Thus, polystyrene scintillator strips with wavelength shifting fibres are used.

Due to increased background levels in the Belle II detector, there is an increase in neutron induced hits. Neutral K_L^0 mesons pass undetected through all of the previous sub-detectors of Belle II. They can, however, be uniquely identified at this level of the detector using a combination of the KLM clusters and ECL clusters measuring particles not associated with a charged track that have an origin near the IP. Muons travel through the KLM, ionising the gas and producing tracks, but do not produce showers. Thus they can be identified with the help of matched charged tracks in the CDC. Charged hadrons generate hadronic

3. The Belle II Experiment

showers when interacting with the iron plates and thus can be differentiated from muons.

3.2.7. Triggering and Data Acquisition

The high instantaneous luminosity, and therefore high background levels at Belle II, results in the need for selection of events of interest at the stage of data collection. This is to reduce the large storage and processing speeds needed for data taking. Triggers are used to efficiently identify dominant background sources previously discussed in Chapter 3.1. Inherited from the previous Belle trigger scheme, a level 1 (L1) trigger uses signals from a number of detectors including the ECL, KLM and tracking detectors to make a quick decision on which events to keep. For example, L1 vetoes Bhabha events.

The increase in background levels since the Belle II upgrade required modifications to the trigger system. A High Level Trigger (HLT) method is implemented whose function is to reconstruct the data events in real time. When data is acquired, the DAQ delivers data from the sub-detectors to the HLT system. The data goes through two trigger selection stages: the filter and the skim stages. The filter trigger stage defines which events are discarded. The effect of the filter trigger can be seen in the difference in data retention in Table 3.2. The number of events that reach the HLT skim stage is reduced, therefore the retention percentage increases. With the leftover events that pass the filter trigger the skim trigger is performed, which tags the events online. It also provides additional selection and further filters Bhabha events with the benefit of allowing effectively 100% of $B\bar{B}$ events to pass the trigger.

Hadron HLT						
Data		MC				
Filter OFF	Filter ON	$B\bar{B}$	$c\bar{c}$	$q\bar{q}$	$\tau\tau$	$\mu\mu$
1.90	7.97	99.54	96.32	79.78	27.26	0.96

Table 3.2.: The retention (%) for the Hadron skims on Run 5710 experiment 12 (FILTER-OFF) and Run 5157 experiment 10 (FILTER-ON), shown with the efficiencies (%) for generic MC using `basf2` software `release-04-02-05`.

In this analysis, to optimise the data sample size and reduce computational load, the dataset is skimmed with a HLT skim called the ‘Hadron skim’. This skim has been in use since the beginning of Phase 3 data taking, and requires that the number of tracks

3. *The Belle II Experiment*

originating from the IP in an event is ≥ 3 and that the event is not a Bhabha event with ≥ 2 tracks. Table 3.2 displays the retention of data and MC after these requirements. This skim does not remove a large portion of the $c\bar{c}$ and $q\bar{q}$ events, resulting in its use as a very general skim, however it is effective in discarding low multiplicity and high-cross section processes such as $\mu\mu$ and $\tau\tau$ events, which are considered nuisance processes for this analysis.

4. Analysis Tools

The ultimate goal of this analysis is to contribute to the optimization of the measurement of the normalisation modes of $R(D)$ and $R(D^*)$ (see Equation 2.8 and 2.9) using Belle II data. The termed ‘normalisation modes’ are $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ and are the signal decays of this analysis. This chapter describes the analysis tools used in the sensitivity study.

4.1. Monte Carlo (MC)

We optimize the analysis by using the collected data from the Belle II experiment and remove any background decays without removing too many signal decays. To successfully do so, it is essential to simulate the data well. Simulated data is produced using the MC method and is used to simulate detector effects and to determine signal efficiencies and background rates at Belle II. This allows optimization of the discrimination of background events. To fully simulate the Belle II data, including signal and background, MC simulation samples of B meson pairs (charged and neutral) and continuum $e^+e^- \rightarrow q\bar{q}$ (where $q = u, d, c, s$) are generated alongside beam scattering, radiative processes and other beam background effects. Simulated data is referred to as MC, and the charged, neutral and continuum samples are considered ‘generic MC’. In this analysis the 14th run-independent MC campaign of Belle II, dubbed MC14a_ri, was used to simulate data. We simulated the number of events of each generic MC sample equivalent to 1 ab^{-1} of data luminosity, given the known cross sections of the simulated processes. The phrase ‘run-independent’ refers to the fact that the MC utilizes simulated background overlays. This means that the simulated backgrounds are subject to inconsistencies with collected data as the background overlays and detector conditions are not applied in proportion to the conditions in data,

4. Analysis Tools

i.e. they have not been tuned or corrected to the real conditions of the beam or the varying status of the Belle II detector that recorded the real data.

The MC14a.ri used in this analysis was produced using `release-05-02` of the `basf2` framework. This MC has early phase 3 geometry ^[i]. The fiducial cross-sections used in the MC14 production are shown in Table 4.1 alongside the corresponding number of events. These differ from that of Table 3.1 due to the limited statistics of the generated events when numerically calculating the cross sections.

Process ($e^+e^- \rightarrow$)	MC name	Cross-section [nb]	Number of events ($\times 10^6$) [per 100 fb $^{-1}$]
$\Upsilon(4S) \rightarrow B^+B^-$ (charged)	charged	0.540	54
$\Upsilon(4S) \rightarrow B^0\bar{B}^0$ (neutral)	mixed	0.510	51
$u\bar{u}$	uubar	1.605	160.5
$d\bar{d}$	ddbar	0.401	40.1
$s\bar{s}$	ssbar	0.383	38.3
$c\bar{c}$	ccbar	1.329	132.9
$\tau^+\tau^-$	taupair	0.919	91.9
$\mu^+\mu^-$	mupair	1.148	114.8
Total		6.835	683.5

Table 4.1.: The production cross-sections of the e^+e^- processes at the $\Upsilon(4S)$ CM energy used for the 14th run-independent MC simulation campaign. The corresponding Belle II MC name is shown for each process simulated in generic MC. Data for this table was collected from internal Belle II calculations and production notes [46].

Generic MC aims to simulate all SM events expected to remain after trigger selection. Thus, we study and measure SM decays of interest using generic MC such as $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$. However, often due to the low branching fraction of these decays, their statistical significance as a subsample of the generated generic MC is inadequate for meaningful studies. Using, for example, the theoretical branching fraction for the $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ decay quoted in Table 2.1, in 1 ab $^{-1}$ of generic MC, we expect approximately 1.6×10^7 events of the 5.1×10^8 $B^0\bar{B}^0$ pairs to be $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$ events before accounting for reconstruction efficiency. Thus, designated samples of specific decays are designed and produced called signal MC.

^[i]For reproduction purposes, this is without PXD L2.

4. Analysis Tools

Signal MC was used to optimize many aspects of this analysis, including selecting cuts that distinguish between signal and background events, optimizing a multi-variate analysis and calculating signal reconstruction efficiencies. For this analysis, 3 sets of signal MC were created for various analysis optimization steps. Firstly, a sample of 40×10^6 events of $B \rightarrow X_C \ell \nu$ decays was used for signal studies of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$. Here X_C is any hadron with a c quark in the quark pair. As this sample contains non-signal decays, the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays were selected using topological generator-level truth information (see Section 4.5.1). The other two samples were specific $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ and $B \rightarrow D^{**} \tau^- \bar{\nu}_\tau$ samples that did not contain superfluous decays. Each of these have 80×10^6 generated events. A summary of these can be found later in Table 4.5.

4.2. The Belle II Data Sample

Belle II started collecting data with the full detector setup in 2018, thus the data available for analysis is not yet the complete dataset projected for the Belle II experiment. The data used in this analysis was collected in the run periods of 2019-2021, reaching a delivered online integrated luminosity of 213.49 fb^{-1} , and a final collected offline integrated luminosity of $(189.8804 \pm 0.0131) \text{ fb}^{-1}$ at the $\Upsilon(4S)$ resonance^[ii] [43]. This dataset was the official dataset used for results for the Moriond 2022 conference and consists of experiments 7 to 18 of on-resonance data. The number of B meson pairs in this dataset has been determined using event-shape variables and has been determined to be $N_{B\bar{B}} = (198.0 \pm 3.0) \times 10^6$. Figure 4.1 displays the total integrated luminosity at Belle II as a function of time. Table 4.2 presents a summary of the dataset used for this analysis.

Data Type	Luminosity	Number of events
Moriond 2022 dataset	$(189.8804 \pm 0.0131) \text{ fb}^{-1}$	$N_{B\bar{B}} = (198.0 \pm 3.0) \times 10^6$

Table 4.2.: A summary of the dataset used for this analysis.

^[ii]This will be abbreviated to 189 fb^{-1} in figures for this thesis.

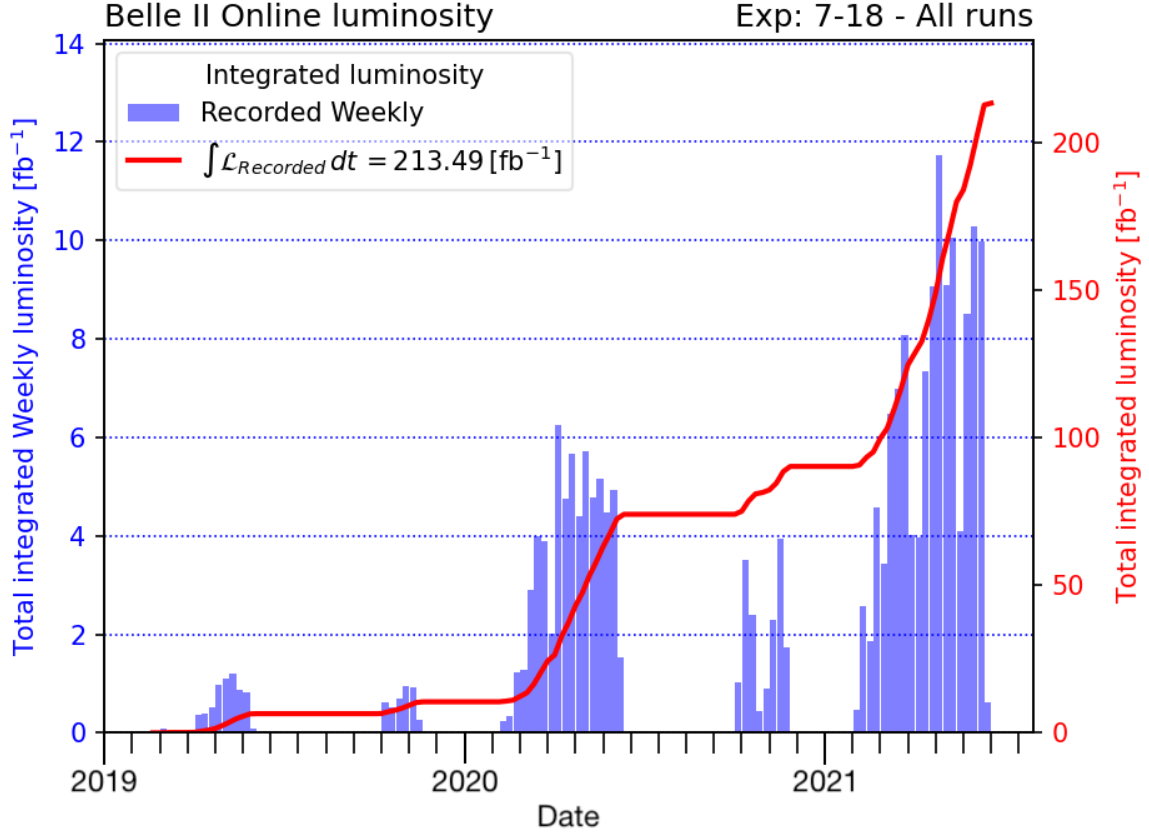


Figure 4.1.: The online luminosity of Belle II; the data used for this analysis. The total integrated daily luminosity is indicated in the (blue) histogram, while the cumulative integrated luminosities are indicated by the (red) line. Figure sourced from internal Belle II Confluence website - image created by Stefano Lacaprara [43].

4.3. The Belle II Analysis Software Framework (basf2)

The `basf2` package is the official software suite of the Belle II experiment [47]. `basf2` has the functionality to simulate particles, to monitor the Belle II detector and to perform physics analyses. It notably incorporates external libraries such as: ROOT, CERN's physics analysis software [48]; EvtGen, a particle decay simulator including those of the B and D mesons relevant to this analysis [49]; and Geant4, a simulator useful for sub-detector interaction and detection of particles [50]. In any Belle II physics analysis, simulated and real data are processed using `basf2`.

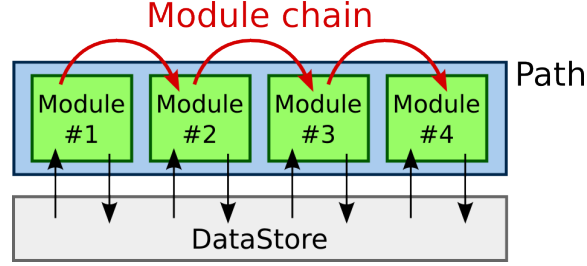


Figure 4.2.: The `basf2` data process chain. Image created by Andreas Moll [47].

4.3.1. `basf2` Structure

`basf2` processes the data and generated MC through the use of software modules. Each module has a separate function, such as reading, manipulating, or writing data. Modules are arranged in a strictly linear path in order to perform a sequence of configurable analysis functions. Analysis scripts define this configuration and control the execution of these modules in the analyst-given path order. The data needed by each module is shared via a `DataStore`, shown as a schematic in Figure 4.2. The `DataStore` stores data-objects such as `ParticleLists`, `ECLClusters`, `KLMClusters` and `Tracks`. The analyst can take these tracks and clusters that have been measured in the detector and create `Particle` objects, which are stored in `ParticleLists`. Next the `Particles` can be combined with other `Particles` to create a parent `Particle` (for example combining a pion π and a kaon K to create a D meson). Throughout the analysis, the analyst can impose selection requirements referred to as ‘cuts’ on the various quantities stored in the `Particle` object. Cuts reject `Particle` candidates that do not satisfy certain conditions. At this point of reconstruction in the analysis, cuts are made to increase the likelihood that the candidate particle(s) chosen for reconstruction is the particle the analyst expects. The analyst can also perform a vertex fit on the `Particle` to find and reconstruct the point at which the particle candidate decayed within the detector. The analyst then has the option of applying various tools and performing various advanced algorithms within `basf2` such as Continuum Suppression, Flavour Tagging and Full Event Interpretation (FEI) [51]. Continuum Suppression uses variables such as CLEO cones [52] and Fox-Wolfram moments [53] to evaluate the shape of an event to distinguish between $B\bar{B}$ events and continuum events. Flavour Tagging is a multivariate method based module that enables the determination of the flavour of one of

the B mesons in a $B\bar{B}$ pair through examining the quark content of its flavoured hadrons descendants. The FEI algorithm is one of the features of this analysis that makes it unique, thus it is discussed in detail in Chapter 4.

4.4. The Full Event Interpretation Algorithm (FEI)

There are various approaches to fully reconstruct an $\Upsilon(4S) \rightarrow B\bar{B}$ event at Belle II. These can be either by a tagged or untagged analysis. An untagged analysis reconstructs a desired signal decay of a B meson and uses the particles in the Rest of the Event (ROE) to constrain the second B meson of the $B\bar{B}$ pair. Here, the ROE is defined as the particles in an event that are not associated with a designated signal particle.

In contrast, a tagged analysis reconstructs or ‘tags’ a B meson (B_{tag}) with well understood (or otherwise beneficial) decays, then selects only the desired signal decay for the other signal B (B_{sig}) from the ROE. The B_{sig} and the companion B_{tag} in an event are then reconstructed into an $\Upsilon(4S)$.

In order to tag a B meson at Belle II, a Fast BDT machine learning software called FEI was developed from the Full Reconstruction algorithm from Belle [54]. This uses hierarchical reconstruction of over 10,000 exclusive B meson decay chains and multivariate classification to tag the B_{tag} meson candidate particles in an event. Every particle decay channel implemented in the FEI algorithm has a designated multivariate classifier which must be trained on the MC simulation of data. The MC used for training must then be removed from further analysis to prevent bias in later application of FEI. For this analysis, FEI was trained on 100 fb^{-1} of generic $B\bar{B}$ MC from the MC14ri_a campaign ^[iii] - a sample size that ensured sufficient training. Following the paths in Figure 4.3, the algorithm:

1. Collects detector data in the form of tracks and neutral E_{ECL} clusters.
2. Identifies tracks, displaced vertices, and neutral clusters that pass basic quality criteria. BDTs are trained to identify long-lived particles in the detector (e^- , μ^- , K^- , p , π^- , γ and K_L) from charged tracks and neutral energy deposits.

^[iii]For replication purposes, the production used was prod 00016661/2.

4. Analysis Tools

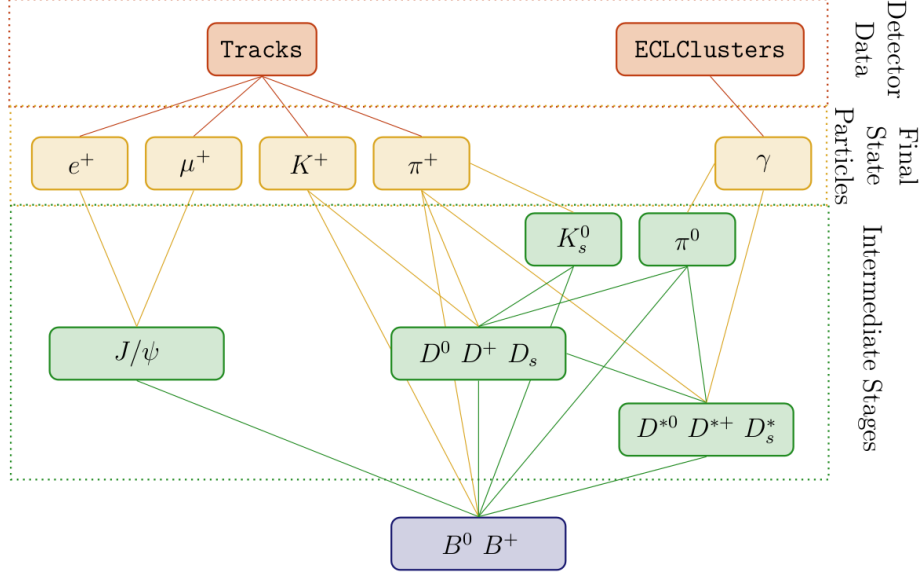


Figure 4.3.: The FEI hierarchical structure showing the progression of the detector information of tracks (and displaced vertices) and neutral clusters (energy deposits in the ECL and KLM) to the final tagged B^+ or B^0 . Figure sourced from Reference [55].

3. Reconstructs stable particles into composite parent particles (π^0 , K_s^0 , J/Ψ , D^0 , D^+ , D_s , Λ_c , Λ and Σ^-). For each composite parent particle in this list, a BDT is trained to identify potential candidates.
4. Reconstructs excited D meson candidates (D^{*0} , D^{*+} , D_s^*) from the candidates in step 2. Separate BDTs are then trained, again to identify potential candidates.
5. Combines all of the information gained from the previous stages to reconstruct potential B_{tag} candidates. A BDT is then trained to identify viable combinations.

The final BDT trains a multi-variate classifier interpreted as the B_{tag} ‘signal probability’ $\mathcal{P}_{B_{\text{tag}}}$. This distinguishes correctly reconstructed B_{tag} candidates from the incorrect combinations of particles by assigning a probability between 0 and 1, where 0 is most background-like (mis-reconstructed B meson candidates) and 1 is most signal-like (a correctly reconstructed B meson candidate).

There are two variations of FEI at Belle II: using a hadronic tag or semileptonic tag. The hadronic tag is the process of using FEI to reconstruct a B_{tag} meson with only its hadronic decay channels. The benefit of using the hadronic decays of the B meson is that

4. Analysis Tools

the kinematics of the reconstructed B_{tag} candidates are fully constrained and the tagged sample is pure in comparison to the semileptonic tag. However, there is a low efficiency in the hadronic tag due to the many levels of reconstruction and BDTs used. Thus, due to the small branching fractions of these processes, a large number of decay modes are necessary to cover a sufficiently high branching fraction. The larger number of hadronic decay modes increases the issue of combinatorial backgrounds, which scale as a factorial of the number of tracks and, as a result, early selections to reduce backgrounds are required that unavoidably cut out signal events too.

The semileptonic tag involves an exclusive reconstruction of a semileptonic B meson such as the $B_{\text{tag}} \rightarrow D^{(*)} \ell \nu_\ell$ decay as the tagged mode. In this case there is a neutrino in the final state which cannot be reconstructed, so the 4-momentum of the corresponding B_{sig} cannot be fully reconstructed. The benefit of the semileptonic tag of the B meson is the relatively large branching fraction. The semileptonic tag is more efficient than the hadronic tag, and so results in a larger tag yield for a given dataset. Thus, it potentially has better statistical sensitivity at the cost of higher backgrounds and lower purity.

As this analysis has a dependency on distinguishing between an event with one and multiple neutrinos, the hadronic version of the FEI is used (see Section 6.2.1). This ensures that neutrinos appearing as missing energy in an event are produced solely by the B_{sig} .

4.4.1. Hadronic FEI Skim

As discussed in 3.2.7, raw data is reduced in size using centralized skimming and to aid in computational load using the trigger and the HLT skim. These are produced centrally by the Belle II collaboration. This analysis is subsequently performed on these reduced datasets.

The FEI hadronic skim consists of a selection of loose ‘pre-cuts’ that reduce the size of data samples with minimal loss of $B\bar{B}$ events and ensures that the FEI considers hadronic $B\bar{B}$ events. Then, it applies the trained FEI algorithm to MC and data. To ensure that tracks have a close proximity to the interaction point, the track parameters are required to be within 2.0 cm of the centre of the detector along the solenoidal-field axis (the z -axis), and within 0.5 cm of the transverse plane d_0 . The track transverse momentum p_t is required

4. Analysis Tools

Decay Mode	Branching Fraction	Decay Mode	Branching Fraction
1. $B^+ \rightarrow \bar{D}^0 \pi^+$	$(4.68 \pm 0.13) \times 10^{-3}$	1. $B^0 \rightarrow D^- \pi^+$	$(2.52 \pm 0.13) \times 10^{-3}$
2. $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0$	5×10^{-4} [i]	2. $B^0 \rightarrow D^- \pi^+ \pi^0$	5×10^{-4} [i]
3. $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0 \pi^0$	1×10^{-3} [i]	3. $B^0 \rightarrow D^- \pi^+ \pi^0 \pi^0$	1×10^{-3} [i]
4. $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^-$	$(5.6 \pm 2.1) \times 10^{-3}$	4. $B^0 \rightarrow D^- \pi^+ \pi^+ \pi^-$	$(6.0 \pm 0.7) \times 10^{-3}$
5. $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^- \pi^0$	No value given.	5. $B^0 \rightarrow D^- \pi^+ \pi^+ \pi^- \pi^0$	No value given.
6. $B^+ \rightarrow \bar{D}^0 D^+$	$(3.8 \pm 0.4) \times 10^{-3}$	6. $B^0 \rightarrow \bar{D}^{*0} \pi^+ \pi^-$	$(6.2 \pm 2.2) \times 10^{-4}$
7. $B^+ \rightarrow \bar{D}^0 D^+ K_s^0$	$(1.55 \pm 0.21) \times 10^{-3}$	7. $B^0 \rightarrow D^- D^0 K^+$	$(1.07 \pm 0.11) \times 10^{-3}$
8. $B^+ \rightarrow \bar{D}^{*0} D^+ K_s^0$	$(2.1 \pm 0.5) \times 10^{-3}$	8. $B^0 \rightarrow D^- D^{*0} K^+$	$(3.5 \pm 0.4) \times 10^{-3}$
9. $B^+ \rightarrow \bar{D}^0 D^{*+} K_s^0$	$(3.8 \pm 0.4) \times 10^{-3}$	9. $B^0 \rightarrow D^{*-} D^0 K^+$	$(2.47 \pm 0.21) \times 10^{-3}$
10. $B^+ \rightarrow \bar{D}^{*0} D^{*+} K_s^0$	$(9.2 \pm 1.2) \times 10^{-3}$	10. $B^0 \rightarrow D^{*-} D^{*0} K^+$	$(1.06 \pm 0.09)\%$
11. $B^+ \rightarrow \bar{D}^0 D^0 K^+$	$(1.45 \pm 0.33) \times 10^{-3}$	11. $B^0 \rightarrow D^- D^+ K_s^0$	$(7.5 \pm 1.7) \times 10^{-4}$
12. $B^+ \rightarrow \bar{D}^{*0} D^0 K^+$	$(2.26 \pm 0.23) \times 10^{-3}$	12. $B^0 \rightarrow D^{*-} D^+ K_s^0$	$(6.4 \pm 0.5) \times 10^{-3}$ [iv]
13. $B^+ \rightarrow \bar{D}^{*0} D^{*0} K^+$	$(6.3 \pm 0.5) \times 10^{-3}$	13. $B^0 \rightarrow D^- D^{*+} K_s^0$	$(6.4 \pm 0.5) \times 10^{-3}$ [iv]
14. $B^+ \rightarrow \bar{D}^{*0} D^{*0} K^+$	$(1.12 \pm 0.13)\%$	14. $B^0 \rightarrow D^{*-} D^{*+} K_s^0$	$(8.1 \pm 0.7) \times 10^{-3}$
15. $B^+ \rightarrow D_s^+ \bar{D}^0$	$(7.6 \pm 1.6) \times 10^{-3}$	15. $B^0 \rightarrow D_s^+ D^-$	$(7.2 \pm 0.8) \times 10^{-3}$
16. $B^+ \rightarrow \bar{D}^{*0} \pi^+$	$(4.90 \pm 0.17) \times 10^{-3}$	16. $B^0 \rightarrow D^{*-} \pi^+$	$0.274 \pm 0.13 \times 10^{-3}$
17. $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0$	5×10^{-4} [i]	17. $B^0 \rightarrow D^{*-} \pi^+ \pi^0$	0.835892% [v]
18. $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0 \pi^0$	5×10^{-4} [i]	18. $B^0 \rightarrow D^{*-} \pi^+ \pi^0 \pi^0$	1×10^{-3} [i]
19. $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^-$	$(1.03 \pm 0.12)\%$ [ii]	19. $B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^-$	$(7.21 \pm 0.29) \times 10^{-3}$
20. $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^- \pi^0$	$(1.8 \pm 10.4)\%$	20. $B^0 \rightarrow D^{*-} \pi^+ \pi^+ \pi^- \pi^0$	$(1.76 \pm 0.27)\%$
21. $B^+ \rightarrow D_s^+ \bar{D}^{*0}$	$(1.71 \pm 0.24)\%$	21. $B^0 \rightarrow D_s^{*+} D^-$	$(7.4 \pm 1.6) \times 10^{-3}$
22. $B^+ \rightarrow D_s^+ \bar{D}^{*0}$	$(8.2 \pm 1.7) \times 10^{-3}$	22. $B^0 \rightarrow D_s^+ D^{*-}$	$(8.0 \pm 1.1) \times 10^{-3}$
23. $B^+ \rightarrow \bar{D}^{*0} K^+$	$(3.63 \pm 0.12) \times 10^{-4}$	23. $B^0 \rightarrow D_s^{*+} D^{*-}$	$(1.77 \pm 0.14)\%$
24. $B^+ \rightarrow D^- \pi^+ \pi^+$	$(1.07 \pm 0.05) \times 10^{-5}$	24. $B^0 \rightarrow J/\psi K_s^0$	$(8.68 \pm 0.30) \times 10^{-4}$
25. $B^+ \rightarrow D^- \pi^+ \pi^+ \pi^0$	2×10^{-3} [i]	25. $B^0 \rightarrow J/\psi K^+ \pi^-$	$(1.15 \pm 0.05) \times 10^{-3}$
26. $B^+ \rightarrow J/\psi K^+$	$(1.006 \pm 0.027) \times 10^{-3}$	26. $B^0 \rightarrow J/\psi K_s^0 \pi^+ \pi^-$	$(4.3 \pm 0.4) \times 10^{-4}$
27. $B^+ \rightarrow J/\psi K^+ \pi^+ \pi^-$	$(8.1 \pm 1.3) \times 10^{-4}$	27. $B^0 \rightarrow \Lambda_c^- p \pi^+ \pi^-$	$(1.02 \pm 0.14) \times 10^{-3}$
28. $B^+ \rightarrow J/\psi K^+ \pi^0$	9.4840×10^{-5} [iii]	28. $B^0 \rightarrow \bar{D}^{*0} p \bar{p}$	$(9.9 \pm 1.1) \times 10^{-5}$
29. $B^+ \rightarrow J/\psi K_s^0 \pi^+$	$(1.14 \pm 0.11) \times 10^{-3}$	29. $B^0 \rightarrow D^- p \bar{p} \pi^+$	$(3.32 \pm 0.31) \times 10^{-4}$
30. $B^+ \rightarrow \Lambda_c^- p \pi^+ \pi^0$	$(1.8 \pm 0.6) \times 10^{-3}$	30. $B^0 \rightarrow D^{*-} p \bar{p} \pi^+$	$(4.7 \pm 0.5) \times 10^{-4}$
31. $B^+ \rightarrow \Lambda_c^- p \pi^+ \pi^- \pi^+$	$(2.2 \pm 0.7) \times 10^{-3}$	31. $B^0 \rightarrow \bar{D}^{*0} p \bar{p} \pi^+ \pi^-$	$(3.0 \pm 0.5) \times 10^{-4}$
32. $B^+ \rightarrow \bar{D}^{*0} p \bar{p} \pi^+$	$(3.72 \pm 0.27) \times 10^{-4}$	32. $B^0 \rightarrow \bar{D}^{*0} p \bar{p} \pi^+ \pi^-$	$(1.9 \pm 0.5) \times 10^{-4}$
33. $B^+ \rightarrow \bar{D}^{*0} p \bar{p} \pi^+$	$(3.73 \pm 0.32) \times 10^{-4}$		
34. $B^+ \rightarrow D^+ p \bar{p} \pi^+ \pi^-$	$(1.66 \pm 0.30) \times 10^{-4}$		
35. $B^+ \rightarrow D^{*+} p \bar{p} \pi^+ \pi^-$	$(1.86 \pm 0.25) \times 10^{-4}$		
36. $B^+ \rightarrow \Lambda_c^- p \pi^+$	$(2.3 \pm 0.4) \times 10^{-4}$		

Table 4.3.: The PDG branching fractions of the decays used in the hadronic FEI tag with corresponding assigned decay mode IDs [5]. No branching fraction in the PDG thus the following branching fraction values are the ones implemented in MC14 production: [i] value kept from Belle. [ii] removed from MC production. [iii] calculated from Isospin: $\mathcal{B}(J/\psi K^+ \pi^0) = \mathcal{B}(J/\psi K^0 \pi^+)/2$. [iv] branching fraction is the sum of both $B^0 \rightarrow D^{*-} D^+ K_s^0$ and $D^- D^{*+} K_s^0$. [v] resonant component is subtracted.

4. Analysis Tools

Category	Cuts
	$E_{\text{vis}} > 4 \text{ GeV}$
	$2 < E_{\text{ECL}} < 7 \text{ GeV}$
Tracks:	$ z_0 < 2.0 \text{ cm}$
‘Good tracks’	$ d_0 < 0.5 \text{ cm}$
	$p_t > 0.1 \text{ GeV}$
	$n_{\text{cleaned tracks}} \geq 3$
ECL clusters:	$0.296706 < \theta < 2.61799$
	$E > 0.1 \text{ GeV}$
	$n_{\text{cleaned ECL clusters}} \geq 3$

Table 4.4.: The centrally validated and applied hadronic FEI cuts. (*omitted decay modes 23 for B^+ and 25 for B^0 which are listed in Table 4.3).

to be larger than 0.1 GeV. The effect of these cuts is that fake clusters in which beam-induced backgrounds interact with tracking components of Belle II are removed. Similarly $B\bar{B}$ events are required to have at least three clusters with total energy $E > 0.1 \text{ GeV}$ within the polar angle CDC acceptance region of $0.296706 < \theta < 2.61799$. This track and cluster selection corresponds to the defined ‘good track’ and cluster selection criteria used throughout this thesis. The visible energy of the event in the CM frame, i.e. the total energy of all of the tracks and neutral clusters in the event, is required to be higher than 4 GeV, with the total energy deposited in the ECL E_{ECL} required to be between 2 and 7 GeV. This is to exclude low multiplicity events such as $e^+e^- \rightarrow e^+e^-$. The hadronic FEI skim pre-cuts also implement a restriction on the number of charged particle tracks per event, which is set to be more than or equal to three. This is due to the fact that the hadronic FEI B meson decay channels have a least three charged particles, so this cut suppresses non- $B\bar{B}$ backgrounds.

These pre-cuts are summarized in Table 4.4. Finally, once these kinematic event-level pre-cuts have been applied, the generic FEI algorithm is applied with the B_{tag} being reconstructed with the hadronic decay channels disclosed in Table 4.3.

These selections were refined over many studies to optimize FEI signal retention. The retention of a skim is the fraction of correctly reconstructed B candidates present after the skim is applied to a sample. The author was instrumental in the development of Phase 3 FEI selections.

4. Analysis Tools

MC type	MC decay code	Events generated	Events post-FEI skim	FEI retention
$\bar{B}^0 \rightarrow D^{(*)+} \tau^- \bar{\nu}_\tau$	1193300007	80×10^6	4.56×10^6	5.7%
$\bar{B}^0 \rightarrow D^{**} \tau^- \bar{\nu}_\tau$	1193300008	80×10^6	5.62×10^6	7.0%
$B^0 \rightarrow X_c \ell \nu$	1193710005	40×10^6	3.15×10^6	7.8%
$B^- \rightarrow D^{(*)0} \tau^- \bar{\nu}_\tau$	1293300007	80×10^6	5.76×10^6	7.2%
$B^- \rightarrow D^{**} \tau^- \bar{\nu}_\tau$	1293300008	80×10^6	6.72×10^6	8.4%
$B^+ \rightarrow X_c \ell \nu$	1293710005	40×10^6	2.91×10^6	7.2%
$\Upsilon(4S) \rightarrow B^+ B^-$	(charged)	0.486×10^9	6.937×10^7	14.3%
$\Upsilon(4S) \rightarrow B^0 \bar{B}^0$	(mixed)	0.459×10^9	5.222×10^7	11.4%
$u\bar{u}$	(uubar)	1.445×10^9	1.615×10^8	11.2%
$d\bar{d}$	(ddbar)	0.361×10^9	3.892×10^7	10.8%
$s\bar{s}$	(ssbar)	0.345×10^9	3.224×10^7	9.3%
$c\bar{c}$	(ccbar)	1.120×10^9	1.875×10^8	16.7%
$\tau^+ \tau^-$	(taupair)	0.827×10^9	0	0%
$\mu^+ \mu^-$	(mupair)	1.033×10^9	0	0%

Table 4.5.: The retention of the Hadronic FEI skim by MC type. The generic MC number of events generated corresponds to 1 ab^{-1} of simulated data. The different signal MC samples used in this analysis have the Belle II signal MC code stated for reproducibility.

Table 4.5 shows that the retention of the FEI skim for signal MC alongside mixed and charged samples is much higher than for continuum, as intended, and is especially effective in removing all $e^+ e^- \rightarrow \tau^+ \tau^-$ continuum events. The hadronic FEI skim was performed on the generic MC, signal MC and data used in this analysis. As 100 fb^{-1} of the generic MC was used to train the FEI, only 900 fb^{-1} was available for use in this analysis. In addition, at the time of final analysis submission to the computing resources, 100 fb^{-1} of generic MC was not available for use. Thus 800 fb^{-1} was used to simulate this full analysis, with little effect on the statistical uncertainties and systematic errors.

The inclusive $B \rightarrow X \ell^- \bar{\nu}_\ell$ decay in a data control sample of 34.6 fb^{-1} was used to validate MC13 in Reference [56]. This mode was chosen due to its large branching fraction of $(10.33 \pm 0.28)\%$ and $(10.65 \pm 0.16)\%$, for B^0 and B^+ , respectively, totalling approximately 20% considering both e and μ modes. The FEI tag side reconstruction efficiency in MC and data differs significantly due to the many multivariate classifiers and the large number

of decay chains used in training and application. The data-MC efficiency of the hadronic FEI at values of $\mathcal{P}_{\mathcal{B}_{\text{tag}}} > 0.001$, $\mathcal{P}_{\mathcal{B}_{\text{tag}}} > 0.01$ and $\mathcal{P}_{\mathcal{B}_{\text{tag}}} > 0.1$ were determined. These are shown in Figure 4.4.

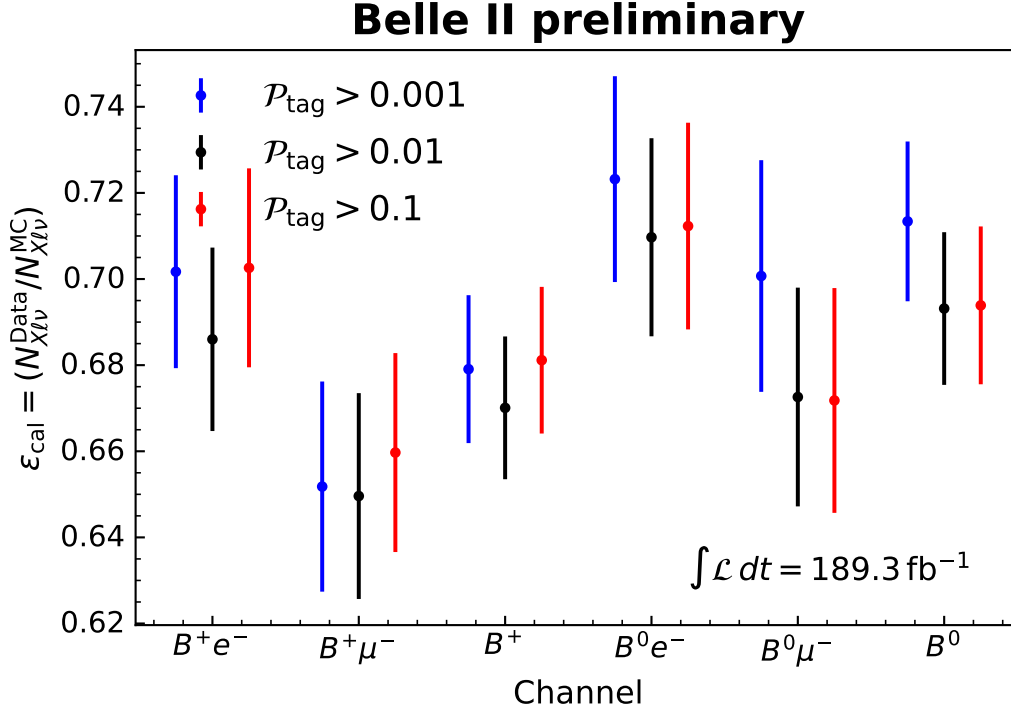


Figure 4.4.: The FEI calibration factors required to correct for the difference in data-MC efficiency as a function of $\mathcal{P}_{\mathcal{B}_{\text{tag}}}$. The calibrations are separated into charged and neutral B mesons, and then electrons, muons and combined electrons and muons. This is a preliminary result, updated from Reference [56].

From this it was recommended that the generic MC samples should be renormalised to data using a factor for calibration, established using the $B \rightarrow X\ell^-\bar{\nu}_\ell$ decay. The calibration factor used to correct for the efficiency is equal to the ratio of the yield of signal channel ($B \rightarrow X\ell^-\bar{\nu}_\ell$) events in data to that of MC. The yields were extracted from a fit that maximised a binned likelihood [56] to the lepton momentum in the CMS frame as shown in Figure 4.5. Preliminary calibration factors calculated for MC14ri with $(189.8804 \pm 0.0131) \text{ fb}^{-1}$ of data have been developed and are shown in Table 4.6. The method differs slightly in the use of $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays for calibration. This decay has a high purity factor and a large branching fraction allowing for an adequate amount of

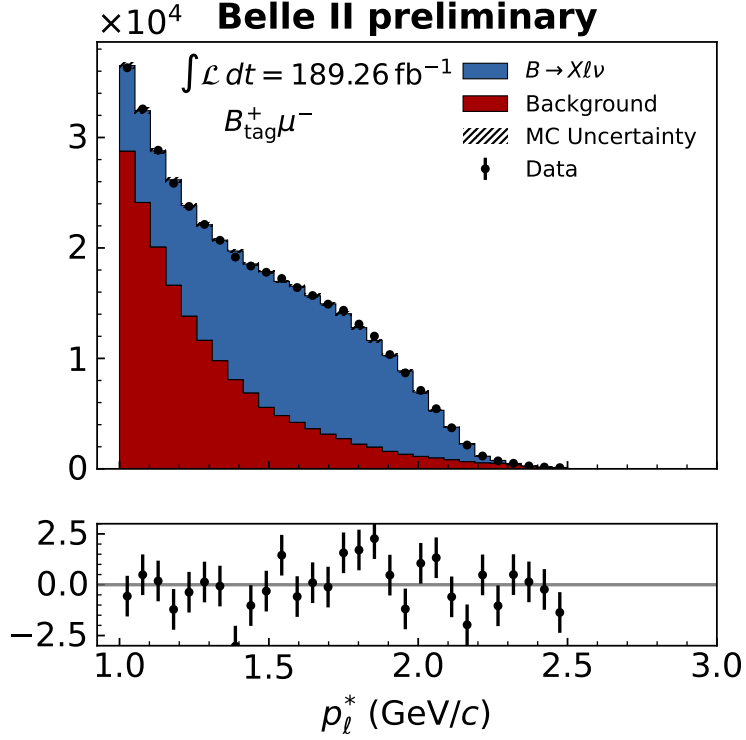


Figure 4.5.: The lepton momentum in the CM frame for $B \rightarrow X \ell^- \bar{\nu}_\ell$ used to extract the FEI calibration factor, in this case using the B_{tag} and μ . Generic MC is classed as $B \rightarrow X \ell^- \bar{\nu}_\ell$, and background. Data is overlaid as points with statistical uncertainties shown as error bars. This result is updated from Reference [56].

statistics for the calibration. The calibration depends on the number of decay channels used in the FEI and the potentially large systematic uncertainties introduced by certain D^* decay modes that are reconstructed, but this is not largely different from the previous more generic calibration channel $B \rightarrow X \ell^- \bar{\nu}_\ell$. The $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ calibration factor results in a significant improvement in the level of data-MC agreement, thus this analysis will use these newly available calibration factors. There are no decay specific or independent validation decays evaluated for the $B \rightarrow D \ell^- \bar{\nu}_\ell$ decay, thus this analysis is restricted to using the $B \rightarrow X \ell^- \bar{\nu}_\ell$ calibration factors.

In conclusion, the overall data-MC efficiency discrepancy caused by the application of the FEI algorithm is corrected for using the calibration factors shown in Table 4.6. In this analysis the purity of the B_{tag} is optimized through the selection of B_{tag} candidates in accordance

Decay Channel Correction	$\bar{B}^0 \rightarrow X\mu^-\bar{\nu}$	$\bar{B}^0 \rightarrow Xe^-\bar{\nu}$	$B^- \rightarrow X\mu^-\bar{\nu}$	$B^- \rightarrow Xe^-\bar{\nu}$
	0.6422	0.6968	0.6715	0.6483
	$\pm 3.24\%$ (sys.)	$\pm 3.88\%$ (sys.)	$\pm 3.10\%$ (sys.)	$\pm 3.71\%$ (sys.)
	$\bar{B}^0 \rightarrow D^*\mu^-\bar{\nu}$	$\bar{B}^0 \rightarrow D^*e^-\bar{\nu}$	$B^- \rightarrow D^*\mu^-\bar{\nu}$	$B^- \rightarrow D^*e^-\bar{\nu}$
	0.6259	0.7422	0.6422	0.6735
	± 0.024 (stat.) $\pm 6\%$ (sys.)	± 0.031 (stat.) $\pm 6\%$ (sys.)	± 0.024 (stat.) $\pm 6\%$ (sys.)	± 0.034 (stat.) $\pm 8\%$ (sys.)

Table 4.6.: The recently updated FEI efficiency corrections for MC14ri_a calibrated for $B \rightarrow D^*\ell^-\bar{\nu}_\ell$ and $B \rightarrow X\ell^-\bar{\nu}_\ell$.

with the most efficient MC14 FEI calibration factor (where `SignalProbability` > 0.001). Further discussion of the pitfalls and future development of the calibration factors can be found in Chapter 9.

4.5. Truth information and truth matching

It is important for this analysis that we have an accurate knowledge of how the decay products interact within the detector, which is why one must simulate these interactions first using MC simulation. The exact same reconstruction and analysis is performed on MC as is performed on data, including the use of the same algorithms. The difference between MC and data is that one can trace the detector response to an MC particle throughout the simulation and know whether the particles that are reconstructed were identified correctly.

Once the lists of final state particle candidates are reconstructed from the track and cluster information, and are reconstructed into the desired signal decay via intermediate particles, truth matching can be performed. This is where the generated particle is compared to the reconstructed particle and is determined to be correctly reconstructed or not. This is referred to as comparing the reconstructed particle with “MC truth” particle. Naturally we cannot determine the truth of data, so improving our reconstruction efficiency using truth matching checks is a key aspect of improving the quality of our data.

Here are a few of the many different errors that can occur during the reconstruction of a particle candidate:

- **Misidentification** e.g. the particle used in the reconstruction was thought to be a

4. Analysis Tools

pion, but was actually a kaon.

- **Missing decay products** e.g. the reconstruction combined a kaon and a pion into a D meson, but actually the D meson decayed as $D \rightarrow K\pi\pi$.
- **A final state radiation photon was missed** e.g. the reconstruction resulted in a D^0 meson, but in reality the particle was a $D^{*0} \rightarrow D^0\gamma$ and a γ was missed.
- **In flight decay** e.g. a pion decayed into a muon, but the track of the pion was reconstructed as a muon.

In `basf2`, a simulated particle is reconstructed in MC simulation as a `Particle` object, which is then related to a `MCParticle` object. When MC matching is performed, the software returns the true PDG identification number of the particle, which has been pre-assigned[5]. For example, the PDG ID of a B^+ is 521, with its charge conjugate B^- taking the negative of this value. This variable is called `mcPDG` in `basf2`, and it is set to the PDG code of the first common mother `MCParticle` of the daughters of the `Particle`.

4.5.1. Topology analysis using TopoAna

Topological analysis is one method that enables comprehensive understanding of the simulation of signal and background of an experiment. Through this understanding, it is possible to optimize signal selection and background suppression for an analysis. The decay of a B meson is relatively straightforward when one considers a certain signal decay consisting of composite particles, but when one then considers the subsequent decays of these composite particles, the topological mapping becomes much more complicated. The topological decay information of a decay in `basf2` is simulated and stored in MC, but it is a gargantuan task for an analyst to decipher manually the decay products of an event due to the large number of combinations and complexities of physics processes. To highlight this, two example topological maps of e^+e^- decays of singular processes are displayed in Figure 4.6.

`TopoAna` is a generic tool for topological analysis of event or candidate based decays in the inclusive MC samples of high energy physics [57]. It is an offline analysis tool that is

4. Analysis Tools

independent of **basf2**. In this analysis, **TopoAna** is used to store the raw topological truth information for an event using an interface between **basf2** and **TopoAna** called **MCGenTopo**. It is specific to generator-level MC truth. In **TopoAna** every event is mapped out using the raw topology truth information: the number of particles in an event, the PDG MC particle codes of each of the particles and the mother indices of those particles. Using this information it categorizes each type of event in a sample, flags each event with its decay type and also counts the number of occurrences of a decay type.

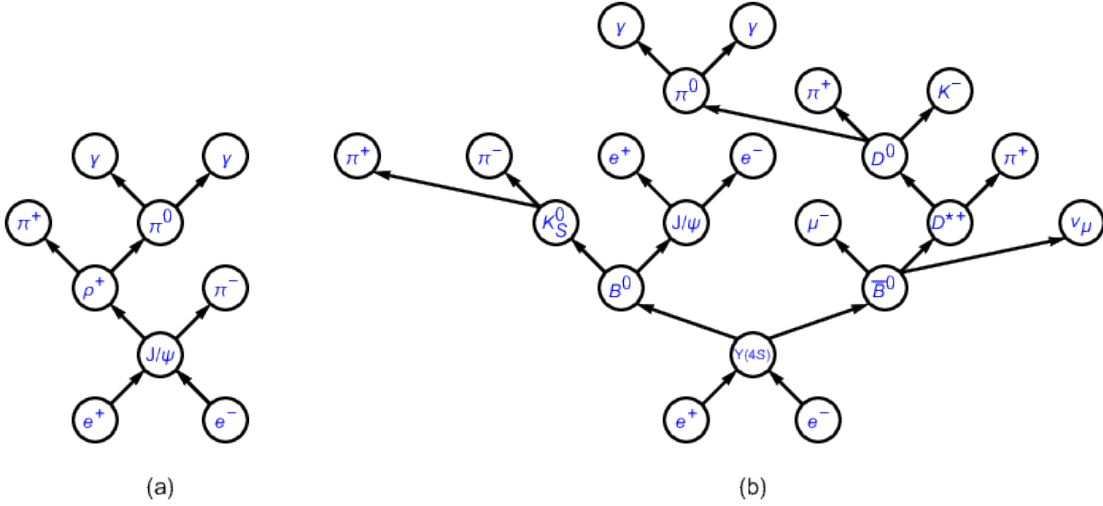


Figure 4.6.: A schematic depicting two topology diagrams of processes occurring at e^+e^- colliders. Diagram (a) is the $e^+e^- \rightarrow J/\psi$ decay, while diagram (b) contains one possible iteration of the decays studied in this analysis: an $Y(4S) \rightarrow B^0 \bar{B}^0$ decay where the $\bar{B}^0$ decays to $D^{*+} \mu^- \bar{\nu}_\mu$ [55].

Detailed knowledge of the topology of B meson decays in a region selected by signal selections allows identification of the main backgrounds from the chosen signal decays. Background events that peak in the same regions of distributions that we expect signal events to peak in are termed ‘peaking backgrounds’. These are often decays that are similar in signature to the chosen signal decay of study. This separation of signal and background allows analysts to further optimize selection criteria and estimate the yields of peaking backgrounds alongside their statistical and systematic uncertainties, even when they cannot be suppressed.

The two main functionalities of **TopoAna**, component analysis and signal identification,

4. Analysis Tools

are used for this analysis. In component analysis, every channel of B^+ and B^0 decays present in the MC are identified and given a flag. In signal identification, the analyst defines which decays **TopoAna** should flag. The benefit of using a combination of the two functionalities is that the entire MC sample can be categorized and analysed using component analysis, whilst the analyst can define their signal decays to include channels that contain extra particles. An example of this is when there is a production of an extra final state radiation (FSR) photon γ in addition to the decay particles of $B \rightarrow D^* \ell^- \bar{\nu}_\ell$. In signal identification the analyst can chose to ignore the FSR to different degrees of strictness, and continue to consider this decay as one of interest to the analysis. With signal identification, inclusive and cascade decays can also be selected. Cascade decays include the definition of multiple, consecutive decays. For example, one can choose to cascade a $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decay with the decay of $D^{*+} \rightarrow D^+ \pi^0$ and inclusively include all subsequent D^+ decays.

TopoAna performs component analysis indiscriminatingly over all truth instances of the same particles in the same events. Due to this and the fact it is independent of **basf2**, it does not consider whether a particle is correctly reconstructed. Therefore particles that have generator truth level information that has been incorrectly reconstructed are still flagged as valid decays.

