

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

FAKULTÄT FÜR PHYSIK

**Inclusive B -Meson Tagging for an $R(D^*)$ Measurement
at Belle II**

**Inklusives B -Meson Tagging für eine $R(D^*)$ Messung
bei Belle II**

Masterarbeit

in der Arbeitsgruppe Experimentelle Flavourphysik

vorgelegt von

Sofia Palacios Schweitzer

betreut von

Prof. Dr. Thomas Kuhr, Dr. Thomas Lück

München, 13. September 2022

Abstract

The world average of previous $R(D^{(*)})$ measurements, defined as $R(D^{(*)}) = \frac{\mathcal{B}(B \rightarrow D^{(*)} \tau \nu_\tau)}{\mathcal{B}(B \rightarrow D^{(*)} \ell \nu_\ell)}$ with $\ell = e, \mu$, shows a 3.3σ deviation from Standard Model predictions, which could indicate some New Physics phenomenon, such as the existence of Leptoquarks. This analysis uses simulations from the Belle II experiment at the SuperKEKB electron-positron collider to measure $R(D^*)$. To account for the challenge of multiple neutrinos as final state particles, an approach is considered, where besides the signal B -meson decay kinematical and topological properties of the other B -meson are reconstructed fully inclusively. In contrast to an exclusive reconstruction of the second B -meson used for previous $R(D^*)$ measurements by the B -factories, this inclusive approach suffers from a larger background, but also offers a higher reconstruction efficiency. As part of the analysis machine learning methods are studied to suppress different background sources. $R(D^*)$ is extracted from a two-dimensional extended maximum likelihood fit.

Contents

1	Introduction	7
1.1	Motivation	7
1.2	Belle II	8
1.3	Reconstruction Strategy	11
2	Theoretical Background	13
3	Data Sample & Simulation	15
3.1	Data	15
3.2	Simulation	16
4	Reconstruction	17
4.1	Signal Selection	18
4.2	Tag Selection	24
4.3	Event Selection	34
4.4	Reconstruction Efficiency	35
5	Background Analysis	37
5.1	Background Components	37
5.2	Background Suppression	40
5.3	Discussion of Applicability	42
6	Fit	48
6.1	Fit Procedure	48
6.2	Fit Results	50
6.3	Fit Validation	53
7	Uncertainties	55
7.1	PDF Estimation	56
7.2	$B \rightarrow D^{**} \ell \nu_\ell$ & Gap Mode	58
7.3	Efficiency	59
7.4	PID Corrections	59
7.5	$\tau \rightarrow \ell \nu_\ell \nu_\tau$ Branching Fraction	60
7.6	Other Systematic Uncertainties	61
8	Conclusion	62
	Bibliography	65
	Appendix	68

1 Introduction

1.1 Motivation

The Standard Model (SM) of particle physics describes the interaction of the most fundamental particles (as of now) and has been proven with success by multiple experiments all over the world.

However, after six decades of High Energy Physics experiments, we are also forced to accept that the picture painted by the SM is incomplete. There are still phenomena such as dark matter in the universe, the matter-antimatter asymmetry etc., that require

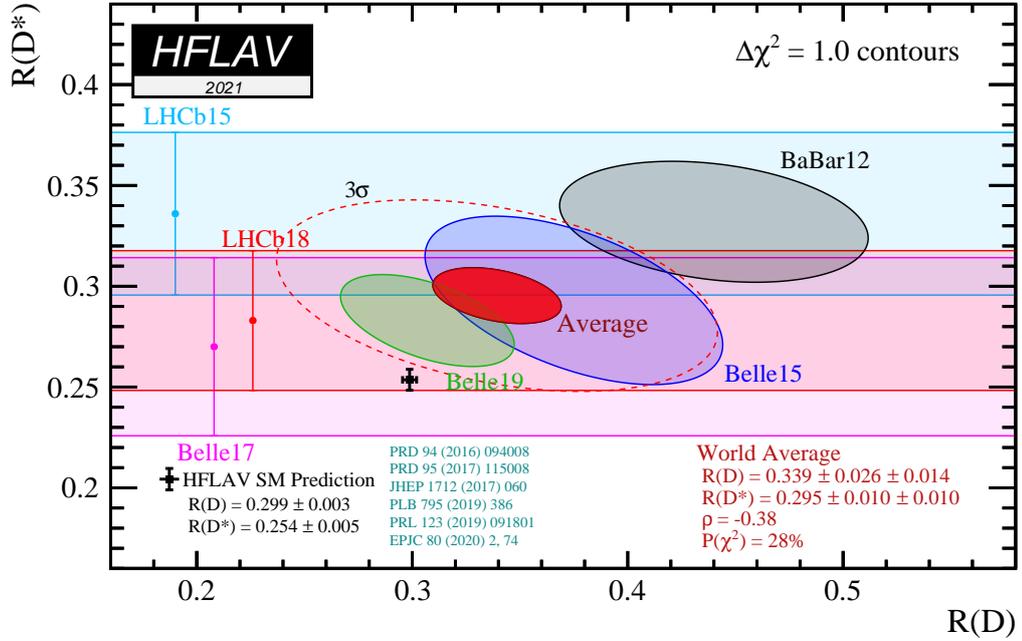


Figure 1.1: Theoretically calculated and experimentally measured values of $R(D^*)$ in dependency of $R(D)$ from [1]. The red dotted line represents the 3σ contour and the solid red ellipse the 1σ contour of the combined $R(D)$ and $R(D^*)$ averaged measurement. Note that the theoretical prediction lies outside of the 3σ region.

physics beyond the SM.

Such theories often rely on New Physics (NP) processes, which should affect physical observables in a measurable way. Therefore, any deviation from theoretical SM predictions and experimental measurements can hint to NP.

An ideal observable to study is $R(D^{(*)})$, defined in Eq. 1.1, with $\ell = e, \mu$.

$$R(D^{(*)}) = \frac{\mathcal{B}(B \rightarrow D^{(*)}\tau\nu_\tau)}{\mathcal{B}(B \rightarrow D^{(*)}\ell\nu_\ell)} \quad (1.1)$$

As most theoretical and experimental uncertainties cancel out, it allows us to precisely probe the SM. According to it, we expect that the W -Boson mediating the B -decay in both the denominator process $B \rightarrow D^{(*)}\tau\nu_\tau$ (referred to as signal mode) and the nominator process $B \rightarrow D^{(*)}\ell\nu_\ell$ (referred to as normalization mode) should couple equally strongly to each generation of leptons - this is the SM concept of lepton universality. This would translate to an $R(D^{(*)})$ -value that differs from one only because of phase space differences ($m_\tau \approx 3500 m_e$) and small form factor dependencies on the lepton's mass terms [2].

A combination of previous $R(D^{(*)})$ measurements from the first generation of the B -factories and the LHCb experiment have shown a 3.3σ deviation between theory and experiment [1], which is visualized in Fig. 1.1. However, next generation experiments are currently running, allowing us to soon access a larger amount of data, pushing our statistical limit even further and with that allowing us to perform a more precise $R(D^{(*)})$ measurement. This analysis focuses on measuring $R(D^{(*)})$ with simulations from the Belle II experiment in Tsukuba, Japan, which will be briefly introduced in Sec. 1.2 followed by a detailed description of the reconstruction strategy in Sec. 1.3. There are brief discussions of the theoretical background of $R(D^{(*)})$ in Ch. 2 and the Belle II data and simulation in Ch. 3. In Ch. 4 the concrete reconstruction selection is described. Dominant background sources and possible ways to suppress them are presented in Ch. 5. A two-dimensional extended maximum likelihood fit is used, to measure $R(D^{(*)})$, described in Ch. 6. Additionally, some systematic uncertainties are considered, which are summarized in Ch. 7, concluding in a discussion and an interpretation of the results in Ch. 8.

1.2 Belle II

The Belle II experiment is the successor of the Belle experiment and it is situated at the SuperKEKB accelerator.

SuperKEKB

The SuperKEKB can accelerate and collide electrons and positrons with asymmetric energies to a center-of-mass energy of $\sqrt{s} = 10.579$ GeV [4], which corresponds to the mass¹ of the $\Upsilon(4S)$ resonance [5].

Among other things, the $\Upsilon(4S)$ resonance, consisting of a $b\bar{b}$ -quark pair, can be produced in such a collision. It then decays further into a B - and a \bar{B} -meson with a branching fraction of above 96% [5]. With this procedure Belle II is expected to produce $5 \cdot 10^{10}$ B -meson pairs corresponding to an integrated luminosity of 50 ab^{-1} within its run time of 8 years [6], making it an ideal facility to investigate B -Physics. The first physics run was in 2019 and since then data corresponding to an integrated luminosity of 424 fb^{-1} [7] was collected. Fig. 1.2 shows a schematic overview of the SuperKEKB.

¹Throughout this thesis natural units $c = \hbar = 1$ are assumed.

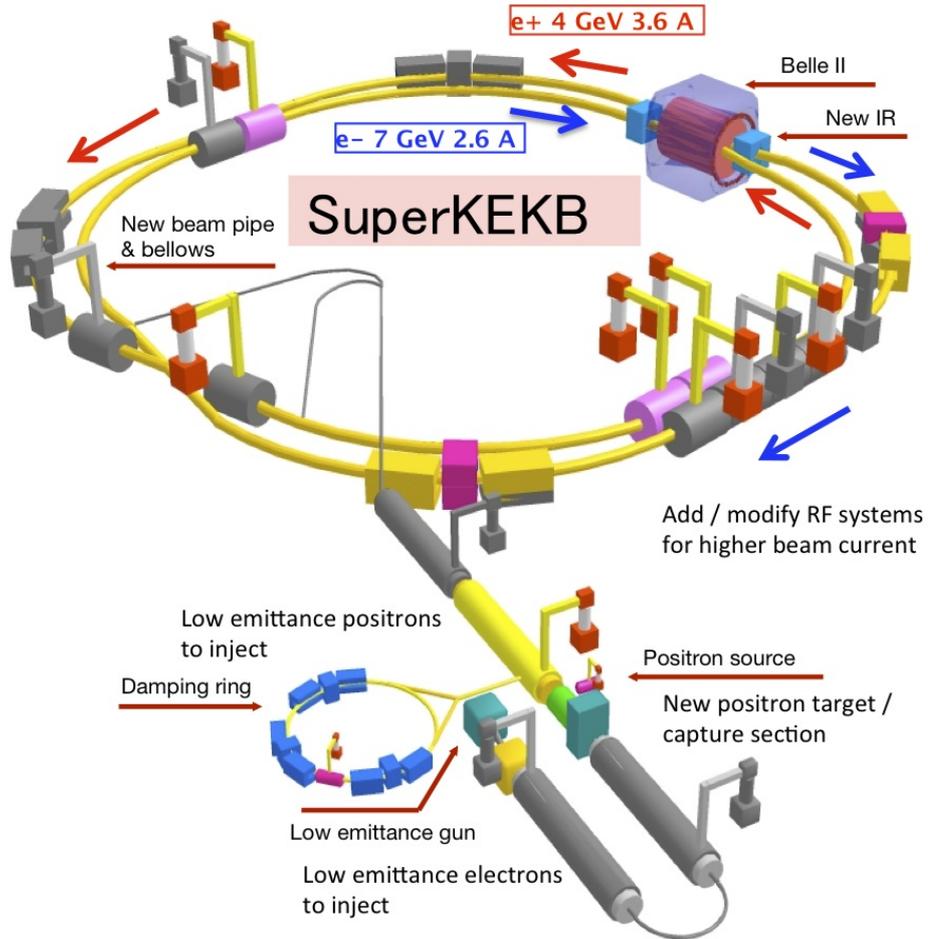


Figure 1.2: Schematic description of SuperKEKB from [3].

Belle II Detector

The Belle II detector, visualized in Fig. 1.3, consists of multiple layers and is described in detail in [6]. Due to the asymmetric energies of the collided electron and positron pair, the momentum of the produced particles in the collision is boosted in the direction of the electron. The region lying in the direction of the boost is referred to as the *forward region*. Most of the particles are expected to be detected here and hence a larger angular range is covered. The region on the opposite side is referred to as *backward region*, whereas the region orthogonal to the electron and positron current is called the *barrel region*.

To measure the B -meson's decay-vertex with high precision a vertex detector is situated close to the e^-e^+ -interaction point (IP). It is made out of two layers of pixilated silicon sensors (PXD) and four layers of double sided silicon strip detectors (DSSD), referred to as the silicon vertex detector (SVD). Silicon is ideal to measure the position

Belle II Detector

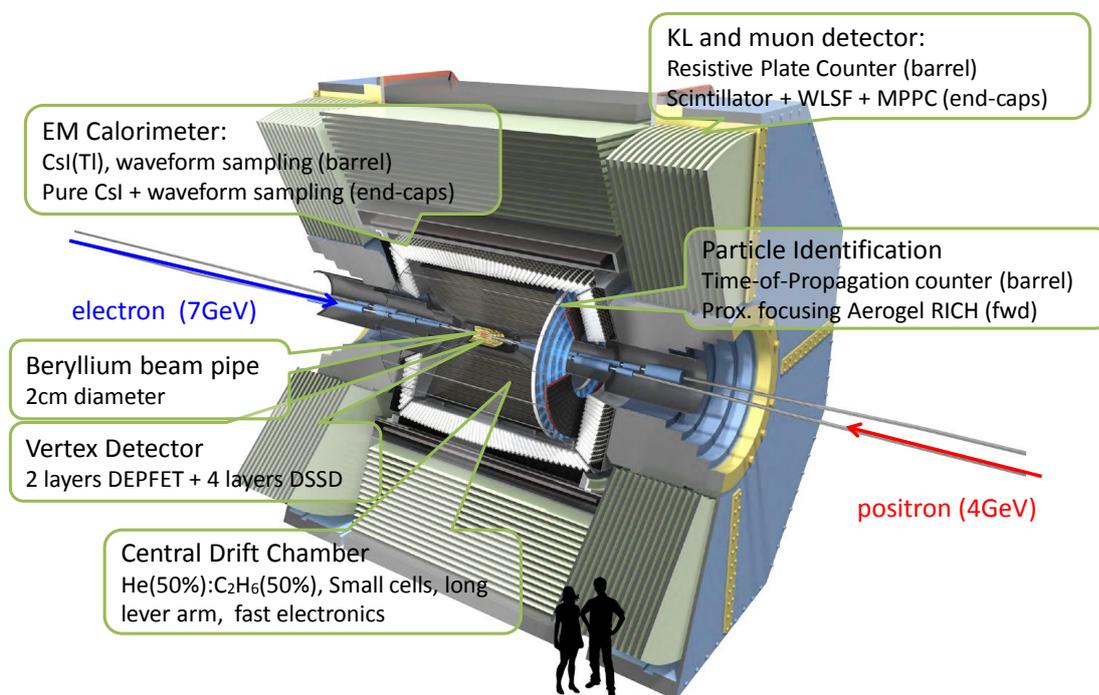


Figure 1.3: Commented overview of the Belle II detector from [3]. Note that the pure CSI and waveform sampling at the end-caps of the ECL has not been implemented yet.

of high energetic charged particles due to its atomic structure and the resulting semi-conductivity.

The vertex detector is followed by a central drift chamber (CDC) filled with a 50:50 mixture of $He - C_2H_6$. In the CDC the trajectories of charged particles are measured as well as their energy loss due to ionization of the gas as a function of their traveled distance $\frac{dE}{dx}$. Among other things this measurement is used for particle identification purposes.

Moving further away from the IP there are two particle identification systems: a time-of-propagation (TOP) counter located in the barrel region and an aerogel ring-image-cherenkov (ARICH) detector in the forward region. Both of them use the Cherenkov effect to distinguish between different charged particles.

They are followed by an electromagnetic calorimeter (ECL) consisting of thallium-doped caesium iodide CsI(Tl) crystals. Here, electromagnetic interacting particles can be detected as well as the energy they lose to electromagnetic showers within the ECL.

Muons, for example, usually surpass the ECL, which is why at the very end there is a K_L -Muon detector (KLM) made out of alternating layers of iron plates and detector material.

Between the ECL and the KLM there is a superconducting solenoid providing a magnetic field of 1.5 T to bend the tracks of charged particles [6].

1.3 Reconstruction Strategy

The experimental challenge of reconstructing semileptonic B -decays and gathering the full kinematical information of the decaying B -mesons lies in taking neutrinos into account. Neutrinos only interact weakly with a very low interaction cross section and therefore, are not detected directly by the Belle II detector.

The strategy used at the Belle II experiment to circumvent this problem is the so called *tagging* approach, which exploits the fact that B -mesons coming from a $\Upsilon(4S)$ -decay are always produced pairwise. So instead of only reconstructing those B -mesons which fulfil signal criteria (B_{sig}), the second B -meson (B_{tag}) is reconstructed as well.

With the kinematical information of the reconstructed B_{tag} -meson, the partially reconstructed B_{sig} -meson and the initial state, which is precisely known, neutrinos can now be detected indirectly: any difference between the four momentum of the initial e^-e^+ -state and the detected final state particles, namely the missing four momentum, should come from neutrinos as described in Eq. 1.2 (assuming a perfect reconstruction efficiency of the respective final state particles). An important related quantity for this analysis is the missing mass squared defined in Eq. 1.3.

$$p_{miss} = p_{e^-e^+} - p_{B_{tag}} - p_{B_{sig,part}} = \begin{pmatrix} E_{miss} \\ \vec{p}_{miss} \end{pmatrix} \quad (1.2)$$

$$M_{miss}^2 = p_{miss}^2 = E_{miss}^2 - |\vec{p}_{miss}|^2 = (E_{miss} - |\vec{p}_{miss}|)(E_{miss} + |\vec{p}_{miss}|) \quad (1.3)$$

However, there exist various tagging approaches, which vary in how exactly the tag-side is reconstructed. The most common approach at Belle II is an algorithm, the *Full Event Interpretation* (FEI), which iterates over multiple, highly resolved exclusive B -meson decays [8]. Those decays can generally be either fully hadronic or semileptonic. In both cases though the FEI only covers a small fraction of the possible B -decay's branching fractions, resulting in a small tag-side reconstruction efficiency, summarized in Tab. 1.1. Therefore, this analysis uses a different tagging approach: Instead of iterating over

Tags	Branching Fraction covered	Tag-Side Reconstruction Efficiency
Hadronic	1.1%	0.46%
Semileptonic	4.0%	2.04%

Table 1.1: FEI performance for hadronic and semileptonic tags from [8] and [9].

exclusive decay modes, the tag-side decay is not specified and the B_{tag} -meson is reconstructed inclusively meaning it is reconstructed directly from its final state particles. As the B_{tag} -meson is not restricted to certain decay modes, this approach promises a higher tag-side reconstruction efficiency. However, no decay-mode-specific cuts can be made, resulting in higher background levels and a worse kinematical resolution.

Because of the very clean signature of the D^{*-} -meson only $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ decays are reconstructed and thus, only neutral B^0 -mesons coming from $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ are of interest. Note that charge conjugated events are implied throughout the entire thesis.

The concrete reconstruction strategy for reconstructing $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ and $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$

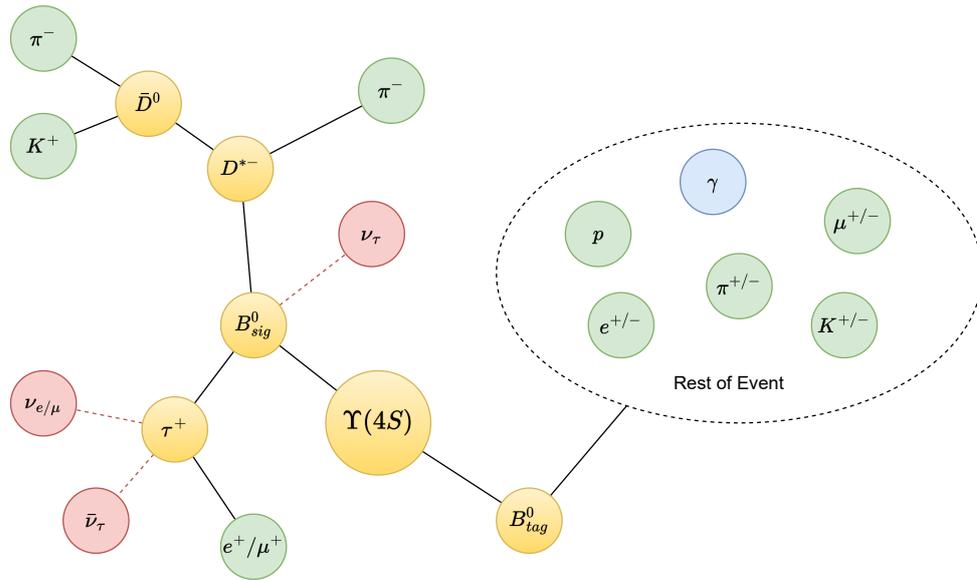


Figure 1.4: Reconstruction strategy for $B \rightarrow D^* \tau \nu_\tau$. All charged final state particles are colored green, neutral final state particles blue, intermediate particles yellow and not detectable particles (neutrinos) red.

simultaneously, schematically visualized in Fig. 1.4, is as follows:

1. The B_{sig}^0 -meson is reconstructed exclusively in $D^{*-} \tau^+ \nu_\tau$, where the D^{*-} -meson and the resulting D^0 -meson decays are fixed to $D^{*-} \rightarrow \bar{D}^0 \pi^-$ and $D^0 \rightarrow K^- \pi^+$ respectively. Additionally, only leptonic τ -decays are considered, so that both normalization and signal mode have the exact same final state particles.
2. The rest of event (ROE), meaning all of the remaining detected tracks and photons surviving certain selection cuts, are now used to reconstruct the B_{tag}^0 -meson without explicitly reconstructing any intermediate daughters (except of $V0$ particles introduced in Sec. 4.2).
3. The $\Upsilon(4S)$ -resonance is reconstructed out of the signal and tag B -meson.

Two similar analyses (using an inclusive tag) done by the Belle Collaboration serve as a rough guideline, where instead of $R(D^*)$ only the branching fraction of $B \rightarrow D^* \tau \nu_\tau$ were measured ([10] and [11]).

2 Theoretical Background

The $B \rightarrow D^* \ell \nu_\ell$ and $B \rightarrow D^* \tau \nu_\tau$ decay is described graphically by Fig. 2.1.

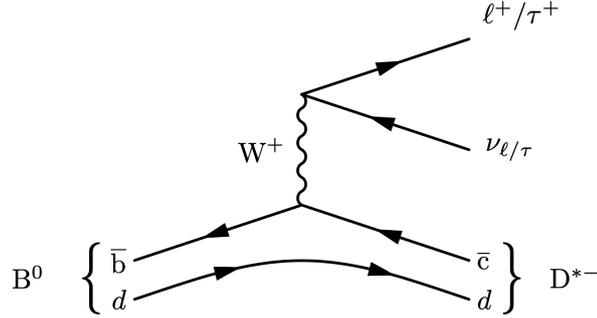


Figure 2.1: Feynman Diagram of $B^0 \rightarrow D^{*-} \ell^+ / \tau^+ \nu_{\ell/\tau}$.

Note that in this section we will have a closer look at the $b \rightarrow c \ell \nu_\ell$ transition as described in the SM, but the conclusions also hold for $b \rightarrow c \tau \nu_\tau$ (as on SM level we assume lepton universality) unless it is explicitly stated otherwise.

The relevant terms for this specific transition in the SM Lagrangian are given in Eq. 2.1. Here, g_w is the weak-coupling constant, V_{cb} is the relevant CKM-Matrix element, γ^μ are the Dirac matrices and $P_L = \frac{1}{2}(1 - \gamma_5)$ is the left-handed projection matrix.

On energy scales much below the mass of the W-Boson, we can approximate this interaction term by a point-like interaction, which then yields Eq. 2.2, where $L^\mu = \bar{\nu}_\ell \gamma^\mu P_L \ell$ is the weak lepton current, G_F the Fermi constant and $H_\mu = V_{cb} \bar{b} \gamma_\mu P_L c$ the weak quark current. On this level the theoretical framework of this problem looks straight forward.

$$\mathcal{L}_{SM} \supset -\frac{g_w}{\sqrt{2}} V_{cb} \bar{b} \gamma^\mu P_L c W_\mu^- - \frac{g_w}{\sqrt{2}} \bar{\nu}_\ell \gamma^\mu P_L \ell W_\mu^+ \quad (2.1)$$

$$\approx -2\sqrt{2} G_F L^\mu H_\mu \quad (2.2)$$

However, in reality mathematical difficulties arise when describing semileptonic B -decays as the quarks are not freely propagating but embedded in a hadronic structure. The internal structures of hadrons described by non-perturbative quantum chromodynamics (QCD) are a complex energy-dependent composition of interacting gluons and quarks. So when looking at the transition matrix of the quark transition, the physical initial and final state are the B - and the D^* -meson respectively, resulting in the hadronic part of the matrix element in Eq. 2.3.

$$\mathcal{M}_{had}^\mu = \langle D^* | H^\mu | \bar{B} \rangle \quad (2.3)$$

Four form factors are introduced to comprise all the QCD dependent terms of the hadronic matrix element and the calculation of those form factors constitutes a great theoretical challenge: As we are in non-perturbative QCD regimes, it is impossible to solve Eq. 2.3 analytically. In fact, we cannot even solve it purely theoretically but need some experimentally measured parameters and/or theories that approximate the SM.

