ETP-KA/2022-01

Karlsruhe Institute of Technology

Preparation of a Measurement of $\mathcal{R}(D^{(*)})$ with Leptonic τ and Hadronic FEI Tag at the Belle Experiment

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M.Sc. Felix Gregor Kuno Metzner

aus Offenbach am Main

Tag der mündlichen Prüfung: 11. Februar 2022
Referent: Prof. Dr. Michael Feindt
Korreferent: Prof. Dr. Florian Bernlochner

Contents

1. Introduction

The B factories [1] KEKB in Tsukuba, Japan [2] and PEP-II in the U.S.A. [3] have produced large data sets of B decays, recorded by the respective particle detectors Belle [4] and *BABAR* [5]. These data sets have been studied for over a decade to reveal physics processes such as CP violation in the B sector [6] and $b \to s\ell\ell$ transitions via penguin diagrams [7, 8], as well as more recent results, such as the evidence for an excess of $B \to D^{(*)}\tau\nu$ decays [9].
Regardless of the aim or scientific impact of an analysis conducted on the basis of these B factory data sets, most analyses rely on the simulation of the complex processes occurring for each bunch crossing at aforementioned electron-positron colliders. From the decay of an $\Upsilon(4S)$ meson into a $B\overline{B}$ meson pair and the subsequent decay of the two B mesons up to the interaction of the final state particles in the detector material, every step must be well described in the Monte Carlo (MC) simulation. Each stage of this process requires external input, such as the detector geometry and the physical properties of the various materials used to build it, in case of the simulation of the detector response. Equally important, if not more so, is the theoretical input for the description of the creation of the B meson pair and especially the multitude of their possible decay processes. This input must describe the very nature of the decay processes and the underlying interactions governed by the Standard Model (SM) of particle physics.
Since the end of data taking at the B factories of the first generation, multiple analyses of the B decay data sets have not stopped to provide evermore detailed insights into the inner workings of the SM and to improve our understanding of the B meson decay processes. This gain of knowledge, however, is not reflected in the MC data samples, which still form the basis of many new analyses of the data sets recorded at the B factories.

In the scope of the presented analysis, I incorporate the latest knowledge on the branching fractions and modeling of the semi-leptonic B meson decays into the Belle MC samples and utilize these updated samples to prepare a measurement of the branching fraction ratios

$$\mathcal{R}(D) = \frac{\mathcal{B}(B \to D\tau\nu_\tau)}{\mathcal{B}(B \to D\ell\nu_\ell)} \qquad \text{and} \qquad \mathcal{R}(D^*) = \frac{\mathcal{B}\left(B \to D^*\tau\nu_\tau\right)}{\mathcal{B}\left(B \to D^*\ell\nu_\ell\right)}, \tag{1.1}$$

with the data set recorded by the Belle experiment. This data set is still the largest B factory data set, and thus in terms of statistical power provides the best access to the semi-leptonic B decays defining the ratios $\mathcal{R}(D^{(*)})$.
These processes are the semi-leptonic B decays into either a ground stats D meson, or the excited D^* mesons, as well as the **light leptons** $\ell = e, \mu$ or the heavier $\boldsymbol{\tau}$ **lepton**, which define the denominator and the numerator of the ratios, respectively.
The ratios $\mathcal{R}(D^{(*)})$ probe the lepton flavor universality (LFU), a prediction of the SM, embedded into the theory by the universality of the electroweak gauge coupling to the three fermion flavors. As a result of LFU, the physical observables are equal for all lepton flavors except for deviations due to the different lepton masses. This leads to a very similar behavior for the light leptons e and μ, whereas the τ lepton behaves differently in many aspects, because of its greater mass. These effects, however, are well described in the SM.

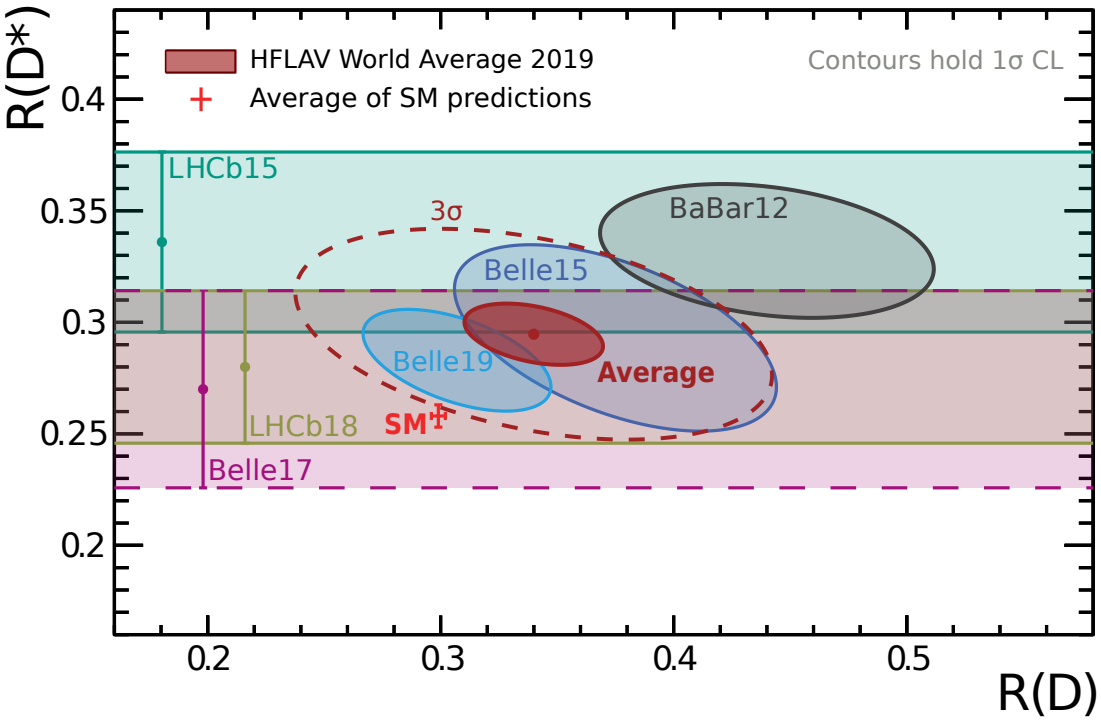


Figure 1.1.: Overview of the measurements of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ conducted at Belle, *BABAR* and LHCb. Detailed information on each measurement and the respective source is listed in Table 1.1. The current SM prediction of the ratios $\mathcal{R}(D^{(*)})$ and the combination of the experimental values (labeled as *Average*), as determined by the Heavy Flavor Averaging Group (HFLAV) [10], are included as well. Adapted from [11].

Hence, precise predictions can be made and experimentally challenged, e.g. via the decay rate ratios $\mathcal{R}(D^{(*)})$.

All thus far conducted analyses which measured both ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ simultaneously show a deviation from this prediction of the SM. An overview of these measurements is provided in Figure 1.1, as well as in Table 1.1. While the latter tabulates the results of the different measurements of the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ together with the respective sources, the Figure 1.1, visualizes the deviation of each result from the SM. Both also include the SM prediction of the two branching fraction ratios and the combination of the experimental measurements determined by the Heavy Flavor Averaging Group (HFLAV) [10]. The comparison of the SM prediction and the experimental average indicates a tension with a significance of $\sim 3.4\sigma$ [10] for this test of LFU. All the combined measurements are statistically limited and can be improved upon with more data or efficient analysis techniques.

Analysis Strategy

The analysis approach using hadronically tagged events followed by the presented analysis, and also employed for the measurements produced by *BABAR* [9] in 2013 and Belle [12] in 2015, exploits the similarity in the event signature of the **signal processes** $B \to D^{(*)}\tau\nu$ and the **normalization processes** $B \to D^{(*)}\ell\nu$. If the τ lepton of the signal process is reconstructed in leptonic decay channels, the signature features the same visible final state particles as the normalization process. However, due to the leptonic τ decay, the signal

Table 1.1.: Measurements of the branching fraction ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ conducted by the Belle, *BABAR* and LHCb collaborations. The table provides an overview of the obtained values and information about the reconstruction methods used for each measurement. The column labeled with τ shows whether the reconstruction of the τ lepton is performed in leptonic or hadronic decay modes. Information about the B meson tagging method used, is given in the column labeled as B_{tag}. The source for the measurement is provided in the last column. Additionally, the average value obtained from the combination of the experimental values and the SM prediction are given in the last two rows.

	τ	B_{tag}	$\mathcal{R}(D^*)$	$\mathcal{R}(D)$	Year	
Belle	leptonic	hadronic	$0.293 \pm 0.038 \pm 0.015$	$0.375 \pm 0.064 \pm 0.026$	2015	[12]
	leptonic	semi-lep.	$0.283 \pm 0.018 \pm 0.014$	$0.307 \pm 0.037 \pm 0.016$	2020	[13]
	hadronic	hadronic	$0.270 \pm 0.035\ ^{+0.028}_{-0.025}$	—	2018	[14]
BaBar	leptonic	hadronic	$0.332 \pm 0.024 \pm 0.018$	$0.440 \pm 0.058 \pm 0.042$	2013	[9]
LHCb	leptonic	—	$0.336 \pm 0.027 \pm 0.030$	—	2015	[15]
	hadronic	—	$0.280 \pm 0.018 \pm 0.029$	—	2018	[16]
HFLAV	Arithmetic Average		$0.295 \pm 0.011 \pm 0.008$	$0.340 \pm 0.027 \pm 0.013$	2019	[11]
SM			0.258 ± 0.005	0.299 ± 0.003	2019	[11]

process results in three neutrinos, which cannot be detected, whereas the normalization process only features one invisible neutrino. This difference in the signal and normalization processes is used to differentiate between the two. In the unique environment provided by the B factories, the invariant mass of the invisible neutrino(s) can be calculated as the missing mass squared of the event:

$$M^2_{miss} = \left(p_{Beam} - p_{B_{tag}} - p_{D^{(*)}} - p_\ell\right)^2 . \tag{1.2}$$

This observable exploits the fact that the e^+e^-–collision at the B factories produce $B\overline{B}$ mesons pairs in a well-known initial state and without further particles. This allows for the reconstruction of the entire collision event as illustrated in Figure 1.2. The B meson decaying via either the signal or normalization process, is henceforth denoted as B_{sig} meson. If the second B meson of the event — the so-called B_{tag} meson — is reconstructed in hadronic decay modes, the neutrino(s) of the B_{sig} decay are the only invisible particles in the event. The above defined observable M^2_{miss} is calculated as the difference of the known center-of-mass (CMS) energy given by the momentum of the colliding particle beams p_{Beam} and all reconstructed particles of the event.
For the normalization process $B \to D^{(*)}\ell\nu$, only a single neutrino remains undetected. The invariant mass of this neutrino is zero, and so is the value of the observable M^2_{miss}. As three neutrinos remain undetected for the signal processes $B \to D^{(*)}\tau\nu$, a non-zero invariant mass is measured for the neutrino system, resulting in a value of $M^2_{miss} > 0\,GeV^2$ for the observable. Furthermore, a light lepton produced in the normalization processes $B \to D^{(*)}\ell\nu$ carries a higher momentum than a light lepton originating from a signal process τ decay, due to the energy loss to the additional neutrinos in the latter case. These two difference are utilized to differentiate between the signal and normalization processes in a fit in the missing mass squared M^2_{miss} and the lepton momentum in the B_{sig} rest-frame p^*_ℓ.

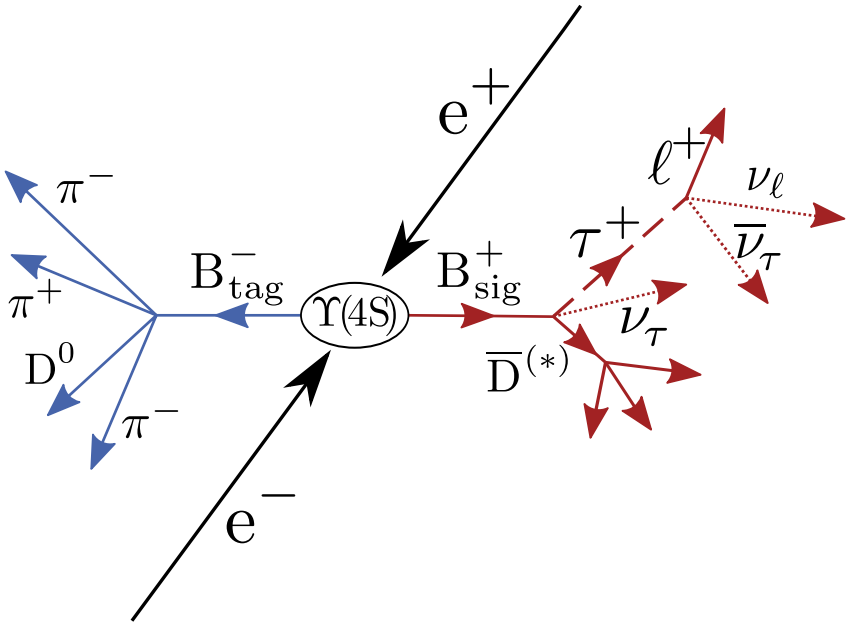


Figure 1.2.: Illustration of the **signal process $\mathbf{B \to D^{(*)}\tau\nu}$** with leptonic τ decay occurring for one of the B mesons of a $\Upsilon(4S)$ decay. This B meson is denoted as $\mathbf{B_{sig}}$. The second B meson produced in the e^+e^-–collision is reconstructed in hadronic decay channels by the B tagging algorithm, and thus called $\mathbf{B_{tag}}$ meson. Without loss of generality, the process is displayed for the case of a pair of charged B mesons.

Crucial for this analysis strategy is the efficiency of the reconstruction of the second B_{tag} meson of an event in hadronic decay modes. For this task, I employ the new Full Event Interpretation (FEI) [17] B tagging algorithm, originally developed for the Belle II experiment, which promises a significant improvement of the B_{tag} meson reconstruction efficiency by a factor of ~ 2. This doubles the number of B mesons available for analysis with respect to the previous Belle analysis [12]. The application of this Belle II algorithm for Belle data is made possible with the help of the `b2bii` data conversion tool [18]. I successfully applied this tagging algorithm already on Belle data for the analysis of $B^+ \to \ell^+\nu_\ell\gamma$ decays [19, 20], which confirmed this improvement.

By employing modern analysis methods such as the FEI tagging algorithm, and by modeling semi-leptonic B meson decays of the normalization and main background processes to reflect our latest knowledge, I aim for a competitive measurement of the decay rate ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with the Belle data sample, which will constitute the final word from Belle on this channel and will replace the current hadronically tagged Belle result from 2015 [21, 12]. This thesis presents the steps of this complex endeavor, starting with the theory foundations necessary for the update of the MC samples, up to the evaluation of the fit performance and the study of data-MC agreement. Due to the high profile of the measurement and the complexity of the analysis and the incorporated improvements, this analysis is at the time of writing in the internal review process of the Belle collaboration and is kept blinded to avoid experimental bias.

Throughout this thesis natural units $c = \hbar = k_{\mathrm{B}} = \epsilon_0 = 1$ are used. If decay processes are specified, the inclusion of charge-conjugate decay modes is implied, if not explicitly mentioned otherwise. When multiple decay modes are jointly referred to in one decay string, the charge and/or labeling of anti-particles may be omitted.

2. Theory

This chapter provides an entry point into the theoretical background of the decay processes studied in this work. After an overview of the decay processes of interest is given, the basic principles of heavy-quark physics [22] describing the properties of the mesons in the initial and final state are briefly explained. Next, the SM Lagrangian describing the decay process is introduced in Section 2.2. The heavy-quark mesons require a special treatment, as a perturbative treatment of the quantum chromodynamics (QCD) interactions is not applicable at the energy scale of the beauty or charm mesons. The common approach [22, 23] of factorization of the matrix elements to this issue is illustrated. This leads to a general expression for the differential decay rate [22, 23, 24] of the process in Section 2.3 in terms of hadronic amplitudes. In Section 2.4 a specific parametrization [25, 26] of the hadronic amplitudes is given, which is used to update the Monte Carlo samples used for this analysis. The chapter ends with a quick overview of the main background to the analysis — the semi-leptonic B decays into D^{**} mesons. This analysis also implements an update of the Monte Carlo description of these background processes. Additional sources for the parametrization of the decay rates particular to these background processes which lay the foundations for the Monte Carlo sample update are provided in Section 2.5. The reweighting procedure of the MC samples to the updated description is elaborated on in Chapter 4.
The decays of interest are the semi-leptonic $\overline{\mathrm{B}} \to \mathrm{D}^{(*)} l \nu_l$ processes illustrated in Figure 2.1, where

$$l = \mathrm{e}, \mu, \tau \tag{2.1}$$

refers to all three lepton flavors. The SM provides a well-understood prediction for the

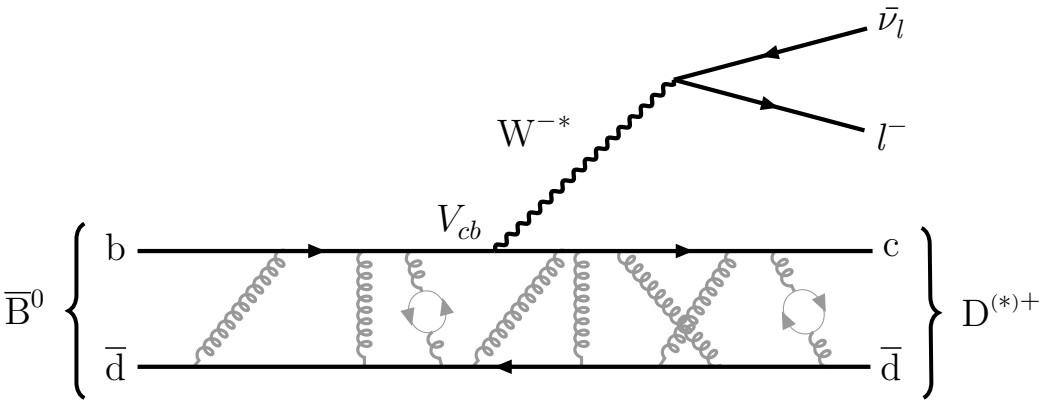


Figure 2.1.: Diagram of the semi-leptonic B meson decay process $\overline{\mathrm{B}}^0 \to \mathrm{D}^{(*)+} l^- \bar{\nu}_l$ in which the charged weak current mediated by a virtual W boson allows for the $\mathrm{b} \to \mathrm{c}$ transition. The virtuality of the off-shell W boson is indicated by a $*$ in the diagram. Spin and parity of the light spectator quark q combined with the gluons and $q\bar{q}$ pairs (light degrees of freedom) characterize the state of the resulting charm meson X_c, which can be a D meson or higher excitations of such. The virtual W boson decays into a $l\nu_l$ pair, rendering the process a semi-leptonic B decay.

leptonic part of this process which arises from the decay of the virtual W boson. The theory of the electroweak interaction predicts a coupling of the charged weak current to the created lepton-neutrino pair $l\nu_l$ that is independent of the three lepton flavors. This universality of the electroweak gauge coupling to the leptons is referred to as lepton flavor universality (LFU). It implies that physical observables of processes involving leptons are independent of the lepton flavor in the SM, except for the effects of the difference in charged lepton masses on the phase space and the leptonic currents. Hence, the following SM calculations are valid for all lepton flavors l. However, the presence of the neutrino, as well as the subsequent decay of the charged lepton itself inside the detector volume, in case of the τ lepton, adds a level of complexity from an experimental point of view. As such, these issues will be addressed in the following chapters of this text.

The interaction of the involved quarks and $\mathrm{b} \to \mathrm{c}$ transition are less well described by the SM, due to the non-perturbative nature of QCD at low energy scales. Hadronization processes can lead to further light hadrons in the final state; however, such processes are not considered here. Furthermore, only leptonic τ decay processes $\tau \to \ell\bar{\nu}_\ell\nu_\tau$ into a light lepton and two additional neutrinos are considered. Thus, the signature of the studied decay processes feature a D meson, or its higher excitations, and a charged light lepton, as well as its corresponding neutrino(s). Before the decay process of interest is described in more detail, a short introduction of the heavy-quark mesons that appear in the initial and final state shall be given, to provide an overview of their properties.

2.1. Heavy-Quark Mesons

The properties of the bound states formed by the heavy quarks Q and light anti-quarks $\bar{q}$ in the initial ($Q = \mathrm{b}$) and final state ($Q = \mathrm{c}$) can be described by heavy-quark symmetry [27, 22]. In the heavy-quark limit $\mathrm{m}_Q \to \infty$ the dynamics of the $Q\bar{q}$ meson is independent of the flavor of the heavy quark Q. Furthermore, the spin s_Q of the heavy quark can be seen as a good quantum number. As the total angular momentum J of the $Q\bar{q}$ meson must be conserved in the bound state, so must be the total spin of the combination of the light quark $\bar{q}$ and the arbitrary number of gluons and $q\bar{q}$ pairs that make up the remainder of the meson. This combination of the light quark and the remainder are collectively called *light degrees of freedom* ℓ, and their total spin quantum number is denominated by s_ℓ. Relevant to the studied decay processes are the heavy-quark mesons listed in Table 2.1 defined by the spin doublet state with

$$J = s_Q \pm s_\ell = \frac{1}{2} \pm s_\ell \tag{2.2}$$

for $s_\ell = 1/2$ and the parity P of the $Q\bar{q}$ pair, which is odd in the ground state as the heavy quark and light anti-quark have opposite intrinsic parity.

In the $\mathrm{b} \to \mathrm{c}l\bar{\nu}_l$ transition, these quantum numbers may change, allowing for different $\mathrm{c}\bar{q}$ meson states in the final state. In case of an additional orbital angular momentum L between the light and heavy quark of the charmed mesons in the final state, the parity is given by $P = (-1)^{L+1}$ and the orbital angular momentum is combined with the spin of the light degrees of freedom depending on the relative orientation. They are collectively referred to as D^{**} mesons. Final states with either of these orbitally excited D^{**} meson states pose an important background to the $\overline{\mathrm{B}} \to \mathrm{D}^{(*)}\tau^-\overline{\nu}_\tau$ processes studied in this analysis. Their properties are tabulated in Table 2.2.

Table 2.1.: Properties of the charged and neutral ground state $Q\bar{q}$ meson doublets for the two possible choices of the heavy quark Q being a charm quark c or a bottom quark b. The former form the D and D^* mesons, while for the latter case, the non-excited B mesons are of relevance to this analysis. The grayed out excited B^* states are listed for completeness. Combining the spin 1/2 of the heavy quark Q and the light degrees of freedom s_ℓ results in doublets with a total angular momentum and parity of $J^P = (0^-, 1^-)$ and spin S. Mass values m are taken from [28].

Quark				$Q = \mathrm{c}$		$Q = \mathrm{b}$	
Content	s_ℓ	J^P	S	Meson	m/MeV	Meson	m/MeV
$Q\overline{\mathrm{d}}$	1/2	0^-	0	D^+	1869.66 ± 0.05	$\overline{\mathrm{B}}^0$	5279.65 ± 0.12
$Q\overline{\mathrm{u}}$	1/2	0^-	0	D^0	1864.84 ± 0.05	B^-	5279.34 ± 0.12
$Q\overline{\mathrm{d}}$	1/2	1^-	1	D^{*+}	2010.26 ± 0.05	$\overline{\mathrm{B}}^{*0}$	5324.70 ± 0.21
$Q\overline{\mathrm{u}}$	1/2	1^-	1	D^{*0}	2006.85 ± 0.05	B^{*-}	5324.70 ± 0.21

2.2. Semi-leptonic B Decays

Independent of the hadronic final state, the $\mathrm{b} \to \mathrm{c} l \bar{\nu}_l$ transition shown in Figure 2.1 is described in the SM by the effective weak Lagrangian

$$\mathcal{L}_\mathrm{W} = -2\sqrt{2}\, G_F\, \mathrm{V_{cb}}\, (\bar{\mathrm{c}}\, \gamma_\mu P_L\, \mathrm{b})(\bar{l}\, \gamma^\mu P_L\, \nu_l) \;+\; \mathrm{h.c.} \tag{2.3}$$

with the left- and right-handed projection operators $P_{L,R} = (1 \mp \gamma^5)/2$ and the Fermi constant $G_F = \sqrt{2}\, g_2^2/(8\, \mathrm{m_W^2}) = (\sqrt{2}\, v^2)^{-1}$, where g_2 denotes the $SU(2)$ weak coupling constant; $v \approx 246.22\,\mathrm{GeV}$ the SM Higgs field vacuum expectation value; and $\mathrm{m_W^2}$ the mass of the W boson. The probability of the change of flavor via the charged weak current is included with the CKM matrix element $\mathrm{V_{cb}}$ for the involved quarks. The fields of the heavy quarks in the initial and final state are represented by $\bar{\mathrm{c}} = \mathrm{c}^\dagger \gamma^0$ and b, respectively. The fields of the lepton and its corresponding neutrino are labelled analogously with l and $\bar{\nu}_l = {\nu_l}^\dagger \gamma^0$.

At leading electroweak order, the resulting matrix element factorizes into the product of the matrix element representing the leptonic current and hadronic currents as [29]

$$\mathcal{M}^{\lambda_l}_{\lambda_{X_\mathrm{c}}}(\mathrm{q}^2, \theta_l) = \frac{G_F}{\sqrt{2}} \mathrm{V_{cb}} \frac{\mathrm{m_W^2}}{\mathrm{m_W^2} - \mathrm{q}^2} \sum_{\lambda_\mathrm{W}} \eta_{\lambda_\mathrm{W}} L^{\lambda_l}_{\lambda_\mathrm{W}}(\mathrm{q}^2, \theta_l) H^{\lambda_{X_\mathrm{c}}}_{\lambda_\mathrm{W}}(\mathrm{q}^2), \tag{2.4}$$

with

$$L^{\lambda_l}_{\lambda_\mathrm{W}}(\mathrm{q}^2, \theta_l) = \epsilon_\mu(\mathrm{q}, \lambda_\mathrm{W})\, \langle l(p_l, \lambda_l) \bar{\nu}_l(p_{\nu_l}) \,|\, \bar{l} \gamma^\mu (1 - \gamma_5) \nu_l \,|\, 0 \rangle \quad \text{and} \tag{2.5}$$

$$H^{\lambda_{X_\mathrm{c}}}_{\lambda_\mathrm{W}}(\mathrm{q}^2) = \epsilon^*_\mu(\mathrm{q}, \lambda_\mathrm{W})\, \langle X_\mathrm{c}(p_{X_\mathrm{c}}, \lambda_{X_\mathrm{c}}) \,|\, \bar{\mathrm{c}} \gamma^\mu (1 - \gamma_5) \mathrm{b} \,|\, \overline{\mathrm{B}}(p_\mathrm{B}) \rangle \tag{2.6}$$

being the leptonic and the hadronic currents, respectively.

The latter depend on the specific hadrons in the initial (B) and final (X_c) state for which the mesons tabulated in Table 2.1 have to be considered for the ratios $\mathcal{R}(\mathrm{D}^{(*)})$. Inserting the D^{**} mesons listed in Table 2.2, which feature a unit of orbital angular momentum, for X_c in the final state leads to the expressions for the main background contributions of this analysis. These background contributions are briefly discussed in Section 2.5.

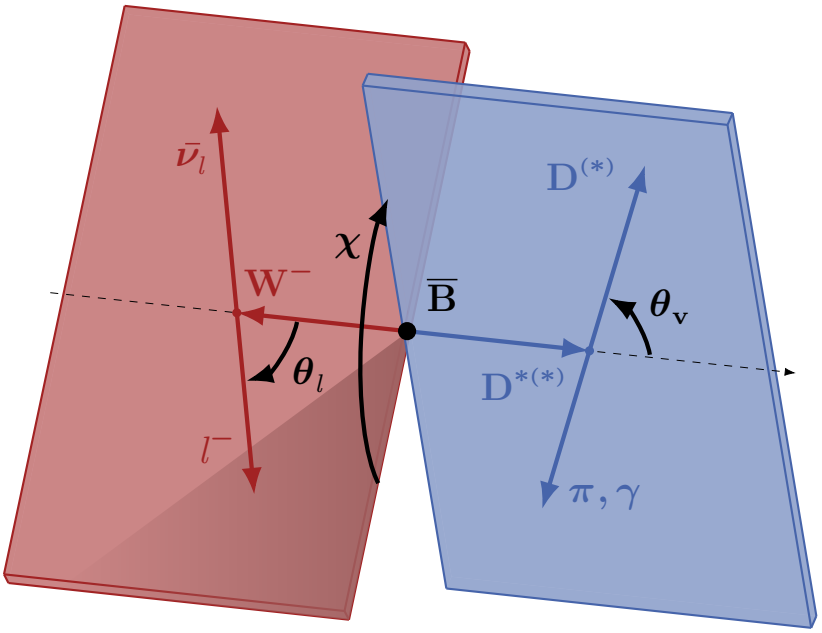


Figure 2.2.: Illustration of the helicity angles of the $B \to D^{*(*)} l\bar{\nu}_l$ decay process. The transition $B \to D^{*(*)}W$ is shown in the rest frame of the B meson, while the creation of the $l\bar{\nu}_l$ pair is shown in the decay plane of the virtual W boson in which the W boson is at rest. Analogously, the two-body decay of the X_c meson ($= D^*, D^{**}$) is shown in the X_c decay plane in which it is at rest. The helicity angle χ is the angle between these two decay planes. The angles θ_l and θ_v are defined relative to the momentum of the X_c meson in the respective rest frames.

All involved particle fields depend on the respective particle four-momentum, where the momenta of the mesons and the virtual W boson can be defined in the rest frame of the B meson as

$$p_B = (m_B, 0, 0, 0), \quad p_{X_c} = (E_{X_c}, 0, 0, |\vec{p}^*_{X_c}|) \quad \text{and} \quad p_W \equiv q = (q^0, 0, 0, -|\vec{p}^*_{X_c}|), \tag{2.7}$$

whereas p_l and p_{ν_l} — the four-momenta of the lepton and its neutrino — are expressed relative to the rest frame of the virtual W boson according to Figure 2.2.
For the expressions of the four-momenta in the B rest frame, one obtains

$$q^0 = \frac{1}{2m_B}(m_B^2 - m_{X_c}^2 + q^2) \quad \text{and} \quad |\vec{p}^*_{X_c}| = \sqrt{q^{0\,2} - q^2}, \tag{2.8}$$

where q^0 is the energy of the virtual W boson and q^2 is its mass, or, from another point of view, the squared four-momentum transfer from the mesons to the lepton system:

$$q^2 = (p_B - p_{X_c})^2 = (p_l + p_{\nu_l})^2. \tag{2.9}$$

Hence, q^2 covers a value range from $q^2{}_{\min} = m_l^2$ set by the mass of the involved lepton l up to $q^2{}_{\max} = (m_B - m_{X_c})^2$, which is maximal for the D meson with $q^2{}_{\max} \approx 11.63\,\mathrm{GeV}^2$, but still small when compared to the square of the on-shell W boson mass $m_W^2 = (80.38\,\mathrm{GeV})^2$.

The matrix element $\mathcal{M}$ depends on the helicities λ of the charged lepton l and the charmed meson X_c in the final state. This dependency is reflected in the factorization into leptonic and hadronic currents, where the respective fields appear as functions of the helicities. The virtual W boson can take on helicity values of $\lambda_\mathrm{W} = \pm, 0, s$, where s refers to the scalar state of the mediator, which also corresponds to a helicity of 0. The helicity λ_W also emerges as an argument of the metric factor

$$\eta_{\{\pm,0,s\}} = \left\{1,\, 1,\, \frac{\mathrm{q}^2 - \mathrm{m}_\mathrm{W}^2}{\mathrm{m}_\mathrm{W}^2}\right\} \overset{\mathrm{q}^2 \ll \mathrm{m}_\mathrm{W}}{\approx} \{1,\, 1,\, -1\}\,, \tag{2.10}$$

as well as in the leptonic $L^{\lambda_l}_{\lambda_\mathrm{W}}$ and hadronic $H^{\lambda_{X_\mathrm{c}}}_{\lambda_\mathrm{W}}$ currents in the form of the polarization four-vector of the virtual W boson $\epsilon_\mu(\mathrm{q}, \lambda_\mathrm{W})$. In the matrix element in Equation (2.4) this dependency on the helicity λ_W is summed over. The helicity of the lepton l can take on the values $\lambda_l = \pm 1$, while the values to be considered for the helicity λ_{X_c} of the X_c in the final state depend on the respective meson. E.g. the D meson as pseudoscalar can assume the helicity $\lambda_\mathrm{D} = s$, whereas the values $\lambda_{\mathrm{D}^*} = \pm 1, 0$ must be considered for D^* mesons with spin 1.

The definition of the polarization vectors depends on sign conventions which do affect the signs in the parametrization of the process to follow, but result in the same physical observables if applied consistently. Following the sign convention used in [30] the polarization vectors $\epsilon_\mu(\mathrm{q}, \lambda_\mathrm{W})$ of the virtual W boson — or current projections — can be defined such that

$$\epsilon^\mu(\mathrm{q}) = \begin{cases} \frac{1}{\sqrt{2}} & (0, \pm 1, -i, 0), & \text{if } \lambda_\mathrm{W} = \pm \\ \frac{1}{\sqrt{\mathrm{q}^2}} & (|\vec{p}^{\,*}_{X_\mathrm{c}}|, 0, 0, -\mathrm{q}^0), & \text{if } \lambda_\mathrm{W} = 0 \\ \frac{1}{\sqrt{\mathrm{q}^2}} & \mathrm{q}^\mu, & \text{if } \lambda_\mathrm{W} = s \end{cases} \tag{2.11}$$

in the B meson rest frame.

Leptonic Matrix Elements

A general representation of the matrix elements of the leptonic current $L^{\lambda_l}_{\lambda_\mathrm{W}}$ using elements of the Wigner D-matrices as $D^j_{m'm}(\chi, \theta_l)$, which also covers the dependency on the helicity angle χ between the two decay planes introduced in Figure 2.2, can be found in [30, Appendix B]. For the purpose of this introduction, the χ-dependency will be omitted and the representation of the leptonic current $L^{\lambda_l}_{\lambda_\mathrm{W}}$ introduced in [29] is used:

$$\begin{aligned} L^+_\pm(\mathrm{q}^2, \theta_l) &= \mp\sqrt{2}\mathrm{m}_l v d_0 \quad & L^+_0(\mathrm{q}^2, \theta_l) &= \sqrt{2}\mathrm{m}_l v(d_+ - d_-) \quad & L^+_s(\mathrm{q}^2, \theta_l) &= -2\mathrm{m}_l v \\ L^-_\pm(\mathrm{q}^2, \theta_l) &= -2\sqrt{\mathrm{q}^2} v d_\pm \quad & L^-_0(\mathrm{q}^2, \theta_l) &= -2\sqrt{\mathrm{q}^2} v d_0 & L^-_s(\mathrm{q}^2, \theta_l) &= 0 \end{aligned} \tag{2.12}$$

with the sign convention from [30] and

$$v = \sqrt{1 - \frac{\mathrm{m}_l^2}{\mathrm{q}^2}}, \quad d_\pm = \frac{1 \pm \cos\theta_l}{\sqrt{2}} \quad \text{and} \quad d_0 = \sin\theta_l, \tag{2.13}$$

which reduce the connection between the polarization vectors and the Wigner D-matrices to expressions depending on the invariant q^2 and θ_l, as well as the helicity states.

Hadronic Matrix Elements
The matrix elements of the hadronic current $H^{\lambda_{X_c}}_{\lambda_W}$ depend on the invariant q^2. Moreover, a dependency on the helicity angles θ_v and χ as defined in Figure 2.2 must be considered for excited D mesons (i.e. D^* and D^{**}). For the purpose of this introduction, however, this dependency is omitted and integrated over for the resulting differential decay rates. A detailed evaluation is presented in [30] or [24].
The hadronic current matrix elements can be split into their vector V^μ and axial-vector A^μ contributions, with

$$V^\mu = \bar{c}\,\gamma^\mu\,b \qquad \text{and} \qquad A^\mu = \bar{c}\,\gamma^\mu\gamma^5\,b. \tag{2.14}$$

They can be described in terms of two or four independent four-vectors and an equal number of independent form factors, depending on whether a ground state or an excited $D^{(*)}$ meson is assumed in the final state, respectively. Various different form factor bases are available. Here, the currents are shown in terms of the heavy-quark symmetry form factor basis $\{h_\pm, h_V, h_{A_{1,2,3}}\}$ [31, 22]. The form factors are, independent of the basis, dimensionless functions of either the Lorentz-invariant four-momentum transfer squared q^2, or the related recoil parameter w defined by

$$w = v \cdot v' = \frac{1}{m_B} p_B^\mu \frac{1}{m_{X_c}} p_{X_c\,\mu} = \frac{m_B^2 + m_{X_c}^2 - q^2}{2 m_B m_{X_c}}, \tag{2.15}$$

where v and v' denote the four-velocities of the B meson and the $D^{(*)}$ meson, respectively. For a more comprehensive overview of different means of parametrizing the hadronic currents, consider [23].
Considering a charged or neutral **D meson** in the final state, one obtains the expressions

$$\langle D(p')|V^\mu|\overline{B}(p)\rangle = \sqrt{m_B m_D}\ \left[h_+(v+v')^\mu + h_-(v-v')^\mu\right] \tag{2.16}$$

$$\langle D(p')|A^\mu|\overline{B}(p)\rangle = 0, \tag{2.17}$$

where the axial-vector current vanishes, due to angular momentum and parity conservation. Assuming the $J^P = 1^-$ state of the **D^* meson** in the final state leads to

$$\langle D^*(p',\epsilon)|V^\mu|\overline{B}(p)\rangle = i\sqrt{m_B m_{D^*}}\ h_V\,\varepsilon^{\mu\nu\alpha\beta}\,\epsilon^*_\nu v'_\alpha v_\beta \tag{2.18}$$

$$\langle D^*(p',\epsilon)|A^\mu|\overline{B}(p)\rangle = \sqrt{m_B m_{D^*}}\ \left[h_{A_1}(w+1)\,\epsilon^{*\mu} - (h_{A_2}\,v^\mu + h_{A_3}\,v'^\mu)\,(\epsilon^*\cdot v)\right] \tag{2.19}$$

with the Levi-Civita tensor $\varepsilon^{\mu\nu\alpha\beta}$ and the polarization four-vectors of the D^* meson ϵ^*_μ which are given by

$$\epsilon^{*\mu}(\lambda_{X_c}) = \begin{cases} \frac{1}{\sqrt{2}} & (0,\mp 1,-i,0), \qquad \text{if } \lambda_{X_c} = \lambda_{D^*} = \pm \\ \frac{1}{\sqrt{m_{D^*}}} & (|\vec{p}^{\,*}_{D^*}|,0,0,E_{D^*}), \qquad \text{if } \lambda_{X_c} = \lambda_{D^*} = 0 \end{cases} \tag{2.20}$$

in the B meson rest frame.

2.3. Differential $\overline{B} \to D^{(*)} l^- \bar{\nu}_l$ Decay Rates

Applying Fermi's Golden Rule to the matrix element in Equation (2.4) under consideration of the definitions above, allows defining the double differential decay rate

$$\frac{d^2\Gamma(\overline{\mathrm{B}} \to \mathrm{D}^{(*)} l^- \bar{\nu}_l)}{dw\, d\cos\theta_l} = \frac{G_F^2 \left|\mathrm{V_{cb}}\right|^2 {\eta_{\mathrm{EW}}}^2 |\vec{p}^{\,*}_{\mathrm{D}^{(*)}}|\, \mathrm{q}^2}{256\,\pi^3\, \mathrm{m_B^2}} \left(1 - \frac{\mathrm{m}_l^2}{\mathrm{q}^2}\right)^2 \Bigg[2\sin^2\theta_l \left|H_0\right|^2 + (1-\cos\theta_l)^2 \left|H_+\right|^2 + (1+\cos\theta_l)^2 \left|H_-\right|^2 + \frac{\mathrm{m}_l^2}{\mathrm{q}^2} \left(\sin^2\theta_l \left(|H_+|^2 + |H_-|^2\right) + 2|H_s + H_0\, \cos\theta_l|^2\right) \Bigg] \tag{2.21}$$

in terms of hadronic amplitudes H_{λ_W}, the four-momentum transfer squared q^2 and the angle θ_l. As already mentioned above, the dependency on the helicity angles θ_v and χ is not considered here, and thus already integrated over. For the fully differential decay rates consider [30] or [24].
Next-to-leading order corrections to the electroweak interaction in semi-leptonic decays [32] are considered by the additional factor $\eta_{\mathrm{EW}} \approx 1 + \alpha/\pi \log(\mathrm{m_Z}/\mathrm{m_B}) \approx 1.0066$ in the differential rate.
Constraining the evaluation of the decay rate to the processes with non-excited D mesons in the final state allows reducing the expression to

$$\frac{d^2\Gamma(\overline{\mathrm{B}} \to \mathrm{D} l^- \bar{\nu}_l)}{dw\, d\cos\theta_l} = \frac{G_F^2 \left|\mathrm{V_{cb}}\right|^2 {\eta_{\mathrm{EW}}}^2 |\vec{p}^{\,*}_{\mathrm{D}}|\, \mathrm{q}^2}{128\,\pi^3\, \mathrm{m_B^2}} \left(1 - \frac{\mathrm{m}_l^2}{\mathrm{q}^2}\right)^2 \Bigg[\sin^2\theta_l \left|H_0\right|^2 + \frac{\mathrm{m}_l^2}{\mathrm{q}^2} |H_s + H_0 \cos\theta_l|^2 \Bigg], \tag{2.22}$$

as the D mesons, being a scalar particle, only allows contributions with $\lambda_\mathrm{W} = 0, s$. The hadronic amplitudes H_{λ_W} embed the non-perturbative QCD dependence at low energy scales. As such, they depend on input from experimental measurements, as well as predictions from lattice QCD calculations. For more details, see e.g. [23]. In the following paragraphs and Section 2.4, common parametrization schemes for the hadronic amplitudes are be presented. The parametrization described in Section 2.4 is utilized for modifications of the MC samples used for this analysis.

Hadronic Amplitudes for $\overline{\mathrm{B}} \to \mathrm{D} l^- \bar{\nu}_l$
Inserting the expressions for the D or D^* meson final states into the definition of the hadronic matrix elements given in Equation (2.6) and matching of the possible helicity states allows for the definition of hadronic amplitudes which simplify the matrix elements. First, the D meson case is considered, which leads to the amplitudes for the processes $\overline{\mathrm{B}} \to \mathrm{D} l^- \bar{\nu}_l$.
A dedicated, alternative form factor basis [33] specific to the $\overline{\mathrm{B}} \to \mathrm{D} l^- \bar{\nu}_l$ decay processes is commonly used to express the hadronic amplitudes for this case. This basis is given by the form factors

$$V_1(w) = h_+(w) - \frac{1-r}{1+r} h_-(w) \quad \text{and} \tag{2.23}$$

$$S_1(w) = h_+(w) - \frac{1+r}{1-r} \frac{w-1}{w+1} h_-(w), \tag{2.24}$$

with $r = \mathrm{m_D}/\mathrm{m_B}$. Utilizing these and inserting the expressions given in the Equations (2.16) and (2.17) via Equation (2.6) into the general formula of the matrix elements of Equation (2.4) yields

$$H_{\pm}(\mathrm{q}^2) = 0 \tag{2.25}$$

$$H_0(\mathrm{q}^2) = \sqrt{\mathrm{m_B m_D}}\,(\mathrm{m_B} + \mathrm{m_D})\,\frac{\sqrt{w^2-1}}{\sqrt{\mathrm{q}^2}}\,V_1(w) \tag{2.26}$$

$$H_s(\mathrm{q}^2) = \sqrt{\mathrm{m_B m_D}}\,(\mathrm{m_B} - \mathrm{m_D})\,\frac{w+1}{\sqrt{\mathrm{q}^2}}\,S_1(w) \tag{2.27}$$

for the amplitudes representing the hadronic current for D mesons in the final state. The amplitude for the cases $\lambda_\mathrm{W} = \pm 1$ vanishes due to the D meson being a spin 0 pseudoscalar particle. Here the relation between the momentum transfer squared q^2 and the recoil parameter w given in Equation (2.15) is applied. Alternatively, the amplitudes can be expressed in terms of w via the inverted relation

$$\mathrm{q}^2(w) = \mathrm{m_B^2} + \mathrm{m_{X_c}^2} - 2\mathrm{m_B m_{X_c}}\,w. \tag{2.28}$$

Hadronic Amplitudes for $\overline{\mathrm{B}} \to \mathrm{D}^* l^- \bar{\nu}_l$

An analogous procedure as described in the previous paragraph is followed to obtain the hadronic amplitudes for the processes $\overline{\mathrm{B}} \to \mathrm{D}^* l^- \bar{\nu}_l$ with the D^* meson in the final state. For this case, the form factor ratios

$$R_1(w) = \frac{h_V}{h_{A_1}}, \tag{2.29}$$

$$R_2(w) = \frac{h_{A_3} + r^* h_{A_2}}{h_{A_1}} \quad \text{and} \tag{2.30}$$

$$R_0(w) = \frac{(w+1)h_{A_1} - (1 - wr^*)h_{A_2} - (w - r^*)h_{A_3}}{(1+r^*)h_{A_1}} \tag{2.31}$$

of the form factors of the heavy-quark symmetry basis are defined. Here, the fraction $r^* = \mathrm{m_{D^*}}/\mathrm{m_B}$ considers the mass of the D^* meson. Inserting the expressions of Equations (2.18) and (2.19) into Equation (2.4) and summing over the helicity states of the virtual W boson leads again to the hadronic amplitudes for the $\overline{\mathrm{B}} \to \mathrm{D}^* l^- \bar{\nu}_l$ processes:

$$H_{\pm}(\mathrm{q}^2) = \sqrt{\mathrm{m_B m_{D^*}}}\,(w+1)\,h_{A_1}(w) \;\mp\; \frac{|\vec{p}^{\,*}_{\mathrm{D}^*}|}{\sqrt{r^*}}\,h_V(w) \tag{2.32}$$

$$H_0(\mathrm{q}^2) = \frac{1}{\sqrt{r^*\mathrm{q}^2}}\left[(\mathrm{m_B^2} - \mathrm{m_{D^*}^2} - \mathrm{q}^2)\frac{w+1}{2} - \mathrm{m_B}|\vec{p}^{\,*}_{\mathrm{D}^*}|\,R_2(w)\right]\,h_{A_1}(w) \tag{2.33}$$

$$H_s(\mathrm{q}^2) = \frac{|\vec{p}^{\,*}_{\mathrm{D}^*}|}{\sqrt{r^*\mathrm{q}^2}}\,R_0(w)\,h_{A_1}(w) \tag{2.34}$$

for the four helicity states $\lambda_W = \pm, 0$ and s of the virtual W boson. The ratio $R_1(w)$ defined in Equation (2.29) can be used to replace $h_V(w)$ in Equation (2.32) to obtain a description of the hadronic amplitudes in terms of the form factor h_{A_1} and the form factor ratios given in Equations (2.32) to (2.34).

2.4. BGL Form Factor Parametrization

With the MC corrections which will be elaborated on in Section 4.2 already in mind, an additional form factor basis, the helicity basis with the form factors $\{g, f, F_1, P_1\}$ is introduced. The form factor parametrization using this basis is referred to as BGL parametrization, after the authors pioneering it [25, 26]. The MC samples for the semi-leptonic $\overline{\mathrm{B}} \to \mathrm{D}^{(*)}\ell^-\bar{\nu}_\ell$ decay processes into the light leptons $\ell = \mathrm{e}, \mu$ used for this analysis are reweighted using this parametrization and the latest experimental results provided in [34].
The relation between the four form factors of the helicity basis and the thus far used heavy-quark symmetry basis can be expressed via the four equations [23]

$$h_V = g\,\mathrm{m_B}\sqrt{r^*}, \tag{2.35}$$

$$h_{A_1} = \frac{f}{\mathrm{m_B}}\sqrt{r^*}(w+1), \tag{2.36}$$

$$h_{A_1} R_0 = P_1 \qquad \text{and} \tag{2.37}$$

$$h_{A_1}\left(w - r^* - (w-1)\,R_2\right) = \frac{F_1}{\mathrm{m_B^2}(w+1)}, \tag{2.38}$$

where the dependency of the form factors and the ratios on w is omitted. Three of the four BGL form factors can be expressed as power series, the parameters of which can then be extracted from fits to experimental data of semi-leptonic $\overline{\mathrm{B}} \to \mathrm{D}^*\ell^-\bar{\nu}_\ell$ decays into light leptons. The light lepton data gives, however, no insight into P_1, as this form factor requires sensibility to the contributions of the leptonic matrix element $L_s^+ \propto \mathrm{m}_l$ defined in Equation (2.12).
The other three form factors can be expressed as power series expansions in the Lorentz-invariant quantity

$$z(w,a) = \frac{\sqrt{w+1} - \sqrt{2}a}{\sqrt{w+1} + \sqrt{2}a} \qquad \text{with} \quad a > 0, \tag{2.39}$$

which is a conformal transformation of w into the interior of the unit disk $|z| < 1$. The parameter a is usually set to $a = 1$, such that $z(w = 1) = 0$. With z defined, the expansions of the BGL form factors are

$$g(z) = \frac{1}{P_V(z)\phi_g(z)} \sum_n a_n^g z^n, \qquad \text{with} \quad \sum_n |a_n^g|^2 \leq 1 \quad \text{and} \tag{2.40}$$

$$F_A(z) = \frac{1}{P_A(z)\phi_{F_A}(z)} \sum_n a_n^{F_A} z^n, \qquad \text{with} \quad \sum_{F_A,\,n} |a_n^{F_A}|^2 \leq 1, \tag{2.41}$$

where F_A refers to both f and F_1 and is summed over in the normalization condition. The functions $P_{g,f,F_1}(z)$ and $\phi_{g,f,F_1}(z)$ are the Blaschke factors and their weighting functions, respectively. Further details on their definition can be extracted from [26, 34].

Table 2.2.: Overview of the D^{**} mesons, which feature one unit of orbital angular momentum between the charm quark and the light anti-quark. The orbital angular momentum L is combined with the spin of the light degrees of freedom yielding the doublet $s_\ell = s'_\ell \pm L = 1/2 \pm 1$, which is then again combined with the spin of the heavy quark. J^P denote the resulting total angular momentum and parity of the meson and S its spin. The listed D^{**} meson mass values m and mass widths do **not** correspond to the numbers available in [28], but reflect the values used in the latest official Belle MC campaign and the updated values used for the MC used in this analysis.

Quark					Official Belle MC		This Work	
Content	s_ℓ	J^P	S	Meson	m/MeV	Width/MeV	m/MeV	Width/MeV
$c\bar{u}$	1/2	0^+	0	D_0^{*0}	2308.0	276	2300.0	274
$c\bar{u}$	1/2	1^+	1	$D_1^{\prime 0}$	2422.3	412	2427.0	384
$c\bar{u}$	3/2	1^+	0	D_1^0	2422.2	18.9	2420.8	31.7
$c\bar{u}$	3/2	2^+	1	D_2^{*0}	2458.9	23.0	2460.7	47.5
$c\bar{d}$	1/2	0^+	0	D_0^{*+}	2308.0	276	2349.0	221
$c\bar{d}$	1/2	1^+	1	$D_1^{\prime +}$	2422.3	412	2427.0	384
$c\bar{d}$	3/2	1^+	0	D_1^+	2427.0	28.0	2423.2	25.0
$c\bar{d}$	3/2	2^+	1	D_2^{*+}	2459.0	25.0	2465.4	46.7

For the production of the official Belle $\overline{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ MC samples the CLN parametrization was used. This parametrization scheme is based on the heavy-quark symmetry basis and the ratios thereof defined in Equations (2.29) to (2.31) and also named after the authors introducing it in [35]. A more detailed description of this form factor parametrization can be found in [24, Section 2.1].
The signal decay processes $\overline{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ have been simulated with the ISGW2 parametrization [36] for the official Belle MC samples. Their form factor parametrization is **not** altered for the presented study.

2.5. $\overline{B} \to D^{**}l^-\bar{\nu}_l$ Background Processes

The orbitally excited charm mesons already mentioned in Section 2.1 pose an important role in the analysis of the ratios $\mathcal{R}(D^{(*)})$. An overview of their properties is given in Table 2.2. They are the two states D_0^* and D_1'[1]of the doublet described by $s_\ell^\pi = 1/2^+$, as well as doublet of D_1 and D_2^* with $s_\ell^\pi = 3/2^+$ and collectively referred to as D^{**} mesons. While the former doublet exhibits rather broad mass widths of the order of $\mathcal{O}(100\,\text{MeV})$, the latter two states are narrow with mass widths of $\mathcal{O}(10\,\text{MeV})$. Here s_ℓ^π combines the spin-parity quantum numbers of the light degrees of freedom $(s_\ell^\pi)'$ and the orbital angular momentum L between the light and heavy quark.
Semi-leptonic B decays with D^{**} mesons in place of the $D^{(*)}$ mesons produce decay signatures, which can mimic the ones of the processes with the ground state $D^{(*)}$ mesons. This is due to the subsequent decay of the different D^{**} mesons into one of the ground states and mostly a single charged or neutral pion. If the additional pion cannot be detected or is assigned to the second B meson (see Figure 1.2), the remaining detectable particles can be

[1]The D_1' state is often also referred to as D_1^*.

misidentified as a $\overline{\mathrm{B}} \to \mathrm{D}^{(*)} l^- \bar{\nu}_l$ process. Especially in the case of the additional neutrino escaping detection, and thus a nonzero missing mass, the signature of the signal processes $\overline{\mathrm{B}} \to \mathrm{D}^{(*)} \tau^- \overline{\nu}_\tau$ is reproduced.
This type of background process, which is henceforth referred to as D^{**} (feed-down) background, is one of the largest contributions to systematic uncertainties in the previous measurements of $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ [9, 12, 13] at B factories. The negative impact of this background component is enhanced by the lack of quantitative knowledge about the B decay processes involving D^{**} mesons. This is in particular true for the nominal mass values of the individual D^{**} and their widths, as well as the branching fractions used for the simulation of the official Belle MC. Furthermore, the parametrization of the form factors used for the description of the $\overline{\mathrm{B}} \to \mathrm{D}^{**} l^- \bar{\nu}_l$ processes can be improved upon. The newer LLSW parametrization [37] includes terms of the order $\mathcal{O}(1/\mathrm{m}_Q)$ in the matrix element, which are not considered by the ISGW2 [36] model used thus far for the simulation of Belle MC.
For this study, the mass values and mass widths of the D^{**} mesons are updated, using the values given in Table 2.2. New MC samples are produced to replace the samples with the outdated values. Due to the state of the Belle software, these samples must be produced with the ISGW2 form factor parametrization and subsequently reweighted to the LLSW parametrization with experimental input from [38]. The reweighting procedure is elaborated on in Chapter 4. For a detailed explanation of the parametrization, as well as the fully differential decay rates, refer to [37] or [38].

2.6. Decay Rate Ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$

Inserting the differential decay rates Equation (2.21) obtained in Section 2.3 into the definition of the ratios as introduced in Chapter 1

$$\mathcal{R}(\mathrm{D}^{(*)}) = \frac{\mathcal{B}\left(\mathrm{B} \to \mathrm{D}^{(*)} \tau \nu_\tau\right)}{\mathcal{B}\left(\mathrm{B} \to \mathrm{D}^{(*)} \ell \nu_\ell\right)}, \tag{2.42}$$

results in the cancellation of many of the pre-factors in the differential decay rates. This includes the Fermi constant G_F, the electroweak correction factor η_{EW}, the B meson mass, as well as the CKM matrix element $|\mathrm{V_{cb}}|$. Thus, many SM parameters, which are subject to uncertainties are removed in these decay rate ratios. This makes the ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ particularly interesting for high precision tests of the Standard Model.

3. Data Samples

The production of MC data sets and the evaluation of a physical property on a recorded data sample represent the beginning and the end of a particle physics analysis. Throughout the analysis these two data sets provide the foundation for every step of the analysis process. In particular the MC data set is utilized to define the analysis procedure and to determine the expected performance of a measurement. Based on the recorded data set, the result of the measurement is determined. It is therefore worthwhile to take the time to understand these two data sets.
In this thesis, the term **recorded** data is used when referring to data samples recorded by the detector, whereas the more general term **data** refers to any kind of information stored on disk, tape or any other memory device and, hence, also simulated events.

3.1. Recorded Data Sample

The analysis presented in this text is based on the full data set recorded with the Belle detector in the time span between 1999 and 2010. During this period, the Belle collaboration collected the equivalent of $711\,\mathrm{fb}^{-1}$ of data at the $\Upsilon(4S)$-resonance, which corresponds to a total of

$$N_{\mathrm{B\overline{B}}} = (771.6 \pm 10.6) \times 10^6 \tag{3.1}$$

$\mathrm{B\overline{B}}$ pairs [39]. Inevitably, the data set also contains non-resonant $\mathrm{e^+e^-} \to q\bar{q}$ processes with $q = \mathrm{u, d, s, c}$ being any of the quarks lighter than the bottom quark. These *continuum* processes are not of interest to B-physics analyses and, thus, considered background. To get a better understanding of this ever-present background, dedicated off-resonance data samples have been recorded at a center-of-mass energy 60 MeV below the energy threshold for the $\Upsilon(4S)$-resonance [1, Chapter 3].

Table 3.1.: Overview of the data set recorded by the Belle experiment. The on-resonance samples are recorded with the beam energies tuned to a center-of-mass energy corresponding to the masses [28] of the respective resonances. The highlighted $\Upsilon(4S)$–data sample provides the best access to B physics and is used for the presented analysis. All off-resonance samples have been recorded at a center-of-mass energy 60 MeV below the respective resonance. [1, Chapter 3]

Resonance	Mass in MeV	Integrated luminosity in fb^{-1}		Number of Υ decays / 10^6
		On-resonance	Off-resonance	
$\Upsilon(1S)$	9460.30	5.7	1.8	102.0 ± 2.0
$\Upsilon(2S)$	10 023.26	24.9	1.7	158.0 ± 4.0
$\Upsilon(3S)$	10 355.2	2.9	0.2	11.0 ± 0.3
$\Upsilon(4S)$	10 579.4	711.0	89.4	771.6 ± 10.6
$\Upsilon(5S)$	10 885.2	121.4	1.7	7.1 ± 1.3

In addition to the $\Upsilon(4S)$-resonance, the Belle collaboration also recorded data at the center-of-mass energies of the resonances $\Upsilon(XS)$ with $X \in \{1, 2, 3, 5\}$, which are not relevant to the presented analyses.

3.2. Simulated Data Samples

The presented analysis is conducted blindly, meaning the optimization of the event-reconstruction and -selection, as well as the fit procedure is not performed on the actual recorded data set, but on simulated events.
Simulated events — also referred to as Monte Carlo (MC) events — are produced to represent the processes observed at the Belle experiment as accurately as possible to enable precision analyses. The physical properties of the decay processes in a simulated event is determined by Monte Carlo generators, such as EvtGen [40], which is dedicated to the peculiarities of decay processes involving B mesons, or PYTHIA [41] (version 6.205), which handles fragmentation and hadronization of quarks. Additionally, the PHOTOS package [42] (version 2.2) is used to estimate radiative photons, which might occur as result of final state radiation. The material interaction of the generated particles within the detector setup is simulated with GEANT 3 [43]. Among other material interactions, the production of bremsstrahlung photons originating from electrons is considered in this step of the event simulation. After the simulation of the detector and trigger system response, the MC events undergo the same treatment as recorded events. The events are reconstructed and studied with the same algorithms and selection procedures, ignoring the additional information available for MC events.
The MC truth information contains the information of the underlying true event produced by the aforementioned MC generators. This generator level information allows for a detailed evaluation of the trigger and reconstruction efficiency for a given decay topology, as well as for the definition of fit components used to extract the observables of interest.
For each recorded data set, multiple MC data sets of matching size are simulated using the same run time parameters. The factor by which the MC sample is increased compared to the recorded sample depends on the simulated decay processes. Processes of importance to an analysis can be simulated with higher statistics in dedicated productions, which allows to reduce the statistical uncertainty of the respective component. To represent the recorded data sample, all MC samples have to be combined under consideration of the respective scaling factor, as well as the sum of the branching fractions of the processes combined in a specific sample, such that the size of the total sample is equivalent to the recorded data sample and the different physical processes are represented with the correct frequency.
The Belle collaboration provides large samples covering the full diversity of B meson decay processes accessible at the B factory, as well as the common $e^+e^- \to q\bar{q}$ continuum backgrounds. These samples are referred to as *generic* Belle MC and cover the processes of interest to this analysis. Details on these samples are elaborated on in Section 3.2.1. For additional studies of the signal and normalization processes, dedicated MC samples are produced.
The $B \to D^{**} \ell \nu_\ell$ background processes are also available in the generic Belle MC, but have been simulated using outdated descriptions of the decay processes, as well as superseded mass and width values for the orbitally excited D^{**} mesons (see Table 2.2). Thus, new MC samples are produced using up-to-date D^{**} mesons masses and mass widths.
Furthermore, a better understanding of the *gap* between the inclusive and exclusive measurements of the semi-leptonic B meson decays was obtained over the last decade. This new knowledge supersedes the description of the gap used for the last offical Belle MC

production. It is also incorporated into the MC sample used for the presented analysis in the form of a new, dedicated gap MC sample. The details on the additional, privately produced Monte Carlo samples for this analysis can be found in Sections 3.2.2 and 3.2.3, which are dedicated to the new $B \to D^{**}\ell\nu$ and gap MC samples, respectively.

3.2.1. Generic Belle Monte Carlo

The official Belle MC production campaign comprises data samples for the majority of decay processes that are produced in the e^+e^-–collision events that occur at the KEKB B meson factory. As these data samples shall reflect the recorded data sample of the Belle experiment, the size of the MC samples is expressed in terms of the size of this recorded data sample. For this purpose, the term *stream* is introduced to represent the event statistics equivalent to the recorded Belle data sample. The following official Belle MC samples are available:

Generic $b \to c$
The electroweak transition from b to c quarks present the majority of the B meson decays occurring for the $B\overline{B}$ pairs produced in via the process $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ at the KEKB B meson factory. This sample also contains the decay processes most relevant to this analysis, namely the $B \to D^{(*)}\tau\nu$ signal processes and the $B \to D^{(*)}\ell\nu$ normalization processes.

Ten streams, equivalent to ten times the statistics of the recorded Belle data sample are available and used for the purpose of the presented analysis. However, some Semi-leptonic B decay processes contained within the generic Belle MC samples modeled incorrectly, and thus are newly produced in dedicated MC samples for this analysis. Events containing such B decays are removed from the generic sample.

$b \to u\ell\nu_\ell$
The B meson decays via the less frequent electroweak transitions into a u quark are provided in a dedicated MC sample with higher statistics. The sample consists of 20 streams, equivalent to 20 times the number of events expected to be in the recorded Belle data sample for these decay processes. Although this sample represents only a small fraction of the events that are recorded by the Belle detector, it must be considered, as the second B meson of the $B\overline{B}$ events in these samples decaies generically. Hence, it can decay via one of the Semi-leptonic B decay processes of interest.

Rare
This sample provides events in which one (or both) B meson originating from the $\Upsilon(4S)$ resonance, decays via a process that is not described by a tree-level electroweak interaction, as it is the case for the two aforementioned sample types. Instead, it covers rare B meson decays, such as the process $B^+ \to \ell^+\nu_\ell\gamma$. Due to the small branching fraction of the decay processes considered in this sample, it contains 50 times the number of events that are expected to be in the recorded Belle data sample for these processes.

Continuum
The continuum MC sample represents the MC expectation of the e^+e^-–collision events which do not result in the production of an $\Upsilon(4S)$ resonance. These are the processes $e^+e^- \to q\bar{q}$ with q = u, d, s, c, which are often misidentified as one of the processes of interest. This is especially comprehensible for the case of q = c, where the charm quark pair can form D mesons or higher excitations of such in the process of hadronization. During the reconstruction and signal selection process of the presented analysis, this

background is addressed by dedicated measures to reduce its contribution. The equivalent of six times of the recorded Belle data sample is available in six streams of MC samples.

Additionally, dedicated samples of the signal and normalization processes are produced for the purpose of the presented analysis. These MC samples are used as independent input to multivariate analysis methods in the course of the analysis procedure.

3.2.2. $B \to D^{**}\ell\nu$ Monte Carlo

As mentioned in Section 2.5, the semi-leptonic B meson decays into higher, orbitally excited D^{**} mesons pose an important background to the signal processes of the presented analysis. These processes are available as part of the generic Belle MC samples. However, as described in Section 2.5, the modeling of these decays, as well as the masses and widths of the individual D^{**} mesons are outdated in the generic Belle MC description. Thus, new samples of the respective decay processes are produced, using the updated masses and width as listed in Table 2.2. These privately produced MC samples contain the equivalent of ten times the number of events expected to be recorded by the Belle experiment for these processes. They replace the fractions of the generic Belle MC samples in which one of the B mesons decays via the respective decay process. This must be considered in the scaling factors of the MC samples, when combining them to a sample representative of the recorded Belledata sample.
In addition to the mass and width updates, additional decay modes are added for the D_1 meson. These are listed in Table 4.1 in the upcoming Section 4.1 and consider decays of said mesons into a $D^{(*)}$ meson and two pions.

3.2.3. Gap Monte Carlo

The inconsistency between the inclusive branching fraction measurement of the charmed semi-leptonic $B \to X_c \ell\nu_\ell$ decays and the sum of the exclusive measurements of these decays suggests that there must be additional semi-leptonic B decay processes. These processes are referred to as the *gap*. An illustration of the gap is shown in Figure 3.1

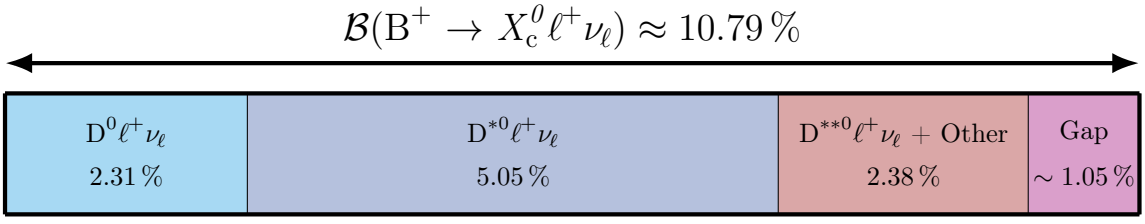


Figure 3.1.: Illustration of the branching fraction *gap* observed between inclusive branching fraction measurement of the semi-leptonic $B \to X_c \ell\nu_\ell$ and the sum of individual exclusive branching fraction measurements. The illustration shows the case of charged B meson decays and is simplified by combining the current knowledge of exclusive branching fraction of the D^{**} processes and other processes with low, but known, branching fraction contributions into on contribution. The *gap* is then the difference observed between the inclusive and the sum of the exclusive branching fraction measurements.

The exclusive branching fraction values used for this illustration are determined as described in Section 3.3.1. The inclusive semi-leptonic $B \to X_c \ell \nu_\ell$ branching fraction is taken from [28].
As this gap is expected to be filled by further, thus far unmeasured semi-leptonic B meson decays, the signature of which will be similar to the processes of interest to this analysis, it must be considered as a background to the analysis and must be simulated to the best knowledge.
As it is not precisely known which decay processes contribute to the gap, two different approaches are followed. These to descriptions of the gap are shown in Table 3.2. The decay processes listed in the column labeled as *New Gap* is the setup that is used throughout the presented analysis, if not explicitly stated otherwise. For both descriptions of the gap, new MC samples are produced for the respective decay processes. If the respective decays are also considered in the generic Belle MC samples, they are removed from those. Furthermore, all decay processes which are not considered as part of the gap description anymore, are removed from the generic Belle MC samples.

Table 3.2.: Semi-leptonic B decay processes used for the two considered descriptions of the gap. The option denoted as *New Gap* is also referred to as *alternative gap* and is used throughout this analysis, if not stated otherwise.

Old Gap	New Gap
$B \to D \eta \ell \nu_\ell$	$B \to D_0^*(\to D\eta)\ell\nu_\ell$
$B \to D^* \eta \ell \nu_\ell$	$B \to D_1'(\to D^*\eta)\ell\nu_\ell$
$B \to D \pi\pi \ell \nu_\ell$	$B \to D_0^*(\to D\pi\pi)\ell\nu_\ell$
$B \to D^* \pi\pi \ell \nu_\ell$	$B \to D_1'(\to D\pi\pi)\ell\nu_\ell$
	$B \to D_0^*(\to D^*\pi\pi)\ell\nu_\ell$
	$B \to D_1'(\to D^*\pi\pi)\ell\nu_\ell$

3.3. Branching Fractions and Monte Carlo Scaling

An important part of the update of the MC samples used for this analysis, are the new branching fractions considered for the processes available in the generic Belle MC samples, as well as the ones contained in the newly generated samples. Furthermore, all samples must be combined again to represent the entire variety of B meson decay processes expected to be in the recorded Belle data sample.
This entails the calculation of scaling factors for each sample, considering the old and new decay branching fraction values of the processes they describe, as well as the fact that some decay processes are removed from samples to be added in the form of dedicated new samples with updated descriptions. This process is simple, if the branching fractions of the altered decay processes were negligibly small, as it is often the case for studies of rare B decay. However, in the case of the presented analysis, the sum of the branching fractions of the updated decay processes is of the order of $\mathcal{O}(1\,\%)$, leading to large fraction of the generic Belle MC samples being replaced by new samples. Moreover, as the $\Upsilon(4S)$ events always contain two B meson decays, replacing a certain B decay process always affects also the branching fraction of the second B meson in this sample.
In the following Sections 3.3.1 and 3.3.2 a brief overview of the methods and sources used to calculate the updated branching fraction values, as well as the concept for the calculation of the scaling weights of the individual updated MC samples is given.

3.3.1. Branching Fractions

The branching fractions of the semi-leptonic B meson decays are updated using the latest information provided by HFLAV [11] and the Particle Data Group [28]. The prescription for the calculation of the branching fractions is described in more detail in [44] or [45].
The branching fractions of the normalization processes $B \to D^{(*)}\ell\nu$ are calculation as the isospin average from the values provided in [28]. Using the SM expectation values of the ratios $\mathcal{R}(D^{(*)})$ [11], the branching fractions of the signal processes $B \to D^{(*)}\tau\nu$ are calculated from the averages for the normalization processes. This embeds the SM expectation for the decay rate ratios into the MC cocktail.
For the main background processes $B \to D^{**}\ell\nu$, the prescription proposed in [38] is followed to obtain corrected branching fraction values based on the values provided in [11]. The branching fractions of the related processes $B \to D^{**}\tau\nu$ with τ leptons in the final state are calculated with the help of the ratios $\mathcal{R}(D^{**})$ provided in [38, Eq. 38].
The resulting branching fraction values are listed in Table 3.3. This table contains also the

Table 3.3.: Listing of the branching fraction values used to scale the generic Belle MC samples, as well as the privately produced MC samples. The table also includes the values for an alternative gap description, which features resonant processes via the broad orbitally excited D^{**} states. The decay processes listed in the last, gray category are not considered in the Monte Carlo cocktail.

Process	B^+	B^0
$B \to D^* \ell \nu_\ell$	5.508 ± 0.127	5.119 ± 0.119
$B \to D \ell \nu_\ell$	2.410 ± 0.071	2.240 ± 0.066
$B \to D_1 \ell \nu_\ell$	0.421 ± 0.027	0.392 ± 0.025
$B \to D_0^* \ell \nu_\ell$	0.420 ± 0.075	0.390 ± 0.070
$B \to D_1' \ell \nu_\ell$	0.420 ± 0.090	0.390 ± 0.084
$B \to D_2^* \ell \nu_\ell$	0.293 ± 0.032	0.273 ± 0.030
$B \to D^* \tau \nu_\tau$	1.421 ± 0.043	1.321 ± 0.040
$B \to D \tau \nu_\tau$	0.721 ± 0.022	0.670 ± 0.021
$B \to D_1 \tau \nu_\tau$	0.042 ± 0.042	0.039 ± 0.039
$B \to D_0^* \tau \nu_\tau$	0.034 ± 0.034	0.031 ± 0.031
$B \to D_1' \tau \nu_\tau$	0.025 ± 0.025	0.023 ± 0.023
$B \to D_2^* \tau \nu_\tau$	0.021 ± 0.021	0.019 ± 0.019
$B \to D_1 (\to D \pi \pi) \ell \nu_\ell$	0.242 ± 0.100	0.225 ± 0.093
$B \to D \eta \ell \nu_\ell$	0.399 ± 0.399	0.397 ± 0.397
$B \to D^* \eta \ell \nu_\ell$	0.399 ± 0.399	0.397 ± 0.397
$B \to D \pi \pi \ell \nu_\ell$	0.062 ± 0.086	0.057 ± 0.080
$B \to D^* \pi \pi \ell \nu_\ell$	0.216 ± 0.102	0.201 ± 0.095
$B \to D_0^* (\to D \eta) \ell \nu_\ell$	0.399 ± 0.399	0.397 ± 0.397
$B \to D_1' (\to D^* \eta) \ell \nu_\ell$	0.399 ± 0.399	0.397 ± 0.397
$B \to D_1' (\to D \pi \pi) \ell \nu_\ell$	0.031 ± 0.031	0.029 ± 0.029
$B \to D_0^* (\to D \pi \pi) \ell \nu_\ell$	0.031 ± 0.031	0.029 ± 0.029
$B \to D_1' (\to D^* \pi \pi) \ell \nu_\ell$	0.108 ± 0.108	0.100 ± 0.100
$B \to D_0^* (\to D^* \pi \pi) \ell \nu_\ell$	0.108 ± 0.108	0.100 ± 0.100
$B \to D_2^* (\to D^* \pi \pi) \ell \nu_\ell$	0.055 ± 0.035	0.051 ± 0.032
$B \to D_s K \ell \nu_\ell$	0.030 ± 0.013	—
$B \to D_s^* K \ell \nu_\ell$	0.029 ± 0.019	—

branching fraction values used for the two definitions of the gap, as well as the branching fraction of the decay processes involving D_1^0 mesons decaying further via $D_1 \to D\pi\pi$. The latter are not included in the original Belle MC samples, and thus produced in dedicated MC sample to be included MC cocktail for the presented analysis. The last three rows of the table list decay processes which are considered in [45] for Belle II, but are not available for Belle analyses.

