

SEARCH FOR THE RARE DECAY $B^+ \rightarrow l^+ \nu_l \gamma$
WITH THE FULL EVENT INTERPRETATION
AT THE BELLE EXPERIMENT

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Moritz J. Gelb
aus Aalen

Tag der mündlichen Prüfung: 20.07.2018

Referent: Prof. Dr. Michael Feindt

Korreferent: Prof. Dr. Florian Bernlochner



This document is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0):

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

Contents

Contents	i
1. Introduction	1
2. Foundations	3
2.1. Introduction	3
2.2. The Rare Decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$	5
2.3. The Decay $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$	7
2.4. Non-leptonic B Decays	8
2.5. Experimental Status	10
3. Experimental Setup	13
3.1. The KEKB Accelerator Facility	13
3.2. The Belle Detector	15
3.2.1. The Belle II Experiment	18
4. Tools	19
4.1. Belle II Analysis Software Framework	19
4.2. B2BII Conversion	20
4.3. Bremsstrahlung Correction	22
4.4. Tagging	22
4.4.1. Concept	23
4.4.2. Implementation	23
4.4.3. The Signal-specific FEI	26
5. Analysis	29
5.1. Data Samples	29

5.2.	Event Reconstruction	31
5.2.1.	Event Selection	32
5.2.2.	Signal-side Selection	33
5.2.3.	Tag-side Reconstruction	34
5.2.4.	$\Upsilon(4S)$ Reconstruction	35
5.2.5.	Continuum Suppression	38
5.2.6.	Peaking Background Suppression	39
5.2.7.	Optimization	41
5.3.	Monte Carlo Corrections	43
5.3.1.	Tag Correction	44
5.3.2.	LID Correction	47
5.3.3.	Branching Fraction Correction	48
5.3.4.	$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ EvtGen Model	49
5.4.	Remaining Background Composition	51
6.	Signal Extraction	55
6.1.	The Binned Maximum Likelihood Fit	55
6.2.	The $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ Fit Model	56
6.3.	The $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ Fit Model	59
6.4.	Significance	61
6.5.	Sample Test	61
6.6.	Toy Studies	62
6.7.	Linearity Check	63
7.	Validation Checks	67
7.1.	Off-Resonance Data	67
7.2.	M_{bc} Sideband	70
7.3.	$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ Validation	73
8.	Results	75
8.1.	Measurement on Data	75
8.2.	Systematics	79
8.3.	Extraction of λ_B and V_{ub}	86
8.4.	Comparison to Previous Analyses	88
8.5.	Outlook for Belle II	90
9.	Conclusion	93
A.	B2BII Validation Plots	97
B.	BDT Control Plots	101
B.1.	Continuum Suppression	101
B.2.	Peaking Background Suppression	107

C. Calibration	111
C.1. Selection Cuts	111
C.1.1. Event Selection	111
C.1.2. Final State Particles	111
C.1.3. $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\ell^+\nu_\ell$	112
C.1.4. $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\ell^+\nu_\ell$	112
C.1.5. $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^-\pi^+)\ell^+\nu_\ell$	113
C.2. Signal Extraction	115
D. Remaining Background Composition	117
List of Figures	121
List of Tables	127
Bibliography	131
Danksagung	137

CHAPTER 1

Introduction

As of today, mankind has a very good understanding what the visible part of the universe is made of. In the second half of the last century the Standard Model of particle physics (SM) was developed to describe nature (except for gravity). Since then, this theory has been extensively validated by many experiments. Despite its great success, there are puzzling questions that remain unanswered; e.g. the strong CP problem, the discrepancy between matter and anti-matter in the universe, and the origin of dark matter. An extension to the SM might give an explanation to those questions.

Collider experiments are looking for empirical hints for such extensions. The experiments at the Large Hadron Collider are searching for new physics at the energy frontier, where new particles might manifest themselves as resonances at very high energy scales. In contrast, the Belle experiment tests the SM at the intensity frontier looking for small deviations from theory predictions, which might reveal the indirect influence of new physics. The experiment was located at the KEKB accelerator, a so-called B-factory, which operated with a center-of-mass energy at the $\Upsilon(4S)$ resonance. The resonance decays almost always into a pair of either charged or neutral B mesons. These mesons, consisting of a heavy bottom quark and a light up or down quark, allow for a huge variety of tests.

Despite the great experimental efforts of Belle and other experiments, no observation of the rare $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay has been reported so far [1, 2]. However, the decay is of great relevance, as it offers a clean path to extract the first inverse momentum λ_B of the light-cone distribution amplitude (LCDA) of the B meson, a parameter difficult to extract from theoretical considerations [3]. Its precise knowledge would help to considerably improve the theoretical understanding of non-leptonic B meson decays. The $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay is experimentally very

challenging, since the full knowledge of the event is essential for the analysis. Whilst the signal-side B meson has a clean signature, the reconstruction of the accompanying second B meson is difficult. To reach an acceptable efficiency it has to be reconstructed by a tagging algorithm in many different known hadronic channels. This is a highly non-trivial task and requires the application of machine learning techniques.

For the upcoming Belle II experiment new analysis tools have been developed, including a significantly improved tagging algorithm. In this thesis, Belle MC and data are processed in the Belle II analysis software framework, thereby enabling the use of these tools to improve the search for the rare decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$. In addition, a new experimental method is presented to extract λ_B which reduces experimental uncertainties.

In Chapter 2, an overview of the fundamentals for the analysis is given and previous searches for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay are summarized. The experimental setup is further discussed in Chapter 3. Important analysis tools used throughout the thesis are presented in Chapter 4. In Chapter 5 the analysis procedure is described and Chapter 6 discusses the fit used for signal extraction, followed by validation checks of the analysis in Chapter 7. The unblinded results are presented in Chapter 8 and an outlook for the Belle II experiment is given.

CHAPTER 2

Foundations

This chapter gives an overview of the foundations of the analysis. A short introduction of the Cabbibo-Kobayashi-Maskawa (CKM) matrix is given in the first section. The theoretical description of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay¹ is presented in Section 2.2. The decay $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ plays an important role as background in the analysis of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decays and is described in Section 2.3. Section 2.4 introduces QCD factorization and the calculation of matrix elements of non-leptonic B decays, for which the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay can give vital input. At the end of this chapter the experimental status of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ measurements is briefly summarized.

2.1 Introduction

The flavor sector of the SM consists of six quarks, which can be grouped into three generations:

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{d} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \mathbf{s} \end{pmatrix} \begin{pmatrix} \mathbf{t} \\ \mathbf{b} \end{pmatrix}. \quad (2.1)$$

The masses of the quarks span over five orders of magnitude, from about 2.3 MeV for the up quark to 173 GeV for the top quark. In the SM, transitions between quarks are only mediated by the weak charged currents interaction, i.e. the exchange of a charged W boson which couples to the weak eigenstates of the quarks. The rotation of the mass eigenstates to the weak eigenstates can be described by a complex unitary 3×3 matrix, the CKM matrix. The matrix can be parameterized

¹The charge conjugated decay is implied throughout this work.

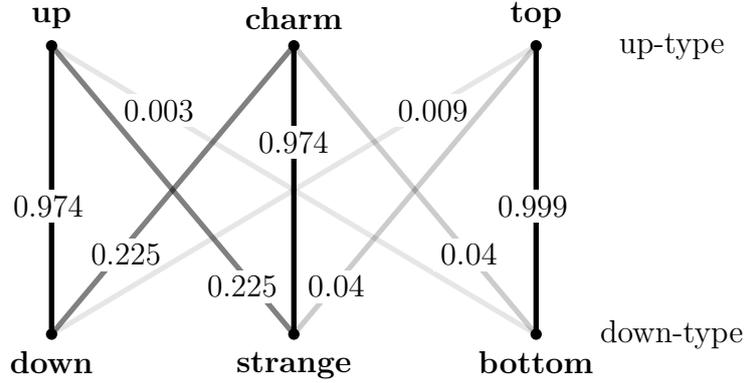


Figure 2.1.: Illustration of quark transitions, the numbers represent the corresponding CKM matrix element [4].

by three angles and a complex phase, which is the source of CP violation in the quark sector of the SM. A clear hierarchy can be seen: Inter-generation transitions are highly favored, whereas transitions between the generations are suppressed. A graphical illustration of the quark transition strength can be seen in Figure 2.1. The SM does not predict the magnitude of the matrix elements, hence they have to be measured by experiment. B decays are sensitive to the CKM matrix elements $|V_{cb}|$ and $|V_{ub}|$. As of today, a tension of about 3σ is observed for inclusive and exclusive measurements of $|V_{ub}|$; Figure 2.2 shows the world averages over time. While both the theoretical and the experimental uncertainties have decreased, the origin of the discrepancy is still not known.

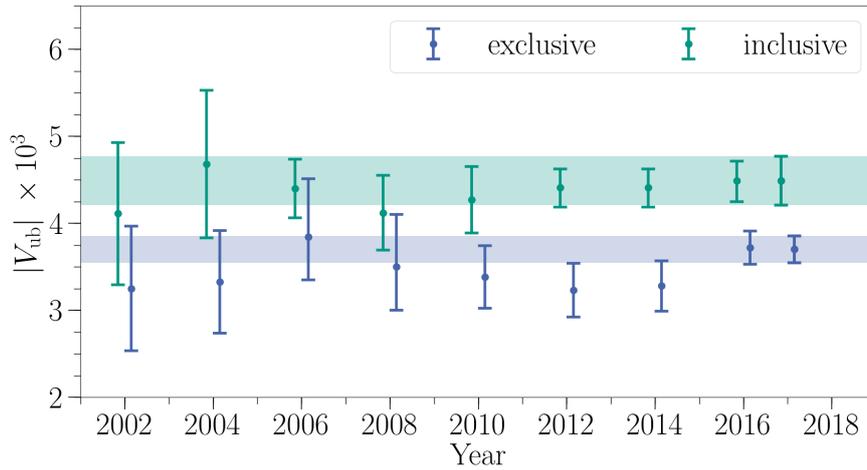


Figure 2.2.: Evolution of the averages of exclusive and inclusive $|V_{ub}|$ measurements, where the error bars represent the combined uncertainty. The *colored bands* correspond to the latest averages. All values are taken from Reference [4] and earlier publications.

2.2 The Rare Decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$

Contrary to the purely leptonic decay $B^+ \rightarrow \ell^+ \nu_\ell$, the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ is not helicity suppressed, but introduces an electromagnetic coupling of the photon. Figure 2.3 shows the dominant Feynman diagram at leading order. A good

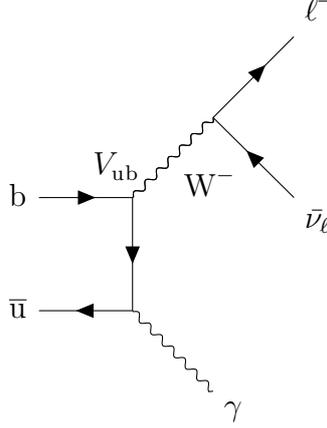


Figure 2.3.: Dominant Feynman graph at tree-level with photon emission from the up quark [5].

description of the underlying theory of the decay can be found in Reference [6] and is presented in the following. The amplitude of the decay can be written in terms of the fields of the involved particles as

$$\mathcal{A}(B^- \rightarrow \ell^- \bar{\nu}_\ell \gamma) = \frac{G_F V_{ub}}{\sqrt{2}} \langle \ell \bar{\nu}_\ell \gamma | \bar{\ell} \gamma^\mu (1 - \gamma_5) \nu \bar{u} \gamma^\mu (1 - \gamma_5) b | B^- \rangle, \quad (2.2)$$

where G_F denotes the Fermi coupling constant and γ^μ and γ^5 denote gamma matrices. The double differential decay rate of the lepton and photon energy depending on the vector F_V and the axial-vector F_A form factor can be determined as

$$\begin{aligned} \frac{d^2\Gamma}{dE_\gamma dE_\ell} &= \frac{\alpha_{em} G_F^2 m_B^3 |V_{ub}|^2}{16\pi^2} (1 - x_\gamma) \\ &\times \left[(1 - x_\nu)^2 (F_A + F_V)^2 + (1 - x_\ell)^2 (F_A - F_V)^2 \right], \end{aligned} \quad (2.3)$$

with the parameterization $x_i = 2E_i/m_B$ with ($i = \gamma, \nu, \ell$) and the mass m_B of the B meson. By integrating over the possible range of lepton energies, the single differential decay rate can be written as

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{em} G_F^2 m_B^4 |V_{ub}|^2}{48\pi^2} x_\gamma^3 (1 - x_\gamma) [F_A^2 + F_V^2]. \quad (2.4)$$

For high energy photons the form factors read as

$$\begin{aligned} F_V(E_\gamma) &= \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + \left[\xi(E_\gamma) + \frac{Q_b m_B f_B}{2E_\gamma m_b} + \frac{Q_u m_B f_B}{(2E_\gamma)^2} \right], \\ F_A(E_\gamma) &= \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + \left[\xi(E_\gamma) - \frac{Q_b m_B f_B}{2E_\gamma m_b} - \frac{Q_u m_B f_B}{(2E_\gamma)^2} + \frac{Q_\ell f_B}{E_\gamma} \right], \end{aligned} \quad (2.5)$$

where Q_i is the charge of the given particle ($i = u, b, \ell$), f_B is the decay constant of the B meson, m_b is the mass of the bottom quark and λ_B is the first inverse momentum of the light-cone distribution amplitude (LCDA) of the B meson. The parameter λ_B also appears in the calculation of non-leptonic B decays what is further discussed in Section 2.4. The functional dependence of the form factors on the photon energy is shown in Figure 2.4.

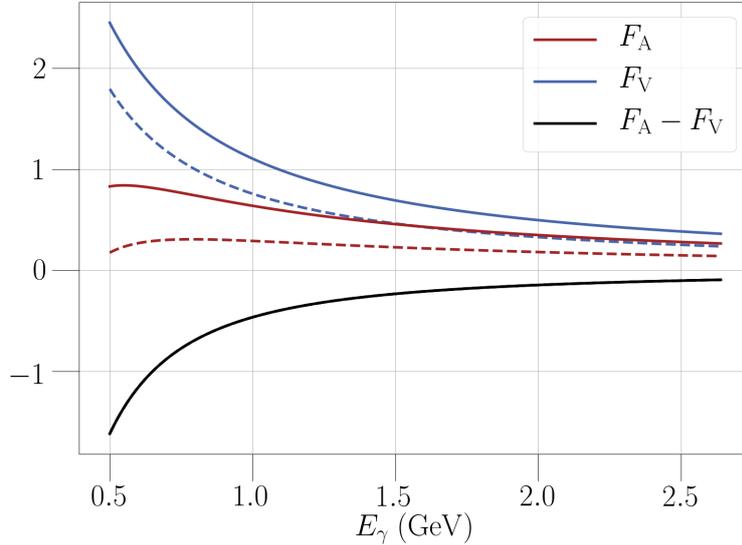


Figure 2.4.: Dependence of the vector and axial-vector form factor on the photon energy for $\lambda_B = 300$ MeV (*continuous lines*) and $\lambda_B = 500$ MeV (*dashed lines*). The black line shows their difference, which is independent of λ_B .

The factor $R(E_\gamma, \mu)$ in Equation (2.5) corresponds to a radiative correction which depends on the photon energy and the energy scale μ . It can be calculated in the heavy-quark expansion (HQE) to first order. The expression in brackets gives the $1/m_b$ power corrections to the leading order term. One can distinguish between *symmetry-preserving* and *symmetry-breaking* contributions, since they are equal and different for both form factors, respectively. The symmetry-preserving term $\xi(E_\gamma)$ is so far unknown, except for the fact that it has to be of $\mathcal{O}(1/m_b)$. The authors of Reference [6] chose an ansatz of $\xi(E_\gamma) = c f_B / 2E_\gamma$, where the factor c was varied between -1 and $+1$. An improvement was presented in

Reference [7]: by applying the framework of HQE the authors could evaluate $\xi(E_\gamma)$ to a potential accuracy of 20%. They applied a method which was originally used for the calculation of form factors for the $\gamma^* \gamma \rightarrow \pi$ transition. The second term in Equation (2.5) corresponds to the power-suppressed photon emission of the bottom quark. The third term represents power corrections of the anti-up quark, and the last term in F_A is the contribution from the lepton emitting the photon. By integrating the differential decay rate over the photon energies, the partial branching fraction can be calculated as

$$\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)_{E_\gamma > E_\gamma^{\min}} = \frac{\tau_{B_d}}{\hbar} \int_{E_\gamma^{\min}}^{\frac{m_B c^2}{2}} \frac{d\Gamma}{dE_\gamma} dE_\gamma. \quad (2.6)$$

The lower integration boundary E_γ^{\min} is defined by the lower cut-off on the photon energy.

The branching fraction was estimated to be in the order of $\mathcal{O}(10^{-6})$ [8].

2.3 The Decay $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$

The semileptonic decay $B \rightarrow \pi \ell \nu_\ell$ is theoretically well understood. Like the previously discussed decay the amplitude is proportional to the CKM matrix element V_{ub} . The differential decay rate depending on the four-momentum transfer $q = p_B - p_\pi$ to the lepton pair can be written in terms of a scalar $f_0(q^2)$ and a vector $f_+(q^2)$ form factor

$$\begin{aligned} \frac{d\Gamma(B \rightarrow \pi \ell \nu_\ell)}{dq^2} &= \frac{G_F^2}{24\pi^3 m_B^2 q^4} |V_{ub}|^2 (q^2 - m_\ell^2)^2 |\vec{p}_\pi| \\ &\times \left[\left(1 + \frac{m_\ell^2}{2q^2} \right) m_B^2 |\vec{p}_\pi|^2 |f_+(q^2)|^2 + \frac{3m_\ell^2}{8q^2} (m_B - m_\pi)^2 |f_0(q^2)|^2 \right], \end{aligned} \quad (2.7)$$

where m_π denotes the mass of the pion. The momentum of the pion in the rest frame of the B meson can be expressed as

$$|\vec{p}_\pi| = \frac{1}{2m_B} \sqrt{(m_B^2 + m_\pi^2 - q^2)^2 - (2m_B m_\pi)^2}. \quad (2.8)$$

For light leptons ($\ell = e, \mu$) one can neglect terms proportional to m_ℓ^2 and write the differential decay rate solely in terms of the vector form factor

$$\frac{d\Gamma(B \rightarrow \pi \ell \nu_\ell)}{dq^2} = \frac{G_F^2}{24\pi^3} |V_{ub}|^2 |f_+(q^2)|^2 |\vec{p}_\pi|^3 \quad (2.9)$$

The vector form factor can be parameterized with the Bourrely-Caprini-Lellouch (BCL) expansion [9, 10]. The parameter q^2 is mapped to the function

$$z(q^2, t_0) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}}, \quad (2.10)$$

using $t_{\pm} = (m_B \pm m_{\pi})^2$ and $t_0 = (m_B + m_{\pi})(\sqrt{m_B} - \sqrt{m_{\pi}})^2$. The form factor can be written in terms of the BCL expansion parameters b_k

$$f_+(q^2) = \frac{1}{1 - q^2/m_{B^*}^2} \sum_{k=0}^{K-1} b_k \left[z^k - (-1)^{k-K} \frac{k}{K} z^K \right], \quad (2.11)$$

with a B^* meson mass of $m_{B^*} = 5.325$ GeV, the number of expansion parameters K and with $z = z(q^2, t_0)$. The expansion parameters were evaluated by fitting the q^2 spectra in data taken by the Belle and BaBar experiments and using the theory calculations from lattice (for high q^2) and light-cone sum rule (for low q^2) calculations. The latest fit of the expansion parameters from HFLAV [11] can be found in Table 2.1. Under the assumption of isospin symmetry a factor of two has to be considered for the charged B meson decay. Using those results, the decay rate is found to be

$$\Gamma(B^+ \rightarrow \pi^0 \ell^+ \nu_{\ell}) = |V_{ub}|^2 \cdot (2.43 \pm 0.17) \times 10^{-12} \text{ GeV}. \quad (2.12)$$

Since the $B^+ \rightarrow \pi^0 \ell^+ \nu_{\ell}$ decay is experimentally and theoretically well known it is ideally suited for the extraction of $|V_{ub}|$.

Table 2.1.: Latest BCL expansion parameters from a fit to experimental data of $B^+ \rightarrow \pi^0 \ell^+ \nu_{\ell}$ and $B^0 \rightarrow \pi^- \ell^+ \nu_{\ell}$ decays and theory calculations [11].

b_0	b_1	b_2
0.421 ± 0.012	-0.390 ± 0.033	-0.650 ± 0.126

2.4 Non-leptonic B Decays

The breakdown of perturbation theory at low energies makes the calculation of decay processes of non-leptonic B decays to two mesons like the $B \rightarrow \pi\pi$ decay very challenging. The framework of QCD factorization (QCDF) was developed to overcome these problems. In this section a short overview of QCDF is given, following Reference [3].

The amplitude of a non-leptonic B meson decay into two mesons M_1 and M_2 can

be written in terms of local operators \mathcal{O}_i as

$$\mathcal{A}(\text{B} \rightarrow M_1 M_2) = \frac{G_F}{\sqrt{2}} \sum_i \lambda_i C_i(\mu) \langle M_1 M_2 | \mathcal{O}_i | \text{B} \rangle(\mu), \quad (2.13)$$

where λ_i is the corresponding CKM factor and $C_i(\mu)$ is a coefficient function representing strong interaction effects above the scale m_b . The calculation of these matrix elements is difficult, due to the corrections from the strong interaction. With the method of QCDF the calculation of the matrix element can be divided into a long-distance (soft) contribution and a short-distance (hard) contribution. In the case of a heavy meson M_1 and a light meson M_2 with masses m_1 and m_2 the matrix element becomes

$$\langle M_1 M_2 | \mathcal{O}_i | \bar{\text{B}} \rangle = \sum_j F_j^{\text{B} \rightarrow M_1}(m_2^2) \int_0^1 du T_{ij}^I(u) \Phi_{M_2}(u), \quad (2.14)$$

if the spectator quark of the B meson goes to the heavy meson. For two light mesons M_1 and M_2 the transition matrix element can be written as

$$\begin{aligned} \langle M_1 M_2 | \mathcal{O}_i | \text{B} \rangle &= \sum_j F_j^{\text{B} \rightarrow M_1}(m_2^2) \int_0^1 du T_{ij}^I(u) \Phi_{M_2}(u) \\ &+ \sum_j F_j^{\text{B} \rightarrow M_2}(m_1^2) \int_0^1 dv T_{ij}^I(v) \Phi_{M_1}(v) \\ &+ \int_0^1 d\xi du dv T_i^{II}(\xi, u, v) \Phi_{\text{B}}(\xi) \Phi_{M_1}(v) \Phi_{M_2}(u), \end{aligned} \quad (2.15)$$

with form factors $F_j^{\text{B} \rightarrow M_{1/2}}(m_{2/1}^2)$ describing the transition between the B and the light mesons, the hard-scattering kernels $T_{ij}^I(u)$ and $T_i^{II}(u)$, and the LCDA $\Phi_X(u)$ of a meson $X = M_1, M_2, \text{B}$. The hard-scattering kernel can be calculated perturbatively in an expansion of the strong coupling constant $\alpha_s(m_b)$. The form factors and the LCDAs have to be calculated by non-perturbative tools or have to be extracted from experiment.

As of today, there is no precise knowledge of the LCDA of the B meson and its first inverse moment

$$\frac{1}{\lambda_{\text{B}}} = \int_0^1 \frac{d\xi}{\xi} \Phi_{\text{B}}(\xi). \quad (2.16)$$

Theoretical calculations and measurements have found different values for λ_{B} ; a summary can be found in Table 2.2. Calculations of branching fractions of the $\text{B} \rightarrow \pi\pi, \pi\rho$ and $\rho_L\rho_L$ decays in the framework of QCDF seem to prefer lower values of λ_{B} . This conclusion is drawn by comparing experimental results with the calculations for different input values of λ_{B} . However, the method is accompanied by large theoretical uncertainties and the results have to be considered with

caution. On the other hand, QCD sum rules favor a significantly higher value. Besides the theoretical tools, the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay allows for the clean extraction of λ_B . As described in the next section, several attempts were made to measure the branching ratio of the decay.

Table 2.2.: Obtained values for first inverse moment of the LCDA of the B meson using theoretical and experimental methods.

	λ_B (MeV)
QCDF [12, 13]	≈ 200
QCD sum rules [14]	460 ± 110
BaBar [6]	> 115
Belle [1]	> 238

2.5 Experimental Status

Searches for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay have been conducted at several e^+e^- collider experiments over the last two decades. Although the experimental methods improved and larger data sets became available over time, no experiment has been able to report evidence for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay. To perform the analysis with a reasonable amount of background events, it is necessary to exploit the full information of the $\Upsilon(4S) \rightarrow B\bar{B}$ event. This requires the application of a dedicated reconstruction algorithm for the second B meson. With increasing data sets it has been possible to apply more advanced reconstruction strategies. An overview of existing measurements can be found in Table 2.3.

The first search was performed by the CLEO collaboration on a data sample of 2.7×10^6 $B\bar{B}$ pairs [15]. A signal candidate was identified by assigning the most energetic photon and lepton to a signal candidate. The second B meson was reconstructed by inclusive tagging (see Section 4.4), i.e. all final state particles not used for the signal-side B meson were used for the reconstruction. The analysis was able to set an upper limit at 90% confidence level (C.L.) of $\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) < 5.2 \times 10^{-5}$ and $\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) < 2.0 \times 10^{-4}$.

In 2009 the BaBar collaboration published an analysis on the full recorded data set of 465×10^6 $B\bar{B}$ pairs [2]. The authors used a hadronic tagging approach (see Section 4.4), where the second B meson is reconstructed exclusively in a hadronic decay mode, which allows for a high purity in the event selection. The authors followed a model-independent approach, using two form factor models for the signal MC generation, since at that time very little was known about the form factor structure. To conduct the analysis model-independent, no cuts on the photon energy and the angles on the signal-side were applied in the reconstruction.

However, as pointed out in Reference [16, p.114f] the approach is not fully model-independent, since the reconstruction efficiencies and other selection variables depend indirectly on the signal model. The authors found upper limits at 90% C.L. of $\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) < 17 \times 10^{-6}$, $\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) < 24 \times 10^{-6}$ and combined $\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 15.6 \times 10^{-6}$. Supplementary, the authors give an upper limit on the partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 14 \times 10^{-6}$ for a cut on the photon energy of $E_\gamma > 1 \text{ GeV}$. The authors of Reference [6] revisited the analysis for the partial branching fraction measurement and found a limit on the first inverse momentum of the LCDA of the B meson of $\lambda_B > 115 \text{ MeV}$.

The Belle collaboration published a search in 2015 which was performed on the full Belle data set of 772×10^6 $B\bar{B}$ pairs [1]. A partial branching fraction was measured for two different cuts on the signal-side photon of $E_\gamma > 0.4 \text{ GeV}$ and $E_\gamma > 1 \text{ GeV}$. The second B meson was hadronically tagged. The analysis was able to set the most stringent upper limits on the partial branching fraction. For the $E_\gamma > 1 \text{ GeV}$ case upper limits at 90% C.L. of $\Delta\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) < 6.1 \times 10^{-6}$, $\Delta\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) < 3.4 \times 10^{-6}$ and $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 3.5 \times 10^{-6}$ were found. The lower limit at 90% C.L. of the first inverse moment of the LCDA of the B meson was improved to $\lambda_B > 238 \text{ MeV}$. The aim of this thesis is to improve on this Belle result.

Table 2.3.: Overview of previous searches for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay.

Experiment	Data set (fb^{-1})	Limit 90% C.L. (10^{-6})	Comment
CLEO (1997) [15]	2.5	$\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) < 52$	–
BaBar (2009) [2]	423	$\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) < 200$	–
		$\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) < 17$	} model-independent
		$\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) < 24$	
Belle (2015) [1]	711	$\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 15.6$	} with $E_\gamma > 1 \text{ GeV}$
		$\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 14$	
		$\Delta\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) < 6.1$	} with $E_\gamma > 1 \text{ GeV}$
		$\Delta\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) < 3.4$	
		$\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 3.5$	

CHAPTER 3

Experimental Setup

The Belle experiment was located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. It ran over ten years from 1999 to 2010. Embedded into the KEKB accelerator facility is was specialized on the measurement of B meson pairs, produced mainly at the $\Upsilon(4S)$ resonance. The Belle experiment was in direct competition with the BaBar experiment located at the Stanford Linear Accelerator Center (SLAC), which was built and operated in the same decade. The physics program of both experiments was very successful and led, besides many other great discoveries, to the Noble Prize in Physics for Kobayashi and Maskawa for the prediction of a third quark generation [17]. In this chapter the experimental setup of the KEKB accelerator facility and the Belle detector is described. In addition, a short outlook on the successor experiment Belle II is given, which started to record the first collisions at the time of writing.

3.1 The KEKB Accelerator Facility

The KEKB accelerator facility was designed as an asymmetric energy e^+e^- collider and was built in the former tunnel of the TRISTAN accelerator [18]. An illustration of the setup can be seen in Figure 3.1. The accelerator consisted of two storage rings with a circumference of about 3 km, the low energy positron ring (LER) and the high energy electron ring (HER). The two beams crossed at the interaction point (IP), where the Belle detector was located. In total 1600 magnets of different types with additional steering magnets were installed along the main ring to deflect the beams [20]. An electron gun was used as source for the

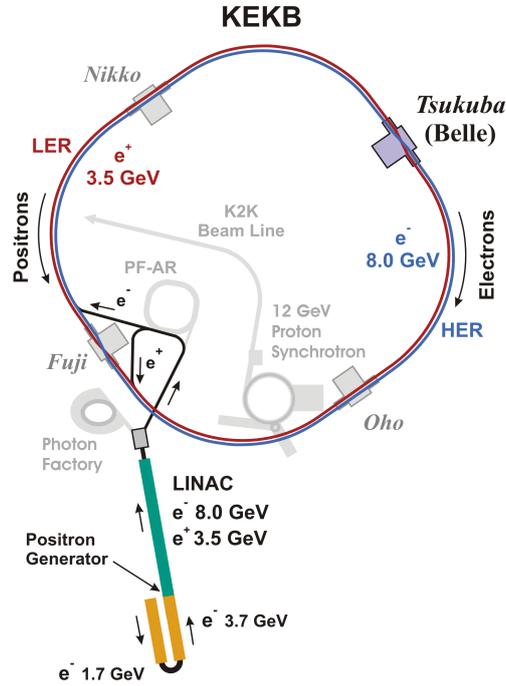


Figure 3.1.: The KEKB accelerator complex. The Belle detector was located in the Tsukuba experimental hall. The beams were injected from the LINAC (*green*). Adapted from [19].

electron beam. Positrons were produced by irradiating a tungsten plate, which was inserted into the electron beam line. Positrons and electrons were separated with a chicane made of four magnets and a beam trap. The beams were accelerated with a linear accelerator (LINAC) and then injected into the storage ring [21]. The LINAC was also part of a previous experiment and was upgraded to meet the requirements of the Belle experiment. An outstanding achievement of the accelerator was the more than doubled design luminosity. More information can be found in Reference [22].

After the shutdown of the Belle experiment, the KEKB accelerator was upgraded to SuperKEKB for the Belle II experiment [24, Chapter 2]. The modifications were necessary to reach the envisaged 40 times higher design luminosity of $8 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$. This included a switch to the nano-beam scheme to reduce the beam size at the IP. The crossing angle was enlarged to reduce the β -functions at the IP. The beam energies were adjusted to 7.0 GeV for the HER and 4.0 GeV for the LER to reduce emittance and shorter beam life times caused by the Touschek effect. In addition, the beam currents were more than doubled. A comparison between the machine parameters at the end of operation time of KEKB and the design parameters of SuperKEKB is shown in Table 3.1.

Table 3.1.: Machine parameters of KEKB at the end of the operation time [23, Chapter 1.3] and the design parameters for SuperKEKB [24].

Parameter	KEKB		SuperKEKB	
	HER	LER	HER	LER
Beam energy (GeV)	8.0	3.5	7.0	4.0
Beam current (A)	1.2	1.6	2.62	3.60
Beam size at IP				
x (μm)	80		7.75	10.2
y (μm)	1.0		0.059	
z (mm)	5.0		5.0	6.0
Luminosity ($\text{cm}^{-2}\text{s}^{-1}$)	2.1×10^{34}		8×10^{35}	
Number of bunches	1584		2503	
Bunch spacing (m)	1.84 (9.4 ns)		1.20 (4 ns)	
Beam crossing angle (mrad)	± 11		± 41.5	

3.2 The Belle Detector

The Belle detector consisted of several sub-detectors which were arranged around the beam axis. Schematic sketches of the detector can be seen in Figures 3.2 and 3.3. The most important parts of the Belle detector are described in the following. For more information the reader is referred to References [23] and [25].

Interaction region At the interaction region (IR) the beam pipe comprised two beryllium cylinders, outside this region the pipe was made of aluminum. The interstice between the cylinders was flood with helium gas for cooling to compensate for beam-induced heating. To allow for a good vertex resolution, the design was optimized to reduce multiple scattering and to place the silicon vertex detector (SVD) as close as possible to the IR. The beams were focused with super-conducting final-focus quadrupole magnets (QCS) in the IR.

Extreme forward calorimeter The extreme forward calorimeter (EFC) was used to monitor the luminosity and the machine background. The EFC consisted of radiation hard bismuth germanate crystals and was placed around the beam pipe in forward and backward direction and covered an angle of $6.4^\circ < \theta < 11.5^\circ$ in forward direction and $163.3^\circ < \theta < 171.2^\circ$ in backward direction.

Silicon vertex detector The objective of the SVD was the measurement of primary and secondary vertices, it was made of three layers of double-sided silicon-strips. With $23^\circ < \theta < 140^\circ$ the first version (SVD1) covered not the full angular acceptance. After three years of operation the SVD1 was

replaced due to radiation damage of the read-out electronics. Its replacement, the SVD2, offered a radiation robust readout chip and was designed with one additional layer. The full acceptance angle of the Belle detector of $17^\circ < \theta < 150^\circ$ was covered.

Central drift chamber The main purpose of the central drift chamber (CDC) was the reconstruction of charged tracks and their momenta. Together with the time-of-flight system (TOF) and the aerogel Cherenkov counter (ACC) the CDC was used for particle identification (PID) and delivered trigger signals for charged tracks. The CDC consisted of 50 cylindrical super-layers, where each super-layer contained three to six axial or stereo layers. A gas mixture of 50% helium and 50% ethane was chosen. This allowed a good momentum resolution of low-momentum tracks (< 1 GeV), which dominate for B decays. The CDC had a very thin inner wall to reduce multiple scattering and contained in total 8400 drift cells, made of gold-plated tungsten sense wires and aluminum field wires.

Time-of-flight system The TOF was built out of 64 modules arranged as barrel. A module comprised two plastic scintillator counters, a trigger scintillator counter and photo multipliers at each side. With the reconstruction information of the CDC and the event time T_0 it measured the time of flight for charged particles. The TOF was optimized for tracks with a momentum of less than 1.2 GeV.

Aerogel Cherenkov counter The ACC complements the PID for high momentum tracks and consisted of 960 single modules in the barrel and 228 modules in the forward end cap. Each module included five aerogel tiles. The tiles were placed in a small aluminum box with photo multipliers attached for the detection of Cherenkov light. The aerogels had different refractive indices to allow for a good separation of kaons and pions.

Electromagnetic calorimeter Photons were measured by the electromagnetic calorimeter (ECL), which was segmented into a barrel, forward and backward region. The ECL was built out of 8736 thallium-doped cesium iodide crystals pointing almost directly towards the IP. At the end of the crystals two silicon photodiodes were attached to measure the scintillation light. Besides the photon detection the ECL allowed to distinguish electrons from charged hadrons by measuring the ratio of deposited energy over the momentum of the charged track. The forward and backward region were also used to measure the luminosity using well known signature of Bhabha events. It had an angular coverage of $17^\circ < \theta < 150^\circ$.

K_L^0 and muon detection system The K_L^0 and muon detection system (KLM) was built of 15 layers of resistive plate counters and 14 layers of iron plates. The

iron layers had 3.9 interaction lengths to form a hadron shower from a K_L^0 . However, due to fluctuations of the shower size, no reliable measurement of the K_L^0 energy was possible. The smaller deflections of muons allowed to separate them from K_L^0 . With the end caps the angular coverage was $20^\circ < \theta < 155^\circ$

Solenoid A 1.5 T magnetic field was created parallel to the beam axis by a superconducting solenoid, located between the ECL and the KLM. The iron plates of the KLM were used as return yoke for the magnetic flux.

Trigger The trigger system was divided into the Level-1 hardware trigger and the Level-3 software trigger. The different sub-detectors gave information on measured tracks and deposited energy to a global decision logic (GDL). Based on this information the GDL decided on different events types, e.g. hadronic or two photon events.

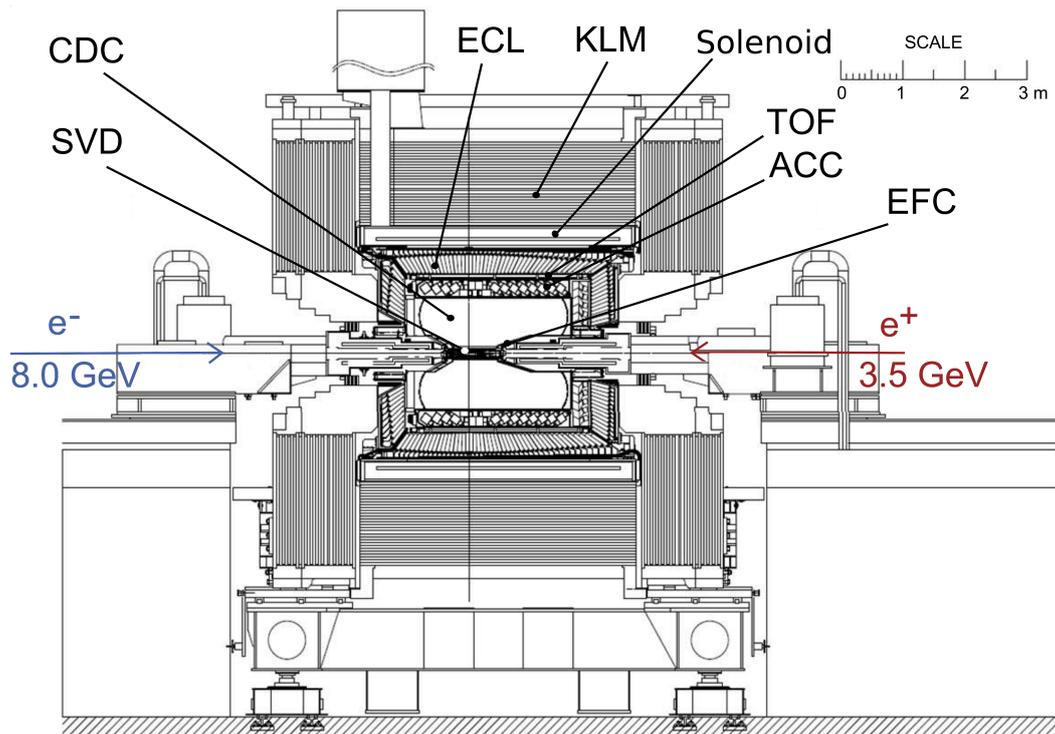


Figure 3.2.: Schematic side view of the Belle detector. The z -axis points along the opposite direction of the e^+ beam. Adapted from [23].