Due to the limited detection efficiencies of both the B_{tag} (Section 4.4) and B_{sig} (discussed later in Section 7.4.1), the purity is very low, resulting in truth information that does not match the reconstructed particle candidates. FEI also typically results in multiple tag candidates, only one (or none) of which are correctly reconstructed. Thus, there are often multiple B_{tag} meson candidates for each event that are not correctly reconstructed.

To mitigate this, **TopoAna** can be combined with reconstruction information. **TopoAna** recently implemented the functionality of being able to additionally require that a particle has been successfully reconstructed using information from **basf2**. In this analysis, **TopoAna** is required to match the index of the truth signal B meson to the reconstructed candidate. Using a combination of indexing and the number of instances of the i^{th} specified particle or decay branch found in each event, this analysis studies correctly reconstructed B_{sig} decays topologically.

5. The measurement of $B \rightarrow D^{(*)}l^{-}\bar{\nu}_l$ at Belle II

This chapter motivates the study of $B \rightarrow D^{(*)}l^{-}\bar{\nu}_l$ at Belle II, then discusses the unique features of this analysis.

5.1. Motivation of the measurement of $B \rightarrow D^{(*)}l^{-}\bar{\nu}_l$ at Belle II

$B \rightarrow D^{(*)}l^{-}\bar{\nu}_l$ ^[i] is one of the best semileptonic decays available for probing the SM for NP via LFU, as introduced in Chapter 2. Belle II has unique capabilities with which to measure the branching fraction of $B \rightarrow D^{(*)}l^{-}\bar{\nu}_l$, and therefore $R(D)$ and $R(D^*)$, to the best precision to-date.

Due to the nano-beam scheme discussed in Section 3.1, the beam induced background in Belle II is much higher than in previous B factories such as BaBar and Belle. This is, in turn, suppressed by utilising the upgrades of the detector alongside improved analysis tools and techniques such as tagging (Chapter 4). Alongside its B factory status, Belle II is a high statistics and clean environment, and therefore an ideal probe for evidence of physics BSM.

This still presents a challenging environment in which to study decays with neutrinos in the final state which traverse through Belle II undetected, also known as decays with missing energy. Access to multineutrino final states of the Belle II detector is essential to the measurement of $B \rightarrow D^{(*)}l^{-}\bar{\nu}_l$ due to the fact that the τ final state can include one

^[i]A reminder that $l = e, \mu, \tau$ and $\ell = e, \mu$.

5. The measurement of $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ at Belle II

or more neutrinos, i.e. from $\tau^+ \rightarrow e^+\nu_e\nu_\tau$ or $\tau^+ \rightarrow \pi^+\nu_\tau$. The disadvantage of trying to measure decays with missing energy is that they are usually limited in their use of kinematic constraints that are otherwise used to reject background events. Fortunately, a speciality of Belle II is that there are unique experimental conditions that allow full reconstruction of the decay kinematics, and therefore access to multineutrino final states of signal B decays.

Firstly, the known production process of $B\bar{B}$ pairs at the $\Upsilon(4S)$ resonance limits the momentum of the $B\bar{B}$ pair. The recoil of the non-signal B meson produced in the event can be reconstructed and utilized to allow precise knowledge of the properties of the signal B , such as flavour and momentum. Secondly, the detector encloses the interaction region almost hermetically (few potentially detectable particles will escape detection). From these two properties, the four-momentum of undetected particles can be inferred from the measured momenta and energies of all of the other particles produced in the e^+e^- collision, through imposing energy and momentum conservation. This is a significant benefit for studying decays with missing energy like $B \rightarrow D^{(*)}l^-\bar{\nu}_l$.

Background separation can also be implemented through access to the beam-energy-constrained mass M_{bc} of the B meson, calculated as

$$M_{bc} = \sqrt{E_{\text{beam}}^{*2} - |\vec{p}_{B_{\text{tag}}}^*|^2}, \quad (5.1)$$

where E_{beam}^* is the beam energy in the CM frame corresponding to half of the energy available in the CM system, and $\vec{p}_{B_{\text{tag}}}^*$ is the B_{tag} momentum in the CM frame. E_{beam} is used instead of the measured B energy to avoid introducing experimental resolution uncertainties. The asterisk indicates that these are CM frame variables. M_{bc} actually corresponds to the invariant mass of the B meson, reconstructed from its daughter particle 4-vectors.

In addition, the energy difference ΔE between the measured momenta and energies of *visible* decay products and the beam energy provides more background separation.

$$\Delta E = E_{B_{\text{tag}}}^* - E_{\text{beam}}^*, \quad (5.2)$$

5. The measurement of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ at Belle II

where $E_{B_{\text{tag}}}$ is the reconstructed energy of the B_{tag} . M_{bc} peaks at the nominal B meson mass and ΔE peaks at 0 for correctly reconstructed B mesons.

Furthermore, Belle II can conduct multiple independent FEI tagged and untagged measurements of $R(D)$ and $R(D^*)$. There are also various methodologies which can potentially be used to perform these measurements. The B_{tag} mesons can be reconstructed hadronically or semileptonically, and it is possible to consider both hadronic and leptonic decay channels of the τ lepton. Therefore, four orthogonal statistically independent tagged measurements of $R(D)$ and $R(D^*)$ can be evaluated, with the lowest systematic uncertainties appearing in the hadronic tagging of the B meson paired with either leptonic or hadronic τ decays. Currently the leptonic τ decays have better background rejection than that of hadronic decays. To-date, only leptonic τ decays have been measured in this analysis, but Belle II anticipates measurements expanding to the τ hadronic channels in the future.

Particle reconstruction at Belle II is highly efficient for charged particles, for example, hadronic reconstructions of charged K and π particles. The tracking components of the Belle II detector, as mentioned in Chapter 3, are also highly efficient. Finally, charged leptons have a highly identifiable signature at Belle II, and can be reconstructed efficiently.

Belle II also has state-of-the-art neutral particle detection. The photon detection efficiency uncertainties are currently under 1% in the barrel region for $E_\gamma > 250$ MeV, alongside newly developed BDTs to distinguish signal photons from beam backgrounds and fake photons [42]. The detection of K_L^0 decays serves primarily as a veto and can be utilised well for missing energy measurements. The efficiency of the KLM is under study and will require much larger datasets. For this analysis, optimisation of D^* and D mesons whose decays include a neutral particle is still underway, as will be seen in Section 6.2.6.

Finally, for measurements of $R(D)$ and $R(D^*)$, the high statistics of the final Belle II dataset compared to previous B factories combined with the cancelled uncertainties from $|V_{cb}|$ and hadronic FF uncertainties (see Section 2.2.3), will allow a decrease in the uncertainties of $R(D^*)$ from more than 20% to 3%, even in a pessimistic scenario [58]. Figure 5.1 displays the expected increase in sensitivity as the size of the Belle II dataset grows. Notice that after approximately 2026 in the pessimistic scenario of Figure 5.1, the uncertainty for $R(D^*)$ plateaus at $\sim 3\%$ while the $R(D)$ ratio takes much longer to reach

5. The measurement of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ at Belle II

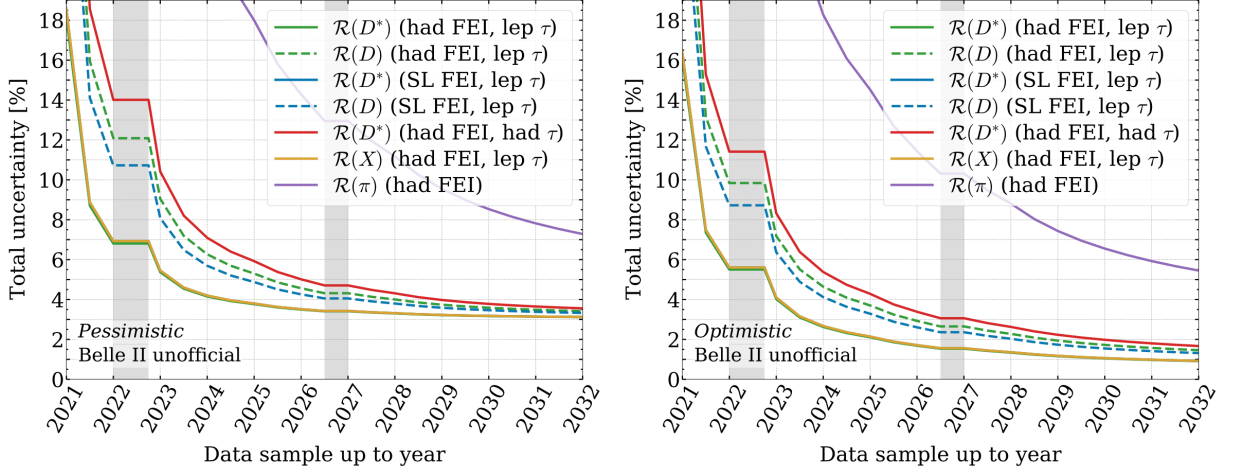


Figure 5.1.: The expected total uncertainty of various branching fraction ratios at Belle II over time. Time represents the expected dataset size and is subject to change. The pessimistic scenario (left) represents a prediction using current statistical and systematic errors that contain the irreducible 3% uncertainty. The optimistic scenario (right) lowered the irreducible systematic to 0.5% and assumes a 50% increase of the effectiveness of the exclusive tagging algorithms. The grey bands represent data collection downtime. The relevant uncertainties for this analysis are the $R(D)$ and $R(D^*)$ (had FEI, lep τ) ratios in green. Figure obtained from Reference [58].

$\sim 3\%$ uncertainty. This is due to the cleaner signature of $R(D^*)$, but also fewer cross-feed contributions that are much more present in $R(D)$. There is promise for the optimistic scenario in Figure 5.1 as there are numerous machine learning and deep learning tools in development that are improving the precision of measurements [59].

Figure 5.2 shows the predicted statistical and systematic uncertainties for the 5 ab^{-1} and 50 ab^{-1} Belle II dataset. Thus with just 5 ab^{-1} of data, Belle II will be able to produce a world leading measurement of $R(D)$ and $R(D^*)$.

5.2. The novel aspects of this analysis

This analysis reports on one of the first tests of LFU at Belle II using exclusive measurements of the denominators of $R(D)$ and $R(D^*)$. It originally contributed to the first published Belle II results of the hadronic tagged $B^0 \rightarrow D^{*+}[D^0 \pi^+] \ell^- \bar{\nu}$ branching fraction using the “golden decay channel” where $D^0 \rightarrow K^- \pi^+$ [1]. This channel has golden decay status due to the high statistics available as a result of a relatively high branching fraction,

5. The measurement of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ at Belle II

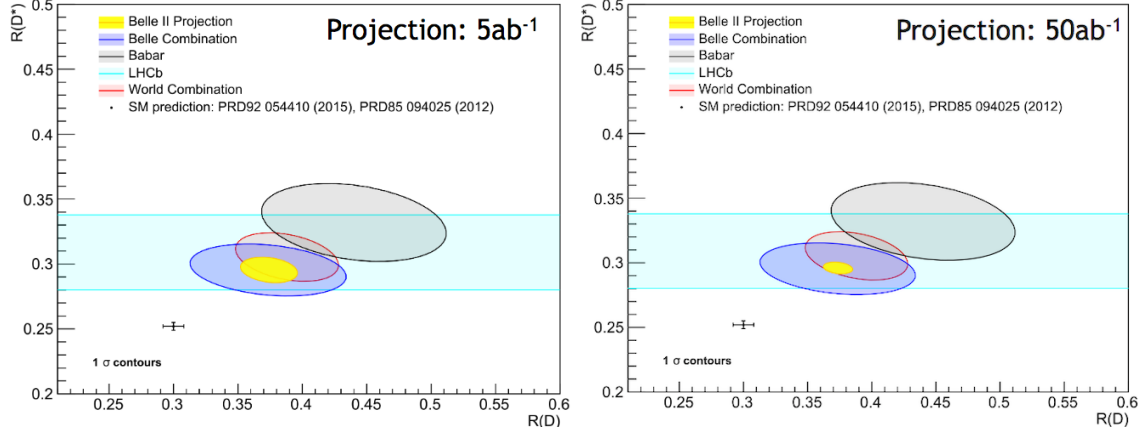


Figure 5.2.: The projection of the $R(D)$ and $R(D^*)$ measurements at Belle II for 5 ab^{-1} (left) and 50 ab^{-1} (right). Figure obtained from Reference [42].

and the best available resolution at Belle II at the time of measurement.

In this thesis the golden decay channel study is expanded to consider supplemental D^* and D^0 decay channels for the B^0 meson. The study is also furthered to include the B^+ meson and its decay channels. The optimal decay channels of the D^* and D mesons used for the exclusive measurement of the $R(D)$ and $R(D^*)$ denominators are detailed in Table 5.1.

This analysis advances the previous study with the use of the larger (and current) Belle II dataset of $(189.8804 \pm 0.0131) \text{ fb}^{-1}$. Amongst the benefits of the use of data with higher statistics, there are also the benefits of the improved understanding of data and MC with the passage of time. For example, for higher (lower) momenta the track reconstruction data/MC ratio was consistent with 1.0 (0.9), with a 0.30% per-track systematic uncertainty for mid-to-high momentum tracking efficiencies. In comparison, the MC used in early stage of Phase 3 of Reference [1] had a 0.91% per-track systematic uncertainty.

In addition, the signal selections for the B^0 and B^0 channels have been studied and optimized using the newer MC14a.ri and the larger dataset than was available for Reference [1]. Optimizations of note include that of continuum suppression, loosened D^* meson mass cuts, improved best candidate selection, PID corrections, and ROE and E_{ECL} development that are further discussed in Chapter 6.

This analysis mirrors and develops a BDT MVA used in the Babar measurements of

5. The measurement of $B \rightarrow D^{(*)}l^- \bar{\nu}_l$ at Belle II

Decay channel			$\mathcal{B}(\%)$	
$D^0 \rightarrow$	$K^- \pi^+$		3.95 ± 0.031	
	$K_S^0 \pi^0$		1.240 ± 0.022	
	$K^- \pi^+ \pi^0$		14.4 ± 0.5	
	$K^- \pi^+ \pi^- \pi^+$		8.23 ± 0.14	
	$K_S^0 \pi^+ \pi^-$		2.80 ± 0.18	
	$K_S^0 \pi^+ \pi^- \pi^0$		5.2 ± 0.6	
	$K^- K^+$		0.408 ± 0.006	
	$\pi^- \pi^+$		0.1455 ± 0.0024	
Total			36.3655 ± 0.8145	
$D^+ \rightarrow$	$K^- \pi^+ \pi^+$		9.38 ± 0.16	
	$K^- \pi^+ \pi^+ \pi^0$		6.25 ± 0.18	
	$K_S^0 \pi^+$		1.562 ± 0.031	
	$K_S^0 \pi^+ \pi^- \pi^+$		3.10 ± 0.09	
	$K_S^0 \pi^+ \pi^0$		7.36 ± 0.21	
	$K_S^0 K^+$		0.304 ± 0.009	
	$K^- K^+ \pi^+$		0.968 ± 0.018	
Total			28.924 ± 0.334	

Decay channel			$\mathcal{B}(\%)$	
$D^*(2007)^0 \rightarrow$	$D^0 \pi^0$		64.7 ± 0.9	
	$D^0 \gamma$		35.3 ± 0.9	
Total			100.0 ± 1.3	
$D^*(2010)^+ \rightarrow$	$D^0 \pi^+$		67.7 ± 0.5	
	$D^+ \pi^0$		30.7 ± 0.5	
Total			98.4 ± 0.7	

Table 5.1.: A summary of the D^* and D meson decay channels and their corresponding branching fractions selected for the analysis. Values obtained from Reference [5].

$B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ [60] by extending to further decay channels, by evaluating the potential of additional training variables, by choosing the optimal type of MVA model utilized and by using signal boosted MC to enhance the performance of the MVA. The performance of the MVA is improved through the use of the larger Belle II data sample and through the use of the Belle II FEI algorithm. Section 6.4 discusses this in detail.

The four branching fractions of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ ($\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$, $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$, $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$) are then measured as the denominators of $R(D)$ and $R(D^*)$ in Chapter 7. This analysis will contribute towards Belle II's first competitive $R(D)$ and $R(D^*)$ measurements, but that measurement is beyond the scope of this work. Publication of the first $R(D)$ and $R(D^*)$ result is predicted by early 2023.

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

This chapter details the selections and methods used in this analysis. Firstly, the components of the analysis are separated into online and offline selections. The online selections are those performed first using `basf2`. The offline selections include additional selections after further studies are performed. A D^{**} background suppression MVA is developed and applied. Finally, the resulting signal and normalisation yields are calculated and presented with application of the optimized D^{**} background suppression MVA classifier selections.

6.1. The analysis procedure

This analysis is an optimization of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ as the normalisation decay channel of the variables $R(D)$ and $R(D^*)$ using hadronic tagging techniques. As discussed in Chapter 2, $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ is the denominator in the calculation of $R(D^{(*)})$. A caveat to this optimization is that in the measurements of $R(D^{(*)})$, the same selections must be applied to both the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ and the $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ decay channels. Thus, the decisions taken in this thesis were performed after studies on both the $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ and the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays^[i].

As discussed in Chapter 4, multiple pieces of software were utilized for this analysis. Table 6.1 summarizes the releases/versions of each of the tools used in this analysis.

As the B_{tag} and B_{sig} of each event are produced approximately at rest in the CMS frame, the sum of their four-momenta would equal the four-momentum of the colliding beams (or the produced $\Upsilon(4S)$). When the B_{sig} decays through the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay channel,

^[i]Normalisation ($B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$) and signal ($B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$) decay channels will be distinguished within MC after the B_{sig} is fully reconstructed. Until that point, figures consider the normalisation and signal combined $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ sample (with $l = e, \mu, \tau$).

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

Software	Release
basf2	light-2110-tartarus
Hadronic FEI skim	FEIv4_2021_MC14_release_05_01_12
Semi-tauonic framework	Version 19

Table 6.1.: A summary of the software and relevant releases/versions used for this analysis.

the two detectable or reconstructable particles at Belle II are the $D^{(*)}$ and ℓ . As previously discussed, it is not possible to measure p_ν directly, thus the neutrino(s) in an event manifest as missing momentum at the Belle II detector. With correct reconstruction, the missing four-momentum vector p_{miss} accurately represents the neutrinos present in each event. By squaring p_{miss} the relation $p_{\text{miss}}^2 = m_{\text{miss}}^2$ can be implemented and one can measure the four-momentum squared of the neutrino through measuring the missing mass of an event squared m_{miss}^2 . This is defined, for this analysis, as

$$m_{\text{miss}}^2 = p_{\text{miss}}^2 = \left(p_{e^+e^-} - p_{B_{\text{tag}}} - p_{B_{\text{sig}}} \right)^2 \quad (6.1)$$

$$= \left(p_{e^+e^-} - p_{B_{\text{tag}}} - p_{D^*} - p_\ell \right)^2, \quad (6.2)$$

where $p_{e^+e^-}$ is the known four-momentum of the colliding beams, and $p_{B_{\text{sig}}}$ and $p_{B_{\text{tag}}}$ are the *measured* four-momenta of the respective B mesons. Due to the experimental constraints, the m_{miss}^2 can also be equated to

$$m_{\text{miss}}^2 = p_{\text{miss}}^2 = \left(p_{B_{\text{sig}}} - p_{D^*} - p_\ell \right)^2, \quad (6.3)$$

to simplify the calculation.

For m_{miss}^2 , one expects two distinct distributions for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ and $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$. In the particular case of correctly reconstructed normalisation events, the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay, there is only one neutrino that cannot be directly detected. This means that the m_{miss}^2 distribution is expected to peak at zero. This is in contrast to $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$, which has a m_{miss}^2 distribution spread in the range -3 GeV to 10 GeV due to its multi-neutrino final state due to the τ decay. This can be better understood through altering Equation

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

6.3 to display the missing momentum contributions explicitly:

$$p_{\text{miss}}^2 = \left(p_{B_{\text{sig}}} - p_{D^*} - p_\ell \right)^2 = \left(p_{\nu_1} + p_{\nu_2} + p_{\nu_3} (+ \dots) \right)^2 \quad (6.4)$$

where the p_{ν_i} ($i = 1, 2, 3, \dots$) represent the possible extra neutrinos present in the event. The m_{miss}^2 here does not quite describe the “mass” of a particle as it does in the single missing neutrino case because it is not the square of the four-momentum of a single particle, however it will still be referred to as such.

Other semi-leptonic decays and combinatorial backgrounds also have a smeared distribution about zero. m_{miss}^2 will be utilized as the yield extraction variable for this analysis due to its background separation and distinct identification of normalisation events compared to signal events.

To mitigate experimental bias, this analysis is performed in the side-band regions of the signal decay such that access to data in the signal region is only allowed once the full $R(D)$ and $R(D^*)$ analysis is finalized. The signal region is defined to be $m_{\text{miss}}^2 > 3 \text{ GeV}^2$, as this is the region in which the $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ decay distribution will be predominantly present. In other words, this analysis considers the side-band region for the $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ decays, which is the same as the signal region for the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decays.

6.2. Online selections

The following section details the online selections for particle candidates and their reconstruction requirements. The online selections are those performed using the **basf2** software. This includes the reconstruction of particles, the use of FEI tagged B mesons, the choice of the B_{sig} decay channel and subsequent daughter decay channels, the application of continuum suppression, and the addition of a few select cuts. For ease of comparison, a summary of all of the online selections is included at the end of the section (Table 6.2).

6.2.1. B_{tag} reconstruction

The first stage of selection is the application of the trained hadronic FEI (see Section 4.4.1) on the data and MC samples. In addition to the standard FEI selections, to further reduce the large number of B candidates per event and the backgrounds to the normalisation and signal modes, each B_{tag} candidate is required to have an M_{bc} greater than 5.27 GeV (Figure 6.1) and $\Delta E \in [-0.15, 0.1]$ GeV (Figure 6.2). M_{bc} and ΔE were previously defined in Section 5.1.

The $\mathcal{P}_{B_{\text{tag}}}$ of each B_{tag} (Figure 6.3) is required to be greater than 0.001^[ii] corresponding to the calibration factor study discussed in Section 4.4.1. The calibration factors in Table 4.6 are applied to all MC samples in this analysis. These additional skim-level selections are summarized in Table 6.2. Furthermore, every track and neutral cluster that is used in the B_{tag} reconstruction is considered to be in the ROE. All remaining tracks and clusters in an event that are used in the B_{tag} reconstruction are then masked and prevented from being reused in the reconstruction of the B_{sig} .

6.2.2. Tracking

In Phase 3 Belle II data, a shift of invariant mass peak positions for D^0 , D^+ , J/ψ , K_S^0 and Λ_C is observed by order of a few percent of their expected invariant mass peaks. The cause of these shifts was found to be predominantly due to a tracking momentum bias derived from the biased measured magnetic field map that was used in the data reconstruction process. As recommended by the Belle II tracking group, an official global correction factor of $0.99971^{+0.00035}_{-0.00055}$ (sys.) is applied to the yields of data.

6.2.3. π selections

Neutral pions (π^0) must be reconstructed by combining two photons (γ) from the pre-defined particle selections as determined by the performance group at Belle II. Specifically in this analysis, π^0 candidates are obtained by using the official Belle II preselected 40% efficiency γ list and 40% efficiency π^0 list shown in Table 6.3. The daughter γ candidates

^[ii] $\log_{10}(0.001) = -3.0$

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

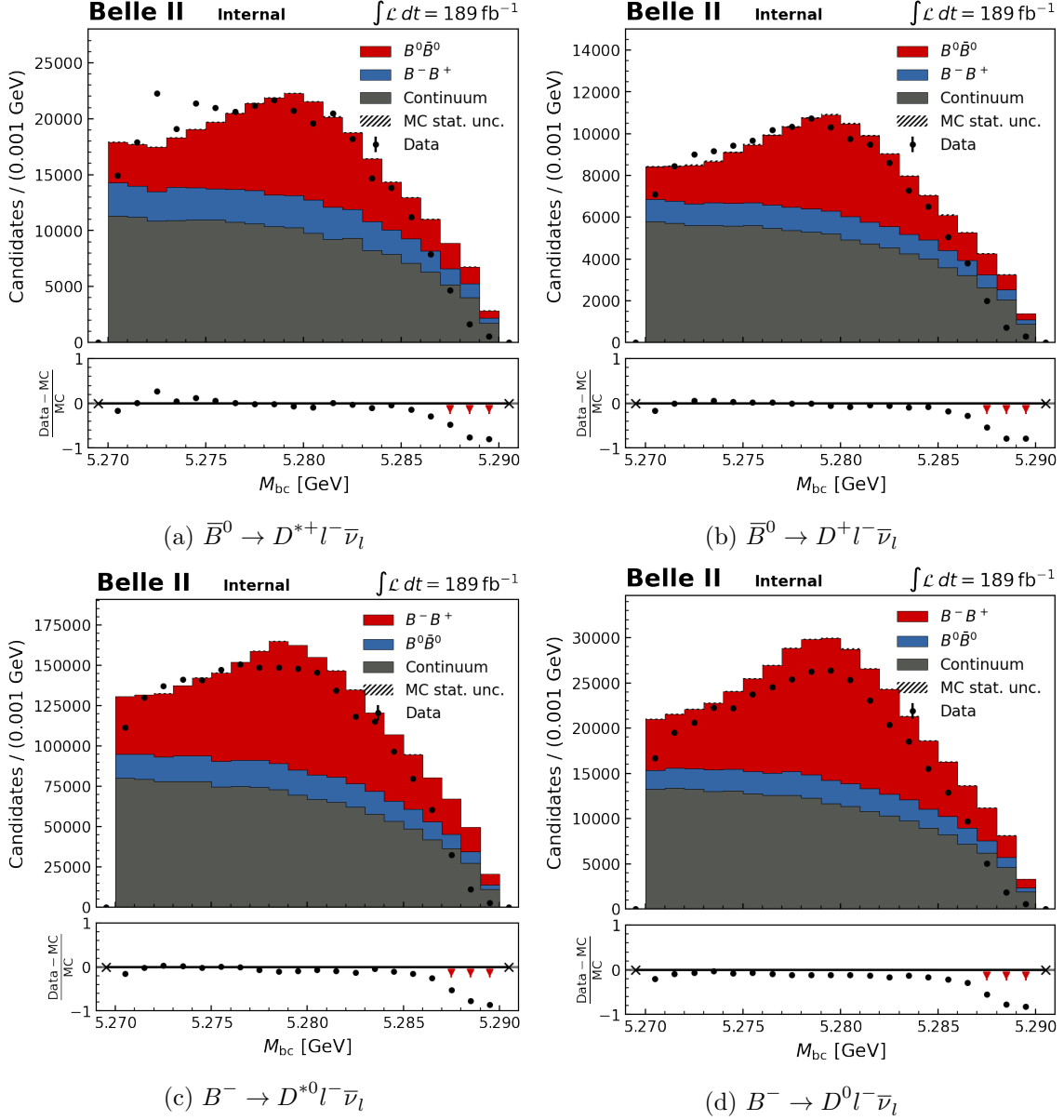


Figure 6.1.: Distributions of M_{bc} for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The red arrows depict the bins where the residual pull is more than 1σ away from data-MC agreement.

must be located in the CDC acceptance region and the weighted sum of all crystals in an isolated ECL cluster must have more than 1.5 hits. An energy cut is applied on the γ with a dependence on detector region:

- Clusters in the ECL forward region (region 1) must have an energy greater than 80 MeV.

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

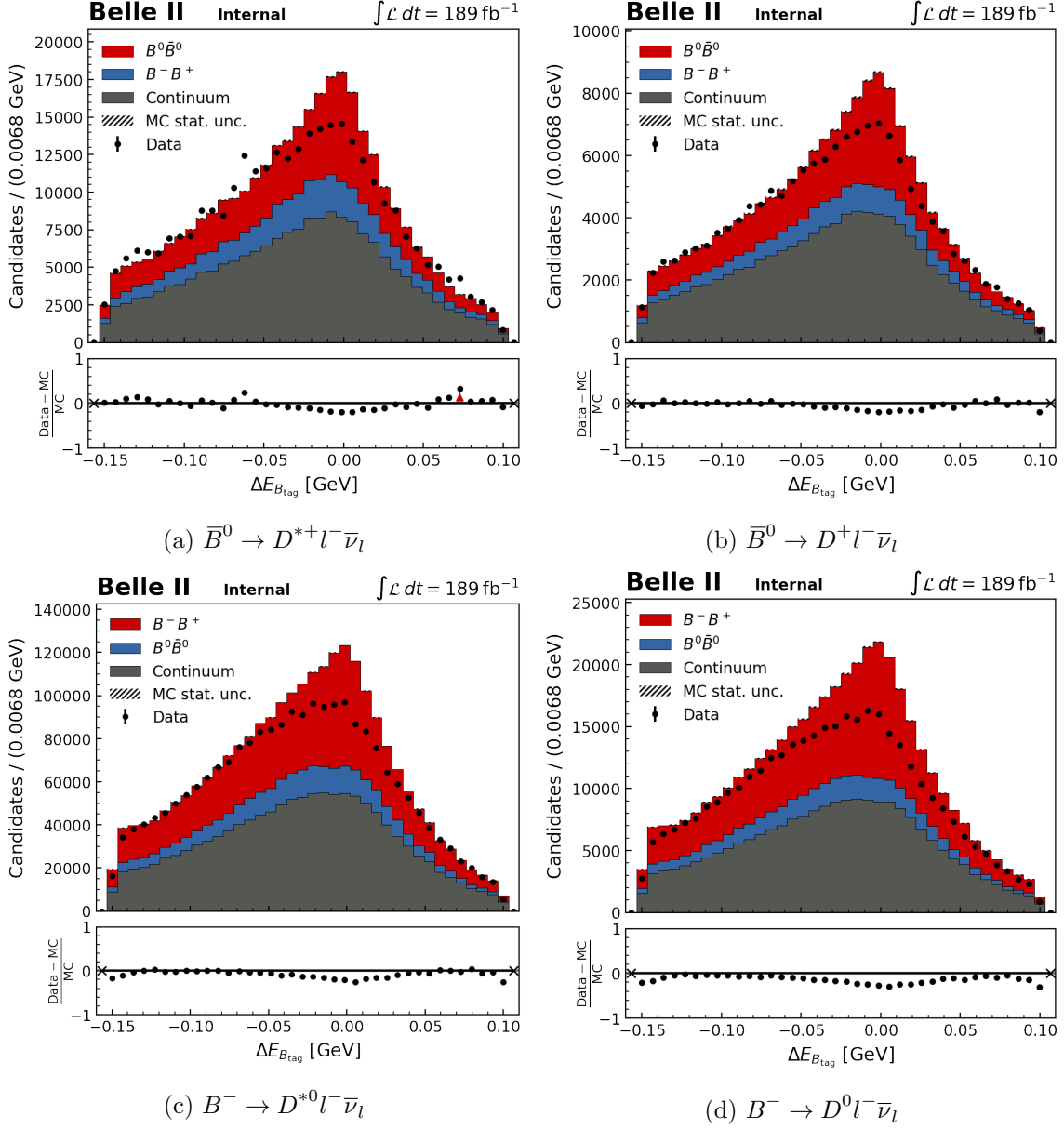


Figure 6.2.: Distributions of ΔE of the B_{tag} in the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The red arrows depict the bins where the residual pull is more than 1σ away from data-MC agreement.

- Clusters in the ECL barrel region (region 2) must have an energy greater than 30 MeV.
- Clusters in the ECL backward region (region 3) must have an energy greater than 60 MeV.

Finally, the π^0 candidate is restricted to the invariant mass of the two photons in the region

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

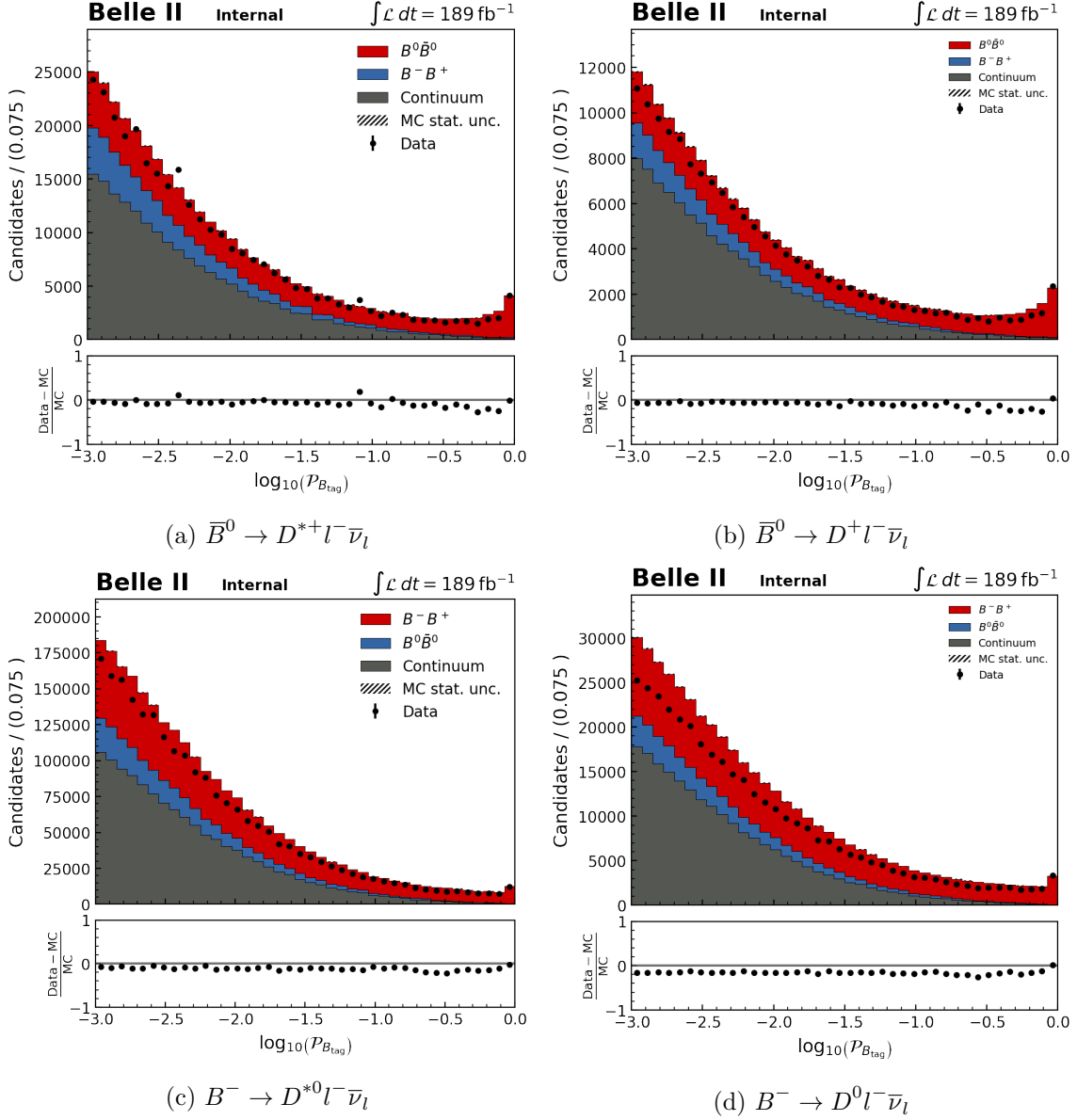


Figure 6.3.: Distributions of the base 10 logarithm of the $\mathcal{P}_{B_{\text{tag}}}$ for $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars.

$0.120 < M_{\gamma\gamma} < 0.145$. These requirements are summarized in Table 6.3. The π^0 candidate is then required to have an absolute cluster timing cut of less than 200 ns. Cluster timing is the difference between the t_0 of the event and the photon timing. This value is calibrated and corrected to include the time of flight of the photons such that photons from the interaction point (IP) correspond to the t_0 of the event. Thus, the cluster timing should

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

be consistent with 0 within errors, predominantly removing photons produced as beam background from candidacy. The π^0 data/MC correction factor for the 40% efficiency particle list is $\frac{\epsilon_{\text{Data}}(\pi^0)}{\epsilon_{\text{MC}}(\pi^0)} = 0.945 \pm 0.004 \text{ (stat.)} \pm 0.041 \text{ (syst.)}$.

In addition to the good track criteria required for all tracks (Table 4.4), a likelihood criteria called PID is required for all hadrons including charged pion ($\pi^\pm$) candidates. PID is the efficiency and fake rate of particle species determined through the use of the detector subsystems. The measure of effectiveness of the PID ranges from a particle candidate i having a PID probability $\mathcal{P}_i = 0$ being the least likely, to $\mathcal{P}_i = 1$ being the most likely. There are two types of PID, global and binary PID.

The global PID is a likelihood ratio

$$\mathcal{P}^{\text{global}} = \frac{\mathcal{L}_{\text{particle}}}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_d}, \quad (6.5)$$

where $\mathcal{L}_i$ is the likelihood of a particular particle candidate i hypothesis. Similarly, a binary PID can be defined when considering just two particles that are often mis-reconstructed as the other due to the similarity of their signatures in the Belle II sub-detectors. An example binary PID is a likelihood ratio such as

$$\mathcal{P}_\pi^{\text{binary}} = \frac{\mathcal{L}_\pi}{\mathcal{L}_\pi + \mathcal{L}_K}. \quad (6.6)$$

. in this analysis it is required that all $\pi^\pm$ candidates have a minimum of $\mathcal{P}_\pi^{\text{binary}} > 0.1$.

6.2.4. Slow π and γ selections

The slow pion π_{slow} is the neutral or charged π daughter produced in the decay of the D^* . The π_{slow}^0 produced in the decay of the D^{*+} and D^{*0} are selected using the π^0 50% efficiency list determined by the performance group at Belle II, similar to the above π^0 40% efficiency list. The selections for the slow π^0 list are shown in Table 6.4. This π_{slow}^0 is also required to have at least one hit in the CDC and an absolute cluster timing cut of $< 200 \text{ ns}$. The π^0 data/MC correction factor for the 50% efficiency list is $\frac{\epsilon_{\text{Data}}(\pi^0)}{\epsilon_{\text{MC}}(\pi^0)} = 0.945 \pm 0.004 \text{ (stat.)} \pm 0.071 \text{ (syst.)}$.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

The tracking efficiency difference between data and MC of the π_{slow}^0 are evaluated by the Belle II performance group. The π_{slow}^0 was the leading systematic in the work of Reference [1], thus this recently developed correction is crucial in the improvement of precision and this current work. Table 6.5 displays the yield correction factors applied in this analysis for π_{slow}^0 with momentum below 0.2 GeV, due to poorer reconstruction strength and resolution at these energies.

In comparison to the π_{slow}^0 , the charged π_{slow} have better agreement between data and MC. The π_{slow}^+ produced in the decay of the D^{*+} are selected to have a momentum higher than 0.05 GeV, have a vertex within a transverse distance of 2.0 cm from the IP ($|d_r| < 2.0$ cm) and have either a vertex or a point of closest approach in the z direction within 4.0 cm of the IP ($|d_z| < 4.0$ cm).

The requirements for the slow photon γ_{slow} produced in the decay of the D^{*0} are that candidates satisfy the γ 40% efficiency list selections and that they pass an absolute cluster timing cut of < 200 ns, shown in Table 6.3.

6.2.5. K_s^0 selection

The K_s^0 candidates are selected using the recommended `KS0:merged` list. A vertex fit is performed on the objects in the event and candidates with $M_{K_s^0} \in [0.450, 0.550]$ GeV that did not fail the vertex fit are kept as K_s^0 candidates. These candidates are then required to have a significance of distance (SOD) from the decay vertex to the interaction point

$$\text{SOD} = \frac{\text{distance}}{\sigma(\text{distance})} > 3. \quad (6.7)$$

This ensures that the flight distance of the K_s^0 is reasonable. Finally, all K are required to satisfy the kaon PID of $\mathcal{P} > 0.1$.

6.2.6. $D^{(*)}$ reconstruction

The four decay channels examined in this thesis contain a $D^{(*)}$ meson: $\bar{B}^0 \rightarrow D^+ l^- \bar{\nu}_l$, $B^- \rightarrow D^0 l^- \bar{\nu}_l$, $\bar{B}^0 \rightarrow D^{*+} l^- \bar{\nu}_l$ and $B^- \rightarrow D^{*0} l^- \bar{\nu}_l$. These D and D^* meson decays are exclusively selected such that both the signal and normalisation channels for the ℓ and

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

τ decay channels have the same reconstructed final state. As mentioned previously, the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ branching fraction measurement is an exclusive measurement, which means that certain decay channels of the D^* are selected, examined and used for the final measurement. The decay channels used are selected through consideration of the size of the branching fraction and through study of particle candidate resolutions and backgrounds. Extensive studies were performed by the author to ensure each channel contributes in a constructive manner to the full analysis. Initially, neutral particles caused resolution issues in the early stages of this analysis. Despite distinct improvements through studies, optimized selections and improved MC campaigns, the generic quality of the reconstruction and simulation at Belle II results in limited π^0 and K_S^0 resolutions and room for improvement in their distinction from backgrounds. This analysis concludes that the channels listed in Table 5.1 are statistically sufficient at the recorded integrated luminosity of $(189.8804 \pm 0.0131) \text{ fb}^{-1}$ and have an adequate signal contribution in comparison to background contributions to be included in the full analysis. Finally, this analysis consists of the largest collection of D^* and D decay channels used in Belle II exclusive hadronic FEI studies of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ branching fraction. In the previous work on this decay channel, only decays of $D^{*+} \rightarrow D^0 \pi^+$ with $D^0 \rightarrow K^- \pi^+$ were used to measure the $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ branching fraction [1].

The D candidates are required to have an invariant mass of $M_D \in [1.78, 1.92] \text{ GeV}$ initially. Figure 6.4 shows the invariant mass of the corresponding D mesons used either as candidates in the $B \rightarrow D\ell^- \bar{\nu}_\ell$ decay channels or as candidates for D^* daughters in the $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ decay channels, after reconstruction of the $\Upsilon(4S)$ candidate.

To construct a D^{*+} candidate, D^0 and D^+ candidates that are a daughter of the D^{*+} are combined with a corresponding π_{slow}^+ or π_{slow}^0 . Using the constraint of a small momentum window for the π daughter, the mass difference between these two candidates must satisfy $\Delta M_{D^{*+}-D} = M_{D^{*+}} - M_D \in [0.130, 0.170] \text{ GeV}$. The distribution for $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ is shown in Figure 6.5a.

Similarly, for a D^{*0} candidate, D^0 candidates that are daughters of the D^{*0} are combined with the corresponding π_{slow}^0 or γ_{slow} and must conform to $\Delta M_{D^{*0}-D} = M_{D^{*0}} - M_D \in [0.100, 0.190] \text{ GeV}$. This distribution is shown in Figure 6.5b.

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

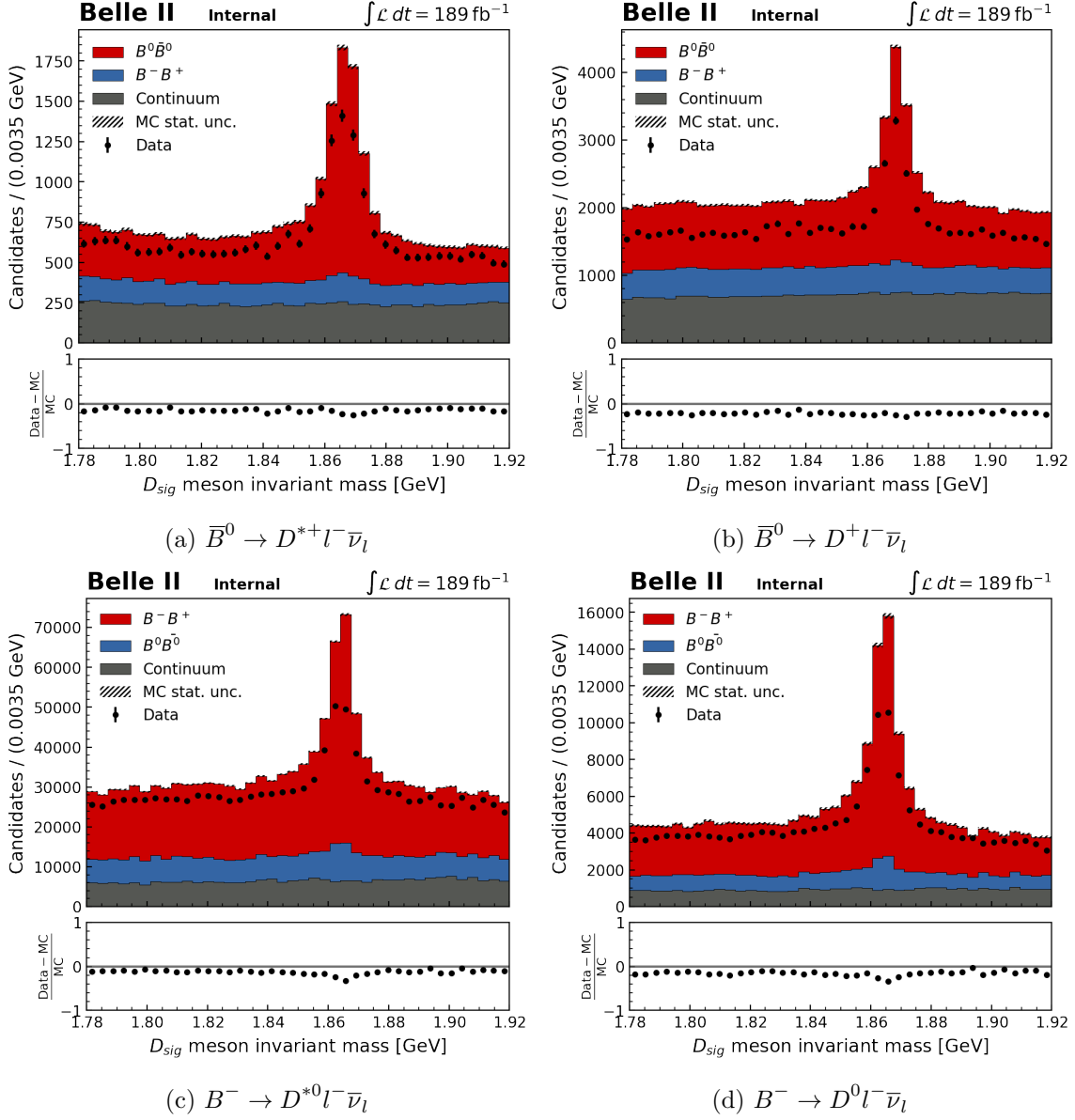


Figure 6.4.: The invariant mass of the B_{sig} daughter D mesons used either as candidates in the $B \rightarrow D l^- \bar{\nu}_l$ decay channels or as candidates for D^* daughters in the $B \rightarrow D^* l^- \bar{\nu}_l$ decay channels. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars.

6.2.7. Lepton selection

At the next stage of reconstruction, the signal-side leptons of the $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ decays are restricted. These charged leptons are required to pass a PID likelihood of more than 0.9 for e and μ candidates. This PID likelihood is determined using typical signatures from each

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

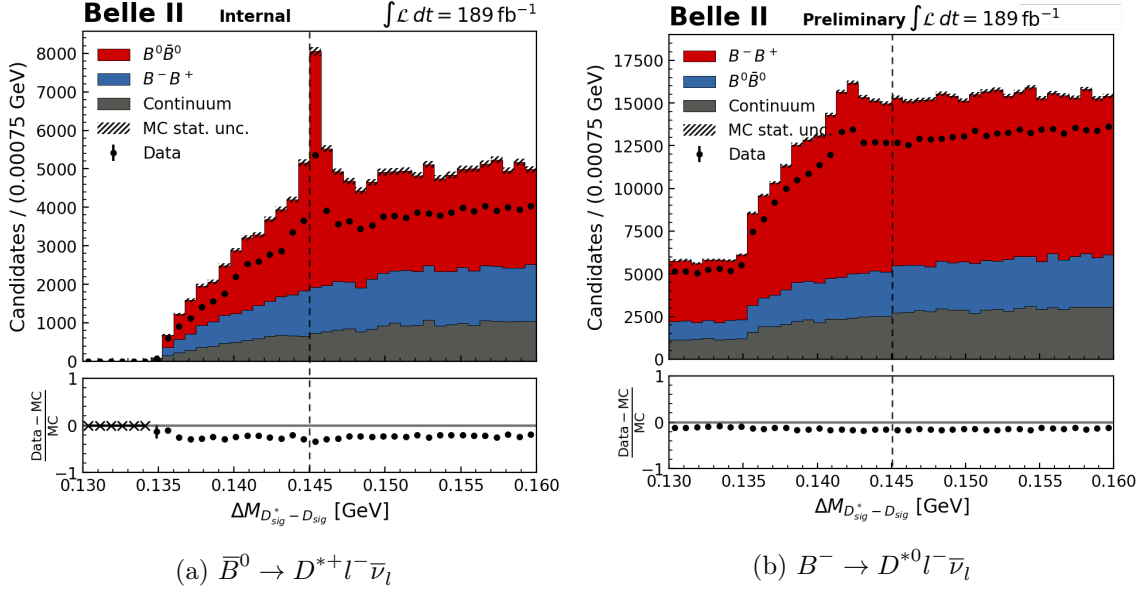


Figure 6.5.: The mass difference between the B meson daughter and granddaughter particles. With restrictions, this variable should represent the mass difference between D^* meson and its daughter D meson from the $B \rightarrow D^* l^- \bar{\nu}_l$ decay, and would peak around 0.145 GeV for the D^{*+} decay.

of the sub-detectors of the Belle II detector. The τ candidate does not have PID applied. All lepton candidates must satisfy the good track criteria. Electrons must be corrected for Bremsstrahlung. This is performed by requiring them to be located within the CDC acceptance region $-0.8660 < \cos(\theta) < 0.9563$, have an energy of more than 0.02 GeV and to have an average of more than 1.5 GeV for the weighted sum of all crystals in an isolated ECL cluster.

The lepton is selected and the momentum spectrum at this stage is shown in Figure 6.6. In the lower lepton momentum region, the data-MC discrepancy is high, leading to the necessity of further selections in continuum suppression and in the lepton momentum itself. These will be addressed in Section 6.3.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

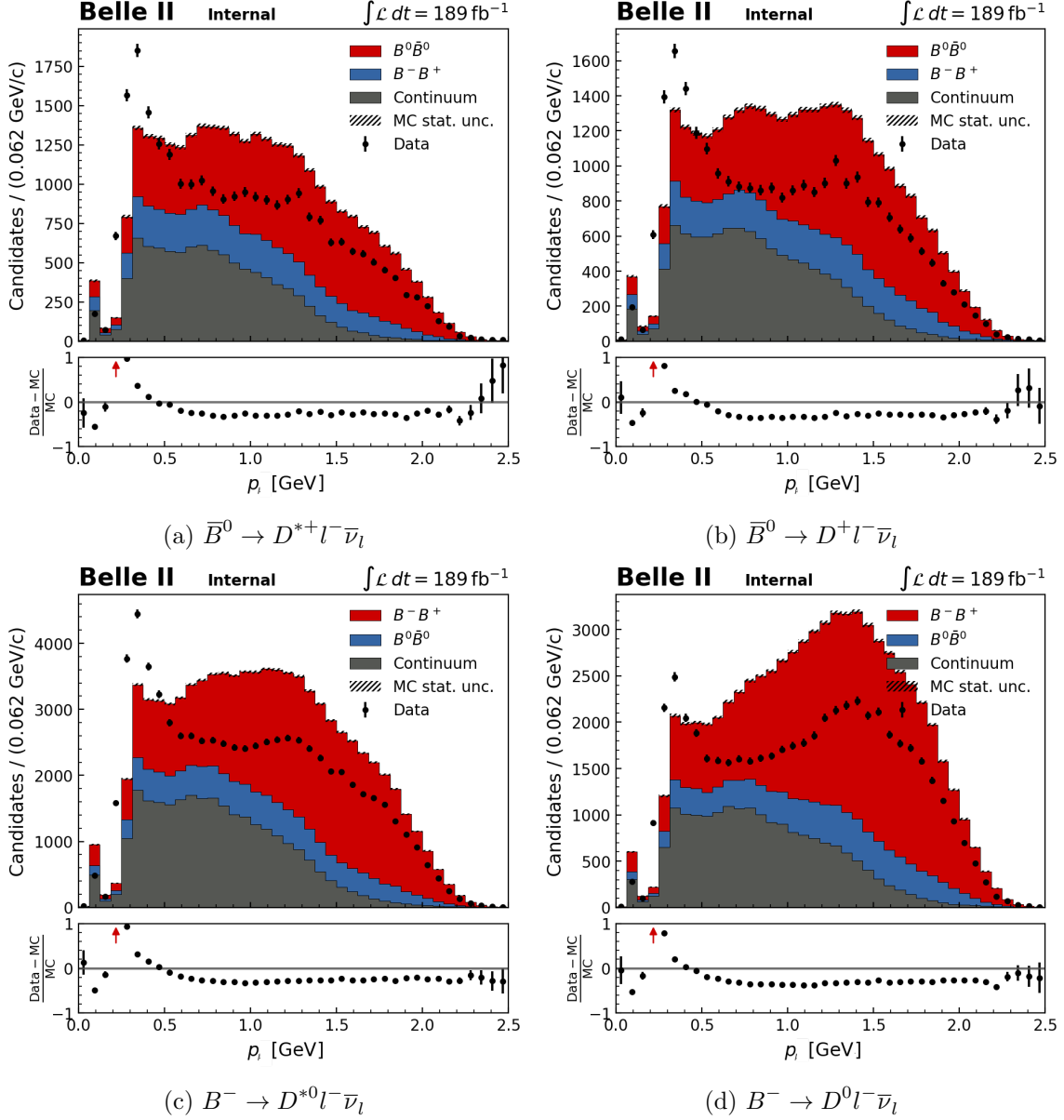


Figure 6.6.: The lepton momentum spectrum for the $B \rightarrow D^{*}\ell^- \bar{\nu}_\ell$ decay channels. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars.

