
Measurement of the Branching Fractions and q^2 Spectra of $B \rightarrow D^{**} \ell \nu$ Decays at Belle II

Eylül Ünlü



Master's Thesis
at the Faculty of Physics
Ludwig-Maximilians-Universität München
Chair of Elementary Particle Physics

Supervisors:
Prof. Dr. Thomas Kuhr
Dr. Thomas Lück

Munich, September 15, 2023

Messung des
Verzweigungsverhältnisses und des q^2
Spektrum von $B \rightarrow D^{**} \ell \nu$ Zerfällen
mit Belle II
Eylül Ünlü



Masterarbeit
an der Fakultät für Physik
Ludwig-Maximilians-Universität München
Lehrstuhl für Experimentelle Teilchenphysik

Betreuer:
Prof. Dr. Thomas Kuhr
Dr. Thomas Lück

München, den 15. September, 2023

Abstract

Previous experimental and theoretical work on semileptonic B meson decays into the four P-wave excited charm mesons $D_1(2420)$, $D_2^*(2460)$, $D_0^*(2300)$ and $D_1'(2430)$ resulted in several unresolved problems. Experimental results contradict the theory predictions that the decays into the narrow states should be dominant over the decays into the wider states. Moreover, there has been some tension between the most recent branching fraction measurements by BaBar and Belle, where Belle observed no semileptonic B decays into the broad D_1' state. Improved understanding of the $B \rightarrow D^{**}\ell\nu$ decays would significantly reduce the systematic uncertainty on the $R(D^{(*)}) = \mathcal{B}(B \rightarrow D^{(*)}\tau\nu)/\mathcal{B}(B \rightarrow D^{(*)}\ell\nu)_{\ell=e,\mu}$ measurement, which hints at a deviation from the Standard Model. This study aims to use simulation from the Belle II experiment to measure the branching fractions of these decays and study the four-momentum transfer (q^2) which is important for comparison to theory. Furthermore, the sensitivity of higher luminosity data samples which Belle II targets to collect in the next decade to the q^2 study was investigated. A hadronic tagging approach was adopted, which allows the tag B momentum to be reconstructed. The fit yields were obtained by an unbinned maximum likelihood fit. For the q^2 study the same fit was implemented in bins of the parameter w , and fitted by the differential decay rate model $d\Gamma/dw$.

Contents

Abstract	v
1 Introduction	1
1.1 Motivation	1
1.1.1 Analysis Idea	3
1.2 The Belle II Experiment	4
2 Theory Background	7
3 Data and Simulation	11
4 Reconstruction	13
4.1 Signal Background Classification	13
4.2 Signal Side	14
4.2.1 Lepton Selection	15
4.2.2 D^{**} Reconstruction	18
4.3 Tag Side	22
4.3.1 Full Event Interpretation (FEI)	23
4.3.2 Other Selection Cuts	23
4.4 Event Selection	24
4.4.1 Missing Energy and Momentum	24
4.4.2 Rest of Event	25
4.5 Best Candidate Selection	26
4.6 Cut Optimisation	27
4.7 Summary and Efficiency	28
4.7.1 Reconstruction Efficiency	29
5 Fit	33
5.1 Fit Procedure	33
5.1.1 Kernel Density Estimation	34
5.1.2 Extended Maximum Likelihood Fit	36
5.2 Fit Results	37
5.3 Fit Validation	38

5.4	Branching Fraction Estimation	39
6	Four Momentum Transfer (q^2) Study	41
6.1	Unfolding	43
6.1.1	Migration Matrix	44
6.1.2	Unfolding Result	47
6.2	Decay Rate Fit	48
6.3	Extrapolation	49
7	Systematic Uncertainties	53
7.1	PDF Estimation Uncertainty	53
7.2	Migration Matrix Estimation Uncertainty	53
7.3	Data MC Correction Uncertainties	57
7.3.1	Slow π Reconstruction	57
7.3.2	Mid to High Momentum Tracking Efficiency	57
7.3.3	LeptonID	59
7.3.4	FEI Calibration	59
7.4	Other Uncertainties	60
7.4.1	Fit Bias	60
7.4.2	$\Upsilon(4S)$, D^{*+} and D^0 Branching Fractions	61
7.4.3	Decay Model Parameters	62
8	Conclusion and Discussion	65
A	Diagrams	69
	Acknowledgements	81
	Declaration	82

Chapter 1

Introduction

1.1 Motivation

The Standard Model (SM) of Particle Physics, despite being supported by numerous experimental findings, still fails to explain some of our observations of the universe. The matter-antimatter asymmetry, the nature of dark matter, and neutrino masses are examples of phenomena that the SM has not accounted for thus far. Current experimental efforts aim at obtaining more precise measurements of the SM parameters and observing possible deviations from the theory predictions. These deviations could hint at New Physics and shed light on unexplained phenomena. In particular, semileptonic B meson decays provide an opportunity to study the structure of weak currents. These decays occur abundantly at tree level with any higher level contributions highly suppressed. This makes it possible to use them to make precise measurements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix elements, V_{cb} and V_{ub} , which are related to the amount of charge-parity (CP) symmetry violation in the SM [54, p. 261]. Moreover, recent measurements of these decays have hinted at the violation of lepton universality [9, p. 183] which is a property of the SM.

Semileptonic B decays involving the $b \rightarrow c$ transition at quark level give rise to numerous hadronic final states. The dominant decays involve either D or D^* mesons. However, there are other resonant and non-resonant final states involving $D^{(*)}$ alongside lighter hadrons. The focus of this analysis are the semileptonic decays into so-called D^{**} states. These are P-wave excited mesons involving one charm and one lighter quark and have been observed to decay into $D^{(*)}$ along with one or two pions. They are classified into two doublets with light quark angular momentum $3/2$ and $1/2$ according to the Heavy Quark Effective Theory (HQET) (see Ch. 2): $\{D_1(2420), D_2^*(2460)\}$, and $\{D_0^*(2300), D_1'(2430)\}$. The elements of the first doublet have widths of order 10 MeV. Whereas, elements of the second are much wider with widths of order 100 MeV, which poses a challenge to observing their resonances.

The not fully understood $B \rightarrow D^{**}\ell\nu$ decays are a background to the measurements of $B \rightarrow D^*\tau\nu$ decays, which makes their modelling a significant systematic uncertainty on the $R(D^{(*)})$ measurement [41, 33, 1]. $R(D^{(*)}) = \mathcal{B}(B \rightarrow D^{(*)}\tau\nu) / \mathcal{B}(B \rightarrow D^{(*)}\ell\nu)_{\ell=e,\mu}$ is a test of lepton universality. The current world average of combined measurements shows 3.3σ devia-

tion from the SM prediction [9, p. 186-187], and any improvement in the uncertainties of future measurements of this value would help clarify this discrepancy.

There have been improvements in the understanding of these decays in the last thirty years, both on theoretical and experimental fronts. Theoretically, the form factors have been calculated up to leading order in the heavy quark expansion ($m_{c,b} \rightarrow \infty$) [43]. The order $\Lambda_{QCD}/m_{c,b}$ corrections to the decay rates have been shown to be significant, especially in the zero recoil case. Additionally, higher order corrections, along with order α_s corrections have been calculated [18]. Experimentally, various observations of these decays were reported by experiments at DESY [7], CESR [10], LEP [3, 4] and Tevatron [2]. More recently in 2008, B factories Belle and BaBar have both measured the branching fractions of these decays. They obtained significantly different results for the decays into the wide D'_1 meson. The BaBar experiment observed D'_1 decays in a 417 fb^{-1} data sample, and reported a branching fraction of $(0.27 \pm 0.04 \pm 0.05) \%$ [11]. Whereas the Belle experiment observed no such decays in a sample of $657 \times 10^6 \text{ } B\bar{B}$ events and only reported an upper limit on the branching fraction [44]. The Belle result was also in contradiction with a previous 2005 DELPHI analysis which found the D'_1 mode to be the dominant contribution to $B \rightarrow D^{**}\ell\nu$ decays [4].

Additionally, there is an ongoing tension between the experimental results and the theory predictions. From QCD sum rules [52] and quark models, an abundance of the narrow $j^P = 3/2^+$ over the wider $j^P = 1/2^+$ states are expected¹. This has been contradicted by experimental measurements so far, and this is known as the "1/2 vs 3/2 Puzzle" [20]. Moreover, the inclusive branching fraction, $\mathcal{B}(B \rightarrow X_c \ell \nu)$ is not fully saturated by decays into $D^{(*)}$ and the D^{**} states. Since theory predictions show that the broad D^{**} states can not account for the remaining branching fraction, the question on the nature of these so-called gap modes persists.

Further investigations of exclusive $B \rightarrow D^{**}\ell\nu$ decays are needed to understand the obscurities described above. Studying the decays into different D^{**} modes separately gives additional insight into the current tension with theory². The Belle II experiment, with the high integrated luminosity goal of 50 ab^{-1} provides an auspicious opportunity to investigate these decays. The aim of this master's thesis is to reconstruct these decays in the $D^{**} \rightarrow D^{*+}\pi^-$ channel, in order to measure the Branching Fractions of D_1 , D_2^* , and D'_1 , and study the q^2 spectra which can be compared to theory. Additionally, an interesting question is how well these decays may be studied with future Belle II data sets. To address this question, the q^2 study, which is performed on a 400 fb^{-1} MC simulation sample, is subsequently extrapolated to higher luminosities.

The rest of this chapter summarises the plan for the analysis and introduces the Belle II experiment, Chapter 2 goes through the necessary theory background, Chapter 3 gives some details on the MC simulation used, Chapter 4 describes the reconstruction and selection of decays, in Chapter 5 an unbinned extended maximum likelihood fit is implemented to extract the signal yields, the q^2 study and the extrapolation are described in Chapter 6, and possible sources of System-

¹Note that the theoretical predictions refer to the *decay amplitudes*. The semileptonic decays into the broad D^{**} states are also kinematically suppressed by a factor of two to three [20]. Hence even if the branching fractions into broad and narrow states are equal, this points to an enhanced decay amplitude for the broad states, which contradicts the theory prediction.

²Additionally, it was advised to study the D^{**} decays into $D^*\pi$ and $D\pi$ separately since this provides more insight for the 1/2 vs 3/2 puzzle [20].

atic uncertainties are discussed in Chapter 7. Finally, the conclusion and discussion of results is provided in Chapter 8.

1.1.1 Analysis Idea

As mentioned in the motivation, the aim of the analysis is to measure the branching fractions of $B^- \rightarrow D^{**}(D^{*+}\pi^-)\ell^-\nu$ decays separately for the D_1 , D_2 , and D'_1 modes, and study their four-momentum transfer (q^2) distributions. The study was conducted on 400 fb^{-1} MC simulation, which roughly corresponds to the luminosity collected by Belle II at the time of the study.

The analysis idea is to exclusively reconstruct one of the B mesons in this channel as illustrated in Fig. 1.1, while the other B meson is reconstructed in hadronic decay channels using the Full Event Interpretation (FEI) algorithm (Subsection 4.3.1). After a series of selections, the yields of signal decays are determined by a maximum likelihood fit in the $\Delta M_{D^{**}} = M_{D^{**}} - M_{D^*}$ distribution. For the q^2 study, the w distribution is investigated for comparability to theory. The $\Delta M_{D^{**}}$ fit is repeated in bins of w to obtain the shape of the w distribution. The reconstructed distribution is corrected for resolution effects by unfolding. The obtained distribution is fitted by the decay rate model by a χ^2 fit. Due to lack of statistics in the D_2^* mode, the q^2 study is limited to the D_1 mode. As a test of the sensitivity of future Belle II data sets, the decay rate fit is repeated on generated distributions corresponding to luminosities of 1 ab^{-1} , 5 ab^{-1} , and 50 ab^{-1} .

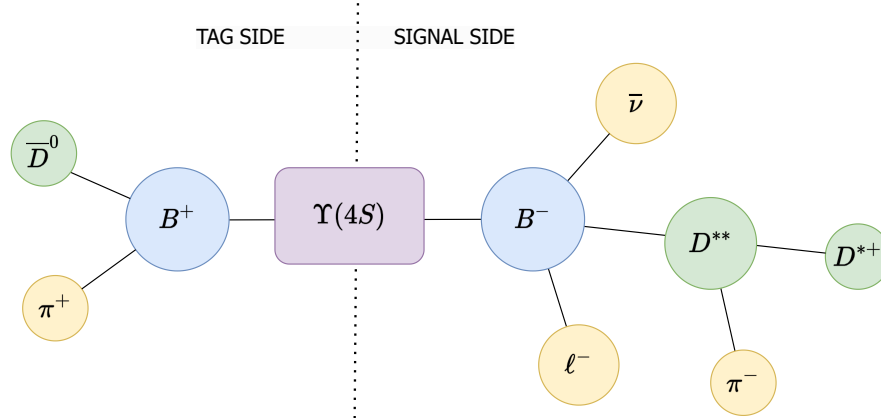


Figure 1.1: Diagram showing how the $\Upsilon(4S)$ is reconstructed for the analysis. The decay channel drawn on the tag side is one example of the many decay channels considered.

1.2 The Belle II Experiment

SuperKEKB

SuperKEKB [28] is a circular electron-positron collider, situated in Tsukuba, Japan. For the majority of its operation time it collides e^-e^+ pairs at a center of mass energy corresponding to the $\Upsilon(4S)$ resonance energy. $\Upsilon(4S)$ decays into a pair of B mesons with a branching fraction larger than 96%, which makes it possible to study B mesons with low background from other decays. The beam energies are asymmetric, with the electron beam at roughly 4 GeV and the positron beam at 7 GeV energy. This gives a boost to the center of mass frame of the experiment, to allow more precise lifetime measurements.

SuperKEKB is an upgrade to KEKB based on the "nano-beam" scheme [23]. The design focuses the particle beams to nanometer scale at the interaction point, by means of strong magnets. With this design feature, SuperKEKB is targeted to reach a luminosity of $8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$.

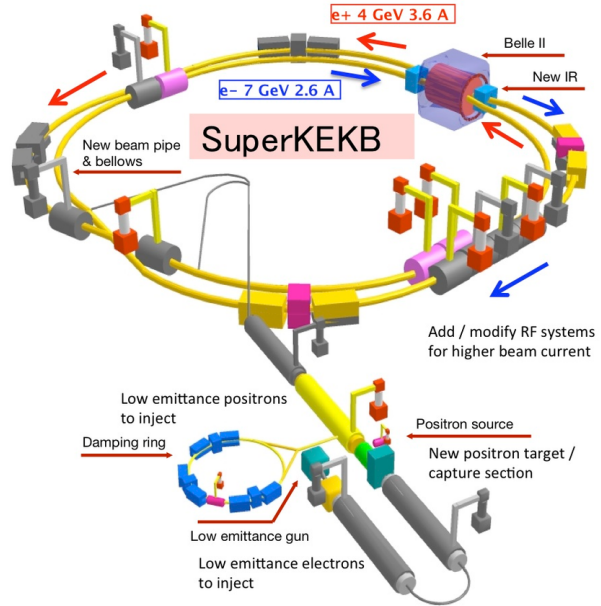


Figure 1.2: Schematic view of SuperKEKB [50].

The Belle II Detector

The Belle II detector is a cylindrical, general purpose spectrometer placed around the e^-e^+ interaction point at SuperKEKB. It is an upgrade from the predecessor Belle, and has been designed to operate at 40 times higher event rates, and 10 to 20 times higher background rates [8]. The part of the detector in the direction of the center of mass boost is referred to as the *forward region*, the opposite end of the detector is the *backward region*, whereas the region transverse to the beam axis is called the *barrel region*. Due to the boost, more particles are expected to be detected in the forward region. The coordinate system of the experiment is set with the z -axis corresponding to the beam axis and with the electron beam directed in the positive z -axis.

The detector is made up of several components. These are listed below in order of proximity to the beam pipe where the collision takes place, and visualised on Fig. 1.3. The following information is based on Ref. [29] where further details may be found.

- The component closest to the interaction point is the **Vertex Detector (VXD)**. It is made up of two layers of the **Pixel Detector (PXD)** with pixelated sensors of the DEPFET type, and four layers of the **Silicon Vertex Detector (SVD)**, equipped with double sided silicon strip sensors. Placing the PXD close to the interaction point where there is a high density of

particles allows precise position measurements in this region. Whereas using strips instead of pixels on the other layers cuts down on the data size.

- The **Central Drift Chamber (CDC)** is used for the detection of charged tracks, as well as measurements of their momenta and energy losses. It is filled with $\text{He-C}_2\text{H}_6$ gas which is ionised with the passing of charged particles through it. The freed electrons drift to the sense wires where the signal is saved.
- The **Time-of-Propagation (TOP)** and **Aerogel Ring Cherenkov (ARICH)** detectors are used for particle identification in the barrel and forward regions respectively. They make use of the Cherenkov effect to measure particle velocities. From this and the momentum information, the mass of the particle is derived.
- The **Electromagnetic Calorimeter (ECL)** is where the energies and angular coordinates of photons are measured. It is also used for the identification of electromagnetically (EM) interacting particles such as electrons. It uses CsI(Tl) crystals as scintillators. EM interacting particles that traverse the sub-detector deposit energy in the form of photon showers in the ECL.
- The outermost layer is dedicated to detecting the long-lived particles which are not stopped in the ECL. **The K-long and Muon detector (KLM)** is made up of alternating layers of iron and detector material. The detector material is comprised of scintillators to detect the shower is produced at the iron layers.

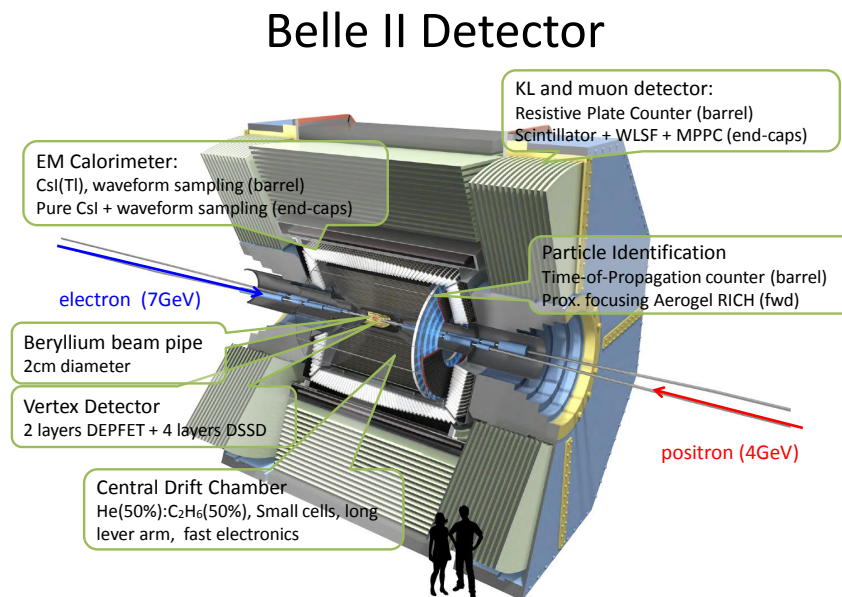


Figure 1.3: 3D model of the Belle II detector with labels indicating the different components [14].

Chapter 2

Theory Background

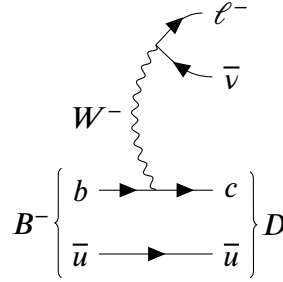


Figure 2.1: Feynman diagram for the semileptonic B meson decay into a D meson

The decays of the B meson into charmed mesons occur through a $V - A$ weak current in the Standard Model. Hence, the decay rates are obtained by calculating the matrix elements of the vector and axial-vector elements between the B and D^1 states:

$$\langle D(v') | V_\mu | B(v) \rangle \text{ and } \langle D(v') | A_\mu | B(v) \rangle \quad (2.1)$$

Here v and v' denote the four-velocities of the mesons. The matrix elements are generally parameterised by functions called *form factors*, which depend on the Lorentz invariant quantity of the four-momentum transfer squared (q^2) (see Eqn. 2.2).

$$q^2 = (p_B - p_D)^2 \quad (2.2)$$

The explicit calculation of the matrix elements is simplified by working in the heavy quark limit ($m_Q \rightarrow \infty$), which is a good approximation for mesons involving one heavy quark ², e.g. B and D . In this limit, the light quark in the meson *sees* the heavy quark field as a static colour source and the light quark field is no longer dependent on the spin or flavour of the heavy quark.

¹ D is used to denote any meson state with a charm quark.

² The term heavy quark describes quarks with mass $m_Q \gg \Lambda_{QCD}$, where $\Lambda_{QCD} \sim 0.2 \text{ GeV}$ is the QCD scale parameter.

[38] This establishes an $SU(2N_h)$ symmetry of the flavour and spin of the heavy quark, where N_h is the number of heavy quark flavours [46].

However, the light quark field is still dependent on the four-velocity of the heavy quark³, and the form factors also depend on the four-velocities of the mesons involved. It turns out that in this limit, all form factors can be represented by a single function of the dimensionless w parameter, defined as in equation 2.3. This function is called the *Isgur-Wise function* $\tau(w)$ [34], and represents the overlap between the states of the light quarks in two mesons with four-momenta v and v' [38].

$$w = v_B \cdot v_D = \frac{m_B^2 + m_D^2 - q^2}{2m_B m_D} \quad (2.3)$$

In the heavy quark limit, some of the form factors vanish at zero recoil ($w = 1$), i.e. when D is at rest in the B rest frame. Hence the $\mathcal{O}(\Lambda_{QCD}/m_{b,c})$ corrections become significant. Since kinematically allowed values of $B \rightarrow D^{**}\ell\nu$ decays ($1 < w \lesssim 1.32$) is close to 1, these corrections are especially significant for the decays into D^{**} mesons. The detailed calculation of these corrections is given in the LLSW study [43] (1998). Here, the Isgur-Wise function, and the decay rates are expanded in powers of $(w - 1)$, as given in equation 2.4. The value and slope of the Isgur-Wise function at zero recoil, $\tau(1)$ and $\tau' = d\tau/dw$ are parameters which need to be measured. Additionally the $\mathcal{O}(\Lambda_{QCD}/m_{b,c})$ corrections break the spin-flavour symmetry and introduce other Isgur-Wise functions, e.g. $\tau_{1,2}$ and $\eta_{1,2,3}$ [18].

$$\tau(w) = \tau(1)(1 + \tau'(w - 1) + \mathcal{O}((w - 1)^2)) \quad (2.4)$$

The LLSW study states the differential decay rates of the $B \rightarrow D^{**}\ell\nu$ decays in w and an angular coordinate. The more recent BLR study [18] (2018) integrates the result over the angular coordinate, and reports the differential decay in w , denoted $d\Gamma/dw$. The decay rate (Eqn.2.6) and the form factors (Eqn. 2.5) stated below are taken from this reference.

³This is equal to the four-velocity of the meson at the heavy quark limit.

$$\begin{aligned}
f_{V_1} &= \frac{1}{\sqrt{6}} \tau(w) \left((1 - w^2) - \epsilon_b(w^2 - 1) \hat{F}_b \right. \\
&\quad \left. - \epsilon_c \left[4(w + 1)(w\bar{\Lambda}' - \bar{\Lambda}) - (w^2 - 1)(3\hat{\tau}_1 - 3\hat{\tau}_2 + 2\hat{\eta}_1 + 2\hat{\eta}_3) \right] \right) \\
f_{V_2} &= \frac{1}{\sqrt{6}} \tau(w) \left(-3 - \epsilon_b 3\hat{F}_b - \epsilon_c \left[(4w - 1)\hat{\tau}_1 + 5\hat{\tau}_2 + 10\hat{\eta}_1 + 4(w - 1)\hat{\eta}_2 - 5\hat{\eta}_3 \right] \right) \\
f_{V_3} &= \frac{1}{\sqrt{6}} \tau(w) \left(w - 2 + \epsilon_b(2 + w)\hat{F}_b + \epsilon_c \left[4(w\bar{\Lambda}' - \bar{\Lambda}) + (2 + w)\hat{\tau}_1 + (2 + 3w)\hat{\tau}_2 \right. \right. \\
&\quad \left. \left. - 2(6 + w)\hat{\eta}_1 - 4(w - 1)\hat{\eta}_2 - (3w - 2)\hat{\eta}_3 \right] \right) \\
f_{V_A} &= \frac{1}{\sqrt{6}} \tau(w) \left(-(w + 1) - \epsilon_b(w - 1)\hat{F}_b \right. \\
&\quad \left. - \epsilon_c \left[4(w\bar{\Lambda}' - \bar{\Lambda}) - 3(w - 1)(\hat{\tau}_1 - \hat{\tau}_2) - (w + 1)(2\hat{\eta}_1 + 3\hat{\eta}_3) \right] \right)^{[18]}
\end{aligned} \tag{2.5}$$

$$\text{where } \hat{F}_b = \bar{\Lambda} + \bar{\Lambda}' - (2w + 1)\hat{\tau}_1 - \hat{\tau}_2 \quad \epsilon_{b,c} = \frac{1}{2m_{b,c}}$$

Here $\hat{\tau}_{1,2}(\hat{\eta}_{1,2,3}) = \tau_{1,2}(\eta_{1,2,3})/\tau(w)$, where $\tau_{1,2}(\eta_{1,2,3})$ are *dimensionful* Isgur-Wise functions added by the $\mathcal{O}(\Lambda_{QCD}/m_{b,c})$ corrections. $\bar{\Lambda}$, and $\bar{\Lambda}'$ denote the kinetic energy of the light degrees of freedom of the $j^P = (1/2)^-$ and $(3/2)^+$ doublets in the heavy quark expansion. The α_s corrections are omitted here. See Ref. [18] for these corrections. The Belle II MC uses an approximation of these form factors with $\eta_1 = \eta_2 = \eta_3 = 0$. Furthermore $\hat{\tau}_1$ and $\hat{\tau}_2$ are treated as constant parameters, which may be fitted. Therefore, this approximation is adopted when implementing the decay rate fit in Chapter 6.

$$\begin{aligned}
\frac{d\Gamma_{D_1}}{dw} = & 2\Gamma_0 r^3 \sqrt{w^2 - 1} \frac{(\hat{q}^2 - \rho_\ell)^2}{\hat{q}^6} \left\{ f_{V_1}^2 \left[2\hat{q}^2 [(w-r)^2 + 2\hat{q}^2] + \rho_\ell [4(w-r)^2 - \hat{q}^2] \right] \right. \\
& + (w^2 - 1) \left(f_{V_2}^2 \left[2r^2 \hat{q}^2 (w^2 - 1) + \rho_\ell [3\hat{q}^2 + 4r^2 (w^2 - 1)] \right] \right. \\
& + f_{V_3}^2 \left[2\hat{q}^2 (w^2 - 1) + \rho_\ell [4(w-r)^2 - \hat{q}^2] \right] \\
& + 2f_A^2 \hat{q}^2 (2\hat{q}^2 + \rho_\ell) \\
& + 2f_{V_1} f_{V_2} \left[2r\hat{q}^2 (w-r) + \rho_\ell (3 - r^2 - 2rw) \right] + 4f_{V_1} f_{V_3} (w-r) (\hat{q}^2 + 2\rho_\ell) \\
& \left. \left. + 2f_{V_2} f_{V_3} \left[2r\hat{q}^2 (w^2 - 1) + \rho_\ell [3w\hat{q}^2 + 4r(w^2 - 1)] \right] \right] \right) \left. \right\}^{[18]}
\end{aligned} \tag{2.6}$$

$$\text{where } \Gamma_0 = \frac{G_F^2 |V_{cb}|^2 m_B^5}{192\pi^3}, \quad r = \frac{m_{D^{**}}}{m_B}, \quad \rho_\ell = \frac{m_\ell^2}{m_B^2}, \quad \hat{q}^2 = \frac{q^2}{m_B^2}$$

Chapter 3

Data and Simulation

Data Taking

Belle II started data taking in 2019 and collected an integrated luminosity of 424 fb^{-1} ¹ before going on the Long Shutdown 1 (LS1) in summer 2022 [15]. This is the first long pause to the running of the experiment, for upgrades to the detector and the accelerator. Data taking is planned to resume in fall 2023.

A plot showing the weekly integrated luminosity collected by Belle II from January 2019 to August 2022 is given in Fig. 3.1. Among the data collected from good runs, 363 fb^{-1} were from collisions at the $\Upsilon(4S)$ resonance energy. The remaining 42 fb^{-1} were below, and 19 fb^{-1} were above this energy [15]. The data sets off the $\Upsilon(4S)$ energy serve various purposes such as hadronic physics or dark sector studies. Particularly, data taken slightly below the $\Upsilon(4S)$ resonance (approximately 60 MeV) is termed the *off-resonance* or *continuum* sample. It allows the study of $e^-e^+ \rightarrow q\bar{q}$ (continuum) background where q denotes quarks lighter than the b quark.

Simulation

The production of the Belle II Monte Carlo (MC) simulations starts with the simulation of physics processes by an event generator, and this is followed by the detector simulation [8]. Belle II uses EvtGen [40] for the simulation of B and D decays. EvtGen is a Monte Carlo event generator which is controlled by a decay table that sets all the possible decay processes, along with their branching fractions and the decay amplitude (model) used to simulate them. Other decays, which are not specified by the decay table are simulated by PYTHIA [49]. Continuum events are also simulated using PYTHIA and KKMC [35]. The final state radiation is added using PHOTOS [27]. Following this, the detector response to the passing of the particles is simulated using the Geant4 software package [6]. The simulation depends on the chosen physics model that describes the interaction of particles with the detector material. Belle II uses the recommendations for high energy physics by the Geant4 group to set this model [8].