There exist multiple of those approximate theories, which differ by their covered kinematical regime and their model assumptions. To describe semileptonic B -decays in the Belle II Monte Carlo (MC) simulation a combination of Heavy Quark Effective Theory (HQET) and Lattice QCD (LQCD) is considered [2], where experimental data is fitted to gather all necessary parameters for the actual calculation of the form factors. In addition, one can exploit analytical properties and QCD dispersion relations of the matrix element in Eq. 2.3 so that the form factors can be extrapolated from a point close to the extreme case of no recoil [2]. Depending on the concrete parameterization there are generally two different theoretical frameworks used to simulate semileptonic B -decays. One named after Boyd, Grinstein, and Lebed (BGL) [12] and another one named after Caprini, Lellouch and Neubert (CLN) [13]. Within the Belle II MC, the BGL model estimates $B \rightarrow D^* \ell \nu_\ell$ with six different parameters taken from [14], whereas $B \rightarrow D^* \tau \nu_\tau$ is described by the CLN model with five different parameters taken from [15].

When looking at the individual branching fraction of the signal or normalization mode, the calculation of the form factors causes the largest theoretical uncertainties. It should, however, cancel out in first order once the ratio of both, namely $R(D^*)$, is considered. Therefore, a more detailed description of the theoretical background of form factors is beyond the scope of this thesis.

3 Data Sample & Simulation

3.1 Data

Belle II data is collected either on-resonance, meaning at a center of mass energy corresponding to the $\Upsilon(4S)$ mass $M_{\Upsilon(4S)}$, or off-resonance, below $M_{\Upsilon(4S)}$.

Off-resonance data is mostly used to study the *continuum background*, in which instead of a $b\bar{b}$ -state other hadronic processes are produced.

The current total integrated luminosity of Belle II data amount to 424 fb^{-1} , its weekly distribution is shown in Fig. 3.1, whereas Belle II on-resonance data only amounts to

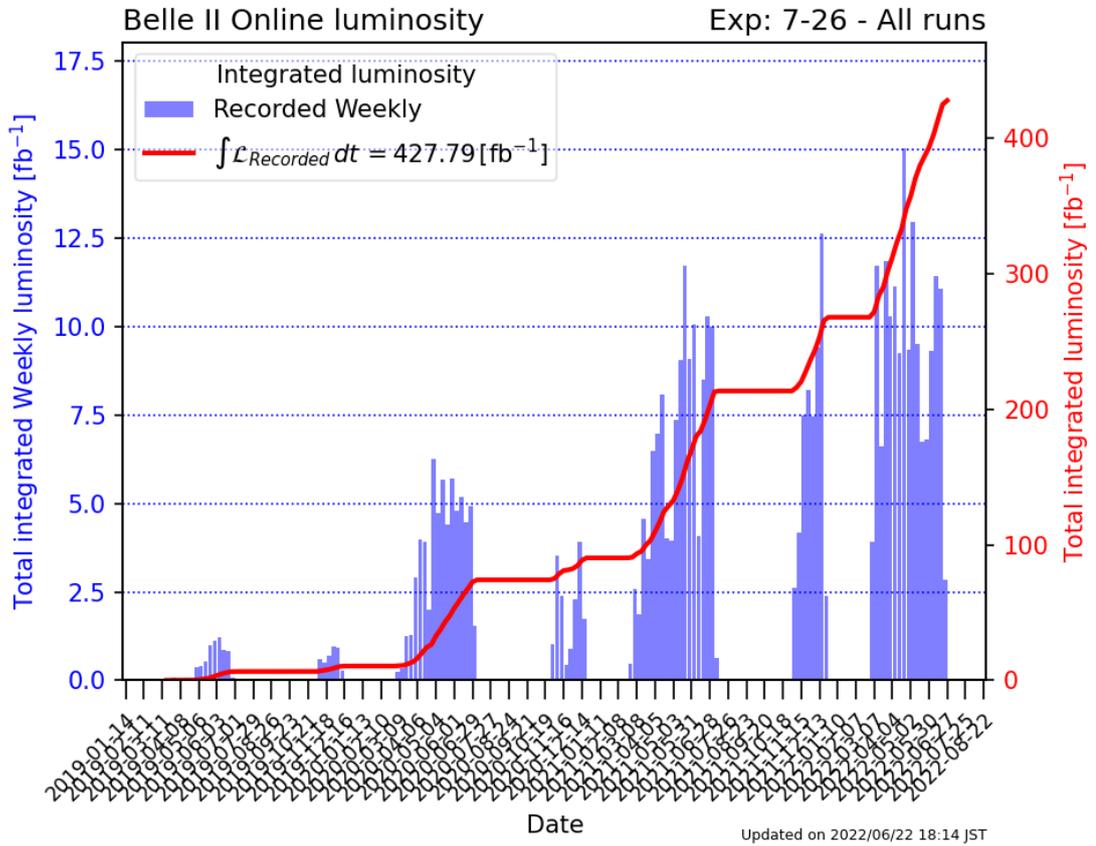


Figure 3.1: Weekly distribution of the total integrated luminosity of the Belle II data for all runs taken from [7].

363 fb^{-1} [7].

However, for analysis purposes for Moriond and ICHEP 2022 only processed data corresponding to 189.88 fb^{-1} were considered [16].

3.2 Simulation

For the Belle II MC simulations, exclusive B - and D -meson decays are generated with EvtGen 2.2.0 [17] following predefined decay tables containing branching fractions, which are regularly updated to the values of [5]. The uncertainties on the measured branching fractions are considered a systematic uncertainty in the context of this analysis. Deviations between branching fractions used for simulation and their PDG value for relevant particle decays need to be corrected, when running over data. For continuum background the KKMC is used as generator [18], hadronic shower processes are simulated with PYTHIA [19] and the simulation of the detector response is performed by Geant4 [20].

Three different kinds of MC simulations are considered to reconstruct $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ according to the reconstruction strategy introduced before.

To gain a first insight into the signal mode and chose first signal selection cuts, signal MC containing 80 million $B^0 \bar{B}^0$ events is used. Here, the B_{sig}^0 -decay is fixed to $B_{sig}^0 \rightarrow X_c^- \tau^+ \nu_\tau$, whereas the X_c^- , being either a D^- - or D^{*-} -meson, the τ^+ - as well as the B_{tag}^0 -decay is generic, meaning all possible decays (of the simulation) are present.

To get a better understanding of the tag-side reconstruction 10000 events are generated, where the signal-side decays to $B_{sig}^0 \rightarrow D^{*-} \ell^+ \nu_\ell$ and the tag-side is also fixed to a single decay listed in the appendix.

To optimize the event selection and apply ideal selection cuts, generic MC is considered, where both the B_{sig}^0 - and the B_{tag}^0 -meson decay generically.

The fit is performed on generic MC corresponding to 200 fb^{-1} to test its statistical sensitivity to the Moriond and ICHEP 2022 data sets of 189.88 fb^{-1} .

All of the MC samples are *run-independent*. In contrast to run-dependent MC, where beam induced background is added using actual data collected by the Belle II experiment, run-independent MC relies on beam background simulated with SAD [21].

4 Reconstruction

The reconstruction and analysis is done with the Belle II analysis software (basf2) [22]. The final selection cuts for signal- and tag-side are determined by taking into account purity P , reconstruction efficiency ϵ , the figure of merit \mathcal{F} and the sample standard deviation of a variable dependent distribution λ , all defined in Eq. 4.1.

$$P = \frac{N_{sig}}{N_{sig} + N_{bkg}}, \quad \epsilon = \frac{N_{sig}}{N_{gen}}, \quad \mathcal{F} = \frac{N_{sig}}{\sqrt{N_{sig} + N_{bkg}}}, \quad \lambda = \sqrt{\frac{1}{N_t - 1} \sum_{i=1}^{N_t} (x_i - \bar{x})^2} \quad (4.1)$$

N_{sig} (N_{bkg}) refers to the number of rightly reconstructed signal events (background events), N_{gen} to the number of generated signal events on MC level and $N_t = N_{sig} + N_{bkg}$ to the total number of events. x_i stands for the value of the i th observation of a variable x and \bar{x} for the respective sample mean.

The purity returns the fraction of signal events over all reconstructed events, which measures how much background contamination is present in the regarded sample. A small purity corresponds to a small number of signal events relative to the total number of reconstructed event.

The efficiency returns the fraction of signal events reconstructed over the number of signal events generated in the regarded sample and with that serves as a measurement on how well signal events are reconstructed.

The figure of merit similar to the purity is a measurement of how much signal compared to background events are present. It gives the number of signal events in terms of the statistical uncertainty loosely approximated by $\sqrt{N_{sig} + N_{bkg}}$.

The last variable defined in Eq. 4.1, the sample standard deviation λ , is a measurement of the resolution of a certain variable distribution. Higher values of λ signify a greater dispersion of a sample around its mean indicating a lower resolution of said variable.

Tightening selection cuts can increase purity, resolution and figure of merit, but will decrease the efficiency. Therefore, optimization should try to keep the loss in efficiency low and the gain in the other performance measurements high.

It is assumed that the number of total events are Poisson distributed with mean N_t , so that the corresponding uncertainty can be estimated by $\sqrt{N_t}$. Meanwhile, the number of selected events N_{sel} from a sample of size N , as an event is either selected or not, follows a binomial distribution. The uncertainty associated to the number of selected events is therefore given by the binomial error $\sigma_{binomial} = \sqrt{N_{sel}(N - N_{sel})}$. The resulting uncertainty in the efficiency is described in Eq. 4.2.

$$\delta\epsilon = \sqrt{\frac{\epsilon(1 - \epsilon)}{N_{gen}}} \quad (4.2)$$

The uncertainty of the sample standard deviation is estimated by $\delta\lambda = \frac{\lambda}{\sqrt{2N_f}}$. For functions of multiple erroneous variables, the uncertainty is computed via the Gaussian error propagation.

4.1 Signal Selection

The leptonic side of the signal mode is considered first. In contrast to the normalization mode, the light leptons coming from the $\tau \rightarrow \ell\nu_\ell\nu_\tau$ decay are so called *secondary* particles, meaning they are no direct daughters of the B -meson. Their absolute three-momentum in the center of mass frame $|\vec{p}_{\ell, CMS}|$ is hence distributed at lower values. Nonetheless, a cut at $|\vec{p}_{\ell, CMS}| > 0.3$ GeV can still significantly reduce beam background

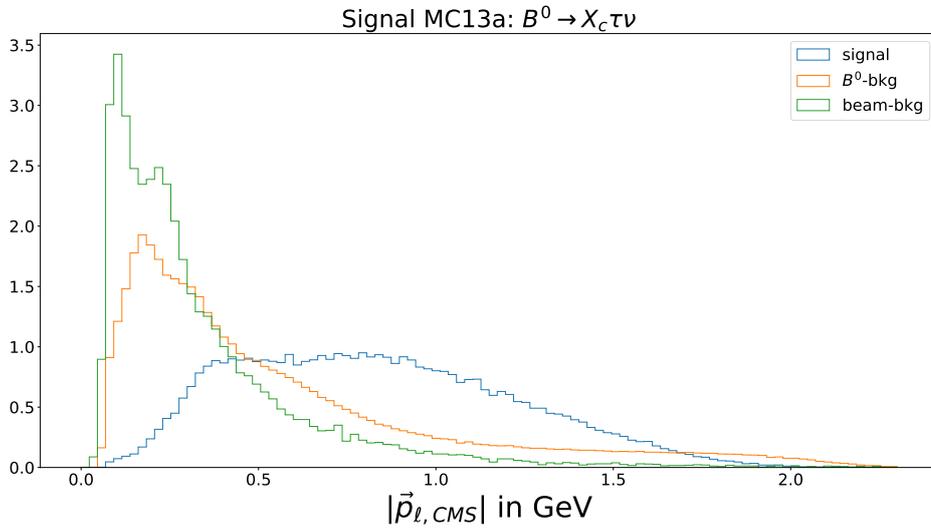


Figure 4.1: Normalized distribution of the absolute value of the lepton's three-momentum in the center-of-mass frame. Signal refers to rightly reconstructed $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ events, whereas B^0 -background to light leptons coming from other B^0 -decays. Additionally, particles not carrying any MC information, which is typical for beam background particles, are also included.

peaking at low lepton momentum, see Fig. 4.1.

Furthermore, using information from all different detectors, one can access a so called *particle identification probability* (PID) for charged final state particles, which is defined as $P_{id} = \frac{\mathcal{L}_P}{\sum_Q \mathcal{L}_Q}$. Here, \mathcal{L}_P is the likelihood of a charged particle being of type P , and the sum over Q runs over all charged final state particles (electron, muon, pion, kaon, proton and deuteron).

In Fig. 4.2 the electronID e_{id} (muonID μ_{id})¹ distribution of the signal electron (muon) is

¹PID corrections introduced in Ch. 7 were computed with PID values, where no SVD information contribute and in addition for the electronID no TOP information is included. Therefore, in this analysis the variables electronID_noSVD_noTOP and muonID_noSVD are used for the lepton selection, but they are still referred to as electronID and muonID respectively.

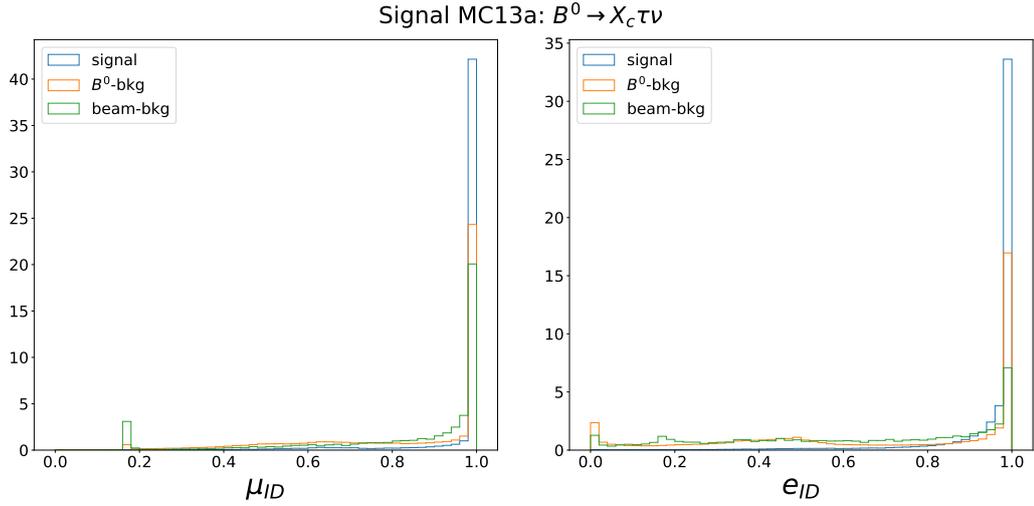


Figure 4.2: Normalized distribution of muonID μ_{id} and electronID e_{id} for the respective lepton channel.

shown, indicating that it could provide some discrimination power between signal and background events, which is why a cut at $\mu_{id} > 0.9$ and $e_{id} > 0.9$ is made.

The light leptons are expected to be produced in a tight region around the IP visible in Fig. 4.3. Therefore, further restrictions on the transversal distance dr as well as on the absolute value of the distance projected on the z-axis dz between the point of closest approach (POCA) and the measured interaction point are made.

Lastly, as a consequence of the limited angular coverage of the detectors, not all tracks are detected. To be more concrete, around 10% of tracks lie outside the angular acceptance of the CDC [6], qualitatively visible in Fig. 4.4, where MC particles are shown from 189000 generic $B^0\bar{B}^0$ events. Therefore, the polar angle θ of the tracks corresponding to the light leptons are to lie within the CDC acceptance region $\theta \in [17^\circ, 150^\circ]$ [6].

The electron is brems-corrected using internal recommendations [23]. This means that photons located in a cone around the electron tracks are assumed to be bremsstrahlung, so that their four momentum is added to the electron's. According to the recommendation, depending on the electron three-momentum, the potential photon must be below a certain energy E_γ and lie within a cone around the leptons direction of flight of a certain opening angle θ_γ :

1. $0.3 \text{ GeV} < |\vec{p}_\ell| \leq 0.6 \text{ GeV}$: $E_\gamma < 0.09 \text{ GeV}$ & $\theta_\gamma < 0.1368 \text{ rad}$
2. $0.6 \text{ GeV} < |\vec{p}_\ell| \leq 1 \text{ GeV}$: $E_\gamma < 0.9 \text{ GeV}$ & $\theta_\gamma < 0.0737 \text{ rad}$
3. $1 \text{ GeV} < |\vec{p}_\ell|$: $E_\gamma < 1.2 \text{ GeV}$ & $\theta_\gamma < 0.0632 \text{ rad}$

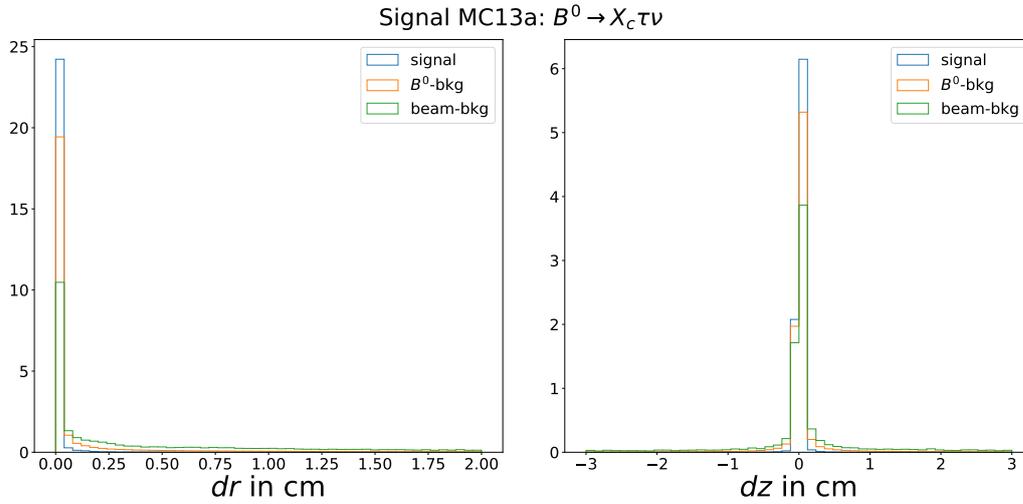


Figure 4.3: Normalized distribution of the transversal and z-distance between POCA and IP of the light lepton.

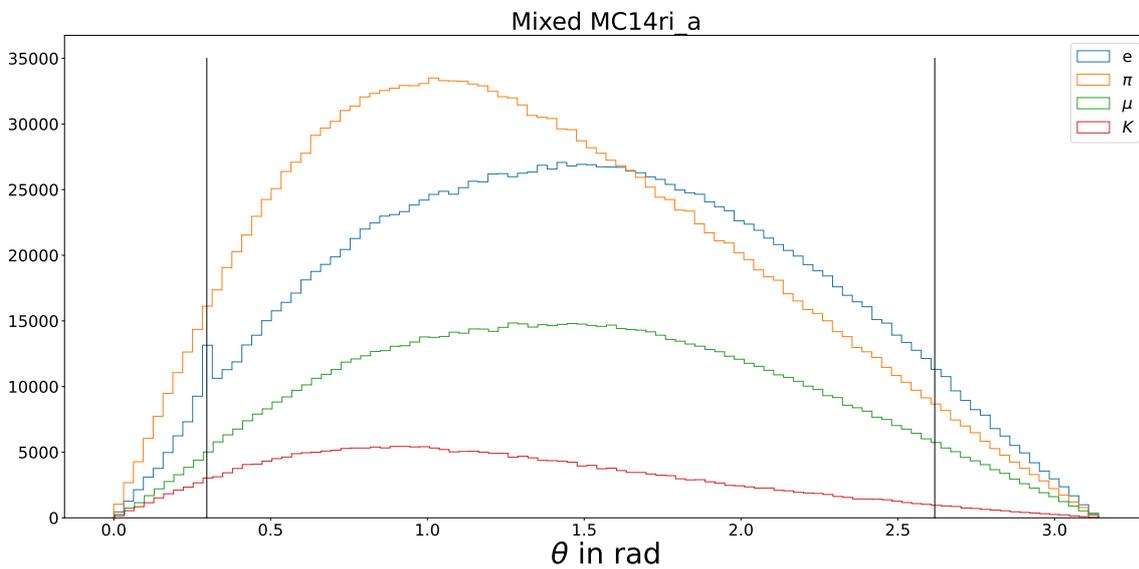


Figure 4.4: The polar angle θ of MC particles for different charged particles. The solid black line correspond to the threshold values of the CDC acceptance region. Tracks outside the region are not detected.

As mentioned before, the tagging approach used for this analysis suffers from high background levels. Therefore, when looking at the hadronic side of the signal decay, only specific D^{*+} and D^0 -decay channels with high purity and resolution are considered in the context of this analysis.

The D^{*+} -meson decays mainly to $D^{*+} \rightarrow D^0 \pi^+$ with a branching fraction of $(67.7 \pm 0.5)\%$ or to $D^{*+} \rightarrow D^+ \pi^0$ with a branching fraction of $(30.7 \pm 0.5)\%$ [5]. As the mass difference between D^{*+} and D^0 (D^+) is relatively small $\Delta M = M_{D^{*+}} - M_{D^0(D^+)} = (145.4258 \pm 0.0017)$ MeV ((140.603 ± 0.015) MeV) [5], one can expect the accompanying pion to have very little momentum in the center of mass frame.

In fact, taking four momentum conservation into account as done in Eq. 4.3, one can compute the absolute value of the pion's three-momentum in the center-of-mass frame of the decaying D^{*+} -meson. Boosting it into the $\Upsilon(4S)$ center-of-mass frame an upper bound can be very loosely approximated by $|\vec{p}_{\pi, CMS}| < 0.4$ GeV.

$$|\vec{p}_{\pi, D^* - CMS}| = \left(\frac{1}{4M_{D^{*+}}^2} (M_{D^{*+}}^2 + M_{\pi}^2 - M_{D^0(D^+)}^2) - M_{\pi}^2 \right)^{1/2} \quad (4.3)$$

Although the tracking efficiency for low momentum charged pions is relatively low

D^0 -Decay Modes	$\mathcal{B}[\%]$	$\epsilon_{sig} [\%]$	P [%]	$\lambda [10^{-4}]$
$K^- \pi^+ \pi^0$	14.4 ± 0.5	4.335 ± 0.016	7.84 ± 0.03	232.56 ± 0.17
$K^- \pi^+ \pi^- \pi^+$	8.22 ± 0.14	7.640 ± 0.027	5.685 ± 0.019	223.64 ± 0.13
$K_S^0 \pi^+ \pi^- \pi^0$	5.2 ± 0.6	2.553 ± 0.020	2.110 ± 0.017	225.53 ± 0.19
$K^- \pi^+$	3.946 ± 0.030	15.08 ± 0.05	20.44 ± 0.08	137.67 ± 0.17
$K_S^0 \pi^+ \pi^-$	2.80 ± 0.18	6.02 ± 0.04	5.68 ± 0.04	237.72 ± 0.28
$K_S^0 \pi^0$	1.239 ± 0.022	5.75 ± 0.06	13.91 ± 0.15	230.1 ± 0.7

Table 4.1: Branching fractions taken from [5] and performance of different D^0 -decay modes measured in signal reconstruction efficiency ϵ_{sig} , purity P and sample standard deviation λ in M_{D^0} (in the region $1.8 \text{ GeV} < M_{D^0} < 1.9 \text{ GeV}$). Note that for the purity calculation continuum background is not included.

[24], reconstructing the π^0 from two low energetic photons via $\pi^0 \rightarrow \gamma \gamma$ introduces much more background as at low energies beam background photons are dominant. Consequently, the D^{*+} -decay mode including a slow π^0 is discarded for this analysis.

For the resulting D^0 -meson six different decay modes are considered listed with their respective branching fractions in Tab. 4.1. The selection of those is mostly inspired by previous $R(D^*)$ measurements ([25] and [26]).

Some loose cuts at $0.13 \text{ GeV} < \Delta M < 0.16 \text{ GeV}$ and $1.8 \text{ GeV} < M_{D^0} < 1.9 \text{ GeV}$ are already in place as well as further selection cuts for the daughter particles: K_S^0 -mesons are to have a mass between $0.468 \text{ GeV} < M_{K_S^0} < 0.528 \text{ GeV}$ and furthermore, they are to pass a pre-implemented Belle selection, based on flight distance, impact parameter, angle between the daughters and vertex information.

π^0 involved are reconstructed in $\pi^0 \rightarrow \gamma \gamma$ and need to have an invariant mass between $0.11 \text{ GeV} < M_{\pi^0} < 0.16 \text{ GeV}$. The two daughter photons are to have a difference of the azimuthal angle of their momenta smaller than $\delta\phi_p < 1$ rad and a difference of

their cluster angles smaller than $\delta\phi_{cluster} < 0.9$ rad. Additionally, cuts on background suppression classifiers are applied to the resulting photons in analogy to Sec. 2, where a detailed description of those classifiers can be found.

For the remaining charged particles, the final selection cuts are applied, summarized at the end of this section.