3.3.2. Scaling Weight Calculation and Application

The entirety of the MC samples used for the analysis must reflect the size of the recorded data sample, i.e. the sum of the MCsample scaling weights should be the number of $B\overline{B}$ events $N_{B\overline{B}}$ recorded by the Belle detector, if no selection is applied. If a subset of the B decay processes in the generic Belle MC samples is replaced by an updated version, the events involving the respective B decay processes must be removed from the generic sample and the new sample must be scaled such that the sum of the weights results in the number of removed events.
Removing the events involving any processes of a set of B decay processes with a combined branching fraction of $\mathcal{B}_{\text{removed}}$ reduces the size of a given data set by

$$N_{\text{data set}} \cdot (2 \cdot \mathcal{B}_{\text{removed}} - \mathcal{B}^2_{\text{removed}}) \tag{3.2}$$

events if the process can occur for both B mesons and the original data set contained $N_{\text{data set}}$ events. If the process occurs with a probability of $\mathcal{B}_{\text{removed}}$ only on one side of the events, the number of removed events is

$$N_{\text{data set}} \cdot \mathcal{B}_{\text{removed}} \tag{3.3}$$

If the B decay branching fraction is updated in a signal specific decay file, without applying this change also to the respective decay definition for the generic B meson decay occurring for the second B meson of the event, the decay appears with two different probabilities for the two B mesons of a simulated $\Upsilon(4S)$ event. In this case, the number of removed events is

$$N_{\text{data set}} \cdot (\mathcal{B}_{\text{removed}} + \mathcal{B}_{\text{removed other side}} - \mathcal{B}_{\text{removed}} \cdot \mathcal{B}_{\text{removed other side}}) \tag{3.4}$$

if the decay process can occur with $\mathcal{B}_{\text{removed}}$ on one side and $\mathcal{B}_{\text{removed other side}}$ for the other B meson.
The events which are removed from one sample are replaced by samples with updated models, different decay processes or larger statistics. The scaling weights of these new samples are given by the number of produced events and the number of events they should replace. Furthermore, the scaling weights of the original samples, from which events have been removed, must be updated. To consider these changes correctly, the condition

$$\sum_i \mathcal{B}_i \stackrel{!}{=} \sum_i \mathcal{B}'_i, \tag{3.5}$$

must be satisfied for all branching fractions which have been touched during the procedure. Here $\mathcal{B}_i$ are all the branching fractions in the original sample (including the removed ones) and $\mathcal{B}'_i$ are all the branching fractions in the new sample (including the newly added ones). As most of the decay processes will be left untouched, this relation must be fulfilled in particular for the branching fractions of the processes which are changed. If, however, the sum of the touched branching fractions is changed, the remaining branching fractions must be rescaled, such that the sum of all branching fractions is one again.

Corrections made to the branching fraction of a generic decay process which can occur in any MC sample must be handled in an event specific approach, as well as in the scaling weight of the respective samples. This concerns in particular the semi-leptonic signal and normalization processes. The event specific B meson branching fraction update is discussed in Section 4.1.
Given the rules defined above, the scaling weights for a MC sample are calculated to

$$w_{\text{scale}} = \frac{N_{\mathcal{B}_{\text{new}}}}{N_{\mathcal{B}_{\text{old}}} - N_{\mathcal{B}_{\text{old, removed}}}} \tag{3.6}$$

using the original branching fractions (old) for the number of events in the denominator and the updated branching fractions (new) for the numerator. This calculation also requires the knowledge about the number of events in each of the original MC samples.

4. Monte Carlo Corrections

The corrections applied to the simulated events of the MC samples used for the presented analysis can be sorted into two main categories:

Model Corrections

The MC corrections that can be summarized under the term model corrections are related to the theoretical input used for the generation of the Belle MC samples. They concern the branching fractions used to determine the rates of the individual decay processes in the MC samples, as well as the event generator model used to describe the dynamics of the decays. Since the last official Belle MC sample production campaign, new knowledge was gained, superseding the theoretical input used for this official MC production.

The updated branching fraction values given in Section 3.3 represent the new information for one of such a MC model correction applied for this analysis. The application of these branching fraction updates is briefly discussed in Section 4.1.

The second MC correction employed for the presented analysis addresses the outdated decay models of the $B \to D^* \ell\nu$ normalization processes and the $B \to D^{**} \ell\nu$ main background processes. The update of these decay models, using new form factor parametrizations introduced in Sections 2.4 and 2.5 for the respective processes, is elaborated on in Section 4.2.

The modification of the gap description and the correction of the D^{**} masses described in Sections 3.2.2 and 3.2.3, respectively, also fall into this category of MC corrections. However, as these updates of the MC entail the production of entirely new MC samples, they are discussed in Section 3.2.

Efficiency Corrections

The MC corrections that can be classified as efficiency corrections attend to the systematic differences in efficiency that are observed between recorded data and simulation. These differences are due to imperfections in the simulation of the particle interaction in the detector and the simulation of the detector response.

These deviations between the MC and the recorded data affect all Belle analyses. Thus, these effects are studied in dedicated analyses, the results of which are commonly used to correct the observed discrepancies. The associated efficiency correction factors are available in the form of correction tables which contain weights and the corresponding uncertainties dependent on the reconstructed particles and their features. The different sources of these systematic discrepancies and their related efficiency correction factors are discussed in Section 4.3.

In the upcoming chapters of this thesis, the importance of the MC description of the studied decay processes will become apparent. Therefore, much effort is undertaken to incorporate the aforementioned MC corrections.

4.1. Branching Fraction Corrections

To update the branching fraction of the processes of interest to this analysis in the generic Belle MC samples, as well as the privately produced samples, event weights are calculated. These correction weights are based on the underlying true decay process of the event and take both B meson decays into account. For each B meson of an event, the weight $w_{\mathcal{B}}$ is calculated as

$$w_{\mathcal{B}} = \frac{\mathcal{B}_{\text{new}}(\text{B} \to X_\text{c}\ell\nu_\ell)}{\mathcal{B}_{\text{old}}(\text{B} \to X_\text{c}\ell\nu_\ell)}, \tag{4.1}$$

where the numerator is the new branching fraction value of the to-be-updated semi-leptonic B decay process and the denominator is the branching fraction of the respective process used in the original MC production. The decay processes considered for this type of branching fraction update are the ones listed in Table 3.3. As the total branching fraction of all semi-leptonic B meson decays is also updated to

$$\mathcal{B}^{\text{incl}}_{\text{new}}(\text{B} \to X_\text{c}\ell\nu_\ell) = (10.65 \pm 0.16)\,\% \qquad [11], \tag{4.2}$$

the change in the total rate for all B mesons must be considered. This is achieved by applying a correction weight to all non-semi-leptonic B meson processes given by

$$w_{\mathcal{B},\,\text{other}} = \frac{1 - \mathcal{B}^{\text{incl}}_{\text{new}}(\text{B} \to X_\text{c}\ell\nu_\ell)}{1 - \mathcal{B}^{\text{incl}}_{\text{old}}(\text{B} \to X_\text{c}\ell\nu_\ell)}, \tag{4.3}$$

where $\mathcal{B}^{\text{incl}}_{\text{old}}(\text{B} \to X_\text{c}\ell\nu_\ell)$ denotes the sum of all semi-leptonic B meson decays in the original Belle MC sample. The above described changes must also be considered for the definition of the scaling weights of the individual MC samples.
In addition to the changes made for the B decay branching fractions, some $\text{D}^{(*(*))}$ meson decay rates are also updated. Most notable are the updates of the decay rates for the two narrow, orbitally excited D^{**} mesons D_1 and D_2^*, which are tabulated in Table 4.1. For the case of these D^{**} meson decays, the branching fraction update is included in the production

Table 4.1.: Branching fraction update of the decay processes of the narrow, orbitally excited D^{**} mesons D_1 and D_2^*. The old and new branching fraction values are given in percent. The original Belle MC sample does not contain three-body D_1 decay processes. New samples containing these processes are added such that the relative rates represent the new branching fractions listed in the table.

Charged	Old	New	Neutral	Old	New
$\text{D}_1^+ \to \text{D}^{*+}\pi^0$	33.33	19.97	$\text{D}_1^0 \to \text{D}^{*0}\pi^0$	33.33	19.97
$\text{D}_1^+ \to \text{D}^{*0}\pi^+$	66.67	39.94	$\text{D}_1^0 \to \text{D}^{*+}\pi^-$	66.67	39.94
$\text{D}_1^+ \to \text{D}^+\pi^+\pi^-$	—	17.19	$\text{D}_1^0 \to \text{D}^0\pi^+\pi^-$	—	17.19
$\text{D}_1^+ \to \text{D}^+\pi^0\pi^0$	—	11.45	$\text{D}_1^0 \to \text{D}^0\pi^0\pi^0$	—	11.45
$\text{D}_1^+ \to \text{D}^0\pi^+\pi^0$	—	11.45	$\text{D}_1^0 \to \text{D}^+\pi^-\pi^0$	—	11.45
$\text{D}_2^{*+} \to \text{D}^{*+}\pi^0$	10.30	13.12	$\text{D}_2^{*0} \to \text{D}^{*+}\pi^-$	26.35	20.90
$\text{D}_2^{*+} \to \text{D}^{*0}\pi^+$	20.90	26.25	$\text{D}_2^{*0} \to \text{D}^{*0}\pi^0$	10.30	13.12
$\text{D}_2^{*+} \to \text{D}^+\pi^0$	22.90	20.21	$\text{D}_2^{*0} \to \text{D}^+\pi^-$	45.90	40.42
$\text{D}_2^{*+} \to \text{D}^0\pi^+$	45.90	40.42	$\text{D}_2^{*0} \to \text{D}^0\pi^0$	22.90	20.21

of the dedicated MC samples. The D_1 decay processes into two pions are considered in a dedicated sample as described in Section 3.2.2.

Updates of the D meson decay branching fractions are conducted considering the information provided in form of a dedicated evaluation of the Belle II charm decay tables [46]. This evaluation discusses all inconsistencies in the D meson decay tables used for the production of Belle II MC. The results of this study are identically applicable to the Belle MC decay tables. However, this would require a reproduction of the entire Belle MC data set. Thus, for the purpose of the presented analysis, only the branching fraction updates are considered as suggested in [46]. This is accomplished by applying weights to each event based on the truth information of the D meson decays appearing in the events. The weight is calculated via

$$w_{\mathcal{B}_\mathrm{D}} = \frac{\mathcal{B}_\mathrm{new}(\mathrm{D} \to \dots)}{\mathcal{B}_\mathrm{old}(\mathrm{D} \to \dots)} \; \frac{\sum \mathcal{B}_\mathrm{old}(\mathrm{D} \to \dots)}{\sum \mathcal{B}_\mathrm{new}(\mathrm{D} \to \dots)}, \tag{4.4}$$

where the sums in the second fraction go over all updated D decay processes, such that the overall normalization is not changed by this branching fraction update.

4.2. Form Factor Corrections

The update of the modeling of the semi-leptonic B decay processes considers the normalization processes $\mathrm{B} \to \mathrm{D}\ell\nu$ and $\mathrm{B} \to \mathrm{D}^*\ell\nu$, as well as the background processes $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ involving the four orbitally excited D^{**} mesons. This update of the Belle MC samples consists of a reweighting of every B meson decay process of the aforementioned categories from one form factor parametrization to another. Hence, this update is referred to as form factor corrections. An overview of the procedure is given in Table 4.2.

This table already alludes to the fact that two different methods are used to achieve the reweighting of the form factors. These methods are, for one, an analytical approach for the normalization processes, which is discussed in Section 4.2.1, as well as a method which is based on ratios of MC distributions used for the reweighting of the main background processes $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$. The latter is explained in Section 4.2.2.

Table 4.2.: Overview of the form factor update applied to the MC samples used for the presented analysis. The update concerns the normalization processes and the $\mathrm{B} \to \mathrm{D}^{**}\ell\nu$ background processes. The column labeled *Form Factor Update* shows the parametrization used in the original Belle MC and the one to which the samples are reweighted to, indicated in the form *old* $\to$ *new*. The reweighting is either conducted analytically via the eFFORT package [47], or with the frequencies observed in MC simulations combined with radial basis functions (RBF). The MC truth variables required for the reweighting procedure are listed in the column *Parameters*.

Process	Form Factor Update			Parameters	Method
$\mathrm{B} \to \mathrm{D}\ell\nu$	CLN	$\to$	BGL	w	eFFORT
$\mathrm{B} \to \mathrm{D}^*\ell\nu$	CLN	$\to$	BGL	w, $\cos\theta_\ell$, $\cos\theta_\mathrm{v}$, χ	eFFORT
$\mathrm{B} \to \mathrm{D}^{**}\ell\nu$	ISGW2	$\to$	LLSW	w, $\cos\theta_\ell$, $\cos\theta_\mathrm{v}$	RBF

4.2.1. Form Factor Reweighting for $B \to D\ell\nu$ and $B \to D^*\ell\nu$

The method denoted by eFFORT uses ratios of analytically calculated differential decay rates provided by the eFFORT package [47]. This package implements the decay rates of the normalization processes $B \to D\ell\nu$ and $B \to D^*\ell\nu$ in terms of the recoil parameter w defined in Equation (2.15) and the helicity angles defined in Figure 2.2. As already discussed in Chapter 2, the parametrization of the decay rates for the decay into a ground state D meson requires only the former, while for the $B \to D^*\ell\nu$ processes all four parameters are necessary. The eFFORT package provides the decay rates in the CLN form factor parametrization [35] as used for the production of the Belle MC, as well as for the BGL parametrization [25, 26].

The reweighting procedure applied to the MC samples of the presented analysis utilizes the values obtained in [34, Table V.] from a fit of the BGL parametrization with the configuration $(1, 1, 2)$ to untagged $B^0 \to D^{*-}\ell^+\nu_\ell$ events reconstructed in Belle data. These parameters are tabulated in Table 4.3. The statistical and systematical correlation matrices associated to these fitted parameters are provided in Appendix A in form of the Tables A.2 and A.3, respectively.

Using the uncertainties in form of the column vectors $\vec{\sigma}_{\text{stat}}$ and $\vec{\sigma}_{\text{sys}}$, as well as their correlation matrices C_{stat} and C_{sys}, variations of the form factor parameters are obtained from the covariance matrix given by

$$V_{\text{BGL}} = \vec{\sigma}_{\text{stat}}\vec{\sigma}_{\text{stat}}^{\intercal}C_{\text{stat}} + \vec{\sigma}_{\text{sys}}\vec{\sigma}_{\text{sys}}^{\intercal}C_{\text{sys}}, \tag{4.5}$$

where $\vec{\sigma}^{\intercal}$ is the transpose of the $\vec{\sigma}$, and thus $\vec{\sigma}\vec{\sigma}^{\intercal}$ the outer product of the column vectors. Pairs of six up and down variations of the form factor parameters are calculated from the eigenvalues and eigenvectors of the thus obtained covariance matrix.

The nominal parameter values and their variations are used to calculate correction factors for each event from the decay rate ratios

$$w_{\text{CLN}\to\text{BGL}}(w, \cos\theta_\ell, \cos\theta_{\text{v}}, \chi; \vec{a}_{\text{BGL}}) = \frac{d\Gamma_{\text{BGL}}(w, \cos\theta_\ell, \cos\theta_{\text{v}}, \chi; \vec{a})}{d\Gamma_{\text{CLN}}(w, \cos\theta_\ell, \cos\theta_{\text{v}}, \chi)} \frac{\Gamma_{\text{CLN}}^{\text{Tot}}}{\Gamma_{\text{BGL}}^{\text{Tot}}(\vec{a})}, \tag{4.6}$$

for which the eFFORT package is used to calculate the decay rates. Here, the vector $\vec{a}$ indicates the dependency of the BGL decay rate on the nominal parameter values and their variations as described above. The rate Γ_{CLN} is calculated using the CLN parametrization applied for the original Belle MC sample, as this is the basis for the reweighting. The

Table 4.3.: Parameters as provided in [34] for the BGL parametrization fitted in the configuration (1,1,2). The table lists the nominal parameter values, as well as their statistical and systematical uncertainties. Please refer to [34] for more details.

	$(a_n \pm \text{stat} \pm \text{sys}) \times 10^3$		
a_0^g	1.00	± 0.02	± 0.02
a_1^g	-2.35	± 0.61	± 0.66
a_0^f	0.511	± 0.004	± 0.013
a_1^f	0.67	± 0.17	± 0.30
$a_0^{F_1}$	0.30	± 0.06	± 0.08
$a_1^{F_1}$	-3.68	± 1.26	± 1.2

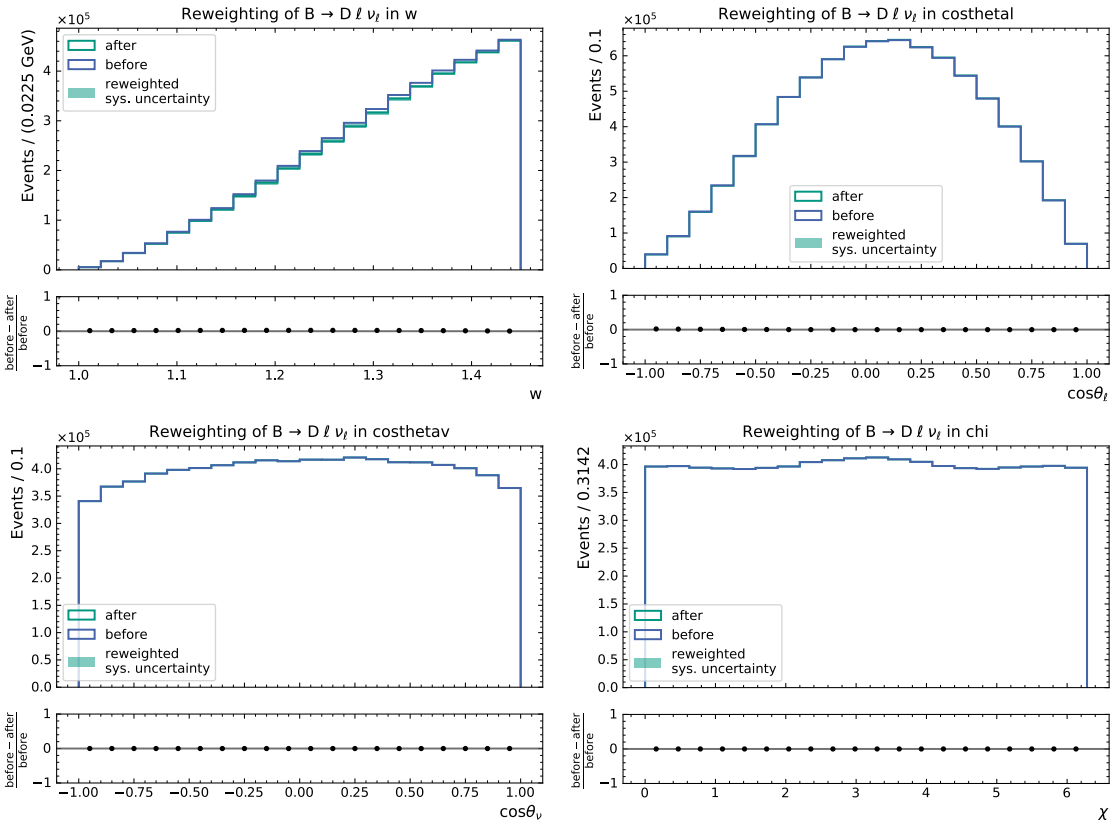


Figure 4.1.: Effect of the CLN-to-BGL form factor reweighting for the normalization processes $B \to D\ell\nu$. The distributions of the recoil parameter w (top left), as well as the three helicity angles $\cos\theta_\ell$ (top right), $\cos\theta_\nu$ (bottom left) and χ (bottom right) are plotted with (after) and without (before) the form factor reweighting applied. The distributions are based on the events obtained with the reconstruction described in Chapter 5. The uncertainty on the reweighted distribution, calculated from the form factor variations, are shown as an error band. For each distribution, the pulls calculated from the bin counts as $(N_{\text{before}} - N_{\text{after}})/N_{\text{before}}$ are included below the respective distribution plot.

ratios of the total decay rates Γ^{Tot} in the second term are added, such that the overall normalization is not changed by the procedure.

$B \to D\ell\nu$

In the case of the processes $B \to D\ell\nu$, the decay rates only depend on the recoil parameter w. The effect of this reweighting on the distributions of the four reweighting parameters w, $\cos\theta_\ell$, $\cos\theta_\nu$ and χ in the events reconstructed for this analysis as described in the upcoming Chapter 5 is shown in Figure 4.1. For these decay processes only small modifications to the MC samples are produced. The shapes of the helicity angles remain unchanged, as expected. Only small changes are visible in the distribution of the recoil parameter w. The uncertainties on the distribution in w due to the form factor variations, indicated in the form of error bands in the respective plot, are small as well.

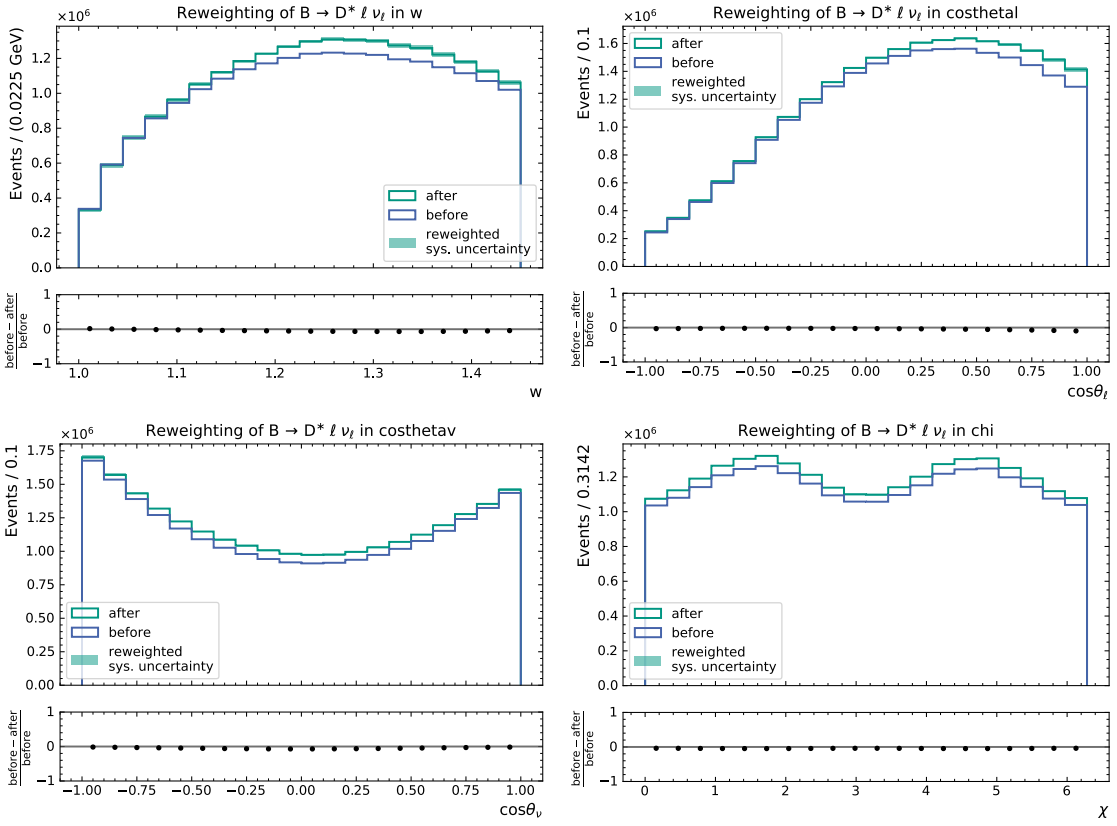


Figure 4.2.: Effect of the CLN-to-BGL form factor reweighting for the normalization processes $B \to D^* \ell\nu$. The distributions of the recoil parameter w (top left), as well as the three helicity angles $\cos\theta_\ell$ (top right), $\cos\theta_v$ (bottom left) and χ (bottom right) are plotted with (after) and without (before) the form factor reweighting applied. The distributions are based on the events obtained with the reconstruction described in Chapter 5. The uncertainty on the reweighted distribution, calculated from the form factor variations, are shown as an error band. For each distribution, the pulls calculated from the bin counts as $(N_{\text{before}} - N_{\text{after}})/N_{\text{before}}$ are included below the respective distribution plot.

$B \to D^* \ell\nu$

The form factor event weights for the processes $B \to D^* \ell\nu$ depend on the recoil parameter w, all three helicity angles $\cos\theta_\ell$, $\cos\theta_v$ and χ, as well as whether the D^* meson decays into a pion or a photon. The effect of the reweighting procedure on the $B \to D^* \ell\nu$ reconstructed for this analysis is illustrated in Figure 4.2 in the form of the distributions of the four reweighting parameters without and with the form factor correction applied. The plots show the contribution of D^* meson decays into a $D\pi$ pair as well as into a D meson and a photon. As the event selection of this analysis is optimized to select $B \to D^* \tau\nu$ events, the distributions shown in Figure 4.2 exhibit a change in their normalization. This is related to the modification of the w distribution due to the form factor weights. In this distribution and the distribution of $\cos\theta_\ell$, the most notable shape changes are produced by the reweighting process. The uncertainties on the reweighted distributions remain small.

4.2.2. Form Factor Reweighting for $B \to D^{**}\ell\nu$

Due to the usage of outdated mass values for the orbitally excited D^{**} mesons in the official Belle MC production, entirely new MC samples for the processes involving these D^{**} mesons are produced in the scope of the presented analysis. However, the updated modeling of these semi-leptonic B decay processes using the LLSW form factor parametrization [37] instead of the ISGW2 parametrization [36] used in the official Belle MC production is not included in this private MC production, as the LLSW model is not available in the Belle analysis software. Therefore, the update to the new LLSW model is also conducted by reweighting the simulated events based on the underlying MC truth information.

The procedure for this form factor reweighting of the main background processes $B \to D^{**}\ell\nu$ was developed in [48] in the scope of this analysis. Instead of utilizing the analytical definition of the decay rates, generator level MC samples are produced with the original parametrization used in the Belle MC production, as well as with the LLSW form factor parametrization. For the latter, the form factor parameters obtained in [38, Table X.] (approximation C), as well as the corresponding uncertainties and correlation matrices are used. For the generation of these generator level MC samples, the Belle II Analysis Software Framework is employed, which implements the LLSW model. Variations of the LLSW form factor parameters are produced similarly to the procedure described in Section 4.2.1. Using these variations as input to the LLSW model, additional generator level MC samples are produced to obtain varied MC truth distributions for the LLSW model.

Based on the generator level MC samples, the event frequency for the ISGW2 and the LLSW model in bins of the MC truth variables w, $\cos\theta_\ell$ and $\cos\theta_\mathrm{v}$ are calculated. The MC truth variables considered for this as well as the number of bins and the variable range used for the definition of these histograms depend on the D^{**} meson type in the final state of the to-be-reweighted decay process. The definitions for the histograms for each D^{**} meson are tabulated in Table 4.4.

Table 4.4.: Information on the distributions used for the MC ratio based reweighting procedure of the $B \to D^{**}\ell\nu$ decay processes. Depending on the D^{**} meson in the final state of the decay process, the generator level MC distributions are evaluated in histograms of up to three of the MC variables w, $\cos\theta_\ell$ and $\cos\theta_\mathrm{v}$. The number of bins (# Bins) considered for each dimension depends on the number of dimensions. The range of each dimension depends on the respective MC truth variable.

D^{**} Meson	Variable	# Bins	Range
D_0^*	w	50	[1.0, 1.425]
D_1'	w	20	[1.0, 1.425]
	$\cos\theta_\ell$	20	[−1.0, 1.0]
D_1	w	10	[1.0, 1.42]
	$\cos\theta_\ell$	10	[−1.0, 1.0]
	$\cos\theta_\mathrm{v}$	10	[−1.0, 1.0]
D_2^*	w	20	[1.0, 1.42]
	$\cos\theta_\ell$	20	[−1.0, 1.0]

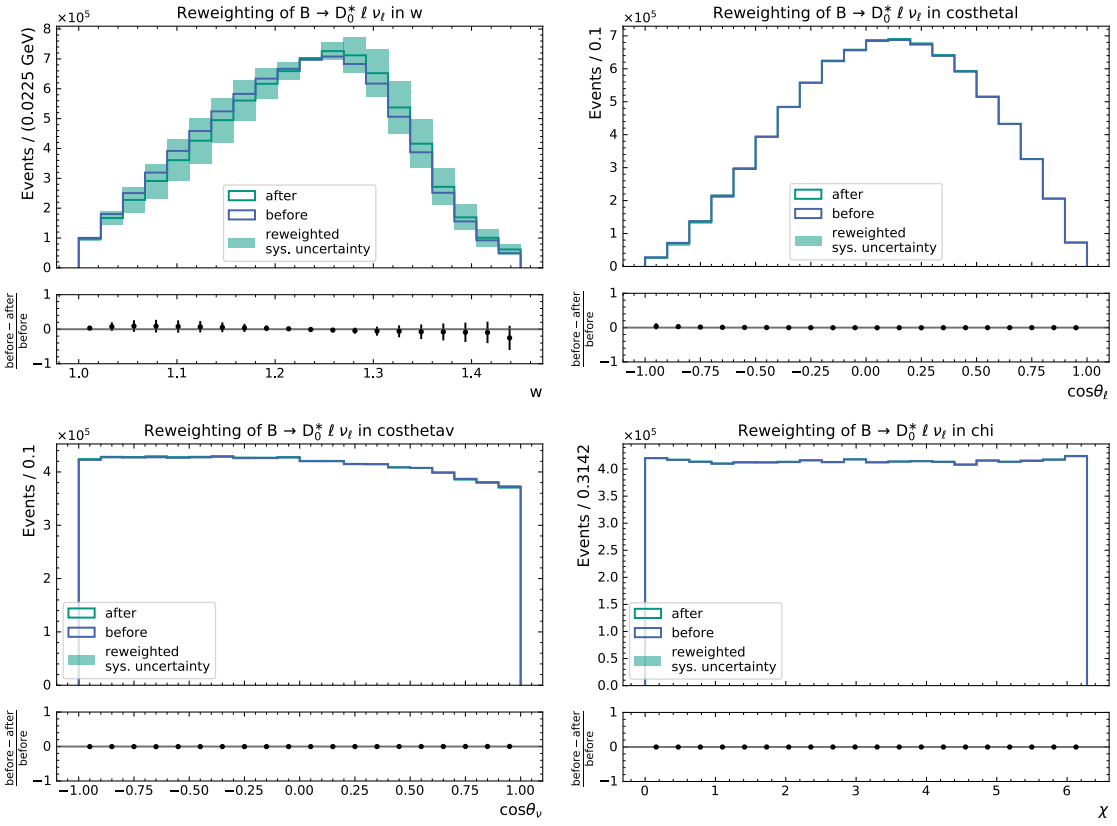


Figure 4.3.: Effect of the ISGW2-to-LLSW form factor reweighting for the processes $B \to D_0^* \ell \nu$. The distributions of the recoil parameter w (top left), as well as the three helicity angles $\cos\theta_\ell$ (top right), $\cos\theta_\nu$ (bottom left) and χ (bottom right) are plotted with (after) and without (before) the form factor reweighting applied. The distributions are based on the events obtained with the reconstruction described in Chapter 5. The uncertainty on the reweighted distribution, calculated from the form factor variations, are shown as an error band. For each distribution, the pulls calculated from the bin counts as $(N_{\text{before}} - N_{\text{after}})/N_{\text{before}}$ are included below the respective distribution plot.

The ratios of the histogrammed MC truth distributions are given by

$$w_{i,j,k} = \frac{N^{\text{LLSW}}_{i,j,k}}{N^{\text{ISGW2}}_{i,j,k}} \tag{4.7}$$

where i, j, k are the bin indices of the up to three dimensions (w, $\cos\theta_\ell$ and $\cos\theta_\nu$) of the D^{**} meson dependent binning, according to Table 4.4. The numerator is given by the events observed in the respective bin for the generator level distributions obtained using the nominal LLSW parametrizations, or the variations thereof. The denominator is based on the generator level distributions based on the ISGW2 model.
The thus obtained ratios are smoothed using radial basis functions as described in [48] to obtain approximate analytical expressions for these ratios in terms of the respective reweighting parameters w, $\cos\theta_\ell$ and $\cos\theta_\nu$.
The Figures 4.3 to 4.6 show the effect of the above described reweighting process of the $B \to D^{**} \ell \nu$ processes on the MC samples of used for the presented analysis. The change due to this reweighting from the ISGW2 model to the LLSW parametrization is more

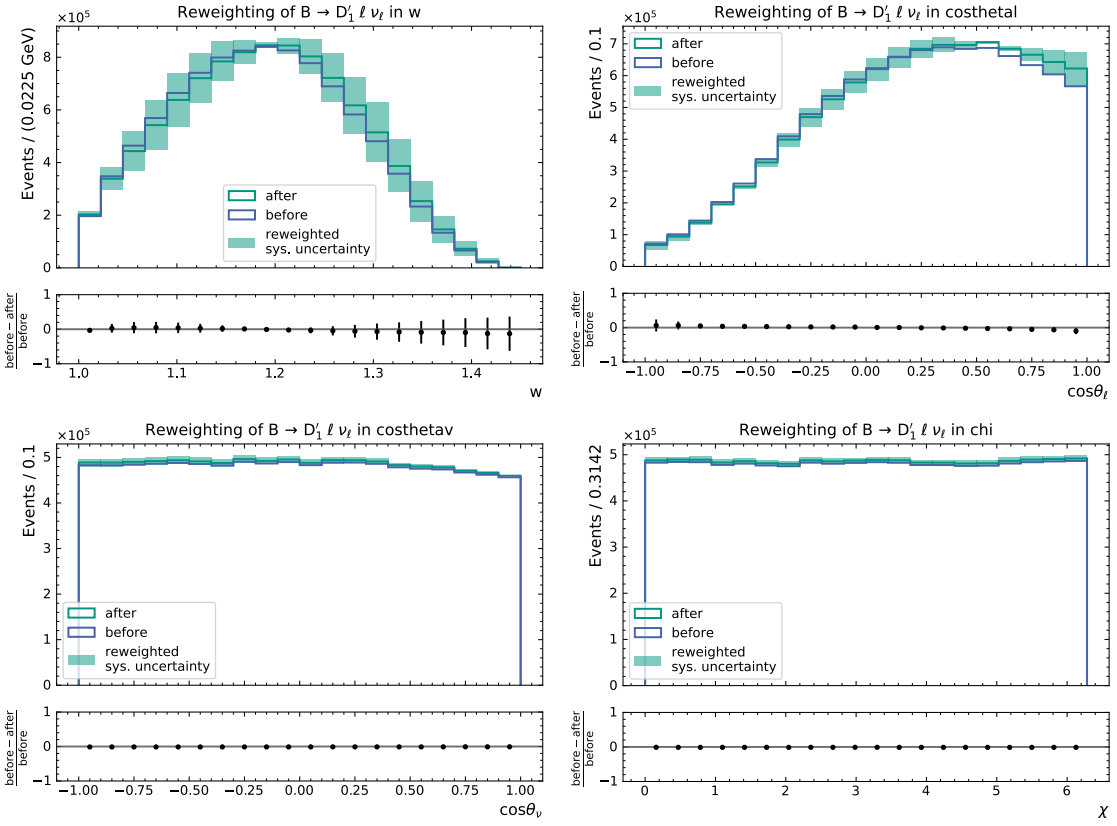


Figure 4.4.: Effect of the ISGW2-to-LLSW form factor reweighting for the processes $B \to D_1' \ell\nu$. The distributions of the recoil parameter w (top left), as well as the three helicity angles $\cos\theta_\ell$ (top right), $\cos\theta_\nu$ (bottom left) and χ (bottom right) are plotted with (after) and without (before) the form factor reweighting applied. The distributions are based on the events obtained with the reconstruction described in Chapter 5. The uncertainty on the reweighted distribution, calculated from the form factor variations, are shown as an error band. For each distribution, the pulls calculated from the bin counts as $(N_{\text{before}} - N_{\text{after}})/N_{\text{before}}$ are included below the respective distribution plot.

visible than what is observed for the reweighting of the normalization processes $B \to D^{(*)}\ell\nu$ discussed in the previous Section 4.2.1. Significant modifications to the MC distributions of the kinematic variables used for the reweighting procedure are observed especially for the process $B \to D_1\ell\nu$ displayed in Figure 4.5. Furthermore, the uncertainty bands on the reweighted distributions, which are calculated from the varied LLSW form factor parameters, indicate a significant uncertainty on the shapes of the $B \to D^{**}\ell\nu$ processes, which can be considered as a systematical uncertainty in the presented analysis.

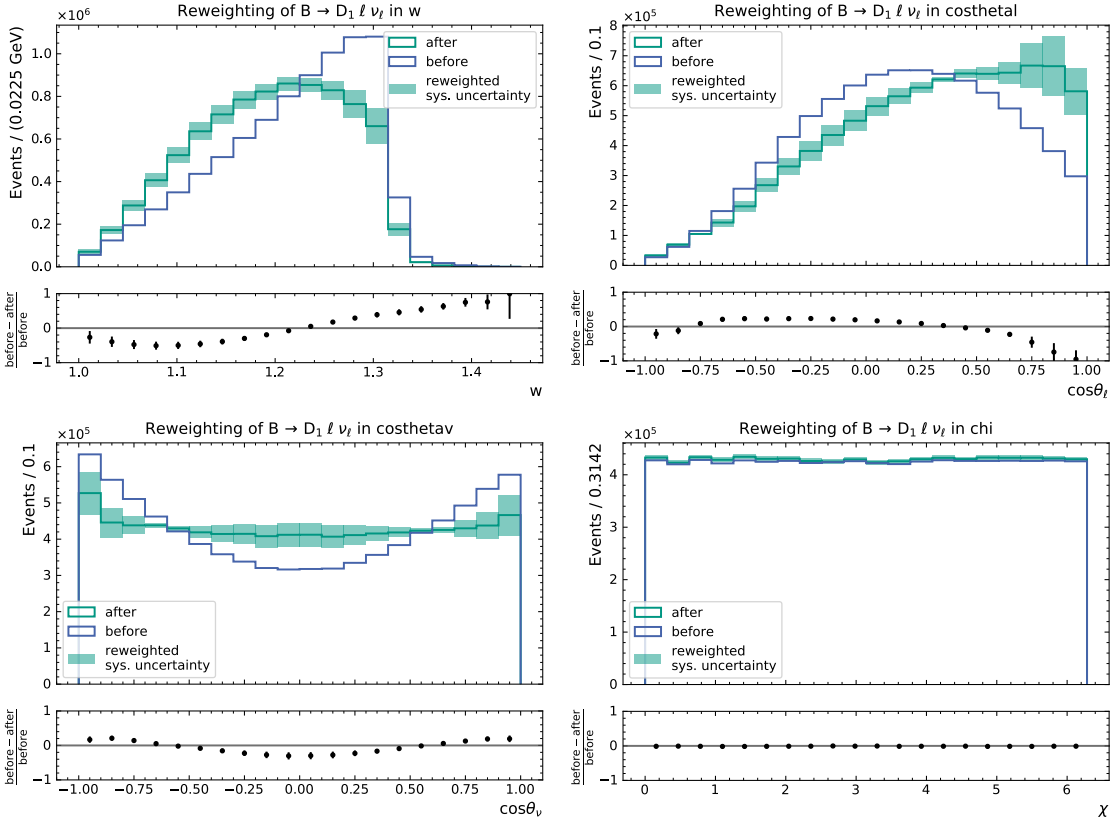


Figure 4.5.: Effect of the ISGW2-to-LLSW form factor reweighting for the processes $B \to D_1 \ell \nu$. The distributions of the recoil parameter w (top left), as well as the three helicity angles $\cos\theta_\ell$ (top right), $\cos\theta_v$ (bottom left) and χ (bottom right) are plotted with (after) and without (before) the form factor reweighting applied. The distributions are based on the events obtained with the reconstruction described in Chapter 5. The uncertainty on the reweighted distribution, calculated from the form factor variations, are shown as an error band. For each distribution, the pulls calculated from the bin counts as $(N_{\text{before}} - N_{\text{after}})/N_{\text{before}}$ are included below the respective distribution plot.

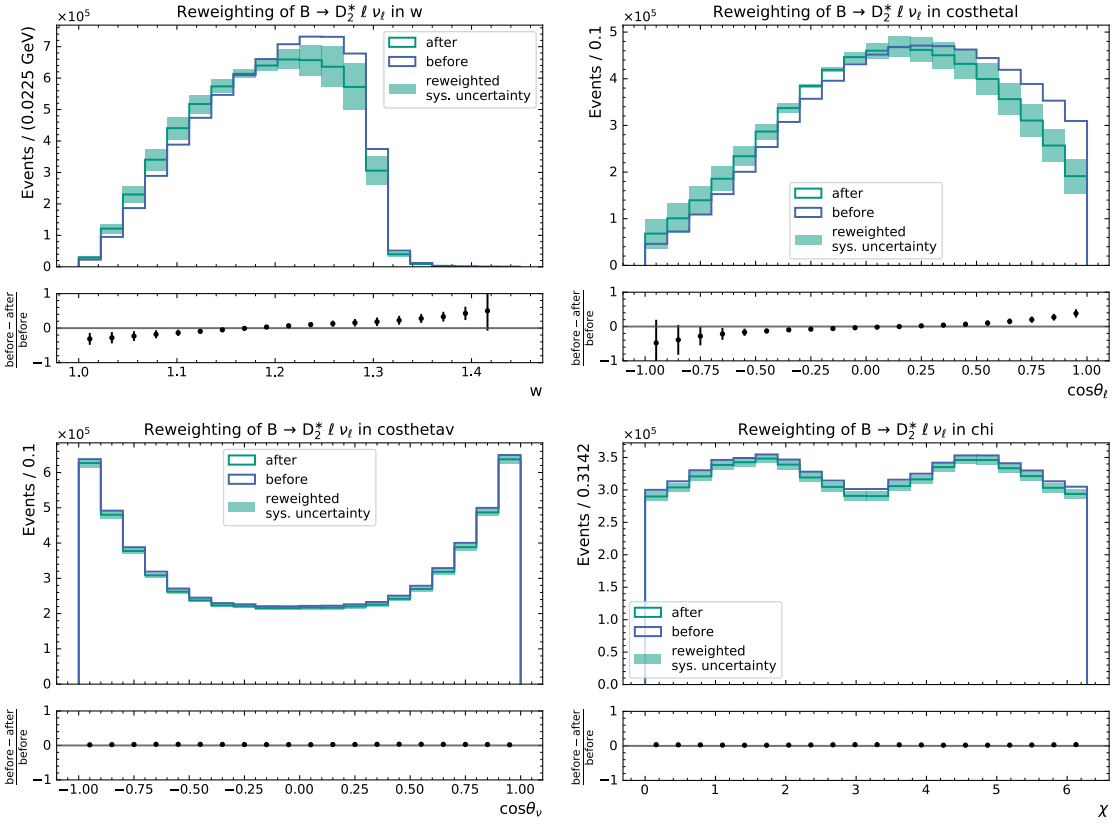


Figure 4.6.: Effect of the ISGW2-to-LLSW form factor reweighting for the processes $B \to D_2^* \ell \nu$. The distributions of the recoil parameter w (top left), as well as the three helicity angles $\cos\theta_\ell$ (top right), $\cos\theta_v$ (bottom left) and χ (bottom right) are plotted with (after) and without (before) the form factor reweighting applied. The distributions are based on the events obtained with the reconstruction described in Chapter 5. The uncertainty on the reweighted distribution, calculated from the form factor variations, are shown as an error band. For each distribution, the pulls calculated from the bin counts as $(N_{\text{before}} - N_{\text{after}})/N_{\text{before}}$ are included below the respective distribution plot.

4.3. Efficiency Corrections

Efficiency correction factors must take into account

- imperfections in the simulation of the particle interaction in the material of the detector,
- imperfections in the detector response to the signals caused by these interactions, as well as
- the difference in the performance of the reconstruction algorithms when facing simulated or recorded events due to said imperfections.

As these effects are dependent on the different particle types which interact with different parts of the detector, as well as their kinematic properties, several efficiency corrections are considered.
The efficiency corrections discussed within this section rely partly on the reconstructed event. Hence, the content of this section may occasionally refer to terms which are introduced in Chapter 5.

4.3.1. Tracking Corrections

The efficiency difference of the Belle tracking algorithm for high momentum particle tracks in recorded versus simulated events is studied in [49]. This study evaluates the efficiency for the reconstruction of particle tracks with a transversal momentum of

$$\mathrm{p_t} > 200\,\mathrm{MeV} \tag{4.8}$$

in partially and fully reconstructed $\mathrm{D}^* \to \mathrm{D}^0\pi$ decays, where the D^0 meson is reconstructed in the decay chain $\mathrm{D}^0 \to (\mathrm{K}^0_\mathrm{S} \to \pi^+\pi^-)\pi\pi$. No significant difference in efficiency is observed between recorded data and MC. Hence, no correction is applied. However, the study suggests applying a systematic uncertainty of 0.35 % per particle track with a transversal momentum above the threshold of 200 MeV. This correction factor is considered for every charged particle track on the signal side B meson decay tree.
A further correction of the Belle tracking algorithm, which affects not the reconstruction efficiency, but the resolution of the track parameters, is discussed in Section 5.1.1.

4.3.2. Slow Pion Correction

The reconstruction efficiency correction for low $\mathrm{p_t}$ particle tracks, with

$$\mathrm{p_t} < 200\,\mathrm{MeV} \tag{4.9}$$

is provided in [50]. For particles of such low momentum, the π is used as a particle hypothesis, as the particle does not reach any particle identification sub-detector. The investigation conducted in [50] uses slow pion tracks produced in the decays of charged or neutral D^* mesons in the processes $\mathrm{B}^0 \to \mathrm{D}^{*-}\pi^+$ and $\mathrm{B}^+ \to \overline{\mathrm{D}}^{*0}\pi^+$. The D^* decays into a D meson and a neutral or charged slow π_s are evaluated to obtain correction factors for both types of slow pions π_s.
The correction factors provided in [50] are given in six equidistant bins in the particle's momentum between 50 MeV and 200 MeV for charged and neutral slow pions π_s. For

each momentum bin and slow pion type, the correction factors ρ_{π_s} and their associated uncertainties are given in the form

$$\rho_{\pi_s\, i} \pm \sigma^{\text{uncorr}}_{\text{stat}\, i} \pm \sigma^{\text{corr}}_{\text{stat}\, i} \pm \sigma^{\text{corr}}_{\text{sys}\, i}, \tag{4.10}$$

where $\sigma^{\text{uncorr}}_{\text{stat}}$ denotes an uncorrelated statistical uncertainty, while $\sigma^{\text{corr}}_{\text{stat}}$ and $\sigma^{\text{corr}}_{\text{sys}}$ are bin-correlated statistical and systematic uncertainties, respectively. The index i denotes the momentum bin.
These uncertainties are used to calculate variations of each of the slow pion correction factors. For this purpose, a covariance matrix C_{π_s} is defined in terms of the six momentum bins. The diagonal elements of this covariance matrix are given by the quadratic sum of all three aforementioned uncertainties:

$$(C_{\pi_s})_{ii} = (\sigma^{\text{uncorr}}_{\text{stat}\, i})^2 + (\sigma^{\text{corr}}_{\text{stat}\, i})^2 + (\sigma^{\text{corr}}_{\text{sys}\, i})^2. \tag{4.11}$$

The off-diagonal elements of C_{π_s} are given by the combinations of the two correlated uncertainties to

$$(C_{\pi_s})_{ij} = \sigma^{\text{corr}}_{\text{stat}\, i}\sigma^{\text{corr}}_{\text{sys}\, j} + \sigma^{\text{corr}}_{\text{stat}\, j}\sigma^{\text{corr}}_{\text{sys}\, i}. \tag{4.12}$$

Using the eigenvalues $\vec{\lambda}$ belonging to C_{π_s}, and the matrix P whose columns are given by the eigenvectors of C_{π_s}, both of which are defined by the eigendecomposition of the covariance matrix C_{π_s}

$$\vec{\lambda} = P^{-1} C_{\pi_s} P, \tag{4.13}$$

variations of the factors $\vec{\rho}_{\pi_s}$ can be calculated via

$$\hat{\vec{\rho}}_{\pi_s} = \vec{\rho}_{\pi_s} \times P^{-1}$$

$$\widetilde{\vec{\rho}}_{\pi_s,\, \pm k} = \hat{\vec{\rho}}_{\pi_s} \pm \sqrt{\lambda_k} \tag{4.14}$$

$$\vec{\rho}\,'_{\pi_s,\, \pm k} = \widetilde{\vec{\rho}}_{\pi_s,\, \pm k} \times P, \tag{4.15}$$

where in Equation (4.14) the kth up and down variation is calculated by adding or subtracting the square root of the kth uncorrelated uncertainty given by the eigenvalue λ_k, as explained in [51]. The thus obtained correction factor variations can be used to evaluate the correlations and shape changes in distributions due to this systematic effect.

4.3.3. K^0_S Reconstruction Efficiency Correction

MC efficiency correction factors and the associated systematic uncertainties for the reconstruction of neutral K^0_S candidates are taken from [52]. In this study of the K^0_S reconstruction efficiency at Belle, a pure sample of fully reconstructed D^* decays is used to evaluate the difference in reconstruction efficiency between recorded data and MC. The D^* mesons are reconstructed in the decay chain $D^* \to D^0\pi$ with $D^0 \to K^0_S\,\pi^+\pi^-$ and the K^0_S being reconstructed from a pair of charged pions.
The study [52] provides the central value of the efficiency correction weight, as well as the statistical and systematic uncertainties thereof, in seven bins of the reconstructed K^0_S particle momentum in the laboratory reference frame. The seven momentum bins are six equidistant bins between 0.5 GeV and 3.0 GeV and an additional bin for candidates with higher momenta. For the purpose of the presented analysis, variations of the K^0_S efficiency correction weights are calculated in the same manner as described in the previous Section 4.3.2. As the K^0_S efficiency correction weights are given in terms of seven particle momentum bins, seven up and down variations of the weights are produced.

4.3.4. Particle Identification Corrections

A further aspect of the event reconstruction is the identification of the particle type of charged tracks. This assignment of a particle hypothesis to a given track candidate is based on the inputs of the different sub-detectors of the Belle experiment. As such, this identification depends also on the simulation of material interaction and detector response for the case of the MC events. The imperfections in the simulation result in discrepancies in the performance of the particle identification efficiencies for recorded data versus MC. These effects are corrected in MC samples with dedicated efficiency correction weights for the different charged particle types. Particle identification efficiency correction factors are available for

Lepton Identification Efficiency
The efficiency correction factors for the lepton identification in MC is determined in the study [53]. This study evaluates the efficiency of the identification of e and μ candidates in the two-photon process $e^+e^- \to e^+e^- \ell^+\ell^-$, where $\ell = e, \mu$. This sample provides large statistics for the determination of the correction factors. However, as the correction is destined to be applied to leptons from B meson decays, the results of the study are crosschecked on inclusive $B \to X J/\psi$ events with $J/\psi \to \ell^+\ell^-$ to determine the effect of the hadronic environment on the efficiency correction factors.

The provided correction factors depend on the applied lepton identification selection criterion, the particle's momentum and polar angle of the particle track, as well as the run time period in which the event was recorded. In addition to the correction factor, the associated statistical uncertainty, as well as two systematic uncertainties are provided. The first systematic uncertainty is determined from the fluctuation of the correction factors in the different run periods and is thus correlated among all runtime periods. The second systematic uncertainty is determined from the crosscheck to the $B \to X J/\psi$ events.

Lepton Fake Rate
The misidentification rate of leptons is studied in the scope of the Belle analysis of semi-inclusive semi-leptonic B decays [54]. This analysis provides lepton fake rate correction factors for e and μ candidates in bins of the particle momentum and the polar angle, as well as for different requirements on the lepton identification. Additionally, the uncertainty on the fake rate correction factor is provided.

Hadron Identification Efficiency and Fake Rate
The evaluation of hadron identification efficiency correction and fake rate correction is described in [55]. In the scope of this study, the decay processes $D^{*+} \to D^0\pi^+$ with a subsequent decay of the D meson into a $K^-\pi^+$ pair are utilized to determine the MC correction factors for the identification of pions and charged kaons, as well as for the rate of misidentifying one as the other. Correction factors and the associated statistical and systematic uncertainty are provided in bins of the particle momentum in the laboratory reference frame and the polar angle of the particle track. The corrections further depend on the selection requirement applied to the respective particle candidate.

For each of the above listed particle identification efficiency correction, 20 variations of the correction factors are obtained by drawing from Gaussian distributions centered about the nominal correction factor value. The variances of the Gaussian distributions are determined from the uncertainties provided for each type of particle identification efficiency correction, as described in [51]. The correlation of uncertainties is considered by choosing the random seeds used for the process accordingly.

4.3.5. B-Tagging Efficiency Correction

The utilization of the hadronic FEI B-tagging algorithm for the automated reconstruction of the second B meson in an event for the presented analysis is described in Section 5.2. In the scope of this Section 4.3, the efficiency correction associated to this tagging algorithm shall be introduced briefly. Hitherto, two calibration studies have been conducted for the efficiency correction of hadronic FEI on recorded Belle data. A simple calibration study is presented in [56], which provides two correction factors, differentiating between charged and neutral B meson FEI reconstruction modes:

$$\begin{aligned} \epsilon_{\text{charged}}^{\text{had. FEI}} &= 0.810 \pm 0.012 \pm 0.054 \\ \epsilon_{\text{neutral}}^{\text{had. FEI}} &= 0.853 \pm 0.059 \pm 0.058, \end{aligned} \tag{4.16}$$

with statistical and systematic uncertainties, respectively. Such correction factors are assumed to cancel in the decay rate ratios $\mathcal{R}(\mathrm{D}^{(*)})$ and are therefore only considered for comparisons of recorded data and MC distribution in this analysis.
A more comprehensive and sophisticated calibration of the hadronic FEI B-tagging algorithm is conducted in the scope of a $|\mathrm{V_{cb}}|$ measurement [57] using the data samples produced as part of this thesis. This FEI calibration uses hadronically tagged, inclusively reconstructed $\mathrm{B} \to X_\mathrm{c}\ell\nu_\ell$ processes in an analysis approach, which is closely coordinated with the presented $\mathcal{R}(\mathrm{D}^{(*)})$ analysis. Furthermore, this FEI calibration utilizes the herein presented updated Belle MC samples. The quality of the calibration factors is improved, as the study provides the factors dependent on the FEI reconstruction modes, as well as on the quality (measured in terms of $\mathcal{P}_{\text{FEI}}$) of the $\mathrm{B_{tag}}$ candidate produced by the B-tagging algorithm. Please refer to Section 5.2 for a more detailed description of the FEI B-tagging algorithm. The results of the FEI calibration study introduced in [57] are not considered in this text, as they were made available only shortly before this thesis is due. However, first insights based on the results of this FEI calibration study will be discussed in the outlook of this thesis, as part of Chapter 11.

5. Event Reconstruction

The n-tuples produced in the scope of this work serve as input for four related analyses, namely

1. **a measurement of the ratios $\mathcal{R}(D^{(*)})$ with leptonic τ-reconstruction,**
2. a measurement of the branching fractions of the processes $B \to D^{**}\ell\nu$,
3. a study of the form factors of the $B \to D^{**}\ell\nu$ processes, and
4. a measurement of $|V_{cb}|$ with hadronically tagged $B \to D^{*}\ell\nu_\ell$ decays;

all of which are conducted with tag side B mesons reconstructed in hadronic decay modes. The first of these analyses is the subject of this text and is still in progress at the time of writing. The studies of the $B \to D^{**}\ell\nu$ processes, with $D^{**} = D_1, D_1', D_0^*, D_2^*$ (see Section 2.5), have been prepared as MC studies in [48] and [58] in close cooperation with this work. They provided important feedback and a better understanding of the MC description of the main background processes. The $|V_{cb}|$ measurement is in Belle-internal review at the time of writing and documented in [57].
In this chapter the event reconstruction for the n-tuples of all four analyses is described. This comprises the reconstruction of entire events for which a B meson decaying via the semi-leptonic processes $B \to D^{(*,**)}\ell\nu_\ell$ can be found. This B meson candidate is henceforth denoted as signal B_{sig} or simply B_{sig} meson; the particles it is reconstructed from, including itself, is referred to as the signal side. $D^{(*,**)}$ represents any of the charmed meson states

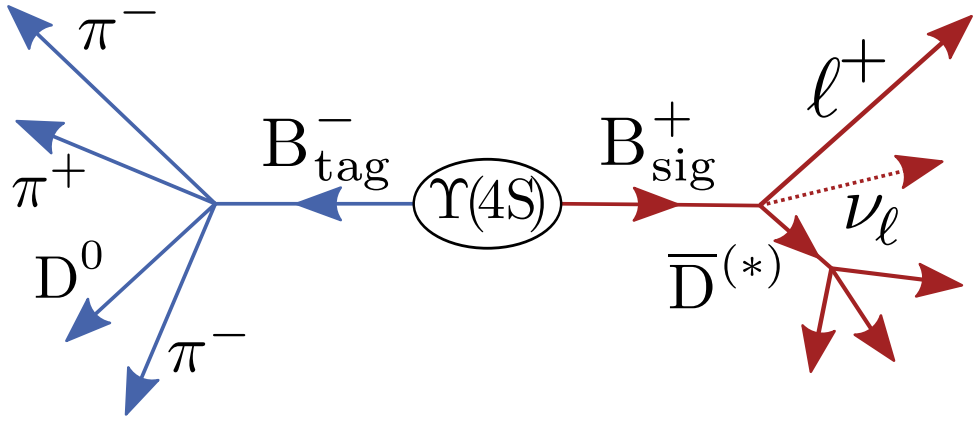


Figure 5.1.: Illustration of an $\Upsilon(4S)$ decay in its rest frame into two back-to-back B mesons. One of the B mesons further decays via a decay process of interest (here shown on the right), and thus is called B_{sig}. The second (left) B meson is reconstructed in a generic hadronic decay mode by a tagging algorithm and referred to as B_{tag}. See Figure 1.2 for the process with a τ lepton on the signal side. Without loss of generality, the process is displayed for the case of a pair of charged B mesons.

D, D^* or D^{**}. The B_{sig} meson originates from the decay of an $\Upsilon(4S)$ meson in the process $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\overline{B}$ as illustrated in Figure 5.1. It is accompanied by a second B meson which is reconstructed in an automated procedure called B tagging. Hence, this B meson is referred to as B_{tag}, and it is the parent particle of the tag side decay tree.
The reconstruction procedure of the entire event can be split into three steps, comprising

1. the reconstruction of the signal side B_{sig} meson in Section 5.1,
2. the reconstruction of the tag side B_{tag} meson in Section 5.2, as well as
3. the combination of the two to the $\Upsilon(4S)$ meson in Section 5.3.

These steps are identical for the four analyses listed above.
The reconstruction of the signal side is performed first, as it is less computationally expensive[1] than the tag side reconstruction due to fewer combinatorial possibilities. Further selection steps described in Chapter 6 refine the event selection for the presented analysis. The specifics of the event selection of the other three analyses are not the subject of this text.

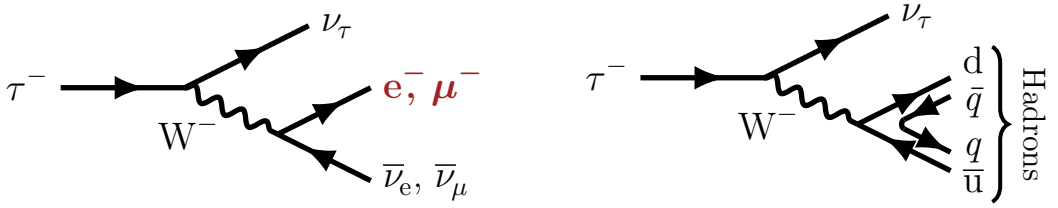
5.1. Signal Side Reconstruction

For all described analyses the signal side B_{sig} meson is reconstructed in the same basic manner. As illustrated in the Decays (5.1) to (5.3), the signal B_{sig} meson is always formed by combining an excited or unexcited D meson with a light lepton ℓ. The neutrino ν_ℓ associated to the lepton ℓ is invisible to the detector, and thus cannot be reconstructed.

$$\begin{array}{lll} B_{sig} \rightarrow D\ell\bar{\nu}_\ell & B_{sig} \rightarrow D^*\ell\bar{\nu}_\ell & B_{sig} \rightarrow D^{**}\ell\bar{\nu}_\ell \\ \quad \hookrightarrow K\pi\pi & \quad \hookrightarrow D\pi & \quad \hookrightarrow D^{(*)}\pi \\ & \quad\quad \hookrightarrow K\pi\pi & \quad\quad \hookrightarrow \ldots \end{array} \qquad (5.1, 5.2, 5.3)$$

The cases involving the two light leptons $\ell = e, \mu$ constitute the **normalization modes** in the ratios $\mathcal{R}(D^{(*)})$. In the context of the Belle detector, the muon μ can be considered as stable particle in the sense that it will not decay within the sensitive volume of the detector. Hence, both, the electron e and the muon μ are reconstructed as final state particles by the Belle tracking algorithms.
The higher mass of the τ lepton of $m_\tau = 1776.86\,\text{MeV}$ [28] provides a larger phase space for the decay and allows for a greater variety of decay products. This implicates the τ lepton's low life time of $\tau_\tau = 290.3 \times 10^{-15}\,\text{s}$ [28], due to which the τ decays within the center of the detector. It can decay into the two lighter leptons via purely leptonic processes as illustrated in Figure 5.2a or under the production of hadronic daughter particles depicted in Figure 5.2b.
The latter hadronic processes involve one τ neutrino ν_τ and a multitude of combinations of light mesons. Combined, they account for $\approx 64.8\,\%$ of all τ decays (see Figure 5.3). The two leptonic decay processes into an electron or muon cover a branching fraction of $\approx 35.2\,\%$

[1] Ideally, the computational effort would be reduced best by the means of a skim, i.e. the production of a subset of the Belle data containing all hadronically tagged events by applying the hadronic FEI in a first step. The production of n-tuples for specific analyses can then be performed based on this smaller data set in a second step. However, such a skim requires additional storage capacity, which is an acceptable compromise only if multiple analyses benefit from the skimmed data. This is not the case for the Belle data set, which is why the described path was chosen.



(a) leptonic τ decay (b) hadronic τ decay

Figure 5.2.: The two Feynman diagrams show the leading order Standard Model processes for the leptonic and hadronic τ decay mediated by a charged W boson. For this analysis only the **leptonic τ decay** is considered, leading to an event signature similar to the normalization modes which requires only a **light lepton $\ell = e, \mu$**.

and entail two additional neutrinos, adding further invisible particles to the decay signature (see Decay (5.4)). Compensating for these disadvantages is the fact, that the resulting detectable end products of the leptonic τ decay are the same as for the normalization modes (see Decays (5.1) and (5.2)). Hence, the signature of the signal and normalization modes differ only in the number of invisible neutrinos involved, namely three and one, respectively. This is utilized for the $\mathcal{R}(D^{(*)})$ measurement, where the similarity in the signatures of signal and normalization modes results in a cancelation of systematical uncertainties in the ratio. This is in addition to the theoretical advantages explained in Chapter 2. The difference in the number of invisible neutrinos allows for a differentiation of the signal and normalization modes:

$$B_{sig} \to D^{(*)} \tau \overline{\nu}_\tau, \quad \tau \to \ell \bar{\nu}_\ell \nu_\tau, \quad D^{(*)} \to \ldots \tag{5.4}$$

Thus, concerning the reconstruction of the B_{sig} meson in signal decay modes, the same approach as for the normalization modes is used, as the visible signature produced by the stable particles in the detector is identical: One (excited) D meson and a light lepton $\ell = e, \mu$.
For the reconstruction of the modes involving higher D meson excitations $D^{**} \in (D_0^*, D_1, D_1', D_2^*)$ (Decay (5.3)) the respective D^{**} mesons are not reconstructed explicitly. Instead, an additional charged or neutral π is added directly to the B_{sig} candidate as illustrated in Decay (5.5).

$$B_{sig} \to \overbrace{D^{(*)} \pi}^{D^{**}} \ell \bar{\nu}_\ell, \quad D^{(*)} \to \ldots \tag{5.5}$$

Despite the D^{**} resonances not being reconstructed explicitly, they can be identified and studied by evaluating the invariant mass of the $D^{(*)}\pi$ pair. Cases of D^{**} decays involving multiple pions are not considered.
The remainder of this section elaborates in detail on the reconstruction steps performed for the recombination of the signal side B meson for the n-tuple production.

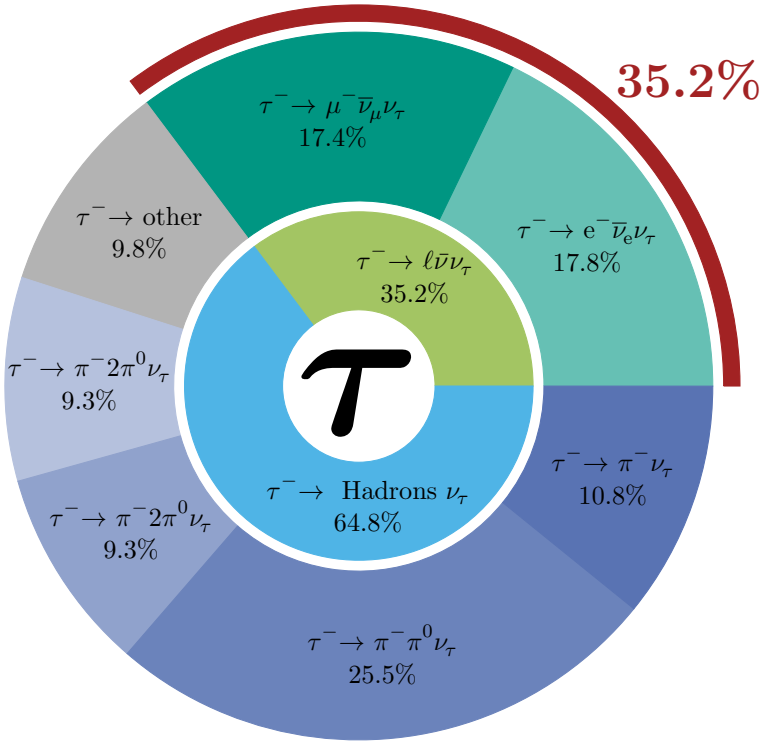


Figure 5.3.: Visualization of the branching fractions of the most common τ decay processes. The green inner circle represents the branching fraction covered by the leptonic τ decays, which is made up by almost equal parts of the exclusive leptonic decay processes into a muon and an electron shown in the outer circle. The sum of the branching fractions of hadronic and other τ decay modes is illustrated by the blue inner circle and the individual fractions of exclusive hadronic decay processes are shown in segments of the outer blue circle. The red arc emphasizes the leptonic decay processes considered in this analysis.

5.1.1. Final State Particles

At the beginning of the analysis reconstruction chain stand the so-called *final state particles*. The collection of particles summarized under this term comprises

charged stable particles detected as tracks in the Belle central drift chamber (CDC): $\mathbf{e^\pm}$, $\boldsymbol{\mu^\pm}$, $\boldsymbol{\pi^\pm}$ **and** $\mathbf{K^\pm}$; as well as

neutral clusters found by the Belle electromagnetic calorimeter (ECL) or the K_L^0 and muon detection system (KLM) without an associated track: $\boldsymbol{\gamma}$ or K_L^0, respectively.

Furthermore, the following short-lived, light particles are provided as pre-recombined candidates by the Belle reconstruction algorithms:

V0-particles whose displaced decay vertices allow for their recombination from charged tracks: $\mathbf{K_S^0}$ **and** $\boldsymbol{\gamma_{V0}}$; and

neutral pions $\boldsymbol{\pi^0}$ recombined from two photon candidates.

Although the last two categories are, strictly speaking, not final state particles, they are treated as such in the context of the signal side reconstruction of this analysis and are therefore included here. Particle candidates for all the above listed categories are provided directly from the data read from Belle `mdst` files and converted by the `b2bii` package. Further particle type dependent selection criteria are imposed on the final state particle candidates to improve the signal-to-noise ratio for each type and thereby reduce the combinatorics to be considered in further steps of the event reconstruction process.
Not relevant to the signal side reconstruction are K_L^0 mesons built from the neutral clusters provided by the Belle KLM detector.

Charged Particles
Charged particle candidates are generated from fitted helical tracks provided by the Belle track finding algorithm *Trasan*[2] and under the respective mass hypothesis. For all charged final state particle candidates the same minimal requirements on the particle's transverse momentum

$$p_t > 0.1\,\text{GeV} \tag{5.6}$$

and on the candidate's track impact parameters

$$|\Delta r| < 2\,\text{cm} \quad \text{and} \quad |\Delta z| < 4\,\text{cm} \tag{5.7}$$

have to be fulfilled. Here, Δr is the distance in the x-y plane of the point of closest approach (POCA) of the track's trajectory to the interaction point (IP) and Δz is the z coordinate of the POCA z_0 with respect to the IP. The former is often also referred to as d_0 in the parametrization of the helix trajectory of a track.
For MC track candidates, a track smearing procedure provided in form of the Belle analysis software function `smear_trk` [61] is employed as part of the `b2bii` data conversion before the selection criteria are applied. This method smears the helix parameters of the tracks in simulated events to improve the agreement to tracks of actual data in the helix parameter resolution. The uncertainty matrix of said parameters is scaled accordingly.

Electron e candidates are selected by requiring values of

$$e_{ID} > 0.6 \quad \text{and} \quad p_e > 0.3\,\text{GeV} \tag{5.8}$$

for the electron particle identification e_{ID} and the particle momentum in the laboratory reference frame, respectively. Due to its low mass the electron is subject to bremsstrahlung which affect the particle's momentum in magnitude and direction as photons are emitted when the electron's trajectory is curved in the magnetic field of the Belle solenoid. This effect is corrected for by a bremsstrahlung correction which collects the emitted bremsstrahlung photons. The details of this algorithm are elaborated on in Section 5.1.2, which is dedicated to this topic.

Muon μ candidates must fulfill the more stringent selection criteria

$$\mu_{ID} > 0.9 \quad \text{and} \quad p_\mu > 0.6\,\text{GeV} \tag{5.9}$$

[2] No official documentation of the Belle track finding algorithm *Trasan* is available, but they are mentioned in several Belle II documents such as [59] and [60].

on the muon particle identification μ_{ID} and their momenta in the laboratory reference frame, as muons must reach the outermost of Belle's sub-detectors — the KLM — to be distinguished from other particle types. The additional selection criterion `muBelleQuality` $= 1$ ensures the presence of KLM hits for a given muon candidate.

For both, electron and muon candidates, a loose selection requirement of $\mathrm{e_{ID}} > 0.1$ and $\mu_{\mathrm{ID}} > 0.1$, respectively, is applied first, to allow for an evaluation of the selection.

Charged Pions $\boldsymbol{\pi^{\pm}}$ are initially only required to pass a loose selection requirement on the pion particle identification of

$$\pi_{\mathrm{ID}} > 0.1 \tag{5.10}$$

in addition to the general requirements on charged particles stated in Equations (5.6) and (5.7). Further, more stringent selection criteria might be required when the $\pi^{\pm}$ candidates are recombined with other particle candidates to form intermediate particle candidates of higher masses.

Charged Kaon $\mathbf{K^{\pm}}$ candidates are selected similarly to $\pi^{\pm}$ candidates by applying only a loose selection requirement of

$$\mathrm{K_{ID}} > 0.1 \tag{5.11}$$

on the kaon particle identification $\mathrm{K_{ID}}$ initially, allowing for channel dependent tighter selection criteria in further reconstruction steps.

Neutral Particles

Photon $\boldsymbol{\gamma}$ candidates are built from ECL clusters with no associated track or by recombining an electron-positron pair with a common displaced vertex. For photon candidates with a related ECL cluster, polar-angle region dependent selection criteria are applied on the cluster energy. These criteria are summarized in the boolean observable `goodBelleGamma` which is defined as

$$\texttt{goodBelleGamma} = \begin{cases} 1, & \text{if } \mathrm{E}_\gamma > 50\,\mathrm{MeV} \quad \text{and} \quad 32.2^\circ < \theta_\gamma < 128.7^\circ \quad \text{Barrel} \\ 1, & \text{if } \mathrm{E}_\gamma > 100\,\mathrm{MeV} \quad \text{and} \quad 12.4^\circ < \theta_\gamma < 31.4^\circ \quad \text{Forward Endcap} \\ 1, & \text{if } \mathrm{E}_\gamma > 150\,\mathrm{MeV} \quad \text{and} \quad 130.7^\circ < \theta_\gamma < 155.1^\circ \quad \text{Backward Endcap} \\ 0, & \text{otherwise} \end{cases} \tag{5.12}$$

depending on whether the ECL cluster related to the photon candidate is found in the barrel region, the forward endcap region or the backward endcap region of the Belle ECL. Photon candidates found via the second scenario are recombined from tracks as V0-objects by assuming the $\mathrm{e^-e^+}$–pair to originate from a photon decaying via pair production. However, photon candidates reconstructed as V0-objects are not considered further for this and the other above-mentioned analyses.

Neutral $\mathbf{K^0_S}$ candidates are reconstructed from oppositely charged pion tracks with a common displaced vertex. This reconstruction channel covers about 69.2 % of the $\mathrm{K^0_S}$ decays[28]. Pre-recombined $\mathrm{K^0_S}$ candidates are provided by Belle's `v0finder` algorithm [62] and made available for use within this analysis with the help of the `b2bii` conversion package. A standard selection criterion on the NeuroBayes [63] based quality estimator `ksnbStandard` for the $\mathrm{K^0_S}$ candidates is applied to reduce combinatorial background, candidates from fake

or curling tracks, as well as other V0-particles such as Λ baryons [64]. In addition to this common selection criterion, further requirements on the mass of the K_S^0 candidate

$$|\Delta M| < 100\,\text{MeV} \quad \text{and} \quad \frac{|M_{\text{inv}} - M_{\text{nom}}|}{\sigma_{M_{\text{inv}}}} < 3.0 \tag{5.13}$$

must be fulfilled. Here, $|\Delta M|$ refers to the absolute value of the difference between the invariant mass M_{inv} of the recombined particle candidate and the nominal mass M_{nom} of the particle. By dividing this difference by the uncertainty of the invariant mass $\sigma_{M_{\text{inv}}}$, one obtains the significance of the deviation from the nominal mass value in units of Gaussian standard deviations. The uncertainty on the reconstructed mass M_{inv} is calculated from the track parameters of the daughter particles.

Neutral Pion π^0 candidates are recombined from two photons with a combined invariant mass roughly matching the nominal π^0 mass $M_{\pi^0} = 134.98\,\text{MeV}$ [28]. This requirement is enforced by the two criteria

$$104\,\text{MeV} < M_{\text{inv}} < 165\,\text{MeV} \quad \text{and} \quad \frac{|M_{\text{inv}} - M_{\text{nom}}|}{\sigma_{M_{\text{inv}}}} < 3.0 \tag{5.14}$$

on the invariant mass of the photon pair and the significance of the deviation from the nominal π^0 mass. Moreover, both of the photons must pass the `goodBelleGamma` requirement defined in Equation (5.12). As the neutral pions decay quickly after they are formed ($\tau_{\pi^0} = 8.43 \times 10^{-17}\,\text{s}$ [28]) close to the IP and mainly into two photons ($\mathcal{B}(\pi^0 \to \gamma\gamma) = 98.82\,\%$ [28]), the quality and efficiency of the π^0 reconstructed is limited by the performance of the ECL. The ECL cannot provide directional information and, hence, the invariant mass of the photon pair is the sole possibility to impose selection requirements on neutral pion candidates. An additional limitation on the efficiency of the π^0 reconstruction is due to photons escaping the ECL undetected in gaps or outside the acceptance region, as well as the ECL clusters of the two photons being inseparable and detected as one cluster.
The selection criteria for each individual type of final state particle are summarized in Table 5.1.
Most final state particles are subject to Monte Carlo corrections which address the issue of efficiency differences in the MC simulation and reconstruction compared to the performance on recorded data. These corrections are discussed in Chapter 4.
Particles with higher masses are recombined from the above listed final state particles, thus forming candidates for the intermediate D mesons or their excited variants, and finally signal B_{sig} meson candidates.

