

# Search for $B \rightarrow \mu\nu_\mu$ at Belle

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

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

genehmigte

DISSERTATION

von

**Markus Tobias Prim**

aus Heidenheim a. d. Brenz

Tag der mündlichen Prüfung: 25. Januar 2019

Referent: Prof. Dr. Michael Feindt

Korreferent: Prof. Dr. Florian Bernlochner



# Contents

<b>1. Introduction</b>	<b>1</b>
<b>I. Search for the Rare Decay <math>B \rightarrow \mu\nu_\mu</math></b>	<b>3</b>
<b>2. The Decay <math>B \rightarrow \mu\nu_\mu</math></b>	<b>5</b>
2.1. The CKM Matrix Element $V_{ub}$	7
2.2. The Radiative Decay $B \rightarrow \mu\nu_\mu\gamma$	8
2.3. Experimental Signature of $B \rightarrow \mu N$ with a Sterile Neutrino $N$	9
2.4. Experimental Signature of a Type-II and Type-III Two Higgs Doublet	10
<b>3. The Belle Experiment</b>	<b>13</b>
3.1. The $\Upsilon(4S)$ -Resonance	13
3.2. The KEKB Accelerator Facility	15
3.3. The Belle Detector	15
3.4. Recorded Data Set	19
<b>4. Signal Reconstruction and Selection</b>	<b>21</b>
4.1. Event Reconstruction	21
4.1.1. Inclusive $B_{\text{tag}}$ Reconstruction	23
4.1.2. $B_{\text{tag}}$ Calibration and Optimization	24
4.2. The Signal Selection Classifier	28
4.2.1. Input Variables	28
4.2.2. MVA-Training	30
<b>5. Control Channel <math>B \rightarrow D[-\rightarrow K\pi]\pi</math></b>	<b>35</b>
5.1. Inclusive Tag Reconstruction	36
5.2. Classifier Output	37
<b>6. Signal Extraction</b>	<b>41</b>
6.1. Goodness of Fit	43
6.2. Significance	44
6.3. Likelihood-Profile	45
6.4. Toy-Experiments and Linearity Test	45
6.5. Determination of Upper Limit	47
6.6. Categorization	48

6.7. Validation . . . . .	48
6.7.1. Fit on Asimov Data . . . . .	49
6.7.2. Fit on Classifier Sideband Data . . . . .	50
6.7.3. Fit on Momentum Sideband Data . . . . .	53
<b>7. Systematics</b>	<b>59</b>
7.1. Limited MC Statistics . . . . .	59
7.2. Continuum Template . . . . .	59
7.3. $b \rightarrow c$ Template . . . . .	59
7.4. Rare Decay Template . . . . .	61
7.5. $B \rightarrow \mu\nu_\mu\gamma$ . . . . .	61
7.6. $b \rightarrow u$ Template . . . . .	64
7.7. Signal Template . . . . .	65
<b>8. Measurement on Data and its Interpretation</b>	<b>67</b>
8.1. Exclusion Limits for Sterile Neutrinos . . . . .	69
8.2. Exclusion Limits for Two Higgs-Doublet Models . . . . .	69
<b>II. Semileptonic <math>B \rightarrow X_u l \nu</math> Decays</b>	<b>77</b>
<b>9. The Decay <math>B \rightarrow X_u l \nu</math></b>	<b>79</b>
9.1. Inclusive Predictions . . . . .	80
9.2. Exclusive Predictions . . . . .	81
9.2.1. The $B \rightarrow \pi l \nu$ Decay in the SM . . . . .	81
9.2.2. The $B \rightarrow \rho/\omega l \nu$ Decays in the SM . . . . .	82
9.2.3. BCL Form Factors . . . . .	83
9.2.4. BCL Form Factors in EvtGen . . . . .	84
<b>10. Improvement of the <math>B \rightarrow \rho l \nu</math> and <math>B \rightarrow \omega l \nu</math> Form Factors</b>	<b>87</b>
10.1. Fitting Procedure . . . . .	87
10.1.1. BCL Coefficient Fit . . . . .	88
10.1.2. Legacy Spectrum . . . . .	88
10.2. Results . . . . .	90
<b>11. <math>B \rightarrow X_u l \nu</math> Hybrid MC</b>	<b>95</b>
11.1. Hybrid Model Design . . . . .	95
11.2. Comparison to Belle's Model . . . . .	96
<b>12. Summary</b>	<b>101</b>

# 1. Introduction

Over two millennia ago the Greek philosophers Leucippus of Miletus and his disciple Democritus of Abdera founded the school of thought called *atomism*, which states that all matter is composed of small and indivisible particles. Although their theory did not survive the test of time, the conception of fundamental particles is still valid and can now be found in the Standard Model of Particle Physics (SM). Over the last decades, the SM developed into an incredibly successful theory which describes the fundamental building blocks of the visible universe and three of its four known fundamental interactions. The last missing piece predicted by the SM was discovered in 2012 with the Higgs boson [1, 2].

Although the SM is a success story, it is not able to answer the remaining open questions, e.g. the strong CP problem, the hierarchy of the CKM matrix, the matter-antimatter asymmetry in the universe and the dark matter content of the universe. In the search for answers to those questions the predictions of the SM are extensively tested, searching for a deviation that can give an indication of physics processes and matter types beyond the ones described by the SM.

One experiment involved in this quest is the Belle experiment in Tsukuba, Japan. It consists of the Belle detector, which is build around the interaction region of the KEKB accelerator. KEKB is an electron-positron collider with a center-of-mass energy equal to the mass of the  $\Upsilon(4S)$  resonance. This bottomonium resonance decays almost exclusively into a  $B\bar{B}$  meson pair. Through studying the decays of those mesons, the prediction of the SM for their decay modes can be tested.

One of the thousands of possible decay modes is the decay  $B \rightarrow \mu\nu_\mu$  which is predicted by the SM with a branching ratio of  $\mathcal{O}(10^{-7})$ . Due to its very small branching ratio, no evidence for the decay could be reported by any present-day experiment. In Part I of this thesis, I describe my search for the decay  $B \rightarrow \mu\nu_\mu$  with the data sample of 770 million  $B\bar{B}$  meson pairs recorded by the Belle experiment: In Chapter 2 the decay  $B \rightarrow \mu\nu_\mu$  is discussed in the context of the SM and how the experimental signature can be modified by physics beyond the SM. A brief overview of the Belle detector and the KEKB accelerator is given in Chapter 3. The event reconstruction is described in Chapter 4, followed by a validation of the reconstruction on an independent control channel in Chapter 5. The signal extraction is explained in Chapter 6, which is followed by a study of the systematic uncertainties in Chapter 7. Chapter 8 presents the result of the search, together with its interpretation.

The semileptonic B meson decays  $B \rightarrow X_u \ell \nu$  play a major role in the search for the decay  $B \rightarrow \mu \nu$ . It is the dominant source of background and its modeling is thoroughly studied in this thesis. The results of those studies are presented in Part II: The description of the semileptonic  $b \rightarrow u$  decays within the SM is briefly discussed in Chapter 9. The improvement of the predictions for the decays  $B \rightarrow \rho \ell \nu$  and  $B \rightarrow \omega \ell \nu$  is described in Chapter 10. An improved model for the simulation of the  $B \rightarrow X_u \ell \nu$  decays is presented in Chapter 11.

Chapter 12 gives a summary of the results in Part I and Part II of this thesis.

Part I.

Search for the Rare Decay

$$B \rightarrow \mu \nu_{\mu}$$



## 2. The Decay $B \rightarrow \mu\nu_\mu$

The branching ratio of the decay  $B \rightarrow \mu\nu_\mu$  can be calculated within the Standard Model (SM) of particle physics. It can be depicted as a Feynman diagram, which encodes the rules on how to calculate the corresponding matrix element of the decay. The tree level Feynman diagram is shown in Figure 2.1. The SM prediction of the branching ratio is given by [3]

$$\mathcal{B}(B \rightarrow \ell\nu_\ell)_{\text{SM}} = \frac{G_F^2 m_B m_\ell^2}{8\pi} \left(1 - \frac{m_\ell^2}{m_B^2}\right)^2 f_B^2 |V_{ub}|^2 \tau_B, \quad (2.1)$$

with the Fermi constant  $G_F$ , the B meson mass  $m_B$ , the lepton mass  $m_\ell$ , the B meson decay constant  $f_B$ , the CKM matrix element  $V_{ub}$  and the B meson lifetime  $\tau_B$ . The numerical values of the constants are tabulated in Table 2.1. The SM prediction using Equation (2.1) and the numerical values of the constants in Table 2.1 are shown in Table 2.2. The interesting properties of Equation (2.1) are discussed in the following.

**Coupling Constant  $G_F$**  The energy scale of the process is given by the mass of the B meson  $m_B$ , which is considerably smaller than the mass of the gauge boson W which mediates the decay. This allows the decay to be treated with the effective weak coupling constant, Fermi's constant  $G_F$ .

**Helicity Suppression** The decay  $B \rightarrow \mu\nu_\mu$  is a decay of a spin 0 particle into two spin  $1/2$  particles. To create a final state spin configuration with  $S=0$ , the spin configuration of the final state particles has to be anti-parallel. The neutrino is, in good approximation, treated as a massless particle which only appears with left-handed helicity. The anti-lepton has a non-zero mass, which allows the anti-lepton to be created in a "wrong" helicity state. The allowed spin configuration is shown in Figure 2.2. The

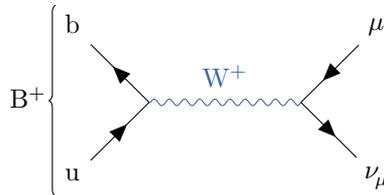


Figure 2.1.: The tree level Feynman diagram. The  $B^+$  decay is mediated by the Standard Model gauge boson  $W^+$  with the  $\mu^+$  and the  $\nu_\mu$  in the final state.

Table 2.1.: The numerical values required to calculate the Standard Model branching ratio from Equation (2.1).

Parameter	Value	Relative Uncertainty	
$G_F$	$1.17 \text{ GeV}^{-2}$	$5.1 \cdot 10^{-7}$	[4]
$m_B$	$5.28 \text{ GeV}$	$2.7 \cdot 10^{-5}$	[4]
$m_\mu$	$105.65 \text{ MeV}$	$2.3 \cdot 10^{-8}$	[4]
$f_B$	$186 \text{ MeV}$	$2.2 \cdot 10^{-2}$	[5]
$\tau_B$	$1.64 \text{ ps}$	$2.4 \cdot 10^{-3}$	[4]
$ V_{ub} _{\text{exc}}$	3.70	$4.2 \cdot 10^{-2}$	[4]
$ V_{ub} _{\text{inc}}$	4.49	$6.2 \cdot 10^{-2}$	[4]
$ V_{ub} _{\text{avg}}$	3.94	$9.1 \cdot 10^{-2}$	[4]

Table 2.2.:  $B \rightarrow \mu\nu_\mu$  Standard Model branching ratios calculated with Equation (2.1) and the numerical values listed in Table 2.1 together with the latest experimental result.

$V_{ub}$	SM Prediction
Exclusive	$\mathcal{B}(B \rightarrow \mu\nu_\mu) = (3.76 \pm 0.36) \cdot 10^{-7}$
Inclusive	$\mathcal{B}(B \rightarrow \mu\nu_\mu) = (5.53 \pm 0.72) \cdot 10^{-7}$
Average	$\mathcal{B}(B \rightarrow \mu\nu_\mu) = (4.26 \pm 0.80) \cdot 10^{-7}$
Latest Measurement	
	$\mathcal{B}(B \rightarrow \mu\nu_\mu) = (6.46 \pm 2.71) \cdot 10^{-7} @ 2.4 \sigma$ [6]

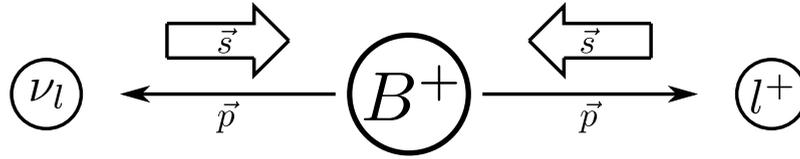


Figure 2.2.: The allowed kinematic configuration of the decay  $B \rightarrow \mu\nu_\mu$  in the reference frame of the parent  $B^+$  meson. The neutrino exists only as left-handed particle. The lepton is created in the wrong helicity state. Figure adapted from [7].

suppression by the requirement of the anti-lepton to be created in the wrong helicity state is visible in the factor  $m_l^2$ . For lighter leptons the helicity suppression becomes stronger.

**Phase Space** The suppression of the decay by the available phase space appears in the phase space factor  $(1 - m_l^2/m_B^2)^2$ , which increases the branching ratio for lighter lepton masses or heavier parent meson masses.

**Decay Constant  $f_B$**  The b and u quarks do not scatter in the initial state but are present in a bound state. This bound state and the annihilation probability of the two quark constituents is described by the decay constant  $f_B$ . It cannot be calculated with perturbation theory in the strong coupling constant  $\alpha_s$ , as the energy scale is the mass of the B meson. However, lattice QCD simulations allow for the calculation of the decay constant.

**CKM Matrix Element  $|V_{ub}|$**  The weak coupling constant is modified by the CKM matrix element  $V_{ub}$ , which is further discussed in Section 2.1.

The decay is a two-body decay. Therefore the lepton momentum in the parent B reference frame is given by a  $\delta$ -peak, which is a very clean experimental signature. The momentum of the lepton in the B reference frame is given by

$$p_l^B = \frac{\sqrt{(m_B^2 - (m_l + m_\nu)^2)(m_B^2 - (m_l - m_\nu)^2)}}{2m_B}. \quad (2.2)$$

Which results in a muon momentum in the parent B reference frame of  $p_\mu^B = 2.639 \text{ GeV}$ .

## 2.1. The CKM Matrix Element $V_{ub}$

The eigenstates of the weak interaction do not coincide with the mass eigenstates of the quark fields but are a mixture of them. This mixture is described by the CKM matrix. The idea was first introduced by Cabibbo [8] for two quark generations and later extended to three quark generations by Kobayashi and Maskawa [9] to generate a mechanism for CP violation in the SM. Although the theoretical mechanism which gives rise to the CKM matrix is well understood, there is no theoretical explanation for the numerical values of the elements of the CKM matrix. Therefore the individual elements have to be determined by experiments.

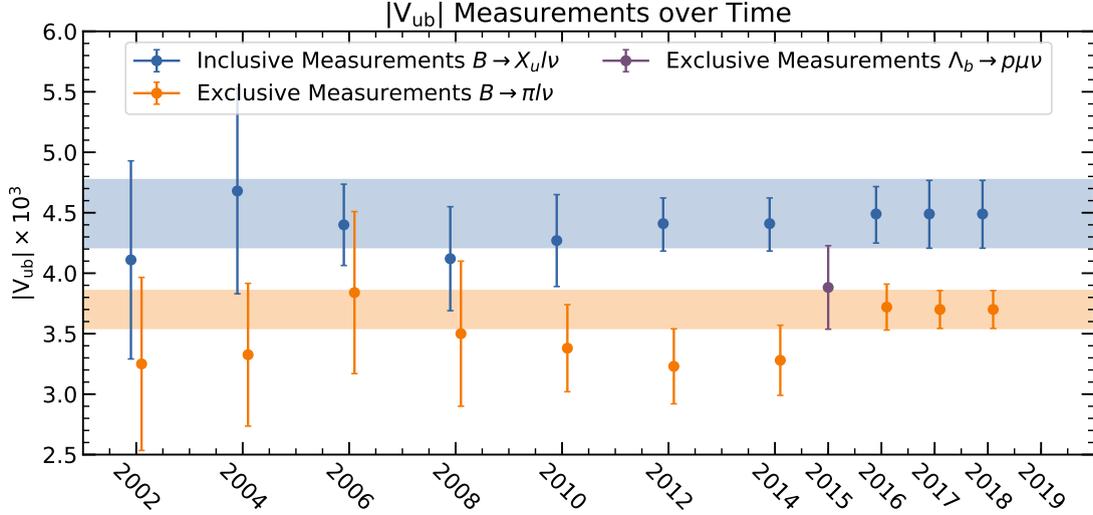


Figure 2.3.: The extracted value for  $|V_{ub}|$  for inclusive  $B \rightarrow u\ell\nu$  and exclusive  $B \rightarrow \pi\ell\nu$  measurements over the last years. With the measurement of the pure leptonic decays, an independent data point can be added to resolve the long standing tension between the  $|V_{ub}|$  measurements. Figure adapted from [10].

Up until now the value of  $|V_{ub}|$  has been determined in two different ways, either by an inclusive measurement of  $B \rightarrow X_u \ell \nu$  or an exclusive measurement of  $B \rightarrow \pi \ell \nu$  or  $\Lambda_b \rightarrow \rho \mu \nu_\mu$ . The determined values of  $|V_{ub}|$  via semileptonic decays are not compatible within their uncertainty and show a tension of approximately  $3\sigma$ . The discrepancy is unresolved up until now and should be kept in mind when the branching ratio of the SM decay  $B \rightarrow \mu\nu_\mu$  is predicted. The evolution of the  $|V_{ub}|$  measurement over the last decade is shown in Figure 2.3. Under the assumption that the leptonic decay of the B is only governed by the SM  $|V_{ub}|$  can be extracted from the decay  $B \rightarrow \mu\nu_\mu$  and provide an independent measurement to help resolve this long standing tension.

## 2.2. The Radiative Decay $B \rightarrow \mu\nu_\mu\gamma$

The three next-to-leading order (NLO) Feynman diagrams in the electromagnetic coupling constant  $\alpha_{\text{em}}$  for the decay  $B \rightarrow \mu\nu_\mu$  are shown in Figure 2.4. The corrections are split into two parts called structure dependent, where the photon is radiated from a quark line, and internal bremsstrahlung, where the photon is radiated from the lepton.

Usually, higher order diagrams are treated as corrections to the tree level process. In the case of  $B \rightarrow \mu\nu_\mu$ , the process  $B \rightarrow \mu\nu_\mu\gamma$  with high energetic photons is studied separately by theory [11, 12] and experiment [13], because by the structure dependent corrections, the B meson can be excited into a virtual  $S = 1$  resonance. This lifts the helicity suppression of the decay  $B \rightarrow \mu\nu_\mu$ , resulting in a large correction to the original branching ratio. The processes  $B \rightarrow \mu\nu_\mu\gamma$ , where the photon has an energy  $E_\gamma > 300 \text{ MeV}$  are treated as background process to  $B \rightarrow \mu\nu_\mu$ .

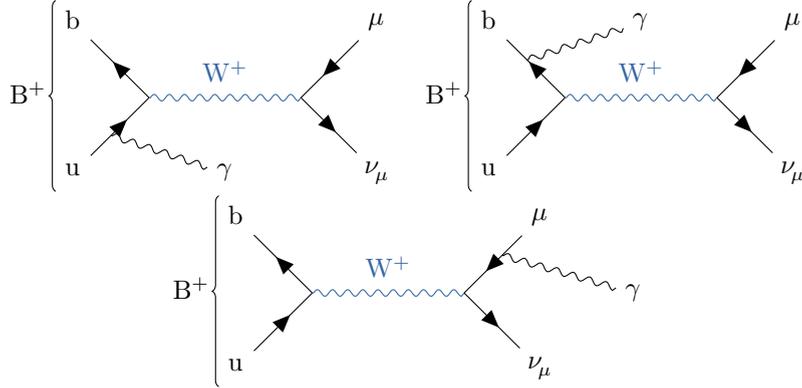


Figure 2.4.: There exist three Feynman diagrams which contribute to the leading order matrix element of the decay  $B \rightarrow \mu\nu_\mu\gamma$ . The photon can either be radiated from one of the two initial quark lines or from the final state lepton.

Corrections arising from the internal bremsstrahlung diagram are treated by PHOTOS [14] and are included in the simulation of the signal events. Photon corrections originating from the initial quark lines with  $E_\gamma < 300$  MeV are neglected.

A more thorough discussion of soft photon and radiative corrections can be found in [15–18].

### 2.3. Experimental Signature of $B \rightarrow \mu N$ with a Sterile Neutrino $N$

The sterile neutrino is a hypothetical particle which is a singlet under the fundamental symmetry group  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  of the SM. It has therefore no color charge, no weak isospin and no weak hypercharge and does not couple to the gauge bosons of the SM. There are two motivations to introduce the sterile neutrino: It can be used to explain the dark matter content of the universe [19] and the smallness of the SM neutrino mass terms [20].

The only possibility for a sterile neutrino to arise in the final state of the  $B \rightarrow \mu\nu_\mu$  decay as  $B \rightarrow \mu N$  is if there exists a non-SM mediator in the decay of the B meson and the mass of the sterile neutrino is  $m_N < m_B - m_\mu$ . The change in the experimental signature is independent of the choice of the model which introduces a new mediator and a coupling to the sterile neutrino. In the following I discuss the two possible changes in the experimental signature if there is a contribution of the process  $B \rightarrow \mu N$ .

The measured branching ratio increases in comparison to the SM, depending on the strength of the coupling of the sterile neutrino to the new mediator and on the mass of the new mediator particle, and is given by

$$\mathcal{B}(B \rightarrow \mu + \text{missing energy}) = \mathcal{B}(B \rightarrow \mu\nu_\mu) + \mathcal{B}(B \rightarrow \mu N), \quad (2.3)$$

where the missing energy is either given by the SM muon-neutrino or the sterile neutrino.

The  $B \rightarrow \mu\nu_\mu$  decay is a two-body decay and has therefore a very clear signature in the lepton spectrum in the B reference frame. This allows to search for a modification of the

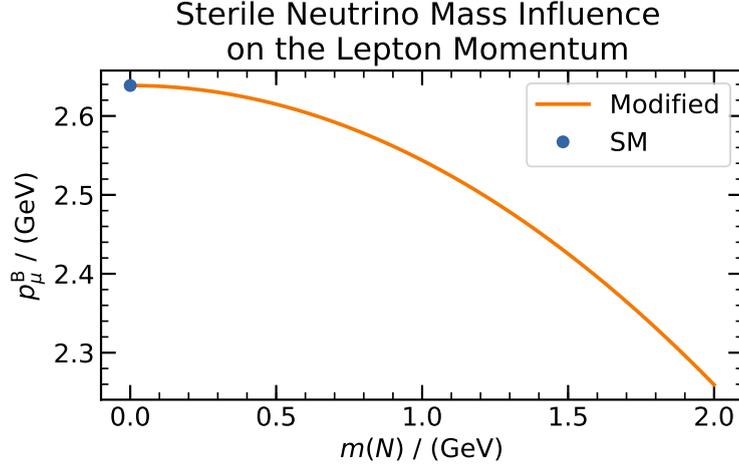


Figure 2.5.: The muon momentum in the B reference frame as a function of the mass of a hypothetical sterile neutrino  $N$ .

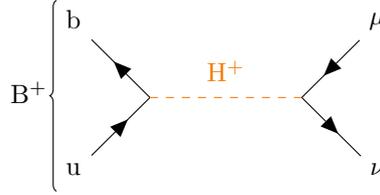


Figure 2.6.: The tree level Feynman diagram. The  $B^+$  decay is mediated by the charged Higgs boson  $H^+$  with the  $\mu^+$  and the  $\nu_\mu$  in the final state.

experimental signature of the SM decay, because the momentum of the muon is directly affected by the mass of the sterile neutrino. The lepton momentum as a function of the sterile neutrino mass can be calculated with Equation (2.2) and is shown in Figure 2.5.

Depending on the new mediator mass, the coupling strength of the new mediator to the sterile neutrino and the mass of the sterile neutrino, it is possible to measure either an enhancement in the branching ratio, two distinct peaks in the lepton spectrum or a combination of both.

## 2.4. Experimental Signature of a Type-II and Type-III Two Higgs Doublet

The two Higgs doublet model (2HDM) introduces a new mediator, a charged Higgs boson  $H^\pm$ , which can also mediate the decay  $B \rightarrow \mu\nu_\mu$ . The relevant Feynman diagram is shown in Figure 2.6. The experimental signature of a type-II or type-III 2HDM is given by a variation of the branching ratio with respect to the SM expectation. The contribution of a charged Higgs can either be constructive or destructive, depending on the model parameters, resulting in a variation of the measured branching ratio.

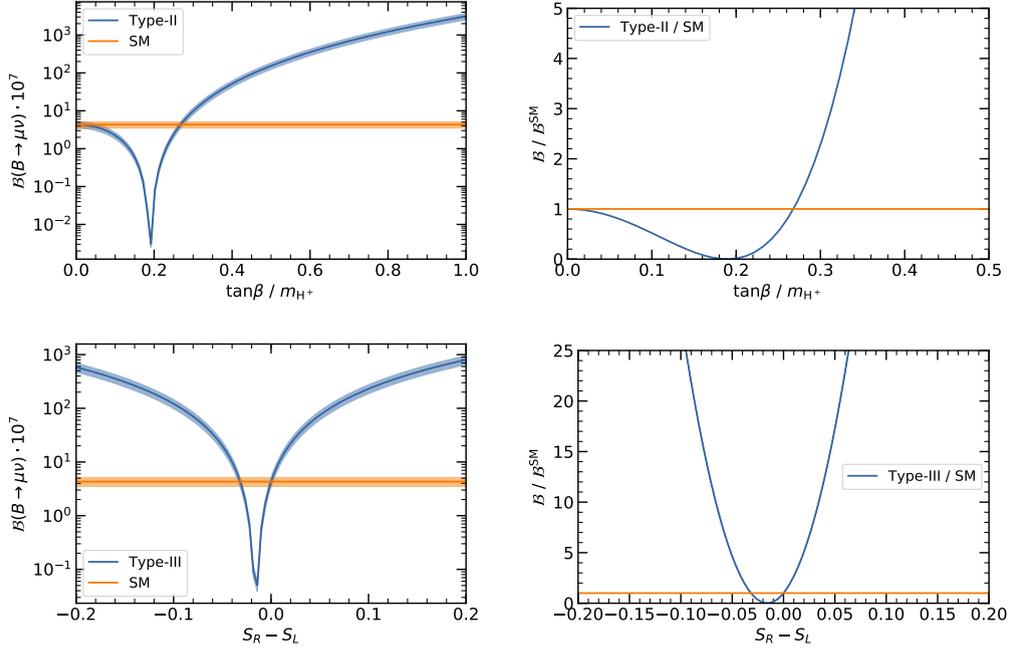


Figure 2.7.: Modification of the expected branching ratio for (top) the type-II and (bottom) the type-III two Higgs doublet model. Both constructive and destructive interference is possible depending on the choice of the model parameters.

The modification of the branching ratio for a type-II 2HDM is given by [3]

$$\mathcal{B}(B \rightarrow \mu\nu_\mu) = \mathcal{B}^{\text{SM}} \times \left| 1 - \frac{m_B^2 \tan^2 \beta}{m_{H^+}^2} \right|^2, \quad (2.4)$$

where  $\tan \beta = v_2/v_1$  is the ratio of the vacuum expectation values and  $m_{H^+}$  the mass of the charged Higgs boson.

The modification of the branching ratio for a type-III 2HDM is given by [21]

$$\mathcal{B}(B \rightarrow \mu\nu_\mu) = \mathcal{B}^{\text{SM}} \times \left| 1 + \frac{m_B^2}{m_b m_\ell} \frac{C_R - C_L}{C_{\text{SM}}} \right|^2, \quad (2.5)$$

where  $m_b$  is the bottom quark mass,  $C_{\text{SM}} = 4G_F V_{ub}/\sqrt{2}$  the SM coupling and  $C_{R(L)}$  the Wilson coefficients encoding the new physics contribution. The new physics contribution can also be expressed in relative terms  $S_{R(L)} = C_{R(L)}/C_{\text{SM}}$ .

The modification of the branching ratio by both models is shown in Figure 2.7 for the absolute branching ratio and the relative change to the SM.



## 3. The Belle Experiment

The Belle Experiment is an asymmetric beam-energy electron-positron collider experiment designed to run with a center-of-mass energy of the  $\Upsilon(4S)$ -resonance mass. Its name-giving part, the Belle detector, is constructed around the interaction region to measure the properties of the decay products that arise during an electron-positron interaction. In Section 3.1 I discuss the  $\Upsilon(4S)$ -resonance and why it is interesting to use it in an experiment. I continue with an overview of the KEKB accelerator facility in Section 3.2 and close this chapter with Section 3.3 where an overview of the Belle detector is presented.

### 3.1. The $\Upsilon(4S)$ -Resonance

The  $\Upsilon(nS)$ -resonances are spin-excited bound states of a  $b\bar{b}$ -quark pair and can, as in our case, be created via the annihilation of an  $e^+e^-$ -pair. Figure 3.1 shows the cross-section of the  $\Upsilon(nS)$ -resonances, measured by the CLEO collaboration [22]. The resonance of interest for this work is the  $\Upsilon(4S)$ -resonance, which has a mass of  $m_{\Upsilon(4S)} = 10.58 \text{ GeV}$  and can be created using an electron-positron collision with a center-of-mass energy of  $\sqrt{s}_{e^+e^-} = m_{\Upsilon(4S)}$ .

With its mass, it is just heavy enough to decay into a  $B\bar{B}$ -meson pair via the strong force. This decay occurs in more than 96% of the cases [4]. Decays via the strong force into states with lighter mesons are suppressed by the OZI-rule. Decays into lighter mesons via the weak or electromagnetic force are suppressed because their associated coupling constants are small compared to the strong coupling constant. The Feynman diagram of the production and decay of the  $\Upsilon(4S)$  is shown in Figure 3.2.

The energy gap between the mass of the  $\Upsilon(4S)$  and the mass of the two B mesons is small enough that no additional particles can be produced during the hadronization of the b quarks in the decay of the  $\Upsilon(4S)$ . In the center-of-mass system of the  $\Upsilon(4S)$  the two B mesons are produced back-to-back with a momentum of  $p_B \approx 332 \text{ MeV}$ , which means that the B's are almost at rest. This kinematic property is exclusive for the  $\Upsilon(4S)$ -resonance. If the  $\Upsilon(4S)$ -resonance is produced via the annihilation of an electron-positron pair, the environment in which the B decays can be studied is very clean because there is no underlying event or pile-up in the detector. The initial state is known and each final state particle in the detector can be assigned to the process under investigation.

The outcome of an  $e^+e^-$  collision is not always given by the creation of the  $\Upsilon(4S)$  resonance. Non-resonant final states are also produced. In the context of this thesis those processes

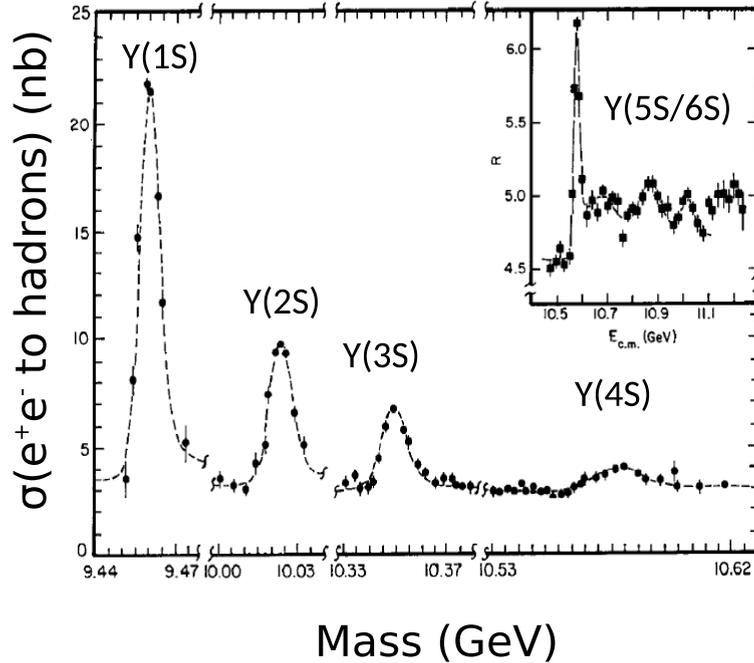


Figure 3.1.: The scan of the first  $\Upsilon$ -resonances by the Cleo Collaboration [22]. The non-resonant contribution to the spectrum is almost constant over the scanned mass range, with the resonances on top. The first three resonances are very narrow, because they can only decay via three gluons, which is suppressed by the OZI-rule. For the  $\Upsilon(4S)$  resonance and the resonances above, the kinematic threshold for the  $B\bar{B}$ -pair production is passed, which results in shorter lifetimes of the resonances.

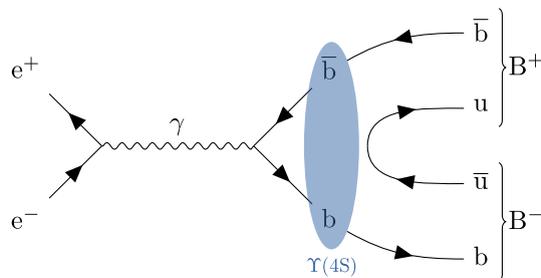


Figure 3.2.: The Feynman diagram of the production and decay of the  $\Upsilon(4S)$ . The  $e^+e^-$  pair annihilates into a virtual photon, which in turn creates a  $b\bar{b}$  pair. This  $b\bar{b}$  pair is a bound state with a fixed set of quantum numbers, here the  $\Upsilon(4S)$ . The  $\Upsilon(4S)$  decays via the strong force producing a  $B\bar{B}$ -meson pair.

are solely background. They are the so-called continuum backgrounds  $e^+e^- \rightarrow q\bar{q}$ , QED  $e^+e^- \rightarrow \bar{\ell}\ell(\gamma)$ , and two-photon  $e^+e^- \rightarrow \bar{\ell}\ell f\bar{f}$  events, with  $q \in \{u, d, s, c\}$ ,  $\ell \in \{e, \mu, \tau\}$  and  $f \in \{q, \ell\}$ .

## 3.2. The KEKB Accelerator Facility

The KEKB accelerator facility was developed to collide electrons and positrons with a center-of-mass energy corresponding to the mass of the  $\Upsilon$ (nS)-resonances and to provide optimal conditions for the measurement of physical quantities of interest with the Belle detector. A schematic overview of the installation is shown in Figure 3.3. In the following I illustrate the individual components of the facility, always referring to the conditions for operating at the  $\Upsilon$ (4S)-resonance. A detailed technical description of the individual components can be found in [23].

The two main components of the facility are the high-energy storage ring (HER) and the low-energy storage ring (LER), which have a circumference of 3 km. The HER stores an electron beam of  $I = 1.64$  A with an energy of 8.0 GeV, while the LER stores a positron beam of  $I = 1.19$  A with an energy of 3.5 GeV. Each beam consists of 1584 individual bunches. The two separate storage rings cross at the interaction point (IP), where the two beams are brought to collision. The created state, resonant or non-resonant, experiences a Lorentz boost  $\beta\gamma = 0.425$  in forward direction due to the asymmetry of the beam energies.

Filling the two storage rings with electrons and positrons and accelerating them to the required energies would require an interruption of operation. Instead, the two storage rings are continuously injected with an 8.0 GeV electron and a 3.5 GeV positron beam from the injector line to allow an uninterrupted data acquisition with the Belle detector.

The injector line is a linear accelerator<sup>1</sup> (LINAC). It is filled with electrons by an electron gun. To generate positrons, the electron beam is targeted at a tungsten plate with an energy of 3.7 GeV. The electrons and captured positrons are then accelerated to the required energy, separated and injected in the respective storage rings.

With the outlined configuration KEKB achieved a peak instantaneous luminosity of  $\mathcal{L} = 2.11 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , which is up-to-date the world's highest luminosity machine. With the  $\Upsilon$ (4S) production cross-section of  $\sigma_{\Upsilon(4S)} = 1.2 \text{ nb}$  this results in 25  $\Upsilon$ (4S) events every second.

## 3.3. The Belle Detector

The design of the Belle detector follows the typical structure of a general purpose  $4\pi$  detector used in high-energy physics and consists of several sub-detector systems. A strong emphasis was put into the vertex resolution, the identification of charged final state particles and the detection of photons. A detailed description of the Belle detector can be found in [24]. A side view of the Belle detector is shown in Figure 3.4. The individual sub-detector systems are listed in the following with a brief description of their design and purpose.