Tab. 4.1 comprises the channel specific performances of efficiency, purity (without continuum background) and the sample standard deviation in M_{D^0} . Differences in

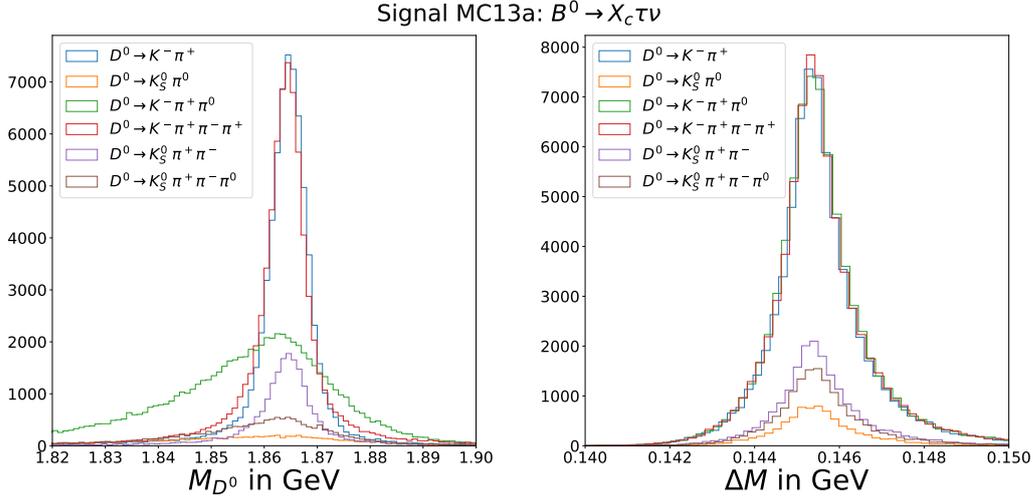


Figure 4.5: Distribution of the invariant mass of the reconstructed D^0 -meson and the mass difference ΔM of the D^{*+} - and D^0 -meson in six different channels. Only correctly reconstructed events are considered.

performances are qualitatively visualized in Fig. 4.5, which shows the mass of the reconstructed D^0 -meson as well as the mass difference ΔM . One can clearly see that the resolution of the former is highly channel dependent, whereas for the latter variable resolution differences mostly cancel.

Taking the measured performance into account, this analysis focuses on the clean signature of the $D^0 \rightarrow K^- \pi^+$ decay. Other modes suffer from either a low purity and/or resolution making background rejection an even bigger problem.

For all the final state particles present in the decay channel chosen, again restrictions on their distance to the IP as well as their allowed polar angle region are made. For kaons (pions) coming from the $D^0 \rightarrow K^- \pi^+$ decay additional cuts on the kaonID K_{id} (pionID π_{id}), see Fig. 4.6, are applied.

As a last step of the signal selection, the mass of the D^0 -meson as well as the mass difference ΔM are restricted. The exact cut value are chosen more loosely by eye rather than by a figure of merit optimization as no data-MC comparison has been done yet and therefore, it is avoided to cut onto the tails of the respective peaking mass distributions. Tighter restrictions can be considered in further studies.

The absolute value of the D^{*+} -meson's three-momentum in the center of mass frame is bounded from above to suppress high energetic D^{*+} -mesons produced in $c\bar{c}$ -background.

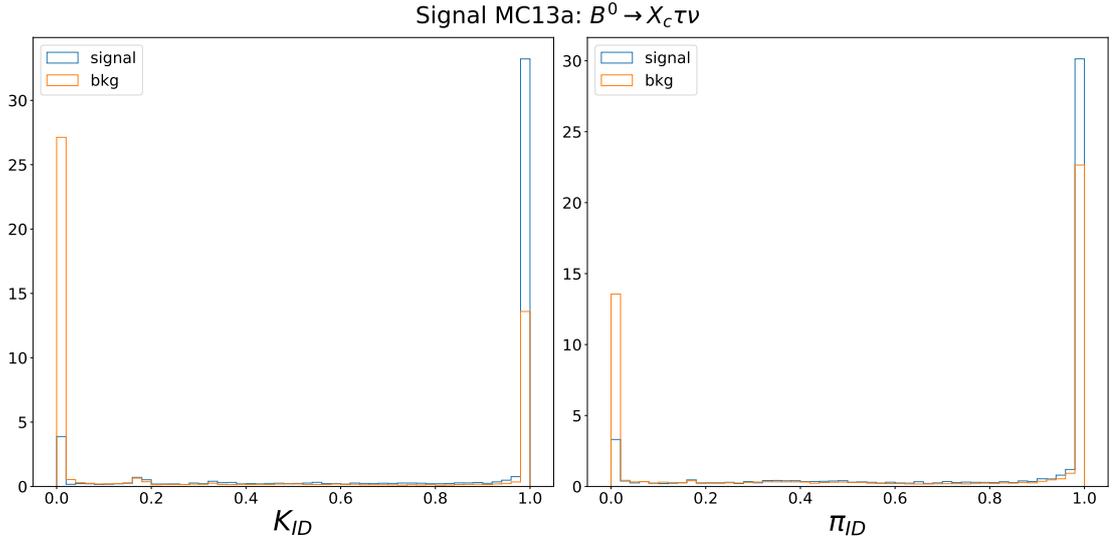


Figure 4.6: Normalized distribution of the kaonID K_{id} and pionID π_{id} coming from $D^0 \rightarrow K^- \pi^+$.

Summarizing the selection of the B_{sig}^0 -meson decay, the following cuts are applied:

- all tracks:
 - $dr < 0.5 \text{ cm} \ \& \ |dz| < 2 \text{ cm}$
 - $17^\circ < \theta < 150^\circ$
- light lepton selection:
 - $|\vec{p}_{\ell, CMS}| > 0.3 \text{ GeV}$
 - $e_{id} > 0.9$ or $\mu_{id} > 0.9$
- light hadron selection:
 - $K_{id} > 0.5$ or $\pi_{id} > 0.1$ (not for slow pion)
 - $|\vec{p}_{\pi, CMS}| < 0.4 \text{ GeV}$ (only for slow pion)
- D^{*-} & D^0 -mesons:
 - $1.84 \text{ GeV} < M_{D^0} < 1.88 \text{ GeV}$
 - $0.14 \text{ GeV} < \Delta M < 0.15 \text{ GeV}$
 - $|\vec{p}_{D^{*+}, CMS}| < 3 \text{ GeV}$

The exact cut values used for the light lepton and hadron selection are chosen to optimize the figure of merit on a generic run-independent MC sample including continuum background and corresponding to 100 fb^{-1} .

Note that the vertex of the B^0 -, D^{*-} - and D^0 -meson are extrapolated via a vertex fit and failed fits are discarded. The resulting distributions of the χ -probability do not provide any additional separation power between signal and background.

4.2 Tag Selection

The tag-side reconstruction is assumed to be independent of the exact decay mode considered on the signal-side if reconstructed correctly. Thus, especially for signal and normalization mode, as they share the exact same final state particles, their tag-side reconstruction efficiency is considered to be the same.

Therefore, within this section the signal-side is reconstructed to $B_{sig}^0 \rightarrow D^{*-} \ell^+ \nu_\ell$, so that higher statistics can be achieved.

As the B_{tag}^0 -meson is reconstructed inclusively, the tag-side selection is less conclusive as the signal selection. One reason being limited access on MC *truth match* information. When reconstructing particles from MC simulations one usually has the option to check whether or not the reconstructed particle corresponds to the actual MC particle, which is called *truth matching*. Hence, one can validate if an event has been reconstructed correctly.

However, the algorithm depends on correct mother-daughter relations, meaning that if no intermediate daughters are reconstructed, no information on the correctness of the reconstruction is available. For a correctly reconstructed signal-side though, by definition all remaining particles are either beam background or coming from the B_{tag}^0 -meson. Within the newest light-releases of basf2 some changes concerning the algorithm have been made to widen its applicability [22]. Nonetheless, for the inclusive reconstruction discussed here, the algorithm still returns non-conclusive results. Consequently, concerning the tag-side selection no truth match information is relied on and with that no figure of merit optimization can be done. Possible systematic uncertainties arising from those circumstances are discussed in Ch. 7.

Instead the selection is done by trying to optimize the signal resolution of the energy difference ΔE_{tag} in the signal region $|\Delta E_{tag}| < 0.5$ GeV, while cutting on the beam constraint mass of the B_{tag}^0 -meson $M_{bc,tag} > 5.27$ GeV. Both of those variables are defined in Eq. 4.4. Here, E_B refers to the reconstructed energy of the B -meson, E_{beam} to the beam energy both in the center of mass frame and $\vec{p}_{B,CMS}$ to the reconstructed three-momentum of the B -meson in the center of mass frame.

$$\Delta E = E_B - E_{beam} \quad M_{bc} = \sqrt{E_{beam}^2 - |\vec{p}_{B,CMS}|^2} \quad (4.4)$$

For B -mesons ΔE is expected to peak around zero and M_{bc} to peak around the B^0 -mass $M_{B^0} = (5.27966 \pm 0.00012)$ GeV [5], if reconstructed correctly.

Another difficulty in an inclusive reconstruction is the lack of channel dependent cuts, instead only cuts on final state particles can be applied.

Those final state particles are either charged tracks or photons, and generally they could have any possible degree of relationship to the B_{tag}^0 -meson. Therefore, selection cuts that keep a wide energetic range of desirable photons and tracks, but discard background effectively are needed.

This poses a big challenge as one can assume that e.g. a primary photon and bremsphotons of a secondary electron behave extremely differently.

The photon and track selection are treated separately. Note that K_L^0 -meson do generally not decay inside the detector, so that they are (ideally) detected directly in the KLM and

have to be treated as final state particles as well. However, given the recently observed data-MC differences concerning the KLM they are not included in the ROE.

Moreover, to discard any events, where additional neutrinos are present on the tag-side, a *lepton veto* is imposed, which will be described in the following subsection.

Lepton Veto

Throughout the photon and track selection sections the lepton veto is already in place, so that no events, which will be discarded, influence the optimization. Therefore, this is discussed first.

As mentioned before, the goal is to suppress tag-side events containing neutrinos. Neutrinos are not detected directly though, meaning that they cannot be suppressed directly either. But as neutrinos are produced most likely in combination of the respective charged lepton, suppressing them is the next best thing. Furthermore, as tauons are not stable only their final state particles in the ROE can be considered, so that tauons cannot be suppressed directly either. As a consequence, the lepton veto can only be imposed

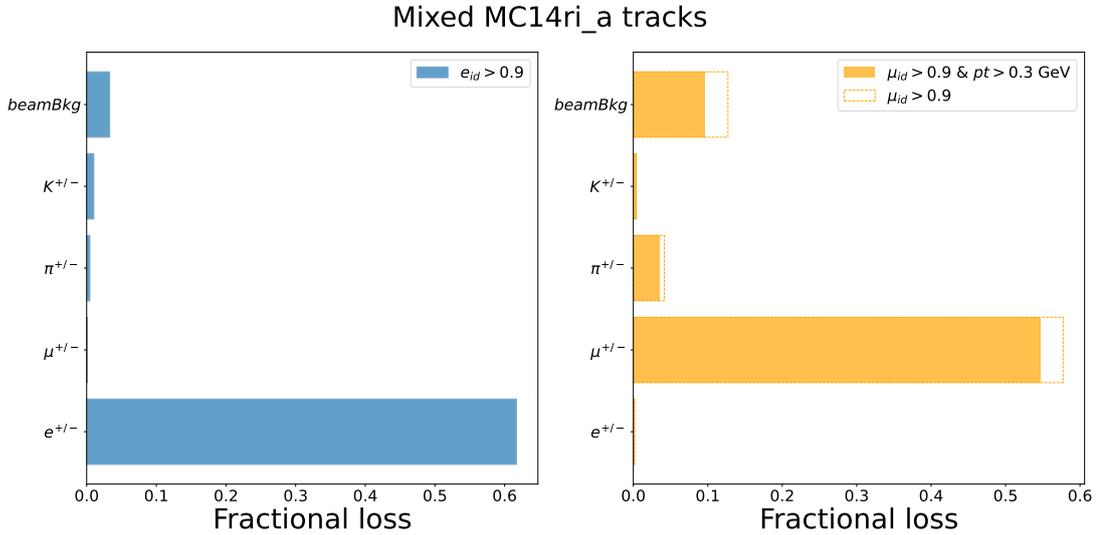


Figure 4.7: Fraction loss of different charged final state particles present in 189000 $B^0\bar{B}^0$ events, when imposing the lepton vetos defined in the text. Note that for the purpose of this diagram the track selection discussed later is already in place.

on light leptons.

At first, the lepton veto is implemented to identify events in which tracks with a high leptonID are present. Those events are then vetoed. Similarly to the the optimization of the leptonIDs in the signal selection, as threshold values $e_{id} = 0.9$ and $\mu_{id} = 0.9$ are chosen.

When evaluating the performance of the lepton veto, run-independent generic MC is used containing 189000 $B^0\bar{B}^0$ -events, from which tracks are reconstructed. Fig. 4.7 shows how many of those tracks lie within the constraint regions sorted by which particle they correspond to on MC level.

It appears that if those threshold values on the leptonIDs are used, around 60% of tracks

corresponding to electrons and 60% of tracks corresponding to muons in the ROE (if the track selection is already in place) are suppressed. The remaining events containing light leptons consequently would not be discarded.

Besides, within the region restricted by the μ_{id} a significant amount of misidentified

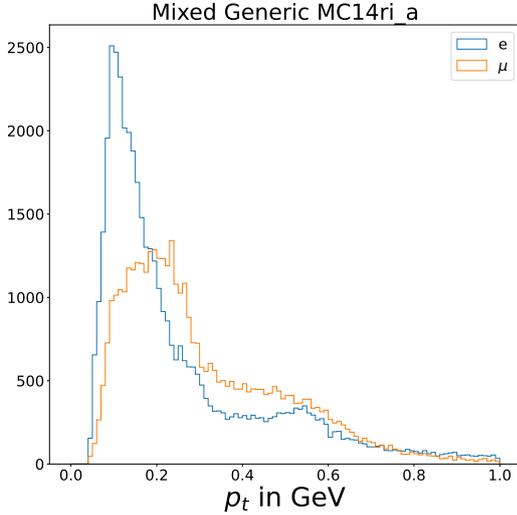


Figure 4.8: Transversal momentum p_t distribution of light leptons surviving the lepton veto.

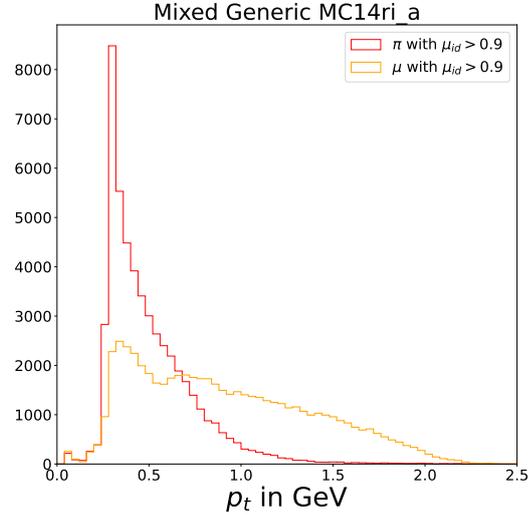


Figure 4.9: Transversal momentum p_t distribution of muons and pions lying in the restricted region.

charged pions are present. Events containing those pions would then be discarded as well.

To further investigate the performance of the lepton veto, the absolute value of the three-momentum of the light lepton lying outside of the restricted region as shown in Fig. 4.8 is considered. Clearly, only low momentum electrons and muons survive the leptonID restrictions. As is visible in Fig. 4.9 misidentified pions falling inside of the restricted region are also of low momentum. To save events containing those pions from rejection, the restricted region is tightened by additionally constraining the tracks with $\mu_{id} > 0.9$ to have a transversal momentum above $p_t > 0.3$ GeV. The fractional loss of different charged final state particles resulting from imposing this modified muon veto is also visualized in Fig 4.7.

Note that although not all leptons are suppressed, the tag is now referred to as being hadronic and inclusive.

Photon Selection

In an e^-e^+ -collision creating an hadronic state, there are a lot of decay processes possible in which photons are produced. Therefore, the average number of photons per B -event is relatively high. However, there are also two dominant background sources to B -meson processes.

Firstly, due to Belle II's high luminosity the experiment suffers from a lot of beam induced background. This background summarizes all kinds of processes in which photons are radiated by the beam.

Secondly, there exists the so called *split off* background, which can be a result of hadrons

interacting with the ECL, but only depositing a part of their total energy in their primary shower. Within the primary shower neutral particles are produced which travel some distance through the ECL after which they produce a displaced *split off* shower, which mimics low energy photons. It can also be caused by mismatched ECL cluster hits.

In an inclusive reconstruction those background sources should be suppressed, otherwise the B_{tag}^0 -meson's energy is overestimated. However, it is equally important to keep as much photons produced in B^0 -decays as possible, otherwise the reversed case would occur, the B_{tag}^0 -meson's energy would be underestimated.

Within the basf2 two pre-implemented multivariate classifier (MVA) were built by members of the collaboration in [27] to optimally suppress beam background as well as split off background individually. To train those classifiers a combination of the following ECL variables served as input variables:

1. Leakage and background corrected energy deposited in all ECL crystals within the cluster `clusterE`
2. Ratio of the energy deposited in the central ECL crystal and the energy deposited in the nine neighboring crystals `clusterE1E9`,
3. Lateral energy distribution `clusterLAT`, defined as:

$$E_{lat} = \frac{\sum_{i=2}^n w_i E_i r_i^2}{(w_0 E_0 + w_1 E_1) r_0^2 + \sum_{i=2}^n w_i E_i r_i^2} \quad (4.5)$$

Here, E_i is the energy deposited in the crystal i sorted in descending order, r_i the transverse distance between the shower center and the i th crystal, w_i the weight assigned to i th crystal and r_0 the distance between two crystals [28]. The weights are defined in basf2 [22].

4. Polar angle of the ECL cluster `clusterTheta`,
5. Time of the ECL cluster `clusterTime` (only used for beam background suppression),
6. Second moment of the cluster `clusterSecondMoment`, defined as in [28]:

$$S = \frac{\sum_{i=0}^n w_i E_i r_i^2}{\sum w_i E_i} \quad (4.6)$$

7. MVA output accumulating the cluster shape information of eleven Zernike moments `clusterZernikeMVA` [29],
8. MVA output distinguishing hadronic and electromagnetic showers by their pulse shape properties `clusterPulseShapeDiscriminationMVA`

For the inclusive reconstruction of the B_{tag}^0 -meson, photons are selected by using those classifier outputs.

Qualitatively, one can start by looking at general B^0 -meson photons, where run-independent generic MC simulation of roughly 189000 $B^0\bar{B}^0$ including beam background is used.

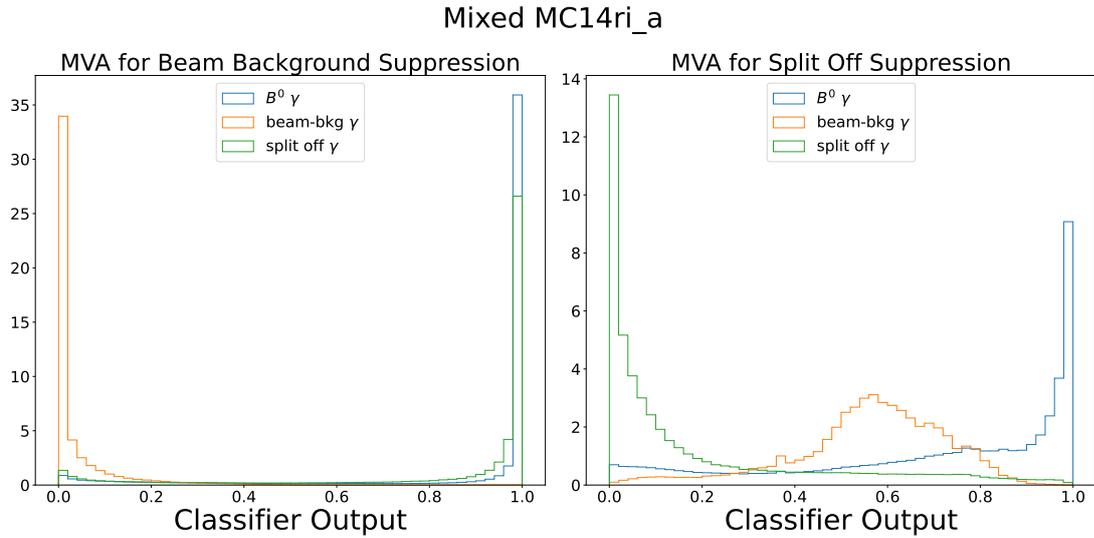


Figure 4.10: Distribution of the classifier outputs of the beam background suppression MVA (left) and the split off MVA (right) compared between B^0 -meson photons, beam background and split off background. Note that the same classification criteria are used as in [27].

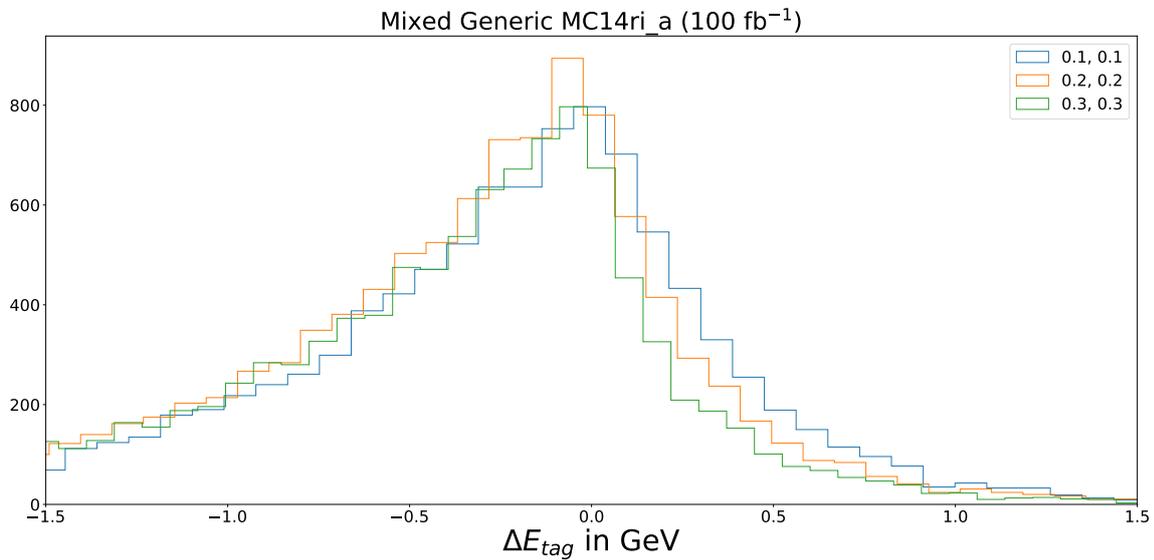


Figure 4.11: Distribution of ΔE_{tag} with three different photon selections. Only events with a rightly reconstructed signal-side are shown.

Note that the MVAs were originally trained on a different generation of MC simulations, so that in this analysis no training data is used when evaluating the MVA distributions.

In Fig. 4.10, based on the different distribution of the classifier outputs for signal and background photons, one can clearly see the separation power of those MVAs.

While selecting tracks with MC variables, which should guarantee no background tracks are selected, the photon selection is done with different cuts on the two photon MVAs. In Fig. 4.11 the distribution of the energy difference ΔE of the B_{tag}^0 -meson are shown for three different exemplary MVAs cuts. In total three different cut values for each MVA output are considered and by comparing their ΔE_{tag} resolutions, the optimal cut values of the photon classifier outputs are chosen to be above 0.1 for both.

Track Selection

In the regarded MC sample reconstructed tracks can either be: 1. charged final state particles coming from the B -meson, 2. so called *clone tracks*, which are low momentum particles curling inside the detector and are hence reconstructed as multiple particles, or 3. a byproduct from beam induced background. In contrast to photons, the background components are very small when compared to the signal components. The exact composition should be sample dependent, but within this particular sample of 189000 generic $B^0\bar{B}^0$ events, $(92.92 \pm 0.07)\%$ are signal tracks, whereas $(3.615 \pm 0.013)\%$ are beam background tracks and $(3.463 \pm 0.013)\%$ clone tracks.

Currently, there exists no track classifier similar to the photon MVAs, designed to use for Belle II analyses. As building such MVAs lies outside the scope of this thesis, only possible input variables are suggested. The transverse momentum p_t as well as the projection of the momentum on the z -axis p_z or the polar angle θ , all of which are shown in Fig. 4.12, show great separation power between the different components. A potential difficulty could arise though, from the fact that signal and background components are represented with different density in the MC simulation.

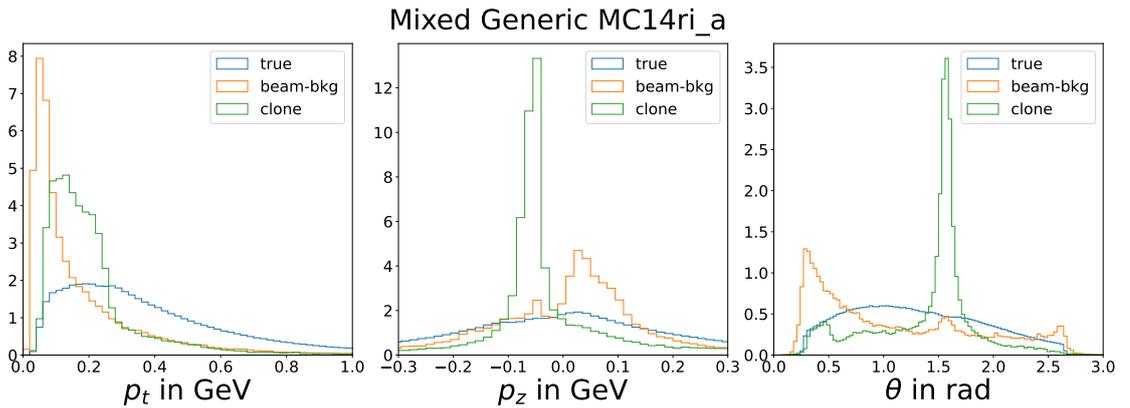


Figure 4.12: Normalized distribution of the transversal momentum p_t , the z -projection of the momentum p_z and the polar angle θ of all charged particles in 189000 simulated $B^0\bar{B}^0$ events. The different components are true tracks, beam background tracks or clone tracks.