5.1.2. Bremsstrahlung Correction

Electrons (and positrons) are subject to bremsstrahlung while they traverse through the magnetic field of the Belle detector. The electrons can experience a significant energy loss when radiating a bremsstrahlung photon γ_{Brems}, resulting in a kink in the trajectory. The possibility for the emission of a bremsstrahlung photon γ_{Brems} is higher in dense matter, such as the beam pipe and the material of Belle's innermost tracking detector, the silicon vertex detector (SVD).
The negative impact of electrons affected by bremsstrahlung on physics analyses is twofold:

- The momentum of the electron is measured without consideration of the energy loss due to the emitted γ_{Brems}. This is propagated to the properties of particles reconstructed from this electron.

- A γ_{Brems} will produce an additional ECL cluster which cannot be correctly associated to the event, if the bremsstrahlung is not taken into account.

Hence, it is worthwhile to identify bremsstrahlung photons γ_{Brems} and combine them with the respective electron candidate. This process is called bremsstrahlung correction and executed by a dedicated algorithm. The algorithm is optimized based on the energy residual

$$\Delta \text{E} = \text{E}_{\text{corr}} - \text{E}_{\text{truth}} \tag{5.15}$$

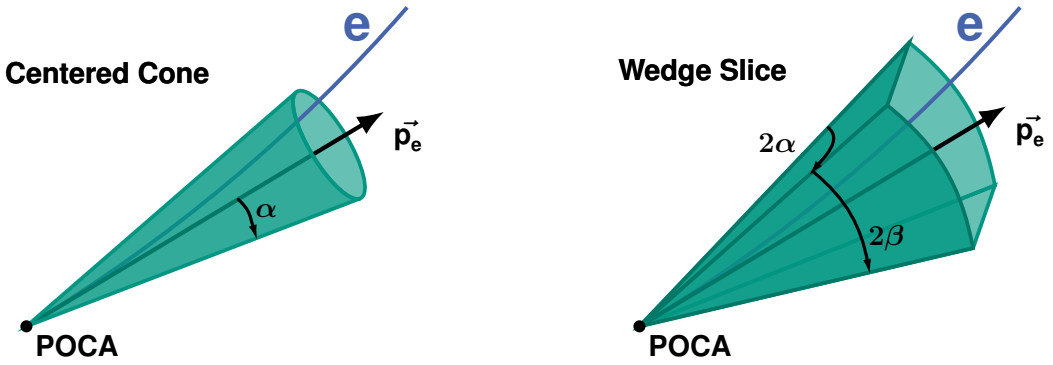
where E_{truth} is the true energy of the electron before the emission of a bremsstrahlung photon and E_{corr} is the energy of the energy of the reconstructed electron after the bremsstrahlung correction is applied. This quantity shall be minimal for electrons which actually experienced bremsstrahlung and for which a correction is necessary, as well as for electrons which were not affected by bremsstrahlung.
For each e candidate, γ_{Brems} candidates are collected in a cone around the electron momentum direction $\vec{\text{p}}_{\text{e}}$ at the POCA as illustrated in Figure 5.4a. The optimal opening angle α of this search cone was found to be 2°. Photons in this cone must fulfill requirements on their energy $\text{E}_{\gamma_{\text{Brems}}}$ and momentum magnitude $|\vec{\text{p}}_{\gamma_{\text{Brems}}}|$ for which the optima were found to be

$$\text{E}_{\gamma_{\text{Brems}}} < 400\,\text{MeV} \qquad \text{and} \qquad |\vec{\text{p}}_{\gamma_{\text{Brems}}}| < 0.4 \cdot |\vec{\text{p}}_{\text{e}}|, \tag{5.16}$$

where the $\text{E}_{\gamma_{\text{Brems}}}$ requirement is defined by a linear dependency relative to the momentum of the reconstructed electron candidate.

Table 5.1.: List of all final state particles and their corresponding selection criteria. The values in parentheses for the p_{ID} requirements are the values used during the initial selection of the candidates.

	Final State Particle	Selection Criteria
Tracks	all	$\lvert\Delta r\rvert < 2\,\text{cm}$ $\lvert\Delta z\rvert < 4\,\text{cm}$ $\text{p}_t > 0.1\,\text{GeV}$
	$\text{e}^{\pm}$	$\text{p}_{\text{e}} > 300\,\text{MeV}$ $\text{e}_{\text{ID}} > 0.6 \quad (0.1)$
	$\mu^{\pm}$	$\text{p}_{\mu} > 600\,\text{MeV}$ $\mu_{\text{ID}} > 0.9 \quad (0.1)$ `muBelleQuality` $= 1$
	$\pi^{\pm}$	$\pi_{\text{ID}} > 0.1$
	$\text{K}^{\pm}$	$\text{K}_{\text{ID}} > 0.1$
Neutrals	γ	`goodBelleGamma` $= 1$
	K^0_{S}	$\lvert\Delta\text{M}\rvert < 100\,\text{MeV}$ $\lvert(\text{M}_{\text{inv}} - \text{M}_{\text{nom}})/\sigma_{\text{M}_{\text{inv}}}\rvert < 3.0$ `ksnbStandard` $= 1$
	π^0	$104\,\text{MeV} < \text{M}_{\text{inv}} < 165\,\text{MeV}$ $\lvert(\text{M}_{\text{inv}} - \text{M}_{\text{nom}})/\sigma_{\text{M}_{\text{inv}}}\rvert < 3.0$ `goodBelleGamma` $= 1$ (for all γ)



(a) (b)

Figure 5.4.: Illustration of the search region for bremsstrahlung photons γ_{Brems} in a cone (a) and a wedge-slice shaped search region (b) around the electron momentum vector at the point of closest approach. In the case of the cone, its opening angle α can be optimized, while the wedge-slice search region has two angles α and β which can be optimized.

All γ_{Brems} candidates matching these conditions are combined with the electron candidate to build a corrected electron candidate e_{corr} with an updated four-momentum

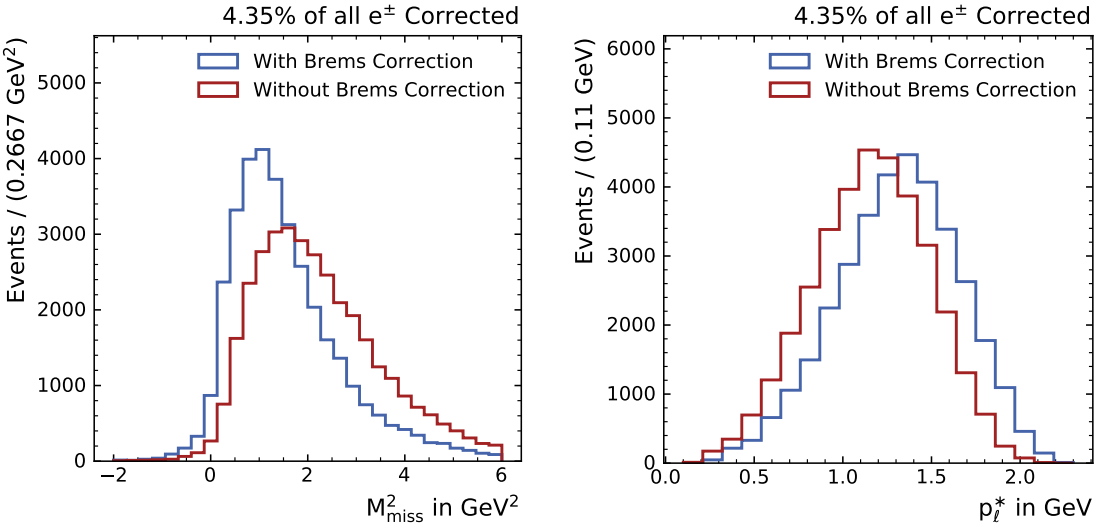
$$\text{p}_{\text{e}_{\text{corr}}} = \text{p}_{\text{e}_{\text{reco}}} + \sum_i \text{p}_{\gamma^i_{\text{Brems}}}. \tag{5.17}$$

This bremsstrahlung corrected candidate e_{corr} competes with uncorrected electrons to be used for the recombination of intermediate particles of higher masses. Thus, if any of the photon candidates used for the correction, or the original uncorrected electron itself is used for a more successful interpretation of the event, the bremsstrahlung corrected candidate is also rejected by design.

The impact of the bremsstrahlung correction can be evaluated for fully reconstructed $\Upsilon(4S)$ events and in particular regarding its effect on the squared missing mass M^2_{miss} defined in Equation (1.2). Of all events reconstructed in the channels relevant to and chosen by the event selection for the $\mathcal{R}(\text{D}^{(*)})$ measurement described in Section 6.3, 4.35 % are bremsstrahlung corrected. The change in the M^2_{miss} distribution due to the bremsstrahlung correction is shown in Figure 5.5. The resolution of the peak in the squared missing mass at 0 GeV is improved by the application of the bremsstrahlung correction. The correction also affects the fit observable p^*_ℓ, as the momentum of the signal side light lepton is increased, if a bremsstrahlung photon is found.

As the momentum direction of the electrons at the POCA is used as the center of the cone in which bremsstrahlung photons are searched for, the approach might not be optimal for low momentum electrons with a high curvature. Other definitions of the search region have been tested. Different orientations of the cone with respect to the electron momentum vector at the POCA were explored. Instead of a cone, a wedge-slice shaped search region Figure 5.4b given by two angles defined with respect to the momentum vector, one spanning an arc in the plane given by the curved track and one perpendicular to this plane, was also tested. For all approaches, the parameters are optimized according to the same criteria as explained above. The best results are obtained with the centered cone method.

For future analyses of Belle II data, a more sophisticated method [65] is available. This new method can make use of the full tracking detector information accessible during the track



(a) (b)

Figure 5.5.: Visualization of the effect of the bremsstrahlung correction on the distribution of the squared missing mass M^2_{miss} on the left (a) and the lepton momentum in the B_{sig} rest frame p^*_ℓ on the right (b) in the selection for the $\mathcal{R}(D^{(*)})$ measurement. Only events involving a bremsstrahlung corrected electron are considered.

reconstruction, while the method described above only utilizes the fitted helix parameter of the track and the ECL clusters.

5.1.3. $D^{(*)}$ Meson Reconstruction

The first part of the signal B_{sig} meson decay tree is now already reconstructed in the form of the electron and muon final state particles. The second reconstructable decay product of the signal B_{sig} meson decay is an unexcited or excited $D^{(*)}$ meson (see Decays (5.1) to (5.4)). As excited D^* mesons always decay into unexcited ones, the unexcited D meson has to be reconstructed for every event.

The D meson can decay via a multitude of different decay channels, which also depend on whether the D meson is charged or neutral. To achieve a high reconstruction efficiency on the signal side, the D mesons are recombined in the eight decay channels with the highest branching fractions for the charged and neutral case, each. The 16 D reconstruction channels and the branching fraction they cover are listed in Table 5.2.
Candidates for D mesons are built from combinations of the final state particles described above. The selection criteria imposed on D meson candidates are reconstruction channel dependent and are included in Table 5.2. To reject obvious background from continuum events, all D meson candidates are required to have a momentum below

$$p_D^{\text{CMS}} < 3.0\,\text{GeV} \tag{5.18}$$

in the center-of-mass reference frame. The invariant mass of a D meson candidate provides separation power to reduce combinatorial background. Again, the difference ΔM of the reconstructed to the nominal mass, as well as the significance of it, are utilized. For D

meson reconstruction channels involving only charged daughter particles, the selection based on the invariant mass is conducted via the significance of the ΔM with respect to the uncertainty on the invariant mass given by the track parameters of the daughter particles, by requiring

$$\frac{|M_{inv} - M_{nom}|}{\sigma_{M_{inv}}} < 4.0 \tag{5.19}$$

for all such reconstruction channels, except for the channel $D^+ \to K^+K^-\pi^+$.

Table 5.2.: Overview of the 16 charged and neutral D meson reconstruction modes considered for the signal side B_{sig} meson reconstruction, as well as their corresponding selection criteria. The selection criterion imposed on the D meson's momentum magnitude $p_D^{CMS} < 3.0\,\text{GeV}$ is not listed explicitly, as it applies for every mode. The column marked $\mathcal{B}$ lists the branching fractions of the respective channel in percent.

	D Reconstruction Channel	Selection Criteria	$\mathcal{B}$ in %
Charged $\sum \mathcal{B} \approx 29.29\,\%$	$D^+ \to K^-\pi^+\pi^+$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leqq 4.0$	9.46
	$D^+ \to K^-\pi^+\pi^+\pi^0$	$\lvert\Delta M\rvert < 40\,\text{MeV}$ $p_{\pi^0}^{CMS} > 200\,\text{MeV}$	6.14
	$D^+ \to K^-\pi^+\pi^+\pi^+\pi^-$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leqq 4.0$	0.58
	$D^+ \to K_S^0\pi^+$	$\lvert\Delta M\rvert < 40\,\text{MeV}$	1.53
	$D^+ \to K_S^0\pi^+\pi^0$	$\lvert\Delta M\rvert < 40\,\text{MeV}$ $p_{\pi^0}^{CMS} > 200\,\text{MeV}$	7.24
	$D^+ \to K_S^0\pi^+\pi^+\pi^-$	$\lvert\Delta M\rvert < 40\,\text{MeV}$	3.05
	$D^+ \to K_S^0K^+$	$\lvert\Delta M\rvert < 40\,\text{MeV}$	0.30
	$D^+ \to K^+K^-\pi^+$	$\lvert\Delta M\rvert < 40\,\text{MeV}$	1.00
Neutral $\sum \mathcal{B} \approx 40.15\,\%$	$D^0 \to K^-\pi^+$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leqq 4.0$	3.93
	$D^0 \to K^-\pi^+\pi^0$	$\lvert\Delta M\rvert < 50\,\text{MeV}$ $p_{\pi^0}^{CMS} > 200\,\text{MeV}$	14.3
	$D^0 \to K^-\pi^+\pi^+\pi^-$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leqq 4.0$	8.07
	$D^0 \to K^-\pi^+\pi^+\pi^-\pi^0$	$\lvert\Delta M\rvert < 40\,\text{MeV}$	4.20
	$D^0 \to K_S^0\pi^0$	$\lvert\Delta M\rvert < 50\,\text{MeV}$ $p_{\pi^0}^{CMS} > 200\,\text{MeV}$	1.20
	$D^0 \to K_S^0\pi^+\pi^-$	$\lvert\Delta M\rvert < 40\,\text{MeV}$	2.85
	$D^0 \to K_S^0\pi^+\pi^-\pi^0$	$\lvert\Delta M\rvert < 40\,\text{MeV}$ $p_{\pi^0}^{CMS} > 200\,\text{MeV}$	5.20
	$D^0 \to K^-K^+$	$\lvert(M_{inv} - M_{nom})/\sigma_{M_{inv}}\rvert \leqq 4.0$	0.40

For all D reconstruction channels involving neutral daughter particles, as well as for the mode $D^+ \to K^+K^-\pi^+$, the selection criterion is based on the absolute value of the difference between the reconstructed and the nominal mass ΔM, on which a reconstruction mode–dependent upper bound of

$$|\Delta M| < 40\,\text{MeV} \quad \text{or} \quad |\Delta M| < 50\,\text{MeV} \tag{5.20}$$

is applied (see Table 5.2). If a neutral pion π^0 is among the daughter particles, an additional

Table 5.3.: Overview of the four considered D^* reconstruction modes and the respective selection criteria. The common criteria on the momentum magnitude in the center-of-mass reference frame is omitted.

D^* Reconstruction Channel	Selection Criteria
$D^{*+} \to D^0\pi^+$	$\Delta M(D, D^*) < 155\,\mathrm{MeV}$ $p^{CMS}_{\pi^\pm} < 400\,\mathrm{MeV}$
$D^{*+} \to D^+\pi^0$	$\Delta M(D, D^*) < 155\,\mathrm{MeV}$
$D^{*0} \to D^0\pi^0$	$\Delta M(D, D^*) < 160\,\mathrm{MeV}$
$D^{*0} \to D^0\gamma$	$130\,\mathrm{MeV} < \Delta M(D, D^*) < 155\,\mathrm{MeV}$

selection requirement on its momentum magnitude in the center-of-mass reference frame

$$p^{CMS}_{\pi^0} > 200\,\mathrm{MeV} \tag{5.21}$$

must be fulfilled for all D meson reconstruction channels, except the neutral D^0 meson channel involving four pions.
To reduce the combinatorics due to the number of D meson candidates to be considered in further reconstruction steps, a best candidate selection is applied. For this purpose, the candidates are ranked according to the ascending absolute value of the difference between the invariant and nominal mass $|\Delta M|$. Only the ten candidates with the lowest $|\Delta M|$ values per event are kept.

The D^* meson offers a much smaller variety of possible decay channels. Neutral D^{*0} mesons decay into a neutral D^0 meson and either a neutral pion π^0 or a photon γ. Charged $D^{*\pm}$ meson decays result mainly in a $D^0\pi^+$ pair, producing a neutral D meson, or a $D^+\pi^0$ pair. Less likely is the charged $D^{*\pm}$ decay mode $D^{*+} \to D^+\gamma$ with a branching fraction of 1.6 % [28]. By taking all but the last decay channels for the reconstruction into account, the D^* meson reconstruction is efficient in this regard. An overview of the four reconstruction modes — two for the charged and two for the neutral case — is given in Table 5.3.
The most powerful selection criterion for D^* candidates is based on the difference of the invariant mass of the reconstructed D^* meson and the daughter D meson:

$$\Delta M(D, D^*) = M_{D^*} - M_D. \tag{5.22}$$

Different requirements are imposed on this quantity, depending on the underlying reconstruction mode as listed in Table 5.3. For the reconstruction mode $D^{*+} \to D^0\pi^+$ an additional selection requirement on the momentum of the charged $\pi^\pm$ in the center-of-mass reference frame

$$p^{CMS}_{\pi^\pm} < 400\,\mathrm{MeV} \tag{5.23}$$

is applied. This criterion is motivated by a physical upper bound in the momentum due to phase space restrictions on the light D^* daughter and to suppresses continuum background. Lastly, the momentum magnitude of all D^* meson candidates is required to be below

$$p^{CMS}_{D} < 3.0\,\mathrm{GeV} \tag{5.24}$$

in the center-of-mass reference frame, similar to the D mesons.

5.1.4. D^{**} Meson Reconstruction

For the semi-leptonic B decays modes into the orbitally excited D mesons, a dedicated reconstruction approach is undertaken. These $1P$ states of the D mesons are the D_1, D_2^*, D_1', and D_0^*, which are commonly referred to as D^{**} mesons (see Section 2.5). All four states usually decay into an excited or unexcited $D^{(*)}$ meson and a single neutral or charged pion. Other decay modes involving multiple pions, a photon γ or an η are also possible, depending on the D^{**} state. However, for the purpose of the herein presented analysis, only the decays involving one charged or neutral pion are considered.
These background decay processes into the D^{**} mesons are reconstructed explicitly for the dedicated analyses of the $B \to D^{**}\ell\nu$ processes, as well as to allow for an evaluation of these processes as background in context of the herein presented $\mathcal{R}(D^{(*)})$ study.
As illustrated by Decay (5.5), the D^{**} mesons are not reconstructed explicitly. Instead, the additional charged or neutral pion is added as a daughter directly to the signal B_{sig} meson. Moreover, no differentiation between the four individual D^{**} states is made. Thus, all D^{**} mesons are treated equally during the reconstruction. The invariant mass of the $D^{(*)}\pi$ pair $M_{inv}(D^{(*)}, \pi)$ can be evaluated post-reconstruction, to allow for an identification of the different D^{**} states in its distribution. The different mass values, as well as mass widths of the two narrow states D_1 and D_2^* and the two wide states D_0^* and D_1' (see Table 2.2) allow for a separation of the states in the observable $M_{inv}(D^{(*)}, \pi)$. At this stage of the reconstruction, however, the separation is insufficient for background suppression, due to the width of the wide states.
For the B_{sig} meson reconstruction in these modes, no additional requirements due to the involvement of the D^{**} meson are applied at this stage, except for a lower bound requirement on the momentum of neutral pions in the center-of-mass reference frame,

$$p_{\pi^0}^{CMS} > 200\,\text{MeV} \tag{5.25}$$

if relevant to the reconstruction mode (see Table 5.4). The details of the B_{sig} meson reconstruction are provided in the next section.

5.1.5. Signal B Meson Reconstruction

The signal side B_{sig} meson candidates are built by combining a $D^{(*)}$ meson and light lepton ℓ candidate. As the light lepton can either be an electron or a muon ($\ell = e, \mu$), it allows for two different reconstruction modes. For the $D^{(*)}$ meson, the unexcited and excited variant are considered, both of which can be charged or neutral, resulting in four different reconstruction modes from the options available for this B_{sig} meson daughter. This gives a total of eight different reconstruction modes, without taking the reconstruction modes dedicated to the D^{**} mesons into account. The separation of reconstruction modes according to the lepton, however, is not made use of by this analysis. This leaves the four reconstruction modes listed in the left column of Table 5.4 for the processes illustrated in Decays (5.1), (5.2) and (5.4), which are relevant to the measurement of the ratios $\mathcal{R}(D^{(*)})$.
For the reconstruction modes dedicated to the D^{**} states illustrated in Decay (5.3), a charged or neutral pion candidate is added to the combination of $D^{(*)}$ meson and light lepton ℓ. Not considering the options for the flavor of the light lepton, this yields eight additional reconstruction modes for the analysis of the B_{sig} decays into D^{**} mesons. These are listed in the middle and right column of Table 5.4 for neutral and charged pions, respectively.
The reconstruction modes considering the D^{**} excitations via the added pions are of use to the $\mathcal{R}(D^{(*)})$ measurements, as candidates for these modes can reject worse-matching

Table 5.4.: Overview of all B_{sig} reconstruction channels, split into the channels involving only an unexcited or exited $D^{(*)}$ meson and a lepton in the left column and the channels including an additional neutral (middle column) or charged pion (right column). For all cases, ℓ indicates one of the two light leptons: $\ell = e, \mu$. The last rows list the selection criteria for each of the reconstruction modes.

$D^{(*)}$ Channels	$D^{(*)}\pi^0$ Channels	$D^{(*)}\pi^\pm$ Channels
$B^- \to D^0 \ell^-$	$B^- \to D^0 \pi^0 \ell^-$	$B^- \to D^+ \pi^- \ell^-$
$B^- \to D^{*0} \ell^-$	$B^- \to D^{*0} \pi^0 \ell^-$	$B^- \to D^{*+} \pi^- \ell^-$
$B^0 \to D^- \ell^+$	$B^0 \to D^- \pi^0 \ell^+$	$B^0 \to \overline{D}^0 \pi^- \ell^+$
$B^0 \to D^{*-} \ell^+$	$B^0 \to D^{*-} \pi^0 \ell^+$	$B^0 \to \overline{D}^{*0} \pi^- \ell^+$
$1\,\text{GeV} < M_{inv} < 6\,\text{GeV}$	$1\,\text{GeV} < M_{inv} < 6\,\text{GeV}$	$1\,\text{GeV} < M_{inv} < 6\,\text{GeV}$
$R_2 < 0.6$	$R_2 < 0.6$	$R_2 < 0.6$
	$p_{\pi^0}^{CMS} > 200\,\text{MeV}$	

candidates reconstructed in the D and D^* channels for the event. Furthermore, the D^{**} reconstruction modes can be used in a simultaneous or speparate fit of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ to constrain the corresponding D^{**} contribution. Likewise, the D and D^* reconstruction channels are used in the D^{**} analysis to constrain the respective components in fits and as normalization mode in branching fraction measurements.

The selection requirements imposed on B_{sig} meson candidates at this stage are rather loose. The reconstructed invariant mass of the B_{sig} meson candidate is required to lie within the mass window given by

$$M_{inv}^{B_{sig}} \in [1.0, 6.0]\,\text{GeV}, \tag{5.26}$$

generously enclosing the nominal mass of both the neutral B^0 and charged $B^\pm$ meson of $M_{B^0} = 5279.7\,\text{MeV}$ and $M_{B^\pm} = 5279.3\,\text{MeV}$, respectively [28]. This loose selection requirement is used to reject obvious combinatorial background early on, without dismissing candidates with lower invariant masses due to the invisible neutrino(s).

Continuum background is suppressed by requiring the ratio of the second order Fox-Wolfram moment [66] with respect to the moment of order zero, called R_2 and described to more detail in Section 6.1, to be below

$$R_2 < 0.6 \tag{5.27}$$

for the signal side B meson.

For the B_{sig} reconstruction modes involving a neutral pion, an additional selection criterion requiring the π^0 momentum magnitude to be above

$$p_{\pi^0}^{CMS} > 200\,\text{MeV} \tag{5.28}$$

in the center-of-mass reference frame is imposed. An overview of all B_{sig} meson reconstruction requirements are given in Table 5.4.

In the final step of the signal side B meson reconstruction, a vertex fit is applied on the obtained B_{sig} candidates. For this, the `TreeFitter` global decay chain vertex fit implementation of the Belle II Analysis Software Framework is used [67]. This tool performs vertex fits on the entire reconstructed decay tree of the B_{sig} candidate. To allow for a stable fit, mass constraints on the $D^{(*)}$ meson candidates in the decay tree are applied. The fit result is **not** used to update the momenta or invariant masses of the reconstructed particles,

but only provides an additional quality indicator for the particle candidate in form of the p-Value $\mathcal{P}_{\text{Vertex}}^{\text{B}_{\text{sig}}}$ of the vertex fit. This property is later used to refine the signal-to-noise ratio of the candidate selection.

5.2. Tag Side Reconstruction

Once a valid B_{sig} candidate is found for an event, the reconstruction of the second B meson accompanying it is necessary to be able to recombine both B mesons to the $\Upsilon(4\text{S})$ of the event, as illustrated in Figure 5.1. This second B meson of an event is referred to as the tag side B meson, or B_{tag} meson. It is reconstructed in an automated manner by the Full Event Interpretation (FEI) algorithm [17]. This high-level analysis tool was developed as part of the Belle II Analysis Software Framework and, as such, can also be used for the analyses of Belle data with the help of the `b2bii` package. It is the successor of Belle's Full Reconstruction (FR) [68], which was used in the first hadronically tagged $\mathcal{R}(\text{D}^{(*)})$ measurement published by the Belle collaboration [12, 21].
The FEI algorithm approaches the reconstruction of the tag side B_{tag} meson by selecting and recombining particle candidates, starting with the reconstructed tracks and clusters from which final state particle candidates are created and building ever heavier particle candidates in further steps. Similar to the steps described for the signal side B reconstruction in Section 5.1, the hierarchy defined by the particle decays and particle masses is followed in reverse order. The level of automatization provided by the FEI tagging algorithm, however, allows for a far greater number of reconstruction channels for each intermediate particle and the final B_{tag} meson.
The hierarchical approach to the B_{tag} meson reconstruction is illustrated in Figure 5.6. It is split into a total of six stages:

1. selection of final state particles from detector information
2. recombination of J/ψ and π^0 candidates
3. selection of K_{S}^0 candidates from neutral pions and V0-objects
4. recombination of D and D_{s} candidates
5. recombination of excited D^* and D_{s}^* mesons
6. recombination of charged and neutral B meson candidates

where in the final stage a differentiation between hadronic and semi-leptonic reconstruction modes is made. For the analyses described in this text, **only the hadronic B_{tag} modes** are considered. That implies that no neutrino should be present in the decay of the tag side B_{tag} meson and all invisible energy can be attributed to the signal side B_{sig} meson, assuming the event is reconstructed correctly.
To determine the validity of each recombined particle candidate in each stage, a boosted decision tree (BDT) is trained for each reconstruction channel of each particle. These BDTs are trained to distinguish between correctly reconstructed candidates and background. Thus, their output can be interpreted as the probability of the particle candidate correctly representing the decay process of a given event. The inputs to the various BDT classifiers depend on the respective decay channel and include the BDT outputs of particle candidates of previous stages used in the current recombination. Finally, the BDT outputs of B_{tag} meson candidates represent the likelihood of the correctness of the tag side candidate. The

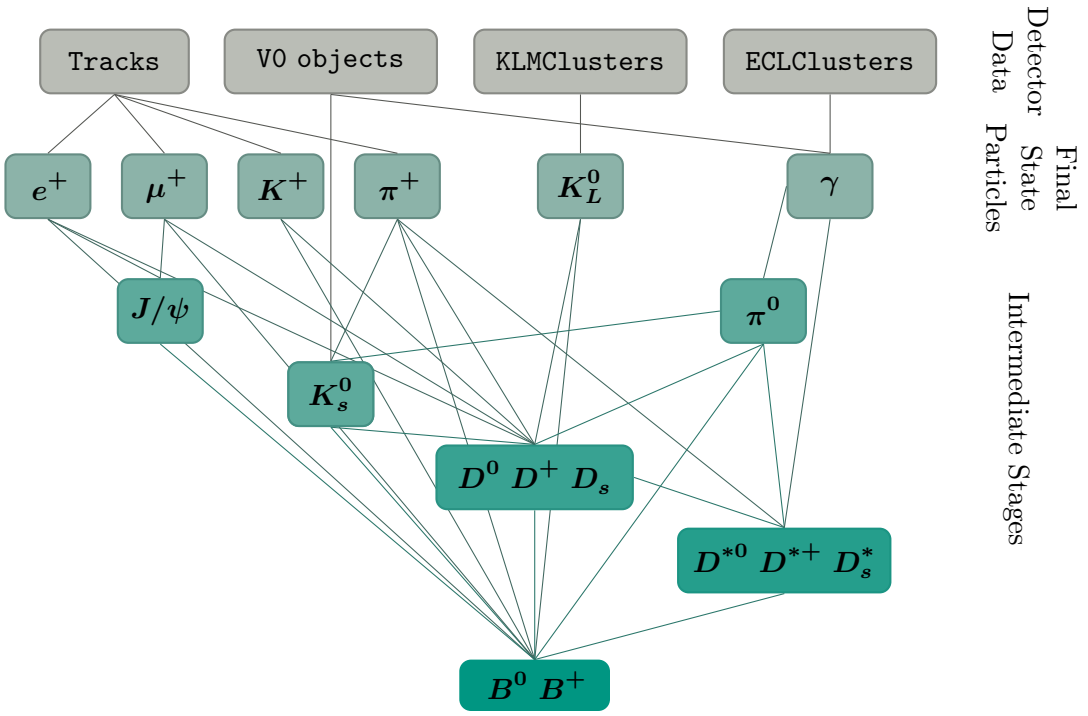


Figure 5.6.: Illustration of the hierarchical structure of the automatization of the B_{tag} meson reconstruction by the FEI algorithm. The different in- and outputs to the stages of the algorithm are displayed as nodes. Beginning with the detector level information at the top, from which final state particles are selected, candidates for ever more massive intermediate particles are recombined in the following stages, as indicated by the edges going from top to bottom. The output of the final stage is the tag B meson candidate shown at the bottom. Adapted from [17].

output of all BDTs of the last stage is referred to as FEI *probability* $\mathcal{P}_{\text{FEI}}$, independent of the B_{tag} meson reconstruction mode. This observable is used in the analyses to quantify the quality of the B_{tag} meson candidate.
The FEI training used for the herein discussed analyses is based on version FEIv4 of the algorithm and has the identifier `FEIv4_2017_MCConverted_Track14_2`. Converted generic Belle MC from stream 0 is used to train the BDT classifiers of this FEI training.
By utilizing the FEI algorithm instead of its predecessor, the FR, an improvement in reconstruction efficiency for the tag side B_{tag} meson by a factor of two can be expected, according to the observations made in [17]. This improvement should also be reflected in an increase in reconstruction efficiency, and therefore a decrease in statistical uncertainty for the $\mathcal{R}(D^{(*)})$ measurements discussed in this text, as compared to the previous Belle result [12].

Tag B Meson Selection
The number of B_{tag} meson candidates provided by the FEI algorithm, as well as the signal-to-noise ratio of those candidates depend on the underlying event. To reduce the number of candidates to be considered, only the 20 best B_{tag} meson candidates ranked by their BDT score are selected.
Using selection requirements based on the properties of the B mesons, unphysical combi-

natorial background is rejected. As for most of the recombined particles described above which do not involve invisible daughter particles, the invariant mass is a good observable to base such a constraint on. For the tag side B_{tag} candidates, however, a more sophisticated approach is usually taken. By utilizing the knowledge about the kinematics of the e^+e^-–collisions and the clean decay of $\Upsilon(4S)$ into a pair of B mesons, additional constraints can be imposed on the calculation of the reconstructed mass of the B meson candidate. Since half of the collision energy can be associated to each of the B mesons of the back-to-back $\Upsilon(4S)$ decay in the center-of-mass reference frame, the energy of the B mesons is known. The definition of the so-called beam-constrained mass M_{bc} exploits this fact by using the beam energy of one beam in the center-of-mass reference frame instead of the reconstructed B meson candidate's energy:

$$M_{bc} = \sqrt{E_{beam}^2 - p_B^{CMS\,2}}. \tag{5.29}$$

Given a correctly recombined hadronic B meson decay, this observable peaks at the nominal B meson mass of $M_B \approx 5.28\,\text{GeV}$. For the B_{tag} meson, the respective momentum p_{tag}^{CMS} is used to define M_{bc}^{tag}. In the center-of-mass reference frame, the energy of each beam E_{beam} corresponds to half of the center-of-mass energy $E_{CMS} = \sqrt{s}$ and is therefore a constant value defined by the beam run time parameters of the accelerator to

$$E_{B_{tag}} = E_{beam} = \frac{1}{2} \cdot E_{CMS} = \frac{1}{2} \cdot \sqrt{s} \tag{5.30}$$

in the center-of-mass reference frame.
Using the same constraint, a deviation from the expected energy value and the reconstructed energy of the B_{tag} meson candidate can be defined as

$$\Delta E_B = E_B^{CMS} - E_{beam} \tag{5.31}$$

with the values taken again in the center-of-mass reference frame. For the specific case of the tag side B_{tag} meson this observable is called ΔE_{tag}, and is calculated using the respective tag side energy $E_{B_{tag}}^{CMS}$.
Based on these observables, two selection criteria must be fulfilled by all B_{tag} meson candidates. For the beam-constrained mass M_{bc}^{tag} of the tag candidate, a loose and tight selection is applied, with

$$M_{bc}^{tag} > 5.22\,\text{GeV} \quad \text{and} \quad M_{bc}^{tag} > 5.27\,\text{GeV}, \tag{5.32}$$

respectively. With the loose selection, a M_{bc}^{tag}-sideband is defined by the window $M_{bc}^{tag} \in [5.22, 5.27]\,\text{GeV}$ right below the value of the tight selection criterion. For the deviation from the expected energy in the center-of-mass reference frame ΔE_{tag} the candidates must lie within the asymmetric window defined by

$$-150\,\text{MeV} < \Delta E_{tag} < 100\,\text{MeV} \tag{5.33}$$

around the expected energy.

5.3. $\Upsilon(4S)$ Reconstruction

After the signal and tag side B meson candidates are gathered, candidates for $\Upsilon(4S)$ mesons can be recombined. This is attempted in each event for which at least one B candidate is available for each side.
The recombination is performed via the two common channels

- $\Upsilon(4S) \to B^+_{sig}\, B^-_{tag}$ + c.c. $(51.4 \pm 0.6)\,\%$ [28] and
- $\Upsilon(4S) \to B^0_{sig}\, \overline{B}^0_{tag}$ + c.c. $(48.6 \pm 0.6)\,\%$ [28],

which cover 100 % of the $\Upsilon(4S)$ decay branching fraction. For the case of a pair of uncharged B^0 mesons, the combinations B^0B^0 and $\overline{B}^0\overline{B}^0$ are considered as valid $\Upsilon(4S)$ candidates as well, to allow for the effect of B meson mixing. Additionally, unphysical combinations of B mesons to $\Upsilon(4S)$ via

- $\Upsilon(4S) \to B^+_{sig}\, B^0_{tag}$ + c.c.
- $\Upsilon(4S) \to B^-_{sig}\, B^0_{tag}$ + c.c.
- $\Upsilon(4S) \to B^0_{sig}\, B^+_{tag}$ + c.c.
- $\Upsilon(4S) \to B^0_{sig}\, B^-_{tag}$ + c.c.

are considered in the event reconstruction. These reconstruction modes are added to allow for the recovery of events in which low momentum charged particles were lost or assigned to the wrong B meson. The unphysical combinations lead to charged $\Upsilon(4S)$ meson candidates. Nonetheless, the properties of the signal side B_{sig} meson can be reflected correctly, making the event still viable for the analyses.
The thus found $\Upsilon(4S)$ meson candidates must not use any track or cluster twice. Furthermore, the invariant mass of the candidate must be within the range

$$M^{\Upsilon(4S)}_{inv} \in [7.0, 13.0]\,\text{GeV} \tag{5.34}$$

around the nominal $\Upsilon(4S)$ mass of 10.579 GeV [28]. This range allows for the invisible mass on the signal side.
Since the whole event should be reconstructed after the recombination of the $\Upsilon(4S)$ meson, all detector objects should in principle be assigned. Hence, the presence of unassigned tracks indicates a mis-reconstruction of the event. Such events are discarded by requiring that no additional, unassigned tracks fulfilling the basic selection criteria of charged particles as defined in Section 5.1 are present. This requirement is referred to as the *completeness constraint*. Track clones due to curling tracks are identified by a dedicated algorithm (described below) and excluded for this requirement. This avoids double counting of curling tracks which would otherwise cause well reconstructed events to be rejected by the completeness constraint only due to cloned tracks.

Curling Track Clone Finder
The algorithm used to determine whether a track candidate is a duplicate of another candidate exploits the similarities of the helix parameters of such candidates. The duplicates (or clones) usually stem from charged tracks with lower transversal momenta. In the magnetic field of the detector's solenoid, the trajectories of such low momentum particles will be bent strongly enough that they reenter the sensitive volume of the tracking detectors or never

leave it. Their helical path curls in the detector. Due to this, the fitted helix parameters of low momentum tracks can be compared to find and exclude duplicates. The upper limit on the transversal momenta p_t of charged particle candidates to be considered for this procedure is

$$\mathrm{p}_t < 275\,\mathrm{MeV}. \tag{5.35}$$

For two track candidates to be considered duplicates of each other, their reconstructed momenta must lie within 20 MeV of each other and the angle between their momentum vectors must not be larger than 5°. If pairs of track candidates fulfilling these criteria are found, the one which has no associated signature in the Belle SVD is marked as the clone. If both tracks do not feature SVD hits, the one with the larger impact parameters, evaluated via

$$\left(\frac{d_0}{\sigma_{d_0}}\right)^2 + \left(\frac{z_0}{\sigma_{z_0}}\right)^2, \tag{5.36}$$

where σ denotes the uncertainty on the respective helix parameter, is marked as the clone candidate.

6. Event Selection

Based on the set of $\Upsilon(4S)$ candidates reconstructed as described in Chapter 5, a reduced data set for the analysis of the ratios $\mathcal{R}(D^{(*)})$ discussed in this text has to be selected. This selection is tailored to improve the signal-to-noise ratio and reduce the number of candidates to one per surviving event. Similar selection criteria are applied for the data sets dedicated to the other analyses mentioned at the beginning of Chapter 5.

The first step in this selection process is independent of the specific analysis and aims to reduce the common continuum background. For this purpose a multivariate classifier is used. The preparation and validation of this continuum suppression classifier is elaborated on in Section 6.1. The general concept of the best candidate selection employed for the described analyses is introduced in Section 6.2. The subsequent Section 6.3 goes into the details of the selection process applied for the $\mathcal{R}(D^{(*)})$ measurement. In addition to the main selection for this measurement, the production of sideband samples for validation purposes is explained as well.

6.1. Continuum Suppression

To reduce the contribution of continuum background originating from $e^+e^- \to q\bar{q}$ processes — where $q = u, d, s, c$ — a dedicated multivariate classifier is utilized. For this purpose, the FastBDT boosted decision tree implementation [69] is used. The BDT is trained to distinguish the jet-like signature of $e^+e^- \to q\bar{q}$ continuum events from the spherical signature of $B\overline{B}$ events. These differences in the event shapes, which shall be exploited by the classifier, are visible in the center-of-mass reference frame in which the B mesons are produced almost at rest. Hence, there is no preferred direction for the decay products of a B meson decay, resulting in an isotropic, spherical distribution of the decay products. The light quarks produced in continuum events on the other hand, do have a high velocity resulting in a common preferential direction for the products of the hadronization as illustrated in Figure 6.1.

A common set of observables is used as input features for the continuum suppression BDT. This collection of observables combines the knowledge of many experimental particle physics collaboration facing similar challenges when describing the shapes of whole events or jet-like signatures. A comprehensive overview of these variables is given in [1, Section 9].

Thrust Variables

Thrust can be defined for a collection of particle candidates, e.g. the particles combined to a B meson candidate. Given such a collection of particles, the thrust axis $\vec{T}$ is defined as the unit vector for which the sum of projections of the particle momenta $\vec{p}_i$ onto itself is maximized:

$$\vec{T} = \max_{|\vec{T}|=1} \left(\sum_i^N \vec{T} \cdot \vec{p}_i \right) \tag{6.1}$$

where the sum is taken over all N final state particles of the collection. The scalar

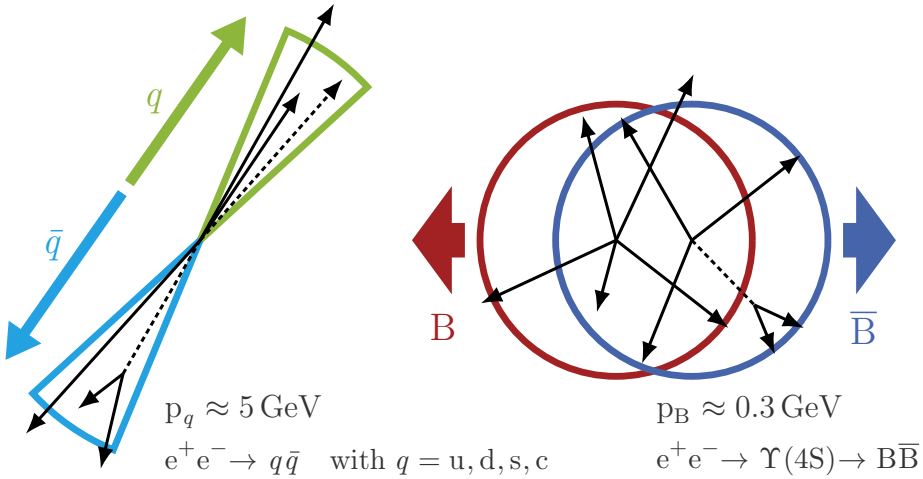


Figure 6.1.: Illustration of the difference in the event signatures of continuum background events on the left and $B\overline{B}$ events on the right in the center-of-mass reference frame. Due to the high amount of energy available for the light quarks $q = u$, d, s, c produced in $e^+e^- \to q\bar{q}$ events, the produced particles exhibit a jet-like signature. These back-to-back jets can be distinguished from the spherical appearance of B meson decays of $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ events. Adapted from [70].

observable thrust T can be calculated, once the thrust direction is known:

$$T = \frac{\sum_{i=1}^{N} |\vec{T}\vec{p}_i|}{\sum_{i=1}^{N} |\vec{p}_i|}. \tag{6.2}$$

This observable quantifies the jet-likeness of the collection of particles, as it will take on values close to 1, if the momenta of all considered final state particles share the same direction, as they would in a hadron jet. On the other hand, the isotropic direction distribution of the decay products of B mesons lead to a thrust value close to 0.5.

For a completely reconstructed $\Upsilon(4S)$ event, the thrust axis of both B meson candidates can be calculated to evaluate their jet-likeness.

Further observables based on the B meson candidates' trust axis can be used to identify continuum background. The cosine of the angle between the thrust axes of the two B meson candidates, $\cos\theta_{\mathrm{TBTO}}$, discriminates between actual B mesons and continuum background, as the back-to-back jet signature of an $e^+e^- \to q\bar{q}$ event in the CMS reference frame is more likely to result in an angle of 180° between the two thrust vectors. In turn, the isotropic distribution of the directions of the momenta of B meson decay products leads to a uniform distribution for $\cos\theta_{\mathrm{TBTO}}$, as the thrust axes point into random directions.

Similarly, the distributions of the observable $\cos\theta_{\mathrm{TB}z}$, based on the angle between B meson candidate thrust axis and the beam axis (= z-axis), allows separating continuum from $B\overline{B}$ events. This is possible, due to the difference in angular distributions resulting from an $\Upsilon(4S)$ (spin 1) decay into two B mesons (spin 0) compared to jet-like signature resulting from the $e^+e^- \to q\bar{q}$ process and the hadronization of the spin-1/2 quarks.

CLEO Cones

The CLEO cones — named after the CLEO collaboration which introduced them in

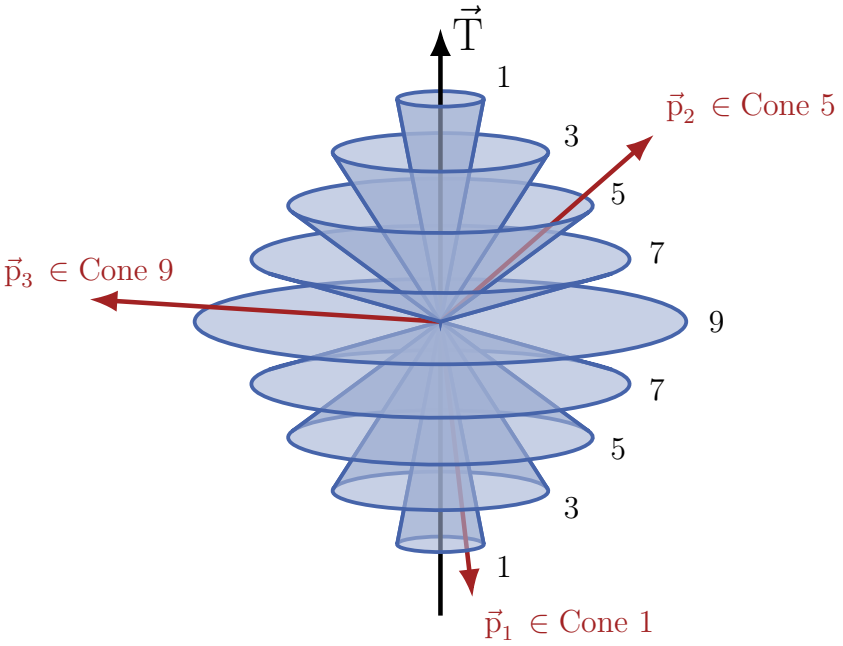


Figure 6.2.: Visualization of the CLEO cones around a given thrust axis $\vec{T}$. Shown are the boarders defined by the cones with increasing opening angles in 10° steps. To simplify the illustration, only every other cone is drawn. The cones are numbered, indicating the related CLEO cone index j. The circle labeled with the number 9 essentially splits the 9th CLEO cone region in half. The particle momenta vectors $\vec{p}_{1,2,3}$ shown in red are added to illustrate association of their momentum flow to the cone regions.

1996 [71] — define a set of observables calculated from the momentum flow through cones around the thrust axis $\vec{T}$ of a B meson candidate. In steps of 10° of the cone opening angle, nine CLEO cones are defined by the sum of particle momenta

$$\text{CLEO Cone}(j) = \sum_{i}^{N} |\vec{p}_i| \quad \begin{aligned} &\text{if } \theta_{\vec{T},\vec{p}_i} \in [(j-1)\cdot 10°, j\cdot 10°] \\ &\text{or } \theta_{\vec{T},\vec{p}_i} \in [180° - (j-1)\cdot 10°, 180° - j\cdot 10°] \\ &\text{for } j \in \{x \in \mathbb{N} \,|\, 1 \leqq x \leqq 9\} \end{aligned} \tag{6.3}$$

where j is an integer between 1 and 9 indicating the angular region covered by the cone. An illustration of the CLEO cones is shown in Figure 6.2. These CLEO cone observables are calculated by considering the momenta of all final state particles of an event ($N = N_{\text{all}}$), or only the particles not included in a given B meson candidate ($N = N_{\text{ROE}}$). For the presented analyses all particles are used for the calculations of CLEO cones.

Fox-Wolfram Moments

Introduced in 1978 by Geoffrey C. Fox and Stephen Wolfram [72, 66], the harmonic moments H_l of two particles — also referred to as Fox-Wolfram moments (FWM) — were originally used for the analysis of event shapes in e^+e^- annihilation. As such, they are predestined to be used to differentiate the event signatures of $B\overline{B}$ signal events and continuum events.

The l-th order Fox-Wolfram moment H_l for a collection of N particle candidates is defined as

$$H_l = \sum_{i,j}^{N} |\vec{\mathrm{p}}_i||\vec{\mathrm{p}}_i| P_l(\cos\theta_{ij}) \tag{6.4}$$

with the help of the l-th order Legendre polynomial $P_l(\cos\theta_{ij})$, where θ_{ij} is the angle between the momentum vectors $\vec{\mathrm{p}}_{i,j}$ of the i-th and j-th particles of the collection. In the original publication by Fox and Wolfram, each element of the sum was weighted by the square of the total energy of the event. When considering the reduced FWM R_k defined as the ratio

$$\mathrm{R}_k = \frac{H_k}{H_0} \tag{6.5}$$

this constant factor cancels. For jet-like structures these reduced FWM R_k take on values close to 0 for odd l and close to 1 for even l. The most commonly used reduced FWM is $\mathrm{R}_2 = H_2/H_0$. Both, the FWM and their reduced versions, can be calculated for the signal and tag side B meson candidates, as well as the whole event. For the purpose of the presented analyses, the tag side based calculation of R_2 is used as input to the continuum suppression BDT.

Modified Fox-Wolfram Moments

For the purpose of continuum suppression at Belle, extended variants of the FWM are calculated, differentiating between particles originating from the reconstructed B meson candidate and the remaining particles. These two categories are denoted by the letters s for *signal* and o for *other*, respectively. When comparing particles from the B meson candidate with the remaining particles (so), an additional differentiation between the particles of the second category based on their charge is made. The particle charge is either designated with c for *charged* or n for *neutral*. Additionally, the index indicating the charge is overloaded to refer to missing particles (the missing momentum) of an event candidate. This case is denoted by the letter m for *missing*. Using this syntax, the modified Fox-Wolfram moments H_{lx}^{so} and H_l^{oo} are defined, where $l \in [0..4]$ again refers to the order of the Legendre polynomial and x refers to one of the three charge categories c, n or m. They are given by the equations

$$H_{ln}^{so} = H_{lm}^{so} = 0 \qquad \text{and} \tag{6.6}$$

$$H_{lc}^{so} = \sum_{i}^{N_s}\sum_{j}^{N_{oc}} Q_i Q_j |\vec{\mathrm{p}}_j| P_l(\cos\theta_{ij}) \qquad \text{for odd } l; \tag{6.7}$$

$$H_{lx}^{so} = \sum_{i}^{N_s}\sum_{j}^{N_{ox}} |\vec{\mathrm{p}}_j| P_l(\cos\theta_{ijx}) \qquad \text{for even } l \tag{6.8}$$

where i runs over all particles of the reconstructed B candidate N_s and j indexes the remaining particles in the categories c, n and the missing momentum m. In the latter category, only the missing momentum $\vec{\mathrm{p}}_m$ calculated for the event using the beam parameters and all visible particles, and the angle θ_{im} between the respective B meson side particle momentum $\vec{\mathrm{p}}_i$ and the missing momentum is considered. The variables $Q_{i,j}$ stand for the charges of the involved particles. These moments, which involve the recombined B meson candidate itself, can be calculated using the momenta of the direct daughter particles of the B meson, or the related final state particles. For the presented analyses, the primary daughters recombined to the B meson candidate are used.

Similarly, the modified FWM for the *oo* case are given by

$$H_l^{oo} = \sum_j^{N_o} \sum_{k \neq j}^{N_o} Q_i Q_j |\vec{\mathrm{p}}_j||\vec{\mathrm{p}}_k| P_l(\cos\theta_{jk}) \qquad \text{for odd } l; \tag{6.9}$$

$$H_l^{oo} = \sum_j^{N_o} \sum_{k \neq j}^{N_o} |\vec{\mathrm{p}}_j||\vec{\mathrm{p}}_k| P_l(\cos\theta_{jk}) \qquad \text{for even } l \tag{6.10}$$

using the same definitions as above, but summing only over pairs of particles not associated to the B meson. For the considered collection of other particles N_o the basic selection criteria for charged tracks, given in the Equations (5.6) and (5.7), and neutral cluster 5.12 must be fulfilled.

To avoid a dependence on the energy difference $\Delta \mathrm{E_B}$ of the B meson candidate used for the calculations, the modified FWM are normalized:

$$H_{lx}^{so} = \frac{\hat{H}_{lx}^{so}}{2 \cdot (\mathrm{E_{CMS}} - \mathrm{E_B^{CMS}})} \quad \text{and} \quad H_l^{oo} = \frac{\hat{H}_l^{oo}}{4 \cdot (\mathrm{E_{CMS}} - \mathrm{E_B^{CMS}})^2}, \tag{6.11}$$

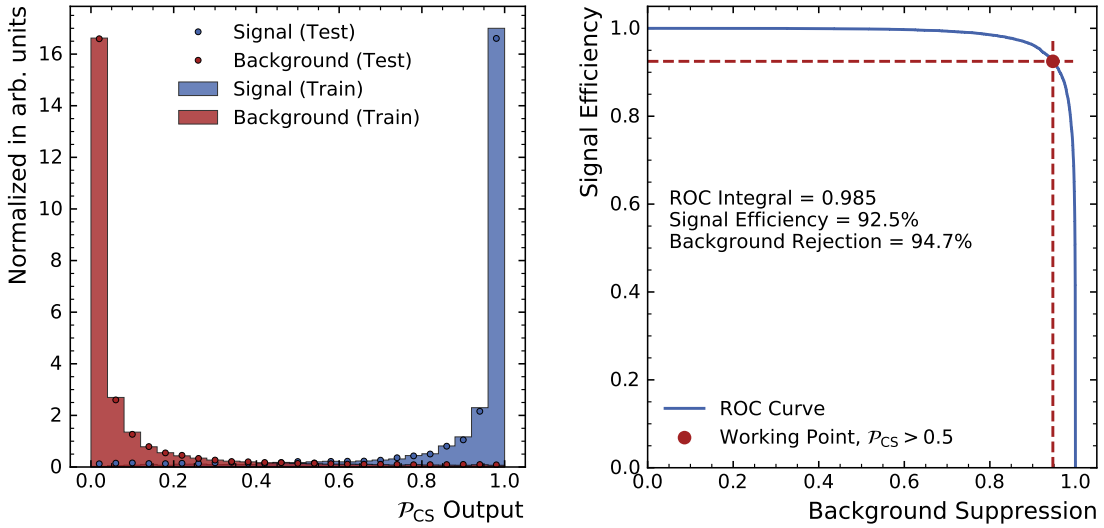
where $\hat{H}$ refers to the respective unnormalized moments defined above. In the denominators, $\mathrm{E_{CMS}} = \sqrt{s}$ is the center-of-mass energy and $\mathrm{E_B^{CMS}}$ is the energy of the B meson candidate in the center-of-mass reference frame.

In total 16 non-zero normalized modified FWM are calculated; five for the *oo* case and $5 + 3 + 3$ for the categories *c*, *n* and *m*, respectively, of the *so* case. Furthermore, the event level squared missing mass $\mathrm{M_m^2}$ and the transverse energy E_t are calculated from the considered particles.

The observables described above are evaluated regarding their correlation to the fit observables $\mathrm{M_{miss}^2}$ and p_ℓ^* to avoid using observables with an obvious dependence. The remaining subset is used as input features for the training of the stochastic gradient-boosted decision trees. As training samples for the continuum suppression BDT, 20 000 independent continuum events from the $\mathrm{c\bar{c}}$– and $q\bar{q}$–samples, each, with $q = \mathrm{u, d, s}$ as described in Section 3.2.1, as well as an equal amount of $\mathrm{B} \to \mathrm{D}^{(*)}\tau\nu$ signal and $\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu$ normalization events are used. A second independent sample of matching size is used for the validation of the classifier. All the samples used for this purpose are excluded from the MC samples for the remaining analysis steps.
The classifier from the FastBDT library is trained with an optimized hyperparameter setup specified in Table 6.1.
The results of the training process and validation are visualized in Figure 6.3, which shows the distribution of the BDT output $\mathcal{P}_{\mathrm{CS}}$, as well as the receiver operating characteristic

Table 6.1.: Hyperparameter setup used for the training process of the FastBDT library–based continuum suppression classifier.

Hyperparameter	Value
# Trees	200
Max Depth	3
Shrinkage	0.1
Sub-Sample per Tree	0.5



(a) CS MVA Output

(b) ROC Curve

Figure 6.3.: Evaluation plots of the performance of the continuum suppression BDT. On the left side the distribution of the BDT output $\mathcal{P}_{CS}$ is shown for the training (bars) and validation (dots) samples. Exellent separation between signal and continuum background candidates can be observed, which is equally good for training and validation samples. On the right-hand side the ROC curve calculated for the validation sample is shown together with the AUC value and the efficiency on signal and background samples.

(ROC) curve. The $\mathcal{P}_{CS}$ distribution indicates a high separation power between continuum background events and signal/normalization events for both, the training sample and the validation sample. No hints for an over-training of the BDT are visible, as the BDT performs equally well on the training and test samples. This separation power is quantified in the shape of the ROC curve illustrating the relation between signal efficiency and background suppression, and, in a more quantitative manner, by the related area under the curve (AUC), which is the integral of the ROC curve. The ROC curve shown in Figure 6.3 is calculated on the independent validation sample and has an AUC of 0.985.

The marked working point for a selection with $\mathcal{P}_{CS} > 0.5$ is used at later stages of the analyses. For this requirement on $\mathcal{P}_{CS}$, a signal efficiency of 92.5 % and background rejection of 94.7 % is obtained on the validation sample. The continuum suppression BDT output $\mathcal{P}_{CS}$ is also used for a further, signal-enriching selection classifiers described in the following subsections.

Beforehand, the continuum suppression BDT output $\mathcal{P}_{CS}$ is used to reject the most obvious $e^+e^- \to q\bar{q}$ background and thereby reduce the size of the otherwise overwhelmingly large continuum MC samples. To achieve this, an initial selection requirement of

$$\mathcal{P}_{CS} > 0.2 \tag{6.12}$$

is applied on all samples before further, more specific selections are applied for the individual analyses. This selection step retains over 97 % of signal candidates, while reducing the continuum background by 87 %.

6.2. Best Candidate Selection

During the reconstruction and selection process a recorded or simulated event is interpreted in various different ways, each of which is handled as a candidate for the event. This multiplicity of an event has to be taken into account for the results of the analyses. Two routes are commonly taken to achieve this. The first one handles the event multiplicity via weights depending on the number of candidates remaining for a given event after the selection process. Option two is a best candidate selection (BCS), which selects one candidate for a given event. This selection is often based on one or multiple observables providing a measure of the quality of a candidate, but can also be done by random choice. For the discussed analysis and the related studies mentioned at the beginning of Chapter 5, the second route is taken. The best candidate is selected based on two observables which are evaluated successively. A candidate is chosen randomly, if the two observables do not yield a conclusive result. The procedure can be split into three steps:

1. **Lowest $\mathbf{nE_{extra}^{ECL}}$**
   The first observable is the sum of the energy of the ECL clusters not associated to the $\Upsilon(4S)$ meson candidate and without a relation to a charged track. This observable is referred to as neutral extra energy nE_{extra}^{ECL}, where the n stands for *neutral*, indicating that only clusters with no relation to charged track are considered. This selection step can be seen as a cluster-based completeness constraint for the event, analogous to the requirement on the number of remaining tracks in the event imposed on the reconstructed $\Upsilon(4S)$ meson candidate in Section 5.3. However, as the ECL reconstruction does not benefit from as many constraints as the track reconstruction while interpreting the detector signatures, more false positives can occur (e.g. due to beam background). Hence, a candidate is not necessarily wrong if one or multiple ECL clusters are not accounted for. Nonetheless, the remaining energy in the calorimeter can be used to reject combinatorial or physical background, as high values of this observable still indicate unused high-energetic photons. This information is utilized for the BCS by choosing the $\Upsilon(4S)$ meson candidate with the lowest nE_{extra}^{ECL} value.

2. **Lowest $|\mathbf{\Delta E_{tag}}|$**
   A second observable is used for the selection of the best candidate in cases in which the neutral extra energy is not conclusive. This can for instance be the case if two candidates for an event utilize the same final state particles, but recombine them differently, yielding other intermediate particle or even B meson candidates. This is often the case for the tag side B_{tag} meson candidate, where the FEI algorithm found multiple ways to combine the available set of final state particles to form a valid candidate. To resolve such conflicts, the deviation from the expected energy ΔE_{tag} for the tag side B meson, as introduced in Equation (5.31), is used as a second BCS observable. Here, the candidate with the lowest absolute value of ΔE_{tag} is chosen.

3. **Random Choice**
   If the selection is still inconclusive after these two steps, a random candidate is selected for the event. This can be necessary for instance if a track object is interpreted in one case as a pion and in another case as a kaon, and vice versa. For such candidates an in-depth evaluation of the most likely assignment of the individual particles is inefficient and does not affect the underlying analysis. Thus, the random selection is sufficient at this point.

The performance of this BCS procedure depends on the selection steps preceding it. Thus,

it is discussed in the following Section 6.3, which elaborates on the signal event selection for the measurement of the decay rate ratios $\mathcal{R}(\mathrm{D}^{(*)})$.

6.3. Signal Sample Selection

Starting from the samples of fully reconstructed $\Upsilon(4\mathrm{S})$ meson candidates obtained as described in Chapter 5 and with the track completeness constraint applied, the goal is now to produce an analysis-specific sample with enriched signal ($\mathrm{B} \to \mathrm{D}^{(*)}\tau\nu$) and normalization ($\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu$) components for the measurement of the fractions $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$. A combination of rectangular selection criteria and a dedicated classifier is employed to reject physical and combinatorial background and thereby improve the signal-to-noise ratio.

For this purpose, a classifier based on stochastic gradient-boosted decision trees from the FastBDT library [69] is trained to distinguish candidates correctly describing signal or normalization events from combinatorial background or decay processes which are not of interest to the measurement. The selection of candidates based on this BDT is optimized such that the uncertainty on the measured ratios $\mathcal{R}(\mathrm{D}^{(*)})$ is minimized. This is achieved by applying the whole selection process with a loose criterion on the selection BDT output first and then performing a grid search to find the optimal requirement on the BDT output. For this optimization, the ratios $\mathcal{R}(\mathrm{D}^{(*)})$ are extracted from Asimov data via the same fit procedure which will later be applied on recorded data to obtain the $\mathcal{R}(\mathrm{D}^{(*)})$ measurements. This fit procedure will be introduced in Chapter 8.

In the following, the training process of the employed BDT is explained, and its performance is evaluated (Section 6.3.1). Subsequently, it is elaborated on how this BDT is optimized (Section 6.3.2) and employed as part of the analysis selection for the $\mathcal{R}(\mathrm{D}^{(*)})$ measurement (Section 6.3.3). Next, the evaluation of the event multiplicities and the application of the best candidate selection are discussed in Section 6.3.4. The section concludes with an evaluation of the selection's performance in Section 6.3.5 and the introduction of sideband selections used for validations of the analysis in Section 6.3.6.

6.3.1. Multivariate Classifier

The BDT employed to enhance the signal-to-noise ratio in the data sets for the $\mathcal{R}(\mathrm{D}^{(*)})$ measurement is trained to identify correctly reconstructed $\mathrm{B} \to \mathrm{D}^{(*)}\tau\nu$ signal and $\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu$ normalization events. The hyperparameter setup of the BDT is the same as the one used for the continuum suppression BDT described in Table 6.1.

For this purpose, the target is defined by the lepton and the excited or unexcited $\mathrm{D}^{(*)}$ meson which are combined to form the signal $\mathrm{B}_{\mathrm{sig}}$ meson candidate. They must be matched to the same B meson and must feature the appropriate charges. No further true, direct daughters particles must be missing from the recombined $\mathrm{B}_{\mathrm{sig}}$ meson, except for final state radiation photons or particles which are invisible to the detector. The exact decay process of the $\mathrm{D}^{(*)}$ meson is not relevant for the target definition. Whether the lepton originates directly from the $\mathrm{B}_{\mathrm{sig}}$ meson or stems from an intermediate τ decay distinguishes between the normalization and signal process, respectively. However, this is not relevant for the target definition of the BDT training.

Input Features

The input features for the $\mathcal{R}(\mathrm{D}^{(*)})$ selection BDT are described in the following paragraphs.

$\mathrm{D}^{(*)}$ Meson Mass $\mathrm{M}_{\mathrm{D}^{(*)}}$

The invariant mass of the D or D^* meson reconstructed as daughter of the $\mathrm{B}_{\mathrm{sig}}$ meson is a good observable to determine the quality of the signal side reconstruction. It is independent of the leptonic part of the decay process and therefore also of the invisible neutrinos. The reconstruction of the $\mathrm{D}^{(*)}$ meson involves the most complex recombination on the signal side. For correctly reconstructed $\mathrm{B}_{\mathrm{sig}}$ meson candidates, the invariant mass of the D or D^* meson has to be within the uncertainty of the nominal mass of the respective $\mathrm{D}^{(*)}$ meson. Hence, this observable allows for the rejection of mis-reconstructed candidates.

$\mathrm{D}^{(*)}$ Daughter Mass $\mathrm{M}_{\mathrm{D}^{(*)}\,\mathrm{Daughter}}$

In case the signal side B meson decay involves an excited D^* meson, this D^* decays further into an unexcited D meson. The invariant mass of this D meson candidate can provide further separation power, similar to the observable $\mathrm{M}_{\mathrm{D}^{(*)}}$ described above. If the $\mathrm{B}_{\mathrm{sig}}$ daughter is an unexcited D meson, this observable provides access to the invariant mass of the D meson candidates' first daughter, which is either a charged or a neutral kaon (see Table 5.2). Additional input features, such as the reconstruction mode IDs described below, provide further information to be combined with this mass observable to achieve a greater separation.

Continuum Suppression Output $\mathcal{P}_{\mathrm{CS}}$

The continuum background suppression classifier provides valuable information to identify mis-reconstructed candidates. In addition to its capabilities to identify continuum events, which are already suppressed by the loose selection criterion 6.12 on $\mathcal{P}_{\mathrm{CS}}$, it also tends to yield lower scores for wrongly recombined $\Upsilon(4\mathrm{S})$ meson events. Due to this separation power, the continuum separation BDT's output $\mathcal{P}_{\mathrm{CS}}$ is also included as an input feature.

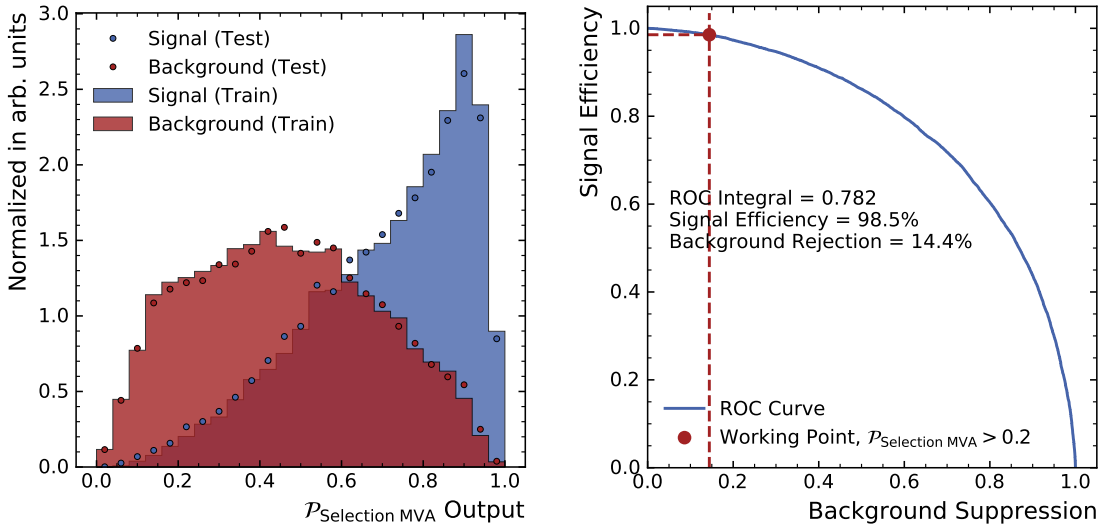
Extra Energy in the Calorimeter $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}}$

Similar to the number of unassociated tracks remaining after the reconstruction of the $\Upsilon(4\mathrm{S})$ meson candidate, the extra energy in the ECL $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}}$ represents unassociated ECL clusters remaining for the event. It is calculated as the sum of the energy of ECL clusters not associated to the reconstructed candidate and fulfilling the basic cluster selection criteria defined in Equation (5.12). As already described for the related observable $\mathrm{nE}^{\mathrm{ECL}}_{\mathrm{extra}}$ utilized by the BCS in Section 6.2, the ECL cluster reconstruction does not allow for a stringent requirement as it is possible for tracks. Nonetheless, a high value for the sum of the remaining energy in the ECL does hint at neutral particles not considered by the candidate at hand. Candidates with many unassociated ECL clusters, and thus higher $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}}$ values, tend to be falsely reconstructed. However, mismodeling of beam background[1] leads to discrepancies between data and MC in this observable. In particular, for low values of $\mathrm{E}^{\mathrm{ECL}}_{\mathrm{extra}} < 0.6\,\mathrm{GeV}$ a surplus of data is observed. This effect is accounted for by setting the observable to 0.6 GeV for candidates which would have values below the threshold. Thus, the BDT is made blind to this problematic region of the observable.

$\mathrm{B}_{\mathrm{sig}}$ Reconstruction Mode Identifier

The identifier of the reconstruction mode of the signal side B meson candidate provides additional information which allows for the separation of e.g. modes using excited and unexcited $\mathrm{D}^{(*)}$ mesons. The reconstruction mode identifiers also encode the charges of the involved particles. Based on this additional information, the BDT is able to optimize the criteria on related input features.

[1] Consider [73, Section 2.1.3] for a detailed discussion of beam background at Belle II, the basics of which are equally applicable for Belle.



(a) MVA Output

(b) ROC Curve

Figure 6.4.: Performance of the $\mathcal{R}(\mathrm{D}^{(*)})$ selection BDT on training and validation data samples. The distribution of the classifier output $\mathcal{P}_{\text{Sel}}$ is shown on the left. It compares the distributions of training (bars) and validation (dots) samples and shows that the BDT generalizes sufficiently enough, such that a similar outcome is achieved for the validation data set. On the right-hand side, the ROC curve of the BDT, calculated using the independent validation sample, is shown. Integration of the curve yields an AUC of 0.782. The working point of $\mathcal{P}_{\text{Sel}} > 0.2$ for a loose selection criterion based on this observable, resulting in a signal efficiency of 98.5 % and background rejection of 14.4 % is indicated as well. For the final selection a more optimized requirement is used.

$\Upsilon(4S)$ Reconstruction Mode Identifier

Analog to the B_{sig} reconstruction mode ID, the identifier of the $\Upsilon(4S)$ reconstruction mode provides additional information which enable the BDT to learn relations between the reconstruction modes and other observables.

The above listed observables used as input features for the selection classifier are evaluated regarding their correlation with the observables used for the fit to determine the ratios $\mathcal{R}(\mathrm{D}^{(*)})$. The result of this evaluation is shown in Appendix B.1. No strong correlation between the input features and the fit observables is observed, rendering the former as suitable.

Selection MVA Performance

The separation power of the thus obtained selection classifier is evaluated on an independent control sample. The results of this evaluation are illustrated in Figure 6.4. In Figure 6.4a the distribution of the BDT's output is plotted; Figure 6.4b displays the ROC curve obtained on the validation sample. The former shows that a separation between the signal and normalization components and the background is achieved equally well for events from the training and validation samples. This indicates that the classifier generalizes sufficiently. The ROC curve and its integral of 0.782 quantify the selection classifier's separation power.

Using a loose selection requirement based on the BDT output of

$$\mathcal{P}_{\text{Sel}} > 0.2 \tag{6.13}$$

rejects 14.4 % of background candidates whilst retaining 98.5 % of the signal candidates. These numbers only provide a rough estimate of the effect of the selection classifier, however, as the full potential will be assessed by the aforementioned grid search which is described in Section 6.3.2 below.

6.3.2. Optimization of the Selection-MVA Requirement

To choose the optimal working point for the multivariate selection classifier, a grid search is performed to find the working point which yields the lowest statistical uncertainties on the branching fraction ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ of interest. For this search, the selection described below in Section 6.3.3, including the best candidate selection (Section 6.3.4), is applied. As a starting point for the selection criterion on $\mathcal{P}_{\text{Sel}}$ the **loose** working point shown in Figure 6.4b is used.
Based on the thus obtained MC data sets, the fitting procedure to determine the ratios $\mathcal{R}(\mathrm{D}^{(*)})$, described in Chapter 8, is performed for various choices of working points. For each choice the statistical uncertainties on $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ are estimated from the results of the fits to the respective Asimov data sets. With this approach the optimal choice for the selection requirement on $\mathcal{P}_{\text{Sel}}$ is determined based on the sum of the uncertainties on the ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$, as the two are obtained simultaneously. The details of this grid search are specified in Appendix B.2. The best results are obtained by applying different criteria on the BDT output depending on the reconstruction modes, differentiating between modes with the ground-state D meson and the excited D^* meson. The optimal selection criterion based on the BDT is found to be

$$\mathcal{P}_{\text{Sel}} > 0.808 \quad \text{for } \mathrm{B} \to \mathrm{D}\ell \tag{6.14}$$
$$\mathcal{P}_{\text{Sel}} > 0.575 \quad \text{for } \mathrm{B} \to \mathrm{D}^*\ell \tag{6.15}$$

yielding statistical uncertainties in the region of

$$\sigma^{\text{stat}}\left(\mathcal{R}(\mathrm{D})\right) \approx 0.057 \tag{6.16}$$
$$\sigma^{\text{stat}}\left(\mathcal{R}(\mathrm{D}^*)\right) \approx 0.031 \tag{6.17}$$

for the ratios. An illustration of the grid search is available in form of the Figures B.3 and B.4 in Appendix B.2. The results of the optimization, namely the statistical fit uncertainty on the measured quantities, depend on the details of the fit strategy, and are thus subject to modifications of the fit strategy. At the time of writing, the final fit strategy has not yet been select.

6.3.3. Selection Requirements

The following rectangular selection requirements are applied to obtain the data samples for the $\mathcal{R}(\mathrm{D}^{(*)})$ measurement. These requirements are partly motivated by the physical properties of the $\mathrm{B} \to \mathrm{D}^{(*)}\tau\nu$ signal process or by common requirements on the underlying reconstruction methods. As such, they are fixed and do not leave any room for an adjustment of the signal-to-noise ratio. The final selection criterion is based on the previously introduced selection classifier output $\mathcal{P}_{\text{Sel}}$ which is tuned to obtain a high signal-to-noise ratio.

Continuum Suppression Classifier Output $\mathcal{P}_{\mathrm{CS}} > 0.5$

A loose selection requirement defined in 6.12, based on the continuum suppression classifier described in Section 6.1 is already applied at an earlier stage to reduce the amount of data that has to be processed. This requirement is tightened to reject further events of the continuum background category.