At the time of writing, the generic MC simulation of Belle II includes non-resonant $B \rightarrow D^{*+}\pi^-\ell\nu$ decay with a branching fraction of 0.1%. Nevertheless, the current recommendation by

¹The number presented here is the luminosity collected from *good* runs.

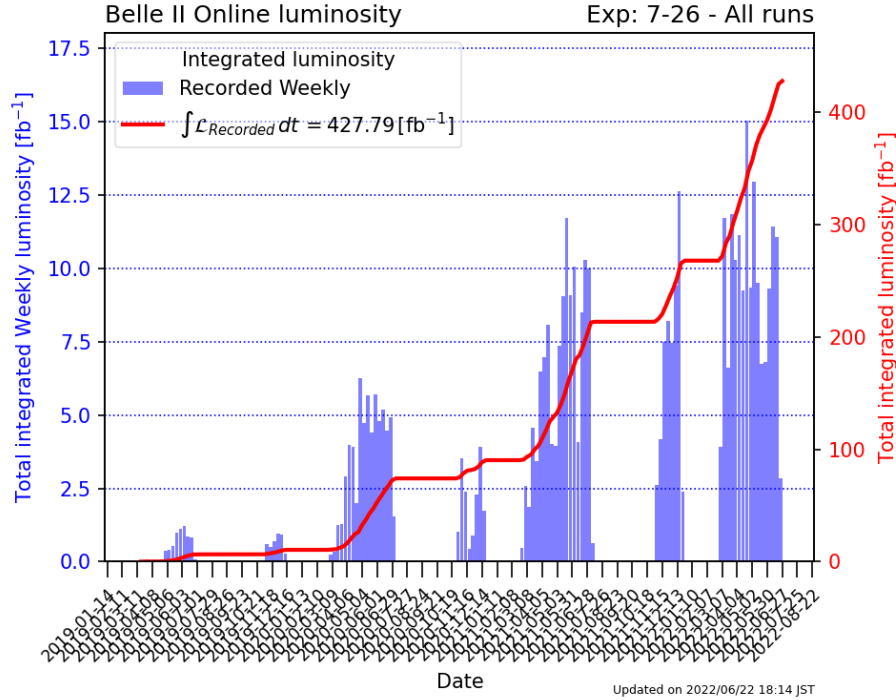


Figure 3.1: The weekly integrated luminosity collected (from *all* runs) at Belle II from 14/01/2019 to 22/08/2022 [15].

the collaboration is to replace these decays with other resonant decays. This is because consideration of past measurements predicts the branching fraction of this decay to be consistent with zero [17]. This correction is not implemented in the current analysis.

Analysis skims

The data or simulation is saved in the mini data summary table (mDST) format, after the initial processing, i.e. reconstruction of tracks and clusters. However, these samples include many events that may not be relevant for specific analyses. Hence it is sensible to subject these to further reconstruction or selection criteria before their use in an analysis. This is called *skimming*. The skimmed data or MC sets are saved in a user data summary table (uDST) format and are generally smaller in size and more efficiently used in analyses.

This current analysis uses skims from the run independent MC simulation ². The sample includes skims from 2.8 ab^{-1} on-resonance, and 1 ab^{-1} continuum simulation. The skim selects events where at least one B meson is reconstructed using the Full Event Interpretation algorithm (Subsec. 4.3.1) in hadronic channels. Some selection cuts are applied after the reconstruction, these include but are not limited to: Signal probability > 0.001 (Subsec. 4.3.1), $M_{bc} > 5.2 \text{ GeV}$, and $|\Delta E| < 0.3 \text{ GeV}$ (Subsec. 4.3.2).

²The campaign name is MC15ri_b, produced with basf2 release 06-01-12.

Chapter 4

Reconstruction

The reconstruction of full decay chains up to the $\Upsilon(4S)$ is carried out using the Belle II Analysis Framework (basf2) [51]. The signal B meson is reconstructed in the exclusive decay channel $B^- \rightarrow D^{*0} \ell^- \nu$, while the other B meson in the $\Upsilon(4S)$ decay, denoted the tag B meson, is reconstructed in hadronic channels using the Full Event Interpretation [37] algorithm. The details of reconstruction and selection will be detailed in this chapter. A summary of the final selection cuts is given in Section 4.7.

4.1 Signal Background Classification

Working with simulation allows one to obtain the true information on the type of particle and its full decay chain. Based on this information the values of the cuts applied are decided with the goal of eliminating as much background as possible while retaining as much signal as possible, in other words, optimising the chosen figure of merit. However, in order to count the true number of signal and background events in the selected sample, first the method of classifying events as such should be introduced.

There is a Monte Carlo truth-matching algorithm implemented in basf2. When applied to a reconstructed particle, it considers recursively all daughters of the particle, and attempts to match each reconstructed daughter to the generated particle (of the same type) which created the signature in the detector. If the matching is successful for all daughters along the decay chain, the particle is marked as *correctly reconstructed*. Otherwise, if the matching is unsuccessful for one of the daughters, e.g. a daughter particle is missing, or of the wrong type¹ the particle is marked as *incorrectly reconstructed*, and if no match is found for the initial particle (to which the truth matching is applied), the output is Not a Number (NaN). Initially, the results of this algorithm were used directly for signal and background classification. However, it was observed that for a significant number of reconstructed $\Upsilon(4S)$ events, the algorithm gave NaN as the result. To be concrete, $\sim 29\%$ of $\Upsilon(4S)$ candidates from a 400 fb^{-1} MC sample were marked as NaN. Upon observing that certain kinematic variables peaked in the signal region for these non-truth-

¹These are only two of the possible reasons why a match could be unsuccessful. The details may be found in Ref. [47]

matched events, and inspecting the generated decay chains corresponding to these candidates, the decision was made to include these events in the signal classification by implementing a custom signal-matching method.

The classification method is as follows:

1. For each reconstructed $Y(4S)$ candidate, write out the corresponding generated decay tree, **ignoring** any photons due to final state radiation.
2. Identify which of the daughter B mesons were reconstructed in the signal decay channel.
3. Check the PDG codes of the daughters of the identified B meson. Classify as background **unless** the following are all true ²:
 - B has three daughters.
 - The first daughter is one of D_1 , D_1' , D_2^* .
 - The second daughter is one of e , μ .
 - The third daughter is a ν .
4. For candidates **not** marked as **background** in the previous step, check the PDG codes of the daughters of the D^{**} meson. Classify as background **unless** the following are all true:
 - D^{**} has two daughters.
 - The first daughter is $D^{\pm*}$.
 - The second daughter is $\pi^\pm$.
5. Classify the $Y(4S)$ satisfying all the criteria in steps 3 and 4 as signal. Save the PDG code of the D^{**} for each candidate to further classify into the three signal D^{**} modes.

Therefore, the signal conditions do not include any criteria on the truth-matching of the daughters of the signal B beyond the $D^{\pm*}$ meson, or the tag B candidates. As a result, the reconstructed value of certain kinematic variables do not agree with the true values as well as they would if the entire $Y(4S)$ decay trees were truth-matched. On the other hand, this allows more of the generated $B \rightarrow D^{**}\ell\nu$ events to be included in the studied sample. For the aim of gathering enough statistics to study three D^{**} modes separately, the method with higher acceptance of signal events is favoured.

4.2 Signal Side

Signal B candidates were obtained by combining track candidates identified as leptons with reconstructed D^{**} mesons while making use of the missing kinematic information due to the single ν in the decay. The D^{**} reconstruction is limited to decays into a charged D^* meson and a charged

²The order of decay products are always the same in the MC generation. This allows one to only check the first daughter for the D^{**} candidate etc.

π . This narrows the focus of this study to three of the four D^{**} modes that decay into this channel. The D^* candidate decays into another charged π and a D^0 meson which was reconstructed in nine hadronic decay channels. The decay chain from the final state particles up to the signal B meson is illustrated on Figure 4.1. The reconstruction and selection of particles will be detailed in the following subsections.

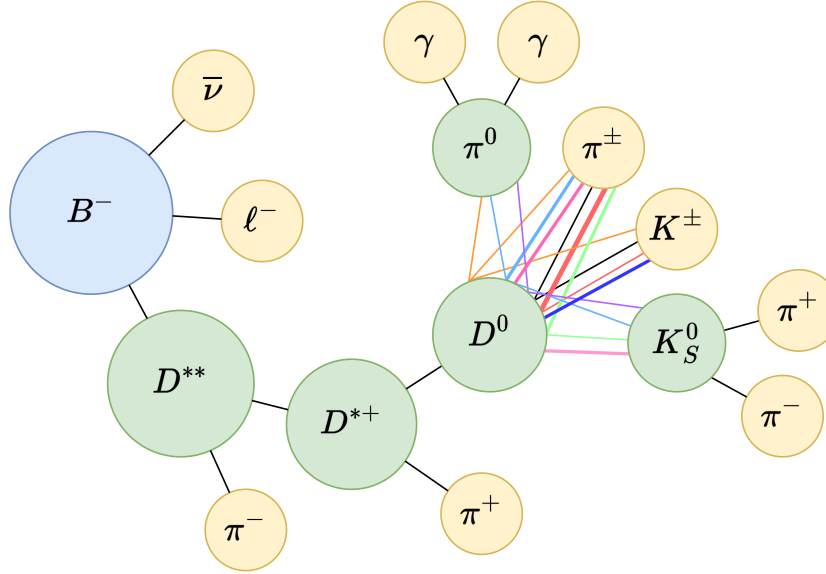


Figure 4.1: Signal B meson reconstruction diagram. The final state and intermediate particles are given in yellow and green respectively. The nine different decay channels of the D^0 meson are indicated by different coloured edges. The width of the edges scale with the number of times the given particle occurs in the decay products of a particular channel.

4.2.1 Lepton Selection

Light leptons (μ and e) are long-lived charged particles, as a result they traverse the VXD and CDC and leave hits in these parts of the detector. These hits are combined into track candidates using Belle II's track finding algorithm [19]. Tracks of charged particles under the solenoidal magnetic field take helical forms, and are described by a set of five parameters, some of which are introduced below. For each reconstructed track, the track parameters are extracted by fitting with the mass hypothesis closest to the mass of the particle of interest [21]. For the lepton candidates used in signal reconstruction, track candidates fitted using a pion mass hypothesis³ were considered, and further selections were applied.

Two derived track parameters⁴ of interest are dr and dz . dr is the signed distance between the point of closest approach (POCA) of a track to the measured interaction point (IP), and the z-axis.

³Currently, the e mass hypothesis is not implemented since the π fit performs better. Since the mass of the μ is close to that of the π , the μ hypothesis is not implemented separately to save resources.

⁴derived track parameters, e.g. dz , are defined w.r.t the measured IP, whereas true track parameters, e.g. z_0 are defined w.r.t the (0,0,0) point.

Whereas dz is the z coordinate of the POCA w.r.t the IP. Since the leptons in the signal decay are primary daughters of the B meson, they are expected to come from within $\sim 130 \mu\text{m}^5$ radius of the IP. [19]. Hence loose cuts of $|dr| < 0.5 \text{ cm}$ and $|dz| < 2 \text{ cm}$ were applied. Additionally, tracks are required to be in angular acceptance of the CDC, i.e. their polar angle, which is calculated from the initial momentum of the particle, is required to be $17^\circ < \theta < 150^\circ$. These track cuts were selected based on internal recommendations.

The momentum of the lepton in the centre of mass frame is another variable of interest. A lower cut on this variable eliminates some of the low momentum leptons from background physics decays, such as leptons from secondary decays. A preliminary cut was placed at $p_l^{CMS} > 0.6 \text{ GeV}/c$, motivated by the 2008 BaBar analysis [11], but this cut was later optimised as described in Section 4.6. On the plot in Fig. 4.6, it is observed that a cut on the lower momentum leptons eliminates more background than signal events. Especially for decays reconstructed in the muon channel, background events are more abundant in the low momentum region. This background could be due to misidentified pions, since pions are more likely to be misidentified as muons. Indeed, checking the signal matching for the decays reconstructed in the $B \rightarrow D^{**}\mu\nu$ channel shows that 42.5% of the background was due to fake leptons. Whereas the same figure for the decays reconstructed in the $B \rightarrow D^{**}e\nu$ channel is 7.1%.

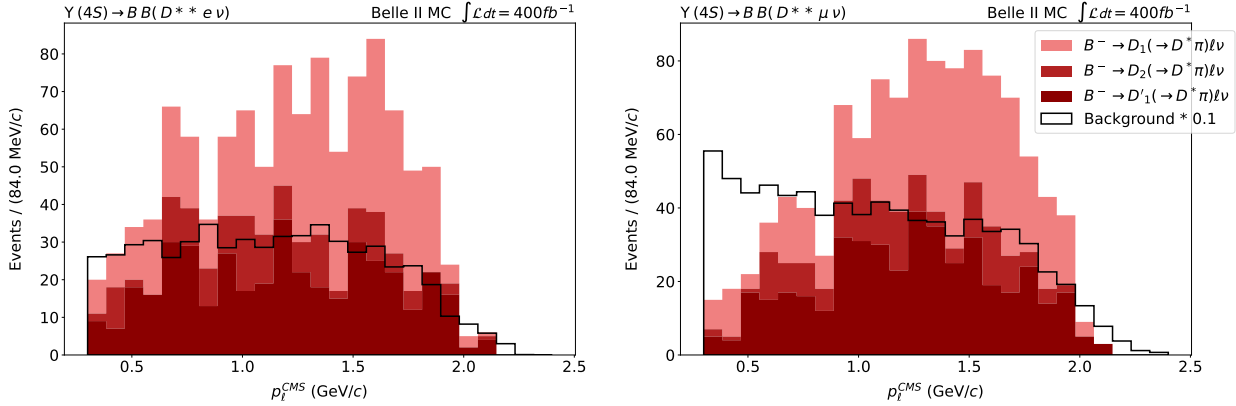


Figure 4.2: Distribution of lepton momentum in the center of mass frame for the electron (left) and muon (right) decay modes, after reconstruction up to $Y(4S)$ events (after leptonID cuts have been applied with no Bremsstrahlung correction). Three signal categories are shown as a stacked histogram and all other decays (background) are added as an unstacked and unfilled histogram.

As charged particles traverse the detector medium they decelerate and emit Bremsstrahlung radiation. The power radiated scales with either $(E/mc^2)^4$ when the acceleration is perpendicular to the velocity of the particle, or with $(E/mc^2)^6$ when it is parallel. Hence, given two particles with the same energy, the lighter one emits more Bremsstrahlung radiation. As a result, this effect is significant for electrons but not for muons (the power emitted by muons is around 200^{-6}

⁵ $130 \mu\text{m}$ is the average displacement of the B meson decay vertices from the IP in the z -direction. The displacement in the other directions is expected to be far lower.

electron p_T (GeV/c)	photon E cut (GeV/c ²)	angle threshold (rad)
$0.3 < p_T \leq 0.6$	< 0.09	0.1368
$0.6 < p_T \leq 1$	< 0.9	0.0737
$1 < p_T$	< 1.2	0.0632

Table 4.1: Cut Values for the Bremsstrahlung correction

to 200^{-4} of that emitted by the electrons), and only the momentum of the electrons need to be corrected for the momentum lost through this radiation. The power radiated is maximal along the direction of motion of the particle. Therefore, only photons with a direction of motion within a cone around the initial electron momentum are considered. The four-momenta of the selected photons are added to the four-momentum of the electron candidates to implement the correction. Following internal recommendations, the angle of this cone (threshold angle) and the energy cuts applied on the selected photons are set unevenly, based on the transverse momentum of the electron. The exact values are given in Tab. 4.1.

Additionally, particle identification (PID) variables are considered. The aim is to quantify the probability that the detected particle is the particle of interest, given the observation of a track in the detector. This can then be used to apply cuts on observed tracks. Through Bayesian logic, this probability scales with the probability that the track observed belongs to the particle of interest, in other words the likelihood of this particle hypothesis. Hence PID variables are defined by this likelihood. The likelihoods are obtained separately for each detector part, and these are multiplied to give the total likelihood of the particle hypothesis.

So-called global PID variables represent the relative likelihood of the hypothesized particle in comparison to the likelihood of the five other particle hypotheses considered at Belle II [13].

$$PID_p = \frac{\mathcal{L}_p}{\mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_p + \mathcal{L}_d} \quad (4.1)$$

Based on recommendations from the collaboration, the likelihood-based PID variable which uses no SVD information was used to select muons. A cut was placed at > 0.9 .

Another method of quantifying the particle identification information from the detector is to use a Boosted Decision Tree (BDT). The BDT is trained on the forementioned likelihood information from the detectors, as well as showershape parameters from the ECL. The development of showers is expected to differ between EM-interacting electrons, minimally ionising particles such as muons, and hadronic particles such as pions. Hadrons, in addition to being minimally ionising, give off hadronic split-offs. This refers to the creation of secondary particles as a hadron interacts with the detector material. Considering showershape parameters exploits these differences for particle identification. For electron selection the BDT-based PID variable is preferred, based on performance evaluations by the dedicated subgroup. As for the muon candidates, a cut is applied at > 0.9

4.2.2 D^{**} Reconstruction

The D^{**} mesons were reconstructed only in the $D^{**} \rightarrow D^{*+}\pi^-$ channel. First, the D^0 meson candidates were reconstructed in nine decay channels. Next, they were combined with a charged π candidate and subsequently a second. The reconstruction of D^0 meson will be introduced first.

D^0 Reconstruction

The nine hadronic decay channels included in the D^0 reconstruction give a total branching fraction of 27.2%. These decay channels are listed in Tab. 4.2. Limiting the analysis to only hadronic decay channels of D^0 ensures there is only one neutrino in the final decay products (coming from the signal B decay). This feature is employed during signal selection, as described later in this section.

Cuts applied on the track parameters to the charged kaon and pion candidates are consistent with those applied to the lepton candidates. The neutral K_S^0 and π_0 candidates need to be reconstructed from final state particles as they generally decay inside the detector or the beam pipe.

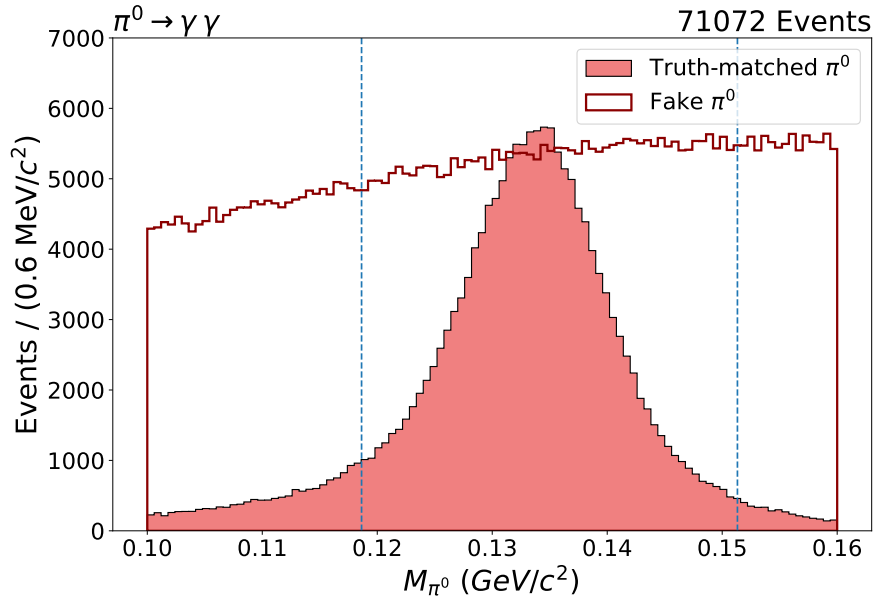


Figure 4.3: The mass distribution for neutral pions reconstructed in the $\pi \rightarrow \gamma\gamma$ channel. The distributions for correctly reconstructed candidates and the rest are given as two unstacked histograms. The cuts applied during the selection are given by the dashed lines.

π^0 Reconstruction: π^0 decays into a pair of photons with a branching fraction of 98.8%. For the reconstruction, two photons were selected with an invariant mass of $0.118 \text{ GeV} < M(2\gamma) < 0.152 \text{ GeV}$ (see Figure 4.3). The value of this cut was determined by fitting a Gaussian to the

mass distribution of reconstructed and truth-matched π_0 candidates (see Figure A.1). Lower and upper cuts were placed at $\pm 2\sigma$ from the nominal mass of π^0 (~ 135 MeV).

The two sources of background during the photon cluster reconstruction are beam background and hadronic split-off. The latter is emitted by hadrons which interact with detector material, and produce clusters in the ECL which are not matched to any tracks. In order to suppress these background processes, outputs from two classifiers are used when selecting the photons. The cuts applied on the MVA outputs are $\text{beamBackgroundSuppression} > 0.5$ and $\text{hadronicSplitOffSuppression} > 0.1$. These cuts were decided based on recommendations from the dedicated group.

K_S^0 Reconstruction: The reconstruction of neutral particles that travel some distance from the IP before decaying, denoted V0, is implemented by a task-specific algorithm in basf2 [39]. All oppositely charged tracks are paired by this algorithm, and for each pair the tracks are extrapolated to the innermost hit of one of the tracks. Afterwards, a vertex fit is implemented, and combinations that fail the extrapolation or the vertex fit are rejected. Accepted candidates are selected by further cuts. One is that the invariant mass of the combination should be within 30 MeV of the nominal V0 candidate mass, for K_S^0 this gives a cut of $0.468 \text{ GeV} < M < 0.528 \text{ GeV}$ (see Figure 4.4).

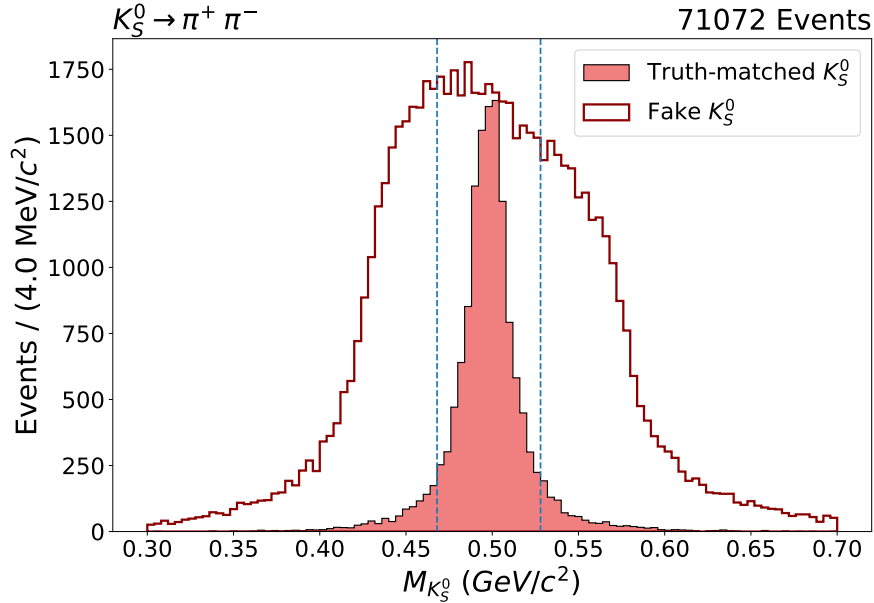


Figure 4.4: The mass distribution for K_S^0 candidates reconstructed in the $K_S^0 \rightarrow \pi^+\pi^-$ channel. The distributions for correctly reconstructed candidates and the rest are given as two unstacked histograms. The cuts applied during the selection are given by the dashed lines.

The reconstructed D^0 candidates are required to have invariant masses within 2σ of the nominal mass of the D^0 meson. The value of the σ (width) was obtained separately for each decay channel by a Gaussian fit to the distribution of truth-matched candidates. The plots with the fits

are shown in Fig. A.2. This separation was necessary since the width of the distribution for some channels were much larger. For instance channels involving neutral pions, due to the mass resolution on these particles. It should be noted that for the decays with lower branching fractions, a lower number of reconstructed signal events are used in this fit, and the width is less well-determined. In addition, the agreement with a Gaussian is in general imperfect, and some of the fits are biased. Nevertheless, the fit results suffice as approximate numbers to place the invariant mass cuts on.

The signal reconstruction efficiencies (defined in Subsec. 4.7.1) for each decay channel are given in the right column of Tab. 4.2. These are calculated based on the number of signal $\Upsilon(4S)$ events obtained from a 400 fb^{-1} MC data sample, using equation 4.5.

Here the generated number of events is calculated based on the number of generated $B\bar{B}$ pairs and the product branching fraction of the decays down to the final state particles. The statistical uncertainties stated are based on the Poisson uncertainty on the number of reconstructed events ($N_{reconstructed}$), which is $\sqrt{N_{reconstructed}}$. Any uncertainty on the number of generated events is neglected here. The uncertainties on the efficiencies are specified since for decays with low numbers of reconstructed events, e.g. $D^0 \rightarrow K_S^0 K_S^0$, the reconstruction efficiency appears to be large, but so is the uncertainty on this value. Overall, $D^0 \rightarrow K^-\pi^+$ is the channel which yields the highest number of signal events.

Decay Channel	Branching Fraction (%) (Belle II MC)	B_{signal} Reconstruction Efficiency (%)
$D^0 \rightarrow K^-\pi^+$	3.9500	3.37 ± 0.19
$D^0 \rightarrow K^-\pi^+\pi^0$	14.4000	0.69 ± 0.04
$D^0 \rightarrow K^-\pi^+\pi^+\pi^-$	5.1072	1.25 ± 0.10
$D^0 \rightarrow K_S^0\pi^+\pi^-$	1.9357	1.49 ± 0.18
$D^0 \rightarrow K_S^0\pi^+\pi^-\pi^0$	0.4147	4.71 ± 0.68
$D^0 \rightarrow K_S^0\pi^0$	0.8572	1.19 ± 0.24
$D^0 \rightarrow \pi^+\pi^-$	0.1455	4.19 ± 1.08
$D^0 \rightarrow K^+K^-$	0.4080	5.78 ± 0.76
$D^0 \rightarrow K_S^0 K_S^0$	0.0097	8.34 ± 5.9
	Total: 27.2281	

Table 4.2: D^0 decay channels included in the reconstruction, along with the branching fractions and reconstruction efficiencies.

It is possible that more background i.e. incorrectly reconstructed events are included by certain D^0 decay channels, than signal. This would reduce the significance of the final sample, where significance is defined as explained below. This is investigated by repeating the reconstruction and best candidate selection (Subsec. 4.5) steps nine times, each time excluding one of these decay channels from the reconstruction. The final significance is evaluated in each case, and compared to the significance obtained when all nine channels are included in the analysis. It was observed that while some channels do not improve the significance, none of them negatively im-

pact this figure of merit, and all of them contribute to the reconstructed signal count. Therefore, the analysis is continued with all nine D^0 decay channels.

Significance

Significance is a figure of merit used to evaluate the performance of the reconstruction. It is defined as given in equation 4.2, and it represents how the number of signal events after selection compares to the fluctuations in the total number of events, i.e. signal plus background.

$$S_{\text{signal}} = \frac{S}{\sqrt{S+B}} \quad (4.2)$$

$S \stackrel{\text{def}}{=} \text{Number of signal events}$

$B \stackrel{\text{def}}{=} \text{Number of background events}$

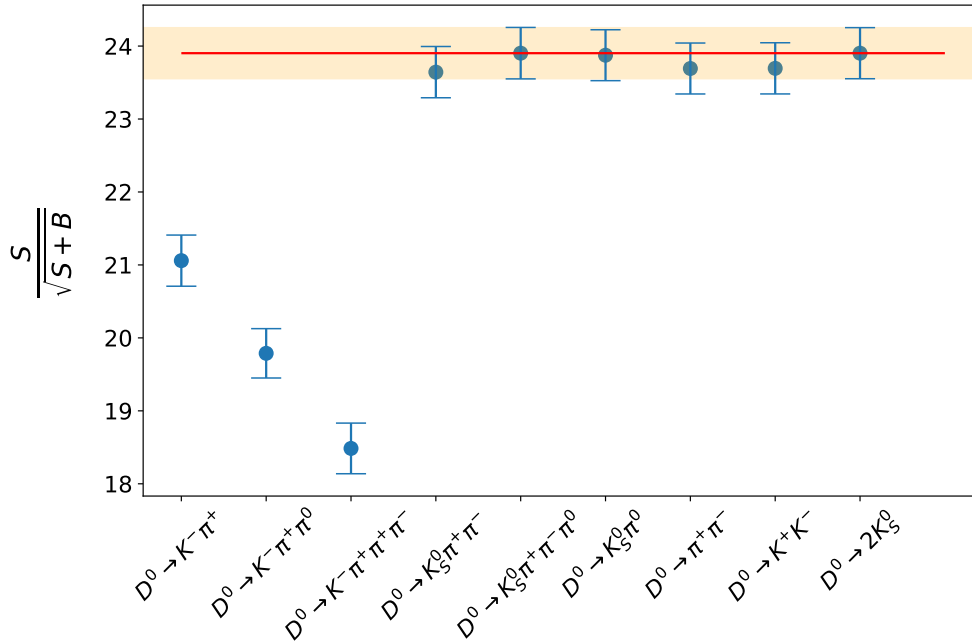


Figure 4.5: The effect of excluding D^0 decay modes on the final significance. The red horizontal line shows the significance when all nine decay modes are included in the reconstruction. The error bars (and the orange shaded region) represent the statistical uncertainty, calculated from the Poisson uncertainty on the number of signal and background events.