Moving on to the concrete track selection used in this analysis, all tracks are to lie in the

angular acceptance region of the CDC, namely $17^\circ < \theta < 150^\circ$, where they are to leave at least one hit.

Additionally their energy is bounded from above to $E < 5.5$ GeV.

Restricting the track's distance to the IP region is not as straight forward as in the signal selection, because now the tracks could be the product of all kinds of different particles. Among others, they could originate from a $V0$ particle, a photon, which when interacting with matter produces a electron-positron pair, or a neutral particle like K_S^0 or Λ^0 , which due to their lifetime can travel some distance away from the IP before potentially decaying into two charged particles.

To be as inclusive as possible while maintaining a good background track rejection, the track selection is updated during reconstruction.

Loose IP cuts, $dr < 10$ cm and $|dz| < 20$ cm, are used for the start, from which $V0$ particles can be reconstructed with the following cuts implemented in [22]:

1. The absolute value of the difference of the reconstructed mass to the nominal mass of the respective $V0$ particle must be below $|dM| < 0.1$ GeV
2. The PID P_{id} value of the respective charged daughter particles P need to be above $P_{id} > 0.2$
3. The charged daughter particles need to be in CDC acceptance, hence the polar angle of their momentum must be between $17^\circ < \theta < 150^\circ$
4. The vertex of the $V0$ particles are fitted and candidates with failed fits are discarded

The photon is reconstructed in $\gamma \rightarrow e^+e^-$, whereas the remaining $V0$ particle are reconstructed in $K_S^0 \rightarrow \pi^+\pi^-$ and $\Lambda^0 \rightarrow p^+\pi^-$.

After reconstruction, they are kept instead of the two tracks originating from them in

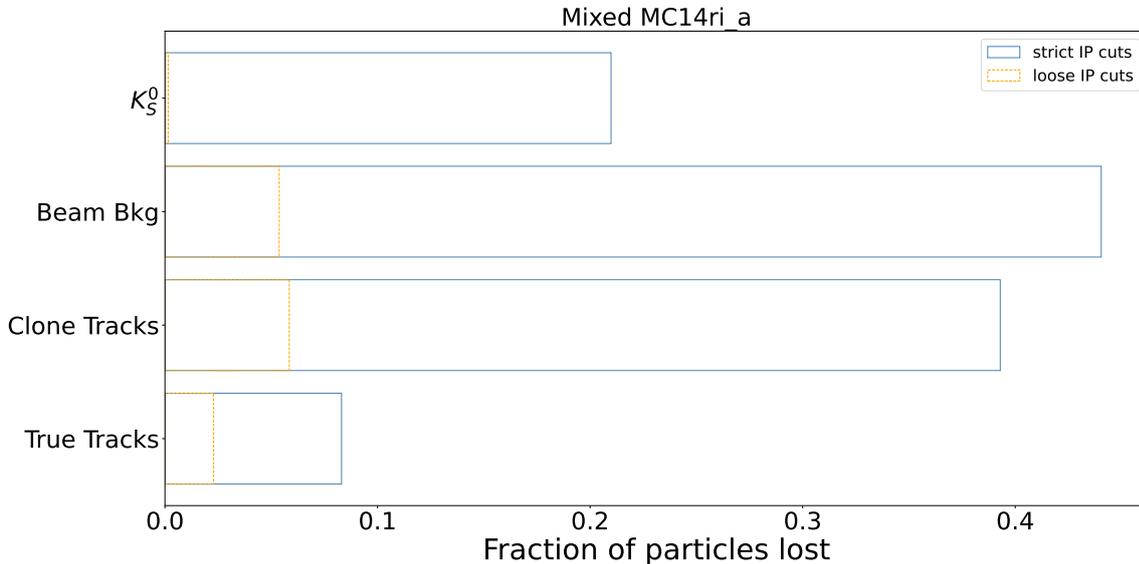


Figure 4.13: Fraction of particles lost when applying loose IP cuts in contrast to tight IP cuts. True tracks refers to B^0 -tracks not originating from a $V0$ particle.

the ROE, so that the track selection can now be tightened to $dr < 2$ cm and $|dz| < 4$ cm

to optimally reject background. Fig. 4.13 shows the fraction of tracks which would be lost with those tight or loose IP-restrictions, for different track sources.

The $K_S^0 \rightarrow \pi^+\pi^-$ decay is used as an example to show the effect of tight IP restrictions to events containing V0 particles. In those events, the kinematical information carried by the charged pion pair would be lost.

Fig. 4.14 shows a comparison of the updated, loose and tight track selection in ΔE_{tag} for

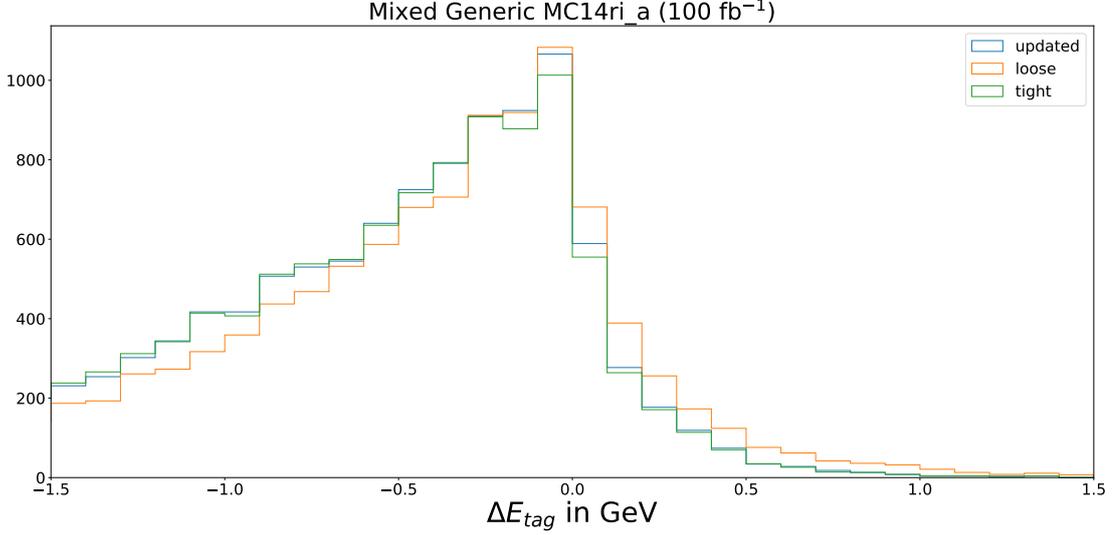


Figure 4.14: Distribution of ΔE_{tag} with three different track selections. Only events with a rightly reconstructed signal-side are shown.

generic MC corresponding to 100 fb^{-1} for rightly reconstructed $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ events. Further studies are needed to improve and optimize the track selection.

It should be also noted that, as there are multiple particles which could correspond to the detected track. Hence, distinguishing signal tracks from background tracks is not the only task when optimizing the track selection. Additional information of which particle is present to use the right mass hypothesis is also necessary.

In this analysis, the tracks are assigned to the particle P corresponding to the highest P_{id} value defined in Sec. 4.1. This can be problematic in cases where two or more PID values are almost equally high or almost equally low, as the PID variables are of finite precision as already seen in the lepton veto description.

When looking at MC samples of 10000 events each, where the signal-side decays to $B_{sig}^0 \rightarrow D^{*-}\ell^+\nu_\ell$ and the tag-side also decays to a single exclusive decay, the performance of the track hypothesis can be estimated.

For this purpose, two different exclusive decay modes are considered, $B_{tag}^0 \rightarrow D^-\pi^+$ and $B_{tag}^0 \rightarrow D^-\pi^+\pi^0$ (the exact decay modes of the daughter particles can be found in the appendix).

In Fig. 4.15 the distribution of the number of charged final state particles reconstructed as charged pions $N_{\pi,tag}$ as well as the total number of all charged particles $N_{c,tag}$ in the ROE are shown. A correctly reconstructed signal-side is required. As both tag-side decays are fixed to one particular decay mode, in case of a perfect track selection, $N_{\pi,B_{tag}}$

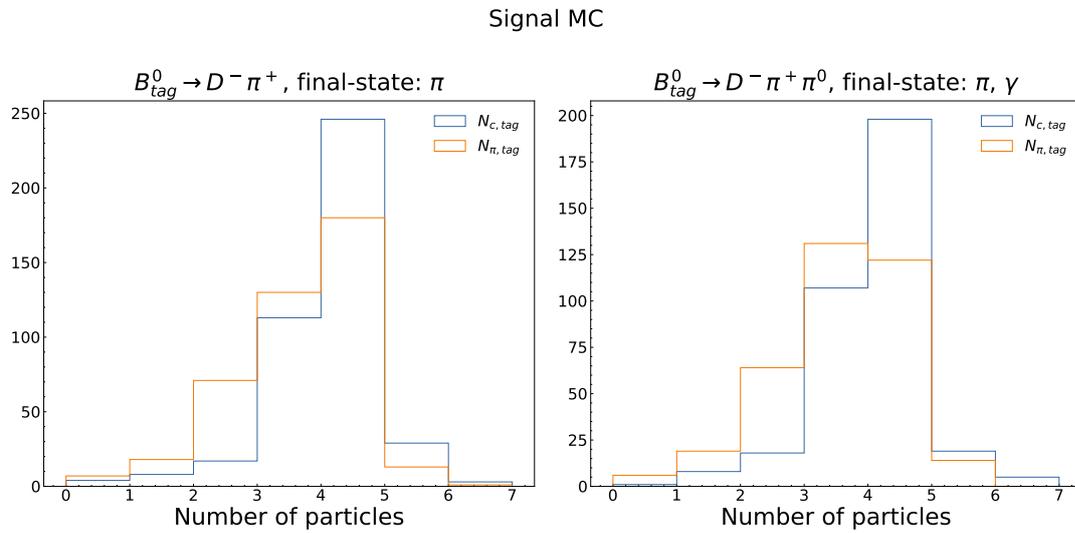


Figure 4.15: Number of charged particles and charged particles reconstructed as pions in the ROE for two different exclusive B_{tag}^0 -decays.

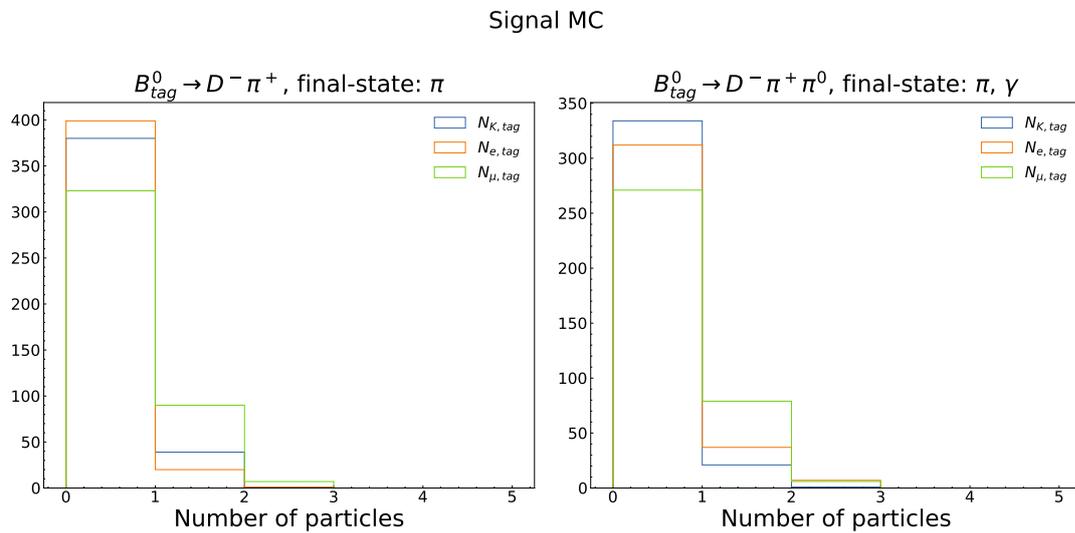


Figure 4.16: Number of charged particles reconstructed as electrons, muons or kaons in the ROE for two different exclusive B_{tag}^0 -decays.

should yield the same number for each candidate, as well as $N_{c,tag}$. Additionally, both decay modes are chosen to only have pions as charged final state particles, so that ideally $N_{\pi,tag} = N_{c,tag}$. However, a discrepancy can be observed indicating that an incorrect mass hypothesis is assigned to the wrongly reconstructed pion, which is confirmed also in Fig. 4.16 showing the number of other charged final state particles. Here, again a discrepancy from the actual value ($= 0$) for some candidates is visible. To be more concrete, in $(42.5 \pm 1.5)\%$ ($(44.9 \pm 1.5)\%$) of the times for events from the $B_{tag}^0 \rightarrow D^- \pi^+$ ($B_{tag}^0 \rightarrow D^- \pi^+ \pi^0$) -sample, the pion is not reconstructed correctly and hence a wrong mass hypothesis is used. A further study on assigning possible PID priors to the track selection is therefore advised.

B-selection

As mentioned B^0 -meson properties are enforced on the inclusive reconstructed tag-side in the form of tight $M_{bc,tag}$ and ΔE_{tag} cuts. Since no MC information of the correctness of the reconstruction is available, consistency must be checked otherwise.

Firstly, one can look at the missing mass squared M_{miss}^2 distribution of the event. The signal-side is reconstructed to $B^0 \rightarrow D^{*-} \ell \nu_\ell$, so that only one neutrino should be present in the event. This would correspond to a M_{miss}^2 peak at zero. Fig. 4.17 confirms this expectation. It also provides a comparison of the M_{miss}^2 distribution for $B^0 \rightarrow D^{*-} \ell \nu_\ell$

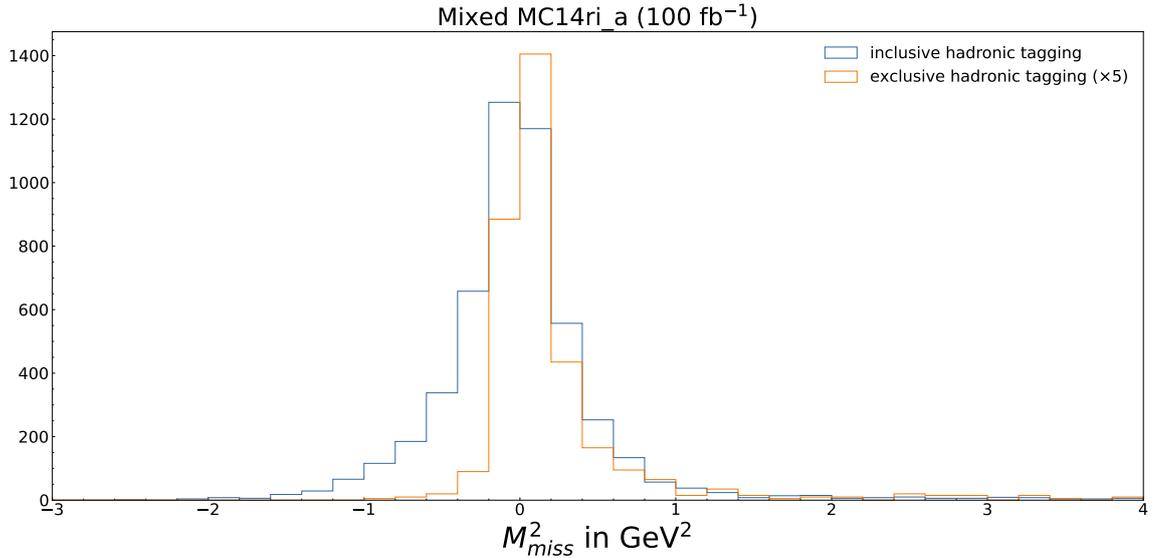


Figure 4.17: Distribution of M_{miss}^2 when using an inclusive hadronic tag or an exclusive hadronic tag (FEI) on 100 fb⁻¹ MC. Note that for comparison the FEI distribution is enlarged by a factor of 5. Only events with a correctly reconstructed signal-side are shown.

reconstruction done with an inclusive and an exclusive hadronic tag. Hence, using an inclusive tag for this analysis and performing a fit in M_{miss}^2 , proves to be consistent. Secondly, one can also compare the direction of the resulting B_{tag}^0 -meson with the MC direction of the rest of the event. Ideally, both of those values should be equal. The

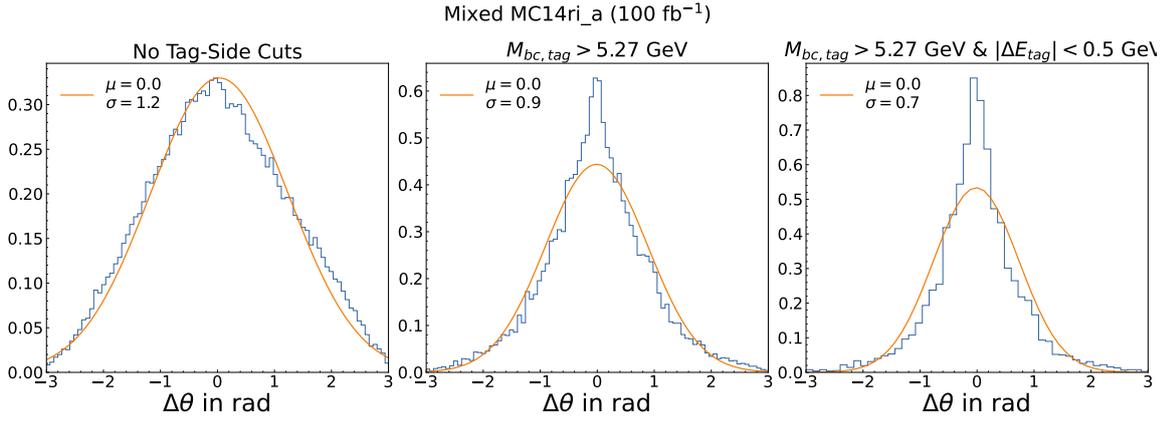


Figure 4.18: Normalized distribution of the difference between the reconstructed three-momentum's polar angle and the MC polar angle $\Delta\theta$ with no cuts on the tag-side, cut on $M_{bc,tag}$ and cut on both $M_{bc,tag}$ and ΔE_{tag} . All distributions are fitted to Gaussian curves, where the fitted mean μ as well as the corresponding standard deviation σ are given in the legend.

resolution of the difference is computed in Fig. 4.18, where the difference of the reconstructed and MC polar angle of the ROE $\Delta\theta$ are shown with a Gaussian fit. With each tag-side cut the resolution (the Gaussian standard deviation) improves, also proving consistency.

At last, it should be mentioned that this analysis does not use the absolute magnitude of the reconstructed tag-side energy or momentum, but rather infers them from the beam energy and PDG information. Consequently, only the direction of the tag-side's three-momentum is of great importance when measuring e.g. the missing mass squared M_{miss}^2 . Hence, a good resolution of $\Delta\theta$ is essential.

Again the vertex of the B_{tag}^0 -meson is fitted and candidates with failed fits are discarded, but similarly to the signal-side, it provides no additional separation power.

4.3 Event Selection

After selecting the optimal B_{sig}^0 - and B_{tag}^0 -meson the $\Upsilon(4S)$ -resonance is reconstructed and all kinematical information about the regarded event including the missing four momentum carried by the undetected neutrinos are accessible.

To only have one $\Upsilon(4S)$ candidate per event, only the candidate with the lowest deviation between the invariant mass of the reconstructed D^{*+} -meson and its nominal mass is kept.

Considering $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ decays, the event can be constrained further, when considering the four momentum carried by the W -boson q^2 responsible for the process. As four momentum is conserved, q^2 is restricted to only a limited kinematical region. To be more specific using Eq. 4.7, where $\cos(\theta_{\tau,\nu})$ refers to the angle between the τ lepton and its accompanying neutrino, and the q^2 distribution visible in Fig. 4.19, a lower bound is

set to $q^2 > 3.5 \text{ GeV}^2$.

$$\begin{aligned}
 q^2 &= (p_B - p_{D^*})^2 \\
 &= (p_\tau + p_\nu)^2 \\
 &= m_\tau^2 + 2(E_\tau - |\vec{p}_\tau| \cos(\theta_{\tau,\nu}))|\vec{p}_\nu| > m_\tau^2
 \end{aligned} \tag{4.7}$$

This cut results in the optimal signal purity, when the normalization mode is excluded.

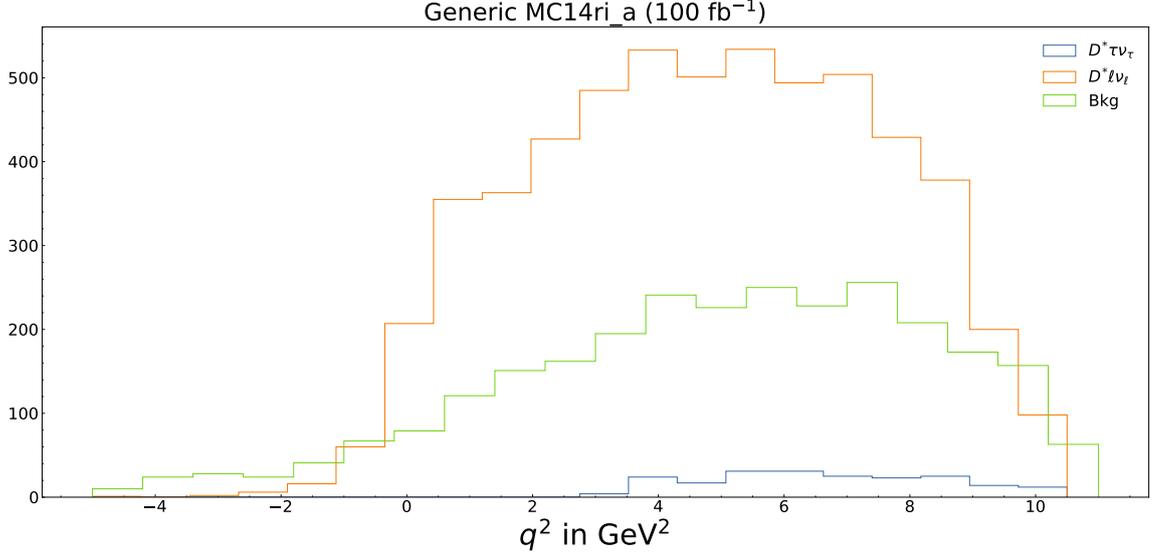


Figure 4.19: Distribution of the squared four momentum carried by the W-Boson q^2 separated in signal, normalization channel and all background candidates (continuum and combinatorial).

Given that for the normalization channel this lower bound does not hold, the cut also reduces the number of normalization events by $(34.1 \pm 0.8)\%$. However, estimating the minimal statistical uncertainty of the number of normalization mode events after the q^2 -cut by $\sqrt{n_{norm}}$, a relative uncertainty of $\frac{\delta n_{norm}}{n_{norm}} = 1.6\%$ is obtained, which is much smaller than the expected uncertainty on the signal mode. As 1.6% is still reasonably low, the q^2 -cut is performed regardless of the loss in normalization mode events. Moreover, the cut also suppresses $(73 \pm 4)\%$ of the continuum background.

4.4 Reconstruction Efficiency

As the tag-side efficiency cannot be determined with MC information, instead of using Eq. 4.1 the tag-side efficiency is calculated by Eq. 4.8.

$$\epsilon_{tag} = \frac{\epsilon_{tot}}{\epsilon_{sig}} = \frac{n_{sig}}{n_{sig,all}} \tag{4.8}$$

Here, $n_{sig,all}$ corresponds to the number of signal events when no tag-side cuts are applied and n_{sig} to the number of signal events after applied tag-side cuts. Furthermore, independent values for the signal- and tag-side efficiencies are assumed and no cut

on q^2 is applied yet. In general, the tag-side efficiency should be dependent on the exact decay mode present on the tag-side. To get an understanding of the performance of the reconstruction for different tag-side decays, again signal MC samples of 10000 $B_{sig}^0 \rightarrow D^{*-} \ell^+ \nu_\ell$ events are considered with a fixed B_{tag}^0 -decay (the whole decay trees can be found in the appendix). The resulting reconstruction efficiencies from those samples are listed in Tab. 4.2. Clearly, the lepton veto in addition with the tight cuts on

B_{tag}^0 -Decay Modes	$\epsilon_{tag}[\%]$	$\epsilon_{sig}[\%]$	$\epsilon_{tot}[\%]$
$D^- K^+$	12.6 ± 0.9	18.6 ± 0.4	2.34 ± 0.16
$D^- D^+ K^{*0}$	8.3 ± 0.7	17.1 ± 0.4	1.43 ± 0.12
$D^- \pi^+$	12.7 ± 0.9	18.7 ± 0.4	2.38 ± 0.15
$D^- \pi^+ \pi^0$	10.4 ± 0.8	18.9 ± 0.4	1.97 ± 0.14
$D^- e^+ \nu_e$	0	18.8 ± 0.4	0
$D^- \mu^+ \nu_\mu$	0	18.3 ± 0.4	0

Table 4.2: Reconstruction efficiencies of the tag- and signal-side as well as the total one for different decays on the tag-side. Other than the two semileptonic B_{tag}^0 -decays at the bottom, the remaining decay modes are purely hadronic.