FEI $\mathrm{B_{tag}}$ BDT Output $\mathcal{P}_{\mathrm{FEI}} > 0.001$

The BDT output of the final stage of the FEI algorithm $\mathcal{P}_{\mathrm{FEI}}$ is an indicator for the quality of the provided $\mathrm{B_{tag}}$ meson candidates. It behaves like a probability, and thus takes on values between zero and one, where larger values represent a high probability for the $\mathrm{B_{tag}}$ candidate being correctly reconstructed. As the quality of the tag side B meson is secondary compared to the one of the signal $\mathrm{B_{sig}}$ meson, only the $\mathrm{B_{tag}}$ meson candidates of the poorest quality are rejected via the requirement on $\mathcal{P}_{\mathrm{FEI}}$. This is done, since a severely wrong reconstructed $\mathrm{B_{tag}}$ meson candidate often implies, due to the completeness constraint, a wrongly reconstructed signal side. Missing slow particles or permuted daughter particles on the tag side resulting in mediocre $\mathcal{P}_{\mathrm{FEI}}$ scores on the other hand, are negligible regarding their effect on the signal side.

Beam-Constrained Mass of the Tag Side $\mathrm{B_{tag}}$ Meson $\mathrm{M_{bc}^{tag}} > 5.27\,\mathrm{GeV}$

An additional observable related to the quality of the $\mathrm{B_{tag}}$ meson candidate provided by the FEI algorithm is the beam-constraint mass of the meson $\mathrm{M_{bc}^{tag}}$ defined in Equation (5.29). The selection requirement based on this observable is tightened now to

$$\mathrm{M_{bc}^{tag}} > 5.27\,\mathrm{GeV} \tag{6.18}$$

enclosing the clearly pronounced mass peak at $\mathrm{M_B} \approx 5.28\,\mathrm{GeV}$ of the hadronically reconstructed tag side B meson candidate.

Momentum Transfer Squared to the Lepton System $\mathrm{q^2} > 4.0\,\mathrm{GeV^2}$

The momentum transfer squared to the lepton system — or the squared invariant mass of the mediating virtual W boson — is an excellent observable to separate the signal processes involving the τ lepton from large fractions of the normalization modes and other backgrounds based on their physical properties. Its minimal and maximal values are

$$\mathrm{M_\ell^2} \leqq \mathrm{q^2} \leqq (\mathrm{M_B} - \mathrm{M_{D^{(*(*))}}})^2 \tag{6.19}$$

given by the mass of the produced lepton or the mass difference between the B meson and the involved D meson. Due to the high mass of the τ lepton, a higher momentum transfer is necessary, compared to the lighter leptons. Thus, the theoretical minimal $\mathrm{q^2}$ value allowed for the signal components is $\mathrm{q^2_{min}} = \mathrm{M_\tau^2} \approx 3.16\,\mathrm{GeV^2}$. With the selection requirement of a minimal squared momentum transfer of $4.0\,\mathrm{GeV^2}$ the majority of signal events are included, while a considerable fraction of background, as well as normalization events are excluded. The reduction of normalization events can be afforded as they are available in abundance.

The sample that is removed by this selection criterion is still of value, as it does contain properly reconstructed candidates of the normalization and $\mathrm{B} \to \mathrm{D^{**}}\ell\nu$ background components. It is one of the sideband data samples described in Section 6.3.6.

Extra Energy in the ECL $\mathrm{E_{extra}^{ECL}} < 3.0\,\mathrm{GeV}$

The extra energy in the ECL $\mathrm{E_{extra}^{ECL}}$ does, as explained in the context of the selection MVA training in Section 6.3.1, quantify the sum of the energy of ECL clusters remaining

unaccounted for in the event. High values for this observable indicate large amounts of energy not considered by the candidate at hand. This allows one to dismiss obvious background candidates by requiring a conservative maximal value of 3.0 GeV for the sum of the remaining energy in the ECL.

$\mathrm{D}^{(*)}$ Meson Mass Deviation $|\Delta \mathrm{M}_{\mathrm{D}^{(*)}}| < 35\,\mathrm{MeV}$

The invariant mass of the reconstructed D or D^* meson is a powerful observable to reject candidates with mis-reconstructed signal side candidates. The difference between the reconstructed and nominal mass allows for the identification of falsely recombined $\mathrm{D}^{(*)}$ mesons. The selection requirement made during the $\mathrm{B}_{\mathrm{sig}}$ meson reconstruction described in Section 5.1 is tightened further for ground-state and excited D mesons by requiring

$$|\Delta \mathrm{M}_{\mathrm{D}^{(*)}}| < 35\,\mathrm{MeV}. \tag{6.20}$$

Lepton Momentum $|\vec{\mathrm{p}}_\ell| < 2.0\,\mathrm{GeV}$

High values of the momentum of the signal side lepton in the laboratory reference frame are an indication of continuum background, as well as $\mathrm{e}^+\mathrm{e}^- \to \ell\bar{\ell}$ processes. The selection requirement on the lepton momentum reduces the contributions from these components.

Successful Signal Side B Meson Vertex Fit

The final selection criterion is based on the signal side B meson vertex fit described in Section 5.1.5. To reject obvious combinatorial and physical background, it is sufficient to require the vertex fit to be successful. Candidates with failed vertex fits can be identified by p-values of -1. Hence, the requirement

$$\mathcal{P}_{\mathrm{Vertex}}^{\mathrm{B_{sig}}} > 0.0 \tag{6.21}$$

is imposed on the p-value of the signal side B meson vertex fit, to reject mis-reconstructed candidates with an unsuccessful vertex fit.

Optimized Requirement on Selection BDT Output $\mathcal{P}_{\mathrm{Sel}}$

The last requirement imposed for the $\mathcal{R}(\mathrm{D}^{(*)})$ sample selection is the optimized criterion specified in Section 6.3.2 above. The criterion is reconstruction-mode dependent with

$$\mathcal{P}_{\mathrm{Sel}} \begin{cases} > 0.808 & \text{for } \mathrm{B} \to \mathrm{D}\ell \\ > 0.575 & \text{for } \mathrm{B} \to \mathrm{D}^*\ell \\ \geqq 0.0 & \text{for } \mathrm{B} \to \mathrm{D}^{(*)}\pi\ell \end{cases} \tag{6.22}$$

and yields minimal statistical uncertainties for the ratios $\mathcal{R}(\mathrm{D}^{(*)})$ obtained from fits to Asimov samples.

6.3.4. Event Multiplicity and Best Candidate Selection

After the application of the above listed selection criteria the total number of $\Upsilon(4\mathrm{S})$ meson candidates is reduced. The number of candidates per surviving event, however, can still be greater than one. For the purpose of the described measurement, the strategy is to allow only one candidate per event. Thus, the best candidate selection as introduced in Section 6.2 is applied. The event multiplicity after every step of the selection, including the best candidate selection and the optimized requirement on a dedicated classifier to improve the signal-to-noise ratio, is shown in Figure 6.5. This illustration demonstrates the necessity of the selection of one best candidate for each event.

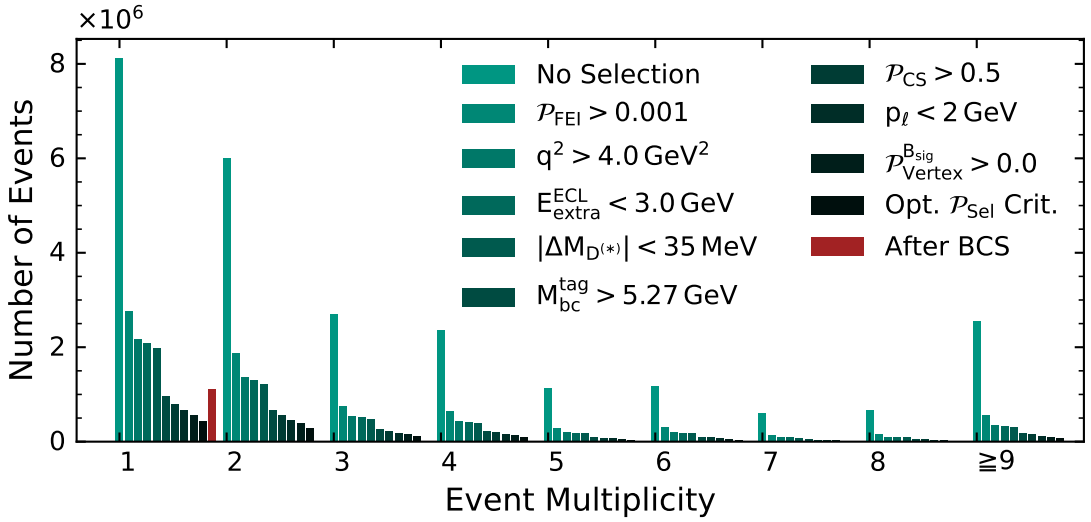


Figure 6.5.: Histogrammed event multiplicities at each step of the selection process for the $\mathcal{R}(D^{(*)})$ measurement, evaluated on MC. The bars indicate the number of events with the given event multiplicity, i.e. for which the respective number of candidates was found. The color of the bars indicate the stage in the selection process, with darker green shades signifying that a higher number of selection requirements are applied. The last applied selection criterion for a given stage is given in the legend. The number of events with high event multiplicities is reduced by each additional selection, and the best candidate selection reduces the number of candidates per event to one for each event, as intended. The status after the application of the best candidate selection is illustrated by the red bar. The last bin also contains the overflow, i.e. the number of events with a multiplicity of 9 or higher.

The effect of the best candidate selection described in Section 6.2 on the event selection is summarized in Table 6.2. In this table the performance of the BCS is listed for signal, normalization and the main background B_{sig} decay categories. This and all following performance evaluations shown in this section are based on the considered MC samples, neglecting the additional correction weights which are applied for the measurement later on. The efficiencies for selecting the candidate matching the actual decay process are approximately 90 % for each of the categories. The lowest efficiencies are observed for the signal process $B \to D\tau\nu$ and the normalization process $B \to D\ell\nu$ involving unexcited D mesons with $\approx 88\,\%$. For events with signal side B decays into D^* mesons, a correctly reconstructed candidate is selected in about 91 % of the cases.

As the main background processes $B \to D^{**}\ell\nu$ are reconstructed in dedicated reconstruction modes, and only one candidate from all reconstruction modes is selected for an event, high efficiencies for correctly identifying these background processes is desirable. By adding these reconstruction modes explicitly, the respective background events are collected in them, and thus are prevented from cluttering the signal-enriched reconstruction modes. For these main background events with $B \to D^{**}\ell\nu$ decay processes on the signal side, a matching candidate is chosen by the BCS in well over 93 % of the cases.

A detailed evaluation of the BCS with respect to the reconstruction modes in which the candidates are found is given in Table 6.3. For this evaluation the three categories signal (Sig), normalization (Norm) and background (Bkg), are defined with respect to the

Table 6.2.: Efficiency of the best candidate selection for different event categories evaluated on MC. The event categories are defined by the B_{sig} meson decay process occurring on the signal side. The table lists the number of events with at least one correctly reconstructed candidate **before** and **after** the best candidate selection. Thus, the **efficiency** calculated from these numbers quantifies how often a correctly reconstructed candidate is chosen for an event for which one is available. For each of the considered categories, this efficiency lies around 90 %, indicating that the procedure for the selection of the best candidate is valid.

Category	Before	After	Efficiency
$D\ \tau\ \nu_\tau$	7118	6248	87.78 %
$D^*\ \tau\ \nu_\tau$	15600	14213	91.11 %
$D\ \ell\ \nu_\ell$	113889	100568	88.30 %
$D^*\ \ell\ \nu_\ell$	472928	433516	91.67 %
$D^{**}\ (\to D^{(*)}\ \pi^0)\ \ell\ \nu_\ell$	27732	26031	93.87 %
$D^{**}\ (\to D^{(*)}\ \pi^\pm)\ \ell\ \nu_\ell$	94401	90824	96.21 %

reconstruction mode given in the column header of the table. Hence,

- the background category refers here to all mis-reconstructed candidates;
- signal refers to $B \to D^{(*)}\tau\nu$ processes for the main reconstruction modes and to $B \to D^{**}\ell\nu$ processes for the dedicated D^{**} reconstruction modes; and
- the normalization category refers to the $B \to D^{(*)}\ell\nu$ processes.

The latter is only defined for the main reconstruction channels. Contributions from $D^* \to D$ feed-down in $B \to D\ell$ reconstruction modes — i.e. candidates for $B \to D^*\ell\nu$ events which are reconstructed in $B \to D\ell$ modes — are counted as correctly reconstructed in the $B \to D\ell$ mode if the lepton and D meson candidates are matched to the same true B_{sig} meson.

Considering these categories, Table 6.3 lists the number of events for which a correctly recombined candidate is available for each of the categories **before** the best candidate selection and per reconstruction mode. The number of events for which a correctly reconstructed candidate was chosen (labeled as **after** in the table), as well as the ratio with respect to the available events is also listed for all categories. The ratio can be interpreted as efficiency for the BCS, and thus is denoted by **Eff** in the table. For each reconstruction mode the best efficiencies are observed in the signal category. The lowest efficiencies for the signal category are encountered for the $B^0 \to D^{*-}\ell^+$ reconstruction modes. For these $B \to D^*\ell\nu$ events, the best candidate is often found to be reconstructed in $B \to D\ell$ modes as $D^* \to D$ feed-down.

The normalization category exhibits efficiency values similar to the ones found for the signal category in the respective reconstruction mode. As desired, the background category containing mis-reconstructed candidates show the lowest efficiencies, indicating that the BCS rejects such candidates. Based on these numbers, the conclusion can be drawn that the best candidate selection fulfills its purpose.

Table 6.3.: Performance of the best candidate selection procedure for each reconstruction mode and the three B_{sig} meson candidate categories signal, normalization and background, evaluated on MC. The B_{sig} categories are defined with respect to the reconstruction mode. Feed-down candidates, i.e. candidates for $B \to D^{*}\ell\nu$ events which are reconstructed in the $B \to D\ell$ mode, are accepted for the respective reconstruction mode.

	$B^0 \to D^- e^+$	$B^0 \to D^- \mu^+$	$B^- \to D^0 e^-$	$B^- \to D^0 \mu^-$
Sig before	1955	743	10864	4880
Sig after	1682	644	7802	3563
Norm before	40533	32364	244614	195695
Norm after	34029	27345	172243	139326
Bkg before	30398	16238	110553	62917
Bkg after	12843	7292	32862	20597
Sig Eff	86.04 %	86.68 %	71.82 %	73.01 %
Norm Eff	83.95 %	84.49 %	70.41 %	71.20 %
Bkg Eff	42.25 %	44.91 %	29.73 %	32.74 %

	$B^0 \to D^{*-} e^+$	$B^0 \to D^{*-} \mu^+$	$B^- \to D^{*0} e^-$	$B^- \to D^{*0} \mu^-$
Sig before	636	380	4639	2560
Sig after	346	217	4006	2201
Norm before	13974	12552	91957	81418
Norm after	7454	6855	77134	69698
Bkg before	11295	9449	115332	82273
Bkg after	3144	2684	34350	26744
Sig Eff	54.40 %	57.11 %	86.35 %	85.98 %
Norm Eff	53.34 %	54.61 %	83.88 %	85.61 %
Bkg Eff	27.84 %	28.41 %	29.78 %	32.51 %

	$B^0 \to D^- e^+ \pi^0$	$B^0 \to D^- \mu^+ \pi^0$	$B^- \to D^0 e^- \pi^0$	$B^- \to D^0 \mu^- \pi^0$
Sig before	3116	2542	11934	9816
Sig after	2633	2168	8713	7365
Bkg before	37749	27279	103256	75971
Bkg after	14541	11021	28900	23125
Sig Eff	84.50 %	85.29 %	73.01 %	75.03 %
Bkg Eff	38.52 %	40.40 %	27.99 %	30.44 %

	$B^0 \to D^{*-} e^+ \pi^0$	$B^0 \to D^{*-} \mu^+ \pi^0$	$B^- \to D^{*0} e^- \pi^0$	$B^- \to D^{*0} \mu^- \pi^0$
Sig before	308	225	3117	2424
Sig after	269	190	2623	2070
Bkg before	3391	2447	35988	25302
Bkg after	1335	959	11129	8562
Sig Eff	87.34 %	84.44 %	84.15 %	85.40 %
Bkg Eff	39.37 %	39.19 %	30.92 %	33.84 %

	$B^0 \to \overline{D}^0 e^+ \pi^-$	$B^0 \to \overline{D}^0 \mu^+ \pi^-$	$B^- \to D^+ e^- \pi^-$	$B^- \to D^+ \mu^- \pi^-$
Sig before	27544	23104	22663	19438
Sig after	20212	16940	20697	17769
Bkg before	144751	110040	96467	72442
Bkg after	44767	35122	41704	32327
Sig Eff	73.38 %	73.32 %	91.33 %	91.41 %
Bkg Eff	30.93 %	31.92 %	43.23 %	44.62 %

	$B^0 \to \overline{D}^{*0} e^+ \pi^-$	$B^0 \to \overline{D}^{*0} \mu^+ \pi^-$	$B^- \to D^{*+} e^- \pi^-$	$B^- \to D^{*+} \mu^- \pi^-$
Sig before	7101	5987	2317	1949
Sig after	6119	5206	2104	1777
Bkg before	47399	34343	8826	5993
Bkg after	17740	13220	3693	2624
Sig Eff	86.17 %	86.96 %	90.81 %	91.17 %
Bkg Eff	37.43 %	38.49 %	41.84 %	43.78 %

Table 6.4.: Event counts and efficiencies for each individual selection step, evaluated per B_{sig} meson decay process category on MC. The event categories are defined by the B_{sig} meson decay process occurring on the signal side. The efficiencies (Eff) are calculated with respect to the previous selection step, except for the total efficiency listed in the last row.

	$D\,\tau\,\nu_\tau$		$D^*\,\tau\,\nu_\tau$		$D\,\ell\,\nu_\ell$	
Criteria	Counts	Eff in %	Counts	Eff in %	Counts	Eff in %
No Selection	46178	—	93718	—	1410297	—
$\mathcal{P}_{\text{FEI}} > 0.001$	22157	47.98	44589	47.58	634867	45.02
$q^2 > 4.0\,\text{GeV}^2$	21512	97.09	44370	99.51	316012	49.78
$E^{\text{ECL}}_{\text{extra}} < 3.0\,\text{GeV}$	21308	99.05	43973	99.11	311424	98.55
$\left\lvert\Delta M_{D^{(*)}}\right\rvert < 35\,\text{MeV}$	20641	96.87	42587	96.85	296236	95.12
$M^{\text{tag}}_{\text{bc}} > 5.27\,\text{GeV}$	17333	83.97	35412	83.15	222833	75.22
$\mathcal{P}_{\text{CS}} > 0.5$	14870	85.79	29165	82.36	216175	97.01
$p_\ell < 2\,\text{GeV}$	14551	97.85	28333	97.15	173884	80.44
$\mathcal{P}^{B_{\text{sig}}}_{\text{Vertex}} > 0.0$	12987	89.25	25573	90.26	153714	88.40
Opt. $\mathcal{P}_{\text{Sel}}$ Crit.	7118	54.81	15600	61.00	113889	74.09
After BCS	6248	87.78	14213	91.11	100568	88.30
Total Eff	—	13.53	—	15.17	—	7.13

	$D^*\,\ell\,\nu_\ell$		$D^{**}\,(\to D^{(*)}\,\pi^0)\,\ell\,\nu_\ell$		$D^{**}\,(\to D^{(*)}\,\pi^\pm)\,\ell\,\nu_\ell$	
Criteria	Counts	Eff in %	Counts	Eff in %	Counts	Eff in %
No Selection	3711347	—	309777	—	1593799	—
$\mathcal{P}_{\text{FEI}} > 0.001$	1649801	44.45	136203	43.97	614270	38.54
$q^2 > 4.0\,\text{GeV}^2$	1331321	80.70	68110	50.01	283684	46.18
$E^{\text{ECL}}_{\text{extra}} < 3.0\,\text{GeV}$	1312008	98.55	67047	98.44	278869	98.30
$\left\lvert\Delta M_{D^{(*)}}\right\rvert < 35\,\text{MeV}$	1259654	96.01	63709	95.02	264557	94.87
$M^{\text{tag}}_{\text{bc}} > 5.27\,\text{GeV}$	1001860	79.53	42448	66.63	148701	56.21
$\mathcal{P}_{\text{CS}} > 0.5$	957565	95.58	36973	87.10	139231	93.63
$p_\ell < 2\,\text{GeV}$	715337	74.70	31975	86.48	120834	86.79
$\mathcal{P}^{B_{\text{sig}}}_{\text{Vertex}} > 0.0$	641137	89.63	27732	86.73	94401	78.12
Opt. $\mathcal{P}_{\text{Sel}}$ Crit.	472928	73.76	27732	100.00	94401	100.00
After BCS	433516	91.67	26031	93.87	90824	96.21
Total Eff	—	11.68	—	8.40	—	5.70

6.3.5. Selection Evaluation

The overall performance of the analysis selection described in this section is quantified in the Tables 6.4 to 6.6. As stated above, this evaluation is based on the full MC sample and neglects the data-MC efficiency correction factors described in Chapter 4. Moreover, the efficiencies of the Belle data acquisition, as well as the event reconstruction described in Chapter 5 are not included in this evaluation. Hence, the total efficiency numbers provided in the tables (i.e. the last rows labeled as **Total Eff**) only reflect the performance of the selection steps described in this section. The overall efficiencies used to determine the decay rate ratios $\mathcal{R}(D^{(*)})$ are calculated after including these additional factors, as described in Chapter 8.

Table 6.4 lists the event counts and efficiencies after each selection step for the different B_{sig} meson decay processes categories. The efficiencies are calculated from the number of events before and after each selection step. The categories are defined by the true B_{sig} decay process that occurred on the signal side of an event and an event must have at

Table 6.5.: Event counts and efficiencies for each individual analysis specific selection step, evaluated per reconstruction mode on MC (part 1 of 2). The Efficiencies (Eff) are listed in percent.

		$B^0 \to D^- e^+$		$B^0 \to D^- \mu^+$		$B^- \to D^0 e^-$		$B^- \to D^0 \mu^-$	
Criteria		Counts	Eff	Counts	Eff	Counts	Eff	Counts	Eff
No Selection	Signal	24603	—	12564	—	67881	—	35937	—
	Norm	773874	—	601913	—	2062722	—	1672784	—
	Bkg	2990753	—	1896186	—	5002610	—	3481256	—
$\mathcal{P}_{\mathrm{FEI}} > 0.001$	Signal	8974	36.48	4921	39.17	33457	49.29	18988	52.84
	Norm	252203	32.59	217219	36.09	944476	45.79	823305	49.22
	Bkg	648001	21.67	499226	26.33	1540713	30.80	1270345	36.49
$q^2 > 4.0\,\mathrm{GeV}^2$	Signal	8780	97.84	4812	97.79	33127	99.01	18793	98.97
	Norm	174166	69.06	154021	70.91	689211	72.97	616740	74.91
	Bkg	565814	87.32	430914	86.32	1307057	84.83	1074035	84.55
$E_{\mathrm{extra}}^{\mathrm{ECL}} < 3.0\,\mathrm{GeV}$	Signal	8689	98.96	4772	99.17	32777	98.94	18632	99.14
	Norm	169906	97.55	152540	99.04	674667	97.89	611604	99.17
	Bkg	520916	92.06	409587	95.05	1235718	94.54	1039817	96.81
$\left\lvert\Delta M_{D^{(*)}}\right\rvert < 35\,\mathrm{MeV}$	Signal	8377	96.41	4597	96.33	31685	96.67	17983	96.52
	Norm	160856	94.67	144663	94.84	644653	95.55	585188	95.68
	Bkg	485128	93.13	381631	93.17	1161291	93.98	977357	93.99
$M_{\mathrm{bc}}^{\mathrm{tag}} > 5.27\,\mathrm{GeV}$	Signal	6274	74.90	3435	74.72	26846	84.73	15345	85.33
	Norm	104654	65.06	96423	66.65	519853	80.64	477879	81.66
	Bkg	220486	45.45	179250	46.97	738612	63.60	641833	65.67
$\mathcal{P}_{\mathrm{CS}} > 0.5$	Signal	5175	82.48	2902	84.48	22185	82.64	12996	84.69
	Norm	99366	94.95	92680	96.12	495625	95.34	460806	96.43
	Bkg	173881	78.86	148936	83.09	644920	87.32	577284	89.94
$p_\ell < 2\,\mathrm{GeV}$	Signal	5089	98.34	2798	96.42	21759	98.08	12483	96.05
	Norm	78492	78.99	69469	74.96	384482	77.58	338436	73.44
	Bkg	151090	86.89	123168	82.70	529180	82.05	448757	77.74
$\mathcal{P}_{\mathrm{Vertex}}^{\mathrm{B_{sig}}} > 0.0$	Signal	4400	86.46	2390	85.42	19732	90.68	11298	90.51
	Norm	66229	84.38	58460	84.15	346726	90.18	305192	90.18
	Bkg	122833	81.30	100253	81.40	467574	88.36	396426	88.34
Opt. $\mathcal{P}_{\mathrm{Sel}}$ Crit.	Signal	1955	44.43	743	31.09	10864	55.06	4880	43.19
	Norm	40533	61.20	32364	55.36	244614	70.55	195695	64.12
	Bkg	40285	32.80	27605	27.54	256016	54.75	191651	48.34
After BCS	Signal	1682	86.04	644	86.68	7802	71.82	3563	73.01
	Norm	34029	83.95	27345	84.49	172243	70.41	139326	71.20
	Bkg	31680	78.64	21683	78.55	170524	66.61	129525	67.58
Total Eff	Signal	—	6.84	—	5.13	—	11.49	—	9.91
	Norm	—	4.40	—	4.54	—	8.35	—	8.33
	Bkg	—	1.06	—	1.14	—	3.41	—	3.72

least one candidate reconstructed in the matching reconstruction mode to be considered. Additionally, $D^* \to D$ feed-down candidates are accepted. The total efficiency of all analysis specific selection steps is given in the last row. The highest efficiencies are observed for events of the $B \to D^{(*)}\tau\nu$ signal categories. It is noteworthy, that the selection criterion based on the optimized selection classifier described in Sections 6.3.1 and 6.3.2 reduces the efficiency for the signal processes $B \to D^{(*)}\tau\nu$ significantly more than for the normalization processes. Despite this, the requirement yields the best uncertainties for the ratios $\mathcal{R}(D^{(*)})$ based on the conducted Asimov study.

Analogously, Tables 6.5 and 6.6 provide an overview of the affect of each selection requirement per reconstruction mode. Again, the three categories signal, normalization (Norm) and background (Bkg) are distinguished. However, here they are examined with respect to the main reconstruction modes. For the reconstruction modes dedicated to the main $B \to D^{**}\ell\nu$ background, a distinction is made only between the two categories signal and background (Bkg). Events are counted to the signal or normalization categories, if candidates are found in the matching reconstruction mode. As above, $D^* \to D$ feed-down candidates are

Table 6.6.: Continuation of event counts and efficiencies for each individual analysis specific selection step, evaluated per reconstruction mode on MC (part 2 of 2). The Efficiencies (Eff) are listed in percent.

		$B^0 \to D^{*-} e^+$		$B^0 \to D^{*-} \mu^+$		$B^- \to D^{*0} e^-$		$B^- \to D^{*0} \mu^-$	
		Counts	Eff	Counts	Eff	Counts	Eff	Counts	Eff
Criteria									
No Selection	Signal	3089	—	1685	—	18838	—	9798	—
	Norm	148506	—	117506	—	609540	—	494484	—
	Bkg	276442	—	169308	—	1488831	—	934400	—
$\mathcal{P}_{\text{FEI}} > 0.001$	Signal	1445	46.78	846	50.21	9547	50.68	5225	53.33
	Norm	67094	45.18	57693	49.10	286721	47.04	249855	50.53
	Bkg	68206	24.67	51783	30.59	380357	25.55	290806	31.12
$q^2 > 4.0\,\text{GeV}^2$	Signal	1394	96.47	820	96.93	9390	98.36	5121	98.01
	Norm	33882	50.50	30067	52.12	203496	70.97	182142	72.90
	Bkg	48932	71.74	35544	68.64	295841	77.78	221632	76.21
$E_{\text{extra}}^{\text{ECL}} < 3.0\,\text{GeV}$	Signal	1385	99.35	816	99.51	9293	98.97	5085	99.30
	Norm	33235	98.09	29848	99.27	199485	98.03	180778	99.25
	Bkg	45111	92.19	33816	95.14	273794	92.55	211140	95.27
$\left\lvert \Delta M_{D^{(*)}} \right\rvert < 35\,\text{MeV}$	Signal	1349	97.40	791	96.94	8987	96.71	4927	96.89
	Norm	31948	96.13	28756	96.34	190789	95.64	173403	95.92
	Bkg	41887	92.85	31409	92.88	252852	92.35	194311	92.03
$M_{\text{bc}}^{\text{tag}} > 5.27\,\text{GeV}$	Signal	1146	84.95	666	84.20	7620	84.79	4269	86.65
	Norm	25631	80.23	23415	81.43	154697	81.08	142570	82.22
	Bkg	19547	46.67	15109	48.10	119858	47.40	94272	48.52
$\mathcal{P}_{\text{CS}} > 0.5$	Signal	957	83.51	572	85.89	6099	80.04	3500	81.99
	Norm	24609	96.01	22709	96.98	147009	95.03	137215	96.24
	Bkg	14839	75.91	12244	81.04	90534	75.53	74696	79.23
$p_\ell < 2\,\text{GeV}$	Signal	943	98.54	551	96.33	5972	97.92	3350	95.71
	Norm	18780	76.31	16300	71.78	111741	76.01	98583	71.85
	Bkg	13463	90.73	10636	86.87	81436	89.95	64528	86.39
$\mathcal{P}_{\text{Vertex}}^{B_{\text{sig}}} > 0.0$	Signal	781	82.82	462	83.85	5399	90.41	2987	89.16
	Norm	15508	82.58	13447	82.50	99663	89.19	87936	89.20
	Bkg	10434	77.50	8185	76.96	68451	84.05	54085	83.82
Opt. $\mathcal{P}_{\text{Sel}}$ Crit.	Signal	636	81.43	380	82.25	4639	85.92	2560	85.70
	Norm	13974	90.11	12552	93.34	91957	92.27	81418	92.59
	Bkg	6219	59.60	5249	64.13	47270	69.06	36031	66.62
After BCS	Signal	346	54.40	217	57.11	4006	86.35	2201	85.98
	Norm	7454	53.34	6855	54.61	77134	83.88	69698	85.61
	Bkg	3144	50.55	2684	51.13	34350	72.67	26744	74.22
Total Eff	Signal	—	11.20	—	12.88	—	21.27	—	22.46
	Norm	—	5.02	—	5.83	—	12.65	—	14.10
	Bkg	—	1.14	—	1.59	—	2.31	—	2.86

accepted as well. For this evaluation of the selection process, the highest efficiencies are also observed for the signal categories. The result for the total efficiencies add up to the numbers shown in Table 6.4, as they should.

6.3.6. Sideband Definitions

Applying particular selection criteria with the requirement inverted such that the value range in which the signal events lie is excluded, provide valuable data samples for validation. Such samples are called *sideband samples* and can be used to compare the effect of the reconstruction and selection approach for data and MC samples. This allows for a validation of the Monte Carlo cocktail which forms the basis of the analysis.
Depending on the selection requirement that is inverted, different aspects of the MC sample can be evaluated. Dedicated sideband selections can be created to validate specific properties, e.g. inverting the selection criterion on the continuum suppression classifier will generate a continuum background enriched sample.
For the analysis at hand, three such sideband samples are considered:

$\mathbf{q^2}$-Sideband: $\mathbf{q^2 < 4.0\,GeV^2}$
The inverted requirement on the momentum transfer to the lepton system provides access to a sample which disregards the $B \to D^* \tau \nu$ signal processes, whilst retaining a significant amount of the $B \to D^* \ell \nu$ normalization processes and other backgrounds.

Continuum-Sideband: $\mathcal{P}_{\mathrm{CS}} < 0.5$
A validation sample which is enriched with the continuum background can be obtained by inverting the criterion on the continuum suppression classifier. This continuum-sideband can be used to validate the modeling of the continuum in the MC cocktail. As the loose pre-selection requirement 6.12 is already imposed before the analysis specific selection is applied, the sample created by this inverted selection does contain events with $\mathcal{P}_{\mathrm{CS}} \in [0.2, 0.5]$. This excludes the obvious continuum background events, but retains a sufficient amount of more signal-like candidates originating from the abundant continuum background for data-MC comparisons.

$\mathbf{M_{bc}^{tag}}$-Sideband: $\mathbf{M_{bc}^{tag} < 5.27\,GeV}$
As the inversion of the requirement in the beam-constrained mass of the tag side B_{tag} meson will provide a sample enriched with mostly wrongly recombined B_{tag} candidates, this sideband allows for the evaluation of the behavior of the combinatorial backgrounds. In contrast to previous studies [12, 9], little use is made of this sideband sample for this analysis, as the relation to the correctly reconstructed events is not straightforward.

Further sideband samples are produced to evaluate the description of the main $B \to D^{**} \ell \nu$ background and continuum background in the MC simulation.

$\mathbf{D^{**}}$-Sideband from $\mathbf{B \to D^{**} \ell}$ Reconstruction Modes
This sideband makes use of the reconstruction modes $B \to D^{**} \ell$ dedicated to the D^{**} background components. As these reconstruction modes are disjunct from the $B \to D \ell$ and $B \to D^* \ell$ modes, they can be used to evaluate said background without compromising the $\mathcal{R}(D^{(*)})$ measurement. They are used to test the modeling of the semi-leptonic B decays into D^{**} mesons which is modified by the reweighting procedure described in Chapter 4. Furthermore, their normalization, and thus the related branching fractions, can be validated in these reconstruction channels. For this purpose, the same selection criteria as described in this section are applied. As defined in Equation (6.22), the optimized criterion on the selection classifier output $\mathcal{P}_{\mathrm{Sel}}$ is not applied for these modes. Additional, specialized selection criteria might be applied for the certain tests in this sideband. Such criteria are specified where necessary.

Off-Resonance-Sideband from Data Recorded at $\sqrt{s} = \mathbf{m_{\Upsilon(4S)} - 60\,MeV}$
The off-resonance data samples, which are introduced in Section 3.1, do not contain any $B\overline{B}$ pair events. The Off-Resonance-Sideband sample is obtained by applying the reconstruction and selection criteria for the $\mathcal{R}(D^{(*)})$ measurement to this recorded data set. As such, it provides a clean sample of recorded continuum events, which reflects the properties of the continuum component in the on-resonance sample used for the measurement. For this purpose, the selection criterion based on the beam-constraint mass of the B_{tag} meson M_{bc}^{tag} defined in Equation (6.18) must be removed, as the center-of-mass energy is below the $B\overline{B}$ threshold in this sample. The effects this might have on the automated tag side reconstruction performed by the FEI algorithm is assumed to be small. Despite the already small reconstruction efficency of the FEI algorithm (see [17]) and the low statistics of the recorded off-resonance data set (see Table 3.1), this sideband can be used to validate the MC description of the $e^+e^- \to q\bar{q}$ continuum processes. The result of this evaluation is presented in Section 9.4.

All of these sideband samples will be made use of in Chapter 9 to validate the data–MC agreement of the samples produced for the presented $\mathcal{R}(D^{(*)})$ analysis.

7. Signal Extraction Methods

The aim of the presented study is the measurement of the branching fraction ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ introduced in Equation (1.1). For this purpose, the number of events remaining for the signal and normalization modes after the presented reconstruction and selection procedure (Chapters 5 and 6) have to be extracted from the distribution of physical observables in the recorded Belle data sample. These event numbers are obtained by fitting the combination of the distributions of all contributing processes to the distribution of the recorded data.

In general, a fit is the optimization of probability density functions (PDFs) describing the individual components (decay processes) and their normalization such that their sum describes the distribution observed in recorded data.

This entails first and foremost that every individual component, i.e. decay process, contributing to the measured distribution is known. Components which are expected to behave similarly in the scope of the physical observable(s) that the fit is based on can be combined to reduce the number of free parameters.

Depending on the knowledge about the underlying distributions of the individual components, the components can be described by analytical PDFs if the shapes of their distributions are well known and can be described by analytical functions. If an analytical description is not possible, PDFs for the components can be generated from the distributions obtained from MC simulation or directly from data (see [74, Section 10.4] for a more elaborate introduction). While in the former case the shape of a PDF can be changed via the parametrization of the analytical function, in the latter the shape is defined by the MC simulation. Due to the inflexibility of the PDF shapes for the latter method, it is often referred to as a template fit.

In both cases, parameters describing the individual PDFs are optimized to minimize the difference between the sum of the contributions and the observed data. In other words, the likelihood that the sum of scaled component PDFs describes the distribution observed in data is maximized. These two points-of-view on the issue at hand describe the two common fit methods, namely the method of least squares and the method of maximum likelihood, in simple terms.

For the purpose of the presented analysis, template PDFs obtained from MC are used to describe the shapes of the components in the two-dimensional space of the fit observables $\mathrm{M}^2_{\mathrm{miss}}$ and p^*_ℓ introduced in Chapter 1. The maximum likelihood method is used to fit these templates to data distributions and thereby obtain the normalization factors of the components.

In the following sections the method of the maximum likelihood fit is introduced (Section 7.1) and its application using template PDFs is described (Section 7.2). A more general description of these methods can be found in [74] or [75]; good overviews are also available in [28, Ch. 40] and [1, Ch. 11]. The summary given here condenses the information given in these sources down to what is relevant to this analysis. The use of Asimov data to obtain estimates for the uncertainties of the fitted parameters is elaborated on in Section 7.4. These tools are used in Chapters 8 and 10 to evaluate the agreement between data and simulation and to estimate the uncertainty on $\mathcal{R}(\mathrm{D}^{(*)})$ using the Asimov technique, respectively.

7.1. Maximum Likelihood Fit

The purpose of a fit is to infer the parameters of a model which is assumed to describe the observed data. In the context of the presented analysis, these parameters are the normalization factors of the signal and normalization components.
Consider the general case of a set of N independent, measured data values $\vec{x} = (x_1, \ldots, x_N)$, which follow the same underlying PDF $f(x; \vec{\varphi})$. The joint PDF $f(x; \vec{\varphi})$ describes the distribution of all components of $\vec{\varphi}$ given a set of M parameters $\vec{\varphi} = (\varphi_1, \ldots, \varphi_M)$. These parameters are (partially) unknown and shall be estimated from the observed data values $\vec{x}$. The values $\vec{x}$ are given by the observed distribution in data and are fixed.
The likelihood $L(\vec{\varphi})$ reflects how well a set of measured values $\vec{x}$ agrees with a given set of parameters $\vec{\varphi}$. Under the condition that the values $\vec{\varphi}$ are independent and identically distributed, this relation can be factorized to

$$L(\vec{\varphi}) = \prod_{i=1}^{N} f(x_i; \vec{\varphi}). \tag{7.1}$$

To obtain the unknown components of $\vec{\varphi}$, the maximum likelihood estimators $\hat{\vec{\varphi}}$, defined as the parameter values which maximize the likelihood function L, are introduced. These $\hat{\vec{\varphi}}$ are found as the values which fulfill the conditions

$$\frac{\partial L}{\partial \varphi_i} = 0 \qquad \text{for all} \quad i \in (1, \ldots, M) \tag{7.2}$$

simultaneously. These equations can often not be solved analytically, and thus numerical methods are usually applied to find the estimates for the unknown $\vec{\varphi}$ values. For the numerical approach it is convenient to use the negative logarithm of the likelihood function, as it will reduce the computational complexity by reducing the product to a sum:

$$-\log L(\vec{\varphi}) = -\sum_{i=1}^{N} \log f(x_i; \vec{\varphi}). \tag{7.3}$$

This transformation is valid, as the logarithm of a function has the same extreme values as the underlying function, due to the strictly monotonously increasing nature of the logarithm. The negative of the expression is used, as minimization algorithms are more readily available. This modified version of the likelihood function is referred to as *negative log-likelihood* and is applicable to all following likelihood definitions.

Extended Maximum Likelihood
As stated in the introduction of this chapter, the normalization factors, or number of events N_k, of the individual processes k are treated as free parameters. For a single PDF $f(x; \vec{\varphi})$ and N data points as discussed above, this additional parameter is included in the extended likelihood function by assuming the number of events N is a Poisson-distributed random variable with the true value ν:

$$L(\vec{\varphi}, \nu) = \mathrm{e}^{-\nu} \, \frac{\nu^N}{N!} \, \prod_{i=1}^{N} f(x_i; \vec{\varphi}). \tag{7.4}$$

If ν is independent of all other parameters $\vec{\varphi}$, Equation (7.2)[1]yields the maximum likelihood estimator $\hat{\nu} = N$.

[1]Equation (7.2) must be extended to also include the derivatives of the additional parameter(s) ν or $\vec{\nu}$.

However, if the normalization parameters are dependent on the other parameters, better estimates can be obtained by including these dependencies in the definition of the likelihood function.

Binned Maximum Likelihood

In particle physics it is common to handle data in the form of histograms, which reduces the computational effort, albeit simultaneously reducing the information contained in the data. The information loss is negligible, if the information on the parameters $\vec{\varphi}$ gained from the variation of the PDF within one bin is small compared to the information gained from the variation over all bins [74]. Considering the simple case of a single PDF $f(x_i;\vec{\varphi})$ and a fixed number of events N again, the likelihood function for the binned data is defined as

$$L(\vec{\varphi}) = N! \prod_{i=1}^{B} \frac{p_i(\vec{\varphi})^{m_i}}{m_i!}, \tag{7.5}$$

where the product iterates over the number of bins $1,\ldots,B$ and m_i is the number of data points contained in the i-th bin. The function $p_i(\vec{\varphi})$ is the integral of the PDF over the bin edges x_i^{min} and x_i^{max} of the i-th bin:

$$p_i(\vec{\varphi}) = \frac{1}{p_{\text{Total}}} \int_{x_i^{min}}^{x_i^{max}} f(x_i;\vec{\varphi})dx, \tag{7.6}$$

where the normalization factor $1/p_{\text{Total}}$ with $p_{\text{Total}} = \sum_{i=1}^{B} p_i(\vec{\varphi})$ has to be included if the scope of the histogrammed data does not cover the full range of the observable x, such that $\sum_i^B p_i = 1$ always holds true.
Extending the binned approach to additionally include an unknown total event number N leads to the likelihood function

$$\begin{aligned} L(\vec{\varphi},\nu) &= \mathrm{e}^{-\nu}\, \frac{\nu^N}{N!}\, N! \prod_{i=1}^{B} \frac{p_i(\vec{\varphi})^{m_i}}{m_i!} \\ &= \prod_{i=1}^{B} \mathrm{e}^{-\nu_i}\, \frac{\nu_i^{m_i}}{m_i!} \qquad (7.7) \\ &= \prod_{i=1}^{B} \mathcal{P}(m_i|\nu_i), \qquad (7.8) \end{aligned}$$

with $N = \sum_i^B m_i$ and using the definition $\nu_i \equiv \nu\, p_i(\vec{\varphi})$ for the expected number of events in the i-th bin given the parameters $\vec{\varphi}$.
The obtained expression is the product of the Poisson probability $\mathcal{P}(m_i|\nu_i)$ of observing m_i events in the i-th bin, given the expectation value ν_i for the events in this bin, over all bins $i = 1,\ldots,B$.

7.2. Template Likelihood Fit

The determination of the decay rates for the ratios $\mathcal{R}(\mathrm{D}^{(*)})$ requires the extraction of the contributions of the signal and normalization processes to the distributions of the fit observables obtained on recorded data. Hence, the number of events n_k — or yield — for

each of the contributing processes (components) k is sought. However, the exact analytical description of the PDFs for each of the processes which constitute the two-dimensional distribution of the considered fit observables $\mathrm{M}^2_{\mathrm{miss}}$ and p^*_ℓ is unknown. Instead, the shapes of the distributions for each component k are taken from simulation. This approach is referred to as a template fit and applicable if the different components can be differentiated based on shape differences alone.
To leading order, the shapes are fixed and, therefore, the parameter vector $\vec{\varphi}$ defining the PDF shapes can be dropped. If the statistical fluctuations of the MC distributions are too large such that the templates are not well-defined, the utilization of an analytical description of the functions $f_k(x)$ should be considered. A further alternative based on the MC distributions are kernel density estimators with the help of which smoothed PDF templates can be obtained. This issue is elaborated on in [74, Section 10.4.1].
For the presented analysis, the binned approach, using templates obtained from the event frequencies given by simulation, is followed. In this case, the probability p_i appearing in Equation (7.5), which represents the probability for an event to fall into the i-th bin, is calculated directly from the binned MC distribution

$$p_i \;=\; \frac{h_i}{\sum_i h_i} \tag{7.9}$$

with h_i being the number of entries in the i-th bin of the MC histogram.
The resulting likelihood function

$$L(\nu) \;=\; \mathrm{e}^{-\nu} \prod_{i=1}^{B} \frac{(\nu p_i)^{m_i}}{m_i!} \;=\; \prod_{i=1}^{B} \mathrm{e}^{-\nu_i} \frac{\nu_i^{m_i}}{m_i!}, \tag{7.10}$$

now only depends on the parameter ν representing the total number of events.

Multiple Dimensions
This method can easily be extended for a fit in D dimensions, with b_d being the number of bins for the observable in dimension d, by flattening the D-dimensional histograms into a single dimension with $B = \prod_{d=1}^{D} b_d$ bins. Despite the simplicity of this extension, this method is quickly affected by the curse of dimensionality for fine binning and multiple dimensions.

Multiple Components
The extension to multiple components ($\equiv$ templates) requires more care. Considering N_{P} different processes contributing to the distribution, the process yields $\vec{n} = (n_1, \ldots, n_{N_{\mathrm{P}}})$, with the true values $\vec{\nu}^{\mathrm{P}} = (\nu_1^{\mathrm{P}}, \ldots, \nu_{N_{\mathrm{P}}}^{\mathrm{P}})$, shall be extracted by the fit. Thus, the expected number of events ν_i in the i-th bin is now a function of the process yields

$$\nu_i(\vec{n}) = \sum_{k}^{N_{\mathrm{P}}} \nu_{ik}(n_k), \tag{7.11}$$

where the sum is taken over the N_{P} components k and ν_{ik} represents the number of events associated to component k in the i-th bin. This contribution of component k to bin i is given by

$$\nu_{ik}(n_k) = n_k f_{ik}(h_k), \tag{7.12}$$

where $f_{ik}(h_k)$ is the fraction of events of component k in the i-th bin relative to the total contribution from this component as given by the underlying MC histogram h_k for component k. This fraction is equivalent to Equation (7.9) in the one component case and represents the shape of component k. It is calculated as the ratio

$$f_{ik}(h_k) = \frac{h_{ik}}{\sum h_{ik}}, \tag{7.13}$$

which is the number of events h_{ik} observed in bin i of the histogram of simulated events for process k relative to the total number of MC events of component k in the considered variable space.

Multiple Reconstruction Modes

Finally, the inclusion of different reconstruction modes or channels j is considered. Assuming a simultaneous fit in a total of N_{Ch} different reconstruction channels, the likelihood function can be written as

$$L(\vec{n}) \;=\; \prod_{j=1}^{N_{\mathrm{Ch}}} \prod_{i=1}^{B} \mathrm{e}^{-\nu_{ij}(\vec{n})} \; \frac{\left(\nu_{ij}(\vec{n})\right)^{m_{ij}}}{m_{ij}!}, \tag{7.14}$$

where an additional product over the channels $j = 1, \ldots, N_{\mathrm{Ch}}$ is included. The likelihood is now a function of the desired yield parameters $\vec{n}$. The dependence on the reconstruction channel is reflected by the index j of the number of observed events m_{ij} in bin i and channel j and the respective expected event counts

$$\nu_{ij}(n_k) = \sum_{k}^{N_{\mathrm{P}}} n_k \epsilon_{jk} f_{ijk}(h_{jk}), \tag{7.15}$$

where the template shape $f_{ijk}(h_{jk})$ and the MC histogram h_{jk} it is obtained from, now also depend on the channel j. The additional factor ϵ_{jk} represents the breakdown of the component (process) k into the different reconstruction modes j. This fraction is also obtained from MC simulation and fulfills the normalization condition $\sum_j \epsilon_{jk} = 1$. The obtained likelihood function allows for the simultaneous determination of the yield parameters $\vec{n}$ of N_{P} processes in N_{Ch} reconstruction modes.

7.3. Further Concepts

This section is meant to introduce additional concepts related to the method of maximum likelihood. In particular, the integration of known, systematic uncertainties affecting the MC distributions making up the foundation of the template fit method is briefly discussed. Following this, common approaches to determine the uncertainties on the fitted parameters are introduced. The latter are not restricted to the template fit method, and thus do apply to any approach based on the method of maximum likelihood.

Template Uncertainties

So far the shapes of the templates are assumed to be fixed by the component distributions in simulation. However, these are themselves statistical distributions and are subject to statistical uncertainties due to the size of the available MC samples. Furthermore, inaccuracies in the modeling or simulation of the MC lead to additional systematic uncertainties

which affect the template shapes or relative rates between different components. These effects are included in the template method by the introduction of so-called nuisances.
The inclusion of nuisances requires knowledge about the systematic uncertainties and the variations they cause in the MC distributions. This information is obtained in various ways, such as parameter variation, toy studies or simply the statistical uncertainty of a bin count. Depending on the type and source of systematic uncertainty, i.e.

- whether it affects template shape or relative normalization, or
- if it exhibits correlation among bins or components,

a systematic uncertainty is integrated as nuisance into the likelihood function.
For a set of N_k^{rate} systematic uncertainties which only affect the rate n_k of a process k, the respective nuisances are included by substituting

$$n_k \to n_k \prod_{l}^{N_k^{\text{rate}}} (1 + \eta_{kl}\theta_{kl}), \tag{7.16}$$

where l denotes the source of each of the relevant systematic uncertainties. The symbol η_{kl} denotes the relative uncertainty induced by the source l on the rate of component k. It is paired with the respective nuisance parameter θ_{kl}, which is a parameter of the fit, constrained by the univariate Gaussian standard normal distribution $\mathcal{N}(\theta_{kl}|0,1)$. This constraint is included into the likelihood function by extending e.g. Equation (7.14) to

$$L(\vec{n},\vec{\theta}) \;=\; L(\vec{n}) \times \prod_{k=1}^{N_{\text{P}}} \prod_{l}^{N_k^{\text{rate}}} \mathcal{N}(\theta_{kl}|0,1), \tag{7.17}$$

where $L(\vec{n})$ is the original likelihood function with the substitution given in Equation (7.16) to which the dependency on the vector of nuisance parameters $\vec{\theta}$ is added.
Systematic uncertainties which affect the template shape of a component k are considered via the substitution

$$f_{ik}(h_k) \;=\; \frac{h_{ik}}{\sum_{b=1}^{B} h_{bk}} \quad\to\quad f_{ik}(h_k) \;=\; \frac{h_{ik}(1+\eta_{ki}\theta_{ki})}{\sum_{b=1}^{B} h_{bk}(1+\eta_{kb}\theta_{kb})} \tag{7.18}$$

for the template shape function f. Here, one systematic shape uncertainty η_{ki} on the template of component k is considered. As it should affect the shape of the template, it depends on the bin index i. This requires one additional nuisance parameter θ_{ki} per template and bin. The uncertainty η_{ki} is considered relative to the expected bin count for the component, and therefore given by the ratio $\eta_{ki} = \sigma_{ki}/h_{ik}$, with σ_{ki} being the uncertainty on the bin count due to the respective systematic uncertainties. The uncertainties on the bin counts are required to be in the form of the covariance matrices V_k for each affected component k and can be the sum of multiple covariance matrices originating from different sources l of systematic uncertainty:

$$V_k = \sum_{l}^{N_k^{\text{shape}}} V_{kl}. \tag{7.19}$$

The σ_{ki} are given by the square root of the diagonal elements of this matrix. The covariance between the bin-wise defined uncertainties on the template shape are included via the constraint on the related nuisance parameters θ_{ki}. For this purpose, the constraint is included

into the likelihood function in form of a multivariate Gaussian distribution $\mathcal{N}(\vec{\theta}_k|0, R_k)$ centered around zero and with R_k being the correlation matrix to V_k. The bin nuisance parameters θ_{ki} relevant to the component k are combined in the vector $\vec{\theta}_k$.
The resulting likelihood function is

$$L(\vec{n}, \vec{\theta}) \;=\; L(\vec{n}) \times \prod_{k=1}^{N_\mathrm{P}} \prod_{l}^{N_k^\mathrm{rate}} \mathcal{N}(\theta_{kl}|0,1) \times \prod_{k=1}^{N_\mathrm{P}} \mathcal{N}(\vec{\theta}_k|0, R_k), \tag{7.20}$$

where both substitutions given in Equations (7.16) and (7.18) have to be applied to the original likelihood function $L(\vec{n})$ to add the dependencies on all nuisance parameters $\vec{\theta}$.

Variance of Estimators
The uncertainties on the parameter estimates produced by the minimization of the negative log-likelihood function are equally important as the estimates themselves.
An estimate of the variance σ_i of a parameter estimate $\hat{\varphi}_i$ which maximizes the likelihood is the square root of the i-th diagonal element of the covariance matrix $V(\hat{\vec{\varphi}})$

$$\sigma_i = \sqrt{\left(V(\hat{\vec{\varphi}})\right)_{ii}}. \tag{7.21}$$

For a sufficiently large data sample, and using the assumption that the likelihood function can be approximated by a Gaussian distribution, this covariance matrix can be estimated as the inverse of the second derivative of the negative log-likelihood function evaluated at the optimized parameter values:

$$V_{ij}^{-1} = \left(\frac{\partial^2 \log L}{\partial\varphi_i \partial\varphi_j}\bigg|_{\vec{\varphi}=\hat{\vec{\varphi}}} \right) = H_{ij}. \tag{7.22}$$

Here, the so-called Hesse matrix is introduced, which contains said second derivatives.
If the above stated conditions are not fulfilled, the estimator variances can still be determined via a computationally more expensive, yet more precise approach: the *profile likelihood* method. For this method, the parameter space of a single parameter φ_i is scanned around the value $\hat{\varphi}_i$ which maximizes the likelihood. If nuisances are considered, the respective nuisance parameters $\vec{\theta}$ are determined by maximizing the likelihood for each value of φ_i in the scanned parameter space. All remaining parameters $\hat{\varphi}_j,\ (j \neq i)$ are set to their optimal value $\vec{\varphi} = \hat{\vec{\varphi}}$. Using this approach, a profile of the negative log-likelihood–ratio

$$\begin{aligned} -2\log\left(\lambda(\varphi_i)\right) &= -2\log\left(\frac{L(\varphi_i, \widetilde{\vec{\theta}}, \hat{\vec{\varphi}})}{L(\hat{\varphi}_i, \hat{\vec{\theta}}, \hat{\vec{\varphi}})} \right) \\ &= -2\log\left(L(\varphi_i, \widetilde{\vec{\theta}}, \hat{\vec{\varphi}}) \right) + 2\log\left(L(\hat{\varphi}_i, \hat{\vec{\theta}}, \hat{\vec{\varphi}}) \right) \end{aligned} \tag{7.23}$$

can be obtained to evaluate the functional form of the likelihood with respect to the parameter φ_i. Here, $\widetilde{\vec{\theta}}$ denotes the vector of nuisance parameters optimized for each scan point in φ_i, while the arguments $\hat{\vec{\theta}}$ of the likelihood function in the denominator are the optimized nuisance parameters from the nominal fit. If no nuisances are considered, these dependencies are dropped and the likelihood function in the numerator is simply evaluated for each point in the parameter space of φ_i.

This likelihood ratio profile can be used to check the validity of the Gaussian approximation, or to determine the lower and upper bounds (σ_i^-, σ_i^+) of the 1σ variance on $\hat{\varphi}_i$ directly from the condition

$$-\log\lambda(\hat{\varphi}_i - \sigma_i^-) = -\log\lambda(\hat{\varphi}_i + \sigma_i^+) = \frac{1}{2}. \tag{7.24}$$

This can yield different values for the lower and upper uncertainty values σ_i^- and σ_i^+, if the variance, and therefore the profile likelihood, is asymmetric.

7.4. Asimov Data Fit

The utilization of Asimov data for the evaluation of the performance of a (maximum likelihood) fit on MC samples, is the subject of this dedicated section due to its importance for this blind analysis.
Formally introduced as *Asimov data* in [76], an artificial data set which reflects the exact expectation of the Monte Carlo simulation can be used to study several aspects of the signal extraction and result calculation of a blind analysis.
In addition to the obvious checks for closure of the fitted branching fractions and ratios thereof, the Asimov data set can also be employed to determine the statistical fit uncertainty for a given fit model. Furthermore, the soundness of the calculation of multiplicative systematic uncertainties can be validated.
The MC study results shown in this text for the ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ are based on Asimov data. Likewise, Asimov data is utilized to validate fit procedures for data-MC agreement tests in sideband regions.
For these purposes, the Asimov sample is always obtained directly from the MC samples which are used to determine the fit templates. No additional scaling or selection is applied. The bin counts of histograms obtained from Asimov samples as artificial data sample input to fit procedure validations are **not** rounded to integer values, as they would be observed in actual data, but used as they are. This means, that Asimov data does exhibit real numbered bin counts, due to scaling and efficiency correction weights.

8. $\mathcal{R}(\mathrm{D}^{(*)})$ on Asimov Data

The properties which this study aims to determine are the ratios of the decay rates of semi-leptonic B meson decays into the heavy τ lepton versus either of the light leptons $\ell = \mathrm{e}, \mu$ for the two cases of a ground state D or exited D^* meson in the final state:

$$\mathcal{R}(\mathrm{D}^{(*)}) = \frac{\mathcal{B}\left(\mathrm{B} \to \mathrm{D}^{(*)}\tau\nu_\tau\right)}{\mathcal{B}\left(\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu_\ell\right)} \qquad \text{with} \quad \ell = \mathrm{e} \text{ or } \mu. \tag{8.1}$$

In this study, these ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ are calculated as ratios of the events observed for the respective decay processes scaled by the associated reconstruction efficiencies, as

$$\mathcal{R}(\mathrm{D}^{(*)}) = \frac{N_{\mathrm{sig}}/\epsilon_{\mathrm{sig}}}{N_{\mathrm{norm}}/\epsilon_{\mathrm{norm}}} = \frac{N_{\mathrm{sig}}}{N_{\mathrm{norm}}}\frac{\epsilon_{\mathrm{norm}}}{\epsilon_{\mathrm{sig}}}, \tag{8.2}$$

where the first definition emphasizes the ratio of two individual branching fractions. In this formula, N_{sig} and N_{norm} denote the number of observed events for the signal decays into τ leptons and the normalization modes with either of the light leptons in the final state, respectively. These observed event numbers are scaled by the respective reconstruction efficiencies ϵ_{sig} and ϵ_{norm} to recover their branching fractions. The efficiencies ϵ are determined from MC simulation according to the definitions given in the dedicated paragraph below. The suffixes *sig* and *norm* denote the true underlying decay process of the event, which is known in MC and used for the determination of the efficiencies and the definition of the templates. These true decay modes, for which the number of events shall be determined on data, are also called components or templates in the context of the fit. Depending on the ratio, the respective suffixes can denote the signal decay processes

$$\mathrm{sig} \coloneqq \begin{cases} \mathrm{B} \to \mathrm{D}\tau\nu & \text{for } \mathcal{R}(\mathrm{D}) \\ \mathrm{B} \to \mathrm{D}^*\tau\nu & \text{for } \mathcal{R}(\mathrm{D}^*) \end{cases} \tag{8.3}$$

and the normalization decay processes

$$\mathrm{norm} \coloneqq \begin{cases} \mathrm{B} \to \mathrm{D}\ell\nu & \text{for } \mathcal{R}(\mathrm{D}) \\ \mathrm{B} \to \mathrm{D}^*\ell\nu & \text{for } \mathcal{R}(\mathrm{D}^*) \end{cases} \tag{8.4}$$

where the signal B meson can be charged or neutral, which also defines the charge of the final state particles.
For the measurement at hand, the number of signal as well as normalization events are determined based on a total of eight reconstruction modes. These event numbers are estimated from template maximum-likelihood fits to distributions of the reconstructed events, for which the electron and muon modes are combined as illustrated in the overview of the reconstruction modes given in Table 8.1. The reconstruction modes are also referred to as channels in the context of the template likelihood fit. The labels used for the reconstruction modes and for the true underlying decay modes can be distinguished by the neutrinos ν, which are included when referring to the true decay, and not included when referring to a reconstruction mode, as the neutrino is not reconstructed.

Table 8.1.: Overview of the reconstruction modes. Charge-conjugated cases are implied. Electron and muon modes are combined for the template fit, reducing the eight modes in which the B_{sig} mesons are reconstructed to four channels which are fitted simultaneously. The four channels depend on whether the B meson or $D^{(*)}$ is charged or neutral, as well as on the excitation of the charmed mesons, i.e. if a D or D^* meson is reconstructed.

	Charged B		Neutral B	
D	$D^0\ e^-$ $D^0\ \mu^-$	$\Big\}\ \mathbf{D^0\ \ell^-}$	$D^+\ e^-$ $D^+\ \mu^-$	$\Big\}\ \mathbf{D^+\ \ell^-}$
D^*	$D^{*0}\ e^-$ $D^{*0}\ \mu^-$	$\Big\}\ \mathbf{D^{*0}\ \ell^-}$	$D^{*+}\ e^-$ $D^{*+}\ \mu^-$	$\Big\}\ \mathbf{D^{*+}\ \ell^-}$

As is apparent from this overview, both light lepton cases are considered. Hence, ℓ denotes here e **and** μ, not either e or μ. This has to be taken into account for the calculation of the ratios $\mathcal{R}(D^{(*)})$. Assuming LFU for the light leptons and their difference in mass to be negligible, this is done by a factor of 1/2 in the denominator of Equation (8.2).
This results in the formula

$$\mathcal{R}(D^{(*)}) = \mathbf{2}\,\frac{N_{\text{sig}}}{N_{\text{norm}}}\frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}} \tag{8.5}$$

for the calculation of the decay rate ratios.
As stated already in Section 5.1, the signal is reconstructed in the same reconstruction modes as the normalization and, as a consequence of this, only the leptonic decay modes of the τ lepton are considered. This fact is taken into account automatically by the reconstruction efficiency ϵ_{sig}.

Reconstruction Efficiencies
The reconstruction efficiencies of the individual components have to incorporate any loss in efficiency due to

- the acceptance of the detector;
- event reconstruction (track finding, misassignment of particle type, ...);
- incomplete coverage of decay processes due to limited number of reconstruction channels;
- imperfections of the B_{tag} reconstruction; and
- the signal-enriching event selection applied for this study.

The last aspect must also include the scope considered for the binning of the fit observables, which will be discussed in the upcoming Section 8.2. The efficiencies for the individual components are taken from MC simulation and calculated via

$$\epsilon_{B\to FS} = \frac{N_{FS}^{MC}}{2\,\mathcal{B}(B\to FS)\;\;\mathcal{B}(\Upsilon(4S)\to B\overline{B})\,N_{B\overline{B}}} \tag{8.6}$$

for a given final state FS, where $\mathcal{B}(\mathrm{B} \to \mathrm{FS})$ denotes the branching fraction for the decay of a charged or neutral B meson into the allowed final state FS. The branching fraction $\mathcal{B}(\Upsilon(4\mathrm{S}) \to \mathrm{B}\overline{\mathrm{B}})$ appearing in the denominator is the decay rate of the $\Upsilon(4\mathrm{S})$ resonance into the required pair of charged or neutral B mesons. It is $(51.4 \pm 0.6)\,\%$ and $(48.6 \pm 0.6)\,\%$, respectively [28], and scales the estimate of the total number of recorded $\mathrm{B}\overline{\mathrm{B}}$ pairs $N_{\mathrm{B}\overline{\mathrm{B}}}$ available in the Belle data sample accordingly. The value of $N_{\mathrm{B}\overline{\mathrm{B}}}$ is given in Section 3.1, Equation (3.1).
The factor of two in the denominator has its origin in the fact that per recorded event each of the two B mesons of the $\Upsilon(4\mathrm{S})$ decay can decay via the process of interest. To be more specific, for the general case of any number of $\mathrm{B}\overline{\mathrm{B}}$ pair events $N_{\mathrm{B\,Pair}}$ containing two (either charged or neutral) B mesons, the expected number of B decays with a given branching fraction $\mathcal{B}_{\mathrm{B}\to X}$ is

$$\begin{aligned} N_{\mathrm{B}\to X} &= \left(2\mathcal{B}_{\mathrm{B}\to X}\,(1-\mathcal{B}_{\mathrm{B}\to X}) + \mathcal{B}^2_{\mathrm{B}\to X}\right)\, N_{\mathrm{B\,Pair}} \\ &= \left(2\mathcal{B}_{\mathrm{B}\to X} - \mathcal{B}^2_{\mathrm{B}\to X}\right)\, N_{\mathrm{B\,Pair}}. \end{aligned} \tag{8.7}$$

The quadratic term in the branching fraction $\mathcal{B}_{\mathrm{B}\to X}$ can often be neglected for processes with small branching fractions. For the study at hand, the branching fractions of some processes are of the order $\mathcal{O}(10^{-2})$, which already leads to sizable deviations from closure if the efficiency is not calculated correctly. However, this relation results in a quadratic equation for the calculation of branching fractions. This would complicate the evaluation of uncertainties. To avoid this, the efficiency must map the number of events after the reconstruction to the number of *candidates* with at least one decay process of interest. This is achieved by dropping the quadratic term in the second line of Equation (8.7), which results in the expression given in Equation (8.6).
If inserted into the formulae for the ratios $\mathcal{R}(\mathrm{D}^{(*)})$ defined in Equation (8.1) or Equation (8.5), the majority of the factors in the efficiency definition 8.6 cancel; e.g. for the case of Equation (8.5) one obtaines

$$\mathcal{R}(\mathrm{D}^{(*)}) = \mathbf{2}\,\frac{N_{\mathrm{sig}}}{N_{\mathrm{norm}}}\,\frac{N^{\mathrm{MC}}_{\mathrm{norm}}\mathcal{B}^{\mathrm{MC}}_{\mathrm{sig}}}{N^{\mathrm{MC}}_{\mathrm{sig}}\mathcal{B}^{\mathrm{MC}}_{\mathrm{norm}}}. \tag{8.8}$$

The values of $N^{\mathrm{MC}}_{\mathrm{sig}}$ and $N^{\mathrm{MC}}_{\mathrm{norm}}$ are taken from the MC samples with all selection requirements, as well as all scaling (Section 3.3) and relevant efficiency correction factors (Chapter 4) applied. The branching fraction values of the signal and normalization processes in the MC samples are denoted by the symbols $\mathcal{B}^{\mathrm{MC}}_{\mathrm{sig}}$ and $\mathcal{B}^{\mathrm{MC}}_{\mathrm{norm}}$, respectively.
In the following sections, the approach used to determine the two remaining quantities N_{sig} and N_{norm} is presented.

8.1. Fit Observables

The number of events remaining after the event selection for the signal and normalization processes are obtained with the help of a template maximum-likelihood fit. As discussed in Chapter 7, the distribution shape of the individual processes is acquired from MC distributions. The normalization parameters of the process templates, i.e. the number of events of the respective process, are free parameters of the fit.
For this approach to function as desired, the templates of the individual components must be distinguishable and well described in MC. The observables selected for this purpose are

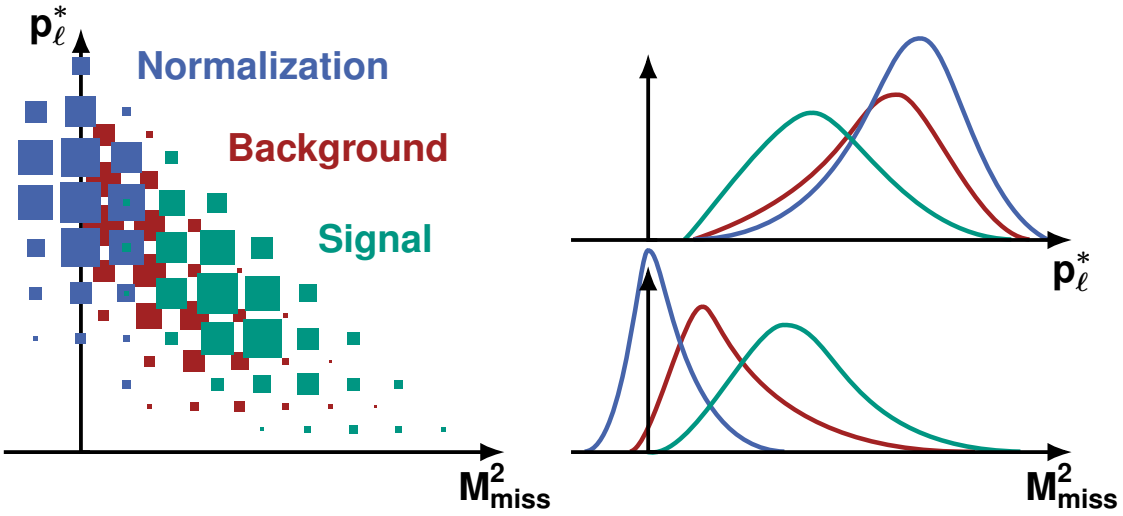


Figure 8.1.: Visualization of the distributions of the $B \to D^{(*)}\tau\nu$ signal and $B \to D^{(*)}\ell\nu$ normalization processes, as well as the main background processes $B \to D^{**}\ell\nu$. The distributions are shown in the 2-D plane spanned by the observables M^2_{miss} and p^*_ℓ on the left. On the right the distributions are visualized in for each of the two observables individually. The shown visualizations are meant to demonstrate the separation power of the chosen observables. They do **not** represent the relative scales of the components, nor are the shapes reproduced perfectly. See Section 8.3 for more accurate plots.

the missing mass squared M^2_{miss} of the event and the momentum p^*_ℓ of the reconstructed light lepton ℓ in the rest frame of the signal side B meson. To boost the lepton momentum into the required reference frame, the negative of the B_{tag} momentum in the center-of-mass reference frame is utilized, as the B_{sig} meson is not fully reconstructed. The missing mass squared is calculated in the center-of-mass reference frame as

$$M^2_{miss} = \Big(p_{Beam} - p_{B_{tag}} - \underbrace{(p_{D^{(*)}} + p_\ell)}_{=p^{visible}_{B_{sig}}}\Big)^2, \tag{8.9}$$

where p_{Beam} is the four-momentum of the e^+e^- system, which is reduced to the beam energy in this frame of reference, and the other four-momenta are associated to the hadronically reconstructed B_{tag} meson and the visible part of the signal side B_{sig} meson. For the calculation of the missing mass squared and p^*_ℓ, the energy component of the B_{tag} four-momentum $p_{B_{tag}}$ in the center-of-mass reference frame is replaced with half of the beam energy.

An illustration of the different shapes for $B \to D^{(*)}\tau\nu$ signal, $B \to D^{(*)}\ell\nu$ normalization and $B \to D^{**}\ell\nu$ background processes in the two-dimensional plane spanned by the observables M^2_{miss} and p^*_ℓ is given in Figure 8.1.

The missing mass squared is chosen as it is expected to be model independent. It represents the invariant mass of the invisible particles of the signal side B meson. In the case of the normalization modes, this is the mass of the respective lepton neutrino ν_ℓ, which is zero. Therefore, the normalization component exhibits a peak at zero in M^2_{miss}. The imperfect resolution of this peak is due to an imperfect reconstruction of the visible particles of the event.

For the signal processes $B \to D^* \tau (\to \ell \bar{\nu}_\ell \nu_\tau) \bar{\nu}_\tau$ the further decay of the τ lepton leads to a total of three invisible neutrinos in the final state. The invariant mass of these three invisible particles is nonzero and results in a broad distribution centered around $M^2_{\text{miss}} \approx 3\,\text{GeV}^2$. Located between these two components is the distribution of the background resulting from semi-leptonic B decays with D^{**} mesons in the final state.
Due to the additional neutrinos in the final state of the signal decay modes, the reconstructed light lepton carries less energy than in the case of the normalization modes. This is particularly apparent in the distribution of the light lepton momentum p^*_ℓ in the signal B meson's rest frame. On average, p^*_ℓ takes on lower values for the signal processes than for the normalization or D^{**} components. This adds further separation power for the different components, when combined with the missing mass squared.

8.2. Fit Procedure

The two observables M^2_{miss} and p^*_ℓ described in Section 8.1 above are used to extract the event counts N for the calculation of the ratios $\mathcal{R}(D^{(*)})$ from two-dimensional template likelihood fits.
The fit is performed simultaneously in the four signal side B meson reconstruction modes $D^0\ell^-$, $D^{*0}\ell^-$, $D^+\ell^-$ and $D^{*+}\ell^-$ listed in Table 8.1.
For each of these channels, templates are defined in the M^2_{miss}–p^*_ℓ space with

- 7 equidistant bins in p^*_ℓ with $p^*_\ell \in [0.1, 2.2]\,\text{GeV}$; and
- 13 bins in M^2_{miss}, given by the bin edges $(-2.0, -1.4, -0.8, -0.2, 0.4, 1.0, 1.6, 2.2, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0)$ in GeV^2, which increase in width for larger M^2_{miss}.

Based on this segmentation of the M^2_{miss}–p^*_ℓ space, templates for the following seven components are defined:

$B \to D\tau\nu$ and $B \to D^*\tau\nu$ Signal Components
The decay processes with leptonic τ decays and either a ground state or an excited $D^{(*)}$ meson in the final state are the signal processes in the numerator of the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$, respectively. For the cases with ground state D mesons, these processes are reconstructed mainly in the matching $B \to D\ell$ channel, with only few such decays resulting in signatures in the $B \to D^*\ell$ reconstruction channel as so-called feed-up. This can occur due to random combinations of D meson candidates and slow pions or photons, which happen to reproduce the D^* meson mass when combined.

A similar, but more pronounced effect is observed for true decay processes of the $B \to D^*\tau\nu$ component, for which the slow pion or photon was misassigned during the reconstruction, or escaped detection. This case, which was already introduced as $D^* \to D$ feed-down for the processes with light leptons previously in this text, is more likely and causes a significant fraction of the $B \to D^*\tau\nu$ component to be reconstructed in the $B \to D\ell$ channel.

For the signal components $B \to D^{(*)}\tau\nu$, events which are correctly reconstructed and end up in the matching channels cannot be distinguished from their cross-feed counterparts, as the distribution in M^2_{miss} is broad and of similar shape for both cases.

For the purpose of this fit, the probabilities for the two cross-feed cases are bound to the correctly reconstructed processes by the frequency with which they are observed in

MC. This constraint is included in the fit in the form of the reconstruction efficiencies calculated as described in the beginning of this chapter.

$B \to D\ell\nu$ and $B \to D^*\ell\nu$ Normalization Components

The two normalization components with light leptons in the final state do behave similar to the signal components, in the sense that they are also subject to cross-feed. However, for the normalization modes this effect does cause distinguishable signatures in the M^2_{miss} distribution. As the normalization modes produce well differentiable peaks at $M^2_{\text{miss}} = 0\,\text{GeV}^2$, the lack or abundance of the slow pion or photon related to the excited D^* meson causes a shift in this observable, which allows for the separation of feed-down or feed-up from the correctly interpreted events.

At the time of writing, these cross-feed components are linked to the components of the correctly recombined normalization processes and are treated as one. The fraction of correctly interpreted events and the respective cross-feed is given again by the relations observed in MC for the processes. For both the signal and the normalization components, this can be justified by the fact that all other efficiencies are also determined on MC.