<sup>1</sup>Technically, the LINAC consists of two linear accelerators connected by a 180 deg bending magnet. The "J" topology is required because the compound is too small.

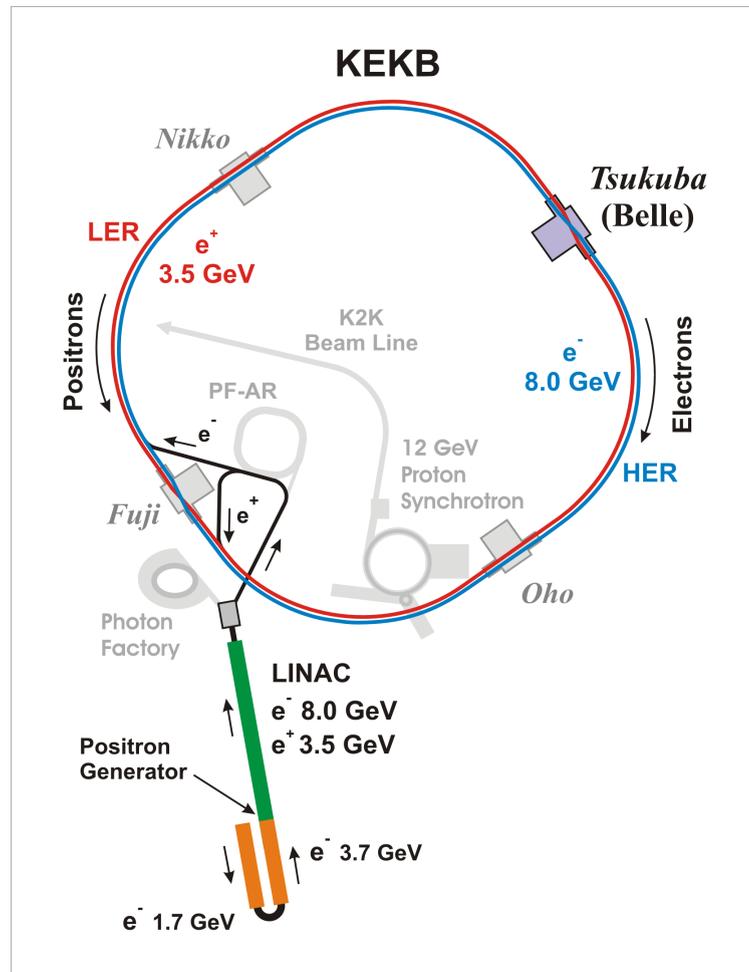


Figure 3.3.: A schematic overview of the KEKB accelerator facility. Electrons are injected into the LINAC and accelerated. Positrons are generated with the electron beam and a tungsten target. The electrons and positrons are accelerated to the nominal energies of the HER and LER and injected into the storage rings. The pre-accelerating LINAC allows to continuously inject beam bunches into the two storage rings. The two beams are brought to collision in Tsukuba Hall, where the Belle detector is located. Figure taken from [23].

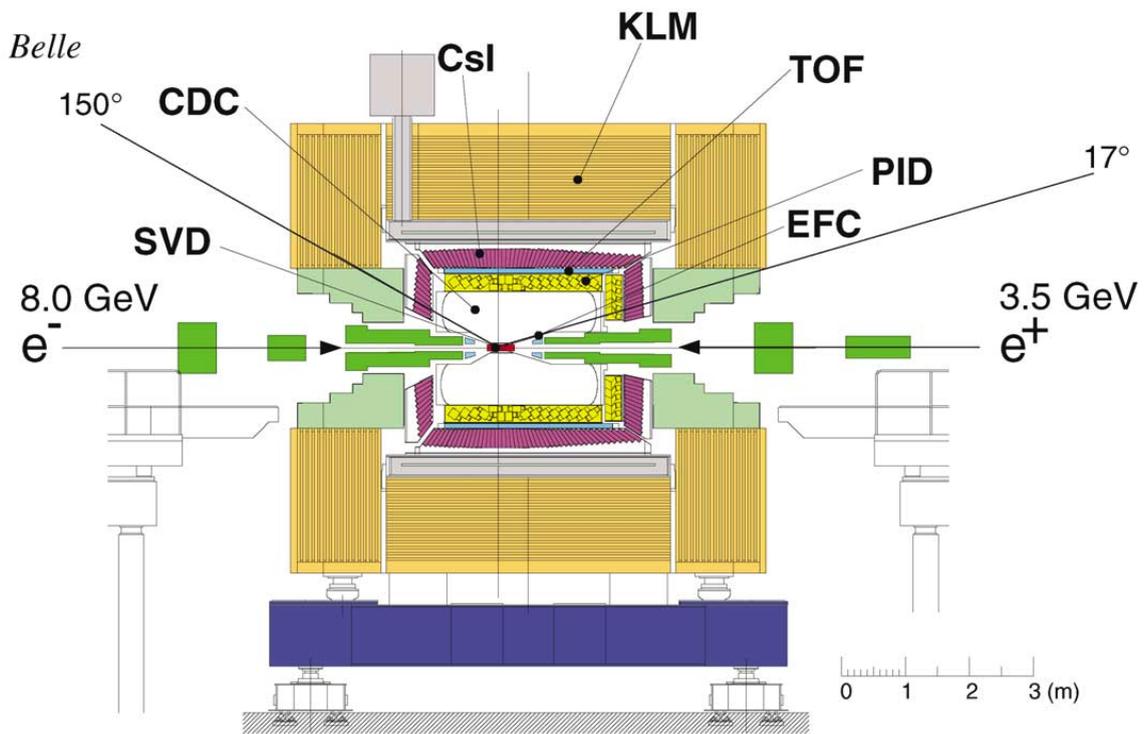


Figure 3.4.: Side view of the Belle detector. The individual sub-detectors are briefly explained in the text. Figure taken from [24].

### Interaction region

In the interaction region, the material of the beam pipe is changed from aluminum to beryllium. This reduces the impact of multiple scattering onto the particle trajectories originating inside the beam pipe. The radius of the beam pipe at the interaction point is reduced, which allows for a close positioning of the silicon vertex detector to the interaction region.

### Extreme forward calorimeter

The extreme forward calorimeter (EFC) covers a polar angle of  $\theta \in [6.4, 11.5]$  deg in forward direction and  $\theta \in [163.3, 171.2]$  deg in backward direction. Apart from providing a coverage outside of the coverage of the other sub-detector systems it is also instrumented as beam monitor. Due to its positioning, the EFC is exposed to a higher dose of radiation. The EFC consists of radiation hard bismuth germanate ( $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ ) crystals to ensure a long lifetime of the detector system.

### Silicon vertex detector

Two different designs of the silicon vertex detector (SVD) were deployed over Belle's lifetime. The first version covers a polar angle of  $\theta \in [23, 140]$  deg with three layers of double-sided silicon strips, whereas the second version covers an increased polar angle of  $\theta \in [17, 150]$  deg and has an additional layer of double-sided silicon strips. The silicon vertex detector is used to improve the resolution of the impact parameters and to reconstruct low momentum tracks which do not reach the central drift chamber.

### Central drift chamber

The central drift chamber (CDC) is designed to precisely measure the momentum of tracks. It consists of 50 layers of sense wires which are grouped into stereo and axial layers. Together with the cathode strip wires, the CDC consists of 8400 individual drift cells. The CDC is filled with a 50:50 mixture of helium and ethane. The CDC covers a polar angle of  $\theta \in [17, 150]$  deg. In combination with the time-of-flight counter and the aerogel Cherenkov counter it also provides particle identification via the measurement of  $dE/dx$ .

### Aerogel Cherenkov counter

The aerogel Cherenkov counter (ACC) is dedicated to the discrimination between high energetic charged pions and kaons. The detector is based on the Cherenkov effect, which is the effect that charged particles traversing a medium with a velocity higher than the speed of light in the medium emit light. The refractive indices  $n \approx 1.02$  of the aerogel are chosen such that pions in a momentum range of  $p \in [1.2, 3.5]$  GeV emit light, whereas kaons in the same momentum range do not.

### Time-of-flight counter

The time-of-flight detector (TOF) is also dedicated to the discrimination between charged pions and kaons and of scintillations counters. The discrimination is based on the measurement of the collision time and the scintillation signal. As particles with different masses have different velocities, the flight-time together with the momentum provided by the CDC can be used to deduce the mass of the particle. It allows to discriminate between kaons and pions with a momentum up to 1.2 GeV.

Table 3.1.: Summary of the integrated luminosity recorded at each resonance by the Belle experiment. The table is adapted from [3].

Resonance	Luminosity fb <sup>-1</sup>	
	On-resonance	Off-resonance
$\Upsilon(5S)$	121.4	1.7
$\Upsilon(4S)$	711.0	89.4
$\Upsilon(3S)$	2.9	0.2
$\Upsilon(2S)$	24.9	1.7
$\Upsilon(1S)$	5.7	1.8
Scan > $\Upsilon(4S)$	-	27.6

### Electromagnetic calorimeter

The electromagnetic calorimeter (ECL) consists of 8736 individual thallium-doped cesium iodide (CsI) crystals with two attached silicon photo diodes. The ECL is able to measure photon energies down to 50 MeV and plays an important role to discriminate electron tracks from other hadron tracks by measuring the energy deposition. The ECL covers a polar angle of  $\theta \in [17, 150]$  deg.

### Solenoid

The superconducting solenoid covers a cylindric volume with a diameter of 3.4 m and length of 4.4 m. It provides a constant 1.5 T magnetic field parallel to the beam axis within this volume. The iron structure of the Belle detector is utilized as return yoke for the magnetic field and absorber material for the KLM.

### $K_L^0$ and muon detection system

The return yoke and support structure of the Belle detector is instrumented with 15 layers of resistive plate counters to form the  $K_L^0$  and muon detection system (KLM). With approximately 4 hadronic interaction lengths the  $K_L^0$ 's create a hadronic shower in the sub-detector. However, due to the structure of hadronic showers, no measurement of the energy is possible. The information is solely used to veto  $K_L^0$  particles. Muon also reach the sub-detector, which can be used, together with the other particle identification systems, to discriminate between muons and other charged particles. The KLM covers a polar angle of  $\theta \in [20, 155]$  deg.

## 3.4. Recorded Data Set

During its ten-year run time, the Belle detector recorded a data sample of  $\mathcal{L} \approx 1 \text{ ab}^{-1}$ . The largest fraction of the data sample was recorded on the  $\Upsilon(4S)$  resonance, but data on the  $\Upsilon(1S)$ ,  $\Upsilon(2S)$ ,  $\Upsilon(3S)$  and  $\Upsilon(5S)$  were also recorded. In addition, a smaller data sample was taken below the energy of each resonance to study the non-resonant contributions at the energy near the resonances. The size of the individual data samples is shown in Table 3.1.

In addition to the recorded data set, the Belle experiment also provides a large sample of simulated events of different categories. The simulated data samples are described in the context of the signal extraction in Chapter 6.



## 4. Signal Reconstruction and Selection

In this chapter I describe how the event reconstruction and selection is performed to retrieve the data sample used for the signal extraction. In Section 4.1 I discuss the event reconstruction with an emphasis on how the inclusive tag is reconstructed, calibrated and optimized. The classifier used for signal selection relies on a properly modeled input space of variables. In Section 4.2 I discuss the input space, how I improved the MC description for it, and the output of the classifier.

### 4.1. Event Reconstruction

To consider an event a possible signal event the following conditions have to be met: At least one charged track has to originate from the interaction region with the impact parameters  $dr < 0.5$  cm and  $|dz| < 2.0$  cm, where  $dr$  is the distance in the  $r - \phi$  plane and  $dz$  the distance in  $z$  direction from the reconstructed interaction point. The track is required to have a center-of-mass<sup>1</sup> momentum  $p_\mu^* > 2.2$  GeV and a muon identification of  $\mu\text{-ID} > 0.9$ . Multiple candidates per event do not occur. Events not fulfilling those criteria are rejected.

The theoretical value of the muon momentum in the parent B reference frame  $p_\mu^B = 2.639$  GeV is given by two-body decay kinematics. The decay of the  $\Upsilon(4S)$  does not provide the two B mesons at absolute rest, but with a momentum of  $p_B^* = 332$  MeV. The muon momentum in the center-of-mass frame is therefore a convolution of the B momentum distribution in the center-of-mass frame and the muon momentum in the B reference frame. To achieve a sharper signal distribution the muon candidate can be boosted into the reference frame of its mother particle by different tagging methods, which are explained in the following. In this reference frame, the lepton momentum is theoretically a delta-peak, but smeared out by the reconstruction resolution. The difference in signal shapes between the center-of-mass and B reference frame is shown in Figure 4.1.

There are different methods available for reconstructing the companion B meson to the signal  $B_{\text{sig}}$ , namely hadronic, semileptonic and inclusive tagging. The hadronic and semileptonic tagging algorithms exclusively reconstruct the companion B meson in specified decay channels. In comparison, the inclusive tagging does not impose any constraint on the decay channel of the companion B meson.

---

<sup>1</sup>The center-of-mass reference frame is the rest frame of the  $\Upsilon(4S)$ . Variables in the center-of-mass reference frame are always denoted with a \*.

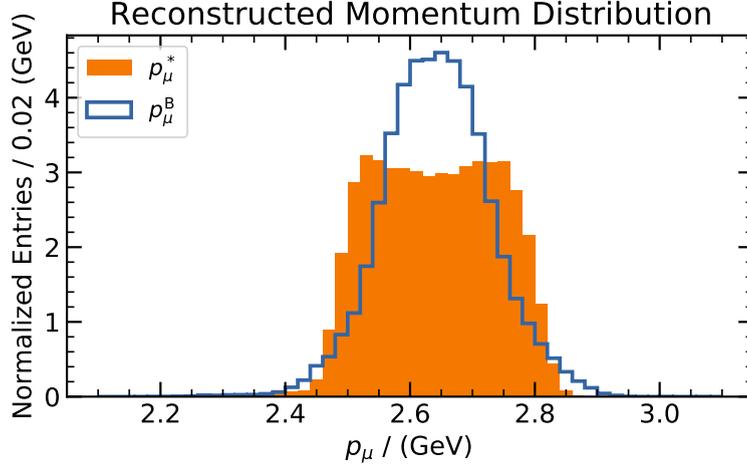


Figure 4.1.: The reconstructed muon momentum in the center-of-mass reference frame  $p_\mu^*$  and the parent B reference frame  $p_\mu^B$  using the reconstructed momentum of the companion B meson with inclusive tagging.

These methods have a trade-off between decreasing information of the companion B due to less stringent constraints on the reconstructed decay and increasing efficiency. The interplay between those tagging methods is illustrated in Figure 4.2. In the full Belle data set, only around 350 signal events are expected. This requires a high reconstruction efficiency. Therefore, for this analysis, the inclusive tagging is chosen due to the absolute requirement of efficiency in the reconstruction.

The hadronic and semileptonic tagging algorithms are explained in [25, 26]. In the following I will go into the details on how an efficient inclusive tag is reconstructed, with whose 4-momentum the lepton momentum in the parent B frame will be calculated.

Inclusive tagging does not impose any specific decay chains of the companion B meson. All leftover tracks are treated under the pion mass hypothesis and all clusters in the electromagnetic calorimeter are treated as photons. To reconstruct the B tag candidate, simply all four-vectors of the pion tracks and photon clusters are combined, except for what is used to reconstruct the signal side. The leftover tracks and clusters in the detector after selecting the tracks and clusters required to reconstruct the signal side are the so-called rest of event (ROE). To improve this inclusive reconstruction, some selections can be performed on the ROE to avoid e.g. double counting from curling tracks or using calorimeter clusters as photons although there is an associated track.

$K_S^0$  candidates are used if they survive the multivariate classification methods from [27]. All tracks originating from a  $K_S^0$  candidate are not used for further selection.

Tracks with a transverse momentum  $p_t < 275$  MeV do not leave the CDC but curl back into the detector. To avoid double counting of those tracks they are combined to form a V0 candidate. When they originate from a common vertex and their track parameters indicate that their momenta are back-to-back, the higher energetic track is kept and the

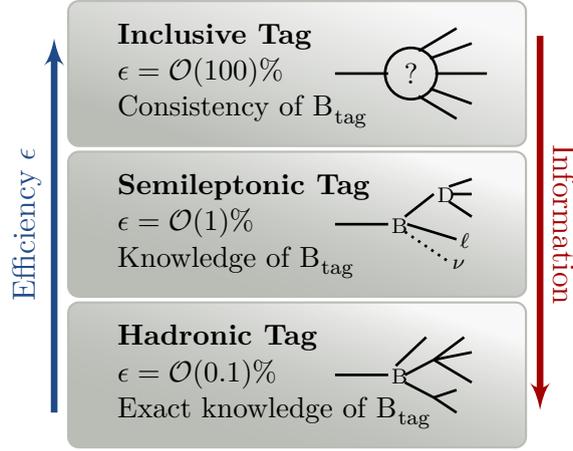


Figure 4.2.: Illustration of the interplay between the different tagging methods. The trade-off is always between information/purity and efficiency. This originates from the constraints on the reconstructed B mesons, e.g. for the hadronic and semileptonic tag candidate a specific decay has to be reconstructed, whereas the inclusive tag candidate is constructed without any requirement on the specific decay. For this analysis, the most important key performance indicator of the tagging variant is efficiency. Figure taken from [25].

lower energetic track is rejected.

Photons are reconstructed from calorimeter clusters where no charged track is located in the proximity.

Particle candidates surviving this selection are used to form a  $B_{\text{tag}}$  candidate.

#### 4.1.1. Inclusive $B_{\text{tag}}$ Reconstruction

After cleansing the ROE from beam remnants and reconstruction artifacts, the remaining tracks and neutral clusters are combined to the inclusive  $B_{\text{tag}}$  candidate. Its four-vector in the center-of-mass frame is given by

$$p_{\text{cms}}^\mu = \begin{pmatrix} \sqrt{\mathbf{p}_{\text{cms}}^2 + m_B^2} \\ \mathbf{p}_{\text{cms}} \end{pmatrix}, \quad (4.1)$$

with  $\mathbf{p}_{\text{cms}} = \sum \mathbf{p}_i \forall \mathbf{p} \in \text{ROE}$ . The momentum magnitude of the four-vector is constrained by the kinematics of the two-body decay  $\Upsilon(4S) \rightarrow B^+B^-$ . This information is used to fix the magnitude of the momentum component  $\mathbf{p}$  to the value of 332 MeV, which yields a much better momentum resolution compared to the reconstructed magnitude of the momentum from the sum of all ROE tracks and clusters. Thus only the direction of the inclusive  $B_{\text{tag}}$  is determined from the reconstructed tracks and clusters.

To further improve the resolution of the inclusive tag candidate, the error of the momentum distribution is studied. There is no information available on the specific decay mode of the tag-side B when using this inclusive approach. Therefore, no information is available

on an event by event basis if any tracks or clusters are missed during the reconstruction or a neutrino is present. However, it is possible to infer information from the statistical distributions of the reconstructed and generated kinematic values. That is when it is likely that a particle was missed. This can be used to improve the resolution of the inclusively reconstructed momentum using a calibration function, what is discussed in the following section.

#### 4.1.2. $B_{\text{tag}}$ Calibration and Optimization

The dominant uncertainty in the inclusive tag reconstruction originates from the fact that the events are boosted in forward direction ( $p_z > 0$ ) and it is likely that charged and neutral particles in forward direction are outside of the Belle detector acceptance. This is reflected by a large bias towards negative  $p_z$ , which can be seen in Figure 4.3 (top row).

Another source for wrongly reconstructed inclusive tags are due to neutrinos being present in the ROE. Neutrinos can not be reconstructed and therefore the fraction of the total reconstructed momentum in the ROE is lost, leading to a biased direction of the inclusive tag momentum vector.

The distribution of the MC truth  $p_z^{\text{MC}}$  in bins of the reconstructed  $p_z^{\text{rec}}$  distribution is shown in Figure 4.4. A clear trend is visible that if the B has a momentum in the forward direction the reconstruction bias becomes larger on average. To correct for this behavior, a calibration function  $f = f(p_z^{\text{rec}})$  is used which maps the reconstructed  $p_z^{\text{rec}}$  momentum onto the mean of the MC truth  $p_z^{\text{MC}}$  distribution. To create a smooth calibration function, a smoothed spline fit is performed on the extracted means per reconstructed  $p_z^{\text{rec}}$  bin. The functional dependency is shown in Figure 4.5.

After reconstructing the inclusive tag momentum, this calibration is applied as

$$\begin{aligned} p_z^{\text{cor}} &= f(p_z^{\text{rec}}), \\ p_t^{\text{rec}} &= \sqrt{p^2 - (p_z^{\text{cor}})^2}, \end{aligned} \quad (4.2)$$

to reconstruct the  $B_{\text{tag}}$  3-momentum. All other kinematic variables can be corrected using these new momentum components. The momentum magnitude  $p$  and the azimuth angle  $\phi$  are agnostic with respect to the calibration. A summary of the kinematic variables of the MC truth, reconstructed, and calibrated inclusive  $B_{\text{tag}}$  including the residuum distributions is shown in Figure 4.3.

After the calibration is applied, the magnitude of the error vector

$$D = \left| \mathbf{p}^{\text{rec}} - \mathbf{p}^{\text{MC}} \right| \quad (4.3)$$

can be calculated. For perfectly reconstructed tag candidates the error is expected to be  $D = 0$  GeV. Therefore, it is desired that the mean of the distribution should also converge to zero. The parameter  $\xi$  is introduced to modify the momentum magnitude

$$D = \left| \xi \mathbf{p}^{\text{rec}} - \mathbf{p}^{\text{MC}} \right| \quad (4.4)$$

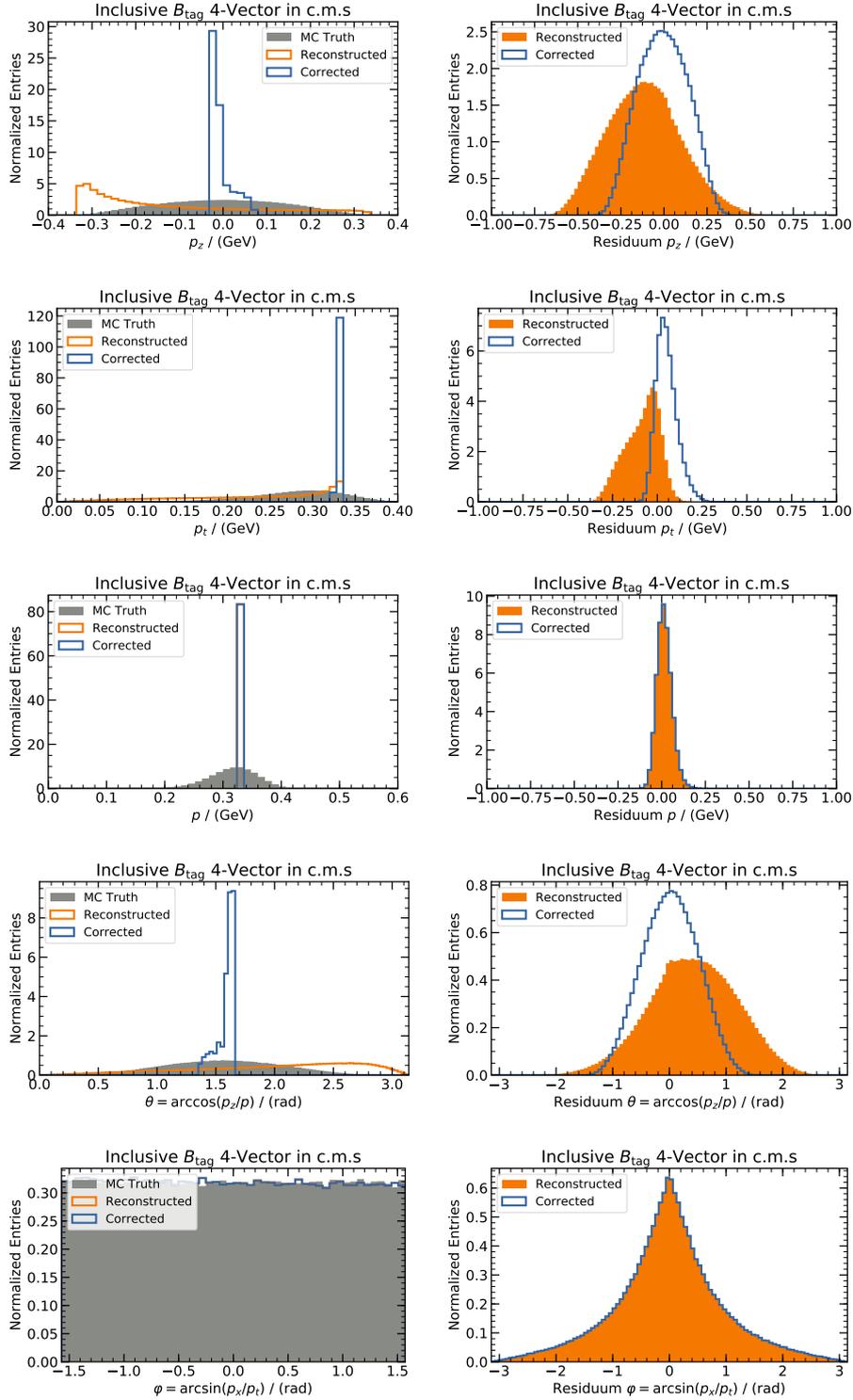


Figure 4.3.: Kinematic variables of the inclusive tag four-vector extracted from the simulation, reconstructed, and after the calibration is applied applied (Left). Residuum distributions (Right) calculated from the distributions shown on the left. Note that the calibration function only corrects  $p_z$  and with it  $p_t$  and the polar angle  $\theta$ . The magnitude and the azimuth angle remain the same.

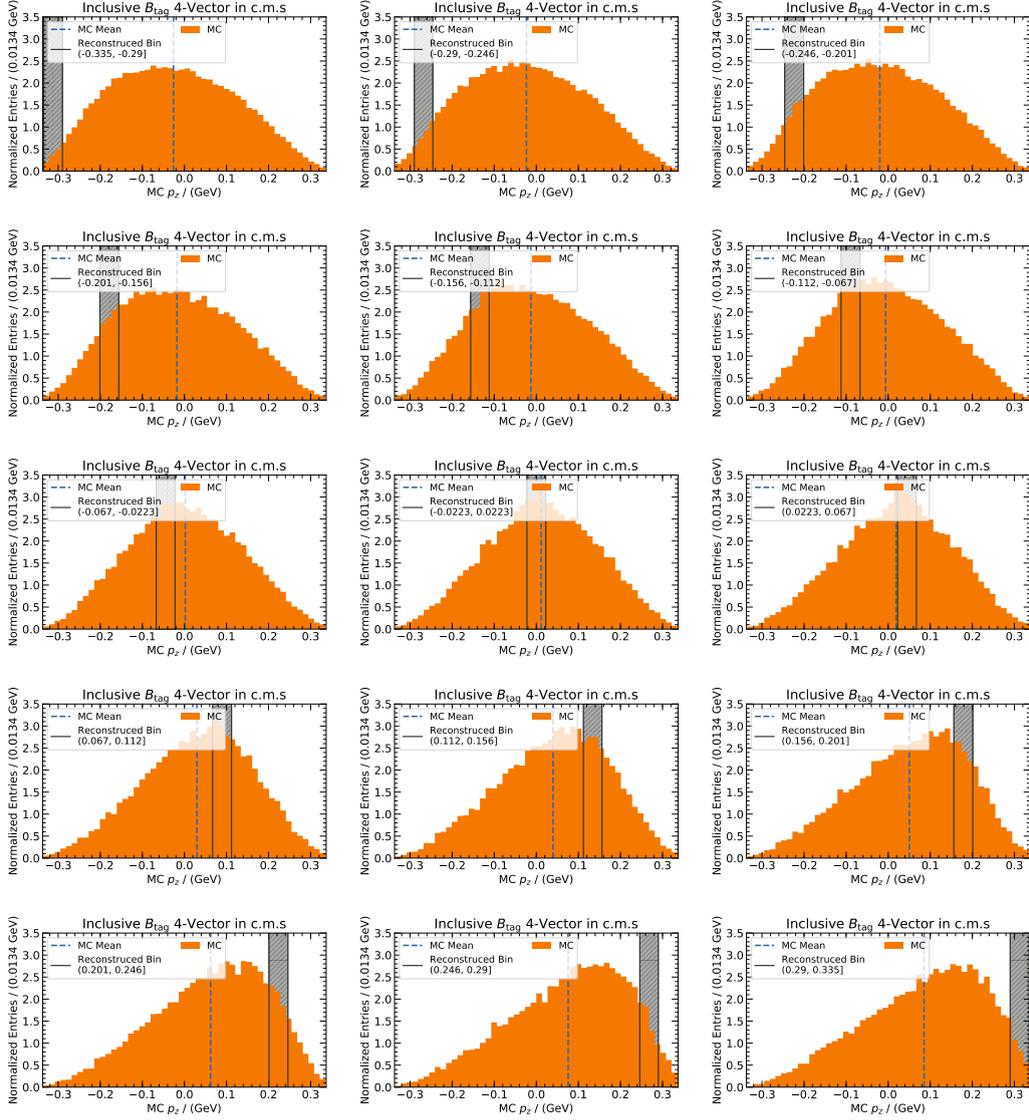


Figure 4.4.: The individual bins used to create the correction function for  $p_z$ . A clear trend of the reconstruction bias is visible. The MC truth distribution skews with increasing reconstruction bias. For a perfect reconstruction of the inclusive tag, the orange distribution would be located in the window of the reconstructed  $p_z$  bin.

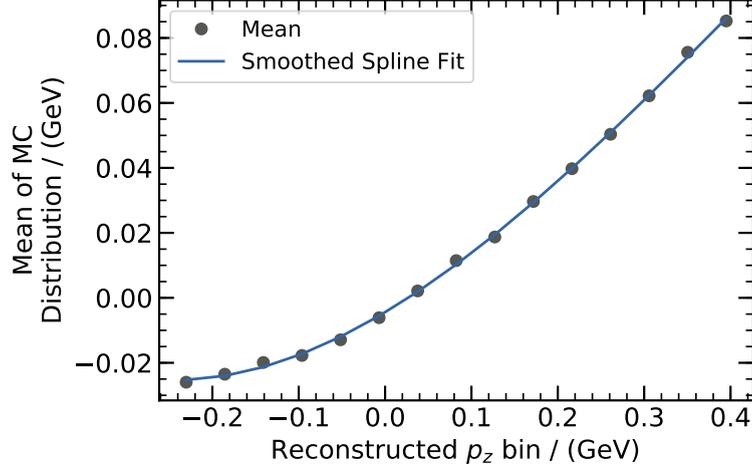


Figure 4.5.: The calibration function of the inclusive tag vector determined on signal MC. Each data point represents the mean of the MC distribution in a bin of the reconstructed  $p_z$  distribution. The blue line is the result of a smoothed spline fit onto the determined means in the reconstructed bins to get a smooth calibration function.

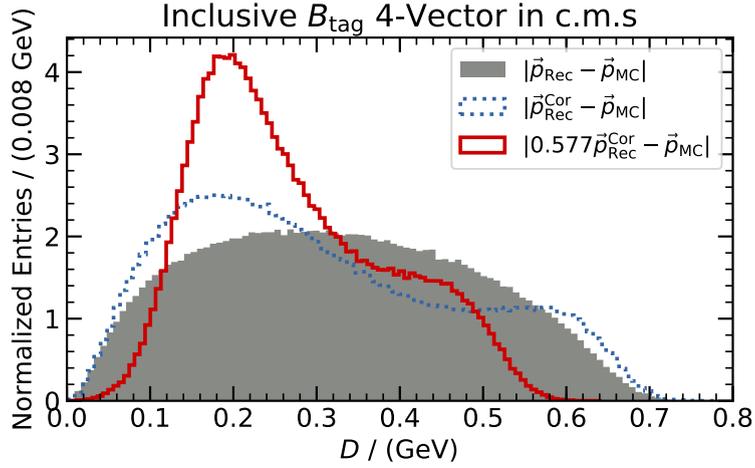


Figure 4.6.: The average error vector for the reconstructed inclusive tag momentum. The gray distribution is without the calibration function applied, the blue distribution with the calibration function applied and the red distribution after the absolute momentum has been corrected with  $\xi = 0.577$  to minimize the mean of the error vector distribution.

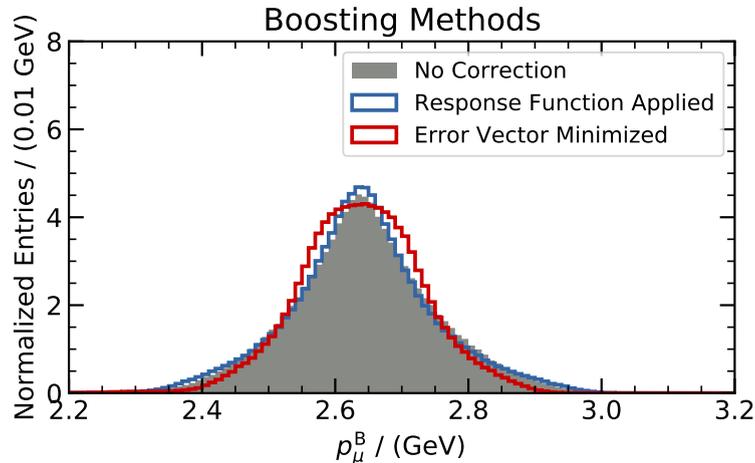


Figure 4.7.: The signal variable after boosting with an uncalibrated inclusive tag vector (gray), with the calibration function applied (blue), and with the average error vector minimized (red).

and is optimized so that the mean of the distribution becomes minimal. The minimum mean is found for  $\xi = 0.577$ . The distribution of  $D$  is shown for reconstructed and calibrated, with and without optimized  $\xi$ , inclusive tags in Figure 4.6.

The effects of the different stages of the inclusive  $B_{\text{tag}}$  candidate reconstruction on the lepton momentum in the  $B_{\text{sig}}$  rest frame are shown in Figure 4.7. The calibration improves the resolution of the lepton momentum per reconstructed  $p_z$  bin whereas the optimization of  $\xi$  modifies the distribution such that the mean error becomes minimal.

With the four-vector of the tag side B meson known, the lepton candidate can be boosted into the center-of-mass frame of the signal side B meson. This yields the lepton momentum  $p_{\mu}^B$  and the angle between the lepton momentum in the  $B_{\text{sig}}$  frame and the momentum of the  $B_{\text{sig}}$  in the center-of-mass frame  $\cos \Theta_{B\mu}$ .

## 4.2. The Signal Selection Classifier

The signal selection is based on a multivariate classification which includes a data vs. MC reweighting to improve the description of the continuum components in the data. Boosted decision trees [28] are chosen as the multivariate method used in the following.

### 4.2.1. Input Variables

The input space for the multivariate method is given two types of variables: Variables describing the topology of the decay and variables calculated on the rest of event (ROE). The difference in event topology between the resonant production of the  $\Upsilon(4S)$  and the non-resonant production is illustrated in Figure 4.8.