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

Category	Cuts
FEI Probability	$\mathcal{P}_{B_{\text{tag}}} > 0.001$
$M_{bc, B_{\text{tag}}}$	$M_{bc, B_{\text{tag}}} > 5.27 \text{ GeV}$
$\Delta E_{B_{\text{tag}}}$	$-0.15 < \Delta E_{B_{\text{tag}}} < 0.1 \text{ GeV}$
γ_{slow}	gamma:eff40_May2020 list and $\ \text{ECL cluster timing}\ < 200 \text{ ns}$
π_{slow}^0	pi0:eff50_May2020 list and $\gamma : \ \text{ECL cluster timing}\ < 200 \text{ ns}$ $\text{nCDCHits} > 1$
π^0	pi0:eff40_May2020 list and $\gamma : \ \text{ECL cluster timing}\ < 200 \text{ ns}$
K_S^0	KS0:merged list significance of distance > 3 (before the B_{sig} vertex fit.)
D	$1.78 < M_D < 1.92 \text{ GeV}$
D^*	$0.130 < \Delta M_{D^*-D} < 0.170 \text{ GeV}$ for D^{*+} $0.100 < \Delta M_{D^*-D} < 0.190 \text{ GeV}$ for D^{*0}
e	Bremsstrahlung correction: $E > 0.02 \text{ GeV}$ and number of ECL cluster hits > 1.5 and $-0.8660 < \cos(\theta) < 0.9563$
Vertex Fit	treeFit(list_name='B0:sig', conf_level=-1, ipConstraint=False, updateAllDaughters=True, massConstraint=[K_S^0, π^0])
Lepton ID	$\mathcal{P}_\mu^{\text{global}} > 0.9, \mathcal{P}_e^{\text{global}} > 0.9$ (no TOP)
Hadron ID	$\mathcal{P}_K^{\text{binary}} > 0.1, \mathcal{P}_\pi^{\text{binary}} > 0.1$
ROE Charged	ROE charge == 0 and ROE number of tracks == 0 $ d_r < 5.0$ and $ d_z < 20.0$
Neutral	number of CDC hits > 20 (0) for hadrons (leptons) gamma:eff40_May2020 list
$\Upsilon(4S)$	Best candidate selection Ranked by $\mathcal{P}_{B_{\text{tag}}}$ and dM for D or $\Delta(\Delta M_{D^*-D})$ for $D^* \rightarrow D$

Table 6.2.: A summary of the online signal selections applied in this analysis. An issue was discovered with the probability density function for the electron ID of the TOP, which is present in **release-05** of **basf2**; the TOP electron ID was removed for this analysis.

Candidates	Cuts
γ (eff40_May2020 list)	[[number of ECL cluster hits > 1.5] and $[0.2967 < \text{ECL cluster angle } \theta < 2.6180]]$, and [[ECL cluster region == 1 and $E > 0.080$] or [ECL cluster region == 2 and $E > 0.030$] or [ECL cluster region == 3 and $E > 0.060$]]
π^0 (eff40_May2020 list)	γ (eff40_May2020 list) γ (eff40_May2020 list) and $0.120 < \text{InvM}_{\gamma\gamma} < 0.145$

Table 6.3.: The π^0 standard particle list with 40% data-MC efficiency defined in May 2020.

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

Candidates	Cuts
γ (eff50_May2020 list)	[[number of cluster hits > 1.5] and [0.2967 < cluster angle θ < 2.6180]] and [[cluster region == 1 and $E > 0.025$] or [cluster region == 2 and $E > 0.025$] or [cluster region == 3 and $E > 0.040$]]
π^0 (eff50_May2020 list)	γ (eff50_May2020 list) γ (eff50_May2020 list) and $0.105 < M_{\gamma\gamma} < 0.150$

Table 6.4.: The π^0 standard particle list with 50% data-MC efficiency defined in May 2020.

Momentum region [GeV]	Correction factor
$0.05 < p < 0.12$	$0.909 \pm 0.032 \pm 0.020$
$0.12 < p < 0.16$	$1.033 \pm 0.026 \pm 0.022$
$0.16 < p < 0.20$	$0.971 \pm 0.027 \pm 0.021$

Table 6.5.: A summary of the π^0 correction factors and corresponding systematic uncertainties as a function of momentum.

6.2.8. B_{sig} reconstruction

To reconstruct the B_{sig} candidates, a lepton candidate is combined with a D^* candidate, matching the charge required for an allowed signal decay. These are constrained with a vertex fit that requires both candidates to originate from a common position in the detector.

The B_{sig} candidates are then fitted using the **basf2** adapted TreeFitter module [61]. This simultaneously fits all of the vertices in the B_{sig} decay chain. It is possible to select the best particle candidates by removing those that fail the fit or have a low confidence level. This was tested using a standard figure of merit (FOM) used in particle physics, defined as:

$$\text{FOM} = \frac{N_{\text{sig}}}{\sqrt{N_{\text{sig}} + N_{\text{bkg}}}}, \quad (6.8)$$

where N_{sig} is the number of signal (and/or normalisation) events, and N_{bkg} is the number of background events. It was ultimately found to be most efficient to keep all particle candidates at this stage of the selection.

The masses of the K_S^0 and the π^0 descendants of each B_{sig} are constrained such that any instances of these particles in the B_{sig} decay chain are assigned the corresponding nominal mass from the PDG [5]. The momenta of the daughter particles are updated with the TreeFitter. Finally the B_{sig} and B_{tag} candidates are ready to be reconstructed into the $\Upsilon(4S)$.

6.2.9. Best Candidate Selection

At this stage in the analysis, a $\Upsilon(4S)$ candidate is reconstructed using the B_{sig} and B_{tag} candidates. The $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ events are selected from generic MC using **TopoAna**. Correctly reconstructed events are defined as those with a truth-matched signal-side lepton. In comparison to Reference [60], the D reconstruction at this stage in Belle II underperforms due multiple reasons. These are:

- Mis-reconstruction of D^* (D) candidates as D (D^*) candidates for the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$.

This is due to mis-reconstruction of daughter candidates or due to extra (missing)

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

daughter candidates being considered for candidacy. There are high D^* fake rates present in the current $B \rightarrow D\ell^- \bar{\nu}_\ell$ selections, however the D fake rates in the $B \rightarrow D^*\ell^- \bar{\nu}_\ell$ selections are negligible.

- Resolution issues perpetrated by constituent daughters such as π^0 candidates.
- Data-MC discrepancy due to modelling of daughter candidates in the Belle II experiment.

Thus, in this study we chose to identify signal and normalisation decays as those in which only the lepton is correctly reconstructed and truth-matched. The use of correctly reconstructed, truth-matched $D^{(*)}$ candidate or B_{sig} candidates to identify signal were extensively studied and had a worse data-MC agreement than the lepton selection.

The presence of multiple candidates per event could give rise to raw biases that are large compared to the usual statistical uncertainties present in a measurement [62]. From a study with signal MC, the average number of $\Upsilon(4S)$ candidate yields at this stage in the analysis is 5.74 (13.84) for the B^0 (B^+) decay channel. Per $D^{(*)}$ decay channel, the average number of candidates per event was found to distinctly vary. It is found that decays with higher multiplicity have more candidates per event, and those with neutral particles have a higher likelihood of more candidates per event. One notable observation is that there is a higher candidate multiplicity for the $B \rightarrow D^*\ell^- \bar{\nu}_\ell$ decay channels, which indicates that the slow pion criteria are a source of increased candidature.

To reduce candidate yields to just one $\Upsilon(4S)$ per event, the arbitration approach is used and the B_{tag} with the highest FEI $\mathcal{P}_{B_{\text{tag}}}$ is selected as the best B_{tag} candidate. At this stage in the analysis it is still possible for an event to have more than one $\Upsilon(4S)$ candidate per event, thus the candidates are ranked again with variables dependent on the type of B_{sig} decay channel. Optimal ranking methods were studied and selected for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decays. For the $B \rightarrow D^*\ell^- \bar{\nu}_\ell$ decays, the event D^* candidate that has the lowest difference between the reconstructed mass difference ΔM_{D^*-D} of the D^* candidate and it's child D candidate to the nominal ΔM_{D^*-D} , $\Delta(\Delta M_{D^*-D})$ is ranked highest. For the $B \rightarrow D\ell^- \bar{\nu}_\ell$ decays, the D candidate that has the closest invariant mass to the nominal D meson mass $\Delta M_{D^{*+}}$ is ranked highest.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

The method of arbitration is only effective when distributions of the correctly reconstructed signal and the other candidates differ in the quantity used for best candidate selection. The distributions for $\mathcal{P}_{\mathcal{B}_{\text{tag}}}$, $\Delta(\Delta M_{D^*-D})$ and ΔM_D are shown to be adequate for the arbitration approach described in Appendix B.

For events with multiple candidates that vary in decay channel, it is possible to obtain more than one best candidate per event. In the MC samples of this analysis, 0.01 – 0.1% of events exhibit this. In these cases, to make a selection between the possible candidates, the candidate with the lowest sum of all of the ranking variables is chosen for the D^* channels, and the candidate with the best ranking is chosen for the D channels. These methods were chosen after studies were made with a focus on optimising signal efficiency and purity.

Finally, each highest ranked B_{tag} coupled with the highest ranked D^* candidate are used to select the best $\Upsilon(4S)$ candidate per event. The results of the best candidate selection are represented in Figures 6.7 and 6.8. Continuum backgrounds and other backgrounds are largely removed from candidacy and the peaks of the D meson masses and the ΔM_{D^*-D} improve in resolution.

The ranking was found to have no adverse effect on the other $B^0 \rightarrow X_C \ell^- \bar{\nu}_\ell$ candidates.

With the best $\Upsilon(4S)$ candidate reconstructed from the optimised B_{sig} and B_{tag} candidate pairs per event, the yield extraction variable - the m_{miss}^2 distribution - is examined. Considering the normalisation channel $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ as the signal decay that is to be measured in this analysis, and other decays (that are not continuum events) as background (including $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$), Figure 6.9 shows the m_{miss}^2 distributions after best candidate selection^[iii].

To reduce the large backgrounds in Figure 6.9, the ROE is considered; if an event has any additional tracks in the ROE after this reconstruction, these events are excluded. The particles in the ROE are required to satisfy the following requirements. Only clusters that satisfy the γ efficiency of 40% criteria listed in Table 6.3 are considered in the ROE. For charged particles in the ROE, the number of hits in drift chambers are required to be more than 20 (0) for hadrons (leptons) and $|d_r| < 5.0$ and $|d_z| < 20.0$ must be satisfied.

^[iii]Note the distinction here between l and ℓ .

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

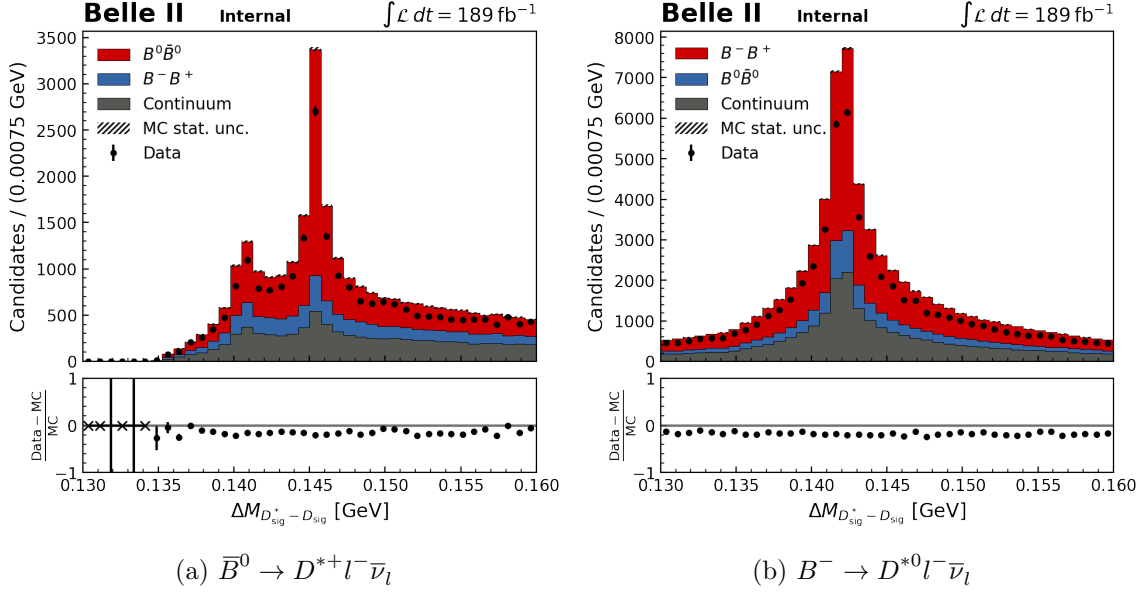


Figure 6.7.: The mass difference between the B meson daughter and granddaughter particles. Generic MC is classed as $B^0 \bar{B}^0$, $B^+ B^-$ and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection.

For neutral particles in the ROE, all events with a non-zero charge and number of tracks are excluded, the distributions of which are shown in Figure 6.10. Finally, the m_{miss}^2 distributions after the ROE exclusions are applied are shown in Figure 6.11.

6.2.10. E_{ECL}

E_{ECL} is the linear sum of the energy of the neutral clusters in the ECL that do not correspond with the reconstructed daughters of either the B_{tag} or the B_{sig} , i.e. they are part of the ROE. For signal and normalisation events, E_{ECL} would be zero if the detector were perfect. However, allowances are made for beam background and detector effects. Background events are distinguishable from signal and normalisation events as their E_{ECL} tends towards the positive due to additional clusters in an event produced by unassigned tracks or from neutral particles present in a mis-reconstructed B_{tag} or B_{sig} .

The distribution of correctly reconstructed events is expected to peak at zero, however due to high levels of beam backgrounds at Belle II, in reality the E_{ECL} peak is significantly shifted towards larger values of E_{ECL} (Figure C.1 in the appendices). This has a strong

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

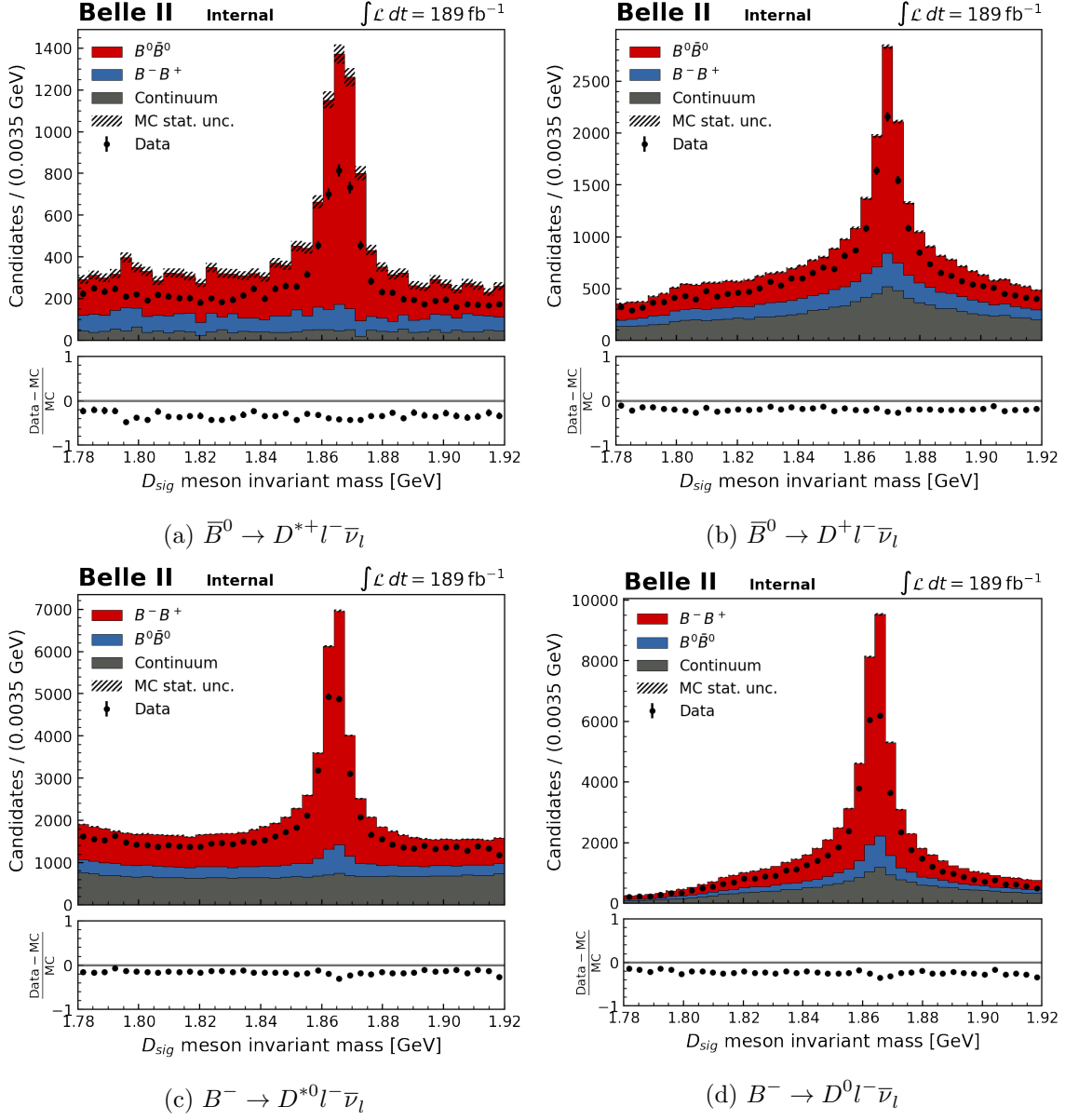


Figure 6.8.: The invariant mass of the B_{sig} daughter D mesons used either as candidates in the $B \rightarrow D \ell^- \bar{\nu}_\ell$ decay channels or as candidates for D^* daughters in the $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ decay channels. Generic MC is classed as $B^0\bar{B}^0$, B^+B^- and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

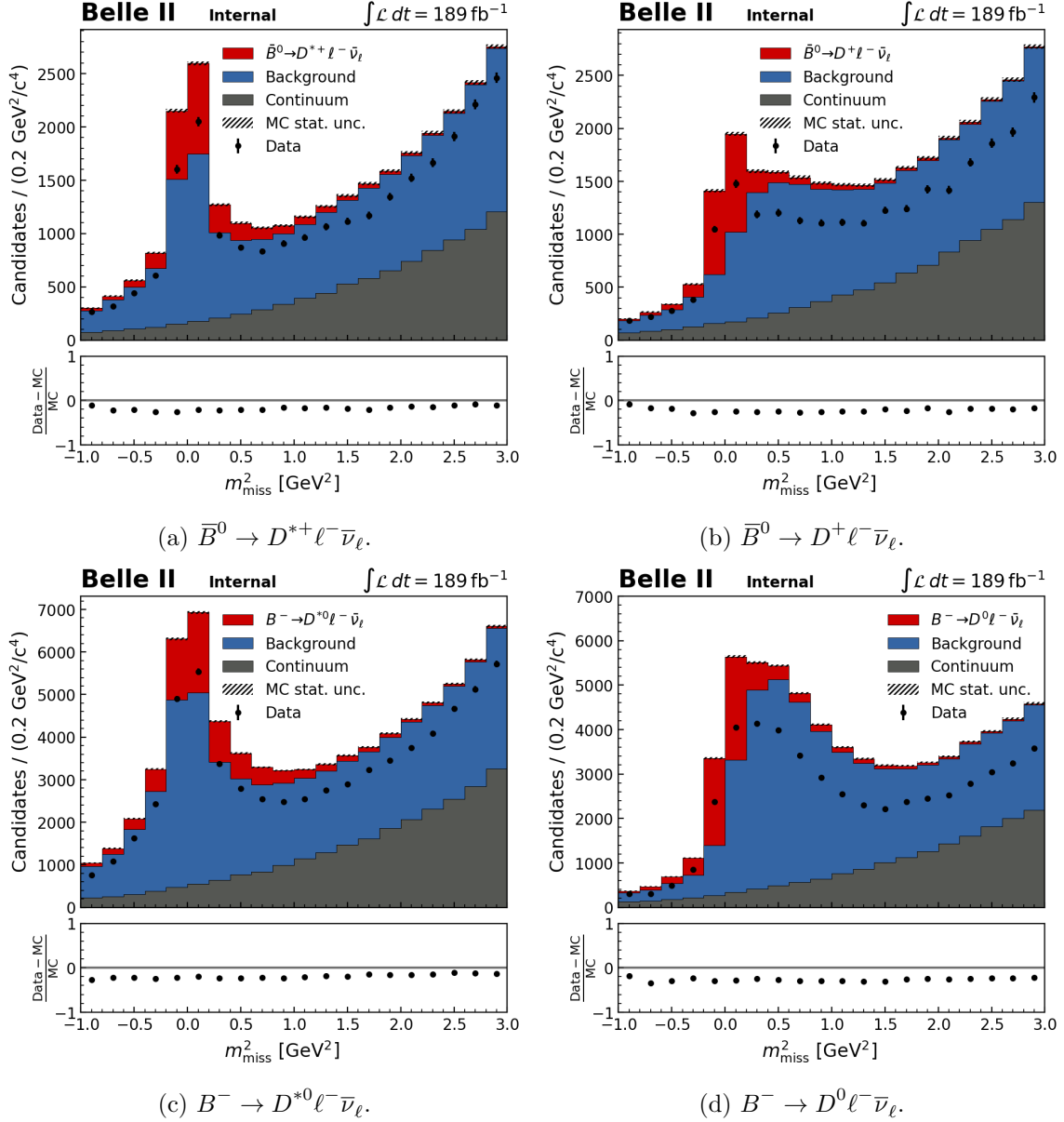


Figure 6.9.: The m_{miss}^2 distributions for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

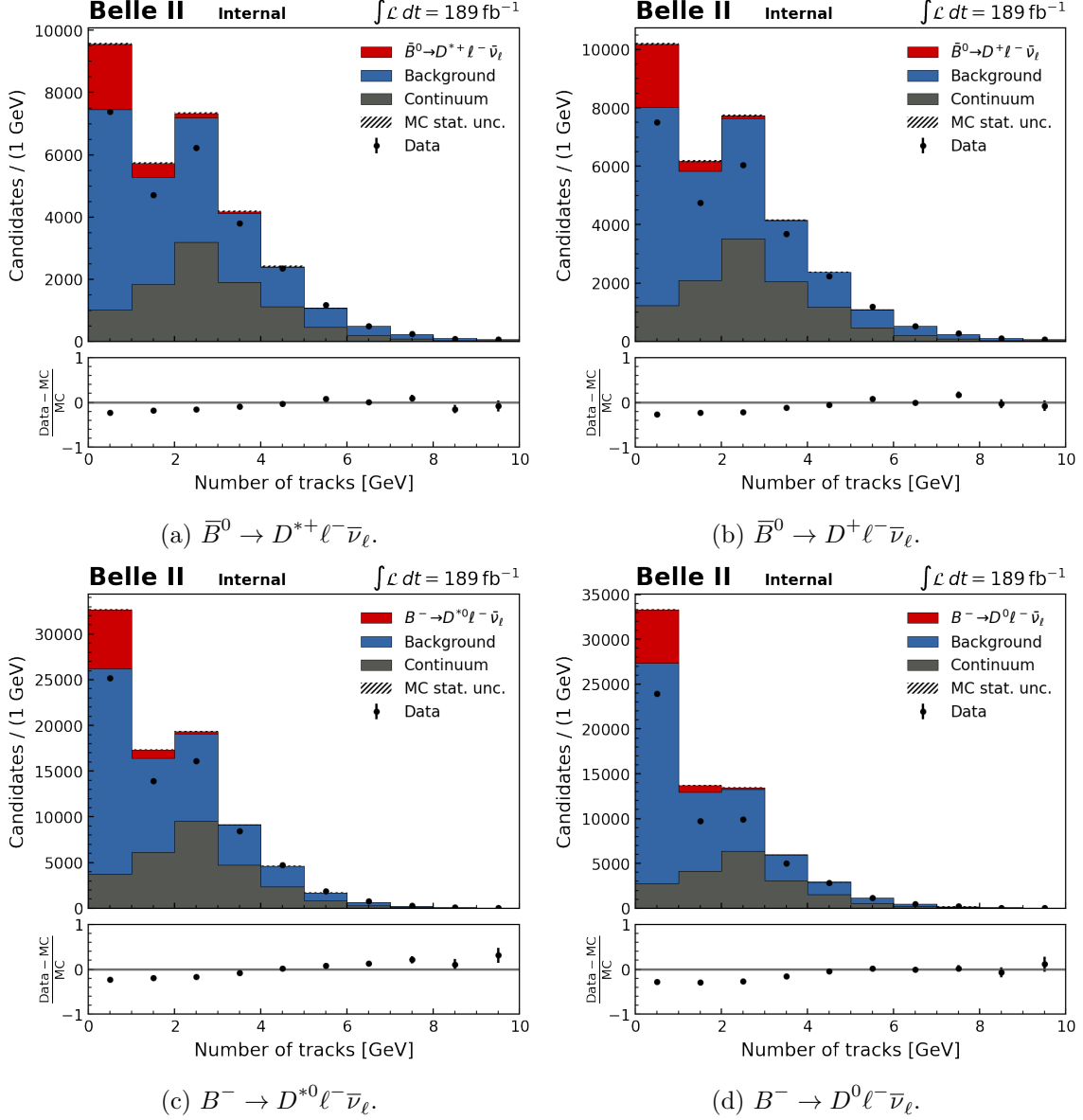


Figure 6.10.: The distribution of the number of tracks in the ROE for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

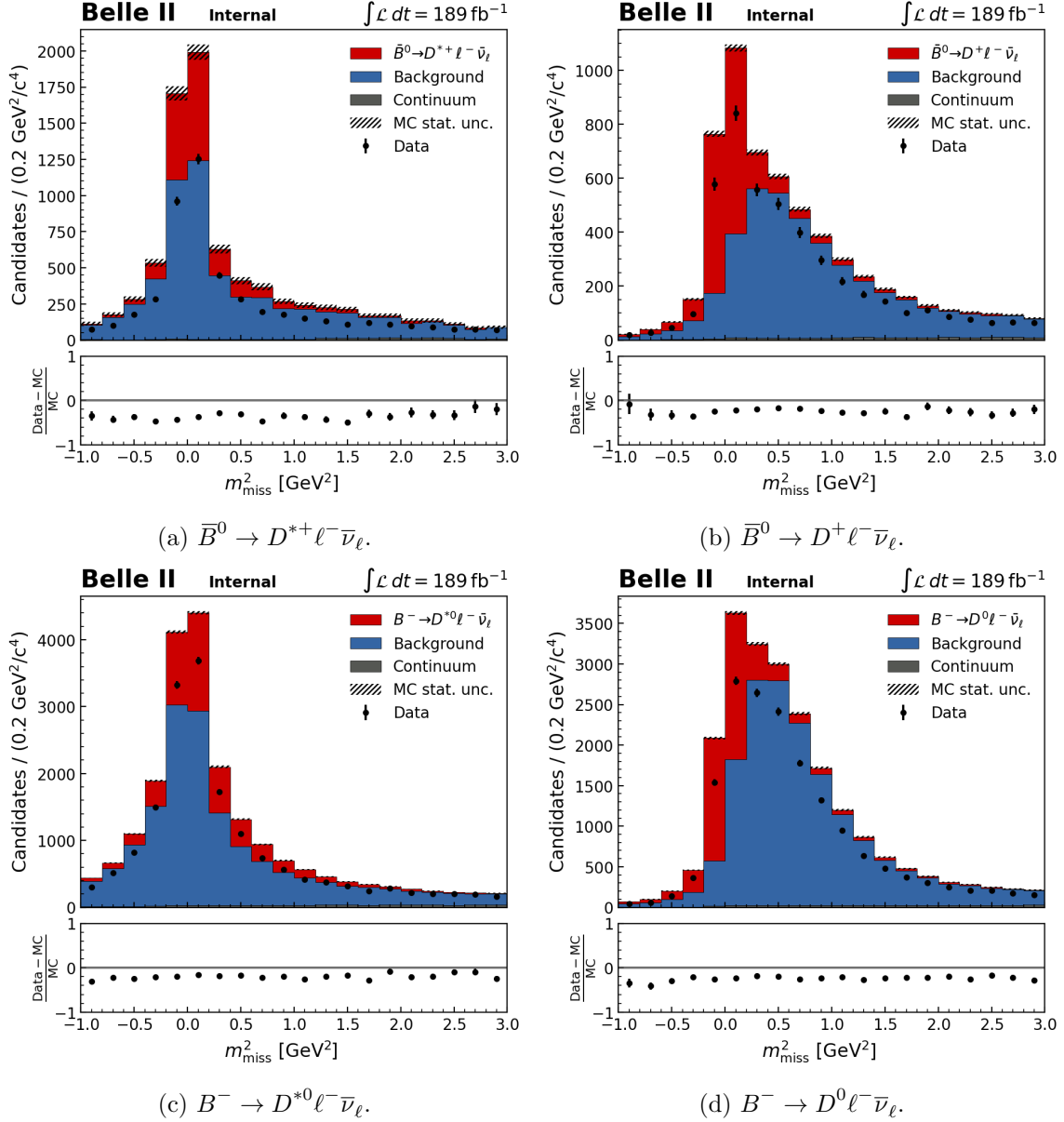


Figure 6.11.: The m_{miss}^2 distributions for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. The distributions shown are after reconstruction of the $\Upsilon(4S)$ candidate and after application of best candidate selection.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

impact on $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ measurements and therefore is an important consideration in this analysis. Due to its branching fraction, $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ has low statistics, and peaking backgrounds are dominant in the m_{miss}^2 signal region. Compared to $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$, the $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ decay channels have poorer background separation.

A shower-shape based MVA was applied to rectify this issue. Shower shape variables were used to train the MVA to distinguish between real photons and the beam backgrounds in the ROE. This MVA was applied to all ROE clusters after the online $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ selections. It results in the reduction of the beam background contribution to E_{ECL} and increases its discriminating power. This MVA plays a crucial role in future $R(D)$ and $R(D^*)$ analyses and has a significant impact on the results of this thesis. This will be discussed in Section 6.4.

6.3. Offline selections and corrections

The following section details the offline selections including further particle candidate selection, data-MC corrections, and development and application of a D^{**} background suppression MVA. At this stage in the analysis, the figures contain the information from the fully reconstructed $\Upsilon(4S)$. The MC is split into the normalisation MC, the D^{**} background MC and, where applicable, any cross-feed background contributions. For a summary of the offline cuts, see Table 6.6.

6.3.1. Continuum Suppression

In an e^+e^- process, event shape analysis is essential in suppressing continuum events. As shown in Table 4.1, the desired $\Upsilon(4S) \rightarrow B\bar{B}$ processes do not have the highest cross-sections of the hadronic processes; there are other non-hadronic e^+e^- processes that have higher cross-sections. A rapid distinction between continuum e^+e^- processes and $B\bar{B}$ processes is through their event shapes. The continuum processes have a jet-like nature, while the $B\bar{B}$ processes appear more spherical.

Tracks and clusters that satisfy the pre-defined ‘good’ Belle II criteria are used to determine an event shape variable: the ratio of the second Fox-Wolfram moment to the zeroth

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

Candidates	Cuts
Continuum Suppression	B_{sig} Fox Wolfram $R_2 < 0.3$
ROE	E_{ECL} in neutral clusters in $\text{ECL} < 1.0$ GeV
$D^{(*)}$	$\bar{B}^0 \rightarrow D^{*+}[D^0\pi^+]\ell\nu: 0.142 < \Delta M_{D^{*-}D} < 0.150$ GeV $\bar{B}^0 \rightarrow D^{*+}[D^0[K\pi^+\pi^0]\pi^+]\ell\nu: 1.842 < M_{D^0} < 1.886$ GeV $\bar{B}^0 \rightarrow D^{*+}[D^0[\text{else}]\pi^+]\ell\nu: 1.853 < M_{D^0} < 1.875$ GeV $\bar{B}^0 \rightarrow D^{*+}[D^+\pi^0]\ell\nu: 0.134 < \Delta M_{D^{*-}D} < 0.148$ GeV $\bar{B}^0 \rightarrow D^{*+}[D^0[K\pi^+\pi^0]\pi^0]\ell\nu: 1.851 < M_{D^0} < 1.887$ GeV $\bar{B}^0 \rightarrow D^{*+}[D^0[\text{else}]\pi^0]\ell\nu: 1.860 < M_{D^0} < 1.878$ GeV $\bar{B}^0 \rightarrow D^+[K^-\pi^+\pi^+]\ell\nu: 1.850 < M_{D^+} < 1.878$ GeV $\bar{B}^0 \rightarrow D^+[\text{else}]\ell\nu: 1.840 < M_{D^+} < 1.888$ GeV $B^- \rightarrow D^{*0}[D^0\pi^0]\ell\nu: 0.141 < \Delta M_{D^{*-}D} < 0.145$ GeV $B^- \rightarrow D^{*0}[D^0\gamma]\ell\nu: 0.138 < \Delta M_{D^{*-}D} < 0.145$ GeV $B^- \rightarrow D^{*0}[D^0 \rightarrow K^-\pi^+\pi^0, K^-\pi^+\pi^-\pi^+, K_s^0\pi^+\pi^-, K_s^0\pi^+\pi^-\pi^0]\ell\nu:$ $1.850 < M_{D^0} < 1.870$ GeV $B^- \rightarrow D^{*0}[D^0 \rightarrow \text{else}]\ell\nu: 1.810 < M_{D^0} < 1.890$ GeV $B^- \rightarrow D^0[K^-\pi^+, K^-\pi^+\pi^-\pi^+, K_s^0\pi^+\pi^-]\ell\nu:$ $1.852 < M_{D^+} < 1.880$ GeV $B^- \rightarrow D^0[\text{else}]\ell\nu: 1.848 < M_{D^+} < 1.90$ GeV
$\Upsilon(4S)$	TopoAna truth matched signal and normalisation modes with truth matched reconstructed lepton

Table 6.6.: A summary of the offline signal selections applied in this analysis. Here ‘else’ refers to the remaining $D^{(*)}$ decays in the mode selections.

moment R_2 . This value is calculated by using all tracks and photons in the event. A cut of $R_2 < 0.3$ is applied to suppress continuum events where $e^+e^- \rightarrow q\bar{q}$ and to reduce the overall discrepancy between data and MC, shown in Figure 6.12. This selection used to be included as a pre-cut in the hadronic FEI, but was recently removed due to impacts on other analyses.

6.3.2. ROE selection

An optimized selection of E_{ECL} in the neutral clusters of an event is restricted to be less than 1.0 GeV to reduce backgrounds and continuum events. The selection is shown in Figure 6.13. Figure 6.13 also represents the outcome of the E_{ECL} MVA discussed in Section

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

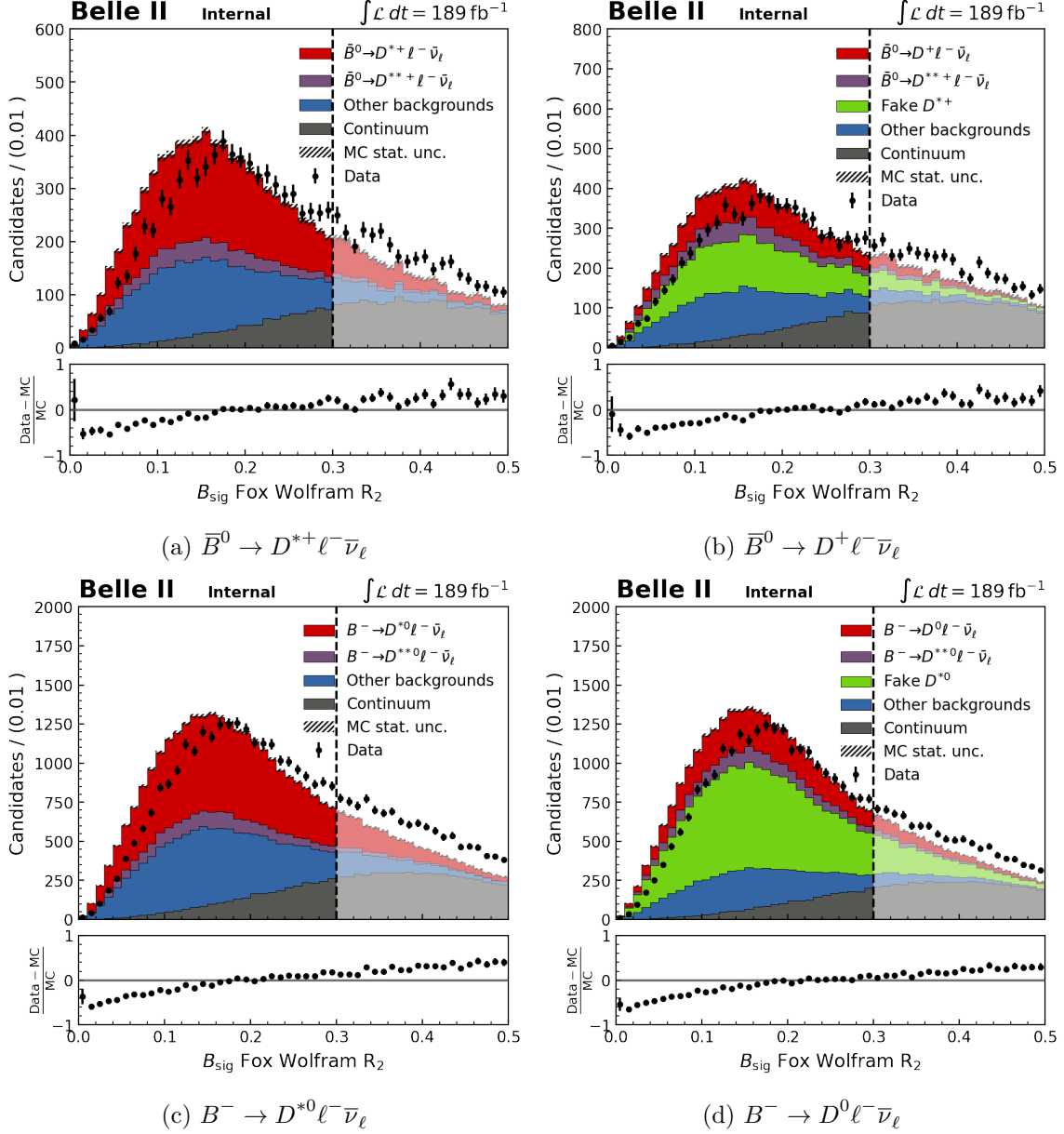


Figure 6.12.: Distributions of R_2 for $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. An offline cut of $R_2 < 0.3$ is projected on the figures.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

6.2.10, and exhibits the reduction of beam backgrounds and correction of the E_{ECL} peak.

The effect on the normalisation signal MC can be seen in more detail in Appendix C.

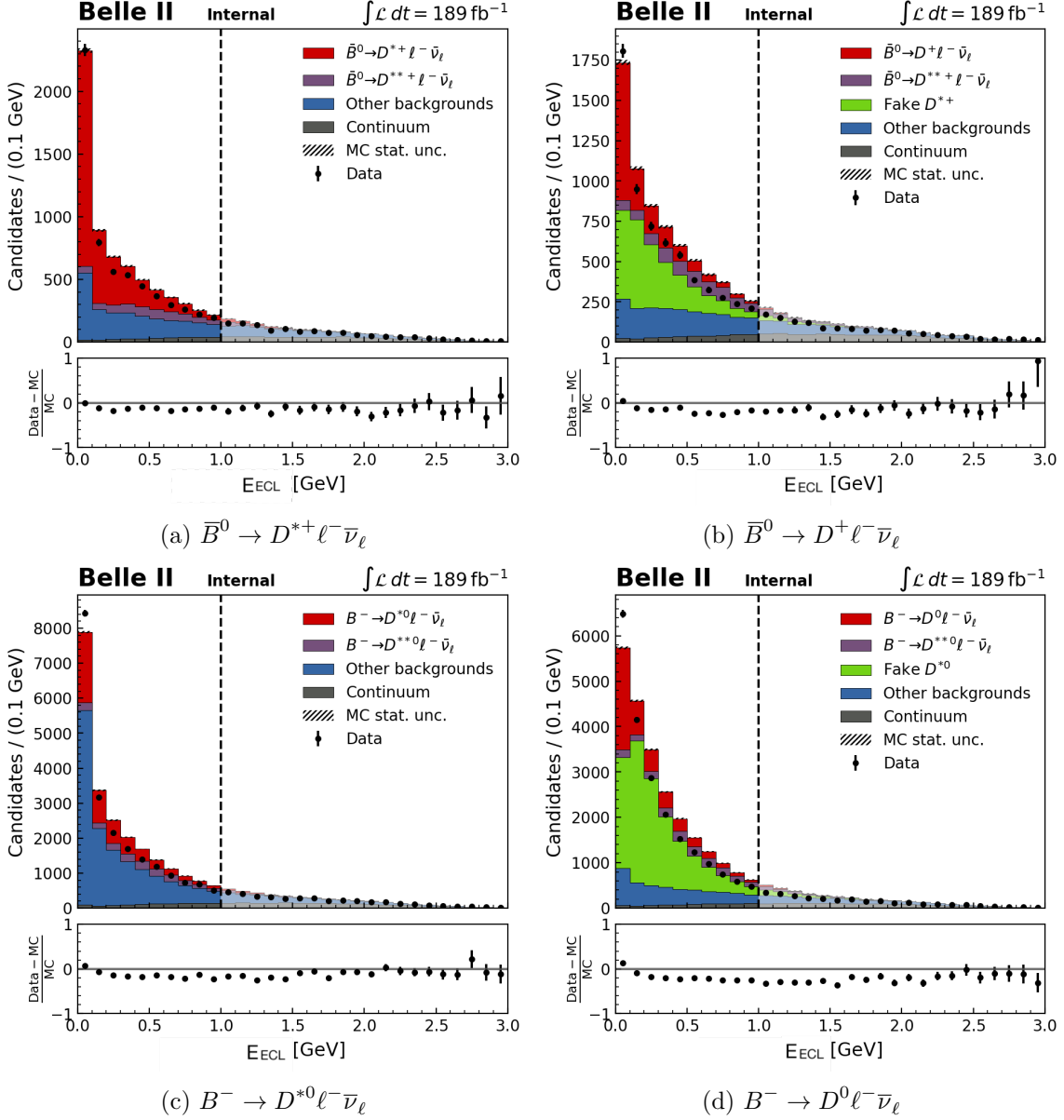


Figure 6.13.: The E_{ECL} distributions after the E_{ECL} beam background suppression is applied. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars. An offline cut of $E_{\text{ECL}} < 1.0$ GeV is projected onto the figures.

6.3.3. $D^{(*)}$

To reduce the number of fake D^* candidates, M_D and ΔM_{D^*-D} cuts are tightened after extensive studies performed on a decay mode by decay mode basis. To select the correctly reconstructed D^* candidates, MC truth matched D^* candidates were studied in the corresponding $B^0\bar{B}^0$ and B^+B^- generic MC samples. The optimal selections for the D^* candidates were chosen using the FOM. Ultimately, the mass of the D candidate is required to satisfy the channel dependent invariant masses specified in Table 6.6. For example, the mass of the D candidates for the $B \rightarrow D\ell^- \bar{\nu}_\ell$ $M_D \in [1.840, 1.888]$ GeV. The $B^0 \rightarrow D^+[K_S^0\pi^+]\ell^- \bar{\nu}$ candidate decay channels are required to have a D invariant mass consistent with the masses of the daughters of the K_S^0 and π : $M_{K\pi} \in [1.858, 1.878]$ GeV. Figure 6.14 displays the D invariant mass for the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ channels after these optimizations.

6.3.4. PID corrections

PID at Belle II is one of the key features in development that will provide opportunity for higher precision measurements. The PID is developed based on the inputs from the majority of the sub-detectors at Belle II.

Using standard decay processes such as $J/\Psi \rightarrow e^+e^-$ and $K_S^0 \rightarrow \pi^+\pi^-$, the performance of the PID is evaluated.

For example, semi-tauonic decays usually dominate the low momentum spectrum of $p < 1$ GeV, however in this range the mis-identification rate (also known as ‘fake rate’) of leptons is larger. To mitigate this the PID algorithm has been developed by the Belle II PID group and updated to differentiate between low momentum leptons and hadrons using the different characteristics of their energy deposits. This update reduced the mis-identification rate of low momentum e by a factor of 10.

Even with this improvement, corrections must be implemented for the PID for lepton and hadron identification and mis-identification. Within the `basf2` package, corrections must be implemented for the various PIDs available. A package (`PIDvar`) [63] was developed by members of the Belle II collaboration to weight true leptons and pions faking leptons

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

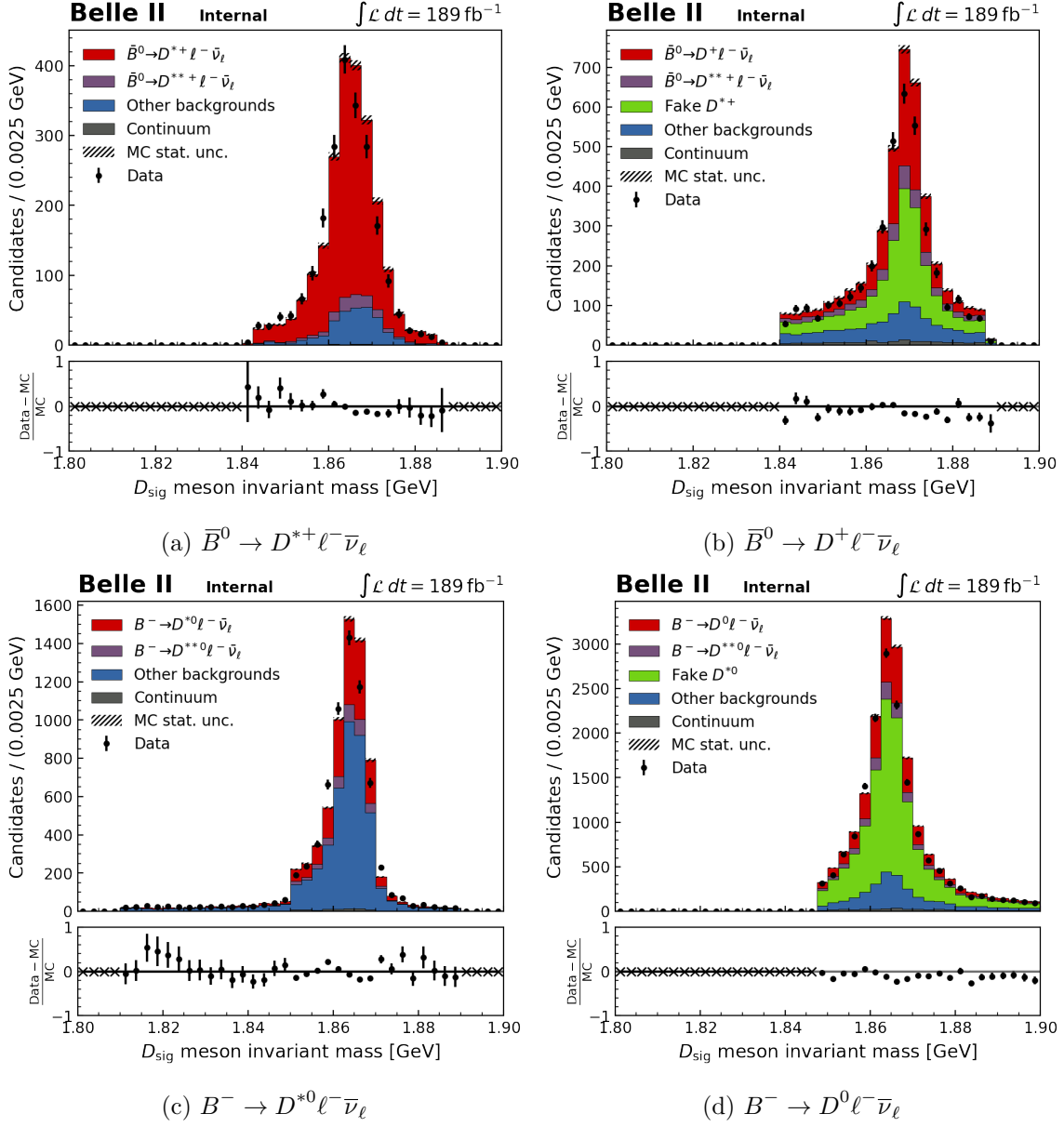


Figure 6.14.: The D meson invariant mass for the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay channels. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$), D^{**} background, other backgrounds and continuum. The D^* optimisation selections have been applied.

using collaboration approved tables of corrections for specific bins of p and θ . This PID correction is implemented at this stage of the analysis. Cases in which the relative data-to-MC systematic and statistical uncertainties are anomalous (i.e. above an arbitrary value of 200 defined in the PID tables) in the correction tables are removed from the tables such

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

that these particles are no longer covered by the corrections.

There are two other situations in which a particle is not covered by the PID correction package. One instance is when a fake particle is not a π , for example when a K is identified as an ℓ . In this case, i.e. when non-covered hadrons are faking leptons, the PID correction factor is set to 1 and all variations are also set to 1. If instead the tables do not cover the p and θ of the particle, the correction is generated from a Gaussian:

$$\bar{\epsilon}_{(\text{Data/MC})}^{\text{variation}} = 1 + \text{Gauss}(0, \sigma_{\text{average}}) \quad (6.9)$$

where the average uncertainty for a given table is calculated using:

$$\sigma_{\text{average}} = \sqrt{\bar{\sigma}_{\text{stat. \& sys.}}^2 + \bar{\epsilon}_{(\text{Data/MC})} - 1^2}. \quad (6.10)$$

There is higher occurrences of non-coverage in the low momentum region. It is a well known feature of the PID that the performances for muons and electrons drops significantly under 1 GeV, hence the coverage of the p bins drops too.