D^* and D^{**} Reconstruction

The D^0 candidates are combined with charged tracks fitted with the pion mass hypothesis, to obtain $D^{*\pm}$ candidates. These are subsequently combined with an additional charged pion to form the neutral D^{**} mesons. The pion originating from the D^* decay is termed the *soft pion*. Since the difference between the rest mass of $D^{*\pm}$ and D^0 is around 145 MeV, and the charged pion mass is around 140 MeV, the momentum of the pion originating from this decay is kinematically constrained to low values. The momentum of this soft pion in the CMS frame is required to be lower than 200 MeV, to reject background from B meson decays with higher energy pions. Due to its low momentum, PID variables do not achieve well differentiability for these particles, and no cut is applied on these variables.

As done at the D^0 reconstruction stage, an invariant mass cut is applied on the decay products of the $D^{*\pm}$ candidates. However this time the $\Delta M_{D^*} = M_{D^*} - M_{D^0}$ variable is used. This variable is determined with higher resolution, and the signal distribution shows a more pronounced peak than the M_{D^*} variable. A cut was placed at 3σ from the nominal value, where σ is again obtained by a Gaussian fit to the ΔM_{D^*} distribution of truth-matched $D^{*\pm}$ candidates as shown in Fig. A.3.

Vertex Fits

After reconstruction of the B_{signal} vertex from the D^{**} and lepton candidates, a vertex fit was implemented using TreeFitter [32]. TreeFitter is one of the vertex fitting algorithms implemented in basf2. The fit takes in as parameters information from entire decay trees, such as the momenta of the final state particles and the four momenta and decay time of the intermediate particles. This analysis made use of two vertex fits, one at the signal B decay vertex, and later one at the $D^* - \ell$ vertex. The second fit was tested since fitting a vertex of two particles was expected to be less computationally extensive and yield more accurate results than including the π track. In the end, both fits were implemented, and the result of the signal B vertex fit was used to apply a cut by rejecting candidates that failed the fit. Meanwhile, the χ^2 result of the $D^* - \ell$ fit was later used in the best candidate selection, described in Sec. 4.5.

4.3 Tag Side

The second B meson in the event was reconstructed in exclusively hadronic decay channels using the Full Event Interpretation algorithm, introduced in subsection 4.3.1. This method of fully reconstructing the second B , so-called *tagging* helps put constraints on the kinematic properties of the signal B meson. In the context of this analysis, this helps determine the momentum of the signal B meson despite the missing momentum due to the ν , which is used in the q^2 study. Additionally, reconstructing the tag B in hadronic channels ensures that the only source of missing energy is the neutrino from the semileptonic signal decay.

4.3.1 Full Event Interpretation (FEI)

FEI is a multivariate classifier (MVC) based algorithm employed widely in Belle II analyses.[37] It is used to reconstruct full decay chains in a hierarchical manner, starting from final state particles. At each reconstruction stage, in other words, when an intermediate particle is reconstructed, an MVC is used to evaluate the probability that the reconstructed candidate describes the true decay. The output of the classifier is passed on to the next level of reconstruction, along with the four momenta of the particles and information from vertex fits. Therefore, the final classifier output for the B meson candidate combines all this information in one value, called *signal probability*. This value is used to reject or accept candidates on the analysis level.

FEI was used to reconstruct the tag B only in hadronic decay channels. Hadronic reconstruction has the advantage of having no missing particles in the final decay products. This allows the kinematic properties of the B meson to be fully reconstructed. One drawback of this method is that it gives lower reconstruction efficiency (0.61%) compared to semileptonic FEI (1.45%) or inclusive tagging approaches since it is limited by the branching fraction of B meson decays in the hadronic decay channels considered [37]. For the purposes of this analysis, this approach was deemed suitable since the reconstruction of the signal B momentum is necessary for the q^2 study.

After the FEI reconstruction, only B candidates with signal probability > 0.001 were included in the analysis. Tighter cuts were tested, however this cut was found to perform the best, as discussed in the section on cut optimisation (Sec. 4.6).

4.3.2 Other Selection Cuts

There are two other variables of interest when evaluating whether the reconstructed candidate is in fact a B meson decay: ΔE and M_{bc} . ΔE is defined as $E_B - \frac{1}{2}E_{CMS}$. For a B meson coming from an $\Upsilon(4S)$ event, where the $\Upsilon(4S)$ energy is the CMS energy, this value is expected to be zero. Whereas M_{bc} is defined as $\sqrt{E_{beam} - |p_B|^2}$, and since beam energy should be equal to the energy of either B meson in a signal event, this value should equal the B meson mass for true signal candidates.

Loose cuts of $-0.1 \text{ GeV} < \Delta E < 0.1 \text{ GeV}$ and $5.27 \text{ GeV} < M_{bc} < 5.29 \text{ GeV}$ were applied to these variables to reject some of the background candidates which were falsely reconstructed. This is likely to be due to missing particles, or using particles from the signal side in the reconstruction, which would give inaccurate total momentum for the B meson. However, it should be noted that the distribution of $B\bar{B}$ background events also peak in the signal region for both of these variables. One explanation of this is the cut applied on the B_{tag} signal probability. Since FEI employs vertex fits to give the final signal probability, it is expected that this cut favours background candidates that fit the momentum requirements for a B meson decay. This would enhance the background distribution in the signal region, while eliminating background in other regions, resulting in a peaking background shape. In addition, the ΔE variable is used in the training of the FEI, which may also cause this peaking background.

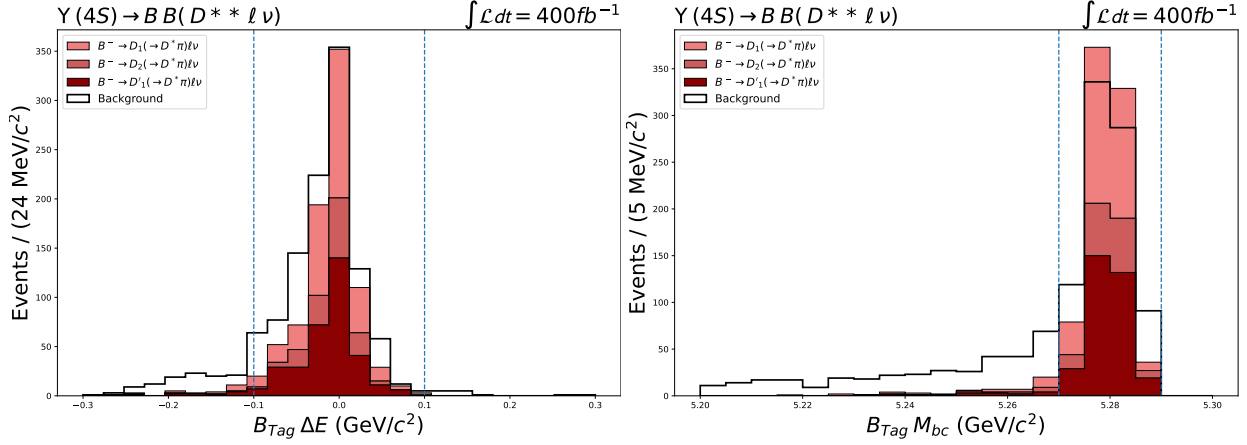


Figure 4.6: ΔE (left) and M_{bc} (right) distributions of reconstructed $Y(4S)$ candidates. All the other cuts, including the signal probability cut have been applied. Candidates are shown in one background and three signal categories.

4.4 Event Selection

4.4.1 Missing Energy and Momentum

The energies and momenta of the reconstructed signal and tag B are not expected to add up to the initial energy and momentum in the collision event, due to the neutrino which leaves no traces in the detector. For signal decays which include only one neutrino, the missing energy can be assigned to this single neutrino. Then, the rest mass calculated from the missing energy and momentum (Eqn.4.3) in the event must be zero for signal events. (the neutrino mass is negligible under detector resolution) This is a distinguishing feature of semileptonic decays with a single neutrino.

$$M_{miss} = \sqrt{E_{miss}^2 - |\vec{p}_{miss}|^2} \quad (4.3)$$

A disadvantage of this variable is that it is model dependent, since $\sqrt{E_{miss}^2 - |\vec{p}_{miss}|^2} = \sqrt{(E_{miss} - |\vec{p}_{miss}|)(E_{miss} + |\vec{p}_{miss}|)}$, and $(E_{miss} + |\vec{p}_{miss}|)$ is a model dependent quantity. A model independent variable may be defined as given in Eqn. 4.4 which, by the same argument peaks at zero for signal events. An initial cut was applied symmetrically around zero, at $-0.1 \text{ GeV} < u < 0.1 \text{ GeV}$. This was later optimised as described in Sec. 4.6, to give a looser, asymmetric cut.

$$u = E_{miss} - |\vec{p}_{miss}| \quad (4.4)$$

Additionally, in order to distinguish decays with a ν from other decays with low missing mass and low missing momentum, a loose cut of $|\vec{p}_{miss}| > 0.05 \text{ GeV}$ was applied.

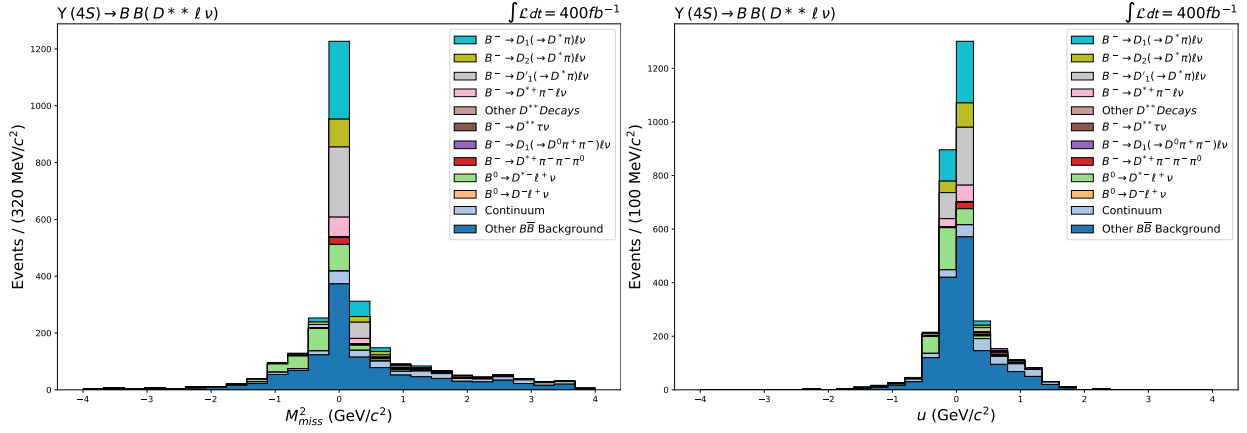


Figure 4.7: M_{miss}^2 and u distributions for $\Upsilon(4S)$ after selection cuts and best candidate selection. The different B_{signal} decay modes are shown as stacked histograms.

4.4.2 Rest of Event

Rest of Event (ROE) is the term given to all the tracks and clusters in the collision experiment, which were not used in reconstructing the particle of interest. In this analysis, the ROE of reconstructed $\Upsilon(4S)$ candidates was studied. Since no additional particle is created in the scattering of $e^+e^- \rightarrow \Upsilon(4S)$, for true signal events ROE should consist of solely beam background. Therefore, the existence of physics background in the ROE is an indicator that the event is not a true $\Upsilon(4S)$ event.

Since beam background is common between signal and background events, better distinguishability is obtained by eliminating these events from the ROE. The cuts applied on ROE tracks and clusters for this purpose are termed *ROE masks*. To eliminate some of the low energy beam background, a cut of $E > 0.1$ GeV was applied to all clusters in the ROE. In addition, the timing of ECL clusters can be used to distinguish particles from the e^-e^+ collision. At Belle II, the timing of a cluster is defined as the fitted time of the signal recording at the highest energy ECL crystal of the cluster [13]. The difference between this measurement and the time of event, after some calibration and corrections, is called *clusterTiming*, and is expected to be zero for particles originating from the e^-e^+ collision. The use of this variable on MC simulation is undertaken with caution since there have been reports of data-MC disagreement in the distributions of this variable by the collaboration. Nevertheless, for this analysis a loose cut of $clusterTiming < 200ms$ was applied.

Moreover, tracks from physics background originate from the interaction point, whereas beam background does not. Hence, particles considered in the ROE are constrained to be close to the IP by applying the cut: $|dz| < 0.5$ cm. Additionally, and for the same purpose of eliminating beam background, a cut of $p_t > 0.08$ was applied. A detailed investigation, and optimisation of these cuts was not conducted.

After applying the ROE masks, two variables are used to differentiate signal $\Upsilon(4S)$ events

from background. One of these is E_{extra}^{ROE} , which is the total energy of the particles in the ROE. In the ideal case, where the ROE masks eliminated all beam background events, this variable would equal zero for signal events. In practice, it is still expected to take on lower values for signal than for background. The cut on this variable was optimised to $< 0.6 \text{ GeV}$ in Sec. 4.6. The second variable considered is the number of tracks in the ROE ($n\text{Tracks}_{ROE}$). Through a similar argument, we expect no ROE tracks to be detected for signal events. Hence this variable is required to be zero.

After the finalisation of the ROE masks, it was suspected that one of the track cuts on the p_t and $|dz|$ may be redundant, since they are both expected to effectively remove beam background. Moreover, the $|dz|$ cut has the disadvantage that it might remove particles originating from K_S^0 decays. At this stage, the ROE masks were kept unchanged. However, in order to investigate this suspicion, the E_{extra}^{ROE} distribution was observed after loosening the cut on $|dz|$, while keeping the p_t cut fixed. This is shown on Fig. 4.8. No change on the distributions of signal and background events was observed, which supports our suspicion that the p_t cut may have already removed most beam background events. This observation also hints at low number of K_S^0 decays in the observed events, since one would expect the dz cut to make a difference otherwise.

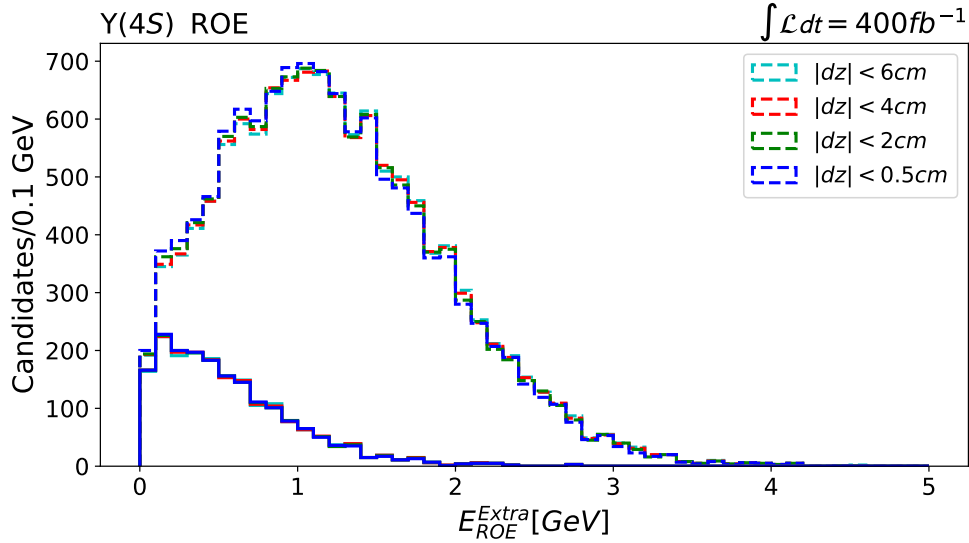


Figure 4.8: The influence on varying the $|dz|$ cut in the ROE mask while keeping all the other cuts the same. The dashed lines give background and full lines signal distributions. It is observed that the variation has close to no impact.

4.5 Best Candidate Selection

There was found to be more than one $\Upsilon(4S)$ candidate which passed all the cuts detailed in the previous sections. After applying the cuts, the average number of candidates per event was $\gtrsim 2$.

The reconstruction process may produce different $Y(4S)$ candidates which share some of their daughter particles. For instance, a single B_{signal} candidate might be remaining after applying all the cuts, however this can be combined with various B_{tag} candidates to produce a multiplicity of $Y(4S)$ candidates. Therefore, a best candidate selection (BCS) is implemented through a series of steps aimed at picking the most signal-like candidate for each particle in the reconstruction. In each step, candidates are ranked according to the value of a given variable (listed below), and the best ranking candidates are selected to move on to the next step. Multiple candidates are allowed to have equal rank. These variables are, in order:

1. Difference between the reconstructed mass of D^0 and the nominal mass. (lowest value is given highest rank)
2. Difference between the reconstructed mass difference of $D^{\pm*}$ and the nominal value. (lowest value is given highest rank)
3. B_{tag} signal probability (highest value is given highest rank)
4. χ^2 probability from the vertex fit of $D^{\pm*}-\ell$ candidates. (highest value is given highest rank)

If events with multiple candidates remain after these four steps, one is chosen at random. The effect of this random selection is negligible since the candidate multiplicity at this stage was ~ 1.02 .

4.6 Cut Optimisation

After the selection of all cut variables, a decision was made to simultaneously optimise the cuts on four chosen variables. The four variables were u , E_{extra}^{ROE} , B_{tag} signal probability, and p_{ℓ}^{CMS} . These were chosen due to the fact that an approximate placement of the cuts based on observing signal and background distributions was deemed non-trivial. The optimisation was limited to four variables to save computing time.

A four dimensional, equally spaced grid was set with each dimension containing possible values of one of the variables. Then a search was implemented on the grid, by evaluating the significance on a 400 fb^{-1} MC sample after applying the cuts at each point in the grid. The point which maximised the significance was chosen to set the values of the cuts on these four variables. The MC sample used was already subjected to all the other cuts mentioned in this chapter, but not the BCS, since this is typically performed at the end to avoid discarding more events than intended.

The grid search was implemented three times, each time with a smaller and finer grid placed at areas which gave higher significance in the previous search. The final search was implemented on the following values :

- u lower cut: 11 values between -0.2 GeV and 0 GeV
- u upper cut: 11 values between 0.1 GeV and 0.3 GeV

- E_{extra}^{ROE} upper cut: 21 values between 0.6 GeV and 0.8 GeV
- B_{tag} signal probability lower cut: 11 values between 0.001 and 0.005
- p_{ℓ}^{CMS} lower cut: 11 values between 0.3 GeV and 0.7 GeV

The optimisation improved the signal significance from 19.21 to 21.15, which corresponds to a 10.13% increase. The difference in significance is 8σ away from zero, which points to a significant improvement. The number of reconstructed signal events increased from 831 to 1483 decays, whereas the number of background events increased from 1041 to 3164. The values of the cuts after the optimisation are shown in Fig. 4.9

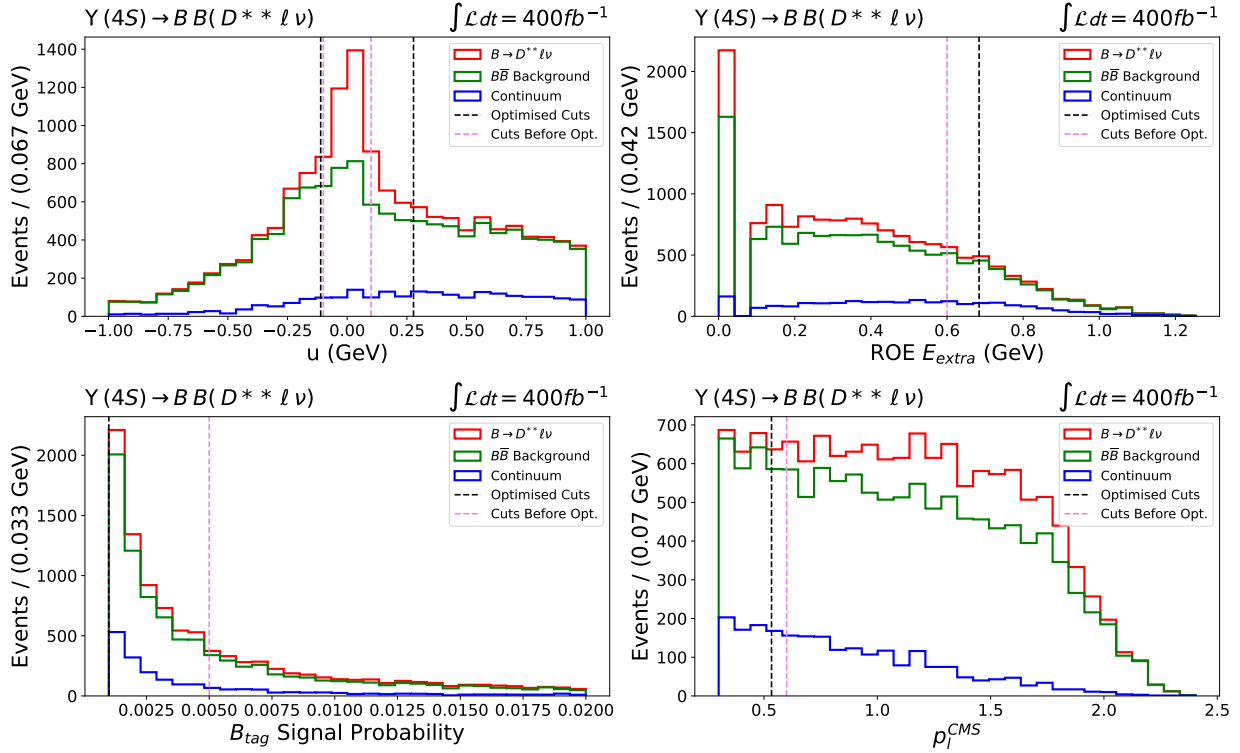


Figure 4.9: The distributions of the four variables considered in the Grid Search Optimisation for signal, $B\bar{B}$ background, and continuum events. The cuts before and after optimisation are shown by the dashed lines.

4.7 Summary and Efficiency

All the cuts applied during the $Y(4S)$ reconstruction are summarised in Tab. 4.4, and ROE cuts are given in Tab. 4.5.

4.7.1 Reconstruction Efficiency

Reconstruction efficiency gives the fraction of generated or produced signal decays which were obtained by the reconstruction (see Eqn. 4.5).

$$\epsilon_{rec} = \frac{N_{reconstructed}}{N_{generated}} \quad (4.5)$$

The total number of generated events which decay into the channels considered can be calculated directly using the number of charged B mesons generated ($N_{B^-B^+}$), and the branching fractions used in the MC simulation. From this, and the number of reconstructed signal events from 2.4 ab^{-1} MC simulation, the total reconstruction efficiencies for each D^{**} decay were calculated.

$$N_{generated} = N_{B^\pm} \times \mathcal{B}(B^- \rightarrow D^{**} \ell \nu) \times \mathcal{B}(D^{**} \rightarrow D^{*+} \pi^-) \times \mathcal{B}(D^{*+} \rightarrow D^0 \pi^+) \times \sum \mathcal{B}(D^0 \text{ decays}) \quad (4.6)$$

Additionally, the reconstruction efficiency may be dependent on the four-momentum transfer during the decay. This is expected since some reconstruction criteria have used variables which are correlated to the q^2 . To estimate this variation, 5000 generator level signal MC samples were generated with no reconstruction applied, and the w distributions were obtained. From the distributions the fractions of events generated in each of the three w bins were obtained (the binning here is the one used in the binned fit detailed in Sec. 6). By multiplying these with the total number of generated events calculated as in Eqn. 4.6, and comparing to the number of reconstructed events in each bin, one obtains the per bin reconstruction efficiencies. These are listed in Tab. 4.3.

		Efficiencies per w bin			
		$\epsilon_{rec}^{total} (\times 10^{-5})$	$\epsilon_{rec}^{bin1} (\times 10^{-5})$	$\epsilon_{rec}^{bin2} (\times 10^{-5})$	$\epsilon_{rec}^{bin3} (\times 10^{-5})$
D_1	9.69	8.55	10.99	10.46	
D_2^*	10.21	7.67	11.49	12.31	
D_1'	9.58	4.92	11.18	15.82	

Table 4.3: Total and per bin reconstruction efficiencies for the three signal D^{**} decays.

It is worth noting that one can assume the tag and signal side reconstruction efficiencies to be independent. This allows one to get an estimate of the reconstruction efficiency of the signal B alone. For this the reported efficiencies should be divided by the tag side (FEI) reconstruction efficiency, which is reported to be 0.61% [37]. Another comment is that there is some uncertainty on the branching fractions used while calculating the number of generated events. However, since the same uncertainty is present in the number of reconstructed events (the same branching fractions were used in the MC generation), this uncertainty cancels out.

		Particle	Cut	Best Candidate Selection
$\Upsilon(4S)$	Signal B	e	electronID (BDT based) > 0.9 $p_{\text{CMS}} > 0.5 \text{ GeV}$ Bremsstrahlung corr. [Tab. 4.1]	Highest chiProb from $D^{\pm*} \ell$ vertex fit
		μ	$\mu\text{ID} (\mathcal{L} \text{ based}) > 0.9$ $p_{\text{CMS}} > 0.5 \text{ GeV}$	
		all tracks	$ dr < 0.5 \text{ cm}$ $ dz < 2 \text{ cm}$ thetaInCDCAcceptance	
		soft $\pi^{\pm}$	$p_{\text{CMS}} < 0.2 \text{ GeV}$	
		π^0 photons	beamBackgroundSuppression > 0.5 hadronicSplitOffSuppression > 0.1	
		π^0	$0.118 \text{ GeV} < M < 0.152 \text{ GeV}$	
		K_S^0	$0.468 \text{ GeV} < M < 0.528 \text{ GeV}$ vertex fit	
		D^0 $D^{\pm*}$	M cut at $\pm 2\sigma$ $0.1414 \text{ GeV} < \Delta M_{D^*} < 0.1484 \text{ GeV}$	M closest to nominal value ΔM_{D^*} closest to nominal value
	Tag B		B_{tag} signal probability > 0.001 $-0.1 \text{ GeV} < \Delta E < 0.1 \text{ GeV}$ $5.27 \text{ GeV} < M_{bc} < 5.29 \text{ GeV}$	Highest signal probability
			$-0.12 \text{ GeV} < u < 0.24 \text{ GeV}$ $p_{\text{miss}} > 0.05 \text{ GeV}$	

Table 4.4: Summary of all cuts and selections applied during the $\Upsilon(4S)$ reconstruction.

ROE	ROE selection	tracks	$p_t > 0.08 \text{ GeV}$ $ dz < 0.5 \text{ cm}$ $nCDCHits > 0$
		clusters	$E > 0.1 \text{ GeV}$ $\text{abs}(\text{clusterTiming}) < 200 \text{ ms}$
	Cuts		$E_{extra}^{ROE} < 0.6 \text{ GeV}$ $nTracks_{ROE} = 0$

Table 4.5: Summary of the selection of ROE particles, and ROE cuts applied on $\Upsilon(4S)$ events.

Chapter 5

Fit

5.1 Fit Procedure

The yields of each $B \rightarrow D^{**}\ell\nu$ decay and background events were estimated by an unbinned extended maximum likelihood fit to the $\Delta M_{D^{**}}$ distribution, performed using the RooFit library [53], included in the ROOT framework. The probability density functions used in the fit were generated from MC simulation by the kernel density estimation method. The methods and results will be detailed in the following subsections.

$B \rightarrow D^{**}\ell\nu \Delta M_{D^{**}}$ Distribution

The $\Delta M_{D^{**}}$ distributions of signal and background events after selection and classification are given in Fig. 5.1, where $\Delta M_{D^{**}} = (M(D^{**}) - M(D^{*+}))$. For the PDF production and the fit, the distribution is split into four components: Three signal B decays into $D_1/D_2^*/D_1'\ell\nu$, and one background component including all other decays. The non-resonant $B \rightarrow D^*\pi\ell\nu$ decay was classified as background. Consequently, it makes up the largest proportion of background decays. As mentioned in Ch. 3, the abundance of this decay channel is under scrutiny, and the background distribution seen in Fig. 5.1 may look different in future analyses.

The $\Delta M_{D^{**}}$ distribution shapes of these four components are distinguishable since the signal distributions have different widths and peak at slightly different positions, whereas the background distribution is non-peaking. Therefore, $\Delta M_{D^{**}}$ was used as the fitting variable.

MC Sample

The fit detailed in this chapter was implemented on hadronic FEI skims (see Ch. 3) from a 400 fb^{-1} run independent MC sample. This corresponds roughly to the integrated luminosity of the data collected by Belle II before Long Shutdown 1. The PDFs were produced using a separate MC sample of 2.4 ab^{-1} for the signal components and 600 fb^{-1} for the background component. The luminosity of the background component was limited by the amount of FEI skims available at the time of this analysis (see Ch. 3).