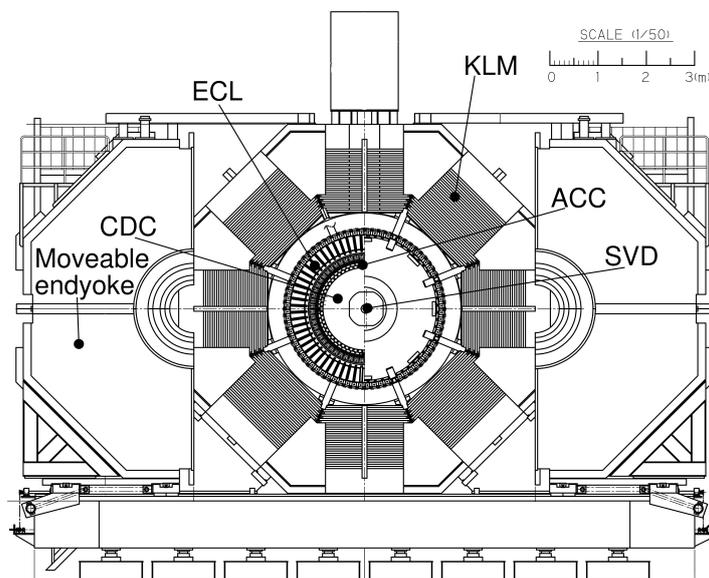


Figure 3.3.: Front view of the Belle detector. Adapted from [23].

3.2.1 The Belle II Experiment

The Belle II detector has to handle increased beam background rates caused by the higher luminosity. A completely new vertex detector was designed, consisting of a pixel detector (PXD) and a silicon strip detector. Due to a reduced beam pipe radius, it can be placed closer to the IP. The PXD comprises two layers of depleted field effect transistors, which can cope with an estimated beam background rate 30 times higher than what was observed with Belle [24, Chapter 5]. The SVD consists of four layers of double-sided silicon strip detectors. The performance of the vertexing is expected to be comparable or better than at Belle. The CDC follows mostly the design of its predecessor. Besides faster read-out electronics the size of the CDC was increased, whereas the drift cell size was decreased. Although the CDC suffers under 20 times higher expected beam background, an improved efficiency can be expected due to better tracking software. The PID system is represented by a Time-Of-Propagation counter (TOP) in the barrel region and an Aerogel Ring-Imaging Cherenkov detector (ARICH) in the forward region. The TOP consists of silica radiator bars with photomultipliers to measure Cherenkov light. The ARICH consists of two aerogel layers with different refractive indices. For the ECL the barrel part is reused — pure cesium iodide crystals are used in the end caps. The readout electronics are replaced. The KLM is further upgraded with scintillator strips in the end cap region to cope with the higher backgrounds. More information about the Belle II design can be found in Reference [24].

CHAPTER 4

Tools

This chapter introduces the most important software tools used in the thesis. Although data of the Belle experiment is analyzed, the analysis is conducted with the software framework of the Belle II experiment. In Section 4.1 an overview of the framework is given. Section 4.2 describes the conversion of recorded data and simulated Monte Carlo (MC) events from the Belle to the Belle II data format. The conversion is necessary, since the data format between the experiments is not compatible. The correction of Bremsstrahlung is discussed in Section 4.3. A new tagging algorithm, allowing to reconstruct an entire $\Upsilon(4S)$ event, is presented in Section 4.4. The algorithm has an improved performance compared to its predecessor and is now, due to the conversion, applicable to Belle data. In the context of the thesis important contributions were made to the aforementioned software tools.

4.1 Belle II Analysis Software Framework

To match the objectives of the Belle II experiment the collider and detector were upgraded. A new software framework, the **Belle II Analysis Software Framework** (BASF2), was developed to cope with the increased demands on the software side. The framework is written in C++ and can be controlled with Python3 steering files, it builds up on ROOT and includes many other third-party libraries, e.g. TensorFlow for deep-learning applications [26]. BASF2 can be used for *online* tasks during data taking (e.g. data acquisition or high-level trigger) and *offline* tasks (e.g. detector calibration or event reconstruction).

The full functionality for physics analyses is provided by BASF2, from event

generation [27] and detector simulation [28] up to advanced analysis techniques such as *flavor tagging* [29] or *B-tagging* (see Section 4.4). The event processing is done within modules, where each module has a particular task. The modules are added to a path and are executed one after another. The data is stored as `ROOT` objects, which can be accessed by the modules.

For instance, to reconstruct a B meson from the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ on analysis level, the user adds modules to the path which first collect the reconstructed final state particles γ , e^+ and μ^+ in the event. These final state particles are reconstructed beforehand from the raw data; i.e. photons are reconstructed from ECL clusters with no charged track assigned and leptons are reconstructed from tracks, which are the result of helix fits to the tracking detector signals using the respective mass hypothesis. The reconstructed particles are stored in so-called `ParticleLists`. The next module combines the collected final state particles to charged B mesons. Combinatorial background can be further reduced by a module which applies selection cuts on physical quantities, e.g. the invariant mass calculated from the candidate. For applications like event skimming it is possible to connect multiple paths with conditions. Different objects can be related, e.g. between a reconstructed particle and a Monte Carlo (MC) particle after applying the MC matching. More implementation details can be found in References [30, 31, 32].

4.2 B2BII Conversion

The Belle experiment stopped data taking in 2010. However, the recorded data set has still the potential for interesting physics analyses. The experiment used the `Belle Analysis Framework (BASF)`, which was developed in the mid-90s [33]. The data was stored in `mDST` files as `PANTHER` tables, a custom solution of the Belle collaboration based on `C/C++` and `Fortran` [34]. The data preservation is compromised on the long run, since this format is incompatible with the Belle II software and the knowledge of maintaining and operating the `BASF` software is lost over time.

As part of `BASF2`, the `b2bii` package was developed to convert data and MC of the Belle experiment and make it available within `BASF2`. The conversion itself is non-trivial, due to the disparate representations of data and partly different sub-detectors. In principle, a conversion of the raw detector data would be possible, but the reconstruction algorithms are optimized for Belle II and cannot be effectively applied to Belle data. This approach would also require the repetition of systematic studies, e.g. for tracking efficiencies. Hence, reconstructed objects are mapped from the `PANTHER` tables to their corresponding representation in `BASF2`. In the following, the basic aspects of the conversion are described; more details can be found in Reference [35, 36].

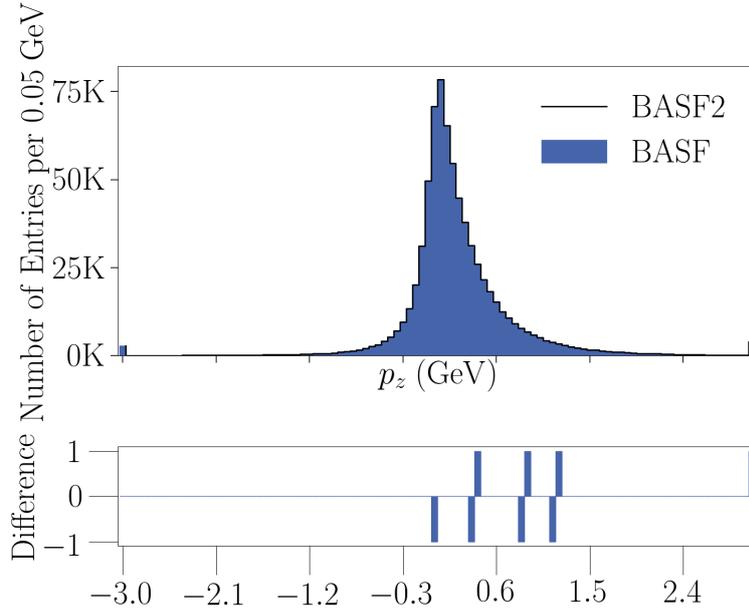


Figure 4.1.: Validation plot of the z momentum component of charged tracks, based on a sample of simulated $b \rightarrow c$ decays. The migration of single tracks between two neighboring bins and the different treatment of special values can be observed at the bin around zero and the overflow bin.

The conversion is divided into three modules in the `b2bii` package. In the first step, a module reads the `PANTHER` tables from the `mdst` files into the memory, it consists predominantly of re-used `BASF` code. The second module applies different corrections to the events, e.g. experiment and run dependent calibration factors to the beam energy, and applies some low-level cuts on the events to remove background events (see also Section 5.2.1); this is also done with `BASF` code. The actual conversion is done with the last module, mapping the reconstructed objects from `PANTHER` tables to their corresponding representation in `BASF2`.

The mapping can be illustrated on the conversion of KLM information: The `Mdst_klm_cluster` and `Mdst_klong` `PANTHER` tables contain reconstructed KLM clusters and K_L^0 candidates, respectively. These `BASF` objects are mapped to the corresponding `BASF2` objects, the `KLMcluster` and the `K_L0:mdst ParticleList`, respectively. No dedicated MC matching algorithm for K_L^0 was available in `BASF` and it was therefore handled by the analyst. Hence, a vague MC matching is performed, where the K_L^0 MC particles are matched to the closest reconstructed K_L^0 candidates within a range of 15 degrees in the θ and ϕ planes.

During the development of `b2bii` a thorough validation of the conversion was necessary. This was done by comparing histograms in more than three hundred different physical quantities on simulated and recorded events, processed with `BASF` and `BASF2`. By comparing these quantities, possible deviations can be investigated.

An example plot for charged tracks can be seen in Figure 4.1; more comparison plots can be found in Appendix A. Numerical imprecision can lead to migration between neighboring bins. Because of a different treatment of NaN (Not a Number) and infinity values, entries in the under- or overflow bin can be found in the bin containing zero. Both effects can be seen in the example plot. In summary, small deviations were observed, but can be explained and were found to be harmless.

4.3 Bremsstrahlung Correction

Electrons can lose a large amount of energy due to Bremsstrahlung along their trajectory through the detector layers. This leads to a tail in the momentum resolution distribution. On analysis level, the momentum resolution can be improved by taking suitable photons into account, which might originate from Bremsstrahlung of electron tracks. Many Belle analyses corrected this in a simple manner by adding the momentum four-vector of a possible Bremsstrahlung photon to the momentum four-vector of the electron.

The approach was re-implemented and improved by introducing the `FSRCorrection` module into `BASF2`. An electron candidate is corrected, if a photon within a given cone (default is 5 degrees) in the θ and ϕ plane is found. The cone is calculated around the momentum of the electron, which is measured at the point of closest approach (POCA) to the IP. Hence, only Bremsstrahlung photons radiated in the region of the IP are taken into account for the correction, unless the track has a high momentum and thus a low curvature. Additionally, an upper photon energy threshold (default is 1 GeV) is applied. If more than one photon is found, the closest photon is taken. A new `ParticleList` is generated, containing the (un)corrected electrons. The original electron and the Bremsstrahlung photon (if found) are added as daughters. This ensures that the photon candidate is not used twice during the event reconstruction. In addition, the error matrix of the corrected electron is updated. The effect of the correction is small but makes the distribution more symmetric. For the Belle II experiment, the resolution can be further improved at the level of track reconstruction, see Reference [37] for a detailed discussion.

4.4 Tagging

At B-Factories the initial state of the $B\bar{B}$ pair is well known. This allows for the application of a very powerful analysis method, which is referred to as *B-tagging*. In this section the general idea and variants of B-tagging are explained, followed by an overview of the implementation used throughout the thesis.

4.4.1 Concept

The $\Upsilon(4S)$ resonance decays with a branching fraction of over 96% into a charged or neutral B meson pair. The mesons are produced back-to-back in the center-of-mass frame, whereof one B meson, the so-called B_{sig} , is reconstructed in the desired signal decay. The accompanying B meson, the so-called B_{tag} , is reconstructed by the tagging algorithm. One can define the *signal-* and *tag-side* as the part of the event, which contains the final state particles corresponding to the B_{sig} and B_{tag} , respectively. In addition, the *rest-of-event* (ROE) with respect to a reconstructed particle can be defined as all final state particles, which were not used for the reconstruction of the specified particle. After the correct reconstruction of the B_{sig} candidate, its ROE should contain all final state particles belonging to the B_{tag} . The situation is depicted in Figure 4.2. By combining both B mesons, the entire visible $\Upsilon(4S)$ event can be reconstructed. A correctly reconstructed event should have no additional final state particles (neglecting small contributions from beam background processes). This constraint can be used to efficiently reject combinatorial background.

Different variants of B-tagging are possible to reconstruct the tag-side, whereby the choice depends heavily on the signal-side of the analysis. In *exclusive B-tagging* a B candidate is reconstructed in specific decay channels. To reach a reasonable efficiency, decay channels with a large branching fraction are chosen. One can distinguish between *hadronic* and *semileptonic* exclusive B-tagging, depending on the reconstructed decay channels. The later allows for more candidates due to the higher semileptonic branching fractions, but has a lower purity compared to the hadronic case. For the semileptonic decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ hadronic tagging is very useful, since it allows to deduce information about the undetectable neutrino. With a suitable variable for the signal extraction, such as M_{miss}^2 (defined in Chapter 5) calculated on the basis of the $\Upsilon(4S)$ candidate, the decay can be measured. Additional neutrinos introduced by a semileptonic tag channel would compromise this approach. *Inclusive B-tagging* reconstructs a B_{tag} candidate by combining all four-momenta of the ROE with respect to the B_{sig} . This approach does not consider any explicit decay channel and allows in almost all cases for a candidate on the tag-side, leading to a high tagging efficiency but also a low purity of the candidates. This approach is chosen for decays like $B^+ \rightarrow \mu^+ \nu_\mu$ which is strongly helicity suppressed ($\mathcal{O}(\mathcal{B}) \sim 10^{-7}$) and a low number of these events is expected in the Belle data set. On the other hand, the decay allows for a very pure signal selection due to the high energetic lepton and the two body decay kinematic.

4.4.2 Implementation

In this thesis an exclusive tagging algorithm is used, which is shipped with BASF2, referred to as Full Event Interpretation (FEI) [38, 36]. The algorithm follows a

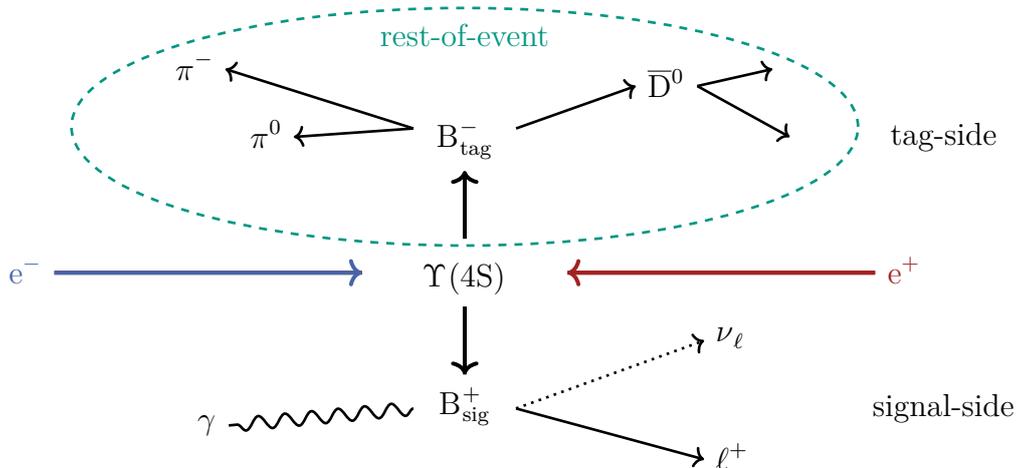


Figure 4.2.: Illustration of signal- and tag-side of a $\Upsilon(4S)$ event. The *green ellipse* indicates the ROE with respect to the B_{sig} candidate. It contains all final state particles which are not used for the B_{sig} reconstruction.

bottom-up approach, depicted in Figure 4.3. It can be used for the reconstruction of charged and neutral B mesons in semileptonic and hadronic decay channels. Currently only decay modes of the $\Upsilon(4S)$ resonance are supported, but in principle the FEI can be easily extended to higher or lower resonances. In the following, a short description of the algorithm is given. For more details the user is referred to the aforementioned references.

The FEI algorithm is structured into stages, following the decay chain of the B meson, starting at the final state particles. At each stage several modules take care of the reconstruction, e.g. by combining particles from previous stages, applying vertex fits and selection cuts. Gradient-boosted decision trees (BDTs) are utilized as multivariate classifiers to separate signal from background candidates and reduce combinatorics [40]. The classifiers assign a signal probability to each candidate, based on variables calculated on the candidates, e.g. the invariant mass and the signal probability of previous stages relevant for the recombined particle. At the final stage, it is up to the analyst to apply a cut on the tagging probability P_{FEI} of the reconstructed B_{tag} candidates to obtain a sample with the desired purity. The cut has to be chosen as a compromise between the required purity and the tagging efficiency. In most cases, the analyst is interested in a high tag-side efficiency, corresponding to a very loose cut on P_{FEI} . The different stages are structured as:

Stage 0 Collection of final state particles reconstructed from clusters and tracks.

Stage 1 Reconstruction of π^0 and J/ψ candidates.

Stage 2 Reconstruction of K_S^0 candidates.

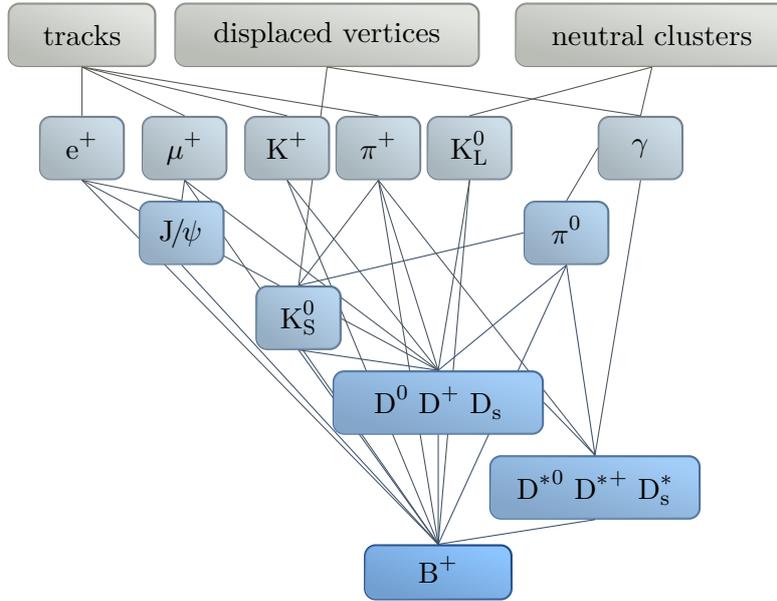


Figure 4.3.: Illustration of the bottom-up approach of the FEI. The graph depicts possible connections between the individual stages to reconstruct B_{tag} candidates in different tag channels. Adapted from [39].

Stage 3 Reconstruction of D candidates.

Stage 4 Reconstruction of D^* candidates.

Stage 5 Reconstruction of B candidates.

The individual BDTs of the various stages have to be trained in consecutive order on simulated MC events. Individual reconstruction channels are discarded if no reasonable training is possible because of an insufficient number of signal or background events. Due to the large amount of possible decay channels of the B meson and the (relatively) low branching fractions of the individual channels, the training process requires a large amount of MC events ($\sim \mathcal{O}(10^7)$). The training takes several days until the classifiers at all stages are trained.

The application of the FEI is fairly easy for the analyst: a path (containing all modules of the trained FEI) has to be added to the analyst's path, providing a `ParticleList` with B_{tag} candidates with an assigned P_{FEI} . These candidates can be used in the following analysis steps.

Although the FEI was developed for the Belle II experiment, it has been adapted to be used for converted Belle MC and data. This allows for a comparison to its predecessor algorithm used within `BASF`, the so-called Full Reconstruction (FR) [41]. The FEI shows an improved performance over the FR. This is, amongst other reasons, due to improved classifiers, the internally applied best-candidate selection and the inclusion of additional tag channels. A comparison of both algorithms

can be seen in Figure 4.4. The graphs show the purity and tag-side efficiency for charged hadronically reconstructed B_{tag} candidates for different cuts on the tag probability.

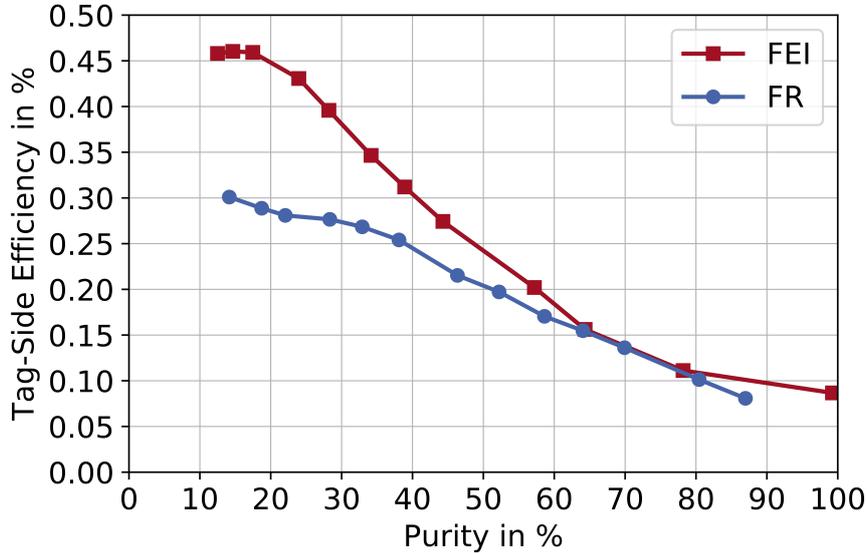


Figure 4.4.: Dependence of the tag-side efficiency and the purity of hadronically reconstructed B_{tag} candidates. Different cuts on the tagging probability of the Belle (FR, *blue line*) and Belle II (generic FEI, *red line*) algorithm were applied. Adapted from [39].

4.4.3 The Signal-specific FEI

The FEI allows for two different modes, referred to as the *generic* and *signal-specific* FEI. Both implementations are based on the same reconstruction steps. The generic FEI is trained on the whole $\Upsilon(4S)$ event and follows the traditional approach of the FR. Once trained, the generic FEI can be used for many different signal channels. In contrast, the signal-specific FEI is trained after the signal-side reconstruction, namely on the ROE of the B_{sig} candidate. The training has to be done specifically for each analysis. The advantage lies in the training on analysis-specific backgrounds, which occur through wrongly reconstructed signal-side candidates. After a correct reconstruction of the B_{tag} candidate no final state particles are left in the event. The event topology can be taken into account since the classifiers are trained on the tag-side of a specific signal decay.

The different performance of the specific and the generic FEI evaluated on $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ signal MC can be seen in Table 4.1. More information on the reconstruction process can be found in Chapter 5. As expected, the signal-specific FEI yields for a higher reconstruction efficiency of about 0.3%.

Table 4.1.: Comparison of the reconstruction efficiency between generic and signal-specific FEI, measured on 10^6 signal MC events for both final states. The best-candidate selection and all final cuts were applied, except for the cuts on the FEI probability, continuum and peaking background suppression (see Chapter 5). The numbers include the tag- and signal-side reconstruction.

Tagging algorithm	Reconstructed $\Upsilon(4S)$ candidates (%)	
	Electron	Muon
Generic FEI	1.50	1.63
Signal-specific FEI	1.80	1.95

Furthermore, significantly less CPU time is required during the application of the specific FEI. The algorithm runs only if a suitable signal candidate was reconstructed and only on a ROE, i.e. a sub-set of the final state particles. This is advantageous if the number of signal-side candidates (and hence the number of ROEs) is low. For signal channels of high candidate multiplicity the generic FEI is more beneficial.

Due to the signal reconstruction beforehand, roughly one order of magnitude more events are required for the training procedure. The classifiers are trained on MC with signal, charged and neutral B decays (see Section 5.1 for a detailed description of the MC samples). For the signal MC (but not for the remaining samples) a correctly reconstructed signal-side is required in training process.

CHAPTER 5

Analysis

In this chapter the analysis procedure is described. The used data samples are discussed in Section 5.1. The event reconstruction steps and the suppression of the backgrounds are presented in Section 5.2. The corrections applied to the MC are discussed in Section 5.3. In the last section the remaining background composition is described.

5.1 Data Samples

During its runtime of more than ten years the Belle experiment recorded a data set on the $\Upsilon(4S)$ resonance containing $(772 \pm 10) \times 10^6$ $B\bar{B}$ pairs. The data set is divided into experiments and runs, where a new experiment was usually started after a long maintenance downtime.¹ After three years of operation the Silicon Vertex Detector (SVD1) was damaged due to radiation exposure and was replaced after experiment 27. An off-resonance data set corresponding to an integrated luminosity of 79.4 fb^{-1} was recorded to allow for background studies by operating 60 MeV below the $\Upsilon(4S)$ resonance. Besides the $\Upsilon(4S)$ data, smaller data sets were collected operating the accelerator on the energies of the $\Upsilon(2S)$, $\Upsilon(3S)$ and $\Upsilon(5S)$ resonances.

The analysis is conducted as a blind analysis, consequently the whole selection process is developed on MC events. This requires a large amount of simulated events. The official Belle MC is structured into *streams*, where one stream of a certain decay type corresponds to the number of events in the recorded data set. The event simulation is divided into two stages. First, the events are generated

¹Belle used only odd experiment numbers.

with the **EvtGen** package according to a given decay table containing the branching fractions to simulate [27]. The package was specially developed for B physics and provides various models to simulate decays. In the second step, the **GEANT3** package simulates the interaction of generated particles within the detector and generates hits along their trajectories [42]. Supplementary background hits are added to incorporate beam background. These hits are generated from background events, which were recorded during every run.

The used MC samples are summarized in Table 5.1 and characterized below:

b \rightarrow **c** The *generic* B decay sample contains b \rightarrow c transitions, which includes both $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ (*mixed*) and $\Upsilon(4S) \rightarrow B^+B^-$ (*charged*) events. In total, ten streams of this type are available.

e⁺e⁻ \rightarrow **q \bar{q}** The *continuum* background arises from the direct transition to lighter quarks $e^+e^- \rightarrow q\bar{q}$ with $q = u, d, s, c$. The quark anti-quark pair fragments into hadrons and the component is by far the most dominant in data. Six streams of this non-B background are available.

Rare Decays of the type b \rightarrow q with $q = u, d, s$ which have very low branching ratios are simulated in an extra sample and scaled to 50 times the expected number of events on data. The signal decay $B^+ \rightarrow \ell^+\nu_\ell\gamma$ is contained in this sample and is removed, since it is simulated separately.

Peaking background To study the most problematic background in this analysis, four samples of so-called peaking background $B^+ \rightarrow \pi^0\ell^+\nu_\ell$ and $B^+ \rightarrow \eta\ell^+\nu_\ell$ with $\ell = e, \mu$ are generated, each containing 5×10^6 events. These decays peak typically in the signal region in the M_{miss}^2 variable which is used for signal extraction. Details on the **EvtGen** model used for the simulation of the $B^+ \rightarrow \pi^0\ell^+\nu_\ell$ decay can be found in Section 5.3.4.

b \rightarrow **u $\ell\nu_\ell$** This sample contains semileptonic decays of charged and neutral B mesons with an up quark. The sample is scaled by a factor of 20. Besides exclusive decays involving an up quark it contains decays via intermediate resonances. Since the $B^+ \rightarrow \pi^0\ell^+\nu_\ell$ and $B^+ \rightarrow \eta\ell^+\nu_\ell$ decays are simulated separately, they are removed from this sample.

B⁺ \rightarrow **$\ell^+\nu_\ell\gamma$** Two signal samples for each final state ($\ell = e, \mu$) containing 110×10^6 events are generated of which the largest part is required for the training of the tagging algorithm. Since there are no precise theory predictions and only upper limits have been experimentally determined the signal events are weighted to a partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma) = 5 \times 10^{-6}$ [43]. An exemplary event display can be seen in Figure 5.1.

Table 5.1.: Applied weights to scale the used MC samples to luminosity or expected (partial) branching fraction.

MC sample	Size	\mathcal{B}	w_{MC}
$b \rightarrow c$	9 streams	–	1/9
$e^+e^- \rightarrow q\bar{q}$	5 streams	–	1/5
Rare $b \rightarrow c$	50 streams	–	1/50
$B^+ \rightarrow \pi^0 e^+ \nu_e$	4905921	7.8×10^{-5}	1.227×10^{-2}
$B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$	4975003	7.8×10^{-5}	1.210×10^{-2}
$B^+ \rightarrow e^+ \nu_e$	4900003	3.8×10^{-5}	5.984×10^{-3}
$B^+ \rightarrow \eta \mu^+ \nu_\mu$	4850003	3.8×10^{-5}	6.045×10^{-3}
$b \rightarrow u \ell \nu_\ell$	20 streams	–	1/20
$B^+ \rightarrow e^+ \nu_e \gamma$	7047812	5.0×10^{-6}	5.474×10^{-4}
$B^+ \rightarrow \mu^+ \nu_\mu \gamma$	7112244	5.0×10^{-6}	5.424×10^{-4}

The described samples are weighted to the expected (partial) branching fractions with

$$w_{\text{MC}} = \frac{N_{B\bar{B}} \times 2 \times \mathcal{B}(\Upsilon(4S) \rightarrow B\bar{B}) \times \mathcal{B}(\text{MC type})}{N_{\text{gen}}}, \quad (5.1)$$

where $N_{B\bar{B}}$ denotes the number of $B\bar{B}$ pairs and N_{gen} is the number of generated events. The applied MC weights can be found in Table 5.1.

5.2 Event Reconstruction

The search for a rare decay like $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ requires a careful selection process to reconstruct signal candidates on data. The selection and reconstruction process consists of several steps. In the first step, final state particles are selected and a B_{sig} candidate is reconstructed. In the second step, the accompanying B_{tag} candidate is reconstructed by the tagging algorithm. Finally, the complete event is reconstructed by combining both B_{sig} and B_{tag} to a $\Upsilon(4S)$ candidate. To reduce combinatorial background, cuts are applied in all of these reconstruction steps. Further background originating from continuum events and the peaking background decays is reduced by making use of multivariate methods.

A similar reconstruction of $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays is applied. This sample is later used in the fit to constrain the irreducible peaking background from $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays in the nominal $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample².

In the following the aforementioned steps are explained in more detail.

²Throughout the thesis the two reconstruction samples are distinguished as the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample and the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample.

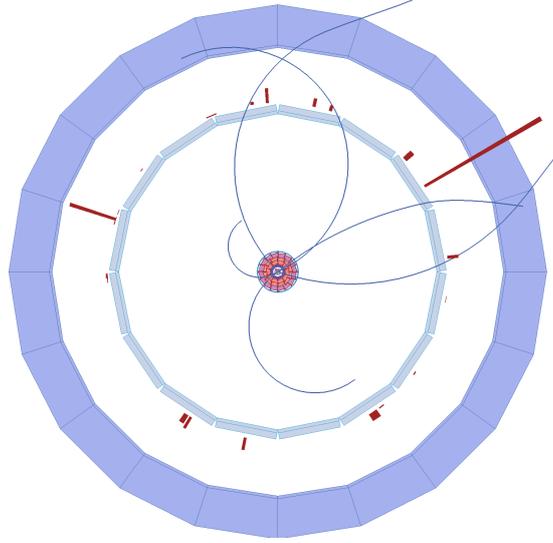


Figure 5.1.: Exemplary event display of a simulated $B^+ \rightarrow e^+ \nu_e \gamma$ event. Reconstructed tracks and ECL clusters are depicted as blue curved lines and red boxes, respectively. The SVD is located in the innermost ring, the middle and outer rings are support structures of the TOP and the ECL.

5.2.1 Event Selection

Prior to the event reconstruction, some basic cuts are applied to remove background events. Beam induced background and two-photon events are reduced with a default skim for hadronic events known as *HadronBJ skim* [23, p.52]. At least three tracks originating from the IP ($|\Delta r| < 2$ cm and $|\Delta z| < 4$ cm) with a transverse momentum p_T of more than 0.1 GeV are required. The visible energy in the event, calculated as the sum of the energies of tracks and photons, must be larger than 20% of the center-of-mass energy E_{CMS} . For more information the reader is referred to the aforementioned reference.

To remove curling tracks a cut of $|\Delta r| < 2$ cm and $|\Delta z| < 4$ cm is applied on charged tracks. For each event a maximum of twelve tracks is allowed, whereof only one is used for the signal-side reconstruction. Figure 5.2 shows the number of tracks for different MC samples after the signal selection without the cut applied. In addition, cuts on the energy of reconstructed photons are applied, depending on the region of the corresponding clusters in the ECL. For the forward (backward) end cap cuts of $E_\gamma > 100$ MeV ($E_\gamma > 150$ MeV) are used, whereas in the barrel region $E_\gamma > 50$ MeV is required. These cuts are referred to as the *goodGamma* cuts and are the default in most Belle analyses. All cuts applied on event selection stage can be found in Table 5.2.

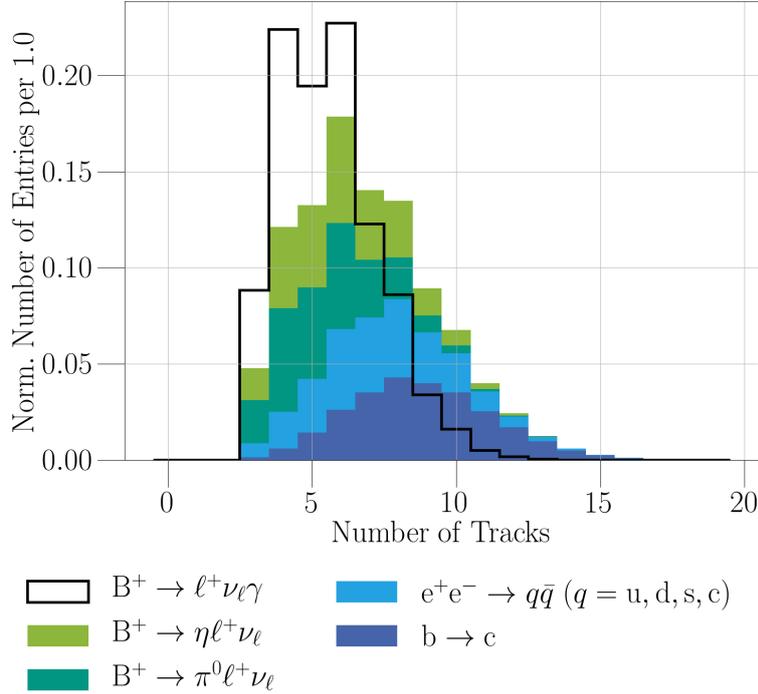


Figure 5.2.: Number of tracks with $|\Delta r| < 2$ cm and $|\Delta z| < 4$ cm per event. Only the background samples are stacked in the histogram. The signal and peaking background samples have on average less tracks, due to only one track on the signal-side.

5.2.2 Signal-side Selection

Lepton candidates are required to have a lepton identification (LID) variable of larger than 0.8. Due to theoretical considerations on the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay (see Section 2.2) only high-energetic photons are selected with $E_\gamma > 1.0$ GeV. Bremsstrahlung can decrease the resolution of the fit variable M_{miss}^2 for the electron final state. Hence, for each electron candidate the four-vector of a photon within a cone of 5.0° and an energy of less than 1.0 GeV is added to the four-vector, see Section 4.3 for more details. A loose mass cut of $M_B \in (1.0, 6.0)$ GeV is applied to remove combinatorial background in the B_{sig} reconstruction.

Similar cuts are applied for the dedicated $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample. The signal-side is reconstructed by replacing the photon by a π^0 candidate in the selection. A cut on the lepton momentum and the pion mass is applied. The cuts are shown in Table 5.3.

Table 5.2.: Cuts applied on the event selection stage.

Variable	Cut
<i>HadronBJ skim</i>	see text
<i>goodGamma</i>	see text
$ \Delta r $	$< 2 \text{ cm}$
$ \Delta z $	$< 4 \text{ cm}$
N_{tracks}	≤ 12

Table 5.3.: Cuts applied on the signal-side reconstruction stage.

Sample	Variable	Cut
$B^+ \rightarrow \ell^+ \nu_\ell \gamma$	<i>eID</i>	> 0.8
	<i>muID</i>	> 0.8
	E_γ	$> 1.0 \text{ GeV}$
	M_B	$\in (1.0, 6.0) \text{ GeV}$
$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$	<i>eID</i>	> 0.8
	<i>muID</i>	> 0.8
	p_ℓ	$\geq 300 \text{ MeV}$
	M_{π^0}	$\in (115, 152) \text{ MeV}$
	M_B	$\in (1.0, 6.0) \text{ GeV}$

5.2.3 Tag-side Reconstruction

The B_{tag} candidate is reconstructed by the B-tagging algorithm. As mentioned before, the signal-specific FEI is used which is explicitly trained for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay. Before reconstructing the tag-side, the corresponding ROE is cleaned up. Photon candidates must have a cluster ratio of $E9E25 > 0.9$, calculated as a ratio of energies of the 3×3 to 5×5 CsI(Tl) crystals with the maximal energy deposition. This variable is a good indicator of the ECL shower shape, where values near 1.0 indicate a sharp shower. Events which do not provide a reasonable tag-side are removed with loose cuts on the energy difference ΔE between the ROE and the beam energy, and the beam-constrained mass M_{bc} calculated on the ROE. The cuts are summarized in Table 5.4.

The technical details of the tagging algorithm are presented in Section 4.4. The training of the specific FEI is done after the signal-side reconstruction. Due to the previously applied B_{sig} selection, a large set of MC events is required to obtain a reasonable amount of events for the training. The algorithm is trained on 10^8 $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ MC events for each final state and one stream of generic mixed

Table 5.4.: Cuts applied for the ROE cleaning.