It should be noted that the PID weights released for analyst use are under continual study and improvement. At this stage, the PID weight tables show a significant inefficiency for the e corrections, and a smaller inefficiency for the μ corrections. This has the result of an asymmetry between the measured e and μ yields, as can be seen later in this study in lepton comparison plots and in tables such as Table 6.12.

6.3.5. $\Upsilon(4S)$

The $\Upsilon(4S)$ is now fully reconstructed with online and offline selections and corrections applied. Figure 6.15 displays the m_{miss}^2 distribution at this stage in the analysis. The normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$) in MC are **TopoAna** selected and correctly reconstructed events. The m_{miss}^2 distributions are also separated by lepton channel in Figures 6.16 and 6.17. Finally the D^* decay channels are displayed in Figures 6.18 to 6.22.

The asymmetry between the separated lepton distributions of the m_{miss}^2 is representative of the differing efficiencies for the e and μ channels in the FEI algorithm. In addition, the

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

PID has significant inefficiencies in the e PID weight tables. Both of these inefficiencies are the main contributors to the approximate 15% yield difference encountered.

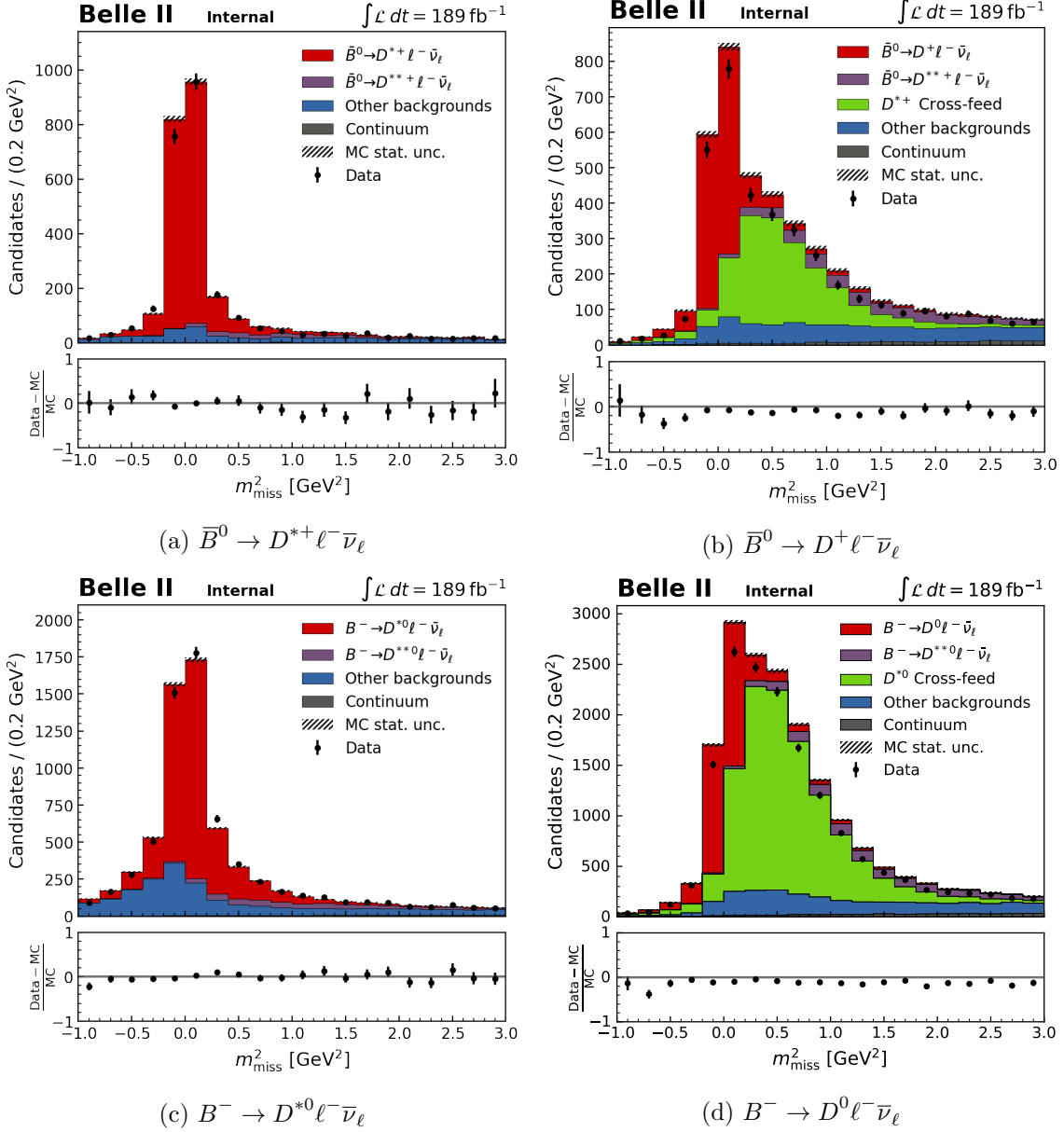


Figure 6.15.: The m_{miss}^2 after online and offline selections and corrections are applied. The distributions are shown after reconstruction of the $\Upsilon(4S)$ candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

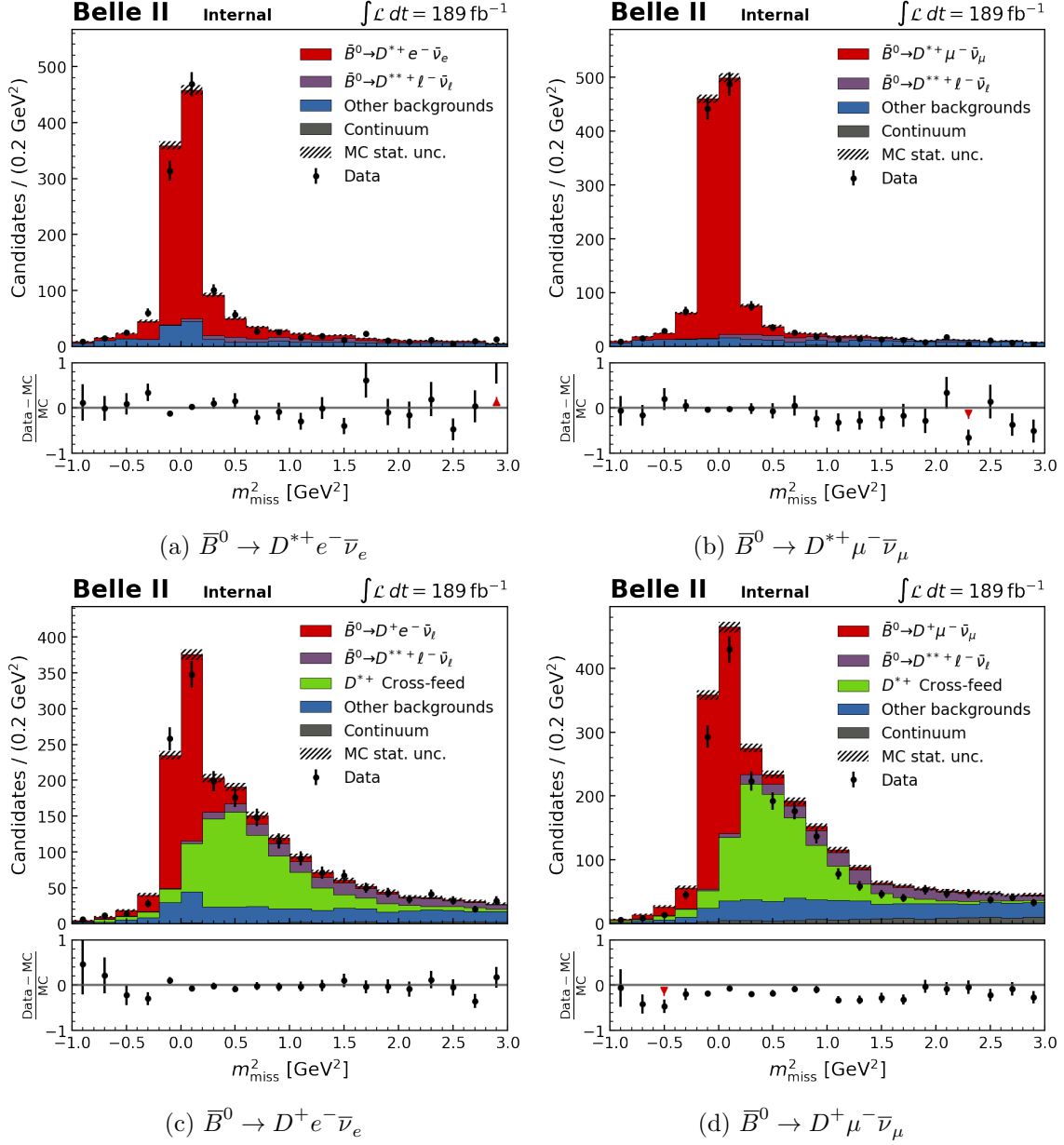


Figure 6.16.: The m_{miss}^2 after online and offline selections and corrections are applied, separated by lepton channel. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

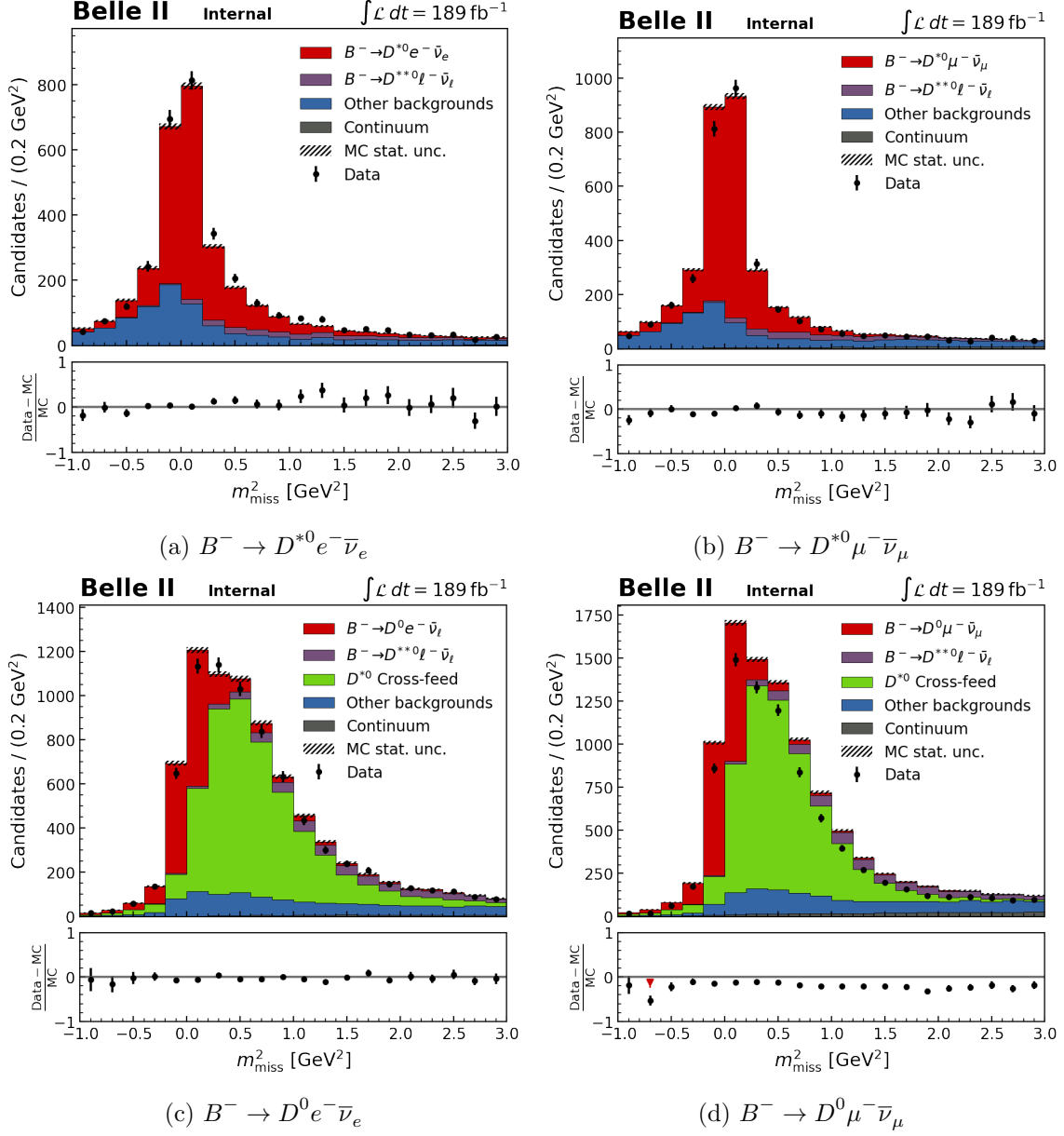


Figure 6.17.: The m_{miss}^2 after online and offline selections and corrections are applied, separated by lepton channel. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

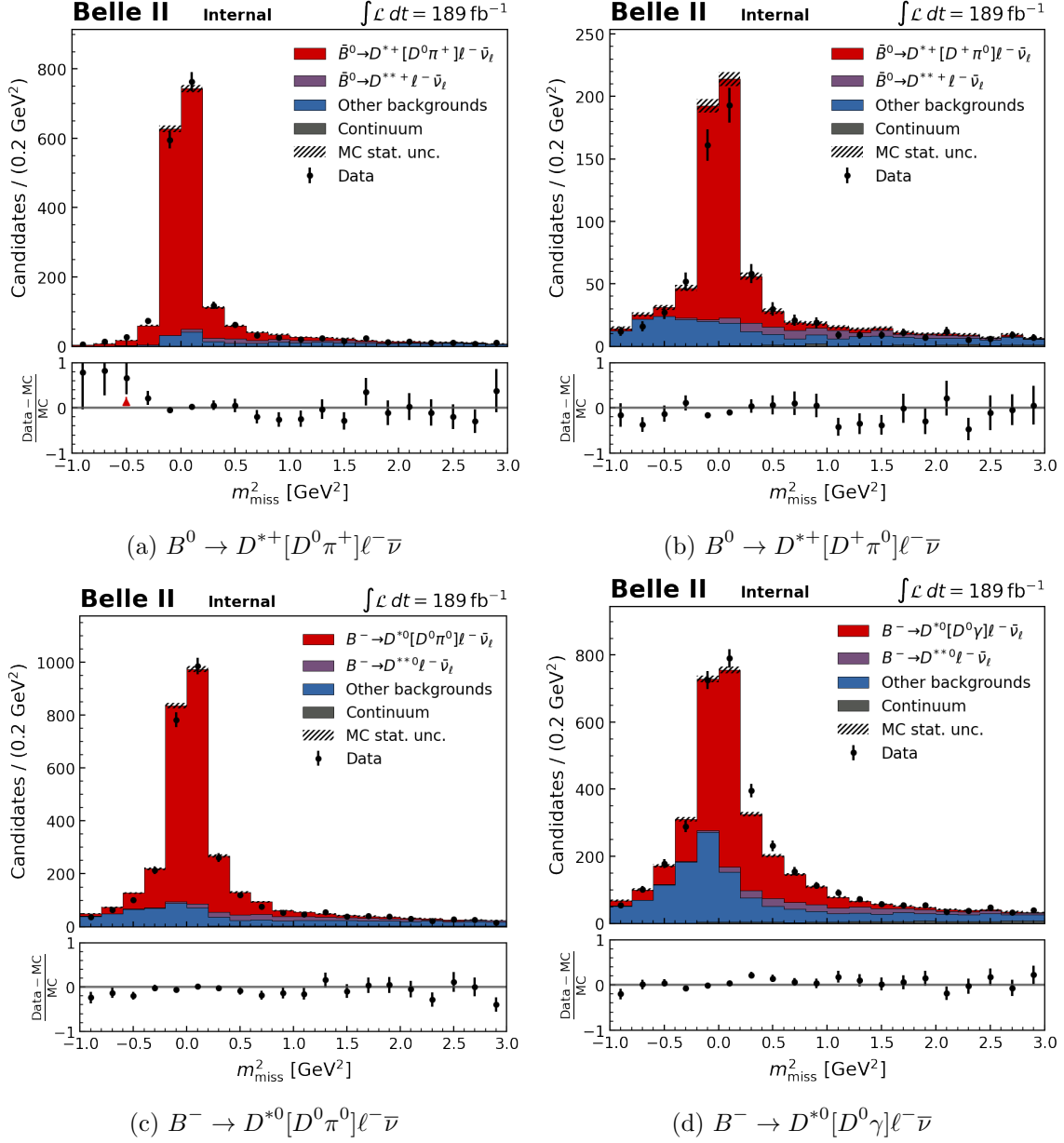


Figure 6.18.: The m_{miss}^2 after online and offline selections and corrections are applied, separated by D^* decay channel. Generic MC is classed as the normalisation channels ($B \rightarrow D^* \ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

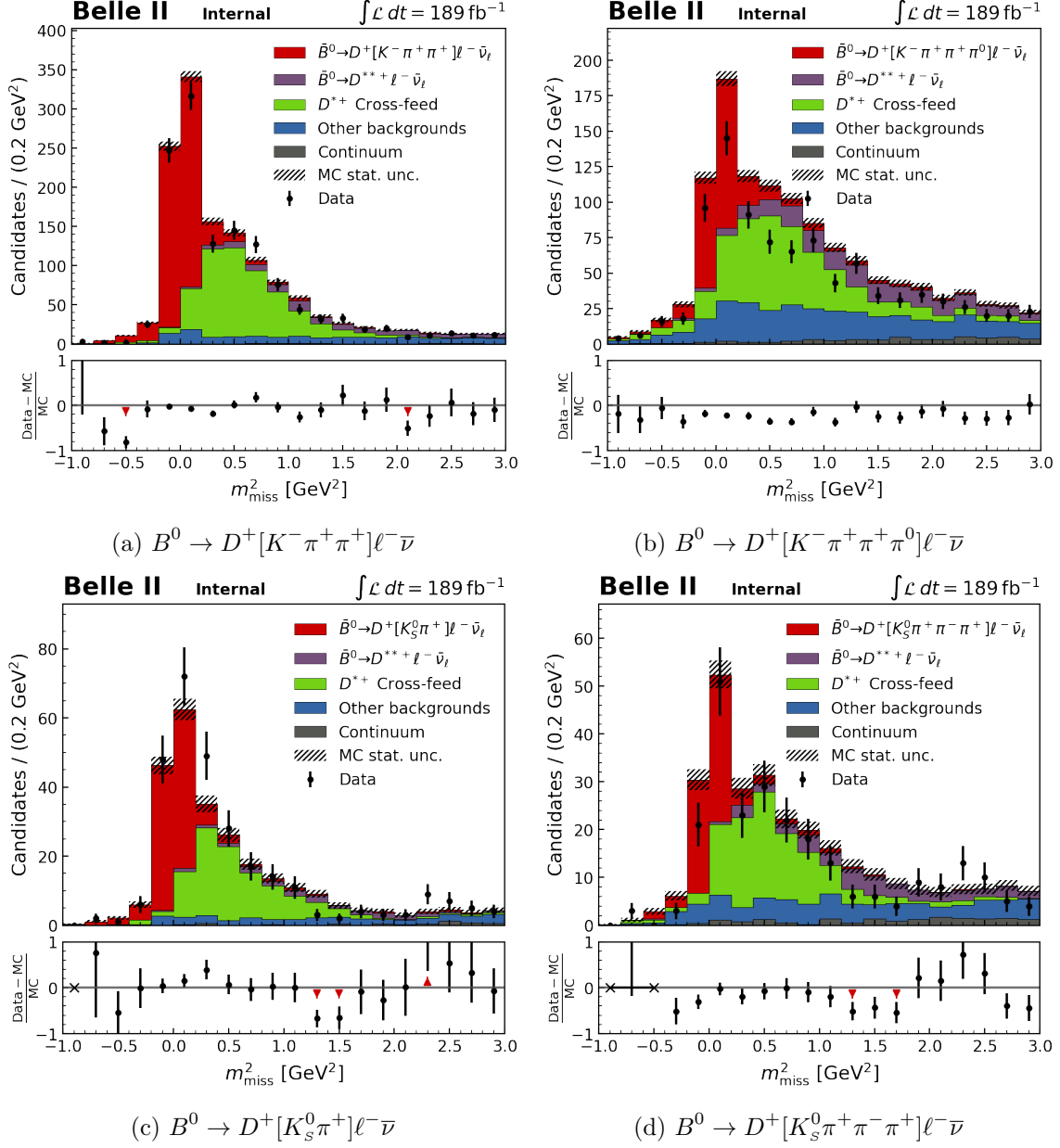


Figure 6.19.: The m_{miss}^2 after online and offline selections and corrections are applied, separated by D decay channel. Generic MC is classed as the normalisation channels ($B \rightarrow D\ell^-\bar{\nu}_\ell$), $B \rightarrow D^*\ell^-\bar{\nu}_\ell$ background, $B \rightarrow D\ell^-\bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

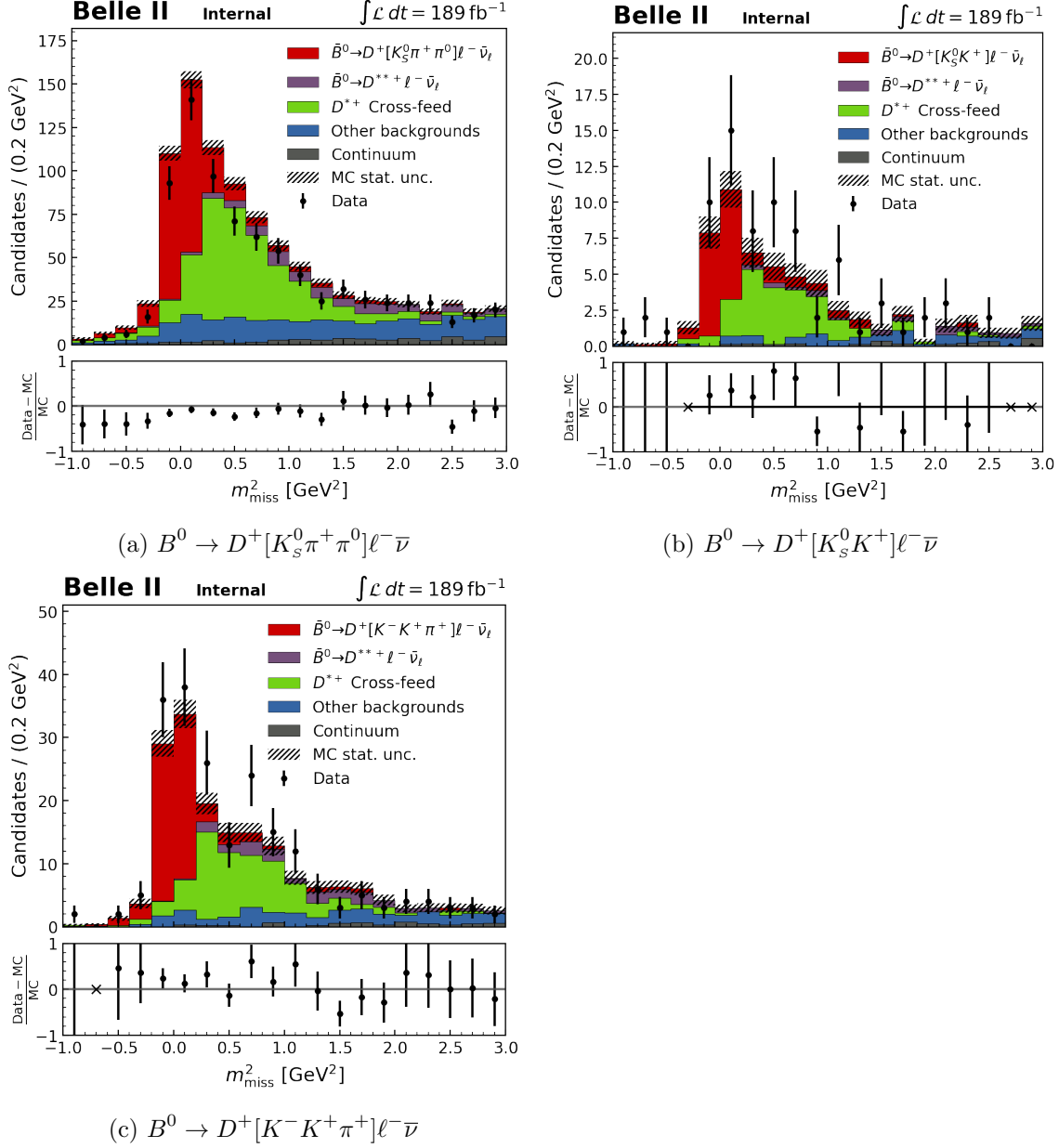


Figure 6.20.: The m_{miss}^2 after online and offline selections and corrections are applied, separated by D decay channel. Generic MC is classed as the normalisation channels ($\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$), $\bar{B}^0 \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background, $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

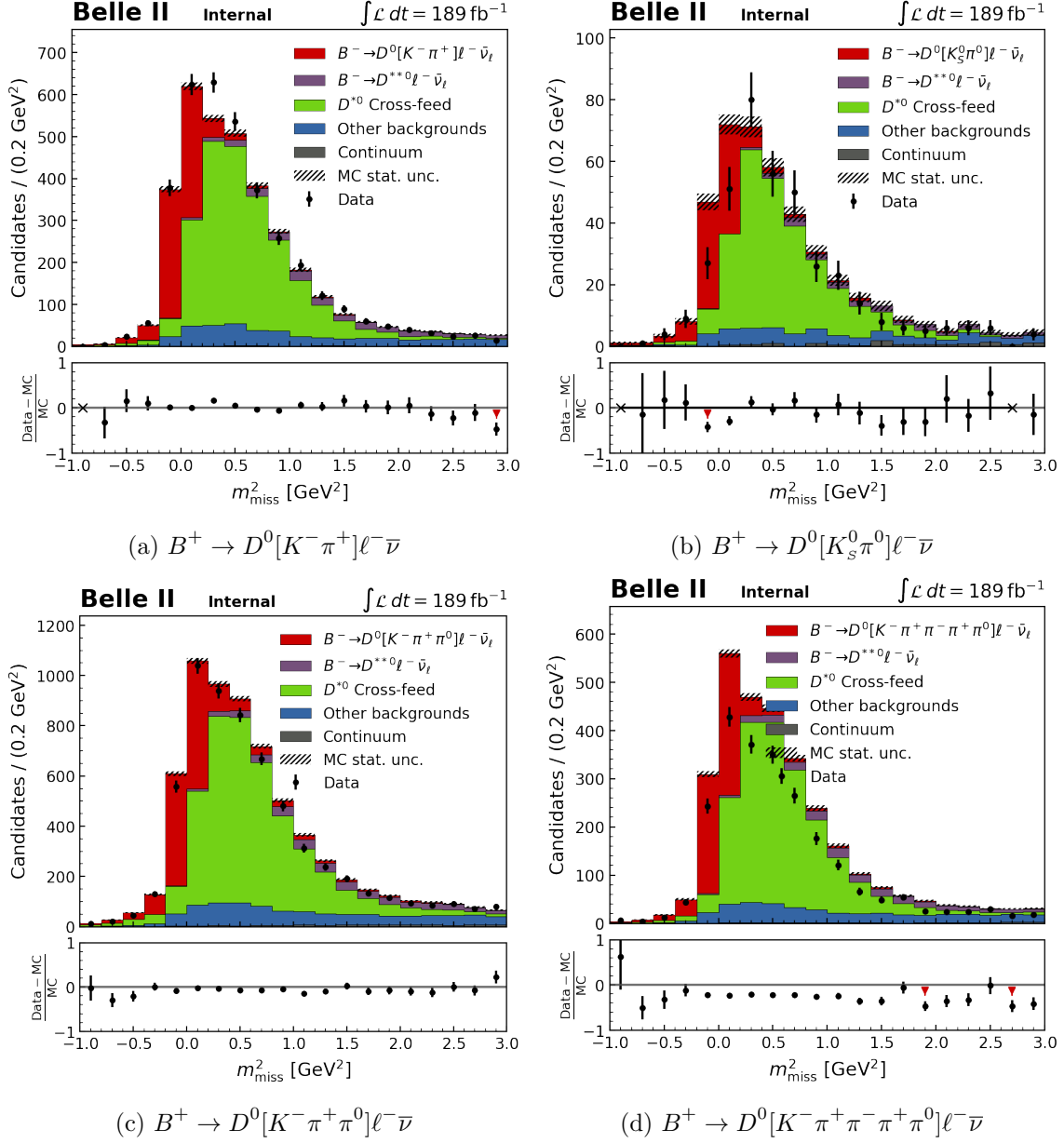


Figure 6.21.: The m_{miss}^2 after online and offline selections and corrections are applied, separated by D decay channel. Generic MC is classed as the normalisation channels ($B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$), $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background, $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

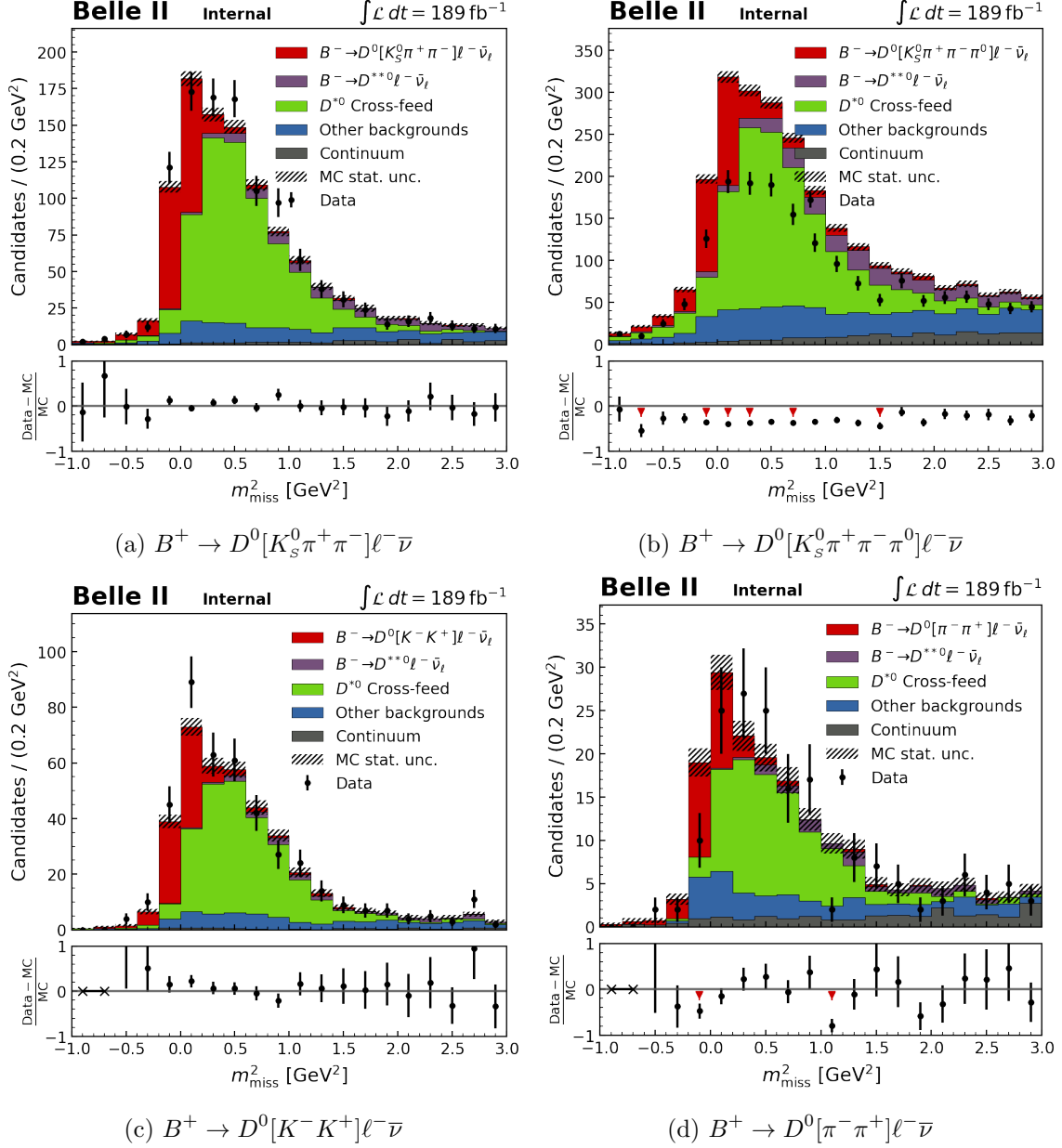


Figure 6.22.: The m_{miss}^2 after online and offline selections and corrections are applied, separated by D decay channel. Generic MC is classed as the normalisation channels ($B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$), $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ background, $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6.4. Multivariate analysis reducing specific peaking backgrounds

A distinct contribution of semileptonic background to the signal region is that of the $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ decay. D^{**} mesons are excited charmed mesons [64] that have higher energies than the $D^*(2010)$ meson, such as the $D^*(2420)^0$ meson. Appendix D describes the D^{**} meson in greater detail.

Mis-reconstructed particles from the D^{**} in the $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ decay channel such as π^0 cause the events to mimic signal and normalisation events, therefore it is a particularly difficult background to suppress. After the aforementioned online and offline signal selections, event yields in generic MC of background $\bar{B}^0 \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ ($\bar{B}^0 \rightarrow D^{**}\tau^- \bar{\nu}_\tau$) channels are a minimum of 6% (7%) of the size of the corresponding normalisation yield for $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ ($\bar{B}^0 \rightarrow D^{*+}\tau^- \bar{\nu}_\tau$), thus $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ is a large background contribution. Table 6.7 shows the size of contributing D^{**} backgrounds represented as a percentage of each relevant signal and normalisation channel of interest (COI).

D^{**} Background Corresponding COI		Size of D^{**} yield as a % of COI yield
$\bar{B}^0 \rightarrow D^{**}\ell^- \bar{\nu}_\ell$	$\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$	6
$\bar{B}^0 \rightarrow D^{**}\tau^- \bar{\nu}_\tau$	$\bar{B}^0 \rightarrow D^{*+}\tau^- \bar{\nu}_\tau$	7
$\bar{B}^0 \rightarrow D^{**}\ell^- \bar{\nu}_\ell$	$\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$	33
$\bar{B}^0 \rightarrow D^{**}\tau^- \bar{\nu}_\tau$	$\bar{B}^0 \rightarrow D^+\tau^- \bar{\nu}_\tau$	29
$B^- \rightarrow D^{**}\ell^- \bar{\nu}_\ell$	$B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$	10
$B^- \rightarrow D^{**}\tau^- \bar{\nu}_\tau$	$B^- \rightarrow D^{*0}\tau^- \bar{\nu}_\tau$	21
$B^- \rightarrow D^{**}\ell^- \bar{\nu}_\ell$	$B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$	31
$B^- \rightarrow D^{**}\tau^- \bar{\nu}_\tau$	$B^- \rightarrow D^0\tau^- \bar{\nu}_\tau$	29

Table 6.7.: The size of the D^{**} background yield in generic MC compared to a corresponding COI yield, represented as a percentage. Yields are after all selections prior to the MVA are applied.

In addition, current knowledge of the D^{**} decay is limited. Though the generic MC simulates the most up-to-date branching fractions, some B meson decays that include D^{**} mesons have a relatively high uncertainty, or have been ‘seen’ but not measured [5]. As this background process to $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ is little known, it is of great importance that Belle II measures the branching ratios in the future to be able to control and mitigate them. This however, will take much larger data samples than currently available at Belle II,

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

necessitating other methods with which to mitigate these backgrounds. Nonetheless, its identification and reduction is critical for this analysis, therefore a method of suppression of the $B \rightarrow D^{**}l^- \bar{\nu}_l$ decays is investigated through the implementation of a MVA.

The straightforward, traditional cut-based approach has been implemented so far in this thesis. These cuts reduced the data and MC to a manageable size and decent signal efficiency. However, multivariate techniques could benefit the analysis at this stage and improve signal efficiency and background rejection, beyond simple cuts on observables. For some observables, there is no signal-background signal separation power for independent cuts. These are therefore excluded from the cut-based method yet still have power within a multivariate analysis.

An MVA needs to be trained on how to classify events. This consists of supplying the model with artificially labelled events (a target variable) and the observable variables available to try to classify these events (training variables). Once tested, the model can be applied to un-labelled MC or data using just the available training variables.

The processes that will be classed as signal for this MVA analysis will be $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ (the normalisation channel) and $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ (the signal channel)^[iv]. As has been mentioned before, though the normalisation channel is the main study of this thesis, it is imperative in the measurements of $R(D)$ and $R(D^*)$ that both $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ and $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ have the same methods applied to each. Thus, this impacts the MVA target variable, as discussed in Section 6.4.2.

Four individual MVAs will have to be trained individually on the four $B \rightarrow D^{(*)}l^- \bar{\nu}_l$ possible combinations: $\bar{B}^0 \rightarrow D^{*+}l^- \bar{\nu}_l$, $\bar{B}^0 \rightarrow D^+l^- \bar{\nu}_l$, $B^- \rightarrow D^{*0}l^- \bar{\nu}_l$ and $B^- \rightarrow D^0l^- \bar{\nu}_l$. This is due to the differing signal selections and performance for the charged and neutral B mesons and the $D^{(*)}$ mesons. Different combinations of training variables are optimal for each decay channel due to the effects of over-training or high (anti-)correlations between variables.

Three separate subsets of the MC14ri.a production campaign are used for this analysis:

- an MC sample used for training the MVA model, referred to as the training sample;

^[iv] Referred to together as $B \rightarrow D^{(*)}l^- \bar{\nu}_l$.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

- an MC sample for evaluating the performance of the MVA model, referred to as the testing sample;
- and the MC and data used for the final measurement in this analysis (excluded from the training and testing samples). The trained MVA will be applied to this dataset.

MC samples equivalent to 800 fb^{-1} are available for this analysis of which 10% (80 fb^{-1}) is used to create testing and training samples (40 fb^{-1} each), and 90% (720 fb^{-1}) is left for the final measurement. In order to adequately estimate the performance of the MVA models the testing or 'validation' MC dataset must resemble expected data and be of equivalent size to the training set.

The final MC is scaled to the 189 fb^{-1} of real data collected by the Belle II experiment, therefore this 720 fb^{-1} sample is adequate for the analysis. The statistical uncertainties for MC are smaller than that of the data, but still significant. The 80 fb^{-1} however supplies a relatively small amount of signal events to be used for the training of the MVA. Thus this 80 fb^{-1} of generic MC is injected with the signal MC samples corresponding to those in Table 4.5. These $\mathcal{O}(10^7)$ signal events represent $\mathcal{O}(1 \text{ ab}^{-1})$ of data (depending on the signal decay channel). This shall be referred to as 'boosted MC' for the rest of this thesis. The boosted MC allows for the MVA model to be better trained to identify signal events. It results in a higher separation power at the training stage of the MVA.

6.4.1. Training variables

Multiple studies were performed to examine optimal training variables for the analysis. The correlations of the initial chosen variables with each other were investigated. Multiple combinations of variables were examined, ultimately to find that as long as the variables were not highly (anti-)correlated, they did not result in over-training and improved signal distinction power of the MVA. The training variables for this analysis first mirror those used in Reference [60] and then extend beyond them to improve the MVA performances. The selection was optimized and improved separately for each of the four MVAs that were trained. The final list of all of the training variables implemented is included in Table 6.8.

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

Variable	Description	MVA
E_{ECL}	Has the highest separation on its own between signal and D^{**} . The distribution of signal events is expected to peak at 0 GeV (Section 6.2.10). Semileptonic background has a flat distribution and continuum increases in the higher E_{ECL} regions.	All MVAs
ΔE	The distribution of signal events are expected to peak at 0 GeV and have a symmetric distribution about it. The asymmetry visible is due to a bias caused by the application of the FEI algorithm which uses this variable for selection.	All MVAs
$m_{D^{(*)}}$	The mass of the first daughter of B_{sig} .	All MVAs
ΔM_{D^*-D}	the difference between the masses of the first B_{sig} daughter and granddaughter (only for $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ decays).	$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$
$m_{D^{(*)}, \text{tag}}$	The mass of the B_{tag} first daughter.	All MVAs
$m_{D^*, \text{tag}} - m_{D, \text{tag}}$	The difference between the masses of the first B_{tag} daughter and granddaughter in the circumstance of a D^* daughter only.	All MVAs
N_{charged}	The number of charged B_{tag} daughters. The higher the number of charged particles, the higher the background yields.	$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$
$\cos(\Delta\theta_T)$	The cosine of the angle between the momentum of the D^* and the lepton frame, and the nominal B_{tag} meson.	All MVAs
M_{bc}	The beam constrained mass of the B_{tag} meson.	All MVAs
DMID of B_{sig}	The decay mode identification of the B_{sig} meson.	All MVAs
DMID of 1 st daughter of B_{sig}	The decay mode identification of the first daughter of the B_{sig} meson.	All MVAs
DMID of 1 st daughter of 1 st daughter of B_{sig}	The decay mode identification of the first daughter of the first daughter of the B_{sig} meson.	$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ & $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$

Table 6.8.: The list of various training variables used in the four MVAs. Variables that are not used to train all four MVAs have their specific usage indicated.

Figures 6.23 and 6.24 show various distributions of the variables used for the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ MVA. These variables were chosen for their discriminating power between signal and normalisation decays and their background $B \rightarrow D^{**}$ decays. A benefit of the intention to mitigate the D^{**} backgrounds is that the separation power between signal and normalisation decays and continuum decays is included in the MVA. Thus only one MVA is needed in the reduction of continuum and D^{**} decays, rather than two separate methods.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

Differing variables are used to the same effect for the $B \rightarrow D^*\ell^- \bar{\nu}_\ell$ and $B \rightarrow D\ell^- \bar{\nu}_\ell$ MVAs due to the structure of **basf2** variables for decays with daughter information at different levels. The variable of greatest ‘importance’ (i.e. gives most training power to the MVA) is the E_{ECL} , followed by $\cos(\Delta\theta_T)$ and $m_{D^{(*)}}$.

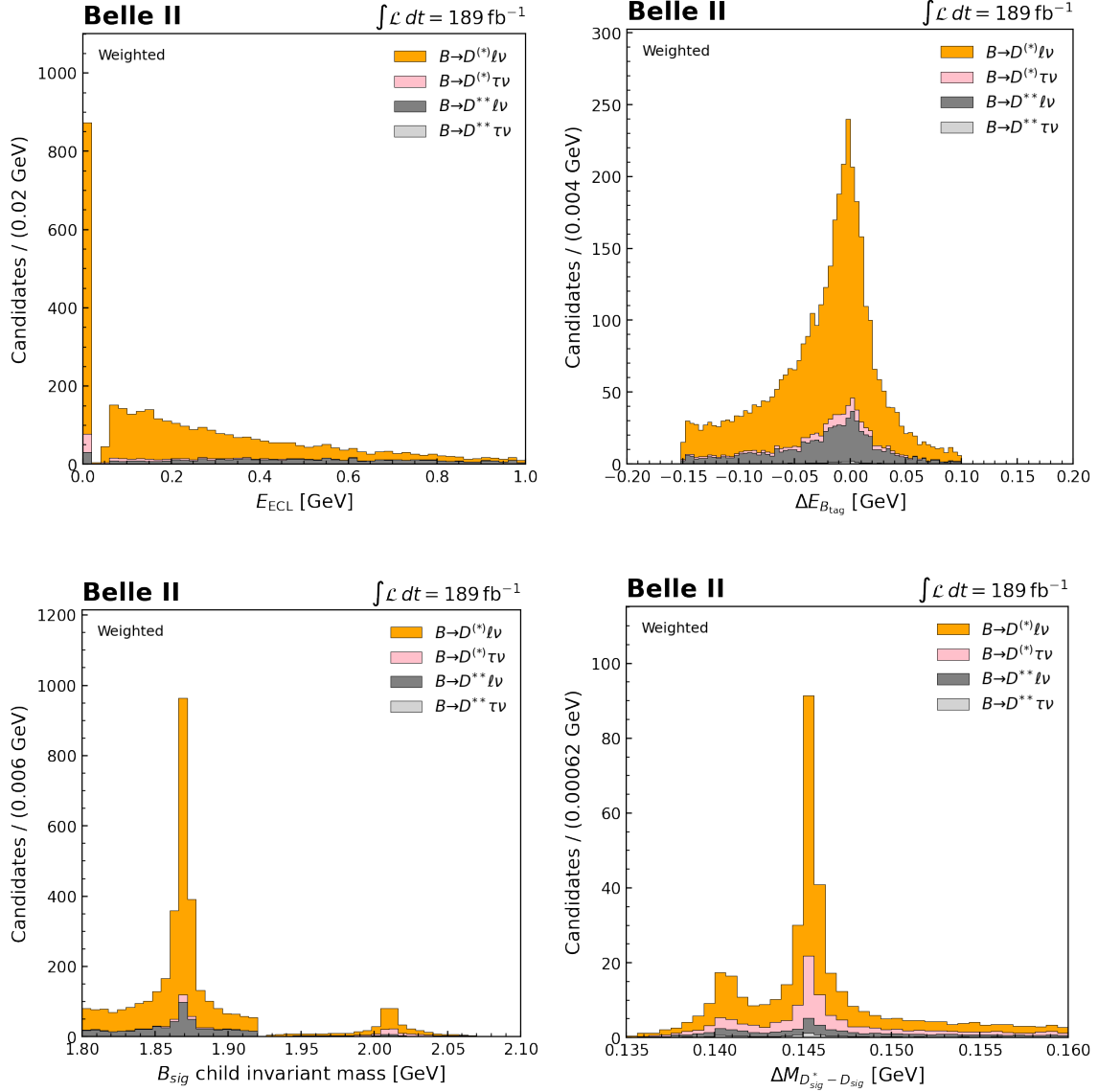


Figure 6.23.: The distributions of some of the MVA training variables for the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$. Signal MC is separated using **TopoAna** into the signal, normalisation and D^{**} backgrounds. ‘Weighted’ refers to the signal MC being weighted to the 189 fb^{-1} of data.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

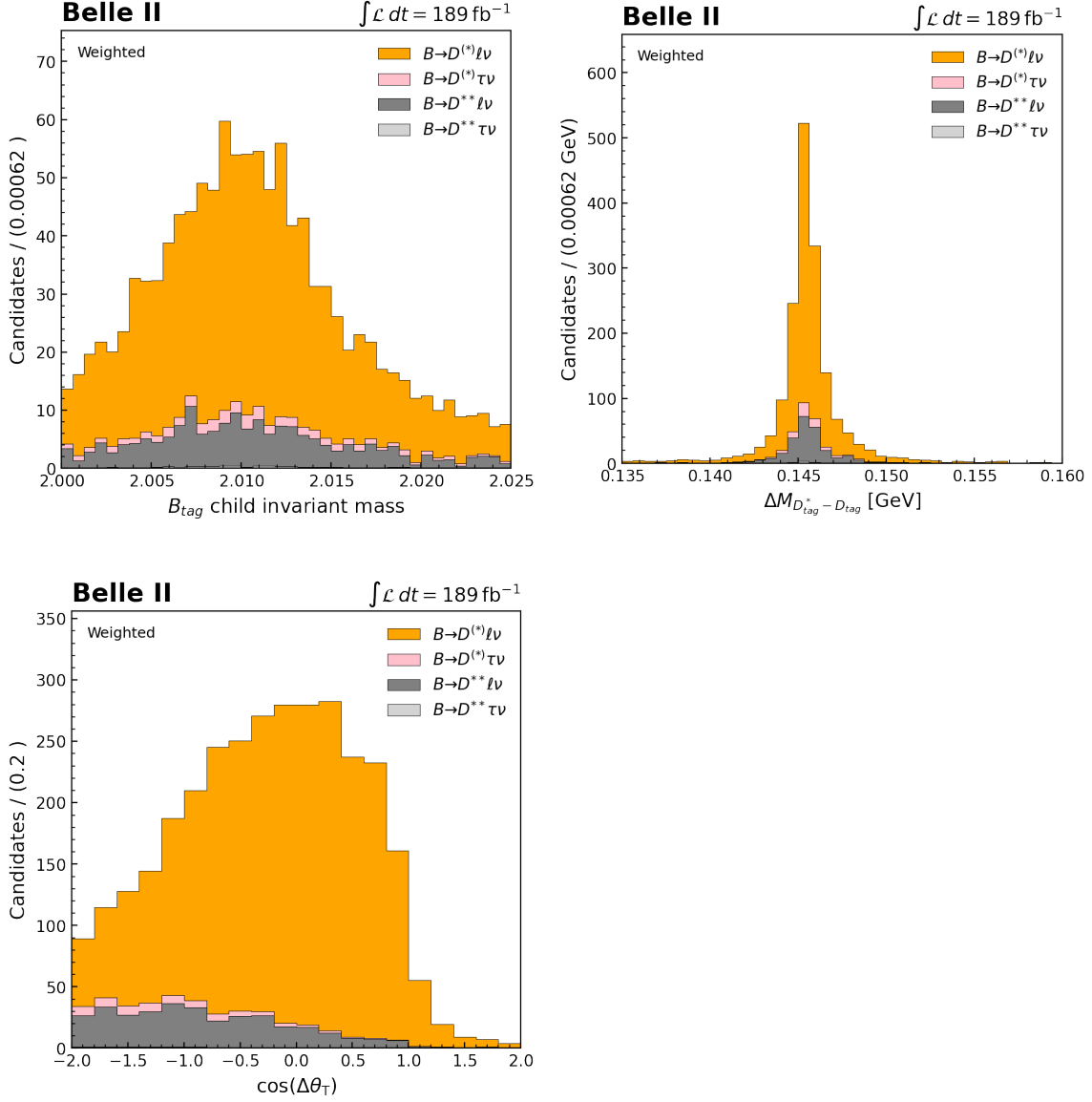


Figure 6.24.: The distributions of some of the MVA training variables for the $\bar{B}^0 \rightarrow D^{*+} l^- \bar{\nu}_l$. Signal MC is separated using **TopoAna** into the signal, normalisation and D^{*+} backgrounds. ‘Weighted’ refers to the signal MC being weighted to the 189 fb^{-1} of data.

6.4.2. Target variable

To inform the MVA models of how to distinguish which of the training events are signal decays, a boolean variable is created using **TopoAna** and is used to label and identify $B \rightarrow D^{*+} l^- \bar{\nu}_l$ and $B \rightarrow D l^- \bar{\nu}_l$ decays. In other words, to ensure use of the MVA on the $R(D)$ and $R(D^*)$ measurements, the MVA is trained to consider both the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

6. Analysis Method of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$

and $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ decays as signal.

At the end of training, the MVA model has learned how to classify events based on the training set, and indicates a probability of each event being either a background event at 0, or a signal event at 1. The classifier output can then be used as any other selection variable to improve the yield in signal versus backgrounds. The optimal selection value is chosen by considering purity as a function of classifier output alongside the FOM. The output from this model is then used on the testing sample with the target variable being withheld to test the competence of the model.

6.4.3. MVA Model

Firstly, multiple machine learning algorithms were evaluated for optimal performance within the MVA. The Receiver Operating Characteristic (ROC) curves [65] of 6 machine learning MVAs are compared: a Decision Tree (DT), a Gradient Boosted DT (GBDT), Neural Network (NN), a Linear Discriminant Analysis (LDA), a Support Vector Machine (SVM), and a Random Forest (RF). Each of the MVA models is trained on this training sample and then evaluated with the testing sample. To evaluate the performance of the MVA model used, the area under the curve (AUC) of the ROC curves are used as a FOM on the test datasets. In general, the larger the AUC (up to a value of 1), the better the performance of the MVA model. Figure 6.25 shows that most of the MVA models tested have comparative performances.