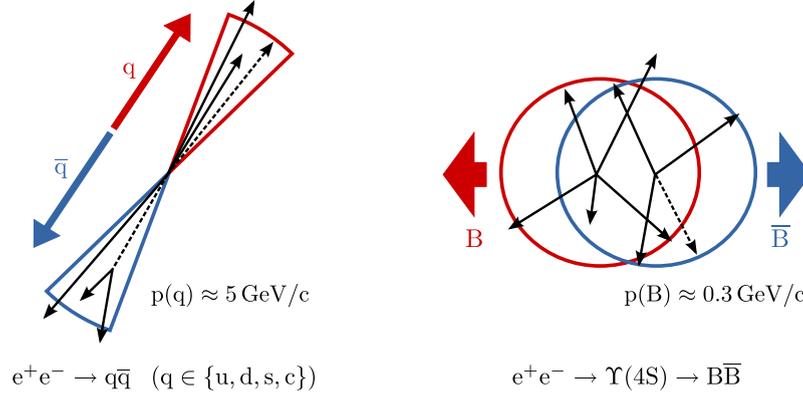


Figure 4.8.: The difference in event topology for resonant and non-resonant interactions in the center-of-mass reference frame. (left) Continuum event. (right)  $\Upsilon(4S)$  event. In the case of a continuum event, the momenta are distributed back-to-back, whereas in the case of the  $\Upsilon(4S)$  event the B mesons, created in the decay of the  $\Upsilon(4S)$ , are almost at rest. The momenta of the B meson decay products are isotropically distributed. The difference in these two event topologies can be quantified with e.g. the Cleo Cones. Figure adapted from [29].

There are several concepts to quantify the difference in the event shape of continuum events and  $\Upsilon(4S)$  decays, which can be used for a topological discrimination of the two. They are discussed in [3] and briefly summarized in the following. Each event consists of a set of  $N$  particles with momenta  $\mathbf{p}_i$ , with  $i \in \{1, 2, \dots, N\}$ .

### Thrust

The thrust  $T$  is defined as as

$$T = \frac{\sum_{i=1}^N |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_{i=1}^N |\mathbf{p}_i|}, \quad (4.5)$$

with the thrust axis  $\mathbf{T}$ , which is defined as the unit vector along which the projection of all momenta is maximal. The thrust takes values between  $1/2$  and  $1$  with a continuum event corresponding to  $T \rightarrow 1$  and an  $\Upsilon(4S)$  event corresponding to  $T \rightarrow 1/2$ .

### $\cos \theta_B$

The angle between the momentum of the reconstructed B meson and the beam axis is  $\cos \theta_B$  and  $1 - \cos^2 \theta_B$  distributed. This distribution originates from the spin  $1 \rightarrow 0 \ 0$  decay of the  $\Upsilon(4S)$ . For continuum events, the distribution is flat, because the B-candidate is created from random combinations of tracks.

### Cleo Cones

The Cleo Cones are defined along the thrust axis with opening angles of  $\Theta \in [\theta, \theta + 10]$  deg. The value of Cleo Cone  $i$  is the total momentum flow of all particles within given cone  $i$ . For continuum events the momentum flow is clustered in the Cleo Cones with small opening angles.

### Fox Wolfram Moments

The Fox Wolfram moments describe the phase-space distribution of energy and

momenta in an event. The  $k$ -th Fox Wolfram moment is defined as

$$H_k = \sum_{i,j}^N |\mathbf{p}_i| |\mathbf{p}_j| P_k(\cos \theta_{ij}), \quad (4.6)$$

where  $\theta_{ij}$  is the angle between  $\mathbf{p}_i$  and  $\mathbf{p}_j$  and  $P_k$  is the  $k$ -th order Legendre polynomial. The reduced Fox-Wolfram moments are given by  $R_k = \frac{H_k}{H_0}$ .

**Super Fox Wolfram Moments** The Super Fox Wolfram moments are calculated by modifying Equation (4.6). Instead of summing over all detected particles, the moments are calculated with respect to the particles belonging to the reconstructed B meson (labeled s) and the rest of event (labeled o). The possible combinations are  $R_k^{oo}$ ,  $R_k^{so}$  and  $R_k^{ss}$ . From the possible combinations only those are used which are uncorrelated to the beam constrained mass  $m_{bc}$ , the missing energy  $\Delta E$  and possible other variables of interest.

In theory the rest of event should form a B meson. Therefore, the reconstructed invariant mass should peak at the nominal B mass as well as the reconstructed missing energies should be zero. Usually the beam constrained mass and the beam constrained missing energy are used:

#### $\Delta \hat{E}^{\text{ROE}}$

The normalized missing energy is calculated with respect to the beam energy  $\Delta \hat{E}^{\text{ROE}} = (E_B - E^*)/E^*$ , with  $E^* = E_{\text{beam}}/2$ . If no neutrino is in the ROE, no missing energy is expected.

#### $\hat{m}_{bc}^{\text{ROE}}$

The beam constrained mass of the B mesons is calculated with respect to the beam energy and the reconstructed momentum in the center-of-mass system  $\hat{m}_{bc}^{\text{ROE}} = \sqrt{(E^{*2} - \mathbf{p}_B^2)/E^{*2}}$ , with  $E^* = E_{\text{beam}}/2$ . The beam constrained mass has a better resolution compared to the invariant mass of the reconstructed B meson, because the beam energy is known precisely.

The variables have been normalized to the beam energy. This allows to use the same variables for off- and on-resonance data and MC.

Additionally, two counting variables are deployed in the training:

#### $n_{\text{tracks}}^{\text{ROE}}$

The number of tracks in the rest of event  $n_{\text{tracks}}^{\text{ROE}}$  is the sum over all tracks surviving the selection cuts.

#### $n_{\text{Leptons}}$

The number of leptons in the event  $n_{\text{Leptons}}$  is given by the number of tracks with an associated muon or electron identification value above 0.9.

### 4.2.2. MVA-Training

Loose selection cuts were performed to reduce the data sample without signal loss before the training. No useful information can be extracted from regions of the input space where only background is present. The applied cuts are tabulated in Table 4.1.

Table 4.1.: Loose selection cuts on the reconstructed data sample before the training of the multivariate classifier. The cuts are chosen to suppress background while retaining approximately 100% signal efficiency.

Description	Selection Cut
Number of tracks in the ROE	$n_{\text{tracks}}^{\text{ROE}} > 2$
Normalized beam-constrained mass	$\hat{m}_{\text{bc}}^{\text{ROE}} > 0.964$
Normalized missing energy	$-0.5 < \Delta \hat{E}^{\text{ROE}} < 0.1$
Number of leptons in the event	$n_{\text{Leptons}} < 3$
Reduced second Fox-Wolfram moment	$R_2 < 0.5$

The classifier method uses a high dimensional input space to discriminate between signal and background events. Mis-modeling of this input space can lead to large discrepancies in performance when applied to data or MC. To improve the description of the continuum component in the MC an additional classifier was trained following the idea from [30]. Using the provided MC samples and the signal free off-resonance data sample, a classifier is trained to discriminate between data and MC events using the input variables to the classifier and the lepton momentum. If the MC is well modeled, the classifier will not be able to discriminate between both, however if there are discrepancies the classifier will learn them. The classifier output can be transformed to reweight the MC events to cancel the shape differences between data and MC using. The weights are defined as

$$w = \frac{p}{1-p}, \quad (4.7)$$

where  $p \in [0, 1]$  is the classifier output and  $w$  is the weight assigned to the MC. Classifier outputs  $p \rightarrow 0$  and  $p \rightarrow 1$  indicate large discrepancies resulting in a weight  $w \rightarrow \infty$ . In comparison, a classifier output of  $p = 0.5$  indicates no difference between data and MC. The associated weight for this case is  $w = 1$ , which shows that the actual MC does properly describe the data and does not need to be modified. With this method it is also possible to retrieve the proper normalization of the continuum contribution from the off-resonance data sample. A distribution of weights is shown in Figure 4.9.

The effect of the reweighter is shown in Figure 4.10 for three exemplary input variables variables and the classifier output. The systematic uncertainty originating from the reweighting is discussed in Section 7.2.

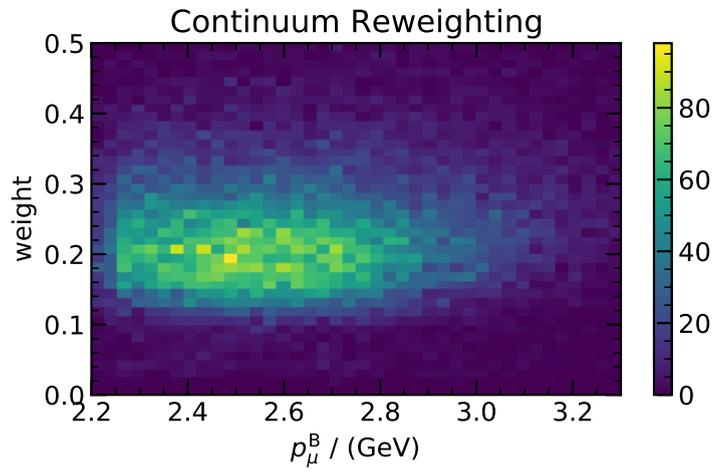


Figure 4.9.: Weights determined vs the lepton momentum. The weights are determined by a training with a MC sample size five times the integrated luminosity of the data sample. This allows to get the overall normalization for the full continuum MC sample and the shape correction from the training.

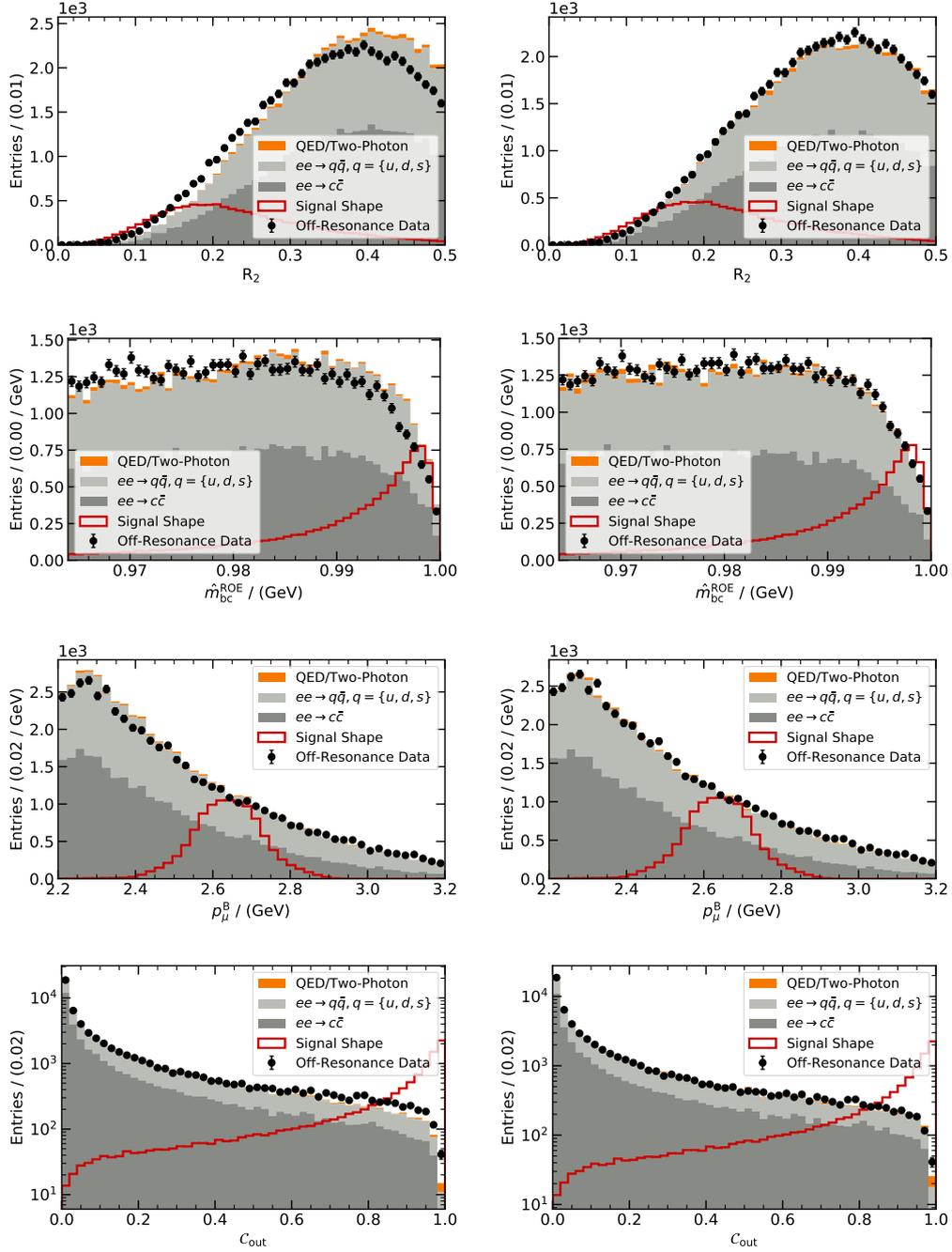


Figure 4.10.: The effect of the data-MC reweighting for three exemplary variables. (left) Before the reweighting is applied and (right) after the reweighting is applied. The improvement ensures that the training of the multivariate classifier is performed on a sample which reflects the true data distribution. The effect is particularly visible in the classifier output (bottom row). The signal shape is only shown for illustrative purpose.



## 5. Control Channel $B \rightarrow D[\rightarrow K\pi]\pi$

The decay channel  $B^+ \rightarrow \bar{D}^0[\rightarrow K^+\pi^-]\pi^+$  and its charge conjugate is used as control channel for this analysis. It has a branching ratio of  $\mathcal{B}(B^+ \rightarrow \bar{D}^0[\rightarrow K^+\pi^-]\pi^+) = 1.82 \cdot 10^{-4}$ , a pure hadronic final state and can be used to mimic the signal decay if the  $\bar{D}^0$  is removed. This allows to validate the consistency of the inclusive tag as well as the resolution of the muon momentum and the efficiency of the signal selection classifier.

The pure hadronic nature of the decay allows for very stringent cuts on the kinematic variables of the involved particles, which allows for a control sample with high purity. In this chapter, the  $\pi^+$  is referred to as the signal track candidate. The selection cuts used to reconstruct the control channel are tabulated in Table 5.1. The variables used are:

**dr, dz** The impact parameters of the signal track candidate with respect to the reconstructed interaction point.

**$p^*$**  The signal track momentum in the center-of-mass frame.

**$\mathcal{P}_{K/\pi}$**  The Kaon over Pion probability, which is calculated from the Likelihood ratio  $\mathcal{L}_K/\mathcal{L}_\pi$  provided by the particle identification detectors.

**$\mathcal{P}_{\pi/K} = 1 - \mathcal{P}_{K/\pi}$ .**

Table 5.1.: Selection cuts on the candidates used in the control channel reconstruction. The semantics of  $x : Y$  is given by particle  $x$  and  $x'$  are used to reconstruct the intermediate particle  $Y$ .

Variable	$\pi^+ : D^0$	$K^- : D^0$	$D^0 : B^-$	$\pi^+ : B^+$	$B^+$
dr / (cm)	< 0.5	< 0.5	-	< 0.5	-
dz  / (cm)	< 2.0	< 2.0	-	< 2.0	-
$p^*$ / (GeV)	> 0.3	> 0.3	-	> 2.1	-
$\mathcal{P}_{K/\pi}$	-	> 0.6	-	-	-
$\mathcal{P}_{\pi/K}$	> 0.6	-	-	> 0.6	-
dM  / (GeV)	-	-	< 0.05	-	< 0.05
\Delta E  / (GeV)	-	-	-	-	< 0.2
$m_{bc}$ / (GeV)	-	-	-	-	> 5.2

Table 5.2.: The momentum resolution determined on data and MC for different observables. No significant difference can be found between the determined resolutions.

	Variable	$\mu \pm \sigma_\mu / (\text{GeV})$	$\sigma \pm \sigma_\sigma / (\text{GeV})$	Median / (GeV)
MC	$\Delta p_\pi^B$	$0.0119 \pm 0.0004$	$0.1108 \pm 0.0003$	0.0025
Data	$\Delta p_\pi^B$	$0.0123 \pm 0.0004$	$0.1113 \pm 0.0003$	0.0033

**dM** The  $D^0$  and  $B^+$  mesons are reconstructed from their respective daughter particles. The mass difference dM is given by the reconstructed invariant mass from the daughter particles to the nominal masses of the mesons.

**$\Delta E$**  The missing energy is calculated with respect to the beam energy  $\Delta E = E_B - E^*$ , with  $E^* = E_{\text{beam}}/2$ . As the decay does not include any neutrinos in the final state, no missing energy is expected.

**$m_{bc}$**  The beam constrained mass of the B mesons is calculated with respect to the beam energy and the reconstructed momentum in the center-of-mass system  $m_{bc} = \sqrt{E^{*2} - \mathbf{p}_B^{*2}}$ , with  $E^* = E_{\text{beam}}/2$ . The beam constrained mass has a better resolution compared to the invariant mass of the reconstructed B meson, because the beam energy is known precisely.

## 5.1. Inclusive Tag Reconstruction

The calibration of the inclusive tag momentum described in Section 4.1 depends on parameters extracted from the simulation. To validate that this information does not affect the reconstruction by introducing e.g. a bias, the calibration method is validated using the control channel.

The decay of the  $\Upsilon(4S) \rightarrow B\bar{B}$  is a two body decay and therefore the momenta are back-to-back in the  $\Upsilon(4S)$  reference frame. In Figure 5.1 the distribution of the reconstructed angle  $\cos\Theta_{B\bar{B}}$  between the  $B_{\text{tag}}$  and  $B_{\text{sig}}$  is shown. As expected, the distribution peaks at 180 deg but the  $B_{\text{tag}}$  is reconstructed with an error on the direction of the momentum up to 180 deg with respect to its companion  $B_{\text{sig}}$  momentum.

By using a decay with a full hadronic signal side, two methods are available to reconstruct the boost vector to boost the signal candidate into the  $B_{\text{sig}}$  reference frame:

**Inclusive tag-side reconstruction** via using the inclusively reconstructed  $B_{\text{tag}}$  candidate. The boost vector into the  $B_{\text{sig}}$  reference frame is determined from the reconstructed four-vector of the  $B_{\text{tag}}$ .

**Exclusive signal-side reconstruction** where the the boost vector into the  $B_{\text{sig}}$  reference frame is determined from the reconstructed four-vector of the  $B_{\text{sig}}$  candidate.

The second method allows for a very precise reconstruction of the  $B_{\text{sig}}$  momentum. The resolution of the reconstructed Pion momentum is shown in Figure 5.2 for simulated and reconstructed events and is tabulated in Table 5.2. No significant differences are observed.

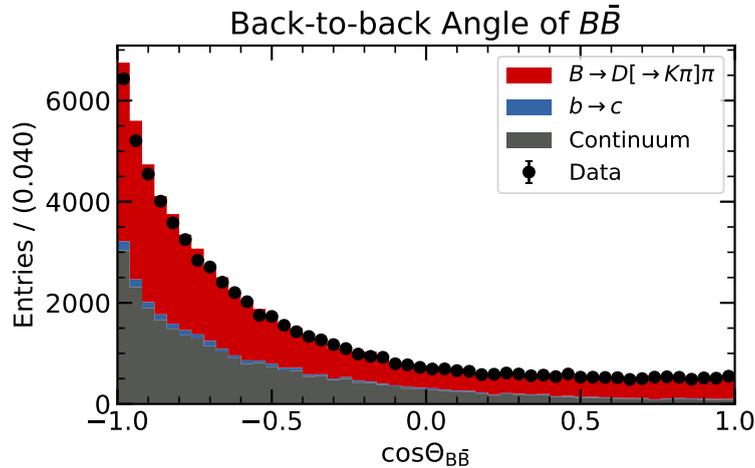


Figure 5.1.: The reconstructed angle between the  $B_{\text{tag}}$  momentum and the  $B_{\text{sig}}$  momentum in the center-of-mass reference frame. The angle  $\cos \Theta_{B\bar{B}}$  is expected to be 180 deg. From the distribution it is visible that the tag momentum is properly reconstructed in most cases, but that the error of the  $B_{\text{tag}}$  momentum direction can be up to 180 deg with respect to the  $B_{\text{sig}}$  momentum.

In addition, the angle between the reconstructed candidate momentum and the  $B_{\text{sig}}$  momentum can be validated. This is shown in Figure 5.3 and no significant difference between the data and MC is observed.

## 5.2. Classifier Output

The response of the classifier described in Section 4.2 can also be validated with the control channel. The topology of the control channel decay and the signal decay are exactly the same when the D meson is treated as a neutrino and removed from the classifier input. The difference is only given by the smaller track momentum of the pion in comparison to the muon in the signal decay  $B \rightarrow \mu\nu_\mu$ . The classifier response of the control channel is shown in Figure 5.4.

In the relevant region of the classifier output used for the signal extraction, the efficiency of the classifier can be determined on reconstructed and simulated events. The determined efficiencies agree within their statistical uncertainties. The efficiencies are summarized in Table 5.3 in the categories defined in Section 6.6.

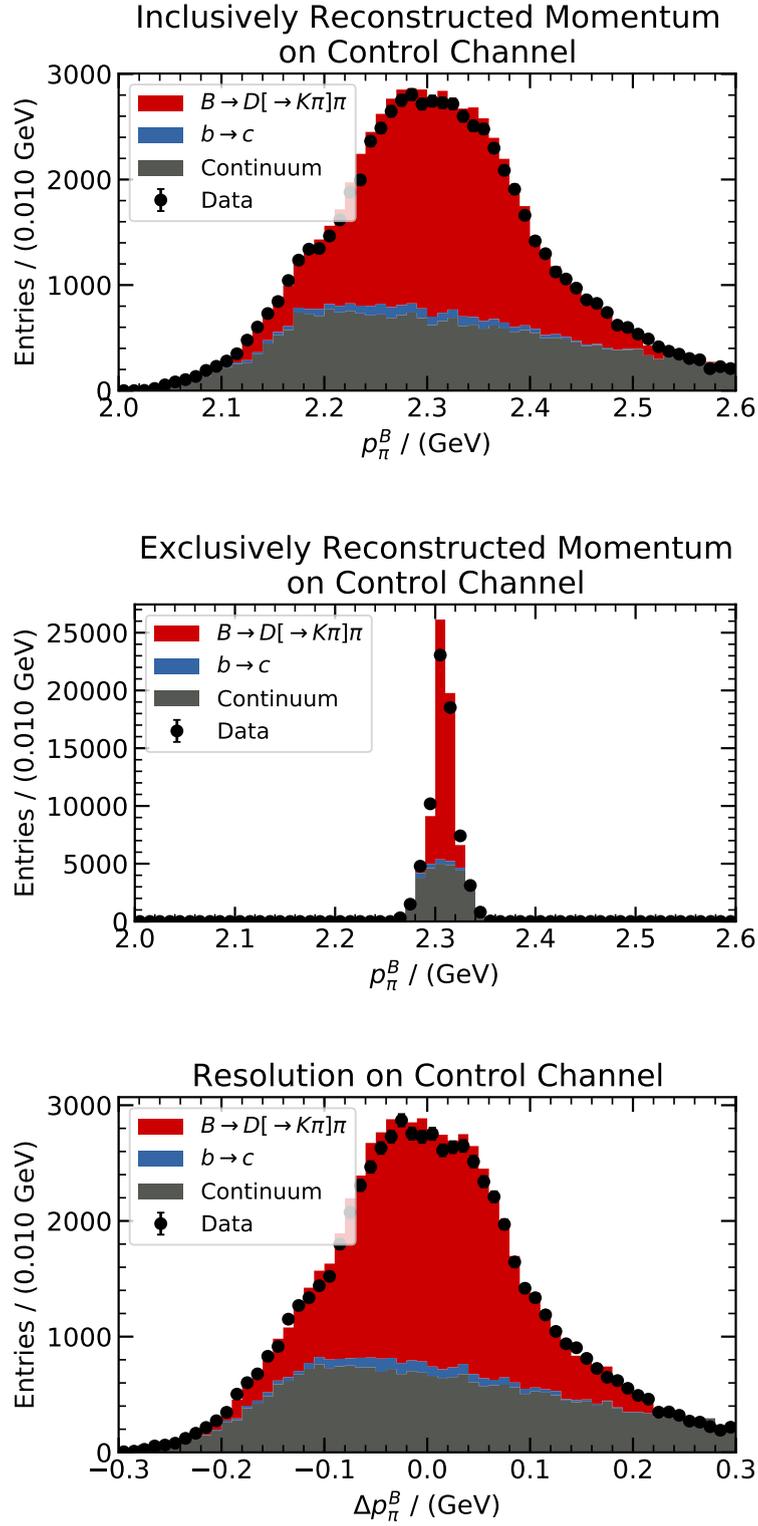


Figure 5.2.: The reconstructed Pion momentum in the signal B frame. The top figure uses the boost vector retrieved from the inclusively reconstructed tag side and the middle figure uses the boost vector retrieved from the exclusively reconstructed signal side. The residuum between both methods is shown in the lower figure. This also allows to determine the momentum resolution purely on data.

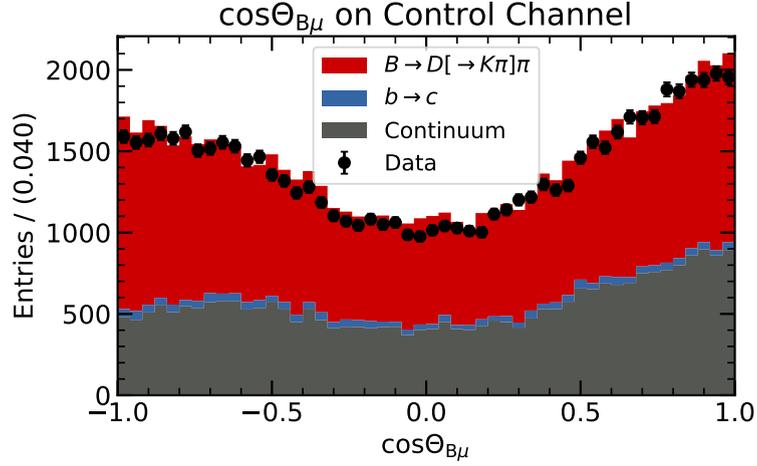


Figure 5.3.: The reconstructed angle between the Pion momentum in the  $B_{\text{sig}}$  frame and the momentum vector of the  $B_{\text{sig}}$  in the center-of-mass frame  $\cos \Theta_{B\pi}$ .

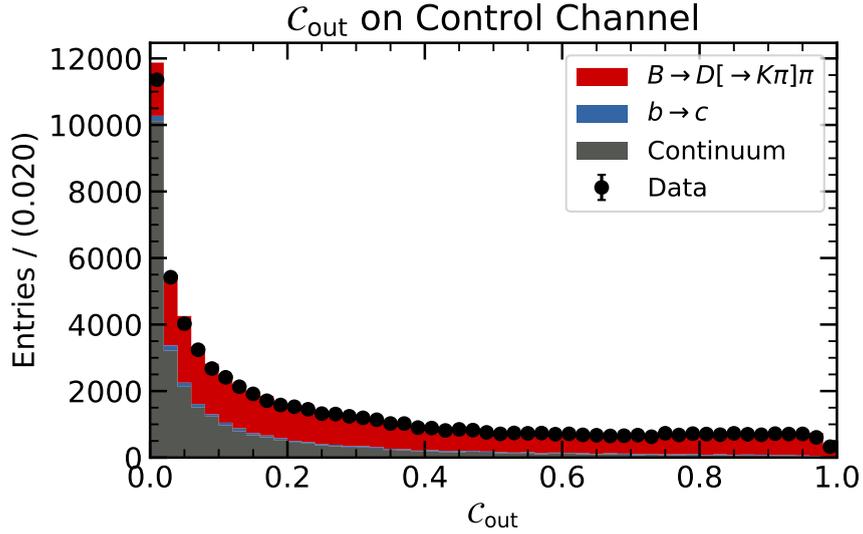


Figure 5.4.: The classifier output determined on the control sample.

Table 5.3.: MVA efficiency on control channel. For categorization of the regions see Section 6.6. The efficiencies on data and MC are compatible within their uncertainties.

Type	Category I-IV	Category I+II	Category III+IV
$\epsilon^{\text{Data}}$	$0.0296 \pm 0.0006$	$0.0047 \pm 0.0003$	$0.0239 \pm 0.0006$
$\epsilon^{\text{MC}}$	$0.0303 \pm 0.0006$	$0.0051 \pm 0.0003$	$0.0252 \pm 0.0006$



## 6. Signal Extraction

The extraction of the number of signal events on the data sample is performed using the method of maximum Likelihood on binned data. The Likelihood function describes the agreement between the measured data and the underlying model as a function of model parameters. Minimizing the negative Likelihood function with respect to the model parameters allows one to estimate their values.

Due to the lack of analytical descriptions of the background components, a template fit is performed. The template PDF's are generated from simulated events with a granularity of 50 MeV bins of the reconstructed lepton momentum in the  $B^{\text{sig}}$  reference frame  $p_{\mu}^B$ .

The composition of the individual templates is explained in the following. The templates are shown in Figure 6.1 in four different signal categories, which are introduced in Section 6.6.

**Signal:** The signal template is generated from a dedicated MC sample of  $2 \cdot 10^5$  events.

**Continuum:** The continuum template consists of three individual components which are composed of: 6 times the integrated luminosity of  $e^+e^- \rightarrow q\bar{q}$  processes, 5 times the integrated luminosity of  $e^+e^- \rightarrow \ell\bar{\ell}$  processes and 5 times the integrated luminosity of  $e^+e^- \rightarrow f\bar{f}f'\bar{f}'$  processes. The combined template is dominated by processes with quarks in the final state.

**b  $\rightarrow$  c:** This sample consists of 10 times the integrated luminosity and mainly composed of  $B \rightarrow D^{(*)}\ell\nu$  decays.

**b  $\rightarrow$  u:** This template consists of 10 times the integrated luminosity. The construction of this MC sample is explained in detail in Chapter 11.

**B  $\rightarrow$   $\mu\nu\gamma$ :** This template is generated from a dedicated MC sample of  $10^7$  events. The branching ratio is set to match the latest measurement by the Belle Collaboration [13]. The normalization is fixed in the fit but the template is allowed to vary within its measured uncertainties.

**Rare:** The template consists of 50 times the integrated luminosity of rare processes, such as flavour changing neutral current decays with  $b \rightarrow s$  transitions, which are suppressed in the SM. The normalization is held fixed in the fit but the template is allowed to vary within its statistical and systematic uncertainties.

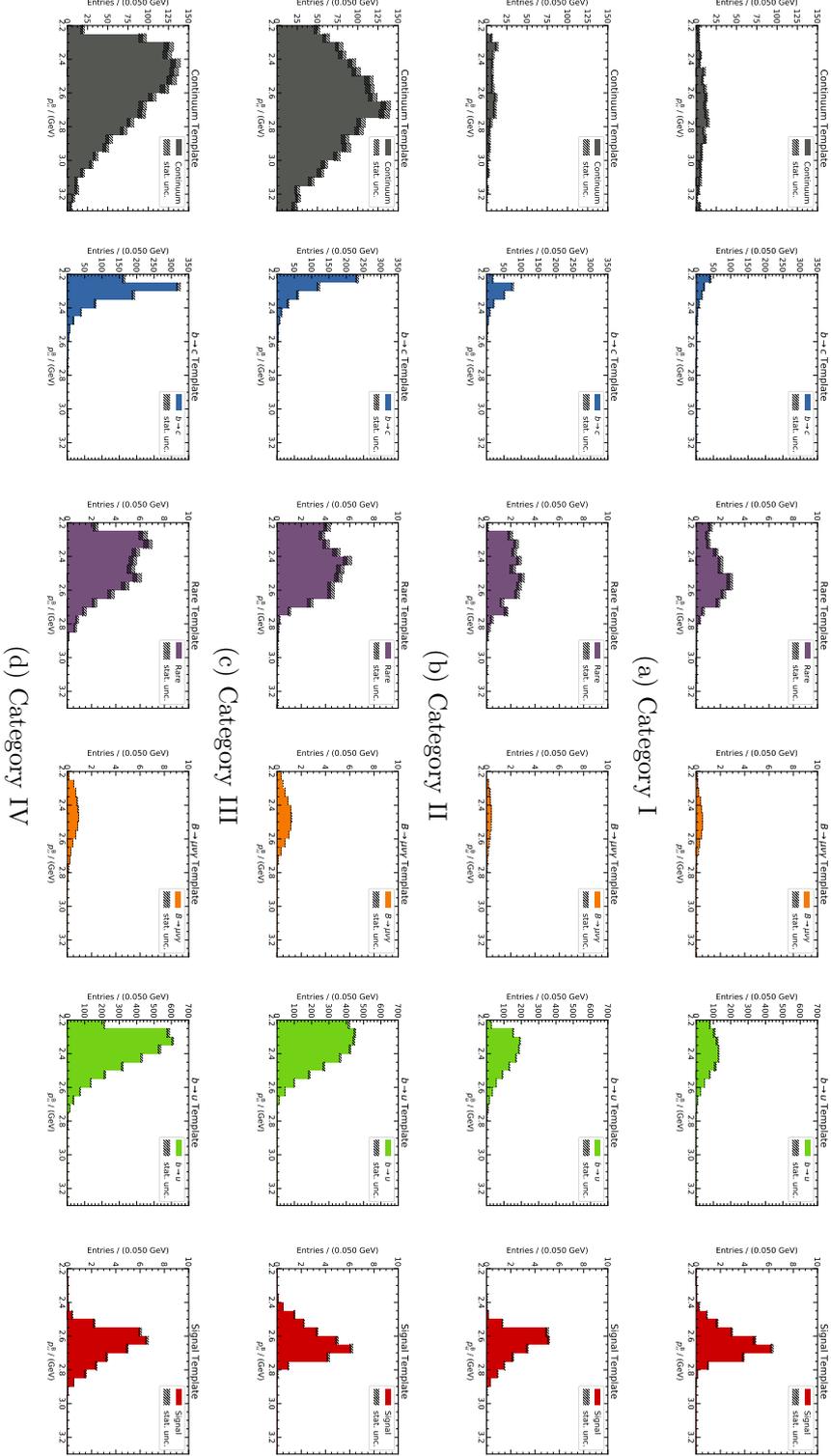


Figure 6.1.: The template PDF's determined on simulated data in the four categories which maximize the Likelihood ratio  $\mathcal{L}_{S+B} / \mathcal{L}_B$ . The error bars show only the statistical uncertainty due to limited MC statistics. The templates including systematic uncertainties are discussed in Chapter 7.

The Likelihood over all signal categories is constructed as

$$\mathcal{L}(\nu) = \prod_i^{\text{bins}} \mathcal{P}(n_i | \nu_i(\theta)) \times \prod_k^{\text{templates}} \mathcal{N}(\theta_k | 0, \Sigma_k), \quad (6.1)$$

with  $\mathcal{P}$  denoting the Poisson distribution,  $n_i$  the number of measured data events in bin  $i$ , and  $\nu_i$  the total number of expected events in bin  $i$ . Further,  $\mathcal{N}$  is the multivariate normal distribution with zero mean and the correlation matrix  $\Sigma_k$  of the statistical and systematic uncertainties for template  $k$ . The multivariate normal distribution imposes Gaussian constraints on the nuisance parameters with the correlations and uncertainties between the lepton momentum bins.

The number of expected events  $\nu_i$  in a given bin  $i$  is

$$\nu_i = \sum_k^{\text{templates}} f_{ik} \eta_k, \quad (6.2)$$

where  $\eta_k$  is the total number of events from template  $k$  and  $f_{ik}$  is the fraction of events of in bin  $i$  in template  $k$ . This fraction is given by

$$f_{ik} = \frac{\eta_{ik}(1 + \theta_{ik}\epsilon_{ik})}{\sum_j^{\text{bins}} \eta_{jk}(1 + \theta_{jk}\epsilon_{jk})}, \quad (6.3)$$

where  $\eta_{ik}$  is the number of expected events from template  $k$  in bin  $i$ ,  $\epsilon_{ik}$  is the relative uncertainty of template  $k$  in bin  $i$  and  $\theta_{ik}$  is a nuisance parameter which allows the fit to pull on the respective uncertainty. Expressing the fraction of events including the nuisance parameters ensures that the overall normalization of the template is not affected by the pull on the systematic uncertainties.