However, implementing the cross-feeds of the normalization components as independent components allows for a better validation of the fit model and therefore the measurement as a whole. Hence, this aspect might be subject to change as the analysis is finalized. An evaluation of this issue based on sideband data is presented in Section 10.2.

$B \to D^{**}\ell\nu$ Background Component

The semi-leptonic B meson decay into orbitally excited D^{**} mesons constitute an important background component, as they produce broad signatures in M^2_{miss} and p^*_ℓ which lie in the same region as the signal processes. This is especially problematic, as these decays are not well studied and thus might not be described well in the simulation. Hence, the $B \to D^{**}\ell\nu$ background component, which comprises processes with the two possible D^{**} decays

- $B \to D^{**}(\to D^{(*)}\pi^0)\ell\nu$
- $B \to D^{**}(\to D^{(*)}\pi^\pm)\ell\nu$

is considered as a separate component in the fit, the normalization of which is handled as a free parameter. This does not include the semi-leptonic B meson decays into a D^{**} meson and a τ lepton, nor the processes in which the D^{**} decays further into a $D^{(*)}$ meson and particles other than a single charged or neutral pion.

As the processes defining this template are not well studied, additional evaluations of their description in MC are undertaken using dedicated reconstruction modes (see Section 9.3). For the Asimov study shown in this chapter, these checks are not of importance. For future improvements to the fit setup for the measurement on data, these sideband studies can be used to constrain this component.

$B\overline{B}$ Background Component

This fit component comprises all other decay processes originating from $B\overline{B}$ events, which are not covered by the aforementioned components. This includes the remaining contributions from semi-leptonic B meson decays involving a D^{**} meson, candidates from processes where a hadron is falsely identified as the light lepton ℓ of the signal side B decay (fake-lepton), as well as combinatorial background. Bundled into one component, these processes are all defined by the combined template shape observed in MC. The normalization of the component is a free parameter in the fit, which allows the fit to adapt to variations in efficiency for the entirety of these processes, but not for relative

differences in reconstruction efficiency between the combined processes. Such effects must be covered by shape uncertainties and can be considered as nuisances in the fit. For the presented Asimov study, these effects are not of relevance, as the Asimov data exactly reflects the MC expectation. They must be studied in sideband samples and with the help of toy studies, which have not been concluded at the time of writing.

Continuum Background Component

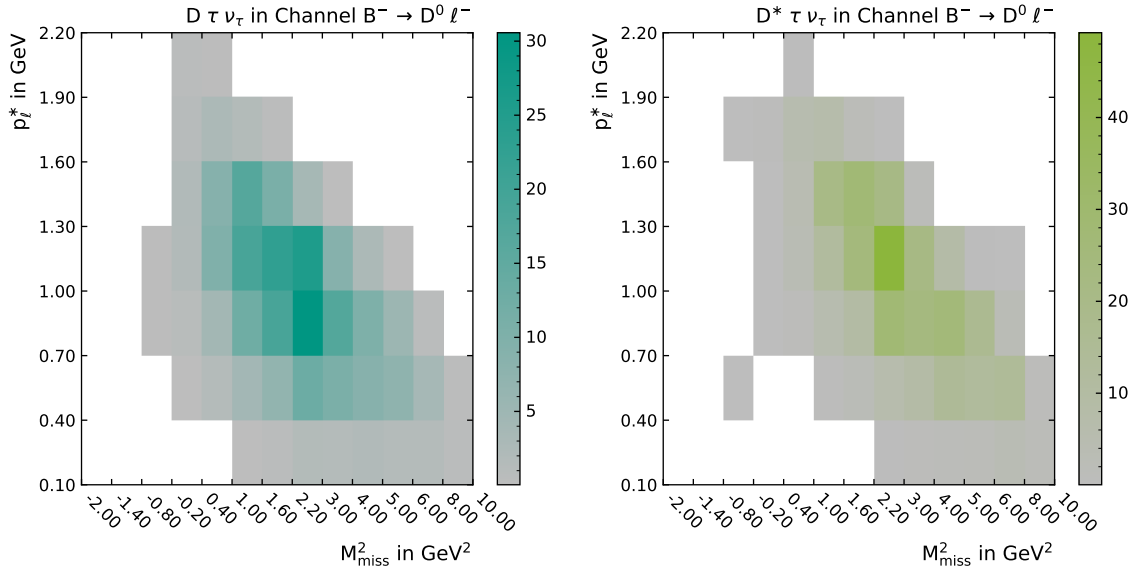
A validation of the MC description of the shape of the continuum background component is given in Section 9.4. The results of this data-MC validation in a dedicated control sample show that the continuum shape is described well in the simulation.

After the event selection process presented in the previous Chapters 5 and 6, however, few candidates originating from $e^+e^- \to q\bar{q}$ continuum processes with $q = u, d, s, c$ are remaining in the obtained data samples. Their contribution in the two-dimensional fit observable space is broad and shaped similar to the one of the signal processes. This becomes apparent in the visualization of the templates in the next Section 8.3. Due to the low statistics for the candidates of this component, the respective fit template might not be able to generalize the shape of these processes sufficiently enough for an application on recorded data. Thus, this component will be combined with the $B\overline{B}$ background component in the future. At the time of writing, however, it is handled as an independent component, the normalization of which is a free parameter in the fit.

The seven normalization parameters of these components are fitted simultaneously in all four reconstruction modes. The contribution of a component to each of the four reconstruction modes is governed by the rates observed in MC. Contributions from $D^* \to D$ feed-down in the ground state D meson reconstruction modes $D^0\ell^-$ and $D^+\ell^-$ are attributed to the $B \to D^*\ell\nu$ component. The template shapes resulting from these component definitions are shown and discussed in the next Section 8.3.
Based on the obtained templates, the negative log-likelihood function is calculated as defined in Section 7.2 and minimized simultaneously for all channels with the help of the MINUIT algorithm [77, 78], yielding the estimates for the number of observed events N of each component, as well as the associated covariance matrix. The propagation of the parameter uncertainties to the uncertainties on the ratios $\mathcal{R}(D^{(*)})$ and the inclusion of systematic uncertainties is discussed in Section 8.4. Results for the application of this fit procedure on an Asimov data set, assuming the SM expectation values for the ratios $\mathcal{R}(D^{(*)})$ (see Table 1.1), are presented in Section 8.5.

8.3. Templates

This section is devoted to the templates resulting from the component definitions presented in the previous Section 8.2. They are shown in Figure 8.2 for the reconstruction mode $B^- \to D^0\ell^-$. Visualizations of the templates obtained for the three remaining channels $D^+\ell^-$, $D^{*0}\ell^-$ and $D^{*+}\ell^-$ are available in Figures C.5 to C.7 in Appendix C.2. In these plots, the color scheme for the different templates matches the colors used in the projections of the fitted distributions shown in Figure 8.3 of the upcoming Section 8.5, except for the $B\overline{B}$ and continuum background components. The latter are shown in different colors in this section for a better illustration.
Instead of showing the normalized distributions, which are the actual templates, the underlying distributions of the different components are shown unnormalized. This allows for the comparison of the relative frequencies of the different components in addition to the shape differences.



(a) $B \to D\tau\nu$ template in $B^- \to D^0\ell^-$ mode

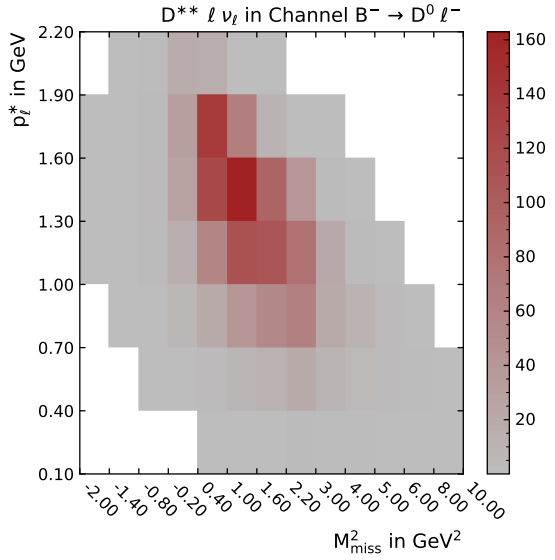
(b) $B \to D^*\tau\nu$ template in $B^- \to D^0\ell^-$ mode

(c) $B \to D\ell\nu$ template in $B^- \to D^0\ell^-$ mode

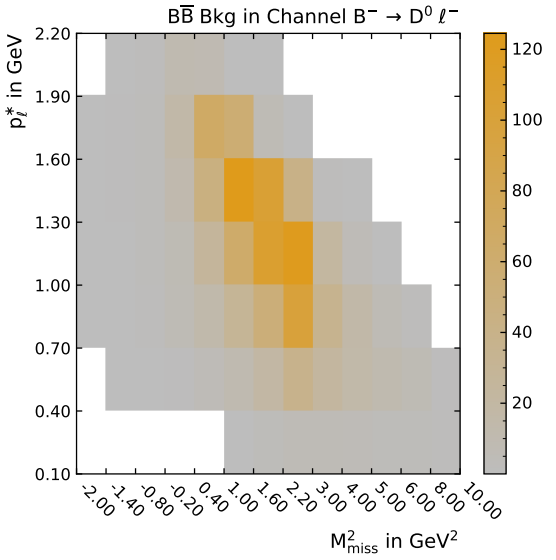
(d) $B \to D^*\ell\nu$ template in $B^- \to D^0\ell^-$ mode

Figure 8.2.: Histogrammed distributions used for the templates for the reconstruction channel $B^- \to D^0\ell^-$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts. Distributions of the three background components for this channel are shown in the continuation of this figure on page 105.

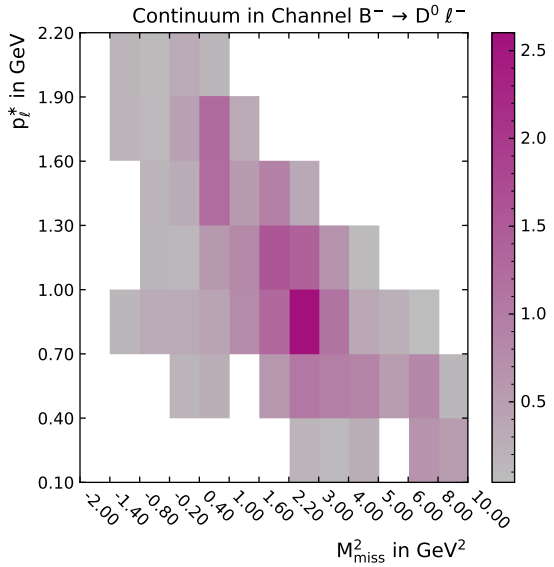
From these visualizations, the properties of the seven components as given in their description in the previous Section 8.2 become apparent. The normalization processes show sharp peaks in the M^2_{miss} dimension, if reconstructed in the correct channel. Cross-feed between



(e) $B \to D^{**}\ell\nu$ template in $B^- \to D^0\ell^-$ mode



(f) $B\overline{B}$ background template in $B^- \to D^0\ell^-$ mode



(g) Continuum background template in $B^- \to D^0\ell^-$ mode

Figure 8.2.: Continuation of Figure 8.2 on page 104.

ground state D and excited D^* meson channels is visible for both signal and normalization processes. In particular, the magnitude of the $D^* \to D$ feed-down contribution is notable when comparing the bin counts in Figures 8.2c and 8.2d. The same effect is visible in Figures C.5c and C.5d for the neutral B reconstruction channels. Equivalent contributions of $D \to D^*$ feed-up are far less likely, which is reflected in the bin counts of the distributions shown in e.g. Figures C.6c and C.6d.

The same cross-feed behavior is observed for the signal processes if Figures 8.2b and 8.2c and Figures C.6b and C.6c are compared in regard to the event counts. These histograms also show the much broader distributions of the signal components which also exhibit on

average larger M^2_{miss} and lower p^*_ℓ values than the normalization components.
The features of the $B \to D^{**}\ell\nu$ and $B\overline{B}$ background components are fairly similar, with a smaller variance and average in the M^2_{miss} dimension than the signal components. The distributions used for their templates for the $D^0\ell^-$ reconstruction mode are shown in Figures 8.2e and 8.2f, respectively. The low statistics for the continuum background component are visible for instance in the low event counts of the histogram plotted in Figure 8.2g. Due to the randomness of the particle combinations resulting in valid signal candidates from continuum processes, the distribution of these candidates is very broad, covering almost the entire considered scope in M^2_{miss} and p^*_ℓ.

8.4. Uncertainty Calculation

The uncertainty on the calculated decay rate ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are determined from the uncertainties on the fitted event counts, as well as the uncertainties on the required efficiency. The former comprise the so-called statistical uncertainty.
The latter are due to systematic uncertainties and, therefore, referred to as such. However, one has to differentiate between such systematic effects which affect the overall normalization of the determined values, and other systematic effects which result in changes in the shape of different components. Systematic uncertainties causing a change in the shape of a distribution are additive systematic uncertainties. They can be considered as part of the likelihood function as discussed in Section 7.3 or by the evaluation of the fit procedure for different variations of the respective systematic uncertainty. However, as the fit discussed in this section is performed on an Asimov data set, no such effects are expected, and thus this type of uncertainty is not considered here. The systematic effects that affect the overall normalization are multiplicative uncertainties. They can also be incorporated into the likelihood function of the fit as described in Section 7.3 to allow for changes in the relative rates of the different fit components. Furthermore, the effect of these uncertainties has to be considered for the calculation of the efficiencies of the different contributions to the ratios $\mathcal{R}(D^{(*)})$. This last contribution is taken into account for the presented Asimov study, while the inclusion into the likelihood function will not affect the result on the Asimov data set, and is thus not included.
In the following two paragraphs the calculation of the statistical and multiplicative systematic uncertainties is described.

Statistical Uncertainties
The covariance matrix V of the fitted parameters, which is obtained from the minimization algorithm in form of the inverse Hesse matrix, as described in Section 7.3, contains the statistical uncertainties on the fitted parameters, namely the number of observed events per fit component. To propagate these uncertainties to the ratios $\mathcal{R}(D^{(*)})$, the Jacobian matrix J is calculated according to

$$J_{ij} = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial x_j}, \tag{8.10}$$

where $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is a function that maps the n-dimensional vector $\boldsymbol{x}$ onto an m-dimensional space.
For the problem at hand, where we consider the two ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$, which depend on the four parameters $N_{\text{sig}}, N^*_{\text{sig}}, N_{\text{norm}}$ and N^*_{norm}, $\boldsymbol{f}$ is given by the ratios and $\boldsymbol{x}$ contains the event counts, and thus the Jacobian is given by

$$J_{ij} = \begin{cases} \frac{1}{N_n} & i = j \neq n \\ -\frac{N_i}{N_n^2} & i \neq j = n \\ 0 & \text{else,} \end{cases} \tag{8.11}$$

where n denotes any of the event count estimates N appearing in the denominator normalization modes. Using the transformation

$$U = JVJ^T, \tag{8.12}$$

the covariance matrix U for the ratios is calculated.

Multiplicative Systematic Uncertainties

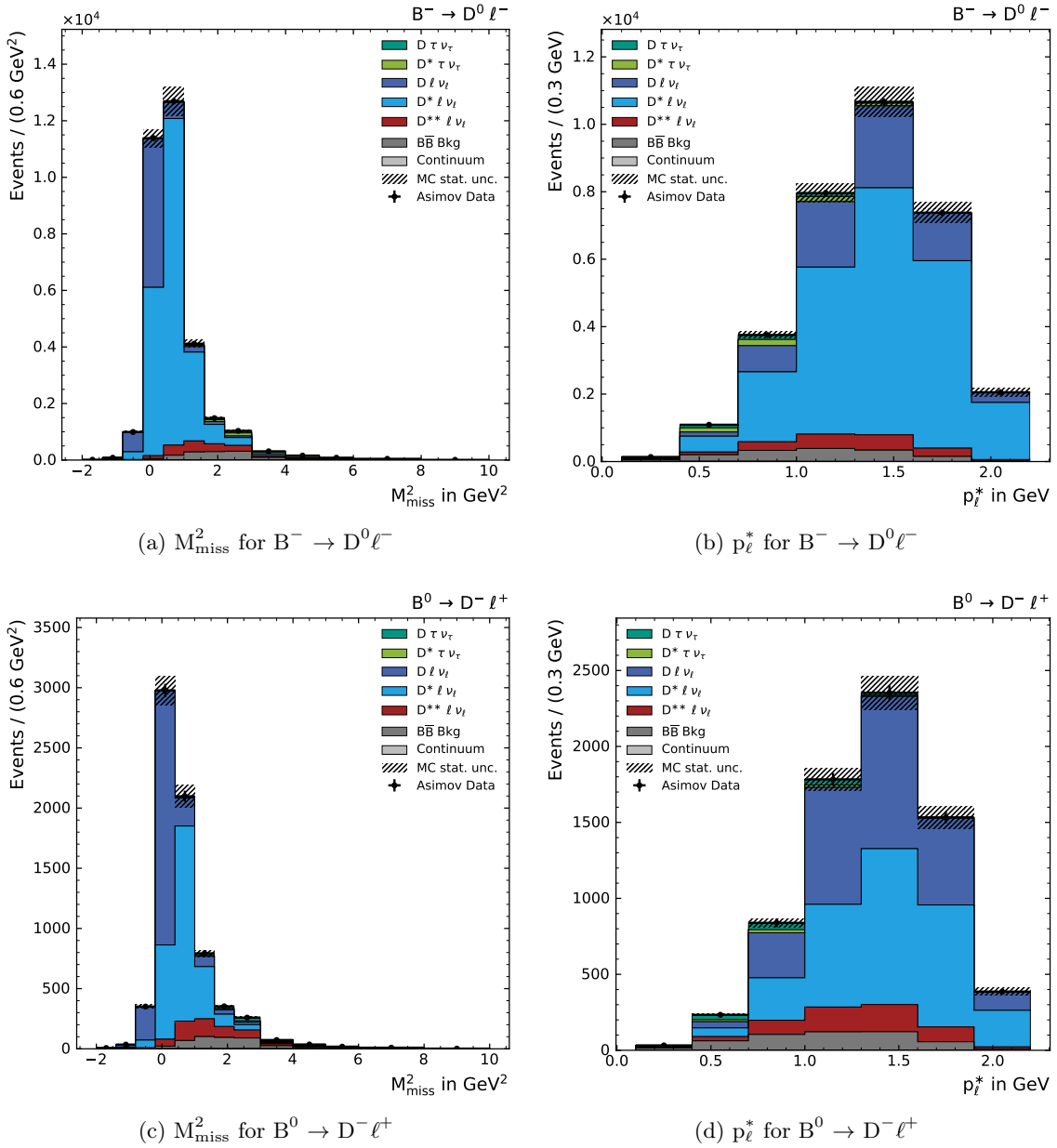
For the multiplicative uncertainties the following systematic effects are considered:

- limited MC statistics for the efficiency calculation;
- data-MC efficiency differences for lepton particle identification;
- data-MC efficiency differences for hadron particle identification ($\pi^{\pm}$, $\mathrm{K}^{\pm}$);
- data-MC differences for hadron particle misidentification rate ($\pi^{\pm}$, $\mathrm{K}^{\pm}$);
- data-MC efficiency differences for slow pion reconstruction;
- data-MC efficiency differences for K^0_S reconstruction;
- uncertainties of $\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu$ form factors.

The limited MC statistics, which affect the uncertainty on the efficiency calculation, are treated as independent uncertainties for each component. The covariance matrix representing this source of uncertainty is thus defined by a diagonal matrix with Poisson uncertainties given by the unweighted MC event count for each component.
For each of the remaining sources of systematic uncertainty, multiple variations of the candidate weights are produced as described in the respective parts of Chapter 4. These weight variations are used to determine the correlation of the uncertainties for the MC event numbers $N_\mathrm{sig}^\mathrm{MC}$ and $N_\mathrm{norm}^\mathrm{MC}$ required for the efficiency calculation as defined in Equation (8.8) in the form of covariance matrices.
For both statistical and multiplicative uncertainties, the correlation matrices can be included in the calculation of ratios, e.g. with the help of [79], to propagate the uncertainty to the ratios. To obtain the uncertainty caused by each individual uncertainty, only the respective covariance matrix is considered. For the total uncertainty, the sum of all covariance matrices is used.

8.5. Asimov-Data Fit

The fit procedure defined in the previous sections of this chapter is now applied on an Asimov data set, which reflects the properties of the underlying MC sample. Hence, the values of $\mathcal{R}(\mathrm{D}^{(*)})$ as used for the scaling of the MC sample are expected as a result of this fit. These are the SM expectation values (Table 1.1). The size of the Asimov sample



(a) M^2_{miss} for $B^- \to D^0\ell^-$

(b) p^*_ℓ for $B^- \to D^0\ell^-$

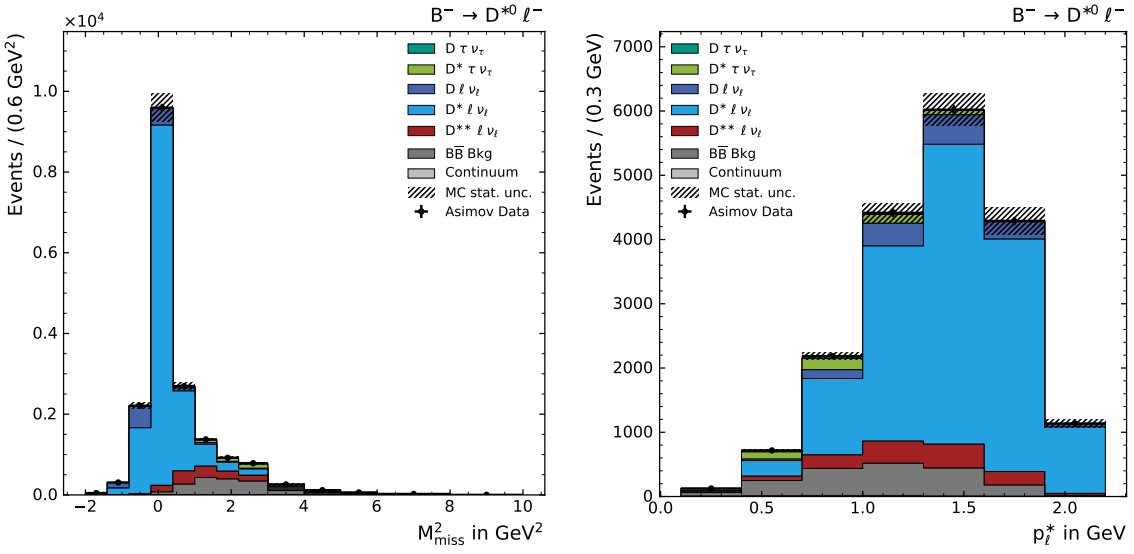
(c) M^2_{miss} for $B^0 \to D^-\ell^+$

(d) p^*_ℓ for $B^0 \to D^-\ell^+$

Figure 8.3.: Projections of the post-fit distributions resulting from the fit on Asimov data. The uncertainties on the MC distribution are indicated as diagonally hatched bands. The distributions in M^2_{miss} and p^*_ℓ are shown on the left and right, respectively. The first row shows the result in the $B^- \to D^0\ell^-$ reconstruction mode and the second row for the $B^0 \to D^-\ell^+$ mode. The results for the remaining two reconstruction modes are shown in the continuation of this figure on page 109.

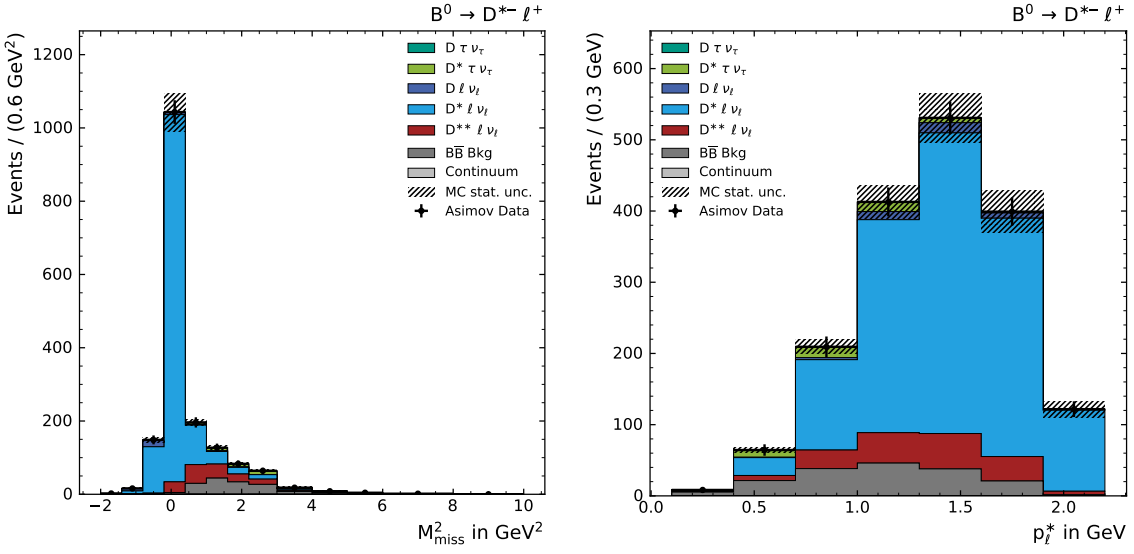
represents roughly the size of the recorded Belle data sample, except for the effect of the FEI reconstruction efficiency difference for recorded data and MC.

As briefly discussed in Section 7.4, this Asimov data fit allows for the validation of the fit procedure regarding closure and to obtain estimates for the expected statistical and the above discussed systematic uncertainties.



(e) M^2_{miss} for $B^- \to D^{*0}\ell^-$

(f) p^*_ℓ for $B^- \to D^{*0}\ell^-$



(g) M^2_{miss} for $B^0 \to D^{*-}\ell^+$

(h) p^*_ℓ for $B^0 \to D^{*-}\ell^+$

Figure 8.3.: Continuation of Figure 8.3 on page 108. The first row shows the result in the $B^- \to D^{*0}\ell^-$ reconstruction mode and the second row for the $B^0 \to D^{*-}\ell^+$ mode.

The projections of the fitted distributions for this Asimov fit are shown in Figure 8.3. These do not yield any surprises, as the artificial data points of the Asimov data set lie exactly on the histograms of the MC expectation. The shown statistical uncertainties on the Asimov data points are smaller than the bin uncertainties of the MC distributions. This is due to the fact that the Asimov data should represent the statistics of the recorded Belle data, while the MC distributions are taken from samples with a multiple of this recorded data samples' size, which are then scaled down as described in Chapter 3.

The resulting values for the decay rate ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ are summarized in Table 8.2. For both cases the obtained values match the SM expectation values, which are used as

input for the definition of the MC samples — closure is achieved.
The uncertainty on the results is dominated by the statistical contribution. This is also the case for the previous, hadronically tagged measurements produced by *BABAR* [9, 2013] and Belle II [12, 2015]. The largest contribution to the considered systematic uncertainties have their origin in the MC statistics and the lepton particle identification efficiency. The form factor variations for the $B \to D\ell\nu$ and $B \to D^*\ell\nu$ modeling affect only the uncertainty of the corresponding ratio. Slow pion efficiency corrections are more notable in $\mathcal{R}(D^*)$, where slow pions are used for the reconstruction of the D^* meson. The majority of the systematic uncertainties originating from particle identification and reconstruction efficiency correction, with the exemption of the lepton identification, are of negligible magnitude.
Bearing in mind, that all background components are currently unconstrained, the result of this Asimov fit study shows promising results. Further constraints from e.g. the D^{**} sideband sample and the inclusion of the continuum background component into the $B\overline{B}$ background component will improve the uncertainties.
Further steps towards the readiness of the presented study towards the fit on data are discussed in the outlook given in Chapter 11. Results of Asimov fit studies based on the old gap description (see Section 3.2.3), as well as with a looser selection requirement on the selection BDT output $\mathcal{P}_{\text{Sel}}$ are included in Appendix C.1.

Table 8.2.: Tabulated results for the branching fraction ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ obtained from a fit to an artificial Asimov data set for which the SM expectation values (see Table 1.1) of the ratios are assumed. The fit reproduces these input values, which indicates that the procedure works as intended. Additionally, the statistical fit uncertainty, as well as the combined multiplicative systematic uncertainties are listed as relative to the nominal results. For the latter, the individual contribution of the various sources of multiplicative systematic uncertainties discussed in Section 8.4 are given as well. The relative uncertainty values are listed with a higher number of decimal places than required, to make also the negligible contributions visible for the purpose of this Asimov study.

	$\mathcal{R}(D)$	$\mathcal{R}(D^*)$
Nominal Value	**0.299**	**0.258**
Stat. Uncertainty	**20.666 %**	**12.202 %**
Mult. Sys. Uncertainty	**1.525 %**	**1.207 %**
MC Statistics	1.299 %	0.916 %
Lepton ID	0.783 %	0.768 %
Pion ID	0.027 %	0.032 %
Fake Pion	0.005 %	0.008 %
Kaon ID	0.038 %	0.040 %
Kaon Fake	0.034 %	0.015 %
Slow Pion	0.010 %	0.028 %
K^0_S Reconstruction	0.005 %	0.002 %
$B \to D\ell\nu$ Form Factors	0.146 %	0.000 %
$B \to D^*\ell\nu$ Form Factors	0.000 %	0.155 %

9. Data-MC Comparison

With the procedure for the event reconstruction, the signal selection, as well as the extraction of the decay rate ratios $\mathcal{R}(\mathrm{D}^{(*)})$ defined and verified on MC, the focus is set on the application of the procedure on recorded data. The first step in this direction is the validation of the MC description of the individual components' distributions by comparing them to what is observed in recorded data. For this purpose, the sideband samples defined in Section 6.3.6 are utilized. These samples are defined such that they contain a negligible amount of $\mathrm{B} \to \mathrm{D}^{(*)}\tau\nu$ signal events, and therefore allow for the evaluation of the recorded data. With these samples, the distributions of the non-signal components are validated in regard to their data-MC agreement, where data refers to recorded data.
The data-MC comparisons are performed for observables which are deemed as important due to their relevance for the event selection and signal extraction. Samples of recorded data with a suitable sideband selection applied, as well as the matching MC samples are used to compare the histogrammed distributions of such observables. If not stated otherwise, this is done in the four main reconstruction channels $\mathrm{D}^0\ell^-$, $\mathrm{D}^{*0}\ell^-$, $\mathrm{D}^+\ell^-$ and $\mathrm{D}^{*+}\ell^-$, which are used for the determination of the ratios $\mathcal{R}(\mathrm{D}^{(*)})$ described in the previous Chapter 8, as well as for the combination of all of these reconstruction modes.
As the overall normalization of the MC samples is not necessarily correct, these samples are scaled with a flat normalization factor such that they match the number of events observed on data. This is done only for the purpose of data-MC agreement studies shown in this chapter. The scaling factors are calculated for each reconstruction mode individually. For the comparison plots using the combination of all reconstruction modes, the relative scaling for each mode is obtained and applied separately such that the overall normalization is unchanged. An additional overall normalization factor is then applied to the combination of all reconstruction modes. The used MC scaling factors are included in each comparison plot in the form MC $\times S$, where S is the scaling factor.
To validate the agreement between the distributions observed for recorded data and MC, the pulls of the bin counts are calculated as

$$\mathrm{Pull} = \frac{N_{\mathrm{Data}} - N_{\mathrm{MC}}}{\sigma_{\mathrm{stat.}}^{\mathrm{Data-MC}}} \tag{9.1}$$

and plotted below the histogrammed distribution. Here N_{Data} and N_{MC} refer to the bin counts of the recorded data and MC distribution, respectively. The denominator is the uncertainty on the difference of these two values. The same uncertainty is also used for the $1\,\sigma$ error bars of these pull values.
As an additional means to quantify the agreement of the two distributions, the p-value of a toy-based χ^2 test is calculated. The χ^2 value, the number of degrees of freedom n_{dof}, and the p-value of this test are included in each data-MC comparison plot.
For some comparison plots, the Pearson correlation of the plotted distributions and the distribution of a second, not plotted observable is included for recorded data and for MC. This second observable to which the correlation is calculated is either the output of the selection classifier $\mathcal{P}_{\mathrm{Sel}}$ or the output of the continuum suppression classifier $\mathcal{P}_{\mathrm{CS}}$. As the different contributions to the MC distributions are weighted, the weighted Pearson correlation is calculated for this case.

Using this approach, the data-MC agreement is tested for the following features:

- continuum suppression classifier output $\mathcal{P}_{\text{CS}}$ in the q^2-Sideband in Section 9.1;
- selection classifier output $\mathcal{P}_{\text{Sel}}$ in the q^2-Sideband in Section 9.2;
- $B \to D^{**}\ell\nu$ background description in $B \to D^{**}\ell$ reconstruction modes in Section 9.3;
- continuum background modeling in the off-resonance sideband in Section 9.4;
- fit observables M^2_{miss} and p^*_ℓ in the q^2-Sideband in Section 9.5.

This list comprises a few selected data-MC agreement validations. Due to the number of figures associated to these data-MC comparisons, the majority of them are moved to the Appendix D.2 to avoid extended interruptions of the text. Many more detailed studies are performed to verify the MC description of important observables, e.g. observables which are used as input features to the classifier trainings described in Section 6.1 and Section 6.3.1. All used observables show no signs of significant mismodeling and are therefore not included in this text.

9.1. Continuum Suppression BDT Output Validation

The first observable to be validated is the output of the continuum suppression classifier $\mathcal{P}_{\text{CS}}$. This validation is important, as the observable is produced by a BDT trained on

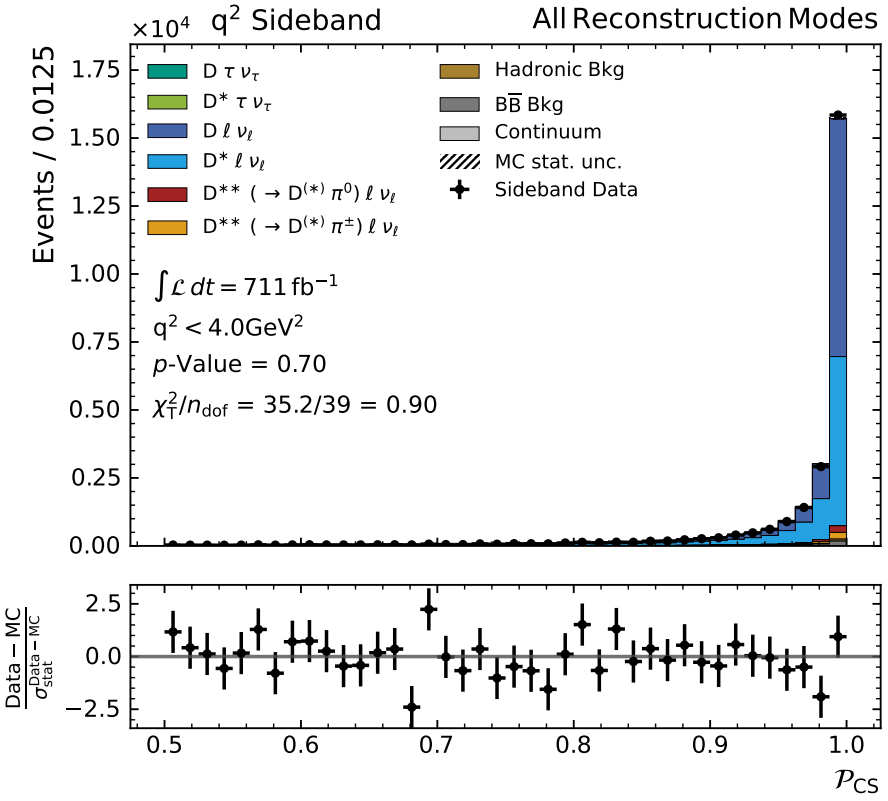


Figure 9.1.: Data-MC comparison of the continuum classifier output $\mathcal{P}_{\text{CS}}$ for all main reconstruction modes in the q^2-Sideband samples. Please refer to the introduction of Chapter 9 for a detailed description of the included information.

simulated events. As such, it might be sensitive to differences in shape or correlation of the used input features. Moreover, the observable $\mathcal{P}_{\text{CS}}$ is also used as input to the selection classifier, which could enhance discrepancies further.

Firstly, the data-MC agreement is validated in the q^2-Sideband. The comparison is shown in Figure 9.1 for the combination of all main reconstruction modes and exhibits good agreement between recorded data and simulation. This is indicated by a p-value of 0.90 and no notable tendencies in the pulls. The results for the comparison in the individual reconstruction channels are shown in Figure D.8 in Appendix D.1. These comparisons also indicate a good agreement for the neutral B^0 reconstruction modes, whereas the data-MC agreement is slightly worse for the charged $\text{B}^\pm$ reconstruction modes. Nonetheless, no channel shows a noteworthy disagreement between the distribution of recorded data and MC.

An additional check of the data-MC agreement in this observable is conducted based on a q^2-Sideband sample with a loose selection requirement of $\mathcal{P}_{\text{CS}} > 0.2$ and a purely random best candidate selection. The results of this test for each of the main reconstruction modes are shown in Figure D.9. For this evaluation based on a looser selection, no obvious data-MC disagreement is observed as well.

As the continuum suppression output $\mathcal{P}_{\text{CS}}$ is used as a feature for the selection classifier, the effect of requirements on $\mathcal{P}_{\text{CS}}$ are studied to evaluate their impact on the fit observables. This is shown for the case of the fit observable p_ℓ^* in Figure D.10 for increasingly tighter requirements on $\mathcal{P}_{\text{CS}}$ from $\mathcal{P}_{\text{CS}} > 0.2$ up to $\mathcal{P}_{\text{CS}} > 0.99$. For this check based on the q^2-Sideband sample, no significant change in the shape of p_ℓ^* is observed. This indicates, that the observable $\mathcal{P}_{\text{CS}}$ behaves similarly for recorded data and simulation. The same evaluation was also conducted for the fit observable M^2_{miss}. The result of this study is not included, as the missing mass squared will be the subject of upcoming discussions in Section 9.5 and Chapter 10.

As a last validation of the continuum suppression classifier output $\mathcal{P}_{\text{CS}}$, the distribution of the observable is studied for increasingly tight cuts on the selection classifier output $\mathcal{P}_{\text{Sel}}$. The results of this study are displayed in Figure D.11 in Appendix D.1. The chosen values for the scan in $\mathcal{P}_{\text{Sel}}$ include the optimized selection requirements defined in Equation (6.22). This test also yields no sign of data-MC disagreement. For high $\mathcal{P}_{\text{Sel}}$ selection requirement values, only the last bins contain events, such that for very tight cuts, only one bin remains. In this case, the normalization of the MC samples hides any disagreement between the data and MC distributions.

9.2. Selection Classifier Output Validation

The selection classifier output $\mathcal{P}_{\text{Sel}}$ is of similar importance to the outcome of this analysis as the continuum suppression classifier output $\mathcal{P}_{\text{CS}}$. It is evaluated by studying the effect that increasingly tight selection criteria on $\mathcal{P}_{\text{Sel}}$ cause in the fit observables. This is done in the same manner as described in the previous Section 9.1, using values between 0.5 and 0.95 for the selection requirement on $\mathcal{P}_{\text{Sel}}$. The values of the optimized selection criterion on $\mathcal{P}_{\text{Sel}}$ defined in Equation (6.22) are also included for this evaluation. For this data-MC comparison, the q^2-Sideband is used.

The results of this validation of $\mathcal{P}_{\text{Sel}}$ for the fit observable p_ℓ^* are shown in Figure D.12 in Appendix D.2. The shape of the p_ℓ^* distribution behaves identically for recorded data and MC for each of the evaluated requirements on $\mathcal{P}_{\text{Sel}}$. For each working point of $\mathcal{P}_{\text{Sel}}$, the p-value, as well as the pulls confirm the visually observed agreement of the distributions.

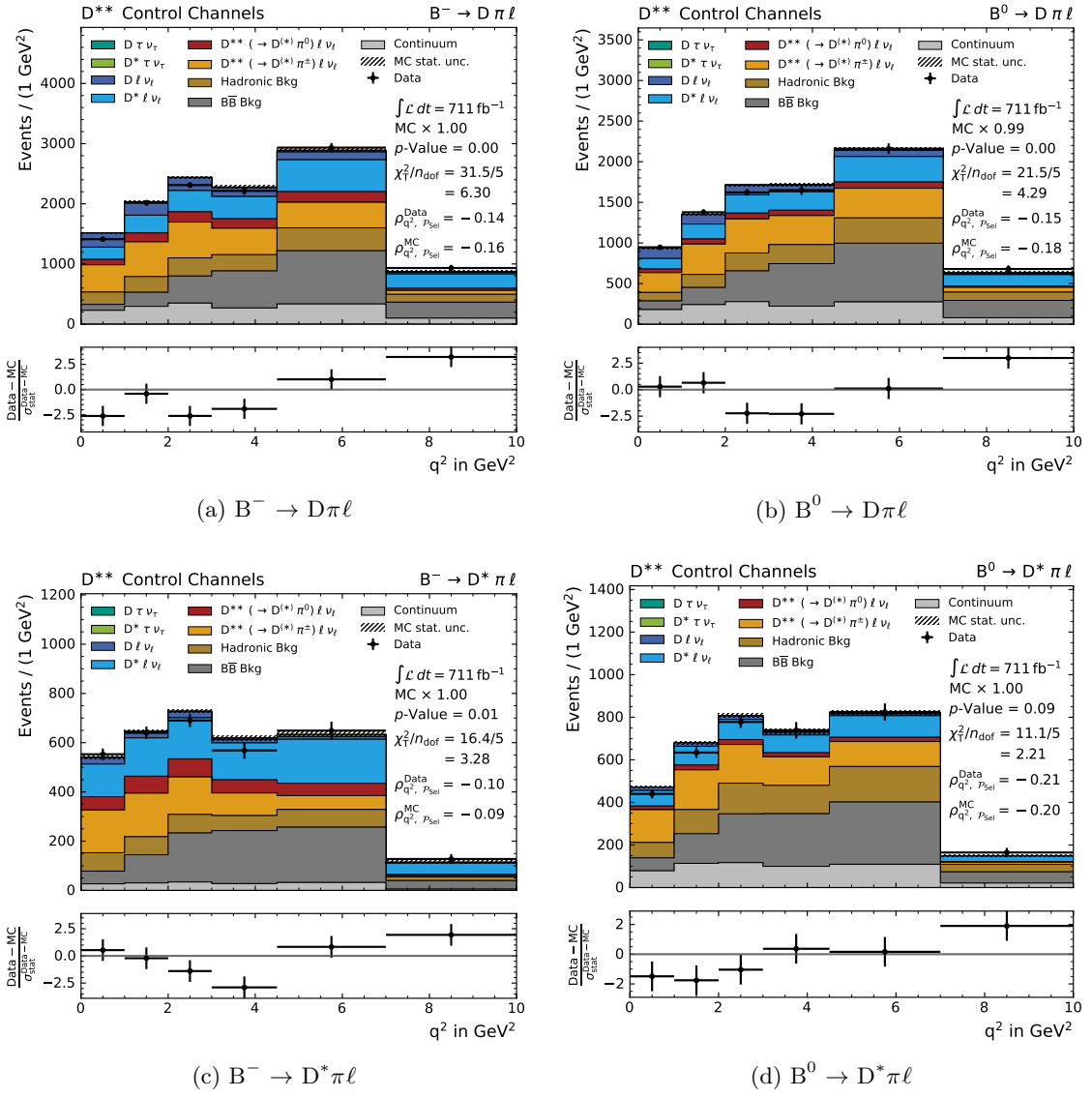
Hence, the fit observable p_ℓ^* is well described by the simulation and neither the selection classifier, nor the continuum classifier disturb this agreement to recorded data.
The same procedure is applied for the fit observable M^2_{miss} and the results are shown in Figure D.13 in Appendix D.2. This data-MC comparison shows a significant disagreement between the shape of M^2_{miss} in recorded data and simulation. Independent of the selection requirement on $\mathcal{P}_{\text{Sel}}$, a distinct pattern is observed in the pulls of the bin counts. The effect occurs at $M^2_{\text{miss}} \approx 0.0\,\text{GeV}^2$. It is attributed to a mismodeling of the resolution of the missing mass squared observable in MC. The plots shown in Figure D.13 are actually produced using a crude method which should already resolve this effect by smearing the M^2_{miss} distribution in MC. In Section 9.5 the pure resolution effect is shown without any method to resolve it applied. A more sophisticated ansatz to resolve this issue than the method used for the plots shown in this section is discussed in Chapter 10. Further findings which might explain the origin of this resolution effect in M^2_{miss} are included in Section 11.1.

9.3. $B \to D^{**}\ell\nu$ Background Model Validation

The $B \to D^{**}\ell\nu$ processes pose the most significant background to this study. They produce signatures similar to the signal processes $B \to D^{(*)}\tau\nu$ in the main reconstruction modes, due to the missing slow pion of the $D^{(*)}$ decay. In addition to this, the knowledge about them is limited, beginning with the branching fractions of the decay processes into the individual orbitally excited $D^{(*)}$ mesons — the normalization of the processes — up to the form factor parametrization of the processes (see Section 2.5), which affects the shape of the distributions in the momentum transfer q^2.
To validate the MC description of these processes, dedicated reconstruction modes are added to the analysis (see Section 5.1.4). A sideband sample based on these reconstruction modes, defined in Section 6.3.6, does not include events originating from the signal processes $B \to D^{(*)}\tau\nu$. This D^{**}-sideband therefore allows one to validate the $D^{(*)}$ background.
To enhance the signal-to-noise ratio for the contribution of the $B \to D^{**}\ell\nu$ processes further in this sample, an additional selection requirement on the invariant mass of the $D^{(*)}\pi$ pair is imposed. This requirement is based on the difference between the invariant mass the $D^{(*)}\pi$ pair, which represents the D^{**}, and the invariant mass of the single $D^{(*)}$ meson:

$$M_{D^{(*)}\pi} - M_{D^{(*)}} > 0.2\,\text{GeV}. \tag{9.2}$$

This exploits the significantly higher masses of the orbitally excited D^{**} mesons.
With this sample, the MC description of the $B \to D^{**}\ell\nu$ processes is validated by comparing the distribution of the reconstructed momentum transfer q^2 for recorded data and simulation. The results are shown in Figure 9.2, where the reconstruction modes are combined based on whether the B meson is charged or neutral and whether a ground state D or excited D^* is reconstructed in the final states. The first observation that is made is the agreement of the overall normalization for recorded data and MC, as the scaling factor of the latter is 1.0 for all channels, except for the mode $B^0 \to D\pi\ell$ for which it is 0.99. Considering the significant contribution from processes producing fake lepton candidates (hadronic Bkg), $B\overline{B}$ and continuum background in these reconstruction modes, this observation also strengthens the confidence in the description of these other three background components.
At first glance, the overall shape of the q^2 distribution is described well in MC. However, the determined p-values show a disagreement between the shapes in particular for the reconstruction modes involving a ground state D meson. This deviation could also be due to the three aforementioned background components, which also might suffer from

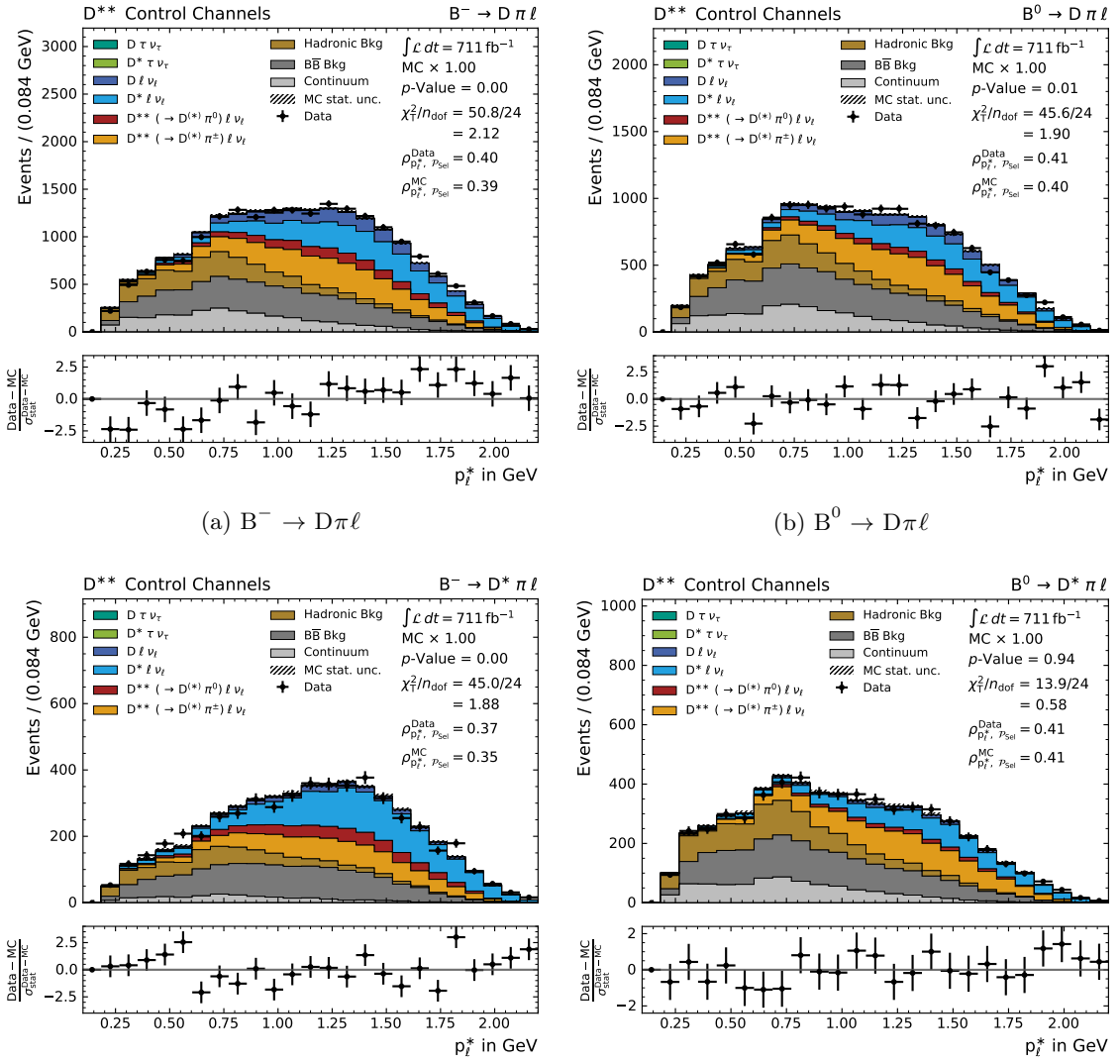


(a) $B^- \to D\pi\ell$ (b) $B^0 \to D\pi\ell$

(c) $B^- \to D^*\pi\ell$ (d) $B^0 \to D^*\pi\ell$

Figure 9.2.: Data-MC comparison of the momentum transfer q^2 in the D^{**}-Sideband samples. Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in the text of Section 9.3.

imperfections in their MC description. To eliminate any doubt regarding the modeling of the $B \to D^{**}\ell\nu$ processes, a fit of the invariant mass $M_{D^{(*)}\pi}$ of the $D^{(*)}\pi$ pair could be conducted to determine the normalization factors of the individual components from data. As the contribution of the $B \to D^{**}\ell\nu$ processes in the fit observables M^2_{miss} and p^*_ℓ is of interest for the determination of the decay rate ratios $\mathcal{R}(D^{(*)})$, the MC description of these two observables is also evaluated in this sideband. The comparison is displayed in Figure 9.3 for p^*_ℓ and Figure 9.4 for M^2_{miss}. In the case of the latter, the new ansatz for the correction of the M^2_{miss} resolution effect, which is described in Chapter 10, is utilized.
Visually, the distributions of p^*_ℓ in recorded data and MC agree well, although the p-values confirm this visual observation only for the reconstruction mode $B^0 \to D^*\pi\ell$ shown in Figure 9.3d. The agreement of the distributions also increase the confidence in the other three background components. The included Pearson correlation values $\rho_{p^*_\ell,\mathcal{P}_{\text{Sel}}}$ for the



(a) $B^- \to D\pi\ell$ (b) $B^0 \to D\pi\ell$

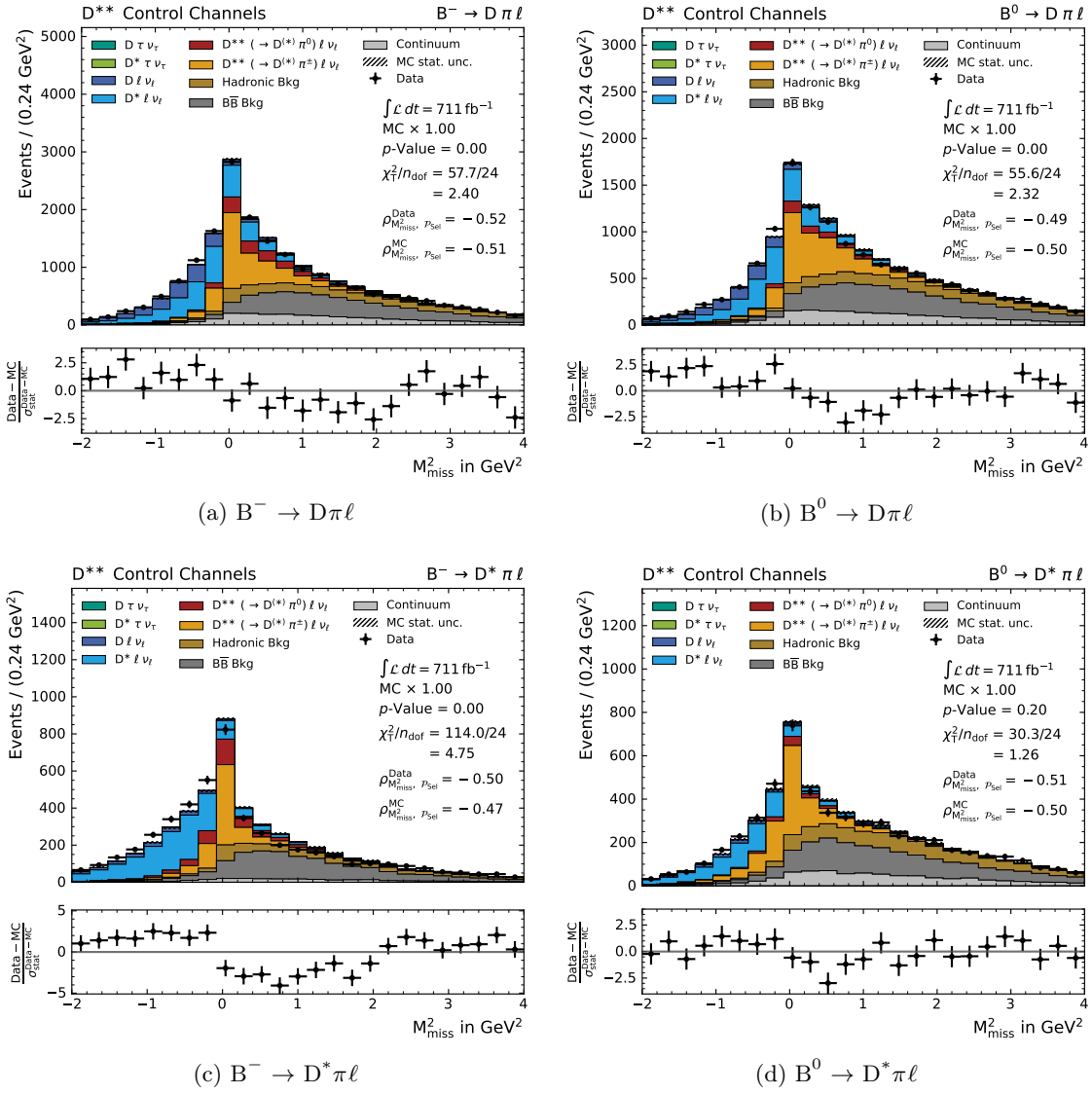
(c) $B^- \to D^*\pi\ell$ (d) $B^0 \to D^*\pi\ell$

Figure 9.3.: Data-MC comparison of the fit observable p^*_ℓ in the D^{**}-Sideband samples. Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in the text of Section 9.3.

displayed p^*_ℓ distribution and the selection classifier output $\mathcal{P}_{\text{Sel}}$ also agree well for recorded data and MC.

The evaluation of the data-MC agreement of the fit observable M^2_{miss}, shown in Figure 9.4, also shows the best agreement in the reconstruction mode $B^0 \to D^*\pi\ell$. As mentioned above, the M^2_{miss} resolution correction method described in Chapter 10 is applied, which improves the overall agreement in shape in the peak region, resulting in a less distinct pattern in the bin count pulls compared to what is observed in Figure D.13. The agreement between the distributions for recorded data and MC is particularly good in the right tail for $M^2_{\text{miss}} > 1.0\,\text{GeV}^2$, which is dominated by the other three background components.

To fully confirm the description of the $B \to D^{**}\ell\nu$ processes in MC, an extraction of the normalization factors of the individual components from a dedicated fit in the D^{**}-Sideband is beneficial. These additional studies of this sideband have not been concluded at the time



(a) $B^- \to D\pi\ell$ (b) $B^0 \to D\pi\ell$

(c) $B^- \to D^*\pi\ell$ (d) $B^0 \to D^*\pi\ell$

Figure 9.4.: Data-MC comparison of the fit observable M^2_{miss} in the D^{**}-Sideband samples. Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in the text of Section 9.3.

of writing.

9.4. Continuum Background Model Validation

The contribution of the ever-present continuum background, originating from $e^+e^- \to q\bar{q}$ processes, where q = u, d, s, c, is evaluated in off-resonance samples recorded at a center-of-mass energy of $\sqrt{s} = m_{\Upsilon(4S)} - 60\,\text{MeV}$ below the threshold for the $B\overline{B}$ production. A more detailed description of this off-resonance-sideband is provided in Section 6.3.6.

Comparing this recorded data sample to a corresponding MC sample for the two fit observables M^2_{miss} and p^*_ℓ yields the plots shown in Figure 9.5 and Figure 9.6. In both cases, a good MC description of the continuum background component can be confirmed. The p-values affirm the visual agreement between recorded data and simulation. The pulls show no significant patterns.

This comparison in the dedicated off-resonance-sideband together with the observations made in the previous Section 9.3 suggest that the shape of the continuum component is well understood.

The normalization of the recorded data and the MC sample differs slightly, however. This is visible in the scaling factor of the MC sample, which is 1.12 for the displayed case of the combination of all main reconstruction modes. To be able to combine the small contribution of the continuum component with the $B\overline{B}$ background component for the fit for the decay rate ratios $\mathcal{R}(D^{(*)})$ as suggested in Section 8.2, this normalization factor has to be understood and considered.

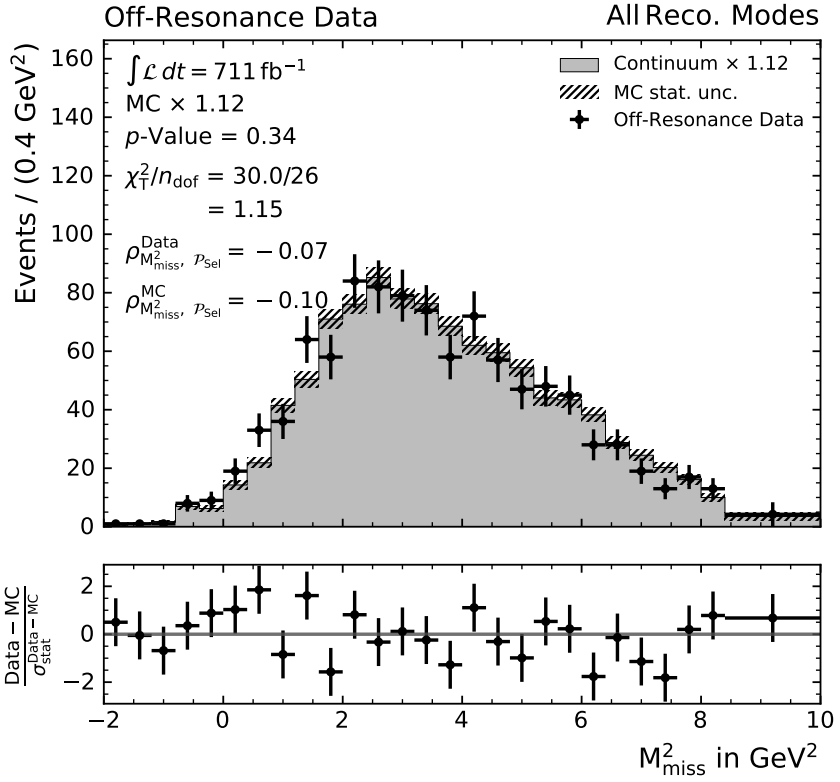


Figure 9.5.: Comparison of the fit observable M^2_{miss} for recorded data and MC off-resonance samples with a center-of-mass energy 60 MeV below the threshold for $B\overline{B}$ production. The comparison uses the combination of all main reconstruction modes. Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in the text of Section 9.4.

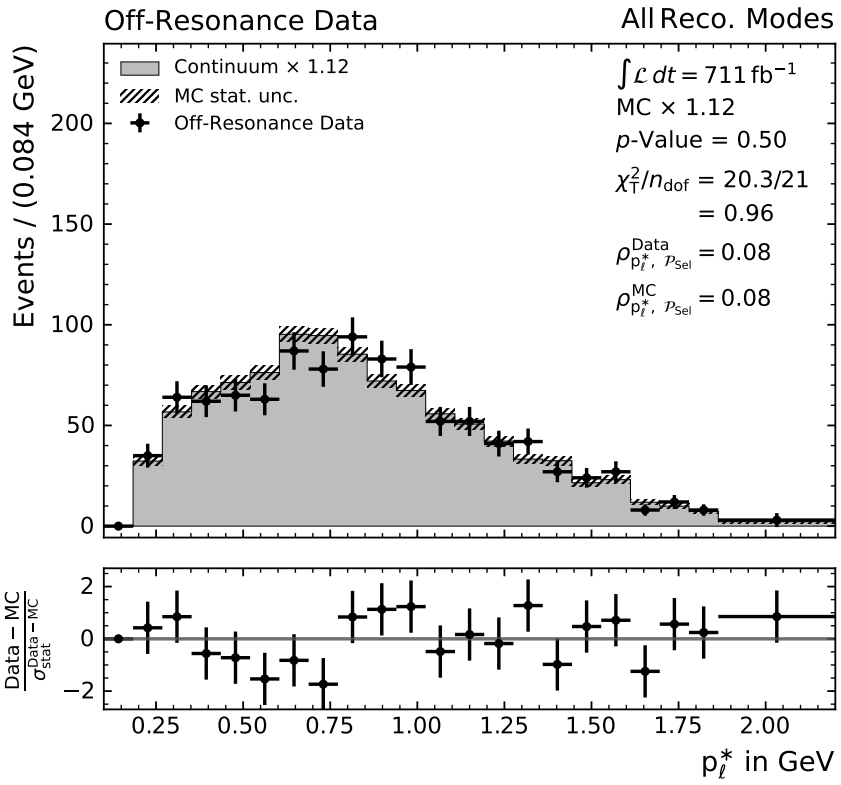


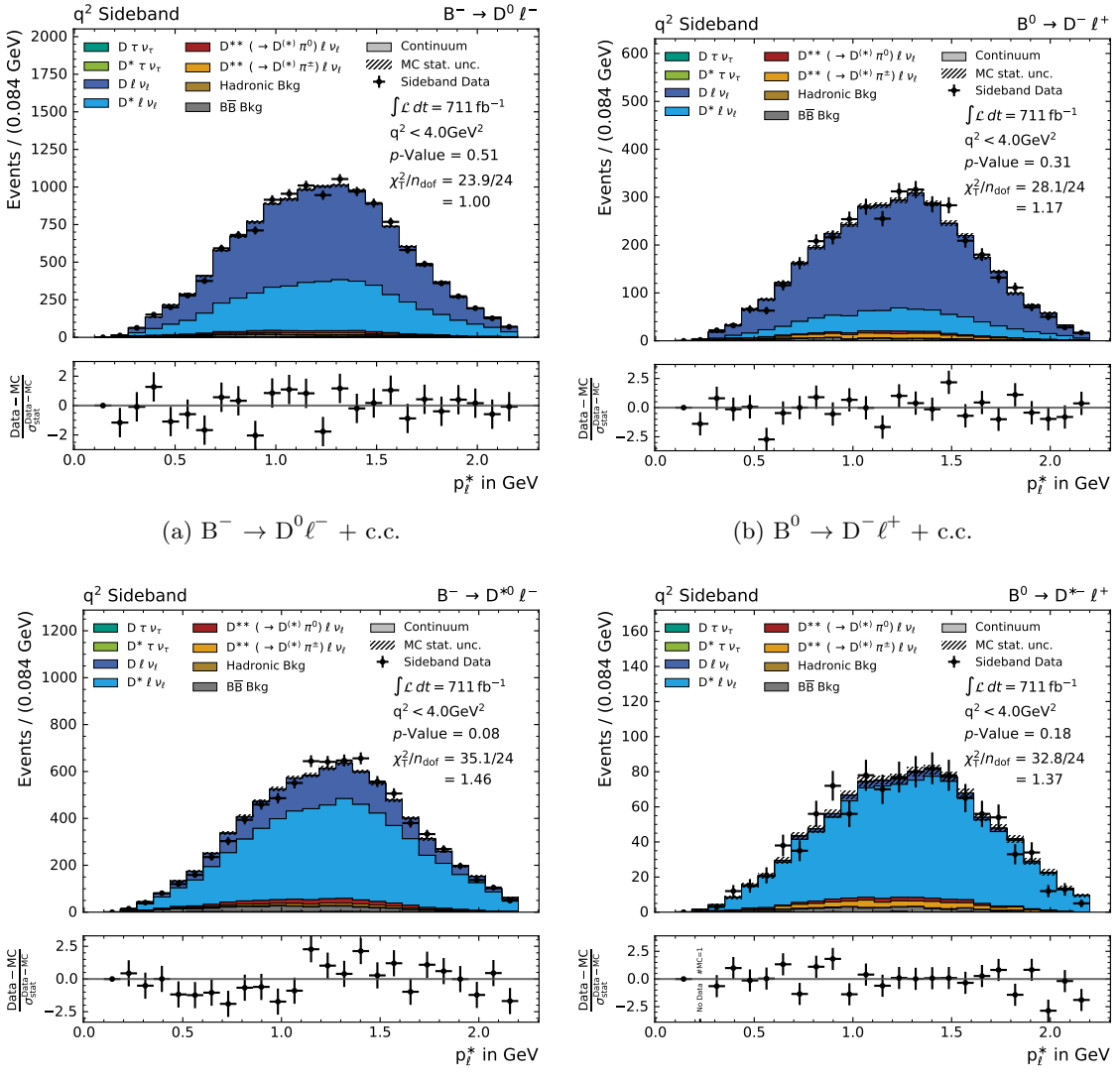
Figure 9.6.: Comparison of the fit observable p^*_ℓ for recorded data and MC off-resonance samples with a center-of-mass energy 60 MeV below the threshold for $\mathrm{B\overline{B}}$ production. The comparison uses the combination of all main reconstruction modes. Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in the text of Section 9.4.

Additional plots for the data-MC comparison in the off-resonance-sideband for the individual reconstruction modes are given in Appendix D.3, where the comparison plots for the fit observables $\mathrm{M}^2_{\mathrm{miss}}$ and p^*_ℓ are shown in Figures D.14 and D.15, respectively. These plots also confirm the observation for each individual reconstruction mode. The results for the channel $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+$ suffer from low statistics. Nonetheless, no significant disagreement between recorded data and MC is observed.

9.5. Fit Observables Validation

Last but not least, the data-MC agreement of the fit observables are evaluated in the q^2-Sideband. This allows in particular for the evaluation of the MC description of the normalization components $B \to D^{(*)}\ell\nu$. The comparison is conducted using the four main reconstruction modes $D^0\ell^-$, $D^{*0}\ell^-$, $D^+\ell^-$ and $D^{*+}\ell^-$. The Figures 9.7 and 9.8 show the results of these comparisons for the observables p^*_ℓ and M^2_{miss}, respectively.

For the observable p^*_ℓ a good agreement between recorded data and MC is observed. The bin count pulls do not show any significant patterns which otherwise might indicate a systematic effect which would require an additional correction of the MC. The visual



(a) $B^- \to D^0\ell^-$ + c.c.

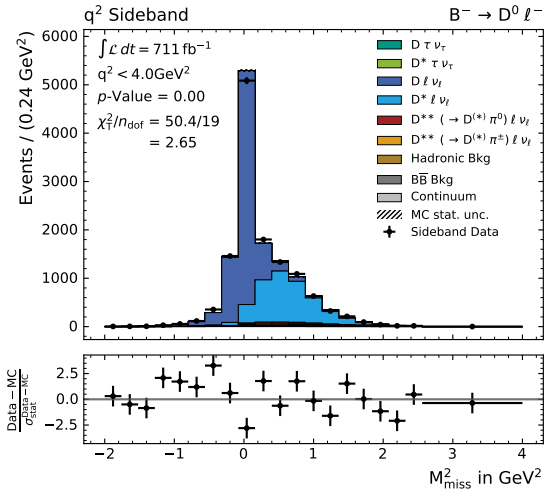
(b) $B^0 \to D^-\ell^+$ + c.c.

(c) $B^- \to D^{*0}\ell^-$ + c.c.

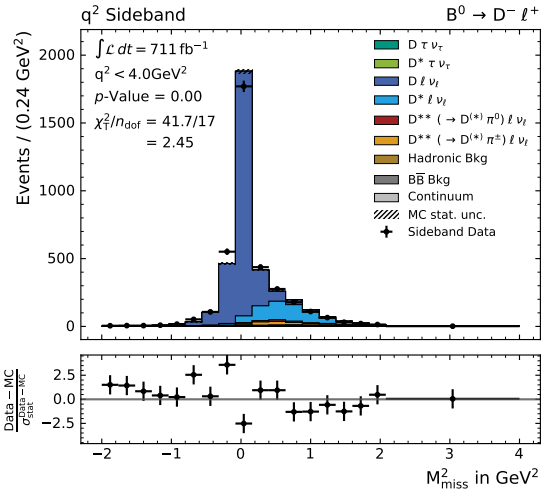
(d) $B^0 \to D^{*-}\ell^+$ + c.c.

Figure 9.7.: Comparison of the fit observable p^*_ℓ for recorded data and MC in the q^2-Sideband with $q^2 < 4.0\,\text{GeV}^2$. The plots show the comparison for each individual main reconstruction mode (see sub-captions). Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in the text of Section 9.5.

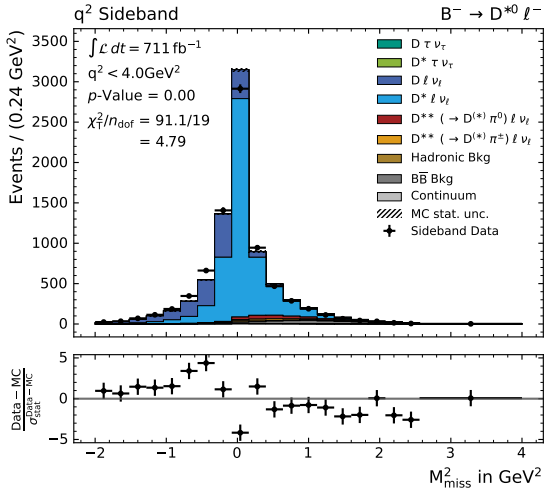
data-MC agreement is confirmed by the p-values obtained for the comparisons in each of the four main reconstruction modes. This result of the data-MC comparisons of the p_ℓ^* distribution, as well as the results obtained for the other sideband evaluates shown in the previous Sections 9.2 to 9.4, affirm the description of p_ℓ^* in MC and its use as a fit observable.
For the data-MC comparison of the observable M_{miss}^2, displayed in Figure 9.8, none of the aforementioned corrections of the resolution effect are applied. This is done to show the extent of this effect. A clear tendency to overestimate the bin counts in the peak region at $M_{\text{miss}}^2 \approx 0.0\,\text{GeV}^2$ is observed for the MC expectation in all four reconstruction modes. This is reflected in a pattern in the pull in the peak region. The p-values indicate a lack of agreement between recorded data and MC for each reconstruction mode as well.
As a proper description of the shape of the fit observable M_{miss}^2 is vital for the measurement of ratios $\mathcal{R}(D^{(*)})$ via the proposed template fit, this effect must be corrected in MC. The next Chapter 10 presents the approach to resolve this resolution issue in M_{miss}^2, as well as a more elaborate evaluation of the corrected M_{miss}^2 resolution via template fits in the q^2-Sideband.



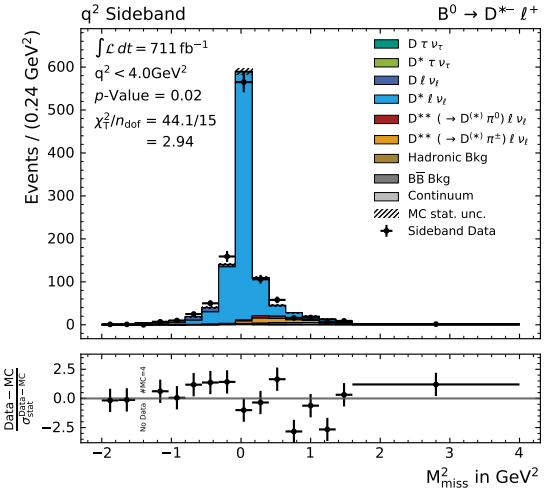
(a) $B^- \to D^0 \ell^-$ + c.c.



(b) $B^0 \to D^- \ell^+$ + c.c.



(c) $B^- \to D^{*0} \ell^-$ + c.c.



(d) $B^0 \to D^{*-} \ell^+$ + c.c.

Figure 9.8.: Comparison of the fit observable M^2_{miss} for recorded data and MC in the q^2-Sideband with $q^2 < 4.0\,\text{GeV}^2$. The plots show the comparison for each individual main reconstruction mode (see sub-captions). Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in the text of Section 9.5.

10. $\mathbf{M^2_{miss}}$ Resolution

The data-MC disagreement in the observable M^2_{miss} observed in the q^2-Sideband in the previous Section 9.5 is a large obstacle for the analysis. The origin of this effect is unknown, and it is not clear if it can be attributed to a mismodeling of the M^2_{miss} in MC.
Regardless of its origin, it causes a misinterpretation of the peaks in the distribution of the normalization modes at $M^2_{\text{miss}} = 0\,\text{GeV}^2$. This mismodeling can cause issues for the template likelihood fit, which relies on the description of the component shapes in MC. Hence, the effect has to be understood, if the missing mass squared shall be used as a fit observable. This chapter describes a method to correct the MC resolution difference, as well as a validation of the resolution.
Similar effects in the resolution of M^2_{miss} are observed in studies using Belle's FULL RECONSTRUCTION [80], as well as analyses which utilize the FEI tagging algorithm on Belle data [81]. In both cases, the attempts made to resolve the difference in the M^2_{miss} resolution observed between recorded data and MC apply a Gaussian smearing to the MC distribution of the observable. This approach has been studied in the scope of this analysis, but did not yield a satisfying data-MC agreement to allow for the usage of the missing mass squared as a fit observable. Instead, the procedure described in the following Section 10.1 is applied. This procedure is the result of many evaluations of the behavior of the observable M^2_{miss}. A validation of the M^2_{miss} resolution, based on fits to the q^2-Sideband data, is presented in Section 10.2.