With hyper-parameter optimisation, the two models chosen for further study were the GBDT and the NN. Later a further variation of the GBDT model, the ‘FastBDT’ [66], was examined. Though the NN had a slightly lower AUC than the DT model, NNs have the ability to handle the same functions as GBDTs, but the difference is that multiple GBDTs would be needed to do the same executive function. Therefore, the NN model was chosen for further study over the DT model. Ultimately, the overall gain in performance of the NN compared to the speed and efficiency of the GBDT was insignificant. Furthermore, within the `basf2` framework, the FastBDT was found to have a negligible difference in performance and has lower run times compared to the GBDT. Thus, the FastBDT model is used in this MVA analysis and is described in detail in the next section.

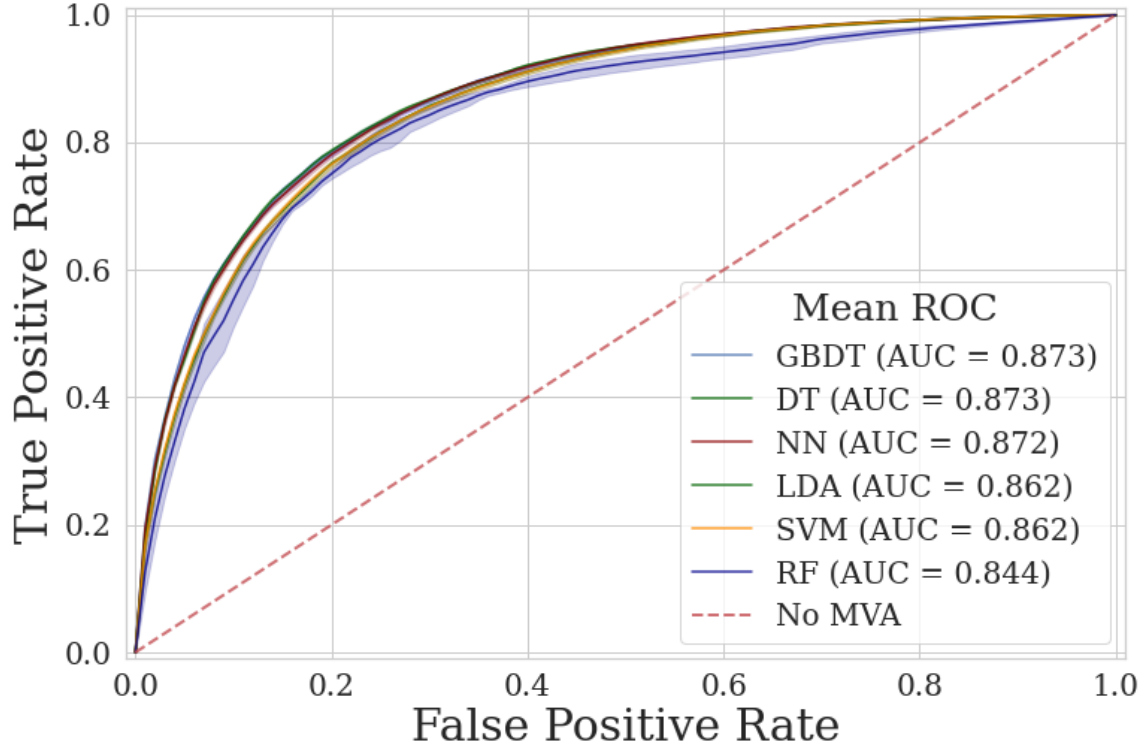


Figure 6.25.: The mean ROC of the preliminary machine learning techniques tested on the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays.

6.4.4. FastBDT

A DT uses supervised learning to map between inputs (‘training variables’) and a target variable. It classifies each event by considering pre-determined consecutive cuts. These cuts are ascertained through fitting to the training sample. At each branch of the DT, two cumulative probability histograms are created to determine the optimal signal-background separation for events that have passed all previous branches.

A stochastic GBDT differs from a simple DT during the fitting of the training data. ‘Boosted’ refers to the fact that multiple DTs with limited branch depths are constructed. Many of these ‘weak-learner’ DTs are then fitted to minimize the expectation value of a negative binomial log-likelihood function. ‘Gradient’ refers to ‘gradient-decent’, an algorithm applied between each boosting step that re-weights the training sample. Finally, ‘stochastic’ refers to the fact that the GBDT only uses an exclusive random sub-sample of training data.

The FastBDT is a stochastic GBDT optimized for reduced execution time at the imple-

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

mentation level [66]. Other than the optimal performance determined in comparison with other machine learning models, the reduced execution time is especially beneficial at the high luminosity Belle II experiment.

The FastBDT does not have the capability to distinguish between the different D decays we classify as signal in our data, so to be able to use this type of MVA on $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$, it is necessary to use four different FastBDTs, one for each decay: $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$, $B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$, $\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$. Therefore, MC samples are separated into the four decays for training, testing and eventual application of the four optimised FastBDTs.

The FastBDT performance was tested using various selections of the exclusive $D^{(*)}$ decays (listed in Table 5.1). Groupings were made to consider each individual decay channel, the effects of channels with and without neutral candidates, and higher multiplicity channels. Ultimately it is found that the effectiveness of the FastBDT allows use of the full decay channel list due to its good discrimination between signal and normalisation, background MC.

The FastBDT in the `basf2` framework was tuned using a grid-based optimization method for 3 available hyper-parameters. Using the AUC alongside the level of over-training and training-testing agreement as the FOMs, the following hyper-parameters were optimized: the number of trees N , the depth of the tree d and the shrinkage parameter λ . The number of trees is equivalent to the number of boosting steps and is set to $N = 200$ for all four FastBDTs. The depth of the tree corresponds to the maximum number of consecutive cuts. Here $d = 11$ ($d = 9$) for the $B \rightarrow D^{*}\ell^- \bar{\nu}_\ell$ ($B \rightarrow D\ell^- \bar{\nu}_\ell$) FastBDTs. The shrinkage power is a value dependent on the learning rate of the FastBDT η which is anti-correlated to the number of trees. Decreasing η means that the best value for N increases. It is set to $\lambda = 0.1$ for all four FastBDTs after iteration tests for a range of values.

With the recent update of the E_{ECL} MVA that now includes the hadronic split-offs, the performance of this MVA is affected. It was found that the purity and efficiency of events was increased and the rate of false positive identifications was reduced. The separation power between signal and background is similar, leading to the same value used for the optimized classifier output cut, however there are more signal events with a higher classification value, thus the MVA has more certainty in the identification of events.

6.4.5. MVA Results

In the following section the performance of each of the four optimized MVAs is reviewed. For succinctness, the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ MVA analysis is chosen as the focus for the figures. When not included in this chapter, equivalent plots for the other three decay channels are available in Appendix E.

Firstly, the training and testing samples are compared. Figure 6.26 (and Figure E.1) displays a comparison of the performance of the MVA on training and testing samples as a check for over-training. The overall good agreement between each of the training and testing samples indicate that the MVA is not over-trained in each of the MVAs.

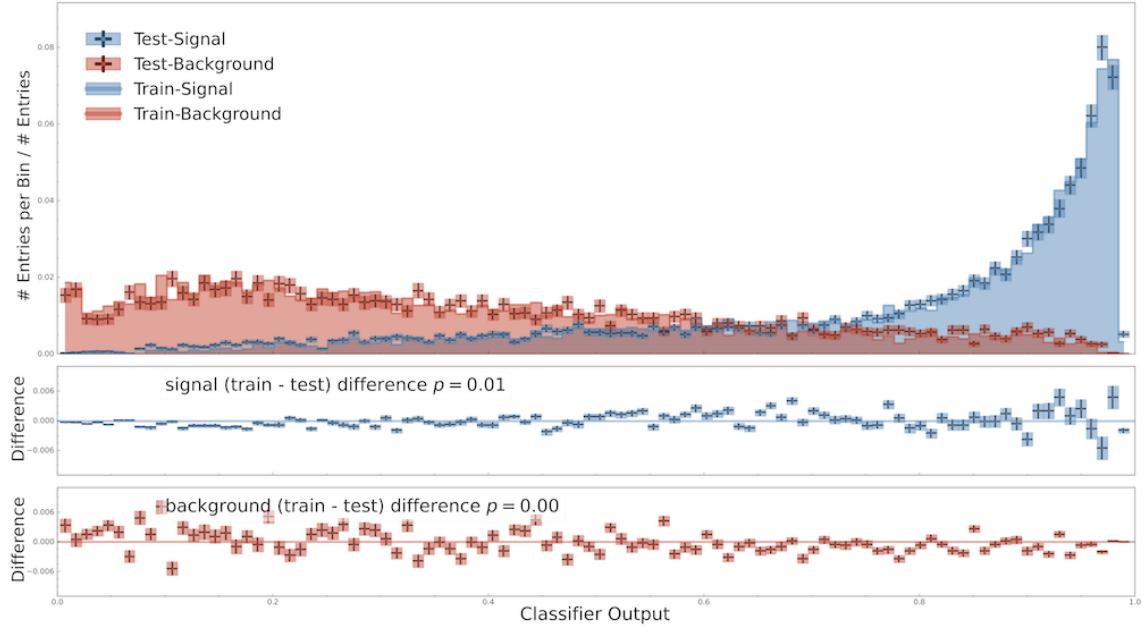


Figure 6.26.: The classifier output for the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ MVA. The red sample represents the backgrounds and the blue sample represents the signal and normalisation samples. The filled histograms are the training data, whilst the points with statistical error bars represent the test data. p represents the pull between testing and training data.

The purities, efficiencies, true positive rates and false positive rates of each MVA as a function of classifier output is presented in Figure ???. The selection made on the classifier output must be chosen carefully to optimize signal and normalisation yields without compromising effects on signal efficiency, purity and false positives. The high false positive rates and modest purities for lower values of the classifier output of the $B \rightarrow D \ell^- \bar{\nu}_\ell$ de-

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

cay channels in Figures E.2 are a result of the high D^* cross-feed present in the data-MC samples.

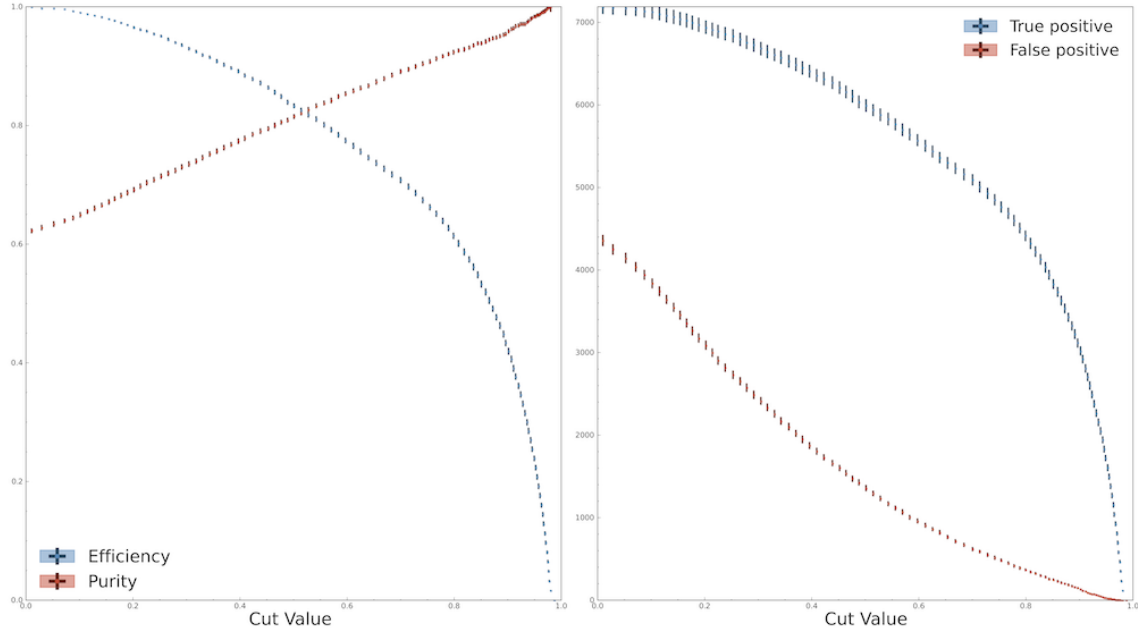


Figure 6.27.: The classification result the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ MVA. The purity (left plot, red) and efficiency are shown (left plot, blue), and the true positive (right plot, blue) and false positive rates (right plot, red) are shown (right), all as a function of classifier output cut.

The application of the MVA to the testing sample allows comparison of its AUC to that of the training sample. All four MVAs show adequate AUC values and agreement, resulting in the conclusion that there is little-to-no over-training in the each of the MVAs. The ROCs and AUC values are shown in the top right corner for the train and test samples for $B \rightarrow D^*\ell^- \bar{\nu}_\ell$, as an example in Figure 6.28.

The standard FOM defined in Equation 6.8 is implemented to indicate potential optimal classifier cuts. Tables 6.9 and 6.10 show the FOM for the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ decay channel with the yields and percentage drop in the yields of the generic MC as a function of a cut on the MVA classifier. As can be seen from the tables, the FOM is a decent guideline for the selection of the value of the classifier, however simply choosing the highest FOM may not be optimal, as the small differences in FOM often have significant impacts on the data. Thus other decision factors are used to choose an optimal classifier cut.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

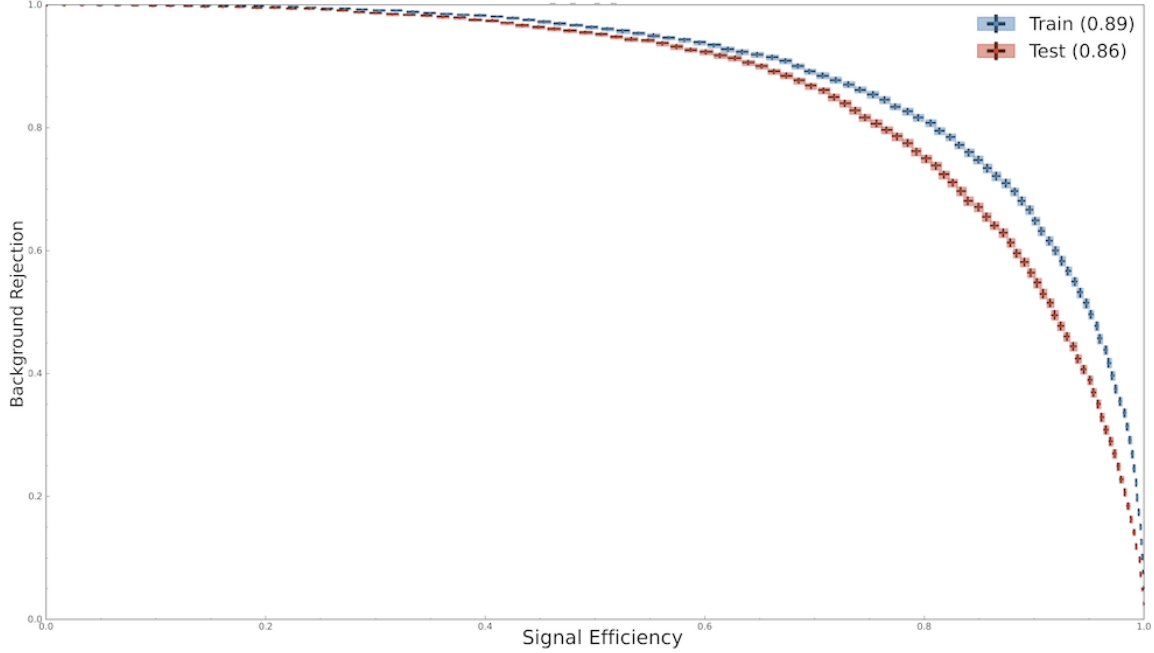


Figure 6.28.: The ROC test result (red) in comparison to the training ROC (blue) for the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ MVA. AUC values are shown and indicate good agreement.

For the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ MVA, there is only a small difference in FOM value for the classifier output cuts between 0.3 and 0.7, however, the $B \rightarrow D^*\tau^- \bar{\nu}_\tau$ percentage drop in yield reaches 24.65% for 0.7 in part due to a low $B \rightarrow D^*\tau^- \bar{\nu}_\tau$ signal yield in generic MC.

The effect of the classifier output on each of the signal and generic MC groupings is then examined in Figure 6.29. Here the percentage drop in yield is presented as a function of classifier cut. The D^{**} background contribution has a higher percentage reduction in yield compared to the normalisation and signal decays, indicating a successful differentiation between them.

The $B \rightarrow D^{(*)}\tau^- \bar{\nu}_\tau$ samples indicate a higher percentage drop at lower classifier cut values than that of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels. This is in part due to the lower branching fraction and therefore lower statistics available for the calculation. In this situation the application of this MVA would have impact the $R(D)$ and $R(D^*)$ measurement at the current data luminosity. Thus for this analysis, signal MC was used to determine an optimal classifier cut. The selection of the classifier output cut must therefore be re-

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$	2041	2040	2037	2030	2015	1991	1948	1881	1781	1509
Background	552	547	519	481	443	394	333	276	222	148
$\bar{B}^0 \rightarrow D^{*+}\tau^- \bar{\nu}_\tau$	34	34	33	33	32	30	28	25	22	17
$\bar{B}^0 \rightarrow D^{**}\ell^- \bar{\nu}_\ell$	140	139	133	124	113	96	77	60	44	23
Continuum	22	20	17	13	10	9	7	5	3	2
FOM	39.91	39.96	40.16	40.41	40.56	40.69	40.72	40.46	39.76	37.05

Table 6.9.: The yield for the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^{*+}\tau^- \bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. Note that the number of events is rounded to an integer value (weighting renders events non-integer values).

Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$	0.00	0.01	0.18	0.51	1.25	2.44	4.55	7.80	12.71	26.05
Background	0.00	0.90	5.96	12.82	19.66	28.63	39.53	49.96	59.71	73.19
$\bar{B}^0 \rightarrow D^{*+}\tau^- \bar{\nu}_\tau$	0.00	0.00	1.19	3.41	6.23	10.43	17.82	24.65	33.45	50.13
$\bar{B}^0 \rightarrow D^{**}\ell^- \bar{\nu}_\ell$	0.00	0.22	4.47	11.54	19.29	30.93	44.87	56.57	68.11	83.36
Continuum	0.00	6.70	21.83	37.92	51.31	58.57	65.37	77.08	82.13	87.74
FOM	39.91	39.96	40.16	40.41	40.56	40.69	40.72	40.46	39.76	37.05

Table 6.10.: The percentage drop in yield for the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^{*+}\tau^- \bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The percentage drop in yields shown are calculated from the non-rounded yields in Table 6.9.

optimized in the future to ensure optimal performance on generic MC and, later, data. This is discussed further in Section 9.3.

The benefit of the MVA being trained against D^{**} backgrounds is that for $B \rightarrow D\ell^- \bar{\nu}_\ell$ decays, this actively trains against D^* cross-feed.

The continuum MC, which has higher presence in the $B \rightarrow D\ell^- \bar{\nu}_\ell$ decay channels, is drastically reduced at low classifier output values.

Finally, a requirement of the classifier output for an event being more than 0.7 for $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ is chosen as the optimal cut due to the clean normalisation channel signature, even though the looser cut on the MVA would allow for a higher retention of $B \rightarrow D^*\tau^- \bar{\nu}_\tau$ decays. This is justified with the comparison to signal MC in Figure 6.29a, in which normalisation and signal channels have a similar shape and percentage drop.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

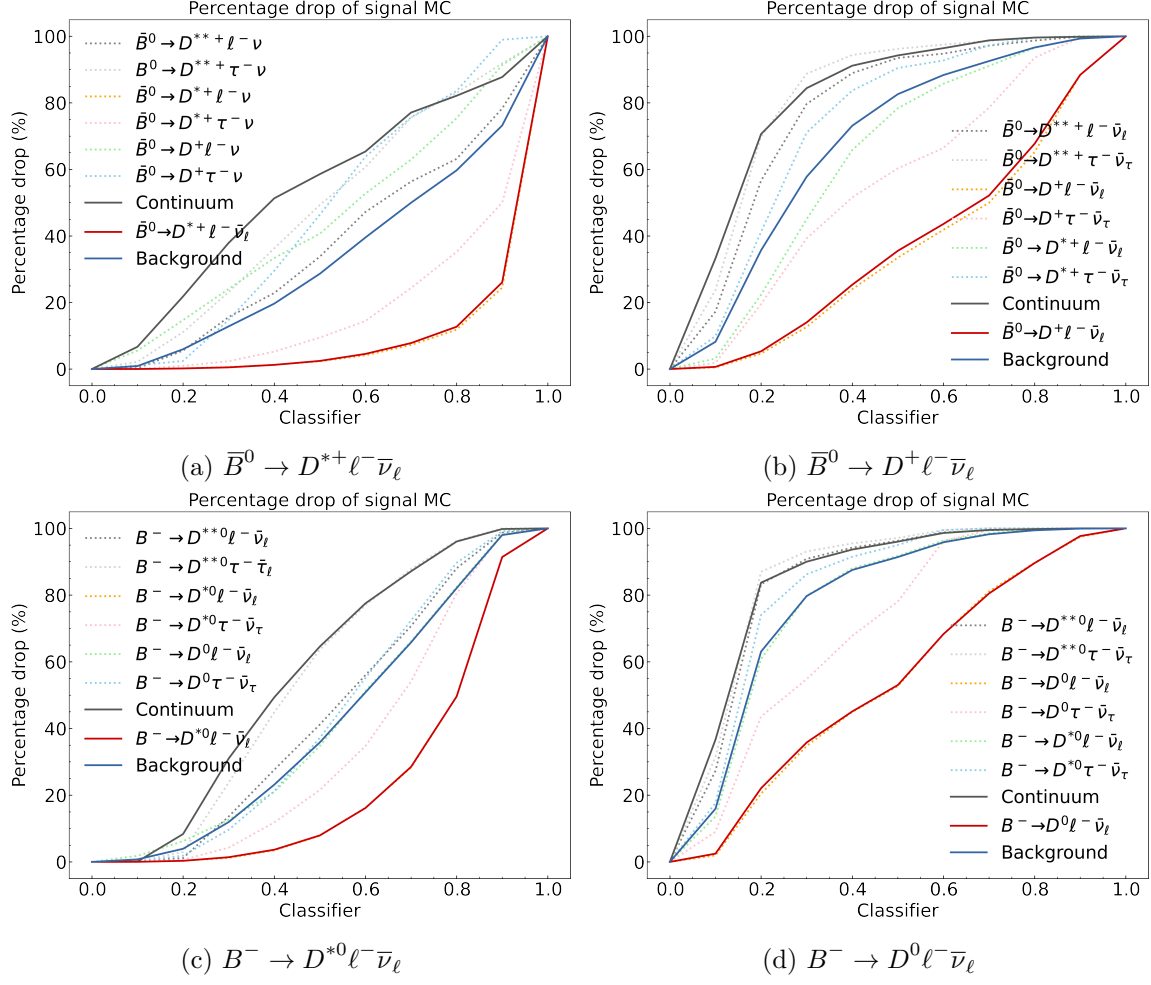


Figure 6.29.: The percentage drop in MC type as a function of MVA output classifier cut. The dotted coloured lines represent dedicated signal MC. The solid coloured lines represent generic MC split into the respective **TopoAna** selected $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decays, background decays and continuum.

The normalisation and signal yields indicate that the higher cuts have more impact on backgrounds and continuum than the normalisation channel. In addition, the data-MC agreement as a function of classifier cut is used to determine the optimal classifier output cut. Figure 6.30 shows an example plot for the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ decay channel. The data-MC agreement slightly improves for the $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$

Overall, the optimal classifier cut is found to be 0.7 (0.4) for $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ ($\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$) and 0.5 (0.2) for $B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$ ($B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$). The classifier selections are represented on the distributions of the output classifiers in Figure 6.31. The effect of each of the optimized classifier cuts on the event yields for normalisation, backgrounds and

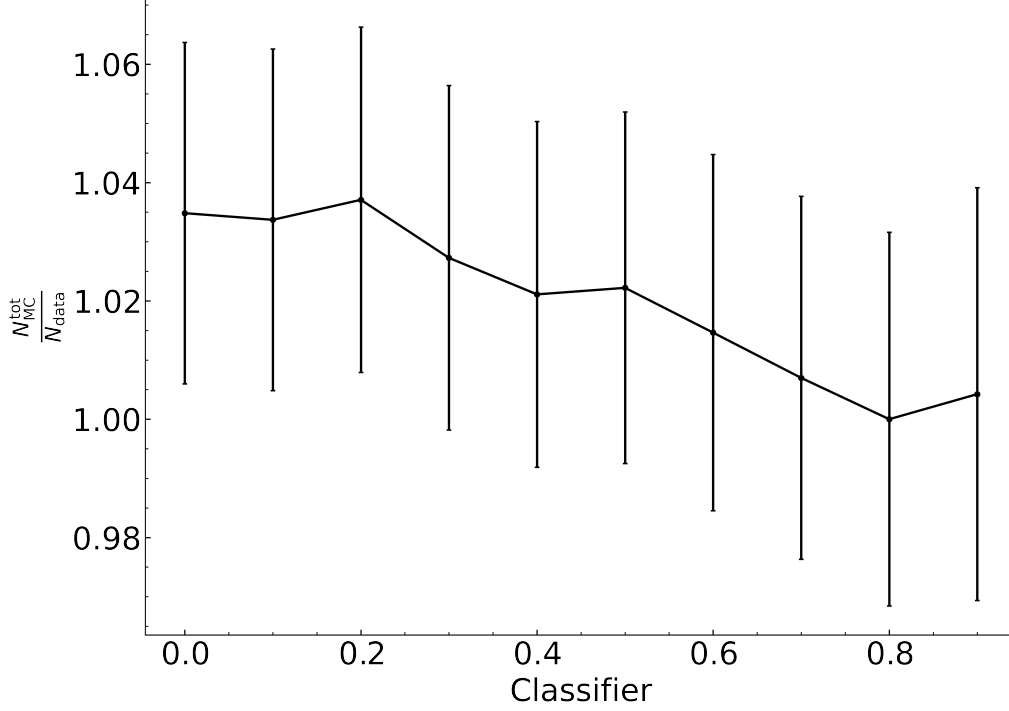


Figure 6.30.: The data-MC ratio as a function of classifier cut for the $\bar{B}^0 \rightarrow D^{*+} l^- \bar{\nu}_l$ decay channel. Errors shown are statistical uncertainties.

continuum events are summarized in Table 6.11.

Overall, the MVAs show good signal-background separation for $B \rightarrow D^* l^- \bar{\nu}_l$ decays, with reduced performance in the MVAs for $B \rightarrow D l^- \bar{\nu}_l$ decays. Here the performance is severely affected by D^* cross-feed present in the $B \rightarrow D l^- \bar{\nu}_l$ samples, referred to as fake D^* candidates. Appendix F discusses further studies performed by the author to mitigate this cross-feed issue. The MVA performance for $B \rightarrow D l^- \bar{\nu}_l$ decay channels can be improved through reduction of fake D^* contributions through a linear cut on m_{miss}^2 and E_{ECL} , however the method cannot be used in the $R(D)$ and $R(D^*)$ measurement due to acute $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ signal reduction. That being said, although fake D^* candidates cause a loss in performance of the MVA on the D meson decay channels, the MVA itself reduces the fake D^* candidates significantly, ultimately improving data-MC agreement.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

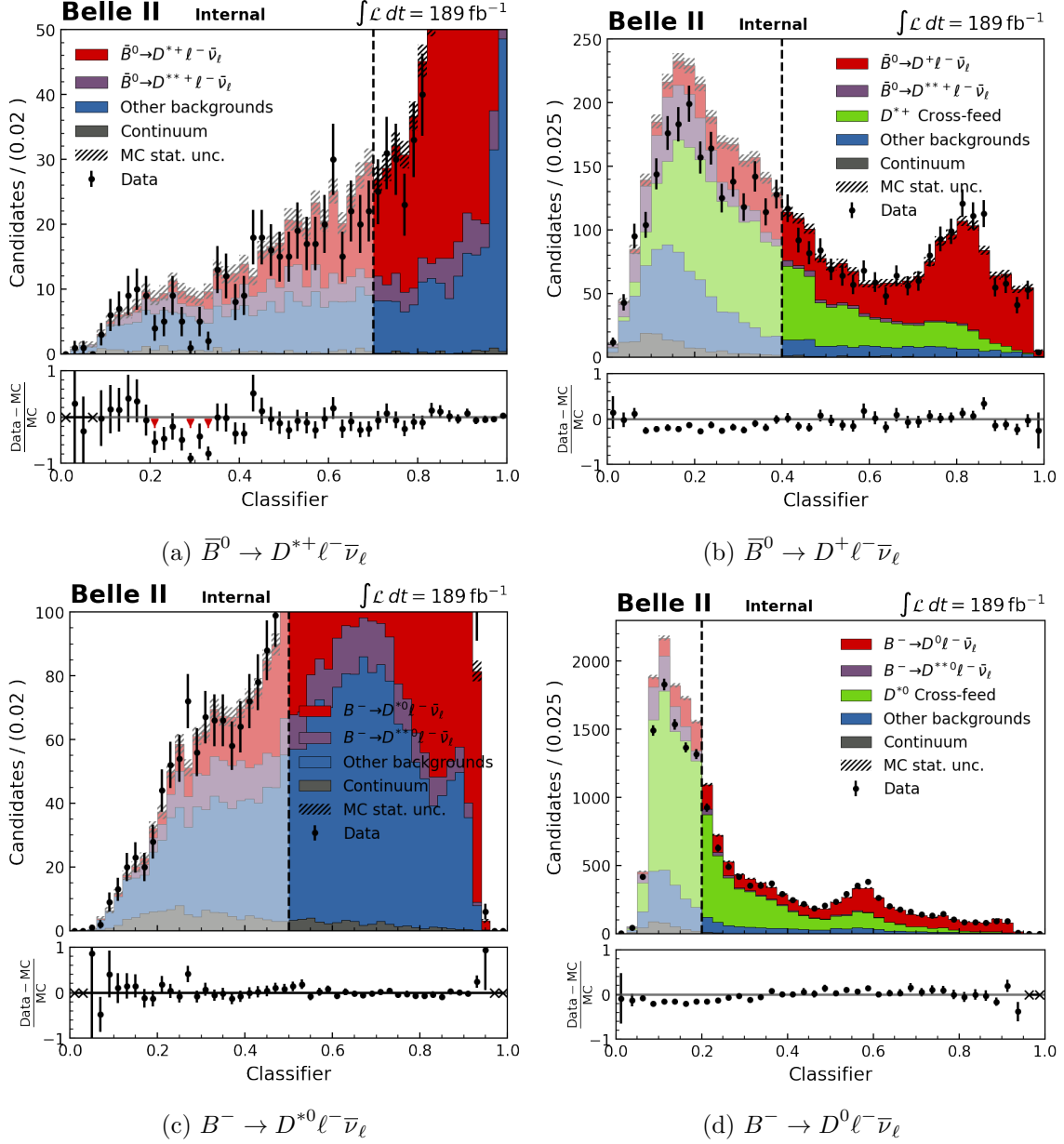


Figure 6.31.: The output classifiers for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ MVAs. Where appropriate, the y-axes have been optimized for visualisation of backgrounds. Generic MC is classified as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

Normalisation Channel	MC type	Yield Before MVA classifier cut	Yield After MVA classifier cut	Percentage Drop (%)
$\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$	$\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$	2041	1881	7.80
	Background	552	276	49.96
	$\rightarrow \bar{B}^0 \rightarrow D^{*+}\tau^- \bar{\nu}_\tau$	34	25	24.65
	$\rightarrow \bar{B}^0 \rightarrow D^{**}l^- \bar{\nu}_l$	140	60	56.57
	Continuum	22	5	77.08
$\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$	$\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$	1379	1029	25.33
	Background	2709	728	73.12
	$\rightarrow \bar{B}^0 \rightarrow D^+\tau^- \bar{\nu}_\tau$	27	11	55.39
	$\rightarrow \bar{B}^0 \rightarrow D^{**}l^- \bar{\nu}_l$	399	34	91.37
	Continuum	140	12	91.16
$B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$	$B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$	4154	3882	6.56
	Background	2305	1534	33.43
	$\rightarrow B^- \rightarrow D^{*0}\tau^- \bar{\nu}_\tau$	73	55	25.55
	$\rightarrow B^- \rightarrow D^{**}l^- \bar{\nu}_l$	404	234	42.01
	Continuum	138	42	69.44
$B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$	$B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$	3620	2821	22.09
	Background	13508	4989	63.07
	$\rightarrow B^- \rightarrow D^0\tau^- \bar{\nu}_\tau$	78	40	48.84
	$\rightarrow B^- \rightarrow D^{**}l^- \bar{\nu}_l$	1121	187	83.30
	Continuum	396	64	83.73

Table 6.11.: The yields per generic MC grouping before and after the MVA selection is applied to each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ channels. The $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values). The percentage drop in yields shown are calculated from the non-rounded values.

6.4.6. The results of the applied MVA

The resulting m_{miss}^2 distributions, after the MVA classifier selections are performed, are presented in Figures 6.32. In comparison with those of Figure 6.15 without the MVA classifier selection, a distinct reduction in background contribution can be seen. Additional m_{miss}^2 distributions with separation into different lepton ℓ modes, e and μ (Figures 6.33 and 6.34), and different D^* decay channels (Figure 6.35) are shown. The corresponding yields are presented in Table 6.12. The current binning is kept to match the [1] binning for ease of comparison of results. Appendix G contains alternate binning histograms for consolidation.

The Figures 6.33 and 6.34 indicate an approximate 15% yield asymmetry between e and μ decay channels. The largest cause of this is the inefficiencies present in the ever-evolving lepton-ID correction factors, which were recently reported to have significant inefficiencies for μ and smaller inefficiencies for e modes, as mentioned in 6.3.4. These are in the process of being updated, and improve with every MC iteration. In addition, though muons and electrons have similar misidentification rates at Belle II, the pion rate of misidentification as electrons and muons differs significantly. The misidentification for pions as electron candidates is $\left(\pi^\pm \rightarrow e^\pm\right) = \left(4.11_{-0.20}^{+0.20}(\text{stat.}) \quad {}^{+<0.01}_{-<0.01}(\text{syst.})\right) \times 10^{-3}$ and for pions as muon candidates is $\left(\pi^\pm \rightarrow \mu^\pm\right) = \left(7.33_{-0.01}^{+0.02}(\text{stat.}) \quad {}^{+0.01}_{-0.01}(\text{syst.})\right) \times 10^{-2}$ [67].

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

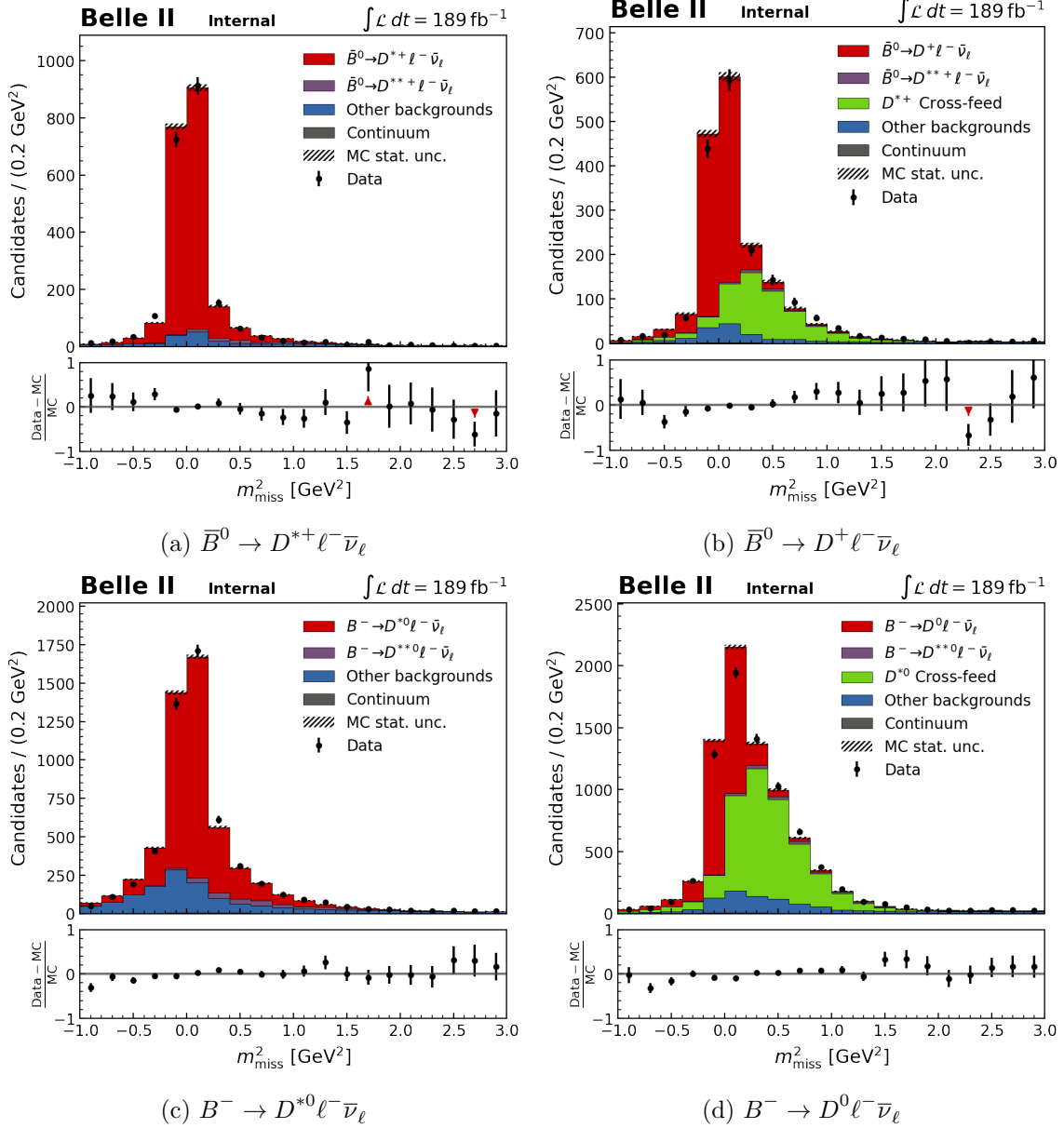


Figure 6.32.: The post-MVA classifier cut m_{miss}^2 for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ MVAs. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum. Errors shown are statistical uncertainties.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

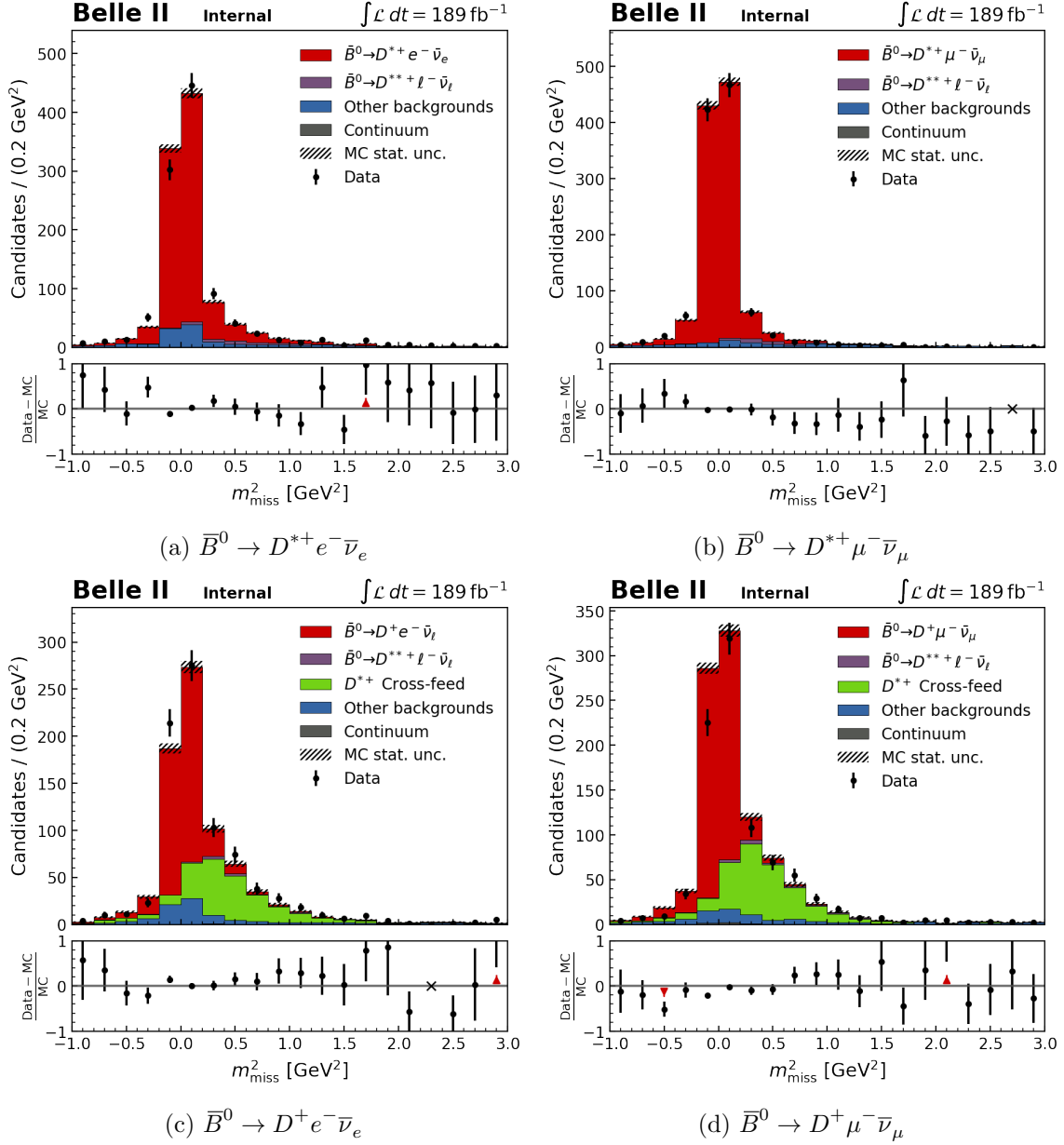


Figure 6.33.: The post-MVA classifier cut m_{miss}^2 for each of the $\bar{B}^0 \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ MVAs separated by lepton candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**}\ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum. Errors shown are statistical uncertainties.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

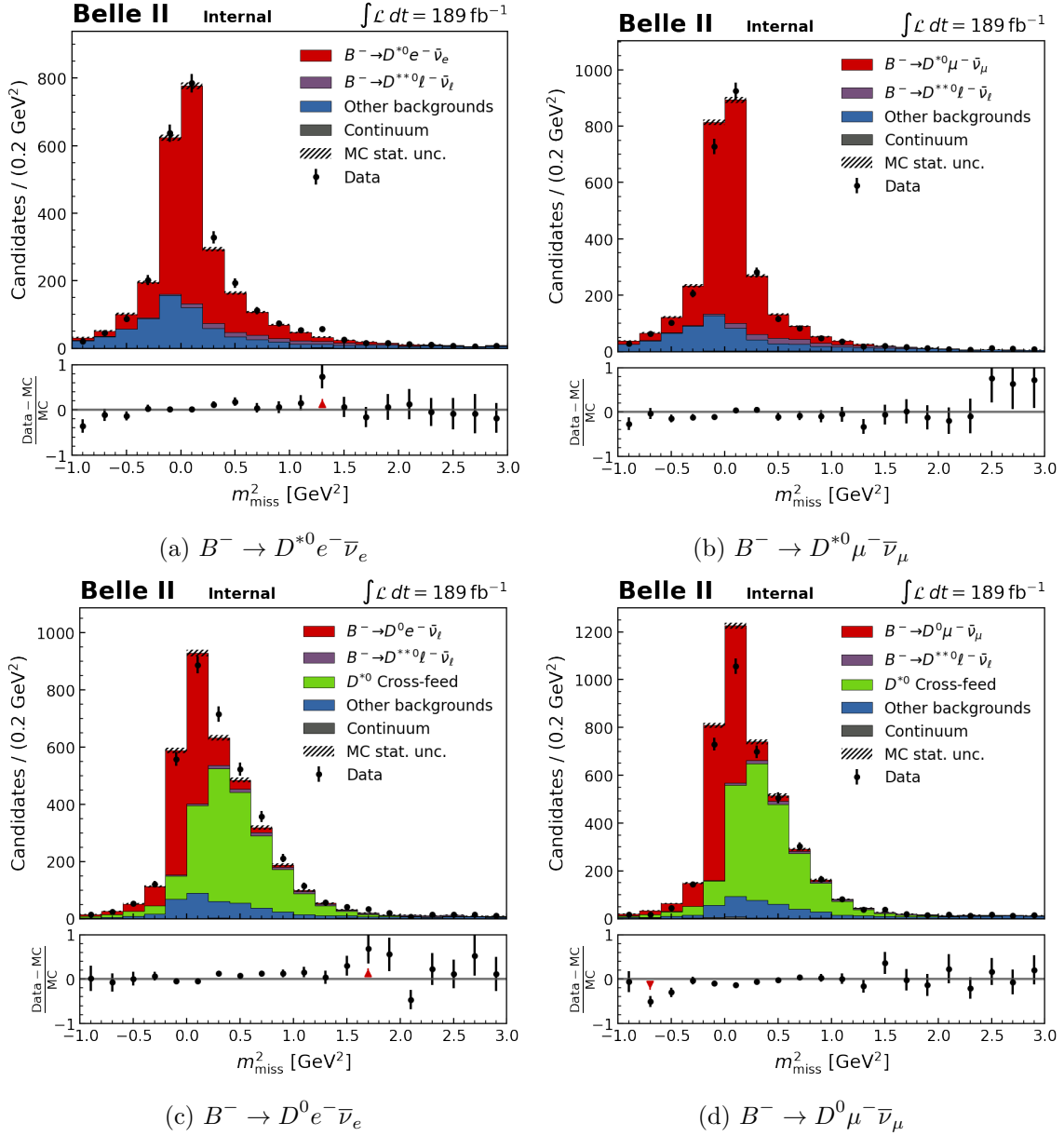


Figure 6.34.: The post-MVA classifier cut m_{miss}^2 for each of the $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ MVAs separated by lepton candidate. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$), $B \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background, $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ with a fake D^* candidate, other backgrounds and continuum. Errors shown are statistical uncertainties.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

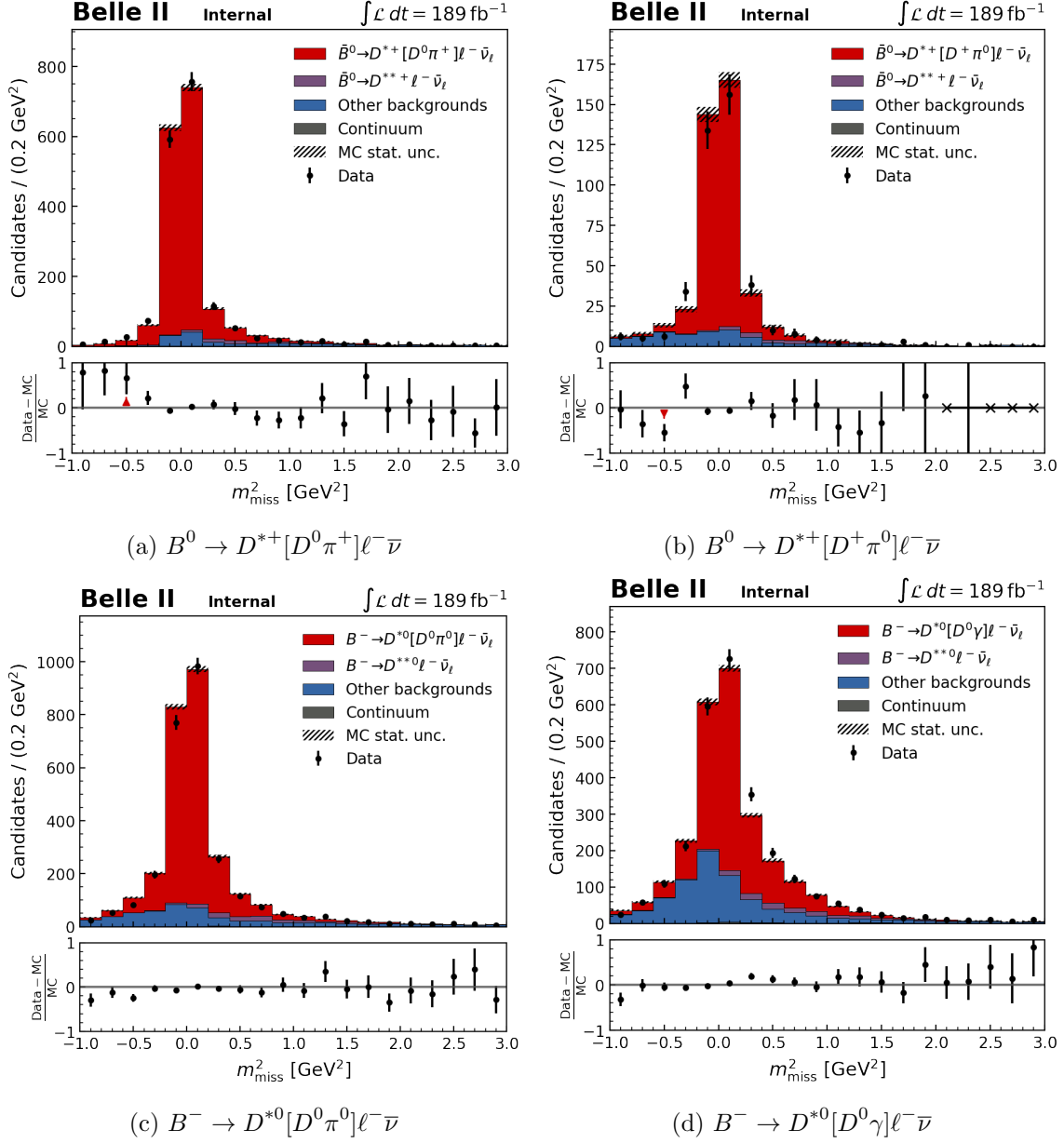


Figure 6.35.: The post-MVA classifier cut m_{miss}^2 for each of the $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ MVAs separated by each D^* candidate decay type. Generic MC is classed as the normalisation channels ($B \rightarrow D^* \ell^- \bar{\nu}_\ell$), $B \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ background, other backgrounds and continuum. Errors shown are statistical uncertainties.

6. Analysis Method of $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$

B channel	$N_{\text{MC}}^{\text{sig}}$	$N_{\text{MC}}^{\text{bkgd}}$	$N_{\text{MC}}^{\text{total}}$	N_{data}
$\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$	1881 ± 23	281 ± 23	2162 ± 23	2147 ± 46
$\bar{B}^0 \rightarrow D^{*+}e^- \bar{\nu}_e$	869 ± 23	162 ± 23	1031 ± 23	1045 ± 46
$\bar{B}^0 \rightarrow D^{*+}\mu^- \bar{\nu}_\mu$	1011 ± 23	118 ± 23	1129 ± 23	1102 ± 46
$\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$	1029 ± 20	740 ± 20	1769 ± 20	1748 ± 41
$\bar{B}^0 \rightarrow D^+e^- \bar{\nu}_e$	439 ± 20	341 ± 20	780 ± 20	836 ± 41
$\bar{B}^0 \rightarrow D^+\mu^- \bar{\nu}_\mu$	590 ± 20	399 ± 20	989 ± 20	912 ± 41
$B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$	3882 ± 36	1577 ± 36	5459 ± 36	5400 ± 73
$B^- \rightarrow D^{*0}e^- \bar{\nu}_e$	1786 ± 36	812 ± 36	2598 ± 36	2672 ± 73
$B^- \rightarrow D^{*0}\mu^- \bar{\nu}_\mu$	2096 ± 36	764 ± 36	2860 ± 36	2728 ± 73
$B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$	2821 ± 43	5053 ± 43	7874 ± 43	7731 ± 87
$B^- \rightarrow D^0e^- \bar{\nu}_e$	1237 ± 43	2375 ± 43	3612 ± 43	3786 ± 87
$B^- \rightarrow D^0\mu^- \bar{\nu}_\mu$	1583 ± 43	2677 ± 43	4260 ± 43	3945 ± 87

Table 6.12.: The expected MC signal yield ($N_{\text{MC}}^{\text{sig}}$), background yield (with continuum included as a background) ($N_{\text{MC}}^{\text{bkgd}}$) and sum of the two ($N_{\text{MC}}^{\text{total}}$), compared to the final measured yield in data (N_{data}) of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels. The yields of the separated electron and muon are included. Errors shown are individually calculated statistical uncertainties.