ΔE_{tag} and $M_{bc,tag}$ discard the undesired semileptonic B_{tag}^0 -decays completely. Besides, relatively high tag-side efficiencies are obtained for $B_{tag}^0 \rightarrow D^- K^+$ and $B_{tag}^0 \rightarrow D^- \pi^+$, where $D^- \rightarrow K_S^0 \pi^-$. Hence, the update of the track selection in the ROE, indeed spares some K_S^0 -mesons from discard. Note that additional studies on those MC samples, when not including beam background on generator level, showed that the tag-side efficiency shrinks by about 50% on average due to beam background.

Returning to generic run-independent MC corresponding to 100 fb^{-1} and reconstructing $B^0 \rightarrow D^{*-} \tau \nu_\tau$, while performing the described selection processes, the final signal-, tag-side and total efficiencies obtained are $\epsilon_{sig} = (20.5 \pm 0.4)\%$, $\epsilon_{tag} = (8.6 \pm 0.6)\%$ and $\epsilon_{total} = (1.78 \pm 0.12)\%$. Hence, the reconstruction efficiency of an inclusively reconstructed hadronic tag-side is bigger than the one of an exclusively reconstructed hadronic tag-side listed in Tab. 1.1 by a factor of almost 20.

After the q^2 -cut the total reconstruction efficiency reduces slightly to $\epsilon_{total} = (1.75 \pm 0.12)\%$.

5 Background Analysis

To achieve a figure of merit competitive with analyses using an exclusive tag, the high background levels introduced when using an inclusive tag need to be suppressed.

The different background components are analyzed and listed in the following section. To classify the different background components, MC information is used via *decay hash*, a tool implemented in basf2 [22]. It can list the entire decay tree associated to an event either on MC or on reconstruction level.

After classifying the background, suppression tools are discussed in Sec. 5.2.

Unfortunately, good suppression performances could not be achieved throughout this analysis and further studies are needed. Possible explanations are given in Sec. 5.3.

5.1 Background Components

In this section generic run-independent MC is used corresponding to an integrated luminosity of 200 fb^{-1} including continuum background. One can subdivide the reconstruction modes present, after applying the selection cuts discussed in the previous chapter, into the following categories sorted by their respective occurrences:

1. Normalization mode ($B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$) ($63.9 \pm 0.7\%$)
2. Combinatorial background ($30.1 \pm 0.5\%$)
3. Signal mode ($B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$) ($3.42 \pm 0.17\%$)
4. Continuum background ($2.56 \pm 0.15\%$)

Taking the low amount of continuum background events into account, the background suppression focuses on the more dominant background source: the so called *combinatorial* background. It consists of reconstructed B -events, where instead of the desired signal or normalization mode, another B -decay is present. In approximately half of those background events, the B -meson is charged.

Decay modes included in the combinatorial background survive all the signal- and tag-side cuts applied, indicating that their kinematical distributions are similar to the signal and/or normalization mode.

To be more concrete, the following decay modes are the dominant components of the combinatorial background:

$B \rightarrow D^{**} \ell \nu_\ell$ This is the most dominant background source (in total $(9.08 \pm 0.28)\%$ of all reconstructed candidates) referred to as D^{**} -background. The D^{**} -meson represents any D state of higher excitation, so with orbital angular momentum of $L = 1$. In the regarded sample D^{**} -mesons include D_1, D'_1, D_2^* and D_0^* -mesons.

If allowed, the D^{**} -meson decays mainly into a D^{*-} -meson and a pion, so that in

most cases the D^{*-} -meson as well as the light lepton are reconstructed correctly. The charge of the D^{**} -meson and with that also that of its daughters depends on the B -meson charge. Besides the additional pion, which is not reconstructed on the signal-side, this background component is most likely to share the same final state particles as the signal and normalization mode. Consequently, most kinematical distributions are very similar.

The main decay modes present in the reconstruction are $D_1^0 \rightarrow D^{*-}\pi^+$ ($D_1^- \rightarrow D^{*-}\pi^0$) and $D_1^{\prime 0} \rightarrow D^{*-}\pi^+$ ($D_1^{\prime -} \rightarrow D^{*-}\pi^0$), where the corresponding D^{**} -meson originate from $B^+ \rightarrow \bar{D}^{*0}\ell^+\nu_\ell$ ($B^0 \rightarrow D^{*-}\ell^+\nu_\ell$). Less frequent decay modes are $D_2^{*0} \rightarrow D^{*-}\pi^+$ ($D_2^{*-} \rightarrow D^{*-}\pi^0$) and $D_0^{*0} \rightarrow D^-\pi^+$ ($D_0^{*-} \rightarrow D^-\pi^0$).

$B \rightarrow D^{(*)}D^{(*)}K^{(*)}$ The second most dominant background source (in total $(3.26 \pm 0.17)\%$ of all reconstructed candidates) are events in which two D -mesons and a kaon of different excitation states are present. Generally, they can originate from either a charged or neutral B -decay.

In most cases one of the D -mesons is the correctly reconstructed D^{*-} -meson, whereas the other decays semileptonically into a lighter hadron accompanied by a lepton-neutrino pair. The additional hadron and the kaon are not reconstructed on the signal-side.

$B \rightarrow D_s^{(*)}D^{(*)}$ The third most dominant background source (in total $(2.73 \pm 0.15)\%$ of all reconstructed candidates) are mesons consisting of a charm and a strange quark, namely D_s -mesons of different excitation states. To be more specific within the regarded reconstruction D_s^+ , D_s^{*+} , D_{s0}^+ and D_{s1}^+ -mesons are present. As such mesons have to be charged because of their quark components, they mainly come from $B^0 \rightarrow D_s^+D^{*-}$, where the D^{*-} -meson is reconstructed correctly.

Among other things, the D_s -mesons can decay purely leptonically, mainly to $D_s \rightarrow \tau^+\nu_\tau$ with a branching fraction of 5.48% in the regraded MC sample [22], so that at the end the final state particles are exactly the same as for the signal mode. Again, similar kinematical distributions are expected.

Rest Thus far only roughly half of the combinatorial background is accounted for. The remaining half include different B -decays, where no individual component occurs often enough (above 1%) to be explicitly mentioned.

Clearly, the D^{**} -background is the most problematic one and significantly higher than any other background source. Therefore, D^{**} -events are reconstructed explicitly by adding a pion to the signal reconstruction, namely modifying the B -decay to $B \rightarrow D^*\pi\ell\nu_\ell$ for charged and neutral B -mesons. With that channel specific variables can be accessed, like the missing mass squared $M_{miss,\pi}^2$ when having an additional pion on the signal-side. The distribution should peak at zero for correctly reconstructed D^{**} -events and at negative values for normalization mode events. This is confirmed in Fig. 5.1, which shows the $M_{miss,\pi}^2$ distribution. The reconstruction of the D^{**} -meson is done in parallel to the signal reconstruction, so that additional variables can be treated as event variables and accessed for each event individual.

On another note, one can conclude that in most cases no fake D^* -mesons and/or fake light leptons are present, meaning that final state particles are mostly shared between different reconstruction modes. To be more concrete the light lepton is reconstructed

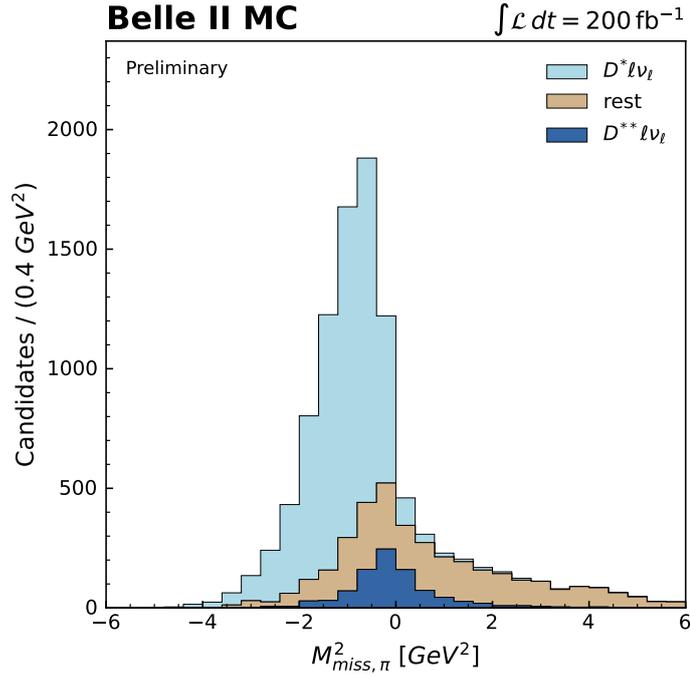


Figure 5.1: Stacked distribution of $M_{miss,\pi}^2$, the missing mass squared distribution when having an additional pion on the signal-side, for three different components, normalization mode, D^{**} -background and remaining decay modes including signal as well as the remaining background modes.

correctly in $(86.0 \pm 0.9)\%$ of the cases, and the D^{*-} -meson in $(88.3 \pm 0.9)\%$ of the cases.

As the signal mode contains three neutrinos, signal events are expected to have a missing mass squared M_{miss}^2 distribution in higher regions, in contrast to the normalization channel peaking at zero, shown in Fig. 5.2. Also visible in Fig. 5.2 is the high contamination of background in higher M_{miss}^2 region as all background modes, in which not all final state particles are reconstructed, contribute here. Hence, the missing mass squared M_{miss}^2 distribution does not allow to distinguish between all the different reconstruction modes, which introduces higher uncertainties to the fit. Therefore, suppression tools are needed to effectively discard background modes, which mimic the signal or normalization mode.

However, a good separation between signal and normalization mode in the M_{miss}^2 distribution allows us to define a *sideband purity*, which is the signal purity for the $M_{miss}^2 > 0.85 \text{ GeV}^2$ region and has a value of $(11.8 \pm 0.6)\%$. For $M_{miss}^2 < 0.85 \text{ GeV}^2$ the normalization channel purity (*peak purity*) is regarded, which is at $(83.8 \pm 1.0)\%$. As the latter is already at a high value, suppression tools are optimized to increase the sideband purity. When evaluating the suppression performance though, both the peak as well as sideband purity are considered.

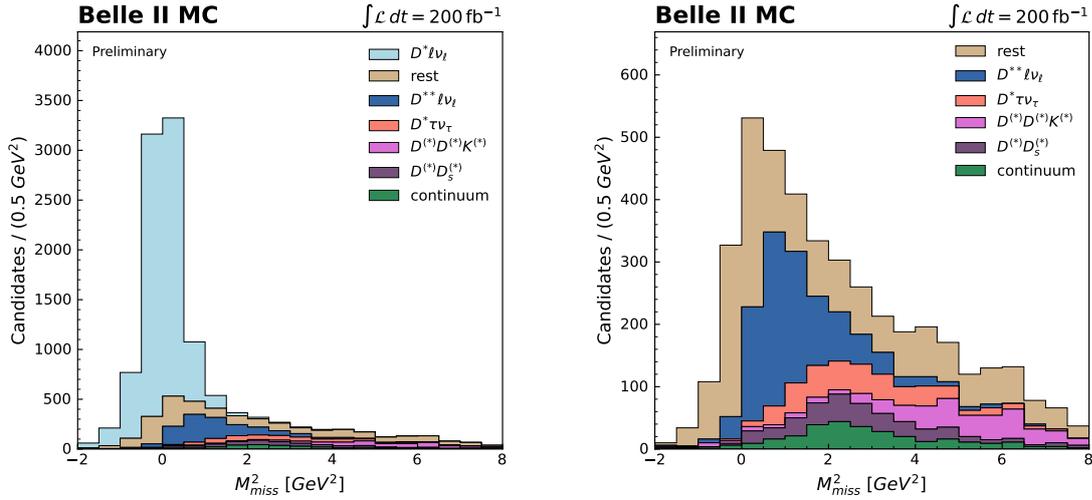


Figure 5.2: Stacked distribution of M_{miss}^2 for seven different decay modes on the left. To better distinguish between signal and background modes the dominant normalization mode is excluded on the right.

5.2 Background Suppression

As the kinematical distribution of signal and background are too similar to naively apply cuts as done in the selection processes, suppression tools used need to be more complex and consider different variables at the same time.

This analysis uses *Fast Boosted Decision Trees* (Fast-BDT), a machine learning algorithm, which is able to perform binary classification between signal and background events. It consists of n_{trees} decision trees, which are boosted by a predefined *shrinkage* value ρ , to prevent overtraining. The depth of each tree n_{depth} as well as the number of regions each input variable is separated into n_{cuts} can also be adapted. A detailed description of the algorithm can be found in [30].

As input variables, variables where different distributions of signal and background are expected, are used.

Two training (90%) and test (10%) samples are created, containing all correctly reconstructed signal events from run-independent signal MC and either all background events or only D^{**} -events from generic MC corresponding to 700 fb^{-1} . To keep the peak purity high, the $M_{miss}^2 < 0.85 \text{ GeV}^2$ region is excluded.

Different sorts of input variables are considered: D^* -kinematics, vertex information of the B - and D -mesons, tag-side kinematics and multiplicities, variables related to the D^{**} -reconstruction, event shape variables, event kinematics etc.

The most discriminating variables found so far when trying to suppress all background modes are plotted in Fig. 5.3 and listed below:

1. The remaining energy in the ECL after reconstructing the signal-side $E_{extra,sig}$. No selection cuts on particles contributing to $E_{extra,sig}$ are applied, which seem to increase the discrimination power. In contrast to background events, with unreconstructed particles, signal and normalization mode events are expected to

have a $E_{extra,sig}$ distribution at smaller values.

2. The energy difference of the B_{sig}^0 -meson ΔE_{sig} as defined in Eq. 4.4.
3. The total visible energy detected in the event in the center of mass frame E_{vis} . It is computed as the sum of the absolute three-momenta of all detected charged tracks and the cluster energy of the ECL not associated to the tracks [31]. As the energy of the three neutrinos coming from signal events is not visible, signal events are expected to lie in lower regions of E_{vis} .
4. The cosine of the angle between the thrust axis of the B_{sig}^0 -meson and the thrust axis of the ROE, here the B_{tag}^0 -meson, $\cos(\theta_{TBTO})$. The thrust axis is defined as the direction in which the projection of all momenta contributing to the respective particle is maximized [4].
5. The number of daughters of the B_{tag}^0 -meson $N_{d,tag}$. Note that because of the tagging approach used the average number of daughters is relatively high, as all final state particles within the ROE are by definition daughters of the B_{tag}^0 -meson. For background events with unreconstructed particles on the signal-side, as those particles are then expected to be matched to the tag-side, $N_{d,tag}$ is expected to lie in higher regions.
6. The number of charged particles in the ROE of the B_{sig}^0 -meson $N_{c,ROE}$ again without applying any selection cuts upon particles added to the ROE as this seem to increase the separation power.

As target variable MC truth-matching information is used. The classifier output then provides a probability value for an event to be signal or background.

However, when looking at Fig. 5.3 it becomes apparent that all those distributions are still too similar in signal and background, so that either the Fast-BDT holds no separation power or it is overtrained, even after fine tuning the hyperparameters to maximize the area underneath the ROC-curve. To compute the ROC-curve the classification of an event provided by different cuts on the classifier output is compared to its actual class, which is known from MC information. The distribution of the corresponding background rejection in dependence of the signal preservation is then the *receiver operating characteristic* (ROC) curve.

For demonstration purposes the described input variables are used to train a Fast-BDT to classify signal (1) and all background (0) events. The resulting distribution of the classifier output on the training sample and on the test sample are shown in Fig. 5.4, which clearly holds little separation power. Although signal and background distribution indicate that the Fast-BDT is overtrained, the effect is still more pronounced for the signal distribution. Hence, to improve the classifier performance the training sample should be enlarged and especially more signal events should be added. However, all available signal MC simulations of the specific MC generation considered are already in use. To estimate the classifier performance the ROC-curve is shown in Fig. 5.5 naming also the corresponding area under the curve, which was maximized for the test sample by the selected hyperparameters. The value is still relatively small though as the classifier performance is relatively bad due to the little separation power of the input variables, the limited size of the train sample and regardless of the exact hyperparameters. Note

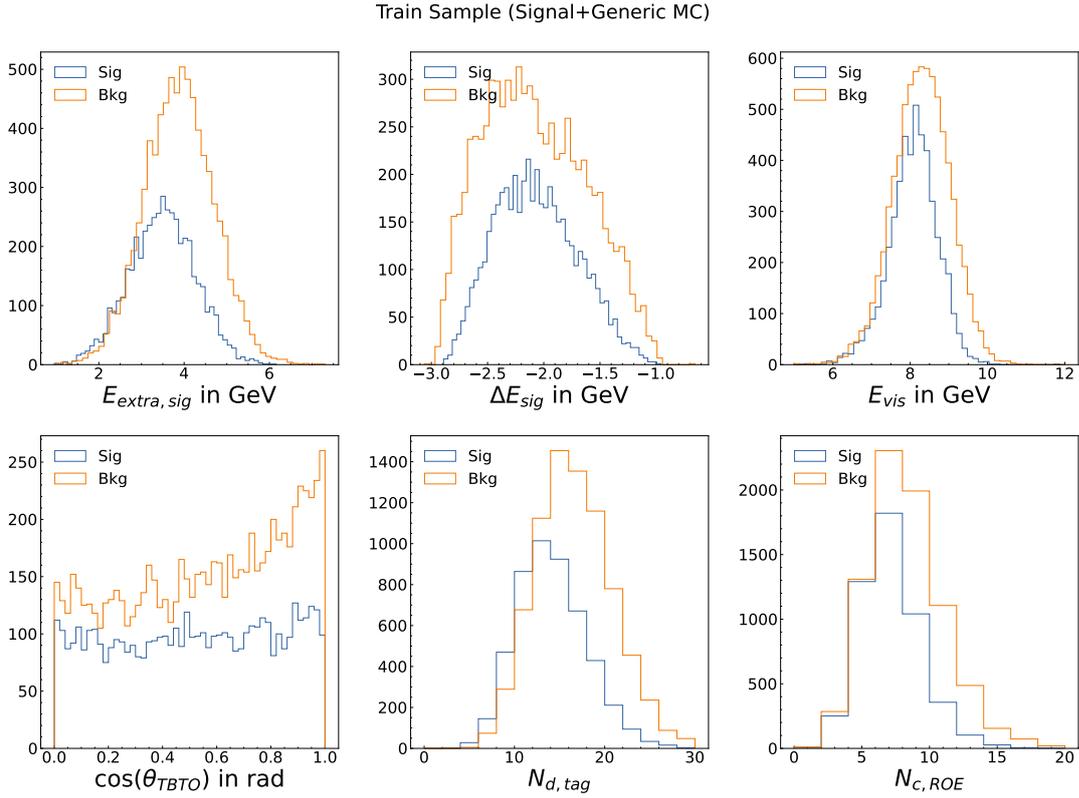


Figure 5.3: Distribution of potential input variables with most separation power found from customized training MC for signal and all background events.

the discrepancy between the classifier performance on train and test sample, which is a result of the overtraining of the Fast-BDT. Additionally, for the ROC-curve shown here, $1 - \text{background rejection}$ is considered.

Some of the input variables are correlated to each other and/or correlated to the missing mass squared M_{miss}^2 and the light lepton's three-momentum $|\vec{p}_{\ell,CMS}|$. If they are used to effectively suppress background events, the effect of the suppression on the fit should be evaluated. In fact, applying the Fast-BDT classification to generic run-independent MC of 200 fb^{-1} , shows another great difficulty connected to the background suppression, which is the binary character of the classifier output. This causes normalization mode events to also be suppressed. Fig 5.6 shows the classifier output for signal, normalization mode and background events. The Fast-BDT clearly tends to classify normalization mode events as background. A cut on the classifier output would therefore not change anything significantly about the combined purity of signal and normalization mode, so that no such cut is applied.

5.3 Discussion of Applicability

Strictly speaking no inclusive measurement of $R(D^*)$ has been done before, where besides $\mathcal{B}(B \rightarrow D^* \tau \nu_\tau)$ also $\mathcal{B}(B \rightarrow D^* \ell \nu_\ell)$ was measured simultaneously. Therefore, the

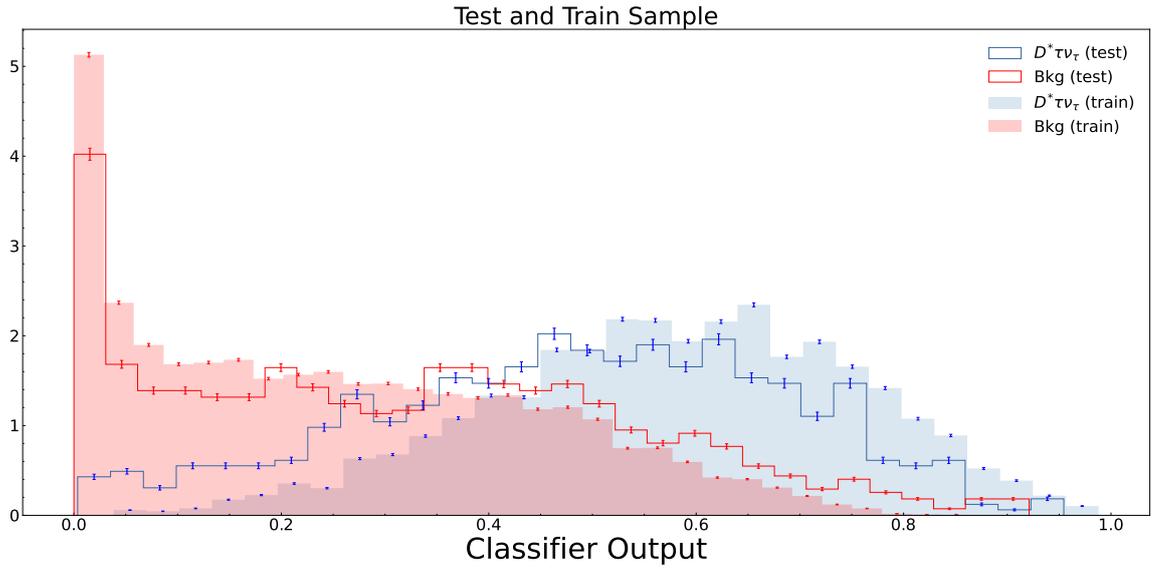


Figure 5.4: Normalized distribution of the classifier output of a Fast-BDT applied to the train and test sample and built with the following hyperparameters: $n_{trees} = 1200$, $\rho = 0.033$, $n_{depth} = 4$, $n_{cuts} = 4$. Correctly reconstructed signal events and all background events are considered.

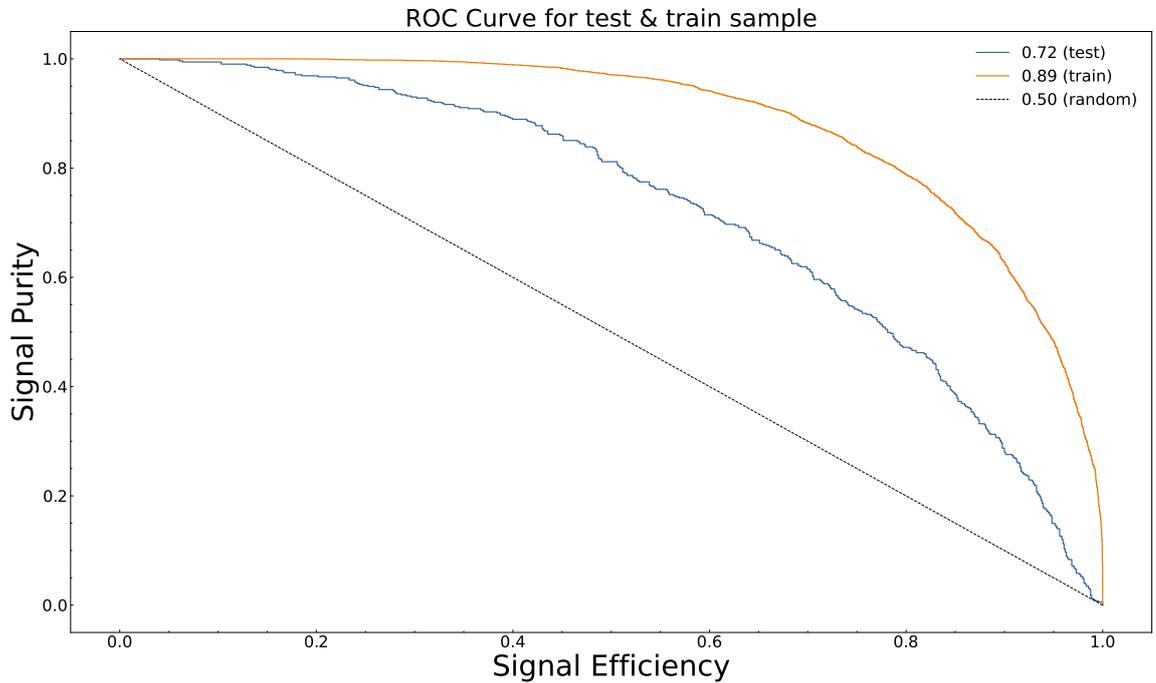


Figure 5.5: ROC-curve of the classification performance of the trained Fast-BDT for train and test sample. The dashed line corresponds to the ROC-curve of a random classifier, i.e. classifying signal and background randomly.