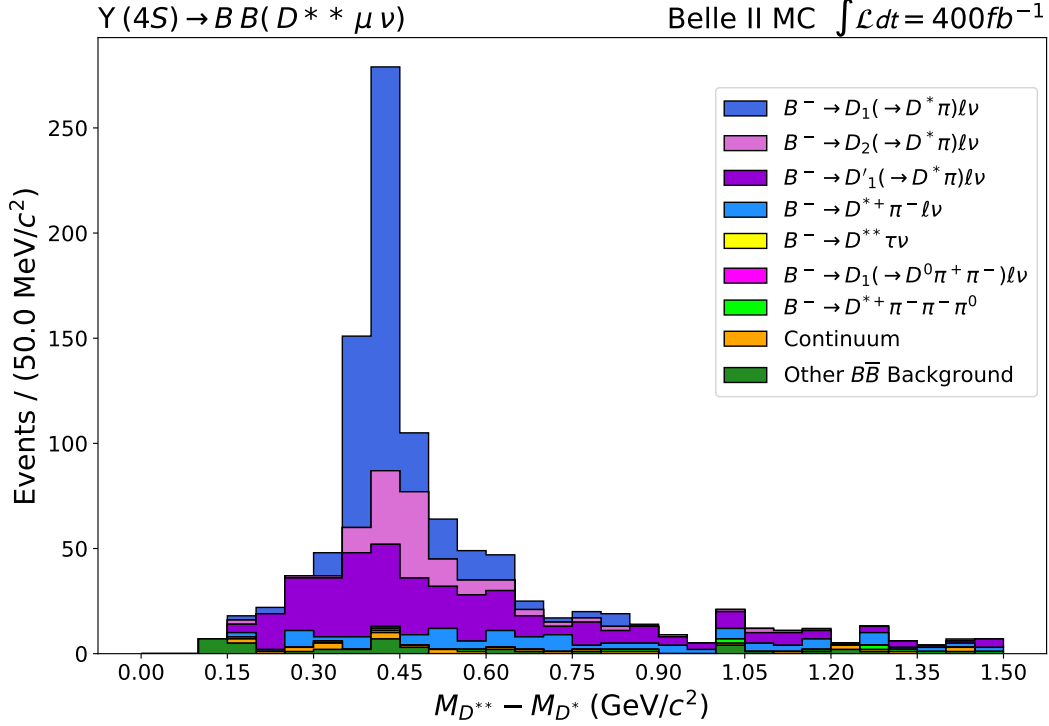


Figure 5.1: The $\Delta M_{D^{**}}$ distribution of $B \rightarrow D^{**}\ell\nu$ candidates after reconstruction and best candidate selection, classified by signal B decay modes.

5.1.1 Kernel Density Estimation

Kernel density estimation [25] is one method of producing a continuous and differentiable probability density function (PDF) which describes discrete data points. It is a non-parametric approach, meaning it does not rely on knowledge of an underlying physical model with tunable parameters and can be applied directly on a given data set¹. This has the advantage that PDF creation is not influenced by the parameterisation of the model. Nevertheless, there is a dependency on the amount of statistics available for PDF creation.

The method places a non-negative, real, and normalisable function (so-called kernel) at each data point in the data set, and adds these kernels to produce an estimator for the PDF. The use of the kernels is to smooth out the data points, and approximate the underlying continuous distribution. The PDF estimator is defined as in equation 5.1 below.

$$\hat{f} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - t_i}{h}\right) \quad (5.1)$$

¹Since MC simulation is used for the PDF estimation, which is based on an underlying decay model, there is still dependence on the accuracy of modelling.

Here $t_{i=1,2,\dots,n}$ are the points in the data set, and h is the so-called *bandwidth*. In order for the method to remain non-parametric, the bandwidth is determined solely based on the data set. The value of this variable determines the smoothness of the final PDF, in other words how much of the small scale fluctuations are included in the PDF. A standard choice of the kernel $K(x)$ is a Gaussian with $\mu = 0$ and $\sigma = 1$. This is the kernel used in this analysis.

$$K(x) \stackrel{\text{def}}{=} \mathcal{N}(0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (5.2)$$

Instead of using a fixed bandwidth parameter, one could tune this parameter individually for each data point. This is termed KDE with *adaptive width*. The bandwidth is calculated as shown below.

$$h_i = \rho \left(\frac{4}{3} \right)^{1/5} \sqrt{\frac{\sigma}{\hat{f}_0(t_i, h_0)}} n^{-1/5} \quad (5.3)$$

Here n is the number of events in the data set, ρ is the so-called *smoothness parameter*, and $\hat{f}_0$ is calculated as in Eqn. 5.1, with a fixed bandwidth parameter. Intuitively, this equation imposes that the width is narrower for regions with a higher density of data points. This way the small scale fluctuations of the distribution are captured more accurately in these regions. The ρ parameter can be tuned to suppress fluctuations globally. It has been found that the method does not have dependence on the initial value of the bandwidth, h_0 , whereas the parameter ρ is optimisable for better results. Moreover, the choice of ρ is accepted to be even more influential on the result than the choice of kernel [22].

ρ Optimisation

A Cross-Validation method was used to obtain an optimal value for ρ . Cross-Validation methods, in general, involve splitting a dataset into training and validation sets iteratively to evaluate a model. In this context, this is done by first obtaining an estimator for the PDF by KDE, using all data points except one, $\hat{f}_i(D - t_i; \rho)$. Then, the PDF is evaluated on the data point which was excluded from training, by means of a loss function, $L(\hat{f}_i(t_i), t_i)$. Repeating this for each point in the dataset and summing the values of the loss function gives the Cross-Validation (CV) term which is minimised by the choice of ρ .

$$\text{CV}(\rho) = \frac{1}{N} \sum_{i=1}^N L(t_i, \hat{f}(D - t_i; \rho)) \quad (5.4)$$

Mean Integrated Square Error (MISE) (defined in Eqn. 5.5) was chosen as the loss function because it is shown to give consistent PDF estimators. In other words, the MISE loss function is asymptotically minimised [22].

$$I(p(x), q(x)) = \int (p(x) - q(x))^2 dx \quad (5.5)$$

With this choice of the loss function and a Gaussian Kernel, the criteria of minimising the CV in Eqn. 5.4 reduces to minimising the quantity defined in Eqn. 5.6. Here t_i and h_i are defined as above, and n denotes the number of points in the data set. This equation is based on the one given in Ref. [22]. During the derivation of Eqn. 5.6, some of the terms which are independent of h_i are left out, which results in negative CV values.

$$\begin{aligned} \text{CV}(\rho) = & \frac{1}{n(n-1)} \sum_{i=1}^n \mathcal{N}(0, \sqrt{2}h_i) + \frac{(n-2)}{n(n-1)^2} \sum_{i \neq j} \mathcal{N}(t_i - t_j, \sqrt{h_i^2 + h_j^2}) \\ & - \frac{2}{n(n-1)} \sum_{i \neq j} \mathcal{N}(t_i - t_j, h_j) \end{aligned} \quad (5.6)$$

After splitting the PDF data set into four fit classes (and w bins for the binned fit), KDE was repeated with 400 evenly spaced ρ values from 0 to 4, and the values which minimise the CV were chosen to be used in the PDF production. The table below shows the obtained ρ values and the corresponding CV values (using PDFs for the full w range). This is compared to the CV obtained when $\rho = 1$, which is the default value in RooFit.

PDF Class	ρ_{OPT}	$\text{CV}(\rho_{OPT})$	$\text{CV}(\rho = 1)$	$(\text{CV}(\rho_{OPT}) - \text{CV}(1)) / \text{CV}(1)$
$B \rightarrow D_1 \ell \nu$	0.37	-6.6724	-6.63579	0.55%
$B \rightarrow D_2^* \ell \nu$	0.53	-4.40801	-4.39163	0.37%
$B \rightarrow D_1' \ell \nu$	0.66	-1.32702	-1.32679	0.02%
Background	0.33	-1.01914	-1.00682	1.22%

Table 5.1: The Cross-Validation values for each of the four PDF classes, using the full w range. The CV values for the optimised values of ρ are given alongside the default value, and the relative difference in CV due to the optimisation.

5.1.2 Extended Maximum Likelihood Fit

The collision experiment (or simulation) can be modelled as a sampling of N points from an underlying PDF. Here the total PDF (Eqn. 5.7) is obtained by summing the individual PDFs ($f_i(x)$) produced as detailed in the previous subsection. The PDF has as floating parameters the coefficients of each component in the sum. To this model one may add the fact that in a collision experiment the total number of events N itself can be modelled as a Poisson distributed random variable with an underlying mean value ν . This case is treated by the *extended maximum likelihood* formalism [24], which gives the log likelihood function given in equation 5.8. Here μ_j denotes the number of events sampled from each component distribution, defined as $\nu\theta_j$. Maximising the function in equation 5.8 (or minimising the negated function) gives maximum likelihood (ML) estimates for these values, which are then used in the branching fraction estimates.

$$f_{\text{TOTAL}}(x) = \sum_{i=1}^4 \theta_i f_i(x) \quad (5.7)$$

$$\log(L(\mu)) = - \sum_{j=1}^4 \mu_j + \sum_{i=1}^N \sum_{j=1}^4 \log(v \theta_j f_j(x_i)) \quad (5.8)$$

For this analysis, an *unbinned* fit was preferred since it results in more accurate fit results. As a downside, this method is more computationally extensive, nevertheless feasible with the size of the data sample used in this analysis.

The minimisation of the negative log likelihood (NLL) function is carried out using the ROOT implementation of MINUIT [36]. Internally, it calls the MIGRAD and HESSE algorithms for finding the parameter values at minimum, and estimating the covariance matrix from the second derivatives of the NLL. It should be noted that for the estimation of the covariance matrix, the NLL function was assumed to be quadratic around the obtained minimum². In other words, any non-linearities were not considered in the error calculation.

5.2 Fit Results

	n_{D_1}	$n_{D_2^*}$	$n_{D_1'}$	n_{bkg}
MC Values	398	124	370	250
Fit Results	349.651 ± 32.619	137.565 ± 31.871	426.870 ± 54.118	228.034 ± 40.043
n_{D_1}	1	-0.424	-0.226	0.095
$n_{D_2^*}$	-0.424	1	-0.303	0.066
$n_{D_1'}$	-0.226	-0.303	1	-0.731
n_{bkg}	0.095	0.066	-0.731	1
global	0.609	0.653	0.829	0.771

Table 5.2: Fit results and correlation coefficients for the decay yields. The top two rows give the true MC values, and the fit results. The remaining rows show the correlation coefficients for the fitted parameters. The global correlation coefficient gives the correlation between the variable and the linear combination (of the other variables) with the highest correlation to it. Here I have changed the number of digits.

The resulting yields from the fit to the full sample are given in Tab. 5.2. The yield of the decay into $D_1 \ell \nu$ was determined with the lowest relative statistical uncertainty of approximately 9.3%, whereas the highest relative uncertainty, as well as the lowest yield was obtained for the decay into $D_2^* \ell \nu$. Overall the highest correlation was between the variables $n_{D_1'}$ and n_{bkg} , with a correlation coefficient of -0.731 . This high correlation is comprehensible since the peak of the

²i.e. no use of MINOS in the error estimation.

PDF for D_1' is not as pronounced as the other D^{**} modes, and the shape closely resembles the background PDF shape in certain regions. This is observed on Figure 5.2, where the fitted sum PDF, along with each PDF component and the MC sample are plotted, and the pull distribution is shown. The pull distribution is defined as: $(x_{\text{PDF}} - x_{\text{MC}})/\sigma_x$.

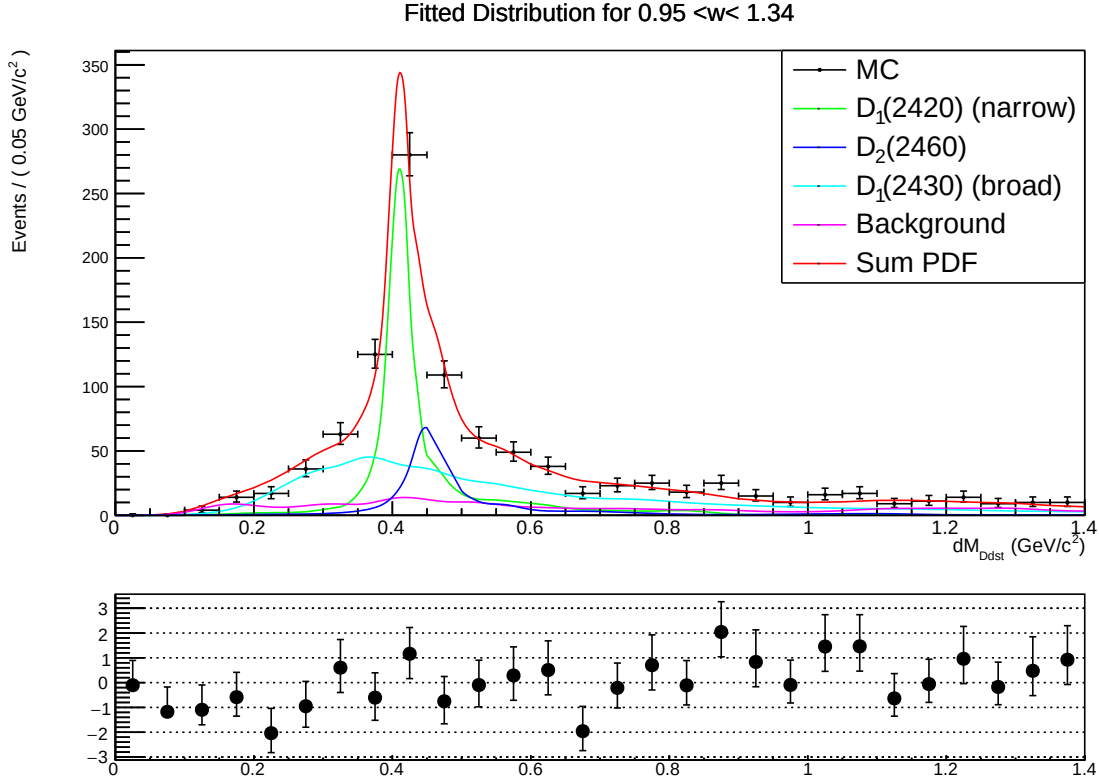


Figure 5.2: Top plot: Fitted sum PDF along with the four components and the MC sample used in the fit. The range $0.95 < w < 1.34$ includes all reconstructed events, i.e. it is the full w range. The legend gives the PDG names of the D^{**} modes, and the narrow and broad D_1 modes are indicated. Bottom plot: The pull distribution of the sum PDF and the MC sample.

5.3 Fit Validation

Monte Carlo studies were conducted in order to test whether the ML fit resulted in unbiased estimators for the yields, and whether the uncertainties were estimated correctly. Using the RooM-CStudy Class from the RooFit library [53], 1000 toy MC samples were created from the PDFs and then fitted. The number of events in each toy sample is chosen from a Poisson distribution with mean equal to the number of events in the "original" 400 fb^{-1} fit sample. From here on, the MC sample which was used in obtaining the reported results will be referred to as the original sample.

The fit results are regarded as biased if the mean of the distribution of results from toy studies differs significantly from the underlying true value³ of the parameter. One way to check this is to define a pull parameter as in equation 5.9, and to inspect its distribution. If a Gaussian fit to the distribution returns a mean consistent with zero, it is concluded that the fit is unbiased. Furthermore, since the pull values are normalised by the estimated uncertainty on the fitted variable, if the width of the Gaussian is 1, this indicates that the uncertainties are estimated correctly.

$$h_{pull} = \frac{\hat{\mu} - \mu}{\hat{\sigma}} \quad (5.9)$$

Figure 5.3 shows the pulls for the four fit parameters plotted with the Gaussians, fitted by a binned ML fit. A binned fit is less computationally extensive and is deemed to be sufficiently accurate in this case. Slight biases are observed for the D_1 and D_2^* fits, corresponding to pull means of -0.113 ± 0.031 and -0.085 ± 0.031 . Multiplying these by the mean fit uncertainty of the toy fits and dividing by the mean fit result gives the relative biases on the fit results. These are $(1.047 \pm 0.288)\%$ and $(2.363 \pm 0.859)\%$ for the D_1 and D_2^* yields respectively. The bias on D_2^* may be a result of statistical fluctuations, since it is still within 3σ from zero. Nevertheless a systematic uncertainty is assigned (Ch. 7) to the final branching fraction results to account for these biases.

5.4 Branching Fraction Estimation

The yields obtained from the fit were used to estimate the product branching fractions using the formula below.

$$\mathcal{B}(B^- \rightarrow D^{**} \ell^- \nu) \times \mathcal{B}(D^{**} \rightarrow D^{*+} \pi^-) = \frac{N_{D^{**}}}{N_{B^- B^+} \times \epsilon_{rec} \times \mathcal{B}(D^{*+} \rightarrow D^0 \pi^+) \times \sum \mathcal{B}(D^0 \text{ decays})} \quad (5.10)$$

where $N_{D^{**}}$ are the yields obtained from the fit, $N_{B^- B^+}$ is the number of charged $B\bar{B}$ pairs in the sample, and ϵ_{rec} is the reconstruction efficiency. When working with data, one does not know the exact value of $N_{B^- B^+}$. It may instead be calculated as:

$$N_{B^- B^+} = N_{Y(4S)} \times \mathcal{B}(Y(4S) \rightarrow B^- B^+) \quad (5.11)$$

Hence, the result depends on how well the branching fraction $\mathcal{B}(Y(4S) \rightarrow B^- B^+)$ is determined. The uncertainty due to this is discussed in Ch. 7.

Additionally, the division by the branching fractions of the D^* and D^0 decays in Eqn. 5.10 is necessary since the yield obtained is limited by the decays considered in the reconstruction. Alternatively, these values could be included in the reconstruction efficiency, and the result would be identical. In this analysis, the product branching fractions are estimated for better comparability to the current PDG values (which are reported as product branching fractions). Furthermore, to estimate the total branching fractions for $B^- \rightarrow D^{**} \ell^- \nu$, one would need to use the branching

³In this case this is the number of events in the input PDF.

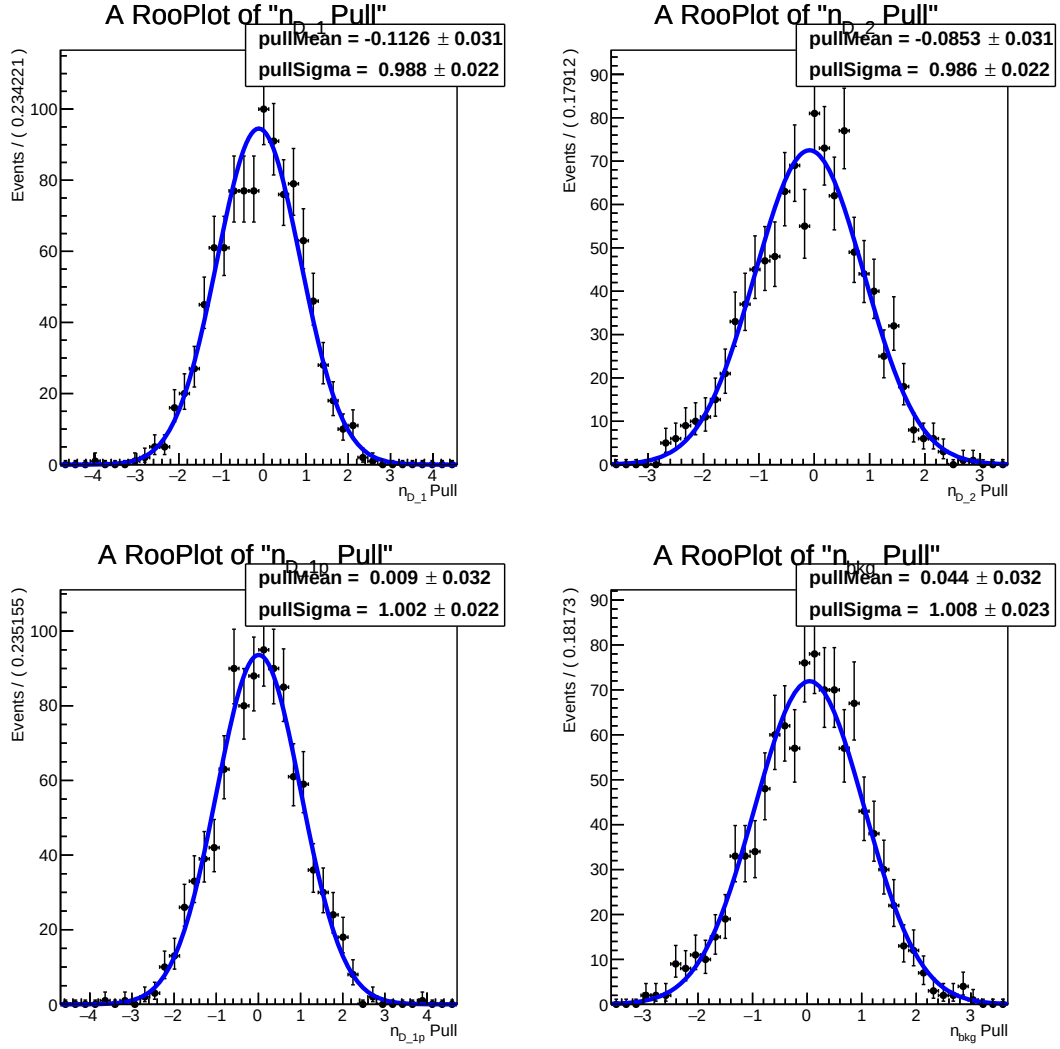


Figure 5.3: The pull distributions for the four fit parameters from fits to 1000 toy MC samples.

fraction for $D^{**} \rightarrow D^{*+}\pi^-$ in the calculation, adding to the uncertainty on the result. The resulting branching fractions are reported and discussed in the conclusion chapter (Ch. 8).

Chapter 6

Four Momentum Transfer (q^2) Study

Binned Fit in w

In order to obtain either the q^2 or the w distribution of the reconstructed decays, the fit which was detailed in the previous chapter (Chapter 5) was implemented in bins of the respective variable. q^2 values for the decays were calculated using equation 2.2 with the reconstructed B and D^{**} four-momenta. w was calculated from this q^2 value, using equation 2.3 with the PDG values of the B and D^{**} masses. Hence w is defined simply as a linear transformation of the reconstructed q^2 . It is worth mentioning that by not using the reconstructed masses of these particles in the w calculation, the finite width of these particles, along with resolution effects during the reconstruction are being neglected.

Since the aim is to extract the shape of the w distribution, one needs to implement the fit in more than one bin in this variable. Using more bins would allow the shape to be more thoroughly determined, however, the number of bins which can be used is limited by statistics. The decision on how to do the binning in w was made by testing the fit with different numbers of uniform bins on a separate fit sample of 700 fb^{-1} run independent MC (MC14)¹. As a test of how well the fit determines the result, the fractional statistical error on the fit results was evaluated in each case. The performance was vastly different for the three $B \rightarrow D^{**}\ell\nu$ decays due to the difference in the sample size. Decays into $D_2^{*}\ell\nu$, which had the fewest events, correspondingly showed large fractional errors for as few as three or four bins. Eventually, the decision was made to limit the fit to three bins.

Additionally, upon observing that the fit generally performed worse for the bins with a lower number of events, a switch from uniform to equal-frequency bins was made. Equal-frequency binning allows there to be sufficient statistics in each bin for a converging fit, although there are still variations in the number of events for each PDF category. Note that as a result, the exact positions of the bin boundaries fluctuate slightly depending on the PDF estimation MC sample used, and may need to be re-evaluated if the analysis is repeated with a different sample. However, the fluctuations are not expected to be drastic as long as the underlying decay model remains the same. As a further test of the number of bins, a fit with four equal-frequency bins was validated

¹This test was implemented in the early stages of the analysis, and not repeated later. This is why the MC sample used here was different than the ones used to obtain the other results reported in this thesis.

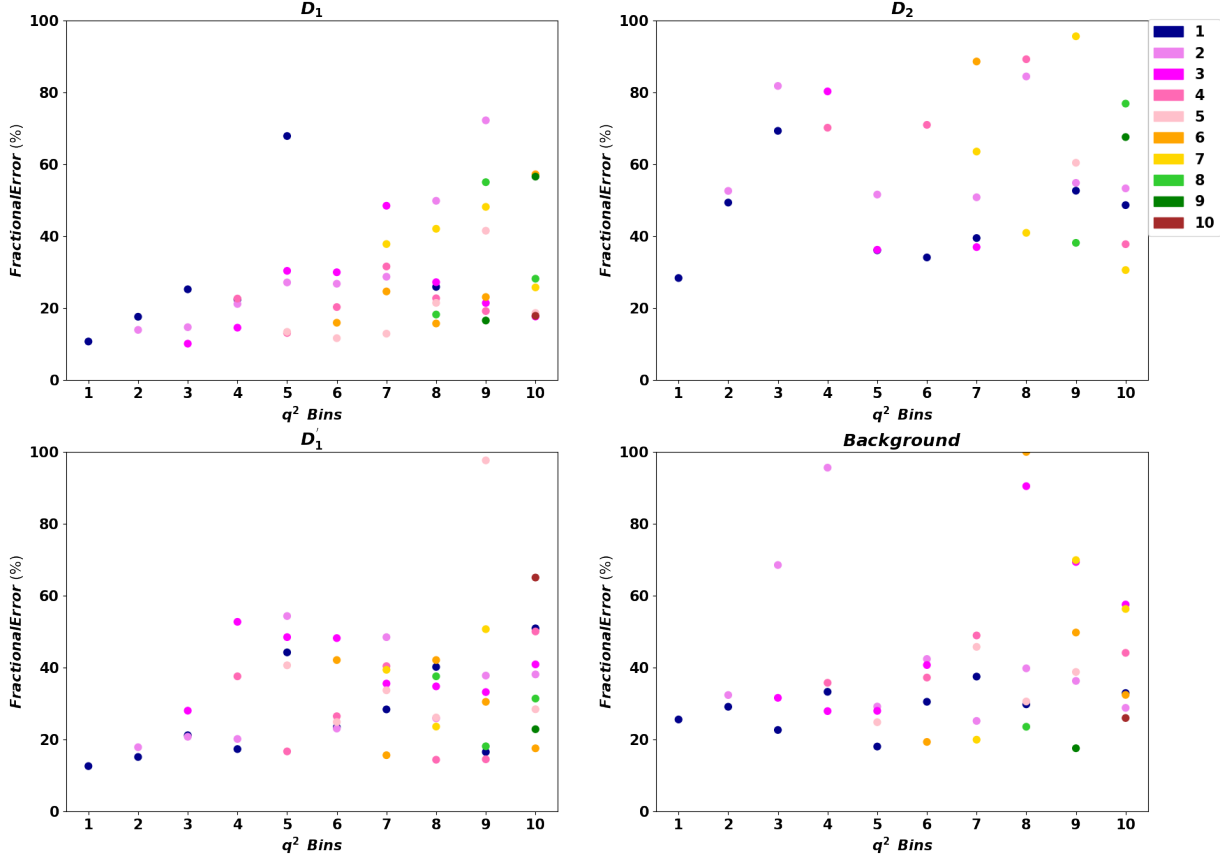


Figure 6.1: The fractional statistical uncertainties on the fit results for $\Delta M_{D^{**}}$ fits with different numbers of q^2 bins. The x-axis represents the total number of bins used in the fit. The uncertainties for the fits in different bins are identified by different colours. The numbering of the bins is from 1 to N (where N is the total number of bins), with increasing q^2 values.

following the method described in Sec. 5.3. Large biases were observed, confirming that the current amount of data is insufficient to perform the fit in more than three bins.

After splitting the PDF and fit datasets into three bins, the KDE and ML fits were repeated with each sample. Another option would be to produce a single PDF using only one of the w bins, or the whole w range. These could be used to fit all the w bins as long as the shape of the $\Delta M_{D^{**}}$ distribution does not differ drastically based on the w values. This option would be sensible if a larger number of bins was used. With only three bins, it was viable to produce separate PDFs for each bin. It was observed that this method was more accurate since the PDF shapes indeed showed differences for the different w bins. The results of the binned fit are given in Tab. 6.1, and the plots of the fitted functions for the three bins are given in Fig. A.2 and Fig. A.4 in the Appendix.

The fit was validated using the methods described in Section 5.3. The resulting pull plots are given in the Appendix (Fig. A.4, A.5, A.6). Some of the fits show biases, which will be taken as systematic uncertainties on the final results. For instance, the toy studies on the D_1 PDF in the

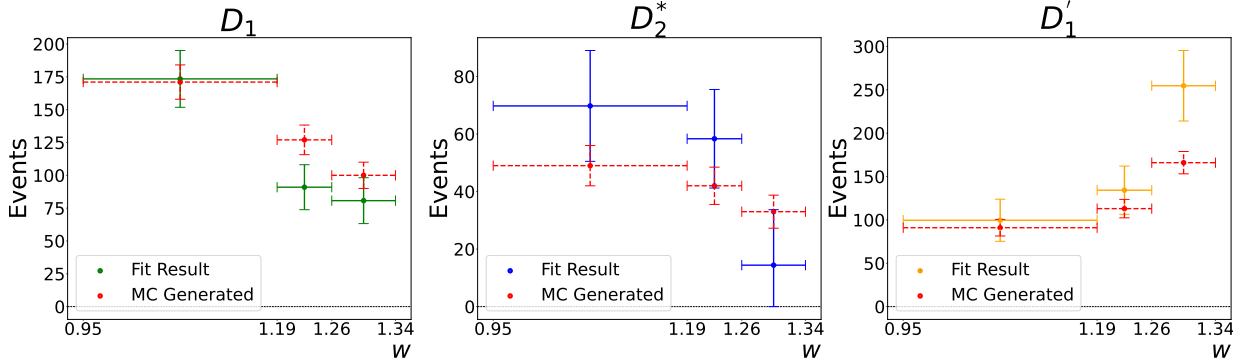


Figure 6.2: The yields obtained from the binned fit plotted alongside the MC generated yields.

w Values		n_{D_1}	$n_{D_2^*}$	$n_{D_1'}$	n_{bkg}
0.95 – 1.19	MC	171	49	91	60
	Fit	173.435 ± 21.630	69.765 ± 19.270	99.615 ± 24.274	28.198 ± 17.859
1.19 – 1.26	MC	127	42	113	66
	Fit	90.950 ± 17.128	58.350 ± 17.131	134.252 ± 27.828	64.471 ± 20.166
1.26 – 1.34	MC	100	33	166	124
	Fit	80.704 ± 17.472	14.408 ± 19.278	254.702 ± 40.675	73.186 ± 27.596

Table 6.1: Fit results for the $\Delta M_{D^{**}}$ fit in three bins of w .

second w bin resulted in a pull distribution with mean -0.136 ± 0.031 . This is a possible reason why the obtained fit result deviates from the MC value by $\sim 2.1\sigma_{stat}$. It is also observed that the fit in the third w bin results in $n_{D_1'}$ and n_{bkg} values which deviate by $\sim +2.2\sigma$ and $\sim -1.84\sigma$ from the MC yields. The correlation coefficient between these two fit parameters is -0.780 , therefore it is sensible that while $n_{D_1'}$ is overestimated, n_{bkg} would be underestimated. Qualitatively, the PDF shapes for these two decay modes are very similar in some regions, which might explain why the fit performs poorly for these two parameters. These inaccuracies in the fit results are not investigated further in this analysis due to time constraints.