7. $R(D)$ and $R(D^*)$ denominator extraction

In this chapter the extraction process for the measurement of the denominators of $R(D)$ and $R(D^*)$ will be described. This includes the validation of the fitting procedure used to extract normalisation event yields from data, and the calculation of the signal efficiency of the analysis. The MVA is applied to the MC and data and the chosen MVA classifier cuts are applied.

7.1. Fitting procedure and software

To calculate the denominators of $R(D)$ and $R(D^*)$, one extracts the yield of the normalisation decay channel from data, N_s . This value is determined using a one-dimensional fit to the m_{miss}^2 distribution. This is achieved through a likelihood fit of generic MC histogram shapes referred to as templates. Each template approximates the probability density function (PDF) of a chosen physics process.

In this analysis, two templates are chosen:

1. **Signal template:** defined as the normalisation mode, $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$. **TopoAna** is used to select generator level, topologically truth matched $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ events. The lepton is required then to be correctly reconstructed using **basf2** truth matching.
2. **Background template:** all other decay channels are considered as background, including:
 - $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ events without a truth matched lepton.
 - Non-topologically matched $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ events.

7. $R(D)$ and $R(D^*)$ denominator extraction

- $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ events.
- $B \rightarrow D^{**} l^- \bar{\nu}_l$ events.
- $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ feed-up (feed-down) events in which the D (D^*) meson in the event is incorrectly reconstructed as a D^* (D) due to reconstruction with an extra (missing) γ or π . This overall effect of misreconstructed $D^{(*)}$ is called ‘cross-feed’.
- All other B meson events.
- Continuum events.

Thus, the background distributions (D^{**} background, fake $D^{(*)}$, other backgrounds and continuum) shown in Section 6.4.5 are combined into one template. This single MC component is an adequate representation of the various backgrounds due to the evaluation of shape variation and uncertainties in the modelling (later discussed in Chapter 8). In addition, the fit continues to be constrained to the side-band analysis of the m_{miss}^2 , in the region -1 GeV^2 to 3 GeV^2 . The templates are constructed with 20 bins such that the bin width is 0.2 GeV and the m_{miss}^2 signal peak is resolved. The number of bins was varied between 10 and 30 with a negligible effect on the signal and background yields in comparison to the uncertainties. As previously mentioned in Chapter 6, the templates are also kept to 20 bins for direct comparison with Reference [1].

7.1.1. Fitting with binfit

The fitting procedure uses the same software package and procedure as that of Reference [1]. The fit to data is performed through the maximisation of a binned Poissonian likelihood, defined as the product of the Poissonian distributions $\mathcal{P}(n_i; \nu_i)$ multiplied by the nuisance parameters (NP) $\mathcal{G}_k$ of each process k :

$$\mathcal{L} = \prod_i^{\text{bins}} \mathcal{P}(n_i; \nu_i) \times \prod_k^{\text{processes}} \mathcal{G}_k. \quad (7.1)$$

Here $\mathcal{P}(n_i; \nu_i) = e^{-\nu_i} \frac{\nu_i^{n_i}}{n_i!}$, n_i is the observed number of events in data for the i^{th} bin and ν_i is the expectation of the number of events in data, estimated from MC. Whenever a

7. $R(D)$ and $R(D^*)$ denominator extraction

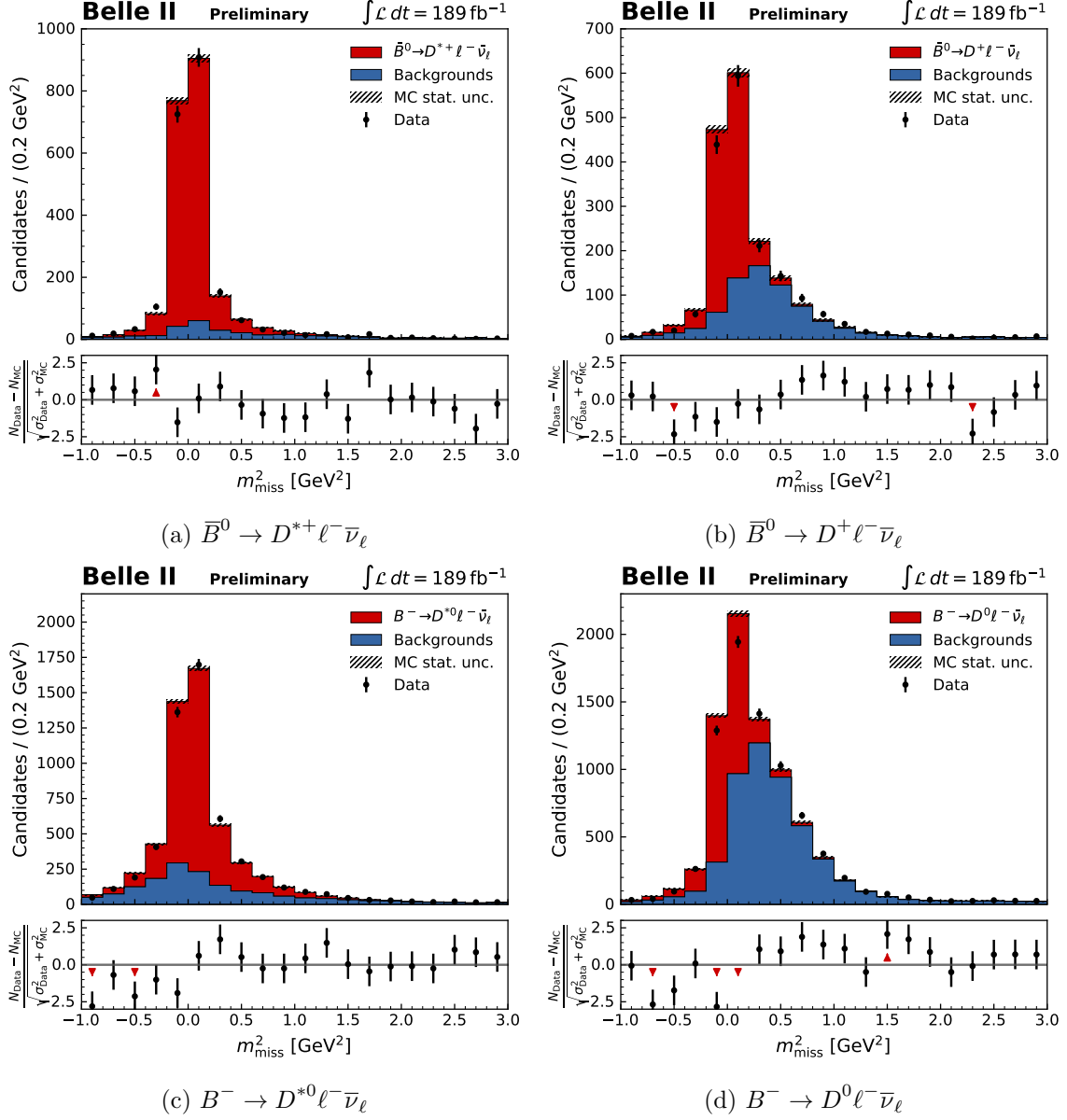


Figure 7.1.: The pre-fit m_{miss}^2 distribution for each of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ channels. Generic MC is classed as the normalisation channel $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ and background. Errors shown are statistical uncertainties. The pull shows the standard deviation of data from MC, with red arrows indicating a difference of more than 1σ .

variation such as a systematic uncertainty is added to the fit, the variation is added as a nuisance parameter.

The fit parameter results give a value and error for each bin, which corresponds to how much the fit differed from the template in a given bin. The error is the error on the pull of

7. $R(D)$ and $R(D^*)$ denominator extraction

the fit. The `binfit` calculates the global covariance matrix for the templates and nuisance parameters.

The templates that are input to the `binfit` package are those shown in Figure 7.1. These are the same histograms as Figure 6.32, with the generic MC redistributed into the chosen fit template definitions.

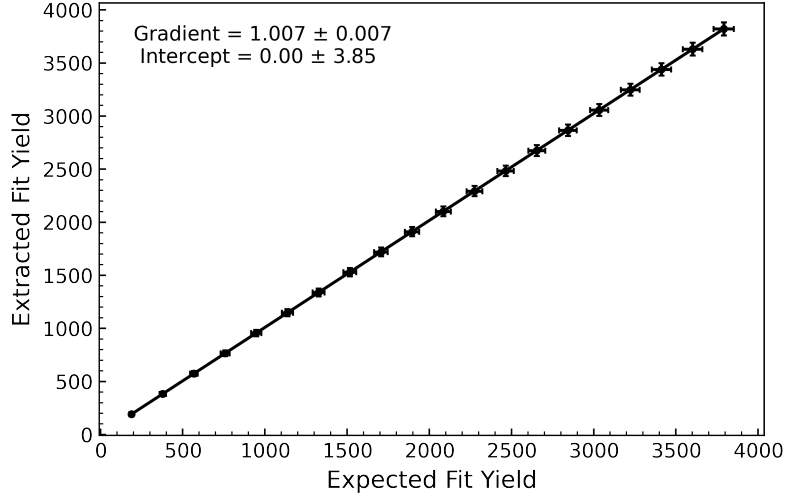
7.2. Fit validation

The fit is first validated before being applied to the normalisation and background templates. The stability of the implemented fitting procedure can be confirmed by the observation of no correlation between the values extracted from the fit and the values expected by the fit. This can be confirmed using a linearity test, in which a stable fit would display linear dependence of the fitted yield parameter on the injected expected yields. The injected expected yields (i.e. MC yields) correspond to the expected fit yields scaled with a factor ranging from 0.1 to 2.0. Figure 7.2a indicates that the fit is stable due to a gradient consistent with 1 and a y -intercept consistent with 0 (both within 1σ). The ratio of the injected extracted fit yields to the expected fit yields in Figure 7.2b confirms this conclusion; the gradient is consistent with 0 and the y -intercept is consistent with 1 (both within 1σ) indicating that there is no bias in the fit.

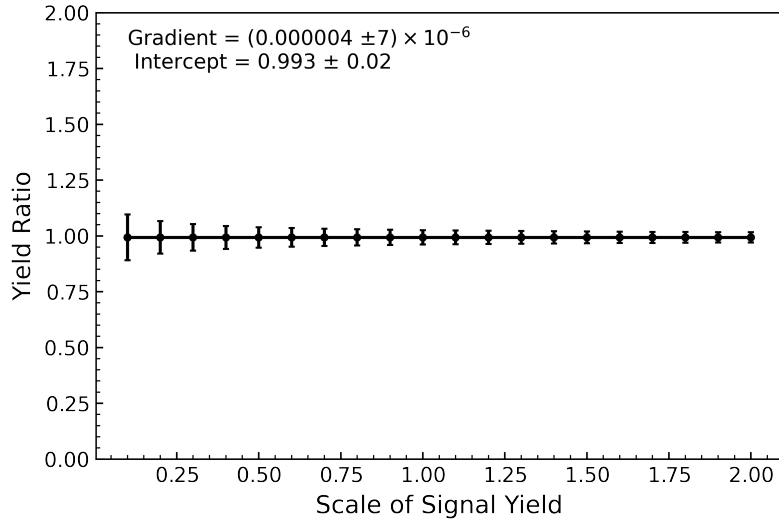
An Asimov dataset, where extractable quantities are set to be equal to their expected values^[i], is then used to test the validity of the fit. In the case of a successful fit, the Asimov data would correspond exactly to the MC data. Thus, the fit is shown to perform perfectly in Figure 7.3, with the Asimov data-MC distributions resulting in a pull distribution of 0 ± 1 (i.e perfectly agreeing). The same procedure was performed on the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$, $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ reconstructed decay channels to re-test the fit to account for fit variations. The fit procedure was found to be successful for all four decay channels.

^[i]”Asimov” refers to the author of *Franchise* in which a single most representative voter is used to predict the outcome of an election and therefore replaces the whole electorate [68].

7. $R(D)$ and $R(D^*)$ denominator extraction



(a) Linearity test of the fit to $B \rightarrow D^* \ell^- \bar{\nu}_\ell$.



(b) Ratio of expected yields to extracted yields for the fit to $B \rightarrow D^* \ell^- \bar{\nu}_\ell$.

Figure 7.2.: The results of the test of linearity of the fit on the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decay channel. Generic MC is classed as the normalisation channel $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ and background. The errors shown are statistical uncertainties.

7.3. Fit results

The fit is applied to each of the four decay channels and the results are presented in Figure 7.4. The final yields extracted from each of the fits are presented in Table 7.1.

7. $R(D)$ and $R(D^*)$ denominator extraction

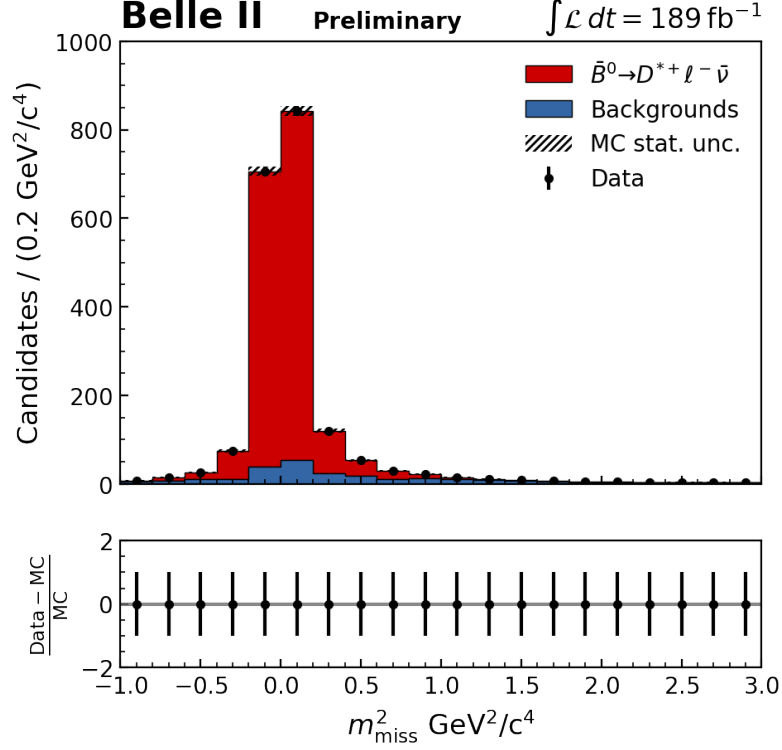


Figure 7.3.: The Asimov test of the fit model on the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decay channel. The errors shown are statistical uncertainties.

7.4. Branching fraction calculation

The denominator of $R(D)$ and $R(D^*)$ for each of the B decay channels are calculated using the equation:

$$\mathcal{B}(B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell) = \frac{N_s \times (1 + f_{+,0})}{4 \times N_{B\bar{B}} \times \epsilon_{\text{tag+sel}}} . \quad (7.2)$$

Here N_s is the normalisation event yield defined as $N_s = N_{\text{MC,fit}}^{\text{total}} - N_{\text{MC,fit}}^{\text{bkgd}}$, and $\epsilon_{\text{tag+sel}}$ is the signal efficiency defined in Section 7.4.1. In addition, $f_{+,0}^{[\text{ii}]}$ is the ratio of the branching

^[ii] Some literature refers to the quantity as $R^{+-/00}$.

7. $R(D)$ and $R(D^*)$ denominator extraction

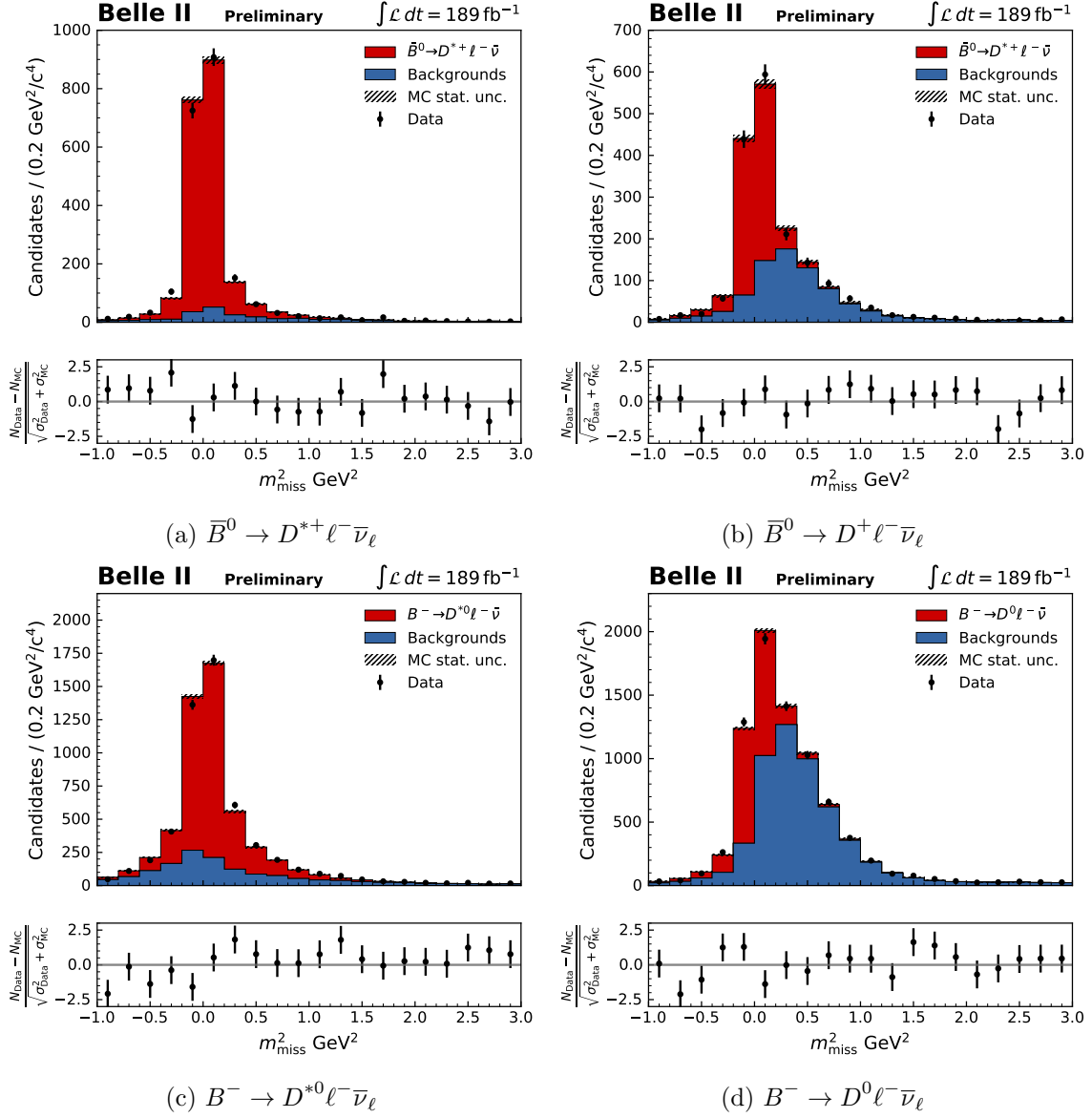


Figure 7.4.: The post-fit m_{miss}^2 distributions for each of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ channels. Generic MC is classed as the normalisation channel $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ and background. Errors shown are statistical uncertainties. The pull shows the standard deviation of data from MC.

fractions of the B^+ to B^0 productions from $\Upsilon(4S)$ decays (or vice versa):

$$f_+ = \frac{f_{+-}}{f_{00}} = \frac{\mathcal{B}(\Upsilon(4S) \rightarrow B^+ B^-)}{\mathcal{B}(\Upsilon(4S) \rightarrow B^0 \bar{B}^0)} = 1.058 \pm 0.024, \text{ or} \quad (7.3)$$

$$f_0 = \frac{f_{00}}{f_{+-}} = \frac{\mathcal{B}(\Upsilon(4S) \rightarrow B^0 \bar{B}^0)}{\mathcal{B}(\Upsilon(4S) \rightarrow B^+ B^-)} = 0.945 \pm 0.024, \quad (7.4)$$

7. $R(D)$ and $R(D^*)$ denominator extraction

B channel	$N_{\text{MC,fit}}^{\text{sig}}$	$N_{\text{MC,fit}}^{\text{bkgd}}$	$N_{\text{MC,fit}}^{\text{total}}$	N_{data}
$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$	1885 ± 21	242 ± 8	2126 ± 22	2196 ± 46
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	939 ± 15	788 ± 14	1728 ± 20	1672 ± 40
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	3945 ± 31	1431 ± 18	5376 ± 36	5389 ± 73
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	2350 ± 24	5359 ± 36	7709 ± 42	7503 ± 86

Table 7.1.: The fitted MC signal yield ($N_{\text{MC,fit}}^{\text{sig}}$), fitted background yield ($N_{\text{MC,fit}}^{\text{bkgd}}$) and total fit yield ($N_{\text{MC,fit}}^{\text{total}}$), compared to the final measured yield in data (N_{data}) of each of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay channels. Errors shown are statistical uncertainties.

quoted from Reference [5]. $N_{B\bar{B}} = (198.0 \pm 3.0) \times 10^6$ is the official Belle II counted $B\bar{B}$ meson pairs produced in the dataset of 189 fb^{-1} . The factor of 4 is a result of the two B mesons in each $\Upsilon(4S)$ decay and the two lepton generations considered in the normalisation modes.

7.4.1. Signal efficiency

The signal efficiency $\epsilon_{\text{tag+sel}}$ is a selection and tagging efficiency ratio. It is calculated as the ratio of correctly reconstructed events to the number of generated events using $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ signal MC (defined in Section 4.1). The correctly reconstructed events are defined as normalisation events identified by **TopoAna** with a **basf2** truth matched lepton.

Table 7.2 presents the signal efficiencies of each of the measured decay channels, calculated using the designated signal MC samples for this analysis.

Normalisation channel	Signal Efficiency
	$\epsilon_{\text{tag+sel}}$
$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$	$(1.908 \pm 0.003) \times 10^{-4}$
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	$(2.319 \pm 0.005) \times 10^{-4}$
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	$(2.867 \pm 0.004) \times 10^{-4}$
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	$(4.905 \pm 0.008) \times 10^{-4}$

Table 7.2.: The signal efficiency of each of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay channels calculated using signal MC. The errors shown are statistical uncertainties.

8. Systematic Uncertainties

Each of the variables and factors in the calculation of the denominators of $R(D)$ and $R(D^*)$ contribute sources of statistical and systematic uncertainties to the final result. From Equation 7.2, there are four variables: N_s , $\epsilon_{\text{tag+sel}}$, $N_{B\bar{B}}$ and $f_{+,0}$, each with varying sources of uncertainty. In this chapter, the systematic uncertainties of the denominators of $R(D)$ and $R(D^*)$ components are considered and evaluated. Each individual systematic uncertainty is then summarized at the end of the chapter in Table 8.5.

To evaluate systematic uncertainties that require a full re-running of the entire analysis such as the application of the systematic uncertainty of the global tracking scale factor (Section 6.2.2), only 200 fb^{-1} of generic MC is used. Though 800 fb^{-1} of MC is used for the main analysis, the impact of addressing systematics with a smaller MC dataset is negligible.

Firstly, the relative systematic uncertainty for the $f_{+,0}$ variable is calculated from the provided systematic uncertainty in Reference [5] to be 2.27%. Secondly, the relative systematic error of $N_{B\bar{B}}$ is 1.5% [69]. It is dominated by the systematic uncertainties of:

- Effects of data/MC discrepancy entering the efficiency of the hadronic event selection criteria for $B\bar{B}$ events, and the variation in non- $B\bar{B}$ efficiencies and cross sections with beam energy being accounted for (0.5%).
- Limited knowledge of luminosities of on-resonance and off-resonance samples (0.9%).
- The limited number of off-resonance data samples available (1.14%).

For the final two variables, N_s and $\epsilon_{\text{tag+sel}}$, there are multiple systematic contributions, which have to be evaluated individually. They both include systematic errors from the FEI tagging calibration, tracking efficiency, π_{slow} reconstruction, π^0 efficiency, assumed charm

8. Systematic Uncertainties

branching fractions, PID, FFs and MVA. One may evaluate the combined impact of these two variables on $R(D^{(*)})$, assuming they are uncorrelated, thus each systematic uncertainty here need only be accounted for once in the calculation for the denominators of $R(D)$ and $R(D^*)$.

Systematics can generally be considered in two categories: those that are flat factors that affect the scale but do not alter the shape of the fit templates, and those that do alter the shape of the fit templates. The non-flat factor contributions of PID and FFs are evaluated as systematic uncertainties of N_s by including them as nuisance parameters in the fit^[i]. To determine their systematic uncertainty contribution to the yield, each nuisance parameter is individually removed from the Asimov fit and the impact of their removal evaluated. In contrast, though the charm branching fraction systematic uncertainties are also non-flat factor contributions to the denominator measurements, they are not evaluated as nuisance parameters of the fit. This is discussed in Section 8.2.2.

The remaining systematic errors contributing to the calculation of the denominators of $R(D)$ and $R(D^*)$ are evaluated using signal MC in their contribution to $\epsilon_{\text{tag+sel}}$. Explicitly, these are the FEI calibration, tracking efficiency, π_{slow} reconstruction, π^0 efficiency and MVA systematic uncertainties.

8.1. $\epsilon_{\text{tag+sel}}$ systematic uncertainties

8.1.1. FEI Calibration

The FEI tagging algorithm and the necessary calibration for data-MC is the largest source of uncertainty in the measurement of the denominators of $R(D)$ and $R(D^*)$. The calibration was performed using a data-driven fit of the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ decay lepton momentum spectrum, as discussed in Chapter 6. The uncertainties are drawn from the fit model, the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ branching fraction uncertainties, the shape of the lepton spectrum, the lepton ID, and the composition of the inclusive semileptonic decays. The largest uncertainties are due to the fit model and the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ branching fractions. In this thesis the

^[i]The fit validation using an Asimov dataset is performed with all non-flat factor systematic uncertainty contributions.

8. Systematic Uncertainties

uncertainties are combined and are assumed to be uncorrelated with the other remaining uncertainties assigned for the efficiencies. The event overlap of the semileptonic D^* events in this analysis and those of the calibration analysis is negligible. Table 8.5 summarizes the relative systematic uncertainties of each of the calibration factors from Table 4.6.

8.1.2. Tracking Efficiency

The Belle II tracking group have provided a tracking systematic uncertainty of 0.30% per K , π and lepton track for the Moriond 2022 dataset (the dataset used in this analysis). The tracking systematics present in the B_{tag} side of the events are accounted for in the FEI calibration and associated systematic, thus only signal side track efficiencies are explicitly calculated. As the signal decay definition is constructed of multiple decay modes, which vary based on the B meson and subsequent $D^{(*)}$ decay mode, the average number of tracks used in the reconstructed of $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ is calculated using the values presented in Tables H.1 and H.2. The π_{slow} track is excluded from this calculation as the systematics associated with the π_{slow} must be considered separately (Section 8.1.3). The average number of tracks per event is calculated and the associated systematic uncertainty on the yields is presented in Table 8.1.

Normalisation channel	Avg. number of signal-side tracks	Relative systematic uncertainty (%)
$\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$	3.33	1.00
$\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$	3.40	1.02
$B^- \rightarrow D^{*0}\ell^-\bar{\nu}_\ell$	3.27	0.98
$B^- \rightarrow D^0\ell^-\bar{\nu}_\ell$	3.29	0.99

Table 8.1.: The average number of tracks per event for each of the four $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ normalisation decay modes and the associated systematic uncertainty.

In addition, the systematic uncertainty on the applied tracking scale factor discussed in Section 6.2.2 is evaluated via variation of the scale factor to $[0.99916, 1.00006]$. These variations result in the yield systematic uncertainties displayed in Table 8.2. These are combined in quadrature with the tracking efficiency systematic uncertainties and are presented in Table 8.5 as the tracking systematic uncertainty.

8. Systematic Uncertainties

Normalisation channel	Relative systematic uncertainty (%)
$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$	0.01
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	0.01
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	0.02
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	0.01

Table 8.2.: The systematic uncertainties as a result of the tracking scale factor for each of the four $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ normalisation decay modes.

8.1.3. Slow π data/MC efficiency

The slow π_{slow}^0 and π_{slow}^+ efficiency corrections and corresponding systematic uncertainties are provided by the Belle II performance group [?]. The systematic uncertainty due to the π_{slow} was evaluated by varying the momentum by $\pm 1\sigma$, and corresponds to 4.1% (4.4%) for B^0 (B^+) of $\epsilon_{\text{tag+sel}}$.

8.1.4. π^0 efficiency

The π^0 efficiency corrections and systematic uncertainties are provided by the Belle II performance group [?]. The systematic error for the π^0 40% efficiency list stated in Section 6.2.3 shows a relative systematic error of 4.34%. Similarly, the systematic error for the π^0 50% efficiency list stated in Section 6.2.4 is 7.51%. This systematic error is evaluated considering $D^{(*)}$ decay channels that contain a π^0 . The overall systematic uncertainty on the yields corresponds to the values presented in Table 8.5.

8.1.5. MVA

To evaluate the systematic uncertainties implicated by the MVA, one can examine the variation of the data-MC agreement as a function of potential classifier cut. Using this variation, one can implement a measure of mis-modelling by then calculating the signal efficiency and the $R(D)$ and $R(D^*)$ denominators, also as a function of classifier cut. Figure 8.1 (and Figure H.1) displays the post-fit variations for $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ [ii]. These lead to a $R(D^*)$ denominator variation depicted in Figure 8.2 (and Figure H.2). The MVA

[ii] Here the classifier cut at a value of 0.9 fails due to low statistics in certain bins leading to a non-accurate covariance.

8. Systematic Uncertainties

contribution to the systematic uncertainties of the denominators of $R(D^{(*)})$ are presented in Table 8.5.

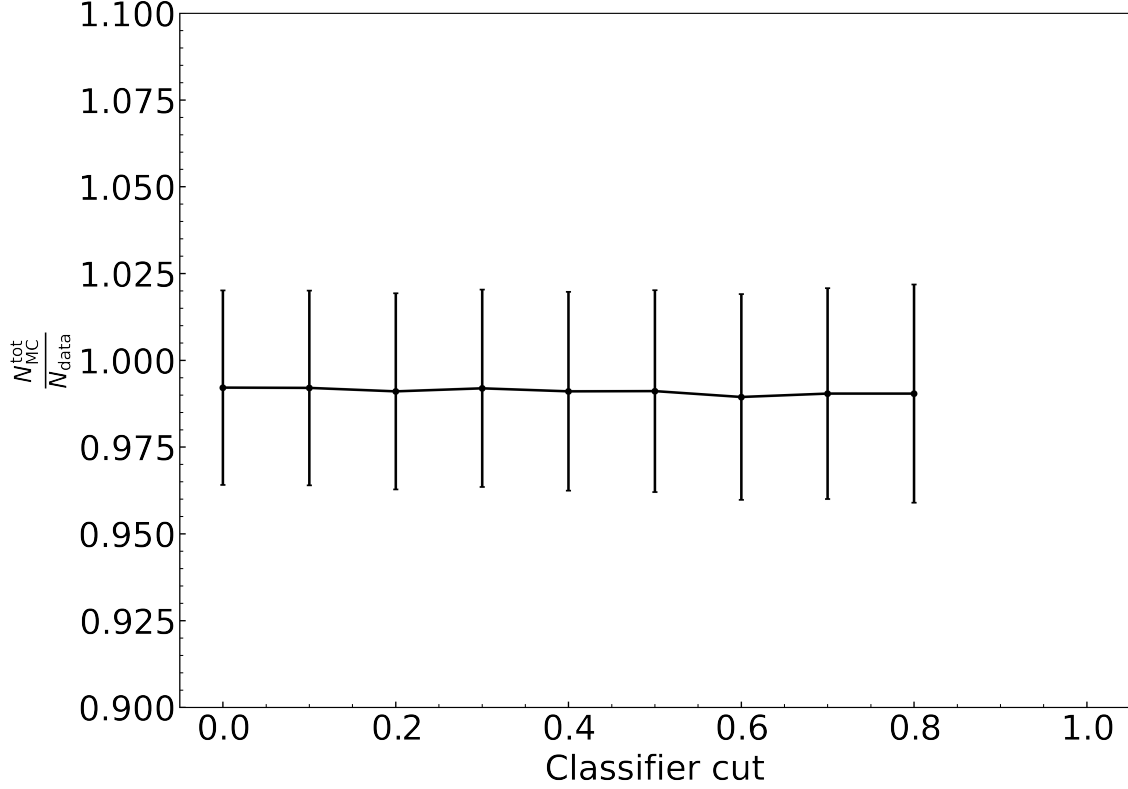


Figure 8.1.: The ratio of MC to data as a function of classifier cut for $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$.

8.2. N_s systematic uncertainties

8.2.1. The fit

The systematic uncertainty introduced by the fit due to MC statistics is estimated through performing the fit on statistically varied subsets of the original dataset using the bootstrap method. A statistically varied subset of events are sampled from the complete dataset. In this it is possible to draw the same event multiple times in a single sub-sample. This procedure is performed 500 times and the standard deviation from the nominal value is taken as the systematic uncertainty. Table 8.3 shows the yield standard deviation in the fit, and the relative systematic uncertainty of the yield.

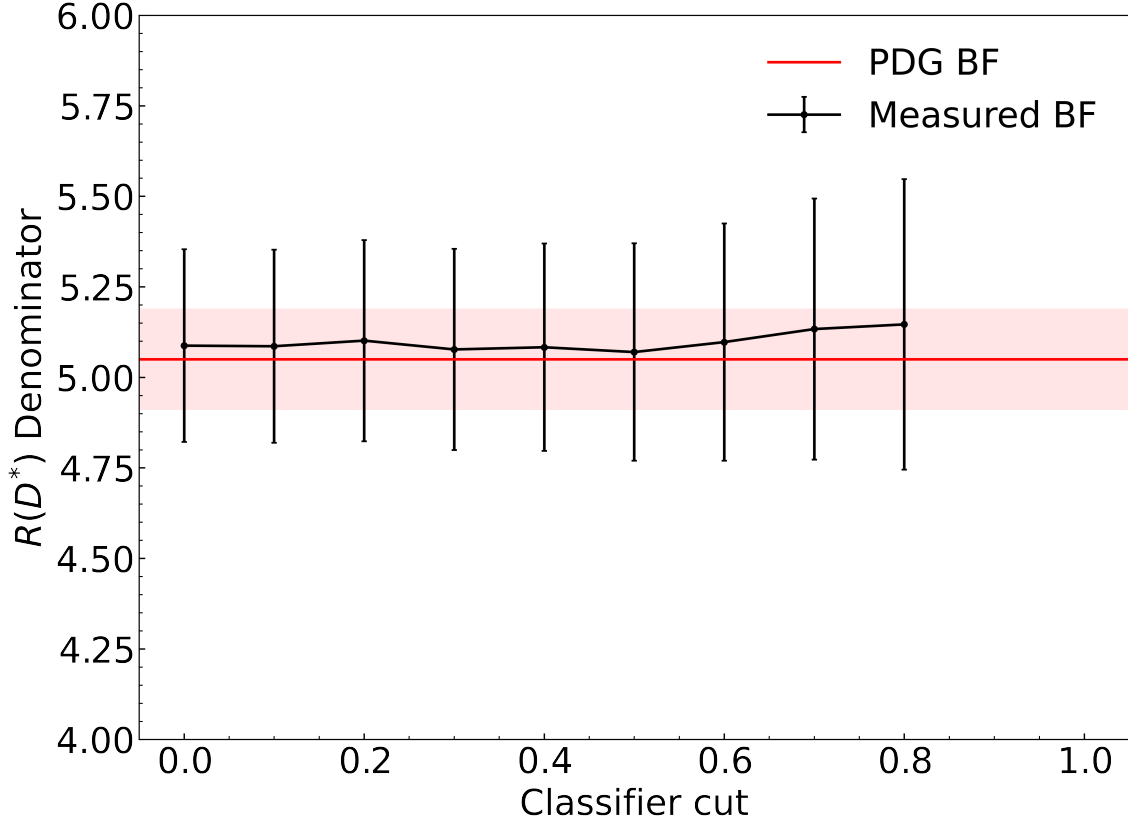


Figure 8.2.: The denominator of $R(D^*)$ as a function of classifier cut for $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$.

Furthermore the D^{**} branching fraction, PID and FF systematic uncertainties are each evaluated via nuisance parameters within the fit. The impact of their variations on the yields are determined individually by comparing the uncertainty with fixed fit parameters to that of unfixed parameters.

The correlations between the different nuisance parameters are difficult to disentangle in a mathematically sound manner. Thus, it is assumed that the systematic uncertainty due to correlations corresponds to the difference between the sum of the systematic uncertainties of the individual nuisance parameter contributions and the overall systematic uncertainty all of the nuisance parameter contributions. This ultimately corresponds to a negligible systematic uncertainty.

8. Systematic Uncertainties

Normalisation channel	$N_{\text{MC,fit}}^{\text{total}}$	Relative Sys. Uncertainty (%)
$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$	$2126 \pm 23 \pm 16$	0.76
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	$1728 \pm 20 \pm 13$	0.79
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	$5376 \pm 36 \pm 37$	0.69
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	$7709 \pm 42 \pm 28$	1.19

Table 8.3.: The total fit yield $N_{\text{MC,fit}}^{\text{total}}$ with statistical and systematic uncertainties included.

8.2.2. Branching fraction uncertainties of charm decays

For the relatively well known D^* branching fraction uncertainties, the systematic uncertainties can be propagated through the branching fraction calculation. The $D^{(*)}$ branching fractions and their systematic uncertainties are obtained from Reference [5] and are displayed in Table 5.1. The relative systematic uncertainties for each of the decay modes calculated for the branching fractions are shown in Table 8.4. These yield total relative uncertainties of 1.85% (0.27%) for the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ ($B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$) channel, and 1.15% (2.24%) for the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ ($B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$) channel.

To account for the more experimentally limited measurements and systematic uncertainties of the D^{**} branching fraction, D^{**} MC is re-weighted with their branching fractions varied by $\Delta \mathcal{B}(B \rightarrow D^{**} \ell \bar{\nu}_\ell) / \mathcal{B}(B \rightarrow D^{**} \ell \bar{\nu}_\ell) = 26.2\%(15.8\%)$, for B^0 (B^+) in accordance with the world average uncertainties [5]. This variation is then propagated through to the fit of the yields and branching fraction calculation. The variation in shape of the D^{**} modelling is incorporated into the large branching fraction uncertainties. The systematic uncertainties due to the $D^{(*)}$ and D^{**} branching fractions for the denominators of $R(D)$ and $R(D^*)$ are presented in Table 8.5.

8.2.3. Particle ID and Fake Rates

The particle ID group provides correction factors for MC particles with different signal identification efficiency targets, with their statistical and systematic uncertainties. Hadron ID and lepton ID uncertainties are considered separately and can be processed using the developed particle identification (PID) package, `PIDvar`. Belle II produces ensembles of

8. Systematic Uncertainties

Decay channel	Relative systematic uncertainty (%)
$D^*(2007)^0 \rightarrow D^0 \pi^0$	1.39
$D^*(2007)^0 \rightarrow D^0 \gamma$	2.55
$D^*(2010)^+ \rightarrow D^0 \pi^+$	0.74
$D^*(2010)^+ \rightarrow D^+ \pi^0$	1.63
$D^0 \rightarrow K^- \pi^+$	0.78
$D^0 \rightarrow K_S^0 \pi^0$	1.77
$D^0 \rightarrow K^- \pi^+ \pi^0$	3.47
$D^0 \rightarrow K^- \pi^+ \pi^- \pi^+$	1.70
$D^0 \rightarrow K_S^0 \pi^+ \pi^-$	6.42
$D^0 \rightarrow K_S^0 \pi^+ \pi^- \pi^0$	11.54
$D^0 \rightarrow K^- K^+$	1.47
$D^0 \rightarrow \pi^- \pi^+$	1.65
$D^+ \rightarrow K^- \pi^+ \pi^+$	1.71
$D^+ \rightarrow K^- \pi^+ \pi^+ \pi^0$	2.88
$D^+ \rightarrow K_S^0 \pi^+$	1.98
$D^+ \rightarrow K_S^0 \pi^+ \pi^- \pi^+$	2.90
$D^+ \rightarrow K_S^0 \pi^+ \pi^0$	2.85
$D^+ \rightarrow K_S^0 K^+$	2.96
$D^+ \rightarrow K^- K^+ \pi^+$	1.86

Table 8.4.: The relative systematic errors of the branching fractions on the normalisation for $D^{(*)}$ meson decays. Values are calculated from Table 5.1 [5].

weights with variations of the statistical and systematic uncertainties of a nominal particle ID table parametrised in several momentum, charge and theta angle bins. Each particle correction factor in each bin has uncorrelated statistical uncertainty and is assumed to have fully independent sources of systematic uncertainty. Within the ensemble, the total uncertainty is then estimated on the selection efficiency using the root-mean-squared of the recovered efficiencies.

Finally, the variations in the weights are evaluated as nuisance parameters in the fitting process. The systematic uncertainties of the PID for the denominators of $R(D)$ and $R(D^*)$ are presented in Table 8.5.

8.2.4. Form Factors (FFs)

The FF uncertainties will predominantly cancel out in the ratios of $R(D)$ and $R(D^*)$ that will be calculated in future analyses. The FFs that are not cancelled out are the FFs

8. Systematic Uncertainties

present in the differential decay rates of $B \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$. FF theoretical uncertainty is also substantially smaller than experimental constraints. Even so, they require comprehensive understanding in a measurement that aims to compare the two, and in the measurement of the denominators of $R(D)$ and $R(D^*)$ in which they do not cancel. This section describes the method of evaluating the FF systematic uncertainties.

In this analysis, the FF parametrization is set by the BGL(1,1,2) model [70] used to simulate the $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays in MC. To evaluate the systematic uncertainties of this model, one must define a new basis of parameters with uncorrelated uncertainties to mitigate experimental correlations between FF parameters. Following this, the new basis of parameters can be combined to form a covariance matrix that describes the correlations between the central values of the FF parameters. The FF parameters can then be varied by $\pm 1\sigma$ in the new basis.

Event-wise correction weights are then calculated using the varied parameters in the normalised differential decay rates:

$$w(\vec{v}) = \frac{\Gamma_{\text{nom}}}{\Gamma_{\text{var}}} \frac{d\Gamma(\vec{v})_{\text{var}}/dv}{d\Gamma(\vec{v})_{\text{nom}}/dv} \quad (8.1)$$

where w is the event weight that is a function of (various) kinematic variables ($\vec{v}$). “nom” refers to the original (differential) decay rate, whereas “var” indicates the varied (differential) decay rate in which the BGL central FF parameters are varied in the new basis by $\pm 1\sigma$. The weights were found to be non-negligible and therefore used to evaluate each of the FF parameter systematic uncertainties in the fitting process. The systematic uncertainties due to the FFs for the denominators of $R(D)$ and $R(D^*)$ are presented in Table 8.5.

8.3. Total relative systematic uncertainties

The summary of relative systematic uncertainties for the measurements of the denominators of $R(D)$ and $R(D^*)$ are shown in Table 8.5.

8. Systematic Uncertainties

Source	$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ (%)	$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ (%)	$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ (%)	$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ (%)
$f_{+,0}$	2.27	2.27	2.27	2.27
$N_{B\bar{B}}$	1.50	1.50	1.50	1.50
FEI calibration (e)	6.00	3.24	8.00	3.10
FEI calibration (μ)	6.00	3.88	6.00	3.71
Tracking	1.00	1.02	0.98	0.99
π_{slow} efficiency	4.10	-	4.40	-
π^0 efficiency	4.08	2.15	4.96	2.31
MVA	4.29	2.80	3.78	8.92
Fit	0.79	0.79	0.79	0.79
D^{**} branching fraction	0.50	0.58	1.58	1.44
$D^{(*)}$ branching fraction	1.85	1.15	0.27	2.24
PID	0.83	0.59	3.14	1.08
FFs	0.51	0.63	1.82	0.61
TOTAL	10.07	6.58	11.48	10.45

Table 8.5.: A summary of the relative systematic uncertainties on the measurements of the denominators of $R(D)$ and $R(D^*)$, separated into the four relevant normalisation decay channels.

9. Results and discussion

In this chapter, the measured branching fraction is presented for each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels. Equation 7.2 is used for the calculation. Table 9.1 presents the final measured branching fractions of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels.

Normalisation channel	Measured branching fraction (%)	PDG branching fraction (%)
$\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$	5.13 ± 0.06 (stat.) ± 0.52 (sys.)	5.05 ± 0.14
$\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$	2.11 ± 0.02 (stat.) ± 0.14 (sys.)	2.31 ± 0.10
$B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$	6.76 ± 0.03 (stat.) ± 0.78 (sys.)	5.66 ± 0.22
$B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$	2.35 ± 0.01 (stat.) ± 0.25 (sys.)	2.35 ± 0.09

Table 9.1.: The measured branching fraction of each of the $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ decay channels. The current world average branching fractions are given for comparison [5].

The measurements of the branching fractions for $\bar{B}^0 \rightarrow D^{*+}\ell^- \bar{\nu}_\ell$, $\bar{B}^0 \rightarrow D^+\ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0\ell^- \bar{\nu}_\ell$ as denominators of $R(D^{(*)})$ are in agreement with the world average experimental measurements of each of the individual $B \rightarrow D^{(*)}\ell^- \bar{\nu}_\ell$ branching fractions within 1σ of uncertainties. The measurement of $B^- \rightarrow D^{*0}\ell^- \bar{\nu}_\ell$ is approximately 2σ from the world average measurement.

Thus, these four results are compatible with the previous experimental measurements of these decay channels. It can be concluded that the methodology used to measure the denominators of $R(D)$ and $R(D^*)$ are in good shape, which bodes well for measurements of $R(D)$ and $R(D^*)$ with this method in the near future.

9.1. The “golden mode”

This section contains a short comparison of the work in this thesis to the work on the ‘golden mode’ in Reference [1], work that the author previously contributed to.

Notably, there are substantial differences between the previous and current analysis methods that alter the signal efficiency, the yields, and therefore the measured branching fraction. The collaboration-wide changes to MC campaign and FEI calibration are unavoidable. With these updates, performance and corrections such as PID are altered, resulting in raw analysis differences. Additionally, the signal selection at the analysis level has been more fine tuned since the publication of Reference [1]. Thus, it is expected that there will be differences between the two analyses that do not allow direct juxtaposition of results.

Nonetheless, an exploration into whether the current analysis improves the “golden mode” measurement by reducing statistical and systematic uncertainties has scientific merit. Consequently, the golden mode measurement is compared to the result of $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ made in this analysis.

The $N_{B\bar{B}}$ yield in data in this analysis is 1% lower than expected if scaled from the previous result to 189 fb^{-1} . This is not necessarily a surprising variation in yield as there is the possibility of fluctuation in effective $B\bar{B}$ production cross-section due to the Belle II beam energy and its spread.

The signal efficiency in the current analysis is measured to be an order of magnitude higher than that of the previous study. This is due to the use of more exclusive decay modes, and improved MC modelling, signal selection and MVA application.

The statistical uncertainty reduces by a factor of 70. This, combined with the reduction in systematic uncertainty from 12% to 10% indicates a significant enhancement of the previous result. The largest reduction in individual systematic uncertainty comes from the reduction in π_{slow} and tracking systematic uncertainties. The comparatively large reduction in statistical uncertainty compared to that of the systematic uncertainty suggests that systematic uncertainties are, and will continue to be, the limiting factor of this measurement in the future. The systematic uncertainties are currently much larger than that of the world

9. Results and discussion

average systematics uncertainties. This means that these results are unlikely to improve the world average much in the future unless these uncertainties can be dramatically reduced. For the purposes of $R(D)$ and $R(D^*)$ measurements, however, this is not problematic because, as previously mentioned, a significant amount of the branching fraction systematic uncertainties will cancel in the ratio.

The branching fraction of the current analysis moves towards the current world average of the branching fraction measurement of $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$. Although there is a mild ($< 2\sigma$) tension in the $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ branching fraction measurement with respect to the world average value, there is no indication of any issue in the denominator branching fractions of $R(D)$ and $R(D^*)$.

In addition to the aforementioned differences in analysis methods, the application of the D^{**} background suppression MVA is arguably the largest difference between the two analyses at the analyst level; so while a direct comparison between the two analyses is nigh impossible, the reduction of systematic uncertainty contribution is justified through the benefits of the D^{**} background suppression.

9.2. The D^{**} suppression

A main facet of this thesis is to suppress D^{**} backgrounds to $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays through the development of an MVA. This section discusses the level of success of the application of the developed MVA in the process of measuring $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ with 189 fb^{-1} of Belle II data.

Firstly, the use of the FastBDT MVA algorithm allowed for reduced computation time in training on the boosted MC sample, but equivalent performance to other machine learning techniques (Figure 6.25). This will be beneficial in future applications and trainings, as with scaling of the data collected at Belle II, the signal MC sample sizes will have to grow significantly too.

Secondly, the training of the MVA using the boosted MC was successful in that it resulted in a lower likelihood of over-training of the samples. The ROC curves in Figures 6.28 and E.3 indicate good signal efficiency and background rejection, as well as agreement between

9. Results and discussion

the training and testing samples.

Thirdly, Figure 6.26 shows that there is good signal-background separation for the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ MVA. However, there is reduced separation for the other three MVAs (Figure E.1). There are a multitude of future paths that can be taken to improve the signal-background separation. These shall be discussed in Section 9.3.

Finally, with the application of an MVA classifier output cut, chosen with consideration of a multitude of factors as discussed in Chapter 6, normalisation efficiencies were kept high without a substantial loss in purity, or loss in signal events. The application of the MVA was effective, with a minimum effect of a reduction of 40% of D^{**} backgrounds, and the maximum reduction being in the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ MVA of $\sim 90\%$. Furthermore, cross-feed contributions and other backgrounds including continuum were significantly reduced (Table 6.11). In addition to the drop in D^{**} yields, the direct effect of this suppression can be seen in the reduction of D^{**} systematic uncertainty contribution to the branching fraction measurement. For example, examining the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decay in the case of the MVA not being applied to the dataset, the branching fraction has a 0.70% systematic contribution from the D^{**} decays. In the case that the D^{**} suppression MVA is applied, the systematic contribution is reduced to 0.50%. The largest effect can be seen for the $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ channel, in which the systematic uncertainty drops by a factor of ~ 5 . This indicates the full power of the MVA before highly optimized signal selections are applied. The relative systematic uncertainties for the other decay channels are presented in Table 9.2. The desired effect of the MVA is that the systematic uncertainty contribution of D^{**} peaking backgrounds is reduced by a minimum of 20% through suppression of these backgrounds.