The Likelihood depends on the 6 expectation values  $\nu_i$  of the signal and background templates and additional 528 nuisance parameters  $\theta_i$  with Gaussian constraints, which parametrize the systematic and statistical uncertainties of the fit templates. The expectation values of the rare and  $B \rightarrow \mu\nu\gamma$  contributions are fixed, but the introduced nuisance parameters parametrize their uncertainties. The minimization of the Likelihood is performed using Sequential Least Squares Programming [31].

## 6.1. Goodness of Fit

The goodness of fit of a statistical model is a measure of how well the theoretical model is describing the measured data by summarizing the discrepancies between theoretical expected and measured values. In regression analyses a common measure for goodness-of-fit is the  $\chi^2$  per degree of freedom of the fit. In the fit model described above, it is given by

$$\chi^2/\text{dof} = \sum_i^{\text{bins}} \frac{(n_i - \nu_i)^2}{\sigma_i^2} / \text{dof}, \quad (6.4)$$

where  $n_i$  is the number of observed events in bin  $i$ ,  $\nu_i$  is the number of expected events in bin  $i$  and  $\sigma_i$  is the uncertainty of bin  $i$ . The degrees of freedom dof are given by the number of bins minus the number of unconstrained parameters in the fit.

The probability to observe a  $\chi^2$  value as large or larger than  $\chi_{\text{fit}}^2$  is given by the p-value:

$$p = \int_{\chi_{\text{fit}}^2}^{\infty} \chi^2(x|k) dx, \quad (6.5)$$

where  $\chi^2(x|k)$  is the  $\chi^2$  distribution with  $k$  number of degrees of freedom.

## 6.2. Significance

To reject the background-only hypothesis and to report evidence or discovery of the signal decay, the significance of the fit has to be determined. The most powerful test statistic  $\lambda(\nu)$  to reject a null hypothesis  $H_0$  with respect to an alternative hypothesis  $H_1$  is given by the Neyman-Pearson lemma [32]

$$\lambda(\nu) = \frac{\mathcal{L}(\mu_1)}{\mathcal{L}(\mu_0)}, \quad (6.6)$$

with the parameters  $\mu_i$  of model  $i$ , where  $\mu_1 \in H_1$  and  $\mu_0 \in H_0$ . In the asymptotic limit, where the sample size  $n$  approaches infinity, the test statistic

$$q_\mu = -2 \ln \lambda(\mu) \quad (6.7)$$

follows a  $\chi^2$  distribution where the difference in dimensionality between  $H_1$  and  $H_0$  defines the degrees  $k$  of freedom of the distribution [33]. For the following discussion, the difference of dimensionality is one, as the alternative hypothesis has one additional parameter given by the signal yield. With the underlying distribution of the test statistic the p-value can be calculated as

$$p = \int_{\chi_{\text{obs}}^2}^{\infty} \chi^2(x|k=1) dx, \quad (6.8)$$

which is translated into the number of standard deviations  $\sigma$  via

$$p \stackrel{!}{=} 1 - \int_{-\sigma}^{\sigma} \mathcal{N}(x|0, 1) dx. \quad (6.9)$$

The measured branching ratio has to be between 0 and 1 to bear any physical meaning. To make a statement about the significance of a positive physical signal, the test statistic  $\lambda(\nu)$  can be modified [34] to:

$$\tilde{q}_0(\mu) = \begin{cases} -2 \ln \lambda(\mu) & \mu \geq 0 \\ 0 & \mu < 0, \end{cases} \quad (6.10)$$

with the distribution of the test statistic  $\tilde{q}_0$  being

$$f(\tilde{q}_0) = \frac{1}{2} \delta(\tilde{q}_0) + \frac{1}{2} \chi^2(\tilde{q}_0|k=1). \quad (6.11)$$

The p-value can again be calculated then by integrating over the underlying distribution of the test statistic

$$p = \int_{\tilde{q}_0^{\text{obs}}}^{\infty} f(\tilde{q}_0) d\tilde{q}_0. \quad (6.12)$$

The two different test statistics answer two different statistical questions:

$\lambda(\mu)$  gives the probability to observe a *positive or negative* measurement as extreme or more extreme as the observation under the assumption of the background only hypothesis  $H_0$ .

$\tilde{q}_0(\mu)$  gives the probability to observe a *positive* measurement as extreme or more extreme as the observation under the assumption of the background-only hypothesis  $H_0$ .

Thus,  $\lambda(\mu)$  provides a two-sided confidence interval whereas  $\tilde{q}_0(\mu)$  states a one-sided confidence interval. The convention for evidence or discovery is the rejection of the background-only hypothesis with  $3\sigma$  or  $5\sigma$  respectively.

### 6.3. Likelihood-Profile

The uncertainty on the extracted parameter of interest can be determined in two ways. Either the Hessian matrix is calculated at the point where the Likelihood is minimized or the Likelihood is profiled in the neighborhood of the minimum. Using the Hessian, one obtains the uncertainties in the Gaussian approximation, while with profiling the Likelihood one can extract asymmetric confidence intervals. All Likelihoods appearing in this thesis are well described by a parabola for which the Gaussian approximation holds with excellent precision.

When profiling the Likelihood for a parameter, e.g.  $\nu^{\text{sig}}$ , the parameter is varied and the Likelihood is maximized for all other parameters given this fixed value of  $\nu^{\text{sig}}$ . This procedure yields the Likelihood profile containing both statistical and systematic uncertainties. To measure the influence of systematic uncertainties on the final result, the Likelihood profile can be performed fixing the nuisance parameters at their best fit values at the maximum of the Likelihood. It is convenient to use

$$\Lambda(\nu) = -2 \ln \mathcal{L}(\nu) - 2 \ln \mathcal{L}_{\min}, \quad (6.13)$$

to illustrate the Likelihood profile. The Likelihood profile for the extracted signal on an Asimov data set with an assumed branching ratio of  $4.26 \cdot 10^{-7}$  for the signal decay  $B \rightarrow \mu\nu_\mu$  is shown in Figure 6.2.

### 6.4. Toy-Experiments and Linearity Test

Toy experiments are a method to test the fit for stability and bias. Toy samples are generated by sampling from a Poisson distribution  $\mathcal{P}$  for each bin with the expectation value given by the MC expectation. The fit is then performed on this toy sample and the parameter of interest is extracted. The pull distribution of the extracted parameter, which is defined as

$$\text{Pull} = \frac{\nu^{\text{fit}} - \nu^{\text{expected}}}{\sigma_\nu}, \quad (6.14)$$

should follow a standard normal distribution. If the distribution has a mean  $\mu \neq 0$  the parameter extraction is biased. If the distribution has a standard deviation  $\sigma \neq 1$  the extracted uncertainty on the parameter is under- or overestimated. The toy study for the signal extraction assuming a SM branching ratio of  $4.26 \cdot 10^{-7}$  for the signal decay  $B \rightarrow \mu\nu_\mu$  is shown in Figure 6.3.

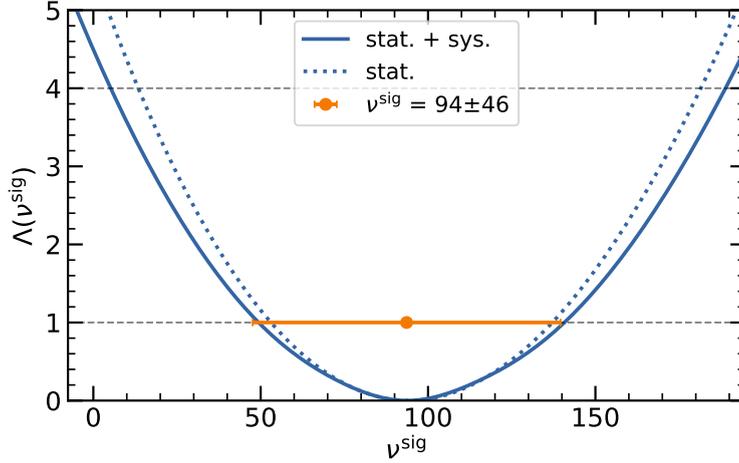


Figure 6.2.: The Likelihood profile for the extracted signal yield in an Asimov fit. The orange marker is the fit result with the uncertainties calculated from the Hessian. The horizontal lines at  $\Lambda = 1$  and  $\Lambda = 4$  define the  $1\sigma$  and  $2\sigma$  confidence intervals. Extracting the  $1\sigma$  confidence interval from the Likelihood profile yields a similar result to extracting the uncertainties from the Hessian. This shows that the Gaussian approximation of the uncertainties is a good approximation. It is also visible that the influence of the systematic uncertainties to the final fit is small. In other words, the measurement is dominated by statistical uncertainties.

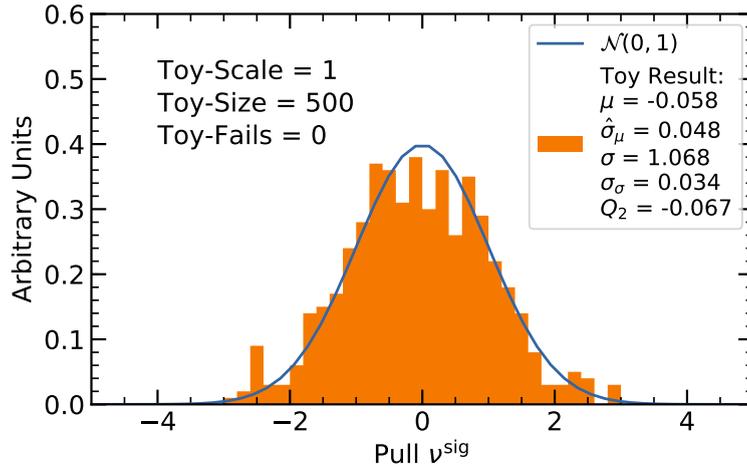


Figure 6.3.: Result of a toy study for the signal extraction. Both the expectation value and the uncertainty are properly estimated by the fit. The legend provides the expectation value and the uncertainty on the expectation value  $\mu \pm \sigma_\mu$ , the standard deviation and the uncertainty on the standard deviation  $\sigma \pm \sigma_\sigma$  and the median  $Q_2$  of the distribution.

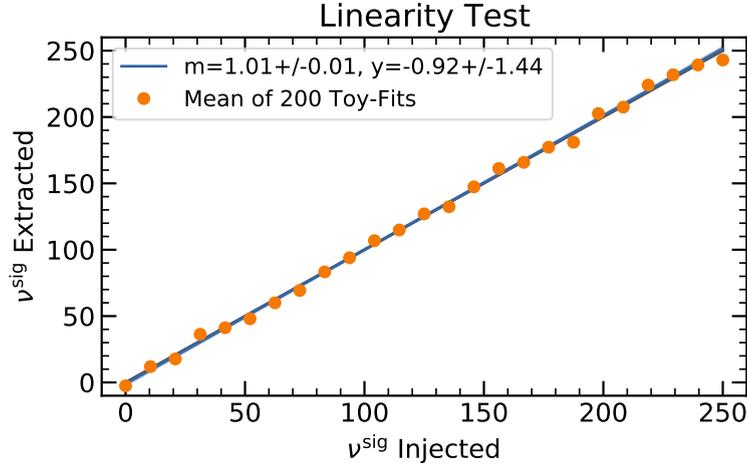


Figure 6.4.: The linearity test for the signal extraction. Each point corresponds to 200 toy experiments. A linear polynomial is fitted to the mean of extracted signal events. The fit yields an unbiased result over a large range of injected signal events.

This procedure can be used to perform a so-called linearity test. It is used to check that the fit is not biased when the parameter of interest deviates strongly from the original expectation. The parameter of interest is scaled to different values and the toy experiments are repeated in the same way with the exception that the original expectation of the parameter is rescaled accordingly. The linearity test for the signal extraction on an Asimov data set is shown in Figure 6.4.

## 6.5. Determination of Upper Limit

If the significance of the observed signal is below  $3\sigma$ , an upper limit on the branching ratio is determined.

The Bayesian upper limit is determined by integrating the properly normalized Likelihood function from 0 to  $x$ , where  $x$  is given by the value when the integral is equal to the desired confidence level. In the approximation of Gaussian uncertainties, the Likelihood function can be written as a Gauss function with the mean equal to the best fit value and the standard deviation equal to the uncertainty determined from the Hesse matrix. For the 90% confidence level, with the prior that the branching ratio is positive, the upper limit  $x$  is then calculated as

$$0.9 = \frac{\int_0^x \mathcal{G}(x' | \mu_{\text{obs}}, \sigma_{\text{obs}}) dx'}{\int_0^\infty \mathcal{G}(x' | \mu_{\text{obs}}, \sigma_{\text{obs}}) dx'} \quad (6.15)$$

The Frequentist upper limit is determined by finding  $x'$  such that

$$0.9 = \int_{\mu_{\text{obs}}}^\infty \mathcal{G}(x' | \mu_A, \sigma_A) dx', \quad (6.16)$$

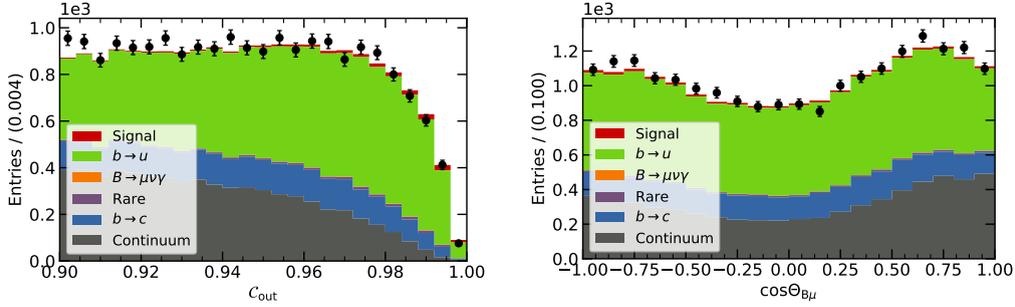


Figure 6.5.: The distributions of  $\mathcal{C}_{\text{out}}$  and  $\cos \Theta_{B\mu}$ . The distributions are split into separate samples to maximize the Likelihood ratio  $\frac{\mathcal{L}_{S+B}}{\mathcal{L}_B}$  when the signal extraction is performed on an Asimov data sample.

with the mean  $\mu_A$  and the standard deviation  $\sigma_A$ , which can be determined on Asimov data sets. The upper limit is thus determined by finding the expected number of events  $\mu_A$ , for which the integral in Equation (6.16) is equal to the confidence level.

The two different limits provide two different statements about the observed result. The Bayesian statement is: Given the observed data and the assumption that there exists a signal, the true signal yield is below the stated upper limit with a probability of at least 0.9. The Frequentist statement is: If the signal yield would have been larger than the stated upper limit, then the process would have been observed with a probability of at least 0.9.

## 6.6. Categorization

The signal extraction was performed in four different mutual exclusive categories. A grid search was performed to search for an optimal split of the signal in the following variables: The classifier output  $\mathcal{C}_{\text{out}}$ , which was introduced in Section 4.2.2, and the angle between the momentum of the  $B_{\text{sig}}$  and the muon  $\cos \Theta_{B\mu}$ , which was introduced in Section 4.1.2. Figure 6.5 shows the variables  $\mathcal{C}_{\text{out}}$  and  $\cos \Theta_{B\mu}$  before the categorization.

For each tuple of possible cut values, the Likelihood ratio  $\frac{\mathcal{L}_{S+B}}{\mathcal{L}_B}$  is determined neglecting systematic uncertainties. The chosen categorization, which was found by maximizing the Likelihood ratio, is given in Table 6.1.

In principle, the grid search can be performed including systematic uncertainties when calculating the Likelihood ratio. However, for each point in the parameter space, the systematic uncertainties would have to be determined anew, resulting in a computationally costly problem. The dominating uncertainty on the final signal extraction originates from the limited number of recorded events. This indicates that omitting the systematics in this optimization has only a small effect on the overall result.

## 6.7. Validation

The fit is performed on different control data samples to validate the signal extraction. The test for closure is performed on an Asimov data sample where the measured data

Table 6.1.: Definition of the four individual categories used for the signal extraction and the two individual categories used in the control region of the classifier sideband. The categories are mutually exclusive.

Category	Classifier Output	Angle
Signal Region		
I	$0.98 < \mathcal{C}_{\text{out}} < 1.00$	$-0.133 < \cos \Theta_{B\mu} < 1.000$
II	$0.98 < \mathcal{C}_{\text{out}} < 1.00$	$-1.000 < \cos \Theta_{B\mu} < -0.133$
III	$0.93 < \mathcal{C}_{\text{out}} < 0.98$	$0.044 < \cos \Theta_{B\mu} < 1.000$
IV	$0.93 < \mathcal{C}_{\text{out}} < 0.98$	$-1.000 < \cos \Theta_{B\mu} < 0.044$
Classifier Sideband Control Region		
I	$0.90 < \mathcal{C}_{\text{out}} < 0.93$	$0.000 < \cos \Theta_{B\mu} < 1.000$
II	$0.90 < \mathcal{C}_{\text{out}} < 0.93$	$-1.000 < \cos \Theta_{B\mu} < 0.000$

is the MC expectation (Section 6.7.1). To validate the proper description of especially the  $B \rightarrow X_u \ell \nu$  and continuum background components the fit is performed in a control region of the classifier output  $\mathcal{C}_{\text{out}} \in [0.90, 0.93]$ , which is enriched by both backgrounds (Section 6.7.2). To check the background description before un-blinding the signal region, the fit was performed in the final selected data sample with the signal region still blinded (Section 6.7.3).

### 6.7.1. Fit on Asimov Data

The Asimov data points correspond to the sum of the individual expectation of all templates in a given bin in each of the categories defined in Table 6.1. The sum of expectation values is rounded to the nearest integer number, because the Poisson distribution is discrete and no well defined analytical continuation for the Poisson distribution is existent. This has a minor effect, but has to be kept in mind when the results are used for closure tests.

The fit result on the Asimov data sample is shown in Figure 6.6. To test the stability of the minimization of the Likelihood function the fit was several times initialized with different starting values. For reasonable starting values, the minimizer always drops into the minimum of the negative log-Likelihood. The goodness-of-fit  $\chi^2/\text{dof} = 1.3/84$  confirms this result. A  $\chi^2 = 0$  is expected when the Asimov data and MC distributions match perfectly, but due to the rounding errors mentioned above a slight discrepancy is found.

In Figure 6.7 the Likelihood profiles and toy experiments for the the fitted expectation values are shown. The Likelihood profiles are stable and the determined uncertainty extracted from the Hessian is a good approximation of the uncertainty extracted from the Likelihood profile. The toy experiments validate that the fitted number of events for each template, especially the signal yield  $\nu^{\text{sig}}$ , can be extracted without bias and with a correctly estimated uncertainty.

The fit does not pull on the nuisance parameters, with exception of the high energy momentum part of the continuum template, as is shown in Figure 6.8. This pull is expected due to the rounding to integer values, which has a large effect when the number of events in a bin are small.

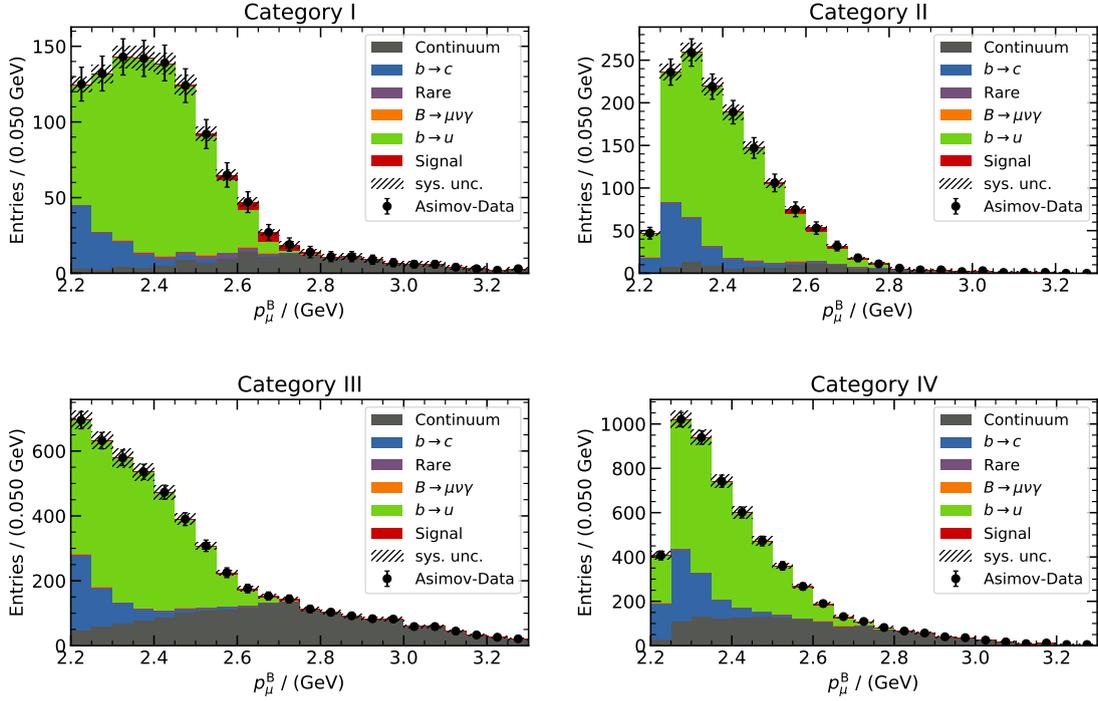


Figure 6.6.: Post-fit distributions after fitting the templates to the Asimov data sample.

The expected median sensitivity determined on the Asimov data sample is:

- Two-sided with the test statistic from Equation (6.7):
  - $2.1\sigma$  with statistical and systematic uncertainties
  - $2.4\sigma$  with only statistical uncertainties
- One-sided with the test statistic from Equation (6.10):
  - $2.4\sigma$  with statistical and systematic uncertainties
  - $2.6\sigma$  with only statistical uncertainties

### 6.7.2. Fit on Classifier Sideband Data

The region  $\mathcal{C}_{\text{out}} \in [0.90, 0.93]$  of the classifier output was defined as a control sample for the background models. The sample was split in the angular variable  $\cos \Theta_{B\mu} > 0$  and  $\cos \Theta_{B\mu} < 0$ . The expected amount of signal events in this control region is small with  $\nu_{\text{MC}}^{\text{sig}} = 15$ . The region can be used to test the signal extraction for closure and to validate the background model.

The fit result on the classifier sideband data is shown on Figure 6.9. The signal yield is negative but compatible with zero. The goodness-of-fit is  $\chi^2/\text{dof} = 30.4/40$ . This indicates that the data is well described by the underlying templates. The Likelihood profiles of the extracted expectation values are shown in Figure 6.10.

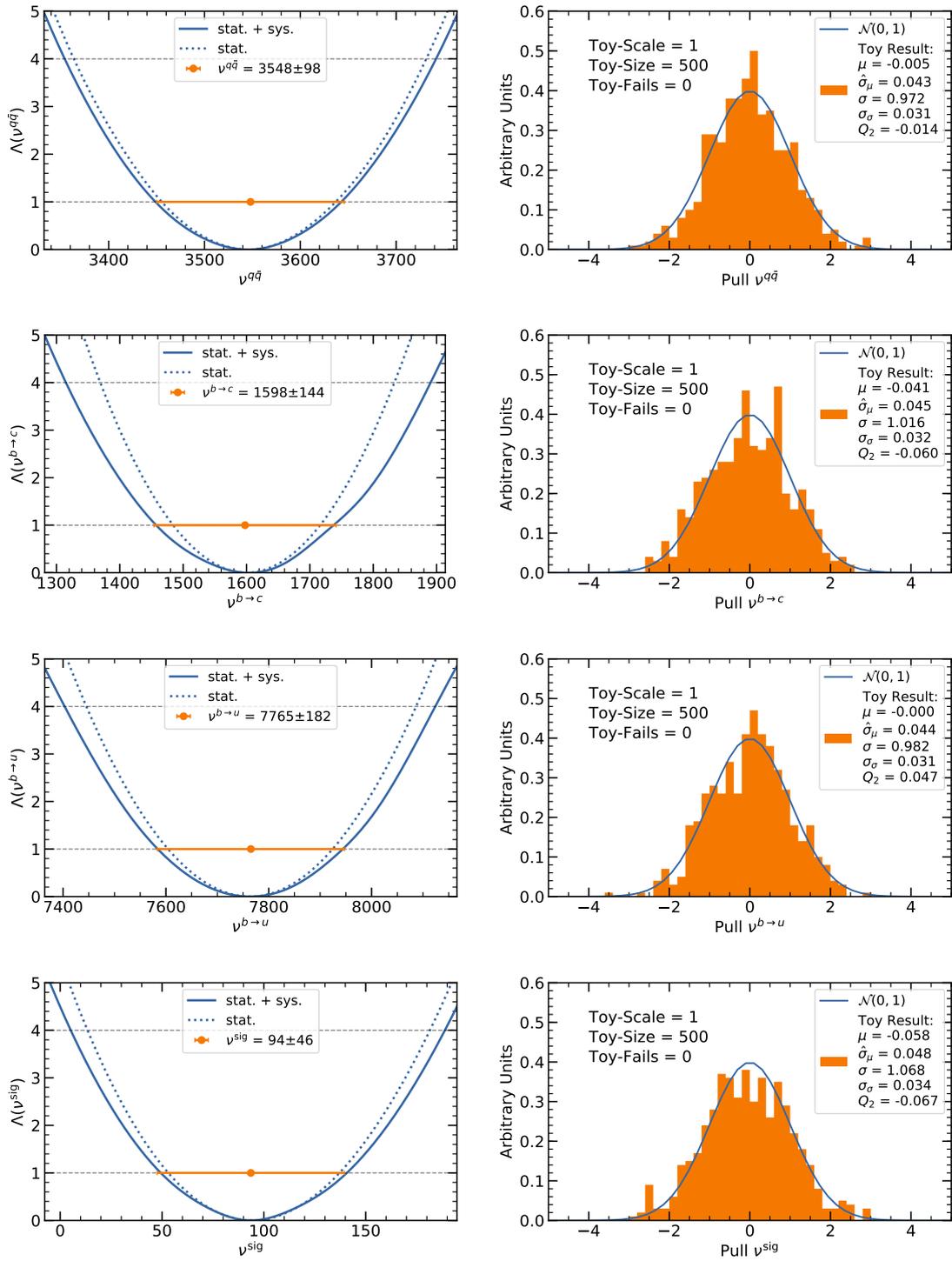


Figure 6.7.: Likelihood profiles and toy experiment results for the extracted parameters on the Asimov data sample.

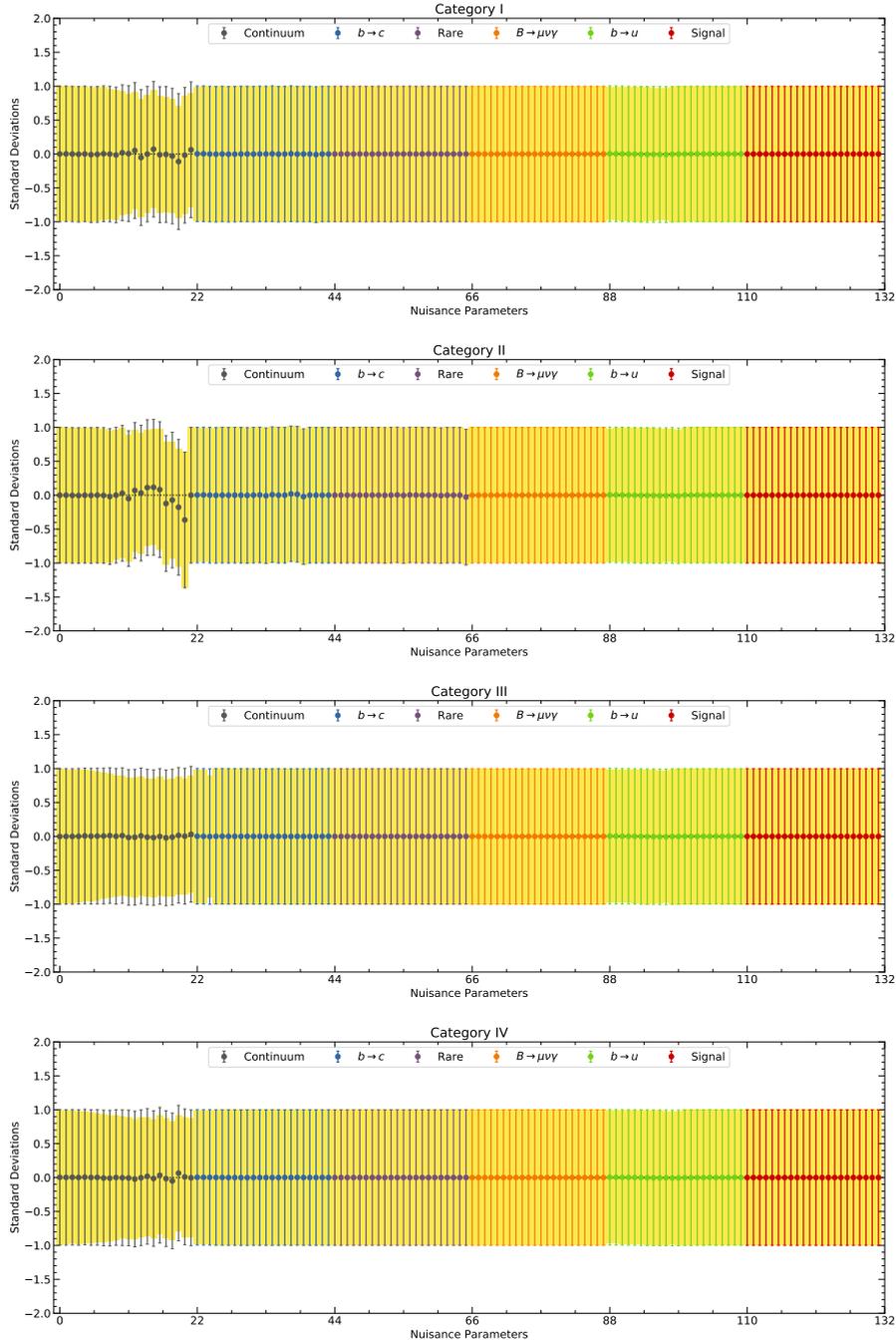


Figure 6.8.: Nuisance parameters after the fit on the Asimov data sample. The error bars show the pre-fit uncertainty associated to the nuisance parameters, the underlying yellow bar shows the post-fit uncertainty on the nuisance parameter.

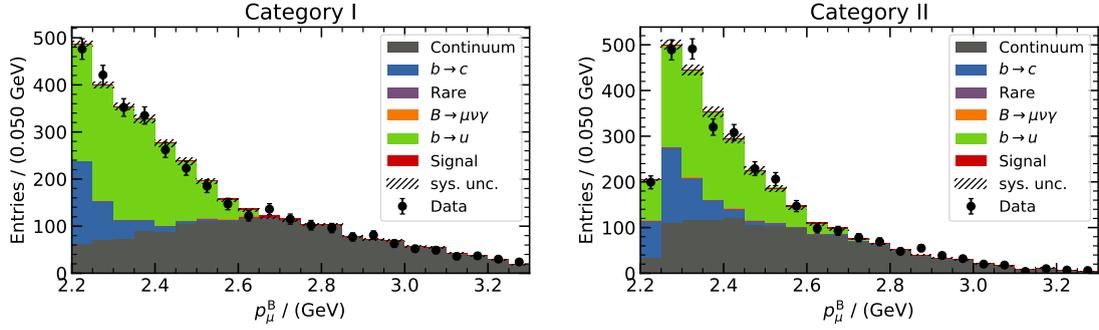


Figure 6.9.: Post-fit distributions after fitting the templates to the classifier sideband data sample.

The pulls on the nuisance parameter are all within one standard deviation, which shows that the fit templates can describe the data adequately.

### 6.7.3. Fit on Momentum Sideband Data

The momentum sideband is defined in the final categories given in Table 6.1 with the region of  $p_\mu^B \in [2.5, 2.85]$  GeV blinded. The signal template and the corresponding nuisance parameters are fixed to zero for this fit. The fit result is shown in Figure 6.12 with the associated Likelihood profiles of the extracted parameters in Figure 6.13. The goodness-of-fit is  $\chi^2/\text{dof} = 60.7/48$ . The pull on the nuisance parameters is well-behaved, as is shown in Figure 6.14.

As an additional cross-check the inclusive branching ratio of  $B \rightarrow X_u \ell \nu$  is extracted on the momentum sideband data. The extracted branching ratio is  $\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.86 \pm 0.07) \cdot 10^{-3}$ . It is compatible with the branching ratio which was used to create the template and with the latest measurement of the BaBar collaboration [35]  $\mathcal{B}(B \rightarrow X_u \ell \nu) = (1.665 \pm 0.087^{+0.103}_{-0.094}) \cdot 10^{-3}$ , which also extracted the inclusive branching ratio at the kinematic endpoint of the spectrum.

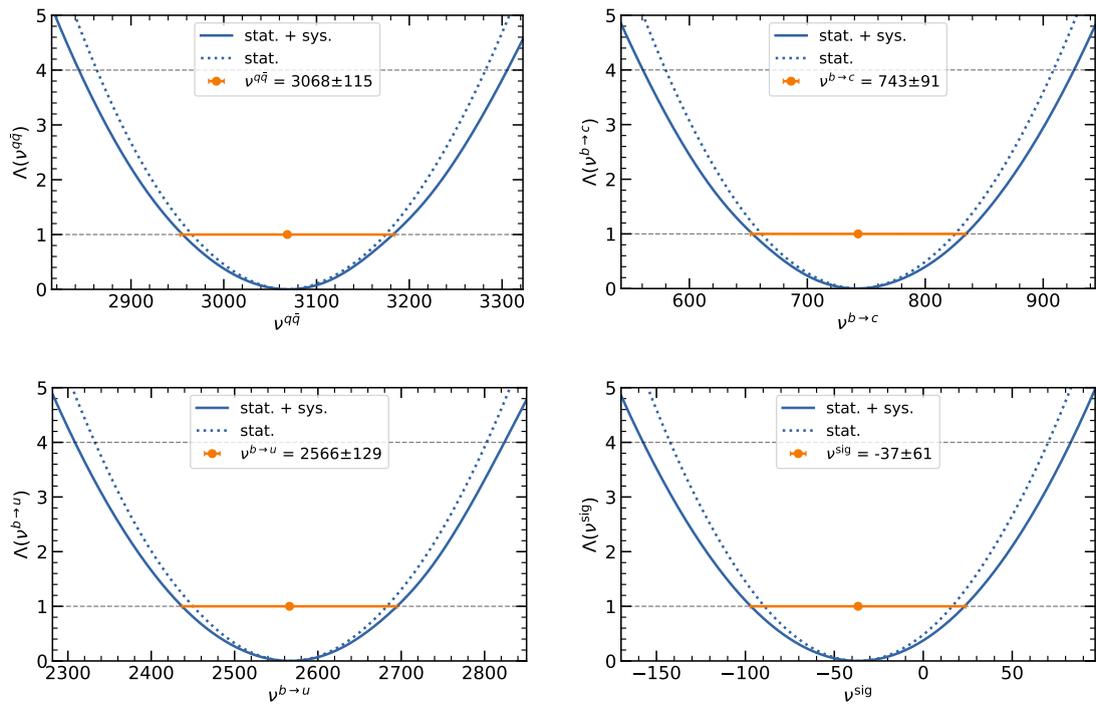


Figure 6.10.: Likelihood profiles for the extracted parameters on the classifier sideband data sample.