6.1 Unfolding

The detector measures the momenta and energy of the particles with finite resolution. Therefore, the reconstructed kinematic properties of the particles are generally different than the generated (true) values. Hence, when observing the binned distribution of a reconstructed variable such as w , as long as the sizes of the bins are comparable to the resolution of the variable, one expects some events to be reconstructed in different bins than the one they were generated in. This effect is called *migration*. This could change the shape of the distribution significantly and needs to be corrected to obtain the true distribution.

6.1.1 Migration Matrix

The sizes of the migrations are represented by an $n \times n$ matrix, where n is the number of bins in the distribution, called the *migration* or *response* matrix M . The elements of the matrix are defined as in Eqn. 6.1, and they represent the frequentist probability that an event generated in bin i is reconstructed in bin j .

$$M_{ij} = \frac{N_{rec}^{(ij)}}{N_{gen}^j} \quad (6.1)$$

where $N_{rec}^{(ij)}$ is the number of events generated in bin j and reconstructed in bin i
 N_{gen}^j is the total number of events generated in bin j

Given the true distribution of the variable and the migration matrix, the reconstructed distribution is given by the matrix product of the two (Eqn. 6.2). Otherwise, if the reconstructed distribution is known, one needs to solve for the true distribution by inverting M_m in a process called *unfolding*.

$$\begin{aligned} N_{rec}^i &= \sum_{j=1}^{N_{bins}} M_{ij} N_{gen}^j \\ \Leftrightarrow \vec{v}_{rec} &= M \vec{v}_{gen} \\ \Leftrightarrow \vec{v}_{gen} &= M^{-1} \vec{v}_{rec} \end{aligned} \quad (6.2)$$

where $N_{rec/gen}^i \stackrel{\text{def}}{=} \text{Number of reconstructed/generated events in bin } i$
 $\vec{v}_{rec/gen} \stackrel{\text{def}}{=} [N_{rec/gen}^1 \dots N_{rec/gen}^{N_{bins}}]^T$

In this analysis, only the w distribution for signal $B \rightarrow D_1 \ell \nu$ decays were unfolded since the decay rate fit was only implemented for this decay. M was estimated using the 2.4 ab^{-1} charged $B\bar{B}$ sample. The generated values for w were calculated using the generated q^2 values and the PDG values of the B and D^{**} masses. Using the same binning as in the $\Delta M_{D^{**}}$ fit, the number of events in each reconstructed and generated bin were counted. The migration matrix obtained is given in Fig. 6.3. Note that the range of the reconstructed (w) and the true variable (w_{MC}) in the matrix are different, since w_{MC} only takes the kinematically allowed values of w , whereas the reconstructed w have been smeared to values outside this range. For the w axis the narrowest range that includes all events is used.

It is observed that the width of the two higher w bins were narrow enough to observe more than 15% migration. The resolution is worsened by the fact that the sample includes some non-truth-matched events that are still classified as signal by the custom signal-matching (Sec. 4.1). This is shown on the scatter plot in Fig. 6.4. One can observe that the decays where the signal B is truth-matched form a narrower diagonal line, and all the outliers are non-truth-matched. One also observes more outliers for higher w values, hence the resolution is w dependent for this sample.

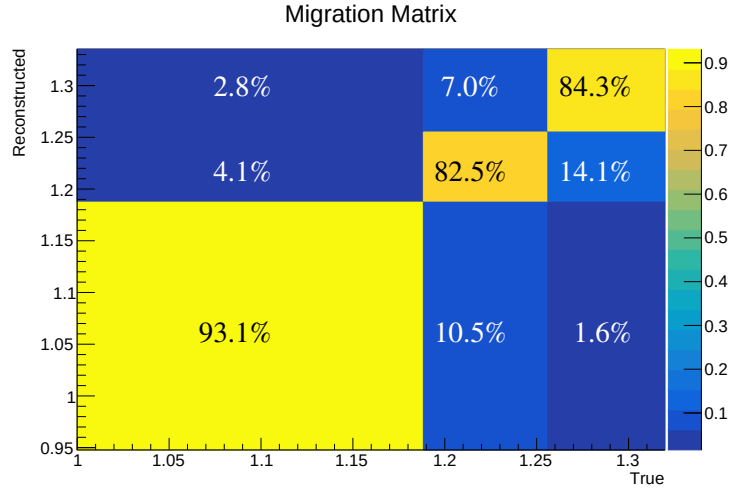


Figure 6.3: The migration matrix plotted as a 2D histogram. The colouring represents the values of the elements. The precise values are written on the bins.

Furthermore, in order to obtain estimates of the w resolution, the $w - w_{MC}$ distributions were fitted by a Gaussian function in each w bin, and the resulting width σ is taken. The fit was implemented separately for decays where signal B was truth-matched, and the remaining decays, since the latter is observed to have worse resolution. The resulting σ values were:

$(8.30 \pm 0.27) \times 10^{-3}$, $(7.51 \pm 0.41) \times 10^{-3}$, $(7.79 \pm 0.39) \times 10^{-3}$ for the truth-matched decays, and $(1.39 \pm 0.12) \times 10^{-2}$, $(2.03 \pm 0.18) \times 10^{-2}$, $(1.09 \pm 0.27) \times 10^{-1}$ for the remaining decays. These numbers should only serve

as an estimate, since for some samples, e.g. the third w bin of non-truth-matched decays, the histogram is not accurately described by the Gaussian. The fits are shown in Fig. 6.5

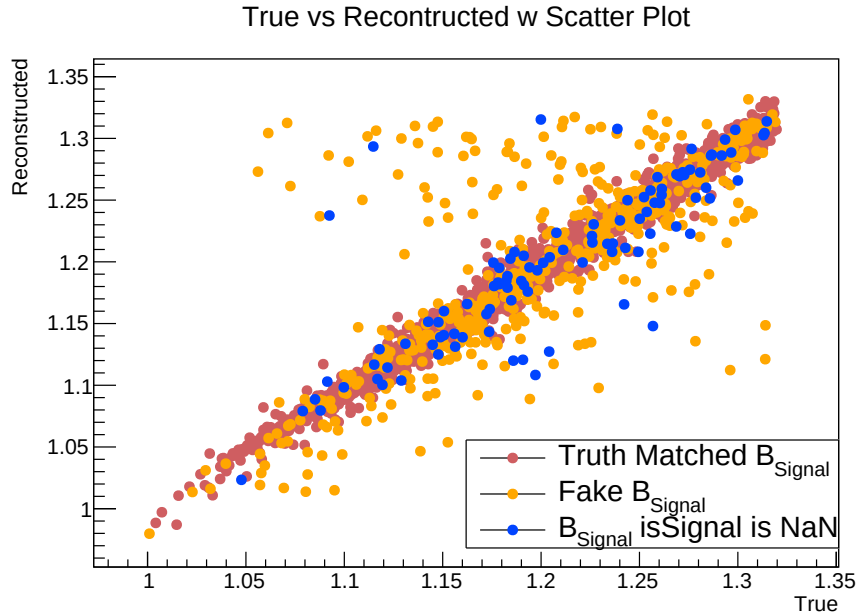


Figure 6.4: Scatter Plot of the true w_{MC} distribution versus the reconstructed w distribution for signal $B \rightarrow D_1 \ell \nu$ decays. Different colours represent different results of the signal B MC truth-matching.

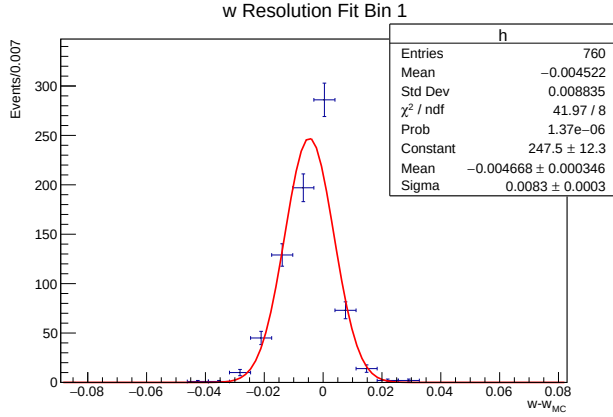
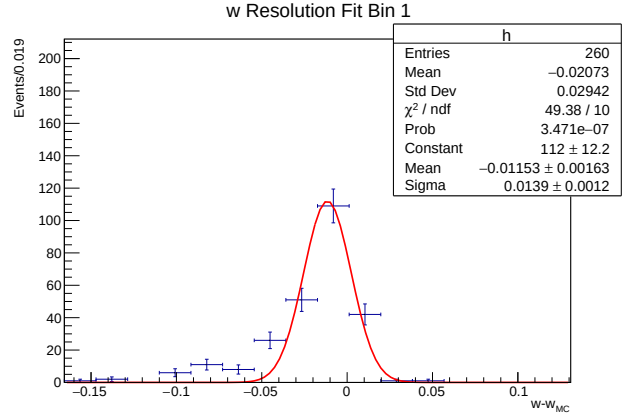
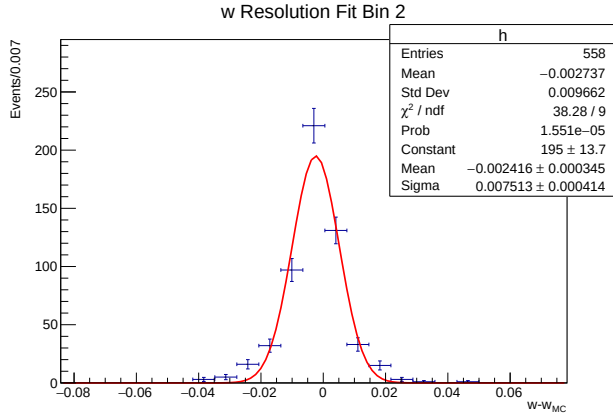
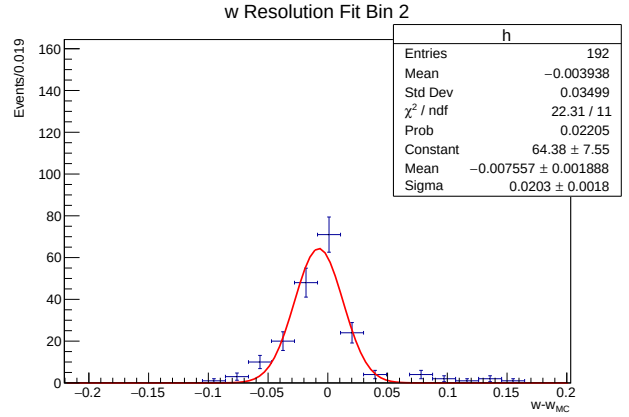
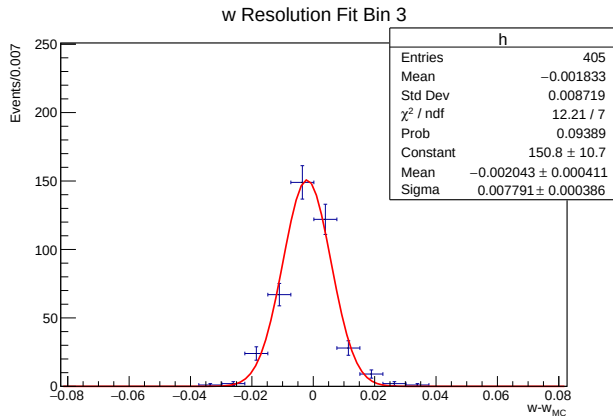
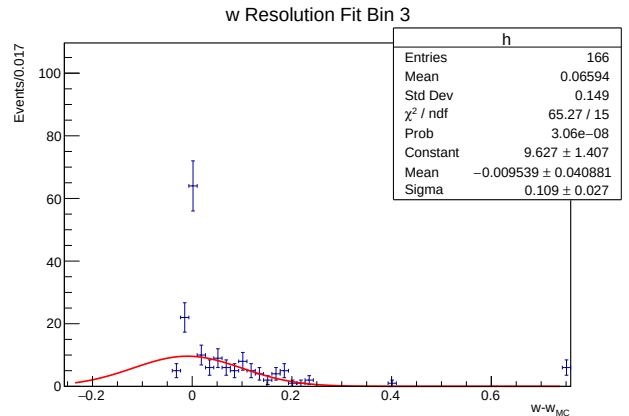
(a) B_{sig} truth-matched w bin 1(b) B_{sig} not truth-matched w bin 1(c) B_{sig} truth-matched w bin 2(d) B_{sig} not truth-matched w bin 2(e) B_{sig} truth-matched w bin 3(f) B_{sig} not truth-matched w bin 3

Figure 6.5: Gaussian Fits to the $w - w_{MC}$ spectrum for (left) truth-matched and (right) not truth-matched $B \rightarrow D_1 \ell \nu$ candidates in 3 w bins.

6.1.2 Unfolding Result

The fitted event yields for the $B \rightarrow D_1 \ell \nu$ decays (Table 6.1) were unfolded by the matrix inversion method shown in Eqn. 6.2. The unfolded distribution is plotted on Fig. 6.6, alongside the fit yields and the true w distribution from MC. Ideally, the unfolding would reproduce the true distribution. However, some bias is observed in the unfolded distribution. The biases for the three bins, defined as $(N_{\text{unfolded}} - N_{\text{true}})$ in this case were: 5.0 ± 23.6 , 33.6 ± 21.7 , and 22.3 ± 21.2 for bins with increasing w . Only the unfolded number of events contribute to the stated uncertainties, with the assumption that N_{True} is a known quantity with zero uncertainty. The calculation of the errors on the unfolded yields is described in Sec. 7.2.

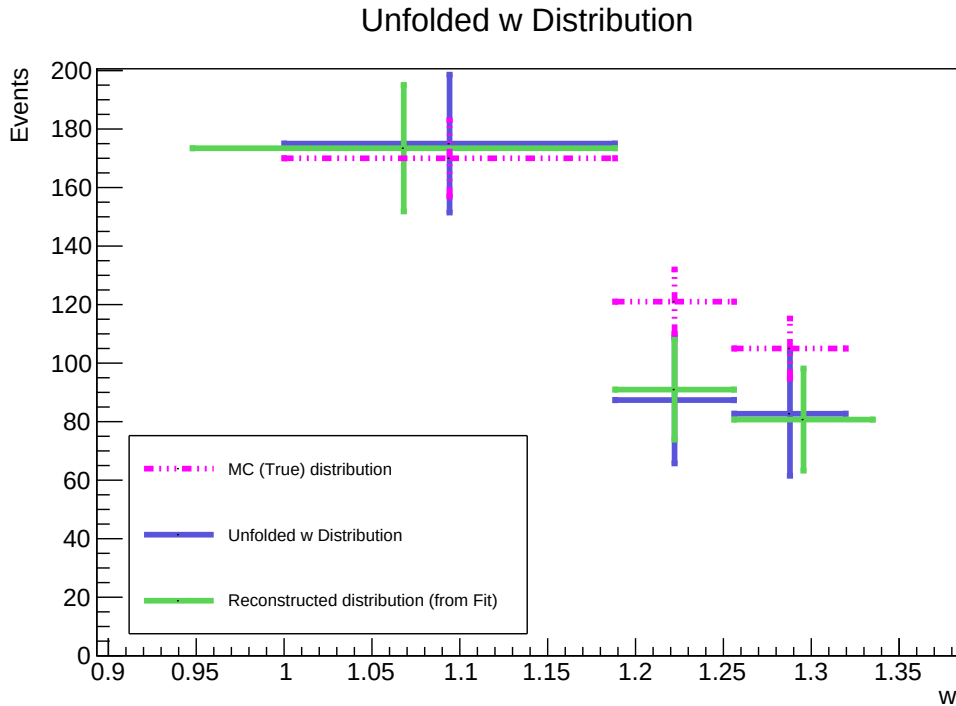


Figure 6.6: The number of events in each w bin obtained by the fit, after unfolding, and in the true MC distribution. The w range is larger for the fitted yields since these represent the reconstructed values, whereas the unfolded and the true distributions represent the true w distribution, which only takes kinematically allowed values.

As discussed in the previous section, the slight bias on the unfolded yields is likely to be introduced by the fit. To confirm that no additional bias is introduced by the unfolding, the reconstructed MC yields from the 400 fb^{-1} fit sample were unfolded using the same method. In this case the biases were 2.5 ± 14.4 , 7.1 ± 14.5 , and 2.5 ± 12.4 , which are all consistent with zero. Hence, the unfolding results are taken as they are.

Matrix inversion is a straightforward method of unfolding, and it yields unbiased results given that the unfolded distribution is sufficiently similar to the one used to estimate the migration matrix. However, for certain problems, e.g. when the migration matrix is singular, one needs

to consider different methods. There exist other methods of unfolding such as singular value decomposition (SVD) [31], Bayesian [26], iterative dynamically stabilised (IDS) [45] methods. Among these, the SVD method was tested at the early stages of the analysis. However, upon observing no improvement in the unfolding bias, this method was not adopted. The remaining unfolding methods might show improvement in the unfolding. However, the testing of these is beyond the scope of this analysis.

6.2 Decay Rate Fit

The w dependence of the $B \rightarrow D_1 \ell \nu$ decay is studied by fitting the decay rate in Eqn. 2.6 to the distribution obtained after the fit, unfolding and the efficiency correction. The model depends on four parameters, $\tau(1)$, τ' , $\hat{\tau}_1$, $\hat{\tau}_2$ ², the definitions of which were introduced in Ch. 2. Out of these parameters τ' , $\hat{\tau}_1$, $\hat{\tau}_2$ affect the shape of the differential decay rate distribution, whereas $\tau(1)$ affects the normalisation. Hence $\tau(1)$ may be calculated based on an assumption or measurement on the branching fraction of the decay, and the other three parameters may be estimated by a fit.

Due to limited statistics, the model is fitted to a distribution with three bins, which makes it challenging to simultaneously fit three parameters. Hence, only τ' is fitted and the other three are fixed to the values used in the MC simulation: $\tau(1) = 0.71$, $\hat{\tau}_1 = -0.5$, $\hat{\tau}_2 = 2.9$. This parameter was chosen because the shape is sensitive enough to this parameter to give a convergent fit. Moreover, the same parameter was fitted in a previous study by Belle, and the result may be compared to the result of that analysis.

The result is obtained by a χ^2 fit³ [12] implemented in ROOT. Before implementing the fit, the function $d\Gamma/dw$ is normalised to the total number of events in the histogram. During the fit, the function is first integrated over the w range of each bin, then divided by the bin width. Additionally, the values of the histogram are divided by the bin width, so that they represent the events rates rather than the event numbers. The resulting value is comparable to the decay rate: $\Delta N/\Delta w \propto d\Gamma/dw$. This also prevents possible problems in the fit due to uneven bin widths.

The fitted function along with the histogram is shown in Fig. 6.7. The result of the fit is $\tau' = -1.738 \pm 0.166$. This corresponds to a relative fit uncertainty of $\sim 9.8\%$. This uncertainty may also be observed on the plot, where the fitted curve is not in very good agreement with the histogram. Nevertheless, the fit resulted in a χ^2 probability or p -value of 0.132, which shows that this sample is still in agreement with the model⁴. Additionally, the fit result is consistent with the true (MC) value of $\tau' = -1.6$ within the fit uncertainty.

² $\tau(1)$ and τ' are more generally denoted $\tau_{3/2}(1)$, $\tau'_{3/2}$ e.g. in Ref. [44], since the values of the parameters are different (but related) for the $j = 1/2$ and $3/2$ states. Here the subscript is dropped since only the spin-3/2 D^{**} state is considered.

³The *modified least square* or *Neyman's χ^2* is used. This is the default in ROOT.

⁴This is based on the general criterion that a p value > 0.05 is not considered a significant deviation from the model.

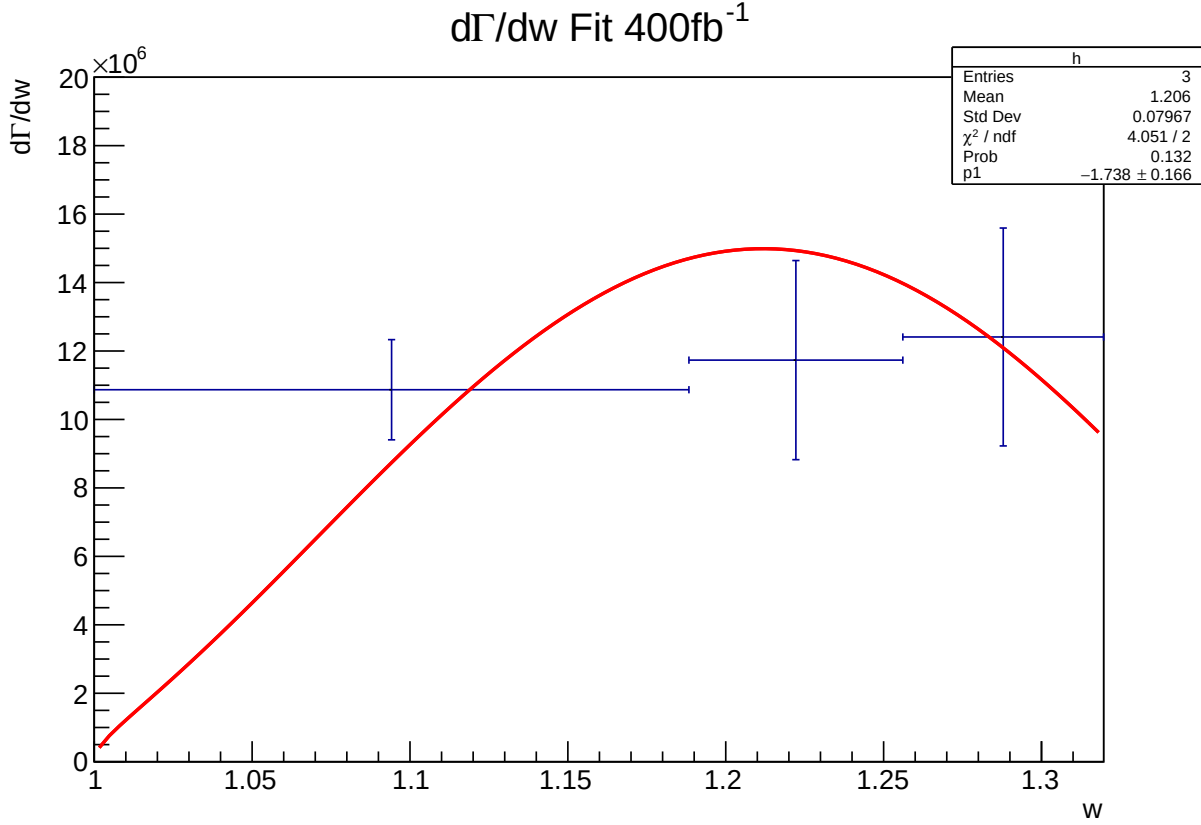


Figure 6.7: The $d\Gamma/dw$ fit result. The histogram represents the number of events in each w bin (after unfolding and efficiency correction) normalised by the bin widths. The red plot is the fitted function.

6.3 Extrapolation

It was mentioned repeatedly that the precision of the fit is limited by the statistics available with the 400 fb^{-1} fit sample. Hence, one might expect increased sensitivity to the τ/ℓ measurement with future Belle II data sets. According to the projections at the time of writing [15], Belle II is targeting to reach integrated luminosities of 1 ab^{-1} in 2024/25, 5 ab^{-1} before 2029, and 50 ab^{-1} around 2035.

To estimate how much improvement could be expected at these luminosities, toy MC samples were generated from the original $\Delta M_{D^{**}}$ PDF, corresponding to 1, 5, and 50 ab^{-1} . Additionally, 400 fb^{-1} toy MC samples were generated to compare to the nominal result. For each luminosity, 1000 samples were generated and fitted using RooMCStudy. The averages of the resulting yields were unfolded and the decay rate fit was applied as described in Sec. 6.1 and Sec. 6.2. The uncertainty on the fit results were taken to be the average of the uncertainties from individual toy fits. The results are given in Tab. 6.2, and the fitted functions and histograms are given in Fig. 6.8 and Fig. 6.9.

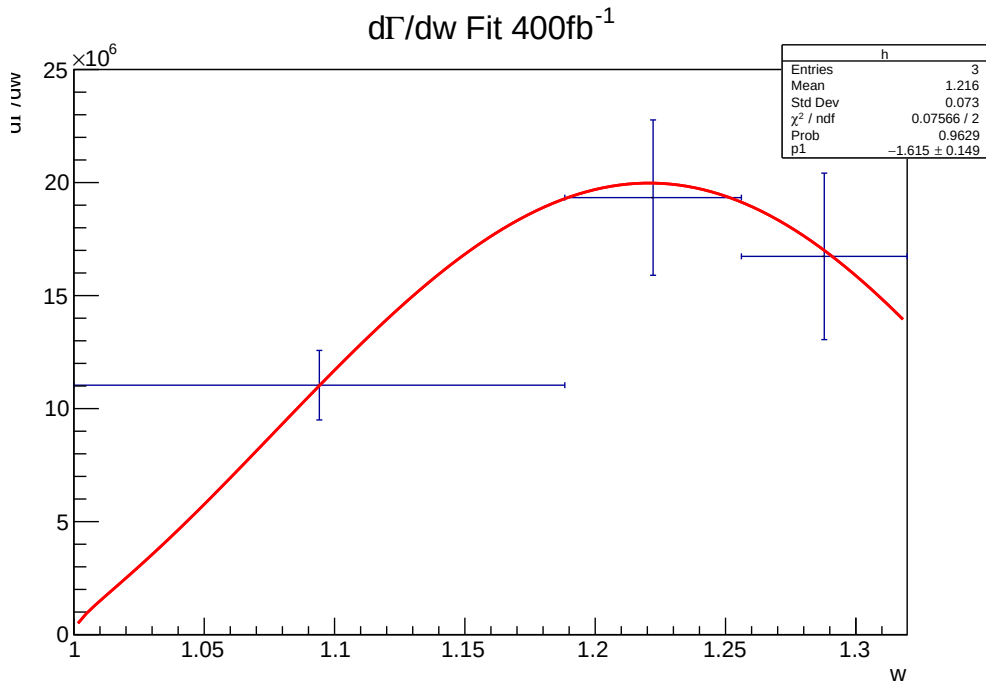
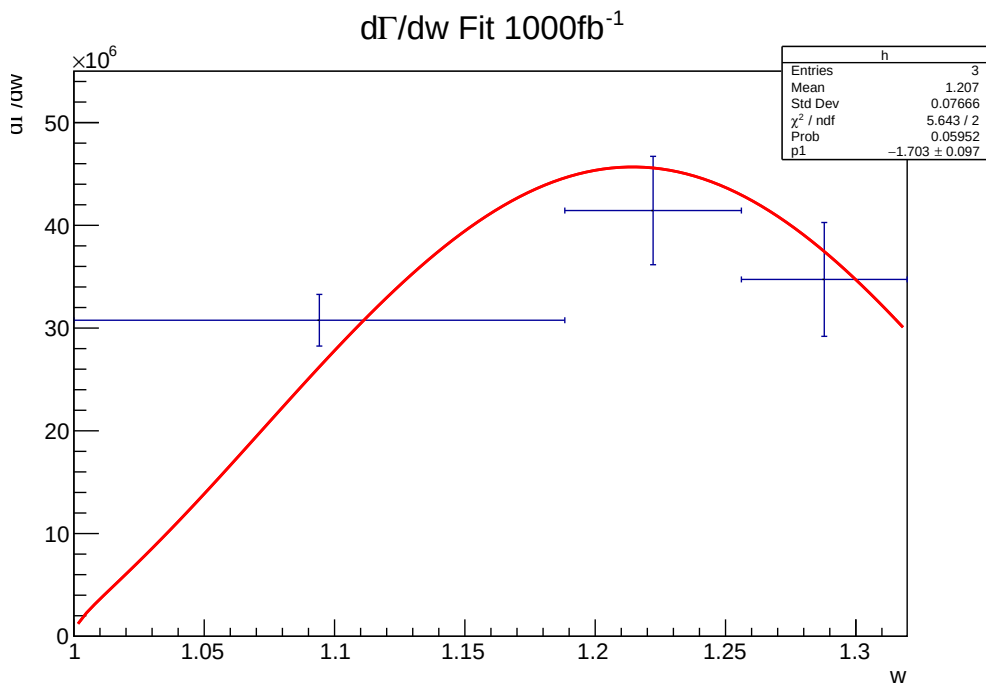
Luminosity (ab^{-1})	τ'	Fit Uncertainty
0.4	-1.615	0.149
1	-1.703	0.097
5	-1.661	0.046
50	-1.621	0.018

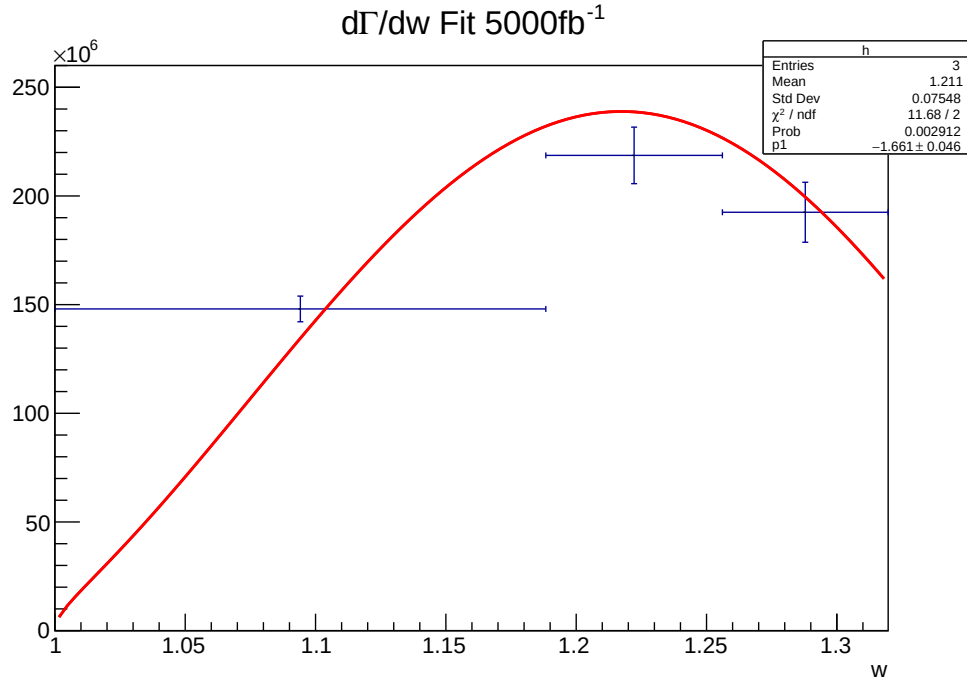
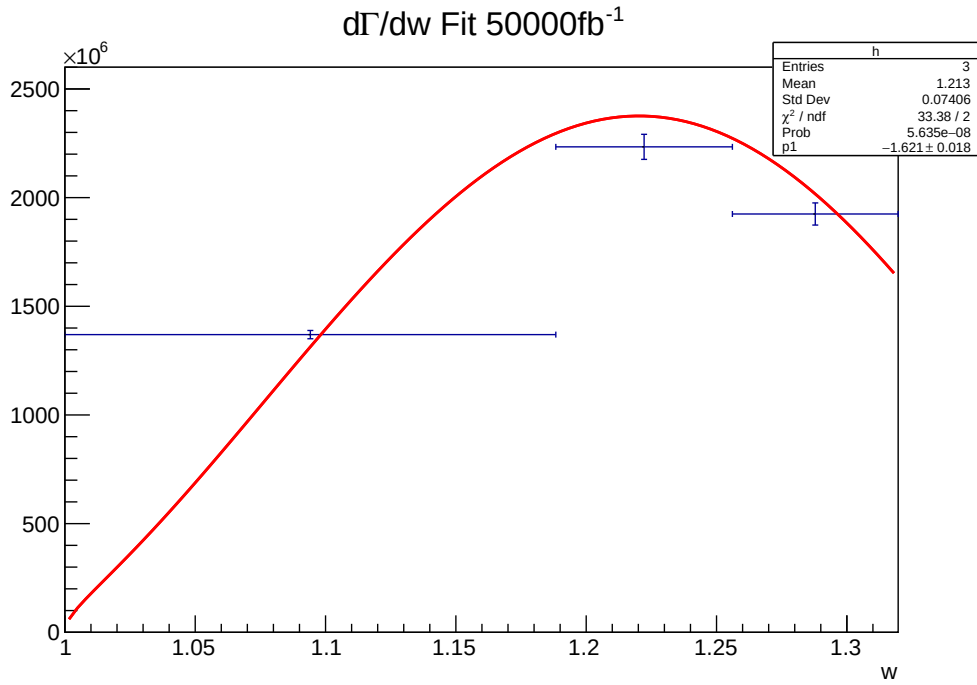
Table 6.2: τ' values obtained from fitting generated MC samples.