Variable	Cut
$E9E25_{\text{ROE}}$	$> 0.9 \text{ GeV}$
$M_{\text{bc, ROE}}$	$> 4.8 \text{ GeV}$
ΔE_{ROE}	$< 2.0 \text{ GeV}$

and charged $b \rightarrow c$ MC events. In this analysis only hadronic channels are used for the B_{tag} reconstruction, leading to one (undetected) neutrino in a correctly reconstructed event. The following tag channels are used:

- $B^+ \rightarrow \bar{D}^0 \pi^+$
- $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0$
- $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^0 \pi^0$
- $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^-$
- $B^+ \rightarrow \bar{D}^0 \pi^+ \pi^+ \pi^- \pi^0$
- $B^+ \rightarrow \bar{D}^0 D^+$
- $B^+ \rightarrow \bar{D}^0 D^+ K_S^0$
- $B^+ \rightarrow \bar{D}^{*0} D^+ K_S^0$
- $B^+ \rightarrow \bar{D}^0 D^{*+} K_S^0$
- $B^+ \rightarrow \bar{D}^{*0} D^{*+} K_S^0$
- $B^+ \rightarrow \bar{D}^0 D^0 K^+$
- $B^+ \rightarrow \bar{D}^{*0} D^0 K^+$
- $B^+ \rightarrow \bar{D}^0 D^{*0} K^+$
- $B^+ \rightarrow \bar{D}^{*0} D^{*0} K^+$
- $B^+ \rightarrow D_s^+ \bar{D}^0$
- $B^+ \rightarrow \bar{D}^{*0} \pi^+$
- $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0$
- $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^0 \pi^0$
- $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^-$
- $B^+ \rightarrow \bar{D}^{*0} \pi^+ \pi^+ \pi^- \pi^0$
- $B^+ \rightarrow D_s^{*+} \bar{D}^0$
- $B^+ \rightarrow D_s^+ \bar{D}^{*0}$
- $B^+ \rightarrow \bar{D}^0 K^+$
- $B^+ \rightarrow D^- \pi^+ \pi^+$
- $B^+ \rightarrow D^- \pi^+ \pi^+ \pi^0$
- $B^+ \rightarrow J/\psi K^+$
- $B^+ \rightarrow J/\psi K^+ \pi^+ \pi^-$
- $B^+ \rightarrow J/\psi K^+ \pi^0$
- $B^+ \rightarrow J/\psi K_S^0 \pi^+$

5.2.4 $\Upsilon(4S)$ Reconstruction

A $\Upsilon(4S)$ candidate can be reconstructed after the reconstruction of the tag- and signal-side. A best candidate selection is performed if more than one $\Upsilon(4S)$

candidate per event is found, choosing the candidate with the highest tagging probability P_{FEI} . The cuts applied to the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ samples are listed in Table 5.5; including cuts on the invariant mass M of the $\Upsilon(4S)$ candidate, the energy difference ΔE between the B_{tag} candidate and the beam energy and the beam-constrained mass M_{bc} of the B_{tag} candidate. In addition, a cut on the angle between the signal-side photon and the missing momentum of $\cos(\theta_{\nu\gamma}) > -0.9$ is required for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample to remove continuum events [16, p. 43].

In Chapter 7 it is described that the off-resonance sample is dominated by non-resonant background events at very low values of P_{FEI} . This background is removed by a moderate cut on the tagging probability of $P_{\text{FEI}} > 0.01$ with reasonable signal efficiency loss, see Table 5.6. The distribution of P_{FEI} for the signal and background MC samples after the selection is shown in Figure 5.3. As expected, most of the background events can be found in the lower P_{FEI} region, since it is more challenging to reconstruct a reasonable B_{tag} candidate with a high tagging probability. The region of larger P_{FEI} values is mainly populated by signal and peaking background events mimicking the signal.

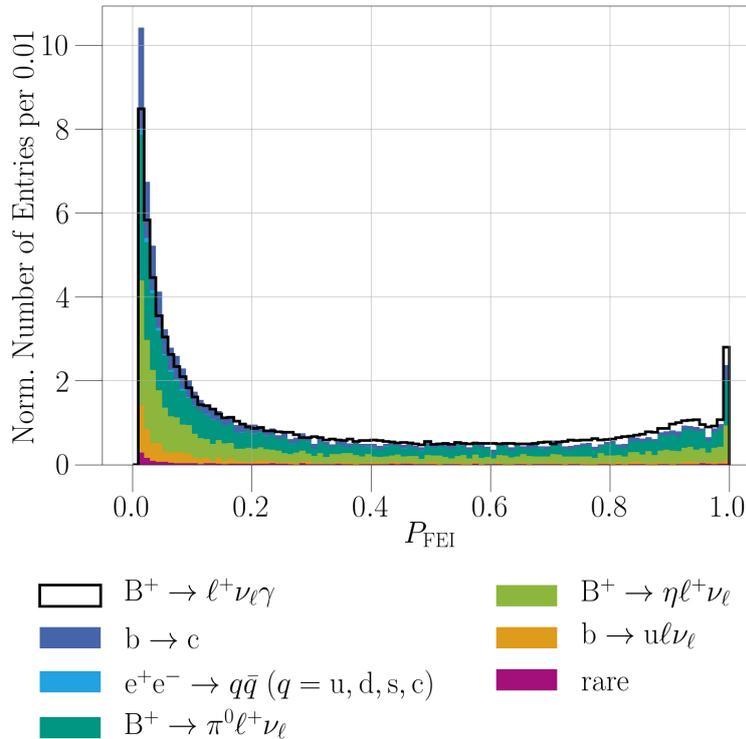


Figure 5.3.: Distribution of the tagging probability for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample.

The signal yield is extracted by making use of the squared missing mass M_{miss}^2 variable calculated as

$$\begin{aligned} M_{\text{miss}}^2 &= (p_{\text{B}_{\text{sig}}} - p_\ell - p_\gamma)^2 \\ &= \left(\left(\begin{array}{c} \frac{E_{\text{CMS}}}{2c} \\ -\vec{p}_{\text{B}_{\text{tag}}} \end{array} \right) - p_\ell - p_\gamma \right)^2, \end{aligned} \quad (5.2)$$

where $p_{\text{B}_{\text{sig}}}$, p_ℓ and p_γ denote the four-momenta of the B_{sig} , lepton and photon candidates, respectively. The four-momentum of the B_{sig} candidate can be replaced by the four-momentum of the B_{tag} candidate with opposite sign for the momentum component, since the $\text{B}\bar{\text{B}}$ pair is produced back-to-back in the center-of-mass frame. The energy of the B_{tag} candidate can be replaced by half of the center-of-mass energy. Since the neutrino is massless, M_{miss}^2 peaks around zero for correctly reconstructed signal events.

To further reject background events the signal-side photon (in the $\text{B}^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample) and the daughter photons of the π^0 candidate (in the $\text{B}^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample) are required to have $E9E25_\gamma > 0.9$. Assuming that the event has been correctly reconstructed, no charged tracks should remain. Hence, no additional tracks are allowed after the $\Upsilon(4\text{S})$ reconstruction. The extra energy E_{ECL} , defined as the sum of the energy deposited in the ECL clusters not used for event reconstruction, has to be lower than 0.9 GeV. The extra energy should be zero if no additional background contributes and the $\Upsilon(4\text{S})$ was correctly reconstructed as no photons are left.

Table 5.5.: Cuts applied on the $\Upsilon(4\text{S})$ reconstruction stage for both samples.

Variable	Cut
M	$\in [7.5, 10.5] \text{ GeV}$
ΔE	$\in [-0.15, 0.1] \text{ GeV}$
M_{bc}	$\in [5.27, 5.29] \text{ GeV}$
E_{ECL}	$\leq 0.9 \text{ GeV}$
M_{miss}^2	$\in (-1.5, 3.0) \text{ GeV}^2$
$E9E25_\gamma$	> 0.9
P_{FEI}	> 0.01
Remaining N_{tracks}	$= 0$

5.2.5 Continuum Suppression

One of the main backgrounds in the analysis originates from continuum events. As noted above, the background is caused by non-resonant $e^+e^- \rightarrow q\bar{q}$ where q denotes a light quark. Continuum events exhibit a different event structure compared to $B\bar{B}$ events due to the underlying kinematics. The light $q\bar{q}$ pair moves back-to-back with large momentum, leading to two jets. In contrast, $B\bar{B}$ pairs decay more isotropically, since they are almost produced at rest in the center-of-mass system. The situation is depicted in Figure 5.4.

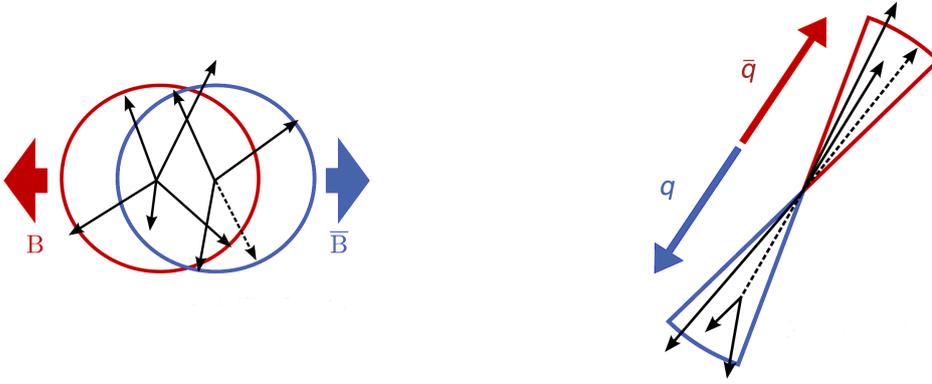


Figure 5.4.: Event topology of a $B\bar{B}$ (*left*) and a continuum (*right*) event. The latter shows a jet-like distribution due to the high momentum of the quark pair. Adapted from [44].

A set of variables describing the event topology can be exploited to identify and suppress continuum events:

$T_{\mathbf{B}}$, T_{ROE} The magnitude of the thrust of the B_{sig} candidate and the ROE, respectively. The thrust T is calculated from the momenta \vec{p}_i of the final state particles as

$$T = \frac{\sum_i^N |\vec{T}\vec{p}_i|}{\sum_i^N |\vec{p}_i|}, \quad (5.3)$$

where \vec{T} denotes the direction of the maximal total momentum.

$\cos \theta_{\mathbf{B},z}$, $\cos \theta_{\mathbf{B},\text{ROE}}$ The angle between the thrust axis of the daughter particles of the B_{sig} candidate and the z -axis and the ROE, respectively. As stated above, continuum events are more jet-like and so large angles are expected between the B_{sig} candidate and its ROE. The distribution is uniform for $B\bar{B}$ events.

R2 To characterize the event shape by energy and momentum flow in the event the so-called Fox-Wolfram Moments were developed [45]. The moments are

calculated as

$$H_l = \sum_{i,j}^N \frac{|\vec{p}_i| |\vec{p}_j|}{s} P_l(\cos(\phi_{ij})), \quad (5.4)$$

where N is the number of particles in the event, s is the squared center-of-mass energy, \vec{p}_x is the momentum of the particle x , ϕ_{ij} is the angle between the particles i and j , and P_l is the l -th Legendre polynomial. The reduced Fox-Wolfram Moment $R2$ is defined as the ratio $R2 = H_2/H_0$.

Kakuno-Super-Fox-Wolfram Moments The improved Fox-Wolfram-Moment were developed by the Belle collaboration [23, p.114]. In total there are 17 such moments.

Cleo Cones In the 90's the CLEO Collaboration introduced the so-called Cleo Cones. Nine cones in 10° steps around the B_{sig} thrust axis are defined. Within these intervals the momentum flow is calculated as the scalar sum of the final state particles pointing in the interval [46].

To suppress continuum events a BDT is used to combine all variables in a final discriminating variable P_{CS} . The multivariate method is trained on an independent set of signal and continuum MC events. Since the full reconstruction of the event already rejects most of the continuum background, the method is trained and applied only on events surviving the selection process, i.e. after the $\Upsilon(4S)$ reconstruction. All aforementioned cuts are applied except for the cuts on M_{bc} , $E9E25_{\gamma_{\text{sig}}}$ and P_{FEI} to retain enough statistics for the training process. One stream of continuum MC and 10^7 signal MC events for each final state are used. The performance of the classifier on an independent test sample for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ selection can be seen in Figure 5.5. The same procedure for continuum suppression is applied for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample. The receiver operating characteristic curve (ROC), an overtraining check and the distributions of the training variables can be found in Appendix B.

5.2.6 Peaking Background Suppression

As mentioned above, another dominant source of background in the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ analysis originates from charged or neutral semileptonic $B \rightarrow X_u \ell^+ \nu_\ell$ decays where X_u indicates a light meson (π^0 , η , ρ^0 , η' , ω , π^+ or ρ^+); mostly dominated by $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ and $B^+ \rightarrow \eta \ell^+ \nu_\ell$ decays. This peaking background is rejected in a two-step process.

In the first step, a π^0 mass veto is applied to suppress $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays. When the light meson decays into two photons, one photon can be misidentified as the signal photon and the other one is assigned to the tag-side. To reject these events, the photon of the signal-side is recombined with any photon of the ROE to create a

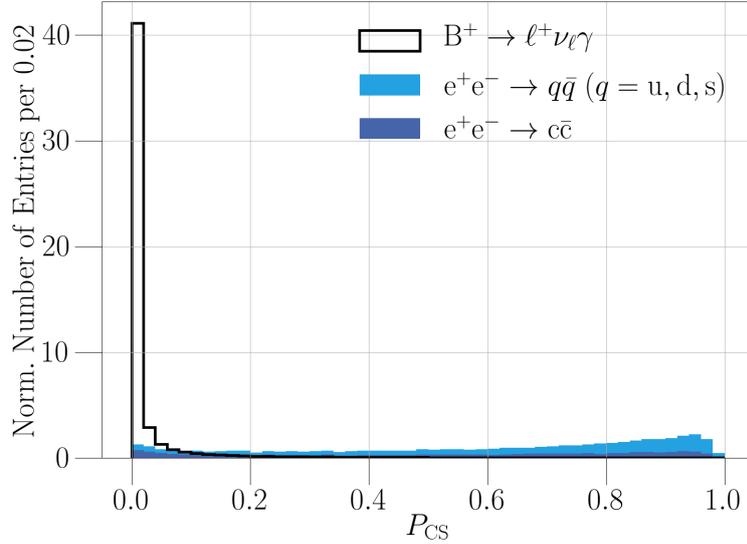


Figure 5.5.: Application of the continuum suppression BDT after the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ selection. Signal events are largely classified as non-continuum events.

π^0 candidate. An event is rejected if a valid π^0 candidate with an invariant mass of $M_{\gamma_{\text{sig}} \gamma_{\text{tag}}} \in (0.11, 0.16)$ GeV is found. The composition of the peaking background in the signal region after the continuum suppression and the applied mass veto can be seen in Figure 5.6. Despite the veto not the entire peaking background can be suppressed. In some cases one of the photons cannot be found, since it flew outside the detector acceptance or was low-energetic and consequently removed during the reconstruction. On the other hand, photons from the π^0 decay which are emitted in the more or less the same direction might be reconstructed as a single ECL cluster. Cluster shape variables such as $E9E25$ help to identify such cases.

In the second step, a BDT is used to identify such events. The following variables were found to discriminate the remaining background:

ECL cluster hits Number of hits associated to the ECL cluster used for the signal-side photon reconstruction.

E9E25 Ratio of energies in inner 3x3 and 5x5 cells of the ECL cluster.

ECL cluster LAT Lateral distribution of the ECL cluster.

$\theta_{\gamma, p_{\text{miss}}}$ Angle of the signal-side photon and the missing momentum \vec{p}_{miss} in the rest frame of the B_{sig} candidate.

E_{ECL} The extra energy.

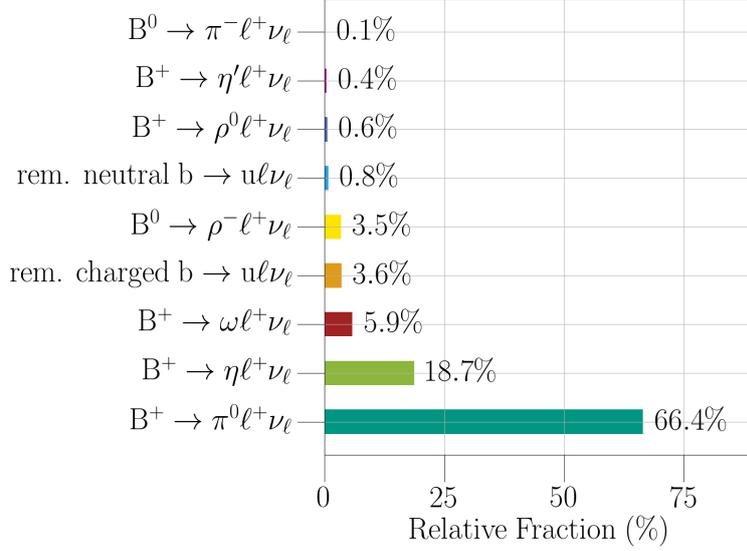


Figure 5.6.: Composition of the $b \rightarrow u \ell \nu_\ell$ background decays in the signal region of $M_{\text{miss}}^2 \in (-0.5, 0.5) \text{ GeV}^2$ after continuum suppression and π^0 mass veto.

Energy asymmetry Energy asymmetry of the daughter particles of the B_{sig} candidate calculated as

$$A_E = \frac{\prod_i E_i}{\sum_i E_i}, \quad (5.5)$$

where i denotes a daughter particle. The variable reveals the asymmetry in the energy distribution of the lepton and photon candidate.

The BDT is trained on samples of 5×10^6 $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$, $B^+ \rightarrow \eta \ell^+ \nu_\ell$ and $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ MC events for each final state. As before, the BDT trained only on events which survived the selection process. The π^0 mass veto and all aforementioned cuts are applied except for the cuts on M_{bc} , $E9E25_{\gamma_{\text{sig}}}$ and P_{FEI} to retain enough statistics. The performance of the classifier on an independent sample can be seen in Figure 5.7. The ROC curve, an overtraining check and the distributions of the training variables can be found in Appendix B.

5.2.7 Optimization

The cuts on the BDT output variables are optimized for both samples, the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ and the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample, within the signal window $M_{\text{miss}}^2 \in (-0.5, 0.5) \text{ GeV}^2$. This is done by maximizing Punzi's figure of merit

$$f.o.m. = \frac{\epsilon}{\frac{\sigma}{2} + \sqrt{B}}, \quad (5.6)$$

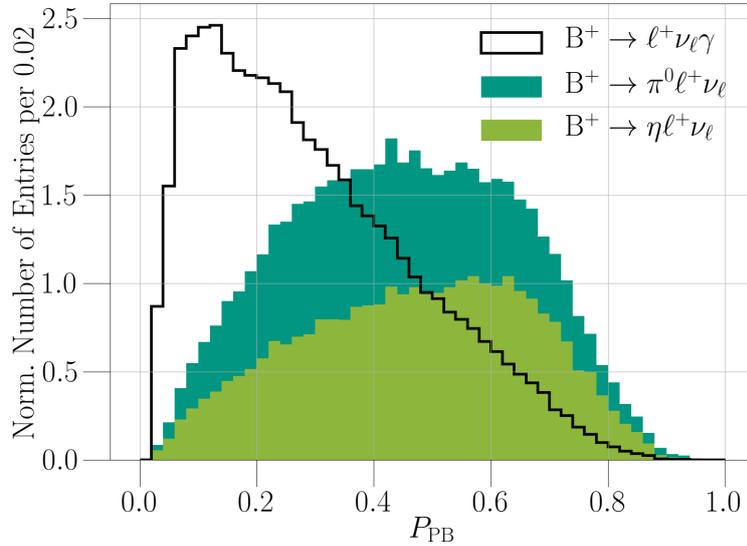


Figure 5.7.: Application of the peaking background suppression BDT on an independent MC sample.

where ϵ is the reconstruction efficiency, σ is the desired significance ($\sigma = 3$) and B is the number of background events [47]. To find the optimal cut on the classifier output the `scipy.optimize` algorithm is used [48].

For the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample a grid-search is done to check the result of the two-dimensional optimization, visualized in Figure 5.8. Optimized cuts on the continuum suppression of $P_{CS} = 0.034$ and on the peaking background suppression of $P_{PB} = 0.44$ are obtained. For the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample a cut of $P_{CS} = 0.008$ is found. The optimized cuts are applied in the final selection step. The signal efficiency and the background rejection rate of the final cuts for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay can be found in Table 5.6. The M_{miss}^2 distributions after all selection cuts for both final states can be seen in Figure 5.9.

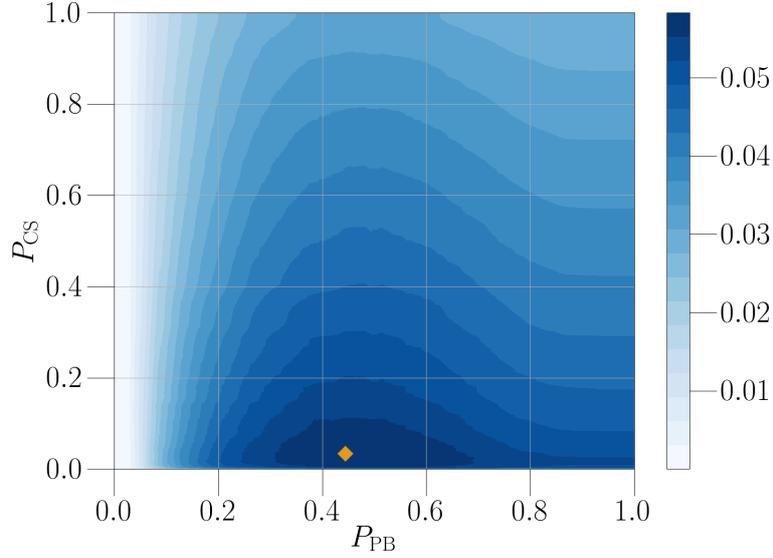


Figure 5.8.: Cut optimization of the continuum and peaking background suppression for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample. The *orange point* represents the result of the optimization algorithm, while the result of the grid search is indicated by the color gradient. The more blueish the higher the figure of merit. The signal component is scaled to a partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$. The background components are scaled to luminosity.

Table 5.6.: Signal efficiency (ϵ) and background rejection ($1 - \epsilon$) for the different cuts applied individually and combined on the individual MC samples.

	Signal efficiency	Background rejection					
	$B^+ \rightarrow \ell^+ \nu_\ell \gamma$	Rare	$b \rightarrow u \ell \nu_\ell$	$B^+ \rightarrow \eta \ell^+ \nu_\ell$	$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$	$e^+ e^- \rightarrow q \bar{q}$	$b \rightarrow c$
E9E25	0.960	0.080	0.050	0.040	0.080	0.20	0.080
$\theta_{\nu\gamma}$	1.000	0.150	0.010	0.020	0.020	0.14	0.010
M_{bc}	0.780	0.820	0.750	0.410	0.370	0.87	0.780
$M_{\pi\text{ veto}}$	0.960	0.290	0.450	0.080	0.680	0.43	0.430
P_{CS}	0.860	0.630	0.340	0.200	0.210	0.95	0.460
P_{PB}	0.860	0.700	0.380	0.440	0.430	0.63	0.520
P_{FEI}	0.640	0.650	0.590	0.440	0.430	0.68	0.650
Comb.	0.406	0.992	0.972	0.839	0.911	0.99	0.989

5.3 Monte Carlo Corrections

The MC simulation suffers from imperfections arising from different sources like imprecise branching fractions of decay channels or insufficiently modeled hadronic decays and detector response. Corrections are applied to address these shortcomings in case they can be identified and measured.

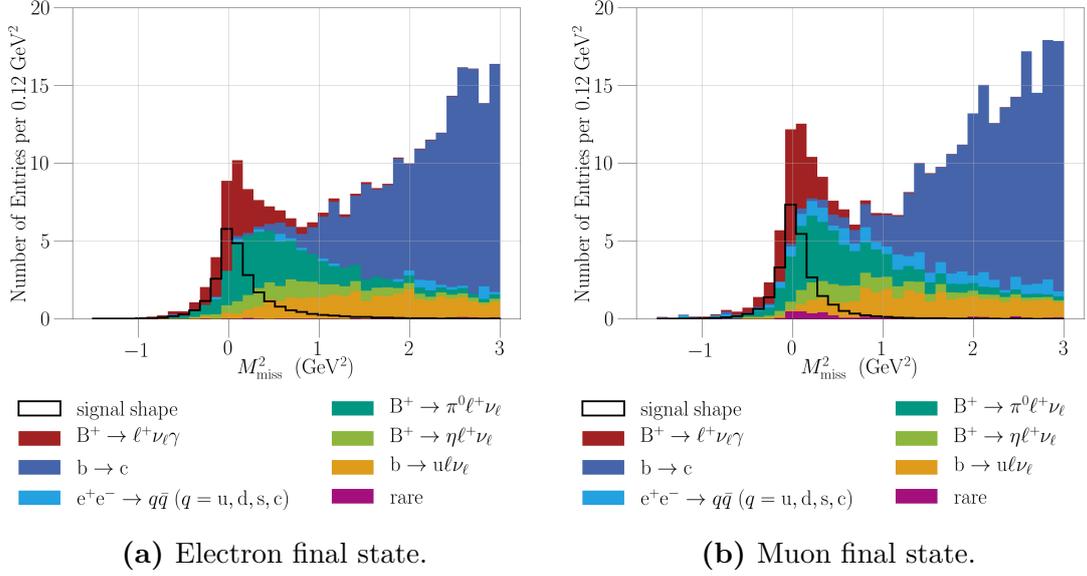


Figure 5.9.: M_{miss}^2 distributions of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample after the final selection with a simulated partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$. The background components are scaled to luminosity. Corrections to the MC are applied.

5.3.1 Tag Correction

Discrepancies between MC and data cause differences in the efficiency of the tagging algorithm. Consequently, the applied algorithm has to be calibrated on data. In the following, the calibration procedure is described³.

A correction factor can be extracted by making use of well-known semileptonic *calibration channels* with relatively large branching fractions:

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

These channels replace the signal channel of the nominal analysis; the applied cuts in the reconstruction can be found in Appendix C. The correction factor is then calculated from the ratio of the extracted signal yield on data and MC as

$$\epsilon_c = \frac{N_c^{\text{data}}}{N_c^{\text{MC}}}, \quad (5.7)$$

³A more detailed description of the calibration procedure can be found in Reference [49].

where c denotes the calibration channel. The signal yield is extracted with a template fit to the M_{miss}^2 distribution. To obtain a robust correction factor which can be used in the analysis, a fit for each of the three calibration channels is performed. The correction factors ϵ_c of the individual calibration channels are then averaged to obtain a global correction factor. In principle, the correction factor should be independent of the calibration channel. The calibration procedure is applied on one stream of generic and continuum MC and on the whole recorded data set.

For each calibration channel MC corrections are applied which refer to the signal-side. These corrections include an update of the branching ratio and efficiency corrections for the LID cuts, which are explained in more detail in the next subsection. The applied branching ratio corrections can be found in Table 5.7.

The calibration also corrects for the fact that in the simulation equal branching ratios of 1/2 for $\Upsilon(4S) \rightarrow B^+B^-$ and $\Upsilon(4S) \rightarrow B^0\bar{B}^0$ decays were assumed. The most recent values from the PDG group are $\mathcal{B}(\Upsilon(4S) \rightarrow B^+B^-) = 0.486 \pm 0.006$ and $\mathcal{B}(\Upsilon(4S) \rightarrow B^0\bar{B}^0) = 0.514 \pm 0.006$ [4].

Different sources of systematic uncertainties are taken into account. For the tracking efficiency an uncertainty of 0.34% per track is assumed (see Section 8.2). Further, a systematic uncertainty of the LID correction (see next subsection) and on the branching ratio correction is considered.

The results of the fit for each calibration channel can be seen in Figure 5.11, the used templates can be found in Appendix C. The individual calibration factors are in good agreement with each other. As additional check calibration factors are determined for each tag channel per calibration channel, as can be seen in Figure 5.10. In most cases the individual calibration factors show a good agreement.

The average calibration factor is found as

$$\epsilon_{\text{all}} = \frac{N_{\text{data}}}{N_{\text{MC}}} = 0.825 \pm 0.014 \pm 0.049, \quad (5.8)$$

where the first uncertainty is statistical and the second is systematic. The overall

Table 5.7.: Branching ratios used for Belle MC simulation and the latest PDG average of the calibration channels [4].

Decay	$\mathcal{B}_{\text{MC}} (10^{-2})$	$\mathcal{B}_{\text{PDG}} (10^{-2})$
$B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$	2.31	2.27 ± 0.11
$D^0 \rightarrow K^- \pi^+$	3.82	3.89 ± 0.04
$D^0 \rightarrow K^- \pi^+ \pi^0$	13.08	14.2 ± 0.5
$D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$	7.21	8.11 ± 0.15

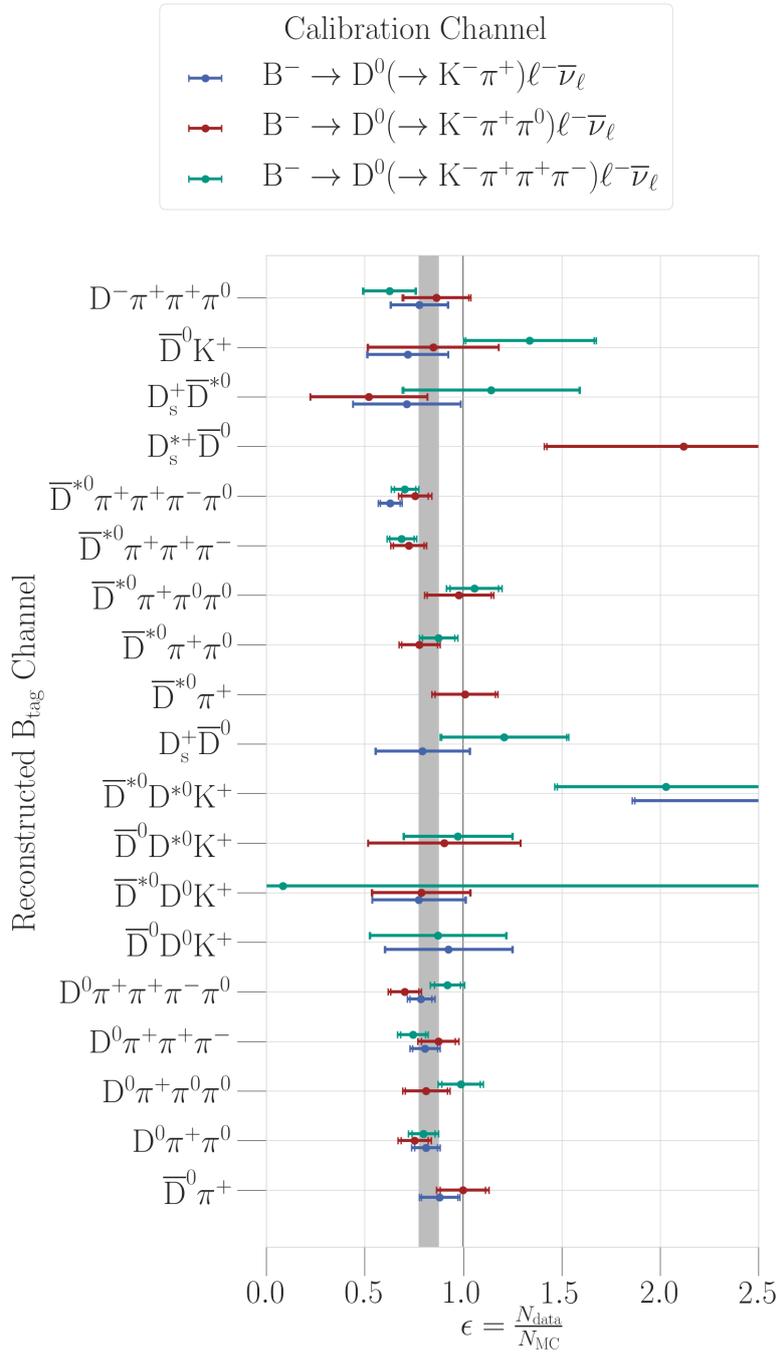


Figure 5.10.: Calibration factors for individual tag and calibration channels of the specific FEL. With statistical (*inner interval*) and systematic uncertainties (*outer interval*). Some tag channels are excluded due to low statistics and are hence not shown. The *gray band* represents the 1σ uncertainty band of the global calibration factor.

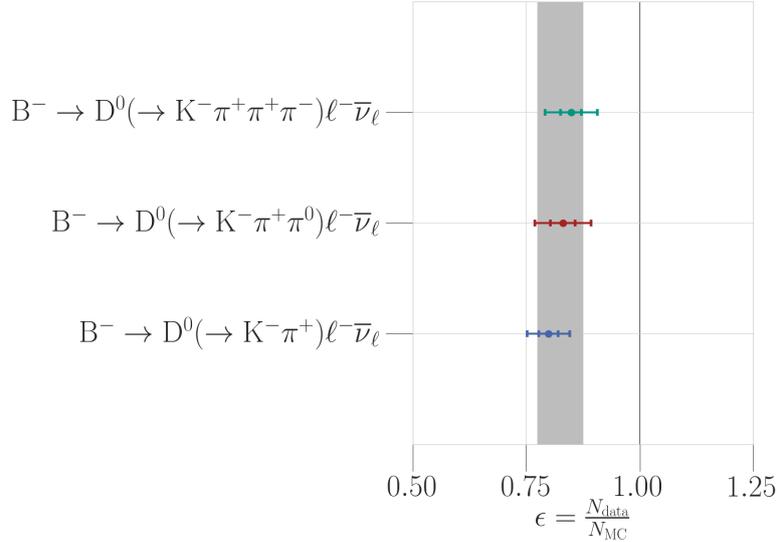


Figure 5.11.: Calibration factors averaged over the tag channels of the specific FEI shown in Figure 5.10. With statistical (*inner interval*) and systematic uncertainties (*outer interval*). For a perfect MC simulation a value of one would be expected. The *gray band* represents the 1σ uncertainty band of the global calibration factor.

calibration factor is used to weight the MC events and reveals that the applied tagging algorithm has a lower performance on data. A correction factor of one would be expected for a perfect data-MC agreement. The result is in good agreement with a similar result obtained for the generic FEI of $\epsilon_{\text{all}}^{\text{gen}} = 0.803 \pm 0.009 \pm 0.050$ [49]⁴.

5.3.2 LID Correction

The correction of the LID efficiency was studied in Reference [50]. The differences between MC and data were investigated on a high statistic sample of the process $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ with $\ell = e, \mu$ for different bins in the polar angle θ , the momentum in the lab frame p_{lab} and the LID cut. To measure a possible influence of a hadronic environment in the event, the latter process was compared to inclusive decays $B \rightarrow X J/\psi (\rightarrow \ell^+\ell^-)$. The found discrepancies were considered as additional systematic uncertainty. The Belle Joint PID group provides tables containing the correction factors with one statistical and two systematic uncertainties [51]. Different tables have to be used depending on the experiment

⁴The results of the generic FEI were obtained with FEI version 4.0. In this thesis additional cuts on the particle ID for K^+ , π^+ candidates; a mass cut on π^0 candidates and a cut on the FEI output probability are applied. All applied cuts can be found in Appendix C.

and run number of the event.

The corrections are applied for the electron (muon) signal-side track for each reconstructed event. The applied weights for the different MC components can be seen in Figure 5.12. The average correction factor of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample is given as 0.978 ± 0.017 (0.951 ± 0.019) for the electron (muon) final state. For the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample 0.981 ± 0.018 (0.957 ± 0.020) is found for the electron (muon) final state.

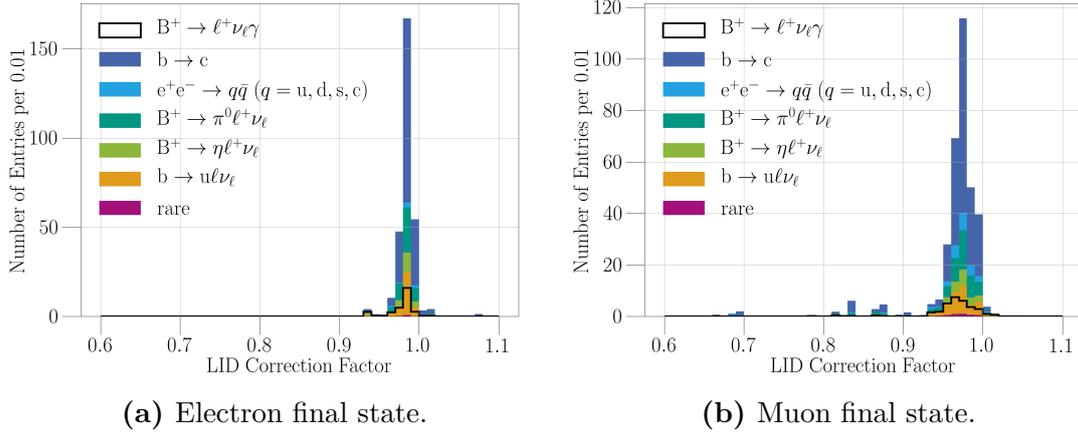


Figure 5.12.: Applied LID correction factors of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample.

5.3.3 Branching Fraction Correction

The branching fractions of the $b \rightarrow u \ell \nu_\ell$ MC samples are updated to the latest PDG world averages. The used branching ratios in the simulation, the latest PDG average and the calculated correction factors for different channels can be found in Table 5.8.

Table 5.8.: Branching ratios used for simulation and for reweighting the $b \rightarrow u \ell \nu_\ell$ MC samples [4, 52].

Decay mode	$\mathcal{B}_{\text{MC}} (10^{-4})$	$\mathcal{B}_{\text{PDG}} (10^{-4})$	Correction
$B^+ \rightarrow \rho^0 \ell^+ \nu_\ell$	1.49	1.58 ± 0.11	1.060 ± 0.074
$B^+ \rightarrow \eta' \ell^+ \nu_\ell$	0.33	0.23 ± 0.08	0.697 ± 0.242
$B^+ \rightarrow \omega \ell^+ \nu_\ell$	1.15	1.19 ± 0.08	1.035 ± 0.078
$B^0 \rightarrow \pi^- \ell^+ \nu_\ell$	1.36	1.47 ± 0.06	1.081 ± 0.044
$B^0 \rightarrow \rho^- \ell^+ \nu_\ell$	2.77	2.94 ± 0.21	1.061 ± 0.076

5.3.4 $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ EvtGen Model

The `EvtGen` package provides a variety of models to simulate decays. Semileptonic decays like $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ can be simulated with the `EvtSLPole` model which is based on calculations of QCD sum rules by *P. Ball and R. Zwicky* [53]. The Belle collaboration used the model for the production of the $b \rightarrow u \ell \nu_\ell$ MC samples.

However, it was found that the form factors were not correctly implemented and lead to an unphysical drop for large values of q^2 . The behavior can be seen in Figure 5.13a. The black line indicates the `EvtGen` implementation, whereas the red line corresponds to the description of the form factors in Reference [53]. The latest BCL prediction is shown as green line.

This flaw can be corrected by reweighting the events to the latest prediction. Unfortunately, the large q^2 region is sparsely populated using the `EvtGen` implementation. This would lead to an disadvantageous situation of very few events with large weights in this region. Instead, the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ MC samples are produced with modified parameters of the `EvtSLPole` model [54]. The different models are illustrated in Figure 5.13b, where the modified model (*blue line*) shows a populated high q^2 region. These samples are used to reweight events with moderate weights to the latest BCL prediction (*green line*) as illustrated in Figures 5.14a and 5.14b. In principle, other modes of the $b \rightarrow u \ell \nu_\ell$ MC using the `EvtSLPole` model are affected. Since these modes give only a negligible contribution to the background they are neglected.