10.1. Missing Mass Squared Resolution Correction

In the course of the evaluation of the observed M^2_{miss} resolution effect, several observables have been studied with regard to a possible correlation between the observable and the extent of the resolution effect. The beam-constraint mass $M^{\text{tag}}_{\text{bc}}$ of the tag side B meson shows the most promising features for this task. The resolution and a possible shift of the peak of the normalization components $B \to D^{(*)}\ell\nu$ in M^2_{miss} is evaluated on recorded data and MC in five bins of $M^{\text{tag}}_{\text{bc}}$. The five regions of $M^{\text{tag}}_{\text{bc}}$ used for this process are shown in Figure 10.1 and are chosen such that they reflect different degrees of reconstruction quality of the tag side B meson candidate.
For each of the five bins in $M^{\text{tag}}_{\text{bc}}$, the M^2_{miss} resolution correction is determined in two steps:

1. Determination of a shift in M^2_{miss} by comparing the distributions for recorded data and MC in the q^2-Sideband with different M^2_{miss} shifts applied to the MC distribution. The **$\mathbf{M^2_{miss}}$ shift value** that minimizes the Pearson χ^2 goodness-of-fit indicator is chosen.

2. The resolution of the M^2_{miss} peak of the normalization components is fitted in recorded data and in the shifted MC distribution using an analytical PDF function. A Gaussian PDF is used to describe the peaking component and a Cruijff distribution for the description of the remaining background components. The difference in the variance of the fitted Gaussian distributions for recorded data and MC is used as a **smearing factor $\boldsymbol{\sigma}$** for the MC M^2_{miss} distribution.

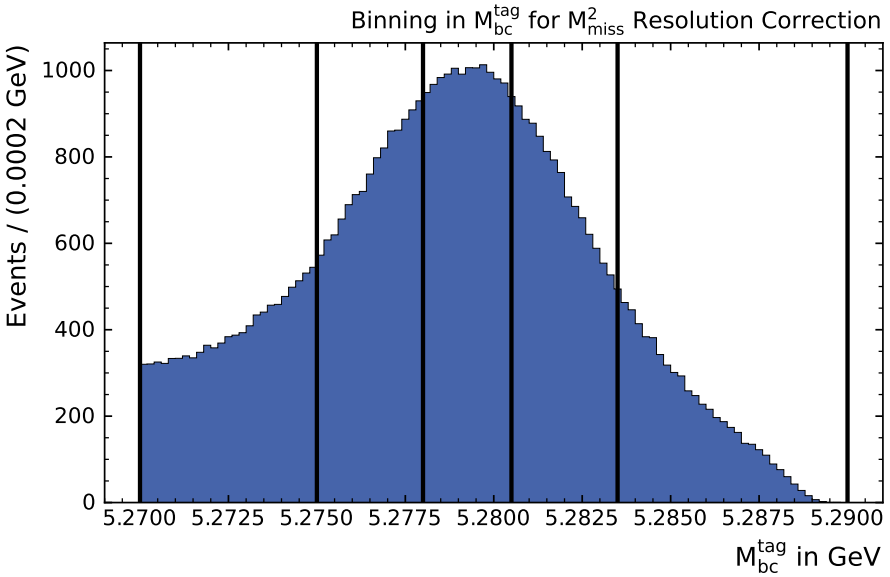
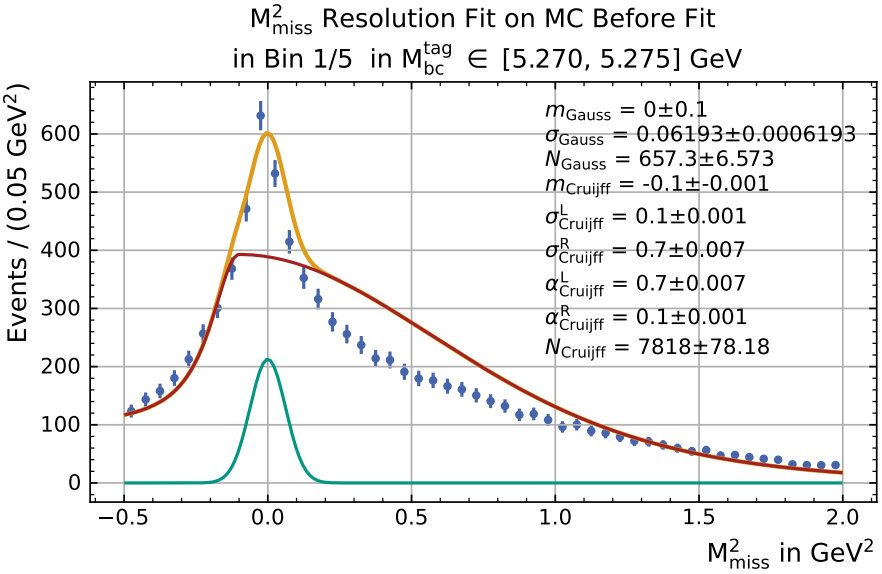


Figure 10.1.: Visualization of the binning in the beam-constraint mass M^{tag}_{bc} used to determine different resolution correction factors for the fit observable M^2_{miss}. The bin locations and widths are chosen to contain different qualities of tag side B meson candidates, with a narrow bin at the nominal B meson mass and bins of increasing width towards higher and lower M^{tag}_{bc} values. The lower bound at 5.27 GeV of the considered scope of M^{tag}_{bc} is defined by the selection criteria of the analysis.

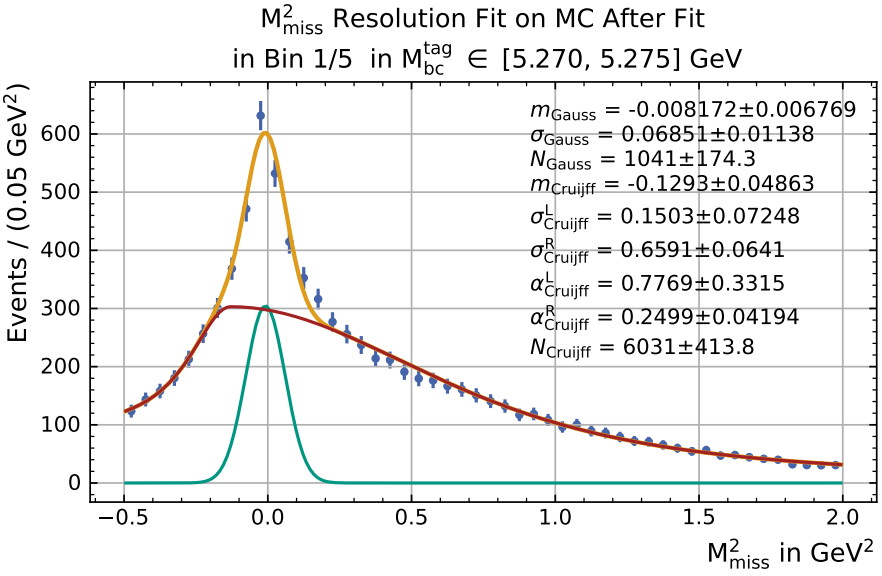
The fit of the second part of this procedure is performed with the help of the `probfit` package [82] and using MINUIT [77, 78] for the minimization of the negative log-likelihood of the fit model. The `probfit` package also provides the analytical description of the unnormalized Cruijff function:

$$f_{\text{Cruijff}}(x; m_0, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = \begin{cases} \exp\left(-\frac{(x-m_0)^2}{2(\sigma_L^2+\alpha_L(x-m_0)^2)}\right) & \text{if } x < m_0, \\ \exp\left(-\frac{(x-m_0)^2}{2(\sigma_R^2+\alpha_R(x-m_0)^2)}\right) & \text{if } x \geq m_0 \end{cases} \tag{10.1}$$

which is used to describe the combined background components. Here, the parameter m_0 denotes the mean of the distribution. The parameters $\sigma_{R,L}$ and $\alpha_{R,L}$ describe the width and tail of the distribution, respectively, to the right (R) and left (L) of the mean.
To improve the stability of the fit, the Gaussian describing the peaking component of the normalization processes $B \to D^{(*)}\ell\nu$ in the q^2-Sideband is first fitted by itself in a truth matched MC distribution, which only contains correctly reconstructed normalization events. The result of this first fit is used to set the starting parameters for the Gaussian in the fit to the full distribution in MC. Examples for this stage of the procedure are shown in Figure 10.2 for the first bin in M^{tag}_{bc}. The fit on recorded data in the q^2-Sideband is performed using the estimates of the fit in MC as starting parameters for the Gaussian peak and the Cruijff background component. The results of this fit to recorded data are displayed in Figure 10.3 for the first bin in M^{tag}_{bc}. The pre- and post-fit distributions for MC and recorded data for all bins in the tag side beam-constraint mass M^{tag}_{bc} are available in Appendix E.1 in the Figure E.17 and Figure E.18, respectively.



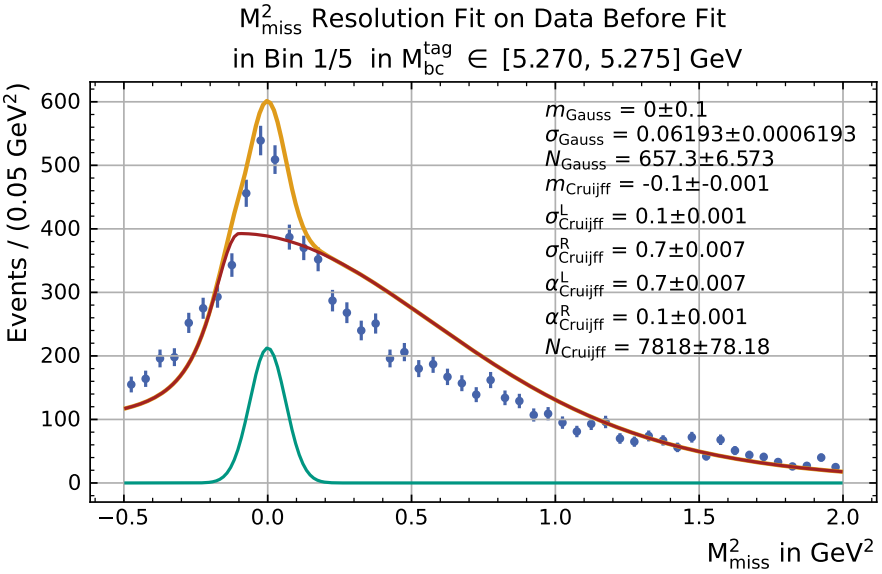
(a) M^2_{miss} resolution fit on **MC** in M^{tag}_{bc} bin 1, **pre-fit**



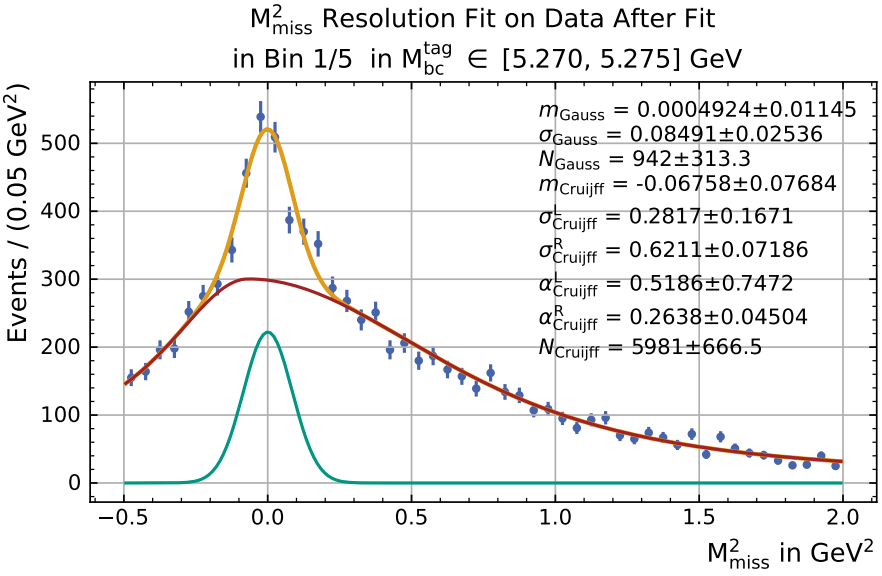
(b) M^2_{miss} resolution fit on **MC** in M^{tag}_{bc} bin 1, **post-fit**

Figure 10.2.: MC fit of M^2_{miss} resolution in the first M^{tag}_{bc} bin and in the q^2-sideband. The upper plot shows the pre-fit distribution (a) and the lower the post-fit distribution (b) with the fitted parameter values. The blue dots are the MC data points, while the solid lines represent the Gaussian signal shape (green), the Cruijff background shape (red) and the combination of the two (orange).

The thus obtained M^2_{miss} shift values and resolution smearing factors σ are used to correct the MC M^2_{miss} distributions. However, a correction of the MC M^2_{miss} resolution using a Gaussian smearing factor for the entire range of the missing mass squared does not yield the desired results (see e.g. the discussion on the M^2_{miss} based evaluation of the stability of the selection classifier output $\mathcal{P}_{Sel}$ in Section 9.2). A more sophisticated approach is required to resolve this inconsistency of the MC description of the M^2_{miss} peak. It is found



(a) M^2_{miss} resolution fit on **data** in M^{tag}_{bc} bin 1, **pre-fit**



(b) M^2_{miss} resolution fit on **data** in M^{tag}_{bc} bin 1, **post-fit**

Figure 10.3.: Data fit of M^2_{miss} resolution in the first M^{tag}_{bc} bin and in the q^2-sideband. The upper plot shows the pre-fit distribution (a) and the lower the post-fit distribution (b) with the fitted parameter values. The blue dots are the recorded data points, while the solid lines represent the Gaussian signal shape (green), the Cruijff background shape (red) and the combination of the two (orange).

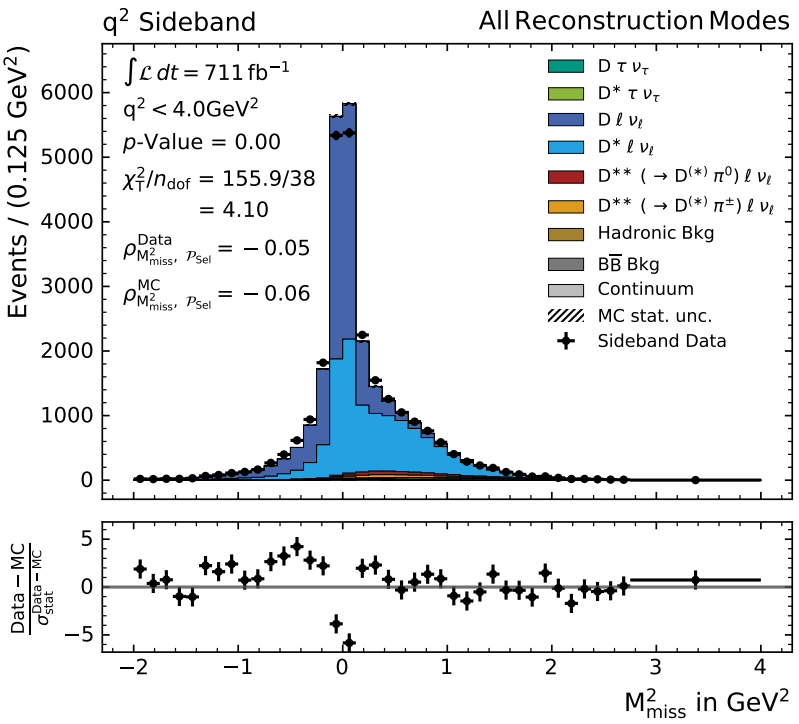
that the effect observed in the M^2_{miss} resolution is asymmetric.

This expresses itself in a consistent overestimation of the peak and underestimation of the tail in the negative M^2_{miss} region in MC. To model this asymmetry, an implementation of

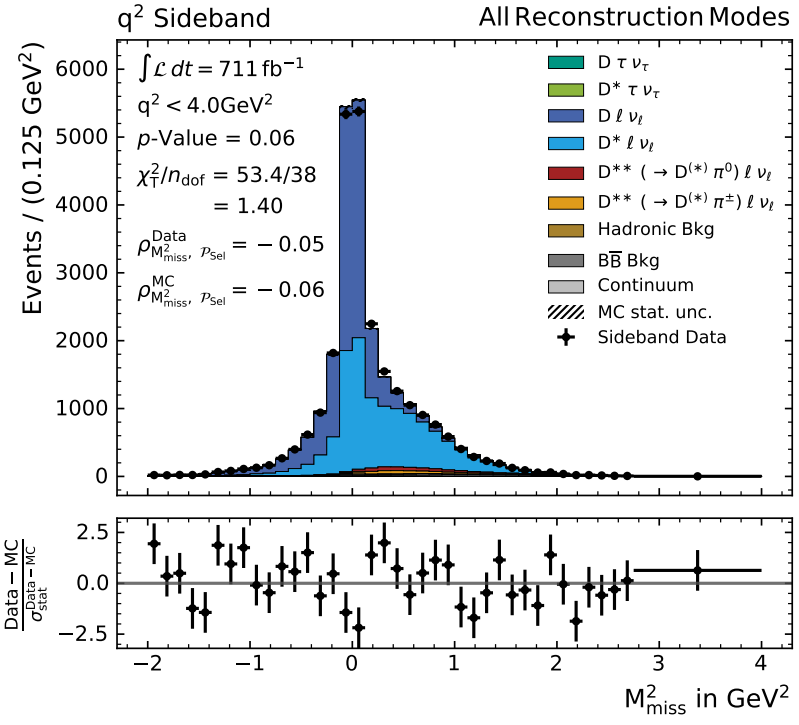
the asymmetric Laplace distribution provided by the `SciPy` package [83] is used:

$$f_{\mathrm{AL}}(x; m, \lambda, \kappa) = \frac{\lambda}{\kappa + 1/\kappa} \begin{cases} \exp\left((\lambda/\kappa)(x-m)\right) & \text{if } x < m, \\ \exp(-\lambda\kappa(x-m)) & \text{if } x \geq m, \end{cases} \tag{10.2}$$

where m denotes the mean of the distribution, λ is a scale parameter, and κ describes the asymmetry of the distribution. For the purpose of the smearing of $\mathrm{M}^2_{\mathrm{miss}}$, the scale parameter λ is a function of the smearing factor σ obtained from the fit to the q^2-Sideband data. The parameter κ is modulated with a reverse logistic sigmoid function depending on the to be smeared $\mathrm{M}^2_{\mathrm{miss}}$ value, such that the smearing is turned on for values of $\mathrm{M}^2_{\mathrm{miss}} < 0.5\,\mathrm{GeV}^2$. The scale or smearing factor σ is modulated with a Gaussian (mean of $-0.05\,\mathrm{GeV}^2$ and variance of $1.0\,\mathrm{GeV}^2$) and a reverse logistic sigmoid function, both depending on $\mathrm{M}^2_{\mathrm{miss}}$, such that the smearing is most prominent just to the left of the peak and is faded out for larger and smaller $\mathrm{M}^2_{\mathrm{miss}}$ values. The behavior of this smearing function is visualized and compared to a simple Gaussian smearing in Figure E.16 in Appendix E.1.
The result of this ansatz to correct the description of the $\mathrm{M}^2_{\mathrm{miss}}$ resolution in MC is displayed in the Figure 10.4. It shows the data-MC comparison in the q^2-sideband for the combinations of all four main reconstruction modes with and without the described resolution correction. The improvement due to the $\mathrm{M}^2_{\mathrm{miss}}$ resolution correction is clearly visible in both the p-values and the bin count pulls. In particular, the pattern in the pulls at $\mathrm{M}^2_{\mathrm{miss}} \approx 0.0\,\mathrm{GeV}^2$ is resolved by the correction.
At the time of writing, a possible source for this resolution effect was discovered. A FEI calibration study, conducted in the scope of the related $|\mathrm{V_{cb}}|$ measurement [57] mentioned in Chapter 5, recently provided new efficiency correction factors for the B tagging algorithm. These correction factors are determined for different ranges of the FEI probability $\mathcal{P}_{\mathrm{FEI}}$ which is an estimate of the quality of the $\mathrm{B_{tag}}$ meson candidate. The results of this calibration study show that $\mathrm{B_{tag}}$ meson candidate with a higher $\mathcal{P}_{\mathrm{FEI}}$ value are subject to a more extreme efficiency correction. This finding could resolve the issue in a more natural manner, as opposed to the artificial correction introduced in this Section 10.1. However, as these new correction factors have been made available only very recently, their effect must be studied before possibly applying them to the analysis. This finding is discussed more elaborately in Section 11.1 as part of the outlook.



(a) without correction



(b) with correction

Figure 10.4.: Data-MC comparison of the M^2_{miss} distribution using the combination of all four main reconstruction modes in the q^2-sideband without (a) and with (b) the M^2_{miss} resolution correction applied. The pattern observed in the bin count pulls (calculated as described in the introduction of Chapter 9) for the case without the correction disappears when the resolution correction is applied. The calculated χ^2-value and p-value also improve significantly. Furthermore, the visual agreement between the recorded data and MC distributions also improves, especially in the region of $M^2_{miss} \approx 0.0\,GeV^2$.

10.2. Ratios of Normalization Modes in q^2-Sideband

To evaluate the effect of the data-MC resolution difference in the region of the normalization component peak at $M^2_{\text{miss}} \approx 0\,\text{GeV}^2$, a fit in the q^2-sideband is performed. This is done by determining the ratio of the two normalization modes

$$\mathcal{R}_{\text{Norm}} = \frac{\mathcal{B}\left(\text{B} \to \text{D}\ell\nu_\ell\right)}{\mathcal{B}\left(\text{B} \to \text{D}^*\ell\nu_\ell\right)}, \tag{10.3}$$

for three different cases of extracting the event count N_{D^*} of the $\text{B} \to \text{D}^*\ell\nu$ processes in the denominator:

- **Matching Only**: in $\text{B} \to \text{D}^*\ell$ channels
- **Feed-Down Only**: $\text{D}^* \to \text{D}$ feed-down in $\text{B} \to \text{D}\ell$ channels
- **Coupled**: in $\text{B} \to \text{D}^*\ell$ channels **and** $\text{D}^* \to \text{D}$ feed-down in $\text{B} \to \text{D}\ell$ channels

In the first case, the $\text{B} \to \text{D}^*\ell\nu$ events are only extracted via the fit component in the matching reconstruction channel, where the component peaks at zero. In the second case, the component used to determine the $\text{B} \to \text{D}^*\ell\nu$ event count is the feed-down contribution of $\text{B} \to \text{D}^*\ell\nu$ events which are reconstructed in the channel with the ground state D meson in the final state. The distribution of this component is located to the right of $M^2_{\text{miss}} = 0\,\text{GeV}^2$ in the $\text{B} \to \text{D}\ell$ mode and is slightly broader. The third case is the combination of the two, where the component in the correct reconstruction channel and the feed-down component is coupled. This approach of coupling the two components is also used in the fit setup described in Chapter 8. For the first two cases, only one fit has to be performed with the two components decoupled from each other. The respective results for the ratio $\mathcal{R}_{\text{Norm}}$ are then computed from the fitted event counts of the two individual fit components.
The calculation of the ratio $\mathcal{R}_{\text{Norm}}$, the therefore required efficiencies, as well as the uncertainties on the quantities is performed analogous to the procedure described in Chapter 8. A template likelihood fit (see Section 7.2) is performed, using the MINUIT algorithm [77, 78] to minimize the negative log-likelihood function.
To determine the performance and validity of the fit procedure, it is performed first on q^2-sideband Asimov samples and then using recorded data. This also provides the opportunity to compare the uncertainties obtained for the Asimov study to the results on recorded data. The q^2-sideband is defined in Section 6.3.6. It is defined such that the description of the normalization components can be validated without unblinding the $\text{B} \to \text{D}^{(*)}\tau\nu$ signal processes.
The same reconstruction modes as in Chapter 8 are used, but with the q^2-sideband selection requirements applied. Fit templates are again defined in the M^2_{miss}–p^*_ℓ space with

- 15 equidistant bins in p^*_ℓ with $p^*_\ell \in [0.2, 2.3]\,\text{GeV}$; and
- 7 bins in M^2_{miss}, given by the bin edges
  $(-2.0, -1.25, -0.5, 0.5, 1.25, 2.0, 3.0, 4.0)$ in GeV^2
  which feature a wider bin at $M^2_{\text{miss}} = 0\,\text{GeV}^2$,

for each of the reconstruction modes.
The four templates per reconstruction mode are defined for the following components:

$B \to D\ell\nu$
The decay processes with the ground state D meson in the final state are used in the numerator of the ratio $\mathcal{R}_{\text{Norm}}$. These decay processes produce few cross-feed events in the $B \to D^*\ell$ reconstruction mode. They are correctly reconstructed in the $B \to D\ell$ modes, where their distribution will exhibit a peak at $0\,\text{GeV}^2$ in the missing mass squared.

$B \to D^*\ell\nu$ and/or its $D^* \to D$ Feed-Down
As described above, the components used to derive the event count for the denominator of the ratio $\mathcal{R}_{\text{Norm}}$ can be coupled, or treated separately. This is made use of, to test the MC description of the resolution of the peak in M^2_{miss} at $0\,\text{GeV}^2$.

$B \to D^{**}\ell\nu$ and $B\overline{B}$ Background Component
The q^2-sideband is dominated by the contributions from the processes considered in the two components defined above. Thus, the remaining events originating from other $B \to D^{**}\ell\nu$ decay processes can be combined into a single background component. This includes lepton fakes and combinatorial background, as well as the contribution from $B \to D^{**}\ell\nu$ processes.

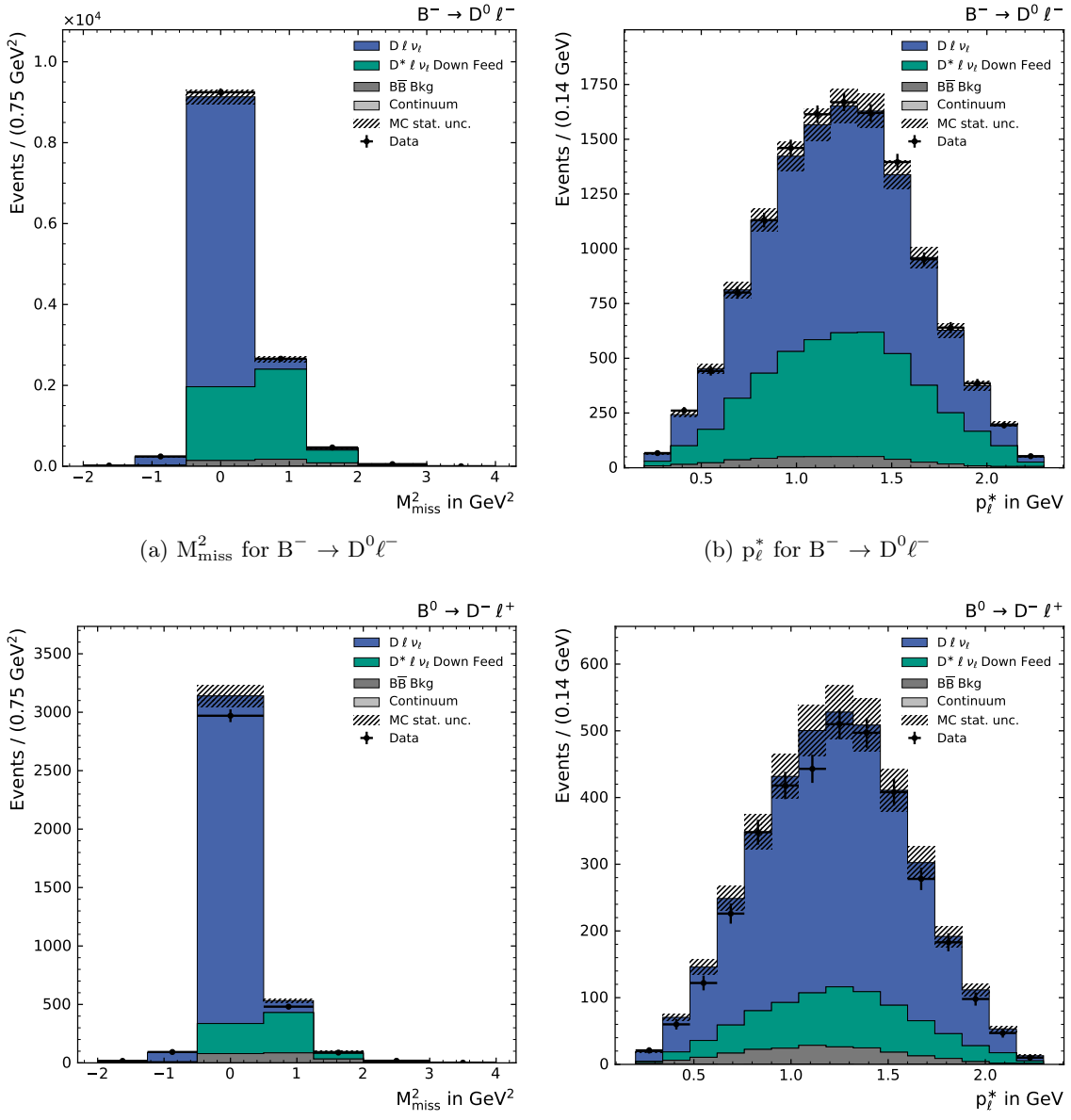
Continuum Background Component
The contribution of $e^+e^- \to q\bar{q}$ continuum processes with $q = u, d, s, c$ is small compared to the other components. Nonetheless, they are considered as a individual background component, due to the uncertainty of the MC description of their reconstruction efficiency.

For a more detailed discussion of the individual processes and their distributions, consider the descriptions given in Section 8.2. Overall, the processes behave in the q^2-sideband similarly to the respective processes in the main selection. Hence, this sideband allows for a representative test of the behavior of the normalization components in the region around $M^2_{\text{miss}} \approx 0\,\text{GeV}^2$ via this fit.
The resulting distributions used for the definition of the templates are shown in Appendix E.2. for the general case of separate components for the processes $B \to D^*\ell\nu$ and $B \to D^*\ell\nu$ $D^* \to D$ feed-down, respectively.
Applying the described fit procedure to the q^2-sideband data results in the post-fit distributions projections shown in Figure 10.5. The fitted templates describe the distribution observed on data well in the p^*_ℓ dimension. However, for the distribution in M^2_{miss}, an overshoot in MC events is observed for both neutral B meson reconstruction channels at $M^2_{\text{miss}} \approx 0\,\text{GeV}^2$. In comparison, the fitted MC distribution agrees better with the event counts observed in recorded data in the charged B meson reconstruction channels.
The values for the ratio $\mathcal{R}_{\text{Norm}}$ of the two normalization mode decay rates obtained from the fit to the Asimov sample, as well as to recorded data, both in the q^2-Sideband, are listed in Table 10.1. This table includes the MC expectation value of this ratio, as well as the fit results on both samples for the three different fit component setups described above. The MC expectation value is calculated from the branching fractions used for the definition of the MC sample as described in Section 3.3. Therefore, the value of this MC expectation is always the same. The uncertainties on the branching fraction values are propagated to this MC expectation of $\mathcal{R}_{\text{Norm}}$ to obtain the uncertainties included in table.
In addition to the evaluation of the optimized selection setup, the ratio $\mathcal{R}_{\text{Norm}}$ is also obtained using the loose selection criterion on the selection BDT output of $\mathcal{P}_{\text{Sel}} > 0.2$ (see Equation (6.22)).
The first notable result of this study is the good agreement between the uncertainty estimate obtained from Asimov fits with the uncertainties observed on recorded data samples. Comparing the three different fit setups for the optimized selection, the best

Table 10.1.: Calculation of the decay rate ratio $\mathcal{R}_{\text{Norm}} = \mathcal{B}(B \to D\ell\nu)/\mathcal{B}(B \to D^*\ell\nu)$ from three different fit setups. A differentiation is made between the for the $\mathcal{R}(D^{(*)})$ measurement optimized selection, denoted by **Opt** and a selection using the loose working point on the selection classifier $\mathcal{P}_{\text{Sel}} > 0.2$ (see Equation (6.13)) listed in the rows labeled with **Loose**. In the column labeled as *MC Expectation*, the ratio $\mathcal{R}_{\text{Norm}}$ calculated from the branching fraction values used for the definition of the MC sample (see Section 3.3) is listed together with the related uncertainty. In the other two columns the results from fits to an Asimov data set and the recorded data set are tabulated. The uncertainties given for the values of $\mathcal{R}_{\text{Norm}}$ from the fits are the statistical and multiplicative systematical uncertainties, respectively.

$\mathcal{R}_{\text{Norm}}$		MC Expectation	q^2-Sideband Asimov	q^2-Sideband Data
Opt	Coupled	0.438 ± 0.016	$0.438 \pm 0.008 \pm 0.003$	$0.445 \pm 0.009 \pm 0.003$
	Matching Only	0.438 ± 0.016	$0.438 \pm 0.008 \pm 0.004$	$0.452 \pm 0.010 \pm 0.004$
	Feed-Down Only	0.438 ± 0.016	$0.438 \pm 0.012 \pm 0.004$	$0.419 \pm 0.012 \pm 0.003$
Loose	Coupled	0.438 ± 0.016	$0.438 \pm 0.008 \pm 0.003$	$0.438 \pm 0.008 \pm 0.003$
	Matching Only	0.438 ± 0.016	$0.438 \pm 0.008 \pm 0.004$	$0.441 \pm 0.009 \pm 0.004$
	Feed-Down Only	0.438 ± 0.016	$0.438 \pm 0.010 \pm 0.003$	$0.429 \pm 0.011 \pm 0.003$

agreement between the expected value and the value obtained on recorded data is found for the coupled fit setup. The setup denoted as Matching Only overestimates the decay rate ratio $\mathcal{R}_{\text{Norm}}$ on data, whereas the Feed-Down Only setup yields a value below the expectation. The results using the loose selection agree slightly better with each other, with the result for the coupled fit setup reproducing the expectation very well. In both cases, the same tendencies for the different fit setups are observed. As these three different approaches are supposed to test the stability of the fit procedure in regard to the M^2_{miss} resolution effect, do the observed differences provide cause for additional investigations. Further studies are planned to get a grasp of the observed effects. Recent findings, which will be discussed in the summary in Section 11.1, provide new paths to be followed, which might also explain the results presented in this Section 10.2.



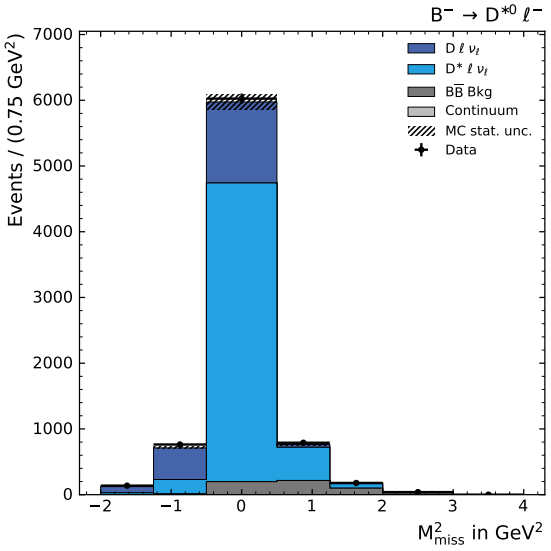
(a) M^2_{miss} for $B^- \to D^0 \ell^-$

(b) p^*_ℓ for $B^- \to D^0 \ell^-$

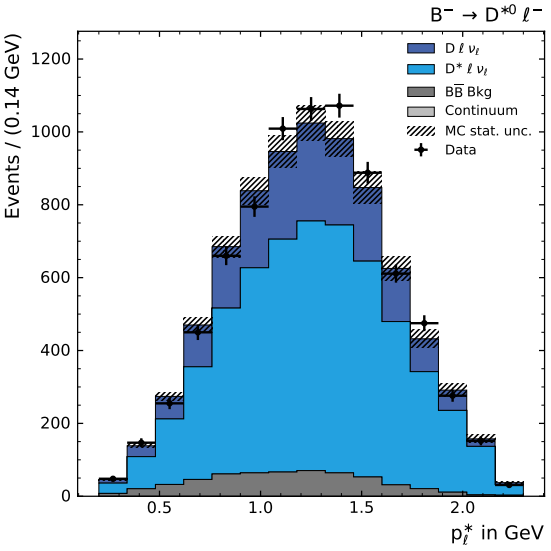
(c) M^2_{miss} for $B^0 \to D^- \ell^+$

(d) p^*_ℓ for $B^0 \to D^- \ell^+$

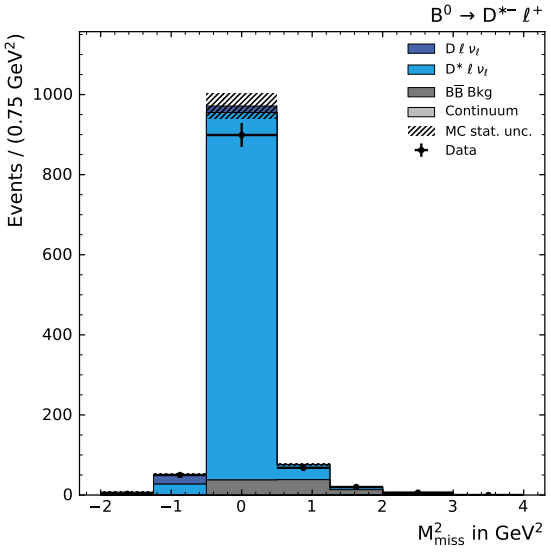
Figure 10.5.: Projections of the post-fit distributions resulting from the fit on q^2-sideband data for the decoupled fit case. The uncertainties on the MC distribution are indicated as diagonally hatched bands. The distributions in M^2_{miss} and p^*_ℓ are shown on the left and right, respectively. The first row shows the result in the $B^- \to D^0 \ell^-$ reconstruction mode and the second row for the $B^0 \to D^- \ell^+$ mode. The results for the remaining two reconstruction modes are shown in the continuation of this figure on page 133. A detailed discussion of the results is provided in the text.



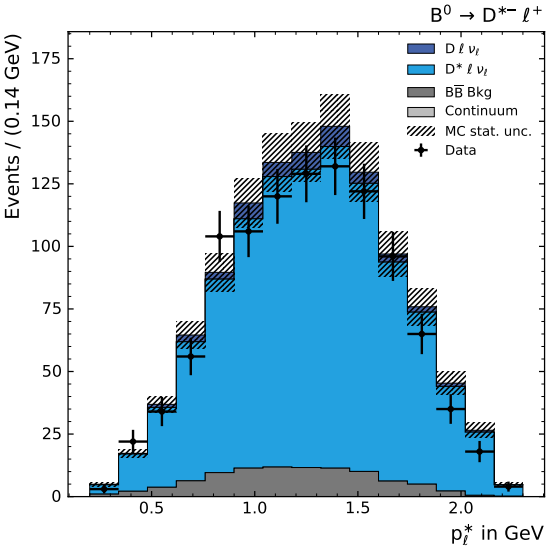
(e) M^2_{miss} for $B^- \to D^{*0}\ell^-$



(f) p^*_ℓ for $B^- \to D^{*0}\ell^-$



(g) M^2_{miss} for $B^0 \to D^{*-}\ell^+$



(h) p^*_ℓ for $B^0 \to D^{*-}\ell^+$

Figure 10.5.: Continuation of Figure 10.5 on page 132. The first row shows the result in the $B^- \to D^{*0}\ell^-$ reconstruction mode and the second row for the $B^0 \to D^{*-}\ell^+$ mode.

11. Conclusion and Outlook

The goal of the presented analysis is to produce the final word from Belle on the test of lepton flavor universality via the measurement of the branching fraction ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$. To make maximal use of the recorded collision events, I employed modern analysis methods and devised new measurement strategies.
The use of the hadronic FEI tagging results in a significant improvement of the reconstruction efficiency. This becomes evident when comparing the SM expected yields of the $B \to D^{(*)}\tau\nu$ signal processes to the previous Belle analysis [12, 21] using a similar analysis strategy but relying on a less efficient tagging algorithm. Table 11.1 shows an improvement of more than a factor of two.

Table 11.1.: Comparison of the expected yields for the signal processes $B \to D^{(*)}\tau\nu$ for this analysis and the previous Belle analysis with hadronic B meson tagging from 2015 [12], both assuming the SM. The expected yields for the previous Belle analysis are taken from [21, Table 9.7].

# Events for	$B \to D\tau\nu$	$B \to D^*\tau\nu$
Belle (2015)	287	470
This Analysis	634	1218
Factor	2.2	2.6

In addition, the modelling of the $B \to D^{**}\ell\nu$ backgrounds and the $B \to D^{(*)}\ell\nu$ normalization modes was improved such that it reflects the latest knowledge of branching fractions and form factors. This results in an improved description of these processes in the signal extraction fit. Particularly, the description of the $B \to D^{**}\ell\nu$ background is crucial, as these processes can mimic the signal $B \to D^{(*)}\tau\nu$ processes and poses one of the largest systematical uncertainties in the measurement of $\mathcal{R}(D^{(*)})$. The new models for both were validated either in sidebands or in case of the $B \to D^{**}\ell\nu$ process in dedicated reconstruction channels and good agreement is found with the new background descriptions.

11.1. Recent Findings

At the time of writing, the analysis is in the internal review process of the Belle collaboration and the signal region remains blinded to avoid any experimental bias. During this review process, the results of a first calibration of the FEI B tagging algorithm using inclusive semi-leptonic decays was derived [57]. The calibration shows a clear dependence of the correction factors to the estimated quality of the B_{tag} candidate provided by the FEI B tagging algorithm. This dependence is visible in Figure F.23 in Appendix F. In a fully reconstructed $B\overline{B}$ event, the quality of the B_{tag} candidate provided by the FEI is correlated to the quality of the signal side B_{sig} meson candidate. Specifically there is a correlation with M^2_{miss} used in the signal extraction. This dependence poses a possible additional systematical effect, which needs to be studied and considered when e.g. correcting the observed differences in the simulated samples and when combining the information of

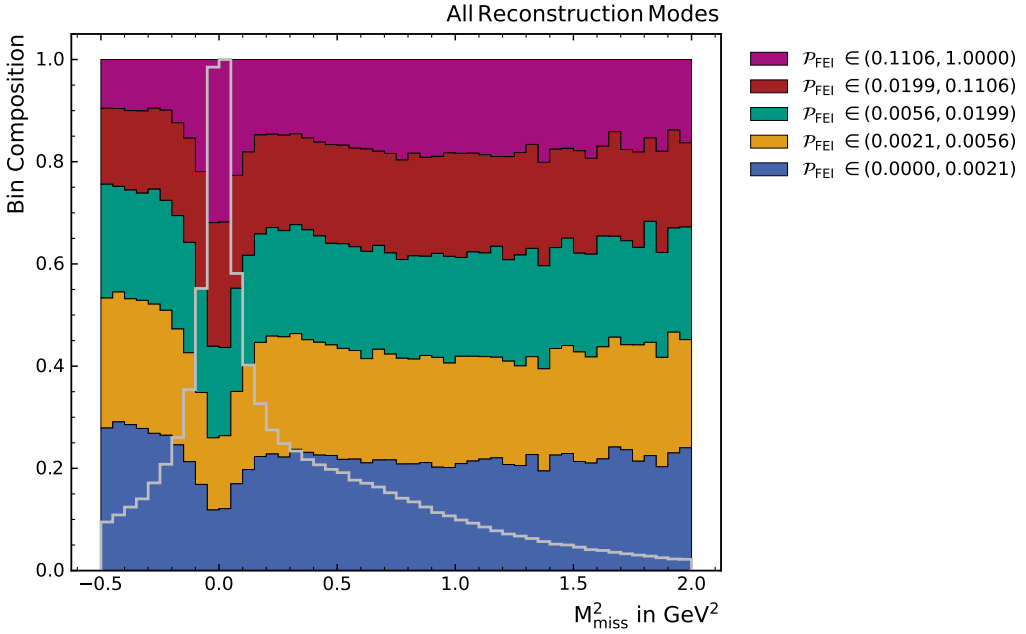


Figure 11.1.: Visualization of the contributions of B_{tag} candidates with different FEI probability $\mathcal{P}_{FEI}$ in bins of the fit observable M^2_{miss}. The shape of the M^2_{miss} distribution is indicated in gray. The colored regions represent the fractions of B_{tag} candidates with a FEI probability $\mathcal{P}_{FEI}$ values within five bins in $\mathcal{P}_{FEI}$ of equal statistics, as given in the legend to the right of the plot. This visualization shows that the contribution of B_{tag} candidates with different $\mathcal{P}_{FEI}$ values depends on M^2_{miss}. Events with $M^2_{miss} \approx 0\,\text{GeV}^2$ feature B_{sig} candidates which are well reconstructed in the $B \to D^{(*)}\ell\nu$ normalization modes. In this M^2_{miss} region, the fraction of B_{tag} candidates with high $\mathcal{P}_{FEI}$ values increases.

charged and neutral B mesons. In the previous Belle analysis [12] such effects could not be considered in detail as no calibration was available.

11.2. Outlook

The final steps planned for the presented analysis comprise

1. the inclusion of the new FEI calibration factors,
2. a final evaluation of the M^2_{miss} resolution effect in the q^2-sideband data samples, as well as
3. the application of constraints to the background components in the fit to further improve the precision on the measurement of the ratios $\mathcal{R}(D^{(*)})$.

I plan to implement these changes and apply the presented procedure to the Belle recorded data immediately. Considering all the aforementioned improvements made within the scope of this thesis, I expect to provide new insights on the test of lepton flavor universality with the measurement of the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ at Belle.

Appendices

A. Supplementary Material for the Form Factor Correction

Table A.2.: Statistical correlation matrix for the fit with BGL parametrization in the configuration (1,1,2) of [34].

	Nom. $\times 10^3$	Stat. Correlation Matrix					
a_0^g	1.00	1.000	−0.937	−0.128	0.069	−0.081	0.161
a_1^g	−2.35	−0.937	1.000	0.127	−0.222	0.110	−0.192
a_0^f	0.511	−0.218	0.127	1.000	−0.800	−0.751	0.624
a_1^f	0.67	0.069	−0.222	−0.800	1.000	0.443	−0.354
$a_0^{F_1}$	0.30	−0.081	0.110	−0.751	0.443	1.000	−0.978
$a_1^{F_1}$	−3.68	0.161	−0.192	0.624	−0.354	−0.978	1.000

Table A.3.: Systematical correlation matrix for the fit with BGL parametrization in the configuration (1,1,2) of [34].

	Nom. $\times 10^3$	Sys. Correlation Matrix					
a_0^g	1.00	1.000	−0.940	−0.132	0.085	−0.077	0.158
a_1^g	−2.35	−0.940	1.000	0.129	−0.228	0.107	−0.189
a_0^f	0.511	−0.132	0.129	1.000	−0.806	−0.755	0.629
a_1^f	0.67	0.085	−0.228	−0.806	1.000	0.452	−0.362
$a_0^{F_1}$	0.30	−0.077	0.107	−0.755	0.452	1.000	−0.977
$a_1^{F_1}$	−3.68	0.158	−0.189	0.629	−0.362	−0.977	1.000

B. MVA Selection Classifier

B.1. MVA Selection Optimization

The observables used as input features for the selection classifier described in Section 6.3.1 are evaluated regarding their correlation with the observables used for the fit to determine the decay rate ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$. The obtained correlation matrix is illustrated in Figure B.2. In addition to the input feature observables and the fit observables $\mathrm{M}^2_{\mathrm{miss}}$ and p^*_ℓ, it also contains the observable calculated from the difference in missing energy $\mathrm{E}_{\mathrm{miss}}$ and missing momentum $\mathrm{p}_{\mathrm{miss}}$. This observable was considered as an alternative to the missing mass squared, but was dismissed. Thus, it is not discussed further in this text.
Focusing on the remaining observables, no strong correlation between the input feature observables and the fit observables is observed. The strong correlation between the two masses $\mathrm{M}(\mathrm{D}^{(*)})$ and M_{D}, where the latter refers to the mass of the daughter of a D^* meson, is expected.

B.2. $\mathcal{R}(\mathrm{D}^{(*)})$ MVA Selection Optimization

For the optimization of the selection criteria applied on the output $\mathcal{P}_{\mathrm{Sel}}$ of the signal selection classifier introduced in Section 6.3.1, the statistical uncertainty on the decay rate ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ obtained from fits to Asimov data is used. The selection as

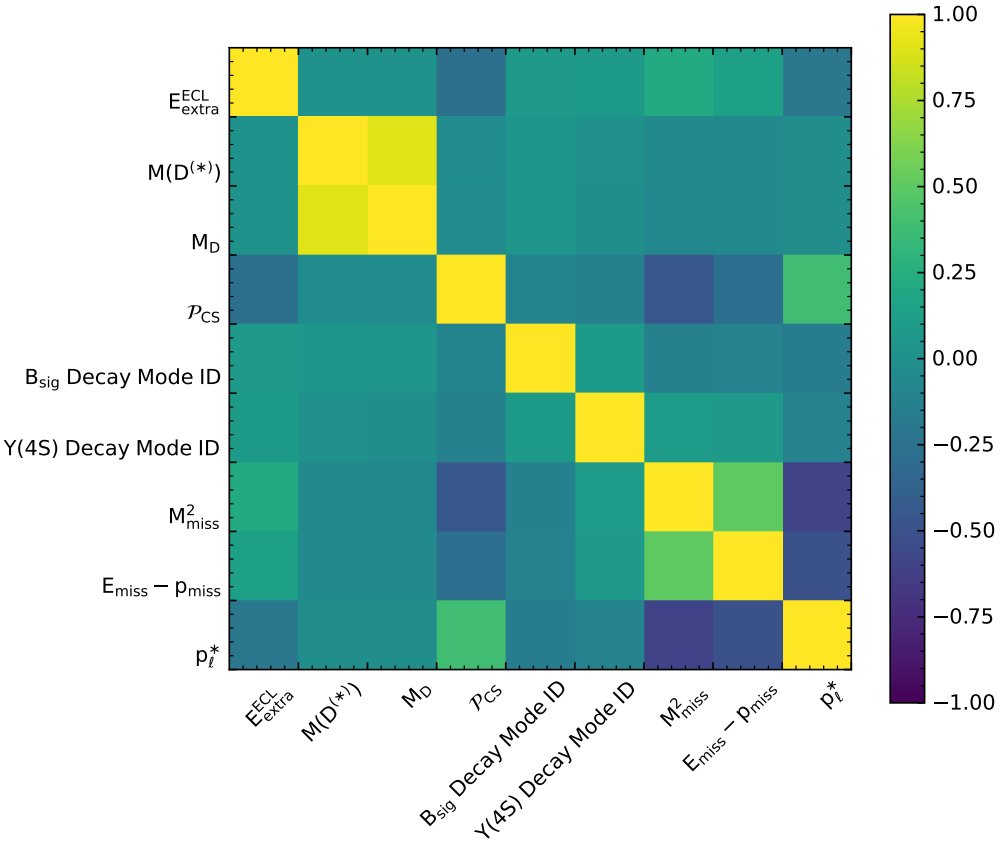


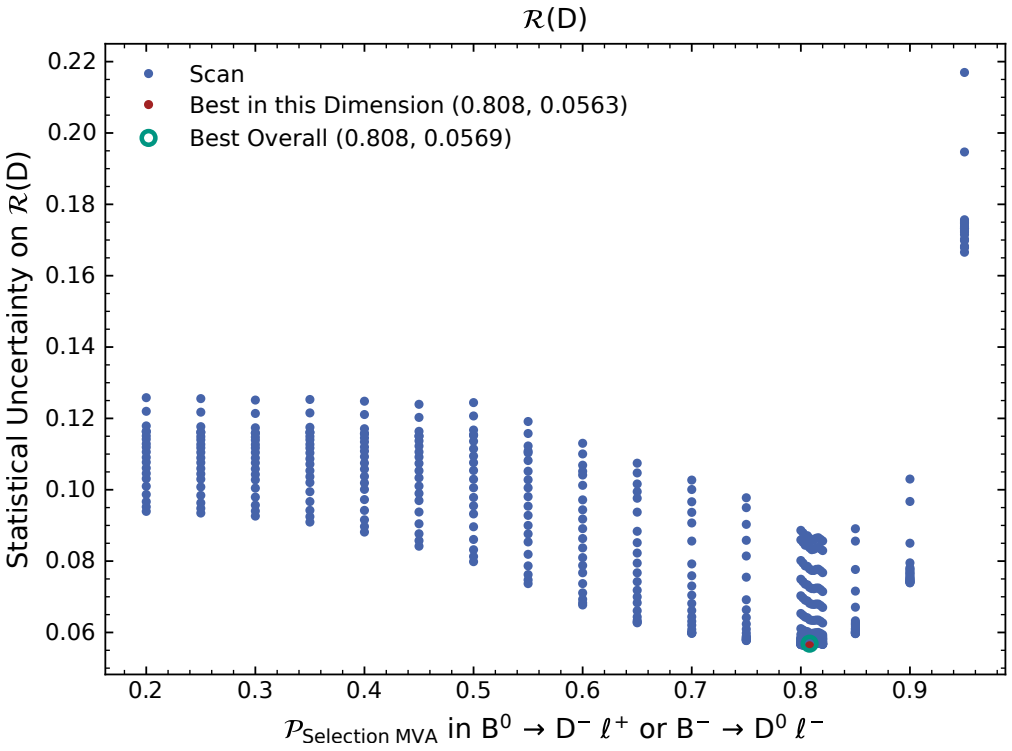
Figure B.2.: Visualization of the correlation matrix of the input feature observables to the selection classifier and the fit observables. No strong correlations between the two categories is observed.

described in Section 6.3 is applied on the simulated data, using an initial, loose requirement on the classifier output of $\mathcal{P}_{\text{Sel}} > 0.2$. Starting from the thus obtained MC sample, the ratios $\mathcal{R}(\text{D}^{(*)})$ are determined with the fit procedure described in Chapter 8. The procedure is repeated with evermore stringent requirements on the classifier output $\mathcal{P}_{\text{Sel}}$, thereby scanning the available value range of $\mathcal{P}_{\text{Sel}}^{\text{Scan}} \in [0.2, 0.95]$. For too stringent requirements on $\mathcal{P}_{\text{Sel}}$, the available statistics do not allow for sensible fit results. Hence, the upper bound of the scan region of $\mathcal{P}_{\text{Sel}} > 0.95$ is used. For each scan point $\mathcal{P}_{\text{Sel}}^{\text{Scan}}$, the statistical uncertainties on $\mathcal{R}(\text{D})$ and $\mathcal{R}(\text{D}^*)$ are added up. The scan point $\mathcal{P}_{\text{Sel}}^{\text{Scan}}$ yielding the lowest value for this sum of uncertainties on $\mathcal{R}(\text{D})$ and $\mathcal{R}(\text{D}^*)$ is deemed as the optimal choice for the selection requirement on $\mathcal{P}_{\text{Sel}}$.

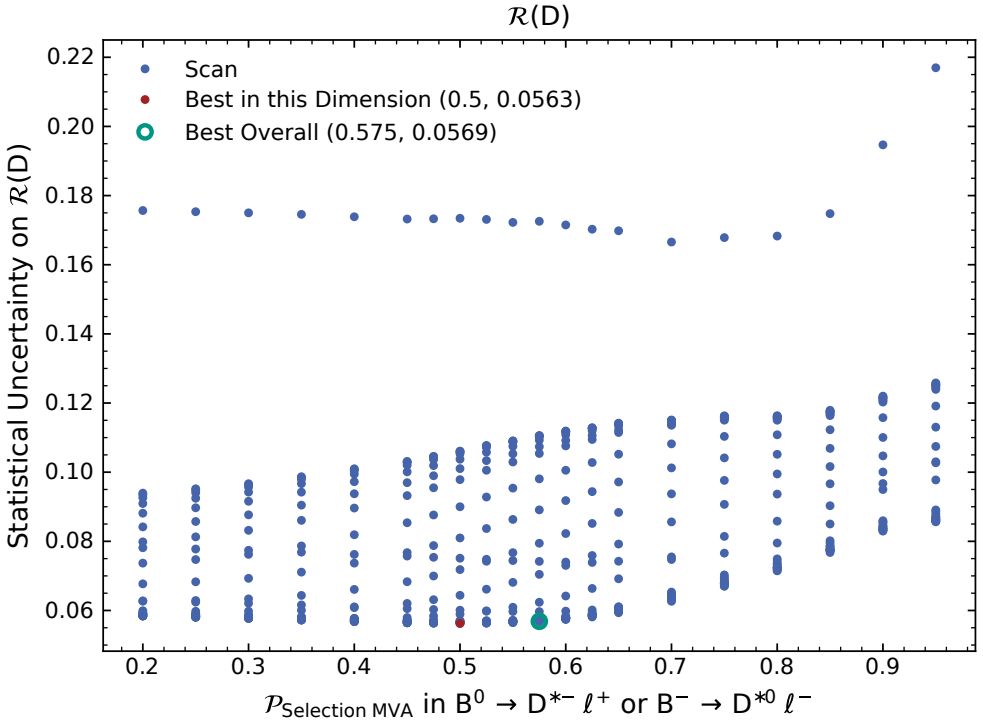
This procedure is repeated using different requirements on $\mathcal{P}_{\text{Sel}}$ depending on the reconstruction mode in which the event was reconstructed. Overall, the best result is obtained, when applying two different requirements on $\mathcal{P}_{\text{Sel}}$ dependent on whether a reconstruction mode with a ground state or an excited D meson is used. The results of this grid search is shown in Figures B.3 and B.4 and the found selection criteria are stated in Section 6.3.2.

In the Figures B.3a and B.3b the statistical uncertainties on $\mathcal{R}(\text{D})$ are shown for the scans for the $\text{B} \to \text{D}\ell$ and $\text{B} \to \text{D}^*\ell$ reconstruction modes, respectively. All values considered during the grid search are indicated as blue dots. The result is dominated by the minimum found in the scan for the $\text{B} \to \text{D}\ell$ modes. In the region of the minimum the scan point density was increase to find the optimal value. The scan of $\mathcal{P}_{\text{Sel}}$ values for the $\text{B} \to \text{D}^*\ell$ reconstruction modes revealed a second more optimal value for the selection criteria for the case that only the reconstruction modes with an exited D^* meson would be used. This value is indicated as red dot (Best in this dimension), whereas the overall best value is indicated as green circle.

The scans for the minimal statistical fit uncertainty on $\mathcal{R}(\text{D}^*)$ is less conclusive and for the overall best value, the uncertainty on $\mathcal{R}(\text{D})$ dominates. Thus, a compromise is made when the statistical uncertainty on both $\mathcal{R}(\text{D})$ and $\mathcal{R}(\text{D}^*)$ are considered together.

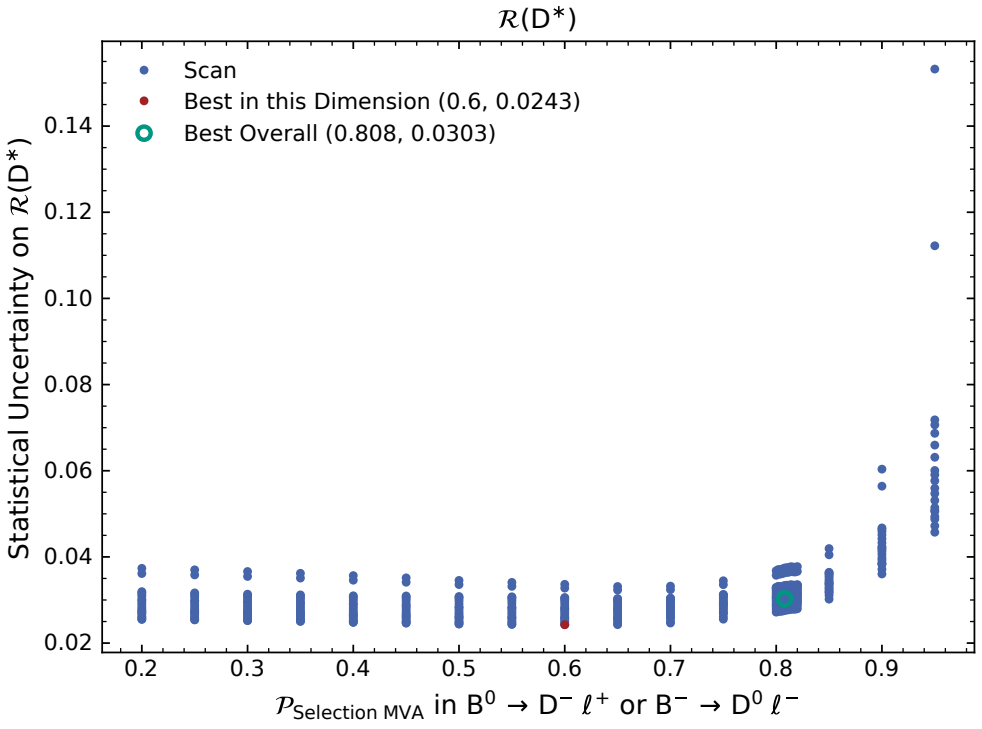


(a) Scan of selection requirement on $\mathcal{P}_{\text{Sel}}$ in the reconstruction modes with **D meson**.

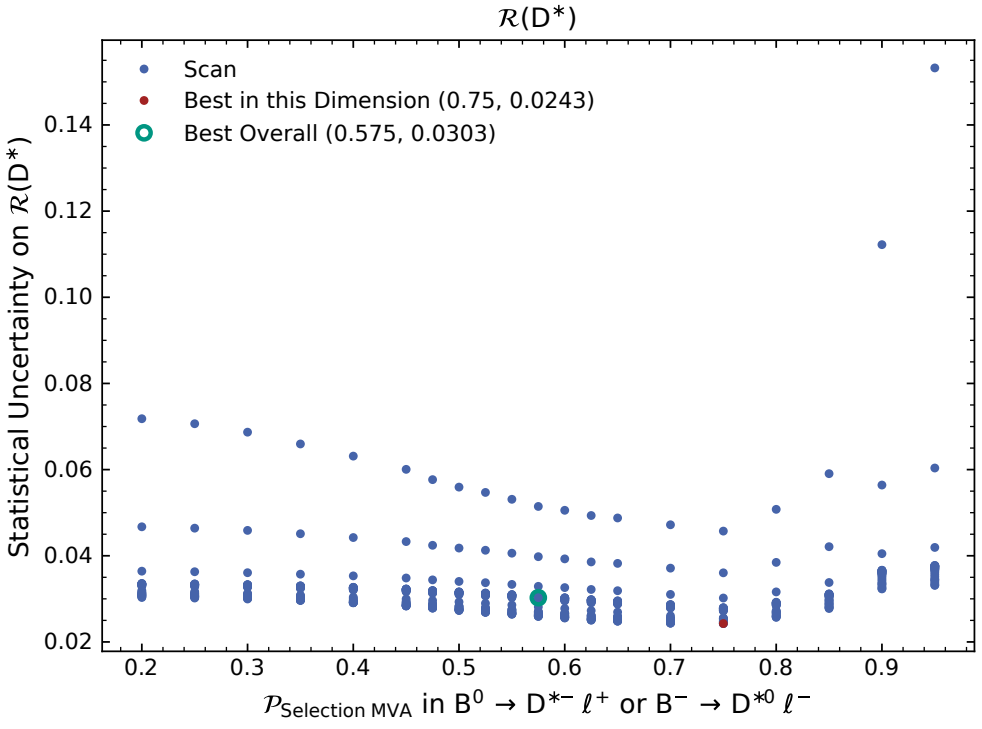


(b) Scan of selection requirement on $\mathcal{P}_{\text{Sel}}$ in the reconstruction modes with **excited D^* meson**.

Figure B.3.: Visualization of the grid search performed for the optimization of the selection criterion on the selection MVA output $\mathcal{P}_{\text{Sel}}$. Shown are the results for the statistical fit uncertainty on **the ratio $\mathcal{R}$(D)**. A detailed description can be found in the text of Appendix B.2. Continued in Figure B.4.



(a) Scan of selection requirement on $\mathcal{P}_{\text{Sel}}$ in the reconstruction modes with **D meson**.



(b) Scan of selection requirement on $\mathcal{P}_{\text{Sel}}$ in the reconstruction modes with **excited D^* meson**.

Figure B.4.: Continuation of Figure B.3: Visualization of the grid search performed for the optimization of the selection criterion on the selection MVA output $\mathcal{P}_{\text{Sel}}$. Shown are the results for the statistical fit uncertainty on **the ratio $\mathcal{R}(D^*)$**. A detailed description can be found in the text of Appendix B.2.

C. Supplementary Material for the $\mathcal{R}(D^{(*)})$ Asimov Fit

This appendix contains complementary material to the MC studies for the ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ using an Asimov data set presented in Chapter 8. It contains results for studies conducted with looser signal selection criteria, as well as using the old gap description (see Section 3.2.3) in Appendix C.1.
Furthermore, the remaining distributions for the templates in the reconstruction modes $D^+\ell^-$, $D^{*0}\ell^-$ and $D^{*+}\ell^-$, which are not included in the main body of this thesis, are supplied in Appendix C.2.

C.1. Results of $\mathcal{R}(D^{(*)})$ Asimov Fit with Alternative Setups

Table C.4 compares the results of Asimov studies using the setup as shown in Section 8.5 and for a MC sample using the old gap description without intermediate D^{**} resonances. Both of these gap definitions are discussed in Section 3.2.3. The difference is only notable in the statistical fit uncertainy of the $\mathcal{R}(D)$ result. The difference between the two gap descriptions will be used to determine systematic uncertainties on the shapes of the gap description in future steps of the analysis. This effect is not considered, however, at the time of writing of this text.
The results of Asimov fits using the loose selection requirement defined in Equation (6.13) are shown in Table C.5 again for both options for the gap description. This study is performed to evaluate the reduction in precision, if the validation of the selection BDT on sideband samples were to show data-MC discrepancies which would render the selection BDT not usable. The increase in statistical fit uncertainty is particular notable in the $\mathcal{R}(D)$ result with a rise of more than 70 %, or about 14 percent points. The effect is less severe for the $\mathcal{R}(D^*)$ result, which is expected, as the optimized selection requirement on $\mathcal{P}_{\text{Sel}}$ is less stringent for the $B \to D^*\ell$ reconstruction modes.

Table C.4.: Tabulated results for the branching fraction ratios $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ obtained from a fit to an artificial Asimov data set for which the SM expectation values (see Table 1.1) of the ratios are assumed. Two definitions for the description of the gap MC samples are compared, where **Old Gap** denotes the description without intermediate D^{**} resonance and **New Gap** is the alternative gap sample which was also used for the main study shown in Section 8.5 The fit reproduces the SM expectation for both setups. Additionally, the statistical fit uncertainty, as well as the combined multiplicative systematic uncertainties are listed as relative to the nominal results. For the latter, the individual contribution of the various sources of multiplicative systematic uncertainties discussed in Section 8.4 are given as well. The relative uncertainty values are listed with a higher number of decimal places than required, to make also the negligible contributions visible for the purpose of this Asimov study.

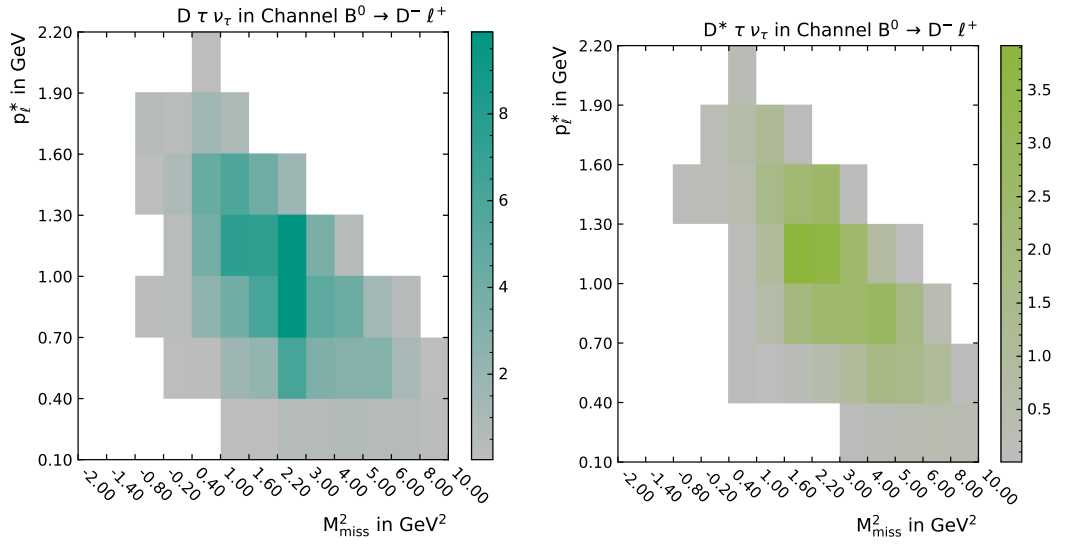
Opt. Selection	**Relative Uncertainty in Percent**			
	$\boldsymbol{\mathcal{R}(D)}$		$\boldsymbol{\mathcal{R}(D^*)}$	
Source	Old Gap	New Gap	Old Gap	New Gap
Nominal Value	0.299	0.299	0.258	0.258
Stat. Uncertainty	19.987 %	20.666 %	12.204 %	12.202 %
Mult. Sys. Uncertainty	1.525 %	1.525 %	1.208 %	1.207 %
MC Statistics	1.299 %	1.299 %	0.916 %	0.916 %
Lepton ID	0.784 %	0.783 %	0.769 %	0.768 %
Pion ID	0.027 %	0.027 %	0.032 %	0.032 %
Fake Pion	0.005 %	0.005 %	0.008 %	0.008 %
Kaon ID	0.038 %	0.038 %	0.040 %	0.040 %
Kaon Fake	0.033 %	0.034 %	0.015 %	0.015 %
Slow Pion	0.010 %	0.010 %	0.029 %	0.028 %
K^0_S Reconstruction	0.005 %	0.005 %	0.002 %	0.002 %
$B \to D\ell\nu$ Form Factors	0.146 %	0.146 %	0.000 %	0.000 %
$B \to D^*\ell\nu$ Form Factors	0.000 %	0.000 %	0.155 %	0.155 %

Table C.5.: Tabulated results for the branching fraction ratios $\mathcal{R}(\mathrm{D})$ and $\mathcal{R}(\mathrm{D}^*)$ obtained from a fit to an artificial Asimov data set for which the SM expectation values (see Table 1.1) of the ratios are assumed and a loose selection requirement of $\mathcal{P}_{\mathrm{Sel}} > 0.2$ is used instead of the optimized criteria given in Equation (6.22). Two definitions for the description of the gap MC samples are compared, where **Old Gap** denotes the description without intermediate D^{**} resonance and **New Gap** is the alternative gap sample which was also used for the main study shown in Section 8.5 The fit reproduces the SM expectation for both setups. Additionally, the statistical fit uncertainty, as well as the combined multiplicative systematic uncertainties are listed as relative to the nominal results. For the latter, the individual contribution of the various sources of multiplicative systematic uncertainties discussed in Section 8.4 are given as well. The relative uncertainty values are listed with a higher number of decimal places than required, to make also the negligible contributions visible for the purpose of this Asimov study.

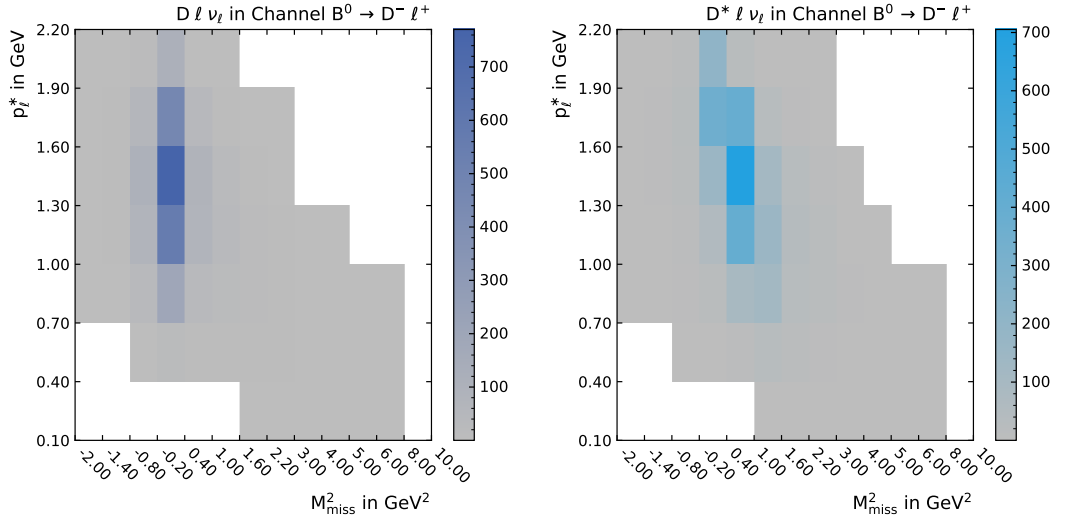
Loose Selection	**Relative Uncertainty in Percent**			
	$\boldsymbol{\mathcal{R}(\mathrm{D})}$		$\boldsymbol{\mathcal{R}(\mathrm{D}^*)}$	
Source	Old Gap	New Gap	Old Gap	New Gap
Nominal Value	0.299	0.299	0.258	0.258
Stat. Uncertainty	34.272 %	33.739 %	13.527 %	12.738 %
Mult. Sys. Uncertainty	1.241 %	1.240 %	1.010 %	1.009 %
MC Statistics	1.011 %	1.011 %	0.718 %	0.718 %
Lepton ID	0.695 %	0.693 %	0.694 %	0.693 %
Pion ID	0.017 %	0.017 %	0.021 %	0.021 %
Fake Pion	0.006 %	0.006 %	0.006 %	0.006 %
Kaon ID	0.043 %	0.043 %	0.036 %	0.036 %
Kaon Fake	0.055 %	0.056 %	0.026 %	0.025 %
Slow Pion	0.013 %	0.013 %	0.016 %	0.016 %
$\mathrm{K}^0_{\mathrm{S}}$ Reconstruction	0.003 %	0.003 %	0.004 %	0.004 %
$\mathrm{B} \to \mathrm{D}\ell\nu$ Form Factors	0.176 %	0.176 %	0.000 %	0.000 %
$\mathrm{B} \to \mathrm{D}^*\ell\nu$ Form Factors	0.000 %	0.000 %	0.143 %	0.143 %

C.2. Templates for Remaining Reconstruction Modes

This appendix shows the component distributions for the remaining reconstruction modes $D^+\ell^-$, $D^{*0}\ell^-$ and $D^{*+}\ell^-$ as stated in Section 8.3. Please refer to the text of this section for a detailed description.