It is observed that for 400 fb^{-1} the relative uncertainty on the result from toy studies is lower than the nominal result, and the fitted value of τ' is closer to the MC value of the parameter. This is within expectations since the fit would perform better on samples generated from the PDF used in the fit. Correspondingly, one could expect the uncertainties for toy studies at higher luminosities to be lower than what should be expected on separate samples. Nevertheless, the steady decrease in the relative uncertainty from the 400 fb^{-1} toy MC samples to the 50 ab^{-1} gives an estimate of the relative reduction in the uncertainty that might be expected with future data sets. Moreover, as a confirmation that the fit performs better with higher luminosity samples, the result of the fit approaches the value of τ' in the simulation, which is -1.6 .

It should be pointed out that the fits at higher luminosities give progressively higher χ^2 values, and as a result, lower p -values. For the 5 ab^{-1} and 50 ab^{-1} samples $p < 0.05$, which indicates that the model fits the data poorly. The increase in χ^2 might be explained by the fact that the systematic errors were neglected at this stage. At low luminosities, since the statistical uncertainty is dominant, this was less influential on the quality of the fit. However, for higher luminosities, the systematic uncertainties become more significant, and adding these could improve the fit. However, this is not investigated in this analysis due to time constraints.

The $\Delta M_{D^{**}}$ fit here was implemented with the same binning for all luminosities for comparison. However, it should be possible to add more degrees of freedom to the fit by dividing the dataset into more bins for higher luminosities. This is expected to further improve the result since the shape of the w distribution may be better described by more bins. It might also be possible to simultaneously fit the other $d\Gamma/dw$ parameters. These potential improvements are beyond the scope of this analysis.

(a) Extrapolation to 0.4 ab^{-1} (b) Extrapolation to 1 ab^{-1} Figure 6.8: The $d\Gamma/dw$ fits with toy samples corresponding to 0.4 , and 1 ab^{-1} .

(a) Extrapolation to 5 ab⁻¹(b) Extrapolation to 50 ab⁻¹Figure 6.9: The $d\Gamma/dw$ fits with toy samples corresponding to 0.4, 1, 5, and 50 ab⁻¹.

Chapter 7

Systematic Uncertainties

In this chapter possible sources of systematic uncertainties are discussed, and the methods of estimating some of them are detailed. List of all the uncertainties considered for the branching fraction and τ' measurements are given on Tab. 7.2 and Tab. 7.1.

7.1 PDF Estimation Uncertainty

The PDFs for the $\Delta M_{D^{**}}$ fit were estimated based on MC simulation. The limited size of this MC sample is a source of systematic uncertainty on the fit results.

The resulting uncertainties on the branching fractions were estimated by bootstrapping the MC sample used for PDF creation to create 1000 samples. Bootstrapping refers to the method of sampling n points from the data sample, where n is the number of elements in the original sample, allowing repetition. The KDE and the $\Delta M_{D^{**}}$ fit on the 400 fb^{-1} MC sample were repeated with each bootstrapped PDF sample. The branching fractions were calculated, using the same reconstruction efficiency each time. The distribution of results for each D^{**} mode was fitted by a Gaussian, and the resulting standard deviation is taken as an estimate of the uncertainty on the result (see Fig. 7.1). The highest relative uncertainty is on the D_2^* branching fraction ($\sim 9.5\%$), which is expected as the MC sample used for the PDF creation is the smallest for these decays.

In order to evaluate the corresponding uncertainty on the τ' result, the same method of re-sampling the PDF sample was implemented for the binned fit, followed by the unfolding and the decay rate fit (see Fig. 7.2).

7.2 Migration Matrix Estimation Uncertainty

The elements of the migration matrix have statistical uncertainties depending on the size of the MC sample used to estimate them. Some simplifying assumptions have been made in order to estimate these uncertainties:

1. The number of events generated in a given w bin and reconstructed in bin i ($i = 1, 2, 3$) were treated as three multinomially distributed random variables, where the number of trials is

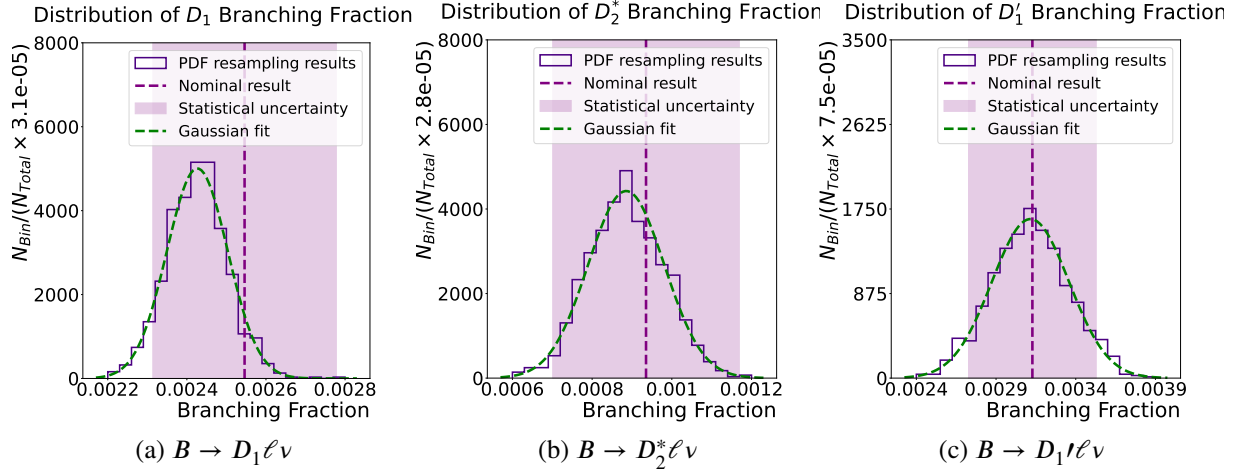


Figure 7.1: The resulting branching fractions after implementing the $\Delta M_{D^{**}}$ fit with 1000 bootstrapped PDF samples. The result of the nominal fit, along with the statistical uncertainties is plotted for comparison. The histogram is scaled so that the area under it is 1.

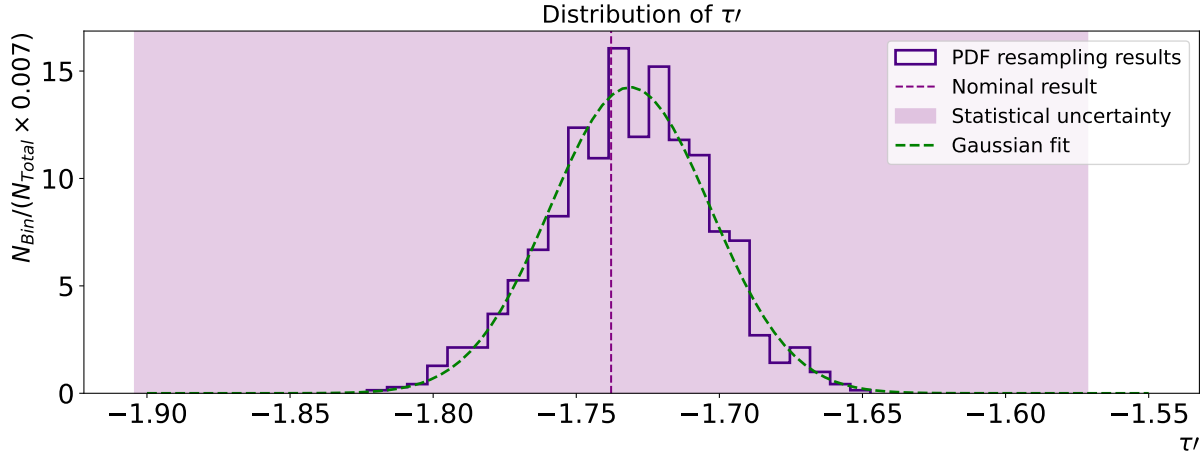


Figure 7.2: The resulting τ' values after implementing the $\Delta M_{D^{**}}$ fit with 1000 bootstrapped PDF samples. The result of the nominal fit, along with the statistical uncertainties are plotted for comparison.

N_{gen}^j (the total number of decays generated in w bin number j).

2. The total number of events generated in different bins were assumed to be uncorrelated i.e. $\text{Corr}(N_{gen}^k, N_{gen}^\ell) = \delta^{k\ell}$
3. The probability of reconstruction in bin i (for an event generated in bin j) was taken to be equal to the migration matrix elements $\left(\frac{N_{rec}^{(ij)}}{N_{gen}^j} \right)$.

Based on these assumptions, the variances and covariances of the $N_{rec}^{(ij)}$ values were calculated using formulas for the multinomial distribution [48]. Using these, the uncertainties on the migration matrix elements were calculated by Gaussian error propagation, described below (Eqn. 7.1). The resulting covariance matrix is given in Fig. 7.3(a).

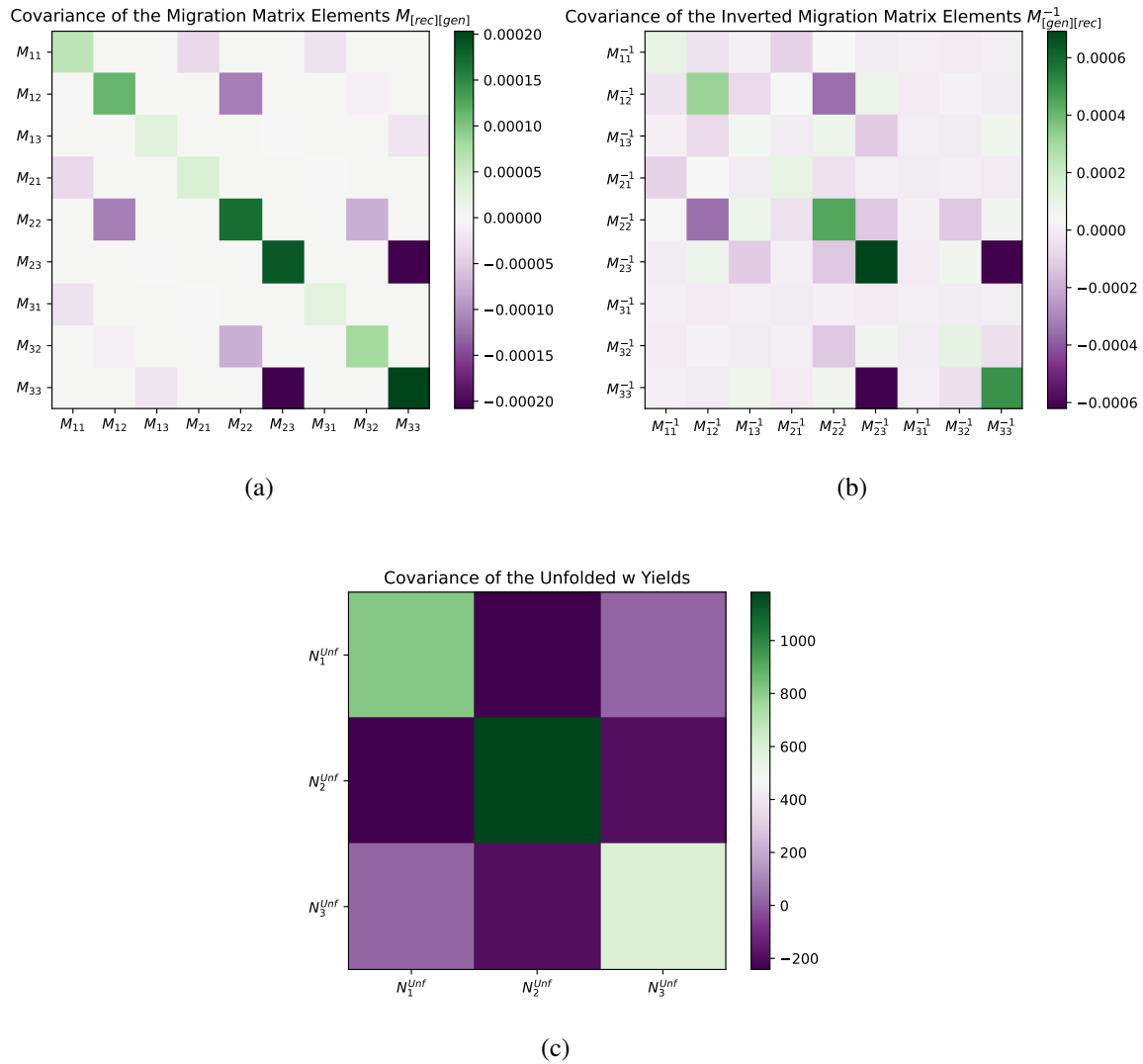


Figure 7.3: The covariance matrix for (a) the elements of the migration matrix, (b) the elements of the inverted migration matrix, (c) the unfolded yields.

Gaussian Error Propagation

Given a differentiable function f of variables $\{x_i\}_{i=1,\dots,N}$ with uncertainties $\{\sigma_i\}$, and covariances $Cov(x_i, x_j)$, the error on f is:

$$\sigma_f = \sqrt{\sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 + \sum_{i \neq j}^N \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) Cov(x_i, x_j)} \quad (7.1)$$

Another method of obtaining the migration matrix estimation uncertainties is by bootstrapping the MC sample used to produce the migration matrix, and see the variation of the matrix elements for the different samples. This is tested with 1000 bootstrapped samples and the results agree with the calculated uncertainties. This is a confirmation that the assumptions made in the calculation accurately describe the data sample.

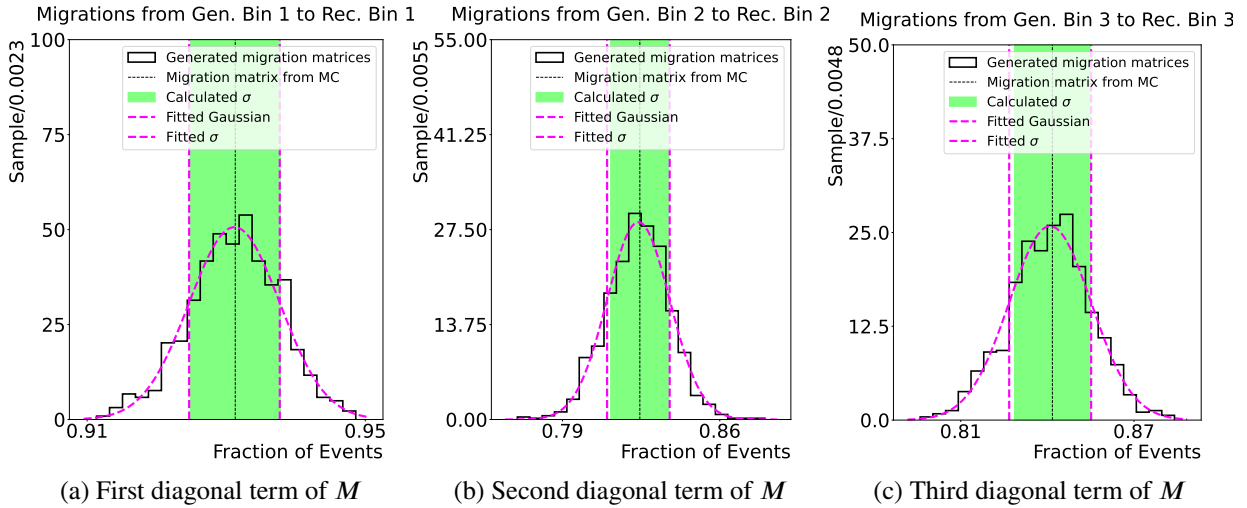


Figure 7.4: The diagonal elements of the 1000 generated migration matrices from the bootstrapped samples. The black dashed line shows the element of the original migration matrix, and the green shaded region shows the uncertainty on this value.

During the unfolding, the migration matrix is inverted, which introduces correlations to some of the non-diagonal elements (see Fig. 7.3(b)). Additionally, the uncertainty on the fit results should be considered when calculating the uncertainty on the unfolding result. These calculations were done using equation (7) and (11) of Ref. [42].

7.3 Data MC Correction Uncertainties

The performance of certain methods used in the reconstruction may vary between data and MC simulation. Consequently, the efficiency and purity of the selection may be mismodelled in the MC simulation, and these generally need to be corrected when working with data. Since the results reported in this analysis are based on simulation only, these corrections do not need to be applied. The uncertainties introduced to the results by these corrections are nevertheless investigated. The corrections considered here are the slow π reconstruction efficiency, the mid to high momentum track reconstruction efficiency, leptonID corrections, and the FEI efficiency calibration.

7.3.1 Slow π Reconstruction

The reconstruction efficiency for the slow π tracks shows slight discrepancy between data and MC in the low momentum region. The correction weights are measured in three bins of the slow π momentum by the collaboration. These are updated regularly with new measurements, and the weights used here are based on the Moriond 2023 data set [16]. The bin boundaries are 0.05, 0.12, 0.16, 0.20 GeV/c, and the correction weights for the bins in order of increasing p_π are 0.996 ± 0.022 , 0.990 ± 0.017 , and 0.987 ± 0.019 . The uncertainties given are statistical and are uncorrelated between the bins. In addition, there is a correlated statistical uncertainty of 0.015, and a systematic uncertainty of 0.003 on each bin, which are assumed to be fully correlated between the bins.

To investigate the uncertainty due to these corrections, 1000 variations of the weights were generated, by randomly sampling from Gaussians with widths equal to the uncertainties listed above. With the uncorrelated uncertainties, the weights were varied independently in each momentum bin, whereas with the correlated uncertainties, every bin was varied by the same randomly sampled amount.

In order to estimate the uncertainty on the branching fractions and the τ/ℓ , first the nominal correction weights are applied to the fit sample. Secondly, the $\Delta M_{D^{**}}$ fit is repeated using the 1000 variations to reweight the PDF sample. The varied weights are only applied on the PDF sample since the aim is to investigate the case where the fit is implemented on data, and estimate the effect of mismodelling of the corrections on the PDF sample, which is obtained from MC. The resulting distributions of the branching fractions and τ/ℓ are given in Fig. 7.5 and 7.6. The distribution mean is shifted compared to the nominal result, since the nominal result is without any corrections applied, whereas the distributions show the variation around the corrected result.

7.3.2 Mid to High Momentum Tracking Efficiency

The performance of the reconstruction of high momentum tracks varies between data and MC simulation. The per-track discrepancy in the efficiencies was measured by the collaboration to be $1 - \epsilon_{data}/\epsilon_{MC} = (0.0607 \pm 0.24)\%$ [30]. Since this is consistent with 0, no correction is applied, i.e. $\epsilon_{data}/\epsilon_{MC} \stackrel{\text{def}}{=} 1$ and only a per-track uncertainty is assigned.

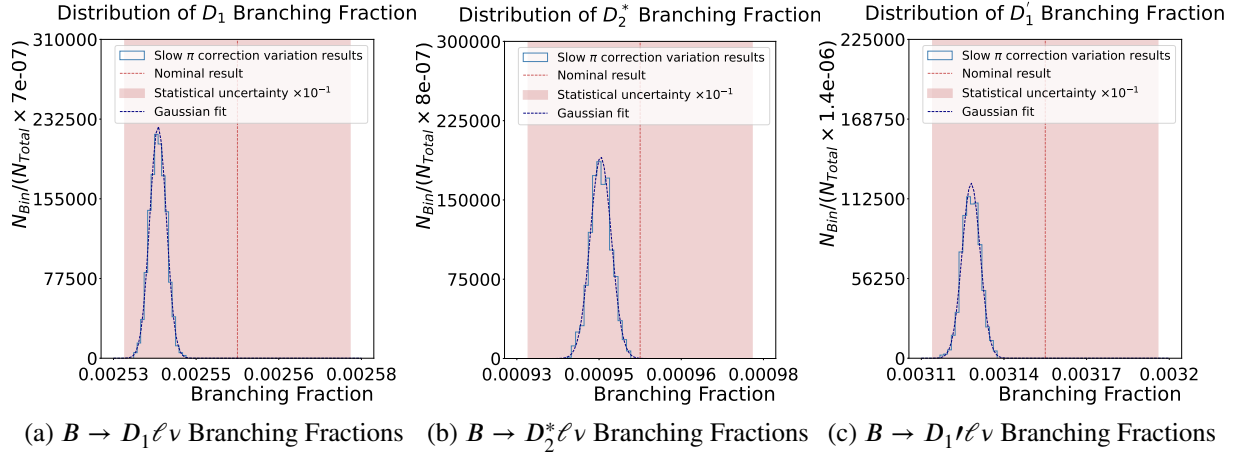


Figure 7.5: The resulting branching fractions after implementing the $\Delta M_{D^{**}}$ fit with 1000 variations of the slow π efficiency correction weights. The result of the nominal fit, along with the statistical uncertainties are plotted for comparison.

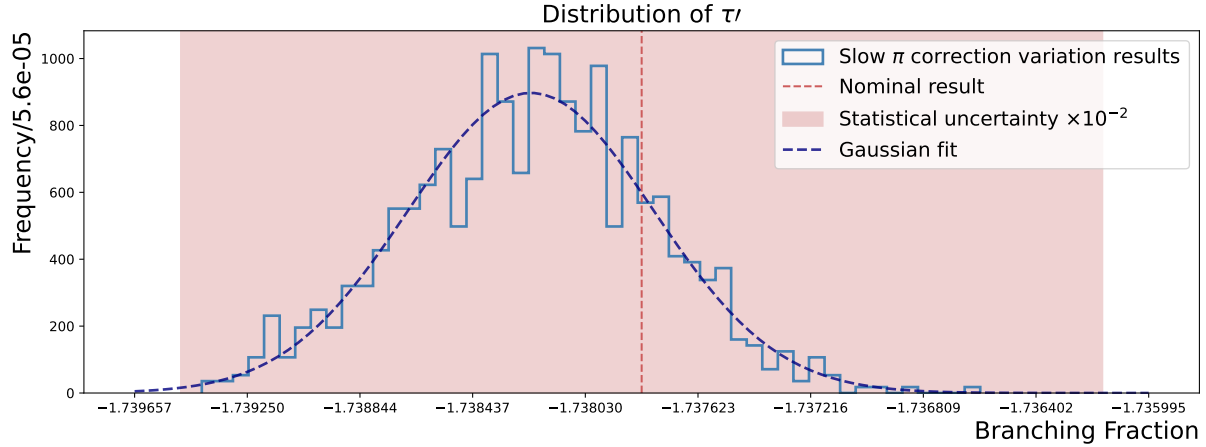


Figure 7.6: The resulting τl values after implementing the $\Delta M_{D^{**}}$ fit with 1000 variations of the slow π efficiency correction weights. The result of the nominal fit, along with the statistical uncertainties are plotted for comparison. The statistical error on the nominal fit is scaled by 10^{-2} .

The resulting correction factor for signal reconstruction efficiency is given by this value taken to the power of the number of tracks in the reconstruction. Here only the tracks on the signal side are taken into account, since the correction for the tag side (including the track reconstruction efficiency) is applied separately. In this analysis, the reconstructed signal channels have either 5 or 7 tracks in the final state. An average of the calibration factors for the two cases is taken, weighted by the number of decay channels for each case (N_5 and N_7). To obtain these numbers, the number of decays with 5 tracks and 7 tracks in the final state were counted in the 2.4 ab^{-1}

PDF sample.

Since the per-track correction factor is 1, so is the correction factor on the signal efficiency. The uncertainty on this value can be estimated by varying the per-track correction up or down by the uncertainty, and checking how much it varies. In other words, equation 7.2 can be solved for σ_e^{sig} .

$$\delta_{sig} \pm \sigma_e^{sig} = \frac{N_5(\delta_{track} \pm \sigma_e^{track})^5 + N_7(\delta_{track} \pm \sigma_e^{track})^7}{N_5 + N_7} \quad (7.2)$$

where $\delta_{sig} = \frac{\epsilon_{Data}}{\epsilon_{MC}} = 1$ and $\delta_{track} = \left(\frac{\epsilon_{Data}}{\epsilon_{MC}} \right)_{\text{per track}} = 1$

This uncertainty is not significant for the τ' result. This is because the uncertainty on the correction is independent of the track momentum for high momentum tracks (neglecting the slow pion correction which was discussed above).

7.3.3 LeptonID

Since cuts have been applied on electronID and muonID during the selection, the uncertainties due to the corrections on these variables should be estimated. The correction weights, defined as $\epsilon_{data}/\epsilon_{MC}$ are given as functions of the lepton's momentum in the laboratory frame (p), polar angle θ , and charge. These weights, along with their uncertainties are provided by the dedicated group. To investigate the uncertainty due to these corrections, 1000 variations of the weights were generated, by random sampling from two Gaussians, the widths of which equal the statistical and systematic uncertainties on the correction weights. The statistical uncertainties are assumed to be uncorrelated for each bin in $p \times \theta$. Whereas the systematic uncertainties are assumed to be fully correlated, and the same random variation is applied to each bin.

The corresponding uncertainties on the branching fractions and τ' are calculated as done for the slow π correction uncertainty (Subsec. 7.3.1). The resulting distributions are shown in Fig. 7.7. The variation of the τ' values are given in Fig. 7.8. The uncertainty on τ' is $\sim 0.1\%$ which is 1% of the statistical uncertainty on this result.

7.3.4 FEI Calibration

As described in Subsec. 4.3.1, the tag side reconstruction by FEI considers a large number of decay channels, and utilises multiple classifiers. As a result, the reconstruction efficiency may differ significantly between data and MC, which needs to be corrected. The value of the correction or calibration factor is measured by the collaboration. Here the result of a 2020 FEI calibration study [5] was taken directly as an estimate of the expected uncertainty. The study used hadronic FEI for tag side reconstruction, and reconstructed $B \rightarrow X\ell\nu$ decays on the signal side. The tag side reconstruction efficiencies were compared for data and MC. The resulting calibration factor for charged B reconstruction was measured to be 0.65 ± 0.02 . Some of the systematic uncertainties included in the result were the uncertainty on the $B \rightarrow X\ell\nu$ branching fractions,

form factor parameters, tracking and particle identification efficiency corrections, and the size of the MC sample used in the study.