9.3. The future of this analysis

A future contribution with the highest potential impact to this analysis would be the fine tuning of the signal selections before the MVA is applied. The development of the “golden mode” ($B^0 \rightarrow D^{*+} [D^0 \pi^+] \ell^- \bar{\nu}$ where $D^0 \rightarrow K^- \pi^+$) analysis resulted in the fine tuning of the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decay channel signal selection, with the further D^* and D

9. Results and discussion

Normalisation channel	Systematic uncertainty without MVA (%)	Systematic uncertainty with MVA (%)
$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$	0.70	0.50
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	1.6	0.58
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	7.26	1.58
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	1.85	1.44

Table 9.2.: The comparison of systematic errors for the D^{**} of the selected $D^{(*)}$ meson decays with and without application of the D^{**} MVA.

daughter decay channels benefiting from this in natural succession. As the analysis grew into including the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ decay channels, more person-power behind these signal selection development was needed.

The effect of this is visible in these decays, where larger continuum and background contributions are present before application of an MVA (Figure 6.15). It is also apparent in the branching fraction results, in which the systematic uncertainties are nearly double for that of the B^+ decay channels. This combined with the measurement of $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ at 2σ from the worldwide average, presents a strong argument for further optimization of the signal selection and MVA. In addition, the data-MC ratios in Figure E.4 for the $B \rightarrow D \ell^- \bar{\nu}_\ell$ decays show a worse data-MC agreement which varies significantly with MVA classifier cut. This indicates that more studies need to be done for a better understanding of the MC simulation of real-world data.

Improving signal selection criteria in the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$, $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decays would also remove further non- D^{**} backgrounds that are mis-identified by the MVA as D^{**} decays and thus improve the performance of the overall analysis. The $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ signal selection improvement is currently underway at Belle II, and the $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ signal selection is an encouraged future endeavour for this analysis.

D^* cross-feed present in the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decay channels is a significant background that affects the MVA performance. Further studies before the application of MVA were performed by the author to reduce these backgrounds and a possibility for future development is presented in Appendix F.

There is room for further investigation into the MVA training variables used, with a

9. Results and discussion

possibility of inclusion of additional variables that could have discriminating power. One suggestion would be the examination of the use of angular variables in the MVA. For example in the frame of the $\tau - \nu$, i.e. the $B_{\text{sig}} - D^*$ frame, one can look at the angle between D^* descendant and the lepton, to investigate if there is any discriminating power.

A key aspect of this thesis is the use of the hadronic FEI algorithm. The FEI is retrained and recalibrated for each MC campaign, and aligning dataset. There are some improvements that can be made to the FEI algorithm to improve its performance. For example, the best candidate selection implemented within the FEI does not always select the purest decay, resulting in higher backgrounds. With the application of FEI to MC and data comes a significant discrepancy. However, with each iteration of MC campaign and FEI training, the procedure for the extraction of calibration factors that mitigate this discrepancy are improved. There are nevertheless inconsistencies with the use of the current calibration factors that must be taken into consideration for this analysis and improved for future analyses. These are the lepton identification (LID) corrections, the calibration methods and the incorrect branching fractions that exist in MC generation.

Firstly, the LID corrections for electrons and muons in MC14ri have limitations that affect the systematic uncertainties of the $R(D)$ and $R(D^*)$ denominators presented in Chapter 8. Currently, the FEI correction factors for the μ and e differ, as represented graphically in Figure 4.4. The effect of the PID and its corrections on the FEI data-MC discrepancy must be investigated, validated and understood further.

Secondly, the calibration method specifically for the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ decays is still in development. There is the question of the validity of using these correction factors derived using $B \rightarrow X\ell^-\bar{\nu}_\ell$ for the exclusive $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ decays. The largest concern is that of purity issues due to the inclusive nature of $B \rightarrow X\ell^-\bar{\nu}_\ell$. There have been recent efforts towards extracting the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ calibration factors from the decay itself, as has been implemented for $B \rightarrow D^*\ell^-\bar{\nu}_\ell$ in this analysis. Further review should be made on the exclusive calibration factors to improve on the systematic uncertainties presented in this analysis.

Finally, the calibration factors also account for, amongst other influencing factors, incorrect branching fractions that exist in MC generation. Alongside this however, there is

9. Results and discussion

a better data-MC agreement for decay channels with low multiplicity and a large disagreement in channels containing $D^{(*)}$ mesons. The data-MC difference is not only present in total branching fractions, but also present in intermediate resonances, as reported by the Belle II FEI task force. This indicates that better MC modelling is necessary for optimization and that intermediate resonances need to be tuned; scaling alone is not sufficient to account for the branching fraction differences. A global correction factor for FEI results in MC and data appearing to have different kinematics, which in turn indicates the need for non-global calibration factors.

These are all essential avenues of further exploration for improvement of this analysis. Full understanding of the FEI, its performance and optimization of its calibration method are required. This is currently in progress at the Belle II collaboration.

The reduction of uncertainties is another crucial aspect of this analysis that can be improved. As more data is collected, this result and its statistical errors will improve. Further, the systematic uncertainties will be improved over time with more understanding of detector and software performance. Moreover, systematic uncertainties of the charm branching ratios can be reduced. In particular, that of the $B \rightarrow D^{**}l^-\bar{\nu}_l$ decay channel, which contributes significant uncertainty to the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ branching fraction measurements, and the focus of this thesis, could still be reduced further. More precise measurements are needed from the physics community of these decay channels to improve this systematic uncertainty. Even with more precise measurements in the future, however, the charm branching fraction systematic is not separable from the total systematic uncertainty of the $B \rightarrow D^{(*)}l^-\bar{\nu}_l$ branching fraction, so cannot simply be updated as the measurements of these decay modes improve. To evaluate the systematic contributions of future charm branching fractions updates, one would have to measure the branching fractions as a function of the hadronic tag modes.

The improvement of the analysis of the denominators of $R(D)$ and $R(D^*)$ is an essential task in the future endeavour of measuring $R(D)$ and $R(D^*)$ themselves. With just 5 ab^{-1} of data, Belle II will be able to measure $R(D)$ and $R(D^*)$ with sufficient experimental resolution to determine the validity of the current discrepancy between theory and experiment through, for example, a 5σ measurement of $R(D^{(*)})$ that will point to physics BSM

9. *Results and discussion*

given the current central value.

10. Conclusion

The measurements of the $R(D)$ and $R(D^*)$ observables could indicate physics beyond the SM through current discrepancy of theory and experimental measurements. Further study, precision measurements and understanding of this discrepancy is of paramount importance in the quest for complete understanding of the universe, its particles and their interactions.

This work was an optimization of the methods used to measure $R(D)$ and $R(D^*)$. The denominators of $R(D)$ and $R(D^*)$ were measured using 189 fb^{-1} of Belle II data. The measurements were performed using the hadronic FEI algorithm, and with an extended list of exclusive decay channels. The data and MC in the side-band regions of the m_{miss}^2 of the $B \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ decay (-1 GeV to 3 GeV) were used. A FastBDT MVA was developed to reduce the largest background yield systematic uncertainty contributions, the D^{**} peaking backgrounds. This was performed successfully with a reduction in systematic uncertainty of a minimum of 20%. The yields of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay channels were extracted from a 1-D fit to the m_{miss}^2 distribution.

The systematic uncertainty contributions towards the measurement of the denominators of $R(D)$ and $R(D^*)$ were assessed with consideration for shape variation in MC modelling. The denominators of $R(D)$ and $R(D^*)$ were calculated with the evaluated signal yields and signal efficiencies and were found to be in agreement with the current worldwide average experimental results for the $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$, $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decay channels. The $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ decay channel was within 2σ of the current average experimental result. The final results are summarized in Table 10.1. This work was an extension of the study in Reference [1], with the result of $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell) = (4.51 \pm 0.41_{\text{stat}} \pm 0.27_{\text{syst}} \pm 0.45_{\pi_{\text{slow}}})\%$.

These measurements indicate possible consistency with the measured experimental dis-

10. Conclusion

Normalisation channel	Measured branching fraction (%)
$\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$	$5.13 \pm 0.06 \text{ (stat.)} \pm 0.52 \text{ (sys.)}$
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	$2.11 \pm 0.02 \text{ (stat.)} \pm 0.14 \text{ (sys.)}$
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	$6.76 \pm 0.03 \text{ (stat.)} \pm 0.78 \text{ (sys.)}$
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	$2.35 \pm 0.01 \text{ (stat.)} \pm 0.25 \text{ (sys.)}$

Table 10.1.: The measured branching fraction of each of the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decay channels.

crepancy of $R(D)$ and $R(D^*)$ seen in Figure 2.8, and excites further measurements of $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ and $R(D^{(*)})$ in the future. This is the first measurement of its kind at the Belle II Experiment. This work will contribute towards future measurements of $R(D)$ and $R(D^*)$.

A. The CKM parametrisation

This appendix is supplementary information for Section 2.1.1.

To conserve probability the CKM matrix must be unitary under normalisation. This then becomes an essential tool for verifying the SM as the unitarity requirement gives two conditions:

$$\sum_i V_{ij} V_{ji}^* = \delta_{jk} \quad (\text{A.1})$$

and

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0. \quad (\text{A.2})$$

Taking the conditions from Equation A.2 and dividing by $|V_{cd}^* V_{cb}|$ gives

$$\frac{V_{ud} V_{ub}^*}{|V_{cd}^* V_{cb}|} - 1 + \frac{V_{td} V_{tb}^*}{|V_{cd}^* V_{cb}|} = 0. \quad (\text{A.3})$$

The CKM matrix can be parametrized by 3 mixing angles and 1 CP violating CKM phase:

$$V_{\text{CKM}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (\text{A.4})$$

$$= \begin{pmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta} \\ -s_{12} c_{23} - c_{12} s_{23} s_{13} e^{i\delta} & c_{12} c_{23} - s_{12} s_{23} s_{13} e^{i\delta} & s_{23} c_{13} \\ s_{12} s_{23} - c_{12} c_{23} s_{13} e^{i\delta} & -c_{12} s_{23} - s_{12} c_{23} s_{13} e^{i\delta} & c_{23} c_{13} \end{pmatrix} \quad (\text{A.5})$$

where $s_{ij} = \sin \theta_{ij}$, $c_{ij} = \cos \theta_{ij}$ and δ is the CP -violating phase for all SM flavour-changing processes. This can be re-written in the Wolfenstein parametrization such that the CKM

A. The CKM parametrisation

matrix is unitary to all orders of $\lambda = s_{12}$ and where $\bar{\rho} = \rho(1 - \lambda^2/2 + \dots)$ and $\bar{\eta} = \eta(1 - \lambda^2/2 + \dots)$:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \quad (\text{A.6})$$

such that $\bar{\rho} + i\bar{\eta} = -(V_{ud}V_{ub}^*) / (V_{cd}V_{cb}^*)$ is independent of phase convention [5]. Thus the unitary triangle in Chapter 2.1.1 can constrain 4 parameters of the SM.

B. Distribution of variables used for the arbitration approach

This appendix gives information on the variables used for the arbitration approach discussed in Section 6.2.9.

In Figures B.1a-B.2d, B_{tag} , $\mathcal{P}_{\mathcal{B}_{\text{tag}}}$, $\Delta(\Delta M_{D^* - D})$ and ΔM_D are shown to have distinct distributions. Here the definition of misreconstructed or correctly reconstructed events depends on the reconstruction of the lepton.

B. Distribution of variables used for the arbitration approach

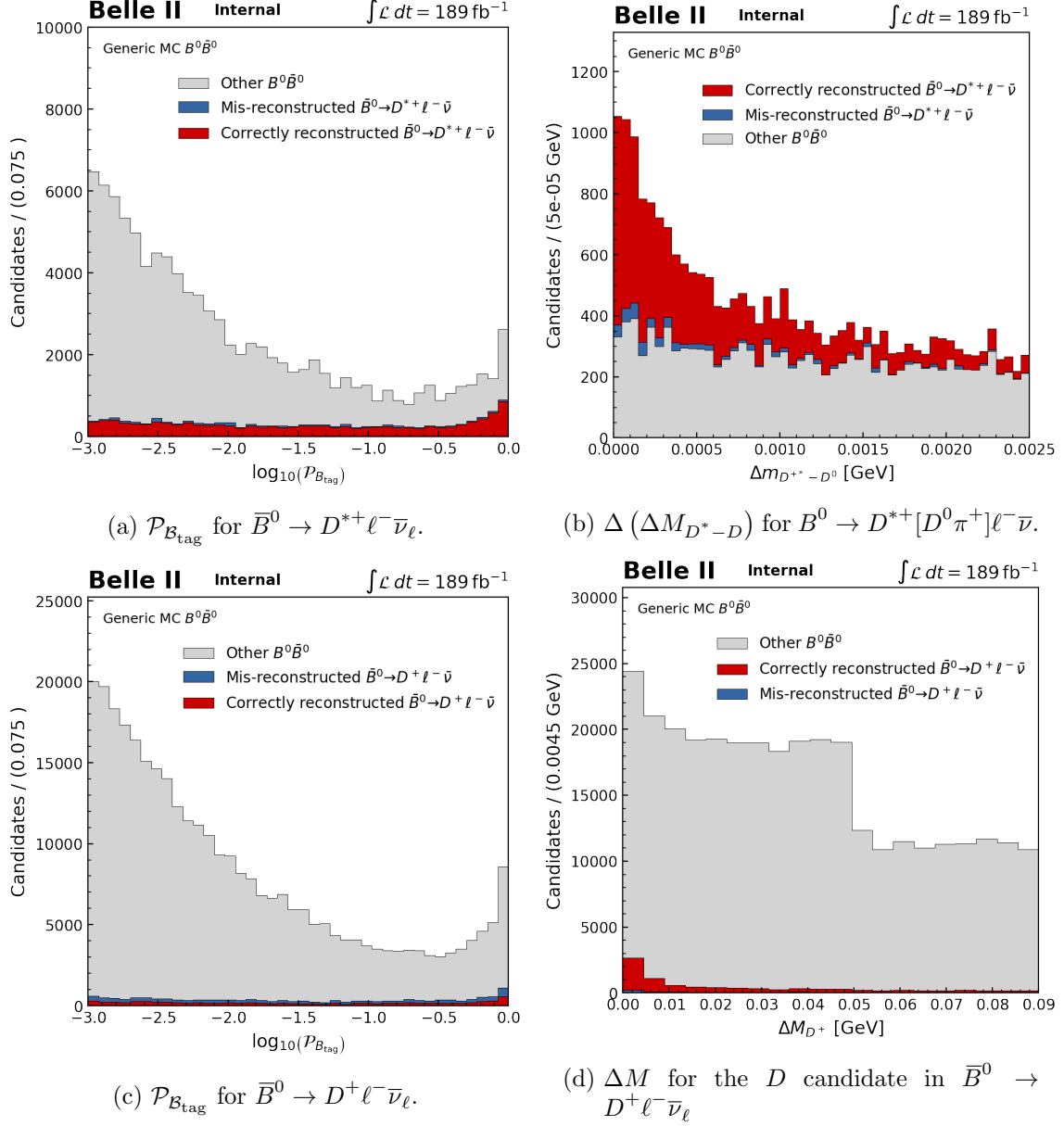


Figure B.1.: Distributions of a selection of variables used for ranking $\bar{B}^0 \rightarrow D^{(*)+} \ell^- \bar{\nu}_\ell$ events by B_{sig} candidate. The $B^0 \bar{B}^0$ generic MC sample is classified using TopoAna and the truth-matched signal-side lepton.

B. Distribution of variables used for the arbitration approach

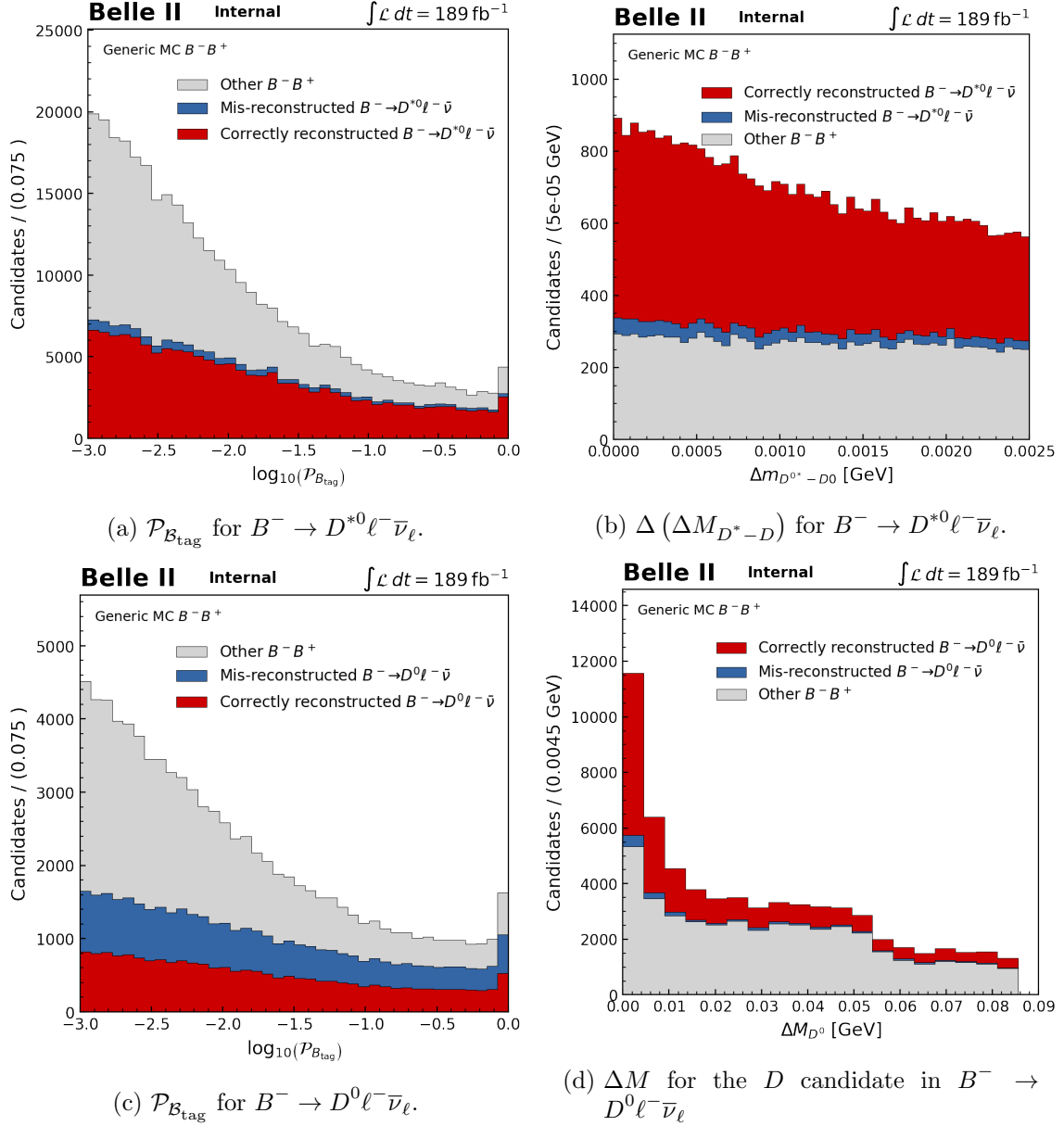


Figure B.2.: Distributions of a selection of variables used for ranking $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ events by B_{sig} candidate. The $B^+ B^-$ generic MC sample is classified using **TopoAna** and the truth-matched signal-side lepton.

C. E_{ECL}

For context, the E_{ECL} distributions in generic MC are shown in Section 6.3.2 before application of the E_{ECL} MVA (Figure C.1). The E_{ECL} distributions are then shown for signal MC before and after application of the E_{ECL} MVA in Figure C.2.

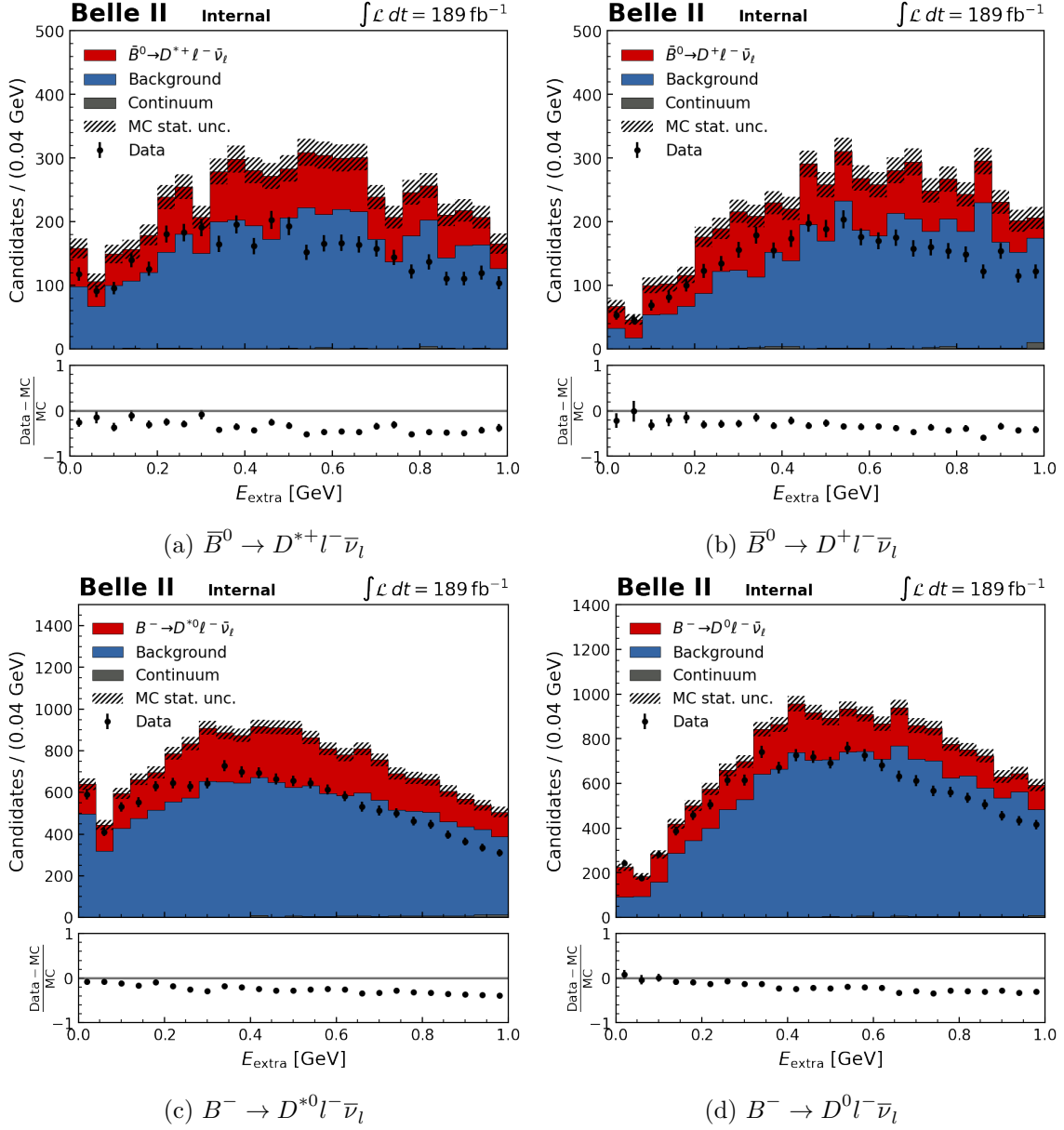


Figure C.1.: The E_{ECL} distributions before the E_{ECL} beam background suppression is applied. Generic MC is classed as the normalisation channels ($B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$), background and continuum. Data is overlaid as points with statistical uncertainties shown as error bars.

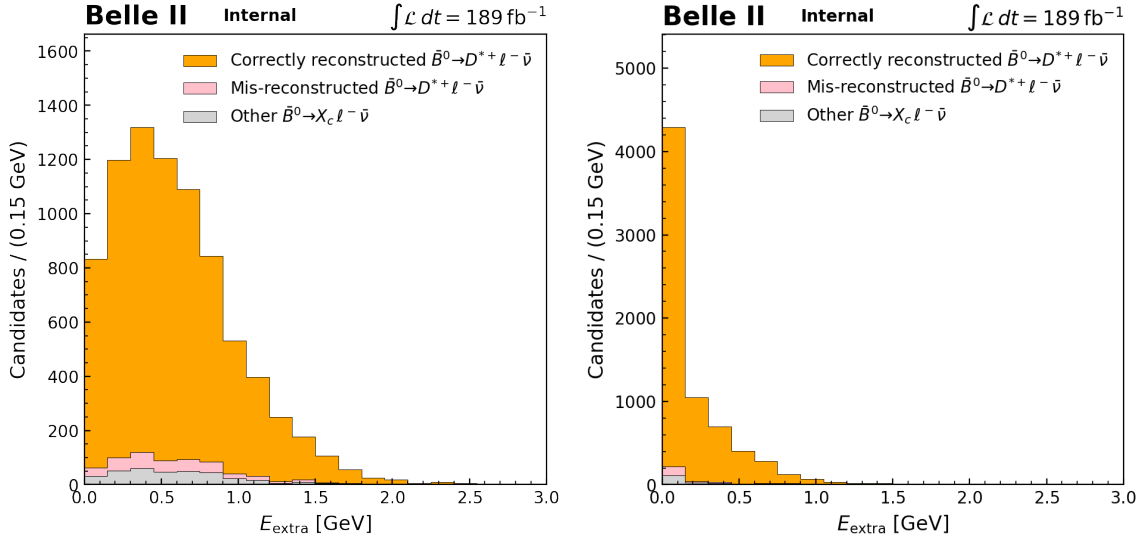


Figure C.2.: Distributions of the E_{ECL} for $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ signal MC before (left) and after (right) the E_{ECL} beam background suppression is applied. The signal MC is classified using **TopoAna** as selected $B \rightarrow D^* \ell^- \bar{\nu}_\ell$ events from generated $B^0 \rightarrow X_C \ell^- \bar{\nu}_\ell$. Correct reconstruction depends on the truth matched signal-side lepton in the event. The other $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ decays show similar changes in their E_{ECL} distributions.

D. The charmed meson D^{**}

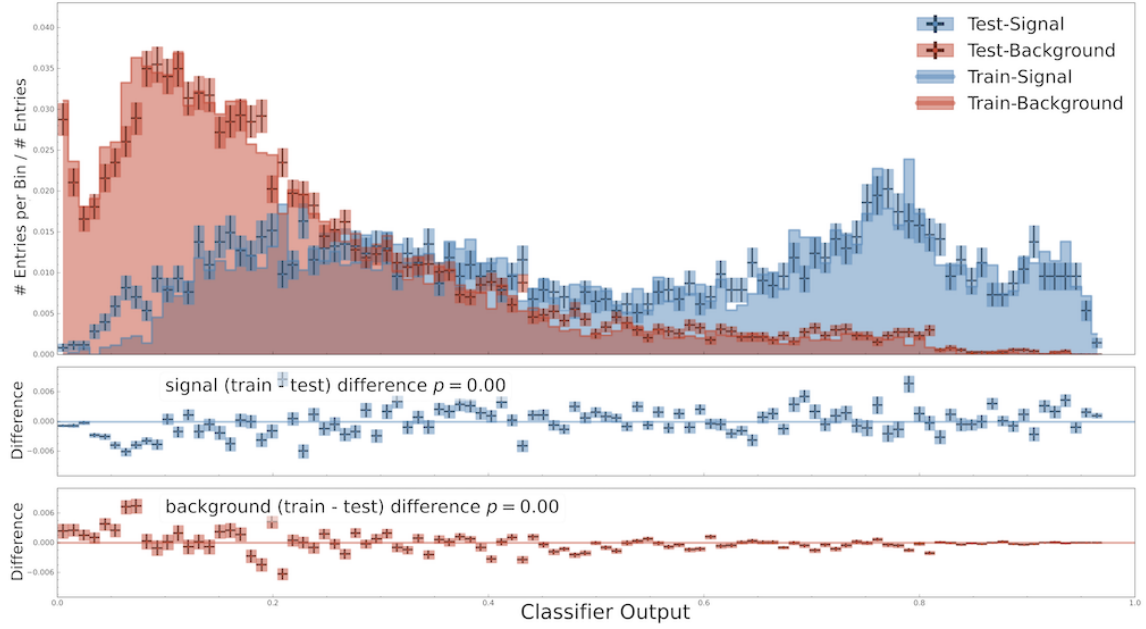
Section 6.4 introduces the D^{**} meson. These mesons are highly relevant to the study of B meson semileptonic decays due to their masses and due to the angular decays of the D^{**} to $D \pi$ and $D^* \pi$. In the measurements of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$, the D^{**} decay is a poorly understood peaking background (a background that peaks in the signal region of the observable measured). As a result, the D^{**} mesons contribute towards the systematic uncertainty of the measurements of the branching fractions of $B \rightarrow D^{(*)} l^- \bar{\nu}_l$.

A heavy-light meson, such as the D^{**} meson, is formed of one heavy quark Q and one light quark q . Due to parity and angular momentum conservation, the possible D^{**} states are restricted to just four D^{**} states. The different doublets of D^{**} mesons can be differentiated through the angular momentum of the light quark $j=S_q+L$ (where S is the spin angular momentum and L is the orbital angular momentum) and their total angular momentum $J=j+S_Q$ [64]. D^{**} mesons make up two doublets with $j=\frac{1}{2}, \frac{3}{2}$.

The $j=\frac{1}{2}$ doublet mesons are labelled by the PDG as $D_0^*(2400)^0$ and $D_1'(2430)^0$ and are predicted to decay only through S-wave decay and are relatively broad. The $j=\frac{3}{2}$ doublet mesons are labelled by the PDG as $D_2^*(2460)^0$ and $D_1(2420)^0$, and are predicted to only decay through D-wave decay and are relatively narrow. This is due to the larger angular momentum needed to be overcome. These are the four D^{**} decay modes simulated in the Belle II MC, using the LLSW model [71].

E. Supplementary MVA resources

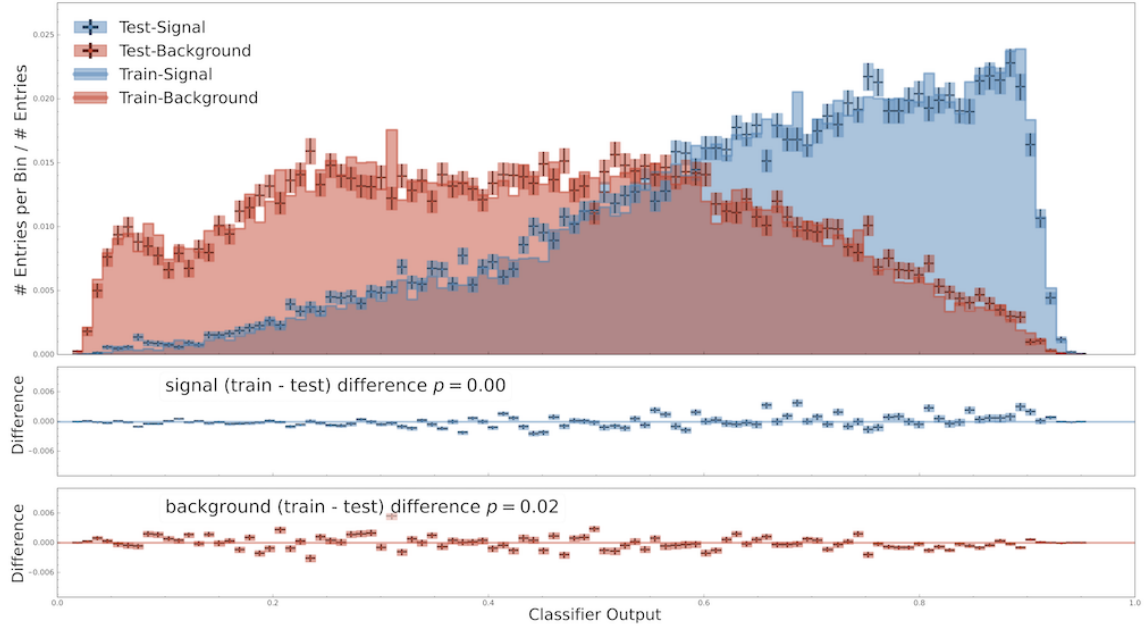
For succinctness of the main thesis text, the corresponding plots for the $\bar{B}^0 \rightarrow D^+ l^- \bar{\nu}_l$, $B^- \rightarrow D^{*0} l^- \bar{\nu}_l$ and $B^- \rightarrow D^0 l^- \bar{\nu}_l$ MVAs from Section 6.4.5 are collated in this section as Figures E.1a to E.4. In addition, the supplementary tables, Tables E.2 to E.6 are included.



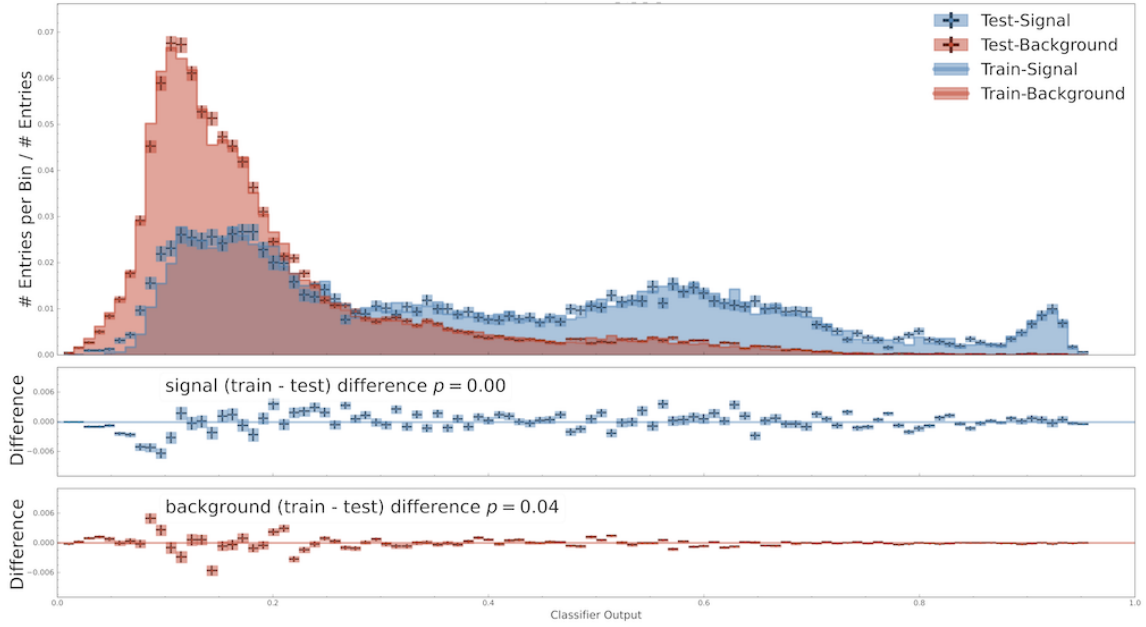
(a) $\bar{B}^0 \rightarrow D^+ l^- \bar{\nu}_l$

Figure E.1.: The classifier output for the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ MVA. The red sample represents the backgrounds and the blue sample represents the signal and normalisation samples. The filled histograms are the training data, whilst the points with statistical error bars represent the test data.

E. Supplementary MVA resources



(b) $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$



(c) $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$

Figure E.1.: The classifier output for each of the $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ MVAs. The red sample represents the backgrounds and the blue sample represents the signal and normalisation samples. The filled histograms are the training data, whilst the points with statistical error bars represent the test data.

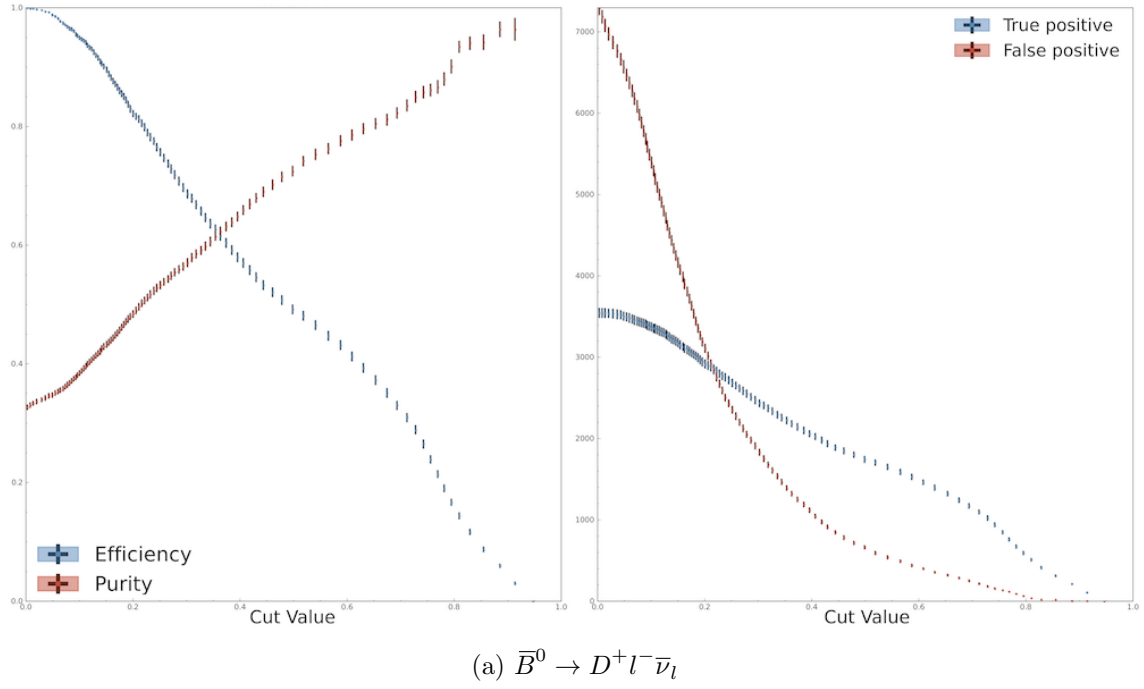
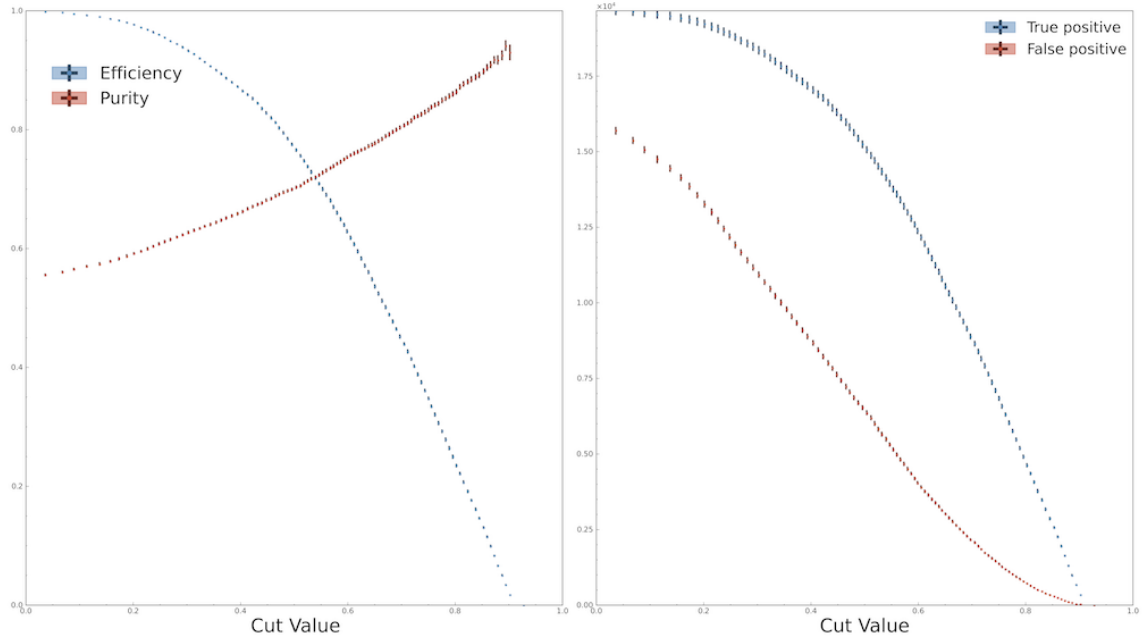
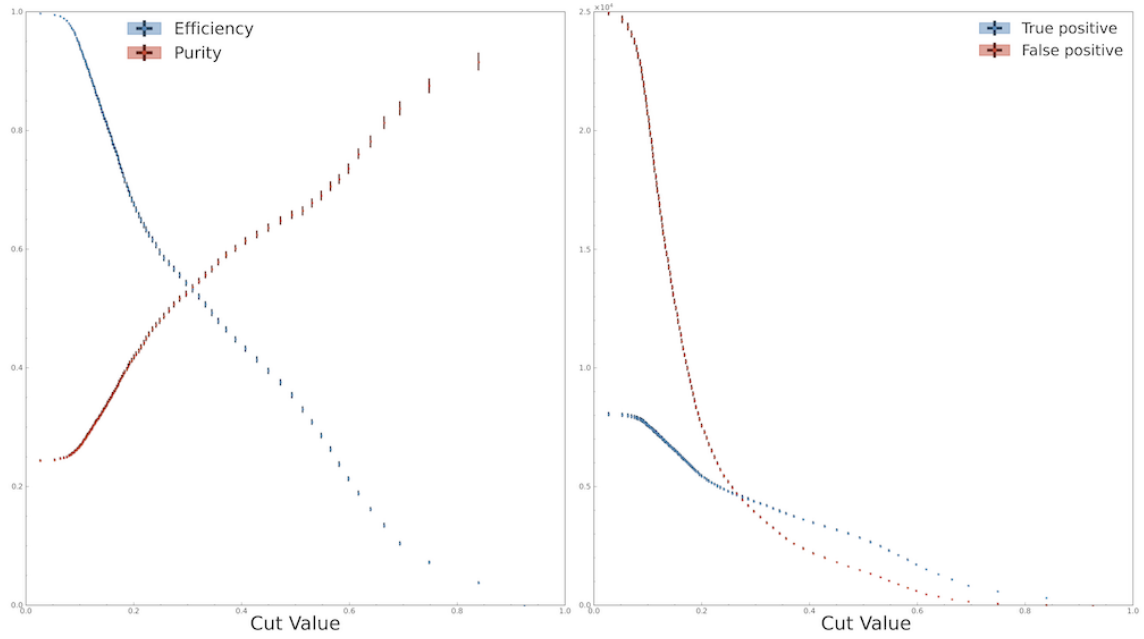


Figure E.2.: The classification result for the $B \rightarrow D \ell^- \bar{\nu}_\ell$ MVA. The purity (left plot, red) and efficiency are shown (left plot, blue), and the true positive (right plot, blue) and false positive rates (right plot, red) are shown (right), all as a function of classifier output cut.

E. Supplementary MVA resources



(b) $B^- \rightarrow D^{*0} l^- \bar{\nu}_l$



(c) $B^- \rightarrow D^0 l^- \bar{\nu}_l$

Figure E.2.: The classification result for each of the $B^- \rightarrow D^{(*)0} l^- \bar{\nu}_l$ MVAs. For the two decay channels the purity (left plot, red) and efficiency are shown (left plot, blue), and the true positive (right plot, blue) and false positive rates (right plot, red) are shown (right), all as a function of classifier output cut.

E. Supplementary MVA resources

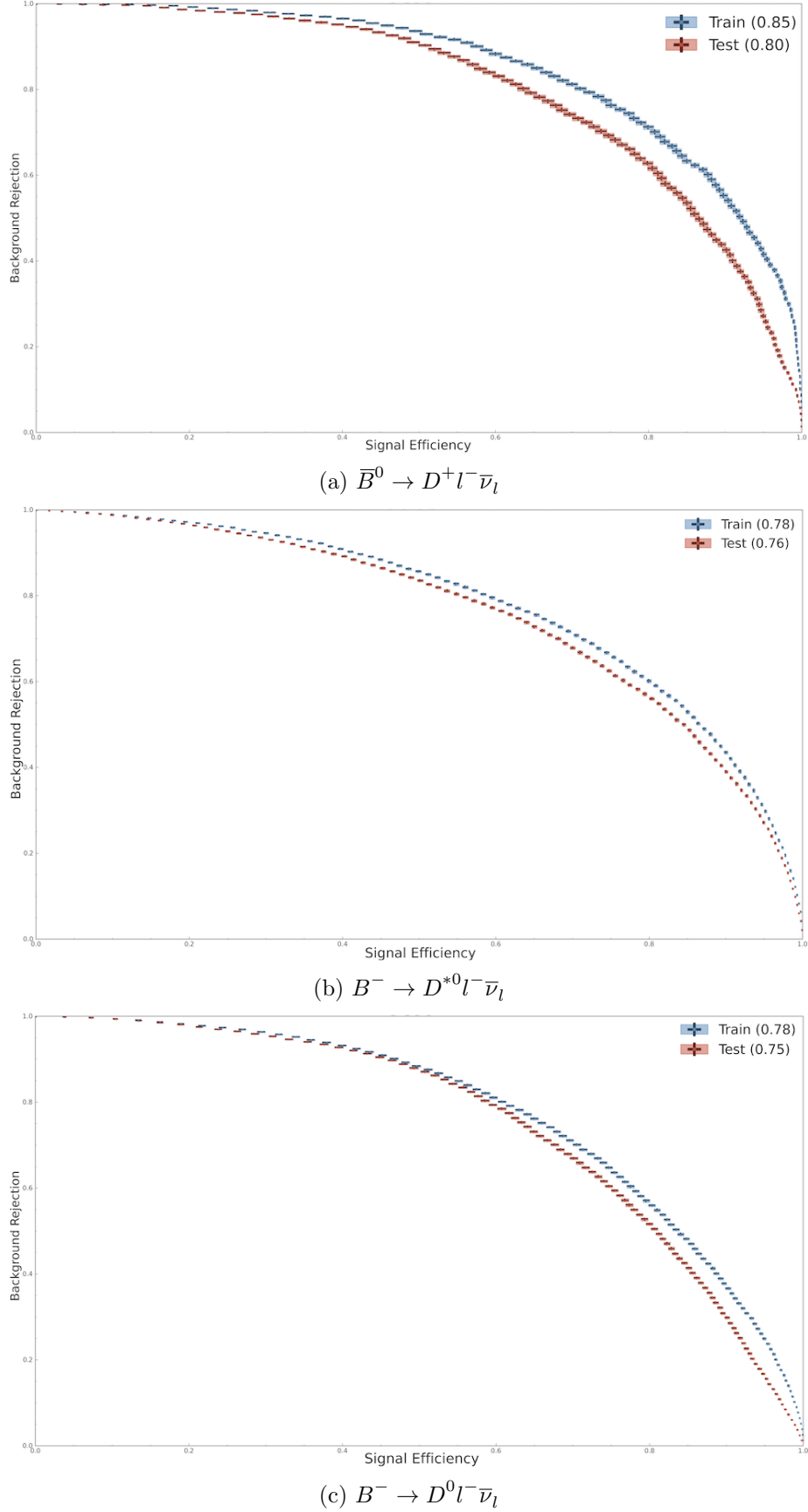


Figure E.3.: The ROC test result (red) in comparison to the training ROC (blue) for each of the $B \rightarrow D^{(*)} l^- \bar{\nu}_l$ MVAs. AUC values are shown and indicate good agreement.

Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	1379	1370	1305	1185	1029	889	777	660	445	159
Background	2709	2487	1742	1143	728	471	315	201	91	18
$\bar{B}^0 \rightarrow D^+ \tau^- \bar{\nu}_\tau$	26	25	19	14	11	9	8	4	0	0
$\bar{B}^0 \rightarrow D^{**} \ell^- \bar{\nu}_\ell$	399	327	155	69	34	21	13	6	2	0
Continuum	140	93	41	21	12	8	5	1	0	0
FOM	21.21	21.8	23.49	24.46	24.48	24.03	23.47	22.46	19.21	11.97

Table E.1.: The yield for the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values).

Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$	0.00	0.65	5.33	14.01	25.33	35.53	43.59	52.13	67.71	88.40
Background	0.00	8.19	35.69	57.79	73.12	82.61	88.35	92.55	96.63	99.32
$\bar{B}^0 \rightarrow D^+ \tau^- \bar{\nu}_\tau$	0.00	4.32	28.16	45.18	55.39	63.15	68.77	82.67	96.45	100.00
$\bar{B}^0 \rightarrow D^{**} \ell^- \bar{\nu}_\ell$	0.00	17.99	61.19	82.70	91.37	94.67	96.69	98.37	99.48	99.96
Continuum	0.00	33.27	70.62	84.42	91.16	94.25	96.43	98.78	99.62	99.86
FOM	21.21	21.80	23.49	24.46	24.48	24.03	23.47	22.46	19.21	11.97

Table E.2.: The percentage drop in yield for the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $\bar{B}^0 \rightarrow D^+ \tau^- \bar{\nu}_\tau$ signal and the $\bar{B}^0 \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. The percentage drop in yields shown are calculated from the non-rounded yields in Table E.1.

Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	4154	4153	4145	4105	4031	3882	3519	2768	1792	433
Background	2305	2295	2220	2033	1796	1534	1166	700	308	48
$B^- \rightarrow D^{*0} \tau^- \bar{\nu}_\tau$	73	73	73	70	62	55	45	30	13	1
$B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$	404	404	394	350	288	234	175	101	42	5
Continuum	139	138	117	86	60	42	26	13	4	0
FOM	51.14	51.18	51.49	52.03	52.54	52.54	51.27	46.92	39.07	19.72

Table E.3.: The yield for the $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^{*0} \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values).

Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$	0.00	0.01	0.21	1.18	2.96	6.56	15.28	33.35	56.85	89.57
Background	0.00	0.40	3.66	11.79	22.05	33.43	49.40	69.60	86.62	97.88
$B^- \rightarrow D^{*0} \tau^- \bar{\nu}_\tau$	0.00	0.00	0.62	5.34	15.23	25.55	37.86	58.53	81.12	98.23
$B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$	0.00	0.00	2.52	13.33	28.73	42.01	56.69	74.93	89.39	98.59
Continuum	0.00	0.74	15.83	37.70	56.31	69.44	81.01	90.23	96.63	99.57
FOM	51.14	51.18	51.49	52.03	52.54	52.54	51.27	46.92	39.07	19.72

Table E.4.: The percentage drop in yield for the $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^{*0} \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. The percentage drop in yields shown are calculated from the non-rounded yields in Table E.3.

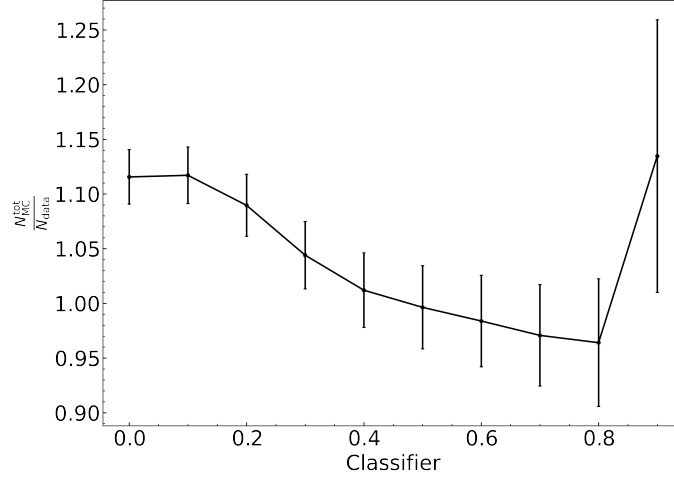
Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	3620	3531	2821	2324	1987	1700	1147	706	376	83
Background	13508	11356	4989	2740	1684	1161	564	237	77	10
$B^- \rightarrow D^0 \tau^- \bar{\nu}_\tau$	78	69	40	30	21	15	3	0	0	0
$B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$	1121	801	187	96	61	41	10	2	0	0
Continuum	396	251	64	39	25	15	5	1	0	0
FOM	27.35	28.7	31.79	32.53	32.69	31.7	27.7	22.97	17.65	8.62

Table E.5.: The yield for the $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^0 \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. Note that the number of events is rounded to an integer value (weighting renders events non-integer values).