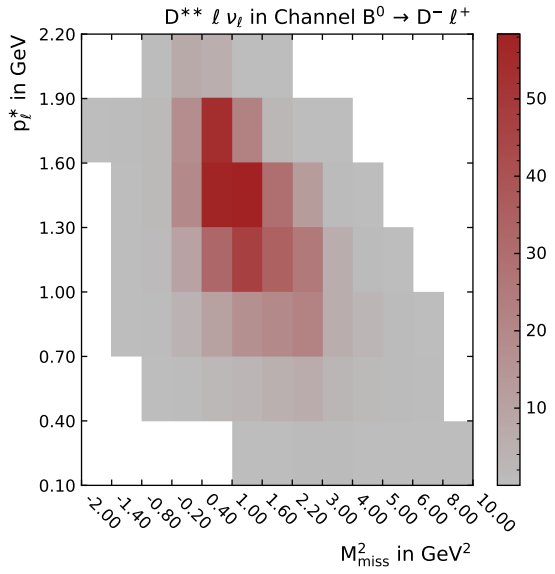


(a) $B \to D\tau\nu$ template in $B^0 \to D^-\ell^+$ mode (b) $B \to D^*\tau\nu$ template in $B^0 \to D^-\ell^+$ mode

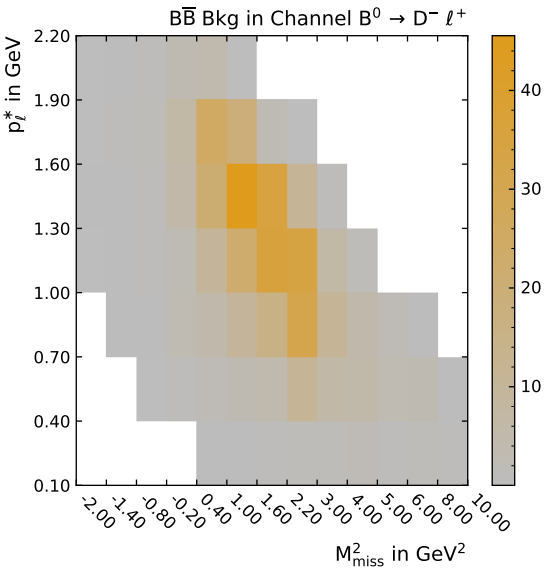


(c) $B \to D\ell\nu$ template in $B^0 \to D^-\ell^+$ mode (d) $B \to D^*\ell\nu$ template in $B^0 \to D^-\ell^+$ mode

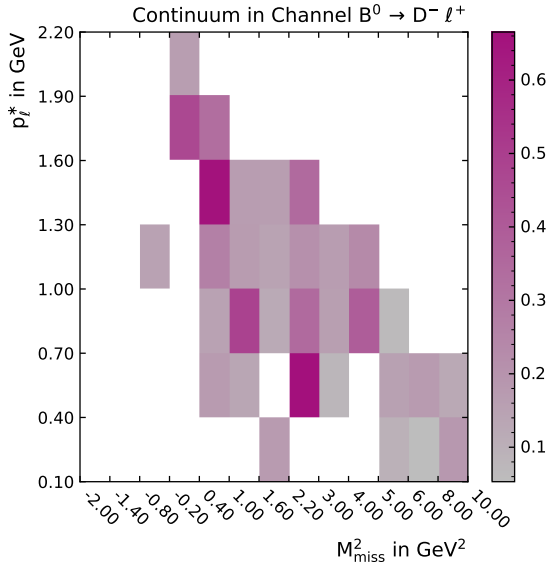
Figure C.5.: Histogrammed distributions used for the templates for the reconstruction channel $B^0 \to D^-\ell^+$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts. Distributions of the three background components for this channel are shown in the continuation of this figure on page 149.



(e) $B \to D^{**}\ell\nu$ template in $B^0 \to D^-\ell^+$ mode

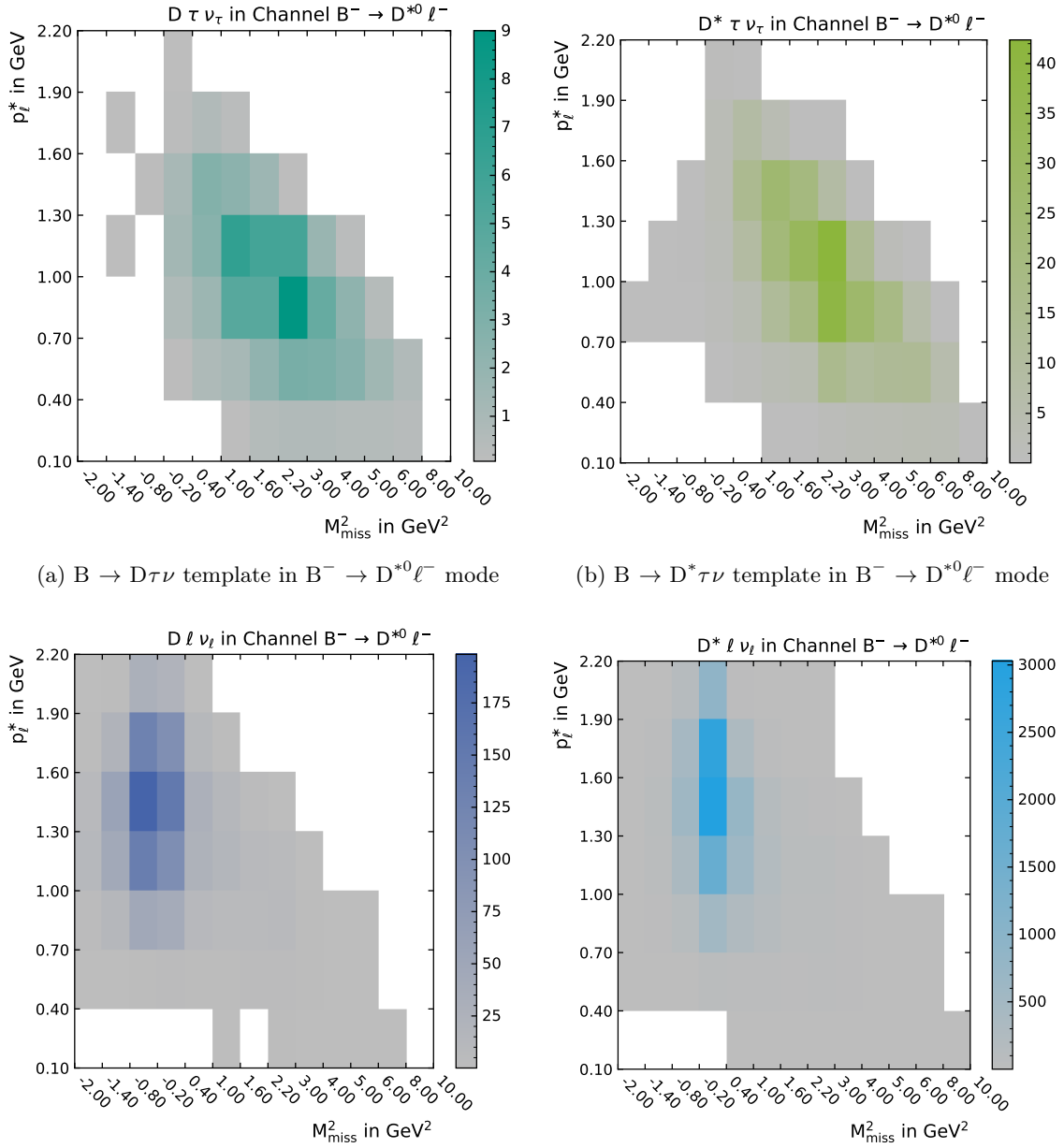


(f) $B\overline{B}$ background template in $B^0 \to D^-\ell^+$ mode



(g) Continuum background template in $B^0 \to D^-\ell^+$ mode

Figure C.5.: Continuation of Figure C.5 on page 148.



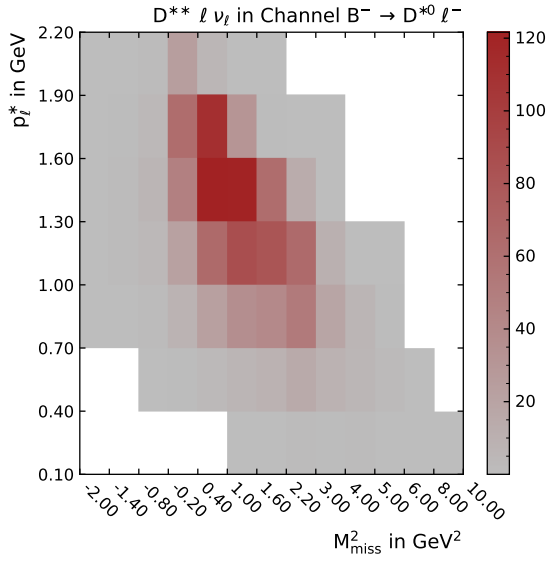
(a) B → Dτν template in $B^- \to D^{*0}\ell^-$ mode

(b) B → D*τν template in $B^- \to D^{*0}\ell^-$ mode

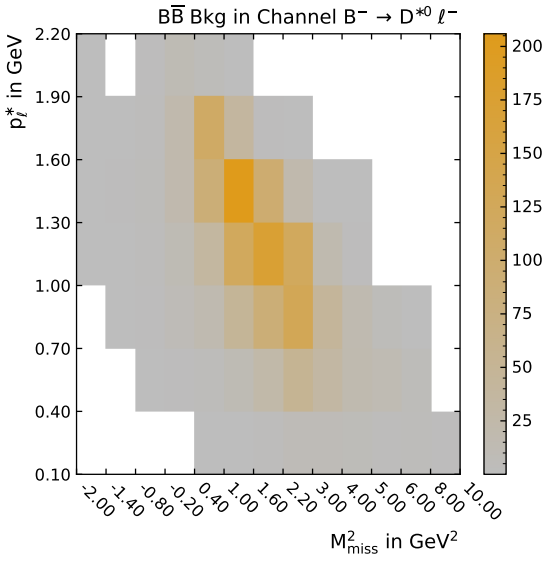
(c) B → Dℓν template in $B^- \to D^{*0}\ell^-$ mode

(d) B → D*ℓν template in $B^- \to D^{*0}\ell^-$ mode

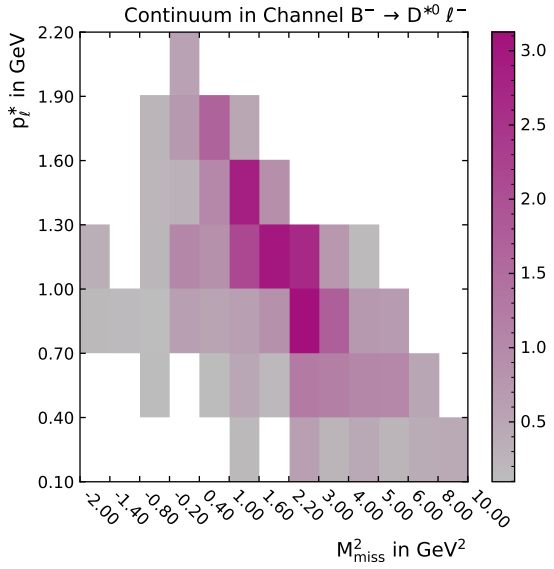
Figure C.6.: Histogrammed distributions used for the templates for the reconstruction channel $B^- \to D^{*0}\ell^-$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts. Distributions of the three background components for this channel are shown in the continuation of this figure on page 151.



(e) $B \to D^{**}\ell\nu$ template in $B^- \to D^{*0}\ell^-$ mode

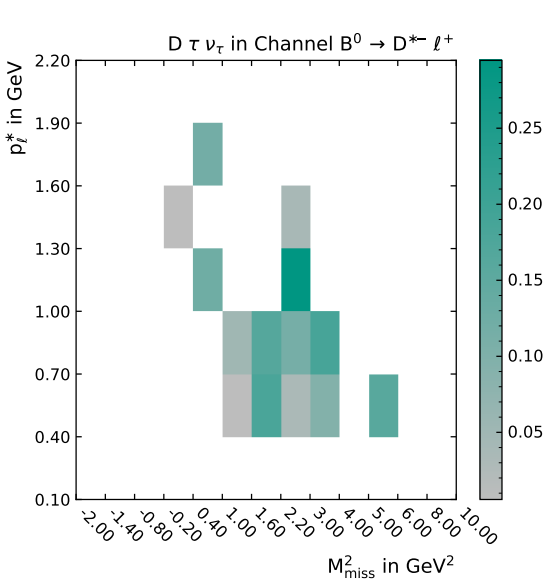


(f) $B\overline{B}$ background template in $B^- \to D^{*0}\ell^-$ mode

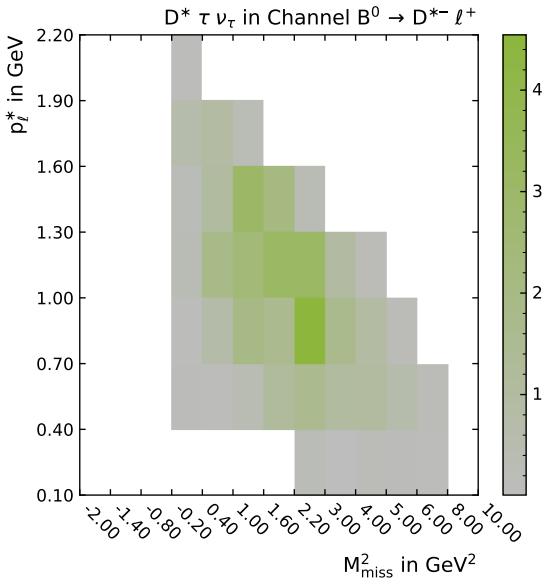


(g) Continuum background template in $B^- \to D^{*0}\ell^-$ mode

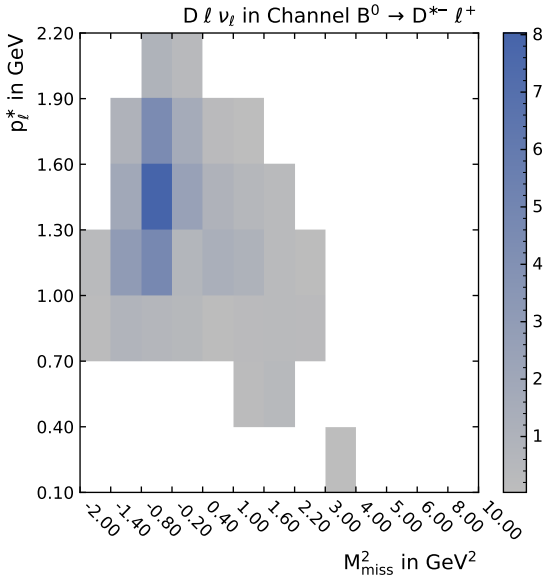
Figure C.6.: Continuation of Figure C.6 on page 150.



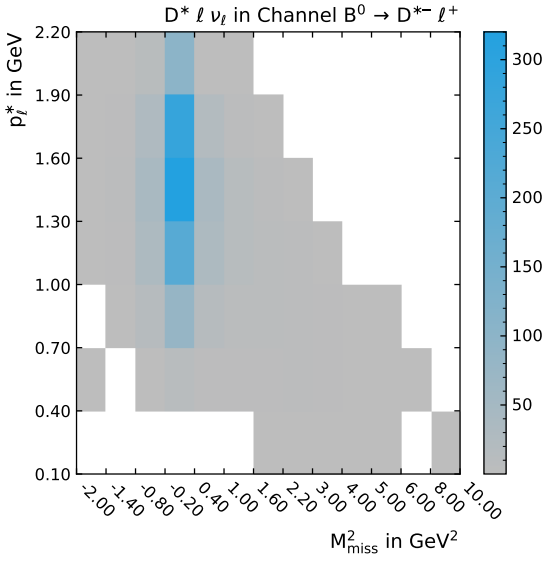
(a) $B \to D\tau\nu$ template in $B^0 \to D^{*-}\ell^+$ mode



(b) $B \to D^*\tau\nu$ template in $B^0 \to D^{*-}\ell^+$ mode

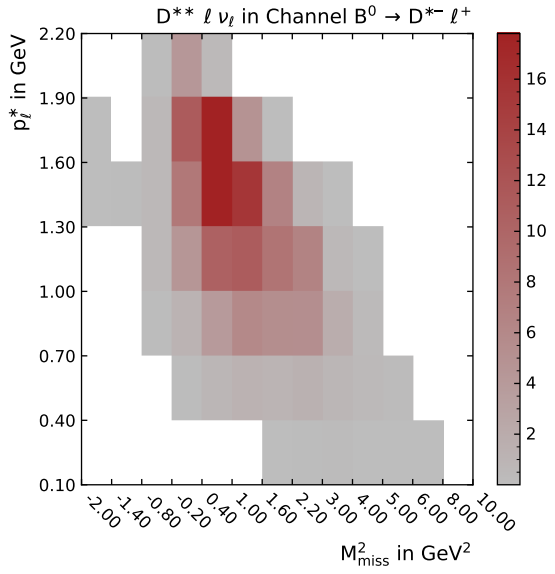


(c) $B \to D\ell\nu$ template in $B^0 \to D^{*-}\ell^+$ mode

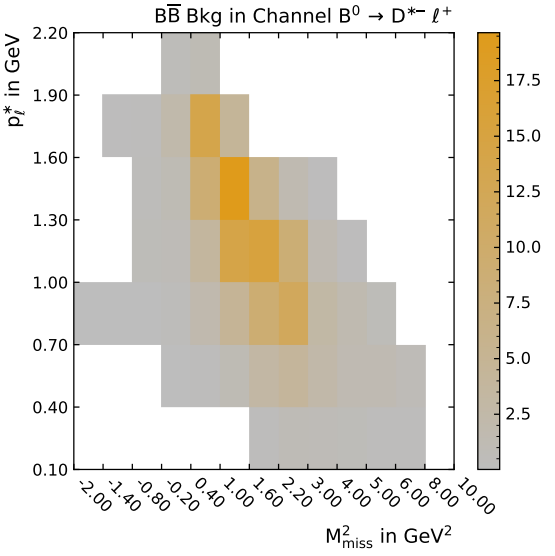


(d) $B \to D^*\ell\nu$ template in $B^0 \to D^{*-}\ell^+$ mode

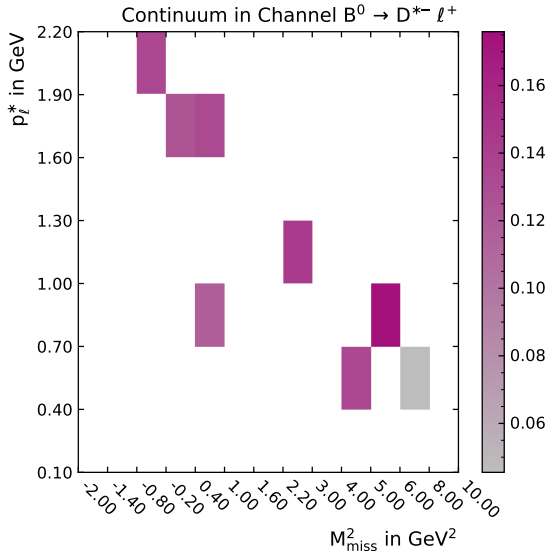
Figure C.7.: Histogrammed distributions used for the templates for the reconstruction channel $B^0 \to D^{*-}\ell^+$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts. Distributions of the three background components for this channel are shown in the continuation of this figure on page 153.



(e) $B \to D^{**}\ell\nu$ template in $B^0 \to D^{*-}\ell^+$ mode



(f) $B\overline{B}$ background template in $B^0 \to D^{*-}\ell^+$ mode

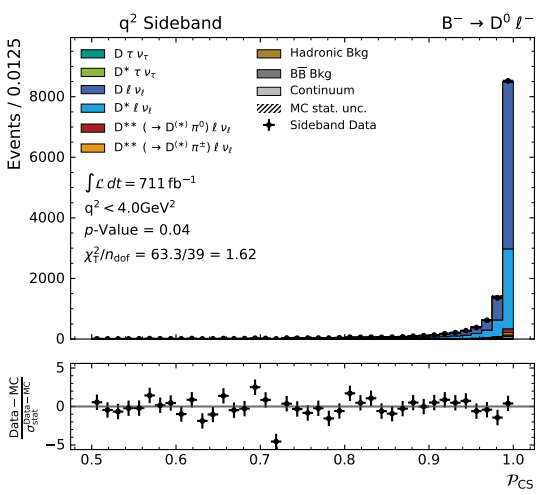


(g) Continuum background template in $B^0 \to D^{*-}\ell^+$ mode

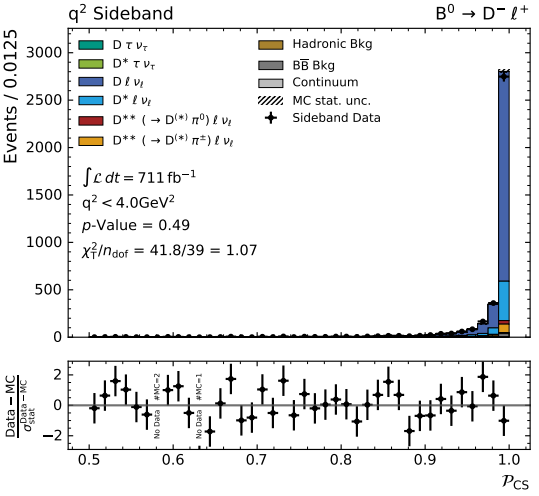
Figure C.7.: Continuation of Figure C.7 on page 152.

D. Additional Data-MC Comparison Results

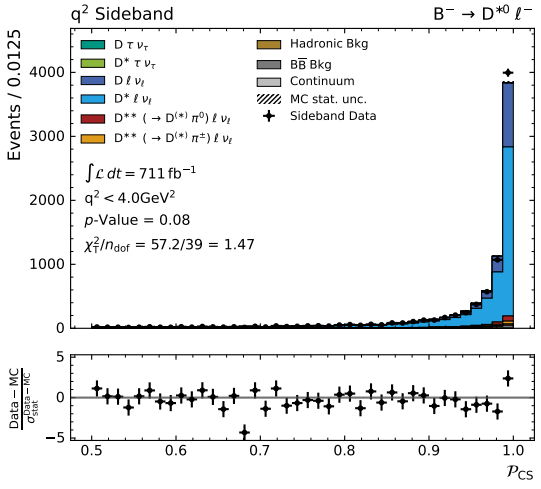
D.1. Additional Continuum Suppression BDT Output Validation Plots



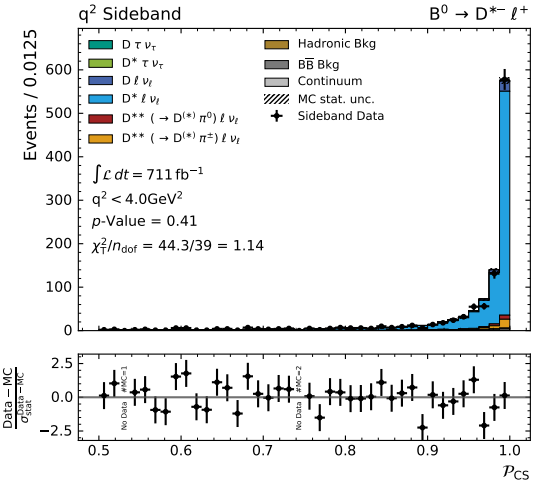
(a) $B^- \to D^0 \ell^-$ + c.c.



(b) $B^0 \to D^- \ell^+$ + c.c.

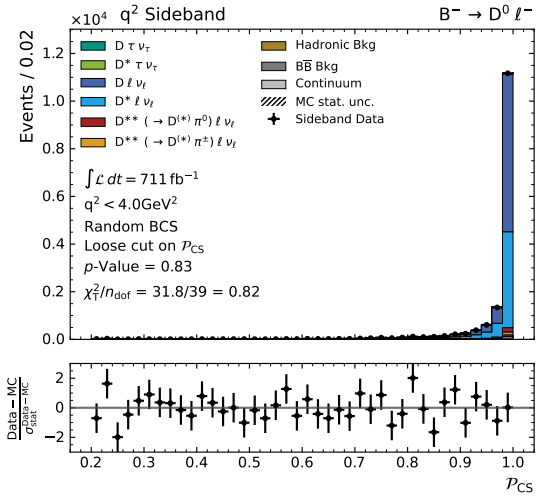


(c) $B^- \to D^{*0} \ell^-$ + c.c.

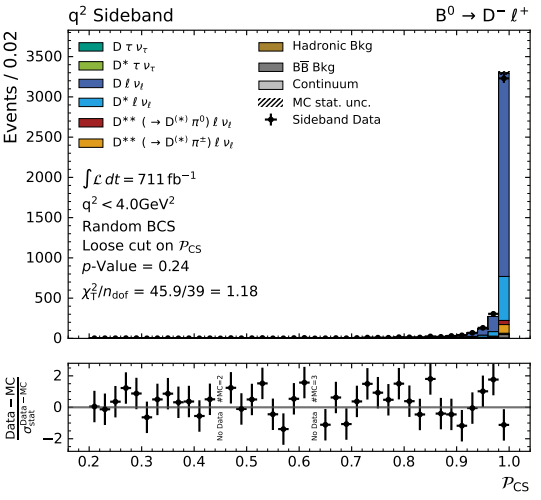


(d) $B^0 \to D^{*-} \ell^+$ + c.c.

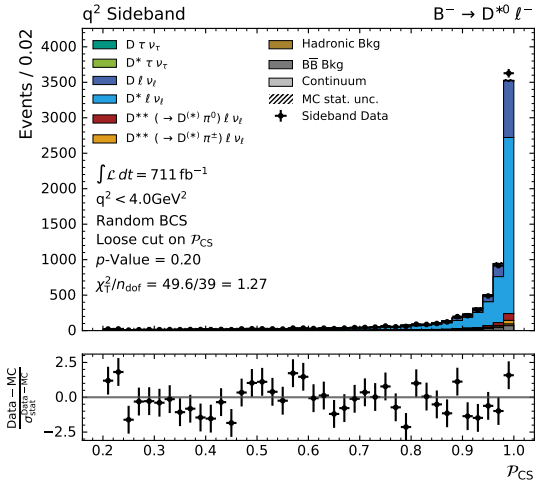
Figure D.8.: Data-MC comparison of the continuum classifier output $\mathcal{P}_{\text{CS}}$ for the individual reconstruction modes in the q^2-Sideband samples. Please refer to the introduction of Chapter 9 for a detailed description of the included information. The interpretation of the plots is given in Section 9.1.



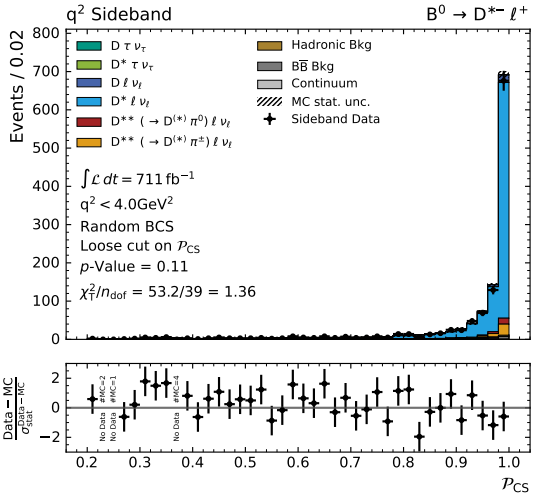
(a) $B^- \to D^0 \ell^-$ + c.c.



(b) $B^0 \to D^- \ell^+$ + c.c.

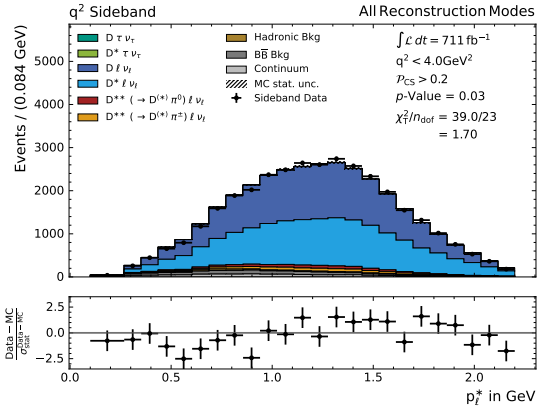


(c) $B^- \to D^{*0} \ell^-$ + c.c.

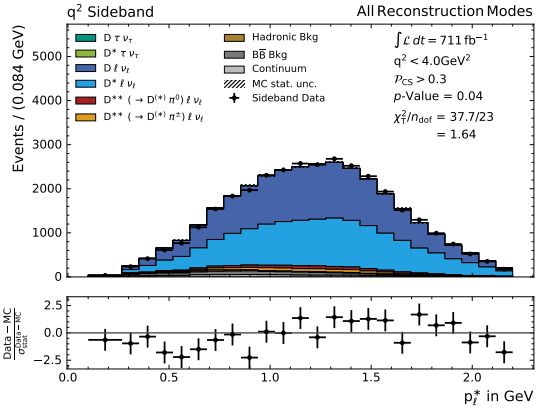


(d) $B^0 \to D^{*-} \ell^+$ + c.c.

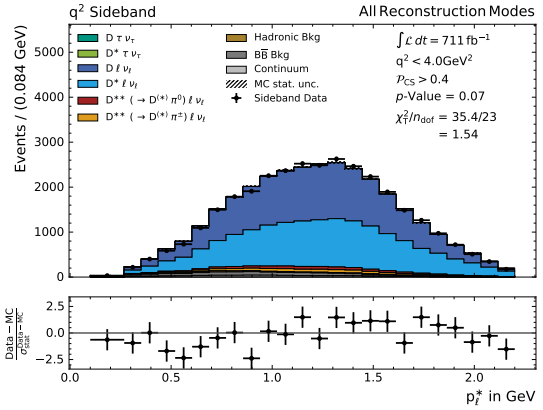
Figure D.9.: Data-MC comparison of the continuum classifier output $\mathcal{P}_{\text{CS}}$ for the individual reconstruction modes in the q^2-Sideband samples. For this comparison, a looser selection requirement on continuum suppression output of $\mathcal{P}_{\text{CS}} > 0.2$ is applied and a purely random best candidate selection is used. Please refer to the introduction of Chapter 9 for a detailed description of the included information. The interpretation of the plots is given in Section 9.1.



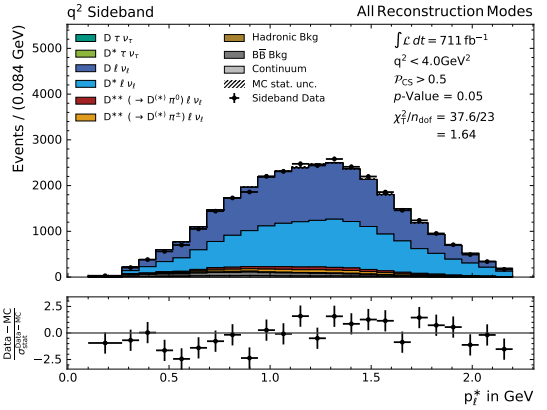
(a) $\mathcal{P}_{\text{CS}} > 0.2$



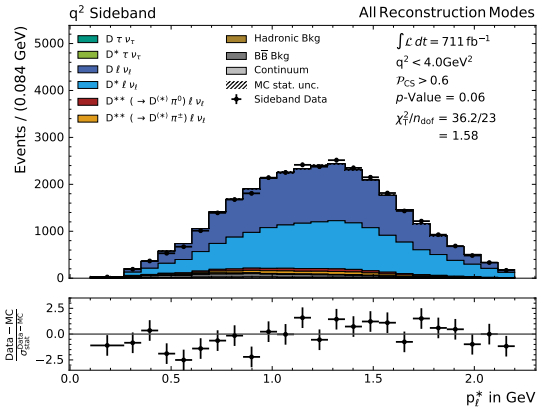
(b) $\mathcal{P}_{\text{CS}} > 0.3$



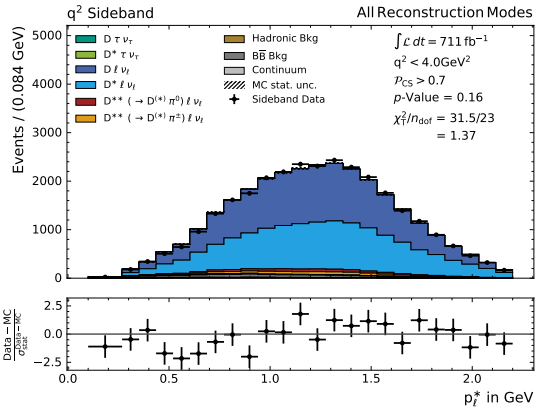
(c) $\mathcal{P}_{\text{CS}} > 0.4$



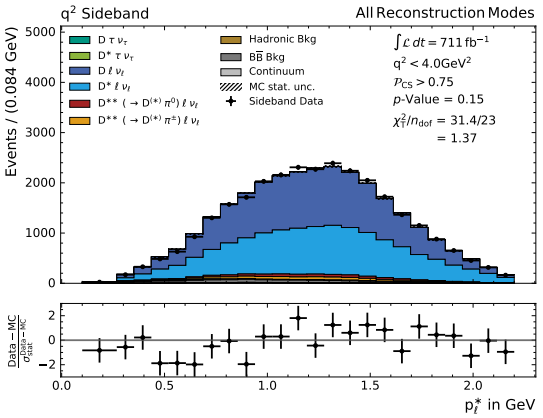
(d) $\mathcal{P}_{\text{CS}} > 0.5$



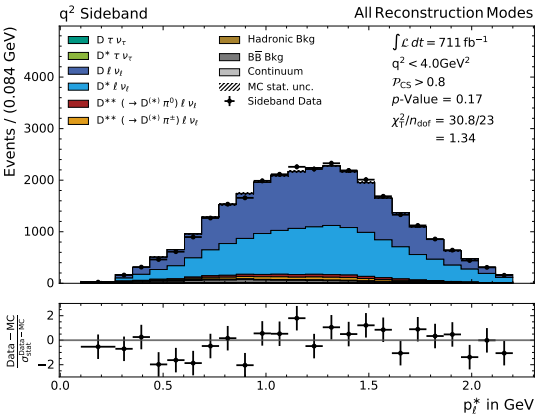
(e) $\mathcal{P}_{\text{CS}} > 0.6$



(f) $\mathcal{P}_{\text{CS}} > 0.7$

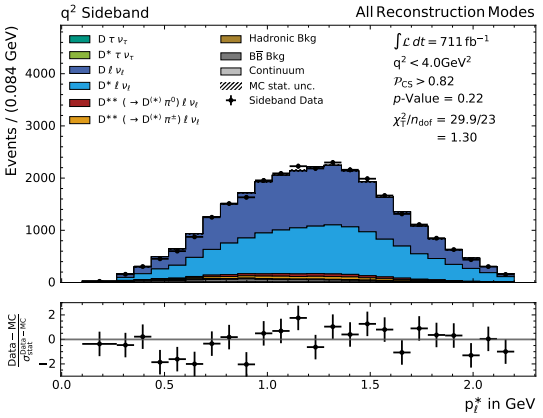


(g) $\mathcal{P}_{\text{CS}} > 0.75$

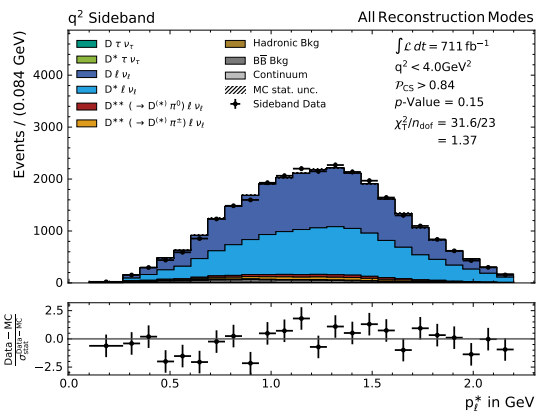


(h) $\mathcal{P}_{\text{CS}} > 0.8$

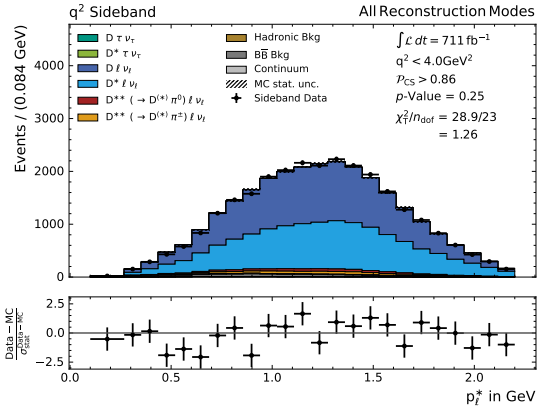
Figure D.10.: See page 159 for caption.



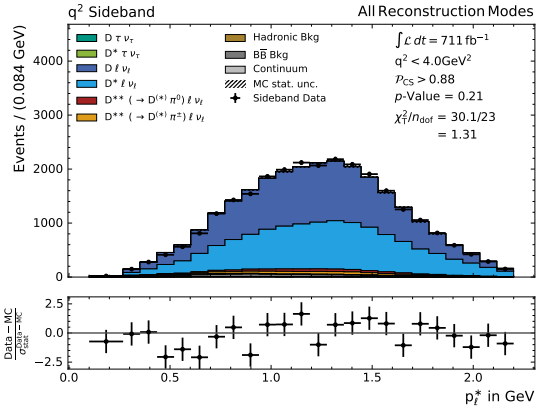
(i) $\mathcal{P}_{\text{CS}} > 0.82$



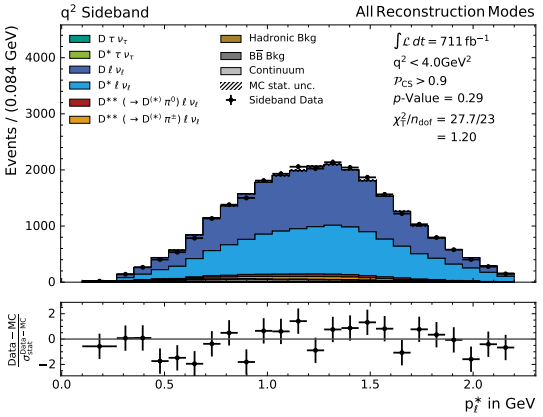
(j) $\mathcal{P}_{\text{CS}} > 0.84$



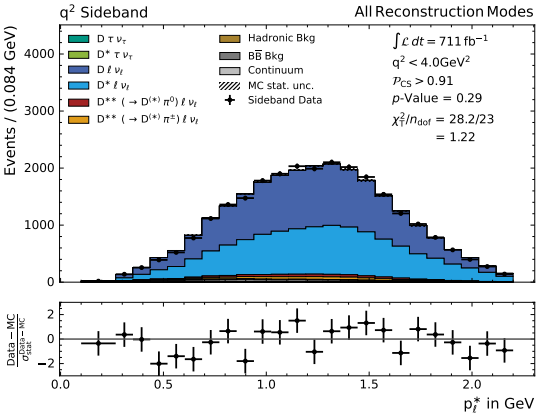
(k) $\mathcal{P}_{\text{CS}} > 0.86$



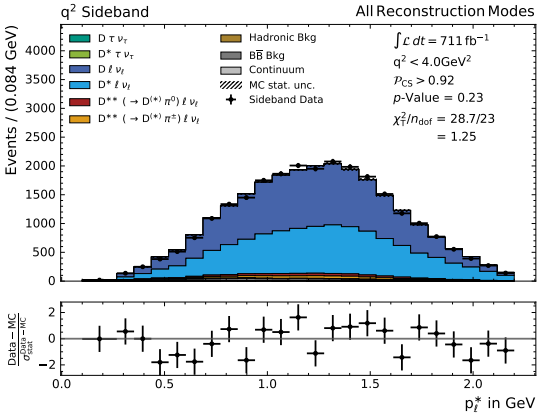
(l) $\mathcal{P}_{\text{CS}} > 0.88$



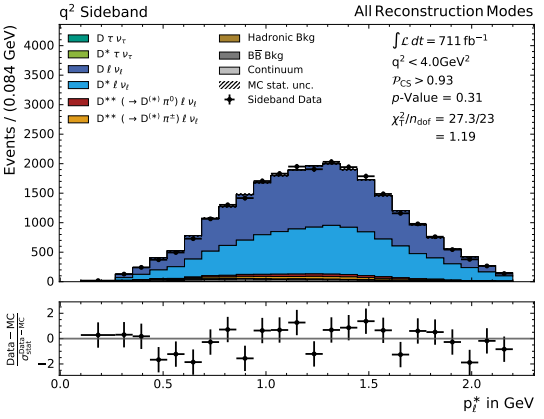
(m) $\mathcal{P}_{\text{CS}} > 0.9$



(n) $\mathcal{P}_{\text{CS}} > 0.91$

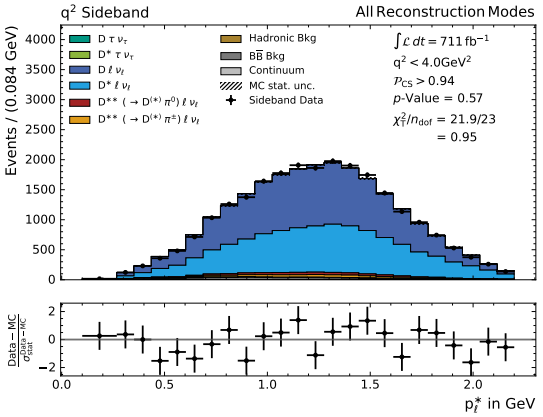


(o) $\mathcal{P}_{\text{CS}} > 0.92$

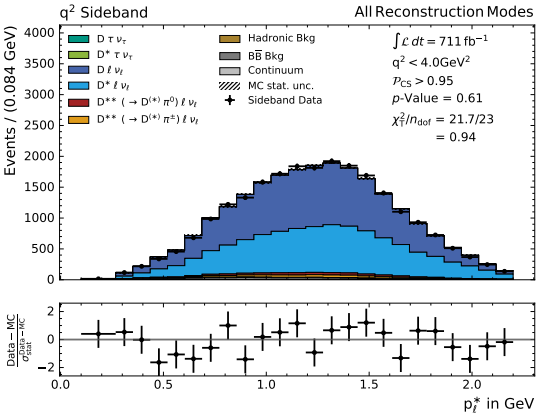


(p) $\mathcal{P}_{\text{CS}} > 0.93$

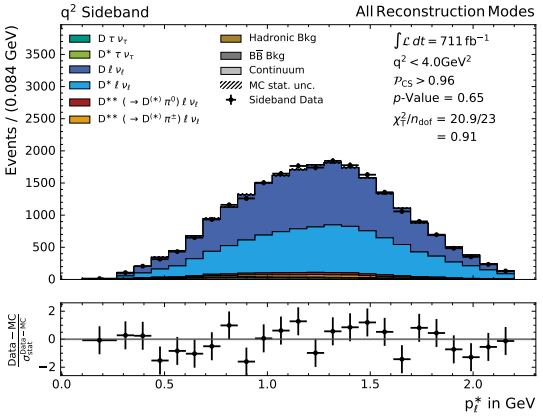
Figure D.10.: See page 159 for caption.



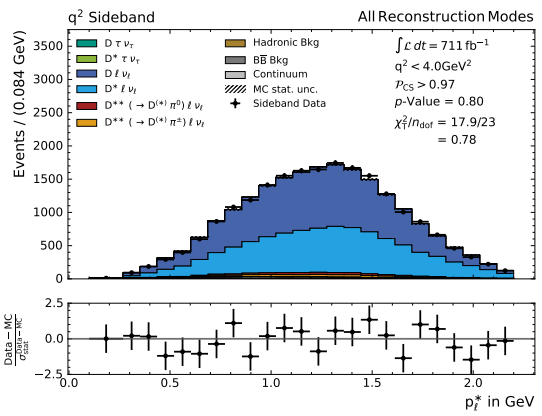
(q) $\mathcal{P}_{\text{CS}} > 0.94$



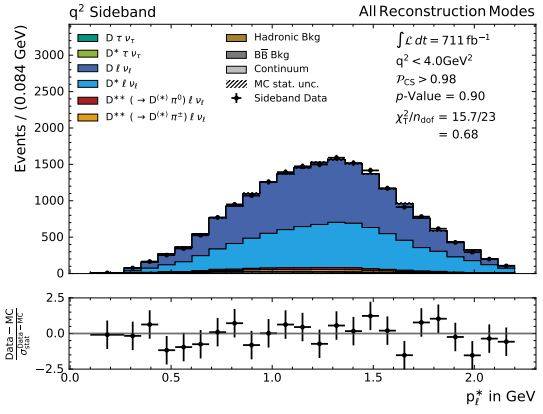
(r) $\mathcal{P}_{\text{CS}} > 0.95$



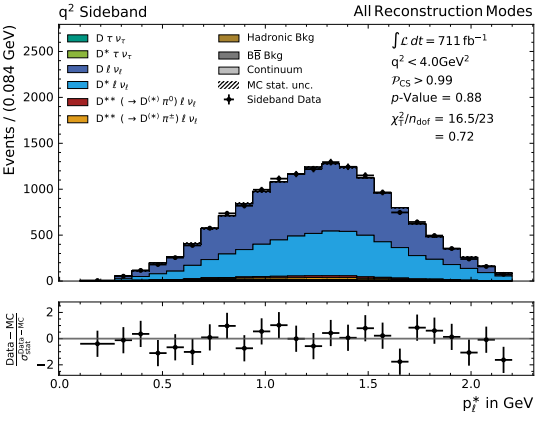
(s) $\mathcal{P}_{\text{CS}} > 0.96$



(t) $\mathcal{P}_{\text{CS}} > 0.97$

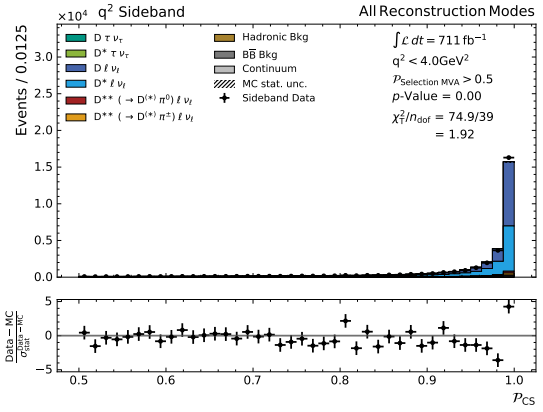


(u) $\mathcal{P}_{\text{CS}} > 0.98$

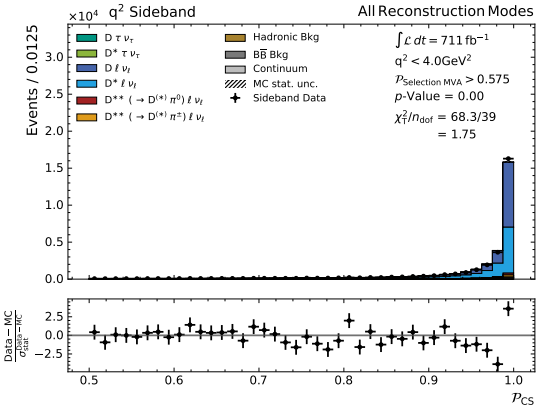


(v) $\mathcal{P}_{\text{CS}} > 0.99$

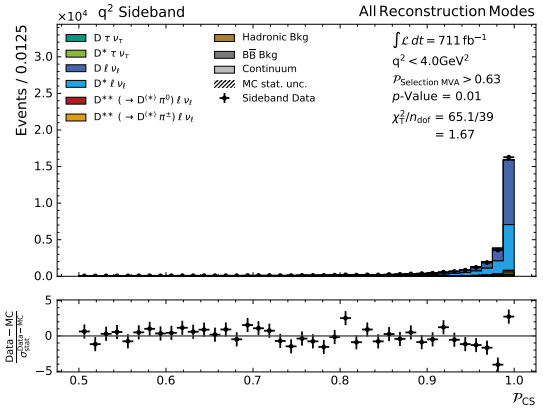
Figure D.10.: Evaluation of the effect of increasingly tight requirements on the continuum suppression classifier output $\mathcal{P}_{\text{CS}}$ on the shape of the fit observable p_ℓ^* in recorded and simulated data. The histogram of the simulated data is rescaled, such that the number of events matches the recorded sample. The thresholds on $\mathcal{P}_{\text{CS}}$ from 0.20 to 0.99 are shown in the sub-captions. No significant change in shape or different behavior for data and MC is visible.



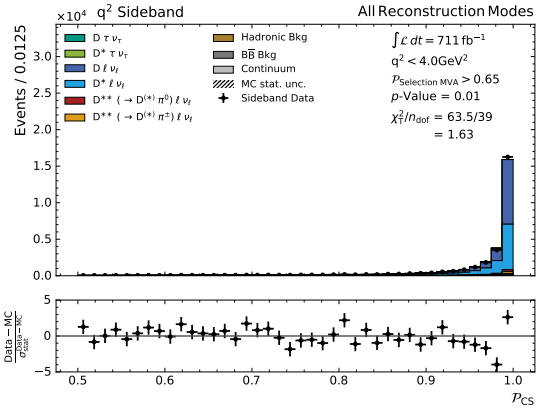
(a) $\mathcal{P}_{\text{Sel}} > 0.5$



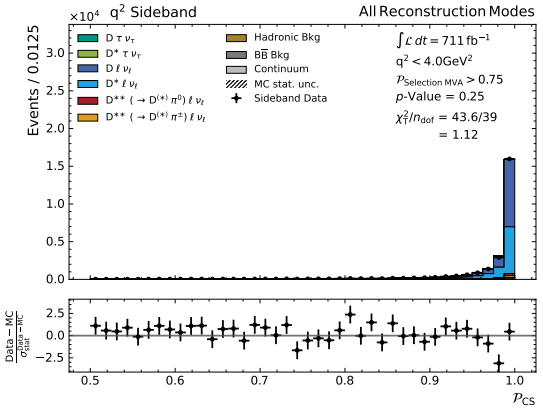
(b) $\mathcal{P}_{\text{Sel}} > 0.575$



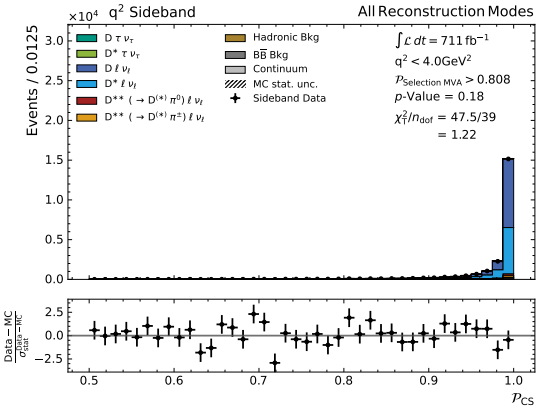
(c) $\mathcal{P}_{\text{Sel}} > 0.63$



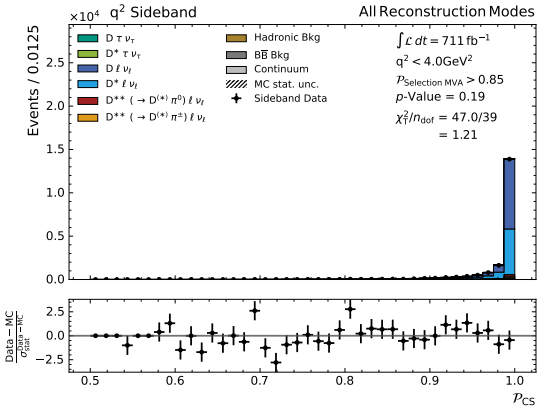
(d) $\mathcal{P}_{\text{Sel}} > 0.65$



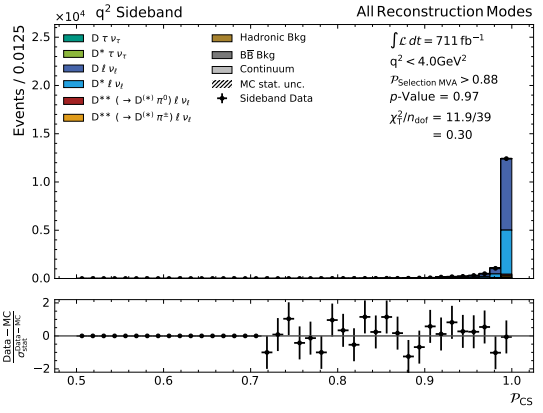
(e) $\mathcal{P}_{\text{Sel}} > 0.75$



(f) $\mathcal{P}_{\text{Sel}} > 0.808$

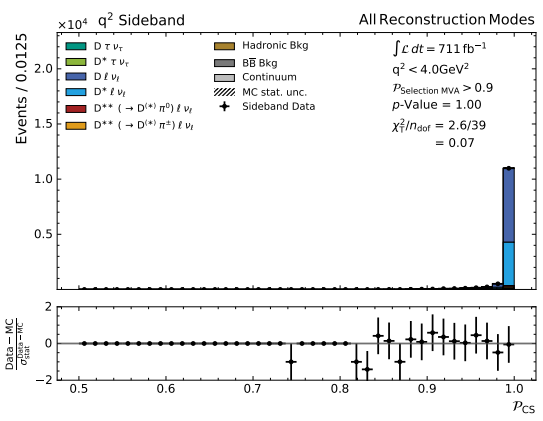


(g) $\mathcal{P}_{\text{Sel}} > 0.85$

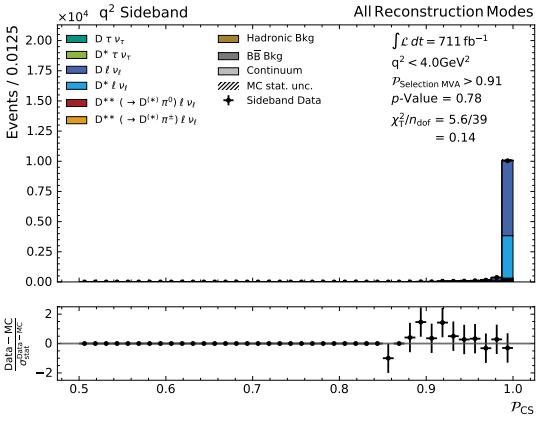


(h) $\mathcal{P}_{\text{Sel}} > 0.88$

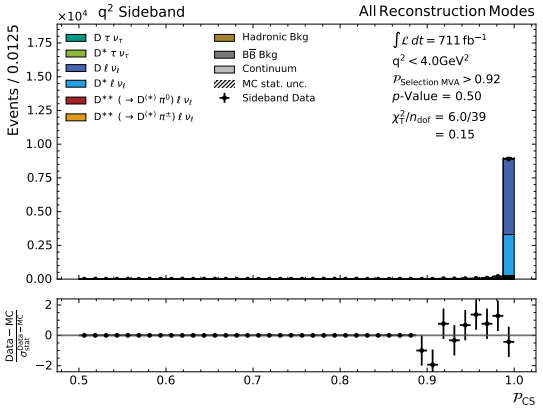
Figure D.11.: See next page for caption.



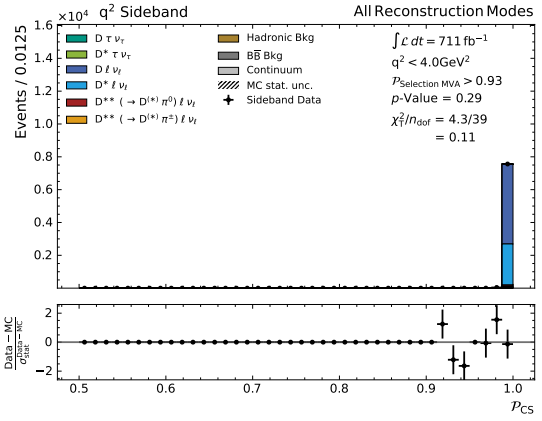
(i) $\mathcal{P}_{\text{Sel}} > 0.9$



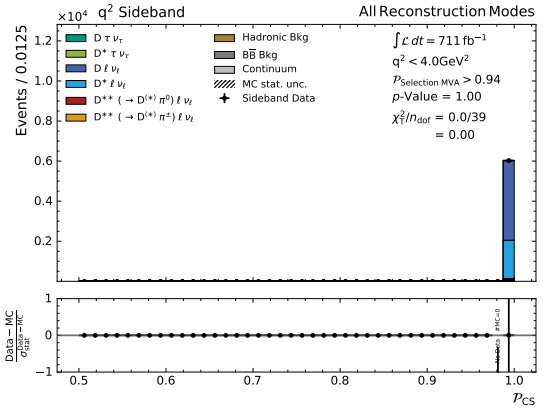
(j) $\mathcal{P}_{\text{Sel}} > 0.91$



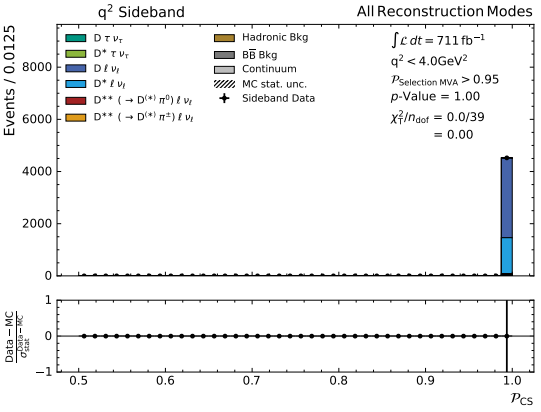
(k) $\mathcal{P}_{\text{Sel}} > 0.92$



(l) $\mathcal{P}_{\text{Sel}} > 0.93$



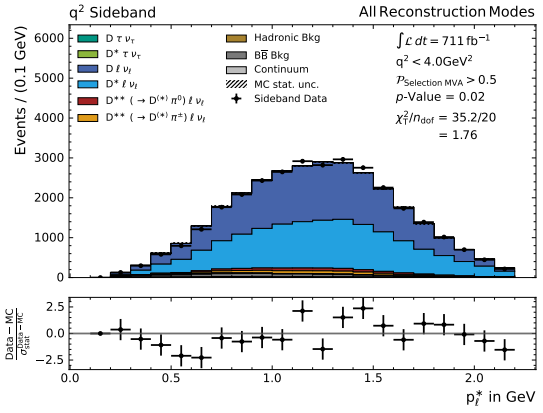
(m) $\mathcal{P}_{\text{Sel}} > 0.94$



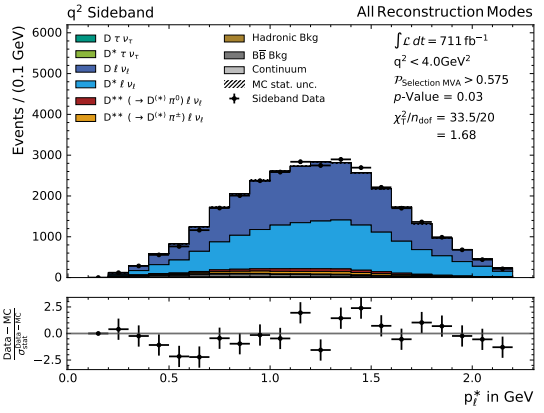
(n) $\mathcal{P}_{\text{Sel}} > 0.95$

Figure D.11.: Evaluation of the effect of increasingly tight requirements on the selection classifier output $\mathcal{P}_{\text{Sel}}$ on the shape of the continuum suppression classifier output $\mathcal{P}_{\text{CS}}$ distribution in recorded and simulated data. The histogram of the simulated data is rescaled, such that the number of events matches the recorded sample. The thresholds on $\mathcal{P}_{\text{Sel}}$ from 0.50 to 0.95 are shown in the sub-captions. No significant change in shape or different behavior for data and MC is visible.

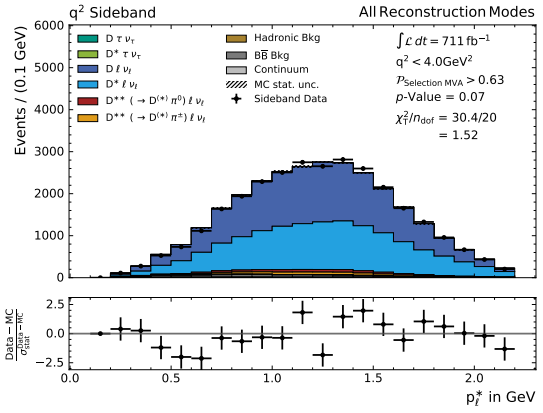
D.2. Additional Selection Classifier Output Validation Plots



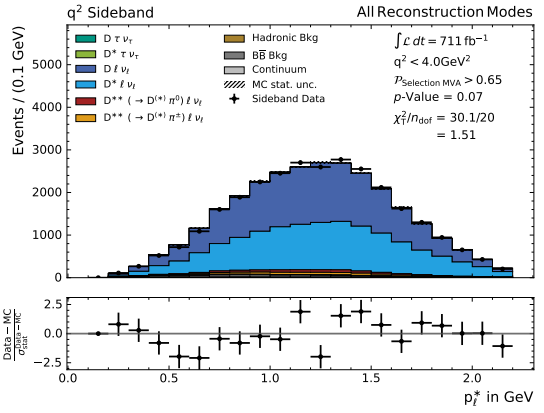
(a) $\mathcal{P}_{\text{Sel}} > 0.5$



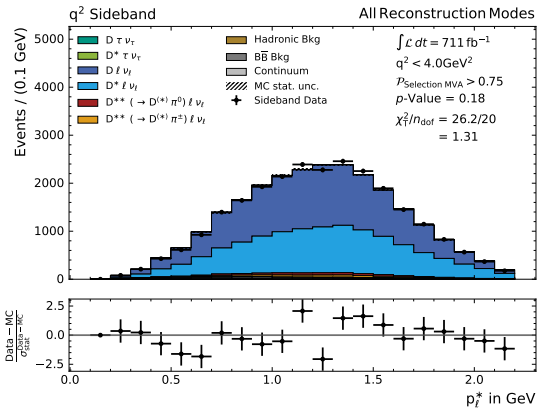
(b) $\mathcal{P}_{\text{Sel}} > 0.575$



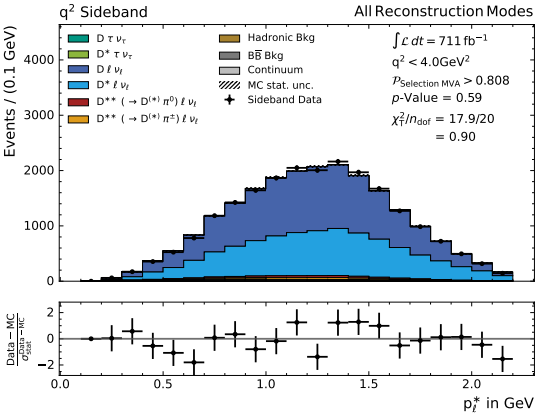
(c) $\mathcal{P}_{\text{Sel}} > 0.63$



(d) $\mathcal{P}_{\text{Sel}} > 0.65$

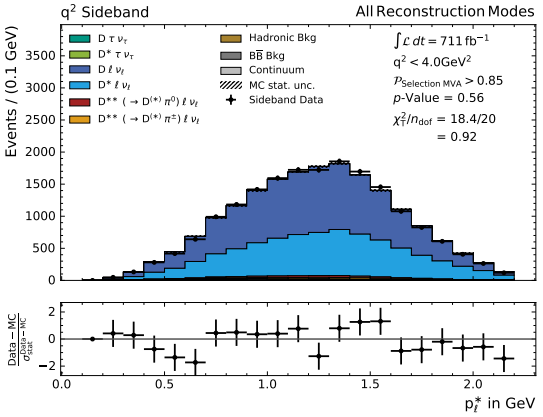


(e) $\mathcal{P}_{\text{Sel}} > 0.75$

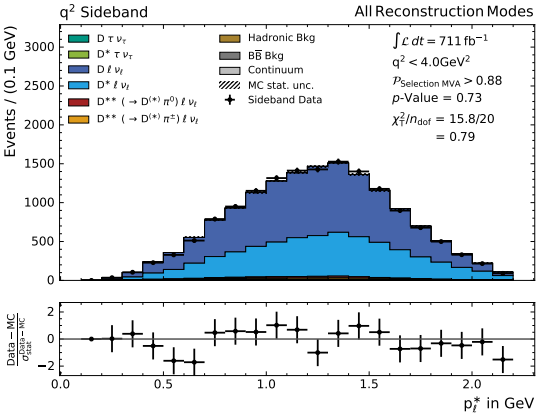


(f) $\mathcal{P}_{\text{Sel}} > 0.808$

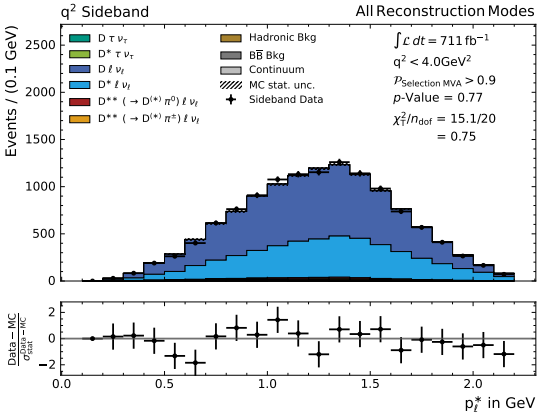
Figure D.12.: Effect of increasingly tight BDT output $\mathcal{P}_{\text{Sel}}$ selection criteria on the shape of the lepton momentum p^*_ℓ in the rest frame of the signal side B_{sig} meson distribution in recorded and simulated data. The histogram of the simulated data is rescaled, such that the number of events matches the recorded sample. The respective thresholds on $\mathcal{P}_{\text{Sel}}$ from 0.50 to 0.95 are shown in the subcaptions. No significant change in shape or different behavior for data and MC is visible. Continued on next page.



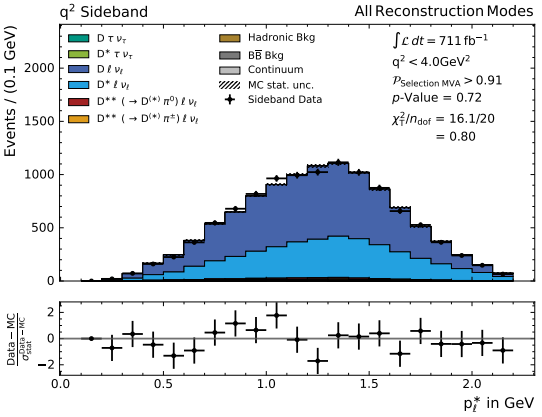
(g) $\mathcal{P}_{\text{Sel}} > 0.85$



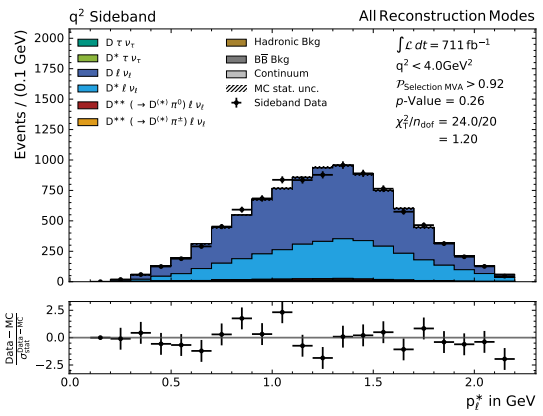
(h) $\mathcal{P}_{\text{Sel}} > 0.88$



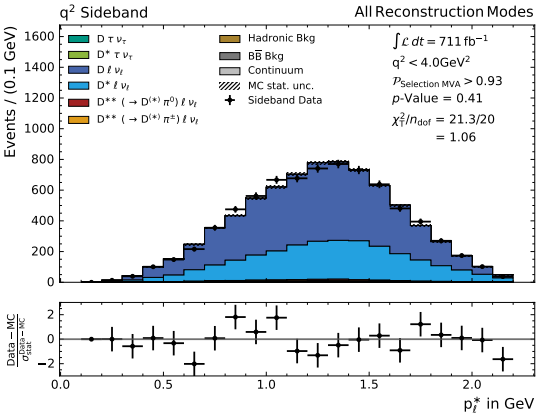
(i) $\mathcal{P}_{\text{Sel}} > 0.9$



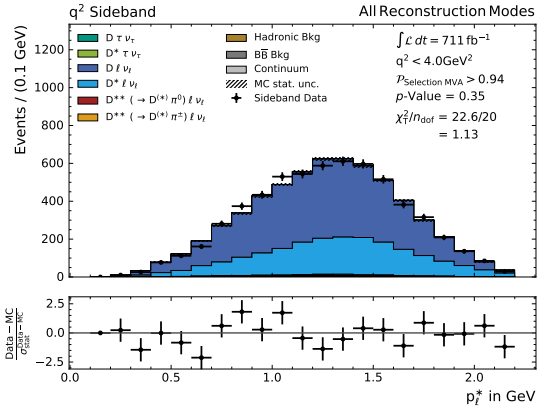
(j) $\mathcal{P}_{\text{Sel}} > 0.91$



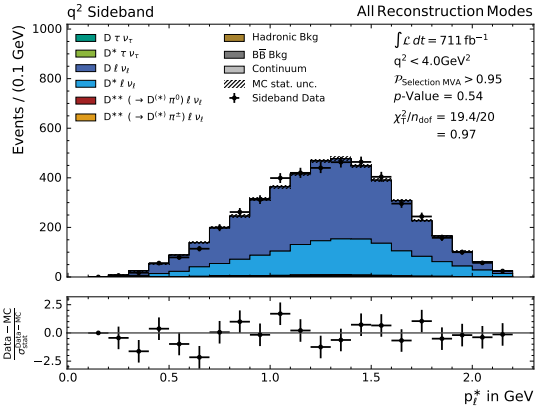
(k) $\mathcal{P}_{\text{Sel}} > 0.92$



(l) $\mathcal{P}_{\text{Sel}} > 0.93$

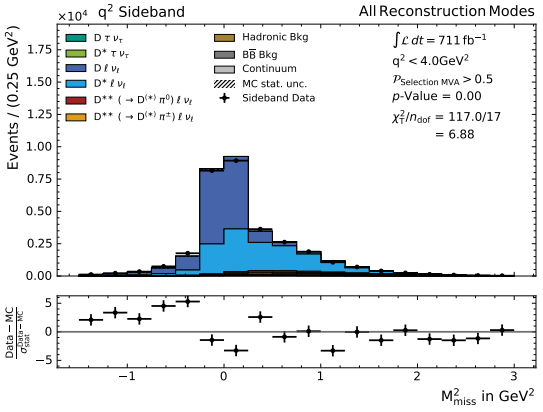


(m) $\mathcal{P}_{\text{Sel}} > 0.94$

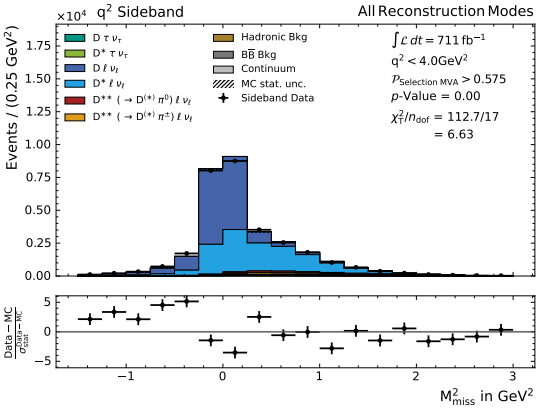


(n) $\mathcal{P}_{\text{Sel}} > 0.95$

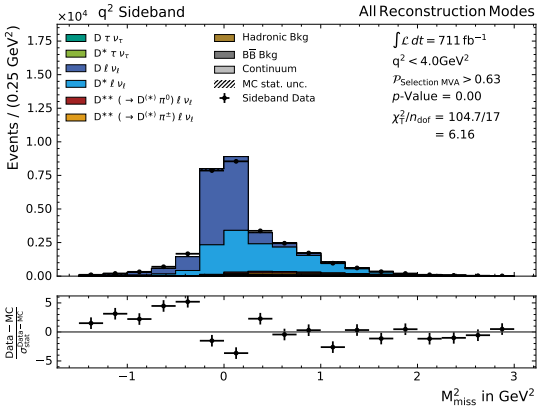
Figure D.12.: See previous page for caption.



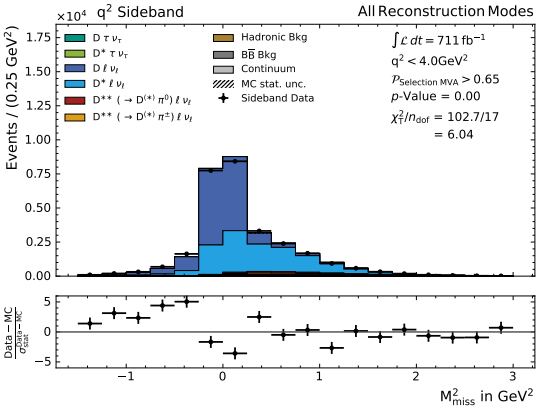
(a) $\mathcal{P}_{\text{Sel}} > 0.5$



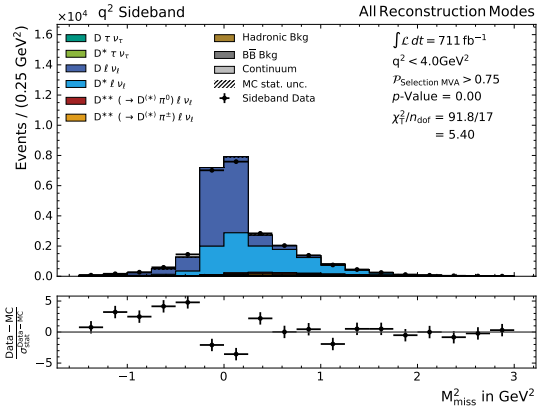
(b) $\mathcal{P}_{\text{Sel}} > 0.575$



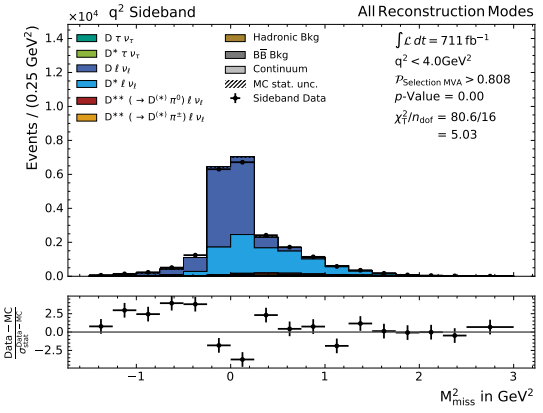
(c) $\mathcal{P}_{\text{Sel}} > 0.63$



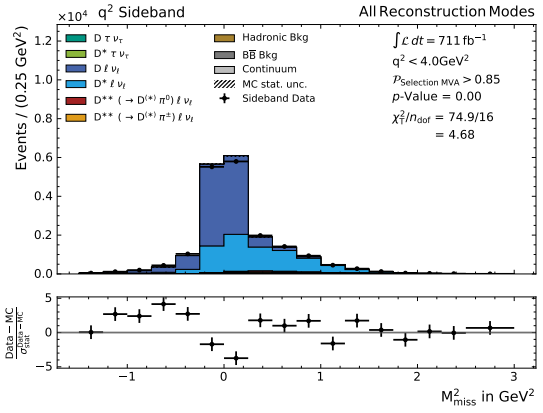
(d) $\mathcal{P}_{\text{Sel}} > 0.65$



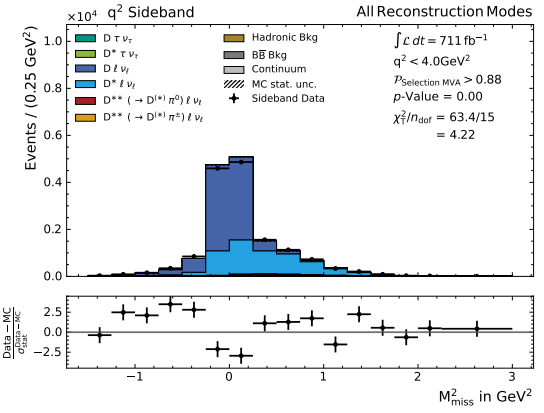
(e) $\mathcal{P}_{\text{Sel}} > 0.75$



(f) $\mathcal{P}_{\text{Sel}} > 0.808$

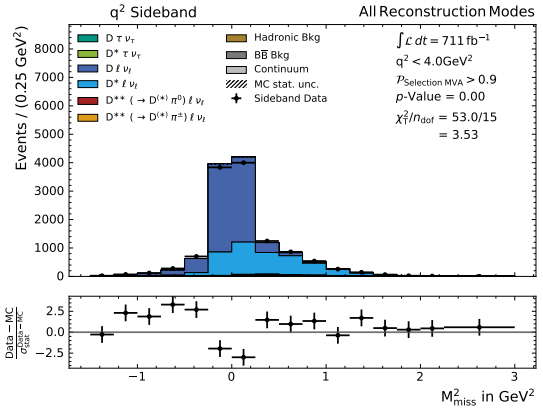


(g) $\mathcal{P}_{\text{Sel}} > 0.85$

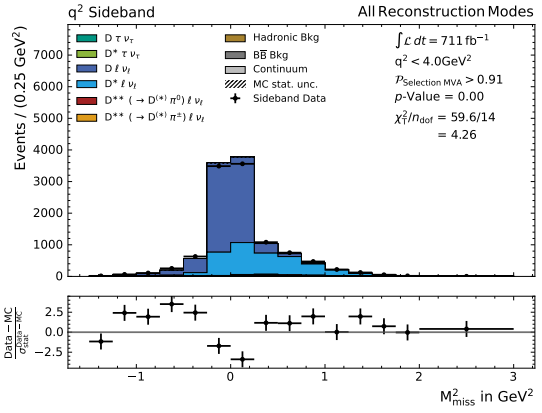


(h) $\mathcal{P}_{\text{Sel}} > 0.88$

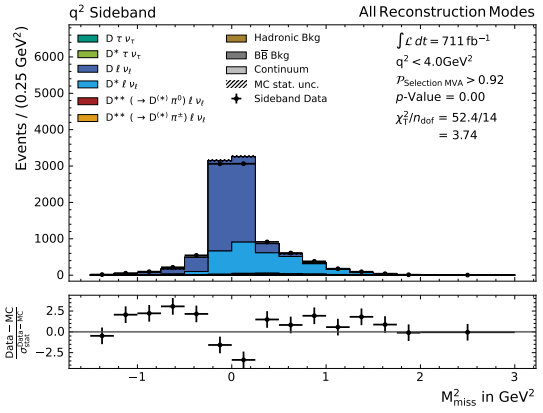
Figure D.13.: See next page for caption.