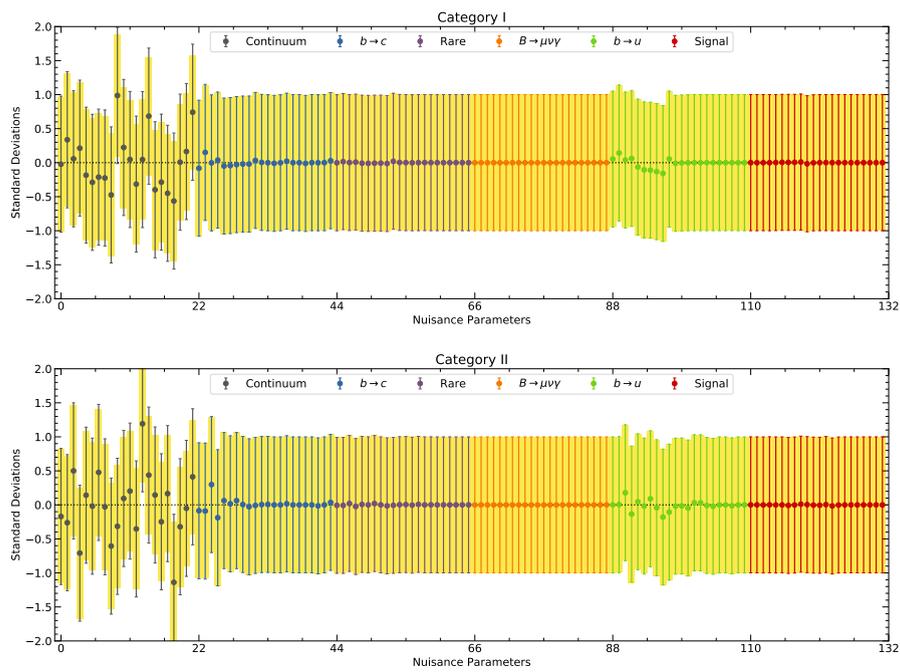


Figure 6.11.: Nuisance parameters after the fit on the classifier sideband data sample. The error bars show the pre-fit uncertainty associated to the nuisance parameters, the underlying yellow bar shows the post-fit uncertainty on the nuisance parameter.

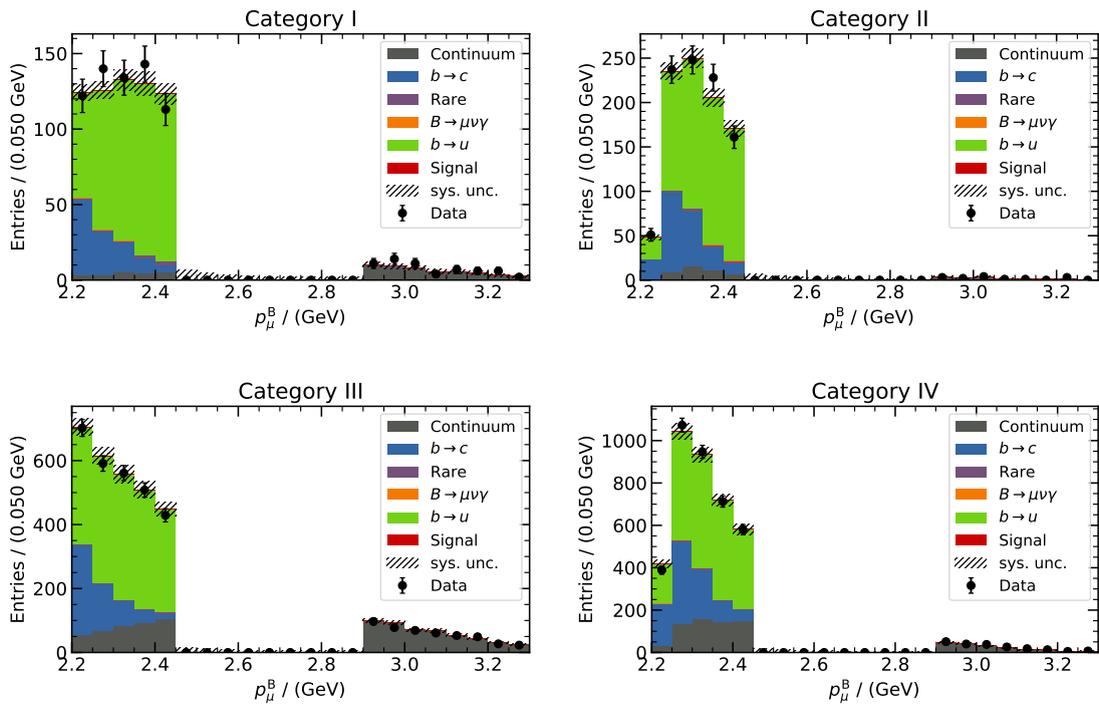


Figure 6.12.: Post-fit distributions after fitting the templates to the momentum sideband data sample.

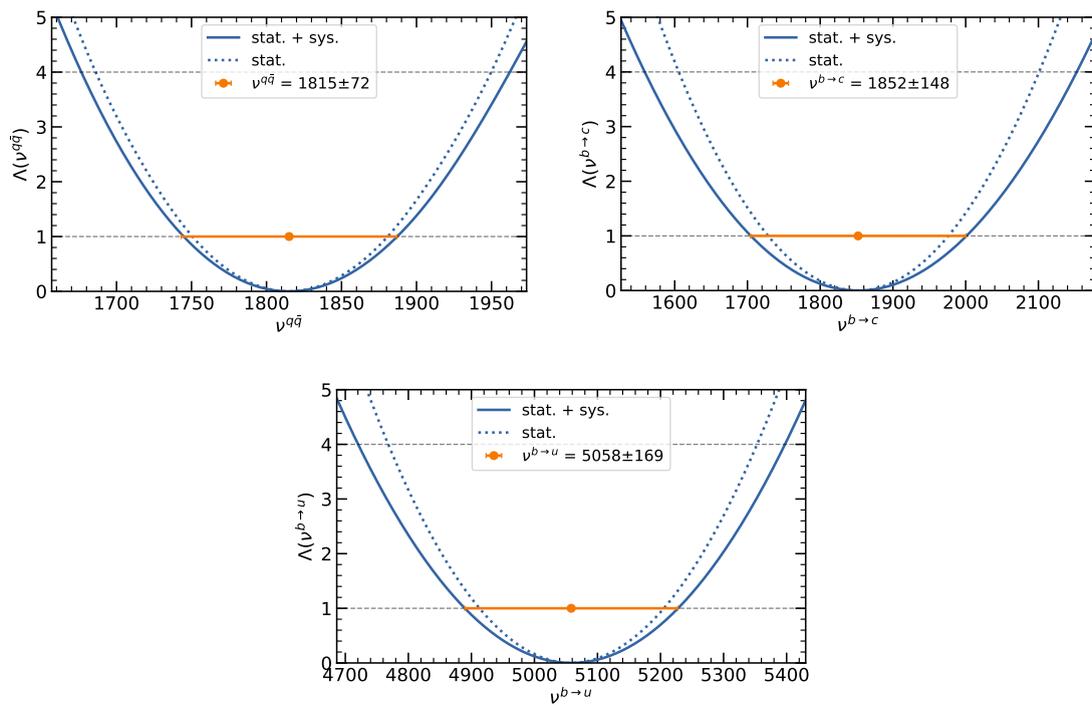


Figure 6.13.: Likelihood profiles for the extracted parameters on the momentum sideband data sample.

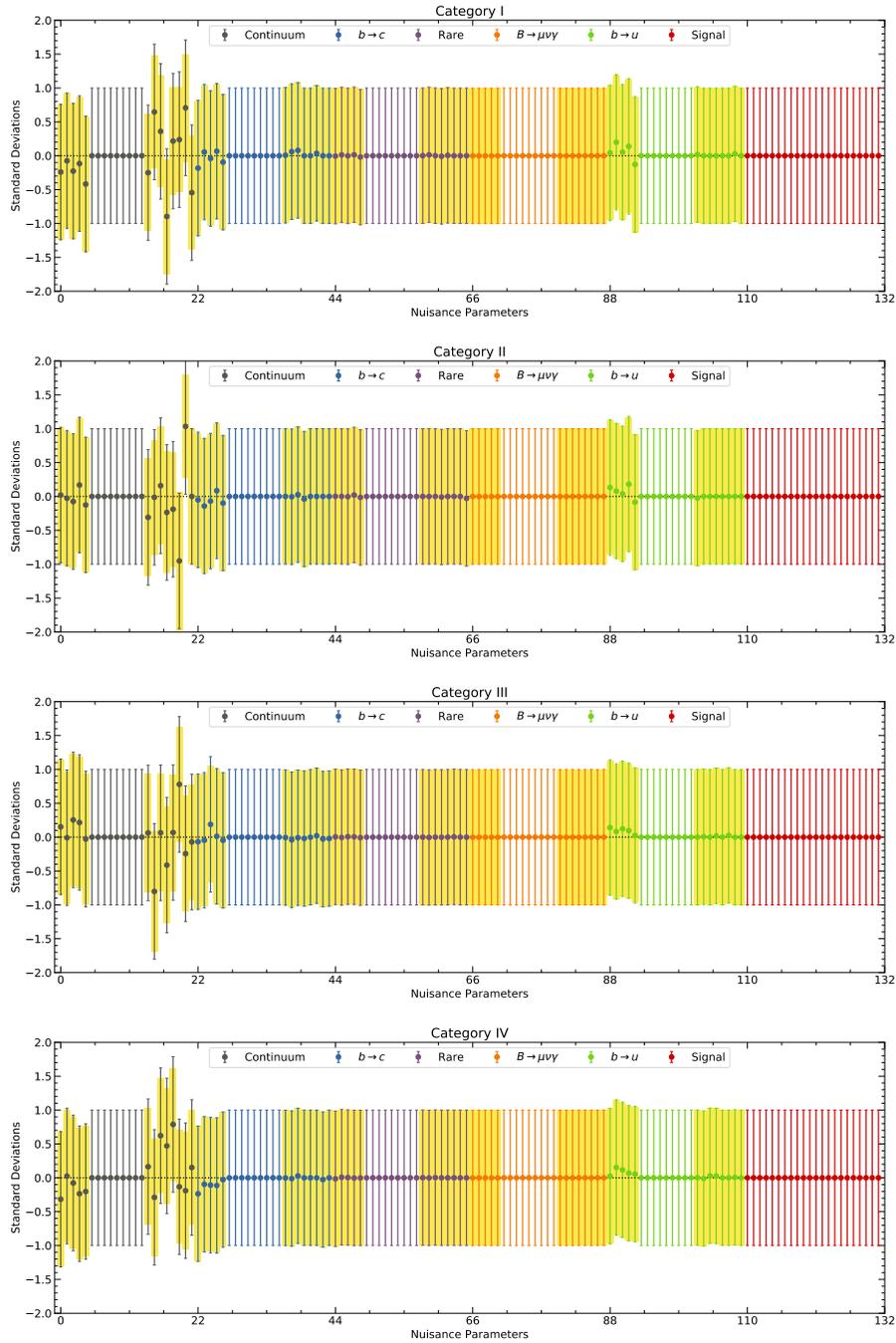


Figure 6.14.: Nuisance parameters after the fit on the momentum sideband data sample. The error bars show the pre-fit uncertainty associated to the nuisance parameters, the underlying yellow bar shows the post-fit uncertainty on the nuisance parameter. The nuisance parameters with a post-fit uncertainty of zero are fixed to zero in the fit.

## 7. Systematics

In this chapter the uncertainties associated with each template are discussed and the construction of the respective covariance matrices is described.

In general, the uncertainty is determined for each template separately in each bin and in each category and a vector  $\sigma = (\sigma_I, \sigma_{II}, \sigma_{III}, \sigma_{IV})$  including the systematic uncertainties of all bins in all categories is constructed. From this, the covariance matrix is built, either as a diagonal matrix for uncorrelated uncertainties between each bin  $C = \text{diag}(\sigma^2)$  or as fully correlated matrix for correlated uncertainties between bins  $C = \sigma \otimes \sigma$ .

### 7.1. Limited MC Statistics

A common systematic uncertainty for all templates is the uncertainty due to the limited MC statistics. The assigned error in each bin is  $\sigma_i = \sum w^2$ , where  $w$  are the weights assigned to the template, and treated as uncorrelated. The covariance matrices are given by  $C_k = \text{diag}(\sigma_k^2)$ . The uncertainty is displayed in Figure 6.1 for all templates.

### 7.2. Continuum Template

The data-MC reweighting is sensitive to the limited statistics of the off-resonance data sample. The effect of this on the weights assigned by the classifier is determined by bootstrapping, which is done in the following way: 50 new same size samples are generated by sampling with replacement from the original off-resonance data sample and the training was repeated. The error in each bin is assigned as the standard deviation of entries  $\sigma(N_1, N_2, \dots, N_{50})$ , where  $N_k = \sum w$  is the number of events for the  $k$ th training and  $w$  are the assigned weights. The errors are treated as uncorrelated as their origin lies in the statistical uncertainty of the off-resonance data. The covariance matrix is given by  $C_{q\bar{q}} = \text{diag}(\sigma_k^2)$ . The continuum template and its uncertainty are shown in Figure 7.1.

### 7.3. $b \rightarrow c$ Template

The  $b \rightarrow c$  template consists of two major contributions on which the systematic uncertainties are determined, namely the semileptonic decays  $B \rightarrow D^{(*)} \ell \nu$ . The branching ratio of the two contributions is varied individually and assigned as a systematic uncertainty. The branching ratios are listed in Table 7.1.

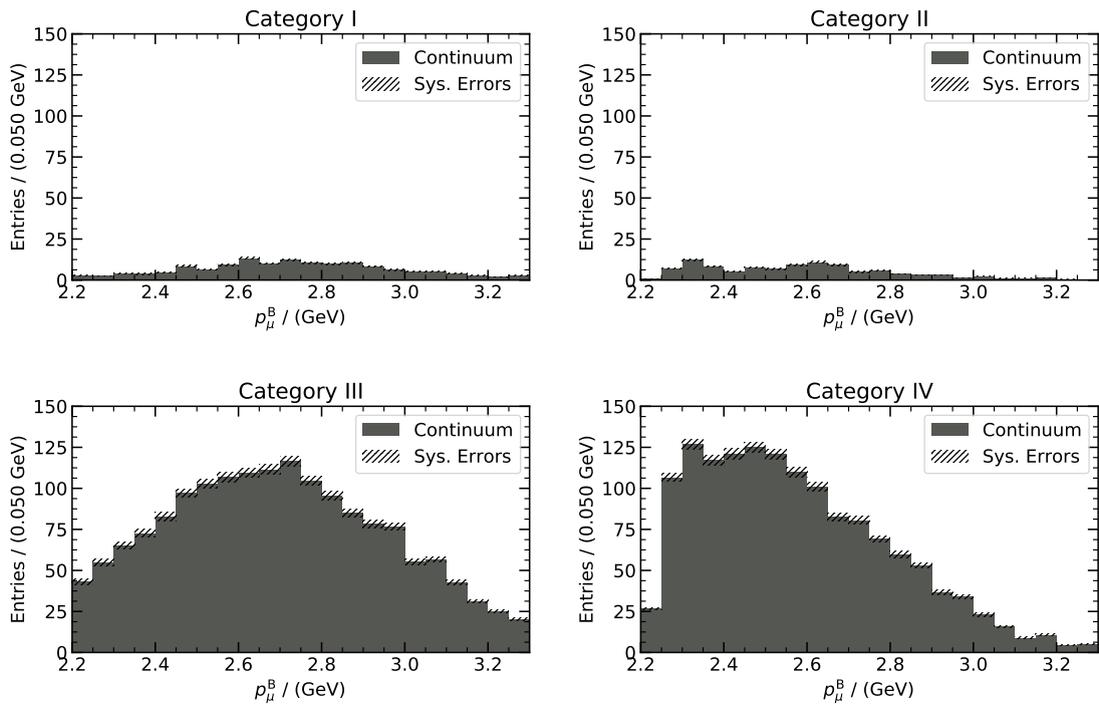


Figure 7.1.: The systematic uncertainty assigned to the continuum template. The origin of the error is the limited off-resonance data sample used for the data-MC reweighting.

Table 7.1.: Corrected branching ratios with the uncertainties from which the systematic uncertainties for the different templates are derived.

Decay	Branching Ratio	
b $\rightarrow$ c Template		
$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$(2.29 \pm 0.08)\%$	[36]
$B^0 \rightarrow D^- \ell^+ \nu$	$(2.13 \pm 0.07)\%$	[36]
$B^+ \rightarrow \bar{D}^{0*} \ell^+ \nu$	$(5.25 \pm 0.11)\%$	[36]
$B^0 \rightarrow D^{-*} \ell^+ \nu$	$(4.88 \pm 0.10)\%$	[36]
Rare Template		
$B^+ \rightarrow K^0 \pi^+$	$(2.37 \pm 0.08) \cdot 10^{-5}$	[4]
B $\rightarrow \mu\nu_\mu\gamma$ Template		
$B \rightarrow \mu\nu_\mu\gamma$	$1.0_{-1.0}^{+1.4} \pm 0.4$	[10]
b $\rightarrow$ u Template		
$B^+ \rightarrow \pi^0 \ell^+ \nu$	$(7.8 \pm 0.27) \cdot 10^{-5}$	[4]
$B^+ \rightarrow \omega \ell^+ \nu$	$(1.19 \pm 0.09) \cdot 10^{-4}$	[4]
$B^+ \rightarrow \rho \ell^+ \nu$	$(1.58 \pm 0.11) \cdot 10^{-4}$	[4]
$B^0 \rightarrow \pi^- \ell^+ \nu$	$(1.50 \pm 0.06) \cdot 10^{-4}$	[4]
$B^0 \rightarrow \rho \ell^+ \nu$	$(2.94 \pm 0.21) \cdot 10^{-4}$	[4]
$B \rightarrow u\ell\nu$ inclusive	$(2.13 \pm 0.31) \cdot 10^{-3}$	[4]

A second source of systematic uncertainty originates from the uncertainty from the form factor model with which the MC is created. The CLN model is reweighted to the more recent BGL parametrization. The model parameters and their uncertainties are taken from [37–40]. For each individual template contribution, the covariance matrix is diagonalized and varied in sequence by each eigenvalue. For each eigenvalue a covariance matrix is determined, where the uncertainty in each bin is assigned by  $\sigma_i = N_{\text{variation}} - N_{\text{central}}$ . The total covariance matrix for the model uncertainties is then given by the sum of the individual contributions  $C = \sum_i^{\text{eigenvalues}} \sigma_i \otimes \sigma_i$ . The b  $\rightarrow$  c template and its uncertainty are shown in Figure 7.2.

## 7.4. Rare Decay Template

The rare template is dominated by  $B \rightarrow K_L^0 \pi$  decays. The template is varied by the uncertainty of the measured branching ratio of this decays and the systematic uncertainty is assigned accordingly. The branching ratio is listed in Table 7.1. The rare template and its uncertainty are shown in Figure 7.3.

## 7.5. $B \rightarrow \mu\nu_\mu\gamma$

The decay  $B \rightarrow \mu\nu\gamma$  is treated as background to the signal decay (see Section 2.2). The background is fixed to the latest Belle measurement [13]  $\mathcal{B}(B \rightarrow \mu\nu\gamma) = (1.0 \pm 1.2 \pm 0.4) \cdot 10^{-6}$

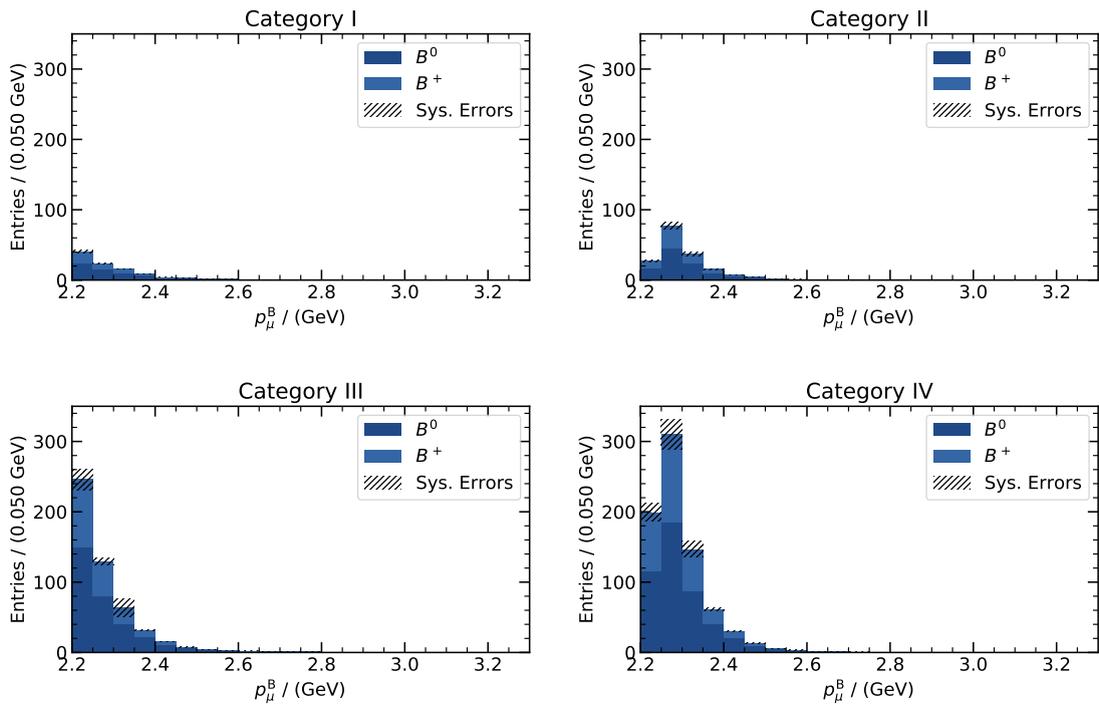


Figure 7.2.: The systematic uncertainty assigned to the  $b \rightarrow c$  template. The uncertainties originate from the uncertainty on the branching ratios and the model uncertainties of the dominant contributions in the template.

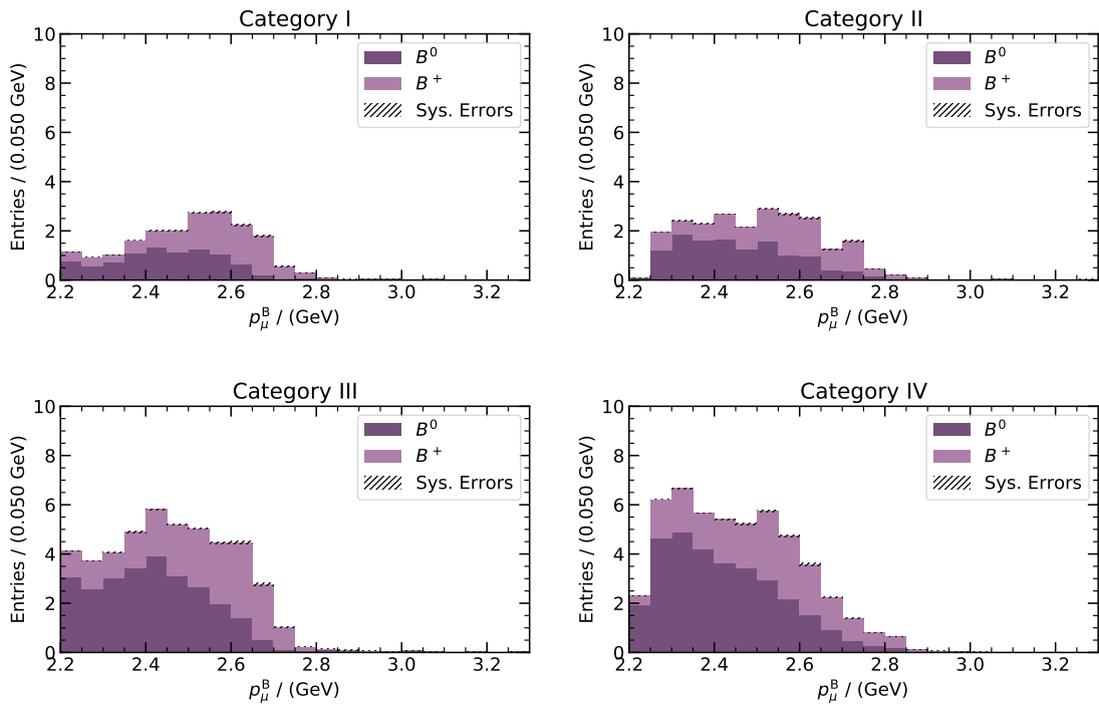


Figure 7.3.: Systematic errors on the rare template. The uncertainties originate from the uncertainty on the branching ratios of the dominant contributions in the template.

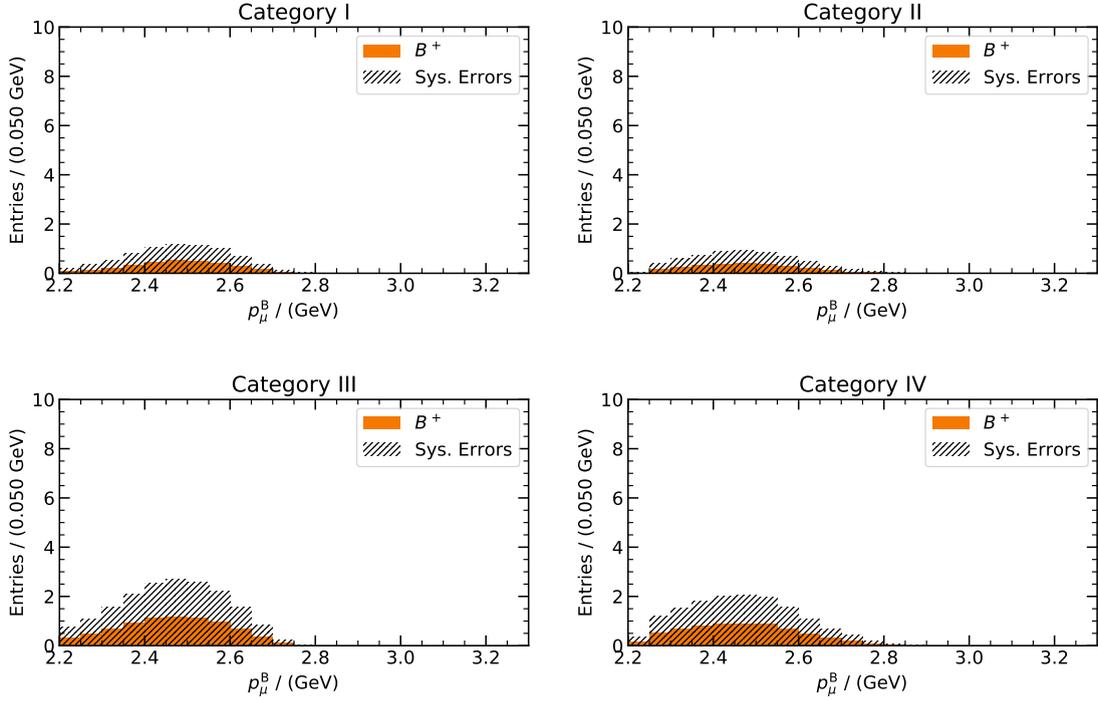


Figure 7.4.: Systematic errors on the  $B \rightarrow \mu\nu\gamma$  template. The uncertainties originate from the uncertainty on the measured branching ratio.

and an uncertainty corresponding to the uncertainty on the measurement is applied. The  $B \rightarrow \mu\nu\gamma$  template and its uncertainty are shown in Figure 7.4.

## 7.6. $b \rightarrow u$ Template

The  $b \rightarrow u$  template consists of four major contributions on which the systematic uncertainties are determined. The individual components are discussed in detail in Part II of this thesis. The  $b \rightarrow u$  template is modeled as the sum of the resonant decays  $B \rightarrow \pi l\nu$ ,  $B \rightarrow \rho l\nu$ ,  $B \rightarrow \omega l\nu$  and the non-resonant decay  $B \rightarrow X_u l\nu$ . The resonant and non-resonant contributions are combined using a Hybrid model. To properly estimate the correlations when varying the branching ratios, the Hybrid model weights were generated anew for each variation of branching ratios of the resonant decays. The uncertainty is determined for each resonant variation individually and summed up. The error on the branching ratio is treated as fully correlated for each individual component. The branching ratios are listed in Table 7.1.

A second source of systematic uncertainty originates from the uncertainty onto the model parameters with which the MC is created. For the  $B \rightarrow \pi l\nu$  and the  $B \rightarrow X_u l\nu$  templates, the model parameters and uncertainties are taken from [36]. For the  $B \rightarrow \rho l\nu$  and  $B \rightarrow \omega l\nu$  templates, new world averages of the form factors were obtained, what is discussed in Chapter 10. For each individual template contribution, the covariance matrix is diagonalized and varied in sequence by each eigenvalue. For each eigenvalue a covariance

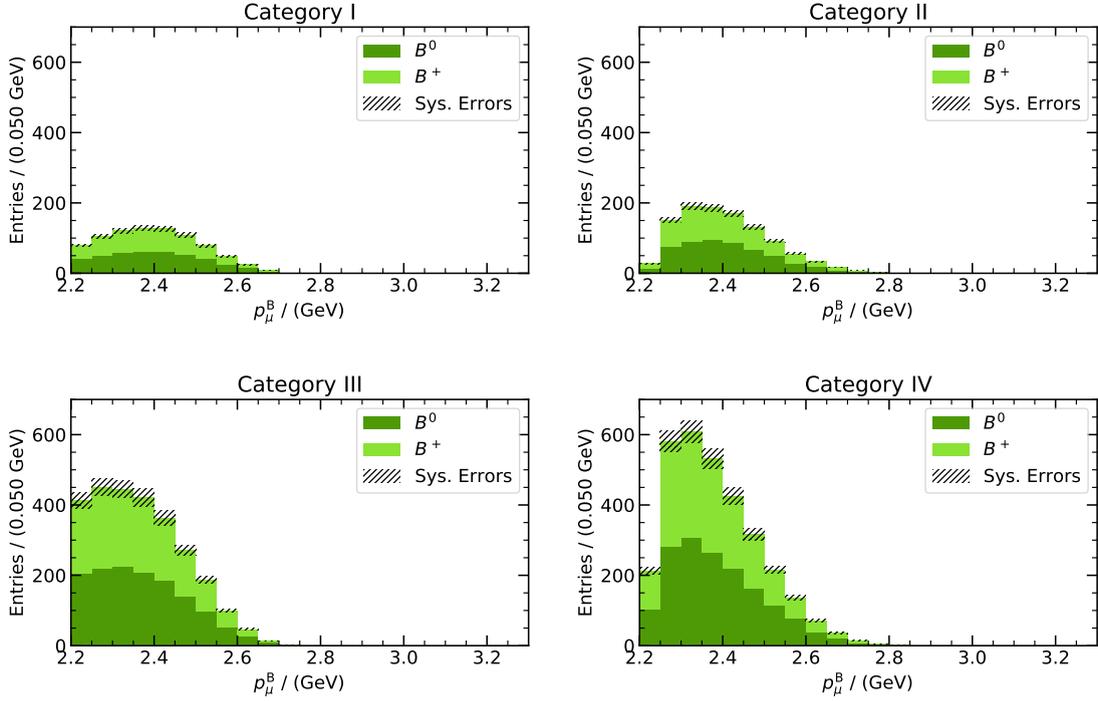


Figure 7.5.: Systematic errors on the  $b \rightarrow u$  template. The uncertainties originate from the uncertainties on the branching ratios and the model uncertainties of the exclusive and inclusive theory predictions.

matrix is determined, where the uncertainty in each bin is assigned by  $\sigma_i = N_{\text{variation}} - N_{\text{central}}$ . The total covariance matrix for the model uncertainties is then given by the sum of the individual contributions  $C = \sum_i^{\text{eigenvalues}} \sigma_i \otimes \sigma_i$ . The  $b \rightarrow u$  template and its uncertainty is shown in Figure 7.5.

## 7.7. Signal Template

Three different systematic uncertainties are assigned to the signal template. Each of the following sources of systematic uncertainty is treated as fully correlated and the covariance matrix is constructed as  $C_{\text{sig}} = \sum_i \sigma_i \otimes \sigma_i$ .

**Lepton Identification** The systematic uncertainty on the lepton identification was studied in [41]. The efficiency correction is applied and a systematic uncertainty of 1.959%, 1.871%, 2.067% and 2.055% is assigned to the signal template in categories I-IV.

**Track Reconstruction Efficiency** The systematic uncertainty on the tracking efficiency was studied in [42]. Bases on this study, a systematic uncertainty of 0.34% is assigned to the signal template in each category.

**Number of  $B\bar{B}$  Events** The number of  $B\bar{B}$  events in the full Belle data set is  $N_{B\bar{B}} = (771.6 \pm 10.6) \cdot 10^6$ . The systematic uncertainty on the signal template  $\sigma_{B\bar{B}}$  is determined by varying the template in respect to the uncertainty on  $N_{B\bar{B}}$ .

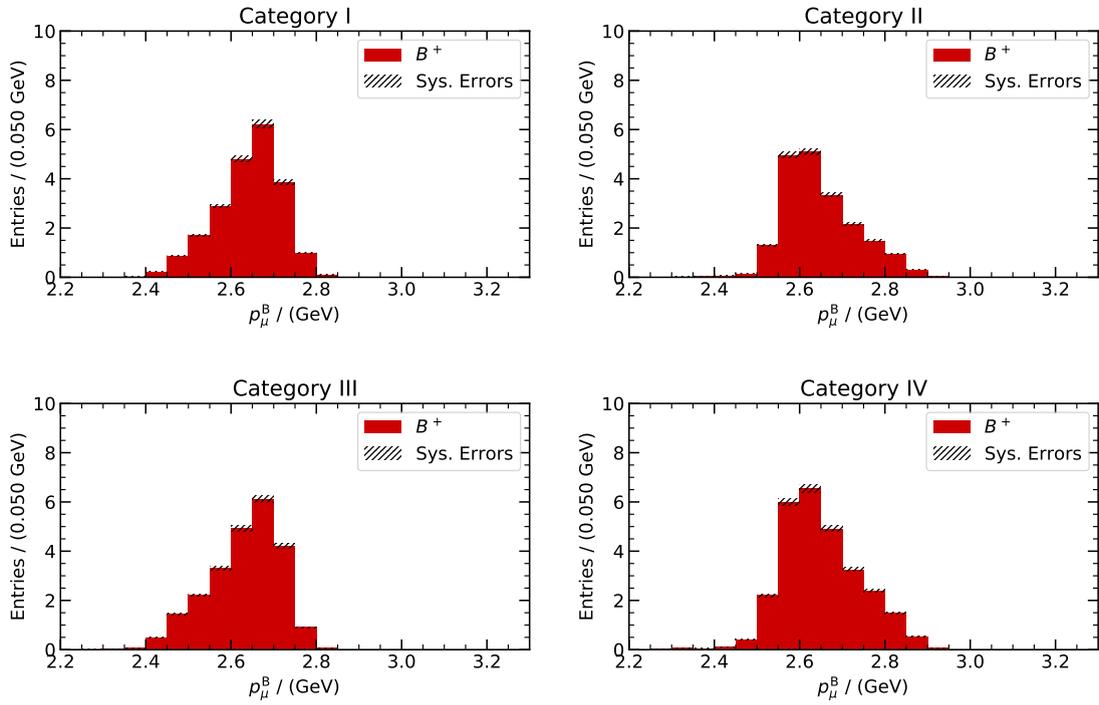


Figure 7.6.: Systematic errors on the signal template. The uncertainties originate from the uncertainties on the number of  $B\bar{B}$  pairs, the tracking efficiency and the lepton identification.

The signal template and its systematic uncertainty is shown in Figure 7.6.

## 8. Measurement on Data and its Interpretation

The fit was performed on the full Belle data sample, resulting in a branching ratio of

$$\mathcal{B}(\text{B} \rightarrow \mu\nu_\mu) = (5.297 \pm 2.169) \cdot 10^{-7}. \quad (8.1)$$

The pre-fit and post-fit distribution are shown in Figure 8.1 and Figure 8.2 respectively. The agreement between the model and the data in terms of the goodness-of-fit is  $\chi^2/\text{dof} = 96.9/84$  before the data is fitted and  $\chi^2/\text{dof} = 58.8/84$  after the Likelihood fit is performed. The Likelihood profiles are shown in Figure 8.3 and the pulls on the nuisance parameters in Figure 8.4.