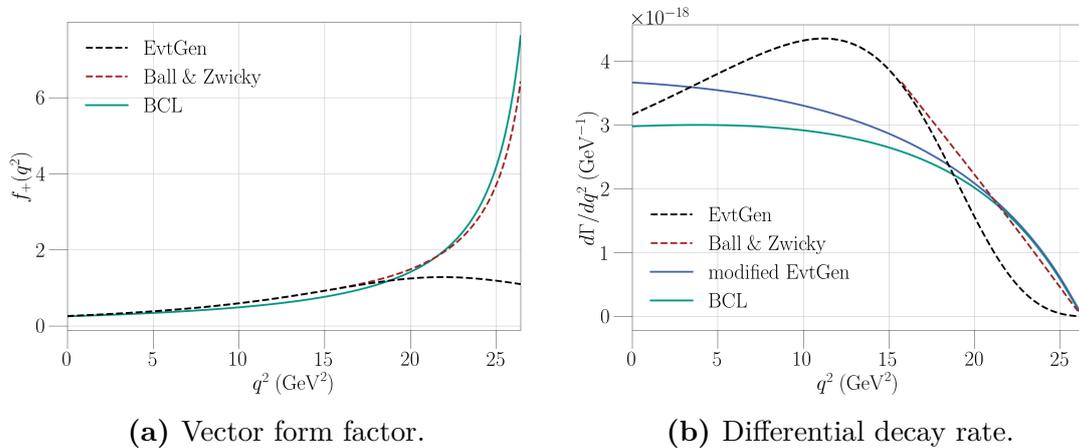


Figure 5.13.: Comparison of different form factor models of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay. The *black line* shows the flawed implementation in `EvtGen`, which was originally used for Belle MC production. The *red line* represents the correct model. The latest BCL prediction is shown as *green line*.

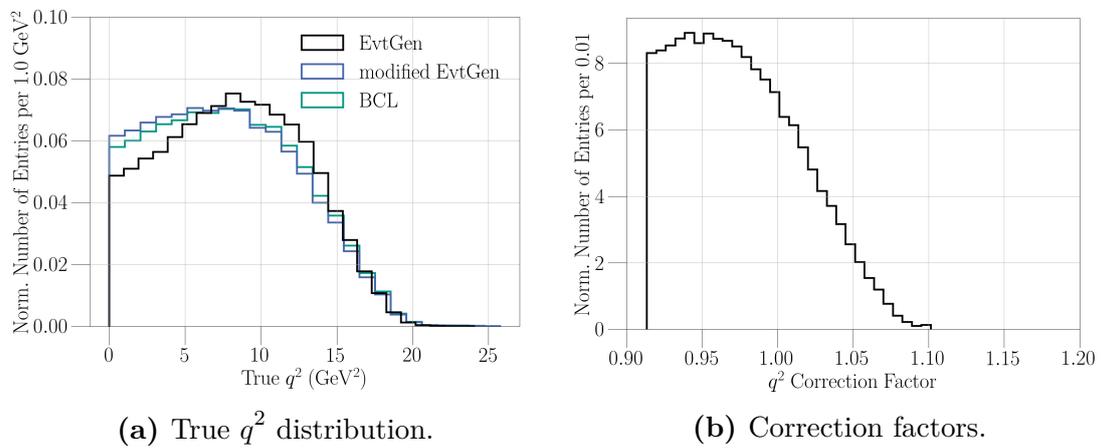


Figure 5.14.: *Left:* The true q^2 distributions of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ MC samples after signal selection. *Right:* The applied correction factors used for reweighting between the modified `EvtGen` and the BCL model.

5.4 Remaining Background Composition

In this section the composition of the remaining background in the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample is presented. The contribution of the different MC samples in the signal window of $M_{\text{miss}}^2 \in (-0.5, 0.5) \text{ GeV}^2$ is shown.

The composition of the electron final state be seen in Table 5.9. The sample dominated by semileptonic $b \rightarrow u$ decays, primarily the peaking backgrounds $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ and $B^+ \rightarrow \eta \ell^+ \nu_\ell$. Decays involving a $b \rightarrow c$ transition, non-resonant continuum decays and rare decays give only a small contribution.

Table 5.10 shows the composition of the muon final state. More decay modes contribute to the background. As for the electron final state, the semileptonic $b \rightarrow u$ decays dominate. In addition, continuum decays contribute 11.4%. The contribution of charged and neutral $b \rightarrow c$ and rare decays is also larger.

The background composition over the full range of $M_{\text{miss}}^2 \in (-1.5, 3.0) \text{ GeV}^2$ can be found in the Appendix D.

Table 5.9.: Background composition of the **electron final state** in the **signal window** of $M_{\text{miss}}^2 \in (-0.5, 0.5) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5% are shown.

MC sample	Decay	Relative fraction (%)
$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$		
65.5 %	$B^+ \rightarrow \pi^0 e^+ \nu_e$	100
$B^+ \rightarrow \eta \ell^+ \nu_\ell$		
17.1 %	$B^+ \rightarrow \eta e^+ \nu_e$	100
$b \rightarrow u \ell^+ \nu_\ell$		
12.1 %	$B^+ \rightarrow \omega \ell^+ \nu_\ell$	49.94
	$B^0 \rightarrow \rho^- \ell^+ \nu_\ell$	21.46
	$B^+ \rightarrow f_2(1270) \ell^+ \nu_\ell$	8.41
	$B^+ \rightarrow X_u^0 \ell^+ \nu_\ell$	7.05
	$B^+ \rightarrow \eta' \ell^+ \nu_\ell$	3.61

Continued on next page

MC sample	Decay	Relative fraction (%)
	$B^0 \rightarrow X_u^- \ell^+ \nu_\ell$	3.56
b \rightarrow c		
3.2 %		
	$B^+ \rightarrow \bar{D}^*(2007)^0 \ell^+ \nu_\ell$	49.89
	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$	40.26
	$B^0 \rightarrow D^- \ell^+ \nu_\ell$	9.85
$e^+e^- \rightarrow q\bar{q}$		
1.7 %	-	-
Rare		
0.5 %		
	$B^+ \rightarrow \pi^+ \pi^0$	25.22
	$B^+ \rightarrow X_{su} \gamma$	24.9
	$B^+ \rightarrow \tau^+ (\rightarrow e^+ \nu_e \bar{\nu}_\tau) \nu_\tau$	12.54
	$B^0 \rightarrow X_{sd} \gamma$	12.47

Table 5.10.: Background composition of the **muon final state** in the **signal window** of $M_{\text{miss}}^2 \in (-0.5, 0.5) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5% are shown.

MC sample	Decay	Relative fraction (%)
$\mathbf{B}^+ \rightarrow \pi^0 \ell^+ \nu_\ell$		
54.8 %	$\mathbf{B}^+ \rightarrow \pi^0 \mu^+ \nu_\mu$	100
$\mathbf{B}^+ \rightarrow \eta \ell^+ \nu_\ell$		
13.7 %	$\mathbf{B}^+ \rightarrow \eta \mu^+ \nu_\mu$	100
$\mathbf{e}^+ \mathbf{e}^- \rightarrow q \bar{q}$		
11.4 %	-	-
$\mathbf{b} \rightarrow \mathbf{u} \ell^+ \nu_\ell$		
8.9 %	$\mathbf{B}^+ \rightarrow \omega \ell^+ \nu_\ell$	44.03
	$\mathbf{B}^0 \rightarrow \rho^- \ell^+ \nu_\ell$	22.21
	$\mathbf{B}^+ \rightarrow X_u^0 \ell^+ \nu_\ell$	8.89
	$\mathbf{B}^+ \rightarrow \rho^0 \ell^+ \nu_\ell$	7.09
	$\mathbf{B}^+ \rightarrow \eta' \ell^+ \nu_\ell$	4.72
	$\mathbf{B}^+ \rightarrow f_2(1270) \ell^+ \nu_\ell$	2.37
	$\mathbf{B}^+ \rightarrow b_1(1235)^0 \ell^+ \nu_\ell$	2.33
	$\mathbf{B}^0 \rightarrow X_u^- \ell^+ \nu_\ell$	2.28
$\mathbf{b} \rightarrow \mathbf{c}$		
5.8 %	$\mathbf{B}^+ \rightarrow \bar{\mathbf{D}}^*(2007)^0 \ell^+ \nu_\ell$	42.47
	$\mathbf{B}^+ \rightarrow \bar{\mathbf{D}}^0 \ell^+ \nu_\ell$	25.48
	$\mathbf{B}^0 \rightarrow \mathbf{D}^- \pi^+$	8.14
	$\mathbf{B}^0 \rightarrow \mathbf{D}^- \ell^+ \nu_\ell$	8.08
	$\mathbf{B}^+ \rightarrow \bar{\mathbf{D}}^0 \pi^+$	7.87

Continued on next page

MC sample	Decay	Relative fraction (%)
	$B^0 \rightarrow D^*(2010)^- \ell^+ \nu_\ell$	4.06
	$B^+ \rightarrow D^*(2007)^0 \pi^+$	3.9
Rare		
5.4 %		
	$B^+ \rightarrow X_{su} \gamma$	17.65
	$B^+ \rightarrow K^+ \pi^0$	12.94
	$B^+ \rightarrow K^*(892)^+ \gamma$	6.23
	$B^+ \rightarrow \eta K^+$	4.81
	$B^+ \rightarrow \eta'(958) K^+$	4.74
	$B^+ \rightarrow K^0 \pi^+ \pi^0$	4.53
	$B^+ \rightarrow K^0 \pi^+ (\gamma)$	4.52
	$B^+ \rightarrow f_2(1270) \pi^+$	3.93
	$B^0 \rightarrow X_{sd} \gamma$	3.17
	$B^+ \rightarrow K^*(892)^0 \pi^+$	2.36
	$B^+ \rightarrow \eta \pi^+$	2.35
	$B^0 \rightarrow K_2^*(1430)^0 \gamma$	2.14
	$B^+ \rightarrow K_0^*(1430)^0 \pi^+$	1.56
	$B^+ \rightarrow K^+ K^0 \pi^0$	1.56
	$B^+ \rightarrow K_2^*(1430)^+ \gamma$	1.55

CHAPTER 6

Signal Extraction

In this chapter the extraction of the partial branching fraction $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ from fitting M_{miss}^2 distributions is described. In Section 6.1 the binned maximum likelihood fit is introduced. As described earlier, the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ background is constrained by fitting its branching fraction simultaneously in the nominal $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample and the dedicated $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample. To be more precise, a simultaneous fit is applied on four samples, the electron and muon final state of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ and $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample. Section 6.2 and Section 6.3 show the used template probability density functions (PDFs) used in the fit. Section 6.4 describes the significance calculation. The fitting procedure is validated with different tests. In Section 6.5 a sample test is presented, where the fit is applied on subsamples of the MC. Toy studies are shown in Section 6.6. Two different linearity tests are conducted in Section 6.7 to check for a possible fit bias.

6.1 The Binned Maximum Likelihood Fit

The partial branching fraction of the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ is extracted with a binned maximum-likelihood fit [23, Chapter 11]. The likelihood is extended by a Poisson probability term to extract the number of events from the fit. To estimate the unknown parameter θ in a binned fit, the number of expected events $\nu_i(\nu_{\text{tot}}, \theta)$ per bin i can be written as

$$\nu_i(\nu_{\text{tot}}, \theta) = \nu_{\text{tot}} \int_{x_i^{\text{low}}}^{x_i^{\text{up}}} \mathcal{P}(x, \theta) dx, \quad (6.1)$$

where ν_{tot} denotes the total number of entries in the binned histogram, x_i^{low} (x_i^{up}) is the lower (upper) bin border and $\mathcal{P}(x, \theta)$ is the PDF containing the Poisson term. For computational reasons the negative log-likelihood is minimized given by

$$\mathcal{L}(\nu_{\text{tot}}, \theta) = -\nu_{\text{tot}} + \sum_{i=1}^N n_i \log \nu_i(\nu_{\text{tot}}, \theta), \quad (6.2)$$

where N is the number of bins.

The branching fraction is calculated using the number of extracted signal events from the fit with

$$\mathcal{B} = \frac{N_{\text{sig},i}}{\epsilon_i \cdot 2 \cdot \mathcal{B}(\Upsilon(4S) \rightarrow B^+ B^-) \cdot N_{B\bar{B}}}, \quad (6.3)$$

where i is the final state, $N_{\text{sig},i}$ is the number of signal events, ϵ_i denotes the reconstruction efficiency and $N_{B\bar{B}}$ is the number of recorded $B\bar{B}$ events in the full data set. The reconstruction efficiency is calculated as the ratio of reconstructed signal events to the number of generated signal events for the final state on MC. For the measurement of the partial branching fraction $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ the reconstruction efficiency is calculated only on signal MC events which fulfill the cut on the signal-side photon energy of $E_\gamma > 1.0$ GeV on generator level.

6.2 The $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ Fit Model

The template PDFs are generated from the MC distributions of the squared missing mass M_{miss}^2 of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample with equally sized bins of 0.15 GeV. A combination of four PDFs for each final state is used:

$B^+ \rightarrow \ell^+ \nu_\ell \gamma$ signal The signal template is generated from a high statistics sample containing 10×10^6 $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ events per final state.

$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ background The template is generated from a high statistics sample containing 5×10^6 $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ MC events per final state.

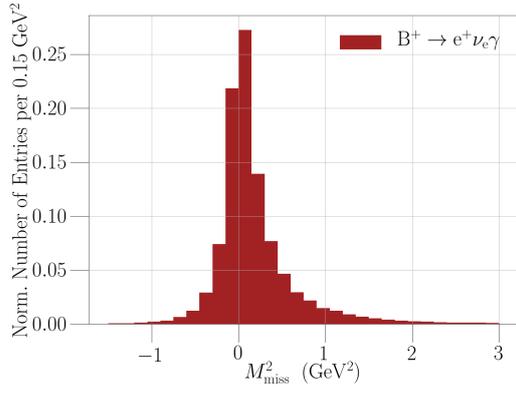
$B \rightarrow X_u \ell^+ \nu_\ell$ background This component contains $B \rightarrow X_u \ell \nu_\ell$ decays, where X_u indicates a light meson ($\eta, \rho^0, \eta', \omega, \pi^+$ or ρ^+). The PDF is generated from the $b \rightarrow u \ell \nu_\ell$ (corresponding to 20 times the luminosity) and the $B^+ \rightarrow \eta \ell^+ \nu_\ell$ MC samples.

Remaining background The remaining background template is generated from nine streams of generic $b \rightarrow c$ decays, five streams of continuum decays, the complete rare MC sample (corresponding to 50 times the luminosity) and the resonant decays of the $b \rightarrow u \ell \nu_\ell$ sample (corresponding to 20 times the luminosity).

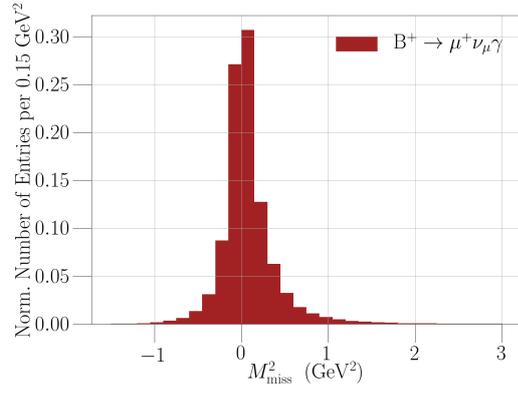
The individual background samples are scaled according to the integrated luminosity of the recorded data. All MC corrections discussed in Section 5.3 are applied. The combined PDF reads as

$$\mathcal{P}(M_{\text{miss}}^2) = \sum_i \left[f_{\text{sig},i} \mathcal{P}_{\text{sig},i}(M_{\text{miss}}^2) + f_{\pi,i} \mathcal{P}_{\pi,i}(M_{\text{miss}}^2) \right. \\ \left. + f_{X_u,i} \mathcal{P}_{X_u,i}(M_{\text{miss}}^2) + f_{\text{rem},i} \mathcal{P}_{\text{rem},i}(M_{\text{miss}}^2) \right], \quad (6.4)$$

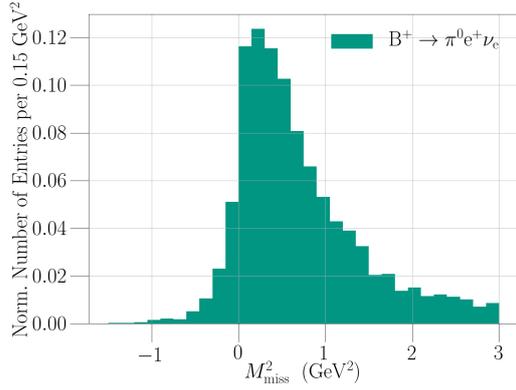
where $\mathcal{P}_{j,i}$ denotes the individual PDF for the specified component j for the final state i . The relative fraction of a component is denoted as $f_{j,i}$. The sum of all fractions $\sum_j f_{j,i}$ per final state is allowed to become smaller or larger than one to allow the fit to adapt to the number of events on data. The background normalizations are allowed to vary in the fit. The signal branching fraction is restricted to positive values by making use of an exponential prior, which penalizes the fit for negative normalizations. The generated templates for both final states can be seen in Figure 6.1.



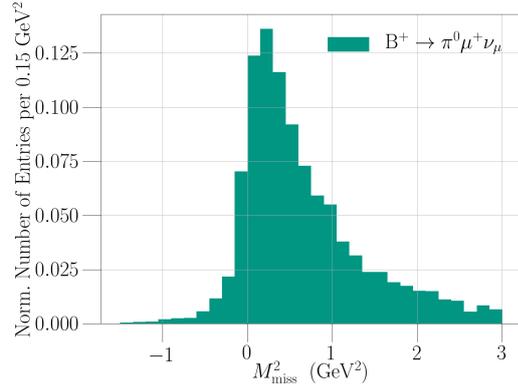
(a) $B^+ \rightarrow e^+ \nu_e \gamma$ signal PDF.



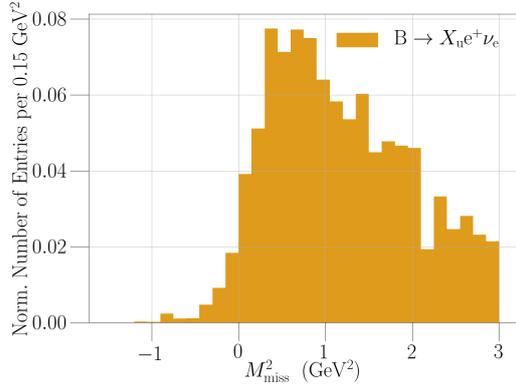
(b) $B^+ \rightarrow \mu^+ \nu_\mu \gamma$ signal PDF.



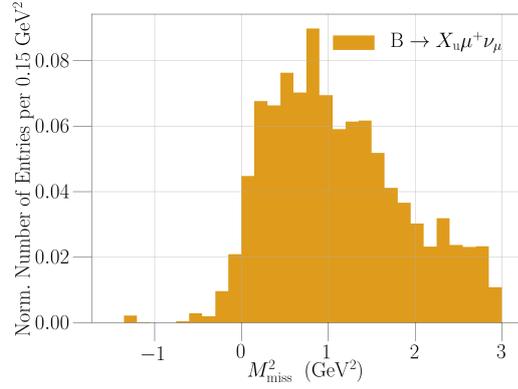
(c) $B^+ \rightarrow \pi^0 e^+ \nu_e$ background PDF.



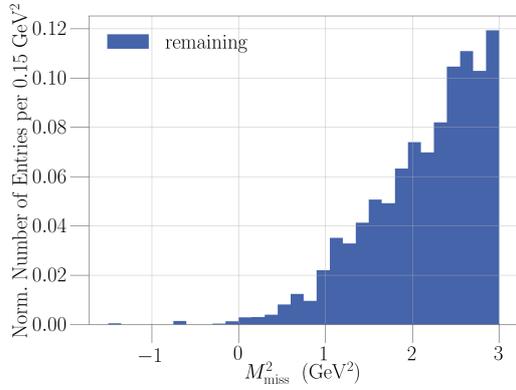
(d) $B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$ background PDF.



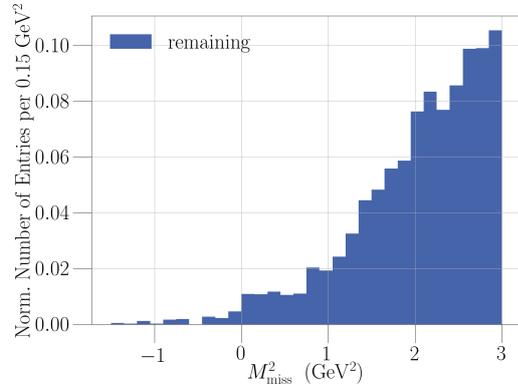
(e) $B \rightarrow X_u e^+ \nu_e$ background PDF.



(f) $B \rightarrow X_u \mu^+ \nu_\mu$ background PDF.



(g) Remaining background PDF.



(h) Remaining background PDF.

Figure 6.1.: Histogram template PDFs used for the signal extraction of the (partial) branching fraction $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ and $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ on the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample. The *left (right) column* shows the electron (muon) final state.

6.3 The $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ Fit Model

The template PDFs are generated from the MC distributions of the squared missing mass M_{miss}^2 of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ selection sample with equal sized bins of 0.15 GeV. A combination of three PDFs for each final state is used:

$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ signal The signal template is generated from a high statistics sample containing 5×10^6 $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ events per final state.

$B \rightarrow X_u \ell^+ \nu_\ell$ background This component contains the $B \rightarrow X_u \ell^+ \nu_\ell$ decays, where X_u indicates a light meson ($\eta, \rho^0, \eta', \omega, \pi^+$ or ρ^+). The sample is generated from the $b \rightarrow u \ell \nu_\ell$ (corresponding to 20 times the luminosity) and $B^+ \rightarrow \eta \ell^+ \nu_\ell$ MC samples.

Remaining background The remaining background template is generated from four streams of generic $b \rightarrow c$ decays, four streams of continuum decays, the complete rare MC sample (corresponding to 50 times the luminosity) and the resonant decays of the $b \rightarrow u \ell \nu_\ell$ sample (corresponding to 20 times the luminosity).

The individual background samples are scaled according to the integrated luminosity of the recorded data. All MC corrections discussed in Section 5.3 are applied. The combined PDF reads as

$$\mathcal{P}(M_{\text{miss}}^2) = \sum_i \left[f_{\pi,i} \mathcal{P}_{\pi,i}(M_{\text{miss}}^2) + f_{X_u,i} \mathcal{P}_{X_u,i}(M_{\text{miss}}^2) + f_{\text{rem},i} \mathcal{P}_{\text{rem},i}(M_{\text{miss}}^2) \right], \quad (6.5)$$

where $\mathcal{P}_{j,i}$ denotes the individual PDF for the specified component j for the final state i . The relative fraction of a component is denoted as $f_{j,i}$. The sum of all fractions $\sum_j f_{j,i}$ per final state is allowed to become smaller or larger than one to allow the fit to adapt to the number of events on data. All normalizations are allowed to vary in the fit. The generated templates for both final states can be seen in Figure 6.2.

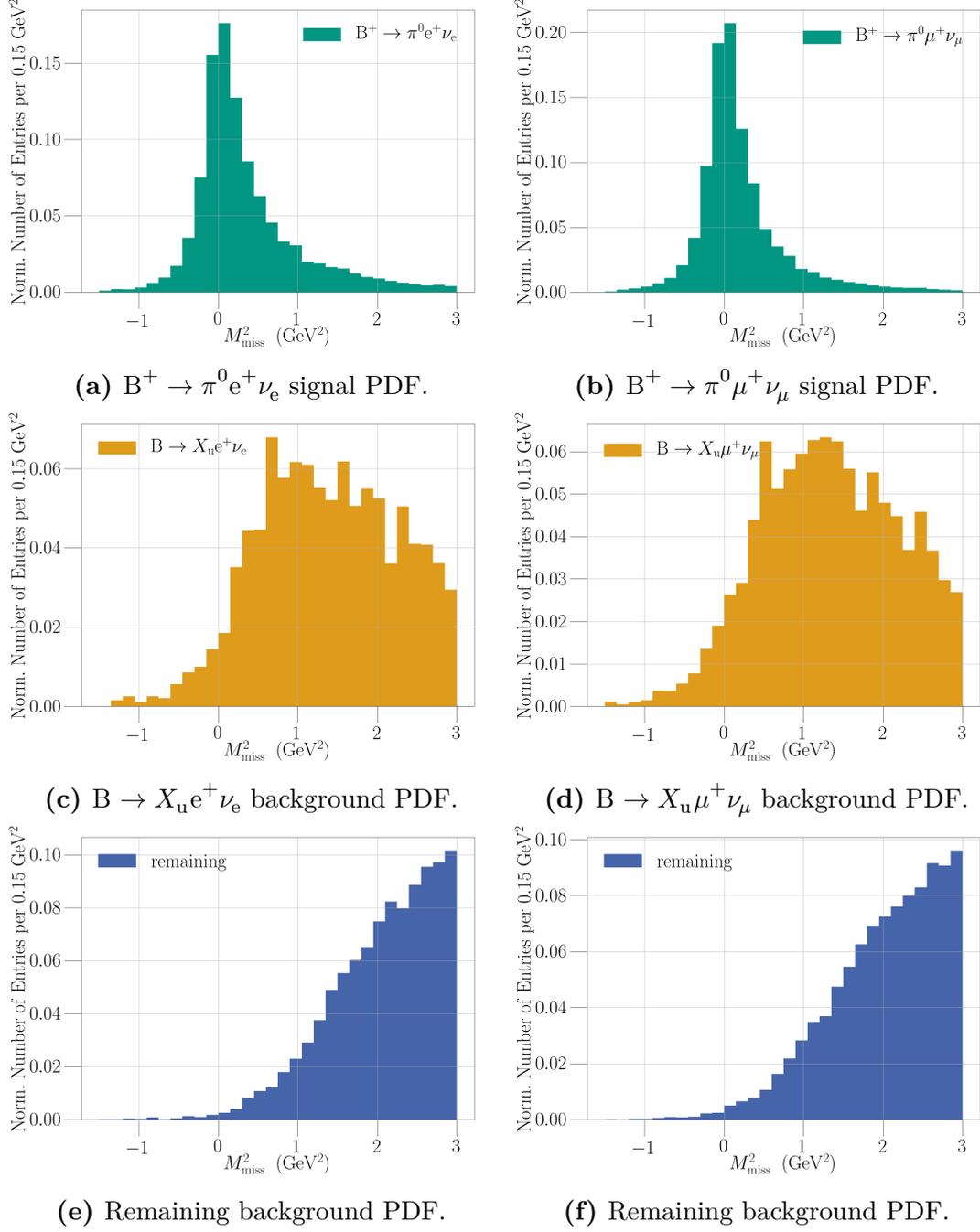


Figure 6.2.: Histogram template PDFs of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample used to constrain the branching fraction measurement of $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$. The left (right) column shows the electron (muon) final state.

6.4 Significance

The statistical uncertainties and the significance of the measurement can be determined by profiling the (log-)likelihood. The parameters of interest (e.g. the branching fraction) are fixed and the likelihood is minimized with respect to the remaining nuisance parameters. The profile has a minimum at the extracted central value. It is more convenient to use the (log-)likelihood ratio

$$\lambda = \log \frac{\mathcal{L}_{\text{S+B}}}{\mathcal{L}_{\text{B}}}, \quad (6.6)$$

where $\mathcal{L}_{\text{S+B}}$ is the likelihood of the signal and background hypothesis and \mathcal{L}_{B} is the likelihood of the background only hypothesis. The significance corresponds to the value of the likelihood ratio, where the parameter of interest is set to zero. By using Wilk's theorem [55], this can be easily transferred into Gaussian standard deviations

$$\Sigma = \sqrt{-2\lambda}. \quad (6.7)$$

The statistical uncertainty is found by scanning the likelihood ratio below and above the central value until the required uncertainty is found. A graphical illustration of this method can be seen in Figure 6.3. The plot shows the likelihood ratio of a fit to MC, which found a global minimum at $\Delta\mathcal{B}(\text{B}^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5.6 \times 10^{-6}$. The lower (upper) statistical uncertainty with respect to the extracted central value is indicated by the dashed gray lines left (right) to the central value. The three lines on each side correspond to the 1σ , 2σ and 3σ uncertainty, respectively.

6.5 Sample Test

The fitting method is validated by a sample test. The available MC of the $\text{B}^+ \rightarrow \ell^+ \nu_\ell \gamma$ and $\text{B}^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ samples are each split randomly into four samples of equal size. The PDF templates are generated on three subsamples and the simultaneous fit is applied on the remaining subsample. Each template and fit subsample is scaled to luminosity. The procedure is iterated such that the fit is performed on each set. The result can be seen in Figure 6.4. For each sample, the input branching fractions of the signal decay and the normalization mode can be extracted within the statistical uncertainty. It has to be noted, that the test suffers under the lower statistics of the background templates. An upper limit has to be determined if the extracted partial branching fraction measurement is not significant. As a test, upper limits on $\Delta\mathcal{B}(\text{B}^+ \rightarrow \ell^+ \nu_\ell \gamma)$ are calculated for the samples under the assumption of Gaussian uncertainties. The results can be found in Table 6.1.

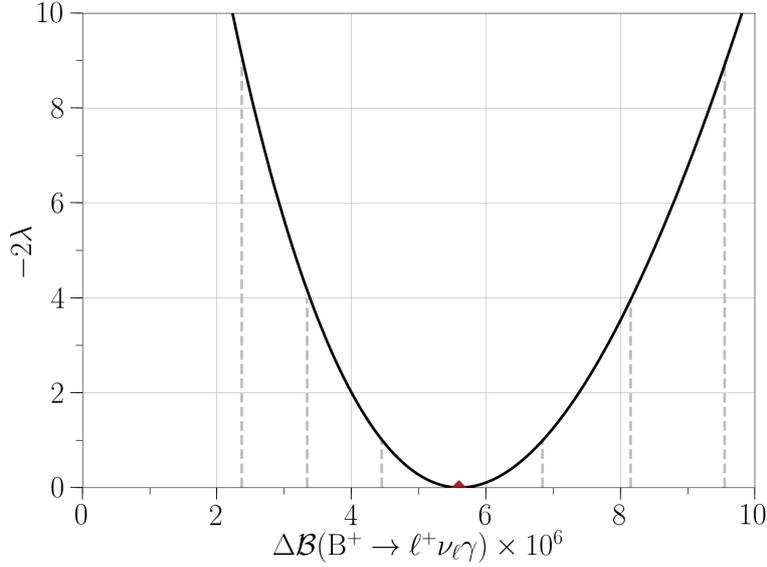


Figure 6.3.: The negative likelihood ratio for Sample 1 of the sample test. The branching fraction was simulated with $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$. A significance of 5.8σ was found. The *gray dashed lines* indicate the lower and upper statistical uncertainties.

Table 6.1.: Estimated upper limits on $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ at 90% C.L. for the sample test. The signal MC is scaled to an expected partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$.

Sample	$\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ limit (10^{-6})
1	< 7.19
2	< 6.33
3	< 6.16
4	< 6.75

6.6 Toy Studies

Toy experiments are performed to check the stability of the fit. The template PDFs are used to draw fake data sets on which the fit can be applied. Repeating this procedure several times allows to test the stability of the fit, e.g. to investigate the impact of the different input branching fractions or to estimate the expected significance for a certain partial branching fraction.

To generate a toy sample the number of expected events n_{exp} from the MC

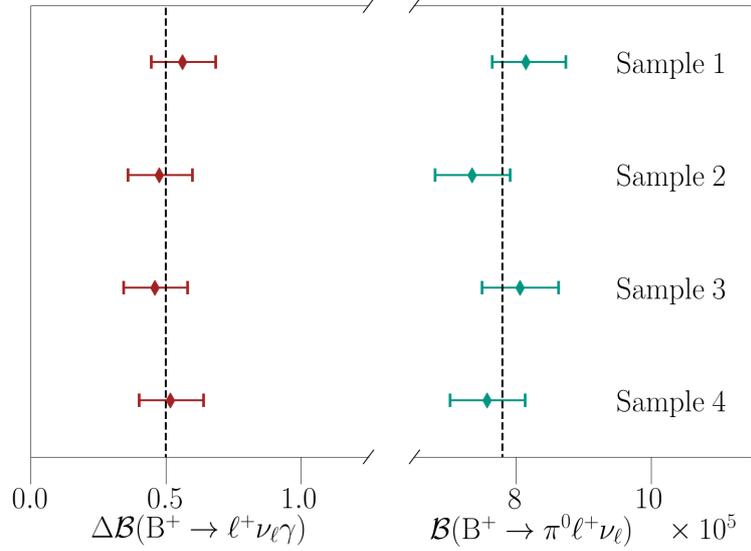


Figure 6.4.: Result of the sample test for the different sets. The available MC is split into four subsamples, whereof three are used for template generation and the remaining one is fitted. The uncertainty is only statistical. Within the uncertainty all results are in agreement with the simulated branching fractions, indicated by the *dashed lines*.

expectation is fluctuated for each PDF by a Poisson distribution as

$$P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{with } \lambda = n_{\text{exp}}, \quad (6.8)$$

where λ denotes the mean. In addition, the number of entries per bin for each PDF template is fluctuated by a Poisson distribution.

Figure 6.5 shows the result of such a study. One thousand toy experiments are generated with a partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma) = 5 \times 10^{-6}$ and $\mathcal{B}(B^+ \rightarrow \pi^0\ell^+\nu_\ell) = 7.8 \times 10^{-5}$. The left plot shows the pull distribution calculated as the difference between the true and the extracted branching fraction divided by the lower (upper) uncertainty for positive (negative) pulls. A shift of the mean from zero would reveal a fit bias; for the given branching fraction no bias is observed. The right plot shows the expected significance. An improvement of 1.9σ compared to the previous Belle measurement [1] is observed for the given branching fraction.

6.7 Linearity Check

A linearity check is performed to check the stability of the fit over a range of possible partial branching fractions. One thousand toy experiments are generated and

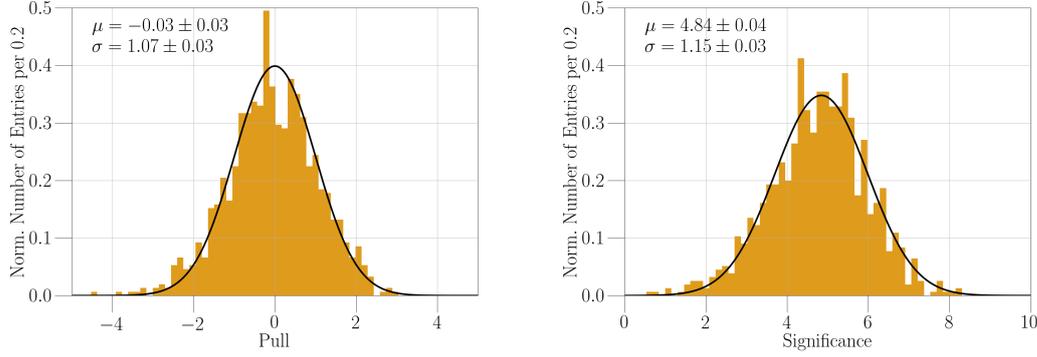


Figure 6.5.: Example of a toy study for a simulated (partial) branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$ and $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) = 7.8 \times 10^{-5}$. One thousand experiments are generated. The *left plot* shows the pull distribution, the *right plot* shows the significance of the extracted $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$. The results of a Gaussian fit to the distributions are shown in the upper left corner.

fitted for 20 different input branching fractions in the range of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) \in (0.5, 10.0) \times 10^{-6}$. The value of $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ is fixed to the current PDG world average. The extracted branching ratios for each sample are fitted with a linear function to check for a possible bias. Lower partial branching ratios are not taken into account, since no significant measurement is feasible in this region. Instead an upper limit would be determined. The result is shown in Figure 6.6. In the range of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) > 1.0 \times 10^{-6}$ the fit is unbiased. Another test is performed to check if the extracted partial branching fraction is correlated to the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ branching fraction. Toy experiments are generated with a fixed branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$ but different branching fractions for $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ in the range of $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) \in (1.8, 12.8) \times 10^{-5}$. The result can be seen in Figure 6.7, no bias is observed.

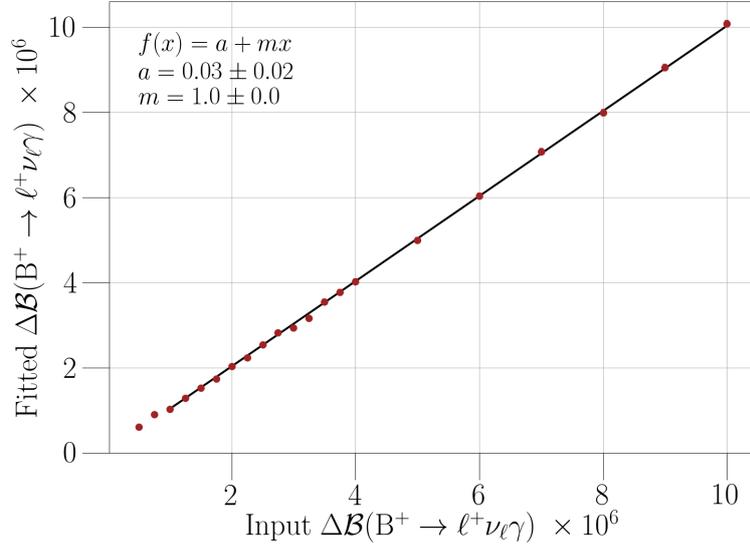


Figure 6.6.: Linearity test for the fitted $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma)$ in the range of $(0.5, 10.0) \times 10^{-6}$. One thousand toy experiments are generated for each point.

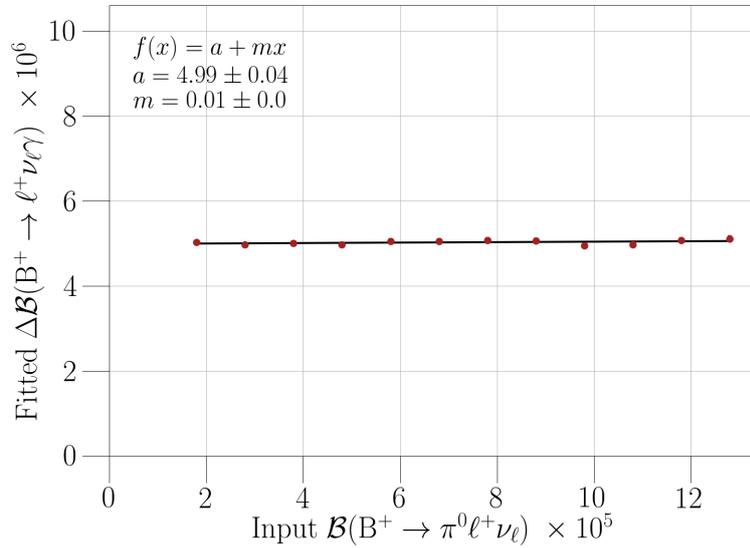


Figure 6.7.: Linearity test to check the correlation between the fitted $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma)$ and the simulated $\mathcal{B}(B^+ \rightarrow \pi^0\ell^+\nu_\ell)$. One thousand toy experiments are generated for each point. The simulated partial branching fraction is fixed to $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma) = 5 \times 10^{-6}$.