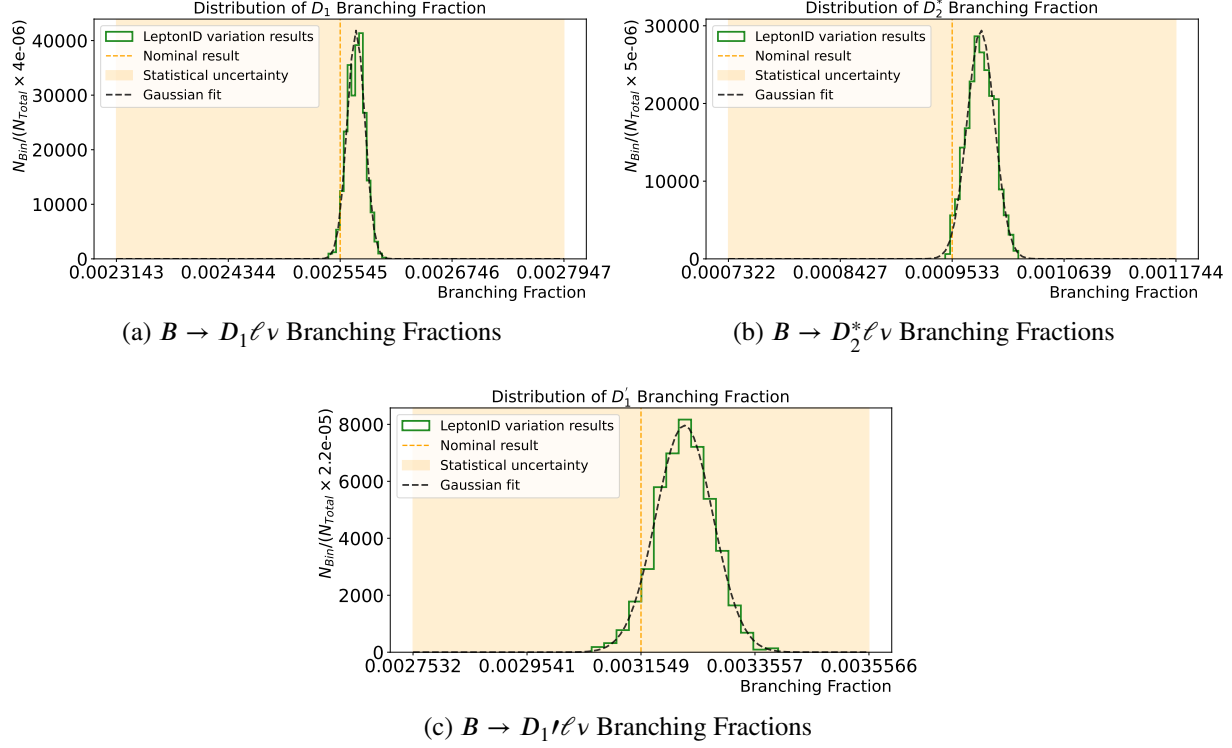


Figure 7.7: The resulting branching fractions after implementing the $\Delta M_{D^{**}}$ fit with 1000 variations of the leptonID weights.

7.4 Other Uncertainties

7.4.1 Fit Bias

As detailed in Subsec. 5.3, the fit validation for the $B \rightarrow D_1 \ell \nu$ and $B \rightarrow D_2^* \ell \nu$ decay yields showed biases in the result. These biases were considered a systematic uncertainty on the branching fraction results. Although the biases are one sided, symmetric uncertainties assigned as estimates. Additionally, the uncertainty on the bias is added to the systematic uncertainty. The resulting relative errors on the branching fractions are 1.34% and 3.22%.

The binned $\Delta M_{D^{**}}$ fit also showed some bias on the $B \rightarrow D_1 \ell \nu$ yields. The observed biases were $(-1.02 \pm 0.41)\%$, $(-2.60 \pm 0.59)\%$, $(-1.25 \pm 0.76)\%$ of the fit results for bins with increasing w . The biases are likely to be correlated, since the same fitting method is applied to each bin. Nevertheless the amount of biases are not the same. This hints that the correlation is not likely to be 100%. For a more accurate estimate of the resulting uncertainty on the τ' result, the χ^2 fit

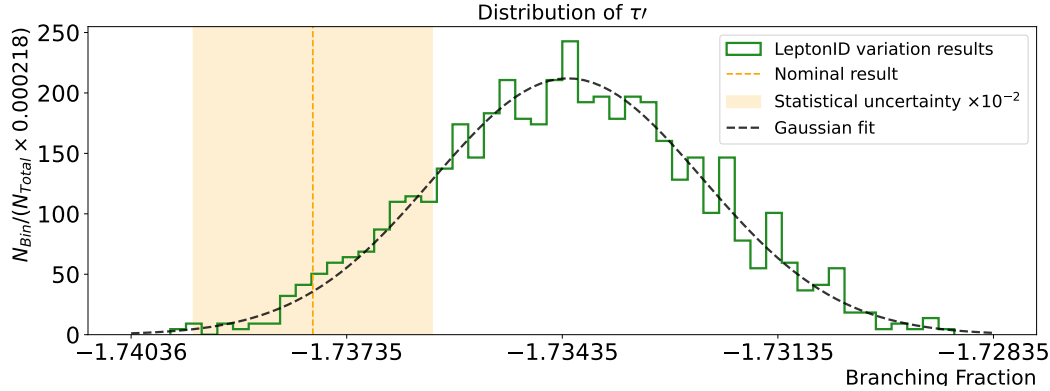


Figure 7.8: The resulting τ' values after implementing the $\Delta M_{D^{**}}$ fit after reweighting the PDF sample with 1000 variations of the leptonID corrections. The result of the nominal fit, along with the statistical uncertainties are plotted for comparison. The statistical error on the nominal fit is scaled by 10^{-2} .

could be repeated, this time taking into account the correlation between the bins, as described in Ref. [24]. This is not implemented due to time constraints, and instead the method described below was followed.

Before making an assumption on how the binned yields are correlated, the effect of varying them in different combinations on the τ' fit result was investigated. Seven possible variations were tested, where the yield in each bin is either varied by the bias plus the uncertainty on the bias, or kept the same. The maximum negative variation on the τ' result was 0.55% of the nominal τ' result and occurred when the yield of the first bin was varied down by the bias, and the yield of the other bins were unaltered. Whereas, the maximum positive variation was 0.43%, and occurred when the yield for the third bin was unaltered and the other two were varied. These results are shown on Fig 7.9.

Eventually, it was decided to assign an uncertainty based on the assumption of uncorrelated bins. Hence the variations in τ' as a result of separately varying each bin (first three points in Fig. 7.9) were added in quadrature to obtain an uncertainty estimate of 0.0121(0.70%).

7.4.2 $\Upsilon(4S)$, D^{*+} and D^0 Branching Fractions

The branching fractions for the $\Upsilon(4S) \rightarrow B^- B^+$, $D^{*+} \rightarrow D^0 \pi^+$ and the nine D^0 decay channels were used in the calculation of the $B \rightarrow D^{**} \ell \nu$ branching fractions (see equation 5.10). These uncertainties cancel out during the calculation of the reconstruction efficiency, since the uncertainty affects both the generated and reconstructed number of events from simulation. However, in the case where the branching fractions are calculated from data, the uncertainty on the $\Upsilon(4S)$, D^{*+} , and D^0 branching fractions would lead to an uncertainty in the result. This is estimated using the relative uncertainties on the PDG values of the branching fractions, and propagating this to the final result.

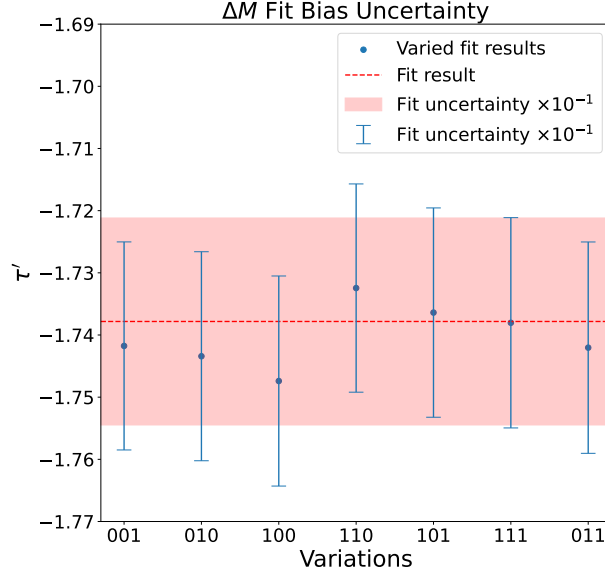


Figure 7.9: The resulting τ' values after varying the $\Delta M_{D^{**}}$ fit yields by the biases. The x-axis labels denote whether the yield from the fit in the w bins were varied by the bias. If there is a 1(0) in position x , the yield in bin x was varied(not varied) by the bias.

7.4.3 Decay Model Parameters

The differential decay rates of the simulated $B \rightarrow D^{**}\ell\nu$ decays depend on the four parameters of the BLR model. The uncertainty on these variables comprises an uncertainty on the reconstruction efficiency, since the efficiency is dependent on w . This uncertainty may be estimated by varying the decay model parameters within their uncertainties and re-weighting the w distribution, or by generating signal MC using varying values for the parameters. These were not included in the current study due to time constraints.

Nevertheless, the effect of form factor uncertainties on the migration matrix and the unfolding result was studied. This effect is not expected to be significant, since the migration matrix depends on the resolution of the detector, and not on the initial model used for generation. To investigate this, 1000 variations of the migration matrix were produced by varying the form factor parameters. In order to obtain the w_{true} distributions, N_{gen} points were generated from the $d\Gamma/dw$ function by varying the four parameters. The parameters were varied by random sampling from a multivariate Gaussian with widths equal to the fit uncertainties given in Ref. [18]. The correlations between the parameters were taken into account.

The generated distributions were smeared by sampling points from the $w - w_{MC}$ histogram obtained from the 2.4 ab^{-1} PDF sample. Initially, it was attempted to fit this histogram by a Gaussian¹, and use this for smearing. However, the generated migration matrices by this method were significantly different than the one used in the analysis. Using the histogram resulted in the

¹Another attempt was made to fit the histogram by a sum of two Gaussian distributions to model the wider, non-truth matched component. This did not yield better results.

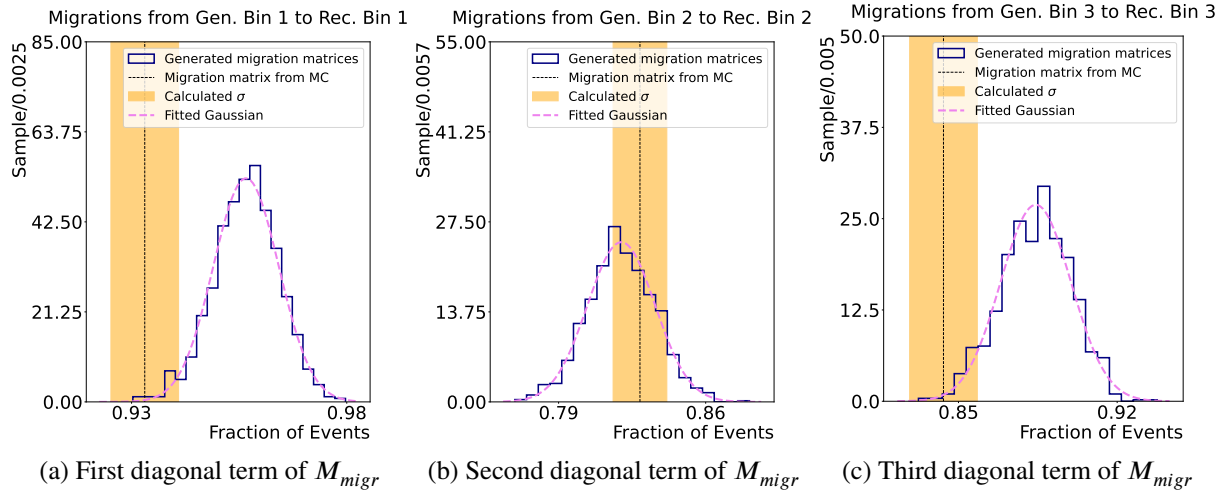


Figure 7.10: The diagonal elements of the 1000 generated migration matrices with randomly varied form factor parameters. The black dashed line shows the element of the original migration matrix, and the orange shaded region shows the uncertainty on this value.

distributions of migration matrix elements some of which are shown in Fig. 7.10. As observed on the plots, it is probable to obtain the original migration matrix from these distributions.

Furthermore, it is observed that the variation of the migration matrix elements with varied decay model parameters is of the same order as the calculated statistical uncertainty on these values. From this one can conclude that the unfolding does not depend significantly on the decay model parameters.

Summary Tables

The systematic uncertainties estimated on the τ' result and the branching fractions are given in Table 7.1 and Table 7.2 below.

Uncertainty on τ'		
PDF	Absolute	Relative (%)
LeptonID Efficiency	0.00188	0.108
Slow π Efficiency	0.00044	0.026
$\Delta M_{D^{**}}$ Fit Bias	0.01210	0.696
Total	0.04242	2.441

Table 7.1: List of systematic uncertainties on the τ' result.

Uncertainty on D^{**} Branching Fractions						
	D_1		D_2^*		D_1'	
	Abs ($\times 10^{-5}$)	Rel(%)	Abs ($\times 10^{-5}$)	Rel (%)	Abs ($\times 10^{-5}$)	Rel (%)
PDF Creation	7.97	3.119	9.01	9.456	24.23	7.682
LeptonID Correction	0.95	0.373	1.36	1.422	5.01	1.588
Slow π Efficiency	0.18	0.069	0.21	0.219	0.32	0.102
Tracking Efficiency	3.36	1.316	1.25	1.316	4.15	1.316
FEI Calibration	7.66	3.000	2.86	3.000	9.46	3.000
$\Delta M_{D^{**}}$ Fit Bias	3.41	1.335	3.07	3.222	0.00	0.000
Branching Fractions	0.34	0.132	0.42	0.132	0.13	0.132
Total	23.87	9.344	17.89	18.766	43.52	13.794

Table 7.2: List of systematic uncertainties on the D^{**} branching fractions.

Chapter 8

Conclusion and Discussion

This analysis aimed at studying the method of measuring $B \rightarrow D^{**}\ell\nu$ branching fractions separately in D_1 , D_2^* and D_1' modes, and studying their q^2 distributions. A hadronic tagging approach was adopted in order to reconstruct the momentum of the tag B. This tagging approach also resulted in having only one neutrino in the event, coming from the signal side, which allows one to use the missing mass squared (M_{miss}^2) or u variables for the selection. After the reconstruction and selection, the yields of signal and background decays were obtained by a maximum likelihood fit to the $\Delta M_{D^{**}}$ distribution, and the branching fractions were estimated. Additionally, the fit was repeated in three bins of w . The obtained yields were unfolded and fitted by the differential decay rate model by a χ^2 fit. Finally, some of the systematic uncertainties on the results were estimated.

The final results and uncertainties obtained are given on Tab. 8.1, along with the MC values for comparison. The analysis was not implemented on data due to time constraints. Hence, some of the questions raised in the Introduction and Motivation chapters need to be investigated by a future analysis on data. The discussion here is limited to the comparison of the values used in the MC simulation to the obtained results, and the comparison of the sensitivity of the results to past analyses.

	$B(B^- \rightarrow D^{**}\ell^-\nu) \times B(D^{**} \rightarrow D^{*+}\pi^-)$		Uncertainties	
D^{**}	MC (%)	Fit (%)	$\sigma_{stat}(\%)$	$\sigma_{sys}(\%)$
D_1	0.3023	0.2555	0.0240 (9.40%)	0.0239 (9.34%)
D_2^*	0.0996	0.0953	0.0221 (23.20%)	0.0179 (18.77%)
D_1'	0.2873	0.3155	0.0402 (12.73%)	0.0435 (13.79%)
Parameter	MC	Fit	σ_{stat}	σ_{sys}
τ'	-1.6	-1.738	0.166 (9.6%)	0.042 (2.4%)

Table 8.1: Summary of final results and uncertainties

The resulting branching fractions for the decays into D_2^* and D_1' are within $0.19\sigma_{stat}$ and $0.7\sigma_{stat}$ of the MC values. Whereas the branching fraction of the decay into D_1 deviates from the MC value by $1.95\sigma_{stat}$. A portion of this is explained by the bias in the fit, which corresponds to a

relative bias of $(1.05 \pm 0.29)\%$ on the branching fraction result. However, correcting the measurement by this bias plus the uncertainty on the bias (1.34%) still gives a discrepancy of $-1.81\sigma_{stat}$ from the MC value. This is not investigated further since the discrepancy is sufficiently low to be due to a fluctuation. The observation that the $B \rightarrow D_1 \ell \nu$ branching fraction result obtained is shifted in the negative direction compared to the distribution of results obtained by bootstrapping the PDF (see Fig. 7.1(a) in the Systematics chapter) supports the prediction that this might be a fluctuation. Additionally, it is worth noting that adding only the systematic uncertainty due to the PDF estimation reduces the discrepancy to $1.42\sigma_{stat}$.

The biases observed in the fits show that there is room for improvement in the fitting method of this analysis. These biases are possibly a consequence of low statistics in the MC sample used for PDF creation. This is supported by the fact that ML estimators are expected to converge to the true values with increasing sample size [24, p. 70]. Using a larger MC sample in a future analysis would likely reduce the biases. Additionally, it would improve the precision of the measurements, since the limited size of the PDF sample produces the leading systematic uncertainty on all the results (see Table 7.1 and Table 7.2). Another viable option is to use a function to fit the ΔM spectrum. This option may be more advantageous than simulating more samples since the MC sample used in this analysis was already sizable at 2.4 ab^{-1} . However, this option would bring about new uncertainties concerning the accuracy of the model used, and these may surpass the PDF estimation uncertainty due to the MC sample size.

The precision obtained on the branching fraction measurements of the decays into D_1 and D'_1 were comparable to the 2008 BaBar analysis conducted on a 417 fb^{-1} data sample [11]. To compare the statistical uncertainties obtained by the present analysis to the ones obtained by BaBar, these need to be scaled by $1/\sqrt{N_{BB}^{\text{BaBar}}/N_{BB}}$ or $1/\sqrt{\mathcal{L}_{BB}^{\text{BaBar}}/\mathcal{L}_{BB}}$. Applying the scaling gives relative statistical uncertainties of 9.2% and 12.5% on the branching fractions of decays into D_1 and D'_1 . BaBar obtained results of $0.29 \pm 0.03(10.34\%) \pm 0.03(10.34\%)$ for the branching fraction of $B \rightarrow D_1 \ell \nu$, and $0.27 \pm 0.04(14.89\%) \pm 0.05(18.56\%)$ for the branching fraction of $B \rightarrow D'_1 \ell \nu$ [11]. Here, and in the following paragraphs, the errors stated first are statistical and the ones stated second are systematic. Hence, the current analysis showed some improvement in the statistical uncertainties. The comparison of systematic uncertainties is non-trivial since BaBar has reported some uncertainties which have not been estimated in this analysis. For instance, they have reported the uncertainty due to the modelling of the broad D^{**} states to be 4.5 – 13.8% [11]. The uncertainty due to the decay model parameters were not estimated in the present analysis, as mentioned in Subsec. 7.4.3, which may alone make up for the lower systematic uncertainty reported here compared to the BaBar result.

The branching fraction of the decay into $D_1 \ell \nu$ may also be compared to the 2008 Belle analysis. They reported the branching fraction to be $0.42 \pm 0.07(16.67\%) \pm 0.07(16.67\%)$ based on an analysis on $657 \times 10^6 \text{ } B\bar{B}$ pairs. Scaling the statistical uncertainty of the present analysis, which was performed with $\sim 216 \times 10^6 \text{ } B\bar{B}$ pairs, gives $\sim 5.4\%$, which is less than one-third of the uncertainty obtained at Belle. Hence the current analysis achieves lower statistical uncertainties than both the 2008 Belle and 2008 BaBar analyses on the branching fraction of $B \rightarrow D_1 \ell \nu$.

On the decay into the other narrow state, Belle reported the product branching fraction to be $0.18 \pm 0.06(33.33\%) \pm 0.03(16.67\%)$. This was the first observation of a semileptonic B decay

to D_2^* , and only obtained a 35 ± 11 yield of these decays. Scaling the uncertainty obtained in this analysis to the luminosity of the Belle dataset results in a relative uncertainty of 13.30%, which gives higher precision than the Belle result.

BaBar reports the branching fraction $\mathcal{B}(B \rightarrow D_2^* \ell \nu) \times \mathcal{B}(D_2^* \rightarrow D\pi^\pm) = 0.15 \pm 0.02(13.33\%) \pm 0.01(6.67\%)$. They use the following ratio which may also be used to compare the result of the current analysis to the one BaBar has reported:

$$\frac{\mathcal{B}(D_2^* \rightarrow D\pi)}{\mathcal{B}(D_2^* \rightarrow D^*\pi)} = 2.2^{[11]} \quad (8.1)$$

Scaling the result of the current analysis and the uncertainties by this ratio, and adjusting the scaling of the statistical uncertainty by the square root of this ratio gives $0.210 \pm 0.032(15.32\%) \pm 0.039(18.7\%)$. Therefore the statistical uncertainty on this result is slightly worse than the one obtained by the BaBar analysis. The reasons for this are not investigated further in this analysis.

The method of tagging allowed the momentum of the tag B to be reconstructed more reliably compared to other possible approaches, such as semileptonic tagging or an inclusive study. Nevertheless, it resulted in lower reconstruction efficiency. As a result, enough sensitivity was not reached with the 400 fb^{-1} sample to implement the q^2 study for the decays into D_2^* . Although a high yield of $B \rightarrow D_1' \ell \nu$ decays were obtained, the q^2 study was not extended to this mode either, due to time constraints.

As a result of implementing the study on the w distribution of decays into D_1 , a τ' value of -1.738 ± 0.166 was obtained on the 400 fb^{-1} MC sample. This result is within $\sim 0.82\sigma$ of the MC value of -1.6 . Again, as a consequence of the limited size of the fit sample, the result has a high statistical uncertainty of 9.6%.

The precision obtained on the τ' result has nonetheless been improved compared to the 2008 Belle measurement. They obtained a result of $-1.8 \pm 0.3(16.67\%)$, from a fit to the decays into the D_2^* and D_0^* modes. A relative statistical uncertainty of 5.49% would be expected from the current analysis at the luminosity used at Belle. The increase in precision obtained by the current analysis follows from the much higher yield of the signal decays obtained compared to the Belle analysis. Since Belle did not report systematic uncertainties on this result, these can not be compared.

Furthermore, the extrapolation of the τ' fit showed that significant improvements can be expected in the precision of the result with future Belle II data sets with integrated luminosities of 1 ab^{-1} , 5 ab^{-1} , and 50 ab^{-1} . The relative fit uncertainty was lowered to $\sim 1.11\%$ with the generated 50 ab^{-1} . Nevertheless, this improvement should be taken with a grain of salt, since the quality of the fit was lowered as the extrapolation was taken to higher luminosities. As discussed in Subsec. 6.3, this might have been an artefact of not including systematic uncertainties in the fit. This is due to any mismatch of the PDF sample and the fitted model becoming more significant as the statistical uncertainty is lowered with increasing MC sample size. The further investigation of the systematic uncertainties on the τ' result, and their influence on the extrapolation study are possible topics for a future analysis.

There are other potential improvements and extensions to the study which could be undertaken by a future analysis. Concerning the reconstruction of $Y(4S)$ decays, the ROE mask cuts could be optimised further. The cuts on the beam background MVA variables could also be optimised.

In order to stay consistent through the course of the analysis, the `hadronicSplitOffSuppression` variable was used in the selection. Nevertheless, the current suggestion by the collaboration is to use the `fakePhotonSuppression` variable instead. This should be adjusted in a potential continuation of the analysis. For all variables used in the selection, the Data-MC agreement was not studied in this analysis. This should be undertaken by a future analysis to validate the use of these variables in the selection.

A detailed study of the types of background decays was not included in the current analysis, and this is justified by the low background obtained by the selection. A big fraction of the so-called background decays is made up of the non-resonant $B \rightarrow D^{*+}\pi^-\ell\nu$ decays. As discussed in the previous chapters, the treatment of these decays may be re-evaluated in future analyses.

Furthermore, it was not thoroughly investigated why the Belle II truth-matching algorithm did not perform as expected with the $B \rightarrow D^{**}\ell\nu$ decays. For this analysis, it was deemed sufficient to use a custom truth-matching instead. Nevertheless, it may be useful input for the improvement of the MC truth-matching algorithm to study the reasons behind its poor performance in this case. At the very least, it might point to an error made in this analysis while interpreting the results of the truth-matching, which might help clarify the usage of the algorithm for future analysts.

Concerning the estimation of results and uncertainties, some improvements are possible to the binning in w if the study is repeated on a larger sample. Additionally, different unfolding methods could be investigated, as mentioned in the respective chapter. On a separate note, it could be investigated whether the PDF estimation uncertainty is dependent on the value of the ρ parameter. The study of this systematic in this analysis only considered fixed and optimal values of ρ , given in Tab. 5.1.

In conclusion, this master's thesis has reported on the method of measuring the branching fractions and the q^2 transfer of the poorly understood $B \rightarrow D^{**}\ell\nu$ decays with the Belle II experiment. It was shown that the analysis method described here could improve on the precision of past measurements. Consequently, significant contributions could be made to the understanding of these decays with current and future Belle II data sets. For the time being, the discrepancies surrounding the measurements and theory predictions related to these decays persist and provide an exciting avenue to improve our understanding of the Standard Model.

Appendix A

Diagrams

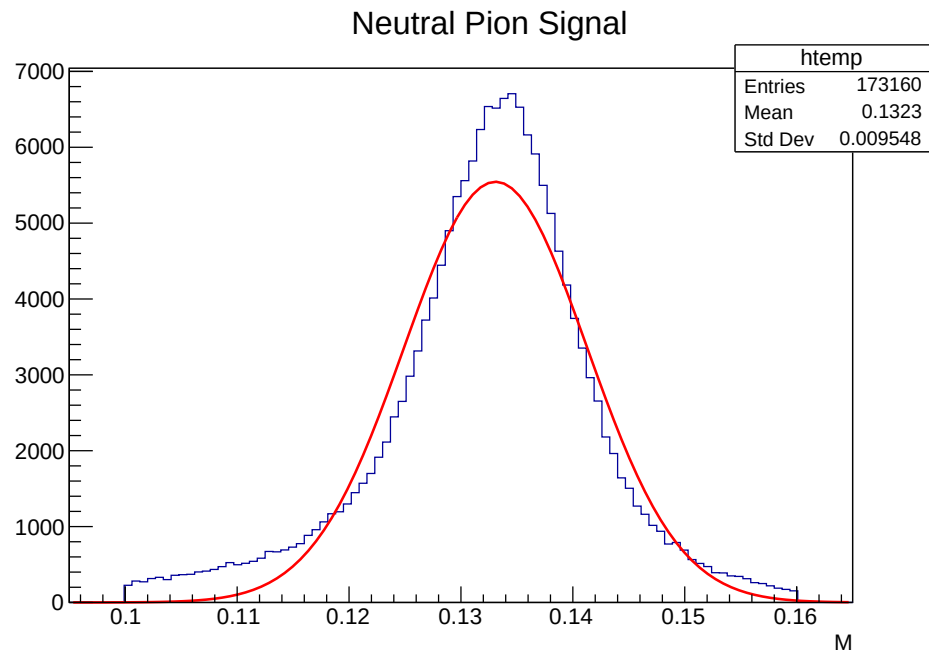


Figure A.1: Gaussian Fit to the π^0 mass spectrum for truth-matched $\pi^0 \rightarrow \gamma\gamma$ decays.