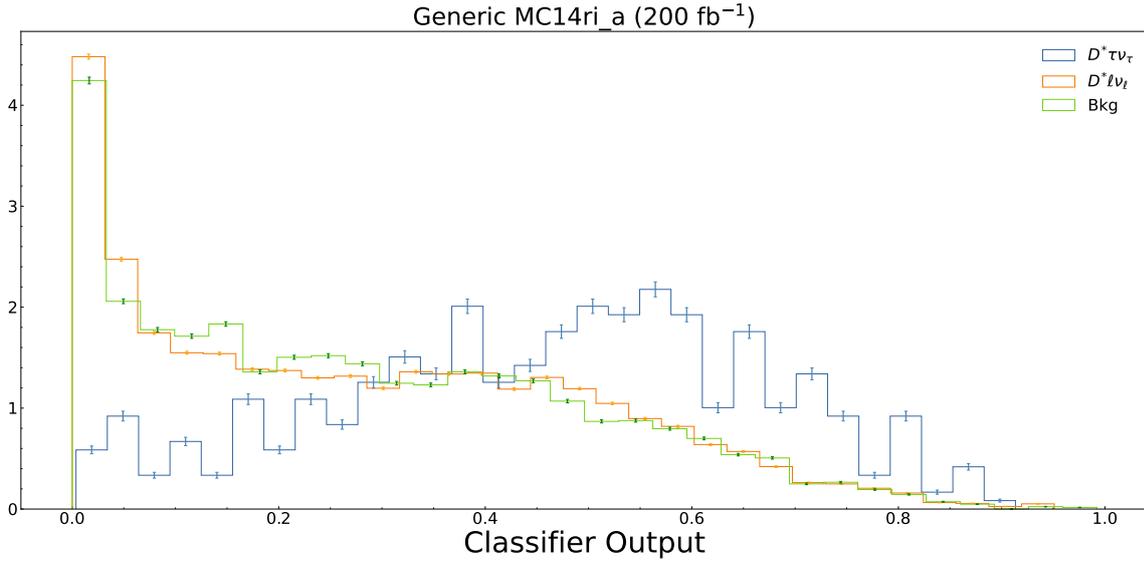


Figure 5.6: Normalized distribution of the classifier output on generic run-independent MC corresponding to 200 fb^{-1} for signal, normalization mode and all background events.

only conclusions regarding the background suppression are drawn from this analysis directly. However, other analyses using either a different tagging approach or measuring a different quantity performed well on background suppression using conventional cutting or machine learning tools. Possible reasons for the difference of performance are discussed in this section.

Firstly, the influence of the tagging approach used is regarded. In 2015, the Belle collaboration measured $R(D^{(*)})$ on the full Belle data corresponding to 711 fb^{-1} using an hadronic tag iterating over 1149 exclusive B -decay modes [25]. Additionally, there exists a BaBar analysis from 2013, where a semi-inclusive tagging approach is used to measure $R(D^{(*)})$. Here, the tag-side is separated into a charged state X , containing a combination of charged and neutral pions and kaons, and a so called *seed meson*, which can be either a $D^{(*)}$, a D_s -meson (including also higher excitation states) or a J/ψ resonance [26].

Whereas, the Belle analysis uses the classifier output of their neural network to fit and extract the number of signal and normalization events, in the BaBar analysis a cut on the BDT classifier output increases the signal purity significantly.

Nonetheless, a big difference between an $R(D^{(*)})$ analysis using an inclusive tag arises from the simple fact that the input variable, described as the most powerful one in both the references, namely E_{extra} , is not accessible when using an inclusively reconstructed tag-side. E_{extra} is defined as the energy in the ECL, which is not associated to the signal- nor the tag-side after applying a photon selection on the ROE. In both the references, photons whose energy contributed to the computation of E_{extra} had to be above a threshold of $E_\gamma > 50 \text{ MeV}$ to exclude beam background.

For background processes, in which the tag-side is reconstructed exclusively and correctly, higher values of E_{extra} stem from not reconstructed particles on the signal-side. Hence, a different distribution of E_{extra} for signal and background is expected.

For example, when considering the D^{**} -background mode in which an additional neutral pion is produced, which will most likely decay into two photons, the energy of those photons should now contribute to E_{extra} . Meanwhile, for signal events with correctly reconstructed tag-sides no additional energy E_{extra} should be present other than beam background or split off photons surviving the initial photon selection of the ROE.

Fig. 5.7 shows the discrimination power of E_{extra} when reconstructing $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$

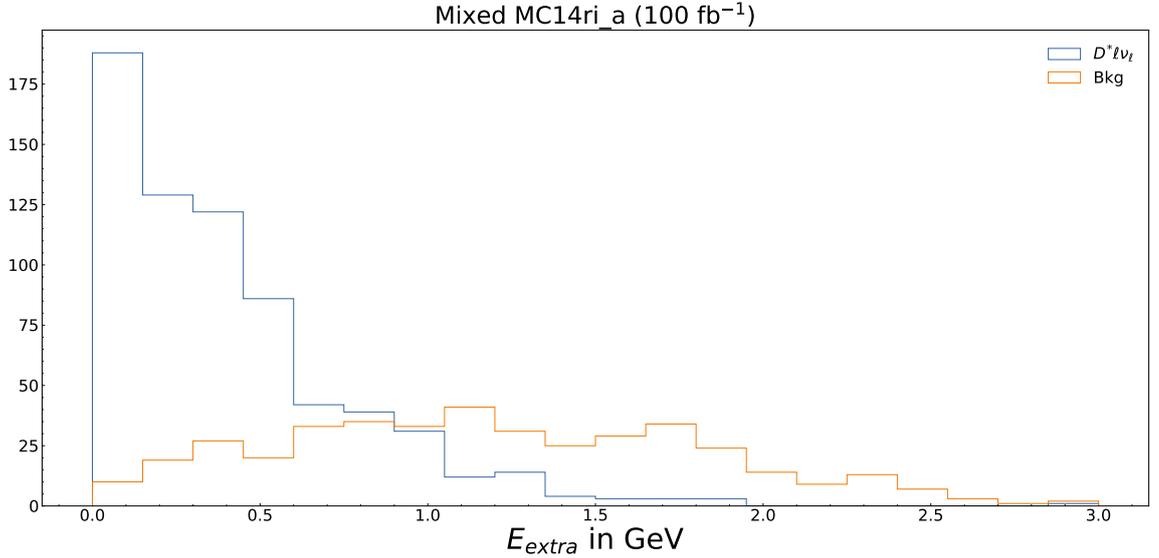


Figure 5.7: Distribution of E_{extra} when using an exclusive hadronic tag (FEI) to reconstruct $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$ on generic $B^0 \bar{B}^0$ -MC. Separation is between correctly reconstructed signal events and background.

with an exclusively reconstructed tag-side (using FEI). Here, the same photon selection for the ROE is used as for the tag-side reconstruction in Sec. 4.2.

However, when performing an inclusive reconstruction of the tag-side, one can only access the energy deposited in the ECL clusters that is not associated to the signal-side and by definition that extra energy, fulfilling certain photon selection criteria (or not), is matched to the tag-side. Hence, the resulting distribution of $E_{extra,sig}$ (with no selection cuts) shown in Fig. 5.3 holds comparably very little separation power.

The extra energy E_{extra} when defined as in the previous analyses would by definition be zero for each reconstruction mode. The only additional photons of the event are the ones not surviving the selection cuts and should therefore ideally correspond only to beam background or split off photons. Consequently, E_{extra} cannot be used as input variable.

Secondly, a problem arises from the fact, that there are two decay modes of interest, the signal and normalization mode. As their kinematical distribution differ, a very wide range of kinematically allowed processes are present. In general this means that a classifier can be build to successfully suppress background process from a $B^0 \rightarrow D^{*-} \tau \nu_\tau$ reconstruction. However, cutting on those classifiers also reduces the number of normalization events, so that no decrease of the background rate is achieved, similarly to what was discussed in the context of the exemplary built Fast-BDT. In the end, no binary

classification between signal and background is of interest, but a trinary classification between signal, normalization and background.

To still use BDTs, variables that are similar for signal and normalization mode but different for background are needed. During this analysis no variable fulfilling this criteria while still maintaining a high discrimination power could be found.

When taking into consideration the Belle analyses from 2007 [10] and 2010 [11], where the branching fractions of $B \rightarrow D^{(*)}\tau\nu_\tau$ are measured using an inclusive tagging approach, it becomes apparent that the signal purity can be of a high value without even using any machine learning tools to suppress background, but by simply applying certain selection cuts.

The most powerful cut according to the analysis is a cut at $X_{miss} > 2$ GeV [10] (for leptonic τ -decays) defined in Eq. 5.1.

$$X_{miss} = \frac{E_{miss} - |\vec{p}_{D^*,CMS} + \vec{p}_{\ell,CMS}|}{\sqrt{E_{beam}^2 - M_B^2}} \quad (5.1)$$

Applying such a cut to an inclusive $R(D^*)$ analysis might increase the sideband purity slightly by approximately 10%, but it also causes a loss of around 100% of normalization mode events. Therefore, no such cut should be applied.

X_{miss} , whose distribution is shown in Fig. 5.8, has also been used as input variable for the background suppression, where it still causes the suppression of too many normalization mode events, and as an alternative variable to be fitted in, however it has no

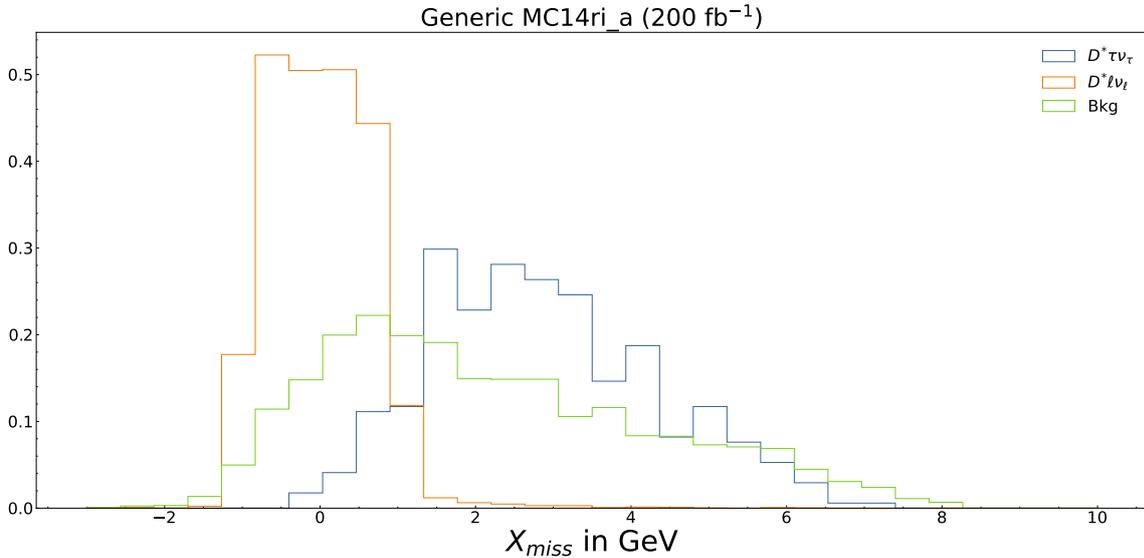


Figure 5.8: Normalized distribution of X_{miss} separated in correctly reconstructed signal, normalization mode events and background.

better fit performance than M_{miss}^2 .

Thirdly, an inclusive analysis at Belle II suffers from more beam background compared to exclusive analyses as well as to analyses done at Belle or BaBar, resulting in a lower resolution.

Regarding possible alternatives, a different approach would be to already use a Fast-BDT in the signal- and tag-side selection and with that drop tight selection cuts. Although this might include undesirable events (lying outside the expected signal region), it would at least ensure that more signal events can be used for training.

Although Fast-BDT have proven to be robust and consistent [30], one could also consider using different machine learning algorithms to perform the classification task at hand.

Another alternative would be to not cut, but fit in the classifier output instead of in the lepton's three-momentum $|\vec{p}_{\ell, CMS}|$, similar to what was done in the Belle analysis from 2015 mentioned [25]. However, one has to be careful of which exact MVA and which fit procedure is preferable in this case: As the classifier output can be of binary character, trying to fit three or more fit component to it can be tricky. To circumvent this problem within the Belle analysis the fit was separated into two different M_{miss}^2 regions corresponding to either the sideband or the peaking region of the normalization mode and the smooth classifier output of a non-binary neural network was used [25].

6 Fit

A two dimensional extended maximum likelihood fit in missing mass squared M_{miss}^2 and the absolute value of the lepton's three-momentum in the center of mass frame $|\vec{p}_{\ell,CMS}|$ is used to extract the number of signal n_{sig} and normalization mode events n_{norm} , which is described in the following section. $R(D^*)$ is then further calculated in Sec. 6.2 according to Eq. 6.1, where in addition of the fit results also the efficiency ratio of both channels enter. Since not all possible τ -decays are considered, the signal efficiency is extrapolated from the leptonic τ -decays.

$$R(D^*) = \frac{n_{sig} \epsilon_{norm}}{n_{norm} \epsilon_{sig}} \quad (6.1)$$

At last the fit is validated in Sec. 6.3.

6.1 Fit Procedure

To describe the fit procedure [32] is used as reference.

The fit is performed for four different components:

1. Signal Mode $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$
2. Normalization Mode $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$
3. D^{**} -Mode $B \rightarrow D^{**} \ell \nu_\ell$
4. Remaining Background (combinatorial as well as continuum)

Fig. 6.1 shows the two dimensional distribution of M_{miss}^2 and $|\vec{p}_{\ell,CMS}|$ for each fit component. Distributions of different variables were tried out for the fit, namely replacing M_{miss}^2 by X_{miss} and/or $|\vec{p}_{\ell,CMS}|$ by $|\vec{p}_{\ell,CMS} + \vec{p}_{D^*,CMS}|$, but they resulted in worse fit performances.

For the fit procedure, instead of artificially dividing the distribution of both variables into discrete bins, Gaussian kernels are used on every observed measurement to estimate the probability density function (PDF) for each component. The resulting PDF is smooth and more robust when having only few data points. This kernel density estimation (KDE) is done on generic run-independent MC corresponding to 700 fb^{-1} (referred to as PDF-sample).

In a next step the PDFs of all components are summed to the total PDF $\mathcal{P}_t(\mathbf{x}, \mathbf{n})$ as in Eq. 6.2.

$$\mathcal{P}_t(\mathbf{x}, \boldsymbol{\theta}) = \theta_{sig} \mathcal{P}_{sig}(\mathbf{x}) + \theta_{norm} \mathcal{P}_{norm}(\mathbf{x}) + \theta_{D^{**}} \mathcal{P}_{D^{**}}(\mathbf{x}) + \theta_{bkg} \mathcal{P}_{bkg}(\mathbf{x}) \quad (6.2)$$

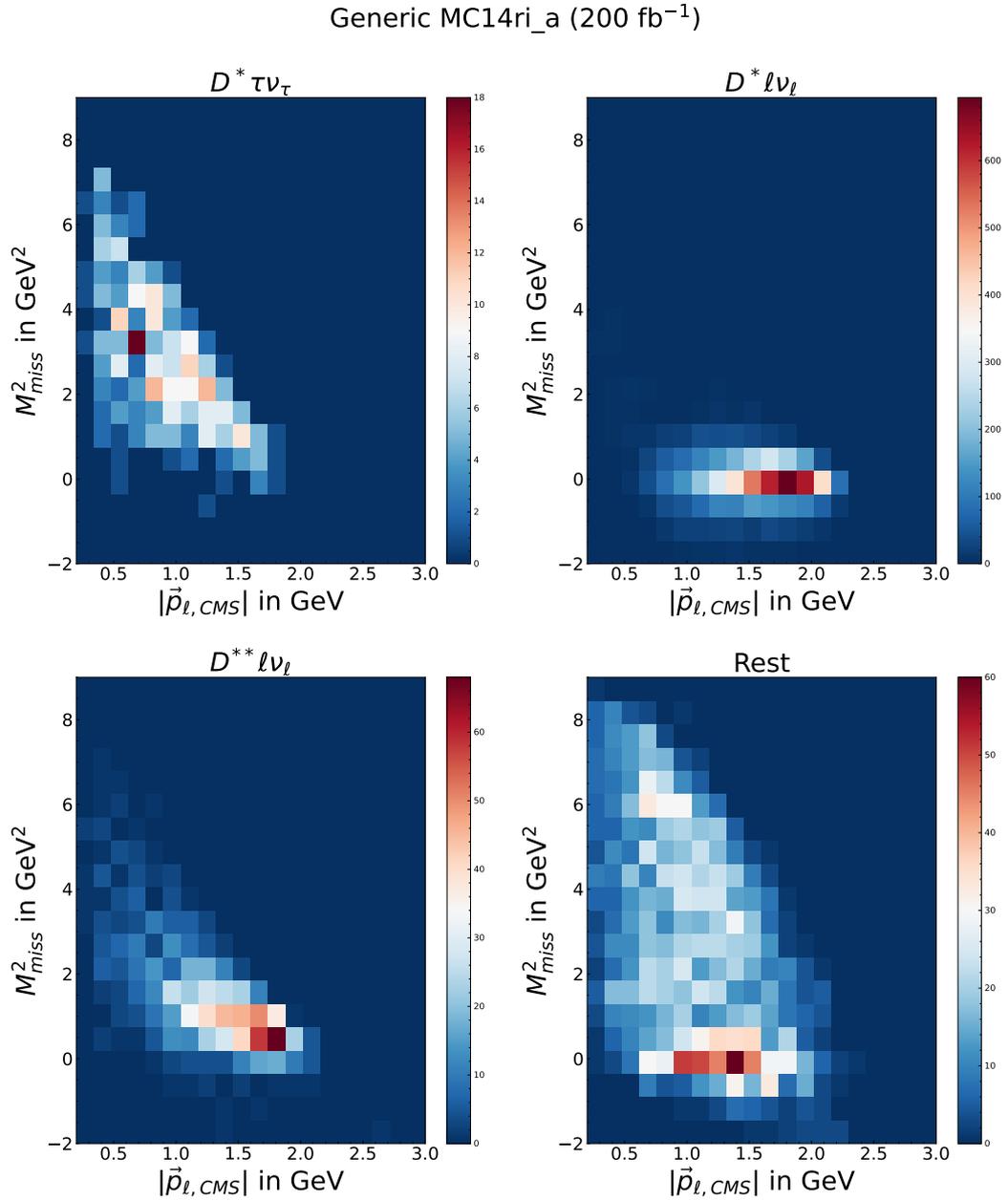


Figure 6.1: Two dimensional distribution of M^2_{miss} and $|\vec{p}_{\ell,CMS}|$ for each fit component from generic MC corresponding to 200 fb⁻¹.

PDFs corresponding to a certain fit component are fixed according to the M_{miss}^2 and $|\vec{p}_{\ell,CMS}|$ -distributions of the fitted sample, here summarized to the two dimensional random variable $\mathbf{x} = (M_{miss}^2, |\vec{p}_{\ell,CMS}|)$.

The unknown parameters are then $\boldsymbol{\theta} = (\theta_{sig}, \theta_{norm}, \theta_{D^{**}}, \theta_{bkg})$. However, for $\mathcal{P}_t(\mathbf{x}, \boldsymbol{\theta})$ to still fulfill PDF-requirements, it can be deduced that the sum of all fit variables must sum to unity, $1 = \sum_i \theta_i$.

As the number of observed events n is approximated to be a Poisson distributed random variable with mean N_t , the corresponding Poisson term is multiplied to the likelihood of the model. This then results to the *extended likelihood* of the modeled total PDF $\mathcal{P}_t(\mathbf{x}, \boldsymbol{\theta})$ defined in Eq. 6.3 as well as the corresponding log-likelihood in Eq. 6.4. The coefficient i runs over all observed n data points in the fitted sample and j over the four different fit components.

$$L(N_t, \boldsymbol{\theta}) = \frac{N_t^n e^{-N_t}}{n!} \prod_i \left(\sum_j \theta_j \mathcal{P}(x_i, \theta_j) \right) \quad (6.3)$$

$$\log L(N_t, \boldsymbol{\theta}) = -N_t + \sum_i \log \left(\sum_j N_t \theta_j \mathcal{P}_j(\mathbf{x}_i) \right) \quad (6.4)$$

$$= - \sum_j n_j + \sum_i \log \left(\sum_j n_j \mathcal{P}_j(\mathbf{x}_i) \right) \quad (6.5)$$

Defining $n_j = N_t \theta_j$ as the number of events of type j , leads to Eq. 6.5. The condition set to the sum of all fit parameters θ_j now translates to $N_t = \sum_j n_j$.

For a run-independent MC sample corresponding to 200 fb^{-1} (fit-sample), the set of $\hat{\mathbf{n}} = (\hat{n}_{sig}, \hat{n}_{norm}, \hat{n}_{D^{**}}, \hat{n}_{bkg})$, which minimizes the negative extended log-likelihood function $\min_{\hat{\mathbf{n}}} (-\log(L(\mathbf{x}, \mathbf{n})))$, is then extracted to calculate $R(D^*)$.

To constrain the D^{**} -component of the fit, the number of D^{**} -events can be determined by doing a sequential fit: Firstly, a one dimensional fit in the missing mass squared distribution when adding an additional pion to the signal-side $M_{miss,\pi}^2$ is done. The fitted number of D^{**} -events as well as their error boundaries factorized by the D^{**} -reconstruction efficiency are then fed to the two dimensional fit.

However, this procedure increases the statistical uncertainty of the D^{**} -component, which is why within this analysis this approach is discarded.

Optimizing the D^{**} -reconstruction and with that the $M_{miss,\pi}^2$ (visible in Fig. 5.1) resolution may make the sequential fit preferable.

RooFit [33] is used as analysis software for the fit.

6.2 Fit Results

The actual and the fitted M_{miss}^2 and $|\vec{p}_{\ell,CMS}|$ projected distribution are shown in Fig. 6.2, with the resulting pull distribution at the bottom. The pull is defined in Eq. 6.6, where μ is the number of events present in the regarded bin of the respective variable, $\hat{\mu}$ its

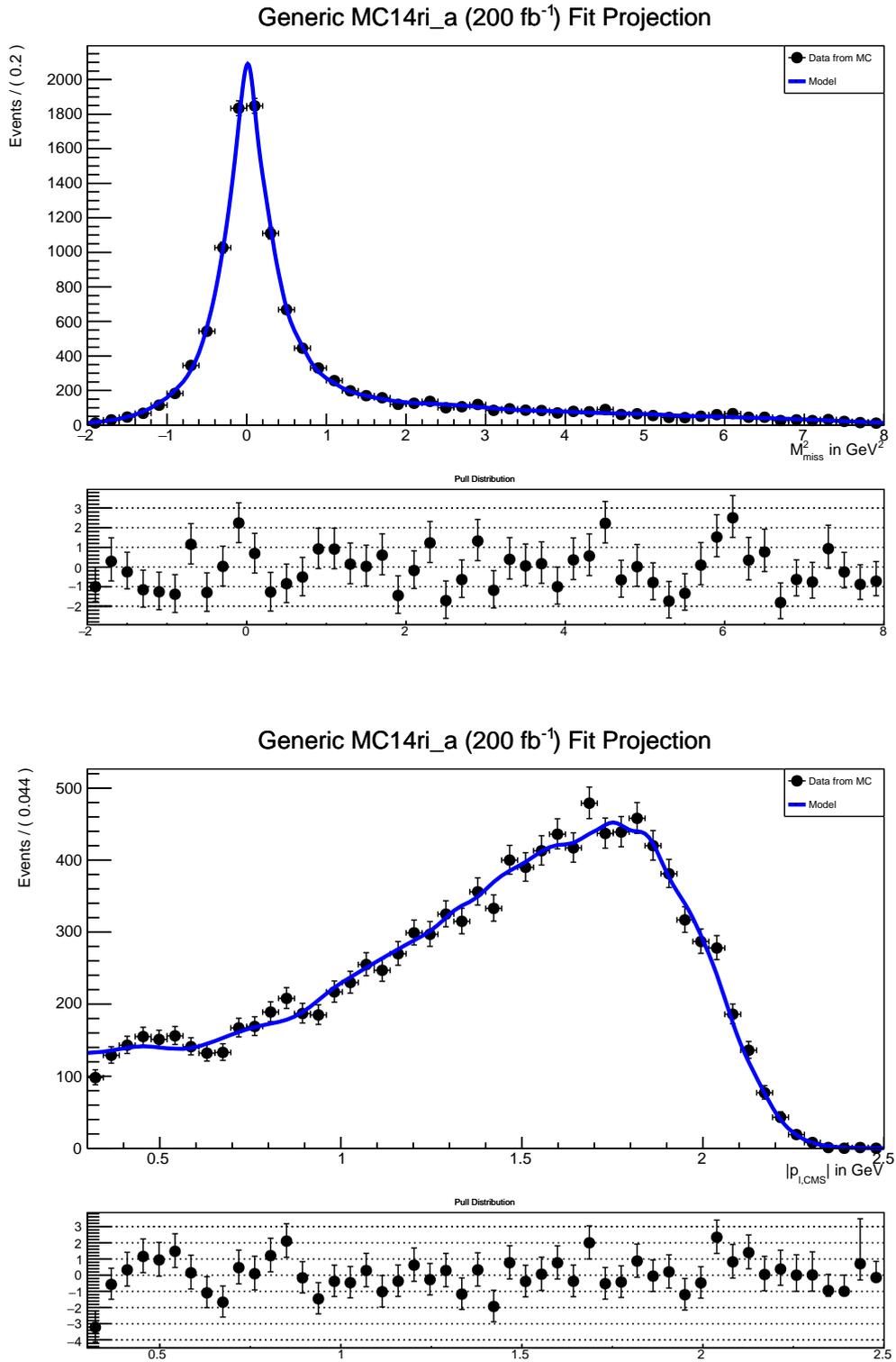


Figure 6.2: Projection of the fit model in M_{miss}^2 and $|p_{\ell,CMS}|$ applied on generic run-independent MC corresponding to 200 fb⁻¹.

estimation from the fitted model, and $\hat{\sigma}$ the statistical uncertainty of the bin content.

$$h_{pull} = \frac{\hat{\mu} - \mu}{\hat{\sigma}} \quad (6.6)$$

The fit results are summarized in Tab. 6.1 compared to the actual values within the sample. Using Eq. 6.1 and the fitted number of signal and normalization events, $R(D^*) = 0.26 \pm 0.04$ is obtained. The statistical uncertainty is calculated with the fit

Fit Component	Expected Number of Events	Fit Result
ℓ-mode		
n_{sig}	394	382 ± 123
n_{norm}	7357	7313 ± 110
$n_{D^{**}}$	1045	1212 ± 89
n_{bkg}	2716	2604 ± 158
e-mode		
n_{sig}	186	134 ± 73
n_{norm}	3553	3510 ± 79
$n_{D^{**}}$	481	554 ± 58
n_{bkg}	1056	1078 ± 113
μ-mode		
n_{sig}	208	192 ± 87
n_{norm}	3804	3792 ± 76
$n_{D^{**}}$	564	645 ± 67
n_{bkg}	1660	1607 ± 104

Table 6.1: Fit results and MC values for the four different fit components from a run-independent MC sample corresponding to 200 fb^{-1} . Different lepton reconstruction modes are considered, including either both light leptons (ℓ -mode) or each individual (e -mode or μ -mode). For the purpose of this table the fit results are rounded to the nearest whole number.

errors of n_{sig} and n_{norm} via the Gaussian error propagation. The PDF-sample is used to compute the efficiency ratio. Within the regarded MC sample $R_{MC}(D^*) = 0.258$, being in a great correspondence to the fitted value. From the correlation matrix listed in Tab. 6.2, one can gather that the correlation of -0.827 between the signal component of the fit and the background component of the fit is relatively high.