CHAPTER 7

Validation Checks

The analysis procedure is developed and optimized on simulated events only. To check the background modeling is therefore crucial. Checks can be made on data on so-called sidebands, where no possible signal is expected. The analysis procedure for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ selection is validated on an off-resonance sample in Section 7.1 and on the M_{bc} sideband in Section 7.2. The fitting procedure is validated on data for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ selection in Section 7.3

7.1 Off-Resonance Data

Besides the on-resonance data set, the Belle experiment recorded an off-resonance sample at an energy 60 MeV below the nominal $\Upsilon(4S)$ resonance. The sample is dedicated to study continuum background, since no $B\bar{B}$ pairs are expected at this energy. The size of 79.4 fb^{-1} corresponds to around 10% of the on-resonance sample.

The data set is used to check the appropriate modeling of the continuum MC simulation after the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ selection by comparing the off-resonance data with the MC expectation. Due to the lower beam energies, variables depending on the center-of-mass energy E_{CMS} are shifted in the off-resonance sample. To address this shortcoming the variables, namely ΔE and M_{bc} , are normalized by the E_{CMS} for the off-resonance sample and corresponding cuts as in the on-resonance sample are applied. Five streams of the continuum MC are used and scaled to the luminosity of the off-resonance sample.

To have sufficient statistics for a comparison of off-resonance data with continuum MC, the cuts on the continuum and peaking background suppression BDTs are

not applied. The off-resonance data sample shows a peaky structure in the fit variable M_{miss}^2 , especially in the signal region. A cut on the tagging probability of $P_{\text{FEI}} > 0.01$ was therefore introduced in the selection process to ensure that this peaking component is suppressed (see also Section 5.2.4). The effect of the cut on the off-resonance data can be seen in Figure 7.1. The number of off-resonance events and the fluctuation between the bins are reduced.

The comparison between the off-resonance data and the continuum MC can be seen in Figure 7.2. No large deviation between the data sample and the MC expectation is observed. This is quantified by a χ^2 -test with equal frequency binning (10 events per bin). The test is applied on the combined final states to have reasonable statistics. A p -value of 0.27 is obtained.

Figure 7.3 shows the comparison with all cuts applied; only in the muon final state one data event can be observed. For both final states the distributions are in agreement with MC expectation, especially in the signal region.

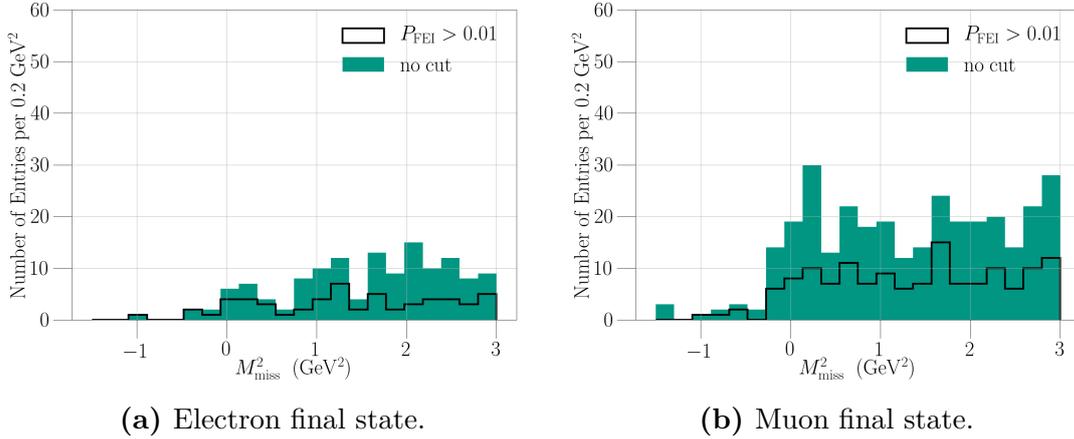


Figure 7.1.: Illustration of the cut on the tagging probability on the off-resonance data. All other cuts are applied except for the cuts on continuum and peaking background suppression.

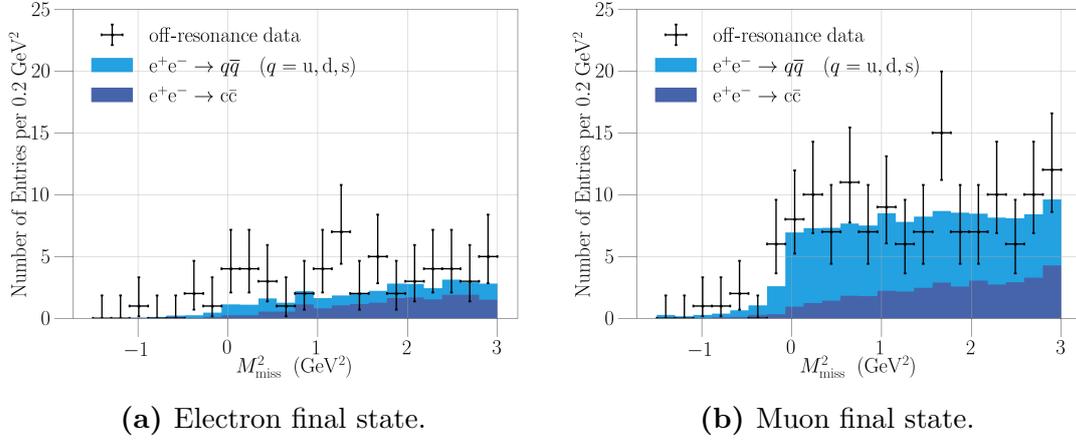


Figure 7.2.: Comparison of off-resonance data with on-resonance continuum MC. All cuts are applied except for the cuts on continuum and peaking background suppression. The on-resonance MC is scaled to the luminosity of the off-resonance sample. Only the statistical uncertainty is shown.

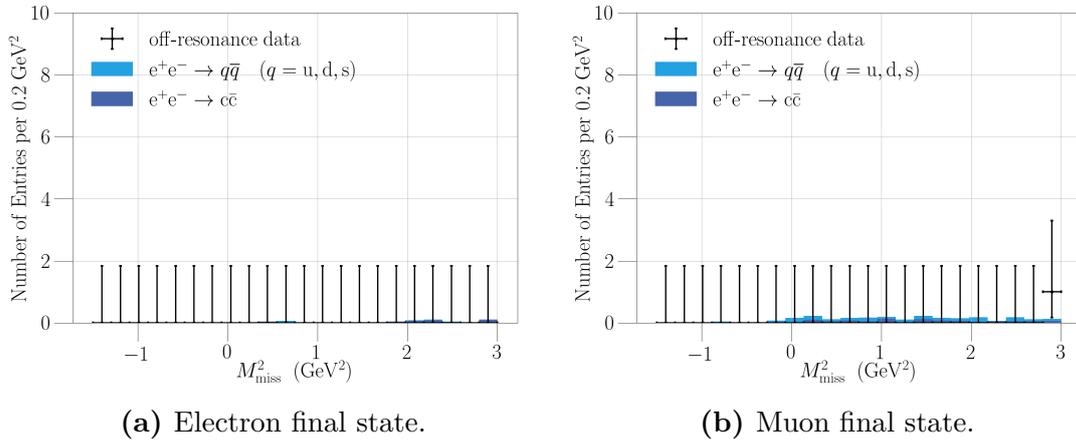


Figure 7.3.: Comparison of off-resonance data with on-resonance continuum MC. All cuts are applied. The on-resonance MC is scaled to the luminosity of the off-resonance sample. Only the statistical uncertainty is shown.

7.2 M_{bc} Sideband

The beam-constrained mass M_{bc} calculated on the hadronically reconstructed B_{tag} is a powerful variable to reject background events. Correctly reconstructed events are expected to peak in the region close to the nominal B mass.

The background modeling of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample can be checked by comparing on-resonance data with the MC expectation in the M_{bc} sideband region of $M_{bc} \in (5.24, 5.27)$ GeV. All remaining selection cuts are applied. A different performance of the tagging algorithm between the signal and the sideband region of M_{bc} is expected, since only wrongly reconstructed B mesons are present in the latter. This issue is addressed by scaling the MC events to the number of data events in the sideband; for the electron (muon) final state a scaling factor of 1.08 (1.23) is found. All other MC corrections are applied except for the tag correction. The resulting distributions of M_{miss}^2 for both final states can be seen in Figure 7.4. The distributions are in a reasonable agreement with the MC expectations. A χ^2 -test with equal frequency binning (10 events per bin) gives a p -value of 0.32 for the electron final state and a p -value of 0.39 for the muon final state.

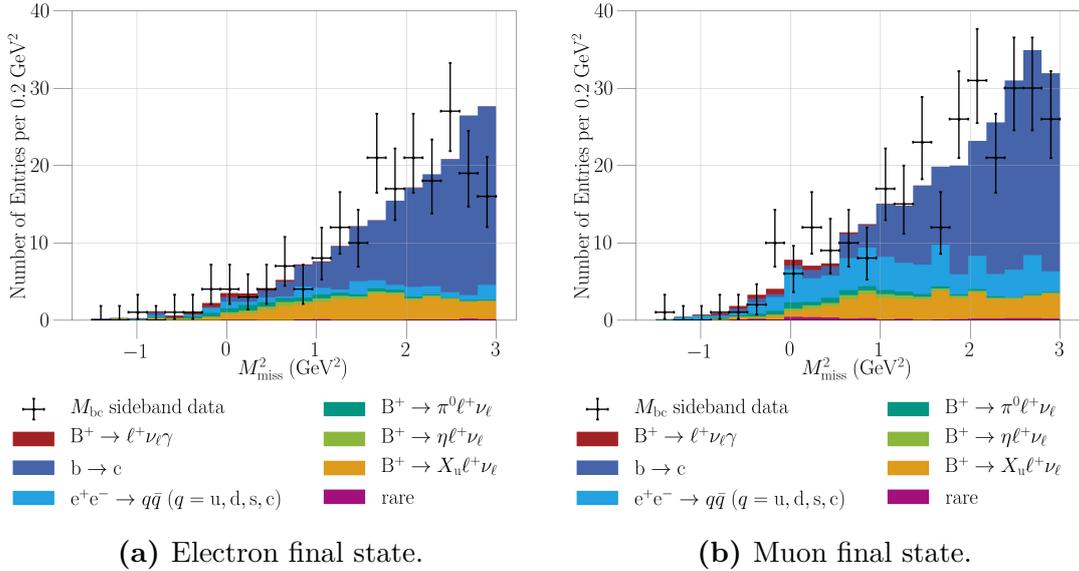


Figure 7.4.: On-resonance data and MC comparison on the M_{bc} sideband. All other selection cuts are applied. Only the statistical uncertainty is shown.

As an additional check, the relative fractions of the reconstructed B_{tag} channels between data and MC are shown in Figure 7.6. A deviation for some of the tag channels with higher relative fraction is observed. This is expected, since the tag correction is only applied globally. The effect of this discrepancy is evaluated by reweighting the events to the fractions of the tag modes on data. Figure 7.5

shows the result of this reweighting; the effect is in general small. A slightly larger deviation can be seen in the higher M_{miss}^2 range, where the MC samples dominate with less events but higher event weights.

Since this deviation is observed on the sideband, it is later treated as a source of uncertainty in the signal region of M_{bc} , see Section 8.2.

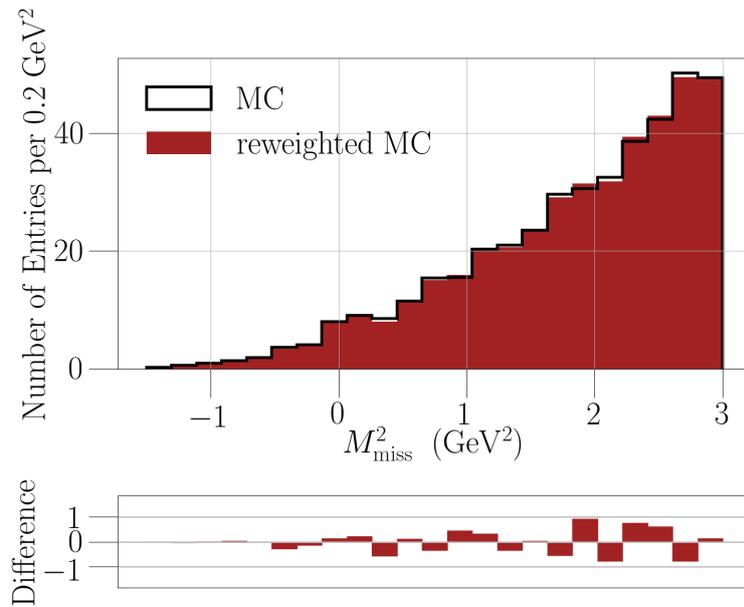


Figure 7.5.: Effect of reweighting the M_{bc} sideband MC to the relative fractions of the tag channels on the M_{bc} sideband data.

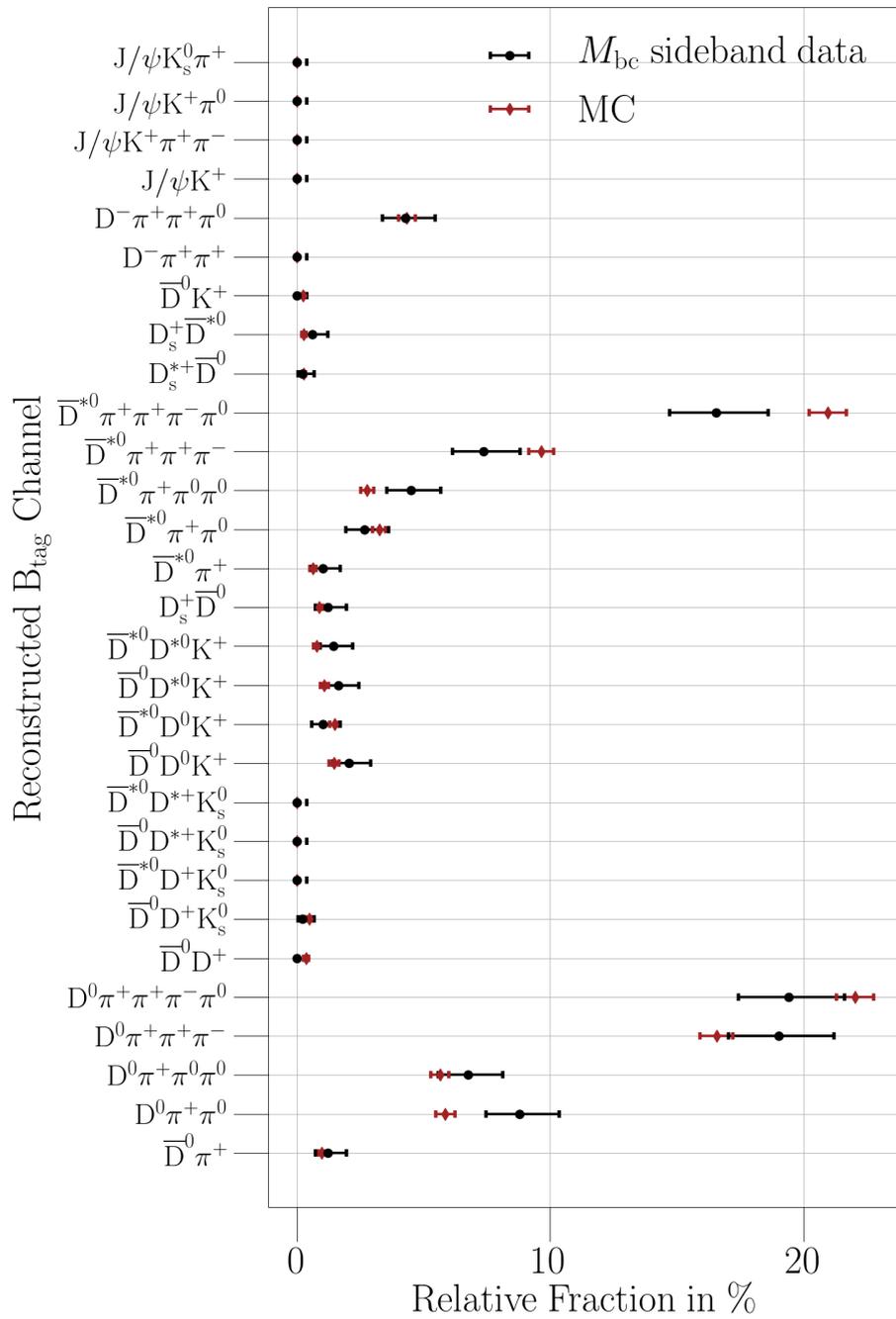


Figure 7.6.: Comparison of the relative fractions of the reconstructed tag channels for the M_{bc} sideband on MC and data. Only the statistical uncertainty is shown.

7.3 $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ Validation

The tagging algorithm and the fitting procedure are validated with a measurement of the branching fraction $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ on the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample. The result should reproduce the current world average of former measurements. The applied selection is discussed in Chapter 5.

The branching fraction is extracted with a simultaneous binned maximum likelihood fit in the electron and muon final state, as described in Section 6.3. The result of the simultaneous fit can be seen in Figure 7.7 and Table 7.1. The uncertainty on the result is only statistical. The branching fraction measurement is in good agreement with the current PDG world average of $\mathcal{B}_{\text{PDG}}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) = (7.8 \pm 0.27) \times 10^{-5}$ [4] and the latest (exclusive) measurement of Belle from 2013 measuring $\mathcal{B}_{\text{excl}}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) = (8.0 \pm 0.8 \pm 0.4) \times 10^{-5}$ [56].

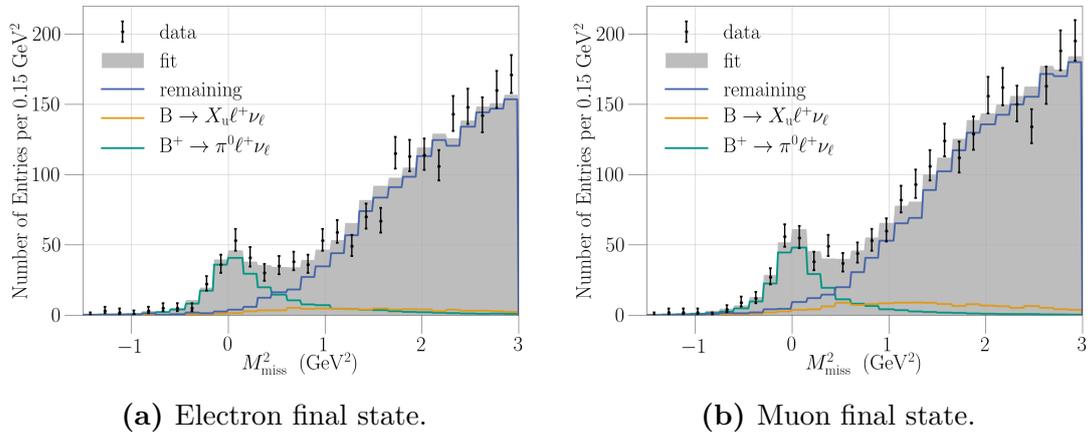


Figure 7.7.: Resulting M_{miss}^2 distributions for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample after the simultaneous fit on data. The *gray component* shows all PDFs stacked.

Table 7.1.: Results of the individual and simultaneous fits for the measurement of $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ on data. Only the statistical uncertainties are given.

Final state	$\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) (10^{-5})$
Electron	$7.9_{-0.9}^{+1.0}$
Muon	$7.5_{-0.8}^{+0.8}$
Both	$7.8_{-0.6}^{+0.6}$

This chapter presents the results of the analysis on data. In the first section the results of the fit are shown. The systematic uncertainties are thoroughly discussed in Section 8.2 and in Section 8.3 the extraction of λ_B and $|V_{ub}|$ is presented. A comparison of the result with the previous measurements can be found in Section 8.4. The Belle II experiment will soon start to take first physics data. Hence, in the last section the possible impact of a larger data sample on the analysis is discussed.

8.1 Measurement on Data

The results of the fit on data for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ and $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ samples are shown in Figures 8.1 and 8.2. As described in Chapter 6, the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample is used to constrain the background in the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample. This background mimics the signal decay in the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample since it has similar shape in the M_{miss}^2 distribution.

The extracted signal yields, branching fractions and significance can be found in Table 8.1. The table contains the fit to only the individual final states and the combined final states. The given significance includes the systematic uncertainty by convolving the likelihood with a Gaussian of mean zero and width of the systematic uncertainty. The full systematic uncertainty is taken into account. The convolution does broaden the likelihood and hence lowers the significance. Similar to the validation check in Section 7.3 the result for the $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ is in good agreement with the current PDG world average of $\mathcal{B}_{\text{PDG}}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) = (7.8 \pm 0.27) \times 10^{-5}$ [4]. Due to the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample, the branching ratio

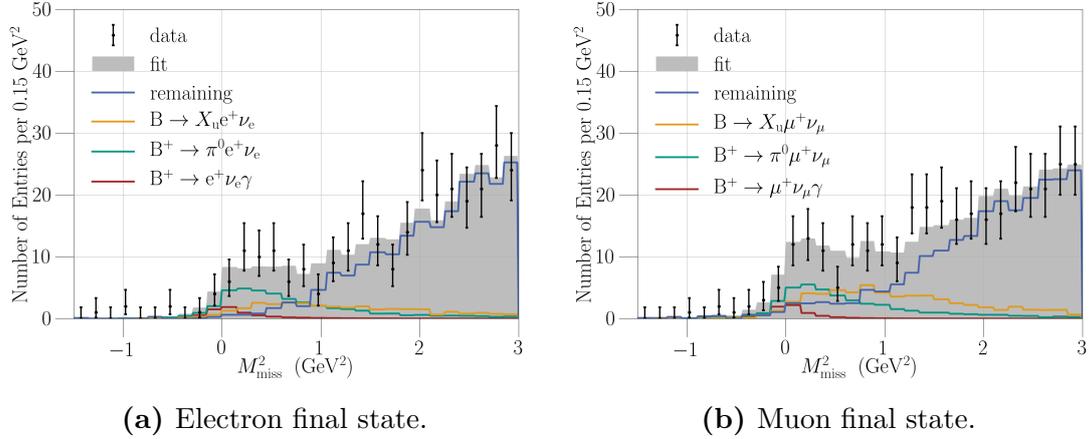


Figure 8.1.: Resulting M_{miss}^2 distributions for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample after the fit on data. The *gray component* shows all PDFs stacked.

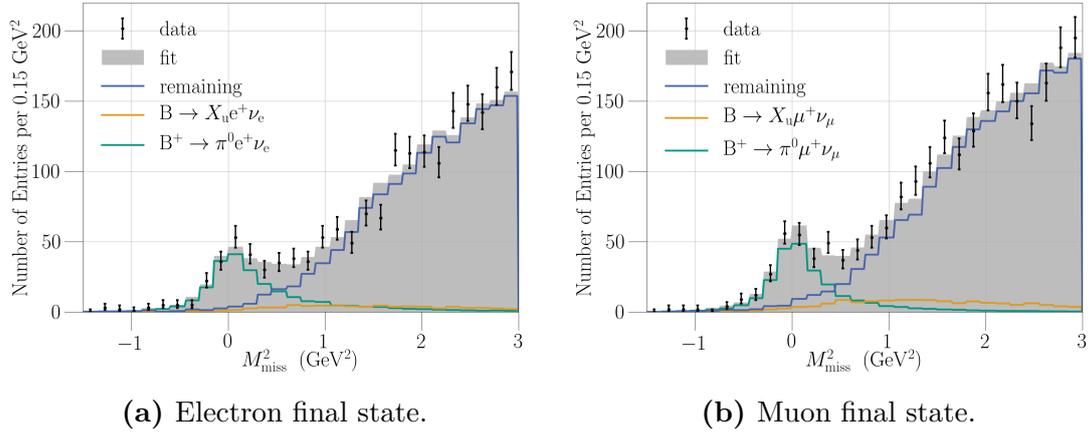


Figure 8.2.: Resulting M_{miss}^2 distributions for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample after the fit on data. The *gray component* shows all PDFs stacked. The samples are used to constrain the $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ during the fit to the nominal sample.

of $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ can be significantly measured in each final state. However, a significant measurement of the partial branching fraction of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ is not possible. The combined measurement including both final states shows the smallest statistical uncertainty and the highest significance. This is expected since the fit is more stable due to the increased statistics.

Table 8.1.: Results of the fit on data. The first uncertainty is statistical, the second is systematic. The significance does include the full systematic uncertainty. The yields correspond to the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample.

Final state	$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$			$B^+ \rightarrow \ell^+ \nu_\ell \gamma$		
	Yield	$\mathcal{B} (10^{-5})$	σ	Yield	$\Delta\mathcal{B} (10^{-6})$	σ
Electron	$41.5^{+4.3}_{-4.1} \pm 4.3$	$8.4^{+0.9}_{-0.8} \pm 0.9$	7.2	$8.5^{+7.9}_{-6.8} \pm 3.2$	$1.7^{+1.6}_{-1.4} \pm 0.7$	1.1
Muon	$38.4^{+4.1}_{-4.0} \pm 3.1$	$7.5^{+0.8}_{-0.8} \pm 0.6$	8.0	$5.3^{+7.5}_{-5.3} \pm 2.2$	$1.0^{+1.4}_{-1.0} \pm 0.4$	0.8
Both	$79.8^{+4.2}_{-4.0} \pm 4.4$	$7.9^{+0.6}_{-0.6} \pm 0.6$	9.7	$13.9^{+7.5}_{-6.9} \pm 2.6$	$1.4^{+1.0}_{-1.0} \pm 0.4$	1.3

Since no significant signal is found for the partial branching fraction, a Bayesian upper limit is determined. The likelihood is therefore integrated in the physical region up to the 90% quantile

$$0.9 = \frac{\int_0^{\Delta\mathcal{B}_{\text{limit}}} \mathcal{L}_{\text{PDF}}(\Delta\mathcal{B}) d\Delta\mathcal{B}}{\int_0^\infty \mathcal{L}_{\text{PDF}}(\Delta\mathcal{B}) d\Delta\mathcal{B}}. \quad (8.1)$$

As for the significance, the full systematic uncertainty is convolved with the likelihood. Figure 8.3 shows a graphical illustration of the upper limit, where the red area corresponds to the remaining 10% quantile.

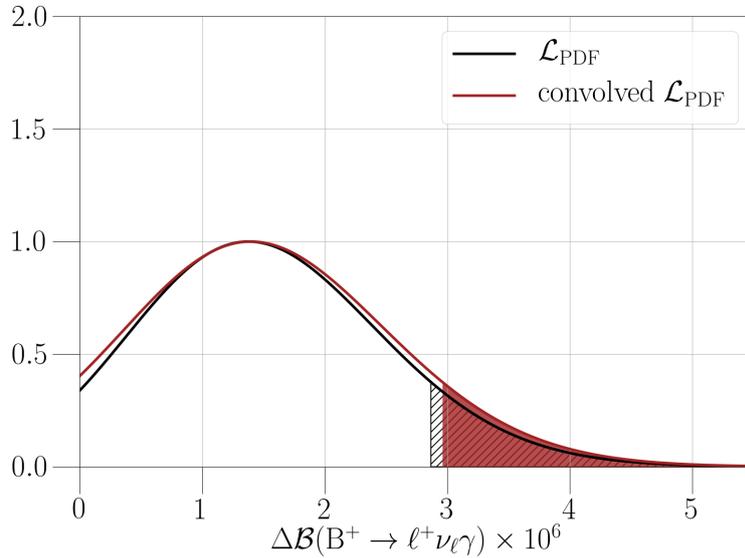


Figure 8.3.: Determination of the upper limit. The marked areas correspond to the remaining 10% quantile of the integrated likelihood PDFs for the combined measurement of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$. The red curve does include the systematic uncertainties.

Table 8.2.: Determined upper limit of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ at 90% C.L.

Final state	$\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ limit (10^{-6})	
	Systematic uncertainty	
	Without	Included
Electron	< 4.1	< 4.3
Muon	< 3.3	< 3.4
Both	< 2.9	< 3.0

Table 8.3.: Events expected on MC and observed on data for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ and the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample.

Final state	$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample		$B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample	
	Expected	Observed	Expected	Observed
Electron	1599 ± 16	1825^{+44}_{-43}	264 ± 4	294^{+18}_{-17}
Muon	1907 ± 17	2244^{+48}_{-47}	301 ± 4	334^{+19}_{-18}

The resulting upper limits with and without systematic uncertainty are summarized in Table 8.2. For the combined final states an upper limit of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) < 3.0 \times 10^{-6}$ is observed at 90% C.L. level.

The number of expected events on MC and the number of observed events can be found in Table 8.3. The uncertainties on the MC events are calculated on weighted events. For both samples a small excess of events is observed, where the normalization of the remaining background is not correctly described by the MC. The fit can adapt for this by up-scaling the background template.

Additionally, a two-dimensional likelihood scan is performed. At each point of the scan, the values of $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ and $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ are fixed and the likelihood is minimized with respect to the remaining nuisance parameters. The resulting values of the likelihood are plotted in Figure 8.4. The central values of the fit are indicated by the black dot, the 39.3% and 68.3% C.L. are drawn as ellipses, where only the statistical uncertainties are included. The small tilt of the ellipses reveals a correlation of $\rho = -0.17$ between both measurements, which is also confirmed by calculating the Hessian matrix. The found upper limit for $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ is drawn as red line.

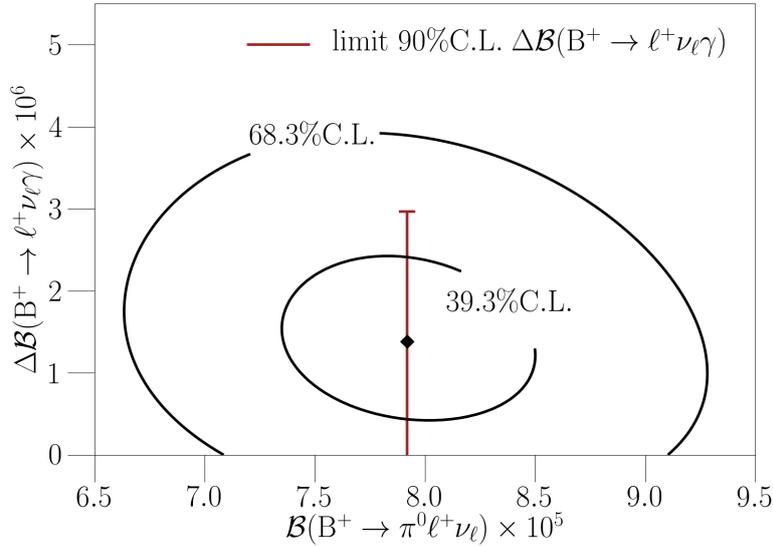


Figure 8.4.: Two dimensional likelihood scan of the fit on data for the combined measurement. The ellipses correspond to the given confidence level, including only the systematic uncertainty. The *red line* indicates the upper limit for $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ including the systematic uncertainty.

8.2 Systematics

Systematic uncertainties reflect the incomplete knowledge on assumptions made during the analysis process. Different sources of uncertainties have to be considered. The multiplicative and additive uncertainties are described in the following subsections. The individual contributions are combined to the total systematic uncertainty by adding them in quadrature. All considered uncertainties are summarized in Tables 8.4 to 8.6.

Number of $B\bar{B}$ Pairs

The number of $B\bar{B}$ pairs in the on-resonance data sample has been measured as $N_{B\bar{B}} = (772 \pm 10) \times 10^6$ [57]. This results in a systematic uncertainty of **1.37%** on the (partial) branching fraction.

LID Corrections

In Section 5.3.2 a correction on MC events for the LID efficiency was introduced. Besides the event-by-event efficiency correction, the authors of Reference [50] provide a method to calculate the global LID efficiency correction and the corresponding uncertainty.

Table 8.4.: Additive and multiplicative systematic uncertainties for the electron final state.

Source	$\mathcal{B}(B^+ \rightarrow \pi^0 e^+ \nu_e) \left(10^{-5}\right)$	$\Delta\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) \left(10^{-6}\right)$
Multiplicative		
Number of $B\bar{B}$ pairs	± 0.11	± 0.02
LID efficiency correction	± 0.16	± 0.03
Track reconstruction efficiency	± 0.03	± 0.01
Tag efficiency correction	± 0.52	± 0.11
Combined	± 0.55	± 0.11
Additive		
Tag mode	± 0.02	± 0.21
Peaking background BDT	± 0.03	± 0.32
PDF templates	± 0.61	± 0.51
$B \rightarrow X_u \ell^+ \nu_\ell$ branching ratios	± 0.01	± 0.02
Signal reconstruction efficiency	± 0.15	± 0.01
$B^+ \rightarrow \ell^+ \nu_\ell \gamma$ model	± 0.00	± 0.04
$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ model	± 0.26	± 0.01
Combined	± 0.68	± 0.64
Total	± 0.87	± 0.65

This approach is followed and a systematic uncertainty of **1.81%** (**1.97%**) is assigned for the combined final states of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ ($B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$) sample.

Track Reconstruction Efficiency

The track reconstruction efficiency and its systematic uncertainty was estimated in Reference [58] by making use of the decay process

$$\begin{aligned}
 D^* &\rightarrow \pi_S D^0 \\
 &\hookrightarrow \pi \pi K_S^0 \\
 &\hookrightarrow \pi^+ \pi^-.
 \end{aligned}
 \tag{8.2}$$

A full and partial reconstruction of the D^* meson was used to investigate the tracking performance. The latter reconstruction approach allowed for one missing charged pion track in the K_S^0 reconstruction. The missing track was extracted from kinematic constraints of the D^* , D^0 and K_S^0 meson. The track finding efficiency was then measured as the ratio of fully and partially reconstructed D^* mesons. A comparison of the efficiency on MC and data allowed to estimate the uncertainty.

Table 8.5.: Additive and multiplicative systematic uncertainties for the muon final state.

Source	$\mathcal{B}(B^+ \rightarrow \pi^0 \mu^+ \nu_\mu) (10^{-5})$	$\Delta\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) (10^{-6})$
Multiplicative		
Number of $B\bar{B}$ pairs	± 0.10	± 0.01
LID efficiency correction	± 0.15	± 0.02
Track reconstruction efficiency	± 0.03	± 0.00
Tag efficiency correction	± 0.46	± 0.06
Combined	± 0.50	± 0.07
Additive		
Tag mode	± 0.04	± 0.06
Peaking background BDT	± 0.02	± 0.25
PDF templates	± 0.23	± 0.33
$B \rightarrow X_u \ell^+ \nu_\ell$ branching ratios	± 0.02	± 0.08
Signal reconstruction efficiency	± 0.13	± 0.00
$B^+ \rightarrow \ell^+ \nu_\ell \gamma$ model	± 0.00	± 0.01
$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ model	± 0.24	± 0.00
Combined	± 0.36	± 0.43
Total	± 0.61	± 0.43

Based on this study the recommended systematic uncertainty of **0.34%** is assigned for the track reconstruction efficiency on the single lepton track on the signal-side.

Tag Efficiency Correction

In Section 5.3.1 a global correction factor for the tagging efficiency is extracted. Each tagged MC event is weighted with a calibration factor to incorporate the discrepancy of tagging power between MC and data. To estimate the total systematic uncertainty on the tag correction the statistical and systematic uncertainties from the calibration are combined. A systematic uncertainty on the tag correction of **6.18%** is assigned.

Relative Fraction of Tag Channels

As shown in Section 7.2, the relative frequencies of reconstructed tag channels are different between MC and data on the M_{bc} sideband.

The relative frequency of the reconstructed tag channels for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample in the M_{bc} signal window is shown in Figure 8.5. The relative fractions agree within the uncertainties for most channels. Nevertheless, this is considered as a

Table 8.6.: Additive and multiplicative systematic uncertainties for the simultaneous fit in both final states.

Source	$\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) \left(10^{-5}\right)$	$\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) \left(10^{-6}\right)$
Multiplicative		
Number of $B\bar{B}$ pairs	± 0.11	± 0.02
LID efficiency correction	± 0.16	± 0.02
Track reconstruction efficiency	± 0.03	± 0.0
Tag efficiency correction	± 0.49	± 0.09
Combined	± 0.52	± 0.09
Additive		
Tag mode	± 0.01	± 0.14
Peaking background BDT	± 0.02	± 0.24
PDF templates	± 0.08	± 0.18
$B \rightarrow X_u \ell^+ \nu_\ell$ branching ratios	± 0.02	± 0.07
Signal reconstruction efficiency	± 0.20	± 0.01
$B^+ \rightarrow \ell^+ \nu_\ell \gamma$ model	± 0.0	± 0.03
$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ model	± 0.25	± 0.01
Combined	± 0.33	± 0.34
Total	± 0.62	± 0.36

systematic uncertainty. The MC events are reweighted accordingly to the fractions on data. This leads to slightly changed M_{miss}^2 distributions of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample, as can be seen in Figure 8.6. Afterwards the modified distributions are used to generate PDF templates and the fit on data is repeated. The difference to the nominal result is taken as systematic uncertainty.

Peaking Background BDT

To suppress the peaking background in the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample a BDT is trained and applied (see Section 5.2.6). A systematic uncertainty for the application of the classifier is evaluated by reweighting the MC samples according to the distributions of the input variables on data. Afterwards the fit on data is repeated with modified PDF templates and reconstruction efficiencies. The procedure is repeated for all input variables of the classifier. The largest deviation to the nominal result is taken as estimate for the systematic uncertainty.