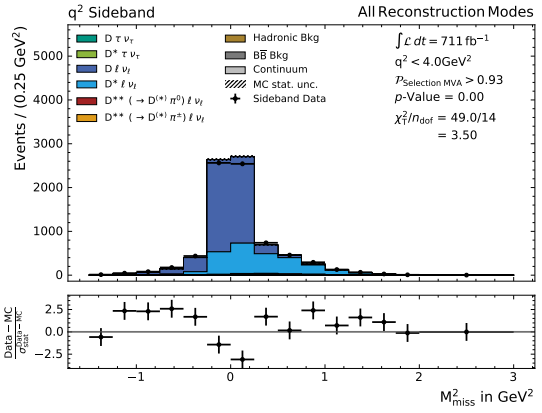
(i) $\mathcal{P}_{\text{Sel}} > 0.9$



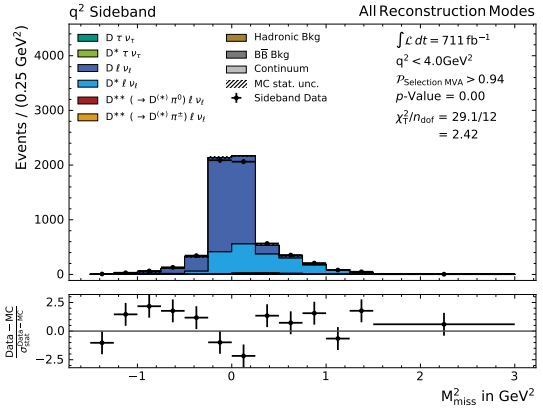
(j) $\mathcal{P}_{\text{Sel}} > 0.91$



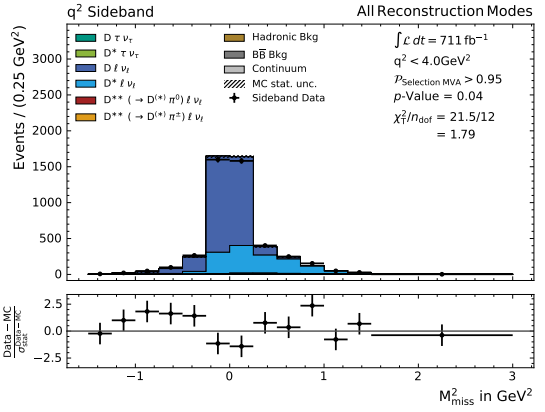
(k) $\mathcal{P}_{\text{Sel}} > 0.92$



(l) $\mathcal{P}_{\text{Sel}} > 0.93$



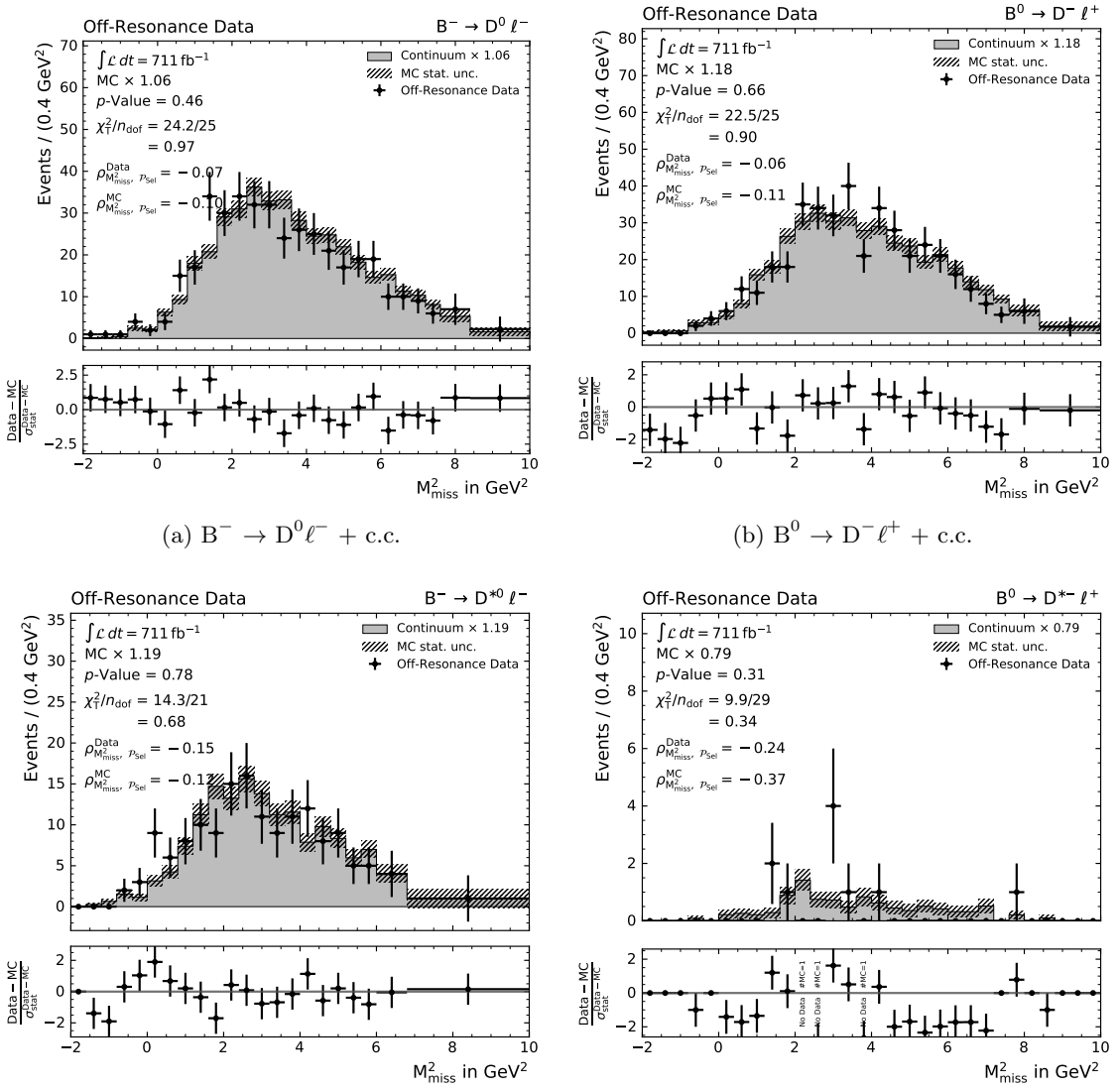
(m) $\mathcal{P}_{\text{Sel}} > 0.94$



(n) $\mathcal{P}_{\text{Sel}} > 0.95$

Figure D.13.: Effect of increasingly tight BDT output $\mathcal{P}_{\text{Sel}}$ selection criteria on the shape of the squared missing mass M^2_{miss} distribution in recorded and simulated data. The histogram of the simulated data is rescaled, such that the number of events matches the recorded sample. The respective thresholds on $\mathcal{P}_{\text{Sel}}$ from 0.50 to 0.95 are shown in the sub-captions. No significant change in shape or different behavior for data and MC is visible. The agreement between data and MC improves slightly for higher thresholds.

D.3. Additional Continuum Background Model Validation Plots



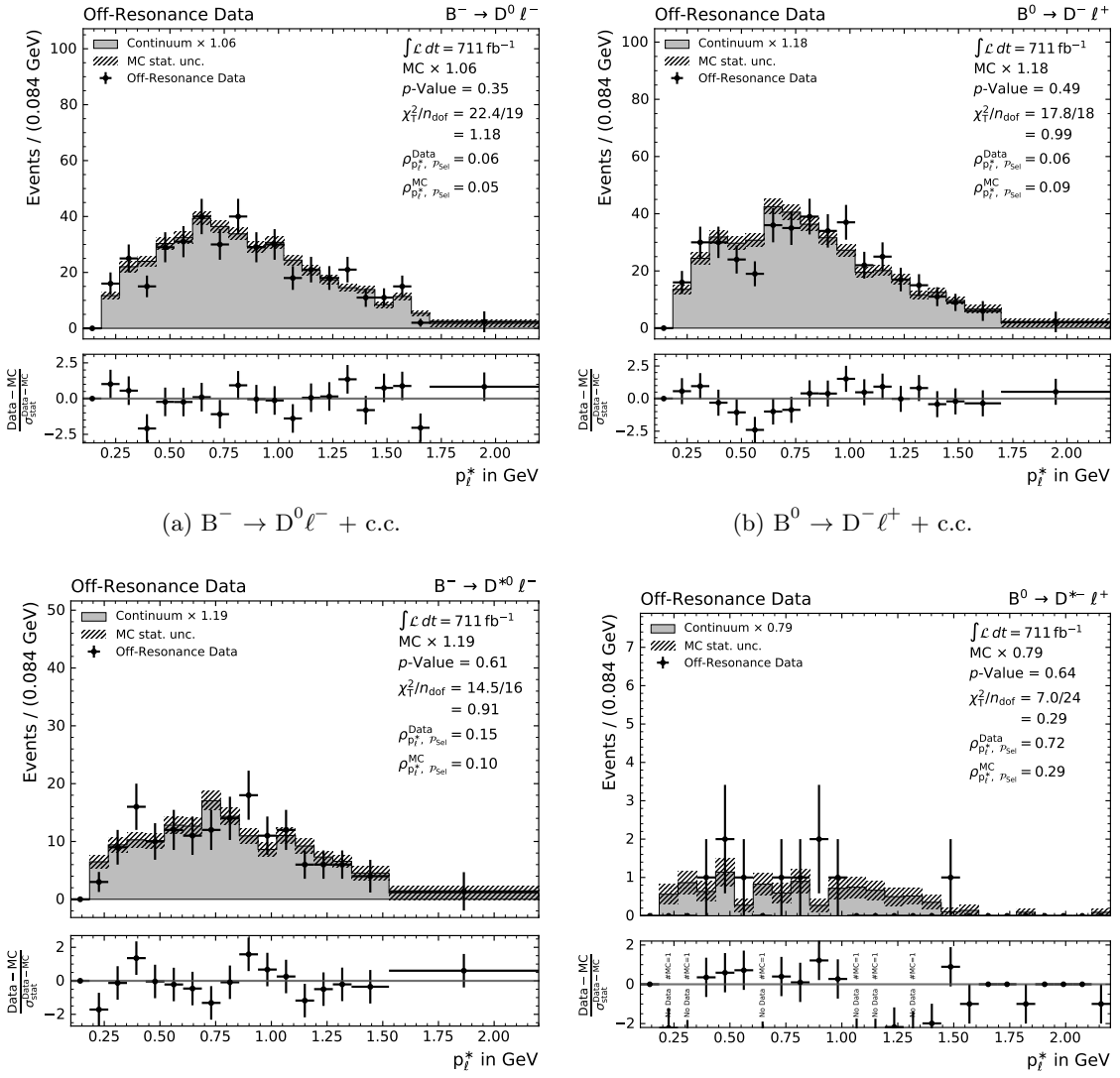
(a) $B^- \to D^0 \ell^-$ + c.c.

(b) $B^0 \to D^- \ell^+$ + c.c.

(c) $B^- \to D^{*0} \ell^-$ + c.c.

(d) $B^0 \to D^{*-} \ell^+$ + c.c.

Figure D.14.: Comparison of the fit observable M^2_{miss} for recorded data and MC off-resonance samples with a center-of-mass energy 60 MeV below the threshold for $B\overline{B}$ production. The plots show the comparison for each individual main reconstruction mode (see sub-captions). Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in Section 9.4.



(a) $B^- \to D^0 \ell^-$ + c.c.

(b) $B^0 \to D^- \ell^+$ + c.c.

(c) $B^- \to D^{*0} \ell^-$ + c.c.

(d) $B^0 \to D^{*-} \ell^+$ + c.c.

Figure D.15.: Comparison of the fit observable p^*_ℓ for recorded data and MC off-resonance samples with a center-of-mass energy 60 MeV below the threshold for $B\overline{B}$ production. The plots show the comparison for each individual main reconstruction mode (see sub-captions). Please refer to the introduction of Chapter 9 for a detailed description of the included information. The results are discussed in Section 9.4.

E. Supplementary Material on M^2_{miss} Resolution

This appendix contains additional material for the evaluation of the M^2_{miss} resolution correction and the fit on data in the q^2-sideband.

E.1. Additional Figures of M^2_{miss} Resolution Correction

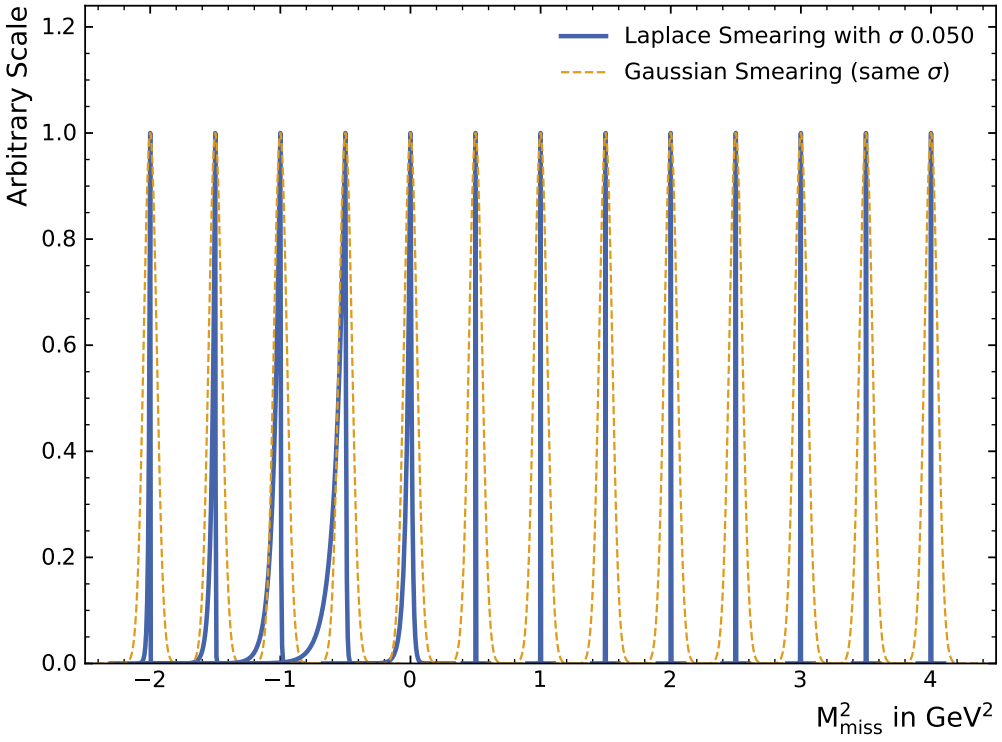
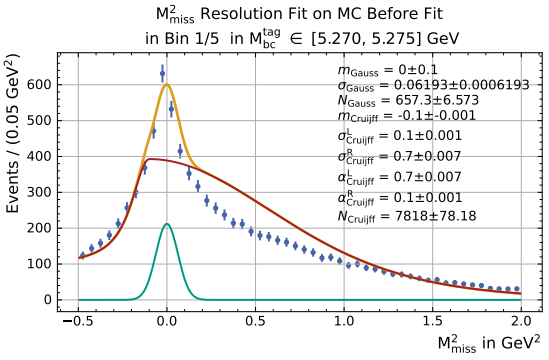
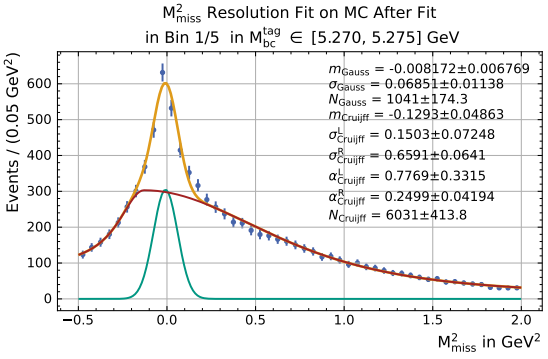


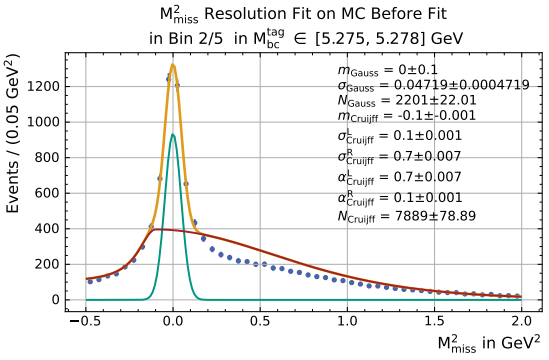
Figure E.16.: Illustration of the smearing method described in Section 10.1 for the case of $\sigma = 0.05$. The smear functions are indicated for every half integer value of the M^2_{miss} distribution between $-2.0\,\mathrm{GeV}^2$ and $4.0\,\mathrm{GeV}^2$. The y-axis shows arbitrary values and the smear PDF function are scaled, such that their widths and shapes can be compared easily. The orange, dashed lines show the smearing achieved by a Gaussian smear function. The blue, solid lines represent the smearing with an asymmetric Laplace distribution, modified as described in Section 10.1. Due to the modulation of the σ values with an additional Gaussian and a reverse logistic sigmoid function, the asymmetric Laplace only affects the M^2_{miss} in the region of $0.0\,\mathrm{GeV}^2$ and for negative values. This artificial smearing is chosen such that the result represents what is observed on recorded data.



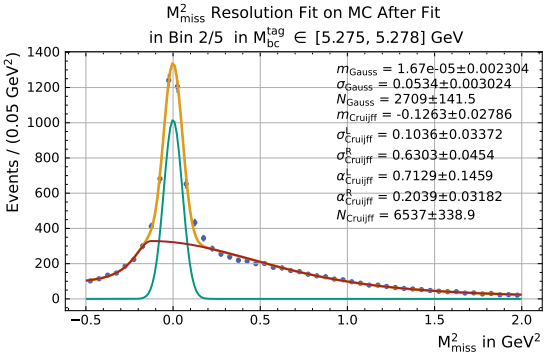
(a) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 1, **pre-fit**



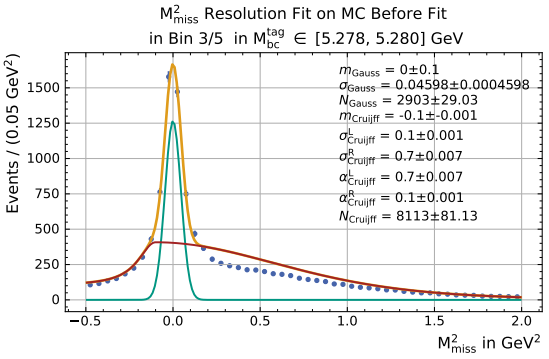
(b) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 1, **post-fit**



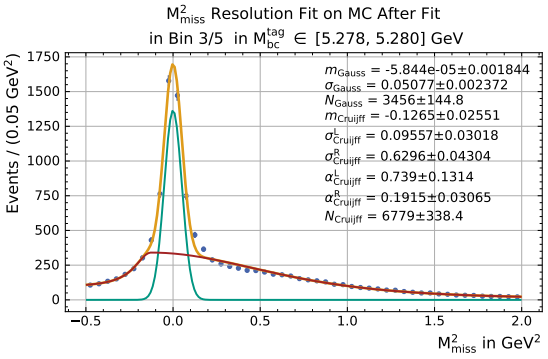
(c) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 2, **pre-fit**



(d) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 2, **post-fit**

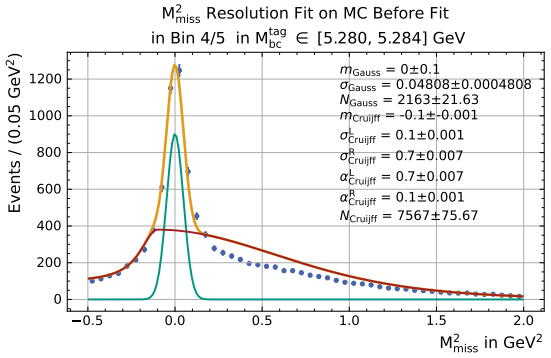


(e) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 3, **pre-fit**

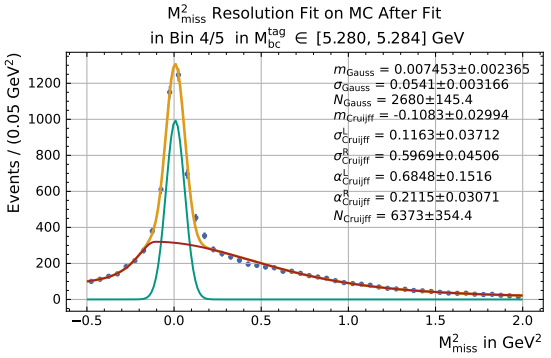


(f) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 3, **post-fit**

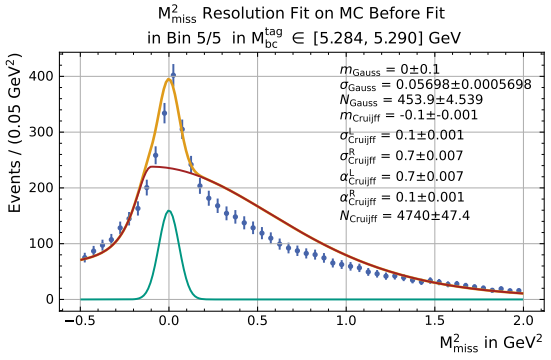
Figure E.17.: MC fit of M^2_{miss} resolution in bins of M^{tag}_{bc} and in the q^2-sideband. The pre-fit distributions are displayed on the left side. The result of the fits are shown in the post-fit distributions on the right side. The blue dots are the MC data points, while the solid lines represent the Gaussian signal shape (green), the Cruijff background shape (red) and the combination of the two (orange). Each row corresponds to the fit in a bin of the beam-constraint mass M^{tag}_{bc} of the tag side B meson. The plots for the remaining bins 4 and 5 are shown in the continuation of this figure on page 171.



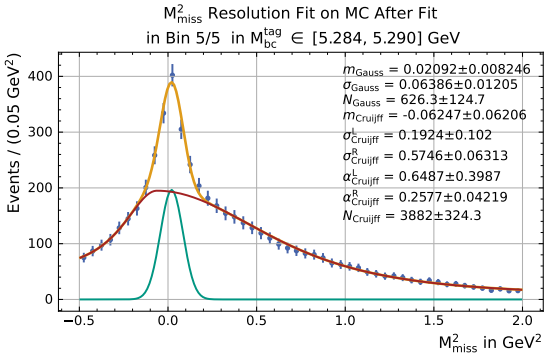
(g) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 4, **pre-fit**



(h) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 4, **post-fit**

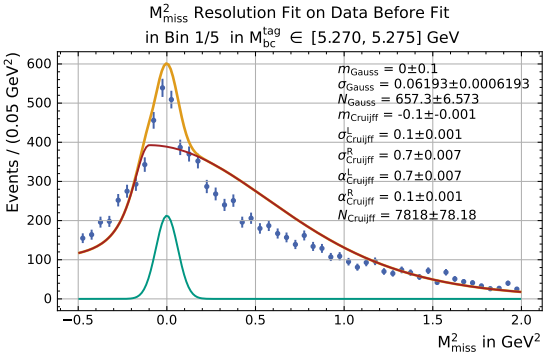


(i) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 5, **pre-fit**

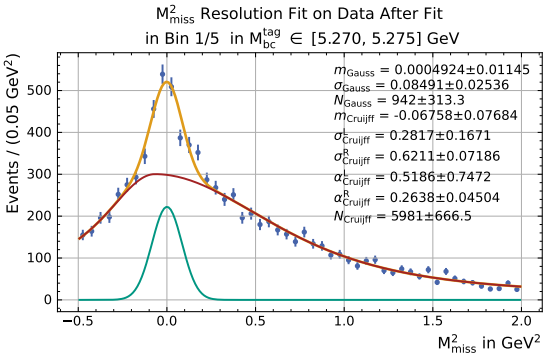


(j) M^2_{miss} res. fit on **MC** in M^{tag}_{bc} bin 5, **post-fit**

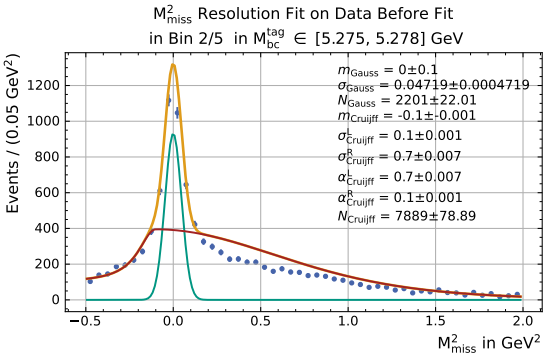
Figure E.17.: MC fit of M^2_{miss} resolution in bins of M^{tag}_{bc} and in the q^2-sideband. The pre-fit distributions are displayed on the left side. The result of the fits are shown in the post-fit distributions on the right side. The blue dots are the MC data points, while the solid lines represent the Gaussian signal shape (green), the Cruijff background shape (red) and the combination of the two (orange). Each row corresponds to the fit in a bin of the beam-constraint mass M^{tag}_{bc} of the tag side B meson. The first part of this figure is displayed on page 170.



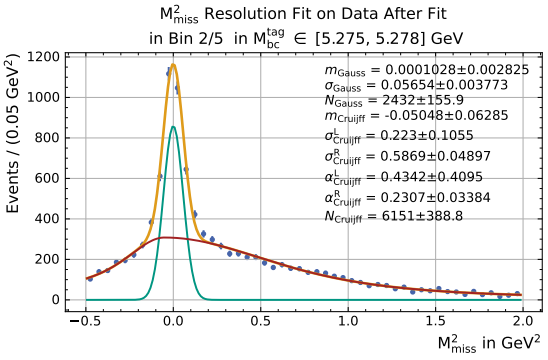
(a) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 1, **pre-fit**



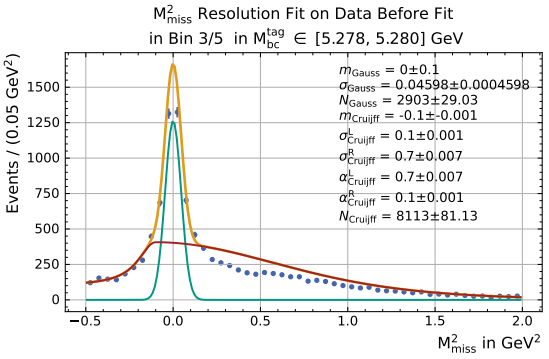
(b) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 1, **post-fit**



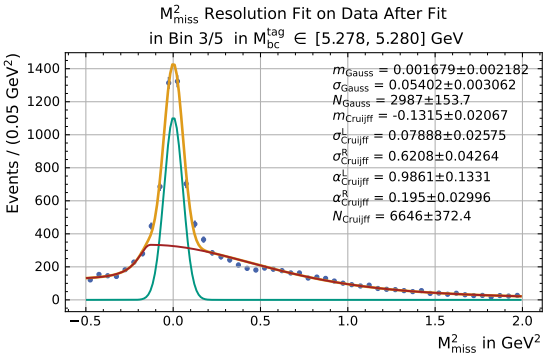
(c) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 2, **pre-fit**



(d) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 2, **post-fit**

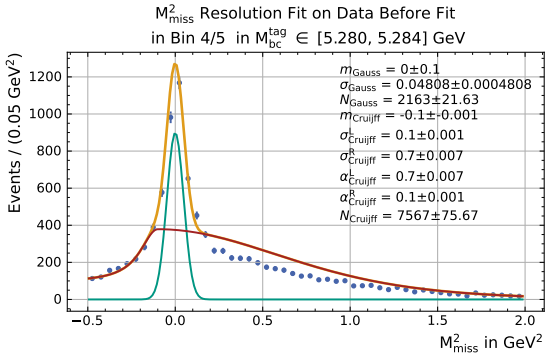


(e) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 3, **pre-fit**

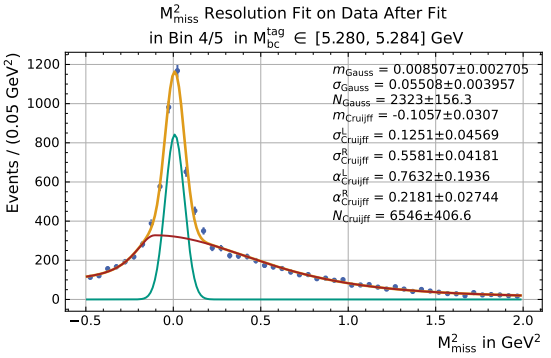


(f) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 3, **post-fit**

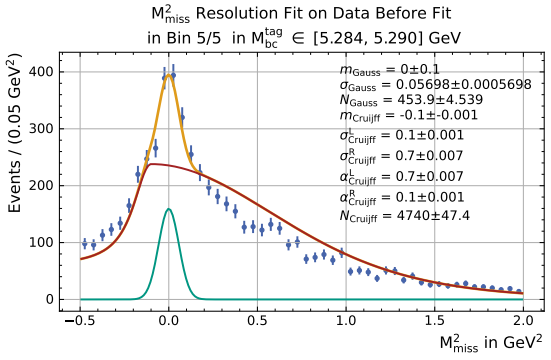
Figure E.18.: Data fit of M^2_{miss} resolution in bins of M^{tag}_{bc} and in the q^2-sideband. The pre-fit distributions are displayed on the left side. The result of the fits are shown in the post-fit distributions on the right side. The blue dots are the recorded data points, while the solid lines represent the Gaussian signal shape (green), the Cruijff background shape (red) and the combination of the two (orange). Each row corresponds to the fit in a bin of the beam-constraint mass M^{tag}_{bc} of the tag side B meson. The plots for the remaining bins 4 and 5 are shown in the continuation of this figure on page 173.



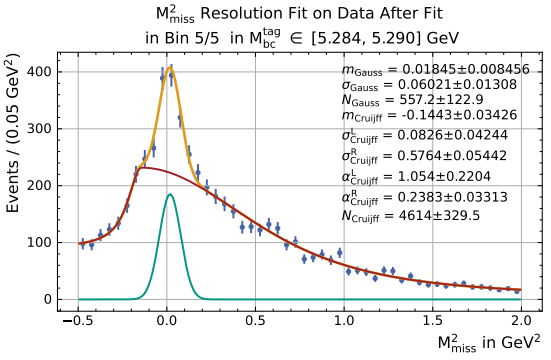
(g) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 4, **pre-fit**



(h) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 4, **post-fit**



(i) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 5, **pre-fit**

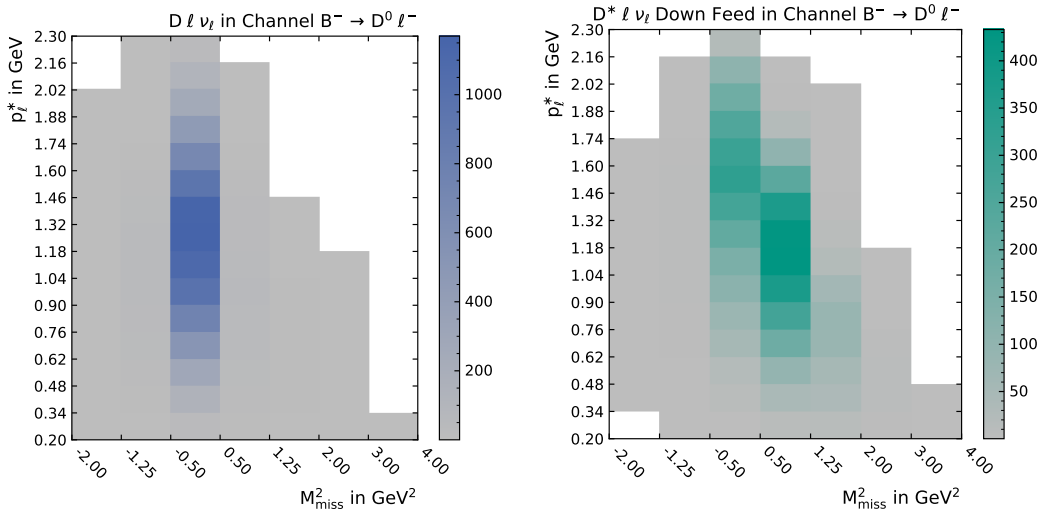


(j) M^2_{miss} res. fit on **data** in M^{tag}_{bc} bin 5, **post-fit**

Figure E.18.: Data fit of M^2_{miss} resolution in bins of M^{tag}_{bc} and in the q^2-sideband. The pre-fit distributions are displayed on the left side. The blue dots are the recorded data points, while the solid lines represent the Gaussian signal shape (green), the Cruijff background shape (red) and the combination of the two (orange). The result of the fits are shown in the post-fit distributions on the right side. Each row corresponds to the fit in a bin of the beam-constraint mass M^{tag}_{bc} of the tag side B meson. The first part of this figure is displayed on page 172.

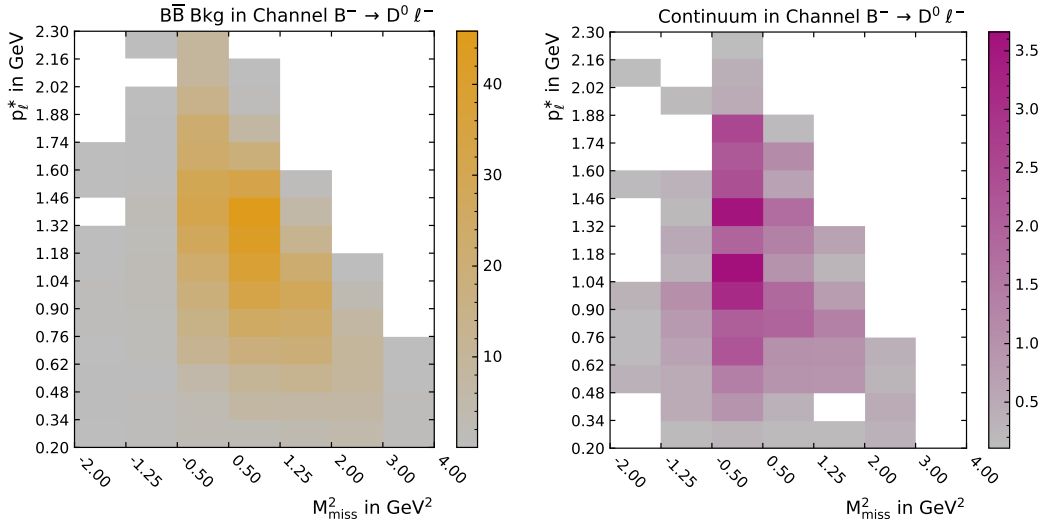
E.2. q^2-Sideband Fit Templates

In this appendix, the templates to the fit on q^2-sideband data for the extraction of the decay rate ratio of the processes $B \to D\ell\nu$ versus the processes $B \to D^*\ell\nu$ are provided. In the reconstruction modes with ground state D mesons the $B \to D^*\ell\nu$ **Feed-Down** templates are defined, whereas the regular $B \to D^*\ell\nu$ templates are defined in the reconstruction channels using the excited D^* mesons.



(a) $B \to D\ell\nu$ template

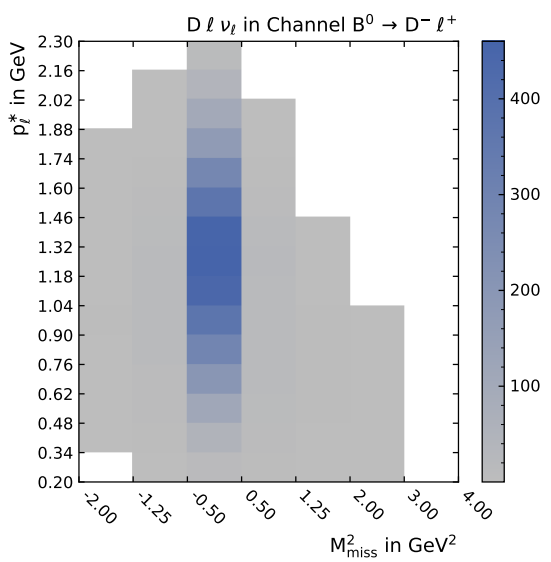
(b) $B \to D^*\ell\nu$ **Feed-Down** template



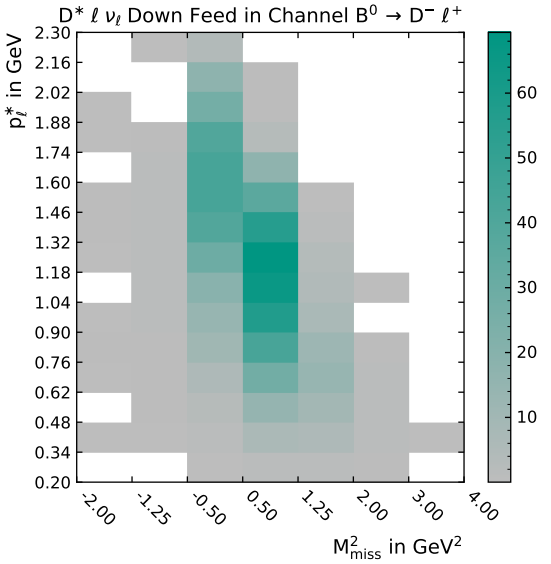
(c) $B\overline{B}$ background template

(d) Continuum background template

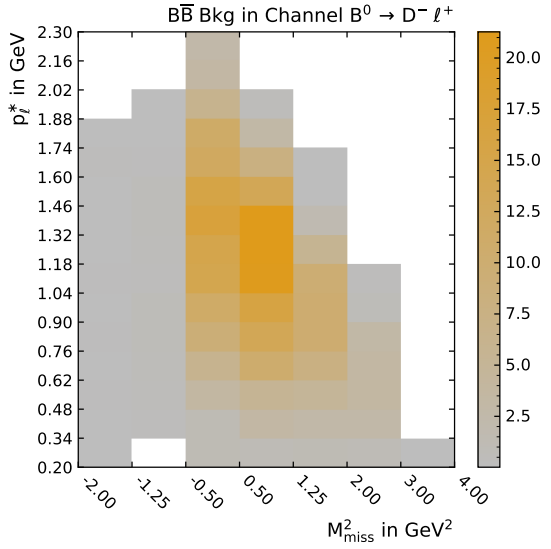
Figure E.19.: Histogrammed distributions used for the templates for the reconstruction channel $B^- \to D^0\ell^-$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts.



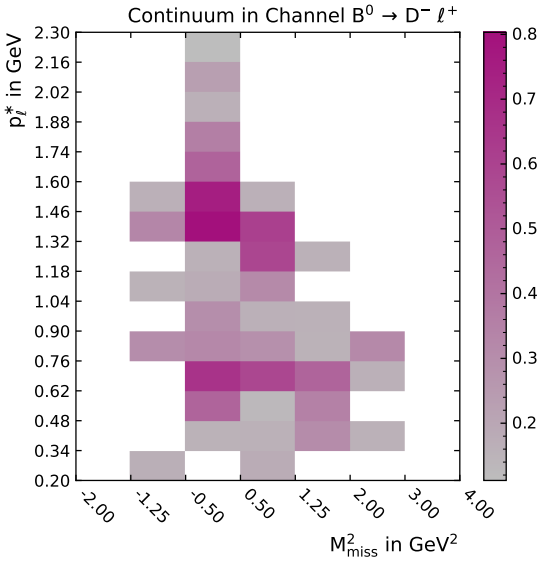
(a) $B \to D\ell\nu$ template



(b) $B \to D^*\ell\nu$ **Feed-Down** template

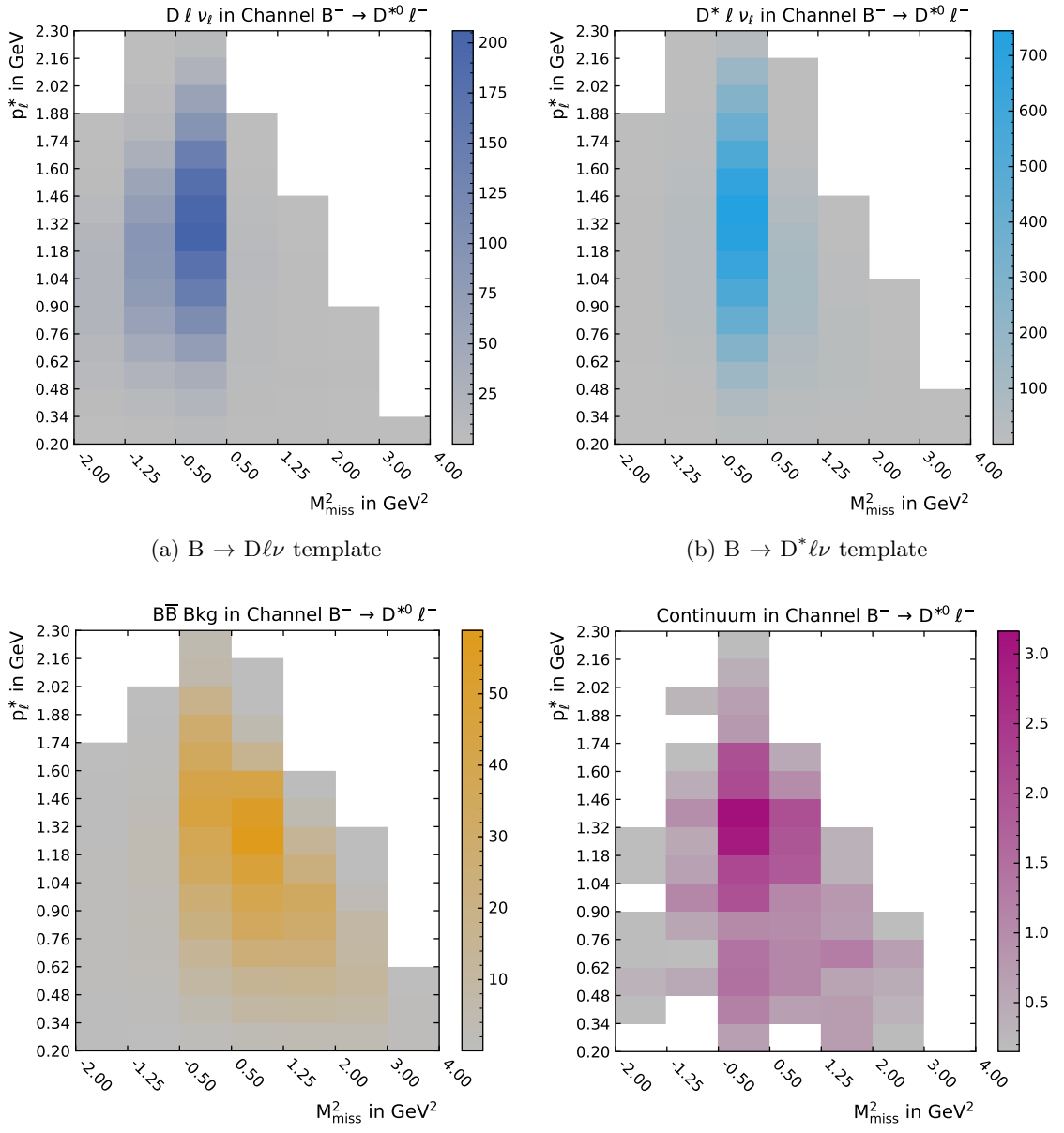


(c) $B\overline{B}$ background template



(d) Continuum background template

Figure E.20.: Histogrammed distributions used for the templates for the reconstruction channel $B^0 \to D^-\ell^+$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts.



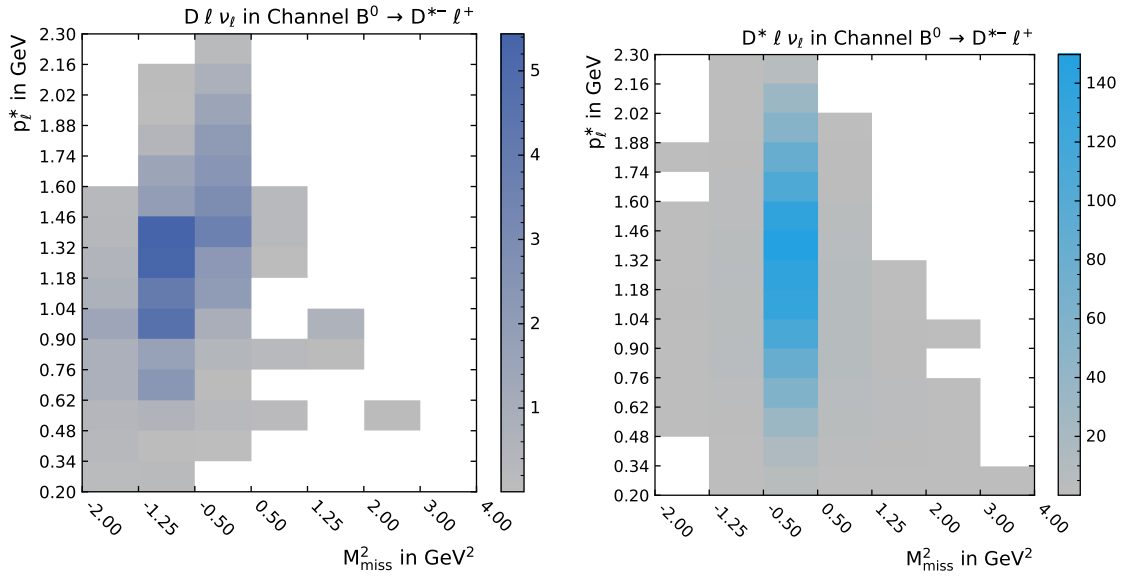
(a) $B \to D\ell\nu$ template

(b) $B \to D^*\ell\nu$ template

(c) $B\overline{B}$ background template

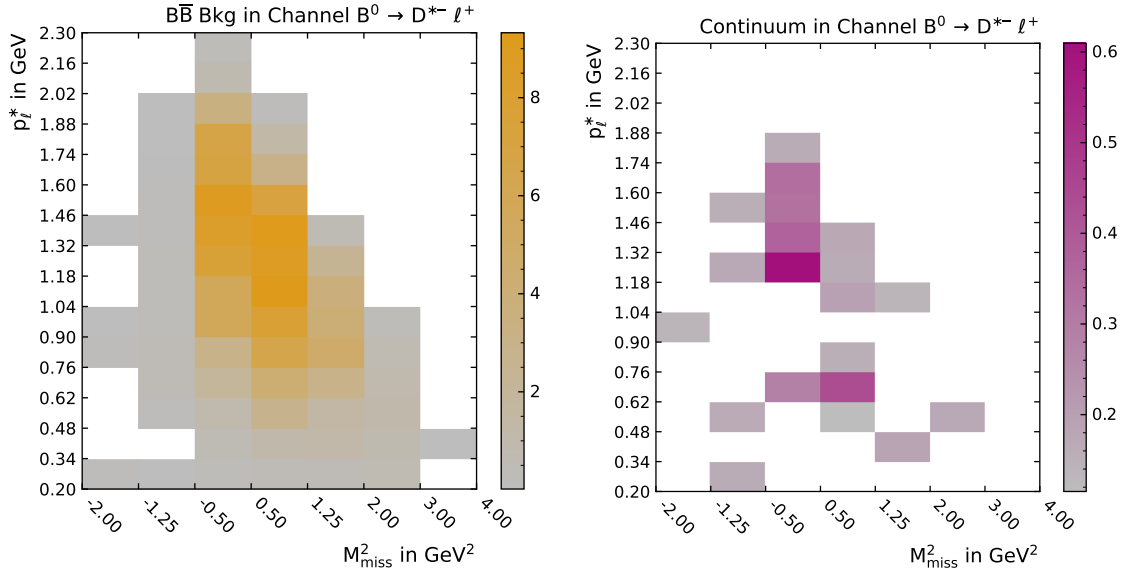
(d) Continuum background template

Figure E.21.: Histogrammed distributions used for the templates for the reconstruction channel $B^- \to D^{*0}\ell^-$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts.



(a) $B \to D\ell\nu$ template

(b) $B \to D^*\ell\nu$ template



(c) $B\overline{B}$ background template

(d) Continuum background template

Figure E.22.: Histogrammed distributions used for the templates for the reconstruction channel $B^0 \to D^{*-}\ell^+$. The unnormalized distributions are shown to allow for a comparison of the relative contributions of the different components. The individual components are labeled in the sub-captions. The x- and y-axis show the bin edges for M^2_{miss} and p^*_ℓ, respectively, where the different bin sizes for the former have to be noted. The gradient to the right of each component shows the color-coded bin counts.

F. New FEI Calibration Factors

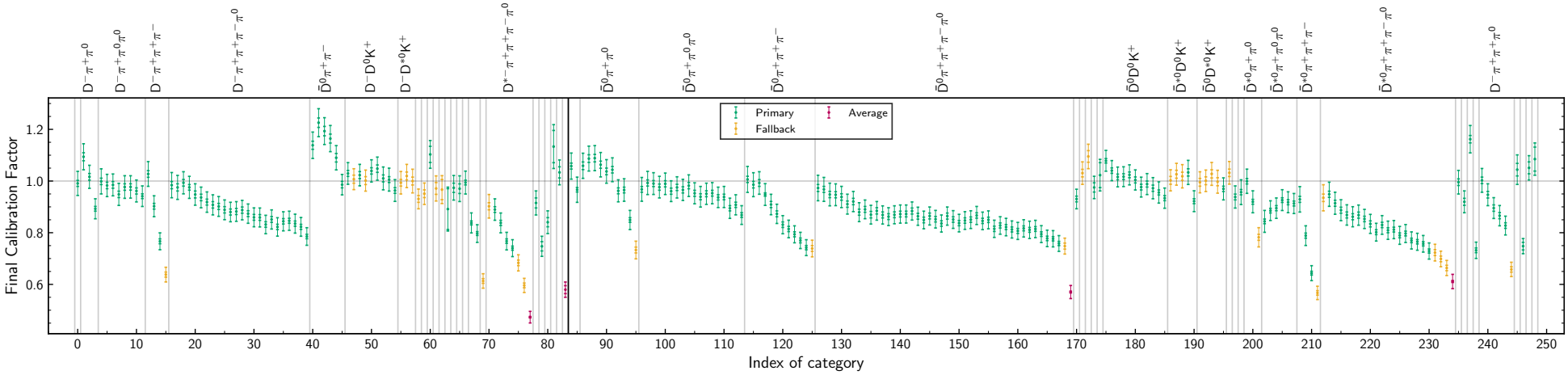


Figure F.23.: Visualization of the FEI B tagging calibration factors produced by a new FEI calibration, which is conducted in the scope of the related (see Chapter 5) $|V_{cb}|$ measurement at Belle [57]. The plot shows the calibration factors for different B_{tag} reconstruction channels, indicated by the vertical lines. The reconstruction channels labels are provided above the plot, if the space allows for it. For reconstruction channels with sufficient statistics, an additional dependency of the calibration factors on the FEI probability $\mathcal{P}_{FEI}$ is introduced. This FEI probability $\mathcal{P}_{FEI}$ quantifies the quality of the B_{tag} candidate produced by the FEI B tagging algorithm. For each suitable FEI reconstruction channel, this additional dependency is included in the form of an additional binning in $\mathcal{P}_{FEI}$, where a higher value of $\mathcal{P}_{FEI}$ results in a higher bin index (The x axis of this plot, labeled as *Index of category*, encodes both the channel and the bin in $\mathcal{P}_{FEI}$). The errors of the data points include the statistical and systematic uncertainties on the calibration factors. The different colors denote different procedures used to obtain the respective calibration factor. For further information, please refer to [57] which also provides the displayed figure.
Of importance to this analysis is the fact that the obtained calibration factors show a dependence on $\mathcal{P}_{FEI}$ and thus the quality of the B_{tag} candidate.

Danksagung

Zu allernächst möchte ich mich bei Prof. Dr. Michael Feindt für die Übernahme des Hauptreferats und Prof. Dr. Florian Bernlochner für die Übernahme des Korreferats bedanken. Michael hat mich durch seine Vorlesungen und interessanten Themen in die Welt der B-Physik geführt und mir einen Platz als Masterand in der Belle-Arbeitsgruppe angeboten. Dadurch führte mein Weg in eine Arbeitsgruppe und eine Welt, die ich nicht aus meiner Vergangenheit wegzudenken möchte. Er erlaubte mir viel Freiheit in der Wahl meiner Themen, wodurch es mir möglich war tief in viele Aspekte von Belle und Belle II einzutauchen. Florian ermöglichte mir meine Erfahrungen die ich auf diese Art und Weise gewonnen habe auf die Analyse der Zerfallsratenverhältnisse $\mathcal{R}(\mathrm{D})$ und $\mathcal{R}(\mathrm{D}^*)$ anzuwenden. Ich möchte mich herzlich für die Betreuung dieses Hauptteils meiner Arbeit bei ihm bedanken. Sein umfangreicher Wissenschatz und seine große Erfahrung in diesem Bereich waren eine Bereicherung für mich und die Qualität dieser Arbeit.

Mein besonderer Dank gilt Dr. Martin Heck, der mir zu Beginn meiner Zeit als Doktorand in der B-Arbeitsgruppe die Entwicklung des Spurfindungsalgorithmus für die Siliziumdetektoren des Belle II–Experiments anvertraut hat. Als Teil der Tracking-Gruppe gewann ich nicht nur ein besseres Verständnis für die Details der Ereignisrekonstruktion mit den Detektoren Belle und Belle II, sondern auch Erfahrungen in der kollaborativen Softwareentwicklung. In dieser Zeit profitierte ich sehr von Martin's Ratschlägen und seinem umfangreichen Wissen, aber auch von den zahlreichen Mitgliedern der Tracking-Gruppe. Insbesondere möchte ich hierbei Dr. Nils Braun, Dr. Thomas Hauth und Dr. Thomas Lück hervorheben, die mir stehts mir Rat und Tat zur Seite standen und mit denen es eine Freude war zusammen zu arbeiten.

Auch im Rahmen der Messung von $\mathcal{R}(\mathrm{D})$ und $\mathcal{R}(\mathrm{D}^*)$ durfte ich mit wunderbaren und hilfsbereiten Kollegen kollaborieren. Meine Masteranden Maximilian Welsch und Patrick Ecker trugen einen wesentlichen Beitrag zu dem Verständnis der Beschreibung des Hauptuntergrunds im Monte Carlo für die Analyse bei. Weiterhin gaben sie, sowie Kilian Lieret und Dr. Markus Prim wichtiges Feedback zu den Datensätzen und Rekonstruktionsmethoden der Analyse.

Bei Dr. Thomas Keck und Dr. Pablo Goldenzweig möchte ich mich für die gute Betreuung vorallem zu Beginn meiner Zeit in der B-Arbeitsgruppe bedanken. Natürlich gilt auch jedem Mitglied dieser Arbeitsgruppe, das ich in meiner Zeit am ETP kennen lernen durfte, mein herzlicher Dank für die wunderbare Zusammenarbeit und das schöne Miteinander. Dies gilt ebenso für die Mitglieder der weiteren Arbeitsgruppen des ETP, welche diese Atmosphäre zusätzlich bereichert haben. Mein bester Dank gebührt auch allen die meine Arbeit Korrektur gelesen haben. Bei dem Admin-Team des ETP möchte ich mich explizit für ihre tadellose Arbeit bedanken, die ein flüssiges Arbeiten am ETP ermöglicht hat. Das selbige kann über die Verwaltung des ETP und damit insbesondere Frau Bräunling und Frau Chen gesagt werden. Bei ihnen bedanke ich mich herzlich dafür, dass sie uns einen Großteil der Bürokratie abnehmen und uns beim Rest mit Rat und Tat zur Seite stehen,

wodurch man sich besser auf die eigene Arbeit konzentrieren kann. Bei Prof. Dr. Torben Ferber möchte ich mich für die Unterstützung auf den letzten Metern zur meiner Promotion bedanken.

Schlussendlich möchte ich auch herzlich meinen Eltern, meinen Brüdern, meinen Freunden und meiner Freundin Naiá danken. Sie alle haben mich durch mein gesamtes Studium und meine Doktorarbeit begleitet und unterstützt. Ich denke oft darüber nach welche Interaktionen und Gespräche mit ihnen mich dahin geführt haben, wo ich nun stehe. Ich möchte keine einzige davon missen und bedanke mich für jede einzelne davon.

Bibliography

[1] A. J. Bevan, B. Golob, T. Mannel, S. Prell, *et al.*, "The Physics of the B Factories", *The European Physical Journal C* **74** no. 11, (Nov, 2014) 3026, Nov, 2014.

[2] **National Laboratory for High Energy Physics**, KEK, "KEKB B-Factory Design Report", 6, 1995.

[3] **PEP-II**, "PEP-II: An Asymmetric B Factory. Conceptual Design Report", 6, 1993.

[4] A. Abashian, K. Gotow, N. Morgan, L. Piilonen, *et al.*, "The belle detector", *Nucl. Instrum. Methods Phys. Res. A* **479** no. 1, (2002) 117–232, 2002.

[5] **BaBar**, D. Boutigny *et al.*, "BaBar Technical Design Report", 3, 1995.

[6] **Belle Collaboration**, K. Abe *et al.*, "Observation of Large CP Violation in the Neutral B Meson System", *Phys. Rev. Lett.* **87** (2001) 091802, 2001.

[7] **The Belle Collaboration**, K. Abe, K. Abe, R. Abe, I. Adachi, *et al.*, "Observation of the Decay B $\to$ Kl^+l^-", *Phys. Rev. Lett.* **88** (Dec, 2001) 021801, Dec, 2001.

[8] **Belle Collaboration**, A. Ishikawa, K. Abe, K. Abe, T. Abe, *et al.*, "Observation of B $\to$ K$^*\ell^+\ell^-$", *Phys. Rev. Lett.* **91** (Dec, 2003) 261601, Dec, 2003.

[9] J. P. Lees *et al.*, "Measurement of an excess of $B^- \to D^{(*)}\tau^-\nu_\tau$ decays and implications for charged Higgs bosons", *Physical Review D* **88** no. 7, (Oct, 2013) 072012, Oct, 2013.

[10] **HFLAV**, Y. S. Amhis *et al.*, "Average of $\mathcal{R}$(D) and $\mathcal{R}$(D*) for 2021". Last accessed Januar 13th 2022.

[11] **HFLAV**, Y. S. Amhis *et al.*, "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018", *Eur. Phys. J. C* **81** no. 3, (2021) 226, 2021.

[12] **Belle Collaboration**, M. Huschle, T. Kuhr, M. Heck, P. Goldenzweig, *et al.*, "Measurement of the branching ratio of $\overline{\mathrm{B}} \to \mathrm{D}^{(*)}\tau^-\overline{\nu}_\tau$ relative to $\overline{\mathrm{B}} \to \mathrm{D}^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle", *Phys. Rev. D* **92** (Oct, 2015) 072014, Oct, 2015.

[13] G. Caria, P. Urquijo, *et al.*, "Measurement of R(D) and R(D*) with a Semileptonic Tagging Method", *Physical Review Letters* **124** (04, 2020) 1, 04, 2020.

[14] **The Belle Collaboration**, S. Hirose *et al.*, "Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\overline{B} \to D^*\tau^-\overline{\nu}_\tau$ with one-prong hadronic τ decays at Belle", *Phys. Rev. D* **97** (Jan, 2018) 012004, Jan, 2018.

[15] **LHCb Collaboration**, R. Aaij *et al.*, "Measurement of the Ratio of Branching Fractions $\mathcal{B}(\overline{\mathrm{B}}^0 \to \mathrm{D}^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{\mathrm{B}}^0 \to \mathrm{D}^{*+}\mu^-\overline{\nu}_\mu)$", *Phys. Rev. Lett.* **115** (Sep, 2015) 111803, Sep, 2015.

[16] **LHCb Collaboration**, R. Aaij *et al.*, "Test of Lepton Flavor Universality by the Measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ Branching Fraction using Three-Prong τ Decays", *Phys. Rev. D* **97** (Apr, 2018) 072013, Apr, 2018.

[17] T. Keck *et al.*, "The Full Event Interpretation", *Computing and Software for Big Science* **3** no. 1, (Feb, 2019) 1 – 10, Feb, 2019.

[18] M. Gelb, T. Keck, *et al.*, "B2BII: Data Conversion from Belle to Belle II", *Comput. Softw. Big Sci.* **2** no. 1, (2018) 9, 2018.

[19] F. Metzner, "Analysis of $B^+ \to \ell^+\nu_\ell\gamma$ Decays with the Belle II Analysis Software Framework", Master's Thesis, Karlsruhe Institute of Technology (KIT), 2016.

[20] **The Belle Collaboration**, M. Gelb, F. U. Bernlochner, P. Goldenzweig, F. Metzner, *et al.*, "Search for the rare decay of $B^+ \to \ell^+\nu_\ell\gamma$ with improved hadronic tagging", *Phys. Rev. D* **98** (Dec, 2018) 112016, Dec, 2018.

[21] M. Huschle, "Measurement of the branching ratio of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ relative to $\overline{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle". PhD Thesis, Karlsruhe Institute of Technology (KIT), 2015.

[22] A. V. Manohar and M. B. Wise, "Heavy quark physics", vol. 10 of *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology*. Cambridge University Press, 2000.

[23] F. U. Bernlochner, M. F. Sevilla, D. J. Robinson, and G. Wormser, "Semitauonic b-hadron decays: A lepton flavor universality laboratory", Jan, 2021.

[24] M. Franco Sevilla, "Evidence for an Excess of $B \to D\tau\nu$ Decays". PhD Thesis, Stanford University, 2012.

[25] C. Glenn Boyd, B. Grinstein, and R. F. Lebed, "Model-Independent Determinations of $B \to D\ell\nu$, $B \to D^*\ell\nu$ Form Factors", *Nuclear Physics B* **461** no. 3, (1996) 493–511, 1996.

[26] C. G. Boyd, B. Grinstein, and R. F. Lebed, "Precision Corrections to Dispersive Bounds on Form Factors", *Phys. Rev. D* **56** (Dec, 1997) 6895–6911, Dec, 1997.

[27] M. Neubert, "Heavy-quark symmetry", *Physics Reports* **245** no. 5, (1994) 259–395, 1994.

[28] **Particle Data Group**, P. A. Zyla *et al.*, "Review of Particle Physics", *Progress of Theoretical and Experimental Physics* **2020** no. 8, (08, 2020) 2092, 08, 2020. 083C01.

[29] K. Hagiwara, A. D. Martin, and M. F. Wade, "Exclusive Semileptonic B Meson Decays", *Nucl. Phys. B* **327** (1989) 569–594, 1989.

[30] J. G. Körner and G. A. Schuler, "Exclusive Semileptonic Heavy Meson Decays Including Lepton Mass Effects", *Z. Phys. C* **46** (1990) 93, 1990.

[31] A. F. Falk and M. Neubert, "Second-order power corrections in the heavy-quark effective theory. I. Formalism and meson form factors", *Phys. Rev. D* **47** (Apr, 1993) 2965–2981, Apr, 1993.

[32] A. Sirlin, "Large m_W, m_Z Behaviour of the $\mathcal{O}(\alpha)$ Corrections to Semileptonic Processes mediated by W", *Nuclear Physics B* **196** no. 1, (1982) 83–92, 1982.

[33] M. Tanaka and R. Watanabe, "Tau longitudinal Polarization in $\overline{\mathrm{B}} \to \mathrm{D}\tau\overline{\nu}_\tau$ and its Role in the Search for the charged Higgs Boson", *Phys. Rev. D* **82** (Aug, 2010) 034027, Aug, 2010.

[34] D. Ferlewicz, P. Urquijo, and E. Waheed, "Revisiting fits to $\mathrm{B}^0 \to \mathrm{D}^{*-}\ell^+\nu_\ell$ to measure $|V_{\mathrm{cb}}|$ with novel methods and preliminary LQCD data at nonzero recoil", *Phys. Rev. D* **103** (Apr, 2021) 073005, Apr, 2021.

[35] I. Caprini, L. Lellouch, and M. Neubert, "Dispersive Bounds on the Shape of $\mathrm{B} \to \mathrm{D}^{(*)}\ell\nu$ Form Factors", *Nuclear Physics B* **530** no. 1, (1998) 153–181, 1998.

[36] D. Scora and N. Isgur, "Semileptonic Meson Decays in the Quark Model: An Update", *Phys. Rev. D* **52** (Sep, 1995) 2783–2812, Sep, 1995.

[37] A. K. Leibovich, Z. Ligeti, I. W. Stewart, and M. B. Wise, "Semileptonic B decays to excited charmed mesons", *Phys. Rev. D* **57** (Jan, 1998) 308–330, Jan, 1998.

[38] F. U. Bernlochner and Z. Ligeti, "Semileptonic $\mathrm{B}_{(s)}$ decays to excited charmed mesons with e, μ, τ and searching for new physics with $R(\mathrm{D}^{**})$", *Phys. Rev. D* **95** (Jan, 2017) 014022, Jan, 2017.

[39] K. Nishimura, "Number of B events in Hadron B(J)". Belle Internal, Last accessed Januar 13th 2022.

[40] D. J. Lange, "The EvtGen particle decay simulation package", *Nucl. Instrum. Meth. A* **462** (2001) 152–155, 2001.

[41] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, *et al.*, "High-energy physics event generation with PYTHIA 6.1", *Comput. Phys. Commun.* **135** (2001) 238–259, 2001.

[42] E. Barberio and Z. Wąs, "PHOTOS – a universal Monte Carlo for QED radiative corrections: version 2.0", *Comput. Phys. Commun.* **79** no. 2, (1994) 291–308, 1994.

[43] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, "GEANT 3: User's Guide; rev. version". CERN, Geneva, 1987.

[44] **Belle Collaboration**, L. Cao, W. Sutcliffe, R. Van Tonder, F. U. Bernlochner, *et al.*, "Measurements of Partial Branching Fractions of Inclusive $\mathrm{B} \to X_{\mathrm{u}}\ell^+\nu_\ell$ Decays with Hadronic Tagging", *Phys. Rev. D* **104** (Jul, 2021) 012008, Jul, 2021.

[45] **Belle II Working Group 1**, S. Duell and F. Bernlochner, "Correction of $\mathcal{B}(\mathrm{B} \to X_{\mathrm{c}}\ell\nu_\ell)$ in Generic MC". `https://confluence.desy.de/pages/viewpage.action?pageId=202394372`. Belle II Internal, Last accessed Januar 24th 2022.

[46] R. A. Briere and E. R. Oxford, "Update of Charm Decay Tables", Mar, 2020. BELLE2-NOTE-PH-2020-008, Internal Document.

[47] M. Prim, "b2-hive/eFFORT v0.1.0", Jul, 2020. `https://doi.org/10.5281/zenodo.3965699`.

[48] M. Welsch, "Measurement of the Branching Fraction of $\mathrm{B} \to \mathrm{D}^{**}(\to \mathrm{D}^{(*)}\pi^0)\,\ell\nu$ in the Context of a $\mathcal{R}(\mathrm{D}^{(*)})$ Analysis at the Belle Experiment", Master's Thesis, Karlsruhe Institute of Technology (KIT), 2019.

[49] B. Bhuyan, "High P_{T} Tracking Efficiency Using Partially Reconstructed D^* Decays", 2010. Belle Note 1165, Internal Document.

[50] W. Dungel, "Systematic Investigation of the Reconstruction Efficiency of Low Momentum $\pi^{\pm}$ and π^0", 2011. Belle Note 1176, Internal Document.

[51] F. Bernlochner, L. Cao, W. Sutcliffe, R. van Tonder, and A. Ermakov, "Determination of $|V_{\mathrm{ub}}|$ and the kinematic distributions of $\mathrm{B} \to X_{\mathrm{u}} \ell \nu_\ell$ decays", 2021. Belle Note 1529, Internal Document.

[52] E. White, "Determination of $\mathrm{K}^0_{\mathrm{S}}$ Efficiency and Systematic Uncertainty", 2011. Belle Note 1207, Internal Document.

[53] L. Hinz *et al.*, "Lepton ID Efficiency Correction and Systematic Error", 2006. Belle Note 954, Internal Document.

[54] P. Urquijo, "Semi-Inclusive Semileptonic B Decays", 2012. Belle Note 1251, Internal Document.

[55] S. Nishida *et al.*, "Study of Kaon and Pion Identification Using Inclusive D^* Sample", 2005. Belle Note 779, Internal Document.

[56] J. Schwab, "Calibration of the Full Event Interpretation for the Belle and the Belle II Experiment", Master's Thesis, Karlsruhe Institute of Technology (KIT), 2017.

[57] M. Prim, F. U. Bernlochner, K. Lieret, T. Kuhr, and F. Metzner, "$|V_{\mathrm{cb}}|$ with hadronically tagged $\mathrm{B}^0 \to \mathrm{D}^* \ell \nu_\ell$ Decays", 2021. Belle Note 1615, Internal Document.

[58] P. Ecker, "Study of $\mathrm{B} \to \mathrm{D}^{**} \ell \nu$ Decays for a $\mathcal{R}(\mathrm{D}^{(*)})$ Measurement at Belle", Master's Thesis, Karlsruhe Institute of Technology (KIT), 2021.

[59] V. Trusov, "Development of Pattern Recognition Algorithms for the Central Drift Chamber of the Belle II Detector". PhD Thesis, Karlsruhe Institute of Technology (KIT), 2016.

[60] B. Kronenbitter, "Measurement of the Branching Fraction of $\mathrm{B} \to \tau \nu$ Decays at the Belle Experiment". PhD Thesis, Karlsruhe Institute of Technology (KIT), 2014. Karlsruhe, KIT, Diss., 2014.

[61] M. Starič, "Track Smearing in MC". Belle Charm Group Meeting June 2nd 2011, Internal Document, Last accessed Januar 13th 2022.

[62] F. Fang, "Study of $\mathrm{K}^0_{\mathrm{S}} \to \pi^+\pi^-$ Selection", 2000. Belle Note 323, Internal Document.

[63] M. Feindt and U. Kerzel, "The NeuroBayes neural network package", *Nucl. Instrum. Meth.* **A559** no. 1, (2006) 190–194, 2006.

[64] H. Nakano, A. Ishikawa, and K. Sumisawa, "$\mathrm{K}^0_{\mathrm{S}}$ Selection with NeuroBayes and nisKsFinder Class", 2012. Belle Note 1253, Internal Document.

[65] P. Ecker, "Study of Bremsstrahlung Finding at the Belle II Experiment", Bachelor's Thesis, Karlsruhe Institute of Technology (KIT), 2018.

[66] G. C. Fox and S. Wolfram, "Event shapes in $\mathrm{e}^+\mathrm{e}^-$ annihilation", *Nuclear Physics B* **149** no. 3, (1979) 413–496, 1979.

[67] J.-F. Krohn, F. Tenchini, P. Urquijo, F. Abudinén, *et al.*, "Global Decay Chain Vertex Fitting at Belle II", *Nucl. Instrum. Methods Phys. Res. A* **976** (2020) 164269, 2020.

[68] M. Feindt, F. Keller, M. Kreps, T. Kuhr, *et al.*, "A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654** no. 1, (2011) 432–440, 2011.

[69] T. Keck, "FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification", *Computing and Software for Big Science* **1** no. 2, (Sep, 2017) 1 – 10, Sep, 2017.

[70] M. Röhrken, "Time-Dependent CP Violation Measurements: Analyses of Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment". Springer Theses. Springer, Cham, 2014.

[71] D. M. Asner, M. Athanas, D. W. Bliss, W. S. Brower, *et al.*, "Search for exclusive charmless hadronic B decays", *Phys. Rev. D* **53** (Feb, 1996) 1039–1050, Feb, 1996.

[72] G. C. Fox and S. Wolfram, "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes", *Phys. Rev. Lett.* **41** (Dec, 1978) 1581–1585, Dec, 1978.

[73] N. Braun, "Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment". Springer Theses. Springer International Publishing, 2019.

[74] O. Behnke *et al.*, "Data analysis in high energy physics: A practical guide to statistical methods". John Wiley & Sons, Ltd, Weinheim, 2013.

[75] G. Cowan, "Statistical Data Analysis", p. 216. Oxford Science Publications. Oxford Clarendon Press, Oxford, 1998.

[76] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *The European Physical Journal C* **71** no. 2, (Feb, 2011) 1554, Feb, 2011.

[77] F. James and M. Roos, "Minuit – a system for function minimization and analysis of the parameter errors and correlations", *Computer Physics Communications* **10** no. 6, (1975) 343 – 367, 1975.

[78] H. Dembinski, P. Ongmongkolkul, *et al.*, "scikit-hep/iminuit", Dec, 2020.

[79] E. O. Lebigot, "Uncertainties: a Python package for calculations with uncertainties". `http://pythonhosted.org/uncertainties/`.

[80] R. Glattauer and C. Schwanda, "Measurement of the decay $B \to D\ell\nu_\ell$ in fully reconstructed events and determination of the CKM matrix element $|V_{cb}|$", 2015. Belle Note 1345, Internal Document.

[81] F. Meier and A. Vossen, "Measurement of $B \to D^{**}\ell\nu_\ell$ branching fractions using Belle data and hadronic Full Event Interpretation", 2021. Belle Note 1569, Internal Document.

[82] P. Ongmongkolkul, C. Deil, C. hsiang Cheng, A. Pearce, *et al.*, "scikit-hep/probfit: Version 1.2.0", Mar., 2021. `https://doi.org/10.5281/zenodo.4595887`.

[83] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", *Nature Methods* **17** (2020) 261–272, 2020.

This thesis was written with the help of LaTeX using an adapted version of a class originally by Roland Bless, Timo Rohrberg and Thorsten Haberecht.
The color code follows the corporate design of the Karlsruhe Institute of Technology (KIT) (`https://www.sek.kit.edu/ci_cd.php`).