The observed sensitivity of the fit is:

- Two-sided with the test statistic from Equation (6.7):
  - $2.53\sigma$  with statistical and systematic uncertainties
  - $3.05\sigma$  with only statistical uncertainties
- One-sided with the test statistic from Equation (6.10):
  - $2.75\sigma$  with statistical and systematic uncertainties
  - $3.11\sigma$  with only statistical uncertainties

The observed sensitivity is not high enough to claim evidence for the decay, therefore an upper limit is determined and listed together with the SM prediction in Table 8.1. The construction of the Frequentist and Bayesian upper limit is shown in Figure 8.5.

Under the assumption that the decay is only governed by the SM  $W^\pm$  gauge boson exchange, the value of  $|V_{\text{ub}}|$  can be extracted. It is found to be  $|V_{\text{ub}}| = (4.37^{+0.82}_{-1.01}) \cdot 10^{-3}$ . The value

Table 8.1.: Determined upper limit of  $\text{B} \rightarrow \mu\nu_\mu$ .

	$\mathcal{B}(\text{B} \rightarrow \mu\nu_\mu)$
Standard Model	$\mathcal{B} = 4.26 \cdot 10^{-7}$
Frequentist Upper Limit	$\mathcal{B} < 8.64 \cdot 10^{-7}$ @ 90% CL
Bayesian Upper Limit	$\mathcal{B} < 8.86 \cdot 10^{-7}$ @ 90% CL

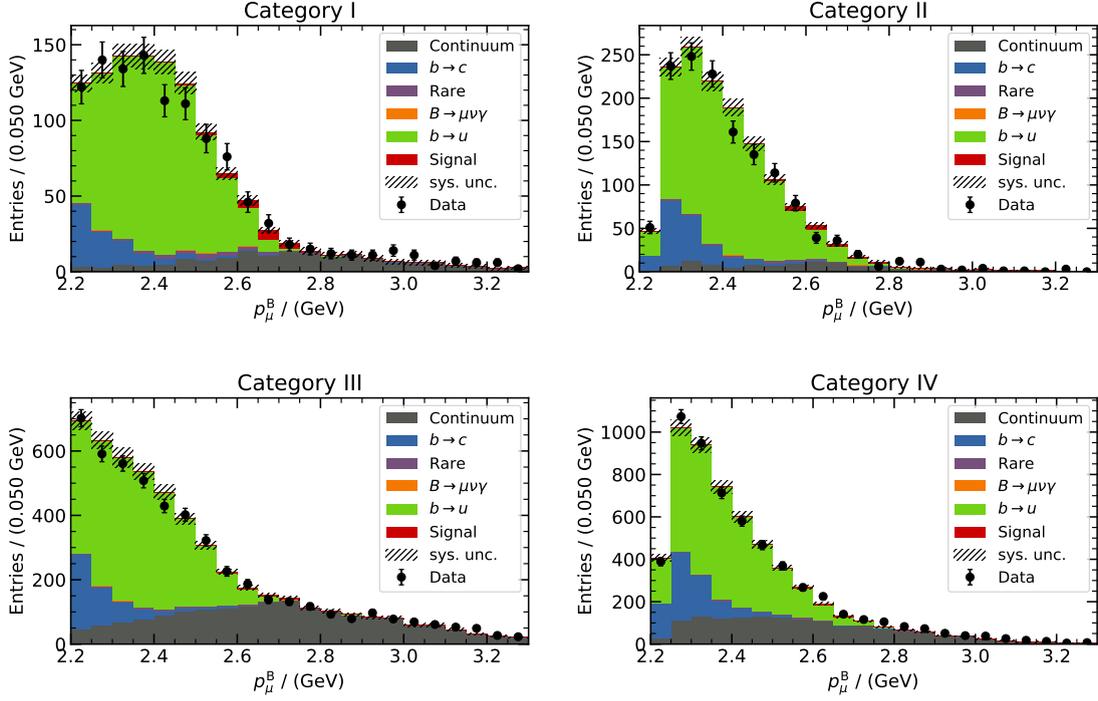


Figure 8.1.: Pre-fit distributions before fitting the templates to the data sample.

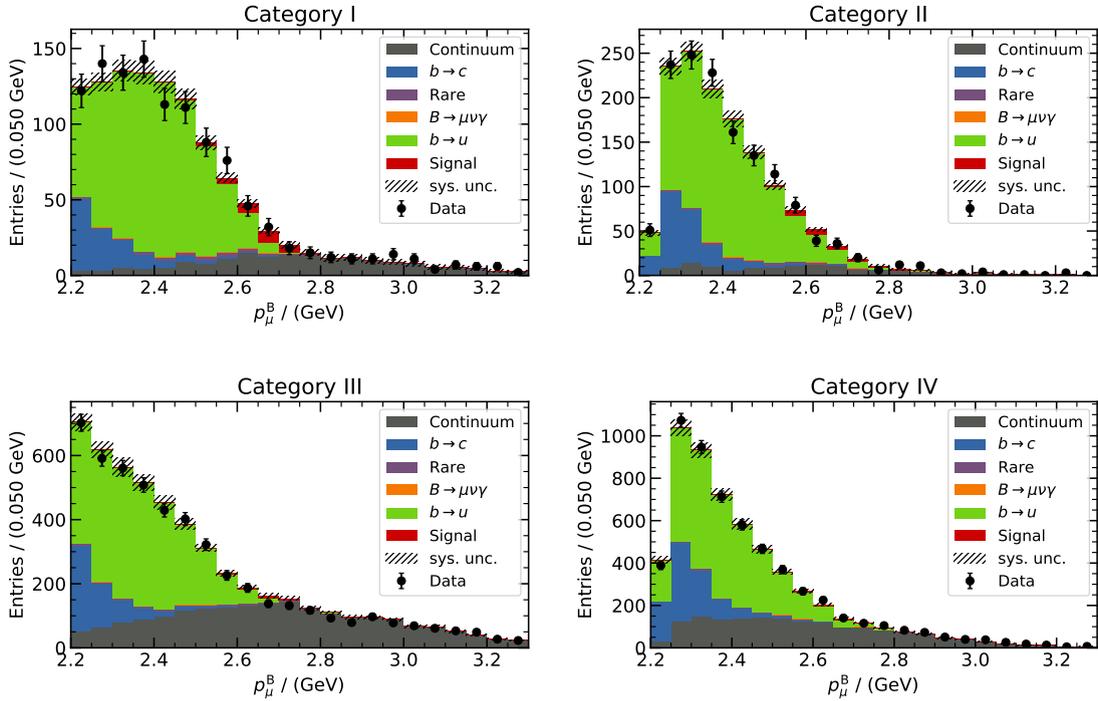


Figure 8.2.: Post-fit distributions after fitting the templates to the data sample.

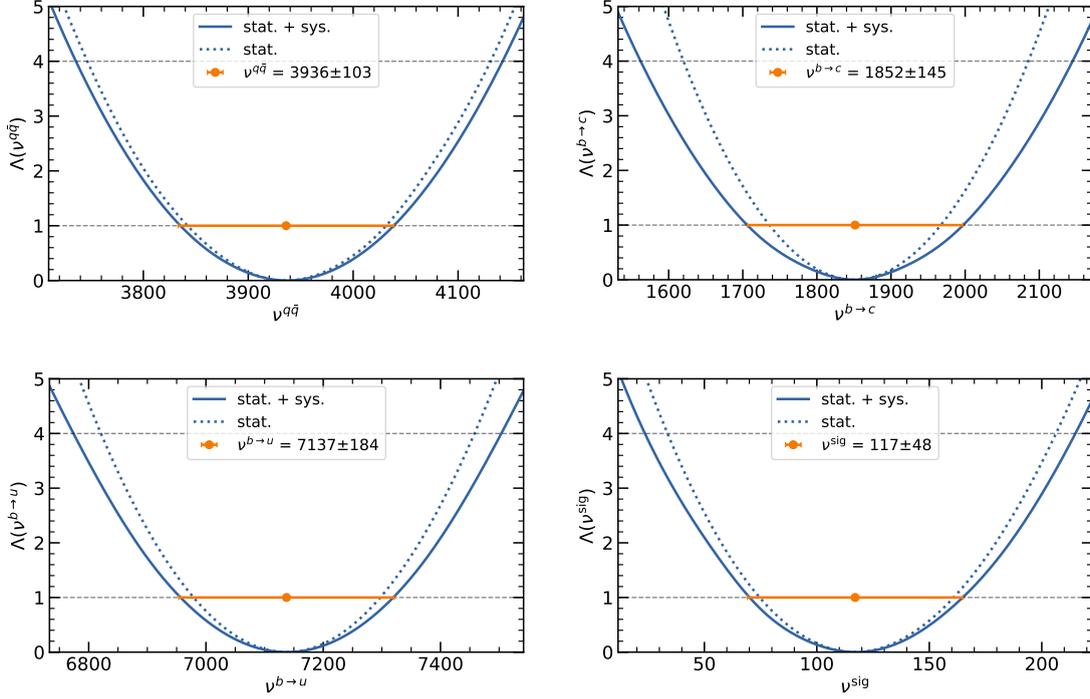


Figure 8.3.: Likelihood profiles for the extracted parameters on the data sample.

of  $|V_{ub}|$  is shown together with the time evolution of  $|V_{ub}|$  from semileptonic B decays in Figure 8.6. The uncertainty on the extracted value of  $|V_{ub}|$  is too large to resolve the discrepancy between inclusive and exclusive measurements of  $|V_{ub}|$ .

## 8.1. Exclusion Limits for Sterile Neutrinos

The scan for the decay  $B \rightarrow \mu N$  with a massive sterile neutrino in the final state was performed by repeating the fit described in Chapter 6 with an additional template containing the  $B \rightarrow \mu N$  decay. The scan was performed twice, once with the assumption that the decay  $B \rightarrow \mu \nu_\mu$  exists with a branching ratio of the SM expectation and once with a branching ratio of zero. The best fit Likelihood values as a function of the sterile neutrino mass are shown in Figure 8.7. The one sided p-value of the best fit, calculated with the test statistic given in Equation (6.10), as a function of sterile neutrino mass is shown in Figure 8.8.

No significant excess is found in the scan with the SM decay  $B \rightarrow \mu \nu_\mu$  being present. The most significant excess has a significance of approximately  $1.4\sigma$ . Without the SM decay included, the best fit point is found at a neutrino mass of  $m_N \approx 0.8 \text{ GeV}$  with a significance of approximately  $3.1\sigma$ .

## 8.2. Exclusion Limits for Two Higgs-Doublet Models

The measured branching ratios of the leptonic decays can be used to determine exclusion limits for the type-II and type-III Higgs doublet models. The measured branching ratios

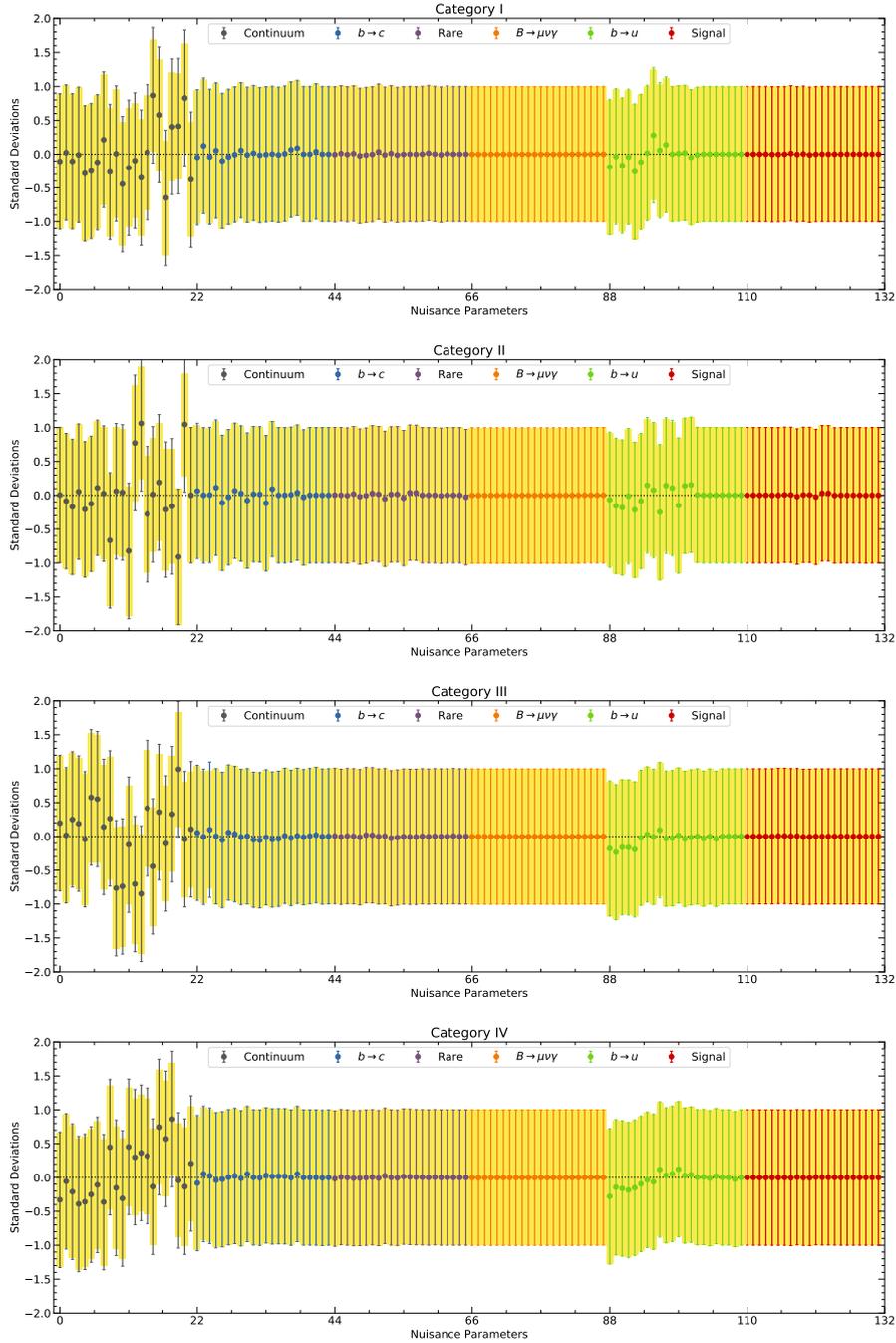


Figure 8.4.: Nuisance parameters after the fit on the data sample. The error bars show the pre-fit uncertainty associated to the nuisance parameters, the underlying yellow bar shows the post-fit uncertainty on the nuisance parameter.

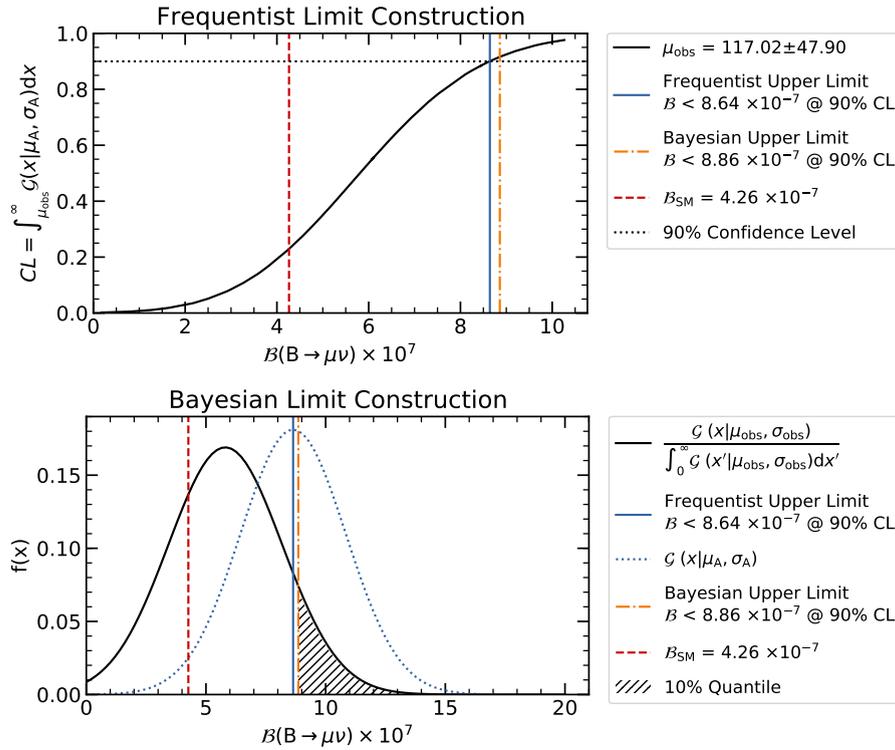


Figure 8.5.: The construction of the Frequentist (top) and Bayesian (bottom) upper limit.

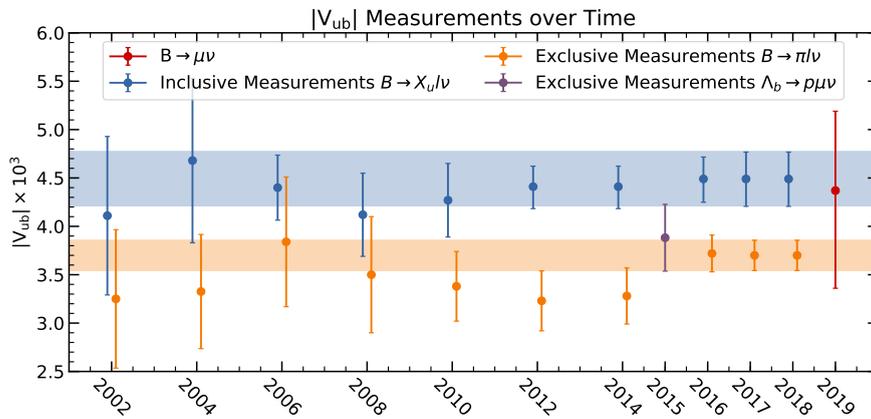


Figure 8.6.:  $V_{ub}$  over time with the result of this work. The uncertainty on the extracted  $V_{ub}$  from this analysis is too large to give a solution to the discrepancy between the inclusive and exclusive semileptonic measurements.

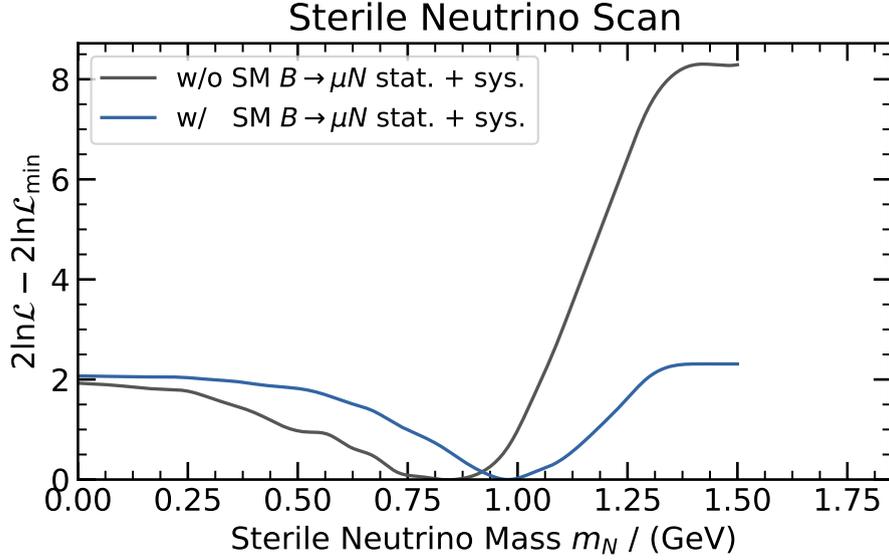


Figure 8.7.: The Likelihood value of the best-fit for different sterile neutrino masses. Although the Likelihood improves for non-zero sterile neutrino masses, no significant deviation from the null hypothesis can be found.

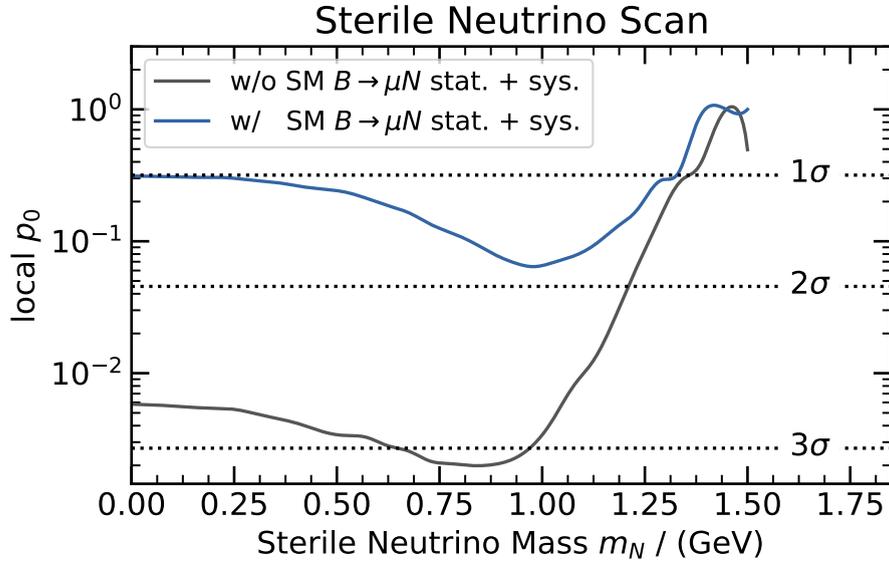


Figure 8.8.: The one-sided local p-value of the fit for different sterile neutrino masses. With the SM decay  $B \rightarrow \mu \nu_\mu$  present, no significant contribution of the decay  $B \rightarrow \mu \nu_\mu$  can be found. Without the SM decay, the most significant result is found at a sterile neutrino mass  $m_N \approx 0.8$  GeV. The result for  $m_N = 0$  GeV is equal to the result of the search for the final state with a standard model neutrino.

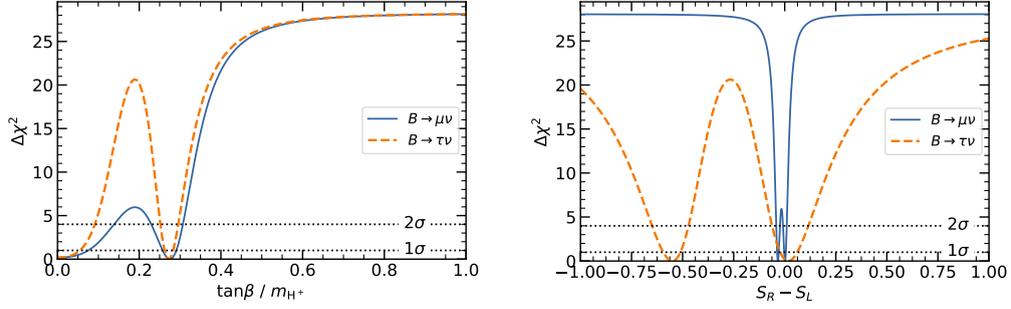


Figure 8.9.:  $\chi^2$  functions for the (left) type-II and (right) type-III two Higgs doublet model from which the exclusion limits on the parameter space of the individual models are extracted.

are

$$\begin{aligned} \mathcal{B}(B \rightarrow \mu\nu_\mu) &= (5.297 \pm 2.169) \cdot 10^{-7}, \\ \mathcal{B}(B \rightarrow \tau\nu_\tau) &= (1.09 \pm 0.25) \cdot 10^{-4}, \end{aligned} \quad (8.2)$$

where the measurement of  $B \rightarrow \tau\nu_\tau$  is taken from [4]. To calculate the exclusion limits, a  $\chi^2$ -function is constructed as

$$\chi^2 = \frac{(\mathcal{B}^{\text{measured}} - \mathcal{B}^{\text{theory}})^2}{\sigma_{\text{measured}}^2 + \sigma_{\text{theory}}^2}. \quad (8.3)$$

The theory prediction is given by either the type-II (Equation (2.4)) or type-III (Equation (2.5)) two Higgs doublet model. The  $\chi^2$  is shown in Figure 8.9 for both models. The extracted exclusion limits from the  $\chi^2$  functions are shown in Figure 8.10 for the type-II and Figure 8.11 for the type-III two Higgs doublet model.

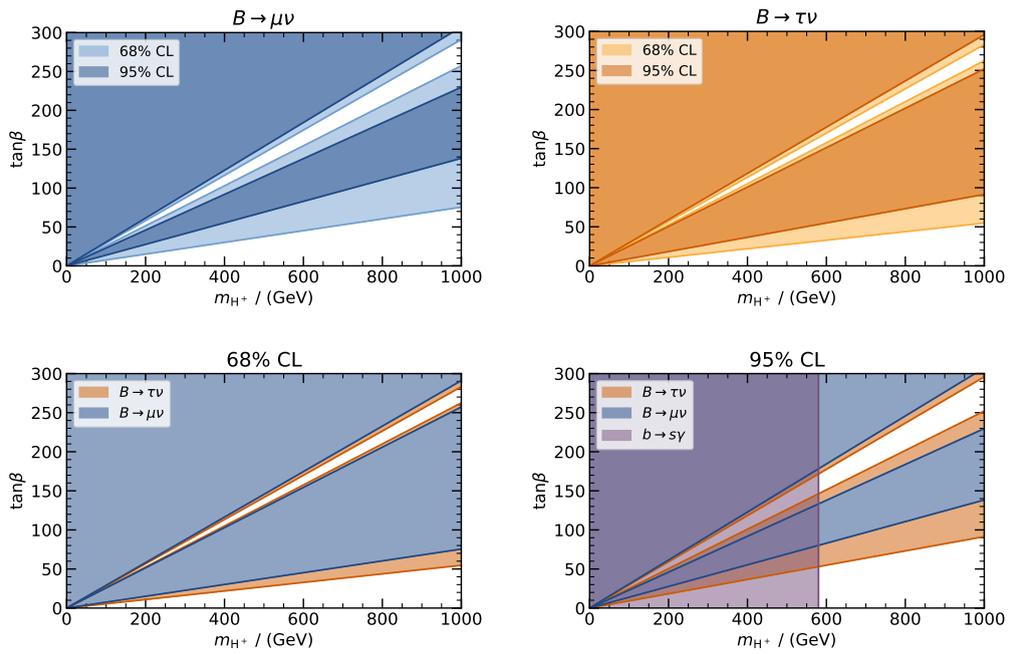


Figure 8.10.: Exclusion limits for the type-II two Higgs doublet model. The measured branching ratio of  $B \rightarrow \mu\nu_\mu$  does not exclude a new region of the parameter space, due to the larger uncertainty on the measured branching ratio. The bottom right plot also shows the exclusion from  $b \rightarrow s\gamma$  measurements [43].

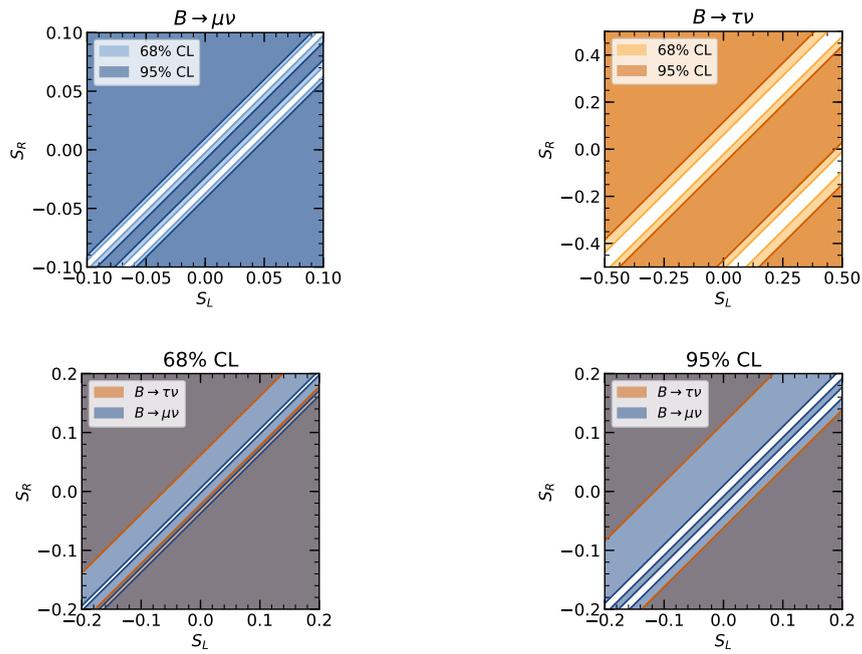


Figure 8.11.: Exclusion limits for the type-III two Higgs doublet model. Although the measurement of  $B \rightarrow \mu\nu_\mu$  has a larger uncertainty, it does provide more stringent exclusion limits due to the dependency of the new physics contribution on the inverse of the lepton mass. The axes are scaled differently for illustrative purpose.



Part II.

Semileptonic  $B \rightarrow X_u \ell \nu$  Decays



## 9. The Decay $B \rightarrow X_u \ell \nu$

The B meson, being the lightest meson containing a b quark, can only decay via the weak interaction. In the following I discuss the semileptonic decay  $B \rightarrow X_u \ell \nu$ , where the final state consists of a hadronic ( $X_u$ ) and a leptonic ( $\ell \nu$ ) system.

At the energy scale of the B meson mass the propagator term of the virtual  $W^\pm$  boson can be integrated out and the weak interaction is described by the effective coupling  $G_F$  together with the corresponding CKM matrix elements. However, at this energy scale the bound state of the two quarks, of which the B meson is composed, is described by non-perturbative QCD. In case the virtual  $W^\pm$  boson decays into a lepton and neutrino pair there exists no strong interaction between the decay products of the  $W^\pm$  and the hadronic system  $X_u$ . Therefore it is possible to factorize the strong and weak interaction contributions and treat them separately.

The effective Standard Model (SM) Lagrangian describing these decays is given by

$$\mathcal{L}_{\text{eff}} = \frac{-4G_F}{\sqrt{2}} V_{ub} (\bar{u} \gamma_\mu P_L b) (\bar{\nu} \gamma^\mu P_L \ell) + \text{h.c.}, \quad (9.1)$$

with Fermi's constant  $G_F$ , the CKM matrix element  $V_{ub}$  and the projection operator  $P_L = (1 - \gamma_5)/2$ . The decay  $B \rightarrow \pi \ell \nu$  is shown at parton level and as an effective diagram in Figure 9.1.

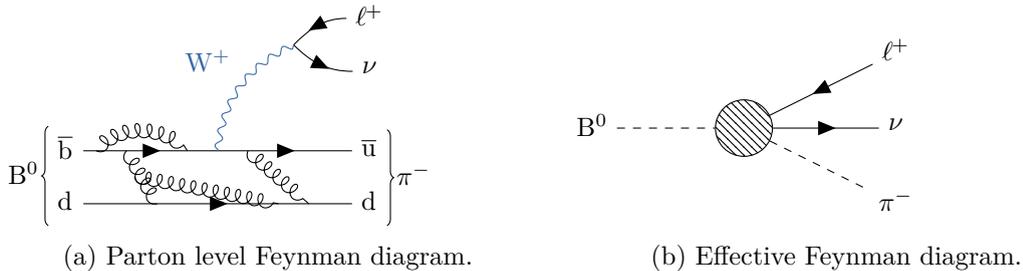


Figure 9.1.: One possible parton level Feynman diagram (a) and the effective Feynman diagram (b). In the effective Feynman diagram, the propagator of the W is integrated out, i.e. the weak interaction is point-like, and the gluon interactions are described by the blob.

Two different questions can be answered when the Lagrangian is studied. Either the exclusive decay rate of one specific resonant hadronic final state or the total inclusive rate, including all hadronic final states, can be calculated.

In Section 9.1 I briefly review the inclusive description of the decay. In Section 9.2 I discuss the exclusive decays.

## 9.1. Inclusive Predictions

This section summarizes the more extensive review found in [4].

The calculation of the inclusive  $B \rightarrow X_u \ell \nu$  decays is based on the Heavy Quark Expansion (HQE). In this treatment, a non-perturbative distribution function occurs, the so-called shape-function, which has an unknown form. The contribution of the shape function is especially relevant in the kinematic endpoint of the lepton energy spectrum.

The single shape function occurring in the leading order calculation can be measured from  $B \rightarrow X_s \gamma$  decays [44, 45]. This is possible, because the leading order shape function is universal for all heavy-to-light quark transitions. At subleading order additional shape functions appear and no trivial relations between different decays of the B meson are possible [46]. The theoretical description of the shape-functions is also non-trivial and different calculations are available [47–49].

The triple differential decay rate for inclusive semileptonic decays is given by

$$\begin{aligned} \frac{d^3\Gamma}{dq^2 dE_l dm_X} = \frac{G_F^2 |V_{ub}|^2}{16\pi^2} (m_B - P_+) & \left( (P_- - P_l)(m_B - P_- + P_l - P_+) \mathcal{F}_1 \right. \\ & + (m_B - P_-)(P_- + P_+) \mathcal{F}_2 \\ & \left. + (P_- - P_l)(P_l - P_+) \mathcal{F}_3 \right) \end{aligned} \quad (9.2)$$

with  $P_l = m_B - 2E_l$  and  $P_{\pm} = E_X \mp |\mathbf{p}_X|$ , where  $m_B$  is the mass of the B meson,  $E_l$  the energy of the lepton,  $E_X$  the energy of the hadronic system, and  $\mathbf{p}_X$  the total 3-momentum of the hadronic system. The  $\mathcal{F}_i$  are the structure functions of the B meson, which include the aforementioned shape functions and perturbative corrections in the strong coupling constant.

The modeling of inclusive final state does not include any knowledge of resonances. A point-wise evaluation of Equation (9.2) therefore fails to predict the resonant structure in the hadronic invariant mass spectrum. To make sound predictions, the phase space is analyzed such that the resonances are integrated out. The prediction of such partial branching fractions can then be compared to the experiment.

The problem that the inclusive predictions can not predict the resonant structures in the hadronic invariant mass spectrum appears again in the context of event generators and is discussed in Chapter 11.

## 9.2. Exclusive Predictions

For resonant final states the hadronic matrix element for the  $b \rightarrow u$  transition can be written as

$$\langle M(p_M) | \bar{u} \gamma^\mu P_L b | B(p_B) \rangle = \sum T_i^\mu F_i(q^2), \quad (9.3)$$

where  $q = p_B - p_M$  is the four-momentum transfer in the decay and  $M$  denotes any light final state meson (I will discuss  $M \in \{\pi, \rho, \omega\}$ ). Further, the  $T_i$  denote tensorial structures of the involved 4-momenta and polarizations in case of vector boson final states, and the  $F_i$  form factors. The sum runs over all allowed<sup>1</sup> tensorial structures. The  $B^0 \rightarrow \pi^- \ell^+ \nu$  matrix element is shown as Feynman diagram in Figure 9.1b. In the diagram the arms are described by the tensorial structures  $T_i$  and the blob is described by the form factors  $F_i$ . The form factors  $F_i$  present in the equation above can not be calculated with perturbation theory in the strong coupling constant and have to be determined using non-perturbative methods.