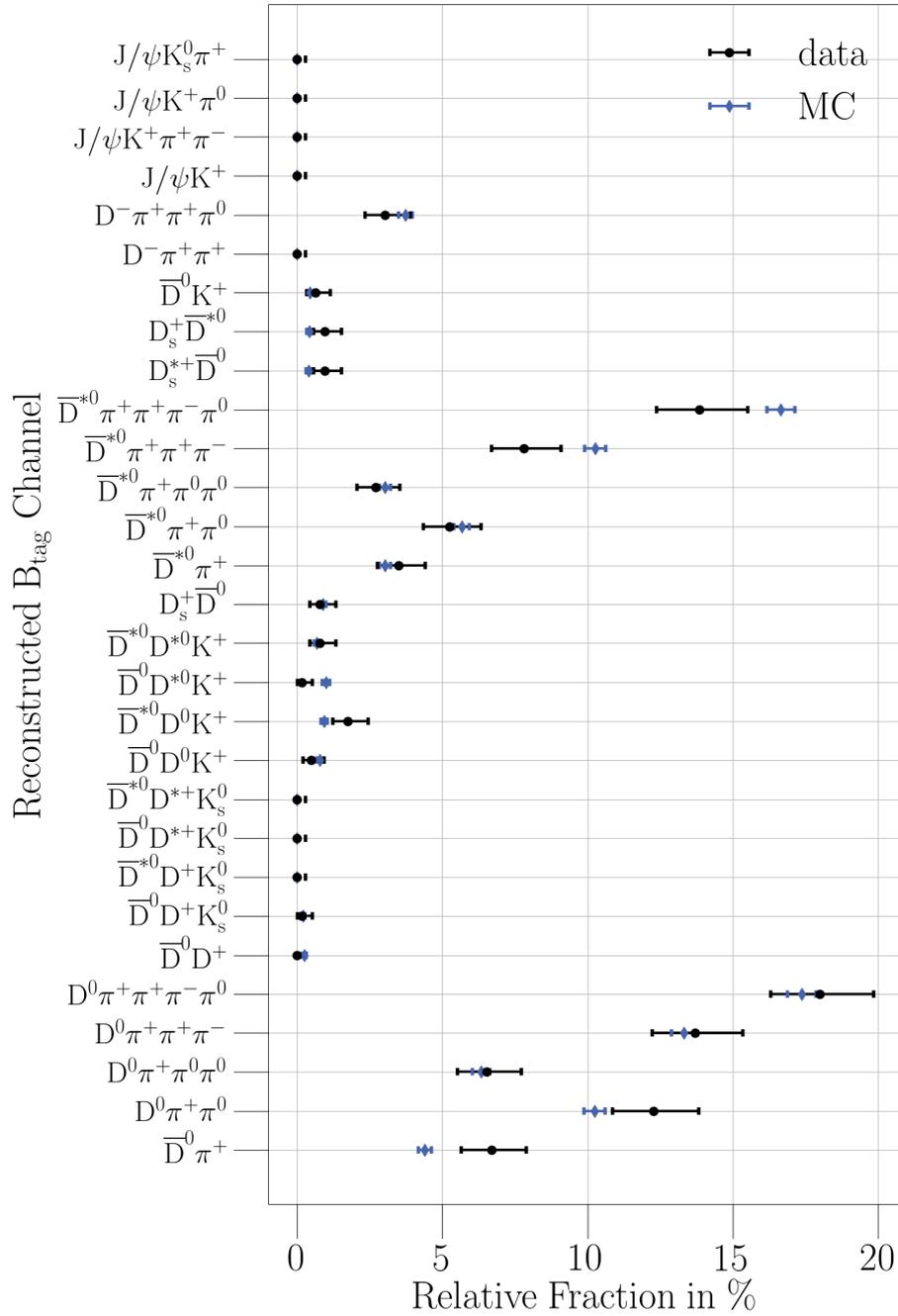


Figure 8.5.: Comparison of the relative fractions of the reconstructed tag channels on MC and data. Only the statistical uncertainty is shown.

PDF Templates

The PDFs used for the signal extraction are drawn from MC samples with limited statistics. The MC samples are varied in each bin by fluctuating the number of

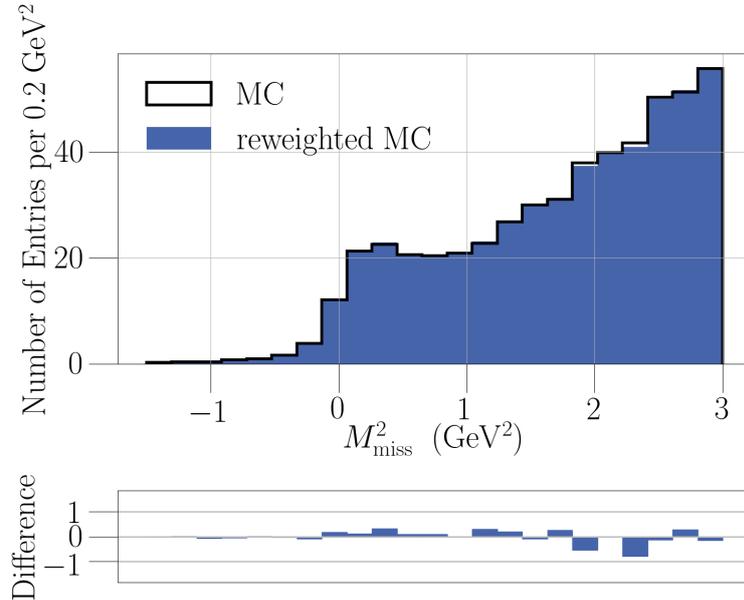


Figure 8.6.: Effect of reweighting the MC samples to the relative fractions of the tag channels on data.

entries per bin by a Poisson distribution as given in Equation (6.8). The modified distributions are then used to generate templates, where one MC type in all samples is varied at once. Afterwards the fit on data is repeated. The procedure is repeated a thousand times for each MC type. The standard deviation of the results is taken as systematic uncertainty. The individual uncertainties for each MC type are added in quadrature.

$B \rightarrow X_u \ell^+ \nu_\ell$ Branching Ratios

In Section 5.3.3 the branching ratios of the $b \rightarrow u \ell \nu_\ell$ MC sample are corrected to the current world averages from PDG. The individual fractions of these modes in the PDF template of the $B \rightarrow X_u \ell^+ \nu_\ell$ background are fixed in the fit on data. This is considered as a systematic uncertainty. The branching ratios are varied by the combined 1σ uncertainty from PDG. Afterwards the fit on data is repeated. The difference to the nominal result is taken as systematic uncertainty for the individual mode, where the average over the up- and downward fluctuation is used. The individual modes are treated independently and eventually combined by adding them in quadrature.

Signal Reconstruction Efficiency

The reconstruction efficiency enters the calculation of the branching fraction and is determined on the signal MC sample. To estimate the uncertainty of the efficiencies binomial uncertainties are assumed [59]. The uncertainty for the true efficiency is given by

$$\sigma_\epsilon = \sqrt{\text{var}(N_{\text{sel}})} \quad (8.3)$$

$$= \sqrt{\epsilon(1-\epsilon)N_{\text{gen}}}. \quad (8.4)$$

The true efficiency is unknown, but its uncertainty can be estimated using the estimated efficiency $\hat{\epsilon} = N_{\text{sel}}/N_{\text{gen}}$ leading to

$$\sigma_{\hat{\epsilon}} = \frac{1}{N_{\text{gen}}} \sqrt{N_{\text{sel}} \left(1 - \frac{N_{\text{sel}}}{N_{\text{gen}}}\right)}. \quad (8.5)$$

The $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ EvtGen Model

The $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decays are simulated with the **LNUGAMMA EvtGen** model, which is based on Reference [60]. A straightforward implementation of the axial and vector form factor is used. The modeling of the form factors can have a decisive impact on the shape of the photon energy distribution and hence on the signal reconstruction efficiency. As a cross-check and to estimate the systematic uncertainty from the used **EvtGen** model, a model with updated form factors based on Reference [6] was implemented. The systematic uncertainty is estimated by reweighting the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ MC samples to the new model¹. Therefore, two samples with 2.5×10^6 events per final state are produced with the old and newly implemented **EvtGen** model. The weights are calculated according to the photon energy on generator level. The true photon distributions for both models and the applied weights can be seen in Figure 8.7. A small shift in the distributions is observed around the 1 GeV cut. The fit on data is repeated and the difference to the nominal result is taken as systematic uncertainty.

The $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ EvtGen Model

The $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ MC sample was produced with a modified q^2 distribution, which is weighted to the latest BCL model prediction (see Section 5.3.4). The form factors have a possible influence on the reconstruction efficiency of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay and the shape of the M_{miss}^2 distribution. The systematic uncertainty from the

¹Due to technical reasons it is not possible to directly produce signal MC with the new model within the Belle AnalysisS Framework.

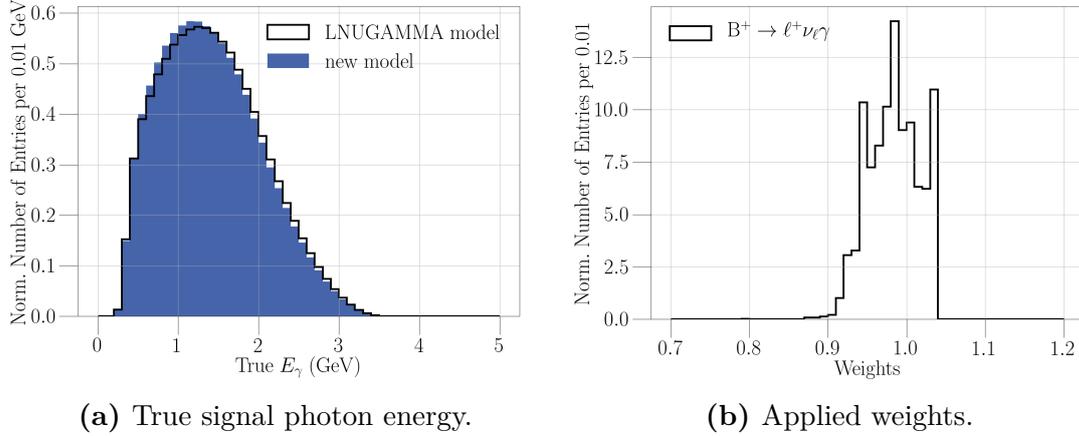


Figure 8.7.: Distributions for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ *EvtGen* model systematic.

BCL model is evaluated by reweighting the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ MC decays, where the individual event weights are varied by their 1σ uncertainty. The fit is repeated and the averaged difference to the nominal fit result is taken as systematic uncertainty.

8.3 Extraction of λ_B and V_{ub}

As introduced in Section 2.2 the axial and vector form factor of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay depend on the first inverse momentum λ_B of the LCDA of the B meson, a parameter which is hard to predict by theoretical calculations. The measurement of the partial branching fraction provides a way to extract this parameter experimentally. In the following two approaches are presented for the extraction of λ_B . In Equation (2.4) it is shown that the branching ratio of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ depends on the CKM matrix element $|V_{ub}|$. The dependence of the partial branching fraction for different values of $|V_{ub}|$ from exclusive and inclusive measurements can be seen in Figure 8.8. The expressions in Equation (2.4) are numerically integrated and solved for λ_B to extract a lower limit from the measured upper limit of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$. A detailed description of those expressions and the used parameters can be found in Reference [6]. To take the theoretical uncertainties into account, the parameters are varied by the proposed uncertainty range in Reference [6]. The individual uncertainties are assumed to be independent and are added in quadrature. This theoretical uncertainty is convolved with the likelihood to obtain a new upper limit, as described before. The obtained limit is higher than the previous one, which includes only the statistical and systematic uncertainties and is used to set the lower limit on λ_B . The results for different values of $|V_{ub}|$ can be found in Table 8.7.

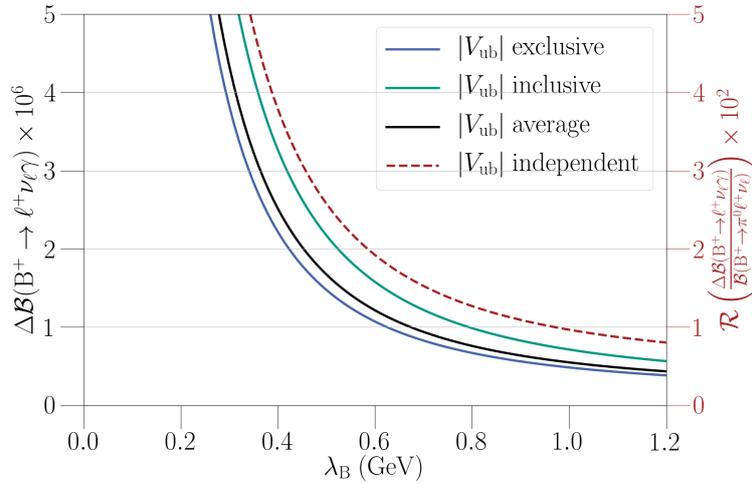


Figure 8.8.: Dependence of λ_B on the measured quantity. The *continuous lines* correspond to the $|V_{ub}|$ dependent method (*left y-axis*). The different values for $|V_{ub}|$ can be found in Table 8.7. The *red dashed line* represents the $|V_{ub}|$ independent method (*right y-axis*).

The second method allows for a $|V_{ub}|$ independent determination of λ_B . Besides the partial branching fraction of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay, the branching fraction of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay is also measured. As shown in Section 2.3, the latter also depends on the CKM matrix element $|V_{ub}|$. By constructing the ratio, the CKM matrix element drops out of the calculation and the equation can be solved for λ_B . The dependence of λ_B on the ratio can be seen in Figure 8.8. This method is beneficial, since a large discrepancy between the inclusive and the exclusive measurements of $|V_{ub}|$ exists (see Section 2.1) and its uncertainty does not have to be considered. Furthermore, some experimental systematic uncertainties cancel in the ratio, mainly the uncertainty on the tagging efficiency correction. On the other hand, a (small) theoretical uncertainty from the theory expression of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ rate and the statistical uncertainty from the measurement have to be incorporated. As for the previous method, the theory expression is varied to obtain the theoretical uncertainties. For the calculation of the ratio Gaussian distributed uncertainties are assumed. Under this assumption an upper limit at 90% C.L. with a Bayesian prior of

$$\mathcal{R} \left(\frac{\Delta \mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)}{\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)} \right) < 4.2 \times 10^{-2}, \quad (8.6)$$

is found. This ratio is used to determine the limit on λ_B . The result can be found in Table 8.7.

As discussed in Section 2.4, QCD factorization results favor low values of $\lambda_B \approx 200$ MeV, whereas QCD sum rules favor values around 500 MeV. All upper limits

Table 8.7.: Calculated lower limits on λ_B for the different approaches. The values for $|V_{ub}|$ are taken from Reference [4]. The limits include the systematic and theoretical uncertainties.

	$ V_{ub} \left(10^{-3}\right)$	λ_B limit (MeV)
Inclusive	$4.49 \pm 0.15^{+0.16}_{-0.17} \pm 0.17$	> 383
Exclusive	$3.70 \pm 0.10 \pm 0.12$	> 316
Average	3.94 ± 0.36	> 333
Independent	–	> 377

found with the two approaches seem to favor the latter value.

The value of λ_B measured from the ratio \mathcal{R} can in principle be used to extract $|V_{ub}|$ from the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ measurement. λ_B is used as input for the theoretical expression which can be solved for $|V_{ub}|$. Figure 8.9 shows the values of $|V_{ub}|$ obtained for a scan over different values for $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ and \mathcal{R} values. The drawn ellipse corresponds to the measured central values. Due to the large statistical uncertainty no sensible measurement or upper limit on $|V_{ub}|$ can be established with the current result. For the measured central values of λ_B and $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ a value of $|V_{ub}^{\text{exp}}| = (4.6 \pm 3.3) \times 10^{-3}$ is obtained, where the uncertainties are combined. As mentioned in Section 2.2, the form factors have a constant contribution for large values of λ_B which corresponds to very small values of the partial branching fraction or the ratio, respectively. In this region, no value for λ_B can be found by numerical integration. This region is indicated by the gray hatched region in Figure 8.9.

8.4 Comparison to Previous Analyses

The partial branching fraction of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ with a cut on the signal-side photon energy of $E_\gamma > 1 \text{ GeV}$ was already measured by the Belle [1] and BaBar [2] collaboration (see Section 2.5). Both analyses were only able to set upper limits. A comparison of the previous results and the results presented in this thesis can be found in Table 8.8. For the muon final state the same upper limit was found as in the previous Belle analysis. In the electron and combined final states the upper limit was improved. The obtained limits are the most stringent upper limits on the partial branching fraction determined so far.

Compared to the previous Belle analysis, the improved tagging algorithm increased the sensitivity for a measurement of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$. For a partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$ the expected signal yield is increased to more than 50 signal events. This corresponds to an improvement by a factor of three compared to the previous Belle analysis – without increasing the expected background [1, Table I]. As shown in Section 6.4, a significance of nearly 5σ was

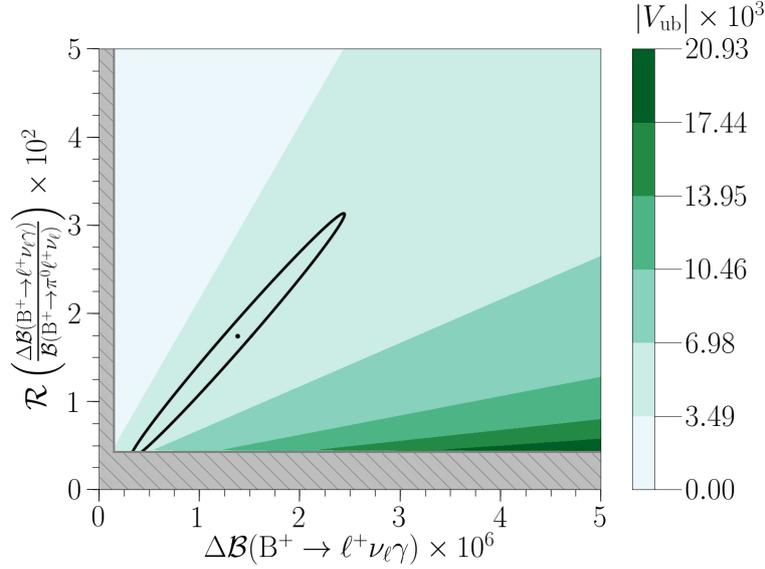


Figure 8.9.: Scan of the partial branching fraction and the ratio for the extraction of $|V_{ub}|$. The *gray hatched area* indicates the region where no value of λ_B and hence $|V_{ub}|$ can be found numerically.

Table 8.8.: Comparison to previous results of the partial branching fraction measurement. All limits correspond to the 90% C.L.

Final state	$\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ limit (10^{-6})		
	BaBar [2]	Belle [1]	This work
Electron	-	< 6.1	< 4.3
Muon	-	< 3.4	< 3.4
Both	< 14	< 3.5	< 3.0

expected. Although no significant signal is observed, the statistical and systematic uncertainties are reduced, resulting in a lower upper limit.

A decisive improvement of the analysis comes from the improved modeling of the peaking background $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$. In Figure 5.14a it is shown that the signal selection (mostly the cut on the signal-side photon energy) favors events with low q^2 . The flawed form factors lead to an overestimation of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ background in this region. In the previous Belle analysis, the yield of this background component was fixed in the fit to the expected number of events from MC. This might have forced the fit on data to underestimate the signal contribution in the signal region. The current analysis solves this issue by an improved decay model and fitting the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ background by making use of an additional sample.

As shown earlier, the extraction of λ_B depends heavily on the used input parameters, e.g. $|V_{ub}|$. In the BaBar analysis the authors give no lower limit on λ_B for the measured partial branching fraction. The authors of Reference [6] revisit the analysis and calculate a limit of $\lambda_B > 115$ MeV. The previous Belle analysis found a lower limit of $\lambda_B > 238$ MeV, where the input parameters of Reference [6] were used, except for the soft correction $\xi(E_\gamma)$, which was taken from Reference [7]. It has to be noted, that latter was calculated for a value of $\lambda_B = 300$ MeV. Hence, the result should be taken with a grain of salt. In addition, values for $|V_{ub}|$ were used, which are significantly lower than the current world averages. The current analysis uses the parameters given in Reference [6], except for the updated values of $|V_{ub}|$. This thesis establishes the most stringent lower limit on λ_B from experimental determination.

8.5 Outlook for Belle II

The search for the partial branching fraction of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ is currently limited by the available data set of B meson decays. In this work an improved tagging algorithm was used, but no significant measurement of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ could be established. However, with a larger data set it likely will be possible to observe the decay. The Belle II collaboration plans to record a data set 50 times larger than its predecessor experiment Belle [61]. The Belle II collaboration envisages to collect a data set of an integrated luminosity of 5 ab^{-1} after approximately two years. By 2025, the collaboration will have collected a data set of 50 ab^{-1} . In the following the impact of Belle II on the search of the partial branching fraction of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ is discussed.

A larger data set will help to reduce the statistical uncertainty of the measurement. These assumptions are of course only valid for comparable signal- and tag-side reconstruction efficiencies. That requires a good understanding of the dominant background processes and capable tools for their suppression. Selection cuts like the cut on the number of remaining tracks after the reconstruction of the $\Upsilon(4S)$ event or the extra energy E_{ECL} are viable for the analysis. Upcoming studies will have to prove that those variables are still well understood for the higher background rates and can be used in the analysis procedure.

Many systematic uncertainties have a large statistical component, which will decrease with a larger data set. The upcoming analysis will profit from reduced systematic uncertainties. With more than 500 publications the Belle collaboration made a huge effort in studying mainly B decays. Many decay modes were more precisely measured and are now being updated in the decay tables used for the MC generation. Additionally, the theoretical understanding of many decays was improved over the last decade, e.g. the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay. Those findings are being used to improve the utilized `EvtGen` models. The improvements will help to

reduce the discrepancies between MC and data which contribute to most systematic uncertainties. The largest multiplicative uncertainty originates from the tag efficiency correction. First of all, improved MC samples will help to equalize the performance of the tagging algorithm on MC and data. Its assigned systematic uncertainty is currently dominated by the uncertainties from MC corrections applied to the signal-side, which will then decrease. Supplementary, the statistical uncertainty will be reduced due to the large data sample. The additive systematic uncertainties in the current analysis are mostly evaluated by reweighting MC distributions and taking the difference to the nominal result as a systematic uncertainty. With improved MC samples, the differences to data should be reduced. A reduction of the statistical and systematic uncertainty might also come from larger background MC samples. In the current analysis it is found that the higher M_{miss}^2 region is dominantly populated by charged or neutral $b \rightarrow c$ decays. Those arise from combinatorial background, where in total ten streams are available for the current Belle analysis, leading to large individual event weights and hence a peaky background description. For Belle II it might be possible to produce more MC samples at reasonable computing costs.

In summary, the statistical and systematic uncertainties will be reduced. However, at Belle II additional sources of systematic uncertainty might have to be considered. The systematic uncertainties will likely dominate. An extrapolation to the expected Belle II data set can be made using the currently measured central values. The statistical uncertainties are scaled according to the expected luminosity. The expected statistical uncertainty in percent with respect to the current measurement can be seen in Table 8.9. It will be drastically reduced. In addition, the expected significance is calculated under the assumption of Gaussian distributed uncertainties. The given number does take the systematic uncertainty into account, which is conservatively expected to decrease by 50%. Under these assumptions, first evidence should be possible after two years of data taking. If a significant measurement can be made with Belle II, it will be possible to set central values for λ_B and $|V_{ub}|$. The introduced method to extract λ_B independently from $|V_{ub}|$ will then aim for the most precise measurement of λ_B , since

Table 8.9.: Statistical uncertainties and significance of the improved Belle analysis and for the expected Belle II data set with respect to the extracted partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)_{E_\gamma > 1.0 \text{ GeV}}$.

	Belle 711 fb ⁻¹	Belle II 5 ab ⁻¹	Belle II 50 ab ⁻¹
Stat. uncertainty	±71%	±27%	±9%
Significance	1.3σ	3.3σ	6.4σ

the theoretical uncertainties on $|V_{ub}|$ can be neglected and the most dominant systematic uncertainties arising from the tag efficiency correction will cancel in the ratio \mathcal{R} . In addition, it might then be possible to establish a new channel to measure $|V_{ub}|$. Figure 8.10 shows the projection for the measurement of λ_B and $|V_{ub}|$ with respect to measured central values. The ellipses correspond to the expected statistical and systematic uncertainties. For the projection symmetric Gaussian uncertainties are assumed. With an increased data set from Belle II the uncertainties on the measured parameters can be drastically reduced by about 90%.

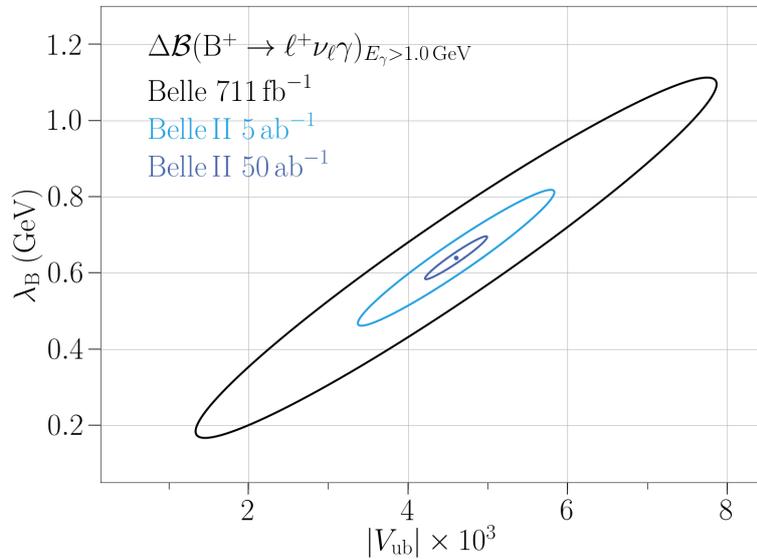


Figure 8.10.: Projection of the extraction of λ_B and $|V_{ub}|$ for the expected Belle II data sets. The ellipses correspond to the expected uncertainty.

At Belle II the partial branching fraction of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ should ideally be measured for several cuts on the signal-side photon energy above 1 GeV. This would reduce the theoretical uncertainties originating from the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ form factors and allow for a more precise measurement of λ_B .

CHAPTER 9

Conclusion

In this thesis a search for the rare decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ with a cut on the signal-side photon of 1 GeV is presented. The analysis is performed on the full Belle data set corresponding to an integrated luminosity of 711 fb^{-1} .

Important contributions of this work to the analysis software allowed to convert Belle MC and data and to use a novel tagging approach. For the first time, a multivariate signal-specific tagging algorithm is applied to reconstruct the tag-side B meson in many thousand hadronic decay channels on reconstructed Belle collision data. The algorithm is specifically trained for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay and calibrated on three independent channels to incorporate differences between MC and data. With this new approach the signal reconstruction efficiency could be increased by a factor of three with respect to the best previous tagged measurement, without increasing the expected background. In addition, the description and extraction of the most challenging background, semileptonic $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decays, is significantly improved.

Despite these enhancements, no significant excess in data over background expectation is observed and a Bayesian upper limit on the partial branching fraction of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ is quoted. The obtained observed upper limits at 90% C.L. for the electron, muon, and combined electron and muon final states are:

$$\begin{aligned}\Delta\mathcal{B}(B^+ \rightarrow e^+ \nu_e \gamma) &< 4.3 \times 10^{-6} \\ \Delta\mathcal{B}(B^+ \rightarrow \mu^+ \nu_\mu \gamma) &< 3.4 \times 10^{-6} \\ \Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) &< 3.0 \times 10^{-6}.\end{aligned}$$

The electron and combined results are the world's most stringent limits on such decays, and supersede the previous Belle results. The quoted limits are sensitive to the first inverse momentum of the light-cone distribution amplitude of the B meson, λ_B .

In this thesis, a novel method was developed to reduce theoretical and experimental uncertainties for the determination of λ_B : the key idea is to measure the ratio of the partial branching fraction of $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ with respect to the branching ratio of the dominant peaking background of $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$. This reduces the overall experimental uncertainties and allows for an extraction of λ_B independent of the measured value of the CKM matrix element $|V_{ub}|$:

$$\lambda_B > 377 \text{ MeV at } 90\% \text{ C.L.}$$

The obtained limit favors values predicted by QCD sum rule calculations, and is in mild tension with the predictions using QCD factorization and measured branching fractions of $B \rightarrow \pi\pi$, $\pi\rho$ and $\rho_L\rho_L$ decays.

If the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ partial branching fraction is of the order of 10^{-6} , the decay will be observable with the Belle II experiment and a 5σ discovery is possible using about 10 ab^{-1} of data. With the full Belle II data set of 50 ab^{-1} a statistical precision of about 9% on the partial branching fraction can be reached, allowing for a precise extraction of λ_B and also to determine $|V_{ub}|$. Figure 9.1 shows the averaged measurements of $|V_{ub}|$ using exclusive and inclusive semileptonic decays of the last 16 years with their growing tension, and their corresponding expected sensitivity with the full Belle II data set [62]. Although a measurement of $|V_{ub}|$ via the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decays will not reach a similar precision as the established measurement methods, its value will help to clarifying their longstanding discrepancy.

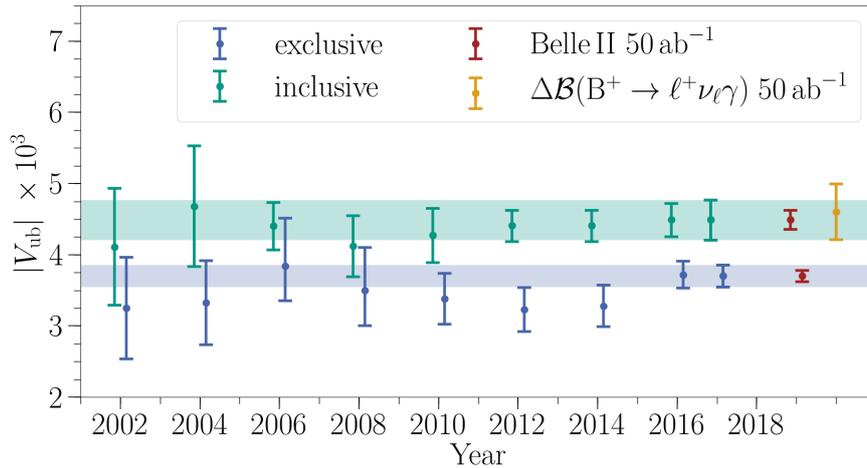


Figure 9.1.: World averages of the exclusive and inclusive $|V_{ub}|$ measurements, the projection for Belle II and the projection for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ measurement.

Appendices

APPENDIX A

B2BII Validation Plots

In this chapter some exemplary validation plots of the `b2bii` package are shown. The same simulated $b \rightarrow c$ events were processed with `BASF` (blue) and the `b2bii` package (black). Different quantities of charged tracks are plotted. See Section 4.2 for a detailed discussion of the observed differences.

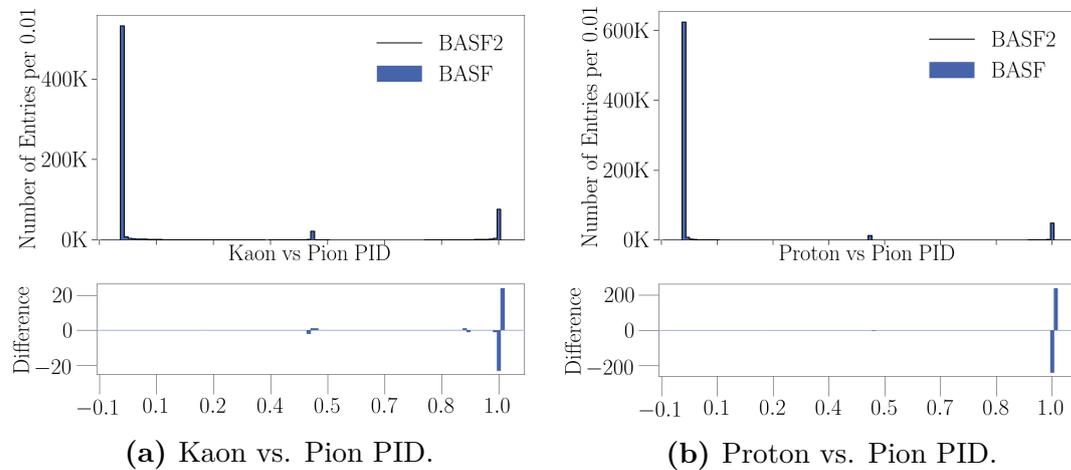


Figure A.1.: B2BII validation plots.

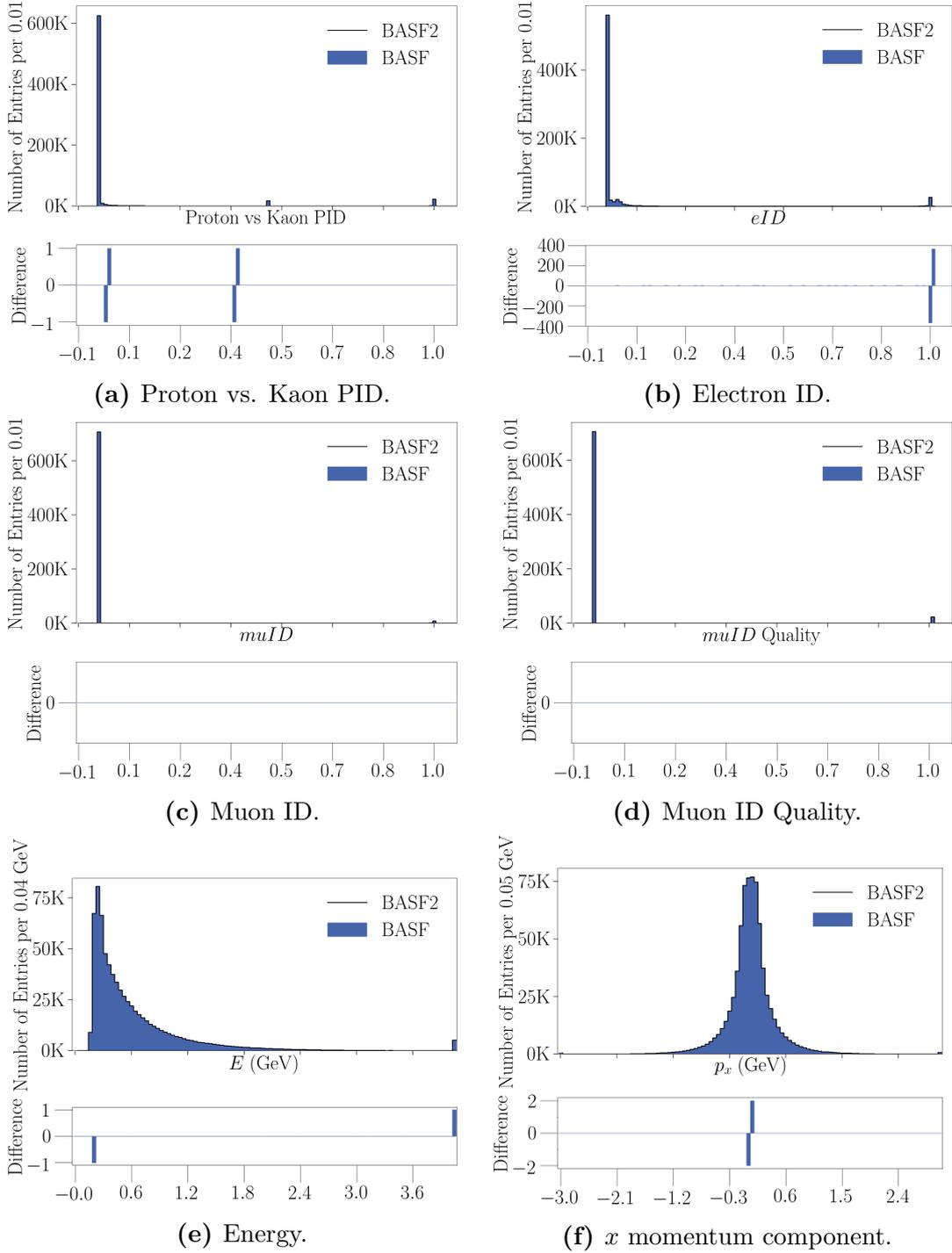


Figure A.2.: B2BII validation plots.

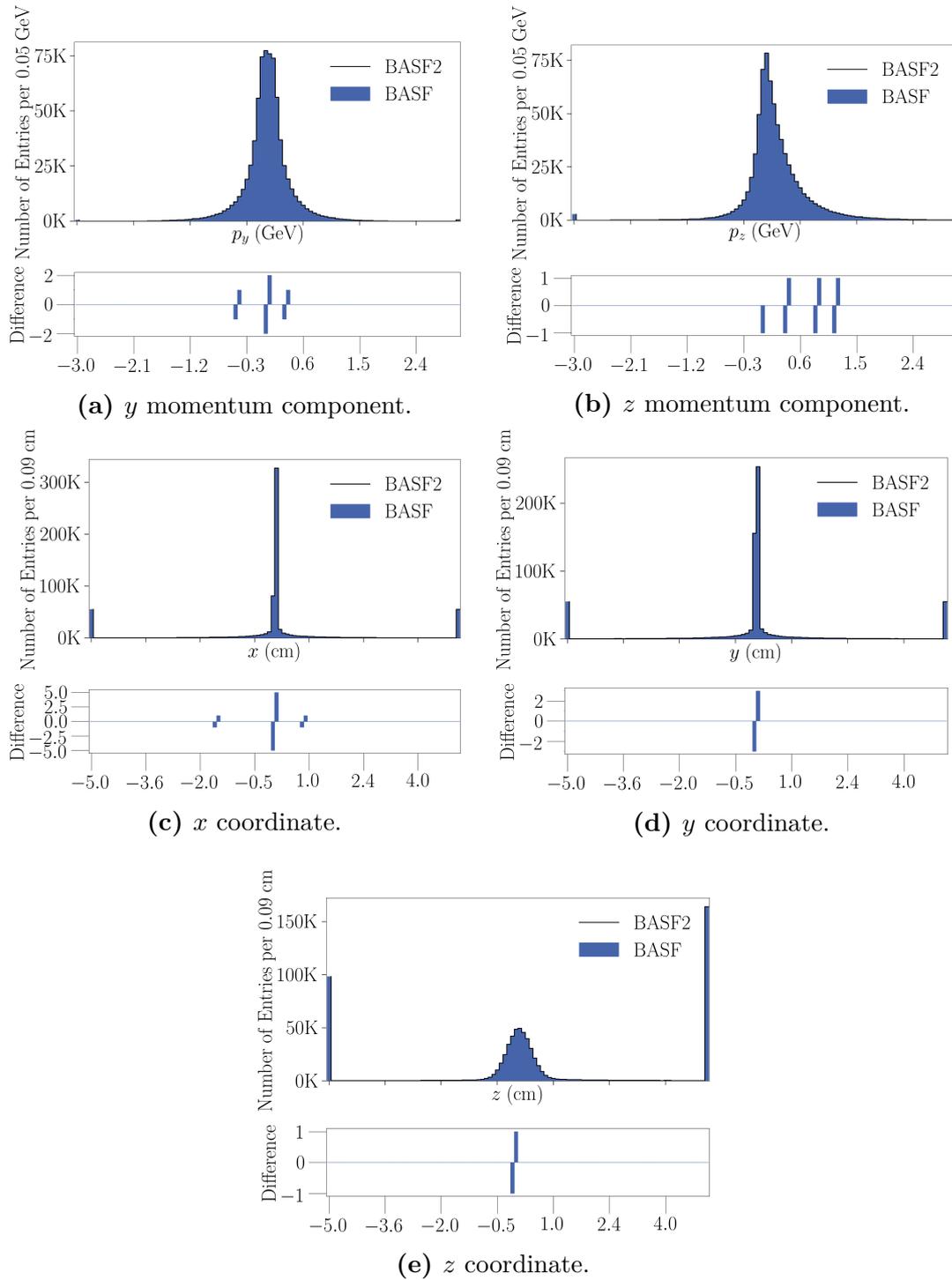


Figure A.3.: B2BII validation plots.

BDT Control Plots

This section contains the control plots for the used multivariate classifiers of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample.