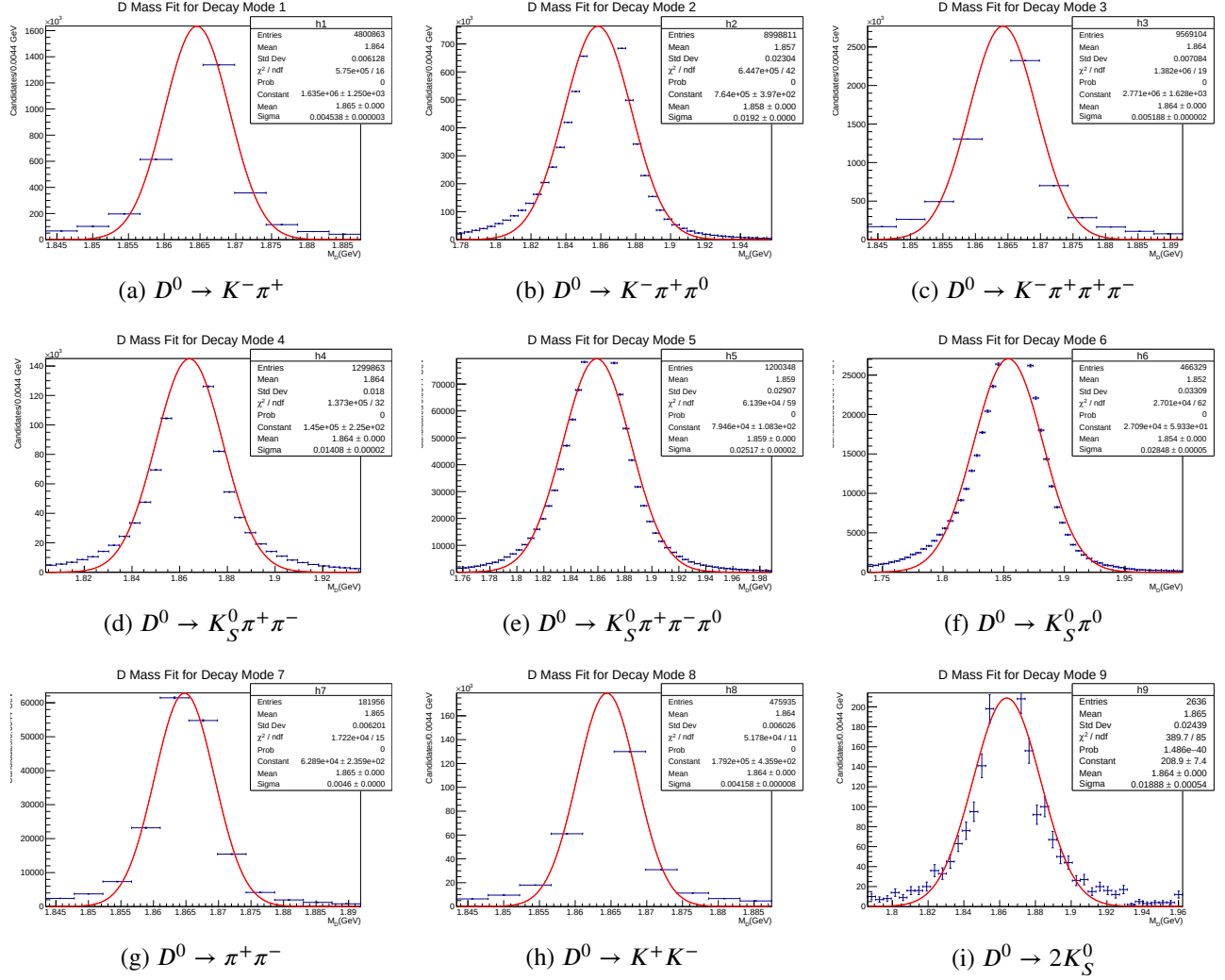


Figure A.2: Gaussian Fits to the D^0 mass spectrum for 9 decay modes, given in the plot labels. The distributions only include truth-matched D^0 decays.

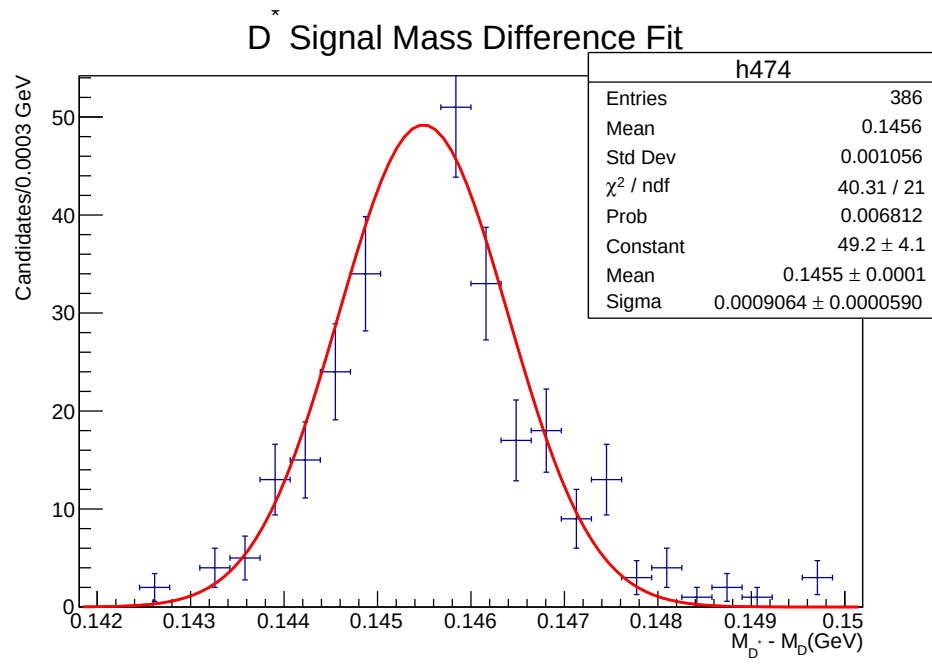


Figure A.3: Gaussian Fit to the D^* ΔM spectrum for truth-matched $D^{*\pm} \rightarrow D^0 \pi^\pm$ decays.

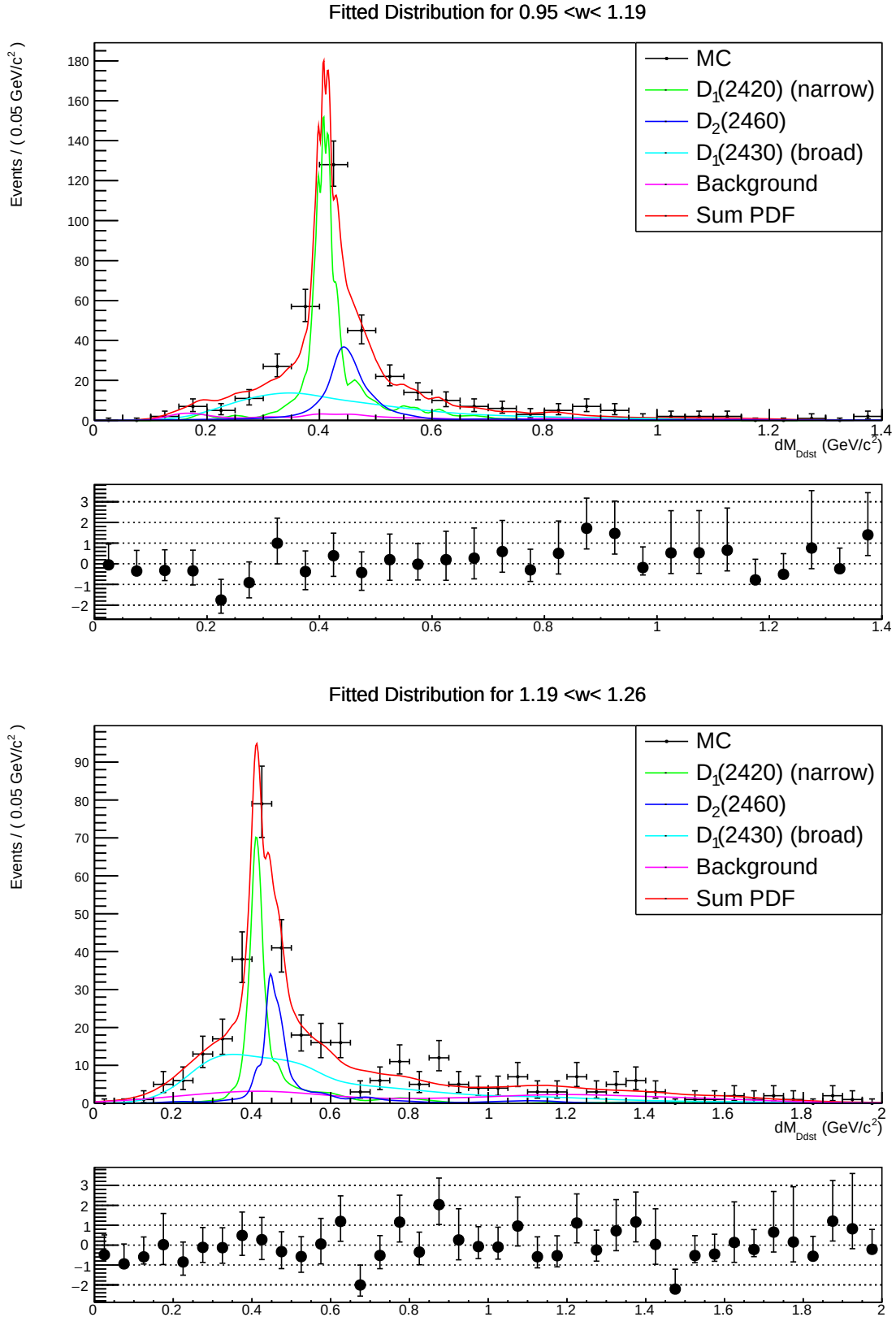


Table A.2: Fitted PDFs and Pull plots for the ΔM fits in the first two bins in w .

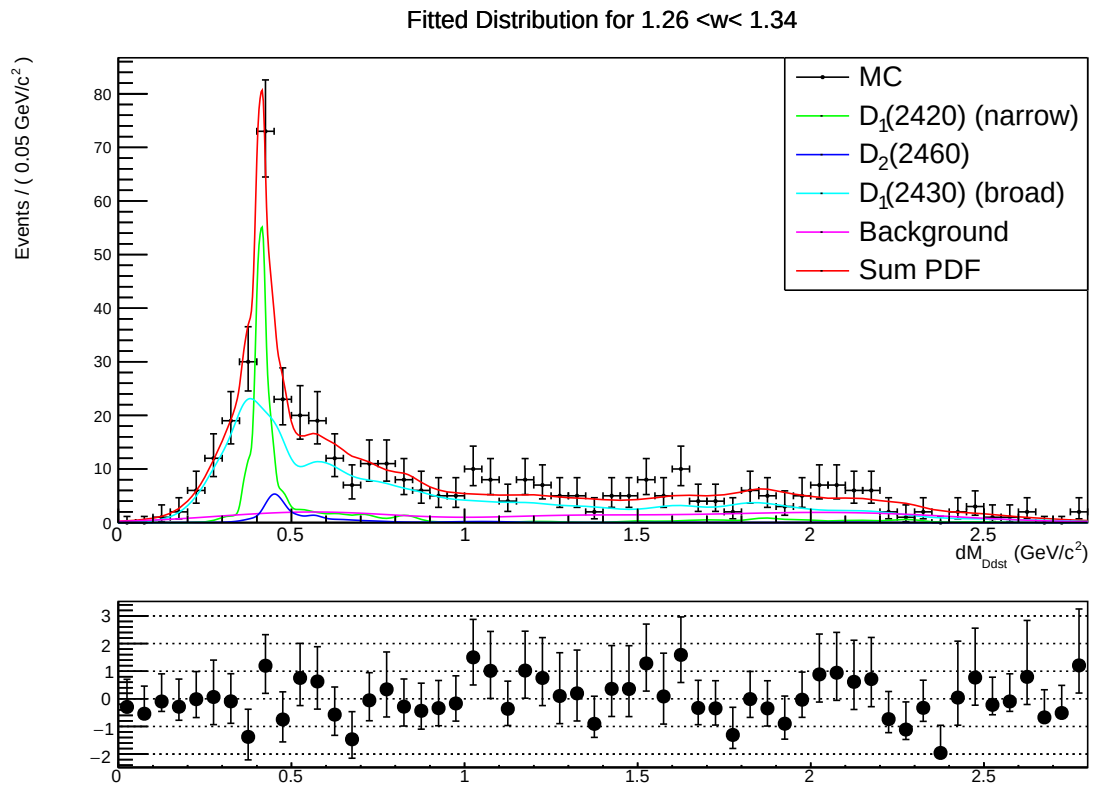


Table A.4: Fitted PDFs and Pull plots for the ΔM fits in the third bin in w .

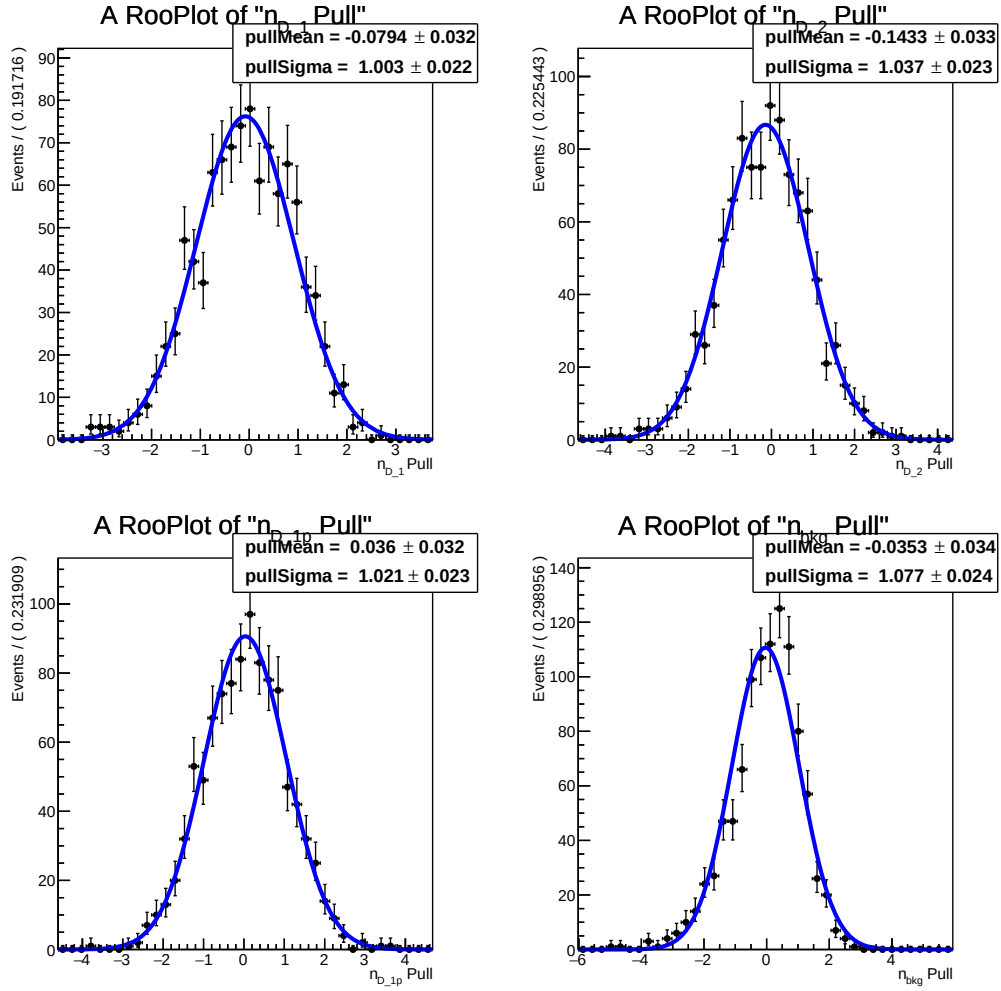


Figure A.4: The pull distributions of the four fit parameters from fits to 1000 toy MC samples for the first w bin, i.e. $0.95 < w < 1.19$.

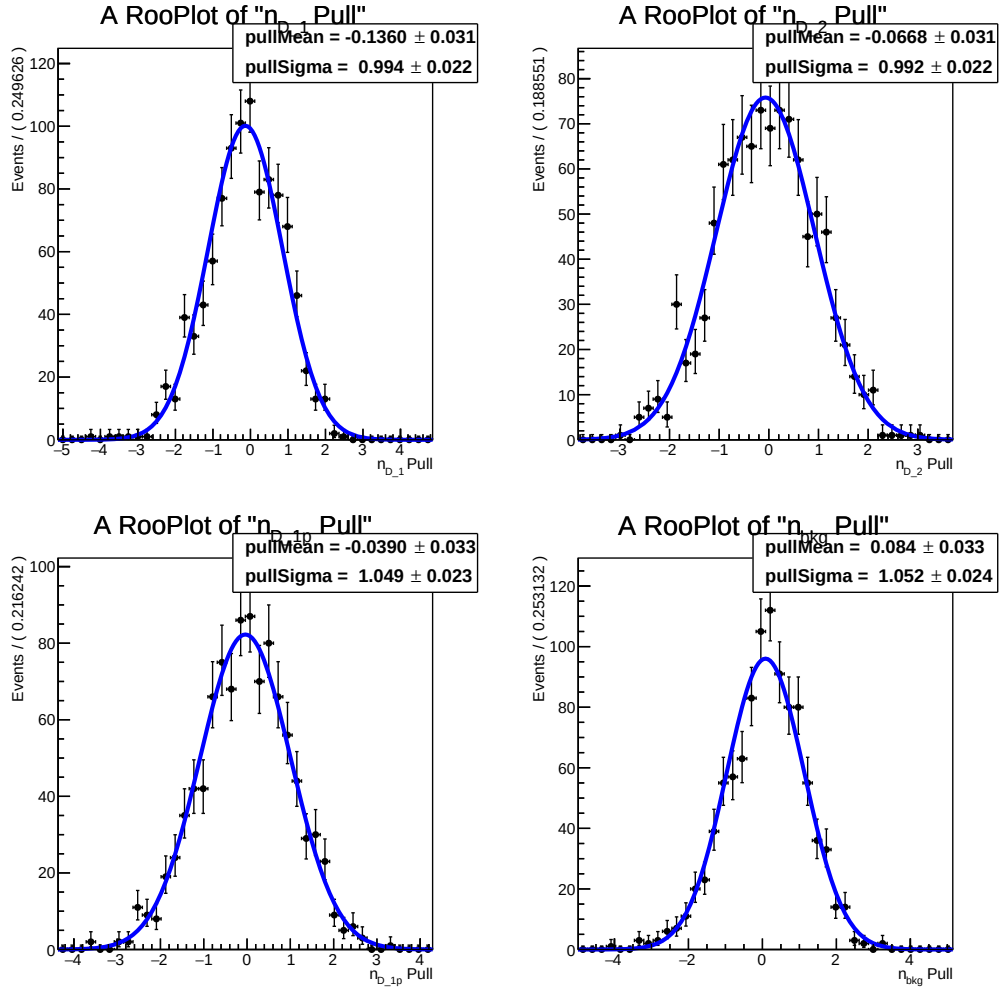


Figure A.5: The pull distributions of the four fit parameters from fits to 1000 toy MC samples for the second w bin, i.e. $1.19 < w < 1.26$.

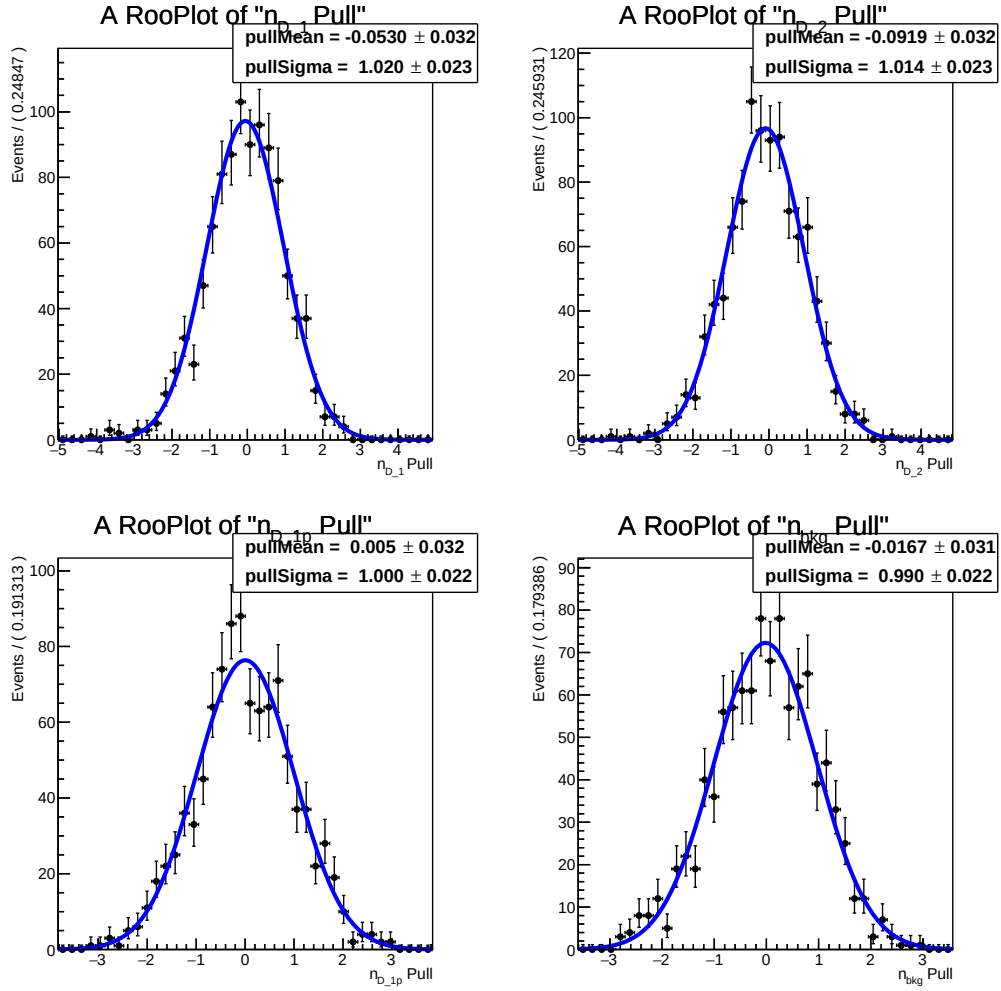


Figure A.6: The pull distributions of the four fit parameters from fits to 1000 toy MC samples for the third w bin, i.e. $1.26 < w < 1.34$.

Bibliography

- [1] R. Aaij et al. Measurement of the ratio of branching fractions $B(\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau)/B(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)$. *Phys. Rev. Lett.*, 115(11):111803, 2015. [Erratum: *Phys.Rev.Lett.* 115, 159901 (2015)].
- [2] V. M. Abazov et al. Measurement of semileptonic branching fractions of B mesons to narrow D^{**} states. *Phys. Rev. Lett.*, 95:171803, 2005.
- [3] G. Abbiendi et al. A Measurement of semileptonic B decays to narrow orbitally excited charm mesons. *Eur. Phys. J. C*, 30:467–475, 2003.
- [4] J. Abdallah et al. Determination of heavy quark non-perturbative parameters from spectral moments in semileptonic B decays. *Eur. Phys. J. C*, 45:35–59, 2006.
- [5] F. Abudinén et al. A calibration of the Belle II hadronic tag-side reconstruction algorithm with $B \rightarrow X \ell \nu$ decays. 8 2020.
- [6] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [7] H. r. a. Albrecht. Investigation of the decays $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}$ and $B \rightarrow D^{**} \ell^- \bar{\nu}$. *Z. Phys. C. Particles and Fields*, 57(4):533–540, Dec. 1993.
- [8] W. Altmannshofer et al. The Belle II Physics Book. *PTEP*, 2019(12):123C01, 2019. [Erratum: *PTEP* 2020, 029201 (2020)].
- [9] Y. Amhis et al. Averages of b -hadron, c -hadron, and τ -lepton properties as of 2021. *Phys. Rev. D*, 107:052008, 2023.
- [10] A. Anastassov et al. Investigation of semileptonic B meson decay to P wave charm mesons. *Phys. Rev. Lett.*, 80:4127–4131, 1998.
- [11] B. Aubert et al. Measurement of Semileptonic B Decays into Orbitally-Excited Charmed Mesons. *Phys. Rev. Lett.*, 103:051803, 2009.
- [12] S. Baker and R. D. Cousins. Clarification of the use of chi-square and likelihood functions in fits to histograms. *Nuclear Instruments and Methods in Physics Research*, 221(2):437–442, 1984.

- [13] Belle II Collaboration. Belle 2 software documentation. Available at <https://software.belle2.org/light-2212-foldex/sphinx/index.html>, release:light-2212-foldex, Accessed: 2023-09-13.
- [14] Belle II Collaboration. Belle2 Archives. Available at <https://www.belle2.org/archives/>, Accessed: 2023-09-13.
- [15] Belle II Collaboration. DESY Confluence Belle II Luminosity. Available at <https://confluence.desy.de/display/BI/Belle+II+Luminosity>, Accessed: 2023-09-13.
- [16] Belle II Collaboration. DESY Confluence-Conference Readiness. Available at <https://confluence.desy.de/display/BI/Conference+readiness#Conferencereadiness-Moriond2023,18-25March2023>(<https://moriond.in2p3.fr/2023/>), Accessed: 2023-09-10.
- [17] Belle II Collaboration. DESY Confluence-Correction of $\text{BF}(B \rightarrow X_c \ell \nu)$ in generic MC. Available at <https://confluence.desy.de/pages/viewpage.action?pageId=202394372>, Accessed: 2023-09-13.
- [18] F. U. Bernlochner, Z. Ligeti, and D. J. Robinson. Model-independent analysis of semileptonic b decays to D^{**} for arbitrary new physics. *Phys. Rev. D*, 97:075011, Apr 2018.
- [19] V. Bertacchi et al. Track finding at Belle II. *Comput. Phys. Commun.*, 259:107610, 2021.
- [20] I. I. Bigi, B. Blossier, A. Le Yaouanc, L. Oliver, O. Pene, J. C. Raynal, A. Oyanguren, and P. Roudeau. Memorino on the ‘1/2 versus 3/2 puzzle’ in $\overline{B} \rightarrow l \overline{\nu} X_c$ - a year later and a bit wiser. *Eur. Phys. J. C*, 52:975–985, 2007.
- [21] T. Bilka et al. Implementation of GENFIT2 as an experiment independent track-fitting framework. 2 2019. arXiv:1902.04405.
- [22] A. W. BOWMAN. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360, 08 1984.
- [23] S. Collaboration. Superb: A high-luminosity asymmetric $e^+ e^-$ super flavor factory. conceptual design report, 2007. arXiv:0709.0451.
- [24] G. Cowan. *Statistical data analysis*. Oxford Science Publications, 1998. ISBN:978-0-19-850156-5.
- [25] K. S. Cranmer. Kernel estimation in high-energy physics. *Comput. Phys. Commun.*, 136:198–207, 2001.
- [26] G. D’Agostini. A Multidimensional unfolding method based on Bayes’ theorem. *Nucl. Instrum. Meth. A*, 362:487–498, 1995.
- [27] N. Davidson, T. Przedzinski, and Z. Was. Photos interface in c++; technical and physics documentation, 2015.

- [28] O. et al. Accelerator design at SuperKEKB. *Progress of Theoretical and Experimental Physics*, 2013(3):03A011, 03 2013.
- [29] T. A. et al. Belle II Technical Design Report, 2010. arXiv:1011.0352.
- [30] A. Glazov, P. Rados, A. Rostomyan, E. Paoloni, and L. Zani. Measurement of the tracking efficiency in phase 3 data using tau-pair events. Feb 2020.
- [31] A. Hocker and V. Kartvelishvili. SVD approach to data unfolding. *Nucl. Instrum. Meth. A*, 372:469–481, 1996.
- [32] 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*, 552(3):566–575, 2005.
- [33] M. Huschle et al. Measurement of the branching ratio of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle. *Phys. Rev. D*, 92(7):072014, 2015.
- [34] N. Isgur and M. B. Wise. Weak transition form factors between heavy mesons. *Physics Letters B*, 237(3):527–530, 1990.
- [35] S. Jadach, B. Ward, and Z. Was. The precision monte carlo event generator for two-fermion final states in collisions. *Computer Physics Communications*, 130(3):260–325, aug 2000.
- [36] F. James and M. Winkler. MINUIT User’s Guide. 6 2004.
- [37] T. Keck et al. The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment. *Comput. Softw. Big Sci.*, 3(1):6, 2019.
- [38] R. D. Kenway. The Isgur-Wise function. *Nucl. Phys. B Proc. Suppl.*, 34:153–158, 1994.
- [39] P. Krishnan, M. Kumar, M. Nayak, and K. Trabelsi. k_S^0 reconstruction using phase 2 data. Jun 2018. BELLE2-NOTE-PH-2018-017.
- [40] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152–155, 2001.
- [41] J. P. Lees et al. Evidence for an excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ decays. *Phys. Rev. Lett.*, 109:101802, 2012.
- [42] M. Lefebvre, R. K. Keeler, R. Sobie, and J. White. Propagation of errors for matrix inversion. *Nucl. Instrum. Meth. A*, 451:520–528, 2000.
- [43] A. K. Leibovich, Z. Ligeti, I. W. Stewart, and M. B. Wise. Semileptonic B decays to excited charmed mesons. *Phys. Rev. D*, 57:308–330, 1998.
- [44] D. Liventsev et al. Study of $B \rightarrow D^{**}\ell\nu$ with full reconstruction tagging. *Phys. Rev. D*, 77:091503, 2008.

- [45] B. Malaescu. An Iterative, dynamically stabilized method of data unfolding. 7 2009. arXiv:0907.3791.
- [46] M. Neubert. B decays and the heavy quark expansion. *Adv. Ser. Direct. High Energy Phys.*, 15:239–293, 1998.
- [47] Y. Sato, S. Cunliffe, F. Meier, and A. Zupanc. Monte Carlo matching in the Belle II software. *EPJ Web Conf.*, 251:03021, 2021.
- [48] G. A. F. Seber. *Multinomial Distribution*, pages 882–884. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [49] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, jun 2015.
- [50] A. Thampi. Charmonium and bottomonium spectroscopy at belle ii, 2021. arXiv:2109.00337.
- [51] The Belle II Collaboration. Belle II Analysis Software Framework (basf2), Aug. 2022. DOI:10.5281/zenodo.6949466.
- [52] N. Uraltsev. Strong interaction effects in semileptonic b decays. *Nuclear Physics B - Proceedings Supplements*, 117:554–557, 2003.
- [53] W. Verkerke and D. P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003.
- [54] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

Acknowledgements

I would like to thank my supervisors Thomas Kuhr and Thomas Lück. Their understanding and encouragement have made the writing of this thesis a genuinely enjoyable experience. A special thanks to Thomas Lück for his close guidance, and patience for all my questions.

I am grateful to have studied in such a welcoming and supportive environment, and for that, I would like to thank everyone in the group.

Finally, I would like to thank my family and friends for their unwavering support.

Erklärung/Declaration

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

I hereby declare that this thesis is my own work, and that I have not used any sources and aids other than those stated in the thesis.

München, 15.09.1999

Eylül Ünlü