The hadronic matrix element for the  $B \rightarrow \pi \ell \nu$  decay is given by [50]

$$\begin{aligned} \langle \pi(p_\pi) | \bar{u} \gamma^\mu P_L b | B(p_B) \rangle &= T_1^\mu f_+(q^2) + T_2^\mu f_0(q^2), \\ \text{with } T_1^\mu &= \left( (p_B + p_\pi)^\mu - \frac{m_B^2 - m_\pi^2}{q^2} q^\mu \right), \\ T_2^\mu &= \left( \frac{m_B^2 - m_\pi^2}{q^2} q^\mu \right). \end{aligned} \quad (9.4)$$

The matrix element for the  $B \rightarrow \rho \ell \nu$  ( $B \rightarrow \omega \ell \nu$ ) decay is given by [50]

$$\begin{aligned} \langle V(p, \eta) | \bar{u} \gamma^\mu P_L b | B(p_B) \rangle &= P_1^\mu F_1 + P_2^\mu F_2 + P_3^\mu F_3 + P_4^\mu F_4, \\ \text{with } P_1^\mu &= 2\epsilon_{\alpha\beta\gamma}^\mu \eta^{*\alpha} p^\beta q^\gamma, \\ P_2^\mu &= i \left( (m_B^2 - m_V^2) \eta^{*\mu} - (\eta^* q)(p + p_B)^\mu \right), \\ P_3^\mu &= i(\eta^* q) \left( q^\mu - \frac{q^2}{m_B^2 - m_V^2} (p + p_B)^\mu \right), \\ P_4^\mu &= i(\eta^* q) q^\mu, \end{aligned} \quad (9.5)$$

with the polarization vector  $\eta^{*\mu}$  and the Levi-Civita tensor  $\epsilon_{\alpha\beta\gamma}^\mu$ . The difference occurs due to that the  $\rho$  and  $\omega$  are  $S = 1$  states, what results in more tensorial structures in Equation (9.3). From these matrix elements, the differential decay rates for  $B \rightarrow M \ell \nu$  transitions can be derived. It is convenient to express the differential decay rates with the form factors in the helicity basis. The transformation is described in [50, Appendix D.1].

### 9.2.1. The $B \rightarrow \pi \ell \nu$ Decay in the SM

The differential decay rate as a function of  $q^2$  for  $B \rightarrow \pi \ell \nu$  decays is given by

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{192\pi^3 m_B^3} q^2 \sqrt{\lambda(q^2)} \left( 1 - \frac{m_\ell^2}{q^2} \right)^2 \left[ \left( 1 + \frac{m_\ell^2}{2q^2} \right) H_0^2(q^2) + \frac{3}{2} \frac{m_\ell^2}{q^2} H_t^2(q^2) \right], \quad (9.6)$$

<sup>1</sup>The  $T_i$  have to respect Lorentz invariance, and e.g. Ward gauge choices can reduce the number of terms further.

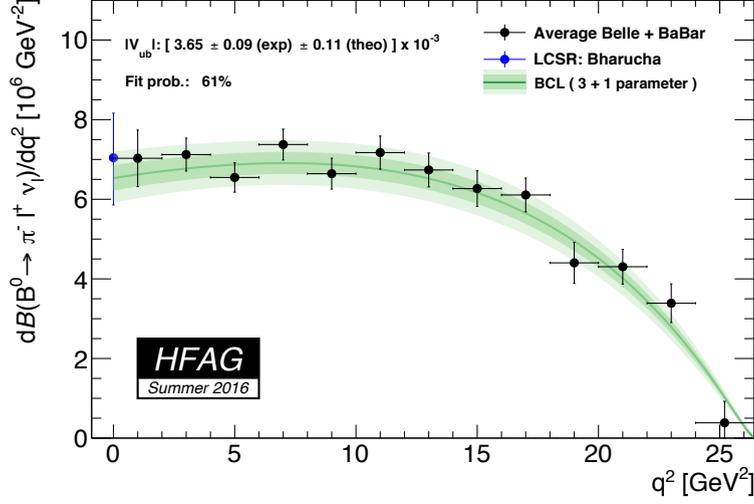


Figure 9.2.: Result for the combined fit of the LCSR predictions, LQCD predictions, BaBar measurements and Belle measurements. Taken from [36, p.102ff.].

with the helicity amplitudes

$$H_0(q^2) = \frac{\sqrt{\lambda(q^2)}}{\sqrt{q^2}} f_+(q^2), \text{ and } H_t(q^2) = \frac{m_B^2 - m_\pi^2}{\sqrt{q^2}} f_0(q^2), \quad (9.7)$$

and the Kaellen function  $\lambda(q^2) = \left( (m_B + m_M)^2 - q^2 \right) \left( (m_B - m_M)^2 - q^2 \right)$ , where  $m_M$  is the mass of the final state meson. For  $\ell = e, \mu$  the zero mass approximation  $m_\ell = 0$  yields

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2 |V_{ub}|^2}{192\pi^3 m_B^3} q^2 \sqrt{\lambda(q^2)} H_0^2(q^2), \quad (9.8)$$

which has excellent accuracy in all practical applications. Using Equation (9.8) one can combine the theory predictions with the experimental measurements. The result of such a combination is given in [36, p.102ff.] and shown in Figure 9.2. The resulting central values of the expansion parameters and their covariance matrix can also be taken from [36, p.102ff.].

### 9.2.2. The $B \rightarrow \rho/\omega \ell \nu$ Decays in the SM

The same discussion as above can be applied to  $B \rightarrow V \ell \nu$  ( $V = \omega, \rho$ ) decays. However, due to the vector structure of the meson, there are in general four helicity amplitudes describing the differential rate

$$\begin{aligned} \frac{d\Gamma}{dq^2} = & |V_{ub}| \frac{G_F^2}{192\pi^3 m_B^3} q^2 \sqrt{\lambda(q^2)} \left( 1 - \frac{m_\ell^2}{q^2} \right)^2 \\ & \times \left[ \left( 1 + \frac{m_\ell^2}{2q^2} \right) \left( H_+^2(q^2) + H_-^2(q^2) + H_0^2(q^2) \right) + \frac{3m_\ell^2}{2q^2} H_s^2(q^2) \right], \end{aligned} \quad (9.9)$$

with the helicity amplitudes

$$\begin{aligned}
H_{\pm}(q^2) &= \sqrt{\lambda(q^2)} \frac{V(q^2)}{m_B + m_V} \pm (m_B + m_V) A_1(q^2), \\
H_0(q^2) &= \frac{8m_B m_V}{\sqrt{q^2}} A_{12}(q^2), \\
H_s(q^2) &= \frac{\sqrt{\lambda(q^2)}}{\sqrt{q^2}} A_0(q^2).
\end{aligned} \tag{9.10}$$

For  $\ell = e, \mu$  the zero mass approximation  $m_\ell = 0$  yields

$$\frac{d\Gamma}{dq^2} = V_{\text{ub}} \frac{G_F^2}{192\pi^3 m_B^3} q^2 \sqrt{\lambda(q^2)} \left[ H_+^2(q^2) + H_-^2(q^2) + H_0^2(q^2) \right], \tag{9.11}$$

which again holds with excellent precision. Using Equation (9.11) one can again combine the theory prediction from light-cone sum rules (LCSR) with the experimental measurements. However, such a combination was not done before. In Chapter 10 I describe the fitting procedure I used for the combination of theory and experiment.

### 9.2.3. BCL Form Factors

The Bourrely-Caprini-Lellouch (BCL) expansion [51] is one possible Ansatz to describe the form factors in a model-independent way. It is based on a fast converging series expansion where  $q^2$  is mapped to

$$z(q^2, t_0) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}}, \tag{9.12}$$

with  $t_+ = (m_B + m_M)^2$  and  $t_0 = (m_B + m_M)(\sqrt{m_B} - \sqrt{m_M})^2$ . The  $B \rightarrow \pi$  form factors can be expanded as

$$\begin{aligned}
f_+(z) &= \frac{1}{1 - q^2/m_{B^*}^2} \sum_{n=0}^{N_z-1} b_n^+ \left( z^n - (-1)^{n-N_z} \frac{n}{N_z} z^{N_z} \right), \\
f_0(z) &= \sum_{n=0}^{N_z-1} b_n^0 z^n,
\end{aligned} \tag{9.13}$$

with the expansion coefficients  $b_n^{+/0}$  and the first pole in the spectrum  $m_{B^*}$ . The  $B \rightarrow \rho$  ( $B \rightarrow \omega$ ) form factors can be expanded as

$$F_i(q^2) = P_i(q^2) \sum_k \alpha_k^i \left( z(q^2) - z(0) \right)^k, \tag{9.14}$$

with the expansion coefficients  $\alpha_k^i$  and the poles  $P_i(q^2) = (1 - q^2/m_R^2)^{-1}$ . The mass of the resonance  $m_R$  in the pole factor is given by the first contributing resonance in the spectrum [50, Table 3].

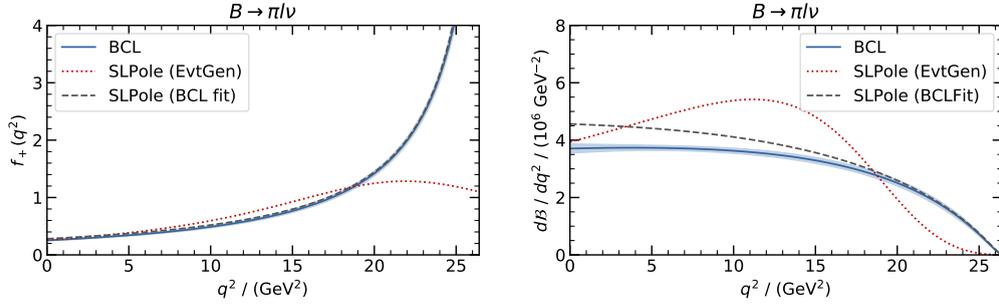


Figure 9.3.: The impact of the wrongly implemented SLPole form factor  $f_+$  on the form factor itself (left) and on the differential decay rate (right). The red dotted curve shows the prediction of the implemented model, the grey dashed curve shows the prediction of the model, when the model parameters are fitted to the prediction of the BCL model, and the blue curve shows the prediction of the BCL model with the model uncertainty of  $\pm 1\sigma$ .

Non-perturbative calculations give predictions for the differential decay rates, which can be mapped onto the BCL series expansion parameters. Predictions are available for  $B \rightarrow \pi$  transitions using light-cone sum rules (LCSR) [52],  $B \rightarrow \pi$  transitions using lattice QCD (LQCD) [5] and  $B \rightarrow \rho/\omega$  transitions using LCSR [50]. There are no LQCD predictions available for the  $B \rightarrow \rho/\omega$  transition. The light-cone expansion is limited to  $q^2 \approx \mathcal{O}(m_B \Lambda_{\text{QCD}})$ . Therefore, the LCSR predictions have to be treated cautiously for  $q^2 > 14 \text{ GeV}^2$ , where their predictive power comes solely from extrapolating from the low momentum transfer regime.

#### 9.2.4. BCL Form Factors in EvtGen

The version of the event generator EvtGen [53] used by Belle does not provide the BCL parametrization. The only available model for the  $B \rightarrow M \ell \nu$ ,  $M \in \{\pi, \rho, \omega\}$  decays is the SLPole model [54]. The form factor  $f_+$  is given by

$$\begin{aligned}
 f_+(q^2) &= \frac{f_+(0)}{1 - a(q^2/m_B^2) + b(q^2/m_B^2)^2}, & q^2 < 14 \text{ GeV}^2, \\
 f_+(q^2) &= \frac{c}{1 - q^2/m_B^{*2}}, & q^2 > 14 \text{ GeV}^2.
 \end{aligned}
 \tag{9.15}$$

The parameters of the model are chosen such that the predictions in the individual  $q^2$  ranges have a continuous transition at  $q^2 = 14 \text{ GeV}^2$ . In the implementation of the form factor model in EvtGen, the second equation was omitted and thus the first equation was used over the whole  $q^2$  range. The impact of this is illustrated in Figure 9.3, where it is clearly visible that for high  $q^2$  values, the prediction of the model fails to reproduce the correct spectrum.

To use the BCL parametrization, the data simulated with the SLPole model is reweighted

using

$$w(q^2) = \frac{\Gamma_{\text{SLPole}}}{\Gamma_{\text{BCL}}} \frac{d\Gamma_{\text{BCL}}(q^2)}{d\Gamma_{\text{SLPole}}(q^2)}, \quad (9.16)$$

with  $d\Gamma$  being the prediction of the differential decay rate via either model. The problem which arises using this equation is the following. The high  $q^2$  phase space is sparsely populated using the SLPole model as is. To cover the high  $q^2$  phase space during event generation, the SLPole model parameters are fitted to match the BCL prediction. This is illustrated in Figure 9.3. The simulated data using the modified SLPole can then be safely reweighted using Equation (9.16).

For the Belle II experiment, the BCL form factors have been implemented and are provided within the Belle II Analysis Software Framework (basf2) together with the improved coefficients determined in Chapter 10.



## 10. Improvement of the $B \rightarrow \rho \ell \nu$ and $B \rightarrow \omega \ell \nu$ Form Factors

In this chapter I present the improvement of the form factors for the  $B \rightarrow \rho \ell \nu$  and  $B \rightarrow \omega \ell \nu$  decays. The general fitting procedure is described in Section 10.1, where I first explain the extraction of the BCL expansion coefficients and second the average of the measured spectra by Belle and BaBar. The measurements of the differential branching ratios are taken from [55–57]. The results of the fits are discussed in Section 10.2.

### 10.1. Fitting Procedure

I use a  $\chi^2$  based optimization procedure to fit the BCL expansion coefficients to the measured differential decay spectra and to create an average of the different measured differential decay rates. The used  $\chi^2$  functions are inspired by the  $B \rightarrow \pi \ell \nu$  fits in [36] and explained in the following subsections for each problem separately. In all cases the covariance matrix of the optimized result is determined by calculating the Hesse matrix at the minimum of the  $\chi^2$  function.

The BaBar measurement has assumed isospin symmetry to average the measured charged and neutral decay modes.

$$\frac{\Delta\mathcal{B}(B^+ \rightarrow \rho^0 \ell^+ \bar{\nu})}{\Delta q^2} = \frac{\tau_{B^0}}{2\tau_{B^+}} \frac{\Delta\mathcal{B}(B^0 \rightarrow \rho^- \ell^+ \bar{\nu})}{\Delta q^2}. \quad (10.1)$$

To take the full systematic correlation of the Belle measurements into account, the measurement vector and its covariance matrix are constructed in the following way:

$$\begin{aligned} \mathbf{x}^{\text{Belle}} &= (\mathbf{x}_{\text{Charged}}, \mathbf{x}_{\text{Neutral}}), \\ \mathbf{C}^{\text{Belle}} &= \begin{pmatrix} \mathbf{C}_{\text{charged}}^{\text{stat}} & 0 \\ 0 & \mathbf{C}_{\text{neutral}}^{\text{stat}} \end{pmatrix} + \sum_s^{\text{sources}} \mathbf{c}_s \otimes \mathbf{c}_s, \quad \text{with } \mathbf{c}_s = \sigma_s \mathbf{x}^{\text{Belle}}, \end{aligned} \quad (10.2)$$

where  $\mathbf{x}$  is the measured spectrum,  $C$  the covariance matrix and  $\sigma_s$  the relative systematic uncertainty.

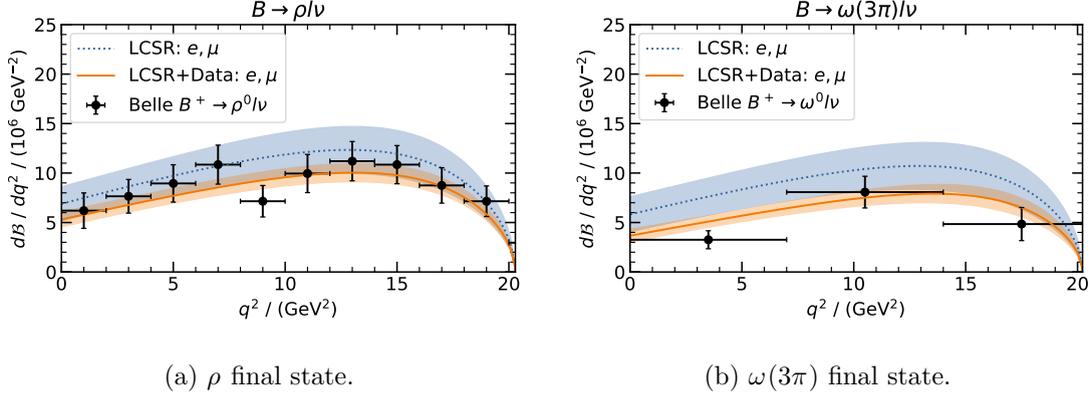


Figure 10.1.: Result of the BCL coefficient fit to the LCSR predictions and the latest Belle and BaBar measurements for the  $B \rightarrow \rho \ell \nu$  (a) and  $B \rightarrow \omega \ell \nu$  (b) final state. The blue curve shows the LCSR predictions, the orange curve the prediction of the differential rate after the fit. The data points reflect the measurements of (a)  $B^+ \rightarrow \rho^0 \ell \nu$  and (b)  $B^+ \rightarrow \omega(3\pi) \ell \nu$  by Belle.

### 10.1.1. BCL Coefficient Fit

In order to enhance the precision of the LCSR prediction over the full  $q^2$  range I performed a binned  $\chi^2$  fit. The  $\chi^2$  function is a function of  $V_{ub}$  and the LCSR expansion coefficients  $\mathbf{a}$  and has the form

$$\begin{aligned} \chi^2(V_{ub}, \mathbf{a}) &= \chi_{\text{LCSR}}^2(\mathbf{a}) + \chi_{\text{Data}}^2(V_{ub}, \mathbf{a}), \\ \chi_{\text{LCSR}}^2(\mathbf{a}) &= \Delta \mathbf{a}^T \mathbf{C}_{\text{LCSR}}^{-1} \Delta \mathbf{a}, \\ \chi_{\text{Data}}^2(V_{ub}, \mathbf{a}) &= \sum^{\text{Exp}} \Delta \mathbf{y}^T \mathbf{C}_{\text{meas}}^{-1} \Delta \mathbf{y}, \end{aligned} \quad (10.3)$$

with  $\Delta \mathbf{a} = (\mathbf{a}_{\text{LCSR}} - \mathbf{a})$  and  $\Delta \mathbf{y} = \left( \frac{\Delta \mathcal{B}_{\text{meas}}}{\Delta q^2} \right) - \left( \frac{\Delta \mathcal{B}(V_{ub}, \mathbf{a})}{\Delta q^2} \right)$ .

The theoretical prediction for the expansion coefficients  $\mathbf{a}_{\text{LCSR}}$  and their covariance matrix  $\mathbf{C}_{\text{LCSR}}$  are taken from [50].

The result of the fit is shown in Figure 10.1a (Figure 10.1b) for the  $B \rightarrow \rho \ell \nu$  ( $B \rightarrow \omega \ell \nu$ ) decay and tabulated in Table 10.1.

### 10.1.2. Legacy Spectrum

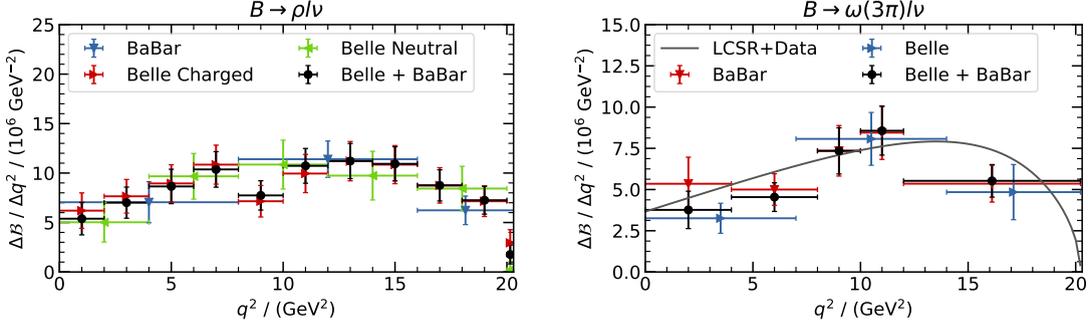
The  $B \rightarrow \rho \ell \nu$  measurements of Belle and BaBar have a compatible binning, which allows to create an averaged differential spectrum of the decay.

The average of the Belle and BaBar measurements is done using a  $\chi^2$  function of the form

$$\begin{aligned} \chi^2(\mathbf{x}^a) &= \sum_m \Delta \mathbf{y}_m^T \mathbf{C}_m^{-1} \Delta \mathbf{y}_m, \text{ with } m \in \{\text{Belle, BaBar}\}, \\ \Delta \mathbf{y}_m &= \begin{pmatrix} x_1^m - (\sum_i^N x_i^a)/N \\ \vdots \end{pmatrix}, \end{aligned} \quad (10.4)$$

Table 10.1.: Tabulated results of the BCL coefficient fit.

Parameter	$B \rightarrow \rho \ell \nu$	$B \rightarrow \omega \ell \nu$
$V_{ub}$	$(3.326 \pm 0.318) \cdot 10^{-3}$	$(3.194 \pm 0.374) \cdot 10^{-3}$
$a_0^1$	$-0.861 \pm 0.188$	$-0.963 \pm 0.284$
$a_0^2$	$1.444 \pm 1.023$	$1.849 \pm 1.199$
$a_1^0$	$0.266 \pm 0.025$	$0.242 \pm 0.031$
$a_1^1$	$0.378 \pm 0.134$	$0.269 \pm 0.217$
$a_1^2$	$0.165 \pm 0.407$	$-0.039 \pm 0.549$
$a_{12}^0$	$0.291 \pm 0.031$	$0.250 \pm 0.038$
$a_{12}^1$	$0.718 \pm 0.167$	$0.517 \pm 0.239$
$a_{12}^2$	$0.384 \pm 0.700$	$-0.045 \pm 0.957$
$v^0$	$0.331 \pm 0.030$	$0.304 \pm 0.038$
$v^1$	$-0.876 \pm 0.177$	$-0.920 \pm 0.266$
$v^2$	$1.907 \pm 0.936$	$1.901 \pm 1.183$
$t_1^0$	$0.274 \pm 0.025$	$0.251 \pm 0.031$
$t_1^1$	$-0.755 \pm 0.139$	$-0.784 \pm 0.205$
$t_1^2$	$1.525 \pm 0.755$	$1.578 \pm 0.957$
$t_2^1$	$0.456 \pm 0.129$	$0.344 \pm 0.208$
$t_2^2$	$0.591 \pm 0.464$	$0.347 \pm 0.551$
$t_{23}^0$	$0.739 \pm 0.072$	$0.646 \pm 0.088$
$t_{23}^1$	$1.831 \pm 0.396$	$1.357 \pm 0.581$
$t_{23}^2$	$2.912 \pm 1.786$	$2.000 \pm 2.174$



- (a) The average of the spectrum can be performed model independent due to the matching bin borders.
- (b) The model line represents the model used for averaging/splitting the bins where the bin borders do not match.

Figure 10.2.: The averaged  $B \rightarrow \rho \ell \nu$  (a) and  $B \rightarrow \omega \ell \nu$  (b) spectra on top of the latest Belle and BaBar measurements. The isospin transformation is applied to the  $B^0 \rightarrow V^- \ell^+ \nu$  measurements. The black data points represent the "Legacy Spectra".

where  $\mathbf{C}_m$  is the covariance of the measurement and  $x_i^m$  the measured differential branching ratio in bin  $i$ . Further  $\mathbf{x}^a$  denotes the averaged spectrum and  $N$  the number of bins used to map to a measured bin. The binning of the average vector is chosen to match the most granular spectrum. The averaged spectrum with the used measurements is shown in Figure 10.2a and tabulated in Table 10.2.

The  $B \rightarrow \omega \ell \nu$  measurements of Belle and BaBar do not have a compatible binning. To create an average of the spectrum, the result of Section 10.1.1 is used as model to split the individual bins where necessary. To match the average bin onto a measurement without matching bin borders, the average bin  $\mathbf{x}_i^a$  is split in the following way:

$$\mathbf{x}_i^a(\text{left}) = I_{\text{left}}/I(1 + \Theta_j \varepsilon_{j,1}), \quad (10.5)$$

$$\mathbf{x}_i^a(\text{right}) = I_{\text{right}}/I(1 - \Theta_j \varepsilon_{j,2}), \quad (10.6)$$

where  $I_{\text{left}}$  ( $I_{\text{right}}$ ) is the area under the model function for the required bin borders,  $I = I_{\text{left}} + I_{\text{right}}$  the area under the model function of the bin,  $\varepsilon$  the uncertainty of the integration given by the model uncertainty and  $\Theta$  the nuisance parameter for the model dependence. The splitting has to be done for two bins. The result should not depend on the nuisance parameters. The averaged spectrum with the used measurements is shown in Figure 10.2b and tabulated in Table 10.2. The result is indeed almost independent of the nuisance parameters, which can be seen in the correlation matrix of the fit shown in Figure 10.3.

## 10.2. Results

The fit of the BCL expansion coefficients to the LCSR predictions and the data allows for an improved prediction of the differential rate of  $B \rightarrow V$ ,  $V = \{\rho, \omega\}$  transitions. The

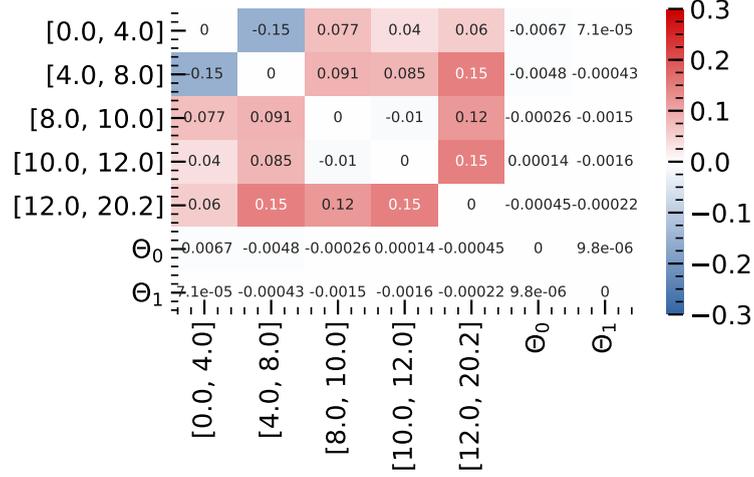


Figure 10.3.: Correlation matrix of the  $B \rightarrow \omega l \nu$  spectrum average fit. The diagonal components are set to 0. The intervals correspond to the  $q^2$  bins, the  $\Theta$ 's to the nuisance parameters. The correlation matrix shows that the result is almost independent of the assumed model of the differential decay rate.

Table 10.2.: Tabulated results of the legacy spectra.

$B \rightarrow \rho l \nu$		$B \rightarrow \omega l \nu$	
$q^2$ bin (GeV)	$\Delta\mathcal{B}/\Delta q^2 \cdot 10^6$	$q^2$ bin (GeV)	$\Delta\mathcal{B}/\Delta q^2 \cdot 10^6$
[0, 2]	$5.386 \pm 1.636$	[0, 4]	$3.765 \pm 1.134$
[2, 4]	$7.007 \pm 1.573$	[4, 8]	$4.540 \pm 0.870$
[4, 6]	$8.647 \pm 1.740$	[8, 10]	$7.345 \pm 1.400$
[6, 8]	$10.367 \pm 1.798$	[10, 12]	$8.565 \pm 1.475$
[8, 10]	$7.740 \pm 1.481$	[12, 20.2]	$5.527 \pm 0.986$
[10, 12]	$10.728 \pm 1.757$	Nuisance parameters	
[12, 14]	$11.215 \pm 1.775$	$\Theta_0$	$-0.011 \pm 1.000$
[14, 16]	$10.931 \pm 1.724$	$\Theta_1$	$0.001 \pm 1.000$
[16, 18]	$8.753 \pm 1.576$		
[18, 20]	$7.254 \pm 1.407$		
[20, 20.3]	$1.768 \pm 0.953$		

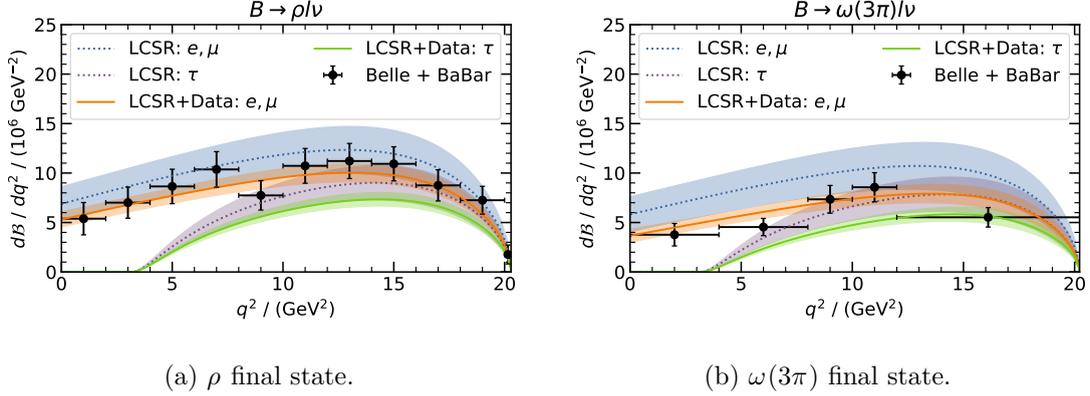


Figure 10.4.: The differential rates of  $B \rightarrow V$  transitions before and after the fit is performed. On the left (a) for the  $\rho$  in the final state and on the right (b) for the  $\omega(3\pi)$  in the final state. The dotted blue and purple lines show the prefit rates, the solid orange and green lines show the postfit rates. The error band corresponds to the  $1\sigma$  interval.

uncertainty was reduced by a factor of  $\approx 2.5$  over the whole  $q^2$  range. The result is illustrated in Figure 10.4. It allows for a significant reduction of the systematic uncertainties in measurements where these transitions are part of the background composition. Additionally, a "legacy spectrum" was created for each final state which allows for easy comparison with future measurements of the differential decay rates, e.g. at Belle II.

Usually, the value of  $|V_{ub}|$  for  $B \rightarrow V$  transitions is only extracted at low  $q^2$  values, due to the fact that the LCSR prediction are only valid for low  $q^2$ . Here, the value of  $|V_{ub}|$  was extracted during the fit over the whole  $q^2$  range and was found to be  $|V_{ub}|_\rho = (3.326 \pm 0.318) \cdot 10^{-3}$  and  $|V_{ub}|_\omega = (3.194 \pm 0.374) \cdot 10^{-3}$ . The result is compatible with the exclusive  $|V_{ub}|_\pi$  measurement from  $B \rightarrow \pi \ell \nu$  decays. The comparison to the current PDG world averages is shown in Figure 10.5.

By integrating the differential rates given in Equation (9.11) and Equation (9.9) over the allowed kinematic range in  $q^2$ , predictions for  $R(\rho)$  and  $R(\omega)$  can be obtained:

$$R(V) = \frac{\mathcal{B}(B \rightarrow V \tau \nu)}{\mathcal{B}(B \rightarrow V \ell \nu)} = \frac{\int_{m_V^2}^{q_{\max}^2} dq^2 d\mathcal{B}(B \rightarrow V \tau \nu)/dq^2}{\int_0^{q_{\max}^2} dq^2 \mathcal{B}(B \rightarrow V \ell \nu)/dq^2}, \quad (10.7)$$

with  $q_{\max}^2 = m_B^2 + m_V^2 - 2m_B m_V$  and the zero mass approximation for the light leptons. The physical constants  $V_{ub}$  and  $G_F$  cancel in this ratio. Using the improved BCL coefficients from Table 10.1 the following predictions for the SM values are obtained:

$$\begin{aligned} R(\rho) &= 0.536 \pm 0.009, \\ R(\omega) &= 0.546 \pm 0.015. \end{aligned} \quad (10.8)$$

In the same manner, but with the lower bound for both integrals in Equation (10.7) chosen

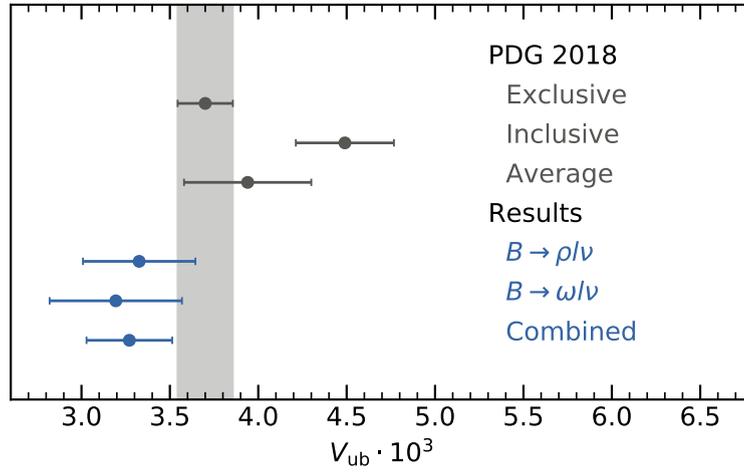


Figure 10.5.: The values of  $|V_{ub}|$  extracted from the fit compared to the current world averages. The combined result is a weighted mean of the two measurements. The exclusive PDG value comes from  $B \rightarrow \pi \ell \nu$  measurements.

to be the  $\tau$  mass squared  $m_\tau^2$ , the ratios are:

$$\begin{aligned} R_r(\rho) &= 0.571 \pm 0.007, \\ R_r(\omega) &= 0.577 \pm 0.013. \end{aligned} \tag{10.9}$$

In comparison to the prediction from LCSR alone, the uncertainties are reduced by 10-20%, when the form factors are extracted over the whole  $q^2$  range and used to calculate the ratio.



## 11. $B \rightarrow X_u \ell \nu$ Hybrid MC

As discussed in Chapter 9, the decay  $B \rightarrow X_u \ell \nu$  is treated depending if one is interested in an exclusive resonant final state or the inclusive rate, including all possible resonant and non-resonant final states. The predicted inclusive rate includes the rates of the resonant contributions, but the resonant structure in the hadronic invariant mass spectrum is lost.

To simulate a data sample which properly reflects the predictions of the rates and the resonant structures in the hadronic invariant mass spectrum  $m_X$ , the lepton energy  $E_l$ , and the momentum transfer  $q^2$ , the inclusive and exclusive predictions have to be properly combined.

For the search for the decay  $B \rightarrow \mu \nu_\mu$ , described in Part I, the modeling of the endpoint of the  $E_l$  spectrum has to be well described. This is where the light resonances contribute the most. In Section 11.2 I discuss briefly the model previously used by the Belle experiment and discuss the need for an alternative model for the endpoint of the lepton momentum spectrum of the  $B \rightarrow X_u \ell \nu$  decays. In Section 11.1 I discuss the Hybrid model and compare the two different models. The principle behind each model is to *fix* the resonant rates and to *modify* the inclusive rate. This is achieved by "subtracting" the resonances from the inclusive rate. This subtracting is done heuristically.

### 11.1. Hybrid Model Design

The Hybrid model originates from the CLEO experiment [58] and is also used by BaBar.

The inclusive distribution described by Equation (9.2) is split into the bins

$$\begin{aligned} m_X &= [0.00, 1.40, 1.60, 1.80, 2.00, 2.50, 3.00, 3.50], \\ q^2 &= [0.00, 2.50, 5.00, 7.50, 10.0, 12.5, 15.0, 20.0, 25.0], \\ E_l^B &= [0.00, 0.50, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 3.00]. \end{aligned} \tag{11.1}$$

The bins are theoretically motivated. The inclusive predictions are only reliable in bins with the resonances integrated out. Therefore the Hybrid model is based on the given bins. The event generator does however perform a point-wise evaluation of the decay rate. For the creation of the Hybrid model, this is ignored and the aforementioned bins are used.

The idea of the Hybrid method is to subtract, in each bin, from the inclusive prediction the partial rates of the resonant contributions. The remainder then only predicts the decay

Table 11.1.: Branching ratios and models of  $B \rightarrow X_u \ell \nu$ ,  $\ell = e, \mu$  decays. The given branching ratios and models are used to create the Hybrid Model. Values taken from [4].

	$\rightarrow \pi \ell \nu$	$\rightarrow \eta \ell \nu$	$\rightarrow \eta' \ell \nu$	$\rightarrow \omega \ell \nu$	$\rightarrow \rho \ell \nu$	$\rightarrow u \ell \nu$ incl.
$B^+$	$7.8 \cdot 10^{-5}$	$3.9 \cdot 10^{-5}$	$2.3 \cdot 10^{-5}$	$1.19 \cdot 10^{-4}$	$1.58 \cdot 10^{-4}$	$2.2 \cdot 10^{-3}$
$B^0$	$1.5 \cdot 10^{-4}$	-	-	-	$2.94 \cdot 10^{-4}$	$2.0 \cdot 10^{-3}$
Model	BCL	ISGW2	ISGW2	BCL	BCL	DFN

rate of the non-resonant final states. The resonant predictions are then added using the exclusive predictions.