B.1 Continuum Suppression

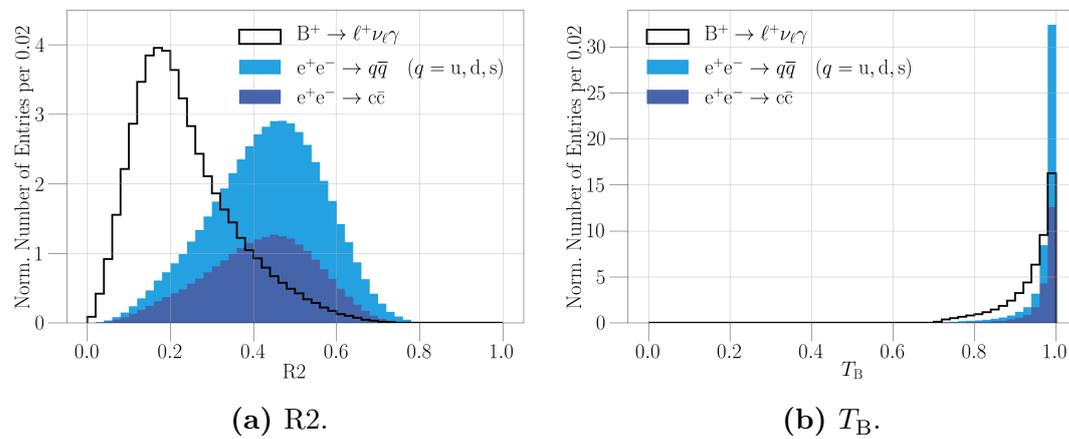
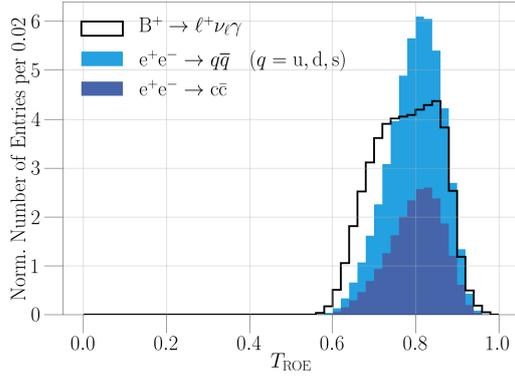
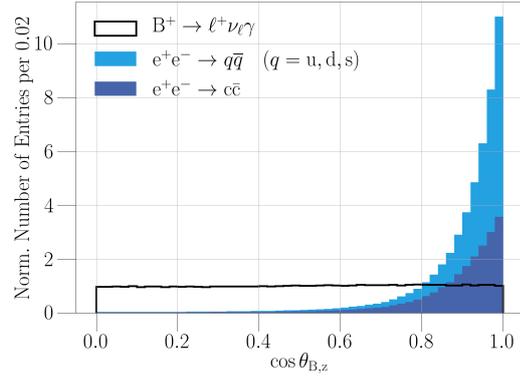


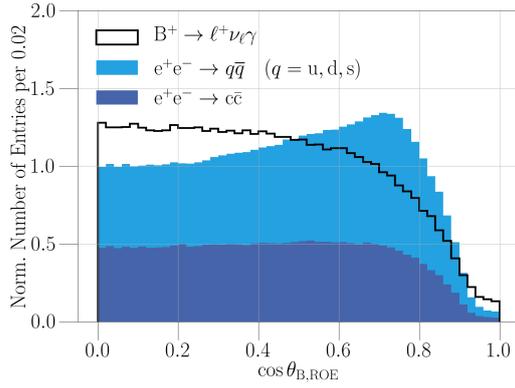
Figure B.1.: Continuum suppression variables.



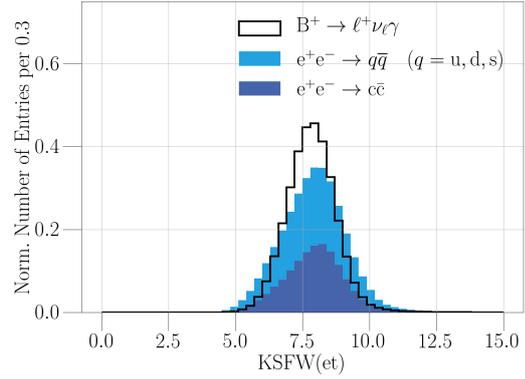
(a) T_{ROE} .



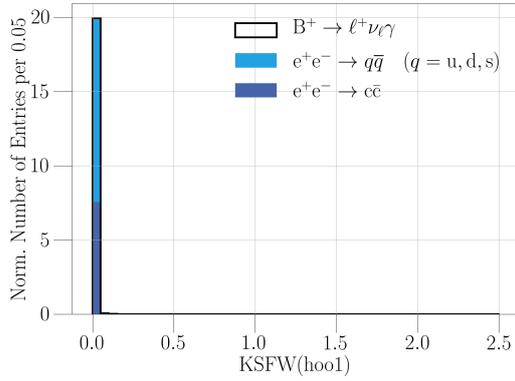
(b) $\cos \theta_{B,z}$.



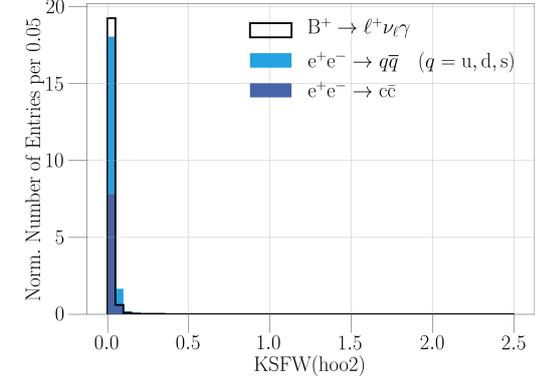
(c) $\cos \theta_{B,\text{ROE}}$.



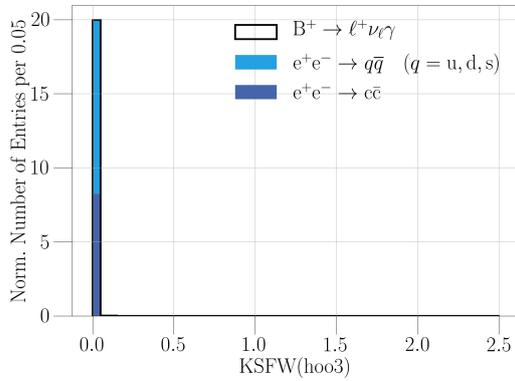
(d) $\text{KSFW}(\text{et})$.



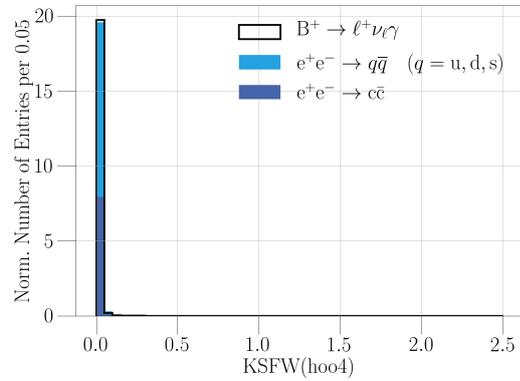
(e) $\text{KSFW}(\text{hoo1})$.



(f) $\text{KSFW}(\text{hoo2})$.

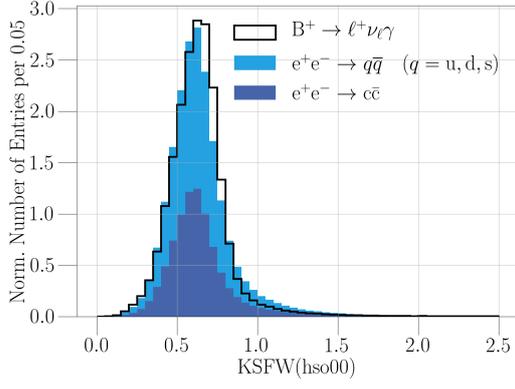


(g) $\text{KSFW}(\text{hoo3})$.

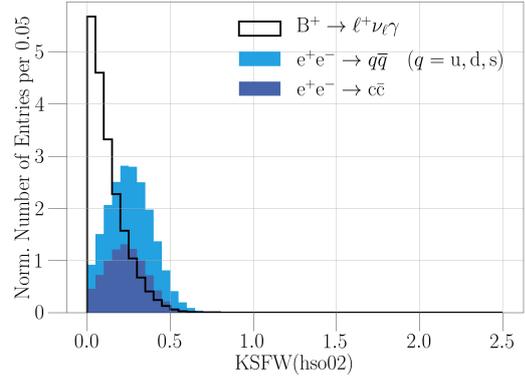


(h) $\text{KSFW}(\text{hoo4})$.

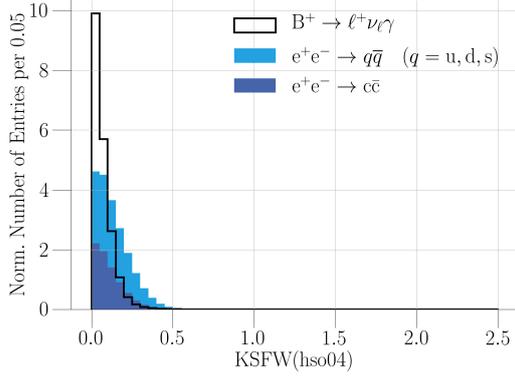
Figure B.2.: Continuum suppression variables.



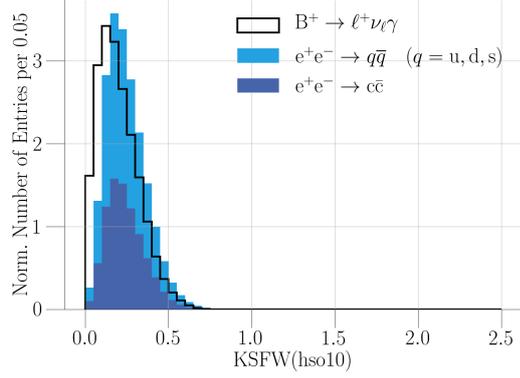
(a) KSFW(hso00).



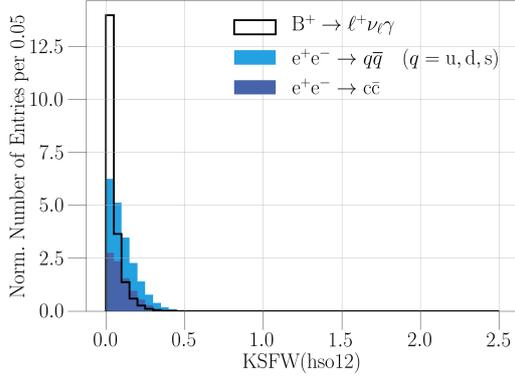
(b) KSFW(hso02).



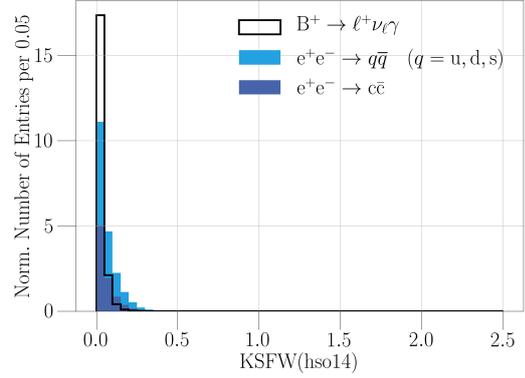
(c) KSFW(hso04).



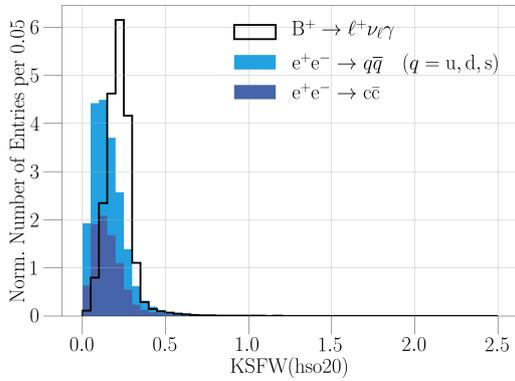
(d) KSFW(hso10).



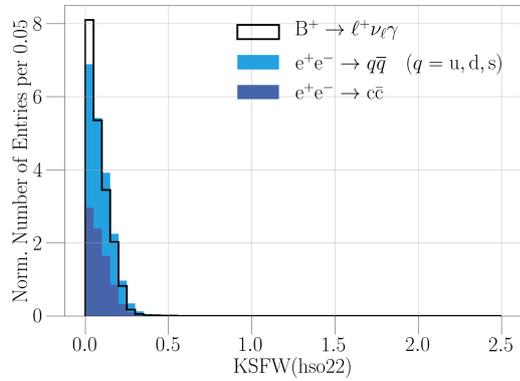
(e) KSFW(hso12).



(f) KSFW(hso14).

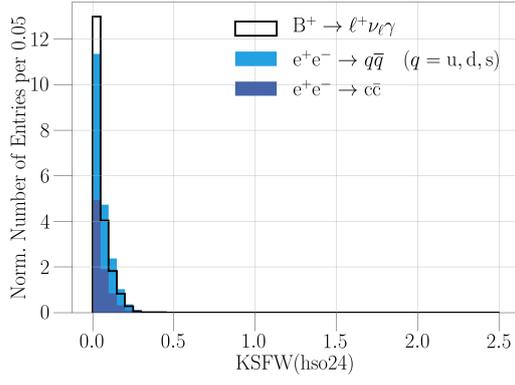


(g) KSFW(hso20).

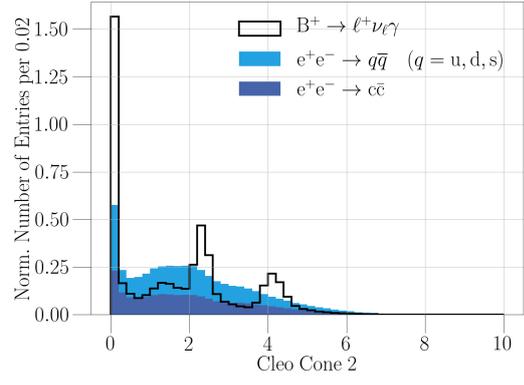


(h) KSFW(hso22).

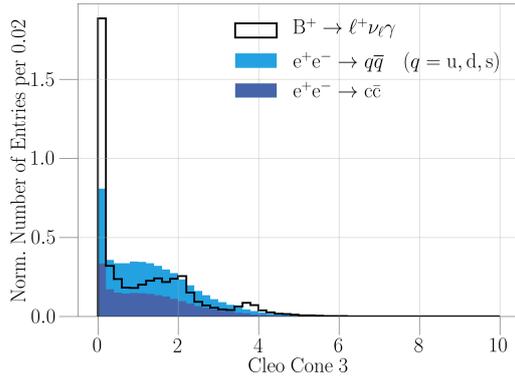
Figure B.3.: Continuum suppression variables.



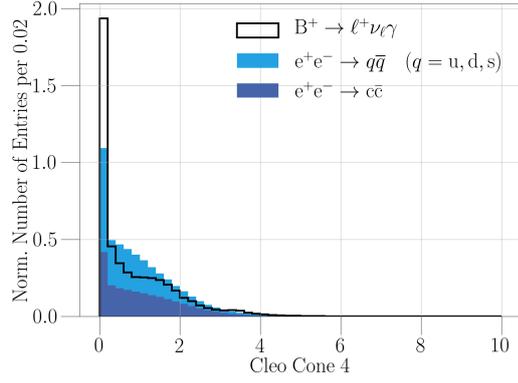
(a) KSFW(hso24).



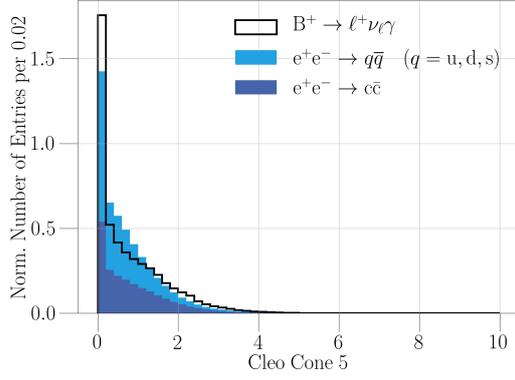
(b) Cleo Cone 2.



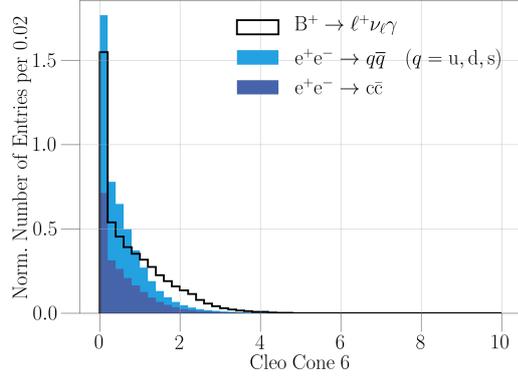
(c) Cleo Cone 3.



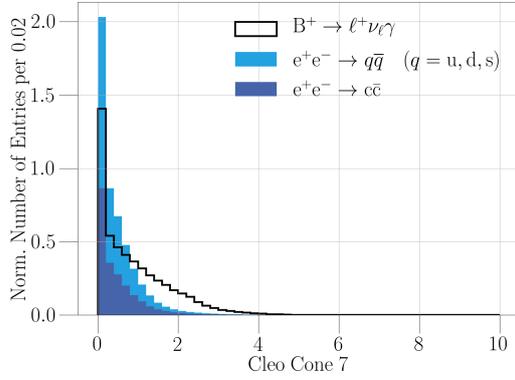
(d) Cleo Cone 4.



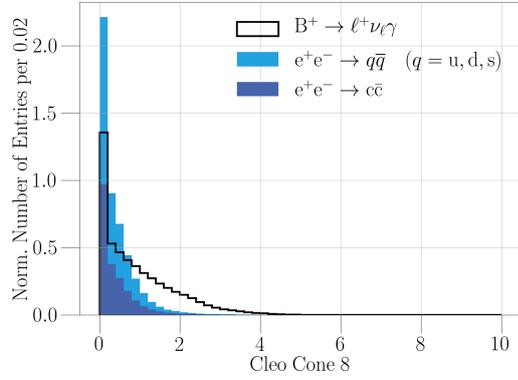
(e) Cleo Cone 5.



(f) Cleo Cone 6.

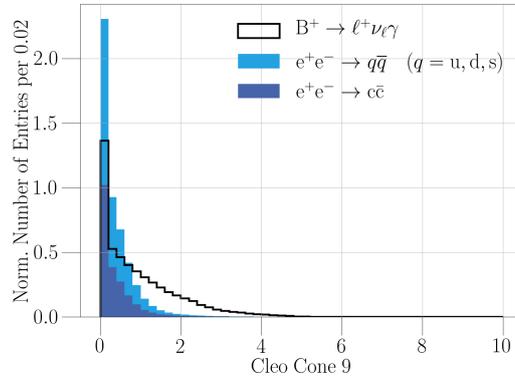


(g) Cleo Cone 7.



(h) Cleo Cone 8.

Figure B.4.: Continuum suppression variables.



(a) Cleo Cone 9.

Figure B.5.: Continuum suppression variables.

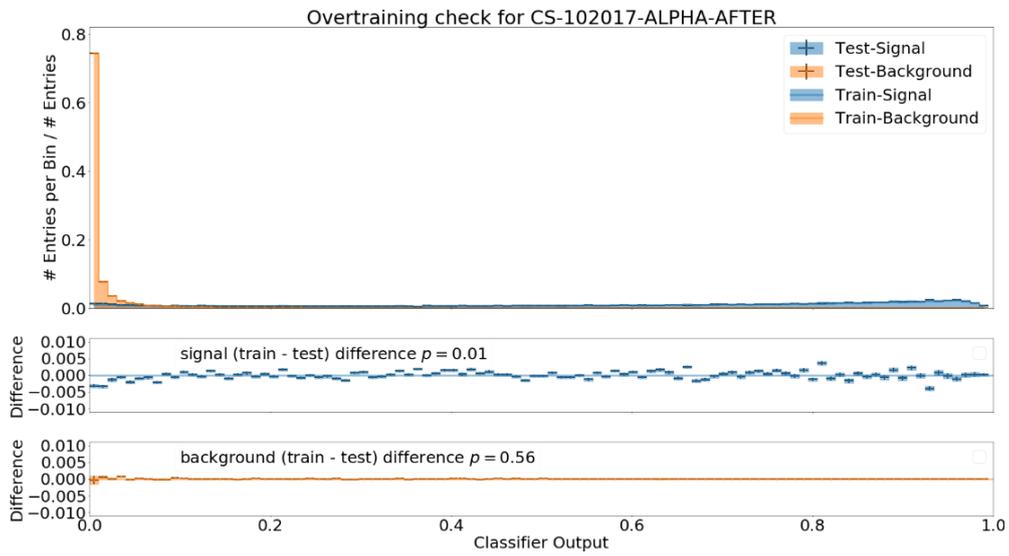


Figure B.6.: Overtraining check for the trained continuum suppression BDT on test and training data set.

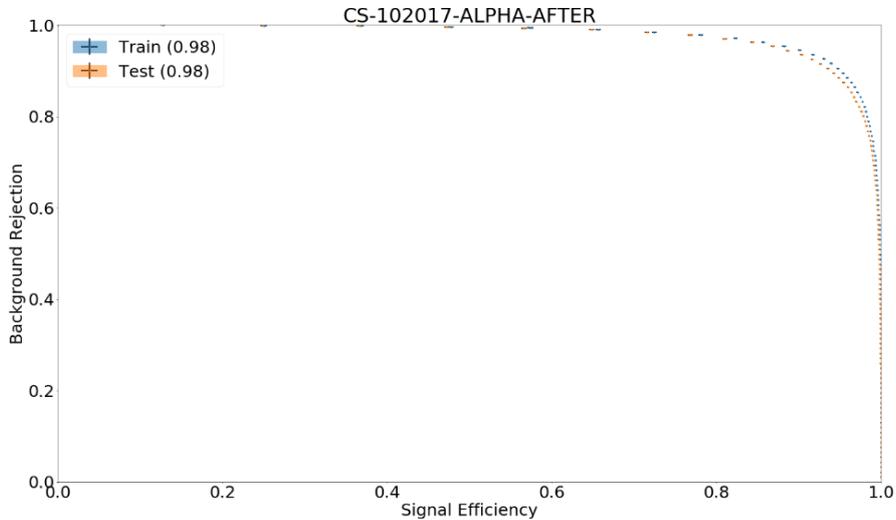
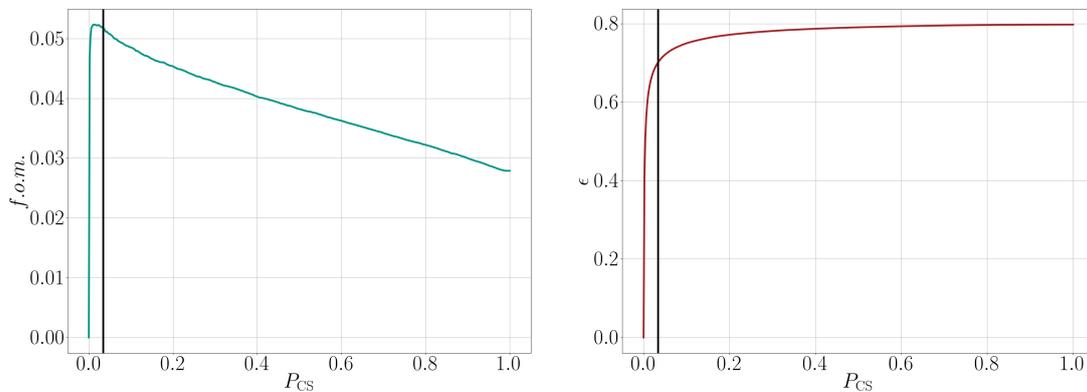


Figure B.7.: Receiver operating curve for the trained continuum suppression BDT on test and training data set.



(a) Punzi FOM.

(b) Signal efficiency.

Figure B.8.: FOM and signal efficiency dependent on the cut of the continuum suppression. All final selection cuts are applied, except for the cut on the peaking background. The *black line* shows the optimal cut found during the optimization with a desired standard deviation of $\sigma = 3$. See Section 5.2.7 for more information.

B.2 Peaking Background Suppression

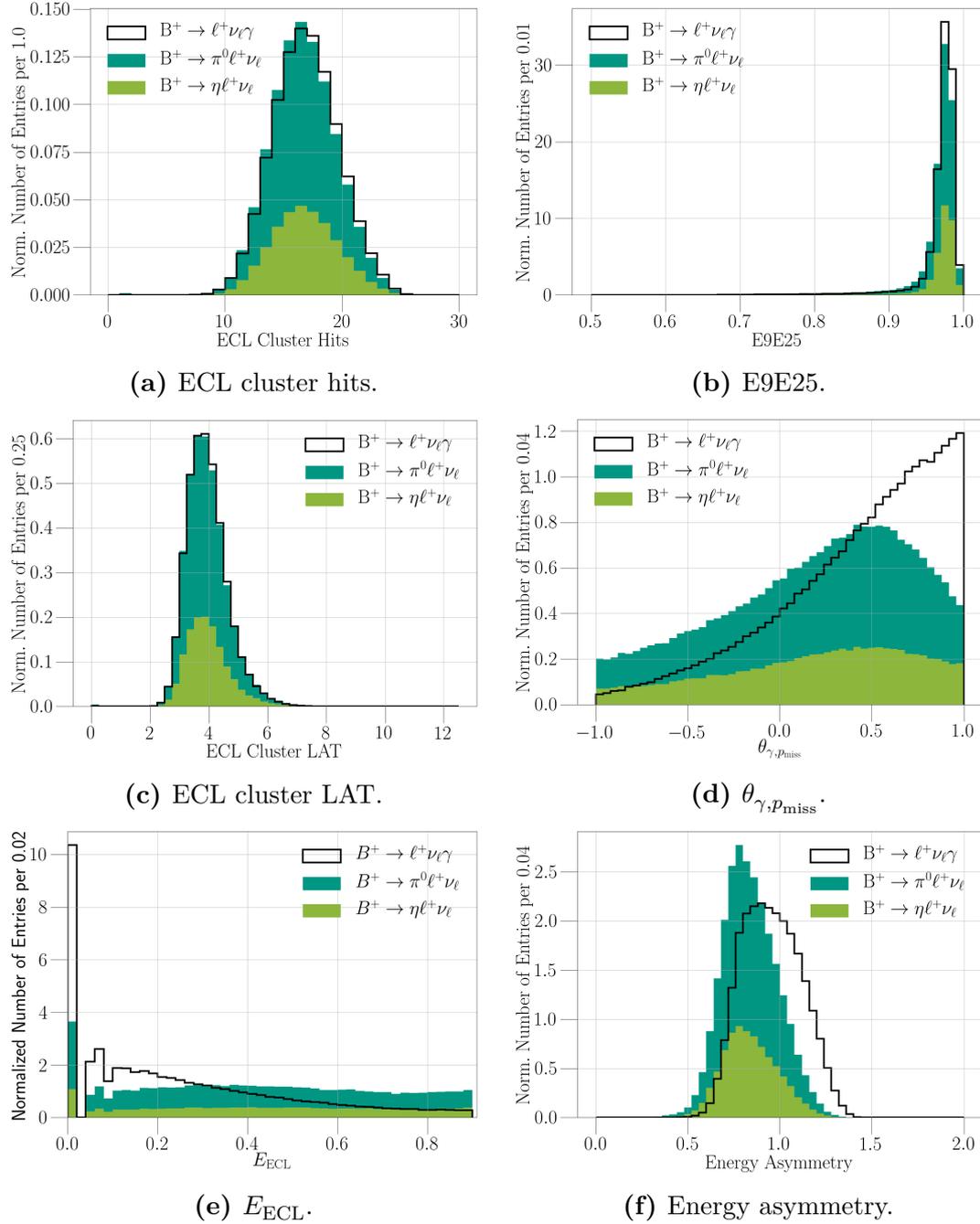


Figure B.9.: Peaking background suppression variables.

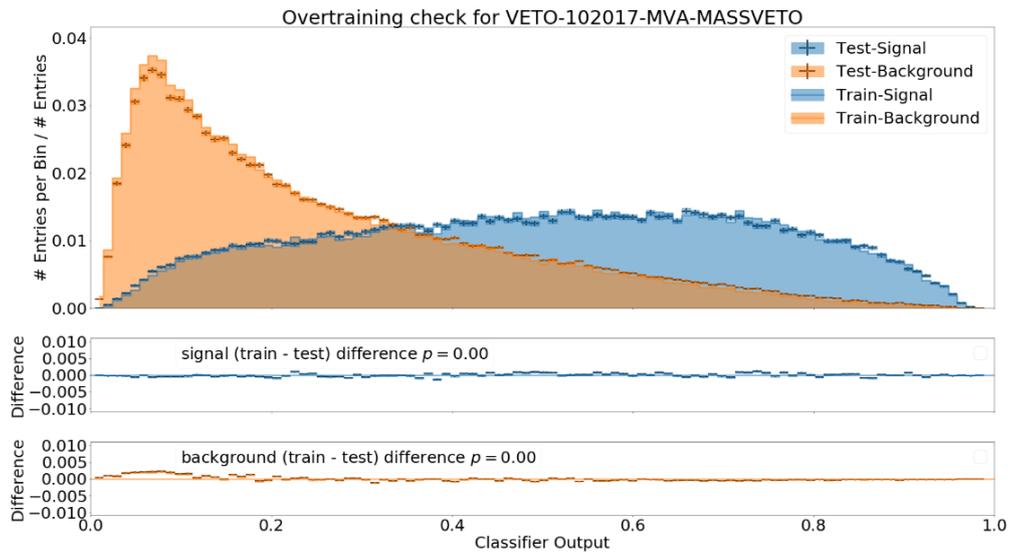


Figure B.10.: Overtraining check for the trained peaking background suppression BDT on test and training data set.

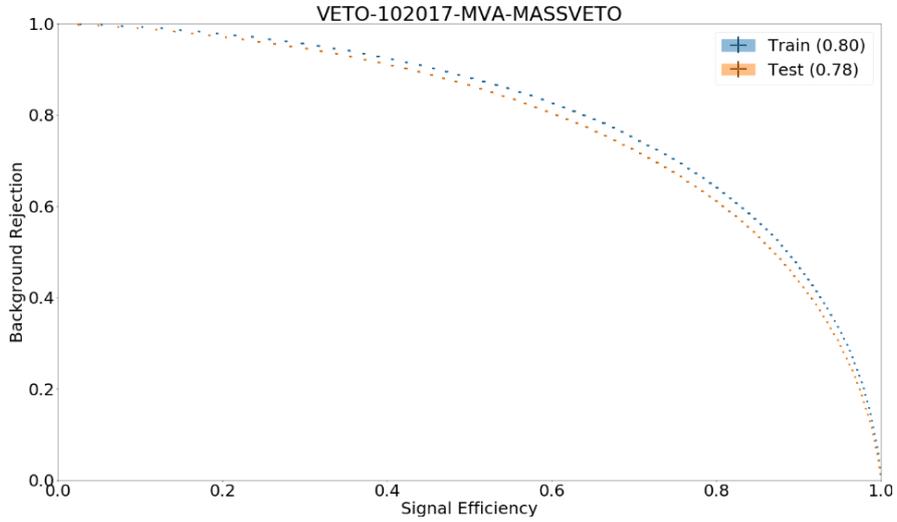


Figure B.11.: Receiver operating curve for the trained peaking background suppression BDT on test and training data set.

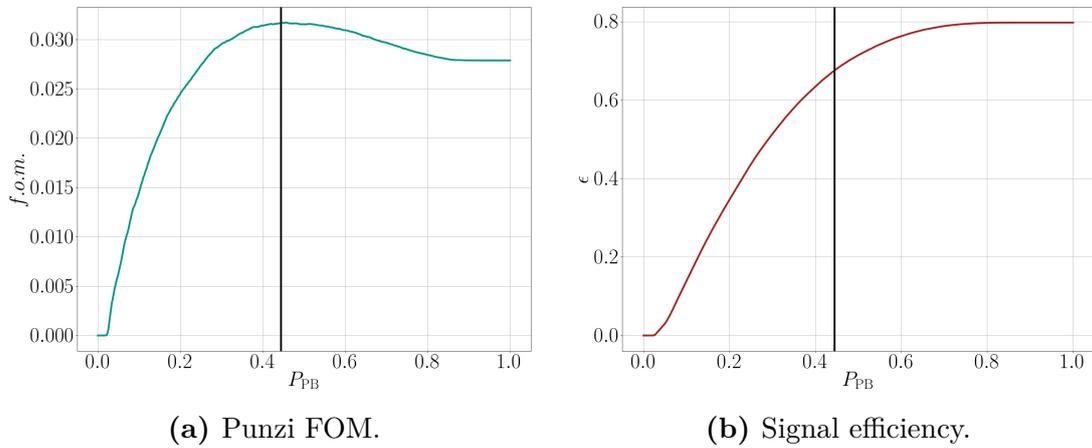


Figure B.12.: FOM and signal efficiency dependent on the cut of the peaking background suppression. All final selection cuts are applied, except for the cut on the continuum background. The *black line* shows the optimal cut found during the optimization with a desired standard deviation of $\sigma = 3$. See Section 5.2.7 for more information.

C.1 Selection Cuts

The following cuts were used for the reconstruction of the calibration channels.

C.1.1 Event Selection

- HadronBJ skim

C.1.2 Final State Particles

- e^+/μ^+
 - $eID > 0.8$
 - $muID > 0.8$
 - $|\Delta r| < 2 \text{ cm}$
 - $|\Delta z| < 4 \text{ cm}$
- γ
 - *goodGamma*
 - $E9E25 > 0.827 \text{ GeV}$
 - *clusterHits* > 5
- π^+

- $actPIDBelle(3, 2) < 0.4$
- $|\Delta r| < 2 \text{ cm}$
- $|\Delta z| < 4 \text{ cm}$

- K^+

- $actPIDBelle(3, 2) > 0.6$

C.1.3 $B^+ \rightarrow \bar{D}^0(\rightarrow K^+ \pi^-) \ell^+ \nu_\ell$

- D^0

- $M_{D^0} \in (1.858, 1.872) \text{ GeV}$
- Unconstrained vertex fit: $P_\chi^2 > 0.01$

- B_{sig}

- $|\cos(\theta_{BD\ell})| < 3.0$
- $M_{bc} > 4.5 \text{ GeV}$
- Unconstrained vertex fit: $P_\chi^2 > 0.01$

- B_{tag}

- $\Delta E \in [-0.15, 0.1] \text{ GeV}$
- $M_{bc} \in [5.27, 5.29] \text{ GeV}$

- $\Upsilon(4S)$

- $M_{miss}^2 \in (-6.0, 8.0) \text{ GeV}^2$
- $E_{ECL} < 1.0 \text{ GeV}$
- Best candidate selection:
 $\Upsilon(4S)$ candidate with highest tagging probability P_{FEI} of the B_{tag} candidate.
- $P_{FEI} > 0.01$
- Remaining $N_{tracks} = 0$

C.1.4 $B^+ \rightarrow \bar{D}^0(\rightarrow K^+ \pi^- \pi^0) \ell^+ \nu_\ell$

- π^0

- $M_{\pi^0} \in (117.8, 150.2) \text{ MeV}$

- D^0

- $M_{D^0} \in (1.849, 1.879) \text{ GeV}$
- Unconstrained vertex fit: $P_{\chi^2} > 0.01$
- B_{sig}
 - $|\cos(\theta_{BD\ell})| < 3.0$
 - $M_{\text{bc}} > 4.5 \text{ GeV}$
 - Unconstrained vertex fit: $P_{\chi^2} > 0.01$
- B_{tag}
 - $\Delta E \in [-0.15, 0.1] \text{ GeV}$
 - $M_{\text{bc}} \in [5.27, 5.29] \text{ GeV}$
- $\Upsilon(4S)$
 - $M_{\text{miss}}^2 \in (-6.0, 8.0) \text{ GeV}^2$
 - $E_{\text{ECL}} < 1.0 \text{ GeV}$
 - Best candidate selection:
 $\Upsilon(4S)$ candidate with highest tagging probability P_{FEI} of the B_{tag} candidate.
 - $P_{\text{FEI}} > 0.01$
 - Remaining $N_{\text{tracks}} = 0$

C.1.5 $B^+ \rightarrow \bar{D}^0(\rightarrow K^+ \pi^- \pi^- \pi^+) \ell^+ \nu_\ell$

- D^0
 - $M_{D^0} \in (1.854, 1.872) \text{ GeV}$
 - Unconstrained vertex fit: $P_{\chi^2} > 0.01$
- B_{sig}
 - $|\cos(\theta_{BD\ell})| < 3.0$
 - $M_{\text{bc}} > 4.5 \text{ GeV}$
 - Unconstrained vertex fit: $P_{\chi^2} > 0.01$
- B_{tag}
 - $\Delta E \in [-0.15, 0.1] \text{ GeV}$
 - $M_{\text{bc}} \in [5.27, 5.29] \text{ GeV}$
- $\Upsilon(4S)$

- $M_{\text{miss}}^2 \in (-6.0, 8.0) \text{ GeV}^2$
- $E_{\text{ECL}} < 1.0 \text{ GeV}$
- Best candidate selection:
 $\Upsilon(4S)$ candidate with highest tagging probability P_{FEI} of the B_{tag} candidate.
- $P_{\text{FEI}} > 0.01$
- Remaining $N_{\text{tracks}} = 0$

C.2 Signal Extraction

Figure C.1 shows the global fit used for the extraction of the calibration factors for each calibration channel.