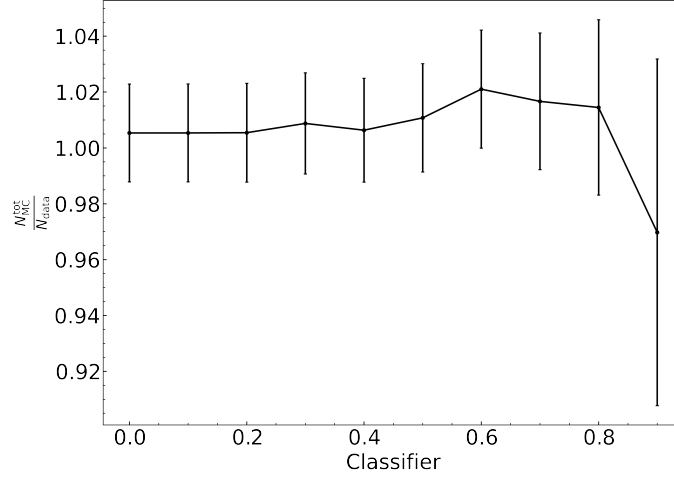
Cut	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$	0	2.47	22.09	35.80	45.11	53.02	68.29	80.48	89.60	97.69
Background	0	15.93	63.07	79.71	87.53	91.40	95.82	98.24	99.42	99.92
$B^- \rightarrow D^0 \tau^- \bar{\nu}_\tau$	0	11.20	48.84	60.79	72.36	80.01	95.72	99.12	100	100
$B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$	0	28.54	83.30	91.35	94.49	96.32	99.04	99.76	99.95	99.99
Continuum	0	36.66	83.73	90.01	93.65	96.03	98.62	99.52	99.79	100
FOM	27.35	28.70	31.79	32.53	32.69	31.70	27.70	22.97	17.65	8.62

Table E.6.: The percentage drop in yield for the $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decay channel as a function of MVA classifier per generic MC type. The $B^- \rightarrow D^0 \tau^- \bar{\nu}_\tau$ signal and the $B^- \rightarrow D^{**} \ell^- \bar{\nu}_\ell$ background is a subset of the background MC in this instance. The FOM is also included for reference. The percentage drop in yields shown are calculated from the non-rounded yields in Table E.5.

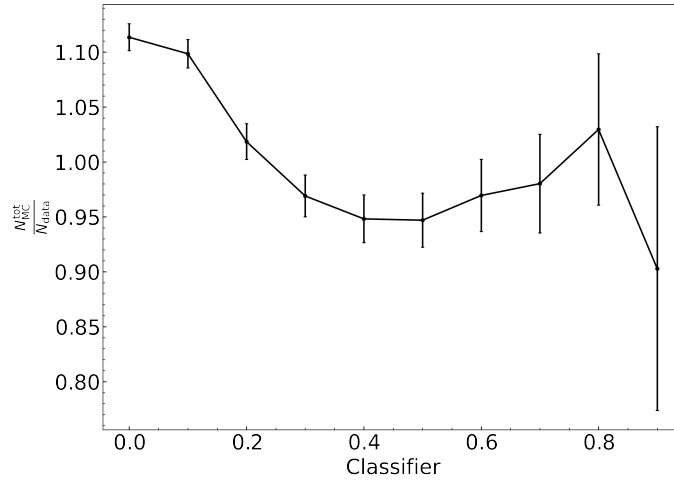
E. Supplementary MVA resources



(a) $\bar{B}^0 \rightarrow D^+ l^- \bar{\nu}_l$



(b) $B^- \rightarrow D^{*0} l^- \bar{\nu}_l$



(c) $B^- \rightarrow D^0 l^- \bar{\nu}_l$

Figure E.4.: The data-MC ratio as a function of classifier cut for the $\bar{B}^0 \rightarrow D^{*+} l^- \bar{\nu}_l$ decay channel. Errors shown are statistical uncertainties.

F. Suppression of D^* cross-feed for

$$B \rightarrow D\ell^-\bar{\nu}_\ell$$

As can be seen in Section 6.4.5, in the histograms representing $B \rightarrow D\ell^-\bar{\nu}_\ell$ such as Figure 6.15b, there is a high proportion of B^0 background compared to the signal decays. Using **TopoAna**, these backgrounds are categorized.

The largest contributing backgrounds to the $B \rightarrow D\ell^-\bar{\nu}_\ell$ decays are the cross-feed of D^* mesons that are misidentified as D mesons. Using this background categorization it is possible to examine the different kinematic distributions between the $B \rightarrow D^*\ell^-\bar{\nu}_\ell$ and $B \rightarrow D\ell^-\bar{\nu}_\ell$ decays to find distinguishing features. Thus the combination of distributions of m_{miss}^2 and p_{lep} were examined. Figure F.1 displays the m_{miss}^2 versus lepton momentum of the $B \rightarrow D\ell^-\bar{\nu}_\ell$ sample. There are two distinct combined shapes, a vertical distribution and a diagonal distribution. Using **TopoAna** truth matching information, Figure F.2 shows a clear distinction between the D meson and the D^* cross-feed. The D^* cross-feed could therefore be suppressed using a m_{miss}^2 dependent p_{lep} cut. Through optimization of the cut using $\frac{s}{\sqrt{s+b}}$ (where s is signal yield and b is background yield), the selection of $p_{\text{lep}} < -3.7 \times m_{\text{miss}}^2 + 2.4$ was implemented. The cut is shown in Figure F.2, and the resulting distributions after implementation can be seen in Figures F.3 and F.4. The proportion of $\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$ events in the MC sample improves from 13.3% before the cut to 77.1% after the cut. Only 15.6% of signal events were lost, while 85.4% of backgrounds were removed.

Though this reduction in D^* cross-feed is highly beneficial in the suppression of $B \rightarrow D\ell^-\bar{\nu}_\ell$ backgrounds, we are left at an impasse due to the impact this specific optimization would have on $B \rightarrow D\tau^-\bar{\nu}_\tau$ events. As discussed in Section 6.1. the signal region for m_{miss}^2 of the $B \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ events is not around 0 as for the $B \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ events. This is due

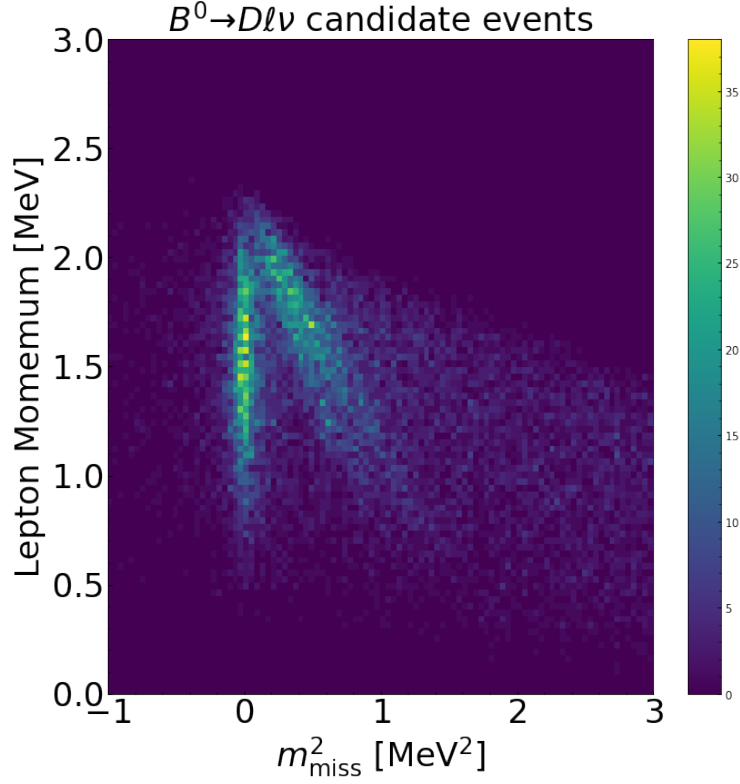


Figure F.1.: A 2D distribution of m_{miss}^2 versus p_{lep} of the candidate $\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$ events.

to the multi-neutrino final state of the decay produced by the τ lepton. Thus this result is not carried forward in the calculation of $R(D^*)$ in this analysis, but retained as knowledge for future studies of the $B \rightarrow D\ell^-\bar{\nu}_\ell$ decays alone.

Instead, the multivariate analysis technique is used to emulate this the same effect but with application to both the $B \rightarrow D\tau^-\bar{\nu}_\tau$ and $B \rightarrow D\ell^-\bar{\nu}_\ell$ decay channels simultaneously.

F. Suppression of D^* cross-feed for $B \rightarrow D\ell^-\bar{\nu}_\ell$

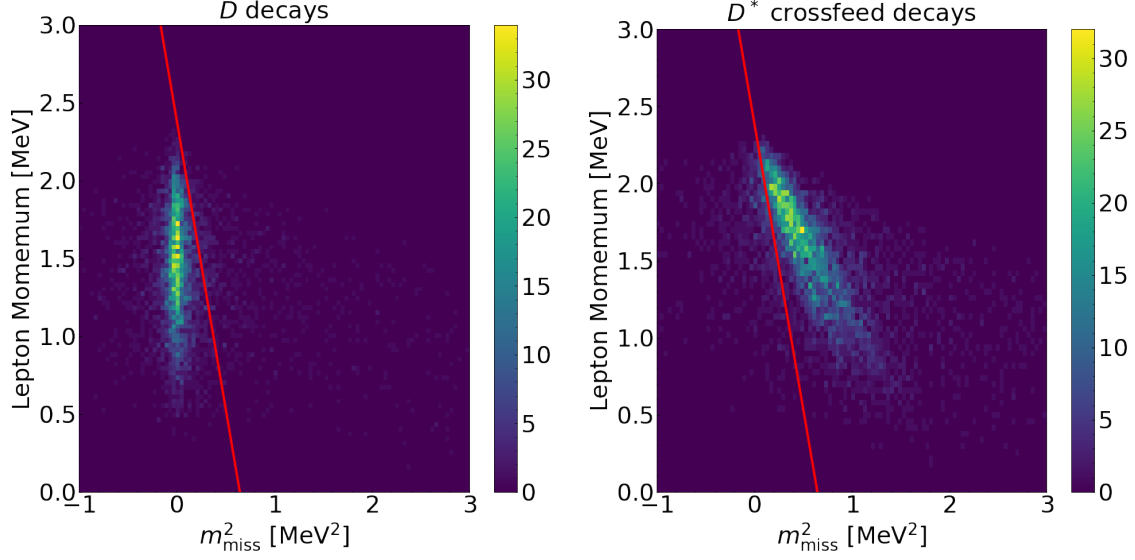


Figure F.2.: The separated 2D distributions of m^2_{miss} versus p_{lep} of the $\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$ candidate events using **TopoAna** truth information. The $\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$ distribution (left) can be clearly seen to be the vertical distribution, while the contributing D^* cross-feed (right) has a diagonal distribution.

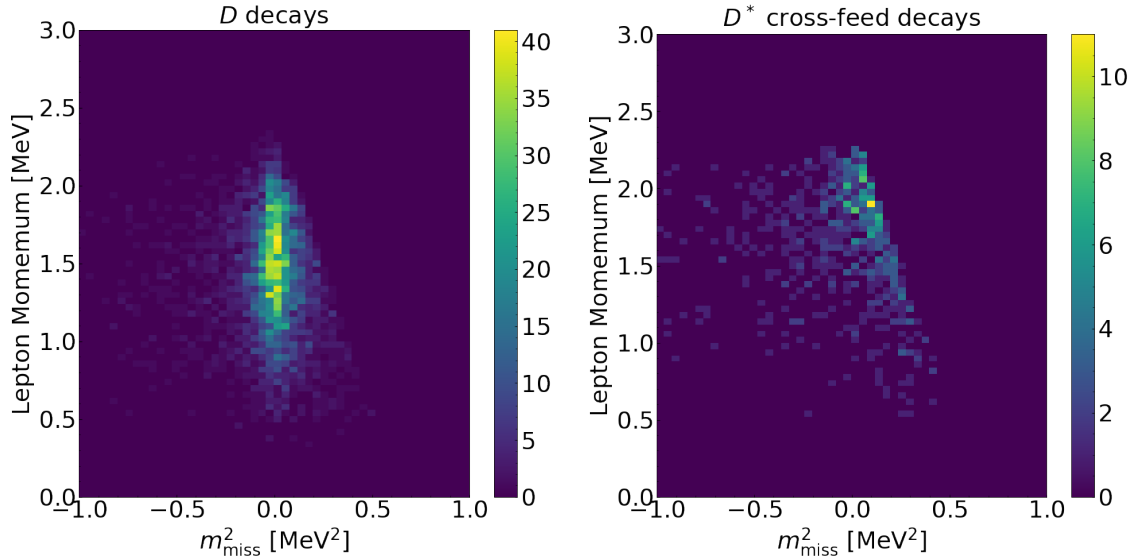


Figure F.3.: The separated 2D distributions of m^2_{miss} versus p_{lep} of the $\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$ decay using **TopoAna** truth information. The $\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$ distribution (left) was minimally affected by the cut, while the contributing D^* cross-feed (right) has a largely reduced contribution to the $B \rightarrow D\ell^-\bar{\nu}_\ell$ signal region.

F. Suppression of D^* cross-feed for $B \rightarrow D\ell^-\bar{\nu}_\ell$

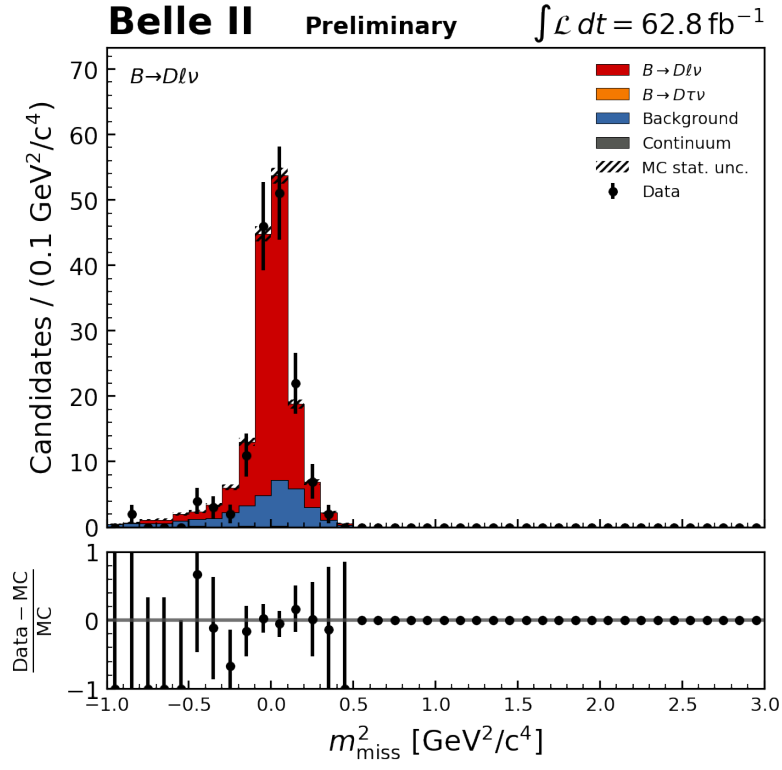


Figure F.4.: The resulting m_{miss}^2 distribution for $\bar{B}^0 \rightarrow D^+\ell^-\bar{\nu}_\ell$ after the suppression of D^* cross-feed events. The study was performed with 62.8 fb^{-1} .

G. Alternate binning plots

The binning of the m_{miss}^2 in this analysis has a historical reasoning for comparison to Reference [1]. This appendix contains the alternate binning distributions for each of the m_{miss}^2 for completion, Figures G.1a and G.1b, discussed in Section 6.4.6.

G. Alternate binning plots

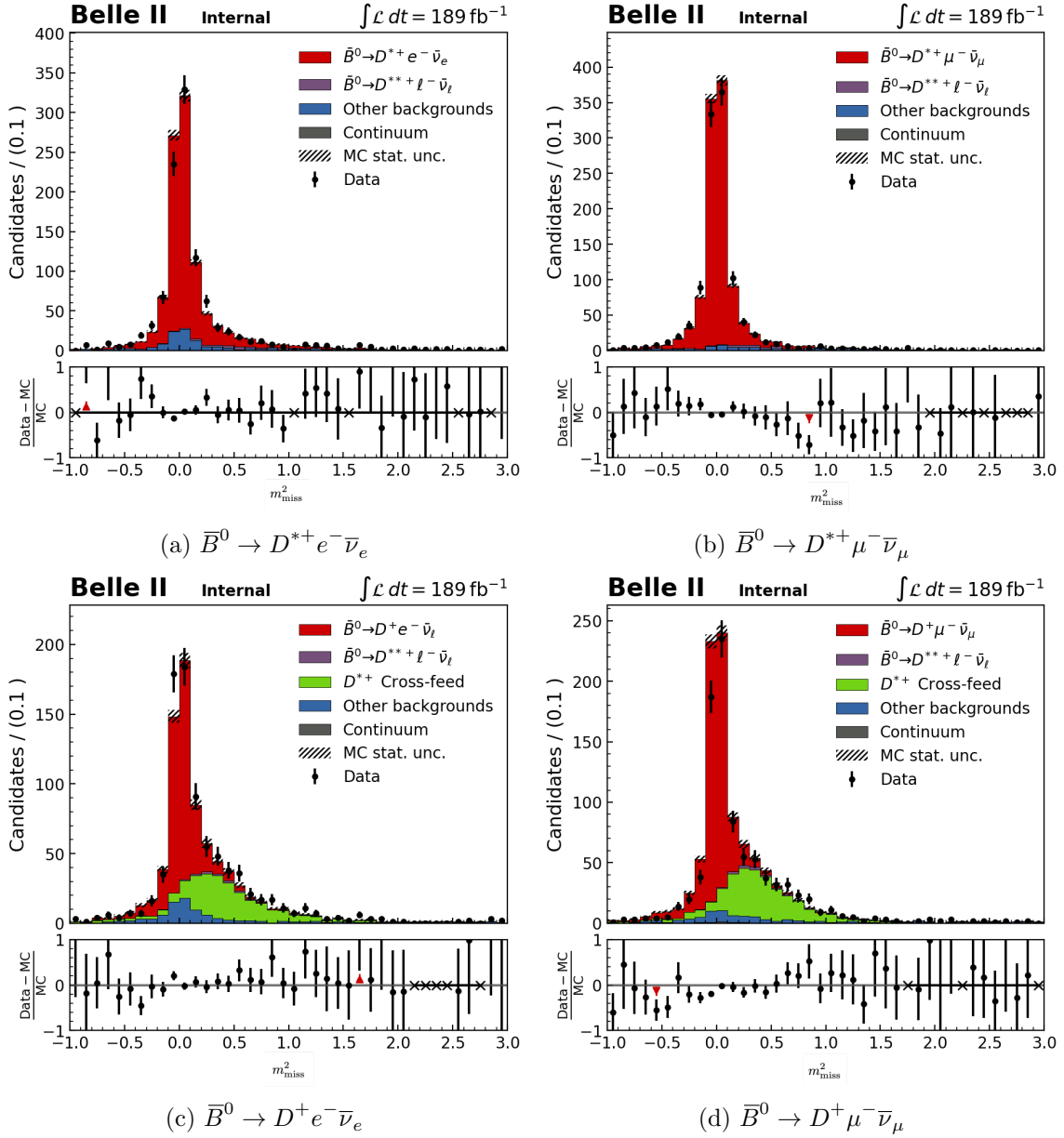


Figure G.1.: The post-MVA classifier cut m_{miss}^2 for each of the $\bar{B}^0 \rightarrow D^{(*)+} \ell^- \bar{\nu}_\ell$ MVAs.

G. Alternate binning plots

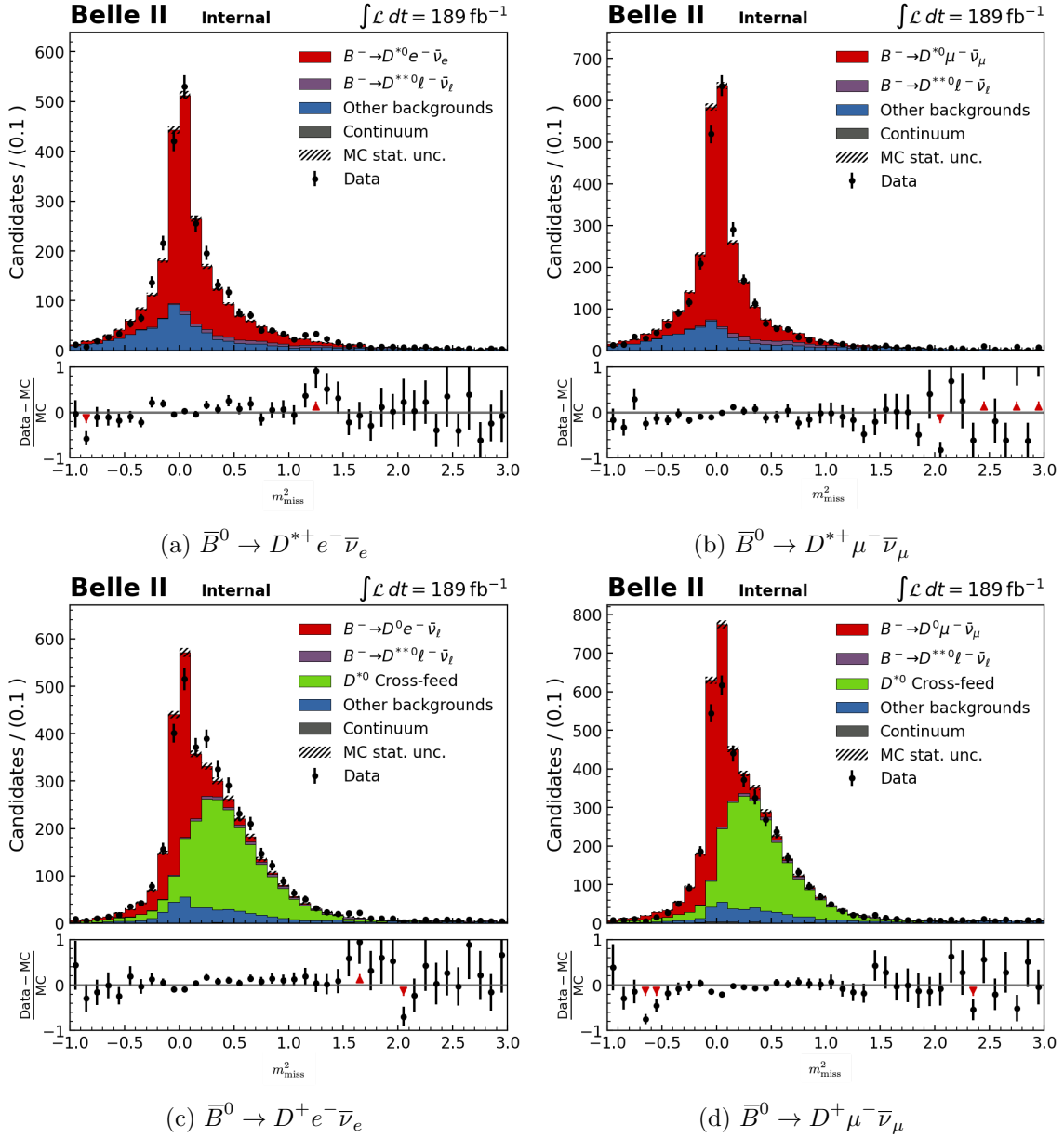


Figure G.2.: The post-MVA classifier cut m_{miss}^2 for each of the $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ MVAs.

H. Systematic uncertainties appendices

Supplementary information for Chapter 8.1.2.

H.1. Tracking efficiency systematics

The fractions of each decay channel present in the $B \rightarrow D^{(*)} \ell^- \bar{\nu}_\ell$ normalisation modes are measured in signal MC and used to calculate the average number of tracks associated with each mode. The number of signal side tracks represents the number of particle tracks present in each B_{sig} meson decay, including the lepton track. These are presented in Tables H.1 and H.2.

H.2. MVA systematics

Supplementary figures for the MVA systematic calculations presented in Chapter 8.1.5. These correspond to the $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$, $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$ and $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$ decays.

H. Systematic uncertainties appendices

Mode	Fraction	Number of signal-side tracks
$D^{*+} \rightarrow D^0 \pi^+ \rightarrow K^- \pi^+$	19.18	3
$K_S^0 \pi^0$	3.61	1
$K^- \pi^+ \pi^0$	40.01	3
$K^- \pi^+ \pi^- \pi^+$	17.67	5
$K_S^0 \pi^+ \pi^-$	5.24	3
$K_S^0 \pi^+ \pi^- \pi^0$	10.81	3
$K^- K^+$	2.42	3
$\pi^- \pi^+$	1.06	3
$D^{*+} \rightarrow D^+ \pi^0 \rightarrow K^- \pi^+ \pi^+$	32.54	4
$K^- \pi^+ \pi^+ \pi^0$	29.46	4
$K_S^0 \pi^+$	5.78	2
$K_S^0 \pi^+ \pi^- \pi^+$	4.9	4
$K_S^0 \pi^+ \pi^0$	21.61	2
$K_S^0 K^+$	1.01	2
$K^- K^+ \pi^+$	4.71	4
$D^+ \rightarrow K^- \pi^+ \pi^+$	38.8	4
$K^- \pi^+ \pi^+ \pi^0$	20.77	4
$K_S^0 \pi^+$	6.46	2
$K_S^0 \pi^+ \pi^- \pi^+$	6.18	4
$K_S^0 \pi^+ \pi^0$	21.65	2
$K_S^0 K^+$	1.44	2
$K^- K^+ \pi^+$	4.69	4

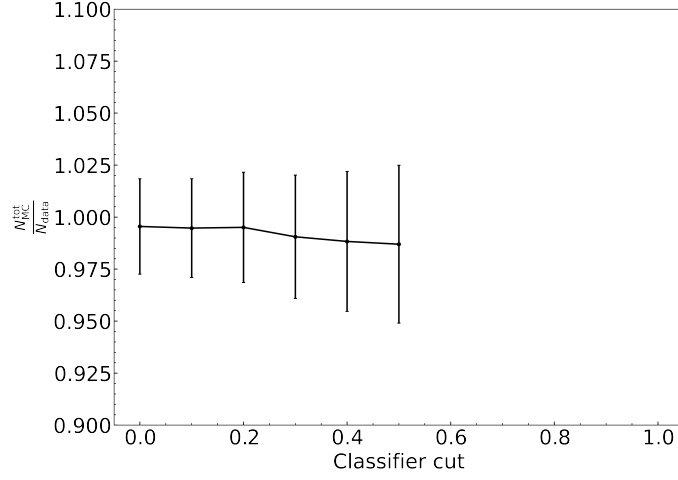
Table H.1.: The fraction of each $D^{(*)}$ decay channel present in $\bar{B}^0 \rightarrow D^{(*)+} \ell^- \bar{\nu}_\ell$ signal MC, and the associated number of signal side tracks per event.

H. Systematic uncertainties appendices

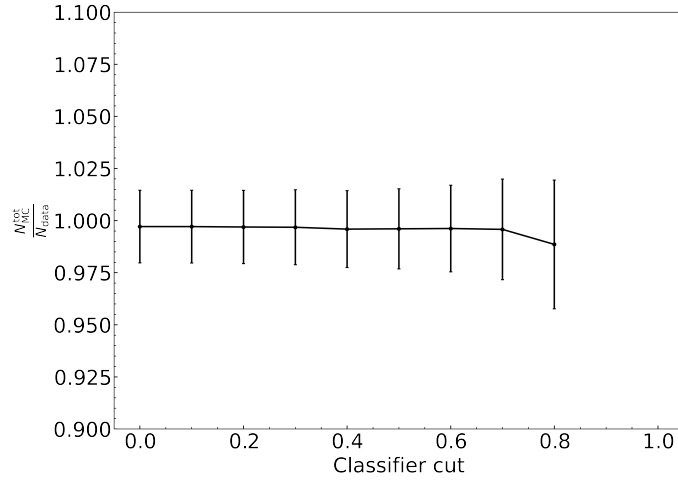
Mode	Fraction	Number of signal-side tracks
$D^{*0} \rightarrow D^0 \pi^0 \rightarrow K^- \pi^+$	8.351	3
$K_S^0 \pi^0$	1.351	1
$K^- \pi^+ \pi^0$	21.247	3
$K^- \pi^+ \pi^- \pi^+$	8.294	5
$K_S^0 \pi^+ \pi^-$	2.777	3
$K_S^0 \pi^+ \pi^- \pi^0$	6.748	3
$K^- K^+$	1.168	3
$\pi^- \pi^+$	0.499	3
$D^{*0} \rightarrow D^0 \gamma \rightarrow K^- \pi^+$	7.935	4
$K_S^0 \pi^0$	1.37	4
$K^- \pi^+ \pi^0$	21.228	2
$K^- \pi^+ \pi^- \pi^+$	7.922	4
$K_S^0 \pi^+ \pi^-$	2.588	2
$K_S^0 \pi^+ \pi^- \pi^0$	7.013	2
$K^- K^+$	1.029	4
$\pi^- \pi^+$	0.48	3
$D^0 \rightarrow K^- \pi^+$	19.18	3
$K_S^0 \pi^0$	3.61	1
$K^- \pi^+ \pi^0$	40.01	3
$K^- \pi^+ \pi^- \pi^+$	17.67	5
$K_S^0 \pi^+ \pi^-$	5.24	3
$K_S^0 \pi^+ \pi^- \pi^0$	10.81	3
$K^- K^+$	2.42	3
$\pi^- \pi^+$	1.06	3

Table H.2.: The fraction of each $D^{(*)}$ decay channel present in $B^- \rightarrow D^{(*)0} \ell^- \bar{\nu}_\ell$ signal MC, and the associated number of signal side tracks per event.

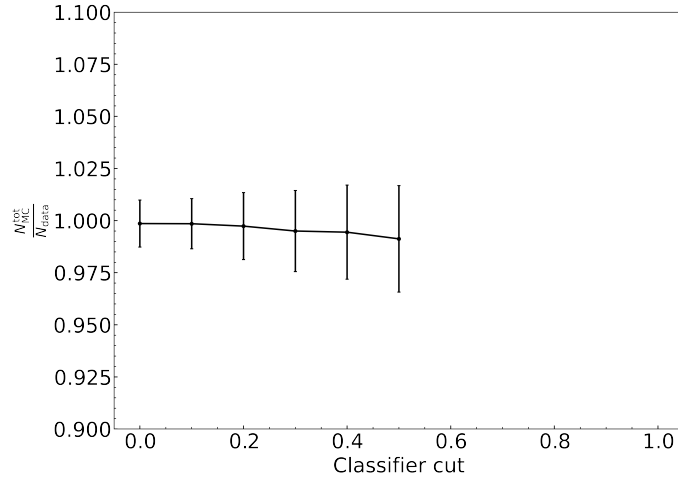
H. Systematic uncertainties appendices



(a) $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$



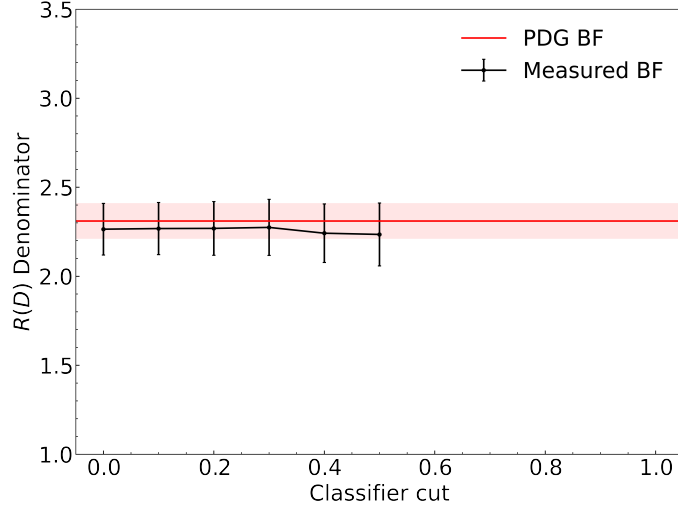
(b) $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$



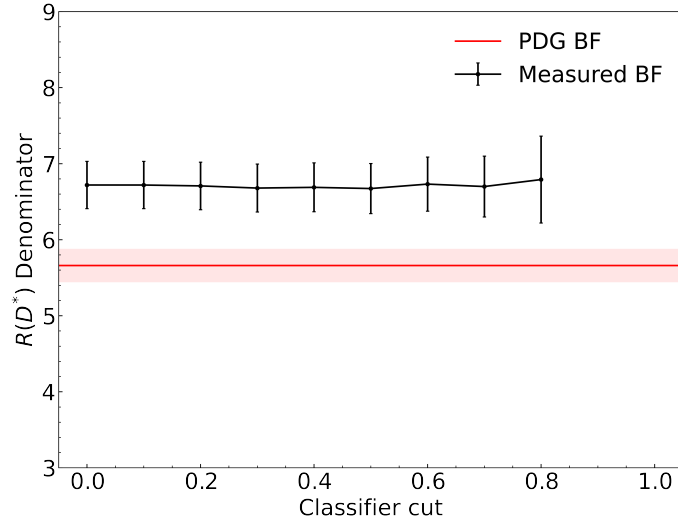
(c) $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$

Figure H.1.: The ratio of MC to data as a function of classifier cut. At higher classifier cuts there is failure due to low statistics in certain bins leading to a non-accurate covariance.

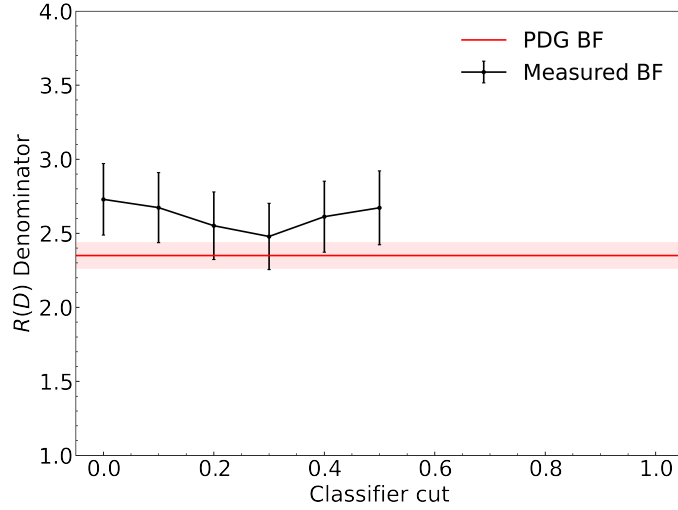
H. Systematic uncertainties appendices



(a) $\bar{B}^0 \rightarrow D^+ \ell^- \bar{\nu}_\ell$



(b) $B^- \rightarrow D^{*0} \ell^- \bar{\nu}_\ell$



(c) $B^- \rightarrow D^0 \ell^- \bar{\nu}_\ell$

Figure H.2.: The denominator of $R(D^*)$ as a function of classifier cut. At higher classifier cuts there is failure due to low statistics in certain bins leading to a non-accurate covariance.

List of Acronymns

- 2HDM: Type II Higgs Doublet Model
- ARICH: Aerogel Ring Imaging Cherenkov
- AUC: Area Under the Curve
- `basf2`: Belle II Analysis Software Framework
- BDT: Boosted Decision Tree
- BSM: Beyond the Standard Model
- CDC: Central Drift Chamber
- CKM: Cabibbo-Kobayashi-Maskawa
- CM: Centre of Mass
- COI: Channel of Interest
- CP : Charge-Parity
- CsI(Tl): Thallium-doped Caesium Iodide
- DAQ: Data Acquisition
- DEPFET: Depleted Field Effect Transistor
- DT: Decision Tree
- ECL: Electromagnetic Calorimeter
- EFT: Effective Field Theory
- FCCC: Flavour Changing Charged-Current
- FEI: Full Event Interpretation
- FF: Form Factor
- FOM: Figure Of Merit
- FSR: Final State Radiation
- GBDT: Gradient Boosted DT
- HFLAV: Heavy Flavour Averaging Group
- HLT: High Level Trigger
- HQET: Heavy Quark Effective Theory
- IP: Interaction Point
- IR: Interaction Region

List of Acronymns

- KLM: neutral Long-lived Kaon and Muon detector
- L1: Level 1
- LDA: Linear Discriminant Analysis
- LEP: Large Electron-Positron
- LFU: Lepton Flavour Universality
- LFUV: Lepton Flavour Universality Violation
- MC: Monte Carlo
- MCP: Micro-Channel Plate
- MVA: Multi-Variate Analysis
- NN: Neural Network
- PDG: Particle Data Group
- PID: Particle Identification
- PMT: Photomultiplier Tube
- PXD: Pixel Vertex Detector
- QCD: Quantum Chromodynamics
- RF: Random Forest
- ROC: Receiver Operating Characteristic
- ROE: Rest Of Event
- SM: Standard Model
- SOD: Significance Of Distance
- SVD: Silicon Vertex Detector
- SVM: Support Vector Machine
- TOP: Time Of Propagation
- **TopoAna**: Topology Analysis
- l : e, μ, τ
- ℓ : e, μ

Bibliography

- [1] The Belle II Collaboration, “Measurement of the semileptonic $\bar{B}^0 \rightarrow D^{*+} \ell^- \nu_\ell$ branching fraction with fully reconstructed B meson decays and 34.6 fb^{-1} of Belle II data,” 2020. arXiv:2008.10299.
- [2] P. Cheema, R. Cheaib, S. H. Robertson, T. Shillington, H. M. Wakeling, *et al.*, “Identification of Beam Background Calorimeter Clusters and Fake Photons at Belle II Using BDT Classifiers,” Tech. Rep. BELLE2-NOTE-TE-2022-015 Draft Version 3.0, Sep 2022.
- [3] T. Abe *et al.*, “Belle II Technical Design Report,” 2010.
- [4] S. Chatrchyan *et al.*, “Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV,” *JHEP*, vol. 06, p. 081, 2013.
- [5] P. D. Group, P. A. Zyla, *et al.*, “Review of Particle Physics,” *Progress of Theoretical and Experimental Physics*, vol. 2020, Aug 2020. 083C01.
- [6] Cush, MissMJ, *et al.*, *Standard Model of Elementary Particles*. Jun 2020.
- [7] Maschen, *QCD Fields Field (physics)*. Sep 2012.
- [8] C. Athanassopoulos *et al.*, “Results on $\nu_\mu \rightarrow \nu_e$ neutrino oscillations from the lsnd experiment,” *Phys. Rev. Lett.*, vol. 81, pp. 1774–1777, Aug 1998.
- [9] F. Abe, H. Akimoto, A. Akopian, M. G. Albrow, *et al.*, “Observation of Top Quark Production in $p-p$ Collisions with the Collider Detector at Fermilab,” *Physical Review Letters*, vol. 74, p. 2626–2631, Apr 1995.

Bibliography

- [10] K. Kodama *et al.*, “Observation of τ neutrino interactions,” *Phys. Lett. B*, vol. 504, pp. 218–224, 2001.
- [11] Y. S. Amhis *et al.*, “Averages of b -hadron, c -hadron, and τ -lepton properties as of 2018,” *Eur. Phys. J.*, vol. C81, p. 226, 2021.
- [12] P. A. M. Dirac, “Quantum theory of emission and absorption of radiation,” *Proc. Roy. Soc. Lond. A*, vol. 114, p. 243, 1927.
- [13] K. G. Wilson, “Confinement of quarks,” *Phys. Rev. D*, vol. 10, pp. 2445–2459, Oct 1974.
- [14] P. Chang, K.-F. Chen, and W.-S. Hou, “Flavor physics and cp violation,” *Progress in Particle and Nuclear Physics*, vol. 97, pp. 261–311, 2017.
- [15] F. U. Bernlochner, M. F. Sevilla, D. J. Robinson, and G. Wormser, “Semitauonic b -hadron decays: A lepton flavor universality laboratory,” *Reviews of Modern Physics*, vol. 94, feb 2022.
- [16] R. Aaij, C. Abellán Beteta, B. Adeva, M. Adinolfi, C. Aidala, Z. Ajaltouni, S. Akar, P. Albicocco, J. Albrecht, F. Alessio, *et al.*, “Search for Lepton-Universality Violation in $B^+ \rightarrow K^+ \ell^+ \ell^-$ Decays,” *Physical Review Letters*, vol. 122, May 2019.
- [17] R. Aaij, C. Abellán Beteta, T. Ackernley, B. Adeva, M. Adinolfi, *et al.*, “Measurement of CP-Averaged Observables in the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ Decay,” *Physical Review Letters*, vol. 125, Jul 2020.
- [18] A. Carvunis, A. Crivellin, D. Guadagnoli, and S. Gangal, “The Forward-Backward Asymmetry in $B \rightarrow D^* \ell \nu$: One more hint for Scalar Leptoquarks?,” 2021.
- [19] B. Abi, T. Albahri, S. Al-Kilani, D. Allspach, L. Alonzi, *et al.*, “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm,” *Physical Review Letters*, vol. 126, Apr 2021.
- [20] A. Bhatta, A. Ray, and R. Mohanta, “Exploring the role of new physics in $b \rightarrow u \tau \bar{\nu}$ decays,” *PTEP*, vol. 2021, no. 7, p. 073B04, 2021.

Bibliography

- [21] J. F. Gunion and H. E. Haber, “CP-conserving two-Higgs-doublet model: The approach to the decoupling limit,” *Physical Review D*, vol. 67, Apr 2003.
- [22] W.-S. Hou, R. Jain, C. Kao, G. Kumar, and T. Modak, “Collider Prospects for Muon $g - 2$ in General Two Higgs Doublet Model,” 2021.
- [23] B. Schrempp and F. Schrempp, “Light leptoquarks,” *Physics Letters B*, vol. 153, no. 1-2, p. 101–107, 1985.
- [24] W.-F. Chang, “One colorful resolution to the neutrino mass generation, three lepton flavor universality anomalies, and the Cabibbo angle anomaly,” *Journal of High Energy Physics*, vol. 2021, Sep 2021.
- [25] H. M. Lee, “Leptoquark option for B -meson anomalies and leptonic signatures,” *Physical Review D*, vol. 104, Jul 2021.
- [26] J. Kriewald, C. Hati, J. Orloff, and A. M. Teixeira, “A combined explanation of the B -meson decay anomalies with a single vector leptoquark,” 2020. arXiv:2012.06315.
- [27] D. A. Faroughy, “Direct searches motivated by recent B -anomalies,” 2018. arXiv:1811.07582.
- [28] J. Albrecht, “Lepton Flavour Universality tests with B decays at LHCb,” 2018. arXiv:1805.06243.
- [29] A. J. Bevan, B. Golob, T. Mannel, S. Prell, B. D. Yabsley, *et al.*, “The Physics of the B Factories,” *The European Physical Journal C*, vol. 74, Nov 2014.
- [30] H. Dijkstra, H. J. Hilke, T. Nakada, and T. Ypsilantis, “LHCb Letter of Intent, LHCb Collaboration,” Jan 1995.
- [31] A. Matyja *et al.*, “Observation of $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decay at Belle,” *Phys. Rev. Lett.*, vol. 99, p. 191807, 2007.
- [32] H. Nowak, “Lifetimes, semileptonic branching ratios of B-hadrons and $|V_{cb}|$ estimation,” *Nuclear Physics B - Proceedings Supplements*, vol. 55, no. 1, p. 63–68, 1997.

Bibliography

- [33] G. Abbiendi, C. Ainsley, P. Åkesson, G. Alexander, J. Allison, *et al.*, “Measurement of the branching ratio for the process $b \rightarrow \tau^- \bar{\nu}_\tau X$,” *Physics Letters B*, vol. 520, no. 1-2, p. 1–10, 2001.
- [34] R. Barate Et Al., “Measurements of $\text{BR}(b \rightarrow \tau^- \bar{\nu}_\tau X)$ and $\text{BR}(b \rightarrow \tau^- \bar{\nu}_\tau D^{*\pm} X)$ and upper limits on $\text{BR}(B^- \rightarrow \tau^- \bar{\nu}_\tau)$ and $\text{BR}(b \rightarrow s \nu \bar{\nu})$,” *The European Physical Journal C*, vol. 19, no. 2, p. 213–227, 2001.
- [35] G. Abbiendi, C. Ainsley, P. Åkesson, G. Alexander, J. Allison, *et al.*, “Measurement of the branching ratio for the process $b \rightarrow \tau^- \bar{\nu}_\tau X$,” *Physics Letters B*, vol. 520, p. 1–10, Nov 2001.
- [36] The LHCb Collaboration, “ $R(D^*)$ and $R(D)$ with $\tau^- \rightarrow \mu^- \nu_\tau \bar{\nu}_\mu$ (LHCb-PAPER-2022-03).” https://indico.cern.ch/event/1187939/attachments/2530158/4355180/DTaunu_CERNSeminar2022.
- [37] S. Kumbhakar, “Signatures of complex new physics in $b \rightarrow c \tau \bar{\nu}$ transitions,” *Nuclear Physics B*, vol. 963, p. 115297, Feb 2021.
- [38] S. Iguro and R. Watanabe, “Bayesian fit analysis to full distribution data of $\bar{B} \rightarrow D^{(*)} \ell \bar{\nu} : |V_{cb}|$ determination and new physics constraints,” *Journal of High Energy Physics*, vol. 2020, Aug 2020.
- [39] R.-X. Shi, L.-S. Geng, B. Grinstein, S. Jäger, and J. M. Camalich, “Revisiting the new-physics interpretation of the $b \rightarrow c \tau \nu$ data,” *Journal of High Energy Physics*, vol. 2019, Dec 2019.
- [40] M. Kobayashi and T. Maskawa, “CP-Violation in the Renormalizable Theory of Weak Interaction,” *Progress of Theoretical Physics*, vol. 49, pp. 652–657, 02 1973.
- [41] A. Wright, “Nobel prize 2008: Nambu, kobayashi and maskawa,” *Nature Physics*, vol. 4, 2008.
- [42] E. Kou, P. Urquijo, W. Altmannshofer, F. Beaujean, *et al.*, “The Belle II Physics Book,” *Progress of Theoretical and Experimental Physics*, vol. 2019, Dec 2019.

Bibliography

- [43] The Belle II Collaboration, “Belle II Confluence Luminosity.”
<https://confluence.desy.de/display/BI/Belle+II+Luminosity>, 2022.
- [44] U. Tamponi, “Charged Particle ID in Belle II.”
https://indico.belle2.org/event/493/contributions/4532/attachments/2841/4301/ChargedPID_StarterKit
- [45] S. Sandilya, “Particle Identification with the TOP and ARICH detectors at Belle II,” 2017. arXiv:1706.08515.
- [46] P. Urquijo and T. Ferber, “Overview of the Belle II Physics Generators,” Tech. Rep. BELLE2-NOTE-PH-2025-006 Draft Version 2.0, Jan 2016.
- [47] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, “The Belle II Core Software,” *Computing and Software for Big Science*, vol. 3, Nov 2018.
- [48] I. Antcheva, M. Ballintijn, B. Bellenot, M. Biskup, R. Brun, *et al.*, “ROOT - a c++ framework for petabyte data storage, statistical analysis and visualization,” *Computer Physics Communications*, vol. 180, pp. 2499–2512, dec 2009.
- [49] D. J. Lange, “The evtgen particle decay simulation package,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 462, no. 1, pp. 152–155, 2001. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.
- [50] S. Agostinelli *et al.*, “Geant4—a simulation toolkit,” *Nucl. Instrum. Meth. A*, vol. ”506”, pp. 250–303, 2003.
- [51] T. Keck, F. Abudinén, F. U. Bernlochner, R. Cheaib, *et al.*, “The Full Event Interpretation,” *Computing and Software for Big Science*, vol. 3, Feb 2019.
- [52] D. M. Asner, M. Athanas, D. W. Bliss, *et al.*, “Search for exclusive charmless hadronic b decays,” *Phys. Rev. D*, vol. 53, pp. 1039–1050, Feb 1996.
- [53] G. C. Fox and S. Wolfram, “Observables for the analysis of event shapes in e^+e^- annihilation and other processes,” *Phys. Rev. Lett.*, vol. 41, pp. 1581–1585, Dec 1978.

Bibliography

- [54] M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, *et al.*, “A Hierarchical NeuroBayes-based Algorithm for Full Reconstruction of B Mesons at B Factories,” *Nucl. Instrum. Meth. A*, vol. 654, pp. 432–440, 2011.
- [55] “Belle II Software Documentation.” <https://software.belle2.org/sphinx/release-06-00-03/>, 2022.
- [56] The Belle II Collaboration, “A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \rightarrow X \ell \nu$ decays,” 2020.
- [57] X. Zhou, S. Du, G. Li, and C. Shen, “TopoAna: A generic tool for the event type analysis of inclusive Monte-Carlo samples in high energy physics experiments,” *Computer Physics Communications*, vol. 258, p. 107540, Jan 2021.
- [58] F. U. Bernlochner, M. F. Sevilla, D. J. Robinson, and G. Wormser, “Semitauonic b -hadron decays: A lepton flavor universality laboratory,” Jan 2021.
- [59] I. Tsaklidis, P. Goldenzweig, I. Ripp-Baudot, J. Kahn, and G. Dujany, *Demonstrating learned particle decay reconstruction using Graph Neural Networks at Belle II*. PhD thesis, Strasbourg, Université de Strasbourg, Karlsruhe, Strasbourg, 2020. Presented on 19 06 2020.
- [60] J. P. Lees, V. Poireau, V. Tisserand, E. Grauges, A. Palano, *et al.*, “Measurement of an excess of $\bar{B} \rightarrow d^{(*)} \tau \bar{\nu}_\tau$ decays and implications for charged higgs bosons,” *Physical Review D*, vol. 88, Oct 2013.
- [61] W. D. Hulsbergen, “Decay chain fitting with a kalman filter,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 552, pp. 566–575, Nov 2005.
- [62] P. Koppenburg, “Statistical biases in measurements with multiple candidates,” 2019.
- [63] W. Sutcliffe, C. Lyu, H. Junkerkalefeld, M. Eliachevitch, and S. Duell, “PID-var: Offline Particle ID efficiency and fake rate corrections for Belle II.” <https://stash.desy.de/users/sutclw/repos/pidvar/>, 2019.

Bibliography

- [64] S. Godfrey and K. Moats, “Properties of excited charm and charm-strange mesons,” *Physical Review D*, vol. 93, Feb 2016.
- [65] A. P. Bradley, “The use of the area under the roc curve in the evaluation of machine learning algorithms,” *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [66] T. Keck, “FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification,” 2016. arXiv:1609.06119.
- [67] The Belle II Collaboration, “Muon and electron identification efficiencies and hadron-lepton mis-identification rates at Belle II for Moriond 2022,” Tech. Rep. BELLE2-CONF-PH-2022-003 Draft Version 2.0, March 2022.
- [68] I. Asimov, *Franchise*, in *Isaac Asimov: The Complete Stories, Vol. 1*. Broadway Books, 1990.
- [69] C. Cecchi, G. D. Nardo, E. Manoni, and M. Merola, “B counting in Moriond 2022/ICHEP 2022 Dataset,” Tech. Rep. BELLE2-NOTE-PH-2022-007 Draft Version 1.0, June 2022.
- [70] D. Ferlewicz, P. Urquijo, and E. Waheed, “Revisiting fits to $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ to measure $|V_{cb}|$ with novel methods and preliminary LQCD data at nonzero recoil,” *Physical Review D*, vol. 103, Apr 2021.
- [71] A. K. Leibovich, Z. Ligeti, I. W. Stewart, and M. B. Wise, “Predictions for $B \rightarrow D_1(2420) \ell \bar{\nu}$ and $B \rightarrow D_2^*(2460) \ell \bar{\nu}$ at Order $\Lambda_{\text{QCD}}/m_{c,b}$,” *Phys. Rev. Lett.*, vol. 78, pp. 3995–3998, May 1997.