The Hybrid model predicts a total number of events in a given bin  $i$  by

$$H_i = R_i + \omega_i I_i, \quad (11.2)$$

where the  $H_i$  is the total number of events in bin  $i$ ,  $R_i$  is the number of resonant contributions,  $I_i$  the number of events contributed by the inclusive distribution, and  $\omega_i = (I_i - R_i)/I_i$  the weight assigned to the inclusive prediction such that  $H_i = I_i$ .

In Table 11.1, the branching ratios for the resonant and inclusive  $B \rightarrow X_u \ell \nu$  decays are tabulated, together with the decay model which was used to simulate each individual contribution.

The resulting distribution of the Hybrid model is shown in Figure 11.1.

## 11.2. Comparison to Belle's Model

In the  $B \rightarrow X_u \ell \nu$  model previously used by the Belle experiment, the subtraction is done by adjusting the normalization and the moments of the inclusive distribution. The resulting distribution for the semileptonic  $b \rightarrow u$  decay of a charged  $B^+$  meson is shown in Figure 11.2.

The difference in the reconstructed endpoint of the lepton momentum spectrum in the context of the analysis described in Part I is shown in Figure 11.3. Belle's model estimates much less events in the kinematic endpoint. The impact of this is shown in Figure 11.4, where Belle's model drastically underestimates the contribution from semileptonic  $B \rightarrow X_u \ell \nu$  decays.

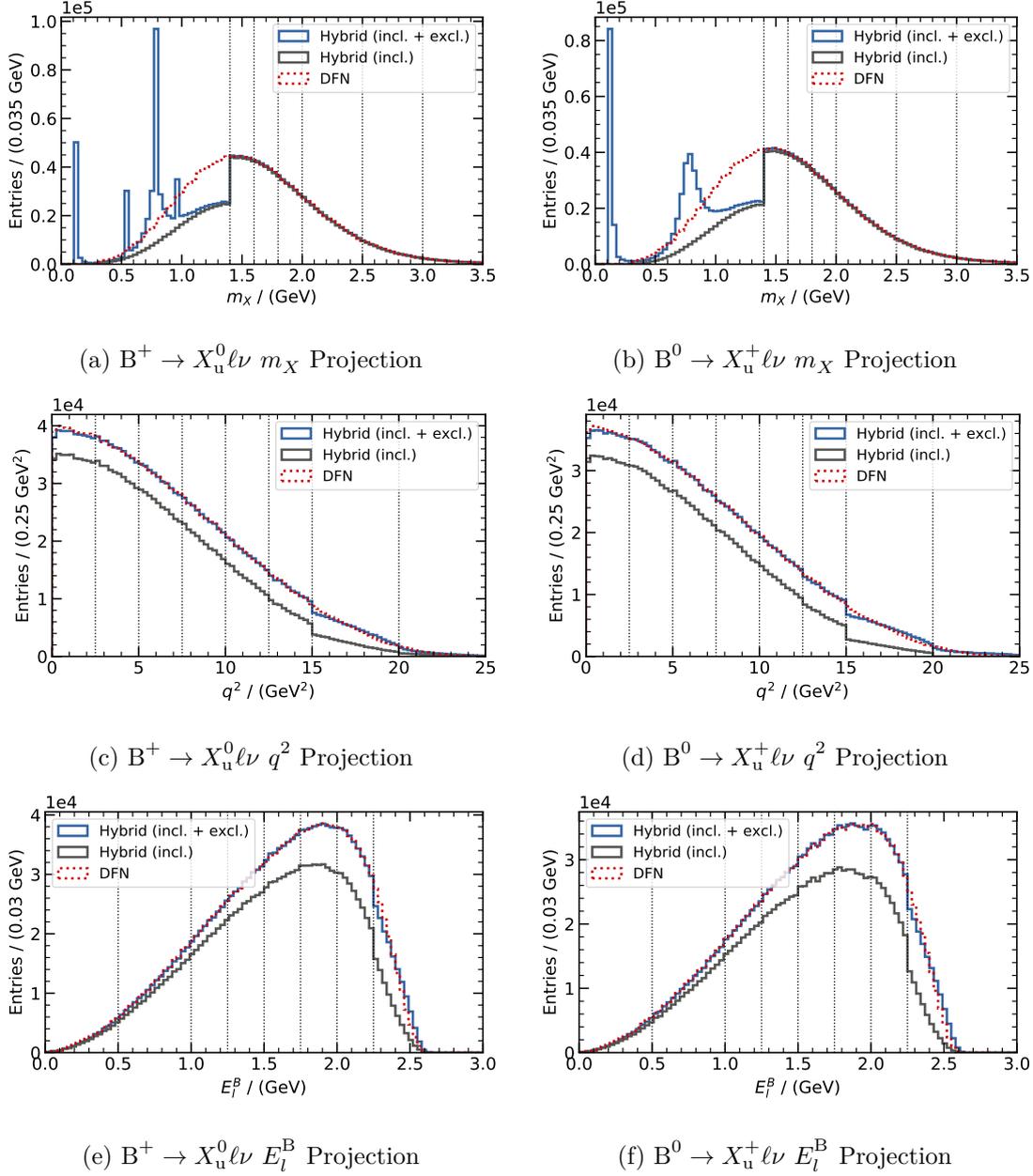


Figure 11.1.: This figure show the projections of the triple differential decay rate for inclusive  $B \rightarrow X_u \ell \nu$  decays for charged (left) and mixed (right) B mesons. The vertical gray lines indicate the chosen binning for calculating the Hybrid weights. The red histograms show the full inclusive prediction, the gray histograms show the inclusive predictions after application of the Hybrid weights and the blue histograms show the full Hybrid model which consists of the reweighted inclusive component and the resonant contributions.

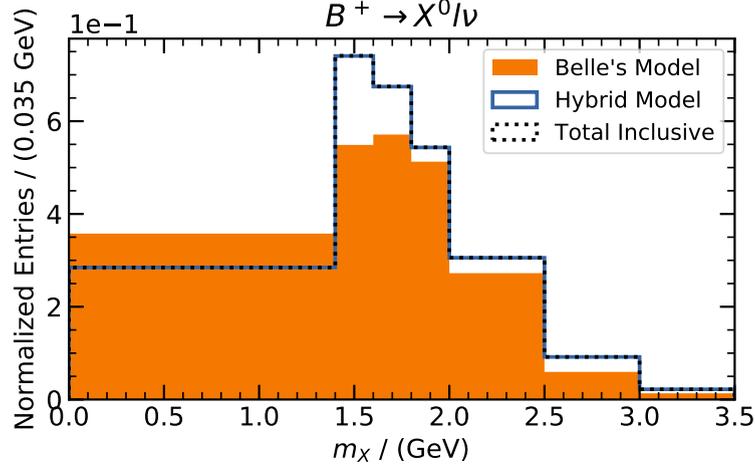


Figure 11.2.: Comparison of the inclusive prediction between Belle's  $B \rightarrow X_u \ell \nu$  model to the newly deployed Hybrid model. The figure shows the total  $B \rightarrow X_u \ell \nu$  MC including the resonances in the binning given in Equation (11.1). The black dotted line is the prediction of the inclusive decay rate given by Equation (9.2). The perfect match of the inclusive prediction and the Hybrid MC is by design.

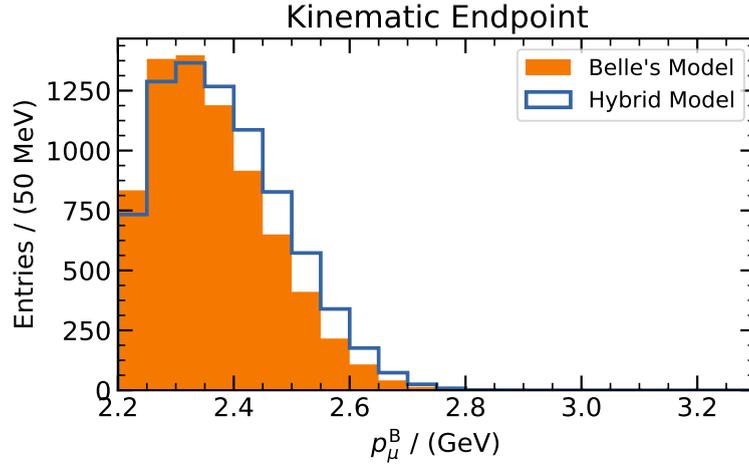


Figure 11.3.: Comparison of kinematic endpoint in the lepton momentum spectrum between Belle's  $B \rightarrow X_u \ell \nu$  model to the newly deployed Hybrid model. The Hybrid model predicts a larger number of events in the very endpoint of the distribution.

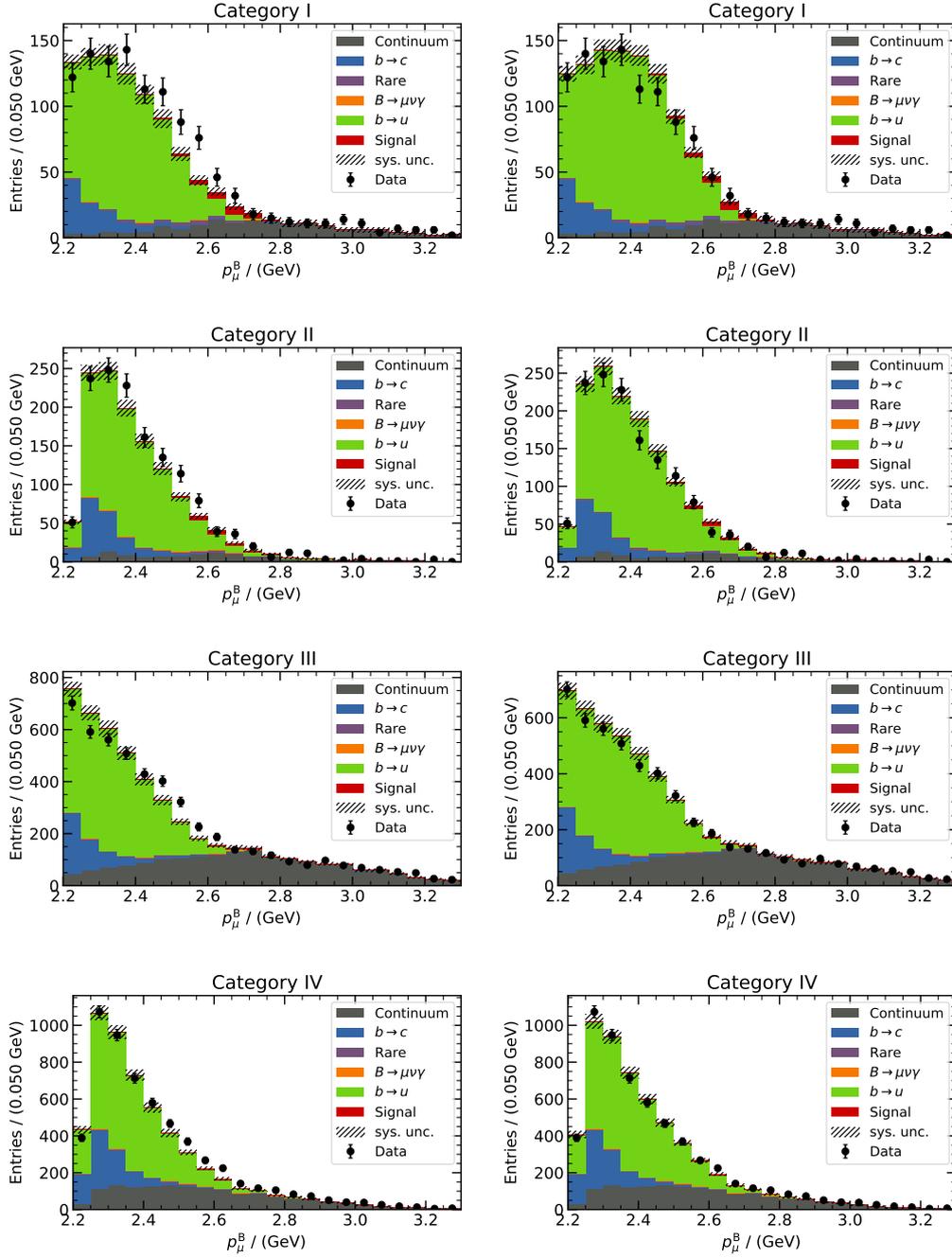


Figure 11.4.: Comparison of Belle's  $B \rightarrow X_u \ell \nu$  model to the newly deployed Hybrid model at the kinematic endpoint of the lepton momentum spectrum (the green template in all plots). The figures show the distributions of the individual MC components and data before any fit on the data is performed. Belle's model (left) does underestimate the rate at the kinematic endpoint in comparison to the Hybrid model (right). For a discussion of the categorization and the individual templates see Chapter 6.



## 12. Summary

In this thesis, the search for the rare decay  $B \rightarrow \mu\nu$  in the full data sample recorded by the Belle experiment was presented. The branching ratio was found to be

$$\mathcal{B}(B \rightarrow \mu\nu_\mu) = (5.297 \pm 2.169) \cdot 10^{-7}.$$

with a significance of  $2.75\sigma$  including statistical and systematic uncertainties and  $3.11\sigma$  including only statistical uncertainties. It is the most precise measurement of this decay to date, improving the previous result of the Belle collaboration [6] by  $0.3\sigma$ . The significance of the observed signal is not high enough to claim evidence for the decay, therefore the upper limit for the branching ratio was determined in the Frequentist and Bayesian interpretation:

$$\text{Frequentist : } \mathcal{B}(B \rightarrow \mu\nu_\mu) < 8.64 \cdot 10^{-7} \text{ @ 90\% CL,}$$

$$\text{Bayesian : } \mathcal{B}(B \rightarrow \mu\nu_\mu) < 8.86 \cdot 10^{-7} \text{ @ 90\% CL.}$$

The measured branching ratio was also interpreted in the context of a type-II and type-III two Higgs doublet model and exclusion limits for the parameters of those models were determined. Additionally, a search for a sterile neutrino in the same final state was performed by scanning the measured muon momentum spectrum. No significant excess for a sterile neutrino over the background only hypothesis was found.

The prediction of the differential decay rate of the  $B \rightarrow \rho\ell\nu$  and  $B \rightarrow \omega\ell\nu$  was improved by extracting the expansion coefficients of the form factor model from previously measured  $q^2$  spectra of Belle and BaBar. This reduces the uncertainty on the prediction of the differential decay rate by a factor of 2 for the whole  $q^2$  range.

With the improved determination of the form factors the prediction for the ratio of semitauonic to semileptonic  $b \rightarrow u$  decays  $R(V) = \frac{\mathcal{B}(B \rightarrow V\tau\nu)}{\mathcal{B}(B \rightarrow V\ell\nu)}$ , where  $V \in \{\rho, \omega\}$ , could be improved by 10%:

$$R(\rho) = 0.536 \pm 0.009,$$

$$R(\omega) = 0.546 \pm 0.015.$$

A new model for the simulation of  $B \rightarrow X_u\ell\nu$  decays was introduced to the Belle and Belle II collaborations. This was essential for a correct background modeling for the search for the  $B \rightarrow \mu\nu_\mu$  decays. The new model will also impact other measurements, such as the ongoing inclusive measurement of the  $B \rightarrow X_u\ell\nu$  branching ratio.

The SM parameter  $|V_{ub}|$  was extracted from the measured branching ratio of the  $B \rightarrow \mu\nu_\mu$  decay:

$$|V_{ub}|_{B \rightarrow \mu\nu_\mu} = (4.37_{-1.01}^{+0.82}) \cdot 10^{-3}.$$

Additionally,  $|V_{ub}|$  was also extracted in the fit of the form factor expansion coefficients for the  $B \rightarrow \rho\ell\nu$  and  $B \rightarrow \omega\ell\nu$  decays. The extracted values are:

$$\begin{aligned} |V_{ub}|_{B \rightarrow \rho\ell\nu} &= (3.33 \pm 0.32) \cdot 10^{-3}, \\ |V_{ub}|_{B \rightarrow \omega\ell\nu} &= (3.19 \pm 0.37) \cdot 10^{-3}. \end{aligned}$$

# Bibliography

- [1] **ATLAS**, G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.* **B716** (2012) 1–29, [arXiv:1207.7214 \[hep-ex\]](#).
- [2] **CMS**, S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.* **B716** (2012) 30–61, [arXiv:1207.7235 \[hep-ex\]](#).
- [3] **Belle**, **BaBar**, A. J. Bevan *et al.*, “The Physics of the B Factories,” *Eur. Phys. J.* **C74** (2014) 3026, [arXiv:1406.6311 \[hep-ex\]](#).
- [4] **Particle Data Group**, M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka, K. Agashe, G. Aielli, C. AMSler, M. Antonelli, D. M. Asner, H. Baer, S. Banerjee, R. M. Barnett, T. Basaglia, C. W. Bauer, J. J. Beatty, V. I. Belousov, J. Beringer, S. Bethke, A. Bettini, H. Bichsel, O. Biebel, K. M. Black, E. Blucher, O. Buchmuller, V. Burkert, M. A. Bychkov, R. N. Cahn, M. Carena, A. Ceccucci, A. Cerri, D. Chakraborty, M.-C. Chen, R. S. Chivukula, G. Cowan, O. Dahl, G. D’Ambrosio, T. Damour, D. de Florian, A. de Gouvêa, T. DeGrand, P. de Jong, G. Dissertori, B. A. Dobrescu, M. D’Onofrio, M. Doser, M. Drees, H. K. Dreiner, D. A. Dwyer, P. Eerola, S. Eidelman, J. Ellis, J. Erler, V. V. Ezhela, W. Fetscher, B. D. Fields, R. Firestone, B. Foster, A. Freitas, H. Gallagher, L. Garren, H.-J. Gerber, G. Gerbier, T. Gershon, Y. Gershtein, T. Gherghetta, A. A. Godizov, M. Goodman, C. Grab, A. V. Gritsan, C. Grojean, D. E. Groom, M. Grünewald, A. Gurtu, T. Gutsche, H. E. Haber, C. Hanhart, S. Hashimoto, Y. Hayato, K. G. Hayes, A. Hebecker, S. Heinemeyer, B. Heltsley, J. J. Hernández-Rey, J. Hisano, A. Höcker, J. Holder, A. Holtkamp, T. Hyodo, K. D. Irwin, K. F. Johnson, M. Kado, M. Karliner, U. F. Katz, S. R. Klein, E. Klempt, R. V. Kowalewski, F. Krauss, M. Kreps, B. Krusche, Y. V. Kuyanov, Y. Kwon, O. Lahav, J. Laiho, J. Lesgourgues, A. Liddle, Z. Ligeti, C.-J. Lin, C. Lippmann, T. M. Liss, L. Littenberg, K. S. Lugovsky, S. B. Lugovsky, A. Lusiani, Y. Makida, F. Maltoni, T. Mannel, A. V. Manohar, W. J. Marciano, A. D. Martin, A. Masoni, J. Matthews, U.-G. Meißner, D. Milstead, R. E. Mitchell, K. Mönig, P. Molaro, F. Moortgat, M. Moskvic, H. Murayama, M. Narain, P. Nason, S. Navas, M. Neubert, P. Nevski, Y. Nir, K. A. Olive, S. Pagan Griso, J. Parsons, C. Patrignani, J. A. Peacock, M. Pennington, S. T. Petcov, V. A. Petrov, E. Pianori, A. Piepke, A. Pomarol, A. Quadt, J. Rademacker, G. Raffelt, B. N. Ratcliff, P. Richardson, A. Ringwald, S. Roesler, S. Rolli, A. Romaniouk, L. J. Rosenberg, J. L. Rosner, G. Rybka, R. A.

- Ryutin, C. T. Sachrajda, Y. Sakai, G. P. Salam, S. Sarkar, F. Sauli, O. Schneider, K. Scholberg, A. J. Schwartz, D. Scott, V. Sharma, S. R. Sharpe, T. Shutt, M. Silari, T. Sjöstrand, P. Skands, T. Skwarnicki, J. G. Smith, G. F. Smoot, S. Spanier, H. Spieler, C. Spiering, A. Stahl, S. L. Stone, T. Sumiyoshi, M. J. Syphers, K. Terashi, J. Terning, U. Thoma, R. S. Thorne, L. Tiator, M. Titov, N. P. Tkachenko, N. A. Törnqvist, D. R. Tovey, G. Valencia, R. Van de Water, N. Varelas, G. Venanzoni, L. Verde, M. G. Vincter, P. Vogel, A. Vogt, S. P. Wakely, W. Walkowiak, C. W. Walter, D. Wands, D. R. Ward, M. O. Wascko, G. Weiglein, D. H. Weinberg, E. J. Weinberg, M. White, L. R. Wiencke, S. Willocq, C. G. Wohl, J. Womersley, C. L. Woody, R. L. Workman, W.-M. Yao, G. P. Zeller, O. V. Zenin, R.-Y. Zhu, S.-L. Zhu, F. Zimmermann, P. A. Zyla, J. Anderson, L. Fuller, V. S. Lugovsky, and P. Schaffner, “Review of particle physics,” *Phys. Rev. D* **98** (Aug, 2018) 030001. <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [5] S. Aoki *et al.*, “Review of lattice results concerning low-energy particle physics,” *Eur. Phys. J. C* **77** no. 2, (2017) 112, [arXiv:1607.00299](https://arxiv.org/abs/1607.00299) [hep-lat].
- [6] **Belle**, A. Sibidanov *et al.*, “Search for  $B^- \rightarrow \mu^- \bar{\nu}_\mu$  Decays at the Belle Experiment,” *Phys. Rev. Lett.* **121** no. 3, (2018) 031801, [arXiv:1712.04123](https://arxiv.org/abs/1712.04123) [hep-ex].
- [7] B. Kronenbitter, *Measurement of the branching fraction of  $B^+ \rightarrow \tau^+ \nu_\tau$  decays at the Belle experiment*. Ph.d. thesis, KIT, 2014. <https://ekp-invenio.physik.uni-karlsruhe.de/record/48604>. KIT, Diss., 2014.
- [8] N. Cabibbo, “Unitary Symmetry and Leptonic Decays,” *Phys. Rev. Lett.* **10** (1963) 531–533.
- [9] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction,” *Prog. Theor. Phys.* **49** (1973) 652–657.
- [10] M. Gelb, *Search for the Rare Decay  $B^+ \rightarrow \ell^+ \nu_\ell \gamma$  with the Full Event Interpretation at the Belle Experiment*. Ph.d. thesis, KIT, 2018. <https://ekp-invenio.physik.uni-karlsruhe.de/record/49050>. KIT, Diss., 2018.
- [11] M. Beneke and J. Rohrwild, “B meson distribution amplitude from  $B \rightarrow \gamma l \nu$ ,” *Eur. Phys. J. C* **71** (2011) 1818, [arXiv:1110.3228](https://arxiv.org/abs/1110.3228) [hep-ph].
- [12] M. Beneke, V. M. Braun, Y. Ji, and Y.-B. Wei, “Radiative leptonic decay  $B \rightarrow \gamma e \nu_e$  with subleading power corrections,” *JHEP* **07** (2018) 154, [arXiv:1804.04962](https://arxiv.org/abs/1804.04962) [hep-ph].
- [13] **Belle**, M. Gelb *et al.*, “Search for the rare decay of  $B^+ \rightarrow \ell^+ \nu_\ell \gamma$  with improved hadronic tagging,” [arXiv:1810.12976](https://arxiv.org/abs/1810.12976) [hep-ex].
- [14] P. Golonka and Z. Was, “PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays,” *Eur. Phys. J. C* **45** (2006) 97–107, [arXiv:hep-ph/0506026](https://arxiv.org/abs/hep-ph/0506026) [hep-ph].
- [15] D. Becirevic, B. Haas, and E. Kou, “Soft Photon Problem in Leptonic B-decays,” *Phys. Lett. B* **681** (2009) 257–263, [arXiv:0907.1845](https://arxiv.org/abs/0907.1845) [hep-ph].

- [16] G. Burdman, J. T. Goldman, and D. Wyler, “Radiative leptonic decays of heavy mesons,” *Phys. Rev.* **D51** (1995) 111–117, [arXiv:hep-ph/9405425 \[hep-ph\]](#).
- [17] G. Chiladze, A. F. Falk, and A. A. Petrov, “Radiative leptonic  $B_c$  decays in effective field theory,” *Phys. Rev.* **D60** (1999) 034011, [arXiv:hep-ph/9811405 \[hep-ph\]](#).
- [18] T. Goldman and W. J. Wilson, “Radiative corrections to leptonic decays of charged pseudoscalar mesons,” *Phys. Rev. D* **15** (Feb, 1977) 709–724. <https://link.aps.org/doi/10.1103/PhysRevD.15.709>.
- [19] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens, and O. Ruchayskiy, “Sterile Neutrino Dark Matter,” *Prog. Part. Nucl. Phys.* **104** (2019) 1–45, [arXiv:1807.07938 \[hep-ph\]](#).
- [20] E. K. Akhmedov, “Neutrino physics,” in *Proceedings, Summer School in Particle Physics: Trieste, Italy, June 21-July 9, 1999*, pp. 103–164. 1999. [arXiv:hep-ph/0001264 \[hep-ph\]](#).
- [21] A. Crivellin, C. Greub, and A. Kokulu, “Explaining  $B \rightarrow D\tau\nu$ ,  $B \rightarrow D^*\tau\nu$  and  $B \rightarrow \tau\nu$  in a 2HDM of type III,” *Phys. Rev.* **D86** (2012) 054014, [arXiv:1206.2634 \[hep-ph\]](#).
- [22] D. Besson and T. Skwarnicki, “ $\Upsilon$  spectroscopy,” *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 333–378.
- [23] Wikipedia contributors, “KEKB (accelerator),” 2018. [https://en.wikipedia.org/wiki/KEKB\\_\(accelerator\)](https://en.wikipedia.org/wiki/KEKB_(accelerator)).
- [24] A. Abashian *et al.*, “The Belle Detector,” *Nucl. Instrum. Meth.* **A479** (2002) 117–232.
- [25] T. Keck, *Machine learning algorithms for the Belle II experiment and their validation on Belle data*. Ph.d. thesis, Karlsruher Institut für Technologie, 2017. <https://ekp-invenio.physik.uni-karlsruhe.de/record/48940>. Karlsruher Institut für Technologie, Diss., 2017.
- [26] T. Keck *et al.*, “The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment,” [arXiv:1807.08680 \[hep-ex\]](#).
- [27] H. Nakano, A. Ishikawa, and K. Sumisawa, “Ks selection with NeuroBayes and nisKsFinder class,” *Belle Note 1253 (internal)*.
- [28] T. Keck, “FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment,” *Computing and Software for Big Science* **1** no. 1, (Sep, 2017) 2. <https://doi.org/10.1007/s41781-017-0002-8>.
- [29] M. Röhrken, “Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charmed Decays at the Japanese Belle Experiment,” <https://ekp-invenio.physik.uni-karlsruhe.de/record/48212>.
- [30] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr, “Advanced event reweighting using multivariate analysis,” *J. Phys. Conf. Ser.* **368** (2012) 012028.
- [31] E. Jones, T. Oliphant, P. Peterson, *et al.*, “SciPy: Open source scientific tools for Python,” 2001–2019. <http://www.scipy.org/>. [Online; accessed 2019-01-07].

- [32] J. Neyman and E. S. Pearson, “Ix. on the problem of the most efficient tests of statistical hypotheses,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **231** no. 694-706, (1933) 289–337. <http://rsta.royalsocietypublishing.org/content/231/694-706/289>.
- [33] S. S. Wilks, “The large-sample distribution of the likelihood ratio for testing composite hypotheses,” *Ann. Math. Statist.* **9** no. 1, (03, 1938) 60–62. <https://doi.org/10.1214/aoms/1177732360>.
- [34] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J.* **C71** (2011) 1554, [arXiv:1007.1727 \[physics.data-an\]](https://arxiv.org/abs/1007.1727). [Erratum: *Eur. Phys. J.* **C73**,2501(2013)].
- [35] **BaBar**, J. P. Lees *et al.*, “Measurement of the inclusive electron spectrum from B meson decays and determination of  $|V_{ub}|$ ,” *Phys. Rev.* **D95** no. 7, (2017) 072001, [arXiv:1611.05624 \[hep-ex\]](https://arxiv.org/abs/1611.05624).
- [36] Y. Amhis, S. Banerjee, E. Ben-Haim, F. Bernlochner, A. Bozek, C. Bozzi, M. Chrzaszcz, J. Dingfelder, S. Duell, M. Gersabeck, T. Gershon, D. Gerstel, P. Goldenzweig, R. Harr, K. Hayasaka, H. Hayashii, M. Kenzie, T. Kuhr, O. Leroy, A. Lusiani, X. R. Lyu, K. Miyabayashi, P. Naik, T. Nanut, A. Oyanguren Campos, M. Patel, D. Pedrini, M. Petric, M. Rama, M. Roney, M. Rotondo, O. Schneider, C. Schwanda, A. J. Schwartz, J. Serrano, B. Shwartz, R. Tesarek, D. Tonelli, K. Trabelsi, P. Urquijo, R. Van Kooten, J. Yelton, and A. Zupanc, “Averages of b-hadron, c-hadron, and  $\tau$ -lepton properties as of summer 2016,” *The European Physical Journal C* **77** no. 12, (Dec, 2017) 895. <https://doi.org/10.1140/epjc/s10052-017-5058-4>.
- [37] D. Bigi, P. Gambino, and S. Schacht, “A fresh look at the determination of  $|V_{cb}|$  from  $B \rightarrow D^* \ell \nu$ ,” *Phys. Lett.* **B769** (2017) 441–445, [arXiv:1703.06124 \[hep-ph\]](https://arxiv.org/abs/1703.06124).
- [38] B. Grinstein and A. Kobach, “Model-Independent Extraction of  $|V_{cb}|$  from  $\bar{B} \rightarrow D^* \ell \bar{\nu}$ ,” *Phys. Lett.* **B771** (2017) 359–364, [arXiv:1703.08170 \[hep-ph\]](https://arxiv.org/abs/1703.08170).
- [39] **Belle**, R. Glattauer *et al.*, “Measurement of the decay  $B \rightarrow D \ell \nu_\ell$  in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element  $|V_{cb}|$ ,” *Phys. Rev.* **D93** no. 3, (2016) 032006, [arXiv:1510.03657 \[hep-ex\]](https://arxiv.org/abs/1510.03657).
- [40] F. U. Bernlochner, Z. Ligeti, M. Papucci, and D. J. Robinson, “Tensions and correlations in  $|V_{cb}|$  determinations,” *Phys. Rev.* **D96** no. 9, (2017) 091503, [arXiv:1708.07134 \[hep-ph\]](https://arxiv.org/abs/1708.07134).
- [41] L. Hinz, “Lepton ID efficiency correction and systematic error,” *Belle Note 954 (internal)*. [http://belle.kek.jp/secured/belle\\_note/gn954/lhinz\\_bn954.ps.gz](http://belle.kek.jp/secured/belle_note/gn954/lhinz_bn954.ps.gz).
- [42] B. Bhuyan, “High  $P_t$  Tracking Efficiency Using Partially Reconstructed  $D^*$  Decays,” *Belle Note 1165 (internal)*. [http://belle.kek.jp/secured/belle\\_note/gn1165/BN1165\\_v1.pdf](http://belle.kek.jp/secured/belle_note/gn1165/BN1165_v1.pdf).
- [43] M. Misiak and M. Steinhauser, “Weak radiative decays of the B meson and bounds on  $M_{H^\pm}$  in the Two-Higgs-Doublet Model,” *Eur. Phys. J.* **C77** no. 3, (2017) 201, [arXiv:1702.04571 \[hep-ph\]](https://arxiv.org/abs/1702.04571).

- [44] M. Neubert, “Analysis of the photon spectrum in inclusive  $B \rightarrow X_s \gamma$  decays,” *Phys. Rev. D* **49** (1994) 4623–4633, [arXiv:hep-ph/9312311](#) [hep-ph].
- [45] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev, and A. I. Vainshtein, “On the motion of heavy quarks inside hadrons: Universal distributions and inclusive decays,” *Int. J. Mod. Phys. A* **9** (1994) 2467–2504, [arXiv:hep-ph/9312359](#) [hep-ph].
- [46] **SIMBA**, F. U. Bernlochner, H. Lacker, Z. Ligeti, I. W. Stewart, F. J. Tackmann, and K. Tackmann, “A model independent determination of the  $B \rightarrow X_s \gamma$  decay rate,” [arXiv:1303.0958](#) [hep-ph]. [PoSICHEP2012,370(2013)].
- [47] F. De Fazio and M. Neubert, “ $B \rightarrow X_u \ell \nu$  decay distributions to order  $\alpha_s$ ,” *JHEP* **06** (1999) 017, [arXiv:hep-ph/9905351](#) [hep-ph].
- [48] B. O. Lange, M. Neubert, and G. Paz, “Theory of charmless inclusive B decays and the extraction of  $V(\text{ub})$ ,” *Phys. Rev. D* **72** (2005) 073006, [arXiv:hep-ph/0504071](#) [hep-ph].
- [49] P. Gambino, P. Giordano, G. Ossola, and N. Uraltsev, “Inclusive semileptonic B decays and the determination of  $|V(\text{ub})|$ ,” *JHEP* **10** (2007) 058, [arXiv:0707.2493](#) [hep-ph].
- [50] A. Bharucha, D. M. Straub, and R. Zwicky, “ $B \rightarrow V \ell^+ \ell^-$  in the Standard Model from light-cone sum rules,” *JHEP* **08** (2016) 098, [arXiv:1503.05534](#) [hep-ph].
- [51] C. Bourrely, I. Caprini, and L. Lellouch, “Model-independent description of  $B \rightarrow \pi \ell \nu$  decays and a determination of  $|V_{\text{ub}}|$ ,” *Phys. Rev. D* **79** (2009) 013008, [arXiv:0807.2722](#) [hep-ph]. [Erratum: Phys. Rev. D **82**, 099902(2010)].
- [52] A. Bharucha, “Two-loop Corrections to the B to  $\pi$  Form Factor from QCD Sum Rules on the Light-Cone and  $|V_{ub}|$ ,” *JHEP* **05** (2012) 092, [arXiv:1203.1359](#) [hep-ph].
- [53] D. J. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth. A* **462** (2001) 152–155.
- [54] P. Ball and R. Zwicky, “Improved analysis of  $B \rightarrow \pi \ell \nu$  from QCD sum rules on the light cone,” *JHEP* **10** (2001) 019, [arXiv:hep-ph/0110115](#) [hep-ph].
- [55] **Belle**, A. Sibidanov *et al.*, “Study of Exclusive  $B \rightarrow X_u \ell \nu$  Decays and Extraction of  $|V_{ub}|$  using Full Reconstruction Tagging at the Belle Experiment,” *Phys. Rev. D* **88** no. 3, (2013) 032005, [arXiv:1306.2781](#) [hep-ex].
- [56] **BaBar**, P. del Amo Sanchez *et al.*, “Study of  $B \rightarrow \pi \ell \nu$  and  $B \rightarrow \rho \ell \nu$  Decays and Determination of  $|V_{ub}|$ ,” *Phys. Rev. D* **83** (2011) 032007, [arXiv:1005.3288](#) [hep-ex].
- [57] **BaBar**, J. P. Lees *et al.*, “Branching fraction measurement of  $B^+ \rightarrow \omega \ell^+ \nu$  decays,” *Phys. Rev. D* **87** no. 3, (2013) 032004, [arXiv:1205.6245](#) [hep-ex]. [Erratum: Phys. Rev. D **87**, no. 9, 099904(2013)].
- [58] C. Ramirez, J. F. Donoghue, and G. Burdman, “Semileptonic  $b \rightarrow u$  decay,” *Phys. Rev. D* **41** (Mar, 1990) 1496–1503. <https://link.aps.org/doi/10.1103/PhysRevD.41.1496>.