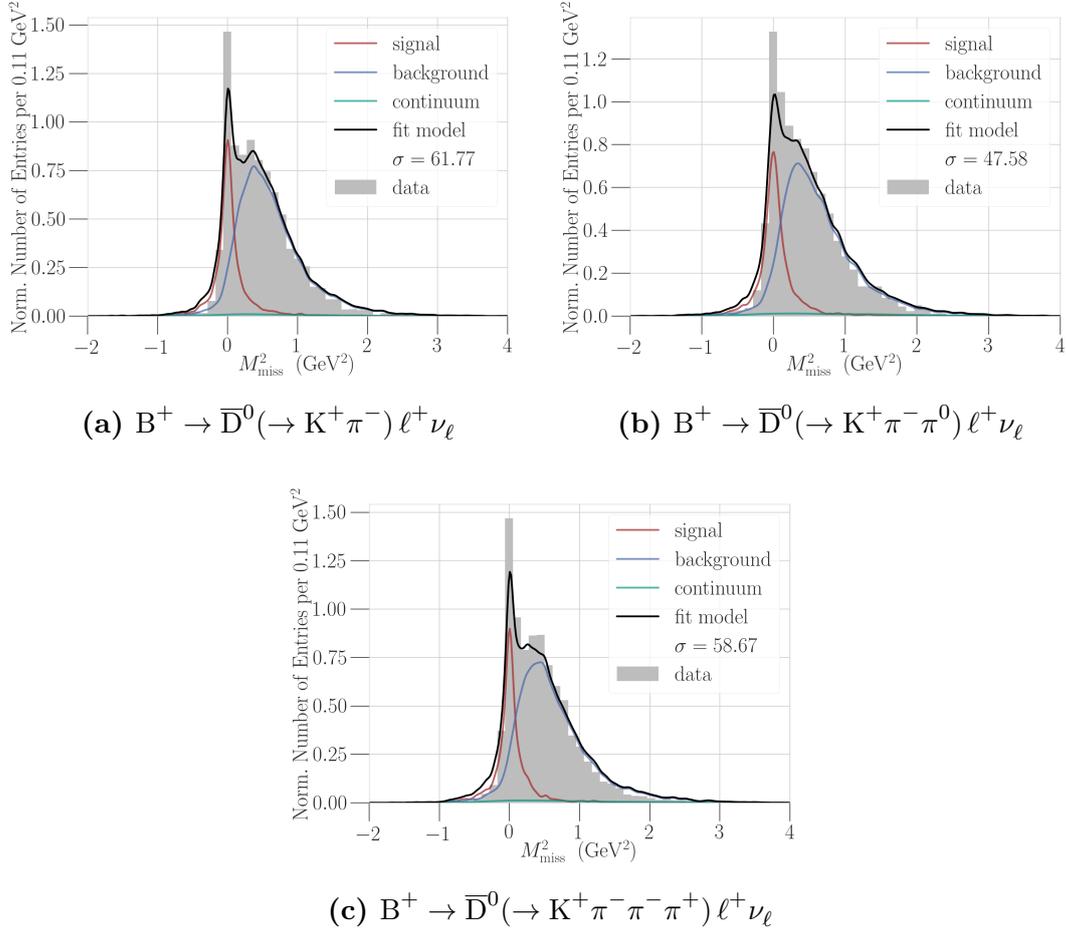


Figure C.1.: Results of the fit to M_{miss}^2 for the different calibration channels. Each calibration channel contains all tag channels.

APPENDIX D

Remaining Background Composition

This section contains the background composition over the full range of $M_{\text{miss}}^2 \in (-1.5, 3.0) \text{ GeV}^2$, see Section 5.4.

Table D.1.: Background composition of the **electron final state** in the **full range** of $M_{\text{miss}}^2 \in (-1.5, 3.0) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5 % are shown.

MC sample	Decay	Relative fraction(%)
b \rightarrow c		
61.4 %		
	$B^+ \rightarrow \bar{D}^*(2007)^0 \ell^+ \nu_\ell$	51.59
	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$	26.3
	$B^0 \rightarrow D^*(2010)^- \ell^+ \nu_\ell$	12.23
	$B^0 \rightarrow D^- \ell^+ \nu_\ell$	5.99
$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$		
16.1 %		
	$B^+ \rightarrow \pi^0 e^+ \nu_e$	100
b \rightarrow u $\ell^+ \nu_\ell$		
13.2 %		

Continued on next page

MC sample	Decay	Relative fraction (%)
	$B^+ \rightarrow \omega \ell^+ \nu_\ell$	25.96
	$B^0 \rightarrow \rho^- \ell^+ \nu_\ell$	20.98
	$B^+ \rightarrow X_u^0 \ell^+ \nu_\ell$	20.33
	$B^0 \rightarrow X_u^- \ell^+ \nu_\ell$	7.85
	$B^+ \rightarrow f_2(1270) \ell^+ \nu_\ell$	5.16
	$B^+ \rightarrow b_1(1235)^0 \ell^+ \nu_\ell$	3.35
	$B^+ \rightarrow \eta' \ell^+ \nu_\ell$	2.96
	$B^0 \rightarrow a_2(1320)^- \ell^+ \nu_\ell$	1.94
	$B^+ \rightarrow a_1(1260)^0 \ell^+ \nu_\ell$	1.67
	$B^+ \rightarrow \rho^0 \ell^+ \nu_\ell$	1.54
$B^+ \rightarrow \eta \ell^+ \nu_\ell$		
6.9 %		
	$B^+ \rightarrow \eta e^+ \nu_e$	99.97
$e^+ e^- \rightarrow q \bar{q}$		
2.0 %	-	-
Rare		
0.4 %		
	$B^+ \rightarrow X_{su} \gamma$	9.16
	$B^+ \rightarrow X_{su} \eta$	7.65
	$B^+ \rightarrow D_S^+ \pi^0$	6.14
	$B^+ \rightarrow \pi^+ \pi^0$	3.09
	$B^+ \rightarrow K^*(1680)^0 \pi^+$	3.06
	$B^+ \rightarrow K^0 \pi^+ \pi^0$	1.68
	$B^+ \rightarrow K_0^*(1430)^0 \pi^+$	1.54
	$B^+ \rightarrow \eta'(958) K^+$	1.54
	$B^+ \rightarrow \tau^+ (\rightarrow e^+ \nu_e \bar{\nu}_\tau) \nu_\tau$	1.54
	$B^0 \rightarrow X_{sd} \gamma$	1.53
	$B^0 \rightarrow K^*(1410)^+ \pi^-$	1.53

Table D.2.: Background composition of the **muon final state** in the **full range** of $M_{\text{miss}}^2 \in (-1.5, 3.0) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5% are shown.

MC sample	Decay	Relative fraction (%)
b \rightarrow c		
59.7 %		
	$B^+ \rightarrow \bar{D}^*(2007)^0 \ell^+ \nu_\ell$	50.67
	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu_\ell$	24.74
	$B^0 \rightarrow D^*(2010)^- \ell^+ \nu_\ell$	11.67
	$B^0 \rightarrow D^- \ell^+ \nu_\ell$	6.1
$B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$		
14.8 %		
	$B^+ \rightarrow \pi^0 \mu^+ \nu_\mu$	100
b \rightarrow u $\ell^+ \nu_\ell$		
11.2 %		
	$B^+ \rightarrow \omega \ell^+ \nu_\ell$	27.47
	$B^+ \rightarrow X_u^0 \ell^+ \nu_\ell$	21.47
	$B^0 \rightarrow \rho^- \ell^+ \nu_\ell$	18.94
	$B^0 \rightarrow X_u^- \ell^+ \nu_\ell$	8.02
	$B^+ \rightarrow \eta' \ell^+ \nu_\ell$	4.16
	$B^+ \rightarrow f_2(1270) \ell^+ \nu_\ell$	3.23
	$B^+ \rightarrow \rho^0 \ell^+ \nu_\ell$	2.32
	$B^+ \rightarrow b_1(1235)^0 \ell^+ \nu_\ell$	2.11
	$B^+ \rightarrow a_1(1260)^0 \ell^+ \nu_\ell$	2.09
	$B^0 \rightarrow a_2(1320)^- \ell^+ \nu_\ell$	1.62
$e^+ e^- \rightarrow q \bar{q}$		
7.0 %	-	-
$B^+ \rightarrow \eta \ell^+ \nu_\ell$		

Continued on next page

MC sample	Decay	Relative fraction (%)
5.9 %	$B^+ \rightarrow \eta \mu^+ \nu_\mu$	100
Rare		
1.4 %	$B^+ \rightarrow X_{su} \gamma$	13.45
	$B^+ \rightarrow K^+ \pi^0$	6.61
	$B^+ \rightarrow \eta'(958) K^+$	4.91
	$B^+ \rightarrow K^*(892)^+ \gamma$	3.59
	$B^+ \rightarrow K^0 \pi^+ \pi^0$	3.53
	$B^+ \rightarrow \eta K^+$	2.46
	$B^0 \rightarrow X_{sd} \gamma$	2.42
	$B^+ \rightarrow K^0 \pi^+ (\gamma)$	2.31
	$B^+ \rightarrow K^+ K^0 \pi^0$	2.01
	$B^+ \rightarrow f_2(1270) \pi^+$	2.01
	$B^+ \rightarrow D_S^+ \pi^0$	1.93
	$B^+ \rightarrow K^*(892)^0 \pi^+$	1.61
	$B^+ \rightarrow K_0^*(1430)^0 \pi^+$	1.59

List of Figures

2.1.	Illustration of quark transitions, the numbers represent the corresponding CKM matrix element [4].	4
2.2.	Evolution of the averages of exclusive and inclusive $ V_{ub} $ measurements, where the error bars represent the combined uncertainty. The <i>colored bands</i> correspond to the latest averages. All values are taken from Reference [4] and earlier publications.	4
2.3.	Dominant Feynman graph at tree-level with photon emission from the up quark [5].	5
2.4.	Dependence of the vector and axial-vector form factor on the photon energy for $\lambda_B = 300$ MeV (<i>continuous lines</i>) and $\lambda_B = 500$ MeV (<i>dashed lines</i>). The black line shows their difference, which is independent of λ_B	6
3.1.	The KEKB accelerator complex. The Belle detector was located in the Tsukuba experimental hall. The beams were injected from the LINAC (<i>green</i>). Adapted from [19].	14
3.2.	Schematic side view of the Belle detector. The z -axis points along the opposite direction of the e^+ beam. Adapted from [23].	17
3.3.	Front view of the Belle detector. Adapted from [23].	18
4.1.	Validation plot of the z momentum component of charged tracks, based on a sample of simulated $b \rightarrow c$ decays. The migration of single tracks between two neighboring bins and the different treatment of special values can be observed at the bin around zero and the overflow bin.	21

4.2.	Illustration of signal- and tag-side of a $\Upsilon(4S)$ event. The <i>green ellipse</i> indicates the ROE with respect to the B_{sig} candidate. It contains all final state particles which are not used for the B_{sig} reconstruction.	24
4.3.	Illustration of the bottom-up approach of the FEI. The graph depicts possible connections between the individual stages to reconstruct B_{tag} candidates in different tag channels. Adapted from [39].	25
4.4.	Dependence of the tag-side efficiency and the purity of hadronically reconstructed B_{tag} candidates. Different cuts on the tagging probability of the Belle (FR, <i>blue line</i>) and Belle II (generic FEI, <i>red line</i>) algorithm were applied. Adapted from [39].	26
5.1.	Exemplary event display of a simulated $B^+ \rightarrow e^+ \nu_e \gamma$ event. Reconstructed tracks and ECL clusters are depicted as blue curved lines and red boxes, respectively. The SVD is located in the innermost ring, the middle and outer rings are support structures of the TOP and the ECL.	32
5.2.	Number of tracks with $ \Delta r < 2$ cm and $ \Delta z < 4$ cm per event. Only the background samples are stacked in the histogram. The signal and peaking background samples have on average less tracks, due to only one track on the signal-side.	33
5.3.	Distribution of the tagging probability for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample.	36
5.4.	Event topology of a $B\bar{B}$ (<i>left</i>) and a continuum (<i>right</i>) event. The latter shows a jet-like distribution due to the high momentum of the quark pair. Adapted from [44].	38
5.5.	Application of the continuum suppression BDT after the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ selection. Signal events are largely classified as non-continuum events.	40
5.6.	Composition of the $b \rightarrow u\ell\nu_\ell$ background decays in the signal region of $M_{\text{miss}}^2 \in (-0.5, 0.5)$ GeV ² after continuum suppression and π^0 mass veto.	41
5.7.	Application of the peaking background suppression BDT on an independent MC sample.	42
5.8.	Cut optimization of the continuum and peaking background suppression for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample. The <i>orange point</i> represents the result of the optimization algorithm, while the result of the grid search is indicated by the color gradient. The more blueish the higher the figure of merit. The signal component is scaled to a partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$. The background components are scaled to luminosity.	43

5.9.	M_{miss}^2 distributions of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample after the final selection with a simulated partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$. The background components are scaled to luminosity. Corrections to the MC are applied.	44
5.10.	Calibration factors for individual tag and calibration channels of the specific FEI. With statistical (<i>inner interval</i>) and systematic uncertainties (<i>outer interval</i>). Some tag channels are excluded due to low statistics and are hence not shown. The <i>gray band</i> represents the 1σ uncertainty band of the global calibration factor.	46
5.11.	Calibration factors averaged over the tag channels of the specific FEI shown in Figure 5.10. With statistical (<i>inner interval</i>) and systematic uncertainties (<i>outer interval</i>). For a perfect MC simulation a value of one would be expected. The <i>gray band</i> represents the 1σ uncertainty band of the global calibration factor.	47
5.12.	Applied LID correction factors of the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample.	48
5.13.	Comparison of different form factor models of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay. The <i>black line</i> shows the flawed implementation in EvtGen , which was originally used for Belle MC production. The <i>red line</i> represents the correct model. The latest BCL prediction is shown as <i>green line</i>	49
5.14.	<i>Left:</i> The true q^2 distributions of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ MC samples after signal selection. <i>Right:</i> The applied correction factors used for reweighting between the modified EvtGen and the BCL model.	50
6.1.	Histogram template PDFs used for the signal extraction of the (partial) branching fraction $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ and $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ on the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample. The <i>left (right) column</i> shows the electron (muon) final state.	58
6.2.	Histogram template PDFs of the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample used to constrain the branching fraction measurement of $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$. The <i>left (right) column</i> shows the electron (muon) final state.	60
6.3.	The negative likelihood ratio for Sample 1 of the sample test. The branching fraction was simulated with $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$. A significance of 5.8σ was found. The <i>gray dashed lines</i> indicate the lower and upper statistical uncertainties.	62
6.4.	Result of the sample test for the different sets. The available MC is split into four subsamples, whereof three are used for template generation and the remaining one is fitted. The uncertainty is only statistical. Within the uncertainty all results are in agreement with the simulated branching fractions, indicated by the <i>dashed lines</i>	63

6.5.	Example of a toy study for a simulated (partial) branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$ and $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell) = 7.8 \times 10^{-5}$. One thousand experiments are generated. The <i>left plot</i> shows the pull distribution, the <i>right plot</i> shows the significance of the extracted $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$. The results of a Gaussian fit to the distributions are shown in the upper left corner.	64
6.6.	Linearity test for the fitted $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ in the range of $(0.5, 10.0) \times 10^{-6}$. One thousand toy experiments are generated for each point.	65
6.7.	Linearity test to check the correlation between the fitted $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ and the simulated $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$. One thousand toy experiments are generated for each point. The simulated partial branching fraction is fixed to $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma) = 5 \times 10^{-6}$	65
7.1.	Illustration of the cut on the tagging probability on the off-resonance data. All other cuts are applied except for the cuts on continuum and peaking background suppression.	68
7.2.	Comparison of off-resonance data with on-resonance continuum MC. All cuts are applied except for the cuts on continuum and peaking background suppression. The on-resonance MC is scaled to the luminosity of the off-resonance sample. Only the statistical uncertainty is shown.	69
7.3.	Comparison of off-resonance data with on-resonance continuum MC. All cuts are applied. The on-resonance MC is scaled to the luminosity of the off-resonance sample. Only the statistical uncertainty is shown.	69
7.4.	On-resonance data and MC comparison on the M_{bc} sideband. All other selection cuts are applied. Only the statistical uncertainty is shown.	70
7.5.	Effect of reweighting the M_{bc} sideband MC to the relative fractions of the tag channels on the M_{bc} sideband data.	71
7.6.	Comparison of the relative fractions of the reconstructed tag channels for the M_{bc} sideband on MC and data. Only the statistical uncertainty is shown.	72
7.7.	Resulting M_{miss}^2 distributions for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample after the simultaneous fit on data. The <i>gray component</i> shows all PDFs stacked.	73
8.1.	Resulting M_{miss}^2 distributions for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ sample after the fit on data. The <i>gray component</i> shows all PDFs stacked.	76

8.2.	Resulting M_{miss}^2 distributions for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ sample after the fit on data. The <i>gray component</i> shows all PDFs stacked. The samples are used to constrain the $\mathcal{B}(B^+ \rightarrow \pi^0 \ell^+ \nu_\ell)$ during the fit to the nominal sample.	76
8.3.	Determination of the upper limit. The marked areas correspond to the remaining 10% quantile of the integrated likelihood PDFs for the combined measurement of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$. The <i>red curve</i> does include the systematic uncertainties.	77
8.4.	Two dimensional likelihood scan of the fit on data for the combined measurement. The ellipses correspond to the given confidence level, including only the systematic uncertainty. The <i>red line</i> indicates the upper limit for $\Delta\mathcal{B}(B^+ \rightarrow \ell^+ \nu_\ell \gamma)$ including the systematic uncertainty.	79
8.5.	Comparison of the relative fractions of the reconstructed tag channels on MC and data. Only the statistical uncertainty is shown.	83
8.6.	Effect of reweighting the MC samples to the relative fractions of the tag channels on data.	84
8.7.	Distributions for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ EvtGen model systematic.	86
8.8.	Dependence of λ_B on the measured quantity. The <i>continuous lines</i> correspond to the $ V_{\text{ub}} $ dependent method (<i>left y-axis</i>). The different values for $ V_{\text{ub}} $ can be found in Table 8.7. The <i>red dashed line</i> represents the $ V_{\text{ub}} $ independent method (<i>right y-axis</i>).	87
8.9.	Scan of the partial branching fraction and the ratio for the extraction of $ V_{\text{ub}} $. The <i>gray hatched area</i> indicates the region where no value of λ_B and hence $ V_{\text{ub}} $ can be found numerically.	89
8.10.	Projection of the extraction of λ_B and $ V_{\text{ub}} $ for the expected Belle II data sets. The ellipses correspond to the expected uncertainty.	92
9.1.	World averages of the exclusive and inclusive $ V_{\text{ub}} $ measurements, the projection for Belle II and the projection for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ measurement.	94
A.1.	B2BII validation plots.	97
A.2.	B2BII validation plots.	98
A.3.	B2BII validation plots.	99
B.1.	Continuum suppression variables.	101
B.2.	Continuum suppression variables.	102
B.3.	Continuum suppression variables.	103
B.4.	Continuum suppression variables.	104
B.5.	Continuum suppression variables.	105

B.6. Overtraining check for the trained continuum suppression BDT on test and training data set.	105
B.7. Receiver operating curve for the trained continuum suppression BDT on test and training data set.	106
B.8. FOM and signal efficiency dependent on the cut of the continuum suppression. All final selection cuts are applied, except for the cut on the peaking background. The <i>black line</i> shows the optimal cut found during the optimization with a desired standard deviation of $\sigma = 3$. See Section 5.2.7 for more information.	106
B.9. Peaking background suppression variables.	107
B.10. Overtraining check for the trained peaking background suppression BDT on test and training data set.	108
B.11. Receiver operating curve for the trained peaking background suppression BDT on test and training data set.	109
B.12. FOM and signal efficiency dependent on the cut of the peaking background suppression. All final selection cuts are applied, except for the cut on the continuum background. The <i>black line</i> shows the optimal cut found during the optimization with a desired standard deviation of $\sigma = 3$. See Section 5.2.7 for more information.	109
C.1. Results of the fit to M_{miss}^2 for the different calibration channels. Each calibration channel contains all tag channels.	115

List of Tables

2.1.	Latest BCL expansion parameters from a fit to experimental data of $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ and $B^0 \rightarrow \pi^- \ell^+ \nu_\ell$ decays and theory calculations [11].	8
2.2.	Obtained values for first inverse moment of the LCDA of the B meson using theoretical and experimental methods.	10
2.3.	Overview of previous searches for the $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decay.	11
3.1.	Machine parameters of KEKB at the end of the operation time [23, Chapter 1.3] and the design parameters for SuperKEKB [24].	15
4.1.	Comparison of the reconstruction efficiency between generic and signal-specific FEI, measured on 10^6 signal MC events for both final states. The best-candidate selection and all final cuts were applied, except for the cuts on the FEI probability, continuum and peaking background suppression (see Chapter 5). The numbers include the tag- and signal-side reconstruction.	27
5.1.	Applied weights to scale the used MC samples to luminosity or expected (partial) branching fraction.	31
5.2.	Cuts applied on the event selection stage.	34
5.3.	Cuts applied on the signal-side reconstruction stage.	34
5.4.	Cuts applied for the ROE cleaning.	35
5.5.	Cuts applied on the $\Upsilon(4S)$ reconstruction stage for both samples.	37
5.6.	Signal efficiency (ϵ) and background rejection ($1 - \epsilon$) for the different cuts applied individually and combined on the individual MC samples.	43
5.7.	Branching ratios used for Belle MC simulation and the latest PDG average of the calibration channels [4].	45

5.8.	Branching ratios used for simulation and for reweighting the $b \rightarrow u\ell\nu_\ell$ MC samples [4, 52].	48
5.9.	Background composition of the electron final state in the signal window of $M_{\text{miss}}^2 \in (-0.5, 0.5) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5% are shown.	51
5.10.	Background composition of the muon final state in the signal window of $M_{\text{miss}}^2 \in (-0.5, 0.5) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5% are shown.	53
6.1.	Estimated upper limits on $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma)$ at 90% C.L. for the sample test. The signal MC is scaled to an expected partial branching fraction of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma) = 5 \times 10^{-6}$	62
7.1.	Results of the individual and simultaneous fits for the measurement of $\mathcal{B}(B^+ \rightarrow \pi^0\ell^+\nu_\ell)$ on data. Only the statistical uncertainties are given.	73
8.1.	Results of the fit on data. The first uncertainty is statistical, the second is systematic. The significance does include the full systematic uncertainty. The yields correspond to the $B^+ \rightarrow \ell^+\nu_\ell\gamma$ sample.	77
8.2.	Determined upper limit of $\Delta\mathcal{B}(B^+ \rightarrow \ell^+\nu_\ell\gamma)$ at 90% C.L.	78
8.3.	Events expected on MC and observed on data for the $B^+ \rightarrow \pi^0\ell^+\nu_\ell$ and the $B^+ \rightarrow \ell^+\nu_\ell\gamma$ sample.	78
8.4.	Additive and multiplicative systematic uncertainties for the electron final state.	80
8.5.	Additive and multiplicative systematic uncertainties for the muon final state.	81
8.6.	Additive and multiplicative systematic uncertainties for the simultaneous fit in both final states.	82
8.7.	Calculated lower limits on λ_B for the different approaches. The values for $ V_{\text{ub}} $ are taken from Reference [4]. The limits include the systematic and theoretical uncertainties.	88
8.8.	Comparison to previous results of the partial branching fraction measurement. All limits correspond to the 90% C.L.	89

-
- 8.9. Statistical uncertainties and significance of the improved Belle analysis and for the expected Belle II data set with respect to the extracted partial branching fraction of $\Delta\mathcal{B}(\text{B}^+ \rightarrow \ell^+ \nu_\ell \gamma)_{E_\gamma > 1.0 \text{ GeV}}$. 91
- D.1. Background composition of the **electron final state** in the **full range** of $M_{\text{miss}}^2 \in (-1.5, 3.0) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5% are shown. 117
- D.2. Background composition of the **muon final state** in the **full range** of $M_{\text{miss}}^2 \in (-1.5, 3.0) \text{ GeV}^2$. The first column shows the MC sample and its overall fraction on the background. The second and third column show the contributing B meson decays for the corresponding sample. Only decays with a fraction of more than 1.5% are shown. 119

Bibliography

- [1] A. Heller et al. Search for $B^+ \rightarrow \ell^+ \nu_\ell \gamma$ decays with hadronic tagging using the full belle data sample. *Phys. Rev. D*, 91:112009, Jun 2015. URL <https://link.aps.org/doi/10.1103/PhysRevD.91.112009>.
- [2] B. Aubert et al. Model-independent search for the decay $B^+ \rightarrow \ell^+ \nu_\ell \gamma$. *Phys. Rev. D*, 80:111105, Dec 2009. URL <https://link.aps.org/doi/10.1103/PhysRevD.80.111105>.
- [3] M. Beneke, G. Buchalla, M. Neubert, and C.T. Sachrajda. QCD factorization for exclusive non-leptonic B-meson decays: general arguments and the case of heavy-light final states. *Nuclear Physics B*, 591(1):313 – 418, 2000. URL <http://www.sciencedirect.com/science/article/pii/S0550321300005599>.
- [4] C. Patrignani and Particle Data Group. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, (2016) and 2017 update. URL <http://stacks.iop.org/1674-1137/40/i=10/a=100001>.
- [5] J. P. Ellis. TikZ-Feynman: Feynman diagrams with TikZ. *Computer Physics Communications*, 210:103 – 123, 2017. URL <http://www.sciencedirect.com/science/article/pii/S0010465516302521>.
- [6] M. Beneke and J. Rohrwild. B meson distribution amplitude from $B \rightarrow \gamma \ell \nu$. *The European Physical Journal C*, 71(12):1818, Dec 2011. URL <https://doi.org/10.1140/epjc/s10052-011-1818-8>.
- [7] V.M. Braun and A. Khodjamirian. Soft contribution to $B \rightarrow \gamma \ell \nu$ and the B-meson distribution amplitude. *Physics Letters B*, 718(3):1014 – 1019, 2013. URL <http://www.sciencedirect.com/science/article/pii/S0370269312012130>.

- [8] S. Descotes-Genon and C.T. Sachrajda. Factorization, the light-cone distribution amplitude of the b-meson and the radiative decay $B \rightarrow \gamma \ell \nu_\ell$. *Nuclear Physics B*, 650(1):356 – 390, 2003. URL <http://www.sciencedirect.com/science/article/pii/S0550321302010660>.
- [9] C. Bourrely, L. Lellouch, and I. Caprini. Model-independent description of $B \rightarrow \pi \ell \nu$ decays and a determination of $|V_{ub}|$. *Phys. Rev. D*, 79:013008, Jan 2009. URL <https://link.aps.org/doi/10.1103/PhysRevD.79.013008>.
- [10] C. Bourrely, L. Lellouch, and I. Caprini. Erratum: Model-independent description of $B \rightarrow \pi \ell \nu$ decays and a determination of $|V_{ub}|$ [Phys. Rev. DPRVDAQ1550-7998 79, 013008 (2009)]. *Phys. Rev. D*, 82:099902, Nov 2010. URL <https://link.aps.org/doi/10.1103/PhysRevD.82.099902>.
- [11] Heavy Flavor Averaging Group. $|V_{ub}|$ from exclusive semileptonic B decays. URL <http://www.slac.stanford.edu/xorg/hflav/semi/summer16/html/ExclusiveVub/exclPilnu.html>.
- [12] M. Beneke, T. Huber, and Xin-Qiang Li. NNLO vertex corrections to non-leptonic B decays: Tree amplitudes. *Nuclear Physics B*, 832(1):109 – 151, 2010. URL <http://www.sciencedirect.com/science/article/pii/S0550321310000763>.
- [13] G. Bell and V. Pilipp. $B^- \rightarrow \pi^- \pi^0 / \rho^- \rho^0$ to next-to-next-to-leading order in QCD factorization. *Phys. Rev. D*, 80:054024, Sep 2009. URL <https://link.aps.org/doi/10.1103/PhysRevD.80.054024>.
- [14] V. M. Braun, D. Yu. Ivanov, and G. P. Korchemsky. B-meson distribution amplitude in QCD. *Phys. Rev. D*, 69:034014, Feb 2004. URL <https://link.aps.org/doi/10.1103/PhysRevD.69.034014>.
- [15] T. E. Browder et al. Search for $B \rightarrow \mu \bar{\nu}_\mu \gamma$ and $B \rightarrow e \bar{\nu}_e \gamma$. *Phys. Rev. D*, 56:11–16, Jul 1997. URL <https://link.aps.org/doi/10.1103/PhysRevD.56.11>.
- [16] A. Heller. *Search for $B^+ \rightarrow l^+ \nu \gamma$ decays with hadronic tagging using the full Belle data sample*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2015. URL <https://ekp-invenio.physik.uni-karlsruhe.de/record/48743>.
- [17] Wikipedia contributors. Kexb (accelerator) — Wikipedia, the free encyclopedia, 2018. URL [https://en.wikipedia.org/wiki/KEKB_\(accelerator\)](https://en.wikipedia.org/wiki/KEKB_(accelerator)). [Online; accessed 21-April-2018].
- [18] S. Kurokawa and E. Kikutani. Overview of the KEKB accelerators. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 499(1):1 –

- 7, 2003. URL <http://www.sciencedirect.com/science/article/pii/S0168900202017710>. KEK-B: The KEK B-factory.
- [19] Wikipedia contributors. KEKB (accelerator) — Wikipedia, the free encyclopedia, 2018. URL [https://en.wikipedia.org/wiki/KEKB_\(accelerator\)](https://en.wikipedia.org/wiki/KEKB_(accelerator)). [Online; accessed 21-April-2018].
- [20] K. Egawa et al. Magnet system for the KEKB main ring. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 499(1):24 – 44, 2003. URL <http://www.sciencedirect.com/science/article/pii/S0168900202017722>.
- [21] I. Abe et al. The KEKB injector linac. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 499(1):167 – 190, 2003. URL <http://www.sciencedirect.com/science/article/pii/S0168900202017874>.
- [22] W. A. Barletta. KEK-B: The KEK B-factory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 499(1):v, 2003. URL <http://www.sciencedirect.com/science/article/pii/S0168900202017709>.
- [23] A. J. Bevan et al. The Physics of the B Factories. *The European Physical Journal C*, 74(11):3026, Nov 2014. URL <https://doi.org/10.1140/epjc/s10052-014-3026-9>.
- [24] T. Abe et al. Belle II Technical Design Report, 2010. URL <https://arxiv.org/abs/1011.0352>.
- [25] A. Abashian et al. The Belle detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 479(1):117 – 232, 2002. URL <http://www.sciencedirect.com/science/article/pii/S0168900201020137>.
- [26] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>. Software available from tensorflow.org.
- [27] D. J. Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152 – 155, 2001. URL <http://www.sciencedirect.com/science/article/pii/S0168900201000894>.
- [28] S. Agostinelli et al. Geant4 - a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,*

- Detectors and Associated Equipment*, 506(3):250 – 303, 2003. URL <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [29] M. Gelb. Neutral B Meson Flavor Tagging for Belle II. Master's thesis, Karlsruhe Institute of Technology (KIT), 2015. URL <https://ekp-invenio.physik.uni-karlsruhe.de/record/48719>.
- [30] A. Moll. The Software Framework of the Belle II Experiment. *Journal of Physics: Conference Series*, 331(3), 2011. URL <http://stacks.iop.org/1742-6596/331/i=3/a=032024>.
- [31] C. Pulvermacher. *Analysis Software and Full Event Interpretation for the Belle II Experiment*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2015. URL <https://ekp-invenio.physik.uni-karlsruhe.de/record/48741>.
- [32] D. Y. Kim. The software library of the Belle II experiment. *Nuclear and Particle Physics Proceedings*, 273-275:957 – 962, 2016. URL <http://www.sciencedirect.com/science/article/pii/S2405601415006380>. 37th International Conference on High Energy Physics (ICHEP).
- [33] R. Itoh. BASF - BELLE Analysis Framework. In *Proceedings, 9th International Conference on Computing in High-Energy Physics (CHEP 1997): Berlin, Germany, April 7-11, 1997*, 1997. URL <http://www.ifh.de/CHEP97/paper/244.ps>.
- [34] S. Nagayama. Panther Reference manual. *Belle Note 131 (internal)*, 1996. URL http://belle.kek.jp/secured/belle_note/gn131/panth2.ps.gz.
- [35] T. Keck, M. Gelb, M. Prim, H. Atmacan, J. Gemmler, R. Itoh, B. Kronenbitter, T. Kuhr, M. Lubej, F. Metzner, C. Park, S. Park, C. Pulvermacher, M. Ritter, and A. Zupanc. B2BII Data conversion from Belle to Belle II. *Manuscript submitted for publication*, 2018.
- [36] T. Keck. *Machine learning algorithms for the Belle II experiment and their validation on Belle data*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2017. URL <https://ekp-invenio.physik.uni-karlsruhe.de/record/48940>.
- [37] M. Prim. Study of material effects in track fitting and improvement of the K_S^0 reconstruction at Belle II. Master's thesis, Karlsruhe Institute of Technology (KIT), 2015. URL <https://ekp-invenio.physik.uni-karlsruhe.de/record/48795>.
- [38] T. Keck. The Full Event Interpretation for Belle II. Master's thesis, Karlsruhe Institute of Technology (KIT), 2014. URL <https://ekp-invenio.physik.uni-karlsruhe.de/record/48602>.

- [39] T. Keck, F. Abudinen, F. U. Bernlochner, R. Cheaib, S. Cunliffe, M. Feindt, T. Ferber, M. Gelb, J. Gemmler, P. Goldenzweig, M. Heck, S. Hollitt, J. Kahn, J.-F. Krohn, T. Kuhr, I. Komarov, L. Ligioi, M. Lubej, F. Metzner, M. Prim, C. Pulvermacher, M. Ritter, J. Schwab, W. Sutcliffe, U. Tamponi, F. Tenchini, N. E. Toutounji, P. Urquijo, D. Weyland, and A. Zupanc. The Full Event Interpretation – An exclusive tagging algorithm for the Belle II experiment, 2018. URL <https://arxiv.org/abs/1807.08680>.
- [40] T. Keck. FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment. *Computing and Software for Big Science*, 1(1):2, Sep 2017. URL <https://doi.org/10.1007/s41781-017-0002-8>.
- [41] M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, D. Zander, and A. Zupanc. A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 654(1):432 – 440, 2011. URL <http://www.sciencedirect.com/science/article/pii/S0168900211011193>.
- [42] R. Brun, F. Bruyant, M. Maire, A.C. McPherson, and P. Zancarini. GEANT3. 1987. URL <http://inspirehep.net/record/252007?ln=en>.
- [43] Y.-Y. Charng and H. Li. B meson wave function from the $B \rightarrow \gamma l \nu$ decay. *Phys. Rev. D*, 72:014003, Jul 2005. URL <https://link.aps.org/doi/10.1103/PhysRevD.72.014003>.
- [44] M. Röhrken. *Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment*. PhD thesis, Karlsruhe Institute of Technology (KIT), 2012. URL <https://ekp-invenio.physik.uni-karlsruhe.de/record/48212>.
- [45] G. C. Fox and S. Wolfram. Observables for the analysis of event shapes in e^+e^- annihilation and other processes. *Phys. Rev. Lett.*, 41:1581–1585, Dec 1978. URL <https://link.aps.org/doi/10.1103/PhysRevLett.41.1581>.
- [46] D. M. Asner et al. Search for exclusive charmless hadronic B decays. *Phys. Rev. D*, 53:1039–1050, Feb 1996. URL <https://link.aps.org/doi/10.1103/PhysRevD.53.1039>.
- [47] G. Punzi. Sensitivity of searches for new signals and its optimization. pages 8–11, 2003. URL <http://arxiv.org/abs/physics/0308063>.
- [48] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL <http://www.scipy.org/>.

- [49] J. Schwab. Calibration of the Full Event Interpretation for the Belle and the Belle II Experiment. Master's thesis, Karlsruhe Institute of Technology (KIT), 2017. URL <https://exp-invenio.physik.uni-karlsruhe.de/record/48931>.
- [50] L. Hinz. Lepton ID efficiency correction and systematic error. *Belle Note 954 (internal)*, 2006. URL http://belle.kek.jp/secured/belle_note/gn954/lhinz_bn954.ps.gz.
- [51] The Belle Joint PID group. LID efficiency correction tables, 2009. URL <http://belle.kek.jp/secured/nbb/nbb.html>.
- [52] Y. Amhis. Averages of b-hadron, c-hadron, and τ -lepton properties as of summer 2016. 77(12):895, Dec 2017. URL <https://doi.org/10.1140/epjc/s10052-017-5058-4>.
- [53] P. Ball and R. Zwicky. Improved analysis of $B \rightarrow \pi e \nu$ from QCD sum rules on the light-cone. *Journal of High Energy Physics*, 2001(10):019, 2001. URL <http://stacks.iop.org/1126-6708/2001/i=10/a=019>.
- [54] P. Hamer and A. Frey. Search for $B^0 \rightarrow \pi \tau \nu$. *Belle Note 1328 (internal)*, 2015. URL http://belle.kek.jp/secured/belle_note/gn1328/BN_v1.5.pdf.
- [55] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9(1):60–62, 03 1938. URL <https://doi.org/10.1214/aoms/1177732360>.
- [56] 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*, 88:032005, Aug 2013. URL <https://link.aps.org/doi/10.1103/PhysRevD.88.032005>.
- [57] The Belle Collaboration. Number of B events in Hadron B(J), 2009. URL http://belle.kek.jp/group/pid_joint/lid/final/.
- [58] B. Bhuyan. High p_T Tracking Efficiency Using Partially Reconstructed D^* Decays. *Belle Note 1165 (internal)*, 2010. URL http://belle.kek.jp/secured/belle_note/gn1165/BN1165_v1.pdf.
- [59] M. Paterno. Calculating efficiencies and their uncertainties. 2004. URL <http://inspirehep.net/record/669498?ln=en>.
- [60] G. P. Korchemsky, D. Pirjol, and T.-M. Yan. Radiative leptonic decays of B mesons in QCD. *Phys. Rev. D*, 61:114510, May 2000. URL <https://link.aps.org/doi/10.1103/PhysRevD.61.114510>.

-
- [61] I. Komarov. Status and perspective of Belle II at SuperKEKB, 2018. URL <http://moriond.in2p3.fr/QCD/2018/MondayMorning/Komarov.pdf>. [Online; accessed 13-May-2018].
- [62] J. Albrecht, F. Bernlochner, M. Kenzie, S. Reichert, D. Straub, and A. Tully. Future prospects for exploring present day anomalies in flavour physics measurements with Belle II and LHCb, 2017. URL <https://arxiv.org/abs/1709.10308>.

Danksagung

I have to understand the world, you see.

Richard Feynman - *Surely You're Joking, Mr. Feynman!*

Ich danke Prof. Dr. Michael Feindt für die Übernahme des Referats und für die Möglichkeit diese Arbeit schreiben zu dürfen. Prof. Dr. Florian Bernlochner möchte ich für die Übernahme des Korreferates danken. Er hatte im letzten Jahr der Doktorarbeit einen entscheidenden Einfluss. Die Zusammenarbeit habe ich als sehr angenehm empfunden.

Insbesondere möchte ich mich bei Dr. Pablo Goldenzweig bedanken, der stets bei Fragen hilfreich zur Seite stand. Vielen Dank auch an meine (ehemaligen) Kollegen für die spannende Arbeitsatmosphäre und die interessanten (manchmal auch kontroversen) Diskussionen. Nils Braun, Dr. Thomas Keck, Markus Prim, Felix Metzner, Jochen Gemmler und Judith Schwab haben durch das Probelesen dieser Arbeit zum Gelingen beigetragen.

Die IT-Infrastruktur des Instituts wird allein von den Doktoranden getragen – auch am Wochenende. Vor allem sind hier Dr. Thomas Keck, Nils Braun, Markus Prim, Felix Metzner und Jochen Gemmler hervorzuheben. Auch Frau Bräunling möchte ich danken, sie sorgt immer für einen reibungslosen organisatorischen Ablauf.

Gerdi und Albert Bender danke ich für die (nicht selbstverständliche) Unterstützung.

Zum Schluss möchte ich mich noch bei meiner Frau Ines bedanken, die mich während des Studiums und der Promotion immer unterstützt hat und mir vor allem im letzten Jahr den Rücken frei gehalten hat.