Additionally, $R_e(D^*)$ ($R_\mu(D^*)$) is then also computed with fitted results of n_{sig} and n_{norm} . This means instead of the averaged branching fraction of $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ for both light leptons, only $B^0 \rightarrow D^{*-}e^+\nu_e$ ($B^0 \rightarrow D^{*-}\mu^+\nu_\mu$) decays contribute to the normalization mode and only $\tau^+ \rightarrow e^+\nu_e\bar{\nu}_\tau$ ($\tau^+ \rightarrow \mu^+\nu_\mu\bar{\nu}_\tau$) decays are considered for the signal mode. Also the PDFs are adapted accordingly.

The computation yields $R_e(D^*) = 0.20 \pm 0.05$ ($R_\mu(D^*) = 0.25 \pm 0.06$).

$R_e(D^*)$ is consistent with the true value $R_{MC}(D^*)$ within 1.2σ and $R_\mu(D^*)$ within 0.13σ . Whereas the deviation of $R_e(D^*)$ from the value present in the MC and also from the fitted value of $R(D^*)$ is slightly greater than 1σ , $R_\mu(D^*)$ has a higher statistical uncertainty. This is due to the much higher amount of background events in the muon channel

	n_{sig}	n_{norm}	$n_{D^{**}}$	n_{bkg}
n_{sig}	1.000	0.417	-0.396	-0.827
n_{norm}	0.417	1.000	-0.337	-0.413
$n_{D^{**}}$	-0.396	-0.337	1.000	0.065
n_{bkg}	-0.827	-0.413	0.065	1.000
e-mode				
n_{sig}	1.000	0.455	-0.232	-0.833
n_{norm}	0.455	1.000	-0.268	-0.465
$n_{D^{**}}$	-0.232	-0.268	1.000	-0.091
n_{bkg}	-0.833	-0.465	-0.091	1.000
μ-mode				
n_{sig}	1.000	0.344	-0.449	-0.779
n_{norm}	0.344	1.000	-0.330	-0.329
$n_{D^{**}}$	-0.449	-0.330	1.000	0.065
n_{bkg}	-0.779	-0.329	0.065	1.000

Table 6.2: Correlation matrices of the fit yields for different lepton reconstruction modes.

listed in Tab. 6.1 alongside the fitted values for each fit component. The entries of the correlation matrices can be found in Tab. 6.2.

6.3 Fit Validation

To check whether the fit result is unbiased and the statistical uncertainty returned describes the statistical uncertainty well, 1000 samples each corresponding to roughly 200 fb^{-1} are randomly generated from the underlying PDF in Eq. 6.2 and fitted using the same procedure as described in Sec. 6.1. The resulting pull distribution of each fit component is then fitted to a Gaussian curve as shown in Fig. 6.3. Ideally the Gaussian curve should have a mean of 0 and a standard deviation of 1.

However, regarding the mean of the individual component of the fit, the signal and normalization mode component seem to be below zero by 3.2σ and 3.9σ respectively. Meanwhile the means of the background and D^{**} -component are greater than zero by 4.1σ and 1.3σ . Note that here σ refers to the uncertainty of the mean estimated with the Gaussian fit in Fig. 6.3. Taking the definition of the pull in Eq. 6.6 into account, the number of signal and normalization mode events tends to be underestimated, while the number of background and D^{**} -events tends to be overestimated. To investigate further if the fit indeed introduces a bias for the corresponding fit components, a higher number of samples should be used.

As for the standard deviation, most are within their 1σ uncertainty boundaries corresponding to one, indicating that the resulting statistical uncertainty is well described. For the signal component however, the statistical uncertainty returned from the fit, seems to be underestimated resulting in a standard deviation greater than one by 7.5σ .

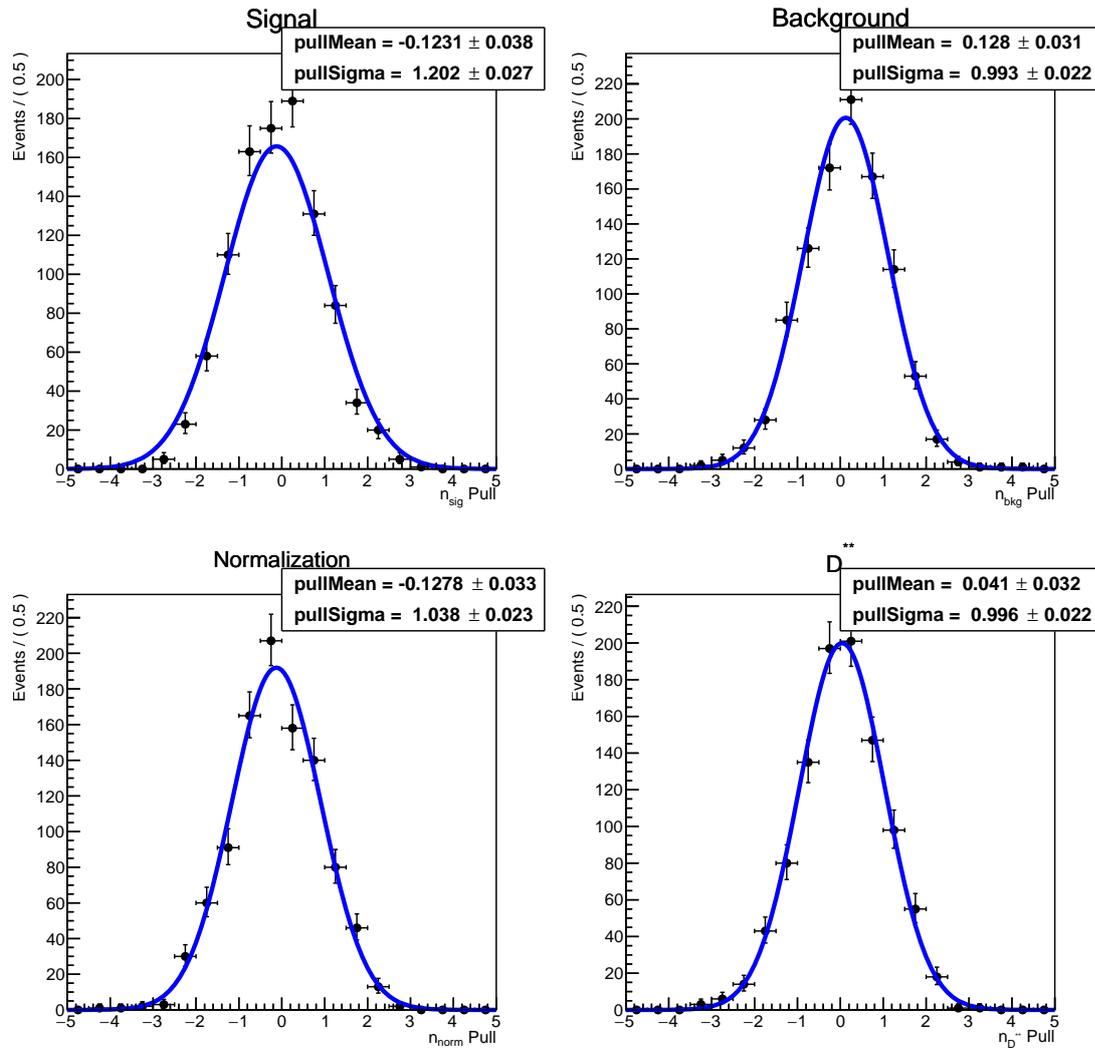


Figure 6.3: Pull distribution of the four fit components computed with 1000 generated and fitted MC samples.

7 Uncertainties

In this chapter systematic uncertainties are discussed, in analogy to [25] and [26], however computing all of them lies outside the scope of this analysis.

In contrast to [10], where the branching fraction of $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ is measured using an inclusive tag, it is not assumed that an additional uncertainty arises from the tag-side reconstruction, since for an $R(D^*)$ measurement the signal branching fraction is normalized with $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$. Therefore, uncertainties concerning the tag-side should

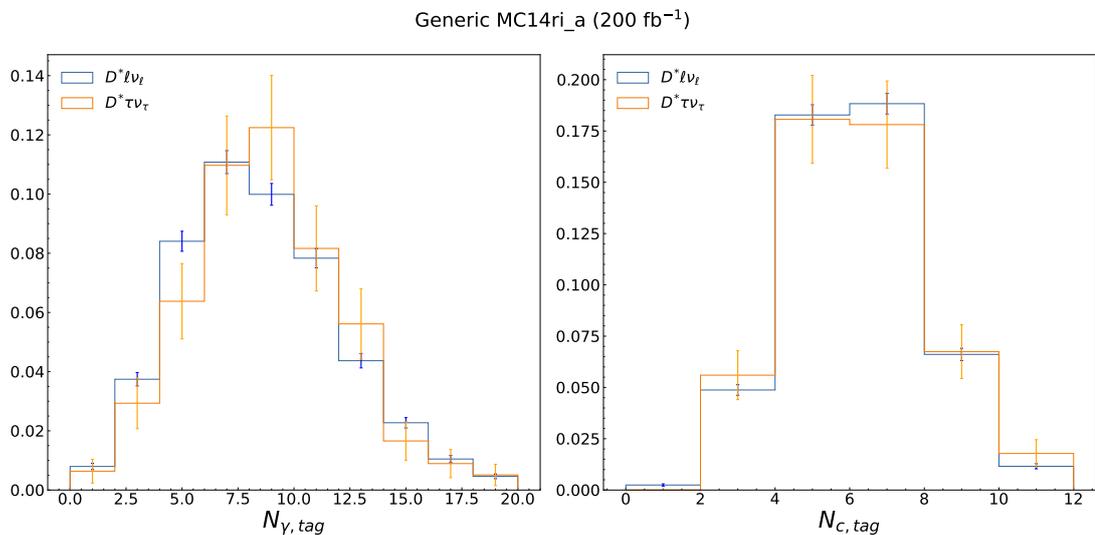


Figure 7.1: Normalized distribution of the number of photons (left) and charged particles (right) coming from the B_{tag}^0 -meson (in the ROE) for signal and normalization mode events.

cancel out. This is qualitatively confirmed when looking at Fig. 7.1, where the tag-side multiplicities for signal and normalization mode are shown to be consistent up to uncertainties.

Additionally, most uncertainties arising from the D^* -reconstruction should cancel out as well, which can be understood when looking at Fig. 7.2, containing the distribution of the absolute three-momentum of the D^* -decay's final state particles for signal and normalization mode. Again, most distributions are, taking uncertainties into account, very similar. Only for the three-momentum of the slow pions is there a significant difference between normalization and signal mode, which thus should be investigated further.

All systematic uncertainties regarded in the following sections as well as the statistical uncertainty from the fit are summarized in Tab. 7.1. The total systematic uncertainty of the fit result is obtained by the individual values corresponding to the different uncer-

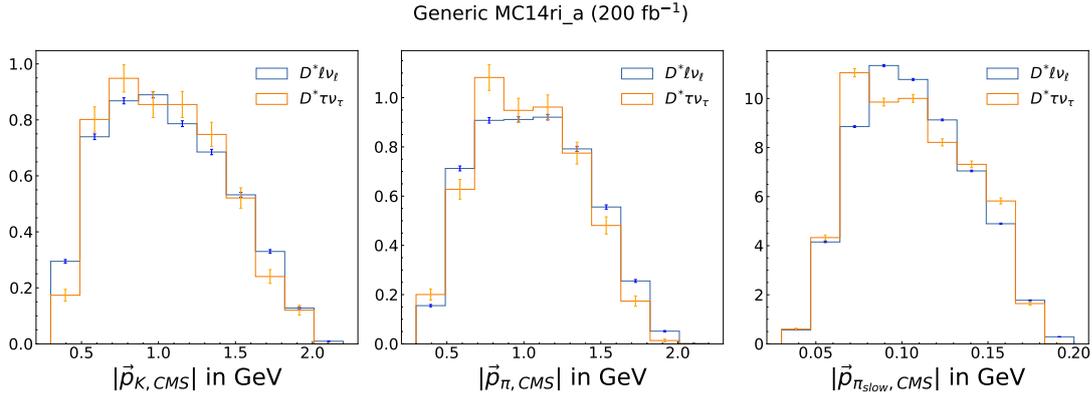


Figure 7.2: Normalized distribution of the absolute three-momentum of all three D^* -final state particles for normalization and signal mode.

Source of Uncertainty	Fractional Uncertainty [%]	Absolute Uncertainty
PDF Estimation	20	0.05
$B \rightarrow D^{**} \ell \nu_\ell$ & Gap Mode	5	0.012
Efficiency	2.7	0.007
PID Corrections	1.0	0.0025
τ -Decay	0.24	0.0006
Total systematic uncertainty (considered)	21	0.06
Total statistical uncertainty	16	0.04

Table 7.1: Absolute and fractional uncertainties with respect to the computed $R(D^*)$ value for different sources of uncertainties.

tainty sources summed in quadrature. This means no correlation between the different systematic uncertainties is assumed.

7.1 PDF Estimation

As described in the Ch. 6, the PDFs of the different fit components are estimated via a KDE. As MC simulations are generated randomly following certain statistical assumptions, different PDFs for different MC samples are expected. The statistical fluctuation of the PDF shapes are smaller the higher the statistics of the PDF-sample are. Although MC corresponding to 700 fb^{-1} are used to extract the PDF shapes of the individual fit components, the number of events is still relatively small due to a relatively small reconstruction efficiency. For the signal component of the fit only 1287 signal events are used to estimate the respective PDF shape \mathcal{P}_{sig} . The statistical fluctuation should therefore be rather large, in contrast to the normalization mode component of the fit, whose PDF \mathcal{P}_{norm} was built with 25321 events.

To account for the resulting systematic uncertainty, the sample from which the different PDFs are generated is re-sampled in each fit component.

The resampling is done with a bootstrap algorithm, where 1000 same sized samples

for each fit component are generated with randomly chosen events from the original sample. Note that the same event can be chosen more than once.

After estimating the PDFs for each of the bootstrap samples with a KDE, the fit is performed on the fit-sample in analogy to Ch. 6. The resulting $R(D^*)$ distributions for each resampled fit component are shown in Fig. 7.3.

Here, the distributions are normalized and fitted to a Gaussian curve. The resulting

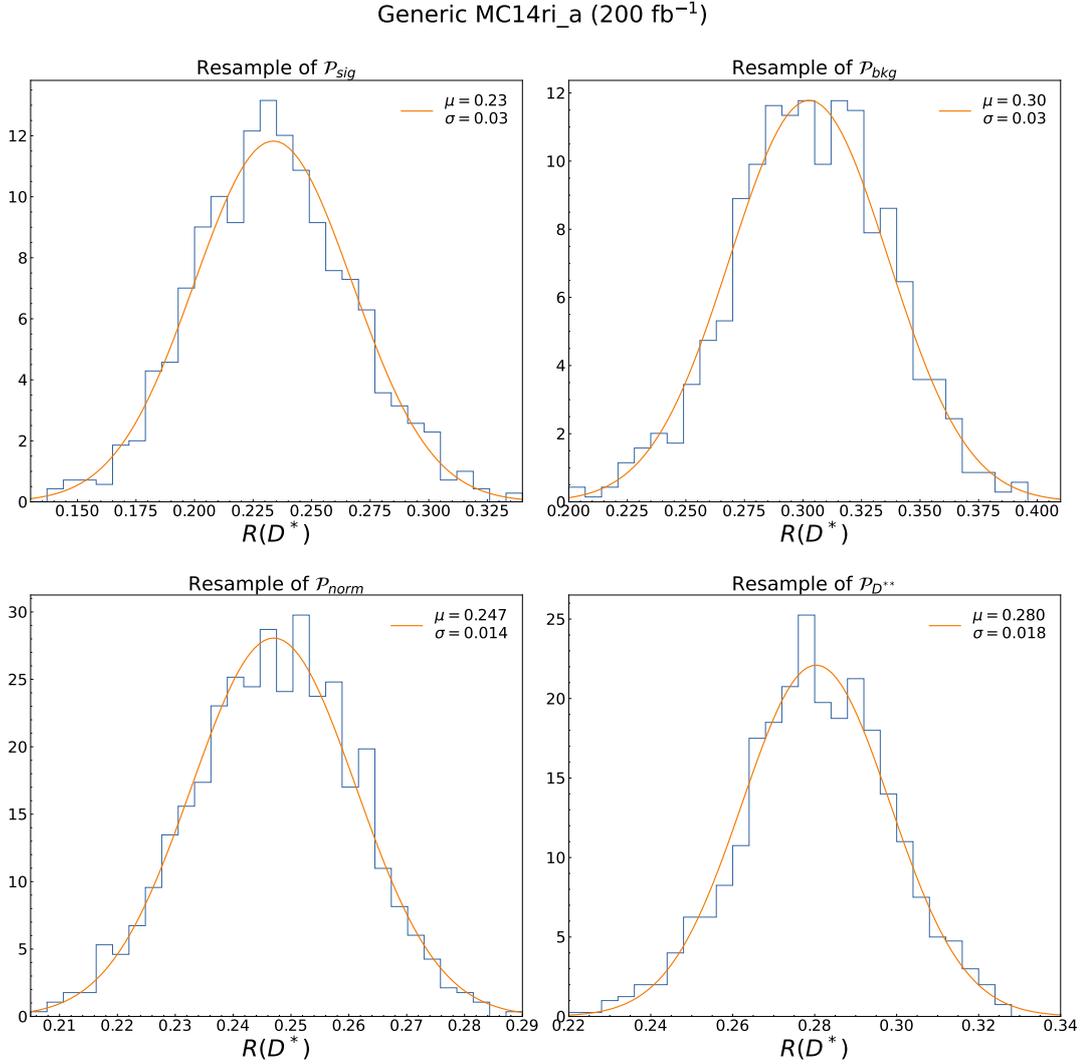


Figure 7.3: Normalized $R(D^*)$ distribution of resampled PDF-samples for each fit component.

standard deviation represents the systematic uncertainty arising from different PDF shapes due to MC statistics for the respective fit component. As the signal statistics is the worst among all fit components a high uncertainty is observed for the shape of \mathcal{P}_{sig} . However, due to the great correlation between the signal and background component, the uncertainty connected to the shape of \mathcal{P}_{bkg} is equally high. Note that some means of the $R(D^*)$ distributions for the resampled samples are not in correspondence with the

nominal $R(D^*)$ result in Ch. 6. However, they are still within the statistical uncertainty boundaries of 1σ .

As a last step, the squared standard deviations of each component are summed, so that the total systematic uncertainty is given by the square-root of the sum.

As listed in Tab. 7.1 the systematic uncertainty regarded is by far the largest (20% of the measured $R(D^*)$ value). However, this uncertainty should decrease with increased MC statistics.

7.2 $B \rightarrow D^{**}\ell\nu_\ell$ & Gap Mode

As the $B \rightarrow D^{**}\ell\nu_\ell$ background consists of four different D -meson states of higher excitation ($L = 1$), the shape of the corresponding fit component differs depending on how often each D^{**} -mode occurs. Therefore, each $B \rightarrow D^{**}\ell\nu_\ell$ decay mode is varied within its branching fraction uncertainty. Those uncertainties, alongside the respective isospin averaged branching fractions for the different D^{**} -modes, can be found in Tab. 7.2.

The listed branching fractions are computed by first averaging over experimental measurements. However, in general those measurements only could be done for particular D^{**} -decays, so that only partial branching fractions of $B \rightarrow D^{**}\ell\nu_\ell$ with $D^{**} \rightarrow D^{(*)}\pi^\pm$ were measured.

Imposing isospin symmetry assumptions on the measured values, one can further extrapolate partial branching fractions of $B \rightarrow D^{**}\ell\nu_\ell$, where the D^{**} -decay is unmeasured [34]. The full branching fractions of $B \rightarrow D^{**}\ell\nu_\ell$ events can then be accessed by simply adding all partial branching fractions. In addition, the branching fractions of the neutral and charged B -meson are isospin averaged. This means that using the isospin relation $\mathcal{B}(B^+) = \frac{\tau_+}{\tau_0} \mathcal{B}(B^0)$, where τ_+ (τ_0) is the lifetime of the B^+ (B^0)-meson, measurements of $\mathcal{B}(B^+)$ and $\mathcal{B}(B^0)$ are combined.

To account for the uncertainty arising from $\mathcal{B}(B \rightarrow D^{**}\ell\nu_\ell)$ decays, the PDF-sample is reweighted according to the variations and the PDFs of the D^{**} fit component is estimated via a KDE for each reweighted sample. The fit is then performed on the unmodified fit-sample. Of the upper and lower variation only the $R(D^*)$ value which

Decay Mode	$\mathcal{B}(B^+) [10^{-3}]$	$\mathcal{B}(B^0) [10^{-3}]$
$B \rightarrow D_1\ell^+\nu_\ell$	6.6322 ± 1.0894	6.1638 ± 1.0127
$B \rightarrow D_0^*\ell^+\nu_\ell$	4.2000 ± 0.7500	3.9033 ± 0.6972
$B \rightarrow D_1'\ell^+\nu_\ell$	4.2000 ± 0.9000	3.9033 ± 0.8366
$B \rightarrow D_2^*\ell^+\nu_\ell$	2.9337 ± 0.3248	2.7265 ± 0.3020
$B \rightarrow D^*\eta\ell^+\nu_\ell$	3.7700 ± 3.7700	4.0920 ± 4.0920
$B \rightarrow D\eta\ell^+\nu_\ell$	3.7700 ± 3.7700	4.0920 ± 4.0920

Table 7.2: Isospin averaged branching fractions with their respective uncertainties of different $B \rightarrow D^{**}\ell^+\nu_\ell$ decay modes and gap mode for charged and neutral B -mesons taken from [34].

differs most from the nominal $R(D^*)$ value calculated in Ch. 6 is kept, so that the uncertainty is symmetric. The systematic uncertainties corresponding to the different D^{**} -modes are then estimated by the difference of nominal and reweighted $R(D^*)$ value.

To additionally account for the difference in the inclusively measured branching fractions and the sum of the exclusively measured branching fractions of $X_c \ell \nu_\ell$ modes, a so called *gap mode* is introduced in the MC simulation, whose branching fraction is defined in Eq. 7.1.

$$\mathcal{B}(\text{gap}) = \mathcal{B}(B \rightarrow X_c \ell \nu_\ell)_{inc} - \sum_{exc} \mathcal{B}(B \rightarrow X_c \ell \nu_\ell)_{exc} \quad (7.1)$$

In the MC sample used the gap mode corresponds to $B^0 \rightarrow D^{*-} \eta \ell \nu_\ell$ and $B^0 \rightarrow D^- \eta \ell \nu_\ell$, decay modes that have not been measured experimentally yet.

The uncertainty of the branching fractions are assumed to be 100%, see also in Tab. 7.2. However the $B^0 \rightarrow D^- \eta \ell \nu_\ell$ do not contribute significantly to the reconstruction modes, so that only $B^0 \rightarrow D^{*-} \eta \ell \nu_\ell$ is considered. The same procedure as for the D^{**} branching fraction is applied to compute the systematic uncertainty arising from the gap mode. As the gap mode is part of the background fit component, in this case the variation should change the shape of the background PDF. Therefore, the KDE is performed on reweighted background samples. Note that there are theoretically possible but experimentally unmeasured B -decays containing e.g. a $D^{(*)}(2S)$ -meson ($L = 2$), which are not included in the MC. Such decays would contribute to the gap and if observed experimentally, their branching fractions should also be varied analogously.

To summarize the total regarded systematic uncertainty, again the square root of the sum of each uncertainty squared is considered.

7.3 Efficiency

Efficiencies calculated as described in Eq. 4.1 are computed by the selected number of events, which are estimated to follow a binomial distribution. Hence, as discussed in Ch. 4, the corresponding uncertainty of the efficiency is given by Eq. 4.2.

For this analysis this uncertainty is considered a systematic uncertainty, which is kept low by using the PDF-sample corresponding to 700 fb^{-1} to calculate the efficiencies.

7.4 PID Corrections

For the light lepton selection, PID cuts are applied to increase purity. However, between MC and real data the particle identification is expected to differ slightly. For that purpose so called *PID corrections* are applied. Those corrections reweight each event according to the light lepton's momentum, angular distribution and charge. The corresponding weights can be accessed internally through PID correction tables, which are computed centrally by the collaboration and are regularly updated.

As the momentum distributions of the light lepton for signal and normalization mode differ, the PID corrections are expected to be different for those two modes and with that also the systematic uncertainty arising from the correction. As a consequence, one cannot assume that this uncertainty cancels out when taking the ratio of signal and normalization mode.

In addition to the PID weight used for the correction, 1000 variations of the weight are computed, which vary within the statistical and systematic uncertainty of the correction. The statistical uncertainty of each bin is seen as uncorrelated to the remaining

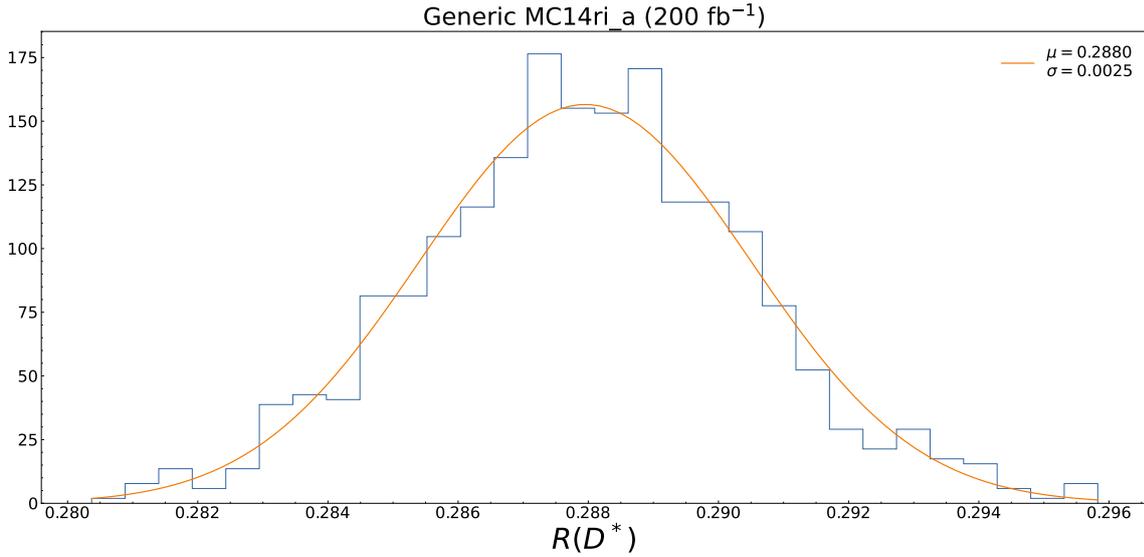


Figure 7.4: Normalized $R(D^*)$ distribution calculated with 1000 different variation of the PID corrections. The PID corrections are applied on the PDF- and fit-sample, whereas different variations only contribute to the efficiencies calculation.

bins, so that the direction of the variation of each bin is chosen randomly. In contrast, the systematic uncertainty is correlated, so that every bin varies in the same direction. The efficiency ratio is then calculated for each of the 1000 reweighted samples and the fit is performed on the PID corrected fit sample.

The corresponding $R(D^*)$ distribution as well as the applied Gaussian fit are shown in Fig. 7.4. The standard deviation is used as the systematic uncertainty arising from the PID corrections.

Note that the purpose of the PID-corrections is to align MC to real data. As no data is used in this analysis, PID corrections are only applied to estimate the corresponding systematic uncertainty in this section, but not for the nominal $R(D^*)$ calculation. Therefore, the mean value of the $R(D^*)$ distribution differs slightly from the value computed in Ch. 6.

Although hadronIDs are also used to select the optimal D^0 -meson, uncertainties concerning the D^* reconstruction cancel out as already mentioned.

7.5 $\tau \rightarrow \ell \nu_\ell \nu_\tau$ Branching Fraction

The branching fractions of the leptonic τ -decays are corrected to their PDG values, $\mathcal{B}(\tau \rightarrow e \nu_e \nu_\tau) = (17.82 \pm 0.04)\%$ and $\mathcal{B}(\tau \rightarrow \mu \nu_\mu \nu_\tau) = (17.39 \pm 0.04)\%$ [5], and varied within their uncertainty boundaries. However, so far the fit is only performed on MC, so the correction effects cancel out.

To still account for the systematic uncertainty arising from the τ -decay branching fractions, $R(D^*)$ is computed with the efficiency ratio from the upper and lower variations corresponding to the respective branching fraction uncertainties.

Similarly as done for the D^{**} systematic, from the upper and lower variation only the one with the greatest difference to the nominal $R(D^*)$ value is kept and this difference is

then estimated to be the systematic uncertainty.

7.6 Other Systematic Uncertainties

The following systematic uncertainties, although partially mentioned in previous $R(D^*)$ measurements such as in [25] and [26], were not computed during this analysis:

Form Factors As mentioned in Ch. 2 the form factor parametrization relies on experimental measurements, which are after all erroneous quantities. All form factor parameters for the relevant decay modes should therefore be varied within their uncertainties and the PDF-sample reweighted according to the respective variation. After performing the fit, the resulting discrepancy of $R(D^*)$ to the nominal value describes the systematic uncertainty.

Background Shape The background component of the fit includes very different decay modes, so that the PDF-shape depends on the exact compositions of that background component. Similarly as done for the D^{**} -background the branching fraction of dominant background components should therefore be varied within their uncertainties.

Detection & Reconstruction As all final state particles are shared between signal and normalization mode, most uncertainties concerning the track reconstruction and final state radiation should cancel out. However, potential differences in e.g. the slow pion as well as the tag-side reconstruction efficiency should still be studied and if present the efficiency ratios should be varied within the respective uncertainties.

8 Conclusion

In this analysis the possibility of measuring $R(D^*)$ using an inclusive tag was studied with Belle II MC by firstly reconstructing the signal and normalization mode, secondly identifying and trying to suppress the dominant background modes and lastly performing an extended two dimensional maximum likelihood fit.

The inclusive reconstruction, although less straight forward than an exclusive reconstruction, turned out to work properly and with a higher tag reconstruction efficiency. Nonetheless, the tag-side reconstruction is not ideal yet, especially concerning the track selection, which is why options for how it could be optimized in the future, were discussed.

Even so, tag-side multiplicities were shown to be similar for signal and normalization mode, indicating that any systematic uncertainties coming from mis-reconstructed B_{tag}^0 mesons when measuring $R(D^*)$ can be discarded.

Furthermore, the dominant background modes were determined, which were mainly combinatorial background. However, no background could be suppressed by using the standard algorithms within basf2, namely Fast-BDTs, but reasons on why this might be the case and also alternative approaches were provided.

Consequently, the big problem of high background levels when using an inclusive tag could not be overcome in this analysis. To try to at least constrain a background component in the fit, D^{**} -events were reconstructed explicitly. Unfortunately, again the problem of having too similar signal and background distributions in the D^{**} -events related variables was encountered, so that the D^{**} -reconstruction did not provide any further insights.

Nonetheless, the fit could still be performed, resulting in $R(D^*) = 0.26 \pm 0.04 \pm 0.06$, where the first uncertainty is statistical and the second is systematic and which within its uncertainty boundaries corresponded to the actual $R_{MC}(D^*) = 0.258$ value present in the MC. In addition, some systematic uncertainties, which according to other analyses are among the most dominant ones, could be estimated. In Fig. 8.1 a comparison of previously measured $R(D^*)$ values is presented. The averaged measurement is also included, but its shown uncertainty corresponds to the scaled uncertainty measured in Ch. 7 for different sizes of MC samples.

As for the question of whether or not this reconstruction procedure is competitive with the standard procedure used, namely iterating over multiple exclusive decay modes with the FEI, no straight answer can be given.

Whereas the statistical uncertainty obtained when measuring $R(D^*)$ is indeed highly competitive, higher background levels are still observed, which could not be suppressed as easily resulting in worse systematics. Additionally, due to lack of time not all systematic uncertainties were detected, which hinders one to fully answer the question of competitiveness even further.

Regardless, some general remarks can be stated. The current integrated luminosity taking on resonance (at a center of mass energy corresponding to $M_{\Upsilon(4S)}$) of the Belle II

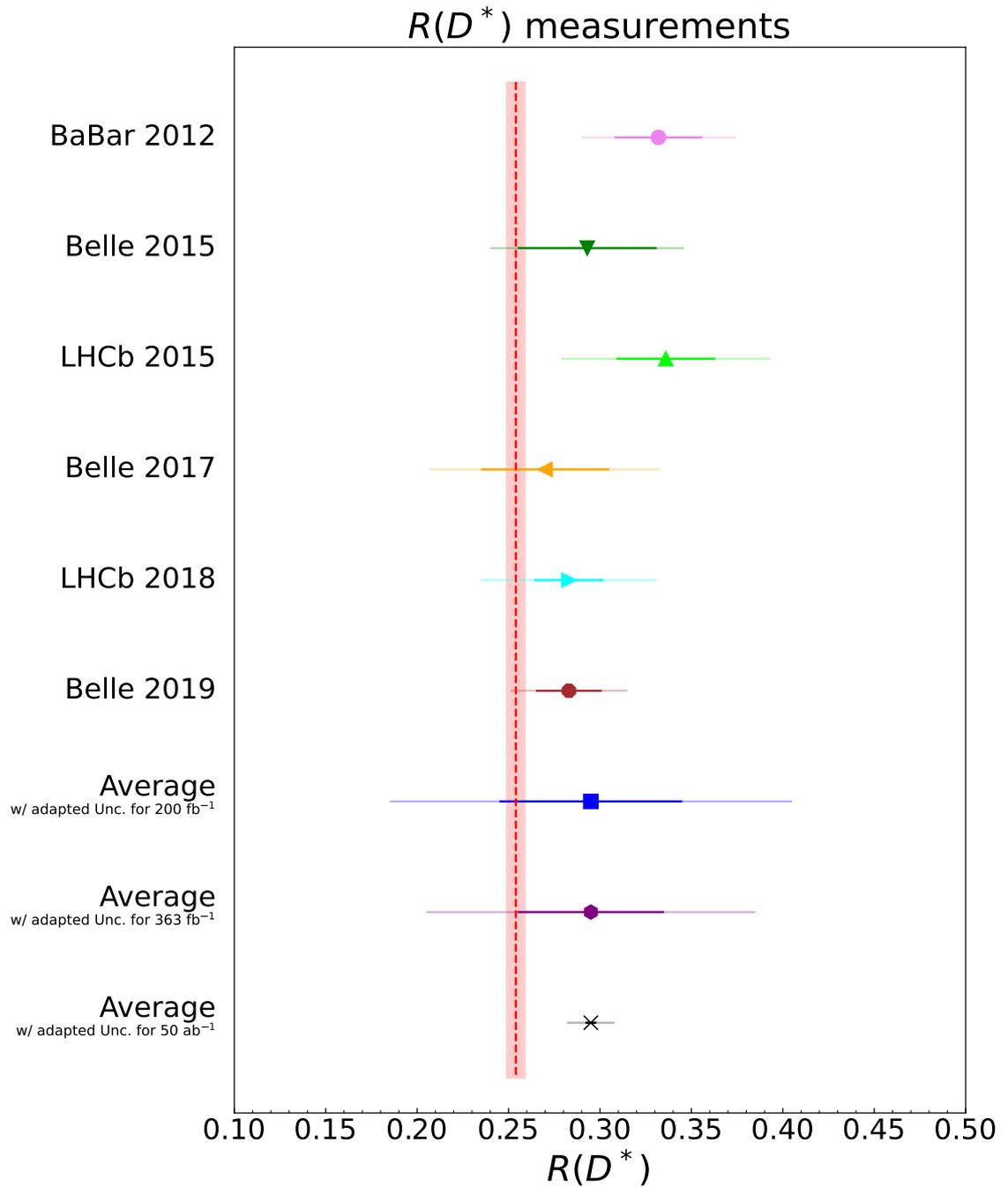


Figure 8.1: Previous $R(D^*)$ measurements taken from [1]. The uncertainty of the measured average is adapted according to the scaled uncertainties from Ch. 7 for fit-samples corresponding to 200 fb^{-1} , 363 fb^{-1} and 50 ab^{-1} . PDF-samples are assumed to always be 3.5 times bigger than the respective fit-sample. The solid error lines correspond to the statistical uncertainties, whereas the transparent lines to the total uncertainties. The vertical red line represents the theoretical average for $R(D^*)$ [1] within its uncertainty boundaries.

data is 363 fb^{-1} [7], therefore the statistics on data is still rather small, making the usage of an inclusive tag highly attractive.

Scaling the statistical uncertainty to the amount of this integrated luminosity, it decreases to 0.03, which corresponds to 11% of the measured $R(D^*)$ value.

Comparing this statistical uncertainty with the one obtained in the Belle analysis from 2015 using an exclusive hadronic tag [25], one can already see that it is smaller, although the Belle analysis worked with data of an even higher integrated luminosity (711 fb^{-1}) and considered multiple D^* - and D^0 -decay modes.

But once the statistical limit is overcome (remember Belle II is designed to end with an integrated luminosity of 50 ab^{-1} [4]) the systematic uncertainties become the bigger issue. For the inclusive tag to stay competitive, the background suppression needs to be optimized and with that also the reconstruction as a whole to hopefully access better resolved variables. However, assuming that the ratio between fit- and PDF-sample of 3.5 is kept, systematic uncertainties arising from limited MC statistics, i.e. from the PDF-shapes and the computed efficiencies, also reduce when using larger samples. Thus, for $3.5 \cdot 363 \text{ fb}^{-1}$ the total systematic uncertainty reduces to 0.04 corresponding to 16% of the measured $R(D^*)$ value.

As an additional result of the limited time of this thesis, no data was used to run over. In spite of that, one should still be careful on how to interpret the results of $R(D^*)$ when applying this reconstruction procedure on real data.

Statistically speaking one is interested in whether the null hypothesis \mathcal{H}_0 holds, which assumes, once data-MC corrections are made, that no discrepancy between the simulated and experimental value of $R(D^*)$ within some uncertainty boundaries should be observed. As the MC simulation uses the theoretical calculated value of $R(D^*)$, one can safely state that the resulting PDFs presented in Ch. 6 should be the actual SM-PDFs. Therefore, the only statistically correct statement one can make, if indeed there exists a discrepancy between theory and experiment, is that the MC simulations are built with wrong assumptions. The exact nature of those wrong assumptions, however, cannot be named.

Of course it could mean some NP process is present, but it could also mean that some SM calculation used in the simulation is incomplete.

To safely probe any NP, one needs to generate a simulation with NP processes and parameters and use the resulting PDFs of this simulation to again measure $R(D^*)$. But as there are multiple possible NP scenarios relying on different theory extensions, this is a rather complicated and computationally expensive procedure.

Therefore, for now we should be happy with probing the known physics of the SM with the experiment, before getting ahead of ourselves.

Bibliography

- [1] Y. Amhis et al. “Averages of b hadron, c -hadron, and τ -lepton properties as of 2021”. In: (June 2022). arXiv: 2206.07501 [hep-ex].
- [2] Florian U. Bernlochner et al. “Semitaquonic b -hadron decays: A lepton flavor universality laboratory”. In: *Rev. Mod. Phys.* 94.1 (2022), p. 015003. doi: 10.1103/RevModPhys.94.015003. arXiv: 2101.08326 [hep-ex].
- [3] Belle II collaboration. “Belle II website”. URL: <https://www.belle2.org/archives/>. (accessed: 01.07.2020).
- [4] E. Kou et al. “The Belle II Physics Book”. In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019). ISSN: 2050-3911. doi: 10.1093/ptep/ptz106.
- [5] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. doi: 10.1093/ptep/ptac097.
- [6] T. Abe et al. “Belle II Technical Design Report”. In: (Nov. 2010). arXiv: 1011.0352 [physics.ins-det].
- [7] Belle II collaboration. “Belle II - DESY Confluence”. URL: <https://confluence.desy.de/display/BI/Belle+II+Luminosity>. (accessed: 31.08.2022).
- [8] T. Keck et al. “The Full Event Interpretation”. In: *Computing and Software for Big Science* 3.1 (Feb. 2019). ISSN: 2510-2044. doi: 10.1007/s41781-019-0021-8. URL: <http://dx.doi.org/10.1007/s41781-019-0021-8>.
- [9] Thomas Keck. “The Full Event Interpretation for Belle II”. MA thesis. Karlsruhe Institute of Technology, Nov. 2014.
- [10] A. Matyja et al. “Observation of $B^0 \rightarrow D^{*-} \tau^+ \nu_\tau$ decay at Belle”. In: *Phys. Rev. Lett.* 99 (2007), p. 191807. doi: 10.1103/PhysRevLett.99.191807. arXiv: 0706.4429 [hep-ex].
- [11] A. Bozek et al. “Observation of $B^+ \rightarrow \bar{D}^0 \tau^+ \nu_\tau$ and Evidence for $B^+ \rightarrow \bar{D}^0 \tau^+ \nu_\tau$ at Belle”. In: *Phys. Rev. D* 82 (2010), p. 072005. doi: 10.1103/PhysRevD.82.072005. arXiv: 1005.2302 [hep-ex].
- [12] C. Glenn Boyd, Benjamin Grinstein, and Richard F. Lebed. “Model-Independent Determinations of $B \rightarrow D \ell \nu$, $D^* \ell \nu$ Form Factors”. In: (1995). doi: 10.48550/ARXIV.HEP-PH/9508211. URL: <https://arxiv.org/abs/hep-ph/9508211>.
- [13] Irinel Caprini, Laurent Lellouch, and Matthias Neubert. “Dispersive bounds on the shape of form factors”. In: *Nuclear Physics B* 530.1-2 (Oct. 1998), pp. 153–181. doi: 10.1016/s0550-3213(98)00350-2. URL: <https://doi.org/10.1016%2Fs0550-3213%2898%2900350-2>.
- [14] Daniel Ferlewicz, Phillip Urquijo, and Eiasa Waheed. “Revisiting fits to $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$ to measure $|V_{cb}|$ with novel methods and preliminary LQCD data at nonzero recoil”. In: *Phys. Rev. D* 103.7 (2021), p. 073005. doi: 10.1103/PhysRevD.103.073005. arXiv: 2008.09341 [hep-ph].

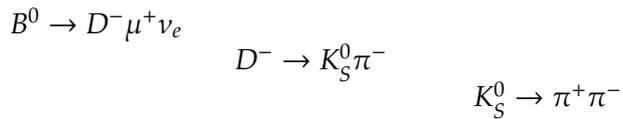
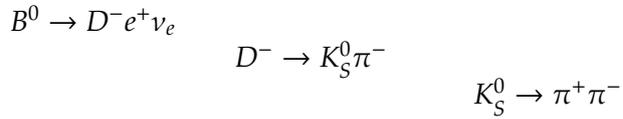
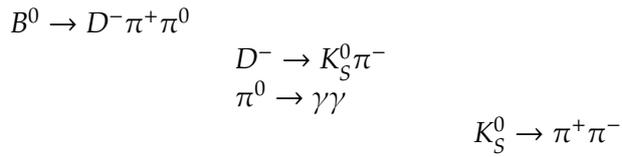
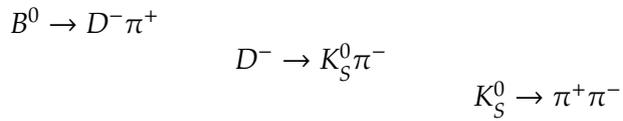
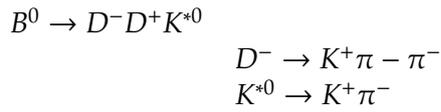
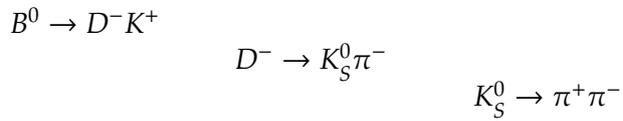
- [15] Y. Amhis et al. “Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016”. In: *Eur. Phys. J. C* 77.12 (2017), p. 895. DOI: 10.1140/epjc/s10052-017-5058-4. arXiv: 1612.07233 [hep-ex].
- [16] Belle II collaboration. “Belle II - DESY Confluence”. URL: <https://confluence.desy.de/pages/viewpage.action?spaceKey=BI&title=Collection+summary>. (accessed: 31.08.2022).
- [17] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.
- [18] S. Jadach, B. F. L. Ward, and Z. Was. “The Precision Monte Carlo event generator KK for two fermion final states in e^+e^- collisions”. In: *Comput. Phys. Commun.* 130 (2000), pp. 260–325. DOI: 10.1016/S0010-4655(00)00048-5. arXiv: hep-ph/9912214.
- [19] T. Sjostrand, S. Mrenna, and P.Z. Skands. “PYTHIA 6.4 Physics and Manual”. In: *JHEP* 05 (2006). DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175.
- [20] S. Agostinelli et al. “GEANT4—a simulation toolkit”. In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.
- [21] “Strategic accelerator design (SAD)”. URL: <https://acc-physics.kek.jp/SAD/>. (accessed: 31.08.2022).
- [22] T. Kuhr et al. “The Belle II Core Software”. In: *Computing and Software for Big Science* 3.1 (Nov. 2018). ISSN: 2510-2044. DOI: 10.1007/s41781-018-0017-9. URL: <http://dx.doi.org/10.1007/s41781-018-0017-9>.
- [23] Henrik Junkerkalefeld. “Bremsstrahlung recommendations for MC14”. 41. Belle II General Meeting - Slides. https://indico.belle2.org/event/6017/contributions/31516/attachments/14845/22307/2022.01.20_HJ_Bremsstrahlung.pdf. Jan. 2022.
- [24] Valerio Bertacchi et al. “Track finding at Belle II”. In: *Comput. Phys. Commun.* 259 (2021), p. 107610. DOI: 10.1016/j.cpc.2020.107610. arXiv: 2003.12466 [physics.ins-det].
- [25] M. Huschle et al. “Measurement of the branching ratio of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle”. In: *Phys. Rev. D* 92.7 (2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. arXiv: 1507.03233 [hep-ex].
- [26] J. P. Lees et al. “Measurement of an Excess of $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons”. In: *Phys. Rev. D* 88.7 (2013), p. 072012. DOI: 10.1103/PhysRevD.88.072012. arXiv: 1303.0571 [hep-ex].
- [27] Kyle Amirie. “An electromagnetic calorimeter cluster classifier as a beam background suppressant in $B^0 \rightarrow D^{*+}\ell^-\nu_\ell$ decays at the Belle II experiment”. MA thesis. McGill University, Nov. 2020.
- [28] Belle II collaboration. “Belle II - Online Book”. URL: <https://software.belle2.org/sphinx/release-06-00-03/index.html>. (accessed: 31.08.2022).
- [29] Alon Hershenhorn, Torben Ferber, and Christopher Hearty. “ECL shower shape variables based on Zernike moments”. In: (Nov. 2020).

- [30] Thomas Keck. “FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification”. In: (2016). DOI: 10.48550/ARXIV.1609.06119. URL: <https://arxiv.org/abs/1609.06119>.
- [31] A. J. Bevan et al. “The Physics of the B Factories”. In: *Eur. Phys. J. C* 74 (2014), p. 3026. DOI: 10.1140/epjc/s10052-014-3026-9. arXiv: 1406.6311 [hep-ex].
- [32] Glen Cowan. “Statistical Data Analysis”. New York: Oxford University Press, 1998. ISBN: 978-0-198-50156-5.
- [33] Wouter Verkerke and David P. Kirkby. “The RooFit toolkit for data modeling”. In: *eConf C0303241* (2003). Ed. by L. Lyons and Muge Karagoz, MOLT007. arXiv: physics/0306116.
- [34] Belle II collaboration. “Belle II - DESY Confluence”. URL: <https://confluence.desy.de/pages/viewpage.action?pageId=202394372>. (accessed: 31.08.2022).

Appendix

Decay Trees

To study the tagging performance 10000 $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$ (with $D^{*-} \rightarrow \bar{D}^0 \pi^-$ and $\bar{D}^0 \rightarrow K^+ \pi^-$) events were generated with the following fixed tag-side decays:



Acknowledgement

I would like to thank Prof. Dr. Thomas Kuhr and his group. I truly felt welcome and enjoyed each lunch and coffee break. A special thanks to Dr. Thomas Lueck for the numerous fruitful discussions and for guiding me patiently throughout the entire thesis. At last, I thank my family and friends for their constant support and inspiration.

Selbstständigkeitserklärung

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

München, 13.09.2022

Unterschrift